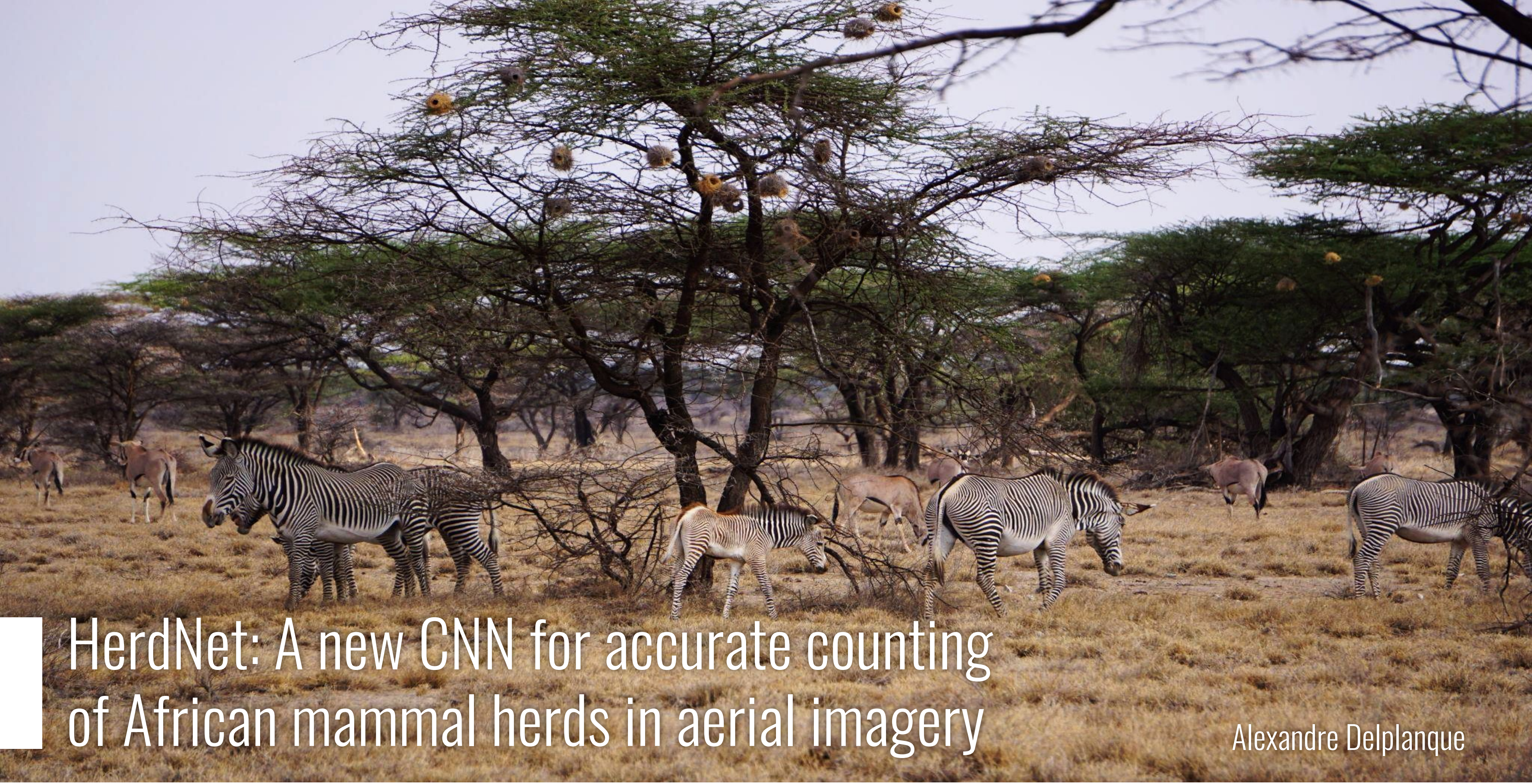


HerdNet: A new CNN for accurate counting of African mammal herds in aerial imagery

Alexandre Delplanque

Forest Is Life seminar on April 07, 2023

LIÈGE université Gembloux Agro-Bio Tech
UdS Université de Sherbrooke
Réserve Naturelle et Culturelle de L'ENNEDI Tchad
fnrs FREEDOM TO RESEARCH

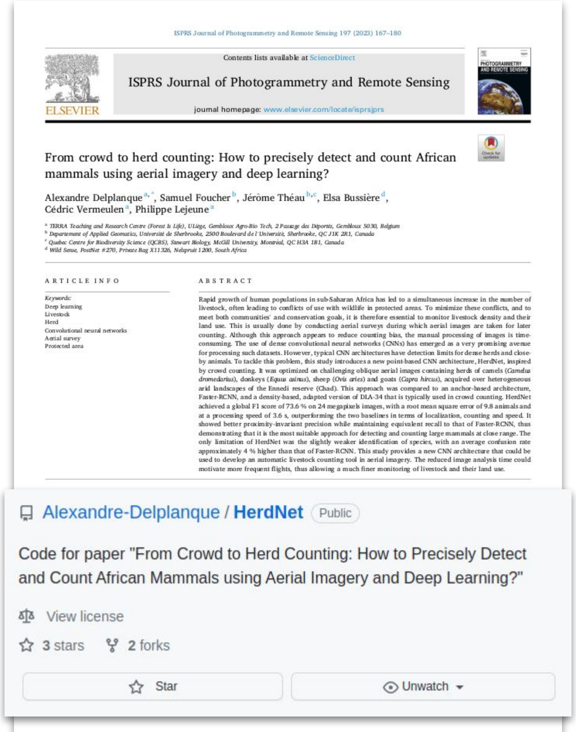
INTRODUCTION

Objective of my PhD thesis | Assess the use of Deep Learning and aerial imagery to automate the detection and counting of large terrestrial mammals in African protected areas

