Ecological Informatics 95 (2026) 103701

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

InsectRoleVision: A computer vision system to study arthropod diversity and functional roles

Song-Quan Ong [a,b,*], Min-Hui Lim [b], Kim Bjerge [c], Francisco Javier Peris-Felipo [d], Rob Lind [e], Toke Thomas Høye [a,*]

[a] *Department of Ecoscience, Aarhus University, C. F. Møllers Allé 8, DK-8000 Aarhus, Denmark*
[b] *Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia*
[c] *Department of Electrical and Computer Engineering, Aarhus University, Finlandsgade 22, DK-8000 Aarhus, Denmark*
[d] *Syngenta Crop Protection, Rosentalstrasse 67, 4058 Basel, Switzerland*
[e] *Syngenta International Research Centre, Jealott's Hill, Berkshire RG42 6EY, UK*

ARTICLE INFO

Keywords:
SSD
Integrated pest management
ScannerVision
Flatbug

ABSTRACT

Background: Arthropods make up the majority of species on Earth. To study their diversity and ecological roles in ecosystems, bulk sampling is commonly used to collect large numbers of specimens. Processing these samples is labor-intensive and time-consuming, often delaying timely decision-making in ecosystem monitoring. Automated detection systems offer a promising alternative to support sample processing; however, most existing systems still have two major limitations. First, the pipelines for localizing and classifying arthropods in images, whether single- or double-stage, are limited. Second, the prediction results often lack information about functional roles. Therefore, we developed *InsectRoleVision*, a more expert-centered and reliable system that enables inference of arthropod diversity based on their functional roles.
Methodology: To develop the system, an image dataset with taxonomic resolution was designed and created to support conclusions about the functional roles of the animals. Both single-stage and double-stage recognition pipelines were compared. For single-stage detection and the first stage of double-stage detection, four YOLO models and a transformer were evaluated to localize and classify the arthropods in each image. In double-stage detection, the region of interest (RoI) was cropped into individual images after localization and used to compare four classification models: InceptionV3, ResNet, MobileNet, and VGG19. A logic block pipeline was connected to the prediction results to further infer the richness and proportionality of each class or taxon with respect to their functional roles.
Result: YOLOv11 was the best-performing model, achieving over 93% mAP, precision, and recall in localizing arthropods in the images. InceptionV3 was the best-performing classifier, achieving 80% precision and recall in classifying more than 43,000 cropped images of arthropods. There was no significant difference between the results of single- and modular double-stage detection strategies. Therefore, the choice between strategies depends on the intended application: single-stage detection provides real-time results and is suitable for real-time detection applications, while double-stage detection allows a human expert to review the detection proposal and refine the classification result. *InsectRoleVision* has adopted the YOLOv11-InceptionV3 architecture, which is more flexible and human-centered, allowing quick access to both arthropod diversity and ecological roles.

1. Introduction

Arthropods constitute the vast majority of known living species on Earth and exhibit remarkable biodiversity, with a wide range of forms and adaptations (Chakravarthy et al., 2016). This group includes insects, crustaceans, arachnids, centipedes, and millipedes, which inhabit nearly every ecosystem on the planet. They play a crucial role in maintaining ecosystem functions and provide numerous benefits to humans. For example, many arthropods pollinate plants (Wardhaugh, 2015), act as natural pest controllers (Sunderland and Samu, 2000), serve as an important food source in many food webs (Gratton and Denno, 2006), and function as scavengers or decomposers, breaking down dead organic

* Corresponding authors at: Department of Ecoscience, Aarhus University, C. F. Møllers Allé 8, DK-8000 Aarhus, Denmark.
E-mail addresses: songguan26@gmail.com, songquan.ong@ums.edu.my (S.-Q. Ong), tth@ecos.au.dk (T.T. Høye).

https://doi.org/10.1016/j.ecoinf.2026.103701
Received 17 November 2025; Received in revised form 8 March 2026; Accepted 8 March 2026
Available online 11 April 2026

material and returning nutrients to the soil (Bagyaraj et al., 2016). However, certain insects are highly invasive, pose a significant threat to global biodiversity, and require strict management measures (Baker et al., 2005). Others directly damage crops and stored products (Ileke et al., 2023) or act as vectors and intermediate hosts for disease-causing organisms (Ong et al., 2024).

To study ecological diversity and the role of insects, various mass sampling methods are used, often involving the collection of large numbers of samples. For example, more than one hundred insect samples can be collected in a single night using light traps for insect monitoring (Kammar et al., 2020), and hundreds or thousands of mosquitoes and other insects can often be collected in a single mosquito trap, such as the Centers for Disease Control and Prevention (CDC) light trap (Ong et al., 2024). These samples often require trained personnel and considerable time to identify the insects and count their numbers to obtain biodiversity data. As human accuracy in insect identification decreases with increased working time and sample numbers (Ratnieks et al., 2016), this limitation may prevent timely and accurate decision-making that helps reduce economic losses, especially in agroecosystems where insect pest numbers exceed the economic control threshold. For medically relevant arthropods, rapid population assessment can provide important insights for early warning systems to reduce vector-borne disease transmission (Pergantas et al., 2021).

To support the sampling process, a computer vision system with a deep learning model offers an excellent alternative for the automatic identification of arthropods. The system typically uses images as input data and employs deep learning models, particularly convolutional neural networks (CNNs), which have demonstrated strong capabilities in extracting features from images, or attention-based mechanisms of transformers to recognize arthropods in images. For example, Wang et al. (2020) developed CPAFNet to classify 20 common insects with an accuracy of 92.63% by comparing different deep neural networks. Ong and Hamid (2022) compared four deep learning models for classifying five different dipteran species and found that each taxonomic level had its own optimal model, suggesting that no single model is universally applicable. Bjerge et al. (2024) demonstrated a pipeline that processes the results of You Only Look Once (YOLO) version 5 recognition into individual image frames and compared six deep learning models for classifying 17 insect groups on flowers in the field.

Apart from the choice of architecture, most studies have focused on using a single-stage detection strategy to classify or recognize insects. In this approach, the input image data is fed into the architecture once, and the result of localization and/or classification is obtained (Abdul-Khalil et al., 2023). However, another strategy for recognizing insects in an image is the use of double-stage detection models (Huang et al., 2022). In arthropod detection, the choice of strategy still largely depends on the application's goal, and in the study of ecological diversity and the role of arthropods, there is no direct comparison to determine which strategy is more effective.

In addition, most studies on the use of computer vision and deep learning in arthropod diversity research have focused on species detection and species richness, often lacking information about the functional roles of arthropods. Information on the functional roles of arthropods is important for managing many ecosystems (Wong et al., 2019). For example, in agroecosystems, the arthropods identified on sticky card images used to inform pest control strategies need to have their roles clarified (Holthouse et al., 2021). When pollinators and natural enemies are highly abundant, no pest control measures should be taken, whereas when pests (herbivores with sucking mouthparts such as aphids) are more abundant, measures should be implemented to reduce economic loss.

