# SEEING THE UNSEEN: AI-POWERED CAMOUFLAGED PEST DETECTION

Henry O. Velesaca[1,3] and Angel D. Sappa[1,2]
[1]*Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, Guayaquil, 090902, Ecuador*
[2]*Computer Vision Center, Campus UAB, 08193 Bellaterra, Barcelona, Spain*
[3]*Software Engineering Department, Research Center for Information and Communication Technologies (CITIC-UGR), University of Granada, 18071, Granada, Spain*

**ABSTRACT**

This paper presents an innovative approach for detecting camouflaged agricultural pests using YOLOv11, a state-of-the-art deep learning architecture, and SINet-v2. The research focuses specifically on cotton bollworm detection, a significant threat to agricultural productivity. Our methodology leverages advanced neural networks to identify subtle patterns distinguishing concealed pests from their surrounding environment. Using a comprehensive dataset of 1,073 cotton bollworm images, the YOLOv11 and SINet-v2 architectures are evaluated against an existing approach. The results obtained by YOLOv11 demonstrate superior performance in the Average Precision metric for detection of 71.3% and segmentation of 60.3%. Compared to recently presented methods, like Mask R-CNN+PVT, our approach shows improved detection capabilities while requiring fewer computational resources. This work contributes to advancing automated pest management systems in agriculture, offering a more efficient and accurate solution for early pest detection.

**KEYWORDS**

Camouflaged Object Detection, Pest Detection, YOLO

## 1. INTRODUCTION

Agricultural pests pose a significant threat to global food security, causing substantial crop losses estimated at billions of dollars annually (https://www.fao.org/plant-production-protection/about/en) (Oerke, 2016). The challenge of detecting these pests is further complicated by their evolutionary adaptations, particularly camouflage mechanisms that make them nearly invisible to conventional detection methods (Skendžić et al., 2021).

In recent years, the emergence of deep learning technologies has revolutionized computer vision applications, offering unprecedented capabilities in automated pest detection and monitoring systems (Karar et al., 2021; Ullah at al., 2022). These advances have transformed traditional agricultural practices by introducing sophisticated neural network architectures like Convolutional Neural Networks (CNNs) (Venkatasaichandrakanthand et al., 2023), Region-based CNNs (R-CNN) (Rong et al., 2022), and You Only Look Once (YOLO) variants (Liu at al., 2020), which have demonstrated remarkable accuracy in identifying and classifying various pest species. The integration of these technologies has enabled real-time monitoring, early detection of pest infestations, and more precise application of pest control measures, leading to significant improvements in crop protection strategies and reduced pesticide usage.

The importance of this work lies in its potential to transform pest management practices, enabling early detection and targeted intervention strategies (Chen at al., 2020). This advancement could substantially reduce pesticide usage while improving crop protection efficiency, thereby contributing to more sustainable agricultural practices. Our approach not only addresses the immediate challenge of pest detection but also provides a scalable framework for future developments in agricultural monitoring systems.

ISBN: 978-989-8704-69-6 © 2025

This research presents an innovative approach for detecting camouflaged agricultural pests using state-of-the-art deep learning techniques. Our methodology leverages advanced convolutional neural networks (CNNs) and feature extraction algorithms to identify subtle patterns and characteristics that distinguish concealed pests from their surrounding environment. By combining high-resolution imaging with sophisticated neural network architectures (Peng et al, 2022).

To address this work in detail, the manuscript is organized as follows. Section 2 introduces the background on the use of pest detection techniques based on deep learning and addresses the problem as a camouflaged object detection approach. Section 3 presents the proposed approach to carry out the identification of pests using deep learning techniques. Then, section 4 shows the experimental results taking as a reference a case study for the camouflage detection of cotton bollworm. Finally, conclusions are presented in Section 5.

## 2. BACKGROUND

As described above, this document presents a deep learning-based approach to camouflaged agricultural pest detection. This section summarizes some of the most relevant techniques related to the present work. First, approaches based on Deep Learning are reviewed, followed by approaches based on Camouflaged Object Detection.

