Few-shot Novel Category Discovery

Chunming Li[1], **Shidong Wang**[2], **Haofeng Zhang**[1*]
[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, China.
[2]School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom.
{chunmingli, zhanghf}@njust.edu.cn, shidong.wang@newcastle.ac.uk

Abstract

The recently proposed Novel Category Discovery (NCD) adapt paradigm of transductive learning hinders its application in more real-world scenarios. In fact, few labeled data in part of new categories can well alleviate this burden, which coincides with the ease that people can label few of new category data. Therefore, this paper presents a new setting in which a trained agent is able to flexibly switch between the tasks of identifying examples of known (labelled) classes and clustering novel (completely unlabeled) classes as the number of query examples increases by leveraging knowledge learned from only a few (handful) support examples. Drawing inspiration from the discovery of novel categories using prior-based clustering algorithms, we introduce a novel framework that further relaxes its assumptions to the real-world open set level by unifying the concept of model adaptability in few-shot learning. We refer to this setting as Few-Shot Novel Category Discovery (FSNCD) and propose Semi-supervised Hierarchical Clustering (SHC) and Uncertainty-aware K-means Clustering (UKC) to examine the model's reasoning capabilities. Extensive experiments and detailed analysis on five commonly used datasets demonstrate that our methods can achieve leading performance levels across different task settings and scenarios. Code is available at: https://github.com/Ashengl/FSNCD.

1 Introduction

Most deep learning methods [He *et al.*, 2016; He *et al.*, 2022; Dosovitskiy *et al.*, 2021] aim to excel in supervised learning, where models assign labels to test samples using knowledge from training. However, this approach poorly reflects real-world scenarios, where unlabelled data may include both known and novel classes. Few-shot Learning (FSL) [Finn *et al.*, 2017; Sung *et al.*, 2018; Fei-Fei *et al.*, 2006; Vinyals *et al.*, 2016; Li *et al.*, 2023] addresses this by mimicking human ability to learn from limited examples. Despite its potential, FSL has a critical drawback: if not all categories are annotated, it forces classification of unseen classes into existing ones, contradicting open-world reasoning.

The advent of Novel Category Discovery (NCD) [Rizve *et al.*, 2022] and Generalized Category Discovery (GCD) [Vaze *et al.*, 2022a] lifts the learning agent to a new level, where the novel examples are allowed to be grouped into a few clusters based on their inherent characteristics. Notwithstanding, the key of these two tasks will degenerate into solving a cross-domain clustering problem if unlabelled data are not presented. A recent research, On-the-fly Category Discovery (OCD) [Du *et al.*, 2023] was proposed to mitigate the limitation of transductive learning involved in NCD and GCD by determining novel classes according to the values derived from decoupling features into symbolic features and absolute features. However, the over-relaxed constraints imposed on the hash encoding for novel classes lead to a phenomenon in which a higher number of clusters formed than initial expectation.

To enhance the adaptability of learning agent to real-world scenarios, we adapt the assumptions and training strategies in NCD. Similar to FSL, we ensure no overlap between training and test label spaces, and the test set's query label space extends beyond the support set to include novel categories. Unlike semi-supervised strategies in NCD [Rizve *et al.*, 2022] and GCD [Vaze *et al.*, 2022a], we adopt episodic meta-learning to better evaluate the model's adaptability to new tasks.

We call this new problem Few-Shot Novel Category Discovery (FSNCD), and it is intersected conception at the FSL [Fei-Fei *et al.*, 2006] and NCD [Rizve *et al.*, 2022], empowering a learning agent with the capability to be inductively adaptive to the discovery of both novel classes and clusters. The key setting of this problem is illustrated in Fig. 1, and it builds upon the existing setting of NCD [Rizve *et al.*, 2022] which was introduced to cluster novel classes by hints of both labelled and unlabelled data. This scenario assumes that a learning agent can both extract features from labeled data and learn patterns from unlabeled data. Our proposed setting challenges this by restricting access to unlabeled data, focusing instead on enhancing the model's adaptability and plasticity to improve generalization.

FSNCD leverages a DINO [Caron *et al.*, 2021] pre-trained ViT-B [Dosovitskiy *et al.*, 2021] model, differing from tra-

*Corresponding Author.

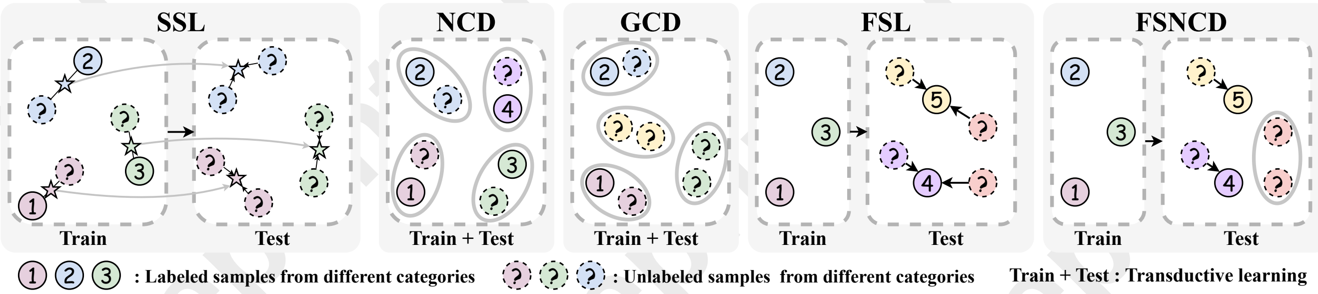

Figure 1: Comparison of Different Task Settings. Solid lines represent labeled data, dashed lines represent unlabeled data, with different colors indicating different categories. Unlabeled data in test phase are samples to be classified. Inductive learning and transductive learning are annotated as “Train/Test” and “Train+Test”. Please note that during the generalization testing phase in inductive learning, FSL and FSNCD are based on a small amount of labeled data.

ditional few-shot learning by tackling novel category discovery, which involves clustering novel classes rather than simple detection. We introduce a hierarchical clustering method with a stopping criterion based on inter-class distances of visible classes, addressing bias from excessive novel categories through balanced sparsity. Additionally, an improved K-means algorithm incorporates uncertainty clustering to handle distributions. Unlike existing methods reliant on prior category counts, our approach restructures K-means and hierarchical clustering to flexibly estimate category numbers in few-shot settings, facilitating the discovery of new categories.

