



Research paper

Harnessing prototype networks for novel plant species and disease classification in open-world scenarios

Jiuqing Dong^{a,b,c}, Wei Jin^e, Alvaro Fuentes^{b,c}, Jaehwan Lee^{b,c}, Yongchae Jeong^b, Sook Yoon^{d,*}, Dong Sun Park^{b,c,**}

^a School of Computer and Information Engineering, Institute for Artificial Intelligence, Shanghai Polytechnic University, 2360 Jinhai Road, Pudong, 201209, Shanghai, China

^b Department of Electronic Engineering, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju, 54896, Jeollabuk-do, South Korea

^c Core Research Institute of Intelligent Robots, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju, 54896, Jeollabuk-do, South Korea

^d Department of Computer Engineering, Mokpo National University, 1666 Yeongsan-ro, Cheonggye-myeon, Muan-gun, 58554, Jeollanam-do, South Korea

^e School of Computer Science and Engineering, Macau University of Science and Technology, Wai Long, Taipa, 999078, Macau, China



ARTICLE INFO

Dataset link: <https://github.com/spMohanty/PlantVillage-Dataset/tree/master/raw>, <https://www.kaggle.com/c/herbarium-2019-fgvc6/data>

Keywords:

Novel class discovery
Anomaly detection
Plant disease recognition
Generalized class discovery

ABSTRACT

Identifying novel plant categories and diseases in open-world scenarios is crucial for modern agricultural production and applications. Recent studies have extended plant disease recognition from closed-set to open-set scenarios, aiming to reject samples that do not belong to known classes. However, beyond rejection, it is necessary to classify unknown samples rather than merely labeling them as "unknown." This paper assumes that images of unknown samples, rejected by open-set recognition algorithms, are available. We aim to classify these unlabeled samples by leveraging prior knowledge from a labeled set. To the best of our knowledge, no existing research has addressed the classification of unknown plant species and diseases. To fill this gap, we propose a novel prototype network that models the category space relationship between known and unknown classes. Specifically, we learn a prototype vector for each known category, enabling samples to obtain distance-based category probabilities by measuring their similarity to these prototypes. This approach captures complex class boundaries more effectively than linear classification models, offering greater flexibility and accuracy. Additionally, we employ a knowledge distillation loss to optimize the category space relationship and calculate a consistency loss to balance the model's classification performance for both known and unknown classes. To further boost performance, we incorporate the pre-trained model dino-v2. Experiments on the large-scale plant specimen dataset Herbarium19 and the plant disease dataset Plant Village demonstrate that our method surpasses baseline approaches, improving novel class accuracy by 1%–30%. This research contributes to intelligent agriculture, and we will release the code to facilitate future work in the field.

1. Introduction

Agriculture is the cornerstone of human civilization and constitutes a significant portion of the global economy. According to the latest United Nations Food and Agriculture Organization statement, the world population is expected to increase significantly, reaching 9.1 billion by 2050 (Alqahtani et al., 2023). Recently, artificial intelligence has been widely applied to automate various tasks related to plant sciences, such as wheat spike height estimation (Zaji et al., 2023), fruit grading (Palei et al., 2023), plant disease diagnosis (Xu et al., 2023). The application

of artificial intelligence in agriculture is expected to help farmers and agricultural experts increase global food production and promote the sustainability of agricultural ecosystems.

Recognizing plant species and diseases is crucial for the development of agricultural ecosystems. Over the past decade, the field of deep learning has achieved great success in applying these methods to plant-related areas. However, many models heavily rely on the availability of a large amount of labeled data for all relevant categories. This reliance

* Corresponding author.

** Corresponding author at: Department of Electronic Engineering, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju, 54896, Jeollabuk-do, South Korea.

E-mail addresses: jqdong@sspu.edu.cn (J. Dong), 3220001972@student.must.edu.mo (W. Jin), afuentes@jbnu.ac.kr (A. Fuentes), dlwo6@jbnu.ac.kr (J. Lee), ycjeong@jbnu.ac.kr (Y. Jeong), syoon@mokpo.ac.kr (S. Yoon), dspar@jbnu.ac.kr (D.S. Park).

¹ These authors contributed equally to this work.

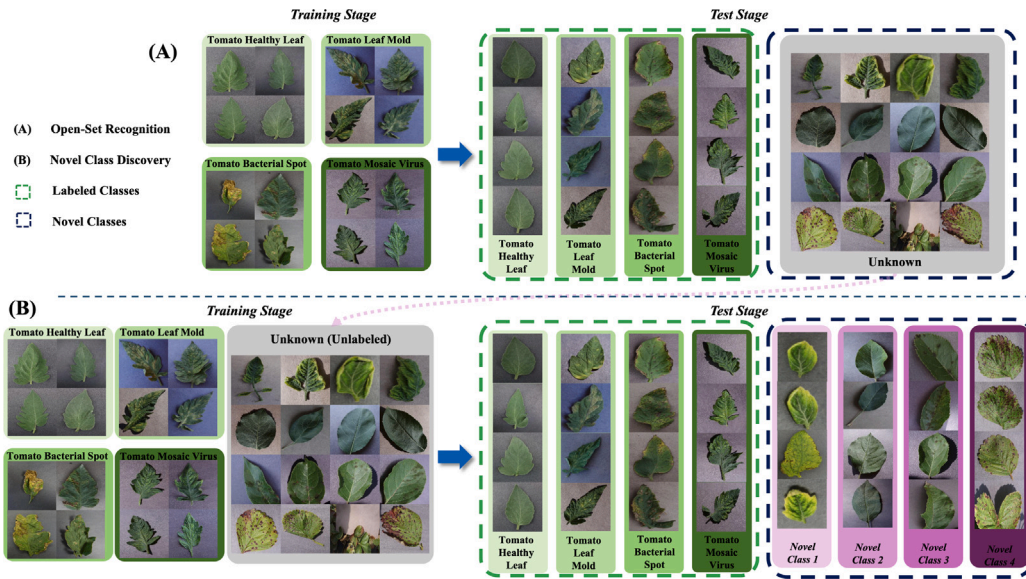


Fig. 1. Comparison of open-set recognition and novel class discovery. (A) Open-set recognition: The training stage involves labeled classes, and the testing stage requires identifying known classes and rejecting unknown classes. (B) Novel class discovery: The training stage includes labeled classes and unknown samples, with the testing stage focusing on classifying both known classes and novel classes.

poses a significant problem: Standard classification models may incorrectly classify instances that do not belong to any known category as belonging to one of the known categories. This increases the decision-making risk for systems deployed in real-world scenarios (Dong et al., 2023a). For example, a close-set model might identify invasive species as one of the known species.

When new entities cause significant problems, remedial measures should be considered, such as new diseases leading to substantial yield losses or invasive species that can cause rapid declines in other species. In these cases, identifying new species and diseases is a fundamental requirement. This task is called plant-related open-set recognition (POSER) (Meng et al., 2023). Open-set recognition requires a model trained on a labeled dataset to recognize and reject samples of unknown categories, as shown in Fig. 1(A). This task rejects out-of-distribution data and categorizes all unknown samples into a single class labeled 'unknown' (Dong et al., 2024). However, it is unreasonable to categorize all unknown samples as 'unknown'. We should address a more practical problem: Classifying unknown samples into specific categories. For example, an advanced system should further classify the samples in the top right corner of Fig. 1. Thus, the question arises: Can a model classify 'unknown' categories at a semantic level based on prior knowledge, similar to human ability?

Most current methods for recognizing unknown categories in plant-related tasks only evaluate the performance of rejecting unknown samples without learning and classifying these unknown samples, potentially leading to a waste of resources. Research (Dong et al., 2023a) has shown that a model trained to detect powdery mildew on tomato leaves can also detect powdery mildew on pepper leaves. This phenomenon inspired us to explore the relationships between plant disease and species categories. This paper aims to further classify these unlabeled data, a problem known as plant-related novel class discovery. Fig. 1(B) illustrates the difference between novel class discovery and open-set recognition. During the testing phase of open-set recognition, the model can easily reject some unknown samples in the test scenario. These samples, which should be further learned, are ignored in the open-set recognition task. In contrast, the novel class discovery task classifies all categories by learning from both labeled and unlabeled data.

After analyzing existing novel class discovery (NCD) algorithms (Finì et al., 2021; Gu et al., 2023; Han et al., 2021; Zhong et al., 2021; Han et al., 2020), we identified their common characteristics. Almost

all NCD algorithms follow a structure consisting of three components: a feature encoder, a linear classification head, and a pseudo-label optimization algorithm. These methods primarily rely on feature extractors and linear classifier heads to classify known classes and assign unknown samples to specific regions. Subsequently, the pseudo-label optimization algorithm enhances NCD performance by improving the quality of the pseudo-labels. However, by visualizing the feature space of linear classifier heads, we observed a mismatch between the number of clusters and the actual number of classes. As shown in Fig. 2, the Uniform Manifold Approximation and Projection (UMAP) for a 10-class tomato disease dataset reveals 13 clusters. We argue that such a mismatch resulted in smaller inter-class distances than intra-class distances, thereby negatively affecting pseudo-labels' initialization and ultimately impacting the NCD performance.

From a representation learning perspective, a linear classifier head essentially establishes class relationships in the semantic space. Dividing decision boundaries in the semantic space may result in a less compact feature space. This phenomenon is particularly pronounced in fine-grained datasets, such as plant disease datasets. Our motivation is that compact feature distributions for known categories generated during the pre-training phase benefit the category discovery process in the second phase. Therefore, we suggest constructing class space relationships directly in the feature space rather than the semantic space. To produce compact feature distributions, this paper proposes a new framework that constructs category spatial relationships using prototype networks. Specifically, a prototype network is a supervised learning neural network. It represents different categories of data by learning a category prototype for each category and adjusting the prototypes during training to maximize the distinction between categories. In the testing phase, new samples are assigned to the category represented by the most similar prototype, enhancing generalization ability. Our approach focuses on modeling the distance between prototypes and samples. Unlike linear classifiers that establish class relationships in the semantic space, our model directly constructs class relationships in the feature space. The prototype classification head allows us to create more compact clusters by minimizing the distance between samples and prototypes. This compact feature distribution facilitates better pseudo-label initialization for unlabeled classes, thereby improving NCD performance on fine-grained datasets. Our main contributions can be summarized as follows:

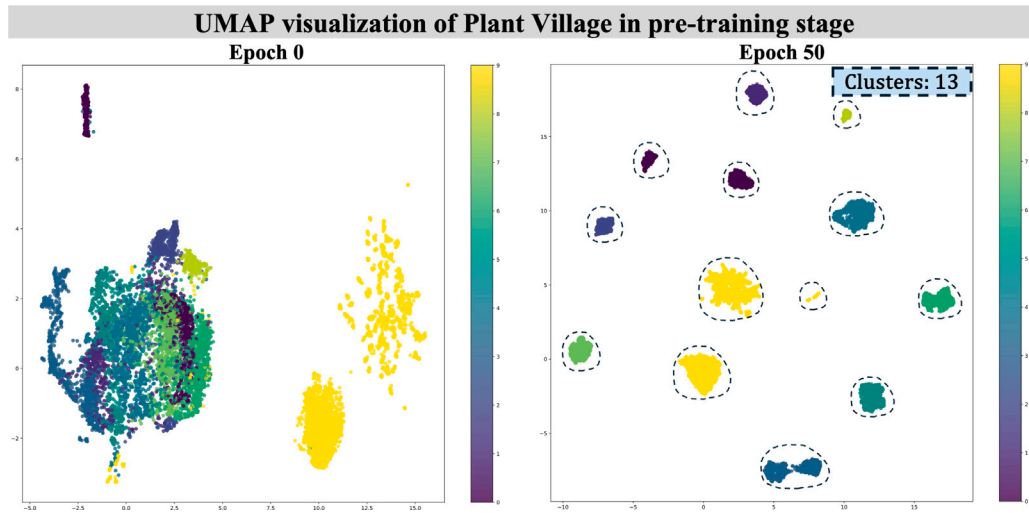


Fig. 2. UMAP visualization of Plant Village dataset in pre-training stage using a linear classification head. The pre-trained model initialized with dino-v2 shows a clearer separation and more distinct clustering. At Epoch 0 (left), the data points are highly overlapped. By Epoch 50 (right), the features form well-separated clusters. Despite the improved clarity, the distribution shows 13 clusters for 10 known categories, which could pose potential risks during the new class discovery process.

