Grenze International Journal of Engineering and Technology, June Issue

Enhancing Insect Pest Detection with Attention-Augmented YOLOv8

Aravind Sathishan[1] and Sonia[2]
[1]Dept. of Mathematics, Chandigarh University, Gharuan, Punjab
Email: aravindsathishan18@gmail.com
[2]Professor, Dept. of Mathematics, Chandigarh University, Gharuan, Punjab
Email: sonia.e8843@cumail.in

***Abstract*— Reliable vision-based crop pest detection is crucial for precision agriculture to prevent economic losses. However, classifying and localizing tiny insect instances is incredibly challenging. In this work, we present an enhanced YOLOv8 model specially tailored for insect pest detection in agriculture. We propose enhancing the popular YOLOv8 object detector with convolutional block attention modules (CBAM) and squeeze-and-excitation (SE) blocks to boost feature representations of small insect pests amidst complex field images. CBAM and SE blocks incorporate attention mechanisms to focus convolutional filters on the most informative regions. We integrate these modules within YOLOv8's feature extractor backbone for heightened awareness of sparse, tiny pest targets. We conduct experiments on a real-world agriculture insect pest dataset containing images collected under various practical field conditions. Custom-made pest dataset was used to train YOLOv8 models incorporated with CBAM and SE Net. Quantitative results showed state-of-the-art pest detection accuracy, with a mean average precision (mAP) of 94.5% and an F1 score reaching 0.91 for the model utilizing both CBAM and SE Net. By integrating these additional modules, YOLOv8 leveraged feature recalibration and scaling to achieve improved performance on this pest detection task. The attention modeling provides noticeable gains in accurately localizing clustered and occluded insects while reducing false detections. Our improved Attention augmented-YOLOv8 model strikes an optimal balance between accuracy gains and on-device computational constraints for viable insect pest monitoring systems to enable sustainable agriculture.**

***Index Terms*— CBAM, SE Net, YOLOv8, Spatial attention, Channel attention, Insect detection.**

I. Introduction

Insect pests inflict substantial damage to crops globally, leading to significant financial losses each year through toxicity introduction, physical harm to plant tissues, and the spread of diseases. This results in both quantitative and qualitative losses in marketable crop yields, escalating farmers' production costs as they invest more in pest management. Industry study data reveals a staggering annual global crop loss of $470 billion as of 2019 due to insect pest devastation [1].. To address the timely and cost-effective prevention of insect pests, especially those specific to certain species, early intervention is crucial. However, traditional manual pest scouting techniques, such as physical field sample analysis and trap assessments, are labour intensive, subjective, and slow for widespread use across vast farmlands. Contemporary technical methods, like spectrum imaging and pheromone/bait traps [2]., analyze proxy signals but lack cross-context generalizability due to their reliance on deep domain knowledge.

Grenze ID: 01.GIJET.10.2.128

Alternatively, data-driven learning algorithms, such as computer vision, enable direct automated analysis of imagery for inconspicuous identification and evaluation of insect populations. Recent studies demonstrate the effectiveness of deep convolutional neural networks (CNN) in automatically classifying pest varieties [3].and estimating severity of damage with an accuracy rate exceeding 96% [4]..

Deep learning-based object detection algorithms, exemplified by YOLOv5 and Faster RCNN, have significantly advanced accurate and real-time automated pest detection in Precision Agriculture management [5].. Despite their proficiency, persistent challenges involve identifying and localizing small insect occurrences amid strong occlusion, irregular forms, grouped appearances, and visual uncertainties [6].. Visual attention modeling is vital for finer discrimination, with CNNs employing channel and spatial attention processes, mimicking human vision [7].[8].. Lightweight attention modules, such as Convolutional Block Attention Modules (CBAM), balance channel and spatial attention, enhancing accuracy without excessive processing overhead [9].. Squeeze-and-Excitation (SE) networks highlight pertinent features while suppressing less valuable ones, effectively modeling channel-wise feature interdependencies [10].. SE blocks adjust feature channels dynamically via channel-wise pooling, excitation, and activation gating. Their integration in convolutional networks enhances performance in computer vision, notably object detection, with minimal computational overhead. SE network topologies demonstrate robustness to scale variation, occlusion, and object deformations, making them well-suited for accurate insect pest identification in challenging field imagery. To fill a significant void in existing pest detection models, addressing challenges like small object scales and appearance ambiguities in agriculture, we propose integrating CBAM. This lightweight solution prioritizes essential areas in detectors, ensuring reliability under real-world uncertainties at minimal cost. However, the absence of systematic analysis across detectors and attention methods for insect challenges calls for focused benchmarking to refine robust modeling strategies. Inspired by the demand for precise pest detection in precision agriculture and acknowledging attention's effectiveness in object recognition, our improved insect detector incorporates channel and spatial attention modeling via CBAM blocks and SE Net. Our contributions encompass:

- Proposing targeted CBAM and SE Net-based architectural modifications to address the limitations of modern insect recognition within the YOLOv8 detector.
- Conducting extensive empirical evaluations using real-world field photos, assessing accuracy-efficiency trade-offs and comparing model performance with cutting-edge insect detection methods.

II. RELATED WORKS

Recent progress in deep CNNs offers significant potential for automated insect recognition without extensive manual feature engineering. A deep learning pipeline integrating channel spatial attention into the CNN backbone and region proposal network achieved an mAP of 87.43% for pest monitoring [11].. Attention mechanisms have been widely explored to augment CNN-based models by focusing on informative visual patterns. For detection in aerial imagery, Zhang and Chen integrated channel and spatial attention in their U-YOLOv5 model to significantly boost small object detection performance on drone datasets, and achieved 5% im-proved mAP compared to baseline model [12].. In the domain of precision agriculture, Mamdouh [13]. leveraged data augmentation and negative sampling to train a robust deep pipeline for detecting olive fruit flies in trap images, achieving 96.68% mean Average Precision. Kim Bjerge deployed an intelligent computer vision system having 89% classification precision across 8 species for continuous monitoring of flower visiting insects, demonstrating ecological insights from longitudinal abundance and behavior data [14].. Zhao and Liu proposed and improved deep CNN including both parallel attention mechanism and residual blocks showcased 98.17% accuracy for pest images. Advanced detectors like Faster R-CNN and YOLO have also been customized for crop pest monitoring applications. Kaur and Kaur [15]. benchmarked various CNN architectures, finding Faster R-CNN with ResNet-101 backbone delivered the best balance of 74.77% mAP and real-time performance for multi-pest recognition. Attention mechanisms have also shown promise for remote sensing change detection [16]. and fine-grained species classification with scarce data [17].. By integrating CBAM and multi scale feature extraction, the MTCNet achieved state of the art in remote sensing [16].. An insect classifier integrating SE Net for focusing on informative fine-grained features, achieving improved species recognition accuracy over alternatives on an Australian atlas insect dataset [17].. Most similar to our work, Wang and Liu enriched YOLOv5 with squeeze-and-excitation and convolutional block attention modules (CBAM) to significantly boost detection of destructive red palm weevils to 90.1% mAP [18].. Specific to the insect classification domain, a fly species recognition method combined improved RetinaNet with CBAM to achieve 90.38% accuracy in identifying fly species [19].. YOLOv8 achieves state-of-the-art object detection. Our study enhances it with attention augmentation for insect pest monitoring, filling a previously identified gap.

III. Materials And Methods

A. Dataset collection

The IP102 dataset comprises over 19,000 images spanning 102 categories, exhibiting a natural long-tailed distribution [20].. These categories encompass members of the Coleoptera, Araneae, Hemiptera, Hymenoptera, Lepidoptera, Odonata, and Diptera families. Additional data were sourced from the internet, selectively choosing pest images filmed in adverse conditions, resulting in 3,486 additional images integrated into the modified IP102 dataset [21].. The combined dataset now totals 23,182 images, divided into training (21,124 images), validation (1,029 images), and testing (1,029 images) sets. Iterative model training and evaluation identified the most efficient model in terms of mean Average Precision (mAP). The study explored the method's generalizability and its ability to identify pests across images with diverse backgrounds, validating its robustness.

