Towards Debiased Generalized Category Discovery

Pengcheng Guo[1], Yonghong Song[1*] and Boyu Wang[2]
[1]School of Software Engineering, Xi'an Jiaotong University, China
[2]Department of Computer Science and the Brain Mind Institute, University of Western Ontario, Canada
bedlexmunaxl@stu.xjtu.edu.cn, songyh@xjtu.edu.cn, bwang@csd.uwo.ca

Abstract

Generalized Category Discovery (GCD) aims at classifying unlabeled training data coming from old and novel classes by leveraging the information of partially labeled old classes. In this paper, we reveal that existing methods often suffer from *competition* between new and old classes, where the focus on learning new classes often results in a notable performance degradation on the old classes. Moreover, we delve into the reason behind this problem: the GCD classifier can be overconfident and biased towards the new class. With this insight, we propose Debiased GCD (DeGCD), a simple but effective approach that mitigates the bias caused by the overconfidence from new categories by a debiased head. Specifically, we first propose semantic calibration loss that aids the GCD classifier in debiasing by enforcing neighborhood prediction consistency with the latent representation of the debiased head. Furthermore, a debiased contrastive objective is proposed to refine the similarity matrix from the GCD classifier and the debiased classifier, suppressing the overconfidence in new classes in unlabeled data. In addition, an alignment constraint loss is designed to prevent damaging the distribution of the old categories caused by overconfidence in the new categories. Experiments on various datasets shows DeGCD achieves state-of-the-art performance and maintains a good balance between new and old classes. In addition, this method can be seamlessly adapted to other GCD methods, not only to achieve further performance gains but also to effectively balance the performance of the new class with that of the old class.

1 Introduction

Deep learning has achieved remarkable success in the field of image recognition [He *et al.*, 2016; Krizhevsky *et al.*, 2017], typically leveraging enormous labeled training samples. However, these deep learning algorithms rely on enormous labeled training samples to achieve state-of-the-art performance, which is expensive and hinders the widespread application of existing deep learning. Semi-supervised learning (SSL) methods [Berthelot *et al.*, 2019; Guo *et al.*, 2024] provide an effective paradigm to alleviate the burden of labeling by leveraging a large corpus of unlabeled samples. However, one underlying assumption in SSL algorithms is that the labeled data and unlabeled data share the same label space, which is impractical in many real-world scenarios where the unlabeled data contains novel classes and old classes [Han *et al.*, 2019; Han *et al.*, 2020; Han *et al.*, 2021; Fini *et al.*, 2021]. Initially, it was studied as Novel Category Discovery (NCD) [Han *et al.*, 2019], which depends on an disjoint assumption between labeled and unlabeled categories. By contrast, Generalized Category Discovery (GCD) [Vaze *et al.*, 2022] relaxes the disjoint assumption between labeled and unlabeled categories, where the unlabelled data can also contain old and new categories from the labelled data. The goal of GCD [Vaze *et al.*, 2022] is to recognize both old and novel categories from a series of unlabeled data by taking advantage of a small amount of labeled data from old categories.

Existing GCD studies mainly includes two major paradigms: clustering-based methods and representation learning methods. Clustering-based methods [Han *et al.*, 2021; Zhang *et al.*, 2023a; Vaze *et al.*, 2022] employ a semi-supervised k-means [MacQueen, 1967] to recognize both old and new categories. But these works are prone to overfitting to old categories [Vaze *et al.*, 2022]. Representation learning methods [Vaze *et al.*, 2022; Wen *et al.*, 2023; Chiaroni *et al.*, 2023] exploit pre-trained self-supervised models (e.g., DINO [Caron *et al.*, 2021]) and partially fine-tune their parameters on the GCD task to recognize both old and new categories.Although effective in leveraging the strong generalization properties of these representations, they may result in overfitting to new classes. In summary, the two paradigms either overly focus on old classes or excessively prioritize new classes.

In this study, we reveal a intractable challenge in GCD: competition between old and new classes (shown in Fig. 1) . Specifically, in Fig. 1, we can easily find that the performance on the old categories peaks early in training, then declines as the new category's performance improves. To diagnose this issue, we analysis confidence distribution, as shown in Fig. 2. We can observe this phenomenon, the confidence distribution

*Corresponding author: Yonghong Song.

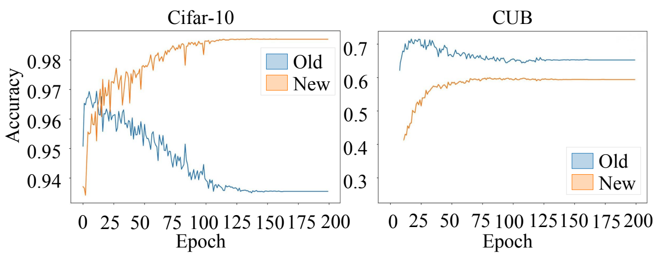

Figure 1: **Accuracy competition of old and new classes in SimGCD on two datasets,** where the performance on the old categories peaks early in training, then declines as the new category's performance improves.

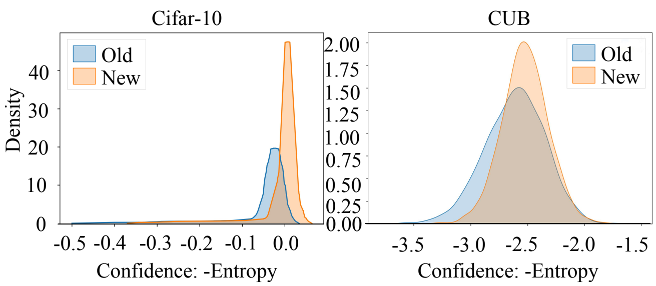

Figure 2: Confidence distribution between new and old categories on SimGCD.

of new classes is higher than that of old classes. This observation leads us to conclude the reason behind the competition is that GCD classifiers are overconfident and biased towards new categories. This over-confidence in new classes, impairing distribution of old classes and resulting in biased predictions. Notably, different from catastrophic forgetting across tasks [Tadros *et al.*, 2022], the competition in GCD occurs within the same task. This finding raises a pivotal question: *Can GCD models achieve unbiased feature learning to alleviate intra-task competitions between new and old classes?*

In this work, to address these issues, we introduce a Debiased Generalized Category Discovery (DeGCD) with a debiased classifier, alleviating competition between old and new categories. To elaborate, we first design a semantic calibration loss to guide the GCD classifier in debiasing by enforcing neighborhood prediction consistency with the latent representation of the debiased head. Second, we propose a debiased contrastive objective to refine the similarity matrix from the GCD classifier and the debiased classifier, suppressing the overconfidence in new classes in unlabeled data. Finally, we introduce an alignment constraint loss to prevent impairing the distribution of the old categories caused by overconfidence in the new categories.

