Virtual Category-Guided Continual Generalized Category Discovery

Jiahui Xiong[1], Qiuxia Lai[2*], and Hongsong Wang[1,3*]

[1]School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[2]State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China
[3]Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China
{hongsongwang, jiahuixiong}@seu.edu.cn, qxlai@cuc.edu.cn

Abstract. Continual Generalized Category Discovery (C-GCD) aims to incrementally identify novel categories from sequential unlabeled data while preserving recognition of known classes, which is an essential capability for open-world visual learning. A major bottleneck lies in ambiguous unlabeled samples that cannot be confidently assigned to known classes nor reliably grouped as novel ones, making pseudo-labeling brittle and often biasing learning toward familiar categories. In this work, we introduce Virtual Category-Guided Continual Generalized Category Discovery by adapting Virtual Category Learning (VCL) to the continual setting. Our method identifies uncertain samples and assigns them to temporary virtual categories, enabling safe and informative learning from unlabeled streams without injecting noisy labels, while improving unlabeled data utilization and mitigating prediction bias. To further stabilize discovery across sessions and enhance class separation, we augment VCL with Expanded Neighborhood Contrastive Learning (ENCL), which exploits extended neighborhood relations and an adaptive margin to learn more discriminative and well-separated representations for both old and emerging classes. Extensive experiments on CIFAR-100, Tiny ImageNet, and ImageNet-100 demonstrate that our approach consistently outperforms state-of-the-art methods, establishing a scalable and effective solution for C-GCD. Code is on: https://github.com/Mrxjh105/VC-CGCD

Keywords: Continual Generalized Category Discovery · Continual Learning · Open-World Recognition

1 Introduction

Modern visual recognition has made striking progress in closed-world settings where the label space is fixed and exhaustively defined [8, 22]. However, real deployments, such as home robots, autonomous driving, and large-scale monitoring,

* Corresponding authors.

operate in open and evolving environments, where novel categories appear continually and must be incorporated without retraining from scratch [9, 24]. This requirement exposes a fundamental challenge for long-lived learners: they must expand their category set over time while retaining competence on previously learned classes, i.e., the stability–plasticity dilemma. When naively fine-tuned on new data, deep networks are prone to catastrophic forgetting [14, 16, 20], making continual open-world recognition particularly unstable.

To address emerging categories, *Novel Category Discovery (NCD)* [7] studies how to discover unseen classes from unlabeled data given labeled known classes, typically under the assumption that the unlabeled pool contains only novel categories. This assumption rarely holds in realistic streams where old and new categories naturally co-exist. *Generalized Category Discovery (GCD)* [23] relaxes the setting by allowing unlabeled data to mix known and novel classes, requiring simultaneous recognition of known instances and clustering of novel ones. However, most GCD methods remain batch-oriented and assume access to the full unlabeled pool, making them ill-suited to continuous data arrival. *Continual Generalized Category Discovery (C-GCD)* [30] pushes category discovery into a truly continual regime: after an offline phase trained on labeled known classes, the model receives a sequence of unlabeled sessions that contain both previously seen and newly emerging categories without retaining past data (e.g., due to storage or privacy constraints). Recent progress explores meta-learning [9], grow-and-merge model updates [30], and debiasing strategies [15], yet practical C-GCD remains challenging.

A central challenge in C-GCD is learning from unlabeled streams under pervasive ambiguity. Online sessions provide no labels, so learning must rely on clustering, neighborhood consistency, or pseudo-labels inferred from the current model. However, unlabeled data are heterogeneous: besides clear known-class instances and discoverable novel-class samples, there exist many ambiguous examples near decision boundaries or in visually overlapping regions. Existing pipelines often force such uncertain samples into either "known" or "novel" groups via hard pseudo-labels or aggressive clustering. In a continual setting, such early mis-assignments are particularly damaging, as they introduce persistent label noise, amplify confirmation bias, and can compound across sessions when earlier errors cannot be corrected by revisiting data. As a result, the learner may underutilize precisely the hard unlabeled samples that are most informative for carving new decision regions.

This observation motivates the main idea of our paper: learning with virtual categories. Rather than treating every unlabeled sample as if it must belong to one of the current semantic categories, we explicitly provide the model with a mechanism to represent unknown content in a controlled manner. Concretely, we adapt *Virtual Category Learning (VCL)* originally proposed in semi-supervised learning [3] to the continual generalized discovery setting. VCL assigns uncertain samples to temporary virtual categories, which serve as a buffer that allows the learner to absorb informative structure from ambiguous samples without injecting hard label noise into the known/novel partition. As representations become

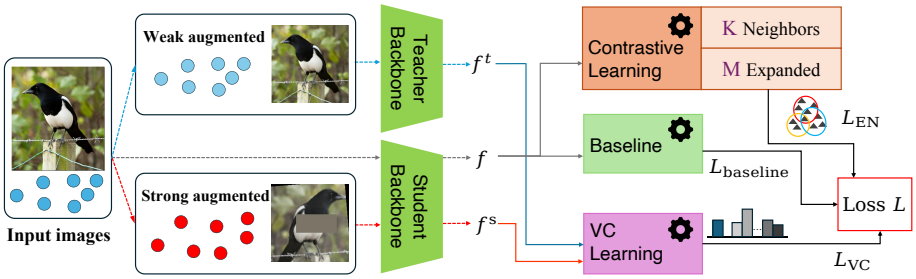

Fig. 1: Pipeline of the proposed framework. Input images are encoded into shared features, where Virtual Category Learning (VCL) serves as the core mechanism to safely incorporate confusing unlabeled samples through virtual categories, stabilizing learning under evolving category boundaries. Guided by this process, ENCL further improves feature separation, while a baseline objective preserves fundamental recognition performance during continual updates. See §3.2 and §3.3 for details.

more reliable over training, samples are less likely to remain associated with virtual categories and can gradually contribute to stable semantic assignments, improving the effectiveness of learning from challenging unlabeled data throughout the continual process. To further reduce class confusion and promote separation in the evolving embedding space, inspired by [2], we introduce Expanded Neighborhood Contrastive Learning (ENCL), which leverages neighbors-of-neighbors and an adaptive margin to encourage more discriminative and well-separated representations for both old and emerging classes. Together, VCL and ENCL form a coherent C-GCD framework that learns safely from uncertain unlabeled samples without over-committing to noisy pseudo-labels, while building more discriminative and well-separated representations that mitigate confusion between old and new categories, as summarized in Fig. 1.

