Computers and Electronics in Agriculture 215 (2023) 108408

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Known and unknown class recognition on plant species and diseases

Yao Meng [a,1], Mingle Xu [a,1], Hyongsuk Kim [a,*], Sook Yoon [b,*], Yongchae Jeong [c], Dong Sun Park [a]

[a] Department of Electronic Engineering, Core Research Institute of Intelligent Robots, Jeonbuk National University, South Korea
[b] Department of Computer Engineering, Mokpo National University, South Korea
[c] Department of Electronic Engineering, Jeonbuk National University, South Korea

ARTICLE INFO

Keywords:
Plant identification
Plant disease recognition
Deep learning
Open set recognition

ABSTRACT

Recognizing plant species and disease is essential to practical applications, such as keeping biodiversity and obtaining a desired crop yield. This study aims to extend the recognition from known to unknown classes in the context of plants, termed Plant-relevant Open-Set Recognition (POSR). In this task, a trained model is required to either classify an input image into one of the known classes or an unknown class, even if the model is only trained with the images of known classes. To achieve this task, we propose a method to obtain a high-performance classifier with compact feature distributions for known classes. To have a high-performance classifier, a ViT model pre-trained in the PlantCLEF2022 dataset is transferred, following an observation that a plant-related source dataset is more beneficial to plant species and disease recognition than other commonly used datasets, such as ImageNet. To have compact feature distributions, we adopt additive margin Softmax loss (AM-Softmax) which brings the distance smaller between the features of the same known class and hence gives more spaces for the unknown class. Extensive experimental results suggest that our method outperforms current algorithms. To be more specific, our method obtains AUROC 93.685 and OSCR 93.256 on average on four public datasets, with an average accuracy of 99.295 on closed-set classification. We believe that our study will contribute to the community and, to fuel the field, our codes will be public[2].

1. Introduction

"The only thing that never changes is that everything changes" – Louis L.Amour, which may be true in terms of ecological environment and crop growth. Changing often suggests that something new appears in systems, the opposite of the disappearing of existing things, and receives attention because of its potential impacts on whole environments, systems, and human societies, one of which is exemplified by COVID-19. In the context of plant science, remedies should be considered when new things are big troubles, such as new crop diseases that may lead to substantial yield loss (Savary et al., 2019) and invasive plant species that may cause quick diminish of other species (Yan et al., 2001). In these scenarios, *recognizing new things is a fundamental requirement*. This task becomes urgent partly because of climate change, a core triggering other non-trivial changes such as in plant traits (Wolf et al., 2022) and biodiversity (Mahecha et al., 2022). This paper aims to probe the potential of finding new plant diseases and species by utilizing optical images and deep learning methods.

Although the recognition of plant disease (Hughes et al., 2015; Thakur et al., 2022a) and species (Goëau et al., 2022; Chen et al., 2022; Beery et al., 2022) has witnessed a significant improvement with deep learning over the last five years, recognizing their new classes is still in its infancy. On the contrary, widely employed deep learning models elusively embrace a Closed Set Assumption (CSA) where the classes in the test process must be included in the training dataset (Geng et al., 2021; Scheirer et al., 2013). This assumption is violated for real-world applications when new classes appear (Scheirer et al., 2013; Bendale and Boult, 2016; Xu et al., 2023a).

To address this issue with CSA, we explicitly embrace Open Set Assumption (OSA) for plant-relevant recognition, termed Plant-relevant Open Set Recognition (POSR) that assumes new classes (unknown classes) existing in the test process. The classes in the training process refer to known classes. In generic, a model to achieve open set recognition (OSR) can only access the known classes in the training process yet is tested with both known and unknown classes. The trained models are required to either distinguish one image from one known class or into the unknown classes. On the contrary, one image belonging to an unknown class will be categorized into one of the known classes in CSA, which is risky in the real world when the unknown class results

* Corresponding authors.
E-mail addresses: hskimk@jbnu.ac.kr (H. Kim), syoon@mokpo.ac.kr (S. Yoon).
[1] Equal contribution.
[2] https://github.com/xml94/POSR

https://doi.org/10.1016/j.compag.2023.108408
Received 4 May 2023; Received in revised form 5 July 2023; Accepted 2 November 2023
Available online 9 November 2023

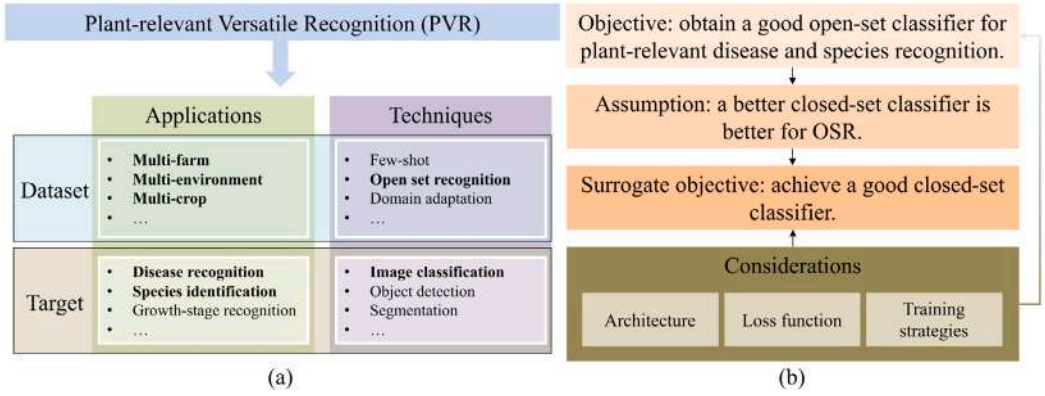

Fig. 1. (a) High-level problem formulation of plant-relevant versatile recognition. The bold suggests the objective of this study and others are taken as future work. (b) Understandings and our considerations to achieve the instantiated objective.

in problems. Besides, deep learning-based models are desired to know what they know and what they do not know; otherwise, these models lack explainability (Samek and Müller, 2019) and are not reliable.

A primary challenge in OSR is the absence of unknown classes in the training process and thus the models can only learn the data from known classes (Zhang et al., 2020). Although extra images are beneficial (Dietterich and Guyer, 2022), such as images from controlling classes (Fuentes et al., 2021) and known unknown images (You et al., 2022), we are interested in a more generic setting, without any images from unknown classes, because prior about the known unknown is limited in real-world applications.[3] In this scenario, *a common strategy is to learn a compact feature space for known classes*, which gives more spaces for unknown classes (Yang et al., 2022; Dietterich and Guyer, 2022; Du et al., 2022). We adopt a simple yet effective loss, additive margin Softmax loss (AM-Softmax) (Wang et al., 2018), which pushes the distance smaller between the features of the same known class.

In addition to POSR, it is interesting to train a model for multiple applications, such as plant identification and disease recognition, inspired by the observation that optical applications in plant science share some visual patterns. However, current research adopts one model to focus on one recognition task, such as either plant species identification (Ghazi et al., 2017; Chen et al., 2022; Ganguly et al., 2022) or plant disease recognition (Mohanty et al., 2016a; Fuentes et al., 2017; Ferentinos, 2018; Xu et al., 2022a), or single dataset, such as PlantVillage (Hughes et al., 2015), PaddyDoctor (Kiruba and Arjunan, 2023), and other private datasets. In contrast, this study explicitly proposes a plant-relevant versatile recognition (PVR), including POSR, which requires designing a single model for multiple recognition tasks and multiple applications as shown in (a) of Fig. 1. More details are described in the next section. To achieve this idea, we follow our previous work (Xu et al., 2022c) to adopt a large-scale plant-related dataset with huge image variations (Xu et al., 2023b), PlantCLEF2022, that encompasses commonness for other plant-relevant versatile recognition. We emphasize that PVR is an ambition and this study instantiates PVR as the boldface in Fig. 1(a) suggested. We hope that our work will promote understanding of finding new classes in plant diseases and species using deep learning and encourages other real-world applications, and show the potential towards plant-relevant versatile recognition.

To conclude, this study makes the following primary contributions:

- We explicitly propose a plant-relevant versatile recognition (PVR) from a unified perspective for real-world applications.
- We instantiate the PVR with the objective to obtain a good open-set classifier for plant-relevant disease and species recognition on different datasets, termed POSR.
- To achieve POSR, we adopt a ViT model pre-trained in the PlantCLEF2022 dataset, rather than CNN models pre-trained in plant-irrelevant datasets, and employ AM-Softmax to learn compact feature spaces for known classes, rather than generic Softmax.
- We execute extensive experiments with non-trivial analysis, which suggests that our method surpasses the current methods with clear margins.

2. Plant-relevant versatile recognition

As shown in Fig. 1(a), our ambition is to achieve a plant-relevant versatile recognition for multiple applications with different techniques from the computer vision field. Both of them can also be further considered from the perspectives of dataset and target. In real-world applications, the datasets have multiple crops cultivated in various farms and the images are taken in different environments such as lighting conditions and backgrounds. Those deviations may lead to huge image variations and further result in challenges to training a model with good generalization ability. The targets in the plant science community may also vary, such as recognizing plant diseases and species. In this case, existing methods are designed for a specific objective (Liu and Wang, 2021; Abade et al., 2021; Thakur et al., 2022b), such as either plant species identification (Xu et al., 2022b; Goëau et al., 2022; Chen et al., 2022) or plant disease recognition (Xu et al., 2022a; Rahman et al., 2020; Nanehkaran et al., 2020; Chen et al., 2021a). Similarly, most of the methods are only validated in a single dataset or single crop (Mohanty et al., 2016a; Liu and Wang, 2021; Xu et al., 2022a; Kendler et al., 2022; Wang et al., 2022; Chen et al., 2021c). Training and deploying these models are time-consuming and are often expensive to get decent performance. We rather highlight the idea that training one trained model for multiple datasets and targets, inspired by the observation that plant-relevant recognition essentially owns commonness across different applications such as the patterns on plant organs. Extra processing for the trained model is allowed and encouraged for different downstream applications. In this study, we instantiate this PVR idea by recognizing plant diseases and species on multiple datasets.

For real-world applications in the plant science community, techniques in the computer vision field can be utilized. The training dataset from the technique perspective may lead to some challenges. For example, a training dataset may only include a few images for each class in the few-shot case (Xu et al., 2022c, 2023a) whereas, in the domain adaptation, the test datasets are not in the similar distribution of the training datasets (Wu et al., 2023). A model may also be desired with diverse targets, such as that a class is the desired output for an image in image classification. Accordingly, a bounding box should be predicted for each class and multiple classes are possible for one image in object detection. In this study, we instantiate the task of PVR into the open set

[3] https://en.wikipedia.org/wiki/There_are_unknown_unknowns.

Table 1
Four utilized datasets and their highlights. These datasets are collected from three countries to verify our idea of PVR. The first three datasets are for plant disease recognition and the last one is for plant species identification.

Dataset	Image	Class	Description
PaddyDoctor	10,407	10	A rice disease dataset with real-world backgrounds in an unbalanced way. One image may include leaves, stems, and heads, not only diseased but also healthy, collected in India.
IVADLTomato	3021	9	A tomato disease dataset, unbalanced and multi-scale, collected from different greenhouses in South Korea.
IVADLRose	3132	6	A rose disease dataset with a similar collecting way as IVADLTomato.
CottonWeed	5187	15	A dataset to identify weeds in cotton field taken from different viewpoints under diverse lighting environments, collected in the United States.

and image classification case. We believe that it is beneficial to more complex tasks such as object detection. The idea is inspired by a current hot topic, foundation models, such as segmenting anything (Kirillov et al., 2023) where a pre-trained model can be utilized to perform multiple tasks. Along with our previous work for few-shot (Xu et al., 2022c), our experiments suggest that the foundation model for plant-relevant recognition shows potential, which therefore encourages the community to put more attention in this direction.

