DETECTING MOSQUITOES WITH CONVOLUTIONAL NEURAL NETWORKS

A Thesis
Presented to
The Academic Faculty

by

Lawrence Moore

In Partial Fulfillment
of the Requirements for the B.S. Degree
in Computer Science with Research Option
in the College of Computing

Georgia Institute of Technology
2017

DETECTING MOSQUITOES WITH CONVOLUTIONAL NEURAL NETWORKS

Approved by:

Dr. James Hays, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. David Hu
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: 5/1/2017

ACKNOWLEDGEMENTS

I would like to thank Dr. Hays and the rest of the lab for their guidance and help throughout the process of writing this thesis. The opportunity to be part of such a talented and hardworking group was a real privilege. It also would not have been possible without the patience and kindness of my father, who stayed up talking with me countless nights throughout my time at Georgia Tech and constantly supported me. My friends were always there to help, and the last four years would have not have been anywhere near as exciting and enjoyable without them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Conv Net	Convolutional Neural Network
LSTM	Long Short-Term Memory

ABSTRACT

Mosquitoes are directly responsible for the death of more than a million people each year. Yet the ability to mitigate their deadly impact or even monitor them in the wild to better understand their behavior remains relatively limited. One of the primary reasons for this lack of progress is the difficulty in locating and tracking an individual mosquito, leading to only estimates for a population as a whole. To address this problem, this research discusses several approaches using computer vision to detect and track the flight of mosquitoes. In particular, we discuss the performance of several convolutional neural network architectures which show promising results. Once these techniques are refined to give a high enough degree of accuracy, this vision system could be used in conjunction with drones to track and eliminate mosquitoes in both an indoor and outdoor setting.

CHAPTER 1: INTRODUCTION

Mosquito-borne diseases kill several million people every year and sicken hundreds of millions more. Investigating methods to control the spread of pathogens through mosquitoes is thus vital for world health. This issue has increased urgency with the arrival of the Zika virus to the United States this past summer. As a result, there is considerable ongoing work to mitigate the problems mosquitoes pose.

Pesticide, insect zappers, and mosquito nets have all been employed over the last several years to combat mosquitoes. Of these, pesticide is the most effective, but comes at the cost of killing other benign insects and animals [1]. Insect zappers have been shown to do little to mosquitoes themselves, instead killing benign insects [2]. Mosquito nets, while effective for some cases like sleeping [3], are too limited to serve as a general method for mosquito control.

More advanced techniques have recently been developed, such as the use of gene editing to eliminate mosquitoes from breeding [4]. The problem with this approach is that it could kill the entire species, leading to unknown damage to the ecosystem. A recent, less invasive system is being researched that uses lasers to eliminate mosquitoes [5]. The system first detects the mosquitoes using infrared light as they pass in front of a retro-reflector and then eliminates them using a laser [5]. The disadvantage of this prototype is that the system requires a large set-up, the range is limited, and the system cannot adapt to a changing environment.

All of the methods presented so far have either serious side effects or are restricted in their application. To avoid both problems, we hope to develop a non-invasive system that

uses cameras to detect mosquitoes and then either drones or lasers to eliminate them. This study will focus specifically on the detection and tracking of mosquitoes.

Great progress has recently been made in object recognition and tracking through the of use deep learning, a machine learning technique [7]. By using large amounts of data to train, deep learning can learn to classify, localize, and track objects with high accuracy. This type of approach has been mainly used for larger, more feature rich objects like people and cars and little work has been done to see how this technique works for smaller objects on the scale of a mosquito. There has been some work on classifying insect types [20], but this was done with an older and less powerful approach than deep learning. The emphasis of drone's vision system will be detecting a small mosquito in a large scene rather than classifying an insect as a specific species given a close up image.

Considering the size and relative lack of features of a mosquito when present in an image, an approach will likely need to utilize the movement of the insect over time to pinpoint its location. To do this, a stack of frames can be used as input to the neural network, giving some sense of the movement over time. Another approach would be to feed the estimated location of the mosquito in the previous frame to the network when evaluating the current frame, as performed by Schenck and Fox do in [16]. This would allow the network to build confidence over time as the movement of the mosquito leaves small signatures in the frames.

