Data in Brief 52 (2024) 110060

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Labeled dataset for bee detection and direction estimation on entrance to beehive

Tomyslav Sledevič*, Dalius Matuzevičius

Department of Electronic Systems, Faculty of Electronics, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania

ARTICLE INFO

Article history:
Received 14 November 2023
Accepted 8 January 2024
Available online 11 January 2024

Dataset link: Labeled dataset for bee detection and direction estimation on beehive landing boards (Original data)

Keywords:
Pattern recognition
Honeybee tracking
Object detection
Segmentation
Pose estimation
Convolutional neural networks
Image processing

ABSTRACT

The datasets for bee detection, pose estimation and segmentation consist of organized folders containing both images and corresponding labels. The detection dataset comprises a total of 7200 individual frames collected at 8 different beehives. The pose dataset contains 400 images of bees annotated with two key points per bee. The first point marks a head, second point marks a stinger. All frames have a resolution of 1920×1080 pixels. The segmentation dataset contains 2300 cropped images of bees. These cropped images are annotated with triangular markers that aid in estimating directional vectors. The labels in all proposed datasets were saved in YOLO format. The labeling process was automated by training YOLOv8 model on a set of manually annotated images for bee detection. After detection, all the labels were visually revised and corrected. Frames were captured using stationary mounted camera 30 cm above beehive landing boards. The data collection period spanned from June to July 2023 in Vilnius district.

* Corresponding author.
E-mail address: tomyslav.sledevic@vilniustech.lt (T. Sledevič).

https://doi.org/10.1016/j.dib.2024.110060

Specifications Table

Subject	Computer Vision and Pattern Recognition
Specific subject area	Computer Vision in Ethology. It pertains to the use of computer vision methods to detect, track, and estimate the direction of bees as they interact at the entrance of a beehive.
Data format	Raw, Annotated
Type of data	Image, Labels
Data collection	The images of hive landing boards were collected in the local apiary in Vilnius distinct at June-July 2023. Frame resolution 1920×1080 px. Frames were captured using smartphone mounted 0.3 m above 8 different beehive landing boards as presented in Fig. 6. Bees in detection dataset were labeled with rectangular bounding boxes using *labellmg* tool. Bees in pose dataset were annotated with two key points per bee that mark a head and a stinger. Segmentation dataset contains cropped images of bees labeled with triangle shape for direction vector estimation using *labelme* tool. Labels were saved in YOLO format.
Data source location	Institution – Vilnius Gediminas Technical University City – Vilnius Country – Lithuania Coordinates – 54.728, 25.358
Data accessibility	Repository name: Mendeley Data identification number doi:10.17632/8gb9r2yhfc.5 Direct URL to data: https://data.mendeley.com/datasets/8gb9r2yhfc

1. Value of the Data

- The publicly provided dataset will be useful to researchers to improve machine learning models for bee detection on native landing boards.
- The data may be reused for conducting experiments related to bee detection, counting, direction estimation, tracking, analysis of bee movements, and behavior patterns recognition.
- The data can be used by ethologists and relevant researchers to study the behavior of individual insect or colony at the entrance to the beehive.
- The dataset is useful to build bee tracking applications which are beneficial for apiarists, farmers, and agriculture industries.

2. Background

If an experienced beekeeper can track the condition of bee family only by looking at the entrance to the hive, then the machine vision algorithm can do it too. To train neural network models that detect, track and count bees at the entrance to the hive, we need a dataset of images, where all the bees are annotated in frames with location and orientation labels. The original motivation for compiling this dataset was to track the condition of multiple hives without disturbing the work of bee families. To improve bee detectability, most existing open-access datasets contain additional boards used in the background of the camera field of view. The lack of annotated images of beehive entrances with native landing boards motivated us to collect, label, and share a dataset through a community of researchers for further development of methods dedicated to recognize bee behavior just by observing the entrance to the hive.

