# COUNTING AFRICAN MAMMAL HERDS IN AERIAL IMAGERY USING DEEP LEARNING: ARE ANCHOR-BASED ALGORITHMS THE MOST SUITABLE?

10th International Conference on Agro-Geoinformatics and
43rd Canadian Symposium on Remote Sensing

July 11 to 14, 2022

Alexandre Delplanque[1]

Samuel Foucher[2], Philippe Lejeune[1] & Jérôme Théau[2,3]

CRSS·SCT

1 LIÈGE université Gembloux Agro-Bio Tech

2 UDS Université de Sherbrooke

3 CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC

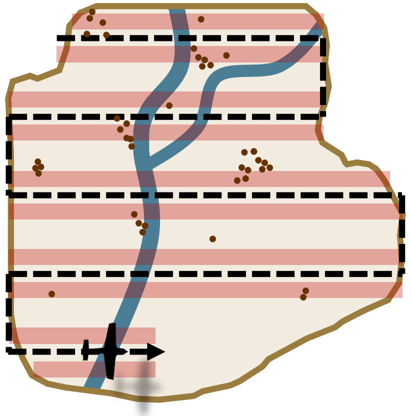
Réserve Naturelle et Culturelle de L'ENNEDI Tchad

fnrs LA LIBERTÉ DE CHERCHER

# WHY USING DEEP LEARNING?

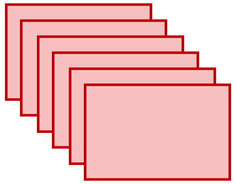
- **Frequent surveys** and **monitoring** of wildlife/livestock are essential for **conservation**
- **Standard survey method**:
  observers in aircraft flying at low altitude following systematic sample units

- × Dangerous
- × Complex logistic
- × Counting errors

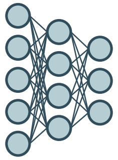


- **Unmanned Aerial Vehicle**: promising alternative

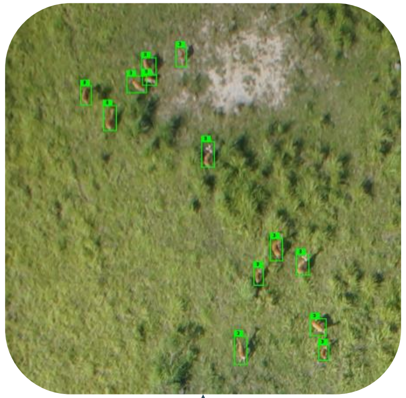
- ✓ Remotely controlled
- ✓ Easier logistic
- ✓ Precise counts



Huge volume of data (images)



