



**HAL**  
open science

## **Invertebrates Detection with YOLOv5: Towards Study of Soil Organisms Using Deep Learning**

Emma Pruvost, Hadrien Tulet, Etienne Delort, Ghulam Sakhi Shokouh, Philippe Montesinos, Baptiste Magnier, Jourdan Christophe, Belaud Emma, Hedde Mickael

► **To cite this version:**

Emma Pruvost, Hadrien Tulet, Etienne Delort, Ghulam Sakhi Shokouh, Philippe Montesinos, et al.. Invertebrates Detection with YOLOv5: Towards Study of Soil Organisms Using Deep Learning. EUVIP 2022 - The 10th European Workshop on Visual Information Processing, Sep 2022, Lisbonne, Portugal. pp.1-6, 10.1109/EUVIP53989.2022.9922675 . hal-04208778

**HAL Id: hal-04208778**

**<https://imt-mines-ales.hal.science/hal-04208778v1>**

Submitted on 10 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Invertebrates Detection with YOLOv5: Towards Study of Soil Organisms Using Deep Learning

Emma Pruvost, Hadrien Tulet, Etienne Delort, Ghulam Sakhi Shokouh, Philippe Montesinos, Baptiste Magnier  
*Euromov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France*

{emma.pruvost, hadrien.tulet, etienne.delort}@mines-ales.org, {ghulam-sakhi.shokouh, baptiste.magnier}@mines-ales.fr

Christophe Jourdan, Emma Belaud

*UMR Eco&Sols, Univ Montpellier, CIRAD, INRAE, IRD, InstitutAgro Montpellier, Montpellier, France*  
*CIRAD, UMR Eco&Sols, F-34398 Montpellier, France*  
{christophe.jourdan, emma.belaud}@cirad.fr

Mickael Hedde

*UMR Eco&Sols, Univ Montpellier, INRAE, CIRAD, IRD, InstitutAgro Montpellier, Montpellier, France*  
mickael.hedde@inrae.fr

**Abstract**—The investigation of the complicated underground life via automatic technique is in high demand in recent days. Using Convolutional Neural Network (CNN) to detect soil invertebrates is an interesting approach, although most studies on the topic have focused on other solutions. The creation of state-of-the-art technique through this work will be a significant step in soil ecology, bio-science and agriculture in effectively exploring the different types of invertebrates, their behaviors and interactions. In this paper, generating and annotating images containing seven classes of invertebrates is firstly presented. Then various automatic detections of the invertebrates using YOLOv5 algorithm on these images are performed and evaluated.

**Index Terms**—Soil fauna, invertebrate detection, small object detection, invertebrate dataset, YOLOv5, cluttered background.

## I. INTRODUCTION AND MOTIVATION

The study of biological diversity and activity in soil is a very challenging topic [1] [2]. Knowledge on organisms' activity and interactions is important, since it is at the core of ecosystem functioning. Soil organisms are involved in most of the functions that support ecosystem services to human well-being. Among them, we can cite soil physical and chemical fertility, maintaining a balanced climate and water cycling, production of food and fibers, etc. Preserving, enhancing and managing soil biodiversity is crucial for a sustainable development of our societies.

Except for some emblematic organisms (as, *e.g.*, earthworms [3] [4], ants or roots [5] [6]), there is unfortunately limited knowledge on soil fauna. Most soil organism monitoring techniques are invasive and require soil excavation (coring, hand sorting, etc.). In particular, biologists lack information on the spatial and temporal distribution of soil organisms. At fine temporal scale, there is almost no data on organisms' daily rhythms or on possible movements in the soil with regard to environmental conditions (*e.g.*, vertical migration linked to daily changes in soil conditions). On a broader temporal scale, the seasonal succession of the different taxa is very laconic and deserves more attention. Indeed, samples are taken once or twice a year, during the expected main period of activity (spring or autumn). Sampling campaigns

often depend on organizational contingencies like agricultural practices (*e.g.*, tillage or pesticide use) or availability and capacity of extraction devices (*e.g.*, number of samples in an extractor). Another major constraint in soil ecology studies is that interactions between organisms are not observed, but have to be inferred from the co-occurrence of individuals in the collected samples. For instance, a trophic interaction would be inferred from the simultaneous presence of a prey (*e.g.*, collembola) and a predator (*e.g.*, myriapoda) in a soil core. Furthermore, the soil is a heterogeneous habitat structured by aggregates and porosity, more or less connected, which limits the effective interactions.

