Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 222 (2023) 415–424

Procedia Computer Science

www.elsevier.com/locate/procedia

International Neural Network Society Workshop on Deep Learning Innovations and Applications (INNS DLIA 2023)

BeeNet: An End-To-End Deep Network For Bee Surveillance

John Yoo[a,*], Rumali Siddiqua[b,*], Xuehan Liu[a], Khandaker Asif Ahmed[d,**], Md Zakir Hossain[a,c,d,**]

[a]*The Australian National University, Canberra ACT 2600, Australia*
[b]*BRAC University, Dhaka 1212, Bangladesh*
[c]*Curtin University, Bentley WA 6102, Australia*
[d]*Commonwealth Scientific and Industrial Research Organisation, Canberra ACT 2601, Australia*

Abstract

Computer vision-based image classification plays a vital role in developing surveillance tools for measuring the biological behavior of bees and their disease detection. Native bees often face numerous environmental threats, ranging from invasive bees to numerous parasitic diseases, which affect not only the existing ecosystem but also the booming honey and wax industries. Numerous ML-based, pre-trained models showed potential in bee classification and monitoring tasks, but heavily curated data-set and closed-set models hinder their applicability in-field monitoring tasks. In this paper, we proposed a deep learning model to obtain improved levels of feature representations of eleven economically important bee species, fine-grained object (e.g. parasite, pollen) detection for bee-health monitoring and gradually progress to an end-to-end model to provide a solution for bee surveillance. Our model can extract learned feature representations from publicly available complex back-grounded images and propose similar usage on other domains through a qualitative analysis to learn appropriate defining features, specifically for morphological classification. In particular, we utilize a variant of the transformer encoder-decoder architecture with the incorporation of extracted image features from a ResNet50 network. Our model obtained 92.45% classification accuracy on the bee species classification task and up to 99.18% on fine-grain object detection sub-tasks. Besides the classification task, our end-to-end model can detect varroa pests and pollen on bee images with 94.50% and 99.18% accuracies. Our model outperformed other existing models for bee surveillance or health monitoring tools. We also discussed the applicability of our BeenNet model in real-time settings. Overall, our end-to-end model has implications in both computer vision and biological computing tasks, such as visual feature extraction, in-domain classification, and sub-task identification. It will also serve as a baseline for future bee monitoring tools and a multi-modal model for disease detection.

Keywords: Bee Classification; Bee Health Monitoring; *Apis mellifera*; Varroa detection; Pollen Detection

* These authors contributed equally
** Corresponding authors: zakir.hossain1@curtin.edu.au; Khandakerasif.ahmed@csiro.au

1877-0509
10.1016/j.procs.2023.08.180

1. Introduction

Bees are the key pollinators of agricultural crops, promoting the long-term development and sustainability of our agriculture. They are important and effective pollinators because of their ability to transport large amounts of pollen grains on their hairy bodies, their reliance on floral resources, and the semi-social or eusocial nature of some species [1]. The potential importance of bees for agricultural pollination has been highlighted as one of the main reasons for wild bee conservation [2,3]. Honey bees are the most significant commercial pollinators of agricultural crops, which account for 35% of global food production and rely on insect pollination for reproduction. There has been a definite decline in both wild and domesticated bee populations around the world in recent years, with risks including habitat loss, increased pesticide usage, parasite and virus expansion, and climate change. Recently, Australian biosecurity authorities discovered the destructive parasite,*Varroa destructor* at Newcastle Port - which feeds on both developing and adult honey bees *(Apis mellifera)*. There has been a rush of research concentrating on the mechanisms of bee decline and the implications of delivering ecological systems due to the decline in bee populations [4]. Identification of bees down to species level would enable a more in-depth understanding of bee population dynamics in particular locations, allowing for more targeted measures to be implemented to preserve and enhance their health and influence on the larger ecosystem.