Our contributions are summarized as follows:

- To the best of our knowledge, this is the first work to reveal competition issue in generalised category discovery. We propose a novel DeGCD approach with a debiased classifier to mitigate competition between old and new categories.
- We propose a semantic calibration objective to facilitate debiasing in the GCD classifier by promoting neighborhood prediction consistency with the latent representation of the debiased head.
- We devise a debiased contrastive objective to refine the similarity matrix from the GCD classifier and debiased head, mitigating overconfidence in new classes within unlabeled data.
- We introduce an alignment constraint loss to prevent damaging the distribution of the old categories caused by overconfidence in the new categories.

2 Related Works

Open-Set Semi-Supervised Learning addresses scenarios where unlabeled data may include outliers that do not belong to any of the labeled categories in the training set. The objective is to train a classifier that accurately recognizes the labeled categories despite the presence of noisy, unlabeled data [Han *et al.*, 2021; Han *et al.*, 2019; Chen *et al.*, 2020b; Guo *et al.*, 2020]. Since the primary focus is on the classification performance of the labeled categories, outliers from novel categories are merely identified and rejected, without requiring further classification.

Novel Category Discovery (NCD) [Troisemaine *et al.*, 2023] was initially formalized as deep transfer clustering[Han *et al.*, 2019], aiming to identify unlabeled new classes using knowledge from labeled ones. Han et al.[Han *et al.*, 2019] employed self-supervision for representation learning and utilized ranking statistics for knowledge transfer. Zhong et al.[Zhong *et al.*, 2021] introduced OpenMix, which applies MixUp [Zhang *et al.*, 2018] to blend old and new classes, mitigating overfitting. UNO [Fini *et al.*, 2021] provides a unified objective that jointly optimizes old and new class predictions via swapped prediction [Caron *et al.*, 2020]. NCD relies on assumption that all unlabeled data belong to novel classes.

Generalized Category Discovery (GCD) [Vaze *et al.*, 2022; Cao *et al.*, 2022] is designed to generalize NCD for open-world scenarios where unlabeled instances belong to old and new categories. GCD aims to simultaneously cluster both old and new classes in unlabeled data. Early studies[Vaze *et al.*, 2022; Zhang *et al.*, 2023a] integrate supervised [Khosla *et al.*, 2020] and unsupervised contrastive learning [Chen *et al.*, 2020a], leveraging semi-supervised K-means [MacQueen, 1967] for clustering. Subsequent works [Pu *et al.*, 2023; Zhang *et al.*, 2023b; Zhao *et al.*, 2023a] enhance feature representations by capturing underlying relationships. Recently, some efforts to improve the performance in new categories [Wen *et al.*, 2023; Hao *et al.*, 2024]. Among them, SimGCD [Wen *et al.*, 2023] rethinks the failure of GCD classifiers and proposes a simple method to achieve promising results in new classes. CiPR [Hao *et al.*, 2024] effectively improves the performance of novel classes with new categories by leveraging cross-instance relations and selective clustering. Although GCD has made great advancements [Vaze *et al.*, 2023; Gu *et al.*, 2023], it inherently suffers from competition between old and new classes, which is intractable due to the over-reliance on new class feature on unlabeled data. In this paper, we propose DeGCD to address this challenge.

Entropy Calibration is a crucial advancement in uncertainty quantification, addressing overconfidence issues and improving model interpretability. Recent studies [Zhang *et al.*, 2022; Braverman *et al.*, 2020; Liu *et al.*, 2024] have demonstrated its effectiveness across various applications, from supervised classification to large-scale language modeling. As machine learning models become more integral to high-stakes decision-making, refining entropy calibration techniques will be essential for ensuring reliability and robustness. However, existing methods assume identical label spaces across domains and are unsuitable for GCD, where competition between old and novel categories exists.

3 Preliminaries and Analysis

Here, we briefly introduce the setting and methods of Generalized Category Discovery (GCD) (Sec. 3.1) and give empirical results to reveal inherent issues (Sec. 3.2), which motivates us to propose our DeGCD in Sec. 4.

3.1 Setup and Training Methods of GCD

Problem definition of GCD. Given a labeled data set $D_l = \{(x_i^l, y_i^l)\}_{i=1}^n$ consisting of n samples with labels belonging to C^l classes, C^l is the set of old classes. An unlabeled data set $D_u = \{x_i^u\}_{i=1}^m$ consisting m unlabeled samples, each of which belongs to one of the classes in C^u. C^u is the set of classes in unlabeled training set. Generally, $m \gg n$. Here, $C^l \subset C^u$, then the set of new classes is $C^n = C^u / C^l$. The classes included in the testing set are denoted as C^u. Feature extractor $\mathcal{F}(\cdot)$ is used to extract the ℓ_2-normalized feature by $h_i = \mathcal{F}(x_i)$. Then, a parameter classifier Φ is utilized to obtain class prediction by $z_i = \Phi(h_i)$. The debiased head is composed of two key components: a multi-layer perceptron (MLP) that transforms the feature representation h_i into a new projection space r_i, an auxiliary classifier Ψ that produces predictions based on the projection representation r_i.

