SPARK: Structural Priors Activated from closed-set Recognition Knowledge for On-the-Fly Category Discovery

Anonymous authors
Paper under double-blind review

Abstract

On-the-Fly Category Discovery aims to identify known classes and continuously distinguish new classes from a streaming test sequence. Existing methods often entangle category discovery with backbone optimization by introducing discovery-specific modules or objectives that reshape the closed-set feature space. In this paper, we ask a simpler question: can a standard closed-set classifier already provide useful structural priors for online discovery? We answer this question affirmatively and propose **SPARK**, SPARK is a non-invasive, backbone-preserving extension of a closed-set classifier that avoids heavy discovery-specific neural modules and activates structural priors through lightweight anchors. Instead of treating known classes as isolated prototypes, SPARK learns *Local Manifold Anchors* around each known class. These anchors preserve inter-class discrimination while providing non-redundant local coverage of intra-class semantic variations. The union of all anchors forms a discrete approximation to a shared superclass manifold. At test time, when a streaming sample is rejected as unknown, SPARK analytically initializes its corresponding prototype under the constraint of this manifold and then updates the prototype online as more samples arrive. Experiments on seven public benchmarks show that SPARK achieves strong overall performance. These results suggest that on-the-fly category discovery does not necessarily require heavy specialized modules and the structural priors already present in closed-set classifiers can be activated for open-world recognition.

1 Introduction

Traditional visual recognition is largely built on the closed-set assumption (Krizhevsky et al., 2012; He et al., 2016), where training and test data share the identical label space. Fueled by large-scale, high-qualit annotations, this setting has underpinned the success of modern classification systems (Szegedy et al., 2013; Minaee et al., 2022). In real environments, however, visual data arrive continuously and the category space evolves over time (Vaze et al., 2022; Ma et al., 2024; Du et al., 2023). A practical model must not only recognize known classes, but also detect and handle previously unseen ones (Zheng et al., 2024; Li et al., 2026). This problem has motivated increasing interest in open-world visual recognition, whose central goal is to preserve discrimination over known classes while responding to the emergence of unknown categories and supporting continual expansion of the label space (Du et al., 2023; Zhao & Mac Aodha, 2023).

On-the-Fly Category Discovery (OCD) (Du et al., 2023) pushes the problem into a more practical online regime. The model is trained offline with only partially labeled data, and is then exposed to a temporally ordered stream of unlabeled samples that must be processed online, in a single pass, and without replay. Rather than fully exploiting the closed-set classifier, existing methods (Du et al., 2023; Zheng et al., 2024; Liu et al., 2025a) focus mainly on developing additional specialized modules to produce expandable descriptors for online identification of novel-classes.

Although such designs partially alleviate the difficulty of streaming discovery, their understanding of novel classes still relies heavily on mappings learned from the training data. Moreover, hashing-based representations are often sensitive to hyperparameters and offer limited interpretability, making it difficult to fully exploit the structure carried by incoming test samples themselves (Du et al., 2023;

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.

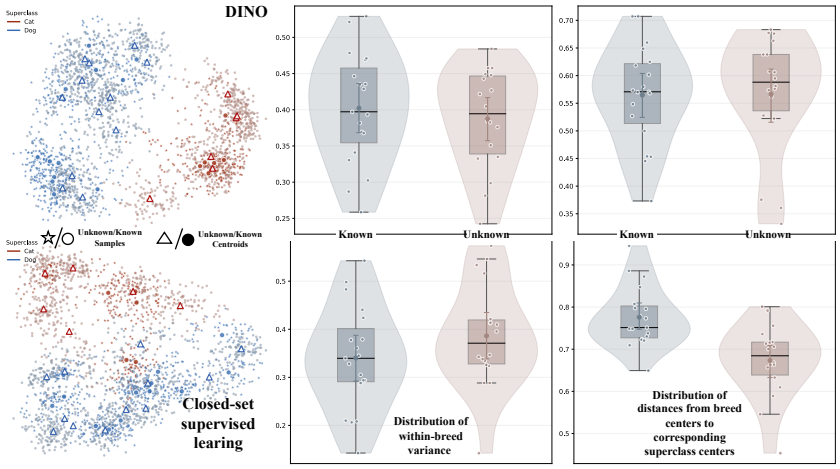

Figure 1: t-SNE / structure visualization and hierarchical statistics: closed-set supervision preserves superclass structure while reshaping fine-grained geometry.

Liu et al., 2025a; Zheng et al., 2024).In fact, additional discovery parameters may be unnecessary. Even under standard closed-set classification, the learned feature space preserves stable geometry and shared semantic structure, providing a structural prior for online category discovery, as shown in Fig. 1. We hypothesize that semantically related known classes are not geometrically isolated, but embedded in a shared superclass manifold (Chen et al., 2025). Therefore, when a novel category appears in the stream, its prototype should be initialized under this manifold constraint, rather than from a random vector or a single instance alone.

Building on this observation, we propose a novel framework that extends any closed-set classifier (Bridle, 1989) to the setting of OCD. Unlike existing methods that rely on extra parameters and auxiliary training objectives for backbones, our starting point is simpler: the feature space learned under standard closed-set supervision already contains structural priors that are useful for online discovering novel classes. In the penultimate feature space, the multiple local representations formed around known categories should not be viewed as isolated class-specific templates. Instead, taken together, they provide a discrete approximation to a shared yet fragmented superclass manifold, which captures higher-level semantic regions across classes. We estimate this shared manifold from the known local representations. When an input is identified as unknown, we solve a manifold-constrained optimization problem to obtain an analytical initialization for its novel-class prototype. We then combine prototype voting, threshold-based decision making, and online prototype-pool updates to unify known-class recognition and online novel-category discovery within a single framework. The resulting method requires no heavy auxiliary objectives on top of the base representation model. Instead, it directly activates the structural knowledge already encoded in the closed-set classifier, offering a simpler and more scalable path toward online open-world recognition.

