Novel nocturnal insect pest monitoring for sustainable crop protection using ensemble augmented deep learning classification

by Zhao, H., Fu, B., Auat Cheein, F., Butler, M., Harris, W.E., Pope, T.W. and Roberts, J.M.

Copyright, publisher and additional information: Publishers' version distributed under the terms of the Creative Commons Attribution License

DOI link to the version of record on the publisher's site

Zhao, H., Fu, B., Auat Cheein, F., Butler, M., Harris, W.E., Pope, T.W. and Roberts, J.M. (2025) 'Novel nocturnal insect pest monitoring for sustainable crop protection using ensemble augmented deep learning classification', *Smart Agricultural Technology,* 12, article 101244.

Smart Agricultural Technology 12 (2025) 101244

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Novel nocturnal insect pest monitoring for sustainable crop protection using ensemble augmented deep learning classification

Hui Zhao [a], Bo Fu [b], Fernando Auat Cheein [c], Matthew Butler [c], W. Edwin Harris [c], Tom W. Pope [c], Joe M. Roberts [c,*]

[a] Dalian Academy of Agriculture Sciences, Dalian, 116021, Liaoning, China
[b] Liaoning Normal University, Dalian, 116081, Liaoning, China
[c] Harper Adams University, Newport, TF10 8NB, Shropshire, United Kingdom

ARTICLE INFO

Keywords:
Vine weevil
Integrated pest management
Protected horticulture
Nocturnal pests
Automated pest detection
Smart monitoring

ABSTRACT

Vine weevil, *Otiorhynchus sulcatus* F. (Coleoptera: Curculionidae), is an economically important pest of soft fruit and ornamental crops globally. Its management has historically relied on broad-spectrum synthetic insecticides, but this has shifted toward integrated pest management compatible methods such as entomopathogenic nematodes and fungi that target soil-dwelling larvae. These methods require reliable pest monitoring tools to be practically effective and economically viable. Existing monitoring methods rely on detecting the nocturnal adult weevils as a proxy for larval presence, however, these are unreliable and time-consuming to implement. This may be addressed by developing an identification algorithm for adult weevils. Here we present results that show improved machine learning models can identify adult vine weevils under laboratory and semi-field conditions. Specifically, we employ a lightweight network model and use ensemble enhancement techniques to address potential issues such as color variations, occlusions, and deformations in the data labels. The proposed framework strategically integrates a lightweight network model with adaptive ensemble augmentation mechanisms to comprehensively address three core data challenges: (1) chromatic variance under varying illumination conditions, (2) partial occlusion from pest aggregation, and (3) morphological deformation during specimen collection. This is the first report of such technologies specifically developed for a nocturnal insect pest. It demonstrates the feasibility of an automated monitoring approach, which could benefit growers as it will provide more timely information about pest populations in their crops and better inform management decisions.

1. Introduction

Vine weevil, *Otiorhynchus sulcatus* Fabricius 1775 (Coleoptera: Curculionidae), is an economically important insect pest of soft fruit and ornamental crops [1]. Once regarded as a sporadic glasshouse pest in Europe and North America [2], this species has emerged as one of the most serious horticultural pests across its geographical range over the past five decades. Its increased economic impact is largely attributed to the expansion of the horticultural sector and the adoption of new growing practices, such as the use of black polythene mulches, that have inadvertently favored vine weevil development. These changes reduce insecticide efficacy and create protective microhabitats shielded from unfavorable climatic conditions [3].

Vine weevil control strategies have increasingly shifted from reliance on persistent, broad-spectrum insecticides to the use of entomopathogenic nematodes (EPNs) and fungi (EPFs) targeting the soil-dwelling larvae [4]. These biologically based methods represent a key advancement toward more sustainable, integrated pest management (IPM). As with all IPM programmes, effective pest monitoring is essential to reduce dependency on synthetic insecticides [5]. Early detection of vine weevil within crops is critical to successful management. Ineffective monitoring often results in growers remaining unaware of economically damaging populations until substantial crop loss has occurred [6]. Monitoring efforts are further complicated by the adults' nocturnal feeding behavior and the larvae and pupae's subterranean lifestyles [3]. Consequently, the development of robust IPM programmes has been hampered by inadequate monitoring techniques [1,7].

Monitoring efforts typically focus on adult weevils, as larval monitoring involves root sampling and this is labor-intensive and potentially harmful to crops [8]. The most common adult monitoring techniques

---

* Corresponding author.
E-mail address: jroberts@harper-adams.ac.uk (J.M. Roberts).

https://doi.org/10.1016/j.atech.2025.101244
Received 3 June 2025; Received in revised form 24 July 2025; Accepted 24 July 2025

Available online 4 August 2025
2772-3755/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

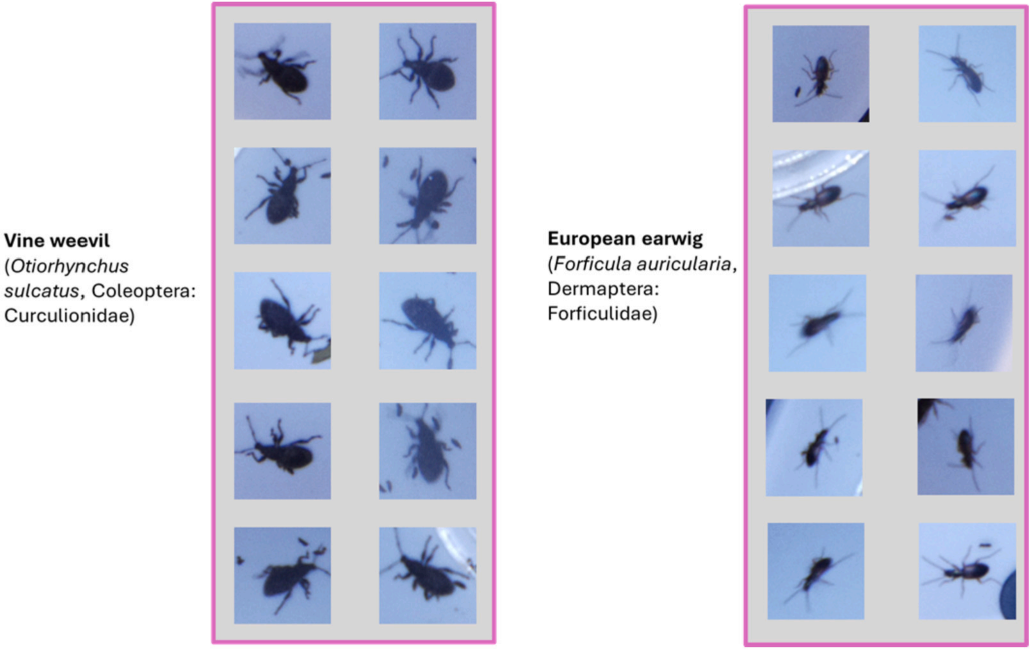


Fig. 1. Several snapshots of the two cultures used in this work: vine weevil and earwig. These images are later used for the training of our detection approaches.