In this study, we introduced *InsectRoleVision*, a computer vision system designed to measure the abundance and roles of arthropods in large samples. The system comprises three components: localization models, which locate the target arthropod in the image; classification models, which assign the region of interest to label classes; and a logic function that translates the output into ecological diversity (species richness and composition) and their roles. Specifically, our objectives were: (1) to compare single-stage and double-stage detection strategies for recognizing arthropods and obtaining data on their ecological diversity and roles; (2) to evaluate the localization model used in the double-stage strategy and the detection models used in the single-stage strategy; and (3) to evaluate the classification model used in the double-stage strategy. Ultimately, this study aims to develop a scalable biodiversity assessment tool that enhances our capacity for ecological monitoring and contributes to biodiversity conservation.

2. Methodology

Fig. 1 shows the complete development pipeline of *InsectRoleVision* and the validation of the system using two external data sets. The individual steps of the process are explained in more detail in Subsections 2.1, 2.2, and 2.3.

2.1. Dataset construction

Arthropods were collected using sticky cards, as described in *ScannerVision* (Ong et al., 2025). In brief, 80 sticky traps were placed evenly across five wildflower bed sites on the campus of Aarhus University in Aarhus, Denmark, covering approximately 20 ha. Sampling was conducted at eight sampling periods covering two consecutive years, 2023 and 2024, during summer and early autumn (July to September). The total of 640 sticky traps were scanned following the protocol described by Ong et al. (2025), using a high-resolution scanner (Epson V850 Pro) at 1200 dpi (10,200 $\times$ 14,039 pixels) with high-contrast settings.

To obtain information on the ecological diversity and roles of arthropods, the data were annotated according to their taxonomic resolution, which provides such information. For example, the order Hemiptera was further subdivided into aphids (Aphidomorpha), leafhoppers (Membracoidea), and true bugs (Heteroptera) based on their morphology and ecological roles. Here, “ecological roles” refers to the main functional roles of arthropod groups within wildflower communities, such as herbivory, pollination, predation, or sap-feeding. For instance, aphids (Aphidomorpha) are primarily phloem-feeding herbivores that influence plant growth and are widely recognized as economically important herbivores in agricultural systems; leafhoppers (Membracoidea) are mainly sap-feeding herbivores with distinct host-plant associations; and many true bugs (Heteroptera) act as predators or omnivores, contributing to biological control. We excluded further subdivision of the large diversity of Coleoptera and Formicidae because the number of samples is small and further sorting would make the class size too small to be trained and tested. Taxonomic annotations were performed by two taxonomists using the Dataloop (https://dataloop.ai/) and CVAT (https://www.cvat.ai/) platforms to draw the Region of Interest (RoI) bounding boxes. All annotations were performed directly at the multi-class taxonomic level, not through a hierarchical or multi-stage process.

2.2. InsectRoleVision development

Fig. 2 presents the process of comparing two strategies – single-stage and modular double-stage detection – to determine which is more effective for *InsectRoleVision*.

To obtain the best-performing deep learning models for *InsectRoleVision*, single-stage and double-stage strategies were compared. In the single-stage approach, four versions of YOLO and one transformer model were evaluated. The models were trained to predict 12 classes, each annotated with their taxonomic name and ecological role. These models represent the evolution of a widely adopted real-time object detection family and are frequently used in ecological and field-based computer vision applications.

On the other hand, in the double-stage strategy, the four YOLO

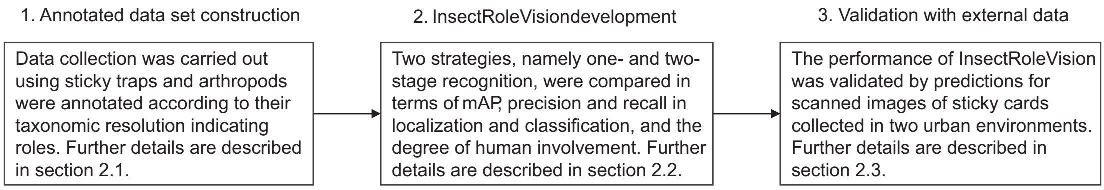

Fig. 1. Workflow of developing and validating the *InsectRoleVision* system.

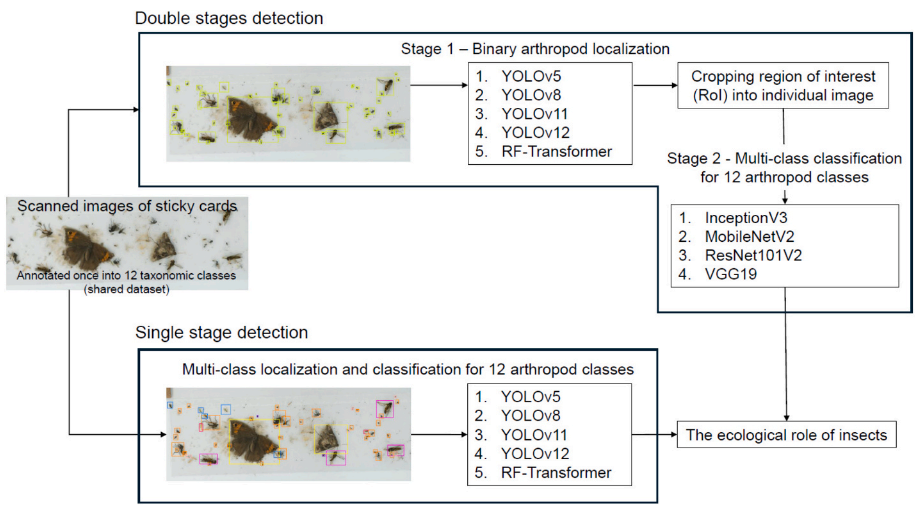

Fig. 2. Overview of the *InsectRoleVision* development and evaluation workflow. In the single-stage detection strategy, arthropods were annotated directly into 12 taxonomic classes, and multiple detection models were trained and compared. In the modular double-stage strategy, the same multi-class annotated dataset was used, but the modeling pipeline was divided into two inference stages: a first-stage binary arthropod localization model and a second-stage classification model that assigns detected regions of interest to one of the 12 predefined classes. This design separates localization and fine-grained classification while maintaining a consistent annotation procedure.

versions and the transformer-based models were trained to perform binary arthropod localization during the modeling stage, utilizing the detection models' localization capabilities. Arthropods were annotated directly into 12 taxonomic classes during dataset preparation, and the binary "arthropod" category was introduced only at the inference stage to separate localization from fine-grained classification. The bounding boxes generated in the first stage were cropped, then subjected to human verification and second-stage classification model development. For the classification model comparison, four state-of-the-art classifiers – ResNet, MobileNet, InceptionV3, and VGG19—were evaluated for their performance in classifying arthropods into 12 classes. The models were selected based on previous studies (Ong et al., 2022, 2024; Ong et al., 2025), which identified models that offered a good balance between performance, model size, and inference time (https://keras.io/api/applications/).

