DEEP LEARNING (AI) FOR EARLY AND ACCURATE IDENTIFICATION OF CROP PESTS AND BENEFICIAL INSECTS IN FIELD IMAGERY

K. A. Tumsare
Assistant Professors,
Department of Zoology,
Amolakchand Mahavidyalaya, Yavatmal
Email : kamleshtumsare10@gmail.com

S. B. Barge
Assistant Professors,
Department of Zoology,
Amolakchand Mahavidyalaya, Yavatmal
Email : kamleshtumsare10@gmail.com

Crossref DOI - https://doi.org/10.63665/rh.v7i2.58

Abstract :

Effective management of agricultural ecosystems requires rapid and precise identification of the entire insect community, encompassing both harmful pests and vital beneficial species. Traditional monitoring is labor-intensive and often leads to non-selective pesticide use, which harms biological control agents. This research addresses this challenge by proposing and evaluating a novel hybrid deep learning architecture, YOLOv8 integrated with a Self-Attention Mechanism (YOLOv8-SAM), designed for real-time object detection in complex field imagery. The model was trained using a composite dataset, integrating the benchmark IP102 dataset with specialized local field images of cotton and soybean crops from Maharashtra, India. The YOLOv8-SAM model demonstrated superior performance, achieving a Mean Average Precision across rigorous Intersection over Union (IoU) thresholds (mAP@[0.5:0.95]) of 0.785 and a high inference speed of 125 Frames Per Second (FPS). These results confirm that the attention mechanism significantly enhances the model's capacity for fine-grained differentiation among visually similar pest and beneficial insect classes while maintaining the low latency essential for real-time edge computing. The proposed system offers a trustworthy and resource-efficient tool for sustainable Integrated Pest Management (IPM), supporting critical decision-making in agricultural production.

Keywords : Object Detection, Integrated Pest Management (IPM), YOLOv8, Vision Transformer, Field Imagery, Precision Agriculture, Real-Time Processing.

Introduction :

1. Agricultural Context and the Need for Precision :

Agricultural production worldwide faces severe threats from insect pests, which reduce crop yield and quality. In Maharashtra, India, the cultivation of cash crops like cotton and soybean suffers heavy losses due to widespread pest attacks. To mitigate these losses, agriculture must shift away from reactive, broad-spectrum chemical interventions toward proactive, precise monitoring systems. Conventional monitoring methods, which rely on human observation and taxonomic expertise, are fundamentally limited by high costs, labour demands, and inherent inaccuracies when dealing with large-scale fields, often preventing the timely intervention needed to prevent economic damage.

2. The Role of AI in Sustainable IPM :

Modern sustainable agriculture hinges on the principles of Integrated Pest Management (IPM), which mandates a comprehensive understanding of the entire arthropod community, including beneficial insects such as predators (e.g., Lady Beetles) and parasitoids. These beneficial species act as natural biological controls, and their presence often dictates non-intervention. Indiscriminate or premature insecticide application, often triggered by inaccurate monitoring, detrimentally reduces these beneficial populations, leading to reduced natural regulation and potential secondary pest outbreaks.

Artificial Intelligence (AI), specifically deep learning (DL) and computer vision techniques, offers a transformative solution for automating detection and diagnosis in entomology and precision agriculture. However, achieving reliable object detection in complex natural environments requires models that can handle extreme visual variations—such as camouflage, variable lighting, and high interspecies similarity (look-alikes)—and still maintain the computational efficiency needed for real-time deployment on mobile and edge computing platforms.

3. Research Objectives and Contribution :

This study proposes to overcome the inherent challenges in fine-grained insect identification by developing a hybrid deep learning model optimized for both speed and precision. The primary objectives are:

1. To develop a robust, field-relevant dataset combining large-scale public benchmarks (IP102) with specialized local imagery from major regional crops (cotton and soybean).
2. To implement a high-speed, high-accuracy hybrid deep learning detector (YOLOv8-SAM) that integrates the efficiency of single-stage detectors with the robust contextual analysis of self-attention mechanisms.
3. To validate the model's performance rigorously, focusing on precise localization accuracy (mAP@[0.5:0.95]) and real-time inference speed (FPS).
4. To demonstrate the model's practical utility in IPM by accurately differentiating harmful pest species from critical beneficial predator and parasitoid species.

