

Prior-Constrained Association Learning for Fine-Grained Generalized Category Discovery

Menglin Wang¹, Zhun Zhong^{2*}, Xiaojin Gong^{3*}

¹School of Computer and Electronic Information, Nanjing Normal University, China

²School of Computer Science and Information Engineering, Hefei University of Technology, China

³College of Information Science and Electronic Engineering, Zhejiang University, China
lynnwang6875@gmail.com, zhunzhong007@gmail.com, gongxj@zju.edu.cn

Abstract

This paper addresses generalized category discovery (GCD), the task of clustering unlabeled data from potentially known or unknown categories with the help of labeled instances from each known category. Compared to traditional semi-supervised learning, GCD is more challenging because unlabeled data could be from novel categories not appearing in labeled data. Current state-of-the-art methods typically learn a parametric classifier assisted by self-distillation. While being effective, these methods do not make use of cross-instance similarity to discover class-specific semantics which are essential for representation learning and category discovery. In this paper, we revisit the association-based paradigm and propose a Prior-constrained Association Learning method to capture and learn the semantic relations within data. In particular, the labeled data from known categories provides a unique prior for the association of unlabeled data. Unlike previous methods that only adopts the prior as a pre or post-clustering refinement, we fully incorporate the prior into the association process, and let it constrain the association towards a reliable grouping outcome. The estimated semantic groups are utilized through non-parametric prototypical contrast to enhance the representation learning. A further combination of both parametric and non-parametric classification complements each other and leads to a model that outperforms existing methods by a significant margin. On multiple GCD benchmarks, we perform extensive experiments and validate the effectiveness of our proposed method.

Code — <https://github.com/Terminator8758/PAL-GCD>

1 Introduction

The success of deep learning models has mostly been driven by the availability of large-scale annotated datasets. However, it is costly and inefficient to annotate all the data, especially as datasets grow larger in an open world. Semi-supervised learning (Oliver et al. 2018) thus emerges to be a promising direction for learning with both labeled and unlabeled data. Typical semi-supervised learning assumes unlabeled data comes from known categories. Nevertheless, data from novel categories frequently appear in the real world, limiting the applicability of semi-supervised learning. As a

relaxation to this assumption, generalized category discovery (GCD) has been proposed (Vaze et al. 2022), allowing unlabeled data to belong to both known and unknown categories. The target of GCD is to recognize images from both old and new categories by learning a model that clusters unlabeled images into distinct semantic groups, making it more practical for discovering novel categories with the assistance of old category data.

Current methods explore the GCD task from two perspectives, representation learning and parametric classification. The initial GCD paper (Vaze et al. 2022) uses self-supervised contrastive learning to learn robust representation, removing the need for parametric classifier. The problem is that it overlooks the intrinsic semantic relations among samples, causing the learned representation to be less discriminative. Indeed, samples belonging to the same potential category call for attraction instead of general repulsion. Later methods (Pu, Zhong, and Sebe 2023; Zhao, Wen, and Han 2023; Zhang et al. 2023) exploit cross-instance similarity relations to discover semantic groups or k -NN positives, and such grouping result guides the contrastive learning towards finding a discriminative feature space. However, the quality of grouping is determined by the design of data association strategy, and severe noise can be incorporated if the association design is not reliable enough. As an alternative, SimGCD (Wen, Zhao, and Qi 2023) revives parametric classifier through self-distillation learning and entropy regularization, which has become a popular baseline. But as the parametric classifier is implicitly regularized, its weights may not capture class-specific semantics well. Also, self-distillation alone may not provide strong enough supervision for the distinction of different classes.

In this paper, we re-examine the previous association-based GCD methods, and identify their weakness in the association design that makes their performance inferior, especially on fine-grained datasets. The aim is that through better association design, more accurate instance groupings can be estimated to facilitate model representation learning. Specifically, Figure 1 provides an intuitive example to illustrate the limitations of previous association designs and our motivation. As shown in the figure, some methods utilize the labeled prior for pre-association refinement, i.e. masking out the distance of images from different known categories. However during the association, the masked image

*Corresponding authors.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

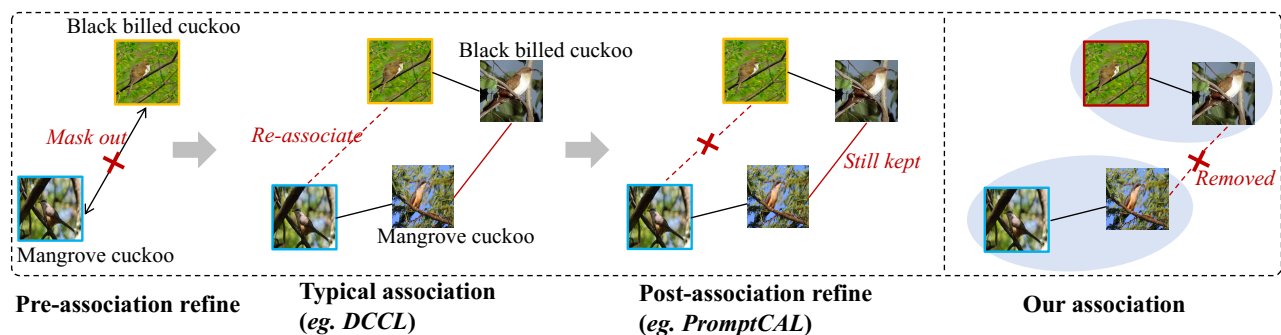


Figure 1: An illustration of implicit false association, and how our proposed association avoids it. Arrowed line represents masking out the distance, solid line indicates direct association of two instances, and dashed line denotes indirect association. Red line indicates false association. Images with colored edges indicate they are from labeled subset.

pair could be re-connected by indirect association of other images. Such association will generate inaccurate groupings of labeled and unlabeled instances. Even if a post-association refinement removes the connection of labeled images, the unlabeled images’ association is still kept, causing undesirable groupings to appear.

