RESEARCH ARTICLE
Indian Journal of Agricultural Research, Volume 60 Issue 9: 1450-1457 (September 2026)

A Novel AI-Driven Approach for Synthetic Data Generation in Camouflaged Insect and Pest Detection

Muralidhar Rao Lokerao[1], K. Basava Raju[1]
10.18805/IJARe.A-6613

ABSTRACT

Background: Data scarcity of agricultural computer vision data in the area of Camouflaged insect and pest detection remains a challenging problem, despite recent advancements in Camouflaged Object Detection (COD) and transformer based architectures. To address this problem, we introduced a novel AI-driven synthetic image generation methodology that leverages computer vision models, Large Language Models(LLMs) and Generative AI to create challenging synthetic camouflaged images of insects and pests. This synthetic data helps enhance the generalisation capabilities of COD models.

Methods: The proposed methodology includes an iterative process involving multiple AI agents performing (1)extraction of pest-crop knowledge using LLMs, (2) collecting pest and plant images in an automated way, (3) separating out the foreground pest from background, (4) generating synthetic images with varied camouflage parameters, (5) evaluating using existing detection models. (6) incorporating images that are not detected into the training dataset, enabling iterative dataset enhancement (7) retraining the model. This approach accelerates the development of intelligent, sustainable and scalable agri-tech solutions by providing critical training data that is inexhaustible and relevant.

Result: We have experimentally validated the proposed iterative methodology on the detection accuracy for camouflaged hornworm pests on tomato plants. In our ablation studies, we have seen that the baseline mAP50 for models trained on web-scraped images was 0.170 when tested with camouflaged images, whereas the models trained with images generated using our methodology have shown a mAP50 value of 0.931 when tested with the same set of camouflaged images. Using the Gradient-weighted Class Activation Mapping (Grad-CAM), we provided evidence on how the iterative blending process in our methodology is able to regularise the model and focus on the biological anatomy and texture of the pests instead of just the edges and boundaries, proving that the methodology not only solves the data scarcity of camouflaged pest images but also improves the robustness of models in complex environments.

Key words: Agricultural computer vision, AI agents, Camouflaged object detection, Large language models, Pest detection, Synthetic data generation.

INTRODUCTION

There are significant economic losses and food shortages due to the reduced agricultural productivity and crop quality affected by the insects and pests. Traditional methods to identify and manage are labour-intensive and not scalable for large agricultural landscapes. Artificial Intelligence (AI) and computer vision brought enhanced efficiency and accuracy in pest management. With object detection, it is now possible to autonomously identify pests and diseases with improved accuracy in classification and precise localization.

One big challenge in pest detection is the camouflage phenomenon that the pests exhibit. Many species of pests blend seamlessly with their surroundings. This makes it difficult both for naked eye as well as computer vision systems to detect them. They exhibit camouflage strategies like background matching, disruptive coloration, transparency, or masquerade.

Background and related work

The related work in this field includes traditional computer vision, sophisticated hybrid CNN-Transformer architectures and multi-modal Vision-Language models (Mehta *et al.,* 2025; Dalal and Mittal, 2025; Hao *et al.*, 2024; Truong *et al.,* 2025; Mondal, 2020). Seasonal variations of crops and diseases were handled using a dual encoder architecture (Sharma and Kumar, 2026). Other specialised models that have made progress in leveraging attention mechanisms and feature fusion for concealed object detection include PRNet (Wang *et al.*, 2023), CamoFormer (Yin *et al.*, 2022), SP-YOLO (Tang *et al.*, 2025), SENet (Hao *et al.*, 2024) and Hybrid-COD. Moreover, the introduction of large-scale multi-modal datasets like MM-CamObj (Ruan *et al.*, 2025), Multimodal Insect Dataset and the integration of advanced

[1]Anurag University Venkatapur, Ghatkesar Medchal-Malkajgiri-500 088, Hyderabad, Telangana, India.

Corresponding Author: Muralidhar Rao Lokerao, Anurag University Venkatapur, Ghatkesar Medchal-Malkajgiri-500 088, Hyderabad, Telangana, India.
Email: 23eg305a25@anurag.edu.in

ORCIDs: https://orcid.org/0009-0000-3790-6234, https://orcid.org/0000-0002-0192-0592

How to cite this article: Lokerao, M.R. and Raju, K.B. (2026). A Novel AI-Driven Approach for Synthetic Data Generation in Camouflaged Insect and Pest Detection. *Indian Journal of Agricultural Research*. **60(9):** 1450-1457. doi: 10.18805/IJARe.A-6613.

