Varroa mite detection using deep learning techniques

Jose Divasón[1], Francisco Javier Martinez-de-Pison[1], Ana Romero[1], Pilar Santolaria[2], and Jesús L. Yániz[2]

[1] University of La Rioja, Spain
{jose.divason,fjmartin,ana.romero}@unirioja.es
[2] University of Zaragoza, Spain
{psantola,jyaniz}@unizar.es

Abstract. The varroa mite is a major problem for beekeeping today because it threatens the survival of hives. This paper develops deep learning methods for detecting varroa in images to monitor the level of infestation of the hives in order to use treatments against varroa in time and save the bees. The ultimate goal is its implementation by beekeepers. Therefore, the deep learning models are trained on pictures taken by smartphone cameras covering the entire board where both pupae and varroas are placed. This makes the object detection task a challenge, since it becomes a small object detection problem. This paper shows the experiments that have been developed to solve this challenge, such as the use of super resolution techniques, as well as the difficulties encountered.

Keywords: varroa mites · beekeping · small object detection

1 Introduction

Varroa destructor, commonly known as varroa mites, are a parasitic species that feed on the bodily fluids of honey bees. These mites (see Figure 1) are considered one of the most significant threats to honey bee health, as they weaken and damage the bees' immune systems, making them more susceptible to disease and other stressors. In addition to causing direct harm to individual bees (both pupae and adults), varroa mites also transmit various viruses that can further weaken colonies and lead to their collapse. In fact, *varroosis* is currently the most damaging disease in beekeeping worldwide. In the European Union it is endemic, being the only beekeeping disease that requires systematic treatment of bee colonies in order to keep parasitization rates below harmful thresholds.

Fig. 1: A varroa destructor

Given the crucial role that honey bees play in pollination and the global food supply, the development of effective and efficient detection methods for varroa mites is therefore critical to the survival of honey bees and the industries that rely on their pollination services. One promising avenue for varroa detection is through the use of deep learning techniques applied to images of honey bees. By training neural networks to recognize the unique characteristics of varroa mites, researchers and beekeepers could quickly and accurately identify infested colonies and take appropriate action to treat and manage the infestation. The development of such methods has the potential to improve the way beekeepers monitor and protect their hives, ultimately helping to protect the future of honey bees and the vital ecosystems they support. Different methods have been proposed to count varroa mites in hives. Three of them are explained below, which can be combined with artificial vision algorithms (see Figure 2).

(a) Sticky board method (b) Sugar shake method (c) Pupae method

Fig. 2: Cropped images of the three methods.

- Bottom board and sticky board: it consists of placing on the bottom of the hive a board and a white card on it with a sticky substance to which the mites that fall on it stick.
- Sugar shake or alcohol wash: invasive methods that require washing the bees with powdered sugar, if they are to be kept alive, or with alcohol or liquid soap, sacrificing them. Bees are collected and then they are washed; a large part of the varroa mites are removed from them, which can be counted on the container in which they remain.
- Bee breeding counting: a hive frame is extracted during the breeding period; then, varroa mites and pupae are counted.

This work is mainly based on the last method. The main advantage of this method is that, in addition to being very reliable, the varroa infestation rate during the breeding season can be known and thus can be treated earlier. The models will also be analyzed with a few images of the sticky board method. Regarding the image analyses concerning the three approaches, all methods share one fact: large images are captured, but very small objects have to be detected.

2 Related work

2.1 Object detection and small object detection

Object detection is a fundamental problem in computer vision that deals with identifying and localizing objects of interest in a digital image. Deep learning techniques, particularly convolutional neural networks (CNNs), have revolutionized this field and achieved state-of-the-art results in recent years [21,16,31]. Small Object Detection (SOD) is a particular instance of object detection that is focused on detecting small-size (or tiny) objects. This issue is particularly relevant in fields such as biology (where small objects like cells are relatively small compared to the input image), satellite images [18], drone scenes [10,28] and more [12,27,29]. Many more challenges appear in small object detection tasks compared to object detection [6].

For instance, *noisy feature problems* could appear since small objects usually have low resolution, and this often causes neuronal networks to have problems learning good representations from their blurred shapes. In addition, due to the typical structure of object detectors (a backbone combined with a detection head), there is usually an *information loss*: the feature extractor component usually reduces the size of the feature maps and tends to learn high-dimensional features. This is particularly critical with small objects, because they are inevitably seen as very few pixels within the network. In fact, the standard Faster R-CNN architecture has an effective stride of 16 i.e., a 16×16 object is seen as a single pixel by the region proposal network (RPN).