Traditional bee morphology research [5] was primarily concerned with the classification of several bee species. Classification is a common task that entails the manual gathering of morphometric data, e.g. - wing, and whole-body images, and the identification of the species by an expert entomologist. There have recently been many DNA-based [6] methods used to lessen the workloads in the bee classification of organisms, however, there is still a lack of approaches for bee morphological studies. Current advancements in machine learning models have provided the ability to precisely extract distinguishing features from general images, which may eliminate the requirement for manual curation in morphological studies in the future. Recent developments in machine learning models [7,8] have made it possible to accurately identify classifying features from general images, thereby eliminating the need for manual curation in morphological studies. Convolutional Neural Networks (CNN) [9,10] for image analysis, in particular, have developed models with identification accuracy equivalent to or better than humans in a variety of visual recognition tasks [11-13]. When employing a general dataset to distinguish between similar subordinate-level categories, CNN models showed good accuracy (90%+) [14]. Convolutional neural networks (CNN) [15,16], in particular, have become the standard method for deep learning models, with success in a wide range of pattern and image recognition applications. Both the DeepABIS [17] and ABIS [7] models used various CNNs to automatically build features from bee wing images in order to categorize them up to species and subspecies level, but both require a highly curated wing data set. Since these curating techniques are very labor-consuming and, depending on the number of bee species, perhaps impossible, data gathering and feasibility for such extensive models become extremely low. Although there are limited and particular images of bee wings, there is a significant public data set of general images of bees in a variety of environments. As such, due to the immutable obstacles and challenges associated with feature-specific data expansion such as bee wings, much of the study questing for efficient bee supervision and preservation has been engrossed by the use of generic bee images. This is particularly prevalent in studies that employ machine learning models for image classification tasks. With the escalation in the need for a more efficient, coherent, and robust framework for bee monitoring and conservation, more extensive work with the aid of image-based machine learning models has become imperative.

Moreover, the image-based machine learning model is capable of identifying bees and their distinctive morphological traits. This will help us to monitor population dynamics and levels of decline more quickly and conveniently, equipping us with the data we need to develop conservation measures. Further, fine-grained image classification [18,19] is an essential sub-task with object detection and species identification applications. Recent advances in the field have seen the implementation of the successful transformer architecture from NLP tasks to image classification with Vision Transformer. As a result of a large amount of data, the correct classification of the bee species is a key factor in providing accurate information to biologists while analyzing it.

In this paper, we developed a transformer-based model, namely BeeNet, which explored the computationally demanding task of training a vision transformer on a relatively smaller data set of bee images and experimented with the model using bee-related data sets and sub-tasks. We developed an end-to-end model incorporating this architecture to learn and classify bee images and identify non-bee images from publicly available data. Additionally, our proposed

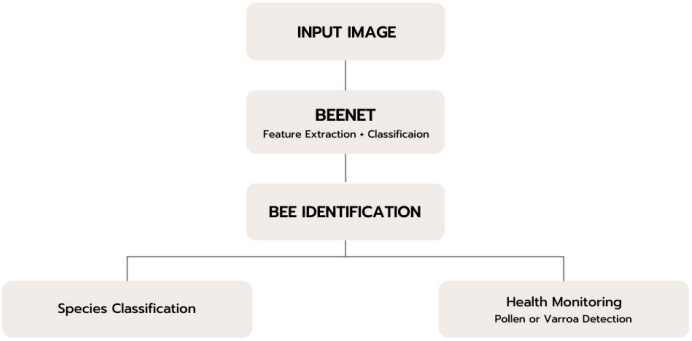

Fig. 1: General idea of BeeNet model

model also engages other sub-tasks which include bee species classification and health conditions depending on the presence of pollen or varroa; curated to make the model more extensive and robust. Moreover, the scope of this study involves the use of convolutional neural networks in conjunction with the transformer-encoder model to perform feature extraction and image classification respectively. Although transformer-based classifiers and conventional CNN models individually demonstrate noteworthy classification performances, their combination can incorporate good results. For this purpose, our work is not confined to just one of the models, helping us to focus on important observations about our classification task. Due to the conjunction, our model outperformed other state-of-the-art CNN and Vision Transformer models in terms of classification accuracy values. Thus, our work can serve as a reliable and robust technique to aid in bee species' health monitoring and preservation.

2. Materials and methods

2.1. Dataset and Data Pre-processing

We have explored the fine-grained image classification task in the domain of bees using three datasets [20]. The first and most focused data set, the classification dataset, uses images from iNaturalist.org, which contains images of various creatures with taxonomic annotations or classifications ranging from 92.3% to 97.3% correct [21]. We have conducted an exhaustive literature search and identified the world's three most important bee genera, namely - *Apis*, *Xylocopa*, and *Bombus*. Based on the number of images available, we chose the top eleven species from these three genera: *A. mellifera, X. virginica, X. micans, X. sonorina, X. tabaniformis, X. violacea, B. griseocollis, B. impatiens, B. pensylvanicus, B. terrestris*, and *B. vosnesenskii*.

The number of images was balanced based on the lowest available images (2,245 images) of *X. micans*, and images were downloaded in bulk using an in-house python script. Finally, there are a total of 24,695 images of bees in the data set. In addition to these bee data set, we also constructed a validation data set of 20,610 images from the Diptera order (the same family as bees) from the same source. Images were randomly selected from a range of 6,512 species. We employ this data set to test the end-to-end model's ability to distinguish bees from non-bees using biologically similar (fly) samples.

