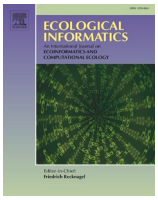
Ecological Informatics 82 (2024) 102744

Contents lists available at ScienceDirect

Ecological Informatics

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



Improving pollen-bearing honey bee detection from videos captured at hive entrance by combining deep learning and handling imbalance techniques

Dinh-Tu Nguyen [a], Thi-Nhung Le [a,f], Thi-Huong Phung [a], Duc-Manh Nguyen [a], Hong-Quan Nguyen [d], Hong-Thai Pham [c], Thi-Thu-Hong Phan [e], Hai Vu [a,b], Thi-Lan Le [a,b,*]

[a] School of Electrical and Electronic Engineering (SEEE), Hanoi University of Science and Technology, Hanoi, Viet Nam
[b] Computer Vision Department, MICA International Research Institute, Hanoi University of Science and Technology, Hanoi, Viet Nam
[c] Research Center for Tropical Bees and Beekeeping, Vietnam National University of Agriculture, Hanoi, Viet Nam
[d] Faculty of Information Technology, Viet-Hung Industrial University, Hanoi, Viet Nam
[e] Department of Artificial Intelligence, FPT University, Da Nang, Viet Nam
[f] Faculty of Information Technology, Vietnam National University of Agriculture, Hanoi, Viet Nam

ARTICLE INFO

Keywords:
Pollen foraging behavior
Pollen-bearing honey bee detection

ABSTRACT

The number of pollen-bearing honey bees serves as a vital indicator for assessing colony balance and health. Despite its significance, prevailing detection techniques still rely heavily on manual observation and annotation, leading to time-consuming processes that cannot sustain long-term, continuous monitoring efforts. To facilitate automatic beehive monitoring, this study introduces an efficient method for pollen-bearing bee detection. Initially, we furnish a comprehensive dataset, dubbed VnPollenBee, meticulously annotated for pollen-bearing honey bee detection and classification. The dataset comprises 60,826 annotated boxes that delineate both pollen-bearing and non-pollen-bearing bees in 2051 images captured at the entrances of beehives under various environmental conditions. To the best of our knowledge, this represents the first dedicated dataset for pollen-bearing bee detection. The VnPollenBee dataset is publicly accessible to the research community at https://comvis-hust.github.io/datasets/pollenbee.html. Subsequently, we propose the incorporation of diverse techniques into two baseline models, namely YOLOv5 and Faster RCNN, to effectively address the imbalance that arises during the detection of pollen-bearing bees due to their number being typically much lower than the total number of bees present at hive entrances. The experimental results demonstrate that our proposed method outperforms the baseline models on the VnPollenBee dataset, yielding Precision, Recall, and F1 score of 99%, 93%, and 95%, respectively. Specifically, the improvements obtained are 3% and 2% in Recall and F1 score when using YOLOv5, and 3%, 2%, and 2% in Precision, Recall, and F1 score when using Faster RCNN. These findings confirm the potential of our approach to facilitate bee foraging behavior analysis and automated bee monitoring.

1. Introduction

Honeybees play a crucial role in ecosystems as pollinators. However, they are highly sensitive to environmental factors such as temperature, lighting, pesticide residues, and the presence of alien species. Colony Collapse Disorder (CCD) represents one of the most severe threats to beehives. To safeguard bee colony health and prevent CCD, beekeepers must regularly monitor beehives, assessing indicators such as the queen's presence, colony size, hive products, and the presence of infestations and predators (Requier, 2019). Despite the importance of these monitoring efforts, widely adopted techniques still rely on manual observation and annotation, making continuous monitoring challenging.

In recent years, efforts have been made to automate bee colony monitoring by evaluating these indicators using various sensors, including visual information (Lee et al., 2023; Ngo et al., 2019; Ngo et al., 2021; Rodriguez et al., 2022; Voudiotis et al., 2022), temperature, relative humidity, beehive weight (Braga et al., 2020; Krishnasamy et al., 2023), and audio information (Kulyukin et al., 2018; Rustam et al., 2024; Truong et al., 2023; Zhao et al., 2021). Each type of data has its own advantages and drawbacks. Visual data, in particular, has gained traction in beehive monitoring due to the decreasing cost of cameras and

* Corresponding author at: School of Electrical and Electronic Engineering (SEEE), Hanoi University of Science and Technology, Hanoi, Viet Nam.
E-mail address: lan.lethi1@hust.edu.vn (T.-L. Le).

https://doi.org/10.1016/j.ecoinf.2024.102744
Received 24 October 2023; Received in revised form 23 July 2024; Accepted 27 July 2024
Available online 31 July 2024

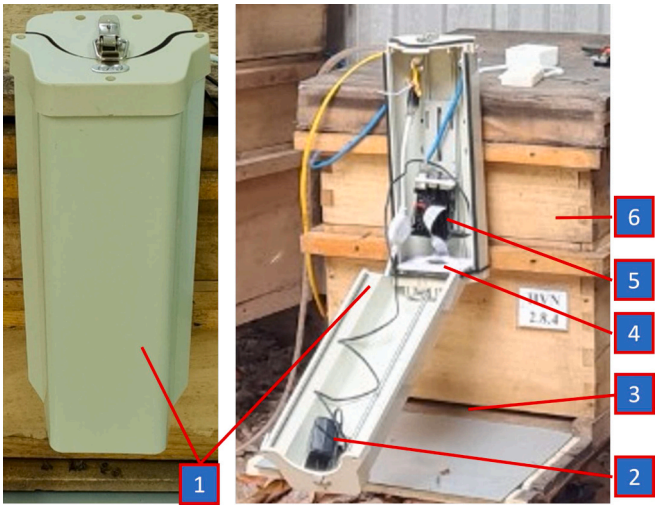


Fig. 1. The image acquisition system: 1 - outdoor surveillance camera weatherproof housing; 2 - power adapters; 3 - beehive entrance; 4 - IMX477HQ camera module with 6 mm CS-mount lens; 5 - Nvidia Jetson Nano; 6 - super beehive.

the successful deployment of various computer vision tasks. From visual images, it is possible to determine metrics such as the number of bees appearing at the beehive entrance (Nguyen et al., 2022; Nguyen et al., 2023), the number of incoming and outgoing bees (Fruet et al., 2023; Krishnasamy et al., 2023; Ngo et al., 2019; Sledević and Plonis, 2023), the ratio of pollen-bearing to non-pollen-bearing bees (Babic et al., 2016; Ngo et al., 2021; Rodriguez et al., 2018a; Rodriguez et al., 2022), the presence of the Varroa mites (Bilik et al., 2024; Yoo et al., 2023), and recognize important behaviors such as foraging, fanning, guarding (Sledević and Plonis, 2023) or dancing (Kongsilp et al., 2024). This study focuses specifically on detecting forager bees with pollen loads (i.e., pollen-bearing bees) returning to the colony, as the number of pollen-bearing bees serves as a crucial indicator of bee population dynamics and overall colony health status. Detecting pollen-bearing bees from images captured at hive entrances poses several challenges, including high bee density and varying lighting conditions. Additionally, the relatively small size of pollen loads in images and its occlusion by bees' bodies further complicate detection efforts. Therefore, most current techniques for automatic pollen-bearing bee detection rely on deep learning methods to address the aforementioned challenges. However, these deep learning approaches typically demand a substantial amount of annotated data. Preparing fully annotated and extensive datasets proves to be a costly and time-consuming endeavor. Furthermore, the number of forager bees with pollen loads is relatively small compared to the total number of bees at the hive entrance. Hence, diverse strategies must be employed to mitigate the imbalance issue in pollen-bearing bee detection. Our research focuses on creating a fully annotated dataset to detect pollen-bearing bees and addressing the issue of data imbalance to enhance the performance of detection models.

The contributions of this study are twofold:

- First, it introduces a dataset for pollen-bearing bee detection and classification named VnPollenBee, which has been meticulously collected and fully annotated. Comprising 2051 images with 60,826 annotated boxes for both pollen-bearing and non-pollen-bearing bees, the dataset captures bees at the entrance of honeybee colonies under various conditions. The VnPollenBee dataset accurately reflects the imbalance issue encountered in real-world applications. Notably, this dataset marks the first publicly available resource for pollen-bearing bee detection in the research community.
- Second, it proposes and integrates three different techniques into two baseline detection models, namely the You Only Look Once v5 (YOLOv5) algorithm and the Faster Region-based Convolutional Neural Network (Faster RCNN), to address the imbalance issue. Comparative analysis with the baseline models reveals that the second method leads to an enhancement in Recall from 85% to 88% and an increase in F1 score from 91% to 93%, whereas the third method achieves improvements of 3%, 2%, and 2% for Precision, Recall, and F1 score, respectively.

2. Related works

In the field of bee health monitoring, a variety of bee detection, tracking, and pollen-foraging behavior monitoring methods have been proposed. In this section, we briefly analyze the relevant works in the literature for bee detection and tracking, and mainly focus on pollen-bearing bees detection and classification.

2.1. Pollen bearing bees detection and classification

Bee detection is considered the first and crucial step in automatic bee counting and behavior monitoring. In (Odemer, 2022), the authors have provided a systematic review of approaches for automated bee detection and counting. The study pointed out that video-based methods have emerged recently thanks to the availability of low-cost image-capturing devices. Within the last decade, different methods have been proposed for honeybee detection and counting. While traditional approaches relied mainly on background subtraction for bee detection or hand-designed features with sliding windows, recent methods were based on different architectures of CNN networks. In (Babic et al., 2016), the authors evaluated different background subtraction methods and suggested employing the Mixture of Gaussian (MG) for honey bee detection thanks to the robustness of this model under different lighting conditions. In (Ngo et al., 2019), the authors first applied background subtraction and then applied morphological operations as a post-processing step to extract the honey bee region from images. The work presented in (Dembski and Szymański, 2019) tried to find out the best color space for honey bee representation and detection. In (Rodriguez et al., 2018b), the authors employed the Part Affinity Fields (PAF) approach that is proposed for human pose estimation to detect the different parts of the

Fig. 2. Frame sample captured by the image acquisition system.