Fig. 3: One image with a varroa mite on a adult bee, taken from [3].

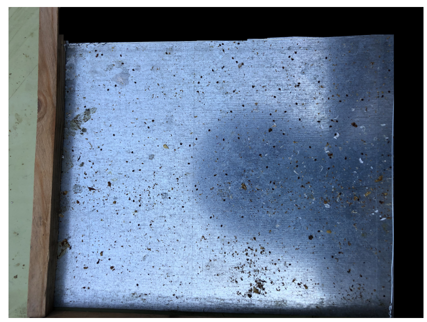
Different approaches have been investigated to try to alleviate these problems [6,5]. For instance, one of the main techniques is the use of specific data-augmentation strategies [30]; indeed, some authors copy instances of the small objects and paste them in different positions of the image with one or several random transformations [13,4]. Moreover, another powerful technique is the use of super-resolution [25,7] to partly reconstruct the blurry appearance of small objects and even the introduction of Generative Adversarial Networks (GANs) to generate new visually similar data to feed the algorithms.

2.2 Varroa mite detection and deep learning

There are some works related to adult honey bees detection, monitorization [26,14], tracking the movement of pollens [22] and others [2,1,19] with deep learning. However, less research has been done for varroa mite detection and counting. Probably, the most similar work to us is the one by Bilil *et al.* [3], in which YOLOv5 and Single Shot Detector (SSD) are used to detect and distinguish between the healthy and the infected bees. Concretely, the starting point is an existing dataset [23] that is modified by including other public images to define six classes: healthy bee, the bee with pollen, drone, queen, varroa mite infected bee and varroa mite. The images have a high resolution and are restricted to adult bees, see [3, Figure 2] and Figure 3. In contrast, our interest relies on varroa mite detection on pupae and working with photos that can be taken easily with a mobile device by the beekeeper. An initial attempt on pupae and varroa mite detection was developed in a master thesis [11]. The results were not fully satisfactory, mainly due to poorly adjusted bounding boxes.

In recent years there have also been mobile applications that use AI to count varroa[3]. However, either the results are not good or they are commercial apps where the models are not publicly available. In addition, none of them is suitable for varroa mite detection on pupae.

(a) One of the 21 images.



(b) One of the 5 images.

Fig. 4: Two images from the dataset.

3 Methods

3.1 Dataset creation

The dataset contained 21 images with a resolution of 4032×3024 pixels and captured by a smartphone camera. The capture method was the third one presented in the introduction, during the months of February to June 2022, and with different lighting conditions. The images contained both pupae and varroa mites from different hives, being the number of varroa mites highly variable in each image (ranging from 2 to 187). A total of 732 varroa mites were identified by experts from these images. Figure 4a is one of such images and Figure 2c a 512×512 pixels crop.

In addition, 5 images of the board method (first method described in introduction) were also available, and obtained under the same conditions as above.

[3] For instance, `https://beemapping.com`, `www.beescanning.com` and `https://apisfero.org`.

These images contained varroa mites (a total of 460), but no pupae appeared. However, they also contained many other noisy artifacts like dust, dirt and soil. Figure 4b is one of such images and Figure 2a a 512×512 pixels crop.

3.2 Metrics

The Intersection over Union (IoU) is a measure of overlap between two bounding boxes, defined as the ratio of the area of overlap between the two bounding boxes to the area of their union. If the IoU between the predicted bounding box and the ground truth bounding box is above a certain threshold, the predicted bounding box is considered a true positive.

The Average Precision (AP) measures the quality of the detection output by computing the area under the precision-recall curve. Obviously, AP is dependent on the IoU threshold used to determine whether a predicted bounding box is a true positive or a false positive. The mean average precision (mAP) is the average of AP of all classes. The mAP score is usually calculated ranging different IoU, i.e., mAP corresponds to mAP@[0.5,0.95,0.05] that is the average AP for IoU from 0.5 to 0.95 with a step size of 0.05. Similarly, mAP50 represents the mAP computed at a fixed IoU threshold of 0.5.

Analogously, the mean average recall (mAR) is the recall averaged over different IoU. This work used both metrics for evaluating the models, based on the `torchmetrics` Python package.