- Evaluate the use of common object detection CNNs
- Develop a specific CNN for accurate animal detection in aerial imagery
- Incorporate CNN model in the aerial surveys data processing
- Develop an approach for easy adaptation of CNN models to new landscapes

Research paper in ISPRS JPRS: https://doi.org/10.1016/j.isprsjprs.2023.01.025

Open-access code on GitHub: https://github.com/Alexandre-Delplanque/HerdNet



ISPRS Journal of Photogrammetry and Remote Sensing 197 (2023) 167–180

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

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

Alexandre Delplanque[a,*], Samuel Foucher[b], Jérôme Théau[b,c], Elsa Bussière[d], Cédric Vermeulen[a], Philippe Lejeune[a]

[a] TERRA Teaching and Research Centre (Forest Is Life), ULiège, Gembloux Agro-Bio Tech, 2 Passage des Déportés, Gembloux 5030, Belgium
[b] Departement of Applied Geomatics, Université de Sherbrooke, 2500 Boulevard de l'Université, Sherbrooke, QC J1K 2R1, Canada
[c] Quebec Centre for Biodiversity Science (QCBS), Stewart Biology, McGill University, Montréal, QC H3A 1B1, Canada
[d] Wild Sense, PostNet #270, Private Bag X11326, Nelspruit 1200, South Africa

ARTICLE INFO

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

ABSTRACT

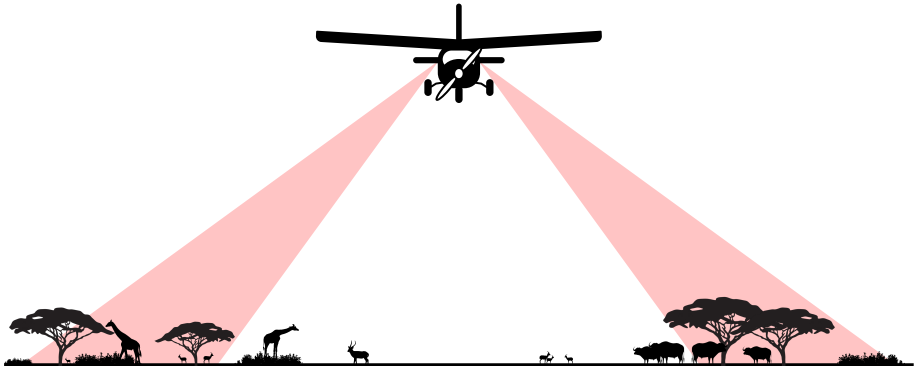
Rapid growth of human populations in sub-Saharan Africa has led to a simultaneous increase in the number of livestock, often leading to conflicts of use with wildlife in protected areas. To minimize these conflicts, and to meet both communities' and conservation goals, it is therefore essential to monitor 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.

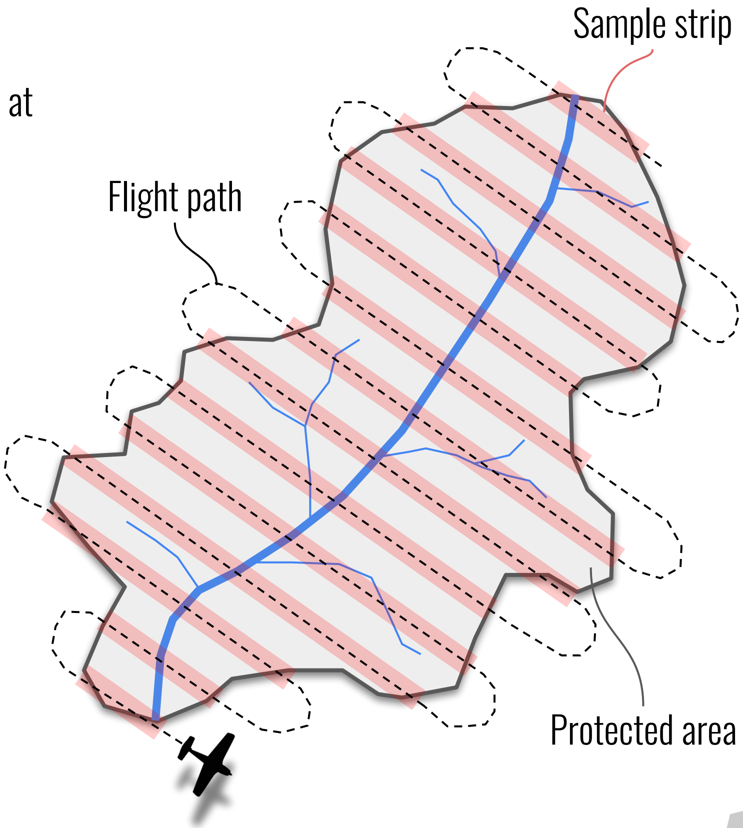
BASIC PRINCIPLES OF AN AERIAL SURVEY

Standard protocol | At least two observers (one on each side) in aircraft flying at low altitude (~300ft.) and following systematic sample strips

Real-time on-sight count of animals

↦ High risk of counting errors, especially for **large groups** (i.e., herds)





USE OF ON-BOARD CAMERAS

Real-time on-sight count is not an easy task and is prone to errors

- Could provide **imprecise** population estimates due mainly to the short observation time (~5s)