Once the system to detect mosquitoes in a video captured by a camera is developed, either a laser or drone could be employed to eliminate the mosquito. Lasers have been demonstrated to be effective in killing insects, and the system in [5] would have greater flexibility and reduced size if it employed a camera for detection as opposed to the IR and

retro-reflector approach. This is because the IR detector is large, and setting up the reflector further limits the spatial environments in which it can be employed. Small drones could also be used. By quickly moving in on a mosquito and killing them by sucking them up, drones could provide similar functionality to the laser system without being limited to one static position. Either of these options would provide an improved and much needed tool to combat the spread of disease through mosquitoes.

CHAPTER 2: LITERATURE REVIEW

The Problem

Mosquitoes kill around 750,000 people each year - almost double as many as humans, the next deadliest animal [19]. Though their bite itself is harmless, mosquitoes carry a number of diseases, such as yellow fever and malaria. These diseases are then spread to the animal when bitten. But despite the serious threat these insects pose, existing tools can do relatively little to stop them without harming the surrounding environment.

Existing Approaches

There is considerable ongoing work to develop methods to reduce the damage mosquitoes can inflict. One common method is the use of pesticide. Pesticide can be deployed over a large area, either on the ground or through aerial vehicles, and is fairly effective in killing mosquitoes [1]. However, due to the lack of precision of this method, the use of pesticide can also eliminate benign insects such as bees [1]. Even fish and wildlife can be affected if the concentration of pesticides is too great [1]. This side effect is a particularly problematic given the rapid decline in bee population the last several years.

Traditional, less invasive methods have limited effectiveness. Mosquito nets can be effective for sleeping but can only be used for small areas [3]. Bug zappers are ineffective in attracting and eliminating mosquitoes, often killing benign insects instead [2]. Further, bug zappers can actually still spread the disease contained in the insect, as the process of being electrocuted releases particles containing bacteria and viruses [6].

A more recent technique works on the genetic level. With this approach, female mosquitoes are infected with a gene that causes 99% of their offspring to become sterile. If enough mosquitoes carried this gene, the mosquito population could be decimated in a

matter of years [4]. While this might be effective in temporarily eliminating the spread of disease, the technique raises a number of issues, both practical and ethical. It is a major ethical question whether we should willfully eliminate an entire species. Further, there could be unanticipated side effects of spreading such a gene.

More precise methods for detection and elimination of mosquitoes have recently been developed. Chen et. al outline a method that uses the frequency of the beating of the insect's wings to classify and detect the type of insect [10]. Using a laser source coupled with phototransistor array, Chen et. al records the wing beating frequency of the insect, and by using this measure along with others such as the time of day and geographic location, classify the insect using k Nearest Neighbors (kNN) [10]. Similarly, Mullen et al. describe a system that detects and eliminates mosquitoes using lasers [5]. The system first detects the mosquitoes using infrared light as they pass in front of a retro-reflector and then eliminates them using a high intensity laser [5]. The disadvantage of both of these approaches is that they require a large set-up and the range is quite limited, rendering them ineffective in more flexible and dynamic situations, such as in a house or a large piece of agricultural land.

Convolutional Networks for Detection and Tracking

One of the biggest constraints in non-invasively eliminating mosquitoes in systems such as the one in [5] is the not the actual elimination mechanism, but rather the ability to accurately detect the mosquitoes in 3D space. The detection approaches in [5] and [10] both require sophisticated equipment and lack flexibility and range. Given a better way of detecting insects, a system utilizing lasers or other means of elimination such as small drones capable of sucking up the insects could become the new benchmark in vector and pest control.

The goal of this research is to develop methods to detect mosquitoes utilizing just a simple video camera. This problem falls under the domain of computer vision. Indeed, computer vision has undergone a revolution in the last half-decade as a result of deep learning, a machine learning technique. Convolutional Neural Networks (CNN's), a subset of deep learning, have been particularly transformative. Previously, features in images were hand selected, such as in [9]. A classifier could then be trained on these image features for tasks such as object recognition. What makes these new techniques so powerful is that they automatically learn the feature representations in images that give the best accuracy for the task at hand.

By using large amounts of data to train, CNN's provide the state of the art performance in localizing and tracking objects [7]. However, this type of approach has been used mainly for larger, more feature rich objects like people or cars. Little work has been done to see how this technique scales for smaller objects like a mosquito in an image. Several deep learning methods can be used for the problem of detecting mosquitoes in a video. Among these, object localization is a viable approach, which is the task the finding a bounding box around an object of interest. Another is object tracking, which is the problem of assigning each pixel in an image to either the background or the object of interest in each frame.

