Data Science Capstone: Relevant Sources Report
Cheadle Center for Biodiversity and Ecological Restoration

Explorations of relevant sources influencing our capstone project

Introduction

Through our Data Science Capstone project work, we have found various sources relating to the identification of bee and insect species. The following sources have either been tested during project development or can provide useful insight for further research beyond the scope of our project.

Bee Wing

Bee Wing is a Python bee species classification tool that accompanies the idBee project from the University of Wisconsin, Madison. Their random forest model can identify over 20 species of bees with an accuracy of 90%. Features are extracted from each bee wing and saved to a CSV file. This data is then run through their species classifier. Sample data is included for testing of their model.

Limitations

The *Bee Wing* GitHub, which was last updated six years ago, is not well documented. There are various Jupyter notebooks that demonstrate preprocessing, feature extraction, and classification efforts without detailed explanation. Since the repository contains many scripts, analyzing their code may be time consuming. Since *DeepWings* builds upon this project, further exploration of this code is unnecessary.

Links

GitHub: https://github.com/machine-shop/beewing?tab=readme-ov-file

Accompanying report: https://idbee.ece.wisc.edu/

DeepWings

DeepWings is a Python bee species classification tool that builds off the code from *Bee Wing* and is inspired by the idBee project from the University of Wisconsin, Madison. The pretrained models support the identification of 21 species from the Agapostemon, Bombus, Ceratina, Lasioglossum, and Osmia genera. Additionally, instructions are provided to train the models for new species.

There are three different models contained in this repository: convolutional neural network (Dense Net 121), artificial neural network, and random forest. Additionally, there is a segmentation feature that identifies the geometric shapes created from the wing vein intersections.

Outcome

We attempted to use the trained Dense Net model and the segmentation feature to train their neural network. As the *DeepWings* repository was last updated five years ago, the models are saved in a file type that is no longer accessible with the current version of TensorFlow. The version of TensorFlow used by *DeepWings* requires an older version of Python, so utilizing their pretrained model is inconvenient.

The segmentation code can be successfully used after minor changes to the code to accommodate package updates. The segmentation process identifies the geometric shapes of the regions created by wing vein intersections. After segmentation, the algorithm attempts to filter the wing regions to only identify the middle cells. However, their region filtering process invalidated our test images for use in their neural network, as the code requires at least 6 valid regions to proceed. As a result, we did not test their neural network.

While the DeepWings models were not able to be successfully tested or implemented, their segmentation process is unique. Their methods can be useful for future exploration of wing geometry. An example of their wing segmentation method output is below.

Figure 1: Cropped test image UCSB-IZC00012323_R.JPG

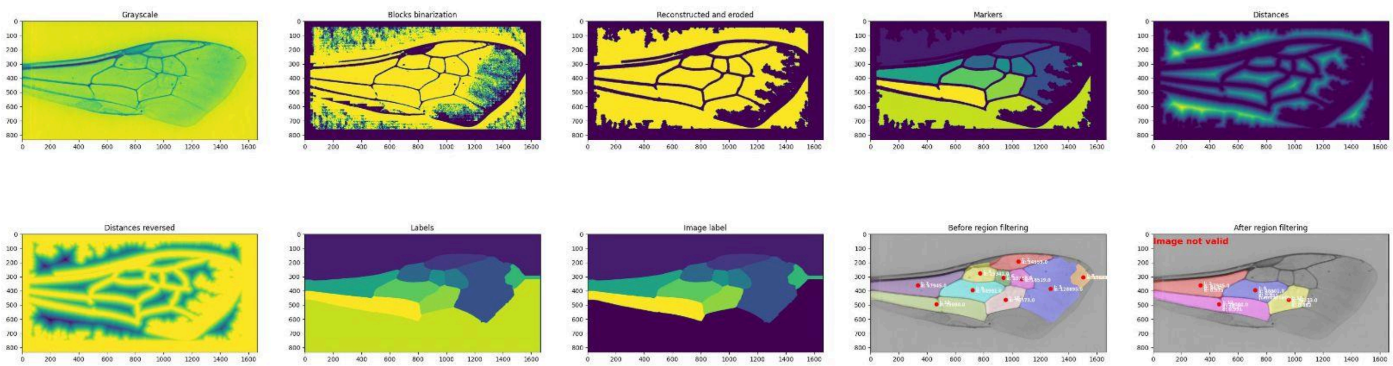

Figure 2: Segmentation outcome for test image

Links

Github: https://github.com/machine-shop/deepwings?tab=readme-ov-file

Accompanying report:
https://github.com/machine-shop/deepwings/blob/master/report/report.pdf

Size and Shape of Insect Wings

This repository accompanies the paper “Size, Shape, and Structure of Insect Wings” by M. Salcedo, J. Hoffmann, S. Donoughe, and L. Mahadevan. The researchers explore various unique ways to analyze and classify wings across insect species. These include: contour (curvature of the wing margin), shape (normalized area and perimeter), venation (summation of all inner vein lengths), vein topology (interconnectedness), and domain shape and size (circularity and size of polygons bound by wing veins). Additionally, the data visualization of their results are exceptional.

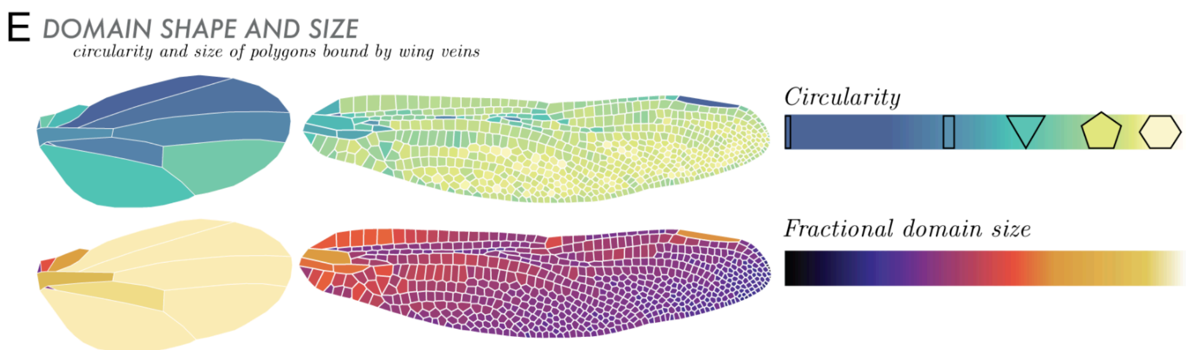

Figure 3: Domain shape and size data visualization example from GitHub

Limitations

This repository was last updated five years ago. The segmentation code is not well documented, so implementing their code in future projects may be time consuming. Additionally, code is not provided to create their graphs and visualizations, which are essential for analyzing results meaningfully.

Links

GitHub: https://github.com/hoffmannjordan/size-and-shape-of-insect-wings

Accompanying report: https://www.biorxiv.org/content/10.1101/478768v1

Ichneumonids of North America

The *Ichneumonids of North America* website is a detailed resource that includes a wasp subfamily classifier based on the frontal view of the head as well as demographic and morphological information of different species of wasps. Users can upload an image of a wasp head, which will return a graph of the probabilities it belongs to each subfamily.

Limitations

The wasp classification code is not publicly available, so more research is necessary to produce a similar facial identification model. However, this website is a unique approach to insect classification, which may be useful for further bee identification research.

Links

Website: https://ichsofna.org/index.html