CLASSIFICATION OF MOTION REGIONS WITH CONVOLUTIONAL NETWORKS, SUPPORT VECTOR MACHINES, AND RANDOM FORESTS IN VIDEO-BASED ANALYSIS OF BEE TRAFFIC

V.A. Kulyukin[1], S. Mukherjee[1], P. Vats[1],
A. Tiwari[1], Y.B. Burkatovskaya[2]
[1] *Utah State University, Logan, USA*
[2] *Tomsk Polytechnic University, Tomsk, RUSSIA*
e-mail: `vladimir.kulyukin@usu.edu, tracey@tpu.ru`

Abstract

Bee traffic is the number of bees moving in a given area in front of a specific hive over a given period of time. Video-based bee traffic analysis has the potential to automate the assessment of bee traffic levels, which, in turn, may lead to the automation of honeybee colony health assessment. In this paper, we evaluate several convolutional networks to classify regions of detected motion as BEE or NO-BEE in videos captured by BeePi, an electronic beehive monitoring system. We compare the performance of several convolutional neural networks with the performance of support vector machines and random forests on the same image dataset.

Keywords: convolutional network, support vector machine, random forest, bee traffic, data science

1 Introduction

Many beekeepers watch bee traffic to ascertain the state of their honey bee colonies, because bee traffic carries information on colony behavior. Bee traffic patterns change in response to stressors such as failing queens, predatory mites, and airborne toxicants. While experienced beekeepers can tell changes in bee traffic levels in stressed colonies, they may not always be able to determine the exact causes of the changes without hive inspections. Unfortunately, hive inspections disrupt the life cycle of bee colonies and put additional stress on the bees. Since beekeepers cannot monitor their hives continuously due to obvious problems with logistics and fatigue, a consensus is emerging among researchers and practitioners that video-based analysis of bee traffic levels can become an integral component of electronic beehive monitoring and help extract critical information on colony behavior and phenology without invasive beehive inspections and considerable transportation costs.

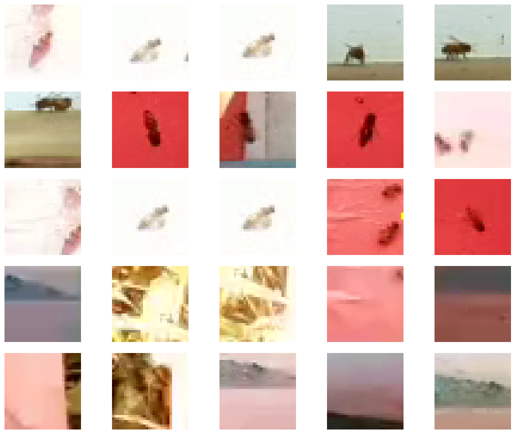

Figure 1: A sample of images from BEE1; the first 3 rows include images classified as BEE; the last 2 rows consist of images classified as NO-BEE.

In this investigation, we focused on training and evaluating three types of classifers – convolutional networks (ConvNets) [1], random forests [2], and support-vector machines (SVM) [3] – to categorize motion regions detected by the video-processing algorithm of BeePi, a multi-sensor electronic beehive monitoring system we designed and built in 2014 [4], and have been iteratively modifying since then [5].

Once mounted on top of a Langstroth beehive, a BeePi monitor captures a 30-second 360x240 video every 15 minutes from 8:00 to 21:00 at a frame rate of 25 frames per second. Each captured video is processed for motion detection. We experimented with three motion detection algorithms available in OpenCV 3.0.0 (www.opencv.org): KNN, MOG, and MOG2. Although all three algorithms performed on par, we found that MOG worked slightly better than either KNN or MOG2, because it was less sensitive to shadows.

The output of the motion detection module is a set of 32x32 image regions centered around detected motion points. Fig. 1 gives a sample of detected motion regions. Each detected motion region is classified by a trained classifier (e.g., a convolutional neural network, a random forest, or an SVM) into two classes – BEE or NO-BEE. Thus, for each video, the video processing algorithm returns an estimate of the number of bees that moved in a given region in front the beehive over a 30-second period.

2 Image Data

The image dataset for this investigation was obtained from the videos captured by four BeePi monitors placed on four Langstroth hives with Italian honeybee colonies. Two monitors were deployed in an apiary in Logan, UT, USA and the other two – in an apiary in North Logan, UT, USA from April 2017 to September 2017.

We randomly selected 40 videos from June and July 2017. The image dataset was then obtained by using the MOG algorithm to automatically extract 54,392 32x32 motion regions from the videos (see Fig. 1). We obtained the ground truth classification by manually labeling the 54,392 32x32 images with two categories - BEE (if it contained at least one bee) or NO-BEE (if it contained no bees or only a small part of a bee). The image dataset from the apiary in Logan, UT was used for model training and testing. The image dataset from the apiary in North Logan, UT was used for model validation.