Our main contributions are summarized as follows:

- We challenge the common belief that OCD requires dedicated modules for foundational representation models, and show that standard closed-set supervision already encodes structural priors for online category discovery.
- We propose a shared superclass-manifold view of features, in which local representations around known categories form a discrete approximation to a fragmented semantic manifold.
- We present a novel OCD framework that extends any closed-set classifier to OCD through manifold-constrained prototype initialization and online prototype adaptation.

2 Related work

2.1 Category Discovery

Category discovery (He et al., 2025) aims to leverage discriminative knowledge from known classes to separate unseen categories. Novel Class Discovery (NCD) (Han et al., 2022; Zhao & Han, 2021)

assumes that the unlabeled set contains only novel classes, while Generalized Category Discovery (GCD) (Vaze et al., 2022; Wen et al., 2023) considers the more realistic case where known and novel classes coexist. Although these methods have achieved strong results on static benchmarks (Liu & Han, 2025; Choi et al., 2024), most of them are still developed under a transductive setting. In this setting, the model can access the entire unlabeled set at once during training or adaptation and exploit its global structure through clustering (Bian et al., 2026), label propagation (Pu et al., 2023), or structural alignment (Rastegar et al., 2024). Such a formulation is well suited for offline discovery, but it does not naturally extend to streaming test scenarios in real-world deployment.

On-the-Fly Category Discovery (Du et al., 2023) targets a more realistic open-world setting. The model is trained with only partially labeled data and must recognize known classes, detect unknowns, and form new categories from a test stream. Existing methods (Du et al., 2023; Zheng et al., 2024) mainly relies on hash-based representations, typically using an MLP to map visual features into discrete hash codes. The appeal of this design is that unseen hash patterns can be naturally treated as novel classes. However, such methods often ignore the evolving structure of the incoming test data and are sensitive to hyperparameters such as code length and decision thresholds. DiffGRE (Liu et al., 2025a) extends this line by using a diffusion model (Rombach et al., 2022) to generate images corresponding to hash codes that are not assigned to known classes, thereby enriching the representation space for unknown categories in advance. However, AGE (Li et al., 2026) moves beyond hashing and mimics a Dirichlet process (Blei & Jordan, 2006) at test time to dynamically expand a Gaussian mixture model as new classes emerge.

2.2 Manifold Learning

Manifold learning (Lin & Zha, 2008; Zhang et al., 2012) is based on the classical assumption that, although data are observed in a high-dimensional space, their semantic variation is often governed by only a small number of latent factors. Samples therefore tend to concentrate near a low-dimensional manifold rather than uniformly fill the ambient space. Classical methods such as Locally Linear Embedding (LLE) (Roweis & Saul, 2000) and Laplacian Eigenmaps (Belkin & Niyogi, 2001) characterize this intrinsic geometry through local linear reconstruction and graph-based spectral embedding, respectively. Prior studies show that features learned with supervised training are not merely isolated class clusters in Euclidean space, but retain measurable and exploitable geometric organization (Ma et al., 2023; Kuoch et al., 2024). Although supervision improves discrimination by compressing class manifolds and reducing their effective radius and dimension, it does not fully erase the structure carried by intra-class variation and inter-class relations. Motivated by this observation, some works explicitly model feature-manifold geometry, while others introduce hyperbolic space when semantic relations exhibit hierarchy (Kwon et al., 2024).

These geometric ideas have also begun to shape category discovery. Recent methods move beyond cluster compactness and instead ask whether representation geometry is sufficient to support stable novel-class formation. For example, some approaches maximize token manifold capacity to preserve intra-class diversity and intrinsic structure; HypCD (Liu et al., 2025b) extend category discovery from Euclidean or spherical space to hyperbolic space to capture latent hierarchies between known and novel classes; BIA (Tang et al., 2026) interprets GCD as a geometric recovery problem for over-compressed representations and mitigates dimensional collapse via isotropy alignment. Taken together, these studies suggest that manifold geometry is a key structural foundation for understanding deep representations and advancing category discovery.

3 Method

3.1 Problem setup

The training set is defined as a labeled support set $\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^{N}$, where each label y_i belongs to the known class space $\mathcal{C}_k$. At test time, the model is exposed to a streaming sequence $\mathcal{D}_u = \{x_t\}_{t=1}^{T}$, whose samples may come from either the known class space $\mathcal{C}_k$ or an unknown class scape $\mathcal{C}_u$, with $\mathcal{C}_k \cap \mathcal{C}_u = \emptyset$. The complete class space is thus defined as $\mathcal{C} = \mathcal{C}_k \cup \mathcal{C}_u$.

