Open-world Semi-supervised Novel Class Discovery

Jiaming Liu , **Yangqiming Wang** , **Tongze Zhang** , **Yulu Fan** , **Qinli Yang** and **Junming Shao**[*]

University of Electronic Science and Technology of China

{liujiaming, leo_wang, zhangtongze, ylfan}@std.uestc.edu.cn, {qinli.yang, junmshao}@uestc.edu.cn

Abstract

Traditional semi-supervised learning tasks assume that both labeled and unlabeled data follow the same class distribution, but the realistic open-world scenarios are of more complexity with unknown novel classes mixed in the unlabeled set. Therefore, it is of great challenge to not only recognize samples from known classes but also discover the unknown number of novel classes within the unlabeled data. In this paper, we introduce a new *Open-world Semi-supervised Novel Class Discovery* approach named OpenNCD, a progressive bi-level contrastive learning method over multiple prototypes. The proposed method is composed of two reciprocally enhanced parts. First, a bi-level contrastive learning method is introduced, which maintains the pair-wise similarity of the prototypes and the prototype group levels for better representation learning. Then, a reliable prototype similarity metric is proposed based on the common representing instances. Prototypes with high similarities will be grouped progressively for known class recognition and novel class discovery. Extensive experiments on three image datasets are conducted and the results show the effectiveness of the proposed method in open-world scenarios, especially with scarce known classes and labels.

1 Introduction

Modern deep learning approaches have received widespread attention and progressed rapidly with labeled datasets. Despite the many strengths, most of the approaches are based on a close-world assumption, where the class distributions remain unchanged in the testing phase. But in the realistic open world, the close-world assumptions can barely hold and novel classes are likely to appear in the unlabeled set. For example, in the field of autopilot, the model needs to not only recognize the pre-trained traffic signs (known classes) but also discover the unknown obstacles (novel classes). For cybersecurity, managers need to detect and classify network intrusions as either existing types or unknown numbers of new types

---

[*]Corresponding author
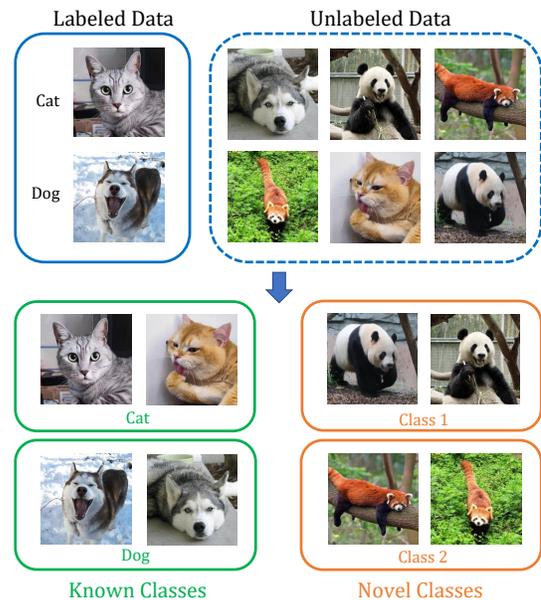
Figure 1: An illustration of the open-world semi-supervised novel class discovery task, where there exist some unknown novel classes in the unlabeled data in semi-supervised learning. The objective is to recognize the samples from known classes while simultaneously discovering the unknown number of novel classes within the unlabeled data.

of attacks. Therefore, as shown in Figure 1, compared to the close-world settings, the open-world scenarios are more complicated with the following challenges.

Challenge 1: *How to recognize the known classes and detect the unknown class samples mixed in the unlabeled dataset?* Traditional semi-supervised learning tasks assume a close-world setting, where the classes in the unlabeled set are the same as the known labeled set. But for open-world scenarios, unknown class samples are mixed in the unlabeled set. To tackle this problem, some robust semi-supervised learning methods [Guo *et al.*, 2020; Huang *et al.*, 2021] are proposed, in which all detected novel samples are simply and crudely regarded as outliers. The recently proposed ORCA [Cao *et al.*, 2022] can simultaneously cluster the different novel classes in the unlabeled set, but with the assumption that the number of novel classes is predefined.

Challenge 2*: How to better cluster the novel classes in the unlabeled set with an extra disjoint labeled dataset?* To address this challenge, some novel class discovery methods [Han *et al.*, 2019b; Han *et al.*, 2019a] are proposed, which try to leverage useful knowledge from an extra labeled set to better cluster the novel classes in the unlabeled data. However, most of them hold the assumption that all samples are from the novel classes in the testing phase. Therefore, these methods do not have the ability to identify known class samples that are mixed with the unknowns in the test set. Recently, GCD [Vaze *et al.*, 2022] generalizes the novel class discovery task to further recognize the known classes in the unlabeled set. But the predictions can only be made in a transductive way, which requires the whole test set.

Challenge 3*: How to estimate the number of unknown classes and match the suitable cluster hierarchy?* Some traditional methods [Pelleg *et al.*, 2000; Hamerly and Elkan, 2003] and recent deep learning methods [Leiber *et al.*, 2021; Ronen *et al.*, 2022] are proposed to estimate the number of clusters in unsupervised clustering tasks. However, it is difficult to determine the clustering level since several different cluster hierarchies often exist in complex datasets. For example, the CIFAR-100 dataset has 100 classes and they can also be grouped into 20 superclasses. For the open-world setting, some labeled samples can be used to better determine the most suitable cluster hierarchy.

