Hey ! Look some stuff of the Pandora's box of **AI**nsects

Insect detection and counting from entomological collections using deep learning methods

Grégoire Noël*[1] , Gautier Black[1], Didier VandenSpiegel [2], Patrick Semal[3], Axel Legay[4] & Frédéric Francis[1]

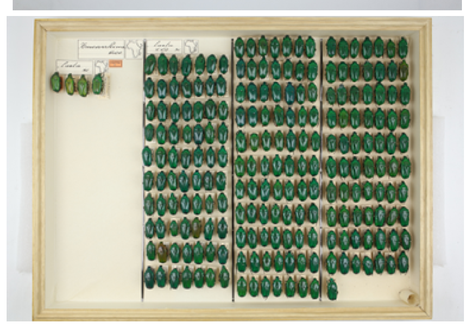
[1]Functional and Evolutionary Entomology–Gembloux Agro-Bio Tech (University of Liège), Gembloux, Belgique, Belgium
[2] Biological and Data Management Unit, Royal Museum for Central Africa, Tervuren, Belgium
[3] Royal Belgian Institute of Natural Sciences, Brussels 1000, Belgium
[4] INGI, ICTEAM, Universite Catholique de Louvain, Place Sainte Barbe 2, LG05.02,01, 1348 Louvain-La-Neuve, Belgium

- Natural history collections are invaluable repositories of biodiversity records !
- The automated detection of an insect specimen from collection box can be a difficult task owing to the remarkable morphological diversity of insects
- Convolutional neural networks (CNNs) have greatly propelled the field of computer vision, especially in object detection and segmentation
- Our central scientific question: Can we use the insect collections as detection and classification tools ? Certainly, but this need to be tested …

Workflow:

Annotation and training preparation

33 Lepidoptera boxes



52 Coleoptera boxes

CÉCI

Cenaero

YOLO V8m

Fast R-CNN

Trained models

YOLO V8m

1) coleo.pt
2) lepido.pt
3) coleo_lepido.pt

Fast R-CNN

1) coleo.pth
2) lepi.pth

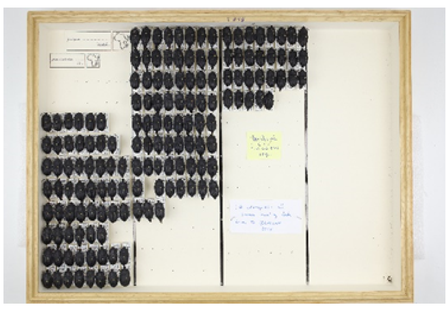
Testing

AFRICA museum

Tests

4 812 boxes

3 365 boxes

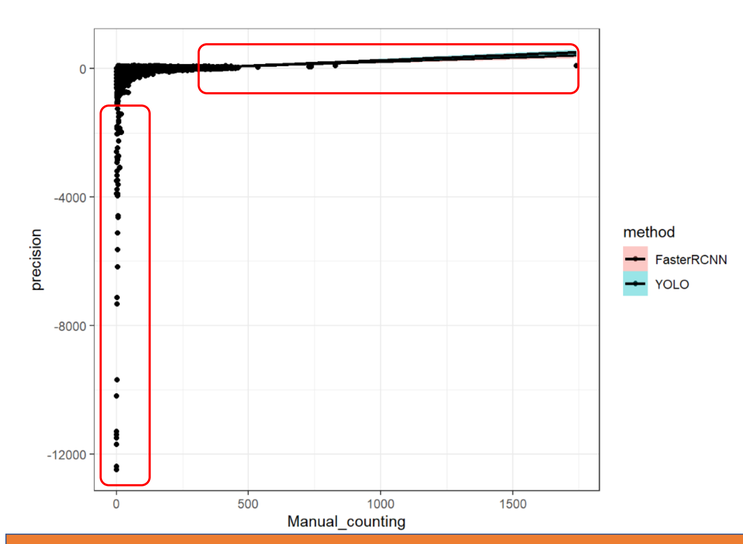


Measured variables: Counting, mean size and insect density

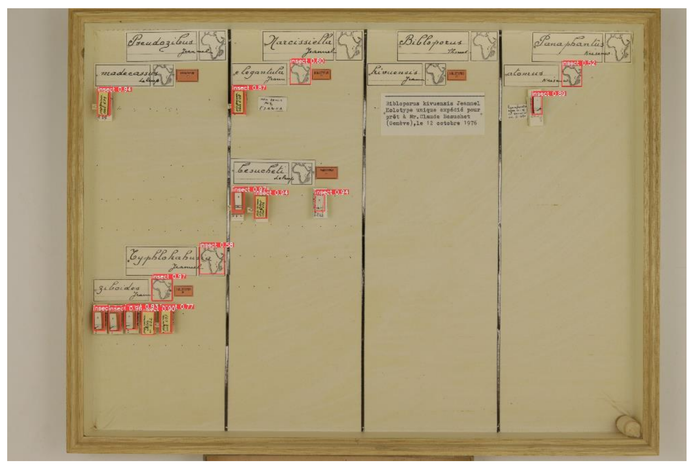
Results: Interesting but to be improved !

Table 1: Precision descriptive statistics for all models

model	method	N	precision	median	sd	se	ci
coleo.pt	YOLO	3272	-30.9966216	-0.8849558	433.3312	7.575535	14.853271
coleo_lepido.pt	YOLO	7006	-50.0284595	-6.5934066	311.9201	3.726562	7.305189
lepido.pt	YOLO	3743	-32.9366131	-8.3333333	134.0942	2.191796	4.297230
model_coleo_f2.pth	FasterRCNN	3263	-27.8434384	4.7368421	380.5738	6.662395	13.062902
model_lepi_1.pth	FasterRCNN	3743	-0.5961714	1.1363636	105.5306	1.724918	3.381870

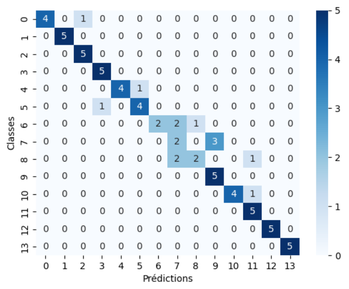


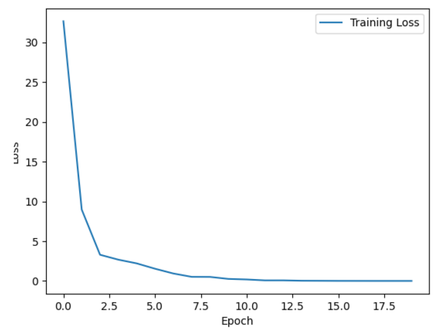
- YOLO V8m overestimates while Fast R-CNN underestimates the insect detection
- At least 20% of outlier to check
- False positive data are present due to the selection of the training datasets



Perspectives: Looks promising !

- Big data management
- Possibility to improve the insect segmentation by checking outliers and a better selection of the training datasets
- After the training, it is really fast (3h to process > 8k boxes) but need a lot of calcul resources
- Fast R-CNN seems to be better to detect insects
- Insect mean size and density per boxes still need to be tested
- At the beginning of the classification process





*Contact: gregoire.noel@uliege.be

Poster presented at International congress of AI in Entomology 03/07/2024

Open access workflow: https://github.com/gb9900/Entomo

LIÈGE université Gembloux Agro-Bio Tech
Entomologie fonctionnelle et évolutive

AI in Entomology

Royal Entomological Society