Related Training Methods. Generalized category Discovery (GCD) [Vaze *et al.*, 2022] introduces to integrate both supervised [Khosla *et al.*, 2020] and self-supervised [Chen *et al.*, 2020a] contrastive learning paradigms, applying them to the labeled subset B^l and the mini-batch B respectively.

$$\mathcal{L}_{con}^{sup} = \frac{1}{|B^l|} \sum_{i \in B^l} \frac{1}{|\mathcal{N}(i)|} \sum_{q \in \mathcal{N}(i)} \log \frac{\exp(\mathbf{h}_i^\top \mathbf{h}_q' / \tau_s)}{\sum_{n \neq i} \exp(\mathbf{h}_i^\top \mathbf{h}_n' / \tau_s)}, \quad (1)$$

$$\mathcal{L}_{con}^{u} = \frac{1}{|B|} \sum_{i \in B} - \log \frac{\exp(\mathbf{z}_i^\top \mathbf{z}_i' / \tau_u)}{\sum_{n \neq i} \exp(\mathbf{z}_i^\top \mathbf{z}_n' / \tau_u)}. \quad (2)$$

where $\mathcal{N}(i)$ denotes the set of same-class samples in the batch, τ_u and τ_s are temperature hyperparameters. The overall contrastive loss $\mathcal{L}_{con} = (1-\lambda)\mathcal{L}_{con}^u + \lambda \mathcal{L}_{con}^l$.

SimGCD [Wen *et al.*, 2023] proposes a parametric prototype-based classification framework, denoted as $\mathcal{C} = \{\mathbf{c}_1, \cdots, \mathbf{c}_K\}$, where $K = K_{old} + K_{new}$ represents the total number of old and novel classes. The posterior probability distribution is formulated as:

$$\mathbf{p}_i^{(k)} = \frac{\exp(\mathbf{h}_i^\top \mathbf{c}_k)/\tau_p}{\sum_{k'} \exp(\mathbf{h}_i^\top \mathbf{c}_k')/\tau_p}. \quad (3)$$

In SimGCD, self-distillation is applied to two views (random augmentations), alongside entropy regularization $H(\cdot)$ for each sample:

$$\mathcal{L}_{cls}^u = \frac{1}{|B|} \ell(\mathbf{q}_i', \mathbf{p}_i) - \lambda_e H(\overline{\mathbf{p}}), \quad (4)$$

Here, $\mathbf{q}_i'$ is a sharpened probability of another view, and $\overline{\mathbf{p}} = \frac{1}{2|B|} \sum_{i \in B} (\mathbf{p}_i + \mathbf{p}_i')$, $\ell(\cdot)$ represents cross-entropy loss. The supervised loss is also employed on $\mathcal{D}_l$ with labels y_i:

$$\mathcal{L}_{cls}^l = \frac{1}{|B^l|} \sum_{i \in B^l} \ell(y_i, \mathbf{p}_i). \quad (5)$$

3.2 Intractable Problems in GCD

We adopt the state-of-the-art SimGCD [Wen *et al.*, 2023] for confidence [Guo *et al.*, 2017] and accuracy analysis, training under practical low-label conditions to investigate real-world GCD challenges.

GCD suffers from severe competition between old and new classes. As shown in Fig. 1, the accuracy of new categories increases while that of old categories decreases, mainly due to the overconfidence in new classes (Fig. 2), which is a key issue for parameterized GCD classifiers and motivates our proposed strategies.

4 Method Overview

We provide the framework of DeGCD in Fig. 3, which has three main components: 1) A semantic calibration loss aids the GCD classifier's debiasing, aligning the neighborhood prediction of the GCD unbiased classifier with the latent representation of the GCD classifier. 2) A debiased contrastive objective is designed to optimize the similarity matrix between the GCD classifier and the debiased classifier, suppressing the influence of noise for new and old classes within unlabeled data. 3) Alignment constraint loss, which is introduced to prevent damaging the distribution of the old categories caused by overconfidence in the new categories.

4.1 Semantic Calibration Objective

In the GCD task, unlabeled data usually contains a large number of noisy samples [Li *et al.*, 2020], which can easily lead to conflicts between the old and new categories in the shared feature space, and thus cause the GCD classifier to generate erroneous pseudo-labels. To mitigate the effects of bias in the parameters classifier on GCD, we introduce a semantic calibration objective that fosters neighborhood-level predictive coherence by aligning with the underlying representations of the GCD debiased head.

Firstly, we use shared feature extractor $\mathcal{F}(\cdot)$ to derive the ℓ_2-normalized feature by $h_i = \mathcal{F}(x_i)$ as initial representation memory bank $\mathcal{M}$, $x_i \in \mathcal{D}_l \cup \mathcal{D}_u$. Meanwhile, a projection head Φ is utilized to obtain class predicted probability memory bank $\mathcal{P}$ by $p_i = \frac{\exp(\Phi(h_i))}{\sum_{j=1}^{m+n} \exp(\Phi(h_j))}$. The memory bank $\mathcal{M}$ and $\mathcal{P}$ are updated in batches, using a momentum update mechanism defined as:.

$$\mathcal{M} \leftarrow \mu \mathcal{M} + (1-\mu)\mathbf{h}_b, \quad \mathcal{P} \leftarrow \epsilon \mathcal{P} + (1-\epsilon)\mathbf{p}_b. \quad (6)$$

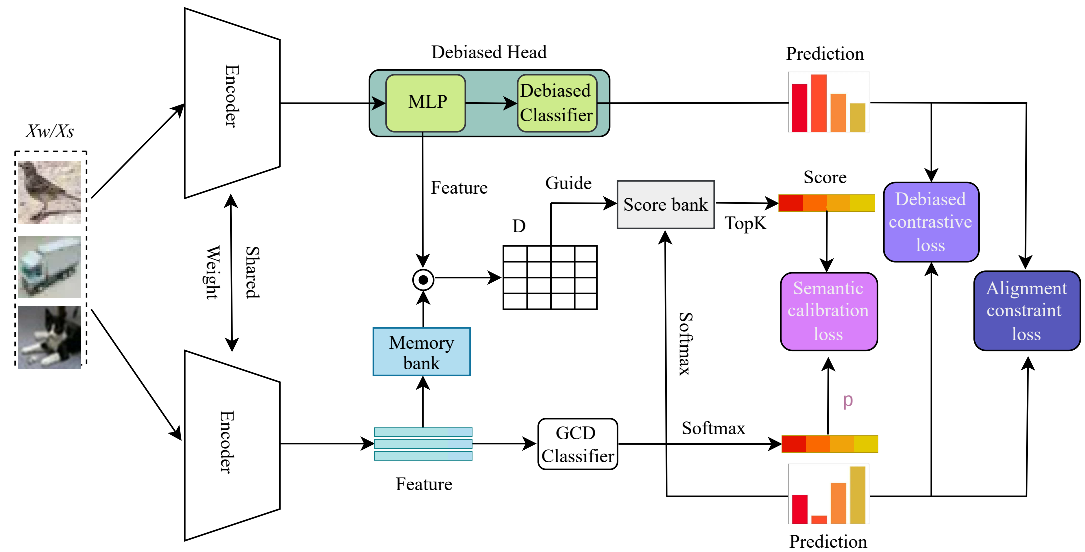

Figure 3: The framework of DeGCD. Models are trained on $\mathcal{D}_l \cup \mathcal{D}_u$ with SimGCD, and select samples in $\mathcal{D}_u$. D is distance matrix.