### 2.1 Deep Learning Approaches

Deep learning approaches, particularly CNN techniques, have demonstrated superior performance in pest detection. As is the case, the work proposed by Rivadeneira et al. (2023), which presented a novel approach that is especially valuable for camouflage scenarios, as it integrates super- resolution techniques with object detection methods like YOLOv8 (Jocher et al., 2023) to address the problem of very low-resolution thermal images. The most relevant feature of the camouflage area is that it allows the detection of objects that could remain undetected in visible spectrum images since it uses improved thermal images through a guided super-resolution strategy that is supported by a high-resolution image of the visible spectrum. This makes the system particularly useful for identifying camouflaged objects that will be difficult or impossible to detect using conventional images alone.

However, there are certain limitations and problems, such as the work presented by Fuentes et al. (2017) identified that these systems require massive amounts of labeled training data and significant computational resources. Their study on tomato disease detection showed that while achieving 96% accuracy under controlled conditions, performance dropped significantly when dealing with new or unseen pest varieties.

### 2.2 Camouflaged Object Detection Approaches

Recent research has specifically addressed the challenge of detecting pests as a Camouflaged Object Detection (COD) problem. Remarkable work by Wang et al. (2023) shows a camouflaged insect segmentation system using a Progressive Refinement Network that employs an innovative multi- stage approach to accurately detect and segment insects that blend into their environment. This architecture implements a hierarchical process that gradually refines visual features, starting with coarse features and progressing to fine details, achieving significant accuracy in detecting camouflaged insects. However, the system faces limitations such as dependence on large and well- labeled data sets, difficulties with variable lighting conditions, and extremely complex camouflage patterns, in addition to requiring considerable computational power for real-time processing.

Finally, the work recently presented by Meng at al. (2024) shows a segmentation system for camouflaged bollworm instances, which combines the Pyramid Vision Transformer (PVT) architecture with Mask R-CNN, creating a hybrid approach that improves detection and segmentation of this pest in agricultural environments. This integration leverages PVT's ability to capture global relationships and hierarchical features, while Mask R-CNN provides accurate segmentation at the instance level, achieving superior results in identifying cotton bollworms even when they are camouflaged in foliage. However, the system presents limitations in terms of required computational resources, sensitivity to varying environmental conditions, and the need for an extensive and diverse training dataset to maintain its effectiveness in different field scenarios.

# 3. PROPOSED APPROACH

This section details the different stages followed to carry out the proposed approach. First, the dataset is detailed, then the techniques used, and finally, the metrics needed for the quantitative evaluation.

## 3.1 Dataset Description

The cotton bollworm dataset represents a specialized collection of images focused on one of the most destructive agricultural pests affecting cotton crops worldwide. This dataset has been specifically created by Meng et al. (2024) to address the challenges in detecting these pests in various scenes and environmental conditions. The dataset is available at the following URL: https://www.kaggle.com/datasets/kexinmeng1/the-dataset-of-cotton-bollworms.

## 3.2 Deep Learning Techniques

For the segmentation task, YOLOv11 (Jocher et al., 2024) is selected as the first neural network architecture. This choice is driven by several key advantages: its exceptional processing efficiency and real-time detection capabilities, coupled with its ability to identify multiple objects simultaneously in a single forward pass. YOLOv11's sophisticated architecture offers remarkable flexibility, allowing for fine-tuned optimization to handle varying environmental conditions, including diverse lighting scenarios, and viewing perspectives. Furthermore, its open-source nature and robust community support ensure both straightforward implementation and sustainable long-term maintenance, making it an ideal choice for this application.

On the other hand, based on the state of the art, a comparison is made with the work proposed by Meng et al. (2024), which uses the Pyramid Vision Transformer (PVT) architecture with Mask R-CNN to detect camouflaged cotton bollworm instances. The dataset proposed by the same authors' team is used and compared with the metrics presented in the original paper.

Finally, it is also chosen to use SINet-v2 (Fan et al., 2021) which is a neural network used for COD tasks in two stages: search and identification. It uses EfficientNet as a backbone and incorporates two modules, neighbor connection decoder (NCD) and group-reversal attention (GRA) that allow it to excel in detecting camouflaged objects. This architecture is chosen because it has demonstrated state-of-the-art results on major reference datasets and is available as open source.