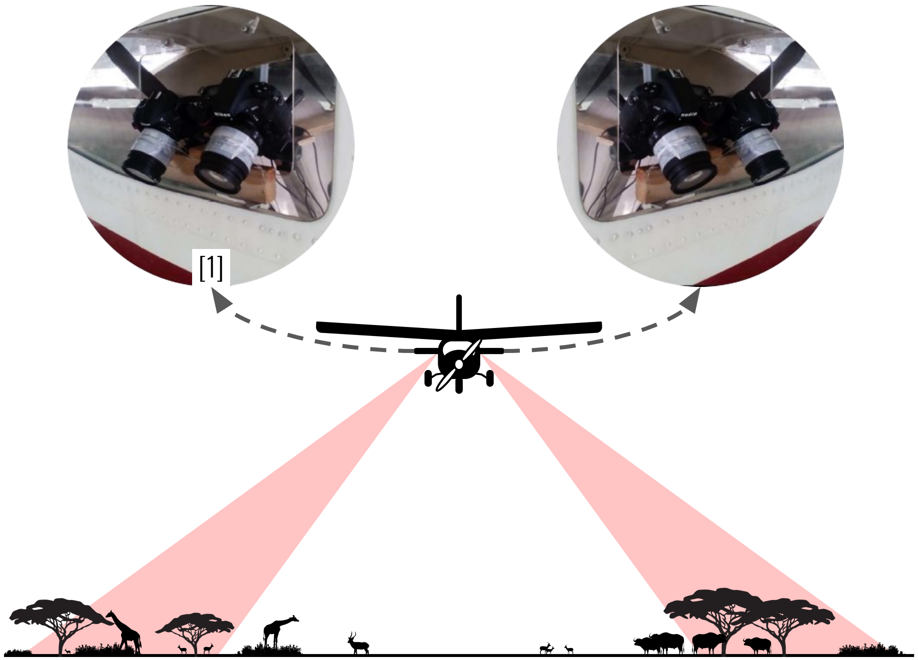
How to reduce such bias?

- By using **on-board cameras**, acquiring imagery at fixed interval

    ↦ Precise the counts
    ↦ Provide more coherent population estimates

    ↦ **Large volume of data** to manually interpret (few seconds to several minutes)

⇨ DEEP LEARNING

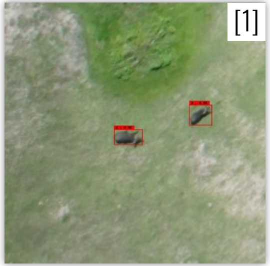


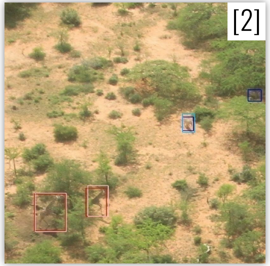
[1] Lamprey et al. (2020): https://doi.org/10.1002/rse2.154

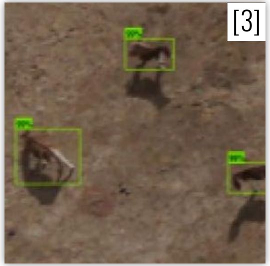
AUTOMATED COUNTING WITH COMMON OBJECT DETECTION CNNs

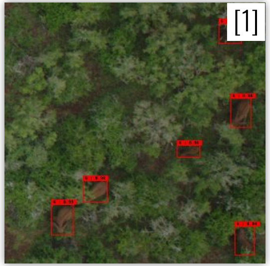
Common object detection CNNs - Faster-RCNN, RetinaNet, etc. - need bounding boxes to be trained

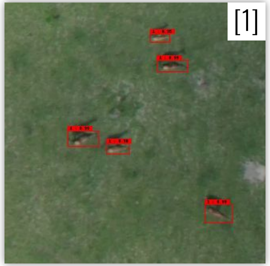
Usually show good performances for **isolated mammals** or **sparse herds**



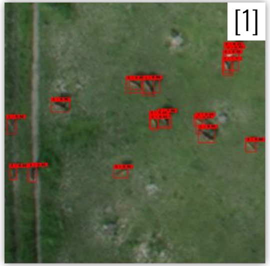


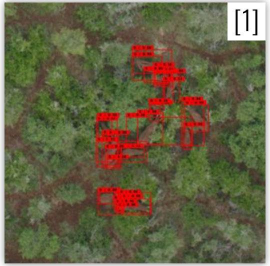


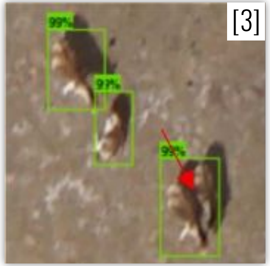


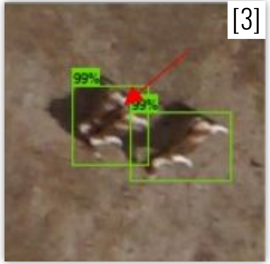


But showed a drop of performances for **dense herds** and **close-by individuals**



