

GOAL: Geometrically Optimal Alignment for Continual Generalized Category Discovery

Jizhou Han¹, Chenhao Ding², SongLin Dong³, Yuhang He^{1*},
Shaokun Wang⁴, Qiang Wang¹, Yihong Gong¹

¹ National Key Laboratory of Human-Machine Hybrid Augmented Intelligence,
National Engineering Research Center for Visual Information and Applications,
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University

² School of Software Engineering, Xi'an Jiaotong University,

³ Shenzhen University of Advanced Technology,

⁴ Harbin Institute of Technology, Shenzhen

jizhou-han@stu.xjtu.edu.cn, heyuhang@xjtu.edu.cn, ygong@mail.xjtu.edu.cn

Abstract

Continual Generalized Category Discovery (C-GCD) requires identifying novel classes from unlabeled data while retaining knowledge of known classes over time. Existing methods typically update classifier weights dynamically, resulting in forgetting and inconsistent feature alignment. We propose GOAL, a unified framework that introduces a fixed Equiangular Tight Frame (ETF) classifier to impose a consistent geometric structure throughout learning. GOAL conducts supervised alignment for labeled samples and confidence-guided alignment for novel samples, enabling stable integration of new classes without disrupting old ones. Experiments on four benchmarks show that GOAL outperforms the prior method *Happy*, reducing forgetting by 16.1% and boosting novel class discovery by 3.2%, establishing a strong solution for long-horizon continual discovery.

Introduction

Continual Generalized Category Discovery (C-GCD) aims to incrementally identify novel categories from unlabeled data while retaining recognition of previously learned classes. Unlike conventional semi-supervised learning, which assumes all categories are predefined and labeled, C-GCD operates in a more realistic open-world setting. In this scenario, only a small set of known classes is labeled in the initial stage, and subsequent sessions provide unlabeled data that include a mixture of known and novel categories. The model must generalize from limited supervision to new concepts while avoiding forgetting of learned knowledge.

Recent efforts in GCD and C-GCD have led to notable advancements. Early approaches (Vaze et al. 2022) used contrastive learning to separate known and unknown classes. Building on this idea, DCCL (Pu, Zhong, and Sebe 2023) and PromptCAL (Zhang et al. 2023) introduced dynamic contrastive objectives and prompt-based affinities to refine decision boundaries. Other methods (Chiaroni et al. 2023; Choi, Kang, and Cho 2024) enhanced representation quality via clustering and mean-shift refinement. In

the continual setting, MetaGCD (Wu et al. 2023) leveraged meta-initialization to improve cross-session transfer, and Happy (Ma et al. 2024a) mitigated bias through entropy-based regularization and prototype sampling.

However, most existing C-GCD frameworks rely on dynamically optimized classifier heads or class prototypes, which present two major challenges. **First, catastrophic forgetting:** The continuous update of class representations across sessions causes newly learned knowledge to overwrite prior information, particularly when novel classes are trained in a purely unsupervised manner. **Second, category confusion:** Without an explicit geometric constraint on the feature space, similar or nearby classes—especially emerging ones—tend to overlap, resulting in ambiguous decision boundaries. As illustrated in Fig. 1b, existing methods often struggle to maintain global class separability. These observations lead to a natural question: *Can we predefine an optimal and unified geometric structure for all classes, enabling consistent feature alignment throughout continual discovery while alleviating forgetting and class confusion?*

We draw inspiration from the Neural Collapse phenomenon (Papayan, Han, and Donoho 2020), which reveals that well-trained classifiers tend to organize features in an Equiangular Tight Frame (ETF), where class means are maximally separated and intra-class variation collapses. This structure provides an optimal configuration for classification, both theoretically and empirically.

Inspired by the Neural Collapse phenomenon, we propose a novel framework named Geometrically Optimal Alignment for Continual Generalized Category Discovery (GOAL), which establishes a globally consistent classification structure by utilizing a fixed ETF, as depicted in Fig. 1c. In contrast to existing approaches that iteratively adjust classifier prototypes across sessions, GOAL predefines an optimal geometric configuration and enforces alignment of both labeled and unlabeled features to this fixed structure throughout the entire continual learning process. The framework is composed of a frozen ETF-based classifier, a learnable feature encoder, and a confidence-guided alignment module. During the base session, labeled samples are explicitly guided to align with their designated ETF direc-

*Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

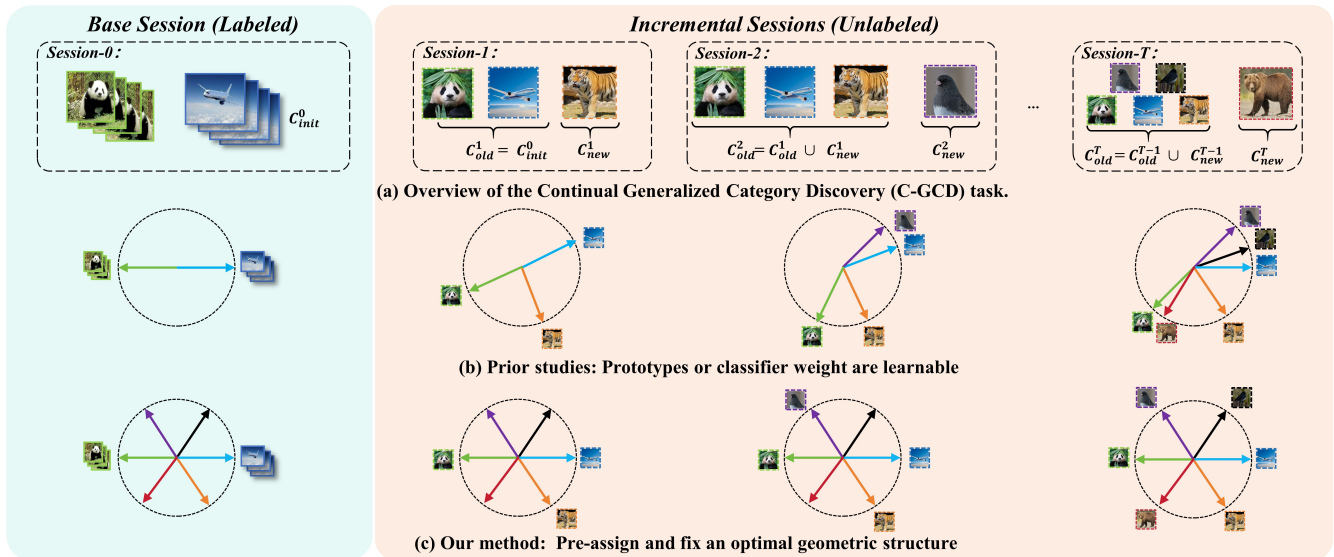


Figure 1: Conceptual illustration of the Continual Generalized Category Discovery (C-GCD) task and classifier design strategies. (a) Overview of the C-GCD task: a model is trained on labeled base classes and incrementally discovers novel categories from unlabeled data in subsequent sessions. (b) Prior studies typically use learnable prototypes or classifier weights. (c) Our method adopts a fixed optimal geometric structure, ensuring better alignment and generalization across continual sessions.