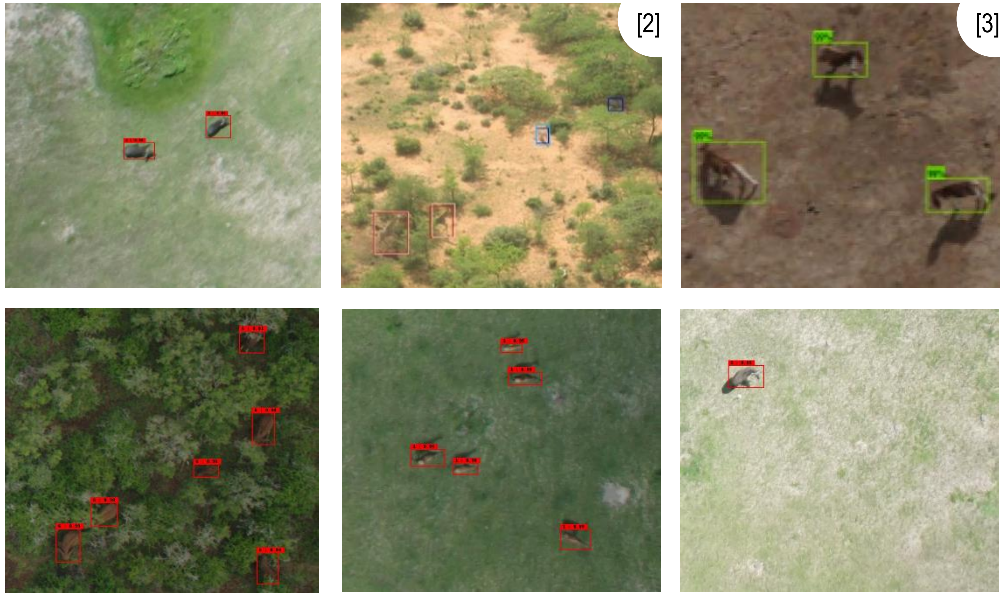
Deep Learning



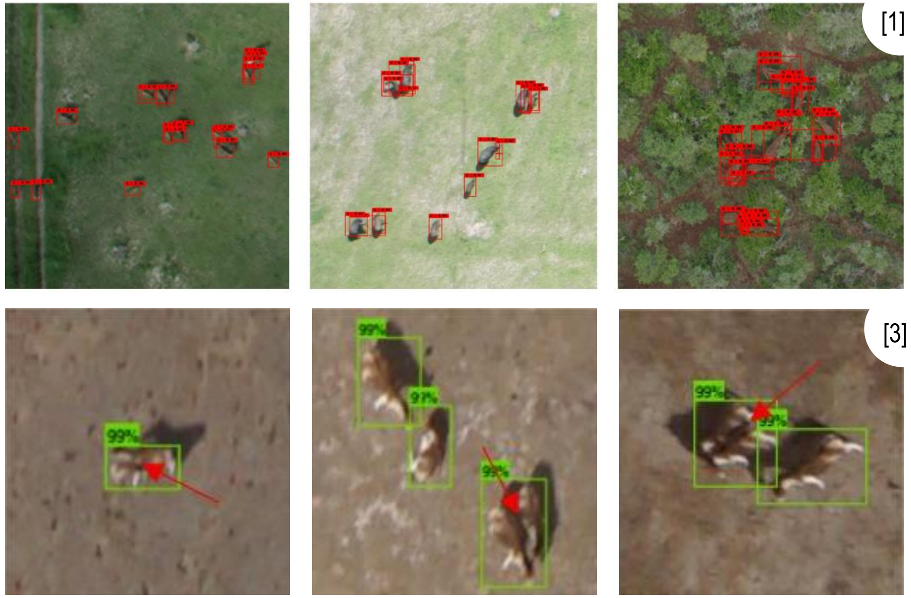
# THE CASE OF HERD COUNTING

- **CNN Anchor-based object detectors** (e.g. Faster-RCNN, RetinaNet)

✓ Good performances for **isolated** mammals and **sparse** herds[1,2,3]




✗ Drop in performances for dense **herds** and **close-by** individuals[1,3]




Are anchor-based object detectors the most suitable for counting large mammals in aerial imagery?

# Aerial Mammal Dataset

- Use of the dataset of Delplanque *et al.* (2021)[1]
  - UAV and nadir aerial images
  - 6 wildlife species
  - Savannah, woodland, open shrubland, grassland
  - GSD between 2.4 and 13.0 cm/pixel
  - 1,297 images
  - 10,239 bounding box annotations

- From bounding box to point …

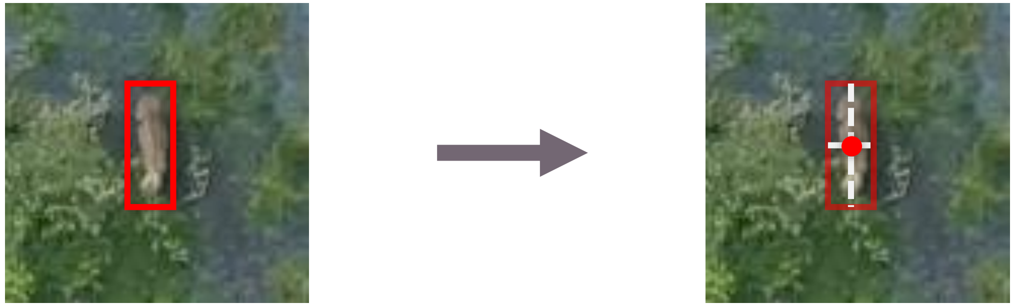


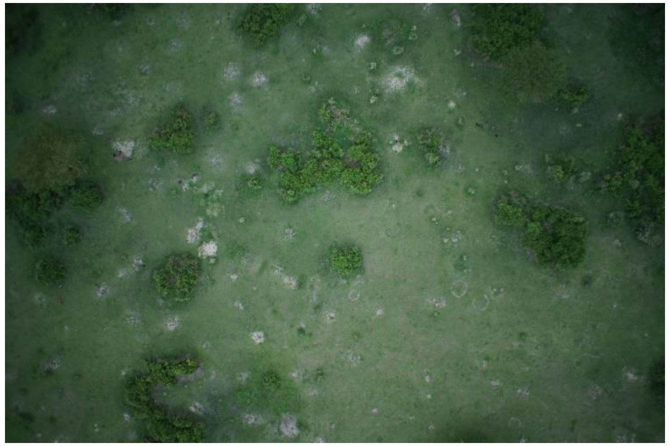
**Table 2.** Number of individuals according to species, training, validation and test sets.