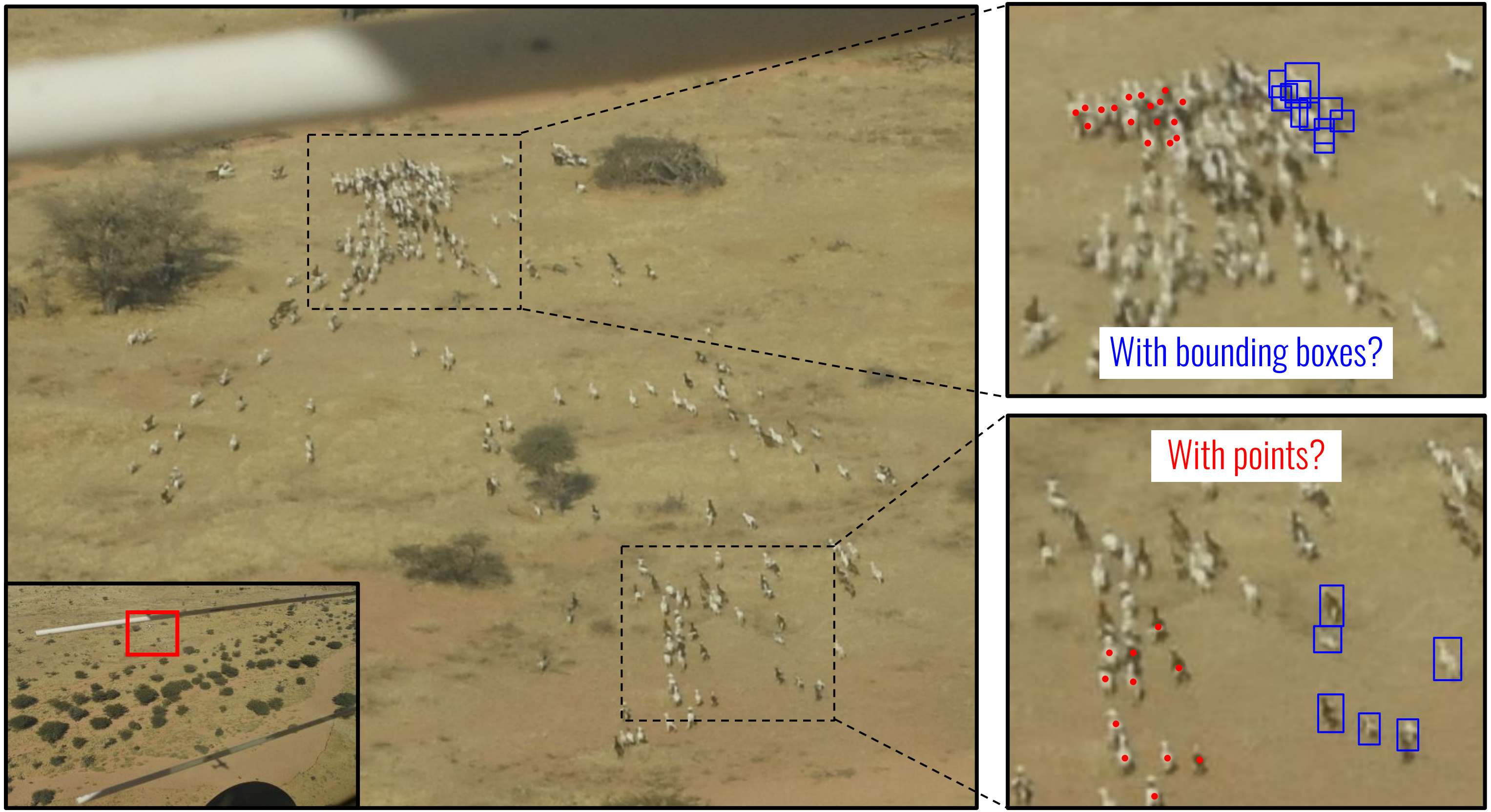






[1] Delplanque et al. (2022): https://doi.org/10.1002/rse2.234 [2] Eikelboom et al. (2019): https://doi.org/10.1111/2041-210X.13277 [3] Peng et al. (2021): https://doi.org/10.1016/j.isprsjprs.2020.08.026

HOW TO EFFICIENTLY ANNOTATE/COUNT A HERD LIKE THIS?

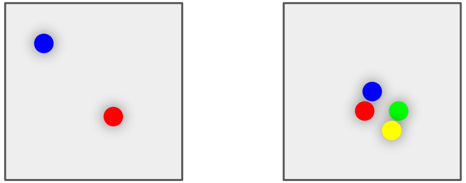


WHAT DOES THAT MAKE YOU THINK OF?

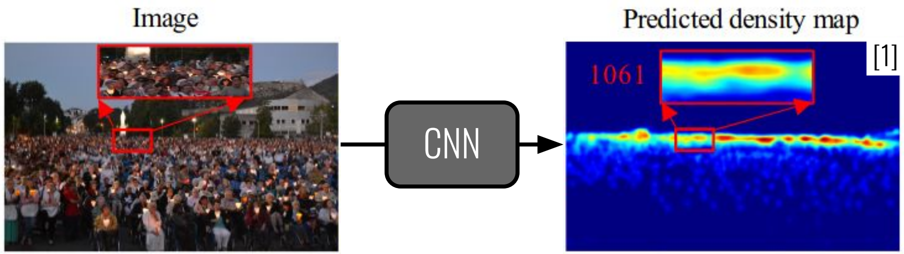
HERD COUNTING ~ CROWD COUNTING

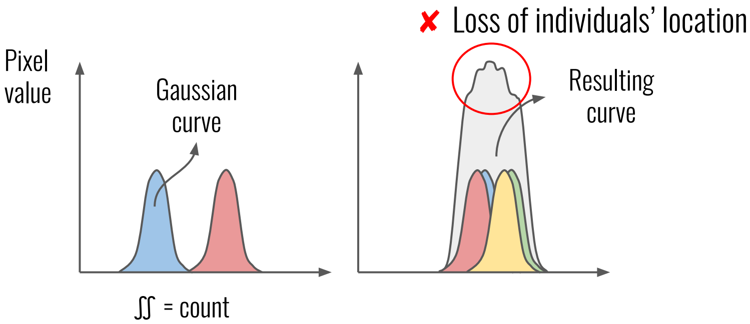
(a) Occlusion (b) Complex background (c) Scale variation (d) Non-uniform distribution