- We introduce a distance-based prototype classification head within a knowledge distillation framework to address the fine-grained task of novel class recognition for plant species and diseases. Our method is simple yet efficient, forming more compact and discriminative clusters while tackling two key challenges: establishing relationships between known and novel categories and estimating the number of novel classes.
- We establish benchmarks for plant-related novel class discovery through diverse data partitions and experiments on the large-scale Herbarium19 (Tan et al., 2019) and Plant Village (Hughes et al., 2015) datasets. Extensive experiments demonstrate that our method consistently outperforms baseline approaches under task-agnostic protocols and challenging conditions, such as when the number of novel classes is unknown or imprecise.
- We provide extensive visualizations and quantitative analyses, laying a solid foundation for novel class discovery in plant-related fields. Our study explores the applicability of the prototype classification head in plant phenotyping, investigates the impact of openness on novel class discovery, and offers comprehensive visual interpretations.

This paper is organized as follows: Section 2 introduces related works, including plant-related unknown class recognition, novel class detection, prototype networks, and knowledge distillation. Section 3 describes the problem definition and introduces the methodology for the NCD task. Section 4 reports the experimental details, data splits, evaluation protocol, main results, and related analysis. Section 5 discusses the effectiveness of the method, some issues related to novel class discovery, and the limitations of this work. Section 6 concludes our work.

2. Related work

2.1. Plant-related unknown class recognition

In recent years, deep learning has made significant advancements in plant-related recognition (Alqahtani et al., 2023; Chang et al., 2024; Singh et al., 2024; Nawaz et al., 2024; Dong et al., 2022, 2023b), yet the recognition of unknown categories remains an area of ongoing exploration. Fuentes et al. (2021) first proposed an open-set domain adaptation method for plant disease recognition, showing improvements in handling unseen data across different domains in tomato plant disease recognition. Studies (Shen et al., 2024; You et al., 2022;

Jiang et al., 2022) further explored open-set recognition (OSR) and out-of-distribution (OOD) detection techniques, achieving high performance in open-set recognition tasks for the tephritid fruit flies dataset and strawberry dataset. Furthermore, Meng et al. (2023) introduced a Plant-relevant Open-Set Recognition (POSR) method using a ViT model pre-trained on the PlantCLEF2022 dataset, achieving superior performance across multiple public datasets by adopting AM-Softmax loss for compact feature distributions. Dong et al. (2023a) were the first to advance the open-set recognition to open-set detection for plant diseases. They deployed an open-world paradigm for plant disease detection, capable of inferring and dynamically learning unseen disease categories, demonstrating robust generalization in cross-species disease detection experiments. These works have achieved decent performance in the task of recognizing unknown diseases. However, these works only categorize unknown classes as ‘unknown’ without further classification, which may be necessary in real-world applications. To this end, Alfarisy et al. (2025) proposed the unsupervised domain-specific open-world recognition framework, which addresses the challenges of identifying, discovering, and continually learning novel classes in a domain-specific open-world setting. Notably, their work demonstrates strong performance on plant species recognition using the VN-PLANT200 dataset (Quoc and Hoang, 2021).

2.2. Novel class discovery

Before delving into the NCD task, it is essential to clarify two key points: 1. The sample space of known classes in labeled data and unknown classes in unlabeled data are disjoint. 2. Known classes and unknown classes share similarities but are distinct. Based on this setting, Hsu et al. (2018) introduced the problem of transferring clustering models over tasks, which is consistent with the NCD problem. A main challenge in the novel category discovery task is that no labels exist for unknown categories. In this case, a common strategy is to use the compact feature space of known categories to guide the clustering of unknown samples. Han et al. (2019) aims to cluster novel classes with the help of known classes. They address the NCD problem proceeding in two steps: a data embedding is learned on the labeled data using a metric learning technique, and then fine-tuned while learning the cluster assignments on the unlabeled data. Interestingly, they also tackle the problem of estimating the number of classes in the unlabeled dataset. This work is also the first formally introduced The concept of novel class discovery. Currently, many NCD methods (Han et al., 2021; Zhong et al., 2021; Fini et al., 2021) adopt the training paradigm

in Han et al. (2019). Some studies (Han et al., 2020, 2021) have found that contrastive learning (Jaiswal et al., 2020; Jing et al., 2023) can significantly increase the accuracy of NCD tasks.

Then, researchers optimized the traditional two-stage method into a single-stage method, allowing labeled and unlabeled data to be trained simultaneously to obtain a more general representation through contrastive learning. For instance, the neighborhood contrastive learning (NCL) (Zhong et al., 2021) method combines contrastive learning and neighboring sample mining, utilizing positive and negative sample pairs to enhance NCD performance and Hard Negative Generation (HNG) to improve contrastive learning effectiveness. Notably, UNO (Fini et al., 2021) uses a multi-view self-labeling strategy (Asano et al., 2020) to generate pseudo labels suitable for known and unknown categories and then constructs a unified cross-entropy loss for training. It eliminates the need for self-supervised pretraining in NCD. Gu et al. (2023) observed that the relationship between unlabeled categories and labeled categories is difficult to capture during the discovery phase. Therefore, they proposed using a knowledge distillation framework to facilitate knowledge transfer from known classes to unlabeled classes, demonstrating the effectiveness of class relations in enhancing novel class discovery performance. Inspired by Gu et al. (2023), our work further explores the field of class relation modeling.

Recent works have extended NCD into more complex scenarios, such as generalized category discovery (An et al., 2024), object detection (Bharadwaj et al., 2023; Hayes et al., 2024), semi-supervised learning (Liu et al., 2023; Xiao et al., 2024), long-tailed recognition (Zhang et al., 2023), continual learning (Huang et al., 2024), and ultra-fine-grained categorization (Liu et al., 2024). While these methods significantly expand the applicability of NCD, they often redefine the problem setting to fit specific domains. For instance, many recent studies assume that novel classes emerge in an open-world setting where unknown categories continuously appear and need to be discovered in a semi-supervised manner and incrementally learned (Xiao et al., 2024; Huang et al., 2024). In contrast, our work adheres to the classical NCD definition where the known and unknown categories are disjoint but share certain semantic similarities. Unlike approaches that integrate knowledge distillation for incremental learning (Huang et al., 2024), our method focuses on modeling class relationships through a prototype-based knowledge distillation framework without assuming a continuous influx of novel categories. This distinction is crucial as it ensures that our approach remains robust within the predefined scope of NCD while avoiding complications introduced by task-specific adaptations.

2.3. Prototype learning

Prototype networks are deep learning models designed to learn prototype vectors that represent different categories. These vectors serve as central points for each category, enabling the mapping of unlabeled samples to the most similar prototype vectors of labeled classes and thereby enhancing the relationship between labeled and unlabeled data. Compared to traditional linear classifiers, prototype networks offer stronger representation and generalization capabilities, making them well-suited for tasks involving fine-grained sample spaces (Yang et al., 2018; Alfariy et al., 2023a,b).

In plant-related recognition tasks, different species or diseases often exhibit similar visual characteristics, making fine-grained classification particularly challenging. Prototype networks are especially effective in addressing these challenges. Yang et al. (2018) introduced convolutional prototype learning, which improves model robustness by learning prototypes that capture class distributions, demonstrating strong generalization in open-world scenarios. Building on this idea, Alfariy et al. (2023a,b) proposed a quad-channel contrastive prototype network, which enhances prototype separation through contrastive loss and multi-view inputs, achieving excellent performance in domain-specific open-set recognition tasks. Furthermore, they introduced a continual multi-channel contrastive prototype network to improve open-set

recognition and class discovery in dynamic environments (Alfariy et al., 2025). Additionally, Zhou et al. (2024) constructs semantic class prototypes to represent category distributions without labeled source data, utilizing a classification score-based prototype learning mechanism. Their findings further validate the role of prototype networks in refining class structures.

Inspired by the robustness and interpretability of these methods, we propose a prototype-based approach within a knowledge distillation framework to model the relationship between known and unknown categories in plant-related recognition tasks.

2.4. Knowledge distillation

Knowledge distillation is a crucial technique for effectively transferring knowledge from a teacher model to a student model, enabling model compactness and improved performance (Gou et al., 2021; Park et al., 2019; Wang and Yoon, 2021). Traditionally, this process involves aligning the outputs of the student model with those of a larger teacher model. In contrast, our approach employs teacher and student networks of identical parameter scales. This design focuses on leveraging knowledge distillation to establish relationships between known and unknown classes. During the pre-training stage, we train the teacher model exclusively on the labeled dataset. In the subsequent discovery stage, the teacher model is frozen and guides the training of the student model. By utilizing the output distributions of the unlabeled samples generated by the teacher model, the student model gains supervision for classifying these unlabeled instances.

To enhance plant-related novel class discovery, we integrate prototype networks with knowledge distillation. The prototype network models spatial relationships between known and unknown categories, while the knowledge distillation framework facilitates the transfer of category relation distributions into a shared feature space. This shared feature space not only regularizes the discovery process but also improves class separation, thereby benefiting clustering methods. By enhancing feature discrimination, our approach ensures that novel and known classes are more distinctly represented, which aids in accurate clustering and robust recognition.

3. Materials and methods

3.1. Problem statement

In the task of novel class discovery (NCD), two distinct datasets are utilized: a labeled dataset $D^l = \{(x_i^l, y_i^l)\}_{i=1}^N$ and an unlabeled dataset $D^u = \{x_i^u\}_{i=1}^M$. The labeled dataset D^l contains N samples, each associated with a label $y_i^l \in \mathcal{Y}^l = \{1, \dots, C^l\}$, where C^l represents the number of known classes. For simplicity and feasibility, we assume that the number of novel classes C^u is known a priori. Consequently, we denote the category of the unlabeled samples as $y_i^u \in \mathcal{Y}^u = \{1, \dots, C^u\}$. Here, $\mathcal{Y}^l$ refers to the known labeled classes, while $\mathcal{Y}^u$ represents the novel classes that the model aims to discover. Importantly, the known and novel classes are disjoint, such that $\mathcal{Y}^l \cap \mathcal{Y}^u = \emptyset$.

During training, the objective is to leverage the labeled dataset D^l to learn transferable semantic knowledge that facilitates the effective clustering of samples within the unlabeled dataset D^u , thereby enabling the discovery of novel classes. In the testing phase, the goal is to classify images from both the known and newly discovered classes. The comprehensive label set for testing is defined as $\mathcal{Y} = \{1, \dots, C^l, C^l + 1, \dots, C^l + C^u\}$, encompassing all C^l known classes and C^u novel classes.

3.2. Preliminary

We adopt the training strategy proposed in UNO (Fini et al., 2021), which incorporates multi-view exchange prediction and a unified loss function. For labeled data, we employ the standard cross-entropy loss, while for unlabeled data, we utilize the self-labeling loss introduced in Asano et al. (2020). The self-labeling approach assigns pseudo-labels to new category data by solving an optimal transport (OT) problem, using the generated pseudo-labels to train the model. To enforce an equipartition constraint and avoid degenerate solutions (e.g., all data points assigned to a single category), the Sinkhorn–Knopp (SK) algorithm (Cuturi, 2013) is applied. Specifically, the model is trained by minimizing the average cross-entropy loss in a supervised setting:

$$E(p|y_i) = -\frac{1}{B} \sum_{i=1}^B \log p(y_i|x_i), \quad (1)$$

where B is the batch size. For the unsupervised setting, we replace the labels with posterior distributions $q(y|x_i)$:

$$E(p, q) = -\frac{1}{B} \sum_{i=1}^B \sum_{j=1}^{C^u} q(y|x_i) \log p(y|x_i). \quad (2)$$

To prevent degenerate solutions, we introduce an equipartition constraint, ensuring B^u samples are uniformly distributed among the C^u categories, as shown in Eq. (3):

$$\min_{p, q} E(p, q) \quad \text{subject to} \quad \forall y : q(y|x_i) \in \{0, 1\} \quad \text{and} \quad \sum_{i=1}^{B^u} q(y|x_i) = \frac{B^u}{C^u}. \quad (3)$$

In our framework, let $l = p(y^u|x_i^u)$ and $L = [l^1, \dots, l^{B^u}]$ denote the prediction logits matrix, and $\hat{Y} = [\hat{y}^1, \dots, \hat{y}^{B^u}]^T$ represent pseudo-labels estimated by the SK algorithm. Following Cuturi (2013), we relax $\hat{Y}$ to be an element of the transportation polytope, as shown in Eq. (4):

$$\hat{Y} \in \{\hat{Y} \in \mathbb{R}_+^{C^u \times B^u} | \hat{Y} \mathbb{1} = \frac{1}{C^u} \cdot \mathbb{1}, \hat{Y}^T \mathbb{1} = \frac{1}{B^u} \cdot \mathbb{1}\}, \quad (4)$$

where $\mathbb{1}$ are vectors of all ones of the appropriate dimensions. We then utilize the Frobenius dot-product for matrices L and $\hat{Y}$, as shown in Eq. (5):

$$\min \langle \hat{Y}, -\log L \rangle, \quad (5)$$

where $\langle \cdot \rangle$ is the Frobenius dot-product between two matrices and $\log$ is applied element-wise. We recommend readers refer to Asano et al. (2020), Fini et al. (2021) for more details on self-labeling.

