SelEx: self-expertise in fine-grained generalized category discovery

Rastegar, S.; Salehi, M.; Asano, M.Y.; Doughty, H.R.; Snoek, M.G.C.; Leonardis, A.; ... ; Varol, G.

SelEx: Self-expertise in Fine-Grained Generalized Category Discovery

Sarah Rastegar[1(✉)], Mohammadreza Salehi[1], Yuki M. Asano[1], Hazel Doughty[2], and Cees G. M. Snoek[1]

[1] University of Amsterdam, Amsterdam, The Netherlands
s.rastegar2@uva.nl
[2] Leiden University, Leiden, The Netherlands

Abstract. In this paper, we address Generalized Category Discovery, aiming to simultaneously uncover novel categories and accurately classify known ones. Traditional methods, which lean heavily on self-supervision and contrastive learning, often fall short when distinguishing between fine-grained categories. To address this, we introduce a novel concept called 'self-expertise', which enhances the model's ability to recognize subtle differences and uncover unknown categories. Our approach combines unsupervised and supervised self-expertise strategies to refine the model's discernment and generalization. Initially, hierarchical pseudo-labeling is used to provide 'soft supervision', improving the effectiveness of self-expertise. Our supervised technique differs from traditional methods by utilizing more abstract positive and negative samples, aiding in the formation of clusters that can generalize to novel categories. Meanwhile, our unsupervised strategy encourages the model to sharpen its category distinctions by considering within-category examples as 'hard' negatives. Supported by theoretical insights, our empirical results showcase that our method outperforms existing state-of-the-art techniques in Generalized Category Discovery across several fine-grained datasets. Our code is available at: https://github.com/SarahRastegar/SelEx.

Keywords: Generalized Category Discovery · Fine-Grained Classification · Hierarchical Representation Learning

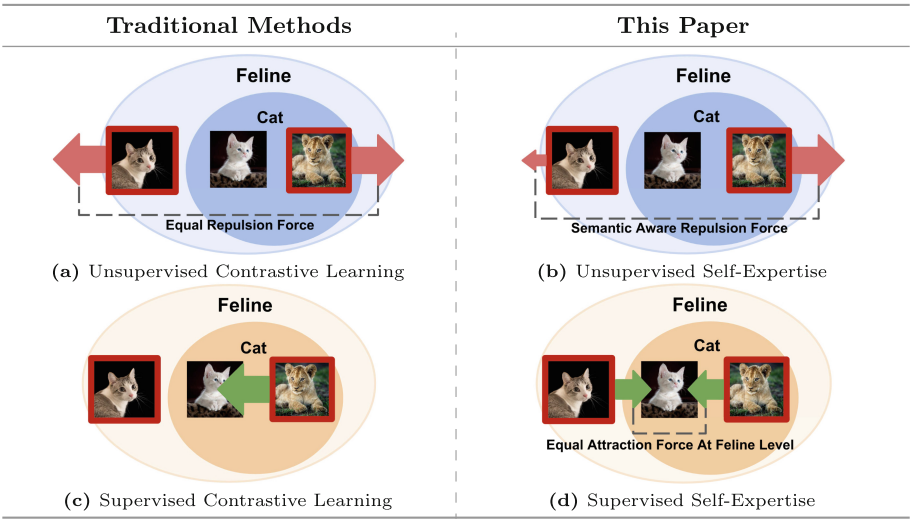
1 Introduction

Supervised learning has proven its effectiveness in classifying predefined image categories [20,21,29,44,45]. However, it struggles significantly when presented with unknown categories, hindering its real-world applicability [4,32,43,48,64]. Generalized Category Discovery (GCD) addresses this limitation by automatically identifying both known and novel categories from unlabeled data

Y. M. Asano—Currently at University of Technology Nuremberg.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-72897-6_25.

A. Leonardis et al. (Eds.): ECCV 2024, LNCS 15128, pp. 440–458, 2025.
https://doi.org/10.1007/978-3-031-72897-6_25



(a) Unsupervised Contrastive Learning

(b) Unsupervised Self-Expertise

(c) Supervised Contrastive Learning

(d) Supervised Self-Expertise

Fig. 1. The motivation for self-expertise *(a) Unsupervised Contrastive Learning.* shows identical repulsion for both misclassified cat and lion. *(b) Unsupervised Self-Expertise.* focuses on distinguishing hard negatives within a category cluster, applying less repulsion to external members, and varying the repulsion for misclassified samples, resulting in milder repulsion for the mislabeled cat and stronger for the lion. *(c) Supervised Contrastive Learning.* focuses on attracting similar category members, leaving others unaffected. *(d) Supervised Self-Expertise.* graduates the attraction of samples based on semantic similarity. Both a misclassified cat and lion are equally attracted to the cat sample at the feline level. Supervised and unsupervised self-expertise together enhance accuracy by attracting similar samples and repelling dissimilar ones.

[1,5,10,19,41,49,50,55,61]. A key approach for handling unknown categories within GCD has been self-supervision through contrastive learning [5,6,9,18,23,30,31,34,59]. However, this method struggles with fragmented clustering and an increased false negative rate, particularly in fine-grained categorization where positive augmented samples may significantly differ from their negative counterparts from the same category, which leads to misclassification [6,12,22,24]. Although supervised contrastive learning [25,49] improves discrimination among known categories, it struggles with unknown categories due to the absence of supervisory signals. Our work navigates this essential trade-off, aiming to merge the discovery of unknown categories with fine-grained classification through self-supervision.

In this paper, we present a novel Generalized Category Discovery approach that combines contrastive learning with pseudo-labeling to uncover novel categories by enhancing self-expertise. We define 'expertise' as the skill to generalize across different abstraction levels, much like an ornithologist distinguishes between species at various levels, a capability that extends beyond ordinary observation. Our method focuses on honing the ability to identify subtle distinctions and achieve broad generalizations. In Fig. 1, we illustrate how our model

improves the detection of fine details and generalization through unsupervised and supervised self-expertise, respectively. We make four contributions:

- We present a hierarchical semi-supervised k-means clustering approach that better initializes unknown clusters using known centers and addresses cluster sparsity by balancing distributions through a stable matching algorithm.
- We propose an unsupervised self-expertise approach that emphasizes hard negative samples with identical labels at each hierarchical level.
- We present supervised self-expertise, which utilizes abstract pseudo-labels to generate weaker positive and stronger negative instances, facilitating rapid initial category clustering and enhancing generalization to novel categories.
- Empirically and theoretically, we demonstrate that our approach facilitates effective generalized category discovery with fine-grained abilities.