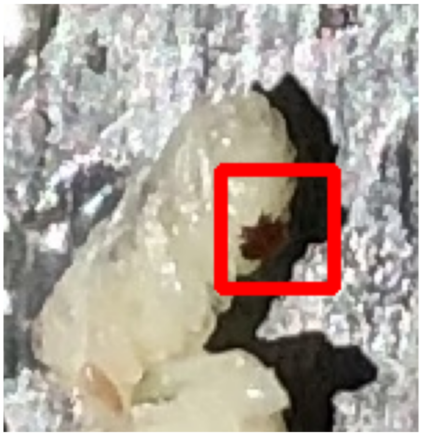
3.3 Deep learning methods

Current state-of-the-art approaches are based on CNN object detectors. There are two main families of detectors: one-stage methods (such as the YOLO family [20], SSD [17] and EfficientDet [24]) and two-stage methods (the R-CNN family [9,8,21]).

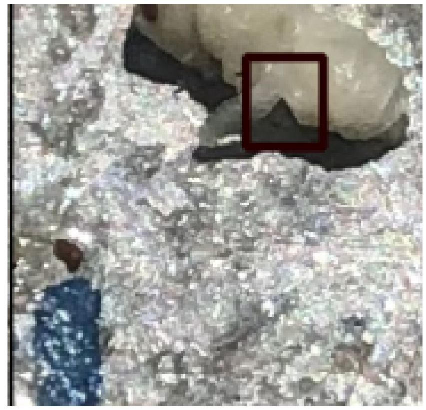
In two-stage detections, one part of the network (the RPN) generates candidate object proposals (the candidate bounding boxes), and the other part analyzes them, ranks their likelihood to be a true positive, and classifies and locates the objects inside. One-stage detectors, on the other hand, directly predict the class and location of objects without the need for a separate proposal generation stage. It is kwown that one-stage detectors are generally faster and more flexible but may sacrifice some accuracy. Two-stage detectors are more accurate, particularly for small objects. However, they are slower and require more training data. Thus, since the varroa mites are a small object detection challenge (with low resolution), it made more sense to use a Faster R-CNN approach.

We also tried several state-of-the-art neuronal networks as different backbones (the part of a neural network that is responsible for feature extraction) for the Faster R-CNN, such as ResNet50 FPN, EfficientNet B0 to B7, MobileNetV3 Large FPN, and so on.

Due to the low resolution of the images, the small objects and the difficulty to distinguish sometimes varroa mites from soil or other artifacts, we tried



(a) A too broad bounding box.



(b) A wrong varroa mite prediction.

Fig. 5: Training with wide bounding boxes.

super-resolution techniques to improve the quality of the images (see Figure 9). Specifically, we used Enhanced Deep Residual Networks [15], to multiply by 2, 3 and 4 the resolution of the input images.

4 Experiments and Results

Experiments were performed on a computing server with two NVIDIA RTX 3090 GPUs.

As a metric we used mAP50, i.e., detection was considered successful when the IoU was at least 50%. We chosen this metric over mAP, because beekeepers are interested in the number of varroa mites, rather than in a perfect fit of the bounding boxes.

A baseline approach with the default Faster R-CNN configuration and parameters achieved a mAP score of 0.1543. Thus, models did not learn well with the initial experiments. Analyzing why this was happening, we detected that the bounding boxes created by the experts were too large. We concluded that the detection of small objects was very sensitive to the quality of the bounding boxes, not only because a small variation of a few pixels greatly affected the metric, but also because too large a bounding box caused the neural networks to learn from the background and not from the desired object.

Thus, bounding boxes had to be manually tuned and adjusted properly. Figure 5a shows an example of a too broad bounding box, which makes neuronal network to learn from the contour of the pupae and the shades, instead of learning from the varroa mite. Figure 5b shows precisely a wrong prediction caused by underadjusted bounding boxes.

Multiple tests were performed to find out the best model configuration. The main decisions and parameters that showed the best performance are the following:

1. As varroa mites tend to be reddish, We performed a preprocessing of the images that proved to improve the quality of the detection, specifically we decreased the intensity of the blue and green channels of each image to an 80%, with no modifications on the red channel. This improved the detection results.
2. The learning rate was set to 0.01, but *ReduceLRonPlateau* technique was used to reduce the value a factor of 0.75 when the metric stopped improving during 10 consecutive epochs, see Figure 6. Stochastic gradient descent (SGD) was used for optimizing the objective function. The maximum number of epochs was 300.
3. We performed data augmentation in each batch. The techniques that worked best were random 224×224 crop (ensuring a balance between the number of crops with and without varroa mites), rotation, horizontal flip, vertical flip, a soft random brightness (limited to 0.05) and random contrast (limited of 0.05).
4. The best backbone was *Resnet50-FPN*, where the anchor generator parameter was set to 35. Some pre-trained weights were tested (such as those ones based on the COCO dataset), but the best results were obtained with no pre-trained weights.