Submitted: 22-06-2026 **Accepted:** 10-07-2026 **Online:** 25-07-2026

modalities like hyper-spectral imaging helped in the efforts to overcome camouflaged effects (Truong *et al.*, 2025; Hupel and Stütz, 2025; Gao *et al.*, 2025). In spite of these advancements, issues like data scarcity (challenges described by Kappali *et al.* (2024), model generalisation, efficiency for edge deployment and the need for Explainable AI remain (Dalal and Mittal, 2025; Shoaib *et al.*, 2025; Tang *et al.*, 2024).

Foundations of computer vision and architectural evolution in agricultural pest management

Early computer vision relied on handcrafted features (colour, texture), which lacked generalisation. Deep learning-specifically CNNs-enabled hierarchical feature learning, achieving high accuracy. Modern agricultural AI now incorporates object detection variants like YOLO series, EfficientNet and hybrid architectures.

Vision Transformers have surpassed LSTMs in capturing global context through self-attention, which is vital for detecting camouflaged objects. Hybrid CNN-Transformer models excel by using CNNs for local features and Transformers for global understanding. Notable architectures include CamoFormer (Masked Separable Attention for segmentation), SP-YOLO (CNN-Transformer for real-time detection) and models like PRNet, Hybrid-COD and ATDMNet.

Multi-modal learning fuses diverse data sources, such as hyperspectral imaging and vision-language models (LVLMs and LLaVA), to penetrate camouflage. Domain-specific models are specialised models like Insect-LLaVA, which can be used for insect understanding and CamObj-LLaVA, which can be used for camouflaged scenes.

Key datasets and benchmarks for camouflaged insect/ pest detection

Developing robust camouflaged pest detection requires diverse, high-quality datasets. Benchmarks like CAMO, COD10K and NC4K provide standard RGB foundations, while IP-102 and BeetPest offer agriculture-specific annotations. Comprehensive resources include Insect-1M (hierarchical taxonomy) and MM-CamObj (Ruan *et al.,* 2025) with image-text pairs. Furthermore, UBCODD supports large-scale fine-tuning and MUDCAD-X expands capabilities beyond RGB into multispectral imaging. Together, these datasets enable the training and evaluation of specialized models capable of overcoming complex, real-world camouflage challenges.

Synthetic data for camouflaged pests

Camouflaged pests make it visually difficult to detect them. To address this challenge for computer vision models, we need more labelled data to train them. The ethical and practical limitations associated with acquiring large, labelled datasets of real-world, highly concealed specimens are difficult. Existing synthetic data strategies include the widely adopted cut-paste and blending techniques (*e.g.*, Generate-Paste-Blend-Detect (Giakoumoglou *et al.*, 2023) and Cut, Paste and Learn (Dwibedi *et al.*, 2017). These strategies can successfully increase the volume and visual diversity of training data, but they introduce sharp unnatural edges that the neural networks use as “Shortcut Learning” (Geirhos *et al.*, 2020). These methods primarily function as open-loop spatial augmentations. They rely heavily on pre-existing libraries of object cutouts. They lack integrated domain-specific intelligence regarding pest ecology, preferred habitat and explicit camouflage strategies. More importantly, they lack a mechanism to systematically diagnose and target model-specific failure modes.

AI agents: Foundations and role in computer vision

AI agents operate autonomously by perceiving environments, reasoning and planning to achieve specific goals. Advances in generative AI and foundation models empower these agents with multimodal processing. In computer vision, they integrate vision-language models with LLMs, enabling autonomous, goal-oriented tasks-such as synthetic data generation-by combining visual detection with natural language reasoning and contextual memory.

The above-mentioned gaps are addressed in our study by introducing a Novel Closed-Loop Data Generation (CLDG) paradigm. This paradigm creates ecologically relevant, high-difficulty training samples adaptively. We used a multi-agent system, commencing with a Large Language Model (LLM) based Knowledge Agent. This agent is tasked with extracting and formalising domain knowledge regarding pest-host interactions and camouflage mechanisms. This foundational intelligence drives a subsequent Generative Agent to perform targeted image synthesis. The agent generates hard-negative scenarios (undetectable camouflaged pests) with high domain specificity. The generated data is then evaluated by a detection model. The central innovation is the Curation Agent, which systematically monitors the model’s performance on the synthetic data. The images with pests not detected by the models are isolated and they form the hard examples. These hard examples are prioritised for retraining and also inform the LLM-based Knowledge Agent to refine and intensify the generation rules for the next iteration. This adaptive, closed-loop feedback mechanism ensures that the model is continuously challenged by its own historical weaknesses. It hence leads to a robust, systematic and exponentially improving detection capability. This approach significantly mitigates the sim-to-real gap for extreme camouflage.