To sum up, our contributions are as follows:

- We propose a new task setting named Few-Shot Novel Category Discovery (FSNCD). It unifies few-shot learning and Novel Category Discovery, realising a selectivity between tasks of identifying examples from labelled classes in a real-time inference manner and clustering novel classes as the query examples increase.
- We introduce two clustering algorithms as baseline methods: Semi-supervised Hierarchical Clustering (SHC) and Uncertainty-aware K-means Clustering (UKC). Both approaches take into account the specific features required to address the complexities of open-set recognition, extending beyond conventional data clustering techniques.
- We conducted experiments on five datasets, analyzing and elucidating the viability of the proposed new setting.

2 Related Work

2.1 Semi-Supervised Learning

In traditional supervised learning, we need a large amount of labeled data to train the model [Chapelle *et al.*, 2009; Yang *et al.*, 2023]. To address the problem of insufficient labeled samples, semi-supervised learning is proposed to utilize labeled and unlabeled data to improve model performance with limited labeled data [Grandvalet and Bengio, 2004; Zhou, 2008; Jiang *et al.*, 2024; Yang *et al.*, 2024]. The most intuitive approach is perhaps Pseudo-Labels self-training [Lee and others, 2013], where a model trained on labeled data generates categorical pseudo-labels for the unlabeled examples. SSL techniques often utilize unlabeled data in various ways, such as employing strategies like consistency regularization [Tarvainen and Valpola, 2017], generative model based methods [Madani *et al.*, 2018], or graph based method [Zhu *et al.*, 2023].

2.2 Novel Category Discovery

Novel Category Discovery (NCD) addresses limited labeled data by leveraging unlabeled samples for automatic classification. Pioneered by DTC [Rizve *et al.*, 2022], the field has expanded with methods like AutoNovel [Han *et al.*, 2020], which combines self-supervised pretraining with pseudo-label generation via ranking. Recent extensions include Novel Category Discovery in semantic segmentation [Zhao *et al.*, 2022], incremental learning frameworks [Liu and Tuytelaars, 2023; Roy *et al.*, 2022], and Generalized Category Discovery (GCD) [Vaze *et al.*, 2022a] that jointly identifies known/novel categories. SimGCD [Wen *et al.*, 2023] enhances GCD with parametric classification, while OCD [Du *et al.*, 2023] introduces real-time hash coding for efficient novel class detection. These advances demonstrate NCD’s growing versatility in handling diverse real-world scenarios.

2.3 Few-Shot Learning

Few-shot learning (FSL) is a method used to address the issue of limited sample size [Fei-Fei *et al.*, 2006; Finn *et al.*, 2017; Snell *et al.*, 2017; Xu *et al.*, 2024; Zhong *et al.*, 2023]. It aims to learn an effective classification model from a few labeled training examples. There are two main types of FSL methods: meta-learning-based [Finn *et al.*, 2017; Ravi and Larochelle, 2016; Tian and Xie, 2023] and metric-learning-based approaches [Snell *et al.*, 2017; Koch *et al.*, 2015]. Additionally, some methods employ a Non-episodic-based strategy [Chen *et al.*, 2019; Dhillon *et al.*, 2020], i.e., fine-tuning a pre-trained model at test time. Although current methods for Few-shot Open-set Recognition [Liu *et al.*, 2020; Jeong *et al.*, 2021] are capable of detecting out-of-distribution samples, they lack the ability to cluster or classify these out-of-distribution samples.

Method	Training	Test	Transferring Manner	Support	Query
SSL	$\mathcal{Y}^{\text{L}} \cup \mathcal{Y}^{\text{L}\dagger}$	$\mathcal{Y}^{\text{L}}$	Inductive	-	-
NCD	$\mathcal{Y}^{\text{L}} \cup \mathcal{Y}^{\text{U}\dagger}$	-	Transductive	-	-
GCD	$\mathcal{Y}^{\text{L}} \cup \mathcal{Y}^{\text{U}\dagger} \cup \mathcal{Y}^{\text{L}\dagger}$	-	Transductive	-	-
FSL	$\mathcal{Y}^{\text{L}}$	$\mathcal{Y}^{\text{U}}$	Inductive	$\mathcal{Y}^{\text{S}}$	$\mathcal{Y}^{\text{S}}$
FSNCD	$\mathcal{Y}^{\text{L}}$	$\mathcal{Y}^{\text{U}}$	Inductive	$\mathcal{Y}^{\text{S}}$	$\mathcal{Y}^{\text{S}} \cup \mathcal{Y}^{\text{N}}$

Table 1: Comparison between different tasks. †denotes providing images during the training phase without accompanying labels.

3 Few-Shot Novel Category Discovery

3.1 Problem Definition

NCD: The training data for NCD is provided in two distinct sets, including a labelled set $\mathcal{D}^{\text{L}} = \{(x_i^l, y_i^l)\}_{i=1}^{N}$ and an unlabeled set $\mathcal{D}^{\text{U}} = \{x_i^u\}_{i=1}^{M}$, where an instance is denoted as x_i^l or x_i^u depending on which set the data is derived from, $y_i^l \in \mathcal{Y}^{\text{L}} = \{1, \ldots, C^{\text{L}}\}$ is the corresponding class label of $\mathcal{D}^{\text{L}}$ and $\mathcal{Y}^{\text{U}}$ is the label space of $\mathcal{D}^{\text{U}}$. The goal is to train a model on both $\mathcal{D}^{\text{L}}$ and $\mathcal{D}^{\text{U}}$, and then aggregate $\mathcal{D}^{\text{U}}$ into C^{U} clusters that are associated with $\mathcal{Y}^{\text{U}} = \{1, \ldots, C^{\text{U}}\}$ where $\mathcal{Y}^{\text{L}} \cap \mathcal{Y}^{\text{U}} = \emptyset$. GCD further relaxes the assumptions of NCD, with an emphasis on identifying both labelled data and novel categories, where the label space of $\mathcal{D}^{\text{U}}$ changes to $\mathcal{Y}^{\text{L}} \cup \mathcal{Y}^{\text{U}}$ instead of $\mathcal{Y}^{\text{U}}$.

