Smart Agricultural Technology 7 (2024) 100405

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

YOLO performance analysis for real-time detection of soybean pests

Everton Castelão Tetila [a,e,*], Fábio Amaral Godoy da Silveira [a], Anderson Bessa da Costa [a], Willian Paraguassu Amorim [a], Gilberto Astolfi [b], Hemerson Pistori [c,d], Jayme Garcia Arnal Barbedo [e]

[a] *Universidade Federal da Grande Dourados, FACET, Rodovia Dourados-Itahum, Km 12, Dourados, 79825-070, Mato Grosso do Sul, Brazil*
[b] *Instituto Federal de Mato Grosso do Sul, Informática/Eng. de Software, R. Taquarí, 831 - Santo Antonio, Campo Grande, 79100-510, Mato Grosso do Sul, Brazil*
[c] *Universidade Federal de Mato Grosso do Sul, FACOM, Av. Costa e Silva - Pioneiros, Campo Grande, 79070-900, Mato Grosso do Sul, Brazil*
[d] *Universidade Católica Dom Bosco, Inovisão, Av. Tamandaré, 6000. Jardim Seminário, Campo Grande, 79117-900, Mato Grosso do Sul, Brazil*
[e] *Empresa Brasileira de Pesquisa Agropecuária, Embrapa Agricultura Digital, Av. Dr. André Tosello, 209 - Cidade Universitária, Campinas, 13083-886, São Paulo, Brazil*

A R T I C L E I N F O

Editor: Stephen Symons

Keywords:
Deep learning
Object detection
Precision agriculture
Soybean pests

A B S T R A C T

In this work, we evaluated the You Only Look Once (YOLO) architecture for real-time detection of soybean pests. We collected images of the soybean plantation in different days, locations and weather conditions, between the phenological stages R1 to R6, which have a high occurrence of insect pests in soybean fields. We employed a 5-fold cross-validation paired with four metrics to evaluate the classification performance and three metrics to evaluate the detection performance. Experimental results showed that YOLOv3 architecture trained with a batch size of 32 leads to higher classification and detection rates compared to batch sizes of 4 and 16. The results indicate that the evaluated architecture can support specialists and farmers in monitoring the need for pest control action in soybean fields.

1. Introduction

Soybean (*Glycine max*) is a plant belonging to the Fabaceae family used in human food (in the form of soy oil, tofu, soy sauce, soy milk, soy protein, soy beans) and animal feed (in the preparation of rations). It is a grain with a good nutritional profile and high economic importance [1]. In the 2019/2020 harvest, Brazil became the world's largest producer of the grain, surpassing its main competitor, the United States. In the 2022/23 harvest, Brazil achieved a production of 154.6 million tons in a planted area of approximately 44 million hectares, with an average productivity of 3,508 kg/ha, breaking historical records for planting area, productivity and production [2].

It is estimated that most direct costs in soybean production are concentrated in fertilizers (27.82%), pesticides (18.24%), machine operations (9.10%), seeds (7.35%) and depreciation of machines and equipment (6.76%) [3]. Since pest control represents a large portion of the production costs, methods capable of making the process more targeted and efficient can result in sizeable cost reductions.

Monitoring pests in all stages of soybean development allows for a more efficient use of pesticides, since inputs can be directed to the right spots and applied in the proper amounts, thus reducing production costs and the environmental impact resulting from the excess use of chemical control, in addition to contributing to human health and food safety [4]. Sampling methods such as beating cloth, sweeping net, visual examination of plants and examination of soil samples have been used to monitor pests and their damage to crops [5].

Regardless of the method adopted, in order to assess the pest infestation in the crop, it is suggested that the number of insects should be recorded at a sufficiently large number of sampling points in order to enable the assessment of the level of infestation in each crop area. The greater the number of samples taken in an area, the greater the certainty of a correct pest infestation prediction. In general, at least six samplings are recommended for crops of up to 10 ha, eight for crops of up to 30 ha and ten for crops of up to 100 ha. For larger areas, dividing into 100 ha plots is recommended [5].

As an alternative to manual sampling methods, there is a growing motivation for using digital images collected in the field, for the devel-

* Corresponding author at: Universidade Federal da Grande Dourados, FACET, Rodovia Dourados-Itahum, Km 12, Dourados, 79825-070, Mato Grosso do Sul, Brazil.

E-mail address: evertontetila@ufgd.edu.br (E.C. Tetila).

https://doi.org/10.1016/j.atech.2024.100405
Received 13 November 2023; Received in revised form 20 December 2023; Accepted 22 January 2024
Available online 1 February 2024

opment of computer vision systems aimed at pest detection. In addition, high spatial resolution cameras can be embedded in Unmanned Aerial Vehicles (UAV) offering a bird's eye view of the crop. However, most studies in the literature are dedicated to pest detection in traps, with only a few focusing on direct detection in the crops.

In this work, we evaluate the YOLOv3 [6] architecture applied to the detection of insect pests in a real field environment, under different lighting conditions, object sizes and backgrounds. The problem of identifying the object's location along with its class is called object detection. This can be achieved by training a multi-label classifier to determine the location and class of each object (encircled in a rectangle called bounding box). The YOLO is one of the most employed object detection models, as it offers accurate results with a relatively low computational burden.

We collected images of the soybean plantation in different days, locations and weather condition, between the phenological stages R1 to R6, which have a high occurrence of insect pests in soybean fields. For training and testing the neural network, we used a 5-fold cross-validation associated with four metrics to evaluate the classification results (precision, recall, F-measure and accuracy) and three metrics to evaluate the detection results (mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2)).

This work is organized as follows. Section 2 presents related work and Section 3 describes important details on materials and methods. Section 5 shows the experimental results, followed by a discussion. Finally, Section 6 presents the conclusions and notes for future work.

2. Related work

In the literature there are several studies comparing image sensors and machine learning methods for pests identification in various crops. Despite this, few studies address the detection of insect pests in the field, using images collected under real conditions. This section presents articles published between 2017 and 2023 that address classification and detection of pests.

2.1. Methods for image classification

Within the classification task, Li et al. [7] created a framework that identifies ten types of pests from different crops. The dataset was collected by downloading a total of 5,629 images from search engines and filming outdoors using an Apple 7 Plus smartphone. In the data preparation phase, GrabCut and Watershed algorithms were implemented to remove complex background from images. Among the tested models, GoogLeNet yielded the best results. On the other hand, GoogLeNet's training was more computationally intensive than other models.

The deep residual learning method was used in [8] to identify 10 classes of agricultural pests in images with complex backgrounds. The method's performance was improved after optimizing, by deep residual learning, the pre-trained ResNetlOl and ResNet50 models on ImageNet. The proposed model achieved 98.67% accuracy, which was significantly higher than traditional SVM and CNN. Despite the promising results, the deep residual net could have been combined with object detection methods such as Faster R-CNN or R-FCN to track pest targets in real time, offering greater practical value of the method in agricultural pest control tasks.