where $\epsilon, \mu \in [0, 1)$ are the momentum coefficient controlling the update rate. The $\mathbf{h}_b$ and $\mathbf{p}_b$ denote the newly obtained feature representations and predicted probability of the current batch, respectively. Then, to avoid conflicts between the old and new categories in the existing GCD classifier, we design a semantic calibration objective between GCD classifier and debiased classifier. Specially, we first compute similarity distance d_b between representations r_b from the current batch in debiased head and representation memory bank.

$$d_b = \frac{\mathbf{r}_b \mathcal{M}^\top}{\tau}, \tag{7}$$

where τ is the temperature scaling factor, $d_b \in \mathbb{R}^{B \times N}$, where B is the batch size and N is the size of the memory bank. Subsequently, for each feature $r_i \in r_b$. We retrieve its k-nearest neighbors in the representation memory bank $\mathcal{M}$ and express $\mathcal{T}_i$ as the index set of the gathered neighbors:

$$\mathcal{T}_i = \underset{j}{argtop_k}\{d_{i,j} | j = 1, ..., N\}, \tag{8}$$

Following the clustering assumption [Jiang *et al.*, 2022], samples that exhibit locally consistent predictions tend to stay away from decision boundaries, thus they are more likely falling into correct clusters. We employ $\mathcal{T}_i^j$ as the index set to select neighborhood-level class predicted probability:

$$\mathcal{S}_i = \frac{1}{k} \sum_{j=1}^{k} p_{\mathcal{T}_i^j}, \tag{9}$$

Here, $\mathcal{T}_i^j$ represents the index of the j-th neighbor of x_i. The semantic calibration objective is designed to minimize the difference between the current predicted probability p_i and its nearest neighbors $\mathcal{S}_i$.

$$\mathcal{L}_{sa} = \frac{1}{m+n} \sum_{i=1}^{m+n} \mathrm{KL}\left(p_i \,\|\, S_i\right). \tag{10}$$

By minimizing KL divergence, the semantic calibration objective L_{sa} enforces prediction consistency among nearest neighbors, promoting coherent predictions and improving bias mitigation in parameterized classifiers. DeGCD further maintains consistency by continuously updating the representation and probability memory banks during training.

4.2 Debiased Contrastive Objective

Current representation learning methods employing traditional contrastive learning (Eq. 2) aggregate samples solely based on similarity, which misguides unlabeled noise samples into incorrect clusters, thereby amplifying noise-induced uncertainty. To address this issue, we devise a debiased contrastive objective to refine the similarity matrix from the GCD head and debiased head, effectively suppressing the influence of noise for new and old classes within unlabeled data. To be more specific, to adaptability refine reliability and comprehensiveness of the predicted logits, we combine GCD predicted ligits $P_i^g = \Phi(h_i)$ and debiased predicted logits $P_i^d = \Psi(r_i)$ using Gaussian distributions[Stacy, 1962] to mix the controlled formulas.

$$P_i^{gd} = \alpha P_i^g + (1-\alpha) P_i^d \qquad \alpha \sim \text{Gamma}(\beta, \beta). \tag{11}$$

where the mixing factor α is sampled from a Gamma distribution parameterized by the β hyper-parameter. P_i^{gd} is blended predicted logits. Sampling the mixing factor α from a Gamma distribution provides flexibility, robust control over variability and sparsity, and improved regularization, enhancing stability and generalization in probabilistic. Then, we propose the debiased contrastive objective using blended prediction logits to suppress noise effects on old and new classes in unlabeled data. Specifically, we first calculate the softmax-normalized similarity as.

$$\hat{p}_i^{\mathrm{pd}} = \frac{\exp((P_i^{gd} \cdot {P_i^{gd}}^\top)/\eta)}{\sum_{i=1}^{m+n} \exp((P_i^{gd} \cdot {P_i^{gd}}^\top)/\eta)}, \tag{12}$$

Here, η is a temperature parameter, y_i^{gd} is the one-hot ground-truth similarity (with logits of 0 for negative pairs and 1 for the positive pair), and the debiased contrastive loss is defined as the cross-entropy between $\hat{p}_i^{gd}$ and y_i^{dc}.

$$\mathcal{L}_{dc} = \mathbb{E}_{x_i \in \mathcal{D}_l \cup \mathcal{D}_u} \text{H}(y_i^{\text{dc}}, \hat{p}_i^{\text{gd}}). \tag{13}$$

In this way, we pull the instance representation closer to the representation of its predicted class, and push it away from the representations of other classes, using the predicted probabilities and their corresponding pseudo-labels.

4.3 Alignment Constraint Loss for Prediction Refining

In the context of GCD, the learning process on unlabeled data excessively focuses on new categories, leading to insufficient constraints on the distribution of old categories, thereby degrading or conflicting with the discriminative ability of old categories. To prevent damaging the distribution of the old categories caused by overconfidence in the new categories, we design an alignment constraint loss for prediction refining. Specifically, based on the generated predicted logits P_i^d on debiased classifier in Section 4.2, we employ a temperature η and max to P_i^d to obtain a refined pseudo-labels $\overline{P}_i^d = max(\frac{P_i^d}{\eta})$. Finally, we introduce our alignment constraint loss, which applies confidence filtering to mitigate the negative impact of unreliable pseudo-labels. Specifically, only predictions with confidence scores above a threshold η contribute to the loss. Formally, we introduce the alignment constraint function as:

$$\mathcal{L}_{\text{ac}} = \mathbb{E}[\ell_{\text{CE}}(p_i^d, \overline{p}_i^d) \cdot \mathbb{I}(\overline{p}_i^d \geq \eta)]. \tag{14}$$

In this way, we can ensure that learning new classes does not excessively disrupt the distribution of old classes.

