SN Computer Science (2026) 7:605
https://doi.org/10.1007/s42979-026-05189-w

ORIGINAL RESEARCH

Honeybee Counting on Comb Images via Part-Level Annotation and Hungarian Matching

Naoya Tsuruta[1] · **Keito Ohara**[2] · **Madoka Hasegawa**[1]

Received: 26 June 2025 / Accepted: 8 June 2026

Abstract

Inspection of combs is a critical task in beekeeping where the condition of the comb and the size of the bee population within a hive are assessed. Manual counting of bees is a cumbersome process that compromises efficiency and accuracy. To address this issue, we propose an automated, machine-learning-based inspection method that uses single images of combs with bees. This method eliminates the need for specialized equipment, such as optical sensors, video cameras, or additional entrance modifications for bee traffic monitoring. The research question was to determine which anatomical region of the honeybee provides the most accurate detection results, and whether combining multiple detection approaches could improve accuracy overall. We annotated three distinct parts of honeybees in comb images: the abdomen, head, and whole-body. The SSD models with VGG16 pretrained backbones and YOLOv11 were fine-tuned to detect each part, and their detection accuracies were compared. We also attempted to combine predicted bounding boxes using the Hungarian method to compensate for those that were not detected due to occlusions. Experimental results showed superior accuracy of the abdomen detector compared to the head and whole-body detectors across different densities of bee images. The integrated approach effectively reduced the number of undetected bees compared to individual detection methods. Furthermore, applying the matching framework to YOLOv11 yielded consistent recall improvements over whole-body detection alone, empirically demonstrating that the proposed matching strategy is detector-agnostic and addresses occlusion-induced false negatives independently of the underlying detection architecture. The proposed machine learning-based method successfully automates bee counting from single comb images. This offers a simple solution for beekeeping comb inspection automation.

Keywords Honeybee colony management · Automated counting · Deep learning · Object detection

Introduction

The western honeybee (*Apis mellifera*, hereinafter referred to as "honeybee") is important in agricultural crop pollination and is indispensable to crop production. Beekeepers manage colonies by inspecting hives to assess comb conditions and estimate the number of adult bees. Furthermore, as pesticide application in nearby fields poses a risk of honeybee mortality, efficacy tests are conducted during pesticide development to evaluate the potential impact on honeybees, with the number of bees in the hive serving as one of the key indicators.

Currently, beekeepers typically estimate honeybee populations by either visually approximating the number of adult bees on a comb or manually counting bees in still images. However, visual estimation is subject to challenges such as reduced visibility owing to beekeeping protective suits, extended inspection times, and inconsistencies among observers. Manual counting from still images is labor-intensive because it requires counting hundreds of bees per comb side. These limitations highlight the need for automation to improve efficiency and accuracy.

Odemer comprehensively reviewed various approaches to automating honeybee counting [1]. In contrast to crop and wildlife observations, honeybee counting is distinguished

✉ Naoya Tsuruta
naoya@icl.is.utsunomiya-u.ac.jp

[1] School of Engineering, Utsunomiya University, Yoto 7-1-2, Utsunomiya, Tochigi 3218585, Japan

[2] Graduate School of Regional Development and Creativity, Utsunomiya University, Yoto 7-1-2, Utsunomiya, Tochigi 3218585, Japan

Published online: 23 June 2026

by using various sensors, including optical, temperature, and capacitance sensors [2, 3]. The condition of the hive can be monitored, and abnormalities can be detected based on the information from these sensors [4]. In recent years, several methods that leverage computer vision and deep learning have been proposed for continuous monitoring [5]. Despite these advancements, many existing methods require specialized hardware, such as multiple sensors and webcams, and additional entrances to the hive for accurate honeybee detection. These requirements render the methods impractical for routine use by beekeepers. Therefore, we proposed a practical and efficient honeybee counting method based on machine learning, using still images that can be easily captured in a single shot of a comb removed from a hive, in this study.

The primary challenge in this approach is the detection of bees in a single comb image. Because we used a single image for detection, it was not possible to use information from the previous and subsequent frames as in the video sequences. Honeybees often cluster tightly on the comb to regulate the temperature of larvae and pupae or insert their heads into cells for feeding or cleaning, which frequently results in the partial occlusion of individual bees.

To address these challenges, we propose a part-based detection and matching framework for automated honeybee counting from single comb images. Although our approach is implemented using SSD as the base detector, the framework's core components—part-level annotation strategy, dual-stage Hungarian matching, and final counting procedure—are detector-agnostic, as demonstrated by validating the complete framework with both SSD and YOLOv11 architectures. Our main contributions are threefold: First, we systematically compare three annotation strategies—abdomen, head, and whole body—and analyze their detection performance under different occlusion conditions. We demonstrate that abdomen detection achieves higher recall due to the distinctive visibility of abdominal stripes even when other body parts are occluded. Second, we extend the part-based matching approach of Qi et al. [14] from pedestrian tracking to honeybee counting by implementing a dual-stage matching strategy that accommodates the unique visibility patterns of honeybee anatomy. Third, we integrate these detections using the Hungarian algorithm to establish correspondences between part-level and whole-body detections, and validate the detector-agnostic nature of this framework by demonstrating consistent recall improvements with both SSD and YOLOv11 architectures.

While Bilik et al. [6] also employed part-level annotations for bee detection in controlled laboratory settings, their work did not address the systematic integration of multiple part detections for counting under severe occlusion. Our approach extends this concept by introducing a dual-stage matching framework that combines abdomen, head, and whole-body detections, and provides quantitative comparisons across different bee densities to demonstrate robustness in field conditions.

Related Work

As previously mentioned, bee colonies are often densely populated; therefore, an approach that annotates the entire body of a single individual is inappropriate. In single-image object counting, a dataset annotated only with human faces is used for crowd counting [7]. While other studies have examined counting animals and plants [8, 9], these annotations depend on the characteristics of the targets to be detected and require considerable human resources.