In this paper, we aim to address the aforementioned three open-world challenges for the recognition of known classes and the discovery of the arbitrary number of novel classes in the unlabeled data. To this end, a progressive bi-level contrastive learning method over multiple prototypes, named OpenNCD, is proposed, which consists of two reciprocally enhanced parts. First, a bi-level contrastive learning method is introduced to maintain the pair-wise similarity in both prototype and prototype group levels for better representation learning. The involved prototypes and prototype groups can be regarded as representative points for similar instances on fine-level (sub-classes hierarchy) and coarse-level (real-classes hierarchy). Then, a new and reliable similarity metric is proposed based on the Jaccard distances of common representing instances. The most similar prototypes will be grouped together to represent the same class. Finally, the prototype groups are associated with real labeled classes for known class recognition and novel class discovery.

Our contributions are summarized as follows:

- We propose a new approach to simultaneously tackle the three challenges in open-world learning.

- We introduce a bi-level contrastive learning approach to achieve better representation learning.

- We design a novel and reliable approach for novel class discovery by progressively grouping the prototypes.

2 Related Works

2.1 Semi-supervised Learning

Traditional semi-supervised learning methods [Lee and others, 2013; Xie *et al.*, 2020; Berthelot *et al.*, 2019; Sohn *et al.*, 2020; Li *et al.*, 2021] assume a close-set scenario, in which the classes of unlabeled samples are the same as the labeled samples. However, in the realistic open world, unknown novel classes exist and are often mixed in the unlabeled set, which is very likely to bring significant performance degradation to the classification of labeled known classes. To deal with this problem, some robust open-set semi-supervised learning methods [Chen *et al.*, 2020b; Yu *et al.*, 2020; Guo *et al.*, 2020; Huang *et al.*, 2021] are proposed. These methods usually assign the detected novel classes as out-of-distribution samples with low weight to decrease the impacts in the training phase, but they cannot identify different novel classes in the detected novel samples. Recently, ORCA [Cao *et al.*, 2022] proposes a new approach to further cluster the novel classes while recognizing the known classes. However, it assumes that the number of unknown novel classes is predefined, and that is just impossible for open-world scenarios.

2.2 Novel Class Discovery

Novel class discovery tasks aim to cluster the unlabeled novel classes with the help of a similar but disjointed extra labeled set, as proposed in several existing approaches [Han *et al.*, 2019b; Han *et al.*, 2019a; Brbić *et al.*, 2020; Zhong *et al.*, 2021]. However, it is assumed that all samples belong to the novel unknown classes in the testing phase. Therefore, while these approaches are able to discover novel classes in the unlabeled set, they cannot identify the known classes which surely exist in the open-world setting. Recently, GCD [Vaze *et al.*, 2022] generalizes the novel class discovery task to further recognize the known classes in the unlabeled set. It relies on a well-train vision transformer (ViT) model [Dosovitskiy *et al.*, 2020] to extract better visual representation. However, in the testing phase, all testing samples are required for clustering to get the final predictions, which is unable to make predictions directly.

2.3 Class Number Estimation

In the open-world scenario, the number of classes or clusters can not be obtained in advance. To estimate the number of classes during the procedure of clustering, most traditional methods [Pelleg *et al.*, 2000; Hamerly and Elkan, 2003; Kalogeratos and Likas, 2012] first initialize a certain number of clusters, and then apply a criterion to determine whether the clusters should be split or merged in an iteration way. In the field of deep clustering, only a few recent works [Leiber *et al.*, 2021; Ronen *et al.*, 2022] include the approach for cluster number estimation. However, the samples in the dataset can be clustered under different hierarchy levels since one class category can be further divided into several sub-categories. Therefore, it is difficult to choose the thresholds of criterion to get the most suitable cluster hierarchy. In the open-world scenario, one can determine the cluster hierarchy in the unlabeled set by matching it with the labeled data to better estimate the number of classes.

3 Proposed Method

3.1 Problem Definition

Given a dataset consisting of the labeled part $\mathcal{D}_l$ and unlabeled part $\mathcal{D}_u$. We denote $\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^m$, where label y_l
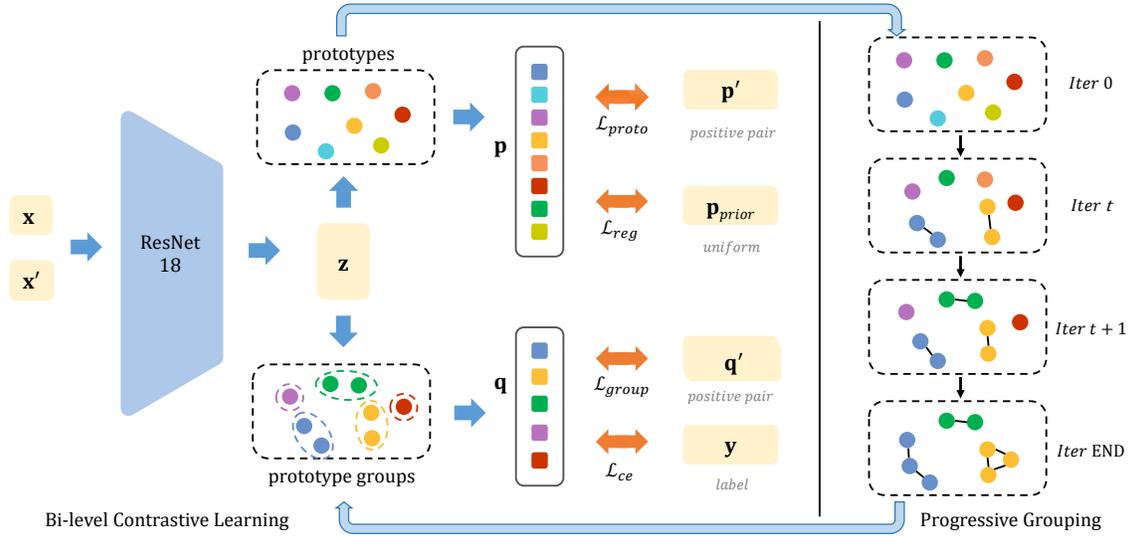
Figure 2: The proposed OpenNCD consists of two reciprocally enhanced parts. First, the bi-level contrastive learning method maintains the similarity in positive pairs x and x' on prototype level $\mathcal{L}_{proto}$ and prototype group level $\mathcal{L}_{group}$ for better representation learning. The prototype prior regularization term $\mathcal{L}_{reg}$ prevents the prototypes from collapsing, while the multi-prototype cross-entropy term $\mathcal{L}_{ce}$ aims to maintain the known class recognition performance. Similar prototypes are grouped progressively as the model's representing ability improves.