4.4 Joint Optimization

The framework of DeGCD is elaborated in Fig. 3. The overall objective of our model is defined as:

$$\mathcal{L}_t = \omega \mathcal{L}_{sa} + \psi \mathcal{L}_{dc} + \omega\, \mathcal{L}_{ac}. \tag{15}$$

where ω and ψ are the weights used to balance the strengths of the three loss functions. More details about ω and ψ in Fig. 4. This work proposes a general debiased method that can be adapted to existing GCD classifier models. For these models, the overall optimization objective is formulated as a combination of the original model's loss function and the loss functions $\mathcal{L}_t$ introduced by the proposed method. This optimization strategy effectively balances the learning of new and old class features to alleviate competitions between them.

5 Experiments

5.1 Experimental Setup

Datasets. We validate the effectiveness of the proposed DeGCD on the generic benchmarks (including CIFAR-10/100 [Krizhevsky *et al.*, 2009] and ImageNet-100 [Deng *et al.*, 2009]), the recently proposed Semantic Shift Benchmark (SSB, including CUB [Wah *et al.*, 2011], Stanford

		Labelled		Unlabelled	
Dataset	Balance	#Image	#Class	#Image	#Class
CIFAR-10 [Krizhevsky *et al.*, 2009]	✓	12.5K	5	37.5K	10
CIFAR-100 [Krizhevsky *et al.*, 2009]	✓	20.0K	80	30.0K	100
ImageNet-100 [Deng *et al.*, 2009]	✓	31.9K	50	95.3K	100
CUB [Wah *et al.*, 2011]	✓	1.5K	100	4.5K	200
Stanford Cars [Krause *et al.*, 2013]	✓	2.0K	98	6.1K	196
Herbarium 19 [Tan *et al.*, 2019]	✗	8.9K	341	25.4K	683

Table 1: Statistics of the datasets we evaluate on.

	CUB			Stanford Cars			Herbarium19		
Method	All	Old	New	All	Old	New	All	Old	New
k-means [Arthur *et al.*, 2007]	34.3	38.9	32.1	12.8	10.6	13.8	13.0	12.2	13.4
RankStats+ [Han *et al.*, 2021]	33.3	51.6	24.2	28.3	61.8	12.1	27.9	55.8	12.8
DTC[Han *et al.*, 2019]	-	-	-	11.8	16.3	16.5	-	-	-
UNO+ [Fini *et al.*, 2021]	35.1	49.0	28.1	35.5	70.5	18.6	28.3	53.7	14.7
GCD [Vaze *et al.*, 2022]	51.3	56.6	48.7	39.0	57.6	29.9	35.4	51.0	27.0
ORCA [Cao *et al.*, 2022]	36.3	43.8	32.6	31.9	42.2	26.9	20.9	30.9	15.5
GPC[Zhao *et al.*, 2023b]	52.0	55.5	47.5	38.2	58.9	27.4	-	-	-
NGCN [Yang *et al.*, 2025]	61.3	60.8	<u>62.1</u>	44.3	58.2	39.1	-	-	-
SimGCD [Wen *et al.*, 2023]	<u>60.3</u>	<u>65.6</u>	57.7	<u>53.8</u>	<u>71.9</u>	<u>45.0</u>	<u>43.0</u>	<u>58.0</u>	<u>35.1</u>
CiPR[Hao *et al.*, 2024]	57.1	58.7	55.6	47.0	61.5	40.1	36.8	45.4	32.6
Ours + SimGCD	64.2	68.4	62.1	**55.3**	**75.2**	**45.7**	**45.9**	**59.3**	**37.4**
Ours + CiPR	**65.3**	**70.9**	**64.1**	51.4	71.3	41.8	41.3	47.5	35.4

Table 2: The results for fine-grained image recognition datasets. Bold values indicate the highest performance, while underlined values denote the second-best results.

Cars [Krause *et al.*, 2013] and the harder Herbarium 19 [Tan *et al.*, 2019]. For each dataset, following [Vaze *et al.*, 2022], we sample a subset of all classes as the labeled ("Old") classes $\mathcal{Y}_l$. Half of the images from these classes form $\mathcal{D}_l$, while the rest are treated as unlabeled data $\mathcal{D}_u$. Table 1 shows the statistics of the datasets we evaluate on.

Evaluation protocol. We evaluate model performance using clustering accuracy (ACC), following standard practice [Vaze *et al.*, 2022]. Given ground truth labels y^* and predicted labels $\hat{y}$, ACC is computed as ACC $= \frac{1}{N} \sum_{i=1}^{N} 1(y_i^* = p(\hat{y}_i))$ where $N = |\mathcal{D}_u|$, and p denotes the optimal permutation aligning predicted clusters with ground truth labels.

Implementation details. Following standard practice in GCD [Vaze *et al.*, 2022], we adopt ViT-B/16 [Dosovitskiy *et al.*, 2021] pre-trained by DINO [Caron *et al.*, 2021] as the backbone, and fine-tune only the last transformer block for all experiments. The output of `[CLS]` token (768-dimensional) serves as the feature representation. We train for 200 epochs using a batch size of 256 and an initial learning rate of 0.1, which follows a cosine decay schedule for each dataset. In this study, the balancing factor ψ is set to 0.35, while the parameter ω is set to 0.2. Following [Wen *et al.*, 2023], the temperature values η is 0.1. η is set to 0.95 follows [Sohn *et al.*, 2020]. Meanwhile, according to [Wen *et al.*, 2023], τ_s is fixed at 0.1, and τ_t is initialized at 0.07, then gradually reduced to 0.04 using a cosine schedule over the first 30 epochs. In this work, we set the Gamma parameter β to 0.5. Following [He *et al.*, 2020], $\epsilon, \mu = 0.99$.

