LEARNING FROM VARIATION FOR OPEN-DOMAIN FEW-SHOT CLASS-INCREMENTAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-Shot Class-Incremental Learning (FSCIL) aims to continually acquire novel classes from limited examples while retaining previously learned knowledge. However, existing approaches typically assume that class increments occur within a fixed real-world domain, overlooking domain variations in open-world environments. In this work, we introduce Open-Domain Few-Shot Class-Incremental Learning (OD-FSCIL), where incremental samples may arise from heterogeneous domains whose composition is not predefined. Such domain heterogeneity makes intra-class variations difficult to capture from few-shot observations and complicates joint recognition of all seen classes. To address these challenges, we propose Learning from Variation (LeVa), which leverages the rich intra-class variations of data-rich base classes as prior knowledge for novel classes. LeVa comprises Variation Recovery (VR) and Variation Exploitation (VE). VR transfers historical intra-class variations to novel classes and uses predicted distributional statistics to retain those compatible with each novel class. VE exploits the recovered variations to model class-specific structures around text anchors, enabling more effective joint recognition of both old and new classes. To facilitate evaluation under open-domain conditions, we further construct Hybrid-miniImageNet, a new benchmark covering diverse visual domains. Extensive experiments on Hybrid-miniImageNet, DomainNet, miniImageNet, and CIFAR100 demonstrate the effectiveness of LeVa under both open-domain and standard FSCIL settings.

1 INTRODUCTION

Human learning is inherently incremental: new concepts are continually acquired from limited observations while prior experience is retained and reused (Zhou et al., 2024). Remarkably, humans can often recognize a newly learned concept across substantially different visual domains even when only a few examples are available. For instance, a child who learns the concept of a "dog" from a cartoon or plush toy may still recognize a real dog despite large visual differences. This motivates a natural question: can variation knowledge acquired from previously learned classes help recognize a new class from limited observations?

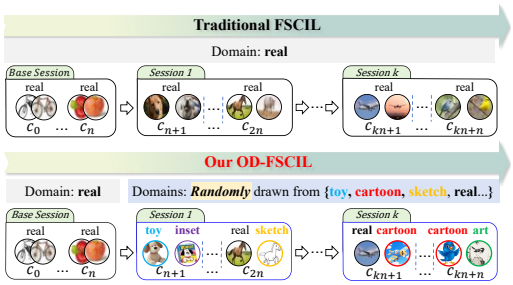


Figure 1: Comparison of traditional FSCIL and the proposed OD-FSCIL setting. Traditional FSCIL assumes a fixed visual domain, whereas OD-FSCIL allows incoming few-shot samples to have non-predefined domain compositions, including heterogeneous mixtures of source and base-unseen domains.

Few-Shot Class-Incremental Learning (FSCIL) shares a similar goal of continually acquiring novel classes from only a few examples while retaining previously learned knowledge. However, existing FSCIL methods (Song et al., 2023; Ahmed et al., 2024), as illustrated at the top of Figure 1, typically assume that class increments occur within a fixed visual domain, where both base and novel classes are represented by real-world images. While cross-domain learning has been explored under predefined domain shifts, open-world

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

class increments introduce a different challenge: the domains associated with incoming classes can be heterogeneous and unpredictable, rather than following a fixed source-to-target configuration.

In this work, we introduce Open-Domain Few-Shot Class-Incremental Learning (OD-FSCIL), as illustrated in Figure 1. OD-FSCIL retains the standard class-incremental protocol but removes prior assumptions on the domain composition accompanying class increments. The few-shot samples of a novel class may come from the source domain, domains unseen during base training, or heterogeneous mixtures of them, with the composition varying across classes and sessions. Thus, "open-domain" refers to the non-predefined domain composition of incoming novel-class observations rather than requiring every incoming domain to be globally unseen. This domain uncertainty introduces two key challenges. First, few-shot observations provide limited evidence of intra-class variations induced by heterogeneous domains, making it difficult to characterize a novel class beyond its observed examples. Second, domain heterogeneity complicates joint discrimination, as domain-induced variations can enlarge within-class differences while reducing inter-class separation. OD-FSCIL is also distinct from related incremental settings. Cross-domain FSCIL considers prescribed domain transitions, while hybrid incremental learning starts from multi-domain base data and jointly considers class and domain changes. In contrast, OD-FSCIL starts from a single-domain base session and introduces only novel classes thereafter, without predefined domain compositions for their few-shot observations, which may be heterogeneous even within the same class.

To address these challenges, we explore whether the rich intra-class variations observed in data-rich base classes can serve as prior knowledge for modeling novel classes from limited observations. Based on this intuition, we propose Learning from Variation (LeVa), which follows a recover-and-exploit strategy through two complementary modules: Variation Recovery (VR) and Variation Exploitation (VE). VR transfers historical intra-class variations to novel classes and predicts their distributional statistics from limited observations and class semantics, retaining variations compatible with each novel class. VE exploits the recovered variations to characterize class-specific visual structures around text anchors, using class semantics as a relatively stable reference for joint recognition under domain-induced visual variations. In summary, our main contributions are as follows.

- We introduce Open-Domain Few-Shot Class-Incremental Learning (OD-FSCIL), which removes the conventional assumption of a fixed or predefined domain configuration during class increments. Incoming few-shot samples may arise from the source domain, domains unseen during base training, or heterogeneous mixtures of them, making intra-class variation substantially more difficult to characterize from limited observations.
- We propose LeVa, a recover-and-exploit framework using historical intra-class variation as prior knowledge for novel-class modeling. Variation Recovery(VR) selects class-compatible transferred variations under distributional and semantic guidance, while Variation Exploitation(VE) exploits these variations to construct text-anchored class structure, using relatively stable class semantics as a reference for joint recognition under domain-induced visual shifts.
- We construct Hybrid-miniImageNet, a new benchmark that instantiates domain uncertainty in FSCIL through heterogeneous and non-predefined domain configurations. Extensive evaluations on Hybrid-miniImageNet, DomainNet, miniImageNet, and CIFAR100 demonstrate the effectiveness of LeVa under both open-domain and standard FSCIL settings.

2 Related Work

Our work lies at the intersection of two research directions: few-shot class-incremental learning, which concerns the continual acquisition of novel classes from limited examples, and few-shot representation learning, which investigates how informative representations of novel classes can be learned from scarce observations.

2.1 Few-Shot Class-Incremental Learning

Few-Shot Class-Incremental Learning (FSCIL) aims to continually acquire novel classes from only a few labeled examples while preserving the recognition of previously learned classes (Ahmed et al., 2024; Chen et al., 2025). The FSCIL setting was first introduced by (Tao et al., 2020), which

employs a neural gas network to preserve the topological structure of learned classes; for example, OrCo (Ahmed et al., 2024) combines feature orthogonality with contrastive learning to enlarge inter-class margins and mitigate interference during incremental learning. Subsequent methods have explored more discriminative representation spaces. More recently, VLM-based approaches (Chen et al., 2025) have exploited pretrained visual-semantic knowledge to improve representation and adaptation in FSCIL. Despite these advances, existing FSCIL methods are predominantly developed under a predefined domain configuration, typically assuming that base and incremental classes are drawn from the same real-image domain. FSHIL (Li et al., 2026) considers hybrid class and domain increments, allowing multi-domain base training and new samples from old classes in later sessions. In contrast, our OD-FSCIL considers domain uncertainty throughout incremental learning, starting from a single-domain base and introducing only novel classes whose associated domains can be heterogeneous and non-predefined rather than fixed.