In light of this, we aim to circumvent such possibly emerging situations by fully incorporating the labeled prior into the association process. As Figure 1 shows, our association keeps track of the updated image groups. When a new image pair appears, we not only examine the image pair, but also check their corresponding groups to determine if the image pair should be associated. In the example, the two groups contain known yet label-conflicting instances, therefore the image pair will not be associated. In this way, both direct and indirect false association can be avoided. By repeating the instance-wise association steps, more faithful instance groupings can be obtained.

With the groupings generated by association, we adopt non-parametric prototypical contrastive learning to strengthen the representation. The association based non-parametric classification serves as a stand-alone and good-performing framework. Specially, the prior knowledge of ground truth class number is not required, which makes the framework flexible in practical applications. Additionally, we also propose to optionally perform association on a subset of all samples to improve the scaling of the association to more data scenarios. We explore the possibility of combining parametric classifier with the proposed association based parametric classifier, which is proven effective when the number of classes is known. Through a two-stage training pipeline, we exploit their complementarity and unify them in one framework to mutually boost each other. This unified model achieves strong performance on multiple datasets.

Our main contributions are summarized as follows:

- We propose a simple yet effective method for fine-grained GCD. By re-examining the role of association, a novel prior-constrained association algorithm tailored for GCD task is proposed.
- With the assistance of proposed association, we unify non-parametric and parametric classification under one single framework, where representation learning and

classifier learning can mutually boost each other.

- Extensive experiments on both fine-grained and generic datasets demonstrate the effectiveness of the proposed method. Compared to previous best method, our method improve the accuracy by 4.4% and 15.3% on CUB and Stanford Cars respectively.

2 Related Work

Generalized Category Discovery (GCD) draws similarity with novel category discovery (NCD) in both containing labeled and unlabeled images and aiming to discover novel categories in the unlabeled set. Initial GCD method (Vaze et al. 2022) learns representations by self-supervised contrastive learning on all data, along with supervised contrastive learning on labeled subset. SimGCD (Wen, Zhao, and Qi 2023) constructs an effective baseline using parametric classifier. A later variant μ GCD (Vaze, Vedaldi, and Zisserman 2024) improves SimGCD by using a teacher network to provide supervision for self-augmented image pairs. More recently, SPTNet (Wang, Vaze, and Han 2024) learns spatial prompts as an alternative to adapt data for better alignment with the model. DCCL (Pu, Zhong, and Sebe 2023) proposes to mine sample relations by generating dynamic conceptions using improved Infomap clustering (Rosvall and Bergstrom 2008), followed by conception and instance-level contrastive learning. Similarly, GPC (Zhao, Wen, and Han 2023) also estimates prototypes by Gaussian mixture model and a split-and-merge to take labeled instances into account. PromptCAL (Zhang et al. 2023) improves the ViT backbone by learning auxiliary prompts, as well as affinity propagation on KNN graph to estimate instance relation. Although labeled data is exploited to assist clustering in these methods, it is often taken as a pre- or post-clustering refinement. As such, the potential benefit of labeled instances are not fully exploited. As a comparison, we fully incorporate the labeled data prior during every step of the association process, empowering reliable association of unlabeled data by taking advantage of the labeled instances as bridges.

Prototypical Contrastive Learning (PCL). In recent years, contrastive learning (Gutmann and Hyvärinen 2010) has proven as an effective technique for self-supervised learning (Wu et al. 2018; He et al. 2020; Chen et al. 2020; Li et al.

2021) and other settings (Khosla et al. 2020; Wang et al. 2021a; Zhao, Wen, and Han 2023). In particular, prototypical contrastive learning compares instances with a set of prototypes encoding class-specific semantic structure, leading to discriminative embedding space. As such, many vision tasks have exploited PCL for method design. ProtoNCE (Li et al. 2021) combines instance-wise contrastive learning and multi-grained PCL for transfer learning. (Ge et al. 2020; Wang et al. 2021b) adopt iterative clustering based PCL for object re-ID. A few methods (Pu, Zhong, and Sebe 2023; Zhao, Wen, and Han 2023) in GCD have also considered PCL to learn discriminative representation. The critical issue for prototypical contrast is how to obtain representative prototypes, which then comes down to designing effective association strategy. Our method also adopts PCL, however, our better utilization of prior and design of semi-supervised association lead to more reliable prototypes, which in turn facilitates learning better representation.

Data Clustering and Association. Clustering has long been used as a way to discover potential semantic groups within the data. Unsupervised clustering methods like K-Means (Hartigan and Wong 1979), DBSCAN (Ester et al. 1996) and hierarchical clustering (Johnson 1967; Murtagh and Contreras 2012) are widely used in many applications (Ge et al. 2020; Wang et al. 2022; Pu, Zhong, and Sebe 2023). Semi-supervised clustering is also studied in some works (Bair 2013; Bilenko, Basu, and Mooney 2004). Basu *et al.* (Basu 2002) propose constrained K-Means by enforcing that labeled instances are assigned to their own cluster during K-Means iteration. COP-Kmeans (Wagstaff et al. 2001) modifies K-Means to make sure no constraints are violated when assigning instances. Constrained DBSCAN (Ruiz, Spiliopoulou, and Menasalvas 2010) and hierarchical clustering (Davidson and Ravi 2005) are also considered. Metric-based methods (Yin et al. 2010; Klein, Kamvar, and Manning 2002; Xing et al. 2002; Lange et al. 2005; Pu, Zhong, and Sebe 2023; Zhang et al. 2023) modify the pairwise distance such that two instances with a "must-link" constraint have a lower distance, and those with a "cannot-link" constraint have a larger distance. Our proposed association is also constraint-based, but the constraints are enforced during a threshold-based group merging process, during which new categories are allowed to be discovered.

