Generalized Biomedicine Discovery

Luyao Tang[1], Yingkai Yang[2], Hanqi Chen[2], Jiewei Zheng[3], Chaoqi Chen[2], and Cheng Chen[†1]

[1] The University of Hong Kong, Hong Kong SAR, China
[2] Shenzhen University, Shenzhen, China
[3] Xiamen University, Xiamen, China
lytang1999@gmail.com, cchen@eee.hku.hk

Abstract. In real-world clinical practice, medical images face open-world shifts: (i) long-tailed rare diseases, (ii) subtle lesions dominated by normal anatomy, and (iii) hierarchical taxonomies. Yet most open-world paradigms assume flat, balanced label spaces, leaving these biomedical demands unresolved. We introduce Generalized Biomedicine Discovery (**GBD**) and a unified benchmark spanning long-tail, anomaly, and taxonomy-aware discovery. Our key insight is that dominant known patterns form a visual manifold that masks subtle novelty. Inspired by expert diagnosis, we propose **SCAN** (Surprise-evoked Complementary AccommodatioN), which follows a cognition-inspired perceptual progression: it applies *predictive suppression* to filter expected norms, triggers *surprise-evoked salience* to highlight unexpected deviations, and performs *complementary accommodation* to integrate these shifts into global representations. Extensive experiments show that SCAN improves novel concept discovery while generally preserving established clinical knowledge, and it plugs into existing architectures to better navigate the known–unknown trade-off in medical imaging. Code is available at github.com/lytang63/generalized-biomedicine-discovery.

Keywords: Generalized biomedicine discovery · Generalized category discovery · Medical image analysis · Cognitive science

1 Introduction

Medical imaging systems are increasingly deployed in *open-world* clinical environments [65, 74], where the test-time concept space is neither closed nor balanced [66]. In clinical practice, medical images exhibit three characteristic open-world signatures: (i) **long-tailed** distributions [32, 67, 75, 77] where clinically critical rare diseases are sparsely observed, (ii) the **visual dominance** of normal anatomy [5, 16, 31, 58] that overwhelms subtle lesions, and (iii) **hierarchical** taxonomies [8, 15, 72] where novel concepts emerge as fine-grained siblings within established diagnostic families. Yet, most open-world paradigms and benchmarks still assume flat and relatively balanced concept spaces, leaving these authentic biomedical needs largely unresolved [14, 21].

[†] Corresponding author.

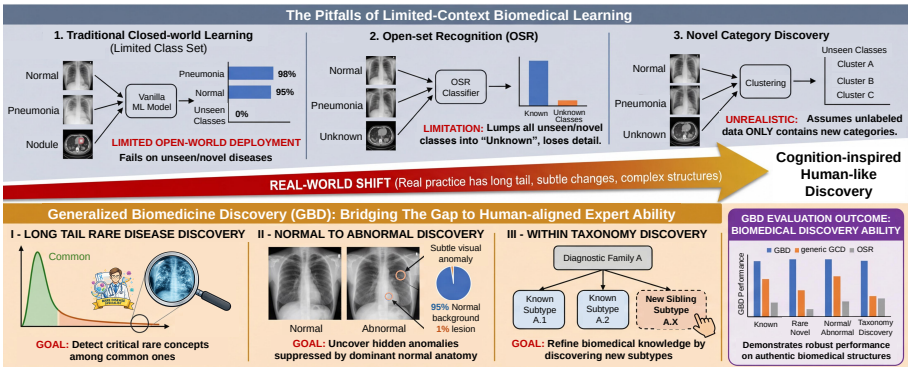


Fig. 1: GBD in practice. Real biomedical data shift beyond the closed world: standard classification fails on unseen diseases, OSR merges heterogeneous novelties into one "unknown", and conventional discovery often assumes unlabeled data contains only new categories. **GBD** formalizes three realistic discovery regimes: (I) long-tail rare disease discovery among common cases, (II) normal-to-abnormal discovery of subtle lesions dominated by normal anatomy, and (III) within-taxonomy discovery that uncovers new sibling subtypes to refine diagnostic hierarchies.

As illustrated in Fig. 1, conventional closed-set learning assumes a static class vocabulary, limiting its use in dynamic clinical environments. Open-Set Recognition (OSR) [22,49,50,60,71] identifies unseen classes but lumps them into a single "unknown" category, while Novel Category Discovery (NCD) [24, 48, 78] clusters unknowns under the unrealistic assumption that unlabeled data contains *only* novel classes. Generalized Category Discovery (GCD) [3,6,29,43,55,56,59,62,76] instead jointly classifies known and discovers novel classes in mixed unlabeled data. However, these paradigms do not capture biomedical data, where class imbalance, spatially dominant normal anatomy, and structured clinical taxonomies jointly obscure novel concepts [35,58,69,75,77]. Consequently, methods validated on natural-image GCD often fail to translate to clinical settings.

This mismatch motivates an expert-aligned target: clinicians discover new concepts *within* established knowledge by recognizing expected patterns, detecting subtle deviations, and situating new subtypes within diagnostic families [1, 23, 52]. Existing open-world paradigms [14, 21, 24, 59] rarely evaluate this capability, especially under long-tail rarity and hierarchical taxonomies [35, 69].

To bridge this gap, we introduce **Generalized Biomedicine Discovery (GBD)**, a realistic open-world paradigm and unified medical-imaging benchmark with three clinically grounded settings: (1) *Long-tail Rare Disease Discovery*, (2) *Normal-to-Abnormal Discovery*, and (3) *Within-Taxonomy Discovery*. Beyond these structural challenges, GBD exposes a critical visual problem: dominant known features (*e.g.*, widespread normal anatomy) suppress subtle, localized novel signals during global representation learning. This bias attenuates the residual evidence needed to discover new diseases, motivating a mechanism that counters visual suppression without prior knowledge of the anomalies.

Inspired by diagnostic cognition [2, 13, 23], we propose **SCAN** (Surprise-evoked Complementary AccommodatioN), a plug-and-play layer operating on a vision backbone's class and patch tokens [7]. SCAN instantiates this progression in three operations: (i) **predictive suppression** filters anticipated patterns under predictive coding [45]; (ii) **surprise-evoked salience** uses residual energy to amplify unexpected, localized deviations [33]; and (iii) **complementary accommodation** integrates the isolated novelty into the global representation while preserving established structure [39].

- We introduce **GBD**, a clinically grounded open-world paradigm and unified benchmark for biomedicine, formulating three grounded settings: Long-tail Rare Disease, Normal-to-Abnormal, and Within-Taxonomy Discovery.
- We propose **SCAN**, a plug-and-play cognitive vision layer that mitigates dominant-feature suppression through predictive filtering, surprise amplification, and complementary feature accommodation.
- Extensive experiments demonstrate that SCAN generally improves conventional GCD methods on the GBD benchmark, with particularly strong gains for SelEx in isolating subtle anomalies and discovering long-tailed rare diseases, while some baseline- and subset-specific trade-offs remain.

