Applying image segmentation method to estimate the number of irregular overlapping objects – Taking the use of YOLOv8 to estimate the quantity of pineapple angiosperms in open farmland as an example

Chien-Hung Lai
Department of Electronic Engineering
National Taipei University of Technology
Taipei, Taiwan
ORCID: 0000-0003-1391-0639

Po-Chun, Lin*
Department of Electrical Engineering and Computer Science
National Taipei University of Technology
Taipei, Taiwan
linpochun13@gmail.com

Yi Lin
Department of Business Administration
Takming University of Science and Technology
Taipei, Taiwan
linyi@takming.edu.tw

Yuh-Shyan Hwang
Department of Electronic Engineering
National Taipei University of Technology
Taipei, Taiwan
yshwang@ntut.edu.tw

***Abstract*—This study proposes a solution that combines drones and the YOLOv8 algorithm to solve the problem of difficult object detection owing to irregular and overlapping leaves of pineapple angiosperms. The leaves of pineapple angiosperms are arranged in spirals, lamellas and naturally spread out in all directions to increasing the area for photosynthesis. But object detection technology faces the challenge of overlapping distinctions during labeling for training. To solve this problem, this study uses drones to capture high-resolution field images and uses the YOLOv8 algorithm to perform image segmentation and object detection. It combines data argumentation and segmentation technologies to effectively improve the model's ability to detect overlapping objects.**

Experimental result show that the solution proposed in this study can quickly and precisely estimate the quantity of pineapple angiosperms, the detection accuracy can reach more than 95% in overlapping situations. The solution provides efficient and reliable technical framework for automated data management in precision agriculture. In addition to overcoming the time-consuming and labor-intensive shortcomings of traditional manual estimation methods, the solution also significantly improves the efficiency of agricultural data collection and processing.

Keywords—aerial photography, irregular, overlapping, image segmentation, object detection

I. Introduction

With the rapid development of precision agriculture, the importance of science and technology application in agricultural management has become increasingly prominent. Precision agriculture uses advanced data collection and analysis tools to allow agricultural practitioners to conduct precise management of crops, thereby improving production efficiency and reducing labor costs. Outstanding among them, drones have become one of the tools of precision agriculture due to the advantages in aerial photography technology.

At present, the application of aerial images provided by drones in precision agriculture covers crop monitoring, pest and disease detection, soil management, land mapping, and yield prediction. Aerial images provided by drones, combined with the development of artificial intelligence, make automated crop identification and data analysis possible.

However, precision agriculture still faces challenges in crop statistics. For example, as an important economic crop, the leaves of pineapple angiosperms are arranged in irregular spiral and laminated shapes, and these leaves will naturally expand outward and overlap. This causes object detection technology to be easily unrecognizable owing to occlusion when performing image labeling and quantity estimating, thereby affecting the data accuracy and management efficiency of farmland.

Fig. 1 to 3 present photos of the three main growth stages of pineapple. It is clearly visible in the photo in Fig. 1 that the leaves of pineapple are the main organs for photosynthesis. Long leaves can increase the area, thereby absorbing more sunlight, providing enough energy for the plant, and serving as a container for storing nutrients and water, so this is a natural strategy used by plants to increase their biomass. [2-4] illustrate the biological characteristics of pineapple.

Fig. 1-3. From left to right. The pineapple at vegetative stage (angiosperms), flowering stage and fruiting stage.

II. OVERVIEW OF YOLOV8

This study combines the image collection capabilities of drones with the image segmentation technology of the YOLOv8[5] algorithm to propose an automated method for estimating the quantity of pineapple angiosperms. Drones can quickly and comprehensively capture open field images, and YOLOv8 has optimized detection capabilities for object overlap. Through this technical framework, this study expected to effectively overcome the difficulty in identification of pineapple angiosperms because overlapping leaves and improve the accuracy and practicality of agricultural management data.