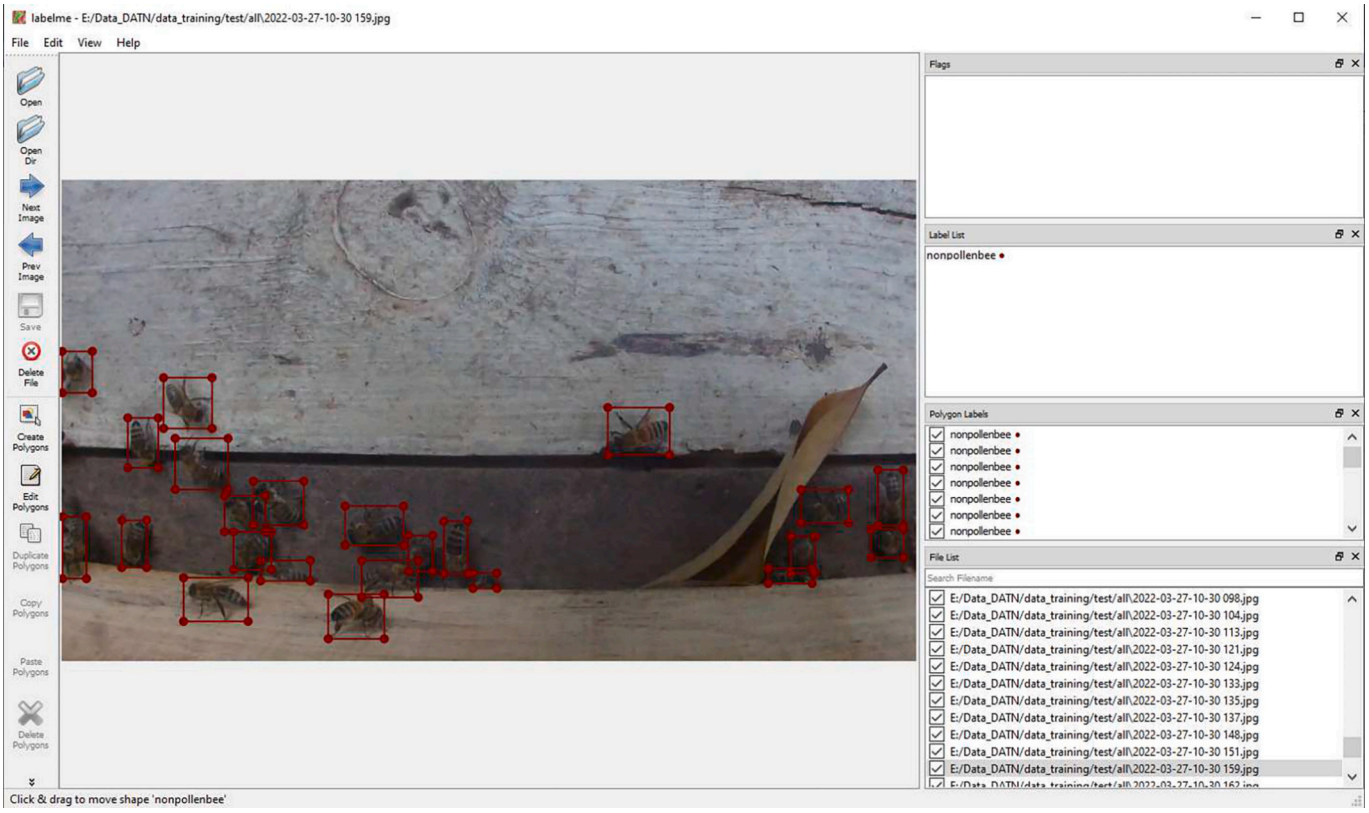


Fig. 3. Example image of the data labeling process.

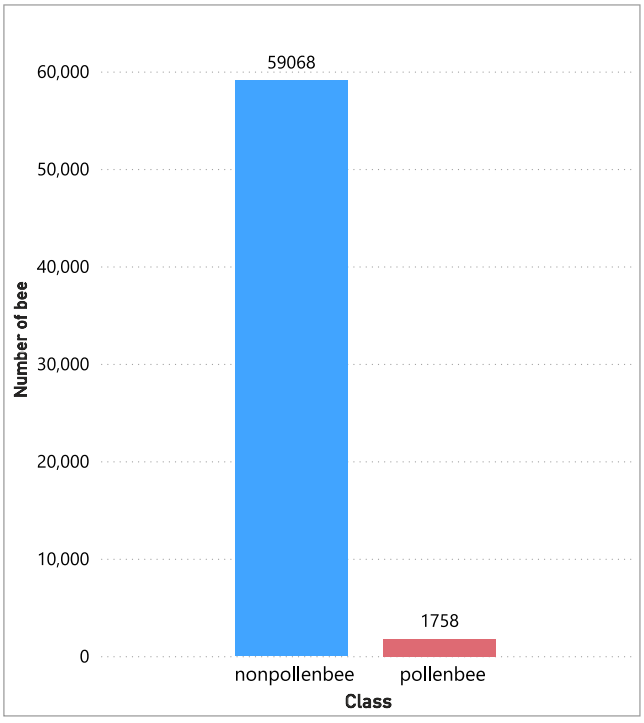


Fig. 4. Number of samples in each class. Nonpollenbee: bees not carrying pollen. Pollenbee: bees carrying pollen.

honeybees. The structural model used for the honeybees in this paper contains 5 parts including the tip of the abdomen, thorax, tip of the head, and tips of the left and right antennas. The proposed method can simultaneously detect and estimate the pose of honey bees. In (Knauer et al., 2022), the authors introduced an open-source software named Bee Tracker for the assessment of the nesting and foraging performance of cavity-nesting solitary bees. However, this software requires a complex setup and can not be integrated into real bee hives. In (Nguyen et al., 2022), a robust bee detection and counting was proposed. First, the YOLO neural network was employed to predict the bee's position on images. However, YOLO is not robust enough in the case of occlusions due to the high density of bees' presence. Therefore, the authors proposed to utilize a kernel-based density estimator for each local region and then count the number of bees in high-density areas by FAMNET (Few Shot Adaptation and Matching Network). In (Nguyen et al., 2023), the authors aimed to improve the performance of bee detection and counting under adverse conditions and noisy labeled data. They utilized the VGG-19 model with a modified backbone to create a density map using Bayesian loss. Experimental results demonstrated high accuracy, with a Mean Absolute Error (MAE) of 3.61 and a Mean Squared Error (MSE) of 4.81. In (Sledević and Plonis, 2023), the authors first used YOLOv8m to detect bees on the landing board, then employed Byte-Track to track them and used heat maps to represent behaviors such as foraging, guarding, and fanning. The proposed bee detection model achieved a mean average precision (mAP@0.5) of 97% and a mAP@0.5:0.95 of 65%. In (Fruet et al., 2023), the authors introduced the ApisFlow system designed to detect worker bees, drones, and potential threats, while also counting incoming and outgoing bees. The system employed YOLOv8 for bee detection, and utilized the Kalman Filter and Hungarian algorithm for bee tracking. To count bees, the system assessed the initial and current positions of each bee to determine if they were within a predefined entrance box. In recent research (Kongsilp et al., 2024), Mask R-CNN with a ResNet-101 backbone was used to detect and segment individual bees, while a Kalman filter was used for tracking. Experiments demonstrated that the proposed bee detection and segmentation model achieved a mean average precision (mAP) of 85%, and the bee tracking model reached 77.48% MOTA and

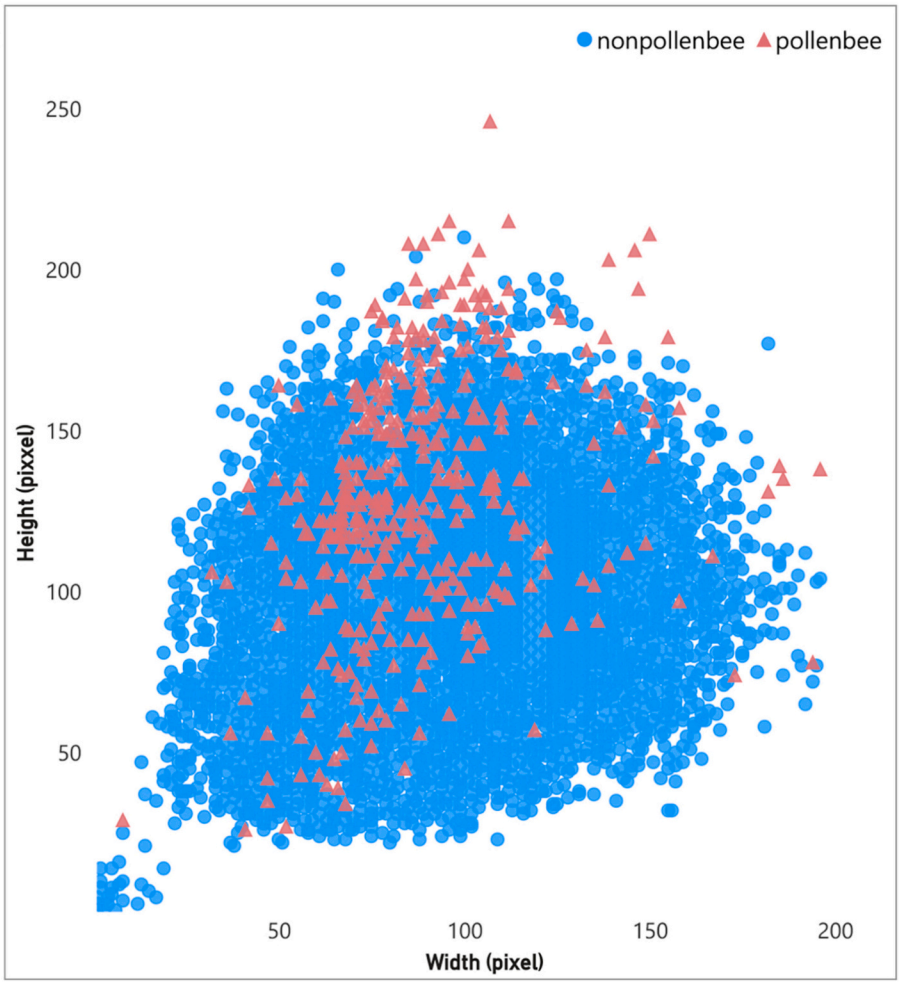


Fig. 5. The distribution of the length and width of the bounding boxes in VnPollenBee dataset.

79.79% MOTP metrics.

Besides the number of returning honey bees, the number of pollen-bearing bees is also an important indicator that the beekeepers rely on to monitor the bee colony's health.

Pollen-bearing bees detection and recognition methods can be divided into two main aspects: (1) sensor-based approaches and (2) image-based approaches.

The sensor-based methods employ some specific physical sensors to detect pollen. In (Kalman et al., 1997), the authors designed an electronic nose, which is a gas sensor array combined with a routine for pattern recognition, together with a pyrolyzing unit that can be used to distinguish pollen from other particles. The main idea is to use different samples together with the pollen that are heated to a high temperature and then the released gases are analyzed by an electronic nose. Finally, the obtained data will be processed with principal component analysis (PCA) and with an artificial neural network (ANN). However, the proposed solution is complex, and the obtained results are very limited.

The image-based methods employ the outputs of bee detection and tracking for pollen-bearing bee detection and classification. The authors in (Babic et al., 2016) used traditional algorithms of image processing such as background subtraction, color segmentation, and morphology to segment honey bees. Then, they used the nearest mean classifier with a simple descriptor consisting of color variance and eccentricity features to classify honey bees into two classes: pollen-bearing honey bees and non-pollen-bearing honey bees. The proposed method can be executed in real-time on an embedded system and achieved 88.7% of the correct classification rate. However, this method is sensitive to the background and brightness changes.

