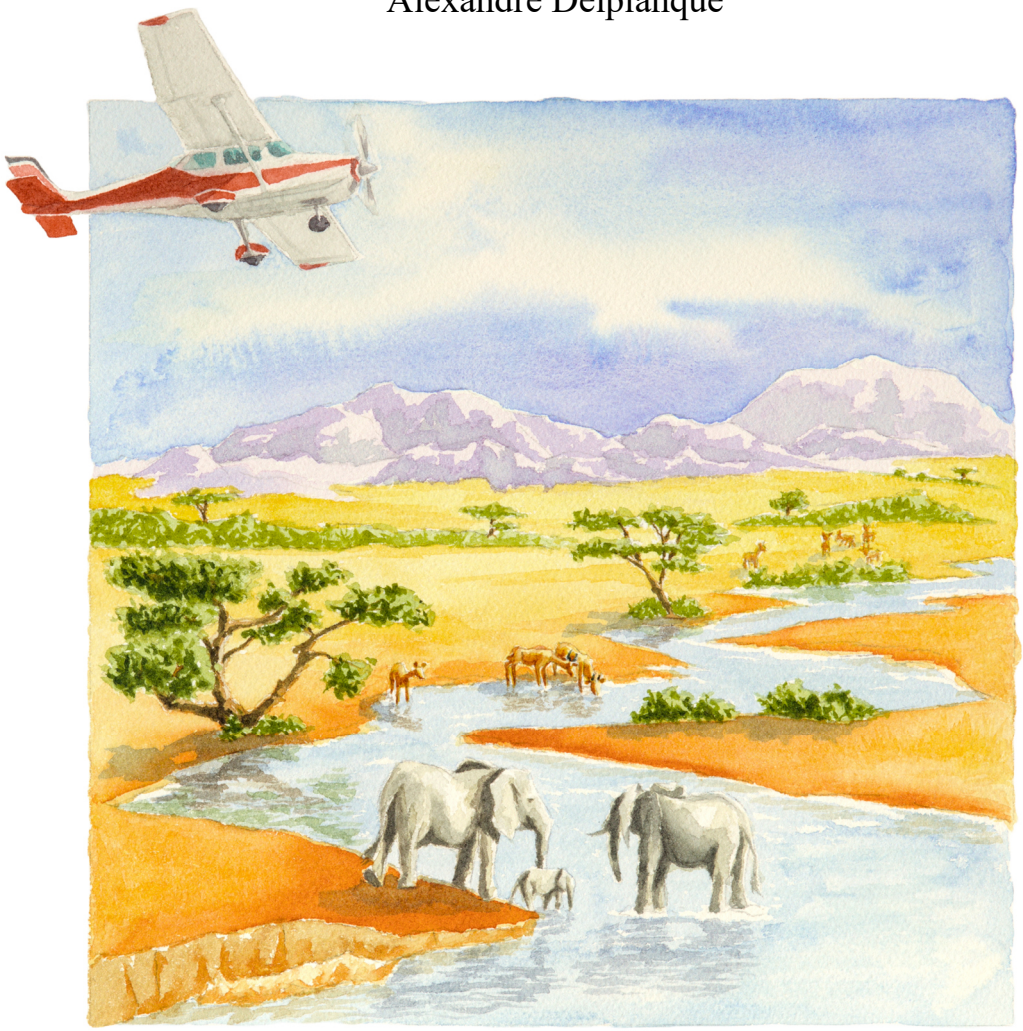
LIÈGE université
Gembloux
Agro-Bio Tech

Integrating remote sensing and deep learning into aerial survey of large African mammals

Alexandre Delplanque



COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

INTEGRATING REMOTE SENSING AND DEEP LEARNING INTO AERIAL SURVEY OF LARGE AFRICAN MAMMALS

Alexandre DELPLANQUE

Dissertation originale présentée en vue de l'obtention du grade de doctorat en sciences agronomiques et ingénierie biologique

Promoteurs : Pr. Philippe LEJEUNE, Pr. Jérôme THÉAU
Année civile : 2024

Abstract

Sub-Saharan Africa is a hotbed of remarkable terrestrial biodiversity, home to a unique diversity of mammals. Unfortunately, this richness is threatened by the growing impact of human activities. While wildlife populations are declining, livestock numbers have been increasing for decades. With significant population growth predicted for the region this century, the pressures on biodiversity are likely to intensify. It is therefore imperative to closely monitor wild and domestic mammal populations. The conventional method of aerial counting using systematic sampling is still widely used to census these populations in open areas. However, the use of on-board photographic sensors on various remote sensing platforms offers the potential to improve and standardize traditional methods. However, processing the large quantities of data generated by these sensors remains a major challenge. In this context, the use of automatic approaches based on deep learning, a branch of artificial intelligence, appears to be a promising solution. The objective of this thesis is therefore to evaluate the effectiveness of the combined use of remote sensing and deep learning for the multi-species census of large African mammals. The research spans several protected areas and considers various mammal species, both wild and domestic.

Firstly, I assessed the potential of pre-existing convolutional neural network architectures to automate the detection and identification of wild species in ultra-high resolution (UHR) images (**Chapter 2**). Three architectures were tested on a dataset comprising six large mammal species. The best model, achieving a mean Average Precision (mAP) of 80%, was applied to an independent dataset from Garamba National Park, Democratic Republic of Congo. It showed superior detection performance to previous studies in similar habitats, opening up promising prospects. However, detection limits were observed for the smallest species (warthog, *Phacochoerus africanus*), and a drop in precision was observed in herd situations for African elephant (*Loxodonta africana*) and buffalo (*Syncerus caffer*), due to a high number of false positives.

Secondly, I developed a novel deep learning architecture named *HerdNet* in response to the limitations of pre-existing models (**Chapter 3**). HerdNet is a point-based object detector inspired by crowd-counting techniques. It has been tested on oblique images of domestic herds of camel (*Camelus dromedarius*), donkey (*Equus asinus*), sheep (*Ovis aries*) and goat (*Capra hircus*) from the Ennedi Natural and Cultural Reserve in Chad. HerdNet demonstrated better detection and counting accuracy than previous methods, on both oblique (+26% of F1 score) and nadir UHR images (+32%). In addition, it solves the problem of false positives in dense herd situations, with proximity-invariant precision. Although species identification could be improved, the practical benefits and potential use of HerdNet were discussed, promising a significant reduction in the human interpretation time associated with aerial surveys.

Thirdly, I evaluated the contribution of oblique UHR imagery and deep learning on systematic aerial sample surveys, in a semi-automatic detection context (**Chapter 4**). I first quantified the reduction in human workload associated with the manual interpretation of oblique images acquired by an on-board camera system on light aircraft. HerdNet was used to detect, count and identify 12 animal species in the Queen Elizabeth Protected Area, Uganda, resulting in a 74% reduction in the number of images to be interpreted by humans. Next, I examined whether a semi-automated approach, incorporating HerdNet, combined with oblique image acquisition, improves the accuracy and precision of population estimates compared with the traditional method. This comparison was carried out for seven key species (e.g. elephant; waterbuck, *Kobus ellipsiprymnus* ssp. *defassa*; western hartebeest, *Alcelaphus buselaphus* ssp. *major*) in Comoé National Park, Côte d'Ivoire, covering 11,500 km². The semi-automatic approach showed significantly higher population estimates for smaller species, i.e. +241% for kob (*Kobus kob* spp. *kob*) and +163% for warthog (*Phacochoerus africanus* ssp. *africanus*), with tighter confidence intervals. However, the obstruction of animals by vegetation had a substantial impact on their detection in the images. Finally, human effort in the semi-automated approach was significantly reduced when compared to fully manual interpretation (estimated at -98%), resolving the major challenge of photographic methods.

In conclusion, this thesis highlights the importance of using remote sensing and deep learning to standardize surveys of large African mammals and efficiently process the growing volume of images generated. Although the approach presented still requires further validation, the results obtained suggest a real potential to revolutionize traditional aerial survey methods. Consequently, the advancement of current aerial survey standards should be considered, as well as the use of other acquisition platforms (e.g. microlight aircrafts), less costly and less challenging to deploy than light aircraft. As for satellites, while recent advancements in image-based ecological monitoring have propelled their potential ahead of other methods, current constraints limit its viability as an immediate alternative. Nevertheless, the use of their images might serve as a valuable complement to organize and deploy other data acquisition platforms, rather than as a standalone survey solution. It is therefore crucial to foster interdisciplinary collaboration to promote these new technological approaches, which will help improve biodiversity monitoring and its long-term preservation.

Résumé

L'Afrique subsaharienne est un foyer de biodiversité terrestre remarquable, abritant une diversité unique de mammifères. Malheureusement, cette richesse est menacée par l'impact grandissant des activités humaines. Alors que les populations d'espèces sauvages déclinent, le nombre de bétail augmente depuis des décennies. Avec la prévision d'une croissance démographique significative dans la région au cours du siècle, les pressions sur la biodiversité risquent de s'intensifier. Il est donc impératif de surveiller de près les populations de mammifères sauvages et domestiques. La méthode conventionnelle de comptage aérien par échantillonnage systématique reste largement utilisée pour inventorier ces populations en milieu ouvert. Cependant, l'utilisation de capteurs photographiques embarqués sur différentes plateformes de télédétection offre un potentiel d'amélioration et de standardisation des méthodes traditionnelles. Pourtant, le traitement de grandes quantités de données générées par ces capteurs reste un défi majeur. Dans ce contexte, l'utilisation d'approches automatiques basées sur l'apprentissage profond, une branche de l'intelligence artificielle, apparaît comme une solution prometteuse. L'objectif de cette thèse est donc d'évaluer l'efficacité de l'utilisation combinée de la télédétection et de l'apprentissage profond pour le recensement multi-espèces des grands mammifères africains. La recherche s'étend sur plusieurs aires protégées et en considère diverses espèces de mammifères, tant sauvages que domestiques.

Premièrement, j'ai évalué le potentiel des architectures préexistantes de réseaux de neurones convolutifs pour automatiser la détection et l'identification des espèces sauvages dans des images à ultra-haute résolution (UHR) (**Chapitre 2**). Trois architectures ont été testées sur un jeu de données comprenant six espèces de grands mammifères. Le meilleur modèle, présentant un mean Average Precision (mAP) de 80%, a été appliqué à un jeu de données indépendant provenant du Parc National de la Garamba, en République Démocratique du Congo. Il a montré des performances de détection supérieures à celles des études précédentes dans des habitats similaires, ouvrant ainsi des perspectives prometteuses. Cependant, des limites de détection ont été observées pour l'espèce de plus faible taille (le phacochère, *Phacochoerus africanus*), et une baisse de précision a été constatée en situation de troupeaux pour les éléphants (*Loxodonta africana*) et buffles d'Afrique (*Syncerus caffer*), en raison d'un nombre élevé de faux positifs.

Deuxièmement, j'ai développé une architecture d'apprentissage profond appelée *HerdNet* en réponse aux limitations des modèles préexistants (**Chapitre 3**). HerdNet est un détecteur d'objets basé sur des points, inspiré des techniques de comptage de foule. Il a été testé sur des images obliques de troupeaux domestiques de dromadaires (*Camelus dromedarius*), ânes (*Equus asinus*), moutons (*Ovis aries*) and chèvres (*Capra hircus*) au sein de la Réserve Naturelle et Culturelle de l'Ennedi au Tchad. HerdNet a démontré une meilleure précision de détection et de comptage par rapport aux méthodes antérieures, tant sur des images à UHR obliques (+26% de F1 score) que nadir (+32%). De plus, il résout le problème des faux positifs en situation de

troupeau dense, avec une précision constante indépendamment de la proximité des individus. Bien que l'identification des espèces puisse être améliorée, les avantages pratiques et le potentiel d'utilisation de HerdNet ont été discutés, promettant une réduction significative du temps d'interprétation humaine associé aux inventaires aériens.

Troisièmement, j'ai évalué la contribution de l'imagerie oblique à UHR et de l'apprentissage profond sur les inventaires aériens par échantillonnage systématique, dans un contexte de détection semi-automatique (**Chapitre 4**). Tout d'abord, j'ai quantifié la réduction de la charge de travail humain liée à l'interprétation manuelle d'images obliques acquises par un système de caméras embarqué sur avion léger. HerdNet a été utilisé pour détecter, compter et identifier 12 espèces d'animaux dans la réserve Queen Elizabeth, en Ouganda, entrainant une réduction de 74% du nombre d'images à interpréter manuellement. Ensuite, j'ai examiné si une approche semi-automatique, intégrant HerdNet, associée à une acquisition d'images obliques, améliore l'exactitude et la précision des estimations de population par rapport à la méthode traditionnelle. Cette comparaison a été effectuée pour sept espèces clés, dont l'éléphant ; le cobe defassa, *Kobus ellipsiprymnus* ssp. *defassa*; le bubale, *Alcelaphus buselaphus* ssp. *major*, dans le Parc national de la Comoé, en Côte d'Ivoire, couvrant 11 500 km². L'approche semi-automatique a montré des estimations de population significativement plus élevées pour les espèces de petite taille, soit +241% pour le cobe de Buffon (*Kobus kob* spp. *kob*) et +163% pour le phacochère (*Phacochoerus africanus* ssp. *africanus*), avec des intervalles de confiance plus étroits. L'obstruction des animaux par la végétation a cependant eu un impact important sur leur détection au sein des images. Enfin, l'effort humain dans l'approche semi-automatique a été significativement réduit par rapport à une interprétation entièrement manuelle (estimé à -98%), résolvant ainsi le défi majeur des méthodes photographiques.

En conclusion, cette thèse met en lumière l'importance de l'utilisation de la télédétection et de l'apprentissage profond pour standardiser les inventaires des grands mammifères africains et traiter efficacement le volume croissant d'images générées. Bien que l'approche présentée nécessite encore des validations supplémentaires, les résultats obtenus suggèrent un véritable potentiel pour révolutionner les méthodes d'inventaire aérien traditionnelles. Par conséquent, la mise à jour des standards actuels d'inventaire aérien devrait être envisagée, ainsi que l'utilisation d'autres plateformes d'acquisition (comme l'ULM), moins coûteuse et moins difficile à mettre en place que les avions légers. Quant aux satellites, bien que les progrès récents en matière de surveillance écologique par imagerie aient propulsé leur potentiel devant d'autres méthodes, les contraintes actuelles limitent leur faisabilité en tant qu'alternative d'inventaire immédiate. Néanmoins, l'usage de leurs images pourrait servir de complément précieux pour organiser et déployer d'autres plateformes d'acquisition de données plutôt qu'une solution d'inventaire à part entière. Il est donc crucial de favoriser la collaboration interdisciplinaire pour promouvoir ces nouvelles approches technologiques, qui contribueront à améliorer la surveillance de la biodiversité et à sa préservation à long terme.

Remerciements

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont contribué à la réalisation de cette thèse. Ce travail représente le fruit de plusieurs années de recherche et d'efforts, et il n'aurait pas été possible sans le soutien et l'aide précieuse de nombreuses personnes. Je souhaite ici les remercier chaleureusement.

Mes plus sincères remerciements vont tout d'abord à mes promoteurs de thèse, Philippe Lejeune et Jérôme Théau, ainsi qu'à Samuel Foucher. Votre réactivité et votre disponibilité ont été de réels catalyseurs pour la réalisation des travaux de cette thèse. Votre expertise, vos conseils avisés et votre soutien constant ont été essentiels à la réussite de ce projet. Votre patience, votre encouragement et votre rigueur scientifique m'ont non seulement aidé à surmonter les défis rencontrés, mais ont également enrichi ma perspective de recherche. Je vous en suis extrêmement reconnaissant.

Je souhaite également remercier les membres du comité d'accompagnement de thèse et les membres du jury, Jérôme Bindelle, Cédric Vermeulen, Hélène Soyeurt, Jean-François Bastin et Devis Tuia. Votre regard critique, vos suggestions constructives et vos commentaires pertinents ont grandement amélioré la qualité de cette thèse. Merci d'avoir accepté de faire partie de cette aventure académique et d'y avoir contribué.

Ensuite, je souhaite remercier toutes les organisations qui ont financé ou collaboré dans le cadre de cette thèse. Particulièrement, le Fond National de la Recherche Scientifique (FNRS) pour avoir cru en mon projet FRIA et financé ces 4 années de doctorat. L'Université de Sherbrooke, pour leur co-promotion, leur collaboration et les opportunités de séjour de recherche au sein de leur établissement. Un immense merci à tous les organismes, gestionnaires de parcs et acteurs de terrain sans qui cette recherche n'aurait pas été possible : African Parks Network, la Réserve Naturelle et Culturelle de l'Ennedi, l'Institut Congolais pour la Conservation de la Nature, le Parc National de la Garamba, le Parc National des Virunga, l'Uganda Conservation Foundation, l'Office Ivoirien des Parcs et Réserves, le Parc national de la Comoé, et Aviation Sans Frontières Belgique.

Un grand merci à toutes les personnes ayant participé de près ou de loin aux nombreux échanges sur le sujet. Vos idées, discussions et retours ont été une source d'inspiration et ont permis d'affiner et d'approfondir mes réflexions. J'aimerais particulièrement remercier Elsa Bussière, Cyril Pélissier, Richard Lamprey, Howard Frederick, Xavier Vincke, Simon Lhoest, Davy Fonteyn, Julie Linchant, et Cédric Vermeulen. Vos contributions ont enrichi ce travail et je vous en suis très reconnaissant.

Merci à tous les collègues de Gembloux pour toutes nos discussions intéressantes et nos moments passés ensemble : Nicolas, Adrien, Jérôme, Romain, Corentin, Arthur, Jo, Marjane, Edwin, Violette, Noé, Justin, Modestine, Gauthier, Marie, Alain, Marie-

Ange, Charlotte, Chloé, Simon T., Alexandre, Capucine, Marie-Pierre, Thierry, Anaïs Simon L., Sarah, Davy, Morgane, Robin et Marius. Merci aussi à tous les collègues de Sherbrooke pour votre accueil et pour les bons moments partagés ensemble : Gafarou, Arifou, Ghazaleh, Simon, Bhanu, Dieu, Ali, Swarna, Helena et Nicolas.

J'aimerais également remercier toutes les personnes qui m'ont accueilli très chaleureusement chez eux ou avec qui j'ai pu vivre quelques temps lors de mes séjours au Québec, je pense particulièrement à José-Normand, Mylen et Tony, Dany, Marc-Alexandre, Samuel et Elaine.

Merci à mes frères, Julien et Nicolas, pour votre aide, vos suggestions et vos conseils d'informaticien. Merci à mes amis, particulièrement Antoine, Nicolas C., Guillaume, Nicolas V., Cyrille et Thomas. Merci pour votre écoute, nos discussions et nos célébrations. Enfin, merci Coline d'être là tous les jours. Ton amour, tes encouragements, et ton humour m'ont permis de tenir le cap jusqu'au bout.

Je voudrais terminer en remerciant du fond du cœur mes parents, Judith et Frédéric, pour leur soutien inconditionnel. Votre amour, votre compréhension et vos encouragements ont été des piliers sur lesquels j'ai pu m'appuyer, et ce, depuis toujours. Merci de m'avoir soutenu sans faille et de m'avoir toujours encouragé à poursuivre mes rêves. Je suis vraiment très fier de vous avoir comme parents.

Table of contents

List of figures

List of tables

List of acronyms

AED	Aerial Elephant Dataset
AI	Artificial Intelligence
ANN	Artificial Neural Network
AP	Average Precision
AVGAS	Aviation Gasoline
CCNN	Counting CNN
CITES	Convention on International Trade of Endangered Species
CNN	Convolutional Neural Network
CNP	Comoé National Park
COCO	Common Object in Context
CPU	Central Processing Unit
DEM	Digital Elevation Model
DL	Deep Learning
DLA	Deep Layer Aggregation
DRC	Democratic Republic of Congo
EBVs	Essential Biodiversity Variables
ENCR	Ennedi Natural and Cultural Reserve
ENI	Even-Number Image
FIDT	Focal Inverse Distance Transform
FM	Foundational/Foundation Model
FN	False Negative
FP	False Positive
FPN	Feature Pyramid Network
FSO	Front-Seat Observer
GE	GeoEye
GEO BON	Group on Earth Observations Biodiversity Observation Network
GIS	Geographic Information Systems
GNSS	Global Navigation Satellite System
GPU	Graphic Processing Unit
GSD	Ground Sampling Distance
HNP	Hard Negative Patch
HOG	Histogram of Gradients
IMU	Inertial Measurement Unit
IoU	Intersect over Union
LiDAR	Light Detection and Ranging
LMDS	Local Maxima Detection Strategy

MAE	Mean Absolute Error
mAP	Mean Average Precision
MIKE	Monitoring the Illegal Killing of Elephants
ML	Machine Learning
MLP	Multilayer Perceptron
MOGAS	Motor Gasoline
MP	Megapixel
MST	Minimum Spanning Tree
NE	North-East
NMS	Non-Maximum Supression
NW	North-West
OCC	Oblique Camera Count
ONI	Odd-Number Image
PA	Protected Area
PAEAS	Pan-African Elephant Aerial Survey
QB	QuickBird
QEPA	Queen Elizabeth Protected Area
RaDAR	Radio Detection and Ranging
RCNN	Region-based CNN
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
RoI	Regions of Interest
RPN	Region Proposal Network
RSO	Rear-Seat Observer
SADL	Semi-Automated Deep Learning
SAL	Semi-Automatic Loop
SE	South-East
SIFT	Scale-Invariant Feature Transform
SRF	Systematic Reconnaissance Flight
SRTM	Shuttle Radar Topography Mission
SSIM	Structural Similarity Index
SURF	Speed Up Robust Features
SVM	Support Vector Machine
SW	South-West
TP	True Positive
UAV	Unmanned/Unpiloted Aerial/Aircraft Vehicle
UHR	Ultra-High Resolution
UK	United Kingdom

VHR	Very-High Resolution
WV	WorldView

1

General introduction

1. Biodiversity loss in an anthropogenic world

Biodiversity (or biological diversity) was defined in the text of the Convention on Biological Diversity as "*the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems*" (CBD, 2011). Biodiversity is a fundamental component of the planet's ecological balance, providing a range of ecosystem services crucial for human well-being (Cardinale et al., 2012; Díaz et al., 2019). However, the global rate of biodiversity loss has accelerated significantly over the past century, raising concerns about its profound implications for ecosystem stability, resilience, and functioning (Ceballos and Ehrlich, 2023; Dirzo et al., 2014).

Number of biologists and ecologists agreed to say that we have entered into a human-driven sixth major episode of biodiversity extinction since life on earth arose (Cowie et al., 2022). Causes of biodiversity loss are complex, encompassing both indirect and direct anthropogenic drivers of change. Indirect drivers are underpinned by societal values and behaviors, such as demography, economy or conflicts, leading to direct drivers related to human consumption, such as habitat destruction/conversion, climate change or pollution (Ceballos and Ehrlich, 2023; Díaz et al., 2019; Foley et al., 2005; Johnson et al., 2017; Weiskopf et al., 2020). The loss of biodiversity has far-reaching consequences for ecosystem functioning and the services they provide. Biodiversity loss can lead to decreased ecosystem stability, reduced resistance to disturbances, and diminished resilience in the face of environmental changes (Hooper et al., 2005). Human activities have led to the extinction of numerous species, with current genera extinction rates estimated to be 354 times higher than the background rate for vertebrates (Ceballos and Ehrlich, 2023). Moreover, climate change further exacerbates the vulnerability of many species and ecosystems (Parmesan, 2006; Weiskopf et al., 2020).

Most land vertebrate genera and monotypic ones are concentrated in tropical and subtropical regions of the Earth, where extinct and endangered species were unfortunately mainly observed (Ceballos and Ehrlich, 2023). Among these regions, sub-Saharan Africa open and semi-open ecosystems show one of the highest mammal species richness with multiple richness hotspots (Ceballos and Ehrlich, 2006), containing some of Africa's iconic species, like the African elephant (*Loxodonta africana*), the white rhinoceros (*Ceratotherium simum*) or the northern giraffe (*Giraffa camelopardalis*). While large wild mammals populations are declining since 1970 (Craigie et al., 2010), livestock population densities have dramatically increased in sub-Saharan Africa during the last decades, following rapid growth of human population and effective sanitary actions on herds (Richard et al., 2019). Eastern Africa has the highest density of livestock and pastoralists in Africa, occupying about two-third of the area (Jenet et al., 2016). Excessive livestock density may have negative effects on ecosystems, such as degradation, resource competition with wildlife and disease spreading (Bengis et al., 2004; Butt and Turner, 2012; De Leeuw

et al., 2001; Georgiadis et al., 2007; Scholte et al., 2022a, 2022b; Vandermeer, 2002), but livestock is also a major source of income for rural populations (Herrero et al., 2013; Jenet et al., 2016).

As a result, conservation actions are essential to avoid further biodiversity loss. An important initiative for biodiversity conservation has been the establishment of protected areas (PAs), which aim to preserve ecological sanctuaries while respecting local communities (United Nations Environment Programme, 1992). PAs are defined by the International Union for Conservation of Nature (IUCN) as: "*clearly defined geographical spaces, recognized, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values*" (Stolton et al., 2013). There are several categories of PAs, each with its own management objectives, features and role in the landscape. These categories are seen as an important global standard for the planning and management of PAs, which facilitate data collection, help reporting on conservation efforts and facilitate comparison between countries (Stolton et al., 2013). As PAs cover 14% of Africa's land (UNEP-WCMC and IUCN, 2024), their effective management is crucial for biodiversity conservation (Riggio et al., 2019; Watson et al., 2018). Even more so considering the impacts of the expected strong growth of sub-Saharan demography during this century (Ezeh et al., 2020) which will represent a real threat for African wildlife by exacerbating current causes of biodiversity loss. The future of PAs, in an increasingly crowded world, may well lie in our ability to maintain sustainable use of natural resources.

2. Vital contribution of aerial surveys for wildlife conservation

Establishing PAs alone is not sufficient to preserve the biodiversity; there is a crucial need for continuous biodiversity monitoring within these regions. Effective biodiversity monitoring should allow for the assessment of conservation strategies, help in identifying emerging threats, and ensure management practices are implemented (Jachmann, 2001; Stolton et al., 2013). By systematically tracking biodiversity variables, such as wildlife abundance, livestock densities or human illegal activities, conservationists can make informed decisions to protect and restore ecosystems. Such information is usually obtained through ecological monitoring programs integrated into an adaptive management process during which system states and variables are estimated through time with repeated sequences of 'monitoring - assessment - decision making' (Nichols and Williams, 2006). In order to provide an harmonized observation system that could form the basis of monitoring programs worldwide, the Essential Biodiversity Variables (EBVs) framework was developed by the Group on Earth Observations Biodiversity Observation Network (GEO BON) in 2013 (Pereira et al., 2013). In the EBVs, species population data recorded in a standardized and systematic way are seen as essential for biodiversity monitoring (Brummitt et al., 2017; Jetz et al., 2019) and are further actively encouraged by the IUCN within the framework of PA management (Stolton et al., 2013).

In the context of a wildlife conservation management objective, regular and periodic counts need to be carried out to obtain at least precise species population estimates, used as an indicator of the state of the system and as a feedback of conservation actions (Jachmann, 2001). For large African mammals, multiple counting methods exist for estimating their populations, depending on the area to survey and the available financial, logistical and human resources (Jachmann, 2001; Norton-Griffiths, 1978). Some use indirect methods, such as the count of indicators of presence, like dung (Barnes et al., 1997) or carcasses (Chase et al., 2016). Others use direct counting methods, using on-sight observations while moving on foot (Waltert et al., 2008), terrestrial vehicle (Ogutu et al., 2006) or aircraft (Schlossberg et al., 2016); or digital observations obtained from sensors like camera traps (Fonteyn et al., 2021), drones (Linchant et al., 2018; Vermeulen et al., 2013), or microphones (Thompson et al., 2010).

For many decades, the most commonly used technique for estimating large mammal populations in African open and semi-open areas is the traditional systematic aerial sample count, using piloted light aircraft and human observers (Grimsdell and Westley, 1981; Jachmann, 2001; Norton-Griffiths, 1978). In this thesis, the term *aerial survey* refers to this technique. The next subsections present the principle of the technique, its source of bias and challenges, and the established standards.

2.1. Systematic aerial sample counting: Principles and practicalities

In this subsection, the main principles of systematic aerial sample counting and its practicalities are described to provide a basic understanding of the method. Extensive details are provided in the reference book of Norton-Griffiths (1978), on which the following text is mainly based. First and foremost, it is important to discern and define the terms *census* and *survey*. The term *census* refers to the entire area, in which the number of animals need to be estimated. During a census, also called a *total count*, the total number of animals is counted. As for the term *survey*, it refers to the use of sample units from which animal counts are extrapolated to cover the census area. A survey is equivalent to a *sample count*. It is therefore based on sample units dispersed randomly or systematically over the census area, the systematic transect scheme being the most popular and commonly used method for covering vast and open areas in Africa.

In systematic aerial sample counting, the transects represent the sample units and are parallel, equidistant, and usually of the same strip width to minimize associated counting bias. Since animals tend to be distributed along major rivers, transects need to be oriented perpendicularly to a baseline that follows the main river (**Figure 1.1a**). In case of vast census areas or uneven animal density, it may be preferred to divide the area into *strata* to obtain subareas of manageable size and/or to reduce the variance of sample units. However, this method may entail constructing a different baseline within each stratum and implies prior information on animal density. The proportion

of the census area to be surveyed, called the *sampling effort*, depends on the precision of population estimates desired whilst being often combined with other factors (e.g. costs and logistics). An increased sampling effort implies a larger sample size, i.e. more units need to be counted, but provide a lower sample error and thus more precise estimates. Nevertheless, the relation between sample size and sample error is not linear; sample error tends to decay in strength with increased sample size. Consequently, from a certain point, the effort and money to be invested in the survey for the slight increase in precision of the population estimates becomes meaningless. Knowing an estimate of the sample variance (from a previous survey for instance), it is then possible to plot this curve and determine the sample size required to achieve the desired precision. However, this information is not always available. In such cases, it is preferable to have a sampling effort around a common value of 15-20% (Norton-Griffiths, 1978) but which may be adapted according to the size of the area (CITES-MIKE, 2020).

On the logistical and practical side, systematic aerial surveys are usually carried out during the dry season, to avoid thick vegetation cover and thus minimize counting bias, and when weather conditions are favorable for flying. Generally, a four- or six-seat light aircraft, an experienced pilot and at least 2 rear-seat human observers are needed. The latter are instructed to count predefined species, usually medium and large mammals, between two streamers attached to the wing struts on each side of the aircraft (**Figure 1.1b**). These streamers theoretically define the desired strip width on the ground at a fixed flight altitude above ground level, following a crucial calibration step. It should be noted that altitude and strip width have an impact on the ability of observers to count animals, with wide strips being more prone to counting bias than narrow ones. In a multispecies survey in open areas, common values are 92 meters (300 feet) for the flight altitude and 200 meters for the strip width (**Figure 1.1c**). During the flight, observers must at least record the species name, the count, the side and the geographic position of the observations they make between the streamers, any additional information being useful (e.g. out-of-strip animals, male:female ratio, proportions of calves, yearlings and sub-adults). It is a common practice to equip observers with single-lens cameras to photograph any group of animal larger than 10 (CITES-MIKE, 2020; Grimsdell and Westley, 1981; Norton-Griffiths, 1978), for later count correction to minimize counting bias. Moreover, it is not rare for an additional front-seat observer (FSO) to join the flight to assist the rear-seat observers (RSOs). The FSO may provide valuable assistance by announcing incoming groups and recording count and position data.

To obtain population estimates and associated standard errors across the census area, the resulting observations are processed using the Jolly's method for unequal sized sampling units (Jolly, 1969). It is also known as the *ratio method* as it is based on the density of animals per sample. It integrates the ratio of animals counted to the total sample unit areas, the variances between animals counted and between sample unit areas, and the covariance between animals counted and the area of each unit. Additionally, 95% confidence limits are usually further calculated (Norton-Griffiths,

1978), simply by multiplying the population standard error by 1.96, for a number of sample units greater than 30. The results of the Jolly's method are very informative as they obviously provide information about population sizes and distribution, but the data could also serve for designing better census.

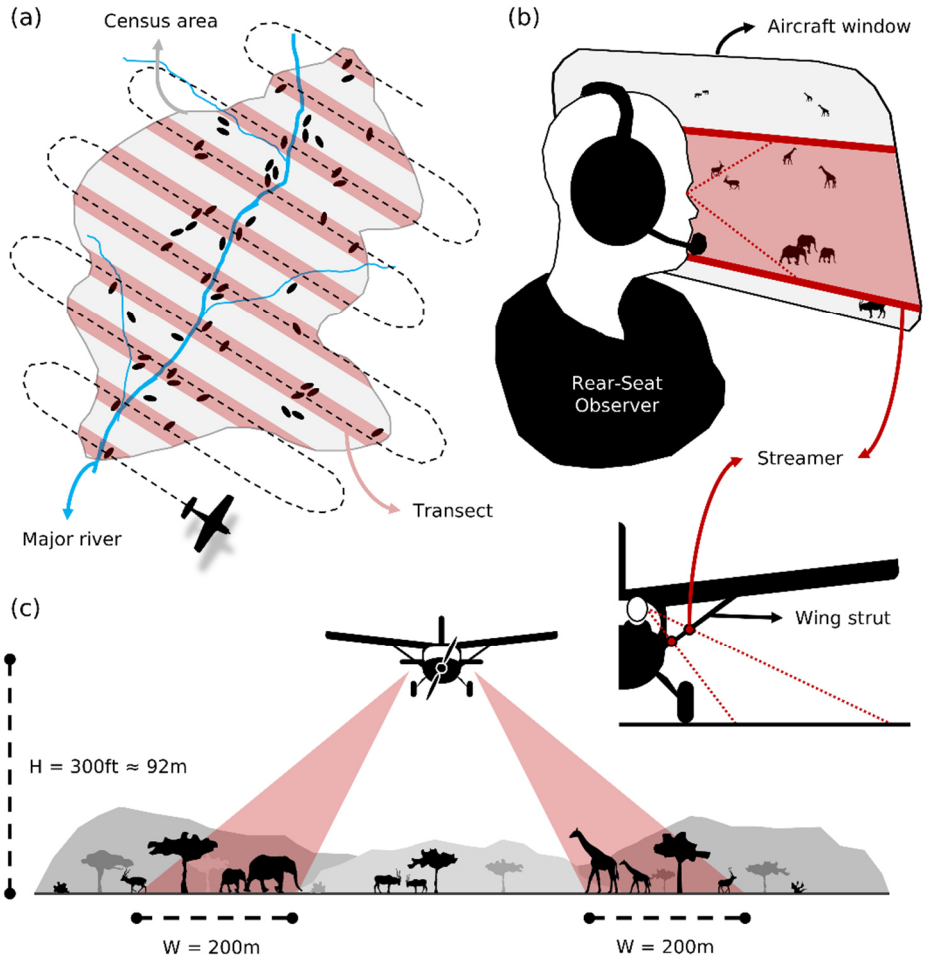


Figure 1.1: Principles of a systematic aerial sample count: (a) Typical parallel flight plan with transects perpendicular to the major river, (b) Rear-seat observer (RSO) visual counting in the sample strip delimited by two streamers, (c) View of the ground strip widths in which the animals are counted by left and right RSOs.

2.2. Source of bias and challenges

Aerial surveys suffer from multiple errors and biases, the latter being errors in a systematic direction. They should be anticipated and minimized, or measurable to be corrected. Biases may originate from survey and environmental factors, both

influencing the most potent source of bias: the observer. Main survey factors include altitude, flying speed, sample strip width or observer experience, positioning in the aircraft and fatigue (Beasom et al., 1981; Caughley, 1974; Jachmann, 2001; Norton-Griffiths, 1978, 1976; Pennycuick and Western, 1972; Schlossberg et al., 2016). Generally, the higher and the faster the aircraft is flying, or the wider the sample strip width, the more animals are missed by the observers. This is intrinsically linked to the human ability to detect animals on the move. Environmental factors cover, among other, the season during which the survey is carried out, time of day, animal size, color and behavior, group size, vegetation type, thickness, and density (Fleming et al., 2008; Goodenough et al., 2024; Griffin et al., 2013; Jachmann, 2002; Norton-Griffiths, 1978; Schlossberg et al., 2016; Wal et al., 2011). These factors influence the visibility and behavior of animals in their habitat and may therefore further complicate the work of observers. Nevertheless, most of these bias factors can be mitigated when planning the survey period and designing the flight plan.

Conventional aerial surveys thus highly depend on crewed aircraft, which poses a whole range of associated challenges. The involvement of human pilots and observers in aerial survey missions exposes individuals to potential risks. The very low flight altitude represents a real life-threatening risk for the flying crew, as the slightest error can be fatal. Aviation accidents were indeed the most significant cause of deaths for wildlife biologists in the United States during the last century (Sasse, 2003). In addition, the fluctuating socio-political context around many PAs in Africa may provoke conflicts, thus increasing the risks for the flying crew. The latter may represent a vulnerable target for armed fighters hidden or inhabiting these areas, often engaging in illegal activities for money (Mulero-Pázmány et al., 2014). Aircraft requirements are also challenging. Four- to six-seat fixed wing aircrafts require appropriate fuel, maintenance and qualified pilots (Bouché et al., 2012), which represent a substantial part of a survey budget (Grimsdell and Westley, 1981). An experienced pilot with a particular interest for wildlife conservation is often recommended (Norton-Griffiths, 1978). He should be able to maintain a stable flight altitude to avoid any sample bias, but this is a hard task that further accentuates the risks. Due to high costs engaged with aerial survey, from material acquisition to salaries, financial support from donors is usually required, making frequent monitoring missions hard to sustain in the long-term (Bouché et al., 2012; Dunham, 2012; Watts et al., 2010). Another challenge is to dispose of trained observers which can provide reliable counts during the flights (Norton-Griffiths, 1978). Counting animals on the move for a substantial amount of time is a hard task. Observers may then under- or overcount large herds, miss or confuse species, or lose attention during long flights (Bouché et al., 2012; Fleming et al., 2008; Grimsdell and Westley, 1981; Jachmann, 2001; Norton-Griffiths, 1976). For these reasons, the use of cameras has been indispensable for decades (Norton-Griffiths, 1974), but continuous image acquisition along transects has received little attention because it generates too much analysis work. It is indeed a time-consuming exercise which may generate

considerable costs, making the approach too expensive for a broader use at the moment (Bröker et al., 2019; Lamprey et al., 2020b).

2.3. Standards and guidelines

In order to produce comparable and uniform results, standards and guidelines were established in 2012 by the Monitoring the Illegal Killing of Elephants (MIKE) system, under the supervision of the Convention on International Trade of Endangered Species (CITES) (Craig, 2012). The document provides suggestions for choice of method and survey design (i.e. guidelines) and minimum requirements for funding contracts (i.e. standards). Initially, these standards were established as a basis for surveying and monitoring elephant trends, but they can and are being used for other species too. The document was discussed and supplemented during the 2014 Pan-African Elephant Aerial Survey (PAEAS) meeting (PAEAS, 2014) and revised in 2019 by the CITES-MIKE program (CITES-MIKE, 2020). It details all the required equipment, the expectations of FSO and RSOs, the parameters to consider when implementing a survey, the observations to record, data analysis, data archiving and advice based on past experience. These standards are a true reference guide in the field, particularly important for guaranteeing the quality of an aerial survey.

3. Remote sensing imagery for improved wildlife monitoring

As detailed in Section 2.2, the most potent source of bias in an aerial survey is counting bias, inherent in a human observer's ability to count animals at a distance while moving. The use of on-board cameras thus rapidly proved indispensable for correcting the observers' animal counts (Grimsdell and Westley, 1981; Norton-Griffiths, 1978, 1974) and is even included in the standards (CITES-MIKE, 2020; PAEAS, 2014), showing that remote sensing has been an integral part of aerial surveys for many decades. Moreover, for several years now, alternatives based on other remote sensing platforms have been proposed to address the risks and challenges associated with traditional aerial surveys, such as the use of drones (Jiménez López and Mulero-Pázmány, 2019; Linchant et al., 2015b; Linchant, 2021) or high-resolution satellites (Sánchez-Díaz and Mata-Zayas, 2019; Wang et al., 2019). Remote sensing is seen as a great tool for ecology and conservation, enabling EBVs to be characterized on a global scale, at high frequency, and in a uniform way (Jetz et al., 2019; Sánchez-Díaz and Mata-Zayas, 2019; Turner et al., 2015). This section aims to define remote sensing, present its different systems and platforms, its potential for wildlife survey and the associated challenges.

3.1. Definition, systems and platforms

The term *remote sensing* gathers all the techniques that involve the collection and interpretation of information about the Earth's surface or atmosphere without direct physical contact. These techniques therefore use various sensors and platforms to acquire data from a distance. The information provided in this section is mainly based

on the books of Campbell and Wynne (2011) and Lillesand et al. (2015). There are several types of remote sensing systems, each designed to capture specific wavelengths of electromagnetic radiation. Electromagnetic radiation is the energy that travels through space in the form of waves, including visible light, infrared radiation, or microwave radiation. These systems can be broadly categorized into active and passive remote sensing (**Figure 1.2a**). Passive remote sensing relies on naturally occurring radiation, such as sunlight radiation, to illuminate the target area. Passive sensors thus detect and record the reflected (e.g. visible) or emitted (e.g. thermal infrared) energy from the Earth's surface. An example of such a sensor is an optical camera, which captures and records visible light to generate images. Active remote sensing involves the emission of radiation by a sensor, which then measures the reflected or scattered signals. Radar (Radio Detection and Ranging) is a typical example of an active remote sensing system, employing microwave frequencies to determine surface characteristics. Another active system worth mentioning is LiDAR (Light Detection and Ranging) that employs laser beams to measure distances and generate detailed topographic information.

Remote sensing platforms refer to the structures that carry the sensors, facilitating data collection from the target area. These platforms can be classified into ground-based, aerial and space-based systems. Ground-based remote sensing involves deploying sensors on the Earth's surface to collect data at a local scale, the most common platforms being hand-held devices, tripods, towers and cranes. Aerial platforms, such as aircraft or drones, operate at higher altitudes and are suitable for high-resolution imaging. Space-based platforms, notably satellites, orbit the Earth at various high altitudes, offering global coverage and the ability to monitor large-scale environments, but often at the expense of lower-resolution imaging (**Figure 1.2a**).

Most of the sensors now produce digital photography/images that can be numerically interpreted, through the use of a two-dimensional array of silicon semiconductor detectors, composed of billions of photosites. The latter convert the electromagnetic radiation into an analog signal being finally digitized and processed to obtain a two-dimensional array of cells (i.e. pixels) with numeric values, i.e. the digital image. The different wavelengths of the receiving radiation may be separated by specific filters (e.g. red, green, and blue for true color images) to obtain images with multiple bands, i.e. multiple two-dimensional arrays of numeric values, each corresponding to a specific range of the electromagnetic spectrum (**Figure 1.2b**). For remote sensing applications, it is usually important to know the size of one pixel on the ground, to assess the spatial resolution of the acquired image. This is given by the Ground Sampling Distance (GSD), representing the distance between the centers of two adjacent pixels in the image, measured in SI units, such as meters or centimeters. GSD is a crucial parameter in remote sensing and digital image interpretation, as it directly influences the level of detail in the imagery. It is determined and influenced by the remote sensing system characteristics, such as the sensor size and the altitude of the platform. A smaller GSD indicates higher spatial resolution, allowing for more detailed features to be captured in the image.

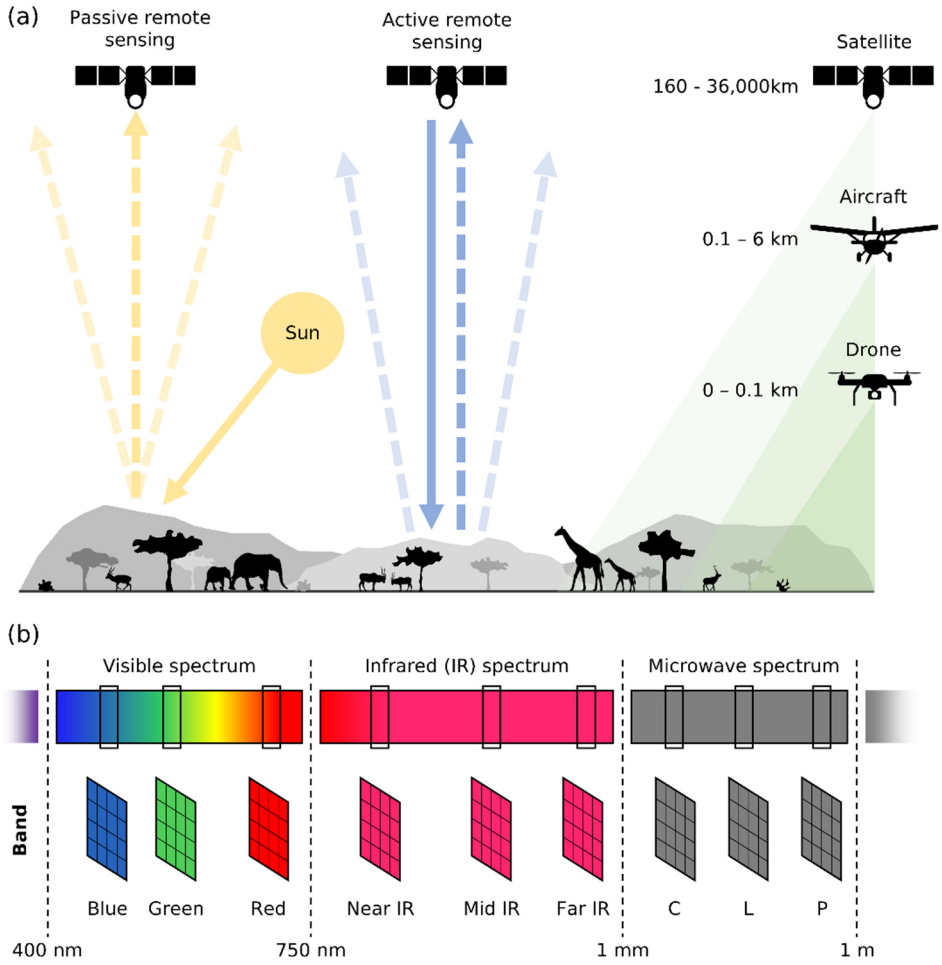


Figure 1.2: Basic concepts of remote sensing: (a) Passive system, active system, and typical altitude range of the aerial platforms, (b) Usual spectra and resulting image bands.

3.2. Potential for large mammal survey

In recent years, advancements in remote sensing technologies, particularly through the utilization of drones and very-high resolution (VHR) satellites (i.e. producing image with sub-meter pixel size), showed great prospects and increasing attention for revolutionizing large mammal survey methodologies (Chabot, 2018; Hodgson et al., 2018; Kuenzer et al., 2014; LaRue et al., 2017; Linchant et al., 2015b; Linchant, 2021; Ocholla et al., 2024; Pettorelli et al., 2014; Vermeulen et al., 2013; Wang et al., 2019). These platforms offer opportunities for monitoring and studying wildlife populations (LaRue et al., 2021; LaRue and Stapleton, 2018; Linchant et al., 2018), habitats (Fretwell and Trathan, 2009; Olsoy et al., 2018; Swinbourne et al., 2018; Wang et al.,

2020), and behaviors (Petso et al., 2021; Wu et al., 2023). In addition, they offer the advantage of accessing remote or challenging terrains that may be difficult or dangerous for researchers to reach (Barber-Meyer et al., 2007; Christie et al., 2016; Krause and Hinke, 2021; Stapleton et al., 2014). This enhances the efficiency and safety of wildlife surveys, especially in areas with rugged topography or dense vegetation. The non-intrusive nature of drones and satellites minimizes disturbance to wildlife during surveys (Christie et al., 2016; Ivosevic et al., 2017; Wang et al., 2019), which may be crucial when studying sensitive or elusive species, as traditional survey methods might cause stress or alter natural behaviors (Efroymson and Suter II, 2001). As regards aircraft platform, the use of on-board cameras taking continuous oblique images has proved particularly effective in producing more precise and accurate population estimates of terrestrial mammals in eastern Africa (Lamprey et al., 2020b, 2020a), and Australia (Lethbridge et al., 2019)

Remote sensing sensors consistently provide standardized data records, ensuring a reliable and reproducible source for wildlife surveys, that can be properly archived and further reviewed or certified by other parties (Lamprey et al., 2020b; Linchant et al., 2015b). In addition, regular and standardized remote sensing imagery allows researchers to track changes in habitats (Wilschut et al., 2018) and species over time (LaRue et al., 2011; Naveen et al., 2012), thus participating in better EBVs monitoring (Brummitt et al., 2017; Jetz et al., 2019; Turner et al., 2015).

3.3. Challenges

While remote sensing shows great potential for wildlife survey, it also faces a series of challenges in this context. The effectiveness of remote sensing may be impacted by adverse weather conditions such as cloud cover, precipitation, and wind. While cloud cover has little impact on aerial platforms, i.e. aircraft or drones (Anderson and Gaston, 2013), it is particularly restrictive for satellite platforms with sensors intercepting the visible spectrum (Asner, 2001; LaRue et al., 2011; LaRue and Stapleton, 2018). Conversely, wind has no impact on satellite acquisitions, has limited impact on drone acquisitions (Goebel et al., 2015; Hodgson et al., 2013), whereas it increases flight risks for crewed aircraft (Sasse, 2003; Watts et al., 2010). The acquisition of VHR remote sensing imagery may incur substantial costs, especially when a large area needs to be covered. It is complicated to assess whether aerial surveys with on-board cameras are more costly than a satellite approach, as this depends on many factors such as the aircraft availability, size of the study area, the sample effort (Norton-Griffiths, 1978) or the human effort required to analyze the images (Lamprey et al., 2020b). Furthermore, aerial surveys may provide much finer and more precise information than satellite imagery, and enable species to be identified (Lamprey et al., 2020b), which is not necessarily yet the case with VHR satellite imagery (Wu et al., 2023). As for the drone, it is presented as a less logistical and less expensive approach (Anderson and Gaston, 2013; Christie et al., 2016; Linchant et al., 2015b; Wang et al., 2019). However, its low endurance remains an obstacle for surveys of large PAs (Christie et al., 2016; Linchant, 2021).

Depending on the platform and sensor, images may be produced at different temporal and spatial resolutions. In general, drones may provide images with ultra-high spatial (GSD < 10 cm) and temporal resolution, with the GSD and acquisition frequency defined by the user (Anderson and Gaston, 2013; Chabot and Bird, 2015; Linchant et al., 2015b). Light aircraft with onboard cameras may also produce ultra-high spatial resolution images and cover large areas (Lamprey et al., 2020a, 2020b). Nevertheless, the temporal resolution is heavily dependent on the available budget, which usually does not permit yearly surveys as a result of the high costs involved (Bouché et al., 2012; Dunham, 2012; Wang et al., 2019; Watts et al., 2010). As for satellites, they may provide images covering very large areas, at a submeter spatial resolution, allowing the detection and counting of large species but not their identification (Wang et al., 2019; Wu et al., 2023). Their temporal resolution may also be high (e.g. 1-2 days) but strongly depends on a commercial system that cannot guarantee the delivery of an image when tasked to a specific location, not to mention the current high cost of images (Wang et al., 2019). Finally, the choice of platform depends on the species studied, the spatial and temporal resolution required, and the available budget.

Regardless of the remote sensing platform, on-board sensors produce large volumes of data, i.e. large numbers of images or huge numbers of pixels. The manual analysis of these images is perhaps the biggest challenge of remote sensing for wildlife surveys, as it is very time-consuming, tedious and involves the assistance of experts (Cubaynes et al., 2019; Lamprey et al., 2020b; Linchant et al., 2015b; Lynch and LaRue, 2014). Although the analysis time depends on a number of factors, such as the experience of the interpreter, the resolution of the image or the degree of heterogeneity of the environment, manual analysis of ultra-high resolution (UHR) aerial imagery (GSD < 10cm) can be carried out at a speed between 0.5 and 2 km²/hour (Lamprey et al., 2020b; Peng et al., 2020), whereas it can be much faster for satellite imagery, i.e. > 5 km²/hour (Corrêa et al., 2022; Cubaynes et al., 2019). While being essential for rapidly validating or establishing conservation actions, results from remote sensing surveys of PAs covering thousands of square kilometers and generating thousands of images can therefore be delayed by several months.

4. The advent of deep learning

Recent advances in machine learning have propelled the perspectives of remotely sensed imagery for wildlife conservation (Tuia et al., 2022). In fact, partial or total automation of image processing will address the main challenge of remote sensing images: the substantial volume of data to be managed. The benefits of such developments would be twofold. First, it will sustain and reinforce the need to use remote sensing for conservation. Second, it will considerably reduce the costs associated with manual analysis.

The aim of this section is to introduce the theoretical elements associated with deep learning (DL), a subfield of artificial intelligence (AI) that has been particularly targeted as promising for animal counting applications. It should be noted that most

of the technical information provided was derived from the book of Goodfellow et al. (2016), the book of Elgendy (2020) and the review paper of LeCun et al. (2015). Furthermore, the potential of DL for remote sensing image processing in the context of large mammal surveys is presented and discussed.

4.1. Definition and principles

Deep learning is a subfield of machine learning methods which are able of automatically extracting patterns from data and making predictions, through the use of computer systems (Goodfellow et al., 2016). Conventional machine learning methods often required hand-engineering to design a *feature extractor* that transforms the raw data into a suitable representation (i.e. a feature vector) to be finally used in the learning algorithm. DL involves the construction and training of artificial neural networks (ANNs) with multiple layers, each composed by non-linear modules that each transform the representation at one level into a higher one (LeCun et al., 2015). These networks were inspired by the functioning of biological brains. In contrast to conventional machine learning methods, ANNs are designed to automatically learn and represent hierarchical features and patterns from natural data in their raw form, enabling the extraction of abstract representations. The depth of these networks (i.e. the high number of layers) facilitates the modeling of complex relationships within the data, allowing for the discovery of representations that may be challenging to capture with conventional machine learning approaches (Goodfellow et al., 2016; LeCun et al., 2015).

In its simplest form, an ANN is made up of interconnected neurons, each of which processes the input information and outputs another, called a *feature*. The whole interconnected system is commonly called the 'architecture'. The simplest one being the *perceptron*, composed by a single neuron (**Figure 1.3a**), while the most complex ones are *multilayer perceptron* (MLP) for processing fixed-dimension structured data (**Figure 1.3b**), or Convolutional Neural Network (CNN) for processing grid-like data (**Figure 1.4**), both composed by multiple layers of neurons. Layers between the first and last layers are commonly called *hidden layers*, and this is where complex features are learned. Mathematically, a perceptron is the weighted sum of the multiple input variable values, plus a bias, followed by an activation function. The latter allows to induce non-linearity of the output, by transforming the weighted sum value. A typical activation function example is the step function that produces a binary output: if the weighted sum value is positive then it outputs 1, else it outputs 0. The size of an ANN's architecture is given by its number of parameters, which consists of its number of weights and biases.

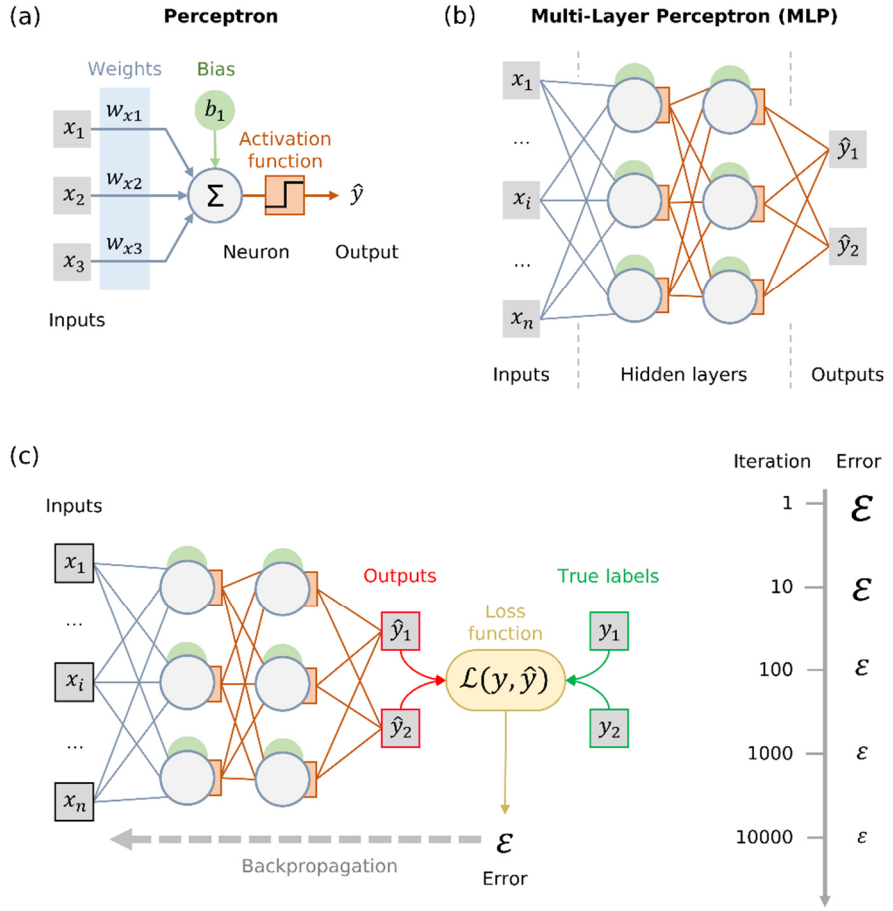


Figure 1.3: Components, architecture and training principle of an Artificial Neural Network (ANN): (a) one neuron called 'perceptron', (b) a simple ANN, the multi-layer perceptron (MLP), (c) training process of an ANN.

One of the most common forms of learning process is *supervised learning*, which consists of training an ANN using labeled data. For instance, in an image classification context, this means having labeled images, i.e. a dataset where the input variable, X (i.e. the image) and the desired output, Y (i.e. the label) are known. Supervised learning relies on a process called *backpropagation* for training the ANNs. During training, the ANN is presented with input data (X), and its initial predictions ($\hat{Y}$) are compared to the actual outcomes (Y). The error (i.e. the loss) between the two is estimated using a *loss function* (**Figure 1.3c**). This error is then backpropagated through the network, and the weights between the neurons are adjusted to minimize this error, using an *optimizer*. The latter calculates gradients to properly modify the ANN's weights. These gradients indicate how much the error would change with a slight increase or decrease in each weight. The weights are then adjusted in the

opposite direction to the gradients. This iterative process continues with multiple rounds of presenting data, calculating errors, and updating weights, optimizing the network's ability to make accurate predictions (**Figure 1.3c**). The objective is for the ANN to generalize patterns and learn complex representations from the training data, enabling it to make accurate predictions on new, unseen data, commonly called the *test* set. In the context of this thesis, the term *architecture* refers to the whole structure of an ANN, and the term *model* to the trained version of an architecture, including post-processes.

4.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs), also known as *ConvNets* (LeCun et al., 1989), are a type of ANN designed to process structured grid-like data, such as images. Thanks to the increased capacity of computer systems in the last decade, CNNs are now a valuable resource for processing remote sensing images (Hoeser and Kuenzer, 2020; Kattenborn et al., 2021; Zhu et al., 2017). CNNs are composed of multiple convolutional layers that, as their name suggests, use a mathematical operation other than the multiplication used in MLPs: the convolution. To put it simply, a convolution is an operation that combines two functions to produce a third, capturing the relationship between them. In the context of CNNs, convolution is discrete and refers to the process of applying a two-dimensional array of weights, i.e. a typical filter (or kernel) of 3x3 pixels, to an array of input data (e.g. an image), resulting in an output called *feature map* (**Figure 1.4a**). This involves sliding the filter across the input, element-wise multiplying the filter values with the overlapping input values, and summing the results.

A convolutional layer typically consists of three main stages (**Figure 1.4a**). First, the convolution stage, in which a set of learnable filters is applied to the input data to produce a set of linear activations capturing local patterns and features. Second, the activation stage, during which a non-linear activation function, commonly the Rectified Linear Unit (ReLU) (Fukushima, 1969), takes as input the linear activations to produce a set of non-linear activations. By introducing non-linearity, this stage helps the network to learn more complex representations in a much faster way (Glorot et al., 2011). Third, the pooling stage, in which the spatial dimension of the non-linear activations is reduced. Max pooling (Zhou and Chellappa, 1988) is commonly used, and consists of dividing the input into small regions (e.g. 2x2 pixel window) and retaining only the maximum value within each region. This last stage helps to merge semantically similar features, to reduce the number of parameters of the architecture, and to make the representation invariant to small translations of the input. These stages, directly inspired by the notions issued from visual neuroscience (Hubel and Wiesel, 1968, 1962), are repeated in multiple layers to create a hierarchical representation of features in the input data, in which higher-level features are obtained by composing lower-level ones (**Figure 1.4b**). This enables the CNN to learn and recognize complex patterns and structures. For instance, in images, the lower-level features represent edges, the combination of the latter shape motifs then motifs form

parts in the mid-level features, and finally the combination of multiple parts reveals objects in higher-level features (LeCun et al., 2015). CNNs have proved particularly useful in the field of computer vision for capturing spatial patterns that have eluded conventional image analysis algorithms. Using CNNs or combining them with MLPs to form DL architecture, complex tasks can be achieved, such as image classification or object detection.

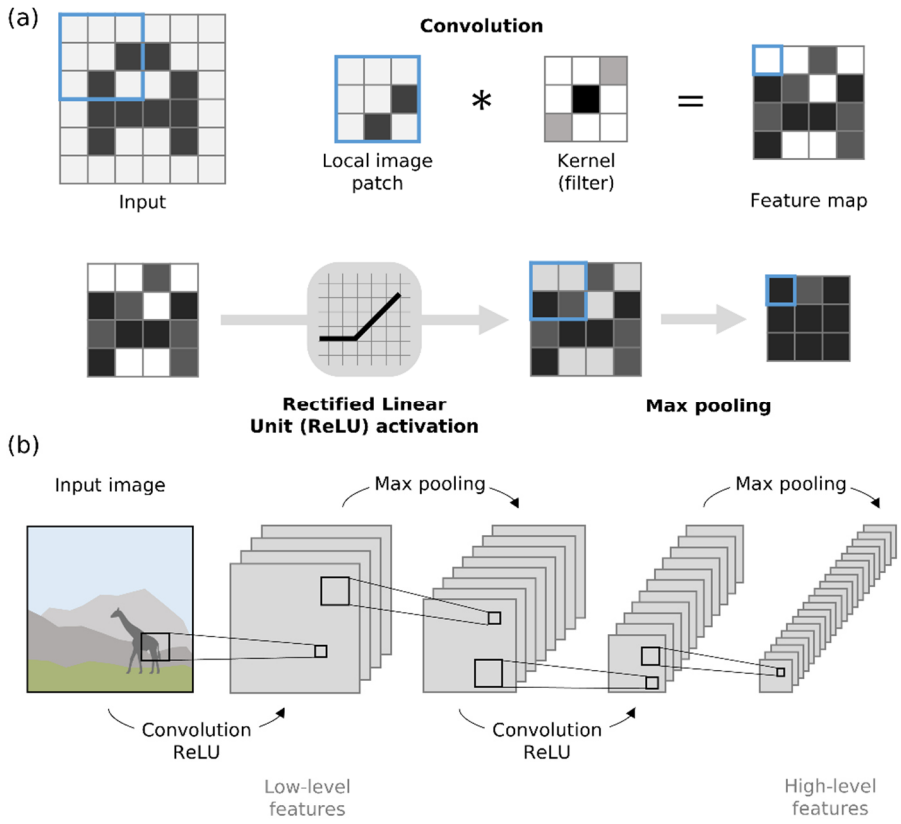


Figure 1.4: Mathematical operations behind a Convolutional Neural Network (CNN) and basic architecture: (a) main stages of a CNN, i.e. convolution, activation and pooling, (b) a simple CNN architecture with multi-level feature representations.

4.3. Object detection using CNNs

In the field of computer vision, object detection (Fischler and Elschlager, 1973) is a task that involves the localization and the category recognition of multiple objects within an image. Traditionally, object detectors used to rely on three main steps (Zou et al., 2023) (**Figure 1.5a**). First, the generation of region proposals, using formerly sliding windows approaches (Viola and Jones, 2001, 2004) to obtain regions of interest (RoI) which potentially contain objects. Second, the features extraction from

RoI, based traditionally on various algorithms like Scale-Invariant Feature Transform (SIFT) (Lowe, 1999), Haar (Lienhart and Maydt, 2002), Histogram of Gradients (HOG) (Dalal and Triggs, 2005), or Speed Up Robust Features (SURF) (Bay et al., 2006). This step served to produce fixed-length features vectors capturing the semantic information of the RoIs. Third, the RoI classification, using formerly Support Vector Machine (SVM) (Hearst et al., 1998), adaboost (Freund and Schapire, 1996) or cascade learning (Viola and Jones, 2004) to decide whether the RoI is an object or not, and/or to associate a category.