The development of image-based monitoring of biodiversity enables to overcome these methodological constraints. The installation of waterproof scanners in fixed locations allows the acquisition of *in situ* images of soil. These image-based monitoring techniques generate larger datasets in terms of size than those usually handled [7]. Therefore, the manual analysis of the images is very time-consuming, especially since many images do not contain any organisms. In addition, there is a bias inducted by the observer with subjective decisions making about the information to extract [8]. Thus, the development of image-based biodiversity monitoring must be accompanied by the development of automated and reliable analysis tools.

In the state of the art, there is currently no work for detecting soil organisms in the soil using Neural Networks (NN) on images. Only detection by X-Ray tomography, or through acoustic systems could be found as in [9]. Besides, some recent work has shown progress in invertebrate automatic detection, but only out of the soil [10].

In this contribution, efforts have been made to automate the detection of objects of interest (soil fauna) in images of truffle soil with little variation in humidity. All the images used are scans of the same soil area, therefore the background is almost the same in all of the images. The YOLOv5 [11] algorithm has been used and trained to automatically detect the invertebrates, and assign a label to each object. This involves image dataset creation, annotations, feeding the annotated data

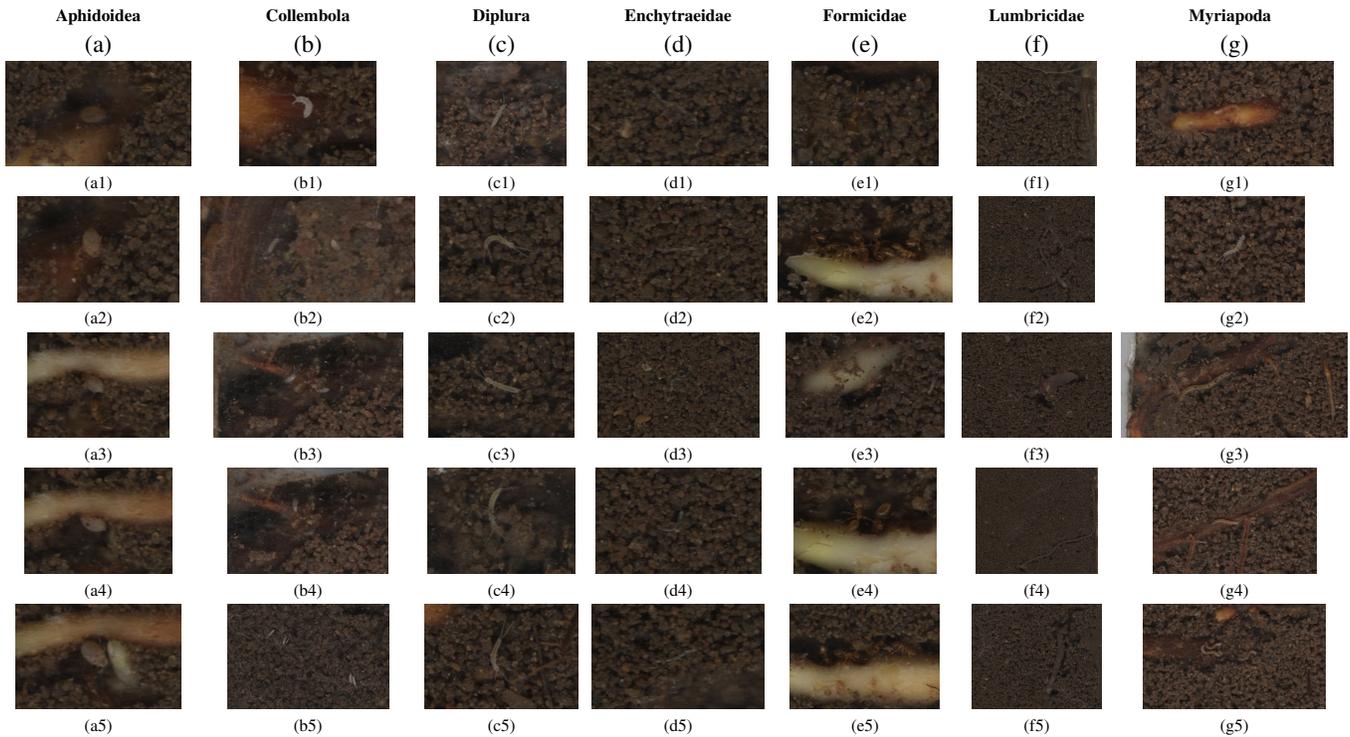


Fig. 1. Organism types/classes demonstrating the inter and intra class variation and complexities.

to the deep neural network (here YOLOv5 models), training the network and consequently detect the invertebrates with the correct labels. The performance of the proposed system is evaluated with standard evaluation metrics, and the obtained results are in free form although usable by non-computer specialists.

The main purpose of this work consists in showing pioneer results in automatic invertebrate detection in soil images. This contribution paves the way for further progress in this field. There are 2 main challenges in this study:

- The size of the invertebrates is very small compared to the image resolution: few pixels are available for the model to deduce characteristics from each type of invertebrates.
