This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

The Devil Is in Gradient Entanglement: Energy-Aware Gradient Coordinator for Robust Generalized Category Discovery

Haiyang Zheng[1] Nan Pu[2*] Yaqi Cai[1] Teng Long[1] Wenjing Li[2] Nicu Sebe[1] Zhun Zhong[2*]
[1]University of Trento, Italy [2]Hefei University of Technology, China
{haiyang.zheng, yaqi.cai, teng.long, niculae.sebe}@unitn.it
{n.pu, wenjingli, zhunzhong}@hfut.edu.cn

Abstract

*Generalized Category Discovery (GCD) leverages labeled data to categorize unlabeled samples from known or unknown classes. Most previous methods jointly optimize supervised and unsupervised objectives and achieve promising results. However, inherent optimization interference still limits their ability to improve further. Through quantitative analysis, we identify a key issue, i.e., **gradient entanglement**, which 1) distorts supervised gradients and weakens discrimination among known classes, and 2) induces representation-subspace overlap between known and novel classes, reducing the separability of novel categories. To address this issue, we propose the Energy-Aware Gradient Coordinator (EAGC), a plug-and-play gradient-level module that explicitly regulates the optimization process. EAGC comprises two components: Anchor-based Gradient Alignment (AGA) and Energy-aware Elastic Projection (EEP). AGA introduces a reference model to anchor the gradient directions of labeled samples, preserving the discriminative structure of known classes against the interference of unlabeled gradients. EEP softly projects unlabeled gradients onto the complement of the known-class subspace and derives an energy-based coefficient to adaptively scale the projection for each unlabeled sample according to its degree of alignment with the known subspace, thereby reducing subspace overlap without suppressing unlabeled samples that likely belong to known classes. Experiments show that EAGC consistently boosts existing methods and establishes new state-of-the-art results. Code is available at https://haiyangzheng.github.io/EAGC.*

1. Introduction

Category discovery—first introduced as Novel Category Discovery (NCD) [14] and later extended to Generalized Category Discovery (GCD) [63]—has recently emerged as an important problem in open-world visual learning. In GCD, a dataset is divided into a labeled subset containing known classes and an unlabeled subset that mixes known and novel categories. The goal is to leverage knowledge from the labeled subset to assign category labels to all samples in the unlabeled subset. Depending on how category labels are inferred, existing approaches can be broadly grouped into non-parametric methods [6, 10, 43, 49, 58, 59, 85, 88, 93], which perform clustering in the representation space, and parametric methods [1–3, 5, 16, 33, 37, 42, 65, 68, 73, 74, 77, 86], which learn explicit classifiers to jointly recognize known and novel categories.

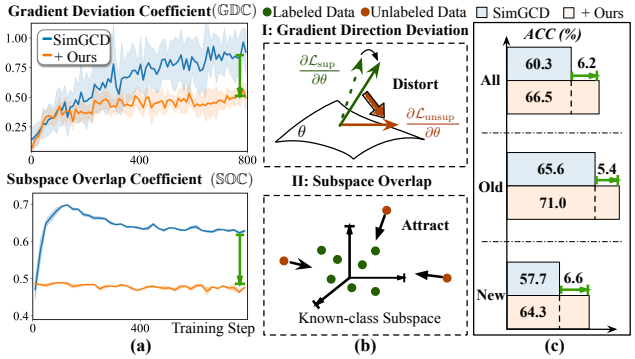


Figure 1. (a) 𝔾𝔻ℂ and 𝕊𝕆ℂ measure **Gradient Entanglement** (GE) from the gradient-direction and feature-subspace perspectives, respectively. Both indices reveal that existing GCD methods (e.g., SimGCD [74]) suffer from non-negligible optimization conflicts. (b) We attribute this issue to two main factors. **I: Gradient Direction Deviation.** Unsupervised gradients aim to discover novel categories but are inherently noisy and unreliable, which **distort** the optimization direction dictated by the supervised objective that focuses on known categories. **II: Subspace Overlap.** Noisy and unreliable unsupervised objectives produce highly divergent gradients, whereas gradients of the supervised objective remain relatively consistent. Joint optimization inevitably lets supervised gradients dominate, which progressively **attract** novel-class representations toward the known-class subspace. (c) Our method effectively mitigates GE, yielding substantial performance gains.

*Corresponding author.

While existing GCD methods have achieved promising results, many rely on a direct combination of supervised and unsupervised objectives, which often leads to an *unsatisfactory trade-off between known and novel classes*. To investigate this issue in depth, we define two quantitative metrics: *Gradient Deviation Coefficient (*$\mathbb{GDC}$*)* and *Subspace Overlap Coefficient (*$\mathbb{SOC}$*)* (see details in Sec. 3). $\mathbb{GDC}$ measures how unsupervised objectives interfere with the optimization direction of supervised gradients, whereas $\mathbb{SOC}$ quantifies the degree of overlap between the representation subspaces of known and novel classes. As shown in Fig. 1(a, b), we analyze the representative baseline SimGCD [74] and find that: **Issue** ① The $\mathbb{GDC}$ value gradually increases as training progresses, indicating that the optimization direction of supervised gradients is increasingly distorted by the unsupervised objectives, thereby weakening discrimination among known classes. **Issue** ② The $\mathbb{SOC}$ rises sharply in the early stage and remains at a high level afterward, suggesting that novel-class representations are progressively drawn toward the known-class subspace, which further reduces their separability. We refer to these under-explored issues in GCD as **Gradient Entanglement (GE)**.

To address these issues, we propose an Energy-Aware Gradient Coordinator (EAGC), which explicitly mitigates the **GE** problem from the perspective of gradient optimization, thereby consistently improving existing GCD methods (see Fig. 1(c)). EAGC comprises two complementary components: Anchor-based Gradient Alignment (AGA) and Energy-aware Elastic Projection (EEP). To address **Issue** ①, AGA leverages a reference model trained under pure supervision as an anchor. It explicitly aligns the gradient direction of labeled samples toward this reference, thereby preserving the discriminative structure of known classes against interference from the unsupervised objectives. To mitigate **Issue** ②, EEP first constructs a Conceptor-based subspace of known classes and projects the gradients of unlabeled samples onto its complement. Considering the GCD characteristic that *some unlabeled samples may also come from known categories*, EEP further introduces an energy-aware adaptive weighting strategy. This strategy determines the projection strength of each sample according to its feature energy ratio within the known-class subspace, preventing the over-suppression of known samples while effectively decoupling novel representations. **Notably**, EAGC operates in a plug-and-play fashion and integrates seamlessly with both parametric and non-parametric GCD frameworks without modifying the backbone architecture or the training objectives of the existing frameworks.

Our main contributions are summarized as:

- We quantitatively analyze the gradient behavior of existing GCD methods and identify a critical yet under-explored issue, *i.e.*, **Gradient Entanglement (GE)**, which substantially constrains their performance.
- We propose a **plug-and-play** gradient-level coordinator, **EAGC**, which effectively mitigates GE and can be seamlessly integrated into existing GCD methods.
- We **derive** two task-specific modules from **theoretical analysis**. Inspired by proximal regularization, AGA primarily preserves the discrimination of known categories. EEP adaptively enhances the separability of novel categories, motivated by an energy-aware subspace analysis.
- Extensive experiments demonstrate that EAGC consistently enhances various parametric and non-parametric GCD methods, achieving new state-of-the-art performance across multiple benchmarks.

2. Related Work

Category Discovery. Novel Category Discovery (NCD) was first introduced by Han et al. [14], who formulated the problem as transferable clustering, *i.e.*, transferring knowledge from labeled (seen) categories to cluster unlabeled (unseen) ones. Since then, extensive efforts [11–13, 15, 26, 29, 30, 34, 39, 46, 60, 70, 84, 87, 95, 96] have been devoted to improving representation learning and clustering reliability under this paradigm. To better reflect real-world scenarios, Generalized Category Discovery (GCD) [63] relaxes the NCD assumption by allowing unlabeled data to contain both known and unknown categories, inspiring a wide range of follow-up research. On one hand, several GCD studies investigate more practical settings, such as federated GCD [44, 71, 83], on-the-fly discovery [9, 32, 91], ultra-fine category discovery [35], multi-modal extensions [41, 57, 67, 78, 92], and domain-shift scenarios [53, 69, 72, 79, 82, 97], reflecting the growing interest in open-world recognition.

