AllGCD: Leveraging All Unlabeled Data for Generalized Category Discovery

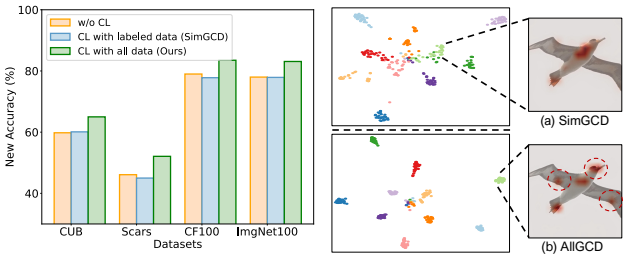
Xinzi Cao[1,2], Ke Chen[1,2], Feidiao Yang[2], Xiawu Zheng[3], Yonghong Tian[2,4], Yutong Lu[1*]
[1] Sun Yat-sen University, [2] Peng Cheng Laboratory, [3] Xiamen University, [4] Peking University
{caoxz, chenk68}@mail2.sysu.edu.cn yangfd@pcl.ac.cn
zhengxiawu@xmu.edu.cn yhtian@pku.edu.cn luyutong@mail.sysu.edu.cn

Abstract

Generalized Category Discovery (GCD) aims to identify both known and novel categories in unlabeled data by leveraging knowledge from labeled datasets. Current methods employ supervised contrastive learning on labeled data to capture known category structures but neglect unlabeled data, limiting their effectiveness in classifying novel classes, especially in fine-grained open-set detection where subtle class differences are crucial. To address this issue, we propose a novel learning approach, ***AllGCD****, which seamlessly integrates* ***all*** *unlabeled data into contrastive learning to enhance the discrimination of novel classes. Specifically, we introduce two key techniques: Intra-class Contrast in Labeled Data (Intra-CL) and Inter-class Contrast in Unlabeled Data (Inter-CU). Intra-CL first refines intra-class compactness within known categories by integrating potential known samples into labeled data. This process refines the decision boundaries of known categories, reducing ambiguity when distinguishing novel categories. Building on this, Inter-CU further strengthens inter-class separation between known and novel categories by applying global contrastive learning to the class distribution in the unlabeled data. By jointly leveraging Intra-CL and Inter-CU, AllGCD effectively improves both intra-class compactness and inter-class separation, effectively enhancing the discriminability between known and novel classes. Experiments demonstrate that AllGCD significantly improves novel classes accuracy,* e.g.*, achieving increases of* ***7.4%*** *on CUB and* ***7.5%*** *on Stanford Cars.*

1. Introduction

Deep learning has achieved outstanding performance in computer vision tasks [4, 7, 15, 31, 32, 36, 40], especially in image classification [14, 16, 20, 33, 34, 36]. However, most conventional methods are developed in a closed setting where all training and test data belong to pre-defined (known) classes. It weakens the capability of these methods to handle novel categories, limiting the scope of their applications [8, 11, 26, 46]. To overcome this issue, the *Generalized Category Discovery* (GCD) paradigm [1, 3, 28, 41, 44, 45, 49, 50] was introduced for open-world learning. It identifies both known and novel categories from unlabeled data by leveraging knowledge learned from labeled data with known classes.



(a) 'New' *acc* w/o CL in SimGCD (b) Attentions w/o unlabeled data

Figure 1. **Comparison of 'New' accuracy and attention w/o unlabeled data in supervised contrastive learning (S-CL).** (a) reveals that S-CL solely with labeled data **fails** to classify novel samples (orange bar *vs.* blue bar) in the baseline SimGCD. In contrast, our method with unlabeled data (green bar) significantly improves accuracy. (b) shows that our AllGCD highlights more class-specific object regions by incorporating unlabeled data.

A key strategy for discovering novel classes is to transfer knowledge from labeled data to promote the learning of novel classes. Based on this idea, previous GCD methods [3, 41, 42, 44, 45] utilize contrastive learning [17, 39] to learn the representations of classification within labeled data. Initially, Vaze *et al*. [41] adopt a two-stage pipeline: contrastive learning for representation, followed by k-means for classifying unlabeled data. However, clustering becomes computationally expensive as data scales. To reduce this, Wen *et al*. [45] proposed SimGCD, a parametric approach that replaces clustering with a classifier and learns class prototypes via joint contrastive learning. Building on this, Wang *et al*. [44] enhanced SimGCD with data prompts for better performance, while Cao *et al*. [3]

*Corresponding author.

addressed SimGCD's limitation in retaining known-class knowledge during transfer by introducing a regularization that improves both known and novel class performance.

Despite advancements in parametric methods, recent studies reveal that supervised contrastive learning (S-CL), a key component of contrastive learning, has a limited effect in discovering novel classes. This is evident from the SimGCD baseline in Fig. 1a, where 'New' accuracy shows little improvement between that without S-CL and S-CL with only labeled data (orange *vs* blue bars) across both fine-grained and general datasets. In particular, accuracy even slightly declines when S-CL is limited to labeled data, as observed in Stanford Cars and CIFAR100. This indicates that current S-CL methods, constrained to labeled data, are inadequate for discovering novel classes, as the labeled data encompass only known classes and omit the novel ones. Therefore, the key to improving the discovery of novel categories is to incorporate unlabeled data into S-CL, as it contains both known and novel samples.

Motivated by this observation, we propose **AllGCD**, a straightforward approach that integrates all unlabeled data into contrastive learning to facilitate the novel category discovery. Specifically, AllGCD comprises two main components: Intra-class Contrast in Labeled Data (Intra-CL) and Inter-class Contrast in Unlabeled Data (Inter-CU). To establish a stronger foundation for Inter-CU, Intra-CL first enhances intra-class compactness within known categories by incorporating potential known samples from unlabeled data into the labeled data. Building on this refined knowledge, Inter-CU further improves inter-class separation between known and novel categories through global contrastive learning on their class distribution in unlabeled data. Finally, by combining Intra-CL and Inter-CU, we significantly improve novel category representations, as shown by the green bar in Fig. 1a.

We validate our approach on six datasets, including three fine-grained datasets (CUB [43], Stanford Cars [18], FGVC-Aircraft [25]), three generic image recognition datasets (CIFAR100 [19] and ImageNet-100 [6]), and the challenging Herbarium-19 dataset [37]. As shown in Fig. 1a, our method outperforms strong baselines in recognizing novel ('New') classes, with a 7.4% improvement over SimGCD on CUB. Additionally, Fig. 1b highlights that AllGCD focuses on more class-specific regions than the baseline. In summary, our main contributions are as follows:

- We propose AllGCD to overcome the limitations of contrastive learning in parametric methods by utilizing all unlabeled data for novel class discovery.
- We introduce Intra-class Contrast in Labeled Data (Intra-CL) to enhance intra-class compactness within known categories by incorporating potential known samples from unlabeled data into the labeled data.
- Building on this refined knowledge, we design Inter-class Contrast in Unlabeled Data (Inter-CU) to improve inter-class separation between known and novel categories by leveraging the global class distribution in unlabeled data.
- Extensive results show that our method significantly improves the performance on novel classes, achieving a 7.4% increase on CUB compared to the baseline.