- The images contain cluttered backgrounds, which makes it harder for the model to make reliable predictions.

## II. DETECTION METHOD USING YOLOV5 ALGORITHM

The choice was made to do supervised machine learning, in other words: images were annotated to train the model.

### A. Dataset Creation

Scanners of size A4 were placed underground for the acquisition of the source images. Each scanner takes an image every six hours, all year round, from 2015 to 2018. The obtained images are TIFF files of resolution  $10192 \times 14039$  pixels. The initial resolution is too large to easily spot the invertebrates, and, more importantly, to be able to feed it as input to the NN, here YOLOv5. To work around the problem, the images were sliced into 90 sub-images of size  $1019 \times 1559$ .

A ground truth dataset of the pre-processed fauna images has been created using Plainsight™ software. The annotations were limited to the 7 most observed types of invertebrates, namely: myriapoda, aphidoidea, diptera, enchytraeidae, lumbricidae, formicidae, and collembola as depicted in Fig. 1. More than 600 images were annotated for a total of 908 organisms' landmarks. 10% fauna-free images were also added to the annotated images afterwards to improve the model performances. As a rule of thumb, the dataset has been divided into the training (80% of the images in the dataset), validation (10%) and test (10%) sets.

Nevertheless, this dataset has a few downsides that should be highlighted. The background of the images could be qualified as “noisy”, due to the granularity of the soil, creating a cluttered background, as illustrated in Fig. 1. Another difficulty is that some plant structures are similar to invertebrates (myriapoda are sometimes hard to distinguish from roots). It is also important to note that in the dataset, the number of organisms is different for each category. Some groups such as aphidoidea, collembola or lumbricidae are poorly represented, creating a problem called class imbalance in machine learning. In that respect, the distribution of the labelled organisms dataset is illustrated in Fig. 2.

### B. Evaluation Metrics

As a base to evaluate models and define evaluation metrics, 4 quantities need to be set:

- True Positive ( $TP$ ): correctly detected object.
- False Positive ( $FP$ ): incorrectly detected object.
- True Negative ( $TN$ ): correctly undetected object.

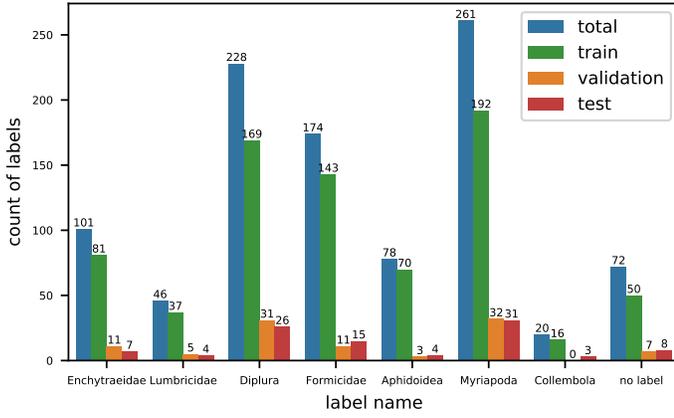


Fig. 2. Composition of the full dataset for train, validation and test.

