Prompt-SSLC: Prompting Semi-Supervised Sequential Leader Clustering for On-the-fly Category Discovery

Anonymous authors
Paper under double-blind review

Abstract

On-the-fly Category Discovery (OCD) enables intelligent systems to perform real-time predictions while adapting to emerging classes without full offline retraining. However, deploying OCD in dynamic environments faces three critical bottlenecks: prototype drift under continuous streams, static feature bias toward base classes, and severe latency caused by sample-by-sample clustering. To address these challenges, we propose **Prompt-SSLC**, a unified framework that tightly couples robust streaming clustering, adaptive prompting, and routing. First, **Semi-Supervised Sequential Leader Clustering (SSLC)** anchors labeled class priors and utilizes a Distance-Aware update rule to maintain explicit decision margins and prevent prototype overlap. Second, a **Dual Prompt** architecture decouples global task context from local feature alignment: a **Task Prompt** guides category discovery, while an **Instance Prompt** dynamically recalibrates representations using neighbor prototypes to prevent base-class collapse without parameter tuning. Third, an **Open-Set Aware (OSA) classifier** uses uncertainty estimation to instantly classify high-confidence known samples, routing only novel or ambiguous samples to SSLC and shielding prototypes from noisy known-class updates. This cohesive integration of streaming clustering, adaptive prompting, and routing establishes a robust framework that balances stability and adaptability. Extensive experiments across generic and fine-grained benchmarks show that Prompt-SSLC delivers strong, competitive accuracy while maintaining the low computational overhead required for real-time OCD deployment.

1 Introduction

While intelligent systems have made substantial progress in closed-world paradigms where testing data align with predefined training categories, open-world learning, which requires models to handle unseen categories or unfamiliar scenarios, has yet to achieve comparable success. Category discovery has emerged as an intriguing open-world learning problem, initially studied as Novel Category Discovery (NCD) (Han et al., 2019; 2021; Zhao & Han, 2021; Jia et al., 2021; Fini et al., 2021) to categorize unlabeled novel-class data by transferring knowledge from labeled data. Generalized Category Discovery (GCD) (Vaze et al., 2022; Zhang et al., 2023; Wen et al., 2023; Vaze et al., 2023; Hao et al., 2024; Wang et al., 2024c; Lin et al., 2024; Zhao et al., 2023; Wang et al., 2024b; Liu & Han, 2025; Liu et al., 2025b; Han et al., 2024) relaxes this setup to unlabeled data with both known and novel classes, while Continual Category Discovery (CCD) (Zhang et al., 2022; Cendra et al., 2024) extends discovery across incremental stages.

Despite encouraging progress, prior studies leave three practical issues under-addressed: **(i)** the model must operate on a collected batch of unlabeled data, whereas practice may present a single instance or a varying stream; **(ii)** the model must be re-optimized whenever unlabeled samples arrive; **(iii)** there is no explicit mechanism for separating novel classes from known ones within the unlabeled stream.

To bridge static evaluation and dynamic real-world operation, On-the-fly Category Discovery (OCD) (Du et al., 2023) has emerged as an appealing paradigm (see Fig. 1). OCD eliminates batch collection, detects emerging categories during continuous operation, and maintains

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

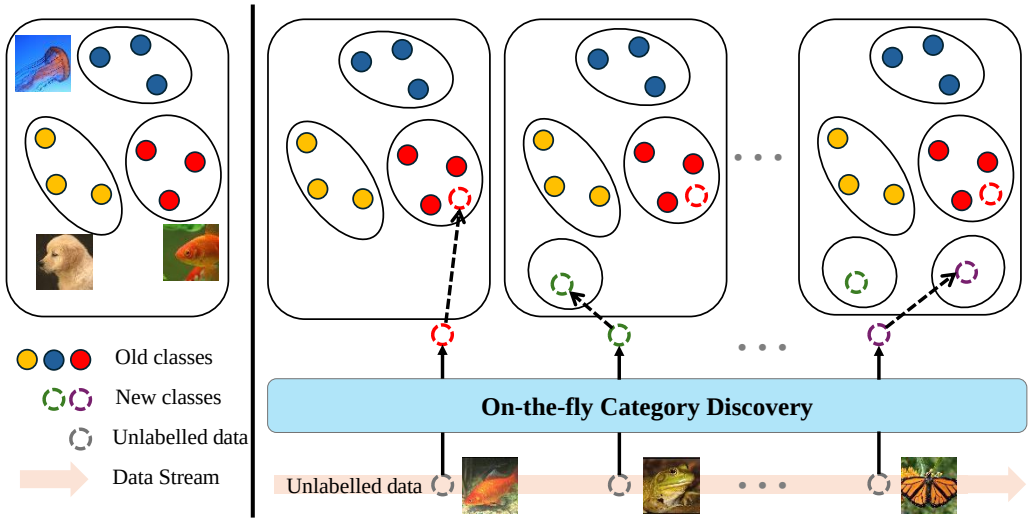


Figure 1: **On-the-fly Category Discovery (OCD).** (Left) During training, the model has access only to labeled data. (Right) During inference, it processes individual instances in real time, assigning them to existing categories or creating new ones as needed—without access to the full unlabeled dataset.

performance without iterative retraining. As an emerging task, OCD has attracted a sparse but growing literature, namely SMILE (Du et al., 2023), PHE (Zheng et al., 2025), DiffGRE (Liu et al., 2025a), AGE (Li et al., 2026), and TALON (Wu et al., 2026). SMILE Du et al. (2023) projects data into a low-dimensional hash space via supervised training on labeled data; novel categories appear as unobserved hash codes at inference. PHE Zheng et al. (2025) decouples performance from hash dimensionality to preserve inter-class discrimination, while DiffGRE Liu et al. (2025a) uses multimodal generative models to simulate novel classes. AGE Li et al. (2026) applies soft class thresholding for known categories and models class densities in a dynamic Gaussian space for real-time novel discovery. TALON Wu et al. (2026) is a hash-free, test-time adaptive OCD framework that updates prototypes and encoder parameters on streaming test data with margin-aware logit calibration. However, existing methods still struggle with three intrinsic challenges that map onto the issues above: **(1) Prototype drift and overlap** under unconstrained continuous streams; **(2) Static feature bias** toward base classes in pre-trained representations; and **(3) Prohibitive computational latency** from adapting and clustering every incoming instance.

To overcome these challenges, we propose **Prompt-SSLC**, a unified framework that tightly couples robust streaming clustering, adaptive prompting, and routing. First, inspired by Sequential Leader Clustering (SLC) Hartigan (1975), we propose Semi-Supervised Sequential Leader Clustering (SSLC), a prototype-driven approach that identifies novel categories in streaming data by comparing incoming patterns to learned class prototypes, enabling continuous adaptation without retraining while preserving labeled constraints. To combat prototype drift and inter-cluster overlap, SSLC incorporates a Distance-Aware (DA) update mechanism that enforces explicit decision margins between adjacent prototypes during online updates. Second, to exploit vision foundation models (*i.e.*, DINOv2 (Oquab et al., 2023)) without inheriting base-class feature bias, we introduce a dual-prompt design: a *Task Prompt* that guides category discovery, and an *Instance Prompt* that dynamically recalibrates features for each input without retraining. Third, to separate seen from unseen unlabeled samples and mitigate per-sample clustering bottlenecks, we introduce an Open-Set Aware (OSA) classifier that routes high-confidence base samples around clustering for immediate inference, while ambiguous or novel instances go to SSLC—reducing latency and shielding clusters from known-class noise. With these designs, Prompt-SSLC faithfully handles on-the-fly category discovery. We thoroughly evaluate Prompt-SSLC on public generic and fine-grained benchmarks. Notably, Prompt-SSLC surpasses PHE (Zheng et al., 2025) by about **10%** in average new-class accuracy on fine-grained benchmarks (CUB, Scars, Herbarium19) (Tab. 2) and outperforms SMILE (Du et al., 2023) by about **20%** in average new-class accuracy on coarse-grained datasets (Tab. 1), while remaining competitive with TALON (Wu et al.,

2026), which retains an edge on several coarse New metrics due to test-time adaptation, and maintaining strict real-time efficiency.

In summary, we make the following contributions: **(i)** We formulate the core bottlenecks of real-world OCD and introduce **Prompt-SSLC**, a unified framework for real-time open-world streaming. **(ii)** Prompt-SSLC is built on three complementary pillars that mirror the bottlenecks above: SSLC with a Distance-Aware update scheme for stable streaming clustering; dual-prompt learning (Task and Instance) for test-time feature recalibration; and an Open-Set Aware (OSA) classifier for efficient instance routing and prototype protection. **(iii)** We thoroughly evaluate Prompt-SSLC on public generic and fine-grained benchmarks, obtaining strong performance that is competitive with, and on many metrics exceeds, prior methods.