is from the set of classes $\mathcal{Y}_l$. We also denote $\mathcal{D}_u = \{(x_i)\}_{i=1}^n$, where the ground-truth label y_u is from the unknown set of classes $\mathcal{Y}_u$. In our open-world setting, the unlabeled set $\mathcal{D}_u$ consists of several novel classes that do not exist in the labeled set, namely, $\mathcal{Y}_l \subset \mathcal{Y}_u$. We consider $\mathcal{Y}_{novel} = \mathcal{Y}_u/\mathcal{Y}_l$ as the set of novel classes. In addition, different from the existing methods that obtain the number of novel classes in advance, in our setting, the number of novel classes $|\mathcal{Y}_{novel}|$ is an unknown value that requires estimation.

3.2 Framework

To handle this challenging task, we introduce a novel progressive bi-level contrastive learning method over multiple prototypes. The multiple prototypes are trained to represent the instances in the feature space. A bi-level semi-supervised contrastive learning approach is designed to learn better representations. A new approach is also proposed to group similar prototypes in a progressive way for unknown novel class discovery.

As shown in Figure 2, the whole framework is composed of a feature extractor f_θ and multiple trainable prototypes $\mathcal{C} = \{\mathbf{c}_1, ..., \mathbf{c}_K\}$ in the feature space. The number of prototypes K is predefined, which far exceeds the number of potential real classes. We also conduct an experiment with the effect of different K. The assignment probabilities from the instances to the prototypes are calculated and further utilized for the bi-level semi-supervised contrastive learning method to learn better representations for class discovery.

To group the unlabeled prototypes and discover the classes, a new metric based on common representing instances is proposed to measure the similarity of two prototypes that belong to the same class. The prototypes are grouped progressively at the training stage. As the representation ability of the encoder networks increases in the iterative training, the instance

from the same class will be more compact and the associated prototypes will also have a higher similarity. In this way, we can find reliable groups and estimate the number of novel classes.

The objective function of the proposed approach includes four components:

$$\mathcal{L} = \mathcal{L}_{proto} + \mathcal{L}_{group} + \lambda_1 \mathcal{L}_{reg} + \lambda_2 \mathcal{L}_{ce} \tag{1}$$

where the first two terms are the bi-level contrastive loss, consisting of the prototype-level similarity loss $\mathcal{L}_{proto}$ and the group-level similarity loss $\mathcal{L}_{group}$. $\mathcal{L}_{reg}$ is the prototype entropy regularization loss and $\mathcal{L}_{ce}$ is the multi-prototype cross-entropy loss.

3.3 Bi-level Contrastive Learning

Prototype-level Similarity. Given an instance feature $\mathbf{z}$ from the feature extractor and a set of prototypes $\mathcal{C} = \{\mathbf{c}_1, ..., \mathbf{c}_K\}$, we denote $\mathbf{p} \in \mathbb{R}^{(m+n) \times K}$ as the assignment probability from $\mathbf{z}$ to the prototype set $\mathcal{C}$ based on the cosine similarity. And the assignment probability from $\mathbf{z}$ to the kth prototype $\mathbf{c}_k$ is given by:

$$\mathbf{p}^{(k)} = \frac{\exp\left(\frac{1}{\tau}\mathbf{z} \cdot \mathbf{c}_k^\top\right)}{\sum_{\mathbf{c}_{k'} \in \mathcal{C}} \exp\left(\frac{1}{\tau}\mathbf{z} \cdot \mathbf{c}_{k'}^\top\right)}, \tag{2}$$

where τ is the temperature scale, and both $\mathbf{z}$ and $\mathbf{c}$ are l_2 normalized.

The pair-wise semi-supervised contrastive learning technique is used for representation learning. Each instance in a training batch is treated as an anchor point. The positive point is obtained either by selecting another random instance from the same class if it is labeled or by selecting its nearest neighbor instance if it is unlabeled. Then the augmented view of the anchor point is chosen to calculate the similarity

with the selected positive point to create more discrepancy for contrastive learning.

