

Category Discovery: An Open-World Perspective

Zhenqi He, Yuanpei Liu, Kai Han

Abstract—Category discovery (CD) is an emerging open-world learning task, which aims at automatically categorizing unlabelled data containing instances from unseen classes, given some labelled data from seen classes. This task has attracted significant attention over the years and leads to a rich body of literature trying to address the problem from different perspectives. In this survey, we provide a comprehensive review of the literature, and offer detailed analysis and in-depth discussion on different methods. Firstly, we introduce a taxonomy for the literature by considering two base settings, namely novel category discovery (NCD) and generalized category discovery (GCD), and several derived settings that are designed to address the extra challenges in different real-world application scenarios, including continual category discovery, skewed data distribution, federated category discovery, etc. Secondly, for each setting, we offer a detailed analysis of the methods encompassing three fundamental components, representation learning, label assignment, and estimation of class number. Thirdly, we benchmark all the methods and distill key insights showing that large-scale pretrained backbones, hierarchical and auxiliary cues, and curriculum-style training are all beneficial for category discovery, while challenges remain in the design of label assignment, the estimation of class numbers, and scaling to complex multi-object scenarios. Finally, we discuss the key insights from the literature so far and point out promising future research directions. We compile a living survey of the category discovery literature at <https://github.com/Visual-AI/Category-Discovery>.

Index Terms—Category Discovery, Open-World Learning, Knowledge Transfer

I. INTRODUCTION

DEEP learning algorithms have demonstrated exceptional performance, often surpassing human performance in tasks such as image classification [1], [2], object detection [3], [4], and semantic segmentation [5], [6]. However, these methods typically operate under a closed-world assumption, relying heavily on large amounts of labelled data and predefined labels, which limits their applicability in real-world scenarios. Semi-supervised learning [7], [8] alleviates this limitation by utilizing a small amount of labelled data alongside a large pool of unlabelled data, reducing the reliance on costly annotations. Yet, semi-supervised learning remains restricted by the assumption of fixed categories. In contrast, Open-World Learning allows models to encounter, recognize, and incorporate unseen classes beyond the predefined label set.

Among open-world learning tasks, Open-Set Recognition (OSR) [9], [10] and Out-Of-Distribution (OOD) Detection [11], [12] focus on flagging novel samples as “unknown” but stop short of grouping them, while Category Discovery goes further by clustering the entire unlabelled dataset into semantically coherent groups. As illustrated in Fig. 1, Category Discovery (CD) distinguishes itself from semi-supervised

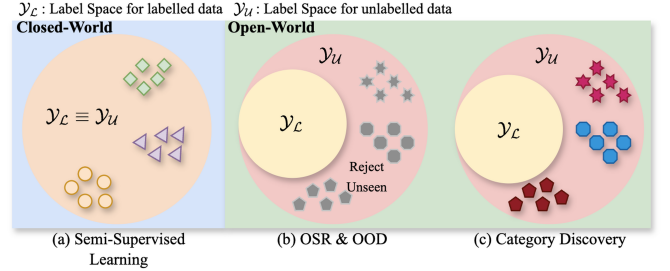


Fig. 1. Comparison of category discovery with semi-supervised learning and OSR&OOD.

learning and OSR&OOD by clustering unlabelled data that includes unseen categories. Drawing inspiration from how humans classify new species based on existing knowledge, CD has proven valuable in many real-world scenarios. For instance, in botanical research, taxonomists use labelled specimens of known species to organize newly collected unlabelled samples into existing taxa or to reveal novel species [13], [14], mirroring the core idea of CD: leverage knowledge from annotated data to cluster and discover both known and previously unknown categories.

In recent years, CD has garnered increasing attention, leading to a proliferation of research exploring various methodologies and settings. CD is initially introduced as Novel Category Discovery (NCD) in [15] to cluster unlabelled novel categories by leveraging knowledge from labelled base categories. It is then extended to Generalized Category Discovery (GCD) [16], which removes the assumption that unlabelled samples only come from novel classes. Instead, GCD accommodates a mixture of both seen and unseen categories in unlabelled data, aligning the task more closely with real-world scenarios. Further advancing category discovery, recent methods have explored a wide range of challenging and realistic settings. Continual Category Discovery (CCD) [17], [18] addresses incremental learning over time, while Federated Category Discovery (FCD) [19] focuses on decentralized model training under privacy constraints. Semantic Category Discovery (SCD) [20] targets assigning semantic labels from an open vocabulary to unlabelled data. Other efforts tackle few-shot scenarios [21], long-tailed distributions [22], and domain shifts [23], enhancing the practicality and robustness of CD methods in real-world applications.

With the rapid advancements in CD, it is increasingly important to track and evaluate the latest advancements in this area. As illustrated in Fig. 2, CD research has surged since 2022 across diverse scenarios. This surge motivates us to provide a thorough overview of existing methods and settings within the CD community, along with comprehensive evaluation results on widely used datasets through standardized reimplementations, facilitating fair comparisons. Although previous surveys [24], [25] have discussed CD, [25] focuses predominantly

Z. He, Y. Liu, and K. Han are with the School of Computing and Data Science, The University of Hong Kong, Hong Kong SAR, China. Corresponding author: Kai Han (kaihanx@hku.hk)

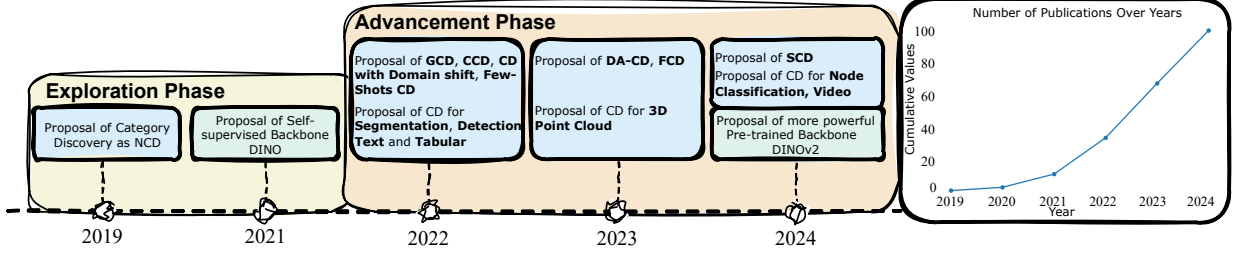


Fig. 2. Roadmap and the cumulative number of publications for the development of Category Discovery from 2019 to 2024.

on NCD, with limited attention to broader and more practical scenarios including GCD, continual category discovery, and other related settings, and [24] presents a review of open-world machine learning, framing category discovery as a central component alongside unknown rejection and continual learning. We provide a focused and in-depth analysis specifically on *Category Discovery*, including systematic analysis of category discovery methods across multiple scenarios, and providing a holistic overview of the current landscape and highlighting future research directions.

Contributions. In this survey, we first provide a systematic review of the current Category Discovery literature with a comprehensive taxonomy. Particularly, we group methods into *Novel Category Discovery* and *Generalized Category Discovery* based on their assumptions on the unlabelled data—we denote this as *base settings*; Building on the base settings, we introduce seven derived settings that adapt category discovery to more practical conditions, including continual learning, decentralized federated training, domain shifts, imbalanced data distributions. To highlight the broader applicability of Category Discovery, we also examine recent advancements extending category discovery beyond visual data. Finally, we conclude with benchmark comparisons across standard datasets, highlighting comparative strengths and limitations to provide actionable insights for future research.

Outline. Fig. 3 outlines the structure of this survey. First, we provide a detailed analysis of problem definitions and related research domains in Sec. II. After that, we present a detailed taxonomy for both base and derived settings in Sec. III. We then move to Sec. IV, which introduces commonly used datasets in CD and evaluation protocols, followed by a performance comparison of these methods. Finally, we explore open questions and potential future research directions in Sec. V, culminating with a conclusion in Sec. VI.

II. BACKGROUND

Overview: This section begins with a clarification of the definitions for base Category Discovery settings in Sec. II-A, followed by a comparison with related research areas including clustering, Semi-Supervised Learning (SSL), Zero-Shot Learning (ZSL), Out-Of-Distribution (OOD) Detection, Open-Set Recognition (OSR), and Open Vocabulary Learning in Sec. II-B with a summary provided in Tab. I.

A. Problem Statements

In the subsequent context, we denote the entire dataset as $\mathcal{D}$, which consists of a labelled subset $\mathcal{D}_L$ and an unlabelled

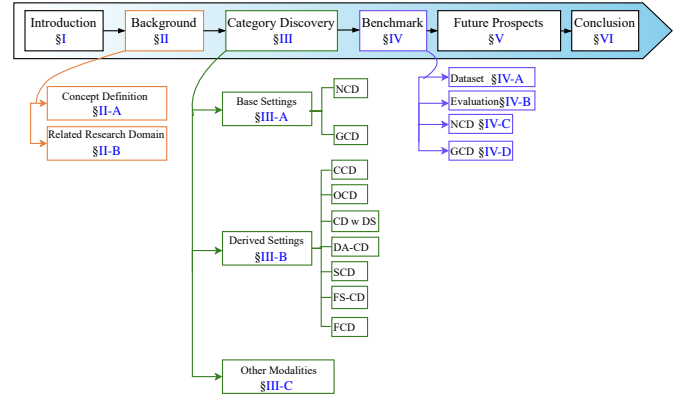


Fig. 3. Overview of this survey.

subset $\mathcal{D}_U$. The corresponding data samples are represented by $\mathcal{X}$, with $\mathcal{X}_L$ and $\mathcal{X}_U$ denoting the labelled and unlabelled portions, respectively. The label spaces associated with the labelled and unlabelled sets are denoted by $\mathcal{Y}_L$ and $\mathcal{Y}_U$. We refer to the sets of base and novel categories as $\mathcal{C}_B$ and $\mathcal{C}_N$.

Novel Category Discovery (NCD). NCD aims to transfer the knowledge learned from base categories to cluster unlabelled unseen categories. This draws inspiration from the observation that *a child can readily differentiate between novel categories, such as birds and elephants, after mastering the classification of base categories like dogs and cats* [15]. Formally, given a dataset $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_U$, where the labelled portion is $\mathcal{D}_L = \{(\mathbf{x}_i, y_i)\}_{i=1}^M \subset \mathcal{X} \times \mathcal{Y}_L$ and the unlabelled portion is $\mathcal{D}_U = \{(\mathbf{x}_i, \hat{y}_i)\}_{i=1}^K \subset \mathcal{X} \times \mathcal{Y}_U$ (with the labels $\hat{y}_i$ being inaccessible during training), the objective of NCD is to leverage the discriminative information learned from the annotated data to cluster the unlabelled data. It presumes that the label spaces of the labelled and unlabelled data are disjoint, i.e., $\mathcal{Y}_L \cap \mathcal{Y}_U = \emptyset$, while also assuming a high degree of semantic similarity between the base and novel categories.

Generalized Category Discovery (GCD). Extending the NCD paradigm, Generalized Category Discovery [16] relaxes the assumption of disjoint label spaces between base and novel categories, offering a more realistic and challenging scenario. In GCD, the labelled and unlabelled datasets share common categories, i.e., $\mathcal{Y}_L \cap \mathcal{Y}_U \neq \emptyset$, where the unlabelled data contains both seen and unseen categories. This broader formulation aligns closely with real-world applications, such as plant species discovery, where an existing database of known species is expanded with newly discovered species, requiring the clustering of both seen and unseen instances.

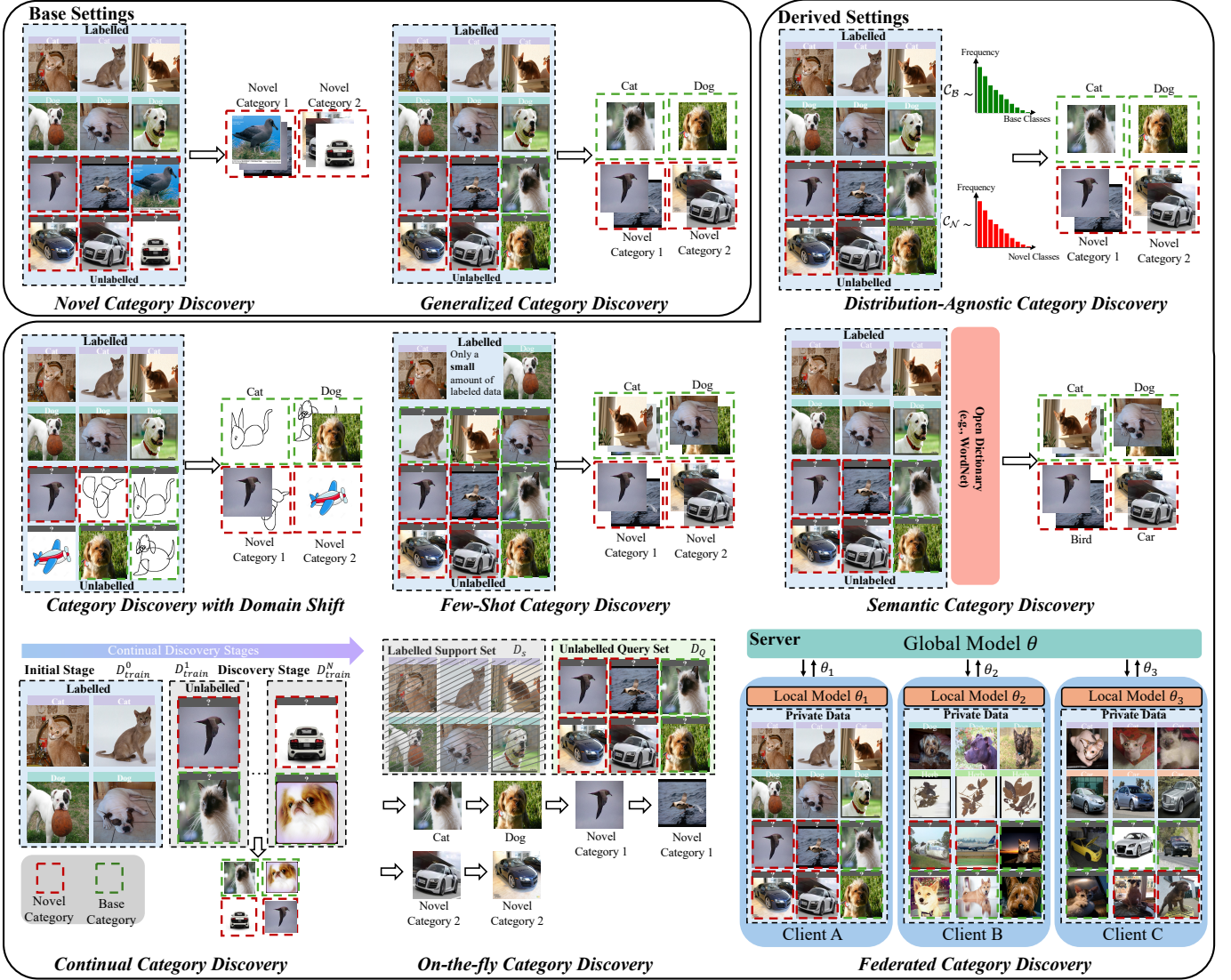


Fig. 4. Overview of the category discovery taxonomy for both *base settings* and *derived settings*. The base settings contain Novel Category Discovery, where only novel categories are present in the unlabelled set, and Generalized Category Discovery, where the unlabelled set contains both base and novel categories. For *derived settings*: Distribution-Agnostic Category Discovery: Existence of imbalanced data distribution; Category Discovery with Domain Shift: Domain shifts within data; Few-Shot Category Discovery: Only a small number of labelled data is available; Semantic Category Discovery: Assignment of semantic labels of unlabelled data; Continual Category Discovery: Discovering novel categories continually; On-the-fly Category Discovery: Instantaneously identifying novel categories under inductive learning; Federated Category Discovery: decentralized training across clients.

Notably, an equivalent formulation has been introduced by Cao *et al.* [26] under the designation of *Open-World Semi-Supervised Learning*. We uniformly use the term *Generalized Category Discovery* in this survey.

B. Related Research Problems

Transfer Learning. Transfer learning leverages knowledge from an annotated source dataset to improve performance on a target dataset by fine-tuning a pre-trained model [27], [28]. This paradigm enables the adaptation of learned representations to new tasks or domains, enhancing generalization when annotated data is available.

Semi-Supervised Learning (SSL). SSL reduces the need for extensive annotations by training on a small labelled set and a large unlabelled set [7], [8]. In contrast to CD, SSL assumes a closed-world setting where both sets share the same categories, *i.e.*, $\mathcal{Y}_L \equiv \mathcal{Y}_U$.

Out-Of-Distribution (OOD) Detection & Open-Set Recognition (OSR). OOD and OSR address open-world challenges by detecting and rejecting test samples that differ from the training distribution [9]–[12]. Particularly, OSR focuses on classifying instances from known categories while rejecting instances from unseen classes, whereas OOD detection focuses on a more general problem of identifying instances deviating from the training data distribution.

Open Vocabulary Learning (OVL). OVL addresses the classification of both base and novel categories by integrating auxiliary vision-aware language vocabularies. This paradigm is typically trained using data-label-vocabulary triplets, for example, $\mathcal{D}_{train} = \{(\mathcal{X}_L, \mathcal{Y}_L, \mathcal{V})\}$, where $\mathcal{V}$ represents auxiliary vocabulary information drawn from an extensive vocabulary space [29], [30]. The primary focus of this task is to leverage pretrained models with aligned visual and textual representations, thereby improving generalization in open-world scenarios.

TABLE I
COMPARISON OF NCD AND GCD WITH RELATED TASKS.

	Training		Testing	
	Labelled	Unlabelled	Novel	Base
SSL	✓	✓	✗	✓
OOD & OSR	✓	✗	Reject	✓
Clustering	✗	✗	✓	✗
ZSL	✓	✗	✓	✗
OVL	✓	✗	✓	✓
NCD	✓	✓	✓	✗
GCD	✓	✓	✓	✓

Clustering. Clustering partitions data into groups based on inherent similarities, without relying on predefined labels or classification criteria [31]. It typically employs distance or similarity measures to reveal latent structures, facilitating the discovery of natural groupings without external annotations.

Zero-Shot Learning (ZSL). ZSL enables recognition of novel categories unseen during training by leveraging auxiliary semantic information [32], [33]. It links representations of base categories to novel ones using semantic attributes, textual descriptions, or other embeddings, which serve as the bridge between seen and unseen classes.

III. CATEGORY DISCOVERY

Overview: In this section, we present a comprehensive taxonomy of category discovery methods, organized from two levels: *base settings* and *derived settings*. In Sec. III-A, we introduce two base settings of Category Discovery - NCD and GCD based on how they define the label space for unlabelled data. In Sec. III-B, we shift focus to the specific objectives and application contexts, grouping them into seven derived settings: CCD, OCD, CD with Domain Shift, DA-CD, SCD, FS-CD, and FCD. Finally, Sec. III-C explores category discovery on other modalities.

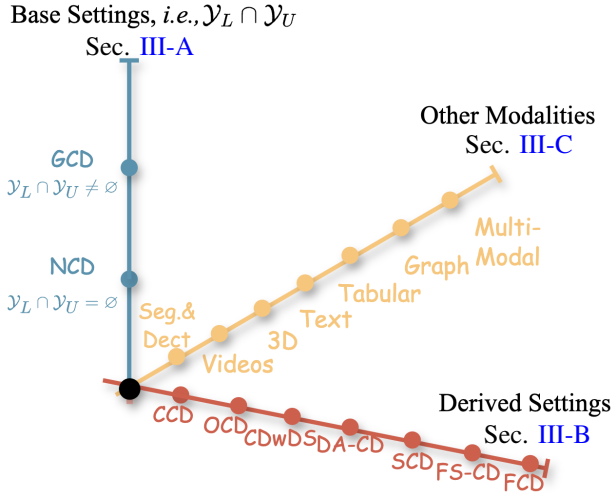


Fig. 5. **Multi-faceted taxonomy.** We present a multi-faceted taxonomy of Category Discovery spanning base settings, derived settings, and other modalities.

Motivation of Taxonomy: A single-faceted taxonomy is insufficient to capture the diversity of methodologies in category discovery. Given the variety of settings, each with its unique challenges and objectives, a more nuanced taxonomy is necessary to categorize these methods effectively. Therefore, we

offer a comprehensive review of the different configurations of category discovery, considering both *base* and more challenging *derived* settings.