On the other hand, a large body of GCD work focuses on improving generalization, calibration, and clustering robustness within the standard GCD setting. Existing approaches can be broadly divided into **parametric** methods [1–3, 5, 16, 33, 37, 42, 65, 68, 73, 74, 77, 86], which rely on a classifier with a predefined number of classes, and **non-parametric** methods [6, 10, 43, 49, 58, 59, 85, 88, 93], which eliminate this constraint. Moreover, recent studies [7, 36, 62] propose plug-and-play modules that can be seamlessly integrated into existing frameworks to further enhance discovery performance. Following this line, our EAGC is also designed as a plug-and-play algorithm. Moreover, *distinct from prior efforts, EAGC analyzes the GCD challenge from an* ***optimization perspective*** *and identifies the inherent Gradient Entanglement issue. By effectively mitigating this problem, EAGC improves existing methods under both parametric and non-parametric paradigms, achieving new state-of-the-art performance.*

Gradient Projection. Gradient projection is a technique designed to regulate the direction of parameter updates by

projecting gradients onto constrained subspaces, thereby mitigating interference between tasks. This approach has been widely applied in multi-task learning [17, 80] (MTL) and continual learning [55] (CL) to reduce gradient conflicts and catastrophic forgetting. For instance, in MTL, Gradient Surgery [80] projects a given task's gradient onto the normal plane of any conflicting task gradient, thus alleviating harmful interference and improving overall optimization. In CL, GPM [55] uses singular value decomposition on the task-activations to identify a core gradient subspace, and then during subsequent task training forces new-task gradients to be orthogonal to that subspace—thereby preserving old-task knowledge. Additionally, PGP [45] integrates prompt-tuning with gradient projection, demonstrating that by ensuring prompt-gradient updates are orthogonal to previous feature subspaces, one can reduce forgetting while using parameter-efficient adaptation. *Unlike prior methods designed for fully supervised settings, our EAGC is* ***tailored*** *for the GCD setting, where 1) the model must learn jointly from both labeled and* ***unlabeled data****, and 2) the unlabeled subset uniquely contains not only known classes but also* ***novel ones****. These characteristics give rise to GCD-specific* ***gradient entanglement*** *issues.*

3. Preliminaries

Problem Setup. In GCD, we are given a labeled dataset $\mathcal{D}_L = \left\{(\mathbf{x}_i^l, y_i^l) \mid (\mathbf{x}_i^l, y_i^l) \in \mathcal{X} \times \mathcal{Y}_L\right\}_{i=1}^{N}$ and an unlabeled dataset $\mathcal{D}_U = \{\mathbf{x}_i^u \mid \mathbf{x}_i^u \in \mathcal{X}\}_{i=1}^{M}$. Here, N and M denote the numbers of labeled and unlabeled samples, respectively, and $\mathcal{Y}_L$ represents the label space of known classes. The unlabeled dataset is drawn from a broader label space $\mathcal{Y}_U$, which contains both known and unknown categories, i.e., $\mathcal{Y}_L \subset \mathcal{Y}_U$. The goal of GCD is to automatically categorize unlabeled samples, distinguishing between known (old) and novel (unknown) classes by leveraging the knowledge learned from labeled data.

3.1. Gradient Entanglement in GCD

Definition. The overall learning objective in GCD is typically expressed as:

$$\mathcal{L}_{\text{GCD}} = \alpha\mathcal{L}_{\text{sup}}(\mathcal{D}_L) + \beta\mathcal{L}_{\text{unsup}}(\mathcal{D}_U), \tag{1}$$

where $\mathcal{L}_{\text{sup}}$ provides reliable supervision from labeled data, while $\mathcal{L}_{\text{unsup}}$ relies on noisy self-supervised or pseudo-label signals—particularly unreliable in the early stage of training. This supervision asymmetry can lead to two coupled effects: *(1) unreliable unsupervised gradients interfere with the optimization of $\mathcal{L}_{sup}$, and (2) dominant supervised gradients pull unlabeled representations toward the known-class subspace.* From the perspective of gradient optimization, we refer to this phenomenon as *Gradient Entanglement* (**GE**). Specifically, the gradient of $\mathcal{L}_{\text{GCD}}$ with respect to model parameters θ can be decomposed as:

$$g = \nabla_\theta\mathcal{L}_{\text{GCD}} = g_L + g_U, g_L = \alpha\nabla_\theta\mathcal{L}_{\text{sup}}, g_U = \beta\nabla_\theta\mathcal{L}_{\text{unsup}}. \tag{2}$$

First, the unsupervised gradient g_U may conflict with or distort the optimization direction of the supervised gradient g_L, thereby weakening the discriminative power of known classes and leading to suboptimal parameter updates. *Second*, since known classes contribute gradients from both labeled supervision (g_L) and the known-class portion of the unlabeled data (g_U^{known}), these gradients tend to dominate the optimization dynamics and pull novel representations toward the known-class subspace, thereby reducing the inter-class separability of novel categories.

Observation ①. Gradient Direction Deviation. Let $\hat{g_L} = \nabla_\theta\mathcal{L}_{\text{sup}}$ denote the reference gradient obtained when optimizing only $\mathcal{L}_{\text{sup}}$ (i.e., setting $\beta = 0$ in Eq. (1)). To quantify the effect of unlabeled data on the supervised optimization dynamics, we define the *Gradient Deviation Coefficient (*$\mathbb{GDC}$*)* as:

$$\mathbb{GDC} = 1 - \frac{\langle \hat{g_L}, g\rangle}{\|\hat{g_L}\|\,\|g\|}. \tag{3}$$

It ranges within $[0, 2]$ and quantifies how unlabeled data interferes with the supervised optimization signal—larger values indicate a more severe GE issue.

Observation ②. Subspace Overlap. Let the feature matrix of labeled (known-class) data be $Z_{\text{old}} \in \mathbb{R}^{N\times d}$ and that of novel classes (from unlabeled data) be $Z_{\text{new}} \in \mathbb{R}^{M_n\times d}$, where M_n denotes the number of novel samples and d is the feature dimension. Let U_k denote the top-k principal components of Z_{old}, and define the projection matrix as $P_{\text{old}} = U_k U_k^\top$, which captures the dominant energy directions of known classes in the feature space. To measure the representation subspace overlap between novel and known classes, we define the *Subspace Overlap Coefficient (*$\mathbb{SOC}$*)* as:

$$\mathbb{SOC} = \frac{\|Z_{\text{new}} P_{\text{old}}\|_F^2}{\|Z_{\text{new}}\|_F^2}. \tag{4}$$

It ranges within $[0, 1]$ and quantifies the proportion of novel-class feature energy projected onto the subspace of known classes—a higher $\mathbb{SOC}$ indicates a stronger tendency of novel features to align with the known-class subspace.