Machine learning models for crowd-counting can be approached in several ways. These methods include object detection, regression models, and methods based on density estimation [10]. Although density estimation-based methods using convolutional neural networks are mainstream, in this study, we decided to use an object detection method, Single Shot MultiBox Detector (SSD) [11], from the viewpoints of accuracy and execution speed with a view to future implementation on edge devices. SSD is based on a forward multilayer convolutional network in which features are detected in the base network, and subsequent convolutional layers assist in object detection using multiscale feature maps. Various improvements have been implemented to enhance the accuracy of SSD, including integration of the attention mechanism, DenseNet, and Feature Fusion [12, 13]. However, the objective of this study was to verify the difference in detection accuracy by annotating the position and matching the estimation results of the whole body and each part. It is important to note that this study was not limited to the SSD as an inference model.

In accordance with the adoption of SSD, the annotation process in this study is provided in a bounding box (BB) format. As an object detection method using BB, Qi et al. proposed a method for tracking a person in a crowd by annotating only the head, which is less prone to occlusion, and tracking the whole body [14]. This approach has been demonstrated to reduce the adverse effects of false negatives and false positives in tracking owing to heavy occlusions. The correspondence between the head and whole-body box of a pedestrian in this method was determined using the Hungarian method [15]. Similarly, Bilik et al. [6] employed part-level annotations (head and thorax) for bee detection in controlled laboratory environments, but did not develop a matching framework for counting applications. This study extends these approaches by proposing a dual-stage Hungarian matching method for honeybee counting that

integrates abdomen, head, and whole-body detections. Our approach differs from Qi et al. [14] in two key aspects: (1) dual-stage matching to accommodate honeybee anatomy, and (2) explicit exclusion of non-overlapping part detections before Hungarian matching to prevent spurious assignments in dense configurations, addressing the challenge of severe occlusion in densely populated field conditions.

Proposed Method

Our proposed method consists of two main components: (1) part-level detection of honeybee anatomical regions (abdomen, head, and whole body), and (2) Hungarian algorithm-based matching to integrate these detections. In this study, we implement the detection component using SSD [11] with VGG16 backbone due to its balance between accuracy and computational efficiency. However, the matching framework and overall methodology are designed independently of the specific detector architecture. The following subsections describe the implementation details and each processing step.

The goals of this study were to label the abdomen and head of honeybees to enable the detection of individuals whose whole body is not visible and to investigate a method for counting the number of bees by mapping detections to the same individuals. An overview of the proposed method is shown in Fig. 1. The input was a single image captured from the front of the hive. First, the input image was divided into overlapping blocks according to the SSD input size. The two VGG16 pre-trained SSD models were then finetuned with head annotated images and abdomen/whole-body annotated images, respectively, and then inference was performed to obtain predictive bounding boxes. The obtained BB coordinates were converted into coordinates of the original input image using a block image integration process. Finally, the abdomen box was matched to the whole-body box, and the head box was matched to the whole-body box of the same individual. The matched whole-body boxes and unmatched abdomen and head boxes constituted the final detection results. In the following sections, we describe the honeycomb image dataset and annotation method, followed by a detailed explanation of each process.

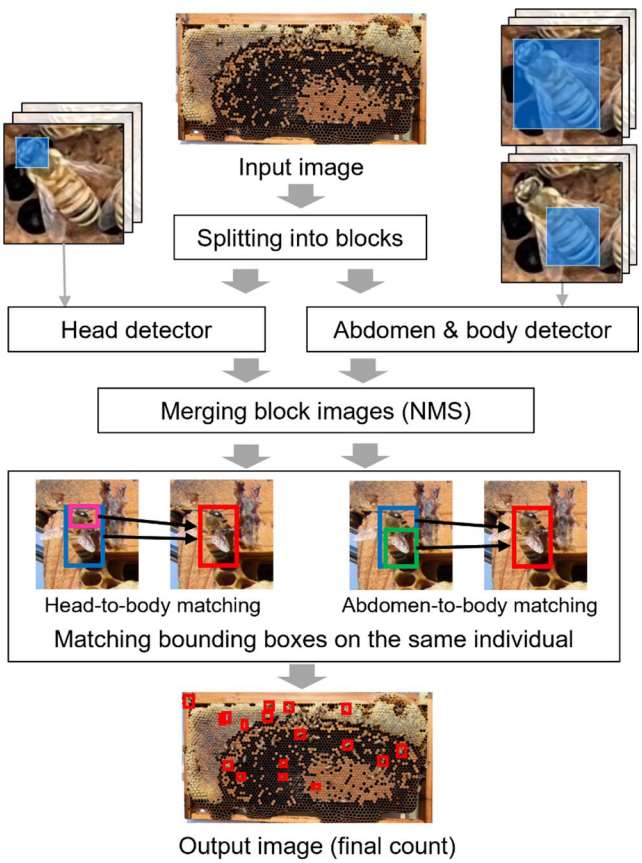

Fig. 1 The overview of our proposed method

Training Images and Annotations

A total of 1000 images were used for the training. These images were manually annotated using the abdomen, head, and overall annotations. The annotation process involved cropping areas from the comb images containing one to several bees to 140 × 140, 200 × 200, and 300 × 300 pixels. Since bees are positioned on an approximately planar surface with minimal depth variation, the scale differences primarily reflect the number of overlapping individuals and natural size variation among bees. Smaller crops (140 × 140) capture isolated or minimally overlapping bees, while larger crops (300 × 300) encompass dense clusters where multiple bees overlap significantly. Examples of training images are shown in Fig. 2. When training the model, the training data were split into training and validation data at a ratio of 9:1.

To improve model robustness, we applied standard data augmentation techniques during training following the SSD pipeline [11]: photometric distortions, random horizontal mirroring, random canvas expansion, random sample cropping, mean subtraction, BGR mean values (104, 117, 123). All augmented images were resized to 300 × 300 pixels to match the SSD input requirements. Validation images underwent only resizing and mean subtraction without augmentation.