FSL: In FSL, a dataset D is usually partitioned into two sets, the training set containing base classes $\mathcal{D}^{\text{Base}}$ and the test set containing novel classes $\mathcal{D}^{\text{Novel}}$. It creates N training tasks $\mathcal{T} = \{T_1, \ldots, T_N\}$ to assemble a finite set of training episodes where each training task $\mathcal{T}_n$ consists of different classes, represented by $\mathcal{T}_n =< \mathcal{S}, \mathcal{Q} >$, where $\mathcal{S}$ and $\mathcal{Q}$ denote examples chosen from the support set and query set, respectively. During testing, the model receives new tasks with the novel classes which have no overlap with those encountered during training.

FSNCD: We approach Few-Shot Novel Category Discovery as a scenario where the dataset is organized under the setting defined in FSL with training set containing base classes $\mathcal{D}^{\text{Base}}$ and the test set containing novel classes $\mathcal{D}^{\text{Novel}}$. However, an episode in FSNCD contains a query set $\mathcal{Q} = \{(x^q, y^q)\} \in \mathcal{X} \times \mathcal{Y}^{\text{Q}}$, and a support set $\mathcal{S} = \{(x_l^s, y_l^s)\}_{l=1}^{NK} \in \mathcal{X} \times \mathcal{Y}^{\text{S}}$ of K image-label pairs for each N classes, referring to an N-way K-shot setting. However, the label space in FSNCD is defined as $\mathcal{Y}^{\text{Q}} = \mathcal{Y}^{\text{S}} \cup \mathcal{Y}^{\text{N}}$ rather than $\mathcal{Y}^{\text{Q}} = \mathcal{Y}^{\text{S}}$ as defined in standard FSL. The primary objective for the learning agent therefore is to identify examples of categories in $\mathcal{Y}^{\text{S}}$ while discovering unknown categories in $\mathcal{Y}^{\text{N}}$.

The characteristics of the proposed setting are summarised in Table 1 and Fig. 1. SSL is an inductive approach leveraging both labeled and unlabeled data within the same label space for test set classification. GCD relaxes the closed-set assumption, allowing the label space of unlabeled data to encompass that of labeled data, following a transductive paradigm. Transitioning from NCD and GCD to FSL introduces FSNCD, which combines few-shot classification with the discovery of novel categories. A key distinction between FSNCD and traditional FSL is that in the testing phase of FSNCD, certain samples may not match any of the support categories. Unlike Few-Shot Open Set Recognition (FSOSR), which only detects outliers, FSNCD requires models to detect and accurately cluster these outliers. Additionally, in contrast to open-world recognition, categories appearing in testing phase of FSNCD are not seen during training, though a limited number of samples are provided during testing.

3.2 Training of FSNCD

The training procedure for a learning agent in FSNCD is divided into two consecutive phases: representation learning and classifier construction. For the former phase, the learning model is dedicated to training a feature extractor $\mathcal{F}$ and a projection head ϕ, $\phi \circ \mathcal{F} : \mathcal{X} \to \mathcal{Z}$, with $\mathcal{Z}$ signifying the embedding space generated from input images $\mathcal{X}$ via $\phi \circ \mathcal{F}$. The latter phase is responsible for constructing a classifier capable of recognizing query examples and discovering novel categories. An overview of our approach is demonstrated in Fig. 2. A detailed explanation of each phase will be presented in the sequel below.

Phase 1: Representation Learning Given that the training process only involves the use of labelled data, the supervised contrastive learning loss [Khosla *et al.*, 2020] is employed to learn more distinctive features for samples belonging to the same category and different categories within each batch. This mitigates the impact of using standard cross-entropy loss for prototype matching which is classification-prone rather than concentrating on the more critical clustering task, and avoids the occurrence of the visible-class bias described in ProtoNet [Snell *et al.*, 2017]. Specifically, given an input image $\mathbf{x}$, it uses the feature extractor $\mathcal{F}(\mathbf{x})$, together with a multilayer perceptron (MLP) projection ϕ to extract the feature embeddings $\mathbf{z}$, denoted as $\mathbf{z} = \phi(\mathcal{F}(\mathbf{x}))$. Formally,

$$\mathcal{L}_{SL} = \sum_{i \in E} \left(-\frac{1}{|\mathcal{N}(i)|} \sum_{q \in \mathcal{N}(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_q / \tau)}{\sum_n \mathbb{I}_{[n \neq i]} \exp(\mathbf{z}_i \cdot \mathbf{z}_n / \tau)} \right), \tag{1}$$

where $\mathcal{N}(i)$ signifies the images belonging to the same category as the input image $\mathbf{x}_i$ in an episode, where $\mathbb{I}$ is an indicator function evaluating to 1 *iff* $n \neq i$.

Phase 2: Classifier Construction Using the previously learned representations, the next task is to construct classifiers with the ability to classify query samples into known classes or group a set of examples into new categories. Due to our aim to propose solutions addressing more realistic problems, the varying number of novel categories and the uncertain number of samples per novel category make it hard to construct a fixed-parameterized classifier. Thus, two clustering algorithms, Semi-supervised Hierarchical Clustering (SHC) and Uncertainty-aware K-means Clustering (UKC), are specifically designed for FSNCD scenarios, enhancing data grouping in open-set scenarios by aligning more closely with their inherent properties.

