Towards Domain Generalized Few-Shot Class Incremental Learning

Jun-Woo Heo[1], Chang-Sik Woo[1], Tae-Young Lee[1]
[1]Department of Artificial Intelligence, Korea University
dinleo11@korea.ac.kr, woocs@korea.ac.kr, tylee0415@korea.ac.kr

Abstract

*Few-Shot Class-Incremental Learning (FSCIL) addresses the challenge of continually acquiring novel classes from limited labeled examples while mitigating catastrophic forgetting. However, conventional FSCIL methods assume that all data originates from a single domain or a stationary distribution. Domain Generalization (DG) tackles distribution shift across domains but assumes a fixed label space, leaving the problem of joint class expansion and domain robustness largely unexplored. In this paper, we introduce Domain-Generalized Few-Shot Class-Incremental Learning (DG-FSCIL), a novel scenario that requires models to incrementally learn novel classes under few-shot supervision while generalizing to unseen target domains. To address this challenge, we propose **PRISM** (**P**rototype **R**efinement via **I**mage-**S**emantic **M**atching), a novel framework that leverages edge-based structural cues to construct robust prototypes in domain-variant environments. Furthermore, we introduce a meta-learning-based visual prompt tuning strategy that exploits the stability of text embeddings in vision-language models to calibrate visual representations under domain shift. Extensive experiments demonstrate that our approach generalizes effectively to unseen target domains and consistently outperforms strong baselines.*

1. Introduction

Real-world recognition systems operate in dynamic environments where both visual conditions and the set of categories of interest evolve over time. Applications such as autonomous driving and industrial inspection frequently encounter the continuous emergence of new classes, including novel obstacles and previously unseen defect types. However, collecting and curating large-scale labeled data for each update remains impractical. Consequently, practical learning systems must be capable of continually incorporating new knowledge from limited supervision while preserving reliable performance on previously learned classes.

Few-Shot Class Incremental Learning (FSCIL) was introduced to formalize this constraint [42]. The standard FSCIL protocol organizes learning into a base session with abundant training data, followed by a sequence of incremental sessions, each providing only a handful of labeled examples per novel class. FSCIL methods aim to address two fundamental challenges: the rapid acquisition of novel categories under severe sample scarcity, and the mitigation of catastrophic forgetting across an expanding label space. This formulation has since become a widely adopted benchmark for continual adaptation under resource constraints.

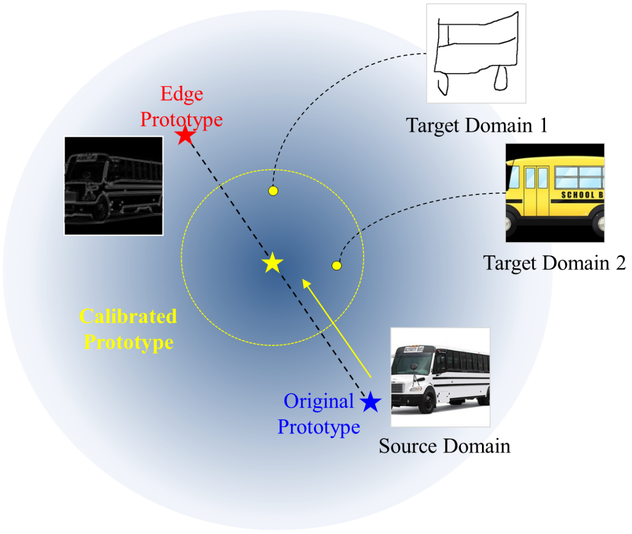


Figure 1. **Method motivation of the proposed DG-FSCIL.** Domain shift primarily affects low-level features (texture, color) while structural information (edges, shapes) remains consistent. Our method extracts domain-invariant structural features to complement prompt-adapted CLIP features and calibrates original prototypes with edge prototypes.

Despite considerable progress, conventional FSCIL methods typically assume that all sessions are drawn from a single domain or from a relatively stationary data distribution. In practice, however, this assumption is frequently violated. Visual inputs can vary substantially over time due to differences in sensors, illumination, background context, visual style, or operating conditions. Under such domain shifts, a learner may overfit to spurious, domain-specific cues present in the limited few-shot data, resulting in unstable class prototypes and degraded performance when evaluated across different environments.