Finally, we use $q(y^u|x_i^u)$ as the optimized pseudo-label. Rewriting Eq. (2), we derive our final clustering loss for unlabeled data, as shown in Eq. (6):

$$\mathcal{L}_{cluster}^u = -\frac{1}{B^u} \sum_{i=1}^{B^u} q(y^u|x_i^u) \log p(y^u|x_i^u). \quad (6)$$

Following the UNO framework, the clustering loss and the supervised cross-entropy loss are unified as Eq. (7):

$$\mathcal{L}_{ce}^{unified} = \mathcal{L}_{ce}^l + \mathcal{L}_{cluster}^u. \quad (7)$$

3.3. Framework

The framework consists of two modules: a pre-training module (upper part) for supervised learning and a discovery training module (lower part) for discovering unlabeled classes. Details are illustrated in Fig. 3.

Pre-Training Module: The pre-training module is designed to learn a feature extractor denoted as $\Omega^T : x \rightarrow z \in \mathbb{R}^{C^l}$. The prototype network in this phase is denoted as $\mathbf{P}^{T(S)}$, where the superscript $T(S)$ indicates its role as a component in either the teacher or student model during knowledge distillation. This module aims to establish a robust feature representation for known classes using supervised learning.

Discovery Training Module: The discovery training module adopts a dual-headed structure with a known class classification head ϕ^l and an unlabeled class discovery head ϕ^u . The known class head ϕ^l connects to the prototype network $\mathbf{P}^S$ to receive class spatial relation information. The unlabeled class head ϕ^u employs a self-labeling method for clustering unlabeled classes. Class spatial relation information is derived from the pre-training module and transferred to the discovery training module through KL divergence regularization on ϕ^l .

Data Flow: The pre-training module only processes labeled classes, generating logits through ϕ^l . Further, the feature encoder is updated by calculating the cross-entropy loss, thereby obtaining a teacher model relevant to downstream tasks. The discovery training module simultaneously processes labeled and unlabeled data, producing logits for C^l (labeled classes) and C^u (unlabeled classes) through ϕ^l and ϕ^u , respectively. The output distributions from the two heads are combined to generate the final prediction results.

Prototype Classification Head: We introduce a novel prototype classification head for labeled classes, where each class is represented by a single prototype $\mu \in \{\mu^1, \dots, \mu^{C^l}\}$. In this framework, similarity between samples and prototypes is measured using a combined metric that leverages both Euclidean distance and cosine similarity. For each sample (x, y) , we calculate its soft assignment to the prototype μ , where the distance serves as an indicator of the probability that the sample x_i belongs to the prototype μ^j .

To measure this similarity, we compute the distance between the feature representation of an instance x_i and the corresponding prototype μ^j as shown in Eq. (8):

$$\begin{aligned} \text{Sim}(\Omega(x_i), \mu^j) &= \text{Sim}_e(\Omega(x_i), \mu^j) - \text{Sim}_c(\Omega(x_i), \mu^j), \\ \text{Sim}_e(\Omega(x_i), \mu^j) &= \frac{1}{d} \|\Omega(x_i) - \mu^j\|_2^2, \\ \text{Sim}_c(\Omega(x_i), \mu^j) &= \Omega(x_i) \cdot \mu^j, \end{aligned} \quad (8)$$

where $\Omega(\cdot)$ denotes the feature extractor before the prototype classifier, and Sim_e and Sim_c represent the Euclidean distance-based dissimilarity and cosine similarity, respectively. The squared Euclidean distance is divided by the feature dimension d for normalization, mitigating the impact of high-dimensional feature spaces on the similarity measure.

The rationale behind constructing $\text{Sim}(\cdot, \cdot)$ as the difference between Sim_e and Sim_c lies in the complementary nature of the two metrics. Cosine similarity captures the directional alignment between the feature vector and the prototype, while Euclidean distance accounts for their magnitude difference. Since a smaller Euclidean distance indicates higher similarity and a larger cosine similarity indicates higher similarity, subtracting Sim_e from Sim_c ensures that larger values of $\text{Sim}(\cdot, \cdot)$ consistently represent higher similarity. This unified similarity measure simplifies subsequent processing and enhances the model's ability to distinguish between prototypes in the feature space. Then, we calculate the probability that an instance x_i belongs to prototype μ^j is calculated as shown in Eq. (9):

$$p(x \in \mu^j | x) = \frac{e^{-\gamma \text{Sim}(\Omega(x), \mu^j)}}{\sum_{n=1}^{C^l} e^{-\gamma \text{Sim}(\Omega(x), \mu^n)}}. \quad (9)$$

Then, we can define the cross-entropy (CE) loss as the objective loss of our framework, as shown in Eq. (10):

$$\mathcal{L}_{ce}^{\text{prototype}}(x) = -\sum_{j=1}^{C^l} y_j \log p(x \in \mu^j | x). \quad (10)$$

By optimizing Eq. (10), the embedding features of the same classes are attracted to each other by the prototype centers and vice versa. Consequently, the feature embeddings obtained through the prototype network are distributed more discriminatively in the latent space, exhibiting more compact intra-class and sparser inter-class distributions. In our proposed pre-training framework, we iterate through all samples and calculate the feature average of the corresponding samples by label to initialize the prototype. The model automatically learns the

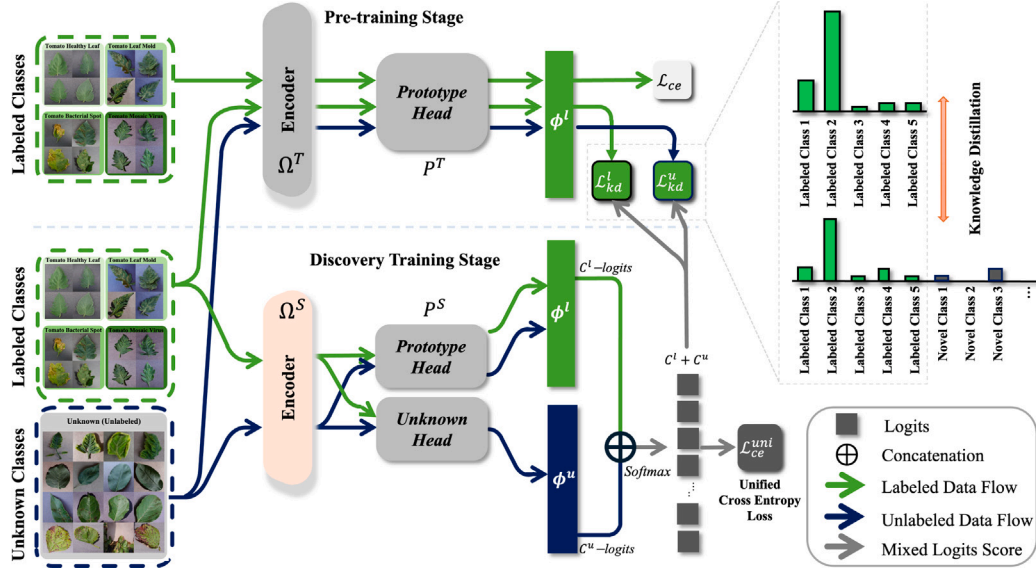


Fig. 3. Overview of our method. First, the teacher model is trained on the labeled dataset, where Ω^T and P^T represent the feature extractor and prototype classification head of the teacher model, respectively. Similarly, Ω^S and P^S denote the feature extractor and prototype classification head of the student model, which are initialized with the parameters of the teacher model. Note that during the discovery training stage, the network from the pre-training stage is fixed.

prototype from the data. The sample category assignment is determined based on distance during inference, making our method more interpretable. Subsequently, we replace the classical cross-entropy loss in the unified loss $\mathcal{L}_{ce}^{unified}$ with our prototype-based cross-entropy loss $\mathcal{L}_{ce}^{prototype}$ and add the knowledge distillation loss $\mathcal{L}_{KD}$. The final objective function is expressed as Eq. (11):

$$\mathcal{L} = \mathcal{L}_{ce}^{prototype} + \mathcal{L}_{cluster}^u + \alpha \mathcal{L}_{KD}, \quad (11)$$

where α is a hyperparameter to balance the knowledge distillation loss.

Knowledge Distillation Loss: The distributions of known classes can be acquired by pre-training the labeled dataset. To measure the relationship between a new class sample and the known classes, we input each new class sample into the supervised pre-training module. This process yields the output distributions of the teacher and student models, as shown in Eqs. (12) and (13):

$$p_u^T = \text{softmax} \left(\frac{P_l^T(\Omega^T(x^u))}{t} \right) \in \mathbb{R}^{c^l}, \quad (12)$$

$$p_u^S = \text{softmax} \left(\frac{P_l^S(\Omega^S(x^u))}{t} \right) \in \mathbb{R}^{c^l}. \quad (13)$$

Here, $\Omega^T(\cdot)$ and $\Omega^S(\cdot)$ represent the feature extractors of the teacher and student models, respectively. The functions $P_l^T(\cdot)$ and $P_l^S(\cdot)$ denote the classifier heads that output logits for the labeled classes. The t is used to soften the probability distributions, providing smoother gradients during training. Both p_u^T and p_u^S are probability distributions over the known class space for the unlabeled sample x^u , enabling the student model to mimic the teacher's predictions.

The knowledge distillation process aims to align the output distribution of the student model with that of the teacher model. By comparing the softened output distributions (Eqs. (12) and (13)), the student model learns to replicate the teacher's predictions for the unlabeled samples. This alignment is quantified using the Kullback-Leibler (KL) divergence, as defined in Eq. (14):

$$\mathcal{L}_{kd}^u = \frac{1}{B} \sum_{i=1}^B KL(p_{u_i}^T \| p_{u_i}^S), \quad (14)$$

where B denotes the batch size. This distillation process enables the student model to capture the distribution of class relationships, thereby facilitating the learning of unknown classes while preserving knowledge about known classes.

One of the benefits of freezing the teacher model parameters is the ability to reduce the influence of unlabeled data on the feature space of labeled data during the discovery phase. For the labeled data, the corresponding loss is defined in Eq. (15):

$$\mathcal{L}_{kd}^l = \frac{1}{B} \sum_{i=1}^B KL(p_{l_i}^T \| p_{l_i}^S), \quad (15)$$

where $p_{l_i}^T$ and $p_{l_i}^S$ are the teacher and student distributions for the labeled samples.

Finally, the overall knowledge distillation loss is computed as a weighted combination of the losses for unknown and known classes, as shown in Eq. (16):

$$\mathcal{L}_{KD} = \eta \mathcal{L}_{kd}^u + (1 - \eta) \mathcal{L}_{kd}^l, \quad (16)$$

where η is a balancing coefficient that controls the trade-off between the distillation losses for known and unknown classes.

3.4. Assumption reasonableness

Novel class discovery aims to leverage the prior knowledge from labeled data to guide the clustering of unlabeled data. Therefore, it is crucial to validate that the prototype classification head can establish category relationships between labeled and unlabeled classes. Figs. 4 and 5 provide empirical support for the hypothesis that a prototype-based classifier is a reasonable approach for this task. As shown in Fig. 4, samples are moved toward their corresponding prototypes to form more compact clusters in the training phase. This ensures that each prototype effectively represents its respective class, creating distinct and well-defined category representations. In the testing phase, input samples are mapped into the latent space. The distances between the input samples and each prototype are calculated. These distances are then converted into a probability distribution through a softmax layer to obtain the final prediction results. This method allows for the efficient classification of new samples based on their proximity to the learned prototypes.

The prototype network's ability to create compact and distinct category representations in the latent space makes it particularly suitable for fine-grained tasks, such as plant disease recognition. By aligning new samples with the closest prototypes, the model leverages the prior knowledge from the labeled data to make informed predictions about

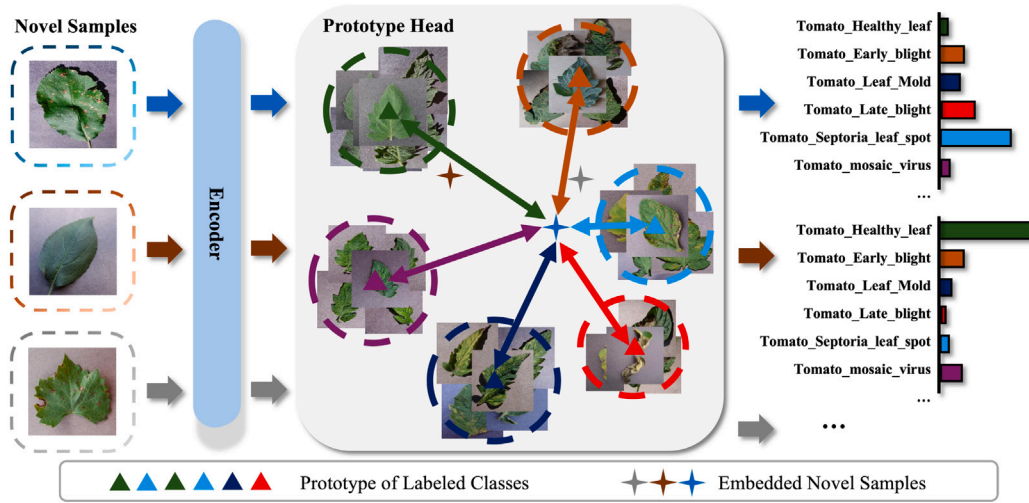


Fig. 4. The illustration of the prototype network. In the training phase, samples are moved toward the corresponding prototype to obtain more compact clusters. In the testing phase, the input samples are mapped into the latent space to calculate the distances from each prototype, respectively. Then, the distance is converted into a probability distribution through the softmax layer to obtain the final prediction result. Images are from the Plant Village dataset.