We executed the ANOVA and MANOVA analyses on the labeled image dataset to determine whether its training (class 0), testing (class 1), and validation (class 2) images are statistically significantly different. We used the following image features as independent variables in our analysis – contrast, energy, and homogeneity. The dependent variables were class 0, class 1, and class 2. The MANOVA analysis with 1 degree of freedom gave the Pillai coefficient of 0.018588, the F value of 107.47, and $Pr(> F) < 2.2e-16$. The ANOVA analysis on contrast gave the mean squared value of 2.3607, the F value of 148.2, and $Pr(> F) < 2e-16$. The ANOVA analysis on energy gave the mean squared value of 6.237, the F value of 313.2, and $Pr(> F) < 2e-16$. The ANOVA analysis on homogeneity gave the mean squared value of 0.18043, the F value of 152.2, and $Pr(> F) < 2e-16$. Since in all cases the P value is < 0.0005, the three training, testing, and validation datasets are significantly different in terms of contrast, energy, and homogeneity.

3 Experiments

We performed an exhaustive search for an optimal ConvNet by starting with 1 hidden layer and 1 max pooling layer and varying the filter size, the number of hidden layers, and the number of nodes in each hidden layer. An addition of a hidden layer was always followed with an addition of a max pooling layer with a kernel size of 2. All hidden layers used the ReLU activation function and the adam optimizer. The learning rate was set to 0.001; the loss function was the categorical cross entropy. All models were trained for 50 epochs with a batch size of 50. Table 1 gives a summary of the ConvNet architectures constructed with exhaustive search. Increasing the number of hidden layers above 6 did not improve the results of the best performing 5-hidden layer model.

Table 1: Best ConvNet models discovered through exhaustive search.

Num. HLs	Test Loss	Test Accuracy	Valid Accuracy
1	0.09	97.93%	50.34%
2	0.03	99.31%	74.58%
3	0.03	99.34%	78.29%
4	0.02	99.46%	81.63%
5	0.03	99.32%	86.57%
6	0.03	99.51%	83.29%

We also designed several ConvNets by hand. The best performing manual ConvNet had 8 layers: input, output, and 6 hidden layers. The first 7 layers use ReLU as their activation function. The last layer uses sofmax and categorical cross entropy. The first two hidden layers use batch normalization. The last 5 layers use a dropout with a keep probability of 0.5. Table 2 gives the confusion matrix of this ConvNet on the validation dataset.

Table 2: Confusion matrix of best hand-crafted ConvNet.

	No-Bee	Bee	Accuracy
No-Bee	1677	15	99.11%
Bee	1	1809	99.99%
Total Accuracy			99.54%

To obtain some standard machine learning benchmarks, we trained and tested SVMs and random forests on the same image dataset. All SVMs used the linear kernel with the max-iter parameter varying from 10 to 1000. The best performing SVM on the validation dataset had a max-iter of 1000 and an accuracy of 51.13%. The confusion matrix for this SVM is given in Table 3.

Table 3: Confusion matrix of linear SVM on validation dataset for max-iter 1000.

	No-Bee	Bee	Total	Accuracy
No-Bee	822	896	1718	47.84%
Bee	828	982	1810	54.25%
Total Accuracy				51.13%

We trained and valided random forests with 10, 50, 80, and 100 trees. The best performing random forest on the validation dataset had 50 trees and achieved an accuracy of 93.67%. In summary, random forests performed much better than the SVMs and all the ConvNet models obtained through exhaustive search but were 6% below the hand-crafted ConvNet.

Table 4: Confusion matrix of random forest with 50 trees on validation dataset.

	No-Bee	Bee	Total	Accuracy
No-Bee	1664	54	1718	96.85%
Bee	169	1641	1810	90.66%
Total Accuracy				93.67%

References

[1] LeCun, Y., Kavukcuoglu, K., Farabet, C. (2010). Convolutional networks and applications in vision. *Proceedings of IEEE International Symposium on Curcuits and Systems*, pp. 253-256.

[2] Breiman, L. (2001). Random Forests. *Machine Learning*, Vol. **45**, pp. 5-32.

[3] Hong, J.H., Cho, S.B. (2008). A probabilistic multi-class strategy of one-vs.-rest support vector machines for cancer classification. *Neurocomputing*, Vol. **71**, pp. 3275–3281.

[4] Kulyukin, V.A., Putnam, M., Reka, SK. (2016). Digitizing buzzing signals into A440 piano note sequences and estimating forage traffic levels from images in solar-powered, electronic beehive monitoring. *Lecture Notes in Engineering and Computer Science: Proceedings of the International MultiConference of Engineers and Computer Scientists*. Vol. **1**, pp. 82-87.

[5] Kulyukin, V.A., Mukherjee, S., Amlathe, P. (2018). Toward audio beehive monitoring: deep learning vs. standard machine learning in classifying beehive audio samples. *Applied Sciences*, Vol. **8**, pp. 1573-1606.