2 Related Work

Open-world Learning. Open-world learning aims to identify and discover novel categories beyond closed-set assumptions. Open-set recognition [22, 26–28, 49, 60, 71] detects unseen classes but groups them into a single "unknown" label, lacking fine-grained discrimination. Novel category discovery [24, 48, 78] clusters unknowns yet unrealistically assumes unlabeled data contains only novel classes. Generalized category discovery [59] advances this by jointly classifying knowns and discovering novelties in mixed unlabeled data, with methods split into parametric [6, 61, 64] and non-parametric frameworks [11, 12, 46, 76]. GCD commonly refines representations through contrastive learning [10, 25, 37, 38], but often assumes separable pre-trained features. Recent variants extend GCD to continual open-world shifts [17, 18, 20], while medical approaches adapt discovery to clinical images [19, 21]. Yet these efforts largely retain flat category spaces, leaving biomedical class imbalance, dominant normal anatomy, and clinical hierarchies underexplored. We instead target dominant-feature suppression with a cognition-aligned framework tailored to biomedical discovery.

Cognitive Sciences in Machine Learning. Cognitive sciences principles have increasingly informed machine learning, leveraging human perceptual and reasoning mechanisms for more robust generalization [4, 9, 57]. Prior work draws on structured cognitive processes (*e.g.*, perceptual grouping, inductive reasoning) [68] to design structured priors [47, 54, 63] and factorized representations [36, 53], enhancing model robustness across tasks [41, 73]. Key cognitive theories: predictive coding [45] (suppressing anticipated visual features), surprise-driven

salience [33] (amplifying anomalous signals), and complementary learning systems [39] have been explored in isolation for visual processing, yet remain underexploited for biomedical diagnostic cognition [2, 13, 23]. We integrate these theories into a unified framework that mirrors clinical diagnostic reasoning, filling this critical gap for biomedical open-world learning.

3 Problem Setting

3.1 Task Backbone: Generalized Biomedicine Discovery (GBD)

For a given dataset, let $\mathcal{X}$ and $\mathcal{Y}$ denote the image and label spaces. We are provided with a labeled subset $\mathcal{D}_l = \{(\mathbf{x}_i^l, y_i^l)\} \subset \mathcal{X} \times \mathcal{Y}_l$ and an unlabeled subset $\mathcal{D}_u = \{(\mathbf{x}_i^u, y_i^u)\} \subset \mathcal{X} \times \mathcal{Y}_u$. y_i^u is used only for evaluation. Only *known* classes are labeled in $\mathcal{D}_l$, while $\mathcal{D}_u$ contains a mixture of known and *novel* classes: $\mathcal{Y}_l = \mathcal{C}_{\text{known}}$ and $\mathcal{Y}_u = \mathcal{C}_{\text{known}} \cup \mathcal{C}_{\text{novel}}$. The objective of GBD is to assign each unlabeled sample to either a known class or a newly discovered novel class.

3.2 Setting I: Long-tail Rare Disease Discovery

This setting instantiates discovery under the *long-tail statistics* prevalent in biomedicine. Let $\mathcal{C}$ denote the set of leaf-level diagnostic categories, and $n(c)$ be the empirical frequency of class $c \in \mathcal{C}$ in the training pool. We define a head–tail split controlled by a parameter $\alpha \in (0, 1)$:

$$\mathcal{C}_{\text{known}} = \text{Head}_\alpha(\mathcal{C}) \;=\; \arg \max_{\substack{\mathcal{S} \subseteq \mathcal{C} \\ |\mathcal{S}| = \lceil \alpha |\mathcal{C}| \rceil}} \sum_{c \in \mathcal{S}} n(c), \qquad \mathcal{C}_{\text{novel}} = \mathcal{C} \setminus \mathcal{C}_{\text{known}}, \tag{1}$$

where ties in empirical frequencies are broken lexicographically. Thus, $\mathcal{D}_l$ contains labels only from $\mathcal{C}_{\text{known}}$, while $\mathcal{D}_u$ mixes samples from $\mathcal{C}_{\text{known}} \cup \mathcal{C}_{\text{novel}}$. This design mirrors the epidemiological reality where annotations focus on frequent conditions, forcing the model to discover *rare but clinically critical* tail diseases. We generally use $\alpha = 0.5$, except for classes explicitly identified as rare in medical references.

3.3 Setting II: Normal-to-Abnormal Discovery

This setting aligns with the workflow of detecting pathology relative to *diverse normal baselines*. We partition the label space into normal and abnormal subsets:

$$\mathcal{C}_{\text{known}} = \mathcal{C}_{\text{normal}}, \qquad \mathcal{C}_{\text{novel}} = \mathcal{C}_{\text{abnormal}}, \qquad \mathcal{C}_{\text{normal}} \cap \mathcal{C}_{\text{abnormal}} = \emptyset. \tag{2}$$

The set $\mathcal{C}_{\text{normal}}$ contains multiple normal subclasses (e.g., distinct anatomical landmarks), whereas $\mathcal{C}_{\text{abnormal}}$ contains multiple pathological categories. By labeling $\mathcal{D}_l$ on $\mathcal{C}_{\text{normal}}$, this setting evaluates a model's capacity to generalize beyond a normal manifold to discover *heterogeneous abnormal patterns*.

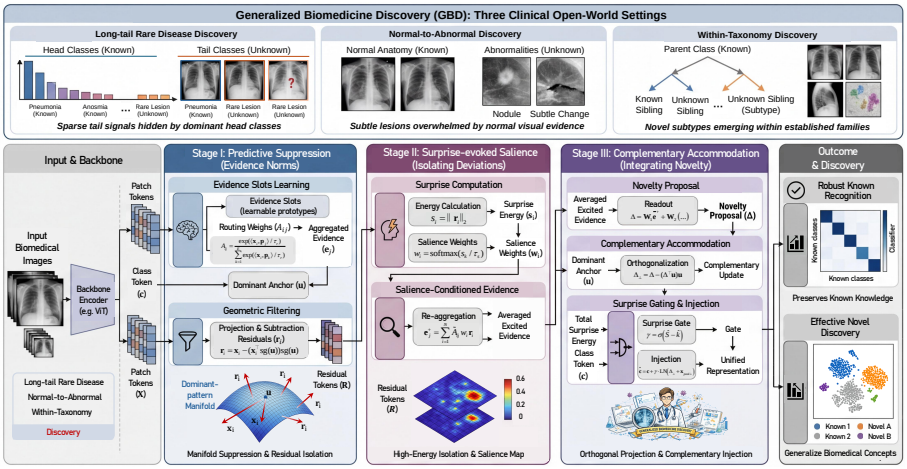


Fig. 2: Method overview of SCAN (Surprise-evoked Complementary AccommodatioN) for GBD. SCAN refines representations in three steps: **(I)** it learns a notion of *dominant evidence* and suppresses predictable visual patterns to expose residual signals; **(II)** it scores the residuals by "surprise" to highlight sparse, high-energy deviations and aggregates them into salient evidence; **(III)** it converts the salient evidence into a complementary update and injects it through a gated fusion, limiting disruption to known concepts while strengthening novel structure.

3.4 Setting III: Within-Taxonomy Discovery

This setting captures previously unseen subtypes *within* established hierarchical taxonomies. Focusing on preterminal nodes, let $\mathcal{P}$ be the set of parent nodes whose children are exclusively leaf classes $\mathcal{L}$. We treat each sibling set as a semantic group: $\mathcal{S}(p) = \text{children}(p)$ for $p \in \mathcal{P}$. We consider sibling groups with at least two leaf classes. For each $p \in \mathcal{P}$, let $n_p = |\mathcal{S}(p)|$ and $k_p = \min\{n_p - 1, \max\{1, \lceil \rho n_p \rceil\}\}$. We choose $\mathcal{K}(p) \subseteq \mathcal{S}(p)$ with $|\mathcal{K}(p)| = k_p$, and define $\mathcal{N}(p) = \mathcal{S}(p) \setminus \mathcal{K}(p)$. This ensures at least one known and one novel class per group. The global partitions are then defined as:

$$\mathcal{C}_{\text{known}} = \bigcup_{p \in \mathcal{P}} \mathcal{K}(p), \qquad \mathcal{C}_{\text{novel}} = \bigcup_{p \in \mathcal{P}} \mathcal{N}(p). \tag{3}$$

Smaller ρ yields fewer labeled siblings per group and more novel siblings $\mathcal{N}(p)$ to discover. This mimics clinical taxonomies, testing the model's ability to isolate fine-grained novelty that is *semantically adjacent* to known diagnostic families.

