國立台灣大學生農學院生物機電工程學系

學士論文

Department of Biomechatronics Engineering

College of Bioresources and Agriculture

National Taiwan University

Bachelor Thesis

應用少量物件辨識方法於茶葉病蟲害辨識模型之建立

Implementation of Few-Shot Object Detection Methods on Tea Diseases Identification

林篆澤

Chuan-Che Lin

指導教授：陳世芳 博士

Advisor: Shih-Fang Chen, Ph.D.

中華民國 111 年 7 月

July 2022

ABSTRACT

Plant disease identification requires a large amount of labeled data to build a robust model, and it is exceedingly time-consuming for plants with various types of diseases. Applying few-shot object detection (FSOD) methods generalize the flexibility of the model to incorporate novel-added classes. The previous study of our group build a dataset of a dataset of 4295 images containing 10 species of tea diseases and harming insects. In this study, 564 images of four new classes were added to form a dataset of 4859 images. This study applies two FSOD methods of the two-stage fine-tuning approach (TFA) and the few-shot object detection via contrastive proposal encoding (FSCE) on tea disease identification and compares their performance with six numbers of shots and two data splits. The results showed that the suggested FSOD method FSCE could efficiently expand the class number for identification without significantly sacrificing performance. It provides a feasible framework for novel class expansion with a lower cost of labor and time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

Tea disease and pest specialist is a scarce profession, and it is hard to fulfill the massive demand from tea farmers to identify tea plant diseases and harming insects. An automatic identification approach is needed to provide real-time responses. Previously, our research group developed a faster region-based convolution neural network (FRCNN) model to identify 12 classes of tea diseases (Chen et al. 2021). However, dozens of species of tea diseases and harming insects occur in Taiwan, and it is essential to include more of them to provide a comprehensive identification service. It is time consuming to collect the huge number of the desired disease images in a short time. Therefore, how to expand novel classes efficiently with limited number of images and acquire acceptable performance at the same time is the bottleneck of the current model. In this study, four common species (sunburn, tea leaf roller, Tetranychoidea, *Jacobiasca formosana*) were collected as the novel classes to be added in the dataset. Among them, *Jacobiasca formosana* is the crucial element to produce the unique taste of Oriental Beauty Oolong Tea that was widely acclaimed and made its way to the world commercial market. Few-shot object detection (FSOD) methods was selected to demonstrate the potential to provide a path towards efficient class expansion.

CHAPTER 2. LITERATURE REVIEW

2.1 Plant Disease Identification

Applying deep learning approaches to plant disease identification requires much labeled data. Open datasets are commonly used in many studies. The PlantVillage dataset (Hughes and Salathe, 2015) contains 54306 images of 26 different classes for 14 crop species, but they were used on classification tasks only due to a lack of annotations. In addition, most of the images were taken under laboratory conditions, hence the trained models might not perform well in complicated field situations. The CropDeep dataset (Zheng et al., 2015) contains 31,147 images with over 49,000 annotated instances of 31 different classes for 19 crop species, and the images were collected under field conditions. Although the number of images is abundant, studies developed with open datasets recognize only major diseases that were included (Brahimi et el., 2017; Ferentinos et el.,2018). However, the diseases that emerged in the field were far beyond the limited categories. Studies that collected their own dataset also encounter similar issues that it lacks some diseases (Fuentes et el., 2017; Hu et el.,2019; Liu et al., 2020). To offer a more applicable identification service, how to expand novel classes efficiently is essential to provide a comprehensive guidance for the farmer. Few-shot learning methods demonstrated the potential to generalize from few samples; therefore, it increase the flexibility of the model to incorporate novel added-sclasses.

2.2 Few-Shot Object Detection

For basic image classification tasks, many have attempted to adopt the concept of meta-learning to improve the performances of novel classes (Finn et al., 2017; Hariharan and Girshick, 2017; Gidaris and Komodakis, 2018). Using simulated few-shot tasks during training helps the model to learn with fewer data from the novel class. Compared to image classification, object detection requires the model to produce localization information besides the object types. It significantly increases the complexity of a few-shot object detection task. To tackle down the under-explored field, some studies followed the meta-learning strategy by adding meta learners to the existing object detection structures (Kang et al., 2019; Yan et al., 2019; Wang et al., 2019). Wang et al. (2020) proposed the two-stage fine-tuning approach (TFA) to focus on the training scheme instead of the model architecture, and it outperforms the earlier state-of-the-art meta-learning-based methods. Based on TFA, the few-shot object detection via contrastive proposal encoding (FSCE) aims to improve the proposal-wise feature selection and further elevates the performance (Sun et al., 2021).

