# Let's Start Over: Retraining with Selective Samples for Generalized Category Discovery

**Zhimao Peng**[1] , **Enguang Wang**[1] , **Xialei Liu**[1*] and **Ming-Ming Cheng**[1]
[1]VCIP, College of Computer Science, Nankai University, Tianjin, China.
{zhimao796, enguangwang}@mail.nankai.edu.cn, {xialei, cmm}@nankai.edu.cn

## Abstract

Generalized Category Discovery (GCD) presents a realistic and challenging problem in open-world learning. Given a partially labeled dataset, GCD aims to categorize unlabeled data by leveraging visual knowledge from the labeled data, where the unlabeled data includes both known and unknown classes. Existing methods based on parametric/non-parametric classifiers attempt to generate pseudo-labels/relationships for the unlabeled data to enhance representation learning. However, the lack of ground-truth labels for novel classes often leads to noisy pseudo-labels/relationships, resulting in suboptimal representation learning. This paper introduces a novel method using Nearest Neighbor Distance-aware Label Consistency sample selection. It creates class-consistent subsets for novel class sample clusters from the current GCD method, acting as "pseudo-labeled sets" to mitigate representation bias. We propose progressive supervised representation learning with selected samples to optimize the trade-off between quantity and purity in each subset. Our method is versatile and applicable to various GCD methods, whether parametric or non-parametric. We conducted extensive experiments on multiple generic and fine-grained image classification datasets to evaluate the effectiveness of our approach. The results demonstrate the superiority of our method in achieving improved performance in generalized category discovery tasks.

## 1 Introduction

Deep neural networks have achieved great success in a variety of mainstream vision tasks, benefiting from the availability of large-scale annotated data [Krizhevsky *et al.*, 2012; He *et al.*, 2016]. Even in semi-supervised scenarios that provide little labeled data and many unlabeled data, the latest work [Fini *et al.*, 2023] can achieve excellent performance on image classification tasks. However, these successes rely on the close-set assumption: test data only comes from categories that have been seen in the model training. In the real world, it is inevitable to encounter unknown categories of data, which presents a huge challenge for model deployment. In order to enable models to handle more realistic scenarios, more and more work attempts to break the constraints of close-set: open-set recognition (OSR) [Scheirer *et al.*, 2012] enables the model to recognize known classes data while rejecting unknown classes, novel category discovery (NCD) [Han *et al.*, 2019] enables the model to recognize unknown classes in the unlabelled data after training on known classes with labeled data. Although NCD can enable the model to recognize unknown classes without any annotation, it assumes that all unlabeled data comes from unknown classes, which is very restrictive. To break this constraint, Generalized Category Discovery (GCD) [Vaze *et al.*, 2022a] has been proposed, which assumes that unlabelled data comes from both known and unknown classes.

As a challenging problem, GCD not only needs to classify images of known classes correctly but also needs to cluster images of novel classes. Inspired by the powerful neighbor classification capabilities of large-scale pre-trained ViTs [Caron *et al.*, 2021], the seminal work [Vaze *et al.*, 2022a] attempted to solve GCD based on non-parametric classifier: supervised/self-supervised contrastive learning on labeled/unlabeled data is used to perform representation learning on Vision Transformers, and then semi-supervised k-means is used to classify unlabelled data. Some follow-up works [Pu *et al.*, 2023; Zhang *et al.*, 2023; Zhao *et al.*, 2023; Hao *et al.*, 2023] attempt to construct pseudo-relationships for all labeled/unlabeled training data so that supervised contrastive learning can be performed for these data to learn unbiased representation, thereby facilitating the separation of unknown classes data in the representation space. Recently, another parametric classifier-based GCD method [Wen *et al.*, 2023] has been proposed, which trains a unified classification head for all known/unknown classes by a single cross-entropy loss through DINO-like [Caron *et al.*, 2021] form of self-distillion. While these methods achieve improved results, without the guidance of unknown classes of ground-truth labels, they inevitably generate many noisy pseudo relationships/labels for unlabelled data, resulting in suboptimal results.

To generate reliable pseudo relationships/labels for unla-

[*] Corresponding author.

belled data, ideally, we should know the ground-truth labels of these samples, but this is impossible because the purpose of the GCD task is to assign labels to them. However, we can obtain a cluster assignment of these unlabelled data by employing an off-the-shelf GCD method, although each sample cluster is expected to contain a different class of unlabelled samples, a significant proportion of these samples should be class-consistent due to semantic clustering. This allows us to select a class-consistent subset for each novel class sample cluster as a pseudo-labeled set for that cluster. By incorporating these pseudo-labeled sets into GCD retraining, we can construct more reliable pseudo relationships/labels than the original GCD training. For this purpose, we propose to use a nearest neighbor distance-aware label consistency criterion to perform reliable sample selection for each cluster.

Specifically, given the cluster assignment generated by the current GCD method, we first find the top-$K$ nearest neighbors for each sample from all training data based on the cosine distance between the feature representation. Then the original class probability distribution is obtained by aggregating the clustering labels of the nearest neighbor samples, and the index corresponding to the maximum probability of this distribution can be used as the corrected pseudo-labels, so that the corrected class probability distribution can be obtained by the corrected pseudo-labels. Finally, the cross-entropy value between the corrected class probability distribution and the original clustering label distribution is used to measure the reliability of the sample. To ensure the same number of samples are selected for novel clusters, we count the number of samples in which the original cluster label and the corrected pseudo-label are agreed upon for each novel class cluster, and a threshold can determine the equal number of samples for each cluster. To address the trade-off between quantity and purity in sample selection, we propose to use supervised representation learning with selected samples for progressive sample selection and take the k-means clustering results of known class validation data as criteria to select the appropriate number of samples.

The main contributions of this paper can be summarized as follows: (1) We propose to construct class-consistent subsets from the cluster assignment of unknown classes generated by current GCD methods as pseudo-labeled sets of these classes and incorporate them into the GCD retraining, thus generating less noisy pseudo relationships/labels during the retraining process. (2) We propose to use the nearest neighbor distance aware label consistency as the criterion and perform supervised representation learning with selected samples for progressive sample selection to achieve a good trade-off between the quantity and purity of the final selected samples. (3) We evaluate our approach on both parametric and non-parametric classifier base GCD methods, achieving superior performance on multiple generic and fine-grained datasets.