4 Method

Although the three GBD regimes differ clinically, they share a failure mode: recurring known structure dominates the global embedding, underrepresenting

weak cues for rare classes, abnormalities, or new sibling subtypes. This motivates two requirements: expose deviations relative to each image's dominant content, then integrate them without indiscriminately rewriting established structure. We follow the analogous clinical progression of forming a predictive prior, attending to deviations, and accommodating informative evidence. We instantiate this progression as **SCAN** in Fig. 2, a lightweight plug-in that updates class and patch tokens through three stages: Predictive Suppression [45], Surprise-Evoked Salience [33], and Complementary Accommodation [39]. Given an input image $\mathbf{x} \in \mathcal{X}$, an encoder outputs a class token and N patch tokens:

$$\mathbf{c},\, \mathbf{X} = F_\theta(\mathbf{x}), \qquad \mathbf{c} \in \mathbb{R}^D,\ \mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^\top \in \mathbb{R}^{N \times D}. \tag{4}$$

SCAN performs a cognition-inspired transformation $(\mathbf{c}, \mathbf{X}) \mapsto \tilde{\mathbf{c}}$ in feature space, leaving the host discovery objective unchanged.

4.1 Stage I: Predictive Suppression

The brain behaves like a "predictor": for *expected* sensory content (e.g., dominant normal anatomy), feedback connections actively suppress predictable responses, freeing capacity for unexpected signals [45]. In medical images, frequently occurring structures can form a strong manifold that suppresses subtle lesions and rare concepts; thus SCAN first builds a sample-conditioned *dominant anchor* and geometrically filters out predictable components.

Dominant evidence and dominant anchor. We softly summarize recurring visual patterns by routing patch tokens into E evidence slots, which serve as a compact, image-adaptive dominant-pattern codebook. Let $\mathbf{P} \in \mathbb{R}^{E \times D}$ be learnable evidence prototypes and τ_r a temperature. We compute routing weights:

$$A_{ij} = \frac{\exp(\langle \mathbf{x}_i, \mathbf{p}_j \rangle / \tau_r)}{\sum_{k=1}^{E} \exp(\langle \mathbf{x}_i, \mathbf{p}_k \rangle / \tau_r)}. \tag{5}$$

We renormalize over tokens so that each slot forms a weighted visual summary independent of its assigned token mass:

$$\tilde{A}_{ij} = \frac{A_{ij}}{\sum_{k=1}^{N} A_{kj}}, \qquad \mathbf{e}_j = \sum_{i=1}^{N} \tilde{A}_{ij}\, \mathbf{x}_i, \qquad \bar{\mathbf{e}} = \frac{1}{E} \sum_{j=1}^{E} \mathbf{e}_j, \tag{6}$$

where $\langle \cdot, \cdot \rangle$ denotes dot product (cosine if tokens are normalized). We then form a sample-specific *dominant anchor* direction:

$$\mathbf{u} = \frac{\mathbf{c} + \bar{\mathbf{e}}}{\|\mathbf{c} + \bar{\mathbf{e}}\|_2}. \tag{7}$$

Fusing global and mean-slot evidence yields an image-adaptive anchor instead of a fixed normal prototype.

Predictive suppression by geometric filtering. We actively remove the predictable component of each patch token along the dominant anchor, producing residuals in the orthogonal complement:

$$\mathbf{r}_i = \mathbf{x}_i - (\mathbf{x}_i^\top \operatorname{sg}(\mathbf{u}))\operatorname{sg}(\mathbf{u}), \qquad \mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_N]^\top. \tag{8}$$

Here $\operatorname{sg}(\cdot)$ denotes stop-gradient and stabilizes the anchor. The projection removes only its aligned component and preserves the remaining token geometry, requiring neither normal labels nor anomaly templates.

4.2 Stage II: Surprise-Evoked Salience

After suppression, the residual norm measures image-conditional mismatch: a token is surprising when the dominant anchor explains it poorly. This mirrors the attentional role of Bayesian surprise [33] without requiring a dataset-specific anomaly threshold.

Surprise energy and salience weights. We quantify patch-wise surprise by the residual energy and convert it into a salience distribution:

$$s_i = \|\mathbf{r}_i\|_2, \qquad w_i = \frac{\exp(s_i/\tau_s)}{\sum_{k=1}^{N} \exp(s_k/\tau_s)}, \tag{9}$$

where τ_s controls concentration. The within-image softmax ranks residual evidence without assuming cross-dataset calibration.

Salience-conditioned evidence of novelty. We re-aggregate residuals using both their original slot assignments and their surprise weights, yielding *excited* evidence that concentrates on coherent, unpredicted deviations:

$$\mathbf{e}_j^+ = \sum_{i=1}^{N} \tilde{A}_{ij}\, w_i\, \mathbf{r}_i, \qquad \bar{\mathbf{e}}^+ = \frac{1}{E}\sum_{j=1}^{E} \mathbf{e}_j^+. \tag{10}$$

Here, $\tilde{A}_{ij}$ preserves slot semantics, whereas w_i controls residual contribution. Eqs. (9)–(10) thus separate *what recurring pattern a token belongs to* from *how unexpectedly it departs*.

Token refinement (broadcast of excited evidence). Beyond global evidence, clinicians also refine a local "mental map" of the image. We broadcast excited evidence back to tokens with a lightweight readout:

$$\mathbf{X}^{out} = \mathbf{X} + \phi(\mathbf{A}\mathbf{E}^+), \qquad \mathbf{E}^+ = [\mathbf{e}_1^+, \ldots, \mathbf{e}_E^+]^\top \in \mathbb{R}^{E\times D}, \tag{11}$$

where $\phi(\cdot)$ is a small MLP applied row-wise.

4.3 Stage III: Complementary Accommodation

Discovery must integrate deviations into the global representation, yet an unconstrained update may distort known structure. Motivated by Complementary Learning Systems theory [39], SCAN constructs an additive update with reduced overlap with the dominant direction.

Novelty vector from excited evidence. We map excited evidence into a global novelty proposal via a simple first- and second-order readout:

$$\Delta = \mathbf{W}_1 \, \bar{\mathbf{e}}^{+} + \mathbf{W}_2 \left(\frac{1}{E} \sum_{j=1}^{E} \mathbf{e}_j^{+} \odot \mathbf{e}_j^{+} \right), \tag{12}$$

where $\odot$ is the Hadamard product and $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{D \times D}$ are learnable. The first-order term retains the signed mean shift; the second records dimension-wise energy that may cancel in the mean. Together, they capture consistent and sparse-but-strong deviations without another attention block.