As for other computer vision tasks, CNNs have emerged as the cornerstone for object detection, due to their ability to automatically learn hierarchical features from a set of raw images (Zou et al., 2023). The milestone was initiated by Girshick et al. in 2014 (Girshick et al., 2014), with the introduction of the Region-based CNN (R-CNN) (Girshick et al., 2016). Since then, the evolution of DL architectures based on CNNs has accelerated at an unprecedented speed. CNN-based object detectors are now mainly grouped into two categories (Zhao et al., 2019; Zou et al., 2023). First, the *two-stage* detectors, which nearly follow the traditional object detection pipeline, i.e. region proposal generation, feature extraction then classification, but using mainly CNNs (**Figure 1.5b**). Second, the *one-stage* detectors, which regard object detection as a regression/classification task, generating objects in a single step. For both categories, the feature extractor is commonly called a *backbone*, while the other blocks built on top of it are called *heads*. One-stage object detectors achieve detection at high processing rate (in real-time for some applications), but their performance is usually poorer than two-stage detectors and may suffer when detecting dense and small objects (Zhao et al., 2019; Zou et al., 2023). In recent years, well-known and commonly used architectures have included Faster R-CNN (Ren et al., 2015), Cascade R-CNN (Cai and Vasconcelos, 2021, 2018) or Libra R-CNN (Pang et al., 2019) for the two-stage detectors, and YOLO (Redmon et al., 2016), RetinaNet (Lin et al., 2017b), or SSD (Liu et al., 2016) for the one-stage detectors. During inference, all these detectors make predictions in the form of labeled rectangles encompassing the detected objects, called *bounding boxes*. While this type of annotation and predictions is the most common one for object detection, other CNN-based architectures promoted the use of points to detect objects (Ribera et al., 2019; Zhou et al., 2019). It has been highlighted as a more suitable and faster approach for annotation, and has already been used for some time to annotate human crowds (Gao et al., 2020; Li et al., 2021), a particular case of detection where the object density makes it quite challenging.

DL has truly revolutionized object detection and is now the dominant method in the field. However, the imbalance problems associated with its use, but not limited to it, may have a non-negligible impact on performance. Imbalance problems even constitute a research area and are receiving particular attention from researchers. For CNN-based object detectors, imbalance occurs at different levels of the whole detection pipeline. This can range from object number imbalance in classes, to task imbalance in the training objective (e.g. regression, classification), as well as

variations in object scale or position in the image (Oksuz et al., 2021). These problems may be particularly acute in applications where remote sensing imagery is used, making the development of object detectors even more challenging.

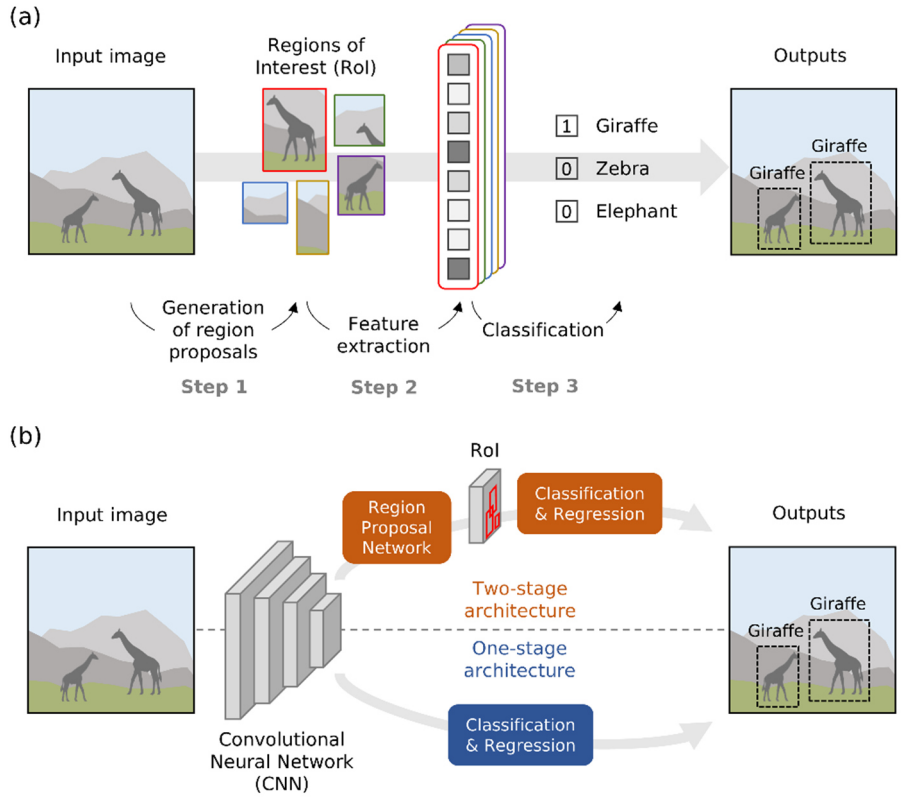


Figure 1.5: Basic principles of object detection: (a) traditional pipeline of object detectors, (b) the two main architectures of CNN-based object detectors.

4.4. Potential of deep learning for surveys of large mammals in Africa

While this was not a priority when cameras were first used intermittently in crewed aircraft, the growing use and interest in drone systems for wildlife conservation has naturally led to the need to automate the processing of the large volume of images generated (Corcoran et al., 2021; Linchant et al., 2015b). The first papers dealing with the (semi-)automated detection of large African mammals on remote sensing imagery using DL were published in the last decade. Pioneer works were the one of Yang et al. (2014) for space-based platforms, Eikelboom et al. (2019) for aerial-based piloted systems, and Kellenberger et al. (2018, 2017) for aerial-based unoccupied systems.

Since then, multiple research studies have been conducted covering different architectures, techniques, camera orientations and species.

Various DL approaches have been used in the literature to count animals on remote sensing imagery, each with its own benefits and limits. A first 'simple' approach was to use image classification architecture to detect the presence of animals within patches cropped from whole images (Barbedo et al., 2020, 2019; Borowicz et al., 2019; Guirado et al., 2019; Moreni et al., 2021; Rahnemoonfar et al., 2019; Rivas et al., 2018; Zhou et al., 2021). This approach can yield good detection results but cannot provide proper counting results (Barbedo et al., 2020; Moreni et al., 2021). Image classification may then be used to remove background areas to substantially reduce the pixel area to analyze. In cases where counting is important, other approaches should be used, such as models producing two-dimensional probability maps (Bowler et al., 2020; Kellenberger et al., 2018, 2019a; Wu et al., 2023) or CNN-based single- or two-stage object detectors (Duporge et al., 2021; Eikelboom et al., 2019; Green et al., 2023; Guirado et al., 2019; Lema et al., 2021; Peng et al., 2020; Sarwar et al., 2021; Torney et al., 2019), from which precise animal positions and counts can be obtained. Other authors have experimented with density-based CNN architectures (Kellenberger et al., 2019a; Padubidri et al., 2021; Qian et al., 2023; Rahnemoonfar et al., 2019). The latter produces density maps of which the integral gives the number of objects within the input image. This approach was initiated by Lempitsky and Zisserman (2010) and is trained using point annotation and a density function convolved over each point, typically a normalized two-dimensional Gaussian. While density-based CNN architectures showed to be particularly efficient for counting dense and small objects, e.g., crowds (Gao et al., 2020; Li et al., 2021), precise object location is lost, especially for close-by objects where the density functions strongly overlap. Eventually, some authors developed point-based CNNs to avoid the use of bounding boxes (Bowler et al., 2020; Gonçalves et al., 2020; Kellenberger et al., 2021; Mücher et al., 2022; Naudé and Joubert, 2019; Sarwar et al., 2021; Wu et al., 2023) that may be challenging to draw and post-process for very small objects or dense object scenes. These architectures typically consist of a fully CNN with an encoding and decoding phase, trained to regressively learn a mapping between the input image and a two-dimensional pixel map from which point positions and labels are extracted. Point-based CNNs then benefit from the advantages of both CNN-based object detectors and point annotation.

Concerning large African mammals, several authors have already studied the use of DL to automate animal detection and counting from remote sensing imagery. To the best of our knowledge, except for Padubidri et al. (2021), all the previous studies addressed the task through the use of object detectors, whether based on CNNs (Duporge et al., 2021; Eikelboom et al., 2019; Fang et al., 2021; Kellenberger et al., 2018, 2019a, 2019b; Mou et al., 2023b; Naudé and Joubert, 2019; Peng et al., 2020; Petso et al., 2021; Torney et al., 2019; Wu et al., 2023) or more primary ANNs (Xue et al., 2017; Yang et al., 2014). Most of the papers have focused on a single species or on a binary detection case, i.e. animal (all species combined) versus background,

but have produced promising performances with some achieving around 90% of detection metrics (e.g., Naudé and Joubert, 2019; Peng et al., 2020). Others have experimented the multispecies configuration and thus incorporated the additional recognition dimension (Eikelboom et al., 2019; Fang et al., 2021; Mou et al., 2023b; Petso et al., 2021), obtaining multilabel detection metrics ranging from 59% to 98%.

The object detection approach therefore appears to be a suitable solution for processing survey images of large African mammals. However, this particular use-case raises or accentuates challenges little or not encountered in conventional computer vision applications. The biggest challenge is to avoid the massive production of false alarms (i.e. false positives) while detecting the maximum number of individuals. False alarms are usually raised by the natural heterogeneity particularly present in African landscapes, which contain many confounding elements such as rocks, tree trunks or shadows (Kellenberger et al., 2018). Another challenge is the imbalance inherent in remote sensing images covering natural scenes. In the case of large mammal detection in wild and vast environments, this imbalance mainly occurs on two levels. First, the so-called *foreground-background* imbalance, which here refers to the imbalance between the pixel surface covered by the animals (i.e. the foreground) and the one covered by the landscape (i.e. the background). Large mammals indeed systematically cover just a tiny fraction of the total pixel area of an image (Fang et al., 2021). Training an object detector roughly on such an unbalanced data configuration will inevitably make it miss most animals (Kellenberger et al., 2018). Second, the *foreground-foreground* imbalance, which refers to the imbalance in the number of instances per animal category (or class). This imbalance problem is intrinsically linked to the natural uneven distribution of animal species in their environments. The risk here is that minority species (i.e. those in smaller numbers) are not correctly identified by the detector. Finally, another challenge that is worth mentioning is the uneven distribution of animals within an image. This may lead to a drop in detector performance between sample images containing scattered animals and those containing clumped groups or close-by individuals (Han et al., 2019; Peng et al., 2020; Rivas et al., 2018). The future of large African mammal surveys using remote sensing imagery therefore relies on the development of DL methods that address these challenges.

5. Research strategy

The ultimate goal of fusioning remote sensing and DL is to enhance the accuracy and precision of traditional aerial surveys, while reducing the risks and costs associated with the latter. Imagery would avoid any observer bias and (semi)automatic models would alleviate the burden of manual image interpretation. The benefits of such approaches would, on the one hand, reduce the costs associated with this task which may be a major stumbling block for conservation agencies (Bröker et al., 2019; Lamprey et al., 2020b). On the other hand, such automatic models and data acquisition would enable standardization of survey results against recorded data. The accuracy and precision of surveys would only improve, and changes in population size would

be more easily perceptible over time. This would enable conservation actions to be taken more promptly and eventually, biodiversity to be better preserved.

There are still many research gaps before an acceptable and reliable solution emerges, not to mention the challenges posed by the modernization of traditional approaches. This section therefore presents the research gaps in this domain as well as the research objectives of this thesis, which started in 2020.

5.1. Research gaps

Although many species have been studied using remote sensing and DL, the case of large African mammals has been little studied. It was not until 2018 that interest grew, with the publication of the paper by Kellenberger et al. (2018). Since then, number of articles have been published, many of them after 2020, but mostly on 'proof-of-concept' cases and 'limited' datasets that are not representative of the real scale involved in a complete survey covering thousands of square kilometers. This is certainly linked to the research that has mainly focused on the drone as an acquisition platform over the last decade. Although promising, their low endurance does not currently meet the needs of surveys of vast open savannahs (Linchant, 2021) and thus covered limited areas. Most of the studies concentrated their research on the development of DL models for the processing of drone and satellite images, rather than aerial images acquired using light aircraft (Wang et al., 2019). Therefore, most of the models were developed to detect large mammals on images acquired with a vertical view (i.e. nadir), and few to be applied on oblique view (Barbedo et al., 2020; Eikelboom et al., 2019; Fang et al., 2021; Torney et al., 2019), although the latter may be particularly useful for avoiding tree occlusion, proper species identification and maximizing ground coverage (Lamprey et al., 2020b). To the best of our knowledge, before 2020, only Eikelboom et al. (2019) had considered species recognition in addition to detection and localization in the image. Fang et al. (2021) and Petso et al. (2021) closely followed the next year. Although more challenging than the binary case (i.e. animal versus background), the multi-species case is therefore a little-studied case for large African mammals, whereas species recognition may represent a particular benefit for the processing of survey images. In addition, previous research has focused mainly on wildlife in PAs, leaving the automatic detection and counting of free-ranging livestock in these regions an unexplored subject. When studying livestock in these environments, individuals can range from the scattered to the densely clumped. This can also be the case with certain wild species (e.g. the buffalo), but is particularly more frequent with livestock like goats, sheep, or cattle. Current DL architectures for object detection (e.g. Faster R-CNN) are basically not developed to handle such diverse cases, and even show trouble precisely counting scenes of very dense objects (Zou et al., 2023). Pending the emergence of long-endurance drones or the broader use of microlight aircraft in PAs, there is a real need to develop multi-species counting methods for automatic processing of oblique aircraft images (Lamprey et al., 2020b, 2020a, 2023).

5.2. Objectives and structure of the thesis

Following the challenges and research gaps described above, the **general objective** of this thesis is to evaluate the combined use of remote sensing and DL for large African mammal multi-species survey through automated methods of detection, counting and recognition. In the context of this thesis, *large African mammals* refers here as wild or domestic terrestrial mammals over 40kg living in or passing through PAs in sub-Saharan Africa.

The core of the thesis is structured into three main chapters (**Chapter 2** to **4**), consisting of 4 published research articles, and concluded by a discussion chapter (**Chapter 5**), drawing partly on a fifth published review article available in the Annex (**Figure 1.6**). The research focuses on passive systems embedded on aerial platforms, in particular aircraft. Although the use of drone is discussed in **Chapter 2**, the main focus of the research was on the use of light aircraft in **Chapter 3** and **Chapter 4**. There are two reasons for this. Firstly, as previously stated, the use of drones in a wildlife monitoring context is a field that has already been extensively researched over the last decade. Secondly, given the current disadvantages of drones, I believe that light aircraft seem to be the most suitable platform for the practical implementation of the developed approaches in the near future. Concerning the sensor used and target spectrum, only optical sensors detecting the visible spectrum are considered, i.e. sensors producing images in the Red-Green-Blue (RGB) bands.

The thesis therefore revolves around UHR aerial imagery (GSD < 10 cm/pixel), and aims to answer the following research question:

"*Does the association of aerial imagery and DL models increase the accuracy and precision of population estimates for large mammals in sub-Saharan protected areas?*"

Thanks to key and growing collaborations since the early stages of the thesis, the latter is fortunate to cover several protected areas with varied, substantial and representative datasets. This variability and large amount of data has been a valuable resource for practical discussion of the approaches developed, in different environments and with different species. The study areas therefore covered multiple open to semi-open savannas from western to eastern Africa, more specifically: the Garamba and Virunga National Parks in Democratic Republic of Congo (DRC), the Ennedi Natural and Cultural Reserve (ENCR) in Chad, the Queen Elizabeth National Park (QENP) in Uganda, and the Comoé National Park (CNP) in Côte d'Ivoire.

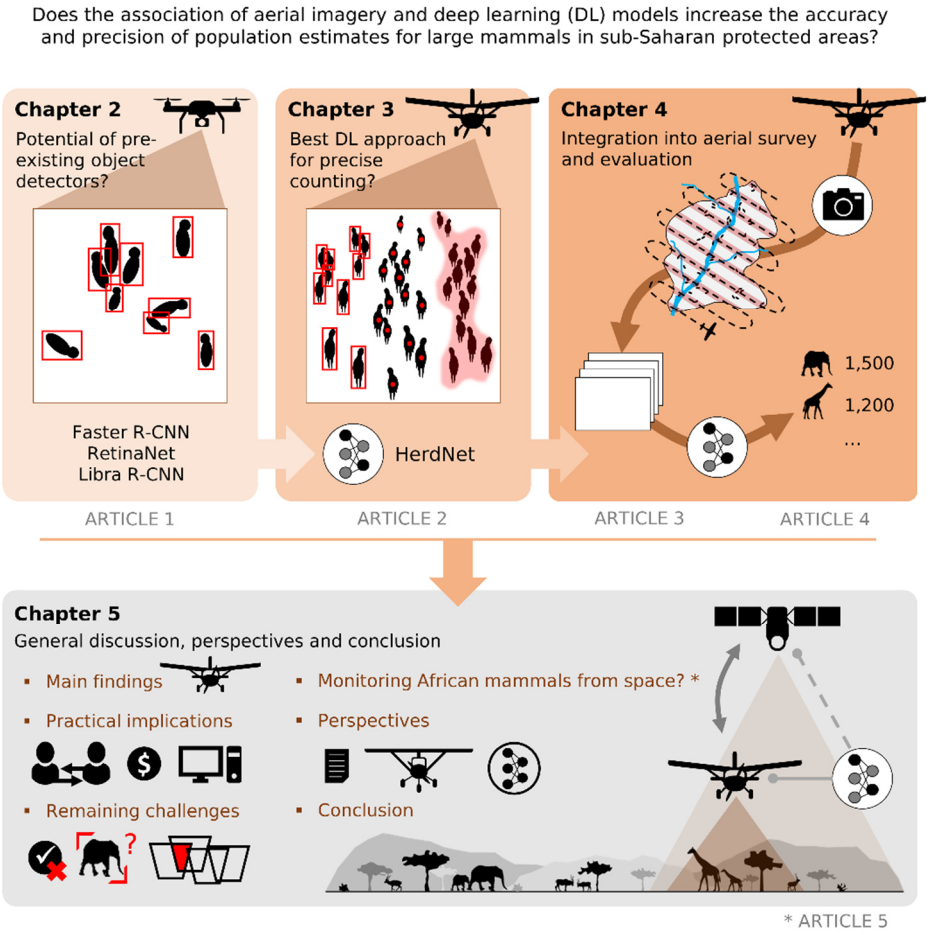


Figure 1.6: General framework and research strategy of the thesis.

To answer the research question, a DL model was developed and evaluated, and then applied in the context of a whole aerial survey. First, in **Chapter 2**, pre-existing CNN-based object detection models that have shown good performance in other computer vision tasks (e.g. Faster R-CNN) were evaluated for multi-species counting of large African mammals on drone imagery. Then, **Chapter 3** describes a DL architecture developed as part of the thesis, named *HerdNet*, designed to precisely count and recognize both scattered and dense herds of large African mammals. It has been first developed and validated at small scale on both nadir images acquired by drone and oblique images acquired by aircraft. Finally, **Chapter 4** presents the results of applying the HerdNet model to oblique images of photographic aerial surveys. The first sub-chapter presents the performance comparison between the model and manual counting for wild species in Uganda. The second sub-chapter presents, to our knowledge, the very first comparison between population estimates obtained via observer data and those via verified detections of a DL model. This was done on the

scale of a PA of around 11,500 km², the Comoé National Park, in which a hybrid aerial survey (i.e. continuous camera and human observers) was conducted in 2022.

In **Chapter 5**, the main findings are presented and discussed, along with the practical implications and limitations of the approaches developed. In addition, the remaining challenges are presented and accompanied by research perspectives, and one section is dedicated to the potential use of satellite imagery for African mammals monitoring. Suggestions for enhancing HerdNet performance are provided and discussed, as well as for upgrading aerial survey standards, integrating continuous imaging and the use of AI for processing the images. Prospects for microlight aircrafts are also put forward, the complementarity of remote sensing platforms for wildlife monitoring and survey is discussed, as well as the development of foundational DL models. The chapter ends with the main conclusions of the thesis.

2

Potential of CNN-based object detectors

Preamble

In this chapter, I focused on the training and application of pre-existing CNN-based object detectors developed within the expansive domain of computer vision. This study specifically targeted their suitability for deployment in the task of wildlife aerial counting. Three distinctive architectures were pre-selected for evaluation, and their performance was compared using aerial imagery obtained from both drone and aircraft, encompassing diverse species of large African mammals. The comparative analysis of the best model's outputs on images acquired from a distinct PA provided an opportunity to discuss its potential for species counting and recognition, as well as highlighting its limits in herd contexts.

Paper 1 | Published

Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks

Alexandre Delplanque, Samuel Foucher, Philippe Lejeune, Julie Linchant & Jérôme Théau

This paper is published in *Remote Sensing in Ecology and Conservation* (IF=3.9), 8(2), 166-179. DOI: 10.1002/rse2.234

Abstract

Survey and monitoring of wildlife populations are among the key elements in nature conservation. The use of unmanned aerial vehicles and light aircrafts as aerial image acquisition systems is growing, as they are cheaper alternatives to traditional census methods. However, the manual localization and identification of species within imagery can be time-consuming and complex. Object detection algorithms, based on convolutional neural networks (CNNs), have shown a good capacity for animal detection. Nevertheless, most of the work has focused on binary detection cases (animal vs. background). The main objective of this study is to compare three recent detection algorithms to detect and identify African mammal species based on high-resolution aerial images. We evaluated the performance of three multi-class CNN algorithms: Faster-RCNN, Libra-RCNN and RetinaNet. Six species were targeted: topis (*Damaliscus lunatus jimela*), buffalos (*Syncerus caffer*), elephants (*Loxodonta africana*), kobs (*Kobus kob*), warthogs (*Phacochoerus africanus*) and waterbucks (*Kobus ellipsiprymnus*). The best model was then applied to a case study using an independent dataset. The best model was the Libra-RCNN, with the best mean average precision (0.80 ± 0.02), the lowest degree of interspecies confusion ($3.5 \pm 1.4\%$) and the lowest false positive per true positive ratio (1.7 ± 0.2) on the test set. This model was able to detect and correctly identify 73% of all individuals (1115), find 43 individuals of species other than those targeted and detect 84 missed individuals on our independent UAV dataset, with an average processing speed of 12 s/image. This model showed better detection performance than previous studies dealing with similar habitats. It was able to differentiate six animal species in nadir aerial images. Although limitations were observed with warthog identification and individual detection in herds, this model can save time and can perform precise surveys in open savanna.

Keywords: African mammals, CNNs, deep learning, multispecies, UAV, wildlife monitoring

1. Introduction

Survey and monitoring of animal populations are key management tools in nature conservation and are essential to help fight the pressures they suffer. Anthropogenic pressures, such as poaching, are encountered mainly in developing countries (including most in Africa) where the pressure on biodiversity is very high (Linchant et al., 2015b). While large African mammals, such as buffaloes (*Syncerus caffe*r) or hippopotamuses (*Hippopotamus amphibius*), play an important role in the dispersion and migration of macro-nutrients within landscapes (Lacher et al., 2019), the average abundance of these populations declined by 59% between 1970 and 2005 (Craigie et al., 2010). Moreover, the latest estimated Living Planet Index indicates a 65% decline in the overall African vertebrate populations between 1970 and 2016 (WWF, 2020). Even though humans are dependent on biodiversity (Isbell et al., 2017; WWF, 2020), their impact on the environment is leading us into a period of mass extinction (Ceballos et al., 2015). Moreover, in view of the disruption of future climate conditions (Pachauri et al., 2014), species not able to adapt rapidly could see their populations decline even further (Hetem et al., 2014; Thuiller et al., 2006).

Most of the time, the size of an animal population is estimated through sample counts which consist of estimating the animal density in sample units selected at random or following a systematic scheme. The size of the population corresponds to the product of the mean density inside sample units per surveyed area surface (Norton-Griffiths, 1978). Unfortunately, counting campaigns of this type can rapidly become expensive (Gaidet-Drapier et al., 2006), particularly for large mammal surveys for which the use of a light aircraft is almost indispensable (Jachmann, 1991). Moreover, these aerial campaigns can become dangerous, and their logistics are very complex for operators (Watts et al., 2010; Witmer, 2005).

Although they cannot cover large areas, the use of UAVs (unmanned aircraft vehicles) is presented as a cheaper, more suitable and safer alternative (Chabot and Bird, 2015; Linchant et al., 2015b; Vermeulen et al., 2013). In addition, there are sensors that can be embedded and which offer the possibility of acquiring very high-resolution images (Linchant et al., 2015b).

Several species of large African mammals have already been studied using UAVs, such as the African elephants (*Loxodonta africana*) (Vermeulen et al., 2013), black (*Diceros bicornis*) and white (*Ceratotherium simum*) rhinos (Mulero-Pázmány et al., 2014), the hippopotamus (Linchant et al., 2018) and many other species (Kellenberger et al., 2018; Rey et al., 2017).

However, counting and identification are not carried out simultaneously and must be deferred when using UAVs. Due to the large amount of data to be analyzed, the size of the study area and the static nature of the animals on the images, counting can become very complex and time-consuming. This problem can be alleviated by utilizing object detection, which finds, locates and classifies objects in images (Zhao et al., 2019). Convolutional neural networks (CNNs) have become the basic elements

of most computer vision processes and have also proven to be extremely effective in the field of remote sensing (Zhu et al., 2017). These networks have been applied to animal detection in aerial and UAV images and have shown encouraging results (Barbedo et al., 2019; Eikelboom et al., 2019; Kellenberger et al., 2018; Moreni et al., 2021; Naudé and Joubert, 2019; Peng et al., 2020). However, almost all of these studies did not distinguish between species nor were they focused on the case of a single species. It would therefore be interesting to develop a multi-species approach in order to further minimize human resources required for the processing of survey data. To our knowledge, only one study of multispecies animal detection on aerial images using object detection has been conducted to date, Eikelboom et al. (2019), who worked on detecting and identifying three African animal species using aerial oblique images and CNN.

The objective of this study is to compare the performances of three object detection algorithms, based on CNNs, to automatically detect and identify six African mammal species in nadir aerial images: African buffalo, kob (*Kobus kob*), topi (*Damaliscus lunatus jimela*), African warthog (*Phacochoerus africanus*), waterbuck (*Kobus ellipsiprymnus*) and African elephant. The best model is then put into a practical perspective on an independent set of UAV images acquired in a different study area.

2. Materials and Methods

2.1. Dataset

2.1.1. Data collection

We used three different aerial datasets to conduct our study (see details in **Table 2.1**). The 'Virunga' and 'Garamba' are two UAV datasets that were taken from a database maintained by the University of Liège, Gembloux Agro-Bio Tech (Belgium). The Aerial Elephant Dataset (AED) is a free dataset provided by Naudé and Joubert (2019).

The Virunga and AED datasets were merged and used as the 'general dataset' to develop the models (training, validation and test), while the Garamba dataset was used as a 'case study' to test the performance of the best model on a complete independent dataset. This was done in order to evaluate the model on a practical use case that did not include all the targeted species and which contained other species.

The species selection was based on the availability of at least 100 individuals in the general dataset to ensure minimal model configurations. In addition, to optimize the speed of model development, images that did not contain animals were not used.

Table 2.1: Dataset specifications and details.

Dataset	**Virunga**	**AED** (Naudé and Joubert, 2019)	**Garamba**
Location	DRC (Virunga national park)	Parks, games, and reserves in Botswana, Namibia and South Africa	DRC (Garamba national park)
Land cover (Mayaux et al., 2004)	Savanna	Deciduous woodland, open deciduous shrubland, closed grasslands	Savanna
Dates	April-June 2016	2014 to 2018	May 2015
Time of day	Early morning	Full day	Early morning
System	Falcon (UAV)	SkyReach BushCar (A/C)	Falcon (UAV)
Camera(s)	Sony-A6000, Sony-Nex7	Canon 6D	Sony-Nex7
Flight altitude	100 m	220 to 2270 m	90 m
Number of flights	9	8	6
Image dimension	6,000 x 4,000 pixels	Various (5,472 x 3,648 pixels; 5,496 x 3,670 pixels; 5,521 x 3,687 pixels; 5,525 x 3690 pixels)	6,000 x 4,000 pixels
GSD	2.4 cm	2.4 to 13.0 cm	2.0 cm
Species	hippopotamus, buffalo, kob, topi, warthog, waterbuck	elephant	hartebeest (*Alcelaphus buselaphus*), hippopotamus, buffalo, kob, warthog, waterbuck, giraffe (*Giraffa camelopardalis*)
Images selected	897	400	All (7034)

GSD, Ground Sampling Distance ; AED, Aerial Elephant Dataset ; DRC, Democratic Republic of Congo ; UAV, Unmanned Aerial Vehicle ; A/C, Aircraft.

2.1.2. General dataset data splitting

The distribution of individuals of each species according to training, validation and test sets is given in **Table 2.2**. The approximate targets of distribution were 70% of the individuals in the training, 10% in the validation and 20% in the test datasets.

Table 2.2: Number of individuals according to species, training, validation, and test sets.

Species	Training	Validation	Test	Total
Buffalo	1058 (70%)	102 (7%)	349 (23%)	1509
Elephant	2012 (68%)	264 (9%)	688 (23%)	2964
Kob	1732 (73%)	161 (7%)	477 (20%)	2370
Topi	1678 (62%)	369 (13%)	675 (25%)	2722
Warthog	316 (73%)	43 (10%)	74 (17%)	433
Waterbuck	166 (69%)	39 (16%)	36 (15%)	241
Total	6962 (68%)	978 (10%)	2299 (22%)	10239

The different rows show the distribution of individuals in each set and the relative percentage (in parentheses).

The distribution of the number of individuals by species and by flight was considered in performing the split. This step was required in order to avoid the splitting of some consecutive images containing the same individuals and to thereby maintain the independence of the three sets.

2.1.3. Ground truth

For the Virunga and Garamba datasets, the annotations (points and labels) were provided with the images. The individuals were previously located and identified manually by two operators on the UAV images using the software WIMUAS (Linchant et al., 2015a). The AED dataset also provided annotations (points) with the images (Naudé and Joubert, 2019). We assumed that all pre-identifications were correct. Bounding boxes were manually defined by a co-author of this study using the Colabeler AI annotation tool (http://www.colabeler.com/).

2.2. Methodology

2.2.1. Detection algorithms and implementation on the general dataset

Three object detection algorithms were tested: Faster-RCNN (Ren et al., 2017), Libra-RCNN (Pang et al., 2019) and RetinaNet (Lin et al., 2017b). These algorithms were selected based on their performance on the benchmark datasets and on the availability of the code at the time of the study.

Faster-RCNN

This object detection algorithm (Ren et al., 2017) takes images as input and constructs feature maps using a CNN (also called backbone). Based on these features maps, a region proposal network generates region proposals and assigns a probability of containing an object to each region. The predicted region proposals are then reshaped and eventually, classification and bounding box regression is performed to predict the presence and location of objects in the input images. These types of networks are commonly called 'two-stage detectors' due to their two-step process (Soviany and Ionescu, 2018). Faster-RCNN was chosen because it is used in many studies as a baseline.

Libra-RCNN

This algorithm, developed by Pang et al. (2019), is also a two-stage detector that does basically the same thing as Faster-RCNN. Its particularity is that it balances the training process at three levels which initially limit the detection performance:

1) the sample level, by balancing the distribution of training samples close to that of challenging samples (called hard negatives). This addresses the problem of the random sampling scheme that often results in selected samples dominated by the easy ones (Pang et al., 2019);
2) the feature level, by balancing the low-level and high-level features of each layer in the backbone, which are complementary for object detection;
3) the objective level, by balancing the tasks of localization and classification, thus avoiding one of the two tasks being overwhelmed by the other.

Thanks to its multi-level balanced approach to training, Libra-RCNN allows for greater precision and recall than Faster-RCNN, which is why it has been selected for comparison.

RetinaNet

This third algorithm is a 'single-stage detector', unlike the first two algorithms presented above. Algorithms of this type treat object detection as a simple regression problem by taking an input image and learning the class probabilities and bounding box coordinates directly (Soviany and Ionescu, 2018). Its architecture is composed of a backbone that takes input images, builds feature maps at different scales and generates region proposals for each scale in the form of anchors (Lin et al., 2017b). These anchors are then used as inputs for two sub-networks, the first one classifies the object and the second one simultaneously performs the regression of the bounding boxes. RetinaNet was used by Eikelboom et al. (2019) for the detection and identification of three African mammal species based on oblique aerial images. This algorithm was therefore chosen to evaluate its performance on nadir UAV images.

For all three algorithms, the backbone consists of a ResNet-101 (He et al., 2016) connected to a feature pyramid network (Lin et al., 2017a). These algorithms were

used through their implementation in the adapted mmdetection toolbox version 1.0.0 (Chen et al., 2019) with PyTorch 1.4.0, TorchVision 0.5.0, OpenCV 4.4.0, MMCV 0.6.0, CuDNN 7.6.3 and Magma 2.5.1 libraries. All the codes and libraries were implemented and transcribed into Jupyter notebooks to run on Google Colaboratory. Training and detection runs were then performed with an NVIDIA Tesla P100-PCIE 16GB GPU running on an Ubuntu 18.04 LTS Colab Linux platform, with CUDA 10.1.243. These were followed by statistical tests conducted using Python's SciPy 1.5.4 library.

2.2.2. Image subdivision and stitching algorithm on the general dataset

All the images were cut into sub-frames of 2000 × 2000 pixels, the maximum size that can be supported by the GPU memory. During the subdivision process, some individuals were cut into several parts and some of them no longer appeared in their entirety. Only individuals whose partial bounding box represented more than 25% of the original surface area were kept. This limit was chosen because below this threshold, individuals are difficult to identify manually.

Only sub-frames containing animals were kept for training (**Figure 2.1**). For the validation and the test sets, the cutting was done with an overlap of 50% on each edge of the sub-frames, and all sub-frames were kept. These steps were taken in order to avoid missing any individuals and to ensure that each individual would appear in its entirety in at least one sub-frame. Moreover, this approach allowed the predictions to be stitched into the initial image frame.

To both eliminate unnecessary partial bounding boxes and to reassemble the sub-frames, a stitching algorithm was constructed. Each image first undergoes a subdivision into overlapped sub-frames of 2000 × 2000 pixels. These sub-frames are then passed through the trained algorithm (i.e. model) to obtain predictions that contain bounding boxes, species names and confidence scores. The coordinates of the predicted bounding boxes of each sub-frame are then modified to be placed in the initial image plan. Next, the NMS (non-maximum suppression) algorithm is applied to a filter the predicted bounding boxes based on the IoU criteria (Everingham et al., 2010):

$$IoU = \frac{area(box\ A \cap box\ B)}{area(box\ A \cup box\ B)}$$

Here, a threshold of 0.5 was chosen. This high threshold was deliberately chosen in order to avoid missing some individuals in herds or some juveniles that are very close to their mothers (**Appendix A2**).

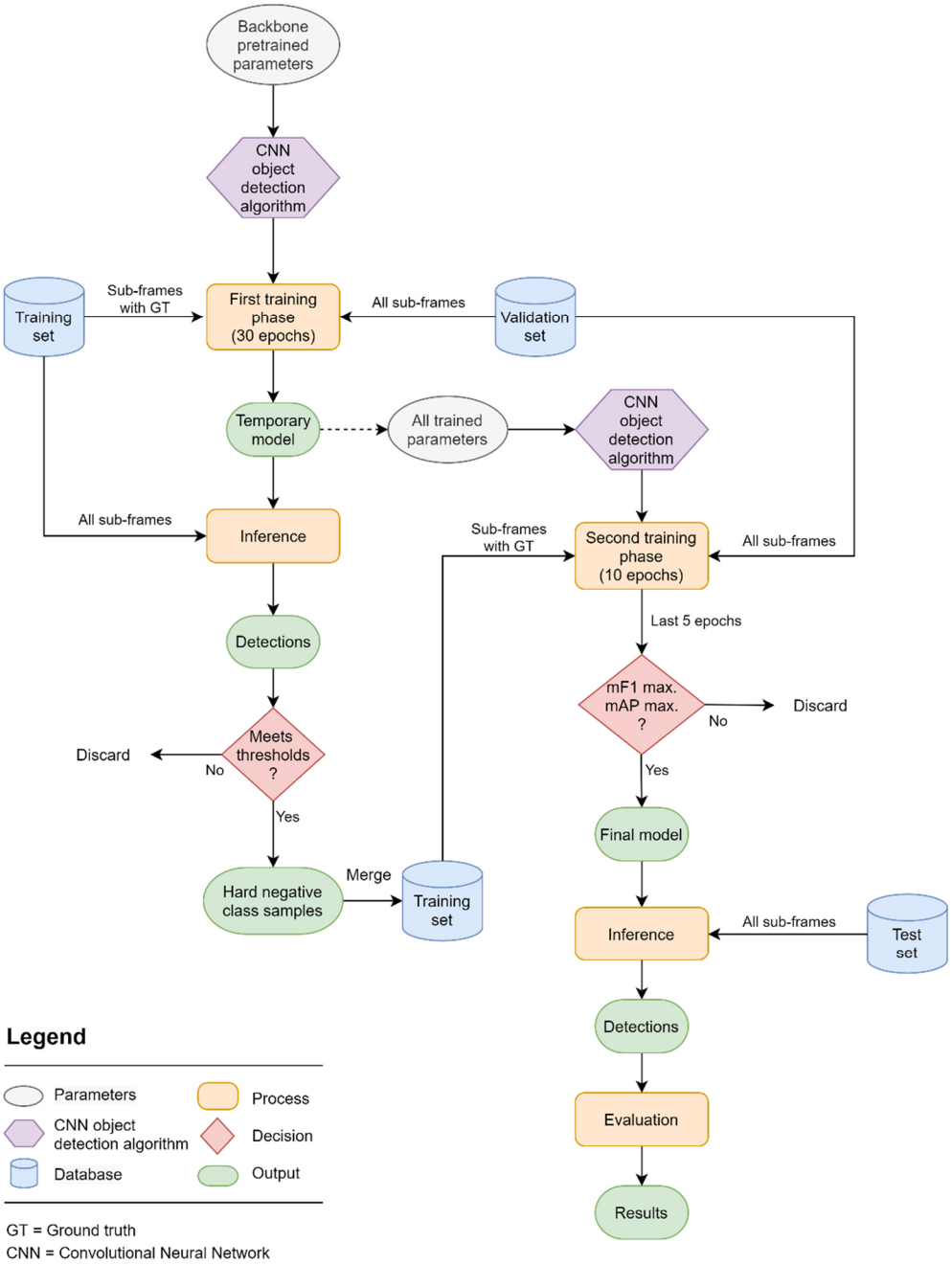


Figure 2.1: Flowchart of the methodology used to train, validate and test each of the three object detection algorithms, using the general dataset. The results after evaluation were then used for comparison.

2.2.3. Training on the general dataset

Because of the imbalance of the different classes, it is possible that during training, the majority species in terms of numbers dominate the others, leading to a decrease in the performance for minority species. Moreover, since the size of the training dataset is relatively small, training the different algorithms from scratch could lead to serious overfitting problems. To overcome these problems, four techniques were used: fine-tuning, data augmentation, class weighting and the hard negative class.

Fine-tuning

Each algorithm's backbone was initialized by pre-trained parameters (**Figure 2.1**) on the ImageNet training dataset (Russakovsky et al., 2015). Next, all the parameters, except the first layer of the backbone, were trained on our dataset, with an adjustment of the number of classes at the head of the network.

Data augmentation

In addition to common strategies (i.e. rotation, mirroring and flipping, horizontal and vertical views), we used other strategies to detect animals in the various situations that can be encountered in aerial images: random blur, random contrast and random brightness.

Class weighting

Used by Kellenberger et al. (2018), this technique led to an improvement in animal detection performance. In our study, satisfactory results were obtained by weighting the species-related terms in the class loss function according to

$$w_i = \frac{\min(\{n_1, \dots, n_i, \dots, n_k\})}{n_i}$$

where n_i is the number of annotations within a class i in the sub-frames training set, and k is the number of classes.

Hard negative class

The hard negative class (Peng et al., 2020) was used to limit the number of false positives (FP) (**Figure 2.1**). This method treats hard negatives (high-scoring FPs) as foreground objects to make the model more sensitive to them. The score threshold was chosen to have a class size of between 2000 and 2200 to avoid a too-high class imbalance. Note that for the validation and test sets, the hard negative class-predicted bounding boxes were discarded and only species classes were maintained. Preliminary analysis showed that the inclusion of the hard negative class increased the models' performance (**Appendix A3**).

The training for the first training phase was done during 30 epochs with the Stochastic Gradient Descent as an optimizer (a momentum of 0.9 and a weight decay of 10^{-3}), and with a learning rate decreasing from 10^{-3} to 10^{-5} by steps of 10 epochs. The hard negative class was included from the 31st epoch (second training phase, **Figure 2.1**,

Appendix A4) and training continued for 10 more epochs with a learning rate of 10^{-4} for the first five epochs and 10^{-5} for the last five epochs.

2.2.4. Evaluation of CNN models

For each complete image, a comparison between ground truth and predictions was performed in order to determine the true positives (TP), FP and false negatives (FN).

A detection was then considered as a TP if the labels between the predicted bounding box and the ground-truth bounding box correspond and if the IoU between these boxes exceed 0.30. When several detections overlapped the same ground truth, the one with the maximum IoU was selected and the others were then considered as FP. Finally, if the labels did not match or if the ground truth was not detected by the model, the ground truth was considered as FN.

Precision/recall curves for each species were constructed to evaluate the performance of each model. These curves were calculated by varying the confidence score threshold associated with each predicted bounding box, between 0 and 1:

$$p(k) = \frac{n_{TP}(k)}{n_{TP}(k) + n_{FP}(k)}$$

$$r(k) = \frac{n_{TP}(k)}{n_{TP}(k) + n_{FN}(k)}$$

where p is the precision and r the recall, k is the confidence score threshold, and n_{TP}, n_{FP} and n_{FN} are the numbers of the TP, FP and FN, respectively.

F1 scores are usually used to define the combination of precision and recall that produces an optimal compromise between the number of FP and FN. An F1 score essentially represents the harmonic mean of precision and recall:

$$F1\ score = \frac{2 * p * r}{p + r}$$

where p and r are the precision and recall, respectively.

In this study, we used a mean F1 score (mF1): F1 scores were calculated for each species and then the whole was averaged. This metric was used because it gives an overall idea of the compromise between the FP and the FN.

The average precision (AP) (Everingham et al., 2010), representing the area under the precision/recall curve, was then calculated for each animal species in order to evaluate the performance of the detection algorithm in detecting a particular species. Finally, the mean average precision (mAP) was calculated to quantify the overall performance of each detection algorithm and thus allow their comparison. The mAP represents the average AP of all the species.

Each algorithm was trained for five runs with different fixed seeds. This step allowed us to control the stochastic aspect related to the training of an object detection

algorithm. From these five runs, paired sample *t*-Student tests and confidence intervals were computed to compare the models and determine if the differences in performance were significantly different.

After each epoch, each trained algorithm (i.e. each model) was saved and tested on the stitched validation image set to verify that it was not falling into overfitting. In addition, for the last five epochs of each three algorithms of interest, the model with the best performance on the validation set was selected for testing (**Figure 2.1**). To determine the best performance, the epoch with the maximum mF1 score was first selected. Next, the mAP corresponding to the epochs that presented an equivalent mF1 value (i.e. with two significant digits retained) was analyzed. From among these, the epoch with the highest mAP was finally selected for testing. This method enabled the selection of a globally efficient model (i.e. a high mAP) with a good compromise between FP and FN (i.e. a high mF1 value).

2.2.5. Processing of the case study dataset

To choose the model to apply to the case study (the Garamba dataset), we first selected the algorithm that showed the best performances on the test set. Then, we selected the best model based on the five tested runs, using the same selection method as in the validation set. The Garamba dataset's images were previously cut into sub-frames according to the same methodology as the validation and test sets (see Section 2.2.2). Detections were then stitched together according to an inference approach using the same stitching algorithm and evaluated using the same evaluation methodology as for the general dataset (see Section 2.2.3). Note that due to the high similarity between the two species (see **Appendix A1**) and the impossibility to distinguish them on UAV images, hartebeests and topis were merged into the same class during the inference step.

3. Results

3.1. Species detection

Topi, buffalo and kob were very well detected by all the trained algorithms (i.e. the three models) studied, with only slightly poorer results for elephants (**Figure 2.2**). Given the results, warthogs and waterbucks appear to be more difficult to detect. Nevertheless, waterbucks were very well detected by Libra-RCNN (AP = 0.89) but very poorly detected by RetinaNet (AP = 0.01). RetinaNet was the model that had the most difficulty in detecting minority species (warthogs and waterbucks).

False positives were particularly high for elephants and warthogs, for all the models, as indicated by the poor precision at the highest recall value of these species (**Figure 2.2**).

Libra-RCNN was the model that presented the highest AP for each of the species, except for elephants, where it equalled the AP of the Faster-RCNN model.

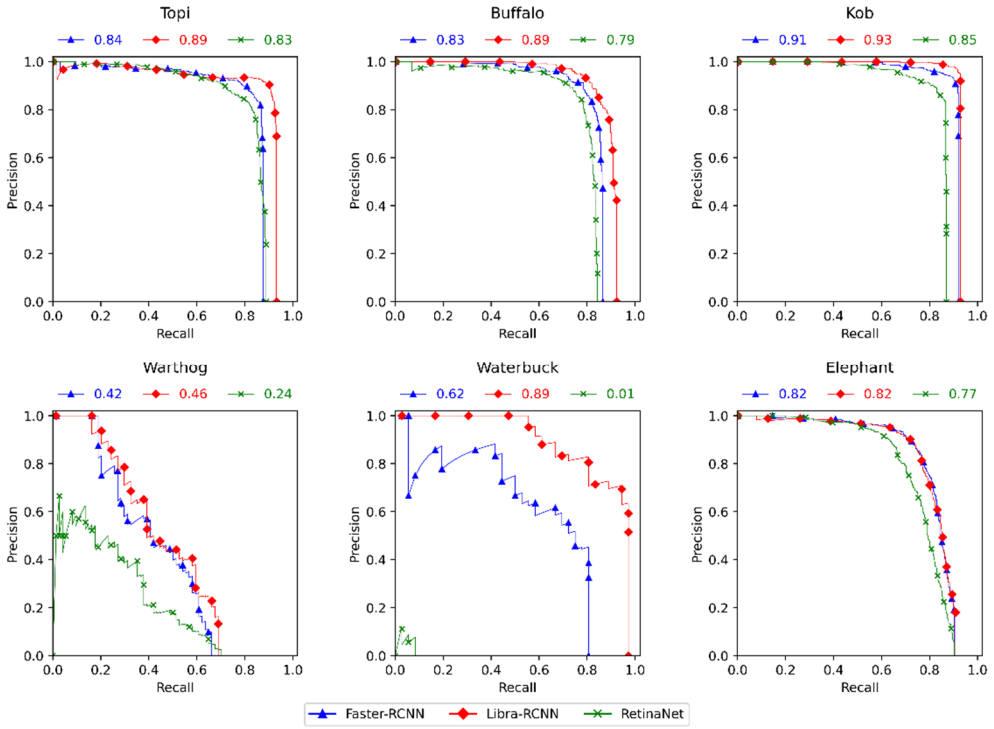


Figure 2.2: Precision/Recall curves of the three detection algorithms for the six targeted species on the test set. Axis legend represents the average precision (AP) of the corresponding curve. These curves were calculated for each of the algorithms using the model with the best mean average precision (mAP) among the five seeds.

3.2. Model comparison

The results of the independent *t*-Student tests showed a significant difference in performance on the test set between the three models for mAP, mF1 and mean interspecies confusion, but not for recall. There was a significant difference in the FP/TP ratio for Faster-RCNN and Libra-RCNN with RetinaNet but not between Faster-RCNN and Libra-RCNN (see **Appendix A5** for details).

The Libra-RCNN model produced the best mAP, the best mF1 score and the lowest average level of interspecies confusion in the test set (**Figure 2.3**). In contrast, the RetinaNet model had the lowest mAP and mF1 values, along with the highest average interspecies confusion score. Finally, Faster-RCNN's performance ranks it between the other two.

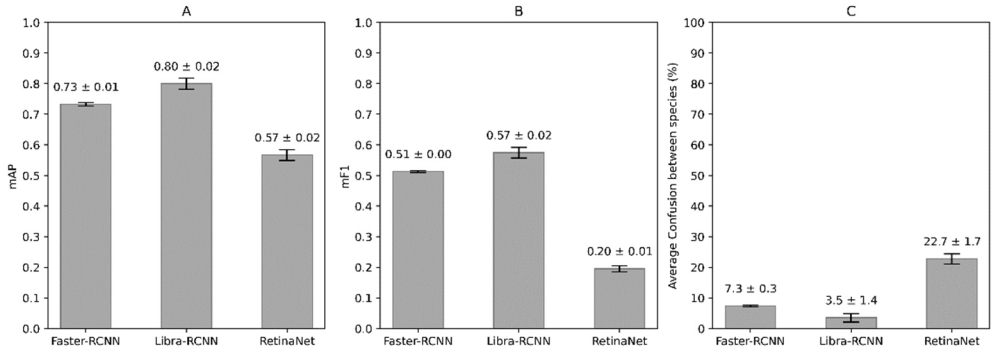


Figure 2.3: Bar plots of mAP (A), mF1 (B) and average interspecies confusion (C) calculated from the detection results of the test set. The error bars represent the 95% *t*-Student confidence interval (4 d.f.), computed from the results of the five seeds. mAP, mean average precision.

Regarding the percentage of animals detected, all three models detected on average the same percentage of animals (true detection rate), with 94.5% (±0.5%) for Faster-RCNN, 94.3% (±0.5%) for Libra-RCNN and 94.6% (±0.3%) for RetinaNet, where the confidence intervals represent the 95% *t*-Student confidence interval (4 d.f.), computed from the results of the five seeds.

Finally, in terms of the FP/TP ratio (binary case), Libra-RCNN presented the lowest value (1.7 ± 0.2), closely followed by Faster-RCNN (1.8 ± 0.1). RetinaNet had the highest ratio with an average of nearly nine FN per TP (9.0 ± 0.7), a very high number of false alarms.

These results suggest that the Libra-RCNN model is more suitable for multi-species animal detection than the other two, with superior detection performance compared to Faster-RCNN and RetinaNet. Therefore, Libra-RCNN was selected and applied to the case study dataset.

3.3. Case study (Garamba dataset)

To evaluate the performance of the best developed model, the Libra-RCNN model was applied on the Garamba dataset. The total processing time (with a single GPU) was 23h26 for all the flight images, with an average of 12 seconds/image. Detections were present in 9% of the images ($607/7034 \approx 0.09$), and among the 180 images containing ground truths, 9% were missed by the model ($16/180 \approx 0.09$). However, no or almost no images were missed for some species (**Table 2.3**). For all six targeted species, 73% were correctly identified, with a relatively wide variation between the species. Furthermore, 64 individuals were correctly detected but misidentified. The same trend in species detection as observed on the test set results can be observed here as well: the majority species are better detected and identified than the minority species (**Table 2.3**).

Table 2.3: Results of the Libra-RCNN model applied on the case study dataset (Garamba) for the six targeted species: hartebeest (considered as topi due to high similarity), buffalo, kob, warthog, waterbuck and elephant.

Species	Number of Images		Individuals					Number of False Positives			
	With GT	With Detections[1]	Missed	GT	Recall	Precision	F1 score	Misclassified	Total	Human-missed[2]	Other species[3]
Hartebeest	29	102	0	151	0.59	0.34	0.43	45	174	4	1
Buffalo	55	148	5	547	0.87	0.44	0.58	7	620	19	10
Kob	62	95	1	321	0.67	0.62	0.64	5	133	26	6
Warthog	24	158	9	82	0.40	0.09	0.14	0	349	6	18
Waterbuck	10	122	1	14	0.14	0.01	0.03	7	144	0	6
Elephant	0	54	0	0	n/a	n/a	n/a	0	171	0	2
All	180	607	16	1115	0.73	0.34	0.46	64	1591	55	43

The last row corresponds to the results of the whole set of six species. Note that the number of images with detections considering all six species (last row) is not equal to the sum of the images with detections by species. This difference is due to the fact that the model sometimes detected several species within the same image, and so these images appear in the detected images of multiple species.

GT, Ground Truth; n/a, not applicable (since this species is absent from the dataset).

[1]Number of images with detections (i.e. that contain predictions).

[2]Number of false positives that were in fact animals missed by human during annotation, but correctly detected and identified by the model.

[3]Number of other species not belonging to the set of six targeted species (i.e., hippopotamus, giraffe, and unknown), but detected by the model.

Among the 305 individuals of other species initially identified during the annotation step, 43 were found by the model: 29 hippopotamuses out of 196, 13 giraffes out of 43 and 1 undetermined species out of 7. In addition, all false positives with an IoU of 0 with the ground truths were reviewed; among these 945 FP detections, 133 were in fact individuals of our six targeted species that were missed during the annotation phase. Of these, 55 were correctly identified by the model (**Table 2.3**).

4. Discussion

The Libra-RCNN model showed better detection performance on the test set than other published models dealing with the detection of mammals in similar habitats and landscapes (Eikelboom et al., 2019; Kellenberger et al., 2018; Rey et al., 2017). Moreover, the models presented here were able to differentiate six animal species on nadir aerial images, which to the best of our knowledge has never been tried before in the literature. The performance of our best model (Libra-RCNN) on the test set surpasses that of the latest multi-species model published (Eikelboom et al., 2019) in terms of global recall, global FP/TP ratios, mAP and F1 scores. Finally, it showed good performance on a complete independent raw dataset from another park (i.e. Garamba) and was able to detect additional individuals, some belonging to other species.

4.1. Species detection

Our best model, the Libra RCNN, showed very good detection, identification and generalization results for the majority species (topi, buffalo and kob) and was even surprisingly good at detecting one minority species, the waterbuck. For topi, buffalo and elephant detection, we observed that all three models were less precise for herds. The lower precision was mainly due to the overlap of the bounding boxes within the herds (**Figure 2.4**). Indeed, this box overlap probably made it more difficult for the algorithms to converge during training. In images containing herds, a large number of boxes were therefore created during the inference step, but despite the application of the NMS, some detections persisted and were therefore qualified as FP, as several boxes defined the same individual. After revision, herding represented about 40% of the FPs for the Libra-RCNN model on the test set, and about 41% for that model on the Garamba dataset.

In addition, for elephants, the images were taken at any time of the day, unlike the other datasets. This led to greater variability of shadows, colours and brightness within the images, and thus to poorer detection results, as observed in Rey et al. (2017). Moreover, this dataset (AED) comes from parks and reserves with varying landscapes and terrain features that differ from those of Virunga, such as denser tree cover in some images. However, training the models on these field variations normally made them more robust to heterogeneous terrain features (Kellenberger et al., 2018).

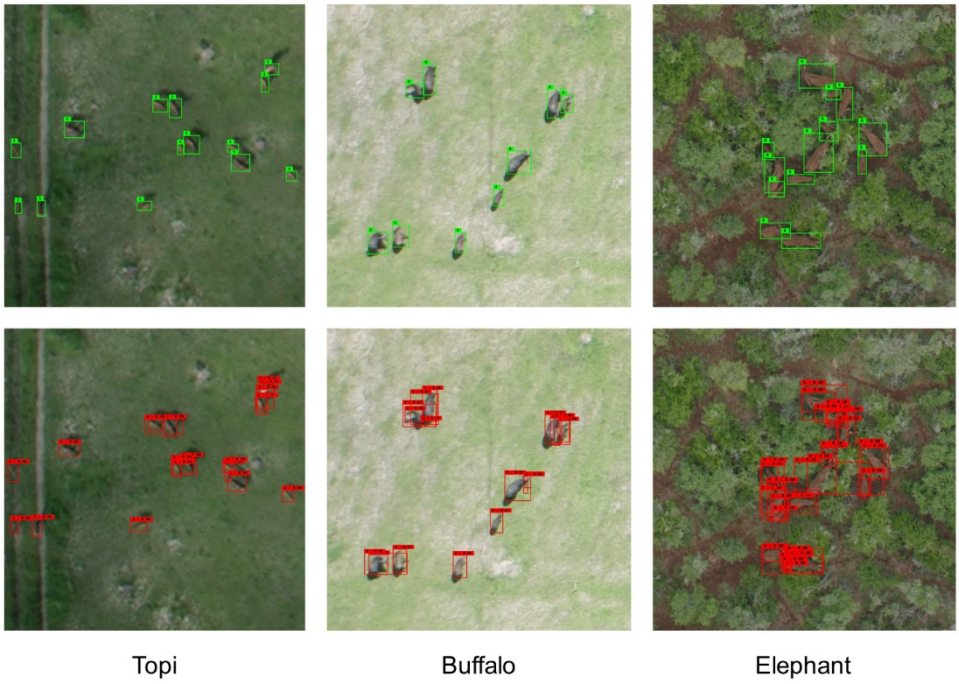


Figure 2.4: Detections examples of the Libra-RCNN model, on partial test images showing the major cause of the high number of false positives. Note that ground truths are in green (first row) and detections are in red (second row).

This difference in the landscape also explains the lower percentage of animals detected by Libra-RCNN on the Garamba dataset. In addition, we observed that these differences in terrain features were also the cause of many FPs within the Garamba images. For example, the model detected a large number of termite mounds as animals. This terrain characteristic was indeed much less present in the training set.

Despite the class weighting during training, the models struggled to correctly identify warthogs, most probably due to a lack of training samples. Furthermore, this animal was the smallest mammal in this study. Its small size generated a large number of FP due to insufficient pixel resolution and because some acquisition drawbacks (blur, contrast) did not allow the model to distinguish some of this small mammal's attributes. It could therefore easily be mistaken for small rocks, common in the African landscapes where this species is found.

The surprisingly high number of the Garamba dataset's FPs (133) that were in fact real animals can be explained by the overlap of the images and the initial methodology of annotating the individuals. Indeed, only 84 individuals were actually real human-missed animals. The other animals had already been tagged during the manual annotation phase in the previous frame or would be in the next frame within a

succession of overlapping images. Annotating everything was not required, although recommended for the purpose of that specific survey, as double counting was not desired at the time of the census, despite the possibilities to differentiate between first observation and double counting in the software. Consequently, attention was focused on the individuals that had not yet been tagged, and so sometimes the individuals present in the periphery were not tagged again. From these 84 new individuals, 55 were correctly identified by the model.

4.2. Model comparison

Two-stage detection models (Faster-RCNN and Libra-RCNN) seemed to detect animal species more precisely than a single-stage model (RetinaNet). This difference in performance was as expected (Soviany and Ionescu, 2018).

The Faster-RCNN and Libra-RCNN models were very similar in terms of their detection performance. The differences that we observed between these two models on the test set (**Figure 2.3**) were probably due to the Libra-RCNN L1-balanced loss and its rebalancing at the training sample distribution level. These components caused the algorithm to focus on difficult cases during training, which leads to better detection and classification performances (Pang et al., 2019).

4.3. Operational implications

The Libra-RCNN model presents interesting perspectives as a good semi-automatic detection and identification tool for African mammal species. It could be used in practice to save human time, create new training data and establish initial, rapid population counts, with human verification of detected individuals as post-processing. However, our experience in reviewing the FPs shows that this screening must necessarily be performed with the animals' surrounding context, which is crucial for decision making by the human eye.

The model developed here can mainly be applied in open savanna or sparsely wooded areas and for the detection of our six studied species. Indeed, our results show that in order to develop a model that can be used in various ecosystems, it would be necessary to have a training set with a large variability of landscapes and terrain features.

Generally, detection performance improves when more training data are used. Unfortunately, the acquisition and pre-processing of aerial animal training data are costly. Developing a semi-automatic animal detection tool, such as those presented here, requires significant upstream work. From the manual identification and location of animals to the dataset training, the workload is quite large and requires significant human resources with highly technical skills. Moreover, as with any deep-learning application, training an algorithm requires a large computing capacity and a huge amount of data. Luckily, more and more open-source data (images and annotations) are being made available (Eikelboom et al., 2019; Kellenberger et al., 2018; Naudé and Joubert, 2019).

Finally, in an attempt to automate the counting of individuals, the thorny problem of images overlap remains an obstacle. Our results from Garamba were presented here without accounting for multiple detections. We observed that this overlap is crucial to detect all possible individuals. Indeed, in Garamba, some individuals were only detected in a few images thanks to a slight change in the viewing angle (**Figure 2.5**). This need for overlap leads the model to slightly overestimate the number of real FN.

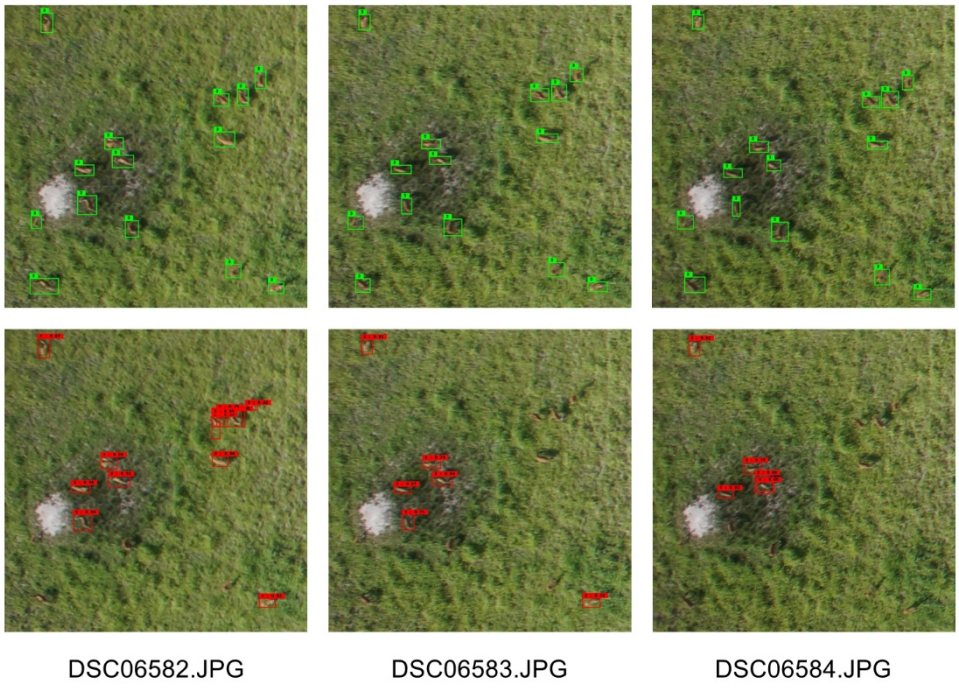


Figure 2.5: Kob detections of the Libra-RCNN model on consecutive Garamba's partial images, showing that the images overlap made it possible to detect a maximum of individuals thanks to the slight viewing angle changes. Note that ground truths are in green (first row) and detections are in red (second row).

4.4. Research perspectives

In surveying animal species, the problem of class imbalance will always be present due to the natural distribution of species within ecosystems. Nevertheless, more and more studies are looking into this recurrent problem in multi-class object detection (Oksuz et al., 2021). There is also the challenging problem of the large number of FP. Newer methods, such as synthetic data generation, could help to address this problem by generating images with heterogenous backgrounds (Beery et al., 2020). In addition, it could be beneficial to consider switching from boxes to points (Ribera et al., 2019) or masks (Xu et al., 2020) to avoid the problems of overlapping boxes in herds and in an attempt to automate the counting. These solutions should be investigated in future works.

5. Appendices

A1: Image samples of the six targeted species

A few samples of each of the six targeted species from the general dataset are shown in **Figure 2.6**. In addition, a few samples of hartebeest from Garamba have been added to the last column to show their similarity with the topis (from Virunga). This illustrates the need to group these two species in the same class when analyzing the Garamba results.

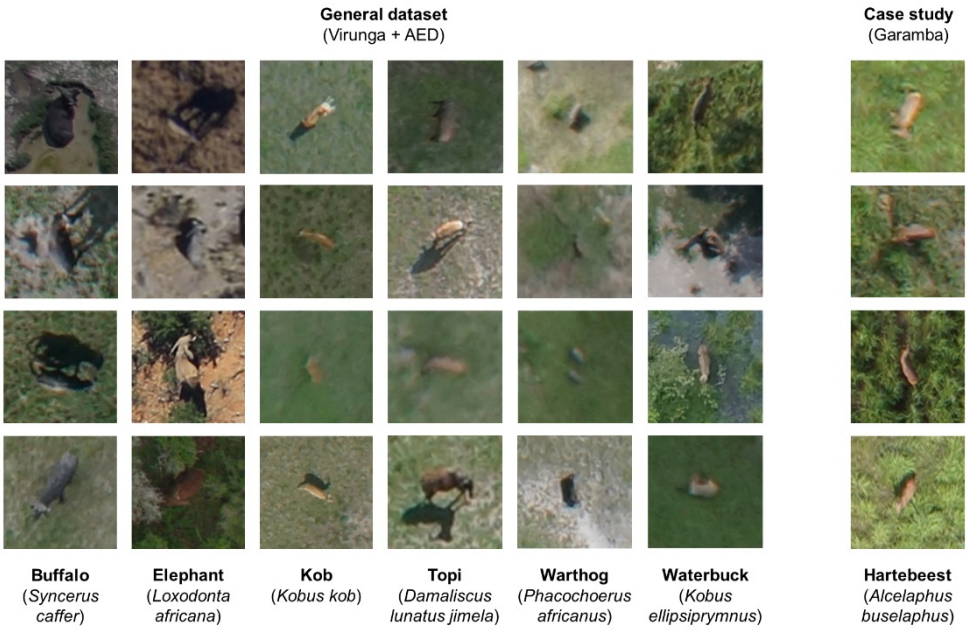


Figure 2.6: Image samples of each of the six targeted species from the general dataset, and a few samples of hartebeest from the case study (Garamba) in the last column.

A2: Preliminary tests for NMS (Non-Maximum Suppression) threshold

To retain a maximum number of individuals, especially close-by animals (i.e. juveniles and herds), the IoU (Intersect-Over-Union) threshold giving the maximum recall on the validation set was selected for NMS during preliminary tests, while keeping an eye on mAP to avoid a too large drop of its value (**Figure 2.7**). In a semi-automated survey context, it was therefore preferred to detect a maximum number of individuals at the expense of a very slight drop in mAP. This IoU threshold was 0.5 for all three models.

