2023 IEEE International Conference on Image Processing (ICIP) | 978-1-7281-9835-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIP49359.2023.10222401

PAIRWISE FEATURE LEARNING FOR UNSEEN PLANT DISEASE RECOGNITION

Abel Yu Hao Chai [a], *Sue Han Lee* [a], *Fei Siang Tay* [a], *Yi Lung Then* [b], *Hervé Goëau* [c], *Pierre Bonnet* [c], *Alexis Joly* [d]

[a] Swinburne University of Technology Sarawak Campus, Kuching, Sarawak, Malaysia
[b] Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia
[c] CIRAD, UMR AMAP, Montpellier, France
[d] INRIA, Montpellier, France

ABSTRACT

With the advent of Deep Learning, people have begun to use it with computer vision approaches to identify plant diseases on a large scale targeting multiple crops and diseases. However, this requires a large amount of plant disease data, which is often not readily available, and the cost of acquiring disease images is high. Thus, developing a generalized model for recognizing unseen classes is very important and remains a major challenge to date. Existing methods solve the problem with general supervised recognition tasks based on the seen composition of the crop and the disease. However, ignoring the composition of unseen classes during model training can lead to a reduction in model generalisation. Therefore, in this work, we propose a new approach that leverages the visual features of crop and disease from the seen composition, using them to learn the features of unseen crop-disease composition classes. We show that our proposed method can improve the classification performance of these unseen classes and outperform the state-of-the-art in the identification of multiple crop-diseases.

Index Terms— Plant disease identification, Feature generation, Unseen instances

1. INTRODUCTION

Although most deep learning models can achieve promising performance for plant species [1, 2, 3] or disease identification, when it comes to identifying unknown classes of multiple crop-disease pairs, performance is severely impacted [4, 5, 6, 7, 8]. The small scale of the publicly available plant disease datasets has also limited the ability of the deep learning models to be able to generalise to large and diverse classes of crop diseases. This prompts us to consider the possibility of exploiting the available data, i.e. labelled crop disease data, to deal with the unseen classes. We consider that since the same pathogen can attack multiple crop species, and those grouped under the similar *common disease* class tend to share similar visual symptoms. Thus, even if we do not have all the disease samples for each crop species, can we leverage the features of the readily available crop-diseases, and adapt them to other unseen crop-disease pairs? Our goal is illustrated in Figure 1(right). With the training set composed of a set of seen crops and diseases composition such as the ***tomato_early blight***, *pepper bell* ***_bacteria spot***, etc., we enrich the class diversity of the training data to encompass the distribution of unseen classes corresponding to the unseen composition of crop and disease, e.g. ***tomato_bacterial spot***.

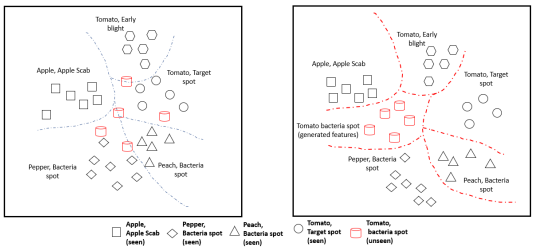

Fig. 1. The illustrated feature space learned (left) without considering the features of unseen classes, and (right) that enriches the class diversity of the training data to encompass the unseen classes.

Existing mainstream methods focus on converting this problem into a general supervised recognition task. They aim to directly predict diseases and crop species from the original visual features, ignoring their entanglement. For example, [7, 6] addressed the problem by training a symptom-oriented feature classifier that only considers diseases without crop species. This approach limits the model's ability to recognize plant diseases, as information on crop species can be useful to generate lists of diseases associated with crop species [9]. Other baseline methods presented in [8] show different configurations for learning crop and disease features, either by two-headed classifiers or by single-headed classifiers. However, experimental results show that by neglecting the composition of unseen classes during model training, the model tends to match individual concepts to seen classes, which limits recognition accuracy. Although conditional multitask

This research is supported by the Fundamental Research Grant Scheme (FRGS) MoHE Grant (Ref: FRGS/1/2021/ICT02/SWIN/03/2), from the Ministry of Higher Education Malaysia; and we gratefully acknowledged the support of NEUON AI with the GPU workstation used for this research.