Our main contributions are threefold:

- We introduce the **Virtual Category-Guided C-GCD framework** that address ambiguity-safe utilization of unlabeled streams and robust feature separation under representation drift.
- We adapt *Virtual Category Learning (VCL)* to C-GCD by assigning uncertain unlabeled samples to temporary virtual categories, which improves the utilization of unlabeled data and mitigates error amplification caused by premature hard assignments.
- We introduce *Expanded Neighborhood Contrastive Learning (ENCL)* with extended neighborhood sampling and an adaptive margin to enhance feature discriminability between learned and emerging categories.

Extensive experiments on CIFAR-100 [11], Tiny-ImageNet [12], and ImageNet-100 [5] demonstrate that our framework consistently improves accuracy, mitigates forgetting, and enhances novel class discovery over strong C-GCD baselines, validating its effectiveness and practical applicability.

2 Related Works

Category Discovery. *Novel Category Discovery (NCD)* aims to discover previously unseen categories by transferring knowledge from labeled known classes to an unlabeled set that is assumed to contain only novel instances. Early NCD pipelines typically learn a representation on labeled data and then apply unsupervised clustering on unlabeled data for novel class grouping [7]. Subsequent studies improve discovery by strengthening representation learning, often leveraging neighborhood- or pseudo-label-based objectives to induce cluster-friendly embeddings under limited supervision [10, 25, 28, 29, 31]. *Generalized Category Discovery (GCD)* relaxes the key NCD assumption by allowing the unlabeled pool to be a mixture of known and novel classes, requiring the model to recognize known instances while clustering the remaining data into novel categories [23]. This formulation better matches realistic deployments, but most existing GCD methods remain *static*, assuming access to the full unlabeled pool and a one-shot training procedure rather than continual updates [1, 4, 13, 18].

Continual Generalized Category Discovery. *Continual learning (CL)* studies how to learn from sequential data streams without catastrophic forgetting, commonly under supervised task- or class-incremental settings via rehearsal or regularization [14, 19]. Bringing category discovery into the continual regime yields *Continual Generalized Category Discovery (C-GCD)*, where a model is initialized with labeled known classes and then must learn from a sequence of unlabeled sessions containing both previously seen and emerging categories, typically without storing past data. This setting couples the stability-plasticity tension of CL with the mixed-label-space uncertainty of GCD. Representative C-GCD methods explore different mechanisms to balance discovery and retention. Grow-and-merge style frameworks expand model capacity or representation diversity and then reconcile conflicts to preserve previously learned knowledge [30]. Meta-learning-based approaches simulate incremental processes to improve continual adaptation under the mixed known/novel regime [27]. Debiasing-oriented frameworks aim to mitigate prediction bias toward known classes and improve robustness under unlabeled continual updates [15]. Despite these advances, C-GCD remains challenging due to ambiguity in fully unlabeled sessions that hinders the reliable utilization of unlabeled data and representation drift that weakens feature discriminability between learned and emerging categories.

Ambiguity and Learning from Unlabeled Data. A central challenge in C-GCD is learning from heterogeneous unlabeled streams that include confident known instances, truly novel instances, and ambiguous boundary samples. Many discovery pipelines rely on pseudo-labeling [26, 28] or clustering-based assignments [7, 23], which can be fragile under ambiguity and may amplify confirmation bias when early errors are reinforced across sessions without the ability to revisit past data. Addressing this requires mechanisms that can utilize ambiguous samples without forcing premature hard commitments. Recent semi-/self-supervised learning studies propose to handle uncertain samples via auxiliary constructs that reduce label noise and stabilize training. In particular, *virtual category* mechanisms introduce temporary placeholders for uncertain instances to

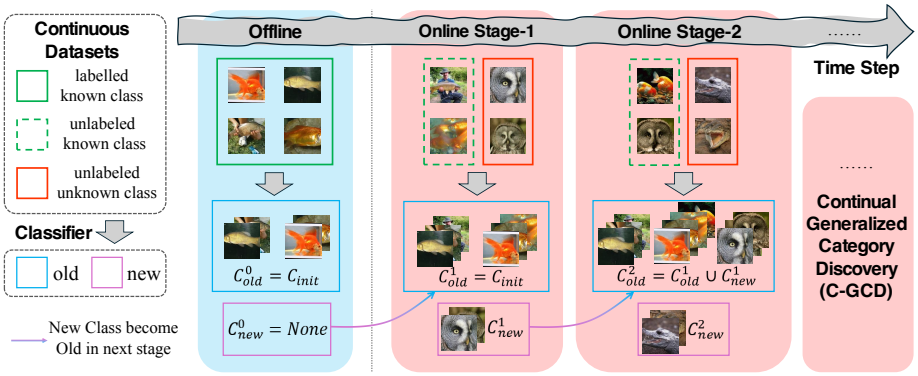

Fig. 2: C-GCD task setting. The model is initialized with labeled classes C_{init} and then receives sequential unlabeled sessions D^t containing both old classes C^i_{old} and novel classes C^i_{new}. The goal is to continually recognize known categories and discover new ones, progressively expanding the known label space across sessions without storing past data. See §3.1 for details.

absorb informative structure while avoiding incorrect semantic commitments [3]. Motivated by this challenge, our method introduces Virtual Category Learning (VCL) to handle uncertain unlabeled samples via temporary virtual categories, avoiding premature semantic commitments under continual drift.