CHAPTER 3. MATERIALS AND METHODS

3.1 Image Collection

From the previous-stage study in our research group, a dataset of 4295 images containing 10 species of tea diseases and harming insects were built (Chen et al. 2021). Four new species were collected from 2020 to 2022 in northern and middle Taiwan (Fig. 3.1). There were 564 images added in the dataset as the novel classes. The merged dataset includes 4859 images of tea leaf lesions and were categorized into 17 classes. Five of them are caused by diseases, and 11 of them are caused by harming insects. Cross sections of stick and leaves with raindrops or dew are defined as one class "other" to minimize the misclassification between healthy leaves. If there are distinct patterns in one specie, it has to be separated into two subclasses. For example, tea mosquito bugs were divided into early and late stages depends on severity; small tea tortrix were divided into bitten wound and rolled leaf due to different lesion causes (Table 3.1).

Figure 3.1 Four newly collected diseases.

Table 3.1 Numbers of images for each class.
(Rows were highlighted in pale green representing the newly collected diseases.)

Class id	Pest species	Stage/pattern	Images	
			Training	***Testing***
brownblight	brown blight	-	1203	281
algal	algal leaf spot	-	260	65
blister	blister blight	-	235	60
sunburn	sunburn	-	63	11
fungi_early	fungi disease	early stage	247	65
roller	oriental tea tortrix	-	214	51
moth	small tea tortrix	bitten wound	148	13
tortrix	small tea tortrix	rolled leaf	361	103
flushworm	tea flush worm	-	103	27
caloptilia	tea leaf roller	-	154	48
mosquito_early	tea mosquito bug	early stage	269	78
mosquito_late	tea mosquito bug	late stage	438	116
miner	leaf miner	-	283	71
thrips	tea thrips	-	311	77
tetrany	Tetranychoidea	-	165	43
formosa	*Jacobiasca formosana*	-	62	23
other	stick cross section, rain drops and sun scald		1023	255
Total			**3887**	**972**

3.2 Model Architecture

3.2.1 Faster Region-Based Convolution Neural Network

A faster region-based convolution neural network (FRCNN) model is divided into two parts—feature extraction and classification (Ren et al., 2015). Feature extraction includes a backbone CNN to extract convolutional feature maps, a region proposal network (RPN) to generate proposal boxes, a region of interest (RoI) pooling layer to obtain fixed-length proposal feature maps, and a RoI feature extractor to extract proposal-wise features. Classification includes a box regressor to provide more accurate bounding boxes, and a box classifier to categorize each bounding boxes (Fig. 3.2). Similar to other deep learning models, FRCNN achieve lower average precision (AP) on novel classes that contain few images.

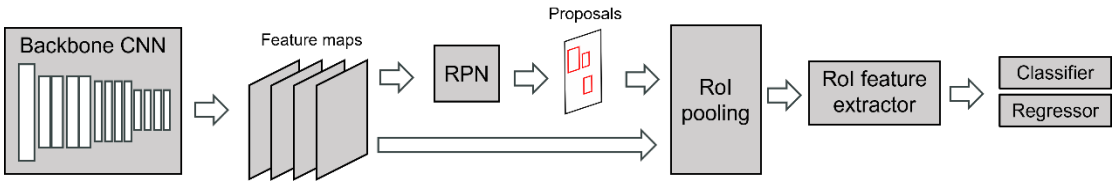

Figure 3.2 Faster region-based convolution neural network (FRCNN).

3.2.2 Two-Stage Fine-Tuning Approach

Based on the FRCNN, the two-stage fine-tuning approach (TFA) was proposed to improve the few-shot performances. Since the feature extraction is class-agnostic, the information can be shared by all classes. If there is a robust feature extractor, merely the class-specific box regressor and box classifier require fine-tuning for novel classes. The TFA divides the training into two stages. In the first stage, the model is trained by the base classes that contain relatively abundant images (about 300 per class). In the second stage, the parameters in the feature extractor are frozen, and only the box

regressor and box classifier are fine-tuned by a small balanced dataset contains the same number of images (less than 60 per class) for both base and novel classes (Fig. 3.3). It prevents overfitting caused by the limited number of novel data and achieves better performances. However, the RoI feature extractor only contains semantic information from base classes, and it would be transferred to the novel classes directly.