For the domain subclass, we used two other datasets, namely BeeImage Dataset [22]. These additional data sets are used to evaluate non-species classification tasks for the models mainly used for bee health monitoring. Firstly, the "Varroa dataset", contains 5172 images of honey-bee, with the presence and absence of a parasitic mite called

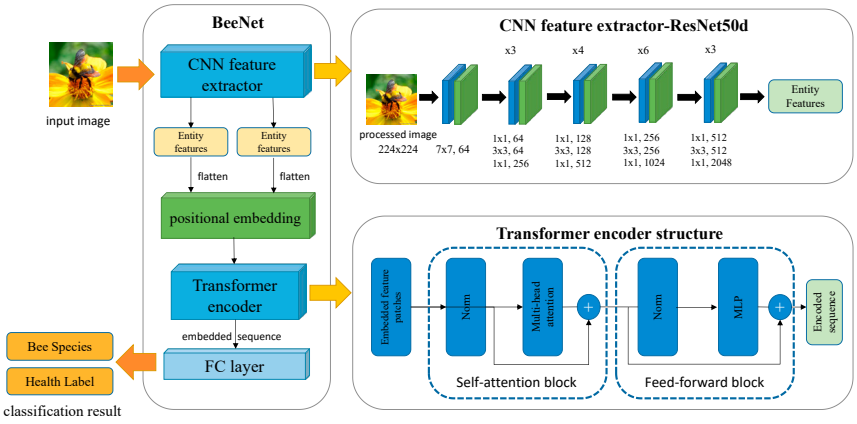

Fig. 2: Our proposed BeeNet model architecture

Varroa destructor. Secondly, the ”pollen dataset” [22], a smaller pollen dataset with 714 images divided into two categories: presence or lack of pollen on the body of honey bees. These two sets of data—for the detection of pollen and varroa—are both manually curated. We resized the images to 224 × 224 pixels and standardized them to fit into pre-trained ImageNet models during pre-processing. For each dataset, 80% and 20% of the processed dataset were used for training and testing purposes.

2.2. Architecture of BeeNet Model

BeeNet is a deep end-to-end network that receives an image as input, detects whether it is a bee or not, and produces a species prediction and a healthy/unhealthy tag as a result in figure 1. The model's fundamental properties are that it employs entity features, recognizes non-bee input and rejects further predictive computations on it, and predicts both species categorization and health state, as well as the presence of varroa or pollen. The model's primary concept is inspired by Vision Transformer, which uses a transformer encoder to apply self-attention to ”patches” of the input image. The overall architecture of the BeeNet model is shown in figure 2. The BeeNet mainly consists of two fundamental components: feature extraction block and classification block.

In the feature extraction block, our BeeNet uses a modified ResNet50 network, to extract features from input images. Further, positional embedding is implemented to strengthen the data. In the classification block, our model uses the encoder block of the transformer model with a fully-connect layer to generate the classification results.As there is a high dimensional problem in the image data and we want to use the extracted features for further classification, in the feature extraction stage, the input image is used as a reference point by the feature extraction network. We use the modified CNN model, which consists of a structure of the ResNet50d network. This Neural network takes a pre-processed 224x224 image as input and produces the final convolutional layer features as output features. In this model, there are multiple convolution layers, and each layer is followed by a max-pooling layer and an activation function at the end of the CNN to output the entity features.After extracting the features, our BeeNet model flattens these entity features by projecting them onto a linear embedding to generate sequence data. Further, positional embeddings were applied on the sequence, which divides ResNet extracted features into patches.

In the classification block, the model uses the transformer encoder and the fully-connect layer to generate the result. The input to the transformer encoder in BeeNet is the strengthened features of the input images extracted from our ResNet CNN. This technique derives from the domain task's difficulty in standardizing imaging conditions. The conventional Vision Transformer method of image patches would underperform the entire dataset. As input, the strengthened sequences were supplied into the transformer encoder.

The encoder is composed of layers for encoding that proceed through the input layer by layer. There are multiple numbers of encoder blocks that comprise the transformer's encoder. Each encoder layer's function is to generate encoding with information about the inputs' components that are interrelated to one another. It provides its encodings as inputs to the following encoder layer. There are two working blocks in each transformer encoder, which are the

self-attention mechanisms and feed-forward neural networks. The self-attention mechanism accepted input encodings from the prior encoder and evaluates the importance of each encoding to the others to produce output encodings. Each output encoding is then individually processed by the feed-forward neural network. These output encodings are then provided as input to the following encoder, or the encoded sequence of the last encoder is put into the FC layer for classification. Finally, the output is sent into a fully-connected layer head to achieve the prediction.