## 2 Related Work

**Generalized Category Discovery.** Generalized Category Discovery (GCD) can be seen as an extension of novel category discovery (NCD) [Han *et al.*, 2019; Han *et al.*, 2020; Zhong *et al.*, 2021; Fini *et al.*, 2021] in real scenarios. From the perspective of classifier form, GCD can be divided into non-parametric classifier-based method and parametric classifier-based method. For the former, [Vaze *et al.*, 2022a] explores the relationship between labeled data through supervised contrastive learning; meanwhile, in order to extend supervised contrastive learning to unlabelled data, [Pu *et al.*, 2023; Zhang *et al.*, 2023; Zhao *et al.*, 2023; Hao *et al.*, 2023] are devoted to establishing the pseudo-relationship between all labeled/unlabeled training data to learn unbiased representation, and then classify samples by semi-supervised k-means or other non-parametric clustering [Hao *et al.*, 2023]. The latter [Rizve *et al.*, 2022; Fini *et al.*, 2023] trains a unified classification head for all classes through DINO [Caron *et al.*, 2021] or SwAV[Caron *et al.*, 2020]-Like form of online clustering. Despite the progress, they inevitably generate many noisy pseudo relationships/labels for unlabelled data, resulting in suboptimal results. We try to select class-consistent sample subsets for the sample clusters generated by current GCD methods, thus generating less-noisy pseudo relationships/labels and incorporating them into retraining to improve model performance.

**Semi-supervised Learning.** Semi-supervised learning (SSL) aims to learn from a limited amount of labeled data and a large amount of unlabelled data. Pseudo-labeling [Lee and others, 2013; Arazo *et al.*, 2020; Rizve *et al.*, 2021] and consistency regularization [Laine and Aila, 2016; Tarvainen and Valpola, 2017; Xie *et al.*, 2020] are two popular methods of SSL in the era of deep learning. The former first uses the model trained with labeled data to generate pseudo-labels for unlabelled data and then adds them to the next training together with the labeled data. The latter forces the model to output consistent predictions for the perturbed unlabeled data, which improves the generalization ability and robustness of the model. We select a class-consistent sample subset from each novel class sample cluster as the labeled data of that class, so that GCD can be treated as a semi-supervised learning task.

**Deep Clustering.** Image clustering is a prevalent proxy task for unsupervised image representation learning. Deep clustering [Caron *et al.*, 2018] is the first method to combine clustering and deep network for unsupervised representation learning. IIC [Ji *et al.*, 2019] introduces a mutual information maximization objective to perform clustering-based unsupervised representation learning. Asano *etal.* [YM. *et al.*, 2020] propose to transform cluster assignment into an optimal transport problem and solve it by Sinkhorn-Knopp algorithm. Based on the above clustering method, Caron *etal.* [Caron *et al.*, 2020] propose a swapped prediction mechanism: the cluster assignment of one view of the same image is used as the pseudo-label for another view. Subsequently, they propose [Caron *et al.*, 2021] a self-distillation mechanism based on a teacher-student network to perform unsupervised representation learning for ViTs.

## 3 Preliminaries

### 3.1 Problem Formulation

For a standard GCD task, the overall training data $\mathcal{D}$ is divided into two datasets: labeled dataset $\mathcal{D}_l$ and unlabeled dataset $\mathcal{D}_u$, where $\mathcal{D}_l = \{\mathbf{x}_i, \tilde{y}_i\}_{i=1}^{n}$, with $\tilde{y}_i \in \mathcal{Y}_l$, and $\mathcal{D}_u = \{\mathbf{x}_j\}_{j=1}^{m}$, with underlying label space $\mathcal{Y}_u$. Specifically, the label space of labelled dataset $\mathcal{Y}_l \subseteq \mathcal{Y}_u$ and the novel classes of unlabelled dataset $\mathcal{Y}_n = \mathcal{Y}_{\mathrm{u}} \backslash \mathcal{Y}_l$. For convenience, we denote $\mathcal{Y}_u$, $\mathcal{Y}_l$ and $\mathcal{Y}_n$ as "All", "Old", and "New" classes respectively. In general, GCD assumes that the number of classes in $\mathcal{D}$ (*i.e.*$|\mathcal{Y}_u|$) is known as a priori[Fei *et al.*, 2022; Zhang *et al.*, 2023; Wen *et al.*, 2023] and can also be estimated by some ready-made methods for practical applications[Han *et al.*, 2019; Vaze *et al.*, 2022a; Pu *et al.*, 2023]. The goal of GCD is not only to correctly classify the old classes samples in $\mathcal{D}_u$ but group the new classes samples in $\mathcal{D}_u$ into $|\mathcal{Y}_n|$ clusters by employing a model with the knowledge from $\mathcal{D}_l$.

### 3.2 Two Paradigms of GCD

**Non-parametric GCD methods.** Given the training data $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$, we assume $\mathbf{x}_i$ and $\mathbf{x}_i'$ are two randomly augmented views of the same image in a mini-batch $\mathcal{B}$. $\mathbf{z_i} = \phi(\mathbf{x_i})$ is the $L_2$-normalized feature embedding, $\mathcal{P}(\mathbf{x})$ is the positive set of feature embedding $\mathbf{z_i}$. $\mathcal{B}^L$ is the labelled subset of $\mathcal{B}$. The seminal work [Vaze *et al.*, 2022a] performed supervised contrastive learning on $\mathcal{D}_l$:

$$\mathcal{L}_{\phi}^{s} = -\frac{1}{|\mathcal{P}(\mathbf{x})|} \sum_{\mathbf{z}_i^+ \in \mathcal{P}(\mathbf{x})} \log \frac{\exp\left(\mathbf{z}_i^\top \cdot \mathbf{z}_i^+ / \tau^s\right)}{\sum_{j \in \mathcal{B}^L, j \neq i} \exp\left(\mathbf{z}_i^\top \cdot \mathbf{z}_j / \tau^s\right)} \tag{1}$$

and self-supervised contrastive learning on $\mathcal{D}$:

$$\mathcal{L}_{\phi}^{u} = -\log \frac{\exp\left(\mathbf{z}_i^\top \cdot \mathbf{z}_i' / \tau^u\right)}{\sum_{j \in \mathcal{B}, j \neq i} \exp\left(\mathbf{z}_i^\top \cdot \mathbf{z}_j / \tau^u\right)} \tag{2}$$