YOLO (You Only Look Once)[1] is an object detection model based on deep learning. It is famous for its speed in completing detection in a single forward propagation. YOLO has undergone many version improvements since its launch, and each iteration has improved detection speed and accuracy. In this study, YOLOv8 was selected as the solution for image segmentation and object detection based on the following considerations:

- Excellent multi-scale detection capabilities: YOLOv8 combines multi-level features by Neural Architecture Search-Feature Pyramid Network (NAS-FPN)[10], can effectively detect objects of different sizes. A Feature Pyramid Network (FPN)[11] for object detection, which designs the optimal feature fusion structure through Neural Architecture Search (NAS)[10] automatically. The purpose is to enhance the fusion ability of multi-level features so that the model can be more accurate when processing objects of different sizes, especially the detection effect of small objects.

- Advantages of handling overlapping objects: Combined with the Focal Loss function, shown in (1) and (2).

$$\mathcal{L}_{FL} = -a_t(1-P_t)^{\gamma}\log(P_t) \tag{1}$$

$$P_t = \begin{cases} P, & ground\ truth = 1 \\ 1-P, & ground\ truth = 0 \end{cases} \tag{2}$$

  (2) indicates the predicted probability of the target category. Where a_t is a balance coefficient used to adjust the proportion of positive samples and negative samples, γ is the adjustment factor used to control the loss weight of difficult samples. YOLOv8 has higher detection accuracy when dealing with overlapping objects and effectively reducing misjudgments.

- Efficient data argumentation technology: The use of Mosaic and MixUp two powerful data argumentation technologies increases the diversity of the data set and improves the generalization ability of the model, which is especially suitable for images under different lighting conditions and shooting angles.

  Mosaic data argumentation method will crop 4 randomly selected images and stitch them into a 2x2 mosaic layout, and then re-label the object's bounding box accordingly. Thus, in addition to expanding the perspective and diversity, also enhances the adaptability of background changes, part of the image is scaled so that small objects appear more frequently, which improves the model's ability to detect small objects.

  MixUp data argumentation method will randomly select two images, mix the two images at the pixel level according to a mixing ratio usually between 0 and 1, and then mix the bounding boxes or classification labels according to the same ratio. Generates an image with mixed properties. This can increase sample diversity, balance class distribution, and reduce the possibility of the model remembering specific sample characteristics, because these generated samples have intermediate characteristics, thereby alleviating the overfitting problem.

- Accurate segmentation and detection performance: YOLOv8 can simultaneously generate object bounding boxes and fine segmentation masks during the detection process. With this, model can calculate the object's area more accurately, thereby achieving accurate quantity estimation.

- Balance speed and accuracy: YOLOv8 achieves a good balance between speed and accuracy, and can handle large-scale image data at a faster processing speed while ensuring high detection accuracy.

III. TECHNICAL ANALYSIS

A. Data augmentation

Data augmentation[6, 7] is a technology that can generates new data to expand the training set by performing various transformations on existing data. The main purpose is to improve the generalization ability of the model and reduce overfitting. Commonly used techniques include:

1) Geometric transformation: Many methods likes rotation, translation, scaling, flipping, etc., to change the spatial structure of the image.

2) Color transformation: Adjust image brightness, contrast, saturation, or perform color dithering operations to change the color attributes of the image.

3) Noise addition: Add random noise to the image for increase the model's robustness to noise.

4) Crop & Fill: Randomly crop a portion of the image, or add a border around the image to change the image's field of view.

By increasing the diversity of data, these technologies let the model to learn richer features, thereby improving performance on unknown data.

B. Segmentation mask

Segmentation mask[8, 9] refers to assigning a label to each pixel in the image to distinguish different objects or regions, which is particularly important in Semantic Segmentation and Instance Segmentation. The main principles include:

1) Feature extraction and convolution operations: In the segmentation task, the convolutional neural network (CNN) extracts the spatial features of the image through multi-layer convolution. Assume that the input image is $X \in R^{H\times W\times C}$, where H is the height, W is the width, and C is the number of channels. When passing through the convolution layer, the calculation formula of the feature map shown as (3).

$$F_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{c=0}^{C-1} K_{m,n,c,k} \cdot X_{i+m,j+n,c} + b_k \quad (3)$$

among them, $K \in R^{M \times N \times C \times K}$ represents convolution kernel, M and N are the convolution kernel sizes, b_k is the offset, $F_{i,j,k}$ is the output value of the k-th channel at the (i, j)-th position of the feature map. Through layer-by-layer convolution, CNN extracts progressively higher-level features in the image, first from edges to structure and then to the overall shape of the object.