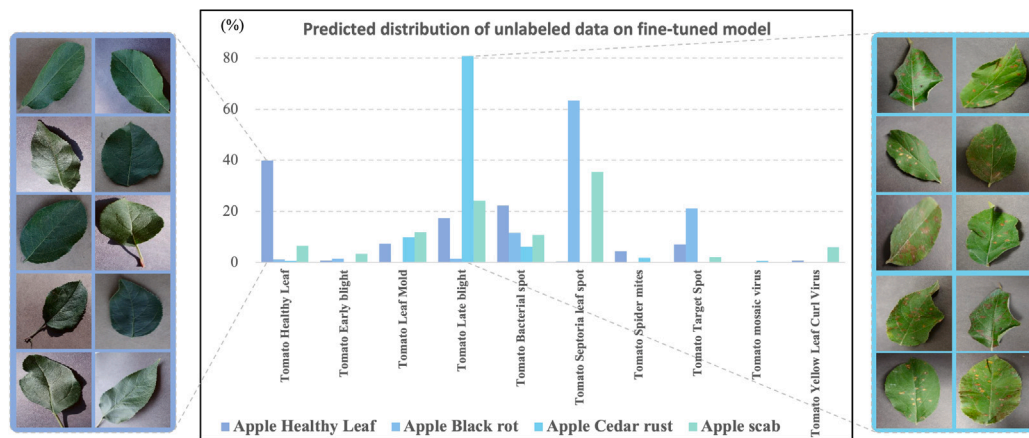


Fig. 5. Predicted distribution of unlabeled apple leaf data on fine-tuned model. The result based on dino-v1 with prototype classification head.

the unlabeled data. We provide the predicted distribution of various unlabeled apple leaf samples when processed by a model fine-tuned on known tomato disease classes, as shown in Fig. 5. On the one hand, the model classifies more than 40% of healthy apple leaves as healthy tomato leaves. This demonstrates the model's ability to generalize the concept of leaf health across different species, suggesting that the learned prototypes for healthy leaves can effectively represent similar health conditions in other species. On another hand, diseased apple leaves are classified into various known tomato disease categories. Although these classifications are not perfect due to the absence of specific apple disease categories in the training set, the model still recognizes disease-related features. For example, more than 80% of apple leaves affected by 'cedar rust' were classified as 'tomato late blight'. This indicates that the learned disease features from the tomato classes are transferable to the apple classes, supporting the idea that disease characteristics share common visual patterns across different species.

These observations validate the hypothesis that there are inherent similarities between known and unknown classes, allowing a prototype-based approach to effectively map unknown samples to the most similar known prototypes. The prototype network leverages these similarities by modeling the spatial relationships in the latent feature space, enabling the classification of new samples based on their proximity to the learned prototypes. This method not only improves the clustering of

new categories but also ensures that the model leverages the underlying structure of the feature space to enhance its generalization capabilities, thereby facilitating the discovery of novel classes in plant disease and species recognition tasks. Overall, these observations support the reasonableness of using a prototype-based approach for NCD.

3.5. Estimating the number of novel classes

Setting an accurate number of novel classes before training often improves clustering accuracy in NCD (Fini et al., 2021; Gu et al., 2023). However, the challenge of estimating the number of novel classes has been largely overlooked in previous studies. In this work, After constructing a compact feature space for labeled datasets using prototype-based classification, the optimal number of clusters can be easily determined using the elbow method (Liu and Deng, 2020). The elbow method essentially identifies the 'elbow point' on the curve of the sum of squared errors (SSE), where the rate of decrease in error slows significantly. SSE measures the distance between each data point and its respective cluster center, reflecting the compactness of the clustering solution. In this work, we propose three elbow-based methods to determine the optimal number of clusters using K-means clustering.

The Max Second Derivative method (Ketchen and Shook, 1996) identifies the point of maximum curvature on the SSE curve by calculating the second derivative, effectively capturing the 'elbow' point

Table 1
Dataset split.

Dataset	Labeled set images			Novel set images		
	Training	Test	Classes	Training	Test	Classes
Herbarium19	17,013	1,335	341	17,212	1,344	342
Plant Village(T/A)	10,785	7,374	10	1,889	1,282	4
Plant Village(T/C)	10,785	7,374	10	2,333	1,519	4
Plant Village(T/G)	10,785	7,374	10	2,428	1,634	4
Plant Village(T/P)	10,785	7,374	10	1,297	855	3
Plant Village(T/H)	10,785	7,374	10	6,111	3,999	7
Plant Village(T/D)	10,785	7,374	10	7,696	5,101	6

where adding more clusters yields diminishing returns in SSE reduction. The Max Distance method (Milligan and Cooper, 1985) calculates the distance of each point on the SSE curve from a straight line connecting the curve's endpoints and selects the point with the greatest distance as the optimal cluster number. The Min Angle method (Thorndike, 1953) identifies the point with the smallest angle between three consecutive points on the SSE curve, indicating a sharp change in the error reduction rate.

In addition to these elbow-based methods, we implement several traditional algorithms for determining the optimal number of clusters. These methods evaluate clustering quality using different validity indices rather than relying directly on the SSE curve. Specifically, the Silhouette Coefficient (Rousseeuw, 1987) assesses both cohesion and separation, achieving its optimal value when intra-cluster distances are minimized and inter-cluster distances are maximized. The Calinski–Harabasz Index (Caliński and Harabasz, 1974) evaluates the ratio of inter-cluster dispersion to intra-cluster dispersion, reaching its maximum when clusters are compact and well-separated. Lastly, the Davies–Bouldin Index (Davies and Bouldin, 1979) calculates the ratio of intra-cluster distances to inter-cluster distances, minimizing when clusters are both tight and well-separated.

4. Experiments

4.1. Implementation details

During the supervised training phase, we trained the models on Herbarium19 and Plant Village datasets for 50 epochs. Subsequently, we conducted joint training for 100 epochs, encompassing both known and new categories, following a methodology similar to UNO (Fini et al., 2021). We utilized an SGD optimizer with a learning rate that starts with linear growth and then undergoes cosine decay. We initialized the backbone network vit-base using dino-v1 (Caron et al., 2021) and dino-v2 (Oquab et al., 2024), with dino-v2 being the primary backbone for the main experimental results. Our approach is implemented based on PyTorch and can be executed on a single GPU device (Nvidia-4090). For a fair comparison, we adopted the same training settings to implement K-means (MacQueen et al., 1967) and Cr-kd (Gu et al., 2023) on these datasets based on their released code, ensuring that all training processes converged. Our visualization methods are based on the ‘umap’ and ‘grad-cam’ libraries.

4.2. Datasets and splits

Novel class discovery has not yet been addressed in plant-related recognition work. Therefore, we first need to construct datasets and set partitioning schemes. This paper involves two significant plant-related themes: species recognition and plant disease identification. We select two representative datasets: the Herbarium19 (Tan et al., 2019) (species recognition) dataset and the Plant Village (Hughes et al., 2015) (plant disease classification) dataset.

The Herbarium19 dataset (Tan et al., 2019) consists of 46,469 high-resolution digital images of herbarium sheets, representing 683 species

from the flowering plant family Melastomataceae, also known as the princess flower family. Challenges associated with this dataset include significant class imbalance and the reliance of expert botanists on subtle features like leaf veins and flower shapes for species identification, which may not be adequately captured by low-resolution imaging methods typically used in image classification. Additionally, there is considerable intraclass variation, where individuals of the same species vary greatly, and minimal interclass variation, making different species visually similar. For Herbarium19, we follow the partitioning strategy of Vaze et al. (2022), dividing 341 classes as labeled categories and the remaining 342 classes as novel categories. We further create subsets of the novel categories based on sample size, selecting the top 25, 50, 100, 200, and 300 classes. These subsets are referred to as Top-25, Top-50, Top-100, Top-200, and Top-300, respectively. The results for these subsets are presented in supplementary discussions rather than the main results.

The Plant Village dataset (Hughes et al., 2015) is a publicly accessible database widely used in agricultural and plant pathology research, especially for identifying and classifying plant diseases. It contains 54,309 images spanning 14 crop species: apple, blueberry, cherry, corn, grape, orange, peach, bell pepper, potato, raspberry, soybean, squash, strawberry, and tomato. The dataset includes images of 17 fungal diseases, four bacterial diseases, two mold diseases, two viral diseases, and one disease caused by a mite. Twelve crop species have images of healthy leaves that are not visibly affected by a disease. We construct different subsets for the Plant Village dataset to divide labeled categories and new classes. The subsets T, A, C, G, P, H, and D in Plant Village correspond to tomato, apple, corn, grape, potato, and other healthy, and other diseased leaves, respectively. Ten classes of tomato diseases serve as labeled categories, while the other subsets are treated as novel categories. More details about the datasets are shown in Table 1.

4.3. Protocol and evaluation metrics

We employ two protocols to evaluate our model: task-aware and task-agnostic. In the task-aware protocol, we follow the transductive learning setting proposed in Gu et al. (2023) to evaluate model performance. Specifically, we assess performance on the **training novel dataset** using ClusterAcc. In this setting, we assume that all obtained unlabeled samples are used for training, and their clustering performance is evaluated. Since the samples are unlabeled, it is reasonable to assess performance on these unlabeled samples. ClusterAcc is defined as:

$$ClusterAcc = \max_{perm \in P} \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{y_i = perm(\hat{y}_i)\}, \quad (17)$$

where y_i and $\hat{y}_i$ denote the ground-truth and predicted labels, respectively, and P represents the set of all permutations. The optimal permutation is calculated using the Hungarian algorithm (Kuhn, 1955).

In the task-agnostic protocol, we follow the task-agnostic setting proposed in Fini et al. (2021) to evaluate the model's performance. Specifically, the accuracy of labeled data is computed using the standard accuracy metric, while the performance of class discovery for unlabeled data is measured using ClusterAcc. In this setting, the evaluation is performed on a **test dataset** containing both known and novel class data. As this protocol does not require prior knowledge about whether a given sample belongs to a known or novel class, it is more suitable for real-world applications compared to the first type of evaluation.

4.4. Main result

The main result is reported in Tables 2 and 3. K-means leverages the prior knowledge of known classes to directly cluster the unknown classes, making it a straightforward solution for novel class discovery.

Table 2

Main results. Note that all methods in this report is based on the **dino-v2** pre-training model. The **bold** indicates the best results. Note that 'Novel' denotes unlabeled classes.

Dataset	Methods	Pre-training	Task-aware (%)	Task-agnostic (%)		
		Accuracy	Novel	All	Novel	Labeled
Herbarium19 (341/342)	K-means	85.18	33.69	52.03	31.77	72.43
	UNO	85.18	33.96	57.20	38.21	76.33
	Cr-kd	85.18	34.71	58.47	39.03	78.05
	Ours	86.43	35.88	59.29	40.18	78.54
Plant Village (T/A)	K-means	99.95	74.01	93.43	57.49	99.67
	UNO	99.95	63.98	93.91	60.57	99.7
	Cr-kd	99.95	63.98	94.00	61.47	99.65
	Ours	99.95	93.70	98.45	91.38	99.68
Plant Village (T/C)	K-means	99.95	69.70	95.61	74.92	99.88
	UNO	99.95	66.95	93.33	63.30	99.52
	Cr-kd	99.95	69.01	94.38	67.54	99.92
	Ours	99.95	90.76	98.19	91.28	99.62
Plant Village (T/G)	K-means	99.95	72.08	94.04	69.46	99.68
	UNO	99.95	91.17	97.88	89.75	99.75
	Cr-kd	99.95	89.11	97.84	88.98	99.87
	Ours	99.95	99.59	99.57	99.48	99.59
Plant Village (T/P)	K-means	99.95	94.76	92.45	65.26	95.64
	UNO	99.95	84.58	96.94	85.03	98.33
	Cr-kd	99.95	86.93	98.22	85.85	99.67
	Ours	99.95	85.39	98.36	86.43	99.76
Plant Village (T/H)	K-means	99.95	42.99	79.85	43.64	99.88
	UNO	99.95	57.01	84.28	56.01	99.90
	Cr-kd	99.95	56.93	83.87	54.89	99.90
	Ours	99.95	67.65	88.06	66.77	99.83
Plant Village (T/D)	K-means	99.95	65.47	79.03	49.21	99.85
	UNO	99.95	66.84	86.00	66.28	99.77
	Cr-kd	99.95	67.02	86.40	67.07	99.90
	Ours	99.95	67.35	87.02	68.75	99.77

Despite its simplicity, K-means shows reasonable performance but often falls short compared to more sophisticated methods. For instance, in the Herbarium19 dataset, K-means achieves a task-aware novel class accuracy of 33.69% and a task-agnostic novel class accuracy of 31.77%, which are significantly lower than the results obtained by our method (35.88% and 40.18%, respectively). Additionally, Cr-kd utilizes knowledge distillation but does not perform as well as our approach in most scenarios. In the Plant Village (A) subset, for example, Cr-kd achieves a task-aware novel class accuracy of 63.98% and a task-agnostic novel class accuracy of 61.47%, while our method reaches 93.70% and 91.38%, respectively.