Here, the relationship between labeled data can be explored by supervised contrastive learning due to the availability of ground truth labels. However, self-supervised contrastive learning only makes the sample close to its own augmented counterpart and far away from all other samples, so the relationship between labeled data and unlabelled data and within unlabelled data is ignored. Several subsequent efforts [Pu *et al.*, 2023; Zhang *et al.*, 2023; Hao *et al.*, 2023] focused on generating pseudo-labels for unlabeled data so that supervised contrastive learning could be used to explore the relationship between all training data as shown in (Fig. 1). However, unlike semi-supervised learning, which provides some ground-truth labels as guidance for all classes, it is not easy to generate accurate pseudo labels for novel class data without any annotation in GCD. Even if current work has designed many ingenious mechanisms to reduce pseudo-label noise, non-negligible undetected noisy pseudo-label still causes the network to learn suboptimal representation, resulting in unsatisfactory performance.

**Parametric GCD methods.** Given the training data $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$, as shown in (Fig. 1), parametric GCD Methods [Wen *et al.*, 2023; Rizve *et al.*, 2022] uses a single cross-entropy loss $\ell$ to train a unified prototypical classification head that includes all old and new classes. Specifically, the labeled samples $x$ can be trained directly by using class labels as supervision information:

$$\frac{1}{|B^l|} \sum_{i \in B^l} \ell\left(\boldsymbol{y}_i, \boldsymbol{p}_i\right), \tag{3}$$

where $\boldsymbol{y}_i$ the one-hot ground-truth label of $\boldsymbol{x}_i$ and $\boldsymbol{p}_i$ is the softmax normalized probability of $\boldsymbol{x}_i$. For unlabelled samples, two random augmented views $(\mathbf{x}, \mathbf{x}')$ of each input image are fed into the network backbone $f(.)$ and prototypical classifier $h$ to obtain the classification logits $(\mathbf{g}, \mathbf{g}')$, softmax function with temperature is then applied to these logits to obtain the probability distribution $(\mathbf{p}, \mathbf{p}')$ of the input image.

Specifically, the process of pseudo-labeling can be represented as follows:

$$\mathbf{q}' = \tau(\mathbf{p}', \mathbf{c}, \eta) \tag{4}$$

where $\tau$ is the pseudo-labelling methods (*e.g.self-distilling, optimal transport*), $\mathbf{c}$ is the implemented context of $\tau$, $\eta$ is entropy regularization term, $\mathbf{q}'$ is the generated pseudo label, which can be used as the target of $\mathbf{p}$ in the cross-entropy loss:

$$\ell(\mathbf{q}', \mathbf{p}) = -\sum_{t=1}^{T} \mathbf{q}_t' \log\left(\mathbf{p}_t\right) \tag{5}$$

where $T$ is the number of sample clusters.

However, due to the strong supervision of the old classes during training, the representation of novel classes samples is easily biased toward the old classes. Meanwhile, due to the random initialization of the prototypical classifier and without the guidance of ground-truth label, the clustering of unlabelled data of novel classes in the feature space is suboptimal and even completely inconsistent with the actual distribution of that classes, since the prototypical classifier may encode spurious correlations for unlabelled data[Fini *et al.*, 2023]. On the contrary, clusters of unlabelled data of old classes could be concentrated on the class centroid represented by the labeled data, which makes the label information of labeled data can be efficiently propagated to unlabeled data.

## 4 Our Method

In this section, we introduce a method for enhancing label consistency by selecting pseudo-labeled samples through a neatest neighbor distance-aware strategy. These samples form subsets for novel class cluster assignments derived from the off-the-shelf GCD method. We then detail our progressive sample selection training approach with supervised representation learning. Finally, we explain the integration of these selected pseudo-labeled sample subsets into the GCD retraining process.

**Enhancing label consistency through nearest neighbor distance-aware sample selection.** To alleviate the biased representation caused by the lack of supervision information of new classes, we propose to select a class-consistent subset from each noisy novel class sample cluster that the current GCD method generates as the "pseudo-labeled set" for that cluster, which allows the unlabelled data of new and old classes to be trained consistently, thus reducing the bias of the learned representation. Although it sounds attractive, selecting such a class-consistent subset is very challenging because

