

Beyond the Static World: Continual Category Discovery under Visual Drift

Wei Feng^{1,2} Yiwen Jiang¹ Sijin Zhou^{1,2} Zongyuan Ge^{1,2}

¹AIM for Health Lab, Monash University ²Airdoc–Monash Research, Monash University

{wf02429, sjzhou1995}@gmail.com, {yiwen.jiang, zongyuan.ge}@monash.edu

Abstract

Generalized Category Discovery (GCD) aims to identify both known and novel classes from unlabeled data with the aid of labeled examples. While promising, most existing GCD methods rely on simultaneous access to labeled and unlabeled datasets—an assumption often impractical in real-world deployments. Continual Category Discovery (CCD) relaxes this requirement by adapting a pre-trained model to streaming unlabeled data, yet it typically assumes domain-consistent data distributions. This places a strong limitation on its applicability. In this work, we study Open Continual Category Discovery (OCCD), where the model must robustly discover previously unseen concepts from real-world data streams that may originate from heterogeneous and shifting domains. To address this, we propose an adaptive framework built on three key ideas. First, we propose a weight-aware separation module, which leverages partial unbalanced optimal transport for instance probability modeling and employs binary response spectrum quantization to generate cues for distinguishing known and unknown categories, enabling automatic sample separation. Second, for known categories, we introduce a cross-domain semantic alignment module that incorporates adversarial learning to perform adaptive prototype matching, thereby enhancing robustness against domain shifts. Finally, for unknown categories, we design a category topology consistency constraint that preserves semantic relationships between known and novel classes during distribution shifts. Experiments show our approach excels at discovering new categories while maintaining strong performance on known ones in evolving domains.

1. Introduction

The rapid advances of deep neural networks in computer vision have significantly accelerated the deployment of AI systems in real-world applications [9–14, 22, 24, 26, 27]. However, these successes largely rely on the closed-world assumption, where models are trained and evaluated on a fixed set of predefined categories [8, 15, 23, 43]. In con-

trast, real-world environments are inherently open, where previously unseen categories may emerge during deployment. To bridge this gap, the task of Generalized Category Discovery (GCD) has gained increasing attention [47, 49]. GCD aims to automatically identify both known and novel categories from unlabeled data, by transferring knowledge from a partially labeled dataset containing only known categories.

Despite notable progress in recent years, most existing GCD methods require simultaneous access to both labeled and unlabeled datasets during training. This assumption limits their applicability in privacy-sensitive or decentralized environments where such data access is infeasible. To address this issue, a new paradigm called Continual Category Discovery (CCD) has emerged [28, 35, 38, 51]. CCD relaxes the data access constraint by assuming that the model leverages a pretrained model to discover categories from continuously arriving unlabeled data. This paradigm shift makes CCD more suitable for real-world applications involving continuous data streams and strict privacy constraints. While CCD alleviates the privacy concern, it still assumes that labeled and unlabeled data originate from the same domain. However, in many real-world scenarios, incoming data streams may simultaneously contain novel categories and arise from heterogeneous domain distributions—for example, medical images collected across different hospitals or street-view photos captured under varying weather conditions. Such shifts induce negative transfer, severely degrading performance and impeding reliable novel-category discovery.

To overcome the limitations of existing settings, we introduce a more realistic and challenging paradigm termed Open Continual Category Discovery (OCCD). OCCD requires models to incrementally learn from streams of unlabeled data that may simultaneously contain novel categories and stem from unseen domains. This setting better reflects real-world open environments. OCCD introduces several core challenges. First, the model must maintain accurate recognition of known categories despite distributional shifts across domains. Second, it must effectively discover and cluster novel categories relying solely on prior knowledge

of known classes, under non-stationary and unlabeled data streams. Notably, traditional unsupervised domain adaptation (UDA) methods risk negative transfer by aligning distributions without considering unknown classes, causing semantic drift. Likewise, CCD approaches assume stable distributions and struggle under domain shifts common in real-world scenarios.

To address these challenges, we propose a novel adaptive category discovery framework tailored to the OCCD setting, comprising three key components: First, we propose a weight-aware separation module (WSM) to automatically distinguish known and unknown samples, thereby avoiding incorrect alignments. It first employs partial unbalanced optimal transport (PUOT) to estimate instance probabilities. Building on this, it then leverages response spectrum quantization to extract informative cues for effective separation by measuring multi-scale marginal transport probabilities. Second, we propose a cross-domain semantic alignment strategy that leverages adversarial learning to align known-category prototypes, thereby improving the model’s robustness to domain shifts in known class recognition. Third, we propose a category topology consistency constraint that explicitly preserves the relational structure between known and unknown classes across domains, facilitating stable semantic transitions and supporting reliable discovery of novel categories. Our contributions can be summarized as follows:

- We formally define and investigate the Open Continual Category Discovery (OCCD) problem and introduce a unified framework to address it.
- We design a three-part solution that includes a weight-aware separation module, a cross-domain semantic alignment strategy, and a category topology consistency constraint, enabling robust category discovery under evolving domain shifts.
- We conduct extensive experiments across multiple OCCD benchmarks and show that our framework consistently surpasses state-of-the-art baselines, achieving strong novel-class discovery while maintaining reliable recognition of known categories in dynamic real-world settings.

2. Related Works

Domain Adaptation Domain Adaptation (DA) tackles the distribution gap between source and target domains. Early methods focused on learning domain-invariant features via statistical alignment [33, 42] or parameter regularization [40]. Adversarial approaches [1, 20] further improved alignment but suffered from instability. To handle real-world deployment, Test-Time Adaptation (TTA) emerged as a paradigm that adapts models to target distributions without requiring access to source data during inference. Typical methods fine-tune only specific components (e.g., batch normalization layers) to avoid overfit-

ting. For example, TENT [45] adapts the model by minimizing prediction entropy via batch norm statistics updates, while SHOT [31] freezes the feature extractor and leverages entropy and diversity regularization to encourage confident and diverse target predictions. Continual Test-Time Adaptation (CTTA) generalizes TTA to handle continuously evolving target domains. CoTTA [48] introduces a student–teacher framework that maintains prediction consistency through temporal ensembling and selective model updates. Building on this, RMT [5] combines symmetric cross-entropy and contrastive learning to improve adaptation robustness and mitigate forgetting during long-term deployment. Recently, Open-Set Source-free DA (OS-SFDA) methods address scenarios with unseen classes and no source data access, using entropy-based rejection [16] or pseudo-labeling [34]. However, existing methods typically focus only on recognizing known classes, overlooking the discovery of novel ones. In contrast, we target the more challenging open-world setting with both domain and category shifts, aiming to discover emerging classes while preserving recognition of known ones.