3 Methodology

3.1 Overview

Under Generalized Category Discovery setting, we consider the problem of clustering images in a dataset among which a subset has known class labels. Assume the dataset \mathcal{D} is comprised of two parts $\mathcal{D}_{\mathcal{L}} = \{(x_i, y_i)\}_{i=1}^N \in \mathcal{X} \times \mathcal{Y}_{\mathcal{L}}$ and $\mathcal{D}_{\mathcal{U}} = \{(x_i, y_i)\}_{i=1}^M \in \mathcal{X} \times \mathcal{Y}_{\mathcal{U}}$, where $\mathcal{D}_{\mathcal{L}}$ is the labeled subset of N images whose labels $\mathcal{Y}_{\mathcal{L}}$ are known, and $\mathcal{D}_{\mathcal{U}}$ is the unlabeled subset of M images whose labels $\mathcal{Y}_{\mathcal{U}}$ are not known. Image labels in $\mathcal{D}_{\mathcal{U}}$ is a superset of image labels in $\mathcal{D}_{\mathcal{L}}$, i.e. $\mathcal{Y}_{\mathcal{L}} \subset \mathcal{Y}_{\mathcal{U}}$. Given dataset \mathcal{D} , the aim is to correctly recognize and cluster the images in $\mathcal{D}_{\mathcal{U}}$ containing known and unknown categories. To address the GCD task, we seek to improve the representation learning by estimat-

Dataset	CUB	StanfordCars	Aircraft	Herbarium19
Estimated class number	76	74	27	228
G.T. class number	100	98	50	342

Table 1: Comparison of class number estimated by pre-clustering refinement + DBSCAN, v.s. ground truth class number in labeled subset of training images.

ing reliable semantic groups as the guidance. To this end, we incorporate the labeled data prior into the association process, and design a prior-constrained greedy association algorithm. Such association generates faithful instance groups as well as class-representative proxies to guide the model representation learning. Finally, to exploit the synergy of non-parametric (prototypical contrastive learning) and parametric classification, we unify them in one framework by joint two-stage optimization.

3.2 Limitation of Previous Methods

There have been some recent attempts (Pu, Zhong, and Sebe 2023; Zhang et al. 2023; Kim et al. 2023; Zhao, Wen, and Han 2023) at estimating the semantic structure by semi-supervised clustering or association, so as to provide stronger and explicit supervision to unlabeled data. Albeit with acceptable performance, we take a closer look at the current association designs in GCD and discover that there exists missing clues and the association can be further optimized with the given labeled data prior.

In GCD task, it is natural to utilize the labeled data from known categories to assist the association of unlabeled instances. Current association-based methods usually adopt the labeled data as a pre- or post-clustering refinement. For pre-clustering refinement, after computing the pairwise distance of instances, those between known yet different categories can be directly masked as disconnected. Then the refined distance matrix would be input to a standard clustering algorithm (Rosvall and Bergstrom 2008; Ester et al. 1996). For post-clustering refinement, after unsupervised clustering, the associations between known different categories are removed as a refinement.

However, both pre- and post-clustering refinement neglect the underlying association process. As Figure 1 shows, inter-class false association can still occur even after simple pre/post-refinement. Simply adopting the standard unsupervised clustering not only fails to address this, but also keeps the incorrect association of unlabeled instances untreated. As an example to verify the problem with pre-clustering refinement, we use the pre-refined inter-instance distance matrix as input to DBSCAN clustering (Ester et al. 1996), and compare the predicted pseudo class number of labeled subset with the ground truth. In Table 1, we observe that the clustering predicts much less class number than ground truth, indicating that instances from different labeled classes have been falsely merged. This demonstrates that simply masking the distance of labeled classes during pre-association is insufficient, as these masked instances can still be mis-connected during association.

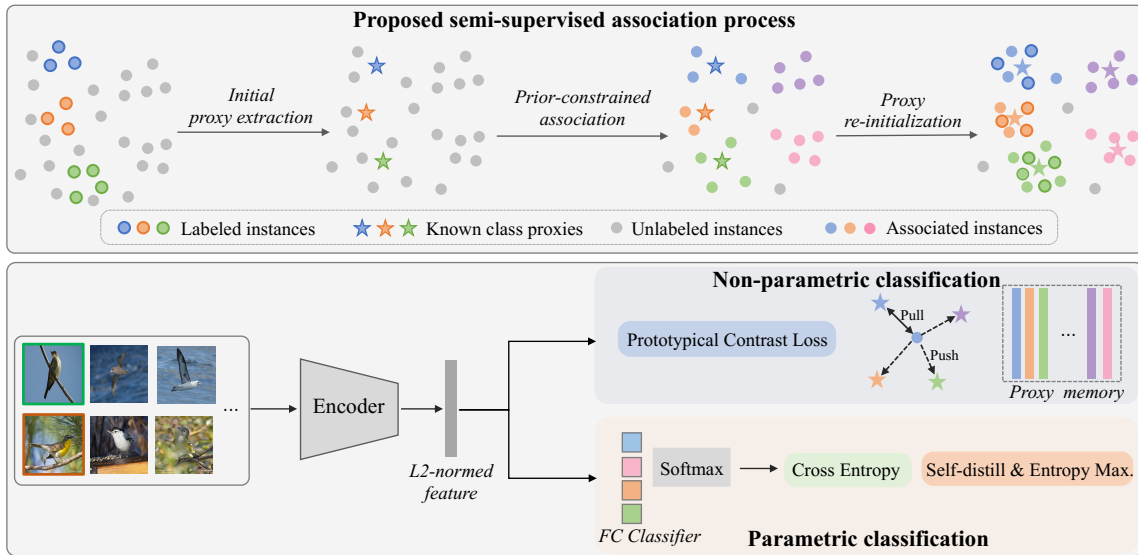


Figure 2: An overview of our method. Non-parametric classification is in the form of prototypical contrastive learning, with prototypes obtained by the proposed semi-supervised association. Parametric classification follows the implementation of SimGCD (Wen, Zhao, and Qi 2023).