There are several approaches to object localization. LeCun *et al.* describe a technique called Overfeat that uses a sliding window to localize an object [11]. To find the bounding box of the object, a window at various scales is slid across the image, keeping track of the predicted label of that window and its associated confidence. At the end, the windows with high confidences are merged to provide non-overlapping windows around

objects. Another more recent technique uses the Regions with CNN Features (R-CNN) approach [13]. In this technique, regions of interest are first extracted, using a method such as selection search [12]. Then for each image, features are extracted through a CNN and are then passed through an SVM classifier [13]. This approach results in higher accuracy than Overfeat and also requires less data [13].

The disadvantage of using object localization to detect an object in each frame is that it generally evaluates each frame of a video independently. Object tracking does not. Object tracking can be approached from different perspectives. Some methods, such as those discussed in [14] and [15], learn a general representation of the object of interest or use features derived from layers in a deep network to represent the object; this allows the object to be detected frame by frame under a wide variety of settings. Other approaches more directly model the temporal relationship between frames. Schenck and Fox outline how liquid can be tracked using recurrent neural networks (RNN's) [16]. Though liquids are relatively featureless, by using the estimated previous location as input to the network along with the current frame, high accuracy can be achieved in tracking liquids through a frame. Similarly, Wei *et. al* show how different body parts can be located in a video even with a large amount of occlusion and variance in spatial layout by propagating the belief map from the previous frame to the next [17]. In a similar vein, Alahi *et. al* demonstrate that RNN's are effective for predicting the movement of an agent in a scene as well [8].

One possible problem with the CNN approach for objects as small as mosquitoes is that the spatial information may be lost as the input and subsequent features are propagated through the network. One possible way to address this is through the use of skip nets. He *et al.* first introduced the concept of skip nets in [18], which lead to the new

state of the art in computer vision tasks across the board. The idea is simple: instead of just passing the output of one layer to the next one in the network, the output "skips" to a deeper layer. This has the effect allowing larger networks to be trained more easily, and also helps maintain spatial information that might be altered by a convolution or pooling layer.

CHAPTER 3: METHODS

There are two components to the system described in this paper. The first is the labeled training data, and the second is the convolutional neural networks employed for the task. Because labeled training data of real mosquitoes in flight has not yet been gathered, synthetic training videos must be employed instead. In addition, three variants of convolutional neural network architectures will be employed. One single frame architecture, which only looks at a single frame as input, and one multi-frame architecture, which looks at chunks of frames at a time.

Data Generation

To provide enough videos for a convolutional neural network to train on, synthetic data was generated[1]. Black circles were superimposed onto static images to simulate the presence of the mosquitoes. The number of circles was randomly chosen between 1 and 5, and the radius of each is randomly chosen between 1 and 3 pixels. To simulate movement, the circles changed location slightly per frame, with 10 frames per second. The circles consisted of pixels of various degrees of darkness, so that the network would have to detect general regions of darkness instead of a perfect circle. The initial direction of movement was chosen randomly and with some small chance the mosquito changes direction. Initially, 90 different images were chosen as possible backgrounds, with 80 being dedicated to training and 10 to testing. The images were taken randomly from Google images by searching for “bedroom”, “living room”, and “patio.” The images were further distorted by white noise in each frame to avoid the network overfitting to the

[1]Full code can be found at https://github.com/Lawrence-Moore/mosquito-tracking

static background. When the videos were created, a corresponding matrix is saved which labels each pixel in a frame as belonging to the background or mosquitoes. Around 16,000 training images and 2,000 testing images were generated.

Network Architectures

All networks used are fully convolutional. A learning rate of 1e-4 with the Adam optimizer was used, and the learning rate was decreased exponentially over time with a rate of 0.1. Batch sizes of 40 were used along with 20,000 total iterations as that is when the loss leveled out. To address the inherent class imbalance problem of this dataset, a weighted softmax loss with weight equal to 500 was employed. Further, the network was trained on crops of size 64x64 that randomly include a mosquito.

Single-Frame Architecture

The single frame network begins by like a traditional conv net, with each convolution followed by a ReLU. For fully connected layers, 1x1 fully convolutional are used, with rectified linear units after each. To generate full size image predictions, three strided deconvolutional layers are used. This is shown in Fig. 1.