2 Related work

Category Discovery Category discovery addresses transferring knowledge from labeled data to categorize unlabeled data containing known and unknown categories. The task was initially studied as Novel Category Discovery (NCD) (Han et al., 2019); conventional NCD methods (Han et al., 2021; Zhao & Han, 2021; Jia et al., 2021; Fini et al., 2021) assume unlabeled data contain only novel classes. Generalized Category Discovery (GCD) Vaze et al. (2022) relaxes this by allowing both known and novel classes, motivating methods for parametric classification (Wen et al., 2023), positive-sample utilization (Hao et al., 2024), spatial prompt learning (Wang et al., 2024c), attention guidance (Lin et al., 2024), GMM-based representation learning (Zhao et al., 2023), debiased curriculum learning (Liu & Han, 2025), hyperbolic representation learning (Liu et al., 2025b), scene context prior (Peng et al., 2025), hierarchical contrastive learning (Rastegar et al., 2024), background tokens prune (Xu et al., 2025). Related settings include Continual Category Discovery (CCD) (Zhang et al., 2022; Cendra et al., 2024), Federated GCD (Pu et al., 2024), Active GCD (Ma et al., 2024), Semantic Category Discovery (Han et al., 2024), and On-the-fly Category Discovery (OCD) (Du et al., 2023). OCD processes streaming data (single or multiple instances) without further model training once trained on labeled data. SMILE (Du et al., 2023) formalizes OCD and employs a hash-based representation for category identification. PHE (Zheng et al., 2025) distinguishes known and unknown classes via prototype hypersphere estimation, while DiffGRE (Liu et al., 2025a) leverages multimodal and generative models for unknown class discovery. AGE Li et al. (2026) decouples sample processing via soft class thresholding for known categories and models class probability density functions in a dynamic Gaussian space to discover emerging novel categories in real time. TALON Wu et al. (2026) is a hash-free, test-time adaptive framework for OCD that continuously updates class prototypes and adapts encoder parameters on streaming test data while using margin-aware logit calibration to prevent category explosion.

Open-Set Recognition Open-Set Recognition (OSR) rejects samples from unfamiliar classes w.r.t. the training data at test time without categorizing unknowns (Scheirer et al., 2012); OpenMax Bendale & Boult (2015) first calibrated deep confidence scores for known/unknown separation. Subsequent work spans generative, reconstruction-based, and prototype-based strategies (Neal et al., 2018; Kong & Ramanan, 2021; Yoshihashi et al., 2019; Sun et al., 2020). As studied in (Vaze et al., 2021; Wang et al., 2024a), the Maximum Softmax Probability (MSP) of a trained classifier is a surprisingly strong OSR baseline; we therefore adopt MSP as the OSA scoring rule for its simplicity and effectiveness. By processing known and open-set samples separately, our approach limits interference with known-class accuracy while continuously discovering novel categories on the fly.

3 Method

We introduce the OCD formulation in Sec. 3.1. Next, we present Prompt-SSLC, designed for robust and efficient OCD. We first propose Semi-Supervised Sequential Leader Clustering Baseline (SSLC-B), a simple yet effective clustering baseline, in Sec. 3.2. This baseline is enhanced in Sec. 3.2 with a Distance-Aware (DA) update scheme to preserve discriminability between adjacent prototypes, yielding the SSLC algorithm. Building upon SSLC, we integrate a dual-prompt tuning strategy to enrich feature representations for emerging categories. In Sec. 3.3, we reformulate closed-set classification as a GCD-style objective using the *Task Prompt* to encode task-specific information. In Sec. 3.3, we further introduce the

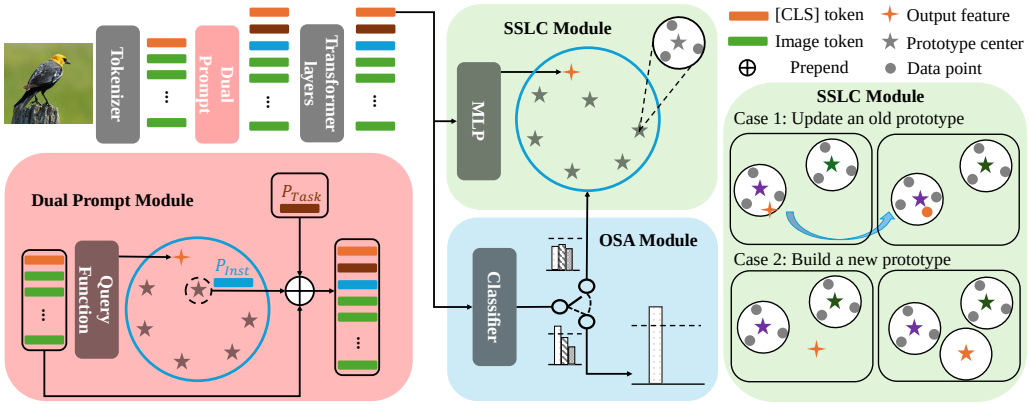


Figure 2: **Overview of Prompt-SSLC.** Dual Prompt prepends a Task Prompt (P_{task}) and an Instance Prompt (P_{inst}) to image patches. The Open-Set Aware (OSA) classifier routes high-confidence known-class samples for immediate prediction; uncertain or novel samples go to Semi-Supervised Sequential Leader Clustering (SSLC) for on-the-fly discovery and prototype updates.

Instance Prompt that captures neighboring prototype relationships. Finally, we design an Open-Set Aware (OSA) classifier (Sec. 3.4) that routes high-confidence known samples around SSLC—cutting per-sample clustering cost—while shielding prototypes from noisy known-class updates. The overview of our framework is shown in Fig. 2.

3.1 Formulation of OCD

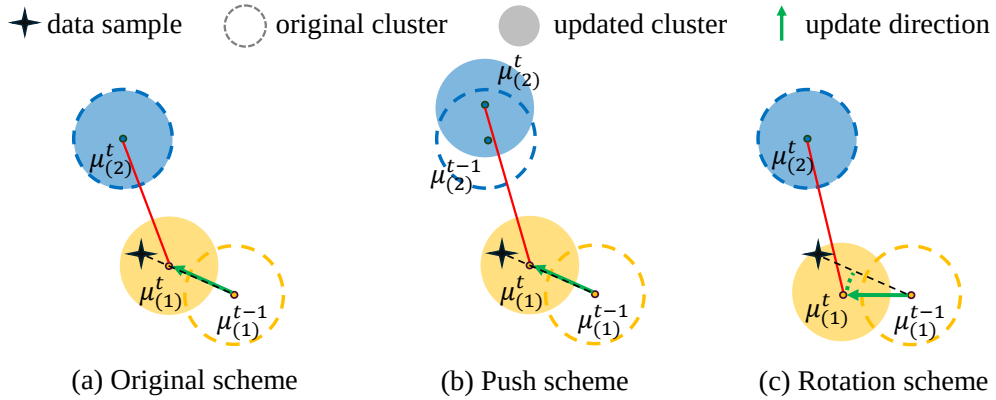
The OCD task leverages labeled data to perform generalized category discovery on streaming unlabeled data without retraining or fine-tuning on the incoming unlabeled data, enabling real-time identification of novel categories alongside classification of unlabeled instances. OCD operates through two distinct phases:

Training phase The model is trained on labeled data $D^{\mathrm{l}} = \{(x_i^{\mathrm{l}}, y_i^{\mathrm{l}})\}_{i=1}^{N}$, where $y_i^{\mathrm{l}} \in C_{\mathrm{old}}$ and $C_{\mathrm{old}} = \{1, 2, \ldots, c\}$ represents the set of known (or old) categories.

Inference phase The model processes an unlabeled data stream $D^{\mathrm{u}} = \{x_i^{\mathrm{u}}\}_{i=1}^{\infty}$ where samples belong to an expanded label space $C = C_{\mathrm{new}} \cup C_{\mathrm{old}}$. At each time step t, the cardinality $\hat{K}_t$ is dynamically estimated via $\hat{K}_t = g(\{x_i^u\}_{i=1}^{t}, C_{\mathrm{old}})$ where $g(\cdot)$ denotes an adaptive clustering mechanism that incrementally adjusts to incoming data. This formulation enables on-the-fly generalized category discovery under real-world constraints, where data arrive sequentially and novel classes emerge unpredictably over time. The approach is highly practical: it eliminates the need for model tuning during inference, for pre-collecting all unlabeled data, and for full dataset access, while continuously estimating new class counts.