are visual assessments, such as identifying characteristic leaf notching or gently shaking plants at night to dislodge individuals [6,9]. While leaf notch assessments may be reliable early in the season, newer damage becomes increasingly difficult to distinguish from older notches over time, and shaking may dislodge fruit or foliage. Moreover, these indirect methods can delay detection, allowing oviposition before control measures can be implemented [3]. Given the impracticality of night-time inspections, more effective and accessible monitoring methods are urgently needed. Although some efforts have focused on acoustic detection of larval stages [8], most research has concentrated on artificial refuges (which allow entry and exit) or traps (which prevent exit) for adult monitoring. In this study, both are collectively referred to as monitoring tools.

Monitoring tools exploit the negative phototaxis exhibited by adult vine weevils, prompting them to seek shelter during daylight hours [10,7]. A variety of designs have been evaluated, including grooved wooden boards placed on the ground [9], pitfall traps, corrugated cardboard around stems or placed on the ground, traps designed for other insect pests [11], and purpose-built vine weevil traps [10]. Despite widespread use, debate persists over the efficacy and reliability of these tools. Key factors influencing performance may include tool size, color, shape, and number of entry points [7]. A semi-field assessment of six monitoring tool designs demonstrated that although all were capable of detecting vine weevil presence, the number of individuals recovered from each tool varied greatly [10]. For other weevil species, tool efficacy and consistency have been improved through the use of semiochemical lures, typically sex or aggregation pheromones, to increase adult attraction (e.g., [12]). However, despite over two decades of research into vine weevil chemical ecology, no effective attractants have been identified. Consequently, the development of improved night-time monitoring methods that do not require manual inspection or direct trapping remains a key objective in advancing vine weevil management.

Monitoring approaches that are fast to deploy and capable of automated pest identification would represent a significant step forward, reducing insecticide use and promoting IPM adoption. One promising avenue involves integrating computer vision and machine learning to recognize and classify pest-related images or video footage. Similar methodologies have already been applied to automated pest monitoring (e.g., [13,14]), although none have yet been tailored for nocturnal beetle pests, to the authors' knowledge. This study, therefore, focuses on implementing contemporary machine learning techniques to address the challenges of nocturnal pest detection and classification. In this manuscript, we develop and evaluate a machine learning-driven monitoring approach that addresses these challenges by combining entomological expertise with automated image-based detection systems.

2. Materials and methods

2.1. Insects

2.1.1. Vine weevil culture

Vine weevil, *Otiorhynchus sulcatus* (Coleoptera: Curculionidae) at various larval stages were collected from commercial strawberry (*Fragaria* × *ananassa* cv. Duchesne) crops grown in Staffordshire (UK) during autumn 2021. Larvae were maintained on strawberry (cv. Elsanta) plants housed within a 47.5 cm^3 white mesh cage (BugDorm-4S4545, MegaView Science Co. Ltd., Taichung, Taiwan) in a controlled environment room (20 °C; 60% relative humidity; 16:8 photoperiod) (Fitotron, Weiss Technik, Ebbw Vale, Wales, UK). Resultant adults from this larval population were maintained under the previously stated environmental conditions using a standard method of placing the weevils in plastic terrariums (30 × 19.3 × 20.6 cm) containing yew (*Taxus baccata*) branches and moist paper towels that were replaced weekly [10]. All adult weevils used in this study were at least one month old and had been confirmed to be reproductively active, ensuring that subsequent monitoring tool tests used a field-representative pest population (Fig. 1).

2.1.2. Earwig culture

Adult European earwigs, *Forficula auricularia* Linnaeus 1758 (Dermaptera: Forficulidae) were collected from an experimental polytunnel containing strawberry plants (cv. Elsanta) at Harper Adams University (Shropshire, UK) during January 2022. Individuals were combined into a single laboratory culture and maintained on strawberry leaves infested with potato aphids, *Macrosiphum euphorbiae* (Hemiptera: Aphididae) housed within a 47.5 m^3 white mesh cage (BugDorm-4S4545) in a controlled environment room (20 °C; 60% relative humidity; 16:8 photoperiod) (Fitotron). This population served as a non-vine weevil species for training a machine learning model (Fig. 1).