3.3 Prior-constrained Greedy Association

In an effort to fix the limitations of the existing association/clustering in GCD, we propose our prior-constrained greedy association algorithm. The initial motivation is that each group should only contain at most one old class. To ensure the constraint is satisfied, we propose to attend to each step of the association. While associating instance pairs in a distance-ascending order, each step goes through a labeled prior based validity check to make sure one group contains no more than one old class. After such association, the generated instance groups are guaranteed to adhere to the ground truth of labeled subset. An overall pipeline of our association process is shown in the upper part of Figure 2.

- *Hybrid feature extraction.* First, we extract feature encodings of all training samples using the backbone network f . For the labeled instances, per-category mean feature is calculated as their class representative proxy. Then, our prior-constrained greedy association is performed among the initial known-class proxies and the rest unlabeled instances.
- *Pairwise distance computation.* With the hybrid features of initial proxies and unlabeled instances, we compute their pairwise Jaccard distance (Zhong, Zheng, and Li 2017), and sort the distances in ascending order. Only pairs whose distance fall below a given threshold ϵ are kept. Those kept pairs are regarded as the association candidates $P = \{(j_1, j_2), (j_3, j_4), \dots\}$.
- *Greedy association with constraint.* In order to incorporate the labeled data prior to constrain the association, we unfold the association in a pair-wise manner. The pseudo algorithm for the association is presented in Alg. 1 of Appendix. Starting from the most similar instance-to-instance (or instance-to-proxy) pair, we obtain the initial grouping $Grp = \{0 : (j_1, j_2)\}$. For the next candi-

date pair (j_3, j_4) , we check for conflicts if this candidate pair is to be associated; If the group of instance/proxy j_3 contains known category different from the group of instance/proxy j_4 , this candidate pair will not be associated. The conflict check is performed for every candidate pair in P . In this way, every step of association updates the grouping result while still fully respects the ground truth class relations of labeled data.

Scaling by subset association. The proposed greedy association works by gradually associating reliable instances into cluster groups. This makes it well-suited for fine-grained datasets like Semantic Shift Benchmark (Vaze et al. 2021), where each category has a moderate number of images. When scaling to large-scale datasets or scenarios with many images per category, we suggest to perform the association on a randomly sampled subset of unlabeled images. Specifically, at each time of association, a fixed ratio of unlabeled instances are randomly sampled from all unlabeled instances, and then associated along with labeled known proxies. Such subset association enjoys *two benefits*: First, it reduces the number of per-category images for more effective association. Second, for large-scale datasets, the computational cost of association (including pairwise distance computation and greedy association) is significantly reduced.

3.4 Non-parametric Classification

With the semantic groups predicted by the proposed association, we construct a proxy memory $\mathcal{K} \in R^{C \times d}$ representing the feature centroid of each semantic group. At the same time, instance-to-group relation is also estimated from the association. To learn the potential semantic structure, we adopt prototypical contrastive learning paradigm (Wu et al. 2018) taking the proxy memory as the prototypes. Given an image x_i , its d -dimensional feature $f(x_i)$ is extracted

through the backbone network f . After l_2 -normalization, the image feature is contrasted with the proxy feature memory, and the prototypical contrastive loss is computed as:

$$\mathcal{L}_{npa} = -\frac{1}{B} \sum_{i=1}^B \log \frac{\exp(\mathcal{K}[y_i]^T f(x_i)/\tau)}{\sum_{j=0}^{C-1} \exp(\mathcal{K}[j]^T f(x_i)/\tau)}, \quad (1)$$

where $\mathcal{K}[j]$ is the j -th entry of the memory, τ is a temperature factor, B is batch size, and C is the number of proxies in \mathcal{K} . The contrastive loss defined in Eq. (1) pulls an instance close to the centroid of its group while pushes it away from the centroids of all other groups. This can be seen as non-parametric classification with the external proxy memory serving as the classifier. Feature representation gets enhanced through optimizing the similarity relation between images and the representative proxies.

After each batch forward, the proxy memory features are updated in a moving-average manner (Xiao et al. 2017) using the online batch image features:

$$\mathcal{K}[\tilde{y}_i] \leftarrow \mu \mathcal{K}[\tilde{y}_i] + (1 - \mu) f_\theta(x_i), \quad (2)$$

where $\mu \in [0, 1]$ is an updating rate. To promote the mutual-beneficial effect of representation learning and data association, the two processes are iteratively performed during training. Better representation improves the quality of data association, and the improved association in turn facilitates stronger representation learning.

As shown in our experiments, prototypical contrastive learning driven by the proposed data association can serve as an effective *stand-alone* GCD framework. One advantage of it is that it does not require the prior knowledge of the ground truth class number, offering more flexibility to application. Nevertheless, when the ground truth class number is known, parametric classification (Wen, Zhao, and Qi 2023) can be integrated as a useful complement to the non-parametric contrastive framework. Next, we briefly introduce the parametric classification, and how the two learning mechanisms can be integrated into one unified framework that produces more powerful model.

3.5 Joint Non-param. and Param. Classification

Parametric Classification. Following the good practice of semi-supervised learning, a representation parametric classification method SimGCD (Wen, Zhao, and Qi 2023) performs self-distillation by mining the prediction consistency between augmented views, using the sharpened prediction of one augmented view as the supervision for another view. The total loss \mathcal{L}_{simgcd} includes the cross entropy loss on the labeled data, self-distillation and entropy maximization on both labeled and unlabeled data, as well as batch supervised/self-supervised contrast (Vaze et al. 2022).

Our association based non-parametric classifier directly learns the globally estimated semantic groups, and presents itself as a complement to parametric classifier. However, combining them in one framework is not so straightforward, as the two types of classifiers are learned with different sampling strategy and at different learning paces. Parametric classifier typically learns at a slower pace, while non-parametric classifier learns faster due to the characteristic of non-parametric classification (Xiao et al. 2017).