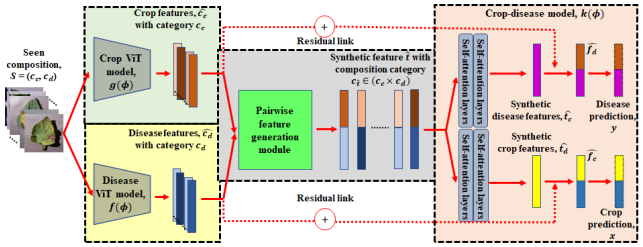

Fig. 2. The overview of the proposed model

learning (CMTL) proposed by [8] simultaneously learns the distribution of crop species and disease features with a conditional linkage between them, generalization of the model to unseen classes has not shown significant improvement. We also note that unlike general zero-shot learning reported on existing object recognition applications [10, 11], the current benchmark dataset for plant disease identification has not yet supported the semantic information about plant disease images. In other words, it is not possible to establish a semantic link between seen and unseen classes, which is essential for generalised zero shot learning.

Thus, considering that only the visual information of the individual concept (crop or disease) is available, we propose a new approach that focuses on the transferability of visual knowledge from a small sample to larger classes of crop diseases. More specifically, we first train embedding generators for crop species and diseases. Then, the output is fed into a pairwise feature generation module to increase the diversity of the training composition. The next module projects the combined plant and disease features into a common feature space that distinguishes between composition classes, including seen and unseen. The relatedness of the two concepts is further reinforced by the self-attention and reference connection with its complementary features. All the aforementioned modules are optimized in an end-to-end manner. In short, this paper has two main contributions. First, our studies provide a new direction on how to effectively use widely available labeled plant disease data to identify unseen classes based on visual information of the two important concepts of crop species and diseases. Second, we show that our proposed method can improve the classification performance of unseen classes of crop disease pairs and outperform the state-of-the-art in the identification of multiple crop-diseases.

2. METHOD

Problem Formulation. In this study, each image in our training and testing set is associated with a composition category, c, which consists of two concepts: the crop category and the disease category. More formally, a composition category can be defined as $c = (c_c, c_d) \in C$ where c_c and c_d correspond to the crop and disease categories respectively. For example, *Tomato* crop and *Bacterial Spot* disease will corresponding to *Tomato_bacterial spot* composition. Besides, same crop concepts or disease concepts can also appear in different composition such as *Tomato_early blight* and *Peach_bacterial spot*. Additionally, $C = S \cup U$ where S refers as seen compositions and U refers as unseen compositions. Our training set only consists of compositions in S but not in U. The goal of our study is to design a model to classify the compositions in U with only information from compositions in S.

Overall Framework. The architecture of our proposed model, as shown in Fig. 2, consists of three main modules: (1) base feature extractor, (2) pairwise feature generation module and (3) crop-disease model. The base feature extractor extracts features of individual concepts (crop features, $\hat{c_c}$ and disease features, $\hat{c_d}$) from the training set, X_s. The pairwise feature generation module generates synthetic composition features, $\hat{t}$ of different pairs that include seen and unseen compositions. Thereafter, the synthetic composition features, $\hat{t}$ will be used to train the crop-disease model for identification task. The details of each module are presented in the following subsections.

2.1. Base Feature Extractor Module

The backbone of our base feature extractor is built from ViT models by [12]. Specifically, it consists of crop ViT model, $g(\phi)$ and disease ViT model, $f(\phi)$. Crop ViT model takes an image x_i from X_s as input, and outputs crop features, $\hat{c_c} = g(x_i)$ while disease ViT model outputs disease features, $\hat{c_d} = f(x_i)$. The features produced ($\hat{c_c}$ and $\hat{c_d}$) will then be used by our next pairwise feature generation module to generate classes that encompass both S and U. Both ViT models consist of 12 layers, 12 attention heads and 768 embedded dimension. The size of the original image is reshaped to 224×224.

2.2. Pairwise Feature Generation Module

To enrich the diversity of classes in the training data, we propose the Pairwise Feature Generation module to generate features for different pairs of synthetic composition features to encompass U. In particular, the module obtains $\hat{c_c}$ and $\hat{c_d}$ from the base feature extractor as inputs, and combines them via feature summation to form the synthetic composition features, $\hat{t} = \hat{c_c} + \hat{c_d}$. The composition category of feature $\hat{t}$, denoted by $c_{\hat{t}}$, is formulated to include both S and U compositions, which means that $c_{\hat{t}} \in c_c \times c_d$, where $c_c \in L_c^{1...K}$ and $c_d \in L_d^{1...M}$ are associated with the crop and disease categories respectively, and K and M are the total number of crop and disease categories.