Liu et al. [9] designed an intelligent autonomous vehicle to acquire images in the natural setting of the farm, and also proposed a Pyralidae pest recognition algorithm. Specifically, by employing the color and shape characteristics of Pyralidae pests, they proposed a targeting algorithm using Inverse Histogram Mapping and Restricted Spatial Otsu methods to target pests. Next, they designed a recognition approach based on Hu Moment Invariant. The proposed method achieved an accuracy of 94.3%. However, the average processing time of each frame was greater than 1 second, delaying the robot's response to the observation results. Pest detection under non-uniform lighting was also not effective.

A dataset containing more than 75,000 images from 102 pest species was introduced in [10]. Eight shallow and four deep models were tested. Experimental results showed low accuracy (<50%) in almost all scenarios, indicating that shallow and deep methods still cannot cope with a large number of classes with a high degree of data imbalance.

Tetila et al. [4] and Tetila et al. [11] evaluated different models of deep learning trained with fine-tuning and transfer learning parameters for tasks of classification and counting of insect pests in soybean fields. First, an image segmentation step with the SLIC Superpixels method was considered to segment the insects in the image. In the classification stage, DenseNet-201 yielded the best results, with an accuracy of 94.89%. In the counting experiment, superpixels were classified using the weights of the best-rated CNN, with counting estimates achieving an accuracy of 90.86%. The results were not compared with other state-of-the-art detection methods.

2.2. Object detection methods

A novel cascade pest detection approach called DeepPest, based on two-stage moving vision, was proposed in [12] for recognition of very small-sized pest species in unbalanced datasets. DeepPest, extracts multi-scale contextual information from images as background knowledge in order to build a contextual information network for initial classification of images into crop categories. Then, a multiprojection pest detection model (MDM) was trained by crop-related pest images. Although the dataset contains 17,192 pest images captured in the field environment with 76,595 pest annotations, the dataset is limited to wheat and rice crops, and no pest images were captured on soybean.

Nam and Hung [13] evaluated three detection and classification methods to identify five insect species trapped in a factory environment. Using a set of 200 original images of 3,026 insects, the following methods were compared: (1) Adaptive Threshold combined with VGG-16, (2) Single Shot MultiBox Detector (SSD) built on top of VGG16 network but replacing fully connected layers, (3) VGG16 with Sliding Window approach. In the experiments, SSD produced the best results, achieving detection and classification rates of 84% and 86%, respectively. In this work, data augmentation was performed by combining the inversion and rotation operations on each photo to obtain more training and testing samples. Because the augmentation was applied before the division into training and test sets, the results reported are likely biased.

In [14] an improved network architecture based on VGG19 was implemented for detection and classification of 24 insect species collected in crop fields such as rice and soybean. The method achieved an accuracy of 89.2% in the MPest dataset, being superior to the traditional state-of-the-art SSD (85.3%) and Fast R-CNN (79.6%) methods. It is worth pointing out that many relevant pest species found in soybean were not considered in this study.

The main focus of the work of Gutierrez et al. [15] is the selection of the best approach for detection and identification of harmful pests in tomato and pepper crops in greenhouse. A dataset with a large number of images of infected tomato plants was created to generate and evaluate machine learning (MLP and k-NN) and deep learning (SSD and Faster R-CNN) models. Deep learning yielded the highest accuracy in distinguishing between *Bemisia tabaci* and *Trialeurodes vaporariorum*. Results also indicated that the detection and identification of eggs is a major challenge. Important evaluation metrics like accuracy and F-measure, were absent in the report.

To carry out a rapid detection and recognition of ten types of insects that affect tea fields in China, Deng et al. [16] created the SIFT-HMAX model inspired by the human visual attention mechanism. First, the Saliency Using Natural (SUN) statistical model was used to generate saliency maps and detect region of interest (ROI) in a pest image. To extract the attributes that represent the appearance of the pest, the Hierarchical Model and X (HMAX) model was improved. The proposed

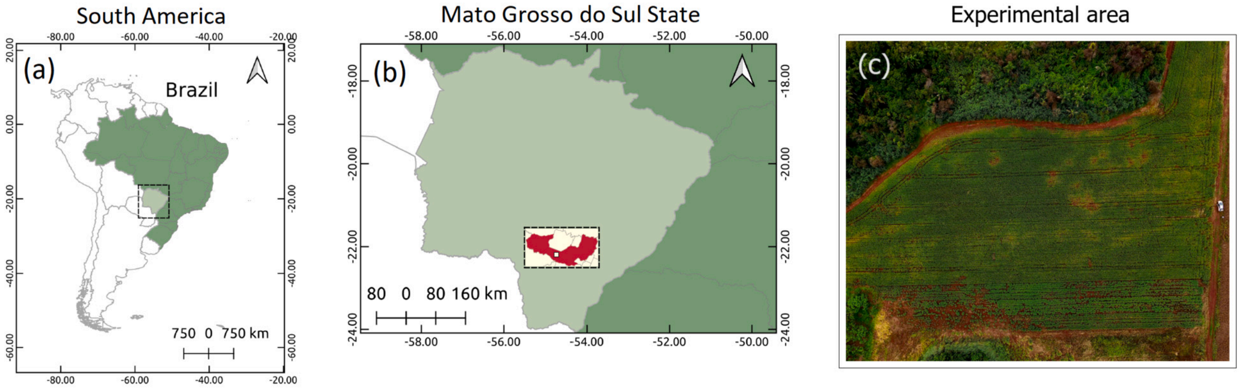

Fig. 1. Study area located in (a) South America and Brazil, (b) Mato Grosso do Sul State, municipality of Dourados highlighted and (c) Experimental area used for soybean planting.

method achieved a recognition rate of 85.5%, showing an advantage over HMAX, Sparse coding and NIMBLE, but being slightly inferior to MatConvNet (86.9%).

In [17–19] the authors analyzed insect pests under real field conditions. In the first two works, the authors used an improved deep learning strategy to detect whiteflies on soybean leaves and eggplant leaves, respectively. In the most recent work, the authors combined hyperspectral imaging technology and a meta-learning algorithm to establish an A-ResNet model, which was used for non-destructive detection of soybeans consumed by *leguminivora glycinivorella matsumura*. Despite promising results, only one species of insect was investigated in each work.

A review characterizing the current state of the art of deep learning applied to soybean crops, detailing the main advancements achieved so far and, more importantly, providing an in-depth analysis of the main challenges and research gaps that still remain were investigated in [20].

Table 1
Total number of insects noted by species and damage caused to soybeans.

Species	Damage	Quantity
Anticarsia gemmatalis	plant leaves	115
Coccinellidae	no damage	120
Diabrotica speciosa	plant leaves	113
Edessa meditabunda	pods and grains	112
Euschistus heros adult	pods and grains	836
Euschistus heros nymph	pods and grains	802
Gastropoda	seedlings	170
Lagria villosa	no damage	67
Nezara viridula adult	pods and grains	125
Nezara viridula nymph	pods and grains	23
Rhammatocerus schistocercoides	plant leaves	37
Spodoptera albula	pods and grains	238
Total		2.758

3. Materials and methods

3.1. Image acquisition

We sowed soybeans in a 2-hectare experimental area specifically to carry out this experiment. Some transgenic soybean cultivars are resistant to certain insect species, so we sow non-genetically modified (non-GMO) soybeans in order to obtain the largest number of species possible. We do not apply any pesticides to the crops to preserve the insects' manifestation. The experimental area shown in Fig. 1 is located in the UFGD farm in the municipality of Dourados, Mato Grosso do Sul State, Brazil, 22°13'57.52"S, 54°59'17.93"W.