3 Methodology

3.1 Preliminaries

Problem Formulation. As illustrated in Fig. 2, C-GCD consists of two phases: an **offline initialization stage** and an **online continual discovery stage**. **(i) Offline stage.** At the initial session S_0, we train the model on a labeled dataset $D^0_{\text{train}} = \{(\boldsymbol{x}^l_i, y_i)\}^{N^0}_{i=1}$. The corresponding label set is denoted by $C^0_{\text{old}} = C_{\text{init}}$. Since this is the initialization phase, no novel unlabeled classes are introduced, i.e., $C^0_{\text{new}} = \varnothing$. **(ii) Online stage.** For each subsequent session S_t, $t \in \{1, 2, \ldots, T\}$, an unlabeled dataset $D^t_{\text{train}} = \{\boldsymbol{x}^u_i\}^{N^t}_{i=1}$ arrives sequentially. Let C^t denote the (unknown) category set contained in D^t_{train}, which is composed of previously encountered classes and newly emerging classes: $C^t = C^t_{\text{old}} \cup C^t_{\text{new}}$. We denote the total number of classes in session t by $K^t = |C^t|$, with $K^t_{\text{old}} = |C^t_{\text{old}}|$ and $K^t_{\text{new}} = |C^t_{\text{new}}|$. Accordingly, $K^t = K^t_{\text{old}} + K^t_{\text{new}}$.

Framework Overview. As illustrated in Fig. 1, our method adopts a multi-branch framework for C-GCD and is instantiated on a baseline. For each input image, we generate weakly and strongly augmented views, which are fed into backbone encoders (e.g., ViT or ResNet) to extract feature representations. These features are then processed by two cooperative modules that address complementary challenges in C-GCD. First, the *Virtual Category Learning*

(VCL) module (§3.2) addresses ambiguous unlabeled samples arising from evolving category boundaries. Rather than forcing early semantic assignments, VCL jointly exploits teacher-student features to identify confusing samples and associate them with temporary virtual categories. This allows ambiguous samples to contribute useful training signals while reducing noisy supervision and stabilizing representation learning across sessions. Second, the *Expanded Neighborhood Contrastive Learning (ENCL)* module (§3.3) improves feature discriminability by leveraging local neighborhoods with expanded semantic relations, thus improving the separation between previously learned and emerging categories. Finally, the training objective (§3.4) integrates the baseline loss with the proposed VCL and ENCL terms in a unified end-to-end optimization framework. Overall, the proposed framework enables the reliable utilization of ambiguous unlabeled data while maintaining robust feature discrimination under continual drift.

3.2 Virtual Category Learning

In C-GCD, unlabeled samples contribute to optimization with highly uneven reliability. While confident instances provide relatively stable semantic guidance, a substantial portion of data remains ambiguous due to evolving category boundaries and incremental distribution shifts. Directly imposing hard semantic assignments on such samples often introduces unstable gradients, causing premature feature attraction toward incorrect categories and gradually distorting the embedding space across sessions. Instead of excluding ambiguous samples or committing them to fixed pseudo labels, we allow uncertain instances to participate in training through *virtual semantic associations*, where their influence is regulated by uncertainty. In this way, ambiguous samples continue to shape representation learning without forcing premature decisions, leading to smoother feature evolution and more stable optimization dynamics throughout continual discovery. Beyond improving sample utilization, this treatment also influences the emergence of neighborhood structure, since representations can organize progressively instead of collapsing toward early predictions.

Confusing Sample. For clarity, we describe VCL from the perspective of a single sample, while the same procedure applies to mini-batch training. As illustrated in Fig. 3, weakly and strongly augmented views are encoded by the teacher encoder T and the student encoder S, producing feature representations for subsequent processing. The teacher classifier computes category probabilities from the projection between the feature vector $(\boldsymbol{f}^t)^\top$ and classifier weights W^t.

Conventional approaches typically use the highest-probability category as a pseudo label for supervision. However, under continual updates, prediction confidence may not reliably reflect semantic correctness, especially when new categories gradually emerge. Treating such predictions as fixed supervision may bias optimization toward transient decision boundaries. VCL instead treats uncertainty as an indicator of incomplete semantic separation rather than noise. After obtaining prediction scores, a potential-category construction step is performed to determine whether the sample admits multiple plausible semantics. If more than one candidate category exists, the sample is identified as a *confusing*

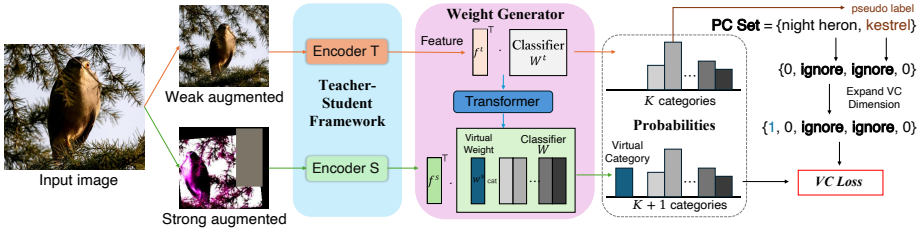

Fig. 3: Illustration of Virtual Category Learning when handling a confusing sample. Weakly and strongly augmented views are encoded by teacher and student networks to produce prediction scores, from which a potential category set is constructed using prediction competition and teacher–student consistency. Instead of assigning a hard pseudo label, the sample is guided toward a temporary virtual category, allowing it to participate in training without enforcing premature semantic commitment. See §3.2.

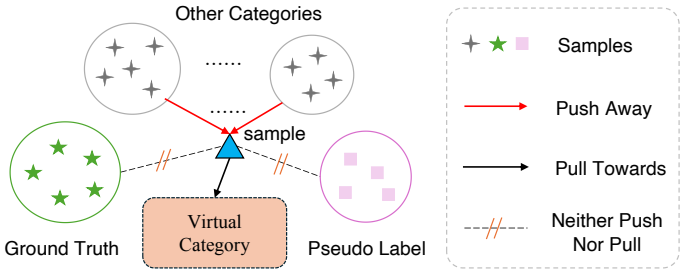

Fig. 4: Feature-space illustration of Virtual Category Learning. The blue triangle denotes a training sample located near multiple candidate categories, while colored clusters represent different classes and arrows indicate optimization directions. Instead of being strongly attracted to any single category center, the sample is guided toward a virtual category (the dashed rectangular), enabling gradual feature adjustment while avoiding incorrect semantic commitment. See §3.2 for details.

sample. Rather than removing these samples, VCL keeps them active during training while avoiding directional bias toward any single uncertain category.