2.2 Few-Shot Representation Learning

Learning informative representations from limited observations is a fundamental challenge in few-shot learning. A common strategy is to leverage knowledge acquired from data-rich base classes to improve the representation of novel classes. Distribution calibration (Yang et al., 2021), for example, transfers Gaussian statistics from related base classes to estimate novel-class distributions and synthesize additional features. LDC (Liu et al., 2023) learns a parameterized module to calibrate novel-class distributions initialized with a shared covariance, while higher-order calibration (Goswami et al., 2024) estimates novel-class covariance using statistics from related base classes. BiMC (Chen et al., 2025) observes that a global covariance dominated by base-class data can impair novel-class recognition and introduces a separate cross-modal nearest-neighbor metric for novel classes. Beyond statistical calibration, recent approaches have explored task-conditioned prototype generation through diffusion (Du et al., 2023) and semantic-attribute-based prototype calibration with dynamic structure matching (Wang et al., 2026). Despite their different formulations, these approaches seek to improve novel-class representations through statistical calibration, prototype refinement, or generative modeling. Our work approaches this problem from the perspective of intra-class variation: few-shot observations reveal only a small portion of how a novel class may vary, whereas data-rich base classes contain substantially richer variation experience. Rather than directly transferring such experience to novel classes, we treat it as prior knowledge and recover only those variations that are compatible with the limited evidence of each novel class. This enables novel-class representations to capture plausible variations beyond those directly observed from a few examples.

3 Methodology

3.1 Problem Formulation

We first formulate the standard FSCIL setting. FSCIL learns from a sequence of sessions $D^{(0)}, D^{(1)}, \ldots, D^{(T)}$, where the base session $D^{(0)}$ contains a set of classes $C^{(0)}$ with abundant labeled samples per class. Each subsequent session $D^{(i)}$ ($i > 0$) introduces a disjoint set of n novel classes $C^{(i)}$, with only k labeled samples available for each class, forming an n-way k-shot incremental task with $|D^{(i)}| = nk$. The class sets are mutually disjoint across sessions, i.e., $C^{(i)} \cap C^{(j)} = \emptyset$ for $i \neq j$. During session i, only the current training set $D^{(i)}$ is accessible; raw training samples from previous sessions are unavailable. After each session, the model is evaluated on all classes encountered so far, with the cumulative label space given by $\mathcal{C}^{(\leq i)} = \bigcup_{j=0}^{i} C^{(j)}$.

Unlike standard FSCIL, which typically assumes that base and incremental classes share the same real-image domain, OD-FSCIL relaxes this restriction by allowing incremental samples to exhibit heterogeneous and non-predefined domain compositions. Specifically, for session $D^{(i)}$, let $\mathcal{U}^{(i)}$ denote its domain set and $u(x) \in \mathcal{U}^{(i)}$ the domain associated with sample x. The base session $D^{(0)}$ is drawn from a single source domain u_0, whereas subsequent incremental sessions may contain samples from u_0, domains unseen during base training, or both. The domain composition is not specified in advance and may vary across classes and sessions; even the k samples of the same novel class may originate from different domains. Thus, "open-domain" refers to uncertainty in the domain composition accompanying class increments, rather than requiring a new domain to appear at every session.

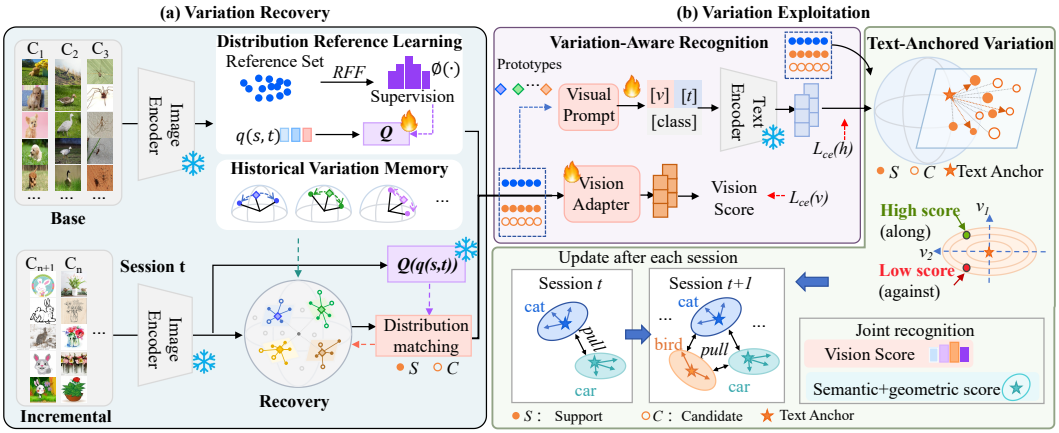


Figure 2: The architecture of LeVa. LeVa comprises Variation Recovery (VR) and Variation Exploitation (VE). VR leverages historical intra-class variations as prior knowledge to recover plausible novel-class variations beyond limited observations. VE exploits the recovered variations to construct text-anchored class structures, using relatively stable class semantics as a reference for joint recognition of all seen classes.

3.2 Overview of LeVa

As illustrated in Figure 2, LeVa builds on a pretrained vision–language model with frozen image and text encoders. During base training, abundant base-class data are used to learn how few-shot evidence relates to richer intra-class variation and to establish a reusable variation memory. During incremental learning, the limited samples of each novel class provide class-specific evidence for recovering plausible variations from the accumulated variation knowledge, which are subsequently exploited for joint recognition. The variation memory is progressively updated using newly observed support samples as new classes arrive, allowing variation knowledge to accumulate across sessions.

3.3 Variation Recovery

Under domain heterogeneity, the few-shot samples of a novel class reveal only a limited portion of its possible intra-class variation. Conventional class prototypes, which summarize these samples by their center, cannot recover such under-observed variation. To address this, we leverage the richer intra-class variations accumulated from data-rich base classes as prior knowledge. However, historical variations cannot be directly transferred to a novel class, since only a subset of them may be compatible with its class-specific characteristics. To identify such compatible variations, we learn from base classes to predict how a class would be distributed under richer observations given only its few-shot samples and semantics. The predicted distributional reference provides a class-specific criterion for selecting historical variations compatible with the novel class.

Learning a Distributional Reference. The limited samples of a novel class provide only a partial view of its intra-class variation, making it difficult to determine which historical variations are plausible for the class. We address this uncertainty by learning from data-rich base classes a distributional reference that relates sparse class evidence to the variation revealed by richer observations. Specifically, for each base class c_e, a few-shot subset S_e provides the sparse evidence, while a disjoint and larger set R_e represents its richer-observation distribution. Rather than reducing R_e to a single prototype or a fixed set of low-order moments, we characterize it using a kernel mean embedding, which provides a nonparametric representation of richer distributional characteristics. For efficient fixed-dimensional approximation, we apply a fixed PCA projection followed by a Random Fourier Feature (RFF) mapping $\phi(\cdot)$ learned from base data. The empirical kernel mean of a nonempty feature set A is:

$$\mu(A) = \frac{1}{|A|} \sum_{z \in A} \phi(z). \tag{1}$$

We learn this sparse-to-rich relationship on base classes by predicting the kernel mean of R_e from the limited evidence in S_e. Specifically, S_e is represented by a class-conditioned descriptor $q(S_e, t_{c_e})$ that summarizes its distributional characteristics, internal structure, and consistency with the fixed class text anchor t_{c_e}. A predictor Q is then learned to map this sparse evidence to the richer-observation kernel mean $\mu(R_e)$ by Eq. (2).

$$Q^\star = \arg\min_{Q \in \mathcal{H}_q} \Big\{ \sum_e \big\| Q\big(q(S_e, t_{c_e})\big) - \mu(R_e) \big\|_2^2 + \lambda_Q \|Q\|_{\mathcal{H}_q}^2 \Big\}, \tag{2}$$

where $\mathcal{H}_q$ denotes the corresponding vector-valued reproducing kernel Hilbert space and λ_Q controls regularization. Once learned from base classes, $Q^\star$ is fixed and transfers this sparse-to-rich relationship to novel classes. Given the few-shot observations S_c and text anchor t_c of a novel class c, we obtain

$$\widehat{\mu}_c = Q^\star(q(S_c, t_c)) \,. \tag{3}$$

The resulting $\widehat{\mu}_c$ serves as a class-specific distributional reference for determining which historical variations are plausible beyond those directly observed in S_c.