Semi-Supervised Hierarchical Clustering. Hierarchical clustering is one of the most commonly used approaches for grouping data, distinguished by its ability to determine the number of clusters with no need to pre-specify the number of clusters. Grouped clusters that are visually represented in a hierarchical tree called a *dendrogram* whose structural characteristics are the key to achieving optimal clustering. How-

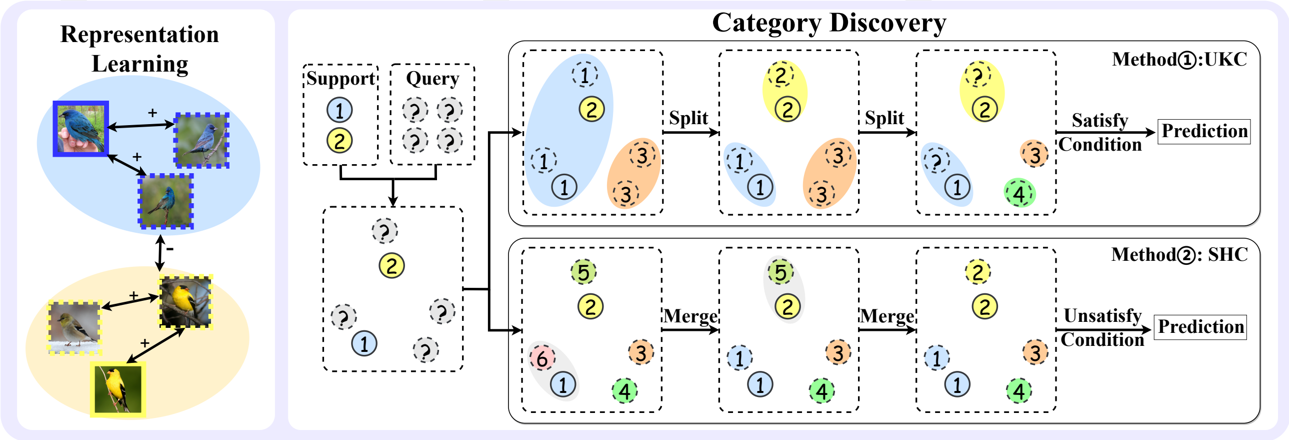

Figure 2: Illustration of our proposed baselines. In the representation learning stage, each episode takes both support and query samples into a min-batch selected from $\mathcal{D}^{\mathrm{L}}$ and trains a feature extractor by using the supervised contrastive learning approach. In the second stage, we test on $\mathcal{D}^{U}$, and for each episode, we utilize the proposed clustering method to classify samples belonging to $\mathcal{Y}^{\mathrm{S}}$ and cluster new categories.

ever, it is always a challenge in practice to determine the best criteria to end the accumulation or division during clustering.

To tackle this issue, we tend to make the most of the information represented by features extracted from the support set which are capable of explicitly characterizing discrepancies between multiple class prototypes. The idea behind this is that hierarchical clustering should cease as chosen prototypes that originate from different classes are merged. Concretely, we introduce an agglomerative hierarchical clustering approach, inspired by the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) where the closest clusters are combined into a higher-level cluster at each step. For each step, the distance between any two clusters C_i and C_j is calculated by averaging the distance between elements within the cluster, expressed by,

$$d_{C_i,C_j} = \frac{1}{|C_i| \cdot |C_j|} \sum_{x \in C_i} \sum_{y \in C_j} d(x, y), \tag{2}$$

where $d(*,*)$ denotes the distance function. In each iteration, the two clusters with the shortest distance are merged, and the clustering between all clusters is updated by,

$$d_{(C_i \cup C_j),C_k} = \frac{|C_i| \cdot d_{C_i,C_k} + |C_j| \cdot d_{C_j,C_k}}{|C_i| + |C_j|}. \tag{3}$$

The proposed clustering aims at merging the most similar classes according to the adopted cosine metric and triggers iteration termination until two different class prototypes are grouped. Furthermore, clustering is conditioned on the number of query samples, i.e., it will be considered a potential cluster P only if the amount of included samples exceeds a certain value, in which case the remaining clusters are designated as R. Moreover, in case there are support prototypes that are not assigned to any cluster, they will be incorporated into the potential class P at the end. With the available class prototypes of the potential classes, the remaining samples will be classified, demonstrating a flexible method of classifying or clustering query samples, called semi-supervised hierarchical clustering.

Uncertainty-Aware K-means Clustering. K-means clustering is an effective method to iteratively partition a certain number of data observations into K clusters, but the need for the pre-specified number of clusters always hinders its generalization to scenarios full of uncertainty. The adoption of K-means on synthetic experimental datasets is based on a *prior* on how human perception of the K values, which is unrealistic when dealing with real-world tasks where such values are difficult to determine. In addition to the heavy reliance on prior knowledge, the performance of standard K-means clustering strongly depends on the choice of the initialized cluster centroids, which is another uncertainty that needs to be addressed. K-means++ [Arthur and Vassilvitskii, 2007] presents a stepwise selection of cluster centers, primarily motivated by the idea that samples farther away from other cluster centers are more likely to be selected as the next cluster center. However, the initial number of categories is still an indispensable prior information.

We introduce a new strategy that builds on the assumption that when clustering is performed in the proposed FSNCD setting, both support and query examples lie in a uniform, high-dimensional feature space, and the uncertainty estimates of this space facilitate the choice of the number of centroids implied by novel categories. Specifically, we consider that the features of prototypes and query examples are located in a shared space, denoted by $\mathbf{z}$, and represent standard K-means as $K\text{-}means\,(\cdot,\cdot,\cdot)$, then the corresponding potential cluster P can be achieved by,

$$P = K\text{-}means\,(\mathbf{z}, n, c)\,, \tag{4}$$

where n is number of categories initialized with $|\mathcal{Y}_{\mathcal{S}}|$, c denotes the centroids and is initialized randomly. As the iteration k increases, the average quantity is calculated based on the previous clustering results, denoted as $m = \frac{1}{|P^k|}\sum_{i=1}^{|P^k|}\left|P_i^k\right|$. The estimation of the number of categories is then transformed into an acceptance criterion for a particular class in P_i^k that surpasses the average quantity m, denoted as α. We then need to split all potential classes where each class is assigned a category count by,

$$n_{P_i^k} = \begin{cases} \delta(P_i^k) & \delta(P_i^k) \geq 2 \\ 2 & \delta(P_i^k) < 2 \ \& \ \mu(P_i^k) \geq \alpha m \\ 1 & \delta(P_i^k) < 2 \ \& \ \mu(P_i^k) < \alpha m \end{cases}, \tag{5}$$

