Deep Learning-Based Prediction of Mosquito Decapitation and Salivary Gland Extraction Success for Automated Harvesting of Mosquito Salivary Glands for Malaria Vaccine Production

Final Report

Team Member

Alan Lai
alai13@jhu.edu

Mentors

Dr. Russell Taylor
rht@jhu.edu

Balazs Vagvolgyi
balazs@jhu.edu

May 11, 2021

Acknowledgements

I would like to thank my mentors Dr. Russell Taylor and Balazs Vagvolgyi, as well as both current and past members of the LCSR-Sanaria team, for their support with this project. Their help and guidance were instrumental in achieving the goals set for this project. I would also like to acknowledge the role of Sanaria, Inc. (Rockville, MD) in this project, and thank them for their support.

Purpose

The goal of this project is to provide ROS-integrated computer vision guidance to an automated mosquito dissection robotic system for live malaria vaccine production. The main tasks of this research project include:

1. Deep learning-based prediction of mosquito decapitation success
2. Deep learning-based prediction of mosquito salivary gland extraction success
3. ROS-integration of aforementioned deep networks as ROS Service and Client nodes
4. Thorough documentation of all algorithms, usage, and design choices

These prediction tasks will occur in the process flow when the mosquito neck has been aligned with the actuating blades. The proposed algorithm(s) will either return a predicted successful decapitation and squeezing, in which case the system will continue as is, or return a predicted failed decapitation and/or squeezing, in which case the system will instruct the robot to reposition the mosquito.

Background & Relevance

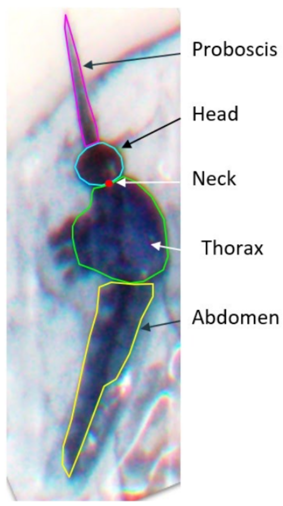


Figure 1. Parts of a mosquito

Malaria is a mosquito-borne disease caused by a single-celled organism of the Plasmodium group that can affect humans. Spread by mosquitoes carrying the parasite, malaria causes symptoms such as fever, tiredness, vomiting, headaches, seizures, and even death. There were over 200 million clinical cases of malaria, over 435,000 deaths, and over $12 billion USD loss in Africa in 2017 alone.[1] Despite the impact that malaria has, there currently exists no effective malaria vaccine available in the market. However, Sanaria, a biotechnology company based in Rockville MD, has recently been successful in developing a live malaria vaccine that has shown to be 100% effective in clinical trials. These vaccines are made from attenuated Plasmodium falciparum sporozoites (PfSPZ), the most common parasite that causes malaria. Because of the nature of these live vaccines, they must be cultivated within live mosquitoes, and hence must also be extracted from mosquito salivary glands, exposed by decapitating the mosquito by its neck, as shown in figure 1., before being able to be used as a vaccine.

Currently, the process to create this vaccine is tedious and slow, as it requires manual extraction of the attenuated PfSPZ from salivary glands using syringes. To scale up production, manual operations that constrain the production of the vaccine must be replaced by automatic systems. Previous work has been done by Schrum et al. to create a semi-automatic system for mosquito dissection.[2] Though training time decreased substantially and processing capacity increased by at least two fold, the process was still manually tedious and labour intensive. Hence, an autonomous robotic system is currently being developed by LCSR in order to automate the process. The proposed full workflow for automated PfSPZ extraction is outlined in figure 2. below.