A. Base Settings

In this subsection, we introduce two base settings based on the definition of the label space for unlabelled data, categorizing methods into NCD and GCD where detailed definitions are provided in Sec. II-A. We then discuss conventional category discovery methods for image classification within this taxonomy, while an overview of methods addressing other scenarios and data modalities is presented in Sec. III-B and Sec. III-C. Both NCD and GCD frameworks generally consist of three core components: *representation learning*, *label assignment*, and *estimation of the number of categories*. We provide a comprehensive understanding of how they address these key aspects of category discovery.

Novel Category Discovery. Han *et al.* [15] introduce Novel Category Discovery (NCD) by adapting Deep Embedded Clustering (DEC) [55] within a transfer learning framework that integrates representation bottlenecks, temporal ensembling, and consistency regularization to cluster novel categories in unlabelled data. KCL [34] and MCL [35] which are originally designed for cross-task transfer, can also be adopted for NCD. They employ similarity prediction networks trained on labelled data to enhance classification on unlabelled categories.

Early NCD methods largely focus on adapting clustering and self-supervised learning techniques to discover novel categories, often integrating transfer learning, representation learning, and pseudo-labelling strategies within unified training frameworks. RankStats (aka AutoNovel) [36], [56] leverages self-supervised learning on both labelled and unlabelled data to mitigate bias toward known categories, generating pseudo-labels using rank statistics followed by fine-tuning. [37] introduces an end-to-end learning framework integrating pairwise similarity learning with self-supervised learning. NCL [39] utilises the local neighbourhood of samples in the embedding space to generate pseudo-positive pairs and hard negatives, enhancing contrastive learning. OpenMix [38] applies MixUp [57] to blend labelled and unlabelled samples, generating more reliable pseudo-labels and uses high-confidence unlabelled samples as reliable anchors to enhance the training process. UNO [40] unifies labelled and unlabelled data classification with a multi-view self-labelling strategy, leveraging the Sinkhorn-Knopp algorithm [58] to generate balanced, high-quality pseudo-labels. DualRS [41] uses a bi-branch learning framework, combining global and local features with dual ranking statistics and mutual knowledge distillation. Squared Mutual Information (SMI) [42] proposes a dependency-based approach that uses SMI to bridge supervised learning and unsupervised clustering. [59] introduce Spacing Loss, a simple yet effective method by improving the separability of categories in the latent space. ResTune [45] introduces a disentangled representation learning framework for NCD, separating the learning process into basic and residual features. The framework optimizes the combined representation using three objectives: clustering for novel categories, knowledge distillation to prevent forgetting, and pairwise labelling to

TABLE II
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED NCD APPROACHES.

Year	Method	Pub.	Backbone	Label Assignment
2018	KCL [34]	<i>ICLR</i>	ResNet	Parametric Classifier
2019	MCL [35]	<i>ICLR</i>	ResNet, VGG, LeNet	Parametric Classifier
	DTC [15]	<i>ICCV</i>	ResNet, VGG	Soft Assignment
2020	RankStats, RankStats+ [36]	<i>ICLR</i>	ResNet	Parametric Classifier
2021	Qing <i>et al.</i> [37]	<i>Neural Networks</i>	ResNet	Parametric Classifier
	OpenMix [38]	<i>CVPR</i>	ResNet, VGG	Parametric Classifier
	NCL [39]	<i>CVPR</i>	ResNet	Parametric Classifier
	UNO [40]	<i>ICCV</i>	ResNet	Parametric Classifier
	DualRS [41]	<i>ICCV</i>	ResNet	Parametric Classifier
2022	SMI [42]	<i>ICASSP</i>	VGG	k -means
	PSSCNCD [43]	<i>T-CYB</i>	N/A	BKBBH k -means
	Li <i>et al.</i> [44]	<i>NeurIPS</i>	ResNet	k -means
2023	ResTune [45]	<i>T-NL</i>	ResNet	k -means
	SK-Hurt [46]	<i>TMLR</i>	ResNet	k -means
	IIC [47]	<i>CVPR</i>	ResNet	Parametric Classifier
	NSCL [48]	<i>ICML</i>	ResNet	k -means
	CRKD [49]	<i>ICCV</i>	ResNet, ViT	Parametric Classifier
	Feng <i>et al.</i> [50]	<i>MICCAI</i>	ResNet	Parametric Classifier
2024	RAPL [51]	<i>CVPR</i>	ResNet	k -means
	SCKD [52]	<i>ECCV</i>	ResNet, ViT	Parametric Classifier
	APL [53]	<i>T-PAMI</i>	ResNet	Parametric Classifier
PrePrint	Hasan <i>et al.</i> [54]	<i>ArXiv</i>	ResNet	Parametric Classifier

preserve structural integrity. PSSCNCD [43] builds a bipartite anchor graph to model data relations and progressively assigns pseudo-labels through self-supervised learning.

More recent approaches further introduce explicit constraints, semantic alignment objectives, and advanced architectures to improve feature separability and enhance the effectiveness of novel category discovery. Inter-Class and Intra-Class Constraints (IIC) [47] improves representation learning by incorporating inter-class and intra-class constraints via symmetric Kullback-Leibler divergence. Li *et al.* [44], [46] investigate how semantic similarity between labelled and unlabelled datasets impacts NCD performance. Their work introduces Sk-Hurt-NCD [46], which quantifies semantic similarity through the transfer flow metric and proposes pseudo transfer flow to guide the effective use of supervised knowledge in NCD tasks. NSCL [48] models labelled and unlabelled data as graph nodes and introduces a Spectral Contrastive Loss to discover novel classes by minimizing the spectral decomposition of the adjacency matrix. CRKD [49] proposes a knowledge distillation framework that transfers class relations from known to novel categories, preserving essential semantic information through a learnable regularization process. Feng *et al.* [50] introduce a framework for identifying unseen skin lesion categories in medical imaging with an uncertainty-aware multi-view cross-pseudo-supervision strategy. Liu *et al.* [51] introduce the Region-Aligned Proxy Learning (RAPL) framework, tailored for ultra-fine-grained novel class discovery. It uses channel-wise region alignment to capture local discriminative features and employs semi-supervised proxy learning to model instance-class relationships. SCKD [52] uses separate representation spaces for known and novel classes, promoting mutual learning through spatial mutual information and self-distillation. Zhang *et al.* [53] propose a two-stage method that first learns features via self-supervised and prototype learning, then trains a classifier using k -means pseudo-labels refined by prototypical similarity. Hasan *et al.* [54] introduce a distribution learning framework that combines Monte Carlo

sampling with neuronal activation patterns to align predicted class probabilities with a Multinoulli distribution.

Label assignment in NCD is addressed through both *parametric* and *non-parametric* classifiers. Non-parametric approaches, such as those in [42]–[46], [48], [51], utilise k -means based clustering techniques. DTC [15] employs soft clustering by computing probability distributions based on Euclidean distance between feature vectors and class centres. In contrast, many other methods [34]–[41], [47], [49], [50], [52]–[54], [56], [60] utilise parametric classifiers, employing classification heads to assign labels to data points.

Estimating the number of novel categories in NCD remains a significant challenge, and only a few methods have explicitly tackled this problem. DTC [15] tackles this by semi-supervised k -means with probe classes, selecting the optimal count via joint optimization of probe accuracy and clustering validity. In contrast, most other methods either assume a predefined number of novel categories [61], overestimate the category number [35], or adopt existing approaches, which simplify the process but limit applicability in open-world scenarios.

NCD methods demonstrate strong potential for open-world generalization by clustering unlabelled data through knowledge transfer from labelled base categories. However, most approaches are trained from scratch, overlooking powerful pre-trained features from self-supervised models like DINO [62] and MAE [63], which may limit representation quality in low-data regimes. More critically, NCD assumes that unlabelled data consists only of novel categories—a convenient but unrealistic setup. In practice, unlabelled sets often mix seen and unseen classes, motivating the shift toward the more practical Generalized Category Discovery.

Generalized Category Discovery. Vaze *et al.* [16] extend NCD [15] to GCD by introducing a decoupled representation learning approach. Unlike NCD methods, which mainly train models from scratch, they use a large-scale self-supervised pre-trained model (*e.g.*, DINO [62]) for initialization. The model is then fine-tuned with both supervised and unsuper-

TABLE III
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED GCD APPROACHES.

Year	Method	Pub.	Backbone	Label Assignment
2022	GCD [16]	<i>CVPR</i>	ViT	k -means
	ORCA [26]	<i>CVPR</i>	ResNet	Parametric Classifier
	ComEx [64]	<i>CVPR</i>	ResNet	Parametric Classifier
	OpenLDN [65]	<i>ECCV</i>	ResNet	Parametric Classifier
	TRSSL [66]	<i>ECCV</i>	ResNet	Parametric Classifier
	NACH [67]	<i>NeurIPS</i>	ResNet	Parametric Classifier
2023	XCon [68]	<i>BMVC</i>	ViT	k -means
	OpenCon [69]	<i>TMLR</i>	ResNet	Prototype-based
	PromptCAL [70]	<i>CVPR</i>	ViT	k -means
	DCCL [71]	<i>CVPR</i>	ViT	Infomap
	OpenNCD [72]	<i>IJCAI</i>	ResNet	Prototype-based
	SimGCD [73]	<i>ICCV</i>	ViT	Parametric Classifier
	GPC [74]	<i>ICCV</i>	ViT	GMM
	PIM [75]	<i>ICCV</i>	ViT	Parametric Classifier
	TIDA [76]	<i>NeurIPS</i>	ResNet	Parametric Classifier
	μ GCD [77]	<i>NeurIPS</i>	ResNet, ViT	k -means
	InfoSieve [78]	<i>NeurIPS</i>	ViT	k -means
	SORL [79]	<i>NeurIPS</i>	ResNet	k -means
	CAC [80]	<i>ICONIP</i>	ViT	Louvain
2024	CiPR [81]	<i>TMLR</i>	ViT	<i>SNC</i>
	AMEND [82]	<i>WACV</i>	ViT	Parametric Classifier
	GCA [83]	<i>WACV</i>	ViT	Guided Cluster Aggregation
	SPT-Net [84]	<i>ICLR</i>	ViT	Parametric Classifier
	LegoGCD [85]	<i>CVPR</i>	ViT	Parametric Classifier
	CMS [86]	<i>CVPR</i>	ViT	Agglomerative Clustering
	TRAILER [87]	<i>CVPR</i>	ResNet, ViT	Parametric Classifier
	ActiveGCD [88]	<i>CVPR</i>	ViT	Parametric Classifier
	TextGCD [89]	<i>ECCV</i>	ViT	CVCS
	Yang <i>et al.</i> [90]	<i>ECCV</i>	ViT	Parametric Classifier
	SelEx [91]	<i>ECCV</i>	ViT	k -means
	LPS [92]	<i>IJCAI</i>	ResNet	Parametric Classifier
	OwMatch [93]	<i>NeurIPS</i>	ResNet	Parametric Classifier
	FlipClass [94]	<i>NeurIPS</i>	ViT	Parametric Classifier
	Contextuality-GCD [95]	<i>ICIP</i>	ViT	Parametric Classifier
2025	MSGCD [96]	<i>Information Fusion</i>	ViT	Parametric Classifier
	CPT [97]	<i>IJCV</i>	ViT	Similarity-based
	PAL-GCD [98]	<i>AAAI</i>	ViT	Parametric Classifier
	DebGCD [99]	<i>ICLR</i>	ViT	Parametric Classifier
	ProtoGCD [100]	<i>T-PAMI</i>	ViT	Parametric Classifier
	MOS [101]	<i>CVPR</i>	ViT	Parametric Classifier
	GET [102]	<i>CVPR</i>	ViT	Parametric Classifier
	AptGCD [103]	<i>CVPR</i>	ViT	Parametric Classifier
	Dai <i>et al.</i> [104]	<i>CVPR</i>	ViT	-
	HypCD [105]	<i>CVPR</i>	ViT	-
	ConGCD [106]	<i>ICCV</i>	ViT	-
	AF [107]	<i>ICCV</i>	ViT	-
PrePrint	SEAL [108]	<i>NeurIPS</i>	ViT	Parametric Classifier
	CLIP-GCD [109]	<i>ArXiv</i>	ViT	k -means
	MCDL [110]	<i>ArXiv</i>	ViT	Parametric Classifier
	PNP [111]	<i>ArXiv</i>	ViT	Infomap
	RPIM [112]	<i>ArXiv</i>	ViT	Parametric Classifier
	OpenGCD [113]	<i>ArXiv</i>	ViT	Parametric Classifier
	ConceptGCD [114]	<i>ArXiv</i>	ViT	Parametric Classifier
	DIG-FACE [115]	<i>ArXiv</i>	ViT	Parametric Classifier
PrePrint	NN-GCD [116]	<i>ArXiv</i>	ViT	SNMF

vised contrastive learning on labelled and unlabelled data. ComEx [64] introduces batch-wise and class-wise experts to manage both base and novel classes, learning discriminative representations across them. NACH [67] mitigates the learning rate disparity between seen and unseen classes by synchronizing learning through pairwise similarity and adaptive distribution alignment. XCon [68] partitions the dataset into sub-datasets using k -means clustering based on DINO [62] representations, followed by contrastive learning to capture fine-grained features.

Recent advances in GCD further explore prototype-driven, prompt-based, hierarchical, and hash-based designs, offering complementary strengths to contrastive learning-based baselines and facilitating more effective category discovery. PromptCAL [70] introduces a two-stage prompt-based framework, combining discriminative prompt regularization

and contrastive affinity learning to iteratively refine semantic representations. DCCL [71] proposes dynamic conceptual contrastive learning, alternating between instance-level and abstract conception-level with dynamically generated conceptions. GPC [74] uses a semi-supervised Gaussian Mixture Model (GMM) with dynamic splitting and merging, along with Prototypical Contrastive Learning, to improve feature representation. PIM [75] introduces bi-level optimization to maximize mutual information between data features and labels while addressing class-balance bias with a parametric family of objectives. μ GCD [77] introduces the Clevr-4 dataset, designed to evaluate GCD methods across different taxonomies, and highlights potential biases in large pre-trained models like DINO [62] and MAE [63]. It adopts a mean-teacher approach to overcome GCD limitations of unstable pseudo-labels. InfoSieve [78] generates binary category codes

through self-supervised contrastive learning and optimizes mutual information between features and codes, organizing them in a hierarchical binary tree for efficient categorization. Extended from NSCL [48], SORL [79] introduces a graph-theoretic framework that uses spectral decomposition to improve data representation by incorporating both self-supervised and supervised signals. CAC [80] employs weak and strong augmentations to generate diverse views of samples, ensuring clustering assignment consistency via feature-prototype similarity alignment. CiPR [81] introduces cross-instance positive relations between labelled and unlabelled data by leveraging a novel selective neighbor clustering algorithm, which generates high-purity pseudo-labels to strengthen representation learning. CMS [86] combines contrastive learning with mean-shift clustering to refine image embeddings by generating mean-shifted representations through nearest neighbors. Yang *et al.* [90] propose a Neighbor Graph Convolutional Network for improved pseudo-labelling by leveraging k -nearest neighbor relationships and iterative contrastive learning at both instance and class levels. SelEx [91] leverages hierarchical semi-supervised k -means pseudo-labelling and self-expertise to refine clustering and adapt contrastive learning, enhancing fine-grained classification and novel category generalization. ConceptGCD [114] employs a three-stage approach that learns known class concepts, composes derivable novel concepts, and discovers underivable ones via contrastive learning and normalization. PNP [111] integrates learnable potential prototypes with self-distillation to refine cluster centers in an end-to-end manner. RPIM [112] extends PIM [75] with regularization to address unconfident predictions and feature refinement using semantic-bias transformations. MSGCD [96] unifies parametric classification and representation learning using a mutual-support mechanism, enhancing both pseudo-labels and feature representations iteratively. NN-GCD [116] reformulates GCD as an optimization problem, proving that optimal k -means clustering can be achieved using Symmetric Non-negative Matrix Factorization and Non-negative Contrastive Learning. DIG-FACE [115] mitigates implicit bias via F-discrepancy and applies explicit debiasing to enhance facial expression recognition in GCD. Dai *et al.* [104] introduces a plug-and-play adaptive part learning module that uses DINO-derived part priors and shared learnable queries to discover consistent object parts without annotations. HypCD [105] pioneers the use of hyperbolic embeddings to capture hierarchical class relations and introduces a general GCD framework operating in hyperbolic space. ConGCD [106] deconstructs an image into competing visual primitives via reconstruction-guided slot attention, then injects a Multiplex Consensus module into the ViT FFN to aggregate them. AF [107] proposes Attention Focusing, a plug-and-play module that combats distracted attention in GCD by learning token importance from labelled data and adaptively pruning low-value image tokens.

SimGCD [73] introduces a popular end-to-end parametric baseline that jointly trains a feature extractor and a parametric classifier with self-distillation and entropy regularization, outperforming previous two-stage methods. It has since become a popular foundation for developing more advanced GCD approaches. Numerous subsequent works extend or build upon

its design to address specific limitations and enhance performance. Building on SimGCD [73], SPT-Net [84] presents a two-stage iterative framework introducing additional spatial prompt tuning, optimizing image patch-based discriminative features. DebGCD [99] introduces an auxiliary debiased classifier to SimGCD [73] trained with hard labels to mitigate label bias, alongside a semantic distribution detector that identifies shifts between known and unknown classes in a separate feature space. AMEND [82] proposes Expanded Neighbourhood Contrastive Learning using both nearest and expanded neighbors for robust contrastive pairing, along with a Class-wise Adaptive Margin Regularizer to improve fine-grained category separation. LegoGCD [85] builds on SimGCD to mitigate catastrophic forgetting by introducing Local Entropy Regularization for stabilizing high-confidence predictions and Dual-view KL Divergence to improve reliable known-sample selection. FlipClass [94] proposes dynamic teacher-student attention alignment, based on Hopfield Network energy functions, by updating the teacher’s attention using student feedback through energy minimization. Contextuality-GCD [95] augments contrastive learning with instance-level context and cluster-level prototypical cues. PAL-GCD [98] introduces a greedy association algorithm for Prior-Constrained Association Learning to SimGCD [73], ensuring each cluster contains at most one known category to avoid incorrect merging. ProtoGCD [100] proposes a unified prototype-based framework that introduces a shared prototype classifier and a dual-level adaptive pseudo-labelling mechanism to mitigate confirmation bias and balance learning across base and novel classes. MCDL [110] uses dual memory banks to assess sample reliability and splits data by intra- and inter-memory consistency for targeted learning. ActiveGCD [88] introduces an active-learning framework to select informative samples from unlabelled data, ensuring robust label assignment through an exponential moving average model. MOS [101] employs an extra zero-shot saliency segmentation model to decouple object and scene regions, and feeds both through a shared backbone with an MLP-based scene-awareness module to model object–scene associations. AptGCD [103] leverages inner-patch Meta Visual Prompts and a Prompt Transformer with global self-attention to efficiently fuse local and global features for distinguishing base and novel classes. SEAL [108] introduces a semantic-aware hierarchical learning framework for GCD, where semantically consistent coarse-to-fine supervision is propagated from labelled to unlabelled data via semantic-aware hierarchical contrastive learning and a cross-granularity consistency module.

In parallel with the development of GCD, a number of methods originating from the Open-World Semi-Supervised Learning (OWSSL) community—conceptually equivalent to GCD—have been proposed to address the challenge of discovering novel categories under partially labelled data. [26] introduces ORCA, an end-to-end framework that adaptively adjusts decision margins based on uncertainty, enabling robust learning across both seen and unseen categories. Building upon this foundation, OwMatch [93] enhances self-labelling by conditioning pseudo-labels on labelled data to reduce confirmation bias, and introduces hierarchical thresholding to balance con-

fidence between base and novel classes. Similarly, LPS [92] focuses on addressing learning imbalance by combining an adaptive margin loss with a contrastive clustering objective tailored for novel class discovery. OpenLDN [65] converts OWSSL into a closed-world task by using regression-based pairwise similarity to generate pseudo-labels and integrate novel samples into the labeled set. OpenCon [69] utilizes prototype-based learning to generate compact and distinguishable representations, applying contrastive loss to both known and novel data within an OWSSL framework. TRSSL [66] introduces a pseudo-labelling method that integrates prior class distribution knowledge and sample uncertainty to generate reliable pseudo-labels for both novel and base classes. OpenNCD [72] utilizes a bi-level contrastive learning approach at both the prototype and group levels to enhance representation learning. TIDA [76] introduces a hierarchical approach that constructs prototypes at multiple levels (sub-class, target-class, super-class) and aligns predictions across these levels using an affinity matrix. TRAILER [87] aligns representations to a fixed Neural Collapse-inspired classifier and refines pseudo-labels via hierarchical sample-target allocation using optimal transport and PU learning. OpenGCD [113] integrates uncertainty-based OSR with GCD to identify unseen categories, and uses semi-supervised k -means++ for automated class discovery.