- False Negative (*FN*): incorrectly undetected object.

Based on these 4 logical values, many evaluation metrics are defined, such as *Precision* and *Recall*:

$$Precision = \frac{TP}{TP + FP} \quad \text{and} \quad Recall = \frac{TP}{TP + FN}. \quad (1)$$

The choice of which metrics to optimize is made according to the desired final application of the neural network.

Here, the *Recall* measure, which measures the proportion of correctly found organisms among all organisms that can be found, was prioritized to evaluate the models. Maximizing this magnitude allows the model to identify as many organisms as possible, even if it means showing false positives. Indeed, the aim of automatically detecting soil fauna here is to gain time on filtering the huge number of images sent by the scanners. The model has to sort out the images containing objects of interest, but these images are all meant to be reviewed by the researchers who need to get information from it. However, it is also important to maintain a reasonable *Recall/Precision* compromise. Indeed, if too many false positives are detected (as roots or granular soil), the time gained thanks to the pre-processing done by the model is lost.

### C. YOLOv5 Description

The chosen CNN model for this study is the YOLOv5 offered by Ultralytics [11]. YOLOv5 is the benchmark model among the current deep CNN object detection architectures (e.g., Faster RCNN, RetinaNet, SPP-net, etc.) thanks to its ease of use and adaptability to specific problems. These relevant factors motivated the choice of YOLOv5 considering the problematic of small and complex object detection.

YOLOv5 is the fifth version or the YOLO (You Only Look Once [12]) algorithm, which is a reliable detector based on Deep Convolutional Neural Network and remains commonly used for real-time object detection. YOLOv5 is a family of object detection architectures pre-trained on the COCO dataset [13]. Each YOLOv5 architecture is a network consisting of three parts: (i) Backbone: CSPDarknet, (ii) Neck: PANet, and (iii) Head: YOLO Layer. The data are first input to

CSPDarknet for feature extraction, and then fed to PANet for feature fusion.

There are 5 different YOLOv5 categories of models, corresponding to 5 levels of architectural complexity. By increasing complexity, these versions are 5n, 5s, 5m, 5l, 5x. For invertebrates' detection, the 5n and 5s models are not suitable, as will be explained later in Sec. II-E. Moreover, 5x are very complex, and therefore, difficult to train on small machines (calculation time and memory required). It is better to discard them at the start of the experiments. In this work, mainly the 5l model was used.

### D. Areas of Optimization

Many ways of affecting the model's performances were identified. Accordingly, multiple parameter combinations were tested to get the best results from the YOLOv5 models.

However, since the dataset was quite small, and since deep learning model performances improve with the size of the datasets (i.e., as data hungry systems, the more the amount of data is high, the better the results are), transfer learning was used. The weights of pre-trained models of YOLOv5 were thus initially taken to optimize the results of the training on our custom dataset. Moreover, due to computer calculation limits, the largest models from YOLOv5 such as yolov5l6 with images of size 1280×1280 as inputs could not be used, and the batch size was limited to 3. Trainings were also not always taken to over-fitting phase because they were very time-consuming (14 hours for 50 epochs).

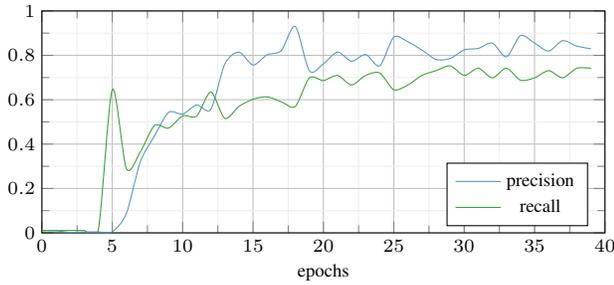
Among the different options that could be tested through the training, there is:

- The complexity of the neural network (YOLOv5s, YOLOv5l, etc.).
- The size of the input images, by cropping initial input images in order to see the effect of image compression by the model. Indeed, when an image has too large dimensions, the model compresses it until both dimensions respect the model's size limit.
- The number of different objects to detect (detecting all 7 invertebrates, removing enchytraeidae, or detecting all the invertebrates as Invertebrates).
- The value of YOLOv5's scale hyper-parameter, which is related to data augmentation to process the images with different scale ratio.
- The Initial Learning Rate (LR0) hyper-parameter of YOLOv5 that controls the speed of convergence of the gradient descent.