B. Data Annotation and Augmentation

The annotation process was meticulously carried out manually using CVAT.ai, ensuring precision in localizing and identifying objects within each image. Bounding boxes were employed to identify and mark various objects in each image, offering flexibility to recognize and label multiple objects within a single image. A corresponding text file, sharing the image's name, was generated for reference and correlation with the source images. This text file detailed information about each recognized object in the image, providing comprehensive details characterizing the bounding box of the detected object, as specified by (1):

$$(class\_identifier, Xcentre, Ycentre, width, height) \qquad (1)$$

Typically, increasing data improves model performance[22].. Yet, acquiring ample training data is challenging. To tackle this, data processing includes generating additional samples through techniques like flipping, Gaussian Blur, and rotation. With 23,182 photos, no further augmentation was required, ensuring the model's generalization

C. YOLOv8

In January 2023, the Ultralytics team developed YOLOv8, the latest addition to the cutting-edge You Only Look Once (YOLO) series, aimed at advancing object detection capabilities. While the public can utilize YOLOv8 for training and testing, the official paper is yet to be released. Fig. 1, provided by GitHub user Range King, offers YOLOv8's architecture [23].. Comprising the Head and the Backbone, YOLOv8 entrusts the Backbone with creating feature pyramids post feature extraction [24].. The YOLO head takes on the prime responsibilities of identifying and displaying bounding boxes along with objectness scores. Renowned for its rapid and accurate detection, YOLO divides images into N grids with S x S sections within each, as part of its single-stage object detection algorithm [22].. YOLOv3 served as the initial baseline within the YOLO family, gaining wide adoption, followed by subsequent variants like YOLOv4, YOLOX, and the recent YOLOv8, pushing the boundaries of state-of-the-art trade-offs. Distinguishing itself from predecessors, YOLOv8 features a dual-path architecture incorporating a central spine and parallel decoupled neck. This design integrates cross-stage partial connections, attention, and spatial pyramid pooling to enrich feature fusion for objects at varying resolutions. The optimized encoder-decoder design coupled with multi-level predictions significantly enhances YOLOv8's capability in handling small objects, a critical aspect for real-world pest detection. Key attributes crucial for insect pest detection in YOLOv8 include a lightweight yet potent CSPDarkNet backbone, ensuring efficiency even for edge devices. The Encoder-decoder pathway preserves both high and low-level spatial details, while the multi-scale specialized neck enhances pyramidal object feature extraction. Anchor-free per-pixel predictions improve localization for cluttered tiny instances, providing tangible boosts in detection accuracy under uncertainties like occlusion and orientation changes, vital for agricultural viability. YOLOv8 further enhances photos during online training, introducing marginally distinct photo variants for each epoch [25].. The use of the mosaic augmentation technique during training, although effective, is disabled for the final ten epochs to prevent a decrease in precision [24]..

D. CBAM Attention Mechanism

The Convolutional Block Attention Module (CBAM) enhances CNN's representational capabilities by recognizing spatial and channel-wise connections in the input. Specifically designed to boost attention mechanisms, CBAM employs twin attention modules: the Spatial Attention Module (SAM) and the Channel Attention Module (CAM), capturing both channel-wise and spatial interdependence. This dual-attention approach enables the CBAM model to focus on critical aspects across various dimensions, as depicted in Fig. 2. Integrated with YOLOv5 for a lightweight YOLO model in pest detection, CBAM demonstrates improved mAP and F1 scores [22]., validating its efficacy in enhancing the performance of attention mechanisms for computer vision tasks.

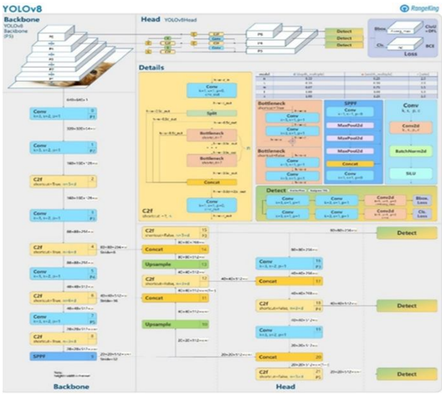

Figure 1. Architecture of YOLOv8 [23]