For model development, the environment was set up by installing the Ultralytics package (YOLOv5, v8, v11, and v12) and the CUDA dependencies, using a Tesla A100 GPU with Python 3.8 on the Google Colab platform. For the object detection model used in both single-stage recognition and the first step of the double-stage recognition model, the annotated dataset, prepared in COCO JSON format, was split into 70% for training, 20% for validation, and 10% for testing. After the data split, all images were preprocessed by resizing them to a resolution of 3200 × 3200 pixels and augmented using rotation, flipping, shearing, and adjustments to hue, saturation, brightness, exposure, and blur to increase variability and reduce overfitting. For the second-stage classification in the double-stage recognition framework, datasets were constructed by extracting regions of interest (RoIs) from the original images using the bounding box annotations in the JSON files. The cropped RoIs were then organized into class-specific directories. Due to class imbalance, with Diptera (Brachycera) and Diptera (Nematocera) overrepresented compared to other classes, random oversampling was applied to the training data to balance class distributions. Oversampling was performed only on the training set to prevent information leakage into the validation and test sets and to minimize the risk of artificially inflated performance. This strategy aims to improve classifier learning for underrepresented classes while maintaining unbiased evaluation. For hyperparameter optimization, an auto-optimizer was used to train the model, with hyperparameters set to a learning rate of 0.000325 and a momentum of 0.9, trained for 50 epochs, with validation-based performance monitoring to reduce the risk of overfitting. Object detection performance was evaluated on the held-out test set using mean average precision (mAP), precision, and recall, while classification performance was assessed using precision and recognition accuracy. The evaluation

Table 1
Definition and formula to calculate the evaluation matrices of object detection in this study.

Matrices	Evaluation focus	Formula
Precision	The proportion of all positive classifications of the model that are actually positive.	$\frac{Tp}{Tp + Fp}$
Recall	The proportion of all actual positive findings that were correctly classified as positive findings.	$\frac{Tp}{Tp + Fn}$
Mean Average precision (AP)	Average precision is calculated as the weighted mean of the precision at each threshold; the weight is the increase in recall over the previous threshold. Mean average precision (mAP) is the average of the AP of each class.	$\frac{1}{N}\sum_{k=1}^{k=n} AP_k$

TP – true positive; TN – true negative; FP – false positive; FN – false negative; p – precision; r – recall.

metrics are summarized in Table 1. To compare performance differences among model architectures and detection strategies, these metrics were statistically analyzed using one-way analysis of variance (ANOVA) with a significance threshold of $p < 0.05$. All models were trained and evaluated using identical datasets, data splits, preprocessing pipelines, and evaluation protocols to ensure a controlled experimental setting. In this design, model architecture and detection strategy were treated as the primary sources of variation, minimizing potential confounding effects from differences in data preparation or evaluation conditions. Prior to ANOVA, statistical assumptions were assessed. Normality of the performance metric distributions was evaluated using the Shapiro–Wilk test, and homogeneity of variances was examined using Levene's test. When ANOVA indicated significant differences among groups, Tukey's honestly significant difference (HSD) post hoc tests were conducted for pairwise comparisons between models. To control inflated Type I error due to multiple comparisons across performance metrics and model configurations, appropriate multiple comparison adjustments were applied. All statistical analyses were conducted to support comparative performance evaluation of the proposed framework rather than causal inference regarding ecological processes. Results are reported with corresponding test statistics and significance values.

2.3. Ecological diversity and roles inference from external datasets

To validate the *InsectRoleVision* system, a separate experiment was conducted to compare the number of individuals and species across different urban environments. The first environment is a warehouse where grain and flour are stored, and the second is a grocery store. UV light traps were deployed simultaneously in both environments from February to March 2025. Sticky cards were replaced every seven days, with each replacement considered an independent sampling interval under standardized exposure conditions (Lim et al., 2025). Traps were installed simultaneously in both environments and left in place for a standardized exposure period of seven days to ensure comparable sampling effort. After retrieval, each sticky card was scanned once using the ScannerVision protocol (Ong et al., 2025), and the resulting images were saved in JPEG format for subsequent system evaluation. In each environment, ten light traps were deployed, and the scanned images were used as input data for validating the *InsectRoleVision* system.

To deploy the system, the Roboflow platform and its workflow were used, which contain a series of function blocks to process the input image and output the prediction as a number (species richness) and the number of classes (species diversity) for an image. Fig. 3 shows the blocks used in Roboflow's workflow. The first block in the workflow is the input block, where the image is imported into the system. This block is connected to a block that uses the developed YOLOv11 to predict the input image. Since the prediction results in the localization of bounding boxes and labels, the subsequent model includes a block that allows the user to visualize the positions of the boxes and labels. Another block is used to aggregate the data into frequencies or counts for each class. The model is used as one of the function blocks in a pipeline. The next function, called "property definition," retrieves the prediction result according to the class, and "data aggregator" calculates the number of individuals per class. To determine the functional role of arthropods, classes with similar functional roles, such as arachnids and Apocrita, both of which serve as natural enemies in agricultural or urban ecosystems, were combined with the number of individuals, and the result was expressed as a proportion according to their functional role.

3. Result

3.1. Annotated dataset

The classification of arthropods followed Krenn (2019), Krenn and Aspöck (2012), and Tsiafouli et al. (2017), who categorized them by morphology as well as ecological and economic functions, with

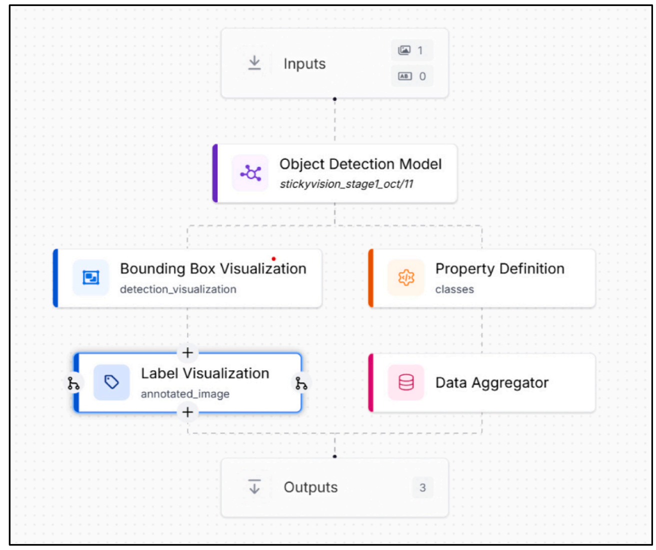

Fig. 3. Overview of the Roboflow workflow illustrating the sequence of functional blocks used in the InsectRoleVision system, including image input, YOLOv11-based detection, visualization, and aggregation of species richness and diversity metrics.

particular emphasis on characteristics such as mouthpart morphology and feeding behavior. This resulted in 43,007 annotated instances across 12 different arthropod classes. These annotations were generated through a structured labeling process to support model training and evaluation.