Machine learning algorithms are widely used in object classification tasks in general as well as in continuous beehive monitoring tasks in particular, typically detecting pollen-bearing bees (Bilik et al., 2024). Algorithms like K nearest neighbor, Naive Bayes, and Support Vector Machine (SVM) are used by (Rodriguez et al., 2018a) in classifying bees that carry pollen and bees that do not. The best accuracy was achieved by using SVM with both linear and Radial Basis Functions (RBF).

In (Vladan Stojnic, 2018), instead of using direct SVM to each image, they used image descriptors as input of the SVM algorithm. SIFT (Scale Invariant Feature Transform) and VLAD (Vector of Locally Aggregated Descriptors) were applied to compute image descriptors. These image descriptors or, in other words, image representations were put through SVM to get a function that can map an image representation to a class label.

Deep learning has grown strongly in the last few years, making computer vision tasks more precise. By using a deep neural network, classification and detection become more accurate and faster. Convolutional neural network (CNN) is a very famous neural network in the field of computer vision. It uses convolution to create features of the image. For instance, (Rodriguez et al., 2018a; Sledevic, 2018) aim to classify pollen-bearing bees/non-pollen-bearing bees. In (Rodriguez et al., 2018a), they used two types of CNN: shallow CNNs and deep CNNs. The shallow CNNs constructed on the basic module: 2D convolution, Relu activation, and Max-pooling. Their complete architecture was built as a sequence of one or two basic modules. The deep CNNs they used are VGG16, VGG19 and ResNet50. The results have shown that the

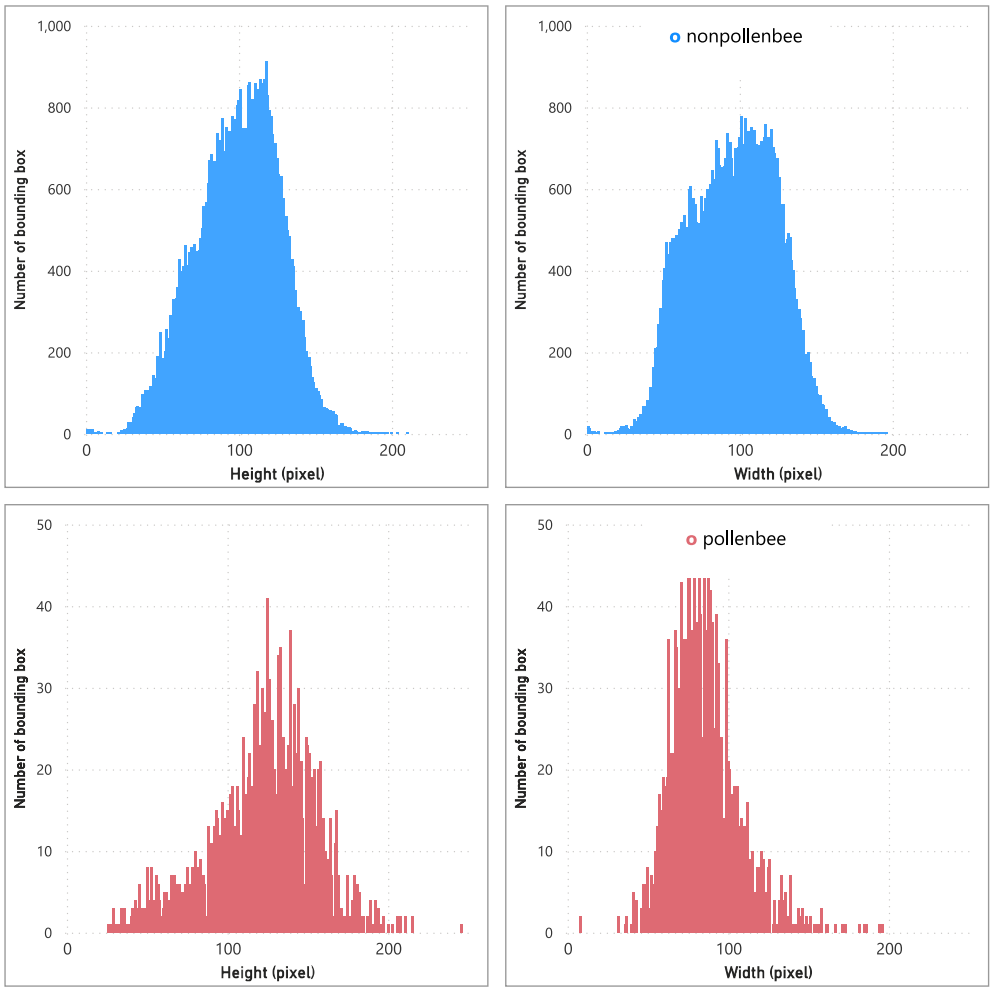


Fig. 6. The distribution of the length and width of the bounding boxes in VnPollenBee dataset.

shallow CNNs allow for getting better results than deep CNNs. The authors in (Rodriguez et al., 2022), first used an automatic centering and orientation method to crop images of individual bees. These images were then used as input to a two-layer shallow network to classify whether the bees were carrying pollen. This method achieved image classification results with a precision of 97.26%, recall of 97.28%, and an F1 score of 97.3%. In (Sledevic, 2018), they also used CNN architecture but with different filter sizes. The obtained results depend on the number of hidden layers, filter size, and the number of filters in convolution layers. The higher the number of hidden layers, filter size, and filters, the higher the accuracy of classification. In (Berkaya et al., 2021), the authors proposed four different models for bee image classification, serving different tasks in beehive monitoring. All these models were based on different Deep Neural Networks (DNNs) that have been pre-trained on the ImageNet dataset (Deng et al., 2009) including AlexNet, DenseNet-201, GoogLeNet, ResNet-101, ResNet-18, VGG-16, VGG-19. In the first model, the authors used a transfer learning method based on these pre-trained DNNs. Meanwhile, in the second, third, and fourth models, the authors used the SVM classifier with deep features, shallow features, and deep+shallow features extracted from the pre-trained DNNs, respectively. For the task of classifying images of pollen-bearing and non-pollen-bearing honeybees, experimental results on the Pollen dataset (Rodriguez et al., 2018) have shown that the proposed models have quite high accuracy. In particular, the model using the transfer learning method with the pre-trained GoogLeNet gave the highest accuracy at 99.07%. In (Le et al., 2023), the authors proposed a new CNN architecture consisting of four convolutional layers, five max-pooling layers, one flatten layer, and one dense layer to classify images in the corrected Pollen dataset. Although the proposed architecture was quite compact, the experimental accuracy was up to 100%. However, research (Berkaya et al., 2021; Le et al., 2023) only stopped at classifying images of pollen-bearing and non-pollen-bearing honeybees, and did not allow detecting and counting the number of pollen-bearing honeybees from videos. In addition, the dataset used for training and testing these models was quite small, including only 714 images, so the robustness of these models should be further verified on other larger datasets.

In (Yang and Collins, 2019), they used the method that combined background subtraction and color thresholding to detect bees and a Faster RCNN network for detecting pollen from a single bee that was detected. Each pollen represented in the image had a bounding box on it. The results from the paper were good, but the time to execute the algorithm was low. And the paper did not mention how to measure pollen weight.

Recently, in (Ngo et al., 2021), the authors formulated the pollen-bearing and non-pollen-bearing bee classification as a two-classes detection problem and then trained a YOLOv3-tiny model to detect pollen-bearing and non-pollen-bearing bees at the hive's entrance. Then Kalman filter and Hungarian algorithm were applied to track these bees and count the incoming and outgoing activity. This method could not solve the problem of unbalance in pollen-bearing bee detection.

In (Narcia-Macias et al., 2023), to capture the overall health status of the beehive, the authors designed a real-time honeybee monitoring system. Accordingly, the YOLOv7-tiny architecture was chosen to be

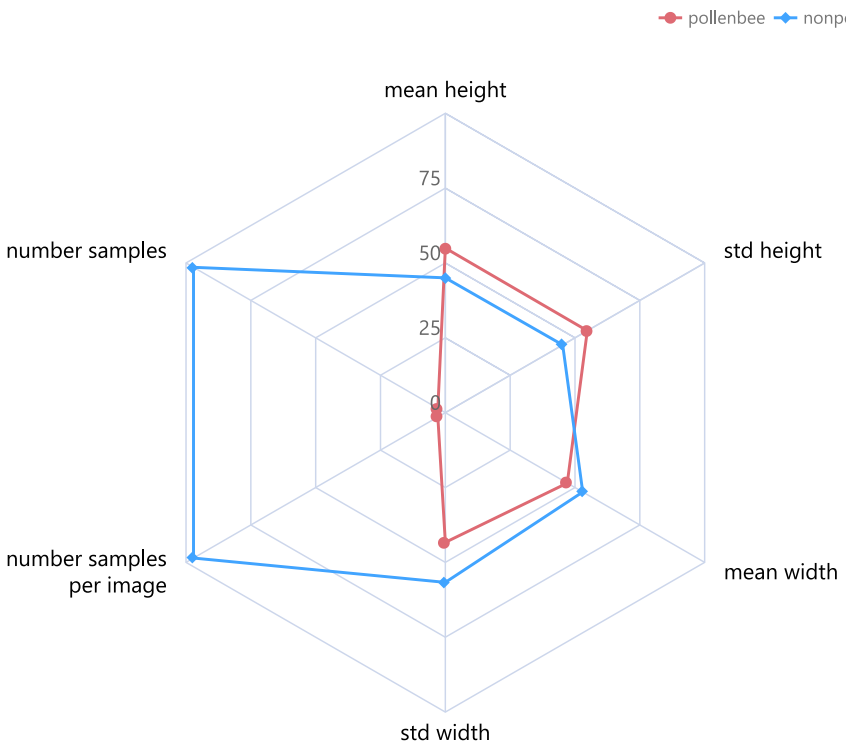


Fig. 7. Descriptive statistics of the data.

Fig. 8. Diversity of pollen-bearing bees in the dataset.

used to detect honeybees, pollen, and Varroa mites. The proposed model gave the F1 score of 0.95. In another study (da Silva et al., 2023), researchers experimented with various methods to detect pollen-bearing bees, such as Cascade R-CNN, Faster R-CNN, Deformable DETR, CenterNet, YOLOX, and YOLOv7. The results indicated that YOLOv7 achieved the highest detection accuracy, with an AP@0.75 of 77%. However, the study did not address the issue of imbalance in the dataset, resulting in overall lower detection performance. BeeNet model for bee colony monitoring was introduced in (Yoo et al., 2023). To identify pollen-bearing bees from images, a modified ResNet50 was used for feature extraction, and a transformer encoder was used for classification. However, the dataset used in this experiment had a limited number of images and did not exhibit an imbalance between pollen-bearing bee images and non-pollen-bearing bee images.