4. Method

Motivation. To address Gradient Entanglement, our motivation is twofold. First, to stabilize the optimization of labeled samples against disturbances from unlabeled objectives, we seek to *anchor the gradient direction of labeled data* and maintain a consistent optimization trajectory for known classes. Second, to mitigate the attraction of novel representations toward the known-class subspace, we further aim to *project the gradient direction of unlabeled data*

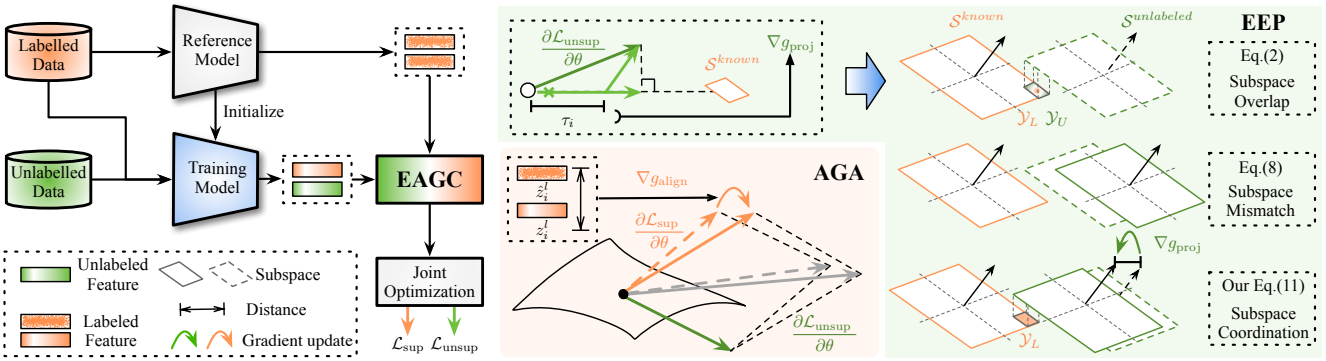


Figure 2. Framework Overview of the proposed Energy-Aware Gradient Coordinator (**EAGC**). EAGC consists of two components: (1) *Anchor-based Gradient Alignment* (**AGA**), which uses feature anchors from a reference model to generate the alignment term ∇g_{align} to stabilize the updates of $\mathcal{L}_{\text{sup}}$, and (2) *Energy-aware Elastic Projection* (**EEP**), which generates the projection term ∇g_{proj} by projecting the unlabeled-gradient component $\nabla \mathcal{L}_{\text{unsup}}$ onto the complement of the known-class subspace $\mathcal{S}^{\text{known}}$ using a sample-adaptive elastic weight τ.

away from the dominant directions of known categories, thereby enhancing their separability.

Framework Overview. We propose an Energy-Aware Gradient Coordinator (EAGC) module to explicitly mitigate Gradient Entanglement in GCD from the perspective of gradient optimization. As illustrated in Fig. 2, EAGC comprises two complementary components designed to address the two manifestations of Gradient Entanglement. (1) *Anchor-based Gradient Alignment (AGA)* focuses on stabilizing the optimization of labeled data. It anchors the gradient direction of labeled samples toward a reference model trained purely on labeled data, thereby preserving stable supervision against disturbances from unlabeled objectives. (2) *Energy-aware Elastic Projection (EEP)* operates on unlabeled data. It performs an elastic soft projection of unlabeled gradients away from the known-class subspace, with an adaptive weighting mechanism that prevents over-suppression of known samples while improving the separability of novel ones. Together, these two components act in a complementary manner—AGA preserves the discriminative boundary of known classes, while EEP encourages novel representations to evolve in a disentangled subspace, resulting in more stable and balanced GCD optimization.

Notation. During backpropagation, the parameter gradient $\nabla_\theta \mathcal{L}$ is computed via the chain rule, $\nabla_\theta \mathcal{L} = \frac{\partial z}{\partial \theta} \cdot \nabla_z \mathcal{L}$, where $\nabla_z \mathcal{L} = \frac{\partial \mathcal{L}}{\partial z}$ denotes the gradient of the loss with respect to the intermediate feature representation z. In our EAGC, we regulate gradient alignment and projection by modulating this feature-level gradient $\nabla_z \mathcal{L}$ before it is propagated back to the encoder parameters.

4.1. Anchor-based Gradient Alignment

To mitigate the interference of unlabeled objectives on the optimization direction of labeled samples during joint training, we introduce a gradient alignment constraint that maintains stable and discriminative learning for known classes.

Reference Model. Given the labeled subset $\mathcal{D}_L$, we first train a reliable *reference model* using the supervised classification objective:

$$\mathcal{L}_{\text{cls}} = \frac{1}{|\mathcal{B}_L|} \sum_{i \in \mathcal{B}_L} \ell(y_i^l, p_i), \tag{5}$$

where $\mathcal{B}_L$ denotes a mini-batch from $\mathcal{D}_L$, $\ell(\cdot)$ is the cross-entropy loss, p_i is the predicted probability distribution of the labeled input x_i^l, and y_i^l is its corresponding label. Specifically, $p_i = \sigma(f(\mathcal{E}(x_i^l))/\tau_s)$, where $\mathcal{E}(\cdot)$ denotes the image encoder, $f(\cdot)$ is a parametric classifier, σ is the softmax function, and τ_s is a temperature parameter. The resulting *reference model*, denoted as $\mathcal{E}_r(\cdot)$, is fixed during GCD training to provide stable feature anchors for known classes.

Gradient Alignment. During GCD optimization with the objective $\mathcal{L}_{\text{GCD}}$ (Eq. (1)), we impose a gradient alignment constraint on labeled samples:

$$\nabla_z \mathcal{L}_{\text{GCD}} = \nabla_{z^l} \mathcal{L}_{\text{sup}} + \underbrace{\nabla_{z^u} \mathcal{L}_{\text{unsup}}}_{\text{Disturbance}} + \underbrace{\lambda_a (z^l - \hat{z}^l)}_{\text{Stabilization}}, \tag{6}$$

where z^l denotes the feature representations of labeled samples, $z_i^l = \mathcal{E}(x_i^l)$, and $\hat{z}_i^l = \mathcal{E}_r(x_i^l)$. Here, $\nabla_{z^l} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial z^l}$ represents the gradient of the loss function with respect to z^l, and λ_a controls the strength of the alignment correction. The alignment term $\nabla_{z^l} g_{\text{align}} = \lambda_a (z^l - \hat{z}^l)$ anchors the gradient direction of labeled samples toward that of the reference model, thereby stabilizing the optimization against disturbances introduced by $\mathcal{L}_{\text{unsup}}$.

Remark. The stabilization term can be interpreted as a form of proximal regularization [40, 81] in the feature space, which effectively defines a local trust region around the supervised optimum $\hat{z}$. This induces a contraction effect that stabilizes the representations of labeled samples and mitigates semantic drift. From the objective perspective, we further verify that such proximal regularization consistently enhances GCD performance, as reported in Appendix C.

4.2. Energy-aware Elastic Projection

During joint optimization, the strong supervision from labeled data tends to generate dominant gradients that steer learning toward known categories, consequently pulling novel representations into the known-class subspace and hindering the formation of distinct clusters. To mitigate this subspace overlap, we explicitly project unlabeled gradients away from the known-class subspace.

Known-Class Subspace Construction. Our goal is to construct a representation subspace for known classes such that novel representations can be projected into regions of the feature space not dominated by known categories. While PCA provides a hard low-rank approximation by retaining only the principal energy directions, it tends to over-compress the feature space and discard low-energy components that may encode subtle semantic variations distinguishing known and novel classes. To address this, we adopt a Conceptor-based [19] representation subspace that models the principal directions of known classes in an energy-weighted and flexible manner. Given the feature matrix of the labeled subset $Z_{\text{old}} = \{\mathcal{E}_r(\mathbf{x}_i^l)\} \in \mathbb{R}^{N\times d}$, the known-class subspace is computed as:

$$\mathcal{S}^{known} = R(R+\eta^{-2}I)^{-1}, \quad R = \frac{1}{N} Z_{\text{old}}^\top Z_{\text{old}}, \tag{7}$$

where η is the *aperture* controlling the softness of subspace capture and I is the identity matrix. The resulting $\mathcal{S}^{known}$ serves as a soft projection operator that preserves the dominant energy directions of known classes. Given a feature vector z_i, its projected component within the known-class subspace is $z_{\text{proj}} = z_i \mathcal{S}^{known}$.

Gradient Projection. For unlabeled samples, we regulate the feature-level gradients to reduce their overlap with the known-class subspace. Let $\nabla_{z^u}\mathcal{L}_{\text{unsup}}$ denote the gradient of the unsupervised loss with respect to unlabeled feature representations z^u. We apply a soft projection defined as:

$$\nabla_{z^u}\mathcal{L}_{\text{unsup}} = \nabla_{z^u}\mathcal{L}_{\text{unsup}} - \lambda_p\big(\nabla_{z^u}\mathcal{L}_{\text{unsup}}\mathcal{S}^{known}\big), \tag{8}$$

where λ_p controls the overall projection strength. However, since the unlabeled set also contains samples from known categories, projecting all unlabeled gradients away from the known-class subspace would undesirably *suppress the learning of known-category samples*.