In OCD, the model must process each incoming sample x_t sequentially, without revisiting previous data. For each sample, it is required to either recognize it as one of the known classes if $x_t \in \mathcal{C}_k$,

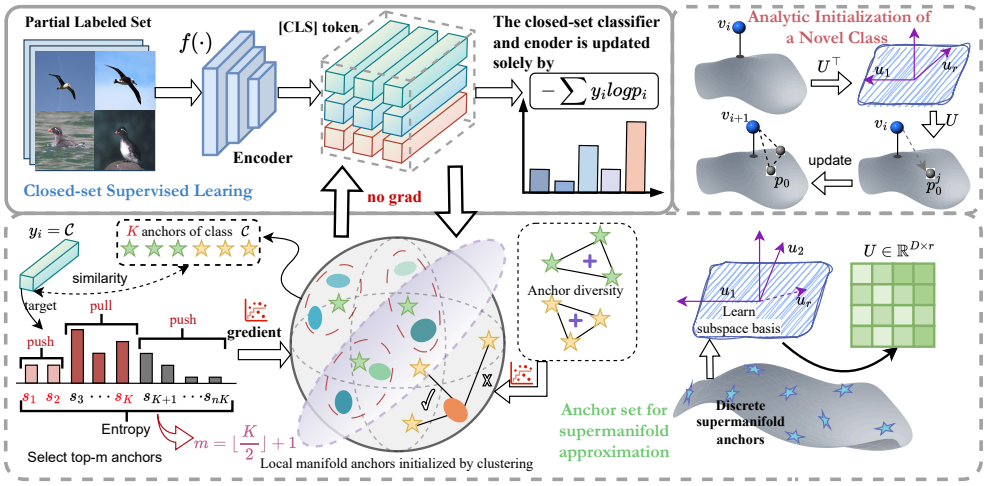

Figure 2: **Overview of the proposed framework.** Known-class features extracted by a standard closed-set classifier are used to construct Local Manifold Anchors without introducing extra gradient supervision on the backbone. By selecting top-m anchors and enforcing entropy and diversity constraints, the anchor set becomes a discrete approximation to a shared superclass manifold. The learned anchors further induce a low-dimensional subspace, which enables analytic initialization of a novel prototype when an unknown sample appears in the test stream, followed by online prototype update for on-the-fly category discovery.

or identify it as belonging to a novel class if $x_t \in \mathcal{C}_u$. Each instance x is encoded by the backbone into a feature vector $v \in \mathbb{R}^d$, followed by ℓ_2 normalization to obtain the final representation used for recognition and discovery.

3.2 Overview

Our method builds on a standard closed-set classifier and constructs Local Manifold Anchors from known-class features, without changing how the backbone and classifier are trained (as shown in Fig. 2). During training, the anchor set is estimated from known-class features, and class-specific thresholds are calibrated from the training data. At test time, each streaming sample is first examined by combining threshold-based gating with anchor voting to distinguish known from unknown classes. Known samples are directly assigned to their predicted known classes. For unknown samples, the model further determines whether they should be merged into an existing unknown prototype pool or instantiated as new categories. In the latter case, a novel prototype is analytically initialized from the Local Manifold Anchors and then updated online as new samples arrive.

3.3 Shared Superclass Manifold and Local Manifold Anchors

We model the closed-set feature space as a shared superclass manifold $\mathcal{M} \subset \mathbb{S}^{D-1}$ on the unit hypersphere. Under this view, known classes do not correspond to isolated point centers or mutually unrelated clusters. Instead, they occupy different local regions of the same shared semantic structure. For the c-th known class, we denote its local region by $\mathcal{M}_c \subset \mathcal{M}$. Therefore, the closed-set training data should be viewed as discrete samples from different local regions of a common high-level semantic manifold, rather than as a collection of unrelated class means.

Based on this formulation, we do not represent class c with a single class center. Instead, we associate it with a set of class-specific vectors

$$\mathcal{A}_c = \{a_{c,1}, a_{c,2}, \ldots, a_{c,K}\}, \qquad a_{c,k} \in \mathbb{R}^d. \tag{1}$$

These vectors are not intended to repeatedly approximate the same class center. Instead, they are expected to capture different local modes, variation directions, or semantic deformations on the class-local manifold $\mathcal{M}_c$. Accordingly, $\mathcal{A}_c$ can be viewed as a discrete support set for $\mathcal{M}_c$. The

union over all known classes, $\mathcal{A} = \bigcup_{c=1}^{|\mathcal{C}_k|} \mathcal{A}_c$, then forms a discrete approximation to the shared superclass manifold $\mathcal{M}$.

3.4 Learning Local Manifold Anchors under Closed-Set Supervision

As the global discriminative backbone, we retain a set of class-level weights as classifier and train them with standard closed-set supervision. Given a training sample (x_i, y_i), the classifier and encoder is optimized by the cross-entropy loss

$$\mathcal{L}_{\text{ce}} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{|\mathcal{C}_k|} y_{ic} \log \frac{\exp\big(\mathbf{w}_c^\top \mathbf{v}_i/\tau\big)}{\sum_{j=1}^{|\mathcal{C}_k|}\exp\big(\mathbf{w}_j^\top \mathbf{v}_i/\tau\big)}, \tag{2}$$

Here, $\mathbf{w}_c$ is the discriminative weight of class c and τ is the temperature. This objective provides a stable global decision boundary and a reliable discriminative space for subsequent anchor learning.