To make the two instances in a positive pair exhibit similar assignment probabilities, we define the prototype-level similarity loss as:

$$\mathcal{L}_{proto} = -\frac{1}{m+n} \sum_{i}^{m+n} \log\langle \mathbf{p}_i, \mathbf{p}'_i \rangle, \qquad (3)$$

where p_i and p'_i denote the assignment probability of the anchor point and its positive point, respectively. $\langle \cdot \rangle$ represents the cosine distance. Note that we directly utilize the instances in the same training mini-batch as positive samples to update the encoder and prototypes, instead of sampling from the entire dataset for negative samples. In this way, the proposed model can be trained and updated online for large-scale datasets.

Group-Level Similarity. The above prototype-level similarity loss makes the instances represented by the fine-grained prototypes. To learn better representation for class identification, the instances should also be represented on the more coarse-grained level to match the real class hierarchy. To this end, we introduce a group-level similarity loss by representing the instances on the level of prototype groups.

Formally, a prototype group $\mathcal{C}_g$ is a set of similar prototypes that are very likely to represent the same class, where $\mathcal{C}_g \subset \mathcal{C}$ and any two prototype groups are disjoint. The prototypes are grouped in a progressive way for class discovery, and the detailed grouping process is illustrated in Section 3.4.

Then, the assignment probability from instance feature $\mathbf{z}$ to prototype group $\mathcal{C}_g$ can be obtained by:

$$\mathbf{q}_i^{(g)} = \frac{\sum_{\mathbf{c}_k \in \mathcal{C}_g} \exp\left(\frac{1}{\tau}\mathbf{z} \cdot \mathbf{c}_k^\top\right)}{\sum_{\mathbf{c}_{k'} \in \mathcal{C}} \exp\left(\frac{1}{\tau}\mathbf{z} \cdot \mathbf{c}_{k'}^\top\right)}. \qquad (4)$$

And the group-level similarity loss can be denoted as:

$$\mathcal{L}_{group} = -\frac{1}{m+n} \sum_{i}^{m+n} (\mathbf{q}'_i \log \mathbf{q}_i + \mathbf{q}_i \log \mathbf{q}'_i), \qquad (5)$$

where q_i and q'_i denote the assignment probability of two instances in a positive pair, which can be regarded as the pseudo label for each other on the group level reciprocally. Note that here we utilize another form of contrastive similarity loss as Equation 3 to extract knowledge from a different perspective to prevent overfitting. Moreover, Equation 5 has the same form as cross-entropy, which makes the model focus more on group or class discrimination.

Prototype Regularization. Since multiple prototypes are utilized to represent the feature distribution, some prototypes or groups with low assignment probabilities might be ignored during the optimization. More seriously, all of the instances are assigned to the same prototype in some cases, leading to model collapse. To solve this problem, we introduce a prototype regularization term to regularize the marginal assignment probability on prototypes $\mathbf{p}_{proto} \in \mathbb{R}^K$ to be close to a balance prior $\mathbf{p}_{prior}$ by Kullback-Leibler (KL) divergence, which is given by:

$$\mathcal{L}_{reg} = KL\left(\mathbf{p}_{proto} \parallel \mathbf{p}_{prior}\right), \qquad (6)$$

where

$$\mathbf{p}_{proto} = \frac{1}{m+n} \sum_{i}^{m+n} \mathbf{p}_i, \qquad (7)$$

We expect to prevent the possible collapsing when all instances are assigned to the same single prototype group, and meanwhile, ensure that all the prototypes are utilized to represent distinct characteristics of the complex class distribution. Therefore, we design the prior $\mathbf{p}_{prior}$ to bring a uniform distribution among the prototype groups and also among the prototypes in one group, which is denoted as:

$$\mathbf{p}_{prior}^{(k)} = \frac{1}{\mathcal{N}_g \times |\mathcal{C}_k|}, \qquad (8)$$

where k denotes the prior of the kth prototype $\mathbf{c}_k$, $\mathcal{N}_g$ is the number of all prototype groups in the current stage, and $|\mathcal{C}_k|$ is the number of prototypes of the group which $\mathbf{c}_k$ belongs to.

Multi-Prototype Cross Entropy. The above objectives mainly focus on unsupervised representation learning. To further improve the capability of known class recognition, the supervised multi-prototype cross-entropy loss is introduced.

Note that the prototype groups are unlabeled in the progressive grouping stage, we hope that the prototype groups can be matched to the real class levels where each prototype group represents one class. To this end, we utilize the Hungarian algorithm [Kuhn, 1955] to match the known classes with the prototype groups. In this way, for the labeled instances, we obtain their ground truth label on the prototype groups. Then, the supervised multi-prototype cross-entropy loss is defined as:

$$\mathcal{L}_{ce} = -\frac{1}{m} \sum_{i}^{m} (\log \mathbf{q}_i^{(y)}), \qquad (9)$$

where $\mathbf{q}_i^{(y)}$ is the assignment probability on the ground-truth group, which is given by:

$$\mathbf{q}_i^{(y)} = \frac{\sum_{\mathbf{c}_k \in \mathcal{C}_y} \exp\left(\frac{1}{\tau}\mathbf{z}_i \cdot \mathbf{c}_k^\top\right)}{\sum_{\mathbf{c}_{k'} \in \mathcal{C}} \exp\left(\frac{1}{\tau}\mathbf{z}_i \cdot \mathbf{c}_{k'}^\top\right)}. \qquad (10)$$

Note that Equation 9 has a similar form as Equation 5. The difference is that Equation 5 regards the group assignment probability of the other view as the pseudo label while Equation 9 utilizes the ground-truth labels directly.