Category Discovery Category discovery aims to organize unlabeled data into semantically meaningful groups, often with partial supervision. Novel Class Discovery (NCD) [23, 52–54] focuses on clustering unlabeled samples from novel classes by transferring knowledge from disjoint labeled classes, initially relying on self-supervised pretraining and ranking-based clustering objectives [23]. Later works integrated multiple goals into end-to-end frameworks for better clustering [11, 17, 21].

Generalized Category Discovery (GCD) extends NCD by considering unlabeled data from both known and unknown classes [7, 43, 47, 49], requiring joint discovery and recognition. Early GCD methods adopted hybrid two-stage pipelines combining supervised and self-supervised learning with semi-supervised clustering [43, 50], while recent single-stage models apply entropy regularization and self-distillation for end-to-end training [2, 7, 47, 49].

Continual Category Discovery (CCD) [28, 35, 38, 51] further introduces sequentially arriving unlabeled data, requiring models to learn emerging classes while retaining prior knowledge. Without ground-truth labels, recent CCD approaches use proxy anchors [28], prompts [4], or energy-based mechanisms [38] to maintain the balance between stability and plasticity. However, most still assume the data stems from known domains, a limitation in realistic open environments.

3. Methodology

Problem Formulation In the task of Open Continual Category Discovery (OCCD), we are given an initial labelled

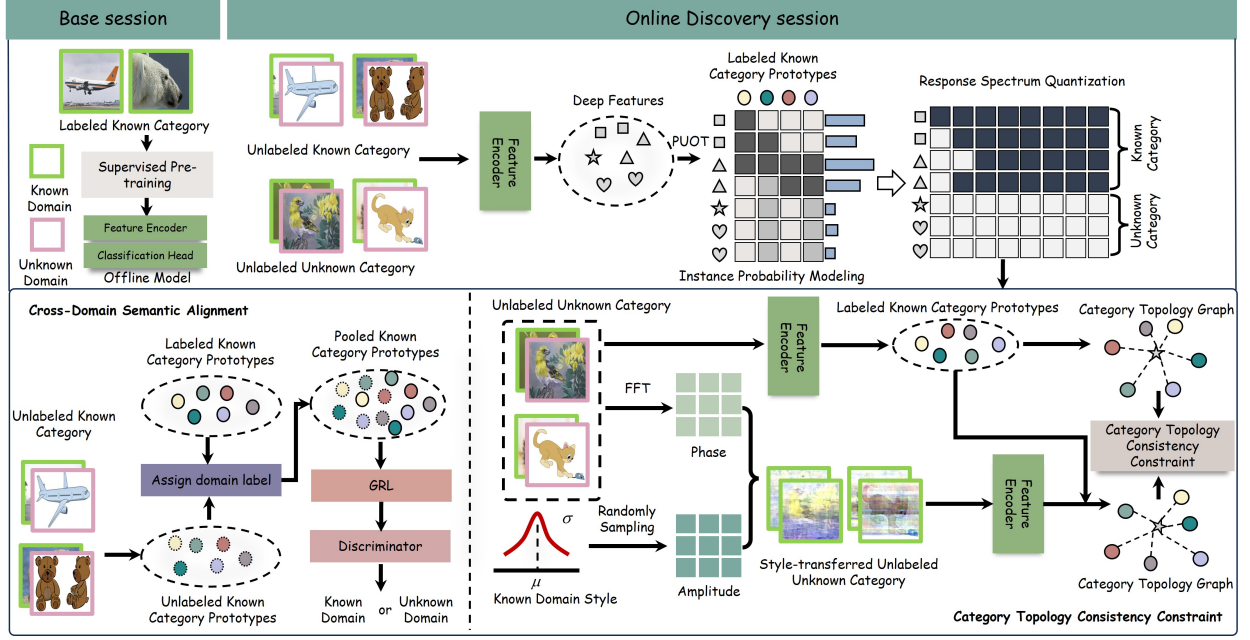


Figure 1. The overall framework of our proposed method.

dataset $\mathcal{D}^l = \{(x_i, y_i)\}_{i=1}^{N^l}$ from a known domain $\mathcal{T}_A$, where $x_i \in \mathcal{X}$ denotes the i -th input sample and $y_i \in \mathcal{Y}^k$ is the corresponding label from a known category set $\mathcal{Y}^k$. This labelled set is used during the initial stage to learn discriminative and transferable feature representations that serve as the foundation for subsequent category discovery. Over the course of T continual discovery stages, the model sequentially receives unlabelled datasets $\mathcal{D}_t^u = \{x_i\}_{i=1}^{N_t^u}$ (for $t = 1, \dots, T$), each containing samples drawn from a mixture of previously known categories and emerging novel categories. Importantly, these unlabelled data batches may also originate from previously unseen domains $\mathcal{T}_B$, introducing distribution shifts across time.

The goal of OCCD is to train a model H_θ , parameterized by θ , such that it can: (1) automatically identify and structure previously unseen categories from unlabelled data (novel class discovery); (2) maintain robust recognition of previously learned categories over time, avoiding catastrophic forgetting (known category retention); and (3) effectively adapt to evolving data distributions while mitigating the negative effects of domain shifts (distribution shift robustness). An overview of our framework is shown in Fig. 1. We first pretrain the backbone on labeled base-session data. During category discovery, a Weight-aware Separation Module distinguishes known and novel samples. For known classes, we enhance cross-domain robustness via Cross-Domain Semantic Alignment. For unknown samples, we impose a Category Topology Consistency Constraint to preserve semantic relationships under distribution shifts, facilitating novel class discovery.

Base Session Supervised Pretraining In the base session, given a labeled dataset $\mathcal{D}^l$ of known categories, we initialize the model using supervised training. The model consists of a feature extractor E and a prototype classifier head F . Specifically, we adopt the cross-entropy loss to pre-train the model and learn robust feature representations, which serves as a foundation for subsequent category discovery. The loss function is defined as:

$$\mathcal{L}_{ce} = -\frac{1}{|\mathcal{D}^l|} \sum_{(x_i, y_i) \in \mathcal{D}^l} y_i \log \sigma(F(E(x_i))), \quad (1)$$

where σ is the softmax function. Once the training is completed, the trained model is transferred to the next stage for online continual category discovery.

Weight-Aware Separation Module (WSM) To effectively distinguish unlabeled samples that belong to known classes from those that originate from unknown categories, we propose a novel two-stage framework, termed Weight-Aware Separation Module (WSM). It includes two stages: Instance Probability Modeling (IPM) to estimate class membership probabilities, and Response Spectrum Quantization (RSQ) to separate known and unknown category instances.