2.2. *Datasets for pollen-bearing bee detection*

To evaluate the bee detection methods in general and pollen-bearing bee detection in particular. Some datasets are collected and annotated for different tasks such as bee detection, tracking, and pollen-bearing bee classification.

The dataset in (Hickert, n.d.) was a result of a project that aims to

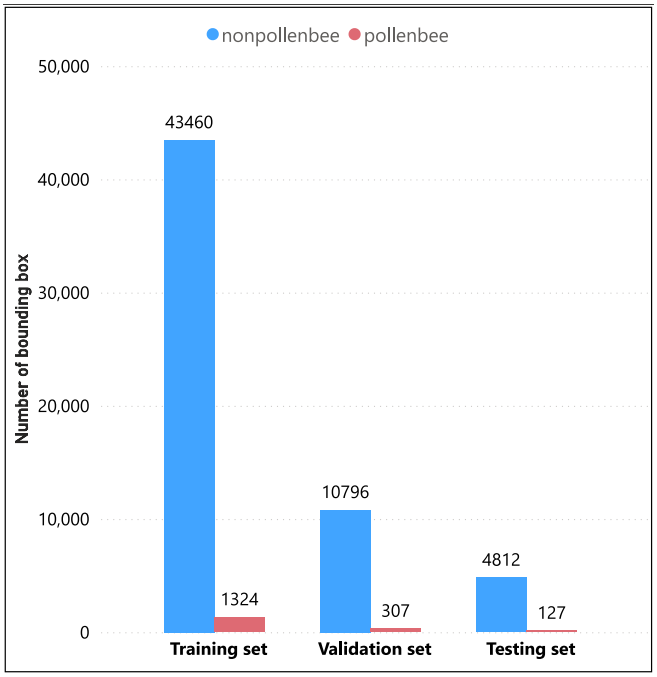


Fig. 9. The number of samples for each class in training, validation, and testing sets.

develop a camera-based bee monitoring system. The dataset contains approximately 7500 images of bees, captured at the entrance of a bee hive from above. Each image contains one bee, rotated to a vertical format with heads or tails up. All images were taken with a green background and the distance to the bees was always the same, thus all bees were the same size. The dataset was designed for multi-class classification including cooling bees, pollen-bearing bees, varroas, and wasps. The cooling bees indicate that the bee is currently cooling the hive. The bee flaps its wings while keeping its position stationary, thereby transporting fresh air into the hive. The pollen-bearing bees indicate that the bee carries a pollen packet. The varroas indicate that the bee is infested with the varroas mite. The varroa mite is a small circular pest that is 1-2 mm in diameter and of brown color. If the mite is carried into the hive, it can seriously damage the bee larva. If left untreated, the whole colony dies. The wasps indicate the species differ from bees.

The dataset in (Yang, 2018) contains 5100+ bee images annotated with location, date, time, subspecies, health condition, caste, and pollen. The original batch of images was extracted from still time-lapse videos of bees. By averaging the frames to calculate a background image, each frame of the video was subtracted against that background to bring out the bees in the forefront. The bees were then cropped out of the frame so that each image had only one bee. Because each video was accompanied by a form with information about the bees and the beehive, the labeling process was semi-automated. Each video resulted in differing image crop quality levels.

Another dataset has been introduced in (Schurischuster and Kampel, 2020). The dataset contains short videos with mite-infected and uninfected (i.e. healthy) bees captured from a camera facing the entrance of a beehive. From these videos, more than 13,000 images of mite-infected and healthy bees were manually labeled.

Recently, in (Bilik et al., 2021), a dataset that contains a total of 803 unique samples, where 500 samples capture bees in the general

Table 1
Main parameters of the classification network.

Layers	Input size	Output size	Parameters
Convolution	300 × 180 × 3	298 × 178 × 8	[3 × 3 conv, strides 1] × 8
Convolution	298 × 178 × 8	296 × 176 × 16	[3 × 3 conv, strides 1] × 16
Pooling	296 × 176 × 16	148 × 88 × 16	2 × 2 max pool
Convolution	148 × 88 × 16	146 × 86 × 32	[3 × 3 conv, strides 1] × 32
Convolution	146 × 86 × 32	144 × 84 × 64	[3 × 3 conv, strides 1] × 64
Dropout	144 × 84 × 64	144 × 84 × 64	Probability 0.5
Pooling	144 × 84 × 64	72 × 42 × 64	2 × 2 max pool
Convolution	72 × 42 × 64	70 × 40 × 128	[3 × 3 conv, strides 1] × 128
Global Average Pooling	70 × 40 × 128	128	–
Dense	128	128	–
Dense	128	128	–
Classification layer (Dense)	128	1	–

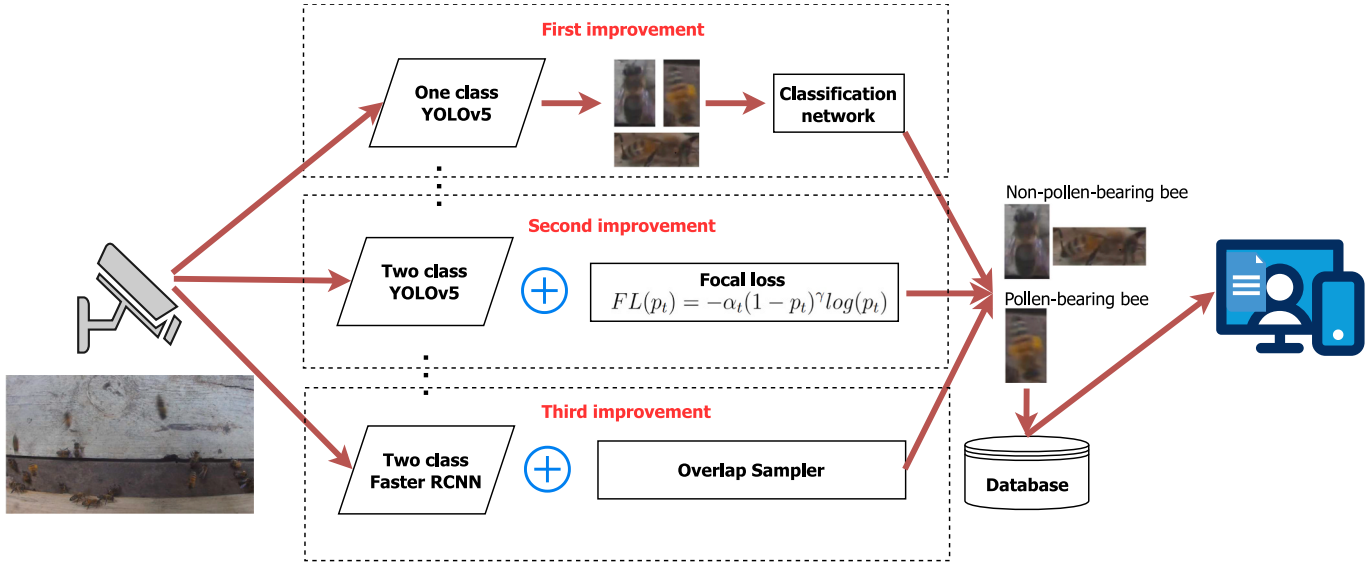


Fig. 10. Overview of our improvements for pollen-bearing bee detection.

Fig. 11. Examples of images in PollenDataset (Rodriguez et al., 2018a): (a) non-pollen-bearing and (b) pollen-bearing bees.

Table 2
Accuracy obtained for each fold and mean accuracy for all folds.

Dataset	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy	97.9	99.3	97.9	99.3	97.8
Mean			98.4		

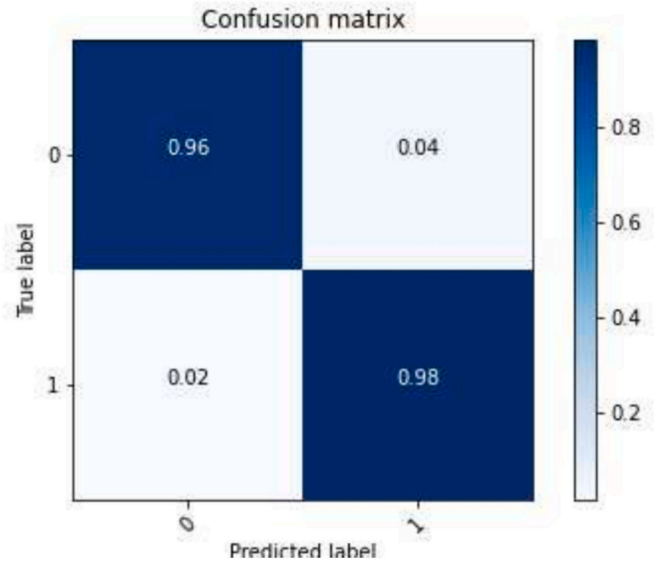


Fig. 12. Normalized confusion matrix obtained with the proposed classification method on the validation set from PollenDataset.

Table 3
Comparison between the proposed and state-of-the-art classification methods on PollenDataset. The highest accuracy is shown in bold font, while the second-highest values are shown in underline font.

Methods	Architecture	Accuracy
Method in (Rodriguez et al., 2018a)	1-Layer	<u>96.4</u>
Method in (Rodriguez et al., 2018a)	VGG16	87.2
Method in (Rodriguez et al., 2018a)	VGG19	90.2
Method in (Sledevic, 2018)	2-Layer	<u>96.4</u>
The proposed classification method	5-Layer	**98.4**

environment and the rest, taken from another dataset has been created. Six classes (1) the healthy bees, (2) bees with pollen, (3) drones, (4) queen, (5) V.-mite-infected bees (5), and (6) V.-mites. To evaluate the proposed method, the authors have also defined three different subsets: The first subset contains two classes: one is the bees (classes 1, 2, 3, 4, 5) and the other is V.-mite (class 6) while in the second subset: one is the healthy bees (classes 1, 2, 3, 4) and the other is infected bees (class 5). The last subset contains only images of V.-mites (only class 6).

With the same objective as our work, the authors in (Rodriguez et al., 2018a) focused on pollen-bearing bees classification. To evaluate the proposed method, the authors have prepared a dataset consisting of 714 image samples captured at the entrance of a bee colony in June 2017 at the Bee facility of the Gurabo Agricultural Experimental Station of the University of Puerto Rico. However, the dataset is captured in highly constrained conditions with high-quality images. Moreover, the dataset was balanced, containing 369 images of pollen-bearing bees and 345 images of bees without pollen loads. Therefore, it could not fully reflect the challenges of pollen-bearing bee detection and classification when applied in real conditions. This motivates us to collect and build a large and fully annotated dataset dedicated to pollen-bearing bee detection and classification.

3. VnPollenBee collection and annotation

This section introduces our image acquisition system, data collection, processing, labeling, and data evaluation protocol for the pollen-bearing bee detection task.