We believe that the proposed PVR is valuable in real-world applications and encourage a holistic perspective to consider these applications and techniques because some commonness is shared. Although the individual ones are listed in Fig. 1, their combinations are more common in the real world. We further instantiate it with the objective of this study to obtain a good open-set image classifier for plant-relevant diseases and species recognition, as shown in (b) of Fig. 1, with other goals remaining as future work. The achieve the instantiated objective, our considerations and methods are described in the next section.

3. Objective, material, and method

3.1. Plant-relevant open set recognition

As mentioned before, we aim to achieve plant-relevant open set recognition (POSR) for diseases and species recognition across different datasets, an instantiation of ambition, PVR. This subsection aims to define open set recognition (OSR) formally. Let the training dataset $\mathcal{D}_{tr} = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{N} \subset \mathcal{X}_{tr} \times \mathcal{Y}_{tr}$ where $\mathcal{X}_{tr}$ refers to the input image space and $\mathcal{Y}_{tr}$ denotes the label space, plant species and disease in this study. N is the number of inputs in the training dataset. Similarly, suppose the test dataset $\mathcal{D}_{te} = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{M} \subset \mathcal{X}_{te} \times \mathcal{Y}_{te}$ where M is its total number of inputs. With the Closed Set Assumption (CSA), training and test datasets share the same label space, namely $\mathcal{Y}_{te} = \mathcal{Y}_{tr}$, which is embraced by most previous works (Xu et al., 2022c,a; Chen et al., 2022; Rahman et al., 2020). However, a new type of class may appear when testing in real-world applications, such as unknown plant species and diseases. In this scenario, current methods will categorize the new type of class into one of the known classes in $\mathcal{X}_{tr}$, which is often risky. Therefore, it is desired to extend the closed set to the open set (Scheirer et al., 2012; Yang et al., 2021; Xu et al., 2023a).

Mathematically, the test dataset in OSR is denoted: $\mathcal{Y}_{te} = \mathcal{Y}_{tr} + \mathcal{Y}_{uk}$ where $\mathcal{Y}_{uk} \neq \infty$ refers to the *unknown* or new classes. A trained model is required to correctly classify a testing image either into the known classes $\mathcal{Y}_{tr}$ or into the unknown class $\mathcal{Y}_{uk}$ (Bendale and Boult, 2016; Yang et al., 2021). We argue that POSR is more complicated than generic OSR in computer vision mainly because the images from different classes in plant science may be very similar. In contrast, general computer vision split datasets, such as CIFAR10 (Krizhevsky et al., 2009) and TinyImageNet (Le and Yang, 2015), to evaluate the OSR (Scheirer et al., 2012; Oza and Patel, 2019; Zhang et al., 2020; Vaze et al., 2022) where one class semantically differentiates another class, such as dogs and cats.

3.2. Datasets

Three plant disease datasets and one plant identification dataset are utilized, as listed in Table 1. Some images from the four datasets are displayed in Figs. 2 to 4. PaddyDoctor (Kiruba and Arjunan, 2023) is originally a completion in Kaggle[4] to distinguish rice diseases in leaves, stems, and heads. The images are collected in the real world and one image has multiple rice organs. In particular, one image includes not only diseased organs but also healthy ones, which is more difficult than the images taken in laboratories (Rahman et al., 2020; Sethy et al., 2020) where one image includes only one dominant visual pattern of a single organ. Despite its difficulty, PaddyDoctor is utilized because of its closeness to a real application. It covers nine diseases and one health in an unbalanced way where some classes have much more images than other classes (Xu et al., 2022a). The original PaddyDoctor training dataset is split into our training and test datasets because the ground truth of the original test dataset is not publicly available. Further, the adopted IVADLTomato and IVADLRose[5] have 3021 images within 9 classes and 3132 images within 6 classes, respectively. The images with different illuminations are collected in multiple greenhouses in South Korea. The multi-scale challenge (Xu et al., 2023b) exists in the two datasets where the distances between cameras and the region of interest vary, with some other image variations such as backgrounds. Some classes in the original datasets have more than 1000 images and we limited the number of each class to less than 520 to make the recognition more complex.

CottonWeed (Chen et al., 2022), a weed dataset collected in the cotton field, is leveraged as plant species recognition. This dataset is collected in 2020 and 2021 under natural light conditions and at diverse growth stages. The weeds are also taken from different viewpoints with heterogeneous lighting environments. We observe that the images tend to have only one species of weeds with a similar clean background. The CottonWeed dataset is seriously imbalanced.

3.3. Method

3.3.1. Baseline to achieve OSR

A strong baseline method is first introduced because OSR is different from generic image classification with the closed-set assumption. Adversarial Reciprocal Points Learning (ARPL) (Chen et al., 2021b; Vaze et al., 2022), one of the current state-of-the-art methods to achieve OSR, is borrowed. K reciprocal points, the opposite of protocol points (Geng et al., 2021), are learned when training an ARPL model where K equals the number of known classes N. Each sample $\boldsymbol{x}$ is first embedded to be a feature $\mathrm{E}(\boldsymbol{x})$ where E is the embedded network. $\mathrm{E}(\boldsymbol{x}_k)$ is learned to be far away from the corresponding reciprocal point $\mathcal{P}_k$ where k is a class. In the test process, a sample will be assigned to be unknown if its embedded feature is close to one of the reciprocal points. Otherwise, it is classified to be the class whose reciprocal point is the farthest from its embedded feature. Fig. 5 displays the test process. For more details refer to Chen et al. (2021b) because this method is not our work. With this baseline, we will adopt two strategies as described below.

[4] https://www.kaggle.com/competitionspaddy-disease-classification/.
[5] https://github.com/IVADL/tomato-disease-detector.

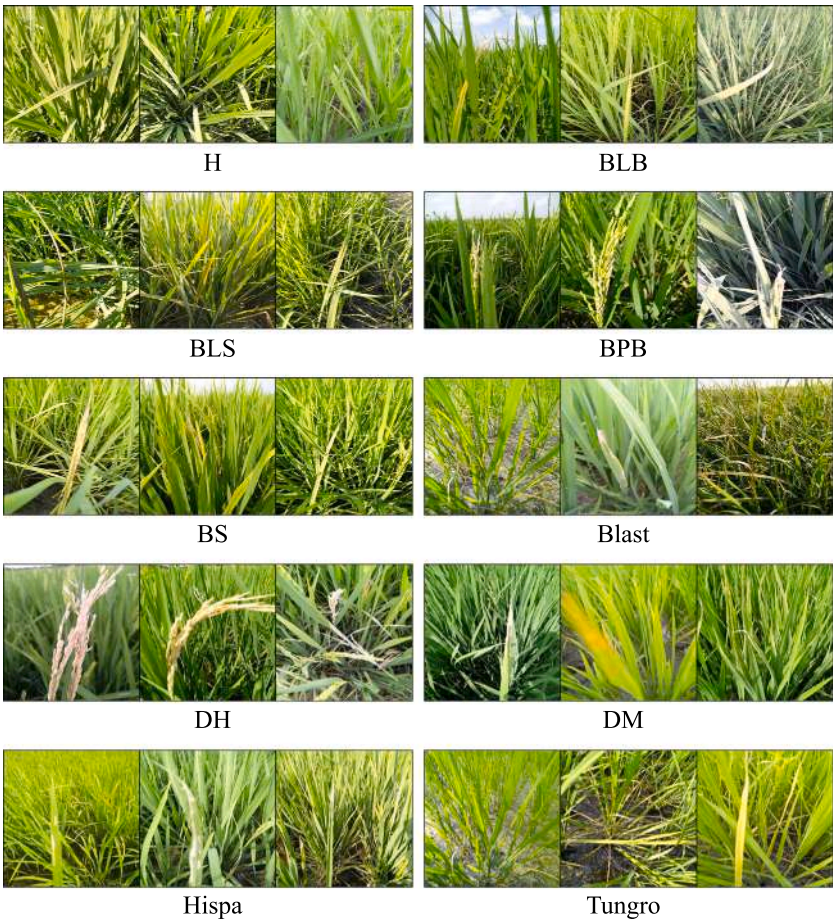

Fig. 2. Image examples in the PaddyDoctor (Kiruba and Arjunan, 2023) dataset. Every triplet denotes the same class. The images in this dataset own may not only have diseased but also healthy ones. The models should learn robust features for distinguishing healthy images from diseased ones.

3.3.2. Our strategies to achieve POSR across multiple datasets

To achieve POSR for different applications across multiple datasets, a straightforward way is to follow the divide-and-conquer paradigm,[6]. For example, individual models and specific training strategies are designed and trained for every heterogeneous task, plant, and dataset, which require enormous resources and are complicated to deploy. Our objective is rather to propose a unified algorithm by utilizing two strategies and considering three things described in the following paragraphs, as shown in (b) of Fig. 1 which are added to the baseline APRL model. Our strategies are illustrated in Fig. 6.

ViT model pretrained in large-scale PlantCLEF2022 dataset with huge image variations Inspired by the observation that a better closed-set classifier tends to be better in an open set (Vaze et al., 2022), we make a surrogate object, obtaining a decent closed-set classifier. To achieve this idea, our first strategy is utilizing transfer learning because it significantly boosts the performance in the downstream tasks (Mohanty et al., 2016b; Liu and Wang, 2021; Fan et al., 2022; Chen et al., 2021d; Zhao et al., 2022; Thakur et al., 2022a; Chen et al., 2022; Xu et al., 2022b; Goëau et al., 2022). In transfer learning, two factors are essential, type of architecture and source dataset (Xu et al., 2022c; Xu, 2023). Our strategy distinguishes from the widely used method in these two factors. A vision transformer (ViT) (Dosovitskiy et al., 2020), rather than a convolution neural network (CNN), is employed. The underlying reasons are from the empirical law that the models having higher performance generally have better transferability (Kornblith et al., 2019) and the observation that ViT models tend to get higher performance in many closed-set tasks, such as in ImageNet (Deng et al., 2009) classification accuracy of 85.9% with ViT-large (He et al., 2022) and 80.62% with ResNet152 (He et al., 2016) (a CNN model).

Second, a ViT model is pre-trained in the PlantCLEF2022 dataset (Goëau et al., 2022; Xu et al., 2022b). This dataset is *plant-related* and the performance in the target task tends to be better if the source and target dataset are similar (Kora et al., 2022; Matsoukas et al., 2022). In contrast, most current methods use ImageNet (Deng et al., 2009) as the source dataset, but it is unrelated to plants. PlantCLEF2022 is on a *large scale with huge image variations*, having 80,000 classes and 2,885,052 images and including multiple organs such as leaves and fruits. In contrast, a plant-related dataset, AIChallenger (Zhao et al., 2022), has much smaller image variations. We emphasize that the PlantCLEF2022 dataset is beneficial not only for a close-set classifier but also for multiple datasets because it allows the pre-trained model to learn many plant-related patterns. A ViT-large model pre-trained in PlantCLEF2022 is directly borrowed from our previous work (Xu et al., 2022c) to save training resources and time. To be more specific, the ViT-large model is first pre-trained in ImageNet in a self-supervised manner and then is pre-trained once more in PlantCLEF2022 dataset in a supervised manner, which facilitates the shortage of GPUs (Xu et al., 2022c).