3.4 From Prototypes to Classes: Progressive Prototype Grouping

In this section, we elaborate on the process of prototype grouping with the aim of progressively discovering the unknown number of novel classes.

First, a novel similarity metric is proposed to judge the similarity between two prototypes, which is shown in Figure 3. Specifically, for an instance feature $\mathbf{z}$, its assignment probability $\mathbf{p}$ is sorted on all prototypes. The prototypes with top κ largest assignment probabilities are regarded as the associated prototypes of $\mathbf{z}$, and in turn, $\mathbf{z}$ is regarded as a representing instance of prototype $\mathbf{c}$. The set of all representing instances of prototype $\mathbf{c}$ is denoted as:

$$\Gamma(\mathbf{c}) = \{\mathbf{z}_i | (\mathbf{c} \in \text{Top}\kappa(\mathbf{q}_i))\}. \qquad (11)$$
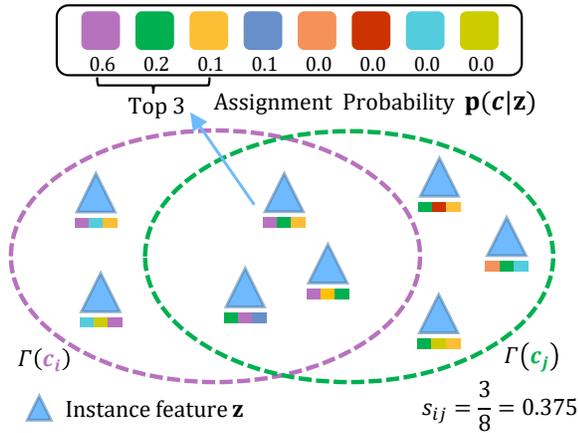
Figure 3: Similarity metric. In this example, each instance has 3 associated prototypes (with top 3 assignment probabilities over all prototypes). The similarity score is computed by Jaccard distance over the representing instance sets of two prototypes.

Intuitively, two prototypes are more likely to belong to the same class if they have more common representing instances. Therefore, we calculate the similarity score of two prototypes by the Jaccard distance over their representing instance sets, given by:

$$s_{ij} = \frac{|\Gamma(\mathbf{c}_i) \cap \Gamma(\mathbf{c}_j)|}{|\Gamma(\mathbf{c}_i) \cup \Gamma(\mathbf{c}_j)|}. \tag{12}$$

At each epoch, the similarity between each two prototypes is calculated to form an affinity matrix. Some graph-based clustering methods can then be utilized to detect the densely connected prototypes, which are regarded as prototype groups. A simple way is to find the prototype groups by linked prototypes, where the similarities are over a certain threshold δ. Some approaches that do not require a predefined number of classes can also work, such as Louvain [Blondel *et al.*, 2008] and affinity propagation [Frey and Dueck, 2007]. To ensure the novel classes are clustered into the same hierarchical level as the known classes, the threshold δ (or the parameter to control the clustering hierarchy) is adjusted automatically by achieving the highest accuracy on the labeled known class samples. With the increasing representation ability of the feature extractor and prototypes, more reliable prototype groups can be obtained gradually.

4 Experiments

4.1 Experimental Setup

Datasets. Our proposed approach[1] is evaluated on three widely-used datasets in the standard image classification tasks including CIFAR-10 [Krizhevsky, 2009], CIFAR-100 [Krizhevsky, 2009], and ImageNet-100, which is randomly sub-sampled with 100 classes from ImageNet [Deng *et al.*, 2009] for its large volume. For each dataset, the first 50% classes are regarded as known and the rest 50% as unknown. Further, only 10% of the known classes are labeled, with the unknown and the rest of the known all unlabeled.

---

[1]Code and appendix at https://github.com/LiuJMzzZ/OpenNCD

Evaluation Metrics. Following the evaluation protocol in [Cao *et al.*, 2022], we evaluate performance on both the known and the novel classes. For known classes, the classification accuracy is measured in the testing phase. For unknown classes, the clustering accuracy is calculated by solving the prediction-target class assignment based on the Hungarian algorithm [Kuhn, 1955]. Moreover, we also calculate the clustering accuracy on all known and novel class samples to measure the overall performance of the proposed model.

Implementation Details. In the proposed method, Resnet-18 is used as the backbone of the feature extractor, which is pre-trained by SimCLR [Chen *et al.*, 2020a] in an unsupervised way. To avoid overfitting, we fix the parameters in the first three blocks of the backbone and only finetune the last block. 50 prototypes are utilized for the CIFAR-10 dataset and 500 for both the CIFAR-100 and ImageNet-100 datasets with a fixed dimension of 32. We adopt an Adam optimizer with a learning rate of 0.002 and fix the batch size to 512 in all experiments. The temperature scale τ is set to 0.1 suggested by most of the previous methods, and the weight of the last two terms in the objective function is set to $\{\lambda_1, \lambda_2\} = \{1, 1\}$. κ is set to 5 in prototype grouping.

4.2 Comparison with Baselines

Baselines. Our proposed method is compared with two baseline methods ORCA [Cao *et al.*, 2022] and GCD [Vaze *et al.*, 2022]. ORCA is proposed for the open-world semi-supervised learning problem with a predefined number of novel classes. For consistency, our experiment settings are following the same protocol of ORCA from its public code. GCD utilizes a more sophisticated pre-trained backbone which exhibits high performance during initialization, and we replace it in the public code with the Resnet-18 network same as in our settings for a fair comparison. Note that in the original paper of ORCA, the ratio of labeled data of the known classes is set to 50%, which is a relatively high ratio for semi-supervised learning. Here we consider comparing in a tougher and general scenario, where only 10% known class data are labeled. We also report the result with the labeled ratio of 50% as an additional result in our appendix[1].