2 Related Works

Generalized Category Discovery was introduced concurrently by Vaze *et al.* [49] and Cao *et al.* [5]. It provides models with unlabeled data from both novel and known categories, placing it within the realm of semi-supervised learning [8,35,38,42,58]. The unique challenge in generalized category discovery is handling categories without any labeled instances alongside already seen categories. There are primarily two approaches to address this challenge. One employs a series of prototypes as reference points, *e.g.*, [1,10,11,19,26,47,53,55–57,60]. The second leverages local similarity as weak pseudo-labels per sample by utilizing sample similarities to form local clusters [14,16,40,41,61,63] or by utilizing mean-teacher networks to address the challenges posed by noisy pseudo-labels [50,52,55,61]. Nonetheless, the foundation of these approaches is contrastive learning, which has previously been shown to falter in fine-grained classification [12] due to strong augmentations in positives in comparison to nuanced visual differences between samples of the same category. To alleviate this, we introduce 'self-expertise' aimed at hierarchical learning of known and unknown categories. Our method is particularly effective in overcoming the limited availability of positive samples per category and enhancing the identification of subtle differences among negative samples, which we deem the biggest challenge in fine-grained classification.

Hierarchical Representation Learning. Different approaches benefit from hierarchical categories. Zhang *et al.* [62] use multiple label levels to enhance their representation through hierarchical contrastive learning. Guo *et al.* [17] extract pseudo labels for hierarchical contrastive learning, where signals are positive within the same cluster. We also use hierarchical pseudo-labels, but instead employ negative samples from the same cluster for generalized category discovery.

Otholt *et al.* [37] and Banerjee *et al.* [3] proposed hierarchical approaches to address generalized category discovery. These works leverage neighborhood structures to delineate refined categories. Rastegar *et al.* [41] learn an implicit category tree, facilitating hierarchical self-coding of categories, which maintains category similarity across all hierarchy levels. Differing from these works, our method

leverages weak supervision from samples within each level of the hierarchy, which reduces misclassification impact on lower levels. Additionally, our focus on hard negatives for unsupervised self-expertise enhances the model's ability to discern nuanced distinctions, leading to better fine-grained classification.

3 Theoretical Framework for Self-Expertise

Notations. We denote the number of total categories with K and the number of samples by N. For each random variable c, we indicate the number of associated samples by $|c|$. We use 'ln' for the natural logarithm and 'lg' for $\log_2$.

Problem Definition. The challenge of generalized category discovery lies in classifying samples during inference as belonging to categories encountered during training or as entirely novel categories. To describe this formally, throughout the training phase, we have access to the input $\mathcal{X}_\mathcal{S}$ for labeled and $\mathcal{X}_\mathcal{U}$ for unlabelled data. However, our access to the labels is limited to $\mathcal{Y}_\mathcal{S}$, which represents the known categories. Our objective is to categorize unlabeled samples. The possible labels for unlabeled input consist of both known and novel categories denoted as $\mathcal{Y}_\mathcal{U}$, where $\mathcal{Y}_\mathcal{S} \subset \mathcal{Y}_\mathcal{U}$. We formulate the generalized category discovery problem as the Bayesian network in Fig. 2. In this network, $\mathbf{x_i}$ and $\mathbf{x_j}$ represent distinct samples or alternative perspectives of the same sample. The associated ground truth category variables, $\mathbf{c_i}$ and $\mathbf{c_j}$, are the focus of our information extraction process. In this Bayesian Network, we assume these two variables determine the distribution of $\mathbf{x_i}$ and $\mathbf{x_j}$. Here z_j indicates the latent representation of the model for category variables, which is derived from $\mathbf{x_j}$. Contrastive training aims to estimate the equal distribution of the ground-truth category random variables accurately. Thus, for any pair of samples i and j, our target is to approximate the following distribution closely:

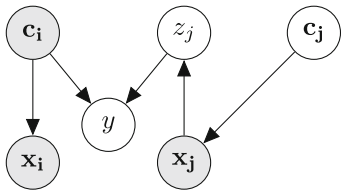


Fig. 2. Bayesian Network for the Generalized Category Discovery. Shaded nodes are observed variables $\mathbf{x_i}$, $\mathbf{x_j}$ corresponds to images i and j, and $\mathbf{c_i}$ and $\mathbf{c_j}$ which are the ground-truth category variable. z_j is the latent category variable extracted from the model.

$$p(y{=}1|\mathbf{c_i}, \mathbf{c_j}){=}\mathbb{1}(\mathbf{c_i}{=}\mathbf{c_j}), \tag{1}$$

where $\mathbb{1}$ signifies the identity operator that yields one when its internal condition is satisfied and zero otherwise. Our approach involves minimizing the Kullback-Leibler (KL) divergence between the actual distribution p and the estimated model distribution $\hat{p}$:

$$D_{\mathrm{KL}}[p(y|\mathbf{c_i}, \mathbf{c_j}) \parallel \hat{p}(y|\mathbf{x_i}, \mathbf{x_j})]. \tag{2}$$

When addressing Generalized Category Discovery in the context of labeled samples, $\mathbf{c_i}$ is treated as observed, while for unlabeled samples, both category variables are considered unobserved.

Supervised Self-Expertise. In supervised contrastive learning, it is assumed that one of the context variables is observable. This assumption facilitates the parameter training process by directing it through the estimation of the conditional probability $\hat{p}(y|\mathbf{c_i}, \mathbf{x_j})$. When dealing with a balanced dataset, the probabilities are uniformly distributed, such that $p(\mathbf{c_i})=\frac{1}{K}$ and $\hat{p}(z_j=k)=\frac{1}{K}$, where K is the total number of classes. Using the Bayesian Network from Fig. 2, we can derive an upper bound for the estimation discrepancy in supervised contrastive learning:

$$D_{\mathrm{KL}}[p(y|\mathbf{c_i}, \mathbf{c_j}) \parallel \hat{p}(y|\mathbf{c_i}, \mathbf{x_j})] \leq \ln \frac{N}{K}. \tag{3}$$

The derivation details are provided in the Appendix. Here, we assume $\mathbf{c_i}$ is fixed, thus considering all the labels for supervised contrastive learning results in:

$$D_{\mathrm{KL}}[p(y|c_i, c_j) \parallel \hat{p}(y|c_i, x_j)] \leq K \ln \frac{N}{K}. \tag{4}$$

In Eq. (4), it is evident that reducing the value of K diminishes the upper bound as long as $K > \ln \frac{N}{K}$, thereby aligning the model's distribution more closely with the true distribution. Using this property facilitates the modulation of abstraction levels across diverse categories. Denoting the upper limit for K categories as $\mathcal{S}_K$, we observe that employing $\frac{K}{2}$ categories, instead of K, results in:

$$\mathcal{S}_K = 2\mathcal{S}_{\frac{K}{2}} - K \ln 2. \tag{5}$$

This suggests that a reduction in K correlates with a decrease in the upper bound $\mathcal{S}_{\frac{K}{2}} > K \ln 2$. Let's consider implementing a hierarchical structure wherein, at each stage, the distribution is approximated by bifurcating into two categories. Subsequently, within each bifurcation, we estimate $\frac{K}{2}$ independently, thus effectively dealing with K categories in a hierarchically extracted manner. We denote the upper bound for this hierarchical scheme as $\hat{\mathcal{S}}_K$, we show in the Appendix:

$$\hat{\mathcal{S}}_K \leq \hat{\mathcal{S}}_{\frac{K}{2}}. \tag{6}$$

Thus, by leveraging a hierarchical approach, we observe a reduction in the upper bound of model error as the granularity of categories increases. This phenomenon underpins our introduction of supervised self-expertise, whereby the model refines its detection granularity. By hierarchically augmenting the resolution of detection (denoted as K), our approach aligns the model's granularity with the ground truth labels, optimizing model performance in categorization.