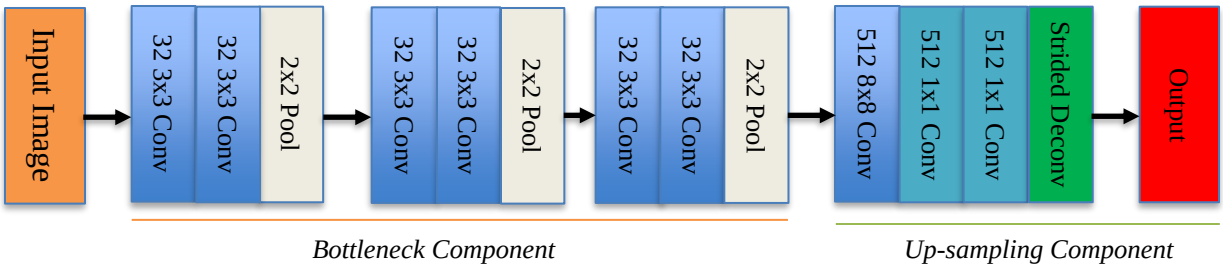

Fig. 1: Single-frame Neural Net Architecture

Multi-Frame Architecture

Twenty frames at a time were fed as input to this architecture. Each frame is first passed through the bottleneck component of the net as shown in fig 1. The resulting frames are then concatenated together channel wise and passed through the 1x1 fully convolutional layers and strided convolutional layers as before. The prediction is made on the 20th frame.

Evaluation

To evaluate the performance of the networks for detecting the mosquitoes, the problem was treated as an object detection task. To generate a prediction in the form of a box containing the mosquito from the pixel by pixel output of the network, a sliding window approach was employed. If a box contained at least 60% mosquito pixels, it was considered a viable prediction. Non-maximum suppression was then used to find unique mosquito predictions. A prediction was considered correct if it was within 10 pixels of the actual mosquito location. A precision-recall curve was then generated.

CHAPTER 4: RESULTS

Figure 2 shows a qualitative measure of the performance of the two architectures when detecting a mosquito with a clear background. The mosquito location is correctly outputted.

Fig. 2: Qualitative Performance. Cropped region from original video (left). Prediction with Single Frame and Multi-Frame Architecture (middle and right).

A quantitative measure of the performance is shown through the precision recall curves for both architectures in figure 3. The multi-frame architecture outperformed the single-frame architecture.

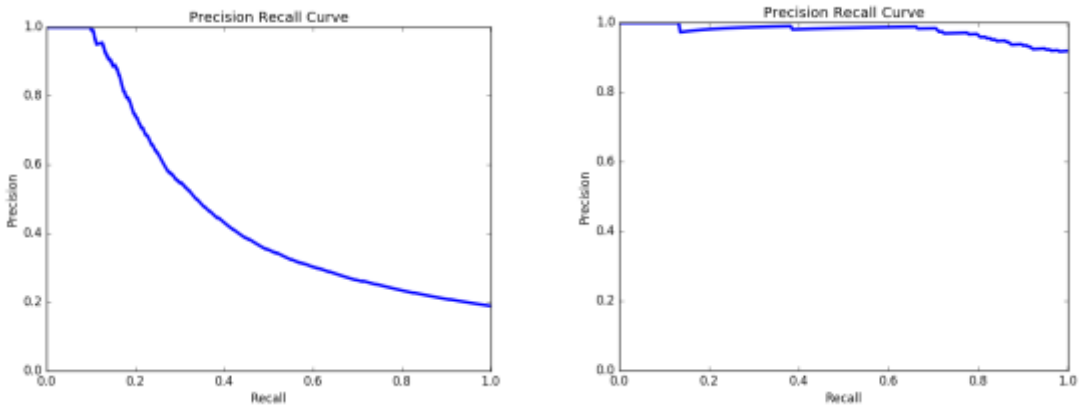

Fig. 3: Precision Recall Curve for Single-Frame Architecture (left) and Multi-Frame Architecture (Right)

CHAPTER 5: DISCUSSION AND FUTURE WORK

As we can see, the networks show promising results in terms of detecting mosquitoes. While the data is synthetically generated, the high confidence predictions almost always correspond to a mosquito. As expected, the multi-frame approach does seem to provide a significant increase in performance, specifically in terms of recall. This is probably because knowledge of previous frames helps eliminate potential false positives that correspond to random noise in an image. Given this, for future work, the use of a recurrent architecture should be tried. Since the multi-frame approach seems to result in higher confidence and lower false positives, it is likely that using a neural net with some sense of time and knowledge of previous location estimates would similarly be effective. This trend in architecture performance has been shown in the literature, so it certainly warrants investigation.