2) Up-sampling and decoding: To generate segmentation masks with the same size as the original image, an up-sampling operation is used to enlarge the feature map. Common methods include bilinear interpolation and transposed convolution. The calculation formula of transposed convolution is as follows (4).

$$Y_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} W_{m,n,k} \cdot F_{i-m,j-n,k} \quad (4)$$

among them, Y is the up-sampled feature map and W is the transposed convolution kernel. This operation restores the compressed feature map to the original size of the input image, facilitating pixel-level classification.

3) Activation functions and classification: Segmentation operations usually use Softmax() activation function for pixel classification. The output for each pixel position (i, j) is calculated as (5).

$$P_{i,j,c} = \frac{exp(z_{i,j,c})}{\sum_{c=1}^{C} exp(z_{i,j,c})} \quad (5)$$

among them, $Z_{i,j,c}$ is the score that the pixel belongs to category c, and $P_{i,j,c}$ represents the probability that the pixel belongs to category c. Therefore, the final classification result of each pixel is shown in (6).

$$\widehat{c_{i,j}} = arg\left(\max_{c}(P_{i,j,c})\right) \quad (6)$$

4) Loss function: Commonly used loss functions include cross-entropy loss and Dice loss, which are used to measure the difference between the predicted mask $\hat{Y}$ and the real mask Y. Allows the model to learn the category to each pixel belongs, thereby generating accurate segmentation masks. The cross-entropy loss is shown in (7) and the Dice loss is shown in (8); in terms of selection, the cross-entropy loss is suitable for the category imbalance problem, while the Dice loss emphasizes the overall shape consistency of the segmented area.

$$\mathcal{L}_{CE} = -\frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{c=1}^{C} Y_{i,j,c} \, log\left(\widehat{Y_{i,j,c}}\right) \quad (7)$$

$$\mathcal{L}_{Dice} = 1 - \frac{2 \times \sum_{i=1}^{H} \sum_{j=1}^{W} Y_{i,j} \widehat{Y_{i,j}}}{\sum_{i=1}^{H} \sum_{j=1}^{W} Y_{i,j}^2 + \sum_{i=1}^{H} \sum_{j=1}^{W} \widehat{Y_{i,j}^2}} \quad (8)$$

In short, image segmentation is the focus of this study, which is used to divide images into regions or objects with semantic meanings, so this study is specially summarized these methods. Depending on the application and method, image segmentation is usually divided into the following main types:

- Semantic Segmentation: Classify each pixel in the image as belonging to a specific category, such as foreground or background. The characteristic is that all pixels belonging to the same category are marked as a group, and different instances cannot be distinguished.
- Instance Segmentation: Not only can identify the category of objects, but can also distinguish different instances in the same category. The characteristics are: compared with semantic segmentation, this method increases the ability to distinguish object instances, so this method is more suitable for object detection in dense scenes.
- Panoptic Segmentation: This method combines semantic segmentation with instance segmentation to assign a category label and object instance identifier to each pixel. Features: can process background and foreground objects at the same time, and is applied to tasks that require a comprehensive understanding of the scene.
- Edge Segmentation: Accurately locate the boundaries of objects in the image, emphasizing the extraction of object contours. Features: This method is used in scenarios that require precise boundary information, such as medical image processing and image annotation.

In this study, the following two image segmentation methods were mainly used to address the challenge of counting pineapple plants:

- ◆ Instance Segmentation: Since pineapple fields are densely populated and leaves often overlap, it is impossible to distinguish each one using only semantic segmentation. This study uses the segmentation head of YOLOv8 to generate instance-level masks, which can accurately identify each object and effectively distinguish overlapping parts. The advantage of this is that instance segmentation can not only identify the range of each object, but also accurately distinguish multiple pineapple angiosperms in overlapping areas, which helps to improve the accuracy of subsequent quantity estimation.
- ◆ Edge Segmentation: To further improve the accuracy, this study combined the boundary information of the mask, and used contour analysis technology to correct the segmentation boundaries, especially in the area where leaves overlap. This method can accurately delineate the boundaries of pineapple angiosperms and avoid misidentifying multiple angiosperms as one. The advantage is that precise boundary segmentation makes the area calculation of each object more accurate, further improving the quantitative estimation capability of the model.