Unsupervised Self-Expertise. In unsupervised contrastive learning, both c_i and c_j are unobserved, necessitating the approximation of this distribution solely based on inputs. This means we can rewrite Eq. (4) as:

$$D_{\mathrm{KL}}[p(y|c_i, c_j) \parallel \hat{p}(y|x_i, x_j)] \leq \frac{K(K+1)}{2} \ln \frac{N}{K}. \tag{7}$$

Notably, the upper bound decreases with the reduction of K. However, in contrast to its supervised counterpart, the presence of K^2 in the upper bound

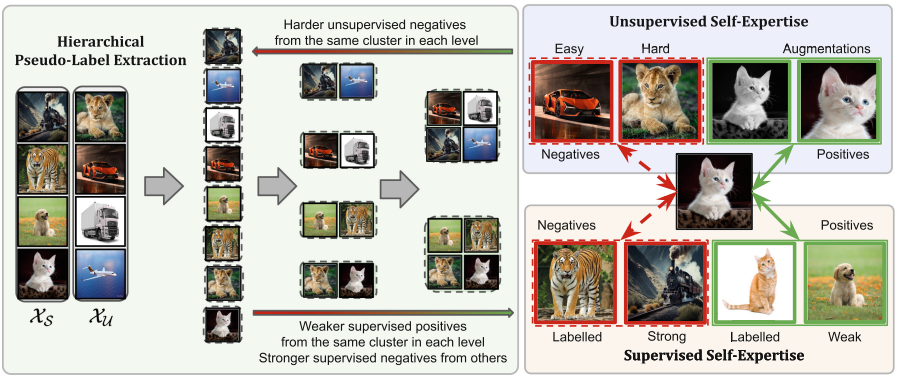


Fig. 3. Self-expertise for Generalized Category Discovery. Our method integrates three key components. The initial component is the Hierarchical Semi-Supervised K-means, which extracts pseudo-labels across multiple levels of expertise. Utilizing these pseudo-labels, the second component adopts unsupervised self-expertise by identifying hard negatives within each pseudo-label for enhanced differentiation across expertise tiers. The final component applies supervised self-expertise, recognizing samples sharing the same pseudo-label as weak positives to boost positive sample frequency while employing external pseudo-labels as strong negatives. This strategy accelerates cluster formation by capturing abstract-level similarities.

precludes the adoption of a hierarchical strategy to mitigate the upper bound by merely increasing K. To address this limitation, we propose an alternative method that refines the consideration of the KL divergence. Specifically, rather than evaluating the KL divergence across all pairs (c_i, c_j), our approach focuses on pairs where $c_i{=}c_j$, which we call hard negatives for the hierarchical approach while considering the rest of pairs in a traditional unsupervised contrastive learning. This modification restricts the analysis to negative samples within the same category, thereby offering a pragmatic approximation of the overall KL divergence while maintaining a tractable upper bound:

$$D_{\mathrm{KL}}[p(y|c_i, c_j, c_i{=}c_j) \parallel \hat{p}(y|x_i, x_j)] \leq K \ln \frac{N}{K}. \tag{8}$$

Given the inaccessibility of c_i and c_j, our approach relies on the premise that z_i and z_j are equivalent for the selection of negative samples. Let's denote the upper bound in Eq. (8) as $\mathcal{U}_K$. In a manner analogous to the procedure employed in the supervised variant, the application of a hierarchical categorization strategy necessitates the introduction of an adjusted upper bound, represented by $\hat{\mathcal{U}}_K$:

$$\hat{\mathcal{U}}_K \leq \hat{\mathcal{U}}_{\frac{K}{2}}. \tag{9}$$

This observation serves as the foundational motivation for our approach, which employs hierarchical hard negatives to develop unsupervised self-expertise. At every hierarchical level, our approach emphasizes the selection of negative samples from identical categories. This strategy mirrors our earlier derivation on

supervised self-expertise, wherein we employed hierarchical labels to systematically reduce the upper bound delineated in Eq. (8). The reduction process continues incrementally until the resolution of the labels matches that of the ground truth. This methodology underpins our effort to enhance the precision of model predictions in an unsupervised learning context, bridging the gap towards achieving granular accuracy that parallels the fidelity of ground truth annotations. Note that while the upper bound discussed here is derived based on ln, transitioning to lg alters the upper bound only by a constant factor. Since halving the category count reduces $\lg K$ by one, we opt for lg over ln in the rest of the paper.

4 Self-Expertise for Generalized Category Discovery

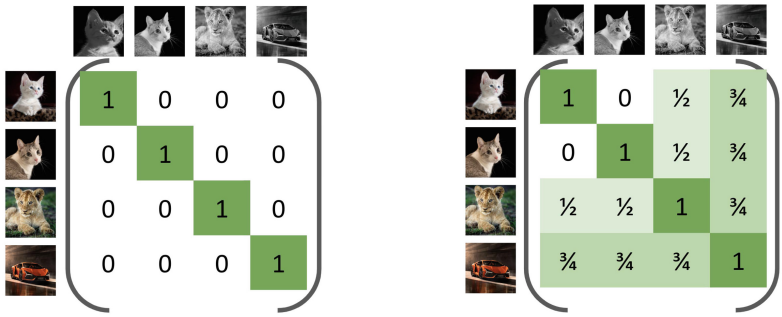
Our proposed method for fine-grained generalized category discovery has three components: hierarchical pseudo-label extraction, unsupervised self-expertise, and supervised self-expertise. As illustrated in Fig. 3, each phase synergistically contributes to achieving discriminative clustering, which is pivotal for the task.

Hierarchical Pseudo-Label Extraction. This component addresses the challenge of optimizing supervisory signals while avoiding the erroneous allocation of unknown category samples to known categories. To achieve this, we implement a multi-tiered approach to pseudo-labeling for unlabeled samples, forming the foundation of our pseudo-label hierarchy.

Pseudo-label Initialization via Balanced Semi-Supervised K-means. We propose the Balanced Semi-Supervised K-means (BSSK) algorithm. This algorithm generates pseudo-labels for the initial level of the subsequently established pseudo-label hierarchy. BSSK starts by establishing K-means centers for known categories by determining cluster centers for already labeled data. For novel categories, we select an equivalent number of random samples as cluster centers, ensuring that each cluster maintains a uniform size. This process yields the base level of our hierarchy, aligning pseudo-labels with the granularity of ground truth categories.