Classifying taxonomy by functional roles allowed for a more application-oriented categorization, especially in contexts such as ecological monitoring and agricultural pest control (Tsiafouli et al., 2017). For example, arachnids, which are predominantly carnivorous, act as predators in ecosystems and are considered natural enemies of pest species. Their role in both natural and urban ecosystems underscores their importance in maintaining ecological balance and reducing dependence on chemical pest control.

Among all arthropod orders, Diptera accounted for the largest number of specimens collected. To further refine the classification and capture ecologically important subgroups, special attention was given to the family Syrphidae (hoverflies), which was separated from the broader Diptera group due to its specific role as a pollinator. This detailed categorization ensures that both the ecological value and potential agricultural impact of arthropod taxa are adequately considered in the downstream analysis. Table 2 presents the results of the annotation and the ecological and economic roles of each arthropod taxon in this study.

3.2. Comparison of single- and modular double-stage detection strategies

A comprehensive evaluation was conducted to compare the performance of single-stage and modular double-stage object detection architectures for arthropod localization and classification in images. Model performance was assessed using mean average precision (mAP), precision, and recall, and differences among model architectures were statistically evaluated using one-way analysis of variance (ANOVA) followed by Tukey's HSD post-hoc tests.

Among the single-stage detection models evaluated, four versions of YOLO (You Only Look Once) and a transformer-based detector (RF-DETR) were compared. The results showed significant differences in detection performance among model architectures (one-way ANOVA, $F(4, 10) = 23.13$, $p < 0.05$). Post-hoc comparisons indicated that the YOLO-based models generally achieved higher performance than the transformer-based detector. In particular, YOLOv11 demonstrated the best overall performance among the single-stage models, achieving a

Table 2
Annotation of classes with the number of annotations and their ecological and economic role.

Taxonomy		No. of annotations	Ecological roles	Economic roles*
Order	Suborder/ Family			
Arachnida		848	Predator	Natural enemy
Coleoptera		1118	Detritivores, Herbivorous	Pest
Diptera	Brachycera (Suborder)	19,895	Detritivores, Pollinator	Pest
	Nematocera (Suborder)	9359	Detritivores, Pollinators	Pest
	Syrphidae (Family)	1112	Pollinator	Pollinator
Hemiptera	Aphidomorpha (Infraorder)	1218	Herbivorous	Pest
	Heteroptera (Suborder)	962	Herbivorous	Pest
	Membracoidea (Superfamily)	1832	Herbivorous	Pest
Hymenoptera	Apocrita (Suborder)	4817	Parasitoid wasp, Pollinators,	Natural enemy, Pollinators
	Formicidae (Family)	947	ND	ND
Lepidoptera		424	Pollinators	Pollinators
	Psocodea (Family)	475	Detritivores	Pest

ND – Not determined for their functional role at a lower taxonomic level.
[*] The economic role of arthropods depends on the location/habitat and scenario, with the definition for this study focusing more on the urban environment.

mean average precision (mAP) of 77.6%, precision of 80.0%, and recall of 73.3%.

The transformer-based model (RF-DETR) showed relatively lower performance compared to the YOLO architectures, despite previous studies reporting strong performance of transformer-based detectors in general object detection tasks (Sapkota et al., 2025). These differences may reflect task-specific challenges associated with small-object detection and domain-specific image characteristics in arthropod monitoring.

Fig. 4 compares the performance of YOLO-based and transformer-based models, highlighting their respective capabilities for identifying and localizing arthropods in scanned images. Detailed training and validation loss curves, as well as confusion matrices showing class-specific prediction performance, are provided in Appendix 1.

In contrast, for the double-stage detection strategy, the models used to localize target arthropods from scanned images showed significant performance differences among architectures (Fig. 5). One-way ANOVA revealed a significant effect of model architecture on localization performance (mAP) (F (4, 10) = 33.87, p < 0.05). Tukey's HSD post hoc comparisons indicated that YOLO-based models achieved significantly higher detection performance than the transformer-based model RF-DETR (p < 0.05).

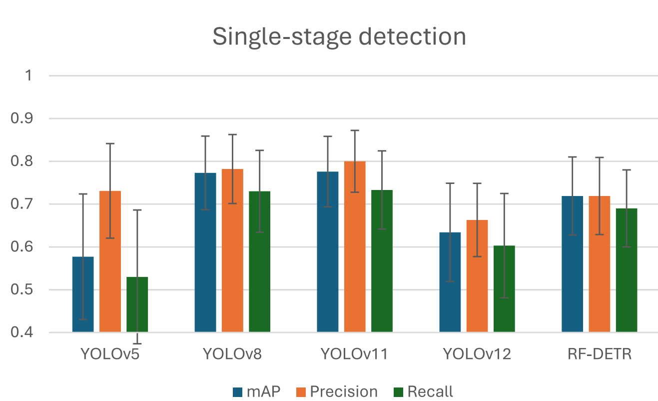

Fig. 4. Comparison of performance results between YOLO-based and transformer-based models in single-stage detection.

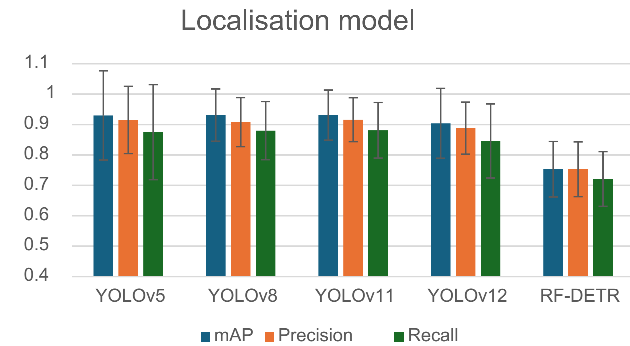

Fig. 5. Comparison of performance results between YOLO-based and transformer-based models in arthropod localization in the images.

Among the localization models evaluated, YOLOv11 outperformed the others. It achieved a mean average precision (mAP) of 93.1%, a precision of 91.6%, and a recall of 88.1%. These metrics highlight the model's strong ability to localize arthropods while minimizing false positives and false negatives.

For a more detailed insight into the model's learning behavior and classification effectiveness, see the training and test loss curves and confusion matrices in Appendix 2.