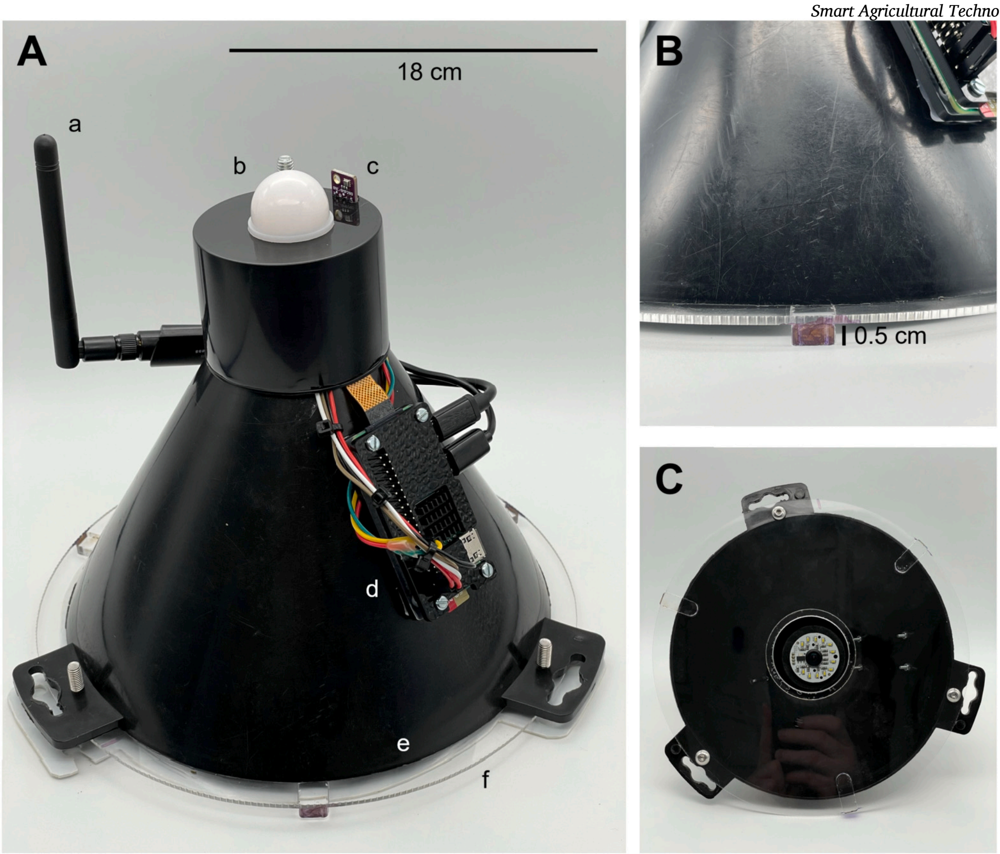


Fig. 2. Modified commercial vine weevil trap to add smart features (A) assembled for laboratory dataset collection with (a) Wi-Fi antenna, (b) BH1750FVI lux sensor, (c) BME280 temperature/humidity sensor, (d) Raspberry Pi Zero 2W, (e) clear acrylic and (f) entry location, (B) close-up of entry location and 0.5 cm stand-offs and (C) internal view of the (g) LED ring light and camera.

2.2. Image data collection and dataset preparation

We utilized a commercially available vine weevil monitoring tool retrofitted with imaging hardware to capture nocturnal insect activity. The monitoring tool was equipped with a Raspberry Pi camera module (V2.1; PiHut, Haverhill, Suffolk) and light emitting diode (LED) ring light (PiHut) connected to a Raspberry Pi Zero 2W microprocessor (PiHut) (Fig. 2). It also included sensors to measure environmental conditions such as light intensity (BH1750FVI lux sensor) and temperature/humidity (BME280) as well as a Wi-Fi antenna for remote data transfer. The camera was positioned to view the trap's interior entry platform, and images were automatically captured every ten minutes throughout the night. A primary objective of image collection was to acquire clear photographs of individual adult weevils entering the trap. However, adult vine weevil exhibited thigmotactic behavior, often clustering together, which made it challenging to isolate and count individuals directly from full-frame images. To address this, we developed an automated region-of-interest (ROI) extraction procedure. The software compared consecutive images and flagged new objects (insects) appearing in the trap by detecting differences between frames. When a newly arrived insect was detected, the algorithm cropped a sub-image of size 100x100 pixels around that ROI for labeling. Additional filtering steps were applied to ensure each cropped sub-image contained a single, centrally positioned insect: specifically, we computed the image's moment of inertia to confirm the subject was roughly centered (i.e., the image was 'balanced'), and we measured the proportion of white background to avoid crops with excessive empty space. Cropped insect images that failed these criteria (off-center or mostly background) were discarded. The resulting set of processed 100 × 100 RGB images was saved to a GitHub repository for subsequent model development and is the basis of the dataset used in this study (Fig. 1). A total of 400 labeled images were obtained using this procedure. Each image was annotated with a class label: "vine weevil" for target pests and "earwig" for non-targets. For model training and evaluation, the dataset was split into training, validation, and test sets. In the baseline (unaugmented) dataset, 314 images (78%) were randomly assigned for training, 46 (11%) for validation (used to tune hyperparameters), and 47 (11%) for final testing. This split ensured that no test images were seen during training and that class proportions were maintained across subsets.

2.3. Machine learning implementation

Our approach employs a lightweight deep learning framework augmented with an ensemble of data transformations to achieve robust pest identification. The task is formulated as an image classification/detection problem on the 100 × 100 ROI images: given a cropped insect image, the model must identify whether the insect is a vine weevil or not (earwig). We evaluated several compact convolutional neural network architectures as candidate backbones for this task, including a custom CNN, AlexNet [15], and three variants of the YOLO object detection family (YOLOv5, YOLOv8, and a similarly small-footprint YOLOv11 experimental model [16,17]. These models were chosen for their balance of accuracy and efficiency, as real-time or in-field deployment would require limited computational overhead. Each model was integrated into our augmented training framework described below. Preliminary trials with standard pretrained architectures (e.g., ResNet [18], Faster R-CNN [19] confirmed that even when fine-tuned on our dataset, off-the-shelf models could achieve basic classification of weevils vs. non-weevils (Table 1). However, their accuracy plateaued due to the limited and biased training data with many training images suffering from unusual orientations or incomplete insect visibility (e.g., only half of a weevil visible due to overlap with another insect). Fig. 3 illustrates several represen-

Table 1
mAP and Recall comparison of different models with varying batch sizes and augmentation strategies.

Model	Batch Size 16 w/o Aug		Batch Size 16 w/ Aug		Batch Size 32 w/o Aug		Batch Size 32 w/ Aug	
	mAP	Recall	mAP	Recall	mAP	Recall	mAP	Recall
CNN	0.803	0.871	0.844	0.862	0.803	0.895	0.841	0.850
AlexNet	0.925	0.774	0.940	0.833	0.915	0.922	0.927	0.754
YOLOv5	0.988	0.947	0.991	0.958	0.991	0.947	0.988	0.960
YOLOv8	0.976	0.946	0.989	0.981	0.989	0.952	0.992	0.973
YOLOv11	0.992	0.970	0.993	0.985	0.993	0.973	**0.994**	**0.987**

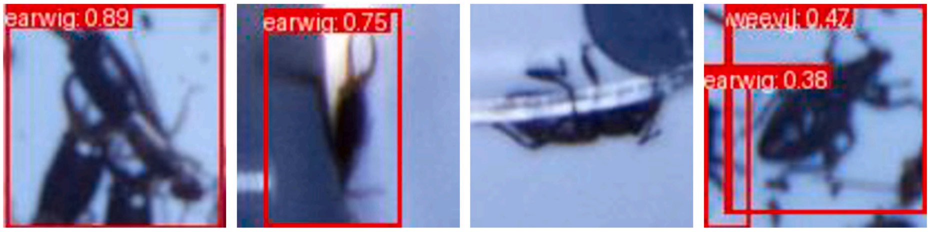


Fig. 3. Quantify schematic diagram after being affected by occlusion and other factors.