From a feature-space perspective, as shown in Fig. 4, confusing samples should be repelled from clearly incorrect categories, yet aggressively pulling them toward any single candidate risks reinforcing errors. Introducing a virtual category provides a neutral optimization direction, allowing the representation to evolve safely while preserving flexibility for later semantic alignment. In continual learning, such treatment allows ambiguous samples to gradually influence feature organization instead of injecting irreversible mistakes early in training, which in turn reshapes local neighborhood structures over time.

Potential Category Set. To realize this idea, we construct a *Potential Category (PC) set*, defined as the collection of candidate categories that remain semantically plausible for an uncertain sample under the current model predictions. Instead of assigning a single pseudo label, the PC set maintains multi-

ple possible category associations, allowing ambiguous samples to contribute to training without enforcing premature decisions. Unlike conventional confidence-threshold filtering, which attempts to eliminate unreliable predictions, the PC formulation preserves ambiguity as informative structural evidence during representation evolution. Maintaining a set of candidates therefore allows optimization to remain flexible while preventing premature semantic commitment.

In practice, the PC set is constructed by jointly considering two complementary uncertainty cues: prediction competition and teacher–student prediction consistency. The first cue measures whether the most likely categories are difficult to distinguish under the current prediction. Specifically, we compute the relative gap between the top predicted probabilities, $p_{1\text{st}}$ and $p_{2\text{nd}}$:

$$D_{\text{top2}} = \frac{p_{1\text{st}} - p_{2\text{nd}}}{p_{2\text{nd}}}, \tag{1}$$

where a smaller value indicates stronger competition between the top candidate categories. However, D_{top2} is not used as a standalone criterion, since prediction uncertainty may also arise from unstable model responses rather than only from a small top-2 gap. Therefore, we further compare the predicted labels y'_t and y'_s from the teacher and student networks. A discrepancy between them indicates prediction inconsistency and provides an additional signal that the sample lies in an uncertain region of the evolving feature space. The categories suggested by these two cues are combined to form the PC set, and samples with multiple plausible candidates are treated as confusing samples. In this way, low-confidence predictions are not filtered or judged solely by the top-2 score; they can still be identified as confusing when teacher–student predictions are inconsistent.

By integrating these complementary cues, the PC formulation captures uncertainty arising from both local prediction ambiguity and temporal representation variation. As training proceeds, samples within the PC set maintain soft associations with multiple candidate categories, allowing semantic structure to consolidate gradually. This set-based representation thus provides a transitional state between unknown and confidently assigned categories, supporting stable feature organization while category boundaries continue to evolve over sessions.

VC Loss. VCL is applied session-wise in the online stage (i.e., at each S_t to mini-batches sampled from D^t_{train}). We first present the loss from the perspective of a single sample in the current session. For notational simplicity, we omit the superscript t in the following equations unless needed.

Given the VC formulation above, where the PC set is instantiated for C-GCD using prediction competition and teacher–student consistency over old and emerging categories, we introduce a virtual category to absorb ambiguous supervision. Specifically, following [3], we extend the classifier weight matrix W with a virtual weight vector $\boldsymbol{w}^v$, which is generated by a self-attention transformer layer. The logits are computed as:

$$\boldsymbol{f}^\top \cdot \left[\boldsymbol{w}^v, \boldsymbol{w}^0, \ldots, \boldsymbol{w}^{K-1}\right] = \left[l^v, l^0, \ldots, l^{K-1}\right], \tag{2}$$

where the l represents the logit vector, and the K denotes the number of predicted categories. For reference, the conventional cross-entropy with pseudo la-

bels obtained through clustering is defined as:

$$\ell_{\mathrm{CE}} = \log(\sum\nolimits_{i=0}^{K} e^{l^i - l^{y'}}). \tag{3}$$

With the virtual category, the objective becomes:

$$\ell_{\mathrm{VC}} = \log(\sum\nolimits_{i=0,\ i\notin \mathrm{PC}}^{K+2} e^{l^i - l^{v}}), \tag{4}$$

where logits corresponding to categories within the uncertain PC set are excluded from direct supervision. This prevents ambiguous candidates from imposing conflicting gradients while still allowing the sample to influence representation learning through the virtual category. In practice, at each online session S_t, the VCL objective is computed by averaging Eq. (4) over confusing samples in the current mini-batch $\mathcal{B}$:

$$\mathcal{L}_{\mathrm{VC}} = \frac{1}{\sum_{\boldsymbol{x}_j \in \mathcal{B}} \delta_j} \sum\nolimits_{\boldsymbol{x}_j \in \mathcal{B}} \delta_j \, \ell_{\mathrm{VC}}(\boldsymbol{x}_j), \tag{5}$$

where $\delta_j \in \{0, 1\}$ indicates whether $\boldsymbol{x}_j$ is identified as a confusing sample. As a result, this optimization encourages consistent feature organization without enforcing premature semantic commitments, which provides a more reliable basis for modeling neighborhood relations as categories continue to evolve.

3.3 Expanded Neighborhood Contrastive Learning

VCL enables ambiguity-safe participation of uncertain samples, improving unlabeled-data utilization but also yielding a feature space that evolves continuously across sessions. To strengthen feature discriminability between learned and emerging categories, we introduce Expanded Neighborhood Contrastive Learning (ENCL), which extends [2] from static GCD to a continual setting to regularize the evolving representation space. Given an image $\boldsymbol{x}_i$, traditional neighborhood contrastive learning defines positive samples using nearest neighbors $N(\boldsymbol{x}_i)$, and the loss can be written as:

$$\ell_N = -\frac{1}{n} \sum\nolimits_{\boldsymbol{q} \in N(\boldsymbol{x}_i)} \log \frac{\exp(\boldsymbol{f}_i \cdot \boldsymbol{f}_q / \tau)}{\sum_{j=0,\ j\neq i}^{n-1} \exp(\boldsymbol{f}_i \cdot \boldsymbol{f}_j / \tau)}, \tag{6}$$

where $\boldsymbol{f}_i$ denotes the l_2-normalized feature representation of $\boldsymbol{x}_i$, τ is the temperature parameter, and $n = |N(\boldsymbol{x}_i)|$.