After the target arthropods were successfully localized and cropped into individual images, a comparative evaluation of four classification models was conducted. Classification performance was assessed using precision and recall, and differences among model architectures were analyzed using one-way analysis of variance (ANOVA) followed by Tukey's HSD post hoc tests.

Significant differences in classification performance were observed among the evaluated models (one-way ANOVA, F (3, 8) = 29.03, p < 0.05). Post hoc comparisons indicated that InceptionV3 achieved significantly higher performance than VGG19 (p < 0.05), while differences between InceptionV3, ResNet, and MobileNet were not statistically significant (p > 0.05).

InceptionV3, ResNet, and MobileNet all achieved precision and recall values above 87%, demonstrating strong performance in identifying arthropod classes from cropped image data. Among these models, InceptionV3 had the highest mean precision and recall values. Fig. 6 shows the comparative performance of the classification models, and additional details, including training and test performance curves and confusion matrices, are provided in Appendix 3.

When comparing the performance of single-stage and double-stage detection strategies, the best-performing single-stage model (YOLOv11) and the modular double-stage pipeline (YOLOv11–InceptionV3) showed similar classification performance.

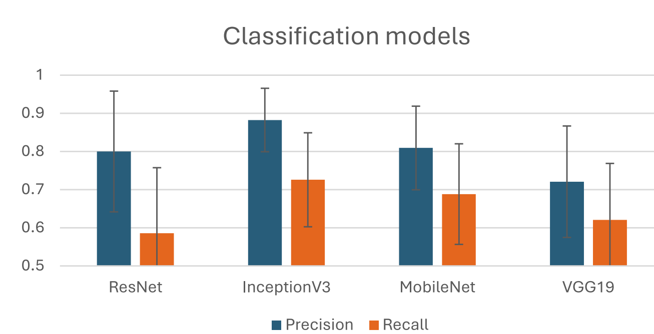

Fig. 6. Comparison of performance results between classification models in the classification of 12 groups of arthropods.

Statistical analysis found no significant differences in precision or recall between the two approaches (one-way ANOVA, precision: F (3, 8) = 30.99, $p > 0.05$; recall: F (3, 8) = 45.12, $p > 0.05$; Tukey's HSD, $p > 0.05$) (Fig. 7). However, the two approaches differed significantly in localization performance. The double-stage pipeline achieved higher mean average precision (mAP) due to its binary target detection configuration in the localization stage, and this improvement was statistically significant (one-way ANOVA, F (3, 8) = 41.74, $p < 0.05$). These results indicate that while both approaches provide similar classification accuracy, the modular double-stage framework improves target localization performance.

3.3. Ecological diversity and roles inference from two urban environments

We validated the ability of the *InsectRoleVision* system to infer ecological diversity and functional roles from scanned images of two urban environments: two department stores and ten grocery stores. To evaluate the system, we compared its results with manual identification and counting of arthropods in the images, then performed an independent *t*-test at $p < 0.05$. Table 3 presents the comparison of the number of arthropods per class in department stores and grocery stores, with most *InsectRoleVision* predictions not significantly different from the manual counts.

For the functional roles of arthropods, the number of predictions was exported from the Roboflow platform as an Excel file. We grouped the arthropods according to their economic roles: pest, natural enemy, or neutral. Table 4 shows the grouping of arthropods by their role per trap (n).

We were also able to access the locations of the traps and visualize the distribution of arthropods in the area. Fig. 8 shows the ratio of pests to natural enemies at each trap location in the two camps. Although many arthropods were collected per trap (Table 2), most were natural enemies. Therefore, if treatment were truly necessary, warehouse-trap 1 would be the location where specific control measures should be implemented to reduce the pest population. In contrast, Fig. 9 shows the distribution of ten grocery stores in Johore, Malaysia, and the ratio of pests to generalist (neutral) arthropods. The figure highlights certain areas with a higher proportion of pests where human intervention or chemical insecticide application could be targeted to reduce pest numbers, allowing for precise application at specific locations.

4. Discussion

Ecological research often aims to understand complex, interconnected interactions among biotic and abiotic factors, rather than to solve narrowly defined prediction tasks (Jeppesen et al., 2014). This contrasts with typical computer vision applications, which usually focus on well-defined objectives such as object detection or classification with explicit ground truth labels (Borji, 2018; Zendel et al., 2017). Therefore, in automated ecological monitoring, recognizing organisms by taxonomy alone may be insufficient to support ecological interpretation. To address this challenge, automated recognition systems must link taxonomic identification to functional roles, as supervised machine learning requires well-defined classes, while ecological inference depends on functional context. Functional roles can often be inferred at specific taxonomic levels; for example, members of the family Syrphidae are generally pollinators, whereas other Diptera–Brachycera, such as Muscidae and Calliphoridae, typically function as detritivores, mechanical vectors, or nuisance pests in natural and managed environments.The creation of an annotated dataset containing 43,007 instances across 12 arthropod taxa is a fundamental contribution of this study. Unlike previous datasets that rely solely on morphology-based classifications, this dataset incorporates ecological and economic functions into the annotation process. This functional taxonomy – particularly the explicit distinction of Syrphidae from Diptera based on their pollination function – offers a more application-oriented approach that is especially valuable for ecological monitoring and integrated pest management (IPM). These ecologically meaningful annotations not only support model training but also enhance the interpretability of results by linking predictions to real ecological functions (Pichler and Hartig, 2023). Additionally, the focus on features such as mouthpart morphology and feeding behavior provides taxonomic resolution relevant for decision-making in agricultural contexts and sets this dataset apart from many open-source image databases that lack consideration of functional roles. By connecting taxonomic classification to ecological impacts, the dataset supports both biodiversity assessment and actionable ecosystem management measures.

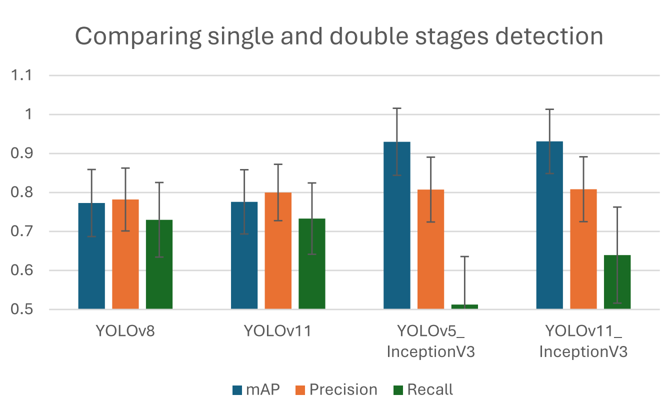

Fig. 7. Comparison of performance results between single-stage detection (YOLOv8 and v11) and modular double-stage detection (YOLOv5-InceptionV3 and YOLOv11-InceptionV3).