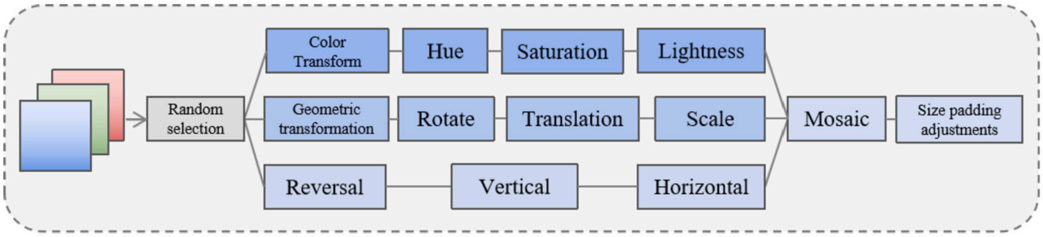


Fig. 4. Schematic diagram of augmentation operation.

tative challenging cases: occluded weevils and multiple insects in one frame that confounded the classifier. These failure cases highlighted the need for an enhanced training strategy to improve generalization.

To address the data limitations, we implemented an ensemble data augmentation strategy. The key idea is to artificially diversify the training samples by applying a random transformation to each image every time it is used for training. This approach exposes the model to a broader range of appearances for the target pest, thereby mitigating overfitting to the small original dataset and improving robustness to the issues of chromatic variation, occlusion, and deformation. We designed a suite of augmentation operations that includes geometric transformations, photometric adjustments, and synthetic occlusion, inspired by the specific challenges of our data. In particular, our augmentation pipeline can perform: (i) Random rotation by rotating the image by a random angle (θ) uniformly sampled in the range [-30°, 30°] about its center (filling any blank areas with the background color); (ii) Random scaling by zooming in or out by a scale factor s uniformly sampled from [0.8, 1.2], followed by resizing back to 100×100 pixels; (iii) Random half-cropping with 50% probability, cropping out either the left or right half of the image (simulating an insect at the edge or partially out-of-frame), then padding the missing region with white to restore the full 100×100 size; (iv) Color jitter that adjusts brightness and contrast by small random offsets (this was implemented although not explicitly listed in the pseudocode, to account for varied lighting); (v) Image inversion by flipping the image horizontally (mirror image) or inverting colors (negative image) in some cases to further increase appearance diversity. These operations were applied in combination: for each training sample, a random subset of transformations is chosen and applied in sequence to produce an augmented sample. A schematic overview of the augmentation process is provided in Fig. 4.

We can formally define the augmentation procedure as follows. Let $x \in \mathcal{X}$ be an input image and $y \in \mathcal{Y}$ its label (vine weevil or non-weevil), with underlying data distribution $P_{\text{data}}(x, y)$. We define a family of stochastic transformation functions $T_\theta : \mathcal{X} \to \mathcal{X}$, parameterized by θ drawn from a distribution Θ. Here, θ represents a particular augmentation operation and its random parameters (e.g., a rotation by 17° or a brightness increase of 20%), and Θ is the overall distribution of augmentation choices (in practice, a mixture of the geometric and photometric transformations described above). For each original sample (x, y), an augmented sample is generated by sampling a $\theta \sim \Theta$ and applying the transformation:

$$x' = T_\theta(x), \qquad \theta \sim \Theta \tag{1}$$

The new augmented data distribution P_{aug} induced by this process can be expressed as:

$$P_{\text{augmented}}(x', y) = \int P_{\text{data}}(x, y) \cdot \Theta(\theta) \cdot \delta(x' - T_\theta(x))\, dx\, d\theta \tag{2}$$

where $\delta(\cdot)$ is the Dirac delta function ensuring x' is exactly the transformed version of x. The learning objective with augmentation is to minimize the expected risk over this augmented distribution rather than the original:

$$\min_\theta E_{(x,y)\sim P_{\text{data}}} \left[E_{\theta\sim\Theta} \left[L(f_\theta(T_\theta(x)), y) \right] \right] \tag{3}$$

where f denotes the model's prediction function (with parameters to be learned) and L is the loss function (here, cross-entropy for classification). In essence, training on augmented data $T_\theta(x)$ encourages the model f to learn features invariant to the transformations T, thus improving generalization. We refer to the resulting model as an ensemble-augmented deep learning classification network, since it effectively trains on an ensemble of transformed versions of each original image. Fig. 5 illustrates the overall framework that consists of two main components: (i) the data augmentation module and (ii) the lightweight classification network. During each training iteration, an input image is stochastically augmented (one of the transformations is applied at ran-

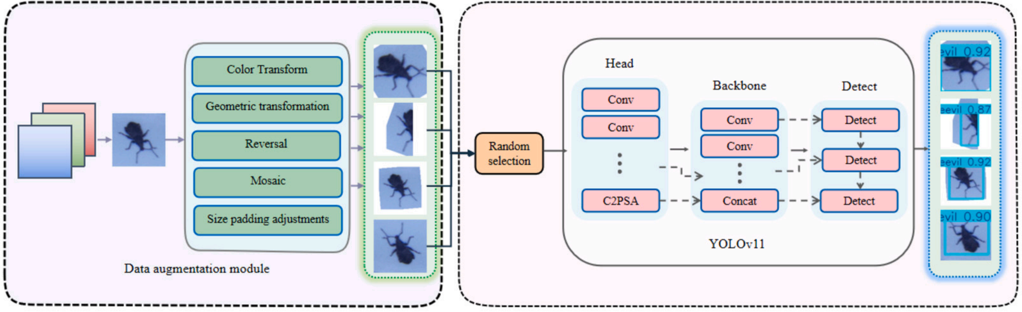


Fig. 5. Schematic diagram of the whole framework.

dom) and the modified image is fed into the network. The network then produces a prediction, which is compared to the true label to compute the loss. The model parameters are updated via backpropagation, and this process repeats for each epoch. By the end of training, the network has seen many altered versions of each insect image, learning to recognize vine weevils even when their appearance or context is varied.