Our method consistently excels in novel class detection across different subsets of the Plant Village dataset. For instance, in the Plant Village (C) subset, our approach attains a task-aware novel class accuracy of 90.76%, outperforming both K-means (69.70%) and Cr-kd (69.01%). Similarly, in the Plant Village (G) subset, our method achieves a task-agnostic novel class accuracy of 99.48%, significantly higher than K-means (69.46%) and Cr-kd (88.98%).

The task-agnostic evaluation protocol is more reflective of real-world applications than the task-aware protocol, as it demonstrates the model's ability to distinguish between known and unknown classes. In practical scenarios, models must effectively identify whether a sample belongs to a known class or a novel one without prior task-specific information. Our method also consistently shows high performance in task-agnostic settings, such as achieving 86.43% in the Plant Village (P) subset, compared to K-means (65.26%) and Cr-kd (85.85%).

The above analysis is based on the dino-v2 pre-trained model. Notably, all methods in [Tables 2](#) and [3](#) utilize the same pre-trained model to ensure a fair comparison. Consequently, the observed performance improvements primarily stem from our distance-based prototype network and knowledge distillation strategy, rather than differences in pre-training models.

In summary, the results demonstrate that our method consistently outperforms existing approaches in both task-aware and task-agnostic

settings across various datasets, regardless of whether dino-v1 or dino-v2 is used for pre-training. Our approach effectively discovers and classifies novel classes while maintaining high accuracy on labeled data, underscoring its advantages over baseline methods such as K-means and Cr-KD. Furthermore, the task-agnostic evaluation reaffirms our method's practical applicability, highlighting its robustness and strong generalization capabilities in distinguishing between known and unknown classes.

4.5. Ablation study

We conducted an ablation study to investigate the contribution of different components in our framework, as shown in [Table 4](#). This study evaluates the impact of using different knowledge distillation losses ($\mathcal{L}_{kd}^u$ and $\mathcal{L}_{kd}^l$) and classifier types (Linear vs. Prototype) on the model's performance. The evaluations were conducted using the dino-v1 and dino-v2 pre-training models across various metrics, including pre-training accuracy, task-aware, and task-agnostic protocol. Ablation experiments are conducted on the Herbarium19 dataset, a highly complex species recognition dataset characterized by significant challenges such as uneven data distribution, limited sample sizes, and a large number of categories.

From the ablation results, several key observations can be made. Firstly, using a linear classifier with the dino-v1 pre-training model, without any knowledge distillation losses, achieves a pre-training accuracy of 75.43%. However, its performance in the task-aware novel class accuracy and task-agnostic settings is relatively low, with scores of 26.73% and 33.97% for novel classes, respectively. Incorporating $\mathcal{L}_{kd}^u$ alone does not significantly enhance the performance, suggesting that additional regularization is needed to improve the model's ability to handle novel classes.

Switching to a prototype classifier with the dino-v1 model and using only $\mathcal{L}_{kd}^u$ increases the pre-training accuracy to 77.12%. This setup shows a slight improvement in task-aware and task-agnostic evaluations, indicating the benefit of using a prototype-based approach.

Table 3

Main results. Note that all methods in this report is based on the **dino-v1** pre-training model. The **bold** indicates the best results. Note that 'Novel' denotes unlabeled classes.

Dataset	Methods	Pre-training	Task-aware (%)	Task-agnostic (%)		
		Accuracy	Novel	All	Novel	Labeled
Herbarium19 (341/342)	K-means	75.43	23.11	43.26	26.71	59.93
	UNO	75.43	24.58	47.63	32.63	63.13
	Cr-kd	75.43	26.77	48.92	33.74	64.19
	Ours	77.12	27.65	50.63	34.93	66.44
Plant Village (T/A)	K-means	99.80	45.84	94.18	63.81	99.46
	UNO	99.80	60.51	89.93	59.98	99.39
	Cr-kd	99.80	63.95	93.91	61.66	99.51
	Ours	99.80	77.58	96.55	79.13	99.57
Plant Village (T/C)	K-means	99.80	81.31	96.29	79.33	99.80
	UNO	99.80	64.67	92.50	63.33	99.35
	Cr-kd	99.80	66.91	94.10	67.12	99.67
	Ours	99.80	88.64	97.63	88.87	99.44
Plant Village (T/G)	K-means	99.80	62.03	91.35	54.59	99.78
	UNO	99.80	77.22	93.21	75.52	99.29
	Cr-kd	99.80	77.76	95.94	79.16	99.78
	Ours	99.80	85.03	96.24	82.80	99.33
Plant Village (T/P)	K-means	99.80	76.56	81.85	54.15	85.10
	UNO	99.80	81.80	97.95	82.60	99.19
	Cr-kd	99.80	82.69	97.81	81.46	99.73
	Ours	99.80	84.50	98.03	85.61	99.49
Plant Village (T/H)	K-means	99.80	64.95	86.95	63.74	99.78
	UNO	99.80	58.63	85.09	57.82	99.75
	Cr-kd	99.80	57.63	84.31	56.41	99.74
	Ours	99.80	59.79	85.88	60.53	99.90
Plant Village (T/D)	K-means	99.80	59.45	77.43	45.29	99.88
	UNO	99.80	63.95	84.01	63.39	99.40
	Cr-kd	99.80	63.62	85.01	63.84	99.79
	Ours	99.80	68.95	86.85	68.70	99.51

Table 4

The impact of different modules on model performance.

$\mathcal{L}_{kd}^u$	$\mathcal{L}_{kd}^l$	Classifier	Pre-training		Task-aware (%)	Task-agnostic (%)		
			Model	Accuracy	Novel	All	Novel	Labeled
$\times$	$\times$	Linear	dino-v1	75.43	26.73	47.35	33.97	60.82
$\checkmark$	$\times$	Linear	dino-v1	75.43	26.77	48.92	33.74	64.19
$\checkmark$	$\times$	Prototype	dino-v1	77.12	27.88	49.03	34.15	64.01
$\checkmark$	$\checkmark$	Prototype	dino-v1	77.12	27.65	50.63	34.93	66.44
$\checkmark$	$\checkmark$	Prototype	dino-v2	86.43	35.88	59.29	40.18	78.54

Table 5

The result of our methods by using different pre-training models.

Dataset	Pre-training		Task-aware (%)	Task-agnostic (%)		
	Model	Accuracy	Novel	All	Novel	Labeled
Herbarium19 (341/342)	dino-v1	77.12	27.65	50.63	34.93	66.44
	dino-v2	86.43	35.88	59.29	40.18	78.54
Plant Village (T/A)	dino-v1	99.80	77.58	96.55	79.13	99.57
	dino-v2	99.95	93.70	98.45	91.38	99.68
Plant Village (T/C)	dino-v1	99.80	88.64	97.63	88.87	99.44
	dino-v2	99.95	90.76	98.19	91.28	99.62
Plant Village (T/G)	dino-v1	99.80	85.03	96.24	82.80	99.33
	dino-v2	99.95	99.59	99.57	99.48	99.59
Plant Village (T/P)	dino-v1	99.80	84.50	98.03	85.61	99.49
	dino-v2	99.95	85.39	98.36	86.43	99.76
Plant Village (T/H)	dino-v1	99.80	59.79	85.88	60.53	99.90
	dino-v2	99.95	63.44	86.70	62.89	99.86
Plant Village (T/D)	dino-v1	99.80	68.95	86.85	68.70	99.51
	dino-v2	99.95	67.35	87.02	68.75	99.77

When both $\mathcal{L}_{kd}^u$ and $\mathcal{L}_{kd}^l$ are applied, the prototype classifier shows further gains, particularly in the task-agnostic labeled class accuracy, which rises to 66.44%. The result demonstrates the effectiveness of the prototype head on the NCD task.

Using larger-scale pre-trained models can further enhance our methods. Notable improvements were observed when using the dino-v2

pre-training model. This configuration achieves the highest pre-training accuracy of 86.43%. In the task-aware setting, the novel class accuracy reaches 35.88%, indicating a substantial improvement over the other configurations. The task-agnostic performance also shows significant enhancement, with a novel class accuracy of 40.18% and a labeled

class accuracy of 78.54%. These results indicate that our method is compatible with larger-scale pre-trained models.

Table 5 further illustrates the impact of using different pre-training models, dino-v1 and dino-v2, on the performance of our method across various datasets. The results show that using dino-v2 significantly improves the model's performance compared to dino-v1. For instance, in the Plant Village (A) subset, task-aware novel class accuracy jumps from 77.58% with dino-v1 to 93.70% with dino-v2, and task-agnostic novel class accuracy increases from 79.13% to 91.38%. Similar trends are observed across other subsets. These results highlight that our method significantly boosts the model's ability to discover and classify novel classes while maintaining high accuracy on known classes.

4.6. Estimation of novel classes

Table 6 presents the results of estimating the number of novel classes using various methods across two datasets: Herbarium19 and Plant Village. The estimated values and their corresponding absolute errors (in parentheses) are shown for multiple subsets. For Herbarium19, the subsets are represented as Top-25, Top-50, Top-100, Top-200, Top-300, and Top-342, while for Plant Village, the subsets are labeled as (A)-4, (C)-4, (G)-4, (P)-3, (H)-7, and (D)-6, with the numbers indicating the true number of novel classes.

Across the Herbarium19 dataset, the Maximum Distance method consistently outperforms other methods, achieving an absolute error of zero for Top-25, Top-50, and Top-342. It also maintains relatively small errors (e.g., 2 for Top-100 and 5 for Top-200), demonstrating its robustness in handling varying levels of complexity. In comparison, methods like the Second Derivative and Minimum Angle show good performance for smaller subsets but suffer from higher errors as the subset size increases (e.g., 60 for Second Derivative and 30 for Minimum Angle in Top-342). The Silhouette Coefficient, Calinski–Harabasz Index, and Davies–Bouldin Index consistently exhibit larger errors, particularly for larger subsets, indicating their limitations in estimating the number of novel classes in complex datasets.

For the Plant Village dataset, both the Second Derivative and Maximum Distance methods show strong and consistent performance across all subsets, with absolute errors typically at or near zero. Other methods, such as the Silhouette Coefficient and Minimum Angle, tend to overestimate the number of classes, resulting in higher errors. These results highlight the superior reliability of the Maximum Distance and Second Derivative methods in simpler datasets compared to other clustering quality indices.

These results suggest that methods leveraging the geometry of the SSE curve, such as Maximum Distance, are better suited for estimating the number of novel classes in diverse datasets. For example, the Maximum Distance method often yields perfect or near-perfect predictions, particularly in larger datasets. This indicates that applying simple elbow-based methods within our framework can achieve accurate estimation of unknown classes, thereby addressing the limitation of assuming the number of unknown classes is known in NCD.

4.7. Visualization

In this section, we provide the class activation maps from the pre-training stage to explain how the model leverages learned disease concepts to guide the clustering of unknown categories. Fig. 6 illustrates the gradient class attention maps for both labeled and unlabeled classes, highlighting the model's attention regions when processing different types of leaf samples. In the upper section of Fig. 6, each leaf sample from the labeled classes is paired with its corresponding gradient class attention map. These attention maps reveal that the model predominantly focuses on distinct regions of the leaves that exhibit significant features, such as lesions, color changes, and morphological anomalies. The concentrated hotspots in these maps suggest

that the model effectively identifies and utilizes the critical characteristics of labeled classes to make accurate classification decisions. This focused attention indicates the model's capability to leverage the provided labeled data to discern critical features that are instrumental in distinguishing between different known categories. The model achieves classification by focusing on disease regions, as evident in the visualizations of labeled classes.