where $\delta(\cdot)$ denotes the way to estimate the number of supported prototypes in a specific cluster, and $\mu(\cdot)$ denotes the calculation of the number of query samples in P_i. Through regrouping the samples of each class into $n_{P_i^k}$ clusters, it yields the centroid $c^k = [c_1^k, ..., c_{n_{P^k}}^k]$, where $n_{P^k} = \sum_{i=1}^{|P^k|} n_{P_i^k}$ signifies each split in P^k that serves as the initialization points for the next iteration denoted as,

$$P_i^k = K\text{-}means(\mathbf{z}, n_{P^k}, c^k). \tag{6}$$

The above iteration will terminate when the conditions $\max \delta(P_i) = 1$ and $\max \mu(P) < \alpha m$ are satisfied simultaneously, which also guarantees convergence under the previously mentioned assumption. Moreover, choosing an appropriate acceptance threshold for this clustering approach facilitates the optimal proficiency to discover novel categories. It is worth noting that when $\alpha \to +\infty$, this method will degenerate to continuously split the potential classes by $\delta(P_i) \geq 2$, which will undoubtedly lead to biases towards classes of existing support sets.

3.3 Scalable Clustering

Compared to the UKC which is an evolution from standard K-means clustering, the proposed SHC exhibits relatively high computational complexity and memory usage due to its need to calculate and store distances between all pairs of data points, which grows exponentially with the size of the large-scale dataset, rendering it both computationally demanding and memory-intensive.

To mitigate the impact of the aforementioned issues, we employ a two-step strategy. In the first step, we randomly sample a smaller subset from the entire dataset, which is assumed to be representative since its distribution is consistent with that of the overall data distribution. The subsequent use of hierarchical clustering on this representative subset yields a set of potential clusters that can be viewed as initial prototypes. In the next step, these established prototypes serve as reference points for classifying the remaining data. Specifically, each data point will be merge to the tree which prototype is closest to it in the shared feature space. This strategy enables hierarchical clustering to be applied at scale, but its performance is influenced by the randomness of sampling and the quality of the initial clustering.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate our methods on two well-known large-scale datasets: CIFAR-100 [Krizhevsky *et al.*, 2009] and ImageNet-100 [Krizhevsky *et al.*, 2012], as well as two fine-grained datasets, including CUB-200 [Reed *et al.*, 2016] and Stanford Cars [Krause *et al.*, 2013]. Each dataset is split into two sets: labeled data is used to train the model,

		C100	I-100	CUB	Cars	Aircraft
Training	$\lvert\mathcal{Y}^{\text{Base}}\rvert$	50	50	100	98	50
	$\lvert\mathcal{D}^{\text{Base}}\rvert$	25k	65k	3k	6.5k	4.4k
Test	$\lvert\mathcal{Y}^{\text{Novel}}\rvert$	50	50	100	98	50
	$\lvert\mathcal{D}^{\text{Novel}}\rvert$	25k	65k	3k	8.1k	4.4k

Table 2: Statistical comparison of data partitions (i.e., training and test) across C-100 (CIFAR-100), I-100 (ImageNet-100), CUB (CUB-200), Cars (StanfordCars) and Aircraft (FGVC-Aircraft).

and unlabeled data is used for testing. SSB (Semantic Shift Benchmark) [Vaze *et al.*, 2022b] provides a detailed evaluation dataset that includes precise "semantic change axes" and provides classifications for $\mathcal{D}_\mathcal{U}$ and $\mathcal{D}_\mathcal{L}$ in a semantically coherent manner. Statistics on the partitioning of adopted datasets are presented in Table 2.

Settings. We mainly report results obtained with two different settings: 5-way 5-shot (5w5s) and 5-way 1-shot (5w1s). For better comparison, we set the number of new classes to 5 for both configurations (5n), denoted as n_{new}, with each class containing 15 images as the query set. Moreover, we initialize a real-time inference evaluation scenarios, which are in line with the objectives enabling the agent to freely switch between categorizing known and clustering novel classes. It is worth noting that for the real-time inference scenario, only one image from an individual class is allowed in the query set per episode.

A vision transformer [Dosovitskiy *et al.*, 2021] (ViT-B-16) pre-trained on ImageNet [Krizhevsky *et al.*, 2012] with DINO [Caron *et al.*, 2021] is used for the feature extraction. Specifically, the outputs of $[CLS]$ token are treated as feature representation which is different from the use of a projection head that has the potential to introduce more biases. The initial learning rate is set to 0.01.

Evaluation Protocol. We assess the performance of the proposed new setting by measuring the accuracies of three folders, they are visible, new classes and overall precision. Concretely, it uses $\hat{\mathcal{Y}}_{All}$, $\hat{\mathcal{Y}}_{Old}$ and $\hat{\mathcal{Y}}_{New} = \hat{\mathcal{Y}}_{All} \setminus \hat{\mathcal{Y}}_{Old}$ to represent the predictions of examples from all, support set and new classes, respectively. Then, the metric for assessing new classes, commonly employed in the GCD approach, has been adapted to suit the FSNCD context. Formally,

$$ACC_{New} = \max_{p \in \mathcal{P}(\hat{\mathcal{Y}}New)} \left(\frac{1}{M_{New}} \sum_{i=1}^{M_{New}} \mathbb{I}(y_i = p(\hat{y}_i)) \right), \tag{7}$$

where $M_{New} = |\hat{\mathcal{Y}}_{New}|$, and $\mathcal{P}(\hat{\mathcal{Y}}_{New})$ defines how the predicted labels for test samples are matched to the true labels. This matching is achieved by using the Hungarian algorithm which identifies the permutation that minimizes the mismatch between predicted and true labels.