We have also given importance to the explainability of the models in order to verify our methodology. We used Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju *et al.*, 2017) to verify that our iterative blending has shifted the models focus from edge noise to the biological anatomy of the pests.

MATERIALS AND METHODS

The research was conducted at the Department of CSE, Anurag University, Hyderabad, during the period November 2025-June 2026, where all activities related to synthetic data generation, model training and performance evaluation took place.

Proposed AI-driven synthetic data generation flow

To concretely address this data scarcity and accelerate the development of more robust camouflaged insect detection models, we propose a novel, AI-driven methodology for synthetic camouflaged insect image generation. This approach leverages the capabilities of advanced AI agents throughout the process, creating a self-improving cycle for dataset enrichment.

The core idea described in this paper is to take images of pests where the pest-detection models can detect the pests. New images will be generated based on these detected pest images to add the effect of camouflage. Various images are generated where the model cannot detect camouflaged pests, which were detected earlier without the camouflage effect. These generated images can then be added to the training data to improve the model. This is depicted in the Fig 1.

The methodology is structured into seven interconnected steps under three AI agents, designed to systematically generate diverse and challenging camouflaged insect images. Fig 2 provides a conceptual illustration of this iterative process, showing the flow of information.

Step-1: Knowledge extraction

LLMs are prompted to compile comprehensive lists of insect pests affecting various crops. For each pest, they extract crucial ecological and morphological details, including.

- Specific plant parts where they are typically found (*e.g.*, underside of leaves, stems, flowers, fruits, roots).

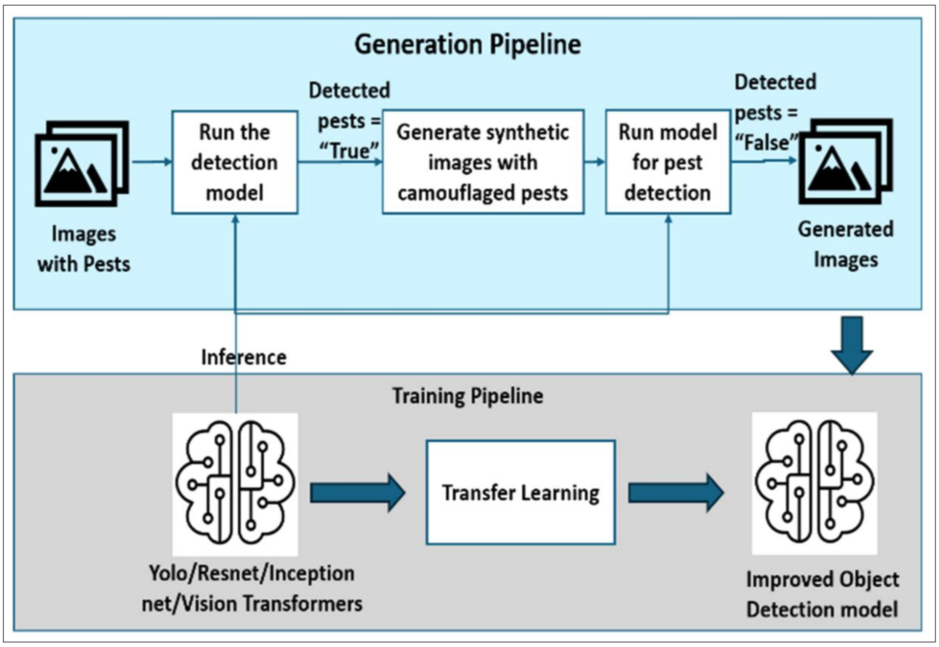

Fig 1: Core idea of synthetic image generation.