Next, we introduce class-specific local manifold anchor representations for modeling local structures. Specifically, let $N = |\mathcal{C}_k|$ denote the number of known classes, and let each class be associated with K local manifold anchors. We stack all anchors into

$$A = [a_{1,1}, \ldots, a_{1,K}, a_{2,1}, \ldots, a_{N,K}]^\top \in \mathbb{R}^{NK\times D}, \tag{3}$$

where $a_{c,k} \in \mathbb{R}^d$ is the k-th local manifold anchor of class c. The anchor-level similarities are

$$z^{\text{anc}}(x) = \frac{f^\top A^\top}{\tau_a} \in \mathbb{R}^{NK}. \tag{4}$$

We further reshape $z^{\text{anc}}(x)$ into $S(x) \in \mathbb{R}^{N\times K}$, where $S_c(x)$ contains the responses of x to the K anchors of class c.

Inspired by ZLPR (Su et al., 2022), instead of forcing a sample to align with all K anchors of its class, we only activate the top-m anchors within the ground-truth class, where $m = \lfloor \frac{K}{2} \rfloor + 1$. Let $\Omega(x) \subset \{1, \ldots, K\}$ and $|\Omega(x)| = m$ denote the indices of the top-m responses inside class y. Based on this set, we define a binary target vector $q(x) \in \{0,1\}^{NK}$ over all anchors, whose (c,k)-th entry is

$$q_{c,k}(x) = \mathbf{1}[c = y]\mathbf{1}[k \in \Omega(x)]. \tag{5}$$

That is, only the selected top-m anchors in the ground-truth class are treated as positives, while all other anchors are treated as negatives.

The Local Anchor Alignment Loss is defined as

$$\mathcal{L}_{\text{laa}}(x) = \log\Big(1 + \sum_{q_{c,k}(x)=0} \exp(z^{\text{anc}}_{c,k}(x))\Big) + \log\Big(1 + \sum_{q_{c,k}(x)=1} \exp(-z^{\text{anc}}_{c,k}(x))\Big). \tag{6}$$

Unlike a single-center formulation, this local selection mechanism only requires each sample to align with a small subset of anchors from its own class. This allows different anchors within the same class to specialize automatically to different local regions, variation directions, or semantic deformations. In other words, $\mathcal{L}_{\text{laa}}$ induces sparse alignment between samples and class-local support sets rather than collapsing the entire class to one point.

Local activation alone, however, does not guarantee effective coverage. Without additional regularization, different anchors may still converge to nearby locations and become redundant copies of one another. To prevent such collapse, we impose an intra-class dispersion constraint. For class c, let $\mathcal{A}_c$ denote its anchor set, and define its average pairwise distance as

$$d_c = \frac{1}{K(K-1)}\sum_{i\neq j}\|a_{c,i} - a_{c,j}\|_2, \qquad \mathcal{L}_{\text{div}} = \frac{1}{N}\sum_{c=1}^{N}\frac{1}{d_c + \varepsilon}. \tag{7}$$

We encourage anchor separation by minimizing $\mathcal{L}_{\text{div}}$. Consequently, $\mathcal{A}_c$ becomes closer to a discrete covering of the class-local manifold $\mathcal{M}_c$.

We further impose an entropy-based balancing term on the batch-wise anchor usage distribution. Let the average soft assignment over a batch of size B be

$$\bar{p} = \frac{1}{B}\sum_{i=1}^{B}\text{softmax}(z^{\text{anc}}(x_i)/\tau), \qquad \text{and} \qquad \mathcal{L}_{\text{me}} = \sum_{j=1}^{NK}\bar{p}_j \log \bar{p}_j \tag{8}$$

Minimizing this term drives $\bar{p}$ toward a uniform distribution and encourages all local manifold anchors to participate in training. In this sense, $\mathcal{L}_{\mathrm{div}}$ spreads anchors apart geometrically, while $\mathcal{L}_{\mathrm{me}}$ ensures that these separated anchors are actually used. Our final training objective is $\mathcal{L} = \mathcal{L}_{\mathrm{ce}} + \mathcal{L}_{\mathrm{laa}} + \mathcal{L}_{\mathrm{div}} + \mathcal{L}_{\mathrm{me}}$.

3.5 Online Recognition and Novel-Class Formation on the Shared Superclass Manifold

After training, we stack all known-class anchors into

$$\mathbf{A} = [\mathbf{a}_{1,1}; \ldots; \mathbf{a}_{C,K}] \in \mathbb{R}^{CK \times d}. \tag{9}$$

Because these anchors have already been optimized to be discriminative, non-redundant, and well activated, $\mathbf{A}$ is more than a collection of discrete prototypes. It can be viewed as a set of support points sampled from a shared superclass manifold. To obtain a compact representation of this structure, we define

$$\boldsymbol{\mu} = \frac{1}{CK} \sum_{c=1}^{C} \sum_{j=1}^{K} \mathbf{a}_{c,j}, \qquad \tilde{\mathbf{A}} = \mathbf{A} - \mathbf{1}\boldsymbol{\mu}^{\top}, \tag{10}$$

and compute the singular value decomposition

$$\tilde{\mathbf{A}} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{U}^{\top}. \tag{11}$$

Let

$$\mathbf{U}_r \in \mathbb{R}^{d \times r} \tag{12}$$

contain the top-r principal directions. The shared superclass manifold is then approximated by the affine subspace

$$\mathcal{M} = \{\boldsymbol{\mu} + \mathbf{U}_r \mathbf{h} \mid \mathbf{h} \in \mathbb{R}^r\}. \tag{13}$$

This step lifts the learned local manifold anchors from discrete class-wise structures to a global geometric prior for test-time inference.