Instance Probability Modeling (IPM) We first extract feature representations $z_i^t = E(x_i^t)$ for unlabeled samples x_i^t using the pretrained feature extractor $E(\cdot)$. Meanwhile, the known class prototypes e are obtained from the prototype classifier of the previous stage. To establish the asso-

ciation between unlabeled features and known class prototypes, we propose a *Partial Unbalanced Optimal Transport (PUOT)* framework, which infers the class membership of each unlabeled instance:

$$\min_{\pi_{ij} \geq 0} \mathcal{J}_{\text{PUOT}} = \langle \mathbf{U}, \pi \rangle + \tau \cdot \text{KL}(\pi \mathbf{1}_{C_{\text{kno}}} \| \mathbf{a}) \quad \text{s.t. } \pi^\top \mathbf{1}_N = \mathbf{b}, \quad (2)$$

where $\pi \in \mathbb{R}^{N \times C_{\text{kno}}}$ is the transport plan between a mini-batch of N unlabeled samples and C_{kno} known prototypes. $\mathbf{a} \in \mathbb{R}^N$ and $\mathbf{b} \in \mathbb{R}^{C_{\text{kno}}}$ are uniform marginals with $a_i = 1/N, b_j = 1/C_{\text{kno}}$. The cost matrix $\mathbf{U}$ is defined as the pairwise cost between target features z^t and the corresponding class prototypes e^o at the o -th iteration: $U_{ij} = e^{\mathcal{E}(z_i^t)} \cdot \left\| z_i^t / \|z_i^t\| - e_j^{(o)} / \|e_j^{(o)}\| \right\|_2^2$, with entropy $\mathcal{E}(\cdot)$ defined by softmax predictions: $\mathcal{E}(x_i^t) = -\sum_{j=1}^{|C_{\text{kno}}|} \sigma(F(E(x_i^t)))[j] \log \sigma(F(E(x_i^t)))[j]$. When the entropy $\mathcal{E}(x_i^t)$ is high, it reflects substantial model uncertainty in assigning unlabeled samples to known categories. The Kullback-Leibler divergence is formulated as: $\text{KL}(x \| z) = \sum_{d=1}^D [x_d \log(x_d/z_d) - x_d + z_d]$. τ is a tunable hyperparameter. Departing from traditional optimal transport, we relax the marginal constraint associated with the target distribution using the above KL penalty to accommodate potential misalignments. To effectively optimize this relaxed objective, we begin by deriving the dual formulation of the PUOT problem.

Dual Form and Optimization The PUOT objective in Eq. 2 admits the following dual formulation:

$$\begin{aligned} \min_{\mathbf{f}, \mathbf{g}, \zeta} & \left[\tau \sum_{i=1}^N a_i \exp\left(-\frac{f_i + \zeta}{\tau}\right) - \sum_{j=1}^{C_{\text{kno}}} b_j (g_j - \zeta) \right] \\ \text{s.t. } & f_i + g_j + s_{ij} = U_{ij}, \quad s_{ij} \geq 0, \end{aligned} \quad (3)$$

where $\mathbf{f}, \mathbf{g}, \mathbf{s}$, and ζ are Lagrange multipliers, and s_{ij} is a slack term. The non-negativity of s_{ij} follows from KKT conditions. This leads to an equivalent transport formulation with sample-level reweighting:

$$\min_{\pi_{ij} \geq 0} \langle \mathbf{U}, \pi \rangle \quad \text{s.t. } \sum_{j=1}^{C_{\text{kno}}} \pi_{ij} = a_i e^{-\frac{f_i^* + \zeta^*}{\tau}}, \quad \sum_{i=1}^N \pi_{ij} = b_j \quad (4)$$

The full proof is provided in the Appendix. This reveals that the marginal probability associated with each unlabeled sample x_i^t becomes: $q_\tau(x_i^t) = a_i \cdot \exp(-(f_i^* + \zeta^*)/\tau)$, which implicitly captures the uncertainty of assigning samples to known prototypes via the Lagrange multipliers f_i^* and ζ^* . Note that it is not necessary to compute the full transport matrix $\pi \in \mathbb{R}^{N \times C_{\text{kno}}}$ explicitly, which is often computationally demanding. Instead, we aim to optimize

the Lagrange multipliers, reducing the complexity from $N \cdot C_{\text{kno}}$ to $N + C_{\text{kno}}$.

To derive an efficient optimization strategy, we employ the classical c -transform from optimal transport theory [32], which provides an upper bound for the dual potential g_j : $g_j = \inf_{k \in [N]} (U_{kj} - f_k)$. Substituting this bound into the dual formulation yields the final objective:

$$\min_{\mathbf{f}, \zeta} \left[\tau \sum_{i=1}^N a_i e^{-\frac{f_i + \zeta}{\tau}} - \sum_{j=1}^{C_{\text{kno}}} \left(\inf_{k \in [N]} (U_{kj} - f_k) - \zeta \right) b_j \right] \quad (5)$$

This reformulation allows the PUOT optimization to be conducted efficiently by focusing solely on the dual potentials $\mathbf{f}$ and ζ . The full derivation and detailed optimization steps are provided in Appendix.

Response Spectrum Quantization We observe that the parameter τ in PUOT directly influences the marginal transport probabilities of unlabeled samples. A larger τ tends to assign higher weights to more unlabeled samples, thus increasing their likelihood of being matched. However, the lack of prior knowledge regarding the proportion of unknown classes in the unlabeled data makes it non-trivial to select an appropriate τ . To address this challenge, we propose a novel *Response Spectrum Quantization* to identify unlabeled samples that are likely to belong to unknown categories. Specifically, for each unlabeled sample x^t , we compute its marginal probability $q_\tau(x^t)$ over a range of τ values: $\tau \in \{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$, and then collect marginal probabilities as $\mathbf{Q} \in \mathbb{R}^{N \times 7}$. This mechanism leverages the tendency of PUOT to assign higher weights to unlabeled samples similar to known classes, while down-weighting those likely from unknown categories. We can construct binary spectrum matrix as: $\mathbf{Q}_{ij} = \text{func}(Q_{ij} - 1/N)$, where $\text{func}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$.

Then we perform K-means clustering on $\mathbf{Q}$ to obtain centers $\mathcal{C}_0, \mathcal{C}_1$. Define: $\mathcal{C}_{\text{known}} = \max(\|\mathcal{C}_0\|_2, \|\mathcal{C}_1\|_2)$, $\mathcal{C}_{\text{unknown}} = \min(\|\mathcal{C}_0\|_2, \|\mathcal{C}_1\|_2)$. Each unlabeled sample is assigned to the closest center. Samples nearer to $\mathcal{C}_{\text{unknown}}$ are labeled as unknown (x_{unk}^t), and the rest as known (x_{kno}^t), forming the final partition: $x^t = x_{\text{unk}}^t \cup x_{\text{kno}}^t$.