tions, thereby forming an initial structured representation space. In the subsequent incremental sessions, unlabeled data containing both previously seen and novel categories are processed by evaluating the prediction confidence of each sample. A fixed proportion of samples with the highest confidence, identified through minimum entropy, are selected and aligned with unassigned ETF vectors. This mechanism facilitates the stable integration of novel categories while preserving the global geometric consistency of the feature space. By maintaining a consistent geometric objective across sessions, GOAL enhances class separability, stabilizes feature alignment, and effectively mitigates forgetting in the absence of explicit supervision for novel categories. Our main contributions are summarized as follows:

- We propose GOAL, a Neural Collapse-inspired framework for continual generalized category discovery, which leverages a fixed ETF classifier to provide a unified and consistent optimization target across all sessions.
- We design a confidence-guided alignment strategy that incrementally assigns high-confidence novel samples to unoccupied ETF directions, preserving feature structure without requiring ETF classifier prototypes updates.
- GOAL achieves strong performance on four C-GCD benchmarks, reducing average forgetting by **16.10%** and improving novel category discovery accuracy by **3.19%** compared to the strongest prior method (Ma et al. 2024a).

Related Work

Generalized Category Discovery

Generalized Category Discovery (GCD) aims to classify known categories while discovering novel ones from partially labeled data. Early works such as VanillaGCD (Vaze

et al. 2022) introduced contrastive learning to distinguish between known and unknown classes. DCCL (Pu, Zhong, and Sebe 2023) and PromptCAL (Zhang et al. 2023) further refined decision boundaries via dynamic similarity modeling and prompt-based affinity propagation. Parametric solutions like SimGCD (Wen, Zhao, and Qi 2023) and GPC (Zhao, Wen, and Han 2023) employed Gaussian mixture modeling or soft classifiers to capture latent structures. Others, including PIM (Chiaroni et al. 2023), CMS (Choi, Kang, and Cho 2024), and AGCD (Ma et al. 2024b), emphasized representation refinement through clustering and mean-shift mechanisms. NCGCD (Han et al. 2025) explores the application of ETF in GCD, but it cannot perform incremental learning. Despite their effectiveness, these methods update prototypes during training, introducing optimization inconsistency that hampers the discovery of novel categories.

Continual Generalized Category Discovery

Continual Generalized Category Discovery (C-GCD) extends the GCD setting to continual learning, where the model receives data in sequential sessions containing a mixture of known and unknown classes. GM (Zhang et al. 2022) proposes a grow-and-merge pipeline, clustering new classes and merging them with prior knowledge. MetaGCD (Wu et al. 2023) leverages meta-learning to initialize models for better cross-stage generalization. Happy (Ma et al. 2024a) tackles prediction bias via entropy-based regularization and prototype resampling. However, current C-GCD methods still face challenges, including **catastrophic forgetting** and the lack of a **global geometric constraint** to preserve inter-class structure across sessions. These issues lead to inconsistent optimization and degraded performance over time.

Neural Collapse and Geometric Structures

Neural Collapse (NC) (Papayan, Han, and Donoho 2020) characterizes the terminal phase of deep classifier training, where within-class features collapse to their means, and class means form a Simplex Equiangular Tight Frame (ETF). This configuration achieves maximum inter-class margin and minimal intra-class variance. NC has been theoretically shown to emerge under both cross-entropy (Wojtowytch et al. 2020; Fang et al. 2021; Graf et al. 2021; Wang et al. 2025) and MSE (Mixon, Parshall, and Pi 2022; Zhou et al. 2022; Han, Papayan, and Donoho 2022) losses. Moreover, NC-inspired formulations have been applied to domains such as imbalanced classification (Zhong et al. 2023), federated learning (Huang et al. 2023), and few-shot class-incremental learning (Yang et al. 2022).

In this work, we are the first to introduce an NC-guided geometric constraint into the C-GCD setting. By assigning a fixed ETF classifier and aligning features through both supervised and confident unsupervised signals, we ensure a **globally consistent optimization target**, which enhances stability, improves novel category separability, and mitigates forgetting across continual sessions.

Preliminaries

Problem Setting of C-GCD

Continual Generalized Category Discovery (C-GCD) extends the conventional GCD task to a continual learning scenario. Formally, the learning process is divided into an initial supervised session (Stage-0) and a sequence of incremental discovery sessions (Stage- t , $1 \leq t \leq T$).

In Stage-0, the model is trained on a small labeled dataset $\mathcal{D}_{\text{train}}^0 = \{(x_i^l, y_i)\}_{i=1}^{N^0}$, where $y_i \in \mathcal{C}_{\text{init}}^0$ and $\mathcal{C}_{\text{old}}^0 = \mathcal{C}_{\text{init}}^0$ denotes the set of initially known categories.

For each incremental stage $t \in \{1, 2, \dots, T\}$, the model receives an unlabeled dataset $\mathcal{D}_{\text{train}}^t = \{x_i^u\}_{i=1}^{N^t}$, which contains instances sampled from a mixture of known and unknown categories. The full category set at stage t is defined as $\mathcal{C}^t = \mathcal{C}_{\text{old}}^t \cup \mathcal{C}_{\text{new}}^t$, where $\mathcal{C}_{\text{old}}^t$ denotes previously discovered categories and $\mathcal{C}_{\text{new}}^t$ represents novel categories appearing at this stage. The number of novel classes $K_{\text{new}}^t = |\mathcal{C}_{\text{new}}^t|$ is assumed to be either given or estimated. After each stage, the model is evaluated on a disjoint test set $\mathcal{D}_{\text{test}}^t = \{(x_i, y_i)\}_{i=1}^{N_{\text{test}}^t}$ that covers all known and newly discovered categories.

Definition of Optimal Geometric Structure

In the final stages of training well-regularized neural networks, especially on balanced classification tasks, a remarkable structural phenomenon known as *Neural Collapse* emerges (Papayan, Han, and Donoho 2020). This behavior reveals that both feature embeddings and classifier weights organize into a highly symmetric configuration, where samples from the same class collapse to a shared center, and class centers are arranged in an optimal geometric layout. This structure is mathematically characterized by a *Simplex Equiangular Tight Frame (ETF)*.

