דו"ח לתכנית מחקר מספר **14-30-0004**

שנת המחקר:____2_______ מתוך___3________ שנים

פיתוח מערכת ראיה ממוחשבת לגילוי סימני מחלת החירכון הנגרמת על ידי החיידק Erwinia amylovora במטעי האגס

Development of a computer vision system for detecting signs of fire blight caused by the bacterium *Erwinia amylovora* in pear trees

מוגש לקרן המדען הראשי במשרד החקלאות

ע"י

רפאל לינקר הפקולטה להנדסה אזרחית וסביבתית, טכניון

מרי דפני ילין מיגל – מו"פ צפון

Raphael Linker, Faculty of Civil aand Environmental Engineering, Technion. Haifa 32000 E-mail: linkerr@tx.technion.ac.il

Mery Dafny Yalin, Northern Agriculture Research & Development, MIGAL Galilee Technology Center. P.O.B. 831 Kiryat Shemona 11016, Israel. E-mail: merydy@migal.org.il; merydy@gmail.com

תקציר

חירכון (Fire blight) הינה מחלה הנגרמת על ידי חיידק גרם שלילי *Erwinia amylovora* הפוגעת במספר צמחים המשתייכים למשפחת הוורדניים. החיידקים חודרים לעצים דרך הפרחים בזמן תקופת ההאבקה על ידי הדבורים ועלולים לפגוע בפריחה, או דרך פצעים ברקמה הצמחית.
התפתחות מחלת החירכון תלויה במצב פנולוגי של העצים ומלווה בסימנים ויזואליים. גילוי המחלה נעשה בעזרת פקחי הגנת הצומח או החקלאים עצמם. בפועל, פעילויות הניטור אינן מבוצעות בתדירות מספקת, דבר הפוגע ישירות ביכולת של החקלאים לטפל בצורה יעילה במחלה.

מאחר וגילוי המחלה מבוסס על סימנים ויזואליים, היפותזת המחקר הינה כי ניתן לייעל את ההתמודדות עם נזקי החירכון על ידי פיתוח מערכת ראיה ממוחשבת שתזהה באופן אוטומטי את אותם הסימנים. לשם כך הפרויקט מתמקד בפיתוח מערכת ראיה ממוחשבת שתזהה באופן אוטומטי (1) את סימני המחלה בשלביה השונים ו (2) את הפריחה סתווית שהיא מקור חשוב בתהליך ההדבקה.

במהלך השנה השנייה של המחקר בוצעו שלושה סוגים של פעילויות: צילומים בשטח על מנת להגדיל את מסד הנתונים, (2) תיוג של תמונות ו (3) פיתוח מודלים מסוג "מכונה לומדת" - machine learning" כדי לזהות באופן אוטומטי את סימני המחלה בתמונות.

לצורך זיהוי סימני המחלה על הגזע המשכנו בפיתוח גישה דו שלבית בה מודל ראשון מזהה את אזור הגזע בתמונה ומודל נוסף מזהה את סימני המחלה עת הגזע. לאחר כיול המודלים הם נבדקו על 10 תמונות שלא השתתפו בכיול. התוצאות מראות שהמודל הראשון תמיד מזהה את אזור הגזע אך לפעמים מסווג גם אזורים אחרים כגזע. המודל השני מסווג בצורה נכונה רק כ 70% משטח הגזע ודרושים שיפורים על מנת לשפר את התוצאות. בשל מורכבות התמונות ומגוון הנגעים, משימה זו הינה ללא ספק המשימה המאתגרת ביותר בפרויקט.

על מנת לזהות עלים נגועים פתחנו גישה דו שלבית מבוססת על בידוד אזורי עניין בעזרת פילטר מבוסס על צבע וסיווג האזורים הנ"ל ע"י מכונה לומדת. המודלים נבדקו על יותר מ 300 תמונות בהן זוהו יותר מ 10000 אזורי עניין. בסה"כ המכונה סווגה נכון (נגוע/לא נגוע) יותר מ %80 מאזורי העניין.

לגבי הזיהוי של פריחה סתווית, ביחס לשנה ראשונה שינינו מעט את הגישה ויישמנו גישה דומה לזו של זיהוי עלים נגועים: סינון ראשוני (גילוי אזורי עניין) על סמך צבע וסיווג אותם האזורים ע"י מכונה לומדת. גישה זו נבדקה על שתי סדרות של תמונות הכוללות בסה"כ כ 250 תמונות. המכונה זיהתה בצורה נכונה כ %80 מהפרחים בתמונות.