| 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%) | 10 239 |

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

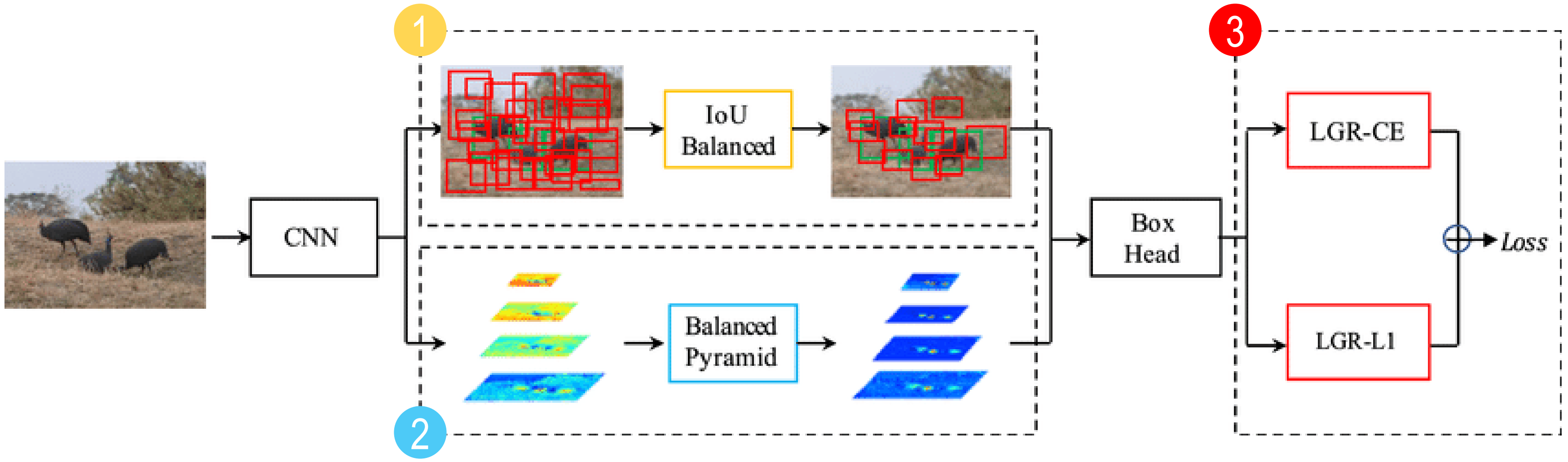
# SOME SAMPLES



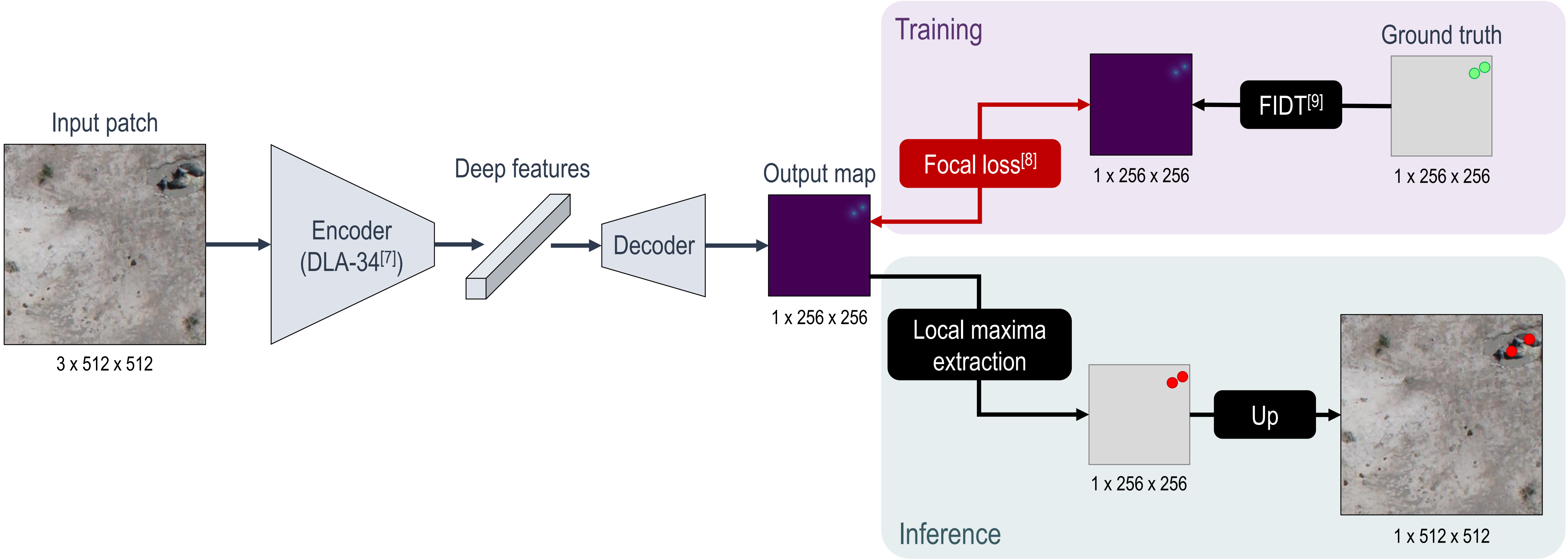
# THE BASELINE: LIBRA-RCNN

- Dataset's **state-of-the-art** model
- **Anchor-based two-stage detector** proposed by Pang *et al.* (2019)[4], balanced version of Faster-RCNN[5]:
  1) Balanced distribution of training samples
  2) Balanced feature levels
  3) Balanced training loss