Additive margin Softmax to have compact feature space within the same known class Our second strategy is using the additive margin Softmax loss function, rather than generic Softmax, to have a compact feature space. Different from closed-set classification, OSR models are entailed to output an unknown score to suggest the possibility of an input image belonging to the unknown, and a threshold can be tuned to finally make a decision with the unknown score. In this

[6] https://en.wikipedia.org/wiki/Divide-and-conquer_algorithm.

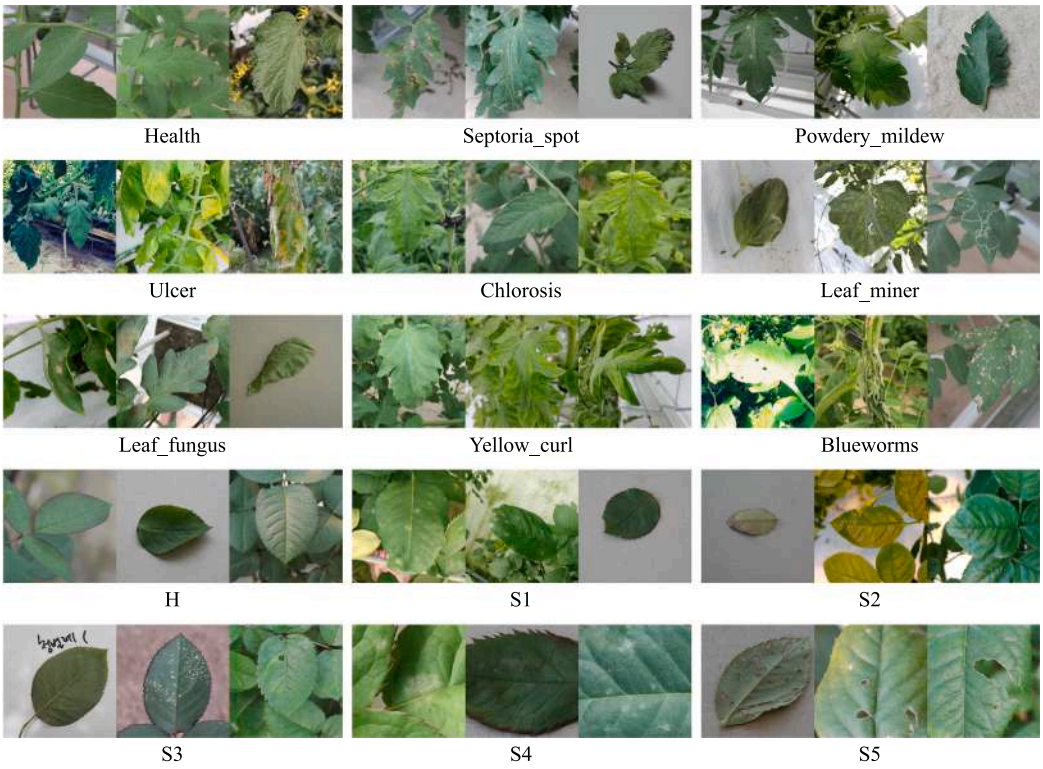

Fig. 3. Image examples in the IVADLTomato (the first three rows) and IVADLRose datasets (the last two rows). Every triplet denotes the same class where H and S refer to healthy and stress respectively.

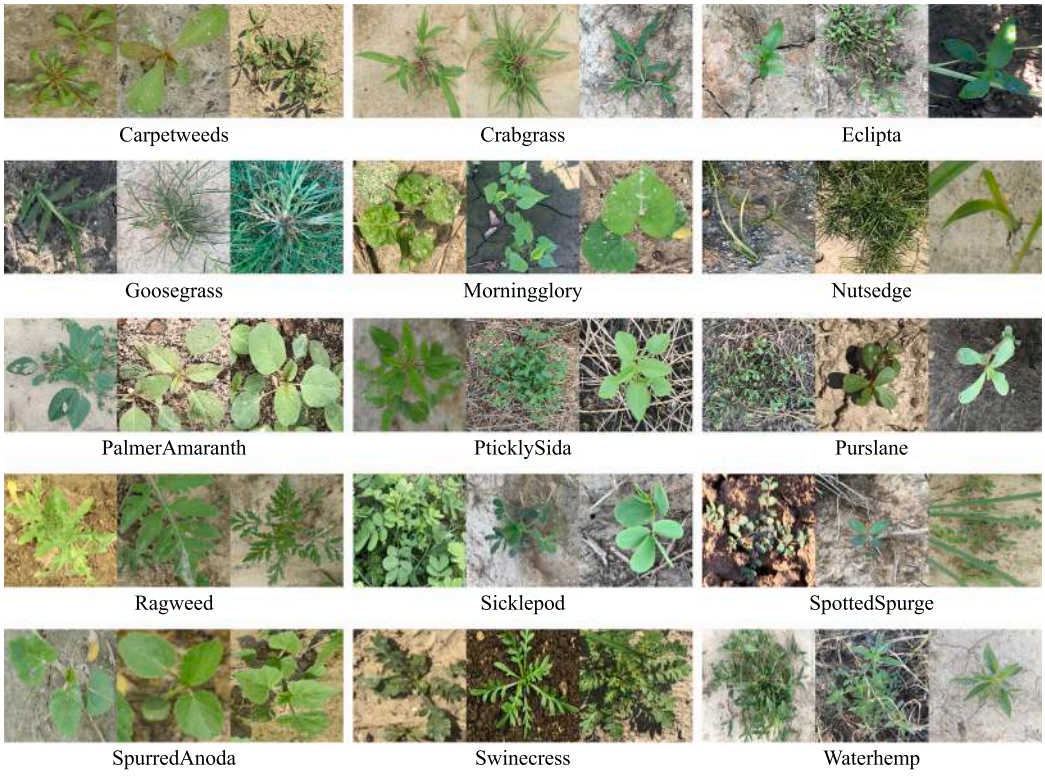

Fig. 4. Image examples in the CottonWeed. Every triplet denotes the same class.

case, a tight feature space for known classes is beneficial because it will give more space for the unknown class. Inspired by this idea, additive margin Softmax (AM-Softmax) (Wang et al., 2018) loss function is employed. Although it is originally designed for face verification, APRL

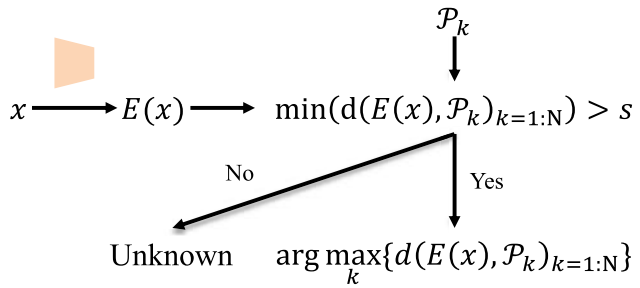

Fig. 5. Test process with an ARPL (Chen et al., 2021b) model, one of the state-of-the-art to achieve OSR. We use it as a baseline and add two more strategies, a ViT pre-trained embedded model in the PlantCLEF2022 dataset and additive margin Softmax loss (AM-Softmax). E is the embedding network and $\mathcal{P}_k$ is a learnable reciprocal point. d is a distance metric function and s is a threshold.

models to achieve OSR has the same core that the central or the reciprocal points are utilized to distinguish images.

To optimize a model to achieve multi-class image classification, cross-entropy loss for one sample can be formalized as

$$H(\boldsymbol{y}, \boldsymbol{p}) = \sum_{k=1}^{C} -y_k \log(p_k), \tag{1}$$

where y_k is "1" for the correct class and "0" for other classes. C is the total number of known classes and p_k is the likelihood for class k the model assigned for a given input and $\boldsymbol{W}_k$ is the corresponding weight vector. In generic Softmax, p_k can be computed as $p_k = \frac{\exp \boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_k}{\sum_{c=1}^{C} \exp(\boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_c)}$, where $\boldsymbol{f}$ denote extracted features for an input image $\boldsymbol{x}$ and $\boldsymbol{W}$ is the weights in the classifier layer. AM-Softmax (Wang et al., 2018) calculates the likelihood in a different way:

$$p_k = \begin{cases} \frac{\exp(s \cdot (\boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_k - m))}{\exp(s \cdot (\boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_k - m)) + \sum_{c=1, c \neq \hat{c}}^{C} \exp(s \cdot \boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_c)} & \text{if } y_k = 1; \\ \frac{\exp(s \cdot \boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_k)}{\exp(s \cdot (\boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_{\hat{c}} - m)) + \sum_{c=1, c \neq \hat{c}}^{C} \exp(s \cdot \boldsymbol{f}^{\mathrm{T}} \boldsymbol{W}_k)} & \text{otherwise}. \end{cases} \tag{2}$$

In Eq. (2), $y_{\hat{c}} = 1$ denotes the correct class. The hyperparameter margin m is only used for the correct class, and s is a scaling hyper-parameter. Importantly, the feature vector $\boldsymbol{f}$ and weight vector $\boldsymbol{W}_k$ are required to be normalized, such that $\boldsymbol{f} = \boldsymbol{f}'/\|\boldsymbol{f}'\|$ ($\|\cdot\|$ is a norm function) and $\boldsymbol{W} = \boldsymbol{W}'/\|\boldsymbol{W}'\|$ where $\boldsymbol{f}'$ and $\boldsymbol{W}'$ are the counterparts before normalization. The loss with this likelihood pushes the features belonging to the same known class tighter, which gives more space for the unknown class in OSR and thus contributes to achieving a better OSR.

3.4. Evaluation metrics

AUROC and CSA. OSR has two fundamental tasks that distinguish the unknown class from the known one and classify the known into the correct class. Because of these tasks, commonly adopted metrics in generic image classification are not reasonable to evaluate OSR (Scheirer et al., 2012). The first challenge to evaluating OSR is that different thresholds to find unknown from known will give different performances. Area Under the Receiver Operating Characteristic (AUROC) curve (Neal et al., 2018), not sensitive to thresholds is one useful metric. To compute AUROC, all known classes are cast as a super known class and the task becomes a binary classification along with the unknown classes as a super unknown class. With these two superclasses, true positive rate (TPR) and false positive rate (FPR) can be computed for the super known class. As the name of AUROC suggests, it can be considered as the area under the curve of TPR_s and FPR_s. Bigger AUROC commonly suggests that the model is better. Although AUROC is helpful to ease the effect of choosing a threshold, it essentially has no ability to evaluate if the model classifies the known classes correctly. Close Set Accuracy (CSA) (Vaze et al., 2022; Neal et al., 2018) can be used together in that a good open-set classifier should not degrade in the closed-set. Only known classes are taken to the trained model and then CSA is directly the generic accuracy. Therefore, we utilize AUROC and CSA together to evaluate different models considering the two tasks simultaneously.

OSCR Dhamija et al. (2018) considers the two OSR tasks simultaneously and aims to output a single value to evaluate trained models. The test datasets are split into known classes $\mathcal{D}_{known}$ and unknown classes $\mathcal{D}_{unknown}$. For samples from $\mathcal{D}_{known}$, the correct classification rate (CCR) is the fraction of the samples whose unknown scores are smaller than a given threshold s, and the learned classifier can also correctly classify them. For samples from $\mathcal{D}_{unknown}$, the false positive rate (FPR) is the fraction of the samples whose unknown scores are smaller than s. Similar to AUROC, OSCR is the area under the CCR_s and FPR_s curve.