Simplex ETF Construction. A set of K prototype vectors $P = [p_1, p_2, \dots, p_K] \in \mathbb{R}^{d \times K}$ forms a simplex ETF if they

satisfy the following condition:

$$P = \sqrt{\frac{K}{K-1}} U \left(I_K - \frac{1}{K} \mathbf{1}_K \mathbf{1}_K^\top \right), \quad (1)$$

where $U \in \mathbb{R}^{d \times K}$ is an orthonormal matrix such that $U^\top U = I_K$, and $\mathbf{1}_K$ is a column vector of ones. Each vector p_k is unit-normalized, i.e., $\|p_k\| = 1$, and the pairwise inner product between any two vectors satisfies:

$$p_k^\top p_j = \begin{cases} 1 & \text{if } k = j \\ -\frac{1}{K-1} & \text{if } k \neq j \end{cases} \quad (2)$$

which ensures equal angular separation between all classes. **Neural Collapse Properties.** The Neural Collapse phenomenon includes four key behaviors:

(NC1) *Within-class Collapse:* All features from the same class converge to their class mean. The within-class scatter tends to vanish, i.e., $\Sigma_W^{(k)} \rightarrow 0$, where $\Sigma_W^{(k)} = \text{Avg} \{(\mu_{k,i} - \mu_k)(\mu_{k,i} - \mu_k)^\top\}$ and $\mu_{k,i}$ is the feature of the i -th instance in class k .

(NC2) *Simplex Configuration:* After centering each class mean by subtracting the global mean $\mu_G = \text{Avg}_{i,k}(\mu_{k,i})$, the normalized class means $\hat{\mu}_k = \frac{\mu_k - \mu_G}{\|\mu_k - \mu_G\|}$ tend to occupy the vertices of a simplex ETF.

(NC3) *Feature-Classifier Alignment:* The centered class means align with the corresponding normalized classifier weights, i.e., $\hat{\mu}_k = \frac{w_k}{\|w_k\|}$, where w_k is the weight vector for class k in the final linear classifier.

(NC4) *Prediction Simplification:* In this regime, classification can be interpreted as a nearest-mean decision rule: $\hat{y} = \arg \min_k \|\mu - \mu_k\| = \arg \max_k \langle \mu, w_k \rangle$, where μ is the test feature representation.

Method

Overview

As illustrated in Fig. 2, our GOAL framework addresses Continual Generalized Category Discovery (C-GCD) by unifying supervised and unsupervised alignment under a fixed geometric structure. GOAL comprises three key components: a frozen ETF prototype set P , a feature encoder $f(\cdot)$, and a session-specific classifier W . These modules are optimized jointly across sessions without storing exemplars.

In the base session, the model is trained with labeled data to establish a structured feature space. Feature embeddings are aligned to fixed ETF directions via dot-product loss, while a supervised contrastive loss further enhances intra-class compactness. A parametric classifier W^0 is also trained to support known-class recognition.

In each incremental session ($t \geq 1$), the model receives unlabeled data containing both seen and novel categories. High-confidence samples are selected via entropy ranking and aligned to unused ETF directions. A clustering-guided strategy is used to initialize new classifier weights, and a pseudo-label-based classification loss guides training. Additionally, unsupervised contrastive learning further improves feature discrimination. By maintaining alignment with a

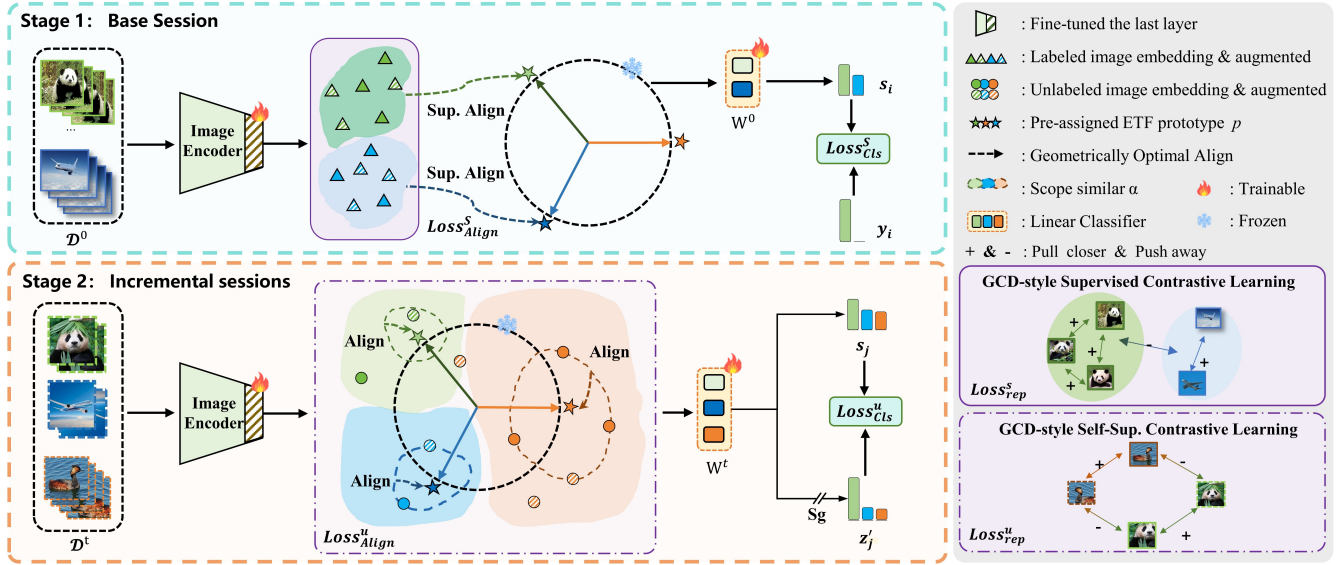


Figure 2: Overview of the GOAL framework. In the base session, labeled features are aligned to fixed ETF prototypes via supervised training. In incremental sessions, high-confidence unlabeled samples are aligned to unallocated prototypes with unsupervised learning. This maintains geometric consistency for continual discovery and retention.

global ETF structure throughout all sessions, GOAL enables progressive discovery of novel classes while preserving knowledge of previous ones—achieving effective C-GCD without any memory buffer or prototype updates.

Base Session

The base session uses labeled data to build a structured feature space by aligning features with predefined ETF prototypes and learning discriminative representations. This establishes a unified geometric foundation for recognizing known classes and discovering novel ones in later sessions.