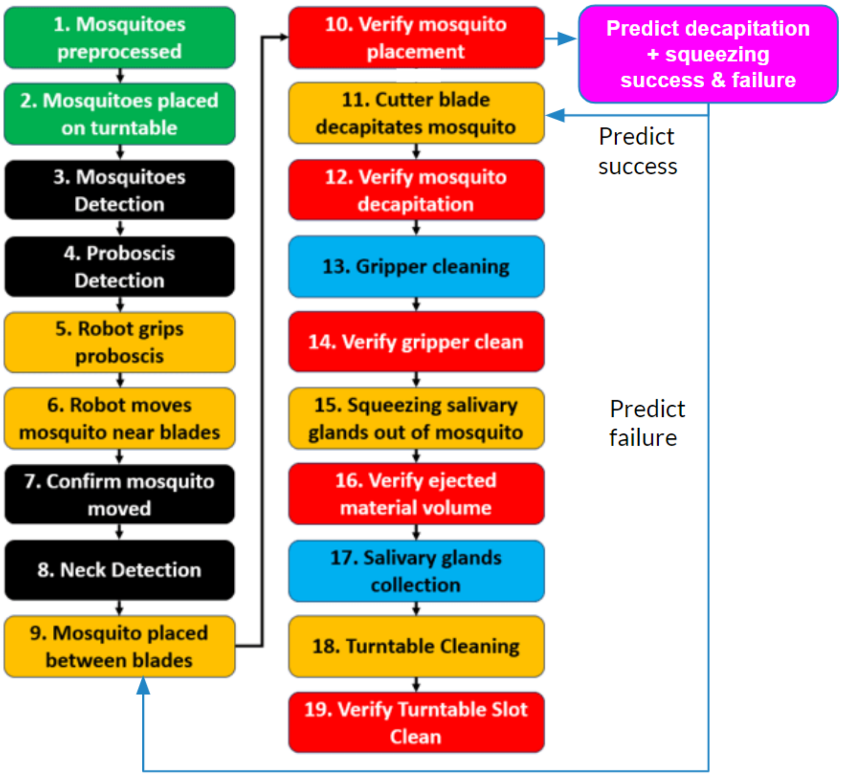


Figure 2. Proposed workflow for automated PfSPZ extraction

The green states in the diagram represent manual preprocessing steps. The black states correspond to completed computer vision tasks, and red states represent incomplete/proposed computer vision tasks. Yellow states represent implemented hardware/robot actuation steps, and blue states represent incomplete/proposed hardware/robot actuation steps. Finally, the magenta state represents the step targeted by this project.

Following the diagram, the mosquitoes are first processed and placed on the rotating wheel, where the mosquito detection and pose estimation algorithms will be used to locate the mosquito and the proboscis (the magenta outlined area in figure 1). The robot grabs the mosquito by the proboscis, and moves it to the staging area, before the cutter. Once the mosquito is moved to the cutter area, the neck is detected, and the robot aligns the neck with the cutter, ensuring that when decapitation occurs, the salivary glands are exposed. In tandem with a series of proposed verification steps, salivary glands are squeezed out of the mosquito, and the turntable cleaned of debris.

From the black states, we can see that previous work has already been done using computer vision to support and guide the current robotic system for automatic mosquito processing. Image processing-based methods include mosquito detection, proboscis detection, and neck detection, while deep learning methods include mosquito orientation classification, mosquito detection, and pose estimation. Many of these methods are already currently in use, as seen in figure 2., providing support to automate the system.

Though those methods have been successful in allowing the robotic system to operate autonomously, comparatively little amounts of effort have been put in with regards to vision-based validation and error prevention of key steps. That is, most existing methods are responsible for guiding the robot's motion to process mosquitoes but only one method, the confirmation of mosquito movement, has been developed to facilitate error checking and recovery. All other error checking and validation shown as red states in the diagram, though proposed to be automated, are currently done manually by the operator and resolved manually as well.