3.2 Semi-Supervised Sequential Leader Clustering (SSLC)

Revisiting Sequential Leader Clustering. The objective of OCD is to predict the label $\hat{y}_i^{\mathrm{u}}$ for each unlabeled instance x_i^{u}. If the predicted label corresponds to an existing class in C, *i.e.*, $\hat{y}_i^{\mathrm{u}} \in C$, the instance is classified into that class. If the predicted label represents a novel class $\hat{y}_i^{\mathrm{u}} \notin C$, category discovery methods such as semi-supervised k-means (Vaze et al., 2022) and pseudo-label refinement (Han et al., 2021; Zhao & Han, 2021; Jia et al., 2021; Fini et al., 2021) prove unsuitable for online scenarios due to their reliance on static datasets and prior knowledge of total category counts. Sequential Leader Clustering (SLC) appears to be a natural baseline for OCD due to its sequential clustering nature. It operates through an online prototype update mechanism that dynamically creates clusters based on distance thresholds. For each incoming instance x_i^{u}, SLC calculates distances to all existing prototypes and assigns the instance to the nearest prototype if their distance falls below a predefined radius. When no prototype satisfies this criterion, SLC creates a new cluster centered at x_i^{u}. This approach eliminates the need for prior knowledge of category counts and supports streaming data processing. However, SLC operates purely unsupervised, disregarding available labeled data that could enhance cluster discriminability and alignment with known classes. To overcome this limitation, we propose Semi-Supervised Sequential Leader Clustering (SSLC), with two benefits: (1) labeled data guides prototype initialization



(a) Original scheme (b) Push scheme (c) Rotation scheme

Figure 3: **Distance-Aware (DA) update schemes.** (a) Original scheme adjusts only the nearest prototype. (b) Push scheme simultaneously attracts the nearest prototype and repels the second-nearest. (c) Rotation scheme enhances cluster separation while preserving the second-nearest prototype.

to avoid semantically inconsistent clusters, and (2) a Distance-Aware prototype update scheme maintains separation between adjacent prototypes.

Semi-Supervised SLC for OCD. We establish the Semi-Supervised Sequential Leader Clustering Baseline (SSLC-B) by **(1)** maintaining prototypes for known classes and **(2)** dynamically creating prototypes for new classes from streaming data, as outlined in Alg. 1. The prototype list is initialized using labeled data, where each known-class prototype is computed as the class-wise mean of feature representations. At each time step t, the system contains $\hat{K}_t$ prototypes, each defined by a center μ and a size n. For an incoming unlabeled sample x_t^{u}, the algorithm identifies the nearest prototype $(\mu_{(1)}, n_{(1)})$ with ℓ_2 distance $d_{(1)}$. If $d_{(1)}$ exceeds a pre-defined threshold τ, a new cluster is initialized with x_t^{u}. Otherwise, the data point is assigned to the nearest prototype, and its center and size are updated. Specifically, SSLC-B updates the center with $\mu_{(1)} = \mu_{(1)} + (x_t - \mu_{(1)}) * 1/(n_{(1)} + 1)$. By predefining prototypes for known classes through this supervised approach, SSLC-B emerges as a robust baseline that strategically preserves prior knowledge while enabling effective novel class discovery.

Algorithm 1 SSLC-Baseline (SSLC-B)

Initialize prototype list $\{\mu_i\}_{i=1}^{|C_{\mathrm{old}}|}$ with labeled data
Initialize $\hat{K}_t = |C_{\mathrm{old}}|$
for $t = 1, 2, \ldots$ **do**
  for $i = 1, 2, \ldots, \hat{K}_t$ **do**
    $d_i = ||x_t^{\mathrm{u}} - \mu_i||_2$
  end for
  Sort d in ascending order to have $d_{(1)}, d_{(2)}, \ldots$
  if $d_{(1)} \geq \tau$ **then**
    $\hat{K}_t \leftarrow \hat{K}_t + 1$
    $\mu_{\hat{K}_t} \leftarrow x_t^{\mathrm{u}}$ and $n_{\hat{K}_t} \leftarrow 1$
  else
    $\mu_{(1)} = \mu_{(1)} + (x_t - \mu_{(1)}) * 1/(n_{(1)} + 1)$
    $n_{(1)} = n_{(1)} + 1$
  end if
end for

Distance-Aware update scheme. The SSLC-B framework exhibits a critical limitation: unconstrained center adjustments under noisy data distributions induce progressive prototype overlap, as visualized in Fig. 3 (left). This degradation occurs when successive updates disproportionately attract prototypes toward outlier instances, collapsing inter-cluster margins. To address this, we incorporate inter-prototype distance constraints during center updates through two schemes, "push" and "rotation", which explicitly regulate prototype spacing to mitigate overlap while maintaining clustering integrity. **Push scheme.** This scheme simultaneously attracts $\mu_{(1)}$ toward x_t and repels $\mu_{(2)}$ from the observed data point, as formalized in Eq. (1).

$$\begin{aligned} \mu_{(1)} &= \mu_{(1)} + (x_t - \mu_{(1)}) * 1/(n_{(1)} + 1) \\ \mu_{(2)} &= \mu_{(2)} - (x_t - \mu_{(2)}) * 1/(n_{(2)} + 1) \end{aligned} \tag{1}$$

Rotation scheme. The rotation scheme regulates the distance between $\mu_{(1)}$ and $\mu_{(2)}$ by enforcing constrained movement, which establishes a rotational trajectory about $\mu_{(2)}$. This mechanism guides $\mu_{(1)}$ toward x_t while preserving inter-prototype spacing, as formalized in Eq. (2).

$$\mu_{(1)} = \mu_{(1)} + (x_t - \mu_{(1)}) * 1/(n_{(1)} + 1) + \lambda * (x_t - \mu_{(2)}) \tag{2}$$

Both update schemes produce significantly more distinguishable prototypes than SSLC-B's original mechanism, with the rotation scheme demonstrating superior performance by balancing prototype adaptation and separation preservation. Consequently, we adopt the **rotation** scheme to establish our SSLC framework.

3.3 Prompt tuning adaptation

***Task Prompt* for GCD simulation.** Existing OCD methods (Du et al., 2023; Zheng et al., 2025) rely on fully supervised training using labeled data composed solely of known classes, thereby reducing the problem to closed-set classification. This approach has a critical shortcoming: the resulting representations become over-specialized for known-class discrimination at the expense of foundation models' inherent generalization, while the closed-set objective misaligns with the OCD goal of novel class identification during inference. To address this limitation, we introduce (1) partial label masking that reformulates the training objective to preserve generalization, and (2) a task-aligned prompting strategy that activates the foundation model's inherent generalization capacity while maintaining discriminative power for known classes. To simulate GCD, partial label masking divides the original classes into two mutually exclusive subsets: C^{old}_{pseudo} and C^{new}_{pseudo}. The pseudo-old dataset comprises instances from C^{old}_{pseudo}, while the pseudo-new dataset contains instances from C^{new}_{pseudo} with their labels masked to reflect unknown categories. This approach enables controlled simulation of novel class discovery within a structured framework. Consequently, supervised contrastive loss (Khosla et al., 2020) is applied to the pseudo-old dataset and self-supervised contrastive loss (Chen et al., 2020) to the pseudo-new dataset ($L = L_{supcon} + L_{InfoNCE}$). Directly fine-tuning the model parameters would compromise the foundation model's inherent generalization capabilities. The *Task Prompt* addresses this by acting as lightweight vectors that prioritize task-relevant features, preserving the foundation model's generalization power and balancing known-class discrimination with novel-class sensitivity. To enhance task-specific learning alongside partial label masking, we integrate the *Task Prompt* (P_{task}) into foundation models. Our framework employs a Vision Transformer (ViT) as its backbone encoder. In this architecture, an input image X is partitioned into n patches $(x_1, \ldots, x_n)$. These patches are projected into a d-dimensional latent space via a shared linear layer e, yielding patch embeddings $E^0 = [e(x_1), \ldots, e(x_n)]$, which are combined with a learnable classification token `CLS` as the layer-0 input $[\texttt{CLS}^0, E^0]$. The attention mechanism is then expressed as:

$$[\texttt{CLS}^i, E^i] = L_i([\texttt{CLS}^{i-1}, E^{i-1}]) \tag{3}$$

where L_i is the i-th transformer layer. Inspired by Visual Prompt Tuning (VPT) (Jia et al., 2022), we prepend learnable Task Prompt vectors P_{task} at each transformer layer (deep prompts), matching the Instance Prompt insertion below. The attention mechanism is modified in Eq. (4)

$$[\texttt{CLS}^i, _, E^i] = L_i([\texttt{CLS}^{i-1}, P^{i-1}_{task}, E^{i-1}]). \tag{4}$$