To improve the effectiveness of joint learning the two classifiers, we propose a two-stage training strategy. In the first “warm-up” stage, we only train the backbone network with non-parametric classification, to better prepare the network for joint training. In the second stage, the parametric classifier is added, and a weight $\beta = 0.1$ is assigned to the non-parametric classifier loss to balance the learning. The training objective of the second stage is the weighted sum of both classifier loss, *i.e.* $\mathcal{L} = \mathcal{L}_{simgcd} + \beta \mathcal{L}_{npa}$.

Discussion. Association based learning has been previously explored for GCD. For example, DCCL (Pu, Zhong, and Sebe 2023), GPC (Zhao, Wen, and Han 2023) and Prompt-CAL (Zhang et al. 2023) all use inter-instance similarity for grouping or pairwise labeling. And prototypical contrastive learning is also adopted by DCCL, GPC and OpenCon (Sun and Li 2022) to enhance representation learning. Our method differs from them in two ways: First, we leverage the labeled data prior in GCD task, and let the prior constrain the data association process, ensuring that labeled instances are grouped respecting their label prior. This enables our association to generate much more reliable instance grouping result. Second, we show that non-parametric classification can be effectively combined with parametric classifier to further advance the performance in discovering novel categories.

4 Experiments

4.1 Experimental Setup

Datasets. We perform experiments on four fine-grained datasets and two generic datasets. Fine-grained datasets include the Semantic Shift Benchmark (Vaze et al. 2021) (CUB-200 (Wah et al. 2011), StanfordCars (Krause et al. 2013), Aircraft (Maji et al. 2013)), and one long-tailed dataset Herbarium19 (Tan et al. 2019). General datasets include Cifar100 (Krizhevsky, Hinton et al. 2009) and ImageNet-100 (Deng et al. 2009). Compared to generic recognition, fine-grained datasets are more challenging due to small inter-class variation, and reflect many real-world cases in visual recognition system. Following common settings (Vaze et al. 2022), a subset of all train classes is sampled as the old classes, the rest are new classes. 50% of the images from known classes are used to construct the labeled subset \mathcal{D}_L , and the rest images constitute \mathcal{D}_U .

Evaluation metric. In accordance with standard practice (Vaze et al. 2022), clustering accuracy (ACC) is utilized to evaluate the model performance. During evaluation, the predicted label \hat{y} is compared with the ground truth label y^* , and ACC is calculated as $ACC = \frac{1}{M} \sum_{i=1}^M \mathbb{1}(y_i^* = p(\hat{y}_i))$, where $M = |\mathcal{D}_U|$, and p is the best permutation of \hat{y} to match the ground truth y^* .

Implementation Details. We adopt ViT-B/16 (Dosovitskiy et al. 2021) pre-trained on DINO (Caron et al. 2021) as the backbone network. Following GCD (Vaze et al. 2022), only the last block of the backbone is fine-tuned. Batch size is 128. Learning rate is 0.01 for the first training stage decayed with a cosine annealed schedule, and 0.1 for the second stage of joint training. More implementation details and dataset statistics can be found in Appendix.

Methods	CUB			Stanford Cars			Aircraft			Herbarium19			Cifar100			ImageNet-100		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
Ground truth number of classes known																		
<i>k</i> -means (1967)	34.3	38.9	32.1	12.8	10.6	13.8	16.0	14.4	16.8	13.0	12.2	13.4	52.0	52.2	50.8	72.7	75.5	71.3
RankStats+ (2021)	33.3	51.6	24.2	28.3	61.8	12.1	26.9	36.4	22.2	27.9	55.8	12.8	58.2	77.6	19.3	37.1	61.6	24.8
UNO+ (Fini et al. 2021)	35.1	49.0	28.1	35.5	70.5	18.6	40.3	56.4	32.2	28.3	53.7	14.7	69.5	80.6	47.2	70.3	95.0	57.9
GPC (2023)	52.0	55.5	47.5	38.2	58.9	27.4	43.3	40.7	44.8	-	-	-	77.9	85.0	63.0	76.9	94.3	71.0
GCD (Vaze et al. 2022)	51.3	56.6	48.7	39.0	57.6	29.9	45.0	41.1	46.9	35.4	51.0	27.0	73.0	76.2	66.5	74.1	89.8	66.3
XCon (Fei et al. 2022)	52.1	54.3	51.0	40.5	58.8	31.7	47.7	44.4	49.4	-	-	-	74.2	81.2	60.3	77.6	93.5	69.7
DCCL (2023)	63.5	60.8	<u>64.9</u>	43.1	55.7	36.2	-	-	-	-	-	-	75.3	76.8	70.2	80.5	90.5	76.2
PromptCAL (2023)	62.9	64.4	62.1	50.2	70.1	40.6	52.2	52.2	52.3	37.0	52.0	28.9	81.2	84.2	75.3	83.1	92.7	78.3
PIM (2023)	62.7	75.7	56.2	43.1	66.9	31.6	-	-	-	42.3	56.1	34.8	78.3	84.2	66.5	83.1	<u>95.3</u>	77.0
μ GCD (2024)	65.7	68.0	64.6	56.5	68.1	50.9	53.8	55.4	53.0	45.8	61.9	37.2	-	-	-	-	-	-
CMS (2024)	68.2	76.5	64.0	56.9	76.1	47.6	56.0	63.4	52.3	36.4	54.9	26.4	82.3	85.7	75.5	84.7	95.6	79.2
SimGCD (2023)	60.3	65.6	57.7	53.8	71.9	45.0	54.2	59.1	51.8	44.0	<u>58.0</u>	36.4	80.1	81.2	77.8	83.0	93.1	77.9
SimGCD [†] (2023)	60.8	65.2	58.5	53.8	70.8	45.6	52.3	59.8	48.6	44.5	57.9	37.3	79.4	82.2	73.9	<u>85.0</u>	94.2	<u>80.3</u>
<i>Ours</i> (param. eval)	<u>67.6</u>	<u>75.5</u>	<u>63.7</u>	<u>66.7</u>	<u>79.1</u>	<u>60.7</u>	59.5	67.2	<u>55.6</u>	47.6	58.6	41.7	<u>82.0</u>	82.4	81.2	86.3	93.0	83.0
<i>Ours</i> (nonparam. eval)	72.6	75.2	71.2	72.2	83.4	66.8	<u>58.8</u>	<u>64.5</u>	56.0	46.8	57.0	41.4	78.5	78.9	<u>77.8</u>	83.0	93.1	78.0
Ground truth number of classes unknown																		
GCD (Vaze et al. 2022)	51.1	56.4	48.4	39.1	58.6	29.7	-	-	-	37.2	51.7	29.4	70.8	77.6	57.0	77.9	91.1	71.3
GPC (2023)	52.0	55.5	47.5	38.2	58.9	27.4	43.3	40.7	44.8	36.5	51.7	27.9	75.4	84.6	60.1	75.3	93.4	66.7
PIM (2023)	62.0	75.7	55.1	42.4	65.3	31.3	-	-	-	<u>42.0</u>	<u>55.5</u>	<u>34.7</u>	75.6	81.6	63.6	83.0	<u>95.3</u>	76.9
CMS (2024)	64.4	68.2	<u>62.4</u>	<u>51.7</u>	<u>68.9</u>	<u>43.4</u>	<u>55.2</u>	<u>60.6</u>	<u>52.4</u>	37.4	56.5	27.1	79.6	<u>83.2</u>	<u>72.3</u>	81.3	95.6	74.2
<i>Ours</i> *	69.9	<u>68.0</u>	70.9	70.5	77.7	67.0	56.2	55.3	56.7	44.8	48.5	42.8	<u>77.3</u>	78.9	74.1	<u>81.7</u>	94.6	<u>75.2</u>