5.2 Comparative Results

To assess the performance of DeGCD, we combine DeGCD with three strong GCD classification methods, including SimGCD [Wen *et al.*, 2023], and CiPR [Hao *et al.*, 2024]. We also compare our models with competitive baselines,

Method	CIFAR-10			CIFAR-100			ImageNet-100		
	All	Old	New	All	Old	New	All	Old	New
k-means [Arthur *et al.*, 2007]	83.6	85.7	82.5	52.0	52.2	50.8	72.7	75.5	71.3
RankStats+ [Han *et al.*, 2021]	46.8	19.2	60.5	58.2	77.6	19.3	37.1	61.6	24.8
DTC[Han *et al.*, 2019]	32.4	42.7	31.8	18.3	31.3	22.9	21.3	25.6	20.8
UNO+ [Fini *et al.*, 2021]	68.6	**98.3**	53.8	69.5	80.6	47.2	70.3	**95.0**	57.9
GCD [Vaze *et al.*, 2022]	91.5	97.9	88.2	73.0	76.2	66.5	74.1	89.8	66.3
ORCA [Cao *et al.*, 2022]	81.8	86.2	79.6	69.0	77.4	52.0	73.5	92.6	63.9
GPC [Zhao *et al.*, 2023b]	90.6	97.6	87.0	75.4	**84.6**	60.1	75.3	93.4	66.7
NGCN [Yang *et al.*, 2025]	96.5	97.6	94.4	74.6	76.5	69.4	78.1	91.3	70.5
SimGCD [Wen *et al.*, 2023]	97.1	95.1	98.1	80.1	81.2	77.8	83.0	93.1	77.9
CiPR [Hao *et al.*, 2024]	97.7	97.5	97.7	81.5	82.4	79.7	80.5	84.9	78.3
Ours + SimGCD	**97.6**	96.3	**98.3**	**83.2**	83.3	**82.9**	**84.8**	94.7	80.1
Ours + CiPR	**98.0**	97.6	98.1	83.0	83.5	80.8	83.0	86.1	**80.9**

Table 3: Results on generic image recognition datasets

i.e., RankStats [Han *et al.*, 2021], UNO [Fini *et al.*, 2021], GCD [Vaze *et al.*, 2022], GPC[Zhao *et al.*, 2023b], ORCA [Cao *et al.*, 2022] and NGCN [Yang *et al.*, 2025].

Comparison on Fine-Grained Datasets. Fine-grained datasets exhibit subtle inter-class differences, making fine-grained visual understanding particularly challenging for GCD. We evaluate the effectiveness of DeGCD by comparing it with existing methods on fine-grained image recognition datasets. As shown in Table 2, DeGCD consistently outperforms all baselines across "All" "Old" and "New" classes. Specifically, for "All" classes, it improves the state-of-the-art performance by 8.3%, 2.8%, and 6.7% on CUB, Stanford Cars, and Herbarium 19, respectively. Moreover, DeGCD achieves notable gains in both "New" and "Old" categories, outperforming the SOTA by 8.1%, 4.5%, and 2.2% on CUB, Stanford Cars, and Herbarium 19 ("Old" category), respectively, demonstrating its effectiveness in capturing unbiased information and reducing competition between categories.

Comparison on Generic Datasets. Table 3 presents a comprehensive comparison, showing that our DeGCD consistently outperforms existing methods on generic image recognition benchmarks. Compared to the prior state-of-the-art (SOTA) approaches, SimGCD or CiPR, DeGCD achieves performance gains of 3.9% on CIFAR-100, 2.2% on ImageNet-100 and 0.3% on CIFAR-10 when considering all categories. The advantage is even more pronounced for "New" classes, with improvements of 4.0%, 3.3%, and 0.2%, respectively. These findings show that the debiased head improves prediction consistency and bias mitigation, leading to more balanced confidence and reduced competition between old and new classes (see Fig. 6 and Fig. 7).

Visualization of the Representation Space. Fig. 5 shows latent representation space of different methods on the CIFAR-10. Fig. 5-(a)-(b) reveals some representations of "New" classes are closer to "Old" classes in SimGCD and CiPR. Fig. 5-(c) demonstrates that the debiased head mitigates the disruption of old class distributions, leading to better separation.

5.3 Ablation Studies

In this section, we conduct comprehensive ablation experiments on CUB and CIFAR-100, as shown in Table 4.

The effect of semantic calibration objective. To explore the benefits of semantic calibration objective, we conduct ablations in Table 4. From the results in Table 4, we can observe that semantic calibration objective has a clear contribution. Specifically, compared with the previous works SimGCD and CiPR, our method obtains consistent improvements in both "Old" and "New" classes (+1.8% and +4.9% in "Old" and "New" on CUB), indicating the importance of semantic calibration. Our approach consistently exceeds baseline on Cifar-100. This improvement proves that it is critical to enhances the consistency between neighborhood predictions and unbiased latent representations for fine-grained classes.

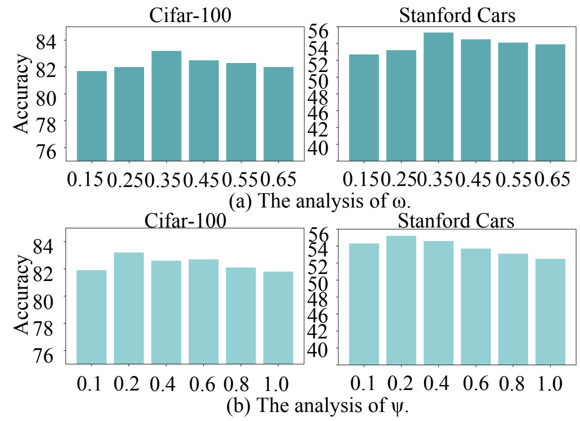

Figure 4: The hyper-parameters analysis.

ID	$\mathcal{L}_{sa}$	$\mathcal{L}_{dc}$	$\mathcal{L}_{ac}$	CIFAR-100			CUB		
				All	Old	New	All	Old	New
(a)	✗	✗	✗	80.1	81.2	77.8	60.3	65.6	57.7
(b)	✓	✗	✗	81.3	81.8	80.1	62.4	66.8	60.5
(c)	✓	✓	✗	82.2	82.3	82.1	63.5	67.7	61.3
(d)	✓	✓	✓	**83.2**	**82.9**	**83.5**	**64.2**	**68.4**	**62.1**

Table 4: Ablations on three key factors, i.e., semantic calibration objective $\mathcal{L}_{sa}$, debiased contrastive objective $\mathcal{L}_{dc}$ and alignment constraint loss $\mathcal{L}_{ac}$ in DeGCD (SimGCD as baseline).