***Instance Prompt* to hint prototype distribution.** Apart from P_{task} that encodes general task information, we propose the *Instance Prompt* (P_{inst}) to leverage near-neighbor prototype information and hint at the data distribution. Training involves two sequential feedforward passes. In the first pass, prompt parameters remain excluded from the input space, and the `CLS` token from the backbone network, denoted as $z_i = f(x_i)$, serves as the queried feature. The ℓ_2 distances between this feature and all cluster centers are computed, followed by identification of the k nearest clusters. The mean parameters of these clusters are concatenated to construct the *Instance Prompt*, that is, $P_{inst} = [\mu_{(1)}, \mu_{(2)}, \ldots, \mu_{(k)}]$. In the second feedforward pass, P_{inst} is integrated into each transformer layer L_i, modifying the input space as follows:

$$[\texttt{CLS}^i, _, E^i] = L_i([\texttt{CLS}^{i-1}, P^{i-1}_{task}, P^{i-1}_{inst}, E^{i-1}]). \tag{5}$$

Table 1: **Comparison with prior methods on coarse-grained datasets.**

Methods	CIFAR100			ImageNet100			CIFAR10		
	All	Old	New	All	Old	New	All	Old	New
SLC Hartigan (1975)	59.9	73.4	34.5	51.3	85.3	34.2	36.5	82.5	13.6
SMILE Du et al. (2023)	59.5	70.3	37.9	51.3	83.1	35.4	36.7	74.6	17.7
AGE Li et al. (2026)	73.6	83.6	52.1	67.4	90.0	58.2	51.3	89.4	32.5
TALON Wu et al. (2026)	75.9	86.2	**55.0**	**72.4**	90.1	**64.4**	**55.2**	89.9	**37.8**
SSLC-B	66.6	76.6	46.5	63.6	88.4	51.2	46.8	86.6	26.9
SSLC	72.2	82.4	51.9	66.5	88.7	55.5	49.6	87.5	30.6
Prompt-SSLC	**76.0**	**87.3**	53.5	70.4	**91.1**	60.0	53.3	**91.0**	34.5

Table 2: **Comparison with prior methods on fine-grained datasets.**

Methods	CUB			Scars			Herb		
	All	Old	New	All	Old	New	All	Old	New
SLC Hartigan (1975)	39.4	78.3	20.0	36.8	71.5	20.6	32.9	56.1	20.4
SMILE Du et al. (2023)	35.2	60.2	22.7	31.8	55.5	20.0	26.3	45.1	16.9
PHE Zheng et al. (2025)	39.4	69.1	24.5	37.2	65.4	23.1	31.3	52.3	20.8
AGE Li et al. (2026)	51.2	89.1	31.2	50.7	85.5	32.7	42.2	70.1	28.9
TALON Wu et al. (2026)	50.5	86.7	31.6	50.2	84.3	31.1	40.6	69.4	27.7
SSLC-B	50.4	88.3	31.5	49.0	85.3	30.8	34.7	58.7	21.8
SSLC	50.9	88.7	31.9	50.2	85.6	32.5	38.3	63.2	24.9
Prompt-SSLC	**52.4**	**90.1**	**33.5**	**50.9**	**86.5**	**33.2**	**45.0**	**72.2**	**30.1**

This design offers a unique advantage: the prototype pool can be extended dynamically during inference without requiring tuning. This capability stems from integrating P_{inst} with SSLC, where the expandable prototype list used in SSLC is shared with the prompt pool. As a result, the combined framework generates more discriminative features and achieves tighter cluster formations than non-prompt-based models.

3.4 Sample routing with Open-Set Aware classifier

To leverage open-set recognition for efficient routing, we introduce an Open-Set Aware (OSA) classifier that separates high-confidence known samples from ambiguous or novel ones. Inspired by (Vaze et al., 2021; Wang et al., 2024a), we adopt the Maximum Softmax Probability (MSP) as the OSA scoring rule, defined as $S(y \in C_{\text{old}}|x) = max_{y \in C_{\text{old}}} P(y|x)$ where $P(\cdot)$ is the softmax output vector. Unlabeled samples with sufficiently high OSA scores, indicating high confidence in belonging to a known class, are directly assigned to the corresponding old class with the highest softmax score. In contrast, low-confidence samples are treated as unknown and routed to the SSLC module for further analysis. This approach preserves the integrity of established prototypes while efficiently isolating novel class instances for subsequent processing.

4 Experiment

4.1 Experiment setup

Datasets. We evaluate on three coarse-grained datasets—CIFAR10 (Krizhevsky et al., 2009), CIFAR100 (Krizhevsky et al., 2009), and ImageNet100 (Russakovsky et al., 2015)—and three fine-grained ones: CUB (Wah et al., 2011), Scars (Krause et al., 2013), and Herbarium19 (Tan et al., 2019). Dataset statistics are summarized in Tab. 3. Following prior OCD splits (Vaze et al., 2022; Du et al., 2023), we sample C_{old} classes for the labeled set; about half of their images form D^{l}, and the remaining images from all classes form the unlabeled stream D^{u}.

Evaluation protocol. The model is trained on D^{l}, while D^{u} is shuffled and presented sequentially at inference. We use a modified Hungarian protocol: old-class prototypes are built from D^{l} features; Old accuracy is the fraction of D^{u} samples that match their old-class prototype, and New accuracy applies Hungarian matching to novel predictions. This separation reduces bias from previously learned class representations when scoring discovery quality.

Implementation details. We use a ViT-B backbone initialized from DINOv2 (Oquab et al., 2023), fine-tuning only the final transformer block. P_{task} and P_{inst} are inserted

Table 3: **Dataset statistics.** Numbers of classes ($|C|$, $|C_{\text{old}}|$) and samples ($|D^{\text{l}}|$, $|D^{\text{u}}|$).

	CUB	Scars	Herb	CIFAR10	CIFAR100	ImageNet100
$\lvert C_{\text{old}}\rvert$	100	98	341	5	80	50
$\lvert C\rvert$	200	196	683	10	100	50
$\lvert D^{\text{l}}\rvert$	1.5k	2.0k	8.9k	12.5k	20.0k	31.9k
$\lvert D^{\text{u}}\rvert$	4.5k	6.1k	25.4k	37.5k	30.0k	95.3k

Table 4: **Ablation study on the components of our approach.**

	Methods	Scars			IM100		
		All	Old	New	All	Old	New
Reference	SLC	36.8	71.5	20.6	51.3	85.3	34.2
	SSLC-B ***w.*** Original scheme	49.0	85.3	30.8	63.6	88.4	51.2
	SSLC-B ***w.*** Push scheme	49.4	85.4	31.4	65.5	88.5	54.0
	SSLC-B ***w.*** Rotation scheme (SSLC)	**50.2**	**85.6**	**32.5**	**66.5**	**88.7**	**55.5**
	SSLC-B ***w.*** Both schemes	49.1	85.4	31.0	64.4	88.4	52.4
Reference	Prompt-SSLC (Ours)	**50.9**	**86.5**	**33.2**	**70.4**	**91.1**	**60.0**
	w.o. Rotation scheme	49.8	86.5	31.5	67.1	90.5	55.4
	w.o. OSA classifier	50.4	85.8	32.7	66.8	88.9	55.7
	w.o. *Task Prompt*	50.6	85.8	33.0	68.5	90.0	57.7
	w.o. *Instance Prompt*	50.5	85.9	32.8	67.7	90.2	56.4

between the `CLS` token and patch tokens; an MLP head (4096-d $L2$-normalized) feeds SSLC, and a linear head produces old-class logits for OSA. Competitors SMILE (Du et al., 2023), PHE (Zheng et al., 2025), AGE (Li et al., 2026), and TALON (Wu et al., 2026) are reproduced with the same DINOv2 extractor (MLP dims 12/64 for SMILE/PHE). We train 100 epochs on fine-grained datasets and 50 on coarse-grained ones with Adam (β_1=0.9, β_2=0.999), initial LR 1e-4, cosine schedule, and batch size 128. Default P_{task} length and Instance Prompt top-k are both 3; rotation coefficient $\lambda = 0.05$.

