Image Segmentation of Meliponine Bee using Mask- RCNN

Afif Nizam, Wan-Noorshahida Mohd-Isa, Aziah Ali

Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

ABSTRACT

Bee keeping has gained quite an attention by local Malaysian especially towards a particular bee species known as Meliponine, which is a tribe of stingless bee. Up until now, the identification of this species still heavily relies on the input of a handful key experts. Many studies have shown that the honey contains beneficial nutrients that can combat against various diseases, which makes it very popular and valuable as compared to the normal honey. This study aims to aid the process of bee keeping by developing a visual-based guidance system.The main process of this system will involve image segmentation. Due to the miniature size of the Meliponine, the conventional way of image segmentation using background subtraction is becoming irrelevant and difficult. In recent years, the convolutional neural network (CNN) has been gaining huge ground in object detection due to its high accuracy and efficiency. In this study, we fine-tuned the Faster R-CNN architecture to perform segmentation on the Meliponine image from its background. Our dataset consists of 400 image frames from videos collected in local Meliponine farm in Malaysia. On these datasets we achieved 74% accuracy, which seems promising for further study. We also incorporated the Mask R-CNN architecture to carry out instance segmentation as opposed to the bounding box level detection provided by Faster R- CNN. Using Mask R-CNN, the computing time for training significantly improved.

Keywords: *Expert system, Image Segmentation, Meliponine, object detection.*

I. INTRODUCTION

The insect of interest for this particular study is known as the Meliponine or "Kelulut" in the Malay language is referred to as a tribe of stingless bee (Fig.1). Recently, the current population of stingless bees in Malaysia has risen up to 68 species [1]. It is thought there are an estimated of around 500 species around the world. As mentioned earlier, the honey of Meliponine is found to be better compared to other types of bees. This is due to the low sugar content as well as having high minerals, moisture and free phenolic acid that can act as a remedy for cancer and dengue [2]. Malaysia is currently taking the initiative to promote the Meliponine honey as a "superfood" [3-4]. This leads to an increasing popularity by the locals towards "Kelulut" farming in Malaysia. As indicated by [5] various types of Meliponine lean towards particular florals and for- aging natural habitat for its ideal wellbeing that in turn produces good quality for human's consumption.

In any case, to identify the types of a Meliponine, local (for the most part amateur) beekeepers need to depend on contributions from just a pool of key specialists. Therefore, the objective of our project is to develop a visual-based guidance system that may aid the process of identification of the Meliponine species. Key findings gathered from such system would enable further studies and contribution to the field of insect identification.

In this paper, we performed image segmentation of a Meliponine image from its background. This study focuses mainly on the process towards performing identification of the species on the Meliponine. We are utilizing the image segmentation algorithm provided by the existing deep learning architecture. In this way, the machines will be able to learn the region of interest (ROI) and segmentize the image of the Meliponine.

Fig. 1. Comparison between Meliponine (left) and European Honey Bee (right). They are not of exact scale

As indicated by [5] various types of Meliponine lean towards particular florals and for- aging natural habitat for its ideal wellbeing that in turn produces good quality for human's consumption .In any case, to identify the types of a Meliponine, local (for the most part amateur) beekeepers need to depend on contributions from just a pool of key specialists. Therefore, the objective of our project is to develop a visual-based guidance system that may aid the process of identification of the Meliponine species. Key findings gathered from such system would enable further studies and contribution to the field of insect identification.

In this paper, we performed image segmentation of a Meliponine image from its background. This study focuses mainly on the process towards performing identification of the species on the Meliponine.

We are utilizing the image segmentation algorithm provided by the existing deep learning architecture. In this way, the machines will be able to learn the region of interest (ROI) and segmentize the image of the Meliponine.

II. LITERATURE REVIEW

There are many existing implementations of computerized identification on butterflies [9], wasps [10], spiders [11], and fruit flies [12]. However, the identification on the Meliponine relies very much on the manual visual inspection. As a reference for this study, we utilized the structured algorithm [6] that is based on the anatomical characteristics of the Meliponine. The anatomical structure can be shown in Fig 2. The image acquisition process, would usually involve capturing a live Meliponine and curing it before an image can be captured. This is due to unpredictable and irregular movement of the Meliponine. Ultimately, this poses a new research challenge especially in the area of image segmentation as well as in object identification.