<u>מעריכים מומלצים לבדיקת הדוח המדעי</u>

1. דר' דני שטיינברג – מכון וולקני
2. דר' ויקטור אלחנתי – מכון וולקני
3. דר' עמוס נאור – מיגל

••

<u>הצהרת החוקר הראשי</u>:

הממצאים בדו"ח זה הינם תוצאות ניסויים.

הניסויים מהווים המלצות לחקלאים: לא (<u>מחק את המיותר</u>)

<u>*במידה וכן, על החוקר להמציא פרטים על הגוף שבאמצעותו מופץ הידע (כמו: שה"ם)</u>

חתימת החוקר ____ תאריך: __28 ינואר 2019___

<u>רשימת פרסומים שנבעו מהמחקר</u>:

1. **<u>אין</u>**

Content

Background and agricultural challenge

Fire blight is a disease caused by the bacterium *Erwinia amylovora* that affects plants belonging to the pome fruit tree, Rosaceae family, and in particular pear trees. The *E. amylovora* bacterium enters the trees through flowers or wounds, and spreads throughout the tree via its vascular system. The rate of infection and the extend of the damage can vary greatly from year to year depending on climatic conditions. Fire blight causes a range of visible symptoms, each at a specific phonological stage, which are quite easy to identify for a trained grower/adviser. However, such visual monitoring is time-consuming and an automated system would enable much more frequent and thorough monitoring.

Research objectives

Develop a vision system for automatic detection of

1. Disease symptoms at various developmental stages
2. Autumn blooming

Progress and main findings

We continued the development of convolutional neural networks (CNNs) for identifying flowers and disease symptoms. Due to the rapid technological developments in the field of CNNs we decided to switch the analysis from Matlab to Python (Pythorch) that enables much more efficient use of graphics processing units and reduces the computing time considerably. All the results reported below were obtained by applying transfer learning to the CNN VGG-16. Other models were tested and no significant performance differences were observed.

<u>Task 2.1 – Detection of disease symptoms in winter</u>

We did not acquire new images this year but focused on the re-analysis of images collected during year 1 in order to develop a procedure that would be more efficient computation-wise. We tagged 181 images according to “trunk area” and “diseased area” using rectangular bounding boxes. Statistical analysis showed that the width of the boxes containing a trunk varied from 200 to 1200 pixels due to irregular tree shape, with most trunk being contained in a 600-pixel wide box. After extracting “empty” (no trunk) regions of similar sizes we trained a first CNN to identify image regions containing a trunk. Altogether 725 images were used (80% for training) and the CNN classification performance was 85%. We further checked the performance on 10 full images not used during the training stages and observed that the trunk region was correctly detected in all the images (no “missed detections”) but some “false positive” detections occurred.

The regions tagged as “disease” were resampled with non-overlapping boxes with size 200*600 pixels (200 pixels being the size of the smallest disease symptoms tagged), leading to close to 1000 samples. A similar number of samples of the same size were extracted at random locations of trunks regions not affected by the disease. The dataset size was increased to over 4000 samples via standard augmentation techniques (image flipping and translation) and a CNN was trained using 75% of the samples. The training accuracy was 72%, the test accuracy was 65% and the overall accuracy was 70%. Figure 1 shows a typical result. The dashed rectangles indicate the three regions that the first CNN identified as “trunk”. The red solid rectangles