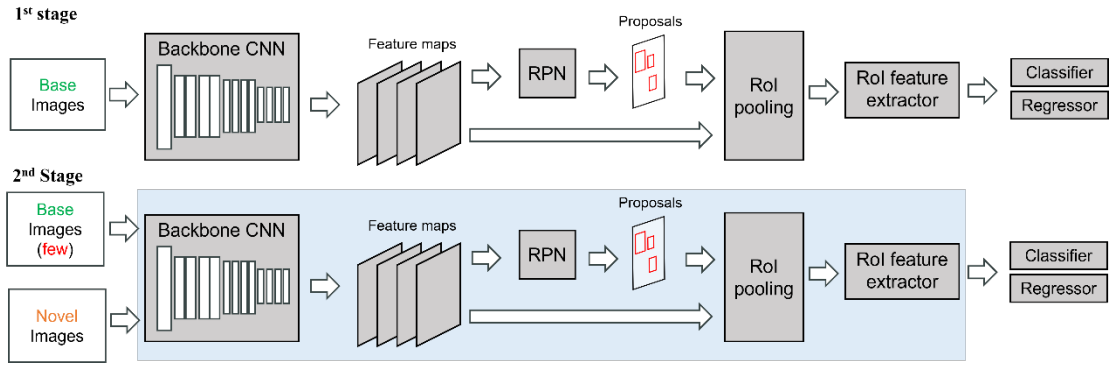

Figure 3.3 Two-stage fine-tuning approach (TFA).
(Processes were highlighted in blue area were the fixed structures.)

3.2.3 Few-Shot Object Detection via Contrastive Proposal Encoding

Based on the TFA, the few-shot object detection via contrastive proposal encoding (FSCE) was further proposed to produce more distinguishable RoI features. The first stage training of FSCE is the same as the TFA. In the second stage, only the backbone and RoI pooling layer are frozen, and it enables the model to learn the semantic information of the novel classes while avoiding the risk of overfitting. In the meantime, a one-layer multi-layer-perceptron (MLP) named contrastive head is added to encode the RoI features into contrastive features, and it guides the model to establish robust feature representations even with limited data (Fig. 3.4). The contrastive head adopts cosine similarity to measure the distance between the contrastive features. Tighter clusters and larger distances between clusters in the cosine projected hypersphere increase the generalizability of the detection model.

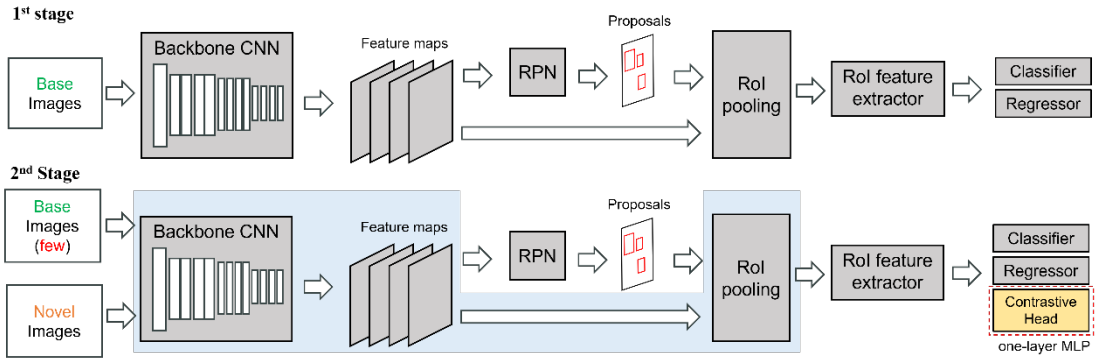

Figure 3.4 Few-shot object detection via contrastive proposal encoding (FSCE). (The blue block indicates the fixed structures)