3.1. Image acquisition system and data collection

To acquire bee images for monitoring and assessing the health of honeybee colonies, we have established an acquisition system depicted in Fig. 1. This system is affixed to a single stage of the hive body, with all equipment housed in a weatherproof surveillance box. Utilizing an IMX477 HQ camera module with a 6 mm CS-Mount lens and an NVIDIA Jetson Nano developer kit, the camera is oriented in a downward-facing position. The IMX477 HQ camera module, equipped with a 6 mm focal length CS-mount lens (65° FoV), can effectively cover the entire width of the hive entrance. Operating at 1920 × 1080 resolution, the camera captures color video frames at 60 frames per second (fps). An example frame captured by the system is illustrated in Fig. 2. Notably, artificial lighting is not employed, and bees are not directed to a large landing platform, as such interventions may influence honeybee behavior.

For data collection, we utilize the OpenCV library and Cron scheduler to automate the scheduling of data collection. Data is continuously collected over several days, commencing at 5:00 a.m., coinciding with the start of the bees' activity. Each day, a one-minute-long video is captured every 30 min from 5:00 a.m. to 6:00 p.m. This approach enables us to obtain images at various times, accounting for differing lighting conditions.

It is worth noting that we use the same acquisition system as presented in (Nguyen et al., 2023). However, the work in (Nguyen et al., 2023) focused on bee counting and bee density estimation. Therefore, images at low, medium, and high bee densities were selected. In this study, with the main aim of counting pollen-bearing bees, we selected images that contain pollen-bearing bees to build the VnPollenBee dataset.

As depicted in Fig. 2, various challenges must be addressed during data collection: (1) lighting conditions exhibit significant variations throughout the observation period; (2) the presence of shadows and extraneous objects (e.g., leaves) complicates image analysis; (3) the size of bees fluctuates as they move closer to or farther from the camera, potentially resulting in blurred images, especially when bees fly at high speeds. Throughout the data collection process, we observed that pollen-bearing bees appear relatively infrequently in images, with typically only one to two pollen-bearing bees scattered across each frame. These

Fig. 13. GradCAM visualization of the classification network decision.

Table 4
Comparison of experimental results between the proposed and baseline models on the VnPollenBee dataset. The best metrics values are shown in bold font.

Method	Evaluation metrics				
	MR	FAR	Precision	Recall	F1-score
Baseline methods					
Yolov5 (Jocher et al., 2022)	0.15	0.009	**0.99**	0.85	0.91
Faster RCNN (Ren et al., 2015)	0.086	0.03	0.96	0.91	0.93
Our proposed methods					
Yolov5 + classification	0.11	0.41	0.58	0.88	0.70
Yolov5 + focal loss	0.12	**0.004**	**0.99**	0.88	0.93
Faster RCNN + Overlap Sampler	**0.07**	0.01	**0.99**	**0.93**	**0.95**

bees are most commonly observed in videos recorded during the morning hours, between 8:30 a.m. and 11:00 a.m. Following data acquisition spanning various periods from March to July 2022, we compiled an image dataset comprising 2051 images.

3.2. Data labeling

The labeling process for obtaining the ground truth of bounding boxes for both pollen-bearing and non-pollen-bearing bees was conducted in two steps. Initially, 400 images were manually annotated using the Labelme Annotation Tool. Each bee in the images was annotated with a rectangle encompassing it from head to tail, categorized as "nonpollenbee" if it lacked pollen and "pollenbee" if it bore pollen (refer to Fig. 3). Upon annotation, a file with the extension ".json" was generated, containing information regarding each image (image name, size) and the coordinates of each bounding box.

Subsequently, an object detection network (YOLOv5) was trained on the manually annotated subset and then applied for detection on the remaining images. However, due to the small and unbalanced nature of the dataset, the detection of pollen-bearing bees was often missed. Consequently, the results underwent manual verification to establish the ground truth bounding boxes. This iterative process yielded a dataset comprising 2051 images, encompassing 1758 pollen-bearing bees and 59,068 non-pollen-bearing bees.

3.3. Characteristic of VnPollenBee datasets

Fig. 4 displays the distribution of bees across different classes, whereas Fig. 5 illustrates the size distribution of the bees' bounding boxes. It is evident that a significant imbalance exists, with the number of non-pollen-bearing bees totaling approximately 59,068, which is approximately 33 times more than the number of pollen-bearing bees (1758). This highlights the necessity for appropriate imbalance techniques to be explored.

Examining Figs. 5 and 6, we observe that the width of the bounding box in the pollen-bearing bee class typically falls within the range of 50 to 110 pixels (equivalent to approximately 0.04 times the width of the image), whereas the height typically ranges from 100 to 170 pixels (equivalent to approximately 0.125 times the height of the image). Conversely, in the non-pollen-bearing bee class, the width predominantly ranges from 25 to 175 pixels (approximately 0.052 times the

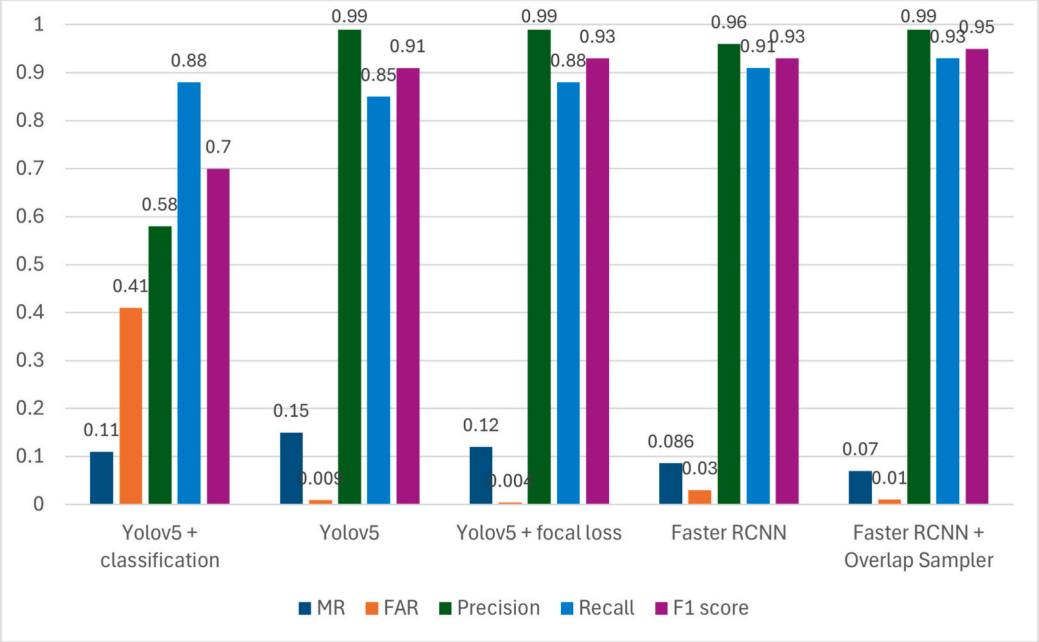


Fig. 14. Evaluation metrics of the pollen-bearing bee detection results.

width of the image), whereas the height is typically concentrated within the range of 25 to 160 pixels (approximately 0.08 times the image height). Fig. 7 presents descriptive statistics of the data. Some examples of pollen-bearing bees are illustrated in Fig. 8.

3.4. Evaluation metric and protocol

The VnPollenBee dataset was partitioned into three sets: a training set, a validation set, and a test set, at a ratio of 0.7:0.2:0.1. To ensure fairness across the datasets, encompassing all classes, and considering the considerably smaller number of pollen-bearing bees compared to non-pollen-bearing bees, the data was divided based on the proportion of pollen-bearing bees. Consequently, the training set comprised 1496 images, the validation set included 381 images, and the test set contained 174 images. The distribution of data among classes within each dataset is illustrated in Fig. 9.

To assess the efficacy of pollen-bearing bee detection methods, we employ common evaluation metrics for object detection tasks, namely Precision and Recall Both Precision and Recall are non-negative numbers less than or equal to one, calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

High precision indicates a high accuracy of predictions, whereas high recall signifies a low rate of missing positive samples. Additionally, to evaluate the method's capability in predicting misses and false positives, we incorporate two additional measures: Miss Rate (MR) and False Alarm Rate (FAR). These metrics are computed using the following formulas:

$$MR = \frac{FN}{FN + TP} \tag{3}$$

$$FAR = \frac{FP}{TP + FP} \tag{4}$$

where TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative respectively.

In addition, we use the F1 score to be able to gauge the balance between Precision and Recall. The F1 score is determined as follows:

$$\frac{2}{F1 - score} = \frac{1}{Precision} + \frac{1}{Recall} \tag{5}$$

4. Proposed methods for pollen-bearing bee detection and classification

4.1. Baseline models

We have selected two widely recognized object detection models, YOLOv5 (Jocher et al., 2022) and Faster RCNN (Ren et al., 2015), as baseline methods for pollen-bearing bee detection. YOLOv5 is a one-stage object detection network known for its rapid processing speed, although its accuracy is relatively lower compared to two-stage object detection networks. Consequently, we have opted for Faster RCNN as the second baseline method due to its higher detection accuracy.

YOLOv5, initially developed by Glenn Jocher, continues to be developed but has already made significant contributions within the YOLO family. It places a strong emphasis on speed and user accessibility for industrial applications. YOLOv5 aims to democratize artificial intelligence, simplifying the neural network training process for users. One notable improvement introduced by YOLOv5 addresses the challenge that users face when training YOLO on their datasets regarding the anchor boxes, which are traditionally trained on the COCO dataset. To mitigate this, YOLOv5 employs a genetic algorithm (GE) that adjusts anchor boxes after an initial k-means identification, thus enabling the selection of anchor boxes tailored to the dataset in use. Additionally, YOLOv5 incorporates a novel module called C3, devised by Glenn, that serves as a replacement for CSP, delivering comparable performance with enhanced speed. Moreover, YOLOv5 introduces a new SPP module named SPP Fast (SPPF), designed to reduce the FLOPS and increase the SPP speed. Furthermore, YOLOv5 adopts the SiLU activation function, unlike YOLOv4's, which utilizes Mish.

Faster RCNN belongs to the RCNN family of object detection networks. Essentially, it comprises two modules: the first module utilizes CNN to propose regions, whereas the second module employs a Fast RCNN model to process these suggested regions. The Region Proposal Network (RPN) represents a significant enhancement to the Faster

Fig. 15. Detection results of the proposed improvement: (above) YOLOv5 + focal loss, (below) Faster RCNN + Overlap sampler. The green and red boxes indicate pollen-bearing and non-pollen-bearing bees, respectively. Values indicate the confidence score. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

RCNN architecture, distinguishing it from other networks within the RCNN family. The RPN takes an input image of any size and outputs a region proposal, consisting of a set of rectangle locations capable of containing objects, along with the corresponding probability of containing the object. An essential aspect of Faster RCNN is the utilization of anchors for generating bounding boxes within the network. Anchors are created to differentiate between positive and negative anchors based on overlap. Furthermore, by comparing the position of predefined anchors with ground-truth bounding boxes (using Intersection over Union rate), the network can predict the location of the output region proposal. The second module of the network operates similarly to the structure of the Faster RCNN network, responsible for both object classification and localization. This module utilizes feature regions generated by the RPN to conduct classification and refine the coordinates of proposed regions based on ground truth bounding boxes.