A honeybee's body comprises three primary anatomical regions: the head, thorax, and abdomen. In both the training and test images, bounding boxes were manually drawn around the honeybee's abdomen, head, and whole body using the open-source annotation tool *labelImg*. Each annotation type was saved as a separate XML file in PASCAL VOC format. Examples of each annotation type are shown

Fig. 2 Examples of training images

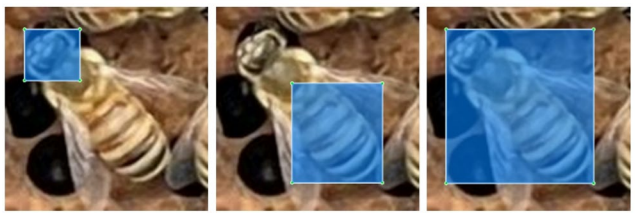

Fig. 3 Example annotations: from left to right: head, abdomen, and whole body

Table 1 The number of bees, image size, and categorization of test images

Image id	Density	Number of bees	Size (px)
1	low	41	2000 × 998
2	low	88	2000 × 1046
3	middle	416	2000 × 1010
4	middle	484	2000 × 1026
5	high	811	2000 × 1109
6	high	985	2000 × 1100

in Fig. 3. Notably, the thorax was not annotated because of the paucity of unique color and shape characteristics. Bounding boxes were drawn around the visible portion of each body part, with the boundary estimated at the point of occlusion for partially visible individuals. Annotations were primarily performed by one trained annotator, with portions of the dataset annotated by additional annotators under their supervision. All annotations were reviewed and finalized by the primary annotator to ensure consistency. We acknowledge that a formal inter-annotator agreement score was not computed, which is a limitation of the current study.

The training dataset contained 8,864 bounding box annotations distributed as follows: 4,393 abdomens (49.5%), 2,819 heads (31.8%), and 2,252 whole bodies (25.4%). The higher proportion of abdomen annotations reflects the biological reality that bee abdomens remain visible even under partial occlusion, while complete whole-body visibility is less frequent in crowded hive conditions. This class imbalance motivated our part-based detection strategy, where abdomens and heads serve as robust indicators for counting individuals whose entire body may not be visible.

Six images were used for the test, including four captured with a Canon EOS Kiss X9 SLR camera (two each at 6000 × 4000 and 3984 × 2656) and two captured with an EOS R5 camera at 8192 × 5464. After cropping the beehive images based on the beehive foundation frame, the images were resized to 2000 pixels on the long side while maintaining the aspect ratio. The six images were categorized into three levels of bee density. The number of bees, image size, and categorization of each image are listed in Table 1. Four ground truths were created for each test image: one for abdominal detection, one for head detection, one for whole-body detection, and one for the final detection results after mapping. All the test images can be found in the appendix.

Image Splitting and Object Detection

The comb image was segmented into 300 × 300-pixel blocks for input into the SSD. The stride was set to 200 pixels such that the blocks overlapped to account for the possibility that a single bee could be divided into two blocks during inference. The stride was set to less than 200, and the overlap was increased at the image borders. This configuration ensured that all blocks were cropped to 300 × 300 pixels, preventing feature loss during inference.

Subsequently, inference using the two trained models was performed on each block image obtained in the previous process to predict the BBs for the abdomen, head, and whole body. The head and abdomen/whole-body models were separated because the output thresholds for the BBs were set to different values. Specifically, the threshold for the head model was set to 0.4, whereas that for detecting the abdomen and whole-body was set to 0.5. The lower threshold for head detection (0.4) was necessary because heads are smaller and have less distinct features than abdomens, making them harder to detect with high confidence. The higher threshold for abdomen/body detection (0.5) reduces false positives while maintaining sufficient recall, as abdomens are more reliably visible due to their distinctive striped patterns.

Merging Block Images

The block-based processing approach results in multiple detections of the same bee when it appears in overlapping

regions of adjacent blocks. To address this, the predicted bounding boxes from each block were first converted from block-relative coordinates to absolute coordinates in the original image using the block's position offset. Overlapping bounding boxes were then removed using non-maximum suppression (NMS) with an IoU threshold of 0.3. This relatively low threshold (compared to the typical 0.5) was chosen to aggressively remove duplicates in dense bee clusters where individuals have high spatial overlap. NMS was applied separately for each detection type (head, abdomen, and whole body) to preserve valid detections of different body parts from the same individual.

Matching Bounding Boxes on the Same Individual

Up to three BBs can be detected per honeybee, i.e., the head, abdomen, and whole body. To avoid counting the same individual multiple times, we adapted the part-based matching approach proposed by Qi et al. [14], which originally matched head and body detections for human counting. While Qi et al. used two-part detection for human, we extended this framework to accommodate honeybee anatomy and implemented a dual matching strategy: first performing abdomen-to-whole-body matching, followed by head-to-whole-body matching for bees where the abdomen matching failed. This modification addresses the unique visibility patterns in honeybee monitoring, where abdomens are frequently visible due to their distinctive coloration and positioning on the comb, while heads may be obscured.

We formulate the matching as an assignment optimization problem using the Hungarian algorithm. The cost between bounding boxes is computed based on their spatial overlap measured by Intersection over Union (IoU). For abdomen-whole-body matching, we construct an $\boldsymbol{m} \times \boldsymbol{n}$ cost matrix where m is the number of abdomen detections and n is the number of whole-body detections:

$$C_{abd}(i,j) = 1 - IoU\left(A_i, B_j\right) \tag{1a}$$

For head-to-whole-body matching, we construct a $\boldsymbol{p} \times \boldsymbol{n}$ cost matrix where $\boldsymbol{p}$ is the number of head detections:

Table 2 Parameters setting used in our experiments

Parameter	Value
Number of epochs	300
Batch size	32
Activation function	Swish
Learning rate	0.001
Optimization method	Stochastic Gradient Descent (SGD)
Momentum	0.85
Weight decay	5×10^{-4}
NMS threshold (IoU)	0.3
BB output threshold	0.4 (head), 0.5 (abdomen and body)

$$C_{head}(i,j) = 1 - IoU\left(H_i, B_j\right) \tag{1b}$$

where A_i denotes the i-th abdomen box, H_i denotes the i-th head box, B_j denotes the j-th whole-body box, and IoU $\in [0,1]$ represents the overlap ratio. A cost of 0 indicates perfect overlap (IoU = 1), while a cost of 1 indicates no overlap (IoU = 0). Before applying the Hungarian algorithm, we remove rows where all costs equal 1 (i.e., $\forall j$, IoU$(A_i, B_j) = 0$ for abdomen matching and IoU$(H_i, B_j) = 0$ for head matching), as these represent part detections with no spatial overlap with any whole-body detection. This preprocessing step is crucial for honeybee counting due to the compact body structure and dense spatial configurations in hive environments. For remaining rows, the Hungarian algorithm finds the globally optimal assignment that minimizes the total matching cost. Both matching stages use the same set of $\boldsymbol{n}$ whole-body boxes.