Energy-Aware Adaptive Weighting. To address this issue, we introduce a sample-specific projection weight determined by each feature's energy ratio within the known-class subspace:

$$\mathbf{E}_{\text{old}}(z_i) = \frac{z_i^\top \mathcal{S}^{known} z_i}{\|z_i\|_2^2} \in [0,1], \tag{9}$$

which measures the proportion of the feature's squared magnitude (i.e., feature energy) residing in the known-class subspace. We then compute the average energy ratio of labeled data, $\overline{\mathbf{E}_{\text{old}}^l}$, and define the adaptive projection weight for each unlabeled feature as:

$$\tau_i = 1 - \frac{\mathbf{E}_{\text{old}}(z_i^u)}{\overline{\mathbf{E}_{\text{old}}^l}}. \tag{10}$$

This normalization calibrates the projection strength relative to the alignment of labeled samples: features that are highly aligned with the known subspace (likely known) receive weaker projection, whereas those less aligned (likely novel) are projected more strongly. The resulting Elastic Gradient Projection is thus formulated as:

$$\nabla_{z^u} g_{\text{proj}} = -\lambda_p\big(\tau \cdot \nabla_{z^u}\mathcal{L}_{\text{unsup}}\mathcal{S}^{known}\big). \tag{11}$$

Remark. Our design exhibits two key properties. (1) The known-class representation subspace is constructed using Conceptor theory [19], which provides a soft, energy-weighted characterization of dominant feature directions. Unlike hard low-rank truncations such as PCA, this formulation offers greater flexibility in modeling the boundary between known and novel categories. (2) Because the unlabeled set contains both known and novel samples, we employ an *Energy-aware Elastic Projection* that adaptively modulates the projection strength based on each sample's alignment with the known-class subspace. This sample-specific adjustment avoids over-suppressing known-category gradients while effectively decoupling novel representations.

4.3. Plug-and-Play Training

We integrate the proposed EAGC module as a plug-and-play component that operates purely at the gradient level, requiring no modification to either the network architecture or the original training objectives. Under EAGC, the overall feature-level gradient update can be formulated as:

$$\nabla_z \mathcal{L}_{\text{GCD}} = \nabla_z \mathcal{L}_{\text{sup}} + \nabla_z \mathcal{L}_{\text{unsup}} + \nabla_{z^l} g_{\text{align}} + \nabla_{z^u} g_{\text{proj}}. \tag{12}$$

The complete training pipeline of EAGC is summarized in Algorithm 1.

5. Experiments

5.1. Experimental Setup

Datasets. We evaluate our method across multiple datasets, including two generic benchmarks, CIFAR-100 [21] and ImageNet-100 [8], as well as the Semantic Shift Benchmark (SSB) [64], which comprises three fine-grained datasets: CUB-200-2011 [66], Stanford Cars [20], and FGVC-Aircraft [38]. Following the protocol of [63], a subset of categories is designated as the labeled set $\mathcal{Y}_L$. Within these classes, 50% of the images are used to form the labeled subset $\mathcal{D}_L$, while the remaining images constitute the unlabeled subset $\mathcal{D}_U$. Dataset statistics are provided in Appendix A.

Algorithm 1: Energy-Aware Gradient Coordinator

Input: Labeled data $\mathcal{D}_L$, unlabeled data $\mathcal{D}_U$, image encoder $\mathcal{E}(\cdot)$, projection head $\phi(\cdot)$, hyperparameters $(\lambda_a, \lambda_p, \eta)$.

/* *Getting reference model* */
Train $\mathcal{E}(\cdot)$ on $\mathcal{D}_L$ using $\mathcal{L}_{cls}$ in Eq. (5) to obtain $\mathcal{E}_r$;
Freeze $\mathcal{E}_r$ to serve as an anchor for labeled data;

/* *Known-class Subspace Construction* */
Compute Z_{old} and $\mathcal{S}^{known}$ using Eq. (7);
Compute the average energy ratio $\overline{\mathbf{E}^l_{\text{old}}}$;

/* *GCD Training with EAGC* */
for $epoch = 1$ ***to*** T **do**
  for *batch* $(x, y, mask_{lab})$ *from* $\mathcal{D}_L \cup \mathcal{D}_U$ **do**
    /* *Anchor-based Gradient Alignment* */
    $z = \mathcal{E}(x), \hat{z} = \mathcal{E}_r(x)$;
    Register backward hook for alignment term: $mask_{\text{lab}} \cdot \lambda_a(z - \hat{z})$;
    /* *Forward pass* */
    Obtain $h = \phi(z)$ and compute $\mathcal{L}_{\text{GCD}} = \alpha\mathcal{L}_{\text{sup}} + \beta\mathcal{L}_{\text{unsup}}$;
    /* *Energy-aware Gradient Projection* */
    Compute projection weights τ via Eq. (10);
    Register backward hook for projection term: $-(1 - mask_{\text{lab}}) \cdot \lambda_p(\tau \cdot \nabla_z \mathcal{L}_{\text{unsup}} \mathcal{S}^{known})$;
    /* *Backward Pass* */
    Update $\mathcal{E}(\cdot)$ and $\phi(\cdot)$ via an optimizer step;
  end
end

Evaluation Metric. Following [63], we report clustering accuracy (ACC), where predicted clusters are matched to ground-truth labels using the Hungarian algorithm.

Implementation Details. We evaluate our EAGC module on both parametric baselines (SimGCD [74], LegoGCD [3], and SPTNet [68]) and the non-parametric baseline SelEx [50]. Unless otherwise specified, experiments adopt a ViT-B/16 backbone pretrained with DINO [4]. All baselines are trained for 200 epochs with a batch size of 128. The learning rate is set to 0.1 for the backbone and 1.0 for the projection head, both scheduled by cosine decay. The reference model follows the same trainable-layer configuration as the baselines and is trained for 30 epochs (only 3 for CIFAR-100 to prevent overfitting on known classes due to its 80/20 split) with τ_s = 0.1 and a cosine-decayed learning rate of 0.02. Other training hyperparameters follow the default settings of the respective baselines. Our EAGC introduces an additional hyperparameter η = 2.0, which controls the aperture of the conceptor. For all baselines and datasets, we uniformly set (λ_a, λ_p) = (0.7, 0.5), which are tuned using SimGCD on the CUB dataset. All experiments are conducted on NVIDIA 3090Ti GPUs, and results are averaged over three runs with different random seeds.

5.2. Quantitative Comparison

We compare our method against a broad range of GCD approaches, including ORCA [2], GCD [63], GPC [88], XCon [10], PromptCAL [85], DCCL [43], μGCD [65], InfoSieve [49], AMEND [1], and CMS [6], as well as baselines SimGCD [74], LegoGCD [3], SPTNet [68], and SelEx [50]. The results are reported in Tab. 1. Across four baselines and five datasets, integrating EAGC improves performance overall, yielding average gains of 3.2% in All ACC and 4.3% in New ACC. Notably, when built upon SelEx, EAGC yields substantial improvements of 9.6% and 6.7% in All ACC on CUB and ImageNet-100, respectively.

Results on Fine-grained Datasets. The results on fine-grained datasets are presented on the left side of Tab. 1. Across the three fine-grained benchmarks, incorporating EAGC improves category discovery accuracy overall. On average across the three fine-grained benchmarks, EAGC brings gains of 4.6% in All ACC and 6.1% in New ACC. In particular, SelEx+EAGC achieves state-of-the-art performance on all three datasets, with average improvements of 8.5% in All ACC and 11.4% in New ACC, demonstrating that EAGC substantially enhances the capability of SelEx.

Results on Generic Datasets. The results on generic datasets are presented on the right side of Tab. 1. Across the two generic benchmarks, incorporating EAGC yields average gains of 1.3% in All ACC and 1.4% in New ACC. The best results on CIFAR-100 and ImageNet-100 are achieved by SPTNet+EAGC and SelEx+EAGC, respectively. These results demonstrate that EAGC, as a plug-and-play module, effectively improves existing GCD baselines and leads to more accurate category discovery.

Additional Results. Using DINOv2 as the backbone, EAGC consistently improves two baselines across three fine-grained datasets, with average gains of 6.2% in All ACC and 8.3% in New ACC. Results on the more challenging Herbarium19 dataset are provided in Appendix B.