Known-Class Recognition and Unknown Rejection. Given a test sample x_t with normalized feature $\mathbf{z}_t$, we first compute its similarities to all anchors and retrieve the top-k_v anchors globally. Mapping these anchors back to their class identities yields a candidate known class $\hat{y}_t$ by majority voting. Let ν_t be the corresponding vote count, and define the strongest in-class anchor response as

$$\rho_t = \max_j \qquad softmax(\mathbf{z}_t^{\top} \mathbf{a}_{\hat{y}_t,j} / \tau_r). \tag{14}$$

The sample is recognized as known if

$$\nu_t \geq \delta_v \qquad \text{and} \qquad \rho_t \geq \delta_{\hat{y}_t}. \tag{15}$$

where δ_v and $\delta_{\hat{y}_t}$ are determined adaptively based on the training set.

Analytic Novel Prototype Initialization. Once a sample is rejected as unknown, we do not initialize its prototype randomly, nor do we simply copy the current feature as a new center. Instead, we absorb it structurally on the shared superclass manifold by solving

$$\mathbf{p}_t^{\star} = \arg\min_{\mathbf{p}} \|\mathbf{p} - \mathbf{z}_t\|_2^2 + \lambda \left\|(\mathbf{I} - \mathbf{U}_r \mathbf{U}_r^{\top})(\mathbf{p} - \boldsymbol{\mu})\right\|_2^2. \tag{16}$$

The first term preserves fidelity to the current sample, while the second penalizes deviation from the manifold induced by known anchors. This problem admits the closed-form solution

$$\mathbf{p}_t^{\star} = \boldsymbol{\mu} + \mathbf{U}_r \mathbf{U}_r^{\top}(\mathbf{z}_t - \boldsymbol{\mu}) + \frac{1}{1+\lambda}(\mathbf{I} - \mathbf{U}_r \mathbf{U}_r^{\top})(\mathbf{z}_t - \boldsymbol{\mu}). \tag{17}$$

Hence, a novel prototype is not created as an unconstrained new center. It is initialized analytically under the geometric prior defined by the learned local manifold anchors.

Table 1: Comparison with state-of-the-art methods on five benchmark datasets.

Method	CUB-200			Herbarium19			Stanford-Cars			Pets			Food101		
	all	old	new	all	old	new	all	old	new	all	old	new	all	old	new
SLC	28.6	44.0	20.9	14.9	27.4	8.1	14.0	23.0	9.7	35.5	41.3	33.1	20.9	48.6	6.8
MLDG	29.5	48.4	20.1	20.8	36.7	12.3	24.0	41.6	15.4	—	—	—	—	—	—
RankStat	21.2	26.9	18.4	13.8	20.6	10.2	14.8	19.9	12.3	33.2	42.3	28.4	22.3	50.7	7.8
WTA	21.9	26.9	19.4	14.6	21.2	11.1	17.1	24.4	13.6	35.2	46.3	29.3	18.2	40.5	6.1
SMILE	32.2	50.9	22.9	22.9	39.3	14.1	26.2	46.7	16.3	41.2	42.1	40.7	24.0	54.6	8.4
PHE	36.4	55.8	27.0	22.5	38.5	14.0	31.3	61.9	16.8	48.3	53.8	45.4	29.1	64.7	11.1
DiffGRE	42.5	54.4	36.5	—	—	—	27.7	48.1	17.8	49.6	50.1	49.3	—	—	—
AGE	46.3	59.8	39.4	30.9	54.7	18.1	34.8	62.7	21.3	61.5	66.5	58.8	30.5	70.0	10.4
ours	45.7	68.7	34.2	31.5	59.1	16.6	35.9	55.7	26.1	65.2	61.1	67.4	32.7	65.6	15.9

Online Unknown Prototype Pool Update. During testing, we maintain an unknown prototype pool $\mathcal{P}_U$. If the current unknown sample is sufficiently similar to an existing unknown prototype, it is merged into that prototype; otherwise, a new unknown class prototype is created. Let $\mathbf{p}_u^{(t)}$ denote the prototype of unknown class u at time t. Its online update is

$$\mathbf{p}_u^{(t)} = \mathrm{Norm}\left(\alpha \mathbf{p}_u^{(t-1)} + (1-\alpha)\mathbf{z}_t\right). \tag{18}$$

This yields a unified test-time procedure for known-class recognition, unknown rejection, and continual novel-class formation. More importantly, the entire process is built directly on the local manifold anchors learned during training. It does not require redesigning the closed-set classifier or introducing an additional offline discovery module, thereby naturally turning a standard closed-set model into a system for on-the-fly category discovery.

4 Experiment

To comprehensively evaluate the effectiveness of the proposed method, we conduct extensive experiments on seven public benchmark datasets. Specifically, our evaluation covers two coarse-grained datasets, i.e., CIFAR100 (Krizhevsky et al., 2009) and ImageNet-100 (Krizhevsky et al., 2012), and five fine-grained datasets, including CUB-200 (Wah et al., 2011), Herbarium19 (Tan et al., 2019), Stanford-Cars (Krause et al., 2013), Oxford-IIIT Pets (Parkhi et al., 2012), and Food101.

Table 2: Comparison with state-of-the-art methods on CIFAR100 and ImageNet-100.