Domain Generalization (DG) addresses a complementary problem: learning representations from multiple source domains that generalize to unseen target domains without requiring target-domain supervision [17]. The DG literature has produced a rich set of techniques for improving robustness to distribution shift. However, standard DG formulations assume a fixed label space and do not accommodate the sequential introduction of novel classes. That is, DG methods are not inherently designed to handle the continual arrival of new categories under few-shot supervision while retaining knowledge of previously learned classes.

To bridge this gap, we introduce a new problem setting: **Domain Generalized Few-Shot Class Incremental Learning (DG-FSCIL)**. In DG-FSCIL, a model undergoes a sequence of class incremental sessions with few-shot supervision for each newly introduced class, while training data may originate from multiple source domains throughout the incremental process. Upon completion of the final session, the model is evaluated on held-out target domains over the cumulative label space, without access to any target-domain labels. This formulation more faithfully captures the demands of realistic deployment scenarios in which both the class vocabulary and the data distribution evolve over time.

DG-FSCIL provides a principled abstraction for numerous safety- and reliability-critical applications. Consider, for instance, an autonomous driving system that must continuously learn to recognize novel objects encountered on the road. Such a system must perform reliably not only under conditions observed during development, such as clear daytime driving, but also under previously unseen operating conditions, including fog, nighttime, and unfamiliar urban environments. Similarly, industrial inspection systems must incorporate new defect categories with minimal annotation effort while maintaining robustness across variable production conditions. In such contexts, the capacity to incrementally acquire new classes must be coupled with resilience to unpredictable domain shifts.

A natural approach to DG-FSCIL would be to combine existing FSCIL and DG techniques. However, naive composition proves insufficient. Few-shot incremental updates amplify prototype noise and accelerate forgetting, while domain shifts induce representation drift that can undermine the stability assumptions of incremental classifiers. Furthermore, these effects compound across sessions: a model that overfits to domain-specific cues when encoding a novel class risks not only poor generalization to unseen domains but also corruption of the shared feature space relied upon for recognizing previously learned classes. Addressing DG-FSCIL therefore requires methods that jointly mitigate incremental interference and domain-induced representational drift.

To tackle these challenges, we propose **E3 (Edge Enhancement Extractor)**, a framework grounded in the hypothesis that structural cues such as edges exhibit greater invariance to domain-specific appearance variations than texture-based features. E3 leverages a combination of standard image filtering operations to extract edge-enhanced representations, which are then integrated into prototype construction. This design yields prototypes that are more robust to domain-dependent variations. Importantly, our objective is not to discard appearance information entirely, but rather to leverage edge-guided signals as a stabilizing complement to learned features under few-shot, domain-shifted conditions.

Additionally, we exploit the strong semantic priors embedded in vision-language models (VLMs) by proposing a meta-learning-based visual prompt tuning strategy tailored for DG-FSCIL. VLMs such as CLIP encode images and text into a shared embedding space, where text embeddings tend to exhibit greater stability under domain shift than their visual counterparts [36]. Building on this observation, we meta-learn visual prompts that calibrate the image encoder to preserve reliable image-text alignment across domain-variant, few-shot incremental sessions. This approach complements existing training-free calibration strategies, such as bi-level modality calibration [9], by introducing a lightweight, learnable mechanism explicitly designed to counteract domain shift while keeping the pretrained backbone largely frozen [19].

Our contributions can be summarized as follows:

- We propose DG-FSCIL, a challenging and practically motivated scenario that unifies few-shot class-incremental learning with domain generalization, requiring models to continually acquire novel classes while generalizing robustly to unseen target domains.
- We present **PRISM** (**P**rototype **R**efinement via **I**mage-**S**emantic **M**atching), a novel framework for DG-FSCIL that combines: (1) bi-level modality calibration built upon CLIP, (2) **Meta-VPDA** (Meta-learned Visual Prompt for Domain Adaptation), and (3) **E**^{3} (Edge Enhancement Extractor) for domain-invariant features.
- We provide extensive empirical evaluation demonstrating that the proposed framework consistently outperforms strong baselines, with particularly pronounced gains in later incremental sessions where the compounding effects of forgetting and domain shift are most severe.