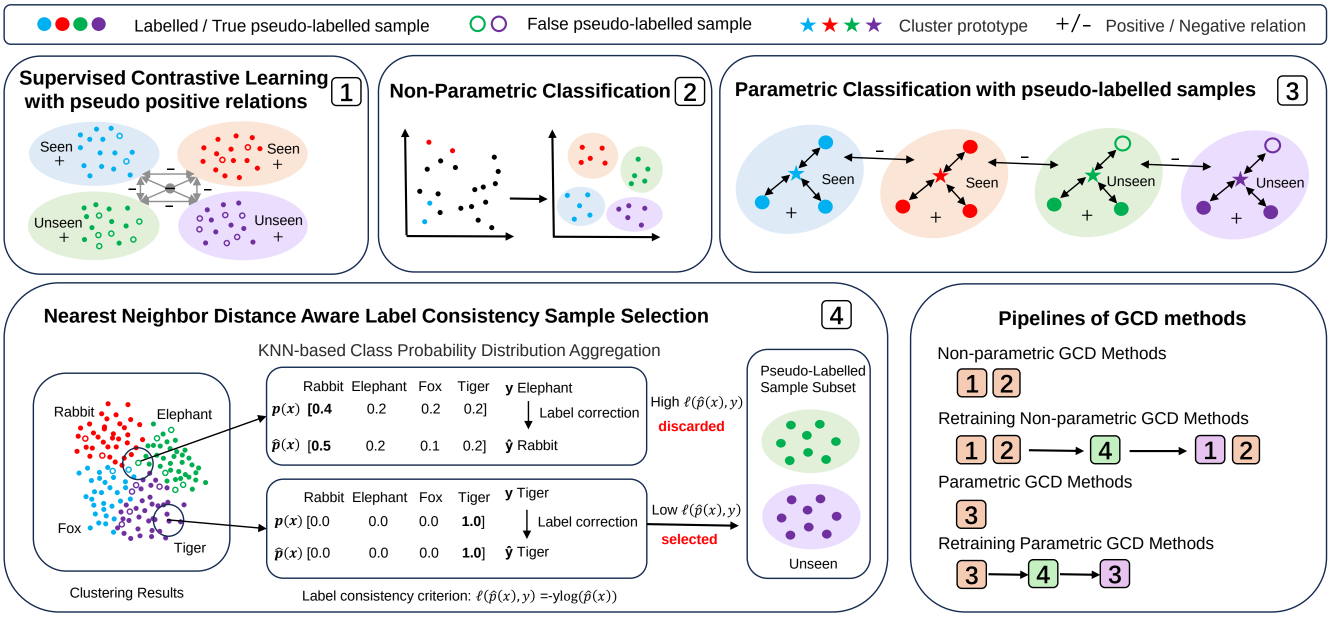
Labelled / True pseudo-labelled sample
False pseudo-labelled sample
Cluster prototype
+/- Positive / Negative relation
Supervised Contrastive Learning with pseudo positive relations 1
Seen +
Seen +
Unseen +
Unseen +
Non-Parametric Classification 2
Parametric Classification with pseudo-labelled samples 3
Seen
Seen
Unseen
Unseen
Nearest Neighbor Distance Aware Label Consistency Sample Selection 4
KNN-based Class Probability Distribution Aggregation
Rabbit
Elephant
Fox
Tiger
Clustering Results
Rabbit Elephant Fox Tiger
p(x) [0.4 0.2 0.2 0.2]
p̂(x) [0.5 0.2 0.1 0.2]
y Elephant
Label correction
ŷ Rabbit
Rabbit Elephant Fox Tiger
p(x) [0.0 0.0 0.0 1.0]
p̂(x) [0.0 0.0 0.0 1.0]
y Tiger
Label correction
ŷ Tiger
High ℓ(p̂(x), y)
discarded
Low ℓ(p̂(x), y)
selected
Pseudo-Labelled Sample Subset
Unseen
Label consistency criterion: ℓ(p̂(x), y) =-ylog(p̂(x))
Pipelines of GCD methods
Non-parametric GCD Methods
1 2
Retraining Non-parametric GCD Methods
1 2 → 4 → 1 2
Parametric GCD Methods
3
Retraining Parametric GCD Methods
3 → 4 → 3

Figure 1: Illustration of the proposed retraining with selective samples for GCD. In the subgraph "Pipelines of GCD methods", "→" represents delimiter for each stage, the process block with purple background represents that the selected novel classes samples are incorporated to the GCD retraining as the pseudo-labeled sample subset.

the noisy samples in the sample cluster are mostly "hard negative". In order to efficiently select less-noisy sample sets for novel classes, inspired by [Ortego *et al.*, 2021], we propose to use the Nearest Neighbor Distance-Aware criterion to select reliable samples for each cluster. For the sample clusters $\mathcal{S} = \left\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \ldots, \mathcal{S}_{|\mathcal{Y}_u|}\right\}$ generated by the GCD method, where $\mathcal{S}_c = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n_c}, c \in \mathcal{Y}_u$, $n_c$ is the sample size of $\mathcal{S}_c$. Given the feature representations $z_i$ of a sample $x_i$ with assigned cluster label $y_i$, we select the top-$K$ nearest neighbor samples $\mathcal{N}_i$ for $x_i$ from all other training data $D \backslash x_i$ according to the cosine similarity between their feature representations. To correct the original assigned noisy cluster labels, the cluster labels of $K$-nearest neighbor samples are aggregated as the class probability distribution for $x_i$ :

$$p_c\left(\boldsymbol{x}_i\right) = \frac{1}{K} \sum_{\substack{k=1 \\ x_k \in \mathcal{N}_i}}^{K} \mathbb{1}\left[y_k = c\right], c \in \mathcal{Y}_u \tag{6}$$

we can take the index corresponding to the maximum probability of $p_c\left(\boldsymbol{x}_i\right)$ as the corrected pseudo-label $\hat{y}_i$. Therefore, the corrected class probability distribution is:

$$\hat{p_c}\left(\boldsymbol{x}_i\right) = \frac{1}{K} \sum_{\substack{k=1 \\ x_k \in \mathcal{N}_i}}^{K} \mathbb{1}\left[\hat{y_k} = c\right], c \in \mathcal{Y}_u \tag{7}$$

Finally, cross-entropy value is used to measure the difference between the pseudo-label probability distribution and the original label probability distribution:

$$\ell\left(\hat{\boldsymbol{p}}\left(\boldsymbol{x}_i\right), y_i\right) = -y_i \log(\hat{p}\left(\boldsymbol{x}_i\right)) \tag{8}$$

The higher the value, the higher the probability that the sample is noisy. Therefore, we can select reliable samples based on the cross-entropy value, and to ensure the number of samples selected for all clusters is consistent, for each novel class sample cluster, we count the number of samples in which the original cluster label $y_i$ and the corrected pseudo-label $\hat{y}_i$ agreed, *i.e.*, $M = \left\{m_{|\mathcal{Y}_l|+1}, \ldots, m_{|\mathcal{Y}_u|}\right\}$, where $m_c = \sum_{i=1}^{n_c} \mathbb{1}\left[\hat{y}_i = y_i\right]\left[y_i = c\right], c \in \left[|\mathcal{Y}_l| + 1, |\mathcal{Y}_u|\right]$. Specifically, the $\gamma$ quantiles of $M$ can be used as the selected samples number $\omega$ for each cluster, where $\gamma \in [0, 1]$. Finally, we can select $\omega$ samples for each novel class sample cluster as class-consistent subset $\mathcal{S}_c^{\omega}$ according to the value of cross-entropy in Eq.8 from small to large and the final selected sample is $\mathcal{S}_{novel} = \cup_{c=|\mathcal{Y}_l|+1}^{|\mathcal{Y}_u|} \mathcal{S}_c^{\omega}$.