## 3.3 Metric Evaluation

For the evaluation of the proposed work, Average Precision (AP) (e.g., for bounding boxes and segmentation tasks), Recall, F1-score, and Pixel Accuracy metrics will be calculated to establish the performance of the proposed approach. Table 1 shows the details of the metrics to be calculated as well as a description of each of them.

Table 1. Metrics used to evaluate proposed work. TP, FP, TN, and FN represent True Positive, False Positive, True Negative, and False Negative, respectively \cite{meng2024camouflaged}

| Metric | Description |
|---|---|
| $AP^{B}$ | Average Precision for predicted bounding boxes, calculated across multiple IoU thresholds from 0.5 to 0.95 in steps of 0.05 |
| $AP_{B50}$ | Average Precision for predicted bounding boxes, evaluated at an IoU threshold of 0.5 |
| $AP_{B75}$ | Average Precision for predicted bounding boxes, evaluated at an IoU threshold of 0.75 |
| $AP^{S}$ | Average Precision for predicted masks, calculated across multiple IoU thresholds from 0.5 to 0.95 in steps of 0.05 |
| $AP_{S50}$ | Average Precision for predicted masks, evaluated at an IoU threshold of 0.5 |
| $AP_{S75}$ | Average Precision for predicted masks, evaluated at an IoU threshold of 0.75 |
| Pixel Accuracy | PA = (TP + TN) / (TP + TN + FP + FN) |
| Recall | R = TP / (TP + FN) |
| F1-Score | F1 = 2 x P x R / (P + R); P = TP/ (TP + FP) |

ISBN: 978-989-8704-69-6 © 2025

# 4. CASE STUDY

This section presents the case study experimental results obtained with the proposed framework. For the performance evaluation of the proposed approach, the metrics described in Table 1 are used.

## 4.1 System Implementation

All experiments are conducted on a high-performance workstation configured with an Intel Core i9 processor (base clock: 3.3 GHz, 16 cores/24 threads) and an NVIDIA Titan XP GPU (12GB GDDR5X, 3840 CUDA cores). This hardware configuration provided the computational resources required for efficient training, validation, and testing of the tested architectures in the context of COD tasks. The system's specifications are chosen to ensure optimal processing of complex neural network operations and handling of large-scale dataset operations.

Figure 1. Example of images from the cotton bollworm dataset presented by Meng et al. (2024)

## 4.2 Dataset

The dataset used for the different experiments carried out is obtained from (Meng et al., 2024). This dataset consists of 1080 images, but inconsistencies are found in 7 of these images, which are removed. Table 2 shows the distribution of the dataset for each of the subsets. Additionally, Figure 1 shows examples of the images that are part of the dataset.

Table 2. Distribution of dataset

| Task | Cotton bollworm | |
|---|---|---|
| | # of Images | # of Instances |
| Training | 856 | 932 |
| Validation | 161 | 172 |
| Testing | 56 | 69 |
| Total | 1073 | 1176 |

## 4.3 YOLOv11 Implementation

To enhance the model's training capabilities, the initial step involved expanding the training dataset through data augmentation techniques. This is accomplished using the Albumentations library (Buslaev et al., 2020), a powerful tool specifically designed for computer vision applications, including the development of object detection models like YOLOv11. The library provides a comprehensive suite of image transformation operations, encompassing techniques such as random cropping, rotational adjustments, brightness modifications, contrast alterations, and various other transformations. These modifications can be systematically applied during the data preprocessing phase, creating additional training samples that maintain realistic characteristics while introducing controlled variations to the original images. This approach effectively enriches the training dataset, promoting better model generalization and robustness. The model "*yolo11m-seg.pt*" with training epochs of 500, batch size of 8, and image size of 640 × 640 pixels are used as execution parameters.