The quality of the images and annotations in our study stood out from previous studies in terms of resolution and detail. For example, previous studies using webcams (Badgujar et al., 2023) and DSLR cameras (Ong et al., 2022) achieved a maximum of 300 dpi with a resolution of 4000 × 8000 pixels. In contrast, the scanned images used in our study had a resolution of 1200 dpi and 14,039 × 10,200 pixels, about four to five times higher than images generated by cameras. This could be an advantage when training the model, especially for small arthropods measuring 2 to 3 mm. Our previous study, *ScannerVision* (Ong et al., 2025), also highlighted the method's ability to obtain high-resolution images for deep learning models.

Regarding the comparison between single-stage and double-stage detection, to our knowledge, this study is the first to directly compare single-stage and modular double-stage detection for arthropod and insect detection. Our results show that YOLOv11 is the best-performing single-stage detector, achieving a strong mAP of 0.776 and high inference speed, which is critical for real-time use in field monitoring systems. This aligns with the findings of Wang et al. (2025), who used YOLOv8n to detect and classify seven classes of insect pests and achieved mAP values ranging from 0.451 to 0.777. Our results for insects also address a gap in the study by Ong et al. (2022), who demonstrated YOLO-SIP (small insect pests) and obtained a mAP of 84.22% for five insects measuring five to ten millimeters. This indicates that our model was able to detect even smaller insects, such as 2 mm non-biting midges and mosquitoes.

On the other hand, the double-stage pipeline with YOLOv11 achieved a localization mAP of 93.1%, consistent with the findings of Zhu et al. (2021) that separating localization and classification into two inference stages can reduce classification noise, especially in scenarios with many visually similar categories, and can improve localization performance. In this framework, arthropods were annotated directly into 12 taxonomic classes during dataset preparation, while the double-stage structure was implemented at the modeling level: the first stage performs binary arthropod localization, and the second stage classifies the detected regions of interest into the predefined classes. From both entomological and practical perspectives, this modeling

Table 3
Comparison of the number of individuals per class (mean ± standard error) in each urban environment between *InsectRoleVision* and manual counts.

	Warehouse ($n = 10$)			Grocery shop (n = 10)		
	InsectRoleVision	Manual	p	*InsectRoleVision*	Manual	p
Arachnida	1.50 ± 0.16	1.25 ± 0.14	0.08	ND	ND	
Coleoptera	324.60 ± 53.95	330.2 ± 48.39	0.12	30.40 ± 5.51	32.30 ± 6.14	0.85
Brachycera	ND	ND		30.70 ± 5.04	29.20 ± 4.78	0.41
Heteroptera	ND	ND		6.50 ± 1.27	7.00 ± 1.33	0.09
Apocrita	560.40 ± 70.84	532 ± 67.16	0.56	ND	ND	
Formicidae	ND	ND		12.70 ± 4.04	23.30 ± 3.83	0.01*
Lepidoptera	4.10 ± 0.90	4.32 ± 1.05	0.09	12.30 ± 3.84	11.90 ± 3.42	0.17
Psocodea	ND	ND		4.40 ± 1.96	9.20 ± 1.68	0.03*

n = number of traps/sticky cards.
ND = not detected.
Categories not shown in the table had zero counts.
[*] $p < 0.05$ indicates a significant difference between system-based and manual counting.

Table 4
Number of arthropods according to their functional role in the urban environment.

	Warehouse		Grocery shop	
	Pest	Natural enemy	Pest	Neutral*
n	Coleoptera + Lepidoptera	Arachnida + Apocrita	Coleoptera + Brachycera	Heteroptera + Formicidae + Lepidoptera+ Psocodea
1	333	128	32	35
2	133	667	94	44
3	456	671	43	59
4	128	541	121	41
5	569	870	58	72
6	451	439	65	54
7	312	665	37	52
8	111	325	57	59
9	216	870	51	53
10	578	431	53	40

[*] Neutral roles of arthropods are neither pests nor natural enemies, but they can serve as components of ecosystem balance.

strategy offers several advantages. First, separating localization and classification enables inspection and validation of cropped regions of interest before classifier training, facilitating class balancing by augmenting underrepresented categories. Second, the modular design allows new taxonomic classes to be added to the classification stage without retraining the localization model, supporting scalability. Third, manual verification and correction of misclassified images can be efficiently performed by reorganizing cropped images within class-specific folders.

The improved mAP in the double-stage pipeline may result from lower task complexity per model: YOLOv11 is optimized for spatial localization in this setup, while InceptionV3 focuses exclusively on classifying pruned specimens – an architecture that parallels hierarchical biological classification methods. Although transformer-based models have generally performed well on general object recognition tasks (Dosovitskiy et al., 2020), they show lower performance in single- and double-level configurations. This may be because they are sensitive to dataset size and require extensive pre-training, making them less suitable for detecting small objects in high-resolution ecological images (Chen et al., 2024). Compared to previous studies on automatic insect classification (e.g., Martineau et al., 2017; Chen and Lin, 2023), this work improves both dataset granularity and recognition classification performance. Most comparable studies either lack ecological context annotations or focus on a narrow selection of taxa. Additionally, they often use only single-stage pipelines or conventional classifiers without fully leveraging the advantages of modular architectures.

Regarding the economic importance of arthropod prediction, the system-based insect counts closely matched manual scoring across sampling periods, indicating that the automated pipeline provides reliable abundance estimates comparable to expert-based assessments. This agreement supports the practical applicability of the system for routine monitoring tasks in economically relevant settings, where manual counting is time-consuming and labor-intensive. Since the study focused more on the validation of the system, other variables such as

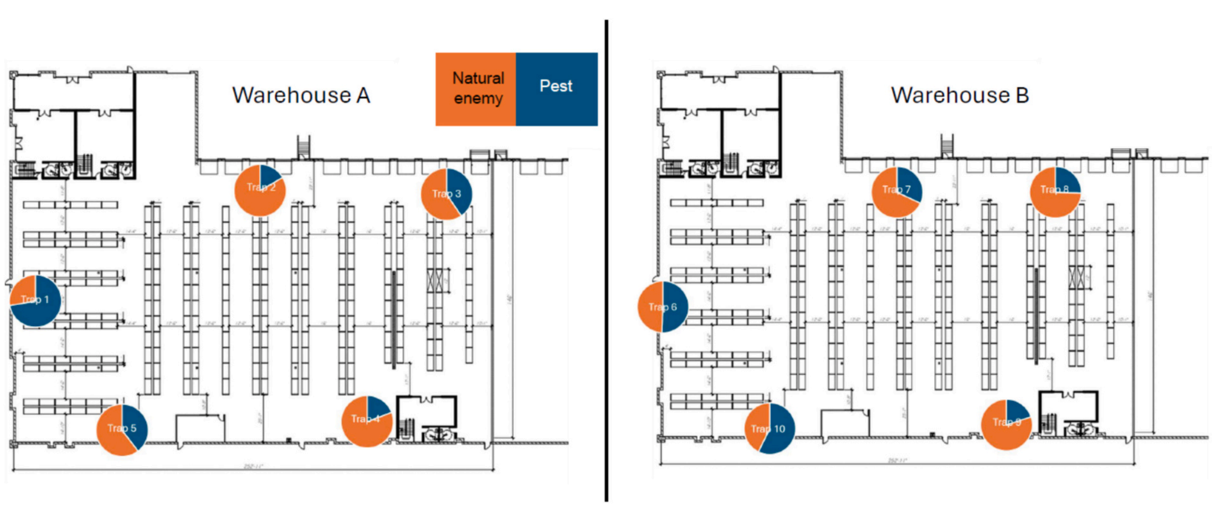

Fig. 8. The proportion of pest or natural enemy at each trap location at two warehouses.