Two key steps in the processing pipeline are the decapitation step and the squeezing step. Though the current vision algorithm for proboscis and neck detection work well, the variety in mosquito shapes, orientation, flexibility, and morphology means that even if keypoints are identified well and the robot moves through the correct motions, the mosquito may not be placed in the correct location, causing issues in decapitation and/or squeezing steps. Even through the proposed verification steps, there is currently no way of resolving these issues if mosquito decapitation or squeezing is ineffective, and the corresponding mosquito would be best case discarded or worst case create problems in the processing pipeline. If an algorithm is capable of predicting, based on mosquito positioning and orientation after the robot has placed the mosquito on the decapitation blades, whether decapitation and/or squeezing will fail, this presents an opportunity for the robot to reposition the mosquito to a valid position such that such wastage would not occur.

This prediction algorithm is the goal of this project, and represented by the magenta state in figure 2. We propose that after the mosquito has been aligned with the actuating blades, that the prediction algorithm performs a prediction classification task using both overhead images taken from the MMS camera, and side view images taken by a side view camera facing the decapitation blades. If the model predicts either a failed decapitation and/or squeezing, the robot will be instructed to reposition the mosquito such that a successful decapitation and squeezing will be more likely. If the model predicts a success for both decapitation and squeezing, the process will continue as is. This way, the number of failed decapitation/squeezes can be minimized, drastically decreasing the potential wastage of mosquitoes, and preventing blockages/issues in the pipeline that may be a result of decapitation and/or salivary gland extraction failures.

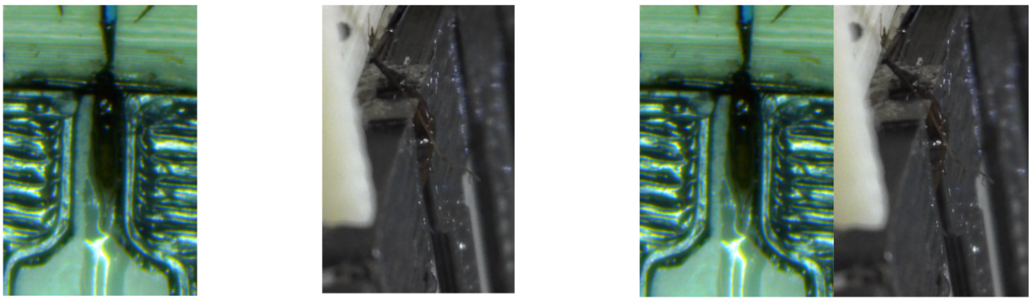
Technical Approach

Deep Learning vs Traditional Image Processing

The reason why deep learning is preferred as a method over image processing with regards to prediction of decapitation/squeezing success/failure is that there are many failure cases that could occur, be it mosquito being too far up, too much to the side, too much in front, which all depend on individual mosquito morphology. It would be impossible to come up with an empirical set of rules for image processing to enforce, and hence deep learning, which allows for greater flexibility and robustness with respect to mosquito morphology, is a more promising solution to this problem. The other advantage of deep learning over image processing is that image processing requires knowledge of the causes of the problems and a method of modeling those causes in order to recognize and resolve the issues. As we currently do not know what causes these issues in decapitation or squeezing, deep learning will be able to allow us to circumvent the need to explicitly model the system.

The other reason as to why we are using deep learning is to be able to leverage the inherent feature representations in deep networks to be able to learn more about what aspects in mosquito morphology and/or placement contributes to successful decapitation and squeezing. By using GradCAM or other class activation map techniques to observe what important features are for the classification task, we hope to be able to delineate these features to be able to aid robotic system design in order to minimize the chances of failed decapitations and squeezings. Though this is not one of the goals of the project, efforts done for this project will facilitate exploration of this analysis.