4.2 Comparison with prior methods

We compare against SLC (Hartigan, 1975), SMILE (Du et al., 2023), PHE (Zheng et al., 2025), AGE (Li et al., 2026), and TALON (Wu et al., 2026) on all six benchmarks (Tabs. 1 and 2). SSLC-B improves over SLC across datasets by initializing known-class prototypes from labeled data. SSLC further improves New by about 3% on average over SSLC-B (Tab. 4) via the Distance-Aware update scheme. On fine-grained datasets (CUB, Scars, Herbarium19), Prompt-SSLC raises average New by about 10% over PHE (Zheng et al., 2025) and about 4% over SSLC-B (Tab. 2); PHE is reported on fine-grained settings only and is omitted from Tab. 1. On coarse-grained tasks, Prompt-SSLC gains about 20% average New over SMILE (Du et al., 2023) and about 9% over SSLC-B, and leads Old on CIFAR100, ImageNet100, and CIFAR10 (Tab. 1). TALON Wu et al. (2026) remains stronger on several coarse New (and some All) metrics due to test-time adaptation.

4.3 Ablation study

Relative to SSLC-B, Prompt-SSLC introduces four ingredients: a DA update scheme, P_{task}, P_{inst}, and an OSA classifier, validated in Tab. 4. Hyperparameter sensitivity is studied in Sec. 4.3. Further experiments like scalability across dataset and model size, robustness to domain shift are shown in Sec. A

Designs for DA scheme. We compare Push and Rotation Distance-Aware updates, and their combination, on top of SSLC-B (Tab. 4, top). Rotation improves New by about 1% over Push: Push can destabilize the second-nearest prototype under noisy inputs, whereas Rotation yields a more stable neighbor geometry. Combining Push with Rotation underperforms Rotation alone, because Push interferes with Rotation's goal of increasing inter-prototype separation while preserving local stability.

Effectiveness of different components. In the bottom block of Tab. 4, removing the OSA classifier drops Old by about 2%, since low-confidence samples are no longer filtered before prototype updates. Ablating P_{task} reduces All by about 1%, and removing P_{inst} reduces New by about 2%, supporting dual-prompt tuning.

Task Prompt design. We compare Task Prompt tokens with Adapter (Houlsby et al., 2019) (r=8) and LoRA (Hu et al., 2022) (r=4). On Scars and CIFAR100, Task Prompt

Table 5: **Task Prompt versus PEFT baselines (Adapter, LoRA).**

PEFT method	Scars			CIFAR100		
	All	Old	New	All	Old	New
Task Prompt	**50.9**	**86.5**	**33.2**	**76.0**	**87.3**	**53.5**
LoRA (Hu et al., 2022)	49.7	86.4	31.4	75.1	86.7	52.0
Adapter (Houlsby et al., 2019)	49.2	85.9	30.9	74.4	86.1	50.9

Table 6: **Sensitivity to Instance Prompt length k (nearest clusters).**

k	Scars			CIFAR100		
	All	Old	New	All	Old	New
1	50.6	86.0	32.9	75.6	87.0	53.2
3	50.9	86.5	33.2	76.0	87.3	53.5
5	50.9	86.7	33.1	76.1	87.6	53.2

Table 7: **Sensitivity to SSLC distance threshold τ (distance percentile).**

SSLC threshold τ	Scars			CIFAR100		
	All	Old	New	All	Old	New
90	50.8	84.5	34.0	75.1	85.6	54.3
95	50.9	86.5	33.2	76.0	87.3	53.5
99	50.0	88.4	30.8	76.3	88.3	52.5

Table 8: **Sensitivity to rotation coefficient λ.**

λ	Scars			IM100		
	All	Old	New	All	Old	New
0.1	50.0	84.4	32.7	66.6	88.1	56.0
0.05	50.2	85.6	32.5	66.5	88.7	55.5
0.02	49.8	85.8	31.6	66.2	88.9	55.2

improves All by about 1–2% and New by up to about 2% (Tab. 5), consistent with encoding category-discovery cues rather than adding generic PEFT capacity alone.

Effects of hyperparameters. We ablate Instance Prompt length k, the SSLC distance-threshold percentile τ, and the rotation coefficient λ (Tabs. 6 to 8). OSA uses Maximum Softmax Probability (MSP) as its scoring rule (Sec. 3.4): high-MSP unlabeled samples go to known classes and low-MSP samples to SSLC; MSP is distinct from the Euclidean radius τ inside SSLC. Defaults are k=3, τ at the 95th percentile of labeled sample–prototype distances at initialization, and λ=0.05. Shorter prompts limit adaptability while longer ones add cost with little gain; higher τ percentiles improve Old but hurt New; λ=0.02 weakens New, whereas $\lambda \in \{0.05, 0.1\}$ remain stable. Further robustness analyses appear in the appendix.

Computational overhead. Although the *Instance Prompt* uses two forward passes, the second pass is short because the prompt is injected after patch embedding. Tab. 9 shows the extra cost is under 40% of a single-pass baseline, remaining competitive with SMILE on CUB and ImageNet-100.

Table 9: **Inference time comparison on CUB and ImageNet-100 (IM-100).**

Method	CUB	IM-100
SMILE	37s	700s
Prompt-SSLC (single pass)	29s	530s
Prompt-SSLC	41s	760s

5 Conclusion

On-the-fly Category Discovery targets continuous unlabeled streams with emerging categories, but deployment is hindered by prototype drift, base-class feature bias, and clustering latency. Prompt-SSLC addresses this motivation by unifying stable streaming clustering, light-weight prompt adaptation, and selective open-set routing.

Taken together, the contributions are threefold. First, SSLC initializes labeled prototypes and uses Distance-Aware updates to maintain margins as unlabeled samples arrive. Second, dual Task and Instance Prompts adapt a frozen backbone to discovery cues and recalibrate samples without backbone updates. Third, OSA routes confident known samples directly and sends only ambiguous or novel cases to SSLC, reducing latency and prototype contamination. Together, these components align stability, adaptability, and efficiency with the demands of streaming discovery.

Experiments on generic and fine-grained benchmarks show strong performance gains on both fine-grained and corase-grained datasets, achieving state-of-the-art for OCD while maintaining real-time streaming efficiency. We discuss limitations and broader impact in Secs. D and E, and view Prompt-SSLC as an effective stepping stone toward scalable, open-world stream perception.

References

Abhijit Bendale and Terrance Boult. Towards open world recognition. In *CVPR*, 2015.

Fernando Julio Cendra, Bingchen Zhao, and Kai Han. Promptccd: Learning gaussian mixture prompt pool for continual category discovery. In *ECCV*, 2024.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning.*, 2020.

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

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

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

Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Autonovel: Automatically discovering and learning novel visual categories. *IEEE TPAMI*, 2021.

Kai Han, Xiaohu Huang, Yandong Li, Vaze Sagar, Jie Li, and Xuhui Jia. What's in a name? beyond class indices for image recognition. In *CVPR Workshops*, 2024.

Shaozhe Hao, Kai Han, and Kwan-Yee K. Wong. CiPR: An efficient framework with cross-instance positive relations for generalized category discovery. *Transactions on Machine Learning Research*, 2024.

John A Hartigan. *Clustering Algorithms*. Wiley, 1975.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, et al. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning.*, 2019.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 2022.

Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *ECCV*, 2022.

Xuhui Jia, Kai Han, Yukun Zhu, and Bradley Green. Joint representation learning and novel category discovery on single-and multi-modal data. In *ICCV*, 2021.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *NeurIPS*, 2020.

Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *ICCV*, 2021.

Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13)*, 2013.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Chunming Li, Shidong Wang, and Haofeng Zhang. Adaptive gaussian expansion for on-the-fly category discovery. In *ICLR*, 2026.

Haonan Lin, Wenbin An, Jiahao Wang, Yan Chen, Feng Tian, Mengmeng Wang, QianYing Wang, Guang Dai, and Jingdong Wang. Flipped classroom: Aligning teacher attention with student in generalized category discovery. *NeurIPS*, 2024.

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*, 2025a.

Yuanpei Liu and Kai Han. Debgcd: Debiased learning with distribution guidance for generalized category discovery. *ICLR*, 2025.

Yuanpei Liu, Zhenqi He, and Kai Han. Hyperbolic category discovery. *arXiv preprint arXiv:2504.06120*, 2025b.