Recovering Variations from Historical Experience. Historical observations provide rich evidence of how visual representations can vary within a class, but directly transferring these features would also carry their source-class content. We therefore represent historical variation as deviations from the corresponding class center rather than as absolute features. Since normalized CLIP features lie on the unit hypersphere, these deviations are expressed in the tangent space at the class center, allowing parallel transport to relocate them between class centers while respecting the spherical feature geometry. Specifically, each historical variation is represented as (ν_b, r_b), where ν_b denotes the source-class center and r_b its tangent deviation. To express this deviation relative to a novel class c, we transport r_b from ν_b to the novel-class center ν_c and map it back to the feature space (Skopek et al., 2020) by Eq. (4), where $\mathrm{PT}_{\nu_b \to \nu_c}$ denotes parallel transport from the tangent space at ν_b to that at ν_c, and Exp_{ν_c} is the exponential map that maps the transported deviation back onto the unit sphere. The resulting $g_{b \to c}$ is a candidate variation expressed relative to the novel class.

$$g_{b \to c} = \mathrm{Exp}_{\nu_c}\left(\mathrm{PT}_{\nu_b \to \nu_c}(r_b)\right). \tag{4}$$

The transferred features represent possible variations of the novel class, but not all historical deviations are compatible with its class identity and expected distribution. We therefore recover variations under two constraints: semantic consistency with the novel class and distributional consistency with the richer-observation reference predicted from its limited evidence. Semantic consistency is enforced using the fixed text anchors of all seen classes, retaining candidates associated with the target class to form the candidate set G. Distributional consistency is measured by how well a set of recovered variations extends the observed class representation toward the predicted reference $\widehat{\mu}_c$:

$$\mathcal{E}_c(G) = \left\| \frac{\sum_{z \in S_c} \phi(z) + \sum_{g \in G} \phi(g)}{|S_c| + |G|} - \widehat{\mu}_c \right\|_2^2, \tag{5}$$

$$g_j = \arg\min_{g \in G \setminus G_{<j}} \mathcal{E}_c(G_{<j} \cup \{g\}) \,. \tag{6}$$

Here, g denotes a candidate feature, G excludes real supports, and $G_{<j} = \{g_1, \ldots, g_{j-1}\}$ contains previously selected variations, with $G_{<1} = \varnothing$. Directly optimizing over all subsets of $\mathcal{G}$ is combinatorial. We therefore progressively construct the recovered set G by adding at each step the candidate that minimizes the resulting distributional discrepancy. Each step selects the remaining candidate whose inclusion makes the joint kernel mean closest to $\widehat{\mu}_c$. After m steps, $G_c = \{g_1, \ldots, g_m\}$ forms the recovered variation set for class c. The recovered variations G_c extend the variation represented by the limited support observations, providing additional class-specific evidence for recognition.

3.4 Variation Exploitation

Variation Recovery provides additional evidence of how a novel class may vary beyond its limited observations. Variation Exploitation (VE) leverages these recovered variations to strengthen class representations and model class-specific variation with text semantics as relatively stable references, enabling more robust recognition under domain-induced visual variation.

Text-Anchored Variation Modeling. Cosine similarity measures the angular distance between a visual feature and a class text anchor, but ignores the direction of its deviation from the anchor. Consequently, features with the same angular distance receive identical similarity scores even when their deviations exhibit different consistency with the class-specific variation pattern. We therefore complement angular similarity with directional variation relative to each text anchor. Such variation is represented in the tangent space at the anchor, where its dominant directions and relative strengths are characterized through a weighted covariance matrix.

For each class c, the weighted covariance Σ_c characterizes its directional variation relative to the text anchor t_c, and we use the top k eigenvectors of Σ_c, denoted by $\{u_{c\ell}\}_{\ell=1}^{k}$, as the dominant variation directions of class c. Observed samples provide direct class evidence and are therefore assigned unit weights, whereas recovered variations may differ in their compatibility with the class and are weighted accordingly. Their weights $\{a_j\}_{j=1}^{m_c}$, where $m_c = |G_c|$, are determined by their collective agreement with the predicted distributional reference $\widehat{\mu}_c$:

$$\mathbf{a}_c^{\star} = \underset{\substack{a_j \geq 0 \\ \sum_{j=1}^{m_c} a_j = \text{const}}}{\arg\min} \left\| \frac{\sum_{z \in S_c} \phi(z) + \sum_{j=1}^{m_c} a_j \phi(g_j)}{|S_c| + \sum_{j=1}^{m_c} a_j} - \widehat{\mu}_c \right\|_2^2 . \tag{7}$$

Thus, the distributional reference that guides variation recovery also determines how strongly the recovered variations contribute to modeling class-specific directional variation. We incorporate the modeled variation directions into text-based matching through the directional score in Eq. (8). The score relaxes the angular penalty along dominant intra-class variation directions, so that deviations consistent with the class-specific variation pattern are better accommodated while deviations in other directions remain penalized.

$$g_c(z) = -\left[\arccos(z^\top t_c)^2 - \sum_{\ell=1}^{k} \left\langle \mathrm{Log}_{t_c}(z), u_{c\ell} \right\rangle^2 \right]. \tag{8}$$

Variation-Aware Recognition. The recovered variations are exploited in both visual and semantic representations for recognizing all seen classes. In the visual branch, a shared residual adapter refines visual features and class representations, producing the visual matching score $v_c(z)$. In the semantic branch, class visual prototypes p_c condition the text prompts to adapt class semantics to the observed visual evidence, yielding the semantic score $s_c(z)$. Beyond adapting the two representations, we make semantic matching aware of class-specific variation by incorporating the directional score derived above in Eq. (9), where γ_g controls the contribution of directional variation. Hence, the semantic branch evaluates a query by both its semantic compatibility with the class and its consistency with the class-specific variation pattern.

$$h_c(z) = s_c(z) + \gamma_g g_c(z), \tag{9}$$

During incremental learning, observed supports and recovered variations provide supervision for novel classes through both branches. The overall classification objective is described in Eq. (10), where $\mathrm{CE}(h)$ and $\mathrm{CE}(v)$ denote the classification losses from the semantic and visual branches, respectively. At inference, their normalized scores are combined to predict among all seen classes.

$$\mathcal{L} = \mathrm{CE}(h) + \mathrm{CE}(v). \tag{10}$$

4 Experiments

We evaluate our method against state-of-the-art approaches under both OD-FSCIL and standard FSCIL settings, using one open-domain dataset and three existing benchmarks. Ablation studies and visualizations further confirm the effectiveness of our method.

Datasets. For standard FSCIL, we evaluate LeVa on miniImageNet (Ravi & Larochelle, 2017) and CIFAR100 (Krizhevsky, 2009), each with 60 base and 40 novel classes under an eight-session, 5-way 5-shot incremental protocol. For OD-FSCIL, we use our proposed **Hybrid-miniImageNet** and **DomainNet** (Peng et al., 2019). Hybrid-miniImageNet preserves miniImageNet's base and

Table 1: **Comparison on Hybrid-miniImageNet. AHM** denotes the average harmonic mean. **FA** denotes the Top-1 accuracy in the final session. Best results are **bolded**; second-best are underlined.