3. Data Description

Fig. 1 presents data structure of the repository. The dataset contains detection, segmentation and pose repositories. The detection dataset contains 8 folders with 7200 frames labeled for bee detection. Images and labels are stored in separate subfolder. The annotations are saved in YOLO

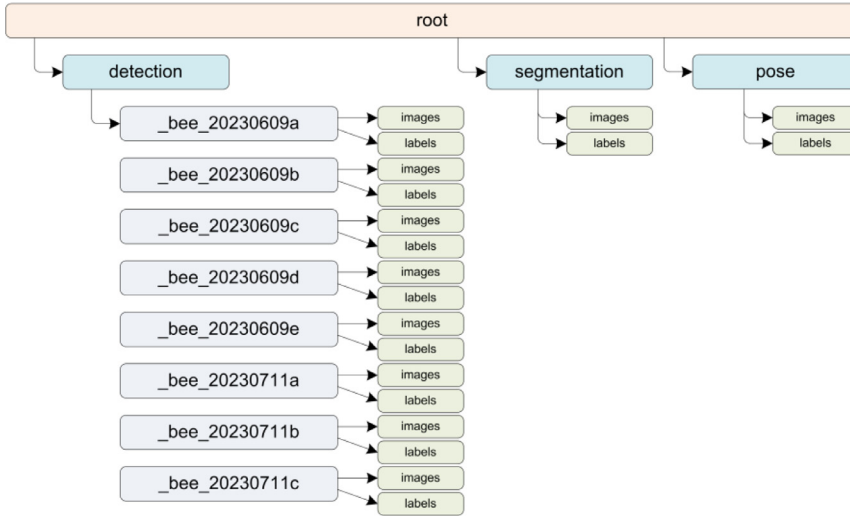

Fig. 1. Dataset repository structure.

format as a text files with the same names as image files. Each text file contains the annotations of bees in the corresponding image file, that is object class, object's center point coordinates in 2D space, width and height normalized to the image resolution.

Fig. 2 shows the samples of 8 different beehive entrances with rectangular shape labeled bees in the bee detection dataset. Bees are labeled in all the locations in the frame independently where they appear, on the hive front wall, on the entrance ramp, or on the grass. If bees partially overlap with each other, then labels overlaps also. If only part of bee is visible due to their transition through entrance hole or if bee appear partially on the border of frame, then only visible part of bee is labeled.

The segmentation dataset contains 2300 cropped images of bees labeled with triangle shape for direction vector estimation. The annotations are saved in YOLO format. Each text file contains single record, that is object class, and coordinates of three point in 2D space. The size of the images varies from 46×27 px to 231×192 px. Bees in the dataset are under various orientations, lighting conditions, background, blurred, and partially overlapped, as shown in Fig. 3.

A samples of triangle shape annotated bees are depicted in Fig. 4a. Red triangles show how the annotations look in 2D space. The most acute angle of the triangle shows the direction of the bee. The segmentation model must be trained on the labeled dataset to estimate the direction of bee. After successful segmentation the model gives a contour with a shape similar to the triangle in Fig. 4a, however, with rounded corners, as shown in Fig. 4b with approximated red contours. The arithmetic mean of all points of a red shape corresponds to a center of mass (centroid) and can be computed using, i.e., *cv2.moments()* function from *OpenCV* library in *Python*. The bee contour (red) is fitted to the minimal area rectangle (blue), i.e., using *cv2.minAreaRect()* function. We can assume that the bee direction vector is parallel to a longer edge of the rectangle. Then the closest corner of the rectangle to the center of the mass is a start point of the direction vector. The end point of the direction vector is the next corner lying on the farthest edge of the rectangle. To visualize bee direction in an output file, the start point of the direction vector is moved to a center of the mass, and the direction vector is added to a frame on top of each detected bee in Fig. 4b.

The pose dataset contains 400 images of bees labeled with rectangular shape bounding box for bee detection and two points for direction vector estimation. Annotation format in pose dataset is as follows: class=0, x, y, w, h, px_1, py_1, px_2, py_2, where x, y are center coordinates, w – width, h – height of bounding box, then px and py mark coordinates of first and second

Fig. 2. Images of 8 different beehive entrances with labeled bees in the bee detection dataset.

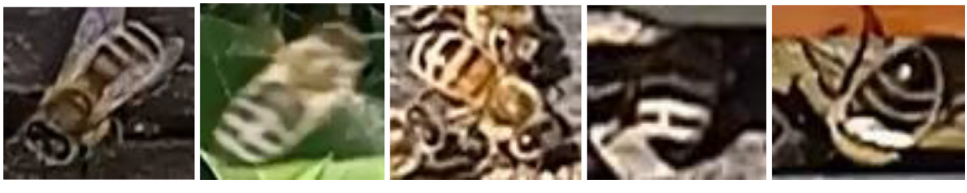

Fig. 3. Cropped images of bees in the segmentation dataset.

