Project Proposal

Using Machine Learning to track the foraging quality of Bumblebees

MSc Computational Methods in Ecology and Evolution

Paul Cabrisy

CID: 02138753

plc21@ic.ac.uk

Supervisor: Peter Graystock

p.graystock@imperial.ac.uk

Department of Life Science

Imperial College London

2022

1. Six keywords to describe the project

Bumblebees, DeepLabCut, foraging, machine-learning, pollen, videos.

2. Introduction to the project idea and proposed questions

Over the past few years, pollinators including bumblebees (*Bombus spp*) have shown sharp declines in population size across Europe (Biesmeijer et al. 2006 & Potts et al. 2011). Bumblebees are important pollinators that forage on a wide variety of plants (Ollerton, Winfree &Tarrant, 2011). Due to their role as pollinators, they provide crucial ecosystem services and maintain plant diversity (Ollerton, Winfree &Tarrant, 2011). About 70% of leading food crops around the world are dependent on pollinators (Klein et al., 2007). Therefore, ensuring the conservation of pollinators like bumblebees is an important factor towards stabilizing global food security.

It is possible to monitor healthy bumblebee populations by looking at their foraging activity (Judd, Huntzinger, Ramirez & Strange, 2020). During the active seasons, large populations have to forage on high quality pollen to survive (Carvel et al. 2006). To investigate the foraging activity of bumblebees, methods such as placing pollen traps can be used and are able to collect large quantity of pollen with minimal labour (Judd et al. 2020). This project investigates the usage of machine learning techniques for collecting data from video footage of foraging bumblebees to investigate the foraging quality of bumblebee colonies.

The first goal of this project is to perform marker-less pose estimation on bumblebees to validate its reliability in detecting individuals within groups using DeepLabCut (DLC). The objective will then be to collect video footage from different bumblebee nests on the Silwood Park campus and using different standardisation methods for the recordings. I will determine the most reliable method based on the DLC performance evaluated in pixel error. Once the algorithm has been trained on one bumblebee nest and with one method it can then be applied to another nest with a different method to compare within-species performance metrics. The second goal is to train the pose estimation software to detect and quantify the size of pollen baskets (corbiculas) on bumblebees as they return from foraging. Pollen traps will be attached to the nests to find the relationship between the pixel size and the collected pollen from the traps. The aim of this project is not only to count how many bumblebees go foraging but quantify the amount of pollen collected. Therefore, I will be modelling the relationship between the data collected from both the videos and the pollen traps to determine and compare foraging efficiency between colonies.

The three hypotheses of this project are as follow:

1) DLC can be trained to identify and track individual bumblebees within groups.

2) The performance of DLC is different between recordings with different camera angles.
3) DLC can be used to measure and quantify the pollen collected by foraging bumblebees.

3. Proposed Methods

I will purchase 4 established bumblebee colonies to set up on campus and record their foraging activity every day from two camera angles over one week. I will then train DLC to identify the bumblebees in the videos and compare recording methods (camera angles) to see which performs better at training DLC to identify individuals within groups. I will also train DLC to measure the size in pixels of pollen collected in the footage and use pollen traps to quantify the amount of the pollen collected.

For the first hypothesis of identifying and tracking bumblebees within groups I will compare two approaches, the bottom-up and top-down methods. The former looks at the full scene and takes the body parts of interest and uses part affinity fields to assign these body parts together to form individuals (Lauer et al. 2021). The later defines bounding boxes and performs the task of pose estimation within that box to predict the pose of the individual inside (Lauer et al. 2021). A top-down method is expected to work best when bumblebees are not in crowded scenes with multiple workers in the same bounding boxes.

To test the second hypothesis of looking at difference in training performance between recording method, I will place GoPro cameras facing the nest entrances from two angles. One camera will directly face the entrance of the nest and another will be placed on the side to record bumblebees going through transparent tubes to get to their nest. I will record the foraging activity of each nest for 1 hour every day over one week. This will give enough footage to compare the DLC performance by training the algorithm on different setups and looking at the outcome to see which method better captures the anticipated results.

For the last hypotheses of using DLC to quantify to amount of pollen collected in the footage I will attach pollen traps to the nest entrances. Pollen traps are 3D printed on site and can be adjusted to different entrance sizes. The traps are designed so that bumblebees must squeeze through a tight tube to enter their nest which knocks off the collected pollen into a container. Pollen traps will be attached in a climate control room before setting up the nests outdoors. After running the algorithm on the Imperial HPC on all remaining videos, I will extract pixel coordinates of the corbiculas for each video in Rstudio and find a calibration for pixel size that best represents the weighed pollen.

4. Anticipated outputs and outcomes

With DLC having been used on a wide variety of taxa including invertebrates (Mattis et al. 2018) I anticipate that it would be able to be trained to recognize bumblebees within groups. I also expect the algorithm to be able to detect the corbicula as an additional body part. After labelling enough frames to validate its performance the algorithm should be able to take new footage of different foraging bumblebees and track the foraging quality of this colony without having to do the training process again. The output for tracking bumblebees will be a pixel error calculated from the frames used in the training and kept for validation to evaluate the performance of DLC. The output for the training to measure corbicula size will be a data-frame in Rstudio with coordinates of specified body parts that were labelled in the training. These coordinates will give the corbicula size in pixel length of each bumblebee entering the nest. I will also have weighed samples of corbiculas collected from the pollen traps that I will use to quantify the amount of pollen collected. I will receive training to manipulate bumblebees nests from my supervisor who will also be guiding me throughout this project.