Cross-Domain Semantic Alignment (CDSA) After separating samples belonging to known categories, since these samples may contain data from unknown domains. To alleviate domain distribution shift, we propose a Cross-Domain Semantic Alignment (CDSA) module that explicitly minimizes the distributional discrepancy between known-class prototypes across domains. Specifically, we begin by extracting known-class prototypes $\mathcal{P}_l = [e_1^{(l)}; \dots; e_{C_{\text{kno}}}^{(l)}]$ from the pretrained prototype classifier F . For the unlabeled data

that are predicted to belong to known classes, inspired by weighted k-means [31], we adopt a confidence-weighted aggregation strategy to estimate class-wise prototypes $\mathcal{P}_u$. This approach accounts for each sample's confidence with respect to each class, resulting in more reliable prototype estimation. The weighted prototype $e_c^{(u)}$ for class c is computed as: $e_c^{(u)} = \frac{\sum_{i=1}^N p_{i,c}^u z_i}{\sum_{i=1}^N p_{i,c}^u}$, where z_i denotes the feature of the i -th sample and $p_{i,c}^u$ is its predicted probability of belonging to class c .

Next, We treat $\mathcal{P}_l$ as source-domain prototypes (labeled as $d = 0$), and $\mathcal{P}_u$ as target-domain prototypes (labeled as $d = 1$). Since the unlabeled known-class samples may include instances from unseen domains, their prototypes $\mathcal{P}_u$ may exhibit domain-specific shifts away from the source distribution. To align the prototype distributions across domains, we employ adversarial learning. A simple CNN-based discriminator D is used to distinguish whether a prototype comes from the known source domain or target domain. The discriminator is optimized using the binary cross-entropy loss:

$$\mathcal{L}_{\text{dis}} = - \sum_n [d \log p(n) + (1 - d) \log(1 - p(n))], \quad (6)$$

where $p(n)$ is the discriminator's prediction for the n -th prototype and $d \in \{0, 1\}$ is its domain label.

We incorporate a Gradient Reversal Layer (GRL) [19] to enable end-to-end adversarial training. While the discriminator learns to differentiate the domain origin of prototypes, the feature extractor is optimized to confuse the discriminator. This promotes the learning of domain-invariant representations. The adversarial objective is formulated as: $\mathcal{L}_{\text{adv}} = \max_P \min_D \mathcal{L}_{\text{dis}}$, where P denotes the prototype set and D is the discriminator. By aligning class-wise prototypes across domains in this manner, our framework effectively reduces domain-induced discrepancies in the feature space, facilitating better generalization on known categories.

Category Topology Consistency Constraint (CTCC)

To enhance the robustness of unknown class recognition among unlabeled samples, we propose a category topology consistency constraint. The core idea aligns with the original motivation of GCD: leveraging the semantic relationships between known and unknown classes to facilitate the discovery of novel categories. We argue that such relational topology should remain invariant under domain shifts. For instance, a palm tree should be more similar to maple tree than to a house, regardless of the domain. To this end, we first expose the model to diverse styles through frequency-based augmentation. Specifically, we pre-compute the amplitude information for each image x_i in the training set from the previous stage. We apply the Fast Fourier Transform (FFT) to obtain its frequency-domain representation:

$\hat{x}_i = \mathcal{F}(x_i)$, where $\mathcal{F}(\cdot)$ denotes the FFT operator. We then extract the corresponding amplitude spectrum $\mathcal{A}_i$:

$$\mathcal{A}_i(u, v) = \sqrt{\mathcal{R}^2(\hat{x}_i(u, v)) + \mathcal{I}^2(\hat{x}_i(u, v))}, \quad (7)$$

where $\mathcal{R}(\hat{x}_i(u, v))$ and $\mathcal{I}(\hat{x}_i(u, v))$ denote the real and imaginary parts of the frequency coefficient at spatial frequency coordinate (u, v) , respectively. We then estimate the mean and variance of the amplitude spectrum across the dataset: $\mu_{\mathcal{A}}(u, v) = \frac{1}{N} \sum_{i=1}^N \mathcal{A}_i(u, v)$, $\sigma_{\mathcal{A}}^2(u, v) = \frac{1}{N} \sum_{i=1}^N (\mathcal{A}_i(u, v) - \mu_{\mathcal{A}}(u, v))^2$.

Given an unknown category sample $x_{i,\text{unk}}^t$, we extract its amplitude $\mathcal{A}_i^t$ and phase $\mathcal{P}_i^t$, and perform a reparameterized AdaIN-style transformation [25] to augment the amplitude:

$$\begin{aligned} \tilde{\mathcal{A}}_i^t(u, v) &= \gamma(u, v) \cdot \frac{\mathcal{A}_i^t(u, v) - \mu(\mathcal{A}_i^t)}{\sigma(\mathcal{A}_i^t)} + \beta(u, v), \\ \beta(u, v) &= \mu_{\mathcal{A}}(u, v) + \epsilon_{\mu} \cdot \sigma_{\mathcal{A}}(u, v), \\ \gamma(u, v) &= \sigma_{\mathcal{A}}(u, v) + \epsilon_{\sigma} \cdot \sigma_{\mathcal{A}}(u, v) \end{aligned} \quad (8)$$

where $\epsilon_{\mu}, \epsilon_{\sigma} \sim \mathcal{N}(0, 1)$. The stylized image is reconstructed via inverse Fourier transform: $\hat{x}_{i,\text{unk}}^t = \tilde{\mathcal{A}}_i^t \cdot e^{j\mathcal{P}_i^t}$, $\hat{x}_{i,\text{unk}}^t = \mathcal{F}^{-1}(\hat{x}_{i,\text{unk}}^t)$. Next, we extract the feature representations of both the original and style-transferred unknown category images, denoted as $z_{i,\text{unk}}^t = f(x_{i,\text{unk}}^t)$ and $\tilde{z}_{i,\text{unk}}^t = f(\hat{x}_{i,\text{unk}}^t)$, respectively. Let $\mathcal{P}_l = [e_1^{(l)}; \dots; e_{C^{\text{kno}}}^{(l)}]$ denote the set of known class prototypes, we compute the pairwise distances between the sample and each prototype: $\hat{w}_{i,c} = \|\tilde{z}_{i,\text{unk}}^t - e_c^{(l)}\|_2$, $w_{i,c} = \|z_{i,\text{unk}}^t - e_c^{(l)}\|_2$. These distances define a sample-specific class-relation topology graph, characterizing the relative positioning of the sample to known semantic centers under different augmentations.

To enforce topological consistency, we minimize the discrepancy between the two graphs by comparing the similarity distributions across augmentations. Following the Gaussian potential-based formulation, we define the category topology consistency loss as:

$$\mathcal{L}_{\text{topo}} = \frac{1}{N_u} \sum_{i=1}^{N_u} \log \left(\frac{1}{C^{\text{kno}}} \sum_c \exp(-m(\hat{w}_{i,c} - w_{i,c})^2) \right), \quad (9)$$

where N_u is the number of unlabeled samples, C^{kno} denotes the number of known classes from the previous stage, and $m = 2$ is a fixed parameter. This loss encourages the model to preserve the relative similarity structure between unknown classes and known classes across different views, providing a structure-aware regularization signal.