Extended Baselines. We also compare the recent novel class discovery methods and robust semi-supervised learning methods by extending them into our open-world settings. Novel class discovery methods can only cluster novel classes but cannot classify the unlabeled samples to the known classes. Here, we compare two novel class discovery methods: DTC [Han *et al.*, 2019b] and RankStats [Han *et al.*, 2019a]. To extend them for known class recognition, we regard the known class samples as unknown and detect them the same way as normal unknown classes. Then we report the accuracy by using the Hungarian algorithm for label matching. The traditional semi-supervised learning methods cannot discover novel classes mixed in the unlabeled data. We select two methods: FixMatch [Sohn *et al.*, 2020] and DS3L [Guo *et al.*, 2020], and extend them for novel class discovery. The samples with lower confidence scores or weights are selected as unknown class samples, which are clustered by k-means, and the clustering accuracy is reported.

Methods	CIFAR-10			CIFAR-100			ImageNet-100		
	Known	**Novel**	**All**	**Known**	**Novel**	**All**	**Known**	**Novel**	**All**
FixMatch†	64.3	49.4†	47.3	30.9	18.5†	15.3	60.9	33.7†	30.2
DS^{3}L†	70.5	46.6†	43.5	33.7	15.8†	15.1	64.3	28.1†	25.9
DTC*	42.7*	31.8	32.4	22.1*	10.5	13.7	24.5*	17.8	19.3
RankStats*	71.4*	63.9	66.7	20.4*	16.7	17.8	41.2*	26.8	37.4
SimCLR*	44.9*	48.0	47.7	26.0*	28.8	26.5	42.9*	41.6	41.5
ORCA	82.8	85.5	84.1	52.5	31.8	38.6	83.9	60.5	69.7
GCD	78.4	79.7	79.1	49.7	27.6	38.0	82.3	58.3	68.2
OpenNCD	**83.5**	**86.7**	**85.3**	**53.6**	**33.0**	**41.2**	**84.0**	**65.8**	**73.2**

Table 1: Accuracy comparison of known, novel, and all classes. The dataset is composed of 50% known classes and 50% novel classes, with only 10% of known classes labeled. Asterisk (∗) denotes that the original method cannot recognize known classes, which are extended by matching the discovered clusters to the classes in the labeled data. Dagger (†) denotes that the original method cannot discover novel classes, which are extended by performing clustering on the out-of-distribution samples.

Methods	PredNum	Acc	NMI
X-means	13.8 ± 6.3	38.3 ± 3.1	37.2 ± 9.3
G-means	28.2 ± 2.1	35.8 ± 0.1	50.8 ± 0.4
DipDeck	8.16 ± 0.8	66.8 ± 4.7	67.0 ± 1.3
Ours(0% label)	$\mathbf{9.4 \pm 1.1}$	$\mathbf{78.3 \pm 1.6}$	$\mathbf{71.2 \pm 2.1}$
Ours(10% label)	9.8 ± 0.4	86.2 ± 0.5	79.6 ± 0.6

Table 2: Class number estimation in the unlabelled data. We report the results on the CIFAR-10 dataset averaged by 5 runs.

Methods	Seen	Novel	All
w/o $\mathcal{L}_{reg}$	11.9	14.2	29.3
w/o $\mathcal{L}_{ce}$	27.9	25.6	23.5
w/o $\mathcal{L}_{group}$	44.0	26.0	33.5
w/o $\mathcal{L}_{proto}$	50.3	31.7	39.7
OpenNCD	**53.6**	**33.0**	**41.2**

Table 3: Ablation analysis on the components of the objective function. We report the results on the CIFAR-100 dataset with 50% known and 50% novel and 10% of the known labeled.

Comparison Results. The results of all compared methods are presented in Table 1, including the accuracy of known classes, novel classes, and all classes. As an additional baseline, we also run k-means directly on the output features of the encoder which is pre-trained by SimCLR and include the obtained results without extra training. Since most baselines cannot deal with the class number estimation task, we assume the number of novel classes is known for all methods. For the extended method, the "all" scores are not the average of the "known" and "novel" scores, because the marked scores (∗ or †) are calculated in an extended and different way. The results in Table 1 demonstrate that simultaneously recognizing known classes and clustering novel classes in the open-world setting is difficult for traditional methods. However, our proposed method can handle these complex tasks effectively with better performance than baselines.

4.3 Estimating the Number of Classes

The previous experiments assume that the number of novel classes is known, so we further conduct an experiment under the scenario where the real number of novel classes is unknown. To this end, the proposed method is compared with some traditional class number estimation methods including X-means [Pelleg *et al.*, 2000], G-means [Hamerly and Elkan, 2003], and a recent deep clustering method DipDeck [Leiber *et al.*, 2021]. For the above three methods, we extract the features of the CIFAR-10 dataset from the pre-trained backbone, which is the same as in our method. Since these three methods are all unsupervised, we report the results of our method in two settings including (a) trained with no labeled data, as in

the three methods, and (b) trained with 50% known and 50% novel and 10% of the known labeled, as in the previous experiments. The results in Table 2 demonstrate that our proposed method can better deal with the class number estimation task even without the labeled data.