Contrastive proposal encoding loss (L_{CPE}) was proposed to enhance the effectiveness of the contrastive head (Eq 3.1). In a mini-batch of N RoI features, L_{z_i} is the loss for each of them (Eq 3.2).

$$L_{CPE} = \frac{1}{N}\sum_{i=1}^{N} f(u_i) \cdot L_{z_i} \tag{3.1}$$

$$L_{z_i} = \frac{-1}{N_{y_i} - 1} \sum_{j=1, j \neq 1}^{N} \mathbb{I}\{y_i = y_j\} \cdot \log \frac{\exp(\tilde{z}_i \cdot \tilde{z}_j / \tau)}{\sum_{k=1}^{N} \mathbb{I}_{k \neq i} \cdot \exp(\tilde{z}_i \cdot \tilde{z}_k / \tau)} \tag{3.2}$$

The u_i is the Intersection over Union (IOU) score; z_i is the MLP-encoded RoI feature for i-th proposal; $\tilde{z}_i = \frac{z_i}{\|z_i\|}$ denotes the normalized feature; y_i is the ground truth label, N_{y_i} is the number of proposals with same label as y_i; τ is the hyper-parameter temperature from InfoNCE (Oord et al., 2018). Optimizing the loss function increases similarity between proposals with the same label and distances between proposals with different labels in the cosine projected space.

The FRCNN contains a L_{rpn}, a L_{cls}, and a L_{reg} for foreground detection, box classification, and box regression, and the L_{CPE} is added to the basic Faster R-CNN loss in the second stage of training. The scaling factor λ is set to 0.5 to balance the scale of each loss (Eq 3.3).

$$\mathbb{L} = L_{rpn} + L_{cls} + L_{reg} + \lambda L_{CPE} \tag{3.3}$$

3.2.4 Modified Faster Region-Based Convolution Neural Network

To evaluate the model performances under the same criterion, we proposed a modified FRCNN that follow the training scheme of TFA but without freezing any model parameters in the second stage of training (Fig. 3.5).

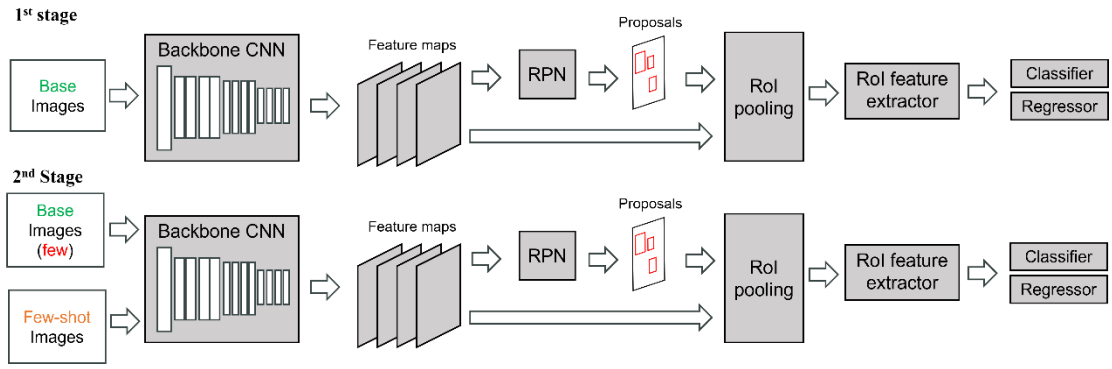

Figure 3.5 The modified FRCNN under TFA training scheme.

3.3 Experimental Construction

In summary, three types of models are trained based on the structure of FRCNN, but different model parameters are fine-tuned during the second stage of the training process (Table 3.2).

Table 3.2 Fine-tuned model structures in the second stage.

Model	Backbone	RPN	RoI feature extractor	Detectors
Modified FRCNN	✓	✓	✓	✓
TFA				✓
FSCE		✓	✓	✓

For all the models in our experiment, FRCNN is chosen as the base detector and ResNet-101 (He et al., 2016) with a feature pyramid network (Lin et al., 2017) is the backbone. A global testing set was saved for the evaluation of all the models. To make sure the result is reproducible under different data distribution, two splits of base and novel classes are tested. The novel classes of the first split are the four newly added classes to better simulate the situation of including new classes in a system. On the other hand, the novel classes of the second split are chosen from the previous 12 categories. By considering their performance in the original FRCNN, four chosen novel classes fall in a various range to provide standards for future estimation (Table 3.3).