Clustering Objective Following prior work [38], we adopt a pseudo-labeling strategy to facilitate label assignment in the absence of true annotations. For data samples from known categories—where the model has already acquired reliable semantic knowledge during earlier

stages—we directly leverage the predictions from the previous model checkpoint to assign pseudo-labels. For samples from previously unseen categories, which do not correspond to any known class, we apply Affinity Propagation [18] to automatically infer cluster memberships. As a non-parametric clustering algorithm, affinity propagation estimates the optimal number of clusters by passing messages between samples based on pairwise similarity, making it particularly suitable for the open-world setting where the number of novel classes is unknown.

The resulting number of clusters is then used to dynamically expand the online classifier, allowing it to incorporate newly emerging categories. During online learning, we update the model incrementally using a standard cross-entropy loss over the combined pseudo-labeled known and novel samples, enabling continual acquisition of novel semantic knowledge without revisiting prior data. The overall optimization objective of our method is given by: $\mathcal{L}_{total} = \mathcal{L}_{ce} + \lambda_1 \mathcal{L}_{adv} + \lambda_2 \mathcal{L}_{topo}$, where λ_1 and λ_2 are balancing hyperparameters.

4. Experiments

4.1. Experimental Setup

Datasets. We conduct extensive experiments on two representative benchmarks: Corrupted Semantic Shift Benchmark (SSB-C) [46] and DomainNet [39]. The SSB-C benchmark is an extension of the original Semantic Shift Benchmark (SSB), which comprises three fine-grained datasets—CUB [44], Stanford Cars [29], and FGVC-Aircraft [36]. To simulate distribution shifts, SSB-C introduces nine corruption types applied to the original images in SSB, each at five severity levels, forming a challenging testbed for evaluating model robustness under both semantic and visual perturbations. DomainNet is a large-scale, real-world benchmark featuring approximately 600,000 images across 345 categories spanning six diverse domains: Real, Clipart, Infograph, Painting, Quickdraw, and Sketch. Following the protocol of [46], we designate the Real domain as the known source domain $\mathcal{T}_A$ and treat each remaining domain as an unknown target domain $\mathcal{T}_B$. In this setup, each stage samples data from a single target domain, while the appendix also reports results for a mixed-target setting where all target domains are combined for sampling. For SSB-C, we adopt a similar configuration: the clean version of each dataset serves as $\mathcal{T}_A$, and its corrupted counterpart serves as $\mathcal{T}_B$. Additionally, a domain-incremental setting—where each stage introduces samples from a different domain—is included in the appendix. For category partitioning, we follow the setup in [4]. Specifically, the model is pre-trained on labeled samples from middle-level categories in the known domain during the initial stage. In subsequent discovery stages, both known and novel samples

from $\mathcal{T}_A$ and $\mathcal{T}_B$ are progressively introduced. Detailed data splits are summarized in the appendix.

Implementation Details and Evaluation Metrics. For all experiments, we use a ViT-B/16 backbone [6] pretrained with DINO [3, 37]. Following [38, 49], we fine-tune the final block using SGD for 50 epochs per stage with a batch size of 128. The learning rate is initialized at 0.1 and decayed to 0.0001 via cosine annealing, and weight decay is set to 5×10^{-5} . Hyperparameters λ_1 and λ_2 are set to 1, and the number of stages T is set to 3 (see Appendix for sensitivity analysis). All experiments are repeated three times with different random seeds, and we report averaged results. Models are implemented in PyTorch and trained on eight NVIDIA RTX 4090 GPUs.

For evaluation, we adopt clustering accuracy (ACC) as the primary metric. Following prior work [43, 46], we compute the optimal label assignment using the Hungarian algorithm [30] and define $ACC = \max_{g \in \mathcal{G}(\mathcal{Y}_U)} \frac{1}{|\mathcal{D}_t^u|} \sum_{i=1}^{|\mathcal{D}_t^u|} \mathbb{1}\{y_i = g(\hat{y}_i)\}$. To evaluate continual discovery performance, we also report continual clustering accuracy (cACC) [4] across streaming sessions, including accuracy on *All*, *Old*, and *New* classes for both known and unknown domains.

4.2. Comparison with Other SOTA Algorithms

We compare our approach with several representative CCD methods, including: Grow & Merge (G&M) [51]; PromptCCD [4]; PA-CGCD [28]; DEAN [38]. In addition, we re-implement several GCD methods to adapt them for the continual discovery setting: GCD [43]; SimGCD [49]; SPTNet [47]. Furthermore, we also include a comparison with HiLo [46], a recent method designed to handle distribution shifts in GCD. As shown in Table 1 and Table 2, when the unlabeled data stream contains domain shifts, both existing GCD and CCD methods suffer a significant performance degradation. Their clustering accuracy drops notably across both known and novel classes. This demonstrates that distribution mismatch significantly hampers the model’s ability to discover novel categories and recognize known ones. While HiLo alleviates the shift to some extent, its improvement remains limited. We suspect this is due to the complexity of online data streams—its patch-mixing augmentation may inadvertently introduce semantic noise, thus destabilizing convergence and impairing representation learning. In contrast, our approach achieves significant performance gains across all datasets. For example, on the CUB-C dataset, our model outperforms the state-of-the-art CCD method PromptCCD by 22.8% in accuracy on all categories in the original domain, and by 19.8% in the corrupted domain. This demonstrates the robustness of our method in identifying both known and unknown classes under continuous data streams. Moreover, on the more challenging

Table 1. Clustering performance on DomainNet benchmark. We use Real as the known domain $\mathcal{T}_A$ and each of the remaining domains as the unknown domain $\mathcal{T}_B$. We report the average All / Old / New accuracy across all stages for both domains.