Therefore, after integrating the above image segmentation methods, this study emphasizes three points:

i. *Combine instance segmentation with boundary segmentation to deal with leaf overlap:* This study innovatively combines instance segmentation with boundary segmentation, which can not only accurately identify each object, but also effectively deal with the false detection problem caused by leaf overlap, which is very important in densely planted crops. It has significant advantages in image analysis.
ii. *Combining segmentation mask with center point separation to improve the accuracy of quantity estimation:* After generating instance masks, this study further adopted a center point detection and separation strategy to split masks with too large an area. This solves the problem of mistakenly merging multiple pineapple angiosperms in densely populated areas.
iii. *Data enhancement strategies that adapt to a variety of scenarios:* Combined with Mosaic and MixUp data enhancement technologies, training samples with diverse scenes and lighting conditions are generated, maintaining the model has high accuracy in different environments.

IV. Experiment and Result

The goal is to estimate the quantity of pineapple angiosperms with overlapping leaves and dense plantings, challenges that can make existing methods inaccurate, including manual estimation. Experiment mainly relies on high-resolution images captured by drones, and uses YOLOv8 algorithm to perform the following steps:

A. Image capture and annotation

The drone shoots vertically over the open field to obtain images covering the entire field, which are then annotated using specificized software. The range of each object is marked with a frame or mask as the "correct answer" for model training, like shown in Fig. 4.

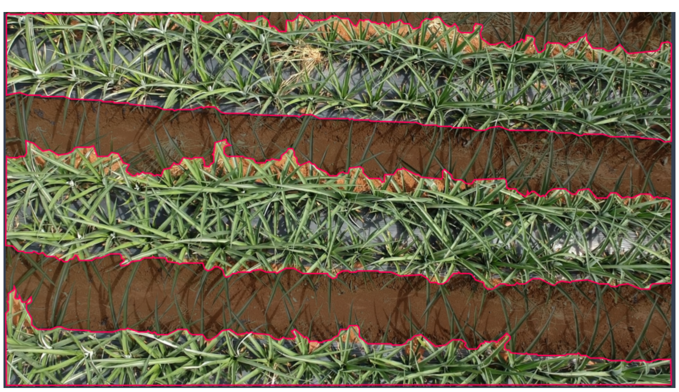

Fig. 4. Manually annotate objects in captured images and know the actual quantity.

B. Training YOLOv8 model

Through a large number of annotated images, the model learns how to identify the position and shape of objects in the image. During the training process, many data argumentation techniques are added to enable the model to adapt the changes in various real-life environments.

C. Solve the problem of overlapping leaves

1) Use segmentation masks: YOLOv8 not only generates bounding boxes, but also generates segmentation masks for each object (equivalent to the "accurate silhouette" of each pineapple). Through these masks, the actual area of each object can be analyzed to avoid overlapping leaves being mistaken for the same plant.

2) Detect duplicate frames: For overlapping frames, use the Non-Maximum Suppression (NMS)[12] algorithm to filter out duplicate or inaccurate frames, then retain the most credible detection results.

The NMS algorithm is a post-processing algorithm used in object detection technology to eliminate redundant bounding boxes. It ensures that for each detected object, only the box with the highest confidence score is retained while suppressing others with significant overlap.

The process will be sorting all boxes by their confidence scores. Then, the box with the highest score is selected as the reference, and the Intersection over Union (IoU) is calculated between this box and the others, the equation shown in (9). Boxes with IoU exceeding a predefined threshold are discarded. This process repeats until no boxes remain, resulting in a final set of non-overlapping, high-confidence detections.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{9}$$

D. Quantity estimate

In general, the experiments in this study used segmentation mask refinement to allow the model to not only know "there is an object here" but also "the specific outline of the object". Even if the leaves overlap or the plants are densely packed, each object can be effectively separated by center point separation and area thresholding. Through data argumentation and post-processing, the model can still maintain high accuracy in complex scenes.