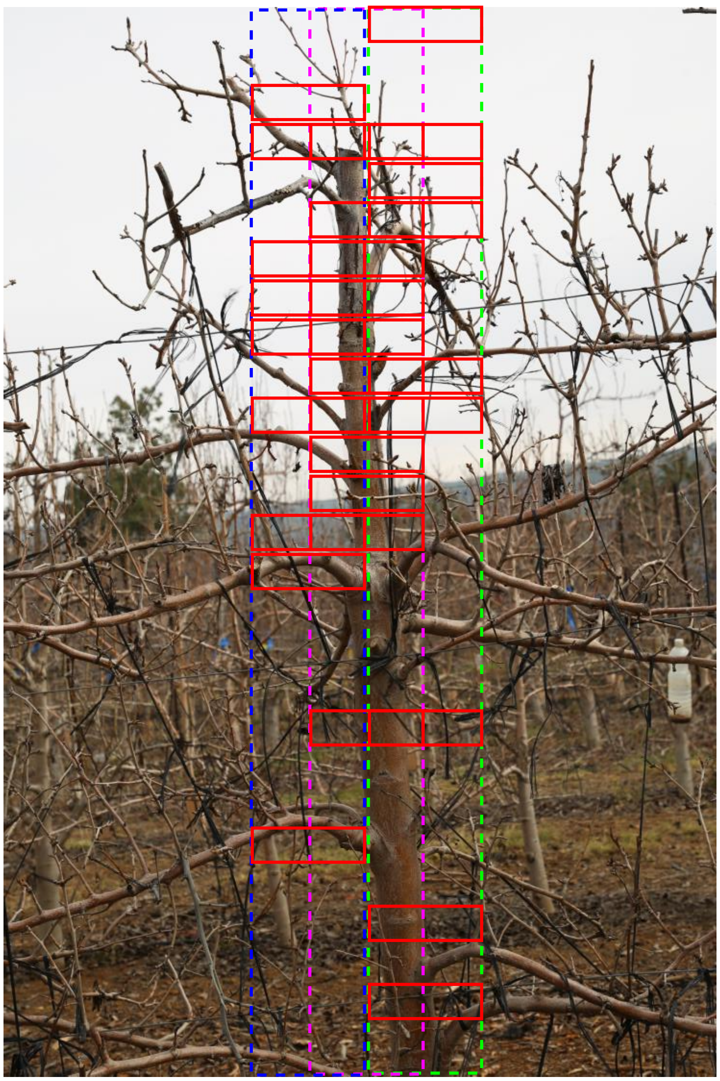

Figure 1: Detection of trunk (dashed rectangles) and fire blight symptoms on trunk (solid rectangles)

indicate the regions that the second CNN identified as affected by fire blight. It can be observed that most areas affected by the disease were indeed detected but there was a number of “false positive” detections.

Task 2.2 – Detection of disease symptoms in canopy

Pictures of trees affected to various degrees by fireblight were acquired on 6 days (See Table 1 below). Each scene was photographed 7 times with varying aperture corrections (from -1 stop to +1 stop at 1/3 stop increments) in order to enable future investigation of the image exposure on the results. Half of the scenes were acquired with a relatively wide angle (24 mm) and the other half were acquired with a narrower angle (50 mm) to determine how the relative size of the affected area affected detection. In each case the camera was placed in a tripod and not moved between the acquisitions of the same scene so that the 7 images would superpose and manual tagging was required only in one image for each scene. Each tagged area has been characterized via two parameters: infection status (1: early infection (only blossom infected), 2: advanced infection (branches blight just started) 3: advanced blight; 4: old infection from previous year) and position (A: image forefront, B: deeper within the image). These parameters were defined in order to enable future in-depth investigation of the relationship between the size and position of the affected area and the probability of detection. However, it must be noted that this information was not used at this point and the results reported below are based on all tagged areas combined.

The tagged areas were extracted and various features were computed in order to identify feature(s) that would enable fast extraction of “suspicious regions”, i.e. regions that have some probability of being infected. The purpose of this step is too greatly reduce the number of regions that have to be analyzed by the much more computation-intensive CNN implemented in the second step. Therefore, this first step was designed such that the number of missed detections was very small, at the expense of high number of “false positive” detections. Simple thresholding of the Hue value (after transforming the RGB in HSV space) was found to provide suitable results. The main drawback of this simple thresholding operation (beside producing many false positives) was that in some cases it produced unrealistically large blobs (mostly due to false detections on branches or soil). Therefore, blobs over some predefined size were eroded iteratively until fragments of suitable sizes were obtained. The blobs that did not overlap with user-tagged areas were extracted to form the “non-infected” training/testing samples. The number of “non-infected” samples was in most cases much larger than the number of “infected” samples and, in order to avoid class imbalance, only some of the “non-infected” samples (selected randomly) were included in the training (75%) and testing (25%) datasets (See Table 1 below). At the testing stage, the original RGB image was transformed to HSV space, H-thresholding was applied, blobs beyond the specified size were iteratively eroded and the resulting “sub-images of interest” were fed to the CNN.

The performances of various models are reported in Table 1. Initially, a separate model was trained with the images of a specific day, mostly as a way to check the consistency of the data. The overall performance was over 85% in most cases although in some cases overfitting occurred due to the small size of the dataset. Next, the images of various days were pooled in order to obtain more robust model. The overall accuracy of this model was over 80% and more detailed results are presented in Table 2. The CNN correctly identified 88% of the infected samples but classified 23% of the non-infected samples as infected.