Shijie Ma, Fei Zhu, Zhun Zhong, Xu-Yao Zhang, and Cheng-Lin Liu. Active generalized category discovery. In *CVPR*, 2024.

Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *ECCV*, 2018.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2023.

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.

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

Sarah Rastegar, Mohammadreza Salehi, Yuki M Asano, Hazel Doughty, and Cees GM Snoek. Selex: Self-expertise in fine-grained generalized category discovery. In *ECCV*, 2024.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 2015.

Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE TPAMI*, 2012.

Xin Sun, Zhenning Yang, Chi Zhang, Keck-Voon Ling, and Guohao Peng. Conditional gaussian distribution learning for open set recognition. In *CVPR*, 2020.

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

Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need? *ICLR*, 2021.

Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *ICCV*, 2022.

Sagar Vaze, Andrea Vedaldi, and Andrew Zisserman. No representation rules them all in category discovery. *NeurIPS*, 2023.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

Hongjun Wang, Sagar Vaze, and Kai Han. Dissecting out-of-distribution detection and open-set recognition: A critical analysis of methods and benchmarks. *IJCV*, 2024a.

Hongjun Wang, Sagar Vaze, and Kai Han. Hilo: A learning framework for generalized category discovery robust to domain shifts. *arXiv preprint arXiv:2408.04591*, 2024b.

Hongjun Wang, Sagar Vaze, and Kai Han. Sptnet: An efficient alternative framework for generalized category discovery with spatial prompt tuning. *ICLR*, 2024c.

Xin Wen, Bingchen Zhao, and Xiaojuan Qi. Parametric classification for generalized category discovery: A baseline study. In *ICCV*, 2023.

Yanan Wu, Yuhan Yan, Tailai Chen, Zhixiang Chi, ZiZhang Wu, Yi Jin, Yang Wang, and Zhenbo Li. Talon: Test-time adaptive learning for on-the-fly category discovery. In *CVPR*, 2026.

Qiyu Xu, Zhanxuan Hu, Yu Duan, Ercheng Pei, and Yonghang Tai. A hidden stumbling block in generalized category discovery: Distracted attention. In *ICCV*, 2025.

Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. Classification-reconstruction learning for open-set recognition. In *CVPR*, 2019.

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.

Xinwei Zhang, Jianwen Jiang, Yutong Feng, Zhi-Fan Wu, Xibin Zhao, Hai Wan, Mingqian Tang, Rong Jin, and Yue Gao. Grow and merge: A unified framework for continuous categories discovery. *NeurIPS*, 2022.

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

Bingchen Zhao, Xin Wen, and Kai Han. Learning semi-supervised gaussian mixture models for generalized category discovery. *ICCV*, 2023.

Haiyang Zheng, Nan Pu, Wenjing Li, Nicu Sebe, and Zhun Zhong. Prototypical hash encoding for on-the-fly fine-grained category discovery. *NeurIPS*, 2025.

Prompt-SSLC: Prompting Semi-Supervised Sequential Leader Clustering for On-the-fly Category Discovery

— Supplementary Material —

A Additional experiment and analysis

Results with variant DINO backbones. We perform the main experiments using the DINO backbone architecture and employ the Hungarian matching algorithm as implemented in GCD (Vaze et al., 2022) and SMILE (Du et al., 2023), distinct from the constrained variant described in our main paper. Our method is evaluated on six benchmark datasets against recent competitors including SLC (Hartigan, 1975), SMILE (Du et al., 2023), and PHE (Zheng et al., 2025). Results for coarse-grained and fine-grained tasks are presented in Table I and Table II. SSLC (Sec. 3.2) outperforms SMILE (Du et al., 2023) and PHE (Zheng et al., 2025) across all metrics, with an average 7% increase in New accuracy, indicating the effectiveness of the DA prototype update scheme. Specifically, Prompt-SSLC only underperforms TALON (Wu et al., 2026) on the ImageNet100 dataset, possibly due to the test time adaptation nature of TALON. For fine-grained datasets, Prompt-SSLC outperforms PHE (Zheng et al., 2025) by 10% in New accuracy and exceeds SSLC by 4%, demonstrating superior generalization to unseen fine-grained categories. For coarse-grained tasks, Prompt-SSLC surpasses SMILE (Du et al., 2023) by 15% in All accuracy. These consistent advances across granularities highlight the adaptability and robustness of our framework.

Table I: **Comparison with prior methods on coarse-grained datasets (DINO backbone).**

Methods	CIFAR100			ImageNet100			CIFAR10		
	All	Old	New	All	Old	New	All	Old	New
SLC Hartigan (1975)	44.4	59.0	15.1	32.9	86.6	5.2	41.5	58.3	33.3
SMILE Du et al. (2023)	51.6	61.6	31.7	33.8	74.2	13.5	49.9	39.9	54.9
PHE Zheng et al. (2025)	56.0	70.1	27.8	39.2	49.3	34.1	53.1	19.3	70.0
AGE Li et al. (2026)	60.8	75.8	30.7	48.2	87.1	28.6	-	-	-
TALON Wu et al. (2026)	64.7	77.4	39.3	82.6	92.0	63.4	65.0	46.1	79.3
SSLC	63.2	79.8	30.0	51.3	89.2	31.4	65.1	60.9	68.4
Prompt-SSLC	**65.4**	**82.3**	**32.1**	**53.9**	**90.2**	**35.7**	**68.2**	**62.4**	**70.2**

Table II: **Comparison with prior methods on fine-grained datasets (DINO backbone).**

Methods	CUB			Scars			Herb		
	All	Old	New	All	Old	New	All	Old	New
SLC Hartigan (1975)	31.3	48.5	22.7	24.0	45.8	13.6	14.9	27.4	8.1
SMILE Du et al. (2023)	32.2	55.8	22.9	26.2	46.7	16.3	22.9	39.3	14.1
PHE Zheng et al. (2025)	36.4	55.8	27.0	31.3	61.9	16.8	25.5	44.3	16.0
DiffGRE Liu et al. (2025a)	42.5	54.4	36.5	27.7	48.1	17.8	-	-	-
AGE Li et al. (2026)	46.3	59.8	39.4	34.8	62.7	21.3	30.9	54.7	18.1
TALON Wu et al. (2026)	43.6	63.5	33.6	37.0	68.1	22.0	-	-	-
SSLC	47.1	72.8	34.2	35.5	65.7	20.7	28.4	52.3	16.8
Prompt-SSLC	**49.4**	**74.6**	**36.9**	**37.7**	**67.5**	**22.8**	**31.2**	**56.7**	**19.5**

Scalability analysis. We provide a scalability analysis across two axes: dataset size and model size. For dataset size, we conduct experiments on the full ImageNet dataset

Table III: **Scalability across backbone sizes (ViT-S/B/L) on Scars.**

Backbone	Method	All	Old	New
ViT-S/14	SMILE	29.2	51.2	18.3
	Prompt-SSLC	47.8	79.8	31.8
ViT-B/14	SMILE	31.8	55.5	20.0
	Prompt-SSLC	50.9	86.5	33.2
ViT-L/14	SMILE	33.9	56.9	22.4
	Prompt-SSLC	53.9	88.3	36.7

Table IV: **Scaling to full ImageNet.**

	ImageNet		
	All	Old	New
SMILE	38.4	59.6	27.7
Prompt-SSLC	50.9	74.7	39.1

Table V: **Comparison with GCD methods (transductive unlabeled evaluation).**

Methods	CUB			IM100		
	All	Old	New	All	Old	New
GCD	47.1	55.1	44.8	72.7	91.8	63.8
SimGCD	61.0	66.0	58.6	81.1	90.9	76.1
Prompt-SSLC (labeled only)	43.9	73.9	28.9	69.8	93.5	56.3
Prompt-SSLC	50.2	74.8	38.8	73.3	94.3	62.7

to evaluate our Prompt-SSLC framework. Both Prompt-SSLC and SMILE use the DINO backbone. The results, as shown in Tab. IV, show that our method outperforms SMILE by over 10% in All accuracy, reinforcing the robustness and scalability of our approach on larger datasets. Regarding model size, we evaluate on the Scars dataset using the DINOv2 backbone with model sizes ranging from ViT-S/14 (21M) to ViT-L/14 (300M), as shown in Tab. III. Prompt-SSLC's All-class accuracy rises from 47.8 to 53.9 (+6.1%). In contrast, the competing method (SMILE) gains only 4.7% (29.2 to 33.9). These results confirm that Prompt-SSLC scales more effectively and benefits more from larger model capacities.