Based on previous studies, the amount of work done on Meliponine is very minimal, let alone into the work of the vision-based system. One study though has explored into the suitability of florals and foraging for Meliponine as reported by [5]. However, there are research being done on the identification of other insects on a generic level by [7-8]. On top of that, we have seen the implementation of insect recognition proved to be beneficial for pest management [13-14]. Alternatively, the work on image segmentation using deep learning has been applied to various areas such as on the computer vision domain. As mentioned by [18], semantic segmentation is the process of pixel to pixel detection which separates the region of interest (ROI) from its background. The author also stressed the importance of semantic segmentation on assisting common applications such as detecting road signs [19] and detecting chronic diseases such as tumors [20]. The recent implementation of image segmentation using Residual Networks (ResNets) together with dilated convolutions has achieved quite a convincing result in terms of its performance [21]. Moreover, the process of semantic segmentation can be divided into unsupervised and supervised approaches. Unsupervised learning methods is considered to be unsuitable for segmentation process, due to its inability to consistently identify region boundaries [22]. Therefore, [23] presented a new enhancement, by introducing a new learning method known as "Semi-supervised" into the segmentation process. The combination of unlabeled data, weakly labelled data and non-real images produced through Generative Adversarial Networks (GAN), provided a resounding accuracy rate of 80%.

Due to the success of applying supervised machine learning approach for object identification. We had developed a slightly similar approach which involves supervised ma- chine learning method via image template and published a paper here [15]. We had to work on a curated sample of Meliponine due to its small size. However, we soon realized that the amount images taken weren't sufficient, that possibly might have caused a significant drop in the accuracy rate.

Therefore, for this particular project, we aim to collect the images and identify the bee species in its natural surroundings. In order for the process of identification to work, we must perform image segmentation to extract the image of Meliponine from its background. We hope this process will aid the effort of preserving the numbers of Meliponine in the wild since we are not required to capture them alive and curing them.

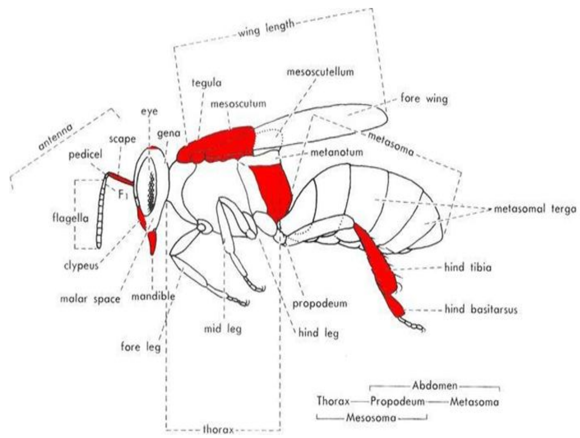

Fig. 2. Anatomical structure of a Meliponine [1].

As shown in Fig 2, the Meliponine consists of distinctive morphological structures that can be used as key features for identification process. Therefore, we decided to utilize the deep learning framework as it thought to be useful for general purpose feature extraction [16]. The framework learns the key features of the image through iterations of training. CNN (Convolutional Neural Network) was also used for its object detection algorithm. The CNN consists of multiple layers of computational process with variety of functionality as well as predefined parameters. In this study, we are using the deep learning framework of Faster R-CNN

(Region-CNN) for the purpose of segmentation of the Meliponine image from its background. The Faster R-CNN provides the functionality of extracting the images using bounding box detection. However, our main goal is to segmentize the region of interest to the level of pixel to pixel detection rather than bounding box. Therefore, we decided to utilize another deep learning architecture known as "Mask R-CNN". Mask R-CNN is an extension of the Faster R-CNN model with an additional branch for the purpose of prediction segmentation mask on the desired object as explained by [17].

III. METHODOLOGY

A. Data sets

We had collected data of Meliponine at a privately-owned farm in the state of Selangor of the Peninsular Malaysia. It housed five species of the Meliponine. Due to the erratic movement of the stingless bees, videos were captured rather than images. We used a Sony HDR-TD30V Full HD 3D video camera. The footages were captured at 60 fps with a resolution of 1920x1080.