Another important direction integrates cross-modal knowledge into GCD, with methods leveraging Vision-Language Models (VLMs) and Large Language Models (LLMs) to incorporate textual information, enhancing category discovery for both base and novel classes. CLIP-GCD [109] is the first to leverage pre-trained VLMs for category discovery by utilizing CLIP’s [117] multi-modal capabilities. It integrates visual features with relevant textual descriptions from a large corpus, enabling joint optimization of image and text embeddings to capture richer semantic information with auxiliary textual cues. TextGCD [89] advances cross-modal learning with a dynamic co-teaching framework, aligning textual and visual features using large language models like GPT-3 [118]. Additionally, GET [102] enhances representation learning by integrating visual and textual features without labelled text, utilizing CLIP’s vision-language alignment to generate pseudo-text embeddings. CPT [97] adapts CLIP to learn “task + class” prompts, ensuring robust category discovery with Vision-Vision and Vision-Language consistency across labelled and unlabelled data.

As with NCD, label assignment plays a critical role in GCD pipelines. Both *parametric* and *non-parametric* classifiers have been explored to support flexible and scalable clustering of novel categories. GCA [83] offers a plug-and-play parametric-classifier solution that seamlessly integrates with existing GCD methods (*e.g.*, [16] and [70]), outperforming traditional clustering methods such as semi-supervised k -means, especially in fine-grained tasks. It employs Guided Cluster Aggregation to hierarchically merge small pure clusters into target categories based on local feature structures. Among non-parametric approaches, OpenCon [69] uses prototype-based assignment, selecting the prototype with the highest softmax-normalized similarity to the projected features. GPC [74] applies a semi-supervised Gaussian Mixture Model (GMM)

for clustering. CMS [86] uses the mean-shift algorithm [119] to iteratively shift data points toward the mode of their neighbours in the embedding space. SMILE [120] employs hash-code-based clustering, while PHE [121] adopts a Hamming ball-based strategy for inference. Both PNP [111] and DCCL [71] utilize Infomap [122] for clustering, and CAC [80] applies the Louvain [123] community detection algorithm for label assignment. Yang *et al.* [90] utilize FINCH [124] and introduces Cross-View Consistency Strategy to refine by comparing clustering results from FINCH [124] and k -means. CiPR [81] utilizes its Selective Neighbor Clustering (SNC), a semi-supervised variant of FINCH [124], to produce hierarchical partitions. PAL-GCD [98] adopts a two-stage framework, using non-parametric clustering for warm-up, then jointly training with a parametric classifier.

Estimating the class number of novel categories remains a core challenge in GCD. GCD [16] addresses this by running k -means with different cluster counts and selecting the optimal number via Brent’s algorithm to maximize accuracy on labelled data. OpenNCD [72] uses Jaccard-based prototype grouping for class estimation and benchmarks it against X-means [125], G-means [126], and DipDeck [127]. GPC [74] integrates class estimation with representation learning using a semi-supervised GMM featuring a dynamic cluster splitting and merging mechanism. CMS [86] applies agglomerative clustering with a Ward linkage criterion to determine optimal clusters during training. CiPR [81] estimates the class number by generating hierarchical partitions through SNC and selecting the optimal level. CAC [80] utilizes the Louvain [123] algorithm for class estimation. PHE enables on-the-fly class discovery by assigning samples to existing categories if their hash lies within a Hamming ball of known centers, or creating a new category otherwise. SMILE [120] also estimates the number of classes with hash-coding by assigning labels to unique binary hashes and counting them as the class count. DCCL [71] and PNP [111] leverage Infomap [122] to detect communities and refine small clusters, determining the final class count. The benchmark for class-number estimation is detailed in the Appendix.

GCD extends NCD to a more realistic scenario where unlabelled data contains both seen and unseen categories, better aligning with real-world conditions and demonstrating the feasibility of open-world learning under partial supervision. Yet, many methods remain biased toward seen classes, limiting novel class discovery, and most studies focus narrowly on general image datasets, leaving domains like medical imaging underexplored. In addition, current approaches often assume balanced distributions and homogeneous domains, which rarely hold in practice. These limitations naturally motivate the development of more complex derived settings.

B. Derived Settings

In the following subsection, we introduce derived settings, which extend category discovery in more practical and challenging tasks to base settings. We identify seven distinct derived settings—CCD, OCD, CD with Domain Shift, DACC, SCD, FS-CD, and FCD. Additionally, we offer an in-depth review of the methods within each setting, highlighting

TABLE IV
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED CCD APPROACHES.

Year	Method	Pub.	Backbone	Scenario	Label Assignment	$\mathcal{Y}_{\mathcal{L}} \cap \mathcal{Y}_{\mathcal{U}}$
2022	NCDwF [128]	ECCV	ResNet	Class Incremental	Parametric Classifier	$\emptyset$
	FRoST [129]	ECCV	ResNet	Class Incremental	Parametric Classifier	$\emptyset$
	GM [17]	NeurIPS	ResNet	All	Parametric Classifier	$\neq \emptyset$
2023	PA-GCD [130]	ICCV	ViT, ResNet	Mixed Incremental	Parametric Classifier	$\neq \emptyset$
	MetaGCD [131]	ICCV	ViT	Mixed Incremental	k -means	$\neq \emptyset$
	iGCD [18]	ICCV	ResNet	Self-Supervised Mixed Incremental	Soft Nearest Neighbor	$\neq \emptyset$
2024	Msc-iNCD [132]	ICPR	ViT	Class Incremental	Parametric Classifier	$\emptyset$
	ADM [133]	AAAI	ResNet	Class Incremental	Parametric Classifier	$\emptyset$
	PromptCCD [134]	ECCV	ViT	Mixed Incremental	GMM	$\neq \emptyset$
	DEAN [135]	ECCV	ViT	Mixed Incremental	Parametric Classifier	$\neq \emptyset$
	CAMP [136]	ECCV	ViT	Self-Supervised Mixed Incremental	Nearest Centroid Classifier	$\neq \emptyset$
	Happy [137]	NeurIPS	ViT	Mixed Incremental	Parametric Classifier	$\neq \emptyset$
2025	VB-CGCD [138]	ICML	ViT	Mixed Class Incremental	Variational Bayes NCM	$\neq \emptyset$
preprint	FEA [139]	ArXiv	ViT	Class Incremental	Parametric Classifier	$\emptyset$

their unique approaches, challenges, and contributions to the category discovery community.

Continual Category Discovery (CCD). CCD [128] provides a continual setting of category discovery in which new categories are identified sequentially while retaining previously acquired knowledge. It begins with an initial stage, where the model is trained on a labelled dataset, denoted as $\mathcal{D}_{train}^0 = \{(\mathbf{x}_i^0, y_i^0)\}_{i=1}^{N^0}$, composed of instances from base categories $\mathcal{C}_B^0$. Subsequently, the model enters a discovery phase in which it is incrementally exposed to new datasets $\{\mathcal{D}_{train}^t\}_{t=1}^T$ over a series of timesteps $\{1, \dots, T\}$. At each timestep, the model is retrained using the current dataset $\mathcal{D}_{train}^t$ to identify novel categories while preserving previously acquired knowledge.

CCD presents several distinct scenarios based on the structure of the incoming data. In the *Class Incremental Scenario* [17], [128], [129], [132], the training set $\mathcal{D}_{train}^t$ contains solely unlabelled instances from novel categories. NCDwF [128] generates pseudo-latent representations to serve as proxies for the discarded labelled data and employs a mutual-information-based regularizer to both enhance novel class discovery and retain performance on known classes. Similarly, FRoST [129] prevents forgetting by storing feature prototypes and applying feature-level knowledge distillation, while Msc-iNCD [132] leverages self-supervised pre-trained models like DINO [62] along with cosine normalization and feature replay to mitigate catastrophic forgetting. ADM [133] decouples representation learning from category discovery using Triplet Comparison and Probability Regularization, and employs Adaptive Model Merging to integrate new knowledge without increasing model size or causing forgetting. FEA further enhances features using a single-head paradigm guided by a prior distribution, incorporating Centroid-to-Samples Similarity and Boundary-Aware Prototype constraints to maintain clear boundaries between known and novel classes.

Class Incremental Scenario assumes that each incoming dataset at a timestep contains only novel categories, which is often an unrealistic premise. In contrast, *mixed-incremental scenario* assumes that $\mathcal{D}_{train}^t$ includes unlabelled data from both base and novel categories, as seen in works like [131], [134], [135], [140]. PA-GCD [130] uses a proxy-anchor approach with affinity propagation [141] for clustering and class estimation, removing the need to predefine category numbers. It also leverages metric-based exemplars to reduce forgetting. MetaGCD [131] presents a meta-learning frame-

work that aligns the training objective with the evaluation metric to prevent forgetting and enhances feature learning via soft neighborhood contrastive learning with adaptive positive sample selection. PromptCCD [134] introduces a Gaussian Mixture Prompting module to dynamically manage prompts for continual category discovery. It combines contrastive learning with GMM-based prompt retrieval to support on-the-fly category estimation and knowledge retention. Class number is estimated via a GMM with a split-and-merge strategy guided by the Hastings ratio. DEAN [135] separates seen and unseen samples via energy-guided discovery, enhances clustering with variance-based augmentation, and applies energy-based contrastive loss to improve novel class discrimination while mitigating forgetting. Class number is estimated via affinity propagation. Happy [137] mitigates prediction and hardness biases via clustering-guided initialization and entropy regularization, and uses Hardness-Aware Prototype Sampling to model class distributions. Class number is estimated by maximizing the silhouette score. VB-CGCD [138] formulates continual category discovery from a Bayesian perspective, modelling each class as a variational Gaussian distribution with covariance-aware nearest-class-mean classification.

Self-supervised mixed-incremental scenario assumes that $\mathcal{D}_{train}^t$ includes both labelled and unlabelled data from base and novel categories. Under this assumption, iGCD [18] combines a non-parametric Soft Nearest-Neighbor classifier with density-based sampling to classify both known and novel categories. It selects density peaks as support samples to reduce forgetting and estimates the class number by counting these peaks. CAMP [136] proposes a three-phase framework combining contrastive learning, semi-supervised clustering, and centroid adaptation to balance novel class discovery and knowledge retention.

Additionally, GM [17] evaluates all these scenarios by introducing an inductive continual category discovery setting. In this approach, the model is first trained on a labelled dataset and then updated across successive datasets, with testing on unlabelled sets at each timestep. It alternates between learning novel categories and merging them into a static branch via distillation, enabling continual discovery while preserving prior knowledge.

On-the-fly Category Discovery (OCD). OCD extends conventional category discovery to an inductive learning paradigm with streaming inference. It trains on a labelled support set D_S

TABLE V
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED OCD APPROACHES.

Year	Method	Pub.	Backbone	Label Assignment
2023	SMILE [120]	<i>CVPR</i>	ViT	Hash-based
2024	PHE [121]	<i>NeurIPS</i>	ViT	Hamming Ball-Based
2024	DiffGRE [142]	<i>ICCV</i>	ViT	Online Clustering

TABLE VI
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED CD WITH DOMAIN SHIFT APPROACHES.

Year	Method	Pub.	Backbone	$\Omega_{\mathcal{U}}$	Label Assignment	$\mathcal{Y}_{\mathcal{L}} \cap \mathcal{Y}_{\mathcal{U}}$
2022	Yu <i>et al.</i> [143]	<i>AAAI</i>	ResNet	Single New Domain	Parametric Classifier	$\emptyset$
	SCDA [144]	<i>ICME</i>	ResNet	Multiple New Domains	Parametric Classifier	$\emptyset$
2023	SAN [145]	<i>ICCV</i>	ResNet	Single New Domain	Parametric Classifier	$\emptyset$
2024	CDAD-Net [146]	<i>CVPRW</i>	ViT	Single New Domain	k -means	$\neq \emptyset$
2025	HiLo [23]	<i>ICLR</i>	ViT	Multiple New Domains	Parametric Classifier	$\neq \emptyset$
Preprint	Wang <i>et al.</i> [147]	<i>ArXiv</i>	ViT	Single New Domain	Parametric Classifier	$\emptyset$

to cluster unlabelled query set D_Q where D_S is unavailable during training and its samples are individually at test time.

SMILE [120] first introduces On-the-Fly Category Discovery, leveraging hash-based descriptors and a sign-magnitude disentanglement architecture to mitigate intra-category variance. Similarly, PHE [121] proposes a robust framework for fine-grained on-the-fly category discovery using hash-based representations with prototypical encoding, combining Category-aware Prototype Generation and Discriminative Hash Encoding. DiffGRE [142] adopts a Generate–Refine–Encode pipeline: cross-space slerp (Stable Diffusion + CLIP) synthesizes novel samples, a diversity-driven filter removes near-known artifacts, and semi-supervised leader prototypes enable high-dimensional online clustering.

Category Discovery with Domain Shift. Category Discovery with Domain Shift addresses the challenge of discovering novel categories when there is a discrepancy between the labelled source domain and the unlabelled target domain. Formally, the labelled data $\mathcal{D}_L$ is assumed to be exclusively drawn from the base domain Ω_B , and let $\mathcal{D}_U$ include samples from both Ω_B and a novel domain Ω_N . The objective is to accurately classify images from the combined domain $\Omega = \Omega_B \cup \Omega_N$, assuming that the base and novel domains are disjoint (*i.e.*, $\Omega_B \cap \Omega_N = \emptyset$). In practice, the novel domain Ω_N may encompass multiple subdomains.

Yu *et al.* [143] propose a self-labelling framework for unsupervised domain adaptation, assuming unlabelled data contains only novel classes from a single target domain. It aligns source and target features via prototypical contrastive learning and mutual information maximization. Zhuang *et al.* [144] propose the Self-supervised Class-Discovering Adapter (SCDA) for open set domain adaptation. SCDA pre-trains a model with adversarial learning to distinguish known from unknown target domain samples, and then alternates clustering-driven class discovery with self-supervised adaptation to dynamically update its classifier. SAN [145] uses soft contrastive learning to reduce domain-induced view noise and introduces an all-in-one classifier that handles both known and unknown categories. However, SAN assumes all unlabelled samples come from a single novel domain and only requires a general “unknown” label for novel categories, limiting its applicability to more complex settings. Similarly, CDAD-Net [146] adopts an entropy-driven adversarial learning approach to align representations between source and target domains, and refines representations using a conditional image inpainting.

CDAD-Net also employs the elbow method [148] for class number estimation. Wang *et al.* [147] propose a style removal module that disentangles domain-specific style information from domain-invariant content features, improving feature alignment. HiLo [23] handles multi-domain unlabelled data by disentangling semantic and domain-specific features, using PatchMix [149] for embedding-level blending and curriculum learning [150] to gradually address complex domain shifts. Among these methods, [143]–[145], [147] are built upon the NCD setting assuming disjoint label spaces between $\mathcal{Y}_L$ and $\mathcal{Y}_U$, while others [23], [146] are developed under the more general GCD setting.

Distribution-Agnostic Category Discovery (DA-CD). DA-CD tackles the challenge of imbalanced class distributions by developing robust methodologies that do not assume uniformity in labelled or unlabelled data, thereby enabling the effective discovery of novel categories in realistic environments. It acknowledges that the data may follow a skewed distribution, such that for certain categories $\mathcal{Y}_i$ and $\mathcal{Y}_j$ within the set $\mathcal{Y}$, it holds that $\mathbb{P}_{\mathcal{Y}_i}(\mathcal{Y}_i) > \mathbb{P}_{\mathcal{Y}_j}(\mathcal{Y}_j)$. In this formulation, the set $\mathcal{Y}_i$ may refer to either the labelled categories $\mathcal{Y}_L$ or the unlabelled categories $\mathcal{Y}_U$.

NCDLR [151] targets long-tailed imbalances with relaxed optimal-transport self-labeling and equiangular prototypes. A k -means-based approach [16] is used to estimate the number of novel classes. Similarly, BaCon [154] introduces a dual-branch framework comprising a contrastive-learning branch that dynamically estimates the data distribution through clustering to enhance feature representations, alongside a pseudo-labelling branch that produces class predictions for both known and novel categories. To address distribution shifts and class imbalance in biomedical imaging, Fan *et al.* [155] use von Mises–Fisher distributions on a hypersphere to decouple semantics from noise, with geometric constraints for better class separation. They estimate class numbers via spectral analysis of the graph Laplacian. BYOP [153] extends the base NCD setting to handle novel classes with imbalanced distributions by iteratively refining the class prior using model predictions and clustering data based on the estimated distribution, with dynamically adjusted temperature settings to improve prediction confidence in uncertain categories. ImbaGCD [152] introduces an optimal transport–based expectation-maximization framework to iteratively align class priors, reducing head-class bias and improving representation and label assignment under

TABLE VII
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED DA-CD APPROACHES.

Year	Method	Pub.	Backbone	Scenario	Label Assignment	$\mathcal{Y}_L \cap \mathcal{Y}_U$
2023	NCDLR [151]	<i>TMLR</i>	ViT	Long-tailed Distribution for $\mathcal{Y}_L \& \mathcal{Y}_U$	Parametric Classifier	$\emptyset$
	ImbaGCD [152]	<i>CVPRW</i>	ResNet	Imbalanced Distribution for $\mathcal{Y}_U$	Parametric Classifier	$\neq \emptyset$
	GCDLR [22]	<i>ICCVW</i>	ResNet	Imbalanced Distribution for $\mathcal{Y}_U$	Parametric Classifier	$\neq \emptyset$
	BYOP [153]	<i>CVPR</i>	ResNet	Imbalanced Distribution for $\mathcal{Y}_U$	Parametric Classifier	$\emptyset$
	BaCon [154]	<i>NeurIPS</i>	ViT	Long-tailed Distribution for $\mathcal{Y}_L \& \mathcal{Y}_U$	k -means	$\neq \emptyset$
2024	Fan <i>et al.</i> [155]	<i>CVPR</i>	ViT	Long-tailed Distribution for $\mathcal{Y}_L \& \mathcal{Y}_U$	Spectral graph	$\neq \emptyset$
Preprint	Zhao <i>et al.</i> [156]	<i>ArXiv</i>	ViT	Long-tailed Distribution for $\mathcal{Y}_L \& \mathcal{Y}_U$	Density-based Clustering	$\neq \emptyset$

TABLE VIII
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED SCD APPROACHES.

Year	Method	Pub.	Backbone	Word Space	Label Assignment	# Unlabelled categories	$\mathcal{Y}_L \cap \mathcal{Y}_U$
2024	SCD [20]	<i>CVPRW</i>	ViT	$\sim$ Open	k -means+Top- k Voting	Known	$\neq \emptyset$
	SNCD [157]	<i>AAAI</i>	ResNet	$C_{base} + C_{novel}$	Parametric Classifier	Known	$\emptyset$
	FineR [158]	<i>ICLR</i>	ViT	$\sim$ Open	Parametric Classifier	Unknown	$\neq \emptyset$

TABLE IX
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED FCD APPROACHES.

Year	Method	Pub.	Backbone	Label Assignment	$\mathcal{Y}_L \cap \mathcal{Y}_U$
2023	FedoSSL [159]	<i>ICML</i>	ResNet	Parametric Classifier	$\neq \emptyset$
2024	FedGCD [19]	<i>CVPR</i>	ViT	GMM	$\neq \emptyset$
2025	GAL [160]	<i>ICCV</i>	ResNet	Parametric Classifier	$\emptyset$

long-tailed distributions. GCDLR [22] extends ImbaGCD by reweighting rare classes and enforcing class-prior constraints for better distribution alignment, but still assumes a known number of novel categories. Zhao *et al.* [156] propose a dual reliable-sample miner, plus a density-peak-driven estimator for the number of novel classes.

Semantic Category Discovery (SCD). Unlike conventional category discovery methods, which only group visually similar samples, SCD extends GCD by directly discovering semantic class names through open-vocabulary reasoning. Specifically, SCD leverages an open vocabulary space to achieve this goal. In this context—such as an open-vocabulary setting—WordNet [161], which comprises approximately 68,000 labels, is used as a comprehensive and unconstrained lexical resource to facilitate the assignment of semantic labels.