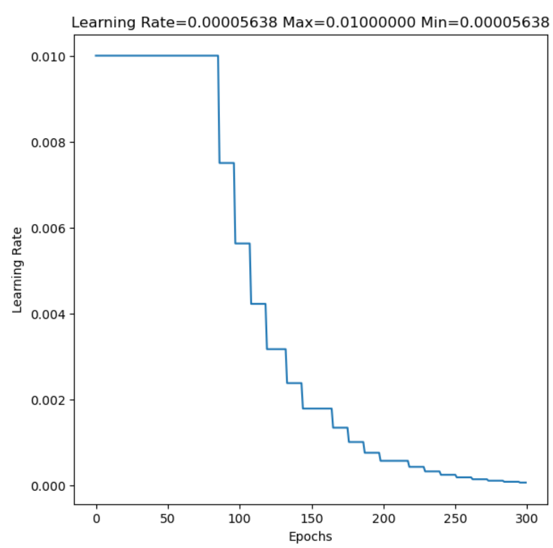


Fig. 6: Reduce Learning Rate on Plateau strategy for the best model.

Figure 7 shows both the mAP and mAR evolution during the training of the best model. The final mAP50 score was 0.7368 and mAR10 was 0.4452.

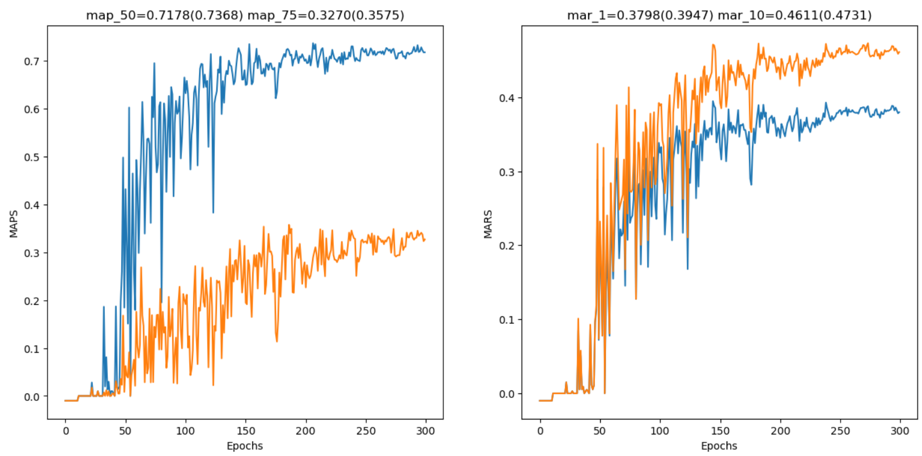
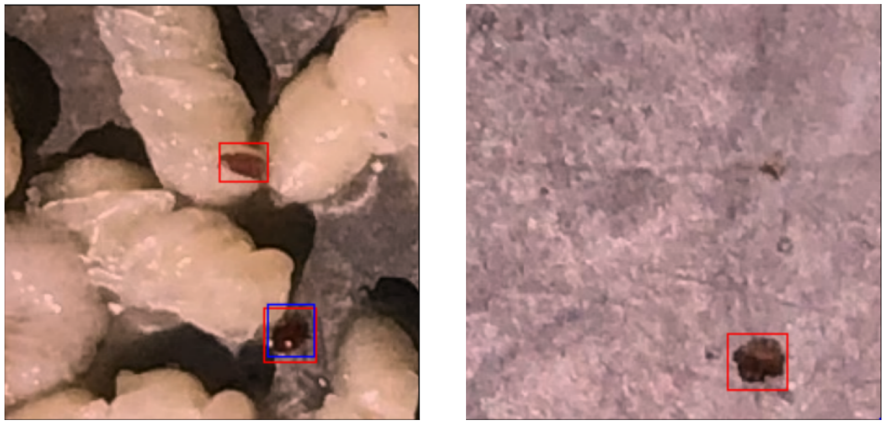


Fig. 7: mAP and mAR evolution during the training of the best model.