Table 3.3 Novel classes for each data split

Split	Novel Classes
1	formosa, caloptilia, sunburn, tetrany
2	tortrix, mosquito_early, roller, flushworm

CHAPTER 4. RESULTS AND DISCUSSIONS

4.1 Model performances

Before the implementation of FSOD methods, a FRCNN model was developed on the full training set consists of 3887 images, and it achieved a mean average precision (mAP) of 78.04%. To evaluate the overall performances of FSOD models concisely, the mAP of the base classes (base AP) and the novel classes (novel AP) are introduced. Six number of shots (1, 3, 5, 10, 30, 60) were tested in our experiment, and both base AP and novel AP of all three models improve as more shots were given to the model in the second stage of training (Fig. 4.1). Since the models with 60 shots gave a reasonable data number and demonstrated a better performance on the first split, they are selected for the following discussions (Table 4.1). The modified FRCNN model achieved the best novel AP, but the 3% dropping of the base AP from the first stage (75.7%) to the second stage (72.7%) revealed the potential overfitting. By contrast, the TFA and FSCE counter overfitting by freezing model parameters; thus, the base AP of FSCE encounters a 0.2% decrease only. In the meantime, FSCE achieves a better novel AP (67.5%) than TFA (30.3%). The degradation of the base AP is possible to be recursively accumulated in each round of class expansion. Therefore, the FSCE model with less base AP drop provides a better option for implementing FSOD on tea disease identification. Besides, the model performances on the second split also demonstrate the similar tendency (Table 4.2).