Wang *et al.* [157] introduce Semantic-guided Novel Category Discovery, which supplements NCD by mapping semantic labels to unlabelled data instead of using discrete one-hot labels. This is achieved by combining a semantic similarity matrix with dynamic visual prototypes, enabling the joint optimization of classification and clustering via mutual information maximization. In a more challenging setting, Han *et al.* [20] introduce an SCD framework where the label space comprises an open vocabulary. By leveraging VLMs such as CLIP [117], their approach aligns visual and textual data to facilitate more effective label assignment from an expansive vocabulary. A voting mechanism within clusters then identifies the most probable semantic labels, and an iterative refinement process continuously enhances label accuracy as the model evolves. FineR [158] proposes a training-free framework for fine-grained visual recognition, aligning with the definition used in SCD. It utilises a VQA model (*e.g.*, BLIP-2 [162]) to extract visual attributes, an LLM (*e.g.*, GPT [118]) to generate candidate labels, and a VLM to build a multimodal classifier for zero-shot recognition.

Few-Shot Category Discovery (FS-CD). FS-CD tackles the challenge of identifying novel categories with limited labelled data by leveraging prior knowledge from known classes and maximizing the value of scarce annotations. It extends conven-

tional category discovery by incorporating few-shot learning principles. In particular, FS-CD adopts an N -way, k -shot framework requiring category discovery with only k labelled examples per base class.

Chi *et al.* [21] extend NCD to a few-shot setting based on the assumption that both base and novel categories share high-level semantic features. Their approach adapts meta-learning techniques - Model-Agnostic Meta-Learning (MAML) [163] and Prototypical Networks (ProtoNet) [164], shifting the focus from classification to clustering tasks—crucial for few-shot category discovery. A key innovation of their method is the clustering-rule-aware task sampler, which ensures consistent clustering rules during training, thereby enhancing the model’s ability to generalize to novel categories despite limited labelled data. However, this method assumes prior knowledge of the number of novel categories.

Federated Category Discovery (FCD). FCD addresses the challenge of collaboratively discovering novel categories from decentralized, privacy-preserving datasets with heterogeneous label distributions, requiring methods that integrate diverse local knowledge while preserving data privacy. It considers N^L local clients, each equipped with its own private labelled dataset $\mathcal{D}_n^L = \{(\mathbf{x}_i, y_i)\} \subset \mathcal{X}_n^L \times \mathcal{Y}_n^L$ and an unlabelled dataset $\mathcal{D}_n^U = \{(\mathbf{x}_i, \hat{y}_i)\}_{i=1}^K \subset \mathcal{X}_n^U \times \mathcal{Y}_n^U$, where the local label space is defined as $\mathcal{Y}_n = \mathcal{Y}_n^L \cup \mathcal{Y}_n^U$. Notably, the label spaces across different clients are not assumed to be identical (*i.e.*, for any two clients i and j , $\mathcal{Y}_i$ may differ from $\mathcal{Y}_j$), thereby reflecting heterogeneous data distributions. The primary objective of FCD is to collaboratively train a generalized global model that aggregates knowledge from all local clients, without exposing any private data.

FedoSSL [159] extends GCD to decentralized learning. It introduces an uncertainty-aware regulation loss to balance learning rates between locally and globally unseen categories, and employs a calibration module for consistent model aggregation. However, FedoSSL assumes prior knowledge of the total number of categories across clients, limiting its use in settings that require category estimation. Fed-GCD [19] introduces associated Gaussian contrastive learning using a

TABLE X
SUMMARY OF ESSENTIAL COMPONENTS FOR REVIEWED CD BEYOND IMAGES.

Task	Year	Method	Pub.	Backbone	Label Assignment
Seg. & Dect.	2022	Zhao <i>et al.</i> [165]	CVPR	ResNet	k -means
		Zhong <i>et al.</i> [166]	CVPRW	ResNet	k -means
	2024	DRM [167]	AAAI	Faster R-CNN	k -means
		PANDAS [168]	CoLLAs	Faster R-CNN	k -means
Videos	2025	Hoang <i>et al.</i> [169]	AAAI	ResNet+SAM	Parametric Classifier
3D.	2023	NEV [170]	ICIP	SlowFast, R2+1D, P3D	Parametric Classifier
	2023	NOPS [171]	CVPR	MinkowskiUNet	k -means
		DNIK [172]	NeurIPS	PointNet	Parametric Classifier
Text	2024	Xu <i>et al.</i> [173]	ECCV	MinkowskiUNet	Parametric Classifier
	2022	An <i>et al.</i> [174]	EMNLP	BERT	k -means
		DPN [175]	AAAI	BERT	k -means
	2023	KNoRD [176]	EMNLP	DeBERTa	GMM
		TAN [177]	AAAI	BERT	k -means
	2024	KTN [178]	AAAI	BERT	GMM
	2025	SDC [179]	AAAI	BERT	Parametric Classifier
Tabular	preprint	SFGRD [180]	ArXiv	BERT	k -means
	2022	TabularNCD [181]	ICKG	Encoder (2 dense hidden layers)	Parametric Classifier
	2023	PBT [182]	DMKN	PBN	k -means
Graph		Colin <i>et al.</i> [183]	ECML PKDD	N/A	k -means
Multi-Modal	2024	SWORD [184]	CIKM	GCN, GAT, GraphSAGE	Parametric Classifier
	2021	Jia <i>et al.</i> [60]	ICCV	ResNet, R3D-18	Parametric Classifier
	2024	Zhou <i>et al.</i> [185]	AAAI	ResNet&BioClinicalBERT	Parametric Classifier
	preprint	Su <i>et al.</i> [186]	AAAI	ViT	Parametric Classifier

learnable GMM. A client semantics association mechanism aggregates local GMMs into a global GMM on the central server, facilitating shared semantic knowledge discovery. It also integrates semi-supervised FINCH [124] to estimate the number of novel categories per client. Wang *et al.* [160] present the GAL framework for FCD in a continual setting. GAL combines modified prototype contrastive learning for robust representation learning, prototype-based filtering for accurate label assignment, and the *potential prototype merge* algorithm to estimate the number of novel categories.

C. Category Discovery on other modalities

Category Discovery in Scene Understandings. Zhao *et al.* [165] extend category discovery to semantic segmentation. They employ a pretrained BASNet [187] saliency model alongside a base segmentation model trained on labelled base categories to isolate novel regions. Pseudo-labels are assigned via k -means clustering. To mitigate clustering noise, they introduce entropy-based uncertainty modeling and self-training, refining pseudo-labels through entropy ranking and dynamic data reassignment.

Zhong *et al.* [166] propose Open-Set Object Detection and Discovery (OSODD), which jointly detects known objects and discovers novel categories. Their two-stage framework first employs an open-set detector to identify both known and unknown objects, followed by contrastive learning and semi-supervised clustering to group unknown objects into novel classes. DRM [167] tackles the bias toward known categories in NCD for object detection. It employs a dual Region Proposal Network (RPN) strategy that combines class-aware and class-agnostic RPNs, enhancing feature representations with semi-supervised instance-level contrastive learning. PANDAS [168] introduces a prototype-based approach for novel class discovery and detection on top of Faster R-CNN. At inference, objects are classified by comparing feature vectors to both base and novel prototypes via a similarity metric, enabling detection without prior knowledge of the number of novel classes. Hoang *et al.* [169] propose a GCD method for

instance segmentation that tackles novel category discovery and class imbalance. The approach employs instance-wise temperature assignment for contrastive learning, soft attention modules to enhance object-specific features, and reliability-based dynamic learning to iteratively refine pseudo-labels with class-wise thresholds.

Category Discovery for Videos. NEV [170] tackles NCD in video action recognition by employing negative learning, entropy regularization, and variance regularization. Additionally, it incorporates view-invariant learning through adversarial training to filter viewpoint-specific information and contrastive learning to align representations of the same action from different viewpoints.

Category Discovery for 3D Objects. NOPS [171] extends category discovery to 3D point cloud under the NCD setting. It leverages online clustering and uncertainty quantification to generate pseudo-labels for novel categories, and further improves feature space partitioning through a class-balanced queuing strategy and a multi-headed segmentation technique. Similarly, Xu *et al.* [173] propose a dual-level adaptive self-labelling framework for 3D point cloud segmentation. It addresses class imbalance and spatial context by leveraging point-level and region-level representations. A key contribution is a dynamic semi-relaxed optimal transport algorithm that adjusts regularization to handle imbalanced novel classes. DNIK [172] presents a different approach by utilizing geometric part-based features from known classes. It constructs a part concept bank to represent 3D shapes and uses a *part relation encoder* to capture spatial relationships. Notably, NOPS [171] and DNIK [172] assume the number of novel categories is known in advance, whereas Xu *et al.* [173] employ an off-the-shelf k -means-based approach for category estimation.

Category Discovery for Text. DPN [175] introduces a *de-coupled prototypical network* that separates the learning of known and novel categories. It employs bipartite matching for prototype alignment between labelled and unlabelled data, and *semantic-aware prototypical learning* to enhance feature dis-

crimination and reduce pseudo-label noise. An *et al.* [174] propose Fine-grained Category Discovery under Coarse-grained supervision (FCDC), using a hierarchical model based on pre-trained BERT [188] to learn coarse and fine-grained features, aided by a self-contrastive module to balance class distances. TAN [177] presents a prototype-based framework in the GCD setting, incorporating knowledge transfer mechanisms and feature alignment strategies to address noisy unlabelled data. It utilises a pretrained BERT [188] model and a filtering algorithm for novel category estimation. KTN [178] applies an entropy-based technique to identify novel category samples, combining prototype-based semantic similarities and logit adjustments for improved pseudo-label accuracy. SDC [179] enhances knowledge transfer using a dual logit adjustment strategy to reduce bias and facilitate transfer between similar known and novel categories. KNoRD [176] extends category discovery to open relation extraction in long-tailed NCD, combining prompt-based training and GMM clustering for semantically aligned relation representations. SFGRD [180] proposes a two-stage framework for open-world relation extraction, reducing entity and context biases while enhancing relational semantics through a cluster-semantic space and iterative novel relation estimation via clustering.

Category Discovery for Tabular Data. TabularNCD [181] extends category discovery to tabular data under the NCD setting, addressing challenges from its heterogeneous structure. It employs self-supervised learning to pre-train latent representations on both labelled and unlabelled data, jointly optimizing a classification task on known classes and a clustering task to discover novel categories. Colin *et al.* [183] develop an interactive interface for category discovery in tabular data, enabling subject-matter experts to analyse results without machine learning expertise. The system integrates NCD methods, such as TabularNCD [181], with traditional clustering algorithms involving k -means and spectral clustering. PBT [182] enhances unsupervised clustering methods such as k -means++ and Spectral Clustering—by incorporating labelled data from known classes to guide the clustering process. It also leverages *cluster validity indices* computed in the latent feature space to estimate the number of novel classes.

Category Discovery for Graph. SWORD [184] introduces category discovery for graph classification. It employs self-training to learn from unlabelled nodes, uses prototype replay to preserve knowledge of existing classes, and incorporates knowledge distillation to mitigate catastrophic forgetting.

Category Discovery Cross Modalities. Jia *et al.* [60] introduce a unified framework for multimodal NCD by combining instance and category contrastive learning with cross-modal contrastive learning. Zhou *et al.* [185] extend UNO [40] to a multimodal framework for chest X-rays. They first apply multimodal contrastive learning to fine-tune image and text encoders, then fuse pseudo-labels from both modalities to improve novel class discovery. Furthermore, Su *et al.* [186] extend GCD to multimodal datasets by employing contrastive learning to align feature spaces across modalities and cross-modal distillation to ensure consistent classification, while incorporating a prototypical classifier for effective category partitioning.

TABLE XI
DETAILED DATASET SPLITTING STATISTICS FOR GCD. FG INDICATES WHETHER THE DATASET IS FINE-GRAINED.

Dataset	FG	Labelled		Unlabelled	
		#Image	#Class	#Image	#Class
ImageNet-100 [189]	✗	31.9K	50	95.3K	100
CIFAR-10 [190]	✗	12.5K	5	37.5K	10
CIFAR-100 [190]	✗	20.0K	80	30.0K	100
CUB [191]	✓	1.5K	100	4.5K	200
Stanford-Cars [192]	✓	2.0K	98	6.1K	196
FGVC-Aircraft [193]	✓	1.7K	50	5.0K	100

IV. BENCHMARK

In this section, we review commonly used datasets and their construction methods, outline the evaluation protocols, and consolidate the performance of the reviewed category-discovery methods on the two base settings—*NCD* and *GCD*. For each task, we select the most commonly used datasets for benchmarking performance. Performance scores are sourced from the original articles unless otherwise specified. More in-depth descriptions of the other datasets are included in the Appendix-V.

A. Dataset

In category discovery, a diverse range of datasets is used, spanning *generic* and *fine-grained* categories across various domains, including natural images, plant species, and biomedical fields. In this subsection, we introduce six widely utilised datasets, comprising three generic and three fine-grained datasets, along with the dataset-splitting methods commonly applied in the field. The statistics for the commonly used datasets are summarized in Tab XI.

The most widely used *generic* datasets: **ImageNet-1K&100** [189] is a widely used dataset for natural image classification. ImageNet-1K contains over one million labelled images across 1,000 object categories. ImageNet-100 is a subset with 100 randomly selected classes. **CIFAR-10 & CIFAR-100** [190] are both natural images sized in 32×32 . CIFAR-10 contains 50,000 images spanning across 10 different classes and CIFAR-100 includes 100 classes, with each class containing 500 images.

The most widely used *fine-grained* benchmark is **SSB** [194], which includes three datasets: CUB [191], Stanford-Cars [192], and FGVC-Aircraft [193]. **CUB** [191] is a widely used fine-grained dataset, particularly focused on bird species recognition. **Stanford-Cars** [192] is a large-scale dataset designed for fine-grained vehicle classification tasks. **FGVC-Aircraft** [193] is a fine-grained visual classification dataset focused on aircraft recognition.

Other Datasets: **Caltech-101** [195] is a collection of images across 101 object categories, and it is re-structured for CCD. **TinyImageNet** [196] is a reduced version of ImageNet with 200 classes, each having 500 images, all resized to 64×64 pixels for computational efficiency.

Dataset Splitting: In category discovery, categories are typically divided into base and novel categories. Given the total number of categories as $|C|$, the number of base categories and novel categories are denoted as $|C_B|$ and $|C_N|$ respectively with the constraint $|C_B| + |C_N| = |C|$. The common splitting ratio of base and novel categories is $|C_B| : |C_N| = 5 : 5$, except

TABLE XII
COMPARISON WITH STATE-OF-THE-ART NCD METHODS. RESULTS ARE REPORTED IN ACC USING TASK-AWARE EVALUATION PROTOCOL.

	Method	CUB	Stanford-Cars	FGVC-Aircraft	CIFAR-10	CIFAR100-20	CIFAR100-50	ImageNet-100	ImageNet-1K
2020	KCL [34]	-	-	-	72.3±0.2	42.1±1.8	-	-	73.8
	MCL [35]	-	-	-	70.9±0.1	21.5±2.3	-	-	74.4
	DTC [15]	-	-	-	88.7±0.3	67.3±1.2	35.9±1.0	-	78.3
	RS [36]	-	-	-	90.4±0.5	73.2±2.1	39.2±2.3	-	82.5
	RS+ [36]	55.3±0.8 [†]	36.5±0.6 [†]	38.4±0.6 [†]	91.7±0.9	75.2±4.2	44.1±3.7	-	82.5
2021	Qing <i>et al.</i> [37]	-	-	-	93.6±0.6	81.3±1.9	-	-	-
	OpenMix [38]	-	-	-	95.3	87.2	-	74.76	85.7
	NCL [39]	48.1±0.9 [†]	43.5±1.2 [†]	43.0±0.5 [†]	93.4±0.5	86.6±0.4	52.7±1.2	-	90.7
	JOINT [60]	-	-	-	93.4±0.6	76.4±2.8	-	-	86.7
	DualRS [41]	-	-	-	91.6±0.6	75.3±2.3	-	-	88.9
	UNO [40]	59.2±0.4 [†]	49.8±1.4 [†]	52.1±0.7 [†]	93.3±0.4	90.5±0.7	62.3±1.4	79.56	90.7
2022	ResTune [45]	-	-	-	89.0±0.6	63.7±1.0	-	-	-
	IIC [47]	71.3±0.6	55.2±0.7	56.0±0.8	99.1±0.0	92.4±0.2	65.8±0.9	80.24	91.9
	CRKD [49]	65.7±0.6	53.5±0.8	55.8±0.9	93.5±0.3	91.2±0.1	65.3±0.6	80.94	-
	NSCL [48]	-	-	-	97.5	85.9	64.0	-	-
	APL [53]	-	-	-	96.0±0.4	78.9±0.9	-	-	88.8
	SCKD [52]	73.1±0.4	56.8±0.7	56.5±0.7	95.6±0.2	92.6±0.6	68.2±0.4	82.18	-

[†] denotes results from [49].

for CIFAR-100, which typically uses a $|\mathcal{C}_B| : |\mathcal{C}_N| = 8 : 2$ ratio. In GCD, since the unlabelled data contains both base and novel categories, it is necessary to further partition the base categories into labelled and unlabelled subsets. The typical ratio for this split is 5 : 5, although a 1 : 9 ratio is sometimes used for more challenging cases [26], [66]. In Tab. XI, we present detailed statistics of commonly used datasets for GCD.

B. Evaluation Protocol

To evaluate the performance of category discovery, the predominant metric employed is the *Clustering Accuracy* on unlabelled dataset $\mathcal{D}_U$. This metric is formulated as

$$ACC = \max_{p \in \mathcal{P}(\mathcal{Y}_U)} \frac{1}{M} \sum_{i=1}^M \mathbb{1}\{y_i = p(\hat{y}_i)\},$$

where $M = |\mathcal{D}_U|$ denotes the total number of samples in $\mathcal{D}_U$, and $\mathcal{P}(\mathcal{Y}_U)$ represents the set of all possible permutations of the class labels. In practice, the Hungarian algorithm [197] is employed to assign the optimal label permutations.

Within the realm of category discovery, three evaluation protocols are adopted to compute clustering accuracy:

- 1) *Task-aware Protocol*: This protocol applies the Hungarian algorithm separately to the novel category set $\mathcal{C}_N$ and the base category set $\mathcal{C}_B$. Originally derived from NCD settings—where the unlabelled data consist solely of novel categories (*i.e.*, $\mathcal{C}_N = \mathcal{Y}_U$)—this approach has been extended in GCD [16] to compute optimal label assignments independently for both sets.
- 2) *Task-agnostic Protocol*: This protocol computes the optimal label assignment over the entire label space $\mathcal{Y}_U$, which includes both novel and base categories, which is the most commonly adopted protocol for GCD.

It is important to note that the majority of category discovery methods operate in a transductive setting, where the evaluation is conducted on the same unlabelled dataset used in training. Nevertheless, some approaches [88] also report performance under an inductive setting using an unseen test set.

Specific Evaluation Protocol for CCD. In the continual learning setting, GM [17] introduces two evaluation metrics, $\mathcal{M}_f$ and $\mathcal{M}_d$, to measure the degree of forgetting in base categories and the ability to recognize novel categories. Specifically, $\mathcal{M}_f = \max_{t \in 1, \dots, T} |ACC_B^0 - ACC_B^t|$ represents the maximum difference in clustering accuracy for base categories

over time, while $\mathcal{M}_d = ACC_N^T$ denotes the final clustering accuracy on novel categories, where ACC is evaluated under the task-agnostic protocol. Additionally, PromptCCD [134] introduces *continual ACC* ($cACC$), which differs from previous metrics by considering not only the initial labelled data but also incorporating unlabelled data from previous timesteps along with their assigned labels. This allows $cACC$ to evaluate performance continuously across all timesteps, rather than focusing solely on the current timestep.