Complementarity. We reduce direct interference with the dominant content by removing the component of the proposal aligned with the anchor:

$$\Delta_{\perp} = \Delta - (\Delta^{\top} \mathbf{u})\mathbf{u}. \tag{13}$$

We further summarize refined tokens by $\mathbf{x}_{pool} = \mathrm{MeanPool}(\mathbf{X}^{out}) \in \mathbb{R}^D$ and project this token evidence into the same complementary subspace:

$$\mathbf{x}_{pool\perp} = \mathbf{x}_{pool} - (\mathbf{x}_{pool}^{\top} \mathbf{u})\mathbf{u}. \tag{14}$$

Projecting both branches into the same subspace prevents the pooled token summary from reintroducing the dominant component removed from the global proposal.

Non-parametric surprise gate. Rather than learning dataset-specific scalar calibrations, we define γ from three cognitively motivated signals: (i) *surprise magnitude* $S = \sum_i w_i s_i$, (ii) *knownness* $k = \cos(\mathbf{c}, \bar{\mathbf{e}})$ (how well the sample aligns with dominant evidence), and (iii) *salience concentration* via peakedness κ:

$$S = \sum_{i=1}^{N} w_i s_i, \qquad k = \cos(\mathbf{c}, \bar{\mathbf{e}}), \qquad \kappa = 1 - \frac{\mathrm{Ent}(\mathbf{w})}{\log N}, \;\; \mathrm{Ent}(\mathbf{w}) = -\sum_{i=1}^{N} w_i \log w_i. \tag{15}$$

Using EMA moments (μ_S, σ_S) and (μ_k, σ_k), we standardize the signals as $\hat{S} = (S - \mu_S)/\sigma_S$ and $\hat{k} = (k - \mu_k)/\sigma_k$. The robust gate is $\gamma = \sigma(\hat{S} - \hat{k}) \cdot \mathrm{clip}(\kappa, 0, 1)$. The three factors are complementary: S measures deviation strength, k discounts samples already explained by dominant evidence, and κ suppresses diffuse residual activation. Accommodation is therefore selective in both residual magnitude and salience concentration.

Stable accommodation. We assimilate novelty as a complementary residual:

$$\tilde{\mathbf{c}} = \mathbf{c} + \gamma \cdot \mathrm{LN}(\Delta_{\perp} + \mathbf{x}_{pool\perp}) \,, \tag{16}$$

The identity path recovers the original representation when γ is small, and $\mathrm{LN}(\cdot)$ controls update scale. This cannot guarantee complete independence from known features, but it biases accommodation away from the dominant direction and limits feature drift. The updated $\tilde{\mathbf{c}}$ enters the unchanged GCD pipeline.

Connection to the GBD. The anchor has a setting-specific interpretation but a shared role. In Setting I, it summarizes head-class patterns, exposing underrepresented tail cues. In Setting II, it reflects recurring normal anatomy, retaining localized pathology in the residual. In Setting III, it captures parent-level morphology, while accommodation emphasizes subtype differences. SCAN thereby targets the common suppression mechanism without setting-specific supervision.

5 Experiments

Through comprehensive experiments, we seek to answer three core questions: (1) Benchmark validity. Does GBD capture the key open-world challenges in medical imaging, and can it differentiate methods across long-tail, anomaly-dominated, and taxonomy-aware settings? (2) Effectiveness and generality. Does SCAN improve generalized biomedicine discovery, and can it serve as a lightweight plug-and-play module that strengthens diverse host baselines without altering their original training protocols? (3) Mechanism and practicality. Which components drive the gains, how do they interact in the full framework, and how robust is SCAN under key hyperparameter choices and compute budgets?

5.1 Experimental Setup

Datasets. We evaluate GBD on four publicly available datasets spanning diverse biomedical imaging domains. Each dataset provides an explicit class label per image (either by folder structure or an official label file), and several of them additionally provide a native taxonomy used for our GBD's settings. **Gastrovision** [34] (endoscopy): multi-center GI endoscopy dataset spanning upper and lower GI tracts; labels are organized by clinically meaningful finding types, including anatomical landmarks/normal findings, pathological abnormalities, and procedure-related categories. **HistoSet-5×14** [44] (histopathology): multi-organ H&E collection covering five organs (breast, colon, lung, oral cavity, ovary); the label space is naturally structured by organ identity and tissue state across 14 tissue classes. **MLL23** [51] (microscopy): expert-annotated peripheral blood single-cell images; labels follow hematopoietic structure with major groups and finer subdivisions into mature and immature cell types/stages. **DERM12345** [70] (dermatology): dermatoscopic skin lesion dataset with a taxonomic tree comprising 5 super-classes, 15 main classes, and 40 subclasses.

Table 1: Comparison on **Setting I** (Long-tail Rare Disease Discovery) across endoscopy , pathology , microscopy , and dermatology images. Results are accuracy on All, Old, and New subsets, with the best results highlighted in bold.

Method	Gastrovision			HistoSet			MLL23			Derm12345			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
SEAL†	56.8	51.2	**60.7**	**81.2**	87.1	77.9	65.4	67.6	54.1	**56.5**	**61.0**	10.6	65.0	66.7	50.8
SelEx	50.1	52.0	47.4	76.4	97.5	70.5	66.4	68.3	61.3	28.9	29.1	27.9	55.5	61.7	51.8
+ SCAN	**71.2**	**83.4**	53.0	76.3	**99.9**	69.7	**80.7**	**89.3**	58.6	35.6	35.6	**35.6**	**66.0**	**77.1**	**54.2**
Δ	**21.1**	**31.4**	**5.6**	-0.1	**2.4**	-0.8	**14.3**	**20.9**	-2.7	**6.7**	**6.5**	**7.7**	**10.5**	**15.3**	**2.4**
SimGCD	37.8	40.6	31.8	74.4	83.8	71.8	56.9	52.9	67.0	25.4	27.9	13.1	48.6	51.3	45.9
+ SCAN	39.7	43.3	31.9	80.0	82.1	**79.4**	64.3	63.1	67.2	25.9	28.2	14.2	52.5	54.2	48.1
Δ	**1.9**	**2.7**	**0.1**	**5.6**	-1.7	**7.6**	**7.4**	**10.2**	**0.2**	**0.5**	**0.4**	**1.1**	**3.8**	**2.9**	**2.2**
LegoGCD	39.8	48.9	26.1	72.5	76.0	71.6	70.8	63.4	**89.7**	26.3	28.5	15.7	52.3	54.2	50.8
+ SCAN	41.3	49.4	29.2	72.5	77.5	71.2	73.2	67.2	88.8	26.7	29.0	14.8	53.4	55.8	51.0
Δ	**1.5**	**0.5**	**3.1**	**0.0**	**1.5**	-0.4	**2.4**	**3.8**	-0.9	**0.3**	**0.6**	-0.9	**1.1**	**1.6**	**0.2**
Avg. Δ	**8.2**	**11.5**	**2.9**	**1.8**	**0.7**	**2.1**	**8.1**	**11.6**	-1.2	**2.5**	**2.5**	**2.7**	**5.1**	**6.6**	**1.6**

Evaluation protocol. For each dataset, we train models on $\mathcal{D}_l \cup \mathcal{D}_u$ *without access to the ground-truth labels* y_i^u *in* $\mathcal{D}_u$. At test time, we measure clustering accuracy between the y_i^u and the model predictions $\hat{y}_i$ over $\mathcal{D}_u$ as:

$$\text{ACC} = \max_{p \in \mathcal{P}(\mathcal{Y}_u)} \frac{1}{|\mathcal{D}_u|} \sum_{i=1}^{|\mathcal{D}_u|} \mathbb{1}\{y_i^u = p(\hat{y}_i)\}\,. \tag{17}$$

where $\mathcal{P}(\mathcal{Y}_u)$ is the set of all permutations of class labels in the unlabeled label space $\mathcal{Y}_u$. The maximization is computed via the Hungarian optimal assignment algorithm. Following standard practice in generalized discovery, we report ACC on all unlabeled instances, and additionally on the known subset (instances in $\mathcal{D}_u$ with labels in $\mathcal{C}_{\text{known}}$) and the novel subset (instances in $\mathcal{D}_u$ with labels in $\mathcal{C}_{\text{novel}}$). Importantly, we compute the Hungarian assignment *once* over all classes in $\mathcal{Y}_u$, and only then evaluate the known/novel subset accuracies under the same assignment. Following the standard generalized discovery protocol [59], we assume the number of novel classes K_{novel} is available for evaluation.