Hierarchical Expansion and Abstraction. Based on BSSK, we introduce Hierarchical Semi-Supervised K-means (HSSK). For each subsequent kth level of abstraction, HSSK clusters the $k-1$th level's seen prototypes into half, effectively creating higher-level abstractions. All seen labels are projected onto these new hyperlabels. This is followed by BSSK, now with doubled cluster size compared to the previous level. This hierarchical structuring allows us to generate progressively abstracted and reliable pseudo-labels across various levels of category granularity. Pseudo-code for BSSK and HSSK is provided in the Appendix.

Unsupervised Self-Expertise. In our approach, we confront the challenges posed by pseudo-labeling in early training stages, where model proficiency with known and unknown labels is limited. Pseudo-labels generated during this phase are often noisy, leading to sub-optimal model training. To mitigate this, we integrate unsupervised contrastive learning. However, this technique focuses



(a) Unsupervised Contrastive Learning **(b)** Unsupervised Self-Expertise

Fig. 4. Illustrating the distinction in target matrix formulation for unsupervised self-expertise *(a) Unsupervised Contrastive Learning.* Each sample's augmented version is deemed positive, with all other samples marked as negative. *(b) Unsupervised Self-Expertise.* On the contrary, our method dynamically adjusts the negativity weight of each sample according to semantic similarity, treating categories with higher similarity (e.g., within the 'cat' category) as more strict negatives. Conversely, semantically different categories (e.g., 'lion' vs. 'cat') incorporate a degree of uncertainty in their negativity, quantified as $\frac{1}{2}$ to reflect the semantic differences between negatives. Since the target matrix represents probabilities, normalization is required to ensure validity.

on distinguishing augmented versions of a sample from others, including those within the same semantic context, potentially aggravating the initial issue.

To address these concerns, we adopt a strategy where the model is instructed to exclusively distance samples within the same clusters (pseudo-labels). This tactic might seem counterintuitive at first glance. However, it is fundamentally based on the notion that distancing a visually similar sample within the same cluster can significantly enhance the purity of that cluster. In contrast, samples that are semantically similar but belong to different clusters are not considered negative instances. This allows the model to either assimilate these samples during the training phase via supervised contrastive learning or to segregate them from other clusters. Our approach also systematically shifts focus towards more abstract category levels while simultaneously diminishing the importance of negative samples from these broader clusters. For instance, in traditional unsupervised contrastive learning, samples i and j are associated with an identity target matrix I, where $I_{ij}{=}\mathbb{1}(i{=}j)$. In contrast, our unsupervised self-expertise necessitates the recalibration of these targets to reflect the semantic similarity between the two samples. To illustrate, we define the pseudo-label for samples i and j at the hierarchical level k as $\mathbf{c_i}^k$ and $\mathbf{c_j}^k$, respectively. Consequently, we introduce an adjusted target matrix Y, comprising elements y_{ij}, calculated as:

$$y_{ij} = \sum_{k=1}^{\lg K} \frac{\mathbb{1}(\mathbf{c_i}^k \neq \mathbf{c_j}^k)}{2^k}. \tag{10}$$

A comparison between the proposed adjusted target matrix and the conventional target matrix is illustrated in Fig. 4. Since the y_{ij}s will be interpreted as probabilities, the final Y target matrix should be normalized. As depicted in Fig. 4a, standard unsupervised contrastive learning treats all negative instances uniformly, thereby ignoring their semantic dissimilarities. Conversely, our unsupervised self-expertise employs a refined target matrix, where cat instances are classified as strict negatives in Fig. 4b, while the negativity of other instances is modulated according to their semantic distance from the positive instance. It is important to note that a linear combination of these target matrices can be employed, allowing for adjustment based on the specific granularity required by the task as:

$$\hat{Y} = \alpha Y + (1-\alpha) I, \tag{11}$$

where α represents the hyperparameter associated with label smoothing. Through empirical analysis, we demonstrate that an increased value of α encourages the model to pay greater attention to more nuanced details. Conversely, a reduced α value renders the model more adept at handling tasks that require a broader, more general approach. As a result, for the contrastive logits P derived from our model, we use the binary cross entropy loss $\mathcal{L}_{\text{BCE}}$ to formulate the unsupervised self-expertise loss, $\mathcal{L}_{\text{USE}}$, as follows:

$$\mathcal{L}_{\text{USE}} = \mathcal{L}_{\text{BCE}}(P, Y). \tag{12}$$

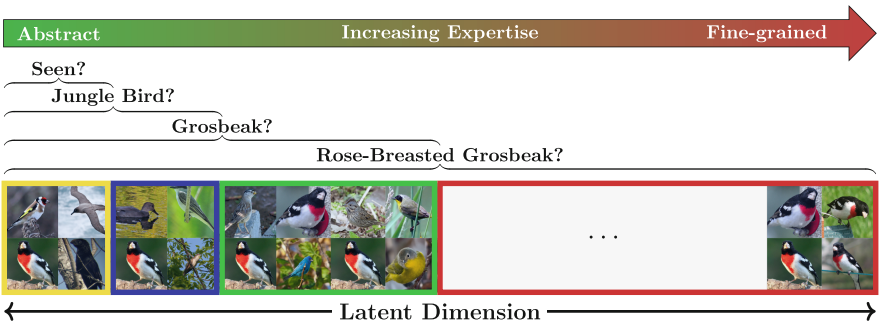


Fig. 5. Supervised self-expertise. Similar to playing a game of twenty questions, our method employs supervised self-expertise to discern sample attributes. As the process evolves, the attributes it discerns between are increasingly specific. Focusing on the Grosbeak classification, the model initially utilizes the leftmost segment of its latent representation (indicated by the yellow square) to ascertain whether a sample is seen. The model then allocates its representation's yellow and blue square parts to identify whether the subject is a jungle bird. Subsequently, it dedicates the latter part (represented by the green rectangle) to determine if this jungle bird is a Grosbeak. (Color figure online)

Supervised Self-Expertise. Generated pseudo-labels from our hierarchical pseudo-label extraction are utilized as supervisory signals for the next epoch. Consider $\mathcal{L}s^k$ as the supervised contrastive learning specific to the pseudo-label level k. The aggregate supervised contrastive learning loss is represented by:

$$\mathcal{L}_{\text{SSE}} = \frac{1}{2}(\sum_{k=0}^{\lg K} \frac{\mathcal{L}_s^k|\frac{D}{2^k}}{2^k}), \tag{13}$$

where the term $\mathcal{L}_s^k|\frac{D}{2^k}$ reflects the supervised loss applied exclusively to the initial segment $\frac{D}{2^k}$ of the embedding vector D. This approach is grounded in the premise that higher hierarchy levels encounter an increased frequency of positive pairs. Yet, the ultimate objective is to learn pseudo-labels aligned with the ground truth labels. Consequently, the model is constrained to use only the first $\frac{D}{2^k}$ segment of the embedding for differentiating pseudo-labels at hierarchy level k. This implies that for distinguishing between various pseudo-labels at level $k-1$, which share a common higher-level pseudo-label at level k, the embedding dimensions from $\frac{D}{2^k}$ to $\frac{D}{2^{k-1}}$ are utilized. When k=0, we only use groundtruth labels for samples in known categories and utilize the full embedding vector for supervised contrastive learning. This ensures accurate label assignment for known categories upon training completion and facilitates the generation of informative pseudo-labels for novel categories. Utilizing different abstraction levels, we apply supervised contrastive learning to samples within the same cluster at different levels of hierarchy. The application of our supervised expertise to the representation is illustrated in Fig. 5. Finally, for the tunable hyperparameter λ, our overall self-expertise loss function is expressed as:

$$\mathcal{L}_{\text{SE}} = (1-\lambda)\mathcal{L}_{\text{USE}} + \lambda\mathcal{L}_{\text{SSE}}. \tag{14}$$