2. Related Works

Novel Category Discovery (NCD) [12] assumes unlabeled data contains only novel classes and was first formalized as deep transfer clustering, where novel classes are clustered using knowledge from known classes. Various methods address this challenge, including Han *et al*. [12, 13] who use self-supervised learning and ranking statistics for knowledge transfer, Zhong *et al*. [52] who propose a mixup strategy [48] to combine old and novel classes, Gu *et al*. [10] who introduce a knowledge distillation framework for regularizing novel class learning, Zang *et al*. [47] who transfer knowledge from a source domain to a label-scarce target domain without label constraints, and Liu *et al*. [23] who apply NCD in ultra-fine-grained visual categorization.

Generalized Category Discovery (GCD) [41] extends NCD by assuming that unlabeled data also contains known class images. Vaze *et al*. [41] formalize this with contrastive learning in representation space during training and clustering at inference. To reduce clustering cost, Wen *et al*. [45] introduce SimGCD, using a parametric classifier to learn class representation centers, serving as a strong baseline. Cao *et al*. [3] note SimGCD's issue of forgetting known classes while learning novel ones, proposing entropy regularization for potential known samples. Wang *et al*. [44] further extend SimGCD by optimizing model and data parameters for better alignment with the GCD task. Despite their performance, these methods rely on limited labeled images, constraining novel class discovery.

Contrastive learning (CL) spans unsupervised [39] and supervised [17] variants for learning from unlabeled and labeled data, respectively. Unsupervised CL treats two augmented views of the same image as positives and all others as negatives, while supervised CL (S-CL) considers samples from the same class as positives. Vaze *et al*. [41] first combine both in GCD to enhance image representations. Choi *et al*. [5] employ mean-shifted embeddings for CL, and Pu *et al*. [28] propose DCCL, which explores conceptual-level CL to model relations between known and novel categories. However, these methods primarily apply S-CL within non-parametric GCD frameworks. In contrast, recent parametric approaches [3, 44, 45] demonstrate that S-CL can effectively train classifiers by learning class prototypes. Nonetheless, these works still rely heavily on limited labeled data, underutilizing the rich supervision signal from abundant unlabeled samples, which ultimately constrains the full potential of contrastive learning.

3. Preliminaries and Analysis

3.1. GCD setting and training

Problem Formulation. Unlike traditional closed-set classification, where training and test classes coincide ($\mathcal{Y}_l$), GCD aims to identify novel classes $\mathcal{Y}_u$ in the test data. Given labeled dataset $\mathcal{D}^l = \{(\boldsymbol{x}_i, y_i)\} \subset \mathcal{X} \times \mathcal{Y}_l$ and unlabeled dataset $\mathcal{D}^u = \{(\boldsymbol{x}_i, y_i)\} \subset \mathcal{X} \times \mathcal{Y}_u$, GCD seeks to classify both known and novel samples in $\mathcal{D}^u$ using knowledge from $\mathcal{D}^l$. The total class number $K = |\mathcal{Y}_l \cup \mathcal{Y}_u|$ is assumed known, following prior work [3, 9, 13, 44, 45].

Architecture. SimGCD [45] consists of a Vision Transformer (ViT) encoder [7], denoted as Φ, along with a classification head f to generate pseudo-labels for self-distillation [2, 4], and a multi-layer perception (MLP) projection head g for representation contrastive learning.

Representation Learning. To learn representations for distinguishing samples, Vaze *et al*. [41] apply unsupervised contrastive learning across all data and supervised contrastive learning (S-CL) on labeled data. Given two views, $\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$, of an identical image as a positive pair in a mini-batch B, the unsupervised contrastive loss encourages their representations to be close in an unsupervised manner:

$$\mathcal{L}^u_{\text{rep}} = \frac{1}{|B|} \sum_{i \in B} -\log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}'_i / \tau_u\right)}{\sum_{n \in B, n \neq i} \exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}'_n / \tau_u\right)}, \quad (1)$$

where $\boldsymbol{z}_i = g \circ \Phi(\boldsymbol{x}_i)$ and τ_u is a temperature. Meanwhile, the supervised contrastive loss minimizes the distance between same-class samples in feature space, defined as:

$$\mathcal{L}^s_{\text{rep}} = \frac{1}{|B^l|} \sum_{i \in B^l} \frac{1}{|\mathcal{N}_i|} \sum_{q \in N_i} -\log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}'_q / \tau_c\right)}{\sum_{n \in B, n \neq i} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}'_n / \tau_c)}, \quad (2)$$

where B^l is the labeled subset of B, and $\mathcal{N}_i$ is the set of images with the same label as $\boldsymbol{x}_i$. The overall contrastive learning loss for representation is:

$$\mathcal{L}_{\text{rep}} = (1 - \lambda)\mathcal{L}^u_{\text{rep}} + \lambda \mathcal{L}^s_{\text{rep}}, \quad (3)$$

where λ controls the balance between unsupervised and supervised contrastive learning.

Parametric Clustering. Instead of the time-consuming k-means, Wen *et al*. [45] propose an efficient parametric classifier f using a set of prototypes $\mathcal{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_K\}$, where K is the total classes number. For an image view $\boldsymbol{x}_i$, the prediction is made by calculating the cosine similarity between its feature, $\Phi(\boldsymbol{x}_i)$, and each prototype $\boldsymbol{c}_k$:

$$\boldsymbol{p}_i^{(k)} = \frac{\exp\left(\cos\left(\Phi\left(\boldsymbol{x}_i\right), \boldsymbol{c}_k\right) / \tau_s\right)}{\sum_{k'} \exp\left(\cos\left(\Phi\left(\boldsymbol{x}_i\right), \boldsymbol{c}_{k'}\right) / \tau_s\right)}. \quad (4)$$

Similarly, the soft pseudo-labels $\boldsymbol{q}'_i$ for another view $\boldsymbol{x}'_i$ are obtained using a smaller τ_t, and the cross-entropy loss is applied for prototype learning using these pseudo-labels $\boldsymbol{q}'_i$:

$$\mathcal{L}^u_{\text{dis}} = -\frac{1}{|B|} \sum_{i \in B} \sum_k \boldsymbol{q}'^{(k)} \log \boldsymbol{p}^{(k)}. \quad (5)$$

Additionally, SimGCD employs a class mean entropy regularization [2], $H(\overline{\boldsymbol{p}}) = -\sum_k \overline{\boldsymbol{p}}^{(k)} \log \overline{\boldsymbol{p}}^{(k)}$, where $\overline{\boldsymbol{p}}$ is the batch-wise class-average prediction, forcing the model to focus on novel classes. The total classifier is:

$$\mathcal{L}_{\text{cls}} = \lambda \mathcal{L}^s_{\text{cls}} + (1 - \lambda)\left(\mathcal{L}^u_{\text{dis}} - \varepsilon H\right), \quad (6)$$

where $\mathcal{L}^s_{\text{cls}}$ is the cross-entropy loss for labeled data.