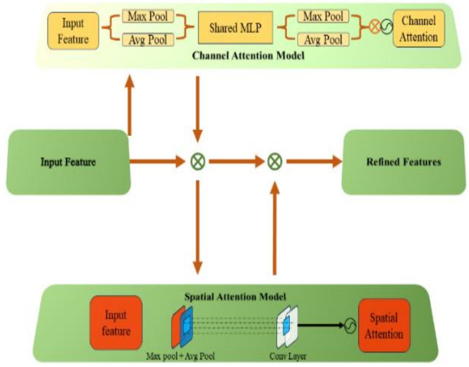

Figure 2. CBAM Architecture

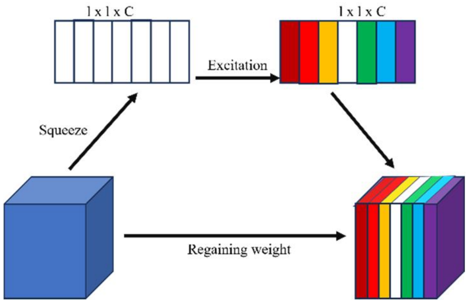

Figure 3. SE-Net

SAM captures spatial dependencies within the input feature map through global average pooling (GAP) and a Multi-Layer Perceptron (MLP). This process highlights crucial spatial locations in the image by deriving spatial attention weights. Conversely, the CAM focuses on channel-wise dependencies using global max pooling (GMP) and an MLP, emphasizing important channel information. The CBAM output combines both spatial attention (SA) and channel attention (CA) weights, contributing to an enhanced overall feature representation, as expressed in (2).

$$CBAM\ Output = SA\ Weight + CA\ Weight \qquad (2)$$

Channel attention compresses feature maps into one-dimensional vectors by pooling within each channel. This process reduces spatial dimensionality through a shared network. Spatial information is combined within feature maps. The channel attention map is formed by summing up each component. CBAM's attention mechanisms effectively handle variations in lighting, background, and pest orientations, contributing to improved generalization across diverse environmental conditions for real-world insect pest detection [9].. Comparative studies consistently show significant accuracy improvements with CBAM integration, lower false positives and negatives, enhancing detection process, precision and recall metrics in insect pest identification processes [26]..

E. Squeeze and Excitation Net

The Squeeze-and-Excitation Network, abbreviated as SE-Net, is an attention mechanism designed to adaptively enhance feature maps, thereby augmenting the representational capacity of Convolutional Neural Networks (CNNs). Introduced by Jie Hu, Li Shen, and Gang Sun in their 2018 paper "Squeeze-and-Excitation Networks," SE-Net has gained widespread acclaim for its utility in various computer vision applications, such as object detection and image categorization [10].. Fig. 3 gives the structure; SE Net enables the network to focus on informative channels, enhancing feature discrimination. By modeling channel-wise feature dynamics, SE blocks have proven effective at learning robust representations invariant to scales, occlusion and deformation - highly desirable for tough cases of tiny, swarming insect pests. Adaptive recalibration allows the model to adapt to varying channel importance, making it more robust to different input variations. Despite its effectiveness, SE Net introduces minimal computational overhead, making it an efficient enhancement.

F. Experimental Setup

During the training phase, 21124 images of 102 classes of insects were used to train and validate each underlying model for insect recognition. We set the photo resolution to 640 x 640 pixels for the training phase. This is because yolo models typically rely on the image size of the data. YOLOv8 from ultranalytics was installed. A number of transformations were applied to the dataset, including random flips in both vertical and horizontal directions, brightness adjustments, and sharpness adjustments. Initially, the YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x models were used to train these data. At first, the models were trained over five epochs. But based on the output graph for these five epochs, it appeared that the model required more balance. The precision and recall grew significantly while the loss curves gradually shrank, indicating the possibility of experimenting with larger epochs. Compared to five epochs, obviously the output of twenty produced better results. The model was trained for a total of 25 epochs in another experiment. Nevertheless, this led to overfitting of the model. Consequently, a 20-epoch model was decided upon.