5.3. Ablation Study

To evaluate the effectiveness of the AGA and EEP components in our EAGC framework, we perform ablation studies on both parametric and non-parametric baselines using the CUB and Stanford Cars datasets. The results are reported in Tab. 3. **(1) AGA.** By anchoring the gradient updates of labeled samples, AGA effectively stabilizes the optimization of known categories. Incorporating AGA (variant I) yields average Old ACC improvements of 4.8% for SimGCD and 3.0% for SelEx. Interestingly, AGA can also improve the discovery of novel classes by providing more stable and transferable feature representations. **(2) EEP.** By

Table 1. Comparison with state-of-the-art GCD methods. Best and second-best results are **bold** and underlined, respectively.

Method	CUB			Stanford-Cars			Aircraft			CIFAR-100			ImageNet-100			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
ORCA [2]	36.3	43.8	32.6	31.6	32.0	31.4	31.9	42.2	26.9	73.5	**92.6**	63.9	81.8	86.2	79.6	51.0	59.4	46.9
GCD [63]	51.3	56.6	48.7	39.0	57.6	29.9	45.0	41.1	46.9	73.0	76.2	66.5	74.1	89.8	66.3	56.5	64.3	51.7
GPC [88]	52.0	55.5	47.5	38.2	58.9	27.4	43.3	40.7	44.8	75.4	84.6	60.1	75.3	93.4	66.7	56.8	66.6	49.3
XCon [10]	52.1	54.3	51.0	40.5	58.8	31.7	47.7	44.4	49.4	74.2	81.2	60.3	77.6	93.5	69.7	58.4	66.4	52.4
PromptCAL [85]	62.9	64.4	62.1	50.2	70.1	40.6	52.2	52.2	52.3	81.2	84.2	75.3	83.1	92.7	78.3	65.9	72.7	61.7
DCCL [43]	63.5	60.8	64.9	43.1	55.7	36.2	–	–	–	75.3	76.8	70.2	80.5	90.5	76.2	–	–	–
μGCD [65]	65.7	68.0	64.6	56.5	68.1	50.9	53.8	55.4	53.0	–	–	–	–	–	–	–	–	–
InfoSieve [49]	69.4	**77.9**	65.2	55.7	74.8	46.4	56.3	63.7	52.5	78.3	82.2	70.5	80.5	93.8	73.8	68.0	78.5	61.7
AMEND [1]	64.9	75.6	59.6	56.4	73.3	48.2	52.8	61.8	48.3	81.0	79.9	83.3	83.2	92.9	78.3	67.7	76.7	63.5
CMS [6]	68.2	76.5	64.0	56.9	76.1	47.6	56.0	63.4	52.3	82.3	85.7	75.5	84.7	95.6	79.2	69.6	79.5	63.7
SimGCD [74]	60.3	65.6	57.7	53.8	71.9	45.0	54.2	59.1	51.8	80.1	81.2	77.8	83.0	93.1	77.9	66.3	74.2	62.0
+ EAGC	66.5	71.0	64.3	62.9	76.0	56.6	57.7	60.4	56.3	83.1	84.1	81.0	83.5	93.9	78.3	70.7	77.1	67.3
LegoGCD [3]	63.8	71.9	59.8	57.3	75.7	48.4	55.0	61.5	51.7	81.8	81.4	82.5	86.3	94.5	82.1	68.8	77.0	64.9
+ EAGC	64.6	72.5	60.7	62.8	76.2	56.3	56.7	61.9	54.0	83.5	83.7	83.0	84.8	94.4	80.1	70.5	77.7	66.8
SPTNet [68]	65.8	68.8	65.1	59.0	79.2	49.3	59.3	61.8	58.1	81.3	84.3	75.6	85.4	93.2	81.4	70.2	77.5	65.9
+ EAGC	67.4	70.5	66.0	62.1	76.1	55.4	57.0	58.1	56.5	**84.0**	84.3	**83.5**	85.3	94.3	80.4	71.2	76.7	68.4
SelEx [50]	73.6	75.3	72.8	58.5	75.6	50.3	57.1	64.7	53.3	82.3	85.3	76.3	83.1	93.6	77.8	70.9	78.9	66.1
+ EAGC	**83.2**	73.9	**87.9**	**65.7**	**83.7**	**57.0**	**65.7**	**65.6**	**65.7**	79.3	84.1	69.7	**89.8**	**96.2**	**86.5**	**76.7**	**80.7**	**73.4**
Avg. △	+4.6	+1.6	+5.9	+6.2	+2.4	+8.1	+2.9	-0.3	+4.4	+1.1	+1.0	+1.3	+1.4	+1.1	+1.5	+3.2	+1.2	+4.3

Table 2. Comparison with DINOv2 as the backbone.

Method	CUB			Stanford Cars			Aircraft		
	All	Old	New	All	Old	New	All	Old	New
SimGCD	71.5	78.1	68.3	71.5	81.9	66.6	63.9	69.9	60.9
+ EAGC	**77.9**	**79.9**	**76.9**	**81.5**	**90.9**	**76.9**	**74.6**	**74.9**	**74.4**
SelEx	87.4	**85.1**	88.5	82.2	**93.7**	76.7	79.8	**82.3**	78.6
+ EAGC	**89.4**	82.3	**92.9**	**85.3**	92.7	**81.7**	**84.9**	81.8	**86.5**
Avg. Δ	+4.2	-0.5	+6.5	+6.6	+4.0	+7.7	+7.9	+2.3	+10.7

Table 3. Ablation study of the proposed components. *AGA* denotes Anchor-based Gradient Alignment, and *EEP* denotes Energy-aware Elastic Projection.

Method	*AGA*	*EEP*	CUB			Stanford-Cars		
			All	Old	New	All	Old	New
SimGCD	✗	✗	60.3	65.6	57.7	53.8	71.9	45.0
I	✓	✗	65.6	**71.1**	62.9	62.2	75.9	55.6
II	✗	✓	65.2	**71.1**	62.2	57.7	74.6	49.6
+ EAGC	✓	✓	**66.5**	71.0	**64.3**	**62.9**	**76.0**	**56.6**
SelEx	✗	✗	73.6	75.3	72.8	58.5	75.6	50.3
I	✓	✗	81.7	73.5	85.9	62.6	83.4	52.6
II	✗	✓	82.3	71.9	87.6	58.8	81.5	47.9
+ EAGC	✓	✓	**83.2**	**73.9**	**87.9**	**65.7**	**83.7**	**57.0**

decoupling known and novel category subspaces at the gradient level, EEP (variant II) enhances the discovery accuracy of novel categories, achieving average New ACC gains of 4.6% for SimGCD and 6.2% for SelEx across the two datasets. In addition, the reduced subspace interference also benefits known-class recognition, yielding an average Old ACC gain of 2.7%. **(3) Full EAGC.** Combining AGA and EEP under the respective optimization objectives $\mathcal{L}_{\text{sup}}$ and $\mathcal{L}_{\text{unsup}}$ achieves the best overall results. Compared with variant I, EAGC improves All ACC by 1.6% on average across the two datasets, and compared with variant II, it brings an additional 3.6% gain in All ACC. These results demonstrate that AGA and EEP act in a complementary manner, jointly improving both known- and novel-class performance.

5.4. Gradient Entanglement Analysis

We evaluate the effectiveness of EAGC in mitigating Gradient Entanglement, as summarized in Tab. 4. The *Gradient Deviation Coefficient* ($\mathbb{GDC}$) measures how unlabeled optimization distorts the gradient direction of labeled samples, whereas the *Subspace Overlap Coefficient* ($\mathbb{SOC}$) quantifies the feature subspace overlap between known and novel categories. For fair comparison, both the baseline and our method adopt the same model configuration, and all metrics are computed as the average over the first 200 training steps. Across all benchmarks, integrating EAGC consistently lowers both $\mathbb{GDC}$ and $\mathbb{SOC}$, highlighting its effectiveness in decoupling the optimization dynamics of labeled and unlabeled data. For the parametric baseline SimGCD, EAGC achieves average reductions of 46.8% in $\mathbb{GDC}$ and 21.4% in $\mathbb{SOC}$, while for the non-parametric baseline SelEx, the reductions are 98.5% and 18.1%, respectively.