We used a Samsung Galaxy S7 smartphone equipped with a 12.2 megapixel SM-G930F rear camera to collect images of insects present in the experimental area. A total of 1,800 images (3024x4032 pixels) were collected in different days and under various weather conditions, at the times 8 am-10 am and 5 pm-6:30 pm. Soybean plants at the reproductive phenological stages R1 to R6 were imaged during the 2017-2018 season, there was no insect sampling during the vegetative stages V1 to V6. During this period, the plants are in the growth phase and the soybean planting lines have not yet been closed. Because of this, most species only appear in the reproductive phase when the plants are larger, offering ideal climatic conditions for their survival (mild temperatures and relatively high humidity). It was found that the exposure of pests on top of plants usually occurs at the beginning of the day or at the end of the afternoon, reinforcing the recommendation that insect sampling should be carried out in the cooler and more humid periods of the day, as reported in [5].

The insects were captured by a digital camera 50 cm away from the target insect with the researcher walking in the soybean field on different days and weather conditions, which causes variations in lighting, occlusion, object overlap and, mainly, similarity of objects with the complex background. These variations make it difficult to recognize the target insect, adding practical value to the soybean pest detection method. Then, we annotated each image using LabelImg https://github.com/tzutalin/labelImg with the support of an entomologist, thus building a reference collection for training and testing the system (see Fig. 2), called INSECT12C-Dataset and is available in [21]. Table 1 presents the total number of insects annotated by species. The imbalanced number of samples reflects the number of occurrences of each pest species under real field conditions.

3.2. Experimental design

Using Scikit-learn [22], a five-fold cross-validation was adopted. In each fold, 60% of the samples were used for training, 20% for validation and 20% for testing. We employed four metrics to evaluate classification results: accuracy, recall, F-measure, and accuracy; and three metrics to evaluate the detection results: mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2). We used ANOVA hypothesis testing to determine whether there is a significant difference in average performance between batch size groups. Batch size refers to the number of data samples that are used in a single iteration through a machine learning model during training. We reported the *p*-value found for each metric and the significance level was set at 5%.

We used the open source implementation of the YOLOv3 architecture that competed in [23]. The following input parameters were used: 608x800-pixel input size, batch sizes ranging from 4 to 32 samples and training with 24,000 iterations. At the same time, we used the SGD op-

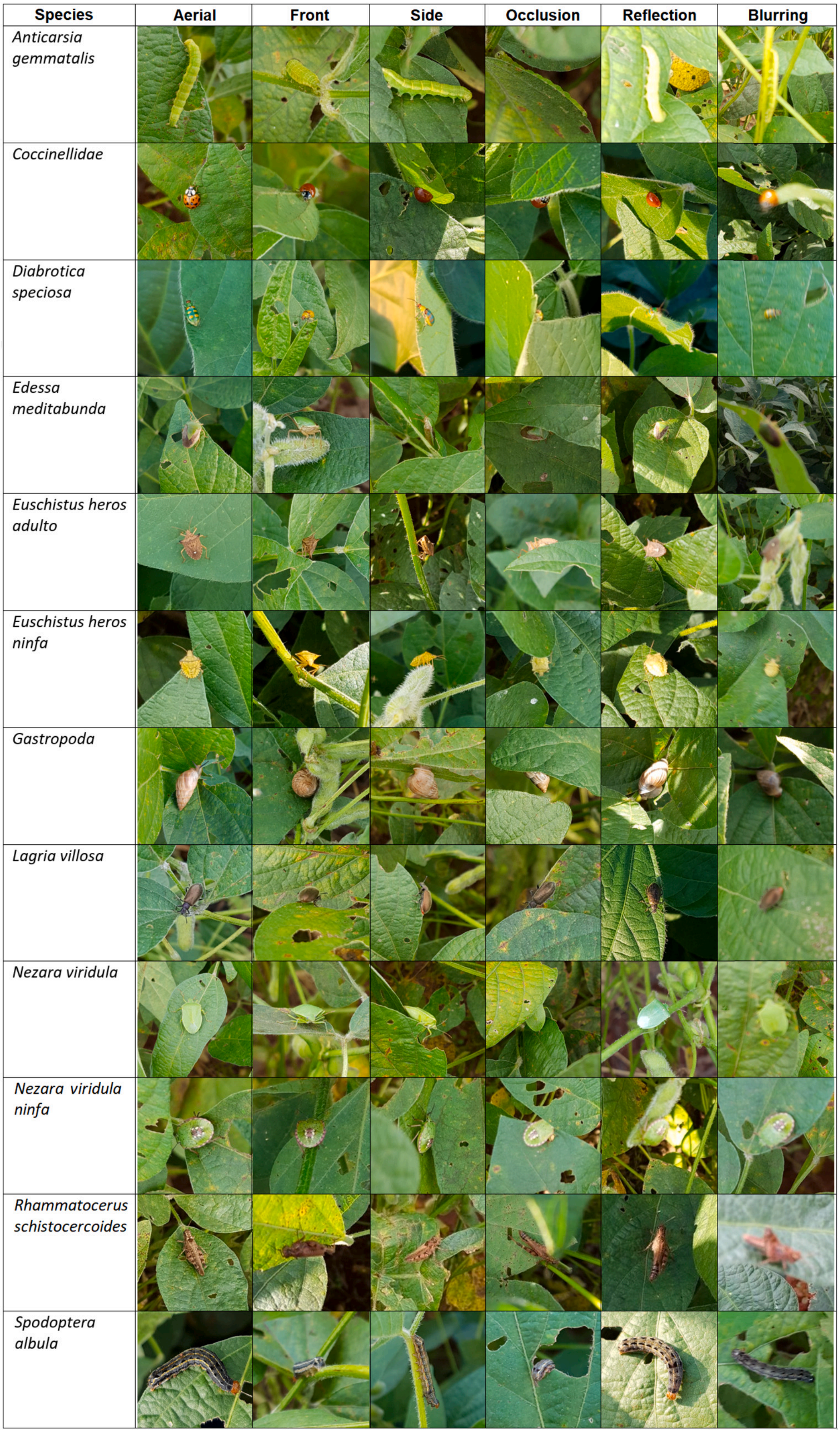

Fig. 2. Sample images from our dataset, divided into 12 species of soybean pests. The images were collected under real field conditions, which provide various lighting conditions, object size and positioning, occlusion, complex background variations and developmental stages.

Table 2

Classification results obtained by YOLO in average percentage in the INSECT12C dataset.