Implementation details. We implement SCAN as plug-and-play that can be seamlessly attached to diverse GCD baselines. All methods use a ViT-B/16 backbone initialized with self-supervised DINO-V2 weights [42]. Models are trained for 200 epochs. In our default configuration, SCAN uses $E = 16$ evidence slots. To evaluate the generality of our approach, we largely follow the ***original training hyperparameters and optimization schedules*** of each host baseline, without method-specific tuning. Following the original GCD configuration [59], we allocate 50% of the samples from the known set to the unlabeled subset. We generally use $\alpha = 0.5$ in GBD's Setting I and use $\rho = 0.5$ in Setting III.

Baselines. We test SCAN with representative GCD baselines for forgetting mitigation and hierarchy-aware learning.

Table 2: Comparison on **Setting II** (Normal-to-Abnormal Discovery) across endoscopy , pathology , microscopy , and dermatology images. Results are accuracy on All, Old, and New subsets, with the best results highlighted in bold.

Method	Gastrovision			HistoSet			MLL23			Derm12345			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
SEAL†	62.6	64.2	**61.0**	**73.5**	**98.7**	48.2	69.2	70.9	60.1	**59.0**	**61.7**	5.5	**66.1**	73.8	43.7
SelEx	50.3	51.9	41.7	64.3	97.7	47.5	58.5	58.0	59.8	29.3	28.2	**40.7**	50.6	58.9	47.4
+ SCAN	**70.1**	**86.0**	46.3	68.6	97.2	**60.7**	**87.1**	**89.1**	**81.7**	33.6	33.3	36.4	64.8	**76.4**	**56.3**
Δ	**19.8**	**34.1**	**4.6**	**4.4**	-0.5	**13.1**	**28.6**	**31.1**	**21.9**	**4.3**	**5.1**	-4.3	**14.2**	**17.5**	**8.8**
SimGCD	39.7	42.9	25.6	61.1	94.1	44.5	57.9	55.9	63.4	26.6	27.2	19.9	46.3	55.0	38.3
+ SCAN	40.1	44.0	22.6	62.0	93.6	46.1	58.3	57.5	60.6	26.4	26.8	21.6	46.7	55.5	37.7
Δ	**0.4**	**1.1**	-3.0	**0.9**	-0.6	**1.6**	**0.5**	**1.7**	-2.8	-0.2	-0.4	**1.7**	**0.4**	**0.5**	-0.6
LegoGCD	49.3	53.7	24.9	62.0	97.5	44.2	68.5	72.1	58.6	28.5	30.8	5.5	52.1	63.5	33.3
+ SCAN	50.5	55.5	23.1	63.9	97.4	47.1	69.4	73.7	57.7	29.0	29.7	21.6	53.2	64.1	37.4
Δ	**1.3**	**1.8**	-1.8	**1.9**	-0.1	**2.9**	**0.9**	**1.6**	-0.9	**0.4**	-1.1	**16.1**	**1.1**	**0.6**	**4.1**
Avg. *Δ*	**7.1**	**12.3**	-0.1	**2.4**	-0.4	**5.9**	**10.0**	**11.5**	**6.1**	**1.5**	**1.2**	**4.5**	**5.2**	**6.2**	**4.0**

– **SimGCD** [64]: a simple one-stage parametric classifier for GCD, using debiased learning and entropy regularization to reduce the tendency.
– **LegoGCD** [6]: a modular add-on that targets catastrophic forgetting in GCD by regularizing predictions to better preserve known-class knowledge while improving novel discrimination.
– **SelEx** [46]: a fine-grained GCD method that builds "self-expertise" via hierarchical pseudo-labeling and complementary supervised/unsupervised training signals to separate subtle categories.
– **SEAL** [30]: a semantic-aware hierarchical framework that leverages naturally available hierarchies. Because SEAL requires additional hierarchical labels, we treat it as an auxiliary performance reference rather than a directly comparable baseline.

5.2 Main Results

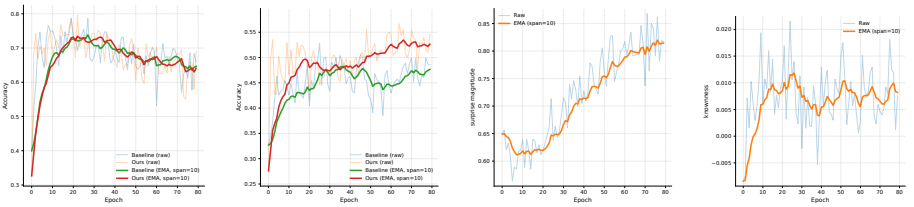
Results on Long-tail Rare Disease Discovery. Table 1 evaluates SCAN on the clinically critical long-tail regime across four imaging domains, where class-frequency imbalance leaves tail categories sparsely represented relative to dominant head classes. SCAN improves the average *All* accuracy of all three host baselines, yielding a notable +**10.5%** gain on SelEx, although several subset-specific regressions remain. The effect is most pronounced on Gastrovision, where *Old* accuracy rises from 52.0% to **83.4%** (+**31.4%**), while on Derm12345 it boosts tail recognition from 27.9% to 35.6% (+**7.7%**). Although SEAL is not directly comparable because it uses additional hierarchical labels, SelEx+SCAN exceeds its reported accuracy in several settings, suggesting that SCAN can recover part of the benefit of hierarchy-aware supervision without explicitly using such labels. These results support our motivation: *predictive suppression* attenuates dominant head-class patterns and *surprise-evoked salience* amplifies underrepresented tail cues.

Table 3: Comparison on **Setting III** (Within-Taxonomy Discovery) across endoscopy , pathology , microscopy , and dermatology images. Results are accuracy on All, Old, and New subsets, with the best results highlighted in bold.