However, when representations evolve under ambiguity-aware supervision, relying solely on immediate neighbors may overlook semantically related samples that emerge through gradual alignment. We therefore expand the neighborhood by incorporating neighbors-of-neighbors. Specifically, for each neighbor $\boldsymbol{q} \in N(\boldsymbol{x}_i)$, we retrieve its m-nearest neighbors, and construct the expanded neighborhood as:

$$E_M(\boldsymbol{x}_i) = N(\boldsymbol{q}); \quad \forall \boldsymbol{q} \in N(\boldsymbol{x}_i). \tag{7}$$

This expanded set enlarges the pool of positive samples while preserving locality, enabling the model to capture indirect semantic relations that stabilize

contrastive learning across sessions. Following Eq. (6), the expanded neighborhood loss is defined as:

$$\ell_{EN} = -\frac{1}{m} \sum\nolimits_{\boldsymbol{q} \in E_M(\boldsymbol{x}_i)} \log \frac{\exp(\boldsymbol{f}_i \cdot \boldsymbol{f}_q / \tau)}{\sum_{j=0, j \neq i}^{m-1} \exp(\boldsymbol{f}_i \cdot \boldsymbol{f}_j / \tau)}, \tag{8}$$

where $m = |E_M(\boldsymbol{x}_i)|$ denotes the number of expanded neighbors. Samples appearing multiple times during expansion naturally receive stronger emphasis, reflecting their closer semantic proximity. Combining the original and expanded neighborhoods with a weighting factor β, the contrastive objective over a mini-batch $\mathcal{B}$ becomes:

$$\mathcal{L}_{\mathrm{EN}} = \sum\nolimits_{i \in \mathcal{B}} (\ell_N + \beta \ell_{EN}). \tag{9}$$

By enlarging neighborhood relations under continually evolving representations, this formulation maintains stable feature structures while allowing gradual separation between previously learned and newly emerging categories.

3.4 Training Objectives

We train the proposed framework with a unified objective that integrates two task-motivated components: Virtual Category Learning (VCL) for ambiguity-safe utilization of unlabeled samples and Expanded Neighborhood Contrastive Learning (ENCL) for robust feature discrimination under continual drift. In implementation, the framework is instantiated within a C-GCD baseline [15,27], whose objective is denoted by $\mathcal{L}_{\mathrm{baseline}}$. For brevity, we do not repeat the full decomposition of $\mathcal{L}_{\mathrm{baseline}}$ here and refer readers to [15]. The overall training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{VC}} + \lambda_1 \mathcal{L}_{\mathrm{EN}} + \lambda_2 \mathcal{L}_{\mathrm{baseline}}, \tag{10}$$

where the coefficient of $\mathcal{L}_{\mathrm{VC}}$ is fixed to 1 as the reference scale, while λ_1 and λ_2 control the relative contributions of the ENCL term and the baseline objective, respectively. In this formulation, $\mathcal{L}_{\mathrm{VC}}$ regulates how ambiguous samples participate in optimization, $\mathcal{L}_{\mathrm{EN}}$ regularizes neighborhood-based feature geometry to promote separation between previously learned and newly emerging categories, and $\mathcal{L}_{\mathrm{baseline}}$ maintains the basic training behavior of the underlying C-GCD pipeline. These terms are jointly optimized in an end-to-end manner, yielding a unified training framework that improves both unlabeled-data utilization and representation discrimination in C-GCD.

4 Experiments

In this section, we evaluate the proposed framework on standard C-GCD benchmarks and examine its effectiveness from three perspectives. We first describe the experimental setups In §4.1, including datasets, evaluation metrics, and implementation details. We then report the main results in §4.2 to compare our method with state-of-the-art methods under the C-GCD setting. Finally, in §4.3, we conduct ablation studies and analysis to isolate the contributions of VCL and ENCL and to further investigate the behavior of the proposed framework.

4.1 Experimental Setups

Datasets. We evaluate our method on three widely used benchmarks for category discovery: CIFAR100 (C100) [11], ImageNet-100 (IN100) [5] and Tiny-ImageNet (Tiny) [12]. Following established C-GCD protocols, each dataset is divided into an offline initialization stage and multiple online continual sessions.

In Stage-0 (offline), the model is trained using labeled data from 50% of the categories, where each selected class contributes 80% of its training samples. These classes form the initial known set C_{init}.

During online stages, the remaining 50% categories are introduced as novel classes through sequential unlabeled sessions. Each session contains a mixture of previously encountered classes and newly emerging ones. For previously observed categories, a portion of samples naturally reappears in later sessions according to the standard benchmark construction, without storing past data explicitly, thereby preserving the non-rehearsal assumption of C-GCD.

After training on the unlabeled training split D^t_{train} of each session, evaluation is conducted on the corresponding test split containing both old classes C^t_{old} and newly introduced classes C^t_{new}.

Evaluation Metrics. We employ **accuracy** (**ACC**) as the primary evaluation metric, reporting performance on: (i) newly introduced classes (*New*), (ii) previously learned classes (*Old*), and (iii) all currently known categories (*All*).

To further analyze continual behavior, we additionally report the **forgetting** metric M_f and the **discovery** metric M_d introduced by [30], which have been widely used in the continual discovery literature. Specifically,

$$M_f = ACC^0_{\text{known}} - \min_{1 \le t \le T} \{ACC^t_{\text{known}}\}, \tag{11}$$

which measures the maximum degradation of previously learned categories across sessions, and

$$M_d = \frac{1}{T} \sum_{1 \le t \le T} ACC^t_{\text{noval}}, \tag{12}$$

which reflects the sustained discovery capability. Together, these metrics characterize the trade-off between knowledge retention and continual discovery.