5 Experiments

5.1 Experimental Setup

Datasets. We assess the efficacy of our approach on four fine-grained datasets: CUB-200 [51], FGVC-Aircraft [33], Stanford-Cars [27] and Oxford-IIIT Pet [39]. Additionally, we demonstrate the adaptability of our method to more coarse-grained datasets CIFAR10 [28], CIFAR100 [28] and ImageNet-100 [13], highlighting its broader applicability beyond fine-grained classification tasks. Finally, in the Appendix experiments section, we report on the challenging Herbarium-19 dataset [46], which is fine-grained and long-tailed, to show that our approach is effective even with non-uniform category distributions. Detailed statistics of the datasets along with their train/test splits are also provided in the Appendix.

Table 1. Comparison with state-of-the-art for fine-grained image classification. Bold and underlined numbers indicate the best and second-best accuracies, respectively. Our method is well suited for fine-grained datasets, profits from stronger backbones, and has strong performance for all three experimental settings (*All*, *Known*, and *Novel*)

	Method	CUB-200			FGVC-Aircraft			Stanford-Cars			Average		
		All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel
DINOv1	ORCA† [5]	36.3	43.8	32.6	31.6	32.0	31.4	31.9	42.2	26.9	33.3	39.3	30.3
	GCD [49]	51.3	56.6	48.7	45.0	41.1	46.9	39.0	57.6	29.9	45.1	51.8	41.8
	GPC [63]	52.0	55.5	47.5	43.3	40.7	44.8	38.2	58.9	27.4	44.5	51.7	39.9
	XCon [15]	52.1	54.3	51.0	47.7	44.4	49.4	40.5	58.8	31.7	46.8	52.5	44.0
	SimGCD [55]	60.3	65.6	57.7	54.2	59.1	51.8	53.8	71.9	45.0	56.1	65.5	51.5
	PIM [10]	62.7	75.7	56.2	–	–	–	43.1	66.9	31.6	–	–	–
	PromptCAL [61]	62.9	64.4	62.1	52.2	52.2	52.3	50.2	70.1	40.6	55.1	62.2	51.7
	DCCL [40]	63.5	60.8	64.9	–	–	–	43.1	55.7	36.2	–	–	–
	AMEND [3]	64.9	75.6	59.6	52.8	61.8	48.3	56.4	73.3	48.2	58.0	70.2	52.0
	μGCD [50]	65.7	68.0	64.6	53.8	55.4	53.0	56.5	68.1	**50.9**	58.7	63.8	56.2
	SPTNet [52]	65.8	68.8	65.1	**59.3**	61.8	**58.1**	**59.0**	**79.2**	49.3	61.4	69.9	57.5
	CMS [11]	68.2	76.5	64.0	56.0	63.4	52.3	56.9	76.1	47.6	60.4	72.0	54.6
	GCA [37]	68.8	73.4	66.6	52.0	57.1	49.5	54.4	72.1	45.8	58.4	67.5	54.0
	InfoSieve [41]	69.4	**77.9**	65.2	56.3	63.7	52.5	55.7	74.8	46.4	60.5	**72.1**	54.7
	TIDA [54]	–	–	–	54.6	61.3	52.1	54.7	72.3	46.2	–	–	–
	SelEx (Ours)	**73.6**	75.3	**72.8**	57.1	**64.7**	53.3	58.5	75.6	50.3	**63.0**	71.9	**58.8**
DINOv2	GCD* [49]	71.9	71.2	72.3	55.4	47.9	58.5	65.7	67.8	64.7	64.3	62.3	65.4
	SimGCD* [55]	71.5	78.1	68.3	63.9	69.9	60.9	71.5	81.9	66.6	69.0	76.6	65.3
	μGCD* [50]	74.0	75.9	73.1	66.3	68.7	65.1	76.1	91.0	68.9	72.1	78.5	69.0
	SelEx (Ours)	**87.4**	**85.1**	**88.5**	**79.8**	**82.3**	**78.6**	**82.2**	**93.7**	**76.7**	**83.1**	**87.0**	**81.3**

* reported from [50] and † reported from [61].

Implementation Details. In our experiments, we adhered to the dataset division proposed by Vaze *et al.* [49], where half of the categories in each dataset are designated as known, except for CIFAR100, where 80% are used as known categories. The labeled set consists of 50% of the samples from these known categories. The remainder of the known category data, along with all data from novel categories, comprise the unlabeled set. Following [49], we use ViT-B/16 as our backbone, which is either pre-trained by DINOv1 [7] on unlabelled ImageNet 1K [29], or pretrained by DINOv2 [36] on unlabelled ImageNet 22K. We use the batch size of 128 for training and set λ=0.35. For label smoothing, we use α=0.5 for fine-grained datasets and α=0.1 for coarse-grained datasets. Different from [49], we froze the first 10 blocks of ViT-B/16 and fine-tuned the last two blocks instead of only the last one to have more parameters given that for each level, only a fraction of the latent dimension is considered.

5.2 Comparison with State-of-the-Art

Fine-Grained Image Classification. We evaluate our model's effectiveness across three fine-grained datasets in Table 1. The results demonstrate our method's capability in handling fine-grained categories, as it consistently outperforms others in both all and novel category classification within these datasets. The success can be attributed to the model's hierarchical approach to category analysis, which is pivotal in differentiating between closely related categories that demand acute attention to specific details. Additionally, as indicated in Table 2, our method also leads in performance for both all and novel categories in the Oxford Pet dataset. Despite its small size, which typically poses a risk of overfitting, our model's strong performance on this dataset further indicates its robustness.

Table 2. Comparison with state-of-the-art for Oxford-IIIT Pet classification. Since the Oxford Pet dataset is small, these results demonstrate our methods' robustness to overfitting.