Predefined ETF Prototypes To promote a well-structured feature space and ensure maximum separation between categories, we construct a fixed set of classifier prototypes based on an ETF. Given a total of K categories, the corresponding ETF prototype matrix is denoted as $\mathbf{P} = \{p_1, p_2, \dots, p_K\} \in \mathbb{R}^{d \times K}$, as defined in Eq. 1. Each prototype p_i is normalized to unit ℓ_2 -norm and satisfies a fixed pairwise cosine similarity:

$$p_i^\top p_j = \frac{K}{K-1} q_i^\top q_j - \frac{1}{K-1}, \quad \forall i, j \in [1, K], \quad (3)$$

where q_i denotes the i -th column of an orthonormal matrix.

This construction provides a stable and optimal alignment structure, serving as a consistent geometric target for both known and novel categories.

Supervised Alignment to ETF Prototypes For each labeled training sample $(x_i^l, y_i^l) \in \mathcal{D}_{\text{train}}^0$, we extract a feature embedding $e_i = f(x_i^l)$ using the image encoder $f(\cdot)$. The feature is then normalized to unit length: $\hat{e}_i = e_i / \|e_i\|_2$, and we align it with the pre-assigned ETF prototype $p_{y_i^l}$ corre-

sponding to the ground-truth label y_i^l . The alignment loss is:

$$\mathcal{L}_{\text{Align}}^s = -\frac{1}{N^0} \sum_{i=1}^{N^0} \langle \hat{e}_i, p_{y_i^l} \rangle, \quad (4)$$

where $\langle \cdot \rangle$ denotes inner product and N^0 is the number of labeled samples in the base session.

Representation Learning and Parametric Classification *Contrastive Representation Learning.* We adopt a GCD-style representation learning (Vaze et al. 2022). For each image $x_i^l \in \mathcal{D}_{\text{train}}^0$, we generate two augmented views x_i and x_i' , and extract their normalized embeddings e_i and e_i' using the encoder $f(\cdot)$. The unsupervised contrastive loss is:

$$\mathcal{L}_{\text{rep}}^u = -\frac{1}{|B|} \sum_{i \in B} \log \frac{\exp(e_i \cdot e_i' / \tau)}{\sum_{j \in B, j \neq i} \exp(e_i \cdot e_j / \tau)}, \quad (5)$$

τ is a temperature hyperparameter and B is the batch size.

For labeled samples, we further encourage semantic consistency through a supervised contrastive loss:

$$\mathcal{L}_{\text{rep}}^s = -\frac{1}{|B_l|} \sum_{i \in B_l} \frac{1}{|H(i)|} \sum_{h \in H(i)} \log \frac{\exp(e_i \cdot e_h / \tau)}{\sum_{j \neq i} \exp(e_i \cdot e_j / \tau)}, \quad (6)$$

where $H(i)$ is the set of positive samples sharing the same label with i . The total contrastive loss is:

$$\mathcal{L}_{\text{rep}}^{\text{Base}} = (1 - \lambda_{\text{rep}}) \mathcal{L}_{\text{rep}}^u + \lambda_{\text{rep}} \mathcal{L}_{\text{rep}}^s. \quad (7)$$

Parametric Classification. We also employ a trainable parametric classifier $W^0 \in \mathbb{R}^{d \times K}$ initialized for the known categories. For labeled samples, the classification loss is defined as standard cross-entropy:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{|B_l|} \sum_{i \in B_l} \log \frac{\exp(w_{y_i^l}^\top e_i)}{\sum_{j=1}^K \exp(w_j^\top e_i)}, \quad (8)$$