C. NCD Performance Benchmarking

We select three fine-grained datasets: CUB [191], Stanford-Cars [192], and FGVC-Aircraft [193]. For these datasets, all results are generated using a DINO [62] pre-trained ViT as the backbone. Regarding the generic datasets, we present results on CIFAR-10 [190], CIFAR100-20 [190], CIFAR100-50 [190], ImageNet-100 [189], and ImageNet-1K [189] for benchmarking the NCD algorithms. Most methods utilise ResNet-18 [1] as the backbone for these datasets, with the exception of KCL [34], DTC [15] and OpenMix [38], which implement a six-layer VGG-like architecture [198] specifically on CIFAR100-20. Tab. XII displays the performance of 17 state-of-the-art NCD methods. The performance is assessed by the average ACC across three runs on the ‘New’ categories utilizing the task-aware evaluation protocol introduced in Sec. IV-B. For most results, we also report the variance of ACC as shown in the included papers.

D. GCD Performance Benchmarking

For the GCD task, we select the most widely-used fine-grained datasets, including CUB [191], Stanford-Cars [192] and FGVC-Aircraft [193] as well as generic datasets including CIFAR-10 [190], CIFAR-100 [190] and ImageNet-100 [189] for benchmarking the GCD algorithms. Tab. XIII shows the performance of 35 state-of-the-art GCD methods on the six datasets based on DINO [62] pre-trained ViT-B [200] backbone. The performance is measured by ACC across the ‘All’, ‘Old’ and ‘New’ categories, as introduced in Sec. IV-B. Additionally, we report the results of selected GCD methods on the same datasets using ViT-B backbones pre-trained with DINOv2 [199] and CLIP [117] on Tab. XIV.

Notably, several methods [26], [65], [66], [69], [72], [76], [79], [87], [92], [93] employ different data-splitting protocols

TABLE XIII
COMPARISON OF STATE-OF-THE-ART GCD METHODS. RESULTS ARE REPORTED IN ACC ACROSS ‘ALL’, ‘OLD’ AND ‘NEW’ CATEGORIES.

		CUB			Stanford-Cars			FGVC-Aircraft			CIFAR-10			CIFAR-100			ImageNet-100		
Method		All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
2022	RankStats+ [61]	33.3	51.6	24.2	28.3	61.8	12.1	26.9	36.4	22.2	46.8	19.2	60.5	58.2	77.6	19.3	37.1	61.6	24.8
	UNO+ [40]	35.1	49.0	28.1	35.5	70.5	18.6	40.3	56.4	32.2	68.6	98.3	53.8	69.5	80.6	47.2	70.3	95.0	57.9
	XCon [68]	52.1	54.3	51.0	40.5	58.8	31.7	47.7	44.4	49.4	96.0	97.3	95.4	74.2	81.2	60.3	77.6	93.5	69.7
	GCD [16]	51.3	56.6	48.7	39.0	57.6	29.9	45.0	41.1	46.9	91.5	97.9	88.2	73.0	76.2	66.5	74.1	89.8	66.3
2023	SimGCD [73]	60.3	65.6	57.7	53.8	71.9	45.0	54.2	59.1	51.8	97.1	95.1	98.1	80.1	81.2	77.8	83.0	93.1	77.9
	PromptCAL [70]	62.9	64.4	62.1	50.2	70.1	40.6	52.2	52.2	52.3	97.9	96.6	98.5	81.2	84.2	75.3	83.1	92.7	78.3
	DCCL [71]	63.5	60.8	64.9	43.1	55.7	36.2	42.9 [†]	44.5 [†]	42.1 [†]	96.3	96.5	96.9	75.3	76.8	70.2	80.5	90.5	76.2
	GPC [74]	52.0	55.5	47.5	38.2	58.9	27.4	43.3	40.7	44.8	90.6	97.6	87.0	75.4	84.6	60.1	75.3	93.4	66.7
	PIM [75]	62.7	75.7	56.2	43.1	66.9	31.6	48.8 [†]	60.0 [†]	43.0 [†]	94.7	97.4	93.3	78.3	84.2	66.5	83.1	<u>95.3</u>	77.0
	μ GCD [‡] [77]	65.7	68.0	64.6	56.5	68.1	50.9	53.8	55.4	53.0	-	-	-	-	-	-	-	-	-
	InfoSieve [78]	69.4	77.9	65.2	55.7	74.8	46.4	56.3	63.7	52.5	94.8	97.7	93.4	78.3	82.2	70.5	80.5	93.8	73.8
	CAC [‡] [80]	58.0	65.0	43.9	47.6	70.6	33.8	-	-	-	92.3	91.4	94.4	78.5	81.4	75.6	81.1	80.3	81.8
≥ 2024	CIPR [81]	57.1	58.7	55.6	47.0	61.5	40.1	44.0 [†]	50.5 [†]	40.8 [†]	97.7	97.5	97.7	81.5	82.4	79.7	80.5	84.9	78.3
	SPTNet [84]	65.8	68.8	65.1	59.0	79.2	49.3	59.3	61.8	58.1	97.3	95.0	98.6	81.3	84.3	75.6	85.4	93.2	81.4
	LegoGCD [85]	63.8	71.9	59.8	57.3	75.7	48.4	55.0	61.5	51.7	97.1	94.3	98.5	81.8	81.4	82.5	86.3	94.5	82.1
	CMS [86]	68.2	76.5	64.0	56.9	76.1	47.6	56.0	63.4	52.3	95.7 [†]	97.0 [†]	94.4 [†]	82.3	<u>85.7</u>	75.5	84.7	95.6	79.2
	Yang <i>et al.</i> [‡] [90]	61.3	60.8	62.1	44.3	58.2	39.1	-	-	-	96.5	97.6	94.4	74.6	76.5	69.4	78.1	91.3	70.5
	AMEND [82]	64.9	75.6	59.6	52.8	61.8	48.3	56.4	73.3	48.2	96.8	94.6	97.8	81.0	79.9	<u>83.3</u>	83.2	92.9	78.3
	GCA [83]	68.8	73.4	66.6	54.4	72.1	45.8	52.0	57.1	49.5	95.5	95.9	95.2	82.4	85.6	75.9	82.8	94.1	77.1
	SelEx [91]	73.6	75.3	72.8	58.5	75.6	50.3	57.1	64.7	53.3	95.9	98.1	94.8	82.3	85.3	76.3	83.1	93.6	77.8
	MSGCD [96]	63.6	70.7	60.0	57.7	75.5	49.9	56.4	64.1	52.6	97.1 [†]	97.4 [†]	97.0 [†]	81.4	82.5	79.0	82.1 [†]	88.9 [†]	78.7 [†]
	FlipClass [94]	71.3	71.3	71.3	63.1	<u>81.7</u>	53.8	59.3	66.9	55.4	98.5	97.6	99.0	85.2	84.9	85.8	86.7	94.3	82.9
	OwMatch [93]	70.5 [†]	81.5[†]	65.0 [†]	58.0 [†]	75.5 [†]	49.3 [†]	52.6 [†]	63.2 [†]	47.3 [†]	96.6 [†]	<u>98.0[†]</u>	95.9 [†]	79.8 [†]	81.2 [†]	79.1 [†]	84.7 [†]	95.6[†]	79.2 [†]
	DebGCD [99]	66.3	71.8	63.5	65.3	81.6	57.4	<u>61.7</u>	63.9	60.6	97.2	94.8	98.4	83.0	84.6	79.9	85.9	94.3	81.6
	Contextuality-GCD [95]	64.5	68.7	61.5	55.8	73.1	47.8	56.8	61.8	54.4	97.3	96.1	<u>98.8</u>	82.0	83.9	78.9	84.6	94.0	81.0
	MOS [101]	69.6	72.3	68.2	64.6	80.9	56.7	61.1	66.9	58.2	-	-	-	-	-	-	-	-	-
	AptGCD [103]	70.3	74.3	69.2	62.1	79.7	53.6	61.1	65.2	59.0	97.3	95.8	98.7	82.8	81.8	85.5	87.8	95.4	84.3
	Dai <i>et al.</i> [104]	68.5	73.1	66.2	62.3	80.7	53.4	60.9	63.5	59.6	-	-	-	-	-	-	-	-	-
	HypCD [105]	79.8	75.8	81.8	62.9	80.0	54.7	65.9	<u>67.3</u>	65.1	96.7	97.6	96.3	82.4	85.1	77.0	<u>86.8</u>	94.6	<u>82.8</u>
	SEAL [108]	66.2	72.1	63.2	65.3	79.3	58.5	62.0	65.3	60.4	97.2	94.7	98.4	82.1	81.7	83.0	84.6	90.9	81.3
	PAL-GCD [‡] [98]	72.6	75.2	71.2	72.2	83.4	66.8	59.5	67.2	55.6	-	-	-	82.0	82.4	81.2	86.3	93.0	82.0
	ConGCD [106]	81.7	<u>80.4</u>	82.4	57.5	77.5	47.9	62.5	70.2	58.7	97.4	95.2	98.5	82.5	85.9	77.3	85.9	93.4	82.5
	AF [107]	69.0	74.3	66.3	67.0	80.7	60.4	59.4	68.1	55.0	<u>97.8</u>	95.9	<u>98.8</u>	82.2	85.0	76.5	85.4	94.6	80.8
	PNP [‡] [111]	67.4	69.2	66.5	57.3	70.2	51.1	47.0	49.5	45.7	90.2	95.0	<u>87.8</u>	80.0	81.7	76.3	82.4	92.9	77.2
	RPIM [‡] [112]	66.8	77.3	61.5	45.8	67.5	35.3	-	-	-	95.3	97.6	94.1	82.7	85.4	77.2	83.2	95.2	77.2
	ConceptGCD [‡] [114]	69.4	75.4	66.5	<u>70.1</u>	81.6	<u>64.6</u>	60.5	59.2	<u>61.1</u>	-	-	-	82.8	84.1	80.1	86.3	93.3	<u>82.8</u>
	MCDL [‡] [110]	68.7	74.2	63.4	61.9	77.9	54.7	57.6	64.6	52.5	97.3	95.4	98.4	<u>84.0</u>	84.7	81.2	83.8	93.8	78.8
	NN-GCD [‡] [116]	72.4	76.3	70.5	65.4	77.5	59.5	60.5	65.3	58.2	97.3	95.0	98.6	81.2	82.4	78.8	85.2	91.5	82.2

[†] denotes results from our implementation. [‡] denotes methods without releasing official codes.

or different backbones, resulting in inconsistent labelled and unlabelled sets and ultimately hindering fair comparisons. Therefore, we reproduce OwMatch [93] as a representative method, with selection details in the Appendix.

E. Benchmarking Takeaways

Based on the above benchmarking and analysis, we summarize the following takeaways:

(i) **Part-level information plays a critical role in category discovery, particularly in fine-grained settings where global features struggle to distinguish visually similar classes.** GCD [16] shows that part-level cues naturally emerge in ViT attention heads and become more semantically specialized after contrastive fine-tuning, with certain heads attending to concrete parts such as windshields, headlights, or wheelhouses in a car. In light of this, GCD methods like [84] and [104] explicitly model localized object parts or discriminative regions, demonstrating stronger cues for separating subtle inter-class differences. Meanwhile, DualRS [41] has also utilised the local parts for knowledge transfer in the NCD context. As shown in Tab. III, SPTNet [84] augments SimGCD [73] with visual prompts to enhance patch-level discriminative cues, lifting *All* accuracy from 60.3 to 65.8, 53.8 to 59.0, and 54.2 to 59.3 for fine-grained datasets respectively; building on this, Dai *et al.* [104] employ adaptive part learning and further improve to 68.5, 62.3 and 60.9.

(ii) **Large-scale self-supervised pretrained vision backbones provide a strong advantage for category discovery.** Early methods often train from scratch using self-supervised strategies such as RotNet [201] and MoCo [202]. DINO [62] brings a clear improvement over previous methods, and DINOv2 [199] further surpasses DINO by over 10% on our benchmarks (Tab. II and Tab. III), underscoring the critical role of high-quality visual representations. These results highlight that stronger pretrained models not only improve feature quality but also enhance the robustness and scalability of GCD methods.

(iii) **Hierarchical information has been shown to be an important inductive bias for category discovery.** SelEx [91] employs hierarchical classification, yielding a 17.9% gain over the GCD baseline [16], and SEAL [108] explicitly models hierarchies via coarse-to-fine labels, boosting the SimGCD baseline [73] by 8.4% on fine-grained datasets (see Tab. XIII). HypCD [105] encodes hierarchy implicitly through hyperbolic embeddings, and further increases over SelEx by 6.5% as shown on Tab. XIII, underscoring the value of incorporating hierarchical information into the discovery process.

(vi) **Auxiliary information beyond raw image features also greatly benefits category discovery.** Methods like MOS [101], GET [102], and TextGCD [89] report measurable gains (see Tab. XIV), and collectively demonstrate that

TABLE XIV
ADDITIONAL COMPARISON OF STATE-OF-THE-ART GCD METHODS USING ViT BACKBONE PRETRAINED BY DINOv2 [199] AND CLIP [117].

		CUB			Stanford-Cars			FGVC-Aircraft			CIFAR-10			CIFAR-100			ImageNet-100		
Method		All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
<i>DINOv2</i>	μ GCD [‡] [77]	74.0	75.9	73.1	76.1	91.0	68.9	66.3	68.7	65.1	-	-	-	-	-	-	-	-	-
	CiPR [81]	78.3	73.4	80.8	66.7	77.0	61.8	59.1 [†]	56.9 [†]	60.2 [†]	99.0	98.7	99.2	90.3	89.0	93.1	88.2	87.6	88.5
	SPTNet [84]	76.3	79.5	74.6	72.3 [†]	82.0 [†]	67.5 [†]	68.0 [†]	75.2 [†]	60.8 [†]	<u>98.9</u> [†]	99.1 [†]	98.8 [†]	89.0 [†]	91.5 [†]	79.2 [†]	90.1	96.1	87.1
	GCD [16]	71.9 [△]	71.2 [△]	72.3 [△]	65.7 [△]	67.8 [△]	64.7 [△]	55.4 [△]	47.9 [△]	59.2 [△]	97.8 [°]	<u>99.0</u> [°]	97.1 [°]	79.6 [°]	84.5 [°]	69.9 [°]	78.5 [°]	89.5 [°]	73.0 [°]
	SimGCD [73]	71.5 [△]	78.1 [△]	68.3 [△]	71.5 [△]	81.9 [△]	66.6 [△]	63.9 [△]	69.9 [△]	60.9 [△]	98.7 [°]	<u>96.7</u> [°]	99.7 [°]	88.5 [°]	89.2 [°]	87.2 [°]	89.9 [°]	95.5 [°]	87.1 [°]
	Selex [91]	87.4	<u>85.1</u>	<u>88.5</u>	<u>82.2</u>	93.7	<u>76.7</u>	79.8	82.3	<u>78.6</u>	98.5 [†]	98.8 [†]	98.5 [†]	87.7 [†]	90.8 [†]	81.5 [†]	90.9 [†]	96.2 [†]	88.3 [†]
	DebGCD [99]	77.5	80.8	75.8	75.4	87.7	69.5	71.9	76.0	69.8	<u>98.9</u>	97.5	<u>99.6</u>	<u>90.1</u>	<u>90.9</u>	88.6	93.2	97.0	91.2
	Dai <i>et al.</i> [104]	75.1	79.1	73.2	73.4	87.6	66.7	68.8	74.1	66.6	-	-	-	-	-	-	-	-	-
	HypCD [105]	90.7	85.3	93.4	83.8	93.3	79.2	83.4	82.0	84.1	98.6	98.1	98.9	88.6	91.5	82.8	<u>92.3</u>	<u>96.4</u>	90.2
	SEAL [108]	76.7	78.3	75.9	77.7	88.7	72.4	74.6	73.2	75.3	98.9	98.1	99.3	89.8	90.4	89.5	91.3	93.3	90.3
<i>CLIP</i>	GCD [16]	57.6 [°]	65.2 [°]	53.8 [°]	65.1 [°]	75.9 [°]	59.8 [°]	45.3 [°]	44.4 [°]	45.8 [°]	94.0 [°]	97.3 [°]	92.3 [°]	74.8 [°]	79.8 [°]	64.6 [°]	75.8 [°]	87.3 [°]	70.0 [°]
	SimGCD [73]	71.7 [°]	76.5 [°]	69.4 [°]	70.0 [°]	83.4 [°]	63.5 [°]	<u>54.3</u> [°]	<u>58.4</u> [°]	<u>52.2</u> [°]	97.0 [°]	94.2 [°]	98.4 [°]	81.1 [°]	85.0 [°]	73.3 [°]	<u>90.8</u> [°]	<u>95.5</u> [°]	<u>88.5</u> [°]
	TextGCD [89]	<u>76.6</u>	80.6	<u>74.7</u>	86.9	87.4	86.7	-	-	-	98.2	98.0	98.6	85.7	86.3	<u>84.6</u>	88.0	92.4	85.2
	GET [‡] [97]	77.0	<u>78.1</u>	76.4	<u>78.5</u>	86.8	<u>74.5</u>	58.9	59.6	58.5	97.2	94.6	98.5	82.1	<u>85.5</u>	75.5	91.7	95.7	89.7
	CLIP-GCD [‡] [109]	62.8	77.1	55.7	70.6	88.2	62.2	50.0	56.6	46.5	96.6	97.2	96.4	<u>85.2</u>	<u>85.0</u>	85.6	84.0	<u>95.5</u>	78.2
	CPT [‡] [102]	70.1	73.5	68.4	74.2	84.3	69.3	-	-	-	96.3	96.2	96.4	81.3	81.3	81.2	89.2	95.2	86.1

[°], [△], and [◇] denote results from [77], [102] and [81] respectively. [‡] denotes methods without releasing official codes. [†] denotes results implemented by us.

incorporating additional signals, whether object-level or textual information, provides valuable semantic grounding that complements visual learning.

V. DISCUSSION AND FORWARD LOOKING

Label Assignment: Parametric vs. Non-Parametric. SimGCD [73] introduces a parametric classifier for GCD, demonstrating strong performance. However, subsequent works such as Selex [91] and CMS [86] have shown consistent improvements with non-parametric clustering-based approaches, highlighting their flexibility in capturing complex category structures. These contrasting findings suggest that there is likely no universally optimal label assignment strategy for CD. Instead, the effectiveness of label assignment appears to be highly dependent on the representation learning paradigm and even the dataset characteristics (*e.g.*, class granularity, distribution skew). Thus, it is worth studying adaptive or hybrid assignment strategies that better align with the learned representations and dataset characteristics, rather than relying on a one-size-fits-all solution.

Efficient and Effective Class Number Estimation. A key challenge in category discovery is determining the number of novel classes. Most existing approaches [16], [73] treat estimation as a post-hoc process, performed after representation learning is complete. More recently, methods such as GPC [74] and CiPR [81] have explored jointly estimating class numbers during representation learning, showing that the two processes can mutually reinforce each other. Nevertheless, open questions remain around efficiency, scalability, and stability in complex scenarios. This calls for future research in developing frameworks that can adaptively infer class numbers during training in an effective and efficient manner.

Cross-Modal Information. With the advancement of multi-modal learning, leveraging complementary modalities, particularly text, has proven effective in various visual tasks [117], [203]. In the CD literature, research has mainly focused on visual-only pipelines, leaving the role of multimodal cues comparatively underexplored. This raises the question of how vision-language models (VLMs) might support category discovery. Models like CLIP [117], pretrained on large-scale image-text pairs, offer rich, semantically grounded representations that can provide strong initialization for category

discovery. Yet, this advantage comes with concerns: recent findings [204] suggest its open-world performance may stem from memorization rather than generalization. This raises the risk of data leakage, especially when evaluation datasets overlap the CLIP training data. To ensure fair benchmarking, it is recommended to *include results on datasets disjoint from CLIP’s training data or across diverse domains with varying overlap* such as the use of NEV in GET [102], which lies outside CLIP’s pretraining.

Empirical comparisons in Tab. XIII and Tab. XIV also show that self-supervised vision models like DINOv2 [199] can outperform CLIP [117] without textual information, suggesting that representations learned purely from visual structures may be better suited for unsupervised category separation. These observations do not rule out the utility of VLMs in category discovery; rather, they highlight the importance of how and when to leverage multimodal signals. Going forward, an interesting direction lies in explicitly incorporating semantic priors extracted from language, such as textual descriptions or attribute names, without overly relying on VLM backbones. This may offer the benefits of multimodal supervision while avoiding potential pitfalls tied to memorization or domain bias.