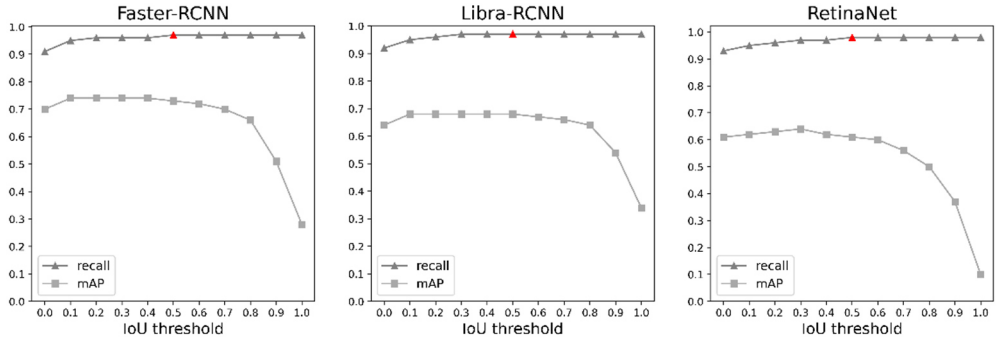


Figure 2.7: Evolution of recall and mAP values obtained on the validation set, according to different IoU thresholds used for the NMS process. The maximum recall value is shown in red. It was obtained with a threshold of 0.5 for each model. The mAP curve shows that this threshold selection did not have much impact on the global performances of the models (slight decrease of mAP values).

A3: Comparison of models' performances with and without the inclusion of the hard negative class

The value of mF1 was significantly higher with the hard negative class for each model, while the average confusion between species was not significantly impacted by this class inclusion (**Figure 2.8A-B**). If the detection is considered as a binary case (animal vs. background), it is clear that the ratio of false positives to true positives was significantly lower for each model (**Figure 2.8C**). However, the percentage of animals detected decreased significantly from 96.6 to 95.7% for Faster-RCNN ($t(4)=9.02$, $p=0.001$), from 97.5 to 96.4% for Libra-RCNN ($t(4)=4.78$, $p=0.009$) and from 97.8 to 97.1% for RetinaNet ($t(4)=5.04$, $p=0.007$). The results of the validation therefore show that the hard negative class significantly improved the results for each of the models studied without a decrease in interspecies average confusion, but at the expense of detecting slightly fewer individuals.

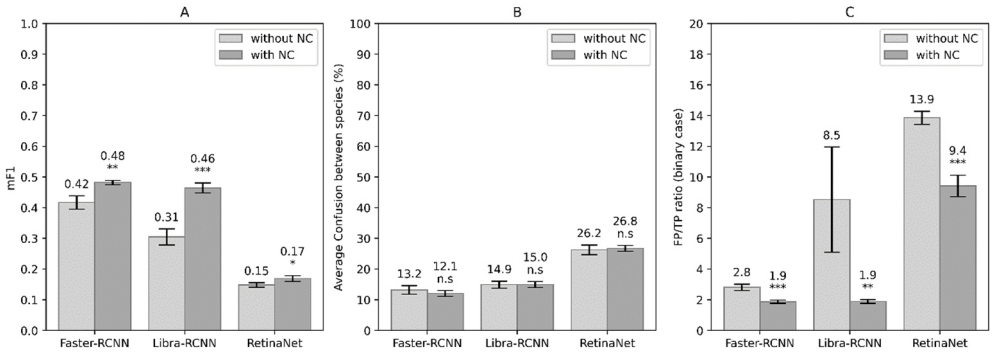


Figure 2.8: Bar plots of mF1 values for (A) average interspecies confusion, (B) the false positives-true positives ratio, and (C) showing the effect of the hard negative class (NC) on the validation set for the three models tested. The stars indicate the level of significance of the paired sample t-Student test (***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$, and 'n.s', $p > 0.05$) and the error bars represent the 95% t-student confidence interval (4 d.f.) computed from the results of the five seeds. Note that the difference distribution of each metric underwent a Shapiro-Wilk test for normality, and each difference distribution accepted the null hypothesis ($p > 0.05$).

A4: Detection algorithms' training details

Table 2.4 gives the details of the training parameters for each of the detection algorithms used. The training was done over 40 epochs with a batch size of one sub-frame. Class weighting was applied to all the algorithms. The reduction of the learning rate (10^{-3} to 10^{-5}) was made by steps of 10 epochs for the first 30 epochs. The data augmentation consists of a horizontal flip with a probability of 0.5, a vertical flip (prob=0.5), a 90-degree rotation (prob=0.5), a random blur with a maximum kernel size of 15 pixels (prob=0.2), and a random adjustment of contrast and brightness (prob=0.2).

Table 2.4: Training details of the detection algorithms.

	Faster-RCNN	**Libra-RCNN**	**RetinaNet**
Backbone	ResNet-101-FPN	ResNet-101-FPN	ResNet-101-FPN
Class loss	Cross entropy	Cross entropy	Focal
Bounding box loss	Smooth-L1	Balanced-L1	Smooth-L1

The hard negative class was applied from epoch 31 with a learning rate of 10^{-4} for the first five epochs, and 10^{-5} for the last five epochs. The method of Peng et al. (2020) has been slightly modified here. Instead of harvesting FPs at each iteration, we harvested FPs only once at the end of the first training step, and then trained the model on them for 10 epochs to minimize disruption of the algorithm during training. Finally, the optimizer is the well-known stochastic gradient descent (SGD), with a momentum of 0.9 and a weight decay of 10^{-3}.

A5: Independent t-Student test results on the test set (general dataset)

The results of the t-Student tests performed on the test set, as well as the verification of the conditions of application (i.e., the normality of the samples and the equality of variances) are presented in **Tables 2.5 and 2.6**. These tests were conducted using the Python's SciPy 1.5.4 library.

The results showed that the difference in performance between the three models was significant for most of the metrics (**Table 2.5**). Only recall values were insignificantly different, as well as the binary ratio (animal vs. background) of false positives to true positives (FP/TP) between Faster-RCNN and Libra-RCNN.

Table 2.5: Model results of the Shapiro-Wilk tests on the values of each of the metrics, obtained following the five runs performed on the test set. The "*sign.*" column indicates the level of significance of the test (***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$, and 'n.s.', $p > 0.05$). Each test was non-significant, meaning that the null hypothesis was accepted (H_0: the values follow a normal distribution.

Model	Metric	df	statistic	p-value	sign.
Faster-RCNN	mAP	4	0.90	0.436	n.s.
	mF1	4	0.82	0.118	n.s.
	confusion	4	0.94	0.689	n.s.
	recall	4	0.96	0.824	n.s.
	FP/TP	4	0.84	0.153	n.s.
Libra-RCNN	mAP	4	0.86	0.232	n.s.
	mF1	4	0.89	0.333	n.s.
	confusion	4	0.82	0.111	n.s.
	recall	4	0.83	0.135	n.s.
	FP/TP	4	0.84	0.177	n.s.
RetinaNet	mAP	4	0.93	0.563	n.s.
	mF1	4	0.98	0.910	n.s.
	confusion	4	0.92	0.534	n.s.
	recall	4	0.94	0.679	n.s.
	FP/TP	4	0.89	0.381	n.s.

Table 2.6: Results of Levene and independent t-Student tests for model comparison. The "*sign.*" column indicates the level of significance of the test (***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$, and 'n.s.', $p > 0.05$). Note that nearly all the Levene tests were non-significant, meaning the null hypothesis was accepted (H_0: the variances of the two group are equal). Since the "*Faster-RCNN vs RetinaNet – FP/TP*" Levene test was significant ($p < 0.05$), a Welch t-test was performed for this case, instead of the standard two-sample t-Student test performed for all the other cases.

Comparison	**Metric**	**df**	**Levene test**			**Independent t-Student test**		
			statistic	p-value	sign.	statistic	p-value	sign.
Faster-RCNN vs Libra-RCNN	mAP	4	1.75	0.222	n.s.	-9.91	< 0.001	***
	mF1	4	1.94	0.201	n.s.	-9.68	< 0.001	***
	confusion	4	1.55	0.248	n.s.	7.36	< 0.001	***
	recall	4	0.02	0.899	n.s.	0.80	0.449	n.s.
	FP/TP	4	0.88	0.377	n.s.	2.07	0.072	n.s.
Faster-RCNN vs RetinaNet	mAP	4	2.57	0.147	n.s.	25.06	< 0.001	***
	mF1	4	2.36	0.163	n.s.	85.42	< 0.001	***
	confusion	4	1.76	0.221	n.s.	-25.09	< 0.001	***
	recall	4	1.18	0.309	n.s.	-0.71	0.496	n.s.
	FP/TP	4	8.11	0.022	*	-29.45 (Welch)	< 0.001 (Welch)	***
Libra-RCNN vs RetinaNet	mAP	4	0.01	0.942	n.s.	25.81	< 0.001	***
	mF1	4	0.32	0.588	n.s.	52.58	< 0.001	***
	confusion	4	0.00	0.955	n.s.	-24.48	< 0.001	***
	recall	4	0.66	0.439	n.s.	-1.57	0.156	n.s.
	FP/TP	4	5.12	0.053	n.s.	-29.00	< 0.001	***

3

Designing a CNN for precise counting of African mammals

Preamble

In **Chapter 2**, I showed that pre-existing CNN-based object detectors have good detection potential but also have certain limits, particularly when it comes to precisely counting animals in densely clumped herds. In response, there was a need to design and develop a DL architecture adapted to the challenges experienced in the context of counting African mammals using aerial imagery, which can be further accentuated with oblique cameras. In this chapter, I thus present the proposed architecture, named *HerdNet*, which was inspired by advances made in the field of crowd counting that bear many similarities to our application case. Its advantages are highlighted, and limitations are discussed. Outperforming alternative CNN-based counting approaches in terms of location, count and processing speed, regardless of proximity between animals, HerdNet is finally being discussed for its practical implications in the realm of aerial surveys.

Paper 2 | Published

From crowd to herd counting: How to precisely detect and count African mammals using aerial imagery and deep learning?

Alexandre Delplanque, Samuel Foucher, Jérôme Théau, Elsa Bussière, Cédric Vermeulen & Philippe Lejeune

This paper is published in *ISPRS Journal of Photogrammetry and Remote Sensing* (IF=10.6), 197, 167-180. DOI: 10.1016/j.isprsjprs.2023.01.025

Abstract

Rapid growth of human populations in sub-Saharan Africa has led to a simultaneous increase in the number of livestock, often leading to conflicts of use with wildlife in protected areas. To minimize these conflicts, and to meet both communities' and conservation goals, it is therefore essential to monitor livestock density and their land use. This is usually done by conducting aerial surveys during which aerial images are taken for later counting. Although this approach appears to reduce counting bias, the manual processing of images is time-consuming. The use of dense convolutional neural networks (CNNs) has emerged as a very promising avenue for processing such datasets. However, typical CNN architectures have detection limits for dense herds and close-by animals. To tackle this problem, this study introduces a new point-based CNN architecture, HerdNet, inspired by crowd counting. It was optimized on challenging oblique aerial images containing herds of camels (*Camelus dromedarius*), donkeys (*Equus asinus*), sheep (*Ovis aries*) and goats (*Capra hircus*), acquired over heterogeneous arid landscapes of the Ennedi reserve (Chad). This approach was compared to an anchor-based architecture, Faster-RCNN, and a density-based, adapted version of DLA-34 that is typically used in crowd counting. HerdNet achieved a global F1 score of 73.6 % on 24 megapixels images, with a root mean square error of 9.8 animals and at a processing speed of 3.6 s, outperforming the two baselines in terms of localization, counting and speed. It showed better proximity-invariant precision while maintaining equivalent recall to that of Faster-RCNN, thus demonstrating that it is the most suitable approach for detecting and counting large mammals at close range. The only limitation of HerdNet was the slightly weaker identification of species, with an average confusion rate approximately 4 % higher than that of Faster-RCNN. This study provides a new CNN architecture that could be used to develop an automatic livestock counting tool in aerial imagery. The reduced image analysis time could motivate more frequent flights, thus allowing a much finer monitoring of livestock and their land use.

Keywords: Deep learning, Livestock, Herd, Convolutional neural networks, Aerial survey, Protected area

1. Introduction

In sub-Saharan Africa, the rapid growth of the human population over the last decades, combined with very effective sanitary actions on herds, has led to a significant increase in the number of heads of different livestock species (Richard et al., 2019). On the one hand, excessive livestock density can have several adverse effects on the environment, such as soil and vegetation degradation, space and grazing competition with wildlife or spread of diseases (Bengis et al., 2004; Butt and Turner, 2012; De Leeuw et al., 2001; Georgiadis et al., 2007; Vandermeer, 2002). On the other hand, livestock is a major source of income and a livelihood strategy for rural populations (Herrero et al., 2013), and it can enhance agricultural sustainability (Ayantunde et al., 2018) and habitat quality for wildlife if well managed (Fynn et al., 2016). Too-high density of livestock may prompt conflicts over important natural resources within a protected area, such as pastures used for grazing by wild and domestic herbivores (Scholte et al., 2022a, 2022b; Toutain et al., 2004). Knowledge of livestock density and land use in these areas is therefore necessary to reach both conservation and local communities' goals.

In large open African areas, livestock and wildlife counting are often carried out by a piloted aircraft, flying at low altitude and following systematic transects while observers count animals in sample strips defined on each side of the aircraft (Caughley, 1977; Grimsdell and Westley, 1981; Norton-Griffiths, 1978). Unfortunately, observers tend to fail to detect and accurately count the true number of animals in the strips, especially when encountering large and dense herds, resulting in biased population estimates (Caughley, 1974; Grimsdell and Westley, 1981; Jachmann, 2002).

For most observers, remote counting from an aircraft becomes inaccurate for groups of 15 or more individuals (Grimsdell and Westley, 1981; Norton-Griffiths, 1978). Photographing large herds has thus become a common practice to improve group size estimates by subsequent counting (Bouché et al., 2012; Craig, 2012; Grimsdell and Westley, 1981; Norton-Griffiths, 1978; Schlossberg et al., 2016). Recently, the use of oblique cameras has been shown to improve wildlife counts, especially for smaller species such as warthog (*Phacochoerus africanus*), Uganda kob (*Kobus kob*), or oribi (*Ourebia ourebi*) (Lamprey et al., 2020b, 2020a). Although nadir imagery is increasingly used for aerial survey of wildlife since the growing interest for drones (Linchant et al., 2015b), oblique imagery remains a relevant and particularly attractive solution for managers of large protected areas. Oblique imagery has the following advantages over nadir imagery, making it a key research area: the better detection of animals under trees, the better identification of species (side view), the larger sampling area at a same flight height, and the similar viewing configuration with onboard observers (facilitation of detection validation). However, the main drawbacks of this method are: 1) the high volume of imagery generated; and 2) the associated intensive photo-interpretation workload. For instance, Lamprey et al. (2020a) acquired 24,000

images for a survey of a 5037 km^2 reserve in Uganda, and it took 6 weeks for 4 people to interpret the images.

Deep learning architectures, through the use of Convolutional Neural Networks (CNNs), now offer the possibility to semi-automatically detect and identify species in aerial images acquired in heterogeneous landscapes using object detection approaches (Delplanque et al., 2022; Eikelboom et al., 2019; Kellenberger et al., 2019a, 2018, 2017; Naudé and Joubert, 2019; Peng et al., 2020; Torney et al., 2019). These recent approaches allow partially-automated processing of the large volumes of images generated during acquisition campaigns. While these seem to work relatively well for isolated individuals or sparse herds, the case of dense herds remains a complex and challenging task (Delplanque et al., 2022).

In oblique images containing dense herds, factors such as mutual occlusions, close-by bodies, complex background, varying scales, and non-uniform distribution of individuals make common object detection approaches cumbersome if not impossible to accurately locate and count the individuals. Common object detectors are usually anchor-based, meaning that they use anchors during the training process, which are a set of prior box proposals with different scales and aspects centered on potential object locations (Ren et al., 2015). Usually, anchors help the network to converge faster and to obtain better detection performance (Lin et al., 2017b; Liu et al., 2016; Redmon and Farhadi, 2017; Ren et al., 2015). However, they are suspected to be the cause of decreased precision in dense herd situations (Delplanque et al., 2022).

The factors mentioned above (i.e., occlusion, complex background, scale variation and non-uniform distribution) are also encountered in crowd detection (Gao et al., 2020), making the task of herd counting very similar to that of crowd counting. While the CNN architectures developed in crowd counting have shown very good results for human counting in densely populated scenes, their transposition to dense terrestrial mammal herd counting in oblique imagery has not yet been explored.

Density-map-based architectures, first proposed by Lempitsky and Zisserman (2010), are popular in crowd counting, due mainly to their improved counting performance compared to detection-based and anchor-based architectures, and for the practicality of dot annotations (Li et al., 2021). Padubidri et al. (2021) have recently shown that density maps can be used to precisely count Steller sea lions (*Eumetopias jubatus*) and African bush elephants (*Loxodonta africana*) in nadir aerial images. Kellenberger et al. (2019a) also proposed density-based approaches that showed great performances using only image-level annotations. However, density-based approaches did not precisely locate individuals in the images, especially in herds; such location capability could be valuable for creating new annotations from unseen images.

This paper presents “HerdNet”, a new dense herd CNN-based counting approach, inspired and adapted from crowd counting approaches, which was compared with an anchor-based and density-based baselines.

2. Background

2.1. Pointing, a more natural and efficient way for herd counting

In addition to being a natural way to count objects for humans, pointing is faster than drawing bounding boxes, especially when large numbers of objects are encountered, as in the case of animal herds. Pointing was first proposed by Lempitsky and Zisserman (2010), who presented it as a very attractive and understudied case. Since then, point annotations have been largely used for labeling crowds in images (Li et al., 2021). In recent years, some CNN point-based approaches have also emerged with promising results. While crowd counting CNN architectures generally use points for density map regression, CNN point-based object detectors are often trained to produce a high-resolution map in an encoder-decoder fashion, where points can then be extracted (Ribera et al., 2019; Zhou et al., 2019). An encoder-decoder framework outputs features over the input image's pixel space to obtain precise localization. The encoder block encodes the images into multi-level features' maps of different resolution (i.e., the down-sampling phase), and then the decoder block decodes the encoded features' map while keeping their spatial information (i.e., the up-sampling phase). Other methods also showed that point detection can be achieved on lower-resolution outputs using a simple encoder (i.e., a CNN) but at the expense of a lower position accuracy (Kellenberger et al., 2021, 2019b, 2018).

2.2. Similarities between crowd and herd counting tasks

In crowd counting, there are some challenges that make the task complex, including occlusion, complex background, scale variation, and non-uniform distribution (Gao et al., 2020). These issues are also encountered in herd counting within oblique aerial imagery, which makes the task of herd counting very similar to that of crowd counting (see **Figure 3.1**):

Occlusion (**Figure 3.1a**) - As the herd density increases, the animals will appear to partially occlude each other. This situation is often observed for gregarious and migratory animals which can be grouped around particular places such as watering holes and resource points, and during some practices such as "tightly bunched herding" (Odadi et al., 2018). Such occlusions could limit the performance of traditional object detection architectures.

Complex background (**Figure 3.1b**) - Aerial survey imagery contains mainly background regions that can include many confusing objects (e.g. shadows, rocks). These can lead to a high number of false alarms and bias the counting result.

Scale variation (**Figure 3.1c**) - In oblique aerial images, the size of animals varies both within the same species by the distance from the camera (i.e., intraspecies variation) and between different species (i.e., interspecies variation), increasing the difficulty for accurate detection and identification.

Non-uniform distribution (**Figure 3.1d**) - Diverse herd distributions and densities may be encountered. The difficulty is further accentuated by the fact that the dataset is dominated by samples containing few individuals, following the patch generation (see Section 3.4.1).

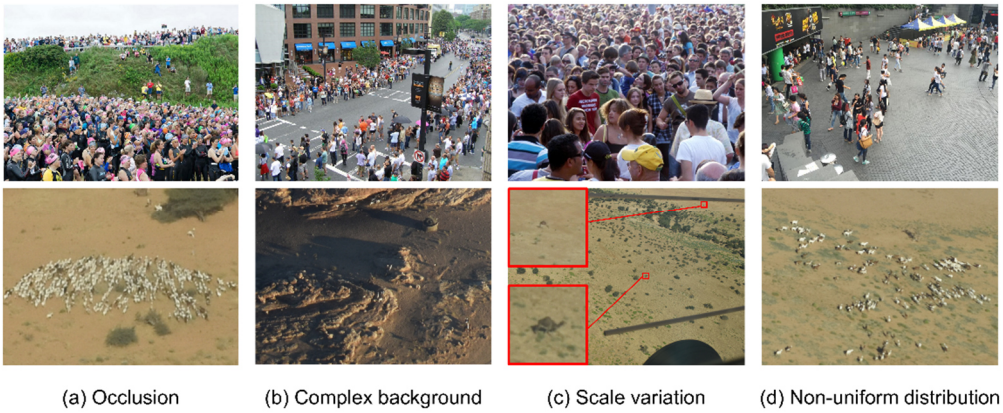


Figure 3.1: Examples of challenges faced by crowd counting (top row), extracted from the Shanghaitech dataset (Zhang et al., 2016) and their equivalents in herd counting (bottom row), extracted from the Ennedi dataset.

These similarities make crowd counting CNN architectures an interesting approach to tackle the challenges of counting dense herds in oblique aerial imagery. However, unlike crowd counting where the problem is binary (human vs. background), herd counting could be a multi-class problem as several species may be targeted in the same area. The original crowd counting CNN architectures must therefore be adapted accordingly.

2.3. Combining detection and counting tasks

Creating an architecture that can accurately locate individuals in a herd could be valuable. It could be used as a tool for obtaining pre-annotations from new data. However, as mentioned above, while traditional CNN-based object detectors can output object locations, they often fail to detect occluded objects. Density-based architectures could then be used, but at the cost of losing precise location information in dense herd regions. The ideal solution would be one that provides both a relatively accurate count of the herd (as usually given by density map approaches) and the position of the individuals in the herd (as given by detectors). Liang et al. (2023) recently proposed such a solution for crowd counting by using a novel Focal Inverse Distance Transform (FIDT) map which replaced the traditional density map. Their experiments demonstrated that this approach outperforms state-of-the-art localization-based methods and showed competitive counting performances while presenting a strong robustness to background and dense scenes samples. Such

robustness is particularly interesting for counting herds in images with a heterogeneous background.

3. Materials and Methods

This section describes the datasets used, the proposed deep learning architecture (called "HerdNet") as well as two standard baselines (anchor-based and density-based), and some details on the data processing utilized in this work. The baselines were used to compare the detection and counting capacity of HerdNet, which was optimized on a dataset that contains challenging herds (Ennedi).

3.1. Study area and dataset

The proposed Deep Learning architecture, HerdNet, was developed on a dataset acquired over the Ennedi Natural and Cultural Reserve (ENCR) located in north-eastern Chad during a 2019 aerial survey (Wacher, 2019). The ENCR covers nearly 50,000 km^2 of arid sandstone landscape surrounded by sandy plains. According to the map of Olson et al. (2001), the ENCR encompasses the following biomes: tropical and subtropical grasslands, savannas, and shrublands; deserts and xeric shrublands. It is managed by the African Parks Network (APN), in partnership with the Government of the Republic of Chad. The ENCR is a vital resource for local semi-nomadic groups who need grazing and water for their camels (*Camelus dromedarius*), goats (*Capra hircus*), sheep (*Ovis aries*), donkeys (*Equus asinus*), and rare cattle (*Bos taurus*). The APN's long-term goal is to get all stakeholders, including local communities that depend on natural resources, to work together to conserve the Sahelo-Saharan heritage of the ENCR, including its archaeological value, while respecting traditions and allowing key species to thrive.

The data were acquired during aerial flights over the ENCR from December 20, 2019 to January 1, 2020. A Cessna 182 equipped with a laser altimeter and external metal strut rod markers calibrated at the observers' eye level to indicate a 200 m band on each side of the aircraft at survey altitudes of 300 and 350 feet captured the images. Two Nikon D5000 SLR cameras, observer-operated by remote release cable and mounted by suction pads on left and right rear windows, were set up to match each observer's view of the strut-mounted sample rods and the ground between them. A total of 19 flights were conducted, covering the core of the reserve (i.e., most of the Ennedi massif and the southwestern plains, representing around 23,000 km^2). Flights were conducted at 350 feet (~107 m) along transects spaced 4 km apart over the massif, and at 300 feet (~91 m) along transects spaced 10 km apart in the south-western plains. Any groups of livestock (camels, donkeys, sheep, and goats) greater than 10 in number were photographed and the images were later used to provide 'corrected' counts. The date and time of image acquisitions were used to match the temporal and spatial data (altitude and GNSS coordinates) acquired by the altimeter during the flights. Thus, the images were associated with their respective transect and flight numbers.

'Corrected' count and observers' group identification were used when establishing the ground truth. The annotations were made on Label Studio 1.3 (Tkachenko et al., 2020) by an expert, and consisted of 22,807 body-centered points in a subset of 914 images at 24 megapixels (6,000 × 4,000 pixels), containing major livestock species, i.e., camels, donkeys, and sheep and goats. Sheep and goats have been grouped in a single class ("sheep/goats") since these two species are not distinguishable and often mixed within herds.

The dataset was split into training, validation and test sets following an allocation of 70, 10 and 20 %, respectively, while considering the species' distribution, the flight and transect number. Dataset independence is thus ensured, whether the same herd is present in several images, and the species distribution is maintained, which is important in a severely unbalanced class distribution like ours. One transect from each flight was selected to construct the test set, resulting in a set of images containing a wide heterogeneity of landscapes from across the reserve. The images and species distribution for each set are given in **Table 3.1**.

Table 3.1: Details of the Ennedi dataset split. The data was split into training (~70 % of all images), validation (~10 %) and test (~20 %) sets while accounting for data heterogeneity (i.e., species distribution, flight and transect) to maintain independence. The numbers in brackets indicate the relative percentage of data in each set. The last row gives the number of patches containing animals extracted from the 24-megapixel images.

Number of	Training	Validation	Test	Total
Camel	2,608 (69.7%)	380 (10.2%)	753 (20.1%)	3,741
Donkey	861 (70.2%)	127 (10.3%)	239 (19.5%)	1,227
Sheep/Goat	12,486 (70.0%)	1,774 (9.9%)	3,579 (20.1%)	17,839
24 MP images	619 (67.7%)	122 (13.4%)	173 (18.9%)	914
512x512 pixel patches	5,826 (75.3%)	1,039 (13.4%)	869 (11.3%)	7,734

MP, megapixel.

3.2. Deep learning architectures

This sub-section provides details about the different deep learning architectures used in this study. These architectures include the following: an anchor-based baseline (Faster-RCNN), a density-based baseline (DLA-34), and the proposed architecture (HerdNet).

3.2.1. Anchor-based Baseline: Faster-RCNN

A naive way to count objects in an image would be to sum the number of detections provided by an object detector. A generic deep learning object detection framework locates and classifies objects within an image through the use of rectangular boxes encompassing the objects, called 'bounding boxes'. Traditional pipelines are anchor-based (Zhao et al., 2019), which means that they rely on anchors, a set of box

proposals with different scales and aspects centered on potential object locations. These were first introduced in Faster-RCNN (Ren et al., 2015) and then used by a number of well-known object detectors like SSD (Liu et al., 2016), YOLOv2 (Redmon and Farhadi, 2017) or RetinaNet (Lin et al., 2017b) because they improved their detection performance.

While anchor-based object detectors have given good detection performances for large mammals detection in aerial images (Eikelboom et al., 2019; Peng et al., 2020; Torney et al., 2019), Delplanque et al. (2022) recently observed a precision drop in herds and close-by animals resulting in overestimated counts.

In crowd counting, the use of anchor-based or even detection-based frameworks is not recommended because of the expensive labeling cost of bounding boxes and the difficulty of training detectors on heavily occluded objects (Li et al., 2021; Liu et al., 2018). Instead, most crowd counting approaches have relied on point annotations since the study of Lempitsky and Zisserman (2010). Nevertheless, anchor-based detectors are widely used in animal detection on aerial images, and thus remain relevant baselines. As it is one of the most-cited object detectors and the most common baseline, Faster-RCNN was chosen as the anchor-based baseline.

Faster-RCNN (Ren et al., 2015) is a two-stage object detector that:

- Generates region proposals using a Region Proposal Network (RPN), which predicts objects' bounds and objectness scores at each position by utilizing anchors; and
- Uses the refinement head of Fast R-CNN (Girshick, 2015) for regions of interest (RoIs) classification and bounding box offset regression.

A RPN is a deep fully convolutional network, and Fast R-CNN is composed of a RoI pooling layer and several fully-connected layers. Both share the same CNN features. For architecture comparison consistency, ResNet-34 (He et al., 2016) has been chosen for feature extraction because it has similar numbers of layers and the same convolutional blocks as the proposed architecture encoder (see Section 3.2.3). This choice will minimize any bias that might be caused by the use of a deeper feature extractor.

3.2.2. Density-based Baseline: Adapted DLA-34

Another way to count objects in an image is to estimate a density map whose integral would give the number of objects within that image. This 'density-based' approach was proposed by Lempitsky and Zisserman (2010) and was a real milestone for crowd counting, thanks to its simple framework for object counting and the introduction of point annotation. Since then, numerous Counting CNN (CCNN) architectures have been deployed and have shown excellent crowd counting performances on benchmark datasets (Gao et al., 2020; Li et al., 2021). Density-based CCNNs use CNN as a feature extractor and are trained to regressively learn a mapping between an image and the density map. Ground truth is produced using a density function, typically a normalized

2D Gaussian, convolved over each annotated point (Lempitsky and Zisserman, 2010). When properly trained, density-based CCNNs estimate the object count by integrating the density map they produce, and they provide spatial information about the objects.

Kellenberger et al. (2019a), Padubidri et al. (2021) have recently shown that density maps can be used for animal counting in nadir aerial images. The former trained an adapted ResNet-18 architecture (He et al., 2016), while the latter trained a U-Net semantic segmentation CNN architecture (Ronneberger et al., 2015) to produce density maps. Unfortunately, precise object location is difficult to obtain from density maps, especially for close-by and occluded objects where 2D Gaussians strongly overlap.

While density-based architectures tend to provide precise object counts in high-density scenes, precise localization is lost. Although the primary goal of aerial surveys is to establish accurate population count, obtaining the precise position of animals in images could be valuable for creating annotations from new data for further model training.

A density-based baseline was therefore established to assess the counting performance of the proposed approach. For comparison consistency, the same feature extractor and decoder as the proposed architecture (i.e., adapted DLA-34, Yu et al., 2018) was selected. In fact, the architecture is that of HerdNet (**Figure 3.2**), except that the classification head has been removed and the main head generates three density maps (one for each species) instead of one localization map. During the test time, for each species, only the pixels with the maximum value among the three predicted density maps were retained. This process prevents the same individuals from being counted as several species. An adaptive threshold of 0.07 was then applied to the density values to eliminate background noise.

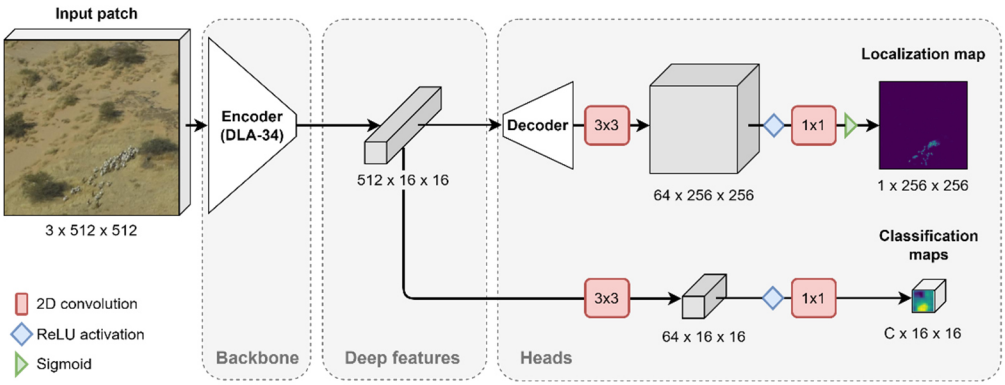


Figure 3.2: HerdNet architecture details.

3.2.3. Proposed Architecture: HerdNet

Since the objective was to develop an architecture to accurately locate and count dense herds, the proposed deep learning architecture, HerdNet, is inspired by both point-based object detectors and crowd counting architectures.

The core of HerdNet is derived from CenterNet, except that only the branch which estimates the objects' center has been retained. This branch corresponds to the localization head. The adapted DLA-34 (Yu et al., 2018) was used as encoder-decoder (**Figure 3.2**) because it gave the best speed vs. accuracy trade-off on the MS COCO (Lin et al., 2014) dataset (Zhou et al., 2019), which is convenient for our application case. As in Zhou et al. (2019), a 3 × 3 convolutional layer was added on top to obtain specialized features maps for each head, and early experiments showed that 64 channels were adequate to obtain good results. A 1 × 1 convolution, preceded by a ReLU activation and followed by a sigmoid activation produces the desired localization map. Early experiments showed that a reduction factor of 2 between input and output sizes gives similar results with fewer network parameters than those obtained by keeping the original patch resolution.

For classification, a second low-resolution head was added onto the deep features layer, with one 3 × 3 convolutional layer with 64 channels on top, as for the location head. Eventually, a 1 × 1 convolution, preceded by a ReLU activation, produces the C classification maps, C being the number of classes including background (**Figure 3.2**). Ablation studies showed that 16 × 16 pixel classification maps were sufficient for species identification, and that including the background class in the training objective helped to better learn the landscape heterogeneity (see **Appendix A1**).

During the testing time, the Local Maxima Detection Strategy (LMDS) proposed by Liang et al. (2023) was used to extract points from the predicted localization map. The LMDS utilizes a 3 × 3 max-pooling operation to obtain candidate points, which are then filtered using an adaptive threshold, set here at 0.3 times the maximum candidate value. An input patch is considered a negative sample when the maximum candidate value is below 0.1, as in the original paper. Next, the classification maps were used to classify the selected points. The softmax function was used on all classes to obtain classification scores. Then, the most confident class was selected among the foreground classes. With this procedure, a selected point could never be classified as background. Finally, the selected points were used to pin the foreground classes to equivalent locations and the class labels and scores were retrieved.

3.3. Data processing

This sub-section describes all the processes implemented for developing the models. Operations were performed on a Windows-10 workstation using Python 3.8.10. The workstation contained a 64 GB AMD Ryzen 9 5900X central processing unit (CPU) and an 8 GB NVIDIA GeForce RTX 3070 graphics processing unit (GPU). All

architectures were implemented in PyTorch 1.11 (Paszke et al., 2019) and experiments were tracked with Weights & Biases 0.10.33 (Biewald, 2020).

3.3.1. Patch generation and stitching

Original 24-megapixel images were cut into patches of 512 × 512 pixels to maintain initial resolution and because experimenting with original-size images exceeds the memory capacity of current GPUs. To ensure that every animal appears in its entirety during training, a patch overlap was used. After manually measuring the largest individuals in the dataset (i.e., camels close to the lower stream bar), it was concluded that a 160-pixel overlap was a good value, as the widest of these was 156 pixels long.

During the testing, the original-size images were scanned with a sliding window to harvest predictions and then stitch them together. To do so, each patch of 512 × 512 pixels was evaluated independently and overlapped region predictions were filtered out. Specifically, the common Non-Maximum Suppression (NMS) method was adopted with an Intersection-over-Union (IoU) threshold (Everingham et al., 2015) of 0.5 (as Delplanque et al., 2022; Peng et al., 2020) and a score threshold of 0.4 for Faster-RCNN predictions (**Appendix A2**). For the adapted DLA-34, predicted density maps were filtered and stitched using Hann windows to reduce the edge-effect, as proposed by Pielawski and Wählby (2020). Next, an adaptive threshold of 0.07 was applied on the stitched image pixel values to eliminate background noise (**Appendix A3**). Finally, overlapped predicted grid values were averaged before using LMDS for HerdNet.

3.3.2. Model training

Hard negative patch mining

Hard negative mining is a training technique used to treat the hard negative samples severely during training (Kellenberger et al., 2018; Liu et al., 2016; Shrivastava et al., 2016). In the animal detection domain, hard negative samples correspond to background elements detected as animals with a high confidence score. A Hard Negative Patch (HNP) mining method was adopted here, following the hard negative mining concept, to further reduce the number of false positives produced by the model. After a first training session where the architecture was trained exclusively on animal patches, the model was run on the 24-megapixel training images. HNPs were then mined from the stitched predictions, which are the patches that contain hard negative instances. These HNPs were eventually used to retrain the model a second time to force it to develop more robust features regarding the most confusing background elements. With this method, only the most complex background patches were selected, which makes the task more efficient and less tedious than training on all the patches, as proposed by Kellenberger et al. (2018).

Faster-RCNN

For the anchor-based baseline (i.e. Faster-RCNN), bounding boxes were generated from annotated points. For this purpose, a subset of the Ennedi dataset was annotated

as bounding boxes. Then, for each species, median height and width were computed per a 200-pixel horizontal strip in the image, and the maximums of each were selected to create square bounding boxes centered on each annotated point.

Training data was augmented artificially using Albumentations' (Buslaev et al., 2020) random horizontal flip and motion blur data augmentations.

During the first training step, the parameters of the features extractor were initialized using ImageNet (Russakovsky et al., 2015) pretrained parameters. The architecture was then trained and validated on animal-only patches for 100 epochs with a batch size of 4 and a weight decay of 0.005 using the Adam optimizer (Kingma and Ba, 2017). Concerning the learning rate, PyTorch's 'ReduceLROnPlateau' learning rate scheduler was used because it made it possible to automatically decrease the learning rate at the most appropriate time during training. The initial learning rate was set to 10^{-5} after a linear warmup of 100 iterations and could then decrease by a factor of 0.1 until 10^{-6} when no improvement was observed on the validation set over a period of 10 epochs. After a reduction, a delay of 10 epochs was imposed to let the architecture adapt to the new learning rate.

At the end of this first training step, the network's parameters that yielded the best performances on the validation set were selected for initializing the second training step. During the latter, we added the HNPs to the training set and validated on 24-megapixel validation images using the same hyperparameters as the first step, except for the number of epochs and the initial learning rate, which were set at 50 and 10^{-6} respectively. 24-megapixel images were used for validation to focus on both localization and counting within real case scenes during this second training step.

Due to the substantial imbalance in species instances, class weighting was used in the bounding boxes' classification loss. Satisfactory results were found by setting the class weights' values to 0.1 for background class, and to the unit rounded value of the ratio of the majority class instances to that of the actual class. All other hyperparameters were left at their default values specified in PyTorch. The parameters of the network that yielded the best performances on the full images of the validation set were then selected for testing.

Adapted DLA-34

The density-based baseline (i.e., adapted DLA-34) ground truth density maps were generated using a 2D Gaussian function, convolved over each annotated point, for each species class, as in Lempitsky and Zisserman (2010):

$$M_{density,c}(i,j) = \sum_{p \in P} \mathcal{N}(i,j\,;p,\sigma^2)$$

where $M_{density,c}\,(i,j)$ is the density map of a class c , p denotes an equivalent low-resolution annotated point $(x'/2, y'/2)$ within the low-resolution image 2D points set P, and $\mathcal{N}(i,j\,;p,\sigma^2)$ represents a normalized 2D Gaussian kernel evaluated at pixel

(i, j), with the mean centered on p, and an isotropic covariance matrix with spread parameter σ, set at 5 pixels. With this definition, integrating each density map produced gives the total count of each species class N_c:

$$N_c = \sum_{(i,j)} M_{density,c}(i, j)$$

The architecture was then trained using the Structural Similarity Index (SSIM) (Wang et al., 2004) loss between the predicted density maps and the ground truth density maps**:**

$$\mathcal{L}_{density}(\hat{Y}, Y) = \frac{1}{C}\sum_{c} w_c(1 - SSIM_c)$$

with:

$$SSIM_c(\hat{y}_c, y_c) = \frac{(2\mu_{\hat{y}_c}\mu_{y_c} + \lambda_1)(2\sigma_{\hat{y}_c y_c} + \lambda_2)}{(\mu_{\hat{y}_c}^2 + \mu_{y_c}^2 + \lambda_1)(\sigma_{\hat{y}_c}^2 + \sigma_{y_c}^2 + \lambda_2)}$$

where Y and $\hat{Y}$ are the ground truth and the predicted density maps, respectively, with y_c and $\hat{y}_c$ their respective class-specific values, C is the number of species classes, w_c is the class weight, μ and σ are the local mean and variance values, respectively, and λ_1 and λ_2 are set to 10^{-4} and $9\text{x}10^{-4}$, respectively.

As for Faster-RCNN, the architecture was trained and validated using the same data augmentations, parameter initialization, hyperparameters, and optimizer. A fixed learning rate of 10^{-5} was used here as a learning rate scheduler gave poorer performances. The HNP mining procedure was discarded here because using it showed an increase in the counting errors.

Class weighting was also applied on the SSIM loss using the same class weights. All other hyperparameters were left at their default values, specified in PyTorch.

HerdNet

Low-resolution FIDT maps (Liang et al., 2023) were adopted as ground truth for training the HerdNet's localization branch:

$$M_{loc}(i, j) = \frac{1}{D(i, j)^{(\alpha \times D(i,j) + \beta)} + k}$$

where $M_{loc}(i, j)$ is the FIDT map, $D(i, j)$ represents the euclidean distance between the pixel (i, j) and its nearest equivalent low-resolution animal location $(\text{x}'/2, \text{y}'/2)$, α and β are FIDT hyper-parameters, set as 0.02 and 0.75 respectively, following Liang et al. (2023), and k is a constant, set to 1 to avoid division by zero. FIDT maps produce local maxima of 1 at each animal's center, with a slow response decay and a background response close to 0.

This branch was trained using the unnormalized penalty-reduced pixel-wise logistic regression with focal loss (Lin et al., 2017b), as proposed by Zhou et al. (2019):

$$\mathcal{L}_{loc}(\hat{Y}_l, Y_l) = -\sum_i \sum_j \begin{cases} (1-\hat{y}_{l,ij})^{\alpha} \log(\hat{y}_{l,ij}), & \text{if } y_{l,ij} = 1 \\ (1-y_{l,ij})^{\beta} (\hat{y}_{l,ij})^{\alpha} \log(1-\hat{y}_{l,ij}), & \text{otherwise} \end{cases}$$

where Y_l and $\hat{Y}_l$ are the ground truth and the predicted localization grids, respectively, and $y_{l,ij}$ and $\hat{y}_{l,ij}$ their values at a specific pixel location (i, j), and α and β are focal loss hyper-parameters, set at 2 and 4, respectively, as indicated in Zhou et al. (2019).

For the classification branch, low-resolution classification maps were produced from equivalent low-resolution animal locations $(x'/32, y'/32)$. Practically, at each equivalent animal location, a 1-pixel border was added and the whole region was defined as the species identifier. This was to ensure a sufficient point coverage area given the low resolution of the classification branch output (16 x 16 pixel). The common cross-entropy loss was used for training this branch:

$$\mathcal{L}_{class}(\hat{Y}_c, Y_c) = -\sum_i \sum_j \sum_c w_c y_{ijc} \log(\hat{y}_{ijc})$$

where Y_c and $\hat{Y}_c$ are the one-hot encoded ground truth and the predicted classification grids, respectively, y_{ijc} and $\hat{y}_{ijc}$ their values at a specific pixel location (i, j) for a particular class c, and w_c is the class weight.

The overall training objective is then:

$$\mathcal{L} = \mathcal{L}_{loc}(\hat{Y}_l, Y_l) + \mathcal{L}_{class}(\hat{Y}_c, Y_c)$$

During the first and second training steps, HerdNet followed the same training procedure and hyperparameters as Faster-RCNN, except that the initial learning was set to 10^{-4}. Again, the best network's parameters were kept based on the performances obtained on the full images of the validation set.

3.3.3. Model evaluation

The trained architectures (or models) were evaluated using both localization and counting metrics. A prediction was defined as a true positive (*TP*) if there was a match with a ground truth and if the animal identification was correct. In the case where several predictions met the two rules, the best one was selected and the others were considered as false positives (*FP*). Finally, if no matches were found or if the identification was incorrect, the ground truth was considered as false negative (*FN*). To define a match, we used the IoU for Faster-RCNN and set a minimum threshold of 0.3, where the best prediction is the one with the highest IoU. In the HerdNet approach, we used the Euclidean distance between points, with a maximum threshold of 5 pixels, where the best prediction is the one with the minimum Euclidean distance.

Recall, precision and F1 score were then computed for each class (i.e. for each species), as well as for the binary case (animal vs. background):

$$\text{recall} = \frac{\sum TP}{\sum TP + \sum FN}$$

$$\text{precision} = \frac{\sum TP}{\sum TP + \sum FP}$$

$$\text{F1 score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$$

As it represents the harmonic mean of recall and precision, the F1 score is a good metric with which to assess the compromise between the number of *FPs* and *FNs*. Therefore, the binary F1 score was used as the performance metric during validation. In addition to these metrics, we also compute, for each species, the foreground interclass confusion, which is equal to 0 when all the predictions are correctly classified:

$$\text{confusion}(c) = 1 - \frac{n_c}{\sum_{i=1}^{C} n_c}$$

where n_c is the number of predictions identified as class c, and C is the number of foreground classes (i.e., the number of species).

Note that localization metrics could not be applied to the adapted DLA-34 due to the loss of localization information caused by the overlap of the 2D Gaussians in dense herd areas.

The Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are used as counting metrics, again computed for each class and for the binary case:

$$\text{MAE} = \frac{1}{I}\sum_{i=1}^{I} |\hat{n}_i - n_i|$$

$$\text{RMSE} = \sqrt{\frac{1}{I}\sum_{i=1}^{I} (\hat{n}_i - n_i)^2}$$

where I is the number of images, and $\hat{n}_i$ and n_i are the predicted and ground truth count of the i-th image, respectively.

Finally, an individual proximity metric was derived by calculating the Minimum Spanning Tree (MST) (Gower and Ross, 1969) on N annotated points in 512 x 512 pixel patches of the test set (**Figure 3.3**). The MST computes a set of $N - 1$ straight line segments joining pairs of points with no loops, forming a tree of minimum length.

MiSTree Python's package version 1.2.0 was used (Naidoo, 2019). In this package, the MST was initially constructed using a k-nearest neighbor graph, here set at $N - 1$, which was then fed to Kruskal's algorithm (Kruskal, 1956). To obtain a metric representative of the proximity of individuals in the patch, we used the median of segment length values instead of the sum. This value was then divided by the threshold value defined above (i.e. 5 pixels) to normalize the metric. Thus, a value close to 1 means a very dense herd where individuals are tightly grouped. Based on this, three proximity classes were defined:

1) High density: patches where the proximity metric varied between 0 and 3;
2) Medium density: patches where the proximity metric varied between 4 and 20; and
3) Low density: patches where the proximity metric was above 20.

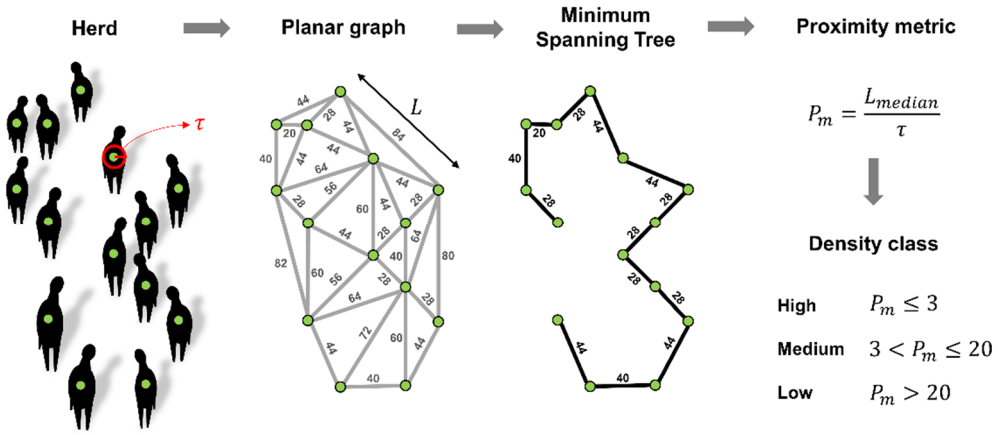


Figure 3.3: Conceptual representation of the Minimum Spanning Tree and proximity metric calculation on a schematic herd. τ represents a circular distance threshold (defined here at 5 pixels) and L represents the Euclidean distance between two individuals.

4. Results

4.1. Hard negative patch mining

The addition of HNPs to the training set increased the precision of Faster-RCNN and HerdNet by more than 18 % and 30 % respectively, despite a decrease in recall of about 8 % and 7 % respectively (**Table 3.2**). This resulted in a better counting performance, with a lower average confusion between species. Consequently, for each of these models, the version trained with HNPs was retained for further analysis on the test set. In contrast, the counting performance of the adapted DLA-34 decreased with the use of this technique (**Table 3.2**). Therefore, the version of the adapted DLA-34 using HNP mining was discarded and only the version without this technique was used for the analyses on the test set. This is to compare the best version of each model.

Table 3.2: Binary (animal vs. background) performances of the three approaches on 24-megapixel images of the validation set, using Hard Negative Patch mining procedure or not. Values in bold indicate the best performance between the two modalities.

Approach	**Anchor-based**		**Density-based**		**Point-based**	
Architecture	**Faster-RCNN**		**DLA-34**		**HerdNet**	
HNP[1]	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
Recall	**64.1%**	56.0%	n/a	n/a	**72.1%**	64.4%
Precision	20.4%	**38.5%**	n/a	n/a	43.5%	**75.4%**
F1 score	30.9%	**45.7%**	n/a	n/a	54.3%	**69.4%**
MAE[2]	40.1	**11.3**	**12.3**	12.3	14.3	**6.1**
RMSE[3]	51.7	**16.5**	**19.1**	23.0	19.4	**10.5**
Average confusion	15.0%	**13.7%**	n/a	n/a	22.4%	**17.8%**
Total counting error	214.4%	**45.4%**	**-0.2%**	-40.8%	65.5%	**-14.6%**

[1]HNP, Hard Negative Patch; [2]MAE, Mean Average Error; [3]RMSE, Root Mean Square Error.

4.2. Model comparison

Overall, HerdNet outperformed the detection and counting performance of the two baselines, Faster-RCNN and the adapted DLA-34, in addition to having a faster processing time (**Table 3.3**). However, Faster-RCNN had a lower average confusion level, and the adapted DLA-34 had a lower absolute total counting error.

Table 3.3: Binary (animal vs. background) performances of the three approaches on 24-megapixel images of the test set. Values in bold indicate the best performance among the architectures.

Approach	Anchor-based	Density-based	Point-based
Architecture	Faster-RCNN	DLA-34	HerdNet
Recall	59.5%	n/a	**70.2%**
Precision	39.4%	n/a	**77.5%**
F1 score	47.4%	n/a	**73.6%**
MAE[1]	15.2	15.9	**6.1**
RMSE[2]	26.2	30.4	**9.8**
Average confusion	**11.1%**	n/a	15.8%
Total counting error	51.2%	**7.6%**	-9.4%
Processing time (seconds)	5.0	5.5	**3.6**

[1]MAE, Mean Average Error; [2]RMSE, Root Mean Square Error.

HerdNet showed a counting performance that was close to true counts while Faster-RCNN tends to overestimate the true number of animals (**Figure 3.4**). As for the adapted DLA-34, it tends to underestimate large groups, and overestimate very small groups.

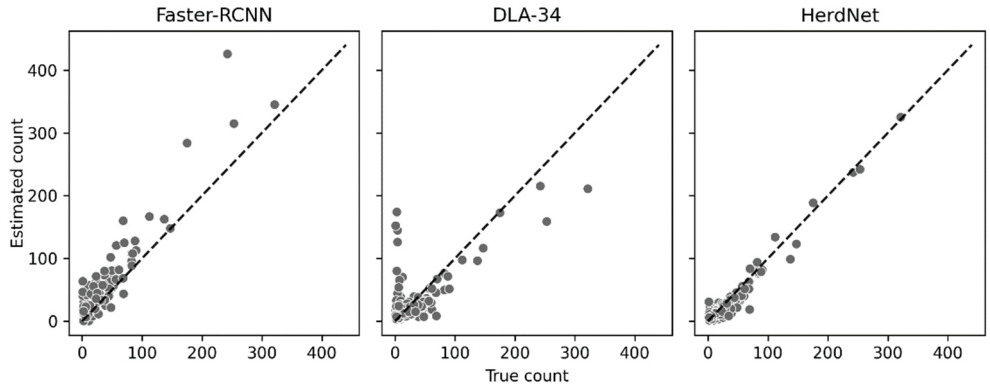


Figure 3.4: Estimated counts produced by each architecture versus the true counts in 24-megapixel images of the test set.

Regarding species identification, HerdNet outperformed Faster-RCNN for all three target species (**Table 3.4**). For camels and donkeys, however, Faster-RCNN showed less confusion between species. The adapted DLA-34 showed lower MAE and RMSE values for donkeys than HerdNet, but higher values for camels and sheep/goats.

Table 3.4: Performances of the three approaches on 24-megapixel images of the test set according to target species. Values in bold indicate the best performance among the architectures.

Approach	**Anchor-based**			**Density-based**			**Point-based**		
Architecture	**Faster-RCNN**			**DLA-34**			**HerdNet**		
Species	**Camel**	**Donkey**	**Sheep /Goat**	**Camel**	**Donkey**	**Sheep /Goat**	**Camel**	**Donkey**	**Sheep /Goat**
n	753	239	3,579	753	239	3,579	753	239	3,579
Recall	57.5%	18.8%	60.6%	n/a	n/a	n/a	**61.8%**	**37.7%**	**70.9%**
Precision	45.8%	9.3%	39.6%	n/a	n/a	n/a	**75.1%**	**59.6%**	**75.3%**
F1 score	51.0%	12.4%	47.9%	n/a	n/a	n/a	**67.8%**	**46.2%**	**73.0%**
MAE[1]	3.3	3.0	13.9	3.6	**1.6**	15.2	**2.6**	2.5	**7.0**
RMSE[2]	5.7	4.3	25.6	6.5	**3.3**	27.8	**4.8**	4.6	**10.7**
Confusion	**0.0%**	**30.8%**	2.4%	n/a	n/a	n/a	7.4%	39.2%	**0.8%**

[1]MAE, Mean Average Error; [2]RMSE, Root Mean Square Error.

Taking into consideration both overall and per-species results, HerdNet is the architecture with the best detection and counting performances, especially for sheep/goats, which represent especially challenging herds (**Figure 3.5**).

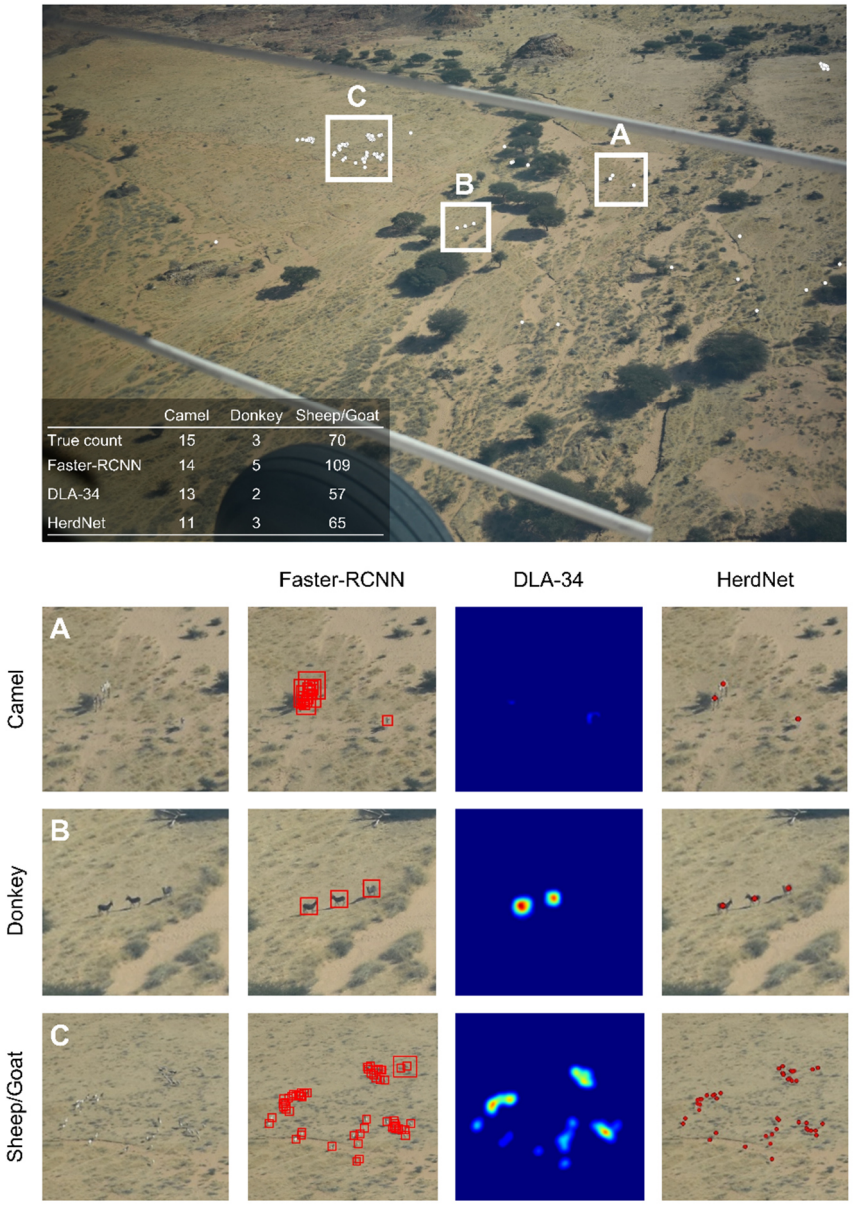


Figure 3.5: Predictions of the three trained architectures on a 24-megapixel image containing the three target species (camel, donkey, and sheep/goat). White points correspond to the annotations, red bounding boxes to Faster-RCNN predictions, density maps to the predictions of adapted DLA-34, and red points indicate the HerdNet predictions.

4.3. Robustness of HerdNet towards animals proximity

Recall and precision were computed for each class of animal proximity defined in section 3.3.3 to assess the robustness of HerdNet towards animal proximity in 512 × 512 pixel patches of test set images. The results indicate that the mean precision of HerdNet was systematically higher than that of Faster-RCNN for each proximity class, while keeping equivalent mean recall values (**Figure 3.6**). This reveals the ability of HerdNet to generate few false positives in both dense (**Figure 3.7**) and sparse herd patterns.

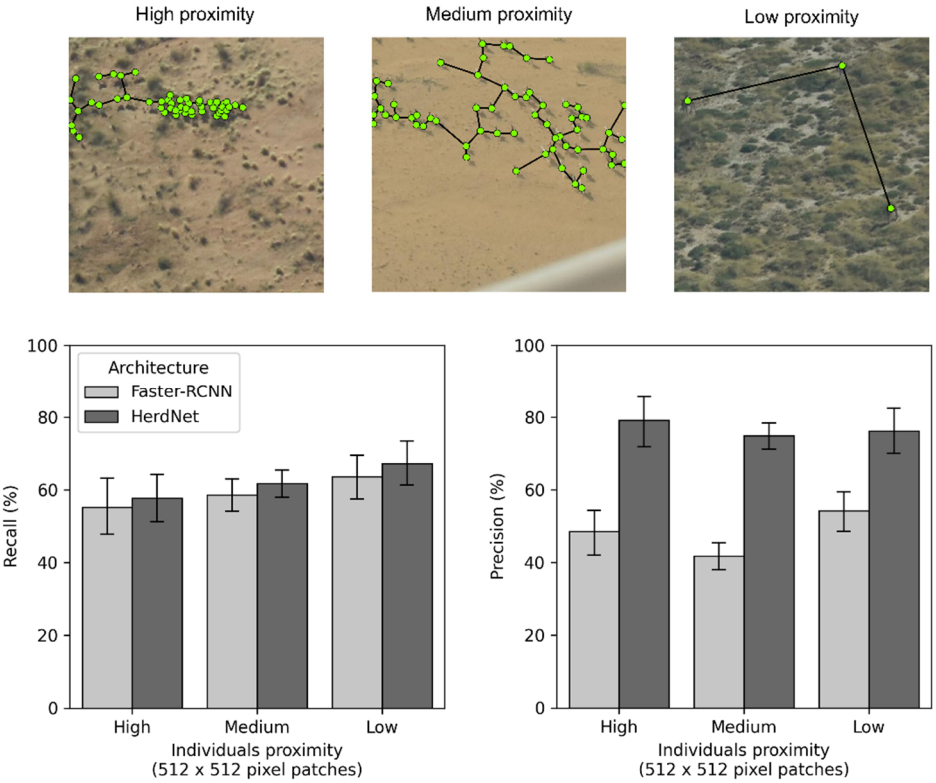


Figure 3.6: Recall and precision mean values of Faster-RCNN and HerdNet, computed on 512 × 512 pixel patches for each class of animal proximity metric based on a minimum spanning tree. The error bars correspond to the 95 % confidence interval.

Figure 3.7: Examples of HerdNet predictions for challenging dense sheep/goat herds. The first row contains sample patches selected from 24-megapixel images, while the second row shows the respective predicted points in red.

5. Discussion

5.1. Best approach for counting dense herds

Three different approaches were compared to precisely detect and/or count animals in herds within oblique aerial imagery: 1) a CNN-anchor-based object detector (Faster-RCNN); 2) a CNN-density-based detector (an adapted version of DLA-34); and 3) a CNN-point-based object detector (called HerdNet). The first two approaches served as baselines because they have already proved their worth in the field of wildlife detection/counting within aerial imagery.

As previously observed (Delplanque et al., 2022; Peng et al., 2020), the anchor-based architecture showed its limitations in precisely detecting close-by individuals. It produced here a high number of false positives in dense herds, even using score thresholding, resulting in systematic over-counting. This raises questions about the use of such models in entire aerial surveys where it is expected to get images with both remote individuals and dense herds.

The CNN-density-based detector (DLA-34) provided better total counting performance but struggled to correctly count large and dense herds. The counting errors are higher than those obtained by Kellenberger et al. (2019a) and Padubidri et al. (2021) on their nadir datasets. In fact, the DLA-34′s counting errors are low for minority species (i.e., camels and donkeys), but much higher for sheep/goats, which are far more gregarious. This could be explained by the change in scale within the image due to the oblique viewing angle, the higher variance in the number of individuals and the greater heterogeneity of the background. These factors can also

limit the performance of crowd counting using density maps (Gao et al., 2020). A solution would be to design a multi-scale architecture such as MCNN, a multi-column architecture that uses different kernel sizes to capture images at different scales (Zhang et al., 2016).

Our CNN-point-based object detector (HerdNet) gave the best detection and counting performances while also being the fastest approach, suggesting that it seems best suited for locating and counting animals in dense herds. Estimating the number of livestock in protected areas is sometimes a politically sensitive issue, as a livestock invasion is detrimental to the biomass of wildlife (Scholte et al., 2022b). Moreover, livestock invasions directly show that the responsible authorities or supporting international non-governmental organizations have failed in their conservation mission. A method that overestimates this figure is therefore undesirable. Hence, HerdNet is the most appropriate and preferred approach for herd counting.

5.2. Species identification limits

In terms of species identification, HerdNet was slightly better than Faster-RCNN for sheep/goats but was about 7–8 % worse for minority species, i.e., camels and donkeys. Thus, the class imbalance seems to impact HerdNet more than Faster-RCNN. After manually analyzing the images with the most significant cases of confusion, overall trends were deduced. First, the size and often the low resolution of the individuals were source of confusion for the model, especially for donkeys (**Figure 3.8**). The latter were usually well identified in the higher resolution areas (i.e., near the lower stream bar), but that the identification degraded with the distance to the aircraft. Regarding camels, the lighter ones located in the low-resolution regions of the image (i.e., near the upper stream bar) were often confused with sheep/goats. Furthermore, identification was sometimes incorrect when the animal was positioned to the side (**Figure 3.8**).

This may be explained in part by the fact that the image resolution and high flight height in the massif sometimes did not allow the species to be accurately distinguished during annotation, especially those far from the aircraft. In such cases, identification was solely based on the observers' survey records. However, as the livestock in this study are typically found in single-species groups, identification of individuals is a bonus and not strictly necessary. Indeed, having a model capable of precisely locating and counting individuals is already a very real help in processing images containing large and dense herds. Identification could be further enhanced by a quick review by the human eye of the surrounding individuals.

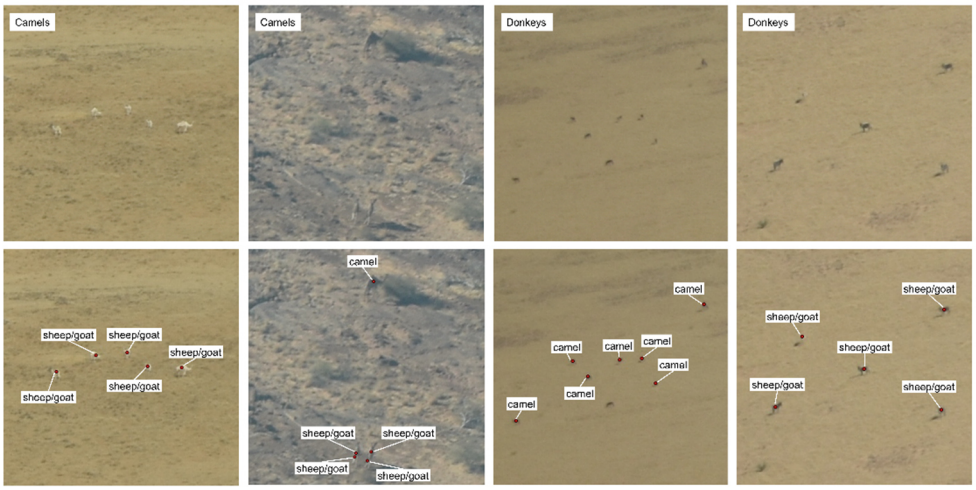


Figure 3.8: Examples of HerdNet predictions for challenging dense sheep/goat herds. The first row contains sample patches selected from 24-megapixel images, while the second row shows the respective predicted points in red.