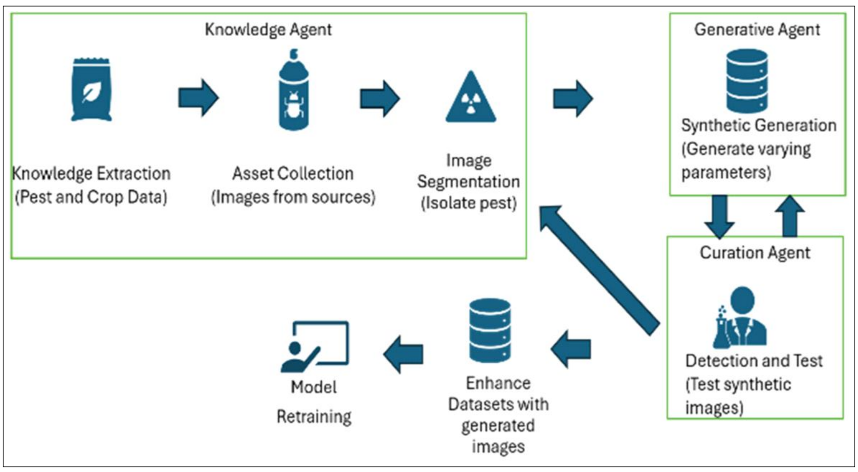

Fig 2: AI Agents orchestration and the steps involved.

- Their typical size range, colour variations and texture characteristics.
- Known camouflage strategies (*e.g*., background matching, disruptive coloration, masquerade, transparency).
- Life cycle stages and associated visual changes.
- Any other important details that can help generate realistic synthetic camouflaged images, such as preferred lighting conditions or common environmental contexts.

A sub-agent under a knowledge agent is used to implement this step. The information about a given pest is generated with the help of an LLM. The LLM that is being used here is Gemini 1.5 Flash. Zero-shot prompting is used in this case. The prompt to LLM requests for information of the pest like life cycle, crops affected, ecological details, morpohological details. The LLM generates the output in JSON format mapping pests to crops affected, their preferred micro-habitats on plants and detailed camouflage-relevant attributes. The knowledge information about the pests retrieved by invoking the LLMs are later used in the image generation steps.

Step-2: Asset collection for pests and plants

With the information generated in Step 1 about the pests, images of the pests and their associated host plants are collected. This includes:

- Images of the pest species in various orientations, lighting conditions and life stages.
- Images of the specific plant parts (leaves, stems, flowers, fruits) that serve as common backgrounds, captured under different lighting, angles and growth stages to ensure environmental variability.

This step is implemented as a tool under the knowledge agent which extracts information needed to get unique pest and crop/plant images from the internet:

- The information extracted for pests include life_stages and host_plants.
- For the plant, information about the parts of the plant where the pests are found are extracted from the JSON.

Once the above information is extracted it creates multiple queries. For each of the queries, it scrapes image URLs from Google Images. The images from these URLs are then extracted to a local path.

Step-3: Image separation (foreground/background segmentation)

For each collected pest image, automatically segment the insect from its original background, generating a clean foreground mask. Similarly, clean background elements (*e.g.,* individual leaves, sections of stems) are extracted from the plant images. This step is critical to ensure precise placement and realistic blending in the synthetic images. This step is implemented as a tool under knowledge agent, that processes each image with the help of a library rembg to remove the background and only keep the pest in the image (Bappy, 2024).

The output of this step is isolated pest foregrounds (with corresponding pixel-level masks).

Step-4: Synthetic image generation with varied parameters

This is the core step that generates images with varying parameters. The isolated pest foregrounds are blended onto the plant backgrounds. For each pest-plant pair, several camouflaged images are generated by systematically varying key parameters like position of pests on plant background, orientation of the pest, scale, blending technique based on pest camouflage strategies, environmental conditions like lighting, occlusion, background noise.

The logic in this step to generate synthetic images using the extracted pest and background from previous steps is implemented under generation agent. We have used Stable Diffusion pipeline that takes instructions and images as inputs to generate the desired synthetic output image. Annotations/labels are automatically generated in yolo format for the generated synthetic images.

The output is a large, diverse dataset of synthetically generated camouflaged insect images, each accompanied by precise bounding box annotations in yolo format.

Step-5: Initial detection and classification by existing models

The newly generated synthetic images are fed through a suite of current detection models. Each image is then classified as ‘detected’ or ‘not-detected’ based on predefined performance thresholds (*e.g*., Intersection over Union (IoU) for bounding boxes/masks, or confidence scores).

The output is a categorized list of synthetic images (detected/not-detected) along with their detection metrics and confidence scores from the existing models.

Step-6: Dataset enhancement with challenging samples

The images that were classified as ‘not-detected’ by the existing models are identified as particularly challenging examples. These images can be added to the existing datasets to enhance them. This step helps in improving the dataset by adding real camouflaged pest images, which are most valuable for improving model robustness.

This step generates an enhanced dataset of camouflaged pest images. This new dataset includes images enriched with difficult-to-detect instances.