2.3. Crop-disease model

Our crop and disease model k takes the combined features $\hat{t}$ as inputs and produces the prediction of the crop, x, and disease, y, via a two-headed classifier. The dual classification

head with individual concepts is deployed here to improve the generalisation of the model by exploiting domain-specific information on crop species and disease concepts contained in the training data. However, simply combining $\hat{c_c}$ and $\hat{c_d}$ to form $\hat{t}$, does not explore the relatedness of the two concepts, especially when the two features may have domain distribution variations. Therefore, given the divergence of disease and crop species features, we exploit their entanglement by enriching the compositional feature representation with self-attention [13]. Furthermore, to preserve its compositional features for pairwise feature classification via a two-headed classifier, we introduce a reference function of its complementary features for each of these individual concepts. This is illustrated in Fig. 2 with the residual links. This is inspired by the CMTL architecture [8] where a conditional link is established between crop and disease features to encourage knowledge sharing and further improve disease identifications. With this, the final synthetic crop and disease features learned before the classifier's head are then $\hat{f_c} = \hat{t_c} + \hat{c_d}$ and $\hat{f_d} = \hat{t_d} + \hat{c_c}$ respectively. The final synthetic features are also normalized and projected into their classifier head with GeLu activation.

2.4. Training Strategy

The three aforementioned modules are trained end-to-end. The features, $\hat{c_c}$ and $\hat{c_d}$ are learned by our crop ViT model, $g(\phi)$ and disease ViT model, $f(\phi)$ using cross-entropy loss. The loss functions are defined as $L_{CE1} = \sum_{i=1}^{n} a_i log(g_i)$ and $L_{CE2} = \sum_{i=1}^{n} b_i log(f_i)$ where a_i and b_i are the truth label for crop and disease respectively. g_i and f_i are the softmax probability for the i^{th} class for crop ViT model and disease ViT model respectively. Another two cross-entropy losses are also applied on synthetic crop-disease feature, $\hat{t}$ by our crop-disease model $k(\phi)$ and the loss functions are defined as $L_{CE1\_com} = \sum_{i=1}^{n} c_i log(k_i)$ and $L_{CE2\_com} = \sum_{i=1}^{n} d_i log(k_j)$ where c_i and d_i are the truth label from original crop and disease label, k_i and k_j are the softmax probability for the i^{th} class. As the final module for learning synthetic crop-disease composition function is highly dependent on the individual features of $\hat{c_c}$ and $\hat{c_d}$, we ensure that our first two individual modules are optimized to a certain weight range in advance before proceeding with the training of the synthetic crop-disease composition function. Specifically, we use the moving weighted sum of all losses to train all modules end-to-end. The loss is defined as follows;

$$L_{final} = \alpha(L_{CE1}+L_{CE2})+(1-\alpha)(L_{CE1\_com}+L_{CE2\_com}) \quad (1)$$

We assign α as a weighting coefficient directly proportional to the number of epochs in our final loss function. Both crop ViT model and disease ViT model are pre-trained from ImageNet. Our model is trained for 15 epochs with an initial learning rate of 0.001 and then decreased by a factor of 10 every 5 epochs. We run the training using an NVIDIA GeForce RTX 3060 graphic card.

3. EXPERIMENTS

Our study is based on the largest available benchmark for multi-crop and disease identification, namely PlantVillage (PV), proposed by [14]. We sampled images of 10 crop-disease compositions from the PV dataset, and from the selected crop-disease compositions, 7 crop-disease compositions were selected as seen composition, S which was used as the training set, while the remaining 3 crop-disease compositions were sampled as unseen composition, U which used only in the testing set. The 7 crop-disease compositions in the training set are *Tomato_bacterial spot*, *Cherry_healthy*, *Grape_black rot*, *Corn_common rust*, *Potato_early blight*, *Squash_powdery mildew* and *Pepper_healthy*. The 3 unseen crop-disease compositions in the testing set are *Tomato_early blight*, *Pepper_bacterial spot* and *Corn_healthy*.