5.3. Potential use of HerdNet

The use of HerdNet on other datasets requires prior training on similar data, i.e., with the same viewing angle, the same mammal species, and similar spectral and spatial resolutions. To assess the potential use of HerdNet architecture, it was trained and evaluated on the wildlife nadir aerial images of Delplanque et al. (2022). Results showed that HerdNet produced far fewer false positives than the state-of-the-art model, which was Libra-RCNN (Pang et al., 2019), while maintaining a high recall value, hence showing better counting performances (see **Appendix A4** for details). This suggests that this architecture is not limited to its use on oblique imagery only but has good potential for various types of aerial image, acquired under different acquisition conditions.

However, HerdNet may need to be modified in the case of dense mixed herds. Despite the results of the sensitivity study (see **Appendix A1**), the low resolution of the classification head could indeed be problematic if different species are within 32 pixels of each other in the input patch. This distance corresponds to one pixel in the 16 × 16-pixel classification maps. This case did not occur in the dataset of this study, as the dense herds were systematically homogeneous in species. Nevertheless, we believe that this should not be an issue for training and using HerdNet on oblique imagery taken with good quality reflex cameras (e.g., 24 MP of resolution), at common camera tilt values (30-45° off nadir) and survey flight heights (300–350 ft). In such case, the ground sampling distance usually does not exceed 3–4 cm/pixel and 6–8 cm/pixel near the lower and upper stream bars, respectively. This means that the low resolution of the classification head could become a concern when two different species would be less than 2–3 m apart in reality, which is rather rare for large wild

terrestrial mammals. However, if such a case arises, the architecture can simply be adapted by adding a decoder at the beginning of the classification head, whose depth will depend on the desired output resolution.

Finally, we observed that applying the model on a too-different dataset without re-training led to poor performance, probably due to a data domain gap. This could be solved by using *transductive transfer learning* techniques, which allow the transfer of knowledge learned from a source domain to a different target domain, considering the same learning task (Pan and Yang, 2010). Kellenberger et al. (2019b) have already proposed such a solution for wildlife detection using a Transfer Sampling criterion, allowing their model to be reused for repeated nadir drone image acquisitions. However, the transfer learning from oblique to nadir animal detection does not seem to have been studied and should be explored in future research. At this stage, we can only suggest that future users of HerdNet re-train the architecture on a data domain close to their own to obtain more satisfactory results.

5.4. Model precision practical implications

Precise counts of large mammals within sampling strips are important to obtain minimum-biased population estimates. Since undercounting is one of the major biases of aerial surveys (Caughley, 1974; Grimsdell and Westley, 1981; Jachmann, 2002), the main expectation of automatic approaches is to obtain a model with a high detection rate (i.e. high recall) and few false positives (i.e. high precision). However, recall and precision are often antagonistic: improving the precision of a model usually reduces its recall and vice-versa. When developing tools to assist protected area managers, the recall/precision trade-off depends on the goal. As a semi-automatic model for background image rejection, recall should be preferred, as the detections will be reviewed by humans afterwards. However, protected area managers do not always have dedicated office staff for such specific tasks. For a fully automatic system, the optimal trade-off should be preferred, and a prior estimation of the possible bias is necessary to correct the counts. In this study, we optimized the model on the F1 score to automate the counting of herds and minimize the error in individual images. In view of the results obtained, the model proved to be a good tool for the automatic counting of individuals in individual oblique images from arid environments containing livestock herds.

5.5. Future work

Three aspects for future research can be identified. First, the species identification capacity of HerdNet, which could be augmented by confronting it with data sets composed of a large number of species and with some that would be very similar but identifiable by humans from the aircraft (e.g. antelopes). This process would assess the limits of HerdNet regarding the human species' identification ability. Future challenges would involve the adaptation of the model for wildlife species living in herds (elephants, buffaloes, wildebeest, giraffes, etc.), and among those of small size living in small groups (kobs, warthogs, etc.). The use of this approach on complete

chains of transect images could then be investigated by automating the management of overlapped images with the aim of obtaining population estimates. Encouraging results will bring us closer to the full automation of aerial surveys. Finally, the generalizability of HerdNet should be further developed by studying its response to background, viewing angle, and species variability, and possible generalization solutions (e.g. using domain adaptation techniques). A general model or an adaptation approach, that should be simple and require limited technical and human resources, would allow its practical use by protected area managers. That sort of approach would enable them to easily adapt the model for use in the savanna, for example, during both the dry and rainy seasons.

6. Conclusion

In large protected areas in Africa, large mammals are usually surveyed by human observers using aircraft. Unfortunately, the difficulty of observers to precisely count large groups has led to the use of aerial imagery. In such images, the manual counting of individuals is time consuming and the latest Deep Learning approaches have shown their limitations in detecting dense herds. Inspired by crowd counting, the point-based Deep Learning architecture proposed in this study, HerdNet, addresses this problem by precisely detecting and counting animals regardless of individual proximity. Outperforming both anchor-based and density-based baselines, the proposed model has proven to be the fastest and the most suitable approach for detecting and counting closed-by large mammals. It could therefore be used as an automatic livestock counting tool on oblique aerial images acquired in arid areas, and it could be extended to other areas and wildlife species after prior retraining.

7. Appendices

A1: HerdNet Classification Head Ablation Studies

Resolution of the Classification Maps

Increasing the resolution of HerdNet's classification head maps resulted in an increase in the number of network parameters (weights and bias). To determine the optimal resolution to obtain reliable identification results of the target species (i.e. camel, donkey, and sheep/goat), three modalities were tested:

1) 16x16 pixel, which is the resolution of the deepest features for an input image of 512x512 pixel;
2) 32x32 pixel; and
3) 64x64 pixel.

A decoder of the same type as for the localization head was added at the beginning of the classification head to increase the resolution of the output classification maps. Each model was trained, validated, and evaluated according to the methodology presented in the paper.

The results showed that the 16x16 pixel resolution gave the best overall identification performance (lowest average confusion) for fewer model parameters and for similar counting and detection performances (**Table 3.5**).

Table 3.5: Binary (animal vs. background) performances of HerdNet on full images of the Ennedi validation at different classification map resolution. Values in bold indicate the best performance between the two experiments.

Size	#parameters	Recall	Precision	F1 score	Average confusion	MAE[1]	RMSE[2]
16x16 pixel	18670685	59.5%	**76.1%**	**66.8%**	**16.4%**	**6.5**	11.7
32x32 pixel	19839069	**61.2%**	73.0%	66.6%	20.2%	6.6	**11.5**
64x64 pixel	20425821	59.7%	73.1%	65.7%	23.4%	6.7	12.0

[1]MAE, Mean Absolute Error; [2]RMSE, Root Mean Square Error.

Moreover, better counting, detection, and identification performance were obtained for minority species (i.e. camels and donkeys) despite slightly worse performance for sheep/goats (**Table 3.6**).

Table 3.6: Identification performances of HerdNet on full images of the Ennedi validation at different classification map resolution. Values in bold indicate the best performance among the three resolutions.

Species	Resolution	Recall	Precision	F1 score	Confusion	MAE[1]	RMSE[2]
Camel	16x16 pixel	**68.7%**	**68.5%**	**68.6%**	**5.1%**	**1.8**	**3.3**
	32x32 pixel	64.2%	65.9%	65.1%	6.2%	2.1	4.2
	64x64 pixel	61.3%	68.1%	64.5%	9.3%	2.1	4.4
Donkey	16x16 pixel	**29.1%**	**41.1%**	**34.1%**	**40.3%**	**2.3**	**3.2**
	32x32 pixel	25.2%	41.0%	31.2%	52.2%	2.4	3.5
	64x64 pixel	22.0%	34.6%	26.9%	57.6%	2.5	3.4
Sheep/ Goat	16x16 pixel	55.4%	**74.7%**	63.6%	3.8%	9.5	14.9
	32x32 pixel	**59.0%**	71.4%	**64.6%**	**2.2%**	8.8	13.8
	64x64 pixel	56.5%	69.7%	62.4%	3.4%	**8.3**	**13.7**

[1]MAE, Mean Absolute Error; [2]RMSE, Root Mean Square Error.

Including the Background Class

To evaluate whether the addition of the background class to the training objective had an impact on HerdNet performance, two modalities were tested on the full images of the validation set:

1) excluding the background class, i.e., ignoring the background cells in the loss calculation; and
2) including the background class in the loss calculation.

The results showed that including the background class in the training objective led to better counting, detection, and identification performance (**Table 3.7**).

Table 3.7: Binary (animal vs. background) performances of HerdNet on full images of the Ennedi validation set. Values in bold indicate the best performance between the two experiments.

Background included	Recall	Precision	F1 score	Average confusion	MAE[1]	RMSE[2]
No	69.9%	35.0%	46.7%	25.1%	19.7	26.7
Yes	72.1%	43.6%	54.3%	22.4%	14.3	19.4

[1]MAE, Mean Absolute Error; [2]RMSE, Root Mean Square Error.

A2: Faster-RCNN Hyperparameters Optimization

Optimal Non-Maximum Suppression (NMS) Threshold

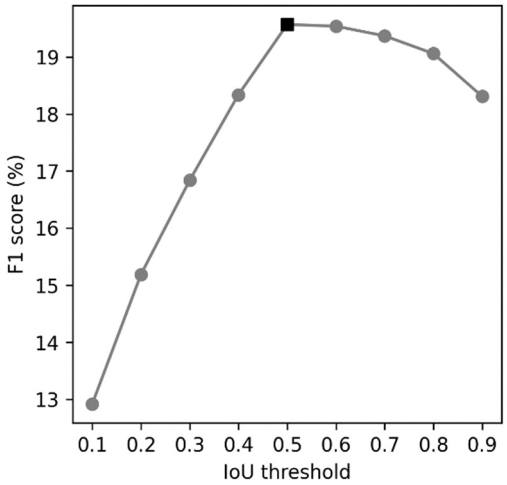


Figure 3.9: Evolution of the F1 score according to different NMS' Intersect-over-Union (IoU) thresholds, calculated on the full images of the validation dataset prior to score thresholding. The best F1 score obtained (19.6%) among the thresholds is indicated as a black square, the corresponding IoU threshold being 0.5.

Optimal Confidence Score Threshold

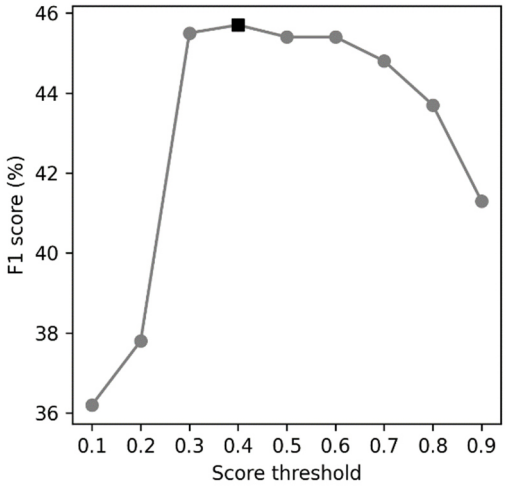


Figure 3.10: Evolution of the F1 score according to different confidence score thresholds, calculated on the full images of the validation dataset. The best F1 score obtained (45.7%) among the thresholds is indicated as a black square, the corresponding confidence score threshold being 0.4.

A3: Adapted DLA-34 Hyperparameters Optimization

Hann Windows for Edge-effect Reduction

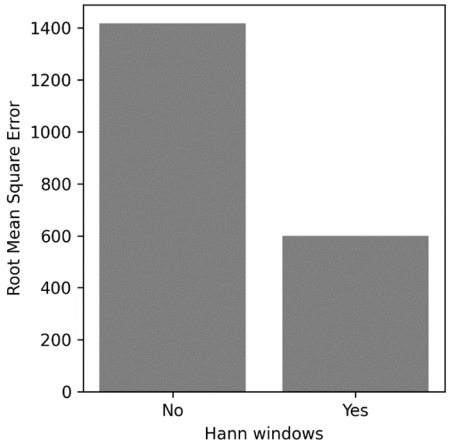


Figure 3.11: Bar plot representing the Root Mean Square Error (RMSE) value obtained on the full image of the validation set before background noise reduction, depending on the use of Hann windows or not. The use of the latter decreased the RMSE by more than half.

Optimal Adaptive Threshold for Background Noise Elimination

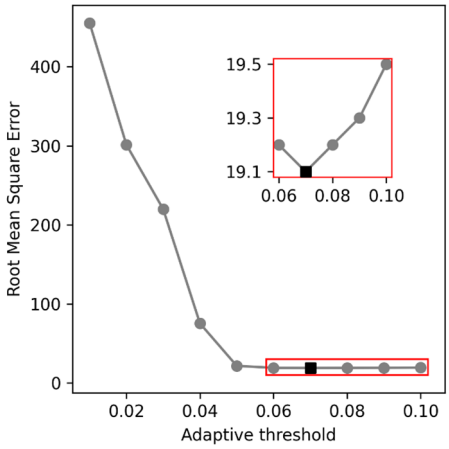


Figure 3.12: Evolution of the Root Mean Square Error (RMSE) according to different adaptive thresholds, calculated on the full images of the validation dataset. The lowest RMSE obtained (19.1) among the thresholds is indicated as a black square, the corresponding adaptive threshold being 0.07.

A4: Performance of HerdNet on Wildlife Nadir Aerial Images

The nadir open-source dataset of Delplanque et al. (2022) was selected to assess the HerdNet architecture on other species, under different African landscapes and from another viewing angle (**Table 3.8**). As the annotations were provided as bounding boxes, points were defined by simply selecting the centers of the boxes.

HerdNet was trained, validated, and tested independently on the dataset of Delplanque et al. (2022) to assess the potential of this architecture to detect and count wildlife in nadir aerial imagery. Training followed the same procedure of HerdNet as presented in section 3.3.2 of the paper, using the same hyperparameter values. Localization and counting metrics were computed as presented in section 3.3.3 of the paper, except that a larger threshold of 20 pixels was used due to the fine ground sampling distance (**Table 3.8**).

HerdNet was then compared to the state-of-the-art model of the dataset, which was the version of Libra-RCNN, a multilevel balanced anchor-based object detector (Pang et al., 2019), that showed the best performance among the five seeds tested by Delplanque et al. (2022)**.** Despite a decrease in overall recall of about 10%, HerdNet's precision was more than twice that of Libra-RCNN, giving much lower count errors (**Table 3.9**) and far fewer false positives, especially for herds (**Figure 3.13**). However, HerdNet was more confusing for species identification, leading to a higher average confusion score.

Table 3.8: Wildlife nadir dataset details. Note that the 'Annotations' row provides the number of annotations per species, in the order in which they are listed in the 'Species' row.

Country	Democratic Republic of Congo, Botswana, Namibia, South Africa
Park/Reserve	Virunga National Park, Hluhluwe-iMfolozi Park, Phinda Private Game Reserve, The Northern Tuli Game Reserve, NG26 concession, Bwabwata National Park, Mudumu National Park, Madikwe Game Reserve
Biome (Olson *et al.*, 2001)	Tropical and subtropical moist broadleaf forests, Montane grasslands and shrublands; Tropical and subtropical grasslands, savannas, and shrublands
Aerial vehicle	Unmanned Aerial Vehicle (Falcon) and Aircraft (SkyReach BushCat)
Camera	Sony-A6000, Sony-Nex7, Canon 6D
Orientation	Nadir
Altitude	100, 220-2270 m
Images	1297
Image dimension	6000 x 4000 pixels, 5472 x 3648 pixels, 5496 x 3670 pixels, 5521 x 3687 pixels, 5525 x 3690 pixels
GSD	2.4-13.0 cm
Species	African buffalo (*Syncerus caffer*), kob (*Kobus kob*), topi (*Damaliscus lunatus jimela*), warthog (*Phacochoerus africanus*), waterbuck (*Kobus ellipsiprymnus*), African bush elephant (*Loxodonta africana*)
Annotations	1509/2370/2722/433/241/2964
Type	Bounding boxes

n/a, not available; GSD, ground sampling distance

Table 3.9: Binary (animal vs. background) performances of the state-of-the-art model (Libra-RCNN) and HerdNet on full images of the Delplanque et al. (2022) test set. Values in bold indicate the best performance among the two architectures.

Architecture	**Libra-RCNN**	**HerdNet**
Recall	**94.6%**	84.4%
Precision	35.4%	**82.5%**
F1 score	51.5%	**83.5%**
MAE[1]	14.9	**1.9**
RMSE[2]	24.4	**3.6**
Average confusion	**2.9%**	7.8%
Total counting error	167.1%	**2.3%**
Processing time (seconds)	12.0	**3.4**

[1]MAE, Mean Absolute Error; [2]RMSE, Root Mean Square Error.

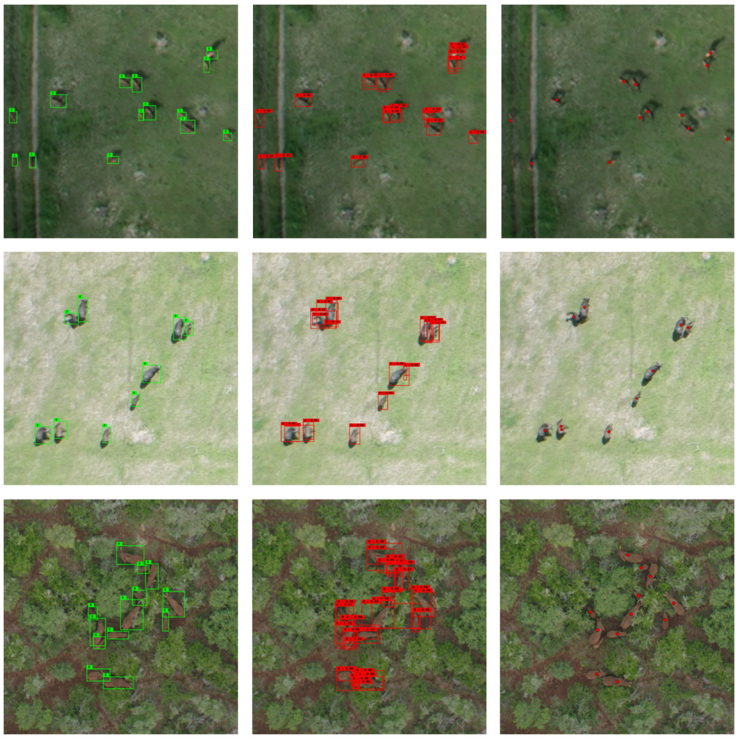


Figure 3.13: Original ground truth (bounding boxes, first column), detection examples of the state-of-the-art model (Libra-RCNN, second column) of Delplanque et al. (2022), and detections of HerdNet (third column) on test patch image samples containing herds

4

Integrating oblique camera systems and deep learning models into aerial survey

Preamble

In **Chapter 3**, I designed a novel DL architecture for precise counting of large African mammals, in response to the limits observed in **Chapter 2** with the use of pre-existing CNN-based object detectors. Across these chapters, the practical implications of such automatic models for species detection, counting and recognition were discussed, prompting considerations regarding their broader applicability and performance in scenarios involving a substantial number of negative images (i.e. images devoid of animals). This fourth chapter thus aims to evaluate and quantify the contribution of continuous imaging and DL to the traditional systematic aerial survey protocol. It is organized into two subchapters. The first subchapter focuses on quantifying the reduction in human workload associated with the manual interpretation of aerial images. The second subchapter investigates whether a semi-automatic model, coupled with continuous oblique imaging, increases the accuracy and/or precision of wildlife population estimates compared to the traditional observer method.

Subchapter 1: Quantifying the reduction of human interpretation

Paper 3 | Published

Surveying wildlife and livestock in Uganda with aerial cameras: Deep Learning reduces the workload of human interpretation by over 70%

Alexandre Delplanque, Richard Lamprey, Samuel Foucher, Jérôme Théau & Philippe Lejeune

This paper is published in *Frontiers in Ecology and Evolution* (IF=2.4), 11, 1270857. DOI: 10.3389/fevo.2023.1270857

Abstract

As the need to accurately monitor key-species populations grows amid increasing pressures on global biodiversity, the counting of large mammals in savannas has traditionally relied on the Systematic-Reconnaissance-Flight (SRF) technique using light aircrafts and human observers. However, this method has limitations, including non-systematic human errors. In recent years, the Oblique-Camera-Count (OCC) approach developed in East Africa has utilized cameras to capture high-resolution imagery replicating aircraft observers' oblique view. Whilst demonstrating that human observers have missed many animals, OCC relies on labor-intensive human interpretation of thousands of images. This study explores the potential of Deep Learning (DL) to reduce the interpretation workload associated with OCC surveys. Using oblique aerial imagery of 2.1 hectares footprint collected during an SRF-OCC survey of Queen Elizabeth Protected Area in Uganda, a DL model (HerdNet) was trained and evaluated to detect and count 12 wildlife and livestock mammal species. The model's performance was assessed both at the animal instance-based and image-based levels, achieving accurate detection performance (F1 score of 85%) in positive images (i.e. containing animals) and reducing manual interpretation workload by 74% on a realistic dataset showing less than 10% of positive images. However, it struggled to differentiate visually related species and overestimated animal counts due to false positives generated by landscape items resembling animals. These challenges may be addressed through improved training and verification processes. The results highlight DL's potential to semi-automate processing of aerial survey wildlife imagery, reducing manual interpretation burden. By incorporating DL models into existing counting standards, future surveys may increase sampling efforts, improve accuracy, and enhance aerial survey safety.

Keywords: wildlife, aerial survey, Deep Learning, remote sensing, convolutional neural networks, animal conservation, livestock, object detection

1. Introduction

As pressures on biodiversity increase across the globe, accurately determining key-species populations is seen as critical in the 'Essential Biodiversity Variables' (EBV) approach to monitoring ecosystem health (Brummitt et al., 2017; Jetz et al., 2019). For over 60 years, the counting of large wildlife species in the expansive savannas of eastern and southern Africa has been addressed using light aircrafts and human counting crews (Gwynne and Croze, 1975; Jachmann, 2001; Norton-Griffiths, 1978). The 'Systematic Reconnaissance Flight' (SRF) technique involves flying an aircraft at low altitude along transects, whilst Rear-Seat-Observers (RSOs) count animals to left and right in strips of terrain defined by markers on the aircraft (Caughley, 1977; Grimsdell and Westley, 1981; Norton-Griffiths, 1978; Stelfox and Peden, 1981). The transects are the sample units, and analysis to derive estimates and margins of error is conducted using the Jolly II Method (Caughley, 1977; Jolly, 1969).

SRF 'counting standards' have been adopted by many eastern and southern African countries to ensure that data meet minimum quality requirements for national and continental-wide trend-analysis of critical flagship such as elephants (CITES-MIKE, 2020; Craig, 2012; Norton-Griffiths, 1978; PAEAS, 2014). These standards define *inter alia* the flying heights and strip-widths for counting, the sampling intensities that should be used, the length of time that RSOs should count animals before rest-breaks, the recording methods and the statistical analysis techniques used. Although these standards can ensure that important technical criteria are met, they cannot account for all human counting bias. Observers may miss cryptic animals, become overstretched when faced with large herds or multi-species groups, and lose concentration in long hot, turbulent flights over monotonous landscapes (Caughley, 1974; Fleming et al., 2008; Jachmann, 2002; Schlossberg et al., 2016). In regard to detection, they have very little time to search and record animals; as the aircraft moves at a ground-speed of 170–180 $km.hr^{-1}$ along the transect, the RSO can hold any particular feature in view for 5–7 seconds (Fleming et al., 2008). For this reason, an optimum RSO strip width of 150 m on each side of the aircraft was derived from experimental studies in the 1970s, and this metric was subsequently embedded within counting standards (Caughley and Goddard, 1975; Norton-Griffiths, 1978; Ottichilo and Khaemba, 2001; Pennycuick and Western, 1972; Stelfox and Peden, 1981).

Despite the long-recognized constraints of RSO-viewing, consistency of method over decades is seen as key in determining trends (Ogutu et al., 2006). Therefore, advances in methods will need to be made incrementally to ensure harmonization with previous surveys. A recent SRF advance in East Africa, known as the 'Oblique-Camera-Count' (OCC), uses digital cameras to record the counting strips to left and right of the aircraft (Lamprey et al., 2020a, 2020b). This replicates the oblique view of the RSOs where animals can be detected under tree canopies. With OCC the observers are not in the aircraft but in the laboratory, and their job is to interpret the many thousands of images obtained in a flight mission.

In recent years, multiple RSO-OCC comparisons have been conducted. Bröker et al. (2019) showed that the abundance estimate of narwhal in Greenland (*Monodon monoceros*) based on oblique-imaging was not significantly different from RSO one. However, Lethbridge et al. (2019) found 30% higher oblique-imaging estimates than RSO ones when surveying Kangaroos in Australia. OCC counts in Kenya and Uganda over the last decade revealed that RSOs had been missing up to 70% of large mammal species, including key cryptic species such as giraffe (Lamprey et al., 2020b). Estimates for smaller animals were greatly increased. In Murchison Falls National Park in Uganda for example, an RSO-based survey estimated 600 oribi (*Ourebia ourebi* ssp. *cottoni*), whilst an OCC survey the following year estimated 12,000 (Lamprey et al., 2020a). Thus the use of cameras is important in resetting baseline population estimates.

The primary advantage of camera-based counts is that time can be spent in the lab to carefully study each image for animals, and that interpreters can cross-check scenes for verification. Conversely, the primary constraint of aerial imaging methods is that thousands of images are acquired that need to be visually interpreted. This is a time-consuming and costly exercise. For example, a standard counting flight transect, involving just 30 minutes of RSO time for detection and recording, would obtain 900 OCC images taken each side of the aircraft. These images will take 4 days to interpret by two interpreters (left and right cameras). It is therefore not surprising that conservation agencies balk at the time and labour costs of OCC counts and other imaging exercises (Bröker et al., 2019; Peng et al., 2020).

Another limitation of the OCC approach is that a very high percentage of aerial images will have no animals. In the arid Tsavo NP in Kenya for example, just 2% of the 160,000 images acquired had animals present (Lamprey et al., 2020b). In Uganda's sub-humid national parks with higher density of wildlife, some 10% of images are positive (Lamprey et al., 2020a). In general, therefore, over 90% of the time of OCC image interpretation is spent on True Negative (TN) images – images with no animals – and if these can be identified and eliminated then there can be significant reductions in human labor.

The next incremental step up from RSO to image-based counting is therefore to accelerate the detection of animals on images. Deep Learning (DL) offers this possibility (Tuia et al., 2022). DL is a subgroup of artificial intelligence approach regrouping machine learning methods based on artificial neural networks, capable of learning and integrating multi-level representation from large datasets (LeCun et al., 2015). Significant progress has already been made in identifying a range of key species in Africa using DL-based object detectors and aerial imagery (Delplanque et al., 2022, 2023a; Eikelboom et al., 2019; Kellenberger et al., 2018; Naudé and Joubert, 2019; Torney et al., 2019). However, DL models produced biased counts because of their current high false positive rate, usually generated by animal-look-alike background objects. Thus, detections still need to be reviewed by humans. Furthermore, the field of animal detection in oblique aerial imagery is not yet as well

developed as that of camera traps, where models trained on large and varied datasets are available for image (pre-)processing (Shepley et al., 2021; Tabak et al., 2019). At the moment, it is therefore often necessary to develop one's own model for application in a given protected area.

Being aware that current DL models need humans for prediction verification, we conducted a study to determine the potential of DL for reducing the interpretation workload of OCC surveys. We asked two specific questions:

1) When the model detects animals in an image that we know are present, how well does it locate, count and identify them?
2) For a 'practical' evaluation to reduce interpretation, can the model discriminate correctly the images which do not contain animals?

2. Methods

We trained a DL model using annotations of a sample of images obtained in an SRF-OCC survey of Queen Elizabeth Protected Area in Uganda. These images had been previously visually interpreted to count animals, with the counts entered into a meta-database. An image could contain nothing and be a TN, or it could be a True Positive (TP) image with (for example) a single warthog, and/or 20 elephants and/or 100 Uganda kob. Having trained the DL model on a range of species from the annotated samples, we then tested the model on a realistic dataset, i.e. visually interpreted images that had not been used in the DL training, which contains both positive and negative images.

2.1. Study area and dataset

The study area is the Queen Elizabeth Protected Area (QEPA) located in southwestern Uganda. The census zone included the Queen Elizabeth National Park and the contiguous Kyambura and Kigezi Wildlife Reserves, covering 2,560 km² of bushed grassland, thicket, open woodlands and forest. Our study is based on aerial imagery acquired for a previous study of wildlife populations of QEPA, conducted in 2018. Only the information necessary for the understanding of the present paper is provided here, for more details the reader is referred to the study of Lamprey et al. (2023).

High-resolution images were acquired using two 24-megapixel Nikon DSLR cameras obliquely mounted at 45° through a camera hatch of a Cessna 182 aircraft. At 600 ft (183 m) above ground level coupled with an aircraft ground speed of 105 knots (194 $km.hr^{-1}$), a 2 second timing interval on cameras provided a continuous sample-strip of 150 m width on the ground ('strip-width') with a 40% overlap between sequential images and frame footprint of 2.1 hectares. The cameras generated sequentially numbered images, stored in incremental folders on the camera cards. Flight transects were spaced at 1 km intervals and a total of 37,000 images were collected with Ground-Sampling Distance (GSD) 2.4 cm at the inner edge and 5.0 cm at the outer edge. These were manually interpreted by a team of four Ugandan

interpreters during a six-week period. For each image, species name and numbers were recorded into a data spreadsheet. Where large herds spanned overlapping images, animals in the overlap area were counted into Even-Number Images (ENIs), while animals were counted in the center portion of Odd-Number Images (ONIs) to avoid any possibility of double counting. Therefore, ENIs contained total counts while ONIs contained partial counts (i.e. only the animals within the gaps between ENIs).

From the manual photo-interpretation, 12 wildlife and livestock species were detected: elephant (*Loxodonta africana*), buffalo (*Syncerus caffer*), topi (*Damaliscus lunatus* ssp. *jimela*), Uganda kob (*Kobus kob* ssp. *thomasi*), waterbuck (*Kobus ellipsiprymnus* ssp. *defassa*), warthog (*Phacochoerus africanus* ssp. *massaicus*), giant forest hog (*Hylochoerus meinertzhageni*), hippopotamus (*Hippopotamus amphibius*), crocodile (*Crocodylus niloticus*), cow (*Bos taurus*), sheep (*Ovis aries*) and goat (*Capra hircus*). Since the management of double counting is beyond the scope of this paper, only ENIs were selected. From all ENIs (18,833), approximately 70% (12,806) were randomly selected for creating annotations, used for training, validation and animal instance-based testing of the DL model, keeping the remaining 30% (6,027) for image-based model testing. Therefore two test sets were established to answer the 2 research questions: 1) the 'animal instance-based' test set, where the annotated points are the ground truth; it was used to answer the first question, and 2) the 'image-based' test set, containing less than 10% of positive images and more than 90% of negative images, where the species counts are the ground truth. This second test set served as a case study and was used to answer the second question.

The animal instance-based dataset was initially annotated as bounding boxes by a team of 4 experienced Ugandan interpreters, using VGG Image Annotator (Dutta and Zisserman, 2019). However, since point annotation has emerged as a faster and better alternative for the detection of animals with DL-based object detectors (Delplanque et al., 2022, 2023a), pseudo-points were created by selecting the center of the bounding boxes. These pseudo-points were finally reviewed by an experienced annotator to obtain body-centered points, as the camera's viewing angle, animal pose or tightness of bounding box drawn may result in a point being outside the animal's body. This has been done using Label Studio software (Tkachenko et al., 2020). The images and points of the animal instance-based dataset were randomly split into training, validation and testing sets following a common allocation of 70%–10%–20% respectively, while taking the species numbers distribution into account (**Table 4.1**). Sheep and goat were amalgamated as a single class due to their great similarity in shape and color given the image resolution.

Table 4.1: Details of the dataset split.

	Animal instance-based dataset				Image-based dataset	
Number of	**Training**	**Validation**	**Test**	**Total**	**Test**	**Prob.[2]**
Elephant	406	58	116	580	299	7.6%
Buffalo	1,258	180	359	1,797	858	23.0%
Topi	172	10	43	225	118	3.0%
Kob	1,526	218	436	2,180	1,137	28.8%
Waterbuck	504	72	143	719	335	9.1%
Warthog	196	28	56	280	172	3.9%
Giant Forest Hog	27	5	8	40	25	0.6%
Hippopotamus	497	71	142	710	351	9.2%
Crocodile	14	2	4	20	16	0.3%
Cow	376	38	227	641	441	9.4%
Sheep/Goat	353	51	100	504	81	5.1%
24MP[1] positive images	717	95	200	1,012	494	-
24MP[1] negative images	0	0	0	11,778	5,533	-

[1]MP, Megapixel; [2]Probability of occurrence in the dataset.

2.2. Deep Learning model

Given its better performances in detecting and counting animals in oblique aerial imagery compared to common DL models, HerdNet (Delplanque et al., 2023a) was chosen to process the dataset. Briefly, HerdNet is a single-stage point-based CNN consisting of two heads, one dedicated to the accurate localization of animals in the image (i.e., points), and the other to their classification, both trained in a pixel-wise manner using the Focal and the Cross-Entropy losses respectively. The training scheme was the same as that presented in Delplanque et al. (2023a) and consisted of two steps: 1) training the architecture using positive patches only, and 2) harvesting and including Hard Negative Patches (HNPs) to further train the model in order to reduce the number of false positives. The patch size was set to 1,024 × 1,024 pixels and following original paper values and early ablation studies, the hyperparameters were set as follows: the learning rate to 10^{-5}, the batch size to 2 and the number of epochs to 100. Horizontal flipping was used for data augmentation, using a 50% probability of occurrence and the Adam optimizer was used for neural network's parameters optimization. During testing, points were obtained by extracting local

maxima from the pixel map produced by the localization head, in which a pixel value close to 1 indicates the presence of an animal. Each point was then used to pin the classification maps and obtain the associated class and confidence score. An image was considered as negative if the maximum pixel value of the localization map did not exceed 0.1. Each full-resolution test image was scanned in a moving-window fashion with a patch overlap was set to 256 pixels. A radial distance threshold of 20 pixels was used to compare ground truths and detections during animal instance-based evaluation. Finally, only detections with confidence score above 50% were retained for image-based evaluation. For more details, the reader is referred to the reference paper. Operations were performed on a Windows-10 workstation using a 64 GB AMD Ryzen 9 5900X central processing unit (CPU) and an 8 GB NVIDIA GeForce RTX 3070 graphics processing unit (GPU).

HerdNet was evaluated in two ways: 1) The 'standard' machine learning way, by calculating common detection metrics on the animal instance-based test set, containing positive images only; and 2) The 'practical' way, by running the model on unseen images of the image-based test set, containing both negative and positive images, and comparing the DL model's counts with interpreters' visual counts. Recall, precision, and F1 score were calculated for each species on the animal instance-based test set for the standard evaluation:

$$\text{recall} = \frac{\#TP}{\#TP + \#FN}$$

$$\text{precision} = \frac{\#TP}{\#TP + \#FP}$$

$$\text{F1 score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$$

where #TP, #FN, and #FP are the number of true positives (i.e., exact detection and identification), false negatives (i.e., missed animals) and false positives (i.e., wrong detections) respectively.

Recall, also referred to as 'true positive rate', measures the proportion of animals correctly detected and identified by the model, while precision measures the proportion of true animals among all detections. The F1 score is the harmonic mean of these two metrics and is higher when recall and precision are balanced.

Concerning the practical evaluation on the image-based test set, only counting comparisons were made as no annotated points were available for calculating the above metrics. The true counting rate, representing the proportion of the human count found by the DL model, and the counting precision, representing the ratio of human count by DL model count, were calculated for each species.

3. Results

3.1. Animal instance-based performance

All species combined, HerdNet reached 85% for both recall, precision and F1 score with little variation in performance according to distance from the aircraft (**Figures 4.1A-B**). Kob, buffalo, waterbuck and elephant were particularly well detected and located, as expressed by recall above 80% in **Figure 4.1C**. Hippopotamus and topi stood just after with a recall close to 60%, and the other species were much less detected. Except for the crocodile and the giant forest hog (i.e., minority species), the precision varied from 44 to 90%, meaning that the model produced respectively between 1.3 and 0.1 false positives per true positive. The least confused species were elephant, hippo and kob while the most confused were cow, warthog and topi. The highest confusions were between cow and buffalo and between topi and kob (**Figure 4.2**).

3.2. Image-based performance

From the image-based test set of 6,027 images, the DL model correctly identified 81.1% of the negative images (4,486/5,533), thus reducing the manual interpretation workload by 74.4% (4,486/6,027). The same tendency was observed when applying the model to the whole set of ENIs: HerdNet identified 80.1% of the negative images (9,487/11,778), reducing the workload by 74.1% (9,487/12,806). In addition, it is worth mentioning that the DL model processed images on the workstation at a rate of about 2.8 seconds per 24-megapixel image, which corresponded to around 10 hours for the entire ENI dataset.

Focusing on detection by species, the model guides the interpreters to 95% or more of the animals for almost all the species studied except warthog, as expressed by the high detection rate in **Table 4.2**. Overall, the model detected 98.2% of animals previously identified in the original 2018 count by interpreters. Meanwhile, the counting precision of the model was low overall at< 50%, but was reasonable for elephant (50.1%) and buffalo (54.1%), and high for topi (92.9%) and cow (90%).

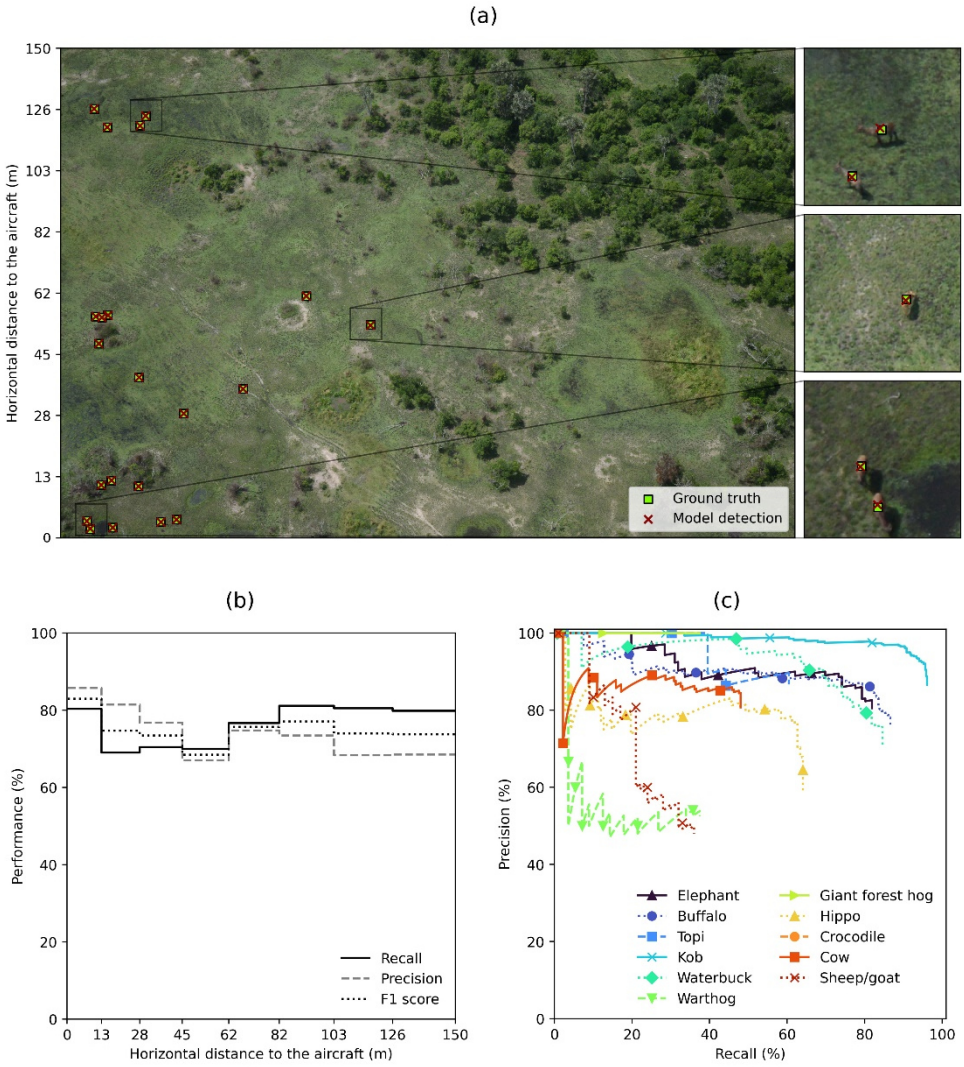


Figure 4.1: Animal instance-based detection performance of the DL model (HerdNet): (A) Example of model detection on a full oblique image, (B) model performance relative to the horizontal distance to the aircraft, and (C) species precision-recall curves.

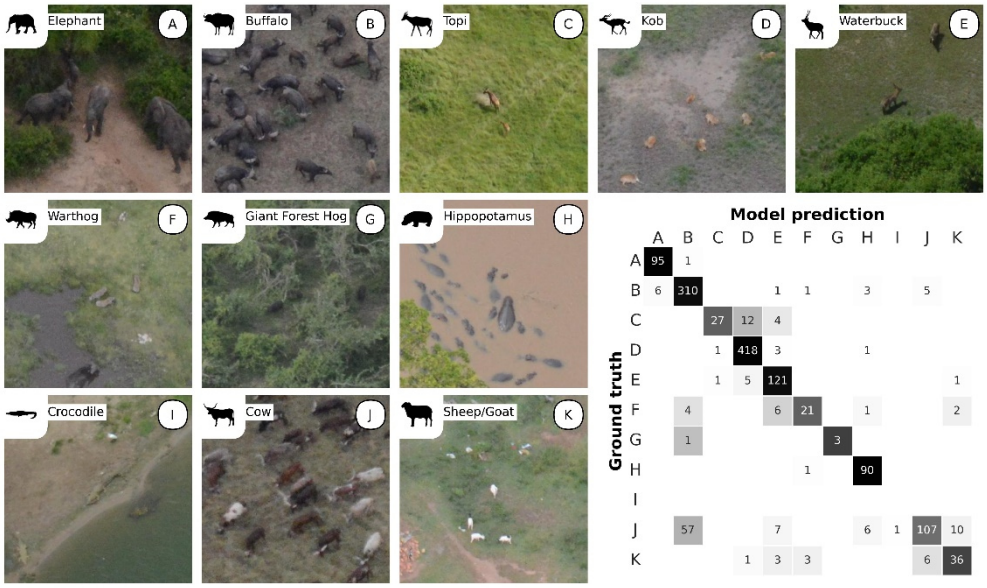


Figure 4.2: Animal instance-based identification performance of the DL model (HerdNet). Each species was assigned a letter for referencing in the confusion matrix (bottom right): (A) Elephant, (B) buffalo, (C) topi, (D) kob, (E) waterbuck, (F) warthog, (G) giant forest hog, (H) hippopotamus, (I) crocodile, (J) cow, and (K) sheep/goat. The confusion matrix shows the comparison between the identification assigned during annotation by the human ('Ground truth') and that predicted by the DL model ('Model prediction').

Table 4.2: Results of the DL model (HerdNet) on the image-based test images (N=6,027).

Species	N_H[1]	$N_{H\|M}$[2]	N_M[3]	$N_{H\|M}$ / N_H[4]	N_H / N_M[5]
Elephant	299 (65)	292 (58)	597 (313)	97.7%	50.1%
Buffalo	858 (51)	852 (46)	1,587 (527)	99.3%	54.1%
Topi	118 (16)	117 (15)	127 (44)	99.0%	92.9%
Kob	1,137 (152)	1,137 (152)	4,092 (1,706)	100.0%	27.8%
Waterbuck	335 (100)	329 (96)	1,348 (841)	98.2%	24.9%
Warthog	172 (61)	143 (46)	701 (514)	83.1%	24.5%
Giant Forest Hog	25 (8)	25 (8)	55 (45)	100.0%	45.5%
Hippopotamus	351 (60)	332 (49)	1,468 (508)	94.6%	23.9%
Crocodile	16 (3)	16 (3)	97 (85)	100.0%	16.5%
Cow	441 (19)	440 (18)	490 (109)	99.8%	90.0%
Sheep/Goat	81 (14)	81 (14)	994 (515)	100.0%	8.1%

[1]Animal numbers in images as determined by human counts ('interpretation') in survey year 2018.
[2]Animal numbers in images from 2018 interpretation, where these images were later classified as animal-positive by the DL model.
[3]Numbers estimated by the DL model; indicating the 'overcount' by the DL model.
[4]True counting rate of the model; the proportion of the 2018 count found by the DL model.
[5]Counting precision of the DL model, where 1/precision is the ratio of the overcount.
The absolute numbers indicated correspond to the number of animals detected, followed by the number of images that contained the species in parentheses.

4. Discussion

In the context of improving multi-species SRF surveys in Africa, we trained a DL model based on aerial imagery of a Ugandan protected area acquired under standardized criteria for OCC surveys, specifically pixel density, camera angles, image footprint size and ground-sampling distance. Our DL model detected human-identified wildlife in positive images at high recall and precision rates (85%). It showed equivalent or better performance than previous DL models developed in similar conditions or habitats (Delplanque et al., 2022, 2023a; Eikelboom et al., 2019; Kellenberger et al., 2018). The CNN used here (i.e., HerdNet) revealed better performance than the study of the original paper (Delplanque et al., 2023a). This may be explained by the higher resolution of current images and their more controlled and standardized acquisition, which should allow for better differentiation of animals in the landscape and within herds and reduced scale variation among individuals.

As previously observed, our model struggles to detect minority species (i.e., crocodile and giant forest hog) certainly due to a lack of training samples for the CNN

to develop robust features. In addition, the inherently small test sample sizes for these species reduce the statistical credibility of the performance. Some of the species showed a low recall while they do not seem more challenging to detect at first sight. This is explained by the difficulty for the model to differentiate visually related species, causing confusion between detected animals. As an example, cow and topi seemed to be poorly detected, but their recall may rise from 47% to 83%, and from 63% to 100% respectively, considering the entire group of confused detected animals (i.e. amalgamated cow and topi). Thus, majority species weighting appears to confuse identification of look-alike species (e.g., cow-to-buffalo confusion). In fact, this phenomenon is common in object detection tasks and is related to 'foreground–foreground class imbalance' (Oksuz et al., 2021), inherent to the imbalance of objects frequencies in nature. Future research should investigate other approaches such as efficient sampling strategies, progressive fine tuning or generative methods (e.g., Wang et al., 2017) to reduce such bias.

We were surprised by the low detection performance of sheep/goat, considering the good results of previous studies involving these species (Delplanque et al., 2023a; Sarwar et al., 2021). We suspect that the use of the HNP mining method during training degraded the sheep/goat detection ability of the model. In this area in Uganda, sheep and goats were mostly found in the villages, where they are not herded (as in arid lands) but roam in small groups around households; villages were an major source of false positives due to the particular bright items found in them, appearing as 'white shapes' of various sizes. Training the model to discard these sheep- or goat-like objects certainly confused the model, as expressed by the 25% drop of recall obtained on the validation set after the second training step.

HerdNet thus correctly detects and counts our studied species in positive images, but what about its performance on a realistic dataset, i.e. containing less than 10% of positive images and more than 90% of negative images? We observed that our DL model succeeded in guiding interpreters to 98.2% of the animals (all species combined). It discriminated more than 80% of negative images, reducing the workload of manual interpretation by 74%. Nevertheless, the counting performance is not yet satisfactory as the model tended to overestimate the true number of animals. This is the result of a high number of false positives, typically generated by unknown or animal-like landscape items such as particular shapes of trunks, shadows, rocks, termite mounds and mud. This model behavior was expected as such landscape items have previously shown to be the main cause of false positives (Delplanque et al., 2022; Kellenberger et al., 2018). Precision could be improved by properly re-training the DL model on these particular landscape items, following a short-time human verification session.

At this time, a sufficient annotated wildlife training dataset acquired of the target area, or of areas with the same wildlife species is required to process all the image data. This training and verification can be accelerated by using point detections,

because adding, deleting or moving points is much faster than adjusting bounding boxes, which makes our model more appropriate for processing aerial surveys images.

Our results confirm and validate that we have entered the era of using DL as a tool to semi-automatically process aerial survey wildlife imagery acquired under standard SRF conditions, with demonstrated effectiveness to reduce human interpretation workloads by over 70%. Humans must remain in the process to study positive images, as filtered by the DL model. Annotated image databases and models will also improve with each new acquisition, and we can therefore anticipate a growing improvement in DL models. Current counting standards such as CITES-MIKE V3 (CITES-MIKE, 2020) can now evolve further to prescribe image-based animal detection based on a combination of manual interpretation and high-performance DL models. Following surveys can invest in increased sampling effort, as the DL model is insensitive to fatigue unlike humans. This can be effected by increasing sampling strip widths, flying higher and using higher resolution cameras, such as the new generation of 40–60 MP mirrorless cameras (Lamprey et al., 2020a, p. 20). On one hand, this would allow for the transfer of the observers' real-time visual counting work to the verification of the model detections. On the other hand, this would decrease the human-life risks associated with traditional aerial surveys while increasing the sampling effort at no extra costs.

In our study we have emphasized the potential use of DL for detection in strip transects. However, the method also has potential for detection in line transects where the population is calculated from a function of the drop-off of observations with distance from a line defined to the side of the aircraft (Buckland et al., 2004; Eberhardt, 1978). To date, problems in measuring distance to aircraft, together with meeting a key assumption of 100% animal detection by observers on the line itself, have precluded the wide use of line transects in Africa (Kruger et al., 2008). However, where pixel position can define the distance from the aircraft, and detection through DL is improved, our approach has the capability to greatly enhance line-transect counts.

Next work will consist of manually verifying detections and producing population estimates. This will enable us to assess the performance of our semi-automated detection model at the scale of an entire aerial survey. On a more general scale, it would be important to develop efficient semi-automated approaches to process large volumes of aerial survey images, integrating Deep Learning and humans with minimal verification time investment, to ensure accurate and precise derived estimates.

Subchapter 2: Comparison of semi-automated and conventional aerial survey estimates

Paper 4 | Published

Will artificial intelligence revolutionize aerial surveys? A first large-scale semi-automated survey of African wildlife using oblique imagery and deep learning

Alexandre Delplanque, Julie Linchant, Xavier Vincke, Richard Lamprey, Jérôme Théau, Cédric Vermeulen, Samuel Foucher, Amara Ouattara, Roger Kouadio, Philippe Lejeune

This paper is published in *Ecological Informatics* (IF=5.8), 82, 102679. DOI: 10.1016/j.ecoinf.2024.102679

Abstract

Large African mammal populations are traditionally estimated using the systematic reconnaissance flights (SRF) with rear-seat observers (RSOs). The oblique-camera-count (OCC) approach, utilizing digital cameras on aircraft sides, proved to provide more reliable population estimates but incurs high manual processing costs. Addressing the urgent need for efficiency, the research explores whether a semi-automated deep learning (SADL) model coupled with OCC improves wildlife population estimates compared to the SRF-RSO method. The study area was the Comoé National Park, in Ivory Coast, spanning 11,488 km² of savannas and open forests. It was surveyed following both SRF-RSO standards and OCC method. Key species included the elephant, western hartebeest, roan antelope, buffalo, kob, waterbuck and warthog. The deep learning model *HerdNet*, priorly pre-trained on images from Uganda, was incorporated in the SADL pipeline to process the 190,686 images. It involved three human verification steps to ensure quality of detections and to avoid overestimating counts. The entire pipeline aims to balance efficiency and human effort in wildlife population estimation. RSO and SADL-OCC approaches were compared using the Jolly II analysis and a verification of 200 random RSO observations. Jolly II analysis revealed SADL-OCC estimates significantly higher for small-sized species (kob, warthog) and comparable for other key species. Counting differences were mainly attributed to vegetation obstruction, RSO observations not found in the images, and suspected RSO counting errors. Human effort in the SADL-OCC approach totaled 111 hours, representing a significant time savings compared to a fully manual interpretation. Introducing the SADL approach for aerial surveys in Comoé National Park enabled us to address the OCC's time-intensive image interpretation. Achieving a significant reduction in human workload, our method provided population estimates comparable to or better than SRF-RSO counts. Vegetation obstruction was a key factor explaining differences, highlighting the OCC

method's limitation in vegetated areas. Method comparisons emphasized SADL-OCC's advantages in spotting isolated, small and static animals, reducing count variance between sample units. Despite limitations, the SADL-OCC approach offers transformative potential, suggesting a shift towards DL-assisted aerial surveys for increased efficiency and affordability, especially using microlight aircraft and drones in future wildlife monitoring initiatives.

Keywords: wildlife, population estimation, aerial surveys, deep learning, biodiversity monitoring, conservation technology, African savanna

1. Introduction

Although biodiversity loss has a significant impact on Earth's ecosystem functions (Cardinale et al., 2012), it is still accelerating following the growth of human population, consumption rates and the continuing pressure humans exert on the biosphere (Ceballos and Ehrlich, 2023). Determining and tracking key-species populations with standardized data collection is seen as critical in the 'essential biodiversity variables' (EBVs) for effective biodiversity assessment and conservation (Jetz et al., 2019). Among the many existing census methods, aerial surveys are still the most economical and quicker way to count large mammals in Africa's large savanna protected areas (PAs) (Norton-Griffiths, 1978).

Counting large terrestrial wildlife species and livestock has traditionally relied on the 'systematic reconnaissance flight' (SRF) method. SRF consists of aircraft flying at low altitude along predefined transects, while rear-seat observers (RSOs) count animals in right and left sample strips defined by markers attached to the aircraft (Grimsdell and Westley, 1981; Norton-Griffiths, 1978). While this technique has been adopted as a standard in African savannas (CITES-MIKE, 2020; Craig, 2012; Norton-Griffiths, 1978; PAEAS, 2014), it suffers from human counting errors as RSOs may under- or overcount large herds, miss species, or lose attention during long flights. Counting animals on sight is challenging. It is often biased by survey factors such as altitude, sample strip width or observer experience (Caughley, 1974; Jachmann, 2001; Norton-Griffiths, 1976), but also by environmental factors such as animal size and color, animal's disturbance caused by an overflying aircraft, group size or vegetation type and density (Griffin et al., 2013; Jachmann, 2002; Wal et al., 2011). To minimize the impact of some of these factors, aerial survey standards for fixed-wing aircraft have been established (CITES-MIKE, 2020; Craig, 2012; PAEAS, 2014). However, the high flight speed (150-190 km/h), essential to ensure the crew's safety at common flying altitude (90-100 m), leaves the observer only a small window of time to scan the terrain and to count animals. This window being estimated at only 5-7 seconds (Lamprey et al., 2020b), observers may be overloaded in high-density animal environments or tired in low-density ones, which could both lead to biased counts (Norton-Griffiths, 1976). Although not always adopted by practitioners, photographing herds for post-processing is a beneficial practice during aerial surveys, as even experienced observers are unable to accurately count groups of more than 20 individuals (Norton-Griffiths, 1978). It is even recommended to photograph any

group of more than 10 individuals in the case of multi-species survey (CITES-MIKE, 2020; Norton-Griffiths, 1978). Counts derived from the images are then used to correct in-sight count and provide unbiased estimates.

To compensate for the limits of the SRF method, the oblique-camera-count (OCC) approach has recently been developed and has proved to increase and precise the estimates of large African mammal species in semi-arid environments (Lamprey et al., 2020a, 2020b). The OCC approach is based on digital cameras placed on both right and left sides of the aircraft, replicating the oblique viewing angle of RSOs which is the most suitable for counting animals in areas with vegetation cover (Lamprey et al., 2020b). These cameras are set to acquire images continuously during the SRF. With this method, the work of observers has shifted from in-sight animal counting in the aircraft to image interpretation in the lab. Nevertheless, counting animals in aerial imagery is a time-consuming exercise which may generate considerable costs, making the approach too expensive for a broader use at present (Bröker et al., 2019; Lamprey et al., 2020b). Previous studies showed that interpreters were able to interpret nearly 150 nadir images per hour from a mono-species drone survey in homogeneous Asian open grasslands (Peng et al., 2020) but only 30 oblique images per hour from a multi-species aerial survey in heterogeneous semi-arid African environments where many variables, including vegetation type, are measured (Lamprey et al., 2020b). While being essential for rapidly validating or establishing conservation actions, results from aerial surveys of PAs covering thousands of square kilometers and generating thousands of images can be delayed by several months using the OCC approach due to the slow but necessary manual processing of images.

Recent advances in machine learning have propelled the perspectives of remotely sensed imagery for wildlife conservation (Tuia et al., 2022), and announced good prospects for the automation of image processing from SRF-OCC surveys (Delplanque et al., 2023b; Eikelboom et al., 2019). Deep learning (DL) is a subgroup of machine learning and artificial intelligence (AI) where artificial neural networks are trained to achieve challenging tasks (e.g. detect animals in aerial imagery) through a complex multi-level representation of information learned from a large amounts of data (LeCun et al., 2015). In the last decade, DL has been widely employed to (semi-)automate the detection and counting of multiple terrestrial mammals on aerial imagery acquired in natural and wild environments, through mainly DL-based object detection approaches (Delplanque et al., 2023a, 2022; Eikelboom et al., 2019; Kellenberger et al., 2018; Naudé and Joubert, 2019; Peng et al., 2020). However, counting results obtained with these approaches remain biased, principally for rare species, due to the high false positive rate of current DL models and to the limited dataset availability. In addition, a time-consuming annotation phase on a subset of acquired images is generally required prior to the development of a model for a specific PA on which an SRF-OCC survey is to be carried out, as a data discrepancy usually appears between different OCC surveys. This is generally caused by both survey (camera angle, image resolution, flight altitude) and environmental factors (natural imbalance of species, landscape heterogeneity).

Pending the development of foundation DL models trained on massive amounts of aerial images, there is a strong need to develop efficient approaches for integrating existing DL models into the aerial survey process. This will reinforce the efficiency of the OCC method and reduce the associated cost by lightening the workload of human interpreters. The goal of this paper was to answer the following research question: Does a semi-automated approach requiring minimal human effort increase the accuracy and/or precision of population estimates compared to the traditional RSO approach? This paper presents the results of the first semi-automated aerial image processing pipeline applied on SRF-OCC images over a large and heterogeneous PA in Ivory Coast.

2. Materials and Methods

2.1. Study area

The study area is the Comoé National Park located in Ivory Coast, which covers 11,488 km², making it the third biggest PA of west Africa. The CNP is covered at 64.3% of shrub savanna, 24.3% of wooded savanna and 7.6% of open forest. In addition, with patches of dense dry forest located in the south of the CNP as well as gallery forests along the shorelines of both Comoé and Iringou rivers, the CNP is an example of transitional habitats between forest and savanna. The park belongs to the 'northern plateaux' geophysical region (average altitude of 300 m) and is locally dominated by a number of reliefs, such as north-south-trending greenstone hills and bars rising to 500-600 m in the north-central and north-western regions; and tabular mounds with armored summits on shale, locally exceeding 500 m, in the south-east. The climate in this region is tropical savanna with dry winters (Aw). Due to its high diversity of habitats, the CNP is an important biodiversity reservoir (Hennenberg et al., 2006) and contains populations of wild terrestrial mammals, such as the roan antelope (*Hippotragus equinus* ssp. *koba*), the western hartebeest (*Alcelaphus buselaphus* ssp. *major*), or the buffalo (*Syncerus caffer* ssp. *brachyceros*), as well as endangered species such as the elephant (*Loxodonta africana*) (Fischer et al., 2002) and the emblematic chimpanzee (*Pan troglodytes* ssp. *verus*).

2.2. Aerial survey

Following previous aerial survey protocols of the CNP, the latter has been divided into four strata (**Figure 4.3**): North-West (NW), North-East (NE), South-West (SW) and South-East (SE). Following standard SRF guidelines, 156 transects of 2 km spacing were oriented north-west to south-east in the northern strata, and north-east to south-west in the southern strata, covering 13% of the area. These orientations followed the ecological gradient of the area (rivers and mountains) while avoiding the aircraft pilot to be glared by the sun during the flights. Twelve days totaling 54 flight hours were needed to cover the entire CNP, starting on April 2 and ending on April 17, 2022.

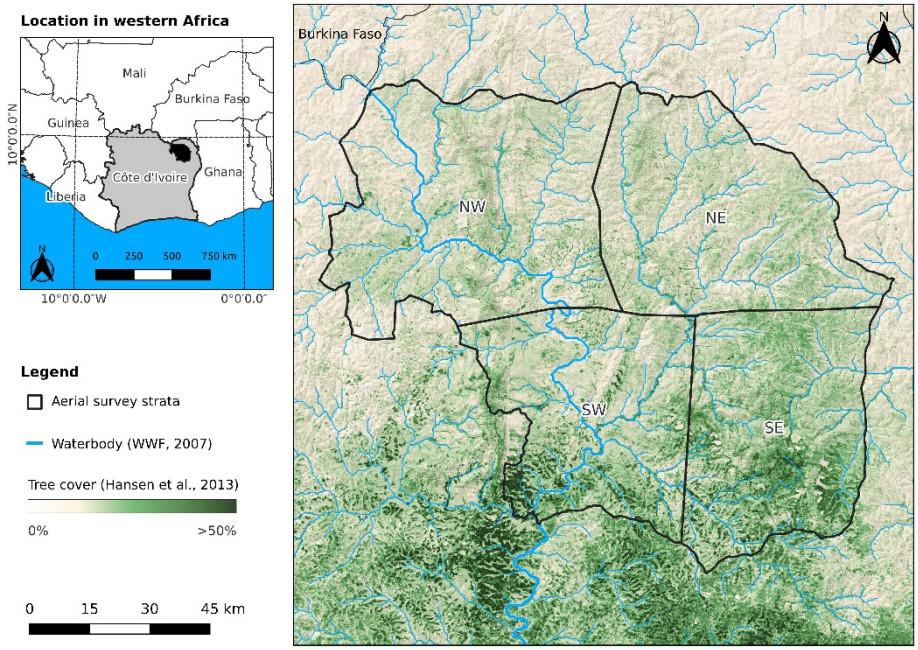


Figure 4.3: Map of the Comoé National Park and survey area strata: North-West (NW), North-East (NE), South-West (SW) and South-East (SE).

The aircraft was a Cessna 206 (registration 5Y-AKP) and the flying crew was composed of: a pilot, an independent front-seat observer, two rear-seat observers (RSOs) and a photography manager. The same crew operated throughout the survey. During flights over the transects, RSOs were instructed to report the number of individuals and the associated species observed between the two strip markers placed on each of the aircraft's struts. The start and end of the transects were announced by the pilot. The front-seat observer was in charge of recording observations from the RSOs and their geo-locations on a tablet computer using the CyberTracker (https://cybertracker.org/) app (v3.520). The photography manager managed the two oblique cameras set to acquire aerial imagery continuously and was also in charge of manually recording observations on papers for back-up. It is worth mentioning that due to a lack of space in the aircraft cabin, additional RSO cameras for large group bias correction (CITES-MIKE, 2020; PAEAS, 2014) have not been used during the survey.

From the multiple species counted during the survey, only seven species usually surveyed by aircraft, in fairly large numbers and/or of conservation interest were selected for this study. These species are referred to as 'key species' in the text and include western hartebeest, buffalo, kob (*Kobus kob* ssp. *kob*), waterbuck (*Kobus*

ellipsiprymnus ssp. *defassa*), elephant, roan antelope and warthog (*Phacochoerus africanus* ssp. *africanus*).

Transect (or sample unit) area were estimated using the height above ground level and a theoretical predefined strip-width of 150 m on each side of the aircraft at a flight altitude of 91.4 m (300 ft) (CITES-MIKE, 2020; Craig, 2012; Norton-Griffiths, 1978; PAEAS, 2014). The height above the ground level as well as associated geo-location were recorded each second during the flights by a LightWare SF30/D laser altimeter. The strip width was calibrated using 20 ground-marks placed 20 m apart on either side of the runway. Thirty crossings of the aircraft at increasing height above ground level (between 55 and 208 m) were carried out, during which the number of marks appearing between the strip markers was counted.

2.3. Cameras and image acquisition

Two Nikon D5600 24-megapixel digital reflex cameras equipped with Nikkor AF-S 18-55 zoom lenses were positioned inside the aircraft, one on each side, using articulated double suction cups fixed on the windows. The cameras were mounted obliquely at an inclination of 36.5° and zoom lens were set and taped at 35 mm to capture a strip of about 150 m width, in accordance with SRF standards (CITES-MIKE, 2020; Craig, 2012; PAEAS, 2014) and recent OCC studies (Lamprey et al., 2020a, 2020b). The camera angle was chosen to be as close as possible to the angle of vision of human observers while intercepting the strip markers at the inner and outer edges of the images. External intervalometers were used and set to acquire images at 2 s intervals, to ensure overlapping coverage at a ground speed of 160 km/h. Based on initial field trials, cameras were set to 'aperture-priority' mode, with aperture set to f/5.0, the auto-ISO was preferred, with a minimum value of 500, and minimum shutter speed was set to 1/2000 s. In total, 190,686 images were saved in 6,000 x 4,000 pixels JPEG format, from which 148,239 appeared on transect after cleaning. All images were geo-referenced in UTM coordinates using the GeoSetter (https://geosetter.de/) software which associated the altimeter's GPS tracklog to the exact acquisition time of each image.

2.4. Deep learning model

The DL architecture HerdNet (Delplanque et al., 2023a) has been selected for processing the aerial survey images given its attractive performances on a previous SRF-OCC study (Delplanque et al., 2023b). HerdNet is a single-stage fully convolutional neural network built with two heads: one dedicated to the accurate point detection of animals in the image and the other to their classification.