Fig. 9. Illustrates the distribution of 10 grocery shops, showing the proportion of pest and neutral arthropods.

temperature, season and environmental parameters were not considered, so further analysis to predict pest outbreaks was not possible. Nevertheless, the validation was able to provide insight into deciding when pests should be controlled, for example, that the number of arthropods in the warehouse is still predominantly occupied by natural enemies and therefore chemical agents should not be used, otherwise a large proportion of the arthropods will be killed indiscriminately.

This study also has its limitations. First, the background of the prediction or the color of the sticky traps was white. As mentioned by Ong et al. (2024), the background color likely influences the model's performance, so *InsectRoleVision*'s predictions could differ on sticky cards of other colors. One way to mitigate this influence is to add a segmentation layer to the pipeline to remove all background, as demonstrated in Flatbug (Svenning et al., 2026). Second, several insect orders were excluded from this study. Many minority orders, such as Odonata, were not considered in the model's development, so the model needs to be updated with more training data to scale to different environments. Finally, performance metrics such as mAP and recall, while informative, do not capture ecological interactions in context-specific applications such as pest outbreak detection. This area should be refined in the future using task-based loss functions or ecological weighting.

Although the training data used in this study were collected from five wildflower sites within a single campus over a two-year period, the primary contribution of this work is the development of a scalable computational framework rather than achieving comprehensive ecological representativeness. The scalability of the proposed system refers to its modular design and its ability to be extended to new datasets and study systems with minimal modification.

To assess generalisability beyond the original training context, the framework was validated using two independent external datasets (Fig. 1), demonstrating that the approach can be applied to data collected under different conditions. Nevertheless, we acknowledge that broader ecological representativeness would require training and evaluation on more diverse habitats, taxa, and geographic regions, which is an important direction for future work.

The external validation using urban sticky-card datasets was conducted with a limited sample size and should therefore be interpreted as a proof-of-concept evaluation rather than a comprehensive assessment of ecological patterns. While the results show that the proposed framework can be applied to independent urban monitoring scenarios, broader inference of functional roles would require larger and more temporally replicated datasets.

In addition, the *InsectRoleVision* system uses a modular, multi-stage design in which Localization, classification, and ecological role inference are performed sequentially. Errors introduced at earlier stages may propagate to downstream analyses and influence ecological interpretation. However, this modularity also enables independent evaluation and improvement of each component, allowing error sources to be identified and mitigated. Future work will focus on quantifying uncertainty propagation across stages and integrating confidence-aware predictions to improve robustness in ecological applications.

5. Conclusion

InsectRoleVision offers a pipeline for arthropod detection and classification using double-stage deep learning strategies. By developing a functionally annotated dataset with over 43,000 instances across 12 arthropod taxa, we highlight the importance of ecological and economic features – such as pollination functions and predatory behavior – for enhancing model utility beyond purely morphological identification.

A comparative analysis showed that YOLOv11 consistently outperforms other models in both single-stage and double-stage configurations. The single-stage YOLOv11 model demonstrated strong performance in real-time recognition tasks, while the double-stage pipeline (YOLOv11 + InceptionV3) achieved higher classification accuracy, particularly in fine-grained tasks where distinguishing between ecologically similar taxa is crucial. This highlights the advantage of modular architecture for domain-specific applications.

Transformer-based models, although powerful for general object detection, were less effective in this context, likely due to the small object sizes and dataset-specific features. These results emphasize the importance of adapting model architecture to domain-specific data constraints and application goals.

Overall, this work makes a valuable contribution to automated biodiversity monitoring and sustainable agricultural management by integrating ecological knowledge, advanced detection models, and comparative assessment. Future research should aim to improve model robustness under field conditions, expand taxonomic coverage, and incorporate temporal data to more effectively track ecological dynamics.

CRediT authorship contribution statement

Song-Quan Ong: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Min-Hui Lim:** Writing – review & editing, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis. **Kim Bjerge:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Conceptualization. **Francisco Javier Peris-Felipo:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Rob Lind:** Writing – review & editing, Software, Project administration, Methodology, Investigation. **Toke Thomas Høye:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

We gratefully acknowledge the support of LivinGro® (Syngenta Crop Protection) by funding the current research. TTH were part-funded by the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No. 101060639 (MAMBO).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103701.

Data availability

All data and code used in this study are publicly available to ensure transparency and full reproducibility of the analyses. The annotated dataset is archived on Figshare DOI link: https://doi.org/10.6084/m9.figshare.30927338, and all scripts for data preprocessing, analysis, and figure generation are hosted in a public GitHub repository link: https://github.com/songguan26/InsectRoleVision.

References

Abdul-Khalil, S., Abdul-Rahman, S., Mutalib, S., Kamarudin, S.I., Kamaruddin, S.S., 2023. A review on object detection for autonomous mobile robot. IAES Int. J. Artificial Intel. (IJ-AI) 12 (3), 1033–1043.

Badgujar, C.M., Armstrong, P.R., Gerken, A.R., Pordesimo, L.O., Campbell, J.F., 2023. Real-time stored product insect detection and identification using deep learning: system integration and extensibility to mobile platforms. J. Stored Prod. Res. 104, 102196.

Bagyaraj, D.J., Nethravathi, C.J., Nitin, K.S., 2016. Soil biodiversity and arthropods: role in soil fertility. Economic and ecological significance of arthropods in diversified ecosystems: sustaining regulatory mechanisms, pp. 17–51.

Baker, R., Cannon, R.A.Y., Bartlett, P., Barker, I.A.N., 2005. Novel strategies for assessing and managing the risks posed by invasive alien species to global crop production and biodiversity. Ann. Appl. Biol. 146 (2), 177–191.

Bjerge, K., Karstoft, H., Mann, H.M., Høye, T.T., 2024. A deep learning pipeline for time-lapse camera monitoring of insects and their floral environments. Eco. Inform. 84, 102861.

Borji, A., 2018. Negative results in computer vision: a perspective. Image Vis. Comput. 69, 1–8.