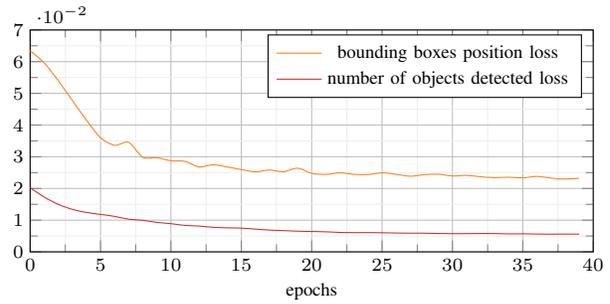
To get a quick idea of the impact of the different parameters, tests on a smaller dataset containing about 270 images were carried out. The number of epochs for these tests was also usually lowered to 25. The final models, presented in more details in this paper, were however trained with 50 epochs on the full dataset to get better results.

### E. Performed Optimization

Concerning the hyper-parameters, two tests were carried out to make a first choice for the values of the scale and initial



(a) Precision and Recall on validation set



(b) Loss measures on validation set

Fig. 3. Metrics for single invertebrate detection model on validation set.

learning rate. Two trainings were performed with similar conditions, but had a different scale value:  $scale = 0.4$  gave better results than  $scale = 0.5$ . The same way, two trainings with different initial learning rate (LR0) values were performed: the comparison between the values  $10^{-3}$  and  $10^{-2}$  was not conclusive because of the limitations of YOLOv5s model used for this experiment. Undoubtedly, there are opportunities for further optimization with the hyper-parameters tuning of both the initial learning rate and scale, but also of all the other available hyper-parameters (learning rate finder, momentum, weight decay, etc.) that were not investigated here.

Meanwhile, better results were obtained with a single label to predict, or when enchytraeidae were removed. The comparison can be made with the results presented in the Figs. 5 and 6, see Sec. III. The choice of the number of labels to detect depends on what will be preferred for network application.

Furthermore, concerning the effect of cropping the images, only a single comparison test could be performed. The results were quite similar between cropped images and original ones, but this is probably due to the chosen scale of 0.4 that renders an image of the same format as the cropped ones. The influence of this parameter along with the scale value could be studied further.

Usually, the analysis of the outputs when distinguishing the 7 soil fauna groups shows that the results (both in *Precision* and *Recall*) are better for myriapoda and diplura. There is an under-representation of aphidoidea, collembola, lumbricidae and technical difficulties in locating enchytraeidae and formicidae to watch out for.

### III. RESULTS: TRAINING, VALIDATION AND TEST

In this section, the results for a model detecting all 7 invertebrates' types, as well as the results for a model detecting a single label are presented. Both were trained on the entire dataset (over 500 images); however, the latter gets, as expected, better results for training. Detection on test images confirms the gap of performances. However, the usage meant for the model in this study implied to keep the detection of the 7 invertebrates, which is why both models' results are further developed here.

#### A. Training and Validation

##### 1) Single Invertebrate Detection:

This training was made with the following parameters:

- The model YOLOv5l6 was used with an image size parameter of 640.
- 50 epochs of batch size 3 were run.
- An initial learning rate of 0.001 and a scale of 0.4 were chosen for the hyper-parameters.