3.2. Limited contrastive learning

Analysis. Although contrastive learning (CL) enhances representation discriminability of known classes in SimGCD, it applies only to labeled data, as shown in the upper-left corner of Fig. 2. In the upper-left corner, same-colored entries along the vertical and horizontal indices correspond to samples from the same class, while different colors denote distinct classes. Self-contrast cases (blue-filled) on the diagonal are masked out. Obviously, abundant unlabeled data, shown with a gray background, are excluded from the CL due to the absence of class labels. As a result, the model lacks direct contrastive supervision on novel samples, making it harder to separate them in feature space.

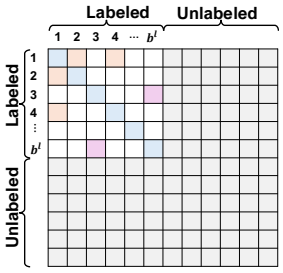


Figure 2. **A class mask**: Samples from the same class are shown in the same color in a batch B.

4. Method

To overcome the limited contrastive learning issue discussed in subsection 3.2, we propose a straightforward approach called AllGCD to integrate supervised contrastive learning across all unlabeled data directly. Specifically, we detail our proposed Intra-class Contrast in Labeled Data (Intra-CL) in subsection 4.1 and Inter-class Contrast in Unlabeled Data (Inter-CU) in subsection 4.2.

4.1. Intra-class Contrast in Labeled Data

Intra-CL first strengthens intra-class compactness within known classes for better separation of known and novel categories in Inter-CU. To achieve this, we use baseline SimGCD [45] to identify potential known samples in unlabeled data and incorporate them into the labeled set. However, as shown by the orange lines in Fig. 4a and Fig. 4b, SimGCD's known-class accuracy declines sharply during training, compromising the quality of potential known

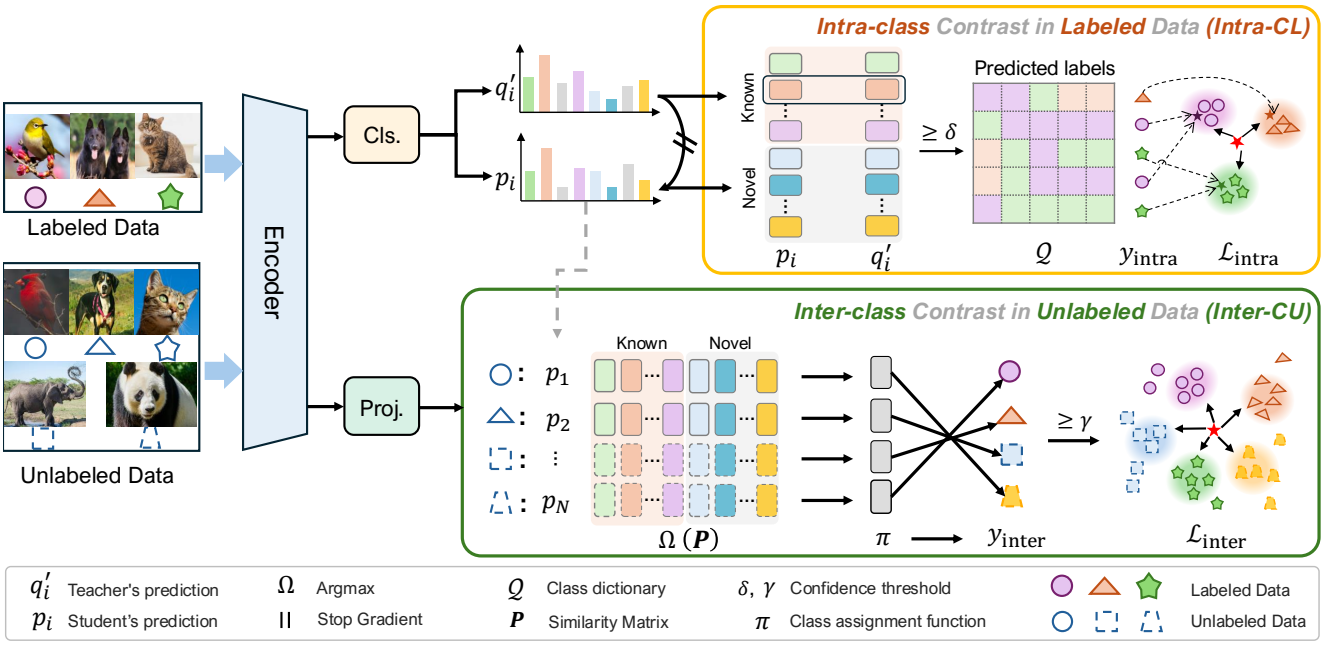


Figure 3. **Illustration of the proposed AllGCD.** It is composed of Intra-CL and Inter-CU. First, Intra-CL (Sec. 4.1) enhances knowledge of known classes by incrementally collecting reliable pseudo-labels $\mathcal{Y}^*$ within unlabeled data. Once this knowledge is strengthened, Inter-CU (Sec. 4.2) further promotes the separation of known and novel classes through contrastive learning between them.

samples. To overcome this, we propose a voting strategy that utilizes the model's prior knowledge to identify potential known samples early. Specifically, Intra-CL comprises **three steps**: (1) Selecting potential known samples within unlabeled data as in [3]; (2) Storing these known samples into a dictionary $\mathcal{Q}$; (3) Voting pseudo-labels $\mathcal{Y}^*$ from $\mathcal{Q}$ based on their most frequent label across training iterations.

Step (1): Selecting potential known samples within unlabeled data.

In this step, we identify potential known samples by selecting instances where both views of the same image exhibit high confidence (δ) and are classified as known categories ($\mathcal{Y}_l$). The specific calculation process is as follows.

Let B^u denote the unlabeled subset of B, where each image in this batch has a unique index $\mathcal{U}_B = \{u_1, u_2, \ldots, u_b\}$. Given the predictions $\boldsymbol{p}_i$ and $\boldsymbol{q}_i'$ from two views of an image in B^u, as defined in Eq. (4). We first identify **high-confidence** samples by selecting those with predictions $\boldsymbol{p}_i$ and $\boldsymbol{q}_i'$ both exceed a confidence threshold δ: $\mathbf{M}_\delta = \mathbb{1}\left(\max(\boldsymbol{p}_i) > \delta \wedge \max(\boldsymbol{q}_i') > \delta\right)$.

Next, we classify a sample as **known** only if the predicted label from the two views match and both belong to known class set $\mathcal{Y}_l$: $\mathbf{M}_{\text{known}} = \mathbb{1}((\hat{y}_i = \hat{y}_i') \wedge (\hat{y}_i \in \mathcal{Y}_l))$.