Instead of directly comparing all predictions to the ground truth like GCD, we adopt a different approach. More specifically, we filter out samples from the same class as the support prototypes and then measure the matches between the pre-

		CIFAR-100			ImageNet-100			CUB-200			StanfordCars			FGVC-Aircraft		
	Methods	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
5way-5shot	ProtoNet [Snell *et al.*, 2017]	48.1	**96.2**	-	47.7	**95.4**	-	48.5	**97.0**	-	42.9	**85.8**	-	43.9	**87.7**	-
	RankStat [Han *et al.*, 2020]	43.6	64.2	23.0	41.1	59.1	23.1	49.8	73.5	26.1	39.4	57.8	21.1	43.5	65.6	21.3
	SimGCD [Wen *et al.*, 2023]	34.6	33.5	35.7	31.1	30.9	31.3	25.1	24.9	25.4	17.1	16.2	18.0	19.2	17.6	20.8
	OCD [Du *et al.*, 2023]	45.9	46.1	45.6	13.5	0.1	26.9	44.9	42.5	47.2	37.3	34.5	40.1	44.9	45.7	**44.1**
	GCD [Vaze *et al.*, 2022a]	63.8	92.7	34.7	62.9	91.9	33.9	68.7	95.0	42.5	48.2	78.6	17.8	50.0	80.5	19.5
	FSNCD (SHC)	73.2	90.7	55.8	74.1	91.5	56.7	73.6	94.5	52.6	**50.3**	64.7	35.9	**51.0**	71.2	30.8
	FSNCD (UKC)	**84.3**	90.9	**77.8**	**84.4**	87.5	**81.4**	**85.8**	92.5	**79.1**	48.8	57.2	**40.3**	49.5	57.3	41.8
5way-1shot	ProtoNet [Snell *et al.*, 2017]	43.9	**87.8**	-	42.6	**85.2**	-	45.3	**90.7**	-	33.0	**66.0**	-	35.5	**71.0**	-
	RankStat [Han *et al.*, 2020]	35.2	46.5	23.8	33.8	43.9	23.6	42.7	59.8	25.7	29.3	37.6	21.0	32.9	43.5	22.3
	SimGCD [Wen *et al.*, 2023]	34.3	32.5	36.1	31.2	29.5	32.9	25.2	23.3	27.1	17.3	15.7	18.8	18.5	16.5	20.4
	OCD [Du *et al.*, 2023]	41.1	34.8	47.5	12.3	0.1	24.5	40.4	34.4	46.5	34.3	28.8	**39.8**	41.0	38.3	**43.8**
	GCD [Vaze *et al.*, 2022a]	63.4	77.3	49.6	65.0	77.6	52.5	66.7	83.1	50.3	41.2	50.1	32.3	44.2	55.1	33.3
	FSNCD (SHC)	67.5	80.1	54.8	69.9	82.1	57.7	72.3	87.1	57.5	**42.4**	48.2	36.7	**45.9**	52.8	38.9
	FSNCD (UKC)	**75.9**	78.9	**72.9**	**76.6**	76.5	**76.6**	**84.2**	88.6	**79.9**	38.9	41.5	36.3	41.8	46.5	37.1

Table 3: Main results of 15 query images for each class. The best result, is highlighted in **bold**.

		CIFAR-100			ImageNet-100			CUB-200			StanfordCars			FGVC-Aircraft		
	Methods	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
Real-time	ProtoNet [Snell *et al.*, 2017]	48.1	**96.2**	-	47.7	**95.4**	-	48.5	**97.0**	-	42.9	**85.8**	-	43.9	**87.7**	-
	RankStat [Han *et al.*, 2020]	59.7	63.6	55.8	59.1	60.1	58.0	61.9	73.7	50.1	53.6	58.9	48.3	54.4	64.3	44.5
	OCD [Du *et al.*, 2023]	66.2	45.2	87.1	53.1	0.1	**87.6**	67.3	41.5	**93.0**	60.1	34.2	86.0	59.2	33.2	85.2
	FSNCD (SHC)	75.7	63.1	**88.2**	82.3	73.0	91.7	74.9	63.7	86.1	57.1	17.7	**96.4**	56.0	15.7	**96.2**
	FSNCD (UKC)	**78.3**	70.0	86.6	**83.3**	79.9	86.7	**78.1**	74.8	81.4	**61.2**	29.6	92.8	**61.0**	30.4	91.6

Table 4: Main result of real-time inference tasks. The best result, excluding ProtoNet, is highlighted in **bold**.

dicted labels and prototypes by

$$ACC_{Old} = \frac{1}{M}\sum_{i=1}^{M}\mathbb{I}(y_i = p_{old}(\hat{y}_i)), \quad (8)$$

where p_{old} represents the one-to-one mapping from the label space of support prototypes $\mathcal{Y}_{Old}$ to $\hat{\mathcal{Y}}_{Old}$.

FSNCD Baselines. We present the following key benchmark methods as the strongest baselines for comparison with the proposed FSNCD. All methods employed ViT-B-16 pretrained on ImageNet with DINO.

(1) ProtoNet [Snell *et al.*, 2017] aims to classify samples into prototypes based on maximum cosine similarity, in contrast, our FSNCD identifies known classes while also considering how to discover novel classes.

(2) OCD [Du *et al.*, 2023] attempts to enhance the generalizability of GCD by using hash codes derived from disentangled features, while our FSNCD exploits a more general feature space to solve analogous generalization problems.

(3) AutoNovel [Han *et al.*, 2020] uses feature ranking to determine whether a given sample belongs to positive pair, while we further extend its indexing ideas to the support prototypes. We use the two closest prototypes as category labels.

(4) SimGCD [Wen *et al.*, 2023] attempts to utilize parametric classifier to solve GCD problem. Each episode of FSNCD can be regarded as a GCD problem with an extremely scarce sample size. We use the weights from Phase 1 and fine-tune using SimGCD to discover novel categories.

(4) GCD [Vaze *et al.*, 2022a] proposes to estimate the number of novel categories with accuracy on labeled samples, and discover categories with semi-supervised K-means. We treat each episode as a GCD task, performing estimation and clustering to discover novel categories.

4.2 Main Results

Table 3 compares method performance across five datasets. We evaluate ”old class” accuracy (support set samples) and ”new class” accuracy (novel query set categories). ProtoNet fails to recognize new classes due to its forced classification into support set categories. Among clustering-based methods, GCD employs semi-supervised K-means for category estimation but struggles with coarse-grained datasets. Our method eliminates category number pre-estimation and achieves state-of-the-art performance, showing 9% (SHC) and 13% (UKC) average old-class accuracy drops versus ProtoNet, yet delivering 5% (SHC) and 10% (UKC) overall accuracy gains over GCD.