Although the metrics show a decent performance, a detailed analysis of the predictions shows that the model makes some blatant errors. In fact, the validation images contains a total of 192 varroa mites, of which the model is able to detect 163. The model fails with artifacts and pupae eyes, indeed there are 96 false positives. If one distinguish the validation images between those with and without pupae, 91 of 95 varroa mites are detected when no pupae appears (with 18 false positives), and 72 of 97 varroa mites are detected when pupae appears, but with 78 false positives. This means that recall is not good enough. Figure 8a shows an example of errors in the prediction, where a pupae eye is mistaken for a varroa mite. Figure 8b shows another error, where a ground stain is mistaken for another varroa mite. These examples show that the problem is truly difficult. In fact, depending on the positions of the pupae, the experts themselves find it difficult to distinguish between eyes or varroa mites, being the main problem the resolution of the images. Although the images have been taken with good resolution smartphone cameras (12 Mpx), the level of detail of small objects is rather low and this causes that there is not enough gradient for the neural network to learn the features.

Trying to overcome this problem, some super-resolution techniques (explained in Section 3.3) were performed. Figure 9 shows an example of a varroa mite before and after the application of the super-resolution method. As can be seen, visually the image quality seems to have improved quite a bit. The varroa mite has a much more defined contour, however it can be appreciated that the neuronal network behind the super resolution has *invented* some parts of the varroa



(a) An eye confused with a varroa mite. (b) A dark spot mistaken for a varroa mite.

Fig. 8: Errors in predictions. Red bounding boxes are the varroa mite predictions. Blue boxes are the true bounding boxes.

mite. Similar behavior was observed in other elements: dark spots, dust, dirt and soil.

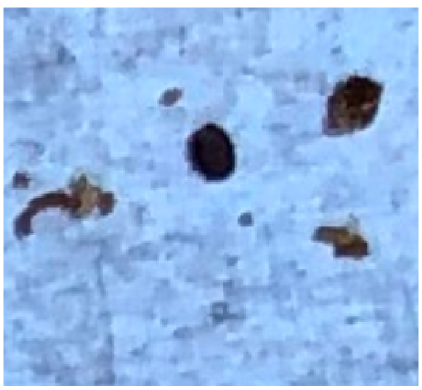
None of the super-resolution techniques ($\times 2$, $\times 3$, $\times 4$) were able to improve the results. Indeed, results were even slightly worse. Super resolution does not improve images enough for the network to learn to distinguish correctly the varroa mites. In fact, visually varroa mites are very similar to some spots and pupae eyes (even if super resolution is applied), being also quite difficult for humans to distinguish among them.

(a) Crop of an input image. (b) Crop after applying super-resolution.

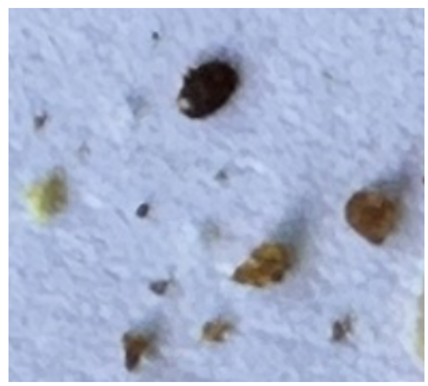
Fig. 9: Before and after super-resolution techniques.

5 Conclusions and further work

This paper presents an approach to varroa mite detection by means of artificial intelligence techniques, from pictures captured with smartphones. The similarity between the eyes of bee pupae and varroa mites, in addition to the low resolution of the input dataset, pose a challenge of small object detection. Different techniques have been employed, including super resolution, to alleviate these problems, obtaining decent results. As future work, to further improve the system and avoid false positives (for instance, when varroa mites are mistaken for dark spots), beekeeping experts are currentling compiling a new dataset by taking closer images, so that each board is divided into several photos (and thus, the resolution is improved). A crop of these new images is shown in Figure 10b, where it is much easier to distinguish the varroa mite from the spots than in Figure 10a.



(a) Crop of an image of the original dataset. It is difficult to distinguish the varroa mite from the soil and spots.



(b) Crop of an image of the new dataset. The resolution is much better and makes it easier to distinguish varroa mites.

Fig. 10: Comparison between old and new images.

Acknowledgments

This work is supported by grants PID2020-112673RB-I00, PID2020-116641GB-I00 and PID2021-123219OB-I00 funded by MCIN/AEI/ 10.13039/501100011033, and the DGA-FSE (grant A07-17R).