Step-7: Training new models with enhanced datasets

The enhanced dataset, which is the output of Step 6 is used to train new camouflaged pest detection models. This training can be done from scratch, or an existing one can be fine-tuned. These enhanced images ensure that subsequent models learn directly from the weaknesses of their predecessors. It will lead to continuous improvement in detection performance and generalisation capabilities. This step produces improved camouflaged pest detection models with better performance on challenging, camouflaged pests.

This structured approach, driven by AI agents at each stage, creates a self-improving cycle for generating high-quality, challenging camouflaged insect data. It directly addresses the data scarcity problem by providing an inexhaustible and intelligently curated source of training data, thereby accelerating the development of more robust and accurate camouflaged insect detection systems for real-world agricultural applications.

RESULTS AND DISCUSSION

The above methodology is used with the pest "weevil" on mango leaves to prepare synthetic images. YOLOv11s model is taken as the base and trained with the synthetically generated camouflaged images.

Comparison of the detections and performance

Fig 3 shows the output of various steps in the methodology for creating the synthetic images. From the images, it is evident that the camouflaged Weevil is better detected after transfer learning with the synthetic generated images.

Fig 4 shows performance comparison of the YOLOv11s original with the Transfer-Learned model (freezing the first 20 layers of the network) using synthetic camouflaged images. As shown in Table and the chart in Fig 4, the transfer-learned YOLOv11s model demonstrates significant improvements across key metrics compared to the baseline YOLOv11s model. The most notable gains are in mAP50, which improved from 0.389 to 0.666 and in Recall (R), which jumped from 0.0606 to 0.54. These results indicate that the synthetic data generated using the novel method greatly enhanced the detection capabilities for the specific task of identifying camouflaged "weevil."

Ablation study: Contribution of the stages

To conduct this ablation study, we have picked the pest as hornworm and the background crop as tomato.

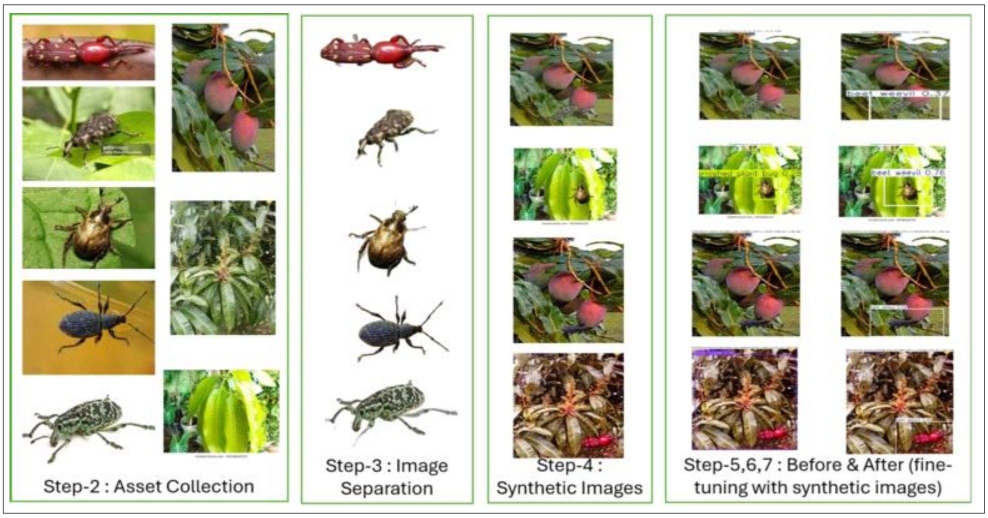

Fig 3: Steps in the methodology.

Model	Box(P)	R	mAP50	mAP50-95
YOLOv11s	0.706	0.0606	0.389	0.228
YOLO11s (Transfer-Learned)	0.883	0.54	0.666	0.351

Fig 4: Performance comparison before and after transfer-learning using synthetic camouflaged images.

There are various stages in the synthetic data generation pipeline. To evaluate the contribution of the important stages like web scraping, cut-paste, iterative blending, we performed an ablation study in a systematic way. We evaluated four different models that are generated with the images acquired from these stages. These models range from a baseline model trained on web scraped images to our final model trained through iterative adversarial blending.

Stage performance benchmarking

We prepared a test dataset that has mix of real and synthetic images with pests displaying camouflage in some of them. In order to evaluate the models generated using images from various stages, we are using this unseen test images. We captured the metrics including mAP50, mAP50-95 and Miss Rate and summarized in Table 1.