4.4 Ablation and Analysis

Ablation Study. In Table 3, the contributions of different parts of the loss functions in our proposed approach are analyzed, including the prototype-level similarly $\mathcal{L}_{proto}$, the group-level similarity $\mathcal{L}_{group}$, the multi-prototype cross entropy $\mathcal{L}_{ce}$ and the prototype regularization $\mathcal{L}_{reg}$. To investigate the importance of these terms, the ablation study is conducted by removing each term separately from the objective function. We can infer from Table 3 that all the components contribute to our proposed method. Moreover, the result with $\mathcal{L}_{group}$ removed demonstrates the importance of grouping, and the result with $\mathcal{L}_{proto}$ removed proves the benefit of multiple prototypes in representation learning.

Effect of the Novel Class Ratio. As shown in Figure 5, we evaluate the performance on the CIFAR-10 dataset with the ratio of novel classes ranging from 0.1 to 0.9. We report the accuracy of all classes by unsupervised clustering accuracy, which can be considered as a good proxy for the overall performance. Despite the inevitable performance degradation due to the increase in the novel class ratio, OpenNCD still performs better than the two strong baselines, especially at higher novel class ratios. The result in Figure 5 indicates that our proposed approach can better face task scenarios with higher openness.
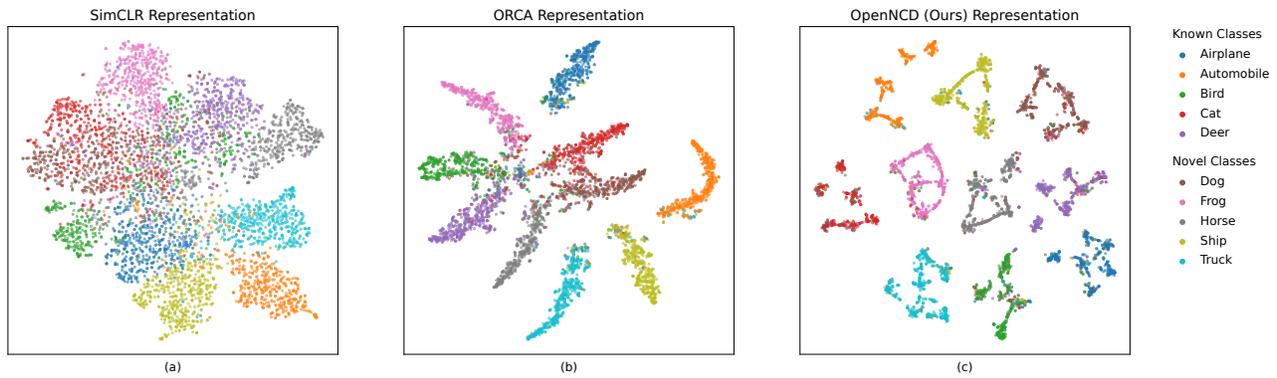
Figure 4: T-SNE visualization of learned feature representation for CIFAR-10 with 50% known classes (10% labeled) and 50% novel classes on (a) pre-trained ResNet-18 by SimCLR, (b) ORCA (Baseline), and (c) OpenNCD (Ours) with 50 prototypes. Colors represent classes.
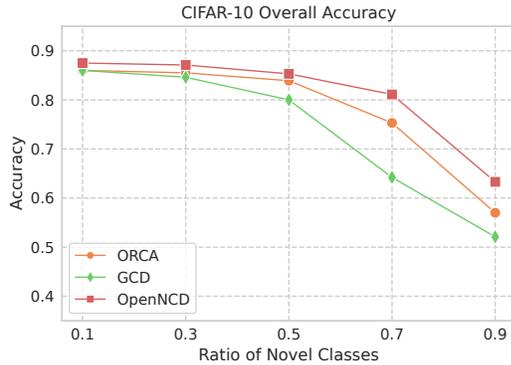
Figure 5: Performance of ORCA, GCD, and the proposed method on the CIFAR-10 dataset with different novel class ratios.
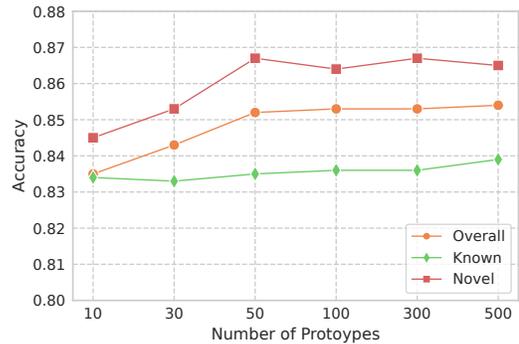
Figure 6: Performance on the CIFAR-10 dataset with number of prototypes varying from 10 (single prototype for each class) to 300.

Benefits of Multiple Prototypes. To investigate the effect of the number of prototypes K, we conducted experiments by increasing it from the actual number of classes (10) to a large number (500) on the CIFAR-10 dataset. The results are shown in Figure 6. It can be observed that using multiple prototypes rather than a single one, to represent the distribution of each class, improves the performance of both known and novel classes. Moreover, representing class distributions using a single prototype requires prior knowledge of the number of new classes. When the number of new classes is unknown, multiple prototypes are employed to represent the class distribution, which will be grouped progressively to discover novel classes. Further increasing the number of prototypes does not result in continuous improvement in representation learning, and the performance tends to stabilize. We can still infer that more prototypes would be beneficial when the data distribution is more complex and harder to approximate.