Architecture	Batch	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
	4	**97.85** ± .0049	51.24 ± .0334	67.21 ± .0288	50.67 ± .0326
YOLO	16	95.47 ± .0171	70.65 ± .0109	81.20 ± .0126	68.36 ± .0178
	32	95.15 ± .0060	**75.79** ± .0234	**84.35** ± .0133	**72.96** ± .0200

Table 3

Insect pest detection results in the INSECT12C dataset for YOLO.

Architecture	Batch	RMSE	MAE	R^2
	4	1.25 ± .0788	0.77 ± .0654	0.14 ± .0254
YOLO	16	0.92 ± .0829	0.48 ± .0460	0.47 ± .0440
	32	**0.83** ± .0606	**0.41** ± .0386	**0.58** ± .0446

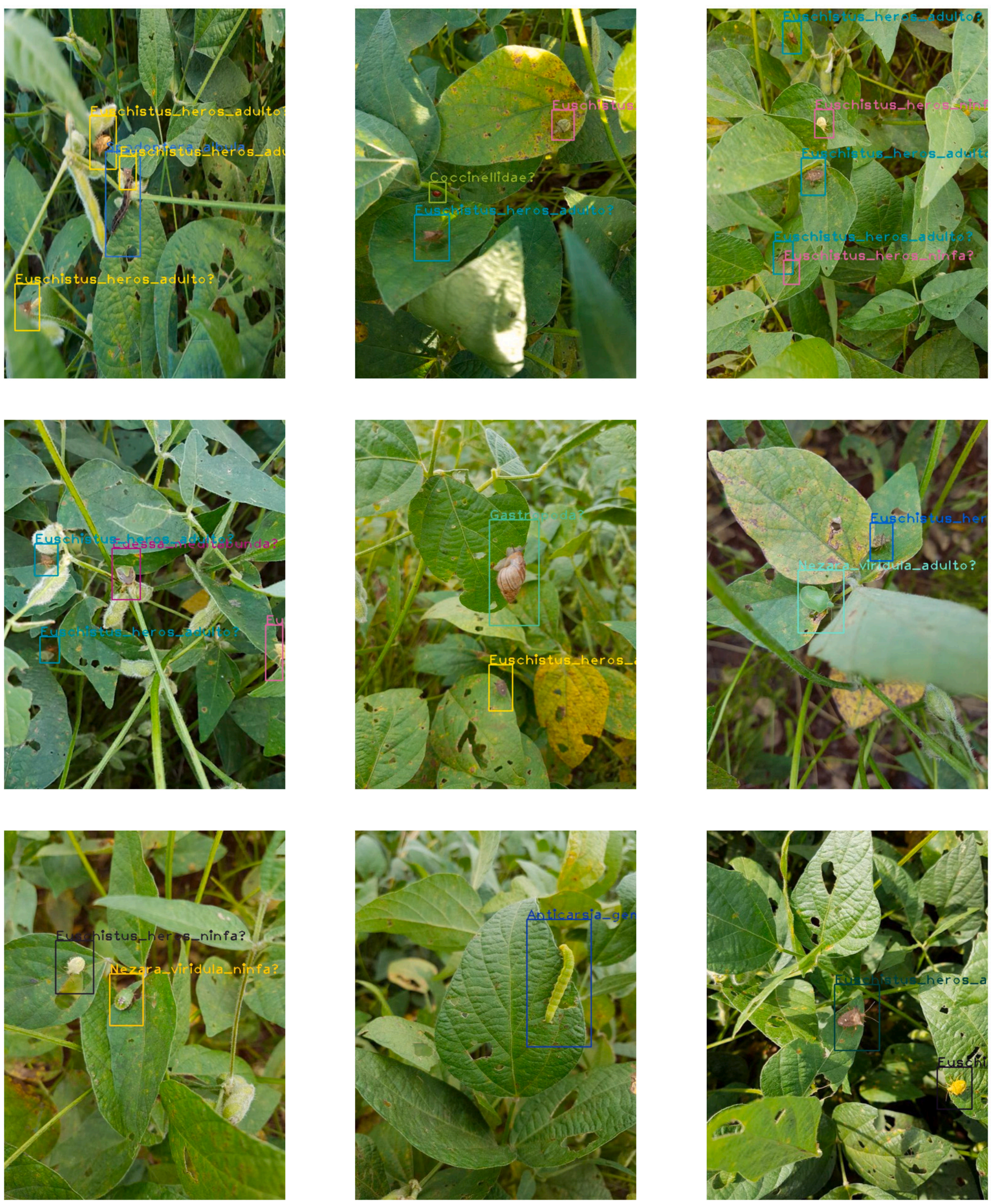

Fig. 3. Examples of insect pest detection in YOLO architecture with batch size defined in 32 samples. Boxes are true positives with a detection score ≥ 0.3.

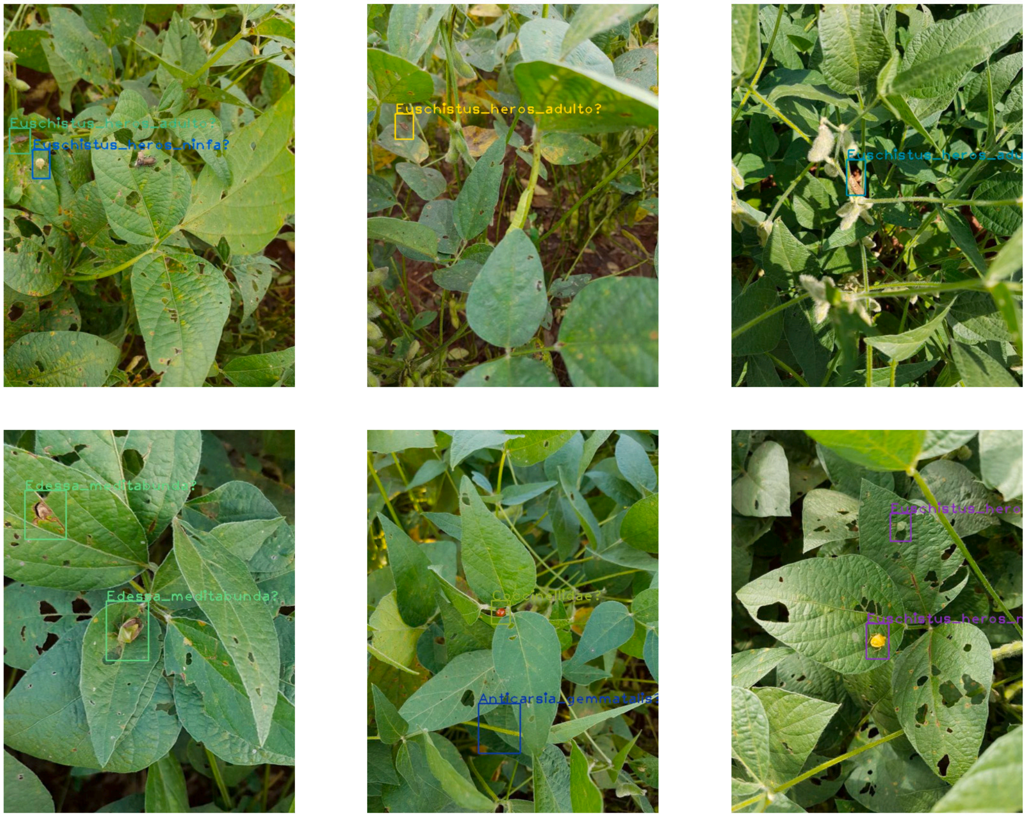

Fig. 4. Examples of insect pest detection failures in the YOLO architecture with a batch size set to 32 samples. The top row shows examples of false negatives, and the bottom row depicts some instances of false positives.