2.3. Evaluation Matrics

We have utilized classification accuracy as the main indicator to compare the performance of our model to that of others. We also perform a qualitative analysis of the model's interpretability and potential for extracting feature representations to aid in domain classification. Further, all layers in the models were fine-tuned for our bee dataset. The deep networks were trained using a mini-batch stochastic gradient descent optimizer with a batch size of 32. Learning rate, momentum, and weight decay were kept at 0.01, 0.9, and 0.0001 respectively. We also employed a dropout[23] value of 0.2 to prevent overfitting. The confusion matrix was used to create the evaluation matrices. The True Positive, True Negative, False Positive, and False Negative values are combined to generate the confusion matrix. The True Positive value stands for "both predicted values and actual values are positive," while the True Negative value stands for "both predicted values and actual values are negative." Model results at nominal values are provided by accuracy (error between predicted and actual values), precision (predicted value dispersion), and F1-score (harmonic mean of precision and recall). Each model was executed for a total of 100 epochs with an NVIDIA 2070 RTX GPU with 8GB onboard memory. The models were executed by combining a modified version of Wightman's Pytorch-image-models [24] with a direct Python code implementation.

3. Results and Discussion

The BeeNet model implemented in this research was fine-tuned and trained using a stochastic gradient descent optimizer with batch size 32, a learning rate of 0.01, a momentum of 0.9, and a dropout value of 0.2 for 100 epochs. Our datasets were split in the ratio 80:20, where 80% of the data was assigned for training and the rest for testing purposes.

3.1. BeeNet Model

Convolutional Neural Networks are particularly outstanding in image classification tasks due to their superior feature extraction capabilities. While CNN frameworks are quite demanding computationally, their ability to process underlying image features extensively with the help of deep networks has become the most crucial aspect in image recognition tasks. For our task in particular, we have made use of the ResNet50d CNN architecture which uses convolution operation on the input image to build feature maps of detected features and max pooling to filter out the most prominent features from the map.

Deep learning models make use of hyper-parameters to control how models learn and perform; the optimization of these parameters helps minimize validation errors. To train and validate our model, we have used the mini-batch stochastic gradient descent optimizer to reduce the loss function. Based on the outputs of the loss function, the optimizer attempts to optimize the model parameters such as weights and learning rate to reduce the loss. Moreover, our Stochastic Gradient Descent Optimizer uses a batch size of 32. A large batch size is thought to degrade the model's ability to generalize and also the performance. However, a smaller batch size (mini-batch) allows the model allows weights to be updated more frequently and yields better results. The ideal batch size is not always very specific and several attempts of testing and tuning are needed to determine the ideal size. Our optimizer also makes use of momentum (value of 0.9) which is used to boost the optimization process by incorporating history or previous gradient results; updates are performed based on previous results. A higher momentum indicates that the next optimization iteration is strongly influenced by previous iterations. As a result, the time taken to minimize the loss function also significantly reduces. Furthermore, we have used a dropout value of 0.2 and a weight decay of 0.0001 for 100 epochs. The number of epochs indicates the number of times the dataset is entirely processed by a model during the training

Table 1: Best CNN models and Vision Transformer against 3 bee data sets

Models/Data sets	Bee identification	Varroa detection	Pollen detection
ResNest	85.08	84.82	96.48
EfficientNet NoisyStudent	84.63	84.44	96.48
Vision Transformer 16x16	90.09	88.46	98.74
Vision Transformer 8x8	90.78	93.24	98.88
BeeNet	**92.45**	**94.50**	**99.18**

phase. The more the number of epochs, the better a model can optimize its learning. The dropout value is used to reduce overfitting and improve generalization. A dropout value of 0.2 indicates that 20% of the nodes are being dropped in each iteration; the drop in certain nodes allows the model to train using different perspectives in every iteration.

3.2. Bee Identification

We have created a validation set that includes the Diptera data set and a mirroring bee data set with 20,610 bee images. Our BeeNet model outperformed previously implemented CNN and transformer-based models by a significant margin with a classification accuracy of 92.45% for the classification of different species of bees, as stated in table 1. The superiority of transformer-based models over state-of-the-art CNN models is noteworthy as well where the "Vision Transformer 8x8" model was recorded with a classification of 90.78% outperforming other CNN models such as ResNest, EfficientNet, and NoisyStudent. The superiority of our model can be attributed to the integration of a CNN architecture along with a transformed-based architecture which facilitated the use of the best features of both types of models.We used an 85% threshold to determine whether the model could successfully distinguish between a bee and a non-bee based on BeeNet's high predictive performance on our main data set. If the model predicted 85% or higher for a particular sample for any bee classes, it was considered a bee prediction, alternatively non-bee. The confusion matrix for the validation task is shown in figure 3. BeeNet was able to identify between bees and non-bees with an accuracy of 92.45%. We can emphasize that this process is just one of many approaches; an arbitrary threshold value is difficult to define, and computing time and cost for such a complex model are highly challenging to manage.