The effect of debiased contrastive objective. As shown in Table 4, we can see that debiased contrastive objective can improve performance. Specifically, compared to baseline, debiased contrastive objective improves accuracy by 1.8% and 1.1% on CUB and CIFAR-100 in all classes, respectively. This improvement highlights the ability of debiased contrastive objective to mitigate the effects of overconfidence in "New" classes within unlabeled data.

The effect of alignment constraint loss. We conduct ablations on two datasets, as shown in Table 4. Specifically, the alignment constraint loss improves accuracy by 1.2% and 1.1% on CUB and CIFAR-100 in all classes, respectively. It can be seen that the alignment constraint loss can prevent damaging the distribution of the "Old" categories caused by overconfidence in the "New" categories.

5.4 Further Analysis

Different Mixed methods in Eq. 6. Our cross-batch mixing mechanism is grounded in momentum-based stability (MoCo[He *et al.*, 2020]) and Mixup[Zhang *et al.*, 2018], operating across batches (vs. Mixup's intra-batch mixing) to enhance the stability of prediction results. The effectiveness of this design is quantitatively validated in Tab. 5.

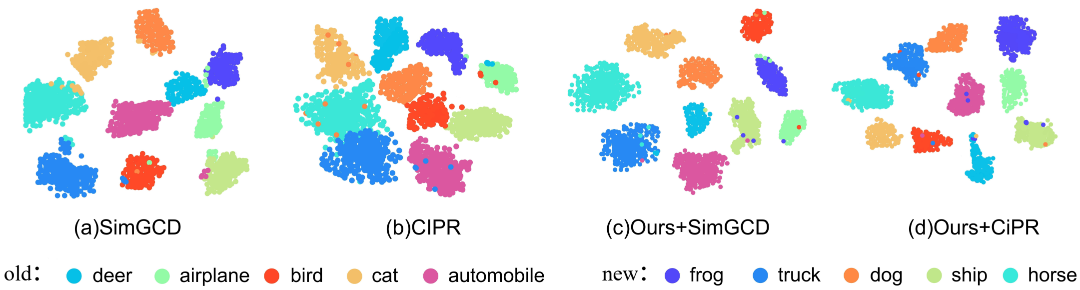

Figure 5: **Visualization on CIFAR-10.** We conduct t-SNE projection on representations extracted by raw SimGCD, CiPR and our approach.

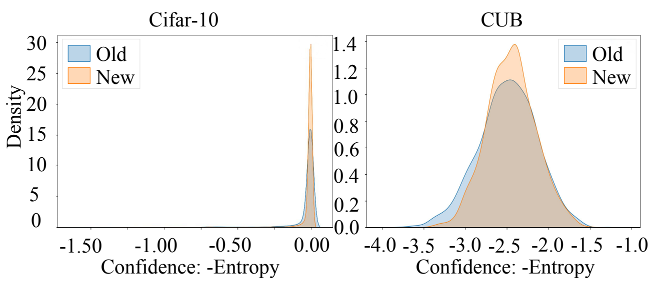

Figure 6: Confidence distribution after DeGCD.

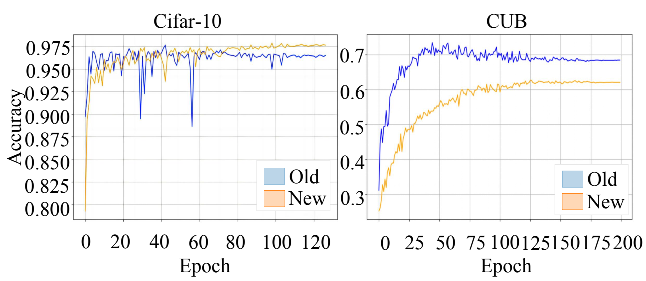

Figure 7: Accuracy after DeGCD.

Analysis of Hyperparameters. We analyze the effect of the Gamma parameter β (Eq. 11) on classification accuracy across datasets (Table 4). On Stanford Cars, "New" accuracy first rises, then slightly drops but stays above SimGCD at 45.5%. On CIFAR-100, it decreases but remains higher than SimGCD (77.9%). To balance maximizing "Old" accuracy while maintaining or surpassing SimGCD on "New", we set $\beta = 0.5$ for both datasets.

We further examine the impact of ω and ψ (Eq. 15) on CIFAR-100 and Stanford Cars (Fig. 4). Optimal results are achieved at $\omega = 0.35$ (Fig. 4-a) and $\psi = 0.2$ (Fig. 4-b).

Evaluation with the number of new class. We also consider the real scenarios with class number $|\mathcal{C}|$ cannot be accessed in advance. We evaluate our DeGCD an off-the-shelf number estimation algorithm [Vaze *et al.*, 2022] to get an estimation of $\mathcal{C}$ in advance, and use it to construct classifiers. As shown in Table 7, DeGCD consistently outperforms SimGCD on both datasets under estimated class numbers.

Method	ImageNet-100	Stanford Cars
Mixup[Zhang *et al.*, 2018]	80.9	48.2
Moco[He *et al.*, 2020]	82.1	52.1
Ours	**84.7**	**55.3**

Table 5: Performance on different mixed methods in Eq. 6.

	Stanford Cars			CIFAR-100		
β	All	Old	New↑	All	Old	New↓
0.2	54.4	67.5	47.3	82.3	80.5	**83.5**
0.3	54.2	69.6	47.5	82.9	82.7	83.3
0.4	55.1	**75.3**	48.1	83.0	82.7	83.0
<u>0.5</u>	**55.3**	75.2	**48.5**	**83.2**	**83.3**	82.9
0.6	54.5	74.7	48.4	82.6	82.5	82.6

Table 6: Ablation study on Gamma parameter β. The <u>underline</u> indicates the selected Gamma parameter.

		CUB			CIFAR-100		
Method	$\vert\mathcal{C}\vert$	All	Old	New	All	Old	New
SimGCD [Wen *et al.*, 2023]	GT (200/100)	60.3	65.6	57.7	80.1	81.2	77.8
DeGCD (Ours)	GT (200/100)	64.2	68.4	62.1	**81.5**	**82.4**	**79.7**
SimGCD [Wen *et al.*, 2023]	Est. (232/108)	61.2	65.8	58.4	80.8	80.7	76.1
DeGCD (Ours)	Est. (231/109)	**65.4**	**70.3**	**62.5**	83.4	81.6	77.4

Table 7: Performance of DeGCD and the baseline SimGCD with an estimated number of categories on CUB and CIFAR-100.