**Progressive sample selection based on representation learning.** For the selected class-consistent subset, our selection strategy naturally leads to the trade-off in quantity and purity: a large $\gamma$ value results in a higher number of selected samples, but the purity of the sample subset decreases because more less-confident samples are selected; a small $\gamma$ value will select a small number of high-purity samples for the sample subset, but they cannot provide rich supervisory information in subsequent training. In order to achieve a good quantity-purity trade-off, we perform supervised representation learning for progressive sample selection. Specifically, we set a $\gamma$ value that allows the model to select a small number of high-purity samples subset $\mathcal{S}_{novel}$ at the beginning of training, and then use these pseudo-labelled novel class samples and the same per-class number of labelled old class samples $\mathcal{S}_{old} = \cup_{c=1}^{|\mathcal{Y}_l|} \mathcal{D}_{l,c}^{\omega}$ as selected labeled training data $\mathcal{S} = \mathcal{S}_{old} \cup \mathcal{S}_{novel}$ for supervised representation learning of the backbone network:

$$\mathcal{L}^{\mathrm{sel}} = \mathcal{L}_{scl}^{\mathrm{sel}} + \lambda_c \mathcal{L}_{ce}^{\mathrm{sel}} \tag{9}$$

$$\mathcal{L}_{scl}^{\mathrm{sel}} = -\frac{1}{|\mathcal{P}_{sel}(\mathbf{x})|} \sum_{\mathbf{z}_i^+ \in \mathcal{P}_{sel}(\mathbf{x})} \log \frac{\exp\left(\mathbf{z}_i^\top \cdot \mathbf{z}_i^+ / \tau^s\right)}{\sum_{j \in \mathcal{B}, j \neq i} \exp\left(\mathbf{z}_i^\top \cdot \mathbf{z}_j / \tau^s\right)} \tag{10}$$

$$\mathcal{L}_{ce}^{\mathrm{sel}} = \frac{1}{|B|} \sum_{i \in B} \ell\left(\boldsymbol{y}_i, \boldsymbol{p}_i\right) \tag{11}$$

where, $\mathcal{L}_{scl}^{\mathrm{sel}}$ is supervised contrastive loss and $\mathcal{L}_{ce}^{\mathrm{sel}}$ is cross-entropy loss, $\lambda_c$ is a balance factor. $\mathbf{x}_i$ is a randomly augmented view of a image in a mini-batch $\mathcal{B}$. $\mathbf{z_i} = \phi(\mathbf{x_i})$ is the $L_2$-normalized feature embedding, $\mathcal{P}_{sel}(\mathbf{x})$ is the postive set of feature embedding $\mathbf{z_i}$. With the benefit of high-quality supervised information of novel classes, the representation capability of the network should be enhanced, which can help the next run of sample selection. We can execute this selection-training strategy iteratively until collect enough samples for each subset. While this sounds appealing, it is unrealistic because we have no prior knowledge of the noise rate of the original sample clusters, which leaves us still unaware of how many samples it is appropriate to select. To solve this problem, we use validation data of old classes as the validation set for the progressive sample selection. At the end of each stage of training, we perform k-means clustering on the validation set and the final number of selected samples can be determined by the best clustering accuracy.

**Clustering refinement learning with selected samples.** Once we have the high-purity sample subset selected from the original novel classes sample cluster generated by existing GCD methods, we can take them as labeled samples of novel classes in GCD retraining. Specifically, for non-parametric GCD methods, since all classes of labeled data could be used in supervised contrastive learning, the network can explore the relationship between all classes, thereby mitigating the bias of the learned representation; for parametric GCD methods, instead of using the pseudo-labels generated by online clustering, we can take the pseudo-label of the selected sample subset as the "ground-truth" label for all the samples of that sample subset, which makes the remaining novel classes unlabelled data can be concentrated on the class centroid generated by the labeled data during the online clustering learning, so that labels can better propagate from labeled data to unlabelled data.

# 5 Experiments

## 5.1 Experimental Setup

**Datasets.** We evaluate our method on two generic datasets (including CIFAR100 [Krizhevsky, 2009] and ImageNet-100 [Deng *et al.*, 2009]) and four challenging fine-grained datasets from Semantic Shift Benchmark [Vaze *et al.*, 2022b] (inculding CUB-200 [Wah *et al.*, 2011], Standford Cars [Krause *et al.*, 2013] and FGVC-Aircraft [Maji *et al.*, 2013] and Oxford-Pet [Parkhi *et al.*, 2012]), and large-scale ImageNet-1K [Deng *et al.*, 2009]. For all training data $D$, following previous work, we split "All" classes into "Old" classes and "New" classes and sample 50% images from base classes and all images from novel classes as unlabelled data $D_u$, while the remaining images are regarded as labelled dataset $D_l$. The dataset statistics are shown in supplemental.

**Evaluation protocol.** Following [Vaze *et al.*, 2022a], we evaluate the clustering accuracy (ACC) on the unlabelled dataset as follows:

$$ACC = \max_{p \in \mathcal{P}(y_u)} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left\{\tilde{y}_i = p\left(\overline{y}_i\right)\right\} \tag{12}$$

Where $\mathcal{P}$ is the set of all permutations that computed with Hungarian algorithm[Kuhn, 1955], $N = |\mathcal{D}_u|$ is the number of samples in the unlabelled dataset, $\tilde{y}_i$ and $\overline{y}_i$ represent the ground-truth label and clustering prediction.

**Implementation details.** Following [Vaze *et al.*, 2022a], the ViT-B-16 pretrained by DINO [Caron *et al.*, 2021] is used as the backbone network for all of our experiments. For the parametric/non-parametric GCD baseline methods that require direct comparison, we perform model training using the same parameter configuration as in the paper to get roughly the same clustering assignment accuracy. In progressive sample selection stage, the network is trained for 120 epochs on each dataset and sample selection is performed every 20 epochs. We use an SGD optimizer with a momentum value of 0.9. We set the initial learning rate to 0.1 and decay it to 0.0001 by cosine annealing every 20 epochs, and then reset the learning rate to the initial value at the start of each sample selection. For all generic and fine-grained image datasets, the value of $K$ is set to 200 and 15 respectively, the value of $\lambda_c$ is set to 1. For all datases, we choose $\gamma$ so that the number of samples selected for each sample subset in the first selection is roughly half the number of per-class labeled samples for the old classes. In the meanwhile, we constraint that the maximum number of samples selected for each sample subset $\omega$ does not exceed the number of per-class labeled samples of the old classes. For the retraining stage, the network is trained for 100 epochs on all datasets. For all datasets, we show the average clustering accuracy of three independent runs. We also provide the performance standard deviation of our method in the supplementary material.