Conversely, the lower section of Fig. 6 presents the gradient class attention maps for the unlabeled classes. In these cases, the attention is more dispersed and less confident compared to the labeled classes. For example, in the first two instances of the last row, despite significant lesions on the leaves, there are few red high-attention areas. This attention dispersion likely reflects the model's uncertainty and the inherent difficulty in processing novel classes without prior labeled examples. Despite this, the model is somewhat adaptable, indicating its potential to generalize beyond the training data. These visualizations help to understand how our method learns class concepts and establishes relationships between known and unknown categories.

We also visualize the unlabeled samples' feature in the discovery stage by UMAP visualization techniques, as shown in Fig. 7. The analysis of the Plant Village dataset demonstrates the effectiveness of our method in distinguishing novel classes across different subsets: Apple, Corn, Grape, Potato, Healthy, and Other Diseases. For the Apple, Corn, Grape, and Potato subsets, our method produces clusters equal to the number of categories, indicating the model's robustness in identifying subtle differences in disease features and its generalization capability across different crop types. For instance, despite the complex inter-class similarities in the grape subset, our method effectively differentiates between various diseases, showcasing the method's strength in handling fine-grained classification tasks. The Other subsets further corroborate these findings with distinct clusters, reflecting the model's proficiency in recognizing diverse disease symptoms. These visualizations show well-defined clusters and clear boundaries, suggesting that the model accurately captures the variations among different diseases and healthy leaves.

However, we also observe a significant aggregation of samples from different categories in the upper right corner of the healthy subset. This phenomenon is primarily due to the difficulty in establishing category relationships. The NCD problem assumes that knowledge from known categories helps classify unknown ones. However, in our setting, the labeled categories are from the tomato disease subset, which includes one healthy category and nine disease categories. Classifying healthy leaves from seven different species using just one healthy category is challenging. Therefore, the model may need to rely on the outline or texture of the leaves to determine the category of unknown samples, but this information is lacking in our setting.

Overall, the visual and quantitative analyses demonstrate the scalability and robustness of our method. It effectively manages an increasing number of novel classes while maintaining superior performance, making it a robust solution for practical applications in biodiversity and agriculture.

5. Discussion

5.1. Prototype and linear classification head

In this section, we discuss why the prototype classification head is a more effective solution from the perspective of feature distribution. Previously, in Fig. 2, we demonstrated the feature distribution of known classes using a linear classification head during the pre-training stage. Despite appearing compact, these clusters revealed 13 clusters instead of the ten actual known categories, indicating that the decision boundaries created by the linear classification head could lead to smaller intra-class distances. This phenomenon could negatively impact the subsequent discovery phase for clustering novel classes.

Table 6
Result of estimating the number of novel classes.

Method	Herbarium19 (Estimated Value and Absolute Error)					
	Top-25	Top-50	Top-100	Top-200	Top-300	Top-342
Second Derivative	26 (1)	52 (2)	104 (4)	180 (20)	260 (40)	282 (60)
Maximum Distance	25 (0)	50 (0)	102 (2)	205 (5)	290 (10)	342 (0)
Minimum Angle	25 (0)	58 (8)	102 (2)	205 (5)	250 (50)	372 (30)
Silhouette Coefficient	20 (5)	44 (6)	80 (20)	165 (35)	240 (60)	262 (80)
Calinski–Harabasz Index	20 (5)	40 (10)	80 (20)	160 (40)	240 (60)	262 (80)
Davies–Bouldin Index	20 (5)	42 (8)	88 (12)	165 (35)	250 (50)	272 (70)
Method	Plant Village (Estimated Value and Absolute Error)					
	(A)-4	(C)-4	(G)-4	(P)-3	(H)-7	(D)-6
Second Derivative	4 (0)	5 (1)	4 (0)	4 (1)	4 (3)	4 (2)
Maximum Distance	5 (1)	4 (0)	5 (1)	4 (1)	6 (1)	5 (1)
Minimum Angle	8 (4)	5 (1)	8 (4)	8 (5)	8 (1)	7 (1)
Silhouette Coefficient	8 (4)	2 (2)	6 (2)	4 (1)	10 (3)	8 (2)
Calinski–Harabasz Index	3 (1)	2 (2)	2 (2)	2 (1)	3 (4)	2 (4)
Davies–Bouldin Index	2 (2)	4 (0)	5 (1)	2 (1)	9 (2)	7 (1)

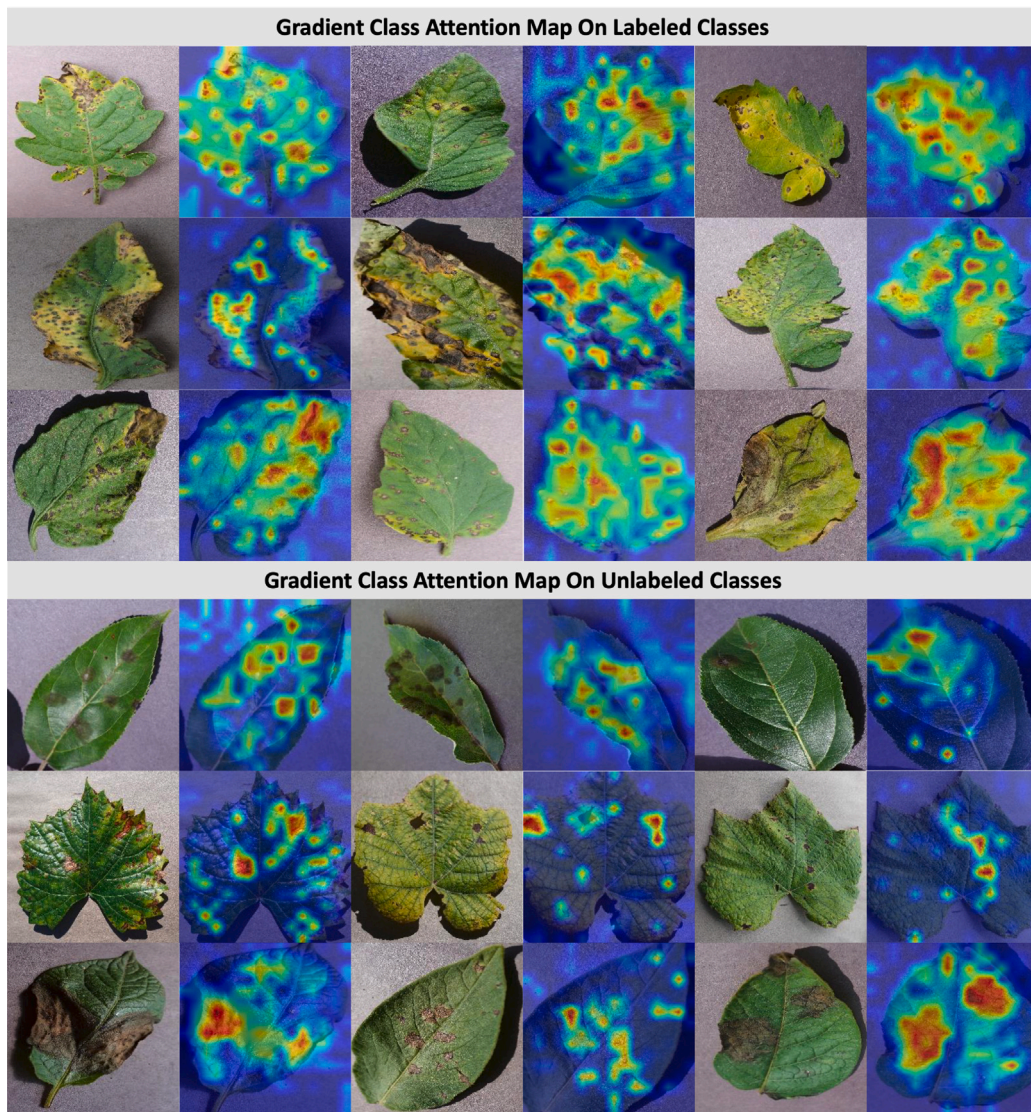


Fig. 6. Gradient class attention map for labeled and unlabeled classes in the pre-training stage. Red areas denote high attention.

In contrast, Fig. 8 showcases the feature distribution when using a prototype classification head for training. For analytical convenience, we focus on the Plant Village dataset to illustrate the effectiveness of

the prototype classification head. The prototype classification head produced ten compact clusters during the pre-training stage, corresponding accurately to the number of known categories. Furthermore, we present

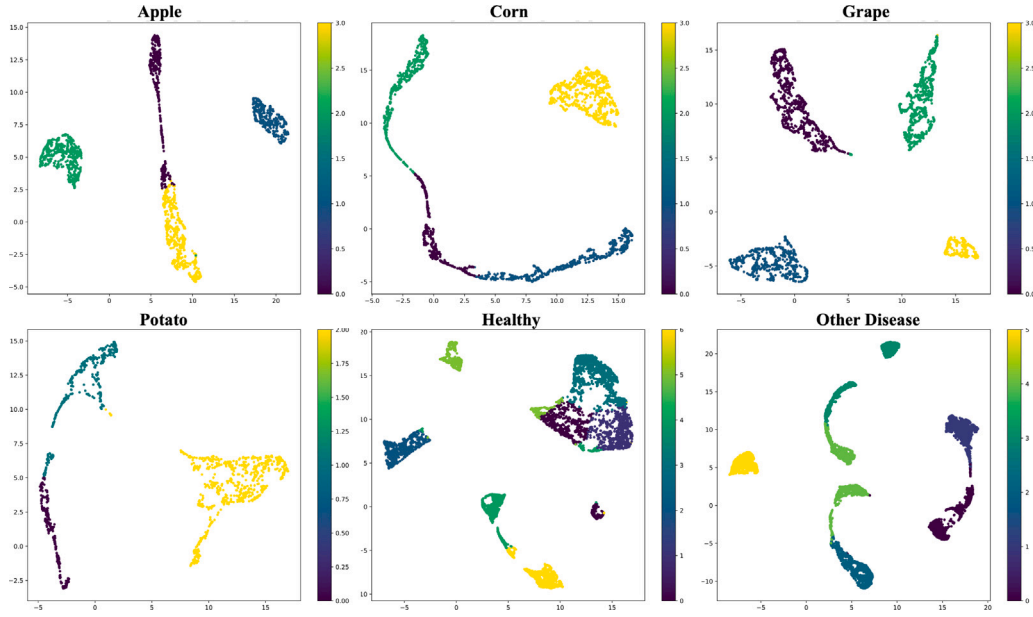


Fig. 7. UMAP visualization for unlabeled samples' feature in the discovery stage. Different colors represent different categories. We present the feature distribution of novel classes in six subsets of the Plant Village dataset.

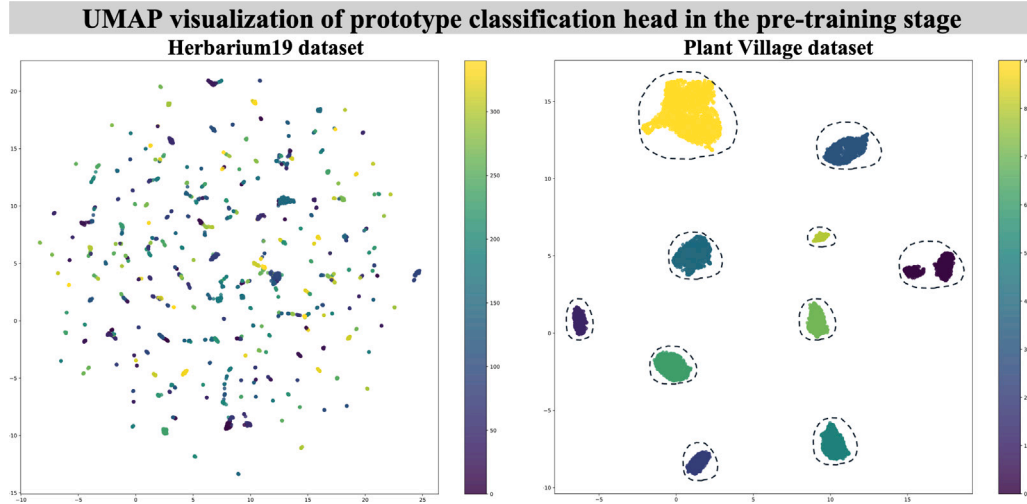


Fig. 8. UMAP visualization of prototype classification head in the pre-training stage. The left and right subplot denote Herbarium19 and Plant Village datasets, respectively. Note that Herbarium19 dataset contains 342 novel classes, each with a relatively small number of samples, which makes the visualization appear cluttered.

how the number of clusters impacts the clustering process during the discovery training stage, as shown in Fig. 9.