Next, we made modifications to the Yolov8x architecture, adding CBAM layers to the core of the YoloV8 architecture. To do this, the CBAM and SE Net layers were added to the yolov8x.yaml configuration file. Furthermore, the YOLOv8/models directory's common.py file now includes both the CBAM and SE Net function. The training process is the same as for the conventional model. The configuration file path needs to be changed in the training parameter. In the end, a model that included both was trained and put into use. A learning rate of "1 × 10−3" was utilized in the training process and AdamW optimizer was employed. After running the model for 20 epochs, the weights are adjusted and fine-tuned. Every model was built using Pytorch. The final output of the model contained the likelihood of a certain class and the position prediction box that gives away the location bounding box of the target insect pest categories. The experiment is done using Nvidia Tesla T4(16GB) GPU enabled Google colab, with V4 16GB RAM. The dataset consists of 1.1 GB of images. Baseline 3.9 version of python is employed. Table I defines each models specific training methods. To evaluate the performance of the underlying models, we used a validation set consisting of 1029 images. To test the models, we employed a total of 1029 unexposed images for the detection of insect pests.

TABLE I. YOLOV8 MODELS UNDER STUDY

YOLO Models	No.of Layers	No.of Parameters	Training Time (Hours)	Inference time (Seconds)
YOLOv8n	225	3346746	9	0.18
YOLOv8s	225	11175074	12	0.24
YOLOv8m	295	25915378	16	0.36
YOLOv8l	365	43708482	19	0.44
YOLOv8x	365	68250834	23	0.5
YOLOv8x-modified	326	62686462	21	0.48

IV. RESULTS AND DISCUSSIONS

Various evaluation metrics were employed to gauge the accuracy of the object detector on the given dataset, with Intersection over Union (IoU) being the primary metric. IoU quantifies the overlap between predicted and ground truth bounding boxes, commonly used in object detection problems. A threshold of 0.5 was set, requiring a minimum 50% overlap for correct detection. Mean Average Precision (mAP) was calculated at IoU = 0.5 and IoU between 0.5 and 0.95 to assess performance across different IoU thresholds. Precision, recall, and F1 scores were also examined. IoU is computed as TP (true positive) if it exceeds the threshold and as FP (false positive) if it is low. Equations (3)–(7) computed precision, recall, and mAP using TP, FP, TN (true negative), and FN (false negative). Precision measures correct detections, while recall reflects successfully detected actual objects. F1 score balances precision and recall. These metrics collectively evaluated the object detector's ability to localize and classify objects accurately in the dataset, providing a comprehensive understanding of its strengths and weaknesses.

$$IoU = Area\ of\ overlap\ /Area\ of\ Union \quad (3)$$

$$Precision = TP/(TP+FP) \quad (4)$$

$$Recall = TP/(TP+FN) \quad (5)$$

$$mAP = (1/N) \times \sum_{i=1}^{N} APi \quad (6)$$

$$F1 = 2\times(Precision\times Recall)/(Precision+Recall) \qquad (7)$$

Following model training, we assessed precision (P), recall (R), mAP@IoU:0.5, and mAP@IoU:0.95. Fig. 4–8 illustrate fluctuations in these metrics, particularly in the initial ten training epochs, prompting an extension of the training epochs to minimize variability. Notably, YOLOv8s and YOLOv8m exhibited declining recall and precision post the 20th epoch, indicating overfitting and reduced performance on the test data. The decline persisted beyond the 20th epoch. Fig. 4-8 depicts the outcomes of six model iterations.