OCD exhibits consistent performance across tasks due to its overly strict hash code categorization (12-bit codes allow up to 2^{12} classes) and underutilization of support feature distributions. While increasing support samples improves hash code precision as noted in [Du *et al.*, 2023], its ultra-fine category division remains fundamentally limited.

RankStat achieves optimal performance when encoding the top two ranks but limits the number of categories to 20, showing strong retention of classification capability for old classes as expected. Its strict partitioning of feature space into 20

	Methods	CIFAR-100 All	CIFAR-100 Old	CIFAR-100 New	ImageNet-100 All	ImageNet-100 Old	ImageNet-100 New
Large scale	ProtoNet [Snell *et al.*, 2017]	49.0	**95.1**	-	47.7	95.4	-
	RankStat [Han *et al.*, 2020]	57.8	61.3	54.4	52.5	50.4	54.6
	OCD [Du *et al.*, 2023]	41.5	39.3	43.7	25.2	22.7	27.6
	FSNCD (SHC)	71.4	89.7	53.1	76.4	94.6	58.2
	FSNCD (UKC)	**89.3**	92.2	**86.3**	**97.5**	**98.1**	**96.8**

Table 5: Main result of large-scale dataset annotation tasks. The best result, excluding ProtoNet, is highlighted in **bold**.

regions based on support samples slightly reduces labeling effectiveness for new classes.

We further conducted tests for real-time inference. As shown in Table 4, OCD easily identifies a sample as a new class, but it also tends to misclassify a substantial number of old classes as new. Although OCD achieves high accuracy in new class discovery, its utility is limited, indicating an excessive inclination toward categorizing instances as new classes. FSNCD with the SHC method achieves a more balanced performance in real-time inference, while FSNCD with the UKC exhibits lower capability in discovering new classes during real-time inference but still maintains excellent performance on old classes.

As shown in Table 5, OCD struggles with large-scale dataset annotation due to its dispersed category distribution, resembling On-the-fly Category Discovery in accuracy. In contrast, UKC leverages sufficient samples to better capture data distributions, achieving superior performance across multiple metrics. Notably, discrepancies between Table 3 and Table 5 are attributed to differing query sample sizes. Specifically, using centroids from small batches to classify large datasets enhances performance, aligning Cosine-trained models with Euclidean clustering. Additionally, in traditional GCD tasks, we found that the strategy of clustering and classification can significantly imporve performance. In extreme cases with limited query samples, it becomes challenging to determine whether a sample belongs to a new category, which is an inherent challenge in FSNCD.

4.3 Ablation Studies

To validate the model's performance across various scenarios, we compared its performance under different settings. For more details on related parameter studies, please refer to the supplementary materials.

Quantity of Query. Due to the discovery of novel categories relying on clustering results, the model may be more sensitive to the distribution of data. Simultaneously, to ensure real-time reasoning capability, we investigated the impact of query quantity on model metrics. Initially, we set the task as 5w5s5n.

Fig. 3 illustrates the accuracy variation with the number of query samples in each class using the 5w5s configuration. The accuracy of UKC change curve indicates that as the number of query samples in each class increases, the data distribution tends to be more reasonable, leading to a continuous improvement. Conversely, due to the weak dependence of hierarchical clustering on data distribution, the accuracy of the SHC based method is more stable.

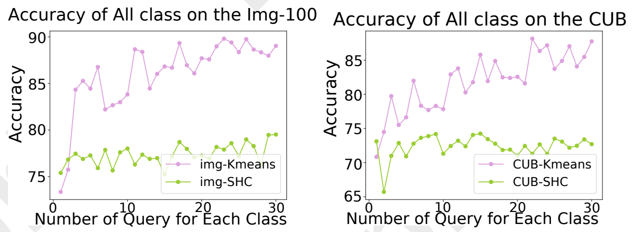

Figure 3: Ablation studies for quantity of query using the 5w5s configuration. (a) The accuracy for Imagenet-100 on All classes. (b) The accuracy for CUB-200 All classes.

	Method	CUB-200 All	CUB-200 Old	CUB-200 New	ImageNet-100 All	ImageNet-100 Old	ImageNet-100 New
(1)	SHC	52.4	**91.9**	42.5	57.9	**93.9**	48.9
	UKC	**61.2**	86.5	**54.9**	**73.6**	91.1	**69.3**
(2)	SHC	59.3	**86.5**	45.7	72.1	**88.7**	63.8
	UKC	**69.4**	75.3	**66.5**	**83.4**	87.0	**81.6**
(3)	SHC	48.8	**84.1**	40.0	63.0	**88.2**	56.7
	UKC	**50.0**	55.8	**48.6**	79.4	80.4	**79.1**

Table 6: Ablation study of different task. (1). 5-way 5-shot 20-new, (2). 10-way 5-shot 20-new, (3). 10-way 5-shot 40-new and both above tasks all use 15 query images of each category.

Effects of Different Settings. In response to distinct task configurations, we conducted further examinations on the outcomes associated with three specific task settings: (1) 5-way 5-shot 20-new, (2) 10-way 5-shot 20-new, and (3) 10-way 5-shot 40-new, and take 15 query images for each category. The results are shown in Table 6. Given that UKC is contingent on the underlying data distribution, the results obtained from UKC exhibit a comparatively stable trend. Our approach demonstrates improvements across all three evaluation metrics as the number of supported sample categories increases. Notably, even with the introduction of additional novel classes, our method consistently maintains a commendable level of accuracy on pre-existing classes. This robust performance underscores the efficacy of our method in handling diverse task settings and accommodating the challenges posed by both an increased number of supported categories and the introduction of new ones.

5 Conclusion

In this paper, we summarize the shortcomings of current Novel Category Discovery (NCD) tasks and Few-Shot Learning (FSL) tasks, considering FSL tasks that are closer to real-world scenarios, where new classes may appear in the query set. To address the aforementioned issues, we propose Few-shot Novel Category Discovery (FSNCD). In the inference stage, we introduce Semi-supervised Hierarchical Clustering (SHC) and Uncertainty-aware K-means Clustering (UKC) through supervised contrastive learning to represent learning in each episode and leverage them to classify old and cluster new classes accurately. We conduct comprehensive experiments on five datasets, validating the model's performance in general settings and its performance in extreme scenarios.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (NSFC) under Grant No. 62371235, and partly by the Key Research and Development Plan of Jiangsu Province (Industry Foresight and Key Core Technology Project) under Grant BE2023008-2.