We compared our models with existing models. In the research, Brian et. al [25] explored images of North-American bumble bees from the different publicly available dataset, and their model got the highest accuracy of 91.7%. The author Barros et al.[26] concentrates on identifying the differences between honeybees and other insect species. They employ a customized CNN-based classifier and two publicly accessible datasets and got a precision of about 94%. Deng et al. [27] proposed a machine-learning model for insect pest detection. Their proposed model demonstrated a recognition accuracy of 85.5%. Karthiga et al. [28] developed a method for classifying the various species of honey bees and detecting their diseases. They used Synthetic Minority Over-sampling Technique (SMOTE) to enhance the data and got the subspecies classification accuracy of 86%. The identification of hive beetles and bees infected by the beetle varroa, which has an impact on the health of bees, though, revealed significant deviations in the model. The comparison of the studies is shown in figure 4.

3.3. Fine-grain Classification

In this research, we have considered the task of fine-grained image classification for a specified domain of entities using a transformer-based classifier. By fine-grained classification, we mean: classifying entities containing a number of relatively small features that are deterministic of their classification, and by a specified domain of entities we define this as a class of entities that may contain sub-classes within the overarching umbrella term, e.g., animal/bug species. For our research we chose to use bee images, noting that this task can be considered a very fine-grained image classification task. We researched for potential public sources of bee images and were able to obtain 3 different data sets of bee images each for a different task in the domain of bees. We examined the concept of pre-training for a domain-specific task using just data from that domain. To this purpose, we presented BeeNet, the vision transformer-inspired end-to-end model that includes bee species classification and health monitoring through the detection of varroa parasites and pollen.

Taking table 1 as the reference, we observe that our BeeNet model was able to perceive and identify bees with Varroa and Pollen much better than other state-of-the-art models with classification accuracy figures of 94.50% and 99.18% when pre-trained on the dataset that constitutes images of bees corresponding to 3 different genera of 11 different species. However, it falls short when our model employs a dataset that is exclusive of bee images for pretraining. Thus, it is reasonable to conclude that our model pre-trained on our employed dataset is able to outperform other CNN and transformer-based models for fine-grain classification tasks. Furthermore, table 1 shows the top-1 accuracy for the varroa and pollen detection using the model pre-trained on the dataset of 11 species from 3 genera in addition to any prior pretraining.

Few studies explored varroa datasets to distinguish between healthy and diseased bees. A semantic segmentation strategy based on AlexNet and ResNet [29], got per-class accuracy of 90%, while Convolutional Neural Network (CNN) architecture [30] achieved the best accuracy of 92.42%. In the study [31], Simon et.al proposed state-of-the-art object identifiers such as YOLOv5, SSD, and deep SVDD for bee identification and Varroa-mite detection in order to discern images of healthy and infected bees. Their training augmented dataset was made up of 24,684 images and their testing set constituted 115 images. Among their proposed models, the YOLOv5 neural network model demonstrated the best performance with the highest precision scores of 0.908 and F1 score of 0.874 when distinguishing between healthy and ill bees.

Babic et al. [32] used a Raspberry Pi at the beehive entrance to demonstrate a video-based imaging technique for identifying pollen and non-pollen bearing honey bees. Using the Nearest Mean Classifier, they were able to classify light-weight features with an accuracy of 88%. The results on the identification of honey bees at the hive entrance and their classification into two types, bearing pollen or not, are encouraging, but there are still many areas for improvement. A self-designed convolutional neural network (CNN) with a classification accuracy of 94 % was presented by Sledevic et al. [33]. Meanwhile, Rodriguez et al. used CNN to automatically classify pollen and non-pollen bearing honey bees [34]. The authors of this particular study curated their own dataset using images extracted from videos of bee activities. The final dataset consisted of a total of 710 images, 354 of which were labeled as pollen-bearing and 346 samples as pollen absent. Comparing baseline models and CNN models, the CNN model VGG19 came out on top with an accuracy figure of 90.2%. However, their study shows that shallow CNN models were far superior in their classification task whereas the 1-layered colorless model achieved the loftiest accuracy figure of 96.4%. While deeper architectures have the potential to improve performance, they did not outperform shallower architectures on this dataset and required longer computations. Stojni'c et al. [35] utilized image descriptor algorithms, namely Scale Invariant Feature Transform (SIFT) and Vector of Locally Aggregated Descriptors (VLAD) to classify pollen and non-pollen containing honey bee classification. The authors observed the highest classification accuracy of 91.5% on the dataset that was prepared using a decorrelated approach to calculate descriptors. The authors recommend the use of deep learning models with a more extensive dataset because this method requires a large amount of data for comprehensive training and their datasets would not be sufficient.