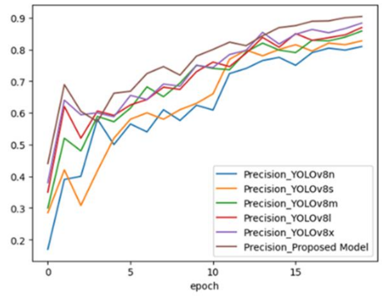

Figure 4. Precision of various models

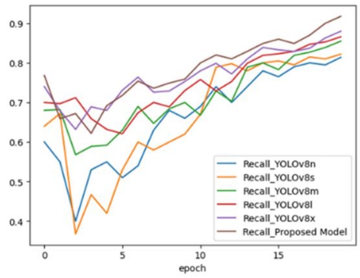

Figure 5. Recall of various models

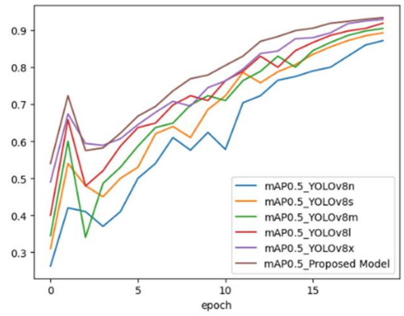

Figure 6. mAP@0.5 of various models

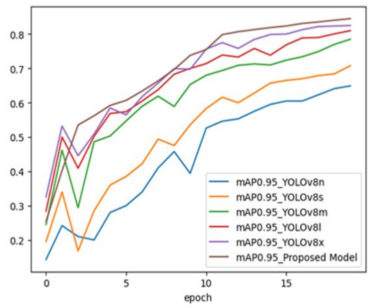

Figure 7. mAP@0.95 of various models

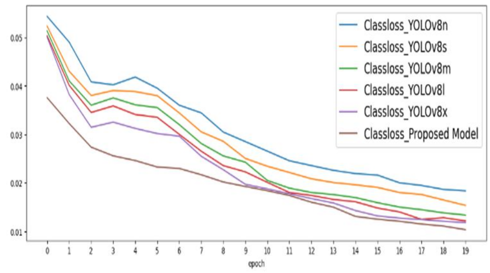

Figure 8. Loss curve of various models

Our experimental results, derived from YOLOv8 models (n, s, m, l, x, and modified x) trained on a single pest dataset, were analyzed using a test set of 1029 unseen photos. Table II provides statistical insight into the detection performance across six distinct YOLO topologies. Hyperparameters were consistently maintained for model comparisons. Notably, YOLOv8n exhibited lower precision, recall, minimum mAP@0.5, mAP@0.5:0.95, F1 score compared to other models. Conversely, the 'YOLOv8x' model demonstrated exceptional recognition accuracy, particularly suited for real-time pest detection in dynamic field environments. Random test set images showcased the model's ability to distinguish insect species across diverse backgrounds, plant types, and lighting conditions are displayed in Fig. 9, making it highly adaptable in practical scenarios. Comparative analysis with other implemented models is detailed in Table III.

This study presents a meticulously designed CNN-attention-based variant of YOLOv8, customized for robust fine-grained insect detection in agricultural images amidst real-world uncertainties. The integrated CBAM boosts accuracy while preserving efficiency, ideal for reliable pest monitoring across diverse field conditions and hardware platforms, supporting intelligent precision agriculture. Future efforts involve extending applications to drone video streams, incorporating online parameter adaptation for dynamic data, and achieving fully automated insect modeling, aiming to enhance agriculture's productivity, sustainability, and efficiency with the refined YOLOv8 model.

TABLE II PERFORMANCE COMPARISON OF ALL SIX MODELS

Model	Training				Validating				Testing			
	P	**R**	**F1**	**mAP**	**P**	**R**	**F1**	**mAP**	**P**	**R**	**F1**	**mAP**
YOLOv8n	0.809	0.814	0.811	0.872	0.807	0.820	0.813	0.873	0.813	0.803	0.808	0.862
YOLOv8s	0.827	0.822	0.824	0.893	0.829	0.824	0.826	0.893	0.833	0.81	0.821	0.885
YOLOv8m	0.858	0.855	0.856	0.905	0.860	0.853	0.856	0.906	0.842	0.838	0.840	0.898
YOLOv8l	0.869	0.866	0.867	0.919	0.867	0.866	0.866	0.920	0.85	0.848	0.849	0.908
YOLOv8x	0.883	0.874	0.878	0.924	0.885	0.873	0.879	0.925	0.869	0.851	0.860	0.912
YOLOv8x modified	0.904	0.917	0.910	0.945	0.905	0.859	0.908	0.934	0.882	0.904	0.893	0.924