Table 1: CNN models for detecting fire blight symptoms in canopy

Model	Dataset	Number of samples (infected/not infected)			Overall accuracy (%)		
		Training	Test	Total	Training	Test	Total
6	15/5	829/771	289/272	1118/1043	92.3	90.8	92.4
5	22/5	559/583	187/198	1006/521	95.4	94.0	94.2
9	27/5	180/181	57/52	236/233	94.3	78.1	92.9
3	31/5	1167/1237	423/366	2124/1069	85.8	85.3	85.0
10	5/6	1285/1292	460/457	1745/1749	83.7	78.0	82.1
14	6/6	126/136	40/42	167/178	98.0	85.9	95.0
17	All dates combined	4748/3597	1646/1196	6394/4793	83.0	81.6	82.5

Table 2: Performance of CNN trained on all datasets combined

Status predicted by CNN	Correct status	
	Infected	Not infected
Infected	4949	1308
Not infected	651	4260

Typical examples of the final detection performance are shown in Figure 2. Red solid boxes correspond to correct detections, blue dashed boxes correspond to "false positive" detections, and white dotted boxes indicate "missed detections" (regions that were selected by the HSV filter but incorrectly rejected by the CNN). The yellow "+" symbols correspond to the center of correct "false negatives" regions (regions that were selected by the HSV filter but correctly rejected by the CNN). These figures show that the high rate of "false positive" detection is partly due to detection of dead leaves on the ground. Future work shall be devoted to avoiding/rejecting the detection of such leaves which are not within the canopy.

Figure 2: Typical results of detections of fire blight symptoms in canopy. The red solid rectangles correspond to correct detections, the blue dashed rectangles correspond to false positive detections, the white dotted rectangle corresponds to missed detection and the + symbols indicate correct negative results, i.e. regions selected by the HSV filter but correctly rejected by the CNN.

Task 2.3 – Detection of autumn blooming using images acquired by unmanned aerial vehicle (drone)

The work started in Year 1 continued using the two datasets acquired in Year 1 and 1667 images captured at an orchard near Rosh Pina on November 11, 2019. In order to apply a unified analysis framework, some of the images of Year 1 were re-tagged according to three levels: (1: fully open large flower, 2: partly visible flower or flower missing a few petals, 3: flower missing most (or all) petals or flower hardly visible (due to distance or orientation). As in Task 2.2, the purpose is to enable future in-depth analysis of the detection capabilities but at this stage all tagged flowers were included in the analysis. The training and analysis pipelines were similar to those described in Task 2.2 but the H-thresholding operation was replaced by thresholding in the Value space to account for the white color of the flowers.

For the images recorded in Metula in 2018, the CCN performance was: training: 83%, testing: 79%, overall: 81%. Figure 3 shows a typical result and it can be observed that some of the "missed detections" were in fact not actually missed but resulted from the way the images were tagged: Each flower was tagged individually and in case the flowers formed a cluster the whole clustered was tagged as well. Figure 3 shows that in some cases although the individual flowers were detected correctly, the CNN failed to recognize the whole cluster, or the opposite, the whole clustered was correctly identified but some of the single flowers were not detected. Therefore, the straightforward computation of the "missed detection rate" which has been used so far to quantify the results (based on a simple one-to-one correspondence between the regions detected by the H-thresholding filter and the CCN results) overestimates the false detection rate. Future work shall be devoted to devising a more appropriate way to quantify the CNN detection performance at the tree level.

The results of the CNN trained with the images recorded in Neot Golan in 2018 were slightly worse: Training accuracy: 81%, test accuracy: 72%, overall accuracy: 81%. These poorer results appear to be due to a combination of three factors: (1) Autumn blooming occurred only very sparingly, so that the number of flower samples for training the CNN is very limited, (2) The images were acquired around noon-time under clearsky conditions so that the images tend to be very bright, which causes the H-thresholding filter to have a “false positive” detection rate much higher than in the Metula dataset (ratio of tagged flowers to false positive 1/35 compared to 1/3 in Metula dataset) and (3) the apparent size of the flowers is significantly smaller in this datatset.

