CS460 Group3 Project Proposal BeeML: BAN MachLA (Bee Annotation Machine Learning Algorithm)

CL Srinivas, M Sreerag and Aritra Mukhopadhyay(TA)

Project Idea

To curate a novel dataset by incorporating noise (images provided by Dr. Rittik Deb) into a pre-existing Kaggle dataset(containing only bee images), and create an ML model trained on this dataset which can identify bee subspecies from photographs. If possible, achieve higher accuracy than previously published model, else create a novel dataset.

Dataset

- https://www.kaggle.com/code/gpreda/honey-bee-subspecies-classification
- Dr.Rittik Deb - Camera trap photograph

Relevant Papers

- De Nart, D., Costa, C., Di Prisco, G. et al. Image recognition using convolutional neural networks for classification of honey bee subspecies. Apidologie 53, 5 (2022). https://doi.org/10.1007/s13592-022-00918-5
- Kelley, W., Valova, I., Bell, D. H., Ameh, O., & Bader, J. (2021). Honey sources: neural network approach to bee species classification. Procedia Computer Science, 192, 650–657. https://doi.org/10.1016/j.procs.2021.08.067
- Spiesman, BJ, C Gratton. RG Hatfield, WH Hsu, S Jepsen, B McCornack, K Patel, G Wang. 2021. Assessing the potential for deep learning and computer vision to identify bumble bee species from images. Scientific Reports 11:7580.

Work Split

All the work will be shared to smaller or larger extent, the work split here quoted is merely indicative of the general idea.

- Sreerag - Running models, annotating odd serial number data points.
- Srinivas - Documentation, annotating even serial number data points.

Baselines

SVM, KNN, Random Forest, CNN

What we have done

- Read relevant papers, defined the problem a hand concretely and finalized the project
- Identified relevant dataset(s) and acquired the raw photograph

What to do by midway

- Complete preparation of the modified dataset (noise in the kaggle dataset) and run some baseline models to find the accuracy
- Decide, based on the results, whether to push for improved accuracy or shift to annotating the complete raw images

Expected Result

- Either creating a improved ML model for identifying bee subspecies from photographs
- Or publishing a novel dataset with annotated photos of bees and their subspecies