TABLE III PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	**Dataset**	**One/multiple-insect inside picture**	**Accuracy**
VGG19 [13].	Olive fruit fly dataset	one	89.22%
PestsLite [3].	Mydataset	multiple	90.07%
DenseNet-121 [27].	Grain insect	multiple	88.06%
CNN [28].	Wang and Xie dataset	one	90%
VGG16 [29].	Rice pest dataset	multiple	88.68%
Improved DeiT [29].	Field and internet	multiple	87.67%
Proposed Model	IP102	multiple	94.5%

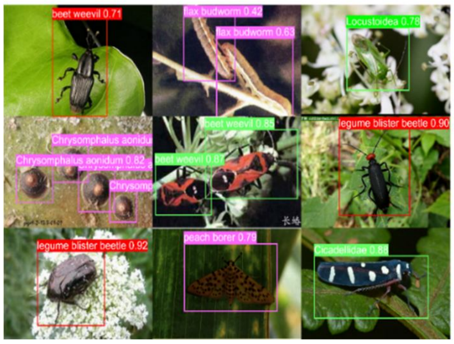

Figure 9. YOLOv8x-modified Output

This study presents a meticulously designed CNN-attention-based variant of YOLOv8, customized for robust fine-grained insect detection in agricultural images amidst real-world uncertainties. The integrated CBAM boosts accuracy while preserving efficiency, ideal for reliable pest monitoring across diverse field conditions and hardware platforms, supporting intelligent precision agriculture. Future efforts involve extending applications to drone video streams, incorporating online parameter adaptation for dynamic data, and achieving fully automated insect modeling, aiming to enhance agriculture's productivity, sustainability, and efficiency with the refined YOLOv8 model.

V. CONCLUSIONS

This study aimed to develop an insect pest identification system capable of accurately identifying pests in image files, utilizing our proposed modified-x and six distinct variants of YOLOv8, ranging from the smallest YOLOv8-Nano to the largest x variant. A custom insect photo dataset was curated from Kaggle and the internet, capturing images in real outdoor settings. The system was trained, validated, and tested on Google Colab using an Nvidia Tesla T4 16GB GPU. Among all YOLO models, the modified YOLOv8x and YOLOv8x demonstrated superior accuracy after extensive training on the custom pest dataset, showcasing suitability for real-time deployment in agricultural field conditions. The integration of selective attention mechanisms in the modified YOLOv8x variant further enhanced detection efficacy. The modified model exhibited overall higher mean Average Precision (mAP) gains of 2% and improved F1 scores by 0.03 over the standard YOLOv8x system. These performance gains underscore the value of attention modeling in our customized detector for reliable insect pest monitoring applications. YOLOv8x emerges as a state-of-the-art model in computational entomology, surpassing previous YOLO models in both accuracy and real-time performance. Looking ahead, the potential integration of YOLOv8x with a vision transformer, such as a Swin transformer, could significantly accelerate detection speed. The established model holds promise for farmers, enabling reliable insect detection in field pest monitoring devices, ultimately reducing the reliance on pesticides and other potentially harmful pest control methods.

REFERENCES

[1] Report: “Global Crop Protection Chemicals Market Growth 2021-2026,” Fortune Business Insights.

[2] Barbedo, and J. G. Arnal, "A review on the main challenges in automatic plant disease identification based on visible range images," Biosystems engineering 144, 2016, pp. 52-60.

[3] Q. Dong, L. Sun, T. Han, M. Cai, C. Gao, “PestLite: A Novel YOLO-Based Deep Learning Technique for Crop Pest Detection,” Agriculture 2024, 14(2):228.

[4] J. P. Gonçalves, F. Pinto, Daniel M., F. M. M. Villar, J. G. A. Barbedo, and E. M. Del Ponte, “Deep learning architectures for semantic segmentation and automatic estimation of severity of foliar symptoms caused by diseases or pests,” Biosystems Engineering, 2021.

[5] I. Ahmad, Y. Yang, Y. Yue, C. Ye, M. Hassan, X. Cheng, Y. Wu, and Y. Zhang, “Deep Learning Based Detector YOLOv5 for Identifying Insect Pests,” Applied Sciences, 2022.

[6] J. Liu, and X. Wang, “Plant diseases and pests detection based on deep learning: a review,” Plant Methods 17, 22 ,2021.

[7] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, “Attention branch network: Learning of attention mechanism for visual explanation,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10705-10714.