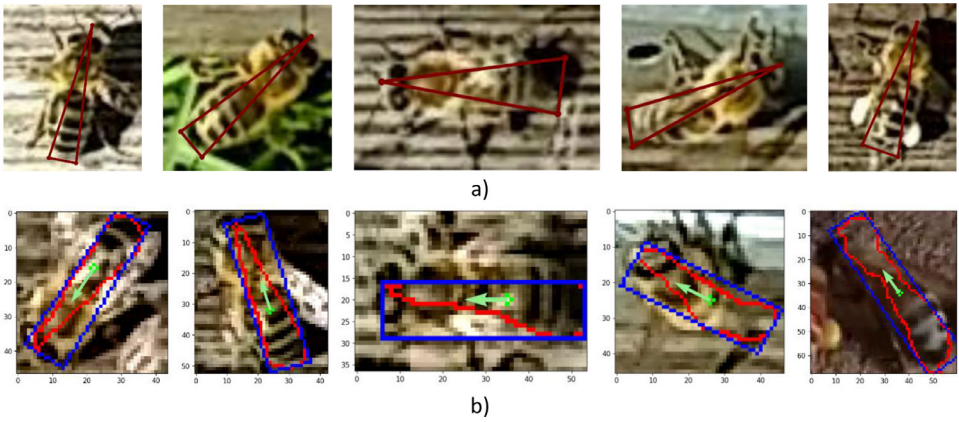

Fig. 4. Triangle shape annotated bees using *labelme* tool for direction estimation (a). Segmented bee contour (red), minimum area rectangle (blue), bee direction vector (green) (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Images of beehive entrances with key points and bounding box labeled bees in the bee pose dataset.

key point. Pose dataset contains 400 images of 8 beehives entrances (50 images per beehive). If bee is fully visible then first point (blue) marks a head, second point (violet) marks stinger, as shown in Fig. 5. If only part of bee is visible then first point marks the front part of bee, second point marks the back.

4. Experimental Design, Materials and Methods

The images were acquired on 8 different beehives 0.3 m above the landing boards not disturbing the foraging process. No extra artificial background was added behind the landing boards. The aim was to acquire records of bee at the entrance of beehives using native landing boards, and then train a convolutional neural network model for pollen detection [1], bee detection, tracking [2] and behavior identification. Camera was stationary mounted to horizontally adjustable stick and a tripod for stability, as shown in Fig. 6. The data was captured in a sunny and cloudy days at June-July 2023. The camera recorded a video files at 50 fps and 1920×1080 px resolution, and then images were extracted from the video records. The detection dataset was labeled using *labelImg* tool. Segmentation dataset was labeled using *labelme* tool. The labeling process was semi-automated by training YOLOv8m model [3] on a set of manually labeled images for bee detection. After automated detection, all the labels were visually revised and corrected. The proposed dataset was used to train bee detection and segmentation models, an-

Fig. 6. Data acquisition on local apiary.

Table 1
Repositories and count of labeled images.

Repository name	Number of images
./detection/_bee_20230609a/	700
./detection/_bee_20230609b/	1000
./detection/_bee_20230609c/	800
./detection/_bee_20230609d/	720
./detection/_bee_20230609e/	1140
./detection/_bee_20230711a/	1320
./detection/_bee_20230711b/	910
./detection/_bee_20230711c/	610
Total in detection dataset:	7200
./segmentation/	2300
./pose/	400

alyze bee movements, and develop algorithms for behavior patterns recognition through visual monitoring of the entrance to the beehive.

Table 1 presents count of labeled images in bee detection, segmentation and pose datasets.

Limitations

Not applicable.

Ethics Statement

The data is available in public. No ethics approval needed for this study. There is no conflict of interest.

Data Availability

Labeled dataset for bee detection and direction estimation on beehive landing boards (Original data) (Mendeley Data)

CRediT Author Statement

Tomyslav Sledevič: Data curation, Validation, Writing – original draft; **Dalius Matuzevičius:** Conceptualization, Methodology, Writing – review & editing.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

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

References

[1] T. Sledevič, A. Serackis, D. Plonis, FPGA implementation of a convolutional neural network and its application for pollen detection upon entrance to the beehive, Agriculture 12 (11) (2022) 1849, doi:10.3390/agriculture12111849.
[2] T. Sledevič, D. Plonis, Toward bee behavioral pattern recognition on hive entrance using YOLOv8, in: 2023 IEEE 10th Jubilee Workshop on Advances in Information, Electronic and Electrical Engineering, IEEE, 2023, pp. 1–4, doi:10.1109/AIEEE58915.2023.10134563.
[3] G. Jocher, A. Chaurasia, J. Qiu, YOLO by Ultralytics (Version 8.0.0) [Computer Software], 2023. https://github.com/ultralytics/ultralytics.