MosquitoNet: A Deep Learning-Based Classifier for Mosquito Identification

Jay Gupta & Aliza Kajani

General Aims

While the global incidence of malaria has been steadily decreasing over the past decade, the World Health Organization estimates that over 200 million new cases still occur each year [1]. Malaria infection is caused by the *Plasmodium* parasite, which is carried by certain species of mosquitoes. Each of these species may be endemic to a unique geographical region, making high-throughput mosquito surveillance methods valuable for public health researchers. Thus, the objective of this project is to develop a deep learning-based classifier for identification of mosquitoes in images collected by researchers in remote locations.

The classifier will incorporate a set of optimizable parameters that correspond to a physical dimension of the image acquisition process. Specifically, pixel size and camera angle will be considered. First, since the images in the mosquito dataset are high-resolution, they will be downsampled to an appropriate size in order to reduce training time and computational costs. This downsampling factor can be empirically determined through individual experiments, or it can be modeled as an optimizable parameter of a deep neural network. Second, each example mosquito in the dataset was captured from four different angles. The data provided to the network can be varied in order to determine the relative importance of each of these different camera angles, describing an optimal arrangement of the physical imaging apparatus.

Discussion

Primary Dataset

The deep learning model will be trained and evaluated using a dataset that was recently collected and annotated by Duke Global Health Institute researchers, led by Dr. Wendy Prudhomme-O'Meara. This dataset consists of images of 2,236 unique mosquitoes. Each mosquito was imaged 4 times (*i.e.* from 4 different directions) giving a total of 8,944 images in the dataset (Table 1).

Table 1. Overview of Primary Dataset

Species	Number of Examples	Number of Images
Anopheles gambiae	1710	6840
Anopheles funestus	107	428
Anopheles demeilloni	17	68
Unidentified	402	1608
Total	2236	8944

Mosquitoes in the example images belong to the *Anopheles* genus and are one of three species (*gambiae*, *funestus*, or *demeilloni*). A fraction of the examples are unidentified. Furthermore, each example mosquito is in one of the following states: unfed, blood-fed, gravid, or half-gravid. These states correspond to stages in the mosquito gonotrophic (reproductive) cycle and can be determined by visual inspection of a mosquito's abdomen [2].

It is important to note that preprocessing strategies will need to be considered for this dataset. Specifically, the dataset consists of images that were captured at a high resolution. Image downsampling will be required in order to reduce computational complexity and training time. There is also a class imbalance in the data, so the dataset may need to be under/over-sampled in order to prevent bias. An ensemble of models trained on differently resampled data may also be considered, in addition to alternative evaluation metrics (*e.g.* balanced accuracy).

Pre-training Dataset

Due to the limited size of the primary dataset, the model may be pre-trained on the IP102 dataset, a large benchmark dataset recently developed for insect pest recognition [3]. IP102 consists of 75,222 images of insects belonging to 102 classes (Figure 1). These insects may share fundamental morphological features with mosquitoes, thus it is possible that pre-training on IP102 may speed up model convergence and accuracy on the primary dataset.

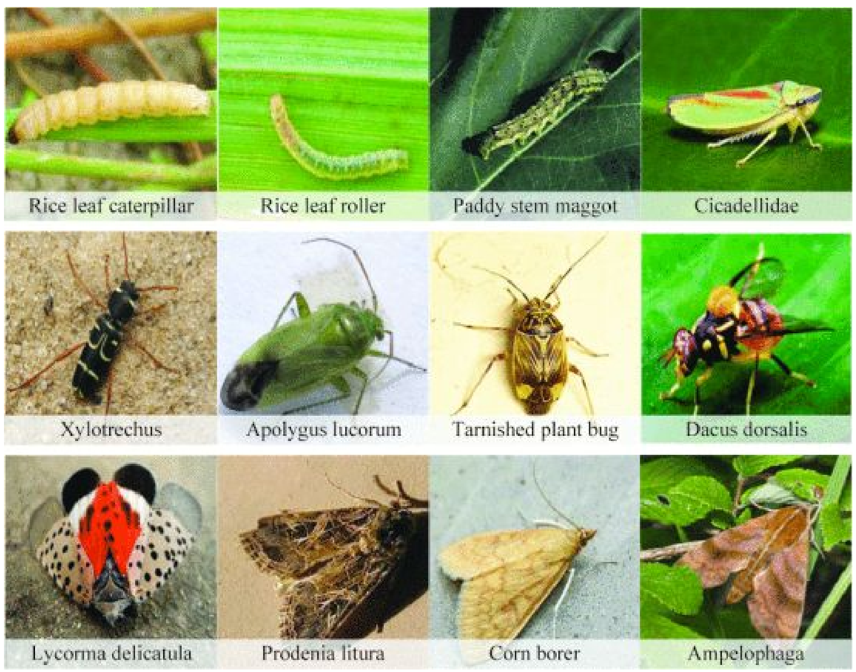

Figure 1. Example images in the IP102 dataset.