Datasets	Methods	S-0	Stage-1			Stage-2			Stage-3			Stage-4			Stage-5			1-5 Stages Avg		
		All	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
C100	VanillaGCD †	90.8	72.3	78.5	41.4	67.0	72.5	34.3	58.0	62.3	28.1	56.6	59.5	33.0	51.4	53.7	30.3	61.1	65.3	33.4
	SimGCD †	90.4	73.4	86.4	8.0	62.6	72.4	3.3	54.2	61.6	2.1	47.6	53.4	1.6	43.5	47.9	4.6	56.2	64.3	3.9
	SimGCD+ †	90.4	75.9	87.0	20.4	67.1	75.3	17.5	58.4	64.3	17.3	54.3	58.7	19.1	50.5	53.9	19.8	61.3	67.9	18.8
	FRoST †	90.4	76.9	79.6	<u>63.3</u>	65.3	68.9	43.9	58.0	61.1	36.5	49.3	50.9	36.2	48.0	48.2	46.8	59.5	61.7	45.3
	GM †	90.4	76.6	79.8	60.5	71.1	74.5	<u>50.6</u>	63.5	68.2	31.0	59.7	62.5	37.6	54.1	54.7	<u>48.4</u>	65.0	68.0	45.6
	MetaGCD †	90.8	76.1	83.6	38.7	69.4	72.8	48.9	62.0	65.8	35.3	58.2	61.2	34.3	55.8	58.5	31.6	64.3	68.4	37.8
	Happy †	90.4	<u>80.4</u>	85.3	56.1	<u>74.1</u>	<u>78.3</u>	49.3	<u>68.2</u>	<u>70.9</u>	<u>49.8</u>	<u>62.3</u>	<u>63.8</u>	50.3	<u>60.0</u>	<u>61.0</u>	51.3	<u>69.0</u>	<u>71.8</u>	<u>51.4</u>
	GOAL (Ours)	91.0	82.8	<u>86.4</u>	65.1	77.8	84.0	62.4	71.8	82.8	53.5	66.6	83.8	<u>45.1</u>	61.5	79.2	43.8	72.1	83.2	54.0
Tiny	VanillaGCD †	84.2	55.9	58.9	41.0	55.0	58.6	33.2	52.8	55.7	32.4	48.8	51.5	27.6	45.9	48.1	26.9	51.7	54.6	32.2
	SimGCD †	85.9	67.0	79.9	2.0	57.8	67.0	2.8	52.7	59.8	2.8	45.0	50.3	2.8	41.6	45.8	3.8	52.8	60.6	2.8
	SimGCD+ †	85.9	70.4	81.8	13.3	62.5	70.8	12.8	54.6	60.5	13.2	48.0	52.5	11.9	43.0	46.5	12.7	55.7	62.4	12.8
	FRoST †	85.9	75.2	78.6	<u>58.1</u>	65.6	67.8	52.5	51.3	54.3	30.4	48.2	52.1	16.9	40.2	42.7	16.9	56.1	59.1	35.0
	GM †	85.9	76.4	82.4	46.5	<u>68.9</u>	<u>73.8</u>	39.2	58.7	63.4	25.4	52.9	57.2	18.1	46.9	50.6	13.4	60.8	65.5	28.5
	MetaGCD †	84.2	60.9	64.9	40.8	57.2	61.0	34.2	54.4	57.2	34.6	50.8	53.6	28.8	48.1	50.2	30.0	54.3	57.4	33.7
	Happy †	85.9	78.9	82.4	61.1	71.3	76.2	42.3	<u>64.7</u>	<u>68.7</u>	<u>36.5</u>	<u>58.5</u>	<u>60.6</u>	<u>41.3</u>	<u>54.6</u>	<u>56.7</u>	<u>35.7</u>	<u>65.6</u>	<u>68.9</u>	<u>43.4</u>
	GOAL (Ours)	86.1	<u>77.1</u>	<u>81.9</u>	52.8	72.1	80.3	<u>51.6</u>	67.3	77.0	51.1	61.5	76.4	42.8	57.2	74.5	39.9	67.1	78.1	47.9
CUB	VanillaGCD †	89.2	64.5	67.1	51.9	58.2	60.7	42.9	54.1	56.4	37.9	50.0	51.3	39.3	46.8	46.6	<u>49.1</u>	54.7	56.4	44.2
	SimGCD †	90.3	73.8	84.5	22.0	63.4	72.4	8.6	55.6	62.0	11.1	49.3	54.6	7.9	44.7	48.7	9.3	57.4	64.4	11.8
	SimGCD+ †	90.3	75.6	85.6	26.0	65.3	73.9	13.7	57.4	63.3	16.3	51.1	55.7	14.3	45.8	49.3	14.3	59.1	65.6	16.9
	FRoST †	90.3	77.0	84.0	43.5	50.8	53.5	34.3	46.4	49.3	26.1	39.4	41.5	23.1	34.6	35.1	29.5	49.6	52.7	31.3
	GM †	90.3	76.2	80.2	56.5	67.9	73.4	34.6	61.1	66.5	23.0	55.9	57.5	43.4	52.0	54.4	30.1	62.6	66.4	37.5
	MetaGCD †	89.2	67.1	70.2	51.9	60.8	61.0	50.9	57.5	59.3	37.8	51.9	52.2	49.4	49.6	50.0	46.4	57.4	58.8	47.3
	Happy †	90.3	81.4	<u>85.1</u>	<u>63.7</u>	74.3	<u>76.0</u>	<u>63.6</u>	<u>67.1</u>	<u>71.1</u>	<u>39.1</u>	<u>62.3</u>	<u>63.8</u>	<u>49.7</u>	<u>59.4</u>	<u>60.5</u>	49.5	<u>68.9</u>	<u>71.3</u>	<u>53.1</u>
	GOAL (Ours)	90.4	<u>81.1</u>	83.1	71.3	<u>73.1</u>	76.4	65.0	69.0	76.1	57.3	64.2	74.6	51.4	61.9	76.0	47.9	69.9	77.2	58.6
IN100	VanillaGCD †	96.0	70.1	72.9	56.2	69.4	73.5	44.8	68.5	70.6	53.6	65.6	67.8	47.2	64.5	67.4	38.4	67.6	70.5	48.0
	SimGCD †	96.2	79.7	91.7	19.6	70.2	78.8	18.6	61.9	67.4	23.2	56.7	60.9	22.6	52.9	56.4	21.4	64.3	71.1	21.1
	SimGCD+ †	96.2	83.1	95.2	22.6	74.6	83.5	21.2	67.6	73.6	25.8	62.1	66.8	24.2	57.6	61.5	23.0	69.0	76.1	23.4
	FRoST †	96.2	87.5	93.0	60.2	79.6	83.4	57.2	76.8	77.0	<u>75.2</u>	66.2	68.7	46.4	63.8	66.4	40.6	74.8	77.7	55.9
	GM †	96.2	89.5	95.0	62.0	82.3	86.9	54.8	78.0	79.2	<u>69.6</u>	72.8	74.7	58.0	71.1	71.8	65.0	78.7	81.5	61.9
	MetaGCD †	96.0	75.3	78.2	60.6	73.8	75.9	54.9	69.4	72.2	49.4	67.2	70.1	44.2	66.7	69.3	43.0	70.5	73.2	50.4
	Happy †	96.2	<u>91.2</u>	<u>95.4</u>	<u>70.4</u>	<u>87.8</u>	<u>90.8</u>	<u>69.8</u>	<u>85.2</u>	<u>86.4</u>	77.0	81.9	<u>83.0</u>	73.4	<u>78.6</u>	<u>79.1</u>	73.8	<u>85.0</u>	<u>86.9</u>	<u>72.9</u>
	GOAL (Ours)	96.2	93.9	95.9	83.8	88.8	94.9	73.7	86.4	94.4	72.9	<u>81.2</u>	92.4	<u>68.4</u>	79.3	93.0	<u>66.6</u>	85.9	94.1	73.1

Table 1: Performance of C-GCD on four benchmarks. Here † denotes results reported by *Happy*.

where w_j is the weight vector for class j , and y_i is the true label for sample x_i . The total training objective for the base session is:

$$\mathcal{L}_{Base} = \mathcal{L}_{Align}^s + \mathcal{L}_{rep}^{Base} + \mathcal{L}_{cls}. \quad (9)$$

Incremental Sessions

After the base session, the model proceeds through a series of incremental sessions $t \in \{1, 2, \dots, T\}$, where it receives unlabeled data $\mathcal{D}_{train}^t = \{x_j^u\}_{j=1}^{N^t}$ consisting of both known and novel categories. The challenge lies in discovering new classes while maintaining performance on old ones, without access to labeled exemplars.

Parametric Classification To handle new classes, we expand the classifier using clustering-guided initialization (Ma et al. 2024a). Specifically, we apply KMeans clustering on the unlabeled dataset \mathcal{D}_{train}^t to obtain a set of ℓ_2 -normalized cluster centers $\{c_i\}$. For each cluster center c_i , we compute its maximum cosine similarity with old classifier weights W_{t-1} . We then select the top C_{new} centers with the lowest similarity scores to serve as new class weights $W_t^{new} = c_{t_j}$.

The classifier weights at stage t are obtained by concatenating the old and new weights:

$$W_t = \text{Concat}[W_{t-1}; W_t^{new}] \in \mathbb{R}^{d \times (C_{old} + C_{new})}. \quad (10)$$

For training, each unlabeled sample x_j^u is passed through the image encoder and the classifier to get a prediction \tilde{s}_j , and its augmented view $x_j^{u'}$ provides a sharper target distribution z_j . The classification loss is:

$$\mathcal{L}_{cls}^u = \frac{1}{|\mathcal{D}_{train}^t|} \sum_{x_j \in \mathcal{D}_{train}^t} \mathcal{H}(z_j, \tilde{s}_j) - \epsilon \mathcal{H}(\bar{s}), \quad (11)$$

\mathcal{H} denotes cross-entropy, ϵ is a regularization parameter and \bar{s} is the mean batch prediction.