Method	Gastrovision			HistoSet			MLL23			Derm12345			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
SEAL†	54.9	51.3	**64.3**	**92.9**	95.7	**89.3**	70.6	83.3	47.2	27.7	62.4	16.4	61.5	73.2	**54.3**
SelEx	50.1	45.7	53.8	77.5	**99.4**	71.4	62.5	69.9	55.6	24.8	65.1	18.2	53.7	70.0	49.8
+ SCAN	**66.2**	**79.6**	46.3	88.5	96.4	83.2	**81.4**	**89.0**	**61.9**	**27.8**	**69.9**	**20.9**	**66.0**	**83.7**	53.1
Δ	**16.1**	**33.8**	-7.5	**11.0**	-3.0	**11.7**	**18.9**	**19.0**	**6.3**	**3.0**	**4.8**	**2.7**	**12.3**	**13.7**	**3.3**
SimGCD	42.9	43.0	42.5	89.4	91.7	87.9	53.2	57.4	49.4	19.3	40.8	15.9	51.2	58.2	48.9
+ SCAN	42.0	42.4	39.4	89.1	90.5	88.2	54.0	58.4	49.9	17.8	36.8	14.7	50.7	57.0	48.1
Δ	-0.9	-0.6	-3.1	-0.3	-1.2	**0.2**	**0.8**	**1.1**	**0.5**	-1.5	-4.0	-1.2	-0.5	-1.2	-0.8
LegoGCD	40.2	39.6	40.6	75.8	84.4	70.0	54.7	61.8	48.2	19.4	41.0	16.0	47.5	56.7	43.7
+ SCAN	40.5	40.4	40.6	80.5	86.2	76.6	60.3	68.7	52.6	20.8	42.4	17.2	50.5	59.4	46.8
Δ	**0.4**	**0.8**	**0.0**	**4.7**	**1.8**	**6.6**	**5.6**	**6.9**	**4.4**	**1.3**	**1.5**	**1.3**	**3.0**	**2.8**	**3.1**
Avg. Δ	**5.2**	**11.3**	-3.6	**5.1**	-0.8	**6.2**	**8.4**	**9.0**	**3.7**	**0.9**	**0.8**	**0.9**	**4.9**	**5.1**	**1.8**

Results on Normal-to-Abnormal Discovery. Table 2 shows that SCAN generally strengthens normal-to-abnormal discovery across the four imaging domains, a clinically crucial setting where normal anatomy can suppress faint pathological cues. With the strong SelEx baseline, SCAN improves the average accuracy by +14.2% on *All* and +8.8% on abnormal *New*, and the impact is most pronounced on Gastrovision, where *Old* increases by **34.1%**, and on MLL23, where *All* and *New* rise by **28.6%** and **21.9%**, respectively. SelEx+SCAN also exceeds SEAL on several reported metrics without extra fine-grained supervision, although the comparison is auxiliary because SEAL uses hierarchical labels. The observed regressions are concentrated in particular host-baseline and subset combinations and may partly reflect keeping the evidence-slot hyperparameter E fixed without per-domain tuning. Overall, the results are consistent with normal-anatomy suppression making subtle abnormalities more salient.

Results on Within-Taxonomy Discovery. Table 3 verifies SCAN in within-taxonomy discovery, where new sibling subtypes differ by subtle, localized cues under a shared parent appearance. With SelEx, SCAN raises the average *All* accuracy from 53.7% to 66.0% (+12.3%), peaking on MLL23 (*All* 62.5% to 81.4%, **+18.9%**) and Gastrovision (*Old* **+33.8%**). SelEx+SCAN also exceeds SEAL on MLL23 *New* accuracy (61.9% vs. 47.2%) without extra fine-grained supervision, although SEAL remains an auxiliary reference because it uses hierarchical labels. The observed drops are concentrated in particular datasets and host baselines and may partly reflect fixing E across domains to stress robustness; overall, the results support our motivation that predictive suppression and surprise salience expose fine-grained subtype evidence.



(a) Known Acc. **(b)** Novel Acc. **(c)** S in Eq. (15) **(d)** k in Eq. (15)

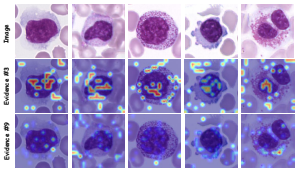
Fig. 3: Training dynamics on Gastrovision (Setting I): Known/Novel accuracy and the evolution of the surprise signal (S) and knownness (k) for SelEx and SelEx+SCAN.

5.3 Mechanism and Qualitative Analysis

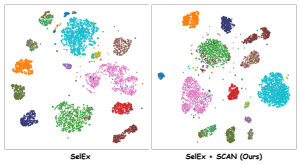
Adaptive Discovery Dynamics. We analyze Gastrovision (Setting I) training dynamics to understand SCAN's balance of discovery. As shown in Fig. 3, SCAN keeps *old* accuracy comparable to SelEx, while *new* accuracy rises steadily and substantially exceeds the baseline. This improvement correlates with rising surprise signals over epochs, a pattern consistent with increased sensitivity to sparse, high-energy deviations. Around 40 epochs, SelEx shows a clear performance drop, whereas SCAN enters a second growth phase with fast accuracy gains. Meanwhile, the knownness indicator k in Eq. (15) slightly drops, corresponding to a less restrictive surprise gate for new samples. These trends suggest that SCAN preserves stable knowledge while reallocating representation capacity as training evolves.

Cross-Sample Consistency of Evidence Slots. To examine the semantic consistency of learned evidence slots in Sec. 4.1, we visualize spatial activations on peripheral blood smear images. As shown in Fig. 4a, the slots appear to capture recurring cellular patterns. Specifically, Evidence slot #3 often highlights high-contrast punctate textures around the perinuclear cytoplasm and salient cell boundaries, whereas Evidence slot #9 tends to cover diffuse, low-intensity staining patterns in broader nuclear or cytoplasmic regions. These qualitative patterns suggest that the learned slots may provide reusable morphological primitives across samples.

Qualitative Analysis. Figure 4b visualizes embeddings on MLL23 (Setting I) with UMAP [40]. SelEx yields dispersed clusters and noticeable overlap between new (stars) and old (dots) regions. With SCAN, the visualization shows tighter within-class neighborhoods and clearer margins among classes. New cells appear more concentrated into coherent groups and more separated from old clusters, suggesting improved organization of subtle tail concepts in this qualitative example.



(a) Evidence activations.



(b) UMAP embeddings.

Models	Params (M)	Training Time (s)	Inference Time (s)
SimGCD	92.10	27.87	10.61
+ SCAN	99.68	30.12	11.32
Δ	+ 7.6	+ 8%	+ 6%

(c) Computational overhead.

Fig. 4: Mechanism, qualitative, and efficiency analyses of SCAN.

Computational Overhead. As shown in Fig. 4c, SCAN adds a small number of parameters and several linear layers on top of the backbone and head. The measured increases in training and inference time are limited, suggesting that SCAN can be integrated into existing discovery pipelines with modest computational overhead.

5.4 Discussion

GBD is not intended as a leaderboard against text-supervised multimodal models (*e.g.*, CLIP or LLMs), which assume a stable language anchor and reliable textual supervision. Instead, it targets upstream clinical *concept formation*: organizing unlabeled archives into coherent cohorts that experts can verify, name, merge, or split before integration into guidelines and curated datasets. This matters because emerging diseases and subtypes often lack consistent nomenclature across sites and time, while clinicians require cohort-level discovery, representative prototypes, and similar-case lists rather than one-off predictions.

6 Conclusion

We study generalized discovery in biomedical imaging under realistic clinical shifts. To support rigorous and comparable evaluation, we establish **Generalized Biomedicine Discovery**, a unified benchmark covering (i) long-tail rare disease discovery, (ii) normal-to-abnormal discovery, and (iii) within-taxonomy discovery across endoscopy, pathology, microscopy, and dermatology. Motivated by our central insight that dominant known patterns, especially normal anatomy, form a visual manifold that suppresses weak novelty cues, we propose **SCAN**, a cognition-inspired three-stage process of predictive suppression, surprise-evoked salience, and complementary accommodation. Extensive analyses and experiments support this insight, showing overall improvements in novel-disease discovery while revealing some dataset- and baseline-specific trade-offs in preserving established knowledge. GBD provides a clinically grounded, taxonomy-aware testbed that reframes biomedical imaging as a measurable concept-formation problem, enabling discovery systems to surface rare and emerging conditions and updateable subtypes with minimal reliance on exhaustive relabeling.