Method	Real → Painting						Real → Sketch						Real → Quickdraw						Real → Clipart						Real → Infograph					
	Real			Painting			Real			Sketch			Real			Quickdraw			Real			Clipart			Real			Infograph		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	51.3	67.2	45.4	27.4	26.7	28.1	52.3	65.7	41.7	9.2	14.5	10.1	38.7	56.2	29.6	5.0	4.7	5.8	46.7	65.7	40.1	14.5	21.2	10.1	39.8	55.3	32.4	8.1	9.8	6.4
SimGCD	48.4	63.9	41.3	22.6	22.4	23.5	48.5	60.2	36.5	7.2	11.3	9.2	32.4	50.3	23.5	4.2	4.0	5.1	40.2	58.8	33.5	10.3	18.8	8.2	33.6	49.2	27.8	6.7	7.8	5.2
SPTNet	49.8	64.5	42.5	24.1	23.5	24.3	49.9	62.3	37.8	7.9	11.7	9.6	34.8	52.6	24.8	4.9	4.6	5.5	43.1	60.3	35.9	11.6	19.3	8.9	35.9	51.4	29.8	7.2	8.0	5.9
HiLo	56.1	70.9	49.2	31.0	31.8	30.8	55.7	71.3	47.2	13.1	17.2	12.4	43.9	61.2	35.4	5.9	5.5	6.7	55.4	70.8	48.1	18.2	25.1	13.0	49.4	65.2	42.1	9.8	12.5	8.1
G&M	47.1	62.3	41.2	26.3	25.5	26.2	48.9	63.4	42.3	10.9	15.1	10.5	34.1	50.2	27.3	3.5	4.1	5.2	40.3	61.1	34.2	12.3	19.2	8.8	32.4	50.1	27.6	7.5	9.2	5.5
PA-CGCD	55.4	70.3	48.1	30.1	30.8	30.2	55.1	70.7	46.6	12.3	16.1	11.2	43.6	60.4	34.2	5.1	5.0	6.0	52.2	70.3	44.6	17.8	24.5	12.3	45.2	61.3	38.1	9.0	11.8	7.1
DEAN	56.0	71.7	47.9	32.8	34.4	31.5	56.7	71.5	47.6	12.9	16.8	11.2	43.6	60.4	34.2	5.1	5.0	6.1	55.1	72.7	47.5	20.3	26.7	15.0	46.7	62.3	40.8	9.5	12.5	7.9
PromptCCD	56.5	71.2	50.3	31.5	32.1	31.2	57.4	73.6	48.6	13.4	17.7	12.1	45.2	62.3	36.7	5.8	5.1	6.5	54.1	71.2	46.7	19.8	26.1	14.4	47.1	63.1	40.2	9.2	12.2	7.8
Ours	60.2	73.3	54.2	38.9	38.5	37.7	59.3	73.4	50.2	16.3	19.3	15.2	53.5	73.1	48.9	6.9	6.2	7.1	57.4	72.5	50.7	23.5	29.8	18.6	59.3	73.5	52.1	10.7	13.3	9.2

Table 2. Clustering performance on SSB-C benchmarks. Each dataset contains both Original and Corrupted settings, and we report the average All / Old / New accuracy across all stages for both domains.

Method	CUB-C						Stanford Cars-C						FGVC-Aircraft-C					
	Original			Corrupted			Original			Corrupted			Original			Corrupted		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	29.4	47.7	23.4	26.8	45.9	20.1	26.4	56.1	21.5	22.3	43.1	11.2	27.7	33.6	24.9	28.8	41.4	28.8
SimGCD	26.6	44.5	21.0	23.4	42.4	17.7	23.1	52.5	18.9	19.3	39.7	9.8	25.4	30.1	22.1	25.2	38.1	25.8
SPTNet	27.8	45.2	22.0	25.1	44.2	18.1	24.9	55.0	20.3	21.1	41.6	9.9	26.1	31.2	23.3	26.9	39.5	26.7
HiLo	32.3	50.6	27.5	28.1	48.1	22.3	30.5	58.8	25.2	25.7	46.8	13.4	32.9	37.1	28.4	33.8	45.6	32.1
G&M	16.4	34.1	10.5	13.7	32.1	7.7	15.7	43.8	12.3	11.4	30.5	6.7	20.5	24.8	17.9	21.6	32.7	22.3
PA-CGCD	28.3	46.5	22.7	25.4	44.7	18.4	25.2	55.1	20.9	21.2	41.5	10.2	26.4	31.4	23.7	27.8	40.1	27.2
DEAN	28.9	47.1	23.0	26.3	46.2	18.2	26.1	58.1	19.4	22.1	41.2	12.9	28.1	32.8	28.9	29.1	40.1	30.3
PromptCCD	30.1	48.1	24.5	27.4	46.1	20.3	27.4	57.4	22.1	23.1	44.4	11.4	29.9	34.5	26.4	30.3	42.9	29.9
Ours	52.9	68.8	47.1	47.2	64.4	40.4	40.1	62.3	32.4	35.5	59.8	25.6	43.3	52.3	42.2	38.9	49.5	37.9

Table 3. Ablation study of different components on **Real → Painting**. We report clustering performance for both domains under different component settings.

Components			Real			Painting		
WSM	CDSA	CTCC	All	Old	New	All	Old	New
✗	✗	✗	54.6	68.7	46.5	28.7	28.1	27.9
✓	✓	✗	57.7	72.2	49.2	34.1	35.4	31.9
✓	✗	✓	57.1	70.4	52.0	32.8	31.3	34.8
✓	✓	✓	60.2	73.3	54.2	38.9	38.5	37.7

DomainNet benchmark, our method again delivers strong performance. For the ‘Real’ and ‘Painting’ domain pair, we observe similar improvements—3.7% and 7.4% gains in accuracy on all categories over PromptCCD in the respective domains. These results further highlight the practical adaptability of our method in open-world continual learning scenarios.

4.3. Analysis and Ablation Study

Component Effectiveness Analysis. We conduct ablation studies to assess the contribution of each module in our framework. The baseline refers to assigning cluster labels directly to samples without any auxiliary components. As shown in Table 3, the baseline performs poorly, highlighting the detrimental impact of distribution shift on

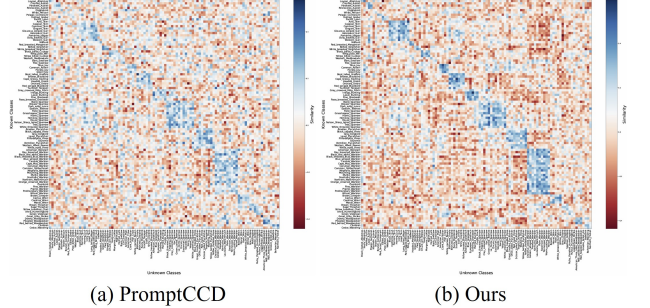


Figure 2. Visualization of relationships between known and discovered categories. Pairwise prototype similarities reveal that categories within the same super-class exhibit stronger relationships, indicating stable semantic structure across domains and supporting reliable novel category discovery.

clustering. Incorporating the WSM module for class separation and adding the CDSA module significantly improves recognition accuracy for known classes, demonstrating their effectiveness in aligning known category distributions and mitigating forgetting in continual learning. Further introducing the CTCC module yields notable gains in novel category discovery, indicating that enforcing cross-domain category-level topological consistency enables robust modeling of semantic relations between known and unknown classes. We further visualize the relationships be-

tween known and newly discovered categories. As shown in Fig. 2, these semantic structures remain stable across domains, and categories belonging to the same super-class exhibit stronger inter-class relationships in the embedding space. This demonstrates the model’s ability to transfer relational knowledge across domains, enabling more reliable discovery of novel categories.