## 4.4 SINet-v2

The SINet-V2 implementation is used for the detection of cotton bollworm camouflage. This architecture has been designed for the COD task and is considered for this work because it is a state-of-the-art architecture that has obtained very good results. The implementation uses PyTorch and features a custom loss function that combines weighted BCE (Binary Cross Entropy) and IoU (Intersection over Union). Training is carried out for 100 epochs with a batch size of 16 and an image size of 352×352 pixels, using the Adam optimizer with an initial learning rate of 0.0001 that is adjusted every 25 epochs.

## 4.5 Experimental Results

As a last stage to evaluate the performance of the proposed approach 56 test images are considered for the calculation of the metrics. Table 3 shows the result of the calculation of the metrics for each of the techniques. The values of the Mask R-CNN+PVT (Meng et al., 2024) technique are obtained from the original paper. In addition, Figure 2 (*3rd column*) uses the images of the results of the article proposed by Meng et al. (2024) for qualitative comparison in this work.

Table 3. Evaluation of metrics for each technique---notation as presented in Table 1

| Technique | Detection | | | Segmentation | | | PA | R | F1 |
|---|---|---|---|---|---|---|---|---|---|
| | $AP^B$ | $AP_{B50}$ | $AP_{B75}$ | $AP^S$ | $AP_{S50}$ | $AP_{S75}$ | | | |
| Mask R-CNN+PVT | 62.5 | 89.7 | 72.9 | 59.5 | 89.2 | 72.9 | **98.61** | **99.12** | **98.95** |
| YOLOv11 | **71.3** | **92.3** | **73.8** | **60.3** | **90.3** | **73.1** | 98.34 | 96.47 | 97.79 |
| SINet-v2 | 54.5 | 71.3 | 59.8 | 44.5 | 56.3 | 49.2 | 90.07 | 65.61 | 47.70 |

Table 3 shows that YOLOv11 architecture achieved notable metrics with an $AP^B$ of 71.3% for detection and an $AP^S$ of 60.3% for segmentation, significantly outperforming the Mask R-CNN+PVT model which achieved 62.5% and 59.5% respectively. This improvement in accuracy is particularly relevant considering the inherent complexity of detecting camouflaged pests and a notable aspect of YOLOv11 is its computational efficiency. The proposed model requires substantially fewer resources, operating at 123.0 GFlops compared to the 154.78 GFlops of the Mask R-CNN+PVT while maintaining a smaller number of parameters (22.34M vs 41.78M) (see Table 4). This optimization in computational resources makes the system more viable for practical implementations in real agricultural environments. These results are particularly impressive considering the complex nature of the test images including various camouflage patterns and varying environmental conditions.

Table 4. Comparison between GFlops and the number of parameters

| Technique | Input size (px) | GFlops | #Param. (M) |
|---|---|---|---|
| Mask R-CNN+PVT | 800 x 640 | 154.78 | 41.78 |
| YOLOv11 | 640 x 640 | 123.00 | 22.34 |
| SINet-v2 | 352 x 352 | 144.48 | 26.98 |

On the other hand, Figure 2 shows the qualitative evaluation of the results obtained by each technique, in which better obtaining of bounding boxes and segmentation masks can be obtained using YOLOv11; these results are consistent with the quantitative metrics obtained for the detection and segmentation tasks. In addition, Figure 3 shows more challenging images when finding the cotton bollworm camouflaged in the different scenarios where they blend in with the environment, and it is more difficult to visualize them with the naked eye. These qualitative results demonstrate the robustness of YOLOv11 in complex scenarios, and this can equal or improve the results of state-of-the-art techniques specialized for COD tasks. Also, SINet-v2 does not obtain good results and is limited to detecting and segmenting salient objects in this particular dataset (see Table 2 *1st row, 2nd column*).

Although the presented dataset is an excellent contribution, it also presents certain inconsistencies in the labeling, for example, Table 2 (*5th row, 2nd column*) shows a labeling error where the GT does not contain the mask of the small worm located on the right side but Mask R-CNN+PVT and YOLO can detect it. Likewise, Table 3 (*1st row, 2nd column*) shows a labeling error where the GT does not show the mask of the small worm located on the left side, but YOLO also detects it (*1st row and 4th column green mask*). This is one reason why the PA, R, and F1 metrics do not present higher values.