Method	Oxford-IIIT Pet		
	All	Known	Novel
k-means [2]	77.1	70.1	80.7
GCD [49]	80.2	85.1	77.6
XCon [15]	86.7	91.5	84.1
DCCL [40]	88.1	88.2	88.0
InfoSieve [41]	91.8	**92.6**	91.3
SelEx (DINOv1)	**92.5**	91.9	**92.8**
SelEx (DINOv2)	**95.6**	**96.5**	**95.1**

Coarse-Grained Image Classification. We also evaluate our model on three coarse-grained datasets: CIFAR10/100 [28] and ImageNet-100 [13]. Table 3 presents a comparative analysis of our proposed method with existing state-of-the-art approaches in generalized category discovery. Our method, originally designed for fine-grained category discovery, demonstrates competitive performance on coarse-grained datasets across both known and novel categories. Despite the potential shallow or absent hierarchical structures in these datasets, our approach shows a notable enhancement in performance over the traditional non-hierarchical baseline method, GCD [49]. Figure 6 presents a radar chart comparing the performance of our proposed method with that of the state-of-the-art methods across various datasets. Specifically, we contrast our approach against InfoSieve [41] for fine-grained datasets and PromptCAL [61] for coarse-grained.

5.3 Ablative Studies

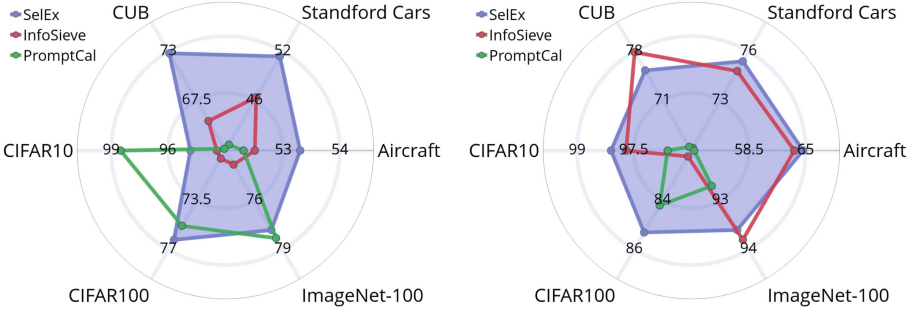
We evaluate the individual effects of method components in this section. All ablative experiments are performed on CUB with the DINOv1 backbone. We present additional ablations, time complexity, and failure cases in the Appendix.

Effect of Each Component . Table 4(a) examines the effect of our three key method components: Hierarchical Semi-Supervised K-means (HSSK), unsupervised self-expertise ($\mathcal{L}_{\text{USE}}$), and supervised self-expertise ($\mathcal{L}_{\text{SSE}}$). The results demonstrate that the Hierarchical Semi-Supervised K-means approach yields the most significant improvements across both known and novel categories. Our unsupervised self-expertise loss, denoted as $\mathcal{L}_{\text{USE}}$, shows a particular affinity for

Table 3. Comparison with state-of-the-art for coarse-grained image classification. Bold and underlined numbers show the best and second-best accuracies. Our method has a consistent performance for the three experimental settings (*All*, *Known*, *Novel*). Our method is especially suitable for known categories in all three datasets.

	Method	CIFAR-10			CIFAR-100			ImageNet-100			Average		
		All	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel
DINOv1	ORCA† [5]	96.9	95.1	97.8	74.2	82.1	67.2	79.2	93.2	72.1	83.4	90.1	79.0
	GCD [49]	91.5	97.9	88.2	73.0	76.2	66.5	74.1	89.8	66.3	79.5	88.0	73.7
	GPC [63]	90.6	97.6	87.0	75.4	84.6	60.1	75.3	93.4	66.7	80.4	91.9	71.3
	XCon [15]	96.0	97.3	95.4	74.2	81.2	60.3	77.6	93.5	69.7	82.6	90.7	75.1
	SimGCD [55]	97.1	95.1	98.1	80.1	81.2	77.8	83.0	93.1	77.9	86.7	89.8	84.6
	PIM [10]	94.7	97.4	93.3	78.3	84.2	66.5	83.1	95.3	77.0	85.4	**92.3**	78.9
	PromptCAL [61]	97.9	96.6	98.5	81.2	84.2	75.3	83.1	92.7	78.3	87.4	91.2	84.0
	DCCL [40]	96.3	96.5	96.9	75.3	76.8	70.2	80.5	90.5	76.2	84.0	87.9	81.1
	AMEND [3]	96.8	94.6	97.8	81.0	79.9	**83.3**	83.2	92.9	78.3	87.0	89.1	**86.5**
	SPTNet [52]	97.3	95.0	**98.6**	81.3	84.3	75.6	**85.4**	93.2	**81.4**	**88.0**	90.8	85.2
	CMS [11]	–	–	–	82.3	**85.7**	75.5	84.7	**95.6**	79.2	–	–	–
	GCA [37]	95.5	95.9	95.2	**82.4**	85.6	75.9	82.8	94.1	77.1	86.9	91.9	82.7
	InfoSieve [41]	94.8	97.7	93.4	78.3	82.2	70.5	80.5	93.8	73.8	84.5	91.2	79.2
	TIDA [54]	**98.2**	97.9	98.5	82.3	83.8	80.7	–	–	–	–	–	–
	SelEx (Ours)	95.9	**98.1**	94.8	82.3	85.3	76.3	83.1	93.6	77.8	87.1	**92.3**	83.0

† reported from [61].



(a) Novel **(b)** Known

Fig. 6. Model Performances Across Diverse Datasets: PromptCAL [61] excels in coarse-grained datasets, while InfoSieve [41] specializes on fine-grained datasets. SelEx is strong on both and is especially proficient in discovering novel fine-grained categories.

enhancing known categories. This is in line with our initial hypothesis, considering that these categories benefit from supervision signals. Such signals facilitate the attraction of semantically similar samples, even if they are initially distant in the embedding space. Concurrently, this approach effectively disregards semantically similar yet distant negative samples, preventing any repulsion until they converge into the same cluster. When integrated with hierarchical semi-supervised k-means, the unsupervised self-expertise loss extends its benefits to novel categories, leveraging the presence of semantic labels. Our supervised self-expertise loss, $\mathcal{L}_{\text{SSE}}$, unsurprisingly excels in aiding novel categories while

Table 4. Ablation study on the effectiveness of each model component and hierarchy using CUB-200. *(a) Effect of Each Component* Each component contributes to our improved performance on known and novel categories. *(b) Effect of different hierarchy levels* With more hierarchy, the model's performance increases. All pseudo-labels from previous levels are also used in each level of contrastive learning.