Method	CIFAR100			ImageNet-100		
	all	old	new	all	old	new
SLC	44.4	59.0	15.1	32.9	86.6	5.2
MLDG	50.6	61.0	29.8	30.6	72.3	9.7
RankStat	35.0	44.0	17.0	31.1	73.3	9.8
WTA	40.8	52.9	16.7	30.8	72.9	9.7
SMILE	51.6	61.6	31.7	33.8	74.2	13.5
PHE	55.9	70.0	27.7	34.1	80.6	10.8
AGE	60.8	75.8	30.7	48.2	87.1	28.6
ours	62.5	74.9	37.8	55.6	86.6	40.0

For comparison, we consider a wide range of state-of-the-art methods for OCD, including DiffGRE, PHE, SMILE, and AGE. In addition, we include SLC (Hartigan, 1975), MLDG (Li et al., 2018), RankStat (Han et al., 2020), and WTA (Jia et al., 2021) as complementary baselines to provide a more comprehensive comparison. By systematically comparing against these strong baselines, we aim to verify whether our method can preserve the discriminative ability for known categories while more effectively discovering, initializing, and absorbing novel categories that emerge from the online test stream.

4.1 Implementation details

We adopt ViT-B/16 (Dosovitskiy et al., 2021) pretrained with DINO (Caron et al., 2021) on ImageNet-1K as the backbone. For fair comparison, we follow the setting used by existing state-of-the-art methods and fine-tune only the last-block of the backbone. All models are trained for 100 epochs. The initial learning rate is set to 0.1 and is decayed to 5×10^{-5} with a cosine schedule. The temperature parameters τ, τ_a, and τ_r are set to 0.1, 0.2, and 0.7, respectively. The subspace dimension is set to 128, and the penalty coefficient λ for analytical manifold regularization is set to 10. The momentum coefficient for online prototype updating is set to 0.7. We provide a detailed analysis of the sensitivity to these hyperparameters in the appendix and the ablation study.

Table 3: Ablation study of training-stage objectives on CUB-200 and CIFAR100.

No.	$\mathcal{L}_{me}$	$\mathcal{L}_{div}$	CUB-200			CIFAR100		
			all	old	new	all	old	new
1	✓	×	39.3	48.6	34.6	57.0	67.5	34.7
2	×	✓	43.4	65.4	32.4	60.6	76.5	29.0
3	✓	✓	45.7	68.7	34.2	62.5	74.9	37.8

Table 4: Ablation study of test-time modules on CUB-200 and CIFAR100.

No.	δ_v	$\delta_{\hat{y}_t}$	Init.	CUB-200			CIFAR100		
				all	old	new	all	old	new
1	×	×	×	36.7	53.2	28.4	59.6	70.1	38.6
2	✓	×	×	41.7	71.1	27.0	60.0	75.4	29.4
3	×	✓	×	43.9	65.2	33.3	59.4	75.3	27.5
4	✓	✓	×	43.6	62.6	31.1	59.9	75.3	29.1
5	×	✓	✓	44.8	67.7	33.4	60.9	78.9	25.1
6	✓	×	✓	43.5	69.8	30.3	62.5	74.9	37.8
7	✓	✓	✓	45.7	68.7	34.2	62.5	74.9	37.8

Following the OCD evaluation protocol, we use clustering accuracy as the main metric. It computes the accuracy after optimally matching predicted clusters to ground-truth categories via the Hungarian algorithm:

$$\mathrm{Acc} = \max_{\pi \in \Pi_K} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{ y_i = \pi(\hat{y}_i) \right\}, \tag{19}$$

where $\hat{y}_i$ and y_i denote the predicted cluster and ground-truth label of x_i, respectively. Π_K is the set of all permutations over K categories, where K is the ground-truth or estimated number of categories.

4.2 Comparison with the state-of-the-art

Tables 1 and 2 present a comprehensive comparison between our method and representative baselines on seven benchmark datasets. Overall, our method achieves the best or highly competitive overall accuracy on most datasets, demonstrating the effectiveness of exploiting structural priors from a closed-set classifier for OCD. Compared with AGE, our method obtains higher overall accuracy on six datasets, with an average improvement of 2.8%. These results suggest that closed-set supervised training already induces useful geometric structure for OCD, and our method leverages this structure without altering the original training process of closed-set supervision.

On CIFAR100, ImageNet-100, Stanford-Cars, Pets, and Food101, our method consistently improves the accuracy on novel categories over AGE, with an average gain of 7.5%. This verifies that local manifold anchors provide effective structural support from known categories, allowing the model to initialize and update novel-class prototypes more accurately when unknown samples emerge in the test stream.

It is worth noting that the CIFAR100 result reported in the main table is obtained under the same hyperparameter setting used for all datasets, without dataset-specific tuning. Under this unified setting, our method already improves the overall accuracy on CIFAR100 to 62.5%, outperforming AGE with 60.8%. Furthermore, when the online prototype updating hyperparameters are selected according to the validation set, setting $\alpha = 0.9$ and $m = 3$ further improves the overall accuracy on CIFAR100 to 65.5%.

4.3 Ablation study

Effects of Training-stage Objectives. Table 3 studies the effects of the two training-stage objectives. When only $\mathcal{L}_{me}$ is used, the model maintains reasonable novel-category recognition, but its old-category accuracy and overall performance are limited. When only $\mathcal{L}_{div}$ is adopted, the old-category accuracy is clearly improved, while the novel-category performance tends to degrade. This indicates that the two objectives play complementary roles. The entropy-based objective encourages the anchors to be sufficiently activated by the data, while the diversity objective promotes discriminative and non-redundant local coverage.