Table 4. Quantitative comparison of the Gradient Deviation Coefficient ($\mathbb{GDC}$) and Subspace Overlap Coefficient ($\mathbb{SOC}$).

Method	CUB		Stanford Cars		Aircraft	
	$\mathbb{GDC}\downarrow$	$\mathbb{SOC}\downarrow$	$\mathbb{GDC}\downarrow$	$\mathbb{SOC}\downarrow$	$\mathbb{GDC}\downarrow$	$\mathbb{SOC}\downarrow$
SimGCD [74]	0.2669	0.6478	0.0909	0.8110	0.0935	0.8643
+EAGC	**0.1525**	**0.4923**	**0.0500**	**0.6208**	**0.0375**	**0.7139**
LegoGCD [3]	0.2646	0.6468	0.0905	0.8112	0.0922	0.8642
+EAGC	**0.1594**	**0.4844**	**0.0515**	**0.6135**	**0.0376**	**0.7112**
SelEx [50]	0.0005	0.5902	0.0007	0.6022	0.0313	0.6033
+EAGC	**0.0003**	**0.4463**	**0.0001**	**0.4470**	**0.0001**	**0.5774**

5.5. Evaluation

Gradient Projection Strategy Evaluation. Our Energy-aware Elastic Projection (EEP) adaptively assigns projection weights to unlabeled samples based on their energy proportions within the known-class subspace. We evaluate this design in Tab. 5, using the best-performing baseline SelEx as the base model. The variant *w/o Energy Adaptation* uniformly projects all unlabeled gradients away from the known-class subspace without considering sample-specific energy alignment. This uniform projection results in inferior performance, with an average drop of 4.2% in All ACC. In particular, the accuracy on known classes decreases by 4.7% on average, indicating that indiscriminate gradient projection hampers the learning of known samples within the unlabeled set.

Table 5. Evaluation of the gradient projection strategy based on SelEx. "w/o EEP" indicates a variant that uniformly projects all unlabeled gradients away from the known-class subspace without applying energy-aware weighting.

Method	CUB			Stanford Cars			Aircraft		
	All	Old	New	All	Old	New	All	Old	New
w/o EEP	80.9	**74.5**	84.2	59.3	76.6	50.9	61.7	57.9	63.6
Ours	**83.2**	73.9	**87.9**	**65.7**	**83.7**	**57.0**	**65.7**	**65.6**	**65.7**

Known-class Representation Subspace Evaluation. In our Energy-aware Elastic Projection, we employ a Conceptor-based soft subspace instead of a PCA-based hard subspace. While PCA captures only the top-K principal components, it overlooks low-energy directions that may encode subtle discriminative cues between known and novel classes. We evaluate this design choice in Tab. 6, where the number of principal components K varies from 16 to 64. All PCA-based variants yield inferior overall results compared with our Conceptor-based formulation. In particular, ours consistently improves performance on novel classes, surpassing the PCA-16 variant by an average of 1.2%.

Table 6. Comparison of subspace construction strategies. "PCA-K" refers to a hard subspace built from the top-K principal components of labeled data.

Method	CUB			Stanford Cars		
	All	Old	New	All	Old	New
PCA-16	82.7	**74.5**	86.8	64.9	**84.0**	55.7
PCA-32	80.9	73.5	84.7	64.5	84.1	55.0
PCA-64	82.1	73.7	86.3	64.0	83.3	54.6
Ours	**83.2**	73.9	**87.9**	**65.7**	83.7	**57.0**

5.6. Hyperparameter Analysis

Aperture. When constructing the known-class subspace, the aperture parameter controls the *softness* of the subspace—larger aperture values correspond to softer subspaces that capture a broader range of directions. This is a shared hyperparameter across the entire EAGC framework. We determine the aperture based on experiments with the parametric baseline SimGCD on CUB and keep it fixed for all other datasets and baselines. As shown in Fig. 3 (a), the performance remains stable across a wide range of aperture values (0.5–4.0). We set $\eta = 2.0$ in all experiments.

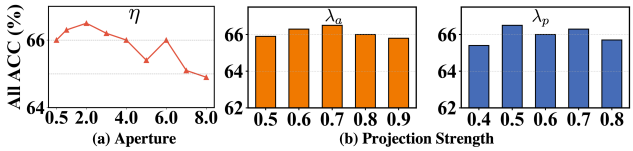


Figure 3. Effects of the aperture η and projection-strength hyperparameters λ_a and λ_p.

Projection Strength. Our EAGC introduces two hyperparameters, λ_a and λ_p, which control the projection strengths of AGA and EEP, respectively. To avoid over-tuning, we select them on CUB using the SimGCD baseline and keep them fixed for all baselines and datasets. The selection process is illustrated in Fig. 3 (b) and (c). Overall, both λ_a and λ_p are relatively stable. We set $\lambda_a = 0.7$ and $\lambda_p = 0.5$.

6. Conclusion

In this work, we identify and quantify a previously underexplored bottleneck in GCD, termed Gradient Entanglement, which fundamentally limits existing methods. To tackle this issue, we propose the Energy-Aware Gradient Coordinator (EAGC), a plug-and-play gradient-level framework that stabilizes labeled gradients to preserve discriminative directions and adaptively decouples unlabeled gradients from the known-class subspace. Extensive experiments across multiple benchmarks show that EAGC consistently improves diverse GCD baselines and achieves new state-of-the-art performance. We expect EAGC to offer a new perspective for advancing GCD and to serve as a general strategy for future GCD methods.

Acknowledgements

This work was funded by the National Natural Science Foundation of China (No. 62572166, No. 62402157), and the Fundamental Research Funds for the Central Universities (No. JZ2025HGTB0219). This work was also supported by the EU Horizon projects ELIAS (No. 101120237) and ELLIOT (No. 101214398), as well as the FIS project GUIDANCE (No. FIS2023-03251). We further acknowledge CINECA and the ISCRA initiative for providing high-performance computing resources.

References

[1] Anwesha Banerjee, Liyana Sahir Kallooriyakath, and Soma Biswas. Amend: Adaptive margin and expanded neighborhood for efficient generalized category discovery. In *WACV*, 2024. 1, 2, 6, 7

[2] Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-world semi-supervised learning. In *ICLR*, 2022. 6, 7

[3] Xinzi Cao, Xiawu Zheng, Guanhong Wang, Weijiang Yu, Yunhang Shen, Ke Li, Yutong Lu, and Yonghong Tian. Solving the catastrophic forgetting problem in generalized category discovery. In *CVPR*, 2024. 1, 2, 6, 7, 8

[4] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, 2021. 6

[5] Florent Chiaroni, Jose Dolz, Ziko Imtiaz Masud, Amar Mitiche, and Ismail Ben Ayed. Parametric information maximization for generalized category discovery. In *ICCV*, 2023. 1, 2

[6] Sua Choi, Dahyun Kang, and Minsu Cho. Contrastive mean-shift learning for generalized category discovery. In *CVPR*, 2024. 1, 2, 6, 7

[7] Qiyuan Dai, Hanzhuo Huang, Yu Wu, and Sibei Yang. Adaptive part learning for fine-grained generalized category discovery: A plug-and-play enhancement. In *CVPR*, 2025. 2

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*. Ieee, 2009. 5, 2

[9] Ruoyi Du, Dongliang Chang, Kongming Liang, Timothy Hospedales, Yi-Zhe Song, and Zhanyu Ma. On-the-fly category discovery. In *CVPR*, 2023. 2

[10] Yixin Fei, Zhongkai Zhao, Siwei Yang, and Bingchen Zhao. Xcon: Learning with experts for fine-grained category discovery. *arXiv preprint arXiv:2208.01898*, 2022. 1, 2, 6, 7

[11] Wei Feng, Lie Ju, Lin Wang, Kaimin Song, and Zongyuan Ge. Towards novel class discovery: A study in novel skin lesions clustering. In *MICCAI*, 2023. 2

[12] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *ICCV*, 2021.