From Single-Object to Complex Multi-Object Scenes. Most CD studies have been developed and evaluated on single-object datasets, a design convenient for benchmarking but unrepresentative of real scenes that typically contain multiple interacting objects, cluttered backgrounds, and occlusions. Consequently, most GCD methods implicitly operate at the image level under a single-salient-object assumption, limiting their ability to localize and disentangle multiple co-occurring instances. Extending CD to complex scenes introduces additional questions: how to disentangle features across instances, which objects should be discovered (potentially conditioned on user interest), and how to associate instance-level evidence with emerging categories. These considerations motivate advancing CD toward realistic multi-object environments.

Toward More Real-world Settings. In real-world deployments, CD systems are often required to operate continually, adapt to new data over time, and handle class imbalance and domain shifts simultaneously. This calls for unified frameworks that can integrate continual learning, cross-domain

generalization, distribution-agnostic modeling, and more. For instance, continual category discovery under imbalanced and heterogeneous domains remains largely underexplored, yet it closely mirrors real-world challenges such as dynamic environments in robotics or evolving data in medical diagnostics. Beyond image-level recognition, it is natural to consider extensions to instance- or pixel-level discovery in multi-object scenes, as well as to related tasks such as 3D depth estimation, video analysis, and action recognition.

VI. CONCLUSION

This survey provides a comprehensive review of the latest developments in the open-world problem of *Category Discovery*, marking the first extensive literature review on this topic. We begin by providing clear conceptual definitions for category discovery and systematically comparing them with related research areas. Existing methods are then organized by base settings and derived settings, followed by an in-depth analysis of their core components: *representation learning*, *label assignment*, and *category number estimation*. We also provide a comparative benchmark using commonly used datasets, offering a quantitative evaluation for all the methods. Finally, we distill key takeaways from prior studies and outline concrete, forward-looking research directions.

REFERENCES

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.
- [2] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *CVPR*, 2021.
- [3] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *CVPR*, 2020.
- [4] G. Zhang, Z. Luo, K. Cui, S. Lu, and E. P. Xing, "Meta-DETR: Image-level few-shot detection with inter-class correlation exploitation," *IEEE TPAMI*, 2022.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.
- [6] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [7] Y. Chen, M. Mancini, X. Zhu, and Z. Akata, "Semi-supervised and unsupervised deep visual learning: A survey," *IEEE TPAMI*, 2024.
- [8] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," in *NeurIPS*, 2018.
- [9] C. Geng, S.-j. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE TPAMI*, 2020.
- [10] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE TPAMI*, 2013.
- [11] J. Yang, P. Wang, D. Zou, Z. Zhou, K. Ding, W. Peng, H. Wang, G. Chen, B. Li, Y. Sun *et al.*, "Openood: Benchmarking generalized out-of-distribution detection," *NeurIPS*, 2022.
- [12] J. Yang, K. Zhou, Y. Li, and Z. Liu, "Generalized out-of-distribution detection: A survey," *arXiv preprint arXiv:2110.11334*, 2021.
- [13] K. Karbstein, L. Kösters, L. Hodač, M. Hofmann, E. Hörandl, S. Tomasello, N. D. Wagner, B. C. Emerson, D. C. Albach, S. Scheu *et al.*, "Species delimitation 4.0: integrative taxonomy meets artificial intelligence," *Trends in Ecology & Evolution*, 2024.
- [14] S. Kutuzova, M. Nielsen, P. Piera, J. N. Nissen, and S. Rasmussen, "Taxometer: Improving taxonomic classification of metagenomics contigs," *Nature Communications*, 2024.
- [15] K. Han, A. Vedaldi, and A. Zisserman, "Learning to discover novel visual categories via deep transfer clustering," in *ICCV*, 2019.
- [16] S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Generalized category discovery," in *CVPR*, 2022.
- [17] X. Zhang, J. Jiang, Y. Feng, Z.-F. Wu, X. Zhao, H. Wan, M. Tang, R. Jin, and Y. Gao, "Grow and merge: A unified framework for continuous categories discovery," in *NeurIPS*, 2022.
- [18] B. Zhao and O. Mac Aodha, "Incremental generalized category discovery," in *ICCV*, 2023.
- [19] N. Pu, Z. Zhong, X. Ji, and N. Sebe, "Federated generalized category discovery," in *CVPR*, 2024.
- [20] K. Han, X. Huang, Y. Li, S. Vaze, J. Li, and X. Jia, "What's in a name? beyond class indices for image recognition," in *CVPR workshop*, 2024.
- [21] H. Chi, F. Liu, B. Han, W. Yang, L. Lan, T. Liu, G. Niu, M. Zhou, and M. Sugiyama, "Meta discovery: Learning to discover novel classes given very limited data," in *ICLR*, 2022.
- [22] Z. Li, C. Meinel, and H. Yang, "Generalized categories discovery for long-tailed recognition," in *ICCV workshop*, 2023.
- [23] H. Wang, S. Vaze, and K. Han, "Hilo: A learning framework for generalized category discovery robust to domain shifts," in *ICLR*, 2025.
- [24] F. Zhu, S. Ma, Z. Cheng, X.-Y. Zhang, Z. Zhang, and C.-L. Liu, "Open-world machine learning: A review and new outlooks," *arXiv preprint arXiv:2403.01759*, 2024.
- [25] C. Troisemaine, V. Lemaire, S. Gosselin, A. Reiffers-Masson, J. Flocon-Cholet, and S. Vaton, "Novel class discovery: an introduction and key concepts," *arXiv preprint arXiv:2302.12028*, 2023.
- [26] K. Cao, M. Brbic, and J. Leskovec, "Open-world semi-supervised learning," in *ICLR*, 2022.
- [27] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, 2021.
- [28] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, 2010.
- [29] J. Wu, X. Li, S. Xu, H. Yuan, H. Ding, Y. Yang, X. Li, J. Zhang, Y. Tong, X. Jiang, B. Ghanem, and D. Tao, "Towards open vocabulary learning: A survey," *IEEE TPAMI*, 2024.
- [30] C. Zhu and L. Chen, "A survey on open-vocabulary detection and segmentation: Past, present, and future," *IEEE TPAMI*, 2024.
- [31] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, 2005.
- [32] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Trans. Intell. Syst. Technol.*, 2019.
- [33] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *CVPR*, 2009.
- [34] Y.-C. Hsu, Z. Lv, and Z. Kira, "Learning to cluster in order to transfer across domains and tasks," in *ICLR*, 2018.
- [35] Y.-C. Hsu, Z. Lv, J. Schlosser, P. Odom, and Z. Kira, "Multi-class classification without multi-class labels," in *ICLR*, 2019.
- [36] K. Han, S.-A. Rebuffi, S. Ehrhardt, A. Vedaldi, and A. Zisserman, "Automatically discovering and learning new visual categories with ranking statistics," in *ICLR*, 2020.
- [37] Y. Qing, Y. Zeng, Q. Cao, and G.-B. Huang, "End-to-end novel visual categories learning via auxiliary self-supervision," *Neural Networks*, 2021.
- [38] Z. Zhong, L. Zhu, Z. Luo, S. Li, Y. Yang, and N. Sebe, "Openmix: Reviving known knowledge for discovering novel visual categories in an open world," in *CVPR*, 2020.
- [39] Z. Zhong, E. Fini, S. Roy, Z. Luo, E. Ricci, and N. Sebe, "Neighborhood contrastive learning for novel class discovery," in *CVPR*, 2021.
- [40] E. Fini, E. Sangineto, S. Lathuilière, Z. Zhong, M. Nabi, and E. Ricci, "A unified objective for novel class discovery," in *ICCV*, 2021.
- [41] B. Zhao and K. Han, "Novel visual category discovery with dual ranking statistics and mutual knowledge distillation," in *NeurIPS*, 2021.
- [42] T. Mukherjee and N. Deligiannis, "Novel class discovery: A dependency approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.
- [43] J. Wang, Z. Ma, F. Nie, and X. Li, "Progressive self-supervised clustering with novel category discovery," *IEEE Transactions on Cybernetics*, 2022.
- [44] Z. Li, J. Otholt, B. Dai, D. Hu, C. Meinel, and H. Yang, "A closer look at novel class discovery from the labeled set," in *NeurIPS Workshop*, 2022.
- [45] Y. Liu and T. Tuytelaars, "Residual tuning: Toward novel category discovery without labels," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [46] Z. Li, J. Otholt, B. Dai, D. Hu, C. Meinel, and H. Yang, "Supervised knowledge may hurt novel class discovery performance," *Transactions on Machine Learning Research*, 2023.
- [47] W. Li, Z. Fan, J. Huo, and Y. Gao, "Modeling inter-class and intra-class constraints in novel class discovery," in *CVPR*, 2023.

- [48] Y. Sun, Z. Shi, Y. Liang, and Y. Li, “When and how does known class help discover unknown ones? provable understandings through spectral analysis,” in *ICML*, 2023.
- [49] P. Gu, C. Zhang, R. Xu, and X. He, “Class-relation knowledge distillation for novel class discovery,” in *ICCV*, 2023.
- [50] W. Feng, L. Ju, L. Wang, K. Song, and Z. Ge, “Towards novel class discovery: A study in novel skin lesions clustering,” in *MICCAI*, 2023.
- [51] Y. Liu, Y. Cai, Q. Jia, B. Qiu, W. Wang, and N. Pu, “Novel class discovery for ultra-fine-grained visual categorization,” in *CVPR*, 2024.
- [52] Y. Wang, Z. Chen, D. Yang, Y. Sun, and L. Qi, “Self-cooperation knowledge distillation for novel class discovery,” in *ECCV*, 2024.
- [53] L. Zhang, L. Qi, X. Yang, H. Qiao, M.-H. Yang, and Z. Liu, “Automatically discovering novel visual categories with adaptive prototype learning,” *IEEE TPAMI*, 2024.
- [54] Z. Hasan, A. Z. M. Faridee, M. Ahmed, S. Purushotham, H. Kwon, H. Lee, and N. Roy, “Novel categories discovery via constraints on empirical prediction statistics,” *arXiv preprint arXiv:2307.03856*, 2023.
- [55] J. Xie, R. Girshick, and A. Farhadi, “Unsupervised deep embedding for clustering analysis,” in *ICML*, 2016.
- [56] K. Han, S.-A. Rebuffi, S. Ehrhardt, A. Vedaldi, and A. Zisserman, “Autonovel: Automatically discovering and learning novel visual categories,” *IEEE TPAMI*, 2021.
- [57] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond empirical risk minimization,” in *ICLR*, 2017.
- [58] M. Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” in *NeurIPS*, 2013.
- [59] K. J. Joseph, S. Paul, G. Aggarwal, S. Biswas, P. Rai, K. Han, and V. N. Balasubramanian, “Spacing loss for discovering novel categories,” in *CVPR Workshop*, 2022.
- [60] X. Jia, K. Han, Y. Zhu, and B. Green, “Joint representation learning and novel category discovery on single- and multi-modal data,” in *ICCV*, 2021.
- [61] K. Han, S.-A. Rebuffi, S. Ehrhardt, A. Vedaldi, and A. Zisserman, “Autonovel: Automatically discovering and learning novel visual categories,” *IEEE TPAMI*, 2021.
- [62] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, “Emerging properties in self-supervised vision transformers,” in *ICCV*, 2021.
- [63] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, “Masked autoencoders are scalable vision learners,” in *CVPR*, 2022.
- [64] M. Yang, Y. Zhu, J. Yu, A. Wu, and C. Deng, “Divide and conquer: Compositional experts for generalized novel class discovery,” in *CVPR*, 2022.
- [65] M. N. Rizve, N. Kardan, S. Khan, F. Shahbaz Khan, and M. Shah, “Openlnd: Learning to discover novel classes for open-world semi-supervised learning,” in *ECCV*, 2022.
- [66] M. N. Rizve, N. Kardan, and M. Shah, “Towards realistic semi-supervised learning,” in *ECCV*, 2022.
- [67] L.-Z. Guo, Y.-G. Zhang, Z.-F. Wu, J.-J. Shao, and Y.-F. Li, “Robust semi-supervised learning when not all classes have labels,” in *NeurIPS*, 2022.
- [68] Y. Fei, Z. Zhao, S. Yang, and B. Zhao, “Xcon: Learning with experts for fine-grained category discovery,” in *British Machine Vision Conference (BMVC)*, 2022.
- [69] Y. Sun and Y. Li, “Opencon: Open-world contrastive learning,” *TMLR*, 2023.
- [70] S. Zhang, S. Khan, Z. Shen, M. Naseer, G. Chen, and F. Khan, “Promptcal: Contrastive affinity learning via auxiliary prompts for generalized novel category discovery,” in *CVPR*, 2023.
- [71] N. Pu, Z. Zhong, and N. Sebe, “Dynamic conceptional contrastive learning for generalized category discovery,” in *CVPR*, 2023.
- [72] J. Liu, Y. Wang, T. Zhang, Y. Fan, Q. Yang, and J. Shao, “Open-world semi-supervised novel class discovery,” in *IJCAI*, 2023.
- [73] X. Wen, B. Zhao, and X. Qi, “Parametric classification for generalized category discovery: A baseline study,” in *ICCV*, 2023.
- [74] B. Zhao, X. Wen, and K. Han, “Learning semi-supervised gaussian mixture models for generalized category discovery,” in *ICCV*, 2023.
- [75] F. Chiaroni, J. Dolz, Z. I. Masud, A. Mitiche, and I. Ben Ayed, “Parametric information maximization for generalized category discovery,” in *ICCV*, 2023.
- [76] Y. Wang, Z. Zhong, P. Qiao, X. Cheng, X. Zheng, C. Liu, N. Sebe, R. Ji, and J. Chen, “Discover and align taxonomic context priors for open-world semi-supervised learning,” in *NeurIPS*, 2023.
- [77] S. Vaze, A. Vedaldi, and A. Zisserman, “No representation rules them all in category discovery,” in *NeurIPS*, 2023.
- [78] S. Rastegar, H. Doughty, and C. Snoek, “Learn to categorize or categorize to learn? self-coding for generalized category discovery,” in *NeurIPS*, 2023.
- [79] Y. Sun, Z. Shi, and Y. Li, “A graph-theoretic framework for understanding open-world semi-supervised learning,” in *NeurIPS*, 2023.
- [80] X. Yang, X. Pan, I. King, and Z. Xu, “Generalized category discovery with clustering assignment consistency,” in *ICONIP*, 2024.
- [81] S. Hao, K. Han, and K.-Y. K. Wong, “Cipr: An efficient framework with cross-instance positive relations for generalized category discovery,” *TMLR*, 2024.
- [82] A. Banerjee, L. S. Kallooriyakath, and S. Biswas, “Amend: Adaptive margin and expanded neighborhood for efficient generalized category discovery,” in *WACV*, 2024.
- [83] J. Otholt, C. Meinel, and H. Yang, “Guided cluster aggregation: A hierarchical approach to generalized category discovery,” in *WACV*, 2024.
- [84] H. Wang, S. Vaze, and K. Han, “Sptnet: An efficient alternative framework for generalized category discovery with spatial prompt tuning,” in *ICLR*, 2024.
- [85] X. Cao, X. Zheng, G. Wang, W. Yu, Y. Shen, K. Li, Y. Lu, and Y. Tian, “Solving the catastrophic forgetting problem in generalized category discovery,” in *CVPR*, 2024.
- [86] S. Choi, D. Kang, and M. Cho, “Contrastive mean-shift learning for generalized category discovery,” in *CVPR*, 2024.
- [87] R. Xiao, L. Feng, K. Tang, J. Zhao, Y. Li, G. Chen, and H. Wang, “Targeted representation alignment for open-world semi-supervised learning,” in *CVPR*, 2024.
- [88] S. Ma, F. Zhu, Z. Zhong, X.-Y. Zhang, and C.-L. Liu, “Active generalized category discovery,” in *CVPR*, 2024.
- [89] H. Zheng, N. Pu, W. Li, N. Sebe, and Z. Zhong, “Textual knowledge matters: Cross-modality co-teaching for generalized visual class discovery,” in *ECCV*, 2024.
- [90] F. Yang, N. Pu, W. Li, Z. Luo, S. Li, N. Sebe, and Z. Zhong, “Learning to distinguish samples for generalized category discovery,” in *ECCV*, 2024.
- [91] S. Rastegar, M. Salehi, Y. M. Asano, H. Doughty, and C. G. M. Snoek, “Selex: Self-expertise in fine-grained generalized category discovery,” in *ECCV*, 2024.
- [92] B. Ye, K. Gan, T. Wei, and M.-L. Zhang, “Bridging the gap: Learning pace synchronization for open-world semi-supervised learning,” *arXiv preprint arXiv:2309.11930*, 2023.
- [93] S. Niu, L. Lin, J. Huang, and C. Wang, “Owmatch: Conditional self-labeling with consistency for open-world semi-supervised learning,” in *NeurIPS*, 2024.
- [94] H. Lin, W. An, J. Wang, Y. Chen, F. Tian, M. Wang, Q. Wang, G. Dai, and J. Wang, “Flipped classroom: Aligning teacher attention with student in generalized category discovery,” in *NeurIPS*, 2024.
- [95] T. Luo, M. Du, J. Shi, X. Chen, B. Zhao, and S. Huang, “Contextuality helps representation learning for generalized category discovery,” in *International Conference on Image Processing (ICIP)*, 2024.
- [96] Y. Duan, Z. Hu, R. Wang, Z. Sun, F. Nie, and X. Li, “Mutual-support generalized category discovery,” *Information Fusion*, 2025.
- [97] M. Yang, J. Yin, Y. Gu, C. Deng, H. Zhang, and H. Zhu, “Consistent prompt tuning for generalized category discovery,” *IJCV*, 2025.
- [98] M. Wang, Z. Zhong, and X. Gong, “Prior-constrained association learning for fine-grained generalized category discovery,” in *AAAI*, 2025.
- [99] Y. Liu and K. Han, “Debgcd: Debaised learning with distribution guidance for generalized category discovery,” in *ICLR*, 2025.
- [100] S. Ma, F. Zhu, X.-Y. Zhang, and C.-L. Liu, “Protogcd: Unified and unbiased prototype learning for generalized category discovery,” *IEEE TPAMI*, 2025.
- [101] Z. Peng, J. Ma, Z. Sun, R. Yi, H. Song, X. Tan, and L. Ma, “Mos: Modeling object-scene associations in generalized category discovery,” in *CVPR*, 2025.
- [102] E. Wang, Z. Peng, Z. Xie, X. Liu, and M.-M. Cheng, “Get: Unlocking the multi-modal potential of clip for generalized category discovery,” in *CVPR*, 2025.
- [103] W. Zhang, B. Zhang, Z. Teng, W. Luo, J. Zou, and J. Fan, “Less attention is more: Prompt transformer for generalized category discovery,” in *CVPR*, 2025.
- [104] Q. Dai, H. Huang, Y. Wu, and S. Yang, “Adaptive part learning for fine-grained generalized category discovery: A plug-and-play enhancement,” in *CVPR*, 2025.
- [105] Y. Liu, Z. He, and K. Han, “Hyperbolic category discovery,” in *CVPR*, 2025.