The final bee count is determined by combining matched and unmatched detections. All whole-body boxes are included in the final detection set, with each counted exactly once regardless of whether they were matched to abdomen or head. This ensures that a whole-body detection matched to both an abdomen and a head is counted only once. Additionally, abdomens that were not matched to any whole-body box are counted as individual bees, representing cases where only the abdomen is visible. Similarly, heads that were not matched to any whole-body box are counted as individual bees, representing cases where only the head is visible. Mathematically, the total bee count is computed as:

$$N = |\text{wholebodyboxes}| + |\text{unmatchedabdomens}| + |\text{unmatchedheads}| \tag{2}$$

This approach ensures each bee is counted exactly once while accounting for partial visibility scenarios common in dense hive environments, where complete body visibility may be obstructed by other bees or honeycomb structures.

Experimental Results

In this section, we present the counting results of the four methods: abdomen-only detection, head-only detection, whole-body detection, and the integration of the three results using the Hungarian method (proposed method). In addition, we calculated the percentage of correct responses for matching between the head and body and between the abdomen and body. The experiments were performed using PyTorch on a system running Windows 10 Pro, equipped with an Intel Core i9-10900K CPU (3.5GHz), an NVIDIA GeForce RTX3090 GPU, and 32 GB RAM. The experimental parameters are listed in Table 2. The computation time required to infer each image was a few seconds, whereas the

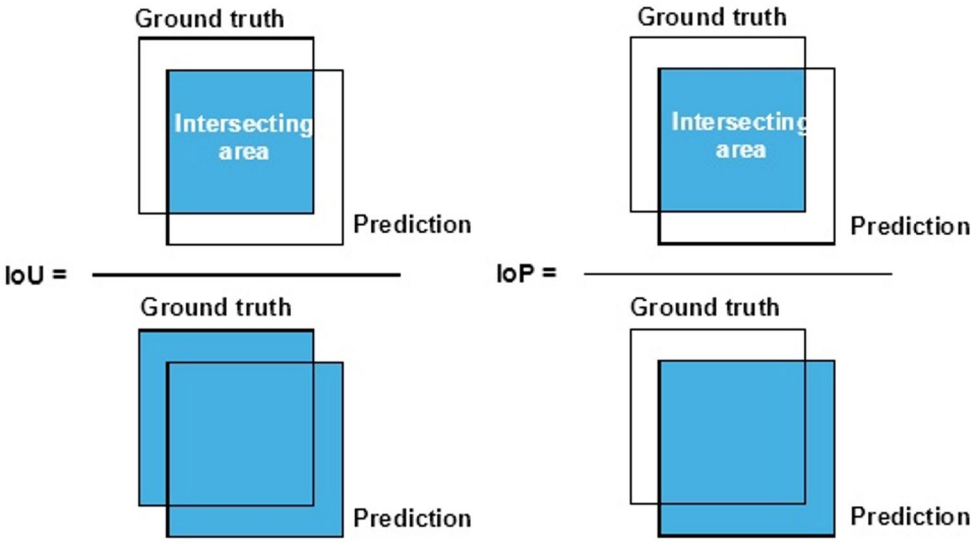

Fig. 4 Definition of IoU and IoP

Table 3 IoP thresholds and the number of true positives

IoP threshold	Image 1: 38 bees	Image 3: 392 bees	Image 5: 736 bees
0.2	38(± 0)	398(+ 6)	736(+0)
0.3	38(± 0)	396(+ 4)	733(− 3)
0.4	38(± 0)	392(+0)	726(− 10)
0.5	38(± 0)	391(−1)	721(− 15)
0.6	37(−1)	386(−6)	708(− 28)
0.7	36(−2)	379(− 13)	659(− 77)
0.8	34(−4)	354(− 38)	576(− 160)

mapping process required 30 s to 1 min in the high-density case.

Evaluation Metrics

We used precision, recall, and F1-scores as evaluation metrics for detection accuracy. These values were calculated as the true positive (TP), false positive (FP), and false negative (FN) values. The value of the intersection over prediction (IoP) determines whether the BB belongs to the TP, FP, or FN. The areas of intersection and prediction between the predicted BB and the ground truth are shown in Fig. 4. Predicted BBs whose IoP is above the threshold are called TP, those whose IoP is below the threshold are called FP, and the BBs in the ground truth whose IoP is below the threshold for any predicted BBs are called FN.

The IoP thresholds were set such that the detection results were close to those of the visual evaluation. The number of TPs was measured for three of the six test images (Image 1, Image 3, and Image 5) when the IoP threshold was varied from 0.2 to 0.8 in 0.1 increments after matching. Table 3 presents the threshold values and number of true positives (differences from manual counts are shown in parentheses).

Table 4 Mean precision, recall, and F1-score for each density category. Bold values indicate the highest value in each row

		Abdomen	Head	Whole body	Matching
Low-density	Precision	0.486	0.480	**0.664**	0.325
	Recall	0.946	0.752	0.612	**0.969**
	F1-score	**0.642**	0.586	0.637	0.486
Medium-density	Precision	**0.990**	0.798	0.936	0.714
	Recall	0.909	0.582	0.341	**0.964**
	F1-score	**0.948**	0.673	0.500	0.820
High-density	Precision	**0.974**	0.873	0.963	0.792
	Recall	0.807	0.492	0.303	**0.873**
	F1-score	**0.883**	0.629	0.461	0.831
Overall average	Precision	0.814	0.716	**0.884**	0.612
	Recall	0.880	0.616	0.421	**0.932**
	F1-score	**0.820**	0.628	0.540	0.711

The IoP threshold was determined as 0.3, as this value resulted in a reduced difference between the detection and manual counts, on average.