[13] Peiyan Gu, Chuyu Zhang, Ruijie Xu, and Xuming He. Class-relation knowledge distillation for novel class discovery. In *ICCV*, 2023. 2

[14] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. In *ICCV*, 2019. 1, 2

[15] Zahid Hasan, Abu Zaher Md Faridee, Masud Ahmed, Sanjay Purushotham, Heesung Kwon, Hyungtae Lee, and Nirmalya Roy. Novel categories discovery via constraints on empirical prediction statistics. *arXiv preprint arXiv:2307.03856*, 2023. 2

[16] Zhenqi He, Yuanpei Liu, and Kai Han. Seal: Semantic-aware hierarchical learning for generalized category discovery. In *NeurIPS*, 2025. 1, 2

[17] Chengyue Huang, Junjiao Tian, Brisa Maneechotesuwan, Shivang Chopra, and Zsolt Kira. Directional gradient projection for robust fine-tuning of foundation models. *ICLR*, 2025. 3

[18] Ziming Huang, Xurui Li, Haotian Liu, Feng Xue, Yuzhe Wang, and Yu Zhou. Anomalyncd: Towards novel anomaly class discovery in industrial scenarios. In *CVPR*, 2025. 6

[19] Herbert Jaeger. Controlling recurrent neural networks by conceptors. *arXiv preprint arXiv:1403.3369*, 2014. 5

[20] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *CVPRW*, 2013. 5, 2

[21] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 5, 2

[22] Yiming Lei, Qiannan Shen, and Junhao Song. A multi-scale graph learning framework with temporal consistency constraints for financial fraud detection in transaction networks under non-stationary conditions. *arXiv preprint arXiv:2603.14592*, 2026. 6

[23] Bingyu Li, Feiyu Wang, Da Zhang, Zhiyuan Zhao, Junyu Gao, and Xuelong Li. Maris: Marine open-vocabulary instance segmentation with geometric enhancement and semantic alignment. *arXiv preprint arXiv:2510.15398*, 2025. 6

[24] Bingyu Li, Da Zhang, Zhiyuan Zhao, Junyu Gao, and Xuelong Li. Fgaseg: Fine-grained pixel-text alignment for open-vocabulary semantic segmentation. *arXiv preprint arXiv:2501.00877*, 2025.

[25] Bingyu Li, Haocheng Dong, Da Zhang, Zhiyuan Zhao, Hao Sun, and Junyu Gao. Exploring efficient open-vocabulary segmentation in the remote sensing. In *AAAI*, 2026. 6

[26] Wenbin Li, Zhichen Fan, Jing Huo, and Yang Gao. Modeling inter-class and intra-class constraints in novel class discovery. In *CVPR*, 2023. 2

[27] Wei Li, Bing Hu, Rui Shao, Leyang Shen, and Liqiang Nie. Lion-fs: Fast & slow video-language thinker as online video assistant. In *CVPR*, 2025. 6

[28] Wei Li, Renshan Zhang, Rui Shao, Jie He, and Liqiang Nie. Cogvla: Cognition-aligned vision-language-action model via instruction-driven routing & sparsification. In *NeurIPS*, 2025. 6

[29] Ziyun Li, Jona Otholt, Ben Dai, Di Hu, Christoph Meinel, and Haojin Yang. A closer look at novel class discovery from the labeled set. In *NeurIPS Workshop*, 2022. 2

[30] Ziyun Li, Jona Otholt, Ben Dai, Di Hu, Christoph Meinel, and Haojin Yang. Supervised knowledge may hurt novel class discovery performance. *TMLR*, 2023. 2

[31] Wenteng Liang*, Feng Xue*, Yihao Liu, Guofeng Zhong, and Anlong Ming. Unknown sniffer for object detection: Don't turn a blind eye to unknown objects. In *CVPR*, 2023. 6

[32] Xiao Liu, Nan Pu, Haiyang Zheng, Wenjing Li, Nicu Sebe, and Zhun Zhong. Generate, refine, and encode: Leveraging synthesized novel samples for on-the-fly fine-grained category discovery. In *ICCV*, 2025. 2

[33] Yuanpei Liu and Kai Han. Debgcd: Debiased learning with distribution guidance for generalized category discovery. In *ICLR*, 2025. 1, 2

[34] Yu Liu and Tinne Tuytelaars. Residual tuning: Toward novel category discovery without labels. *TNNLS*, 2023. 2

[35] Yu Liu, Yaqi Cai, Qi Jia, Binglin Qiu, Weimin Wang, and Nan Pu. Novel class discovery for ultra-fine-grained visual categorization. In *CVPR*, 2024. 2

[36] Yuanpei Liu, Zhenqi He, and Kai Han. Hyperbolic category discovery. In *CVPR*, 2025. 2

[37] Shijie Ma, Fei Zhu, Xu-Yao Zhang, and Cheng-Lin Liu. Protogcd: Unified and unbiased prototype learning for generalized category discovery. *TPAMI*, 2025. 1, 2

[38] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013. 5, 2

[39] Tanmoy Mukherjee and Nikos Deligiannis. Novel class discovery: A dependency approach. In *ICASSP*, 2022. 2

[40] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006. 4

[41] Rabah Ouldnoughi, Chia-Wen Kuo, and Zsolt Kira. Clip-gcd: Simple language guided generalized category discovery. *arXiv preprint arXiv:2305.10420*, 2023. 2

[42] Zhengyuan Peng, Jinpeng Ma, Zhimin Sun, Ran Yi, Haichuan Song, Xin Tan, and Lizhuang Ma. Mos: Modeling object-scene associations in generalized category discovery. In *CVPR*, 2025. 1, 2

[43] Nan Pu, Zhun Zhong, and Nicu Sebe. Dynamic conceptional contrastive learning for generalized category discovery. In *CVPR*, 2023. 1, 2, 6, 7

[44] Nan Pu, Wenjing Li, Xingyuan Ji, Yalan Qin, Nicu Sebe, and Zhun Zhong. Federated generalized category discovery. In *CVPR*, 2024. 2

[45] Jingyang Qiao, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yong Peng, and Yuan Xie. Prompt gradient projection for continual learning. In *ICLR*, 2024. 3

[46] Yuanyuan Qing, Yijie Zeng, Qi Cao, and Guang-Bin Huang. End-to-end novel visual categories learning via auxiliary self-supervision. *Neural Networks*, 2021. 2

[47] Wentao Qu, Yuantian Shao, Lingwu Meng, Xiaoshui Huang, and Liang Xiao. A conditional denoising diffusion probabilistic model for point cloud upsampling. In *CVPR*, 2024. 6

[48] Wentao Qu, Jing Wang, YongShun Gong, Xiaoshui Huang, and Liang Xiao. An end-to-end robust point cloud semantic segmentation network with single-step conditional diffusion models. In *CVPR*, 2025. 6

[49] Sarah Rastegar, Hazel Doughty, and Cees Snoek. Learn to categorize or categorize to learn? self-coding for generalized category discovery. *NeurIPS*, 2023. 1, 2, 6, 7

[50] Sarah Rastegar, Mohammadreza Salehi, Yuki M Asano, Hazel Doughty, and Cees GM Snoek. Selex: Self-expertise in fine-grained generalized category discovery. In *ECCV*. Springer, 2024. 6, 7, 8, 3, 5

[51] Bin Ren, Yahui Liu, Yue Song, Wei Bi, Rita Cucchiara, Nicu Sebe, and Wei Wang. Masked jigsaw puzzle: A versatile position embedding for vision transformers. In *CVPR*, 2023. 6

[52] Bin Ren, Xiaoshui Huang, Mengyuan Liu, Hong Liu, Fabio Poiesi, Nicu Sebe, and Guofeng Mei. Masked clustering prediction for unsupervised point cloud pre-training. In *AAAI*, 2026. 6

[53] Sai Bhargav Rongali, Sarthak Mehrotra, Ankit Jha, Mohamad Hassan N C, Shirsha Bose, Tanisha Gupta, Mainak Singha, and Biplab Banerjee. Cdad-net: Bridging domain gaps in generalized category discovery. In *CVPRW*, 2024. 2

[54] Zhendong Ruan, Teng Yan, Yaobang Cai, Yu Han, Leyi Zheng, and Yang Zhang. Q-value regularized decision convformer for offline reinforcement learning. In *ROBIO*. IEEE, 2024. 6

