Towards Precise Recognition of Pollen Bearing Bees by Convolutional Neural Networks

Fernando C. Monteiro[1 [0000−0002−1421−8006]], Cristina M. Pinto, and José Rufino[1⋆ [0000−0002−1344−8264]]

Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
{monteiro,rufino}@ipb.pt
https://cedri.ipb.pt

Abstract. Automatic recognition of pollen bearing bees can provide important information both for pollination monitoring and for assessing the health and strength of bee colonies, with the consequent impact on people's lives, due to the role of bees in the pollination of many plant species. In this paper, we analyse some of the Convolutional Neural Networks (CNN) methods for detection of pollen bearing bees in images obtained at hive entrance. In order to show the influence of colour we preprocessed the dataset images. Studying the results of nine state-of-the-art CNNs, we provide a baseline for pollen bearing bees recognition based in deep learning. For some CNNs the best results were achieved with the original images. However, our experiments showed evidence that DarkNet53 and VGG16 have superior performance against the other CNNs tested, with unsharp masking preprocessed images, achieving accuracy results of 99.1% and 98.6%, respectively.

Keywords: Pollen bearing bees · Convolutional neural network · Deep learning.

1 Introduction

Honey bees and other insect pollinators are an essential component of ecosystems, being necessary for the successful reproduction of a wide variety of flowering plants, including agricultural crops [8]. The recognition of bee foraging behaviour brings an important input to identify the colony balance and health [4].

The scientific observation of honey bees activities within and outside their colonies began nearly a century ago [7]. In this seminal work, Lundie referred to the extreme difficulty to obtain accurate records by counts made at the entrance of the hive. Therefore, he claimed the use of some mechanical means to automatically register the exit and return of the bees over long periods of time. Despite his visionary proposals, the traditional technique still remains the human observation and manual annotation, as it is the only approach that enables the

⋆ This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UIDB/05757/2020.

extraction of a wide range of behaviours available to bee specialists [1]. As such, developing image and video acquisition at the entrance of the hive, and recognizing bees bearing pollen, enables automatizing bee foraging behaviour analysis, which is of great interest for ecological and ethological studies [2,10,13].

Pattern recognition from images has a long and successful history. In recent years, deep learning, and in particular Convolutional Neural Networks (CNNs), has become the dominant machine learning approach in the field of computer vision and, to be specific, in image classification and recognition. While still being a classification based on a combination of key image features, this new approach involves a model determining and extracting the features itself, rather than these being predefined by human analysts. Since the number of labelled pollen bearing bees images in the publicly available datasets is too small to train a CNN from scratch, transfer learning can be employed.

This paper analyzes the performance of 9 state-of-the-art deep learning CNNs (VGG16, VGG19, ResNet50, ResNet101, InceptionV3, Inception-ResNetV2, Xception, DenseNet201 and DarkNet53) in classifying images of pollen bearing and non-bearing bees. As colour is *a priori* a relevant feature for the presence of pollen, CNNs were trained and tested using different colour image preprocessing techniques. Besides original images, grayscale images were used, as well as Contrast Limit Adaptive Histogram Equalization and Unsharp Masking techniques.

The rest of the paper is organized as follows: in the next section, related work is presented; Section 3 provides an overview of the CNNs used; Section 4 describes the proposed approach and the experimental setup; this is followed by the results analysis in Section 5, and conclusions and future work in Section 6.

2 Related Works

In this section, we review some papers that provide pollen bearing bees recognition systems and define computer vision techniques as baselines.

Babic [2] used background subtraction, colour segmentation and morphology methods for honey bees segmentation. Classification of pollen bearing bees and bees that do not have pollen load is done using a nearest mean classifier, with a simple descriptor consisting of colour variance and eccentricity features. This achieved a correct classification rate of 88.7% with 50 training images per class.

Stojnić [14] proposed an approach that starts by segmenting honey bees from images using two segmentation methods based on colour descriptors. Then, Support Vector Machines (SVM) are trained on few variations of VLAD and SIFT descriptors to classify the images into two classes (with or without pollen). Finally, the classification results are evaluated using Area Under a Curve (AUC), obtaining a score of 91.5% for a dataset of 1000 images.

Rodriguez [10] proposed a colour vision system for detecting pollen-bearing bees, using a Convolutional Neural Network (CNN) for detecting pollen on bees entering the hive. To highlight most of the pollen balls in the dataset, the colour information is filtered using a Gaussian model. Using three models of supervised

classification (KNN, Naive Bayes and SVM), and three deep CNN architectures (VGG16, VGG19 and ResNet50), accuracy varies from 50% to 96%.

Sledevic [13] presented the classification of images with pollen bearing bees using a simple scratched CNN with a sufficient configuration required for future implementation on a low-cost FPGA. The three hidden layer CNN was selected as a trade-off between performance and number of required arithmetic operations, yielding an accuracy of 94%.