Comparison with GCD methods. We conduct comparative experiments with GCD (Vaze et al., 2022) and SimGCD (Wen et al., 2023) on CUB and ImageNet100 datasets. To ensure fair comparisons, all methods utilize the estimated number of categories as proposed in GCD (Vaze et al., 2022) and adopt the DINO backbone architecture. Two variants of Prompt-SSLC are evaluated: the first variant is trained exclusively on labeled data D^{l}, aligning with the OCD setting, while the second variant sees both D^{l} and D^{u} during training to assess GCD performance. The results (as shown in Tab. V and Tab. VI) demonstrate the effectiveness of our Prompt-SSLC. Prompt-SSLC learned on D^{l} achieves a 10% improvement in old-class accuracy over GCD and SimGCD, demonstrating robust classification capabilities. With access to both labeled and unlabeled training data, the new-class accuracy increases by 5% compared to the labeled-only variant, resulting in an overall accuracy that surpasses GCD by approximately 2%. This shows that Prompt-SSLC can approach GCD-level overall accuracy without collecting all unlabeled data for offline clustering.

Varying ratios of labeled to unlabeled classes. By default, the number of labeled classes is set to half the total number of classes. To evaluate the robustness of our Prompt-SSLC framework under different labeled-to-unlabeled class ratios, we conduct experiments on the Stanford Cars (Krause et al., 2013) and ImageNet-100 datasets. We test configurations including 10% labeled to 90% unlabeled, 30% labeled to 70% unlabeled, and the baseline 50% labeled to 50% unlabeled settings, presented in Tab. VII. These experiments show that Prompt-SSLC remains robust across diverse labeled-data proportions.

Prototype drift. Prototype drift is a known challenge in SLC and other clustering methods like K-Means, where older members may no longer align with updated cluster prototypes. To empirically validate the effectiveness of our DA update scheme in preventing drift, we conduct an experiment comparing cluster assignment consistency between SSLC and the original SLC. We measure the percentage of samples receiving different cluster assignments between the midpoint and the end of a streaming data sequence. The results show that SSLC reassigns only approximately 8% of samples, compared to over 14% for original SLC. This demonstrates that our DA update scheme significantly mitigates prototype drift, ensuring more stable and meaningful cluster assignments over time.

Threshold selection. We investigate the impact of replacing the global threshold value τ with class-specific thresholds. Another threshold selection strategy is to use cluster-specific thresholds for old classes while updating τ for novel classes. We adopt the common practice

Table VI: **Comparison with GCD methods on held-out unlabeled test data.**

Methods	CUB			IM100		
	All	Old	New	All	Old	New
GCD	52.5	56.2	46.7	78.1	91.9	64.5
SimGCD	64.3	66.7	59.3	84.3	91.2	77.4
Prompt-SSLC (labeled only)	50.8	73.9	30.0	75.7	93.8	59.3
Prompt-SSLC	57.2	75.2	41.4	79.1	94.7	63.9

Table VII: **Robustness under varying labeled-to-unlabeled class ratios (All / Old / New).**

Ratio (old / new)	Scars			IM100		
	All	Old	New	All	Old	New
50:50	50.9	86.5	33.2	70.4	91.1	60.0
30:70	43.5	75.8	27.4	63.8	83.4	54.0
10:90	35.0	64.4	20.4	56.1	74.6	46.9

of setting each threshold to the 95th percentile of the distances to the class prototype and compare the shared threshold versus cluster-specific threshold, as presented in Tab. VIII. The results show that cluster-specific thresholds yield a modest improvement over the global radius approach.

Effect of second-nearest prototype. To assess the effectiveness of the current second-nearest prototype rotation scheme, we conduct an ablation comparing it against the rotation scheme that incorporates the top-3 nearest prototypes, as presented in Tab. IX. Experimental findings indicate that using solely the second-nearest prototype consistently outperforms the top-3 aggregation approach across all evaluated accuracy metrics. Moreover, this simpler strategy substantially lowers the computational overhead associated with prototype updates, thereby improving overall efficiency without compromising predictive performance.

Parameter count. Tunable parameters of Prompt-SSLC mainly consist of backbone parameters and prompt parameters. The amount of backbone parameters is about 7.09 Million, which is the same as previous works (Du et al., 2023; Zheng et al., 2025). The extra prompt parameters (including both Task Prompt and Instance Prompt) account for 55296, which is only 0.71% of the backbone parameter. With only a small addition of prompt parameters, Prompt-SSLC achieves significant gains on all metrics compared to previous methods, demonstrating the efficiency and effectiveness of the prompt module.

Robustness to domain shift. To evaluate the robustness of Prompt-SSLC under distribution shifts between the labeled training data and the unlabeled streaming data, we conduct additional experiments on domain-shifted variants of the ImageNet-100 benchmark. The labeled dataset D^l is constructed exactly as in the main experiments using the standard ImageNet-100 images. For the unlabeled dataset D^u, we employ class-consistent subsets of ImageNet-R and ImageNet-C restricted to the 100 overlapping classes with ImageNet-100, which considers different artistic renditions and corruption types. With all other settings identical to the described configuration, we compare Prompt-SSLC against SMILE. Results for the ImageNet-R and ImageNet-C subsets are reported in Tab. X. Prompt-SSLC consistently outperforms SMILE under both stylistic and corruption-induced shifts, achieving absolute gains of approximately 9–10% in new-class accuracy, confirming that Prompt-SSLC maintains strong generalization in the presence of domain shift.

Extensions to other modality. To evaluate the applicability of Prompt-SSLC beyond purely RGB-based vision tasks, we conduct experiments in a depth continual learning setting using the NYU-Depth V2 dataset and another audio setting using the ESC-50 dataset. For the depth case, we adopt the 9-base and 4-novel class split. The model is first tuned on labeled RGB images of the base classes. In subsequent phases, it processes unlabeled RGB-depth pairs, with depth maps converted to grayscale and channel-concatenated with RGB for on-the-fly inference. The result is shown on Tab. XI. Prompt-SSLC surpasses the SMILE baseline by approximately 3% across all metrics, confirming the framework's effectiveness on depth modality. For the audio setting, we adopt the ESC-50 dataset consisting of 50 environmental sound classes. Half of the classes form the labeled set D^l,

Table VIII: **Shared versus cluster-specific threshold τ.**

Threshold	Scars			IM100		
	All	Old	New	All	Old	New
shared	50.2	85.6	32.5	66.5	88.7	55.5
cluster-specific	50.5	86.0	32.7	66.8	89.2	55.6

Table IX: **Effect of second-nearest versus top-3 prototype rotation.**

Update scheme	Scars			IM100		
	All	Old	New	All	Old	New
SSLC-B w. Original	49.0	85.3	30.8	63.6	88.4	51.2
w. Push	49.4	85.4	31.4	65.5	88.5	54.0
w. Push (top-3)	48.7	85.3	30.3	63.3	88.1	50.9
w. Rotation	50.2	85.6	32.5	66.5	88.7	55.5
w. Rotation (top-3)	48.9	85.3	30.6	63.7	88.2	51.5

while the remaining clips from all classes constitute the unlabeled streaming set D^u. Each audio clip is transformed into a log-mel spectrogram, resized to 224×224, and replicated across three channels to resemble visual input. The comparison of Prompt-SSLC and SMILE on audio modality is shown on Tab. XII. Prompt-SSLC outperforms the SMILE baseline by 8% on New accuracy. The framework requires no pipeline modifications to extend to audio, complementing the depth-modality results in Tab. XI and demonstrating that Prompt-SSLC's design is inherently cross-modal.

Convergence analysis. We provide a rigorous convergence analysis for the Semi-Supervised Sequential Leader Clustering (SSLC) module at the core of Prompt-SSLC. The analysis demonstrates that the clustering procedure terminates with a stable and finite set of prototypes. The following assumptions are made, all of which are satisfied in the experimental setting of the paper and in realistic on-the-fly category discovery deployments. We assume the unlabeled data stream $D^u = \{x_t^u\}_{t=1}^N$ is finite, with $N < \infty$. This assumption aligns directly with the evaluation protocol used. Under these assumptions, the SSLC algorithm—whether using the original update, the push scheme, or the rotation scheme—terminates after exactly N steps and returns a fixed finite set of prototypes whose positions no longer change. The proof proceeds as follows. Since the stream contains precisely N samples, the algorithm executes a single sequential pass over the data. At each step $t = 1, \ldots, N$, either a new prototype is created or an existing prototype is updated. A new prototype can be initialized at most once per sample, so the total number of prototypes satisfies $\hat{K}_N \leq N + |C_{\text{old}}|$. After the final sample has been processed, no additional inputs arrive. Consequently, the prototype count $\hat{K}_t$ stabilizes at $\hat{K}_N$, and all prototype centers μ_i and cluster sizes n_i remain constant for all subsequent time steps. The clustering output is therefore fixed after a finite number of steps. For completeness, we briefly consider the idealized case of an infinite stream ($N = \infty$) under the additional natural assumption that the number of novel categories remains finite. In such a scenario, there exists (almost surely) a finite random time T beyond which every arriving feature lies within distance τ of an existing prototype, implying almost-sure convergence of each μ_i to a stationary point.