Implementation Details. We adopt a ViT-B/16 [6] backbone pre-trained with DINO initialization. The loss weighting coefficients are set to $\{\beta, \lambda_1, \lambda_2\}$ as $\{0.1, 0.8, 1.0\}$. The offline stage is trained for 100 epochs, while each online session is optimized for 30 epochs using a batch size of 128 and a learning rate of 0.01. All experiments are conducted on NVIDIA GeForce RTX 4090D GPUs. Unless otherwise noted, optimization settings follow commonly adopted configurations in C-GCD frameworks to ensure comparable training dynamics across methods.

4.2 Main Results

We report the ACC results of different methods in Table 1. In general, our work achieves the best performance in terms of ACC in most cases, which demonstrates the effectiveness of our work for the C-GCD task. It is observed that, compared with previous methods: 1) KMeans [17] on pre-trained features, 2) GCD

Table 1: Main results of accuracy across offline and 5 online stages on Cifar100(C100), TinyImageNet(Tiny), ImageNet-100(IN100) compared with other works. See §4.2.

Datasets	Methods	Offline	Session-1			Session-2			Session-3			Session-4			Session-5		
		All	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
C100	KMeans	66.16	40.27	41.76	32.80	37.14	38.33	30.00	36.20	37.63	26.20	36.66	38.30	23.50	35.69	36.79	25.80
	VanillaGCD	90.82	72.32	78.50	41.40	67.04	72.50	34.30	57.99	62.26	28.10	56.60	59.55	33.00	51.36	53.70	30.30
	SimGCD	90.36	73.37	86.44	8.00	62.56	72.43	3.30	54.17	61.61	2.10	47.62	53.37	1.60	43.53	47.86	4.60
	FRoST	90.36	76.87	79.58	63.30	65.31	68.88	43.90	58.01	61.09	36.50	49.27	50.90	36.20	48.03	48.17	46.80
	GM	90.36	76.58	79.80	60.50	71.10	74.52	50.60	63.51	68.16	31.00	59.74	62.51	37.60	54.11	54.74	48.40
	MetaGCD	90.82	76.12	83.60	38.70	69.40	72.82	48.90	61.95	65.76	35.30	58.22	61.21	34.30	55.78	58.47	31.60
	Happy	90.36	80.40	**85.26**	56.10	74.13	78.27	49.30	68.23	70.86	49.80	62.26	63.75	**50.30**	59.99	60.96	**51.30**
	Ours	**90.64**	**83.22**	83.56	**81.50**	**76.33**	**85.26**	**54.00**	**70.55**	**79.74**	**55.23**	**64.99**	**81.94**	43.80	**61.28**	**62.70**	48.50
Tiny	KMeans	61.70	35.42	35.46	35.20	34.99	35.75	30.40	34.80	36.07	25.90	34.77	35.90	24.90	34.62	35.63	25.50
	VanillaGCD	84.20	55.93	58.92	41.00	54.96	58.58	33.20	52.82	55.74	32.40	48.81	51.46	27.60	45.94	48.06	26.90
	SimGCD	85.86	66.95	79.94	2.00	57.81	66.98	2.80	52.70	59.83	2.77	45.01	50.29	2.80	41.59	45.79	3.80
	FRoST	85.86	75.15	78.56	58.10	65.64	67.83	52.50	51.32	54.31	30.40	48.22	52.14	16.90	40.15	42.73	16.90
	GM	85.86	76.42	82.40	46.50	68.87	73.82	39.20	58.68	63.43	25.40	52.86	57.21	18.10	46.90	50.62	13.40
	MetaGCD	84.20	60.88	64.90	40.80	57.20	61.03	34.20	54.36	57.19	34.60	50.83	53.59	28.80	48.14	50.16	30.00
	Happy	**85.26**	76.67	**81.72**	51.40	69.50	**82.32**	37.45	61.78	**79.94**	31.50	56.17	**78.36**	28.43	51.99	**76.60**	27.38
	Ours	85.15	**77.15**	80.88	**58.50**	**70.54**	81.38	**43.45**	**63.52**	78.88	**37.93**	**57.56**	77.68	**32.40**	**53.25**	75.98	**30.52**
IN100	KMeans	85.56	54.90	57.04	44.20	54.73	56.37	44.90	54.67	56.66	40.80	54.63	56.25	41.70	53.92	56.18	33.60
	VanillaGCD	95.96	70.13	72.92	56.20	69.37	73.47	44.80	68.50	70.63	53.60	65.56	67.85	47.20	64.54	67.44	38.40
	SimGCD	96.20	79.67	91.68	19.60	70.23	78.83	18.60	61.90	67.43	23.20	56.67	60.92	22.60	52.90	56.40	21.40
	FRoST	96.20	87.50	92.96	60.20	79.63	83.37	57.20	76.78	77.00	75.20	66.18	68.65	46.40	63.82	66.40	40.60
	GM	96.20	89.53	95.04	62.00	82.34	86.93	54.80	77.97	79.17	69.60	72.80	74.65	58.00	71.08	71.76	65.00
	MetaGCD	95.96	75.27	78.20	60.60	73.79	75.93	54.90	69.35	72.20	49.40	67.22	70.10	44.20	66.68	69.31	43.00
	Happy	96.16	91.03	95.16	**70.40**	85.46	93.56	65.20	83.07	92.92	**66.67**	76.96	**93.24**	56.60	74.98	**92.40**	57.56
	Ours	**96.21**	**91.10**	**95.52**	69.00	**86.31**	**94.72**	**65.30**	**83.33**	**93.96**	65.60	**78.60**	92.24	**61.55**	**76.72**	91.88	**61.55**

methods: VanillaGCD [23], SimGCD [26], 3) novel C-GCD works: FRoST [21], GM [30], MetaGCD [27], Happy [15]. Our method maintains higher accuracy throughout the continual process while avoiding rapid degradation in later sessions. A clear improvement is observed on newly emerging categories. In contrast to methods that rely on hard pseudo-labeling or fixed neighborhood assumptions, ambiguity-aware supervision allows uncertain samples to contribute without enforcing premature semantic assignments. As a result, newly introduced classes are discovered more reliably, leading to improved New accuracy across most stages. For example, our approach improves the overall performance by 1.29% on C100, 1.26% on Tiny, and 1.74% on IN100, with particularly noticeable gains in new categories in later sessions.