(a) Effect of Each Component

HSSK	$\mathcal{L}_{\text{USE}}$	$\mathcal{L}_{\text{SSE}}$	All	Known	Novel
			46.1	49.2	44.5
✓			62.6	71.9	58.2
	✓		47.0	52.2	44.4
		✓	55.5	53.5	56.6
✓	✓		54.7	66.7	48.6
✓		✓	68.7	72.3	66.9
	✓	✓	56.7	51.8	59.1
✓	✓	✓	**73.6**	**75.3**	**72.8**

(b) Effect of Hierarchy Levels

Hierarchy	Pseudo-labels	All	Known	Novel
Baseline (None)	+0	62.6	71.9	58.2
Level 1	+200	63.8	74.9	58.3
Level 2	+100	69.2	**76.8**	65.3
Level 3	+50	70.0	74.5	67.9
Level 4	+24	71.0	74.4	69.4
Level 5	+12	69.0	72.5	67.2
Level 6	+6	72.5	75.4	71.1
Level 7	+2	**73.6**	75.3	**72.8**

also contributing positively to known ones. We attribute this to the fact that, although hierarchical structures are advantageous for known categories with robust label-based supervision, novel categories lack such ground-truth labels. As a result, pseudo-labels at finer granularities may introduce noise. However, as we ascend the hierarchy, these pseudo-labels for novel categories gain reliability, offering more effective supervision. In conclusion, the combination of all three components – hierarchical semi-supervised k-means, unsupervised self-expertise, and supervised self-expertise – yields the most optimal results for both known and novel categories, as demonstrated in our experiments.

The Effect of Hierarchy Level. In Table 4(b), we compare model performance across varying hierarchy levels. These hierarchy levels are incorporated into the training phase for all three model components. Specifically, the Baseline component employs supervised contrastive learning using only the ground-truth labels, which are limited to samples that have been labeled. Level 1 is identified as the base level of the hierarchy, utilizing pseudo-labels that offer semantic detail comparable to ground-truth labels, thereby enriching our dataset with an additional 200 pseudo-labels for samples without labels. As we ascend through the hierarchy levels, the quantity of pseudo-labels decreases by half, as detailed in the accompanying table, until reaching the apex level. This topmost level introduces the most abstract categorization, distinguishing between 'seen' and 'unseen' samples. The results depicted in Table 4(b) indicate a notable trend: integrating additional hierarchical levels appears to be particularly advantageous for unknown categories. This observation can be attributed to increased granularity between categories at finer hierarchy levels, resulting in heightened uncertainty and noise in pseudo-labels. This phenomenon underscores the efficacy of our model in handling complex, hierarchical category structures, especially in scenarios involving unknown category distinctions.

Effects of Smoothing Hyperparameter. In our unsupervised self-expertise, we adopted a smoothing hyperparameter α to modulate the uncertainty threshold for negative samples outside a given cluster. Specifically, when α=1, the model exclusively incorporates negative samples from its own cluster. Conversely, setting α=0 equalizes the treatment of all negative samples, aligning with traditional unsupervised contrastive learning. We conducted experiments with varying α values, as detailed in Fig. 7. Our findings indicate an enhancement in the model's performance on novel categories as α increases. This improvement is attributed to the fact that standard unsupervised contrastive learning indiscriminately distances all non-matching samples, including those with semantic similarities.

In contrast, our unsupervised self-expertise concentrates on cluster-specific samples. This allows semantically related samples outside the cluster to be less repelled, which is particularly beneficial for novel categories since they do not have ground-truth labels to counteract this repelling through supervised contrastive learning. Hence, a higher α enhances novel category identification performance. It is essential to highlight that dataset granularity can influence the choice of the hyperparameter α. Specifically, a more fine-grained dataset necessitates a larger value of α to discern the subtle differences between samples. We demonstrate the impact of α to balance the probability and uncertainty on various datasets in the Appendix.

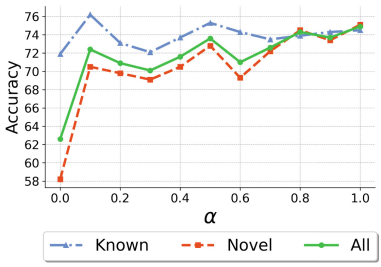


Fig. 7. Effect of smoothing hyperparameter constant. A higher smoothing hyperparameter strengthens the focus on negative samples within the cluster; this enhances performance on novel categories while decreasing it on known categories.

6 Conclusion

This work presents self-expertise in identifying and categorizing known and previously unknown categories, focusing on fine-grained distinctions. We introduce a method that utilizes hierarchical structures to effectively bridge the gap between labeled data for known categories and unlabeled data for novel categories. This is achieved by generating hierarchical pseudo-labels, which guide both supervised and unsupervised learning phases of our self-expertise framework. The supervised phase is designed to incrementally increase the complexity of differentiation tasks, thereby accelerating the training process and enhancing the formation of distinct clusters for unknown categories. This strategy improves the model's ability to generalize to novel categories. In the unsupervised phase, we integrate a label-smoothing hyperparameter, compelling the model to concentrate on negative samples within a localized context and to make finer distinctions. This approach enhances the model's fine-grained categorization capabilities. Overall,

our work demonstrates the effectiveness of self-expertise in handling unknown and fine-grained categorization tasks. In the Appendix section titled 'Discussions,' we outline the limitations of our work and propose directions for future research.

Acknowledgments. This work is part of the project Real-Time Video Surveillance Search with project number 18038, which is (partly) financed by the Dutch Research Council (NWO) domain Applied and Engineering/ Sciences (TTW).

References

1. An, W., Tian, F., Zheng, Q., Ding, W., Wang, Q., Chen, P.: Generalized category discovery with decoupled prototypical network. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37 (2023)
2. Arthur, D., Vassilvitskii, S.: K-means++ the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035 (2007)
3. Banerjee, A., Kallooriyakath, L.S., Biswas, S.: Amend: adaptive margin and expanded neighborhood for efficient generalized category discovery. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2101–2110 (2024)
4. Boult, T.E., Cruz, S., Dhamija, A.R., Gunther, M., Henrydoss, J., Scheirer, W.J.: Learning and the unknown: surveying steps toward open world recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 9801–9807 (2019)
5. Cao, K., Brbic, M., Leskovec, J.: Open-world semi-supervised learning. In: Proceedings of the International Conference on Learning Representations (2022)
6. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: Proceedings of the European Conference on Computer Vision, pp. 132–149 (2018)
7. Caron, M., et al.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9650–9660 (2021)
8. Chapelle, O., Scholkopf, B., Zien, A.: Semi-supervised learning. IEEE Trans. Neural Netw. **20**(3), 542–542 (2009)
9. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
10. Chiaroni, F., Dolz, J., Masud, Z.I., Mitiche, A., Ben Ayed, I.: Parametric information maximization for generalized category discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1729–1739 (2023)
11. Choi, S., Kang, D., Cho, M.: Contrastive mean-shift learning for generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024)
12. Cole, E., Yang, X., Wilber, K., Mac Aodha, O., Belongie, S.: When does contrastive visual representation learning work? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14755–14764 (2022)
13. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)