Methods	Publications	Backbone	Base Acc	Session-wise Harmonic Mean (%) ↑								AHM	FA
				1	2	3	4	5	6	7	8		
SAVC	CVPR'23	ResNet	80.97	29.65	24.11	27.47	24.10	21.38	19.16	17.57	17.90	22.67	51.25
OrCo	CVPR'24	ResNet	83.93	55.78	45.15	48.42	43.58	38.77	35.62	33.42	34.27	41.88	52.81
Comp-FSCIL	ICML'24	ResNet	81.15	41.00	40.39	39.32	35.43	32.43	28.71	26.86	27.98	34.02	51.57
YourSelf	ECCV'24	ResNet	84.00	47.53	35.38	40.15	24.57	31.16	25.90	25.99	25.69	32.05	54.20
CLOSER	ECCV'24	ResNet	77.43	22.53	20.90	26.82	23.93	21.30	19.00	17.56	16.79	21.10	48.85
LRT	TPAMI'25	ResNet	90.12	47.60	38.33	39.83	36.15	30.30	30.43	28.26	26.98	34.74	59.22
ADBS	AAAI'25	ResNet	81.58	18.37	26.13	31.23	26.71	24.55	22.32	20.59	22.07	24.00	49.70
ConCM	ICLR'26	ResNet	83.47	56.98	47.06	48.19	43.08	39.32	35.26	32.98	32.55	41.93	53.99
BiMC†	CVPR'25	CLIP	95.00	91.88	88.93	89.13	87.09	87.31	85.42	85.50	85.98	87.66	87.99
LeVa (Ours)	-	CLIP	**96.95**	**95.76**	**93.92**	**93.13**	**90.50**	**90.49**	**88.56**	**88.52**	**89.04**	**91.24**	**90.77**

Table 2: **Comparison on DomainNet. AHM** denotes the average harmonic mean. **FA** denotes the Top-1 accuracy in the final session. Results are averaged across the three subsets and reported in percentages. Best results are **bolded**; second-best are underlined.

Methods	Publications	Backbone	Base Acc	Session-wise Harmonic Mean (%) ↑								AHM	FA
				1	2	3	4	5	6	7	8		
SAVC	CVPR'23	ResNet	79.09	35.78	31.02	28.43	27.76	26.11	24.73	23.12	23.06	27.50	49.07
CLOSER	ECCV'24	ResNet	76.19	27.70	23.60	23.02	21.44	20.09	18.81	17.97	17.61	21.28	45.69
Comp-FSCIL	ICML'24	ResNet	76.04	38.53	34.31	34.13	31.46	30.21	27.98	26.11	26.12	31.11	47.18
LRT	TPAMI'25	ResNet	76.21	27.67	24.88	23.93	26.04	23.92	21.75	20.17	19.95	23.54	40.45
ADBS	AAAI'25	ResNet	78.33	30.68	28.96	28.57	28.03	27.42	28.07	27.43	27.93	28.39	49.21
ConCM	ICLR'26	ResNet	80.13	63.76	57.04	53.44	48.90	45.58	43.43	40.90	40.11	49.15	49.65
BiMC†	CVPR'25	CLIP	70.72	70.15	68.92	67.27	66.13	64.83	64.82	64.28	64.08	66.31	64.24
LeVa (Ours)	-	CLIP	**87.01**	**78.46**	**76.08**	**73.74**	**72.25**	**70.98**	**71.13**	**70.15**	**70.22**	**72.88**	**73.38**

validation sets, class split, and incremental protocol. Each novel class spans four domains (real, painting, sketch, and clipart), with 150 images per domain and 600 per class. Further details on dataset construction are in the Appendix. DomainNet contains 345 classes across six domains: Clipart, Real, Infograph, Sketch, Painting, and Quickdraw. We use only the Real domain for base training, while incremental domain compositions are not predefined; even the few-shot supports of one novel class may span multiple domains. This provides a controlled OD-FSCIL evaluation using a finite collection of visual domains. We randomly construct three 100-class subsets, each with 60 base and 40 novel classes, and evaluate them using the same incremental protocol.

Implementation details. We adopt CLIP with a ViT-B/16 image encoder as the backbone. The visual adapter applies a linear projection, LeakyReLU, and ℓ_2 normalization, followed by residual addition to the input and another normalization. The prototype-conditioned prompt generator is a four-layer MLP with dropout after the first linear layer and ReLU before the final projection. We use SGD with an initial learning rate of 0.001 and cosine annealing. We set $\gamma_g = 0.3$ based on the sensitivity analysis in Appendix Figure 6, which shows stable performance across a broad range of values. All experiments run on an NVIDIA GeForce RTX 3090 GPU.

4.1 Experimental Results

Details of Compared Methods. We compare LeVa with state-of-the-art FSCIL methods based on ResNet (He et al., 2016) or CLIP (Radford et al., 2021), including SAVC (Song et al., 2023), OrCo (Ahmed et al., 2024), Comp-FSCIL (Zou et al., 2024), YourSelf (Tang et al., 2024), CLOSER (Oh et al., 2024), LRT (Zhao et al., 2025), ADBS (Li et al., 2025), BiMC (Chen et al., 2025), and ConCM (Wang et al., 2026). Among these methods, LRT distills knowledge from CLIP, whereas BiMC adopts a training-free CLIP-based approach that incorporates GPT-generated auxiliary textual descriptions. For BiMC, we use its best-performing variant, BiMC†, in our comparisons.

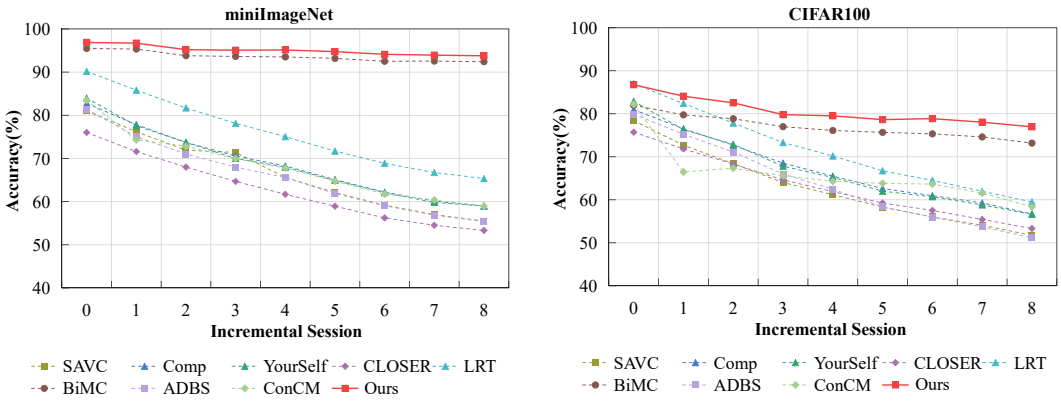


Figure 3: Classification accuracy trends of LeVa compared with SOTA methods across all sessions on CIFAR100 and miniImageNet under the standard FSCIL setting.

Table 3: **Ablation of VE.** TAVM: text-anchored variation modeling; VM: visual matching; PCP: prototype-conditioned prompt.

TAVM	VM	PCP	$Avg_{Novel}\uparrow$	$Avg_{Base}\uparrow$	AHM ↑
Zero-Shot CLIP			81.35	90.36	85.51
✓	✓		84.27	95.56	89.44
✓		✓	<u>85.86</u>	95.32	90.21
	✓	✓	85.06	**96.52**	<u>90.34</u>
✓	✓	✓	**86.78**	<u>96.36</u>	**91.24**

Table 4: **Alternative Designs for Variation Recovery.** Best and second-best results are shown in **bold** and <u>underlined</u>, respectively.