4.2. Our proposed methods

When applying two baseline models for pollen-bearing bee detection, the small size of the pollen sac presents a challenge, as the external appearance of pollen-bearing and non-pollen-bearing bees is similar. Consequently, deep learning networks may encounter difficulty in distinguishing between these two objects. Furthermore, upon observing the image data collected from bee hives, it becomes apparent that there is an imbalance in the number of pollen-bearing bees compared to non-pollen-bearing bees. In light of these challenges, this study introduces three improvements to the two baseline models, namely YOLOv5 and Faster RCNN. Our first improvement involves initially detecting bees from images, wherein the bee class encompasses both pollen-bearing and non-pollen-bearing bees. Subsequently, we conduct classification to distinguish between pollen-bearing and non-pollen-bearing bees using a convolutional neural network. The second and third improvements are specifically designed whereby the Focal Loss function will be used in the YOLOv5 model, and the Overlap Sampler will be combined

Fig. 16. Results of detecting pollen-bearing bees and non-pollen-bearing bees using the Faster RCNN + Overlap sampler model under unfavorable conditions.

Table 5
Comparison of experimental results between the proposed method and recent state-of-the-art models on the VnPollenBee dataset. The best metrics values are shown in bold font, while the second-highest values are shown in underline font.

Method	Evaluation metrics				
	MR	FAR	Precision	Recall	F1-score
Other state-of-the-art methods					
EfficientDet (Tan et al., 2020)	0.078	0.13	0.87	0.92	0.89
RetinaNet (Lin et al., 2017)	0.078	0.13	0.87	0.92	0.89
DETR (Carion et al., 2020)	**0.062**	0.04	0.95	**0.93**	0.94
Our proposed method					
Faster RCNN + Overlap Sampler	0.07	**0.01**	**0.99**	**0.93**	**0.95**

Table 6
Comparison of the number of parameters, GFLOPs, and inference time of models in our experiments.

Method	Evaluation metrics		
	Parameters (M)	GFLOPs	Time (ms)
Baseline methods			
Yolov5 (Jocher et al., 2022)	86.18	203.8	52.9
Faster RCNN (Ren et al., 2015)	41.20	446.7	60.0
Other state-of-the-art methods			
EfficientDet (Tan et al., 2020)	6.55	2.8	322.0
RetinaNet (Lin et al., 2017)	37.96	47.8	698.0
DETR (Carion et al., 2020)	41.30	97.1	79.4
Our proposed methods			
Yolov5 + classification	86.35	204.2	22.0
Yolov5 + focal loss	86.18	203.8	52.3
Faster RCNN + Overlap Sampler	41.30	1097.9	110.0

with Faster RCNN to address the data imbalance issue. Fig. 10 illustrates our improvements. Accordingly, images that are captured from a camera installed at the entrance of the beehive are fed into the detection models. As a result, pollen-bearing and non-pollen-bearing bees on each image will be detected. The labels and locations of these bees are then stored in a database. Finally, a web page is designed to show the detection results. (See Fig. 11.)

4.3. First improvement: One class YOLOv5 + classification network

The first method comprises two steps. Initially, YOLOv5 is applied to detect bees in the image. Subsequently, in the second step, a classification network is developed to categorize the detected bees into either pollen-bearing or non-pollen-bearing bee classes. The architecture of the classification network draws inspiration from the AlexNet network, adhering to the principle of incorporating one pooling layer for every two convolutional layers. The primary parameters of the classification network are outlined in Table 1. Five convolutional layers with a filter size of 3 × 3 are employed to identify features within the image. Utilizing a 3 × 3 filter aids in reducing network parameters, thereby minimizing irrelevant features and preserving fine details. To decrease the computational cost, a max-pooling layer is incorporated to downsize the output feature images while retaining crucial features. The Dropout layer is integrated to mitigate overfitting. Additionally, the Global Average Pooling layer summarizes features, generating input for the fully connected layer. This layer also facilitates the visualization of regions relied upon by the network for predictions. The subsequent three fully connected layers generate the predicted class for the original image.

4.4. Second improvement: Two class YOLOv5 + focal loss

In this enhancement, an object detection network is trained to detect two classes: pollen-bearing and non-pollen-bearing bees. However, due to data imbalance, where one class significantly outnumbers the other, we propose the implementation of a focal loss function to address this issue. Focal loss (Lin et al., 2017) is a technique designed to mitigate imbalance problems in object detection networks by assigning higher weights to samples that are challenging or prone to misclassification. These samples typically include those with complex backgrounds or only partial representations of the objects of interest. Conversely, weights are reduced for simpler samples.

Consequently, focal loss diminishes the impact of straightforward samples on training loss while amplifying the significance of challenging ones. The focal loss function is calculated as follows:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t) \quad (6)$$

where:

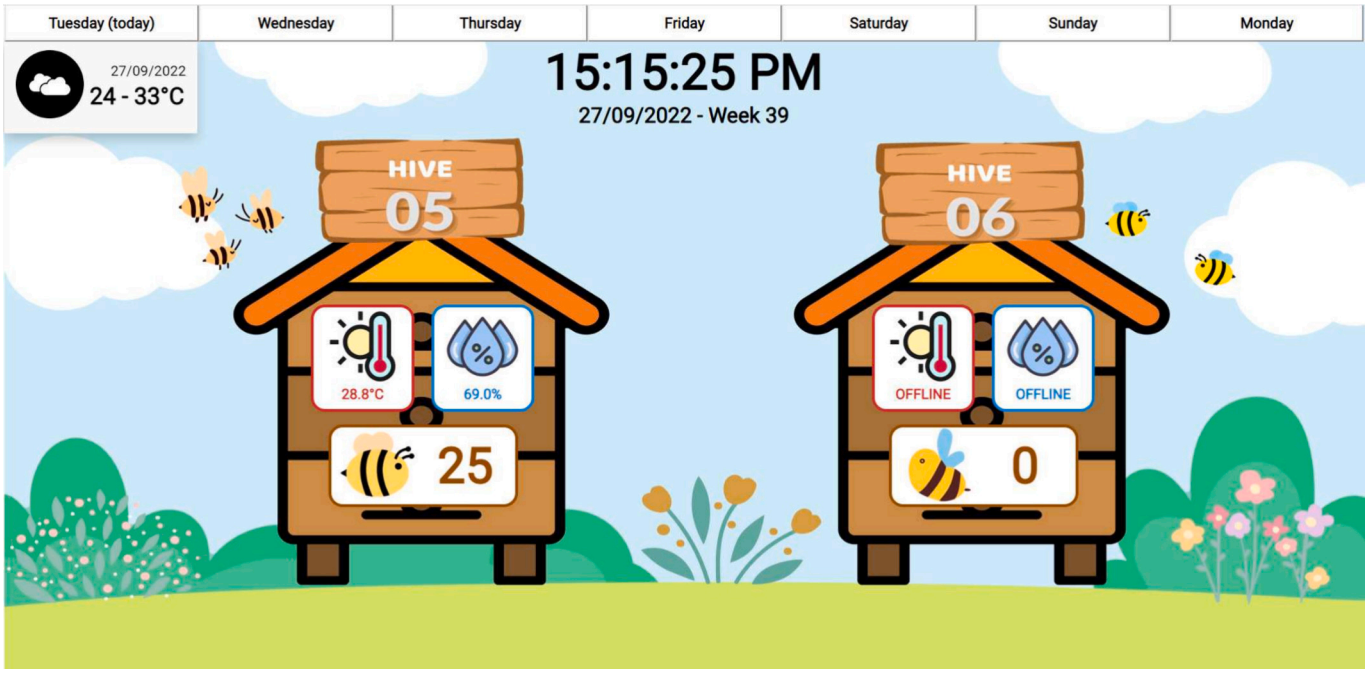


Fig. 17. Home interface of the beehive monitoring system.

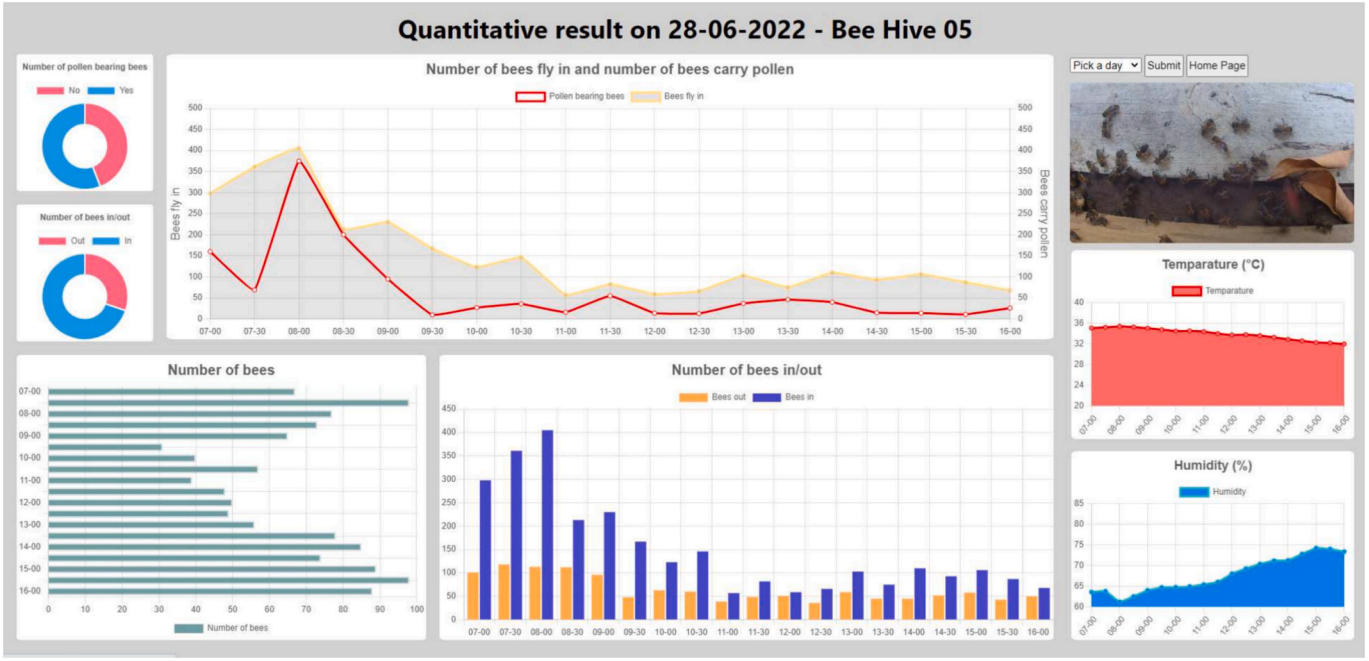


Fig. 18. Interface of the beehive monitoring system showing results of the bee and pollen-bearing bee counting.