CROWD COUNTING CNN

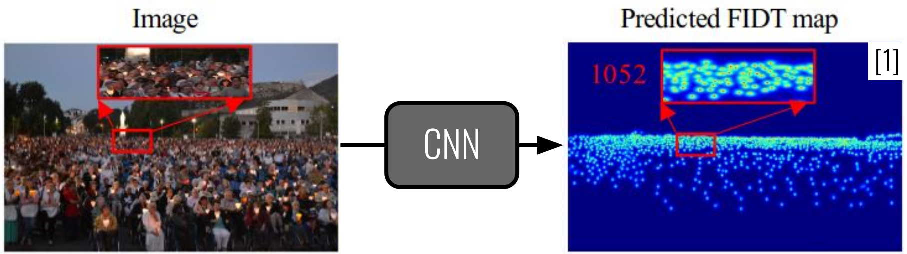


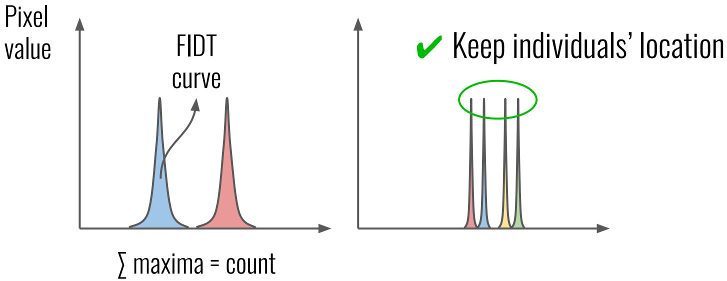
Most used approach | CNN trained with Gaussian map - produces density





Recent approach | CNN trained with FIDT[1] map - produces points

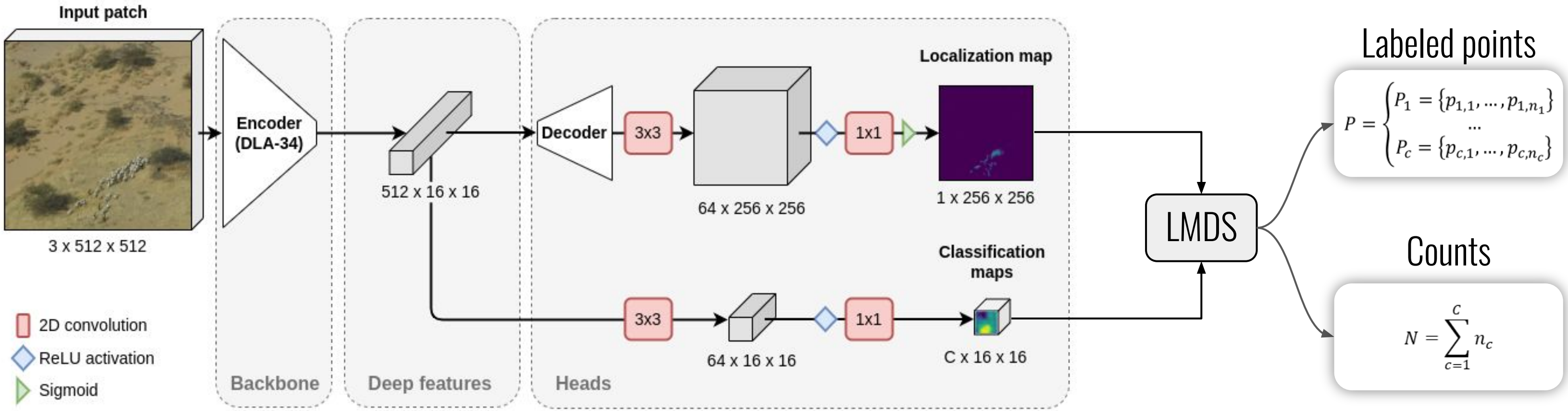




[1] Liang et al. (2021): http://arxiv.org/abs/2102.07925

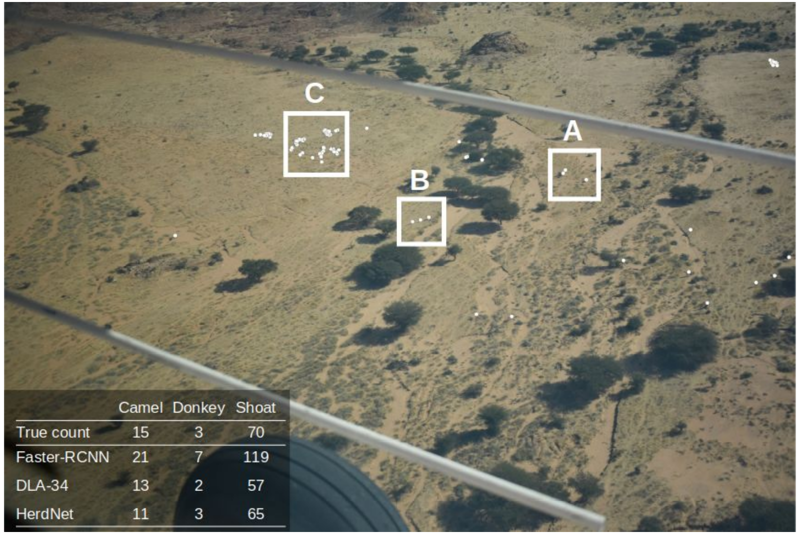
PROPOSED APPROACH: HerdNet - Architecture

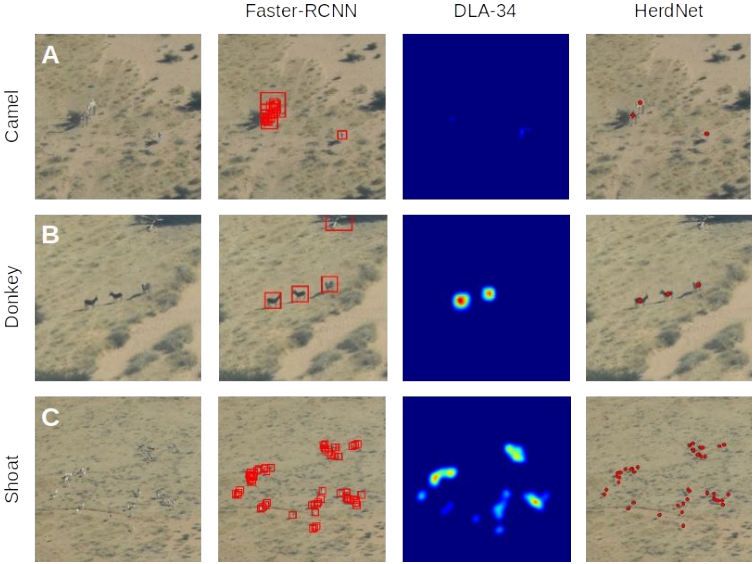
Inspired by **crowd counting** and **point-based** object detectors