4. Experiments and discussions

4.1. Implementation details

We split each dataset into training and testing sets at a ratio of 7:3. Six known classes are randomly chosen for PaddyDoctor and IVADLTomato datasets while four and ten known classes are randomly chosen for IVADLRose and CottonWeed, respectively. The other classes across these datasets are left as unknown classes. For each dataset, RandAugment (Cubuk et al., 2020) with AUG_m of 30, AUG_n of 2 is utilized and images are resized to 224. For the AM-Softmax loss function, according to preliminary experiments, the values of s and m are set to 10.0 and 0.5, respectively. We set the batch size to 32 and the number of workers to 16 with one RTX 3090 GPU (24 GB memory). We train the network for 100 epochs with a learning rate of 0.0001. Following the generic experimental setting of OSR, each model is trained on five different random splits of known and unknown on each dataset and the mean and variance of the performance are reported.

4.2. Compared methods

We designed several comparisons with different strategies and models. For the selection of methods for comparison, the following characteristics exist: CNN-based or ViT-based, and supervised or self-supervised, with or without transfer learning, whether training with PlantCLEF2022 (Xu et al., 2022c) or not. The compared methods are listed in Table 2 and described as follows:

- CNN-based
  - RN50. A ResNet50 model is trained from scratch with the target datasets listed in Table 1.
  - RN50-IN1k. A ResNet50 model is pre-trained with the ImageNet-1k (IN1k) dataset in a supervised way and then fine-tuned in the target datasets.
  - RN50-MoCov2. A ResNet50 model is pre-trained with the IN1k dataset in a self-supervised way and then fine-tuned in the target datasets.
- ViT-based
  - ViT-L. A ViT-large (Dosovitskiy et al., 2020) model is trained from scratch with the target datasets.
  - ViT-IN1k. A ViT-large model is pre-trained with the IN1k dataset in a supervised way and then fine-tuned in the target datasets.
  - MAE. A ViT-large model is pre-trained with the IN1k dataset in a self-supervised way (He et al., 2022).
- Ours. We emphasize that our methods are based on ViT. To explain easier, we put our methods at the same level as CNN- and ViT-based.

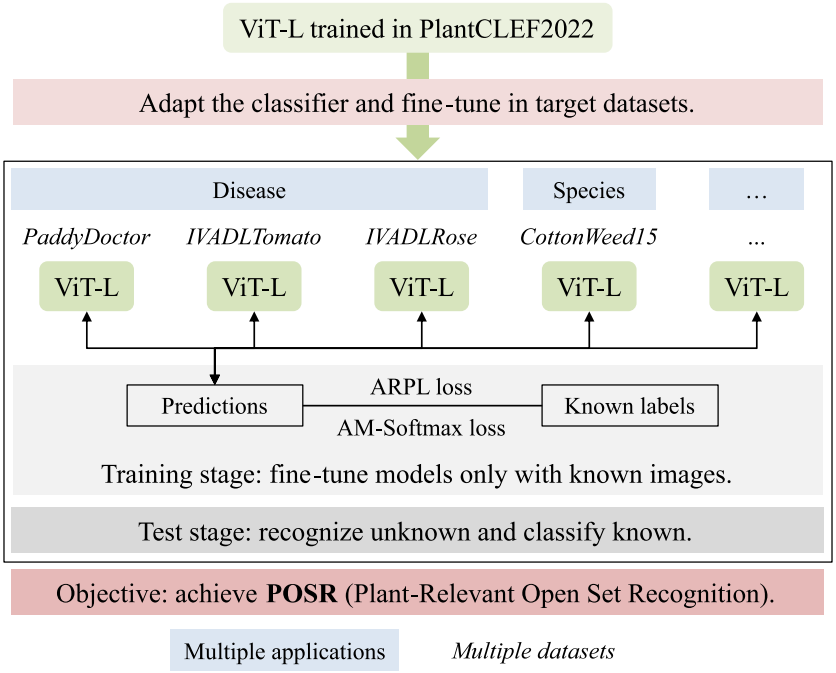

Fig. 6. Illustration of our strategies considering three things: model architecture, training strategy, and loss function. A ViT model (architecture) is pretrained in the PlantCLEF2022 dataset (training strategy). The last classification layer in the pretrained ViT-L model is adapted for downstream tasks and the whole model is finetuned using APRL and AM-Softmax loss (loss function) using individual datasets. Our ambition is to achieve PVR in multiple applications and multiple datasets.

Table 2
Characteristics of the compared methods. CE and AM represent Cross Entropy loss with generic Softmax and AM-Softmax, respectively. w/o denotes without using the corresponding datasets.

Architecture		Training strategies		Loss function
Model	Method	ImageNet-1k	PlantCLEF2022	CE/AM
CNN-based	RN50	w/o	w/o	CE/AM
	RN50-IN1k	Supervised	w/o	
	RN50-MoCov2	Self-supervised	w/o	
ViT-based	ViT-L	w/o	w/o	CE/AM
	ViT-IN1k	Supervised	w/o	
	MAE	Self-supervised	w/o	
Ours	MAE-PlantCLEF-CE	Self-supervised	Supervised	CE
	MAE-PlantCLEF-AM	Self-supervised	Supervised	AM

- MAE-PlantCLEF-CE. We pre-train the ViT-large model from MAE with the PlantCLEF2022 dataset once again in a supervised way and then fine-tuned it with the target datasets.
- MAE-PlantCLEF-AM. In MAE-PlantCLEF-CE, Softmax loss is replaced with AM-Softmax.

We want to emphasize that more comparisons are encouraged to highlight the contributions of our proposed strategies. Especially, a pretrained RN-50 model in PlantCLEF2022 dataset is highly desired to show the impact of architectures. However, pretraining models is not a trivial thing that often requests access to GPUs and tuning the hyperparameters. This situation becomes more challenging if the dataset is on a large scale, such as PlantCLEF2022. Due to this issue, we attempt to directly utilize existing pretrained models, and thus pretraining RN-50 in PlantCLEF2022 is not compared. Although it may be beneficial to the downstream performance, an empirical law, that the models having higher performance generally have better transferability (Kornblith et al., 2019), suggests that it will be inferior to ViT models pretrained in PlantCLEF2022. To be specific, ViT-based models outperform CNN-based models by a clear margin in the official report of the PlantCLEF2022 challenge (Goëau et al., 2022).

4.3. Main result

As our main objective is to achieve versatile plant-relevant open-set recognition across multiple datasets, we first compare our method to other strategies. Table 3 denotes the main results of different methods over four datasets mentioned in Table 1.

As shown in Table 3, experimental results show that our method significantly outperforms other methods on all datasets. Especially on the IVADLRose dataset, our method MAE-PlantCLEF-AM achieves 91.52 AUROC and 91.14 OSCR, 19.48 AUROC, and 19.89 OSCR higher than the MAE method. Moreover, our method outperforms the second-best method RN50-IN1k by 7.54 AUROC and 8.06 OSCR on the IVADLRose dataset. In the CottonWeed dataset, only PlantCLEF2022 pre-training (MAE-PlantCLEF-CE) can obtain 99.51 AUROC and 99.29 OSCR, 9.30 AUROC and 9.83 OSCR higher than the MAE method. Furthermore, AM-Softmax loss improves the classification accuracy on known data and achieves comparable results on AUROC and OSCR.

Regarding the impact of transfer learning, both the CNN-based and the ViT-based methods achieve much higher POSR results than ResNet50 and ViT models trained from scratch in all four datasets. Another important phenomenon can be seen from the experimental results. The ResNet50 model can achieve better POSR results than the ViT model on disease datasets (such as PaddyDoctor and IVADLRose datasets) as shown in Fig. 7. Still, when we apply PlantCLEF pre-training to the ViT model, the POSR capability of the ViT model can be significantly improved.

Due to the AM-Softmax loss, intra-class compactness and inter-class separability are achieved, so better closed-set classification can be obtained and open-set recognition is also improved. It can also be seen from Table 3 that the AM-Softmax loss function also plays a crucial role in promoting POSR in most datasets. Specifically, AM-Softmax loss improves AUROC and OSCR by 15.21 on the IVADLRose dataset.

The number of parameters and computation complexity are also displayed in Table 4 for two basic architectures, CNN and ViT. ViT-L model has much more learnable parameters and computations. The training time for every dataset is also shown in the table but we point out that the training time is impacted by not only models but also

Table 3
Main results on four datasets from different methods. TDS denotes the target data set. PD, IVADLT, IVADLR, and CW are PaddyDoctor, IVADLTomato, IVADLRose, and CottonWeed datasets. The boldface represents the best results in every dataset for the specific evaluation metrics.

TDS	Method	Loss	CSA	AUROC	OSCR
PD	RN50	CE	60.97 ± 0.03	60.68 ± 0.06	41.73 ± 0.04
	RN50-IN1k	CE	97.52 ± 0.01	87.57 ± 0.01	86.26 ± 0.01
	RN50-MoCov2	CE	97.97 ± 0.01	90.02 ± 0.01	88.90 ± 0.02
	ViT-L	CE	68.29 ± 0.08	62.15 ± 0.08	47.64 ± 0.09
	ViT-IN1k	CE	97.12 ± 0.01	85.95 ± 0.03	84.53 ± 0.03
	MAE	CE	97.11 ± 0.01	86.28 ± 0.03	84.87 ± 0.03
	MAE-PlantCLEF-CE	CE	98.50 ± 0.01	90.59 ± 0.02	89.73 ± 0.02
	MAE-PlantCLEF-AM	AM	**98.70 ± 0.01**	**91.69 ± 0.01**	**91.07 ± 0.01**
IVADLT	RN50	CE	68.90 ± 0.02	59.83 ± 0.13	47.79 ± 0.11
	RN50-IN1k	CE	95.67 ± 0.00	87.13 ± 0.02	84.74 ± 0.02
	RN50-MoCov2	CE	96.94 ± 0.01	86.17 ± 0.04	84.86 ± 0.04
	ViT-L	CE	68.76 ± 0.01	68.58 ± 0.04	53.34 ± 0.03
	ViT-IN1k	CE	98.81 ± 0.01	91.13 ± 0.03	90.55 ± 0.03
	MAE	CE	93.78 ± 0.02	84.32 ± 0.03	81.46 ± 0.03
	MAE-PlantCLEF-CE	CE	98.62 ± 0.01	90.98 ± 0.04	90.36 ± 0.04
	MAE-PlantCLEF-AM	AM	**99.29 ± 0.00**	**92.02 ± 0.05**	**91.75 ± 0.04**
IVADLR	RN50	CE	74.97 ± 0.04	58.96 ± 0.08	48.00 ± 0.06
	RN50-IN1k	CE	97.95 ± 0.01	83.98 ± 0.09	83.08 ± 0.09
	RN50-MoCov2	CE	98.69 ± 0.01	81.09 ± 0.10	80.58 ± 0.10
	ViT-L	CE	78.78 ± 0.04	59.65 ± 0.08	52.03 ± 0.07
	ViT-IN1k	CE	99.20 ± 0.01	74.12 ± 0.17	73.75 ± 0.17
	MAE	CE	97.88 ± 0.01	72.04 ± 0.19	71.25 ± 0.19
	MAE-PlantCLEF-CE	CE	99.23 ± 0.01	76.31 ± 0.18	75.93 ± 0.18
	MAE-PlantCLEF-AM	AM	**99.39 ± 0.01**	**91.52 ± 0.09**	**91.14 ± 0.09**
CW	RN50	CE	64.39 ± 0.02	60.12 ± 0.02	43.32 ± 0.02
	RN50-IN1k	CE	97.86 ± 0.01	86.82 ± 0.03	85.69 ± 0.03
	RN50-MoCov2	CE	98.50 ± 0.00	86.61 ± 0.05	85.82 ± 0.05
	ViT-L	CE	69.65 ± 0.03	63.95 ± 0.09	49.27 ± 0.05
	ViT-IN1k	CE	99.12 ± 0.00	94.88 ± 0.01	94.29 ± 0.01
	MAE	CE	98.54 ± 0.00	90.21 ± 0.03	89.46 ± 0.02
	MAE-PlantCLEF-CE	CE	99.75 ± 0.00	**99.51 ± 0.00**	**99.29 ± 0.00**
	MAE-PlantCLEF-AM	AM	**99.80 ± 0.00**	99.26 ± 0.01	99.10 ± 0.01