For our experiments, we trained each of the five selected models (CNN, AlexNet, YOLOv5, YOLOv8, YOLOv11) under two training regimes: with the augmentation pipeline enabled, and with no augmentation (for baseline comparison). We also examined two batch sizes (16 and 32) to assess the effect of batch gradient estimation on performance, especially given our small image size and the presence of small object features. All models were trained for 500 epochs on an NVIDIA GPU. Stochastic gradient descent was used with an initial learning rate of 0.01, gradually reduced to 1×10^{-5} by the end of training (a cosine decay schedule was used). A brief 3-epoch warm-up at the start helped stabilize training. We applied a weight decay of 5×10^{-4} to regularize the CNN weights. A fixed random seed (4) was used in all training runs to ensure reproducibility of the data splitting and augmentation sampling. When augmentation was enabled, the effective number of training samples doubled (each original image could yield two distinct augmented variants per epoch on average). For instance, starting from 314 original training images, each epoch's training set was expanded to 628 augmented samples; in practice we rounded this to 634 to keep class proportions balanced, and similarly increased the validation set from 46 to 80 and test set from 47 to 79 by augmenting those images for consistency in evaluation.

2.4. Evaluation metrics

We evaluated model performance using precision, recall, and mean average precision (mAP) computed on the held-out test set. A positive prediction was defined as the model correctly identifying a vine weevil. In our context (binary classification/detection of weevil vs. earwig), a true positive (TP) occurs when a weevil is correctly identified, a false positive (FP) when a non-weevil is misclassified as weevil, and a false negative (FN) when a weevil is missed (classified as non-target). Precision and Recall are given by:

$$\text{Precision} = \frac{TP}{TP+FP}, \qquad \text{Recall} = \frac{TP}{TP+FN}. \tag{4, 5}$$

Since we are effectively dealing with two classes (weevil vs. non-weevil), we calculate the average precision (AP) for the vine weevil class and report mAP as the mean of the AP values across both classes (or equivalently, AP of weevil and AP of non-weevil). In general form, if AP_c is the average precision for class c and C is the number of classes, then:

$$\text{mAP} = \frac{1}{C}\sum_{c=1}^{C} AP_c. \tag{6}$$

3. Results

3.1. Quantitative performance

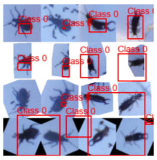
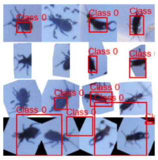
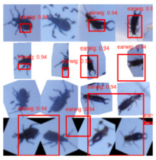
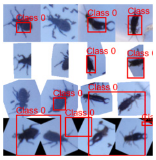
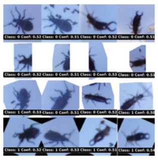
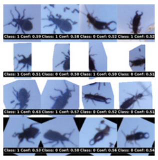
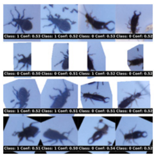
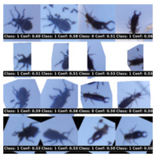
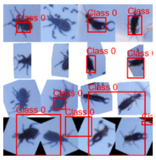
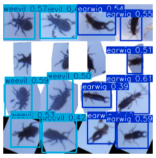
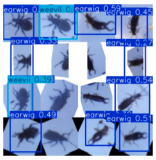
Overall, the proposed ensemble augmentation strategy yielded consistent improvements in detection accuracy across nearly all models and settings. When the five compact backbones were trained on our insect dataset, augmentation lifted performance in every case in terms of mAP and, in most cases, for recall as well (Table 1). For example, the YOLOv5 model (batch size 16) achieved an mAP of 0.988 and recall of 0.947 without augmentation, which increased to 0.991 mAP and 0.958 recall when trained with our augmentation pipeline. YOLOv11 saw its already high accuracy further improved to 0.993–0.994 mAP and 0.985–0.987 recall with augmentation (depending on batch size), making it the best performing classifier tested. Simpler models also benefited from augmentation with the baseline CNN's mAP rising from 0.803 to 0.844 (batch 16) while AlexNet saw improvements in mAP from 0.925 to 0.940 (with a modest recall increase from 0.774 to 0.833). These improvements indicate that enriching the training set with diversified examples helps the networks generalize better to the test data. Doubling the batch size from 16 to 32 had a smaller effect on raw accuracy metrics than augmentation, but we observed that it produced more stable training and slightly improved recall in some cases (e.g., CNN recall 0.871 at batch 16 vs. 0.895 at batch 32, without augmentation) (Fig. 4; Table 1). The augmented models consistently show higher median mAP and recall with reduced variability in performance (Fig. 6).

3.2. Qualitative detection

Beyond the numerical gains, we inspected the detection outputs from each model to qualitatively assess how augmentation and model choice impacted the quality of pest localization and identification. Table 2 presents representative detection outcomes for each model under different training scenarios (with/without augmentation and batch size 16 vs 32). Several clear trends can be identified from these visual comparisons. First, the YOLO-based detectors (v5, v8, v11) produced markedly more precise bounding boxes around the weevils than the simpler CNN or AlexNet. In cases with no augmentation, the CNN and AlexNet often struggled and their output images show either incomplete localization (e.g., bounding boxes that do not tightly fit the insect) or false-positive highlights on background elements that resemble insect texture. In contrast, the YOLO models, even without augmented training, delineated the weevil bodies much more sharply and rarely confused background patterns for insects (see Fig. 6).

Training with the ensemble augmentation framework significantly improved each model's ability to handle difficult cases. Augmented models robustly detected partially occluded or oddly illuminated weevils that their unaugmented counterparts missed or misclassified. For instance, in Table 2 we include scenarios where a vine weevil is half-obscured by another insect or only faintly lit; the augmented YOLOv8 and YOLOv11 still successfully identify the weevil (drawing a correct box around the visible part of the insect), whereas the non-augmented

Table 2
Qualitative results in the visual aspect.