PROPOSED APPROACH: HerdNet - Detection results

Study area and dataset | Ennedi Reserve (Chad) - Oblique aerial images of the 2019 aerial survey (n=914)
Target species | Domestic mammals: camels, donkeys, sheep and goats
Baselines | Faster-RCNN (common object detector) & Adapted DLA-34 (usual crowd counting approach)





Faster-RCNN

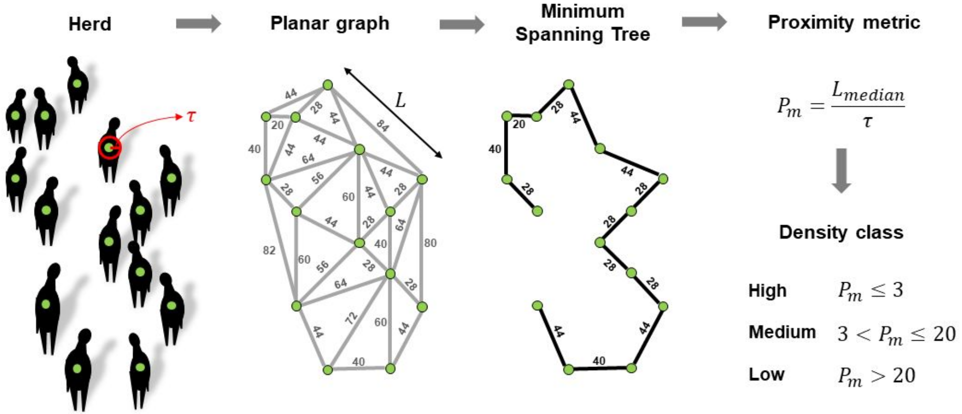
Recall	60%
Precision	39%
F1 score	47%
Confusion	11%

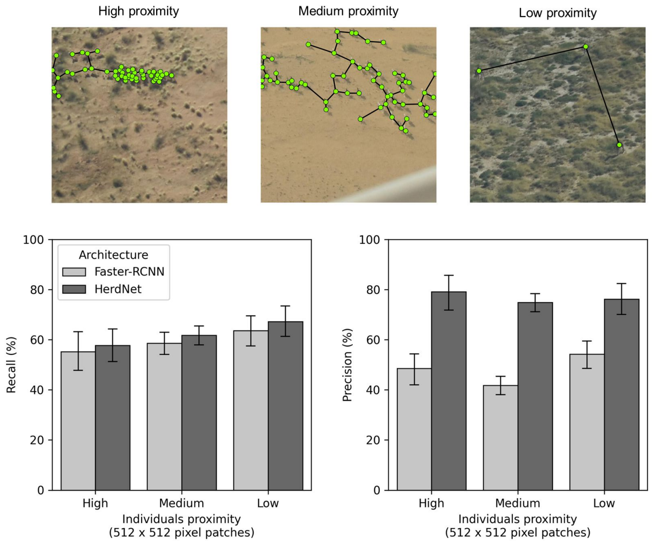
HerdNet

Recall	70%
Precision	78%
F1 score	74%
Confusion	16%

PROPOSED APPROACH: HerdNet - Detection results

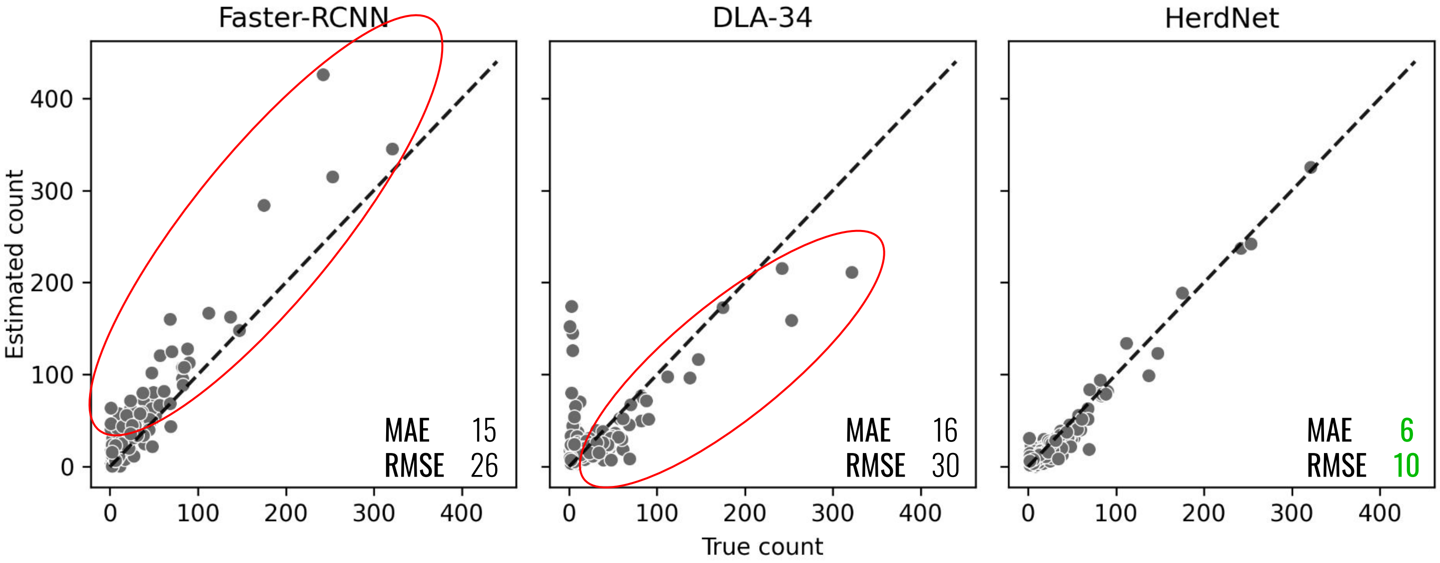
Study area and dataset | Ennedi Reserve (Chad) - Oblique aerial images of the 2019 aerial survey (n=914)
Target species | Domestic mammals: camels, donkeys, sheep and goats
Baselines | Faster-RCNN (common object detector) & Adapted DLA-34 (usual crowd counting approach)