Table 2: Performance comparison with SoTA methods. *Ours* (param. eval) and *Ours* (nonparam. eval): Our full model evaluated with pseudo label predicted by parametric classification logits, or by the proposed association. *Ours**: Our model trained with only non-parametric loss. [†] denotes reproduced results. Best and second best results are marked by **Bold** and underline.

	Training			Evaluation		CUB			Stanford Cars			Aircraft			Herbarium19		
	PcA	Param. Cls	2-stage	A&A	Param. Cls	All	Old	New	All	Old	New	All	Old	New	All	Old	New
(1)		✓			✓	60.8	65.2	58.5	53.8	70.8	45.6	52.3	59.8	48.6	44.5	57.9	37.3
(2)	✓			✓		<u>69.9</u>	68.0	<u>70.9</u>	<u>70.5</u>	77.7	67.0	56.2	55.3	56.7	44.8	48.5	42.8
(3)	✓	✓	✓		✓	67.6	75.5	63.7	66.7	<u>79.1</u>	60.7	59.5	67.2	55.6	47.6	58.6	41.7
(4)	✓	✓	✓	✓		72.6	<u>75.2</u>	71.2	72.2	83.4	<u>66.8</u>	<u>58.8</u>	64.5	56.0	46.8	57.0	41.4
(5)	✓	✓			✓	65.1	70.7	62.3	56.8	74.3	48.4	57.1	61.0	55.2	45.1	<u>57.4</u>	38.5
(6)	✓	✓		✓		66.9	72.0	64.4	60.0	76.6	51.9	54.8	62.8	50.9	44.1	54.3	38.6

Table 3: Ablation study on the main components of our method. ‘PcA’ denotes the proposed Prior-constrained Association. ‘Param. Cls’ denotes the parametric classifier. ‘A&A’ denotes evaluating the model by our Association and Assign.

4.2 Comparison with State of The Arts

In Table 2, we compare with the state-of-the-art GCD methods under two settings: ground truth number of classes known or unknown.

Ground truth number of classes known. Under this setting, we combine the association based non-parametric classification with the parametric classification, where the ground truth class number is utilized as a prior in the latter. In Table 2, our proposed method achieves state-of-the-art performance on fine-grained datasets, whether using parametric or non-parametric classifier for evaluation. On CUB and Stanford Cars, our method surpasses the previous best method CMS by 4.4% and 15.3% on ‘All’ accuracy. On generic datasets, our method performs on par with SoTA methods. Additionally, we notice a consistent improvement on ‘New’ classes, proving our method is good at discovering and clustering new categories.

Ground truth number of classes unknown. Not knowing the ground truth class number is a more practical setting but causes the model learning to be more challenging. In Ta-

ble 2, we compare our method with others that does not require the ground truth class number. The results show that with solely non-parametric loss, our method achieves much higher accuracy on all fine-grained datasets, and also delivers consistent performance on generic datasets. The comparisons prove the effectiveness of our method and its flexibility to work under class-unknown setting.

4.3 Ablation Study

To investigate how each component affects the model performance, we perform ablation experiments and present the results in Table 3.

Effectiveness of the prior-constrained association. Table 3 (1) lists the accuracy of training and evaluation with parametric classifier (Wen, Zhao, and Qi 2023). Compared with (1), our association based non-parametric classification as denoted by (2) achieves better performance on each dataset. Noticeably on CUB and Stanford Cars, (2) improves the All Acc by 9.1% and 16.7% respectively.