3.1. Comparing with State-of-the-Arts

Table 1. Performance comparison between SOTA models and our proposed model.

Model	Unseen Average Acc(%)			Seen Average Acc (%)
	P	D	PD	PD
ViT single network [12]	66.83	26.33	3.00	**100.00**
CMTL-ViT [8]	69.67	**31.17**	2.67	**100.00**
Proposed model	**77.50**	26.50	**11.67**	**100.00**

P is crop species identification. *D* is disease symptoms identification. *PD* is identification for both crop and disease concepts together (crop-disease identification).

In this experiment, we analyse the performance of our proposed model against ViT single network from [12] and CMTL network architecture proposed by [8]. ViT single network consists of two separate single networks for crop ViT model and disease ViT model (16 image patches with 224×224 image size) to perform plant and disease identification individually. The CMTL network architecture is able to simultaneously predict the concepts of plant species and diseases with a conditional link between them. To enable a fair comparison, we replace its base CNN model with ViT, and named it CMTL-ViT so that the depth of the feature learning layers is the same as our model and the performance comparison can focus on the design of the classification head. The results were tabulated in Table 1. The accuracy of plant disease identification for all models is calculated by post-prediction. Specifically, the plant disease classification will only be considered correct if the prediction of both crop species and disease is considered to be the top-1 accuracy for each concept.

From the Table 1, all models able to perform well in seen composition and our proposed model outperformed all SOTA models in unseen composition plant disease identifications. ViT single network and CMTL-ViT only able to achieve top-1 accuracy of 3.00% and 2.67% respectively in unseen composition crop disease identifications which are relatively poor when compared with our proposed model. This therefore proves that by taking into account the features of unseen

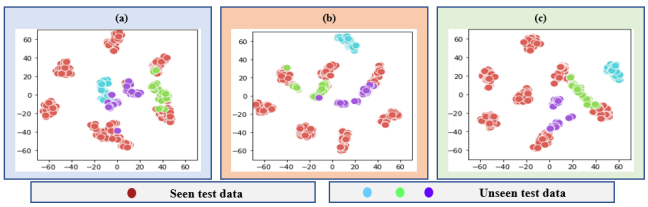

Fig. 3. Feature embedding visualization of (a) ViT single network, (b) CMTL-ViT and (c) proposed model using t-SNE.

compositional classes, the transferability of the models can be improved. Furthermore, our pairwise feature generation module is able to generate reliable data for unknown compositions, thus increasing the diversity of the training data, which improves model generalisation.

Qualitative analysis. From Fig. 3, it can be seen that the features learned in our proposed model are more semantically separable than the existing ViT single network and CMTL-VIT methods. This indicates that the features learned by our method are more discriminative compared to the existing methods. It also shows the importance of including the unseen classes to enrich the class diversity in the training data to improve the model generalization.

3.2. Ablation Study

Table 2. Ablation studies for our proposed model with different network elements and training schemes.

Proposed model	Unseen Ave Acc(%)		
	P	D	PD
two-head classifier	77.50	26.50	11.67
two-head classifier + RC	61.17	21.17	20.33
two-head classifier + RC + MWS	74.50	32.67	22.00
single-head classifier	77.00	34.33	23.83
single-head classifier + RC	71.33	37.33	28.00
single-head classifier + RC + MWS	**80.50**	**60.17**	**51.50**

RC is residual connection. *MWS* is moving weighted sum.

Two-head vs. single-head classifier In this section, we perform a performance comparison and ablation study with the implementation of a single-head classifier. A single-head classifier classifies crop and disease features under a single objective function combining both crop and disease concepts. From Table 2, we notice that the single-head classifier is able to further improve unseen composition crop disease identification (PD) from 11.67% to 23.83%. This is explained by the disease identification performance of the single-head classifier, which is able to achieve a higher accuracy of 34.33% compared to the two-head classifier of 26.50%. This is probably due to the fact that, in this scenario, the differentiation of disease symptoms may appear to be more difficult than that of crop species, as the visual appearances of the different disease symptoms may be very similar [8]. The combined crop and disease features formed by the single-head classifier could then, in turn, serve as an important cue to make the crop-disease features more distinctive between different classes.