Finally, we rank all samples by prediction entropy and select the top $\alpha\%$ with lowest entropy to form the confident subset \mathcal{D}_{conf}^t for ETF alignment in the next step.

Unsupervised ETF Alignment Following the confidence guided strategy, we cluster the selected high-confidence samples \mathcal{D}_{conf}^t and match the resulting centroids to unused

Methods	CIFAR100		CUB-200		TinyImageNet		ImageNet-100		All Datasets Avg	
	$\mathcal{M}_f \downarrow$	$\mathcal{M}_d \uparrow$	$\mathcal{M}_f \downarrow$	$\mathcal{M}_d \uparrow$	$\mathcal{M}_f \downarrow$	$\mathcal{M}_d \uparrow$	$\mathcal{M}_f \downarrow$	$\mathcal{M}_d \uparrow$	$\mathcal{M}_f \downarrow$	$\mathcal{M}_d \uparrow$
FRoST (ECCV 2022)	42.19	45.34	55.14	31.30	43.13	34.96	29.80	55.92	42.57	41.88
GM (NeurIPS 2022)	35.62	45.62	35.86	37.51	35.24	28.52	24.44	61.88	32.79	43.38
MetaGCD (CVPR 2023)	32.35	37.76	39.24	47.27	34.04	33.68	26.65	50.42	33.07	42.28
Happy (NeurIPS 2024)	29.40	51.36	29.77	53.13	29.20	43.38	17.09	72.88	26.37	55.19
GOAL (Ours)	11.82	53.97	14.36	58.57	11.66	47.88	3.24	73.08	10.27	58.38
<i>Improv. over Happy</i>	<i>+17.58</i>	<i>+2.61</i>	<i>+15.41</i>	<i>+5.44</i>	<i>+17.54</i>	<i>+4.50</i>	<i>+13.85</i>	<i>+0.20</i>	<i>+16.10</i>	<i>+3.19</i>

Table 2: Forgetting and discovery performance on four datasets. We report forgetting rate (\mathcal{M}_f) and final discovery rate (\mathcal{M}_d).

Datasets	Methods	S-0	S-1	S-2	S-3	S-4	S-5	S-6	S-7	S-8	S-9	S-10	1-10-stages Avg		$\mathcal{M}_f \downarrow$	
		All	All	All	All	All	All	All	All	All	All	All	All	Old		New
C100	MetaGCD	90.82	81.07	76.55	74.26	67.64	64.45	61.58	59.13	60.13	56.91	56.51	65.82	68.49	31.82	32.81
	Happy	90.36	85.62	81.88	79.82	74.01	71.81	68.46	64.05	62.14	61.38	57.81	70.70	71.89	52.34	31.00
	GOAL (Ours)	90.60	87.22	84.50	83.20	77.66	75.55	73.72	70.07	68.04	66.48	63.06	74.95	82.63	62.62	11.86
	<i>Improv. over Happy</i>	<i>+0.24</i>	<i>+1.60</i>	<i>+2.62</i>	<i>+3.38</i>	<i>+3.65</i>	<i>+3.74</i>	<i>+5.26</i>	<i>+6.02</i>	<i>+5.90</i>	<i>+5.10</i>	<i>+5.25</i>	<i>+4.25</i>	<i>+10.74</i>	<i>+10.28</i>	<i>+19.14</i>
Tiny	MetaGCD	84.20	68.87	65.48	62.92	60.81	58.21	56.16	54.56	52.48	50.57	48.92	57.90	59.91	30.76	34.35
	Happy	85.86	80.75	76.92	73.34	69.77	66.33	62.75	57.56	54.73	53.02	50.69	64.59	66.24	43.04	34.91
	GOAL (Ours)	86.14	80.96	79.23	76.43	72.40	68.85	64.95	62.28	58.63	56.77	53.03	67.35	79.15	47.23	11.20
	<i>Improv. over Happy</i>	<i>+0.28</i>	<i>+0.21</i>	<i>+2.31</i>	<i>+3.09</i>	<i>+2.63</i>	<i>+2.52</i>	<i>+2.20</i>	<i>+4.72</i>	<i>+3.90</i>	<i>+3.75</i>	<i>+2.34</i>	<i>+2.77</i>	<i>+12.91</i>	<i>+4.19</i>	<i>+23.71</i>

Table 3: C-GCD performance with 10 incremental stages on CIFAR100 and TinyImageNet.

ETF prototypes. Each sample $x_j \in \mathcal{D}_{\text{conf}}^t$ is assigned to a prototype $p_{\phi(j)}$, and the alignment loss is defined as:

$$\mathcal{L}_{\text{Align}}^u = -\frac{1}{|\mathcal{D}_{\text{conf}}^t|} \sum_{x_j \in \mathcal{D}_{\text{conf}}^t} \langle \hat{e}_j, p_{\phi(j)} \rangle, \quad (12)$$

where \hat{e}_j is the normalized embedding and $\phi(j)$ is the index of the assigned ETF prototype. This mechanism anchors novel features to the global structure.

Incremental Objective To improve representation quality under unlabeled conditions, we retain unsupervised contrastive learning for each sample, denoted as $\mathcal{L}_{\text{rep}}^{\text{Inc}}$. The overall objective is:

$$\mathcal{L}_{\text{Inc}} = \lambda_A \mathcal{L}_{\text{Align}}^u + \mathcal{L}_{\text{rep}}^{\text{Inc}} + \mathcal{L}_{\text{cls}}^u, \quad (13)$$

where λ_A balances alignment strength.

Experiment

Experimental Setup

Datasets. We evaluate GOAL on four C-GCD benchmarks: CIFAR100 (Krizhevsky and Hinton 2009), TinyImageNet (Deng et al. 2009), ImageNet-100 (Tian, Krishnan, and Isola 2020), and CUB (Wah et al. 2011). Following (Vaze et al. 2022), each dataset is split into a labeled base session (50% classes) and $T = 5$ incremental sessions with unlabeled novel classes and previously seen classes.

Evaluation Protocol. At each stage t , the model is trained on unlabeled training data $\mathcal{D}_{\text{train}}^t$ and evaluated on a disjoint labeled test set $\mathcal{D}_{\text{test}}^t$. We adopt an inductive evaluation, where ground-truth labels are only used for computing accuracy. Predictions are matched to true classes using the Hungarian algorithm over all classes seen so far. We report accuracy for three subsets: ‘‘All’’ (overall), ‘‘Old’’, and ‘‘New’’.