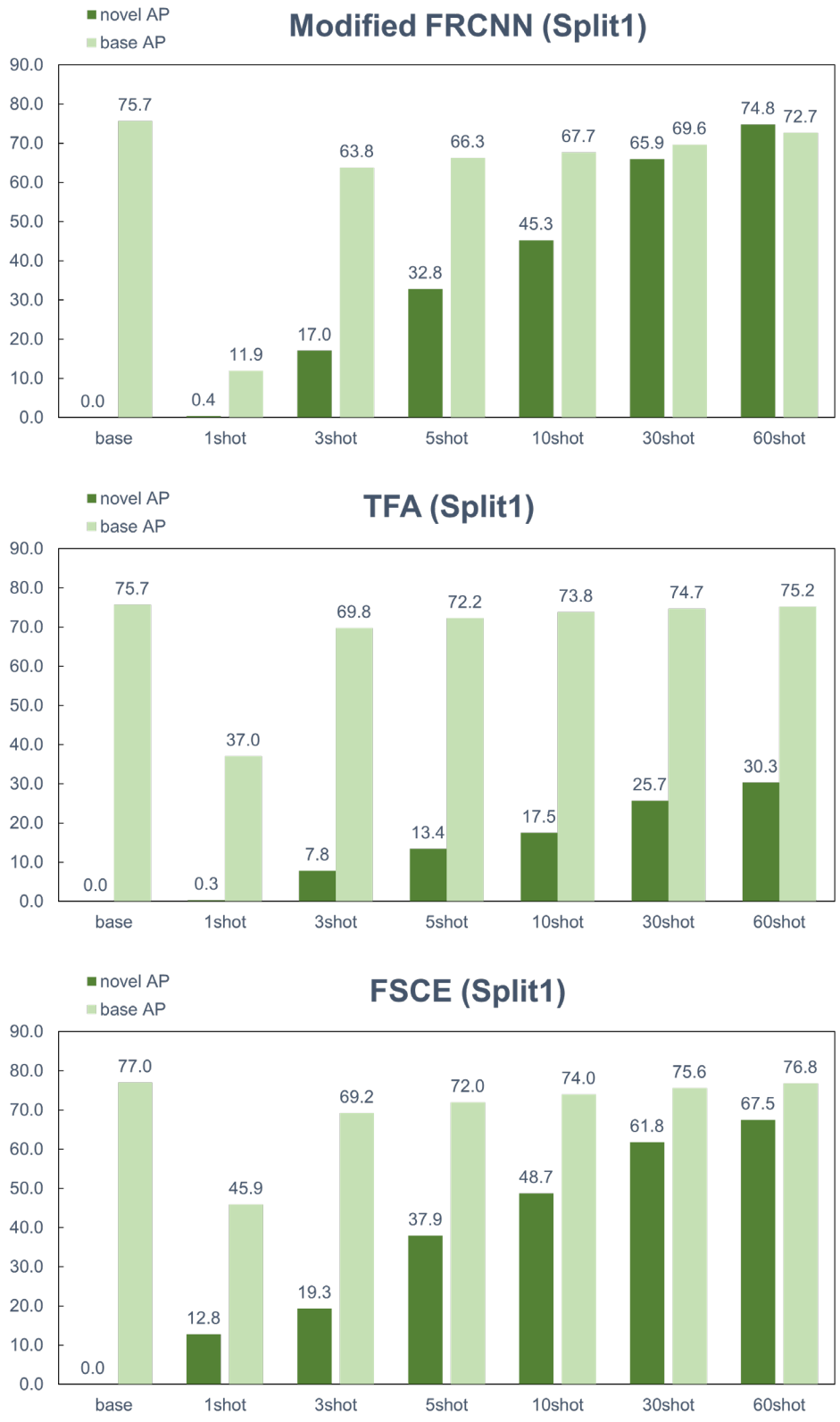

Figure 4.1 Comparisons of model performances with different shots on the first split.

Table 4.1 Model performances on the first split.

Model	AP	1st stage	2nd Stage					
			Shots					
			1	*3*	*5*	*10*	*30*	*60*
Modified FRCNN	base	75.7	11.9	63.8	66.3	67.7	69.6	*72.7*
	novel	0.0	0.4	17.0	32.8	45.3	65.9	*74.8*
TFA	base	75.7	37.0	69.8	72.2	73.8	74.7	*75.2*
	novel	0.0	0.3	7.8	13.4	17.5	25.7	*30.3*
FSCE	base	77.0	45.9	69.2	72.0	74.0	75.6	*76.8*
	novel	0.0	12.8	19.3	37.9	48.7	61.8	*67.5*

Table 4.2 Model performances on the second split.

Model	AP	1st stage	2nd Stage					
			Shots					
			1	*3*	*5*	*10*	*30*	*60*
Modified FRCNN	base	76.6	9.4	66.5	67.2	68.0	71.8	74.3
	novel	0.0	2.1	13.7	20.9	31.1	46.3	57.0
TFA	base	77.3	37.8	72.5	74.2	75.8	76.1	76.5
	novel	0.0	2.3	12.0	14.6	20.2	27.8	32.4
FSCE	base	77.4	60.9	72.3	73.5	74.3	75.7	76.2
	novel	0.0	14.7	27.9	29.3	37.2	47.7	53.8

4.2 Accumulated Degradation Test

In order to demonstrate the severe influence of the accumulated degradation on the base AP, two rounds of class expansion were conducted with 60 shots of novel images. Eight classes that were excluded from the both splits were chosen as the base for the test, and the base AP of the eight classes were evaluated. In the two rounds of expansion, the two splits was added in a reverse order to better simulate the actual situation. The modified FRCNN model drops 2.2% of the base AP from 76.2% to 74.0% in the first round, and a 2.4% dropping also happens in the second round (74.0% to 71.6%). The accumulated degradation on base AP is 4.6% for the modified FRCNN, and it proves that further expansion would recursively damage the base AP. In the meantime, the TFA only decrease 0.2% of base AP, and the FSCE even increase 0.6% after two rounds of expansion (Table 4.3). The result supports that the modified FRCNN with the best novel AP performances is not suitable for class expansion task.

Table 4.3 Comparisons of base AP with 60 shots.

Model	Base (8 classes)	12 classes	16 classes
Modified FRCNN	76.2	*74.0*	*71.6*
TFA	76.2	75.8	76.0
FSCE	77.1	77.2	77.7

4.3 Line Bot Application

A convenient identification tool is necessary to help farmers making proper management strategies. Mobile devices and messaging apps are widely-used in modern society. Thus, a Line Bot application was deployed to provide a real-time tea diseases identification service in practical fields. After receiving the image from the farmers, the application displays the locations and the classes of each disease and offers the suggestions on how to deal with them (Fig. 4.2).