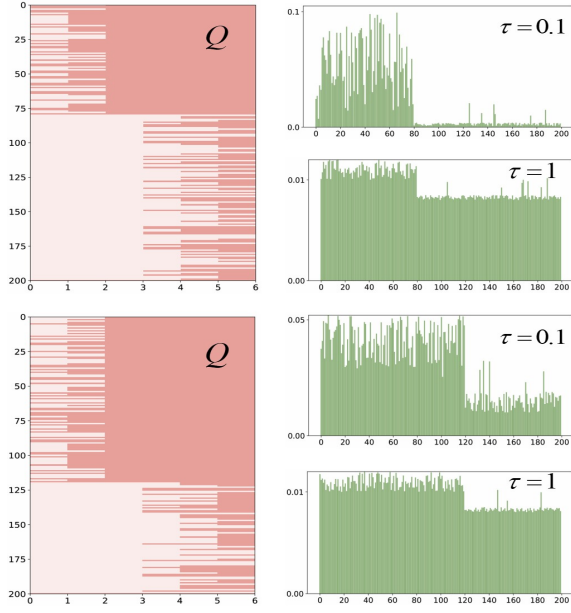


Figure 3. Visualization of the weight-aware separation results on 200 unlabeled CUB-C samples, where 40% or 60% of the samples are from unknown categories. We apply Instance Probability Modeling to estimate sample-level weights and construct the response matrix $\mathbf{Q}$, which facilitates the separation of known and unknown samples.

Table 4. Ablation study on separation strategy.

Method	Real			Painting		
	All	Old	New	All	Old	New
entropy-based	54.9	69.1	46.9	29.4	28.8	28.3
energy-based	55.6	69.7	47.8	30.1	28.9	29.4
Ours	60.2	73.3	54.2	38.9	38.5	37.7

Effectiveness of the Separation Module. We also conduct a dedicated analysis on the effectiveness of our proposed separation module. Specifically, we compare our PUOT-based weight-aware separation strategy with two alternative baselines: (1) an entropy-based separation [41], and (2) an energy-based separation method [38]. As shown in Table 4, our method consistently outperforms these baselines. The proposed PUOT relaxes the rigid alignment constraint, enabling a more flexible and accurate partitioning of unlabelled samples. By measuring **multi-scale marginal probabilities**, it better adapts to partially mismatched distributions, leading to improved clustering performance for

both known and novel categories. Furthermore, as illustrated in Fig. 3, the value of Q for unknown class samples are notably lower, allowing them to be easily distinguished. This indicates that our method is capable of adapting to varying proportions of unknown class samples, even in the absence of any prior knowledge.

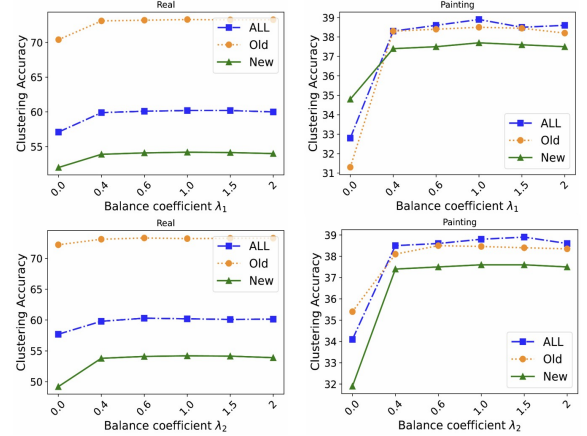


Figure 4. Sensitivity analysis of key hyperparameters.

Hyperparameter Sensitivity Analysis. To evaluate the robustness of our method with respect to the loss-balancing coefficients λ_1 and λ_2 , we perform a sensitivity study on DomainNet using *Real* as the known domain and *Painting* as the unknown domain (Figure 4). We systematically vary both coefficients and report clustering accuracies for all, old, and new classes. When both coefficients are set to zero, performance drops markedly, confirming the necessity of the corresponding components. In contrast, assigning non-zero values within a reasonable range yields consistently stable and high accuracy across category types. Overall, the results demonstrate that our method is largely insensitive to the specific choices of λ_1 and λ_2 , highlighting its robustness to hyperparameter variations.

5. Conclusion

In this paper, we tackle the challenging Open Continual Category Discovery (OCCD) task by proposing a robust framework composed of three key modules. The Weight-aware Separation Module enables automatic differentiation of known and unknown samples via partial unbalanced optimal transport and response spectrum quantization. The Cross-Domain Semantic Alignment module ensures reliable prototype matching for known classes under domain shifts, while the Category Topology Consistency Constraint (CTCC) preserves semantic relations for unknown class discovery. Extensive experiments demonstrate that our approach effectively discovers novel categories while maintaining strong performance on known classes across evolving domains.

References

- [1] Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, and Mario Marchand. Domain-adversarial neural networks. *arXiv preprint arXiv:1412.4446*, 2014. 2
- [2] Xinzi Cao, Xiawu Zheng, Guan hong Wang, Weijiang Yu, Yunhang Shen, Ke Li, Yutong Lu, and Yonghong Tian. Solving the catastrophic forgetting problem in generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16880–16889, 2024. 2
- [3] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 6
- [4] Fernando Julio Cendra, Bingchen Zhao, and Kai Han. Promptccd: Learning gaussian mixture prompt pool for continual category discovery. In *European conference on computer vision*, pages 188–205. Springer, 2024. 2, 6
- [5] Mario Döbler, Robert A Marsden, and Bin Yang. Robust mean teacher for continual and gradual test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7704–7714, 2023. 2
- [6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 6
- [7] Yu Duan, Junzhi He, Runxin Zhang, Rong Wang, Xuelong Li, and Feiping Nie. Prediction consistency regularization for generalized category discovery. *Information Fusion*, page 102547, 2024. 2
- [8] Wei Feng and Zongyuan Ge. Generalized category discovery under domain shift: A frequency domain perspective. *arXiv preprint arXiv:2511.00573*, 2025. 1
- [9] Wei Feng, Lin Wang, Lie Ju, Xin Zhao, Xin Wang, Xiaoyu Shi, and Zongyuan Ge. Unsupervised domain adaptive fundus image segmentation with category-level regularization. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*, pages 497–506. Springer, 2022. 1
- [10] Wei Feng, Meihan Duan, Bingjie Wang, Yu Du, Yiran Zhao, Bin Wang, Lin Zhao, Zongyuan Ge, and Yuntao Hu. Automated segmentation of choroidal neovascularization on optical coherence tomography angiography images of neovascular age-related macular degeneration patients based on deep learning. *Journal of Big Data*, 10(1):111, 2023.
- [11] Wei Feng, Lie Ju, Lin Wang, Kaimin Song, and Zongyuan Ge. Towards novel class discovery: A study in novel skin lesions clustering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 24–33. Springer, 2023. 2
- [12] Wei Feng, Lie Ju, Lin Wang, Kaimin Song, Xin Zhao, and Zongyuan Ge. Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 623–631, 2023.
- [13] Wei Feng, Qiuqing Huang, Tong Ma, Lie Ju, Zongyuan Ge, Yuzhong Chen, and Peiquan Zhao. Development and validation of a semi-supervised deep learning model for automatic retinopathy of prematurity staging. *Iscience*, 27(1), 2024.
- [14] Wei Feng, Duong Nhu, Alison Anderson, Daniel Thom, Sarah N Barnard, Robert Zeibich, Emma Foster, Mark Howard, Susannah T Bellows, Rosemary Burgess, et al. Integrative deep learning of genomic and clinical data for predicting treatment response in newly diagnosed epilepsy. *Neurology*, 105(10):e214315, 2025. 1
- [15] Wei Feng, Sijin Zhou, Yiwen Jiang, and Zongyuan Ge. Prism: Progressive robust learning for open-world continual category discovery. In *The Fourteenth International Conference on Learning Representations*, 2026. 1
- [16] Zeyu Feng, Chang Xu, and Dacheng Tao. Open-set hypothesis transfer with semantic consistency. *IEEE Transactions on Image Processing*, 30:6473–6484, 2021. 2
- [17] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9284–9292, 2021. 2
- [18] Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007. 6
- [19] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015. 5
- [20] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016. 2
- [21] Peiyan Gu, Chuyu Zhang, Ruijie Xu, and Xuming He. Class-relation knowledge distillation for novel class discovery. *lamp*, 12(15.0):17–5, 2023. 2
- [22] Haris Hakeem, Wei Feng, Zhibin Chen, Jiun Choong, Martin J Brodie, Si-Lei Fong, Kheng-Seang Lim, Junhong Wu, Xuefeng Wang, Nicholas Lawn, et al. Development and validation of a deep learning model for predicting treatment response in patients with newly diagnosed epilepsy. *JAMA neurology*, 79(10):986–996, 2022. 1
- [23] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Autonovel: Automatically discovering and learning novel visual categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6767–6781, 2021. 1, 2
- [24] Wenlong Hang, Wei Feng, Shuang Liang, Lequan Yu, Qiong Wang, Kup-Sze Choi, and Jing Qin. Local and global structure-aware entropy regularized mean teacher model for 3d left atrium segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I* 23, pages 562–571. Springer, 2020. 1