Implementation Details. We use a pre-trained DINO ViT-B/16 (Caron et al. 2021; Dosovitskiy et al. 2021) as the backbone, fine-tuning only the last block and a 3-layer MLP projection head (hidden size 2048, output dim 768). Stage-0 is trained for 100 epochs with labeled data; each incremental session runs for 30 epochs using a batch size of 128 and learning rate 0.01. We set $\lambda_A = 0.7$ and $\lambda_{\text{rep}} = 0.35$. All experiments are conducted on NVIDIA RTX 3090 GPU.

Comparison with State of the Art Methods

We compare GOAL with representative approaches in Generalized Category Discovery (GCD) and its continual variant (C-GCD). These include VanillaGCD (Vaze et al. 2022), SimGCD (Wen, Zhao, and Qi 2023), SimGCD+ (with LwF), FRoST (Roy et al. 2022), GM (Zhang et al. 2022), MetaGCD (Wu et al. 2023), and Happy (Ma et al. 2024a). To ensure fairness, we follow MetaGCD’s protocol for GM and use sampled features instead of exemplars. Full results are reported in Table 1, Table 2, and Table 3.

As summarized in Table 1, GOAL establishes a new state of the art in C-GCD across all benchmarks. Our method achieves consistent and significant gains over the best prior approach, Happy. On CIFAR100, GOAL improves overall accuracy by +3.1%, with clear benefits for both old +11.4% and new +2.6% classes. On TinyImageNet, GOAL narrows the old-new performance gap and yields a +4.5% boost for novel categories. For CUB, GOAL significantly enhances new class discovery +5.5% while maintaining strong retention. These results confirm the effectiveness of our confidence-driven alignment and fixed geometric structure in supporting balanced, continual category discovery.

Sup ETF Alignment	Unsup ETF Alignment	CUB			CIFAR100			TinyImageNet			ImageNet-100			Avg		
		All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
\times	\times	64.8	73.1	49.9	69.6	79.9	47.5	63.3	72.5	44.2	80.1	87.9	68.0	69.5	78.4	52.4
\checkmark	\times	64.7	77.5	51.1	70.4	83.7	46.2	63.4	<u>73.7</u>	44.7	82.3	94.3	68.7	70.2	<u>82.3</u>	52.7
\times	\checkmark	67.9	73.8	<u>57.3</u>	<u>71.2</u>	81.2	52.4	<u>65.6</u>	72.8	46.1	84.7	89.5	72.1	72.4	79.3	57.0
\checkmark	\checkmark	69.9	<u>77.2</u>	58.6	72.1	<u>83.2</u>	54.0	67.1	78.1	47.9	85.9	<u>94.1</u>	73.1	73.8	83.2	58.4
<i>Improv. over baseline</i>		+5.1	+4.1	+8.7	+2.5	+3.3	+6.5	+3.8	+5.6	+3.7	+5.8	+6.2	+5.1	+4.3	+4.8	+6.0

Table 4: Ablation study on the effect of supervised and unsupervised ETF alignment.

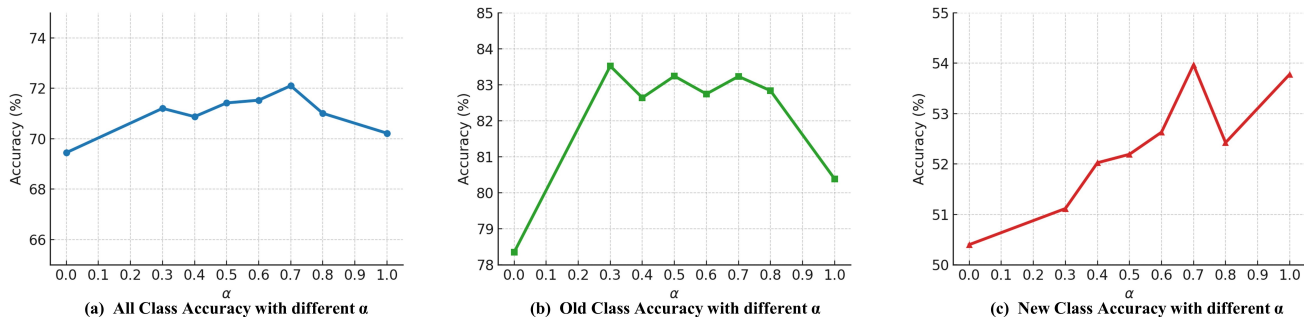


Figure 3: Accuracy trend across different values of α on CIFAR100.

Forgetting and Discovery Performance. To evaluate the model’s continual learning ability, we adopt two metrics: the forgetting rate (\mathcal{M}_f) and the final discovery rate (\mathcal{M}_d) in (Zhang et al. 2022). The forgetting rate measures how much performance on previously learned classes degrades over time, while the discovery rate reflects the model’s accuracy in identifying novel classes introduced in later sessions.

As shown in Table 2, GOAL consistently outperforms prior methods, achieving both the lowest forgetting and highest discovery scores across all benchmarks. Specifically, compared to *Happy*, GOAL reduces \mathcal{M}_f by more than **17%** on CIFAR100 and TinyImageNet, while improving \mathcal{M}_d by +2.6% and +4.5%, respectively. On CUB and IN100, GOAL also achieves substantial gains in both metrics. Averaged across all datasets, GOAL achieves a **16.1%** reduction in forgetting and a **3.2%** improvement in discovery, demonstrating its strong capability to retain old knowledge while continuously discovering new categories.

C-GCD with More Continual Stages. We evaluate GOAL under a 10 stage setting on CIFAR100 and TinyImageNet. As shown in Table 3, GOAL improves new class accuracy over *Happy* by **+10.28%** and **+4.19%**, respectively, while reducing forgetting by **19.14%** and **23.71%**. These results highlight GOAL’s strength in discovering novel categories and mitigating forgetting, making it especially suitable for long-horizon continual discovery.

Ablation Study

Effect of ETF Alignment. To assess the contributions of *Supervised* and *Unsupervised ETF Alignment*, we evaluate four configurations: (1) Baseline without alignment, (2) supervised only, (3) unsupervised only, and (4) both enabled. Results in Table 4 show consistent gains from both modules. *Unsupervised ETF Alignment*. This component notably

boosts novel class discovery, improving new class accuracy by +4.6% on average. The improvement is especially prominent on fine-grained datasets, where modeling inter-class similarity is challenging. It guides confident samples toward unused ETF prototypes, enhancing discriminability.