[55] Gobinda Saha, Isha Garg, and Kaushik Roy. Gradient projection memory for continual learning. *arXiv preprint arXiv:2103.09762*, 2021. 3

[56] Qiannan Shen and Jing Zhang. Ai-enhanced disaster risk prediction with explainable shap analysis: A multi-class classification approach using xgboost. 2025. Preprint, Version 1, posted December 31, 2025. 6

[57] Yuchang Su, Renping Zhou, Siyu Huang, Xingjian Li, Tianyang Wang, Ziyue Wang, and Min Xu. Multimodal generalized category discovery. *arXiv preprint arXiv:2409.11624*, 2024. 2

[58] Yiyou Sun and Yixuan Li. Opencon: Open-world contrastive learning. *TMLR*, 2022. 1, 2

[59] Yiyou Sun, Zhenmei Shi, and Yixuan Li. A graph-theoretic framework for understanding open-world semi-supervised learning. In *NeurIPS*, 2023. 1, 2

[60] Yiyou Sun, Zhenmei Shi, Yingyu Liang, and Yixuan Li. When and how does known class help discover unknown ones? provable understandings through spectral analysis. In *ICML*, 2023. 2

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

[62] Luyao Tang, Kunze Huang, Chaoqi Chen, Yuxuan Yuan, Chenxin Li, Xiaotong Tu, Xinghao Ding, and Yue Huang. Dissecting generalized category discovery: Multiplex consensus under self-deconstruction. *arXiv preprint arXiv:2508.10731*, 2025. 2

[63] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *CVPR*, 2022. 1, 2, 5, 6, 7, 3

[64] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. In *ICLR2022*. OpenReview.net, 2022. 5

[65] Sagar Vaze, Andrea Vedaldi, and Andrew Zisserman. No representation rules them all in category discovery. In *NeurIPS*, 2024. 1, 2, 6, 7

[66] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 5, 2

[67] Enguang Wang, Zhimao Peng, Zhengyuan Xie, Xialei Liu, and Ming-Ming Cheng. Get: Unlocking the multi-modal potential of clip for generalized category discovery. In *CVPR*, 2025. 2

[68] Hongjun Wang, Sagar Vaze, and Kai Han. Sptnet: An efficient alternative framework for generalized category discovery with spatial prompt tuning. In *ICLR*, 2024. 1, 2, 6, 7

[69] Hongjun Wang, Sagar Vaze, and Kai Han. Hilo: A learning framework for generalized category discovery robust to domain shifts. In *ICLR*, 2025. 2

[70] Jingyu Wang, Zhenyu Ma, Feiping Nie, and Xuelong Li. Progressive self-supervised clustering with novel category discovery. *IEEE Transactions on Cybernetics*, 2022. 2

[71] Lixu Wang, Chenxi Liu, Junfeng Guo, Jiahua Dong, Xiao Wang, Heng Huang, and Qi Zhu. Federated continual novel class learning. In *ICCV*, 2025. 2

[72] Yicheng Wang, Feng Liu, Junmin Liu, and Kai Sun. Exclusive style removal for cross domain novel class discovery. *arXiv preprint arXiv:2406.18140*, 2024. 2

[73] Yu Wang, Zhun Zhong, Pengchong Qiao, Xuxin Cheng, Xiawu Zheng, Chang Liu, Nicu Sebe, Rongrong Ji, and Jie Chen. Discover and align taxonomic context priors for open-world semi-supervised learning. In *NeurIPS*, 2024. 1, 2

[74] Xin Wen, Bingchen Zhao, and Xiaojuan Qi. Parametric classification for generalized category discovery: A baseline study. In *ICCV*, 2023. 1, 2, 6, 7, 8, 3

[75] Feng Xue, Yicong Chang, Tianxi Wang, Yu Zhou, and Anlong Ming. Indoor obstacle discovery on reflective ground via monocular camera. *International Journal of Computer Vision*, 132(3), 2024. 6

[76] Teng Yan, Xiaohong Zhou, Jiamin Long, Wenxian Li, and Yang Zhang. Pandas: Prediction and detection of accurate slippage. In *IROS*. IEEE, 2025. 6

[77] Muli Yang, Yuehua Zhu, Jiaping Yu, Aming Wu, and Cheng Deng. Divide and conquer: Compositional experts for generalized novel class discovery. In *CVPR*, 2022. 1, 2

[78] Muli Yang, Jie Yin, Yanan Gu, Cheng Deng, Hanwang Zhang, and Hongyuan Zhu. Consistent prompt tuning for generalized category discovery. *IJCV*, 2025. 2

[79] Qing Yu, Daiki Ikami, Go Irie, and Kiyoharu Aizawa. Self-labeling framework for novel category discovery over domains. In *AAAI*, 2022. 2

[80] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *NeurIPS*, 2020. 3

[81] Jihun Yun, Aurélie C Lozano, and Eunho Yang. Adaptive proximal gradient methods for structured neural networks. *NeurIPS*, 34, 2021. 4

[82] Zelin Zang, Lei Shang, Senqiao Yang, Fei Wang, Baigui Sun, Xuansong Xie, and Stan Z. Li. Boosting novel category discovery over domains with soft contrastive learning and all-in-one classifier. In *ICCV*, 2023. 2

[83] Jie Zhang, Xiaosong Ma, Song Guo, and Wenchao Xu. Towards unbiased training in federated open-world semi-supervised learning. In *ICML*, 2023. 2

[84] Lu Zhang, Lu Qi, Xu Yang, Hong Qiao, Ming-Hsuan Yang, and Zhiyong Liu. Automatically discovering novel visual categories with adaptive prototype learning. *IEEE TPAMI*, 2024. 2

[85] 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 *CVPR*, 2023. 1, 2, 6, 7

[86] Wei Zhang, Baopeng Zhang, Zhu Teng, Wenxin Luo, Junnan Zou, and Jianping Fan. Less attention is more: Prompt transformer for generalized category discovery. In *CVPR*, 2025. 1, 2

[87] Bingchen Zhao and Kai Han. Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. In *NeurIPS*, 2021. 2

[88] Bingchen Zhao, Xin Wen, and Kai Han. Learning semi-supervised gaussian mixture models for generalized category discovery. In *ICCV*, 2023. 1, 2, 6, 7

[89] Dong Zhao, Qi Zang, Nan Pu, Shuang Wang, Nicu Sebe, and Zhun Zhong. Secov2: Semantic connectivity-driven pseudo-labeling for robust cross-domain semantic segmentation. *TPAMI*, 2025. 6

[90] Dong Zhao, Qi Zang, Nan Pu, Wenjing Li, Nicu Sebe, and Zhun Zhong. Open-vocabulary domain generalization in urban-scene segmentation. *arXiv preprint arXiv:2602.18853*, 2026. 6

[91] Haiyang Zheng, Nan Pu, Wenjing Li, Nicu Sebe, and Zhun Zhong. Prototypical hash encoding for on-the-fly fine-grained category discovery. In *NeurIPS*, 2024. 2

[92] Haiyang Zheng, Nan Pu, Wenjing Li, Nicu Sebe, and Zhun Zhong. Textual knowledge matters: Cross-modality co-teaching for generalized visual class discovery. In *ECCV*, 2024. 2

[93] Haiyang Zheng, Nan Pu, Wenjing Li, et al. Generalized fine-grained category discovery with multi-granularity conceptual experts. *arXiv preprint arXiv:2509.26227*, 2025. 1, 2

[94] Haiyang Zheng, Nan Pu, Wenjing Li, Teng Long, Nicu Sebe, and Zhun Zhong. Open-world deepfake attribution via confidence-aware asymmetric learning. In *AAAI*, 2026. 6

[95] 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 *CVPR*, 2020. 2

[96] Zhun Zhong, Enrico Fini, Subhankar Roy, Zhiming Luo, Elisa Ricci, and Nicu Sebe. Neighborhood contrastive learning for novel class discovery. In *CVPR*, 2021. 2

[97] Jingyu Zhuang, Ziliang Chen, Pengxu Wei, Guanbin Li, and Liang Lin. Open set domain adaptation by novel class discovery. In *IEEE International Conference on Multimedia and Expo*, 2022. 2