Effects of Test-time Modules. Table 4 evaluates the contribution of each test-time module. Using any single module alone leads to an imbalanced trade-off between recognition and discovery. Analytical initialization improves the reliability of novel prototype construction, while voting and threshold-based decisions help suppress erroneous absorption and prototype drift. The full model achieves the best overall and novel-category accuracy on both datasets, showing that these modules are complementary for stable online category discovery.

Table 5: Ablation study on whether anchor gradients are back-propagated to the backbone.

Setting	CUB			CIFAR100		
	All	Old	New	All	Old	New
w/ grad	35.2	70.5	17.5	60.2	73.6	33.6
w/o grad	45.7	68.7	34.2	62.5	74.9	37.8

Table 6: Ablation study on different sources for subspace construction.

Source	CUB			CIFAR100		
	All	Old	New	All	Old	New
Class Prototypes	36.5	56.3	26.6	62.2	73.2	**40.1**
Instance Features	36.4	55.6	26.9	61.8	73.2	39.1
Local Manifold Anchors	45.7	68.7	34.2	62.5	74.9	37.8

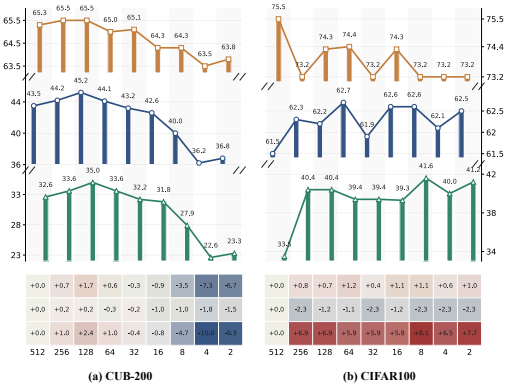

Figure 3: The impact of different sub-space dimensions on Performance

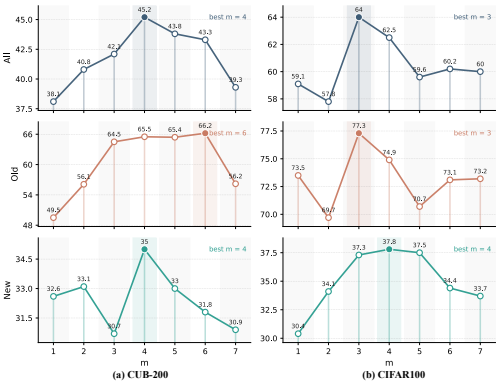

Figure 4: The impact of different m on Performance

Effects of Anchor Gradients. Table 5 investigates whether the gradients from local manifold anchors should be back-propagated to the backbone. The results show that stopping the anchor gradients consistently outperforms directly updating the backbone with them. On CUB-200, the overall accuracy increases from 35.2 to 45.7, and the accuracy increases from 17.5 to 34.2 on New. On CIFAR100, the overall accuracy also improves from 60.2 to 62.5. This validates a key principle of our design: local manifold anchors should serve as structural analyzers on top of the closed-set feature space, rather than perturbing the original closed-set supervised training process.

Effects of Subspace Construction. Table 6 compares different sources for constructing the analytical subspace. Directly using class prototypes or instance features makes it difficult to achieve a good balance between overall performance and fine-grained novel-category discovery, especially on CUB-200. In contrast, local manifold anchors achieve the best overall accuracy on both datasets and substantially improve both old- and novel-category accuracy on CUB-200. Local manifold anchors provide a more stable structural abstraction between these two extremes.

Effects of Hyperparameters. We further study the sensitivity of key hyperparameters in subspace construction and online prototype updating (as shown in Fig. 3 and Fig. 4). For d, performance is strongest around 128, suggesting that a moderate subspace preserves useful manifold structure while filtering redundant directions. For m, both too small and too large neighborhoods degrade performance: the former provides insufficient local support, while the latter may introduce irrelevant categories. We use $d = 128$ and $m = 4$ by default, unless otherwise specified.

5 Conclusion

We revisited whether On-the-Fly Category Discovery must rely on additional discovery-specific modules beyond a closed-set classifier. Our study shows that standard closed-set supervision does not merely learn isolated decision boundaries. It also preserves shared semantic geometry in the feature space, which can serve as a structural prior for open-world recognition. Motivated by this observation, we introduced **SPARK**, which organizes local variations around known classes with *Local Manifold Anchors* and uses them as a discrete approximation to a shared superclass manifold. At test time, this manifold provides a geometric prior for unknown samples. Instead of randomly initializing a novel prototype or copying a single instance, SPARK obtains an analytical initialization under a manifold-consistency constraint and then updates the prototype pool online.

AI Use Statement

In this work, we used generative AI tools to assist with the preparation and refinement of the manuscript, including language polishing, translation, improving clarity and readability, restructuring selected sections, and formatting LaTeX content. We also used generative AI tools to assist with literature search, summarization of related work, and limited code editing and debugging during experimentation.

Generative AI tools were not used to autonomously generate experimental results, fabricate data, or make final scientific conclusions. The core research ideas, methodological design, experimental protocol, and interpretation of the reported results were developed and validated by the authors.

All AI-assisted text, code, references, and technical content were reviewed by the authors. Suggested references were manually checked against the original publications, AI-assisted code was inspected and tested before use, and all reported experimental results were obtained from our own implementations and experiments. We additionally verified the consistency between the manuscript descriptions and the actual experimental settings and implementations.