Effectiveness of joint training. The full model indicated

Association	CUB			Stanford Cars			Aircraft		
	All	Old	New	All	Old	New	All	Old	New
Semi-Kmeans	61.0	50.6	66.2	49.7	58.9	45.2	38.2	36.2	39.2
Semi-DBSCAN	68.3	64.8	70.1	65.2	74.7	60.6	47.4	47.8	47.2
Semi-DBSCAN w/ constraint	71.1	72.8	70.3	<u>68.6</u>	<u>77.1</u>	<u>64.4</u>	<u>53.7</u>	<u>53.6</u>	<u>53.8</u>
Ours w/o constraint	67.9	61.5	71.1	65.9	72.3	62.9	47.5	52.9	44.7
Our association	<u>69.9</u>	<u>68.0</u>	<u>70.9</u>	70.5	77.7	67.0	56.2	55.3	56.7

Table 4: Comparison of models trained with different association algorithms. Semi-Kmeans is proposed in (Vaze et al. 2022). Semi-DBSCAN is based on the clustering algorithm DBSCAN (Ester et al. 1996) and inter-class distances among known instances are masked before clustering. Semi-DBSCAN w/ constraint: adds our proposed prior constraint into the semi-DBSCAN clustering. Ours w/o constraint: our association but with the prior constraint removed.

by (3) and (4) jointly trains with association-based non-parametric classifier and parametric classifier. Compared to (1) and (2), the result in (3) improves the parametric classifier to a large extent, and the accuracy in (4) also consistently boosts over the non-parametric classifier alone. This shows that the joint training is indeed able to benefit both classifiers by mining their complementarity.

Effectiveness of two-stage training. To validate the necessity of two-stage training, we also provide the results of jointly training non-parametric and parametric classifier in one single stage, as indicated by (5) and (6). Compared with (1) and (2), the single-stage training promotes the accuracy of parametric classifier, but drops the performance of non-parametric classifier, indicating that a warming-up stage is necessary to better prepare the model for joint training.

4.4 Analysis on The Proposed Association

In this subsection, we conduct analysis on the proposed association from different aspects. To focus on the association part, we only adopt the association-based non-parametric classification loss when reporting the performances.

How does the model perform with other association algorithms? In Table 4, we explore the option of adopting other common clustering algorithms including Semi-Kmeans (Vaze et al. 2022) and Semi-DBSCAN (Ester et al. 1996), both during training and evaluation. From the table, we observe that Semi-DBSCAN shows competitive performance compared to Semi-Kmeans, but stills underperforms the proposed association on all three datasets.

How much contribution does the prior constraint make in association? The prior constraint serves as the key element in our proposed association. We validate its effectiveness in two ways: First, we demonstrate its integration into the classic DBSCAN algorithm which merges the instances greedily. As shown in Table 4, adding the prior constraint to Semi-DBSCAN leads to steady improvement on all three datasets, even surpassing our association method on CUB. This highlights the generalizability of the proposed prior constraint. With the constraint incorporated, our association achieves better accuracy than other algorithms on Stanford Cars and Aircraft.

Subset size	CUB			Aircraft			Cifar100		
	All	Old	New	All	Old	New	All	Old	New
100%	69.9	68.0	70.9	50.8	55.8	48.4	74.4	81.8	59.7
50%	60.8	52.5	64.9	56.2	55.3	56.7	77.0	79.6	71.8
30%	45.2	41.8	46.8	48.7	38.7	53.7	77.3	78.9	74.1

Table 5: Comparison of model performance with different subset size for association.

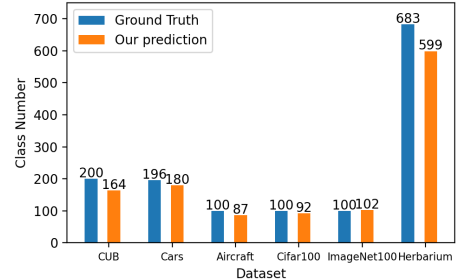


Figure 3: Estimated class number on each dataset.

When is subset association necessary? For large-scale datasets or when number of images per category is high, we propose to perform the subset association. To verify the effect of subset association, we provide in Table 5 the performance on representative datasets with subset and full-set association. CUB, with an average of 30 images per-category, shows better accuracy with full-set association, which is reasonable as reducing subset size further leads to insufficient data for association. In contrast, Aircraft and Cifar100, with an average of 67 and 500 images per category, benefit from subset association, likely because our association works best with a small amount of representative samples, and involving too many samples per-category brings more noise to association, thus harming the performance.

How well can the proposed method predict the class number? When the non-parametric classification loss is utilized alone, our method does not require the prior knowledge of ground truth class number. In Figure 3, we compare the estimated *v.s.* ground truth class numbers. In most cases, the association generates fewer pseudo classes than ground truth. Overall, the estimated class number is close to ground truth, with a maximum error rate of 18%.

5 Conclusion

In this paper, we have proposed a simple yet effective method for generalized category discovery. By mining the labeled data prior under GCD setting, we propose a prior-constrained greedy association algorithm to estimate reliable semantic groups for representation learning. Assisted by the association, the non-parametric prototypical contrastive learning can not only work alone to achieve good performance, but also be effectively integrated with the parametric classifier to mutually benefit each other, leading to further enhanced accuracy. Extensive experiments on multiple benchmarks demonstrate the effectiveness and superiority of the proposed method.