We further evaluate forgetting and discovery capability in Table 2. The M_d values of our method all achieve the best results on both C100 and Tiny by $10.2 \sim 15.2\%$. Meanwhile, competitive M_f values show that the model preserves knowledge of old classes while incorporating new categories. These results suggest that preventing premature commitments helps reduce confirmation bias and stabilizes learning across sessions.

To examine long-term behavior, we extend the experiment to 10 stages, as shown in Table 3. The proposed method maintains a higher average accuracy over extended continual updates.

Table 2: Comparison of forgetting M_f and discovery M_d metrics which demonstrate improved stability and sustained discovery capability of the proposed method. See §4.2.

Methods	C100		Tiny	
	M_f (↓)	M_d (↑)	M_f (↓)	M_d (↑)
VanillaGCD [23]	37.12	33.42	36.14	32.22
FRoST [21]	42.19	45.34	43.13	34.96
MetaGCD [27]	32.35	37.76	34.04	33.68
Happy [15]	29.40	51.36	**8.66**	35.23
Ours	**27.94**	**56.61**	9.17	**40.56**

Table 3: Average "All" accuracy across 10 continual sessions, evaluating long-term performance under extended online updates. Our method shows improved robustness to continual distribution shifts and sustained category discovery ability. See §4.2.

Data	Methods	1	2	3	4	5	6	7	8	9	10	Avg
C100	Happy	85.20	81.63	77.94	74.57	69.16	66.25	62.34	59.72	54.49	52.89	68.42
	Ours	85.60	83.35	80.51	75.24	71.17	68.23	63.73	61.03	59.51	54.87	**70.32**
Tiny	Happy	79.35	75.63	71.28	66.83	62.53	60.05	55.09	52.01	49.79	45.77	61.83
	Ours	79.25	76.13	72.28	68.21	64.25	61.74	55.95	53.28	50.60	47.42	**62.91**

4.3 Ablation Study and Analysis

Component Analysis. To better understand the performance gains observed in the main results, we report ablation results in Table 4 by removing the VCL and ENCL on CIFAR-100 and Tiny-ImageNet. Removing either component consistently degrades performance, indicating that stable continual discovery relies on both ambiguity regulation and representation refinement during training. Without ambiguity-aware supervision, uncertain samples introduce noisier updates, while removing neighborhood expansion weakens the gradual organization of features across sessions. When both components are enabled, the model achieves the best performance across *All*, *Old*, and *New* categories, suggesting that the utilization of confusing samples and evolving neighborhood relations jointly supports more stable category formation over time.

Confusing Ratio. The proportion of confusing samples among all training data indicates how uncertainty is handled throughout training. Since confusing samples are identified by ambiguity rather than compact class membership, this ratio provides an indirect batch-level view of how VCL buffers and regulates these samples over time. As shown in Fig. 5, the ratio fluctuates at the beginning of Session 1 (the blue curve) and gradually stabilizes in later sessions. Similar but weaker variations appear afterward, suggesting that confusing samples play a larger role during early adaptation and become progressively regulated as representations mature. Although the ratio slightly decreases in later stages,

Table 4: Ablation study on the main modules, reporting the average accuracy (ACC) across all stages for C100 and Tiny. The results highlight the contributions of VCL and contrastive neighborhood refinement to stable C-GCD. See §4.3 for details.

		CIFAR-100			Tiny-ImageNet		
VC	EN	All	Old	New	All	Old	New
×	×	69.00	71.82	51.36	63.22	79.79	35.23
✓	×	70.04	80.06	50.45	63.13	79.92	35.14
×	✓	69.57	80.96	51.10	64.22	78.93	40.22
✓	✓	**71.27**	78.64	56.61	**64.40**	78.96	40.56

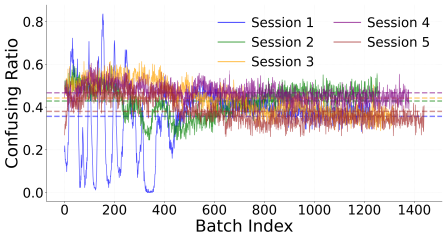

Fig. 5: Ratio of confusing samples used in training across sessions. This metric indicates how uncertain samples are handled during continual updates, reflecting the model's ability to utilize ambiguous data without forcing premature decisions.

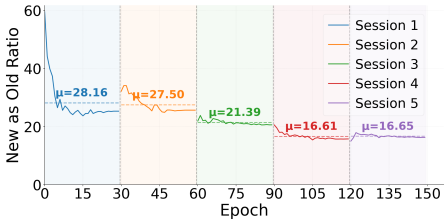

Fig. 6: Proportion of novel-class samples misclassified as old classes in each session. This metric measures the separability between old and emerging categories, showing how feature organization improves over time. See §4.3 for details.

it remains within a stable range, indicating balanced utilization of uncertain samples throughout continual updates.

New as Old Ratio. We further measure the proportion of novel-class samples misclassified as old classes to evaluate category separability. Fig. 6 shows that this ratio consistently decreases within each session, while the session-wise mean gradually stabilizes. This trend suggests that feature organization improves over time, reducing bias toward previously learned categories and enabling clearer separation between old and emerging classes.

Hyperparameter Sensitivity Analysis. All hyperparameters are selected via validation and kept unchanged across all experimental settings to ensure fair comparison. To examine the robustness of our framework and justify our default parameter choices, we conduct sensitivity studies on two core components: the ambiguity threshold for PC set construction and the neighborhood size adopted in ENCL. As reported in Table 5, our default settings (threshold 0.15, neighborhood size 3) consistently achieve the best average accuracy among nearby values. We note that the PC set retains samples satisfying both the top-2 prediction ambiguity and teacher–student consistency criteria, and the virtual-category weights are learned automatically during training instead of being manually tuned.