Model	Batch Size 16 w/o Aug	Batch Size 16 w/ Aug	Batch Size 32 w/o Aug	Batch Size 32 w/ Aug
CNN				
AlexNet				
YOLOv5				
YOLOv8				
YOLOv11				

versions either failed to detect it at all or placed an incorrect box. Similarly, augmentation helped reduce false negatives for the simpler networks. AlexNet with augmentation identified some weevils in images containing several overlapping vine weevils that it overlooked without augmentation. These qualitative improvements align with the earlier metric gains in recall: the augmented models have learned to recognize vine weevils, even when the input deviates from ‘ideal’ images, by focusing on key morphological features such as portions of the distinctive snout or leg shape present even in partial views. Increasing the batch size from 16 to 32 further enhanced the detection stability. In the augmented models, a larger batch yielded slightly tighter and more confident bounding boxes in the qualitative results. We observed fewer misaligned boxes when multiple weevils were present, indicating that batch = 32 training helped the network better distinguish individual insects in groups (Table 2). Among all the model configurations, the YOLOv11 model trained with augmentation and a batch size of 32 delivered the best overall performance. It consistently detected all vine weevils in the test images, even under the most challenging conditions (such as multiple overlapping weevils, or individuals touching the trap edge). Its bounding boxes were also the most accurate, tightly fitting the insect outlines with minimal background included, and it produced virtually no false positives on blank areas.

4. Discussion

Recent advances in machine learning have facilitated the development of automated pest detection and identification tools with numerous studies applying deep learning to this problem, but these are largely focused on flying lepidopteran and dipteran pests. The present study demonstrates that a lightweight-network detector augmented with an ensemble-augmentation pipeline can identify a nocturnal beetle pest with accuracy in line with the best single-species models published to date. When five compact backbones were trained on our dataset, augmentation lifted performance in every case: for YOLOv5, mAP went from 0.988 to 0.991 and recall from 0.947 to 0.958 at batch-size 16, while YOLOv11 reached 0.994 mAP and 0.987 recall under the same conditions. Doubling the batch size to 32 produced qualitatively tighter bounding boxes by stabilizing gradient estimates for small objects. These gains confirm the value of our augmentation strategy, which (1) applies targeted geometric and color perturbations, (2) injects synthetic occlusions/overlaps, and (3) purposefully simulates partial visibility so that the network learns to focus on local, discriminative cues rather than complete outlines. The efficacy of the proposed ensemble-augmented technique is empirically validated through systematic experimentation and the data in Table 1, demonstrating consistent performance gains when applied to five prevalent categories of lightweight deep learning-based image classification models compared to their baseline counterparts without ensemble-augmented integration. This improvement primarily stems from the technique’s capacity to enhance data diversity in training databases, enabling classification networks to effectively learn critical patterns from challenging scenarios inadequately addressed in standard preprocessing pipelines, particularly partial insect visibility due to localization inaccuracies or overlapping specimens. Table 2 also demonstrates that implementing increased batch sizes coupled with our proposed augmentation framework enables classification networks to achieve enhanced bounding box localization accuracy across diverse insect imagery, consistently performing robustly irrespective of specimen integrity. As illustrated in Table 2, our augmented network model demonstrates superior performance in identifying occluded or cropped insects, cases where traditional recognition methods typically fail. This

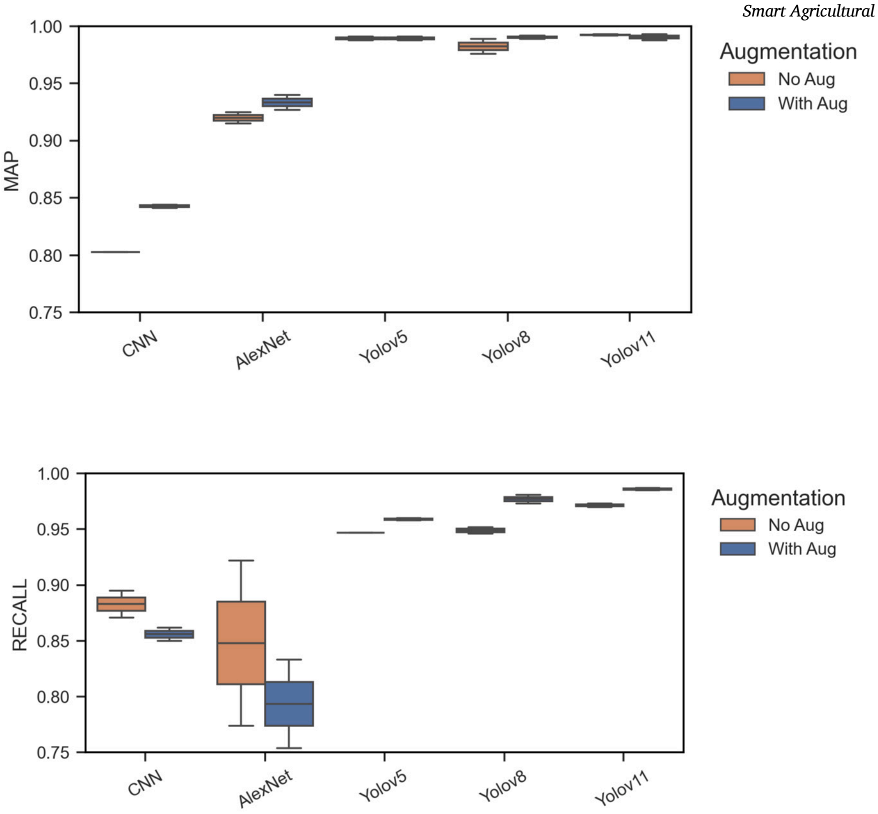


Fig. 6. Boxplot comparison of mean Average Precision (mAP) and Recall for the five models under two training conditions: without augmentation (brown) and with the proposed augmentation (blue). For each classifier (CNN, AlexNet, YOLOv5, YOLOv8, YOLOv11), the median performance with augmentation is equal or higher than without, and the spread (interquartile range) is generally reduced.

improvement is evident both in higher recognition probabilities and more accurate selection rates. The performance improvement stems from our augmentation techniques, which enhance the model's ability to learn from challenging cases including occluded, partially visible, and rotationally distorted insect specimens, ultimately yielding more robust recognition capabilities.