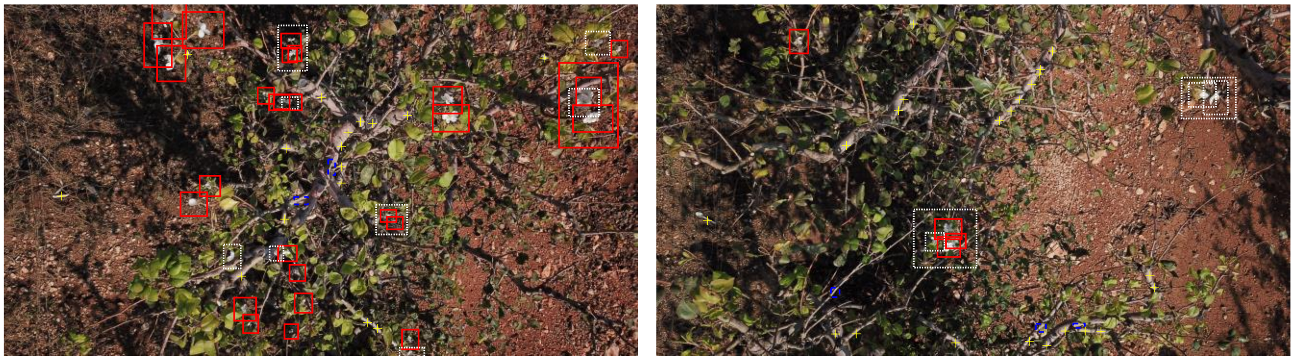

Figure 3: Typical results of detections of autumn blooming. The red solid rectangles correspond to correct detections, the blue dashed rectangles correspond to false positive detections, the white dotted rectangles correspond to missed detection and the + symbols indicate correct negative results, i.e. regions selected by the RBG filter but correctly rejected by the CNN.

Task 2.4 – Visualization of detection results in GIS

All the images collected in the project are geo-tagged so that it is straightforward to present the results in map format. As an example, Figure 4 shows the number of flower detected in every fifth image recorded by the UAV while flying over a specific row. The actual number of flowers in these images is also shown for comparison. The X and Y axes correspond to longitude and latitude, respectively, so that the points can be readily superimposed to an existing map of the orchard (if available).

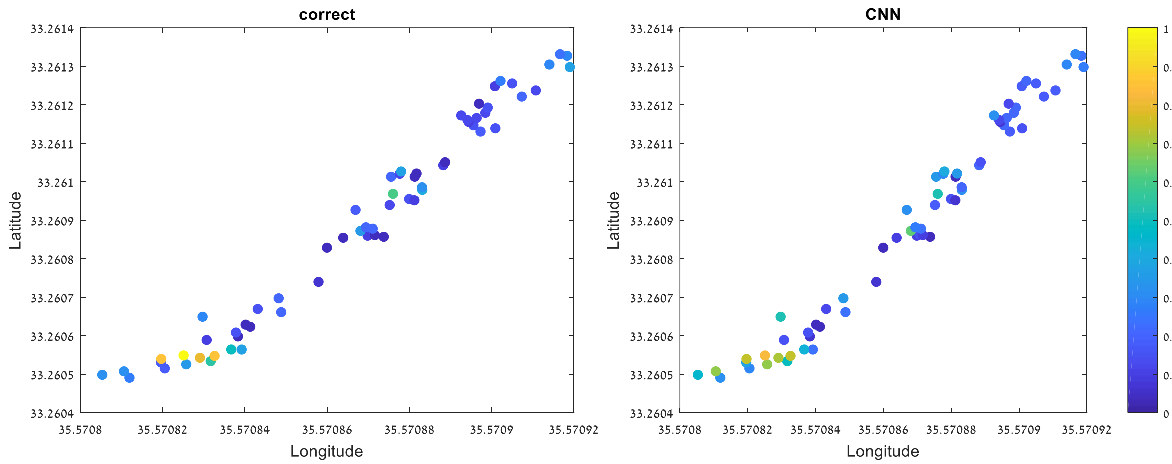

Figure 4: Map of autumn blooming according to automatic detection by CNN (left) and according to manual tagging of the images (right).

Departures from planned workplan and workplan for Year 3

As in Year 1, the occurrence of oozing was very limited (both in terms of locations and time) and we were not able to acquire enough images to create a database suitable for machine learning. The attempt to detect oozing will not be continued in Year 3.

With the exception of oozing, we have successfully trained CNNs to identify all the fire blight symptoms on which this project focuses: autumn blooming, symptoms on main trunk in winter and leaves/flowers affected in spring. However, there is still need for improvements, especially with regard to detection of symptoms on trunks which is the most challenging task. Also, as mentioned above, there is a need to define a performance index that reflects more appropriately the detection performance. Accordingly, the third will be devoted not only to validating the existing models but also analyze (and improve) them. In particular, we will investigate the impact of various factors on the detection performance such as apparent size of the area affected by symptoms, position with the image, values of the threshold used at the first stage of detection. As mentioned above, image tagging was done in such a way that these investigations could be conducted but it was not possible to perform this analysis in Year 2 due to time constraints.