The next step is to see how these networks perform on data of real mosquitoes in flight. This will not only directly gauge the feasibility of the system for the intended task but also inform how the simulation could be updated to better reflect the difficulties of real data. For instance, to add more realism, actual footage of mosquitoes can be used in the simulations by cropping out the mosquito and super-imposing it on one of the many background images currently being used. It is unlikely that enough real footage can be gathered to solely train the networks, so this hybrid approach of using a mixture of simulated and real data will likely be the most effective moving forward. It also likely that once the data is sufficiently complex, a recurrent or multi-frame net will exhibit even greater advantages over the single frame approach, as the expressive power will be more fully utilized.

REFERENCES

1. R.I Rose, “Pesticides and public health: Integrated methods of mosquito management”, *Emerg. Infect. Dis.,* 7 (2001).

2. P. Alonso, S. Lindsay, J. Armstrong, et al, “A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of the Gambia, West Africa. 2: Mortality and morbidity from malaria in the study area,” *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 87, Suppl. 2), 1993.

3. University Of Florida, Institute Of Food & Agricultural Sciences, "Snap! Crackle! Pop! Electric Bug Zappers Are Useless For Controlling Mosquitoes, Says UF/IFAS Pest Expert," *ScienceDaily*, 1997.

4. A. Hammond, R. Galizi, K. Kyrou, et al, “A CRISPR-Cas9 gene drive system-targeting female reproduction in the malaria mosquito vector Anopheles gambiae”, *Nature Biotechnology*, Vol:34, 2016.

5. E. Mullen, P. Rutschman, N. Pegram, J. Patt, J. Adamczyk, E. Johanson, "Laser system for identification, tracking, and control of flying insects," *Opt. Express 24*, 2016.

6. J.E Urban, A Broce, “Killing of flies in electrocuting insects traps releases bacteria and viruses”, *Curr. Microbiol.*, 41, 2000.

7. Y. LeCun, Y. Bengio, G. Hinton, “Deep Learning”, *Nature,* 521, (2015).

8. A. Alahi, K.Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, S. Savarese, “Social LSTM: Human Trajectory Prediction in Crowded Spaces”, *CVPR 2016*, 2016.

9. D. Lowe, “Distinctive Image Features from Scale-invariant Keypoints”, *IJCV*, 60 (2), 2004.

10. Y. Chen, A. Why, G. Batista, A. Mafra-Neto, E. Keogh, “Flying Insect Detection and Classification with Inexpensive Sensors.” *J. Vis. Exp.* (92), 2014.

11. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks”, *ICLR 2014*, 2014.

12. J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, “Selective search for object recognition”, *IJCV*, 2013 .

13. R. B. Girshick, J. Donahue, T. Darrell, and J. Malik,“Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation”, *Computer Vision and Pattern Recognition, CVPR* 2014.

14. N. Wang and D.Y. Yeung, “Learning a deep compact image representation for visual tracking”, *NIPS*, 2013.

15. C. Ma, J. Huang, X. Yang, and M. Yang, “Hierarchical convolutional features for visual tracking”, *ICCV*, 2015.

16. C. Schenck and D. Fox, “Detection and tracking of liquids with fully convolutional networks,” *arXiv:1606.06266,* 2016.

17. S.E. Wei, V. Ramakrishna, T. Kanade, Y. Sheikh, “Convolutional Pose Machines”, *CVPR 2016,* 2016.

18. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, *arXiv:1512.03385,* 2015.

19. The Gates Foundation.

20. A. Lu, C. Hou, C. Liu, “Insect Species Recognition Using Sparse Representation”, *Proceedings of the British Machine Vision Conference*, 2010.

21. K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”, *CoRR, abs/1409.1556*, 2014.

22. Greff, Klaus, Srivastava, Rupesh Kumar, Koutnik, Jan, Steunebrink, Bas R, and Schmidhuber, Jurgen, “Lstm: A search space odyssey”, arXiv preprint arXiv:1503.04069, 2015.