Visualization of the Feature Space. In Figure 4, we show the learned latent space of the raw pre-trained ResNet-18, ORCA, and the proposed OpenNCD on the CIFAR-10 dataset. The high-dimensional latent features are reduced to 2D by T-SNE [Van der Maaten and Hinton, 2008]. As shown in Figure 4(a), the raw ResNet-18 features are mixed with each other and do not form clear clusters. In the latent space of ORCA shown in Figure 4(b), the instance features are more

separated, but they are of irregular shapes and do not cluster compactly enough. Comparatively, Figure 4(c) shows the latent space of the proposed OpenNCD, where the features are represented by 50 prototypes and 10 prototype groups represented by different colors. We can see the prototypes are linked in each group and features distribute closely around the associated prototypes and groups, forming very compact and clear clusters, which benefits our progressive prototype grouping approach for discovering novel classes.

5 Conclusion

In this paper, we tackle the three challenges simultaneously in the open-world setting, including known class recognition, novel class discovery, and class number estimation. To this end, we propose a novel method named OpenNCD with two reciprocally enhanced parts. First, a bi-level contrastive learning method maintains the pair-wise similarity of the prototypes and the prototype group levels for better representation learning. Then, a progressive prototype grouping method based on a novel reliable similarity metric groups the prototypes, which are further associated with real labeled classes for novel class discovery. We conduct extensive experiments and the results demonstrate that our proposed method can deal with the challenges effectively in the open-world setting and outperform the previous methods.

Acknowledgments

References

[Berthelot *et al.*, 2019] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019.

[Blondel *et al.*, 2008] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[Brbić *et al.*, 2020] Maria Brbić, Marinka Zitnik, Sheng Wang, Angela O Pisco, Russ B Altman, Spyros Darmanis, and Jure Leskovec. Mars: discovering novel cell types across heterogeneous single-cell experiments. *Nature methods*, 17(12):1200–1206, 2020.

[Cao *et al.*, 2022] Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-world semi-supervised learning. In *International Conference on Learning Representations*, 2022.

[Chen *et al.*, 2020a] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[Chen *et al.*, 2020b] Yanbei Chen, Xiatian Zhu, Wei Li, and Shaogang Gong. Semi-supervised learning under class distribution mismatch. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3569–3576, 2020.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[Dosovitskiy *et al.*, 2020] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[Frey and Dueck, 2007] Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.

[Guo *et al.*, 2020] Lan-Zhe Guo, Zhen-Yu Zhang, Yuan Jiang, Yu-Feng Li, and Zhi-Hua Zhou. Safe deep semi-supervised learning for unseen-class unlabeled data. In *International Conference on Machine Learning*, pages 3897–3906. PMLR, 2020.

[Hamerly and Elkan, 2003] Greg Hamerly and Charles Elkan. Learning the k in k-means. *Advances in neural information processing systems*, 16, 2003.

[Han *et al.*, 2019a] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically discovering and learning new visual categories with ranking statistics. In *International Conference on Learning Representations*, 2019.

[Han *et al.*, 2019b] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8401–8409, 2019.

[Huang *et al.*, 2021] Junkai Huang, Chaowei Fang, Weikai Chen, Zhenhua Chai, Xiaolin Wei, Pengxu Wei, Liang Lin, and Guanbin Li. Trash to treasure: Harvesting ood data with cross-modal matching for open-set semi-supervised learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8310–8319, 2021.

[Kalogeratos and Likas, 2012] Argyris Kalogeratos and Aristidis Likas. Dip-means: an incremental clustering method for estimating the number of clusters. *Advances in neural information processing systems*, 25, 2012.

[Krizhevsky, 2009] Alex Krizhevsky. Learning multiple layers of features from tiny images. *Master's thesis, University of Toronto*, 2009.

[Kuhn, 1955] Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[Lee and others, 2013] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013.

[Leiber *et al.*, 2021] Collin Leiber, Lena GM Bauer, Benjamin Schelling, Christian Böhm, and Claudia Plant. Dip-based deep embedded clustering with k-estimation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 903–913, 2021.

[Li *et al.*, 2021] Junnan Li, Caiming Xiong, and Steven CH Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9475–9484, 2021.

[Pelleg *et al.*, 2000] Dan Pelleg, Andrew W Moore, et al. X-means: Extending k-means with efficient estimation of the number of clusters. In *Icml*, volume 1, pages 727–734, 2000.

[Ronen *et al.*, 2022] Meitar Ronen, Shahaf E Finder, and Oren Freifeld. Deepdpm: Deep clustering with an unknown number of clusters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9861–9870, 2022.

[Sohn *et al.*, 2020] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel,

Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[Vaze *et al.*, 2022] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.

[Xie *et al.*, 2020] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *Advances in Neural Information Processing Systems*, 33:6256–6268, 2020.

[Yu *et al.*, 2020] Qing Yu, Daiki Ikami, Go Irie, and Kiyoharu Aizawa. Multi-task curriculum framework for open-set semi-supervised learning. In *European Conference on Computer Vision*, pages 438–454. Springer, 2020.

[Zhong *et al.*, 2021] Zhun Zhong, Linchao Zhu, Zhiming Luo, Shaozi Li, Yi Yang, and Nicu Sebe. Openmix: Reviving known knowledge for discovering novel visual categories in an open world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9462–9470, 2021.