We take full responsibility for the final content of this work, including all text, claims, experimental results, and artifacts produced with the aid of generative AI tools.

References

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NeurIPS*, volume 14, 2001.

Yuwei Bian, Shidong Wang, Yazhou Yao, and Haofeng Zhang. Foundation-adaptive integrated refinement for generalized category discovery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(4):2453–2461, Mar. 2026.

David M. Blei and Michael I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1):121 – 143, 2006.

John Bridle. Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In *NeurPIS*, volume 2, 1989.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, pp. 9650–9660, 2021.

Huitong Chen, Yu Wang, Yan Fan, Guosong Jiang, and Qinghua Hu. Reducing class-wise confusion for incremental learning with disentangled manifolds. In *CVPR*, pp. 10121–10130, June 2025.

Sua Choi, Dahyun Kang, and Minsu Cho. Contrastive mean-shift learning for generalized category discovery. In *CVPR*, pp. 23094–23104, June 2024.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021.

Ruoyi Du, Dongliang Chang, Kongming Liang, Timothy Hospedales, Yi-Zhe Song, and Zhanyu Ma. On-the-fly category discovery. In *CVPR*, pp. 11691–11700, June 2023.

Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically discovering and learning new visual categories with ranking statistics. In *ICLR*, 2020.

Kai Han, Sylvestre-Alvise Rebuffi, Sébastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Autonovel: Automatically discovering and learning novel visual categories. *TPAMI*, 44(10): 6767–6781, 2022.

John A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., USA, 99th edition, 1975. ISBN 047135645X.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, June 2016.

Zhenqi He, Yuanpei Liu, and Kai Han. Category discovery: An open-world perspective, 2025. URL https://arxiv.org/abs/2509.22542.

Xuhui Jia, Kai Han, Yukun Zhu, and Bradley Green. Joint representation learning and novel category discovery on single- and multi-modal data. In *ICCV*, pp. 590–599, 2021.

Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *ICCV workshops*, pp. 554–561, 2013.

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

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NeurIPS*, volume 25, 2012.

Michael Kuoch, Chi-Ning Chou, Nikhil Parthasarathy, Joel Dapello, James J. DiCarlo, Haim Sompolinsky, and SueYeon Chung. Probing biological and artificial neural networks with task-dependent neural manifolds. In *Conference on Parsimony and Learning*, volume 234, pp. 395–418, 2024.

Hyeongjun Kwon, Jinhyun Jang, Jin Kim, Kwonyoung Kim, and Kwanghoon Sohn. Improving visual recognition with hyperbolical visual hierarchy mapping. In *CVPR*, pp. 17364–17374, June 2024.

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

Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018.

Tong Lin and Hongbin Zha. Riemannian manifold learning. *TPAMI*, 30(5):796–809, 2008.

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*, pp. 1078–1087, October 2025a.

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

Yuanpei Liu, Zhenqi He, and Kai Han. Hyperbolic category discovery. In *CVPR*, pp. 9891–9900, June 2025b.

Shijie Ma, Fei Zhu, Zhun Zhong, Xu-Yao Zhang, and Cheng-Lin Liu. Active generalized category discovery. In *CVPR*, pp. 16890–16900, June 2024.

Yanbiao Ma, Licheng Jiao, Fang Liu, Shuyuan Yang, Xu Liu, and Lingling Li. Curvature-balanced feature manifold learning for long-tailed classification. In *CVPR*, pp. 15824–15835, June 2023.

Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *TPAMI*, 44(7):3523–3542, 2022.

Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, and CV Jawahar. Cats and dogs. In *CVPR*, pp. 3498–3505, 2012.

Nan Pu, Zhun Zhong, and Nicu Sebe. Dynamic conceptional contrastive learning for generalized category discovery. In *CVPR*, pp. 7579–7588, 2023.

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

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pp. 10684–10695, June 2022.

Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

Jianlin Su, Mingren Zhu, Ahmed Murtadha, Shengfeng Pan, Bo Wen, and Yunfeng Liu. Zlpr: A novel loss for multi-label classification, 2022. URL https://arxiv.org/abs/2208.02955.

Christian Szegedy, Alexander Toshev, and Dumitru Erhan. Deep neural networks for object detection. In *NeurIPS*, volume 26, 2013.

Kiat Chuan Tan, Yulong Liu, Barbara Ambrose, Melissa Tulig, and Serge Belongie. The herbarium challenge 2019 dataset, 2019.

Luyao Tang, Kunze Huang, Chaoqi Chen, and Cheng Chen. Bures-isotropy alignment: Manifold learning of generalized category discovery. In *ICLR*, 2026.

Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *CVPR*, pp. 8769–8778, 2018.

Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *CVPR*, pp. 7492–7501, June 2022.

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

Xin Wen, Bingchen Zhao, and Xiaojuan Qi. Parametric classification for generalized category discovery: A baseline study. In *ICCV*, pp. 16544–16554, 2023.

Zhenyue Zhang, Jing Wang, and Hongyuan Zha. Adaptive manifold learning. *TPAMI*, 34(2): 253–265, 2012.

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

Bingchen Zhao and Oisin Mac Aodha. Incremental generalized category discovery. In *ICCV*, pp. 19137–19147, October 2023.

Haiyang Zheng, Nan Pu, Wenjing Li, Nicu Sebe, and Zhun Zhong. Prototypical hash encoding for on-the-fly fine-grained category discovery. In *NeurIPS*, volume 37, pp. 101428–101455, 2024.