PROPOSED APPROACH: HerdNet - Counting results

Study area and dataset | Ennedi Reserve (Chad) - Oblique aerial images of the 2019 aerial survey (n=914)
Target species | Domestic mammals: camels, donkeys, sheep and goats
Baselines | Faster-RCNN (common object detector) & Adapted DLA-34 (usual crowd counting approach)



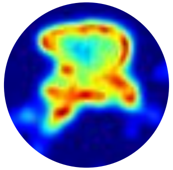
PROPOSED APPROACH: HerdNet - Results on challenging herds

BEST APPROACH FOR COUNTING DENSE HERDS?

Three approaches were evaluated:

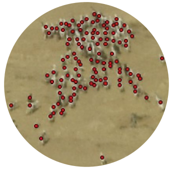
1) ■ CNN-**anchor**-based object detector: Faster-RCNN

- Drop of precision in dense herd, as previously observed
- Systematic over-counting



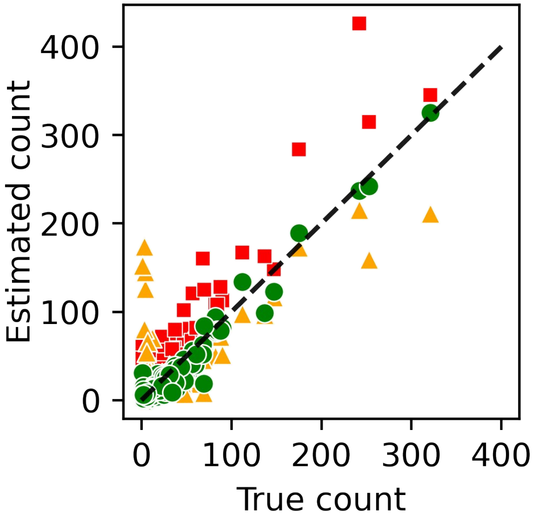
2) ▲ CNN-**density**-based detector: adapted DLA-34

- Under-counting
- Probably caused by high variance in the number of animals



3) ● CNN-**point**-based object detector: HerdNet

- Best detection and counting performances
- 1.4x Faster (3.6s/24MP image)
- Reliable estimated count per image

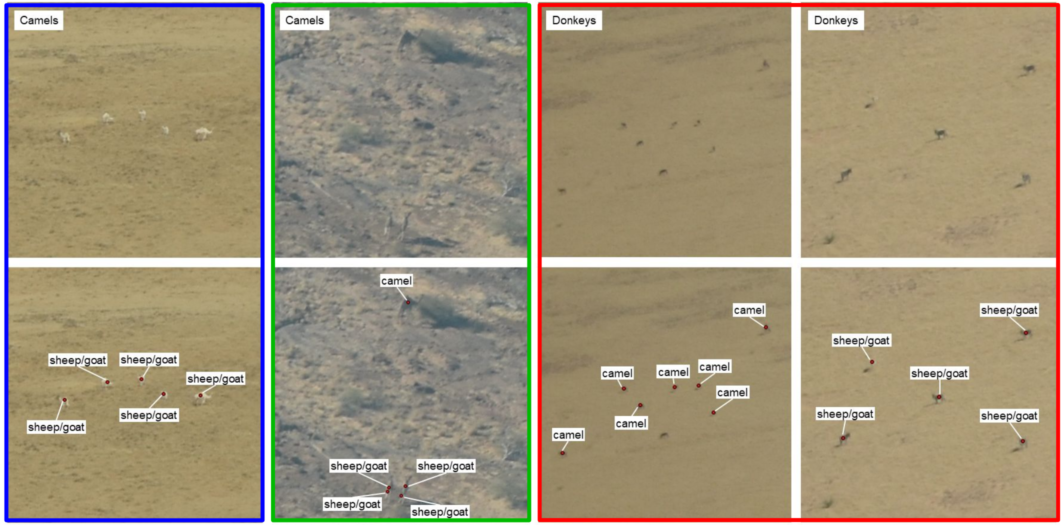


SPECIES IDENTIFICATION

HerdNet better identified the majority species (sheep/goats) but was not so good for minority ones (i.e., camels and donkey)

- Might be explained by variance in fur color, standing position and size of the animal

- Identification was sometimes difficult when annotating (help of the observers' flight sheet)



POTENTIAL USE OF HERDNET - Drone

HerdNet has been trained on a **drone dataset** to assess its potential use (see Appendix S4 for more details)

Dataset | Nadir drone images of our previous paper[1] (n=1297)