Comparison of Detection Methods

The averages of precision, recall, and F1-scores for each density category are listed in Table 4. Different ground truths were used for each detection method, and the number of bees was different because some individuals had only their heads or abdomens visible. For all densities and overall averages, the F1-score was highest for abdominal-only detection. This indicated that the detection of abdominal stripes, an important characteristic of honeybees, is effective for counting the number of individuals. The results obtained after the matching process had a higher recall rate and fewer undetected cases than those of the other detection methods. A comparison of the results for the high-density

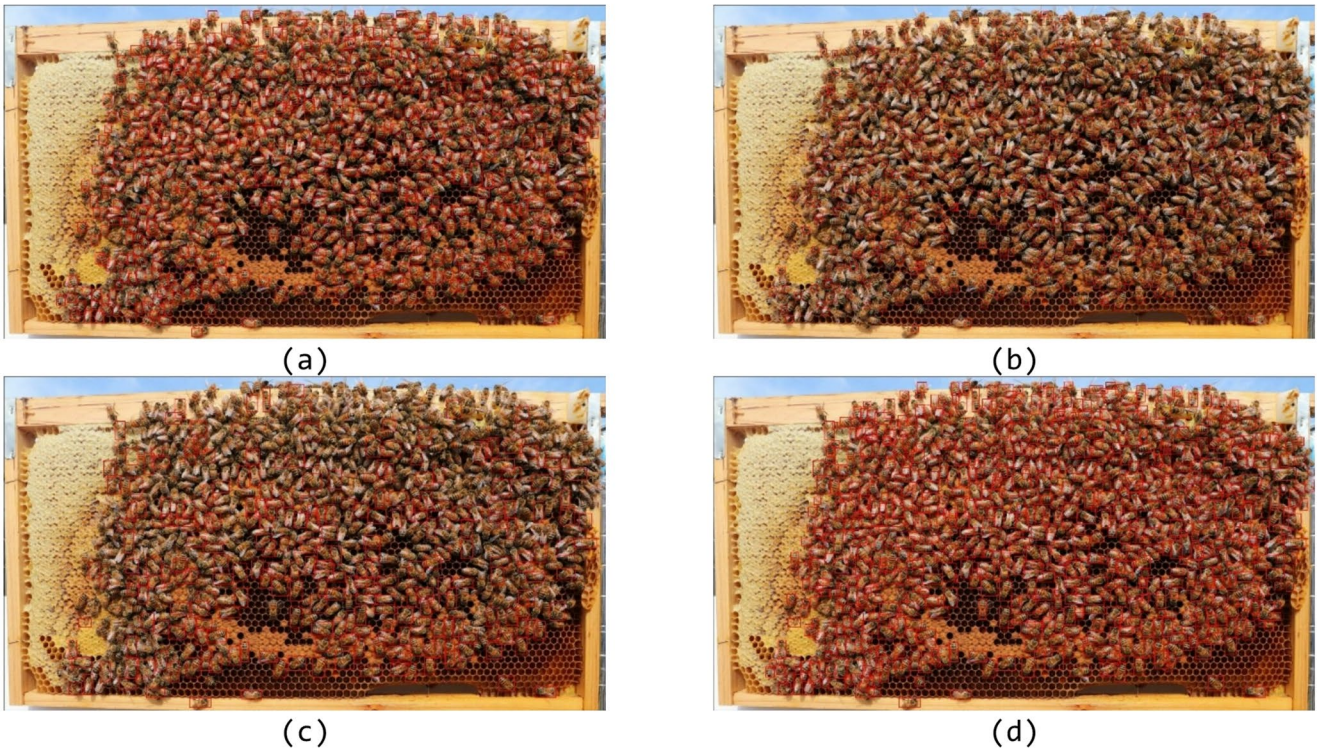

Fig. 5 Comparison of the detection results in high-density comb images: (**a**) head detector, (**b**) abdomen detector, (**c**) body detector, (**d**) result after matching process (proposed method)

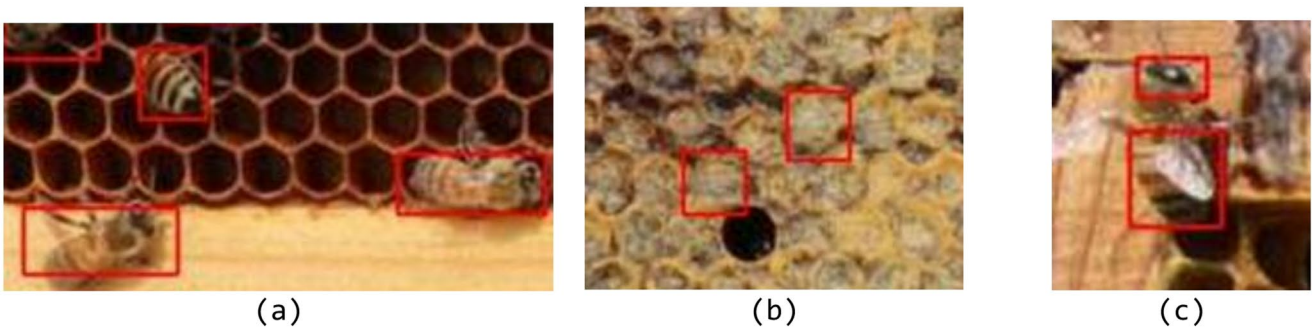

Fig. 6 Enlarged view of the detection results: (**a**) Honeybees not displaying the whole body were detected. (**b**) Honey was mistakenly detected as a bee in a low-density image (**c**) Two boxes (head and abdomen) in one honeybee because the whole body was not detected

images is shown in Fig. 5. Figure 6a shows an enlarged view of Fig. 5d, demonstrating that the abdominal detector can detect honeybees with their heads in the holes.