The proposed ensemble-augmented deep learning classification framework demonstrates significant advantages in pest recognition scenarios by dynamically combining diverse augmentation operations tailored to dataset characteristics. This approach enhances data diversity for deep image classification models through three key mechanisms as follows. Firstly, the system intelligently ensembles geometric transformations and occlusion simulations based on target pest morphology and occlusion patterns, optimizing feature representation learning. Secondly, by synthetically generating realistic pest variations (partial occlusion, overlapping specimens), it reduces dependency on large-scale manual annotations while maintaining model accuracy. Thirdly, the augmentation pipeline specifically addresses pest overlap challenges by simulating partial visibility scenarios, forcing the model to focus on discriminative local features rather than complete specimen shapes. When compared to the current state of automated pest monitoring, our study results are at the upper end of what has been reported for single-species detectors. The CNN pipeline used by Ding and Taylor [13] for codling-moth pheromone traps achieved around 87% mAP on trap images captured under mixed field lighting, whereas TPest-RCNN used by Li et al. [20], optimized for whiteflies and thrips on yellow sticky traps, reached 95.2% mAP on a two-class greenhouse dataset. Once the class count rises, performance falls steeply: the Pest24 benchmark, which spans 24 taxa and more than 190,000 labeled insects, reports 69.59% mAP and 77.71% recall even after architectural tuning with the Pest-YOLO variant [21]; the IP102 dataset, with 102 classes and a natural long-tailed distribution, remains a severe challenge for deep classifiers despite its 75,000 images. Against this landscape the present detector's ≈ 99% mAP illustrates that, for a well-defined nocturnal beetle, accuracy comparable with the best lepidopteran and whitefly systems is attainable on a much smaller training dataset.

High recall has direct agronomic value. Adult vine-weevils typically begin oviposition three to four weeks after emergence [1]; missing that early window compromises the efficacy of larval biocontrols such as entomopathogenic nematodes. With recall approaching 0.96, the system is expected to miss fewer than one adult in twenty, enabling just-in-time interventions instead of prophylactic treatments. Because the Raspberry Pi camera records images every ten minutes, growers could receive a quantitative time-series of nightly activity that is impossible to obtain through manual scouting [22]. Such data can not only guide treatment scheduling but also allow the efficacy of control measures to be audited, a key requirement for certification schemes that reward evidence-based IPM.

Several limitations nevertheless remain. Augmented images, while diverse, cannot reproduce every real visual contingency, leaving the model vulnerable to subtle domain shifts. Because the detector is binary, non-target beetles, earwigs or large leaf fragments sometimes trigger false positives; an explicit "unknown" channel coupled with hard negative mining will be required before broad deployment. The principal sources of error are heavy occlusion and specular glare, which occasionally suppress detections or distort boxes despite augmentation efforts. Practical deployment also raises additional engineering questions. Illumination must be bright enough for the camera yet benign to the insect; diffuse near-infra-red LEDs or filtered yellow wavelengths are the most promising options, but long runs in humid glasshouses mandate lens heaters and hydrophobic coatings. Edge inference on the current prototype draws less than three watts, suggesting that a solar-battery package similar to those described for light-trap systems powered entirely by photovoltaic modules is feasible. Component costs of approximately US $200–300 per trap should be offset in high-value horticulture by reduced labor and the avoidance of whole-crop losses; moreover, the rich digital audit trail offers an intangible asset that conventional monitoring lacks. Analysis of failure cases indicates three opportunities for immediate improvement. First, a lightweight verification classifier that focuses on the beetle's distinctive rostrum would suppress false positives from elongate debris. Second, temporal smoothing over successive frames can recover individuals missed in a single image owing to occlusion. Third, an out-of-distribution filter based on predictive entropy would allow the camera to flag unfamiliar objects, an essential safeguard once additional pest classes are introduced.

Future research should pursue three converging paths. Season-long validation in multiple protected cropping environments to determine monitoring tool performance under variable humidity, crop architecture and lighting; incremental inclusion of new taxa using hierarchical detectors and datasets such as Pest24 will broaden utility; and model pruning plus quantization will deliver sub-50 ms inference on microcontrollers, enabling full video analysis for behavioral classification as well as detection. Parallel advances in non-visual sensing using acoustic monitoring could support multimodal smart tools where vision, pheromones and sound are combined into a single, low-maintenance IPM tool [8].

5. Conclusion

This study presents the first automated identification method for nocturnal vine weevils using a lightweight neural network enhanced with ensemble augmentation techniques. The framework specifically addresses three critical challenges in field-collected pest imagery: (1) chromatic variations under uneven lighting, (2) occlusion from clustered specimens, and (3) morphological deformations. Laboratory validation confirms the model's effectiveness in overcoming traditional limitations, achieving 99% mAP and 0.987 recall—performance comparable to the best single-species detectors while using substantially fewer training images. The implications of this technology extend far beyond technical metrics. For IPM, automated monitoring addresses a fundamental challenge: implementing threshold-based decisions for nocturnal pests. With a recall approaching 96%, the system reliably detects adult vine weevil presence to facilitate the timely deployment of biological control agents like entomopathogenic nematodes. A continuous, quantitative data stream allows growers to evaluate control efficacy and adjust strategies based on evidence rather than calendar dates, as would be traditionally the case in pest management. Ecologically, the shift from prophylactic to targeted interventions preserves beneficial arthropod communities and pollinators typically impacted by broad-spectrum insecticides. This aligns with global sustainability goals, including pesticide reduction targets and biodiversity conservation mandates. The technology directly supports certification schemes that reward evidence-based IPM, as the digital audit trail provides verifiable proof of monitoring-based decisions. For growers, this translates to premium market access and compliance with increasingly stringent environmental regulations. Economic viability underpins practical adoption. While component costs of $200-300 per unit represent initial investment, returns manifest through multiple channels: reduced labor for night-time scouting, optimized biocontrol timing, decreased crop losses, and lower input costs from eliminating unnecessary treatments. For high-value protected horticulture, where vine weevil damage can devastate entire crops, the economic case is evident. The lightweight neural network's success with limited training data suggests that it could be rapidly adapted to other pest species, potentially revolutionizing monitoring across diverse agricultural systems. Integration with emerging technologies—acoustic sensors, pheromone detection, environmental modeling—points toward comprehensive surveillance networks providing both landscape-scale insights and field-level precision. By demonstrating that accurate, affordable monitoring of challenging nocturnal pests is achievable, this work catalyzes a fundamental shift in pest management capability.

CRediT authorship contribution statement

Hui Zhao: Writing – original draft, Visualization, Formal analysis. **Bo Fu:** Writing – original draft, Visualization, Formal analysis. **Fernando Auat Cheein:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis. **Matthew Butler:** Writing – review & editing, Funding acquisition, Formal analysis, Data curation. **W. Edwin Harris:** Writing – review & editing, Funding acquisition, Data curation, Conceptualization. **Tom W. Pope:** Writing – review & editing, Writing – original draft, Investigation, Funding acquisition. **Joe M. Roberts:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Funding acquisition, 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.