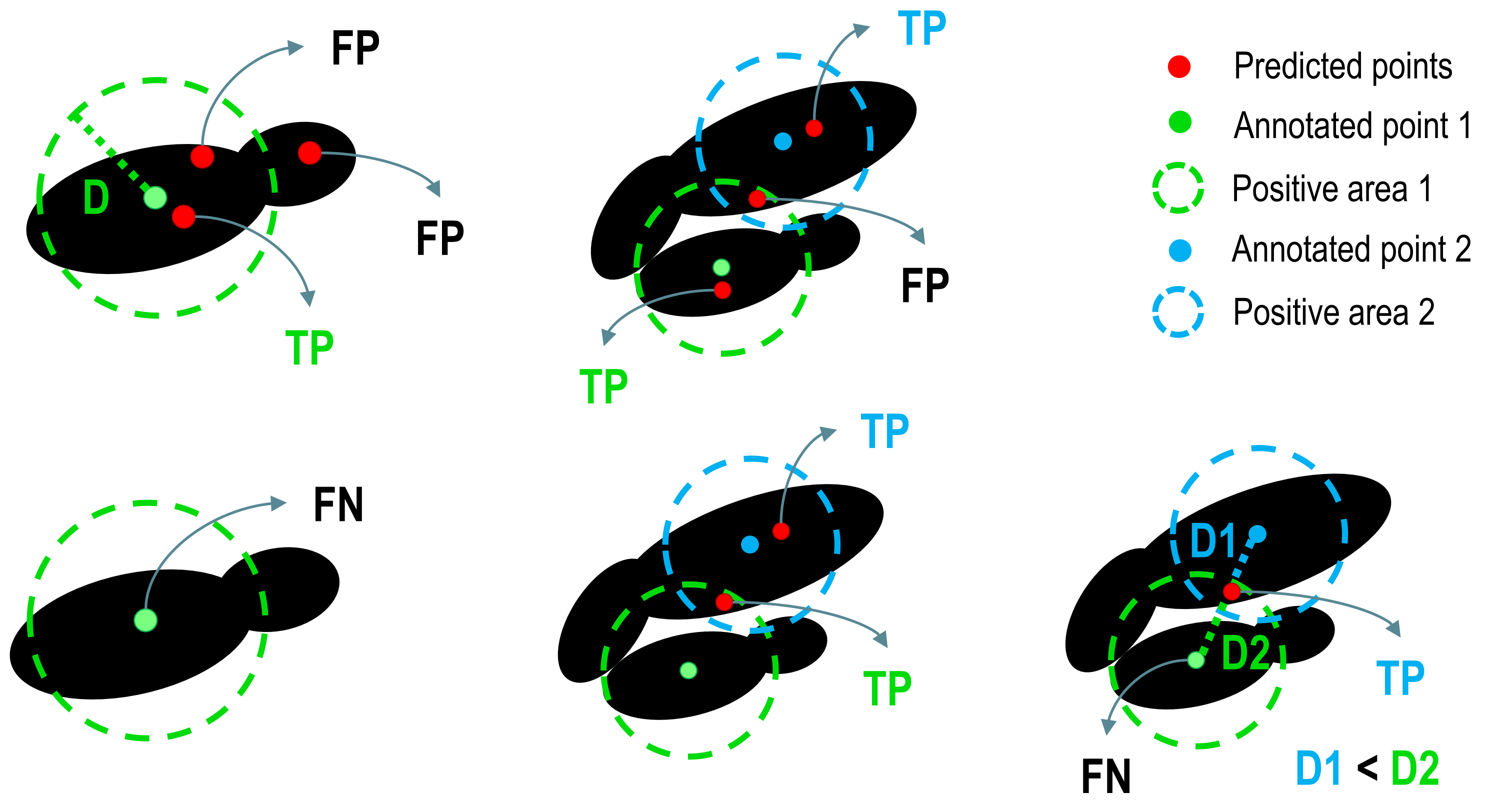
# Proposed Approach

- **Point-based single stage detector**, adapted from CenterNet[6]



# EVALUATION

- **True Positive (TP):** nearest prediction within circular area of radius D (= 20 pixels)
- **Estimated count ($\widehat{C}$):** number of predictions within image




LOCALIZATION

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

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

COUNTING

$$MAE^{a} = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{C}_i - C_i\right|$$

$$RMSE^{b} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{C}_i - C_i\right)^2}$$

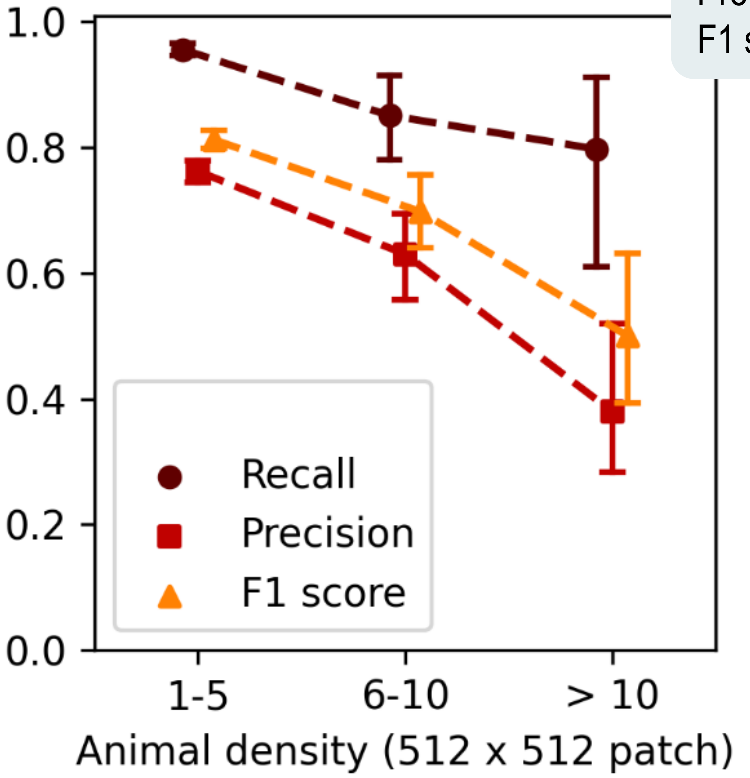
[a]Mean Absolute Error
[b]Root Mean Square Error