References

- Bair, E. 2013. Semi-supervised clustering methods. *Wiley Interdisciplinary Reviews: Computational Statistics*.
- Basu, S. 2002. Semi-supervised clustering by seeding. In *International Conference on Machine Learning*.
- Bilenko, M.; Basu, S.; and Mooney, R. J. 2004. Integrating constraints and metric learning in semi-supervised clustering. In *Proceedings of the twenty-first international conference on Machine learning*.
- Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; and Joulin, A. 2021. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*.
- Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *International Conference on Machine Learning*.
- Chiaroni, F.; Dolz, J.; Masud, Z. I.; Mitiche, A.; and Ben Ayed, I. 2023. Parametric information maximization for generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Choi, S.; Kang, D.; and Cho, M. 2024. Contrastive Mean-Shift Learning for Generalized Category Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Davidson, I.; and Ravi, S. 2005. Clustering with constraints: Feasibility issues and the k-means algorithm. In *Proceedings of the SIAM international conference on data mining*. SIAM.
- Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*.
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*.
- Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*.
- Fei, Y.; Zhao, Z.; Yang, S.; and Zhao, B. 2022. Xcon: Learning with experts for fine-grained category discovery. *British Machine Vision Conference*.
- Fini, E.; Sangineto, E.; Lathuiliere, S.; Zhong, Z.; Nabi, M.; and Ricci, E. 2021. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Ge, Y.; Chen, D.; Zhu, F.; Zhao, R.; and Li, H. 2020. Self-paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID. In *Advances in Neural Information Processing Systems*.
- Gutmann, M.; and Hyvärinen, A. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *AISTATS*.
- Han, K.; Rebuffi, S.-A.; Ehrhardt, S.; Vedaldi, A.; and Zisserman, A. 2021. Autonovel: Automatically discovering and learning novel visual categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Hartigan, J. A.; and Wong, M. A. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*.
- He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Johnson, S. C. 1967. Hierarchical clustering schemes. *Psychometrika*.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised Contrastive Learning. *arXiv preprint arXiv:2004.11362*.
- Kim, H.; Suh, S.; Kim, D.; Jeong, D.; Cho, H.; and Kim, J. 2023. Proxy anchor-based unsupervised learning for continuous generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Klein, D.; Kamvar, S. D.; and Manning, C. D. 2002. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *International Conference on Machine Learning*.
- Krause, J.; Stark, M.; Deng, J.; and Fei-Fei, L. 2013. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*.
- Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.
- Lange, T.; Law, M. H.; Jain, A. K.; and Buhmann, J. M. 2005. Learning with constrained and unlabelled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE.
- Li, J.; Zhou, P.; Xiong, C.; and Hoi, S. 2021. Prototypical Contrastive Learning of Unsupervised Representations. In *International Conference on Learning Representations*.
- MacQueen, J.; et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA.
- Maji, S.; Rahtu, E.; Kannala, J.; Blaschko, M.; and Vedaldi, A. 2013. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*.
- Murtagh, F.; and Contreras, P. 2012. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.
- Oliver, A.; Odena, A.; Raffel, C. A.; Cubuk, E. D.; and Goodfellow, I. 2018. Realistic evaluation of deep semi-supervised learning algorithms. *Advances in neural information processing systems*.
- Pu, N.; Zhong, Z.; and Sebe, N. 2023. Dynamic Conceptual Contrastive Learning for Generalized Category Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

- Rosvall, M.; and Bergstrom, C. T. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the national academy of sciences*.
- Ruiz, C.; Spiliopoulou, M.; and Menasalvas, E. 2010. Density-based semi-supervised clustering. *Data mining and knowledge discovery*.
- Sun, Y.; and Li, Y. 2022. Opencon: Open-world contrastive learning. *Transactions on Machine Learning Research*.
- Tan, K. C.; Liu, Y.; Ambrose, B.; Tulig, M.; and Belongie, S. 2019. The herbarium challenge 2019 dataset. *Workshop on Fine-Grained Visual Categorization*.
- Vaze, S.; Han, K.; Vedaldi, A.; and Zisserman, A. 2021. Open-set recognition: A good closed-set classifier is all you need? *arXiv preprint arXiv:2110.06207*.
- Vaze, S.; Han, K.; Vedaldi, A.; and Zisserman, A. 2022. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Vaze, S.; Vedaldi, A.; and Zisserman, A. 2024. No Representation Rules Them All in Category Discovery. *Advances in Neural Information Processing Systems*.
- Wagstaff, K.; Cardie, C.; Rogers, S.; Schrödl, S.; et al. 2001. Constrained k-means clustering with background knowledge. In *International Conference on Machine Learning*.
- Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The caltech-ucsd birds-200-2011 dataset.
- Wang, G.; Zhan, Y.; Wang, X.; Song, M.; and Nahrstedt, K. 2022. Hierarchical semi-supervised contrastive learning for contamination-resistant anomaly detection. In *European Conference on Computer Vision*. Springer.
- Wang, H.; Vaze, S.; and Han, K. 2024. SPTNet: An Efficient Alternative Framework for Generalized Category Discovery with Spatial Prompt Tuning. In *The Twelfth International Conference on Learning Representations*.
- Wang, M.; Lai, B.; Chen, H.; Huang, J.; Gong, X.; and Hua, X.-S. 2021a. Towards Precise Intra-camera Supervised Person Re-Identification. In *IEEE/CVF Winter Conference on Applications of Computer Vision*.
- Wang, M.; Lai, B.; Huang, J.; Gong, X.; and Hua, X.-S. 2021b. Camera-aware Proxies for Unsupervised Person Re-Identification. In *AAAI Conference on Artificial Intelligence*.
- Wen, X.; Zhao, B.; and Qi, X. 2023. Parametric classification for generalized category discovery: A baseline study. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Wu, Z.; Xiong, Y.; Yu, S. X.; and Lin, D. 2018. Unsupervised Feature Learning via Non-Parametric Instance Discrimination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Xiao, T.; Li, S.; Wang, B.; Lin, L.; and Wang, X. 2017. Joint Detection and Identification Feature Learning for Person Search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Xing, E.; Jordan, M.; Russell, S. J.; and Ng, A. 2002. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*.
- Yin, X.; Chen, S.; Hu, E.; and Zhang, D. 2010. Semi-supervised clustering with metric learning: An adaptive kernel method. *Pattern Recognition*.
- Zhang, S.; Khan, S.; Shen, Z.; Naseer, M.; Chen, G.; and Khan, F. S. 2023. 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*.
- Zhao, B.; Wen, X.; and Han, K. 2023. Learning Semi-supervised Gaussian Mixture Models for Generalized Category Discovery. *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Zhong, Z.; Zheng, L.; and Li, S. 2017. Re-ranking Person Re-identification with k-Reciprocal Encoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.