HerdNet was trained multiple times during image processing, for progressively fine-tuning it to the study area landscape and species. For each training stage, we constructed an unbiased validation set comprising 20% of the current dataset's images. Each set was carefully designed to maintain a similar distribution of species and to ensure independence by keeping images from the same transect grouped together.

This avoided as far as possible any performance bias related to the natural imbalance of species, and any spatial bias related to the overlap of images. As for the hyperparameters, we set training patch size to 512x512 pixels, the minibatch size to 8 patches, the learning rate to 10^{-6}, the weight decay to $5x10^{-4}$ and the number of epochs to 200 or 50, depending on the training schedule (see section 2.5.2 below). Horizontal flipping and motion blur have been used as data augmentation, with a 50% probability of occurrence. To avoid any risk of overfitting at each training stage, we selected the model relative to the epoch that gave the best performance on the validation set. During inference, the patch size was set to 1,024x1,024 pixels to accelerate the process. Further information on the fine-tuning process is described in section 2.5.2.

2.5. Image processing

Images have been processed through the use of the DL model coupled with human manual interpretation steps. This 'human-assisted' DL-based image processing pipeline has been designed to minimize human effort while maximizing the quality of counting results. In the following sections, this approach is referred to as the Semi-Automated Deep Learning (SADL) model, and SADL-OCC refers to the integration of the SADL model with the OCC technique. This section therefore presents the main components and steps of this developed approach.

2.5.1. Semi-automatic loop

The core component of the pipeline was the Semi-Automatic Loop (SAL), which integrates both the DL model and a human-expert interpreter. The SAL operated by taking aerial images as input and passing them through the DL model to harvest point detections. Subsequently, it conducted a 256x256 pixel crop, centered on each detection, generating thumbnails that received a rapid examination during the initial human verification step. This verification step entails manually classifying each thumbnail as either False Positive (FP), True Positive (TP), or uncertain object (**Figure 4.4**).

This first human verification step played a crucial role in significantly reducing the number of full-size images requiring review, thereby minimizing the overall analysis time. Additionally, this step served as a guide for the interpreter, directing attention to the most relevant detections (i.e. TP). After this step, the relevant detections were projected back into their original full-size images for a second verification. In the second human verification step, the interpreter thoroughly examined the entire image to point out any potentially missed animals and, if necessary, had the option to rectify the predicted identification (species name) and/or point coordinates (**Figure 4.4**). The second step has been done on Label Studio 1.3 (Tkachenko et al., 2020) through a custom template.

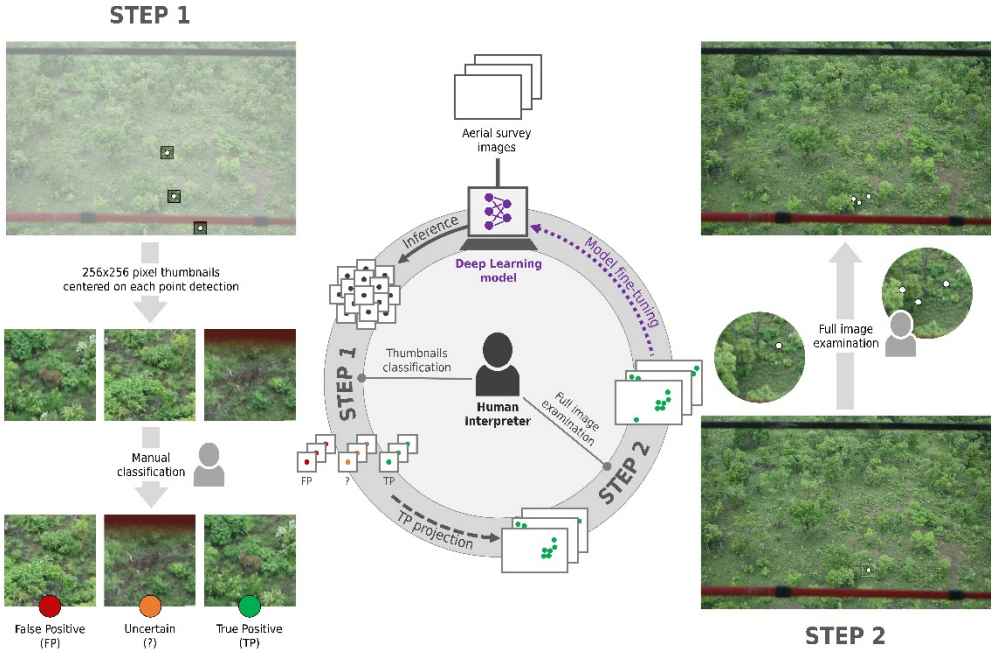


Figure 4.4: Overview of the semi-automatic loop (SAL). The central part of the figure is a schematic representation of the loop, and the sides illustrate the two main steps on a sample image of the aerial survey. TP and FP referred to True Positive and False Positive respectively.

2.5.2. Model fine-tuning and inference

In aerial surveys of PA using the OCC technique, the availability of a region-specific DL model, encompassing the PA's unique species and landscape characteristics, remains limited as these approaches are still emerging and such data are often sensitive. Consequently, a pre-trained DL model is needed. Such a model has been trained on images from a different source (e.g. another PA) but following a similar task (e.g. detecting terrestrial mammals in oblique aerial imagery). Ideally, the pre-trained model should originate from a similar PA containing similar target species and following the same acquisition standards to achieve optimal results. Nevertheless, it is essential to recognize that while these pre-trained models may yield reasonably accurate predictions at times, they are not entirely reliable and risk losing accuracy as the discrepancy between source and target data widens. Therefore, a crucial step involves fine-tuning the model to suit the targeted PA. As some researchers shared their model in recent years (Delplanque et al., 2023a, 2022; Eikelboom et al., 2019), we propose a simple yet effective method that could be applied across various cases.

It is essential to select the most densely populated region to guarantee a sufficient number of instances per species for the optimal fine-tuning of the DL model. In this

study, the SW stratum was selected, which encompassed three distinct flights conducted over three consecutive days.

Regarding the pre-trained model, we utilized the DL model developed by Delplanque et al. (2023b) that was initially fine-tuned on images acquired in a survey of Queen Elizabeth National Park, Uganda (Lamprey et al., 2023) following OCC procedures developed in Murchison Falls National Park, Uganda (Lamprey et al., 2020a). The data acquisition conditions closely resembled the current study's data acquisition process, encompassing similar wildlife species.

The pre-trained model underwent inference and fine-tuning for 4 iterations using the entire SW stratum employing the SAL. This iterative process served to enhance the model's performance and gather samples pertaining to each key species present in the region. The training procedure for the two first fine-tuning iterations was the one proposed in the original paper (Delplanque et al., 2023a) which consisted of two main steps: 1) training the architecture using positive patches for 200 epochs, and 2) collecting and including hard negative patches, which are patches containing false positives, to further train the model for 50 epochs in order to reduce the number of false positives. During the two last fine-tuning iterations, only the second step of the training procedure was used. Hard negative patches were created using false positives that emerged from the thumbnail classification (step 1 of the SAL). To avoid a too severe imbalance between positive and negative patches the batch was equally balanced between the latter during training.

Once the fine-tuning process was done, the model was inferred on images from the other strata. The detections resulting from this inference were subjected to verification using the SAL. To both maximize the probability of detecting all species individuals and take advantage of the collected verified detections, the model was trained one last time on the entire set of verified images and then inferred on all 148,239 transect images for a final verification. From this process, the previously unseen images were verified using the SAL.

2.5.3. Duplicate removal

Due to overlapping coverage of images, the same animal may be present in multiple images, which may lead to an overestimation of the true number of individuals. It was therefore necessary to carefully manage consecutive images to avoid double counting. This has been done on Label Studio 1.3 (Tkachenko et al., 2020) by a human operator who manually reviewed consecutive images and assigned an additional label to the detections to distinguish and discard duplicates.

2.6. Data analysis and population estimate

Prior to comparing RSO and SADL-OCC approaches, counting bias was checked between right and left RSOs. For each species and stratum, the number of groups encountered as well as the number of animals counted in the groups were compared

using a chi-square test and a Mann-Whitney U-test, respectively (CITES-MIKE, 2020).

RSO and SADL-OCC counting results were analyzed using the Jolly II method for unequal sized sample units (Jolly, 1969), where the transects were the sample units, following the guidelines of Norton-Griffiths (1978). As the survey area was stratified, the method of Jolly II was applied on each stratum. The population estimates and variances were calculated for each stratum, and these were then added together to obtain estimates for the whole survey area. The global standard error was calculated by taking the square root of the summed variances (Norton-Griffiths, 1978). To compare the RSO and SADL-OCC surveys, the null hypothesis that estimates were not significantly different (α=0.05) was tested by calculating *d* as:

$$d = \frac{\hat{Y}_S - \hat{Y}_R}{\sqrt{\sigma_S^2 + \sigma_R^2}}$$

Where $\hat{Y}_S$ and $\hat{Y}_R$ are the population estimates of SADL-OCC approach and RSOs respectively, and σ_S^2 and σ_R^2 are their variance (Norton-Griffiths, 1978). Where d > 1.96, the result is statistically significant at alpha=0.05.

RSO and SADL-OCC approaches were also compared using the paired *t*-test and the non-parametric Wilcoxon signed-ranks test, where the samples were the transects (Lamprey et al., 2020b).

Finally, in order to evaluate and explain potential counting differences between SADL-OCC and RSO, 50 RSO observations were randomly selected for each of the following group size class announced by RSOs: 1) 1 to 5 animals, 2) 6 to 10 animals, 3) 11 to 20 animals and 4) 21 and more animals. At each of these 200 locations, an experienced human operator compared the RSO count with the SADL-OCC count. An explanation of the differences observed was provided following a visual analysis of the matching images (**Figure 4.5**) as follows: 1) part of the group is probably hidden by vegetation, 2) a suspected counting error of RSOs, 3) part of the group is out-of-strip on the matching images, 4) the group was missed by the SADL-OCC approach or 5) the group observed by RSO does not appear on the matching images.

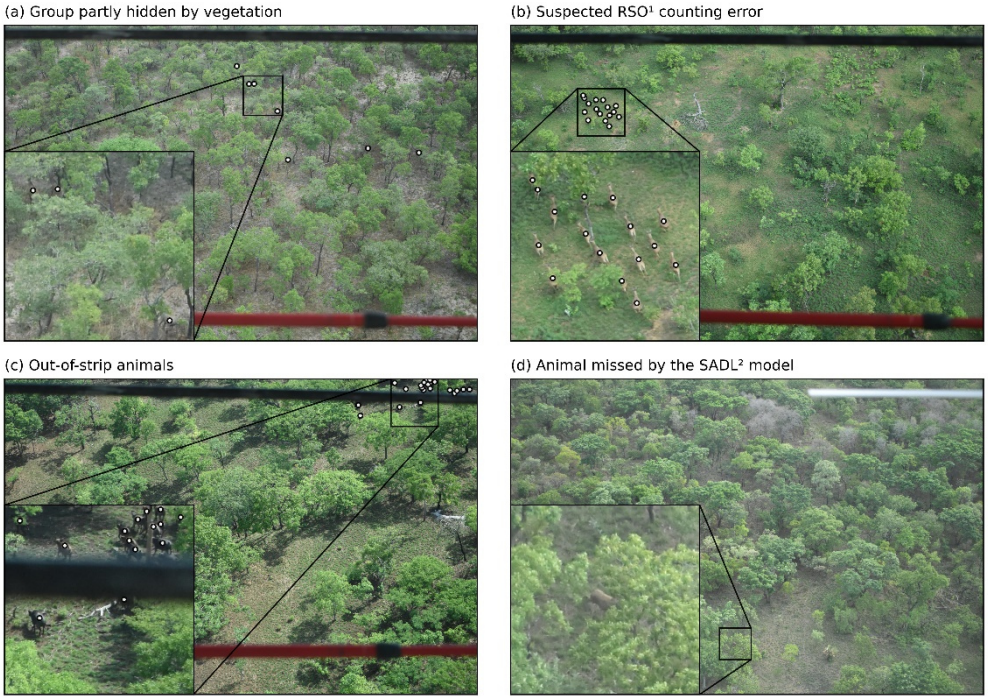


Figure 4.5: Illustration of differences observed between the SADL-OCC and RSO approaches for 200 random RSO observations: (a) a group of 7 roan antelopes detected by the SADL-OCC approach, where some individuals were probably hidden by trees since the RSO announced a group of 35 individuals, (b) a group of 17 western hartebeests estimated at 20 individuals by the RSO indicating a probable RSO counting error, (c) a group of buffalo where most of the individuals appeared out-of-strip in the SADL-OCC approach, but where all individuals were counted in-the-strip by the RSO, and (d) an example of image containing a roan antelope missed by the SADL-OCC approach.

3. Results

3.1. RSOs consistency and Jolly II analysis

Testing consistency between right and left RSO indicated no significant differences in encounter rates for the western hartebeest, kob, waterbuck, elephant and warthog. The exceptions were for roan antelope in SW stratum with 21 (right) and 7 (left) encounters (χ^2=7.00, d.f.=1, P=0.008) and buffalo in SE stratum with 5 (right) and 0 (left) encounters (χ^2=5.00, d.f.=1, P=0.025). Concerning the number of animals reported, only buffalo in SE stratum and waterbuck in SW stratum showed a significant difference. Median buffalo counts of right and left RSO were 35 and 0, respectively (U=0, n1=5, n2=0, P<0.001) while median waterbuck counts were 1 and 3 (U=1, n1=5, n2=3, P=0.025).

The Jolly II analysis showed that SADL-OCC population estimates were significantly higher than RSO ones for small-sized species, i.e. kob and warthog, and not significantly different for the other key species (i.e. elephant, buffalo, western hartebeest, roan antelope and waterbuck). Similarly, results of the paired transect *t*-test and the non-parametric Wilcoxon signed-ranks test indicated the same trend (**Table 4.3**). For kob and warthog, the difference in estimates was highly significant ($p < 0.001$) with tighter confidence intervals, indicating that the SADL-OCC approach counted much more individuals than RSOs and that the counts were more consistent across the survey area. SADL-OCC estimates for kob, warthog and buffalo were respectively 240%, 163% and 17% higher than RSO estimates, while being lower for roan antelope (-19%), western hartebeest (-7%), and waterbuck (-2%). While the elephant population was unfortunately too small for drawing valid consideration, the results showed that the SADL-OCC approach found and correctly counted the two groups observed during the aerial survey.

The SADL-OCC approach estimates are systematically higher than RSOs for each key species in western strata (i.e. NW and SW), while the inverse trend was observed for eastern strata (i.e. NE and SE), except for buffalo, kob and warthog where the results vary. For instance, SADL-OCC buffalo estimates are nearly two times higher than RSO ones in the NW stratum, but nearly three times lower in the SE stratum (**Table 4.3**).

3.2. Counting differences

Based on the additional analysis of 200 randomly-selected RSO observations, 35 instances (17.5%) of mutual agreement with the SADL-OCC approach were observed, ranging from 1 observed animal to a group of 16 individuals. This leaves 165 instances (82.5%) where differences were observed. These differences were mainly explained (35.8%) by the presence of relatively dense vegetation (**Figure 4.6**), hiding some individuals of the group in the image. The second most observed situation (26.7%) was a group of animals observed by RSOs but not found on images, which often contained highly vegetated scenes. The third explanation (24.2%) of differences was the suspected error of RSOs when estimating the group of animals they observed. Finally, animals of the group that appeared out-of-strip on the image(s) explained around 9.7% of the differences, and animal or group of animals missed by the SADL-OCC approach (total of 22 animals) explained 3.6%. It should be noted that the presence of vegetation is an explanatory cause that cannot be excluded for the others and may be a secondary explanation of the difference observed.

Table 4.3: Jolly II estimates ($\hat{Y}$) and standard error (SE) for SADL-OCC[1] ($\hat{Y}_S$) and RSO[2] ($\hat{Y}_R$) surveys of key species in Comoé National Park, using the stratified statistical scheme, and results of the d-statistic (Norton-Griffiths, 1978) and the paired transect t-test (df = 154) and Wilcoxon signed-ranks test for comparison. The final column indicates the extent to which SADL-OCC estimates are superior to RSO estimates, and was calculated as $\Delta\% = (\hat{Y}_S/\hat{Y}_R) - 1$ (Lamprey, Pope, et al., 2020).

	North-West		North-East		South-West		South-East		Total				SADL-OCC vs RSO			
Species	$\hat{Y}_S$ (SE)	$\hat{Y}_R$ (SE)	$\hat{Y}_S$ (SE)	$\hat{Y}_R$ (SE)	$\hat{Y}_S$ (SE)	$\hat{Y}_R$ (SE)	$\hat{Y}_S$ (SE)	$\hat{Y}_R$ (SE)	$\hat{Y}_S$ (SE)	$CI_{95\%}$	$\hat{Y}_R$ (SE)	$CI_{95\%}$	d-stat (p)	t-stat (p)	W (p)	Δ%
Western hartebeest	5243 (713)	4972 (869)	2287 (274)	2630 (428)	7716 (546)	7424 (901)	2316 (356)	3793 (621)	17562 (1005)	± 11%	18819 (1461)	± 15%	-0.709 (0.479)	0.979 (0.329)	2918 (0.386)	-7%
Buffalo	425 (200)	220 (105)	9 (6)	9 (6)	2669 (754)	1852 (710)	284 (190)	813 (476)	3387 (803)	± 46%	2894 (861)	± 58%	0.419 (0.676)	-0.308 (0.759)	140 (0.775)	17%
Kob	1743 (381)	520 (187)	454 (126)	213 (107)	7766 (799)	2102 (425)	181 (54)	142 (70)	10143 (896)	± 17%	2977 (482)	± 32%	7.045 (<0.001)	-4.592 (<0.001)	432.5 (<0.001)	241%
Waterbuck	249 (77)	73 (44)	250 (76)	694 (261)	893 (160)	275 (123)	168 (66)	542 (176)	1559 (204)	± 26%	1585 (341)	± 42%	-0.064 (0.949)	0.123 (0.902)	535.5 (0.159)	-2%
Elephant	0 (0)	0 (0)	0 (0)	0 (0)	275 (133)	225 (109)	0 (0)	0 (0)	275 (133)	± 95%	225 (109)	± 95%	0.290 (0.772)	-1.419 (0.158)	0 (0.157)	22%
Roan antelope	930 (255)	820 (239)	500 (88)	833 (236)	1560 (210)	1535 (300)	755 (210)	1432 (404)	3745 (401)	± 21%	4621 (605)	± 26%	-1.206 (0.230)	1.408 (0.161)	1011 (0.687)	-19%
Warthog	849 (158)	278 (99)	111 (29)	46 (30)	1785 (209)	584 (125)	200 (71)	213 (97)	2946 (273)	± 18%	1121 (189)	± 33%	5.498 (<0.001)	-4.078 (<0.001)	328 (<0.001)	163%

[1]SADL-OCC, semi-automated deep learning oblique-camera-count; [2]RSO, rear-seat observer.

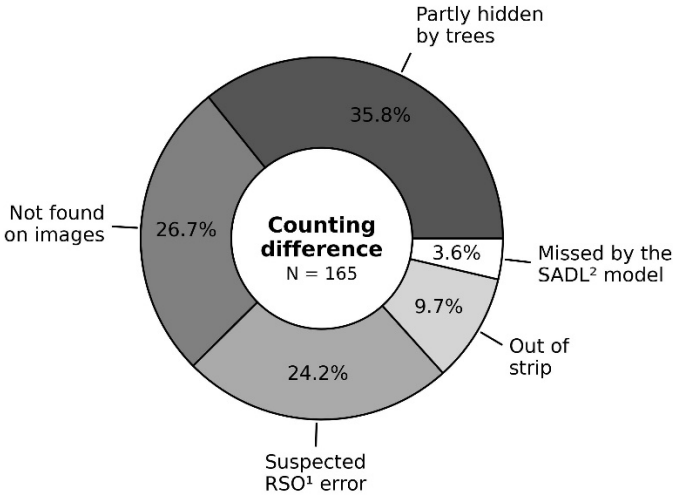


Figure 4.6: Distribution of explanatory causes for differences in counts observed between the SADL-OCC and RSO approaches. The percentages were calculated from the 165 random observations showing differences in counts. Mutual agreements (i.e. no differences) were observed for 35 observations.

Comparing the sample of 200 RSO count values with those derived from the SADL-OCC approach for each key species, it was observed that in most cases large groups were underestimated by the SADL-OCC approach (**Figure 4.7**). This was mainly due to the vegetation cover, the absence of the group in the acquired images and because part of the groups appeared out-of-strip in the images. The resulting differences were particularly severe for the groups observed in the SE and NE strata and for large groups (> 20 animals) of western hartebeest, buffalo, waterbuck and roan antelope (**Figure 4.7**).

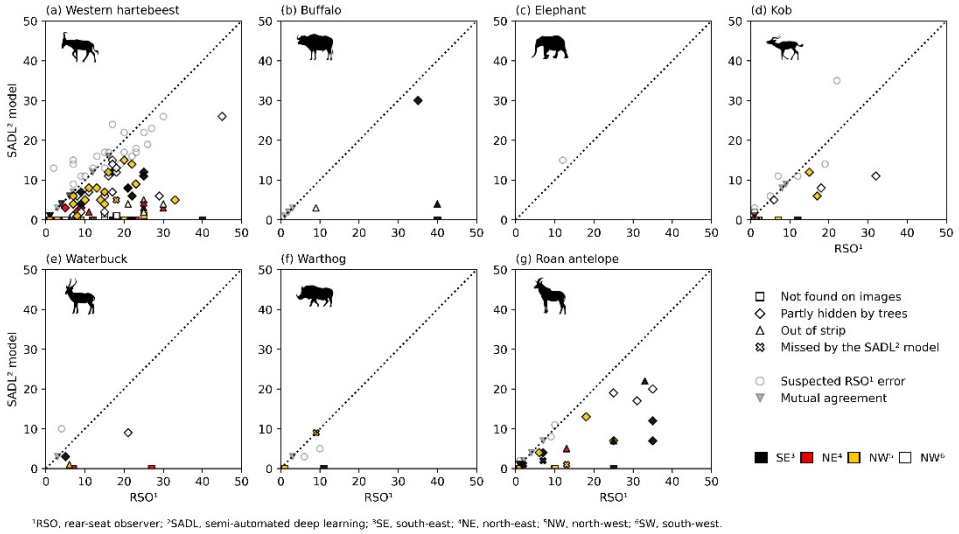


Figure 4.7: Scatter plots between count values announced by RSOs and those derived from the SADL-OCC approach, for each key species. These plots were constructed on the basis of the 200 random RSO observations examined visually. Point markers are differentiated according to the most likely explanatory cause and shaded according to the strata.

3.3. Human effort

The human time investment in the SADL-OCC approach was around 111 hours. More than half of this time was devoted to full examination of the 24 megapixel (MP) images (i.e. step 2 of the SAL), a third to classifying the thumbnails (i.e. step 1 of SAL) and around 10% to removing duplicates in overlapping image areas (**Table 4.4**). Considering a 8-hours working day, the total time is equivalent to 14 working days for one person. Nevertheless, most of the work was done by one machine through the DL model, which devoted around 530 hours to all the required processing, i.e. inference and fine-tuning.

Table 4.4: Detail of the human workload involved in the SADL-OCC detection verification process.

	Number of items		**Allocated time**	
Human task	**First pass**	**Final pass**	**Total (relative share)**	**8h-workday equivalent**
Thumbnails classification	85,779 thumbnails	93,472 thumbnails	24.0 hours (33%)	4.7 days
Full 24MP image examination	3,188 images	529 images	64.3 hours (58%)	8.0 days
Duplicate removal	1,739 images	163 images	9.5 hours (10%)	1.1 days

Assuming a manual interpretation time of a few minutes per 24MP image (Lamprey, Ochanda, et al., 2020), it would take thousands of hours for one person to process the 148,239 transect images. The use of the SADL model thus represents a significant time saving compared to a fully manual interpretation. Furthermore, when comparing the total cumulative counting time of the 3 observers (162h) with the human time invested in the SADL-OCC approach (54h for the photography manager and 111h for DL model's detections verification), the SADL-OCC approach required similar human effort than the traditional SRF approach.

4. Discussion

As DL models are not yet ready for fully automatic use on aerial survey images, we propose a semi-automatic DL approach that tackles the main limitation of the OCC technique: the considerable time required by humans to interpret images. To the best of our knowledge, this is the first time that DL has been integrated into an aerial camera survey at such a large-scale study area in Africa to produce population estimates. Our results showed that the SADL model significantly reduced the human interpretation workload while providing as good or even better population estimates than those obtained from RSOs counts. The SADL-OCC approach seems to be well adapted to count small-sized static species, as revealed by the high estimates for kob and warthog. However, it is difficult to draw conclusions about the larger and more mobile species.

4.1. Population estimates

Unlike previous African OCC studies (Lamprey et al., 2020a, 2020b), we did not observe a systematic significant positive difference between RSO population estimates and those derived from imagery counts for each key species. While the DL model performance could be the first likely explanation, our thorough comparison of 200 RSO and SADL matching counts highlighted that the vegetation was the main cause of the observed differences and that model errors were the least one. The CNP is indeed a vegetated PA and thus differs from arid and semi-arid areas where animals are much more easily captured by oblique cameras, even when running, as they are less prone to tree occlusion. We hypothesize that SADL-OCC approach performance should increase in open areas. Nonetheless, as no additional cameras were used to correct RSO counts for large groups, we may not reject the possibility of biased estimation from RSOs.

The SADL-OCC approach gave lower estimates compared to RSOs in eastern strata, particularly for western hartebeest, buffalo, waterbuck and roan antelope. These differences are mainly explained by the vegetation which occluded the few groups observed by RSOs during the survey. Given the lower animal density in these regions, the observed difference had a significant impact on the overall population estimates.

It should also be added that western hartebeest, buffalo and roan antelope often showed a running reaction to the passing aircraft, making them more detectable to

RSOs due to movement. Whereas an RSO can easily estimate a group of moving animals because he has a continuous view of the scene, the OCC method has a fixed sampled view. This would explain why kob and warthog were better estimated by the SADL-OCC approach, as these species have a much more static behavior. This staticity also complicates the task of RSOs, and increases the risk of missing individuals during flights.

4.2. Method comparison

Our semi-automatic approach went far in addressing the main limitation of the OCC method, i.e. the time-consuming burden of image interpretation (Bröker et al., 2019; Lamprey et al., 2020b), whilst providing at least similar population estimates than the traditional RSO approach. In addition, this combination of DL model and OCC method seems to better detect small-sized and static animals, reducing the variance in counts between transects and therefore tightening the confidence interval of the estimate. Thanks to the cameras and the SAL, what has been counted and identified during the aerial survey was recorded, increasing the validity of the estimates obtained and enabling further checking and potential certification.

The role of human interpreter in the semi-automatic approach is crucial, as he thoroughly verifies the DL model predictions and therefore gives confidence in the final count values. While the use of a DL model reduces traditional aerial survey bias such as animal size and color, group size or density, and observer fatigue (Griffin et al., 2013; Jachmann, 2002; Norton-Griffiths, 1976; Wal et al., 2011), the work of the human interpreter helps to reduce DL model counting bias appearing in heterogeneous scenes where many false alarms may be generated (Delplanque et al., 2023b).

More than 70% of the counting difference between RSOs and the SADL-OCC approach were explained by environmental and acquisition factors, and less than 4% by the DL model performance. These results highlight some shortcomings of our OCC protocol. First, as stated in section 4.1, the fixed and sampled time window of the OCC method makes it impossible to count animals running under sparse canopy, unlike RSOs, which seemed to easily adapt to animal movements through continuous observation. A shorter time interval between image footage might enable better capture of animal movement, and therefore better estimation. Secondly, even though the RSOs have been instructed to count the animals in the strip perpendicular to the line of flight, the front-seat observer's announcement of a group instinctively guided them to look slightly forward to give a better estimate. This, combined with vegetation cover and animal's perturbation due to the overflight aircraft, could explain why we didn't find part or entire groups on concurrent images with the OCC method. While these assumptions were impossible to validate in our study, they could be verified in future research by using additional RSO cameras (CITES-MIKE, 2020; PAEAS, 2014). This will allow refine the counts and reject RSO observations absent from the images. Finally, turbulence during flight at 92 m (300ft) height had obviously an impact on image footage and may explain the proportion of animals counted 'out-of-strip' by the SADL-OCC approach. Nevertheless, we observed that when a strip

marker crossed a large group of animals, RSOs had difficulty estimating the number of animals in and out the strip, often leading to an overestimation of animals counted in the strip. This effect might be exacerbated by turbulence, causing the aircraft to rock and the strip to vary.

Concerning the effect of vegetation, future work should further study the relationship between vegetation cover and animal counting in aerial images. In addition, video recording and analysis might be considered in highly vegetated areas to better capture the movement of groups under spare canopies.

4.3. New insights for aerial surveys

Our study opens up promising perspectives for frequent monitoring and mitigation effort in PAs since OCC survey results may now be obtained rapidly with the use of our semi-automatic approach. Given that the SADL-OCC approach gave similar or even better estimates for small-sized species compared to RSOs, and following previous OCC results (Delplanque et al., 2023b; Lamprey et al., 2020a, 2020b, 2023), we suggest that aerial survey standards are moved forward to embrace new technologies.

We believe that the observer work could be migrated from on-sight count to DL model detection verification, which could considerably reduce associated costs given that during an aerial survey, RSOs are generally mobilized full time for several weeks. Verifying DL model predictions (i.e. points) is an easier task than on-sight counting and does not require highly experienced interpreters who can be easily and rapidly trained. Furthermore, unlike on-sight counting, detection verification may be spread over several people and spaced out over time to avoid any effect of human fatigue on counting results. However, for an autonomous use by PAs, the proposed approach requires a workstation with a good Graphic Processing Unit (GPU) (i.e. at least 8GB of memory), and a dedicated and experienced person in charge of model fine-tuning and inference.

Pending the development of long-endurance Unpiloted Aerial Vehicles (UAVs), our proposed method has a great potential for the use of microlight aircrafts in aerial surveys. Compared to 4-(6-)seat Cessna light aircraft, microlight aircrafts are a much affordable option for PA managers since they are cheaper, they require less expensive fuel and maintenance, and they have less stringent pilot licensing regulations. The main obstacles to their use in wildlife aerial surveys have been their limited capacity of 1 or 2 people and their poor stability at low altitude, making it impossible to apply the traditional method with RSOs. However, our results and those of previous image-based studies (Lamprey et al., 2020b; Lethbridge et al., 2019) showed that observers may be replaced by oblique cameras, since image interpretation burden should now mainly be handled by semi-automated DL models. Thanks to high-resolution cameras (e.g. 36MP), it is then possible to fly higher, which will 1) ensure flight stability and therefore human safety, 2) increase the sampling rate at no extra flying time and costs, thus providing more accurate estimates (Norton-Griffiths, 1978) and 3) mitigate the

effect of running animals thanks to a large image footprint and thus a greater scope for movement. Coupling these cameras with Inertial Measurement Units (IMU) and Global Navigation Satellite Systems (GNSS) would enable image ground projection from which more precise transect area estimates could be derived (Lisein et al., 2013), thus eliminating the need for strip markers.

5. Conclusions

Will AI revolutionize wildlife aerial survey? Our results suggest that we are heading in this direction. Most of our observations regarding the differences observed between RSO and SADL-OCC approaches point to the need to refine the OCC protocol more than improving the semi-automatic approach. While the proposed methodology needs to be validated in other PAs and our OCC protocol further refined, the significant time saving compared to a fully manual image interpretation is a major step towards revolutionizing aerial surveys in Africa.

5

General discussion, perspectives, and conclusion

1. Main findings

In this thesis, I explored the use of remote sensing imagery and DL to address current challenges and research gaps observed in multi-species census of large African mammals in sub-Saharan PAs. I initially evaluated the use of pre-existing CNN-based object detectors, i.e. Faster R-CNN, RetinaNet and Libra R-CNN, to detect and count six species of wild African mammals on drone images acquired mainly in Garamba and Virunga national parks (**Chapter 2**). The Libra R-CNN model outperformed other models published at that time (i.e. Eikelboom et al., 2019; Kellenberger et al., 2018; Rey et al., 2017) in detecting African mammals within similar habitats and landscapes, demonstrating notable differentiation of species in nadir aerial images. It also surpassed the latest multi-species model in most of the performance metrics (i.e. Eikelboom et al., 2019), showing robust performance on an independent dataset. These results suggest that the balancing techniques employed within the Libra R-CNN architecture were particularly interesting for wildlife detection. The model exhibited high precision for major and isolated species but lower precision for herds due probably to bounding box overlap, with herding scenarios representing a significant portion of false positives (40% in our cases). Despite operational implications and challenges in data acquisition, the Libra R-CNN model offered promising perspectives for semi-automatic detection and identification of African mammal species, particularly in open savanna or sparsely wooded areas. Nevertheless, the precision limits observed with dense herds led me to consider other alternatives such as developing CNN-based object detectors based on points to improve counting and overcome challenges posed by highly overlapping boxes in herds, which probably affect the CNN during training. Consequently, I developed *HerdNet*, a point-based CNN better suited to counting animals on aerial imagery (**Chapter 3**). It was directly inspired by approaches developed in crowd counting, a field sharing many similarities with oblique animal counting, and was optimized on challenging herds of free-ranging livestock. I compared HerdNet with two other CNN-based counting approaches: Faster R-CNN (anchor-based), and DLA-34 (density-based). It appeared that previous methods relying on density or bounding boxes prove less suitable for precise animal counting while other authors suggested them as promising approaches (Eikelboom et al., 2019; Padubidri et al., 2021; Peng et al., 2020). The results obtained also further emphasize the inefficiency of anchor-based models for close-by animals already observed in **Chapter 2**. HerdNet not only offers superior detection and counting accuracy but also makes annotation and verification processes easier, which could be valuable in a semi-automated process. Furthermore, HerdNet was trained and evaluated on nadir images used in **Chapter 2**. It surpassed the performance of the first model developed in this thesis, i.e. Libra R-CNN, by drastically reducing counting errors, such as RMSE, which went from 24 to 4 animals per image with HerdNet. However, species identification, particularly for minority species, displayed limitations and therefore, for the time being, humans are still needed to ensure the quality of species recognition and count values by verifying model detections. **Chapter 3** also highlighted the trade-off between recall and precision in developing

tools for PA managers. Since undercounting is one of the major biases of aerial surveys (Caughley, 1974; Grimsdell and Westley, 1981; Jachmann, 2002), the main expectation of automated detection methods is to obtain a model with a high detection rate (i.e. high recall) and few false positives (i.e. high precision). However, recall and precision are often antagonistic: enhancing one aspect typically comes at the expense of the other.

In **Chapters 2 and 3**, the performance of the models was mainly evaluated on images containing animals (i.e. positive images), on small areas and without considering image overlap. By applying a model to a set of images taken continuously over transects, we get closer to the practical use of applying such automated methods. The risk of false positives is however increased with the higher number of confounding elements in the landscape, and so the claimed precision is overestimated. As a result, I have migrated from evaluating the performance of HerdNet on positive images only (i.e. the machine learning way) to obtaining population estimates for an entire PA (**Chapter 4**). Being aware that the current model needs humans for detection verification, we first estimated the human workload reduction using a set of oblique images acquired in Uganda containing less than 10% of positive images and more than 90% of negative images. After training HerdNet to detect 12 wildlife and livestock species on a small set of positive annotated images, it was evaluated using a significant and meaningful set of about 6,000 test images. The results showed a 74% reduction of manual interpretation workload while guiding human interpreters to about 95% of the animals. Identification showed to be difficult for some visually related species, like buffalo and domestic cow, and false positive rate was notable, both emphasizing again the need to keep humans in the whole process for verification. Nevertheless, these results have opened up new avenues for advancing aerial surveys by integrating DL and continuous oblique imaging into aerial survey standards. Consequently, we experimented with a hybrid aerial survey in the Comoé National Park (CNP) in Côte d'Ivoire, i.e. simultaneously using human observers and an automatic photographic approach. To the best of our knowledge, this was the first time that DL has been integrated into an aerial continuous camera survey at such a large-scale African PA to produce population estimates. To do so, I designed a pipeline, integrating a pre-trained version of HerdNet and involving 3 verification steps, that aims to balance efficiency and human effort in 7 key-species population estimation. It represented a drastic time savings compared to a fully manual interpretation of the images. The Jolly II analysis (Jolly, 1969; Norton-Griffiths, 1978) enabled us to compare estimates derived from observer counts with those derived from the model's verified detections. This analysis revealed that using the semi-automated method, estimates were 2.6 to 3.4 times higher for small-sized species (kob, warthog) and comparable for other key species, as well as tighter confidence intervals. Counting differences between the two approaches appeared to be mainly attributed to vegetation obstruction, counts announced by observers that do not appear in the images, and presumed counting errors during the flights.

2. Practical implications of these emerging technologies

Every technological advance aims to address previous challenges, but paradoxically creates new ones. The research in this thesis is based on the integration of remote sensing imagery into African aerial survey, an idea that is not new but has become possible due to the rapid evolution of technology. However, this has also brought with it a number of constraints, mainly in terms of image processing requirements, a point that I have particularly examined following on from the incremental improvement of an existing, relatively effective method (**Figure 5.1a-b**). To put this in context, here is first a brief review of these key incremental steps. For over 50 years now, imagery has been part of the systematic aerial sample count through the use of observers' cameras, photographing large groups for latter counting. These cameras were mainly used for intermittent shooting, even though camera-only systems with continuous imaging were already proposed for replacing observers (Caughley, 1974; Leedy, 1948; Siniff and Skoog, 1964). In those days, observers used film cameras, which required a great deal of care (e.g. changing films, referencing images, developing films) and generated handling errors (Norton-Griffiths, 1978), probably slowing down the idea of continuous-imaging systems. Digital cameras as well as technological and computer advances in the last 20 years have simplified matters, and have led to a whole host of wildlife monitoring applications, first punctually (e.g. Vermeulen et al., 2013) and very recently at large scale with oblique camera systems (e.g. Lamprey et al., 2020b) or VHR satellite systems (e.g. Wu et al., 2023). Although these methods are seen as revolutionary for resetting baselines and better tracking African mammal populations, the time-consuming and tedious processing of images has been a real obstacle and still causes reluctance for integration in aerial survey protocols due to the cost involved. The development of (semi-)automatic image processing approaches was and is still therefore crucial to ensure the sustainability of these new methods.

In response, I showed that the use of DL models alongside continuous UHR aerial imagery has demonstrated superior performance compared to conventional aerial survey methods, constituting a pivotal advancement. Using HerdNet as a tool for sorting thousands of images and pre-detecting animals in them has tackled the main limitation of oblique camera systems: the considerable time required by humans to interpret images. These results are particularly valuable in the realm of cost-effectiveness, as a significant saving on manual labor is in fact a saving which could be allocated, for instance, to increasing the sampling rate or supporting conservation actions. The deployment of these technological approaches not only translates to some financial savings but also contributes significantly to enhanced wildlife monitoring capabilities. Echoing the introductory emphasis on EBVs framework (Brummitt et al., 2017; Jetz et al., 2019), the combination of remote sensing and DL should facilitate wildlife observation and hence biodiversity monitoring through standardized, systematic and comparable records. Additionally, it should fortify the foundation for data validation, quality assessment, and certification processes.

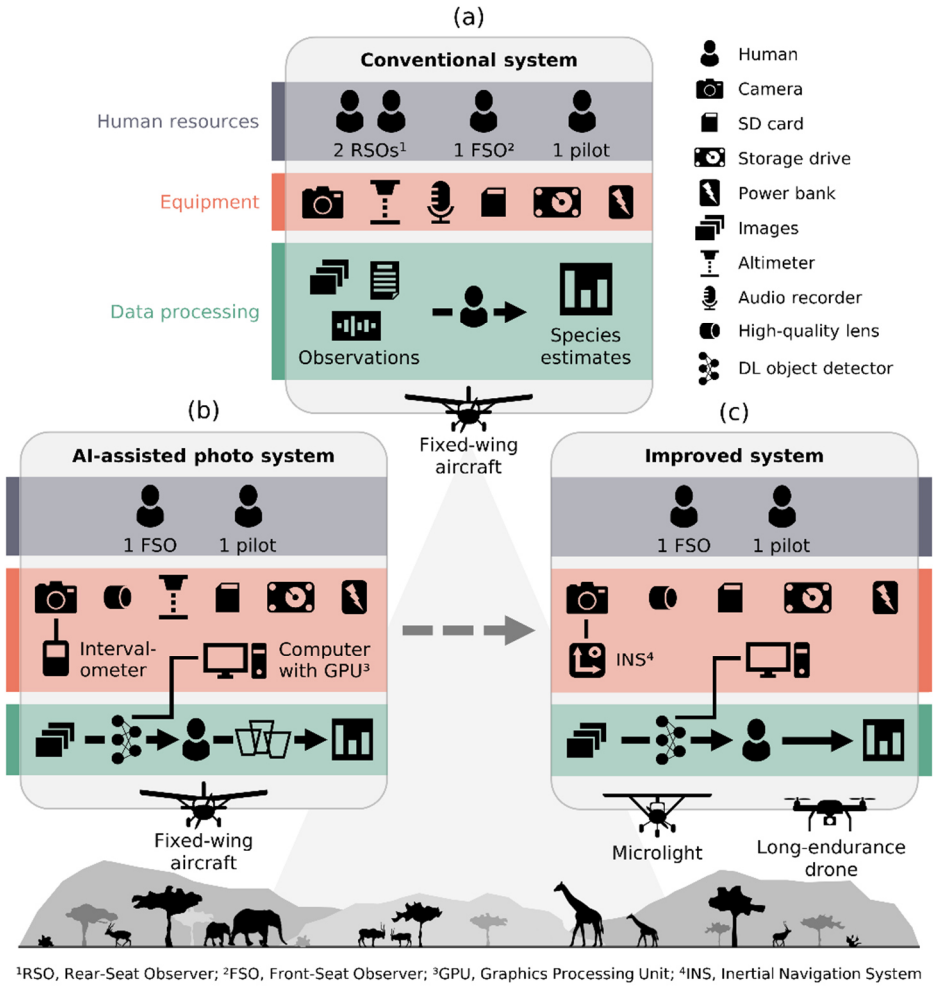


Figure 5.1: Advances in the conventional aerial survey system and human, technical and practical requirements: (a) the conventional observer system; (b) the AI-assisted photo system, developed and tested in the context of the thesis; and (c) an improved system based on the recommendations and perspectives of the thesis.

The most likely source of bias in the conventional approach is the observer, who is bound by several factors influencing the counting performance. Camera systems, in contrast, operate mechanically and are much less influenced by survey and environmental factors, making them better 'eyes' for counting wildlife. The DL model would refer to the 'brain', processing visual information. Although, like the observer, it is subject to omissions, over- or under-estimations, it may be predictable as precise and repeatable patterns can be extracted from its behavior on various datasets, unlike humans, who show sometimes unpredictable behavior and great variability. These trends may serve for correcting errors (Eikelboom et al., 2019). Alternatively,

detections can also be verified by humans, as was done in **Chapter 4.2**. Detection verification, like interpretation in similar studies (Frederick et al., 2003; Lamprey et al., 2020b; Terletzky and Ramsey, 2016; Xue et al., 2017), is a task that may be performed by anyone with a minimum of interest and attention to detail after a short training session. In fact, I would even suggest that this could be the new task of conventional system observers. Their job would thus migrate from counting animals on sight from the aircraft in the conventional system, to verifying the detections of a model in the AI-assisted photo system (**Figure 5.1a-b**). This is even more justified given that counting by sight from the aircraft is a risky task that may additionally cause airsickness to observers.

Despite the advantages just mentioned, it is imperative to acknowledge certain technical implications for practical use by PA managers. The use of remote sensing and machine learning for wildlife applications requires multidisciplinary work and knowledge. While particularly interesting in the field, such profiles are still rare for the moment and are more present in the field of scientific research since the application of these technologies is still in its infancy. It remains a semantic gap between the remote sensing, ecology and computer science communities, which may create collaborative challenges (Kuenzer et al., 2014; Pettorelli et al., 2014), but which should be bridged to train multidisciplinary staff. Remote sensing and DL also require equipment, which may generate some initial costs and regular maintenance. There are two main types of equipment to be distinguished: those used to acquire the images, and those used to process them. In a move towards a more technological approach, the AI-assisted photo system has the advantage of inheriting much of the hardware already used in the conventional system (**Figure 5.1a-b**). However, it still requires the purchase of high-resolution digital cameras (I suggest at least 24MP), lenses of sufficient quality to identify the target species, intervalometers for photo-shooting at regular intervals, power banks, but also several SD cards and storage drives for data back-up (**Figure 5.1b**). Providing a precise absolute cost for all this equipment is difficult as it may vary according to the model of equipment used, but as an example, the total cost of aerial photo equipment for the 2022 survey of the Comoé National Park was around 6200€. For image processing equipment, high-capacity computers with large-memory Graphics Processing Unit (GPU) are required, particularly for training DL models. From experience, I suggest a GPU with a minimum memory of 8GB. High-performance computing clusters are also available via cloud platforms such as *Google Compute Engine* or *Amazon Web Services*, with prices that may be adapted to computing needs. However, as its name suggests, cloud computing requires a good internet connection, which is not always the case in some African PAs. Of course, a model might be trained on such platforms upstream by partners or collaborators, but I would still suggest having a machine with a GPU to perform local computing or inference. It is an investment of around 2,000€ that would avoid having to transfer hundreds of thousands of images elsewhere for processing. However, as mentioned above, this requires qualified personnel to maintain both the computer and the consequent database produced.

3. Remaining challenges for automated aerial counting

Although current DL models show very interesting results for the detection and counting of animals on UHR aerial images, there are still several challenges to be overcome and aspects to be improved to increase the automation of the whole workflow, i.e. from image acquisition to population estimates delivery. In this section, I develop what I consider to be the main elements to work on in the years to come.

3.1. False positive reduction

Reducing the number of false positives is perhaps the greatest challenge in wildlife counting using remote sensing imagery. Unlike most computer vision applications, the imbalance between the pixel area covered by animals (i.e. the foreground) and their environment (i.e. the background) is extremely important (Kellenberger et al., 2018). Due to its natural heterogeneity, particularly strong in African savanna, this environment is often riddled with confusing animal-like objects, like tree trunks, rocks or termite mounds, increasing significantly the number of false positives generated by a DL model when applied on large areas, as systematically observed across **Chapters 2 to 4**. In response, *hard negative mining* techniques have already been employed to reduce false positives and shown to provide improved model performance (Kellenberger et al., 2018; Peng et al., 2020). I also developed a similar approach in **Chapter 3**. Although this minimized the counting error at the image scale, the number of false positives remains high at the scale of transects, as observed in **Chapter 4.1**. Future research should continue experimentation in this area, as it is considered essential to lighten the task of downstream human verification. Nevertheless, this task was considerably reduced following a manual classification of thumbnails centered on each model detection during our study of the Comoé 2022 aerial survey (**Chapter 4.2**). This first step served as a guide for the interpreter, directing his attention to the most relevant detections while discarding numbers of false positives, thus reducing the number of full-size images to be verified subsequently in the second step. However, this first thumbnail classification step takes a considerable amount of time (33% of total human verification time in our study). It is therefore not surprising that previous studies explored the use of an image classification model, trained to distinguish image patches containing animals from background-only ones (Guirado et al., 2019; Rahnemoonfar et al., 2019). The main risk with this type of two-step automated approach is that errors in the first step (mainly false negatives) cannot be rectified by the second step. Although this method seems interesting, my advice would be to focus on improving a single object detection model rather than combining several models to avoid error propagation.

3.2. Species identification

While the case of multi-species detection faces practical needs, most studies have focused on a single species, or have not distinguished between species and grouped

them into a single category (e.g. 'animal', or 'mammal') (Xu et al., 2024). Accurate species identification is an inherent challenge for DL models, caused by the naturally unbalanced distribution of species, consequently creating a significant imbalance in training datasets. As a result, species that are predominantly present in PAs are well represented and recognized by DL models, whereas species that are found in smaller numbers are under-represented and therefore poorly recognized by DL models. Not to mention the consequently small test sample sizes for these species that reduce the statistical credibility of the model performance. The main problem with this imbalance issue is that it is usually the less abundant species that are targeted by conservation agencies, as they are often endangered or on the increase. In **Chapter 3**, HerdNet showed slightly poorer species identification performance for livestock than the baseline (i.e. Faster R-CNN), suggesting mandatory human verification to correct this and avoid a too high level of confusion. This was also confirmed for the identification of wildlife species in **Chapter 4.1**, where strong confusion was observed between look-alike species (e.g. buffalo and cow). One hypothesis, apparently shared with other authors (e.g. Xu et al., 2024), is that the relatively small pixel size of animals on aerial images creates features that are difficult to distinguish between species. This was illustrated and discussed in **Chapter 3**, where we observed that identification performance degraded with distance from the aircraft, i.e. in areas where animals are represented by fewer pixels. Future research should focus on the optimization and design of feature extractors to better identify small objects that are at least recognizable by humans. Nevertheless, while at first we might naively have thought that if a human is able to properly discern species on RGB aerial images, then a DL model should be able to as well; the results of this thesis show that this is not necessarily the case even on images with GSD smaller than 5 cm (see **Chapter 2**). Humans seem to be better at identifying species, even if they can be prone to fatigue or mistakes in challenging conditions (e.g. occlusion). Based on my own experience and probably that of others accustomed to the task of annotation and identification, it appears that our eye takes several elements into account when making its decision, such as the surrounding elements of the potential animal, the shape and behavior of the herd, other images overlapping the target area, the proximity of other species, or the spatial location of the image in the PA. All these features might be taken into account by a DL model (see Section 5.1), especially with the advent of *foundational models* (FMs, see Section 5.5), which should be experimented with to improve current identification performance.

3.3. Image overlap

Due to overlapping coverage of images generated by continuous camera systems, the same animal may be present in multiple images, which may lead to an overestimation of the true number of individuals if not managed properly. It is therefore necessary to carefully manage overlap to avoid double counting. In their study, Lamprey et al. (2020b) chose to count all the animals in the even-number images, and count only animals appearing in the non-overlap portion in the odd-number images. During the 2022 Comoé aerial survey (**Chapter 4.2**), I manually

checked each batch of consecutive images for the same individuals present in several images. This took around 10% of the total human verification time, representing about 10 working hours. I personally used a custom template on *Label Studio* (Tkachenko et al., 2020) for this task, without outlining overlapping areas, but recent tools like *Scout* (*WildMe*) proposed such a feature. Automatic approaches could perhaps be explored for managing double counts due to image overlap (**Figure 5.1c**), as the time investment could prove much greater for surveys in PAs with high animal density. A simple idea would be to use basic computer vision techniques, such as *feature detection and matching*, to outline areas of image overlap and potentially layer the detections of the relevant images. Briefly, this method independently detects features and assigns them a descriptor in the images to be compared and then searches for likely matching candidates (Szeliski, 2022). While this might be a solution for static species (e.g. Corcoran et al., 2019), other methods based on image projection and bipartite graphs may be used for moderately moving animals (Shao et al., 2020; Soares et al., 2021). Unfortunately, such methods have been mainly tested on vertical imagery and thus could be of little interest for oblique imagery containing species that are running during the passage of an aircraft. This was a common occurrence, for example, with hartebeest and buffalo during the 2022 Comoé aerial survey. Some animals were found in more than two images because they were running parallel to the aircraft. Such behavior makes the problem even more complex. Although I believe that this is not an absolute priority at the moment, future research should focus on the development of suitable automatic approaches for this task.

3.4. Transect area estimates

In the conventional method, the area of the transects is estimated following a prior calibration of the strip width, theoretically delimited by the streamers as a function of the flying height, which is then multiplied by the length of the transect (CITES-MIKE, 2020; Frederick, 2012; Norton-Griffiths, 1978; PAEAS, 2014). Calibration is undoubtedly one of the most crucial aspects of an aerial survey, but it can unfortunately be subject to errors that have a major impact on survey results, such as a change in observer or streamer position after calibration (Frederick, 2012). During conventional calibration, markers are placed on the ground at regular distances (e.g. 20 m) and the aircraft makes several passes at different altitudes (recorded by an a radar or laser altimeter) during which the observers count the number of markers visible between the streamers. To avoid the use of altimeters, which can lead to errors and require frequent checking (Frederick, 2012), flight height can be estimated using the GPS-DEM method (Lamprey et al., 2020b) thanks to recent developments in GPS technology and Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM). This method measures the flying height as the difference in elevation above mean-sea-level between the aircraft navigation GPS and the terrain below. For oblique camera systems, recent studies have performed calibration by randomly selecting geo-referenced transect images and superimposing them on very-high resolution satellite images in *Google Earth* to measure the strip width or image footprint (Lamprey et al., 2020a, 2020b). However, whether using the traditional method or the oblique camera

system, turbulence may momentarily cause wing tilt when flying over transects, resulting in a considerable increase in the transect area (Lamprey et al., 2020b; Pennycuick and Western, 1972). In the traditional observer approach, this effect is unfortunately not taken into account. For the oblique camera system, a correction factor for tilt can be derived from a random analysis of several batches of consecutive images, again using Google Earth. Experiments of Lamprey et al. (2020b) showed, for example, an increase of around 7.4% on the estimated sample area based on conventional calibration results. Such an increase is bound to have an impact on the Jolly II method, and consequently on the population estimates derived from it. It thus suggests a call for the development of more accurate methods for estimating transect area. Solutions may be found by drawing ideas from existing methods developed for vertical aerial imagery (e.g. Lisein et al., 2013; Soares et al., 2021) that use photogrammetry elements and Geographic Information Systems (GIS). Such methods often require image geo-referencing and ground projection through the use of GNSS sensor and Inertial Measurement Units (IMU) (Verykokou and Ioannidis, 2018; Wolf et al., 2014), combined to become an *Inertial Navigation System* (INS) (**Figure 5.1c**), which should be increasingly precise and affordable.

4. Monitoring African mammals from space: reality or fantasy?

This thesis focused exclusively on aerial platforms and passive RGB acquisition systems, but the use of remote sensing for wildlife monitoring could be theoretically extended to space platforms coupled with other systems and targeted spectra. However, is current satellite imagery already suitable for surveying or monitoring large African mammals? This is the key question for upscaling the methods developed in this thesis and imagining the prospects for large-scale applications. This question, but extended to all wildlife, has been the subject of a review article published in *GIScience & Remote Sensing* (IF=6.7), available in the **Annex** to this thesis, on which the following text is based and derived. Examining 49 peer-reviewed papers, the analysis reveals trends in publications, targeted species, studied biomes, sensors, resolutions, and methods used for detection, counting and surveying. However, I believe that the practical application of satellite imagery for large African mammal survey is limited and not a realistic alternative for the moment. This is due to various constraining factors associated with the use of current satellites (**Figure 5.2**).

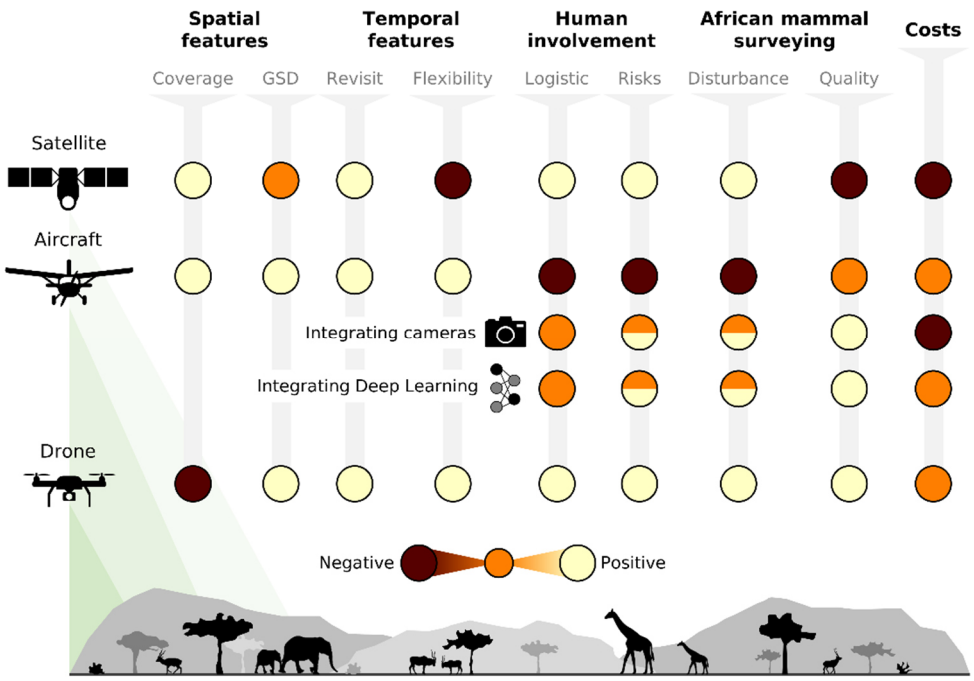


Figure 5.2: Comparison of the technical, practical and economic considerations of the 3 platforms mainly discussed in the thesis. Note that a two-color cell means a potential transition to a more positive state, depending on implementation details (e.g. higher flight height).

Firstly, although the GSDs of today's VHR satellites are attractive (30-50 cm/pixel) because they can apparently detect some key conservation species (e.g. African elephant, *Loxodonta africana*) (Duporge et al., 2021), differentiation between species is difficult, if not impossible (Wu et al., 2023; Yang et al., 2014). In addition, as one animal is covered by just a few numbers of pixels, differentiation between animals and environmental features is trickier than with aerial imagery. This is further accentuated by the complexity and the number of confusing landscape features in African savannas (Irvine et al., 2019; Xue et al., 2017). Based on recent personal technical experience, it is sometimes very difficult to decide on certain manually detected features. In fact, it is not always possible to reject the hypothesis of an animal for certain objects if we do not have a reference image of the same area taken at a near date. Regardless of the financial aspects, I feel that this is the most constraining aspect when it comes to considering the practical use of satellite imagery for wildlife monitoring.

Secondly, the large spatial and temporal coverage of satellite images means that large areas may be covered rapidly and frequently, offering possibilities for total count. Moreover, this can be of particular interest in reducing the double-counting

bias of the conventional aerial counting method, as well as potentially increasing both the precision and accuracy of population estimates. However, the unpredictable availability and current cost of VHR satellite imagery severely limit these prospects (**Figure 5.2**). On the one hand, the weather can easily render an image unusable, for example in overcast conditions. Furthermore, it is not possible to set the time of day at which the image is acquired, which may not correspond to the time and conditions when mammals are most visible. On the other hand, the commercial system on which these VHR satellites are based operates on priority and license principles, which can prevent an image from being acquired at a given time and place, and its sharing to third parties. Combined with the fact that the higher the image's spatial resolution, the smaller its swath width, this means that covering a PA of several thousand square kilometers requires a greater number of satellite passes or acquisition requests. This has repercussions not only on the complexity of image mosaicking, but also on the costs involved. Pricing varies according to several factors such as the image provider, type of demand (archive or tasking), image resolution and spectral characteristics, level of processing, coverage area, licensing terms, and intended use. It is therefore difficult to give a precise estimate, as they are also often linked to requests for quotations. Prices may range from several hundreds to thousands of euros per minimum-size ordered scene (25 to 100 km²), which may result in prohibitive costs for covering large areas (Apollo Mapping, 2023; LAND INFO Worldwide Mapping, LLC, 2023). Nevertheless, we may look forward to positive developments in this aspect, as several VHR satellite image providers (e.g. Airbus, Maxar) are offering access opportunities at lower or no cost.

Finally, image processing is an aspect that should not be overlooked and which can also generate significant additional costs, depending on the degree of human investment in the process. Although, as presented in the **Annex** and in this thesis, automatic methods are emerging, they do not yet make it possible to fully automate image analysis. Human intervention remains essential to first develop or train automated methods, and then to validate their predictions and obtain reliable count data. As stated in **Section 2** of this chapter, the development of automated approaches requires expertise at the crossroads of different fields, mainly ecology, computer science and GIS. Unfortunately, the licensing scheme of satellite image providers does not allow data to be shared freely, thus limiting scientific research, community collaborations and the development of robust models trained on a large volume of images. To ensure the practical application of these tools by PA managers, this multidisciplinary knowledge is however crucial and needs to be passed on to local operators. This is essential to ensure sustainable use of these new tools, and to avoid dependence on external researchers and developers.

In conclusion to this section, while recent advancements in image-based ecological monitoring have propelled the potential of satellite ahead of other methods, current limitations, including the still-too-low resolution and the inability to identify species, underscore the method's reliance on ground-truthing data for validation. The efficacy is notably high in open and homogeneous environments but falters in more complex

ecosystems with numbers of confusing elements, such as in African savannas. Emphasizing the indispensability of field data at present, I would argue that satellite images may serve as valuable complements to organize and deploy other platform data acquisition (e.g. aircraft) rather than standalone solutions (see **Section 5.3**). While the general idea hold promise, the existing constraints hinder its emergence as a viable alternative, necessitating cost reduction, continued research, and technological refinement to unlock its full potential in ecological monitoring. In the interests of positive change towards operational approaches, we have identified and presented several research priorities and recommendations, which are extensively detailed in the **Annex**: (1) Establishing wildlife-dedicated VHR satellite constellations providing freely available imagery at high spatial and temporal resolutions; (2) Developing sampling methods in tandem with advancements in remote sensing and image processing; (3) Creating foundational DL models for processing diverse wildlife monitoring data; (4) Strengthening initiatives for sharing and collaborative annotation platforms; and (5) Promoting efforts to enhance events, training, publications, and funding programs that merge interdisciplinary approaches.

5. Perspectives

In this final section, I put forward some perspectives related to the findings of this thesis. Five topics are particularly covered: 1) the potential enhancements to HerdNet; 2) the upgrade of aerial survey guidelines and standards; 3) the use of microlight aircraft as an alternative to light aircraft; 4) the combination of the different remote sensing platforms considered in the thesis; and 5) the development of foundational DL models.

5.1. Enhancing HerdNet

Although HerdNet has shown promising counting and detection performances to cut down on the time-consuming interpretation of aerial photo survey images (**Chapter 3-4**), there is still room for improvement and further validation. A first aspect that could undoubtedly be improved is species identification. In **Chapter 3**, I showed that this was the major weakness of HerdNet. There are several tracks of improvement that I believe could help to better identify species. From an 'architecture' perspective and its training, different levels can be explored, such as the addition of extra convolution layers for the classification head, the testing of other cost functions, performing additional balancing, adaptive fine-tuning or *incremental learning* (Belouadah et al., 2021; Oksuz et al., 2021; Wu et al., 2019). Other avenues of research would be to look at *few-shot learning* techniques where models are able to be trained efficiently on little or no training data (Xin et al., 2024). This may be obtained with the help of synthetic data for instance (Nikolenko, 2021), which showed promising results for rare wildlife species recognition in camera trap images (e.g. Beery et al., 2020) or for rare object detection in satellite imagery (e.g. Martinson et al., 2021). In addition, it would be very worthwhile to evaluate the combination of images with other species-related variables as inputs, such as herd structure, habitat, vegetation type or species

relationships. This additional information should certainly improve the accuracy of model identification.

Easy and rapid adaptation of the model to another landscape is also an essential aspect to address for its practical use. From experience, I noticed that when I applied a version of the model trained on one PA (e.g. QENP, Uganda) to another PA (e.g. Comoé, Côte d'Ivoire), recall and precision systematically dropped, even though some species were present in both areas. This is due to a number of technical and natural factors, such as the camera model, its angle of inclination, a change in landscape or species, a different GSD, or a change in contrast and brightness. These factors result in a change in data distribution and/or feature spaces between the source dataset, i.e. the one on which the model has been trained, and the target dataset, the application case. The discrepancy between the input data distributions of the two datasets or the difference in their feature spaces can lead to a degradation in model performance. On the one hand, if only a limited dataset is available, solutions exist to minimize the gap between source and target distributions, like *transfer learning*. The latter aims to transfer the knowledge learned to perform a task in the annotated source dataset, in order to perform the same or another task in the unannotated target dataset (Pan and Yang, 2010). Another complementary perspective would be to train a backbone in a *self-supervised* way on a large set of heterogeneous images, leveraging both images with and without animals. *Self-supervised learning* is a type of machine learning where the model learns to predict part of its input from other parts, using the input data itself as a form of supervision (Jing and Tian, 2021). This method leverages large amounts of unlabeled data to generate useful features or representations without the need for extensive manual labeling. Such an approach provides pre-trained models that could be versatile bases for subsequent fine-tuning steps. On the other hand, if we have a large heterogeneous annotated dataset, i.e. with many different landscapes, animal species and instances, it is theoretically possible to train a model on the whole to learn how to generalize correctly. Such a model would have a much larger feature space and a much wider data distribution, enabling it to be used more widely, thus avoiding the need to systematically revert to a from-scratch annotation phase when new aerial images are acquired. As an example, I recently had the opportunity to test HerdNet on independent datasets acquired in another PAs (i.e. Zakouma National Park, Chad and Murchison Falls National Park, Uganda), following a prior training on all the aerial African mammal datasets available in our lab. Although the acquisition conditions were somehow similar, the model showed early signs of generalization. Qualitatively, the model was able to detect species absent from the training set, such as lions, ostriches, and various bird species. Quantitatively, we estimated an overall recall of over 70% and a positive image detection rate (i.e. with animals) of over 95%. These first results suggest the idealistic idea of a general high-performance model trained on numerous PAs and on which we could be much more confident about its predictions. While these tests are a good starting point for validating the approach proposed in this thesis, they should be multiplied to highlight its weaknesses and strengths. I still think that human intervention and verification

remain essential, whether to correct the identification and counts of new species, to avoid double counting or to spot partially occluded individuals in substantial vegetated areas.