The core contribution is a robust, deployable solution tailored to the specific technical demands of real-time entomological monitoring in resource-constrained agricultural settings.

Review of Literature :

1. Object Detection Architectures and the Speed-Accuracy Trade-off :

Early efforts in digital insect monitoring using traditional computer vision struggled in complex field conditions due to issues like low contrast and variable lighting. The shift to Deep Learning, particularly Convolutional Neural Networks (CNNs), enabled automated feature learning necessary for modern agriculture.

Deep learning object detection architectures are typically classified into single-stage (e.g., YOLO series) and two-stage (e.g., Faster R-CNN) models. Single-stage detectors like YOLO are preferred for field deployment due to their exceptional speed and low latency, making them ideal for edge computing. For instance, comparative studies show YOLOv8 achieving a GPU latency as low as 1.3 ms, significantly faster than two-stage models like Faster R-CNN (latency ~54 ms).

However, traditional CNN-based YOLO models, while fast, can sometimes struggle with global contextual reasoning in cluttered scenes. Vision Transformers (ViTs) use self-attention mechanisms to aggregate context across the entire image, which is vital for fine-grained classification where subtle differences distinguish species. The current state-of-the-art involves hybrid models, which embed attention mechanisms or transformer modules into efficient YOLO backbones to combine the speed of CNNs with the high contextual expressiveness of transformers.

2. Challenges in Entomological Detection and Data Requirements :

Automated insect identification faces significant visual challenges, including high intraspecies dissimilarity, interspecies similarity (look-alikes), and the need to identify insects across various life stages and under diverse imaging conditions. This difficulty is pronounced when the system must distinguish between look-alike insects where one is a predator (e.g., Asian lady beetle) and the other is a pest (e.g., Mexican bean beetle).

To address these challenges, deep learning models require vast, high-quality datasets. The IP102 dataset, comprising over 75,000 images across 102 insect categories, with approximately 19,000 images annotated with bounding boxes, serves as a crucial benchmark for training robust pest recognition models. Additionally, models must be trustworthy, possessing the capability to avoid making predictions when confidence is low, thus facilitating human intervention when necessary.

3. Evaluation Metrics for Real-Time Systems :

Performance evaluation relies on metrics that rigorously assess both classification and localization accuracy. The Mean Average Precision (mAP) is the primary metric for overall performance. While mAP@0.5 measures basic detection quality, the critical metric for fine-grained tasks like early insect detection is mAP@[0.5:0.95], which averages the Average Precision (AP) across 10 Intersection over Union (IoU) thresholds (from 0.5 to 0.95). This metric confirms the model's ability to precisely localize small or irregularly shaped objects, essential for successful intervention. Alongside accuracy, inference speed (Frames Per Second, FPS) is mandatory for assessing a model's suitability for real-time field deployment.

Methodology :

1. Study Area, Duration, and Data Compilation :

The study focused on agricultural fields in Yavatmal District, Maharashtra, India, targeting cotton and soybean crops during the Kharif season of 2024–2025. The dataset was compiled from two sources:

1. **Benchmark Data :** A subset of the large-scale, taxonomically verified IP102 dataset.
2. **Local Field Imagery :** Real-time images collected from fields near Yavatmal,

Maharashtra, introducing variations typical of the regional environment.

The total dataset comprised approximately 15,000 expertly annotated images used for object detection training. The data was partitioned into a 60% training set, a 10% validation set, and a 30% independent test set.

2. Target Insect Taxonomy and Annotation :

The classification objective was to reliably differentiate 20 agriculturally significant insect classes, equally divided between harmful pests (e.g., Bollworms, Aphids, Whiteflies) and beneficial insects (e.g., Lady Beetles, Assassin Bugs, Green Lacewings). This dual focus supports informed IPM decision-making where predator presence dictates management strategy. All images were annotated with bounding boxes, and species identification was verified by professional entomologists.