Variant	$Avg_{Novel}\uparrow$	$Avg_{Base}\uparrow$	$Avg_{HM}\uparrow$
Data Augmentation	83.07	<u>96.54</u>	89.21
Base Cov.	<u>85.19</u>	**96.59**	<u>90.46</u>
Random	85.01	96.51	90.33
LeVa	**86.78**	96.36	**91.24**

Comparisons under the OD-FSCIL setting. We evaluate LeVa on Hybrid-miniImageNet and three DomainNet subsets using base-session accuracy, session-wise harmonic mean (HM), average HM (AHM), and final-session Top-1 accuracy (FA). DomainNet results are averaged across the three subsets. *Results on Hybrid-miniImageNet.* As shown in Table 1, LeVa achieves the highest base-session accuracy, AHM, and FA of 96.95%, 91.24%, and 90.77%, respectively. It surpasses BiMC† by 3.58 and 2.78 percentage points in AHM and FA, with HM gains of 3.02–4.99 points across all eight incremental sessions. *Results on DomainNet.* As shown in Table 2, LeVa achieves the highest base-session accuracy, AHM, and FA of 87.01%, 72.88%, and 73.38%, respectively. It outperforms BiMC† by 6.57 and 9.14 percentage points in AHM and FA, with HM gains of 5.87–8.31 points across all incremental sessions. These results support LeVa's effectiveness in joint base- and novel-class recognition under domain uncertainty, consistent with its recover-and-exploit design: VR enriches sparse novel-class evidence with historical intra-class variations, while VE incorporates this evidence into text-anchored class structures.

Comparisons under the standard FSCIL setting. We further evaluate LeVa on CIFAR100 and miniImageNet. As shown in Figure 3, LeVa maintains strong performance across incremental sessions on both datasets and consistently outperforms the compared methods.

From Single-Domain to Open-Domain FSCIL. The paired miniImageNet and Hybrid-miniImageNet results further examine the transition from conventional single-domain FSCIL to OD-FSCIL under the same class-incremental protocol. LeVa's final accuracy decreases from 93.78% (Fig. 3) to 90.77% (Tab. 1), compared with 92.4%(Fig. 3) to 87.99% (Tab. 1) for BiMC. This suggests stronger robustness to heterogeneous domain compositions.

4.2 Ablation Study

Ablation of Variation Exploitation. As shown in Table 3, we include zero-shot CLIP with the same pretrained ViT-B/16 backbone as a reference. LeVa surpasses it by 6.00%, 5.43%, and 5.73% in Avg_{Base}, Avg_{Novel}, and AHM, respectively. Removing TAVM reduces Avg_{Novel} and AHM

by 1.72% and 0.90%, respectively, despite a slight increase in Avg_{Base}. This suggests that modeling class-specific variations around text anchors complements visual matching and prototype-conditioned prompting. Removing VM or PCP decreases AHM by 1.03% or 1.80%, respectively, while also reducing both Avg_{Base} and Avg_{Novel}. The full model achieves the highest Avg_{Novel} and AHM, supporting the complementary roles of the three components in joint recognition.

Alternative Designs for Variation Recovery. Table 4 compares LeVa with three alternatives: image augmentation (Basic, RandAugment, and AugMix), Base Cov. means sampling from the nearest base-class covariance, and random transfer of historical variations. LeVa achieves the highest average novel-class accuracy and AHM of 86.78% and 91.24%, outperforming Base Cov. Its AHM also exceeds image augmentation and random variation transfer by 2.03 and 0.91 points. These results demonstrate that VR improves novel-class recognition while maintaining competitive base-class performance. More ablations are provided in the appendix.

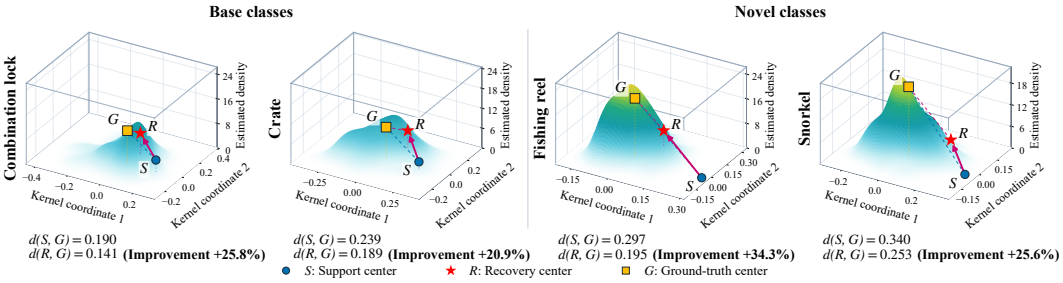


Figure 4: Visualization of Variation recovery on Base and Novel classes. S, R, and G denote the support center, predicted recovery center, and empirical ground-truth center, respectively. Distances are computed before projection; Improvement denotes the relative reduction from $d(S, G)$ to $d(R, G)$.

4.3 Visualization of Variation Recovery

To assess whether estimation learned from base classes transfers to novel classes, Figure 4 compares the support center S, recovery center R, and empirical ground-truth center G. Here, S is the empirical kernel mean of supports, R is the kernel mean predicted by the base-trained predictor, and G is computed from reference samples for base classes and test samples for novel classes. Distances $d(S, G)$ and $d(R, G)$ are computed in the kernel-feature space, and Improvement denotes the relative distance reduction over S. The surfaces show projected reference/test sample densities, while arrows indicate the correction from S to R. In these examples, R is closer to G than S, indicating reduced estimation error from sparse supports. The novel-class improvements provide the evidence that the base-trained predictor produces distributional references better aligned with the empirical statistics of unseen classes.

5 Conclusion

In this work, we introduced Open-Domain Few-Shot Class-Incremental Learning (OD-FSCIL), which extends FSCIL to class increments with heterogeneous and non-predefined domain configurations. To address the limited intra-class variation observable from few-shot samples under such conditions, we proposed Learning from Variation (LeVa), which leverages historical intra-class variation as prior knowledge for novel-class modeling. LeVa recovers plausible variations for novel classes under a predicted distributional reference and further exploits them to model class-specific variation around relatively stable text semantics, facilitating joint recognition of old and new classes. We also constructed Hybrid-miniImageNet to facilitate evaluation under open-domain conditions. Extensive experiments on Hybrid-miniImageNet, DomainNet, miniImageNet, and CIFAR100 demonstrate the effectiveness of LeVa under both open-domain and standard FSCIL settings.

AI USE STATEMENT

In this work, we used generative AI tools to generate auxiliary class descriptions required by the BiMC baseline. Additionally, we used generative AI tools for language polishing of the manuscript. We have reviewed all AI-assisted content, including the generated class descriptions and the revised text. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Noor Ahmed, Anna Kukleva, and Bernt Schiele. Orco: Towards better generalization via orthogonality and contrast for few-shot class-incremental learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 28762–28771. IEEE, 2024. doi: 10.1109/CVPR52733.2024.02717.

Yiyang Chen, Tianyu Ding, Lei Wang, Jing Huo, Yang Gao, and Wenbin Li. Enhancing few-shot class-incremental learning via training-free bi-level modality calibration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 9881–9890. Computer Vision Foundation / IEEE, 2025. doi: 10.1109/CVPR52734.2025.00923.

Victor Guilherme Turrisi da Costa, Enrico Fini, Moin Nabi, Nicu Sebe, and Elisa Ricci. solo-learn: A library of self-supervised methods for visual representation learning. *Journal of Machine Learning Research*, 23(56):1–6, 2022. URL http://jmlr.org/papers/v23/21-1155.html.

Yingjun Du, Zehao Xiao, Shengcai Liao, and Cees Snoek. Protodiff: Learning to learn prototypical networks by task-guided diffusion. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 46304–46322. Curran Associates, Inc., 2023. doi: 10.52202/075280-2008. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/911dd89c81efc624c4e1c39381179505-Paper-Conference.pdf.

Dipam Goswami, Bartłomiej Twardowski, and Joost Van De Weijer. Calibrating higher-order statistics for few-shot class-incremental learning with pre-trained vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 4075–4084, June 2024.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, Dawn Song, Jacob Steinhardt, and Justin Gilmer. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8320–8329, 2021. doi: 10.1109/ICCV48922.2021.00823.

Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009. URL https://api.semanticscholar.org/CorpusID:18268744.