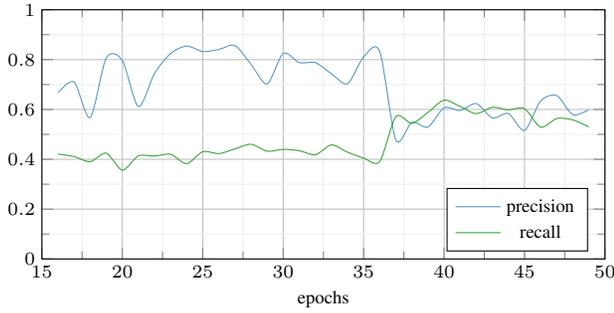
The Fig. 3 shows the evolution of *Precision*, *Recall* and multiple loss metrics during the training of the single invertebrate detection. Due to computer limitations, the training had to be divided into 2 sequences. On Fig. 3, the results for the first part (40 out of 50 epochs) are displayed. From the Fig. 3(b), i.e., the loss measures, it can be deduced that the training could be continued since the Loss measures are still decreasing. This was also the case when the 50th epoch was run. In Fig. 3(a), the measure *Recall* is almost increasing as fast as *Precision*, but stays slightly inferior. The model converges to a *Precision* of about 0.8 and a *Recall* of 0.7 according to the measures on the validation set. Since the metrics were not improving a lot after the 10th epoch, and because the experiment was very time-consuming, the training was interrupted at the 50th epoch before reaching over-fitting. This proves that the results could be better. However, the slow improvement of the metrics indicates that the main limitation to get a good detection model is the quality and quantity of labeled images (as expected).

##### 2) Multiple Invertebrates' Detection:

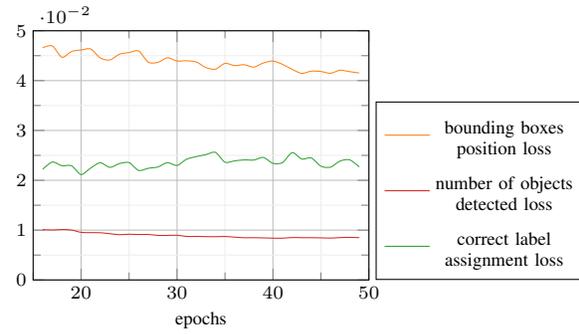
This training was made with the following parameters:

- The YOLOv5l algorithm was used with an image size parameter of 640.
- 50 epochs of batch size 3 were run.
- An initial learning rate of 0.001 and a scale of 0.4 were chosen for the hyper-parameters.

As for the single invertebrates' detection, the training of the 50 epochs has been divided into 15 and 35 epochs. The measures on the validation set for the last 35 epochs are displayed in Fig. 4. Similar observations on the evolution of the curves can be made: the biggest augmentation for *Precision* and *Recall*, and the biggest decrease in losses happen during the 10 first epochs. For the 35 epochs presented, the losses keep on decreasing, so the model is still not over-fitted. However, the evolution for *Precision* and *Recall* is interesting to focus on for the last epochs: the *Precision* suddenly drops while the *Recall* rises to the same value as *Precision*. No



(a) Precision and Recall on validation set



(b) Loss measures on validation set

Fig. 4. Metrics for 7 invertebrates' detection model on validation set.

explanation to this observation could be found. Nevertheless, it indicates that even with multiple object detection, the model is able to improve itself on not missing the invertebrates. At the end of the training, *Precision* and *Recall* are both at about 0.6, which is as expected lower than the results for single object detection. It should be reminded that, as mentioned in Sec. II-E, the global results could be already better with the same training conditions if the enchytraeidae were removed.

## B. Test

The curves displayed in this section were obtained with the test set.

### 1) Single Invertebrate Detection:

Figs. 5(a) and (b) represent respectively *Precision* and *Recall* results by the *Confidence* threshold. The *Confidence* threshold can be chosen afterwards, depending on the application of the model by analyzing those 2 curves. The higher the *Confidence* threshold, the lower the *Recall* (and *FP*) and the higher the *Precision* (and *FN*). Here, the *Confidence* threshold does not need to be higher than 0.72. Indeed, according to Fig. 5(a), above this value all the invertebrates detected are assigned to the right label. In opposition, Fig. 5(b) shows that even by lowering the *Confidence* threshold to 0, all invertebrates will not be detected by the model. This underlines the limitations of doing object detection on small objects with such a cluttered background.

To get equivalent results for both metrics, a *Confidence* threshold around 0.35 can be set. This enables to get a *Precision* and a *Recall* both of approximately 0.7. The Fig. 5(c) can help to make a compromise between *Precision* and *Recall*. It can also be noticed that this result is coherent with the results obtained on the validation set, which confirms that the model is not over-fitted.