Finally, the selected known samples within the unlabeled subset B^u are determined as:

$$\hat{\mathcal{Y}}_{\text{select}} = (\mathbf{M}_\delta \odot \mathbf{M}_{\text{known}}) \odot \hat{\mathcal{Y}}, \tag{7}$$

where $\mathbf{M}_\delta$ and $\mathbf{M}_{\text{known}}$ are binary (0-1) vectors of length b, $\odot$ is element-wise multiplication, and $\hat{\mathcal{Y}} = \{\hat{y}_i \,|\, i \in [1, b]\}$ represents the predicted labels across the batch of unlabeled samples. Thus, $\hat{\mathcal{Y}}_{\text{select}}$ identifies the known samples ($\mathbf{M}_{\text{known}} = 1$) that also have high confidence ($\mathbf{M}_\delta = 1$).

Step (2): Storing above predicted known samples.

To leverage the model's early ability to recognize known classes, we initialize a dictionary $\mathcal{Q}$ to store the pseudo-labels $\hat{\mathcal{Y}}_{\text{select}}$ of each image based on its unique index $\mathcal{U}_B$.

Concretely, for each unlabeled subset B^u, we update the dictionary $\mathcal{Q} : \mathcal{U}_B \rightarrow \hat{\mathcal{Y}}_{\text{select}}$ as:

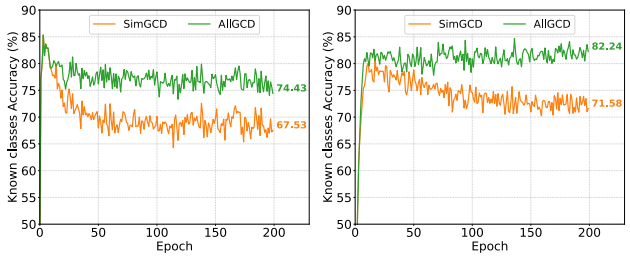
$$\mathcal{Q}(u_i) = \begin{cases} \mathcal{Q}(u_i) \cup \{\hat{y}_i \in \hat{\mathcal{Y}}_{\text{select}}\} & \text{if } u_i \in \text{dom}(\mathcal{Q}), \\ \{\hat{y}_i \in \hat{\mathcal{Y}}_{\text{select}}\} & \text{if } u_i \notin \text{dom}(\mathcal{Q}). \end{cases} \tag{8}$$

Step (3): Voting on pseudo-labels for unlabeled data based on frequency.

To obtain a reliable pseudo-label for each unlabeled image, we apply a voting mechanism over the stored pseudo-labels $\mathcal{Q}$ across training batches. Specifically, for a batch of unlabeled samples, we select the most frequently assigned pseudo-label from its historical records in $\mathcal{Q}(u_i)$:

$$\mathcal{Y}^* = \arg\max_y \left(\text{freq}(\hat{y} \in \mathcal{Q}(u_i)), \forall u_i \in \mathcal{U}_B\right), \tag{9}$$

where $\mathcal{Y}^* = \{y_1^*, y_2^*, \ldots, y_b^*\}$ denotes the pseudo-labels for unlabeled samples in B^u. The updated mini-batch labels, incorporating both original labeled data $\mathcal{Y}_l$ and pseudo-labels $\mathcal{Y}^*$, are given by $\mathcal{Y}_{\text{intra}} = \mathcal{Y}_l \cup \mathcal{Y}^*$. Consequently,



(a) CUB (b) Stanford Cars

Figure 4. **Known-class accuracy on unlabeled data during training.** SimGCD, using S-CL without unlabeled data, shows a sharp drop on CUB and Stanford Cars. In contrast, Intra-class leverages potential known samples in unlabeled data, improving the knowledge of known classes.

the supervised contrastive loss in Eq. (2) is updated with pseudo known samples from unlabeled data as:

$$\mathcal{L}_{\text{intra}} = \frac{1}{|B^{\text{intra}}|} \sum_{i \in B^{\text{intra}}} \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} -\log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}'_j / \tau_c\right)}{\sum_n \mathbb{1}_{[n \neq i]} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}'_n / \tau_c)}, \tag{10}$$

where B^{intra} is the newly labeled subset of mini-batch B with labels $\mathcal{Y}_{\text{intra}}$, and $\mathcal{N}_i$ indexes same-label images as $\boldsymbol{x}_i$ in the batch. After applying Intra-CL, the accuracy of known classes in unlabeled data remains more stable (Fig. 4), indicating more compact representations within known classes.

4.2. Inter-class Contrast in Unlabeled Data

While Inter-CL enhances discrimination among known classes, it lacks contrastive interactions with novel classes, limiting novel category accuracy (Fig. 1a). To address this, we propose Inter-CU, which enforces global contrastive learning on the unlabeled class distribution. Specifically, once the model has stabilized on known and novel categories from epoch e, we use the Hungarian algorithm [21] on the predictions of all unlabeled data to optimize the class distribution of known and novel categories. To ensure a high-quality distribution, we retain only high-confidence predictions ($\boldsymbol{P} > \gamma$) for global contrastive learning.

Formally, given the unlabeled dataset $\mathcal{D}^u = \{(\boldsymbol{x}_i, y_i)\} \in \mathcal{X} \times \mathcal{Y}_u$, we obtain classifier predictions $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N]$ via Eq. (4), where N is the total number of unlabeled samples. The minimum-cost assignment function π is defined as:

$$\pi^* = \arg\min_{\pi} \sum_{i=1}^{N} -P_{i,\pi(i)}, \quad \pi(i) \in \{1, 2, \ldots, K\}, \tag{11}$$

where π^* assigns each sample in $\mathcal{D}^u$ to one of K categories, minimizing the sum of costs $-P_{i,\pi(i)}$. The Hungarian algorithm efficiently solves this assignment problem by finding the optimal mapping with minimal total cost. Therefore, the final assignments for unlabeled data are:

$$\mathcal{Y}_{\text{inter}} = \{\hat{y}_i = \pi^*(i) \mid i = 1, 2, \ldots, N\} \odot \mathbf{M}_{\text{inter}}, \tag{12}$$

where $\mathbf{M}_{\text{inter}}$ is a binary mask that applies a confidence threshold γ to filters out low-confidence predictions: $\mathbf{M}_{\text{inter}} = \{i \mid \max(\boldsymbol{p}_i) > \gamma\}$. Only high-confidence samples contribute to the global contrastive loss:

$$\mathcal{L}_{\text{inter}} = \frac{1}{|B^{\text{inter}}|} \sum_{i \in B^{\text{inter}}} \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} -\log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}'_j / \tau_c\right)}{\sum_n \mathbb{1}_{[n \neq i]} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}'_n / \tau_c)}, \tag{13}$$

where B^{inter} corresponds to the refined labeled subset with pseudo-labels $\mathcal{Y}_{\text{inter}}$. This approach enables global contrastive learning on both known and novel samples, enhancing category discovery, particularly for novel classes.

Overall. Utilizing all unlabeled data in supervised contrastive learning significantly boosts class discovery, especially for novel classes. This overall training loss is updated as:

$$\mathcal{L}_{\text{all}} = (1 - \lambda)\mathcal{L}_{\text{intra}} + \lambda \mathcal{L}_{\text{inter}} + \mathcal{L}_{\text{cls}}, \tag{14}$$

where λ controls the trade-off between Intra-CL and Inter-CU. Algorithm 1 details one training step of AllGCD.