5. Project feasibility supported by a timeline of tasks

This hybrid lab and desk-based project will be carried out over 5 months and involves quantitative and computational skills that are within the bounds of my course.

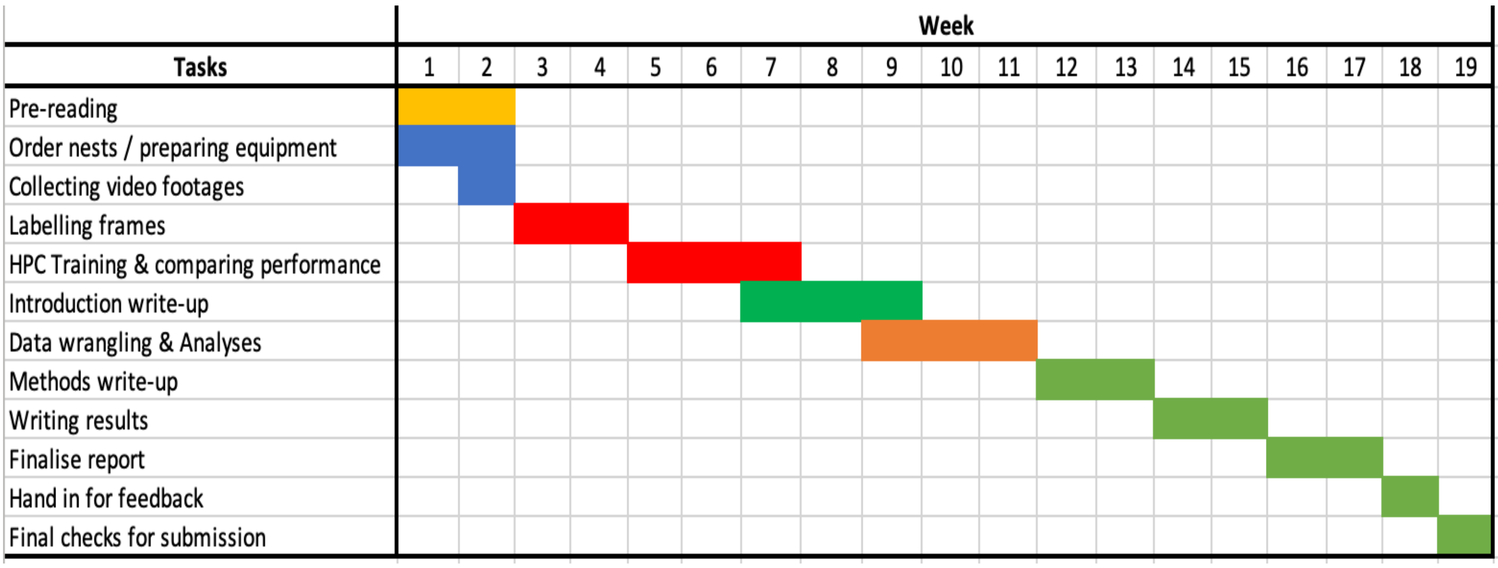

Figure1. Gantt chart of the project timeline in weeks from the11th-April to the 29th-August (Yellow: pre-reading, Blue: data collection, Red: DLC training, Orange: data analysis, Green: report writing).

6. An itemized budget

4x bumblebee nests @ £84.99ea: £424.95 – Four bumblebee nests will allow to collect enough footage to compare different recording methods and has the potential to ask further ecological questions after the main hypotheses have been tested.

10kg Pollen @ £20: to feed bees

PLA white plastic @ £20: This is to 3D print the pollen traps to be attached to the hives.

Clear tubes@ £5:- Transparent tubes that will be attached to the hives to record bumblebees entering.

Total: £469.95

7. References

Biesmeijer, J.C. (2006). Parallel Declines in Pollinators and Insect-Pollinated Plants in Britain and the Netherlands. Science, 313 (5785), pp. 351–354.

Carvell, C., Westrich, P., Meek, W.R., Pywell, R.F. & Nowakowski, M. (2006). Assessing the value of annual and perennial forage mixtures for bumblebees by direct observation and pollen analysis. *Apidologie*. 37 (3), pp. 336-340.

Judd, H., Huntzinger, C., Ramirez, R. and Strange, J., 2020. A 3D Printed Pollen Trap for Bumble Bee (*Bombus*) Hive Entrances. Journal of Visualized Experiments, (161).

Klein, A. M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T, (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings.Biological Sciences / the Royal Society.* 274 (1608), pp. 303-313.

Lauer, J., Zhou, M., Ye, S., Menegas, W., Nath, T., Rahman, M.M., Di Santo, V., Soberanes, D., Feng, G., Murthy, V.N., Lauder, G., Dulac, C., Mathis, M.W. and Mathis, A. (2021). Multi-animal pose estimation and tracking with DeepLabCut.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W. and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, [online] 21(9), pp.1281–1289.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W.E. (2010). Global pollinator declines: trends, impacts and drivers. Trends in Ecology & Evolution, 25 (6), pp. 345–353.

Ollerton, J., Winfree, R. & Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos*. 120 (3), pp. 321-326.

I have seen and approved the proposal and the budget

Dr Peter Graystock

11/04/2022