- [106] L. Tang, K. Huang, C. Chen, Y. Yuan, C. Li, X. Tu, X. Ding, and Y. Huang, "Dissecting generalized category discovery: Multiplex consensus under self-deconstruction," in *ICCV*, 2025.
- [107] Q. Xu, Z. Hu, Y. Duan, E. Pei, and Y. Tai, "A hidden stumbling block in generalized category discovery: Distracted attention," in *ICCV*, 2025.
- [108] Z. He, Y. Liu, and K. Han, "Seal: Semantic-aware hierarchical learning for generalized category discovery," in *NeurIPS*, 2025.
- [109] R. Ouldnoughi, C.-W. Kuo, and Z. Kira, "Clip-gcd: Simple language guided generalized category discovery," *arXiv preprint arXiv:2305.10420*, 2023.
- [110] Y. Tu, Z. Zhong, Y. Li, and H. Zhao, "Memory consistency guided divide-and-conquer learning for generalized category discovery," *arXiv preprint arXiv:2401.13325*, 2024.
- [111] Y. Wang, Y. Wang, Y. Wu, B. Zhao, and X. Qian, "Beyond known clusters: Probe new prototypes for efficient generalized class discovery," *arXiv preprint arXiv:2404.08995*, 2024.
- [112] Z. Tan, C. Zhang, X. Yang, J. Sun, and K. Huang, "Revisiting mutual information maximization for generalized category discovery," *arXiv preprint arXiv:2405.20711*, 2024.
- [113] F. Gao, W. Zhong, Z. Cao, X. Peng, and Z. Li, "Opengcd: Assisting open world recognition with generalized category discovery," *arXiv preprint arXiv:2308.06926*, 2023.
- [114] C. Zhang, P. Gu, X. Yu, and X. He, "Composing novel classes: A concept-driven approach to generalized category discovery," *arXiv preprint arXiv:2410.13285*, 2024.
- [115] T. Luo, Y. Liu, Y. Liu, A. Zhang, X. Wang, Y. Zhan, C. Tang, L. Liu, and Z. Chen, "Dig-face: De-biased learning for generalized facial expression category discovery," *arXiv preprint arXiv:2409.20098*, 2024.
- [116] Z. Ji, S. Yang, J. Liu, Y. Pang, and J. Han, "A fresh look at generalized category discovery through non-negative matrix factorization," *arXiv preprint arXiv:2410.21807*, 2024.
- [117] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *ICML*, 2021.
- [118] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," in *NeurIPS*, 2020.
- [119] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE TPAMI*, 1995.
- [120] R. Du, D. Chang, K. Liang, T. Hospedales, Y.-Z. Song, and Z. Ma, "On-the-fly category discovery," in *CVPR*, 2023.
- [121] H. Zheng, N. Pu, W. Li, N. Sebe, and Z. Zhong, "Prototypical hash encoding for on-the-fly fine-grained category discovery," in *NeurIPS*, 2024.
- [122] M. Rosvall, D. Axelsson, and C. T. Bergstrom, "The map equation," *The European Physical Journal Special Topics*, 2009.
- [123] N. Dugué and A. Perez, "Directed louvain: maximizing modularity in directed networks," Ph.D. dissertation, Université d'Orléans, 2015.
- [124] M. S. Sarfraz, V. Sharma, and R. Stiefelhofen, "Efficient parameter-free clustering using first neighbor relations," in *CVPR*, 2019.
- [125] L.-Z. Guo, Z.-Y. Zhang, Y. Jiang, Y.-F. Li, and Z.-H. Zhou, "Safe deep semi-supervised learning for unseen-class unlabeled data," in *ICML*, 2000.
- [126] G. Hamerly and C. Elkan, "Learning the k in k-means," in *NeurIPS*, 2003.
- [127] C. Leiber, L. G. Bauer, B. Schelling, C. Böhm, and C. Plant, "Dip-based deep embedded clustering with k-estimation," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021.
- [128] K. J. Joseph, S. Paul, G. Aggarwal, S. Biswas, P. Rai, K. Han, and V. N. Balasubramanian, "Novel class discovery without forgetting," in *ECCV*, 2022.
- [129] S. Roy, M. Liu, Z. Zhong, N. Sebe, and E. Ricci, "Class-incremental novel class discovery," in *ECCV*, 2022.
- [130] H. Kim, S. Suh, D. Kim, D. Jeong, H. Cho, and J. Kim, "Proxy anchor-based unsupervised learning for continuous generalized category discovery," in *ICCV*, 2023.
- [131] Y. Wu, Z. Chi, Y. Wang, and S. Feng, "Metagcd: Learning to continually learn in generalized category discovery," in *ICCV*, 2023.
- [132] M. Liu, S. Roy, Z. Zhong, N. Sebe, and E. Ricci, "Large-scale pre-trained models are surprisingly strong in incremental novel class discovery," in *ICPR*, 2024.
- [133] G. Chen, P. Peng, Y. Huang, M. Geng, and Y. Tian, "Adaptive discovering and merging for incremental novel class discovery," in *AAAI*, 2024.
- [134] F. J. Cendra, B. Zhao, and K. Han, "Promptgcd: Learning gaussian mixture prompt pool for continual category discovery," in *ECCV*, 2024.
- [135] K.-H. Park, H. Lee, K. Song, and G.-M. Park, "Online continuous generalized category discovery," in *ECCV*, 2024.
- [136] G. Rypś, D. Marczak, S. Cygert, T. Trzcinski, and B. Twardowski, "Category adaptation meets projected distillation in generalized continual category discovery," in *ECCV*, 2024.
- [137] S. Ma, F. Zhu, Z. Zhong, W. Liu, X.-Y. Zhang, and C.-L. Liu, "Happy: A debiased learning framework for continual generalized category discovery," in *NeurIPS*, 2024.
- [138] H. Dai and J. Chauhan, "Continual generalized category discovery: Learning and forgetting from a bayesian perspective," in *ICML*, 2025.
- [139] Y. Yu, S. Wang, Y. He, J. Chen, and Y. Gong, "Continual novel class discovery via feature enhancement and adaptation," *arXiv preprint arXiv:2405.06389*, 2024.
- [140] S. Kim, D. Kim, M. Cho, and S. Kwak, "Proxy anchor loss for deep metric learning," in *CVPR*, 2020.
- [141] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, 2007.
- [142] X. Liu, N. Pu, H. Zheng, W. Li, N. Sebe, and Z. Zhong, "Generate, refine, and encode: Leveraging synthesized novel samples for on-the-fly fine-grained category discovery," in *ICCV*, 2025.
- [143] Q. Yu, D. Ikami, G. Irie, and K. Aizawa, "Self-labeling framework for novel category discovery over domains," in *AAAI*, 2022.
- [144] J. Zhuang, Z. Chen, P. Wei, G. Li, and L. Lin, "Open set domain adaptation by novel class discovery," in *IEEE International Conference on Multimedia and Expo*, 2022.
- [145] Z. Zang, L. Shang, S. Yang, F. Wang, B. Sun, X. Xie, and S. Z. Li, "Boosting novel category discovery over domains with soft contrastive learning and all-in-one classifier," in *ICCV*, 2023.
- [146] S. B. Rongali, S. Mehrotra, A. Jha, M. H. N. C. S. Bose, T. Gupta, M. Singha, and B. Banerjee, "Cdad-net: Bridging domain gaps in generalized category discovery," in *CVPR workshop*, 2024.
- [147] Y. Wang, F. Liu, J. Liu, and K. Sun, "Exclusive style removal for cross domain novel class discovery," *arXiv preprint arXiv:2406.18140*, 2024.
- [148] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, 1953.
- [149] Y. Hong and Y. Chen, "Patchmix: patch-level mixup for data augmentation in convolutional neural networks," *Knowledge and Information Systems*, 2024.
- [150] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *ICML*, 2009.
- [151] C. Zhang, R. Xu, and X. He, "Novel class discovery for long-tailed recognition," *Transactions on Machine Learning Research*, 2023.
- [152] Z. Li, B. Dai, F. Simsek, C. Meinel, and H. Yang, "Imbagcd: Imbalanced generalized category discovery," in *CVPR workshop*, 2023.
- [153] M. Yang, L. Wang, C. Deng, and H. Zhang, "Bootstrap your own prior: Towards distribution-agnostic novel class discovery," in *CVPR*, 2023.
- [154] J. Bai, Z. Liu, H. Wang, R. Chen, L. Mu, X. Li, J. T. Zhou, Y. Feng, J. Wu, and H. Hu, "Towards distribution-agnostic generalized category discovery," in *NeurIPS*, 2023.
- [155] J. Fan, D. Liu, H. Chang, H. Huang, M. Chen, and W. Cai, "Seeing unseen: Discover novel biomedical concepts via geometry-constrained probabilistic modeling," in *CVPR*, 2024.
- [156] B. Zhao and K. Han, "Generalized category discovery under the long-tailed distribution," *arXiv preprint arXiv:2506.12515*, 2025.
- [157] W. Wang, T. Lei, Q. Chen, and Y. Liu, "Semantic-guided novel category discovery," in *AAAI*, 2024.
- [158] M. Liu, S. Roy, W. Li, Z. Zhong, N. Sebe, and E. Ricci, "Democratizing fine-grained visual recognition with large language models," in *ICLR*, 2024.
- [159] J. Zhang, X. Ma, S. Guo, and W. Xu, "Towards unbiased training in federated open-world semi-supervised learning," in *ICML*, 2023.
- [160] L. Wang, C. Liu, J. Guo, J. Dong, X. Wang, H. Huang, and Q. Zhu, "Federated continual novel class learning," in *ICCV*, 2025.
- [161] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.
- [162] J. Li, D. Li, S. Savarese, and S. Hoi, "Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," in *ICML*, 2023.
- [163] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *ICML*, 2017.
- [164] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *NeurIPS*, 2017.
- [165] Y. Zhao, Z. Zhong, N. Sebe, and G. H. Lee, "Novel class discovery in semantic segmentation," in *CVPR*, 2022.
- [166] J. Zheng, W. Li, J. Hong, L. Petersson, and N. Barnes, "Towards open-set object detection and discovery," in *CVPR workshop*, 2022.

- [167] J. Feng, Y. Yang, Y. Xie, Y. Li, Y. Guo, Y. Guo, Y. He, L. Xiang, and G. Ding, "Debiased novel category discovering and localization," in *AAAI*, 2024.
- [168] T. L. Hayes, C. R. de Souza, N. Kim, J. Kim, R. Volpi, and D. Larlus, "Pandas: Prototype-based novel class discovery and detection," in *The Third Conference on Lifelong Learning Agents - CoLLAs 2024*, 2024.
- [169] C. M. Hoang, Y. Lee, and B. Kang, "Generalized class discovery in instance segmentation," in *AAAI*, 2025.
- [170] Z. Hasan, M. Ahmed, A. Z. M. Faridee, S. Purushotham, H. Kwon, H. Lee, and N. Roy, "Nev-ncd: Negative learning, entropy, and variance regularization based novel action categories discovery," in *IEEE International Conference on Image Processing (ICIP)*, 2023.
- [171] L. Riz, C. Saltori, E. Ricci, and F. Poiesi, "Novel class discovery for 3d point cloud semantic segmentation," in *CVPR*, 2023.
- [172] T. Weng, J. Xiao, and H. Jiang, "Decompose novel into known: Part concept learning for 3d novel class discovery," in *NeurIPS*, 2023.
- [173] R. Xu, C. Zhang, H. Ren, and X. He, "Dual-level adaptive self-labeling for novel class discovery in point cloud segmentation," in *ECCV*, 2024.
- [174] W. An, F. Tian, P. Chen, S. Tang, Q. Zheng, and Q. Wang, "Fine-grained category discovery under coarse-grained supervision with hierarchical weighted self-contrastive learning," in *EMNLP*, 2022.
- [175] W. An, F. Tian, Q. Zheng, W. Ding, Q. Wang, and P. Chen, "Generalized category discovery with decoupled prototypical network," in *AAAI*, 2023.
- [176] W. Hogan, J. Li, and J. Shang, "Open-world semi-supervised generalized relation discovery aligned in a real-world setting," in *EMNLP*, H. Bouamor, J. Pino, and K. Bali, Eds., 2023.
- [177] W. An, F. Tian, W. Shi, Y. Chen, Y. Wu, Q. Wang, and P. Chen, "Transfer and alignment network for generalized category discovery," in *AAAI*, 2024.
- [178] W. Shi, W. An, F. Tian, Y. Chen, Y. Wu, Q. Wang, and P. Chen, "A unified knowledge transfer network for generalized category discovery," in *AAAI*, 2024.
- [179] W. An, H. Lin, J. Nie, F. Tian, W. Shi, Y. Wu, Q. Wang, and P. Chen, "Unleashing the potential of model bias for generalized category discovery," in *AAAI*, 2025.
- [180] J. Wang, L. Zhang, J. Liu, T. Guo, and W. Wu, "Learning from semi-factals: A debiased and semantic-aware framework for generalized relation discovery," *arXiv preprint arXiv:2401.06327*, 2024.
- [181] C. Troisemaine, J. Flocon-Cholet, S. Gosselin, S. Vaton, A. Reiffers-Masson, and V. Lemaire, "A method for discovering novel classes in tabular data," in *IEEE International Conference on Knowledge Graph, ICKG*. IEEE, 2022.
- [182] C. Troisemaine, A. Reiffers-Masson, S. Gosselin, V. Lemaire, and S. Vaton, "A practical approach to novel class discovery in tabular data," in *International Conference on Data Mining and Knowledge Discovery*, 2024.
- [183] C. Troisemaine, J. Flocon-Cholet, S. Gosselin, A. Reiffers-Masson, S. Vaton, and V. Lemaire, "An interactive interface for novel class discovery in tabular data," in *Machine Learning and Knowledge Discovery in Databases: Applied Data Science and Demo Track*, 2023.
- [184] Y. Hou, X. Chen, H. Zhu, R. Liu, B. Shi, J. Liu, J. Wu, and K. Xu, "Nc-ncd: Novel class discovery for node classification," in *Conference on Information and Knowledge Management*, 2024.
- [185] J. Zhou, Y. Liu, and Q. Chen, "Novel class discovery in chest x-rays via paired images and text," in *AAAI*, 2024.
- [186] Y. Su, R. Zhou, S. Huang, X. Li, T. Wang, Z. Wang, and M. Xu, "Multimodal generalized category discovery," *arXiv preprint arXiv:2409.11624*, 2024.
- [187] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "Basnet: Boundary-aware salient object detection," in *CVPR*, 2019.
- [188] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019.
- [189] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *IJCV*, 2015.
- [190] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Master's thesis, Department of Computer Science, University of Toronto*, 2009.
- [191] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The Caltech-UCSD Birds-200-2011 Dataset," California Institute of Technology, Tech. Rep., 2011.
- [192] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3d object representations for fine-grained categorization," in *International IEEE Workshop on 3D Representation and Recognition (3dRR)*, 2013.
- [193] S. Maji, E. Rahtu, J. Kannala, M. Blaschko, and A. Vedaldi, "Fine-grained visual classification of aircraft," *arXiv preprint arXiv:1306.5151*, 2013.
- [194] S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Open-set recognition: a good closed-set classifier is all you need?" in *ICLR*, 2022.
- [195] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE TPAMI*, 2006.
- [196] Y. Le and X. Yang, "Tiny imagenet visual recognition challenge," *CS 231N*, 2015.
- [197] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, 1955.
- [198] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.
- [199] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. HAZIZA, F. Massa, A. El-Nouby *et al.*, "Dinov2: Learning robust visual features without supervision," *TMLR*, 2024.
- [200] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *ICLR*, 2020.
- [201] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *ICLR*, 2018.
- [202] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *CVPR*, 2020.
- [203] Y. Zhang, X. Ding, K. Gong, Y. Ge, Y. Shan, and X. Yue, "Multi-modal pathway: Improve transformers with irrelevant data from other modalities," in *CVPR*, 2024.
- [204] P. Mayilvahanan, R. S. Zimmermann, T. Wiedemer, E. Rusak, A. Juhos, M. Bethge, and W. Brendel, "In search of forgotten domain generalization," in *ICLR*, 2025.
- [205] G. Perez, D. Sheldon, G. Van Horn, and S. Maji, "Human-in-the-loop visual re-id for population size estimation," in *ECCV*, 2024.
- [206] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. V. Jawahar, "Cats and dogs," in *CVPR*, 2012.
- [207] K. C. Tan, Y. Liu, B. Ambrose, M. Tulig, and S. Belongie, "The herbarium challenge 2019 dataset," in *Workshop on Fine-Grained Visual Categorization*, 2019.
- [208] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.
- [209] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, 2015.
- [210] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *NeurIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.
- [211] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *ICCV*, 2019.
- [212] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *ECCV*, 2010.
- [213] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *CVPR*, 2017.
- [214] X. Peng, B. Usman, N. Kaushik, J. Hoffman, D. Wang, and K. Saenko, "Visda: The visual domain adaptation challenge," *arXiv preprint arXiv:1710.06924*, 2017.
- [215] G. Van Horn, E. Cole, S. Beery, K. Wilber, S. Belongie, and O. Mac Aodha, "Benchmarking representation learning for natural world image collections," in *CVPR*, 2021.
- [216] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, 2015.
- [217] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*, 2014.
- [218] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," in *ICCV*, 2019.
- [219] Y. Pan, B. Gao, J. Mei, S. Geng, C. Li, and H. Zhao, "Semanticpos: A point cloud dataset with large quantity of dynamic instances," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020.
- [220] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *CVPR*, 2015.

- [221] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su *et al.*, “Shapenet: An information-rich 3d model repository,” *arXiv preprint arXiv:1512.03012*, 2015.
- [222] M. A. Uy, Q.-H. Pham, B.-S. Hua, T. Nguyen, and S.-K. Yeung, “Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data,” in *CVPR*, 2019.
- [223] J. Reizenstein, R. Shapovalov, P. Henzler, L. Sbordone, P. Labatut, and D. Novotny, “Common objects in 3d: Large-scale learning and evaluation of real-life 3d category reconstruction,” in *ICCV*, 2021.
- [224] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev *et al.*, “The kinetics human action video dataset,” *arXiv preprint arXiv:1705.06950*, 2017.
- [225] H. Chen, W. Xie, A. Vedaldi, and A. Zisserman, “Vggsound: A large-scale audio-visual dataset,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.
- [226] K. Soomro, A. R. Zamir, and M. Shah, “Ucf101: A dataset of 101 human actions classes from videos in the wild,” *arXiv preprint arXiv:1212.0402*, 2012.
- [227] S. Larson, A. Mahendran, J. J. Peper, C. Clarke, A. Lee, P. Hill, J. K. Kummerfeld, K. Leach, M. A. Laurenzano, L. Tang, and J. Mars, “An evaluation dataset for intent classification and out-of-scope prediction,” in *EMNLP*, 2019.
- [228] X. Liu, A. Eshghi, P. Swietojanski, and V. Rieser, “Benchmarking natural language understanding services for building conversational agents,” in *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*, 2021.
- [229] K. Kowsari, D. E. Brown, M. Heidarysafa, K. Jafari Meimandi, M. S. Gerber, and L. E. Barnes, “Hdltext: Hierarchical deep learning for text classification,” in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017.
- [230] I. Casanueva, T. Temčinas, D. Gerz, M. Henderson, and I. Vulić, “Efficient intent detection with dual sentence encoders,” in *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, 2020.
- [231] J. Xu, P. Wang, G. Tian, B. Xu, J. Zhao, F. Wang, and H. Hao, “Short text clustering via convolutional neural networks,” in *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 2015.
- [232] S. Li, H. Ji, and J. Han, “Open relation and event type discovery with type abstraction,” in *EMNLP*, 2022.
- [233] G. Stoica, E. A. Platanios, and B. Póczos, “Re-tacred: Addressing shortcomings of the tacred dataset,” in *AAAI*, 2021.
- [234] X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, and M. Sun, “Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation,” *arXiv preprint arXiv:1810.10147*, 2018.

Overview. In this appendix, we provide an off-the-shelf benchmark for class number estimation, offering a unified comparison of existing methods. Additionally, we include further experimental details and results, as well as extended descriptions of the datasets commonly used in category discovery tasks.

APPENDIX A

BENCHMARK OF ESTIMATING THE NUMBER OF UNKNOWN CATEGORIES

We further report the estimated number of categories for six commonly used datasets. Table A1 presents the estimated number of classes (K) and the corresponding error rates ($\text{Err}\%$) across various datasets for several methods, including GCD [16], DCCL [71], PIM [75], CiPR [81], GPC [74], PAL-GCD [98], and CMS [86]. The ground truth number of classes for each dataset (CIFAR-10 [190], CIFAR-100 [190], ImageNet-100 [189], CUB [191], Stanford-Cars [192], and FGVC-Aircraft [193]) is provided in the first row for reference.

Beyond methods developed within the GCD community, other domains have also proposed effective strategies for estimating the number of categories in unlabelled datasets. For example, Perez *et al.* [205] introduce a class number estimation approach for Re-ID tasks based on nested importance sampling and human-in-the-loop feedback. Their method involves limited human supervision to determine whether a given image pair belongs to the same class, which aligns well with our partially labelled setting. To evaluate its applicability in GCD, we reimplement their method using a DINOv2 [199] ViT-Base backbone and apply it to both fine-grained and generic datasets. For fine-grained datasets, we sample only 0.1% of all possible image pairs from the labelled data, while for CIFAR-10 and CIFAR-100, we use just 0.01%. For ImageNet-100, due to its larger image count, we only employ 5000 pairs, which is around 10^{-6} of all possible image pairs from the labelled data. The estimation results are reported in Table A1.

APPENDIX B

BENCHMARK OF CONTINUAL CATEGORY DISCOVERY (CCD)

We additionally report the benchmark results for CCD methods on fine-grained datasets, including CUB [191], Stanford-Cars [192] and FGVC-Aircraft [193] as well as generic datasets including CIFAR-100 [190], ImageNet-100 [189], TinyImageNet [196] and Caltech-101 [195] based on DINO [62] pre-trained ViT-Base [200] backbone. Table A4 shows the performance of seven state-of-the-art GCD methods. The performance is measured by $cACC$ across the ‘All’, ‘Old’ and ‘New’ categories, as introduced in Section IV (Evaluation Protocol).