Yang and Collins [17] applied deep learning techniques to pollen sac detection and measurement on honeybee monitoring video. The pollen sac detection model is built using a faster RCNN architecture with VGG16 core network. This pollen detection model is then combined with a bee tracking model, so that each flying bee tracked on successive video frames is identified as bearing pollen or not. The classification score obtained was 93%.

3 Convolutional Neural Networks Architectures

Convolutional Neural Networks (CNN) is a type of deep learning model for processing data with a grid pattern (e.g., images), inspired by the organization of animal visual cortex and designed to automatically and adaptively learn spatial hierarchies of features through back-propagation by using multiple building blocks, like convolution layers, pooling layers, and fully connected layers from low to high level patterns. It is especially suited for image processing as it uses 2D hidden layers to convolve the features with the input data. The main strength of CNN is that it suppresses the need for feature extraction by automatically extracting the more discriminant features of a set of training images.

Next, we provides an overview of the main features of the CNNs used in this study. We choose nine popular architectures due to their performance on several classification tasks as well as on ImageNet dataset [11].

VGG16 and **VGG19** architectures [12] are 16 and 19 layers deep networks, respectively, on 224×224 RGB images input. They use 3×3 kernels in every convolution layer. They have five convolutional blocks where the first two blocks have two convolution layers and one max-pooling layer in each block. The remaining three blocks of the network have three fully-connected layers equipped with the rectification (ReLU) non-linearity and the final softmax layer. The main difference between VGG16 and VGG19 is the number of convolutional layers.

ResNet50 and **ResNet101** [5] also apply to 224×224 RGB image inputs, and have some design similarities with the VGG architectures with convolutional layers with 3×3 filters with convolution stride of one pixel, except the first convolutional block, whose filter is 7×7 with convolution stride of two pixels. The batch normalization is adopted right after each convolution layer and before activation. These architectures introduce the *residual block* whose aim is to address the degradation problem observed while training the networks. In the residual block the identity mapping is performed, adding the output from the previous layer to the non-linear layer ahead. The network finishes with a global average pooling layer and a fully-connected layer with the softmax function.

In **Inception-V3** [16], the first three convolutional layers have 3×3 filters with convolution strides, respectively, of two, one, and one pixel. These convolutional layers are followed by max-pooling and other three convolutional layers with 3×3 filters. This network has three inception modules where the resulting output of each module is the concatenation of the outputs of three convolutional filters with different sizes. The goal of these modules is to capture different visual patterns of different sizes and approximate the optimal sparse structure. Finally, before the final softmax layer, an auxiliary classifier acts as a regularization layer.

Inception-ResNet-V2 [15] combines residual connections and the Inception architecture. Since Inception networks tend to be very deep, they are hard to train because of the notorious vanishing gradient problem - as the gradient is back-propagated to earlier layers, repeated multiplication may make the gradient indefinitely small, so Inception-ResNet-V2 replaced the filter concatenation stage of the Inception architecture with residual connections as in ResNet.

The **Xception** architecture [3] has an initial convolutional layer with 3×3 filters with a convolution stride of 2 pixels on 299×229 RGB image input. The architecture has three blocks in a sequence where are carried out convolution, batch normalization, ReLU, and max pooling operations. Besides, at the output of each block, residual connections are made as in ResNet50.

DenseNet201 [6] is built from dense blocks and pooling operations, where each dense block is an iterative concatenation from all previous layers. Within those blocks, the layers are densely connected together: each layer gets the input from all preceding layers and passes on its own feature maps to all subsequent layers. This extreme reuse of residuals creates deep supervision because every layer receives more supervision from the previous layer and thus the loss function will react accordingly which makes it a more powerful network. Like on ResNet, the first block has 7×7 filters with a convolution stride of two pixels on 224×224 RGB images input; however, max-pooling in this convolutional block is carried out over a 3×3 pixel window.

Darknet53 [9] has 53 layers deep and acts as a backbone for the YOLOv3 object detection approach for a 256×256 image input. This network uses successive 3×3 and 1×1 convolutional layers with shortcut connections (introduced by ResNet to help the activations propagate through deeper layers without gradient vanishing) to improve the learning ability. Batch Normalization is used to stabilize training, speed up convergence, and regularize the model batch.

3.1 Transfer Learning

Constraints of practical problems, such as the limited size of training data, refrain the performance of deep CNNs, trained from scratch, to be satisfactory. Since there is so much work that has already been done on image recognition and classification, we can use transfer learning technique to solve the problem. With transfer learning, instead of starting the learning process from scratch, we can start from patterns that have been learned when solving a similar problem.