Model 0 is built by transfer learning YOLOv11s model using web scraped images of hornworm on tomato plants. This model clearly has the scarcity issues as finding camouflaged hornworm on web is difficult. There is a domain gap as the images obtained from web scraping do not represent all the actual cases in the farm. These issues resulted in a model with prohibitive 83% miss rate.

The scarcity issue is addressed in Model 1 with the Cut-Paste method. This stage creates huge number of hornworm pests on tomato plants, but the generated images have unnatural edges which makes the model learn using shortcut learning (Geirhos *et al.*, 2020). The Model 1, provided a significant improvement in the mAP50 to 0.710. The miss rate has also reduced to 29%.

With our proposed Iterative Blending (Model 2), it can be seen that the model has improved its mAP50 to 0.920 and the miss rate has drastically reduced to a 8%, providing the most critical improvement. Our final adversarial refinement (Model 3) clocked a mAP50 of 0.931 and a miss rate of 7%. This improvement is the result of learning the pest anatomy with texture based detection instead of edge based as in cut-paste stage.

PR curve analysis

The Precision-Recall curves for the web scraping, cut-paste and iterative blending are shown in Fig 5. These curves give clear indication of the pest detection precision at varying recall values. The model from web scraped images showed drop in precision even at a low recall value. The cut-paste model maintained high precision at low recall, but declined at higher recall which indicated failure to detect camouflaged or low-contrast target pests. The models from iterative blending, Models 2 and 3 maintained high precision even at high recall values increasing the area under the curve value, demonstrating that the model has learned to maintain high precision even when detecting difficult to detect camouflaged pests.

Explainability *via* Grad-CAM matrix

To bring in the explainability dimension for the pest detection models, we have used Gradient-weighted Class Activation

Table 1: Quantitative ablation results across four training iterations.

Model iteration	Training data source	mAP50	mAP50:95	Miss rate
Model 0	Web-scraped (Baseline)	0.170	0.070	83%
Model 1	Cut-paste	0.710	0.511	29%
Model 2	Blending iteration 1	0.920	0.717	8%
Model 3	Blending iteration 2	0.931	0.729	7%