# LOCALIZATION PERFORMANCE

### LIBRA-RCNN

FULL IMAGE

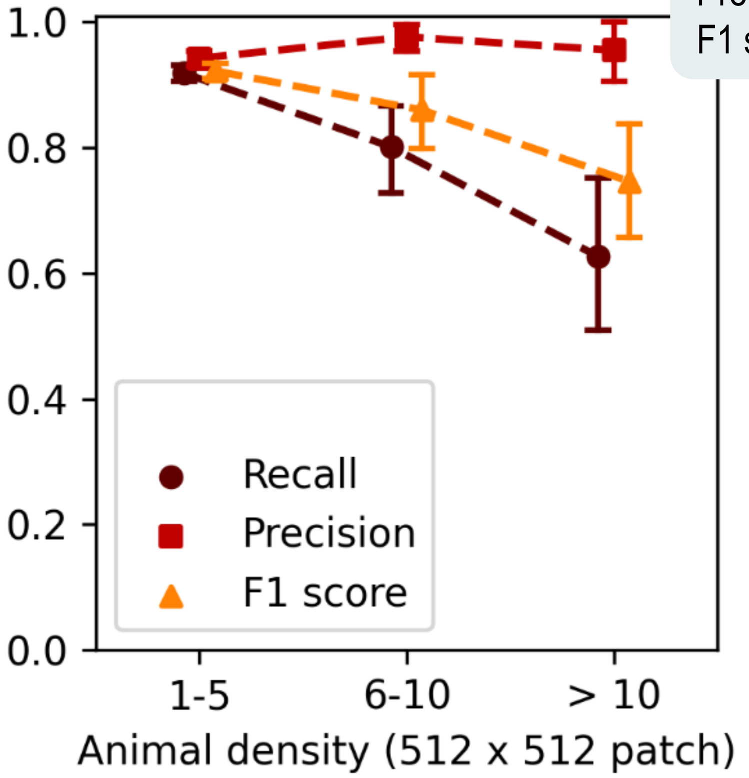
Recall = 95%
Precision = 37%
F1 score = 53%