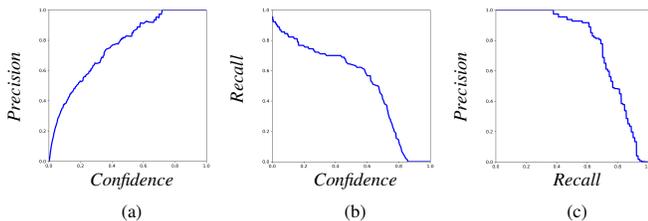


Fig. 5. Metrics for single invertebrate detection model on test set.

### 2) Multiple Invertebrates' Detection:

In Fig. 6, *Precision - Confidence*, *Recall - Confidence* and *Precision - Recall* curves were also used to evaluate the 7 invertebrates' detection model. As mentioned during the description of the dataset, there were not enough images containing collembola, lumbricidae and even aphidoidea. This explains why few points are available in Figs. 6 for those invertebrates. The results for these 3 animals are thus not reliable. It is also interesting to notice that for enchytraeidae, the smallest studied species, *Precision* and *Recall* are always lower than the average. Meanwhile, for formicidae, the second smallest species, which are more represented in the dataset as the enchytraeidae, the results are better than average. This confirms that the NN is hindered by both too small objects to detect, and objects that are difficult to distinguish because of the noisy background. Meanwhile, it confirms that improvements can be obtained by increasing the population of a class. The best results are obtained for myriapoda, and results for diplura are quite similar to those of formicidae. Their proportion in the dataset was the highest, and they are both not as small as enchytraeidae.

A better dataset and further training would be needed to see how much the constraints linked to the detection of invertebrates such as enchytraeidae is preventing from getting better results. At the moment, the model barely enables to get a compromise of 0.5 in *Precision* and *Recall* for them (see Fig. 6(c)). In opposition, myriapoda, which gets the best results, can be detected with a *Precision* and *Recall* of almost 0.8. A similar (and even better) result could be expected as the average of all invertebrates' detection, if their sizes were close to the myriapodas' one, if the dataset was bigger, the categories better distributed, and if the training was further continued.

Some visual output examples of the solution developed are shown in Fig. 7. These images are cropped from the original outputs, being images of size  $14039 \times 10192$ . *FN* were difficult to find with a *Confidence* threshold equal to 0.25 for instance, which is the default threshold of our solution (to prioritize *Recall* over *Precision*). For the examples of Fig. 7, a *Confidence* threshold of 0.5 has been chosen. Furthermore, the values displayed near the labels of the bounding boxes are the *Confidence* with which the model has made the prediction.

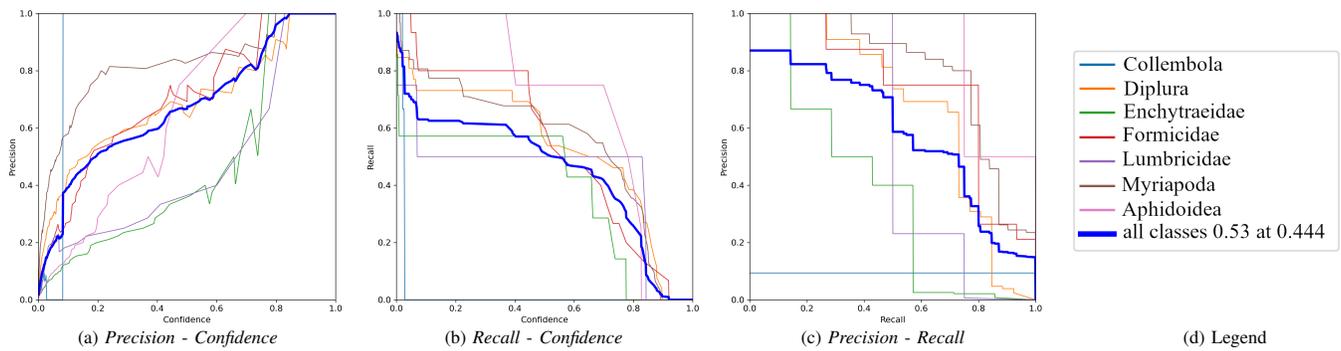


Fig. 6. Metrics for 7 invertebrates' detection model on test set.