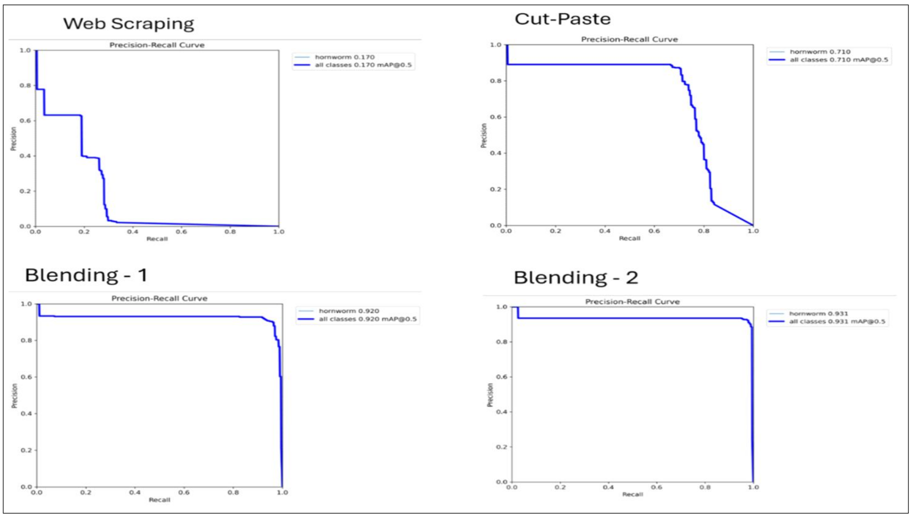

Fig 5: Precision recall curve analysis

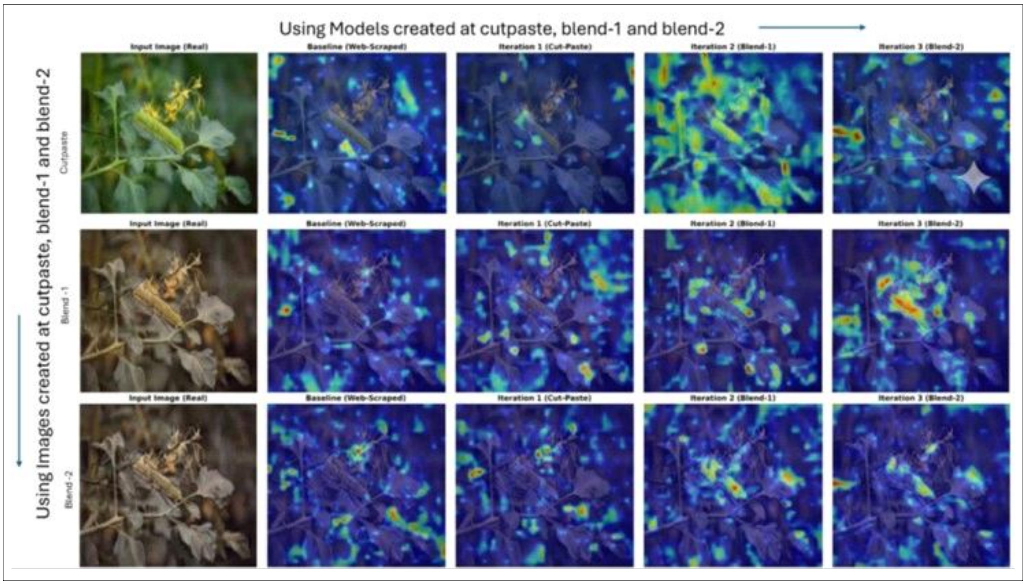

Fig 6: GRADCAM matrix - stage image complexity vs stage model iterations.

Mapping (Grad-CAM) (Selvaraju *et al.* (2017)). We created a matrix/table (Fig 6) with columns showing the four model iterations(web scraped, cut-paste, blend-1, blend-2) against three levels of image complexity: raw Cut-Paste (no camouflage), moderate Blending-1 (camouflaged) and advanced Blending-2 (more camouflage) .

Shortcut learning in cut-paste model

The model trained on the cut-paste images (Model 1) showed a critical vulnerability. The Grad-CAM hotspots show focus on the synthetic boundary artifacts or the cut edges. This confirms the model was using the shortcut learning which relies on the artificial edges instead of the biological features. The models generated from iterative blended images (Models 2 and 3), show a more holistic attention map on the pest. This indicates that our blending pipeline has forced the model to ignore the artificial edges and regularized the model.

Texture-based detection in blending models

In moderate camouflage pests, Models 0 and 1 exhibited activation scattered across background leaf structures. Model 2, has shown a marked improvement where the background noise is suppressed and activation hotspot is created over the length of the hornworm. This proves that the focus has shifted towards the pest texture and away from the artificial edges.

Deep camouflage (Row 3: Blend-2)

The third row shows highest camouflage, where the pest is camouflaged seamlessly into the background. For this test:

- **Model 1** has failed as the edge shortcuts are not present in the camouflaged test image.
- **Model 2** is able to detect the target but shows attention on plant stems as well.
- **Model 3 (Adversarial)** achieves the most localized and accurate activation. By training on the failures of the previous model, Model 3 learned internal biological patterns that distinguish the pest from its host, even in the absence of edges or boundary contrast.

The findings in the ablation experiments have shown a clear hierarchy of learning in the models. The Cut-Paste stage provided the detection based on the edges while the Iterative Blending stages provided the pest understanding. Our adversarial loop helped the model learn internal anatomy by denying the shortcut learning, resulting in a model that is resilient to real-world camouflage and varied lighting environments.

CONCLUSION

The domain of camouflaged insect and pest detection through computer vision and multi-modal transformers represents a critical frontier in precision agriculture. Significant improvements have been made transitioning from traditional image processing to deep learning and more recently to hybrid CNN-Transformer architectures and multi-modal LVLMs. In spite of these improvements, the fundamental challenge of data scarcity persists.

To address this data bottleneck, this paper proposed a novel, AI-driven methodology for synthetic data generation. We are using multiple agents to perform the tasks autonomously. These agents leverage LLMs for information extraction, automate asset collection, perform advanced segmentation and iterative synthetic image generation and evaluation. This approach creates a self-improving cycle for dataset enrichment. It addressed the limitations of manual annotation and accelerated the training of more robust detection models.

Building fully autonomous solutions for pest detection and making them universally deployable need more future efforts. Unexplored opportunities beyond data generation include computationally efficient edge deployments, explainable AI implementation, potential of unsupervised and weakly supervised learning, deeper vision-language integrations and leveraging advanced sensing modalities beyond the visible spectrum.

ACKNOWLEDGEMENT

Not applicable.

Disclaimers

The views and conclusions expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated institutions. The authors are responsible for the accuracy and completeness of the information provided, but do not accept any liability for any direct or indirect losses resulting from the use of this content.