When compared to existing state-of-the-art, our model outperformed in the identification of bee health monitoring with the detection of pollen. The proposed model provides the best accuracy of 99.18 % for pollen detection for bee health monitoring, according to the experimental results. We propose that the low quality of the pollen data and low variation in samples resulted in easy to predict images, which may be a result of the low quality of the pollen data and low variance in samples. We also emphasize that performance can be influenced by hyperparameters such as optimizer, weight decay, batch size, and data augmentation magnitude, in addition to the architecture used. We have attempted to do experimental runs using the same relevant parameters to the greatest extent possible with the computational resources available.

3.4. Real-world Applicability

To determine the real-time feasibility of our proposed model, we compared our results with a few other studies that have used real-time data for their models. While a few existing studies [36,37] utilized embedded cameras inside bee-hive, pre-trained deep algorithms, and got the highest accuracies of 77% and 90% for varroa and pollen detection, our end-to-end BeeNet model best performed among all these models. Further, the framework proposed in this research is computationally extensive due to the incorporation of a transformer-based encoder within the deep network architecture. This allowed the framework to extract very small features of bees facilitating improved classification. However, with the aid of transfer learning, the computational complexity of the framework can be mitigated as learned

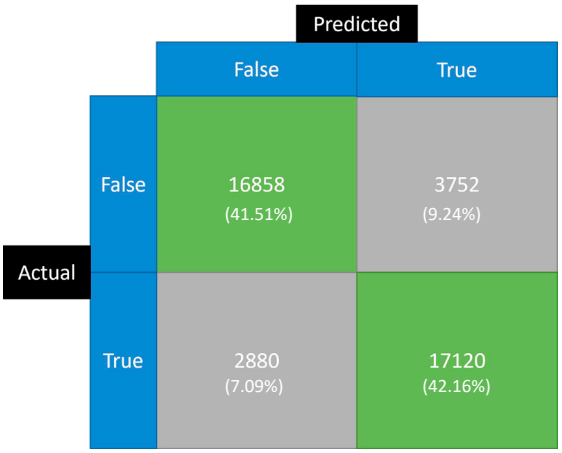

Fig. 3: Confusion matrix for BeeNet verification

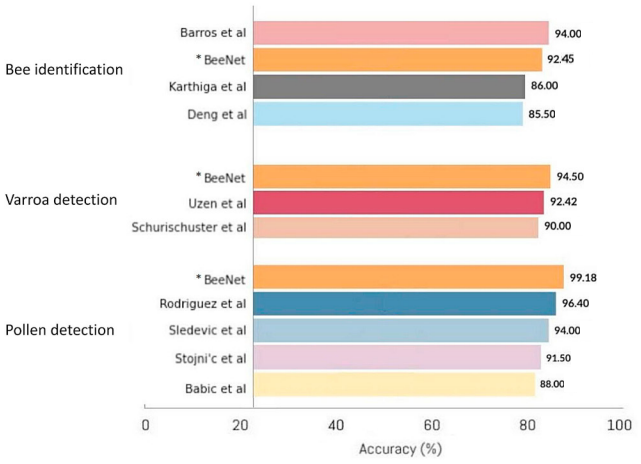

Fig. 4: Comparison of BeeNet with other existing models

features from a particular task can then be employed to perform other tasks without the need to regulate the entire architecture.

3.5. Limitations and Future Works

In the case of BeeNet, we intend to do experiments with more complicated CNN feature extractors, which could result in objectively more comprehensive input to the transformer encoder block. By fine-tuning specific parameters and increasing computing power or the number of sample images, there is still much potential for improvement of accuracy. While we have implemented RandAug, we would like to seek methods of extending smaller data sets for our task, such as in the case of our varroa and pollen data set, that would still aid in improving generalization for BeeNet, with more powerful computational resources we would like to explore a variety of data augmentation methods to further optimize on our bee data-set. We mentioned in this research that the CNN and transformer encoder blocks in BeeNet have the potential to achieve visual interpretability; we intend to obtain this approach and perform experiments on the efficacy of visual features from the CNN and self-attention heatmaps from the transformer encoder. We intend to use an expanded data set to explore a more robust and justifiable approach to verifying BeeNet's ability to distinguish between bees and non-bees. The establishment of a more efficient and reliable approach to bee supervision and preservation will accommodate a more sustainable and favourable ecosystem for different species of bees, which in turn will facilitate population expansion. Consequently, the honey and wax industries will discover themselves relishing a prime period in their business. Few bee classification studies [38,39] explored transfer learning with deep learning models and got good accuracies. In future work, we would like to explore state-of-the-art transfer learning, which will eventually reduce computational time and requirements for a large training dataset.