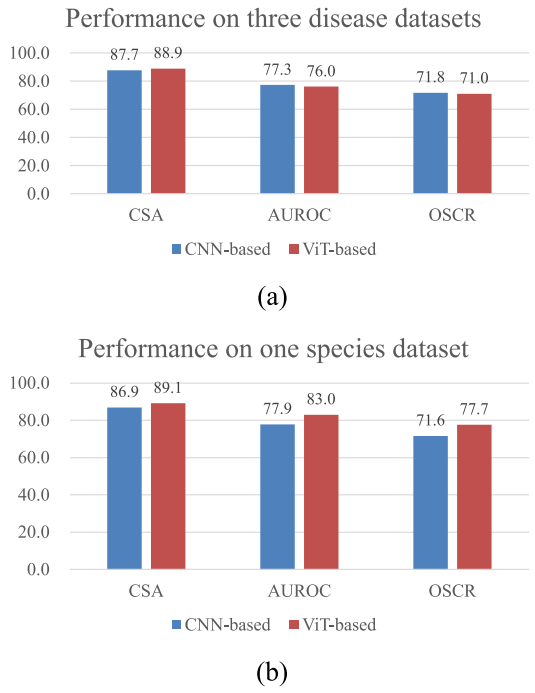

Fig. 7. Performance comparison on two model groups with the same loss function CE. (a) and (b) denote average performance on disease and species datasets, respectively. Compared to CNN-based methods, ViT-based methods achieve comparable open-set performance on disease datasets and better performance on species datasets.

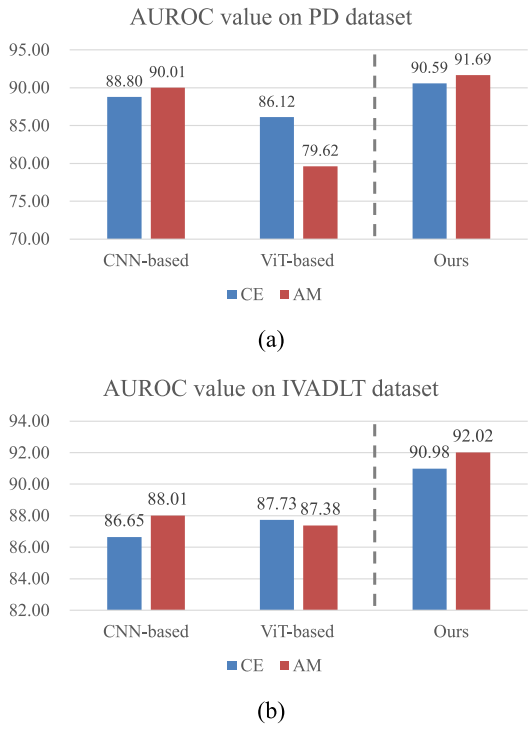

Fig. 8. AUROC comparison based on different methods. CE and AM represent Cross Entropy loss and AM-Softmax loss, respectively.

hardware and training settings, such as the number of workers to read data using PyTorch. From the table, ViT-L models spend a longer time but is tolerant. In terms of the FPS, ViT models can achieve the real-time requirement, 20 FPS.

4.4. Ablation study and visualization

To evaluate our method, in this subsection, we mainly analyze the impact of the architecture, AM-Softmax loss function, and pre-trained PlantCLEF2022 dataset. As mentioned before, one image has multiple plant organs, not only diseased but also healthy ones in the PaddyDoctor dataset as shown in Fig. 2. These characteristics make the PaddyDoctor dataset more complex. Moreover, it can be seen from the IVADLTomato dataset shown in Fig. 3 that different diseases are very similar, such as Health, Chlorosis, and Yellow_curl. Another important phenomenon is that the Powdery_mildew classes have very few pictures (49 pictures) in the dataset. Hence, we mainly choose these two datasets to analyze the ablation study. Moreover, we believe that the pre-training of the PlantCLEF2022 dataset will be helpful for CNN-based models, but there is currently no CNN-based pre-training model on the PlantCLEF2022 dataset. Therefore we only show the ViT-based pre-trained model on the PlantCLEF2022 dataset.

As shown in Table 5 and Fig. 8, a common phenomenon obtained from experiments on the PaddyDoctor and IVADLTomato datasets is that the AM-Softmax loss function can promote the performance of open sets in CNN-based methods. On the contrary, in the ViT-based models, AM-Softmax loss will reduce the classification effect of open sets. The ViT model with more capacity has a strong dependence on feature reuse (Matsoukas et al., 2022), while ImageNet-1k is far away from plant disease data, therefore if the features from ImageNet feature reuse are made more compact, the results of disease classification will decrease. However, it can be seen from Fig. 8 that after AM-Softmax loss is combined with the ViT model (MAE-PlantCLEF-AM) pre-trained on the plant-related PlantCLEF2022 dataset, the open-set recognition effect will be significantly improved. MAE-PlantCLEF-AM can obtain

Table 4
Comparisons of the number of parameters (params) of models, floating-point operations (FLOPs), training time with the hour (H) unit, and frames per second (FPS) in inference time. Although ViT models have much more parameters and require a longer time to be trained, their FPSs are still in real-time mainly because of the simplicity of image classification.

Model	Params (M) ↓	FLOPs (G) ↓	Dataset	Training time (H) ↓	FPS ↑
ResNet50	23.533	4.109	PD	0.35	190.61
			IVADLT	0.18	192.05
			IVADLR	0.18	193.04
			CW	1.78	187.26
ViT-large	311.296	61.603	PD	2.46	80.99
			IVADLT	1.00	80.84
			IVADLR	0.97	80.88
			CW	2.13	80.70

Table 5
Ablation study on the PaddyDoctor (PD) and IVADLTomato (IVADLT) datasets.

TDS	Method	Loss	CSA	AUROC	OSCR
PD	RN50-IN1k	CE	97.52 ± 0.01	87.57 ± 0.01	86.26 ± 0.01
	RN50-IN1k	AM	97.77 ± 0.01	88.77 ± 0.01	87.56 ± 0.01
	RN50-MoCov2	CE	97.97 ± 0.01	90.02 ± 0.01	88.90 ± 0.02
	RN50-MoCov2	AM	97.83 ± 0.00	91.25 ± 0.01	90.16 ± 0.01
	ViT-IN1k	CE	97.12 ± 0.01	85.95 ± 0.03	84.53 ± 0.03
	ViT-IN1k	AM	98.26 ± 0.00	84.44 ± 0.08	83.67 ± 0.08
	MAE	CE	97.11 ± 0.01	86.28 ± 0.03	84.87 ± 0.03
	MAE	AM	97.63 ± 0.01	74.79 ± 0.10	73.84 ± 0.09
	MAE-PlantCLEF-CE	CE	98.50 ± 0.01	90.59 ± 0.02	89.73 ± 0.02
	MAE-PlantCLEF-AM	AM	98.70 ± 0.01	91.69 ± 0.01	91.07 ± 0.01
IVADLT	RN50-IN1k	CE	95.67 ± 0.00	87.13 ± 0.02	84.74 ± 0.02
	RN50-IN1k	AM	96.84 ± 0.01	88.47 ± 0.04	86.83 ± 0.04
	RN50-MoCov2	CE	96.94 ± 0.01	86.17 ± 0.04	84.86 ± 0.04
	RN50-MoCov2	AM	96.91 ± 0.01	87.54 ± 0.04	86.04 ± 0.03
	ViT-IN1k	CE	98.81 ± 0.01	91.13 ± 0.03	90.55 ± 0.03
	ViT-IN1k	AM	97.83 ± 0.01	90.96 ± 0.02	89.92 ± 0.02
	MAE	CE	93.78 ± 0.02	84.32 ± 0.03	81.46 ± 0.03
	MAE	AM	94.18 ± 0.02	83.80 ± 0.03	81.28 ± 0.03
	MAE-PlantCLEF-CE	CE	98.62 ± 0.01	90.98 ± 0.04	90.36 ± 0.04
	MAE-PlantCLEF-AM	AM	99.29 ± 0.00	92.02 ± 0.05	91.75 ± 0.04

99.29 CSA, 92.02 AUROC, and 91.75 OSCR on the IVADLTomato dataset.

Fig. 9 shows that when the AM-Softmax loss is applied, the intra-class and inter-class feature spaces become more compact and separated, respectively. To quantify the compactness of feature space after applying the AM-Softmax loss function, we introduce the Silhouette Coefficient Score (SCS) metric as shown in Fig. 9. The SCS ranges between −1 and 1, where a higher silhouette coefficient refers to a model with more coherent clusters (Belyadi and Haghighat, 2021). Meanwhile, the MAE-PlantCLEF-AM and RN50-IN1k-AM methods can achieve 0.71 SCS and 0.54 SCS, respectively, much higher than using the cross-entropy loss. Furthermore, as can be seen from the red frame in Fig. 9, although unknown classes are identified as known classes, the overlap between known and unknown classes is reduced. The more interesting phenomenon is that the feature space of the unknown classes is also more compact and far away from the feature space of known classes, which makes the recognition rate of POSR higher.

Comparing MAE-PlantCLEF-CE with MAE in Table 5, it can be seen that the pre-trained PlantCLEF2022 dataset plays a key role in the POSR task. The MAE-PlantCLEF-CE method achieves 98.50 CSA, 90.59 AUROC, and 89.73 OSCR on the PaddyDoctor dataset 5, improving by 1.39 CSA, 4.31 AUROC, and 4.86 OSCR over the MAE method, respectively. The improvement of the POSR recognition effect on the IVADLTomato dataset is more obvious. Compared with MAE, MAE-PlantCLEF-CE improves 4.84 CSA, 6.66 OSCR, and 8.9 OSCR, respectively.

To illustrate the effect of AM-Softmax loss and PlantCLEF2022 pre-training on classification results in more detail, we obtained the confusion matrix of the known classes on the IVADLTomato dataset. From Fig. 10, it can be seen that both CNN-based and ViT-based methods can achieve better classification results when AM-Softmax loss is applied. Especially on the RN50-IN1k method, the classification accuracy of Leaf_miner disease has been significantly improved. Comparing MAE-PlantCLEF-AM and RN50-IN1k-AM, the ViT-based method can achieve an accuracy of 1.0 in most known diseases classification. Since Chlorosis and Yellow_curl disease are very similar, the CNN-based method misclassifies Chlorosis as Yellow_curl disease, while the ViT-based method can largely classify it correctly. As shown in Fig. 11, when the true positive rate is 95%, our method can better identify unknown classes (TN) and after utilizing AM-Softmax loss, the false positive ones are reduced a lot.