References

[Arthur and Vassilvitskii, 2007] David Arthur and Sergei Vassilvitskii. k-means++: the advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, page 1027–1035, 2007.

[Caron *et al.*, 2021] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.

[Chapelle *et al.*, 2009] O. Chapelle, B. Scholkopf, and A. Zien, Eds. Semi-supervised learning. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.

[Chen *et al.*, 2019] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *International Conference on Learning Representations*, 2019.

[Dhillon *et al.*, 2020] Guneet Singh Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A baseline for few-shot image classification. In *International Conference on Learning Representations*, 2020.

[Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, and Sylvain Gelly. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[Du *et al.*, 2023] Ruoyi Du, Dongliang Chang, Kongming Liang, Timothy Hospedales, Yi-Zhe Song, and Zhanyu Ma. On-the-fly category discovery. In *IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[Fei-Fei *et al.*, 2006] Li Fei-Fei, Robert Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006.

[Finn *et al.*, 2017] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017.

[Grandvalet and Bengio, 2004] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, page 529–536, 2004.

[Han *et al.*, 2020] 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*, 2020.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[He *et al.*, 2022] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15979–15988, 2022.

[Jeong *et al.*, 2021] Minki Jeong, Seokeon Choi, and Changick Kim. Few-shot open-set recognition by transformation consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 12561–12570, 2021.

[Jiang *et al.*, 2024] Bingbing Jiang, Xingyu Wu, Xiren Zhou, Yi Liu, Anthony G. Cohn, Weiguo Sheng, and Huanhuan Chen. Semi-supervised multiview feature selection with adaptive graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(3):3615–3629, 2024.

[Khosla *et al.*, 2020] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Advances in neural information processing systems*, volume 33, pages 18661–18673, 2020.

[Koch *et al.*, 2015] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, 2015.

[Krause *et al.*, 2013] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 554–561, 2013.

[Krizhevsky *et al.*, 2009] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Technical report, Univeristy of Toronto, 2009.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 2012.

[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*, page 896, 2013.

[Li *et al.*, 2023] Wenbin Li, Ziyi Wang, Xuesong Yang, Chuanqi Dong, Pinzhuo Tian, Tiexin Qin, Jing Huo, Yinghuan Shi, Lei Wang, Yang Gao, et al. Libfewshot: A comprehensive library for few-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Liu and Tuytelaars, 2023] Yu Liu and Tinne Tuytelaars. Residual tuning: Toward novel category discovery without labels. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10):7271–7285, 2023.

[Liu *et al.*, 2020] Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos. Few-shot open-set recognition using meta-learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8795–8804, 2020.

[Madani *et al.*, 2018] Ali Madani, Mehdi Moradi, Alexandros Karargyris, and Tanveer Syeda-Mahmood. Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1038–1042, 2018.

[Ravi and Larochelle, 2016] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International conference on learning representations*, 2016.

[Reed *et al.*, 2016] Scott Reed, Zeynep Akata, Honglak Lee, and Bernt Schiele. Learning deep representations of fine-grained visual descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 49–58, 2016.

[Rizve *et al.*, 2022] Mamshad Nayeem Rizve, Navid Kardan, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Openldn: Learning to discover novel classes for open-world semi-supervised learning. In *European Conference on Computer Vision*, pages 382–401, 2022.

[Roy *et al.*, 2022] Subhankar Roy, Mingxuan Liu, Zhun Zhong, Nicu Sebe, and Elisa Ricci. Class-incremental novel class discovery. In *European Conference on Computer Vision*, pages 317–333, 2022.

[Snell *et al.*, 2017] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in neural information processing systems*, volume 30, 2017.

[Sung *et al.*, 2018] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018.

[Tarvainen and Valpola, 2017] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems*, volume 30, 2017.

[Tian and Xie, 2023] Pinzhuo Tian and Shaorong Xie. An adversarial meta-training framework for cross-domain few-shot learning. *IEEE Transactions on Multimedia*, 25:6881–6891, 2023.

[Vaze *et al.*, 2022a] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7492–7501, June 2022.

[Vaze *et al.*, 2022b] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: a good closed-set classifier is all you need? In *International Conference on Learning Representations*, 2022.

[Vinyals *et al.*, 2016] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016.

[Wen *et al.*, 2023] Xin Wen, Bingchen Zhao, and Xiaojuan Qi. Parametric classification for generalized category discovery: A baseline study. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16590–16600, 2023.

[Xu *et al.*, 2024] Jian Xu, Bo Liu, and Yanshan Xiao. A multitask latent feature augmentation method for few-shot learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):6976–6990, 2024.

[Yang *et al.*, 2023] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):8934–8954, 2023.

[Yang *et al.*, 2024] Xihong Yang, Xiaochang Hu, Sihang Zhou, Xinwang Liu, and En Zhu. Interpolation-based contrastive learning for few-label semi-supervised learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(2):2054–2065, 2024.

[Zhao *et al.*, 2022] Yuyang Zhao, Zhun Zhong, Nicu Sebe, and Gim Hee Lee. Novel class discovery in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4340–4349, 2022.

[Zhong *et al.*, 2023] Xian Zhong, Cheng Gu, Mang Ye, Wenxin Huang, and Chia-Wen Lin. Graph complemented latent representation for few-shot image classification. *IEEE Transactions on Multimedia*, 25:1979–1990, 2023.

[Zhou, 2008] Zhi-Hua Zhou. Semi-supervised learning by disagreement. In *IEEE International Conference on Granular Computing*, pages 93–93, 2008.

[Zhu *et al.*, 2023] Meiqi Zhu, Xiao Wang, Chuan Shi, Yibo Li, and Junping Du. Towards adaptive information fusion in graph convolutional networks. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):13055–13069, 2023.