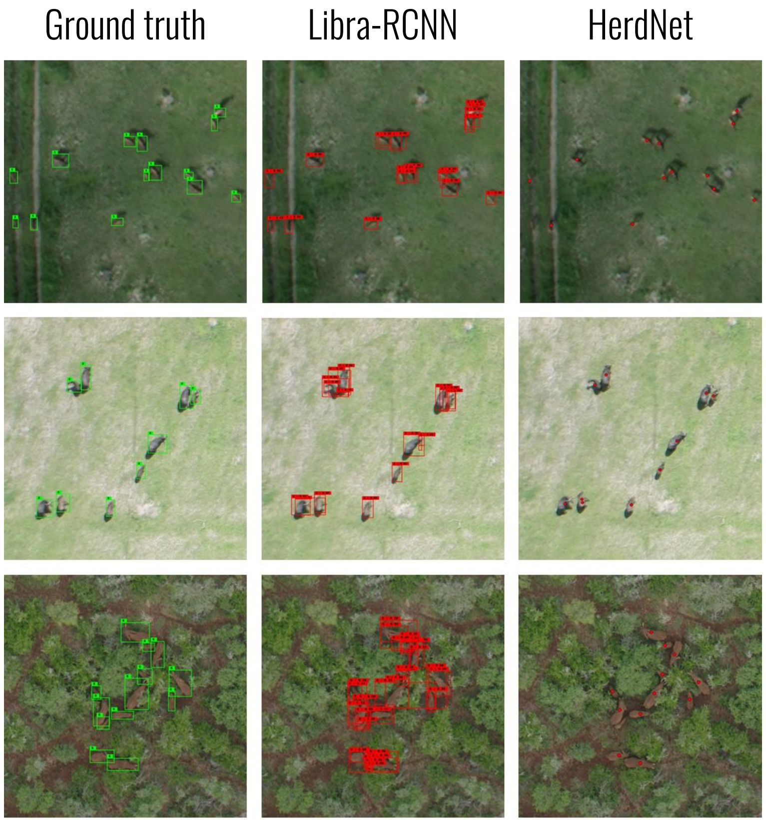
Target species | Wildlife: topi, buffalo, kob, warthog, waterbuck and elephant

Baseline | Libra-RCNN (state-of-the-art model)

Table 5 - 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.



[1] Delplanque et al. (2022): https://doi.org/10.1002/rse2.234

POTENTIAL USE OF HERDNET - Mixed herds

Identification approach may become a concern in the case of dense mixed herds (i.e., when different species are close to each other)

- Classification maps of 16x16 pixel were sufficient for our case (distance between different species > 32 pixels in input patch)

- If distance < 32 pixels (i.e., < 2-3m in real life, with similar camera), resolution of the classification head should be increased

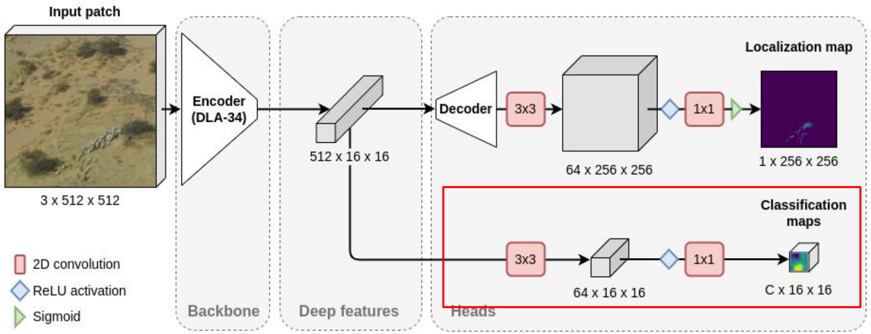


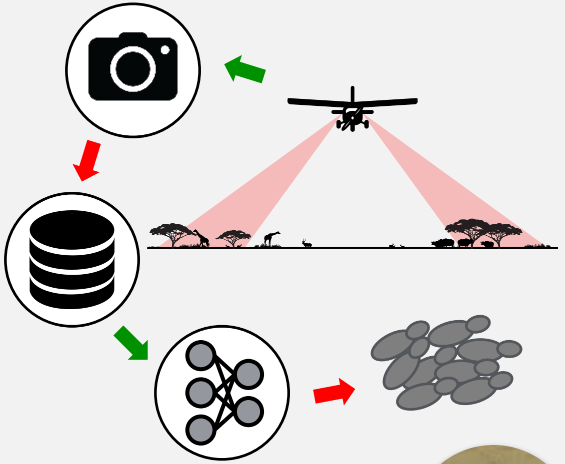
Table 2 – Identification performances of HerdNet on full images of the Ennedi validation at different classification map resolution. Values in bold indicate the best performance among the three resolutions.

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.

CONCLUSION

Context



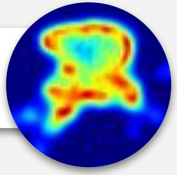
Common object detection CNNs struggle to detect/count dense herds

Method

Anchor-based CNN (Faster-RCNN)

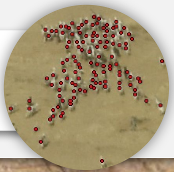
VS

Density-based CNN (DLA-34)

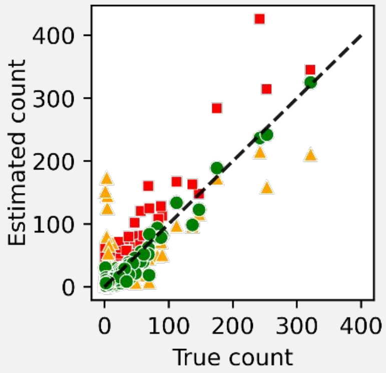


VS

Point-based CNN (HerdNet)



Results



±10

73.6%

3.6 sec

Thank you for your time, any questions?

Alexandre Delplanque

Forest Is Life seminar on April 07, 2023