Figure 2. Prediction results using Mask R-CNN+PVT, YOLOv11, and SINet-v2

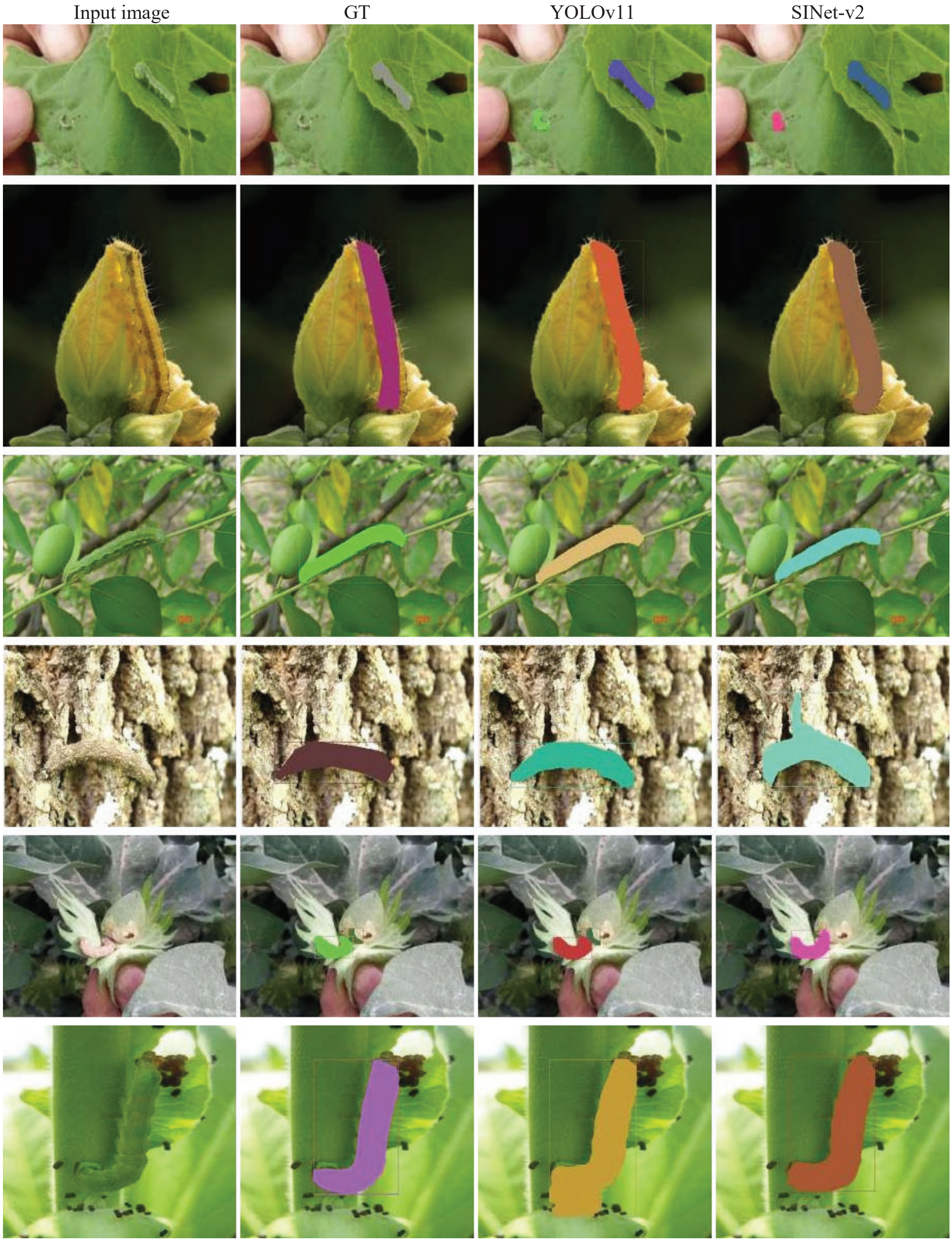


Figure 3. Prediction results using YOLOv11 and SINet-v2 on challenging camouflage images