Fig. 12 illustrates image samples of incorrectly recognized unknown and known classes by the MAE-PlantCLEF-AM and RN50-IN1k-AM methods. Some diseases, especially early-stage diseases or diseases without obvious features, are wrongly recognized by both two methods. The MAE-PlantCLEF-AM method has a better recognition ability and can completely correctly classify Ulcer disease.

4.5. Extra understandings towards POSR

Openness tells us that knowing more is better to recognize new types of diseases. To analyze the impact of openness on POSR, we introduce openness based on the ratio of the numbers of classes in training and test sets (Zhang et al., 2020).

$$openness = 1 - \sqrt{\frac{k_{train}}{k_{test}}}, \tag{3}$$

where k_{train} and k_{test} are the number of classes in the training dataset and the test dataset, respectively.

To better analyze the impact of known classes on POSR, we define a novel method for sampling the number of known and unknown classes. Same with before experiments, we randomly split 'known and unknown' classes for five trials, moreover, we select a *fixed number*

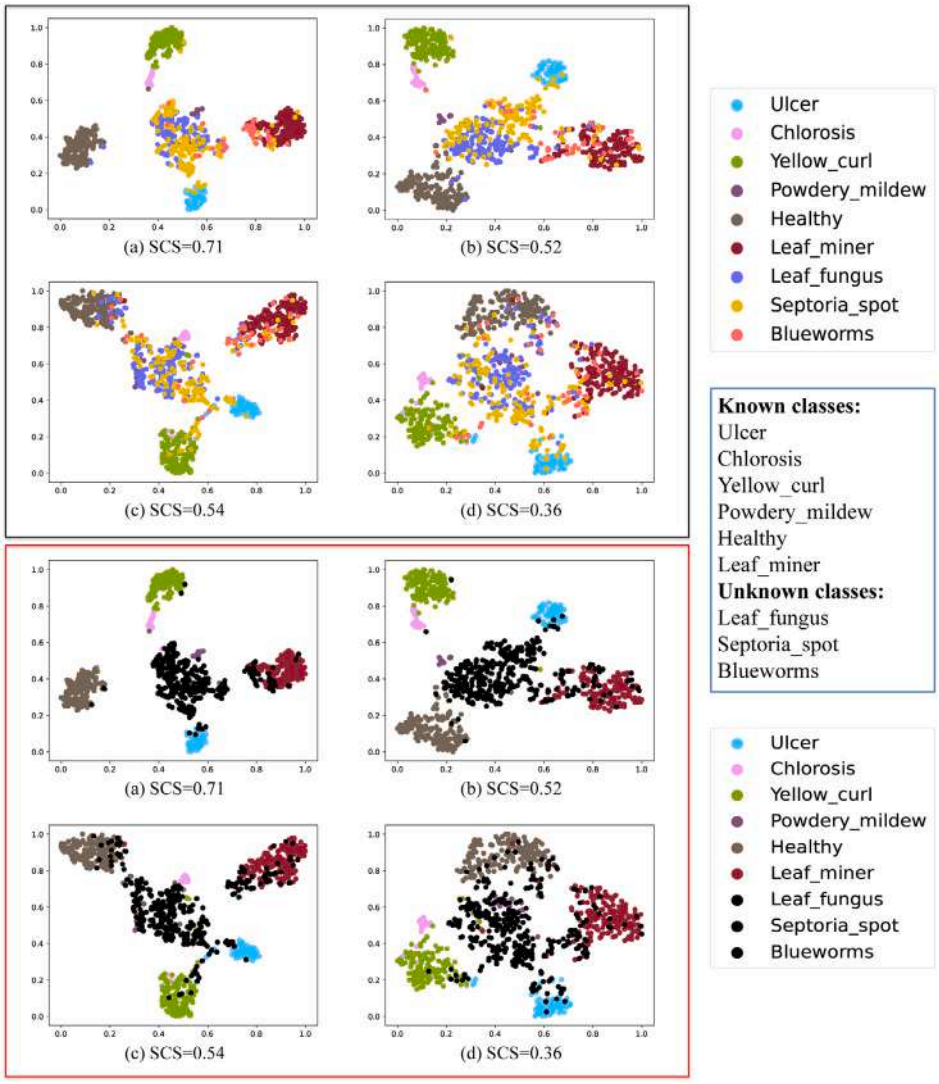

Fig. 9. Visualization using t-SNE on the IVADLTomato dataset. Six categories are from known data and three categories are from unknown data. Every known and unknown class is displayed with different colors in the top part and all unknown classes are displayed with the same black color in the down part. (a), (b), (c), and (d) denote our method MAE-PlantCLEF-AM, MAE-PlantCLEF-CE, RN50-IN1k with AM-Softmax loss (RN50-IN1k-AM), and RN50-IN1k with Cross Entropy loss (RN50-IN1k-CE), respectively. Please zoom in to see the details.

Table 6
Performance based on different openness on the PaddyDoctor dataset.

Openness	CSA	AUROC	OSCR
22.54%	98.70 ± 0.01	91.69 ± 0.01	91.07 ± 0.01
25.46%	99.14 ± 0.01	90.93 ± 0.03	90.53 ± 0.02
29.29%	99.15 ± 0.01	90.29 ± 0.04	89.91 ± 0.04
34.53%	99.19 ± 0.01	84.15 ± 0.06	83.87 ± 0.06
42.26%	99.72 ± 0.00	79.47 ± 0.04	79.40 ± 0.04
55.28%	100.00 ± 0.00	25.37 ± 0.06	25.34 ± 0.06

Table 7
Performance based on different openness on the IVADLTomato dataset.

Openness	CSA	AUROC	OSCR
18.35%	99.29 ± 0.00	92.02 ± 0.05	91.75 ± 0.04
20.94%	98.96 ± 0.01	91.79 ± 0.04	91.29 ± 0.04
24.41%	98.98 ± 0.01	92.75 ± 0.05	92.27 ± 0.05
29.29%	99.25 ± 0.00	91.56 ± 0.06	91.20 ± 0.05
36.75%	98.87 ± 0.01	90.97 ± 0.04	90.66 ± 0.04
50.00%	100.00 ± 0.00	15.29 ± 0.05	15.18 ± 0.05

of unknown classes for each trial and decrease the number of known classes to get different openness values. We fix four and three unknown classes for the PaddyDoctor dataset and the IVADLTomato dataset, respectively. In other words, the number of unknown classes is four, and the number of known classes decreases from six to one in the PaddyDoctor dataset. The fewer the number of known classes, the higher the openness.

As shown in Table 6, the higher the openness, the higher the accuracy of closed-set classification. On the contrary, because there are fewer known classes, it poses a great challenge to POSR, resulting in lower AUROC and OSCR. A noteworthy phenomenon is that when the openness is 22.54%, 25.46%, and 29.29%, that is, the number of known classes is greater than or equal to the number of unknown classes, the variation range of AUROC and OSCR is relatively small. But when the number of known classes is less than the number of unknown classes, such as the openness of 34.53% and 42.26%, the variation range of AUROC and OSCR starts to become apparent. Especially when there is only one known class (openness is 55.28%), AUROC and OSCR drop sharply.

While similar variations can also be seen in the IVADLTomato dataset, as shown in Table 7, some features emerge due to the presence of some very similar classes in the IVADLTomato dataset. When the

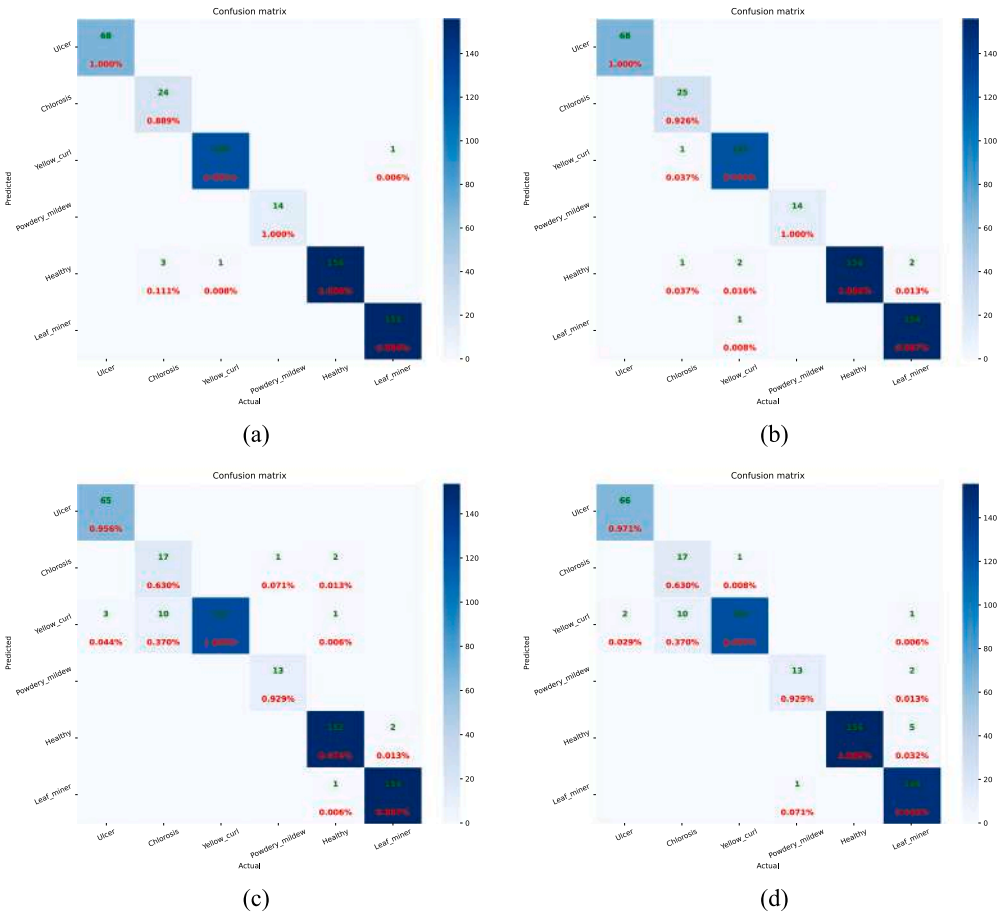

Fig. 10. One of the examples of confusion matrix of known classes in the IVADLTomato dataset. (a), (b), (c), (d) indicate our method MAE-PlantCLEF-AM, MAE-PlantCLEF-CE, RN50-IN1k-AM, and RN50-IN1k-CE, respectively. Please zoom in to see the details.

number of known categories is reduced, some similar diseases may disappear, resulting in higher AUROC and OSCR, for example, when the openness is 24.41%, AUROC, and OSCR are higher than when the openness is 18.35% and 20.94%. The same phenomenon is that when the number of known classes is one (the openness is 50.00%), the performance drops rapidly.

What kind of classes impact the performance of POSR The similarity of plant diseases has a large impact on the performance of POSR. To analyze the impact, we implemented open-set classification accuracy curves based on different true positive rates on the IVADL-Tomato dataset. As shown in Fig. 13, id 0 to id 4 represent five different "known/unknown" trials, where the known and unknown classes are different and chosen randomly. In Fig. 13(a), we follow the previous experimental settings, six classes are known classes and three classes are unknown. From the results, we can see that different 'known/unknown' combinations have a large influence on the performance of open-set classification. To understand what kind of classes affect performance, three unknown classes of id 1 are evaluated separately. Specifically, we select one of the three unknown classes as an unknown class and then use the model trained in (a) for validation. From Fig. 13(b), we can see that chlorosis has a large impact on performance because chlorosis is similar to Yellow_curl and health.