Linhao Li, Yongzhang Tan, Siyuan Yang, Hao Cheng, Yongfeng Dong, and Liang Yang. Adaptive decision boundary for few-shot class-incremental learning. In Toby Walsh, Julie Shah, and Zico Kolter (eds.), *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, pp. 18359–18367. AAAI Press, 2025. doi: 10.1609/AAAI.V39I17.34020.

Yan Li, Yuzhu Shi, Kan Zhou, Shu Zhang, Diqi He, Dingwen Zhang, and Junwei Han. Few-shot hybrid incremental learning:continually learning under data scarcity and task uncertainty. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 32334–32344, June 2026.

Binghao Liu, Boyu Yang, Lingxi Xie, Ren Wang, Qi Tian, and Qixiang Ye. Learnable distribution calibration for few-shot class-incremental learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):12699–12706, 2023. doi: 10.1109/TPAMI.2023.3273291.

Junghun Oh, Sungyong Baik, and Kyoung Mu Lee. CLOSER: towards better representation learning for few-shot class-incremental learning. In Ales Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part XLIX*, volume 15107 of *Lecture Notes in Computer Science*, pp. 18–35. Springer, 2024. doi: 10.1007/978-3-031-72967-6_2.

Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pp. 1406–1415. IEEE, 2019. doi: 10.1109/ICCV.2019.00149.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021.

Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

Ondrej Skopek, Octavian-Eugen Ganea, and Gary Bécigneul. Mixed-curvature variational autoencoders. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020. URL https://openreview.net/forum?id=S1g6xeSKDS.

Zeyin Song, Yifan Zhao, Yujun Shi, Peixi Peng, Li Yuan, and Yonghong Tian. Learning with fantasy: Semantic-aware virtual contrastive constraint for few-shot class-incremental learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pp. 24183–24192. IEEE, 2023. doi: 10.1109/CVPR52729.2023.02316.

Yu-Ming Tang, Yi-Xing Peng, Jingke Meng, and Wei-Shi Zheng. Rethinking few-shot class-incremental learning: Learning from yourself. In Ales Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part LXI*, volume 15119 of *Lecture Notes in Computer Science*, pp. 108–128. Springer, 2024. doi: 10.1007/978-3-031-73030-6_7.

Xiaoyu Tao, Xiaopeng Hong, Xinyuan Chang, Songlin Dong, Xing Wei, and Yihong Gong. Few-shot class-incremental learning. *CoRR*, abs/2004.10956, 2020.

Haohan Wang, Songwei Ge, Zachary Lipton, and Eric P Xing. Learning robust global representations by penalizing local predictive power. In *Advances in Neural Information Processing Systems*, pp. 10506–10518, 2019.

Qinzhe Wang, Zixuan Chen, Keke Huang, Xiu Su, Chunhua Yang, and Chang Xu. Consistency-driven calibration and matching for few-shot class incremental learning. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 1255–1279, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/030bfd0d78556d4cf3301cb61866600b-Paper-Conference.pdf.

Xinrui Wang and Jinze Yu. Learning to cartoonize using white-box cartoon representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

Shuo Yang, Lu Liu, and Min Xu. Free Lunch for Few-Shot Learning: Distribution Calibration. In *International Conference on Learning Representations*, 2021. URL https://mlanthology.org/iclr/2021/yang2021iclr-free/.

Yifan Zhao, Jia Li, Zeyin Song, and Yonghong Tian. Language-inspired relation transfer for few-shot class-incremental learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(2):1089–1102, 2025. doi: 10.1109/TPAMI.2024.3492328.

Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Class-incremental learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):9851–9873, 2024. doi: 10.1109/TPAMI.2024.3429383.

Yixiong Zou, Shanghang Zhang, Haichen Zhou, Yuhua Li, and Ruixuan Li. Compositional few-shot class-incremental learning. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

APPENDIX

This supplementary material includes the following:

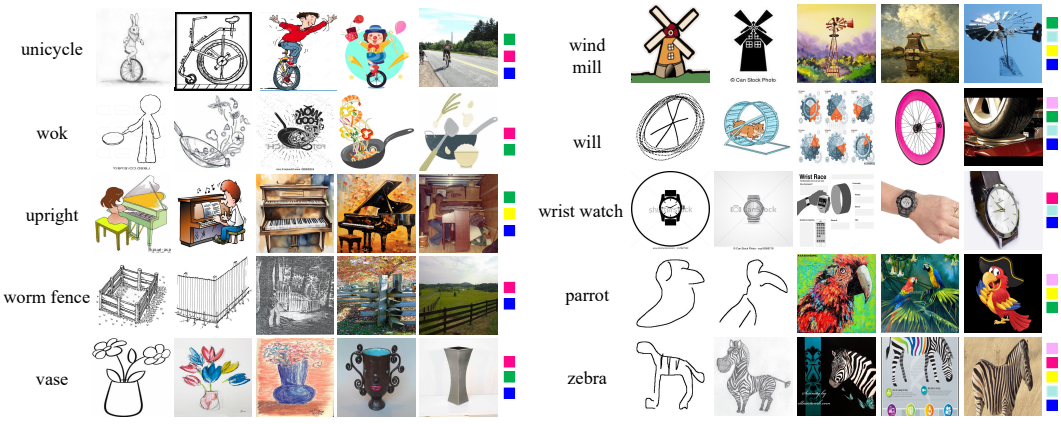
- Section A summarizes comprehensive descriptions of the datasets used in our experiments.
- Section B outlines the training procedures in greater detail.
- Section C shows more ablation studies.
- Section D elaborates on additional experimental results.

A DATASETS

We conduct a comprehensive evaluation of the proposed LeVa on 4 datasets: *miniImageNet* (Ravi & Larochelle, 2017), *CIFAR100* (Krizhevsky, 2009), *DomainNet* (Peng et al., 2019), and a multi-domain extended version of miniImageNet, *Hybrid-miniImageNet*. The stylized samples in Hybrid-miniImageNet are sourced from style-transfer datasets, as illustrated in Figure 5 (a).

1) For standard FSCIL, we evaluate on miniImageNet and CIFAR100, following the common 8-session, 5-way 5-shot protocol with 60 base and 40 novel classes.

2) For OD-FSCIL, we use the proposed Hybrid miniImageNet and standard DomainNet. Hybrid-miniImageNet, built from diverse sources, covers various visual domains (real, sketch, cartoon, and painting). Real images are randomly sampled from the original miniImageNet novel set, while stylized images are collected from style transfer datasets (Hendrycks et al., 2021; Wang et al., 2019; Wang & Yu, 2020). Hybrid-miniImageNet follows the same class split and incremental protocol as miniImageNet while introducing heterogeneous domains for novel classes, providing paired single-domain and open-domain settings. DomainNet contains 345 classes across six domains: Clipart, Real, Infograph, Sketch, Painting and Quickdraw (see Figure 5 (b)). Only real images are used for base training, with incremental phases following the Hybrid-miniImageNet protocol. DomainNet is randomly split into 3 subsets of 100 classes (each 60 base, 40 novel), following the standard 8-session, 5-way 5-shot setting. Final results are averaged over the 3 splits. The detailed splits for the three DomainNet subsets are as shown in Table 5.



(a) *Hybrid-miniImageNet* novel set: randomly sampled examples.

(b) *DomainNet* novel set: randomly sampled examples.

Figure 5: Randomly sampled examples from the novel sets of (a) *Hybrid-miniImageNet* and (b) *DomainNet*, where colors denote different domains and the rightmost column lists the domains contained in each category.

Comparison Methods. We compare LeVa against state-of-the-art FSCIL methods, including SAVC (Song et al., 2023), OrCo (Ahmed et al., 2024), Comp-FSCIL (Zou et al., 2024), Yourself (Tang et al., 2024), CLOSER (Oh et al., 2024), ADBS (Li et al., 2025), LRT (Zhao et al., 2025), BiMC (Chen et al., 2025) and ConCM (Wang et al., 2026). BiMC utilizes CLIP as its backbone. BiMC

Table 5: The detailed splits for the three DomainNet subsets: D_{sub1}, D_{sub2}, D_{sub3}.