3. YOLOv8-SAM Model Architecture and Training :

The core of the proposed system is the YOLOv8-SAM architecture. The highly efficient YOLOv8 framework was chosen as the base due to its superior speed. To enhance its precision in visually ambiguous field conditions, a Self-Attention Mechanism (SAM), inspired by Vision Transformer designs, was integrated into the feature extraction backbone. This modification allows the model to process global contextual dependencies, which is critical for fine-grained differentiation among look-alike species, overcoming the limitations of pure local feature extraction by conventional CNNs.

Input images were uniformly resized to 640 times 640 pixels. To maximize robustness against field variability, extensive data augmentation, including geometric transformations and Mosaic loading, was applied during training. The model was trained for 300 epochs using the Adam optimizer.

Observation, Data Analysis, and Calculation :

1. Flowchart: Deep Learning Pipeline :

The operational pipeline ensures rapid processing from image acquisition to final management decision.

Flowchart 1: Deep Learning Pipeline for Real-Time Field Insect Identification

Step 1: Data Captures	**Step 2: Preprocessing & Augmentation**	**Step 3: Model Training (YOLOv8-SAM)**	**Step 4: Inference & Prediction**	**Step 5: IPM Decision Support**
Field Image Acquisition (Mobile/Drone Camera) ↓	Image Loading & Resizing (640×640) ↓ Data	Optimized Loss Function (CIOU/Focal Loss) ↓	Real-Time Inference (Low Latency) ↓	Output: Species ID, Location (Bounding Box), Count ↓

Image Labeling (Bounding Boxes) ↓ Dataset Split (60% Train, 10% Val, 30% Test)	Augmentation (Blurring, Mosaic, Scale) ↓ Normalization & Color Space Adjustment	Feature Extraction (YOLO Backbone + SAM) ↓ Backpropagation & Weight Update	Non - Maximum Suppression (NMS) ↓ Confidence Score Output	IPM Action: Localized Spraying (Pests > Threshold) OR Monitoring (Beneficials Present) ↓ Trustworthiness Check (Abstention if low confidence)

2. Comparative Performance Results :

The analysis focused on comparing the proposed YOLOv8-SAM against two established baselines on the 30% independent test set: the efficient single-stage YOLOv3 and the high-accuracy two-stage Faster R-CNN.

Table 2: Comparative Performance Metrics of Proposed Model and Baseline Detectors on Field Imagery

Metric	Proposed YOLOv8 - SAM	YOLOv3 Baseline	Faster R-CNN Baseline	Interpretation
Precision	0.941	0.887	0.925	Rate of correct positive detections (TP/ (TP + FP))
Recall	0.938	0.875	0.911	Rate of all positive examples found (TP/ (TP + FN))
F1-Score	**0.940**	0.881	0.918	Harmonic mean of Precision and Recall
mAP@0.5	0.972	0.921	0.955	Overall detection quality (IoU threshold 0.5)
mAP@[0.5:0.95]	**0.785**	0.690	0.742	Average precise localization
Inference Speed (FPS)	125 FPS	90 FPS	20 FPS	Real-time capability

Latency (ms)	8.0 ms	11.1 ms	50.0 ms	Time required per image for inference

3. Fine-Grained Identification Analysis :

The YOLOv8-SAM model demonstrated superior performance, achieving an F1-Score of 0.940 and a breakthrough mAP@[0.5:0.95] of 0.785. This 0.785 score significantly surpasses both the Faster R-CNN (0.742) and YOLOv3 (0.690) baselines, confirming that the integrated Self-Attention Mechanism successfully enhanced the model's spatial accuracy and bounding box precision. This improvement is essential for reliably detecting small, early-stage insects and distinguishing fine-grained morphological differences, such as those separating beneficial species from visually similar pests. By achieving high accuracy in both pest and predator identification, the system provides reliable ecological data necessary for precise IPM decision-making.

