Detection and classification of insects on stick-traps in a tomato crop using Faster R-CNN

Ard Nieuwenhuizen
Agro Food Robotics
Wageningen University and Research
Droevendaalsesteeg 1, 6708 PB Wageningen
The Netherlands
ard.nieuwenhuizen@wur.nl

Jochen Hemming
jochen.hemming@wur.nl

Hyun Suh
davidsuh79@gmail.com

Abstract

In this paper we present the method and performance to detect tomato whitefly and its predatory bugs on yellow sticky traps. These traps are imaged in controlled light conditions with a digital single lens reflex camera and in uncontrolled environment with smartphone camera. The method consists of the following steps. First, image sub setting and data labelling by manual annotation. Secondly, training a deep learning convolutional neural network. Third step is classification of the images. Final step is comparison with hand counted data of insects. The weighted averaged accuracy for deep learning detected insects was 87.4%. The correlation of hand counted insects with deep learning counted insects was over 0.95 for the smartphone images. The methods used show that the training data used on controlled conditions could be transferred to uncontrolled smartphone imaging conditions for the data provided.

1. Introduction

In Europe *Trialeurodes vaporariorum* and *Bemisia tabaci* (greenhouse whitefly) in tomatoes are listed in the top 10 of the most problematic pests in greenhouse vegetable crops [5]. Within the framework of the CIPM PeMaTo-Europep project attention is given to integrated pest management to control these pests. Biological control in Belgian and Dutch tomato greenhouses is based on the release of the predatory bug *Macrolophus pygmaeus* and *Nesidiocoris*. This predator interacts with every pest and beneficial in the greenhouse. Therefore the level of the predator and the problematic insect should be monitored accurately. Monitoring of the insects in the greenhouse gained a lot in interest and is the basis of integrated pest management. Insects can be attracted to traps and monitored on these traps. Counting and classification of insects is however time consuming and error prone as it is only partly automated by classical thresholding and blob counting algorithms. Current methods for counting rely on specific hardware for recording the so called yellow sticky traps in a Scoutbox [1] and count and identify insects by hand or partly automated procedures. However more accurate numbers of insects are required for population models and other cameras like smart phone cameras would increase the uptake of this technology by farmers more easily. Therefore the idea was launched that trained operators provide labeled training data for use in a deep learning convolutional network to detect multiple types of insects. In addition to labeling and training on images recorded under controlled conditions, also images were fed to the trained network that were recorded under uncontrolled conditions recorded with smartphone cameras.

2. Methodology

Fuentes *et al*. [2] compared the performance of different deep network architectures for plant diseases and pests detection and showed that Faster R-CNN could effectively recognise plant diseases and pests with the ability to deal with complex scenarios from a plant surroundings area. In this project one of the most recent deep neural networks, Faster R-CNN with inception Resnet v2 [6], was used for insect detection and counting on yellow sticky traps.

2.1. Scoutbox images

Images were recorded with a Scoutbox, under controlled conditions with a resolution of 5184 x 3456 pixels. These images were recorded on two greenhouse locations in Belgium. From the 6900 recorded images a subset of 225 images was randomly chosen to represent variability in insect populations to be expected. Each image was split into six section of 1720 x 1220 pixels to fit to an image size being

Figure 1. Three example insects macrolophus, whitefly and nesidiocoris from left to right respectively.

accepted by the CNN for training and classification.

2.2. Smartphone images

In addition to the Scoutbox images, also a dataset was available being recorded with smartphone camera. The dataset consisted of 90 images recorded with resolution of 4608 x 3456 pixels. These images originated from five different greenhouse locations compared to the images recorded by the Scoutbox. The images were also split into six sections each to be accepted in the deep learning and classification pipeline.

2.3. Human image annotation

For image annotation by experts from greenhouse research stations in Belgium and Spain, the open source program LabelImg was used. Three insects were labeled, whitefly (WF), macrolophus (MR) and nesidiocoris (NC) as shown in Figure 1. In the Scoutbox images dataset that consisted of 225 x 6 = 1350 image slices, the following number of insects were present. WF 5611, MR 1314 and NC 511. These numbers represent the natural population densities of these insects in commercial tomato greenhouse crops.