Algorithm 1 One Training Procedure for AllGCD

Input: Encoder Φ, classifier head f, projector g, labeled data $\mathcal{D}^l$, unlabeled data $\mathcal{D}^u$, hyperparameters δ, e.
Output: Φ, f, and g;

Initialize dictionary $\mathcal{Q} \leftarrow \emptyset$;
for epoch = 1 to total epochs **do**
  if epoch $\geq e$ **then**
    Get assignment $\mathcal{Y}_{\text{inter}}$ for D^u using Eq. (12);
  end if
  for $(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{D}^l \cup \mathcal{D}^u, y \in \mathcal{Y}_l, y_{\text{inter}} \in \mathcal{Y}_{\text{inter}}$ **do**
    $\boldsymbol{z} \leftarrow g \circ \Phi(\boldsymbol{x}), \boldsymbol{z}' \leftarrow g \circ \Phi(\boldsymbol{x}')$;
    $\boldsymbol{p} \leftarrow f \circ \Phi(\boldsymbol{x}), \boldsymbol{q}' \leftarrow f \circ \Phi(\boldsymbol{x}')$;
    Compute $\mathcal{L}^s_{\text{cls}}(\boldsymbol{p}, y) + \mathcal{L}^s_{\text{cls}}(\boldsymbol{q}', y)$; ▷ For $\mathcal{D}^l$
    Compute $\mathcal{L}^u_{\text{dis}}(\boldsymbol{p}, \boldsymbol{q}')$ using Eq. (5);
    Compute Max class Entropy $H(\boldsymbol{p})$;
    Compute $\mathcal{L}_{\text{cls}} \leftarrow \lambda \mathcal{L}^s_{\text{cls}} + (1 - \lambda)(\mathcal{L}^u_{\text{dis}} - \varepsilon H)$;
    Step 1: Generate $\hat{y} \leftarrow (\boldsymbol{p}, \boldsymbol{q}', \delta)$; ▷ For $\mathcal{D}^u$
    Step 2: Store $\hat{y}$ into $\mathcal{Q}$ using Eq. (8);
    Step 3: Vote for y^* from $\mathcal{Q}$ using Eq. (9);
    Combine with pseudo-labels $y_{\text{intra}} \leftarrow y \cup y^*$;
    Compute $\mathcal{L}^u_{\text{rep}}(\boldsymbol{z}, \boldsymbol{z}')$ using Eq. (1);
    Compute $\mathcal{L}_{\text{intra}}(\boldsymbol{z}, \boldsymbol{z}', y_{\text{intra}})$ using Eq. (10);
    Compute $\mathcal{L}_{\text{inter}}(\boldsymbol{z}, \boldsymbol{z}', y_{\text{inter}})$ using Eq. (12);
    $\mathcal{L}_{\text{all}} = (1 - \lambda)\mathcal{L}^u_{\text{rep}} + \lambda(\mathcal{L}_{\text{intra}} + \mathcal{L}_{\text{inter}}) + \mathcal{L}_{\text{cls}}$;
    Update Φ, f, and g by SGD [29];
  end for
end for

Table 1. Comparison with State-of-the-Art GCD Methods, Evaluated *with* and *without* Ground-Truth K for Clustering. **Bold** denotes the best results.

Method	CIFAR100			ImageNet-100			CUB			Stanford Cars			FGVC-Aircraft			Herbarium 19		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
*(a) Clustering **with** the ground-truth number of classes K given*																		
k-means[24]	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	13.0	12.2	13.4
RS+[13]	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	27.9	55.8	12.8
UNO+[9]	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	28.3	53.7	14.7
ORCA[22]	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	20.9	30.9	15.5
GCD [41]	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	35.4	51.0	27.0
DCCL [28]	75.3	76.8	70.2	80.5	90.5	76.2	63.5	60.8	64.9	-	-	-	43.1	55.7	36.2	-	-	-
PromptCAL [50]	81.2	84.2	75.3	83.1	92.7	78.3	62.9	64.4	62.1	50.2	70.1	40.6	52.2	52.2	52.3	37.0	52.0	28.9
SimGCD [45]	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	44.0	58.0	36.4
InfoSieve [30]	78.3	82.2	70.5	80.5	93.8	73.8	**69.4**	**77.9**	**65.2**	55.7	74.8	46.4	56.3	63.7	52.5	41.0	55.4	33.2
SPTNet [44]	81.3	84.3	75.6	85.4	93.2	81.4	65.8	68.8	65.1	59.0	**79.2**	49.3	**59.3**	61.8	**58.1**	43.4	**58.7**	35.2
LegoGCD [3]	81.8	81.4	**82.5**	86.3	94.5	82.1	63.8	71.9	59.8	57.3	75.7	48.4	55.0	61.5	51.7	45.1	57.4	**38.5**
CMS [5]	82.3	**85.7**	75.5	84.7	**95.6**	79.2	68.2	76.5	64.0	56.0	63.4	52.3	56.9	**76.1**	47.6	36.4	54.9	26.4
AllGCD (ours)	**82.3**	82.3	82.2	**86.5**	94.7	**82.3**	68.4	75.1	65.1	**60.5**	77.0	**52.5**	57.4	62.4	54.9	**45.2**	58.3	38.0
*(b) Clustering **without** the ground-truth number of classes K given*																		
GCD [41]	70.8	77.6	57.0	77.9	91.1	71.3	51.1	56.4	48.4	39.1	58.6	29.7	-	-	-	37.2	51.7	29.4
GPC [51]	75.4	84.6	60.1	75.3	93.4	66.7	52.0	55.5	47.5	38.2	58.9	27.4	43.3	40.7	44.8	36.5	51.7	27.9
SimGCD [45]	80.1	81.2	77.8	81.7	91.2	76.8	61.5	66.4	59.1	49.1	65.1	41.3	-	-	-	-	-	-
CMS [5]	79.6	83.2	72.3	81.3	**95.6**	74.2	64.4	68.2	62.4	51.7	68.9	43.4	55.2	60.6	52.4	37.4	56.5	27.1
AllGCD (ours)	**81.5**	**83.3**	**78.4**	**84.4**	94.3	**80.2**	**66.7**	**72.1**	**64.0**	**58.8**	**75.7**	**50.6**	**56.7**	**64.2**	**53.0**	**45.7**	**58.7**	**38.5**

5. Experiments

5.1. Experimental Setup

Datasets. We evaluate our approach on three fine-grained datasets: CUB [43], Stanford Cars [18], and FGVC-Aircraft [25], as well as two generic image recognition datasets, CIFAR100 [19] and ImageNet-100 [6]. Additionally, we test on a more challenging long-tailed fine-grained Herbarium 19 [37]. Following [41], we randomly sample 50% of seen categories as labeled data $\mathcal{D}^l$, with the remaining images from seen and novel categories forming the unlabeled set $\mathcal{D}^u$. Details are in the supplementary material.