[8] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images." Medical image analysis 53, 2019, pp. 197-207.

[9] S. Woo, J Park, J. Y. Lee, I. S. Kweon, "CBAM: Convolutional block attention module," European conference on computer vision, Springer, Cham, 2018, pp. 3-19.

[10] J. Hu, L. Shen and G. Sun, “Squeeze-and-excitation networks,” In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132-7141.

[11] D. Rajalakshmi, V. Monishkumar, S. Balasainarayana, M. S. R. Prasad, “Deep Learning Based Multi Class Wild Pest Identification and Solving Approach Using Cnn,” Annals of RSCB, 2021, pp. 16439–16450.

[12] G. Zhang, X. Chen, X. Tan, J. Zhang, and X. Lan, “U-YOLO: Improved YOLOv5 for Small Object Detection on UAV-Captured Images,” in International confrence on Cognitive Computation and systems,2022.

[13] N. Mamdouh and A. Khattab, “YOLO-Based Deep Learning Framework for Olive Fruit Fly Detection and Counting,” in IEEE Access 2021, 9, 84252–84262.

[14] K. Bjerge, H. M. R. Mann, and T. T. Høye, “Real-time insect tracking and monitoring with computer vision and deep learning”, Remote Sensing in Ecology and Conservation, 2022.

[15] Arun, R. Arumuga, and S. Umamaheswari, “Effective and Efficient Multi-crop Pest Detection Based on Deep Learning Object Detection Models,” Journal of Intelligent & Fuzzy Systems, vol. 43, 2022.

[16] W. Wang, X. Tan, P. Zhang and X. Wang, "A CBAM Based Multiscale Transformer Fusion Approach for Remote Sensing Image Change Detection," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, 2022.

[17] Y. J. Park, G. Tuxworth and J. Zhou, "Insect Classification Using Squeeze-and-Excitation and Attention Modules - a Benchmark Study," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3437-3441.

[18] S. Wu, J. Wang, L. Liu, D. Chen, H. Lu, C. Xu, R. Hao, Z. Li, and Q. Wang Q, “Enhanced YOLOv5 Object Detection Algorithm for Accurate Detection of Adult Rhynchophorus ferrugineus,” Insects. 2023.

[19] Y. Chen, X. Zhang, W. Chen, Y. Li and J. Wang, "Research on Recognition of Fly Species Based on Improved RetinaNet and CBAM," in IEEE Access, vol. 8, 2020, pp. 102907-102919.

[20] IP102 Available online: https://www.kaggle.com/mistag/arthropod-taxonomy-orders-object-detection-dataset

[21] Insect images availabe online: https://www.insectimages.org/

[22] N. Kumar, Nagarathna, and F. Flammini, “YOLO-Based Light-Weight Deep Learning Models for Insect Detection System with Field Adaption,” Agriculture 2023, 13, 741.

[23] YOLOv8: https://github.com/ultralytics/ultralytics/issues/189

[24] J. Terven, and D. A. Cordova-Esparza, “Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond,” arXiv preprint arXiv:2304.00501, 2023.

[25] S. Tamang, B. Sen, A. Pradhan, K. Sharma and V. K. Singh, “Enhancing COVID-19 Safety: Exploring YOLOv8 Object Detection for Accurate Face Mask Classification,” In International Journal of Intelligent Systems and Applications in Engineering (2023), 11(2).

[26] S. Zhang, L. Wen, X. Bian, and Z. Lei, “Single-shot Refinement Neural Network for Object Detection,” in IEEE conference on computer vision and pattern recognition, 2018.

[27] J. Redmon, and A. Farhadi, ”YOLO9000: Better, faster, stronger,” In Proceedings of the CVPR, Honolulu, HI, USA, 2017;pp. 7263–7271.

[28] T. Kasinathan, D. Singaraju, and S.R. Uyyala, ”Insect classification and detection in field crops using modern machine learning techniques.” Information Processing in Agriculture, 2020.

[29] X. Yu, and J. Zheng, "A Review on the Application of Deep LearningMethods in Detection and Identification of Rice Diseases and Pests," Computers, Materials & Continua 78.1 (2024), 2024.