ISBN: 978-989-8704-69-6 © 2025

## 5. CONCLUSION

This research demonstrates the effectiveness of the YOLOv11-based approach for camouflaged agricultural pest detection, significantly outperforming recent methods such as Mask R-CNN+PVT in both performance and computational efficiency. Another finding is that SINet-v2 exhibits lower performance than YOLOv11, despite SINet-v2 being a specialized architecture for COD tasks. The YOLOv11 evaluation architecture not only achieves better pest detection and segmentation but also requires fewer computational resources, making it more practical for deployments in real agricultural environments. Systems based on these types of technologies not only promise to reduce herbicide use but also offer more precise and environmentally friendly weed control solutions. The robustness of the system is evidenced by its ability to maintain high performance under various challenging scenarios with different camouflage patterns and varying environmental conditions, thus establishing a solid foundation for future developments.

## ACKNOWLEDGEMENT

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-24-1-0206; in part by the ESPOL project CIDIS-003-2024-T; and in part by the University of Granada.

## REFERENCES

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M. & Kalinin, A.A., (2020). Albumentations: fast and flexible image augmentations. Information, 11(2), p.125.

Chen, C.J., Huang, Y.Y., Li, Y.S., Chang, C.Y. & Huang, Y.M., (2020). An AIoT based smart agricultural system for pests detection. IEEE Access, 8, pp.180750-180761.

Fan, D.P., Ji, G.P., Cheng, M.M. & Shao, L., (2021). Concealed object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10), pp.6024-6042.

Fuentes, A., Yoon, S., Kim, S.C. & Park, D.S., (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors, 17(9), p.2022.

Jocher, G., Chaurasia, A. & Qiu, J., (2023). Ultralytics yolov8. Available at: https://github.com/ultralytics/ultralytics

Jocher, G. & Qiu, J., (2024). Ultralytics yolo11. Available at: https://github.com/ultralytics/ultralytics

Karar, M.E., Alsunaydi, F., Albusaymi, S. & Alotaibi, S., (2021). A new mobile application of agricultural pests recognition using deep learning in cloud computing system. Alexandria Engineering Journal, 60(5), pp.4423-4432.

Liu, J. & Wang, X., (2020). Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. Frontiers in Plant Science, 11, p.898.

Meng, K., Xu, K., Cattani, P. & Mei, S., (2024). Camouflaged cotton bollworm instance segmentation based on PVT and Mask R-CNN. Computers and Electronics in Agriculture, 226, p.109450.

Oerke, E.C., (2006). Crop losses to pests. The Journal of Agricultural Science, 144(1), pp.31-43.

Peng, Y. & Wang, Y., (2022). CNN and transformer framework for insect pest classification. Ecological Informatics, 72

Rivadeneira, R.E., Velesaca, H.O. & Sappa, A., (2023), November. Object detection in very low-resolution thermal images through a guided-based super-resolution approach. In: 2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE, pp.311-318.

Rong, M., Wang, Z., Ban, B. & Guo, X., (2022). Pest Identification and Counting of Yellow Plate in Field Based on Improved Mask R-CNN. Discrete Dynamics in Nature and Society, 2022(1), p.1913577.

Skendžić, S., Zovko, M., Živković, I.P., Lešić, V. & Lemić, D., (2021). The impact of climate change on agricultural insect pests. Insects, 12(5), p.440.

Ullah, N., Khan, J.A., Alharbi, L.A., Raza, A., Khan, W. & Ahmad, I., (2022). An efficient approach for crops pests recognition and classification based on novel deeppestnet deep learning model. IEEE Access, 10, pp.73019-73032.

Venkatasaichandrakanthand, P. & Iyapparaja, M., (2023). Pest detection and classification in peanut crops using CNN, MFO, and EViTA algorithms. IEEE Access, 11, pp.54045-54057.

Wang, J., Hong, M., Hu, X., Li, X., Huang, S., Wang, R. & Zhang, F., (2023). Camouflaged insect segmentation using a progressive refinement network. Electronics, 12(4), p.804.