4. Conclusion

In this research, we have explored the task of fine-grained image classification and both investigated and developed existing and new methods to interpret and tackle a bee species classification task. Fine-grained image classification is a critical task with relations to many different fields, investigating methods to improve upon applications of models and uncovering greater uses can be especially influential. Bees are an incredibly important insect family that upholds the greater ecosystem, industries such as food and agriculture are heavily reliant on the balance of our ecosystem. It is therefore of utmost interest to preserve and observe bee populations around the world, species classification would aid in population management and observation efforts. We have researched visual interpretability methods that can pave the way for improved classification efforts. We have also developed an end-to-end model BeeNet that can not only classify bee species but conduct additional fine-grained classification tasks such as parasite and pollen detection.

Our results are promising regarding BeeNet model performance and broadly in the methods introduced with CAMs, however, there are many aspects and challenges to consider with the availability of appropriate data and accuracy of interpretability approaches.Comparison with our research suggests that our proposed framework has the potential to produce more efficient and reliable results on real-time bee image data due to the outstanding classification values achieved.

Acknowledgements

The authors acknowledge the Optus-Curtin Centre of Excellence in Artificial Intelligence, Curtin University, Bently, WA - for their support in this research.

References

[1] Klein, A.M., Boreux, V., Fornoff, F., Mupepele, A.C. and Pufal, G., (2018). Relevance of wild and managed bees for human well-being. Current Opinion in Insect Science, 26, pp.82-88.
[2] Gill, R.J., Baldock, K.C., Brown, M.J., Cresswell, J.E., Dicks, L.V., Fountain, M.T., Garratt, M.P., Gough, L.A., Heard, M.S., Holland, J.M. and Ollerton, J., (2016). Protecting an ecosystem service: approaches to understanding and mitigating threats to wild insect pollinators. In Advances in ecological research (Vol. 54, pp. 135-206). Academic Press.
[3] Klein, A.M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C. and Tscharntke, T., (2007). Importance of pollinators in changing landscapes for world crops. Proceedings of the royal society B: biological sciences, 274(1608), pp.303-313.
[4] Goulson, D., Nicholls, E., Botías, C. and Rotheray, E.L., (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science, 347(6229), p.1255957.
[5] Schroder, S., Wittmann, D., Drescher, W., Roth, V., Steinhage, V. and Cremers, A.B., (2002). The new key to bees: automated identification by image analysis of wings. Pollinating bees–the Conservation Link Between Agriculture and Nature, Ministry of Environment, Brasilia, pp.209-218.
[6] Gibbs, J., (2018). DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. Genome, 61(1), pp.21-31.
[7] Arbuckle, T., Schröder, S., Steinhage, V. and Wittmann, D., (2001, October). Biodiversity informatics in action: identification and monitoring of bee species using ABIS. In Proceedings of the 15th International Symposium Informatics for Environmental Protection (Vol. 1, pp. 425-430).
[8] De Moor, B., De Gersem, P., De Schutter, B. and Favoreel, W., (1997). DAISY: A database for identification of systems. JOURNAL A, 38(4), p.5.
[9] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). IEEE.
[10] Ciregan, D., Meier, U. and Schmidhuber, J., (2012, June). Multi-column deep neural networks for image classification. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3642-3649). IEEE.
[11] Taigman, Y., Yang, M., Ranzato, M.A. and Wolf, L., (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).
[12] Sun, Y., Chen, Y., Wang, X. and Tang, X., (2014). Deep learning face representation by joint identification-verification. Advances in neural information processing systems, 27.
[13] Wan, L., Zeiler, M., Zhang, S., Le Cun, Y. and Fergus, R., (2013, May). Regularization of neural networks using dropconnect. In International conference on machine learning (pp. 1058-1066). PMLR.
[14] Lin, Z., Jia, J., Gao, W. and Huang, F., (2020). Fine-grained visual categorization of butterfly specimens at sub-species level via a convolutional neural network with skip-connections. Neurocomputing, 384, pp.295-313.
[15] Chauhan, R., Ghanshala, K.K. and Joshi, R.C., (2018, December). Convolutional neural network (CNN) for image detection and recognition. In 2018 first international conference on secure cyber computing and communication (ICSCCC) (pp. 278-282). IEEE.
[16] Srinivas, S., Sarvadevabhatla, R.K., Mopuri, K.R., Prabhu, N., Kruthiventi, S.S. and Babu, R.V., (2016). A taxonomy of deep convolutional neural nets for computer vision. Frontiers in Robotics and AI, 2, p.36.
[17] Buschbacher, K., Ahrens, D., Espeland, M. and Steinhage, V., (2020). Image-based species identification of wild bees using convolutional neural networks. Ecological Informatics, 55, p.101017.
[18] He, X. and Peng, Y., (2017). Fine-grained image classification via combining vision and language. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5994-6002).
[19] Peng, Y., He, X. and Zhao, J., (2017). Object-part attention model for fine-grained image classification. IEEE Transactions on Image Processing, 27(3), pp.1487-1500.
[20] Yoo, J., Hossain, M.Z. and Ahmed, K.A., 2021. A Machine Learning Based Approach to Study Morphological Features of Bees. Proceedings 2021, 1, 0.
[21] Unger, S., Rollins, M., Tietz, A. and Dumais, H., (2021). iNaturalist as an engaging tool for identifying organisms in outdoor activities. Journal of Biological Education, 55(5), pp.537-547.