## 5.2 Comparison to the State-of-the-Art

We apply our method to GCD[Vaze *et al.*, 2022a] and SimGCD[Wen *et al.*, 2023], a strong non-parametric baseline and a strong parametric baseline, to compare with the state-of-the-art methods: including k-means [Arthur and Vassilvitskii, 2007], RS+ [Han *et al.*, 2020], UNO+ [Fini *et al.*, 2021], OCRA [Cao *et al.*, 2021], DCCL [Pu *et al.*, 2023], GPC [Zhao *et al.*, 2023]. The clustering accuracy on generic and fine-grained datasets are reported in Table 1. The clustering accuracy on ImageNet-1K dataset are reported in Table 2. "+all" refers to incorporating all samples from cluster assignment to the GCD retraining as the pseudo-labeled sets. "+sel" refers to using our method to select the class-consistent

| Method | CIFAR100 | | | ImageNet-100 | | | CUB | | | Stanford Cars | | | FGVC-Aircraft | | | Oxford-Pet | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | Old | New | All | Old | New | All | Old | New | All | Old | New | All | Old | New | All | Old | New |
| k-means | 52.0 | 52.2 | 50.8 | 72.7 | 75.5 | 71.3 | 34.3 | 38.9 | 32.1 | 12.8 | 10.6 | 13.8 | 16.0 | 14.4 | 16.8 | 77.1 | 70.1 | 80.7 |
| RS+ | 58.2 | 77.6 | 19.3 | 37.1 | 61.6 | 24.8 | 33.3 | 51.6 | 24.2 | 28.3 | 61.8 | 12.1 | 26.9 | 36.4 | 22.2 | – | – | – |
| UNO+ | 69.5 | 80.6 | 47.2 | 70.3 | **95.0** | 57.9 | 35.1 | 49.0 | 28.1 | 35.5 | 70.5 | 18.6 | 40.3 | 56.4 | 32.2 | – | – | – |
| ORCA | 69.0 | 77.4 | 52.0 | 73.5 | 92.6 | 63.9 | 35.3 | 45.6 | 30.2 | 23.5 | 50.1 | 10.7 | 22.0 | 31.8 | 17.1 | – | – | – |
| DCCL | 75.3 | 76.8 | 70.2 | 80.5 | 90.5 | 76.2 | 63.5 | 60.8 | **64.9** | 43.1 | 55.7 | 36.2 | – | – | – | 88.1 | **88.2** | 88.0 |
| GPC | 77.9 | 85.0 | 63.0 | 76.9 | 94.3 | 71.0 | 55.4 | 58.2 | 53.1 | 42.8 | 59.2 | 32.8 | 46.3 | 42.5 | 47.9 | – | – | – |
| GCD | 73.0 | 76.2 | 66.5 | 74.1 | 89.8 | 66.3 | 51.3 | 56.6 | 48.7 | 39.0 | 57.6 | 29.9 | 45.0 | 41.1 | 46.9 | 80.2 | 85.1 | 77.6 |
| + all | 72.5 | 78.6 | 60.1 | 72.8 | 86.1 | 66.2 | 49.0 | 51.9 | 47.6 | 39.3 | 56.0 | 31.3 | 44.0 | 44.5 | 43.8 | 81.2 | 78.3 | 82.7 |
| + sel (Ours) | 74.2 | 76.7 | 69.0 | 75.0 | 85.4 | 69.9 | 52.3 | 55.9 | 50.6 | 41.4 | 58.6 | 33.1 | 45.4 | 50.3 | 42.9 | 83.9 | 82.8 | 84.4 |
| SimGCD | 80.1 | 81.2 | 77.8 | 83.0 | 93.1 | 77.9 | 60.3 | 65.6 | 57.7 | 53.8 | 71.9 | 45.0 | 54.2 | **59.1** | 51.8 | 88.8 | 82.4 | 92.1 |
| + sel (Ours) | **80.7** | **81.4** | **79.1** | **85.4** | 92.8 | **81.7** | **65.1** | **66.2** | 64.5 | **55.8** | **72.2** | **47.9** | **54.7** | 58.4 | **52.8** | **92.2** | 88.0 | **94.4** |

Table 1: Results on generic and fine-grained image recognition datasets.

| Method | ImageNet-1K | | |
|---|---|---|---|
| | All | Old | New |
| GCD | 52.5 | 72.5 | 42.2 |
| SimGCD | 57.1 | **77.3** | 46.9 |
| +sel (Ours) | **58.3** | **77.3** | **48.7** |

Table 2: Results on ImageNet-1K dataset

| Method | CIFAR100 | | | CUB | | |
|---|---|---|---|---|---|---|
| | All | Old | New | All | Old | New |
| None | 80.1 | 81.2 | 77.8 | 60.3 | 65.6 | 57.7 |
| Random | 80.5 | 81.0 | 79.4 | 61.3 | 63.6 | 60.1 |
| Nearest Neighbor w/o constraint | 80.5 | 81.0 | **79.6** | 61.5 | 61.6 | 61.5 |
| Nearest Neighbor | **80.7** | **81.4** | 79.1 | **65.1** | **66.2** | **64.5** |

Table 3: Ablation study of sample selection strategies.

sample subsets from the cluster assignment and incorporating them to the GCD retraining as the pseudo-labeled sets.

In Table 1, it can be seen that SimGCD+sel outperforms all the state-of-the-art methods on both generic and fine-grained datasets. Compared with SimGCD, SimGCD+sel achieved 0.6% and 2.4% improvement on generic CIFAR100 and Imagenet100 datasets in the case of high baseline results and achieved 4.8% and 3.4% improvement on fine-grained CUB and Oxford-Pet datasets. Compared with the non-parametric baseline method GCD and GCD+all, GCD+sel achieves improvement on all six datasets. These results validate the effectiveness of incorporating the selected class-consistent sample subset into GCD retraining. In addition, we can find that the improvement of "All" acc mainly depends on the improvement of "New" acc, which well demonstrates that our method generated less noisy pseudo relationships/labels for novel classes, thus facilitating representation learning.

In Table 2, it can be seen that SimGCD+sel also achieves the best results, which well demonstrates the effectiveness of our method on large scale dataset.