The average duration of the footages were 30 seconds. Due to the small size of these Meliponine, we had to make use of the macro setting, which was provided by the camera manufacturer.

These videos undergo a slicing process, where it was sliced into individual frames. There were around 3600 images frame extracted with each species consisting of 900 images. Every individual image frame was exposed to a resizing procedure. This helps the process of training, since the larger the image frames, the longer time would be required to train them. Before we were to feed the images into our deep learning implementation, we had to label the image individually using a labelling tool known as "VGG Image Annotator".

The processed image frames after resizing were divided into training and testing image frames. The image frames were separated into 80:20 train to test percentage ratio. The total amount of image frames were 400 image frames consisting of 320 training image frames and 80 testing image frames. Each image frames were labeled manually before becoming inputs to our proposed deep learning framework.

B. Deep Learning for Image Segmentation

We implemented the deep learning framework of Faster R-CNN via Tensor Flow model and utilizes the Mask RCNN algorithm for the instance segmentation process. Faster R-CNN has two key components; the Region Proposal Network (RPN) and Fast region- based CNN (Fast R-CNN). In a Faster R-CNN, both RPN and Fast R-CNN are combined to create a single CNN, which are faster than the Fast R-CNN. The resulting Faster R-CNN implements an end-to-end process via looking into the high and low resolutions pixels of the image frames.

Faster R-CNN "learns" the region by creating a convolutional feature map. It starts with a predicted region proposal, which is then modified and enhanced to create another region proposal. The image frames, as inputs, go through multiple layers of network with predicted region proposals. The enhancement of the region proposals is done by way of a pooling layer based on a Region of Interest (ROI). The layer creates a bounding box surrounding the proposed region and predicts its offset values.

Mask R-CNN or most commonly known as (regional convolutional neural network). It is the continuation from the Faster R-CNN. As mentioned earlier, Faster R-CNN is well known for its object detection capability. Mask R-CNN extends the existing model with the capability of carrying out instance segmentation process. Fundamentally, Mask R-CNN consists of few main modules namely; the backbone, Region Proposal Network (RPN), ROI classifier and Bounding Box Regressor.

We ran our training using a GPU of NVIDIA GTX 950m on an Intel Core i7-4720 HG CPU @ 2.6 GHz. For training the TensorFlow model, each training image frames would need to have a generated TFRecords. Each TFRecords is a .xml file consisting of labelled data, which would be the input data for the TensorFlow model.

By the 80:20 ratio of train-to-test data, we used 320 image frames for training our model. The initial value of the learning rate was set at 0.05 as we aim to get a 95% accuracy rate. Training went on till the average loss was almost constant. For evaluation, we looked at several loss functions namely: the RPN objectness localization and classification loss, the RPN localization loss, the total loss and the clone loss. Then, using the trained model, we performed a test on 80 image frames of the Meliponine from the genus of Heterotrigona. Detection scores were assigned to represent the confidence level for each predicted region.

The processes involved in our implementation for training of image frames using the Faster R-CNN and Mask R-CNN are illustrated as in Fig.3.

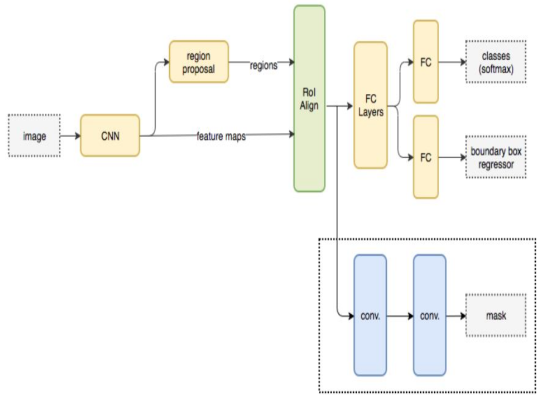

Fig. 3. The processes involved in training using Faster R-CNN and Mask R-CNN in this paper [24].

IV. RESULTS AND DISCUSSION

Table 1 gives out the results of training. From Table 1, we could see that at the smoothing values of less than 0.05, the average computing time per image is 0.65s. This average time was considered quite fast considering the GPU specification that we used was considered to be a below average specification than usual implementation by others. In terms of the total time taken for training, it took us nearly 16 hours for 87,000 iterations before the loss function consistently went below 0.05 rate. The time is considered a slow training time compared to usual standard.