Split	Categories
	D_{sub1}
Base	airplane, angel, anvil, arm, asparagus, backpack, barn, binoculars, bird, blackberry, bracelet, brain, bread, bridge, bus, butterfly, camel, campfire, canoe, carrot, cat, cell phone, compass, couch, crab, crayon, crown, diamond, diving board, duck, dumbbell, eyeglasses, finger, firetruck, fish, flamingo, flying saucer, foot, giraffe, hamburger, hand, helicopter, hospital, hourglass, hurricane, jacket, kangaroo, key, knee, lighthouse, lightning, lollipop, map, marker, microphone, microwave, mosquito, mountain, mouth, oven
Novel	owl, paint can, palm tree, panda, parachute, parrot, peanut, penguin, pickup truck, picture frame, police car, pool, potato, sandwich, scissors, scorpion, shoe, skateboard, snake, spider, square, squiggle, squirrel, stethoscope, streetlight, suitcase, sun, swan, tiger, toe, tornado, The Eiffel Tower, toothbrush, triangle, vase, violin, washing machine, watermelon, wheel, zigzag
	D_{sub2}
Base	ambulance, animal migration, ant, anvil, apple, asparagus, bandage, barn, bathtub, beach, bed, bench, birthdaycake, blackberry, bracelet, broom, bulldozer, cactus, camel, candle, canoe, car, castle, cello, circle, cloud, compass, cookie, crocodile, door, dragon, elbow, eraser, face, fan, fence, finger, flashlight, flipflops, fryingpan, golf club, headphones, helicopter, hospital, hot dog, hot tub, hourglass, hurricane, keyboard, lantern, leaf, line, lion, map, matches, megaphone, mosquito, nail, necklace, ocean
Novel	paintbrush, panda, penguin, picture frame, pineapple, potato, purse, rabbit, rain, rainbow, remote control, saxophone, see saw, shoe, sink, skateboard, snowman, sock, spider, squiggle, stitches, strawberry, swan, swing set, syringe, teddy bear, television, tennis racquet, tent, The Great Wall of China, toe, toothpaste, tractor, train, tree, underwear, violin, washing machine, watermelon, whale
	D_{sub3}
Base	animal migration, beard, binoculars, bird, birthday cake, blackberry, blueberry, bread, broom, bush, butterfly, camouflage, castle, cell phone, chandelier, computer, cookie, diamond, dishwasher, duck, elbow, elephant, envelope, eye, eyeglasses, face, fan, feather, flying saucer, foot, garden hose, goatee, grass, hamburger, harp, headphones, horse, jacket, ladder, leaf, leg, light bulb, line, lion, lipstick, lobster, mosquito, motorbike, moustache, owl, paint can, paintbrush, palm tree, panda, parrot, passport, peas, penguin, pillow, pizza
Novel	police car, popsicle, postcard, remote control, rifle, rollerskates, saw, saxophone, shoe, shovel, sink, smiley face, snail, snorkel, snowflake, spreadsheet, squiggle, steak, stop sign, string bean, sweater, swing set, telephone, television, tent, The Mona Lisa, toe, toothbrush, tornado, tractor, train, tree, truck, umbrella, watermelon, wheel, windmill, wine glass, wristwatch, zebra

is a training-free method that improves performance through GPT-generated auxiliary textual descriptions. For ADBS, we use the authors' released implementation, while for BiMC, we adopt the best-performing BiMC† variant.

B TRAINING DETAILS

For the *Hybrid-miniImageNet* dataset, we directly use the parameter settings based on the miniImageNet dataset provided in the original papers, as these have been verified by the authors to be the best-performing configurations.

For the *DomainNet* dataset: **CLOSER** (Oh et al., 2024): The learning rate is set to 0.005, and the model is trained for 200 epochs. **BiMC** (Chen et al., 2025): We report the results of its best-performing variant, BiMC†. As BiMC relies on GPT-generated auxiliary descriptions, we follow the original paper and official code to generate DomainNet class texts (same prompts and settings), and manually verify that the outputs are semantically valid and class-consistent. The training settings for **SAVC** (Song et al., 2023), **OrCo** (Ahmed et al., 2024), **Comp-FSCIL** (Zou et al., 2024), **ADBS** (Li et al., 2025), Yourself (Tang et al., 2024), and ConCM (Wang et al., 2026) follow their respective

configurations used for miniImageNet. For ConCM (Wang et al., 2026), we followed the instructions in its README and used solo-learn (da Costa et al., 2022) to pretrain a separate model for each of the DomainNet1, DomainNet2, and DomainNet3 datasets. We then used these models for the corresponding experiments.

C Additional Ablation Studies

C.1 Additional Component Ablations

As shown in Table 6, enabling PCP, VM, or TAVM individually improves AHM by 3.36%, 2.92%, and 2.47%, respectively, over the baseline without these components. PCP achieves the highest novel-class accuracy and AHM among the individual variants, while VM yields the highest base-class accuracy. All three components improve both base- and novel-class accuracies, indicating that each contributes independently to joint recognition.

Table 6: **Individual Contributions of Variation Exploitation Components.** TAVM: text-anchored variation modeling; VM: visual matching; PCP: prototype-conditioned prompting. ✓ denotes an enabled component; blank entries indicate disabled components.

TAVM	VM	PCP	Avg_{Novel} ↑	Avg_{Base} ↑	AHM ↑
Zero-Shot CLIP			81.35	90.36	85.51
		✓	**83.29**	<u>95.54</u>	**88.87**
	✓		82.15	**95.96**	<u>88.43</u>
✓			<u>82.29</u>	94.84	87.98

C.2 Sensitivity to the γ_g.

We examine the influence of the geometric weight γ_g on novel-class recognition and balanced incremental performance. Compared with $\gamma_g = 0$, setting $\gamma_g = 0.3$ improves Avg. Novel accuracy and AHM by 1.72 and 0.90 percentage points, respectively. Larger weights do not consistently improve performance, suggesting that a moderate geometric contribution helps complement semantic matching.

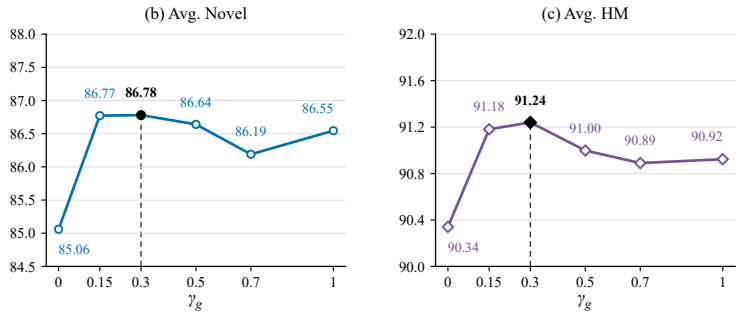


Figure 6: Sensitivity to the geometric weight γ_g on Hybrid miniImageNet. We report average Novel accuracy and average harmonic mean (AHM) over S1-S8. Bold values and vertical dashed lines mark the highest value among the displayed settings in each panel.

D Additional Results on Overall Accuracy

Number of Recovered Variations. As shown in Table 7, recovering five variations improves novel-class accuracy and AHM by 3.74 and 2.06 percentage points over $m = 0$, while base-class

accuracy decreases by only 0.16 percentage points. These results support the motivation behind VR: historical intra-class variations provide useful evidence beyond scarce novel-class observations, improving novel-class recognition while largely preserving base-class performance. Increasing the number of recovered variations yields no consistent further improvement, with $m = 5$ and $m = 15$ achieving the same highest AHM. Thus, within the evaluated range, a small set of recovered variations is sufficient to obtain these gains, while additional variations are not necessarily beneficial.