- [25] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017. 5
- [26] Yiwen Jiang, Deval Mehta, Wei Feng, and Zongyuan Ge. Enhancing interpretable image classification through llm agents and conditional concept bottleneck models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12285–12297, 2025. 1
- [27] Lie Ju, Yicheng Wu, Wei Feng, Zhen Yu, Lin Wang, Zhuoting Zhu, and Zongyuan Ge. Universal semi-supervised learning for medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 355–365. Springer, 2024. 1
- [28] Hyungmin Kim, Sungho Suh, Daehwan Kim, Daun Jeong, Hansang Cho, and Junmo Kim. Proxy anchor-based unsupervised learning for continuous generalized category discovery. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16688–16697, 2023. 1, 2, 6
- [29] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 554–561, 2013. 6
- [30] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955. 6
- [31] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020. 2, 5
- [32] Weiming Liu, Xiaolin Zheng, Chaochao Chen, Jiahe Xu, Xinting Liao, Fan Wang, Yanchao Tan, and Yew-Soon Ong. Reducing item discrepancy via differentially private robust embedding alignment for privacy-preserving cross domain recommendation. In *Forty-first International Conference on Machine Learning*, 2024. 4
- [33] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017. 2
- [34] Yadan Luo, Zijian Wang, Zhuoxiao Chen, Zi Huang, and Mahsa Baktashmotlagh. Source-free progressive graph learning for open-set domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):11240–11255, 2023. 2
- [35] Shijie Ma, Fei Zhu, Zhun Zhong, Wenzhuo Liu, Xu-Yao Zhang, and Cheng-Lin Liu. Happy: A debiased learning framework for continual generalized category discovery. *Advances in Neural Information Processing Systems*, 37:50850–50875, 2024. 1, 2
- [36] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013. 6
- [37] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 6
- [38] Keon-Hee Park, Hakyung Lee, Kyungwoo Song, and Gyeong-Moon Park. Online continuous generalized category discovery. In *European Conference on Computer Vision*, pages 53–69. Springer, 2024. 1, 2, 5, 6, 8
- [39] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1406–1415, 2019. 6
- [40] Artem Rozantsev, Mathieu Salzmann, and Pascal Fua. Beyond sharing weights for deep domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 41(4):801–814, 2018. 2
- [41] Bardia Safaei, VS Vibashan, Celso M De Melo, and Vishal M Patel. Entropic open-set active learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 4686–4694, 2024. 8
- [42] Baochen Sun, Jiashi Feng, and Kate Saenko. Correlation alignment for unsupervised domain adaptation. In *Domain adaptation in computer vision applications*, pages 153–171. Springer, 2017. 2
- [43] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7492–7501, 2022. 1, 2, 6
- [44] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 6
- [45] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020. 2
- [46] Hongjun Wang, Sagar Vaze, and Kai Han. Hilo: A learning framework for generalized category discovery robust to domain shifts. *arXiv preprint arXiv:2408.04591*, 2024. 6
- [47] Hongjun Wang, Sagar Vaze, and Kai Han. Sptnet: An efficient alternative framework for generalized category discovery with spatial prompt tuning. *arXiv preprint arXiv:2403.13684*, 2024. 1, 2, 6
- [48] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7201–7211, 2022. 2
- [49] Xin Wen, Bingchen Zhao, and Xiaojuan Qi. Parametric classification for generalized category discovery: A baseline study. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16590–16600, 2023. 1, 2, 6
- [50] Sheng Zhang, Salman Khan, Zhiqiang Shen, Muzammal Naseer, Guangyi Chen, and Fahad Shahbaz Khan. Promptcal: Contrastive affinity learning via auxiliary prompts for generalized novel category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3479–3488, 2023. 2

- [51] Xinwei Zhang, Jianwen Jiang, Yutong Feng, Zhi-Fan Wu, Xibin Zhao, Hai Wan, Mingqian Tang, Rong Jin, and Yue Gao. Grow and merge: A unified framework for continuous categories discovery. *Advances in Neural Information Processing Systems*, 35:27455–27468, 2022. [1](#), [2](#), [6](#)
- [52] Bingchen Zhao and Kai Han. Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. *Advances in Neural Information Processing Systems*, 34:22982–22994, 2021. [2](#)
- [53] Zhun Zhong, Enrico Fini, Subhankar Roy, Zhiming Luo, Elisa Ricci, and Nicu Sebe. Neighborhood contrastive learning for novel class discovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10867–10875, 2021.
- [54] Zhun Zhong, Linchao Zhu, Zhiming Luo, Shaozi Li, Yi Yang, and Nicu Sebe. Openmix: Reviving known knowledge for discovering novel visual categories in an open world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9462–9470, 2021. [2](#)