References

1. Alves, T.S., Pinto, M.A., Ventura, P., Neves, C.J., Biron, D.G., Junior, A.C., De Paula Filho, P.L., Rodrigues, P.J.: Automatic detection and classification of

honey bee comb cells using deep learning. Computers and Electronics in Agriculture **170**, 105244 (2020)

2. Bilik, S., Bostik, O., Kratochvila, L., Ligocki, A., Poncak, M., Zemcik, T., Richter, M., Janakova, I., Honec, P., Horak, K.: Machine learning and computer vision techniques in bee monitoring applications. arXiv preprint arXiv:2208.00085 (2022)
3. Bilik, S., Kratochvila, L., Ligocki, A., Bostik, O., Zemcik, T., Hybl, M., Horak, K., Zalud, L.: Visual diagnosis of the varroa destructor parasitic mite in honeybees using object detector techniques. Sensors **21**(8), 2764 (Apr 2021). https://doi.org/10.3390/s21082764
4. Chen, C., Zhang, Y., Lv, Q., Wei, S., Wang, X., Sun, X., Dong, J.: Rrnet: A hybrid detector for object detection in drone-captured images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 0–0 (2019)
5. Chen, G., Wang, H., Chen, K., Li, Z., Song, Z., Liu, Y., Chen, W., Knoll, A.: A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. IEEE Transactions on Systems, Man, and Cybernetics: Systems **52**(2), 936–953 (2022). https://doi.org/10.1109/TSMC.2020.3005231
6. Cheng, G., Yuan, X., Yao, X., Yan, K., Zeng, Q., Han, J.: Towards large-scale small object detection: Survey and benchmarks. arXiv preprint arXiv:2207.14096 (2022)
7. Deng, C., Wang, M., Liu, L., Liu, Y., Jiang, Y.: Extended feature pyramid network for small object detection. IEEE Transactions on Multimedia **24**, 1968–1979 (2021)
8. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
9. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. pp. 580–587 (2014). https://doi.org/10.1109/CVPR.2014.81
10. Gupta, H., Verma, O.P.: Monitoring and surveillance of urban road traffic using low altitude drone images: a deep learning approach. Multimedia Tools and Applications pp. 1–21 (2022)
11. Higuera Pinillos, N.: Detección de varroa y pupas de abejas mediante procesamiento de imágenes y aprendizaje profundo (2022), Master Thesis, Universidad de La Rioja
12. Huang, H., Tang, X., Wen, F., Jin, X.: Small object detection method with shallow feature fusion network for chip surface defect detection. Scientific reports **12**(1), 3914 (2022)
13. Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., Cho, K.: Augmentation for small object detection. arXiv preprint arXiv:1902.07296 (2019)
14. Kulyukin, V., Mukherjee, S.: On video analysis of omnidirectional bee traffic: Counting bee motions with motion detection and image classification. Applied Sciences **9**(18), 3743 (2019)
15. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)
16. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)

17. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. pp. 21–37. Springer (2016)
18. McAllister, E., Payo, A., Novellino, A., Dolphin, T., Medina-Lopez, E.: Multispectral satellite imagery and machine learning for the extraction of shoreline indicators. Coastal Engineering p. 104102 (2022)
19. Ngo, T.N., Rustia, D.J.A., Yang, E.C., Lin, T.T.: Automated monitoring and analyses of honey bee pollen foraging behavior using a deep learning-based imaging system. Computers and Electronics in Agriculture **187**, 106239 (2021)
20. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
21. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)
22. Rodriguez, I.F., Megret, R., Acuna, E., Agosto-Rivera, J.L., Giray, T.: 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. IEEE (2018)
23. Schurischuster Stefan, Kampel, M.: Varroa dataset (2020), `https://zenodo.org/record/4085044`
24. Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10781–10790 (2020)
25. Wang, Z., Chen, J., Hoi, S.C.: Deep learning for image super-resolution: A survey. IEEE transactions on pattern analysis and machine intelligence **43**(10), 3365–3387 (2020)
26. Yang, C.R.: The use of video to detect and measure pollen on bees entering a hive. Ph.D. thesis, Auckland University of Technology (2018)
27. Yu, X., Gong, Y., Jiang, N., Ye, Q., Han, Z.: Scale match for tiny person detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1257–1265 (2020)
28. Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., Ling, H.: Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 7380–7399 (2021)
29. Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., Hu, S.: Traffic-sign detection and classification in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2110–2118 (2016)
30. Zoph, B., Cubuk, E.D., Ghiasi, G., Lin, T.Y., Shlens, J., Le, Q.V.: Learning data augmentation strategies for object detection. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16. pp. 566–583. Springer (2020)
31. Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J.: Object detection in 20 years: A survey. Proceedings of the IEEE (2023)