14. Du, R., Chang, D., Liang, K., Hospedales, T., Song, Y.Z., Ma, Z.: On-the-fly category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11691–11700 (2023)
15. Fei, Y., Zhao, Z., Yang, S., Zhao, B.: Xcon: learning with experts for fine-grained category discovery. In: British Machine Vision Conference (2022)
16. Gao, F., Zhong, W., Cao, Z., Peng, X., Li, Z.: Opengcd: assisting open world recognition with generalized category discovery. arXiv preprint arXiv:2308.06926 (2023)
17. Guo, Y., et al.: Hcsc: hierarchical contrastive selective coding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9706–9715 (2022)
18. Han, K., Rebuffi, S.A., Ehrhardt, S., Vedaldi, A., Zisserman, A.: Automatically discovering and learning new visual categories with ranking statistics. In: Proceedings of the International Conference on Learning Representations (2020)
19. Hao, S., Han, K., Wong, K.Y.K.: Cipr: an efficient framework with cross-instance positive relations for generalized category discovery. arXiv preprint arXiv:2304.06928 (2023)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
21. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
22. Huynh, T., Kornblith, S., Walter, M.R., Maire, M., Khademi, M.: Boosting contrastive self-supervised learning with false negative cancellation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2785–2795 (2022)
23. Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F.: A survey on contrastive self-supervised learning. Technologies **9**(1), 2 (2020)
24. Khorasgani, S.H., Chen, Y., Shkurti, F.: Slic: self-supervised learning with iterative clustering for human action videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16091–16101 (2022)
25. Khosla, P., et al.: Supervised contrastive learning. Adv. Neural Inf. Process. Syst. **33**, 18661–18673 (2020)
26. Kim, H., Suh, S., Kim, D., Jeong, D., Cho, H., Kim, J.: Proxy anchor-based unsupervised learning for continuous generalized category discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 16688–16697 (2023)
27. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3d object representations for fine-grained categorization. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 554–561 (2013)
28. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images. Technical Report. University of Toronto, Toronto, Ontario (2009). https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf
29. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **25** (2012)
30. Li, J., Zhou, P., Xiong, C., Hoi, S.: Prototypical contrastive learning of unsupervised representations. In: International Conference on Learning Representations (2021)
31. Liu, X., et al.: Self-supervised learning: generative or contrastive. IEEE Trans. Knowl. Data Eng. **35**(1), 857–876 (2021)

32. Mahdavi, A., Carvalho, M.: A survey on open set recognition. In: 2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pp. 37–44 (2021)
33. Maji, S., Rahtu, E., Kannala, J., Blaschko, M., Vedaldi, A.: Fine-grained visual classification of aircraft. arXiv preprint arXiv:1306.5151 (2013)
34. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 69–84. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_5
35. Oliver, A., Odena, A., Raffel, C.A., Cubuk, E.D., Goodfellow, I.: Realistic evaluation of deep semi-supervised learning algorithms. Adv. Neural Inf. Process. Syst. **31** (2018)
36. Oquab, M., et al.: DINOv2: learning robust visual features without supervision. Trans. Mach. Learn. Res. (2024)
37. Otholt, J., Meinel, C., Yang, H.: Guided cluster aggregation: a hierarchical approach to generalized category discovery. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2618–2627 (2024)
38. Ouali, Y., Hudelot, C., Tami, M.: An overview of deep semi-supervised learning. arXiv preprint arXiv:2006.05278 (2020)
39. Parkhi, O.M., Vedaldi, A., Zisserman, A., Jawahar, C.: Cats and dogs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3498–3505 (2012)
40. Pu, N., Zhong, Z., Sebe, N.: Dynamic conceptional contrastive learning for generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
41. Rastegar, S., Doughty, H., Snoek, C.G.M.: Learn to categorize or categorize to learn? self-coding for generalized category discovery. Adv. Neural Inf. Process. Syst. (2023)
42. Rebuffi, S.A., Ehrhardt, S., Han, K., Vedaldi, A., Zisserman, A.: Semi-supervised learning with scarce annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 762–763 (2020)
43. Salehi, M., Mirzaei, H., Hendrycks, D., Li, Y., Rohban, M.H., Sabokrou, M.: A unified survey on anomaly, novelty, open-set, and out of-distribution detection: Solutions and future challenges. Trans. Mach. Learn. Res. (2022)
44. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations (2015)
45. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
46. Tan, K.C., Liu, Y., Ambrose, B., Tulig, M., Belongie, S.: The herbarium challenge 2019 dataset. arXiv preprint arXiv:1906.05372 (2019)
47. Tan, Z., Zhang, C., Yang, X., Sun, J., Huang, K.: Revisiting mutual information maximization for generalized category discovery. arXiv preprint arXiv:2405.20711 (2024)
48. Troisemaine, C., Lemaire, V., Gosselin, S., Reiffers-Masson, A., Flocon-Cholet, J., Vaton, S.: Novel class discovery: an introduction and key concepts. arXiv preprint arXiv:2302.12028 (2023)
49. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022)

50. Vaze, S., Vedaldi, A., Zisserman, A.: No representation rules them all in category discovery. Adv. Neural Inf. Process. Syst. **37** (2023)
51. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The Caltech-UCSD birds-200-2011 dataset (2011)
52. Wang, H., Vaze, S., Han, K.: SPTNet: an efficient alternative framework for generalized category discovery with spatial prompt tuning. In: Proceedings of the International Conference on Learning Representations (2024)
53. Wang, Y., Wang, Y., Wu, Y., Zhao, B., Qian, X.: Beyond known clusters: probe new prototypes for efficient generalized class discovery. arXiv preprint arXiv:2404.08995 (2024)
54. Wang, Y., et al.: Discover and align taxonomic context priors for open-world semi-supervised learning. Adv. Neural Inf. Process. Syst. (2023)
55. Wen, X., Zhao, B., Qi, X.: Parametric classification for generalized category discovery: a baseline study. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 16590–16600 (2023)
56. Xiao, R., et al.: Targeted representation alignment for open-world semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 23072–23082 (2024)
57. Yang, M., Wang, L., Deng, C., Zhang, H.: Bootstrap your own prior: towards distribution-agnostic novel class discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3459–3468 (2023)
58. Yang, X., Song, Z., King, I., Xu, Z.: A survey on deep semi-supervised learning. IEEE Trans. Knowl. Data Eng. **35**, 8934–8954 (2022)
59. Zhai, X., Oliver, A., Kolesnikov, A., Beyer, L.: S4l: self-supervised semi-supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1476–1485 (2019)
60. Zhang, L., Qi, L., Yang, X., Qiao, H., Yang, M.H., Liu, Z.: Automatically discovering novel visual categories with self-supervised prototype learning. arXiv preprint arXiv:2208.00979 (2022)
61. Zhang, S., Khan, S., Shen, Z., Naseer, M., Chen, G., Khan, F.: 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 (2023)
62. Zhang, S., Xu, R., Xiong, C., Ramaiah, C.: Use all the labels: a hierarchical multi-label contrastive learning framework. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16660–16669 (2022)
63. Zhao, B., Wen, X., Han, K.: Learning semi-supervised gaussian mixture models for generalized category discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2023)
64. Zhu, F., Ma, S., Cheng, Z., Zhang, X.Y., Zhang, Z., Liu, C.L.: Open-world machine learning: a review and new outlooks. arXiv preprint arXiv:2403.01759 (2024)