[22] "The BeeImage Dataset: Annotated Honey Bee Images," www.kaggle.com. https://www.kaggle.com/jenny18/honey-bee-annotated-images (accessed Nov. 30, 2022).
[23] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), pp.1929-1958.
[24] Wightman, R. (2019) "PyTorch Image Models," GitHub. https://github.com/rwightman/pytorch-image-models.
[25] Spiesman, B.J., Gratton, C., Hatfield, R.G., Hsu, W.H., Jepsen, S., McCornack, B., Patel, K. and Wang, G., (2021). Assessing the potential for deep learning and computer vision to identify bumble bee species from images. Scientific reports, 11(1), pp.1-10.
[26] Barros, C.M., de Freitas, E.D.G., Braga, A.R., Bomfim, I.G.A. and Gomes, D.G., (2021, July). Applying convolutional neural networks in images for automated recognition of honey bees (Apis mellifera l.). In Proceedings of the XII Workshop on Computing Applied to the Management of the Environment and Natural Resources (pp. 19-28). SBC.
[27] Deng, L., Wang, Y., Han, Z. and Yu, R., (2018). Research on insect pest image detection and recognition based on bio-inspired methods. Biosystems Engineering, 169, pp.139-148.
[28] Karthiga, M., Sountharrajan, S., Nandhini, S.S., Suganya, E. and Sankarananth, S., (2021, March). A Deep Learning Approach to classify the Honeybee Species and health Identification. In 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII) (pp. 1-7). IEEE.
[29] Schurischuster, S. and Kampel, M., (2020, November). Image-based classification of honeybees. In 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6). IEEE.
[30] Üzen, H., Yeroğlu, C. and Hanbay, D., (2019, September). Development of CNN architecture for Honey Bees disease condition. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-5). IEEE.
[31] Bilik, S., Kratochvila, L., Ligocki, A., Bostik, O., Zemcik, T., Hybl, M., Horak, K. and Zalud, L., (2021). Visual diagnosis of the varroa destructor parasitic mite in honeybees using object detector techniques. Sensors, 21(8), p.2764.
[32] Babic, Z., Pilipovic, R., Risojevic, V. and Mirjanic, G., (2016). Pollen bearing honey bee detection in hive entrance video recorded by remote embedded system for pollination monitoring. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3, p.51.
[33] Sledevič, T., (2018, November). 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). IEEE.
[34] Rodriguez, I.F., Megret, R., Acuna, E., Agosto-Rivera, J.L. and Giray, T., (2018, March). 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.
[35] Stojnić, V., Risojević, V. and Pilipović, R., (2018, March). Detection of pollen bearing honey bees in hive entrance images. In 2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH) (pp. 1-4). IEEE.
[36] Voudiotis, G., Moraiti, A. and Kontogiannis, S., (2022). Deep Learning Beehive Monitoring System for Early Detection of the Varroa Mite. Signals, 3(3), pp.506-523.
[37] Marstaller, J., Tausch, F. and Stock, S., (2019, October). DeepBees-Building and Scaling Convolutional Neuronal Nets For Fast and Large-Scale Visual Monitoring of Bee Hives. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) (pp. 271-278). IEEE.
[38] Berkaya, S.K., Gunal, E.S. and Gunal, S., (2021). Deep learning-based classification models for beehive monitoring. Ecological Informatics, 64, p.101353.
[39] Chawane, S., 2022. Image based bee health classification (Master's thesis, University of Twente).