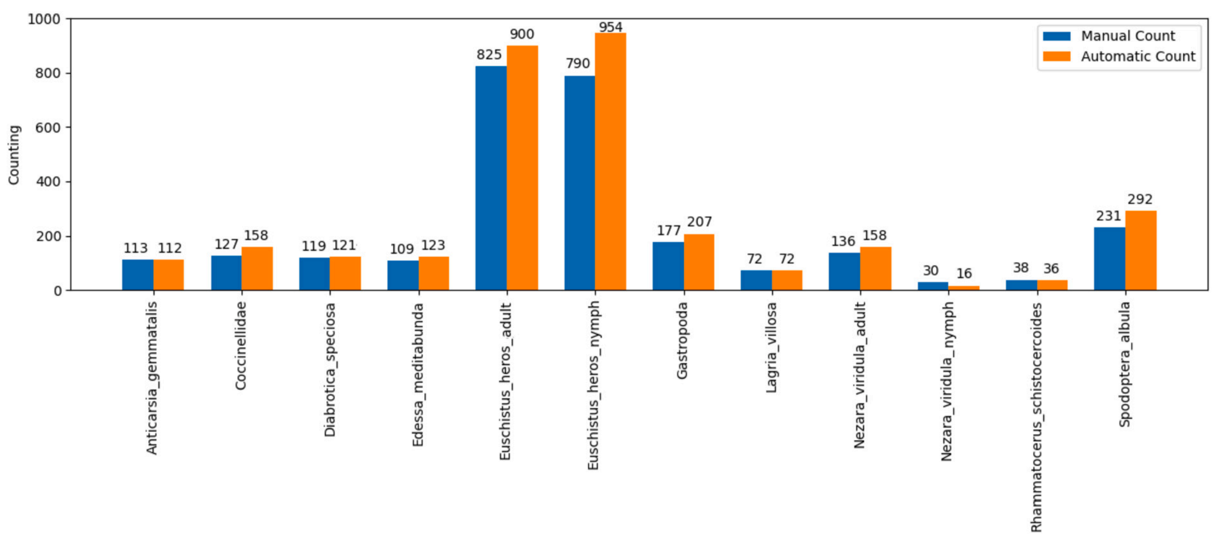

Fig. 5. A bar-graph showing the ground-truth, in blue, and the counts produced by YOLO, in orange.

timizer [24] with a learning rate set of 0.001 and momentum of 0.9. We used the data augmentation to supplement the training data by applying random rotation between -90° and 90° and changing the image brightness from -10% to 10%. This procedure aims to reinforce rotation and illumination invariance during detection.

In all of our experiments we used Colab, a Google Research service that enables writing and running Python codes through a browser, while providing free GPU resources. However, for our work, we used the Google Colab Pro, which provides priority access to more powerful GPU resources and high-memory virtual machines.

3.3. Evaluation metrics

We used four metrics to evaluate the classification results (equations (1) to (4)). True positives (TP) happen when the insect is detected and classified correctly with a score of at least 0.3. False positives (FP) occur

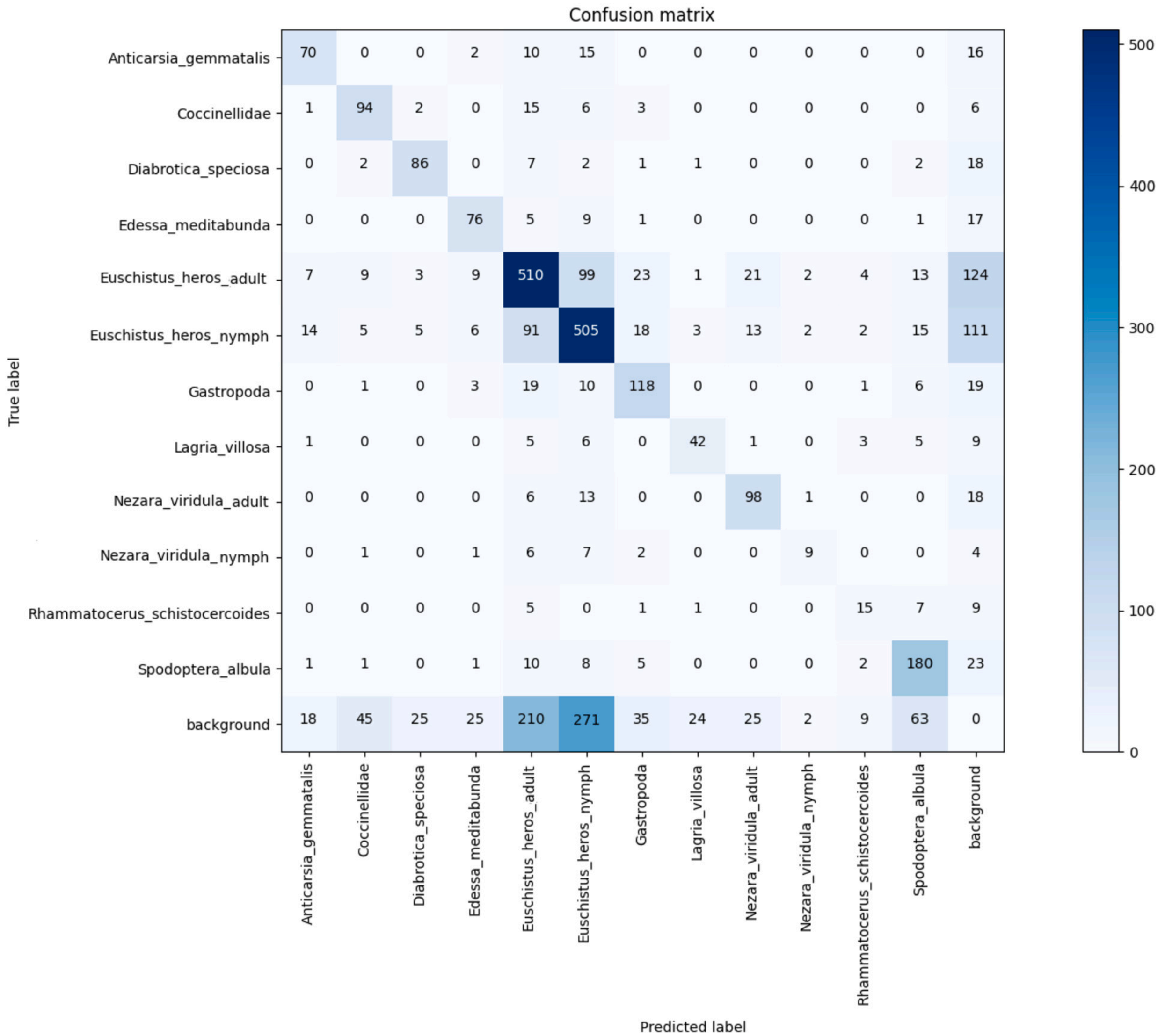

Fig. 6. Confusion matrix showing the predicted versus true classifications among species and between a species and the background (last columns and last line).

when a spurious object is identified as an insect. False negatives (FN) can be due to an insect being either undetected or misclassifield.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - measure = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \tag{4}$$

Furthermore, we consider three metrics to evaluate the detection results:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{7}$$

where y_i is the observed value, $\hat{y}_i$ is the estimated (forecast) value, $\bar{y}$ is the average of observations and n is the number of observations.