2.4. Deep learning convolutional network training

The Scoutbox dataset of 1350 image slices was used to perform a training and classification with the deep learning convolutional network. The dataset was split in three parts, training, testing and validation with a portion of 0.6, 0.2 and 0.2 of the size of the data respectively. The Faster RCNN network was applied in Tensorflow version 1.5 with the object detection module [3]. When training the network, transfer learning was used. The starting point for the training was the trained network on the MS-COCO dataset [4]. The endpoints, or classes for continuing the training were adjusted to three classes, WF, MR and NC. Training of the network was stopped after maximum of 200.000 epochs, or when the total loss was converged at 0.15, the latter was the case in this study. The training took about 50 hours on a single Nvidia 1080Ti Gpu. At that point the network was saved as a frozen inference to do classification on the validation dataset.

instances		predicted				
		wf	mr	nc	none	total
labelled	wf	355	0	0	242	597
	mr	1	60	0	14	75
	nc	0	0	6	2	8
	none	206	9	3	0	218
	total	562	69	9	258	
percentage		predicted				
		wf	mr	nc	none	total
labelled	wf	59.5	0	0	40.5	100
	mr	1.3	80.0	0	18.7	100
	nc	0	0	75.0	25.0	100
						71.5

Table 1. Number of instances and percentage of detected insects in validation dataset.

2.5. Classification performance

To perform the classification on unknown data several metrics need to be defined to compare classification results with other research. In this study only classifications with a confidence over 90% were considered valid. Furthermore, in accepting a detection as correct the intersection over union (IoU) with the ground truth had to be over 50%. Three types of insects are detected by the classifier, this results in a multiclass confusion matrix explaining the accuracies reached in detection.

3. Results and discussion

Table 1 gives an overview of the results of classifying the validation set of the Scoutbox images. The classification shows that 59% of whitefly is correctly classified. From the total of 597 whitefly in the ground truth, 242 whitefly instances have not been detected. The network detected 206 whiteflies that were not available in the ground truth images. The number of instances of MR and NC are relatively low. The weighted classification accuracy is calculated as well as a function of true positive (TP) and total number of insects. This is calculated as ((TP WF + TP MR + TP NC) / number of insects)*100% and results in 61.9 %. However, this analysis showed that the ground truth quality had to be improved, as many insects were not hand labeled at all, 206 whiteflies were missing that were detected and 242 were not detected due to improper hand drawn labels causing rejects by low intersections over unions. After improving the ground truth hand labeled images, the averaged accuracy of the classification increased to 87.4%. A complete classified yellow sticky trap is shown as example in Figure 2.

For the smartphone images no hand labeled ground truth annotations are available. However the total number of insects on the traps had been counted by human observations on the 90 yellow sticky traps. These manual observations

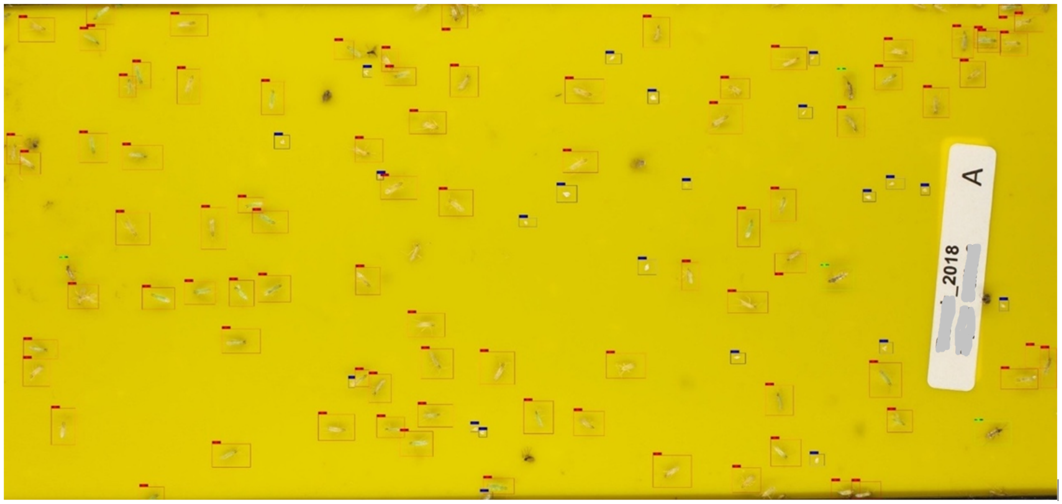

Figure 2. Example results for insect detection on complete yellow sticky trap.