2. Related Works

Continual Learning. Continual Learning aims to train models that can continually acquire new knowledge over a stream of tasks while retaining performance on previously learned tasks, thereby mitigating catastrophic [14, 30, 31] caused by distribution shifts across time. Classical approaches fall into three major categories: regularization-based methods [1, 21, 26, 51], which constrain changes to parameters deemed important for past tasks; architecture-based methods [20, 24, 28, 29, 37, 39, 40, 45, 50, 52], which allocate dedicated parameter subspaces to different tasks; and rehearsal-based methods [4–8, 18, 34, 38, 49], which retain a subset of past samples or generative memories for replay. More recently, the availability of large-scale pre-trained models (*e.g.* ViT [13] and CLIP [35]) has led to a surge of rehearsal-free approaches [27, 32, 41, 44, 46, 47, 53, 54] that exploit strong generalization priors to mitigate forgetting without storing old data. Within this paradigm, Few-Shot Class-Incremental Learning (FSCIL) [10–12, 16, 25, 33] extends these pre-trained model–based strategies to the few-shot regime, where each incremental session provides only a handful of samples. However, existing FSCIL works assume that incremental sessions share the same domain and thus cannot handle jointly evolving classes and domains. We instead study a setting where both class and domain shift occur together, a scenario largely overlooked in prior FSCIL research.

Domain Generalization. Domain Generalization aims to train models that remain robust when deployed on unseen target domains, typically by learning domain-invariant representations, aligning feature distributions, or applying adversarial regularization. While classical DG methods rely on feature alignment or distribution matching [15, 23], recent benchmark studies such as DomainBed have shown that well-tuned ERM can already be highly competitive [17]. To further exploit cross-domain structure, meta-learning approaches [2, 22] and recent arithmetic meta-learning formulations [43] aim to guide updates toward directions that generalize across multiple source domains. In parallel, DG research with vision–language models such as CLIP has explored prompt learning and prompt generation to enhance robustness under domain shift [3, 48]. Despite the practical importance of domain generalization in real-world deployment, continual learning research has largely overlooked domain shifts in the inference phase. We address this gap by incorporating DG principles into a FSCIL setting.

Visual Prompt Tuning. Visual Prompt Tuning (VPT) adapts pre-trained vision or vision–language models by inserting a small set of learnable prompts into the input or intermediate feature space [19]. Unlike full fine-tuning, VPT updates only these prompts while keeping the backbone frozen, enabling lightweight and parameter-efficient transfer. This simple mechanism has been widely adopted as an alternative to heavyweight adaptation methods in various downstream vision tasks.

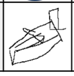
				C_0	C_1	C_2	C_3
Train	D_0	Real		240 Full-shot			
	D_1	Infograph			35 5-shot		
	D_2	Painting				35 5-shot	
	D_3	Sketch					35 5-shot
Test	D_4	Clipart		345			
	D_5	Quickdraw		345			

Figure 2. **Scenario description of the proposed DG-FSCIL.** We propose DG-FSCIL, a challenging scenario where models undergo a sequence of class incremental sessions with few-shot supervision for each newly introduced class, while training data may originate from multiple source domains throughout the incremental process.

3. Problem Formulation: DG-FSCIL

Domain Generalized Few-Shot Class-Incremental Learning (DG-FSCIL) extends conventional Few-Shot Class-Incremental Learning (FSCIL) by introducing domain shifts and requiring generalization to unseen target domains. The overall task consists of a sequence of sessions $\{\mathcal{T}^0, \mathcal{T}^1, \ldots, \mathcal{T}^t, \ldots\}$. Each session $\mathcal{T}^t$ provides a set of new classes $\mathcal{C}^t$, a labeled training set $\mathcal{S}^t$, and an associated domain $\mathcal{D}^t$.

The class sets are disjoint across sessions:

$$\mathcal{C}^i \cap \mathcal{C}^j = \emptyset, \quad \forall i \neq j. \tag{1}$$

The cumulative label space up to session t is defined as

$$\mathcal{C}^{0:t} = \bigcup_{j=0}^{t} \mathcal{C}^j. \tag{2}$$

The training set of session t is

$$\mathcal{S}^t = \left\{(\mathbf{x}_i^t, y_i^t)\right\}_{i=1}^{N^t}, \quad \mathbf{x}_i^t \in \mathcal{X},\ y_i^t \in \mathcal{C}^t, \tag{3}$$

where $\mathbf{x}_i^t$, y_i^t, and N^t denote the i-th image, its class label, and the number of training samples, respectively. Samples in $\mathcal{S}^t$ are drawn from the session-specific domain distribution $p_{\mathcal{D}^t}(x, y)$.