4. You only look once (YOLO)

The YOLO detection method evaluated in this work is briefly described below. The source code found in https://github.com/qqwweee/keras-yolo3 was used as basis for our implementation.

YOLO addresses object detection as a problem of direct pixel regression for bounding box coordinates and class probabilities. The input image is divided into $S \times S$ blocks. For each block, YOLO predicts bounding boxes using dimension groups as anchor boxes. For each bounding box, an objectivity score is estimated using logistic regression, which indicates the chance that the bounding box has an object of interest. In addition, class C probabilities are estimated for each bounding box, indicating the classes it may contain. In our case, each bounding box can contain the species of an insect or the background of the crop (uninteresting object). Thus, each estimation in YOLO is composed of four bounding box parameters (coordinates), an objectivity score, and class C probabilities. To provide high accuracy, YOLO uses Darknet-53 as its backbone, requiring fewer operations compared to other architectures.

Fig. 7. Cropped close-up images of adults (left) and nymphs (right) from the species *Euschistus heros* (top) and *Nezara viridula* (down).

5. Results and discussion

5.1. Classification assessment

The classification results obtained by YOLO for precision, recall, F-measure and accuracy are presented in Table 2. The percentage values represent the average of the five folds in the test set. The best results were obtained with a batch size of 32 according to all metrics except precision. More importantly, this batch size resulted in the fewest number of false negatives, witch is arguably the most damaging type of error in this kind of application.

The ANOVA test results indicate that there is a significant difference in mean performance between batch size groups at a 5% significance level, using precision (p-value$=2.46e^{-04}$), recall (p-value$=7.99e^{-09}$), F-measure (p-value$=2.3e^{-08}$) and accuracy (p-value$=1.37e^{-08}$) as metrics.

5.2. Detection of insect pests in soybean

Table 3 shows the detection results. In the experiments, batch size 32 produced the best results, with MAE and RMSE rates of 0.41 and 0.83, respectively. The results show low average error rates in almost all scenarios, indicating that the YOLO architecture can handle multispecies pest detection, even in imbalanced datasets with inter- and intraclass variance. Some examples of insect pest detection in YOLO are shown in Fig. 3. Bounding boxes are true positives with a detection score ≥ 0.3.

However, YOLO led to some false negatives (insects not detected or detected incorrectly) and false positives (when an object that is not of interest is identified as an insect), showing that detections can fail under certain field conditions such as non-uniform lighting over insect, insect-like complex background (e.g. herbivory and leaf lesions), partial insect occlusion and, mainly, low representation of species with few samples (Fig. 4).

5.3. Discussion

Dewi et al. [25] report a 99.40% accuracy using ResNet to classify 11 pest species. It is much higher than our best accuracy of 72.96%, however, they have used a dataset with very high-definition images of insect close-ups and with background information that may be correlated with species and make the problem much easier. The background in their dataset was not composed only of soybean crop images. In our case, the background is uniform, as all the images were taken from the same crop. Besides, the classification results that we present are derived from an object detection task, where more than one insect can appear in the same image, turning the problem even harder than a simple classification.

In a recent work that dealt with insect detection in soybean crop, as ours, Chamara et al. [26] concluded that insect detection is a tough task reporting a mean average precision (mAP) of just 2%, even not trying to detect particular species. As they used a camera that was more distant from the crop than ours, the insects were smaller, making the problem harder, which reflected in their worse results. Park et al. [27] collected images from soybean crops using an unmanned ground vehicle (UGV) with GoPro CAM at a distance similar to ours. Applying three object detectors based on YOLOv3, MRCNN and Detectron2 they have achieved mAPs above 90%. However, while our work aimed at detecting and classifying 10 species, two of them in 2 different stages (nymph and adult), giving a total of 12 classes, Park et al. [27] had a simpler, one-class problem, to detect the *R. pedestris* pest.

Above 90% accuracy results have also been reported by Farah et al. [28], but again, their datasets represent a simpler problem of 2-class classification and not object detection. They aimed to classify images taken from a greater distance than ours, using information such as the holes in soybean leaves left by caterpillars, in the classes healthy or infested. They have experimented with caterpillars and *Diabrotica speciosa*. None of the recent works addresses the problem using such a large number of species and so, their better results, cannot be directly compared to ours.

The bar-graph in Fig. 5 shows the ground-truth and automatic counting per classes. For most of the cases the proposed approach overcounted the number of insects, but in the case of *Anticarsia gemmatalis*, *Lagria villosa*, *Nezara viridula* nymph and *Rhammatocerus schistocercoides*, we have no error or a small undercounting. Proportionally, the worst case happened with *Rhammatocerus schistocercoides*, with an overcounting of approximately 46% and the best case, with an exact counting, was with *Lagria villosa*. Another three classes, *Coccinellidae*, *Euschistus heros* nymph and *Spodoptera albula*, with a proportional counting error above 20%, seem to present greater challenges for automatic counting.

In order to further investigate the disproportion of counting errors among species, Fig. 6 shows a confusion matrix that includes a new class representing the background. The numbers in the last line indicate that an insect has been completely missing and those in the last column indicate that an insect has been detected where there were none. The diagonal indicates the correct classifications and the numbers out of the diagonal, excepting the last line and column, indicate a miss-classification: insect detected but with a wrong species or stage (adult or nymph) attribution. It is clear from the matrix that distinguishing between an adult and a nymph *Euschistus heros* is a hard task, with many confusions in this block. Interestingly, the same does not happen with *Nezara viridula*, with just one case of confusion between its adult and nymph stages.

Fig. 7 shows close-up cropped images from an adult (left) and a nymph (right) of a *Euschistus heros* sample, in the first row, and an adult and a nymph image of a *Nezara viridula*, in the second row. In the case of *Euschistus heros*, the difference in color, from adult to nymph, seems to be an important feature. However, for the *Nezara viridula*, several shape features differ as well. We argue that the network, due to the prevailing green background, ended up favoring shape features over color to distinguish among different species which may lead to more errors with *Euschistus heros* than *Nezara viridula* when trying to choose between adult or nymph.