Comparison with State-of-the-Art Detector

To demonstrate that our part-based matching framework is detector-agnostic, we implemented the complete matching pipeline with YOLOv11 [16] in addition to our SSD-based implementation. YOLOv11 was trained on whole-body, abdomen, and head annotations using the same training data, with equivalent hyperparameters where applicable, with an input size of 288 × 288 pixels.

Table 5 shows the comparison results across different density levels. YOLOv11 achieves an overall F1-score of 0.781 with 0.689 recall and 0.901 precision for whole-body detection, demonstrating superior performance compared to our SSD baseline (F1 = 0.540). Applying our part-based matching framework to YOLOv11 further improves the overall F1-score to 0.790, with recall increasing from 0.689 to 0.732. This improvement is consistent with the recall improvement observed in the SSD-based proposed method (+51.1 percentage points over SSD whole-body), confirming that the matching framework provides recall improvements regardless of the underlying detector architecture.

Table 5 Comparison with state-of-the-art object detector

Method		Precision	Recall	F1-score
Low density	SSD baseline (whole-body)	0.664	0.612	0.637
	SSD + Matching (Ours)	0.325	0.969	0.486
	YOLOv11 (whole-body)	0.875	0.729	0.788
	YOLOv11 (+Matching)	0.460	0.946	0.619
Medium density	SSD baseline (whole-body)	0.936	0.341	0.500
	SSD + Matching (Ours)	0.714	0.964	0.820
	YOLOv11 (whole-body)	0.902	0.742	0.815
	YOLOv11 (+Matching)	0.879	0.876	0.878
High density	SSD baseline (whole-body)	0.963	0.303	0.461
	SSD + Matching (Ours)	0.792	0.873	0.831
	YOLOv11 (whole-body)	0.905	0.653	0.759
	YOLOv11 (+Matching)	0.926	0.645	0.760
Overall average	SSD baseline (whole-body)	0.884	0.421	0.540
	SSD + Matching (Ours)	0.612	0.932	0.711
	YOLOv11 (whole-body)	0.901	0.689	0.781
	YOLOv11 (+Matching)	0.858	0.732	0.790

The improvement is particularly evident in medium-density images (F1 = 0.878), where occlusion is most prevalent. These results empirically validate our claim that the part-based matching framework is detector-agnostic: applying the same matching strategy to a stronger baseline detector (YOLOv11) yields consistent recall improvements over whole-body detection alone.

Ablation Study

Table 6 presents an ablation study analyzing the contribution of each detector component within the YOLOv11-based framework. Among individual detectors, the abdomen detector achieves the highest overall F1-score (0.804) and recall (0.772), consistent with the SSD-based results in Table 4. This confirms that abdomen detection is the most effective single indicator for honeybee counting regardless of the base detector architecture. The head detector alone yields a substantially lower F1-score (0.470), reflecting the difficulty of detecting small and frequently occluded head regions.

Combining the wholebody detector with the abdomen detector (Wholebody + Abdomen) achieves an overall F1-score of 0.812 with recall of 0.791, representing a substantial improvement over wholebody detection alone (F1 = 0.781, Recall = 0.689). This demonstrates that abdomen-to-body matching effectively recovers bees missed by the wholebody detector. Adding the head detector (Wholebody + Abdomen + Head) further increases recall to 0.732 in the full proposed configuration, though the overall F1-score (0.790) is slightly lower than Wholebody + Abdomen alone (0.812), due to additional false positives introduced by head detections in medium- and high-density images. This trade-off between recall and precision is consistent with the findings observed in the SSD-based experiments.

These ablation results indicate that the abdomen detector contributes most to the matching framework, while the head detector provides marginal additional recall at the cost of reduced precision. The consistent trends observed across both SSD and YOLOv11 architectures further support the detector-agnostic nature of the proposed matching framework.

Matching Error Analysis

Error rates were calculated for each image to examine the effect of the Hungarian matching process in detail. The error rate is the number of incorrect pairs divided by the total number of pairs. Table 7 shows the results of abdomen-total matching, and Table 8 shows the results of head-total matching for all six test images. Note that these values are calculated by including BBs that are false positives.

Table 6 Ablation study of detector combinations using YOLOv11-based framework

Method	Precision	Recall	F1-score	Low F1	Med F1	High F1
Abdomen only	0.838	0.772	0.804	0.545	0.893	0.786
Head only	0.487	0.453	0.470	0.307	0.432	0.526
Wholebody only	0.901	0.689	0.781	0.788	0.815	0.759
Wholebody + Abdomen	0.834	0.791	0.812	0.555	0.892	0.800
Wholebody + Head	0.768	0.629	0.692	0.442	0.773	0.685
Wholebody + Abdomen + Head	0.858	0.732	0.790	0.619	0.878	0.760

Table 7 Abdomen-to-body matching error rates

Density	Low		Medium		High	
Image id	Image #1	Image #2	Image #3	Image #4	Image #5	Image #6
Matched pairs	28	68	90	227	259	300
Ground truth	26	57	84	219	245	276
Incorrect matchings	2	11	6	8	14	24
Error rates	7.1%	16.2%	6.7%	3.5%	5.4%	8.0%

Table 8 Head-to-body matching error rates

Density	Low		Medium		High	
Image id	Image #1	Image #2	Image #3	Image #4	Image #5	Image #6
Matched pairs	25	66	90	188	243	283
Ground truth	23	48	70	143	203	220
Incorrect matchings	2	18	20	45	40	63
Error rates	8.0%	27.3%	22.2%	23.9%	16.5%	22.3%

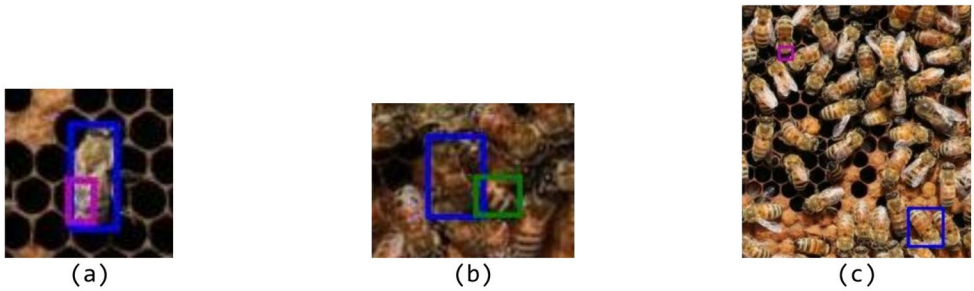

Fig. 7 Examples of incorrect matching. The blue, green, and magenta rectangles are the body, abdomen, and head, respectively: (**a**) pair with false detection, (**b**) slightly overlapped boxes, (**c**) non-overlapped boxes