In deep learning, transfer learning is a technique whereby a CNN model is first trained on a large image dataset with a similar goal to the problem that is being

solved. One or more layers from the trained model are then used in a new CNN, trained with sampled images for the current task. This way, the learned features in re-used layers may be the starting point for the training process and adapted to predict new classes of objects. Transfer learning has the benefit of decreasing the training time for a CNN model and can result in lower generalization error due to the small number of images used in the training process.

The weights in re-used layers may be used as a starting point for the training process and adapted in response to the new problem. This usage treats transfer learning as a type of weight initialization scheme. This may be useful when the first related problem has a lot more labelled data than the problem of interest and the similarity in the structure of the problem may be useful in both contexts.

4 Pollen Bearing Bees Dataset

The PollenDataset[1] [10] used in this work was created in 2018 for classifying pollen bearing honeybees obtained at hive entrance. This dataset features high resolution images (180×300) (as illustrated in Figure 1) of segmented honey bees with a total of 714 labelled images (369 Pollen and 345 Non-Pollen bearing).

Fig. 1. PollenDataset samples: bees with (1st row) and without (2nd row) pollen bag.

As described in [10], the images were extracted and manually annotated from videos using a protocol defined to avoid near-duplicate samples, remove

[1] https://github.com/piperod/PollenDataset

misaligned samples and ensure a balanced and representative dataset. The orientation of the bees was compensated to ensure in all image samples that the bee is facing upwards. The authors did not preprocessed the images.

5 Experimental Setup

The nine CNN architectures adopted were set up to use the fine-tuning strategy as well as the stochastic gradient descent with momentum optimizer at their default values; also, dropout rate was set at 0.5, early-stopping was used to prevent over-fitting, and the learning rate was established at 0.0001. Additionally, the CNNs batch size was configured to be 12 and they were trained with 30 epochs. This specific batch size was chosen to consume less memory and train the CNNs faster since it allows to update the network weights more often. During training, all images go through a heavy data augmentation which includes horizontal and vertical flipping, 360^o random rotation, rescaling factor between 70% and 130%, and horizontal and vertical translations between -20 and $+20$ pixels. All CNNs were trained using the MatConvNet MATLAB toolbox in a virtual machine with a pair of NVIDIA RTX 2080 Ti GPUs attached, hosted at the CeDRI cluster.

We adopted 5-Fold Cross-Validation on the dataset, such that in each fold the dataset was split on 70% (499 images) for training and 30% (215 images) for test, allowing the CNN models to be iteratively trained and tested on different sets. Since the testing set is composed of images not seen by the model during the training step, this allows to anticipate the CNN behaviour against new images.

5.1 Colour preprocessing techniques

As colour is *a priori* a relevant feature for the presence of pollen, we used different colour image preprocessing techniques, as exemplified in Fig. 2. Grayscale images where used to remove the colour information. Contrast limit adaptive histogram equalization (CLAHE) improved the low contrast issue, without producing noise, by changing the slope of the function that is relating input image intensity value to desired resultant image intensities. Unsharp masking is a linear image processing technique which sharpens the image. The sharp details are identified as a difference between an original image and its blurred version.

6 Results and Discussion

This section shows the performance of the deep learning networks trained on individual bee images. Each network was applied for pollen bearing bee detection on the 215 images validation set (111 pollen bee images and 104 non-pollen bee images). To study the influence of colour features in the recognition process, besides using the original images, CNNs were also trained and tested with the results produced by different colour image preprocessing techniques applied on those images: grayscale images, local contrast enhancing and unsharp masking.

(a) (b) (c) (d)

Fig. 2. Colour image preprocessing. (a) Original version. (b) Grayscale version. (c) After contrast limit adaptive histogram equalization. (d) After unsharp masking.

The training step produces a network model for pollen bearing bee detection images. The network is applied for pollen detection on the validation images producing correct classification (true positive or true negative) results and incorrect classification (false positive or false negative) results, as shown in Fig. 3.

(a) (b) (c) (d)

Fig. 3. Example of pollen recognition results. (a) A true positive result. (b) A true negative result. (c) A false positive result. (d) A false negative result.

An interesting observation is that the false negative samples have smaller pollen bags than the true positive samples, which confirms the intuition that the reduced size of pollen bags increases the difficulty of the classification.

The classification results for Precision, Recall, F1-score and Accuracy for the nine state-of-the-art CNN architectures considered, with different colour preprocessing techniques, are presented in Table 1. The numbers exhibited in bold indicate the best F1-score and Accuracy results obtained for each architecture.

The highest F1-score and Accuracy, both of 99.1%, were achieved by the DarkNet53 architecture with the unsharp masking preprocessing technique, show-

Table 1. Classification results (in percentage) through 5-Fold Cross-Validation on the test set for different CNNs and preprocessing techniques considered.