In order to ensure image quality and data reliability, the drone's sailing altitude, speed, lens viewing angle, as well as the restrictions and configuration of environmental conditions have been carefully designed and adjusted:

- The flight height is set to 2.5 meters; This choice is based on object size and image clarity. Because pineapple angiosperms grow on the ground, making them visually relatively short, flying the drone too high will reduce the resolution and details may not be clearly presented. If the drone flies too low, the field of view may not cover enough and the shooting time may be extended. Selecting this flight height can balance image clarity and coverage, ensuring that each frame of image fully captured the details of the pineapple angiosperms and leaves.

- The flying speed is fixed and set to 0.3m/second; The flying speed of the drone is fixed and set to avoid image blur caused by too fast speed, especially when shooting in low light or high wind speed. Stable speed ensures clear images while giving the camera enough time to capture frame-by-frame.
- The lens adopts a vertical overhead shot (90 degrees); This can minimize perspective distortion caused by angles and keep the size and shape of objects consistent in the image that facilitates subsequent model training.
- Camera: Use a camera with at least 4K resolution.

Besides, the limitations and responses to environment:

- The best shooting time is two hours after sunrise and before sunset; During this period, the light is even and soft to avoid shadows caused by strong light at noon and interfering with the identification of the boundaries of objects.
- Limit the wind speed to less than 5m/second; High wind speed will affect the stability of the drone and may cause image blur or flight path deviation.
- The site is set as a flat, barrier-free open field; Make sure the crops are evenly distributed and avoid undulating terrain or obstacles that may affect the integrity of the captured images.
- Set up geofencing in the flight plan; Ensure the drone always flies within the intended area.
- The flight path planning is set to a serpentine path; Ensure that the drone can cover the entire open field with the shortest possible flight, and that there is at least 30% to 50% overlap between adjacent images. Such a setting can be used for image stitching in the later stage to form a complete field coverage map.

E. Result

Fig. 5 shows the confusion matrix of the model trained with YOLOv8 in this study. The effect of object detection is shown in Fig. 6. During inference, aerial images are input and the quantity is estimated by the calculated area quickly, as shown in Fig. 7.

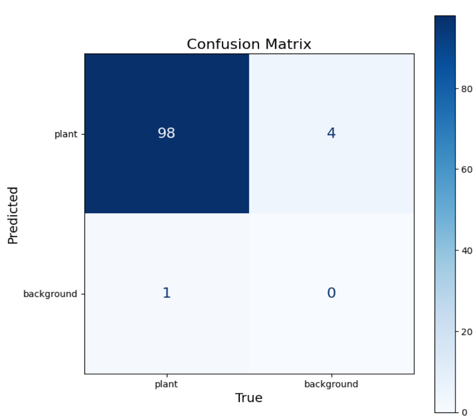

Fig. 5. Confusion matrix of the model in this study.

Fig. 6. After object detection, the area of the object is estimated.

```
plant confidence:0.8643535375595093 Area:105927.5
plant confidence:0.768988847732544 Area:68228.5
Total area: 174156.0 Estimated quantity:42 Time spent:97ms
plant confidence:0.8783074021339417 Area:106365.0
plant confidence:0.8244731426239014 Area:64300.0
Total area: 170665.0 Estimated quantity:41 Time spent:67ms
plant confidence:0.8563641905784607 Area:105611.0
plant confidence:0.7867398262023926 Area:61202.0
Total area: 166813.0 Estimated quantity:40 Time spent:86ms
```

Fig. 7. Estimate the quantity based on the calculated area.

In terms of accuracy, when the deviation of quantity is set within the range of ±0, the accuracy reaches 12.50%. When the deviation is set within the range of ±1, the accuracy reaches 54.17%. When the deviation is set within the range of ±2, The accuracy rate reaches 93.75%. When the deviation is set within the range of ±3, the accuracy rate reaches 100%. The result shows that the model has certain accuracy over a wide range.