Finally, **Chapter 4.2** highlighted the counting limitations of the AI-assisted photo system in vegetated areas, compared to the observer who tend to better estimate the true number of animals as they run. This hypothesis had been put forward but could not be fully confirmed, as we could not rule out the possibility that certain flaws in the acquisition protocol may have limited the possibility of animal detection in these areas. The results of the 2024 aerial survey of Zakouma National Park (Chad), during which we carried out a hybrid survey in collaboration with African Parks, highlighted this same observation. Despite the continuous imaging of the photo system and the images taken by observers, the comparison of photo counts and those announced by observers differed inconsistently in heavily vegetated areas. In some cases, even no animals were visible in the images, probably because they are moving too fast and hiding directly under the trees (E. Bussière, personal communication, July 24, 2024). To tackle this challenge, we suggested the alternative of video recording in **Chapter 4.2**. However, such an approach brings with it a series of logistical and technical challenges. The resolution of the images composing the videos is generally lower, and the cameras are more expensive. In addition, the volume of data generated is significant, and management could be trickier when considering survey flights of 3-5h in general. For example, a 4K camera (4,096 x 2,160 pixels at maximum resolution) can generate between 30 and 50 GB per hour, compared with around 12 GB/hour for images of equivalent resolution taken at 1-second intervals. Although video recording can be interesting for detecting animals more easily during an automated or a manual analysis, in my opinion it is of little use in open areas and therefore generates worthless data for most of a savanna survey. Without considering costs as a constraint, an intermediate solution could be to take punctual video recordings in heavily vegetated areas. However, if we are heading for uncrewed acquisition platforms, we need to think about an automatic shooting mode when flying over highly vegetated area within the transects. I think this should be feasible if a vegetation map is available and the acquisition platform is continuously geo-referenced. In this way, a signal could be sent to the camera when the platform approaches and leaves a densely vegetated area.

5.2. Advancing aerial survey guidelines and standards

Given the increased accuracy and precision of oblique camera systems compared with the traditional approach (Lamprey et al., 2020a, 2020b, 2023), and the positive results of DL in significantly reducing human interpretation of images (**Chapter 4**), aerial survey standards and guidelines should move forward to include these new technologies. The transition to automatic approaches should nevertheless be gradual, in my opinion, and remain an alternative for the time being given the expertise and practical implications this entails (**Section 2** of this chapter). The guidelines and standards documents (e.g. CITES-MIKE, 2020) are indeed true references for

practitioners, and are the result of decades of continual refinement and improvement. They are really important because they guarantee the quality of the results of an aerial survey. I think it's important to add at least basic protocols for installing oblique camera systems for continuous image acquisition, given the growing interest in these systems in recent years. In my opinion, this will have two beneficial consequences: 1) the spread of the method's use, leading to the creation of multiple databases that are crucial for the development of large-scale and robust DL models; and 2) the continuous improvement of protocols and the promotion of exchanges between organizations, which might lead to an accepted and standard protocol. Nevertheless, manual interpretation is seen as the biggest challenge, as it is very labor-intensive and therefore very costly at present. As the results of this thesis show, automatic DL models like HerdNet are certainly the solution to this problem but including them in the guidelines may be premature, as no free, effective and easy-to-use tool is currently available. Furthermore, AI is a rapidly evolving field, with new techniques or CNN architecture coming out every week, if not every day. I believe this is an aspect that should not be overlooked. I feel that what should be included in the guidelines are more general considerations, such as techniques for verifying and correcting the predictions of an automatic approach, whatever its architecture.

5.3. Microlight aircrafts

The combination of oblique imaging and DL has great potential for the use of long-endurance drones and microlight aircrafts (or simply microlights, also called ultralight aircrafts in some countries), eliminating or reducing the risks for the flying crew in the conventional protocol (**Figure 5.1c**). The current main disadvantage of drones for wildlife survey applications is its low endurance and therefore poor spatial coverage compared to the use of light aircraft (Linchant, 2021). Future technical advances will undoubtedly improve their endurance, such as continuous solar power supply (Hassanalian and Abdelkefi, 2017; Rajabi et al., 2021), and will offer a real alternative to light aircraft. Pending this, microlights appear to be a best candidate, combining lower cost and efficiency (Linchant, 2021). Light aircraft are very expensive, and are subject to a very strict legal framework, such as extensive pilot training, technical certification of every single component, and frequent aircraft maintenance. All this makes up a large part of an aerial survey budget (Bouché et al., 2012; Grimsdell and Westley, 1981). Furthermore, the aircraft must be operated at an airspeed no slower than about 40% to 60% above the stall speed to remain safe (Lamprey et al., 2020b). At a common survey altitude of 300 feet, ground speed of at least 160 kmph is thus recommended (CITES-MIKE, 2020). Microlights are by definition aircraft with a maximum take-off mass of 450 kg and a stall speed of 65 kmph for a two-seat landplane (Civil Aviation Authority, 2024). They are less regulated than light aircraft, offer better maneuverability and consume motor gasoline (MOGAS), some 20% cheaper than aviation gasoline (AVGAS) used by light aircraft (AirNav, 2021). Microlights also offer better flexibility in PAs as they can take off and land outside conventional aerodromes, as well as carrying sensors outside the engine (Latte et al., 2020). They are in fact already used in PAs but not for traditional aerial surveys, since

their limited capacity of 1 or 2 people is unadapted to the observer method. However, the results presented in **Chapter 4** offer good prospects for transposing the oblique camera system to this platform. The 'big data' limitation previously highlighted (Linchant, 2021), regarding the development of this platform as an alternative to the drone or light aircraft, is now lifted thanks to the use of DL models. However, as the approach is semi-automatic, this would not jeopardize the work of the observers. In my opinion, the use of such models will inevitably still involve human verification to ensure optimal quality of survey results, at least for the next few years. The previous work of observers, i.e. risky and hasty sight-counting, would therefore shift to risk-free and thorough laboratory work, i.e. model detection verification.

5.4. Combining remote sensing platforms

As discussed in **Section 4**, the use of VHR satellites is not currently a viable solution for large mammal census in heterogeneous environments, as is the case for most PAs in sub-Saharan Africa. The use of light aircraft is very popular and effective, particularly with the addition of oblique camera systems (**Chapter 4**), but its logistics, associated risks, regulations and costs have led to the emergence of alternatives such as microlights and drones. The drone is indeed interesting, especially when combined with the DL to manage the large quantity of images generated (**Chapter 2-3**), but its low endurance and efficiency restrict it to limited sampling for the time being. However, I believe that the strengths of these three platforms could be combined to develop multi-level monitoring for more effective active conservation strategies. For example, a time series of satellite images could be collected upstream at points of interest in a PA to study the seasonality of gregarious or migratory species at macro-level (e.g. Wu et al., 2023). This would make it easier to plan and set up drone or microlight acquisition missions downstream to first obtain precise and verifiable estimates at meso- and micro-level, and then to carry out appropriate conservation actions. This multi-level combination also has prospects for use in monitoring the co-existence of wildlife and livestock in PAs, which may be a source of territorial, health and economic conflict (Butt and Turner, 2012; Georgiadis et al., 2007; Herrero et al., 2013; Scholte et al., 2022a, 2022b). For example, Vrieling et al. (2022) have recently shown that it is possible to track the spatio-temporal patterns of livestock night-time enclosures at macro-level using satellite imagery at 3 m/pixel resolution. Although improvements are needed, Wilson et al. (2022) have also developed a DL model for segmenting cattle camps in South Sudan based on 10 m/pixel resolution satellite imagery. Downstream, such tracking and mapping tools could be used to validate the movements of transhumant livestock herds at meso-level, using reconnaissance flights by light or microlight aircrafts. This would enable a better understanding of the spatio-temporal patterns of livestock, and thus support initiatives for better environmental protection or human-livestock-wildlife cohabitation. While these examples focused directly on animals, other applications, complementary to censuses, may also be envisaged, such as using satellites to create vegetation maps, which might be particularly useful for designing survey strata or targeting areas of ecological interest for certain species. The multi-level spatial and temporal combination of these

platforms therefore opens up a whole new field of perspectives that promises to improve the adaptive management process of biodiversity.

5.5. Foundational deep learning models

Wildlife monitoring is inherently a multimodal task, with a wide range of distal data sources, like remote sensing platforms acquiring imagery, as well as proximal ones, such as camera trap, GNSS collar, or in-situ microphones. Large-scale data obtained from these various modalities might be centralize in a so-called *Foundational Model* (or *Foundation Model*, FM), which once trained, may operate as a basis and can be adapted to a wide range of downstream tasks (e.g. animal detection in aerial image or species recognition from audio record) with zero or very few training samples (Awais et al., 2023; Bommasani et al., 2022). This would enable a model to perform all the functions expected of an observer in addition to counting and identifying animals, such as detecting illegal activity in a PA. The basic component of FMs are deep neural networks, transfer learning and self-supervised learning techniques (Bommasani et al., 2022), which have been particularly useful when data and model size have been massively scaled-up with recent *large language models* (Zhao et al., 2023), like GPT-3 (Brown et al., 2020). Following this, research has explored FMs whose visual inputs can be provided, such as SAM for segmenting any type of object (Kirillov et al., 2023), but also other FMs, like FLAVA (Singh et al., 2022), ImageBind (Girdhar et al., 2023) or AudioCLIP (Guzhov et al., 2022) that align multiple data modalities to learn meaningful representations useful for different downstream tasks. FMs are seen as the future for all automation tasks concerning remote sensing, like image interpretation. In fact, such models have already been developed and show good prospects for remote sensing applications, such as RSPrompter (Chen et al., 2024) or RingMo (Sun et al., 2023). For wildlife monitoring, FMs are also beginning to emerge, such as KI-CLIP (Mou et al., 2023a) for the rapid identification of wildlife on proximal images, or CLAP for facilitating bioacoustic tasks (Miao et al., 2023). Although these models emphasize the potential progress for wildlife monitoring, they contain billions of parameters that need to be trained on huge databases before being fully versatile on various tasks. Unfortunately, aerial survey images are still sparse and not widely shared. Given that FMs require the cross-referencing of massive amounts of data, the scarcity of aerial images might be the limiting factor. I could only suggest even more data sharing and community collaboration to grow databases of wildlife remote sensing imagery. Despite all this, I think it is important to be aware that these models require massive computing clusters, running continuously for days or weeks to be trained, and then receiving numerous requests every day. The carbon footprint of these supercomputers is not to be overlooked (Allen, 2022). If the use of these tools impacts the environment more than they help to protect it, is it really essential to develop and use them?

6. Conclusion

Wildlife monitoring and census applications using remote sensing have recently emerged with the rapid technological advances including digital imagery and computer processing. The massive production of associated data has however instantly revealed a major bottleneck: the tedious and time-consuming aspects of manual image interpretation. The general objective of this thesis was therefore to evaluate the combined use of remote sensing and DL models for large African mammal multi-species census applications.

In addressing the central research question of whether the combination of aerial imagery and DL models enhances the accuracy and precision of population estimates for large mammals in sub-Saharan PAs, the research conducted in this thesis presents an early response. While **Chapters 2 and 3** primarily evaluated model performance on positive images, in **Chapter 4**, the focus shifted towards obtaining population estimates for entire PAs, incorporating continuous transect images. Despite the need for human verification in species identification and false positive reduction, the study demonstrated a substantial reduction in manual interpretation workload using HerdNet, paving the way for upgrading aerial surveys. The integration of DL into a large-scale aerial camera survey marked a groundbreaking approach, achieving remarkable time savings compared to fully manual interpretation. The Jolly II analysis further validated the semi-automated method, revealing increased estimates for small-sized species and comparable results for other key species, with tighter confidence intervals. This thus reflects a significant step towards advancing aerial surveys and obtaining more accurate and precise population estimates for diverse wildlife species. Nevertheless, although the results presented are supportive and bring good prospects for cheaper platforms like microlights, one study is not enough to fully answer this question. I have presented here the results of a DL model in a specific environment with a certain diversity and density of species. We have observed, for example, that vegetation is a sensitive factor for the proposed approach. Further studies in other landscapes, with other species densities and diversity, should be carried out to target reliable patterns in the use of DL models for potential integration into aerial survey guidelines and standards. I also detailed and gave some perspectives for four main remaining challenges to further improve and automate the workflow: 1) further reducing false positives; 2) improving species identification; 3) managing image overlap; and 4) better estimating transect areas. False positives persist due to environmental complexities, necessitating further research into hard negative mining techniques. Additionally, accurately identifying species remains a challenge, especially for less common species, prompting exploration into optimizing feature extractors and employing few-shot learning techniques. Managing image overlap and estimating transect areas are little studied and may be achieve using computer vision techniques, orientation and position sensors and calibration methods.

In a world where biodiversity is under threat, advancing of conventional methods is crucial to obtain a more standardized and frequent observation system. Such a system

will certainly lead to better monitoring of biodiversity indicators and, ultimately, more targeted, and effective conservation actions. However, the journey towards fully realizing the potential of remote sensing and DL models in census applications is far from over. While this thesis contributes significant insights and advancements for large mammals in sub-Saharan PAs, I hope it will also serve as a catalyst for further research and innovation, driving towards more efficient, accurate, and comprehensive animal census methodologies.

References

AirNav, 2021. Fuel Price Report [WWW Document]. URL https://airnav.com/fuel/report.html (accessed 3.15.24).

Allen, M., 2022. The huge carbon footprint of large-scale computing. Phys. World 35, 46. https://doi.org/10.1088/2058-7058/35/03/32

Anderson, K., Gaston, K.J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Frontiers in Ecology and the Environment 11, 138–146. https://doi.org/10.1890/120150

Apollo Mapping, 2023. Download Imagery & DEM Price Lists - Apollo Mapping | The Image Hunters. URL https://apollomapping.com/download-imagery-dem-price-lists (accessed 4.15.23).

Asner, G.P., 2001. Cloud cover in Landsat observations of the Brazilian Amazon. International Journal of Remote Sensing 22, 3855–3862. https://doi.org/10.1080/01431160010006926

Awais, M., Naseer, M., Khan, S., Anwer, R.M., Cholakkal, H., Shah, M., Yang, M.-H., Khan, F.S., 2023. Foundational Models Defining a New Era in Vision: A Survey and Outlook. https://doi.org/10.48550/arXiv.2307.13721

Ayantunde, A.A., Duncan, A.J., van Wijk, M.T., Thorne, P., 2018. Review: Role of herbivores in sustainable agriculture in Sub-Saharan Africa. Animal 12, s199–s209. https://doi.org/10.1017/S175173111800174X

Barbedo, J.G.A., Koenigkan, L.V., Santos, P.M., 2020. Cattle Detection Using Oblique UAV Images. Drones 4, 75. https://doi.org/10.3390/drones4040075

Barbedo, J.G.A., Koenigkan, L.V., Santos, T.T., Santos, P.M., 2019. A Study on the Detection of Cattle in UAV Images Using Deep Learning. Sensors 19, 5436. https://doi.org/10.3390/s19245436

Barber-Meyer, S., Kooyman, G., Ponganis, P., 2007. Estimating the relative abundance of Emperor Penguins at inaccessible colonies using satellite imagery. Polar Biology 30, 1565–1570. https://doi.org/10.1007/s00300-007-0317-8

Barnes, R.F.W., Beardsley, K., Michelmore, F., Barnes, K.L., Alers, M.P.T., Blom, A., 1997. Estimating Forest Elephant Numbers with Dung Counts and a Geographic Information System. The Journal of Wildlife Management 61, 1384–1393. https://doi.org/10.2307/3802142

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Features, in: Leonardis, A., Bischof, H., Pinz, A. (Eds.), Computer Vision – ECCV 2006, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 404–417. https://doi.org/10.1007/11744023_32

Beasom, S.L., Hood, J.C., Cain, J.R., 1981. The Effect of Strip Width on Helicopter Censusing of Deer. Journal of Range Management 34, 36–37. https://doi.org/10.2307/3898449

Beery, S., Liu, Y., Morris, D., Piavis, J., Kapoor, A., Joshi, N., Meister, M., Perona, P., 2020. Synthetic Examples Improve Generalization for Rare Classes. Presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 863–873.

Belouadah, E., Popescu, A., Kanellos, I., 2021. A comprehensive study of class incremental learning algorithms for visual tasks. Neural Networks 135, 38–54. https://doi.org/10.1016/j.neunet.2020.12.003

Bengis, R.G., Leighton, F.A., Fischer, J.R., Artois, M., Mörner, T., Tate, C.M., 2004. The role of wildlife in emerging and re-emerging zoonoses. Scientific and Technical Review 497–511.

Biewald, L., 2020. Experiment tracking with weights and biases.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J.Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D.E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P.W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X.L., Li, X., Ma, T., Malik, A., Manning, C.D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J.C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J.S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A.W., Tramèr, F., Wang, R.E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S.M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K., Liang, P., 2022. On the Opportunities and Risks of Foundation Models. https://doi.org/10.48550/arXiv.2108.07258

Borowicz, A., Le, H., Humphries, G., Nehls, G., Höschle, C., Kosarev, V., Lynch, H.J., 2019. Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. PLOS ONE 14, e0212532. https://doi.org/10.1371/journal.pone.0212532

Bouché, P., Lejeune, P., Vermeulen, C., 2012. How to count elephants in West African savannahs? Synthesis and comparison of main gamecount methods. Biotechnol. Agron. Soc. Environ. 16, 77–91.

Bowler, E., Fretwell, P.T., French, G., Mackiewicz, M., 2020. Using Deep Learning to Count Albatrosses from Space: Assessing Results in Light of Ground Truth Uncertainty. Remote Sensing 12, 2026. https://doi.org/10.3390/rs12122026

Bröker, K.C.A., Hansen, R.G., Leonard, K.E., Koski, W.R., Heide-Jørgensen, M.P., 2019. A comparison of image and observer based aerial surveys of narwhal. Marine Mammal Science 35, 1253–1279. https://doi.org/10.1111/mms.12586

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language Models are Few-Shot Learners, in: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 1877–1901.

Brummitt, N., Regan, E.C., Weatherdon, L.V., Martin, C.S., Geijzendorffer, I.R., Rocchini, D., Gavish, Y., Haase, P., Marsh, C.J., Schmeller, D.S., 2017. Taking stock of nature: Essential biodiversity variables explained. Biological Conservation, SI:Measures of biodiversity 213, 252–255. https://doi.org/10.1016/j.biocon.2016.09.006

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2004. Advanced Distance Sampling: Estimating abundance of biological populations. OUP Oxford.

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A., 2020. Albumentations: Fast and Flexible Image Augmentations. Information 11, 125. https://doi.org/10.3390/info11020125

Butt, B., Turner, M.D., 2012. Clarifying competition: the case of wildlife and pastoral livestock in East Africa. Pastoralism: Research, Policy and Practice 2, 9. https://doi.org/10.1186/2041-7136-2-9

Cai, Z., Vasconcelos, N., 2021. Cascade R-CNN: High Quality Object Detection and Instance Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 1483–1498. https://doi.org/10.1109/TPAMI.2019.2956516

Cai, Z., Vasconcelos, N., 2018. Cascade R-CNN: Delving Into High Quality Object Detection. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6154–6162.

Campbell, J.B., Wynne, R.H., 2011. Introduction to Remote Sensing, Fifth Edition. Guilford Press.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact on humanity. Nature 486, 59–67. https://doi.org/10.1038/nature11148

Caughley, G., 1977. Sampling in Aerial Survey. The Journal of Wildlife Management 41, 605–615. https://doi.org/10.2307/3799980

Caughley, G., 1974. Bias in Aerial Survey. The Journal of Wildlife Management 38, 921–933. https://doi.org/10.2307/3800067

Caughley, G., Goddard, J., 1975. Abundance and distribution of elephants in the Luangwa Valley, Zambia. African Journal of Ecology 13, 39–48. https://doi.org/10.1111/j.1365-2028.1975.tb00122.x

CBD, 2011. Convention on Biological Diversity: text and annexes.

Ceballos, G., Ehrlich, P.R., 2023. Mutilation of the tree of life via mass extinction of animal genera. Proceedings of the National Academy of Sciences 120, e2306987120. https://doi.org/10.1073/pnas.2306987120

Ceballos, G., Ehrlich, P.R., 2006. Global mammal distributions, biodiversity hotspots, and conservation. Proceedings of the National Academy of Sciences 103, 19374–19379. https://doi.org/10.1073/pnas.0609334103

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances 1, e1400253. https://doi.org/10.1126/sciadv.1400253

Chabot, D., 2018. Trends in drone research and applications as the Journal of Unmanned Vehicle Systems turns five. J. Unmanned Veh. Sys. 6, vi–xv. https://doi.org/10.1139/juvs-2018-0005

Chabot, D., Bird, D.M., 2015. Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in? J. Unmanned Veh. Sys. 3, 137–155. https://doi.org/10.1139/juvs-2015-0021

Chase, M.J., Schlossberg, S., Griffin, C.R., Bouché, P.J.C., Djene, S.W., Elkan, P.W., Ferreira, S., Grossman, F., Kohi, E.M., Landen, K., Omondi, P., Peltier, A., Selier, S.A.J., Sutcliffe, R., 2016. Continent-wide survey reveals massive decline in African savannah elephants. PeerJ 4, e2354. https://doi.org/10.7717/peerj.2354

Chen, K., Liu, C., Chen, H., Zhang, H., Li, W., Zou, Z., Shi, Z., 2024. RSPrompter: Learning to Prompt for Remote Sensing Instance Segmentation Based on Visual Foundation Model. IEEE Transactions on Geoscience and Remote Sensing 62, 1–17. https://doi.org/10.1109/TGRS.2024.3356074

Chen, K., Wang, Jiaqi, Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, Jingdong, Shi, J., Ouyang, W., Loy, C.C., Lin, D., 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. https://doi.org/10.48550/arXiv.1906.07155

Christie, K.S., Gilbert, S.L., Brown, C.L., Hatfield, M., Hanson, L., 2016. Unmanned aircraft systems in wildlife research: current and future applications of a transformative technology. Frontiers in Ecology and the Environment 14, 241–251. https://doi.org/10.1002/fee.1281

CITES-MIKE, 2020. Monitoring the Illegal Killing of Elephants: Aerial Survey Standards for the MIKE Programme. Version 3.0. Presented at the Convention on International Trade in Endangered Species - Monitoring the Illegal Killing of Elephants Programme (CITES-MIKE), United Nations Environment Programme, Nairobi, Lenya.

Civil Aviation Authority, 2024. Microlights [WWW Document]. URL https://www.caa.co.uk/general-aviation/aircraft-ownership-and-maintenance/types-of-aircraft/microlights/ (accessed 3.15.24).

Corcoran, E., Denman, S., Hanger, J., Wilson, B., Hamilton, G., 2019. Automated detection of koalas using low-level aerial surveillance and machine learning. Sci Rep 9, 3208. https://doi.org/10.1038/s41598-019-39917-5

Corcoran, E., Winsen, M., Sudholz, A., Hamilton, G., 2021. Automated detection of wildlife using drones: Synthesis, opportunities and constraints. Methods in Ecology and Evolution 12, 1103–1114. https://doi.org/10.1111/2041-210X.13581

Corrêa, A.A., Quoos, J.H., Barreto, A.S., Groch, K.R., Eichler, P.P.B., 2022. Use of satellite imagery to identify southern right whales (Eubalaena australis) on a Southwest Atlantic Ocean breeding ground. Marine Mammal Science 38, 87–101. https://doi.org/10.1111/mms.12847

Cowie, R.H., Bouchet, P., Fontaine, B., 2022. The Sixth Mass Extinction: fact, fiction or speculation? Biological Reviews 97, 640–663. https://doi.org/10.1111/brv.12816

Craig, G.C., 2012. Aerial Survey standards for the MIKE Programme. Version 2.0. CITES MIKE programme, Nairobi.

Craigie, I.D., Baillie, J.E.M., Balmford, A., Carbone, C., Collen, B., Green, R.E., Hutton, J.M., 2010. Large mammal population declines in Africa's protected areas. Biological Conservation 143, 2221–2228. https://doi.org/10.1016/j.biocon.2010.06.007

Cubaynes, H.C., Fretwell, P.T., Bamford, C., Gerrish, L., Jackson, J.A., 2019. Whales from space: Four mysticete species described using new VHR satellite imagery. Marine Mammal Science 35, 466–491. https://doi.org/10.1111/mms.12544

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Presented at the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 886–893 vol. 1. https://doi.org/10.1109/CVPR.2005.177

De Leeuw, J., Waweru, M.N., Okello, O.O., Maloba, M., Nguru, P., Said, M.Y., Aligula, H.M., Heitkönig, I.M.A., Reid, R.S., 2001. Distribution and diversity of wildlife in northern Kenya in relation to livestock and permanent water points. Biological Conservation 100, 297–306. https://doi.org/10.1016/S0006-3207(01)00034-9

Delplanque, A., Foucher, S., Lejeune, P., Linchant, J., Théau, J., 2022. Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks. Remote Sensing in Ecology and Conservation 8, 166–179. https://doi.org/10.1002/rse2.234

Delplanque, A., Foucher, S., Théau, J., Bussière, E., Vermeulen, C., Lejeune, P., 2023a. From crowd to herd counting: How to precisely detect and count African mammals using aerial imagery and deep learning? ISPRS Journal of Photogrammetry and Remote Sensing 197, 167–180. https://doi.org/10.1016/j.isprsjprs.2023.01.025

Delplanque, A., Lamprey, R., Foucher, S., Théau, J., Lejeune, P., 2023b. Surveying wildlife and livestock in Uganda with aerial cameras: Deep Learning reduces the workload of human interpretation by over 70%. Front. Ecol. Evol. 11. https://doi.org/10.3389/fevo.2023.1270857

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.-J., Visseren-Hamakers, I., Willis, K.J., Zayas, C.N., 2019. Pervasive human-driven decline of life on Earth points to the need for transformative change. Science 366, eaax3100. https://doi.org/10.1126/science.aax3100

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B., Collen, B., 2014. Defaunation in the Anthropocene. Science 345, 401–406. https://doi.org/10.1126/science.1251817

Dunham, K.M., 2012. Trends in populations of elephant and other large herbivores in Gonarezhou National Park, Zimbabwe, as revealed by sample aerial surveys. African Journal of Ecology 50, 476–488. https://doi.org/10.1111/j.1365-2028.2012.01343.x

Duporge, I., Isupova, O., Reece, S., Macdonald, D.W., Wang, T., 2021. Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. Remote Sensing in Ecology and Conservation 7, 369–381. https://doi.org/10.1002/rse2.195

Dutta, A., Zisserman, A., 2019. The VIA Annotation Software for Images, Audio and Video, in: Proceedings of the 27th ACM International Conference on Multimedia, MM '19. Association for Computing Machinery, New York, NY, USA, pp. 2276–2279. https://doi.org/10.1145/3343031.3350535

Eberhardt, L.L., 1978. Transect Methods for Population Studies. The Journal of Wildlife Management 42, 1–31. https://doi.org/10.2307/3800685

Efroymson, R.A., Suter II, G.W., 2001. Ecological Risk Assessment Framework for Low-Altitude Aircraft Overflights: II. Estimating Effects on Wildlife. Risk Analysis 21, 263–274. https://doi.org/10.1111/0272-4332.212110

Eikelboom, J.A.J., Wind, J., van de Ven, E., Kenana, L.M., Schroder, B., de Knegt, H.J., van Langevelde, F., Prins, H.H.T., 2019. Improving the precision and accuracy of animal population estimates with aerial image object detection. Methods in Ecology and Evolution 10, 1875–1887. https://doi.org/10.1111/2041-210X.13277

Elgendy, M., 2020. Deep Learning for Vision Systems. Simon and Schuster.

Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2015. The Pascal Visual Object Classes Challenge: A Retrospective. Int J Comput Vis 111, 98–136. https://doi.org/10.1007/s11263-014-0733-5

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal Visual Object Classes (VOC) Challenge. Int J Comput Vis 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4

Ezeh, A., Kissling, F., Singer, P., 2020. Why sub-Saharan Africa might exceed its projected population size by 2100. The Lancet 396, 1131–1133. https://doi.org/10.1016/S0140-6736(20)31522-1

Fang, Y., Du, S., Boubchir, L., Djouani, K., 2021. Detecting African hoofed animals in aerial imagery using convolutional neural network. IJRA 10, 133. https://doi.org/10.11591/ijra.v10i2.pp133-143

Fischler, M.A., Elschlager, R.A., 1973. The Representation and Matching of Pictorial Structures. IEEE Transactions on Computers C–22, 67–92. https://doi.org/10.1109/T-C.1973.223602

Fleming, P.J.S., Tracey, J.P., Fleming, P.J.S., Tracey, J.P., 2008. Some human, aircraft and animal factors affecting aerial surveys: how to enumerate animals from the air. Wildl. Res. 35, 258–267. https://doi.org/10.1071/WR07081

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global Consequences of Land Use. Science 309, 570–574. https://doi.org/10.1126/science.1111772

Fonteyn, D., Vermeulen, C., Deflandre, N., Cornelis, D., Lhoest, S., Houngbégnon, F.G.A., Doucet, J.-L., Fayolle, A., 2021. Wildlife trail or systematic? Camera trap placement has little effect on estimates of mammal diversity in a tropical forest in Gabon. Remote Sensing in Ecology and Conservation 7, 321–336. https://doi.org/10.1002/rse2.191

Frederick, H., 2012. Aerial Wildlife Survey Manual- Aerial Procedures Manual v 0.9 (Uganda).

Frederick, P.C., Hylton, B., Heath, J.A., Ruane, M., 2003. Accuracy and variation in estimates of large numbers of birds by individual observers using an aerial survey simulator. forn 74, 281–287. https://doi.org/10.1648/0273-8570-74.3.281

Fretwell, P.T., Trathan, P.N., 2009. Penguins from space: faecal stains reveal the location of emperor penguin colonies. Global Ecology and Biogeography 18, 543–552. https://doi.org/10.1111/j.1466-8238.2009.00467.x

Freund, Y., Schapire, R.E., 1996. Experiments with a New Boosting Algorithm, in: Proceedings of the Thirteenth International Conference on Machine Learning. pp. 148–156.

Fukushima, K., 1969. Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements. IEEE Transactions on Systems Science and Cybernetics 5, 322–333. https://doi.org/10.1109/TSSC.1969.300225

Fynn, R.W.S., Augustine, D.J., Peel, M.J.S., de Garine-Wichatitsky, M., 2016. Strategic management of livestock to improve biodiversity conservation in African savannahs: a conceptual basis for wildlife–livestock coexistence. Journal of Applied Ecology 53, 388–397. https://doi.org/10.1111/1365-2664.12591

Gaidet-Drapier, N., Fritz, H., Bourgarel, M., Renaud, P.-C., Poilecot, P., Chardonnet, P., Coid, C., Poulet, D., Le Bel, S., 2006. Cost and Efficiency of Large

Mammal Census Techniques: Comparison of Methods for a Participatory Approach in a Communal Area, Zimbabwe. Biodivers Conserv 15, 735–754. https://doi.org/10.1007/s10531-004-1063-7

Gao, G., Gao, J., Liu, Q., Wang, Q., Wang, Y., 2020. CNN-based Density Estimation and Crowd Counting: A Survey. arXiv:2003.12783 [cs].

Georgiadis, N.J., Ihwagi, F., Olwero, J.G.N., Romañach, S.S., 2007. Savanna herbivore dynamics in a livestock-dominated landscape. II: Ecological, conservation, and management implications of predator restoration. Biological Conservation 137, 473–483. https://doi.org/10.1016/j.biocon.2007.03.006

Girdhar, R., El-Nouby, A., Liu, Z., Singh, M., Alwala, K.V., Joulin, A., Misra, I., 2023. ImageBind: One Embedding Space To Bind Them All. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15180–15190.

Girshick, R., 2015. Fast R-CNN. Presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2016. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 38, 142–158. https://doi.org/10.1109/TPAMI.2015.2437384

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep Sparse Rectifier Neural Networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Presented at the Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, pp. 315–323.

Goebel, M.E., Perryman, W.L., Hinke, J.T., Krause, D.J., Hann, N.A., Gardner, S., LeRoi, D.J., 2015. A small unmanned aerial system for estimating abundance and size of Antarctic predators. Polar Biol 38, 619–630. https://doi.org/10.1007/s00300-014-1625-4

Gonçalves, B.C., Spitzbart, B., Lynch, H.J., 2020. SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. Remote Sensing of Environment 239, 111617. https://doi.org/10.1016/j.rse.2019.111617

Goodenough, A.E., Berry, D.L., Carpenter, W.S., Dawson, M., Furlong, N., Lamb, R.J., MacTavish, L., O'Reilly, N., Toms, H., Whitehead, L.H., Hart, A.G.,

2024. Do you see what I see? Variation in detection, identification and enumeration of mammals during transect surveys. African Journal of Ecology 62, e13205. https://doi.org/10.1111/aje.13205

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Gower, J.C., Ross, G.J.S., 1969. Minimum Spanning Trees and Single Linkage Cluster Analysis. Journal of the Royal Statistical Society. Series C (Applied Statistics) 18, 54–64. https://doi.org/10.2307/2346439

Green, K.M., Virdee, M.K., Cubaynes, H.C., Aviles-Rivero, A.I., Fretwell, P.T., Gray, P.C., Johnston, D.W., Schönlieb, C.-B., Torres, L.G., Jackson, J.A., 2023. Gray whale detection in satellite imagery using deep learning. Remote Sensing in Ecology and Conservation 9, 829–840. https://doi.org/10.1002/rse2.352

Griffin, P.C., Lubow, B.C., Jenkins, K.J., Vales, D.J., Moeller, B.J., Reid, M., Happe, P.J., Mccorquodale, S.M., Tirhi, M.J., Schaberl, J.P., Beirne, K., 2013. A hybrid double-observer sightability model for aerial surveys. The Journal of Wildlife Management 77, 1532–1544. https://doi.org/10.1002/jwmg.612

Grimsdell, J.J.R., Westley, S., 1981. Low-level aerial survey techniques. International Livestock Centre for Africa.

Guirado, E., Tabik, S., Rivas, M.L., Alcaraz-Segura, D., Herrera, F., 2019. Whale counting in satellite and aerial images with deep learning. Sci Rep 9, 14259. https://doi.org/10.1038/s41598-019-50795-9

Guzhov, A., Raue, F., Hees, J., Dengel, A., 2022. Audioclip: Extending Clip to Image, Text and Audio, in: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 976–980. https://doi.org/10.1109/ICASSP43922.2022.9747631

Gwynne, M.D., Croze, H., 1975. East African habitat monitoring practice: a review of methods and application. Presented at the Proceedings of the Seminar on the evaluation and mapping of tropical African rangelands, Bamako, Mali, pp. 95–135.

Han, L., Tao, P., Martin, R.R., 2019. Livestock detection in aerial images using a fully convolutional network. Comp. Visual Media 5, 221–228. https://doi.org/10.1007/s41095-019-0132-5

Hassanalian, M., Abdelkefi, A., 2017. Classifications, applications, and design challenges of drones: A review. Progress in Aerospace Sciences 91, 99–131. https://doi.org/10.1016/j.paerosci.2017.04.003

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intelligent Systems and their Applications 13, 18–28. https://doi.org/10.1109/5254.708428

Hennenberg, K.J., Fischer, F., Kouadio, K., Goetze, D., Orthmann, B., Linsenmair, K.E., Jeltsch, F., Porembski, S., 2006. Phytomass and fire occurrence along forest–savanna transects in the Comoé National Park, Ivory Coast. Journal of Tropical Ecology 22, 303–311. https://doi.org/10.1017/S0266467405003007

Herrero, M., Grace, D., Njuki, J., Johnson, N., Enahoro, D., Silvestri, S., Rufino, M.C., 2013. The roles of livestock in developing countries. Animal 7, 3–18. https://doi.org/10.1017/S1751731112001954

Hetem, R.S., Fuller, A., Maloney, S.K., Mitchell, D., 2014. Responses of large mammals to climate change. Temperature 1, 115–127. https://doi.org/10.4161/temp.29651

Hodgson, A., Kelly, N., Peel, D., 2013. Unmanned Aerial Vehicles (UAVs) for Surveying Marine Fauna: A Dugong Case Study. PLOS ONE 8, e79556. https://doi.org/10.1371/journal.pone.0079556

Hodgson, J.C., Mott, R., Baylis, S.M., Pham, T.T., Wotherspoon, S., Kilpatrick, A.D., Raja Segaran, R., Reid, I., Terauds, A., Koh, L.P., 2018. Drones count wildlife more accurately and precisely than humans. Methods in Ecology and Evolution 9, 1160–1167. https://doi.org/10.1111/2041-210X.12974

Hoeser, T., Kuenzer, C., 2020. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. Remote Sensing 12, 1667. https://doi.org/10.3390/rs12101667

Hooper, D.U., Chapin III, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J., Wardle, D.A., 2005. Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge. Ecological Monographs 75, 3–35. https://doi.org/10.1890/04-0922

Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. The Journal of Physiology 195, 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160, 106-154.2.

Irvine, J.M., Nolan, J., Hofmann, N., Lewis, D., Simpamba, T., Zyambo, P., Travis, A.J., Hemami, S., 2019. Estimating the Population of Large Animals in the Wild Using Satellite Imagery: A Case Study of Hippos in Zambia's Luangwa River, in: 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). Presented at the 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1–8. https://doi.org/10.1109/AIPR47015.2019.9174564

Isbell, F., Gonzalez, A., Loreau, M., Cowles, J., Díaz, S., Hector, A., Mace, G.M., Wardle, D.A., O'Connor, M.I., Duffy, J.E., Turnbull, L.A., Thompson, P.L., Larigauderie, A., 2017. Linking the influence and dependence of people on biodiversity across scales. Nature 546, 65–72. https://doi.org/10.1038/nature22899

Ivosevic, B., Han, Y.-G., Kwon, O., 2017. Monitoring butterflies with an unmanned aerial vehicle: current possibilities and future potentials. j ecology environ 41, 12. https://doi.org/10.1186/s41610-017-0028-1

Jachmann, H., 2002. Comparison of Aerial Counts with Ground Counts for Large African Herbivores. Journal of Applied Ecology 39, 841–852.

Jachmann, H., 2001. Estimating Abundance of African Wildlife: An Aid to Adaptive Management. Springer Science & Business Media.

Jachmann, H., 1991. Evaluation of four survey methods for estimating elephant densities. African Journal of Ecology 29, 188–195. https://doi.org/10.1111/j.1365-2028.1991.tb01001.x

Jenet, A., Buono, N., Di Lello, S., Gomarasca, M., Heine, C., Mason, S., Nori, M., Saavedra, R., Van Troos, K., 2016. The Path to Greener Pastures: Pastoralism, the Backbone of the World's Drylands. https://doi.org/10.2139/ssrn.3888381

Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M., Geller, G.N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F.E., Pereira, H.M., Regan, E.C., Schmeller, D.S., Turak, E., 2019. Essential biodiversity variables for mapping and monitoring species populations. Nat Ecol Evol 3, 539–551. https://doi.org/10.1038/s41559-019-0826-1

Jiménez López, J., Mulero-Pázmány, M., 2019. Drones for Conservation in Protected Areas: Present and Future. Drones 3, 10. https://doi.org/10.3390/drones3010010

Jing, L., Tian, Y., 2021. Self-Supervised Visual Feature Learning With Deep Neural Networks: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 4037–4058. https://doi.org/10.1109/TPAMI.2020.2992393

Johnson, C.N., Balmford, A., Brook, B.W., Buettel, J.C., Galetti, M., Guangchun, L., Wilmshurst, J.M., 2017. Biodiversity losses and conservation responses in the

Anthropocene. Science 356, 270–275. https://doi.org/10.1126/science.aam9317

Jolly, G.M., 1969. Sampling Methods for Aerial Censuses of Wildlife Populations. East African Agricultural and Forestry Journal 34, 46–49. https://doi.org/10.1080/00128325.1969.11662347

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

Kellenberger, B., Marcos, D., Lobry, S., Tuia, D., 2019a. Half a Percent of Labels is Enough: Efficient Animal Detection in UAV Imagery using Deep CNNs and Active Learning. IEEE Trans. Geosci. Remote Sensing 57, 9524–9533. https://doi.org/10.1109/TGRS.2019.2927393

Kellenberger, B., Marcos, D., Tuia, D., 2019b. When a Few Clicks Make All the Difference: Improving Weakly-Supervised Wildlife Detection in UAV Images, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Long Beach, CA, USA, pp. 1414–1422. https://doi.org/10.1109/CVPRW.2019.00182

Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. Remote Sensing of Environment 216, 139–153. https://doi.org/10.1016/j.rse.2018.06.028

Kellenberger, B., Veen, T., Folmer, E., Tuia, D., 2021. 21 000 birds in 4.5 h: efficient large-scale seabird detection with machine learning. Remote Sensing in Ecology and Conservation 7, 445–460. https://doi.org/10.1002/rse2.200

Kellenberger, B., Volpi, M., Tuia, D., 2017. Fast animal detection in UAV images using convolutional neural networks, in: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Presented at the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, Fort Worth, TX, pp. 866–869. https://doi.org/10.1109/IGARSS.2017.8127090

Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/arXiv.1412.6980

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollar, P., Girshick, R., 2023. Segment Anything. Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026.

Krause, D.J., Hinke, J.T., 2021. Finally Within Reach: A Drone Census of an Important, But Practically Inaccessible, Antarctic Fur Seal Colony. Aquat Mamm 47, 349–354. https://doi.org/10.1578/AM.47.4.2021.349

Kruger, J.M., Reilly, B.K., Whyte, I.J., 2008. Application of distance sampling to estimate population densities of large herbivores in Kruger National Park. Wildl. Res. 35, 371–376. https://doi.org/10.1071/WR07084

Kruskal, J.B., 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. Proc. Amer. Math. Soc. 7, 48–50. https://doi.org/10.1090/S0002-9939-1956-0078686-7

Kuenzer, C., Ottinger, M., Wegmann, M., Guo, H., Wang, C., Zhang, J., Dech, S., Wikelski, M., 2014. Earth observation satellite sensors for biodiversity monitoring: potentials and bottlenecks. International Journal of Remote Sensing 35, 6599–6647. https://doi.org/10.1080/01431161.2014.964349

Lacher, T.E., Jr., Davidson, A.D., Fleming, T.H., Gómez-Ruiz, E.P., McCracken, G.F., Owen-Smith, N., Peres, C.A., Vander Wall, S.B., 2019. The functional roles of mammals in ecosystems. Journal of Mammalogy 100, 942–964. https://doi.org/10.1093/jmammal/gyy183

Lamprey, R., Ochanda, D., Brett, R., Tumwesigye, C., Douglas-Hamilton, I., 2020a. Cameras replace human observers in multi-species aerial counts in Murchison Falls, Uganda. Remote Sensing in Ecology and Conservation 6, 529–545. https://doi.org/10.1002/rse2.154

Lamprey, R., Pope, F., Ngene, S., Norton-Griffiths, M., Frederick, H., Okita-Ouma, B., Douglas-Hamilton, I., 2020b. Comparing an automated high-definition oblique camera system to rear-seat-observers in a wildlife survey in Tsavo, Kenya: Taking multi-species aerial counts to the next level. Biological Conservation 241, 108243. https://doi.org/10.1016/j.biocon.2019.108243

Lamprey, R.H., Keigwin, M., Tumwesigye, C., 2023. A high-resolution aerial camera survey of Uganda's Queen Elizabeth Protected Area improves detection of wildlife and delivers a surprisingly high estimate of the elephant population. https://doi.org/10.1101/2023.02.06.525067

LAND INFO Worldwide Mapping, LLC, 2023. Buying Satellite Imagery: Pricing Information for High Resolution Satellite Imagery [WWW Document]. LAND INFO Worldwide Mapping, LLC. URL https://landinfo.com/satellite-imagery-pricing/ (accessed 4.15.23).

LaRue, M., Salas, L., Nur, N., Ainley, D., Stammerjohn, S., Pennycook, J., Dozier, M., Saints, J., Stamatiou, K., Barrington, L., Rotella, J., 2021. Insights from the first global population estimate of Weddell seals in Antarctica. Science Advances 7, eabh3674. https://doi.org/10.1126/sciadv.abh3674

LaRue, M.A., Rotella, J.J., Garrott, R.A., Siniff, D.B., Ainley, D.G., Stauffer, G.E., Porter, C.C., Morin, P.J., 2011. Satellite imagery can be used to detect variation in abundance of Weddell seals (Leptonychotes weddellii) in Erebus Bay, Antarctica. Polar Biol 34, 1727. https://doi.org/10.1007/s00300-011-1023-0

LaRue, M.A., Stapleton, S., 2018. Estimating the abundance of polar bears on Wrangel Island during late summer using high-resolution satellite imagery: a pilot study. Polar Biol 41, 2621–2626. https://doi.org/10.1007/s00300-018-2384-4

LaRue, M.A., Stapleton, S., Anderson, M., 2017. Feasibility of using high-resolution satellite imagery to assess vertebrate wildlife populations. Conservation Biology 31, 213–220. https://doi.org/10.1111/cobi.12809

Latte, N., Gaucher, P., Bolyn, C., Lejeune, P., Michez, A., 2020. Upscaling UAS Paradigm to UltraLight Aircrafts: A Low-Cost Multi-Sensors System for Large Scale Aerial Photogrammetry. Remote Sensing 12, 1265. https://doi.org/10.3390/rs12081265

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L., 1989. Handwritten Digit Recognition with a Back-Propagation Network, in: Advances in Neural Information Processing Systems. Morgan-Kaufmann.

Leedy, D.L., 1948. Aerial Photographs, Their Interpretation and Suggested Uses in Wildlife Management. The Journal of Wildlife Management 12, 191–210. https://doi.org/10.2307/3796415

Lema, D.G., Pedrayes, O.D., Usamentiaga, R., García, D.F., Alonso, Á., 2021. Cost-Performance Evaluation of a Recognition Service of Livestock Activity Using Aerial Images. Remote Sensing 13, 2318. https://doi.org/10.3390/rs13122318

Lempitsky, V., Zisserman, A., 2010. Learning To Count Objects in Images, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Lethbridge, M., Stead, M., Wells, C., 2019. Estimating kangaroo density by aerial survey: a comparison of thermal cameras with human observers. Wildl. Res. 46, 639–648. https://doi.org/10.1071/WR18122

Li, B., Huang, H., Zhang, A., Liu, P., Liu, C., 2021. Approaches on crowd counting and density estimation: a review. Pattern Anal Applic 24, 853–874. https://doi.org/10.1007/s10044-021-00959-z

Liang, D., Xu, W., Zhu, Y., Zhou, Y., 2023. Focal Inverse Distance Transform Maps for Crowd Localization. IEEE Transactions on Multimedia 25, 6040–6052. https://doi.org/10.1109/TMM.2022.3203870

Lienhart, R., Maydt, J., 2002. An extended set of Haar-like features for rapid object detection, in: Proceedings. International Conference on Image Processing. Presented at the Proceedings. International Conference on Image Processing, p. I–I. https://doi.org/10.1109/ICIP.2002.1038171

Lillesand, T., Kiefer, R.W., Chipman, J., 2015. Remote Sensing and Image Interpretation. John Wiley & Sons.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature Pyramid Networks for Object Detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944. https://doi.org/10.1109/CVPR.2017.106

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017b. Focal Loss for Dense Object Detection. Presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common Objects in Context, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Linchant, J., 2021. Reinventing wildlife census with unmanned aerial systems: new survey designs for ungulate counts in the vast African semi-open ecosystems. ULiège - Université de Liège, Gembloux, Belgium.

Linchant, J., Lhoest, S., Quevauvillers, S., Lejeune, P., Vermeulen, C., Ngabinzeke, J.S., Belanganayi, B.L., Delvingt, W., Bouché, P., 2018. UAS imagery reveals new survey opportunities for counting hippos. PLOS ONE 13, e0206413. https://doi.org/10.1371/journal.pone.0206413

Linchant, J., Lhoest, S., Quevauvillers, S., Semeki, J., Lejeune, P., Vermeulen, C., 2015a. WIMUAS: Developing a tool to review wildlife data from various UAS flight plans. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-3-W3, 379–384. https://doi.org/10.5194/isprsarchives-XL-3-W3-379-2015

Linchant, J., Lisein, J., Semeki, J., Lejeune, P., Vermeulen, C., 2015b. Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges. Mammal Review 45, 239–252. https://doi.org/10.1111/mam.12046

Lisein, J., Linchant, J., Lejeune, P., Bouché, P., Vermeulen, C., 2013. Aerial Surveys Using an Unmanned Aerial System (UAS): Comparison of Different Methods for Estimating the Surface Area of Sampling Strips. Tropical Conservation Science 6, 506–520. https://doi.org/10.1177/194008291300600405

Liu, J., Gao, C., Meng, D., Hauptmann, A.G., 2018. DecideNet: Counting Varying Density Crowds Through Attention Guided Detection and Density Estimation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5197–5206.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. SSD: Single Shot MultiBox Detector, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Lowe, D.G., 1999. Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision. Presented at the Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 1150–1157 vol.2. https://doi.org/10.1109/ICCV.1999.790410

Lynch, H.J., LaRue, M.A., 2014. First global census of the Adélie Penguin. The Auk 131, 457–466. https://doi.org/10.1642/AUK-14-31.1

Martinson, E., Furlong, B., Gillies, A., 2021. Training Rare Object Detection in Satellite Imagery With Synthetic GAN Images. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2769–2776.

Mayaux, P., Bartholomé, E., Fritz, S., Belward, A., 2004. A new land-cover map of Africa for the year 2000. Journal of Biogeography 31, 861–877. https://doi.org/10.1111/j.1365-2699.2004.01073.x

Miao, Z., Elizalde, B., Deshmukh, S., Kitzes, J., Wang, H., Dodhia, R., Ferres, J.M.L., 2023. Zero-Shot Transfer for Wildlife Bioacoustics Detection (preprint). In Review. https://doi.org/10.21203/rs.3.rs-3180218/v1

Moreni, M., Theau, J., Foucher, S., 2021. Train Fast While Reducing False Positives: Improving Animal Classification Performance Using Convolutional Neural Networks. Geomatics 1, 34–49. https://doi.org/10.3390/geomatics1010004

Mou, C., Liang, A., Hu, C., Meng, F., Han, B., Xu, F., 2023a. Monitoring Endangered and Rare Wildlife in the Field: A Foundation Deep Learning Model Integrating Human Knowledge for Incremental Recognition with Few Data and Low Cost. Animals 13, 3168. https://doi.org/10.3390/ani13203168

Mou, C., Liu, T., Zhu, C., Cui, X., 2023b. WAID: A Large-Scale Dataset for Wildlife Detection with Drones. Applied Sciences 13, 10397. https://doi.org/10.3390/app131810397

Mücher, C.A., Los, S., Franke, G.J., Kamphuis, C., 2022. Detection, identification and posture recognition of cattle with satellites, aerial photography and UAVs

using deep learning techniques. International Journal of Remote Sensing 43, 2377–2392. https://doi.org/10.1080/01431161.2022.2051634

Mulero-Pázmány, M., Stolper, R., Essen, L.D. van, Negro, J.J., Sassen, T., 2014. Remotely Piloted Aircraft Systems as a Rhinoceros Anti-Poaching Tool in Africa. PLOS ONE 9, e83873. https://doi.org/10.1371/journal.pone.0083873

Naidoo, K., 2019. MiSTree: a Python package for constructing and analysing Minimum Spanning Trees. Journal of Open Source Software 4, 1721. https://doi.org/10.21105/joss.01721

Naudé, J., Joubert, D., 2019. The Aerial Elephant Dataset: A New Public Benchmark for Aerial Object Detection. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 48–55.

Naveen, R., Lynch, H.J., Forrest, S., Mueller, T., Polito, M., 2012. First direct, site-wide penguin survey at Deception Island, Antarctica, suggests significant declines in breeding chinstrap penguins. Polar Biol 35, 1879–1888. https://doi.org/10.1007/s00300-012-1230-3

Nichols, J.D., Williams, B.K., 2006. Monitoring for conservation. Trends in Ecology & Evolution 21, 668–673. https://doi.org/10.1016/j.tree.2006.08.007

Nikolenko, S.I., 2021. Synthetic Data for Deep Learning, Springer Optimization and Its Applications. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-75178-4

Norton-Griffiths, M., 1978. Counting Animals, J.J.R Grimsdell. ed, Handbook No. 1. African Wildlife Leadership Foundation, Nairobi, Kenya.

Norton-Griffiths, M., 1976. Further Aspects of Bias in Aerial Census of Large Mammals. The Journal of Wildlife Management 40, 368–371. https://doi.org/10.2307/3800445

Norton-Griffiths, M., 1974. Reducing counting bias in aerial censuses by photography*. African Journal of Ecology 12, 245–248. https://doi.org/10.1111/j.1365-2028.1974.tb00119.x

Ocholla, I.A., Pellikka, P., Karanja, F.N., Vuorinne, I., Odipo, V., Heiskanen, J., 2024. Livestock detection in African rangelands: Potential of high-resolution remote sensing data. Remote Sensing Applications: Society and Environment 33, 101139. https://doi.org/10.1016/j.rsase.2024.101139

Odadi, W.O., Riginos, C., Rubenstein, D.I., 2018. Tightly Bunched Herding Improves Cattle Performance in African Savanna Rangeland. Rangeland Ecology & Management 71, 481–491. https://doi.org/10.1016/j.rama.2018.03.008

Ogutu, J.O., Bhola, N., Piepho, H.-P., Reid, R., 2006. Efficiency of strip- and line-transect surveys of African savanna mammals. Journal of Zoology 269, 149–160. https://doi.org/10.1111/j.1469-7998.2006.00055.x

Oksuz, K., Cam, B.C., Kalkan, S., Akbas, E., 2021. Imbalance Problems in Object Detection: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 3388–3415. https://doi.org/10.1109/TPAMI.2020.2981890

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51, 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Olsoy, P.J., Shipley, L.A., Rachlow, J.L., Forbey, J.S., Glenn, N.F., Burgess, M.A., Thornton, D.H., 2018. Unmanned aerial systems measure structural habitat features for wildlife across multiple scales. Methods in Ecology and Evolution 9, 594–604. https://doi.org/10.1111/2041-210X.12919

Ottichilo, W.K., Khaemba, W.M., 2001. Validation of observer and aircraft calibration for aerial surveys of animals. African Journal of Ecology 39, 45–50. https://doi.org/10.1111/j.1365-2028.2001.00268.x

Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., Dubash, N.K., Edenhofer, O., Elgizouli, I., Field, C.B., Forster, P., Friedlingstein, P., Fuglestvedt, J., Gomez-Echeverri, L., Hallegatte, S., Hegerl, G., Howden, M., Jiang, K., Jimenez Cisneroz, B., Kattsov, V., Lee, H., Mach, K.J., Marotzke, J., Mastrandrea, M.D., Meyer, L., Minx, J., Mulugetta, Y., O'Brien, K., Oppenheimer, M., Pereira, J.J., Pichs-Madruga, R., Plattner, G.-K., Pörtner, H.-O., Power, S.B., Preston, B., Ravindranath, N.H., Reisinger, A., Riahi, K., Rusticucci, M., Scholes, R., Seyboth, K., Sokona, Y., Stavins, R., Stocker, T.F., Tschakert, P., van Vuuren, D., van Ypserle, J.-P., 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, EPIC3Geneva, Switzerland, IPCC, 151 p., pp. 151, ISBN: 978-92-9169-143-2. IPCC, Geneva, Switzerland.

Padubidri, C., Kamilaris, A., Karatsiolis, S., Kamminga, J., 2021. Counting sea lions and elephants from aerial photography using deep learning with density maps. Animal Biotelemetry 9, 27. https://doi.org/10.1186/s40317-021-00247-x

PAEAS, 2014. Aerial Survey Standards and Guidelines for the Pan-african Elephant Aerial Survey. Vulcan Inc, Seattle, USA.

Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 22, 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., Lin, D., 2019. Libra R-CNN: Towards Balanced Learning for Object Detection, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, pp. 821–830. https://doi.org/10.1109/CVPR.2019.00091

Parmesan, C., 2006. Ecological and Evolutionary Responses to Recent Climate Change. Annual Review of Ecology, Evolution, and Systematics 37, 637–669. https://doi.org/10.1146/annurev.ecolsys.37.091305.110100

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., Ye, H., 2020. Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. ISPRS Journal of Photogrammetry and Remote Sensing 169, 364–376. https://doi.org/10.1016/j.isprsjprs.2020.08.026

Pennycuick, C.J., Western, D., 1972. An investigation of some sources of bias in aerial transect sampling of large mammal populations. African Journal of Ecology 10, 175–191. https://doi.org/10.1111/j.1365-2028.1972.tb00726.x

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential Biodiversity Variables. Science 339, 277–278. https://doi.org/10.1126/science.1229931

Petso, T., Jamisola, R.S., Mpoeleng, D., Bennitt, E., Mmereki, W., 2021. Automatic animal identification from drone camera based on point pattern analysis of herd behaviour. Ecological Informatics 66, 101485. https://doi.org/10.1016/j.ecoinf.2021.101485

Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., Turner, W., 2014. Satellite remote sensing for applied ecologists: opportunities and

challenges. Journal of Applied Ecology 51, 839–848. https://doi.org/10.1111/1365-2664.12261

Pielawski, N., Wählby, C., 2020. Introducing Hann windows for reducing edge-effects in patch-based image segmentation. PLOS ONE 15, e0229839. https://doi.org/10.1371/journal.pone.0229839

Qian, Y., Humphries, G.R.W., Trathan, P.N., Lowther, A., Donovan, C.R., 2023. Counting animals in aerial images with a density map estimation model. Ecology and Evolution 13, e9903. https://doi.org/10.1002/ece3.9903

Rahnemoonfar, M., Dobbs, D., Yari, M., Starek, M.J., 2019. DisCountNet: Discriminating and Counting Network for Real-Time Counting and Localization of Sparse Objects in High-Resolution UAV Imagery. Remote Sensing 11, 1128. https://doi.org/10.3390/rs11091128

Rajabi, M.S., Beigi, P., Aghakhani, S., 2021. Drone Delivery Systems and Energy Management: A Review and Future Trends, in: Fathi, M., Zio, E., Pardalos, P.M. (Eds.), Handbook of Smart Energy Systems. Springer International Publishing, Cham, pp. 1–19. https://doi.org/10.1007/978-3-030-72322-4_196-1

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

Redmon, J., Farhadi, A., 2017. YOLO9000: Better, Faster, Stronger. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Rey, N., Volpi, M., Joost, S., Tuia, D., 2017. Detecting animals in African Savanna with UAVs and the crowds. Remote Sensing of Environment 200, 341–351. https://doi.org/10.1016/j.rse.2017.08.026

Ribera, J., Guera, D., Chen, Y., Delp, E.J., 2019. Locating Objects Without Bounding Boxes. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6479–6489.

Richard, D., Alary, V., Corniaux, C., Duteurtre, G., Lhoste, P., 2019. Dynamique des elevages pastoraux et agropastoraux en Afrique intertropicale. Ed. Quae; CTA, Versailles.

Riggio, J., Jacobson, A.P., Hijmans, R.J., Caro, T., 2019. How effective are the protected areas of East Africa? Global Ecology and Conservation 17, e00573. https://doi.org/10.1016/j.gecco.2019.e00573

Rivas, A., Chamoso, P., González-Briones, A., Corchado, J.M., 2018. Detection of Cattle Using Drones and Convolutional Neural Networks. Sensors (Basel) 18, 2048. https://doi.org/10.3390/s18072048

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sánchez-Díaz, B., Mata-Zayas, E.E., 2019. Remote sensing as indispensable technology in ecology to support the protection of biodiversity: a review. International Journal of Conservation Science 10, 811–820.

Sarwar, F., Griffin, A., Rehman, S.U., Pasang, T., 2021. Detecting sheep in UAV images. Computers and Electronics in Agriculture 187, 106219. https://doi.org/10.1016/j.compag.2021.106219

Sasse, D.B., 2003. Job-Related Mortality of Wildlife Workers in the United States, 1937-2000. Wildlife Society Bulletin (1973-2006) 31, 1015–1020.

Schlossberg, S., Chase, M.J., Griffin, C.R., 2016. Testing the Accuracy of Aerial Surveys for Large Mammals: An Experiment with African Savanna Elephants (Loxodonta africana). PLOS ONE 11, e0164904. https://doi.org/10.1371/journal.pone.0164904

Scholte, P., Kari, S., Moritz, M., 2022a. Thousands of pastoralists seek refuge in Waza National Park, Cameroon. Oryx 56, 330–330. https://doi.org/10.1017/S0030605322000217

Scholte, P., Pays, O., Adam, S., Chardonnet, B., Fritz, H., Mamang, J.-B., Prins, H.H.T., Renaud, P.-C., Tadjo, P., Moritz, M., 2022b. Conservation

overstretch and long-term decline of wildlife and tourism in the Central African savannas. Conservation Biology 36, e13860. https://doi.org/10.1111/cobi.13860

Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., Naemura, T., 2020. Cattle detection and counting in UAV images based on convolutional neural networks. International Journal of Remote Sensing 41, 31–52. https://doi.org/10.1080/01431161.2019.1624858

Shepley, A., Falzon, G., Meek, P., Kwan, P., 2021. Automated location invariant animal detection in camera trap images using publicly available data sources. Ecology and Evolution 11, 4494–4506. https://doi.org/10.1002/ece3.7344

Shrivastava, A., Gupta, A., Girshick, R., 2016. Training Region-Based Object Detectors With Online Hard Example Mining. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 761–769.

Singh, A., Hu, R., Goswami, V., Couairon, G., Galuba, W., Rohrbach, M., Kiela, D., 2022. FLAVA: A Foundational Language and Vision Alignment Model. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15638–15650.

Siniff, D.B., Skoog, R.O., 1964. Aerial Censusing of Caribou Using Stratified Random Sampling. The Journal of Wildlife Management 28, 391–401. https://doi.org/10.2307/3798104

Soares, V.H.A., Ponti, M.A., Gonçalves, R.A., Campello, R.J.G.B., 2021. Cattle counting in the wild with geolocated aerial images in large pasture areas. Computers and Electronics in Agriculture 189, 106354. https://doi.org/10.1016/j.compag.2021.106354

Soviany, P., Ionescu, R.T., 2018. Optimizing the Trade-Off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction, in: 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). Presented at the 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), IEEE, Timisoara, Romania, pp. 209–214. https://doi.org/10.1109/SYNASC.2018.00041

Stapleton, S., LaRue, M., Lecomte, N., Atkinson, S., Garshelis, D., Porter, C., Atwood, T., 2014. Polar Bears from Space: Assessing Satellite Imagery as a Tool to Track Arctic Wildlife. PLOS ONE 9, e101513. https://doi.org/10.1371/journal.pone.0101513

Stelfox, J.G., Peden, D.G., 1981. The aerial survey programme of the Kenya rangeland ecological monitoring unit: 1976-79. International Livestock Centre for Africa.

Stolton, S., Shadie, P., Dudley, N., 2013. Guidelines for applying protected area management categories including IUCN WCPA best practice guidance on recognising protected areas and assigning management categories and governance types, Best Practice Protected Area Guidelines Series. IUCN, Gland, Switzerland.

Sun, X., Wang, P., Lu, W., Zhu, Z., Lu, X., He, Qibin, Li, J., Rong, X., Yang, Z., Chang, H., He, Qinglin, Yang, G., Wang, R., Lu, J., Fu, K., 2023. RingMo: A Remote Sensing Foundation Model With Masked Image Modeling. IEEE Transactions on Geoscience and Remote Sensing 61, 1–22. https://doi.org/10.1109/TGRS.2022.3194732

Swinbourne, M.J., Taggart, D.A., Swinbourne, A.M., Lewis, M., Ostendorf, B., 2018. Using satellite imagery to assess the distribution and abundance of southern hairy-nosed wombats (Lasiorhinus latifrons). Remote Sensing of Environment 211, 196–203. https://doi.org/10.1016/j.rse.2018.04.017

Szeliski, R., 2022. Computer Vision: Algorithms and Applications. Springer Nature.

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., Di Salvo, P.A., Lewis, J.S., White, M.D., Teton, B., Beasley, J.C., Schlichting, P.E., Boughton, R.K., Wight, B., Newkirk, E.S., Ivan, J.S., Odell, E.A., Brook, R.K., Lukacs, P.M., Moeller, A.K., Mandeville, E.G., Clune, J., Miller, R.S., 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution 10, 585–590. https://doi.org/10.1111/2041-210X.13120

Terletzky, P.A., Ramsey, R.D., 2016. Comparison of Three Techniques to Identify and Count Individual Animals in Aerial Imagery. Journal of Signal and Information Processing 7, 123–135. https://doi.org/10.4236/jsip.2016.73013

Thompson, M.E., Schwager, S.J., Payne, K.B., Turkalo, A.K., 2010. Acoustic estimation of wildlife abundance: methodology for vocal mammals in forested habitats. African Journal of Ecology 48, 654–661. https://doi.org/10.1111/j.1365-2028.2009.01161.x

Thuiller, W., Broennimann, O., Hughes, G., Alkemade, J.R.M., Midgley, Guy.F., Corsi, F., 2006. Vulnerability of African mammals to anthropogenic climate change under conservative land transformation assumptions. Global Change Biology 12, 424–440. https://doi.org/10.1111/j.1365-2486.2006.01115.x

Tkachenko, M., Malyuk, M., Holmanyuk, A., Liubimov, N., 2020. Label Studio: Data labeling software.

Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M., Kohi, E.M., Hopcraft, G.C., 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images.