The first and second rows of Fig. 9 display the feature representations of novel classes during the discovery training stage using a linear and prototype classification head. We used Apple, Corn, and Grape as novel class subsets. The prototype classification head consistently produced clusters equal to the number of novel categories, aligning with observations from the pre-training stage. However, the linear classification head resulted in more clusters than the actual novel categories. For example, in the Apple disease leaf subset, around 7–8 clusters were generated despite there being only four novel categories. We observed that some samples were closer to clusters of different categories rather than their own, a phenomenon also observed in the Corn and Grape subsets (showing 7 and 6 clusters, respectively). This issue likely arises from suboptimal class representations with a linear classification head during the pre-training stage.

On the other hand, the prototype classification head constructs more precise decision boundaries by leveraging prototype vectors of known

categories, which helps maintain distinct and accurate clusters. This method better captures the semantic relationships between classes, providing higher flexibility and accuracy for classification tasks, especially in complex scenarios involving numerous novel categories. For example, in the Herbarium19 dataset, the prototype classification head also yields well-defined boundaries, as shown in the left of Fig. 8. Although the visualization appears chaotic, distinct decision boundaries and clusters are still discernible. Thus, the prototype classification head proves to be a superior choice for new class discovery tasks, ensuring robustness and reliability in the clustering process. We will discuss more results from the Herbarium19 dataset in the following section.

5.2. The impact of openness

Discussing the impact of openness on clustering performance is of significant importance (Yuan et al., 2022; Scheirer et al., 2012). Therefore, this section explores the effect of openness on clustering performance based on experiments conducted on the Herbarium19

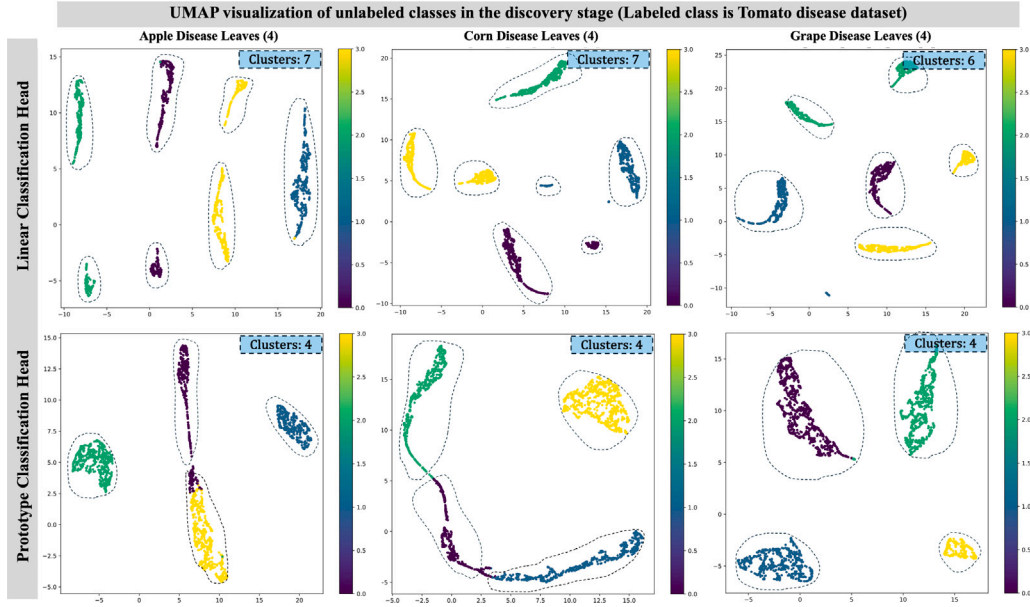


Fig. 9. Comparison of linear and prototype classification head in the discovery training stage. The number of clusters is provided in the top-right corner of each subplot.

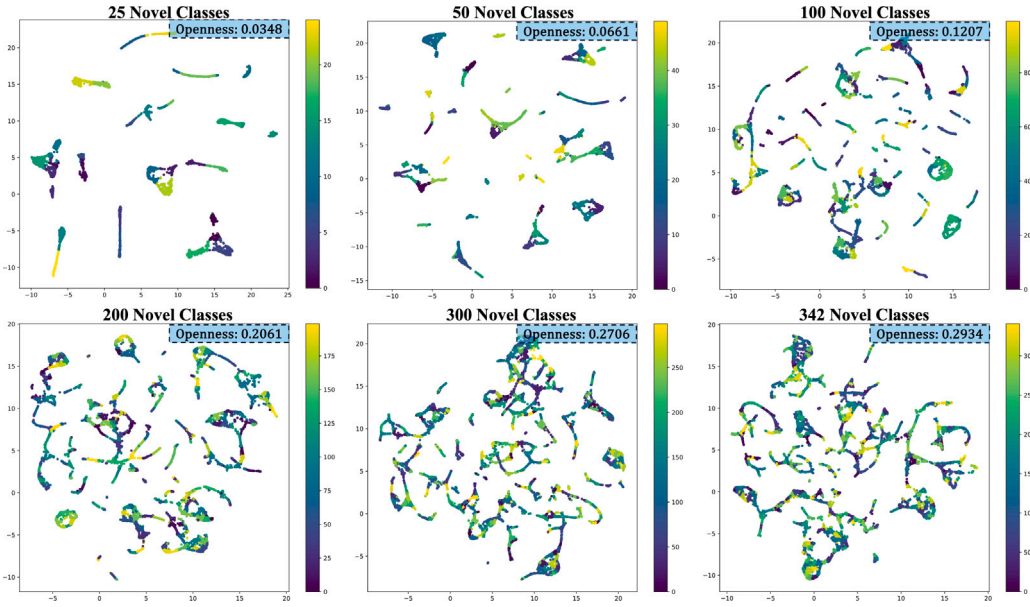


Fig. 10. UMAP visualization on discover stage. We visualized the feature distribution of novel class samples when the number of new classes is 25, 50, 100, 200, 300, and 342, respectively. Note that the pre-training model is dino-v2.

dataset. Following the definition provided in Scheirer et al. (2012), we calculated the openness for different data splits of the Herbarium19 dataset, which is displayed in the top-right corner of each subplot in Fig. 10. We provide a visual representation of the feature distribution of novel class samples using UMAP visualization during the discovery stage, as depicted in Fig. 10. As openness increases, the clustering task becomes significantly more challenging. Specifically, for lower levels of openness, where fewer novel categories are present relative to the known categories, clusters are relatively well-defined and separated. As openness grows and the proportion of novel categories increases, the clusters become more intertwined.

Fig. 11 illustrates the performance trends of three methods – K-means, Cr-KD, and our method – under different levels of openness across various evaluation protocols. The three subplots represent the task-aware novel accuracy, task-agnostic novel accuracy, and task-agnostic overall accuracy, respectively. Under the task-aware protocol,

our method outperforms Cr-KD and K-means as openness increases in most settings. At lower openness levels (e.g., 0.05), all methods achieve relatively high accuracy, with our method and Cr-KD showing similar performance. However, as openness increases, the accuracy of all methods declines due to the increased complexity introduced by more novel classes. K-means perform the worst, reflecting the limitations of the training-free method in handling complex scenarios with higher proportions of novel classes.

The task-agnostic protocol poses a more challenging setting, as no distinction is made between known and unknown classes during evaluation. In this scenario, training-based methods (Ours and Cr-kd) show a significant advantage over K-means. It suggests that training on unlabeled data containing both known and unknown classes is an effective strategy to improve the clustering accuracy of unknown categories, rather than using cluster algorithms like K-means.

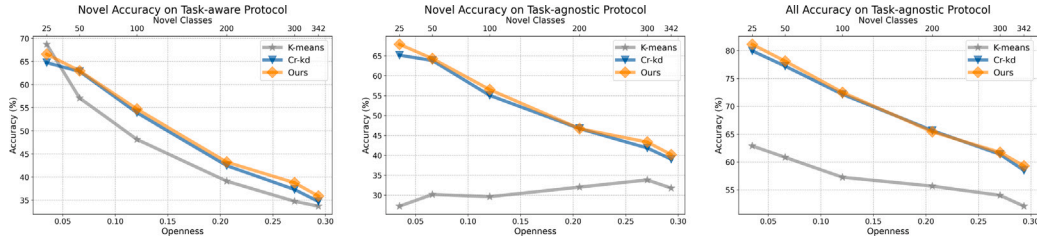


Fig. 11. The performance of three methods under different levels of openness.

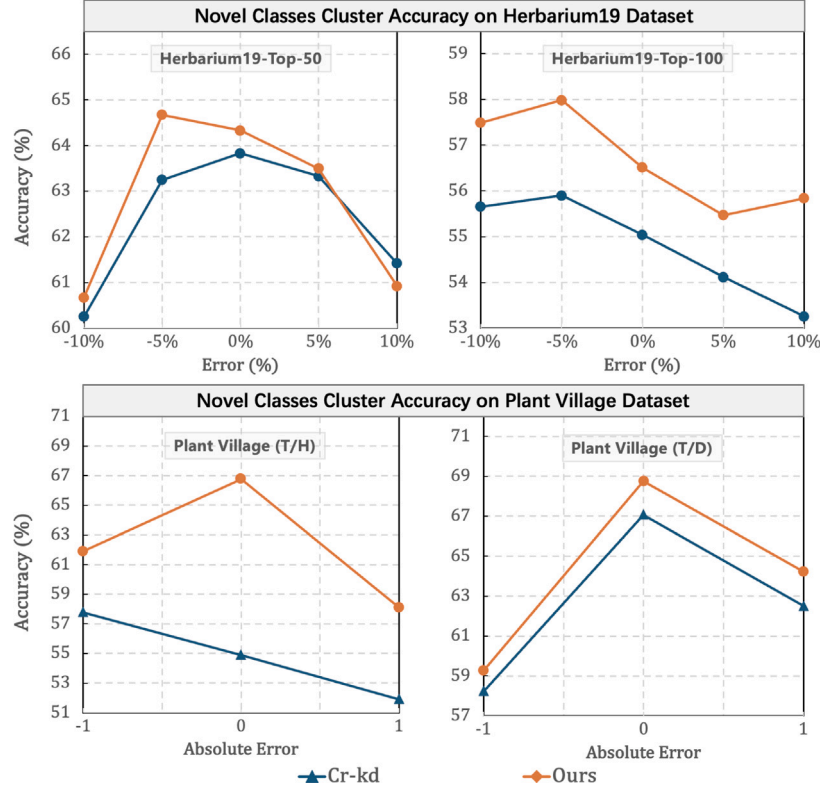


Fig. 12. The x-axis in these plots represents the error rate or absolute error of the estimated number of novel clusters. For the Herbarium19-Top100 dataset, which has 100 novel classes, -10% and 10% denote underestimated and overestimated 10 classes, respectively.

Overall, compared to Cr-KD, our method achieves improvements under different levels of openness, although the gains appear modest (0%–2%). However, given the high complexity of the Herbarium19 dataset, such an improvement is already quite remarkable.

5.3. The number of clusters is unknown

The above experiments assume that the number of novel classes is known. However, this assumption may not hold in practical scenarios. Therefore, to further validate the effectiveness of our method in real-world situations, we conduct experiments under the condition of an unknown number of classes. We assume that the deviation between the estimated and true class numbers ranges from -10% to +10%, which aligns with the results of our class number estimation experiments. For instance, in the case of 100 novel classes, a -10% error rate means that the estimated number of classes is 10 fewer than the true number, while a +10% error rate indicates that the estimated number of classes is 10 more. Since the subsets in the Plant Village dataset contain fewer categories, we use absolute errors instead of percentages. The results are reported on the test novel dataset.

As shown in Fig. 12, our method performs better than the baseline method in most cases. Moreover, the more accurate the class estimation, the higher the clustering accuracy of novel classes tends to be. This

phenomenon is particularly evident in datasets with fewer categories, such as Plant Village. In subsets (T/H) and (T/D), accuracy peaks when the absolute error is 0 and declines significantly as the error increases to ± 1 . Our method demonstrates superior performance, with notable margins over Cr-KD across all error conditions.

We also find that, for the Herbarium19 dataset (Top-50 and Top-100 subsets), clustering accuracy peaks when the number of novel clusters is underestimated (-5% error). Both underestimation (-10%) and overestimation (+10%) result in a noticeable decline in accuracy, with overestimation having a more detrimental impact. We speculate that when the class estimation error is relatively large, multiple novel classes may merge or split, making the relationship between the novel and known classes noisy and not conducive to learning relations.

In summary, Fig. 12 highlights the critical importance of accurate cluster estimation for NCD performance. Our method consistently achieves higher accuracy and demonstrates greater resilience to estimation errors across all datasets and subsets, underscoring its robustness in handling both underestimation and overestimation scenarios. These findings emphasize the practical value of methods that can tolerate inaccuracies in cluster number estimation without significant degradation in performance.