Discussion and Conclusion :

1. Real-Time Viability and IPM Impact :

The achieved inference speed of 125 FPS (or 8.0 ms latency) is crucial for the practical viability of the system. This high-speed performance ensures the YOLOv8-SAM architecture is suitable for deployment on low-cost, resource-constrained edge devices, such as mobile phones or drones. This resource efficiency is key to lowering the technical barriers to adoption for smallholder farmers, thereby promoting equitable access to precision agricultural tools.

By achieving high-fidelity detection and localization (mAP@[0.5:0.95] = 0.785) across pest and beneficial classes, the system enables informed, threshold-based IPM. Farmers can confidently monitor the presence and quantity of natural enemies, which reduces the dependence on chemical interventions, conserves biological control agents, and minimizes environmental harm. The ability of the model to avoid predictions when uncertain further enhances its trustworthiness in critical agricultural scenarios.

2. Limitations and Future Directions :

Despite its superior performance, the system faces limitations inherent to field deployment, particularly reduced recognition effect under extreme environmental conditions like heavy rain or reduced daylight. Future research should focus on:

1. **Multi-Modal Data Integration :** Integrating environmental data (temperature, humidity) with visual data to enable proactive predictive modeling of pest outbreaks.
2. **Generalization :** Conducting extensive longitudinal trials across different growing seasons and regional climates to ensure universal applicability.

In conclusion, the YOLOv8-SAM hybrid detector successfully balances speed and precision for real-time, fine-grained insect identification. The model's superior mAP@[0.5:0.95] and high FPS make it a transformative, trustworthy solution for implementing sustainable Integrated Pest Management, contributing significantly to improved crop protection and productivity in regions like Maharashtra and beyond.

References :

- Ahmad, I., Yang, Y., Yue, Y., Ye, C., Hassan, M., Cheng, X., Wu, Y., & Zhang, Y. (2022). Deep Learning Based Detector YOLOv5 for Identifying Insect Pests. *Applied Sciences*, *12*(19), 10167. https://doi.org/10.3390/app121910167
- Alabama Cooperative Extension System. (2024). *Beneficial insects in cotton: Predators*.
- Bjerge, K., Iosifidis, A., & Høye, T. T. (2024). InsectNet: Real-time identification of insects using an end-to-end machine learning pipeline. *PNAS Nexus*, *4*(1), pgae575. https://doi.org/10.1093/pnasnexus/pgae575
- Braukmann, T. W. A., et al. (2019). Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, *19*(3), 711–727.
- Elbrecht, V., et al. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, *7*, e7745.
- Gui, J., Xu, H., & Fei, J. (2023). Detection of *Leguminivora glycinivorella* Matsumura Based on an A-ResNet Meta-Learning Model. *Sensors*, *23*(2), 678.
- Høye, T. T., Ärje, J., Bjerge, K., Hansen, O. L. P., Iosifidis, A., Leese, F., Mann, H. M. R., Meissner, K., Melvad, C., & Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences*, *118*(2), e2002545117.
- Kasinathan, T., Singaraju, D., & Uyyala, S. R. (2020). Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture*.
- Meshram, A., Meshram, K., Vanalkar, A., Badar, A., et al. (2025). Deep learning for cotton pest detection: Comparative analysis of CNN architectures. *Indian Journal of Entomology*. https://doi.org/10.55446/IJE.2025.2742
- Mississippi State University Extension. (n.d.). *Common beneficial arthropods found in field crops*.
- Tannous, M., Stefanini, C., & Romano, D. (2023). A Deep-Learning-Based Detection Approach for the Identification of Insect Species of Economic Importance. *Insects*, *14*(2), 148.
- Wagner, D. L. (2020). Insect declines in the Anthropocene. *Annual Review of Entomology*, *65*, 457–480.
- Wu, X., Zhan, C., Lai, Y., Cheng, M.-M., & Yang, J. (2019). *IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- Xie, C., Wang, Z., Chen, G., et al. (2015). Automatic classification for field crop insects via multiple features. *Information Processing in Agriculture*, *8*(3), 446–457.
- Zahan, I. I., et al. (2024). Ecological and socio-economic trade-offs of AI-based plant disease detection systems in agriculture. *PMC*.