Residual connection. We propose a residual connection between the seen crop features, $\hat{c_c}$ and synthetic disease features, $\hat{t_d}$, and vice versa, in order to promote a deeper knowledge sharing between the two concepts. The results presented in Table 2 show that the residual connection is able to improve our proposed model for both two-head classifier (11.67% to 20.33%) and single-head classifier (23.83% to 28.00%). We also note that despite the decrease in disease detection (D) for the two-headed classifier with residual connection, the overall crop disease identification was improved, confirming that the features with residual connection are more generalised and can better adapt to the unseen composition.

Moving weighted sum. We compare our final loss, L_{final} with that without the moving weighted sum loss, which is $L^{'}_{final} = L_{CE1} + L_{CE2} + L_{CE1\_com} + L_{CE2\_com}$. The result in Table 2 shows that our proposed model with moving weighted sum achieved the highest best performance for both the two-head classifier (20.33% to 22.33%) and the single-head classifier (28.00% to 51.50%). This therefore implies the importance of optimizing both the crop and disease ViT model up to a certain stage before starting to learn the synthetic crop-disease composition.

Batch size. We perform a grid search for each batch size from 2 to 128. We found that our model achieved best performance around batch size of 42, and it suffers performance drops when the batch size is reduced. This is probably due to the fact that the total composition, C available in our training set is 42 and our model can better learn the synthetic combined features when it sees all compositions including seen and unseen ones in each training batch.

4. CONCLUSION

To the best of our knowledge, this paper is the first to address unseen classes for multiple crop-disease identification. We have shown that our proposed model, which is able to enrich the class diversity in the training data by leveraging visual crop and disease features from the seen classes, can improve the model generalisation and outperform the existing approaches. We have also conducted ablation studies to analyse and improve the proposed model. For future work, we will extend our work to integrate the open-set and continual learning to adapt to a more challenging and realistic setting.

Acknowledgment

This research is supported by FRGS MoHE Grant (Ref: FRGS/1/2021/ICT02/SWIN/03/2) from the Ministry of Higher Education Malaysia; GPU is supported by NEUON AI.

5. REFERENCES

[1] Sue Han Lee, Yang Loong Chang, Chee Seng Chan, and Paolo Remagnino, "Hgo-cnn: Hybrid generic-organ convolutional neural network for multi-organ plant classification," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 4462–4466.

[2] Sue Han Lee, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in *2015 IEEE international conference on image processing (ICIP)*. IEEE, 2015, pp. 452–456.

[3] Sue Han Lee, Chee Seng Chan, Simon Joseph Mayo, and Paolo Remagnino, "How deep learning extracts and learns leaf features for plant classification," *Pattern Recognition*, vol. 71, pp. 1–13, 2017.

[4] Edna Chebet Too, Li Yujian, Sam Njuki, and Liu Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.

[5] Junde Chen, Jinxiu Chen, Defu Zhang, Yuandong Sun, and Yaser Ahangari Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, pp. 105393, 2020.

[6] Sue Han Lee, Hervé Goëau, Pierre Bonnet, and Alexis Joly, "Attention-based recurrent neural network for plant disease classification," *Frontiers in Plant Science*, vol. 11, pp. 1897, 2020.

[7] Sue Han Lee, Hervé Goëau, Pierre Bonnet, and Alexis Joly, "New perspectives on plant disease characterization based on deep learning," *Computers and Electronics in Agriculture*, vol. 170, pp. 105220, 2020.

[8] Sue Han Lee, Herve Goeau, Pierre Bonnet, and Alexis Joly, "Conditional multi-task learning for plant disease identification," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 3320–3327.

[9] Melissa B Riley, Margaret R Williamson, and Otis Maloy, "Plant disease diagnosis," *The Plant Health Instructor*, vol. 10, 2002.

[10] Yuval Atzmon, Felix Kreuk, Uri Shalit, and Gal Chechik, "A causal view of compositional zero-shot recognition," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1462–1473, 2020.

[11] Wenjia Xu, Yongqin Xian, Jiuniu Wang, Bernt Schiele, and Zeynep Akata, "Attribute prototype network for zero-shot learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21969–21980, 2020.

[12] Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations*, 2021.

[13] Hengshuang Zhao, Jiaya Jia, and Vladlen Koltun, "Exploring self-attention for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10076–10085.

[14] Sharada P Mohanty, David P Hughes, and Marcel Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, pp. 1419, 2016.