CNN	Original	Segmented	Seg. Equal.	Seg. CLAHE
VGG16	94.8	99.1	96.9	96.7
VGG19	98.2	98.2	**98.2**	**98.1**
ResNet50	86.6	99.1	92.4	91.6
ResNet101	94.7	96.4	**95.5**	**95.4**
Inception V2	92.0	93.7	**92.9**	**92.6**
Inception V3	98.0	87.4	92.4	92.6
Xception	86.0	82.9	84.4	84.2
DenseNet201	95.6	97.3	**96.4**	**96.3**
DarkNet53	99.1	96.4	97.7	97.7

ing that the rates of true positives, true negatives, false positives and false negatives did not present large distortions. On the opposite extreme, the Xception model achieved the lowest results of all CNNs, in all the experiments.

The generality of the CNNs used in this study considerably improve the results presented in [10] for the same dataset. In that work, researchers obtained the best results with an accuracy of 96%. However, they removed several images from the dataset without given any motives or information on that removal.

The validation score obtained with the DarkNet53 network for the images in the test dataset has a mean value of 96.7% with a variance of 1.2% for non-pollen images, and a mean of 98.8% with a variance of 0.4% for pollen images. This network produces only one false positive and one false negative with scores of 96.3% for the false positive and 68.7% for the false negative.

The grayscale preprocessing technique produced the worst evaluation results in all networks due to its lack of colour information, achieving the best result with ResNet101 with an Accuracy of 94.4%.

The CLAHE approach achieved good results, but its increasing of brightness enhanced the yellow regions of the bees body, which can be confused with pollen bags, thus biasing the training process.

As the unsharp masking technique enhances image details, this may help the training process, allowing the learning of different image characteristics. Thus, this technique achieved the best results for most of the networks, and good results even when the networks produced the best results for the original images.

Although the number of images in the dataset is still low, we foresee increasing the classification capacity of the tested approaches by expanding the number of images available via data augmentation with rotation, translation and scaling.

Also, the high values for the four evaluation metrics in all CNNs show that the number of correctly identified images is high when compared to the number of tested images. We believe that a 99.1% value for all the four evaluation metrics is enough to build an automatic classification system of pollen bearing bees, since the visual classification of that bees performed by humans is a hard task.

7 Conclusion

In this paper, an automated pollen bearing bee recognition approach is proposed. To promote research in the automation of pollen bearing bees classification, we report performance for nine pre-trained CNN topologies, applied in four different colour preprocessing techniques. We identify evidence that DarkNet53 has superior performance against other CNNs tested, specially with unsharp masking preprocessing technique, achieving an F1-score and Accuracy of 99.1%. VGG network produces the best evaluation results for original images with an Accuracy of 98.2% against the 97.7% of DarkNet53. This study proves that using the CNN architecture defined for the PollenDataset image collection, allows good classification results when used in a transfer learning approach. In the future, we plan to combine different CNNs in order to further improve the performance.

References

1. Abou-Shaara, H.: The foraging behaviour of honey bees, apis mellifera: a review. Veterinarni Medicina **59**(1), 1–10 (2014)
2. Babic, Z., Pilipovic, R., Risojevic, V., Mirjanic, G.: Pollen bearing honey bee detection in hive entrance video recorded by remote embedded system for pollination monitoring. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences **III-7**, 51–57 (Jun 2016)
3. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1800–1807 (2017)
4. Frias, B., Barbosa, C., Lourenço, A.: Pollen nutrition in honey bees (apis mellifera): impact on adult health. Apidologie **47**, 15–25 (2016)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
6. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2261–2269 (2017)
7. Lundie, A.: The flight activities of the honneybee. United States Department of Agriculture **1328**, 1–38 (1925)
8. Madras Majewska, B., Majewski, J.: Importance of bees in pollination of crops in the European Union countries. In: International Conference Economic Science for Rural Development. pp. 114–119 (2016)

9. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. ArXiv p. 1804.02767 (2018)
10. 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. pp. 314–322 (2018)
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A., Fei-Fei, L.: ImageNet large scale visual recognition challenge. Int. J. of Computer Vision **115**, 211–252 (2015)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Int. Conference on Learning Representations. pp. 1–14 (2015)
13. Sledevič, T.: The application of convolutional neural network for pollen bearing bee classification. In: 2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE). pp. 1–4 (2018)
14. Stojnić, V., Risojević, V., Pilipović, R.: Detection of pollen bearing honey bees in hive entrance images. In: 17th International Symposium INFOTEH-JAHORINA (INFOTEH). pp. 1–4 (2018)
15. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inceptionresnet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 4278–4284 (2017)
16. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2818–2826 (2016)
17. Yang, C., Collins, J.: Deep learning for pollen sac detection and measurement on honeybee monitoring video. In: 2019 International Conference on Image and Vision Computing New Zealand. pp. 1–6 (2019)