Table 7: **Effect of the Number of Recovered Variations.** $m = |G_c|$ denotes the number of recovered variations retained per novel class.

m	Avg_{Novel} ↑	Avg_{Base} ↑	AHM ↑
0	83.04	**96.52**	89.18
5	<u>86.78</u>	<u>96.36</u>	**91.24**
10	86.08	96.26	90.80
15	**86.85**	96.28	**91.24**
20	86.42	96.31	91.01

Overall classification accuracy. Tables 8 and 9 report per-session overall classification accuracy on Hybrid-miniImageNet and the three DomainNet subsets. Average accuracy is computed over sessions 0–8, while PD measures the accuracy drop from session 0 to session 8.

On Hybrid-miniImageNet, LeVa achieves the highest accuracy in every session among the reported results. Its average and final-session accuracies reach 93.82% and 90.77%, exceeding BiMC by 2.51 and 2.78 percentage points, respectively. LeVa also achieves the lowest PD of 6.18 percentage points, compared with 7.01 for BiMC, indicating a smaller decline in overall accuracy as new classes are introduced.

On the three DomainNet subsets, LeVa likewise achieves the highest reported accuracy in every session. Its average accuracies reach 79.63%, 78.56%, and 77.77%, exceeding BiMC† by 10.86, 12.08, and 11.96 percentage points, respectively. Its final-session accuracies are 74.05%, 73.21%, and 72.87%, corresponding to gains of 9.39, 8.84, and 9.18 percentage points over BiMC†. Although BiMC† has smaller PD values on these subsets, LeVa maintains higher absolute accuracy throughout the evaluated sessions. Since PD measures the decline relative to each method's own initial accuracy, it should be interpreted alongside average and final-session accuracies.

These results support the effectiveness of LeVa for joint recognition under domain heterogeneity. They are consistent with its recover-and-exploit design: Variation Recovery enriches sparse novel-class observations with compatible historical variations, while Variation Exploitation uses the enriched evidence for text-anchored variation modeling and visual–semantic adaptation.

Table 8: **Comparison on Hybrid-miniImageNet under the OD-FSCIL setting. Avg.** denotes the average accuracy over sessions 0-8. **PD** denotes the accuracy drop from session 0-8. Best results are **bolded**; second-best are underlined.

Methods	Publications	Accuracy (%) in each session ↑									Avg. (↑)	PD (↓)
		0	1	2	3	4	5	6	7	8		
SAVC	CVPR'23	80.97	75.15	70.49	66.99	63.16	59.51	56.33	53.47	51.25	64.15	29.72
OrCo	CVPR'24	83.93	76.57	71.77	68.61	65.08	60.99	57.68	54.56	52.81	65.78	31.12
Comp-FSCIL	ICML'24	81.15	75.66	71.11	66.89	62.79	59.48	56.17	53.26	51.57	64.23	29.58
YourSelf	ECCV'24	84.00	77.60	73.00	70.30	66.30	62.50	58.80	56.40	54.20	67.01	29.80
CLOSER	ECCV'24	77.43	71.97	67.30	64.27	60.58	57.00	53.86	51.19	48.85	61.38	28.58
LRT	TPAMI'25	90.12	85.29	80.04	76.20	72.18	67.88	65.14	61.88	59.22	73.11	30.90
BiMC	CVPR'25	95.00	94.49	93.17	92.53	91.07	90.51	88.79	88.21	87.99	91.31	7.01
ADBS	AAAI'25	81.58	74.31	69.51	66.11	62.05	57.79	54.57	51.88	49.70	63.06	31.88
ConCM	ICLR'26	83.47	75.14	71.19	69.12	65.96	62.58	59.17	56.25	53.99	66.32	29.48
LeVa (Ours)	-	**96.95**	**96.77**	**95.76**	**95.19**	**93.60**	**93.07**	**91.39**	**90.84**	**90.77**	**93.82**	**6.18**

Table 9: **Classification accuracy (%) of LeVa versus SOTA methods on D_{sub1}, D_{sub2}, and D_{sub3} under the OD-FSCIL setting. aAcc** denotes the average accuracy over sessions 0-8, and **PD** denotes the accuracy drop from session 0-8. Best and second-best available results are shown in **bold** and underlined, respectively.

Methods	Publications	Accuracy (%) in each session ↑									aAcc (↑)	PD (↓)
		0	1	2	3	4	5	6	7	8		
D_{sub1}												
SAVC	CVPR'23	79.75	73.85	69.05	65.11	60.80	57.22	54.00	51.51	49.22	62.28	30.53
Comp-FSCIL	ICML'24	76.31	71.42	66.32	62.25	57.71	54.15	51.06	48.81	46.71	59.42	29.60
CLOSER	ECCV'24	77.23	71.42	66.41	62.26	57.50	53.89	50.69	48.31	46.04	59.31	31.19
ADBS	AAAI'25	79.02	72.27	68.11	64.16	59.67	55.83	53.58	51.52	49.67	61.54	29.35
BiMC†	CVPR'25	72.64	72.63	71.07	69.93	69.61	66.73	66.51	65.17	64.66	68.77	**7.98**
LRT	TPAMI'25	72.27	65.32	58.79	52.72	48.62	44.47	41.71	39.28	37.57	51.19	34.70
ConCM	ICLR'26	80.87	71.10	65.64	63.91	60.52	56.95	54.16	51.57	49.22	61.55	31.65
LeVa (Ours)	-	**88.05**	**84.36**	**81.81**	**80.62**	**79.55**	**76.96**	**76.38**	**74.92**	**74.05**	**79.63**	14.00
D_{sub2}												
SAVC	CVPR'23	79.84	73.54	69.63	64.98	60.58	57.38	54.49	51.34	48.85	62.29	30.99
CLOSER	ECCV'24	76.89	71.10	66.85	62.31	57.79	54.64	51.81	48.41	45.77	59.51	31.12
Comp-FSCIL	ICML'24	77.26	72.51	68.96	64.81	60.03	57.05	53.86	50.44	47.83	61.42	29.43
LRT	TPAMI'25	76.52	70.40	65.45	60.09	55.10	51.38	47.49	43.83	41.21	56.83	35.31
ADBS	AAAI'25	79.14	71.43	67.67	63.06	58.89	56.27	53.72	50.89	48.76	61.09	30.38
BiMC†	CVPR'25	69.19	68.75	68.24	67.08	65.68	65.06	65.26	64.71	64.37	66.48	**4.82**
ConCM	ICLR'26	80.65	70.04	67.76	64.51	61.35	58.57	56.17	52.87	50.26	62.46	30.39
LeVa (Ours)	-	**86.95**	**83.68**	**81.31**	**79.19**	**76.96**	**75.98**	**75.74**	**73.98**	**73.21**	**78.56**	13.74
D_{sub3}												
SAVC	CVPR'23	77.68	72.23	67.46	63.83	61.08	57.93	54.96	51.34	49.15	61.74	28.53
CLOSER	ECCV'24	74.44	68.91	63.66	60.13	57.05	53.74	50.74	47.52	45.27	57.94	29.17
Comp-FSCIL	ICML'24	74.54	69.04	64.46	60.95	58.22	55.20	52.21	48.56	46.99	58.91	27.55
LRT	TPAMI'25	79.83	50.04	50.75	50.19	49.37	47.82	45.73	44.07	42.56	51.15	37.27
ADBS	AAAI'25	76.82	70.18	65.50	62.19	59.66	56.72	54.46	51.15	49.21	60.65	27.61
BiMC†	CVPR'25	70.34	68.81	67.94	66.23	64.30	64.03	63.45	63.53	63.69	65.81	**6.65**
ConCM	ICLR'26	78.86	67.21	64.85	61.43	59.71	57.34	54.41	51.36	49.47	60.52	29.39
LeVa (Ours)	-	**86.02**	**82.01**	**80.71**	**79.06**	**76.61**	**75.35**	**74.11**	**73.23**	**72.87**	**77.77**	13.15