Acknowledgements

This work was funded through a joint BBSRC and AHDB Farm Sustainability Fund (Grant No.: 91140082).

Data availability

The data that has been used is confidential.

References

[1] T.W. Pope, J.M. Roberts, Vine weevil, otiorhynchus sulcatus (coleoptera: Curculionidae), management: current state and future perspectives, Annu. Rev. Entomol. 67 (2022) 221–238, https://doi.org/10.1146/annurev-ento-071221-060822.

[2] S. Gill, J. Lutz, P. Shrewsbury, M. Raupp, Evaluation of biological and chemical control methods for black vine weevil, otiorhynchus sulcatus (Fabricius) (coleoptera: Curculionidae), in container grown perennials, J. Environ. Hortic. 19 (2001) 166–170, https://doi.org/10.24266/0738-2898-19.3.166.

[3] W.H.M. Van Tol, D.J. Bruck, F.C. Griepink, W.J.D. Kogel, Field attraction of the vine weevil otiorhynchus sulcatus to kairomones, J. Econ. Entomol. 105 (2012) 1096–1103, https://doi.org/10.1603/EC11248.

[4] M.A. Ansari, F.A. Shah, T.M. Butt, The entomopathogenic nematode steinernema kraussei and metarhizium anisopliae work synergistically in controlling overwintering larvae of the black vine weevil, otiorhynchus sulcatus, in strawberry growbags, Biocontrol Sci. Technol. 20 (2009) 99–105, https://doi.org/10.1080/09583150903420031.

[5] S.K. Dara, The new integrated pest management paradigm for the modern age, J. Integr. Pest Manag. 10 (2019) 1–9, https://doi.org/10.1093/jipm/pmz010.

[6] S.Y. Li, S.M. Fitzpatrick, D.E. Henderson, Grooved board traps for monitoring the black vine weevil (coleoptera: Curculionidae) in raspberry fields, J. Entomol. Soc. B.C. 92 (1995) 97–100.

[7] E. Fezza, J.M. Roberts, T.J.A. Bruce, L.E. Walsh, M.T. Gaffney, T. Pope, Optimising vine weevil, otiorhynchus sulcatus f. (Coleoptera: Curculionidae), monitoring tool design, Insects 13 (2022) 80, https://doi.org/10.3390/insects13010080.

[8] R.W. Mankin, J.R. Fisher, Acoustic detection of black vine weevil, otiorhynchus sulcatus (Fabricius) (coleoptera: Curculionidae) larval infestations in nursery containers, J. Environ. Hortic. 20 (2002) 166–170, https://doi.org/10.24266/0738-2898-20.3.166.

[9] S.C. Gordon, J.A. Trefor Woodford, A. Grassi, M. Zini, T. Tuovinen, I. Lindqvist, J.W. McNicol, Monitoring and importance of wingless weevils (otiorhynchus spp.) in European red raspberry production, IOBC Bull. 26 (2003) 55–60.

[10] J.M. Roberts, A. Jahir, J. Graham, T.W. Pope, Catch me if you can: the influence of refuge / trap design, previous feeding experience, and semiochemical lures on vine weevil (Coleoptera: Curculionidae) monitoring success, Pest Manag. Sci. 76 (2020) 553–560, https://doi.org/10.1002/ps.5545.

[11] T.W. Pope, G. Hough, C. Arbona, H. Roberts, J. Bennison, J. Buxton, G. Prince, D. Chandler, Investigating the potential of an autodissemination system for managing populations of vine weevil, otiorhynchus sulcatus (Coleoptera: Curculionidae) with entomopathogenic fungi, J. Invertebr. Pathol. 154 (2018) 79–84, https://doi.org/10.1016/j.jip.2018.04.002.

[12] J.V. Cross, H. Hesketh, C.N. Jay, D.R. Hall, P.J. Innocenzi, D.I. Farman, C.M. Burgess, Exploiting the aggregation pheromone of strawberry blossom weevil anthonomus rubi Herbst (Coleoptera: Curculionidae): part 1. Development of lure and trap, Crop Prot. 25 (2006) 144–154, https://doi.org/10.1016/j.cropro.2005.04.002.

[13] W. Ding, G. Taylor, Automatic moth detection from trap images for pest management, Comput. Electron. Agric. 123 (2016) 17–28, https://doi.org/10.1016/j.compag.2016.02.003.

[14] Y. Li, H. Wang, L. Dang Minh, A. Sadeghi-Niaraki, M. H, Crop pest recognition in natural scenes using convolutional neural networks, Comput. Electron. Agric. 169 (2020) 105174, https://doi.org/10.1016/j.compag.2019.105174.

[15] A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, in: NIPS, 2012.

[16] J. Redmon, A. Farhadi, Yolov3: an incremental improvement, CoRR, arXiv:1804.02767, 2018, http://arxiv.org/abs/1804.02767.

[17] A. Bochkovskiy, C. Wang, H.M. Liao, Yolov4: optimal speed and accuracy of object detection, CoRR, arXiv:2004.10934, 2020, https://arxiv.org/abs/2004.10934.

[18] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016, IEEE Computer Society, 2016, pp. 770–778.

[19] X. Zhao, W. Li, Y. Zhang, T.A. Gulliver, S. Chang, Z. Feng, A faster rcnn-based pedestrian detection system, in: IEEE 84th Vehicular Technology Conference, VTC Fall 2016, Montreal, QC, Canada, September 18–21, 2016, IEEE, 2016, pp. 1–5.

[20] W. Li, D. Wang, M. Li, Y. Gao, J. Wu, X. Yang, Field detection of tiny pests from sticky trap images using deep learning in agricultural greenhouse, Comput. Electron. Agric. 183 (2021) 106048, https://doi.org/10.1016/j.compag.2021.106048.

[21] C. Wen, H. Chen, Z. Ma, T. Zhang, C. Yang, H. Su, H. Chen, Pest-yolo: a model for large-scale multi-class dense and tiny pest detection and counting, Front. Plant Sci. 13 (2022) 973985, https://doi.org/10.3389/fpls.2022.973985.

[22] T.T. Høye, J. Arje, O.L.P. Hansen, A. Isoifidis, F. Leese, H.M.R. Mann, K. Meissner, C. Melvad, J. Raitoharju, Deep learning and computer vision will transform entomology, Proc. Natl. Acad. Sci. 118 (2021) e2002545117, https://doi.org/10.1073/pnas.2002545117.