Therefore, the error rate will also be high if there are many false positives. The error rate was less than 10% for matching between the abdomen box and the whole-body box, except for image #2. Conversely, the error rate for matching between the head box and the whole-body box was approximately 20%, except for Image #1. This difference can be attributed to the small number of head and body detections. It appears that the disadvantage associated with the diminished performance of the mapping method, resulting from the elevated number of false positives, exceeds the benefit of identifying honeybees with concealed abdomens. Examples of incorrect matches are shown in Fig. 7.

Discussion

Our experimental results demonstrate that the proposed part-based matching framework effectively addresses occlusion-induced false negatives in honeybee counting, independently of the underlying detector architecture. The SSD-based implementation substantially improves recall from 0.421 to 0.932 compared to SSD whole-body detection alone, confirming that part-level information is essential for robust counting in dense hive environments. More importantly, applying the same matching strategy to YOLOv11 yields a consistent recall improvement from 0.689 to 0.732, empirically validating the detector-agnostic nature of the framework. These findings suggest that the matching framework can be combined with any detector that produces bounding box predictions, and that pairing it with a stronger baseline such as YOLOv11 is expected to further improve overall counting accuracy.

Annotation Parts of the Honeybee

We discuss the parts of the bee that should be annotated based on the results of the three detectors. As shown in Table 4, the abdominal annotation demonstrated superior performance, as evidenced by the highest F1-score of the abdominal detector across all densities. In particular, all evaluation metrics were higher than 0.9 for medium-density images, indicating that highly accurate detection was achieved. However, the recall was the highest for the proposed method. Integrating the results from the head, abdomen, and body detectors for counting effectively reduces the number of undetected bees. The differences between the abdominal detector and the proposed method were 0.023, 0.055, and 0.066 for low, medium, and high densities, respectively. This finding suggested that the effect of the matching process increases as density increases.

Neither the head nor the body detector exhibited a very high detection accuracy. The head detector exhibited a higher F1-score than the body detector for medium and high densities. The body detector had fewer false positives, but the number of undetected bees increased as the density increased. A large number of undetected bees led to an increase in the number of false positives in the proposed method.

The detection results for low-density images exhibited lower precision and F1-scores than those for dense images, regardless of the detection method used. In low-density images, the color and shading of the comb, as well as the condition of the cells (especially nectar), can produce patterns that closely resemble those of some bees. This resulted in more false positives. The proposed method incorporated the abdomen and head boxes, which did not correspond to the body boxes. Consequently, it also included false positives in the three detectors.

The YOLOv11-based experiments further corroborate these findings. As shown in Table 6, the abdomen detector

achieves the highest F1-score (0.804) among individual YOLOv11-based detectors, consistent with the SSD-based results in Table 4. The head detector alone yields a lower F1-score (0.470), reflecting the same difficulty in detecting small and occluded head regions observed with SSD. Applying the matching framework to YOLOv11 improves recall from 0.689 to 0.732 compared to wholebody detection alone, demonstrating that the recall improvement from part-based matching persists across different detector architectures. These consistent trends across both SSD and YOLOv11 provide empirical evidence that the proposed matching framework is detector-agnostic, addressing the fundamental challenge of occlusion-induced false negatives independently of the underlying detection architecture.

Matching Process Issues

The causes of misdetection due to matching include (1) undetected and incorrect predictions at each detector and (2) incorrect matching. In Fig. 7a, the body box is matched with the incorrect detection of the head box. These errors may lead to incorrect matching of head boxes, resulting in redundant detections and subsequent double counting. Such issues are influenced by the accuracy of the object detection model and indicate a need for improvements in both training data and model architecture.

In addition, there were fewer detections in the whole body at high densities. As shown in Fig. 6c, the abdomen and head, but not the body, were detected. Therefore, even after matching, two boxes were detected for each bee. One of these two boxes was considered an incorrect detection. This is one of the main reasons for the low precision of the proposed method. The body of an insect can be divided into three parts. Determining whether the head and abdomen belong to the same individual is challenging because of the absence of anatomical continuity between these parts. It would be advantageous to establish a method for matching boxes based on their size and positional relationship independent of the body box.

Figure 7b shows an example of matching a bee detected by the body detector and the abdominal box of a neighboring bee. When using the Hungarian method, the rows in the cost matrix where all IoUs are zero are removed to exclude bounding boxes with no overlap from the matching; however, this error occurs because bounding boxes with even a small overlap are still targets of the matching. There are cases in which boxes of different individuals far apart match, as shown in Fig. 7c. Despite excluding nonoverlapping bounding boxes from the cost matrix, the following may cause such an error: The Hungarian method is an algorithm that finds the maximum matching in a bipartite graph, and one-to-many matching (i.e., multiple boxes corresponding to one box) is not possible. When a body box is matched to another individual's head box, as shown in Fig. 7b, other distant bodies and head boxes may remain as candidates for mapping. They were then mapped even though their IoUs were 0. These errors can be improved using the Hungarian method by adjusting the threshold of the IoU, which determines which BBs are excluded from the cost matrix, and by re-checking the overlap after matching.

Conclusions

In this study, we proposed a part-based detection and matching framework for automated honeybee counting from single comb images. We implemented the framework using SSD and compared detection accuracies with three different annotations for bee counting using a single comb image as the input. We also developed a method for integrating the results of each detection using the Hungarian algorithm. The experimental results indicated that the abdominal detector exhibited the highest F1-score and was the most effective for bee counting. The proposed method, which integrated the results of three detectors through a matching process, effectively reduced the number of undetected bees. However, this also includes false positives for each detector.