#### IV. CONCLUSION

In this paper, the automatic detection of invertebrates in soil images using YOLOv5 has been tested and evaluated. This pioneer work sets a baseline to deepen the study of soil fauna via image analysis combined with CNNs. The best model obtained for the single object detection is characterized by values of 0.7 for both *Precision* and *Recall*. It is not satisfying enough for the outputs of the model to be reliable. More importantly, it does not fully satisfy the needs of underground biology researchers, since it still makes a significant number of errors, and it does not distinguish species. However, this result is higher than what was first expected, and we are confident for further research and testing to produce usable results. Indeed, there are several ways that could much improve the results. Firstly, other CNN frameworks may be more adapted. Secondly, even with YOLOv5, a broader optimization of the parameters, the usage of fine-tuning, as well as increasing the number of epochs for the training might enhance the detection quality a lot. Finally, increasing the dataset size, balancing the labels, and improving the annotating protocol with an Inter-Annotator Agreement (that could help to minimize annotation errors due to wide images, cluttered background, invertebrates having similar shapes and colors as roots or granular soil) would be great improvements.

#### REFERENCES

[1] G. Centenaro, C. Hudek, A. Zanella, and A. Crivellaro, "Root-soil physical and biotic interactions with a focus on tree root systems: A review," *Applied Soil Ecology*, vol. 123, pp. 318–327, 2018.

[2] M. Bonkowski, C. Villenave, and B. S. Griffiths, "Rhizosphere fauna: the functional and structural diversity of intimate interactions of soil fauna with plant roots," *Plant and Soil*, vol. 321, pp. 213–233, 2009.

[3] V. Nuutinen and K. R. Butt, "The mating behaviour of the earthworm *lumbricus terrestris* (oligochaeta: Lumbricidae)," *Journal of Zoology*, vol. 242, pp. 783–798, 1997.

[4] T. Djerdj, D. K. Hackenberger, D. K. Hackenberger, and B. K. Hackenberger, "Observing earthworm behavior using deep learning," *Geoderma*, vol. 358, p. 113977, 2020.

[5] E. Han, A. G. Smith, R. Kemper, R. White, J. A. Kirkegaard, K. Thorup-Kristensen, and M. Athmann, "Digging roots is easier with ai," *Journal of experimental botany*, vol. 72, no. 13, pp. 4680–4690, 2021.

[6] A. G. Smith, J. Petersen, R. Selvan, and C. R. Rasmussen, "Segmentation of roots in soil with u-net," *Plant Methods*, vol. 16, no. 1, pp. 1–15, 2020.

[7] T. Hoye, J. Arje, K. Bjerger, O. Hansen, A. Iosifidis, F. Leese, H. Mann, K. Meissner, C. Melvad, and Raitoharju., "Deep learning and computer vision will transform entomology," in *Proceedings of the National Academy of Sciences 118*, 2021.

[8] A. Dell, J. Bender, K. Branson, I. Couzin, G. Polavieja, L. Noldus, A. Perez-Escudero, P. Perona, A. Straw, M. Wikelski, and U. Brose, "Automated image-based tracking and its application in ecology," *Trends in Ecology and Evolution*, p. 417–428, 2014.

[9] R. Mankin, "Current and potential uses of acoustic systems for detection of soil insect infestations," *J. of The Acoustical Society of America*, 1998.

[10] T. Kasinathan, D. Singaraju, and S. R. Uyyala, "Insect classification and detection in field crops using modern machine learning techniques," *Information Processing in Agriculture*, vol. 8, no. 3, pp. 446–457, 2021.

[11] G. Jocher, "ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements," *Zenodo*, 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.4154370>

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE CVPR*, 2016, pp. 779–788.

[13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*. Springer, 2014, pp. 740–755.



Fig. 7. Examples of invertebrates' detection in underground scanned images: from left to right, true positives (myriapods and diplura –diplora–), false positives (2 roots detected as diplura), and false negative (undetected myriapod, circled in red). Images of size 449×760 , 434×908 and 356×573 respectively.