Fig. 8 shows a normalized version of the confusion matrix where each line adds up to 100%. In this way, the unbalanced nature of the problem is diminished and we can see, for instance, that *Nezara viridula* nymph has just 30% of correct classifications and the *Spodoptera albula*

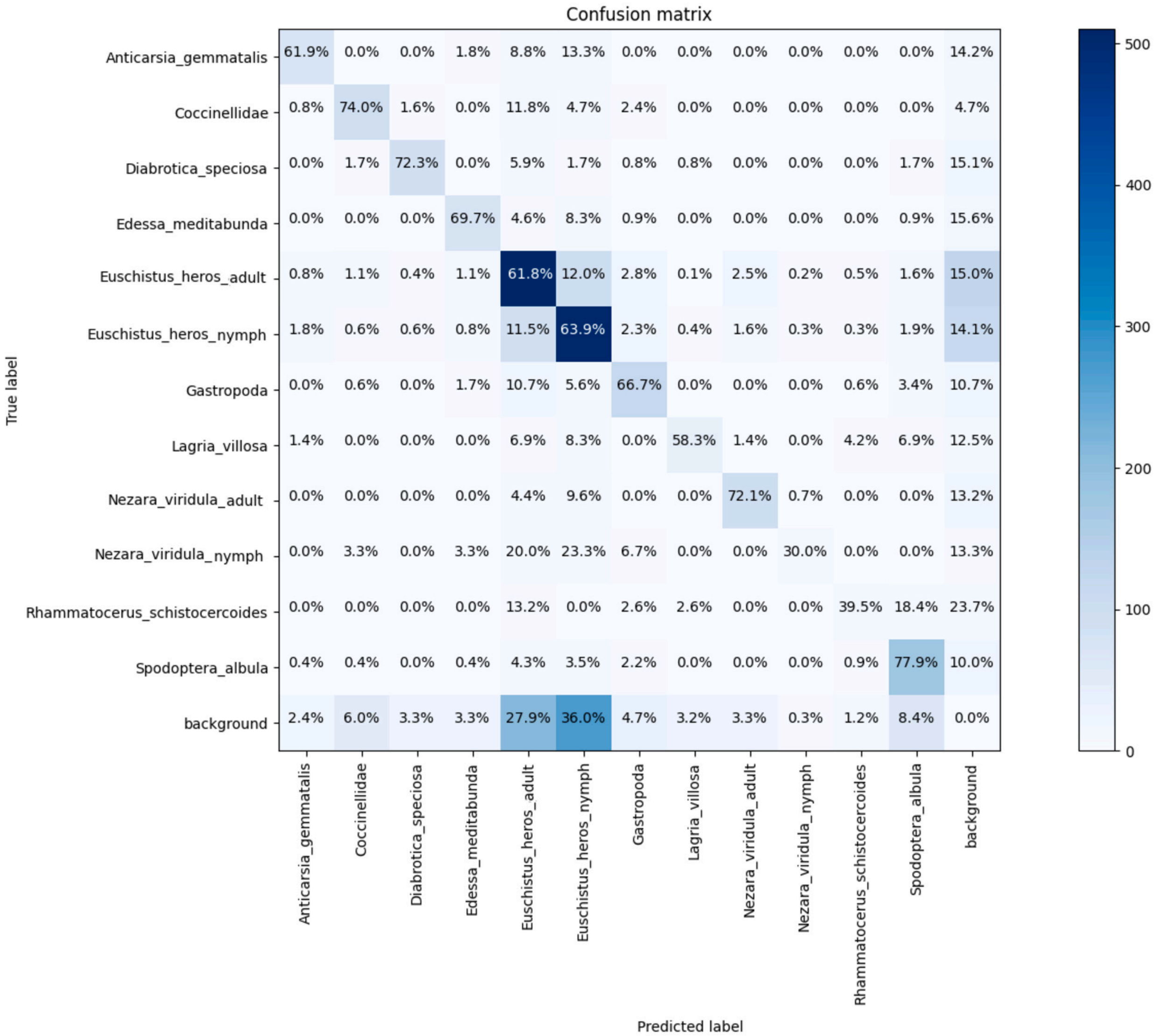

Fig. 8. Normalized confusion matrix showing the predicted versus true classifications using a proportion between correct classifications and the number of samples per class (each row adds up to 100%).

has the highest classification score of 77.9%. But most of the confusion errors are still concentrated in the *Euschistus heros* classes, indicating that unbalancing may be a crucial factor, as these are the most frequent classes in our dataset. Better training balancing techniques should be tested in the future.

6. Conclusion

In this paper we evaluated the performance of the YOLO architecture for real-time detection of insect pests in soybeans. We also created a new dataset called INSECT12C-Dataset, composed of 2,758 annotated insects from 12 species and made it available for academic research. INSECT12C-Dataset can serve as a baseline for real-time detection of insect pests by species in soybeans. Experimental results showed that the YOLO architecture trained with batch size 32 leads to higher classification and detection accuracy compared to batch sizes 4 and 16. However, the method failed under conditions such as areas with complex lighting conditions, herbivory and leaf lesions, partial insect occlusion and low representation of species with few samples. The proportion of true positives with respect to the total predicted positives achieved did not present major distortions, indicating that YOLO allows tracking of pest targets in real time, offering greater practical value for pest control tasks.

As part of future work, we intend to evaluate new state-of-the-art object detection architectures paired with higher resolution cameras embedded in the UAV. We also intend to employ the oversampling technique to supplement the training data of underrepresented classes, adjusting the proportion of samples between the different classes. Finally, we plan to implement the models into an end-to-end system, in the form of pesticide application maps.

CRediT authorship contribution statement

Everton Castelão Tetila: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Fábio Amaral Godoy da Silveira:** Data curation, Software, Validation, Writing – original draft. **Anderson Bessa da Costa:** Writing – review & editing, Data curation, Formal analysis. **Willian Paraguassu Amorim:** Writing – review & editing, Data curation, Formal analysis. **Gilberto Astolfi:** Data curation, Software, Supervision, Validation, Writing – review & editing. **Hemerson Pistori:** Conceptualization, Formal analysis, Project administration, Supervision, Writing – review & editing. **Jayme Garcia Arnal Barbedo:** Investigation, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Process: 423256/2021-1, Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT) – Process: 83/007.898/2023 TO: 11/2023 SIAFEM: 32834, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) – Process: 2023/03870-8 and Process: 2022/09319-9.

References

[1] J. Hou, C. Wang, X. Hong, J. Zhao, C. Xue, N. Guo, J. Gai, H. Xing, Association analysis of vegetable soybean quality traits with ssr markers, Plant Breed. 130 (4) (2011) 444–449, https://doi.org/10.1111/j.1439-0523.2011.01852.x.

[2] CONAB, Acompanhamento da safra brasileira grãos V. 10 - SAFRA 2022/23 - N. 12 - Décimo segundo levantamento, https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos, Setembro 2023, ISSN 2318-6852.

[3] CONAB, Compêndio de estudos conab: Evolução dos custos de produção de soja no brasil, vol. 2, ISSN 2448-3710, https://www.conab.gov.br/institucional/publicacoes/compendio-de-estudos-da-conab/item/download/2512_c2638f76696e3b926ab22e93f9549d21, 2016.