Methods in Ecology and Evolution 10, 779–787. https://doi.org/10.1111/2041-210X.13165

Toutain, B., Visscher, M.-N.D., Dulieu, D., 2004. Pastoralism and Protected Areas: Lessons Learned from Western Africa. Human Dimensions of Wildlife 9, 287–295. https://doi.org/10.1080/108071200490505963

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse, B., Mathis, A., Mathis, M.W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I.D., van Horn, G., Crofoot, M.C., Stewart, C.V., Berger-Wolf, T., 2022. Perspectives in machine learning for wildlife conservation. Nat Commun 13, 792. https://doi.org/10.1038/s41467-022-27980-y

Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A.K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L.P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M., Woodcock, C., 2015. Free and open-access satellite data are key to biodiversity conservation. Biological Conservation 182, 173–176. https://doi.org/10.1016/j.biocon.2014.11.048

UNEP-WCMC, IUCN, 2024. Protected Planet: The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM) [WWW Document]. Protected Planet. URL https://www.protectedplanet.net (accessed 1.18.24).

United Nations Environment Programme, 1992. Convention on biological diversity, June 1992.

Vandermeer, J.H., 2002. Tropical Agroecosystems. CRC Press.

Vermeulen, C., Lejeune, P., Lisein, J., Sawadogo, P., Bouché, P., 2013. Unmanned Aerial Survey of Elephants. PLOS ONE 8, e54700. https://doi.org/10.1371/journal.pone.0054700

Verykokou, S., Ioannidis, C., 2018. Oblique aerial images: a review focusing on georeferencing procedures. International Journal of Remote Sensing 39, 3452–3496. https://doi.org/10.1080/01431161.2018.1444294

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Presented at the Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, p. I–I. https://doi.org/10.1109/CVPR.2001.990517

Viola, P., Jones, M.J., 2004. Robust Real-Time Face Detection. International Journal of Computer Vision 57, 137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

Vrieling, A., Fava, F., Leitner, S., Merbold, L., Cheng, Y., Nakalema, T., Groen, T., Butterbach-Bahl, K., 2022. Identification of temporary livestock enclosures in Kenya from multi-temporal PlanetScope imagery. Remote Sensing of Environment 279, 113110. https://doi.org/10.1016/j.rse.2022.113110

Wacher, T., 2019. Aerial survey of the Ennedi Massif. Zoological Society of London.

Wal, E.V., Mcloughlin, P.D., Brook, R.K., 2011. Spatial and Temporal Factors Influencing Sightability of Elk. The Journal of Wildlife Management 75, 1521–1526.

Waltert, M., Meyer, B., Shanyangi, M.W., Balozi, J.J., Kitwara, O., Qolli, S., Krischke, H., Mühlenberg, M., 2008. Foot Surveys of Large Mammals in Woodlands of Western Tanzania. The Journal of Wildlife Management 72, 603–610. https://doi.org/10.2193/2006-456

Wang, D., Shao, Q., Yue, H., 2019. Surveying Wild Animals from Satellites, Manned Aircraft and Unmanned Aerial Systems (UASs): A Review. Remote Sensing 11, 1308. https://doi.org/10.3390/rs11111308

Wang, D., Song, Q., Liao, X., Ye, H., Shao, Q., Fan, J., Cong, N., Xin, X., Yue, H., Zhang, H., 2020. Integrating satellite and unmanned aircraft system (UAS) imagery to model livestock population dynamics in the Longbao Wetland National Nature Reserve, China. Science of The Total Environment 746, 140327. https://doi.org/10.1016/j.scitotenv.2020.140327

Wang, X., Shrivastava, A., Gupta, A., 2017. A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2606–2615.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612. https://doi.org/10.1109/TIP.2003.819861

Watson, J.E.M., Venter, O., Lee, J., Jones, K.R., Robinson, J.G., Possingham, H.P., Allan, J.R., 2018. Protect the last of the wild. Nature 563, 27–30. https://doi.org/10.1038/d41586-018-07183-6

Watts, A.C., Perry, J.H., Smith, S.E., Burgess, M.A., Wilkinson, B.E., Szantoi, Z., Ifju, P.G., Percival, H.F., 2010. Small Unmanned Aircraft Systems for Low-Altitude Aerial Surveys. The Journal of Wildlife Management 74, 1614–1619. https://doi.org/10.1111/j.1937-2817.2010.tb01292.x

Weiskopf, S.R., Rubenstein, M.A., Crozier, L.G., Gaichas, S., Griffis, R., Halofsky, J.E., Hyde, K.J.W., Morelli, T.L., Morisette, J.T., Muñoz, R.C., Pershing, A.J., Peterson, D.L., Poudel, R., Staudinger, M.D., Sutton-Grier, A.E., Thompson, L., Vose, J., Weltzin, J.F., Whyte, K.P., 2020. Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource

management in the United States. Science of The Total Environment 733, 137782. https://doi.org/10.1016/j.scitotenv.2020.137782

Wilschut, L.I., Heesterbeek, J.A.P., Begon, M., de Jong, S.M., Ageyev, V., Laudisoit, A., Addink, E.A., 2018. Detecting plague-host abundance from space: Using a spectral vegetation index to identify occupancy of great gerbil burrows. International Journal of Applied Earth Observation and Geoinformation 64, 249–255. https://doi.org/10.1016/j.jag.2017.09.013

Wilson, T., Crispell, J., Harris, T., 2022. Technical report: Predicting cattle camp locations in South Sudan from Sentinel 2 satellite imagery.

Witmer, G.W., 2005. Wildlife population monitoring: some practical considerations. Wildl. Res. 32, 259–263. https://doi.org/10.1071/WR04003

Wolf, P.R., Dewitt, B.A., Wilkinson, B.E., 2014. Elements of Photogrammetry with Applications in GIS, 4th Edition. ed. McGraw-Hill Education.

Wu, Y., Chen, Y., Wang, L., Ye, Y., Liu, Z., Guo, Y., Fu, Y., 2019. Large Scale Incremental Learning, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, pp. 374–382. https://doi.org/10.1109/CVPR.2019.00046

Wu, Z., Zhang, C., Gu, X., Duporge, I., Hughey, L.F., Stabach, J.A., Skidmore, A.K., Hopcraft, J.G.C., Lee, S.J., Atkinson, P.M., McCauley, D.J., Lamprey, R., Ngene, S., Wang, T., 2023. Deep learning enables satellite-based monitoring of large populations of terrestrial mammals across heterogeneous landscape. Nat Commun 14, 3072. https://doi.org/10.1038/s41467-023-38901-y

WWF, 2020. Living Planet Report 2020: Bending the Curve of Biodiversity Loss, Almond, R.E.A., Grooten M. and Petersen, T. ed. WWF, Gland, Switzerland.

Xin, Z., Chen, S., Wu, T., Shao, Y., Ding, W., You, X., 2024. Few-shot object detection: Research advances and challenges. Information Fusion 107, 102307. https://doi.org/10.1016/j.inffus.2024.102307

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., Schneider, D., 2020. Automated cattle counting using Mask R-CNN in quadcopter vision system. Computers and Electronics in Agriculture 171, 105300. https://doi.org/10.1016/j.compag.2020.105300

Xu, Z., Wang, T., Skidmore, A.K., Lamprey, R., 2024. A review of deep learning techniques for detecting animals in aerial and satellite images. International Journal of Applied Earth Observation and Geoinformation 128, 103732. https://doi.org/10.1016/j.jag.2024.103732

Xue, Y., Wang, T., Skidmore, A.K., 2017. Automatic Counting of Large Mammals from Very High Resolution Panchromatic Satellite Imagery. Remote Sensing 9, 878. https://doi.org/10.3390/rs9090878

Yang, Z., Wang, T., Skidmore, A.K., Leeuw, J. de, Said, M.Y., Freer, J., 2014. Spotting East African Mammals in Open Savannah from Space. PLOS ONE 9, e115989. https://doi.org/10.1371/journal.pone.0115989

Yu, F., Wang, D., Shelhamer, E., Darrell, T., 2018. Deep Layer Aggregation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2403–2412.

Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y., 2016. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 589–597.

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., Wen, J.-R., 2023. A Survey of Large Language Models. https://doi.org/10.48550/arXiv.2303.18223

Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X., 2019. Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems 30, 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865

Zhou, Chellappa, 1988. Computation of optical flow using a neural network, in: IEEE 1988 International Conference on Neural Networks. Presented at the IEEE 1988 International Conference on Neural Networks, pp. 71–78 vol.2. https://doi.org/10.1109/ICNN.1988.23914

Zhou, M., Elmore, J.A., Samiappan, S., Evans, K.O., Pfeiffer, M.B., Blackwell, B.F., Iglay, R.B., 2021. Improving Animal Monitoring Using Small Unmanned Aircraft Systems (sUAS) and Deep Learning Networks. Sensors 21, 5697. https://doi.org/10.3390/s21175697

Zhou, X., Wang, D., Krähenbühl, P., 2019. Objects as Points. https://doi.org/10.48550/arXiv.1904.07850

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine 5, 8–36. https://doi.org/10.1109/MGRS.2017.2762307

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object Detection in 20 Years: A Survey. Proceedings of the IEEE 111, 257–276. https://doi.org/10.1109/JPROC.2023.3238524

Annex

GISCIENCE & REMOTE SENSING
2024, VOL. 61, NO. 1, 2348863
https://doi.org/10.1080/15481603.2024.2348863

Wildlife detection, counting and survey using satellite imagery: are we there yet?

Alexandre Delplanque[a], Jérôme Théau[b,c], Samuel Foucher[b], Ghazaleh Serati[b,c], Simon Durand[b,c] and Philippe Lejeune[a]

[a]TERRA Teaching and Research Centre (Forest Is Life), ULiège, Gembloux Agro-Bio Tech, Gembloux, Belgium; [b]Departement of Applied Geomatics, Université de Sherbrooke, Sherbrooke, QC, Canada; [c]Quebec Centre for Biodiversity Science (QCBS), Stewart Biology, McGill University, Montréal, QC, Canada

ABSTRACT

Wildlife surveys are key to assessing the health of global biodiversity. Traditional field and aerial methods however have significant limitations, including high costs, substantial time investment, and potentially biased estimates. The increasing availability of high-throughput monitoring sensors in recent years has opened new perspectives for wildlife studies. Very-high-resolution (VHR) satellite sensors promise large spatial and temporal coverage while seemingly being less costly than traditional methods. Deep learning (DL) has shown increasingly impressive capabilities for processing remote sensing imagery, suggesting good prospects for imagery-based wildlife surveys. We reviewed all taxa and geographic area studies that use satellite imagery for wildlife detection, counting and surveys. Through an analysis of 49 peer-reviewed papers, this study examined the sensors and resolutions employed along with the methods used to detect, count and survey wildlife in various biomes. Results have revealed an increasing trend of publications. Mammals and birds are the focus of most of the papers, mainly in polar/alpine and pelagic ocean waters biomes. Visual interpretation is the most common method used for wildlife detection and counting while total count is mostly used for surveying. Most of the papers present a proof of concept to detect, count and survey wildlife. Technological advances are expected to enhance the spatial and temporal resolutions of satellite imagery, as well as image processing capabilities. Three main bottlenecks preventing the development of on-demand operational approaches for wildlife surveys were identified: 1) the business model of VHR satellite imagery providers is not conducive to wildlife studies; 2) satellite imagery is rarely shared; and 3) the training of multidisciplinary highly qualified personnel is underdeveloped. In response, this review presents key research priorities for advancing remote sensing for wildlife monitoring. They include wildlife-dedicated satellite constellations at enhanced spatial and temporal resolutions, increased data accessibility and sharing, adapted survey strategy, development of foundational DL model and multidisciplinary integration. We believe that progress in these directions will foster new survey strategies that are certain to revolutionize wildlife monitoring in the decades to come.

ARTICLE HISTORY
Received 21 December 2023
Accepted 24 April 2024

KEYWORDS
Wildlife; remote sensing; satellite imagery; survey; deep learning

1. Introduction and background

Biodiversity loss is one of the most significant environmental crises, threatening the survival of human civilization (Ceballos, Ehrlich, and Raven 2020). Wildlife surveys are key data for characterizing and monitoring biodiversity, but current tools and methods make it difficult to rapidly survey large areas and often provide potentially incomplete and biased estimates (Tuia et al. 2022; Turner 2014).

Most survey data are acquired using traditional field methods, which are costly and time-consuming, and present important limitations related to the accessibility of the territory and the areas covered (Davis et al. 2020; Seidlitz et al. 2021; Tuia et al. 2022). For several decades, aerial surveys have been used to survey species distributed over large areas, especially those that are not easily accessible or over rugged terrain (Davis et al. 2022; Krebs 2006). Aerial surveys are generally limited to direct visual detection (and occasional imagery) and are subject to biases associated with the subjectivity of human observation and observer disturbance, in addition to posing a significant risk of accidents (Schlossberg et al. 2016; Tuia et al. 2022). The main counting errors associated with aerial surveys are usually related to false negatives; observers often miss individuals, especially species in small groups (Lamprey et al. 2020). Although much work has been done to reduce

CONTACT Alexandre Delplanque alexandre.delplanque@uliege.be
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15481603.2024.2348863

these errors, the precision and accuracy of counts remain limited and impact the effectiveness of management actions for some populations and species (Brack et al. 2018; Davis et al. 2022).

Over the past few years, the increasing availability of numerous *in situ* sensors has opened new perspectives for wildlife surveys. Camera traps, geolocation tracking devices, drones, sound sensors, cellphones, and environmental DNA analysis are increasingly used as survey methods, creating unique opportunities for wildlife monitoring (Hughey et al. 2018; Lahoz-Monfort and Magrath 2021; Turner 2014; Whitford and Klimley 2019). However, despite this growing availability, these *in situ* devices still require intensive field effort for deployment (e.g. camera trap installation, access the territory to flight drones), have a low sampling rate, and require maintenance in varying conditions (e.g. cold, humidity, rain) that can affect their performance (Dyo et al. 2012; Newey et al. 2015). The use of these sensors can also impact the behavior of some species (Ditmer et al. 2015; Vas et al. 2015) or even their survival (Arnemo et al. 2006; LeTourneux et al. 2022). Moreover, the ratio of generated to useful data is very high and leads to a very high amount of data in multiple formats to manage and process (Lahoz-Monfort and Magrath 2021; Tuia et al. 2022).

Earth observation satellite sensors have provided images since the 1980s, with increasing spatial, temporal, and spectral resolutions. Following the advent of satellites capable of providing imagery at sub-meter resolutions (i.e. very high resolution, or VHR), several studies have focused on the use of this type of imagery for wildlife surveys, primarily to detect terrestrial and marine mammals (Hollings et al. 2018; D. Wang, Shao, and Yue 2019). Despite the tantalizing potential of these images thanks to their global terrestrial coverage, their potential for high acquisition frequency (e.g. daily) and the archiving of historical images, and their relatively low acquisition cost compared to field data acquisition, there are still limitations to wildlife detection. LaRue et al. (2017) identified 3 minimum necessary criteria for the detection of wildlife using VHR images: 1) an open landscape; 2) a sufficient body size to be detected or a positive indicator of the targeted species' presence; and 3) a contrasting color of the animal with the landscape. In addition, there are other limitations related to the availability of good quality images (e.g. cloud-free) at the targeted periods and for the targeted regions, as well as the high costs of some VHR images, especially over large territories (D. Wang, Shao, and Yue 2019).

In parallel with the development of sensors, the field of machine learning and especially deep learning (DL) has produced a stunning acceleration of massive data processing capabilities (LeCun, Bengio, and Hinton 2015). Specifically, in the field of imagery applied to Earth observation (Hoeser and Kuenzer 2020; Hoeser, Bachofer, and Kuenzer 2020; Zhao et al. 2019) and wildlife detection (Christin et al. 2019; Delplanque et al. 2022; Eikelboom et al. 2019; Kellenberger et al. 2021; Lee et al. 2021; Peng et al. 2020), approaches based on object detection using convolutional neural networks (CNNs) have the potential to automate the detection and counting of individuals with higher detection rates than conventional surveys, while significantly reducing costs and analysis time (Norouzzadeh et al. 2018; Tuia et al. 2022). Although these approaches have thus far been applied mainly on proximal (e.g. camera traps) and aerial (e.g. drones) imagery, their potential combined with the increasing availability of satellite imagery at very high spatial and temporal resolutions could represent a major advance in wildlife detection and survey techniques.

Several review papers on wildlife detection, counting or survey using remote sensing imagery have been published in the last decade (Butcher et al. 2021; Clarke et al. 2021; Corcoran et al. 2021; Delisle et al. 2023; Edney and Wood 2021; Goddijn-Murphy et al. 2021; Hollings et al. 2018; Jiménez López and Mulero-Pázmány 2019; Kuenzer et al. 2014; LaRue, Stapleton, and Anderson 2017; Linchant et al. 2015; Nazir and Kaleem 2021; Petrou, Manakos, and Stathaki 2015; Petso, Jamisola, and Mpoeleng 2021; Pettorelli et al. 2014; Sánchez-Díaz and Mata-Zayas 2019; Wang, Shao, and Yue 2019; Weinstein and Prugh 2018). However, none of them focused systematically and specifically on the use of satellite imagery, nor did any attempt to cover all taxa and geographic areas (Appendix A1). Moreover, a high number of papers have been published on these topics since the last systematic reviews were applied on papers from 2018 and earlier (40% of papers selected in the present review were published after 2018). Considering the very rapid evolution of image processing approaches combined with the increasing availability of satellite imagery at very high spatial, temporal, and spectral resolutions, a systematic and up-to-date literature review is needed.

The objectives of this paper are: (1) to provide a systematic review of existing studies that used satellite imagery to detect, count and survey animal populations; (2) to identify bottlenecks to efficient wildlife detection, counting and surveys using satellite imagery; and (3) to offer valuable perspectives and identify key research priorities for the next decade.

This review paper is organized into 5 main sections: (1) "Methods," in which we present our paper search strategy, selection criteria and definition of important terms; (2) "Results," in which we examine spatial and temporal publication trends, followed by an analysis of studied species, biomes and image processing methods; (3) "Discussion," in which we focus on and discuss sensors, resolutions, and methods employed for wildlife monitoring. It covers detection criteria, Ground Sampling Distance (GSD), spatial and temporal aspects, cost considerations and data sharing practices; (4) "Perspectives," in which we present the key research priorities identified, covering data resolution, accessibility, survey strategies, deep learning and multimodal integration; (5) "Summary and Conclusions," where we summarize and highlight the main bottlenecks and key priorities for advancing remote sensing for wildlife monitoring.

2. Methods

A comprehensive peer-reviewed paper search was performed using the Scopus database. Three concept combinations using boolean operators (AND between concepts and OR between synonyms) were defined as follows and used as keywords in the databases search: Concept 1: satellite, remote sensing, remotely sensed; Concept 2: wildlife, animal, bird, fish, mammal; Concept 3: counting, survey, detection. The preliminary list of papers was completed by reviewing the lists of references in each selected paper. The paper search was performed on works published up to September 2023.

A final selection was performed after applying the following six criteria, determined prior to the research: (1) only papers written in English were selected; (2) papers dealing with indirect counting were selected (e.g. wombat warrens, bird nests) when the objective of the study was to provide a direct relationship with population size; (3) reviews without a case study and non-peer-reviewed papers were excluded; (4) only papers focusing on non-microscopic and moving animals were selected (i.e. excluding groups of species such as corals and zooplankton); (5) papers focusing only on habitat or Global Navigation Satellite Systems (GNSS) localization of individuals were excluded; and (6) papers using only platforms other than satellite were excluded.

It is important to highlight that, for the context of this study, the terms detection, counting, and surveying have been dissociated and defined as follows:

Detection: The process of searching for and pinpointing individuals or groups of individuals belonging to a species within a satellite image. While the results may yield count values, this aspect is not performed systematically. Detection may be confined to the approximate location of a group of individuals or to a presence indicator.

Counting: The estimation of the number of individuals present within a predetermined portion of a satellite image or the entire image. If the counting method employed encompasses the entire area intended for surveying, counting may be deemed equivalent to surveying.

Surveying: The estimation of the population size of a species within the scope of its habitat or living area. Surveys may involve the utilization of spatial or temporal sampling techniques.

3. Results

The paper search and selection process yielded 49 peer-reviewed papers that employed satellite imagery for the purposes of wildlife detection, counting, or surveying (Appendix B). The results of the analysis are presented in the four following sections: 1) spatial and temporal trends observed in the selected publications; 2) studied species and biomes; 3) sensors and resolutions used; and, 4) methods used for detection, counting, and surveying.

3.1. Spatial and temporal publication trends

Most of the publications come from America (Figure 1), with 49% of the articles published. More precisely, 45% come from the United States of America (USA), followed by Europe (37%), with 22% from the United Kingdom (UK), Asia (10%), and Australia (4%).

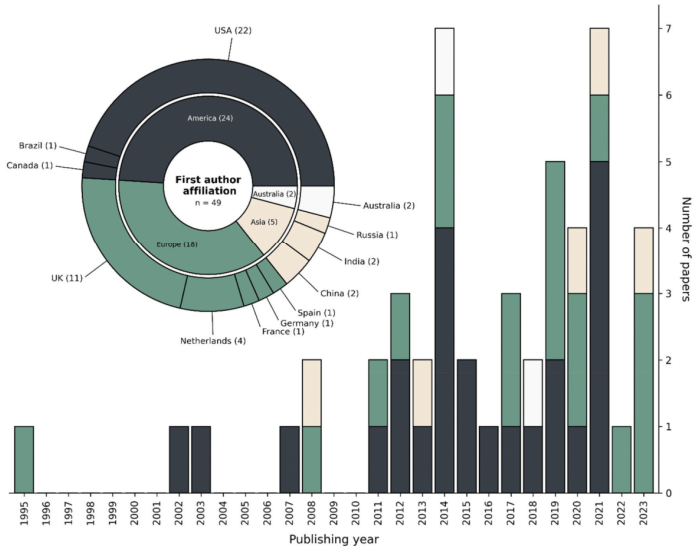


Figure 1. Historical trend of publications and overview of first author affiliation grouped by country and continent. Number of papers published are indicated in parentheses.

The temporal evolution of publications shows an increasing trend, the first being that of Guinet et al. published in 1995, in Europe. From 2011–2021, research works from America were published every year, while research from Europe was discontinuous (gaps of 1 and 2 years) before 2019, and then continuous until 2023. Australian and Asian teams published sporadically during this period. Two publication peaks occurred, in 2014 and 2021, both with 7 papers, dominated by American researchers. Only one paper was published in 2022.

3.2. Species and biomes studied

Among the 49 selected publications, two animal classes have been studied: the mammals class (Mammalia), studied in 33 papers; and the birds class (Aves), studied in 17 papers (Figure 2). More than 25 mammalian species were studied, spread into 11 families: right whales (Balaenidae), rorquals (Balaenopteridae), bovids (Bovidae), elephants (Elephantidae), equids (Equidae), hippopotamus (Hippopotamidae), monodontids (Monodontidae), earless seals (Phocidae), bears (Ursidae), squirrels (Sciuridae) and wombats (Vombatidae). Regarding birds, more than 13 species were studied, spread into 5 families: anatids (Anatidae), albatrosses (Diomedeidae), flamingos (Phoenicopteridae), penguins (Spheniscidae) and sulids (Sulidae). The papers of Guirado et al. (2019) and Kapoor et al. (2023), did not mention the species studied but only the order, which was cetaceans (Cetacea).

The species families most studied using satellite imagery were penguins (appearing in 24% of the papers), bovids (16%) and earless seals (12%), closely followed by rorquals (10%), bears and right whales (8% each). Penguins and earless seals have been mostly studied on the Antarctica coastline, making this continent the most studied to date (Figure 2). In fact, the polar/alpine (cryogenic) biome appeared in 37% of the papers, nearly equaled by the pelagic ocean waters biome (39%) which includes the sea ice functional group (Keith et al. 2022). Polar bears were studied in northern Canada and in the

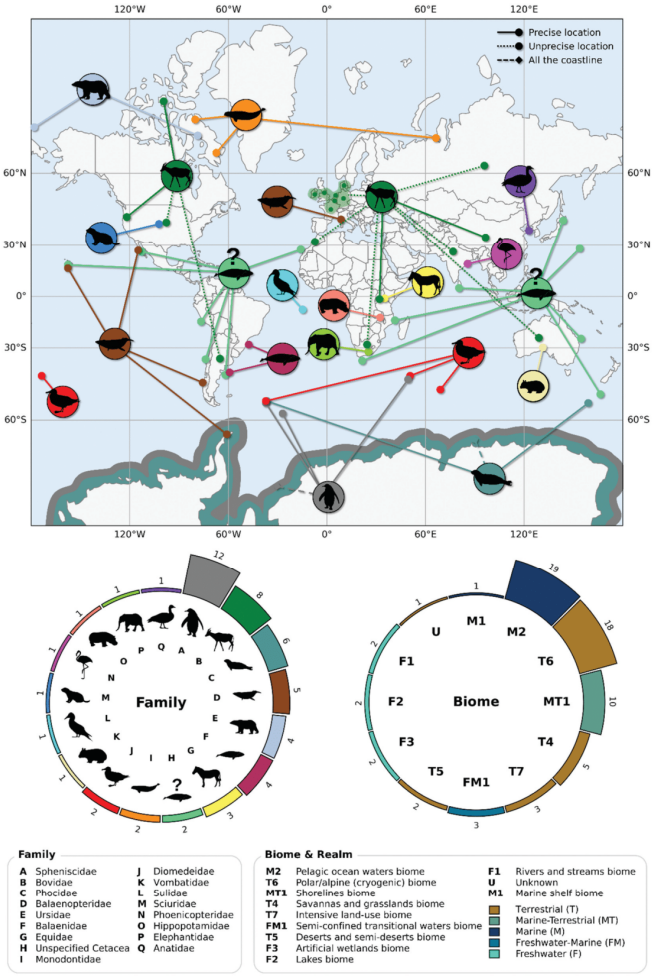


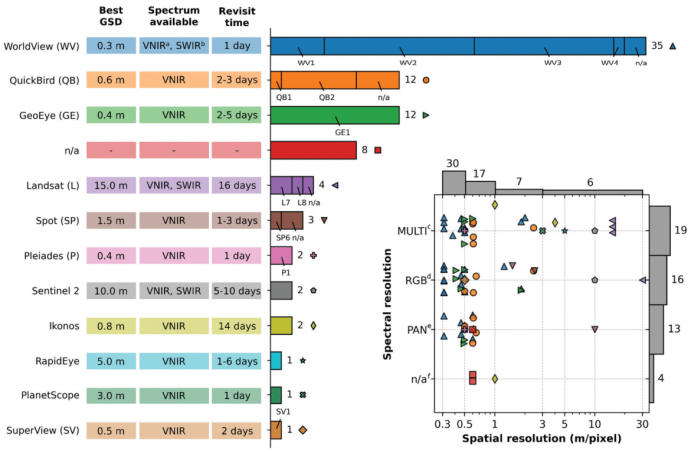
Figure 2. Spatial distribution and overview of the studied species and biomes. The location of study areas was determined using the information in the papers (i.e. geographical coordinates and/or use of the figures and places mentioned). The biomes were determined by selecting the most representative biome of each study area, using the IUCN global ecosystem typology (v2.1) maps (Keith et al. 2022). The numbers above each bar correspond to the number of published papers.

northwestern Russian Federation in these two biomes (Figure 2). The shorelines biome (20%) is linked to coastal species such as albatrosses, sulids, and certain cetacean species. Cetaceans (i.e. right whales, rorquals and monodontids) were mainly found in the marine realm, not far from coasts or islands around the world, while water birds (i.e. anatids and flamingos) were found in freshwater realms. Bovids, encompassing a wide range of terrestrial and aquatic biomes, were studied extensively due to their broad global distribution across diverse geographical regions (Figure 2).

It appeared that most of the papers (88%) focused on homogeneous and open habitats such as polar regions, waters, or shorelines leading to a generally acceptable contrast with the targeted species. Few papers studied heterogeneous landscapes (Duporge et al. 2021; Wu et al. 2023; Xue, Wang, and Skidmore 2017; Yang et al. 2014), likely due to the added complexity this poses for detection.

3.3. Sensors and resolutions

A total of 11 sensor types were employed for space-based animal detection (see Figure 3); the type used most frequently was WorldView (WV), used in 71% of the papers, followed by QuickBird (QB) and GeoEye (GE), each used in 24% of the papers. These three sensor types, alongside Pleiades (4%), possess a submeter resolution panchromatic band, which is often leveraged to enhance the resolution of other spectral bands through pan-sharpening techniques. Consequently, most of the papers examined animals at a very high resolution (<1 m/pixel), using multiple spectral bands (see Figure 3). Lower resolution sensors (>1 m/pixel) were commonly employed for detecting larger animals (e.g. cetaceans) or identifying presence indicators of specific species, for example penguin guano (Schwaller, Southwell, and Emmerson 2013), wombat warrens (Swinbourne et al. 2018) or prairie dog burrow mounds (Sidle



[a]visible and near-infrared (400 – 1100 nm); [b]short-wave infrared (1 – 2.5 µm); [c]multispectral, i.e. > 3 bands; [d]red, green and blue bands; [e]panchromatic; [f]not available

Figure 3. Overview of sensors used in the studies and the spatial and spectral resolutions of imagery used. The numbers next to or above each bar indicate the number of papers. 'n/a' means that the information was not available. Note that since some articles used several sensors with different spatial and/or spectral resolutions, the number indicated above the bars of the scatter plot does not necessarily correspond to the number of points.

et al. 2002). Finally, 8 papers (16%) did not specify the sensor used; instead, the authors mentioned the application employed to obtain satellite imagery (e.g. Google Earth), the commercial company from which the images were purchased (e.g. DigitalGlobe), or provided no information on this matter.

In terms of spectral resolution, multispectral (i.e. >3 spectral bands) images were used in 39% of the paper, followed by optical (i.e. red-green-blue bands) images (33%) and panchromatic (one-band) images (27%) (Figure 3). While most species appeared to be studied with optical and multispectral imagery, the spectral bands used rarely exceed red-green-blue and or red-green-blue and near-infrared. Beyond these bands, only Fretwell et al. (2014, 2019). have used a wider spectrum to detect whales, i.e. 9 bands including coastal bands.

Regarding the temporal aspect, close to three-quarters of the studies (76%) incorporated images from multiple dates, while single-date studies accounted for 24% of the total. Some researchers used multiple images to monitor populations over several months or years (12%), but only a few species have been monitored over time, such as Weddell seals (Ainley et al. 2015; LaRue et al. 2011), southern right whales (Corrêa et al. 2022), penguins (LaRue et al. 2014; Naveen et al. 2012) or wildebeests (Wu et al. 2023). Other authors also used multiple images to identify target species (8%) like polar bears (LaRue et al. 2015; Stapleton et al. 2014) or wildebeests and zebras (Wu et al. 2023; Xue, Wang, and Skidmore 2017) by distinguishing them using a reference image devoid of animals. However, multiple images were primarily used to allow the coverage of the entire study area (59% of the papers).

3.4. Methods used for detection, counting and surveying

The various methods used for animal detection, counting, and surveying using satellite imagery are listed and categorized by validation methods and main limitations identified by the authors in those papers (Table 1). It should be emphasized that the limits stated in Table 1 are only those put forward by the authors of the selected papers. As our aim in this section is to present the results of paper analysis, we have decided not to interpret limits that were not highlighted by the authors. All the studies performed detection, while 80% extended to counting and 45% to surveying. A total of 8 method categories were identified for detection, 5 for counting and 3 for surveying.

3.4.1. Detection

The main detection methods utilized by these studies were visual interpretation, used in 55% of the papers, supervised pixel classification (37%), supervised object detection (8%) and change detection (8%). It should be noted that authors of the selected papers mainly used change detection as a guide to facilitate manual interpretation (Stapleton et al. 2014; Wu et al. 2023; Xue, Wang, and Skidmore 2017). Only LaRue et al. (2015) have evaluated this approach as an automatic detection method. Visual interpretation seems to be a powerful method for detecting animals, especially small-sized ones (e.g. Bowler et al. 2020), but it requires experts. Such methods can only be properly validated under specific conditions, i.e. the exclusive presence of the species in a given and known location, as well as the availability of ground truth data. Supervised pixel classifiers or supervised object detection were either used for positive indicator detection (e.g. penguin guano stains, LaRue et al. 2014) or for direct animal detection (e.g. wildebeests, Wu et al. 2023). They are trained on labeled data to use spectral information from the satellite image to search for pixels or groups of pixels defining target objects (e.g. animals). In theory, a high spectral resolution provides a better discriminating power to detect specific objects, which is why pan-sharpening is commonly used to keep both spatial and spectral information. Detection methods were mostly validated using ground and/or aerial survey data or by testing them on independent imagery.

The main limitations of non-automated detection methods (i.e. visual interpretation and change detection) were the high time investment, the need for experienced interpreters and the need for reference images to distinguish animals from landscape features. Regarding automated methods, the main limitation trends were the confusion with landscape features, the difficulties in differentiating species, and the reliance on specific environmental conditions to achieve adequate performance. While multispectral instead of panchromatic imagery was recommended for better detection of wildlife (Barber-Meyer, Kooyman, and Ponganis 2007; LaRue et al. 2015), it

Table 1. Overview of methods used for detecting, counting and surveying wildlife on satellite imagery, and their description, validation methods and main limits as described in the 49 reviewed papers. Note that as visual interpretation was usually used to create ground truth, only papers using this method to produce detection, counting or survey results were considered.

Task	Method	N[a]	Description	Validation methods[b]	Main limitations[c]
Detection	Visual interpretation	27	Visual interpretation of images by one or more human interpreter(s), sometimes with the help of reference data, in order to locate animals within each image.	Comparison with ground/aerial survey data Comparison of interpretations from multiple observers	Time-consuming Requires experienced interpreters Difficulty in identifying/differentiating species Detectability relies on environmental conditions
	Supervised pixel classification	18	Assignment of each pixel to a class by a classifier, trained on a labeled dataset.	Cross-validation Manual verification Test on independent imagery	Less reliable than visual interpretation Confusion with landscape features Difficulty in identifying/differentiating species
	Supervised object detection	4	Detection, localization and classification of objects of interest (i.e. animal or group of animals) within an image by a detector, trained on a labeled dataset.	Test on independent imagery	Detectability performance relies on environmental conditions
	Change detection	4	Use of an image without animals as a reference to detect changes (i.e. potential animals) in the image of interest.	Comparison with ground/aerial survey data	Time-consuming manual review Availability of concurrent overlapping images
	Histogram thresholding	3	Selection of an optimal threshold to be applied on a histogram of pixel values to maximize the signal of animals.	Comparison with manual detection Comparison with ground survey data	Less reliable than visual interpretation for small groups Confusion with landscape features Difficulty in identifying/differentiating species Detectability relies on environmental conditions
	Supervised image classification	2	Assignment of a class to the image by a classifier, trained on a labeled dataset.	Cross-validation Test on independent imagery	Confusion with landscape features Difficulty in identifying/differentiating species Detectability relies on environmental conditions
	Unsupervised object detection	1	Detection, localization and classification of objects of interest (i.e. an animal or group of animals) within an image by a detector, trained on an unlabeled dataset.	Test on independent imagery	Confusion with landscape features Detectability relies on environmental conditions
	Unsupervised pixel classification	1	Assignment of each pixel to a class by a classifier, trained on an unlabeled dataset.	Test on independent imagery	Not mentioned in the paper.
Counting	Visual interpretation	18	Visual interpretation of images by one or more human interpreter(s), sometimes with the help of reference data, in order to locate and count animals within each image.	Comparison with ground/aerial survey data Comparison of interpretations from multiple observers	Time-consuming Requires experienced interpreters Difficulty in identifying/differentiating species Detectability relies on environmental conditions
	Detection method results	17	Use of the detection method results for counting, or at least as an aid to counting.	Validation method(s) of the corresponding detection approach (listed in the detection section).	Main limits of the corresponding detection approach (listed in the detection section).
	Regression model	10	Use of a regression model to predict estimated counts from pixels occupied by animals and ground/aerial counts.	Cross-validation Comparison with population data from previous study(ies)	Relies on precise ground truth estimates Requires concurrent ground truth data Species interactions not considered
	Density extrapolation	2	Multiplication of a known density value by a surface area to obtain an estimated count.	Not mentioned in the papers.	Requires a precise known density value
	Pixel brightness value based	1	Use of a counting model based on pixel brightness and the probability of pixel occupancy by an animal.	Comparison with ground counts	Not mentioned in the paper.

(*Continued*)

Table 1. (Continued).

Task	Method	N[a]	Description	Validation methods[b]	Main limitations[c]
Surveying	Total count	19	The entire targeted study area is surveyed.	Comparison with other survey estimates Comparison with previous estimates Use of Monte Carlo procedure	Difficulty to estimate the availability bias for aquatic animals (i.e. the ratio of submerged individuals) Difficulty to estimate the natural variability of the population Risk of overestimation due to the persistence of presence indicator Inconsistency of the imagery acquisition time with the peak activity of the target species
	Sample count	5	Part(s) of the study area is(are) surveyed and the results are then used to obtain estimates for the entire area.	Comparison with ground survey estimates Comparison with estimates from total count approach	Difficulty to guarantee the representativity of the sample
	Mark and recapture	1	Usage of each observer's detections as an independent sampling period to generate capture histories, and eventually population estimates.	Comparison with aerial survey estimates	Absolute confirmation of presumed animals is impossible

[a]Number of papers that used the associated method; [b]Validation methods used in the papers; [c]Main limitations observed by the authors of papers.

has been shown that supervised pixel classifiers struggled to differentiate animals in habitat with similar spectral signatures (Barber-Meyer, Kooyman, and Ponganis 2007; Cubaynes et al. 2019; Fretwell et al. 2019; Yang et al. 2014), were temporally inconsistent (Fretwell et al. 2014; Labrousse et al. 2022) and were prone to produce a high number of false positives (Fretwell et al. 2014, 2019; Lynch, Schwaller, and Schumann 2014). These considerations may also be valid for the histogram thresholding method, as LaRue et al. (2015) and Laliberte and Ripple (2003) observed that the surrounding landscape of terrestrial animals (polar bears and cattle, respectively) showed similar reflectance values to their bodies. Nevertheless, Fretwell et al. (2014) showed that thresholding the coastal band (400–450 nm) was the best approach to detect whales compared to unsupervised pixel classification methods. To overcome this animal-landscape spectrum similarity concern, image differencing (i.e. change detection), in which values of a reference image are subtracted from values of a target image, could be the solution. However, this method requires two orthorectified overlapping images taken at relatively close time intervals. It has thus far been shown to be effective for automatically detecting polar bears on relatively flat and open terrain (LaRue et al. 2015).

The use of deep learning is very recent, with the first paper published in 2019, and is therefore still in its infancy. To date, 10 peer-reviewed papers have applied deep learning to detect wildlife from satellite imagery, with the target species being: cetaceans (Borowicz et al. 2019; Green et al. 2023; Guirado et al. 2019; Kapoor, Kumar, and Kaushal 2023), albatrosses (Bowler et al. 2020), cattle (Mücher et al. 2022), African elephants (Duporge et al. 2021), wildebeests (Wu et al. 2023), seals (Gonçalves, Spitzbart, and Lynch 2020), and penguins (Le et al. 2022). Borowicz et al. (2019) trained a CNN-based image classifier, ResNet-152 (He et al. 2016), on down-scaled aerial image patches to discriminate the presence of whales in satellite tiles. Related to this idea, Guirado et al. (2019) trained a CNN-based image classifier, Inception-v3 (Szegedy et al. 2016), to discriminate whales from water, submerged rocks and ships, and then added a second step to locate and count individuals in the resulting tiles using Faster- R-CNN (Region-based CNN), a CNN-based object detector (Ren et al. 2017). This object detector was also used by Duporge et al. (2021) to directly detect and count African elephants on VHR satellite images. Kapoor et al. (2023) used another object detector called "Tiny YOLO (You-Only-Look-Once) v3" (Redmon and Farhadi 2018) to detect cattle and Green et al. (2023 used YOLO v5 (Jocher, Stoken, and Borovec 2021) to detect gray whales. Other works used the U-Net architecture (Ronneberger, Fischer, and Brox 2015) to detect albatrosses (Bowler et al. 2020) and wildebeests (Wu et al. 2023), or an adapted version of it to detect and count pack-ice seals (SealNet, Gonçalves, Spitzbart, and Lynch 2020) or to segment penguin colonies (PenguinNet, Le et al. 2022).

3.4.2. Counting

The most common counting methods were visual interpretation, used in 46% of papers conducting counts, the use of detection method results to estimate counts (44%) and the use of regression models (26%) generally fitted to reliable ground truth estimates. Except for the use of detection method results, counting methods were usually validated by comparing their results to ground and/or aerial counts, or to previous population data. The primary limitations of counting methods were analogous to those of detection methods. The need for precise, reliable and concurrent ground truth estimates was critical for the success of regression and extrapolation methods.

3.4.3. Surveying

Finally, total counting, which accounted for 86% of the papers conducting surveys, sample counting (23%) and mark and recapture (5%) were the three methods used for surveying. Sample counting relies on the use of sample units selected over the census area. For instance, LaRue and Stapleton (2018) used full non-overlapping satellite images and LaRue et al. (2015) used plots selected from non-overlapping satellite images. The latter assessed the sampling requirements and investigated the effect of sample plot size on polar bear population estimates on Rowley Island. Their findings suggested that sampling 50% of the study area could strike a balance between reliable results and the associated cost of using VHR satellite imagery. They also observed that plot size did not significantly impact the reliability of the results. Mark and recapture was only used by Stapleton et al. (2014) who used the counting results of two independent interpreters and treated each result as an independent sampling period to generate capture histories for mark-recapture analysis. The

abundance estimate they obtained was like the one derived from aerial surveys conducted at nearby dates.

The results of survey methods were typically validated by comparing them with other estimates or with previous estimates. Survey methods were primarily constrained by the challenge of estimating population parameters, including factors such as availability and variability, as well as ensuring the representativeness of the sample.

4. Discussion

4.1. Detection criteria

The relevance of using satellite imagery for wildlife monitoring or survey directly stems from the feasibility of detecting the target species in its surrounding habitat. To evaluate the feasibility of satellite imagery for wildlife studies, LaRue et al. (2017) established a set of eight detection criteria, categorized as primary or secondary. The primary criteria encompass three essential conditions that must be satisfied by a prospective system: 1) the presence of an open landscape; 2) a discernible color contrast between the target species and the surrounding environment; and 3) a target species possessing a detectable size or displaying positive indications of its presence. According to the authors, secondary criteria serve to enhance the utility of satellite imagery: 4) species-landscape differentiation, which entails a significant distinction between the target species' visual appearance and the surrounding landscape; 5) habitat associations, indicating the consistent presence of the species at specific locations; 6) temporal exclusivity, wherein the target species exclusively occupies an area during a specific time period; 7) coloniality, referring to the congregation of the target species in herds or groups; and 8) ground truthing, which involves the availability of accurate population data or ground validation for the detected species.

We observed that most of the selected papers reached primary criteria, while they varied among species and study areas for secondary criteria. Studies involving birds, whales or seals generally fulfilled all the criteria, while studies involving large terrestrial mammals (e.g. elephants, wildebeests) appeared to reach fewer secondary criteria, due mostly to poor temporal exclusivity, poor habitat associations and/or no ground truthing. Our trend results for the biomes studied revealed that most papers focused on homogeneous and open habitats. Recent studies have however demonstrated noteworthy levels of accuracy (approximately 80%) in detecting terrestrial mammals within heterogeneous landscapes using DL models (Duporge et al. 2021; Wu et al. 2023). Their results highlight the potential to overcome previous limitations related to heterogeneity and emphasize that continued advances in DL may further improve detection in diverse landscapes. In contrast, certain studies that satisfied most of the criteria exhibited inferior detection results because of heterogeneous coloration of animals (Fretwell et al. 2019), lack of animal-landscape differentiation (LaRue, Stapleton, and Anderson 2017), or many confusing landscape elements produced by other species or vegetation (Lynch, Schwaller, and Schumann 2014 This could be addressed using improved detection methods. These criteria are indeed predicated upon a human-centric detection paradigm, disregarding the potential processing capabilities of a computer that can effectively analyze and extract information from more extensive spatial and spectral data. Therefore, we propose that complementary characteristics, related primarily to satellite images processing and acquisition, need to be considered and are therefore addressed in the following sections.

4.2. Ground sampling distance

At spectral level, the main criterion for proper animal detection is a sufficient contrast between the target species and its surrounding landscape. At spatial level, GSD can be considered as the most critical criterion to detect animals as it is directly related to the level of details provided in imagery. Nevertheless, satellite design must deal with multiple trade-offs between spatial resolution and data volume, spectral resolution, and noise (Al-Wassai and Kalyankar 2013). Since the 2010s, there have been significant advances in spatial resolution with the launch of Worldview-1, which provides 50 cm resolution in the panchromatic band, and subsequently WV-4, which provides 30 cm resolution (Khan et al. 2023). As illustrated in Figure 4, a decimetric GSD is critical for the identification of medium-size species or individuals, particularly in high-density contexts. Several pixels are necessary to identify an animal on imagery, nine to ten pixels

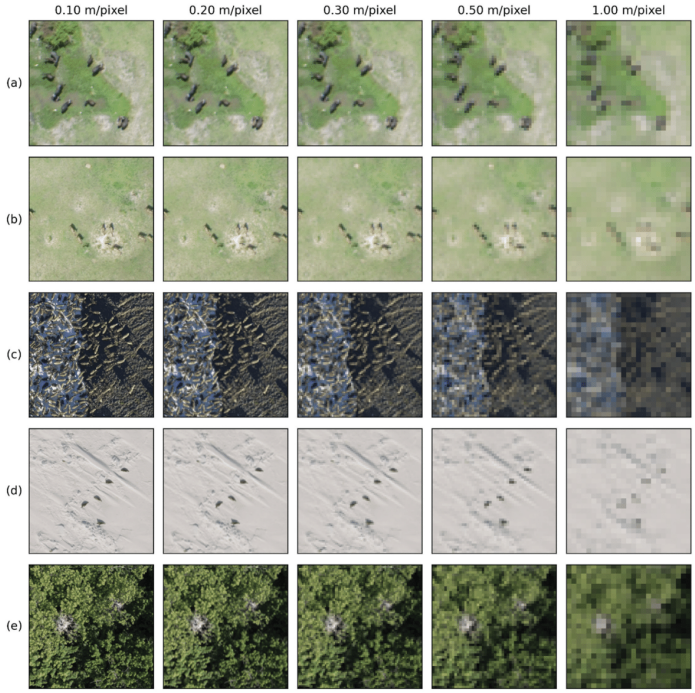


Figure 4. Spatial representation of multiple wildlife species under varying environments and ground sampling distances, simulating various satellite spatial resolutions and their impact on image clarity, species identification and high-density individual distinction: (a) African buffalo (*syncerus caffer*), (b) topi (*damaliscus lunatus jimela*), (c) caribou (*Rangifer tarandus*), (d) harp seal (*pagophilus groenlandicus*), and (e) nests of great blue heron (*Ardea Herodias*). Note that ultra-high resolution aerial images (< 5cm) were artificially down-sampled to simulate these different satellite resolutions. Images of African buffalo and topi (a, b) are samples from the dataset of Delplanque et al. (2022) with permission of the authors. Images of caribou, harp seal and great blue heron were shared by the Alaska Department of Fish and Game (c), fisheries and oceans Canada-québec (d), and the government of Quebec and CERFO (e).

being previously identified as a minimum size for detectability on visible (Wu et al. 2023) and thermal imagery (Burke et al. 2019). Considering that the highest resolution available is 30 cm/pixel, only large animals are directly detectable (e.g. whales, elephants), smaller ones being detectable using indicators of presence (e.g. guano, warrens), as shown in the reviewed papers (Figure 5). Small (i.e. under nine to ten pixels) animals also appear to be detectable, but this relies on prior knowledge of the species location, its surrounding habitat and its temporal behavior. For example, albatrosses have been detected and counted by insider personnel during their nesting period (Bowler et al. 2020), but would only appear as white spots on satellite imagery to inexperienced personnel (Figure 5a).

In addition, even if a species is detectable, differentiation between species of the same size seems difficult, if not impossible (Bamford et al. 2020; Wu et al. 2023; Yang et al. 2014). Fine-scale features (i.e. similar size of the target species) such as individual trees, small water bodies, or vegetation structure may

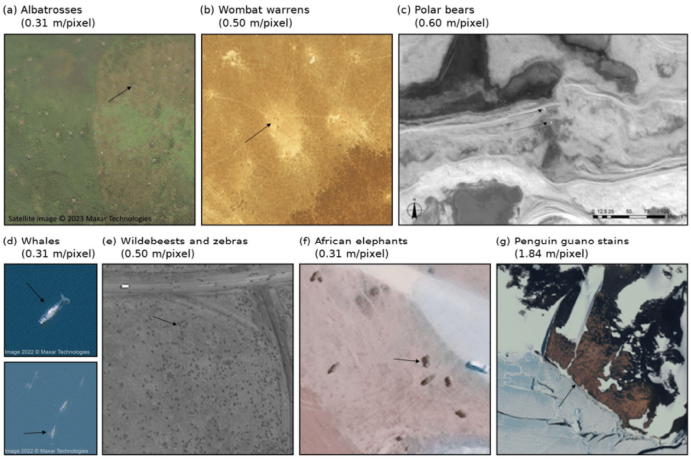


Figure 5. Examples of species studied using VHR satellite imagery: (a) Albatrosses (*diomedea Exulans*). Image from Bowler et al. (2020), printed with permission from the authors, copyright (2023), maxar technologies. (b) Wombat (*lasiorhinus latifrons*) warrens. Figure reprinted from Swinbourne et al. (2018), copyright (2018), with permission from Elsevier. (c) Polar bears (*Ursus maritimus*). Figure reprinted from LaRue et al. (2015), copyright (2015), with permission from John Wiley and sons. (d) Right whale (*Eubalaena australis*) and gray whales (*Eschrichtius robustus*). Images from Cubaynes et al. (2019), printed with permission from the authors, copyright (2022), maxar technologies. (e) Wildebeests (*connochaetes taurinus*) and zebras (*equus quagga*). Figure reprinted from Xue et al. (2017), copyright (2017), maxar technologies. (f) African elephants (*Loxodonta africana*). Figure from Duporge et al. (2021), copyright (2021), maxar technologies. (g) Penguin guano stains. Figure reprinted from Le et al. (2022), copyright (2021), with permission from John Wiley and sons.

also be difficult to discern (Fretwell et al. 2014; Irvine et al. 2019; Laliberte and Ripple 2003; McMahon et al. 2014; Xue, Wang, and Skidmore 2017), limiting the accurate assessment of population distributions. Multispectral imagery may nevertheless be helpful to distinguish species from confusing background features, or even among other species (Lynch et al. 2012). The lack of spatial resolution can impede the detection and monitoring of cryptic or elusive species that rely on camouflage or occupy densely vegetated and obstructed environments. Finally, the inability to discriminate between closely spaced individuals (Corrêa et al. 2022; Fretwell et al. 2019; Xue, Wang, and Skidmore 2017), between adults and calves (Cubaynes et al. 2019; Stapleton et al. 2014), and/or the persistence of presence indicators (Hughes, Martin, and Reynolds 2011) may also hinder the estimation of population densities and demographic parameters.

4.3. Spatial coverage

Satellite imagery covers relatively large areas compared to other types of imagery (e.g. aerial), allowing the acquisition of snapshots over very large territories. This enables animals to be counted at several spatial scales, ranging from a few tens of square kilometers to cover local areas (e.g. Bowler et al. 2020) to several thousand to cover vast territories (e.g. Cubaynes et al. 2019; Wu et al. 2023). Available VHR imagery can cover swath widths between 12 to 20 km (Khan et al. 2023), which can provide data not only on animals but also on their habitats. In these cases, although panchromatic bands can provide some radiometric and textural information, the additional use of multispectral bands (higher GSD) in visible and near-infrared domains is usually required for characterizing habitats or background (Goddijn-Murphy et al. 2021; Wang, Shao, and Yue 2019).

The spatial coverage of satellite images also gives access to any location on Earth, removing all limitations linked to the accessibility or dangerousness of a study area. This type of imagery also causes no disturbance to wildlife, and considerably reduces the deployment of field logistics. These characteristics are widely exploited in existing studies, as the vast majority of published articles focus on environments that have very low (polar/alpine (cryogenic) biome: 37%) or low (pelagic ocean water biome: 39%, shoreline biome: 20%) accessibility.

These spatial characteristics also limit certain survey biases. Wide territorial coverage means that a larger sample of a population can be surveyed compared to traditional methods, theoretically providing more accurate estimates (LaRue et al. 2015). This coverage, combined with the absence of disturbance, also limits certain detection biases linked to animal movement (e.g. double counting between flight transects). In the case of aggregating species (e.g. migratory ungulates, penguin colonies), satellite imagery opens opportunities for total count (e.g. Bamford et al. 2020; Labrousse et al. 2022, Wang et al. 2020; Wu et al. 2023). The latter would greatly increase the accuracy of population abundance estimations, currently limited by statistical constraints associated with the sampling of this type of heterogeneous and autocorrelated spatial distribution (Wu et al. 2023).

4.4. Temporal resolution

The revisit rate of the nine satellite sensors offering very high spatial resolution images currently ranges from less than 3 days to 2 times a day (Khan et al. 2023). This frequency therefore provides more than daily theoretical coverage of the Earth's surface and allows specific periods to be targeted with great precision. As wildlife surveys are often carried out during specific periods of the annual population cycle like open water season for whales (e.g. Charry et al. 2021), African ungulate migrations (e.g. Wu et al. 2023), seal breeding (e.g. Ainley et al. 2015), albatross nesting (e.g. Bowler et al. 2020), or flamingo wintering (e.g. Sasamal et al. 2008), a high revisit rate favors the availability of imagery in these time slots. Although most papers use satellite imagery at specific points in time to detect individuals or populations, some (LaRue et al. 2015; Stapleton et al. 2014; Wu et al. 2023; Xue, Wang, and Skidmore 2017) exploit this revisit rate in the detection approach itself by analyzing the temporal changes to identify individuals (i.e. moving targets) and to eliminate static confusing objects.

The continuous coverage of a territory over time also makes it possible to monitor populations over time at a relatively high frequency compared to traditional methods involving field logistics (Ainley et al. 2015, Corrêa et al. 2022; LaRue et al. 2011, 2014; Naveen et al. 2012; Wu et al. 2023). Without completely replacing traditional surveys, satellite imagery could increase their frequency while providing better reproducibility related to the objective nature of the information contained in the images and the use of automatic detection approaches. The use of traditional census approaches would remain important to validate the results obtained by image processing but could be carried out less frequently and over reduced areas (Wu et al. 2023).

However, these revisit rates remain theoretical, and several factors can influence the availability of images. For example, the shorter the targeted acquisition period, the lower the availability of imagery. As for traditional ground or aerial survey, weather conditions (e.g. sea state, cloud cover) may limit data collection during critical periods by disturbing or obscuring satellite views (Bamford et al. 2020; LaRue and Stapleton 2018; LaRue et al. 2011; Lynch et al. 2012). Finally, mosaics of scenes acquired at different dates are not adapted to wildlife surveys since wildlife is mobile.

4.5. Cost and availability of imagery

Although large web-based processing platforms such as Google Earth Engine and Microsoft Planetary Computer have democratized advanced image processing, the availability of VHR imagery remains very restricted and strongly limits the availability of images and their use for wildlife surveys. The cost of VHR satellite imagery may be a major obstacle to wildlife studies. Satellite imagery pricing varies according to several factors such as the image provider, type of demand (e.g. archiving, tasking, priority tasking), image resolution and spectral characteristics, level of processing, coverage area, licensing terms, and intended use. While it is difficult to give a precise

estimate, as they are often linked to requests for quotations, prices can range from several hundreds to thousands of dollars per minimum-size ordered scene (25 to 100 km^2), resulting in prohibitive costs for covering large areas (Apollo Mapping 2023; LAND INFO Worldwide Mapping, LLC 2023).

Current sale conditions of satellite images also significantly limit their application to wildlife surveys. Large animals, the type most targeted by satellite imagery, generally occupy large and sparsely inhabited territories in low density, and their spatial distribution varies over time while being relatively unpredictable. The availability of images must therefore respond to these constraints by covering large territories, considering that most of them do not contain animals, and by acquiring images at specific times, often determined with very short advance notice. Although less expensive than tasked images, the availability of archived images is limited because satellites usually acquire imagery only when tasked by customers. Archive catalogs do not allow for a full resolution preview which makes it impossible to evaluate the presence of animals before purchase. Archived imagery is therefore of limited interest if specific periods and sites are targeted. On the other hand, tasking imagery also has several limitations such as: high prices, a low level of priority which does not guarantee their acquisition (highest priority being given to military and commercial applications), an acquisition window of several weeks (with no control on a specific acquisition date), and a limit on coverage areas and acquisition periods.

However, these acquisition constraints are not raised in the reviewed papers. Most of them presented a proof of concept regarding species detection or population estimation using satellite imagery that requires relatively few constraints on image acquisition, given that the study usually covered a relatively small area and the site and target period can be flexible. The few articles (Ainley et al. 2015; Fretwell et al. 2012; LaRue et al. 2011; Lynch and LaRue 2014; Wu et al. 2023) that carried out operational studies (i.e. total count, temporal monitoring) focused on fixed and known study areas (e.g. Antarctic nesting sites, wildlife corridors) associated with relatively large observation windows, which favor image availability. These acquisition constraints thus remain underestimated in the literature but represent a major obstacle to the development of future operational survey tools.

4.6. Image processing

Reducing the tedious and costly workload associated with the manual interpretation of satellite imagery is a high-priority achievement that would enable larger-scale wildlife monitoring, more frequent surveys and consequently more robust population estimates (Cubaynes et al. 2019; Fretwell et al. 2014, 2019; LaRue et al. 2015). Hence, various (semi-)automated detection and counting methods have been applied on satellite imagery containing wildlife, often providing promising results, but with limitations. Several authors suggest that object-based detection methods might be more appropriate to detect and count wildlife on satellite imagery because such methods use a combination of shape, texture and spectral characteristics to detect objects (Cubaynes et al. 2019; Fretwell et al. 2019; LaRue et al. 2015; Yang et al. 2014). In recent years, these characteristics have proven to be automatically and particularly well-leveraged by CNNs, a type of artificial neural network used in various DL approaches that has demonstrated great success in the detection of objects in images (LeCun, Bengio, and Hinton 2015). While earlier object detection methods, i.e. not using DL, struggled with detecting small-sized objects, recent advances in DL have shown increasing promise for small object detection tasks (Tong, Wu, and Zhou 2020). Objects occupying just a few pixels, like animals on remote sensing imagery, may be then detected by such DL methods (e.g. Delplanque et al. 2023; Sarwar et al. 2021). DL is undoubtedly the future for all image processing tasks and would leverage the massive amount of remote sensing data, but it is still in its infancy for satellite-based wildlife monitoring. Expert-based visual interpretation may nevertheless still provide value for detecting species against complex landscapes containing numbers of confusing elements. In such cases, a hybrid approach combining DL and human experts should be more effective. DL might handle scalability by directing human attention to areas of interest and experts might verify DL model predictions and provide additional training data. The expertise of visual analysts will thus likely continue playing a role even as automated image processing techniques progress.

As described in section 3.4, different DL approaches have already been considered for the detection and counting of animals in satellite imagery, each with promising and sometimes stunning

results. Training supervised deep learning models starting from random weights requires however a large amount of labeled data, which is not easy to produce for wildlife. As an example, a common dataset used for everyday object detection is Microsoft Common Object in Context (COCO), a large-scale dataset containing more than 200,000 labeled images with 1.5 million object instances (Lin et al. 2014). Unfortunately, the size of the datasets we usually encounter is much smaller because of the significant cost and time involved for labeling. For this reason, in most surveyed papers, deep learning object detectors are typically derived from pre-trained backbones built on large computer vision training sets such as ImageNet (Deng et al. 2009). While this is a reasonable approach when training data are limited, there is evidence that spatial resolution and data preprocessing are not always appropriate for satellite imagery (Corley et al. 2023). This technique is commonly called "fine-tuning" and is widely used in various research domains. Nonetheless, one might ask how many samples are needed for the proper detection of wildlife in satellite images. This is not an obvious question, but the literature suggests some answers. Shahinfar et al. (2020) studied the effect of training sample size on the accurate classification of wildlife by CNNs in camera trap imagery. They observed that 150–500 images per class is sufficient to achieve reasonable performance when using fine-tuning. Even if it is somewhat similar, image classification differs from object detection, and we may still wonder about the minimum number of samples and annotated objects to perform satisfactory detection. Future research should clarify this aspect, but results of previous studies using deep learning for wildlife detection on satellite imagery still provide some indications. As an example, Guirado et al. (2019) reached a detection performance of 81% by using fine-tuning and 700 training samples per class, containing 945 animals. Similarly, Duporge et al. (2021) used only 188 training satellite tiles containing 1,125 animals and achieved an overall detection performance of 75% for both homogeneous and heterogeneous landscapes. Therefore, it seems that a few hundred training samples and around 1000 animal objects per class would be sufficient for the acceptable detection of wildlife by deep learning and satellite imagery.

4.7. Data sharing and multidisciplinarity

Sharing satellite imagery and annotations would certainly promote the development of automated or semi-automated detection models. Unfortunately, satellite images are often licensed by the selling companies (e.g. DigitalGlobe), which severely limits data sharing. In fact, among the 16 papers that announced the availability of their data, more than half gave the product identifier to purchase the image in the vendor's catalog, and only Yang et al. (2014) made the image used in their study freely available. As for the availability of the code for processing the satellite images, only 6 of the 49 reviewed papers made it freely available. However, as these 6 are recent (after 2018), we can hope that this will become a common practice.

In addition to data sharing, the collaboration between remote sensing and ecology communities remains an obstacle to the development of wildlife remote sensing, which mobilizes multidisciplinary expertise. For a long time, these communities evolved in silos, creating collaborative challenges linked to semantic gaps, reference frame gaps, as well as differences in needs and constraints regarding data and targeted results (Kuenzer et al. 2014; Pettorelli et al. 2014). As an indication, the first publications combining the keywords "remote sensing" and "biodiversity" date back to the early 1990s, and only 65 articles were published between 1990 and 2000 (compared to more than 200 every year recently) (Wang and Gamon 2019). This period also corresponds to the first articles on remote sensing of wildlife using satellite imagery. The recent advent of Earth observation big data combined with the development of processing approaches based on machine learning has propelled these disciplines toward each other, opening new perspectives for wildlife characterization at different spatio-temporal scales (Tuia et al. 2022).

5. Perspectives

Based on the research projects made in the last decades, we believe that future developments of wildlife detection and survey using satellite imagery will be

related to developments in 6 main axes: 1) spatial and temporal resolutions; 2) image accessibility and availability; 3) survey strategy; 4) deep learning and multimodal integration; 5) data and code sharing; and 6) training and multidisciplinarity.

5.1. Spatial and temporal resolution of images

Given that GSD is a determining factor in the detectability of animals on satellite images, the future availability of imagery at resolutions of less than 30 cm is a key factor in the widespread use of this type of data. We therefore believe that the future direction of research and technology should be toward low cost solutions such as Lower Earth Orbits (LEOs) satellites, High-Altitude Pseudo-Satellites (HAPS) or High Altitude and Long Endurance (HALE) drones. Missions such as Albedo[1] are currently underway to acquire imagery at a GSD of 10 cm (visible) and 2 m (thermal) using LEO satellite. At the same time, microsatellite technology for Very Low Earth orbits (VLEOs) between 250 and 500 km is quickly advancing. The deployment of constellations of dozens of small, low-cost satellites, each less than a meter in diameter will potentially improve the radiometric performance of optical, LIDAR and radar instruments as well as the temporal coverage. The number of constellations of micro and small satellites has greatly increased with nearly 1,000 spacecrafts in orbit forecasted for 2022 (Curzi, Modenini, and Tortora 2020). These constellations, with their high revisit rate – multiple times daily – will enable more accurate, comprehensive and timely mapping, providing a clearer understanding of conditions on the ground. Decimetric spatial resolutions could be envisioned at the cost of smaller swath widths, therefore requiring more revisiting orbits to cover a targeted area. As these spacecrafts are deployed in greater numbers and in less-traditional circular orbits, constellations can be formed that can offer more frequent revisit opportunities and thus improved temporal resolution. However, atmospheric drag will significantly reduce the sensor lifespan, which can impact data continuity.

As for HAPS and HALE drones, they can maintain a fixed position in the stratosphere (10 to 50 km), between satellites and conventional aircrafts (Guérard, Baudin, and Hertzog 2016). HAPS can be in the form of lightweight platforms such as airplanes, airships, or balloons, and are moving rapidly toward maturity, thanks to trends in solar power, battery storage, and artificial intelligence (AI). They are designed to operate at high altitudes using solar energy but have limited payloads and cannot operate well at extreme latitudes. Some notable examples of HAPS platforms that have been in development for several years include the Airbus Zephyr platform (Robinson 2022), the BAE Phasa-35 (Thisdell 2020) and the Leonardo Skydweller (Skydweller Aero Inc 2022).

In addition to sensor improvements and lower orbits, computational techniques have emerged as a powerful tool for improving spatial resolution. Specifically, push-frame satellites, such as Planet's SkySat (Murthy et al. 2014), can observe Earth's locations multiple times, creating short videos of up to 40 frames. Subsequently, multi-image super resolution techniques can be used to increase the effective spatial resolution by a factor of 2 by merging multiple observations (Nguyen et al. 2022).