Compared with [4], the results of this study require simpler resources than [4], if multiple drones are used to be responsible for planning area; meantime, the results of this study do not require geographic information processing capabilities for images obtained from aerial photography, due to the software can be relied upon to sum up individual quantities.

V. Conclusion

This study breaks through traditional obstacles and proposes a segmentation strategy based on center point separation and area threshold based on the characteristics of overlapping leaves of pineapple angiosperms and uneven planting density in the field, which improves the accuracy of detection. By realizing the application of precision agriculture and providing efficient and practical technical solutions, this study can significantly reduce the labor and time costs in agricultural management by fully automating image processing and quantity statistics, laying the foundation for the popularization of precision agriculture.

By adopting YOLOv8 to process multi-level technology applications and using data argumentation methods such as Mosaic and MixUp, the model's adaptability to diverse scenarios is enhanced. Meantime, the accuracy of small object detection is improved through technologies such as NAS-FPN and Focal Loss. The results of this study can also be extended to multi-species applications, such as estimating the number of densely planted fruit trees such as bananas and apples. In the future, the impact of different crop characteristics on the results of this study will also be explored.

VI. Future work

Improving accuracy is a key goal for the future development of this study, which can further reduce false detections (False Positive) and missed detections (False Negative), and improve the model's ability to identify objects and their characteristics.

1) Introduce multi-modal data; Combine RGB images and multi-spectral images (such as infrared or near-infrared), so that the model can simultaneously use visible light and crop physiological characteristics (such as health status or moisture content) to detect. In this way, crops and backgrounds can be effectively distinguished for multi-modal data, especially in situations with low lighting condition or complex backgrounds.

2) Optimize segmentation mask generation; Use higher-order semantic segmentation models (such as DeepLabv3+ or Segment Anything Model) to replace the segmentation function of YOLOv8. These models segment object boundaries more accurately, and improve the detection rate of smaller objects.

3) Focus on the main areas (Attention Mechanism); The attention mechanism is embedded in YOLOv8, allowing the model to focus more on the characteristic areas of objects, reducing interference with the background, reducing false detections, and further improving the stability of the model in complex scenes.

4) Enhance the diversity of the data set; Expand the data set to cover more images with different lighting conditions, angles and environments, and introduce Synthetic Data Augmentation (SDA) technique to improve the model's generalization ability to unknown environments and reduce the probability of missed detections.

Acknowledgment

This study benefited from Link Join Agriculture in Taiwan, the field and images provided by the Chiayi County Da Mao Fruit and Vegetable Production Cooperative.

References

[1] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[2] V.M. Maia, F.S. Oliveira, R.F. Pegoraro, B.A.M. Souza, L.B. Ferreira, I. Aspiazu, “Vegetative growth stages of irrigated ‘Pérola’ pineapple”, Acta Horticulturae, 2016.

[3] J.A. Souza, E.J. Oliveira, M.S. Lima, E.M. Souza, “Natural flowering in pineapple: inhibition by growth regulators”, Fruits, 2001.

[4] Jennifer Hobbs, Prajwal Prakash, Robert Paull, Harutyun Hovhannisyan, Bernard Markowicz, Greg Rose, “Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation”, Frontiers in Plant Science, 2020.

[5] Glenn Jocher et al. (2023). YOLOv8. GitHub repository.

[6] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60.

[7] Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T. Y., Shlens, J., & Le, Q. V. (2020). Learning data augmentation strategies for object detection. arXiv preprint arXiv:1906.11172.

[8] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

[9] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

[10] Golnaz Ghiasi, Tsung-Yi Lin, Ruoming Pang, Quoc V. Le, “NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[11] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie, “Feature Pyramid Networks for Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[12] Philipp Fischer, Alexey Dosovitskiy, Thomas Brox, “Learning Non-Maximum Suppression”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[13] Po-Chun Lin, Yuh-Shyan, Hwang, Chien-Hung Lai, “Research on Estimating the Number of Pineapple Plants Using Image Segmentation”, National Taipei University of Technology, College of Electrical Engineering and Computer Science, master, 2025.