[4] Everton Castelão Tetila, Bruno Brandoli Machado, Gilberto Astolfi, Nícolas Alessandro de Souza Belete, Willian Paraguassu Amorim, Antonia Railda Roel, Hemerson Pistori, Detection and classification of soybean pests using deep learning with uav images, Comput. Electron. Agric. (ISSN 0168-1699) 179 (2020) 105836, https://doi.org/10.1016/j.compag.2020.105836, https://www.sciencedirect.com/science/article/pii/S016816991831055X.

[5] C.B. Hoffmann-Campo, B.S. Corrêa-Ferreira, F. Moscardi, Soja: manejo integrado de insetos e outros Artrópodes-praga, Embrapa Soja, ISBN 978-85-7035-139-5, 2012, http://www.cnpso.embrapa.br/artropodes/Capitulo9.pdf.

[6] Joseph Redmon, Ali Farhadi, Yolov3: an incremental improvement, http://arxiv.org/abs/1804.02767, 2018.

[7] Yanfen Li, Hanxiang Wang, L. Minh Dang, Abolghasem Sadeghi-Niaraki, Hyeonjoon Moon, Crop pest recognition in natural scenes using convolutional neural networks, Comput. Electron. Agric. (ISSN 0168-1699) 169 (2020) 105174, https://doi.org/10.1016/j.compag.2019.105174.

[8] Xi Cheng, Youhua Zhang, Yiqiong Chen, Yunzhi Wu, Yi Yue, Pest identification via deep residual learning in complex background, Comput. Electron. Agric. (ISSN 0168-1699) 141 (2017) 351–356, https://doi.org/10.1016/j.compag.2017.08.005.

[9] Boyi Liu, Zhuhua Hu, Yaochi Zhao, Yong Bai, Yu Wang, Recognition of pyralidae insects using intelligent monitoring autonomous robot vehicle in natural farm scene, CoRR, arXiv:1903.10827 [abs], 2019, http://arxiv.org/abs/1903.10827.

[10] X. Wu, C. Zhan, Y. Lai, M. Cheng, J. Yang, Ip102: a large-scale benchmark dataset for insect pest recognition, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 8779–8788.

[11] Everton Castelão Tetila, Bruno Brandoli Machado, Geazy Vilharva Menezes, Nícolas Alessandro de Souza Belete, Gilberto Astolfi, Hemerson Pistori, A deep-learning approach for automatic counting of soybean insect pests, IEEE Geosci. Remote Sens. Lett. 17 (10) (2020) 1837–1841, https://doi.org/10.1109/LGRS.2019.2954735.

[12] Fangyuan Wang, Rujing Wang, Chengjun Xie, Po Yang, Liu Liu, Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition, Comput. Electron. Agric. (ISSN 0168-1699) 169 (2020) 105222, https://doi.org/10.1016/j.compag.2020.105222.

[13] Nguyen Tuan Nam, Phan Duy Hung, Pest detection on traps using deep convolutional neural networks, in: Proceedings of the 2018 International Conference on Control and Computer Vision, ICCCV '18, New York, NY, USA, Association for Computing Machinery, ISBN 9781450364706, 2018, pp. 33–38.

[14] Denan Xi, Peng Chen, Bing Wang, Jun Zhang, Chengjun Xie, Insect detection and classification based on an improved convolutional neural network, Sensors (ISSN 1424-8220) 18 (12) (2018), https://doi.org/10.3390/s18124169.

[15] Aitor Gutierrez, Ander Ansuategi, Loreto Susperregi, Carlos Tubío, Ivan Rankić, Libor Lenža, A benchmarking of learning strategies for pest detection and identification on tomato plants for autonomous scouting robots using internal databases, J. Sens. (2019), https://doi.org/10.1155/2019/5219471.

[16] Limiao Deng, Yanjiang Wang, Zhongzhi Han, Renshi Yu, Research on insect pest image detection and recognition based on bio-inspired methods, Biosyst. Eng. (ISSN 1537-5110) 169 (2018) 139–148, https://doi.org/10.1016/j.biosystemseng.2018.02.008.

[17] Rubens de Castro Pereira, Edson Hirose, Osmar Luiz Ferreira de Carvalho, Ronaldo Martins da Costa, Díbio Leandro Borges, Detection and classification of whiteflies and development stages on soybean leaves images using an improved deep learning strategy, Comput. Electron. Agric. (ISSN 0168-1699) 199 (2022) 107132, https://doi.org/10.1016/j.compag.2022.107132, https://www.sciencedirect.com/science/article/pii/S0168169922004495.

[18] Arantza Bereciartua-Pérez, Laura Gómez, Artzai Picón, Ramón Navarra-Mestre, Christian Klukas, Till Eggers, Insect counting through deep learning-based density maps estimation, Comput. Electron. Agric. (ISSN 0168-1699) 197 (2022) 106933, https://doi.org/10.1016/j.compag.2022.106933, https://www.sciencedirect.com/science/article/pii/S0168169922002502.

[19] Jiangsheng Gui, Huirong Xu, Jingyi Fei, Non-destructive detection of soybean pest based on hyperspectral image and attention-resnet meta-learning model, Sensors (ISSN 1424-8220) 23 (2) (2023), https://www.mdpi.com/1424-8220/23/2/678.

[20] Jayme Garcia, Arnal Barbedo, Deep learning for soybean monitoring and management, Seeds (ISSN 2674-1024) 2 (3) (2023) 340–356, https://www.mdpi.com/2674-1024/2/3/26.

[21] Everton Castelão Tetila, INSECT12C-Dataset - Soybean Pest Image Dataset, UFGD, 2023, https://github.com/EvertonTetila/INSECT12C-Dataset.

[22] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay, Scikit-Learn: Machine Learning in Python, 2018.

[23] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.

[24] Léon Bottou, Olivier Bousquet, The tradeoffs of large scale learning, in: J. Platt, D. Koller, Y. Singer, S. Roweis (Eds.), Advances in Neural Information Processing Systems, vol. 20, Curran Associates, Inc., 2008, https://proceedings.neurips.cc/paper/2007/file/0d3180d672e08b4c5312dcdafdf6ef36-Paper.pdf.

[25] C. Dewi, H.J. Christanto, G.W. Dai, Automated identification of insect pests: a deep transfer learning approach using resnet, Acadlore Trans. Mach. Learn. 2 (4) (2023) 194–203.

[26] Nipuna Chamara, Geng Bai, Yufeng Ge, Aicropcam: deploying classification, segmentation, detection, and counting deep-learning models for crop monitoring on the edge, Comput. Electron. Agric. 215 (2023) 108420.

[27] Yu-Hyeon Park, Sung Hoon Choi, Yeon-Ju Kwon, Soon-Wook Kwon, Yang Jae Kang, Tae-Hwan Jun, Detection of soybean insect pest and a forecasting platform using deep learning with unmanned ground vehicles, Agronomy 13 (2) (2023) 477.

[28] Niklas Farah, Nicolas Drack, Hannah Dawel, Ricardo Buettner, A deep learning-based approach for the detection of infested soybean leaves, IEEE Access (2023).