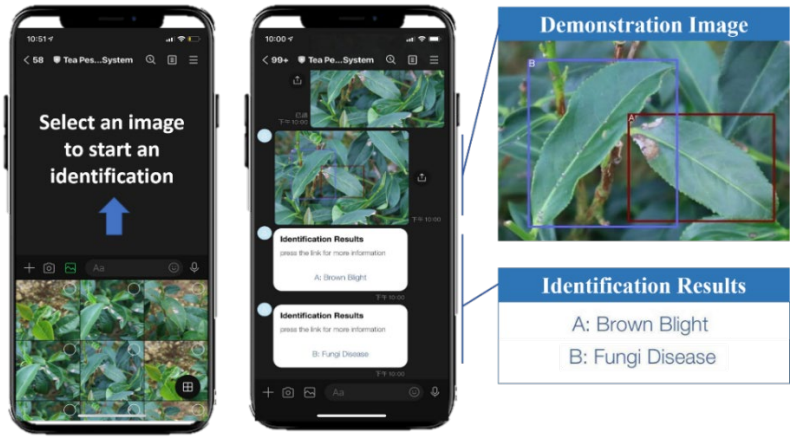

Figure 4.2 Line Bot demonstration

CHAPTER 5. CONCLUSION

This study aims to solve the bottleneck of developed plant identification systems and focus on the methods to efficiently expand novel classes. By decreasing the required amount of images to develop a model that maintains the original performance, the repetitive process of image collecting can be reduced. Tea diseases were selected as the target in the experiment. The proposed tea diseases identification system can detect on 16 categories of diseases after 564 images of four new classes were added in our study, and a FRCNN model can achieved a mAP of 78.04% on the dataset. Two FSOD methods of TFA and FSCE were implemented, and a modified FRCNN model was proposed to compare the results comprehensively. By comparing the performances of three FSOD models with six numbers of shots and two data splits, the TFA model can only achieves 30.3% novel AP in the best scenario. The modified FRCNN achieved best novel AP, but the accumulated degradation test demonstrates that it is not capable of maintaining a stable performance on base AP. The suggested model FSCE can achieve 67.5% of novel AP and only decrease 0.2% on base AP with 60-shot. The result shows that the adoption of FSOD on tea disease identification is feasible, and it saves time and lowers the cost of data collection in practical consideration.

REFERENCES

Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence*, 31(4), 299-315.

Chen, X., Lin, C., Lin, S., & Chen, S. (2021). Application of region-based convolution neural network on tea diseases and harming insects identification. *2021 ASABE Annual International Virtual Meeting, July 12-16, 2021*. https://doi.org/10.13031/aim.202100872

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311-318. https://doi.org/10.1016/j.compag.2018.01.009

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126-1135). PMLR.

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, 17(9), 2022.

Gidaris, S., & Komodakis, N. (2018). Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4367-4375).

Hariharan, B., & Girshick, R. (2017). Low-shot visual recognition by shrinking and hallucinating features. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3018-3027).

Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*.

Hu, G., Yang, X., Zhang, Y., & Wan, M. (2019). Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustainable Computing: Informatics and Systems*, *24*, 100353. https://doi.org/10.1016/j.suscom.2019.100353

Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., & Darrell, T. (2019). Few-shot object detection via feature reweighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8420-8429).

Liu, B., Ding, Z., Tian, L., He, D., Li, S., & Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Frontiers in Plant Science*, 11, 1082.

Oord, A. V. D., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Sun, B., Li, B., Cai, S., Yuan, Y., & Zhang, C. (2021). Fsce: Few-shot object detection via contrastive proposal encoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7352-7362).

Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., & Lin, L. (2019). Meta r-cnn: Towards general solver for instance-level low-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9577-9586).

Wang, X., Huang, T. E., Darrell, T., Gonzalez, J. E., & Yu, F. (2020). Frustratingly simple few-shot object detection. *arXiv preprint arXiv:2003.06957*.

Wang, Y. X., Ramanan, D., & Hebert, M. (2019). Meta-learning to detect rare objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9925-9934).

Zheng, Y. Y., Kong, J. L., Jin, X. B., Wang, X. Y., Su, T. L., & Zuo, M. (2019). CropDeep: the crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors*, 19(5), 1058.