Deep Learning Based Prediction of Mosquito Decapitation and Salivary Gland Extraction Success



Overhead view | Side view | Combined view

Figure 3. Input Image Styles for DL-Based Prediction

Top view and side view images (shown above in figure 3.) will be taken after the robot has placed the mosquito onto the decapitation blades, but before the blades are actuated. If possible, mosquito images on the turntable prior to robot manipulation may also be taken and incorporated into the training. After the blades are actuated, the operator will manually determine whether the decapitation was a success. Then, the mosquitoes will be further processed and salivary glands extraction will be attempted via squeezing. Once the mosquitoes have been squeezed, the operator will then manually determine whether the squeezing was a success. The decapitation and/or squeezing success will then be written to a file that links these results to the original top and side view images taken of the mosquitoes prior to decapitation.

These images will then serve as training images for training our neural network. PyTorch will be used for training, and transfer learning, using PyTorch's pretrained networks, will be used. This will simply be a classification problem, predicting classes:

1. Success for both decapitation and squeezing
2. Failure for decapitation
3. Failure for squeezing

Note that more classes may be added in the future that may further delineate differences in how successful the squeezing step was.

Images will be split 75-25 for training and validation sets, and parameters such as optimizer, learning rate, batch size, and other potential data augmentation schemes will be attempted in order to both (1) determine the feasibility of prediction of mosquito processing success (2) create a predictor that attempts to predict the result of the decapitation and/or squeezing steps. Models will be trained using the overhead images only, using the side view images only, and using a concatenation of side and overhead views, as shown in figure 3. After training occurs, an inference pipeline will be created, and then used for testing with a new test set. Subsequent to testing, a ROS test client and server will be created in preparation of integration with the larger system.

Materials & Methods

Deep Learning Datasets

The dataset used in this training consists of 122 image pairs consisting of a top-view image taken by the overhead MMS camera, and a side view image taken by a blade-facing side view camera, as seen in figure 3. These images are taken after the mosquito has been aligned with the actuating blades, and labels of the images consist of three classes:

1. Successful decapitation and successful squeezing (51 total)
2. Successful decapitation and failed squeezing (42 total)
3. Failed decapitation (29 total)

Note that there is no need to further quantify failed decapitation as the squeezing step is not attempted after a failed decapitation. The images shown in figure 3 are cropped versions of the whole image captured by the camera, as only information within those regions of interest are useful for our task. In order to combine the side view and overhead images, the images were concatenated horizontally. Classification labels for the training images were labeled solely by Wanze Li, who was in charge of data collection. Labels are initially embedded as part of the file names of the images, but are then manually entered into a CSV file that serves as an annotation file to all the training images.

Deep Learning Training Pipeline

PyTorch is used as the deep learning framework to train the models, due to my familiarity of using PyTorch in previous computer vision tasks in development of the mosquito microdissection system. Transfer learning is also preferred, due to our very limited dataset. Furthermore, previous work done on orientation classification of mosquitoes has shown that transfer learning results in far superior classification results as compared to training a model from random initialization. For this initial training, a standard ResNet18 model is used, due to its smaller capacity suited for our small dataset, and also because of the ease of implementation and training. Because of the limited number of images for training, many data augmentation techniques were used to bring more variety into the training data, including normalization of intensities, random erasing of portions of the images, and random horizontal flips of the images. As the problem is a classification problem, cross entropy loss is used as the loss function, and the Adam optimizer with default momentum and learning rate parameters used for training.

Deep Learning Evaluation

Evaluation of prediction classification was done by computing the number of correct classifications compared to the total number of images. Furthermore, at each training iteration, the confusion matrix is also computed, so that further metrics, such as specificity and sensitivity, can also be computed.

Results & Discussion

Decapitation Prediction

Overall, with the overhead images, a trained ResNet18 model was able to obtain 92% accuracy in making the correct classification of whether decapitation would be successful or not. The confusion matrix corresponding to this is shown below in table 1. Side view images on the other hand were not as effective, reaching only 88% accuracy, while the combined images showed a similar 92% accuracy performance compared to using the overhead images.