Acknowledgments

The work described in this paper is supported by grants from HKU Startup Fund and HKU Seed Fund for Basic Research.

References

1. Abu-Nasser, B.S.: Medical expert systems survey. International Journal of Engineering and Information Systems **1**(7), 218–224 (2017)
2. Van den Berge, K., Mamede, S.: Cognitive diagnostic error in internal medicine. European journal of internal medicine **24**(6), 525–529 (2013)
3. Bian, Y., Wang, S., Yao, Y., Zhang, H.: Foundation-adaptive integrated refinement for generalized category discovery. AAAI **40**(4), 2453–2461 (Mar 2026)
4. Blasi, D.E., Henrich, J., Adamou, E., Kemmerer, D., Majid, A.: Over-reliance on english hinders cognitive science. Trends in cognitive sciences **26**(12), 1153–1170 (2022)
5. Bushberg, J.T., Boone, J.M.: The essential physics of medical imaging. Lippincott Williams & Wilkins (2011)
6. Cao, X., Zheng, X., Wang, G., Yu, W., Shen, Y., Li, K., Lu, Y., Tian, Y.: Solving the catastrophic forgetting problem in generalized category discovery. In: CVPR. pp. 16880–16889 (2024)
7. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: ICCV. pp. 9650–9660 (2021)
8. Chen, H., Miao, S., Xu, D., Hager, G.D., Harrison, A.P.: Deep hiearchical multi-label classification applied to chest x-ray abnormality taxonomies. Medical image analysis **66**, 101811 (2020)
9. Chen, L.: Topological structure in visual perception. Science **218**(4573), 699–700 (1982)
10. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: ICML. pp. 1597–1607. PmLR (2020)
11. Chiaroni, F., Dolz, J., Masud, Z.I., Mitiche, A., Ben Ayed, I.: Parametric information maximization for generalized category discovery. In: ICCV. pp. 1729–1739 (2023)
12. Choi, S., Kang, D., Cho, M.: Contrastive mean-shift learning for generalized category discovery. In: CVPR. pp. 23094–23104 (2024)
13. Croskerry, P., Campbell, S.G., Petrie, D.A.: The challenge of cognitive science for medical diagnosis. Cognitive Research: Principles and Implications **8**(1), 13 (2023)
14. Das, A., Gautam, C., Agrawal, P., Yang, F., Liu, Y., Savitha, R.: Medgcd: Generalized category discovery in medical imaging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 427–437. Springer (2025)
15. Dimitrovski, I., Kocev, D., Loskovska, S., Džeroski, S.: Hierarchical annotation of medical images. Pattern Recognition **44**(10-11), 2436–2449 (2011)
16. Feeman, T.G.: The mathematics of medical imaging. Springer (2010)
17. Feng, W., Jiang, Y., Zhou, S., Ge, Z.: Beyond the static world: Continual category discovery under visual drift. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25032–25042 (2026)

18. Feng, W., Jiang, Y., Zhou, S., Qi, Z., Xu, Z., Wang, Z., Tang, F., Ge, Z.: Seeing through the shift: Causality-inspired robust generalized category discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17766–17775 (2026)
19. Feng, W., Wang, B., Wang, Z., Zhou, S., Ge, Z.: Leveraging image-text pairs for generalized category discovery in medical image classification. IEEE Transactions on Medical Imaging (2026)
20. Feng, W., Zhou, S., Jiang, Y., Ge, Z.: Prism: Progressive robust learning for open-world continual category discovery. In: The Fourteenth International Conference on Learning Representations (2026)
21. Feng, W., Zhou, S., Jiang, Y., Tang, F., Ge, Z.: Neighbor-guided unbiased framework for generalized category discovery in medical image classification. IEEE Journal of Biomedical and Health Informatics (2025)
22. Geng, C., Huang, S.j., Chen, S.: Recent advances in open set recognition: A survey. IEEE transactions on pattern analysis and machine intelligence **43**(10), 3614–3631 (2020)
23. Gilhooly, K.J.: Cognitive psychology and medical diagnosis. Applied cognitive psychology **4**(4), 261–272 (1990)
24. Han, K., Vedaldi, A., Zisserman, A.: Learning to discover novel visual categories via deep transfer clustering. In: ICCV. pp. 8401–8409 (2019)
25. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: CVPR. pp. 9729–9738 (2020)
26. He, L., Cheng, D., Wang, H., Zhu, X., Yang, X., Wang, N., Gao, X.: Task-driven subspace decomposition for knowledge sharing and isolation in lora-based continual learning. In: Forty-third International Conference on Machine Learning (2026)
27. He, L., Cheng, D., Xu, D., Wang, H., Wang, N.: Harnessing textual semantic priors for knowledge transfer and refinement in clip-driven continual learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 21645–21653 (2026)
28. He, Z., Li, L., Chen, L.: Flowcomposer: Composable flows for compositional zero-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12396–12405 (2026)
29. He, Z., Liu, Y., Han, K.: Category discovery: An open-world perspective. arXiv preprint arXiv:2509.22542 (2025)
30. He, Z., Liu, Y., Han, K.: Seal: Semantic-aware hierarchical learning for generalized category discovery. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025)
31. He, Z., Unberath, M., Ke, J., Shen, Y.: Transnuseg: A lightweight multi-task transformer for nuclei segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 206–215. Springer (2023)
32. Holste, G., Wang, S., Jiang, Z., Shen, T.C., Shih, G., Summers, R.M., Peng, Y., Wang, Z.: Long-tailed classification of thorax diseases on chest x-ray: A new benchmark study. In: MICCAI Workshop on Data Augmentation, Labelling, and Imperfections. pp. 22–32. Springer (2022)
33. Itti, L., Baldi, P.: Bayesian surprise attracts human attention. Vision research **49**(10), 1295–1306 (2009)
34. Jha, D., Sharma, V., Dasu, N., Tomar, N.K., Hicks, S., Bhuyan, M.K., Das, P.K., Riegler, M.A., Halvorsen, P., de Lange, T., Bagci, U.: Gastrovision: A multi-class endoscopy image dataset for computer aided gastrointestinal disease detection. In: ICML Workshop on Machine Learning for Multimodal Healthcare Data (ML4MHD 2023) (2023)