- p_t is the probability of belonging to the class t
- α represents the ratio of generated boxes containing background and foreground information, aiding in balancing the disparity between background and foreground when generating boxes.
- γ represents the "concentration" of indistinguishable regions; the larger γ, the smaller the error values in the distinguishable regions, and the lower the contribution to the total loss of the model.

4.5. Third improvement: Two class faster RCNN + overlap sampler technique

Faster RCNN (Ren et al., 2015) operates as a two-stage object detection network. In the initial stage, the network focuses on proposing regions that may contain objects of interest. Subsequently, the second stage is dedicated to classifying these proposals and determining their precise locations. This two-stage approach enables Faster RCNN to achieve high accuracy across various tasks. However, it comes at the cost of computational time. To address imbalance issues inherent in pollen-bearing bee detection, we incorporate the Overlap Sampler method

(Chen et al., 2020) into the Faster RCNN network. Positioned at the end of the first stage of the Faster RCNN network, the Overlap Sampler method serves to enhance the generation of valuable proposed regions during the training process.

5. Experimental results

5.1. Setup

In our experiments, we used the Python 3.10 programming environment along with popular libraries such as PyTorch, TensorFlow, and OpenCV. The detection models were trained on a server equipped with an Intel® Core™ i7–8700 CPU @ 3.20GHz (32GB DDR4–2666 memory) and an NVIDIA GeForce RTX 3070 GPU (8GB GDDR6 memory).

The methods utilizing YOLOv5 will undergo training for 300 epochs, whereas the classification network in method 1 will undergo training for 200 epochs. Here, one epoch signifies training the model on all data once. For YOLOv5 training, fundamental parameters such as the learning rate will be set to 0.02, with weight decay and momentum values of 0.001 and 0.9, respectively. In contrast, methods employing Faster RCNN will be trained with 90 thousand iterations. Each iteration involves training the model on a batch of images. The fundamental parameters for Faster RCNN training mirror those of YOLOv5.

For evaluating the performance of the object detection models, we used a server equipped with an Intel(R) Xeon(R) CPU E5–2620 v2 @ 2.10GHz (16GB DDR3–1066 memory) and an NVIDIA GeForce GTX 1080 Ti GPU (11GB GDDR5X memory). The Intersection over Union (IoU) threshold was set to 0.5, and the confidence threshold was also set to 0.5.

5.2. Ablation study

As detailed in the Related Work section, the PollenDataset was initially introduced in (Rodriguez et al., 2018a). Some examples of the PollenDataset are illustrated in Fig. 11 The images comprising this dataset were captured at the entrance of a bee colony in June 2017 at the Bee facility of the Gurabo Agricultural Experimental Station of the University of Puerto Rico. This dataset encompasses 714 honey bee bounding boxes, consisting of 369 bee images with pollen loads and 345 bee images without pollen loads. Given its focus on pollen-bearing and non-pollen-bearing bee classification, we utilize this dataset to assess the effectiveness of the classification network devised in the first improvement.

We utilize K-fold cross-validation to partition the dataset into 5 distinct subsets. Each subset undergoes evaluation, and the final result is derived from the average of these evaluations. Parameters governing the training process include 200 epochs, a learning rate of 0.001, a batch size of 22, and the utilization of the Adam optimizer. The results attained through the proposed method are presented in Table 2, whereas the normalized matrix is depicted in Fig. 12. Experimental results demonstrate that the classification method outperforms other networks, yielding a mean accuracy of 98.4%.

To gauge the efficacy of the proposed network, we compare its classification results with those of other pertinent studies on the PollenDataset. The classification accuracy results are outlined in Table 3. Accordingly, our proposed model has 2% higher accuracy than the second-ranked methods in terms of accuracy. Additionally, to elucidate the regions contributing to the network's decision-making process, several GradCAM images are provided in Fig. 13. These images are represented as heatmaps, wherein areas of greater interest to the model are denoted by red, whereas those of lesser interest are denoted by blue. As depicted in the figure, the model predominantly focuses on the pollen-laden areas of the bee to differentiate between pollen-bearing and non-pollen-bearing bees.

5.3. Pollen bearing bee detection results

To demonstrate the superior performance of our proposed methods compared to the baseline methods, we first conducted experiments to assess the performance of five methods, comprising two baseline methods and three enhanced methods, using the VnPollenBee dataset. The results obtained are presented in Table 4 and Fig. 14.

It is evident that among the three improvements, the third enhancement (Faster RCNN + Overlap Sampler method) yields the most favorable outcomes. These results exhibit significantly higher performance compared to other methods across both Recall and MR, demonstrating robust stability across all measures. Despite the impressive results achieved by the classification network on the PollenDataset, as demonstrated in Section 5.2, its performance in pollen-bearing bee detection on the more challenging VnPollenBee dataset remains suboptimal, with Precision, Recall, and F1 score at 58%, 88%, and 70%, respectively. Remarkably, the second and third improvements surpass their respective baseline models. The second enhancement reduces MR from 15% to 12%, while achieving a + 2% increase in F1 score compared to the YOLOv5 baseline model. In comparison to the baseline Faster RCNN method, the third enhancement (Faster RCNN + Overlap Sampler) yields superior results, boasting a precision of 99%, recall of 93%, and F1 score of 95%. This indicates that focusing on more challenging samples enhances the model's performance. Regarding MR and FAR, the proposed methods exhibit the lowest values, with an MR of 7% and FAR of 0.4%. This implies that the proposed methods effectively identify pollen-bearing bees with minimal false positives.

Furthermore, Fig. 15 depicts some detection outcomes of the second proposed enhancement (YOLOv5 + focal loss) and the third enhancement (Faster RCNN + Overlap Sampler). It is noticeable that in certain instances, the second enhancement overlooks some pollen-bearing bees, whereas the third enhancement successfully detects them.

Fig. 16 shows that, under conditions of high bee density, similarity in appearance among bees, low light, similarity between the color of the bees' bodies and the color of the landing board, the presence of leaves, and especially the significant disparity between the number of pollen-bearing bees and non-pollen-bearing bees, the proposed Faster RCNN + Overlap Sampler model is still capable of accurately identifying pollen-bearing bees and non-pollen-bearing bees.

We conducted experiments to compare the proposed method with three state-of-the-art models: EfficientDet (Tan et al., 2020), DETR (Carion et al., 2020), and RetinaNet (Lin et al., 2017). EfficientDet (Tan et al., 2020) uses a weighted BiFPN and a compound scaling method to build a scalable and efficient object detector. DETR (Carion et al., 2020) is an end-to-end object detection method based on a transformer encoder-decoder architecture, which improves the object detection process by eliminating some hand-designed components and using bipartite matching. RetinaNet (Lin et al., 2017) is a simple dense detector that uses a focal loss function to address the class imbalance problem. All three methods have achieved high detection performance on the challenging COCO benchmark. The results of these methods on the VnPollenBee dataset are shown in Table 5. Our proposed model demonstrates superior performance compared to the other models. Specifically, compared to the second-best model (DETR), our proposed model's FAR, Precision, and F1 score are higher by 3%, 4%, and 1%, respectively. Although the MR measurement of our proposed model is higher than that of DETR, the difference is not significant.

Additionally, the number of parameters, GFLOPs, and the average inference time per image for each model in our experiments are computed and shown in Table 6.

5.4. Application to the problem of counting pollen-bearing bees

Counting the number of pollen-bearing bees entering the hive enables beekeepers to monitor the hive's food status and promptly replenish it if necessary. Recognizing this benefit, we integrated the

proposed pollen-bearing bee detection method into a monitoring system. Illustrated in Fig. 17 is the home interface of our application website, where beekeepers can select the hive of interest to view its status, including humidity, temperature, and the number of pollen-bearing bees at different intervals.

Leveraging the superior performance of the third improvement (Faster RCNN + Overlap Sampler), we applied it to count the number of pollen-bearing bees in videos captured at the hive entrance. As pollen-bearing bee detection occurs in each frame, an additional object-tracking algorithm was implemented to facilitate bee counting. Essentially, this tracking algorithm assigns the same ID to the same pollen-bearing bee detected across different frames. Quantitative results for each beehive are depicted in Fig. 18, with the number of pollen-bearing bees represented by a red curve.

5.5. Discussions

Pollen serves as a crucial food source for bees, offering abundant protein, vitamins, and minerals. Foragers bring pollen into the beehive, where processing workers prepare it for storage. As nursing bees consume pollen, their hypopharyngeal and mandibular glands develop and secrete royal jelly to feed larvae and the queen, facilitating the colony's robust development. Consequently, as the number of bees returning to the colony with pollen loads increases, the beehive grows stronger and healthier (W. M. L, 1991).

In recent years, several methods have emerged with the primary objective of automatically determining the number of bees with pollen loads. For instance, a method proposed in (Babic et al., 2016) relies on background subtraction for bee detection, making it susceptible to variations in background, the lighting conditions, and requiring precise image acquisition setups. To alleviate these constraints, in 2021, Ngo et al. proposed employing deep learning methods for detecting bees and pollen-bearing bees (Ngo et al., 2021). In comparison with the approach presented in (Ngo et al., 2021), the methods proposed in this paper address imbalance issues in pollen-bearing bee detection, enabling the detection of pollen-bearing bees with low FAR and MR. This capability facilitates further monitoring and assessment of beehive conditions. While the proposed methods show promising results on the VnPollenBee dataset, integrating these methods into a beehive monitoring system reveals challenges. Due to the low camera frame rate, the methods struggle to detect bees when they move quickly or change direction suddenly, leading to potential missed detections. In such cases, incorporating temporal information is essential to ensure reliable detection results.

The VnPollenBee dataset represents the first collection of pollen-bearing and non-pollen-bearing honeybee images in Vietnam dedicated to training and evaluating honeybee detection models. Collected over multiple days and various times, the dataset includes images captured under diverse natural lighting conditions, including challenging scenarios such as low light and foliage obstructing the landing platform. As such, this dataset serves as a valuable benchmark for assessing pollen-bearing bee detection models. Despite a significant imbalance between pollen-bearing and non-pollen-bearing bee images, making it suitable for addressing data imbalance issues, the dataset remains limited in size and comprises images of only one bee species. Future efforts will focus on expanding and diversifying the dataset.

6. Conclusions and future works

This work contributes a step toward an automatic honeybee monitoring system, focused on detecting and counting the number of pollen-bearing bees. Three improvements have been introduced to address the imbalance issue in pollen-bearing bee detection. Experimental results have demonstrated that the proposed methods outperform the baseline models. Compared with the baseline model, the second method increased the Recall from 85% to 88% and the F1 score from 91% to 93%, whereas the third method led to improvements of 3%, 2%, and 2% in Precision, Recall, and F1 score, respectively. Furthermore, the proposed methods achieved the lowest values of FAR and MR. Additionally, a new dataset named VnPollenBee, comprising 2051 images with 60,826 annotated boxes for both pollen-bearing and non-pollen-bearing bees captured under various conditions, has been fully annotated. This dataset is publicly available for the research community. Although the obtained results are promising, further work is needed to optimize the proposed methods for devices with limited resources. Integrating the results of bee and pollen-bearing bee detection with other key indicators essential for assessing the health condition of beehives is also a focus of our future work. Additionally, once bees are detected and tracked, their activities could be recognized using methods similar to human activity recognition (Çalışkan, 2023).