Table 1. Confusion Matrix for Decapitation Prediction with Overhead Images

		Ground Truth	
		Success	Fail
Prediction	Success	20	1
	Fail	1	3

With 92% accuracy for a prediction classification problem, the algorithm works very well in predicting whether a decapitation will be successful or not. This will definitely allow the system to be able to reposition mosquitoes if they are predicted to have unsuccessful decapitations.

To further improve on the accuracy of the model, more data needs to be collected and used for training. Though it is a simple binary classification problem, because of the nature of deep learning, 122 images is not enough to train a fully robust system. Furthermore, the lack of training samples also makes evaluation of the models difficult - with only 25 images in the validation set, a misclassification of a single image results in a 4% change in accuracy. Hence such a small dataset is unable to fully reflect incremental gains that specific augmentations and parameters may provide. This may also be what is contributing to the lack of improvement as seen by using both overhead and side view images in a combined manner. Without more data, it is difficult to find ways to combine the images in such a way that the network is able to get relevant information from both images to make a better informed classification decision.

Another area of improvement would be in the models that we use. Currently, all models trained are ResNet18 models, as their capacity is not too large to allow it to easily overfit to our small dataset. In the future, with more data, models with higher capacity, such as ResNet152 or other variations, can be used to see if improvements in accuracy can be obtained. Furthermore, alterations to the standard models can also be considered. An example is instead of concatenating the overhead and side view images into a single image for classification, two neural networks can be used for training, where both overhead and side view images can be trained on their individual models, with their final feature vectors combined in order to make a classification. This allows each model to specialize in a specific image, potentially allowing more relevant, image-specific features to be identified and used for classification.

Salivary Gland Extraction Prediction

On the other hand, results for salivary gland extraction are not as impressive. With the overhead images, a trained ResNet18 model was able to obtain only 76% accuracy in making the correct classification, matched by the combined images, and again with a lower 70% accuracy when using the side view images. Shown below in table 2 is a sample confusion matrix corresponding to the validation set used for training the network.

Table 2. Confusion Matrix for Salivary Gland Extraction Prediction with Overhead Images

		Ground Truth	
		Success	Fail
Prediction	Success	8	2
	Fail	2	5

We can clearly see unimpressive results for this problem, especially considering that random guessing in a binary classification problem results in a 50% accuracy. We can infer from this that salivary gland extraction success prediction is a much harder problem, which makes sense given that we are attempting to predict an outcome several steps downstream in the processing pipeline, where it is dependent on both the mosquito, hardware components, and software parameters. Furthermore, this squeezing prediction is hampered even further by the limited dataset problem, since only 93 image pairs have squeezing labels, as compared to 122 images total. Hence this model faces the same issues as the decapitation prediction, with difficult training, difficult evaluation, and inability to use higher capacity models due to the limited dataset size.

Another reason for the poor performance of this prediction is related to the veracity of the labels for the dataset. Currently, all images are labeled by a single student, which in terms of consistency is good. However, the student has no experience with actual processing of collected salivary gland material, and hence bases the success on visual inspection of squeezing results. There may be cases where the extrudate visually looks good, but may be unusable in terms of vaccine production. Expertise is needed in order to make accurate judgments about whether the squeezing step is truly a success or not. Hence in the future, once the squeezing station design has been finalized, images should be sent to Sanaria/those with processing expertise to judge and label.

Another issue with labeling is label consistency. The labels that we have had so far have been collected through experiments spanning a period of over two months. Throughout this time, the squeezing station has undergone many different iterations. As these labels depend on the squeezing station methodology, label consistency is lost when the squeezing station is changed. For example, a squeezing failure in an earlier iteration of the squeezing station may be a squeeze success if a later iteration of the squeezing station is used. Hence the lack of label consistency means that accurate/good training results for squeezing should not be expected. Related to this, decapitation results are also highly dependent on implemented CV algorithms. Because parameters of these CV algorithms may change between testing, there may also be a lack of consistency in terms of image labels for mosquito decapitation. As the hardware becomes finalized, subsequent training with new samples will allow more robust and accurate models to be trained.