Out of the 80 image frames tested, at 99% confidence level, 59 image frames were correctly detected for segmentation of Meliponine from its background, refer Fig.4. Thus, producing only 73.75% rate of segmentation accuracy. There were 21 image frames that were incorrectly detected. Based on our inspection, this may be because of the difficulty for the model to detect the Meliponine from its background due to the color contrast. As can be seen from Fig.4, there existed similarities in color between the background and the Meliponine. Meliponine, being small and dark in nature were unable to be detected and had incorrectly been detected by the dark patches that existed in its surrounding.

Fig. 4. The sample of output for the image segmentation process

Table 2 shows the result of training using Mask R-CNN, we could clearly see that the model performed quite well in terms reaching below the value of 0.05 at around 4 hours. The model also showed quite a promising result with regards to how good the model is at locating objects within the image; presented as RPN bounding box loss graph in Table 2. However, the validation loss seems to be quite high considering that substantial amount of training time has been made. The lowest validation score extracted after 1 day of training was 2.823. We suspected that this may be due to the miniature scale of data set and perhaps the model is experiencing the issue of over fitting.

We could see that for future implementation, we may need to adjust the color contrast of the inputs. However, the accuracy value of segmentation (74%) seems promising as this is our first attempt in this problem domain.

Table 1. Results of Training using Faster R-CNN

Loss Functions	***Steps***	***Smooth Values***	***Processing Time (h:m:sec)***
RPN Objectness Localisation Loss	12,110	4.745×10^{-2}	4:53:17
RPN Objectness Classification Loss	25,320	4.744×10^{-2}	6:47:17
RPN Localisation Loss	44,820	4.896×10^{-2}	9:35:17
Total Loss	85,890	4.749×10^{-2}	15:29:17
Clone Loss	86,590	4.576×10^{-2}	15:35:17
Average Computing Time per Image (sec)			**0.65**

Table 2. Results of Training using Mask R-CNN

Loss Functions	***Steps***	***Smooth Values***	***Processing Time (d:h:m:sec)***
Loss for Mask R-CNN bounding box refinement	2	3.89×10^{-2}	1:11:47
RPN bounding box loss graph	20	1.11×10^{-2}	11:56:02
Total Loss	7	4.872×10^{-2}	4:09:44
Validation Loss	3	2.823	1:0:4:34

V. CONCLUSION AND FUTURE WORK

The result achieved have shown to be promising, in which the accuracy rate is al- most 74% at a smoothing value of less than 0.05. The average computing that the GPU and CPU required to train one image (0.65) was surprisingly fast. With the implementation of Mask R-CNN, the total training time significantly improved slightly above the 4-hour mark. Future possible work would be to implement a color contrast correction to the input image frames before applying image segmentation via Faster R-CNN. As mentioned previously, the Mask R-CNN that we implemented might be suffering from the issue of over fitting. Therefore, we are exploring the possibility of utilizing various data augmentation method to improve the validation result. We are also looking into the possibility of utilizing better GPU and CPU specification for specifically for training procedure.

ACKNOWLEDGEMENT

We acknowledge the receipt of funding from Malaysia Ministry of Education under the FRGS grant scheme with a grant no: MMUE/160024.

REFERENCES

[1] Jalil, Abu Hassan, and Ibrahim Shuaib, "*Beescape for Meliponines: Conservation of Indo-Malayan Stingless Bees. Singapore*": Patridge" Publishing, 2014.

[2] Roowi, Suri, Siti Aisyah Muhammad, Hamdan Sipon, Mohd fahimee Jaafar, Nazrul Hashim Daud, and Rosnah Othman. "*Free phenolic acids in kelulut honey*." Bulletine Technology MARDI, 2, pp. 145- 147, 2012.

[3] "*Madu kelulut superfood pertama di Malaysia*." Utusan Malaysia,published 7 Sept 2016. http://www.utusan.com.my/bisnes/usahawan/madu-kelulut-i-superfood-i-pertama-di-malaysia- 1.379430

[4] "*Sweet profit in stingless bee farms*." The Star Malaysia, published 3 Sept 2016.