Chakravarthy, A.K., Kammar, V., Shashank, P.R., 2016. Arthropods: evolution and ecology. Economic and ecological significance of arthropods in diversified ecosystems: sustaining regulatory mechanisms, pp. 1–16.

Chen, Y., Liu, S., Li, Y., Tian, L., Chen, Q., Li, J., 2024. Small Object Detection with Small Samples Using High-Resolution Remote Sensing Images. In. In: *Journal of Physics: Conference Series*, Vol. 2890, No. 1. IOP Publishing.

Chen, C.K., Lin, T.T., 2023. An automated online learning framework for insect pest image classification model enhancement. In: In *2023 ASABE Annual International Meeting* (p. 1) American Society of Agricultural and Biological Engineers. Paper No. 2300292 DO. https://doi.org/10.13031/aim.202300292. PY.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Houlsby, N., 2020. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Gratton, C., Denno, R.F., 2006. Arthropod food web restoration following removal of an invasive wetland plant. Ecol. Appl. 16 (2), 622–631.

Holthouse, M.C., Spears, L.R., Alston, D.G., 2021. Comparison of yellow and blue sticky cards for detection and monitoring parasitoid wasps of the invasive Halyomorpha halys (Hemiptera: Pentatomidae). J. Insect Sci. 21 (5), 1.

Huang, H., You, Z., Cai, H., Xu, J., Lin, D., 2022. Fast detection method for prostate cancer cells based on an integrated ResNet50 and YoloV5 framework. Comput. Methods Prog. Biomed. 226, 107184.

Ileke, K.D., Nwosu, L.C., Emeribe, E.O., 2023. Control of stored products arthropod pests. Forensic Entomology 75.

Jeppesen, E., Meerhoff, M., Davidson, T., Trolle, D., Sondergaard, M., Lauridsen, T., Nielsen, A., 2014. Climate change impacts on lakes: an integrated ecological perspective based on a multi-faceted approach, with special focus on shallow lakes. J. Limnol. 73.

Kammar, V., Rani, A.T., Kumar, K.P., Chakravarthy, A.K., 2020. Light trap: a dynamic tool for data analysis, documenting, and monitoring insect populations and diversity. In: Innovative Pest Management Approaches for the 21st Century: Harnessing Automated Unmanned Technologies, pp. 137–163.

Krenn, H.W., 2019. Form and function of insect mouthparts. In: Insect Mouthparts: Form, Function, Development and Performance, pp. 9–46.

Krenn, H.W., Aspöck, H., 2012. Form, function and evolution of the mouthparts of blood-feeding Arthropoda. Arthropod Struct. Dev. 41 (2), 101–118.

Lim, M.H., Chan, H.H., Ong, S.Q., 2025. An annotated image dataset of urban insects for the development of computer vision and deep learning models with detection tasks. Data Brief 60, 111673.

Martineau, C., Conte, D., Raveaux, R., Arnault, I., Munier, D., Venturini, G., 2017. A survey on image-based insect classification. Pattern Recog. 65, 273–284.

Ong, S.Q., Hamid, S.A., 2022. Next generation insect taxonomic classification by comparing different deep learning algorithms. PLoS One 17 (12), e0279094.

Ong, S.Q., Nair, G., Yusof, U.K., Ahmad, H., 2022. Community-based mosquito surveillance: an automatic mosquito-on-human-skin recognition system with a deep learning algorithm. Pest Manag. Sci. 78 (10), 4092–4104.

Ong, S.Q., Pinoy, N., Lim, M.H., Bjerge, K., Peris-Felipo, F.J., Lind, R., et al., 2025. ScannerVision: scanner-based image acquisition of medically important arthropods for the development of computer vision and deep learning models. Curr. Res. Parasitol. Vector-Borne Dis. 100268.

Ong, S.Q., Safree, A.S.A., Ismai, N.B.A., 2024. Vector-borne diseases on Borneo island: a scoping review. Malaria J. 23 (1), 337.

Pergantas, P., Papanikolaou, N.E., Malesios, C., Tsatsaris, A., Kondakis, M., Perganta, I., et al., 2021. Towards a semi-automatic early warning system for vector-borne diseases. Int. J. Environ. Res. Public Health 18 (4), 1823.

Pichler, M., Hartig, F., 2023. Machine learning and deep learning—a review for ecologists. Methods Ecol. Evol. 14 (4), 994–1016.

Ratnieks, F.L., Schrell, F., Sheppard, R.C., Brown, E., Bristow, O.E., Garbuzov, M., 2016. Data reliability in citizen science: learning curve and the effects of training method, volunteer background and experience on identification accuracy of insects visiting ivy flowers. Methods Ecol. Evol. 7 (10), 1226–1235.

Sapkota, R., Cheppally, R.H., Sharda, A., Karkee, M., 2025. RF-DETR object detection vs YOLOv12: a study of transformer-based and CNN-based architectures for single-class and multi-class greenfruit detection in complex orchard environments under label ambiguity. arXiv preprint arXiv:2504.13099.

Sunderland, K., Samu, F., 2000. Effects of agricultural diversification on the abundance, distribution, and pest control potential of spiders: a review. Entomol. Exp. Appl. 95 (1), 1–13.

Svenning, A., Mougeot, G., Alison, J., Chevalier, D., Molina, N.C., Ong, S.Q., et al., 2026. A general method for detection and segmentation of terrestrial arthropods in images. Methods Ecol. Evol. 17 (3), 727–739.

Tsiafouli, M.A., Drakou, E.G., Orgiazzi, A., Hedlund, K., Ritz, K., 2017. Optimizing the delivery of multiple ecosystem goods and services in agricultural systems. Front. Ecol. Evol. 5, 97.

Wang, J., Li, Y., Feng, H., Ren, L., Du, X., Wu, J., 2020. Common pests image recognition based on deep convolutional neural network. Comput. Electron. Agric. 179, 105834.

Wang, N., Fu, S., Rao, Q., Zhang, G., Ding, M., 2025. Insect-YOLO: a new method of crop insect detection. Comput. Electron. Agric. 232, 110085.

Wardhaugh, C.W., 2015. How many species of arthropods visit flowers? Arthropod Plant Interact. 9 (6), 547–565.

Wong, M.K., Guénard, B., Lewis, O.T., 2019. Trait-based ecology of terrestrial arthropods. Biol. Rev. 94 (3), 999–1022.

Zendel, O., Honauer, K., Murschitz, M., Humenberger, M., Fernandez Dominguez, G., 2017. Analyzing computer vision data-the good, the bad and the ugly. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1980–1990.

Zhu, X., Qiu, T., Qu, W., Zhou, X., Atiquzzaman, M., Wu, D.O., 2021. BLS-location: a wireless fingerprint localization algorithm based on broad learning. IEEE Trans. Mobile Comput. 22 (1), 115–128.