Table 5: Sensitivity analysis of ambiguity threshold and neighborhood size on C100.

Threshold	0.07	0.1	**0.15**	0.2	0.3	Neighborhood Size	2	**3**	4
Avg	68.63	70.53	**71.27**	70.12	68.42	Avg	70.52	**71.27**	69.09

Table 6: Computational overhead of ENCL on C100 with ViT-B/16.

Metric	w/o ENCL	Full model
Training Time (s/epoch)	338	384
GPU Memory (MiB)	17,936	18,824

Computational Efficiency. Our k-NN in ENCL is implemented with GPU-accelerated FAISS for efficiency. As summarized in Table 6, compared with "w/o ENCL", our full model increases training time from 338s to 384s per epoch, and GPU memory from 17,936 MiB to 18,824 MiB, indicating a moderate overhead.

5 Conclusion

In this paper, we present an ambiguity-aware framework for Continual Generalized Category Discovery that improves learning from unlabeled streams under continual distribution shifts. Instead of forcing uncertain samples into premature semantic assignments, our approach introduces Virtual Category Learning (VCL) as a principled mechanism to safely absorb ambiguous data through temporary virtual categories, enabling stable knowledge accumulation while mitigating error propagation across sessions. By allowing uncertain samples to contribute to representation learning without injecting noisy supervision, the proposed framework improves data utilization and alleviates prediction bias toward previously learned classes. Built upon this foundation, an expanded neighborhood contrastive strategy further refines the learned embedding space, promoting clearer separation between evolving categories and enhancing discovery stability over time. Extensive experiments across multiple benchmarks demonstrate consistent gains in overall recognition, novel class discovery, and forgetting mitigation, validating the effectiveness of learning with virtual categories for continual open-world recognition. Our results suggest that explicitly modeling uncertainty is a key step toward scalable and reliable continual category discovery systems.

Acknowledge

This work was supported by National Science Foundation of China (62302093, 62306292, 52441503), Jiangsu Province Natural Science Fund (BK20230833) and the Open Research Fund of the State Key Laboratory of Multimodal Artificial Intelligence Systems under Grant E5SP060116. We thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations.

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, pp. 12527–12535 (2023)
2. 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)
3. Chen, C., Han, J., Debattista, K.: Virtual category learning: A semi-supervised learning method for dense prediction with extremely limited labels. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(8), 5595–5611 (2024)
4. 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)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255. Ieee (2009)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
7. Han, K., Vedaldi, A., Zisserman, A.: Learning to discover novel visual categories via deep transfer clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8401–8409 (2019)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
9. Javed, K., White, M.: Meta-learning representations for continual learning. Advances in Neural Information Processing Systems **32** (2019)
10. Jia, X., Han, K., Zhu, Y., Green, B.: Joint representation learning and novel category discovery on single-and multi-modal data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 610–619 (2021)
11. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images. Technical report, University of Toronto (2009)
12. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N **7**(7), 3 (2015)
13. Li, Z., Dai, B., Simsek, F., Meinel, C., Yang, H.: Imbagcd: Imbalanced generalized category discovery. arXiv preprint arXiv:2401.05353 (2023)
14. Lopez-Paz, D., Ranzato, M.: Gradient episodic memory for continual learning. Advances in Neural Information Processing Systems **30** (2017)
15. Ma, S., Zhu, F., Zhong, Z., Liu, W., Zhang, X.Y., Liu, C.L.: Happy: A debiased learning framework for continual generalized category discovery. Advances in Neural Information Processing Systems **37**, 50850–50875 (2024)
16. McCloskey, M., Cohen, N.J.: Catastrophic interference in connectionist networks: The sequential learning problem. In: Psychology of Learning and Motivation, vol. 24, pp. 109–165 (1989)
17. McQueen, J.B.: Some methods of classification and analysis of multivariate observations. In: Proc. of 5th Berkeley Symposium on Math. Stat. and Prob. pp. 281–297 (1967)

18. 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)
19. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: icarl: Incremental classifier and representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2001–2010 (2017)
20. Rizve, M.N., Kardan, N., Khan, S., Shahbaz Khan, F., Shah, M.: Openldn: Learning to discover novel classes for open-world semi-supervised learning. In: European Conference on Computer Vision. pp. 382–401. Springer (2022)
21. Roy, S., Liu, M., Zhong, Z., Sebe, N., Ricci, E.: Class-incremental novel class discovery. In: European Conference on Computer Vision. pp. 317–333. Springer (2022)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
23. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7492–7501 (2022)
24. Wang, H., Sun, A., Gui, J., Wang, L.: Data-free class-incremental gesture recognition with prototype-guided pseudo-feature replay. IEEE Transactions on Image Processing **35**, 3623–3632 (2026)
25. Wang, J., Ma, Z., Nie, F., Li, X.: Progressive self-supervised clustering with novel category discovery. IEEE Transactions on Cybernetics **52**(10), 10393–10406 (2021)
26. 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)
27. Wu, Y., Chi, Z., Wang, Y., Feng, S.: Metagcd: Learning to continually learn in generalized category discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1655–1665 (2023)
28. Yu, Q., Ikami, D., Irie, G., Aizawa, K.: Self-labeling framework for novel category discovery over domains. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 3161–3169 (2022)
29. Zang, Z., Shang, L., Yang, S., Wang, F., Sun, B., Xie, X., Li, S.Z.: Boosting novel category discovery over domains with soft contrastive learning and all in one classifier. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11858–11867 (2023)
30. Zhang, X., Jiang, J., Feng, Y., Wu, Z.F., Zhao, X., Wan, H., Tang, M., Jin, R., Gao, Y.: Grow and merge: A unified framework for continuous categories discovery. Advances in Neural Information Processing Systems **35**, 27455–27468 (2022)
31. Zhong, Z., Fini, E., Roy, S., Luo, Z., Ricci, E., Sebe, N.: Neighborhood contrastive learning for novel class discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10867–10875 (2021)