6 Conclusions and Limitations

This study addresses the key challenge of competition between new and old classes in GCD caused by classifier overconfidence. This work proposes a DeGCD, which mitigates bias by introducing a semantic calibration loss to ensure neighborhood prediction consistency and an alignment constraint loss to prevent the distribution of old categories from being affected by overconfidence in new categories. Experiments confirm SOTA performance on various GCD benchmarks, demonstrating robust handling of both new and existing classes. Additionally, it can be easily integrated with other GCD methods to further enhance performance and maintain balance between new and old classes. There are two main limitations. First, while computational overhead is manageable, further acceleration is needed. Second, generalization under extreme data imbalance requires further study.

Acknowledgments

This research is supported in part by National Natural Science Foundation of China (Grant No.: 61973245, 61771199).

References

[Arthur *et al.*, 2007] David Arthur, Sergei Vassilvitskii, et al. k-means++: The advantages of careful seeding. In *Soda*, volume 7, pages 1027–1035, 2007.

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

[Braverman *et al.*, 2020] Mark Braverman, Xinyi Chen, Sham Kakade, Karthik Narasimhan, Cyril Zhang, and Yi Zhang. Calibration, entropy rates, and memory in language models. In *International Conference on Machine Learning*, pages 1089–1099. PMLR, 2020.

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

[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. 2021.

[Chen *et al.*, 2020a] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. 2020.

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

[Chiaroni *et al.*, 2023] Florent Chiaroni, Jose Dolz, Ziko Imtiaz Masud, Amar Mitiche, and Ismail Ben Ayed. Parametric information maximization for generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1729–1739, 2023.

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

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

[Fini *et al.*, 2021] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. 2021.

[Gu *et al.*, 2023] Peiyan Gu, Chuyu Zhang, Ruijie Xu, and Xuming He. Class-relation knowledge distillation for novel class discovery. *lamp*, 12(15.0):17–5, 2023.

[Guo *et al.*, 2017] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.

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

[Guo *et al.*, 2024] Pengcheng Guo, Yonghong Song, Boyu Wang, Jiaohao Liu, and Qi Zhang. Plbr: A semi-supervised document key information extraction via pseudo-labeling bias rectification. *IEEE Transactions on Knowledge & Data Engineering*, (01):1–12, 2024.

[Han *et al.*, 2019] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. 2019.

[Han *et al.*, 2020] K Han, SA Rebuffi, S Ehrhardt, A Vedaldi, and A Zisserman. Automatically discovering and learning new visual categories with ranking statistics. In *Proceedings of the 8th Intennational Conference on Learning Representations, ICLR 2020*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[Han *et al.*, 2021] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Autonovel: Automatically discovering and learning novel visual categories. 2021.

[Hao *et al.*, 2024] Shaozhe Hao, Kai Han, and Kwan-Yee K Wong. Cipr: An efficient framework with cross-instance positive relations for generalized category discovery. *Transactions on Machine Learning Research*, 2024.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. 2016.

[He *et al.*, 2020] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[Jiang *et al.*, 2022] Zhen Jiang, Yongzhao Zhan, Qirong Mao, and Yang Du. Semi-supervised clustering under a “compact-cluster” assumption. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):5244–5256, 2022.

[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. *Advances in neural information processing systems*, 33:18661–18673, 2020.

[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. 2009.

[Krizhevsky *et al.*, 2017] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 2017.

[Li *et al.*, 2020] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*, 2020.

[Liu *et al.*, 2024] Yuchi Liu, Lei Wang, Yuli Zou, James Zou, and Liang Zheng. Optimizing calibration by gaining aware of prediction correctness. *arXiv preprint arXiv:2404.13016*, 2024.

[MacQueen, 1967] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[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.

[Sohn *et al.*, 2020] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. 2020.

[Stacy, 1962] Edney W Stacy. A generalization of the gamma distribution. *The Annals of mathematical statistics*, pages 1187–1192, 1962.

[Tadros *et al.*, 2022] Timothy Tadros, Giri P Krishnan, Ramyaa Ramyaa, and Maxim Bazhenov. Sleep-like unsupervised replay reduces catastrophic forgetting in artificial neural networks. *Nature communications*, 13(1):7742, 2022.

[Tan *et al.*, 2019] Kiat Chuan Tan, Yulong Liu, Barbara Ambrose, Melissa Tulig, and Serge Belongie. The herbarium challenge 2019 dataset. *arXiv preprint arXiv:1906.05372*, 2019.

[Troisemaine *et al.*, 2023] Colin Troisemaine, Vincent Lemaire, Stéphane Gosselin, Alexandre Reiffers-Masson, Joachim Flocon-Cholet, and Sandrine Vaton. Novel class discovery: an introduction and key concepts. *arXiv preprint arXiv:2302.12028*, 2023.

[Vaze *et al.*, 2022] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. 2022.

[Vaze *et al.*, 2023] Sagar Vaze, Andrea Vedaldi, and Andrew Zisserman. No representation rules them all in category discovery. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 19962–19989, 2023.

[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*, pages 16590–16600, 2023.

[Yang *et al.*, 2025] Fengxiang Yang, Nan Pu, Wenjing Li, Zhiming Luo, Shaozi Li, Nicu Sebe, and Zhun Zhong. Learning to distinguish samples for generalized category discovery. In *European Conference on Computer Vision*, pages 105–122. Springer, 2025.

[Zhang *et al.*, 2018] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018.

[Zhang *et al.*, 2022] Jun Zhang, Wen Yao, Xiaoqian Chen, and Ling Feng. Understanding negative calibration from entropy perspective. In *International Symposium on Methodologies for Intelligent Systems*, pages 305–314. Springer, 2022.

[Zhang *et al.*, 2023a] Chuyu Zhang, Ruijie Xu, and Xuming He. Novel class discovery for long-tailed recognition. *Transactions on Machine Learning Research*, 2023.

[Zhang *et al.*, 2023b] 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.*, 2023a] Bingchen Zhao, Xin Wen, and Kai Han. Learning semi-supervised gaussian mixture models for generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16623–16633, October 2023.

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

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