4.6. Social impacts and limitations

New class recognition is a fundamental requirement for many practical applications. We explicitly proposed plant-relevant open-set recognition and conducted preliminary experiments to recognize new plant species and diseases. An ambition, PVR, is highlighted simultaneously that a model or some strategies can be shared for different downstream applications, inspired by the observation of commonness among them. Our results suggest that pretraining big models such as ViTs in a large-scale dataset related to plants is promising. However, this ambition should be validated by more applications. Further, a straightforward question is the boundary of this ambition. One of the essential factors in our method is to pretrain a model in the PlantCLEF2022 dataset. Even though it is large-scale, it may not be large enough. Collecting a better dataset may be beneficial to the field yet is definitely not easy. A major limitation of PlantCLEF2022 for downstream tasks is only related to RGB images. Therefore, the pretrained models are expected to contribute less to other modalities, such as infrared images suitable for nights. Although POSR and PVR proposed in this study are very limited, we believe that our work has displayed possibilities and potential. With this point, more applications beyond plant disease and species recognition are highly encouraged. On the other hand, deploying POSR in real-world applications will be interesting, such as what a weed-removing robot should do when it finds unknown weeds. We believe that extending our work for different practical applications is charming.

To achieve POSR and PVR, we elusively embraced a perspective that they can be considered as a common task in general computer vision (Xu et al., 2023a). With this embrace, the techniques are borrowed from the latter to the former. However, plant-related tasks may have their dominant characteristics (Xu et al., 2023a), such that the similarities between plant diseases and species may be big. Another example is the scale and severity of plant disease. The widely adopted strategies in general computer vision may not always be optimal for this challenge. For example, human experts may look closely and carefully

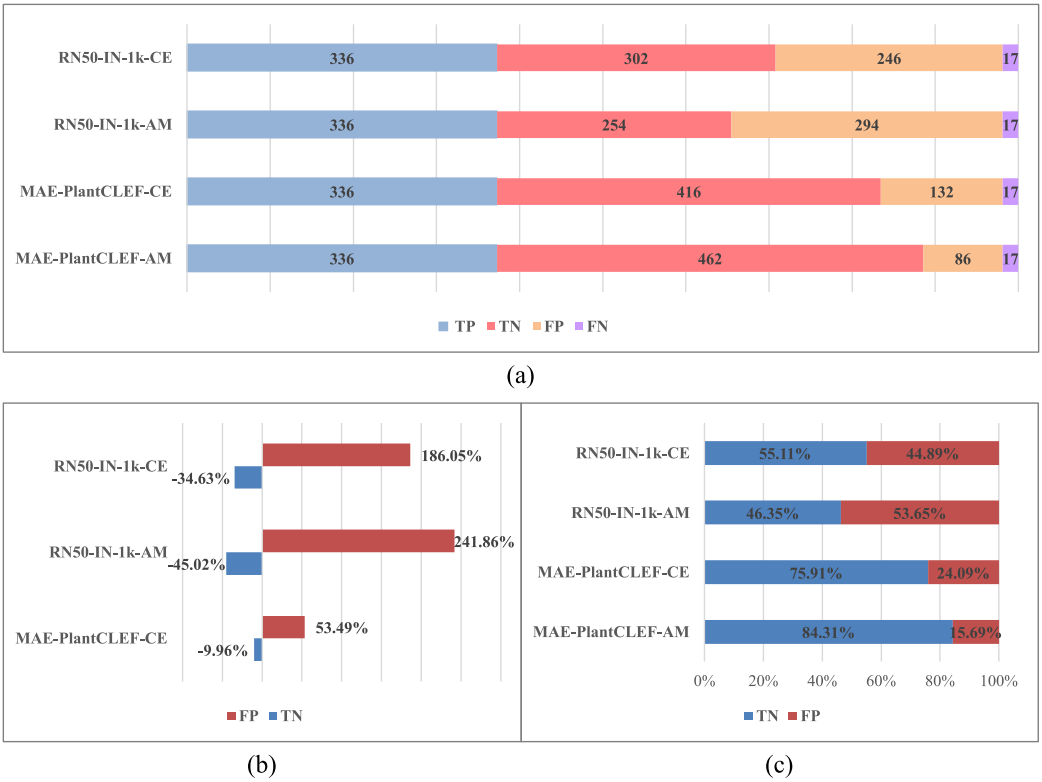

Fig. 11. One of the examples of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) samples of known and unknown classes at a 95% true positive rate ($TPR = \frac{TP}{TP+FN}$ is 95%) in the IVADLTomato dataset, where the known class is deemed positive. (a) the TP, TN, FP, and FN values are based on different methods. (b) compared with the MAE-PlantCLEF-AM method, the other three methods reduce the correct identification of unknown classes and increase the false recognition of known classes. (c) represents the proportion of different methods to identify TN and FP ($\frac{TN}{TN+FP}/\frac{FP}{TN+FP}$).

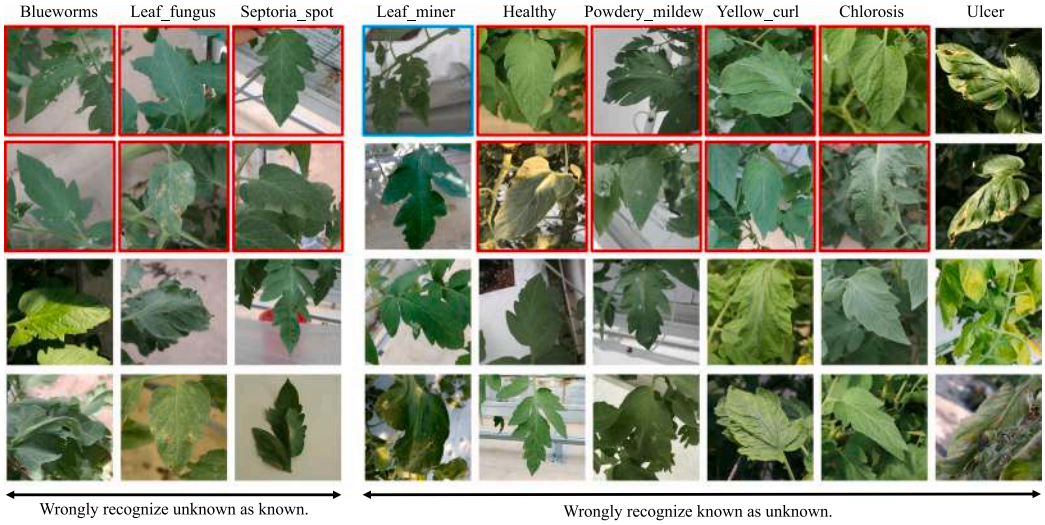

Fig. 12. Examples of misidentified images on the IVADLTomato dataset. The images with red boundaries are the wrong recognization for both MAE-PlantCLEF-AM and RN50-IN1k-AM, the image with blue boundary was the only wrong recognition by our model, MAE-PlantCLEF-AM, and the images without colorful boundaries are the only wrong recognition by RN50-IN1k-AM.

but current models often employed a fixed resolution of input images. Even though our method obtains decent performance in four public datasets, we cannot guarantee that it is beneficial for other datasets especially when the style to take pictures changes. However, varying the input resolution and using multiple observations may mitigate this issue when a low performance is observed. Preliminary results with the ResNet50 model are suggested in Table 8. We emphasize that the performance with high resolution tends to be higher but still lower than ours. In this case, employing a higher resolution for our model is encouraging and left as future work.

5. Conclusions

In this study, we described an ambition, PVR (plant-relevant versatile recognition), and instantiated it as POSR (plant-relevant open set recognition) for two specific tasks, plant disease, and species recognition. To achieve POSR, two strategies were utilized and incorporated

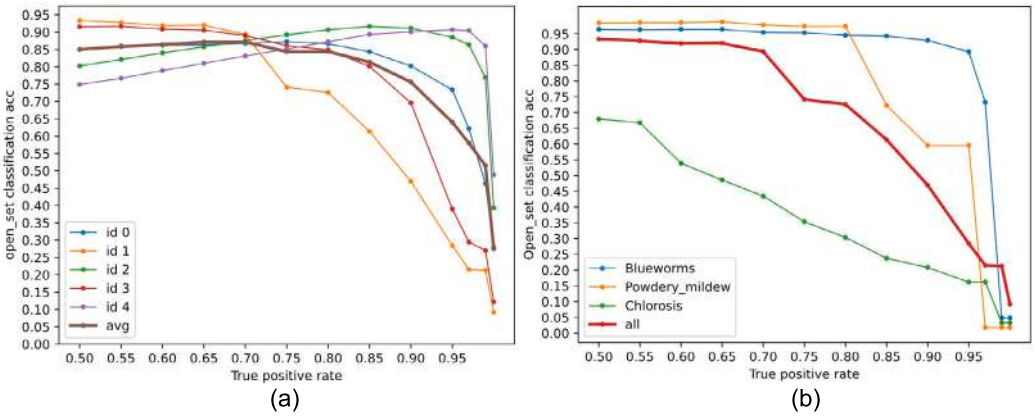

Fig. 13. Open-set classification accuracy with different true positive rates on the IVADLTomato dataset. (a) denotes five 'known/unknown' splits and the average result, respectively. (b) To illustrate what kind of classes affect the performance, three unknown classes of split one are evaluated separately.

Table 8
Preliminary experiments to probe the impact of input resolutions with the RN50-IN1k-CE method. Generally, the performance tends to be better when the resolution increases and the gain become smaller, but not always such as the case in the IVADLRose dataset. Besides, AUROC in The IVADLTomato dataset slightly increases when the input image size becomes larger. We conjecture that it is because the image scales are small-scale, compared to the CottonWeed dataset.

Dataset	Image size	CSA	AUROC	OSCR
PD	224	97.52	87.57	86.26
	336	97.89	87.77	86.68
	384	97.78	87.48	86.34
IVADLT	224	95.67	87.13	84.74
	336	96.77	88.99	87.20
	384	97.22	89.03	87.38
IVADLR	224	97.95	83.98	83.08
	336	98.56	79.51	78.89
	384	98.65	78.20	77.63
CW	224	97.86	86.82	85.69
	336	98.53	89.57	88.78
	384	98.76	90.41	89.71

with the state-of-the-art OSR method, APRL. To be more specific, a ViT-based model pretrained in a large-scale and plant-related dataset, PlantCLEF2022, was transferred for the downstream tasks, compared to the commonly used CNN-based models pretrained in either plant-irrelated or small-scale datasets. An AM-Softmax loss was employed to have a tight intra-class feature space that is beneficial to detect the unknown class. Our strategies were executed on four public datasets and the experimental results and ablation studies validate their effectiveness. The results also suggest that POSR and PVR are promising. In spite of the decent performance and some basic understanding of POSR, our model is desired to be improved for real-world applications. Social impacts and some limitations are summarized in the previous section. We hope that our work will contribute to the community and encourage more work and, to fuel the field, our codes will be public at https://github.com/xml94/POSR.