Evaluation protocol. We use clustering accuracy (*ACC*) to evaluate model performance, following [41, 45]. During evaluation, we compare the ground-truth labels y_i with the predicted labels $\hat{y}_i$, and calculate *ACC* as $ACC = \frac{1}{|\mathcal{D}^u|}\sum_{i=1}^{|\mathcal{D}^u|}\mathbb{1}\left(y_i = p\left(\hat{y}_i\right)\right)$, where p is determined using the Hungarian optimal assignment algorithm [21].

Implementation details. AllGCD is built upon SimGCD [45], using a ViT-B/16 backbone [7] pre-trained with DINO [4]. We use the class token to train a classifier and fine-tune only the last attention block. We use a batch size of 128 and an initial learning rate of 0.1, decayed over 200 epochs using a cosine schedule. Following [41], we set $\lambda = 0.35$, $\tau_u = 0.07$, and $\tau_c = 1.0$ for representation learning, and use $\tau_s = 0.1$, with τ_t decayed from 0.07 to 0.04 over 30 epochs for classifier training. Pseudo-label thresholds δ and γ range from 0.6 to 0.9. Based on the convergence time of 'New' classes in SimGCD (Fig. 7a), the Inter-CU start epoch e is selected from $\{0, 20, 40, 60, 80\}$. All experiments are conducted using PyTorch on NVIDIA Tesla V100 GPUs.

5.2. Comparison With the State of the Arts

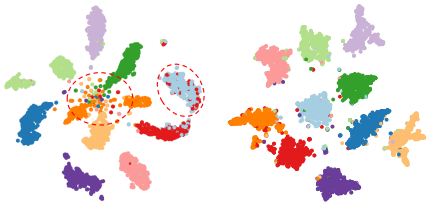
To demonstrate the effectiveness of AllGCD, we evaluate it under both known and unknown total class numbers K: (1) In Table 1(a), we assume K is known, consistent with prior works [3, 9, 13, 44, 45]; (2) In Table 1(b), we report results where K is unknown, estimating it using clustering techniques as in [5, 41, 51].

Evaluation with ground-truth class number K. Table 1(a) compares state-of-the-art methods assuming a known class number K. Our method significantly enhances 'New' category accuracy compared to the baseline SimGCD, with notable improvements of **7.4**%, **7.5**%, and **4.4**% on CUB, Stanford Cars, and ImageNet-100, respectively. Furthermore, AllGCD remains competitive with SPTNet, surpassing it by **2.6**% and **1.5**% on the 'All' classes in CUB and Stanford Cars, respectively. Importantly, our method attains comparable performance to SPTNet using only 200 training epochs, compared to SPTNet's 1000 epochs, showing superior computational efficiency. Though AllGCD trails CMS [5] by 2.7% in 'Old' accuracy on CIFAR100, it outperforms CMS by **6.7**% in 'New' accuracy, highlighting its strength in novel category discovery.

Evaluation with estimated number of clusters. Table 1 (b) presents a comparison conducted without access to the ground-truth class number K, instead using an estimated K derived from an off-the-shelf method [41], as shown in Table 2. Table 2 shows our approach provides a more accu-

Method	CIFAR100		ImageNet-100		CUB		Stanford Cars		FGVC-Aircraft		Herbarium 19	
	K	Err(%) ↓	K	Err(%) ↓	K	Err(%) ↓	K	Err(%) ↓	K	Err(%) ↓	K	Err(%) ↓
GT	100	-	100	-	200	-	196	-	100	-	683	-
GCD [41]	100	0	109	9	231	15.5	230	17.3	-	-	520	23.8
DCCL [28]	146	46	129	29	172	9	192	0.02	-	-	-	-
SimGCD [45]	100	0	109	9	231	15.5	230	17.3	-	-	520	23.8
GPC [51]	100	0	103	3	212	6	201	0.03	-	-	-	-
CMS [5]	95	5	116	16	170	15	156	20.4	90	10	622	8.9
AllGCD (ours)	95	5	108	8	213	6.5	210	7.1	109	9	636	6.9

Table 2. Estimated class number and error rate for K using off-the-self method [41].

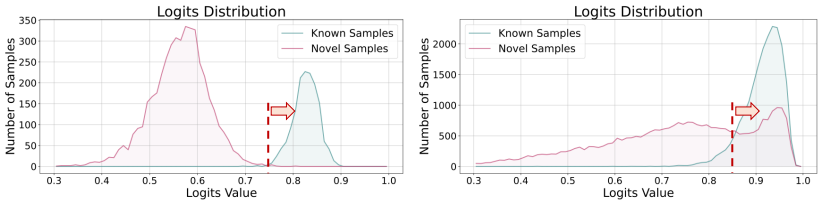


(c) SimGCD [45] (d) AllGCD

Figure 5. T-SNE visualization of 10 random novel classes from CIFAR100.

δ	CUB			CIFAR100		
	All	Old	New	All	Old	New
0.70	68.2	74.6	64.9	80.8	81.3	80.5
0.75	**68.4**	75.1	**65.1**	81.4	81.8	81.3
0.80	68.2	**75.7**	64.5	82.2	**83.1**	81.8
0.85	67.2	74.7	63.4	**82.3**	82.3	**82.2**
0.90	66.2	73.3	62.5	81.6	81.9	81.5

Table 3. Effect of confidence δ in Intra-CL.



(a) CUB with $\delta = 0.75$ (b) CIFAR100 with $\delta = 0.85$

Figure 6. Logit distributions of unlabeled data.

rate estimation of K, with fewer errors in determining cluster centers than SimGCD and CMS, except for CIFAR100 in SimGCD. Under this improved estimation, our approach consistently outperforms state-of-the-art methods across all datasets. Notably, performance gains are particularly significant on the Stanford Cars, ImageNet-100, and Herbarium-19 benchmarks, where AllGCD achieves **7.2**%, **6.0**%, and **9.1**% higher accuracy in the 'New' category compared to the second-best results. These findings highlight the effectiveness of our method in utilizing all available unlabeled data to learn more discriminative representations.

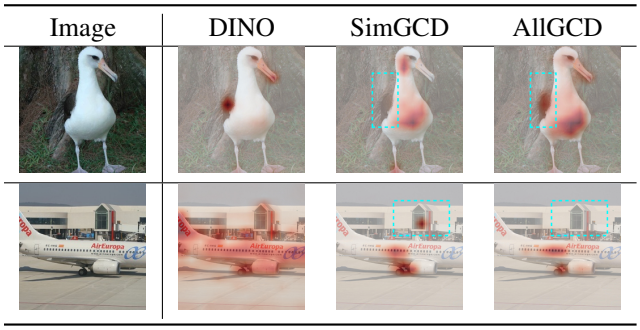
Feature Visualizations and Attention Distributions. To demonstrate the effectiveness of our approach in representation learning, we visualize 10 randomly selected novel classes from CIFAR100. As shown in Fig. 5, AllGCD produces more compact and well-separated clusters, whereas SimGCD exhibits considerable cluster overlap. Additionally, we visualize attention on both known and novel samples from unlabeled data in Table 4. Ours focuses on class-specific regions (*e.g.*, the bird's wing) and suppressing background noise (*e.g.*, the window in the second row), highlighting that leveraging unlabeled data enhances attention to relevant regions and improves class discriminability.