Informed consent

Not applicable for this study.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this article. No funding or sponsorship influenced the design of the study, data collection, analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Bappy, M.A. (2024). AI-powered background removal for enhancing image processing in digital applications. *Research Gate*. doi: 10.13140/RG.2.2.10008.81927.

Dalal, M. and Mittal, P. (2025). A systematic review of deep learning-based object detection in agriculture: Methods, challenges and future directions. *Computers, Materials and Continua*. **84(1):** 57-91.

Dwibedi, D., Misra, I. and Hebert, M. (2017). Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV): 1310-1319.

Gao, A., Dong, Y., Liu, D., Li, A., Lin, Z. and Li, Y. (2025). SSHFormer: Optimizing spectral reconstruction with a spatial-spectral hybrid transformer. *Remote Sensing*. **17(9):** 1585.

Gao, S., Feng, Y., Wang, Q., Hong, L., Zhou, X., Fei, L., Wang, Y. and Zhang, W. (2025). MSVCOD: A large-scale multi-scene dataset for video camouflage object detection. *arXiv Preprint arXiv:* 2502.13859.

Geirhos, R., Jacobsen, J.H., Michaelis, C. *et al.* (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*. **2:** 665-673.

Giakoumoglou, N., Pechlivani, E.M. and Tzovaras, D. (2023). Generate-Paste-Blend-Detect: Synthetic dataset for object detection in the agriculture domain. *Smart Agricultural Technology*. **5:** 100258.

Hao, C., Yu, Z., Liu, X., Xu, J., Yue, H. and Yang, J. (2024). A simple yet effective network based on vision transformer for camouflaged object and salient object detection. *ArXiv preprint arXiv*: 2402.18922.

Hupel, M. and Stütz, P. (2025). Optimized spectral indices for camouflage detection in multispectral imagery. *Journal of Applied Remote Sensing*. **62(1):** doi: 10.1080/15481603.2025.2508574.

Kappali, R.H., Sadyojatha, K.M. and Prashanthi, S.K. (2024). Computer vision and machine learning in paddy diseases identification and classification: A review. *Indian Journal of Agricultural Research*. **58(2):** 183-187. doi: 10.18805/IJARe.A-6061.

Mehta, R.A., Kumar, P., Prem, G., Aggarwal, S. and Kumar, R. (2025). Leveraging artificial intelligence for disease diagnosis in agricultural crops: A review. *Indian Journal of Agricultural Research*. **59(5):** 681-690. doi: 10.18805/IJARe.A-6363.

Mondal, A. (2020). Camouflaged object detection and tracking: A survey. *International Journal of Image and Graphics*. **20(4):** 2050028.

Ruan, J., Yuan, W., Lin, Z., Liao, N., Li, Z., Xiong, F., Liu, T. and Fu, Y. (2025). MM-CamObj: A Comprehensive Multimodal Dataset for Camouflaged Object Scenarios. AAAI Publications.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017). Grad-CAM: visual explanations from deep networks *via* gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV): pp 618-626.

Sharma, A. and Kumar, A. (2026). Dual-encoder variational auto encoder for detection and classification of plant leaf diseases. *Indian Journal of Agricultural Research*. **60(6):** 912-918. doi: 10.18805/IJARe.A-6451.

Shoaib, M., Sadeghi-Niaraki, A., Ali, F., Hussain, I. and Khalid, S. (2025). Leveraging deep learning for plant disease and pest detection: A comprehensive review and future directions. *Frontiers in Plant Science*. **16:** 1538163.

Tang, K., Qian, Y., Dong, H., Huang, Y., Lu, Y., Tuerxun, P. and Li, Q. (2025). SP-YOLO: A Real-Time and efficient multi-scale model for pest detection in sugar beet fields. *Insects*. **16(1):** 102.

Tang, L., Jiang, P.T., Shen, Z.H., Zhang, H., Chen, J.W. and Li, B. (2024). Chain of Visual Perception: Harnessing Multimodal Large Language Models for Zero-shot Camouflaged Object Detection. Proceedings of the 32nd ACM International Conference on Multimedia: pp 8805-8814.

Truong, T.D., Hoang-Quan, N., Xuan-Bac, N., Ashley, D., Xin, L., Khoa, L. (2025). Insect-foundation: A foundation model and large multimodal dataset for vision-language insect understanding. *arXiv preprint arXiv:* 2502.09906.

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

Yin, B., Zhang, X., Hou, Q., Sun, B., Fan, D. and Gool, L.V. (2022). CamoFormer: Masked separable attention for camouflaged object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **46:** 10362-10374.