### 5.3 Ablation Study

In this section, we adopt SimGCD as the baseline method and conduct ablation experiments on CIFAR100 and CUB datasets to verify the effectiveness/sensitivity of the main components of our method.

**Different sample selection strategies.** We evaluate the effect of different sample selection strategies. The results using different selection strategies are shown in Table 3. It can be observed that nearest neighbor sample selection with quantitative constraints achieves the best results. In contrast, randomly selecting the same number of samples as the nearest neighbor method as the pseudo-labeled dataset achieves inferior results, which well demonstrates the necessity of nearest neighbor aware label consistency sample selection. Meanwhile, all retraining methods based on selected samples achieve better "New" classes ACC than the baseline, which proves that using the class centroid generated by pseudo-labeled data (even if it contains a certain amount of noise) instead of randomly initialized clustering prototype does help the clustering of novel classes.

| $K$ | CIFAR100 | | | $K$ | CUB | | |
|---|---|---|---|---|---|---|---|
| | All | Old | New | | All | Old | New |
| 50 | 80.6 | 81.2 | **79.6** | 5 | 62.5 | 63.7 | 61.9 |
| 100 | **80.8** | 81.4 | 79.4 | 10 | 64.4 | **67.0** | 63.1 |
| 200 | 80.7 | 81.4 | 79.1 | 15 | **65.1** | 66.2 | 64.5 |
| 250 | 80.7 | **81.7** | 78.8 | 20 | 65.0 | 65.6 | **64.7** |
| 500 | 80.7 | **81.7** | 78.8 | 30 | 61.8 | 62.6 | 61.4 |

Table 4: Ablation study on the value of $K$.

**Impact of $K$ values.** We evaluate the choice of different $K$ values. Multiple sample selection is performed by setting different $K$ values in KNN-based class probability distribution aggregation. The results of SimGCD retrained with samples obtained by setting different $K$ values for sample selection are shown in Table 4. It can be observed that for CIFAR100 dataset, once $K$ is not too small, the final clustering results are not sensitive to the $K$ value. In contrast, the clustering results of CUB dataset are different under different $K$ values. This is because the sample size of each class in CUB dataset is very small, setting too small or too large $K$ value will make the selected samples unreliable, and using these samples for retraining will lead to the decline of clustering results.

| $\mathcal{L}_{scl}^{\text{sel}}$ | $\mathcal{L}_{ce}^{\text{sel}}$ | CIFAR100 | | | | CUB | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | All | Old | New | clean/overall (prec) | All | Old | New | clean/overall (prec) |
| ✓ | | 79.7 | **83.4** | 72.5 | 990/1000 (99%) | 61.7 | 64.5 | 60.4 | 478/598 (80%) |
| | ✓ | 80.4 | 82.8 | 74.5 | 990/1000 (99%) | 62.7 | **66.9** | 60.6 | 812/1093 (74%) |
| ✓ | ✓ | **80.7** | 81.4 | **79.1** | 4326/5000 (87%) | **65.1** | 66.2 | **64.5** | 1015/1386 (73%) |

Table 5: Ablation study of supervised contrastive loss and cross-entropy loss. “clean” refers to the number of clean samples selected. “overall” refers to the number of all samples selected. “prec” refers to percentage.

**The effectiveness of supervised representation learning.** We evaluate the effectiveness of supervised representation learning. Different sample selection is performed by using different losses in supervised representation learning. The results SimGCD retrained with samples obtained by using different losses for sample selection are shown in Table 5. It can be observed that the best results are achieved by using both supervised contrastive loss and cross-entropy loss, this is because the selected samples achieve a better quantity-quality trade-off: the method using two loss terms selects significantly more samples for novel classes than the method using a single loss (supervised contrastive loss) (5000 vs. 1000 for CIFAR100, 1386 vs. 598 for CUB), even though the sample quality decrease (99% vs. 87% for CIFAR100, 80% vs. 73% for CUB); for CUB dataset, comparing with the method using a single loss (cross-entropy loss), the method using two loss terms selects significantly more samples for novel classes (1386 vs. 1093) while maintaining nearly the same sample quality (73% vs. 74%).

| Method | CUB | | | Oxford-Pet | | |
|---|---|---|---|---|---|---|
| | All | Old | New | All | Old | New |
| XCon | 51.0 | 57.8 | 47.6 | 82.1 | 81.7 | 82.4 |
| SimGCD | 61.5 | 66.4 | 59.1 | 82.8 | 87.3 | 80.5 |
| +sel (Ours) | **65.5** | **69.8** | **63.3** | **85.4** | **87.3** | **84.5** |

Table 6: Results on CUB and Oxford-Pet with estimated class number $k$. Following Xcon [Fei *et al.*, 2022], we use the number of classes estimated by the off-the-shelf method in [Vaze *et al.*, 2022a], where $k = 231$ for CUB and $k = 34$ for Oxford-Pet.

## 5.4 Further Analyses

**Sample selection analysis.** To further verify the effectiveness of our sample selection strategy, we perform SimGCD retraining on the CUB dataset using each sample selected during the progressive sample selection process and the final clustering results are shown in Figure 2. Finally, our method uses the samples obtained from the second sample selection as pseudo-labeled set of the novel classes. It can be observed that SimGCD retraining based on our sample selection strategy achieves the best clustering results. After that, with the progress of sample selection training, the model gradually fits noise samples, so that the selected clean samples gradually decrease and the selected noisy samples gradually increase, resulting in the clustering accuracy gradually decreases.

**Performance with estimated class number.** To evaluate the performance of our method in more realistic scenarios, we conduct experiments under the assumption that the number of new classes is unavailable. Specifically, following Xcon [Fei *et al.*, 2022], we first employ an off-the-shelf estimation method from [Vaze *et al.*, 2022a] to obtain the number of classes of all training data, and then use the estimated class number $k$ to perform GCD training. The comparison of our method with SimGCD and Xcon is reported in Table 6. As can be seen, our method achieves the best results and has a significant improvement compare with SimGCD on CUB and Oxford-Pet datasets, which well confirms the robustness of our method in more realistic scenarios.