Notably, we employ the data splitting strategy in PromptCCD [134] with details reported in Table A3.

APPENDIX C

ADDITIONAL EXPERIMENTAL DETAILS OF GCD BENCHMARK

In this section, we provide results on additional datasets - Oxford-Pets [206], Herbarium19 [207] and ImageNet-1K [189]. The performance is measured by ACC across the ‘All’, ‘Old’ and ‘New’ categories. Table A5 shows the performance of 12 state-of-the-art GCD methods on this two datasets based on DINO [62] pre-trained ViT-B [200] backbone.

APPENDIX D

NOTES ON REPRODUCTION SETTINGS AND PROTOCOL DIFFERENCES

As discussed in Section IV (GCD Benchmarking) of the main paper, several methods [26], [65], [66], [69], [72], [76], [79], [87], [92], [93] adopt different data-splitting protocols or backbones, leading to inconsistencies in the composition of labelled and unlabelled sets. In this section, we provide further details on these differences and explain our choice to reproduce OwMatch [93] as a representative method using the commonly adopted data-splitting protocol and the DINO-pretrained ViT-Base backbone.

Several methods [26], [65], [66], [69], [72], [76], [79], [87], [92], [93] utilize ResNet-18 or ResNet-50 [208] as backbone networks, in contrast to most GCD methods which adopt ViT-Base [200] as the standard backbone. In addition, these methods follow different dataset-splitting protocols, leading to inconsistencies in the composition of labelled and unlabelled data. Several of them [65], [66], [76], [92], [93] also adopt an inductive evaluation setting, where performance is measured on a separate, unseen test set. Among these, [66], [76] use a 1:9 ratio for labelled-to-unlabelled data, whereas the majority of GCD methods employ a 5:5 split. In contrast, [26], [69], [72], [79] follow a transductive evaluation protocol consistent with most GCD methods, but apply a distinct dataset-splitting strategy that results in different labelled and unlabelled sets. Notably, these four methods share the same alternative splitting protocol.

Table A7 presents a comparison of the aforementioned methods. To ensure fair benchmarking, we select the best-performing method under each splitting protocol—OwMatch [93], TIDA [76], and SORL [79]—for further evaluation. Since all three originally use ResNet [208] as their backbone, we reimplement them using a ViT-Base backbone pretrained with DINO [62] to align with the standard setup. The reimplement results are shown in Table A8. Based on these results, OwMatch [93] achieves the best performance across both CIFAR-10 and CIFAR-100. Consequently, we choose OwMatch for reimplement and evaluation on six commonly used datasets—CUB [191], Stanford Cars [192], FGVC-Aircraft [193], CIFAR-10 [190], CIFAR-100 [190], and ImageNet-100 [189]—as reported in the main paper.

APPENDIX E

DATASET

In this section, we present a more detailed introduction to the datasets used across different CD settings. Additional

TABLE A1
ESTIMATING THE NUMBER OF CLASSES K AND THE ERROR RATE (ERR%).

Method	CIFAR10		CIFAR100		ImageNet-100		CUB		Stanford Cars		FGVC Aircraft	
	K	Err(%)	K	Err(%)	K	Err(%)	K	Err(%)	K	Err(%)	K	Err(%)
Ground truth	10	-	100	-	100	-	200	-	196	-	100	-
GCD [16]	9	1	100	0	109	9	231	15.5	230	17.3	-	-
DCCL [71]	14	4	146	46	129	29	172	14	192	0.02	-	-
PIM [75]	10	0	95	5	102	2	227	13.5	169	13.8	-	-
CIPIR [81]	12	2	103	3	100	0	182	9	182	7.14	-	-
GPC [74]	10	0	100	0	103	3	212	6	201	0.03	-	-
PAL-GCD [98]	-	-	-	-	-	-	176	12	172	12.2	77	23
CMS [86]	-	-	97	3	116	16	170	15	156	20.4	98	2
DINOv2												
GCD [16]	10	0	100	0	100	0	211	5.5	215	9.69	105	5
Perez <i>et al.</i> $\ddagger$ [205]	10	0	100	0	101	1	202	1	194	1.02	123	23

$\ddagger$ denotes our implementation using the DINOv2 pretrained ViT-Base backbone.

TABLE A2
STATISTICS OF THE CCD BENCHMARK DATASETS FOLLOWING THE SPLITS IN TAB. A3.

Stages	CIFAR100 [190]		ImageNet100 [189]		TinyImageNet [196]		Caltech101 [195]		Aircraft [193]		SCars [192]		CUB [191]	
	#Class	#Image	#Class	#Image	#Class	#Image	#Class	#Image	#Class	#Image	#Class	#Image	#Class	#Image
Stage 0 (D^1)	70	30.45K	70	77.46K	140	60.90K	71	4.70K	70	1.98K	130	4.62K	140	3.65K
Stage 1 (D^2_1)	80	5.95K	80	15.14K	160	11.90K	81	0.73K	80	0.37K	152	0.98K	160	0.71K
Stage 2 (D^2_2)	90	6.55K	90	16.66K	180	13.10K	91	0.65K	90	0.43K	174	1.13K	180	0.79K
Stage 3 (D^2_3)	100	7.05K	100	17.94K	200	14.10K	101	1.12K	100	0.55K	196	1.38K	200	0.85K

TABLE A3
DATA SPLITS FOR CCD.

Class splits	D^1	D^2_1	D^2_2	D^2_3
$\{y_i \mid y_i \leq 0.7 * \mathcal{Y} \}$	87%	7%	3%	3%
$\{y_i \mid 0.7 * \mathcal{Y} < y_i \leq 0.8 * \mathcal{Y} \}$	0%	70%	20%	10%
$\{y_i \mid 0.8 * \mathcal{Y} < y_i \leq 0.9 * \mathcal{Y} \}$	0%	0%	90%	10%
$\{y_i \mid 0.9 * \mathcal{Y} < y_i \leq \mathcal{Y} \}$	0%	0%	0%	100%

TABLE A4
COMPARISON OF STATE-OF-THE-ART CCD METHODS. RESULTS ARE REPORTED IN $cACC$ ACROSS ‘ALL’, ‘OLD’ AND ‘NEW’ CATEGORIES.

Method	CIFAR-100			ImageNet-100			TinyImageNet			Caltech-101			FGVC-Aircraft			Stanford-Cars			CUB		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
ORCA † [26]	60.91	66.61	58.33	40.29	45.85	35.40	54.71	63.13	51.93	76.77	82.80	73.20	30.77	25.71	32.44	20.79	33.40	17.60	41.73	66.19	34.14
GCD † [16]	58.18	72.27	52.83	69.41	81.56	65.65	55.20	65.87	51.61	78.27	86.60	72.92	47.37	61.43	42.53	39.21	58.29	33.45	54.98	75.47	48.15
SimGCD † [73]	25.56	38.76	20.43	31.38	40.47	27.44	33.40	29.11	34.74	33.65	37.53	31.62	29.03	35.72	25.61	21.01	40.93	16.48	39.89	59.25	33.75
GM † [17]	57.43	63.68	55.31	67.84	75.10	66.60	52.14	59.68	49.96	75.75	83.66	71.59	31.06	33.33	30.78	21.90	35.29	18.17	38.87	65.00	30.29
MetaGCD † [131]	55.49	69.38	48.98	66.41	80.54	60.65	55.26	66.12	50.79	80.75	89.02	75.86	44.63	59.05	39.39	35.98	56.97	29.96	44.59	74.40	35.40
PA-CGCD † [130]	58.25	87.11	49.04	64.79	91.15	57.83	51.13	74.95	43.52	77.96	94.75	69.66	48.24	73.09	40.60	48.88	80.43	33.54	52.48	77.26	44.74
PromptCCD [134]	64.17	75.57	60.34	76.16	81.76	74.35	61.84	66.54	60.26	82.44	89.08	79.72	52.64	60.48	50.23	44.07	66.36	36.83	55.45	75.48	48.56

† denotes results from [134].

TABLE A5
ADDITIONAL EXPERIMENTS COMPARISON FOR GCD METHODS ON HERBARIUM19 [207] AND OXFORD-PETS [206].

Method	Oxford-Pets [206]			Herbarium19 [207]		
	All	Old	New	All	Old	New
ORCA [26]	-	-	-	24.6	26.5	23.7
GCD [16]	80.2	85.1	77.6	35.4	51.0	27.0
XCon [68]	86.7	91.5	84.1	-	-	-
OpenCon [69]	-	-	-	39.3	58.9	28.6
DCCL [71]	88.1	88.2	88.0	-	-	-
SimGCD [73]	91.7	83.6	96.0	44.0	58.0	36.4
μ GCD [77]	-	-	-	45.8	61.9	37.2
InfoSieve [78]	90.7	95.2	88.4	40.3	59.0	30.2
Selex [91]	92.5	91.9	92.8	39.6	54.9	31.3
DebGCD [99]	93.0	86.4	96.5	44.7	59.4	36.8
HypCD [105]	92.7	91.5	93.3	45.1	60.1	36.9
SEAL [108]	92.9	88.9	95.0	46.9	45.8	48.2

TABLE A6
ADDITIONAL EXPERIMENTS COMPARISON FOR GCD METHODS ON IMAGENET-1K [189].

Method	ImageNet-1K [189]		
	All	Old	New
GCD [16]	52.5 †	72.5 †	42.2 †
SimGCD [73]	57.1	77.3	46.9
Contextuality-GCD [95]	59.5	79.9	49.5
LegoGCD [85]	62.4	79.5	53.8
DebGCD [99]	65.0	82.0	56.5

† denotes results from [73].

statistics on the dataset splits for GCD are provided in Table A9.

A. Generic Datasets

ImageNet-1K [189] & ImageNet-100: ImageNet [189] is a widely used dataset for natural image classification in computer vision. It contains over one million labelled images, spanning 1,000 different object categories. ImageNet-100 is constructed by randomly subsampling 100 classes.

TinyImageNet [196] is a scaled-down version of the ImageNet [189] dataset designed for benchmarking machine learning models. It contains 200 classes, with each class represented by 500 training images and 50 validation images, totaling 100,000 images. All images are resized to 64x64 pixels, providing a challenging yet computationally efficient dataset.

CIFAR-10 & CIFAR-100 [190]: CIFAR-10 contains 50,000 images spanning across 10 different classes and CIFAR-100 includes 100 classes, with each class containing 500 images. All images in CIFAR-10 and CIFAR-100 are sized in 32×32 .

OmniGlott [209] comprises 1,623 handwritten characters from 50 unique alphabets, partitioned into two distinct subsets. The background set includes 964 characters from 30 alphabets, while the evaluation set consists of 659 characters from 20 alphabets. Each character represents a distinct category with 20 example images.

TABLE A7
COMPARISON OF METHODS UNDER DIFFERENT DATA SPLITTING PROTOCOL. ALL METHODS BELOW EMPLOY THE RESNET [1] BACKBONE.

	Method	CIFAR10			CIFAR100			ImageNet-100		
		All	Old	New	All	Old	New	All	Old	New
Ind. $\mathcal{D}_C$ 50%	OpenLDN [65]	95.4	95.7	95.1	59.3	74.1	44.5	79.1	89.6	68.6
	LPS [92]	92.4	90.2	93.4	54.3	64.5	49.9	84.5	91.3	81.3
	OwMatch [93]	96.8	96.5	97.1	71.9	80.1	63.9	85.5	91.5	79.6
Ind. $\mathcal{D}_C$ 10%	TRSSL [66]	92.2	94.9	89.6	60.3	68.5	52.1	75.4	82.6	67.8
	TIDA [76]	93.8	94.2	93.4	65.3	73.3	56.6	83.4	77.6	71.2
Transductive	ORCA [26]	89.7	88.2	90.7	48.1	66.9	43.0	77.8	89.1	71.2
	OpenNCD [72]	90.1	88.4	90.6	49.3	69.7	43.4	81.6	90.0	77.5
	OpenCON [72]	90.4	89.3	91.1	52.8	69.1	47.8	83.8	90.6	80.8
	SORL [79]	93.5	94.0	92.5	56.1	68.2	52.0	-	-	-

TABLE A8
ADDITIONAL EXPERIMENTS COMPARISON FOR THE ABOVE SELECTED GCD METHODS ON CIFAR10&100 [190].

Method	CIFAR10 [190]			CIFAR100 [190]		
	All	Old	New	All	Old	New
TIDA [76]	92.9	93.3	92.6	72.1	73.4	71.5
SORL [79]	95.8	97.6	95.0	76.3	79.1	70.6
OwMatch [93]	96.6	98.0	95.9	79.8	81.2	79.1

TABLE A9
ADDITIONAL ILLUSTRATION OF DETAILED STATISTICS ABOUT DATASET SPLITTING FOR GCD.

Dataset	Balance	Labelled		Unlabelled	
		#Image	#Class	#Image	#Class
ImageNet-100 [189]	✓	31.9K	50	95.3K	100
ImageNet-1K [189]	✓	321K	500	960K	1000
CIFAR-10 [190]	✓	12.5K	5	37.5K	10
CIFAR-100-50 [190]	✓	12.5K	50	37.5K	100
CIFAR-100-20 [190]	✓	20.0K	80	30.0K	100
CUB [191]	✓	1.5K	100	4.5K	200
Cars [192]	✓	2.0K	98	6.1K	196
FGVC-Aircraft [193]	✓	1.7K	50	5.0K	100
Oxford-Pets [206]	✓	0.9K	19	2.7K	37
Herbarium 19 [207]	✗	8.9K	341	25.4K	683

SVHN [210] is a large collection of real-world digit images obtained from house number signs in Google Street View images. It includes over 70 thousand images in the training set and over 26 thousand in the test set.

B. Fine-Grained Datasets

CUB [191] is a widely used benchmark dataset for fine-grained visual classification tasks, particularly focused on bird species recognition. It contains 6,033 images of 200 bird species with semantic details for bird species classification.

Stanford-Cars [192] is a large-scale dataset designed for fine-grained vehicle classification tasks. It contains 16,185 images of 196 different car models, primarily spanning various makes, models, and years.

FGVC-Aircraft [193] is a fine-grained visual classification dataset focused on aircraft recognition. It contains 10,000 images spanning 100 different aircraft model variants, with each image labelled by its corresponding model.

Oxford-Pets [206] is a large, fine-grained dataset designed for pet image classification and segmentation tasks. It contains 7,349 images of 37 pet breeds, including both cats and dogs, with approximately 200 images per breed.

Herbarium19 [207] is a large-scale image collection focused on plant species identification, particularly for herbarium specimen recognition. It contains over 460,000 images representing more than 32,000 species, offering a rich and diverse set of labelled plant samples.

C. Domain Shift Datasets

DomainNet [211] is a large-scale, multi-domain dataset designed for domain adaptation and transfer learning tasks. It consists of approximately 600,000 images across six distinct domains: clipart, infograph, painting, quickdraw, real, and sketch, with 345 object categories common to all domains.

Office31 [212] is a widely used benchmark for domain adaptation in computer vision tasks. It consists of 4,110 images categorized into 31 classes spanning three distinct domains - Amazon, Webcam, and DSLR.

OfficeHome [213] is a popular benchmark for domain adaptation, consisting of 15,500 images categorized into 65 classes of everyday objects such as furniture, appliances, and office supplies. The dataset spans four different domains: Art, Clipart, Product, and Real-World.

VisDA [214] is a large-scale benchmark designed for evaluating domain adaptation algorithms, particularly in the context of image classification. It comprises over 280,000 images spanning 12 object categories, collected from two distinct domains: synthetic renderings and real-world photographs, thereby enabling rigorous assessment of cross-domain generalization.

D. Continual dataset

iNatIGCD [18] is specifically designed for Continual Category Discovery (CCD). This dataset is based on real-world data from the iNaturalist [215] platform and is structured

to reflect a more natural, incremental learning environment. It features four temporal stages, each corresponding to a different geographical region and time period, providing a realistic simulation of how species data grows over time. iNatIGCD includes challenges like geographical domain shifts and fine-grained categorization, making it significantly more challenging and suitable for evaluating models under realistic conditions compared to previous artificial dataset splits.

E. Datasets for Scene Understanding

PASCAL VOC [216] is a benchmark in computer vision, providing annotated images across 20 object categories—such as vehicles, animals, and household items—for tasks including object detection, classification, and segmentation.

PASCAL-5ⁱ [216] is a benchmark for few-shot semantic segmentation, constructed from the PASCAL VOC 2012 [216] dataset.

MS COCO is a large-scale benchmark, encompassing over 328,000 images with more than 2.5 million annotated instances across 91 object categories, designed to advance object recognition and scene understanding through detailed per-instance segmentations and contextual scene information.

COCO-20ⁱ [217] is a few-shot semantic segmentation benchmark derived from the MS COCO [217] dataset, where 80 object classes are evenly divided into four splits, each containing 20 classes.

F. Datasets for 3D

SemanticKITTI [218] is a large-scale benchmark for semantic scene understanding in autonomous driving, providing dense point-wise annotations for over 43,000 LiDAR scans across 22 sequences, covering a 360-degree field-of-view and encompassing 28 semantic classes, including distinctions between moving and static objects.

SemanticPOSS [219] is a 3D point cloud benchmark for semantic segmentation, comprising 2988 LiDAR scans collected in a campus environment.

ModelNet [220] is a large-scale benchmark for 3D object recognition, comprising over 127,000 synthetic CAD models across 662 object categories.

ShapeNetCore [221] is a richly annotated subset of the ShapeNet [221] repository, encompassing approximately 51,300 unique 3D models across 55 common object categories, each linked to synsets in WordNet, and featuring manually verified category and alignment annotations to support research in 3D object recognition and modeling.

ScanObjectNN [222] is a real-world benchmark for 3D point

cloud classification, comprising approximately 15,000 objects categorized into 15 classes with 2,902 unique object instances.

Co3D [223] is a large-scale collection designed to facilitate research in category-specific 3D reconstruction and novel view synthesis, comprising approximately 1.5 million frames from nearly 19,000 videos capturing objects from 50 MS COCO categories, each annotated with camera poses and ground truth 3D point clouds.

G. Datasets for Video

Kinetics-400 [224] is a large-scale benchmark for human action recognition, comprising approximately 306,245 short video clips sourced from YouTube, each lasting around 10 seconds, categorized into 400 distinct human action classes with at least 400 clips per class.

VGG-Sound [225] is a large-scale audio-visual collection comprising over 210,000 videos spanning 310 audio classes, curated from YouTube to ensure audio-visual correspondence.

UCF101 [226] is a comprehensive collection for human action recognition, comprising 13,320 realistic video clips sourced from YouTube, categorized into 101 distinct action classes, and encompassing a wide variety of activities such as human-object interactions, body-motion actions, human-human interactions, playing musical instruments, and sports.

H. Datasets for text

CLINC [227] is a benchmark for intent classification in task-oriented dialogue systems, comprising 22,500 in-domain utterances across 150 intent classes spanning 10 domains, along with 1,200 out-of-scope queries, designed to evaluate the performance of classifiers in identifying both supported intents and out-of-scope inputs.

HWU64 [228] is a multi-domain, fine-grained intent classification benchmark comprising 11,036 user-generated utterances spanning 64 intents across 21 domains related to personal assistant interactions.

WOS [229] is a document classification corpus comprising 46,985 scholarly articles categorized into 134 subject areas under 7 parent categories, designed to facilitate research in hierarchical text classification.

BANKING [230] is a fine-grained intent detection benchmark in the banking domain, comprising 13,083 customer service queries labelled across 77 distinct intent classes.

StackOverflow [231] is a comprehensive collection of user-generated content from the Stack Overflow platform, encompassing millions of programming-related questions, answers, comments, votes, tags, and badges.

TACRED [232] is a large-scale benchmark for relation extraction, comprising 106,264 sentences annotated with 41 relation types, constructed from newswire and web text

Re-TACRED [233] is a thoroughly re-annotated version of the original TACRED dataset, comprising over 91,000 sentences across 40 relation types, developed to address labelling inaccuracies and enhance the reliability of relation extraction model evaluations.

FewRel [234] is a large-scale benchmark for few-shot relation classification, comprising 70,000 sentences annotated with 100 distinct relations derived from Wikipedia.