CRediT authorship contribution statement

Dinh-Tu Nguyen: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Duc-Manh Nguyen:** Software, Data curation. **Hong-Quan Nguyen:** Visualization, Validation, Software. **Hong-Thai Pham:** Writing – review & editing, Validation, Funding acquisition, Data curation. **Thi-Thu-Hong Phan:** Supervision, Resources, Funding acquisition, Conceptualization. **Hai Vu:** Validation, Supervision, Conceptualization. **Thi-Lan Le:** Writing – review & editing, Writing – original draft, Validation, Data curation, Conceptualization.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by the national research project titled "Study and application of Industry 4.0 technologies in the management of honey bee production for export and national consumption", Grant Number: KC4.0-20/19-25. The funders had no role in designing experiments, collecting and processing data, deciding to publish, or preparing the manuscript.

References

Babic, Z., Pilipovic, R., Risojevic, V., Mirjanic, G., 2016. Pollen bearing honey bee detection in hive entrance video recorded by remote embedded system for pollination monitoring. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3, p. 51.

Berkaya, S.K., Gunal, E.S., Gunal, S., 2021. Deep learning-based classification models for beehive monitoring. Eco. Inform. 64, 101353.

Bilik, S., Kratochvila, L., Ligocki, A., Bostik, O., Zemcik, T., Hybl, M., Horak, K., Zalud, L., 2021. Visual diagnosis of the varroa destructor parasitic mite in honeybees using object detector techniques. Sensors 21, 2764. https://doi.org/10.3390/s21082764.

Bilik, S., Zemcik, T., Kratochvila, L., Ricanek, D., Richter, M., Zambanini, S., Horak, K., 2024. Machine learning and computer vision techniques in continuous beehive monitoring applications: a survey. Comput. Electron. Agric. 217, 108560.

Braga, A.R., Gomes, D.G., Freitas, B.M., Cazier, J.A., 2020. A cluster-classification method for accurate mining of seasonal honey bee patterns. Eco. Inform. 59, 101107.

Çalışkan, A., 2023. Detecting human activity types from 3d posture data using deep learning models. Biomed. Sign. Proces. Control 81, 104479.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: European Conference on Computer Vision. Springer, pp. 213–229.

Chen, J., Luo, B., Wu, Q., Chen, J., Peng, X., 2020. Overlap sampler for region-based object detection. In: EEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 756–764.

da Silva, D.A., Bomfim, I.G.A., Braga, A.R., Gomes, D.G., 2023. Applying computer vision models to detect in real time the pollen flow at the input of honeybee hives (apis mellifera l.). In: Anais do XIV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais. SBC, pp. 21–30.

Dembski, J., Szymański, J., 2019. Bees detection on images: Study of different color models for neural networks. In: International Conference on Distributed Computing and Internet Technology. Springer, pp. 295–308.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.

Fruet, G.V., Bomfim, I.G.A., Domingues, R.C., Braga, A.R., Gomes, D.G., 2023. Apisflow: A real-time automated tool to detect, classify and count honey bees castes at the hive entrance. In: Anais do XIV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais. SBC, pp. 1–10.

Hickert, F.. Dataset for a Camera Based Bee-Hive Monitoring (2021). URL. https://github.com/BeeAlarmed.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Fang, J., Michael, K., Montes, D., Nadar, J., Skalski, P., et al., 2022. ultralytics/yolov5: v6. 1-tensorrt, tensorflow edge tpu and openvino export and inference. Zenodo.

Kalman, E.-L., Winquist, F., Lundström, I., 1997. A new pollen detection method based on an electronic nose. Atmos. Environ. 31 (11), 1715–1719.

Knauer, A.C., Gallmann, J., Albrecht, M., 2022. Bee tracker—an open-source machine learning-based video analysis software for the assessment of nesting and foraging performance of cavity-nesting solitary bees. Ecol. Evol. 12 (3), e8575.

Kongsilp, P., Taetragool, U., Duangphakdee, O., 2024. Individual honey bee tracking in a beehive environment using deep learning and kalman filter. Sci. Rep. 14 (1), 1061.

Krishnasamy, V., Sridhar, N., Niranjan, L., 2023. An iot-based beehive monitoring system for real-time monitoring of *apis cerana indica* colonies. Sociobiology 70 (4), e9352.

Kulyukin, V., Mukherjee, S., Amlathe, P., 2018. Toward audio beehive monitoring: deep learning vs. standard machine learning in classifying beehive audio samples. Appl. Sci. 8 (9), 1573.

Le, T.-N., Thi-Thu-Hong, P., Nguyen, H.-D., Thi-Lan, L., et al., 2023. A novel convolutional neural network architecture for pollen-bearing honeybee recognition. Int. J. Adv. Comput. Sci. Appl. 14 (8).

Lee, H.G., Kim, M.-J., Kim, S.-B., Lee, S., Lee, H., Sin, J.Y., Mo, C., 2023. Identifying an image-processing method for detection of bee mite in honey bee based on keypoint analysis. Agriculture 13 (8), 1511.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988.

Narcia-Macias, C.I., Guardado, J., Rodriguez, J., Rampersad-Ammons, J., Enriquez, E., Kim, D.-C., 2023. Intellibeehive: An automated honey bee, pollen, and *varroa destructor* monitoring system. arXiv preprint arXiv:2309.08955.

Ngo, T.N., Wu, K.-C., Yang, E.-C., Lin, T.-T., 2019. A real-time imaging system for multiple honey bee tracking and activity monitoring. Comput. Electron. Agric. 163, 104841.

Ngo, T.N., Rustia, D.J.A., Yang, E.-C., Lin, T.-T., 2021. Automated monitoring and analyses of honey bee pollen foraging behavior using a deep learning-based imaging system. Comput. Electron. Agric. 187, 106239 https://doi.org/10.1016/j.compag.2021.106239.

Nguyen, H.C., Nguyen, D.-T., Phung, T.-H., Nguyen, T.-L.-A., Nguyen, T.V., Thai, P., Phan, T.-H., Le, T.-L., Nguyen, X.D., Thi, N.L.T., Hai, V., 2022. A method for automatic honey bees detection and counting from images with high density of bees. In: 2022 IEEE Ninth International Conference on Communications and Electronics (ICCE), pp. 406–411. https://doi.org/10.1109/ICCE55644.2022.9852024.

Nguyen, D.-T., Nguyen, D.-M., Pham, D.-T., Than, K., Pham, H.-T., Vu, H., 2023. Bayesian method for bee counting with noise-labeled data. In: Proceedings of the 12th International Symposium on Information and Communication Technology, pp. 401–408.

Odemer, R., 2022. Approaches, challenges and recent advances in automated bee counting devices: a review. Ann. Appl. Biol. 180 (1), 73–89. https://doi.org/10.1111/aab.12727.

Ren, S., He, K., Girshick, R.B., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Neural Information Processing Systems (NIPS), pp. 91–99.

Requier, F., 2019. Bee colony health indicators: synthesis and future directions. In: CAB Reviews Perspectives in Agriculture Veterinary Science Nutrition and Natural Resources, 14, pp. 1–13. https://doi.org/10.1079/PAVSNNR201914056.

Rodriguez, I.F., Megret, R., Acuna, E., Agosto-Rivera, J.L., Giray, T., 2018a. Recognition of pollen-bearing bees from video using convolutional neural network. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 314–322. https://doi.org/10.1109/WACV.2018.00041.

Rodriguez, I., Branson, K., Acuna, E., Agosto, J.L., Giray, T., Mégret, R., 2018b. Honeybee Detection and Pose Estimation Using Convolutional Neural Networks, in: Congrès Reconnaissance Des Formes, Image, Apprentissage et Perception (RFIAP).

Rodriguez, I.F., Chan, J., Alvarez Rios, M., Branson, K., Agosto-Rivera, J.L., Giray, T., Mégret, R., 2022. Automated video monitoring of unmarked and marked honey bees at the hive entrance. Front. Comp. Sci. 3, 769338.

Rustam, F., Sharif, M.Z., Aljedaani, W., Lee, E., Ashraf, I., 2024. Bee detection in bee hives using selective features from acoustic data. Multimed. Tools Appl. 83 (8), 23269–23296.

Schurischuster, S., Kampel, M., 2020. Image-based classification of honeybees, 2020 tenth international conference on image processing theory. Tools Appl. (IPTA) 1–6.

Sledevic, T., 2018. The application of convolutional neural network for pollen bearing bee classification, 6th workshop on advances in information, electronic and electrical engineering, Vilnius. Lithunia. https://doi.org/10.1109/AIEEE.2018.8592464.

Sledević, T., Plonis, D., 2023. Toward bee behavioral pattern recognition on hive entrance using yolov8. In: 2023 IEEE 10th Jubilee Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE). IEEE, pp. 1–4.

Tan, M., Pang, R., Le, Q.V., 2020. Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781–10790.

Truong, T.H., Du Nguyen, H., Mai, T.Q.A., Nguyen, H.L., Dang, T.N.M., et al., 2023. A deep learning-based approach for bee sound identification. Eco. Inform. 78, 102274.

Vladan Stojnic, R.P., 2018. Vladimir Risojevic, Detection of Pollen Bearing Honey Bees in Hive Entrance Images, 17th International Symposium INFOTEH-JAHORINA, 21–23 March 2018. https://doi.org/10.1109/INFOTEH.2018.8345546.

Voudiotis, G., Moraiti, A., Kontogiannis, S., 2022. Deep learning beehive monitoring system for early detection of the varroa mite. Signals 3 (3), 506–523.

W. M. L, 1991. The Biology of the Honey Bee. Harvard University Press.

Yang, J., 2018. The BeeImage Dataset: Annotated Honey Bee Images. https://www.kaggle.com/jenny18/honey-bee-annotated-images/.

Yang, C., Collins, J., 2019. Deep learning for pollen sac detection and measurement on honeybee monitoring video. In: 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1–6.

Yoo, J., Siddiqua, R., Liu, X., Ahmed, K.A., Hossain, M.Z., 2023. Beenet: An end-to-end deep network for bee surveillance. Procedia Comp. Sci. 222, 415–424.

Zhao, Y., Deng, G., Zhang, L., Di, N., Jiang, X., Li, Z., 2021. Based investigate of beehive sound to detect air pollutants by machine learning. Eco. Inform. 61, 101246. https://doi.org/10.1016/j.ecoinf.2021.101246. URL. https://www.sciencedirect.com/science/article/pii/S1574954121000376.