The following two factors should be considered to further enhance the detection accuracy of the proposed method: (1) the accuracy of each object detector and (2) the enhancement of the matching algorithm. It is possible to expand the training data, modify the structure of the object detection model, and optimize various hyperparameters to address the first point. In response to the second point, modifications to the cost matrix are necessary to prevent matching between boxes with minimal overlap. Additionally, devising a method to match the head and abdomen directly would help increase the accuracy of bee counting. While our implementation uses SSD as the base detector, the matching framework's design is independent of the specific detection architecture. Experiments with YOLOv11 confirmed that applying the same matching strategy to a stronger baseline detector consistently improves recall over whole-body detection alone, empirically validating the detector-agnostic nature of the proposed framework. Future work includes further optimization of the matching cost matrix to reduce false positives, particularly in low-density conditions, and evaluation on a larger and more diverse dataset to improve statistical robustness.

Appendix

The test images are listed below (six in total): low-density combs are shown in Fig. 8, medium-density combs in Fig. 9, and high-density combs in Fig. 10.

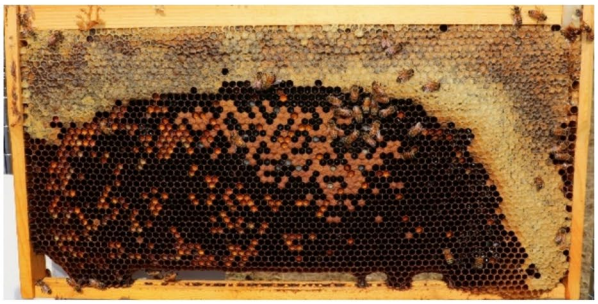

Image id #1

Image id #2

Fig. 8 Low density test images

Image id #3

Image id #4

Fig. 9 Medium density test images

Image id #5

Image id #6

Fig. 10 High density test images

Author contributions Conceptualization, K.O., M.H.; writing-original draft preparation, N.T., K.O.; writing-review and editing, N.T., M.H; methodology, K.O., M.H.; formal analysis, N.T., K.O., M.H.; funding acquisition, M.H.; All authors have read and agreed to the published version of the manuscript.

Funding This study was supported by Grants-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (JSPS), Grant Numbers 18K11346 and 21K11931.

Data Availability The code, pre-trained models, and test images used in this study are publicly available at https://github.com/tsurutana/partlevel_bee_detection.

Declarations

Conflict of Interest The authors have no competing interests to declare.

References

1. Odemer R. Approaches, challenges and recent advances in automated bee counting devices: a review. Ann Appl Biol. 2020;180(1):73–89. https://doi.org/10.1111/aab.12727.
2. Norouzzadeh MS, Morris D, Beery S, Joshi N, Jojic N, Clune J. A deep active learning system for species identification and counting in camera trap images. Methods Ecol Evol. 2021;12(1):150–61. https://doi.org/10.1111/2041-210X.13504.
3. Farjon G, Huijun L, Edan Y. Deep-learning-based counting methods, datasets, and applications in agriculture: a review. Precision Agric. 2023;24(5):1683–711. https://doi.org/10.1007/s11119-023-10034-8.
4. Davidson P, Steininger M, Lautenschlager F, Kobs K, Krause A, Hotho A. Anomaly detection in beehives using deep recurrent autoencoders. In: Proceedings of the 9th International Conference on Sensor Networks. SCITEPRESS - Science and Technology Publications; 2020: 142–149. https://doi.org/10.5220/0009161201420149.
5. Bilik S, Zemcik T, Kratochvila L, Ricanek D, Richter M, Zambanini S, et al. Machine learning and computer vision techniques in continuous beehive monitoring applications: a survey. Comput Electron Agric. 2024;217:108560. https://doi.org/10.1016/j.compag.2023.108560.
6. Bilik S, Janakova I, Ligocki A, Karel Horak DF. Computer vision approaches for automated bee counting application. IFAC-PapersOnLine. 2024;58(9):43–8. https://doi.org/10.1016/j.ifacol.2024.07.369.
7. Sindagi V, Yasarla R, Patel V. Pushing the frontiers of unconstrained crowd counting: New dataset and benchmark method. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE; 2019: 1221–1231 https://doi.org/10.1109/ICCV.2019.00131.
8. Arteta C, Lempitsky V, Zisserman A. Counting in the wild. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision – ECCV 2016. Springer International Publishing; 2016: 9911. https://doi.org/10.1007/978-3-319-46478-7_30.
9. Lu H, Cao Z, Xiao Y, Zhuang B, Shen C. TasselNet: Counting maize tassels in the wild via local counts regression network. Plant Methods. 2017;13(1):79. https://doi.org/10.1186/s13007-017-0224-0.
10. Gao G, Gao J, Liu Q, Wang Q, Wang Y. CNN-based density estimation and crowd counting: a survey; 2020. https://doi.org/10.48550/arXiv.2003.12783. arXiv:2003.12783.
11. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. SSD: single shot multibox detector. In Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision – ECCV 2016. Springer International Publishing; 2016:9905. https://doi.org/10.1007/978-3-319-46448-0_2.
12. Zhai S, Shang D, Wang S, Dong S. DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion. IEEE Access. 2020;8:24344–57. https://doi.org/10.1109/ACCESS.2020.2971026.
13. Zhou S, Qiu J. Enhanced SSD with interactive multi-scale attention features for object detection. Multimed Tools Appl. 2021;80(8):11539–56. https://doi.org/10.1007/s11042-020-10191-2.
14. Qi Z, Zhou M, Zhu G, Xue Y. Multiple pedestrian tracking in dense crowds combined with head tracking. Appl Sci. 2022;13(1):440. https://doi.org/10.3390/app13010440.
15. Kuhn HW. The Hungarian method for the assignment problem. Naval Res Log Quart. 1955;2(1–2):83–97. https://doi.org/10.1002/nav.3800020109.
16. Jocher G, Qiu J, Chaurasia A. Ultralytics YOLO11; 2024. https://github.com/ultralytics/ultralytics.