Computational overhead. Inference wall-clock timings for SMILE and Prompt-SSLC (single- and dual-pass) on CUB and ImageNet-100 are reported in the main paper (Tab. 9).

Robustness to non-uniform stream ordering. Realistic on-the-fly streaming environments often present class-correlated or bursty data rather than uniformly shuffled data. To simulate these conditions, we evaluate the framework using Dirichlet-distributed sampling ($\alpha = 0.1$). As shown in Tab. XIII, Prompt-SSLC performs even better under class-correlated streams, increasing new class accuracy by approximately 4% on both Scars and ImageNet-100. When samples of the same class arrive in proximity, the SSLC module stabilizes prototypes more rapidly, minimizing drift.

Notation standardization. For absolute clarity, a standardized index of all key mathematical notations used in the formulation of the Prompt-SSLC architecture is provided in Tab. XIV.

B Further visualization

Table X: **Performance under domain shift on ImageNet-R and ImageNet-C subsets.**

Method	IM-R			IM-C		
	All	Old	New	All	Old	New
SLC	38.2	71.6	14.9	35.7	68.9	12.4
SMILE	41.7	69.8	19.5	39.5	67.1	18.9
Prompt-SSLC	**56.8**	**84.1**	**35.9**	**53.7**	**82.9**	**32.5**

Table XI: **Results on the depth modality (NYU-Depth V2).**

Method	NYU-Depth V2		
	All	Old	New
SMILE	37.2	44.5	32.3
Prompt-SSLC	40.2	50.2	36.2

DA update scheme prevents overlap. We analyze the inter-prototype distance distribution with and without the Distance-Aware (DA) update scheme by computing the ℓ_2 distance between each prototype and its nearest neighboring prototype on StanfordCar (Krause et al., 2013), as shown in Fig. I. The original SLC update scheme exhibits smaller inter-prototype distances, indicating prototype overlap and reduced discriminability. The rotation scheme effectively mitigates this issue, increasing inter-prototype distances to establish clearer decision boundaries.

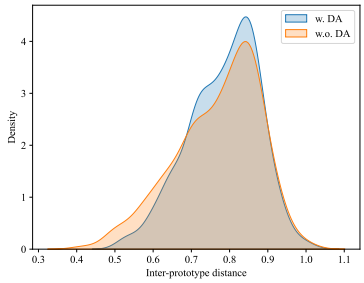


Figure I: **Inter-prototype distance distribution.** SSLC with the Distance-Aware (DA) update scheme yields larger inter-prototype distances and clearer decision boundaries.

Dual Prompt enhances representations. We analyze feature representations on CIFAR-10 (Krizhevsky et al., 2009) through t-SNE in Fig. II, comparing models with and without the dual-prompt tuning. The baseline lacking dual prompts exhibits degraded feature separation, causing significant overlap between seen and unseen classes under OCD. In contrast, the proposed dual-prompt mechanism yields more discriminative representations with tighter class clusters.

C Evaluation framework of main experiments

In this section, we formalize the constrained Hungarian matching that addresses feature bias toward old classes during evaluation. The core idea is to use D^{l} to find mapping between features and old categories and to utilize this mapping to guide Hungarian matching. (*e.g.*, in SMILE (Du et al., 2023) case, each old class is uniquely represented by a hash code feature.). We denote f_i for $i = 1, 2, \ldots, c$ to be unique features corresponding to each old class. For all features extracted from D^{u}, they are compared against all f_i to check if they belong to old classes. This feature is assigned to the old class if this mapping exists. Otherwise, remaining features perform a Hungarian matching with C_{new}. The old-class accuracy is defined as the percentage of old-class samples in D^{u} that match correctly to corresponding features relative to the total old-class samples. For novel categories, Hungarian matching is exclusively applied to align model predictions with ground-truth labels, enabling new class accuracy computation. Previous literature (Vaze et al., 2022; Du et al., 2023) computes accuracy by matching all extracted features across $C = C_{\text{new}} \cup C_{\text{old}}$. Importantly, this allows the well-built old clusters to be mapped to new categories. We argue that this biases measured performance and does not faithfully reflect the OCD setting.

D Limitations and future work

Results are reported primarily for a DINOv2 ViT-B backbone on standard OCD image benchmarks; defaults (k, τ, λ) are validated but not claimed universal across architectures or domains. Severe label noise, long-tailed arrivals, and strongly bursty class imbalance are only partially stress-tested (*e.g.*, Dirichlet stream-order experiments), and remain important

Table XII: **Results on the audio modality (ESC-50).**

Method	ESC-50		
	All	Old	New
SMILE	57.9	77.5	39.2
Prompt-SSLC	62.5	78.9	45.1

Table XIII: **Robustness under different stream orders.**

Stream	Scars			IM-100		
	All	Old	New	All	Old	New
Random	50.9	86.5	33.2	70.4	91.1	60.0
Dirichlet order ($\alpha = 0.1$)	53.8	87.4	37.0	72.7	90.9	63.5

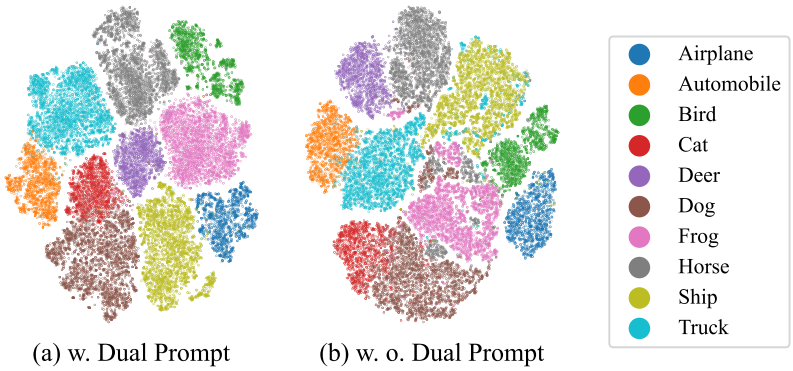
failure modes for future study. Extending Prompt-SSLC to continually expanding taxonomies with human-in-the-loop verification is a natural next step.

E Broader impact

On-the-fly category discovery can aid adaptive perception in robotics, monitoring, and scientific imaging by surfacing previously unseen categories without exhaustive labels. The same capability could enable unwanted surveillance or biased discovery if deployed without safeguards. Responsible use should prioritize consent, privacy, fairness audits, and human review of newly discovered categories before high-stakes action.

Table XIV: **Standardized definitions of mathematical notations.**

Notation	Definition
x_i^l, y_i^l	Labeled instance and corresponding ground-truth label.
C_{old}	Set of known categories.
D^u	Unlabeled data stream processed during the inference phase.
x_i^u	Individual unlabeled instance from the data stream.
$\hat{y}_i^u$	Predicted label for an unlabeled instance x_i^u.
C_{new}	Set of novel categories.
K	Ground-truth cardinality of all categories.
$\hat{K}_t$	Dynamically estimated cardinality of categories at step t.
μ_i	Prototype center for class i.
n_i	Number of instances assigned to prototype i.
λ	The rotation coefficient.
$d_{(1)}$, $d_{(2)}$	Distances to the nearest and second-nearest prototypes, respectively.
τ	SSLC Euclidean distance threshold (radius) to trigger creation of a new cluster.
$S(y \in C_{\text{old}} \mid x)$	OSA score via Maximum Softmax Probability (MSP); distinct from τ.
P_{task}	Task Prompt, which encodes general task-specific information.
P_{inst}	Instance Prompt, which captures neighboring prototype relationships.
k	Number of nearest clusters used to construct the Instance Prompt.



(a) w. Dual Prompt (b) w. o. Dual Prompt

Figure II: **t-SNE visualization of CIFAR-10 features.** (a) With Dual Prompt; (b) without Dual Prompt. Without Dual Prompt, Old- and New-class features exhibit substantial confusion.