5.3. Ablation Study

In this section, we present ablation studies to assess the effectiveness of Intra-CL and Inter-CU in our AllGCD model. The evaluation includes fine-grained datasets such as CUB and Stanford Cars, alongside generic image recognition datasets CIFAR100 and ImageNet-100.

Effect of different confidence δ. Table 3 shows the effect of varying the confidence threshold δ in Intra-CL on CUB and CIFAR100. Overall, the accuracy for both 'Old' and 'New' classes initially improves with increasing δ but slightly drops beyond the optimal point. The fine-grained CUB dataset performs best at $\delta = 0.75$, while the generic dataset CIFAR100 peaks at $\delta = 0.85$. This difference arises from the varying proportions of unlabeled samples preserved in fine-grained and generic datasets. Higher thresholds, such as 0.85, help retain sufficient known samples in CIFAR100 (see Fig. 5b), but may exclude too many in CUB (see Fig. 5a), resulting in suboptimal performance for the fine-grained dataset. Therefore, we set $\delta = 0.75$ for fine-grained datasets and $\delta = 0.85$ for generic datasets.

Table 4. **Average attention across 12 heads in the final layer of the image encoder on CUB and FGVC-Aircraft.** The known sample (top row) and novel sample (bottom row) in AllGCD show improved focus on object regions with reduced background noise.



Effect of different confidence γ. Similar to δ in Intra-CL for selecting high-confidence known samples, γ filters high-quality distributions for inter-class contrastive learning in Inter-CU. Table 5 shows its effect on Stanford Cars and

Table 5. Ablation study on different thresholds γ in Inter-CU.

γ	Stanford Cars			ImageNet-100		
	All	Old	New	All	Old	New
0.60	60.0	76.8	52.2	85.6	93.8	81.5
0.65	**60.5**	**77.0**	**52.5**	86.0	94.1	81.7
0.70	59.8	76.7	52.1	**86.5**	94.7	**82.3**
0.75	58.9	75.1	51.2	86.3	**94.9**	82.1
0.80	57.6	73.7	50.1	86.1	94.4	81.9

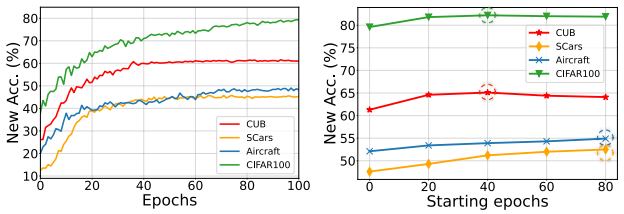
Table 6. Effectiveness of each component of our method. Intra-CL and Inter-CU are described in Sec. 4.1 and 4.2, respectively. Row (1) shows the baseline SimGCD results.

Index	SimGCD	Intra-CL	Inter-CU	CUB			CIFAR100		
				All	Old	New	All	Old	New
(1)	✓			60.3	65.6	57.7	80.1	81.2	77.8
(2)	✓	✓		65.2	73.1	61.3	81.2	81.8	79.6
(3)	✓		✓	67.1	72.8	63.8	81.6	81.5	81.7
(4)	✓	✓	✓	**68.4**	**75.1**	**65.1**	**82.3**	**82.3**	**82.2**

ImageNet-100. Specifically, we fix $\delta = 0.75$ for Stanford Cars and $\delta = 0.85$ for ImageNet-100 based on the δ ablation. Intuitively, Stanford Cars performs best with $\gamma = 0.65$, while ImageNet-100 performs best with $\gamma = 0.7$. Similar to Intra-CL, this difference is mainly due to the varying number of selected samples at each threshold: 6.1k *vs.* 95.3k (see Data in appendix). ImageNet-100 retains more high-confidence samples at a stricter threshold (*e.g.*, 0.75), while fine-grained datasets yield fewer, affecting Inter-CU's effectiveness. Therefore, we set $\gamma = 0.65$ or 0.6 for fine-grained datasets and $\gamma = 0.70$ or 0.75 for generic ones.

Effect of each proposed component. Table 6 presents the ablation study on our proposed Intra-CL and Inter-CU. Compared to SimGCD (Row (1)), Intra-CL (Row (2)) improves 'Old' class accuracy by 7.5% on CUB and 0.6% on CIFAR100, demonstrating its effectiveness in refining known sample selection for supervised contrastive learning. This also leads to a 3.6% and 1.8% improvement in 'New' accuracy on CUB and CIFAR100, respectively. Inter-CU (Row (3)) further enhances 'New' accuracy by 6.1% on CUB and 3.9% on CIFAR100 by fully utilizing unlabeled data for novel class discovery. Combining both (Row (4)) achieves the best overall performance, confirming their complementary benefits in improving recognition of both known and novel classes.

Effect of starting epoch e in Inter-CU. Fig. 7b reports 'New' accuracy for different Inter-CU start epochs e. Fig. 7a shows that SimGCD struggles to separate known and novel classes early (*e.g.*, during the first 20 epochs). This difficulty leads to an unreliable class distribution, which negatively impacts Inter-CU since it depends on well-separated inter-class relationships. To address this, we evaluate $e \in \{0, 20, 40, 60, 80\}$, where $e = 0$ denotes All-



(a) 'New' ACC. in SimGCD [45] (b) 'New' ACC. @ Epochs

Figure 7. **'New' class accuracy on unlabeled data.** (a) In SimGCD, 'New' ACC. remains low in the early training stages (*e.g.* below 40% for FGVC-Aircraft and Stanford Cars at epoch$\leq$20.) (b) Inter-CU improves accuracy more effectively in the middle and late stages (*e.g.* $e \geq 40$).

GCD without Inter-CU (only Intra-CL). Results show that Stanford Cars and FGVC-Aircraft peak at later epochs (*e.g.*, $e = 80$), while CUB and CIFAR100 peak at $e = 40$. This aligns with CUB's higher early-stage known class recognition compared to Stanford Cars and FGVC-Aircraft (60% *vs.* $\approx$43%), enabling earlier effective separation of known and novel samples. CIFAR100 is less sensitive to e, reaching approximately 70% accuracy by epoch 40, yielding stable class distributions for Inter-CU.

6. Conclusion

This paper identifies that previous supervised contrastive learning (S-CL) in parametric GCD algorithms limits novel category discovery due to its restricted use of labeled data. To address this, we propose AllGCD, a straightforward approach that leverages all unlabeled data containing novel classes into S-CL. Specifically, we introduce two strategies named Intra-class Contrast in Labeled Data (Intra-CL) and Inter-class Contrast in Unlabeled Data (Inter-CU). Intra-CL first enhances the intra-class compactness within known classes by incorporating potential known samples from unlabeled data into the original S-CL. Building on this, Inter-CU further separates known and novel classes by applying S-CL to a global class distribution. By combining these strategies, AllGCD maximizes the potential of S-CL, achieving substantial improvements in the discovery of novel categories. Extensive experiments show that AllGCD significantly outperforms baselines and achieves competitive results compared to other GCD methods.