CRediT authorship contribution statement

Yao Meng: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Mingle Xu:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Hyongsuk Kim:** Writing – review & editing, Project administration. **Sook Yoon:** Formal analysis, Writing – review & editing, Supervision. **Yongchae Jeong:** Writing – review & editing, Supervision. **Dong Sun Park:** Formal analysis, Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

We appreciate the work and valuable suggestions of the editor and reviewers. This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Korea Smart Farm Foundation(KosFarm) through the Smart Farm Innovation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs(MAFRA), South Korea and Ministry of Science and ICT(MSIT), South Korea, Rural Development Administration (RDA), South Korea (421005-04). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2019R1A6A1A09031717). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2021R1A2C1012174).

References

Abade, A., Ferreira, P.A., de Barros Vidal, F., 2021. Plant diseases recognition on images using convolutional neural networks: A systematic review. Comput. Electron. Agric. 185, 106125.

Beery, S., Wu, G., Edwards, T., Pavetic, F., Majewski, B., Mukherjee, S., Chan, S., Morgan, J., Rathod, V., Huang, J., 2022. The auto arborist dataset: A large-scale benchmark for multiview urban forest monitoring under domain shift. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21294–21307.

Belyadi, H., Haghighat, A., 2021. Unsupervised machine learning: clustering algorithms. In: Machine Learning Guide for Oil and Gas using Python.

Bendale, A., Boult, T.E., 2016. Towards open set deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1563–1572.

Chen, J., Chen, J., Zhang, D., Nanehkaran, Y.A., Sun, Y., 2021a. A cognitive vision method for the detection of plant disease images. Mach. Vis. Appl. 32, 1–18.

Chen, D., Lu, Y., Li, Z., Young, S., 2022. Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. Comput. Electron. Agric. 198, 107091. http://dx.doi.org/10.1016/j.compag.2022.107091.

Chen, G., Peng, P., Wang, X., Tian, Y., 2021b. Adversarial reciprocal points learning for open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 44 (11), 8065–8081.

Chen, J., Wang, W., Zhang, D., Zeb, A., Nanehkaran, Y.A., 2021c. Attention embedded lightweight network for maize disease recognition. Plant Pathol. 70 (3), 630–642.

Chen, J., Zhang, D., Suzauddola, M., Nanehkaran, Y.A., Sun, Y., 2021d. Identification of plant disease images via a squeeze-and-excitation MobileNet model and twice transfer learning. IET Image Process. 15 (5), 1115–1127.
Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V., 2020. Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 702–703.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.
Dhamija, A.R., Günther, M., Boult, T., 2018. Reducing network agnostophobia. Adv. Neural Inf. Process. Syst. 31.
Dietterich, T.G., Guyer, A., 2022. The familiarity hypothesis: Explaining the behavior of deep open set methods. Pattern Recognit. 132, 108931.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
Du, X., Wang, Z., Cai, M., Li, S., 2022. Towards unknown-aware learning with virtual outlier synthesis. In: International Conference on Learning Representations.
Fan, X., Luo, P., Mu, Y., Zhou, R., Tjahjadi, T., Ren, Y., 2022. Leaf image based plant disease identification using transfer learning and feature fusion. Comput. Electron. Agric. 196, 106892.
Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145, 311–318.
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).
Fuentes, A., Yoon, S., Kim, T., Park, D.S., 2021. Open set self and across domain adaptation for tomato disease recognition with deep learning techniques. Front. Plant Sci. 12, http://dx.doi.org/10.3389/fpls.2021.758027.
Ganguly, S., Bhowal, P., Oliva, D., Sarkar, R., 2022. BLeafNet: A Bonferroni mean operator based fusion of CNN models for plant identification using leaf image classification. Ecol. Inform. 69, 101585.
Geng, C., Huang, S.-J., Chen, S., 2021. Recent advances in open set recognition: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (10), 3614–3631. http://dx.doi.org/10.1109/TPAMI.2020.2981604.
Ghazi, M.M., Yanikoglu, B., Aptoula, E., 2017. Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing 235, 228–235.
Goëau, H., Bonnet, P., Joly, A., 2022. Overview of PlantCLEF 2022: Image-based plant identification at global scale. In: CLEF 2022-Conference and Labs of the Evaluation Forum, Vol. 3180. pp. 1916–1928.
He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16000–16009.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.
Hughes, D., Salathé, M., et al., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060.
Kendler, S., Aharoni, R., Young, S., Sela, H., Kis-Papo, T., Fahima, T., Fishbain, B., 2022. Detection of crop diseases using enhanced variability imagery data and convolutional neural networks. Comput. Electron. Agric. 193, 106732.
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al., 2023. Segment anything. arXiv preprint arXiv:2304.02643.
Kiruba, B., Arjunan, P., 2023. Paddy doctor: A visual image dataset for automated paddy disease classification and benchmarking. In: Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD). pp. 203–207.
Kora, P., Ooi, C.P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W.Y., Meenakshi, K., Swaraja, K., Plawiak, P., Acharya, U.R., 2022. Transfer learning techniques for medical image analysis: A review. Biocybern. Biomed. Eng. 42 (1), 79–107.
Kornblith, S., Shlens, J., Le, Q.V., 2019. Do better imagenet models transfer better? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2661–2671.
Krizhevsky, A., Hinton, G., et al., 2009. Learning multiple layers of features from tiny images.
Le, Y., Yang, X.S., 2015. Tiny ImageNet visual recognition challenge.
Liu, J., Wang, X., 2021. Plant diseases and pests detection based on deep learning: a review. Plant Methods 17 (1), 1–18.
Mahecha, M.D., Bastos, A., Bohn, F.J., Eisenhauer, N., Feilhauer, H., Hartmann, H., Hickler, T., Kalesse-Los, H., Migliavacca, M., Otto, F.E., et al., 2022. Biodiversity loss and climate extremes—study the feedbacks. Nature 612 (7938), 30–32.
Matsoukas, C., Haslum, J.F., Sorkhei, M., Söderberg, M., Smith, K., 2022. What makes transfer learning work for medical images: feature reuse & other factors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9225–9234.
Mohanty, S.P., Hughes, D.P., Salathé, M., 2016a. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, http://dx.doi.org/10.3389/fpls.2016.01419.
Mohanty, S.P., Hughes, D.P., Salathé, M., 2016b. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419.
Nanehkaran, Y., Zhang, D., Chen, J., Tian, Y., Al-Nabhan, N., 2020. Recognition of plant leaf diseases based on computer vision. J. Ambient Intell. Humaniz. Comput. 1–18.
Neal, L., Olson, M., Fern, X., Wong, W.-K., Li, F., 2018. Open set learning with counterfactual images. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 613–628.
Oza, P., Patel, V.M., 2019. C2ae: Class conditioned auto-encoder for open-set recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2307–2316.
Rahman, C.R., Arko, P.S., Ali, M.E., Iqbal Khan, M.A., Apon, S.H., Nowrin, F., Wasif, A., 2020. Identification and recognition of rice diseases and pests using convolutional neural networks. Biosyst. Eng. 194, 112–120.
Samek, W., Müller, K.-R., 2019. Towards explainable artificial intelligence. In: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer, pp. 5–22.
Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., Nelson, A., 2019. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. 3 (3), 430–439.
Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E., 2012. Toward open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35 (7), 1757–1772.
Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E., 2013. Toward open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35 (7), 1757–1772. http://dx.doi.org/10.1109/TPAMI.2012.256.
Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K., 2020. Deep feature based rice leaf disease identification using support vector machine. Comput. Electron. Agric. 175, 105527.
Thakur, P.S., Khanna, P., Sheorey, T., Ojha, A., 2022a. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Syst. Appl. 208, 118117. http://dx.doi.org/10.1016/j.eswa.2022.118117.
Thakur, P.S., Khanna, P., Sheorey, T., Ojha, A., 2022b. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Syst. Appl. 118117.
Vaze, S., Han, K., Vedaldi, A., Zisserman, A., 2022. Open-set recognition: A good closed-set classifier is all you need? In: International Conference on Learning Representations (ICLR).
Wang, F., Cheng, J., Liu, W., Liu, H., 2018. Additive margin softmax for face verification. IEEE Signal Process. Lett. 25 (7), 926–930.
Wang, F., Rao, Y., Luo, Q., Jin, X., Jiang, Z., Zhang, W., Li, S., 2022. Practical cucumber leaf disease recognition using improved swin transformer and small sample size. Comput. Electron. Agric. 199, 107163.
Wolf, S., Mahecha, M.D., Sabatini, F.M., Wirth, C., Bruelheide, H., Kattge, J., Moreno Martínez, Á., Mora, K., Kattenborn, T., 2022. Citizen science plant observations encode global trait patterns. Nat. Ecol. Evol. 6 (12), 1850–1859.
Wu, X., Fan, X., Luo, P., Choudhury, S.D., Tjahjadi, T., Hu, C., 2023. From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. Plant Phenom. 5, 0038.
Xu, M., 2023. Enhanced Plant Disease Recognition with Limited Training Dataset Using Image Translation and Two-Step Transfer Learning (Ph.D. thesis). Jeonbuk National University.
Xu, M., Kim, H., Yang, J., Fuentes, A., Meng, Y., Yoon, S., Kim, T., Park, D.S., 2023a. Embracing limited and imperfect training datasets: opportunities and challenges in plant disease recognition using deep learning. Front. Plant Sci. 14.
Xu, M., Yoon, S., Fuentes, A., Park, D.S., 2023b. A comprehensive survey of image augmentation techniques for deep learning. Pattern Recognit. 137, 109347. http://dx.doi.org/10.1016/j.patcog.2023.109347.
Xu, M., Yoon, S., Fuentes, A., Yang, J., Park, D.S., 2022a. Style-consistent image translation: A novel data augmentation paradigm to improve plant disease recognition. Front. Plant Sci. 12, http://dx.doi.org/10.3389/fpls.2021.773142.
Xu, M., Yoon, S., Jeong, Y., Lee, J., Park, D.S., 2022b. Transfer learning with self-supervised vision transformer for large-scale plant identification. In: International Conference of the Cross-Language Evaluation Forum for European Languages. Springer, pp. 2253–2261.
Xu, M., Yoon, S., Jeong, Y., Park, D.S., 2022c. Transfer learning for versatile plant disease recognition with limited data. Front. Plant Sci. 13, http://dx.doi.org/10.3389/fpls.2022.1010981.
Yan, X., Zhenyu, L., Gregg, W.P., Dianmo, L., 2001. Invasive species in China—an overview. Biodivers. Conserv. 10, 1317–1341.
Yang, H.-M., Zhang, X.-Y., Yin, F., Yang, Q., Liu, C.-L., 2022. Convolutional prototype network for open set recognition. IEEE Trans. Pattern Anal. Mach. Intell. 44 (5), 2358–2370. http://dx.doi.org/10.1109/TPAMI.2020.3045079.
Yang, J., Zhou, K., Li, Y., Liu, Z., 2021. Generalized out-of-distribution detection: A survey. arXiv preprint arXiv:2110.11334.
You, J., Jiang, K., Lee, J., 2022. Deep metric learning-based strawberry disease detection with unknowns. Front. Plant Sci. 13, http://dx.doi.org/10.3389/fpls.2022.891785.
Zhang, H., Li, A., Guo, J., Guo, Y., 2020. Hybrid models for open set recognition. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. Springer, pp. 102–117.
Zhao, X., Li, K., Li, Y., Ma, J., Zhang, L., 2022. Identification method of vegetable diseases based on transfer learning and attention mechanism. Comput. Electron. Agric. 193, 106703.