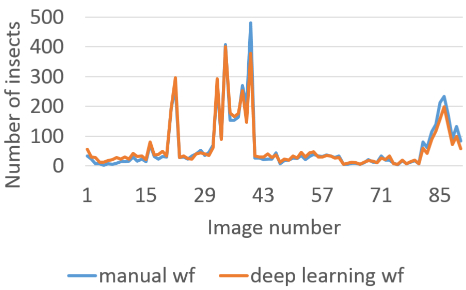

Figure 3. Number of whiteflies detected manually and by deep learning in 90 individual yellow sticky traps imaged with smartphone in uncontrolled conditions.

resulted in 5574 whiteflies, 1592 macrolophus and 26 nesidiocoris over the 90 traps. The deep learning classification resulted in 5521, 390 and 9 insects respectively. The number of detected whiteflies and macrolophus correlate well ($R^2 > 0.8$) with the human counted number of insects on the sticky traps, visualised in Figure 3.

The pipeline to train and classify based on images of yellow sticky traps has been prepared and used in this research project. The project was successful in detecting the insects. However, the quality of hand labeled data and annotations influences the classification accuracy and makes it hard to quantify the success. The trained and frozen network could be easily used for detection of insects in smartphone images. The numbers of insects found in these images correlate well with human counted insects on the images. Future detection tasks can now be sped up by pre labeling images to be checked by human operators. Furthermore the detection and counting of insects will be automated and presented to commercial greenhouse growers as a web service to empower their decision making in integrated pest management.

4. Acknowledgements

We want to thank colleagues from Belgium and Spain within the CIPM Pemato project for their support in annotating many insects in the images that were recorded and used in the research. This work is part of the ERA-Net research programme C-IPM and the project PeMaTo-EuroPep. The work presented here was funded by the Dutch research organization NWO (ALW.FACCE.5) and the Dutch Topsector Tuinbouw en Uitgangsmaterialen (EU-2017-06).

References

[1] Agrocares. Scoutbox specifications, https://www.agrocares.com/en/products/scoutbox/, 2018. 1

[2] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors (Switzerland)*, 2017. 1

[3] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017. 2

[4] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014. 2

[5] G. J. Messelink, J. Bennison, O. Alomar, B. L. Ingegno, L. Tavella, L. Shipp, E. Palevsky, and F. L. Wäckers. Approaches to conserving natural enemy populations in greenhouse crops: Current methods and future prospects, 2014. 1

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, K. He, X. Zhang, S. Ren, J. Sun, A. Rajkomar, S. Lingam, A. G. Taylor, M. Blum, J. Mongan, Y. Gong, K. H. T. Leung, A. T. Toshev, S. Ioffe, Y. Jia, K. He, X. Zhang, S. Ren, J. Sun, Z. Z. Z. Wang, X. X. Wang, G. Wang, O. M. Parkhi, A. Vedaldi, A. Zisserman, Others, O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, Others, L. Wang, S. Guo, W. Huang, Y. Xiong, Y. Qiao, C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, K. He, X. Zhang, S. Ren, J. Sun, G. Böning, J. F. Kahn, D. Kaul, R. Rotzinger, P. Freyhardt, M. Pavel, F. Streitparth, K. Simonyan, A. Zisserman, L. C. Gansel, N. Bloecher, O. Floerl, J. Guenther, K. Kang, H. H. Li, J. Yan, X. Zeng, B. Yang, T. Xiao, C. Zhang, Z. Z. Z. Wang, R. Wang, X. X. Wang, Others, X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, W. Liu, Z. Z. Z. Wang, X. Liu, N. Zeng, Y. Y. Liu, F. E. Alsaadi, H. H. Li, L. Zhang, X. Zhou, B. Huang, A. G. Trofimov, B. M. Velichkovskiy, S. L. Shishkin, S. Wu, S. Zhong, Y. Y. Liu, X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, R. Ramachandra, S. Y. McGrew, J. C. Baxter, J. R. Howard, K. S. Elmslie, N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the IEEE international conference on computer vision*, 2017. 1