5.4. Limitation

Despite the promising results, our method has several limitations. First, it relies heavily on large-scale pre-trained models like dino-v2, which may not be feasible in resource-constrained environments due to their computational and memory demands. Secondly, we have not extended our validation to other datasets, and doing so may require significant additional work beyond this research's scope. Our observations and motivations for this work are indeed grounded in experiments conducted on plant-related datasets, as this study focuses on addressing challenges specific to this domain. However, we believe the methodologies and principles presented here could potentially inspire adaptations or applications in other domains, and we welcome future studies to explore these possibilities.

6. Conclusion

In this study, we present a framework for novel class discovery in plant species and disease recognition. Our method leverages the strengths of prototype networks and knowledge distillation to effectively balance learning new classes while preserving the recognition ability of known classes. The results show that simple methods like K-means are effective when the number of novel classes is small and under straightforward evaluation protocols such as class-aware settings. However, our method excels in handling more complex and practical open-world scenarios. Extensive experiments have demonstrated the effectiveness of the prototype classifier and knowledge distillation loss in extracting the relationship between novel and known classes. Our method consistently outperforms baseline approaches under class-agnostic protocols, even when the number of novel classes is unknown or inaccurately estimated. Overall, our method provides a powerful and flexible solution for plant-related NCD tasks. These findings pave the way for further advancements in the field. Future work will focus on refining the framework to handle even larger and more diverse datasets and exploring its applicability to other areas of ecological and agricultural research.

CRedit authorship contribution statement

Jiuqing Dong: Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Wei Jin:** Visualization, Software, Methodology, Conceptualization. **Alvaro Fuentes:** Writing – review & editing, Validation, Investigation. **Jaehwan Lee:** Resources, Data curation. **Yongchae Jeong:** Writing – review & editing, Supervision. **Sook Yoon:** Writing – review & editing, Supervision, Resources. **Dong Sun Park:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Ethical and informed consent for data used

This study adheres to ethical guidelines and obtained informed consent for the data used. All data sources and participants involved in the research provided their consent for the collection, analysis, and publication of the data. We confirm that all aspects of our study have been conducted in strict accordance with the relevant guidelines and legislation. This compliance covers institutional protocols as well as national and international regulations about plant research. This declaration affirms our commitment to conducting research in an ethical manner and upholding the principles of informed consent.

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.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Taiwan (No. RS-2019-NR040079); by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT). (RS-2024-00360581); and by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT). (NRF-2021R1A2C1012174).

Data availability

The datasets utilized in this study are publicly available and can be accessed at <https://github.com/spMohanty/PlantVillage-Dataset/tree/master/raw> and <https://www.kaggle.com/c/herbarium-2019-fgvc6/data>. Researchers are encouraged to cite the respective datasets when utilizing them in their work. Please adhere to the terms of use and citation requirements as specified on the respective dataset pages.

References

- Alfarisy, G.A.F., Malik, O.A., Hong, O.W., 2023a. Quad-channel contrastive prototype networks for open-set recognition in domain-specific tasks. *IEEE Access* 11, 48578–48592.
- Alfarisy, G.A.F., Malik, O.A., Hong, O.W., 2023b. Enhancing continual deep open-set recognition with perceptive unknown feature search. In: 2023 6th International Conference on Applied Computational Intelligence in Information Systems. ACIIS, IEEE, pp. 1–6.
- Alfarisy, G.A.F., Malik, O.A., Hong, O.W., 2025. Towards unsupervised domain-specific open-world recognition. *Neurocomputing* 619, 129141.
- Alqahtani, Y., Nawaz, M., Nazir, T., Javed, A., Jeribi, F., Tahir, A., 2023. An improved deep learning approach for localization and recognition of plant leaf diseases. *Expert Syst. Appl.* 230, 120717.
- An, W., Tian, F., Shi, W., Chen, Y., Wu, Y., Wang, Q., Chen, P., 2024. Transfer and alignment network for generalized category discovery. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 10856–10864.
- Asano, Y., Rupprecht, C., Vedaldi, A., 2020. Self-labelling via simultaneous clustering and representation learning. In: International Conference on Learning Representations. pp. 1–13.
- Bharadwaj, R., Naseer, M., Khan, S., Khan, F.S., 2023. Enhancing novel object detection via cooperative foundational models. *CoRR*.
- Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Comm. Statist. Theory Methods* 3 (1), 1–27.
- Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9650–9660.
- Chang, B., Wang, Y., Zhao, X., Li, G., Yuan, P., 2024. A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification. *Expert Syst. Appl.* 237, 121638.
- Cuturi, M., 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Adv. Neural Inf. Process. Syst.* 26.
- Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1* (2), 224–227.
- Dong, J., Fuentes, A., Yoon, S., Kim, H., Jeong, Y., Park, D.S., 2023a. A new deep learning-based dynamic paradigm towards open-world plant disease detection. *Front. Plant Sci.* 14, 1243822.
- Dong, J., Fuentes, A., Yoon, S., Kim, H., Park, D.S., 2023b. An iterative noisy annotation correction model for robust plant disease detection. *Front. Plant Sci.* 14, 1238722.
- Dong, J., Lee, J., Fuentes, A., Xu, M., Yoon, S., Lee, M.H., Park, D.S., 2022. Data-centric annotation analysis for plant disease detection: Strategy, consistency, and performance. *Front. Plant Sci.* 13, 1037655.
- Dong, J., Yao, Y., Jin, W., Zhou, H., Gao, Y., Fang, Z., 2024. Enhancing few-shot out-of-distribution detection with pre-trained model features. *IEEE Trans. Image Process.* 33, 6309–6323.
- Fini, E., Sangineto, E., Lathuilière, S., Zhong, Z., Nabi, M., Ricci, E., 2021. A unified objective for novel class discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9284–9292.
- 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, 758027.
- Gou, J., Yu, B., Maybank, S.J., Tao, D., 2021. Knowledge distillation: A survey. *Int. J. Comput. Vis.* 129 (6), 1789–1819.
- Gu, P., Zhang, C., Xu, R., He, X., 2023. Class-relation knowledge distillation for novel class discovery. In: 2023 IEEE/CVF International Conference on Computer Vision. ICCV, IEEE Computer Society, pp. 16428–16437.

- Han, K., Rebuffi, S., Ehrhardt, S., Vedaldi, A., Zisserman, A., 2020. Automatically discovering and learning new visual categories with ranking statistics. In: *Proceedings of the 8th International Conference on Learning Representations. ICLR 2020, Schloss Dagstuhl-Leibniz-Zentrum für Informatik*, pp. 1–11.
- Han, K., Rebuffi, S.-A., Ehrhardt, S., Vedaldi, A., Zisserman, A., 2021. Autonovel: Automatically discovering and learning novel visual categories. *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (10), 6767–6781.
- Han, K., Vedaldi, A., Zisserman, A., 2019. Learning to discover novel visual categories via deep transfer clustering. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 8401–8409.
- Hayes, T.L., de Souza, C.R., Kim, N., Kim, J., Volpi, R., Larlus, D., 2024. PANDAS: Prototype-based novel class discovery and detection. *CoRR*.
- Hsu, Y.-C., Lv, Z., Kira, Z., 2018. Learning to cluster in order to transfer across domains and tasks. In: *International Conference on Learning Representations*. pp. 1–13.
- Huang, H., Gao, F., Sun, J., Wang, J., Hussain, A., Zhou, H., 2024. Novel category discovery without forgetting for automatic target recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 17, 4408–4420.
- 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*.
- Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F., 2020. A survey on contrastive self-supervised learning. *Technologies* 9 (1), 2.
- Jiang, K., You, J., Dorj, U.-O., Kim, H., Lee, J., 2022. Detection of unknown strawberry diseases based on openmatch and two-head network for continual learning. *Front. Plant Sci.* 13, 989086.
- Jing, M., Zhu, Y., Zang, T., Wang, K., 2023. Contrastive self-supervised learning in recommender systems: A survey. *ACM Trans. Inf. Syst.* 42 (2), 1–39.
- Ketchen, D.J., Shook, C.L., 1996. The application of cluster analysis in strategic management research: an analysis and critique. *Strat. Manag. J.* 17 (6), 441–458.
- Kuhn, H.W., 1955. The hungarian method for the assignment problem. *Nav. Res. Logist. Q.* 2 (1–2), 83–97.
- Liu, Y., Cai, Y., Jia, Q., Qiu, B., Wang, W., Pu, N., 2024. Novel class discovery for ultra-fine-grained visual categorization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 17679–17688.
- Liu, F., Deng, Y., 2020. Determine the number of unknown targets in open world based on elbow method. *IEEE Trans. Fuzzy Syst.* 29 (5), 986–995.
- Liu, J., Wang, Y., Zhang, T., Fan, Y., Yang, Q., Shao, J., 2023. Open-world semi-supervised novel class discovery. *arXiv preprint arXiv:2305.13095*.
- MacQueen, J., et al., 1967. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, Oakland, CA, USA, pp. 281–297.
- Meng, Y., Xu, M., Kim, H., Yoon, S., Jeong, Y., Park, D.S., 2023. Known and unknown class recognition on plant species and diseases. *Comput. Electron. Agric.* 215, 108408.
- Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50, 159–179.
- Nawaz, M., Nazir, T., Javed, A., Amin, S.T., Jeribi, F., Tahir, A., 2024. CoffeeNet: A deep learning approach for coffee plant leaves diseases recognition. *Expert Syst. Appl.* 237, 121481.
- Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., HAZIZA, D., Massa, F., El-Nouby, A., et al., 2024. DINOv2: Learning robust visual features without supervision. *Trans. Mach. Learn. Res.*
- Palei, S., Behera, S.K., Sethy, P.K., 2023. A systematic review of citrus disease perceptions and fruit grading using machine vision. *Procedia Comput. Sci.* 218, 2504–2519.
- Park, W., Kim, D., Lu, Y., Cho, M., 2019. Relational knowledge distillation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 3967–3976.
- Quoc, T.N., Hoang, V.T., 2021. Vnplant-200—a public and large-scale of vietnamese medicinal plant images dataset. In: *Integrated Science in Digital Age 2020*. Springer, pp. 406–411.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.
- Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boulton, T.E., 2012. Toward open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (7), 1757–1772.
- Shen, Y., Hossain, M.Z., Ahmed, K.A., Rahman, S., 2024. An open set model for pest identification. *Comput. Biol. Chem.* 108, 108002.
- Singh, A.K., Rao, A., Chattopadhyay, P., Maurya, R., Singh, L., 2024. Effective plant disease diagnosis using vision transformer trained with leafy-generative adversarial network-generated images. *Expert Syst. Appl.* 124387.
- Tan, K.C., Liu, Y., Ambrose, B., Tulig, M., Belongie, S., 2019. The herbarium challenge 2019 dataset. *arXiv preprint arXiv:1906.05372*.
- Thorndike, R.L., 1953. Who belongs in the family? *Psychometrika* 18 (4), 267–276.
- Vaze, S., Han, K., Vedaldi, A., Zisserman, A., 2022. Generalized category discovery. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7492–7501.
- Wang, L., Yoon, K.-J., 2021. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (6), 3048–3068.
- Xiao, R., Feng, L., Tang, K., Zhao, J., Li, Y., Chen, G., Wang, H., 2024. Targeted representation alignment for open-world semi-supervised learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 23072–23082.
- Xu, M., Kim, H., Yang, J., Fuentes, A., Meng, Y., Yoon, S., Kim, T., Park, D.S., 2023. Embracing limited and imperfect training datasets: opportunities and challenges in plant disease recognition using deep learning. *Front. Plant Sci.* 14, 1225409.
- Yang, H.-M., Zhang, X.-Y., Yin, F., Liu, C.-L., 2018. Robust classification with convolutional prototype learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3474–3482.
- You, J., Jiang, K., Lee, J., 2022. Deep metric learning-based strawberry disease detection with unknowns. *Front. Plant Sci.* 13, 891785.
- Yuan, Y., He, X., Jiang, Z., 2022. Adaptive open domain recognition by coarse-to-fine prototype-based network. *Pattern Recognit.* 128, 108657.
- Zaji, A., Liu, Z., Xiao, G., Bhowmik, P., Sangha, J.S., Ruan, Y., 2023. Wheat spikes height estimation using stereo cameras. *IEEE Trans. AgriFood Electron.* 1 (1), 15–28.
- Zhang, C., Xu, R., He, X., 2023. Novel class discovery for long-tailed recognition. *Trans. Mach. Learn. Res.*
- Zhong, Z., Fini, E., Roy, S., Luo, Z., Ricci, E., Sebe, N., 2021. Neighborhood contrastive learning for novel class discovery. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 10867–10875.
- Zhou, L., Li, N., Ye, M., Zhu, X., Tang, S., 2024. Source-free domain adaptation with class prototype discovery. *Pattern Recognit.* 145, 109974.