5.2. Image accessibility and availability

While developments of Earth observation applications have greatly benefited from open-source satellite imagery such as the Landsat and Sentinel collections, VHR imagery availability remains very restricted for the time being. Even with precise tasking, the mere definition of acquisition parameters does not ensure the retrieval of an image that is useful in terms of the presence of animals. The next advancement in this field is likely to be smart tasking, where image acquisition is predicated on the presence of specific objects within the image. Such downstream data processing services are already offered by some satellite companies[2] and can detect objects of interest such as roads and buildings, or simply alert the user about changes between two acquisition dates. These strategies could be easily extended to other objects of interest such as the presence of animals. Upstream, at the data acquisition level, AI chipsets for edge computing continue to improve (Momose, Kaneko, and Asai 2020) and may become part of the satellite payload. This strategy will both greatly reduce the bandwidth required by high temporal frequency constellations and simplify image management and tasking. Instead of providing large volumes of raw images, satellites will directly supply high-level information streams regarding events or

objects of interest. Already, on-board processing with AI chips has facilitated the recognition of distinct features in an image, such as volcanic eruptions (Del Rosso et al. 2022). Intel has provided AI processing for PhiSat-1, guiding the onboard retrieval of cloud-free images and is currently being extended to flood event detection (Mateo-Garcia et al. 2021). With the advent of intelligent remote sensing (Zhang et al. 2022), we can foresee fleets of low-cost smart satellites dedicated to specific missions such as wildlife monitoring. The availability of such constellations providing open data dedicated exclusively to wildlife monitoring is critical, given the specific acquisition constraints associated to these targets (e.g. movement, low densities, unpredictable and large spatial distribution) which are not compatible with multi-application VHR missions such as WorldView, GeoEye, and QuickBird.

5.3. Survey strategy

In cases where species do not exhibit period-specific aggregation behavior, achieving a total count using VHR satellite imagery may not be feasible given the current limits of VHR satellites. Therefore, appropriate sampling methods need to be developed and should evolve simultaneously with advances in remote sensing imagery. For instance, sampling strategies might incorporate covariates (Meng et al. 2022), such as previous species distribution from ground or aerial surveys, or habitat suitability models (Singh et al. 2009) to identify locations of interest within the study area. At the moment, we assume that existing methods used in ecology may be applied or adapted in some cases. For instance, the sampling method of Jolly (1969), commonly employed in aerial survey standards (Craig 2012; Norton-Griffiths 1978), may be adapted to obtain population estimates over vast areas using satellite imagery. The sample units might be non-overlapping full images or plots selected from the latter (e.g. LaRue and Stapleton 2018; LaRue et al. 2015). Nevertheless, satellite imagery estimates should still be combined with ground efforts to ensure accurate assessment of population trends. This combination is necessary for image interpretation and because quantifying detection errors remains challenging (Ainley et al. 2015; Fretwell et al. 2012; LaRue and Stapleton 2018; LaRue et al. 2011; LaRue, Stapleton, and Anderson 2017; Pettorelli et al. 2018; Swinbourne et al. 2018; Wu et al. 2023). Furthermore, VHR satellite imagery may be used complementarily to identify and evaluate interesting or unsurveyed areas for future ground or aerial counts.

5.4. Deep learning and multimodal integration

Given the large amount of satellite data available and anticipated, there is an opportunity to build self-supervised sensor-specific models instead of simply fine-tuning pre-trained models as the availability of annotations becomes a bottleneck. One strategy is to adopt unsupervised or self-supervised techniques that allow neural networks to build better sensor-specific representations. Large datasets can be leveraged, reaching performances superior to pre-trained models (Tao et al., 2022). Generative techniques could also potentially help alleviate the lack of training samples by generating entirely new samples (Ramesh et al. 2022). Still, the same challenge remains, as most available generative models are also trained on massive proximal computer vision datasets (Koh, Fried, and Salakhutdinov 2024). In this regard, the development of specific generative models based on overhead imagery could be a promising research avenue.

Animals on satellite imagery can appear as a small group of pixels on satellite imagery. Given the results obtained by previous studies (Bowler et al. 2020; Wu et al. 2023), we argue that pixel-based object detection CNNs should provide the best detection performance for small-size (few pixels) animal detection in satellite imagery. Attractive point-based architectures developed on aerial images, and which provided good detection results for small animal detection, such as HerdNet (Delplanque et al. 2023), the seabird CNN detector of Kellenberger et al. (2021) or the sheep CNN detector of Sarwar et al. (2021), should be experimented with in the future.

The task of wildlife monitoring is inherently multimodal, with a wide range of possible data sources such as satellite, airborne, and drone imagery, as well as proximal data such as camera trap images, GNSS collars, in-situ microphones, and so on. Independently, so-called "Foundational Models," trained on very large and diverse training sets in a self-supervised and unsupervised way, have emerged, first in natural language processing, and also in computer vision (Bommasani et al. 2022). We believe that the future of wildlife monitoring relies on

such models and that future research should focus on this. Remarkably, these models are "few-shots" learners and can be readily applied to downstream tasks with very few training examples (Moor et al. 2023). Some of them are also multimodal and can handle speech, text, and computer vision, allowing the user to interact directly in text format. Large language models are providing an underlying structure to relate information from different sources and models (Shen et al. 2023). This capability is already being used in various domains, such as general medicine (Moor et al. 2023). These approaches could replicate what a photo interpreter would do when analyzing an image, taking into account a larger context of multimodal information. Pending the development of platforms that would handle multimodal data for training foundational models, large volumes of data from a wide range of sources are already available on online portals. These include for example Wildlife Insights[3] for camera trap images, Movebank[4] for wildlife GPS data and trajectories, AWIR[5] for aerial wildlife imagery, or BioAcoustica[6] for bioacoustic recordings. These data may also be cross-referenced with past satellite imagery acquisitions, providing ground truth for the development or validation of automatic methods. As a result, there is an opportunity for the emergence of specialized multimodal approaches that blend multi-sensor imagery, sound, and textual reports that can aid in the conduct of wildlife surveys.

5.5. Data and code sharing

Sharing data and code would further help the expansion and development of automated detection approaches. Moreover, building a large "wildlife satellite imagery" database similar to ImageNet or COCO is crucial and would lead to pre-trained CNN parameters, usable for various wildlife detection tasks. In this vein, Cubaynes and Fretwell (2022) have created an open-access dataset of satellite images containing annotated whales. This is bound to motivate other researchers to do the same in the near future. In recent years, multiple annotation tools have emerged, such as AIDE (Kellenberger et al. 2020) or Label-Studio (Tkachenko et al. 2020), and even a protocol to correctly annotate wildlife on satellite imagery (Cubaynes et al. 2023). Such tools should promote data sharing and collaborative work for future wildlife research. Pending an open database of wildlife satellite images or foundational wildlife models described in section 5.4, alternatives like using Web images (Chabot, Stapleton, and Francis 2022) or down-scaled aerial images should be developed (Borowicz et al. 2019).

5.6. Training and multidisciplinarity

The need for interdisciplinary integration has become obvious in the study and monitoring of biodiversity, as evidenced by the development of essential biodiversity variables (Jetz et al. 2019) and the development of global monitoring networks such as GeoBon,[7] which bring together scientists from a wide range of backgrounds. This interdisciplinary integration must continue and even be strengthened to accelerate the development of tools in this field, notably through the creation of open resources (e.g. best practices, data, code). This effort must also be reflected in the training of highly qualified personnel, through the development of more multidisciplinary programs combining geomatics, ecology, and computer science. This new generation of data scientists trained outside traditional disciplinary silos is certainly one of the most promising prospects for advancing knowledge in wildlife remote sensing.

6. Summary and conclusions

Satellite wildlife monitoring has emerged in recent years with the increasing availability of high- and very high-resolution satellite imagery. Several proofs of concepts have since demonstrated the potential of this new technology to detect large mammals or large bird colonies, mainly in open and homogeneous areas. To reduce the burden of manual interpretation, several automated image processing methods have been applied. The recent advent of deep learning opens important perspectives for increasing both the precision and the efficiency of image processing, while allowing multimodal data integration. New satellite acquisition platforms are being developed, anticipating the increasing availability of high spatial and temporal resolution. A revolution in wildlife monitoring techniques is therefore theoretically possible, but are we there yet?

The development of operational approaches that enable on-demand wildlife surveys and temporal monitoring is currently severely limited by three major bottlenecks: (1) The business model of VHR image providers is currently not adapted to wildlife studies;

(2) Current VHR satellite imagery is rarely shared, as it is limited by commercial license, even though it is essential for the development of robust machine learning approaches; (3) Training of multidisciplinary highly qualified personnel (geomatics, ecology, computer science) and interdisciplinary research is needed but still limited by traditional discipline-oriented training and communities. Once these bottlenecks are addressed, satellite wildlife remote sensing should enter a new era and will revolutionize wildlife monitoring.

Therefore, our key research priorities and recommendations are: (1) Wildlife-dedicated VHR satellite constellations should be developed and designed to offer freely available imagery at high spatial and temporal resolutions; (2) Sampling methods need to be developed and should evolve simultaneously with advances in remote sensing imagery and image processing methods; (3) Foundational DL models should be developed for processing data from various wildlife monitoring projects; (4) Initiatives to develop sharing and collaborative annotation platforms need to be further strengthened; (5) Initiatives to increase the number and quality of events, training, publications and funding programs dedicated to merge these disciplines should be encouraged.

Notes

1. https://albedo.com/
2. https://www.planet.com/products/analytics/
3. https://www.wildlifeinsights.org/
4. https://www.movebank.org/
5. https://projectportal.gri.msstate.edu/awir/
6. https://bio.acousti.ca/
7. https://geobon.org/

Acknowledgment

We are grateful to Laurie McLaughlin for the English revision.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Fonds de la Recherche Scientifique (FNRS) as part of Alexandre Delplanque's Fund for Research Training in Industry and Agriculture (FRIA) grant and the Natural Sciences and Engineering Research Council of Canada (NSERC) – Discovery Grants.

ORCID

Alexandre Delplanque http://orcid.org/0000-0003-3722-7076

Author contributions

Alexandre Delplanque: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original Draft, Review & Editing, Visualization. **Jérôme Théau**: Conceptualization, Methodology, Investigation, Writing – Original Draft, Review & Editing, Supervision. **Samuel Foucher**: Conceptualization, Writing – Original Draft, Review & Editing. **Ghazaleh Serati**: Investigation, Writing – Review & Editing. **Simon Durand**: Investigation, Writing – Review & Editing. **Philippe Lejeune**: Conceptualization, Writing – Review & Editing, Supervision.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

Ainley, D. G., M. A. Larue, I. Stirling, S. Stammerjohn, and D. B. Siniff. 2015. "An Apparent Population Decrease, or Change in Distribution, of Weddell Seals Along the Victoria Land Coast." *Marine Mammal Science* 31 (4): 1338–1361. https://doi.org/10.1111/mms.12220.

Al-Wassai, F. A., and N. V. Kalyankar. 2013. "Major Limitations of Satellite Images. (arXiv:1307.2434). arXiv. https://doi.org/10.48550/arXiv.1307.2434.

Apollo Mapping. 2023. "Download Imagery & DEM Price Lists– Apollo Mapping | the Image Hunters." Accessed April 15, 2023. https://apollomapping.com/download-imagery-dem-price-lists.

Arnemo, J. M., P. Ahlqvist, R. Andersen, F. Berntsen, G. Ericsson, J. Odden, S. Brunberg, P. Segerström, and J. E. Swenson. 2006. "Risk of Capture-Related Mortality in Large Free-Ranging Mammals: Experiences from Scandinavia." *Wildlife Biology* 12 (1): 109–113. https://doi.org/10.2981/0909-6396(2006)12[109:ROCMIL]2.0.CO;2.

Bamford, C. C. G., N. Kelly, L. Dalla Rosa, D. E. Cade, P. T. Fretwell, P. N. Trathan, H. C. Cubaynes, et al. 2020. "A Comparison of Baleen Whale Density Estimates Derived from Overlapping Satellite Imagery and a Shipborne Survey." *Scientific Reports* 10 (1): Article 1. https://doi.org/10.1038/s41598-020-69887-y.

Barber-Meyer, S. M., G. L. Kooyman, and P. J. Ponganis. 2007. "Estimating the Relative Abundance of Emperor Penguins at Inaccessible Colonies Using Satellite Imagery." *Polar Biology* 30 (12): 1565–1570. https://doi.org/10.1007/s00300-007-0317-8.

Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, et al. 2022. "On the Opportunities and

Risks of Foundation Models. (arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258.

Borowicz, A., H. Le, G. Humphries, G. Nehls, C. Höschle, V. Kosarev, H. J. Lynch, and P. Pławiak. 2019. "Aerial-Trained Deep Learning Networks for Surveying Cetaceans from Satellite Imagery." *Public Library of Science ONE* 14 (10): e0212532. https://doi.org/10.1371/journal.pone.0212532.

Bowler, E., P. T. Fretwell, G. French, and M. Mackiewicz. 2020. "Using Deep Learning to Count Albatrosses from Space: Assessing Results in Light of Ground Truth Uncertainty." *Remote Sensing* 12 (12): Article 12. https://doi.org/10.3390/rs12122026.

Brack, I. V., A. Kindel, L. F. B. Oliveira, and K. Scales. 2018. "Detection Errors in Wildlife Abundance Estimates from Unmanned Aerial Systems (UAS) Surveys: Synthesis, Solutions, and Challenges." *Methods in Ecology and Evolution* 9 (8): 1864–1873. https://doi.org/10.1111/2041-210X.13026.

Burke, C., M. Rashman, S. Wich, A. Symons, C. Theron, and S. Longmore. 2019. "Optimizing Observing Strategies for Monitoring Animals Using Drone-Mounted Thermal Infrared Cameras." *International Journal of Remote Sensing* 40 (2): 439–467. https://doi.org/10.1080/01431161.2018.1558372.

Butcher, P. A., A. P. Colefax, R. A. Gorkin, S. M. Kajiura, N. A. López, J. Mourier, C. R. Purcell, et al. 2021. "The Drone Revolution of Shark Science: A Review." *Drones* 5 (1): Article 1. https://doi.org/10.3390/drones5010008.

Ceballos, G., P. R. Ehrlich, and P. H. Raven. 2020. "Vertebrates on the Brink As Indicators of Biological Annihilation and the Sixth Mass Extinction." *Proceedings of the National Academy of Sciences* 117 (24): 13596–13602. https://doi.org/10.1073/pnas.1922686117.

Chabot, D., S. Stapleton, and C. M. Francis. 2022. "Using Web Images to Train a Deep Neural Network to Detect Sparsely Distributed Wildlife in Large Volumes of Remotely Sensed Imagery: A Case Study of Polar Bears on Sea Ice." *Ecological Informatics* 68:101547. https://doi.org/10.1016/j.ecoinf.2021.101547.

Charry, B., E. Tissier, J. Iacozza, M. Marcoux, and C. A. Watt. 2021. "Mapping Arctic Cetaceans from Space: A Case Study for Beluga and Narwhal." *Public Library of Science ONE* 16 (8): e0254380. https://doi.org/10.1371/journal.pone.0254380 .

Christin, S., É. Hervet, N. Lecomte, and H. Ye. 2019. "Applications for Deep Learning in Ecology." *Methods in Ecology and Evolution* 10 (10): 1632–1644. https://doi.org/10.1111/2041-210X.13256.

Clarke, P. J., H. C. Cubaynes, K. A. Stockin, C. Olavarría, A. de Vos, P. T. Fretwell, and J. A. Jackson. 2021. "Cetacean Strandings from Space: Challenges and Opportunities of Very High Resolution Satellites for the Remote Monitoring of Cetacean Mass Strandings." *Frontiers in Marine Science* 8. https://doi.org/10.3389/fmars.2021.650735.

Corcoran, E., M. Winsen, A. Sudholz, and G. Hamilton. 2021. "Automated Detection of Wildlife Using Drones: Synthesis, Opportunities and Constraints." *Methods in Ecology and Evolution* 12 (6): 1103–1114. https://doi.org/10.1111/2041-210X.13581.

Corley, I., C. Robinson, R. Dodhia, J. M. L. Ferres, and P. Najafirad. 2023. "Revisiting Pre-Trained Remote Sensing Model Benchmarks: Resizing and Normalization Matters. (arXiv:2305.13456). arXiv. https://doi.org/10.48550/arXiv.2305.13456.

Corrêa, A. A., J. H. Quoos, A. S. Barreto, K. R. Groch, and P. P. B. Eichler. 2022. "Use of Satellite Imagery to Identify Southern Right Whales (Eubalaena Australis) on a Southwest Atlantic Ocean Breeding Ground." *Marine Mammal Science* 38 (1): 87–101. https://doi.org/10.1111/mms.12847.

Craig, G. C. 2012. *Monitoring the Illegal Killing of Elephants: Aerial Survey Standards for the MIKE Programme. Version 2.0.* Nairobi, Kenya: CITES MIKE programme.

Cubaynes, H. C., P. J. Clarke, K. T. Goetz, T. Aldrich, P. T. Fretwell, K. E. Leonard, and C. B. Khan. 2023. "Annotating Very High-Resolution Satellite Imagery: A Whale Case Study." *MethodsX* 10:102040. https://doi.org/10.1016/j.mex.2023.102040.

Cubaynes, H. C., and P. T. Fretwell. 2022. "Whales from Space Dataset, an Annotated Satellite Image Dataset of Whales for Training Machine Learning Models." *Scientific Data* 9 (1): Article 1. https://doi.org/10.1038/s41597-022-01377-4.

Cubaynes, H. C., P. T. Fretwell, C. Bamford, L. Gerrish, and J. A. Jackson. 2019. "Whales from Space: Four Mysticete Species Described Using New VHR Satellite Imagery." *Marine Mammal Science* 35 (2): 466–491. https://doi.org/10.1111/mms.12544.

Curzi, G., D. Modenini, and P. Tortora. 2020. "Large Constellations of Small Satellites: A Survey of Near Future Challenges and Missions." *Aerospace* 7 (9): Article 9. https://doi.org/10.3390/aerospace7090133.

Davis, A. J., D. A. Keiter, E. M. Kierepka, C. Slootmaker, A. J. Piaggio, J. C. Beasley, and K. M. Pepin. 2020. "A Comparison of Cost and Quality of Three Methods for Estimating Density for Wild Pig (Sus Scrofa)." *Scientific Reports* 10 (1): Article 1. https://doi.org/10.1038/s41598-020-58937-0.

Davis, K. L., E. D. Silverman, A. L. Sussman, R. R. Wilson, and E. F. Zipkin. 2022. "Errors in Aerial Survey Count Data: Identifying Pitfalls and Solutions." *Ecology and Evolution* 12 (3): e8733. https://doi.org/10.1002/ece3.8733.

Delisle, Z. J., P. G. McGovern, B. G. Dillman, R. K. Swihart, T. Sankey, and G. V. Laurin. 2023. "Imperfect Detection and Wildlife Density Estimation Using Aerial Surveys with Infrared and Visible Sensors." *Remote Sensing in Ecology and Conservation* 9 (2): 222–234. https://doi.org/10.1002/rse2.305.

Delplanque, A., S. Foucher, P. Lejeune, J. Linchant, J. Théau, T. Sankey, and A. Carter. 2022. "Multispecies Detection and Identification of African Mammals in Aerial Imagery Using Convolutional Neural Networks." *Remote Sensing in Ecology and Conservation* 8 (2): 166–179. https://doi.org/10.1002/rse2.234.

Delplanque, A., S. Foucher, J. Théau, E. Bussière, C. Vermeulen, and P. Lejeune. 2023. "From Crowd to Herd Counting: How

to Precisely Detect and Count African Mammals Using Aerial Imagery and Deep Learning?" *ISPRS Journal of Photogrammetry and Remote Sensing* 197:167–180. https://doi.org/10.1016/j.isprsjprs.2023.01.025.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Ditmer, M. A., J. B. Vincent, L. K. Werden, J. C. Tanner, T. G. Laske, P. A. Iaizzo, D. L. Garshelis, and J. R. Fieberg. 2015. "Bears Show a Physiological but Limited Behavioral Response to Unmanned Aerial Vehicles." *Current Biology* 25 (17): 2278–2283. https://doi.org/10.1016/j.cub.2015.07.024.

Duporge, I., O. Isupova, S. Reece, D. W. Macdonald, T. Wang, N. Pettorelli, and G. Buchanan. 2021. "Using Very-High-Resolution Satellite Imagery and Deep Learning to Detect and Count African Elephants in Heterogeneous Landscapes." *Remote Sensing in Ecology and Conservation* 7 (3): 369–381. https://doi.org/10.1002/rse2.195.

Dyo, V., S. A. Ellwood, D. W. Macdonald, A. Markham, N. Trigoni, R. Wohlers, C. Mascolo, B. Pásztor, S. Scellato, and K. Yousef. 2012. "WILDSENSING: Design and Deployment of a Sustainable Sensor Network for Wildlife Monitoring." *ACM Transactions on Sensor Networks* 8 (4): :29:1–:29:33. https://doi.org/10.1145/2240116.2240118.

Edney, A. J., and M. J. Wood. 2021. "Applications of Digital Imaging and Analysis in Seabird Monitoring and Research." *Ibis* 163 (2): 317–337. https://doi.org/10.1111/ibi.12871.

Eikelboom, J. A. J., J. Wind, E. van de Ven, L. M. Kenana, B. Schroder, H. J. de Knegt, F. van Langevelde, and H. H. T. Prins. 2019. "Improving the Precision and Accuracy of Animal Population Estimates with Aerial Image Object Detection." *Methods in Ecology and Evolution* 10 (11): 1875–1887. https://doi.org/10.1111/2041-210X.13277.

Fretwell, P. T., J. A. Jackson, M. J. U. Encina, V. Häussermann, M. J. P. Alvarez, C. Olavarría, C. S. Gutstein, and A. Fujimura. 2019. "Using Remote Sensing to Detect Whale Strandings in Remote Areas: The Case of Sei Whales Mass Mortality in Chilean Patagonia." *Public Library of Science ONE* 14 (10): e0222498. https://doi.org/10.1371/journal.pone.0222498.

Fretwell, P. T., M. A. LaRue, P. Morin, G. L. Kooyman, B. Wienecke, N. Ratcliffe, A. J. Fox, et al. 2012. "An Emperor Penguin Population Estimate: The First Global, Synoptic Survey of a Species from Space." *Public Library of Science ONE* 7 (4): e33751. https://doi.org/10.1371/journal.pone.0033751.

Fretwell, P. T., I. J. Staniland, J. Forcada, and T. Gilbert. 2014. "Whales from Space: Counting Southern Right Whales by Satellite." *Public Library of Science ONE* 9 (2): e88655. https://doi.org/10.1371/journal.pone.0088655.

Goddijn-Murphy, L., N. J. O'Hanlon, N. A. James, E. A. Masden, and A. L. Bond. 2021. "Earth Observation Data for Seabirds and Their Habitats: An Introduction." *Remote Sensing Applications: Society & Environment* 24:100619. https://doi.org/10.1016/j.rsase.2021.100619.

Gonçalves, B. C., B. Spitzbart, and H. J. Lynch. 2020. "SealNet: A Fully-Automated Pack-Ice Seal Detection Pipeline for Sub-Meter Satellite Imagery." *Remote Sensing of Environment* 239:111617. https://doi.org/10.1016/j.rse.2019.111617.

Green, K. M., M. K. Virdee, H. C. Cubaynes, A. I. Aviles-Rivero, P. T. Fretwell, P. C. Gray, D. W. Johnston, C.-B. Schönlieb, L. G. Torres, and J. A. Jackson. 2023. "Gray Whale Detection in Satellite Imagery Using Deep Learning." *Remote Sensing in Ecology and Conservation* 9 (6): 829–840. https://doi.org/10.1002/rse2.352.

Guérard, J., F. Baudin, and A. Hertzog. 2016 May. "High Altitude Drones for Science. Near Space in the Near Future." *SONDRA 4th Workshop*. https://hal.science/hal-01993992.

Guinet, C., P. Jouventin, and J. Malacamp. 1995. "Satellite Remote Sensing in Monitoring Change of Seabirds: Use of Spot Image in King Penguin Population Increase at Ile Aux Cochons, Crozet Archipelago." *Polar Biology* 15 (7): 511–515. https://doi.org/10.1007/BF00237465.

Guirado, E., S. Tabik, M. L. Rivas, D. Alcaraz-Segura, and F. Herrera. 2019. "Whale Counting in Satellite and Aerial Images with Deep Learning." *Scientific Reports* 9 (1): Article 1. https://doi.org/10.1038/s41598-019-50795-9.

He, K., X. Zhang, S. Ren, and J. Sun 2016. "Deep Residual Learning for Image Recognition." *Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hoeser, T., F. Bachofer, and C. Kuenzer. 2020. "Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications." *Remote Sensing* 12 (18): Article 18. https://doi.org/10.3390/rs12183053.

Hoeser, T., and C. Kuenzer. 2020. "Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends." *Remote Sensing* 12 (10): Article 10. https://doi.org/10.3390/rs12101667.

Hollings, T., M. Burgman, M. van Andel, M. Gilbert, T. Robinson, A. Robinson, and J. McPherson. 2018. "How Do You Find the Green Sheep? A Critical Review of the Use of Remotely Sensed Imagery to Detect and Count Animals." *Methods in Ecology and Evolution* 9 (4): 881–892. https://doi.org/10.1111/2041-210X.12973.

Hughes, B. J., G. R. Martin, and S. J. Reynolds. 2011. "The Use of Google EarthTM Satellite Imagery to Detect the Nests of Masked Boobies Sula Dactylatra." *Wildlife Biology* 17 (2): 210–216. https://doi.org/10.2981/10-106.

Hughey, L. F., A. M. Hein, A. Strandburg-Peshkin, and F. H. Jensen. 2018. "Challenges and Solutions for Studying Collective Animal Behaviour in the Wild." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373 (1746): 20170005. https://doi.org/10.1098/rstb.2017.0005.

Irvine, J. M., J. Nolan, N. Hofmann, D. Lewis, T. Simpamba, P. Zyambo, A. J. Travis, and S. Hemami. 2019. "Estimating the Population of Large Animals in the Wild Using Satellite Imagery: A Case Study of Hippos in Zambia's Luangwa

River." *2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 1–8. https://doi.org/10.1109/AIPR47015.2019.9174564.

Jetz, W., M. A. McGeoch, R. Guralnick, S. Ferrier, J. Beck, M. J. Costello, M. Fernandez, et al. 2019. "Essential Biodiversity Variables for Mapping and Monitoring Species Populations." *Nature Ecology & Evolution* 3 (4): Article 4. https://doi.org/10.1038/s41559-019-0826-1.

Jiménez López, J., and M. Mulero-Pázmány. 2019. "Drones for Conservation in Protected Areas: Present and Future." *Drones* 3 (1): Article 1. https://doi.org/10.3390/drones3010010.

Jocher, G., A. Stoken, and J. Borovec. 2021. *Ultralytics/yolov5: V5.0 - YOLOv5-P6 1280 Models, AWS, Supervise.Ly and YouTube Integrations [Computer Software]*. Zenodo. https://doi.org/10.5281/zenodo.4679653.

Jolly, G. M. 1969. "Sampling Methods for Aerial Censuses of Wildlife Populations." *East African Agricultural and Forestry Journal* 34 (sup1): 46–49. https://doi.org/10.1080/00128325.1969.11662347.

Kapoor, S., M. Kumar, and M. Kaushal. 2023. "Deep Learning Based Whale Detection from Satellite Imagery." *Sustainable Computing: Informatics and Systems* 38:100858. https://doi.org/10.1016/j.suscom.2023.100858.

Keith, D. A., J. R. Ferrer-Paris, E. Nicholson, M. J. Bishop, B. A. Polidoro, E. Ramirez-Llodra, M. G. Tozer, et al. 2022. "A Function-Based Typology for Earth's Ecosystems." *Nature* 610 (7932): Article 7932. https://doi.org/10.1038/s41586-022-05318-4.

Kellenberger, B., D. Tuia, D. Morris, and L. Graham. 2020. "AIDE: Accelerating Image-Based Ecological Surveys with Interactive Machine Learning." *Methods in Ecology and Evolution* 11 (12): 1716–1727. https://doi.org/10.1111/2041-210X.13489.

Kellenberger, B., T. Veen, E. Folmer, D. Tuia, N. Horning, and K. Scales. 2021. "21 000 Birds in 4.5 H: Efficient Large-Scale Seabird Detection with Machine Learning." *Remote Sensing in Ecology and Conservation* 7 (3): 445–460. https://doi.org/10.1002/rse2.200.

Khan, C. B., K. T. Goetz, H. C. Cubaynes, C. Robinson, E. Murnane, T. Aldrich, M. Sackett, et al. 2023. "A Biologist's Guide to the Galaxy: Leveraging Artificial Intelligence and Very High-Resolution Satellite Imagery to Monitor Marine Mammals from Space." *Journal of Marine Science and Engineering* 11 (3): Article 3. https://doi.org/10.3390/jmse11030595.

Koh, J. Y., D. Fried, and R. R. Salakhutdinov. 2024. "Generating Images with Multimodal Language Models." *Advances in Neural Information Processing Systems* 36: 21487–21506.

Krebs, C. J. 2006. "Mammals." In *Ecological Census Techniques: A Handbook*, edited by W. J. Sutherland, 351–369. 2nd ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511790508.011.

Kuenzer, C., M. Ottinger, M. Wegmann, H. Guo, C. Wang, J. Zhang, S. Dech, and M. Wikelski. 2014. "Earth Observation Satellite Sensors for Biodiversity Monitoring: Potentials and Bottlenecks." *International Journal of Remote Sensing* 35 (18): 6599–6647. https://doi.org/10.1080/01431161.2014.964349.

Labrousse, S., D. Iles, L. Viollat, P. Fretwell, P. N. Trathan, D. P. Zitterbart, S. Jenouvrier, M. LaRue, N. Pettorelli, and T. Kuemmerle. 2022. "Quantifying the Causes and Consequences of Variation in Satellite-Derived Population Indices: A Case Study of Emperor Penguins." *Remote Sensing in Ecology and Conservation* 8 (2): 151–165. https://doi.org/10.1002/rse2.233.

Lahoz-Monfort, J. J., and M. J. L. Magrath. 2021. "A Comprehensive Overview of Technologies for Species and Habitat Monitoring and Conservation." *BioScience* 71 (10): 1038–1062. https://doi.org/10.1093/biosci/biab073.

Laliberte, A. S., and W. J. Ripple. 2003. "Automated Wildlife Counts from Remotely Sensed Imagery." *Wildlife Society Bulletin* 31 (2): 362–371.

Lamprey, R., F. Pope, S. Ngene, M. Norton-Griffiths, H. Frederick, B. Okita-Ouma, and I. Douglas-Hamilton. 2020. "Comparing an Automated High-Definition Oblique Camera System to Rear-Seat-Observers in a Wildlife Survey in Tsavo, Kenya: Taking Multi-Species Aerial Counts to the Next Level." *Biological Conservation* 241:108243. https://doi.org/10.1016/j.biocon.2019.108243.

LAND INFO Worldwide Mapping, LLC. 2023. "Buying Satellite Imagery: Pricing Information for High Resolution Satellite Imagery." LLC: LAND INFO Worldwide Mapping. Accessed April 15, 2023. https://landinfo.com/satellite-imagery-pricing/.

LaRue, M. A., H. J. Lynch, P. O. B. Lyver, K. Barton, D. G. Ainley, A. Pollard, W. R. Fraser, and G. Ballard. 2014. "A Method for Estimating Colony Sizes of Adélie Penguins Using Remote Sensing Imagery." *Polar Biology* 37 (4): 507–517. https://doi.org/10.1007/s00300-014-1451-8.

LaRue, M. A., J. J. Rotella, R. A. Garrott, D. B. Siniff, D. G. Ainley, G. E. Stauffer, C. C. Porter, and P. J. Morin. 2011. "Satellite Imagery Can Be Used to Detect Variation in Abundance of Weddell Seals (Leptonychotes Weddellii) in Erebus Bay, Antarctica." *Polar Biology* 34 (11): 1727–1737. https://doi.org/10.1007/s00300-011-1023-0.

LaRue, M. A., and S. Stapleton. 2018. "Estimating the Abundance of Polar Bears on Wrangel Island During Late Summer Using High-Resolution Satellite Imagery: A Pilot Study." *Polar Biology* 41 (12): 2621–2626. https://doi.org/10.1007/s00300-018-2384-4.

LaRue, M. A., S. Stapleton, and M. Anderson. 2017. "Feasibility of Using High-Resolution Satellite Imagery to Assess Vertebrate Wildlife Populations." *Conservation Biology* 31 (1): 213–220. https://doi.org/10.1111/cobi.12809.

LaRue, M. A., S. Stapleton, C. Porter, S. Atkinson, T. Atwood, M. Dyck, and N. Lecomte. 2015. "Testing Methods for Using High-Resolution Satellite Imagery to Monitor Polar Bear Abundance and Distribution." *Wildlife Society Bulletin* 39 (4): 772–779. https://doi.org/10.1002/wsb.596.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521 (7553): Article 7553. https://doi.org/10.1038/nature14539.

Lee, P. Q., K. Radhakrishnan, D. A. Clausi, K. A. Scott, L. Xu, and M. Marcoux. 2021. "Beluga Whale Detection in the Cumberland Sound Bay Using Convolutional Neural Networks." *Canadian Journal of Remote Sensing* 47 (2): 276–294. https://doi.org/10.1080/07038992.2021.1901221.

Le, H., D. Samaras, H. J. Lynch, N. Pettorelli, and T. Kuemmerle. 2022. "A Convolutional Neural Network Architecture Designed for the Automated Survey of Seabird Colonies." *Remote Sensing in Ecology and Conservation* 8 (2): 251–262. https://doi.org/10.1002/rse2.240.

LeTourneux, F., G. Gauthier, R. Pradel, J. Lefebvre, and P. Legagneux. 2022. "Evidence for Synergistic Cumulative Impacts of Marking and Hunting in a Wildlife Species." *Journal of Applied Ecology* 59 (11): 2705–2715. https://doi.org/10.1111/1365-2664.14268.

Linchant, J., J. Lisein, J. Semeki, P. Lejeune, and C. Vermeulen. 2015. "Are Unmanned Aircraft Systems (UASs) the Future of Wildlife Monitoring? A Review of Accomplishments and Challenges." *Mammal Review* 45 (4): 239–252. https://doi.org/10.1111/mam.12046.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. 2014. "Microsoft COCO: Common Objects in Context." In *Computer Vision – ECCV 2014*, edited by D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, 740–755. Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48.

Lynch, H. J., and M. A. LaRue. 2014. "First Global Census of the Adélie Penguin." *The Auk* 131 (4): 457–466. https://doi.org/10.1642/AUK-14-31.1.

Lynch, H. J., M. R. Schwaller, and G. J.-P. Schumann. 2014. "Mapping the Abundance and Distribution of Adélie Penguins Using Landsat-7: First Steps Towards an Integrated Multi-Sensor Pipeline for Tracking Populations at the Continental Scale." *Public Library of Science ONE* 9 (11): e113301. https://doi.org/10.1371/journal.pone.0113301.

Lynch, H. J., R. White, A. D. Black, and R. Naveen. 2012. "Detection, Differentiation, and Abundance Estimation of Penguin Species by High-Resolution Satellite Imagery." *Polar Biology* 35 (6): 963–968. https://doi.org/10.1007/s00300-011-1138-3.

Mateo-Garcia, G., J. Veitch-Michaelis, L. Smith, S. V. Oprea, G. Schumann, Y. Gal, A. G. Baydin, and D. Backes. 2021. "Towards Global Flood Mapping Onboard Low Cost Satellites with Machine Learning." *Scientific Reports* 11 (1): Article 1. https://doi.org/10.1038/s41598-021-86650-z.

McMahon, C. R., H. Howe, J. V. D. Hoff, R. Alderman, H. Brolsma, M. A. Hindell, and Y. Ropert-Coudert. 2014. "Satellites, the All-Seeing Eyes in the Sky: Counting Elephant Seals from Space." *Public Library of Science ONE* 9 (3): e92613. https://doi.org/10.1371/journal.pone.0092613.

Meng, C., E. Liu, W. Neiswanger, J. Song, M. Burke, D. Lobell, and S. Ermon. 2022. "IS-Count: Large-Scale Object Counting from Satellite Images with Covariate-Based Importance Sampling." *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (11): Article 11. https://doi.org/10.1609/aaai.v36i11.21462.

Momose, H., T. Kaneko, and T. Asai. 2020. "Systems and Circuits for AI Chips and Their Trends." *Japanese Journal of Applied Physics* 59 (5): 050502. https://doi.org/10.35848/1347-4065/ab839f.

Moor, M., O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar. 2023. "Foundation Models for Generalist Medical Artificial Intelligence." *Nature* 616 (7956): Article 7956. https://doi.org/10.1038/s41586-023-05881-4.

Mücher, C. A., S. Los, G. J. Franke, and C. Kamphuis. 2022. "Detection, Identification and Posture Recognition of Cattle with Satellites, Aerial Photography and UAVs Using Deep Learning Techniques." *International Journal of Remote Sensing* 43 (7): 2377–2392. https://doi.org/10.1080/01431161.2022.2051634.

Murthy, K., M. Shearn, B. D. Smiley, A. H. Chau, J. Levine, and M. D. Robinson. 2014. "SkySat-1: Very High-Resolution Imagery from a Small Satellite." *Sensors, Systems, and Next-Generation Satellites XVIII* 9241:367–378. https://doi.org/10.1117/12.2074163.

Naveen, R., H. J. Lynch, S. Forrest, T. Mueller, and M. Polito. 2012. "First Direct, Site-Wide Penguin Survey at Deception Island, Antarctica, Suggests Significant Declines in Breeding Chinstrap Penguins." *Polar Biology* 35 (12): 1879–1888. https://doi.org/10.1007/s00300-012-1230-3.

Nazir, S., and M. Kaleem. 2021. "Advances in Image Acquisition and Processing Technologies Transforming Animal Ecological Studies." *Ecological Informatics* 61:101212. https://doi.org/10.1016/j.ecoinf.2021.101212.

Newey, S., P. Davidson, S. Nazir, G. Fairhurst, F. Verdicchio, R. J. Irvine, and R. van der Wal. 2015. "Limitations of Recreational Camera Traps for Wildlife Management and Conservation Research: A practitioner's Perspective." *AMBIO: A Journal of the Human Environment* 44 (4): 624–635. https://doi.org/10.1007/s13280-015-0713-1.

Nguyen, N. L., J. Anger, A. Davy, P. Arias, and G. Facciolo. 2022. "Self-Supervised Super-Resolution for Multi-Exposure Push-Frame Satellites." *Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1848–1858. https://doi.org/10.1109/CVPR52688.2022.00190.

Norouzzadeh, M. S., A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune. 2018. "Automatically Identifying, Counting, and Describing Wild Animals in Camera-Trap Images with Deep Learning." *Proceedings of the National Academy of Sciences* 115 (25): E5716–E5725. https://doi.org/10.1073/pnas.1719367115.

Norton-Griffiths, M. 1978. *Counting Animals (J.J.R Grimsdell)*. Nairobi, Kenya: African Wildlife Leadership Foundation.

Peng, J., D. Wang, X. Liao, Q. Shao, Z. Sun, H. Yue, and H. Ye. 2020. "Wild Animal Survey Using UAS Imagery and Deep Learning: Modified Faster R-CNN for Kiang Detection in Tibetan Plateau." *ISPRS Journal of Photogrammetry and Remote Sensing* 169:364–376. https://doi.org/10.1016/j.isprsjprs.2020.08.026.

Petrou, Z. I., I. Manakos, and T. Stathaki. 2015. "Remote Sensing for Biodiversity Monitoring: A Review of Methods for Biodiversity Indicator Extraction and Assessment of

Progress Towards International Targets." *Biodiversity and Conservation* 24 (10): 2333–2363. https://doi.org/10.1007/s10531-015-0947-z.

Petso, T., R. S. Jamisola, and D. Mpoeleng. 2021. "Review on Methods Used for Wildlife Species and Individual Identification." *European Journal of Wildlife Research* 68 (1): 3. https://doi.org/10.1007/s10344-021-01549-4.

Pettorelli, N., W. F. Laurance, T. G. O'Brien, M. Wegmann, H. Nagendra, W. Turner, and E. J. Milner-Gulland. 2014. "Satellite Remote Sensing for Applied Ecologists: Opportunities and Challenges." *Journal of Applied Ecology* 51 (4): 839–848. https://doi.org/10.1111/1365-2664.12261.

Pettorelli, N., H. Schulte to Bühne, A. Tulloch, G. Dubois, C. Macinnis-Ng, A. M. Queirós, et al. 2018. "Satellite Remote Sensing of Ecosystem Functions: Opportunities, Challenges and Way Forward." *Remote Sensing in Ecology and Conservation* 4 (2): 71–93. https://doi.org/10.1002/rse2.59.

Ramesh, A., P. Dhariwal, A. Nichol, C. Chu, and M. Chen. 2022. "Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv preprint arXiv:2204.06125 https://arxiv.org/abs/2204.06125.

Redmon, J., and A. Farhadi. 2018. *YOLOv3: An Incremental Improvement* (arXiv:1804.02767). arXiv. https://doi.org/10.48550/arXiv.1804.02767.

Ren, S., K. He, R. Girshick, and J. Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39: 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031 .

Robinson, T. 2022. "Zephyr – Down but Definitely Not Out." Royal Aeronautical Society. Accessed April 15, 2023. https://www.aerosociety.com/news/zephyr-down-but-definitely-not-out/.

Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, edited by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, 234–241. Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.

Rosso, M. P. D., A. Sebastianelli, D. Spiller, and S. L. Ullo. 2022. "A Demo Setup Testing Onboard CNNs for Volcanic Eruption Detection." *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 719–724. https://doi.org/10.1109/MetroXRAINE54828.2022.9967684.

Sánchez-Díaz, B., and E. E. Mata-Zayas. 2019. "Remote Sensing As Indispensable Technology in Ecology to Support the Protection of Biodiversity: A Review." *International Journal of Conservation Science* 10 (4): 811–820.

Sarwar, F., A. Griffin, S. U. Rehman, and T. Pasang. 2021. "Detecting Sheep in UAV Images." *Computers and Electronics in Agriculture* 187:106219. https://doi.org/10.1016/j.compag.2021.106219.

Sasamal, S. K., S. B. Chaudhury, R. N. Samal, and A. K. Pattanaik. 2008. "QuickBird Spots Flamingos off Nalabana Island, Chilika Lake, India." *International Journal of Remote Sensing* 29 (16): 4865–4870. https://doi.org/10.1080/01431160701814336.

Schlossberg, S., M. J. Chase, C. R. Griffin, and A. L. Roca. 2016. "Testing the Accuracy of Aerial Surveys for Large Mammals: An Experiment with African Savanna Elephants (Loxodonta Africana)." *Public Library of Science ONE* 11 (10): e0164904. https://doi.org/10.1371/journal.pone.0164904.

Schwaller, M. R., C. J. Southwell, and L. M. Emmerson. 2013. "Continental-Scale Mapping of Adélie Penguin Colonies from Landsat Imagery." *Remote Sensing of Environment* 139:353–364. https://doi.org/10.1016/j.rse.2013.08.009.

Seidlitz, A., A. F. Wayne, M. C. Calver, N. J. Armstrong, K. A. Bryant, K. A. Bryant, N. J. Armstrong, M. C. Calver, A. F. Wayne, and A. Seidlitz. 2021. "Sign Surveys Can Be More Efficient and Cost Effective Than Driven Transects and Camera Trapping: A Comparison of Detection Methods for a Small Elusive Mammal, the Numbat (Myrmecobius Fasciatus)." *Wildlife Research* 48 (6): 491–500. https://doi.org/10.1071/WR20020.

Shahinfar, S., P. Meek, and G. Falzon. 2020. ""How Many Images Do I need?" Understanding How Sample Size per Class Affects Deep Learning Model Performance Metrics for Balanced Designs in Autonomous Wildlife Monitoring." *Ecological Informatics* 57:101085. https://doi.org/10.1016/j.ecoinf.2020.101085.

Shen, Y., K. Song, X. Tan, D. Li, W. Lu, and Y. Zhuang. 2023. "HuggingGPT: Solving AI Tasks with ChatGPT and Its Friends in Hugging Face (arXiv:2303.17580). arXiv. https://doi.org/10.48550/arXiv.2303.17580.

Sidle, J. G., D. H. Johnson, B. R. Euliss, and M. Tooze. 2002. "Monitoring Black-Tailed Prairie Dog Colonies with High-Resolution Satellite Imagery." *Wildlife Society Bulletin* 30 (2): 405–411.

Singh, N. J., N. G. Yoccoz, Y. V. Bhatnagar, and J. L. Fox. 2009. "Using Habitat Suitability Models to Sample Rare Species in High-Altitude Ecosystems: A Case Study with Tibetan Argali." *Biodiversity and Conservation* 18 (11): 2893–2908. https://doi.org/10.1007/s10531-009-9615-5.

Skydweller Aero Inc. 2022. "Luxembourg's Directorate of Defence, Skydweller Aero and Leonardo Announce Collaboration Agreement to Support Flight Test Programme for Ultra-Persistent, Solar-Powered, Unmanned Aerial Platform." June 14. Accessed May 25, 2023. http://gouvernement.lu/en/actualites/toutes_actualites/communiques/2022/06-juin/14-bausch-uas.html.

Stapleton, S., M. LaRue, N. Lecomte, S. Atkinson, D. Garshelis, C. Porter, T. Atwood, and Y. Ropert-Coudert. 2014. "Polar Bears from Space: Assessing Satellite Imagery as a Tool to Track Arctic Wildlife." *Public Library of Science ONE* 9 (7): e101513. https://doi.org/10.1371/journal.pone.0101513.

Swinbourne, M. J., D. A. Taggart, A. M. Swinbourne, M. Lewis, and B. Ostendorf. 2018. "Using Satellite Imagery to Assess the Distribution and Abundance of Southern Hairy-Nosed Wombats (Lasiorhinus Latifrons)." *Remote Sensing of Environment* 211:196–203. https://doi.org/10.1016/j.rse.2018.04.017.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2016. "Rethinking the Inception Architecture for Computer

Vision." *Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, USA, 2818–2826. https://doi.org/10.1109/CVPR.2016.308.

Tao, C., J. Qi, W. Lu, H. Wang, and H. Li. 2022. "Remote Sensing Image Scene Classification with Self-Supervised Paradigm Under Limited Labeled Samples." *IEEE Geoscience and Remote Sensing Letters* 19:1–5. https://doi.org/10.1109/LGRS.2020.3038420.

Thisdell, D. 2020. "BAE Joins High-Altitude Race with Maiden PHASA-35 Flight." Flight Global. Accessed June 21, 2023. https://www.flightglobal.com/aerospace/bae-joins-high-altitude-race-with-maiden-phasa-35-flight/136767.article .

Tkachenko, M., M. Malyuk, A. Holmanyuk, and N. Liubimov. 2020. "Label Studio: Data Labeling Software." *Computer Software*. https://github.com/heartexlabs/label-studio.

Tong, K., Y. Wu, and F. Zhou. 2020. "Recent Advances in Small Object Detection Based on Deep Learning: A Review." *Image and Vision Computing* 97:103910. https://doi.org/10.1016/j.imavis.2020.103910.

Tuia, D., B. Kellenberger, S. Beery, B. R. Costelloe, S. Zuffi, B. Risse, A. Mathis, et al. 2022. "Perspectives in Machine Learning for Wildlife Conservation." *Nature Communications* 13 (1): Article 1. https://doi.org/10.1038/s41467-022-27980-y.

Turner, W. 2014. "Sensing Biodiversity." *Science* 346 (6207): 301–302. https://doi.org/10.1126/science.1256014.

Vas, E., A. Lescroël, O. Duriez, G. Boguszewski, and D. Grémillet. 2015. "Approaching Birds with Drones: First Experiments and Ethical Guidelines." *Biology Letters* 11 (2): 20140754. https://doi.org/10.1098/rsbl.2014.0754.

Wang, R., and J. A. Gamon. 2019. "Remote Sensing of Terrestrial Plant Biodiversity." *Remote Sensing of Environment* 231:111218. https://doi.org/10.1016/j.rse.2019.111218.

Wang, D., Q. Shao, and H. Yue. 2019. "Surveying Wild Animals from Satellites, Manned Aircraft and Unmanned Aerial Systems (UASs): A Review." *Remote Sensing* 11 (11): Article 11. https://doi.org/10.3390/rs11111308.

Wang, D., Q. Song, X. Liao, H. Ye, Q. Shao, J. Fan, N. Cong, X. Xin, H. Yue, and H. Zhang. 2020. "Integrating Satellite and Unmanned Aircraft System (UAS) Imagery to Model Livestock Population Dynamics in the Longbao Wetland National Nature Reserve, China." *Science of the Total Environment* 746: 140327. https://doi.org/10.1016/j.scitotenv.2020.140327.

Weinstein, B. G., and L. Prugh. 2018. "A Computer Vision for Animal Ecology." *Journal of Animal Ecology* 87 (3): 533–545. https://doi.org/10.1111/1365-2656.12780.

Whitford, M., and A. P. Klimley. 2019. "An Overview of Behavioral, Physiological, and Environmental Sensors Used in Animal Biotelemetry and Biologging Studies." *Animal Biotelemetry* 7 (1): 26. https://doi.org/10.1186/s40317-019-0189-z.

Wu, Z., C. Zhang, X. Gu, I. Duporge, L. F. Hughey, J. A. Stabach, A. K. Skidmore, et al. 2023. "Deep Learning Enables Satellite-Based Monitoring of Large Populations of Terrestrial Mammals Across Heterogeneous Landscape." *Nature Communications* 14 (1): Article 1. https://doi.org/10.1038/s41467-023-38901-y.

Xue, Y., T. Wang, and A. K. Skidmore. 2017. "Automatic Counting of Large Mammals from Very High Resolution Panchromatic Satellite Imagery." *Remote Sensing* 9 (9): Article 9. https://doi.org/10.3390/rs9090878.

Yang, Z., T. Wang, A. K. Skidmore, J. D. Leeuw, M. Y. Said, J. Freer, and M. Cristani. 2014. "Spotting East African Mammals in Open Savannah from Space." *Public Library of Science ONE* 9 (12): e115989. https://doi.org/10.1371/journal.pone.0115989.

Zhang, B., Y. Wu, B. Zhao, J. Chanussot, D. Hong, J. Yao, and L. Gao. 2022. "Progress and Challenges in Intelligent Remote Sensing Satellite Systems." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15:1814–1822. https://doi.org/10.1109/JSTARS.2022.3148139.

Zhao, Z.-Q., P. Zheng, S.-T. Xu, and X. Wu. 2019. "Object Detection with Deep Learning: A Review." *IEEE Transactions on Neural Networks and Learning Systems* 30 (11): 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865.

Appendix A. Overview of previous review papers' focus

Reference	Focus on animal detection/ counting/ survey	Focus on satellite imagery	Review of papers after 2018	Systematic review	All taxa	All regions
Butcher et al. (2021)			✓			✓
Clarke et al. (2021)	✓	✓	✓			✓
Corcoran et al. (2021)	✓		✓	✓	✓	✓
Delisle et al. (2023)	✓		✓	✓	✓	✓
Edney and Wood (2021)	✓		✓			✓
Goddijn-Murphy et al. (2021)	✓	✓	✓			✓
Hollings et al. (2018)	✓				✓	✓
Jiménez López and Mulero-Pázmány (2019)				✓	✓	✓
Kuenzer et al. (2014)		✓			✓	✓
Larue et al. (2017)	✓	✓			✓	✓
Linchant et al. (2015)	✓			✓	✓	✓
Nazir and Kaleem (2021)			✓		✓	✓
Pettorelli et al. (2014)		✓			✓	✓
Petrou et al. (2015)					✓	✓
Petso et al. (2021)	✓		✓		✓	✓
Sánchez-Díaz and Mata-Zayas (2019)		✓			✓	✓
Wang et al. (2019)	✓			✓	✓	✓
Weinstein and Prugh (2018)	✓			✓	✓	✓

Appendix B. List of papers selected for the systematic review

Ainley, D. G., Larue, M. A., Stirling, I., Stammerjohn, S., & Siniff, D. B. (2015). An apparent population decrease, or change in distribution, of Weddell seals along the Victoria Land coast. *Marine Mammal Science*, *31*(4), 1338–1361. https://doi.org/10.1111/mms.12220

Bamford, C. C. G., Kelly, N., Dalla Rosa, L., Cade, D. E., Fretwell, P. T., Trathan, P. N., Cubaynes, H. C., Mesquita, A. F. C., Gerrish, L., Friedlaender, A. S., & Jackson, J. A. (2020). A comparison of baleen whale density estimates derived from overlapping satellite imagery and a shipborne survey. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-69887-y

Barber-Meyer, S., Kooyman, G., & Ponganis, P. (2007). Estimating the relative abundance of Emperor Penguins at inaccessible colonies using satellite imagery. *Polar Biology*, *30*, 1565–1570. https://doi.org/10.1007/s00300-007-0317-8

Begall, S., Červený, J., Neef, J., Vojtěch, O., & Burda, H. (2008). Magnetic alignment in grazing and resting cattle and deer. *Proceedings of the National Academy of Sciences*, *105*(36), 13451–13455. https://doi.org/10.1073/pnas.0803650105

Borowicz, A., Le, H., Humphries, G., Nehls, G., Höschle, C., Kosarev, V., & Lynch, H. J. (2019). Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLOS ONE*, *14*(10), e0212532. https://doi.org/10.1371/journal.pone.0212532

Bowler, E., Fretwell, P. T., French, G., & Mackiewicz, M. (2020). Using Deep Learning to Count Albatrosses from Space: Assessing Results in Light of Ground Truth Uncertainty. *Remote Sensing*, *12*(12), Article 12. https://doi.org/10.3390/rs12122026

Charry, B., Tissier, E., Iacozza, J., Marcoux, M., & Watt, C. A. (2021). Mapping Arctic cetaceans from space: A case study for beluga and narwhal. *PLOS ONE*, *16*(8), e0254380. https://doi.org/10.1371/journal.pone.0254380

Corrêa, A. A., Quoos, J. H., Barreto, A. S., Groch, K. R., & Eichler, P. P. B. (2022). Use of satellite imagery to identify southern right whales (Eubalaena australis) on a Southwest Atlantic Ocean breeding ground. *Marine Mammal Science*, *38*(1), 87–101. https://doi.org/10.1111/mms.12847

Cubaynes, H. C., Fretwell, P. T., Bamford, C., Gerrish, L., & Jackson, J. A. (2019). Whales from space: Four mysticete species described using new VHR satellite imagery. *Marine Mammal Science*, *35*(2), 466–491. https://doi.org/10.1111/mms.12544

Duporge, I., Isupova, O., Reece, S., Macdonald, D. W., & Wang, T. (2021). Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *Remote Sensing in Ecology and Conservation*, *7*(3), 369–381. https://doi.org/10.1002/rse2.195

Fretwell, P. T., Cubaynes, H. C., & Shpak, O. V. (2023). Satellite image survey of beluga whales in the southern Kara Sea. *Marine Mammal Science*, *n/a*(n/a). https://doi.org/10.1111/mms.13044

Fretwell, P. T., Jackson, J. A., Encina, M. J. U., Häussermann, V., Alvarez, M. J. P., Olavarría, C., & Gutstein, C. S. (2019). Using remote sensing to detect whale strandings in remote areas: The case of sei whales mass mortality in Chilean Patagonia. *PLOS ONE*, *14* (10), e0222498. https://doi.org/10.1371/journal.pone.0222498

Fretwell, P. T., LaRue, M. A., Morin, P., Kooyman, G. L., Wienecke, B., Ratcliffe, N., Fox, A. J., Fleming, A. H., Porter, C., & Trathan, P. N. (2012). An Emperor Penguin Population Estimate: The First Global, Synoptic Survey of a Species from Space. *PLOS ONE*, *7*(4), e33751. https://doi.org/10.1371/journal.pone.0033751

Fretwell, P. T., Scofield, P., & Phillips, R. A. (2017). Using super-high resolution satellite imagery to census threatened albatrosses. *Ibis*, *159*(3), 481–490. https://doi.org/10.1111/ibi.12482

Fretwell, P. T., Staniland, I. J., & Forcada, J. (2014). Whales from Space: Counting Southern Right Whales by Satellite. *PLOS ONE*, *9*(2), e88655. https://doi.org/10.1371/journal.pone.0088655

Gonçalves, B. C., Spitzbart, B., & Lynch, H. J. (2020). SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. *Remote Sensing of Environment*, *239*, 111617. https://doi.org/10.1016/j.rse.2019.111617

Green, K. M., Virdee, M. K., Cubaynes, H. C., Aviles-Rivero, A. I., Fretwell, P. T., Gray, P. C., Johnston, D. W., Schönlieb, C.-B., Torres, L. G., & Jackson, J. A. (2023). Gray whale detection in satellite imagery using deep learning. *Remote Sensing in Ecology and Conservation*, *n/a*(n/a). https://doi.org/10.1002/rse2.352

Guinet, C., Jouventin, P., & Malacamp, J. (1995). Satellite remote sensing in monitoring change of seabirds: Use of Spot Image in king penguin population increase at Ile aux Cochons, Crozet Archipelago. *Polar Biology*, *15*(7), 511–515. https://doi.org/10.1007/BF00237465

Guirado, E., Tabik, S., Rivas, M. L., Alcaraz-Segura, D., & Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-50795-9

Hughes, B. J., Martin, G. R., & Reynolds, S. J. (2011). The use of Google EarthTM satellite imagery to detect the nests of masked boobies Sula dactylatra. *Wildlife Biology*, *17*(2), 210–216. https://doi.org/10.2981/10–106

Irvine, J. M., Nolan, J., Hofmann, N., Lewis, D., Simpamba, T., Zyambo, P., Travis, A. J., & Hemami, S. (2019). Estimating the Population of Large Animals in the Wild Using Satellite Imagery: A Case Study of Hippos in Zambia's Luangwa River. *2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 1–8. https://doi.org/10.1109/AIPR47015.2019.9174564

Kapoor, S., Kumar, M., & Kaushal, M. (2023). Deep learning based whale detection from satellite imagery. *Sustainable Computing: Informatics and Systems*, *38*, 100858. https://doi.org/10.1016/j.suscom.2023.100858

Labrousse, S., Iles, D., Viollat, L., Fretwell, P., Trathan, P. N., Zitterbart, D. P., Jenouvrier, S., & LaRue, M. (2022). Quantifying the causes and consequences of variation in satellite-derived population indices: A case study of emperor penguins. *Remote Sensing in Ecology and Conservation*, *8*(2), 151–165. https://doi.org/10.1002/rse2.233

Laliberte, A. S., & Ripple, W. J. (2003). Automated Wildlife Counts from Remotely Sensed Imagery. *Wildlife Society Bulletin (1973–2006)*, *31*(2), 362–371.

LaRue, M. A., Lynch, H. J., Lyver, P. O. B., Barton, K., Ainley, D. G., Pollard, A., Fraser, W. R., & Ballard, G. (2014). A method for estimating colony sizes of Adélie penguins using remote sensing imagery. *Polar Biology*, *37*(4), 507–517. https://doi.org/10.1007/s00300-014-1451-8

LaRue, M. A., Rotella, J. J., Garrott, R. A., Siniff, D. B., Ainley, D. G., Stauffer, G. E., Porter, C. C., & Morin, P. J. (2011). Satellite imagery can be used to detect variation in abundance of Weddell seals (Leptonychotes weddellii) in Erebus Bay, Antarctica. *Polar Biology*, *34* (11), 1727. https://doi.org/10.1007/s00300-011-1023-0

LaRue, M. A., & Stapleton, S. (2018). Estimating the abundance of polar bears on Wrangel Island during late summer using high-resolution satellite imagery: A pilot study. *Polar Biology*, *41*(12), 2621–2626. https://doi.org/10.1007/s00300-018-2384-4

LaRue, M. A., Stapleton, S., & Anderson, M. (2017). Feasibility of using high-resolution satellite imagery to assess vertebrate wildlife populations. *Conservation Biology*, *31*(1), 213–220. https://doi.org/10.1111/cobi.12809

LaRue, M. A., Stapleton, S., Porter, C., Atkinson, S., Atwood, T., Dyck, M., & Lecomte, N. (2015). Testing methods for using high-resolution satellite imagery to monitor polar bear abundance and distribution: Remote Sensing of Polar Bears. *Wildlife Society Bulletin*, *39*(4), 772–779. https://doi.org/10.1002/wsb.596

LaRue, M., Salas, L., Nur, N., Ainley, D., Stammerjohn, S., Pennycook, J., Dozier, M., Saints, J., Stamatiou, K., Barrington, L., & Rotella, J. (2021). Insights from the first global population estimate of Weddell seals in Antarctica. *Science Advances*, *7*(39), eabh3674. https://doi.org/10.1126/sciadv.abh3674

Le, H., Samaras, D., & Lynch, H. J. (2022). A convolutional neural network architecture designed for the automated survey of seabird colonies. *Remote Sensing in Ecology and Conservation*, *8*(2), 251–262. https://doi.org/10.1002/rse2.240

Lynch, H. J., & LaRue, M. A. (2014). First global census of the Adélie Penguin. *The Auk*, *131*(4), 457–466. https://doi.org/10.1642/AUK-14-31.1

Lynch, H. J., & Schwaller, M. R. (2014). Mapping the Abundance and Distribution of Adélie Penguins Using Landsat-7: First Steps toward an Integrated Multi-Sensor Pipeline for Tracking Populations at the Continental Scale. *PLOS ONE*, *9*(11), e113301. https://doi.org/10.1371/journal.pone.0113301

Lynch, H. J., White, R., Black, A. D., & Naveen, R. (2012). Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar Biology*, *35*(6), 963–968. https://doi.org/10.1007/s00300-011-1138-3

McMahon, C. R., Howe, H., Hoff, J. van den, Alderman, R., Brolsma, H., & Hindell, M. A. (2014). Satellites, the All-Seeing Eyes in the Sky: Counting Elephant Seals from Space. *PLOS ONE*, *9*(3), e92613. https://doi.org/10.1371/journal.pone.0092613

Mücher, C. A., Los, S., Franke, G. J., & Kamphuis, C. (2022). Detection, identification and posture recognition of cattle with satellites, aerial photography and UAVs using deep learning techniques. *International Journal of Remote Sensing*, *43*(7), 2377–2392. https://doi.org/10.1080/01431161.2022.2051634

Naveen, R., Lynch, H. J., Forrest, S., Mueller, T., & Polito, M. (2012). First direct, site-wide penguin survey at Deception Island, Antarctica, suggests significant declines in breeding chinstrap penguins. *Polar Biology*, *35*(12), 1879–1888. https://doi.org/10.1007/s00300-012-1230-3

Platonov, N. G., Mordvintsev, I. N., & Rozhnov, V. V. (2013). The possibility of using high resolution satellite images for detection of marine mammals. *Biology Bulletin*, *40*(2), 197–205. https://doi.org/10.1134/S1062359013020106

Sasamal, S. K., Chaudhury, S. B., Samal, R. N., & Pattanaik, A. K. (2008). QuickBird spots flamingos off Nalabana Island, Chilika Lake, India. *International Journal of Remote Sensing*, *29*(16), 4865–4870. https://doi.org/10.1080/01431160701814336

Schwaller, M. R., Southwell, C. J., & Emmerson, L. M. (2013). Continental-scale mapping of Adélie penguin colonies from Landsat imagery. *Remote Sensing of Environment*, *139*, 353–364. https://doi.org/10.1016/j.rse.2013.08.009

Sidle, J. G., Johnson, D. H., Euliss, B. R., & Tooze, M. (2002). Monitoring Black-Tailed Prairie Dog Colonies with High-Resolution Satellite Imagery. *Wildlife Society Bulletin (1973–2006)*, *30*(2), 405–411.

Stapleton, S., LaRue, M., Lecomte, N., Atkinson, S., Garshelis, D., Porter, C., & Atwood, T. (2014). Polar Bears from Space: Assessing Satellite Imagery as a Tool to Track Arctic Wildlife. *PLoS ONE*, *9*(7), e101513. https://doi.org/10.1371/journal.pone.0101513

Swinbourne, M. J., Taggart, D. A., Swinbourne, A. M., Lewis, M., & Ostendorf, B. (2018). Using satellite imagery to assess the distribution and abundance of southern hairy-nosed wombats (Lasiorhinus latifrons). *Remote Sensing of Environment*, *211*, 196–203. https://doi.org/10.1016/j.rse.2018.04.017

Wang, D., Song, Q., Liao, X., Ye, H., Shao, Q., Fan, J., Cong, N., Xin, X., Yue, H., & Zhang, H. (2020). Integrating satellite and unmanned aircraft system (UAS) imagery to model livestock population dynamics in the Longbao Wetland National Nature Reserve, China. *Science of The Total Environment*, *746*, 140327. https://doi.org/10.1016/j.scitotenv.2020.140327

Witharana, C., LaRue, M. A., & Lynch, H. J. (2016). Benchmarking of data fusion algorithms in support of earth observation based Antarctic wildlife monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, *113*, 124–143. https://doi.org/10.1016/j.isprsjprs.2015.12.009

Wu, Z., Zhang, C., Gu, X., Duporge, I., Hughey, L. F., Stabach, J. A., Skidmore, A. K., Hopcraft, J. G. C., Lee, S. J., Atkinson, P. M., McCauley, D. J., Lamprey, R., Ngene, S., & Wang, T. (2023). Deep learning enables satellite-based monitoring of large populations of terrestrial mammals across heterogeneous landscape. *Nature Communications*, *14*(1), Article 1. https://doi.org/10.1038/s41467-023-38901-y

Xue, Y., Wang, T., & Skidmore, A. K. (2017). Automatic Counting of Large Mammals from Very High Resolution Panchromatic Satellite Imagery. *Remote Sensing*, *9*(9), Article 9. https://doi.org/10.3390/rs9090878

Yang, Z., Wang, T., Skidmore, A. K., Leeuw, J. de, Said, M. Y., & Freer, J. (2014). Spotting East African Mammals in Open Savannah from Space. *PLOS ONE*, *9*(12), e115989. https://doi.org/10.1371/journal.pone.0115989

Zhao, P., Liu, S., Zhou, Y., Lynch, T., Lu, W., Zhang, T., & Yang, H. (2021). Estimating animal population size with very high-resolution satellite imagery. *Conservation Biology*, *35*(1), 316–324. https://doi.org/10.1111/cobi.13613