Supervised ETF Alignment. This module improves retention, enhancing old class accuracy by +3.9%. By aligning labeled features to fixed ETF directions, it anchors known categories and reduces forgetting. While its effect on novel class accuracy is smaller, the joint use of both alignment strategies yields the best overall performance.

Influence of the α Coefficient. We examine the effect of the ETF scope coefficient α on CIFAR100. As shown in Fig. 3, increasing α consistently improves novel class accuracy, reaching 53.97% at $\alpha = 0.7$. All-class accuracy also peaks at 72.10% when $\alpha = 0.7$, while old class accuracy remains stable across the range. These results indicate that a proper choice of α enhances novel discovery without affecting old class retention.

Conclusion and Future Work

In this paper, we propose GOAL, a unified framework for Continual Generalized Category Discovery. By aligning features to fixed ETF prototypes and guiding novel class discovery through confidence-aware selection, GOAL achieves consistent improvements in both accuracy and stability across incremental sessions. Extensive experiments on four benchmarks demonstrate its superior performance in new class discovery and forgetting mitigation, especially under longer task horizons. In future work, we will explore adaptive ETF expansion for dynamically estimating the number of novel classes and extend the framework to multimodal settings with textual or semantic cues.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant No.U21B2048, No.62302382 and No.62576224, the Shenzhen Key Technical Projects under Grant CJGJZD20220517141605013 and the China Postdoctoral Science Foundation No.2024M752584.

References

- Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; and Joulin, A. 2021. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9650–9660.
- Chiaroni, F.; Dolz, J.; Masud, Z. I.; Mitiche, A.; and Ben Ayed, I. 2023. Parametric information maximization for generalized category discovery. In *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, 1729–1739.
- Choi, S.; Kang, D.; and Cho, M. 2024. Contrastive mean-shift learning for generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 23094–23104.
- Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 248–255.
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houshy, N. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Fang, C.; He, H.; Long, Q.; and Su, W. J. 2021. Exploring Deep Neural Networks via Layer-Peeled Model: Minority Collapse in Imbalanced Training. *Proceedings of the National Academy of Sciences*, 118(43).
- Graf, F.; Hofer, C.; Niethammer, M.; and Kwitt, R. 2021. Dissecting Supervised Contrastive Learning. In *Proceedings of the International Conference on Machine Learning*, 3821–3830.
- Han, J.; Wang, S.; He, Y.; Ding, C.; Wang, Q.; Gao, X.; Dong, S.; and Gong, Y. 2025. Consistent Supervised-Unsupervised Alignment for Generalized Category Discovery. *arXiv preprint arXiv:2507.04725*.
- Han, X.; Pappas, V.; and Donoho, D. L. 2022. Neural Collapse Under MSE Loss: Proximity to and Dynamics on the Central Path. In *International Conference on Learning Representations*.
- Huang, C.; Xie, L.; Yang, Y.; Wang, W.; Lin, B.; and Cai, D. 2023. Neural Collapse Inspired Federated Learning with Non-iid Data. *arXiv preprint, arXiv:2303.16066*.
- Krizhevsky, A.; and Hinton, G. 2009. Learning multiple layers of features from tiny images. Technical report, University of Toronto.
- Ma, S.; Zhu, F.; Zhong, Z.; Liu, W.; Zhang, X.-Y.; and Liu, C.-L. 2024a. Happy: A debiased learning framework for continual generalized category discovery. *Advances in Neural Information Processing Systems (NeurIPS)*, 37: 50850–50875.
- Ma, S.; Zhu, F.; Zhong, Z.; Zhang, X.-Y.; and Liu, C.-L. 2024b. Active generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16890–16900.
- Mixon, D. G.; Parshall, H.; and Pi, J. 2022. Neural collapse with unconstrained features. *Sampling Theory, Signal Processing, and Data Analysis*, 20(2): 11.
- Pappas, V.; Han, X.; and Donoho, D. L. 2020. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40): 24652–24663.
- Pu, N.; Zhong, Z.; and Sebe, N. 2023. Dynamic Conceptual Contrastive Learning for Generalized Category Discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7579–7588.
- Roy, S.; Liu, M.; Zhong, Z.; Sebe, N.; and Ricci, E. 2022. Class-incremental novel class discovery. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 317–333. Springer.
- Tian, Y.; Krishnan, D.; and Isola, P. 2020. Contrastive representation distillation. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Vaze, S.; Han, K.; Vedaldi, A.; and Zisserman, A. 2022. Generalized category discovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 7492–7501.
- Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The Caltech-UCSD Birds-200-2011 Dataset. Technical report, California Institute of Technology.
- Wang, Q.; He, Y.; Dong, S.; Song, X.; Han, J.; Luo, H.; and Gong, Y. 2025. DualCP: Rehearsal-Free Domain-Incremental Learning via Dual-Level Concept Prototype. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, 21198–21206.
- Wen, X.; Zhao, B.; and Qi, X. 2023. Parametric classification for generalized category discovery: A baseline study. In *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, 16590–16600.
- Wojtowitsch, S.; et al. 2020. On the emergence of simplex symmetry in the final and penultimate layers of neural network classifiers. *arXiv preprint arXiv:2012.05420*.
- Wu, Y.; Chi, Z.; Wang, Y.; and Feng, S. 2023. MetaGCD: Learning to continually learn in generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 1655–1665.
- Yang, B.; Lin, M.; Zhang, Y.; Liu, B.; Liang, X.; Ji, R.; and Ye, Q. 2022. Dynamic support network for few-shot class incremental learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 2945–2951.
- Zhang, S.; Khan, S.; Shen, Z.; Naseer, M.; Chen, G.; and Khan, F. S. 2023. Promptcal: Contrastive affinity learning

via auxiliary prompts for generalized novel category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3479–3488.

Zhang, X.; Jiang, J.; Feng, Y.; Wu, Z.-F.; Zhao, X.; Wan, H.; Tang, M.; Jin, R.; and Gao, Y. 2022. Grow and merge: A unified framework for continuous categories discovery. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 27455–27468.

Zhao, B.; Wen, X.; and Han, K. 2023. Learning semi-supervised gaussian mixture models for generalized category discovery. In *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, 16623–16633.

Zhong, Z.; Cui, J.; Yang, Y.; Wu, X.; Qi, X.; Zhang, X.; and Jia, J. 2023. Understanding imbalanced semantic segmentation through neural collapse. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19550–19560.

Zhou, D.-W.; Wang, F.-Y.; Ye, H.-J.; Ma, L.; Pu, S.; and Zhan, D.-C. 2022. Forward Compatible Few-Shot Class-Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9046–9056.