Expected Convolutional Neural Network

The convolutional neural network (CNN) is expected to be developed using existing state-of-the-art models, such as AlexNet. This network employs “five convolutional layers, some of which are followed by max-pooling layers, and two globally connected layers with a final 1000-way softmax,” as shown in Figure 2 [4]. AlexNet has been applied in previous work on insect pest detection, and may therefore be a good choice for the mosquito classification task described here [5]. Other groups have also utilized residual network architectures (*e.g.* ResNet-50) and these may be considered as well if AlexNet is unsuccessful [6]. Of note, AlexNet contains over 60 million trainable parameters. Due to the limited size of the dataset, the model ultimately implemented may need to be simplified in order to retain generalizability.

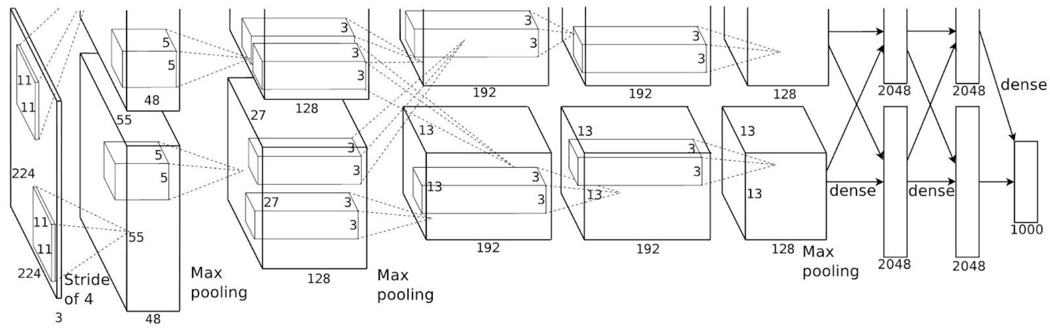

Figure 2. Illustration of AlexNet CNN model architecture.

Expected Simulations

The following are examples of simulations that are expected for this dataset and the chosen model architecture, independent of the physical dimensions of the image acquisition process that are also manipulated:

1. AlexNet trained on primary dataset
2. AlexNet trained on an oversampled primary dataset to reduce imbalance effects
3. AlexNet trained on an undersampled primary dataset to reduce imbalance effects
4. Ensemble of AlexNets trained on differently resampled primary datasets
5. AlexNet pre-trained on IP102 and fine-tuned on primary dataset
6. Train models to classify mosquitoes by species, independent of gonotrophic state
7. Train models to classify mosquitoes by gonotrophic state, independent of species
8. Train models to classify mosquitoes by both species and gonotrophic state

Physical Layer Simulations and Quantitative Analysis

As described in the general aims, pixel size and camera angle will be considered as physical dimensions of the image acquisition system.

First, optimal pixel size will be determined by either running the model on datasets that were resized using different downsampling factors, or by including image resizing as the first CNN layer with an optimizable downsampling factor. In the first case, model accuracy will be plotted for each of the downsampling factors tested. In the second case, model accuracy and loss will be plotted for each epoch of training and validation. The optimized downsampling factor will be reported.

Second, optimal camera angle will be determined by training the network multiple times, on subsets of the primary dataset. Only images taken at the same camera angle will be shown to the network. Model accuracy after training with each specific camera angle will be compared in order to determine whether one view outperforms the others. Furthermore, an ensemble of four models, each trained on data taken from a different camera angle, could also be considered. The final layer of the network will make a prediction by simultaneously using a weighted average of classification results for all four images. Analysis of the quantitative weight values could shed light on the relative importance of each camera angle.

References

[1] World Health Organization, “Global Health Observatory (GHO) data,” *WHO*, 2019.

[2] M. Service, *Medical Entomology for Students: Fifth Edition Mike*. 2012.

[3] X. Wu, C. Zhan, Y.-K. Lai, M.-M. Cheng, and J. Yang, “IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition,” 2019.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” 2012.

[5] L. Nanni, G. Maguolo, and F. Pancino, “Research on insect pest image detection and recognition based on bio-inspired methods,” Oct. 2019.

[6] F. Ren, W. Liu, and G. Wu, “Feature Reuse Residual Networks for Insect Pest Recognition,” *IEEE Access*, vol. 7, pp. 122758–122768, 2019, doi: 10.1109/ACCESS.2019.2938194.