### POINT-BASED APPROACH

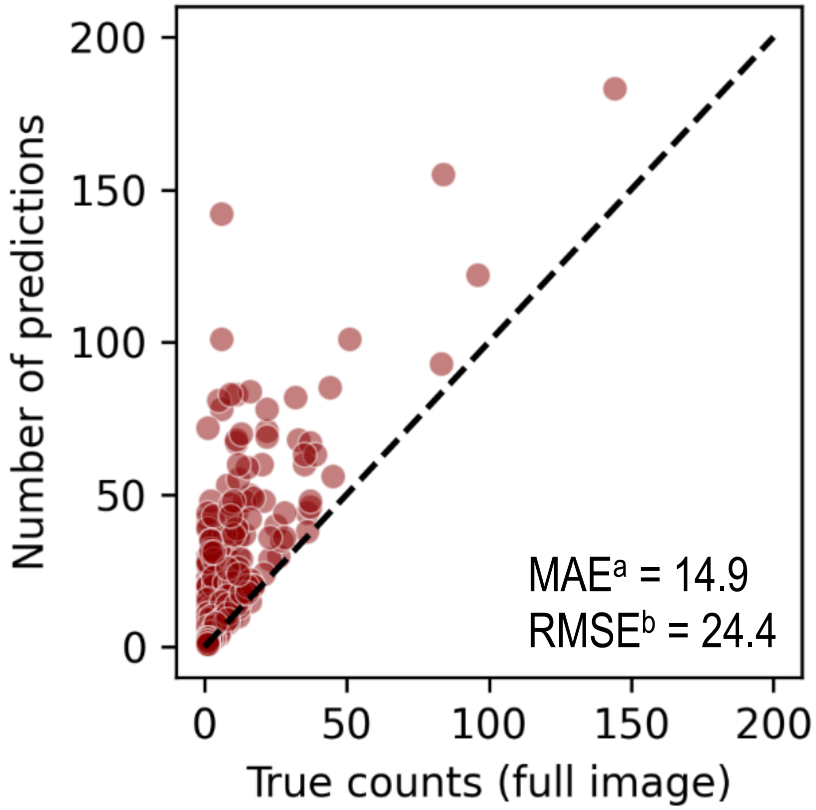
FULL IMAGE

Recall = 89%
Precision = 70%
F1 score = 78%

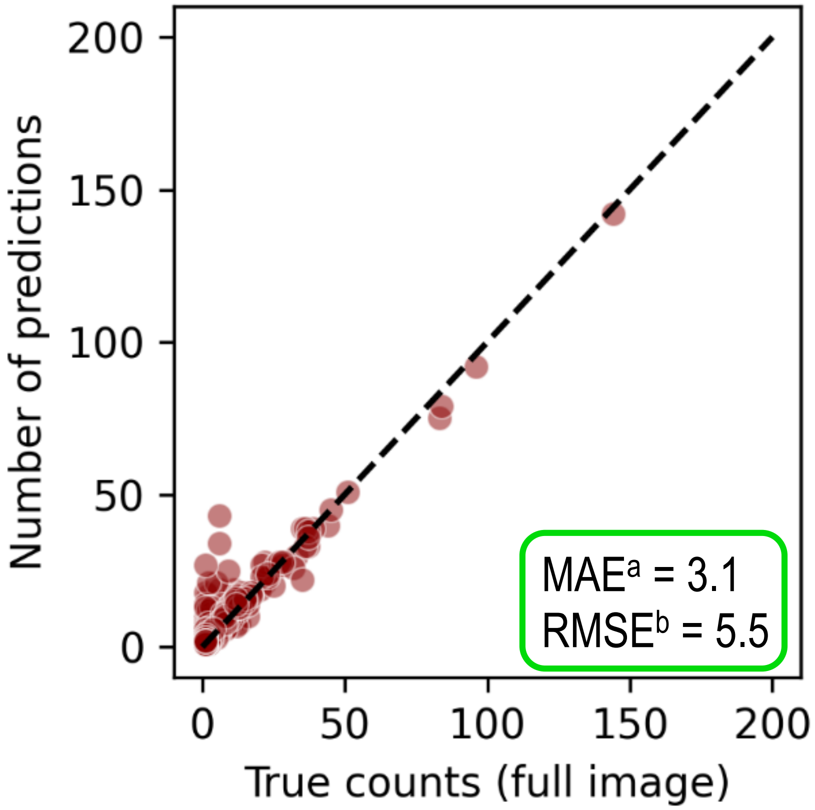


# COUNTING PERFORMANCE

**LIBRA-RCNN**

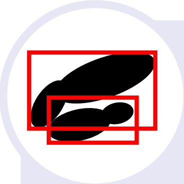



**POINT-BASED APPROACH**




[a]Mean Absolute Error
[b]Root Mean Square Error

# Conclusion & Future work



**ANCHOR-BASED DETECTOR**

- ✓ Higher recall
- ✗ Precision decreases with increasing animal density
- ✗ Over-counting
- ✗ Time-consuming annotations

**POINT-BASED DETECTOR**

- ✗ Lower recall
- ✓ High precision regardless of animal density
- ✓ Low counting error
- ✓ Faster annotations

Better adapted for animal **detection** and **counting** ?

Seems to be but need to be tested on other datasets to draw general conclusions…

# REFERENCES


[1] Delplanque, A., Foucher, S., Lejeune, P., Linchant, J. and Théau, J. (2022), Multispecies detection and identification of African mammals in aerial imagery using convolutional neural networks. Remote Sens Ecol Conserv, 8: 166-179. https://doi.org/10.1002/rse2.234

[2] Eikelboom, J.A.J., Wind, J., van de Ven, E., et al. (2019), Improving the precision and accuracy of animal population estimates with aerial image object detection. Methods Ecol Evol., 10: 1875– 1887. https://doi.org/10.1111/2041-210X.13277

[3] 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

[4] Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., & Lin, D. (2019). Libra r-cnn: Towards balanced learning for object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 821-830).

[5] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.

[6] Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as points. arXiv preprint arXiv:1904.07850.

[7] Yu, F., Wang, D., Shelhamer, E., & Darrell, T. (2018). Deep layer aggregation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2403-2412).

[8] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

[9] Liang, D., Xu, W., Zhu, Y., & Zhou, Y. (2021). Focal inverse distance transform maps for crowd localization and counting in dense crowd. arXiv preprint arXiv:2102.07925.

# THANK YOU FOR YOUR ATTENTION, ANY QUESTIONS ?

10th International Conference on Agro-Geoinformatics and
43rd Canadian Symposium on Remote Sensing

July 11 to 14, 2022

Alexandre Delplanque[1]

Samuel Foucher[2], Philippe Lejeune[1] & Jérôme Théau[2,3]

alexandre.delplanque@uliege.be

CRSS·SCT

1 LIÈGE université Gembloux Agro-Bio Tech

2 UDS Université de Sherbrooke

3 CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC

L'ENNEDI

fnrs LA LIBERTÉ DE CHERCHER