Acknowledgments

This work was supported by the Guangdong Science and Technology Program (Grant No. 2024B0101040005), the National Natural Science Foundation of China (Grant No. 62332002 and 61825101), the Shenzhen High-Level Talent Team Program, and the Shenzhen Science and Technology Innovation Commission (Grant No. KQTD20240729102051063).

References

[1] Wenbin An, Feng Tian, Qinghua Zheng, Wei Ding, QianYing Wang, and Ping Chen. Generalized category discovery with decoupled prototypical network. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 12527–12535, 2023. 1

[2] Mahmoud Assran, Mathilde Caron, Ishan Misra, Piotr Bojanowski, Florian Bordes, Pascal Vincent, Armand Joulin, Mike Rabbat, and Nicolas Ballas. Masked siamese networks for label-efficient learning. In *Proceedings of the European conference on Computer Vision (ECCV)*, pages 456–473, 2022. 3

[3] Xinzi Cao, Yutong Lu, and Yonghong Tian. Solving the catastrophic forgetting problem in generalized category discovery. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16880–16889, 2024. 1, 2, 3, 4, 6

[4] 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 IEEE/CVF International Conference on Computer Vision (CVPR)*, pages 9630–9640, 2021. 1, 3, 6

[5] Sua Choi, Dahyun Kang, and Minsu Cho. Contrastive mean-shift learning for generalized category discovery. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23094–23104, 2024. 2, 6, 7, 1

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009. 2, 6

[7] 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 *Proceedings of International Conference on Learning Representations (ICLR)*, 2021. 1, 3, 6

[8] Yu Du, Fangyun Wei, Zihe Zhang, Miaojing Shi, Yue Gao, and Guoqi Li. Learning to prompt for open-vocabulary object detection with vision-language model. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14064–14073, 2022. 1

[9] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9264–9272, 2021. 3, 6

[10] Peiyan Gu, Chuyu Zhang, Ruijie Xu, and Xuming He. Class-relation knowledge distillation for novel class discovery. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16428–16437, 2023. 2

[11] Xiuye Gu, Tsung-Yi Lin, Weicheng Kuo, and Yin Cui. Open-vocabulary object detection via vision and language knowledge distillation. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2022. 1

[12] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8400–8408, 2019. 2

[13] Kai Han, Sylvestre-Alvise Rebuffi, Sébastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Autonovel: Automatically discovering and learning novel visual categories. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(10):6767–6781, 2022. 2, 3, 6

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 1

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2980–2988, 2017. 1

[16] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 1

[17] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Proceedings of Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 1, 2

[18] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of IEEE/CVF International Conference on Computer Vision Workshops (ICCV)*, pages 554–561, 2013. 2, 6, 1

[19] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 2, 6, 1

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of Conference on Advances in Neural Information Processing Systems (NeurIPS)*, pages 1106–1114, 2012. 1

[21] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955. 5, 6

[22] Jiaming Liu, Yangqiming Wang, Tongze Zhang, Yulu Fan, Qinli Yang, and Junming Shao. Open-world semi-supervised novel class discovery. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI*, pages 4002–4010. ijcai.org, 2023. 6

[23] Yu Liu, Yaqi Cai, Qi Jia, Binglin Qiu, Weimin Wang, and Nan Pu. Novel class discovery for ultra-fine-grained visual categorization. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17679–17688, 2024. 2

[24] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of*

the fifth Berkeley symposium on mathematical statistics and probability, pages 281–297, 1967. 6

[25] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew B. Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013. 2, 6, 1

[26] Matthias Minderer, Alexey A. Gritsenko, Austin Stone, Maxim Neumann, Dirk Weissenborn, Alexey Dosovitskiy, Aravindh Mahendran, Anurag Arnab, Mostafa Dehghani, Zhuoran Shen, Xiao Wang, Xiaohua Zhai, Thomas Kipf, and Neil Houlsby. In *Proceedings of Conference on European Conference on Computer Vision (ECCV)*, pages 728–755, 2022. 1

[27] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision. *Trans. Mach. Learn. Res.*, 2024, 2024. 2

[28] Nan Pu, Zhun Zhong, and Nicu Sebe. Dynamic conceptional contrastive learning for generalized category discovery. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7579–7588, 2023. 1, 2, 6, 7

[29] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151, 1999. 5

[30] Sarah Rastegar, Hazel Doughty, and Cees Snoek. Learn to categorize or categorize to learn? self-coding for generalized category discovery. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 6

[31] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 91–99, 2015. 1

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015. 1

[33] Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, 2018. 1

[34] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[35] Yiyou Sun and Yixuan Li. Opencon: Open-world contrastive learning. *Trans. Mach. Learn. Res.*, 2023. 1

[36] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. 1

[37] Kiat Chuan Tan, Yulong Liu, Barbara Ambrose, Melissa Tulig, and Serge Belongie. The herbarium challenge 2019 dataset. *arXiv preprint arXiv:1906.05372*, 2019. 2, 6, 1

[38] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *Proceedings of the European conference on Computer Vision (ECCV)*, pages 776–794, 2020. 1

[39] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748, 2018. 1, 2

[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017. 1

[41] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7482–7491, 2022. 1, 2, 3, 6, 7

[42] Sagar Vaze, Andrea Vedaldi, and Andrew Zisserman. No representation rules them all in category discovery. In *Proceedings of Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 1, 2

[43] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 2, 6, 1

[44] Hongjun Wang, Sagar Vaze, and Kai Han. Sptnet: An efficient alternative framework for generalized category discovery with spatial prompt tuning. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2024. 1, 2, 3, 6

[45] 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. 1, 2, 3, 6, 7, 8

[46] Jianzong Wu, Xiangtai Li, Shilin Xu, Haobo Yuan, Henghui Ding, Yibo Yang, Xia Li, Jiangning Zhang, Yunhai Tong, Xudong Jiang, Bernard Ghanem, and Dacheng Tao. Towards open vocabulary learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 46(7):5092–5113, 2024. 1

[47] Zelin Zang, Lei Shang, Senqiao Yang, Fei Wang, Baigui Sun, Xuansong Xie, and Stan Z. Li. Boosting novel category discovery over domains with soft contrastive learning and all in one classifier. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11824–11833, 2023. 2

[48] Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2018. 2

[49] Jie Zhang, Xiaosong Ma, Song Guo, and Wenchao Xu. Towards unbiased training in federated open-world semi-supervised learning. In *Proceedings of International Conference on Machine Learning, (ICML)*, pages 41498–41509, 2023. 1

[50] Sheng Zhang, Salman H. 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 (CVPR)*, pages 3479–3488, 2023. 1, 6

[51] 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, 2023. 6, 7, 1

[52] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9462–9470, 2021. 2