Conclusion and Significance

It is clear that work done in this project has demonstrated the feasibility of the use of deep neural networks to predict both decapitation and salivary gland extraction success. With more and more varied labeled data collected in the future, more robust and accurate models can easily be trained from the frameworks set up from this project for use in the final system. With 92% accuracy achieved by the decapitation success prediction model, model training for that classification task is considered a success. On the other hand, salivary gland extraction, with 76% accuracy, is somewhat disappointing. However, the issues with the limited dataset and labels that contribute to this lower number have been identified in the results and discussion section, and hence future efforts will be aimed at reducing the impacts of those limitations.

Because of the limitations due to the limited data collection and aforementioned issues with regards to the label consistency, much effort was diverted away from obtaining good results towards building a framework that will enable future training to be much more easily done. Both the training framework and ROS server/client scripts have been set up such that training and operation parameters are controlled by configuration files that make modification of parameters much easier. Students/team members who wish to continue efforts done in this project simply need to update parameters in the relevant configuration files to achieve desired training and ROS integration results.

With the ultimate goal of mass-producing Sanaria's live malaria vaccine in mind, the progress made in this project with the vision components of the dissection system represents a large step towards a fully/semi-automated system. The prediction of downstream process success will definitely enable more cases of failures to be caught and dealt with before they happen, limiting the number of mosquitoes wasted and possible blockages to the pipeline. This will surely help streamline the mass production of the malaria vaccine, enabling us to save millions of lives worldwide.

Management Summary

Deliverables

- Working algorithm + training framework on mosquito decapitation success prediction
- Working algorithm + training framework on salivary gland extraction success prediction
- ROS integration test client/server for prediction algorithms
- Documentation on all algorithms and usage

All planned deliverables were achieved. Note that the initial goal of analyzing trained networks for clustering and mosquito morphology analysis for guidance on robot design was made impossible due to the lack of data; hence that deliverable was removed prior.

Future Work

It goes without saying that with more data in the future, better and more robust models can be trained. Furthermore, models that are trained in the future can be used to analyze mosquito morphology and/or position, using GradCAM or other class activation map techniques. This will enable us to interpret class-determining features that will help guide the design and decisions about physical hardware and robot control software, hopefully further decreasing the chances of decapitation and salivary gland extraction failures of occurring.

Lessons Learnt

- Technical skills learnt (deep learning and image processing techniques; ROS integration)
- Importance of data and data augmentation in training deep networks
- Limitations in data can have a large role in driving design choices
- Error handling and error checking are vital for the correct operation of a integrated system

References

1. “Sanaria's Mission.” Sanaria, 2020, sanaria.com/about/.
2. M. Schrum, A. Canezin, S. Chakravarty, M. Laskowski, S. Comert, Y. Sevimli, G. S. Chirikjian, Stephen L. Hoffman, and R. H. Taylor, "An Efficient Production Process for Extracting Salivary Glands from Mosquitoes", arXIV, 2019, http://arxiv.org/abs/1903.02532.
3. Wu, H., Mu, J., Da, T., Xu, M., Taylor, R. H., Iordachita, I., & Chirikjian, G. S. (2019). Multi-mosquito object detection and 2d pose estimation for automation of PfSPZ malaria vaccine production. In 2019 IEEE 15th International Conference on Automation Science and Engineering, CASE 2019 (pp. 411-417). [8842953] (IEEE International Conference on Automation Science and Engineering; Vol. 2019-August). IEEE Computer Society. https://doi.org/10.1109/COASE.2019.8842953

Technical Appendix

Documentation

Documentation for this project can be found on the following GitLab wiki page.
Please note that only mentors for our project and the team members have access to these Wiki pages.

Source Code

Source code for this project can be found here.
Please note that only mentors for our project and the team members have access to these Git repositories.