35. Lester, H., Arridge, S.R.: A survey of hierarchical non-linear medical image registration. Pattern recognition **32**(1), 129–149 (1999)
36. Liang, P.P., Deng, Z., Ma, M.Q., Zou, J.Y., Morency, L.P., Salakhutdinov, R.: Factorized contrastive learning: Going beyond multi-view redundancy. Advances in Neural Information Processing Systems **36**, 32971–32998 (2023)
37. Liu, Y., Han, K.: DebGCD: Debiased learning with distribution guidance for generalized category discovery. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=9B8o9AxSyb
38. Liu, Y., He, Z., Han, K.: Hyperbolic category discovery. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 9891–9900 (2025)
39. McClelland, J.L., McNaughton, B.L., O'Reilly, R.C.: Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review **102**(3), 419 (1995)
40. McInnes, L., Healy, J., Saul, N., Grossberger, L.: Umap: Uniform manifold approximation and projection. The Journal of Open Source Software **3**(29), 861 (2018)
41. Muennighoff, N., Wang, T., Sutawika, L., Roberts, A., Biderman, S., Le Scao, T., Bari, M.S., Shen, S., Yong, Z.X., Schoelkopf, H., et al.: Crosslingual generalization through multitask finetuning. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 15991–16111 (2023)
42. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)
43. Pu, N., Zhong, Z., Sebe, N.: Dynamic conceptional contrastive learning for generalized category discovery. In: CVPR. pp. 7579–7588 (2023)
44. Rafi, A.H., Bhowmik, P.: HistoSet-5×14: A Collection of Balanced Multi-Organ Histopathology Datasets (2025). https://doi.org/10.17632/nc8k63t7mp.1, https://doi.org/10.17632/nc8k63t7mp.1
45. Rao, R.P., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience **2**(1), 79–87 (1999)
46. Rastegar, S., Salehi, M., Asano, Y.M., Doughty, H., Snoek, C.G.: Selex: Self-expertise in fine-grained generalized category discovery. arXiv preprint arXiv:2408.14371 (2024)
47. Roessle, B., Barron, J.T., Mildenhall, B., Srinivasan, P.P., Nießner, M.: Dense depth priors for neural radiance fields from sparse input views. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12892–12901 (2022)
48. Roy, S., Liu, M., Zhong, Z., Sebe, N., Ricci, E.: Class-incremental novel class discovery. In: ECCV. pp. 317–333. Springer (2022)
49. Saito, K., Kim, D., Saenko, K.: Openmatch: Open-set semi-supervised learning with open-set consistency regularization. Advances in Neural Information Processing Systems **34**, 25956–25967 (2021)
50. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E.: Toward open set recognition. IEEE transactions on pattern analysis and machine intelligence **35**(7), 1757–1772 (2012)
51. Shetab Boushehri, S., Gruber, A., Sun, X., Kazeminia, S., Matek, C., Hehr, M., Spiekermann, K., Pohlkamp, C., Haferlach, T., Marr, C.: A large publicly available single-cell peripheral blood dataset (mll23) (Jan 2023). https://doi.org/10.5281/zenodo.14277609, https://doi.org/10.5281/zenodo.14277609

52. Shortliffe, E.H.: Medical expert systems—knowledge tools for physicians. Western Journal of Medicine **145**(6), 830 (1986)
53. Song, Y., Keller, A., Sebe, N., Welling, M.: Flow factorized representation learning. Advances in Neural Information Processing Systems **36**, 49761–49782 (2023)
54. Sung, M., Kim, V.G., Angst, R., Guibas, L.: Data-driven structural priors for shape completion. ACM Transactions on Graphics (TOG) **34**(6), 1–11 (2015)
55. Tang, L., Huang, K., Chen, C., Chen, C.: Bures-isotropy alignment: Manifold learning of generalized category discovery. In: The Fourteenth International Conference on Learning Representations
56. Tang, L., Huang, K., Chen, C., Yuan, Y., Li, C., Tu, X., Ding, X., Huang, Y.: Dissecting generalized category discovery: Multiplex consensus under self-deconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 297–307 (2025)
57. Thagard, P.: Cognitive science. In: The Routledge companion to philosophy of science, pp. 597–608. Routledge (2013)
58. Thirion, J.P., Prima, S., Subsol, G., Roberts, N.: Statistical analysis of normal and abnormal dissymmetry in volumetric medical images. Medical Image Analysis **4**(2), 111–121 (2000)
59. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Generalized category discovery. In: CVPR. pp. 7492–7501 (2022)
60. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Open-set recognition: A good closed-set classifier is all you need. In: ICLR2022. OpenReview.net (2022)
61. Vaze, S., Vedaldi, A., Zisserman, A.: No representation rules them all in category discovery. In: NeurIPS. vol. 36 (2024)
62. Wang, H., Vaze, S., Han, K.: Sptnet: An efficient alternative framework for generalized category discovery with spatial prompt tuning. In: ICLR (2024)
63. Wen, G., Cao, P., Bao, H., Yang, W., Zheng, T., Zaiane, O.: Mvs-gcn: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. Computers in biology and medicine **142**, 105239 (2022)
64. Wen, X., Zhao, B., Qi, X.: Parametric classification for generalized category discovery: A baseline study. In: ICCV. pp. 16590–16600 (2023)
65. Yang, Y., Wang, Z.Y., Liu, Q., Sun, S., Wang, K., Chellappa, R., Zhou, Z., Yuille, A., Zhu, L., Zhang, Y.D., et al.: Medical world model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8319–8329 (2025)
66. Yang, Y., Zhang, H., Gichoya, J.W., Katabi, D., Ghassemi, M.: The limits of fair medical imaging ai in real-world generalization. Nature medicine **30**(10), 2838–2848 (2024)
67. Yang, Z., Pan, J., Yang, Y., Shi, X., Zhou, H.Y., Zhang, Z., Bian, C.: Proco: Prototype-aware contrastive learning for long-tailed medical image classification. In: International conference on medical image computing and computer-assisted intervention. pp. 173–182. Springer (2022)
68. Yang, Z., Dong, L., Du, X., Cheng, H., Cambria, E., Liu, X., Gao, J., Wei, F.: Language models as inductive reasoners. In: Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 209–225 (2024)
69. Yao, Q., He, Z., Li, Y., Lin, Y., Ma, K., Zheng, Y., Zhou, S.K.: Adversarial medical image with hierarchical feature hiding. IEEE Transactions on Medical Imaging **43**(4), 1296–1307 (2023)
70. Yilmaz, A., Yasar, S.P., Gencoglan, G., Temelkuran, B.: Derm12345: A large, multisource dermatoscopic skin lesion dataset with 40 subclasses. Scientific Data **11**(1), 1302 (2024)

71. Yoshihashi, R., Shao, W., Kawakami, R., You, S., Iida, M., Naemura, T.: Classification-reconstruction learning for open-set recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4016–4025 (2019)
72. Yu, Z., Nguyen, T.D., Ju, L., Gal, Y., Sashindranath, M., Bonnington, P., Zhang, L., Mar, V., Ge, Z.: Hierarchical skin lesion image classification with prototypical decision tree. NPJ Digital Medicine **8**(1), 26 (2025)
73. Zhang, R., Xu, Q., Yao, J., Zhang, Y., Tian, Q., Wang, Y.: Federated domain generalization with generalization adjustment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3954–3963 (2023)
74. Zhang, S., Zhang, Q., Zhang, S., Liu, X., Yue, J., Lu, M., Xu, H., Yao, J., Wei, X., Cao, J., et al.: A generalist foundation model and database for open-world medical image segmentation. Nature Biomedical Engineering pp. 1–16 (2025)
75. Zhang, Y., Kang, B., Hooi, B., Yan, S., Feng, J.: Deep long-tailed learning: A survey. IEEE transactions on pattern analysis and machine intelligence **45**(9), 10795–10816 (2023)
76. Zhao, B., Wen, X., Han, K.: Learning semi-supervised gaussian mixture models for generalized category discovery. In: ICCV. pp. 16623–16633 (2023)
77. Zheng, Q., Zhao, W., Wu, C., Zhang, X., Dai, L., Guan, H., Li, Y., Zhang, Y., Wang, Y., Xie, W.: Large-scale long-tailed disease diagnosis on radiology images. Nature Communications **15**(1), 10147 (2024)
78. Zhong, Z., Zhu, L., Luo, Z., Li, S., Yang, Y., Sebe, N.: Openmix: Reviving known knowledge for discovering novel visual categories in an open world. In: CVPR. pp. 9462–9470 (2021)