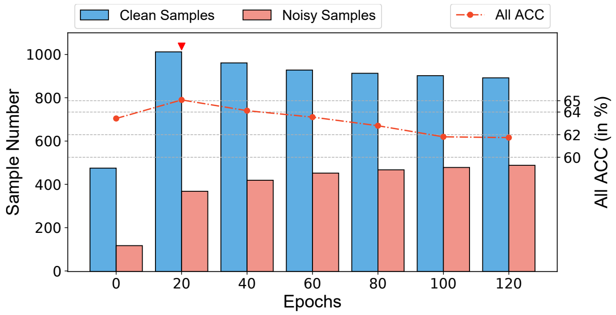

Figure 2: The trend of clustering performance of SimGCD retraining on the CUB dataset, with the sample subsets of novel classes selected each time during the sample selection training. The red inverted triangle indicates the results of our method.

| | CF100 | IN-100 | CUB | Scars | Aircraft | Oxford-Pet | IN-1K |
|---|---|---|---|---|---|---|---|
| Runtime | 0.54s | 1.67s | 0.057s | 0.063s | 0.07s | 0.044s | 158.1s |

Table 7: Runtime of K-nearest neighbor sample selection.

**The computational resource cost of K-nearest neighbor sample selection.** Our KNN sample selection method is performed on faiss-gpu[1] code library, which can achieve more than 100 times acceleration compared to KNN search in sklearn, enabling the proposed method to quickly perform sample selection during the model training. Table 7 reports the runtime for one KNN selection on all training samples.

## 6 Conclusion

In this paper, we propose to generate less noisy pseudo relationships/labels for unlabeled data during GCD training by using a class-consistent sample subset of novel classes. To achieve this, we propose to use a nearest neighbor distance aware label consistency criterion to conduct sample selection. To achieve a good trade-off between the quantity and purity of the final selected samples, we perform progressive sample selection training with selected samples. By using the selected sample as a pseudo-labeled sample set, the pseudo relationships/labels with less noise are generated for the training data in GCD retraining. Extensive experimental results on multiple generic and fine-grained datasets for both parametric and non-parametric GCD methods demonstrate the validity of our proposed method.

[1] https://github.com/facebookresearch/faiss

## Acknowledgments

This work is funded by NSFC (NO. 62206135, 62225604), Young Elite Scientists Sponsorship Program by CAST (2023QNRC001), and the Fundamental Research Funds for the Central Universities (Nankai Universitiy, 070-63233085). Computation is supported by the Supercomputing Center of Nankai University.

## References

[Arazo *et al.*, 2020] Eric Arazo, Diego Ortego, Paul Albert, Noel E O'Connor, and Kevin McGuinness. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.

[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*, pages 1027–1035, 2007.

[Cao *et al.*, 2021] Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-world semi-supervised learning. *arXiv preprint arXiv:2102.03526*, 2021.

[Caron *et al.*, 2018] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 132–149, 2018.

[Caron *et al.*, 2020] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.

[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.

[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.

[Fei *et al.*, 2022] Yixin Fei, Zhongkai Zhao, Siwei Yang, and Bingchen Zhao. Xcon: Learning with experts for fine-grained category discovery. *arXiv preprint arXiv:2208.01898*, 2022.

[Fini *et al.*, 2021] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9284–9292, 2021.

[Fini *et al.*, 2023] Enrico Fini, Pietro Astolfi, Karteek Alahari, Xavier Alameda-Pineda, Julien Mairal, Moin Nabi, and Elisa Ricci. Semi-supervised learning made simple with self-supervised clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3187–3197, 2023.

[Han *et al.*, 2019] 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.

[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 (ICLR)*, 2020.

[Hao *et al.*, 2023] Shaozhe Hao, Kai Han, and Kwan-Yee K Wong. Cipr: An efficient framework with cross-instance positive relations for generalized category discovery. *arXiv preprint arXiv:2304.06928*, 2023.

[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*, pages 770–778, 2016.

[Ji *et al.*, 2019] Xu Ji, Joao F Henriques, and Andrea Vedaldi. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9865–9874, 2019.

[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.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[Krizhevsky, 2009] Alex Krizhevsky. Learning multiple layers of features from tiny images. *Technical report*, 2009.

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

[Laine and Aila, 2016] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.

[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*, page 896, 2013.

[Maji *et al.*, 2013] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013.

[Ortego *et al.*, 2021] Diego Ortego, Eric Arazo, Paul Albert, Noel E O'Connor, and Kevin McGuinness. Multi-objective interpolation training for robustness to label noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6606–6615, 2021.

[Parkhi *et al.*, 2012] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, and CV Jawahar. Cats and dogs. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3498–3505. IEEE, 2012.

[Pu *et al.*, 2023] Nan Pu, Zhun Zhong, and Nicu Sebe. Dynamic conceptional contrastive learning for generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7579–7588, 2023.

[Rizve *et al.*, 2021] Mamshad Nayeem Rizve, Kevin Duarte, Yogesh S Rawat, and Mubarak Shah. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv preprint arXiv:2101.06329*, 2021.

[Rizve *et al.*, 2022] Mamshad Nayeem Rizve, Navid Kardan, and Mubarak Shah. Towards realistic semi-supervised learning. In *European Conference on Computer Vision*, pages 437–455. Springer, 2022.

[Scheirer *et al.*, 2012] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012.

[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. *Advances in neural information processing systems*, 30, 2017.

[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*, pages 7492–7501, 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.

[Wah *et al.*, 2011] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[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.

[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.

[YM. *et al.*, 2020] Asano YM., Rupprecht C., and Vedaldi A. Self-labelling via simultaneous clustering and representation learning. In *International Conference on Learning Representations (ICLR)*, 2020.

[Zhang *et al.*, 2023] Sheng Zhang, Salman Khan, Zhiqiang Shen, Muzammal Naseer, Guangyi Chen, and Fahad Shahbaz Khan. Promptcal: Contrastive affinity learning via auxiliary prompts for generalized novel category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3479–3488, 2023.

[Zhao *et al.*, 2023] Bingchen Zhao, Xin Wen, and Kai Han. Learning semi-supervised gaussian mixture models for generalized category discovery. *arXiv preprint arXiv:2305.06144*, 2023.

[Zhong *et al.*, 2021] Zhun Zhong, Enrico Fini, Subhankar Roy, Zhiming Luo, Elisa Ricci, and Nicu Sebe. Neighborhood contrastive learning for novel class discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10867–10875, 2021.