[5] https://www.thestar.com.my/news/nation/2016/09/03/sweet-profit-in-stingless-bee-farms-mardi-eager-to-push-kelulut-honey-as-new-entry-into-superfood-ma/

[6] Muradian, Ligia Biicuda Almeida. "*Comparison between Apis mellifera bee and stingless bee (Tetragoonisca angustula) honey*." Presentation slides, University of Sao Paulo, Paulo, Brazil. Retreived 3 Feb 2015.

[7] Abu-Hasan, A.J.. "*Pictorial key to Indo-Malayan stingless bee genera*." In: Ibrahim, S. Eds, 17th Congress of the International Union for the Study of Social Insects (IUSSI), 2014.

[8] Wang, Jiangning, Congtian Lin, Ligian Ji, and Alping Linag. "*A new automatic identification system of insect images at the order level*." Knowlede-Based Systems, 33, pp. 102-110, 2012.

[9] Yaakob, Shahrul Nizam and Lakhmi Jain. "*An insect classification analysis based on shape features using quality threshold ARTMAP and moment invariant.*" Applied Intelligence, 37, pp. 12-30, 2012.

[10] Wang, Jiangning, Liqiang Li, Aiping Liang, and Decheng Yuan. ***"The identification of butterfly families using content-based image retrieval***." Biosystems Engineering, III, pp. 24-32, 2012.

[11] Weeks, P.J.D., M.A. O'Neill, K.J. Gaston, and I.D. Gauld. "*Species-identification of wasps using principl component asssociative memories*." Image and Vision Computing, 17, pp. 861-866, 1999.

[12] Ticay-Rivas, Jaime R., Marcos del Pozo-Banos, William G. Eberhard, and Jeusu B. Alonso. "*Spider specie identification and verification based on pattern recognition of its cobweb*." Expert System with Applications, 40, pp. 4213-4225, 2013.

[13] Faria, F.A., P. Perre, R.A. Zuchi, L.R. Jorge, T.M. Lewinsohn, A. Rocha, and R. da S. Torres. "*Automatic identification of fruit flies (Diptera: Tephritidae)*." Journal of Visual Communication and Image Representation, 25, pp. 1516-1527, 2014.

[14] Wen, Chenglu and Daniel Guyer. "*Image-based orchard insect identification and classification method.*" Computers and Electronics in Agriculture, 89, pp.110-115, 2012.

[15] Wen, Chenglu, Daniel Guyer, and Wei Li. "*Local feature-based identification and classification for orchard insects*." Biosystems Engineering, 104, pp. 299-307, 2009.

[16] Nizam, Afif, Wan-Noorshahida Mohd-Isa, Aziah Ali. "*Identification of the genus of the stingless bee via image template*." Journal of Fundamental and Applied Sciences, 10(6S), pp. 812-821, 2018.

[17] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. ***"Deep learning***." nature 521, no. 7553 (2015): 436.

[18] Kaiming He, Georgia Gkioxari ,Piotr Dollar, Ross Girshick . "*Mask R-CNN*". Facebook AI Research (FAIR)

[19] Xiaolong Liu1,Zhidong Deng,Yuhan Yang, "*Recent progress in semantic image segmentation*" 2018

[20] Maldonado-Bascon S, Lafuente-Arroyo S, Gil-Jimenez P, Gomez-Moreno H, López-Ferreras F (2007) *Roadsign detection and recognition based on support vector machines. IEEE Trans Intell Transp Syst 8(2):264–278*

[21] Moon N, Bullitt E, Van Leemput K, Gerig G (2002) "*Automatic brain and tumor segmentation*". Med Image Comput Comput Assist Interv MICCAI 2002:372–379

[22] K. He, X. Zhang, S. Ren, and J. Sun, "*Deep residual learning forimage recognition*," Retrieved August 3, 2019 arXiv preprint arXiv: 1512.03385, 2015.

[23] Harri Valpola, "*From Neural PCA to Deep Unsupervised Learning*" Retrieved August 3, 2019 https://arxiv.org/pdf/1411.7783, 2015

[24] Nasim Souly, Concetto Spampinato ,Mubarak Shah (2017) , "*Semi Supervised Semantic Segmentation Using Generative Adversarial Network*"

[25] Jonathan Hui (2018, April 20). Image segmentation with Mask R-CNN [Web log post]. Retrieved August 3, 2019, from https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-ebe6d793272