In the base session $\mathcal{T}^0$, the model is trained on base classes $\mathcal{C}^0$ with full-shot samples from domain $\mathcal{D}^0$. In incremental sessions $\mathcal{T}^t$ with $t \geq 1$, the model encounters novel classes $\mathcal{C}^t$ with only few-shot samples per class. Specifically, for each new class $c \in \mathcal{C}^t$, only K labeled samples are provided (with $K = 5$ in this work), so $N^t = K \cdot |\mathcal{C}^t|$ for $t \geq 1$. The domain $\mathcal{D}^t$ changes across sessions, which induces domain shift during the incremental learning process.

DG-FSCIL follows a domain generalization protocol. Let $\mathcal{D}_{src}$ be the set of source domains observed during the training sessions, and let $\mathcal{D}_{tgt}$ be the set of unseen target domains used only for evaluation. We define

$$\mathcal{D}_{src} = \{\mathcal{D}^j \mid j \in \{0, 1, \ldots, T\}\}, \tag{4}$$

and assume that target domains $\mathcal{D}_{tgt}$ are disjoint from the observed source domains:

$$\mathcal{D}_{src} \cap \mathcal{D}_{tgt} = \emptyset. \tag{5}$$

No labeled data from $\mathcal{D}_{tgt}$ is available during training.

Let θ^t denote the model parameters after learning up to session t. At each session, the model is updated using only the current session data $\mathcal{S}^t$. After the final training session T, the model is evaluated on target-domain test sets from $\mathcal{D}_{tgt}$ over the cumulative label space $\mathcal{C}^{0:T}$. The goal is to maximize the classification performance on unseen target domains while retaining knowledge of previously learned classes.

4. Method: PRISM

We present **PRISM** (**P**rototype **R**efinement via **I**mage-**S**emantic **M**atching), a novel framework for DG-FSCIL that combines: (1) bi-level modality calibration built upon CLIP, (2) **Meta-VPDA** (Meta-learned Visual Prompt for Domain Adaptation), and (3) **E**^{3} (Edge Enhancement Extractor) for domain-invariant features. Figure 3 illustrates the overall architecture.

4.1. Base Model: Bi-Level Modality Calibration

Our foundation extends the BiMC framework [9] built upon CLIP [35], constructing robust classifiers through intra-modal and inter-modal calibration without additional training.

4.1.1. Intra-Modal Calibration

Textual Calibration. We leverage LLM-generated descriptions to enrich zero-shot CLIP classifiers. For class c with n_c descriptions $\{\mathbf{t}_{c,j}\}_{j=1}^{n_c}$, the calibrated text prototype is:

$$\tilde{\boldsymbol{\mu}}_c^T = (1 - \lambda_T)\mathbf{w}_c + \lambda_T \left(\frac{1}{n_c} \sum_{j=1}^{n_c} \frac{g(\mathbf{t}_{c,j})}{\|g(\mathbf{t}_{c,j})\|_2} \right) \tag{6}$$

where $\mathbf{w}_c$ is the original CLIP zero-shot weight and $g(\cdot)$ is the text encoder.

Visual Calibration. For class c, we compute the naive visual prototype from training samples:

$$\boldsymbol{\mu}_c^I = \frac{1}{|\mathcal{D}_c|} \sum_{\mathbf{x} \in \mathcal{D}_c} \frac{f(\mathbf{x})}{\|f(\mathbf{x})\|_2} \tag{7}$$

Novel classes ($c \notin \mathcal{Y}_0$) suffer from estimation bias due to limited samples. We calibrate using base class prototypes:

$$\tilde{\boldsymbol{\mu}}_c^I = \begin{cases} \boldsymbol{\mu}_c^I & \text{if } c \in \mathcal{Y}_0 \\ (1 - \lambda_I)\boldsymbol{\mu}_c^I + \lambda_I \sum_{b \in \mathcal{Y}_0} s_{b,c} \boldsymbol{\mu}_b^I & \text{otherwise} \end{cases} \tag{8}$$

where $s_{b,c} = \frac{\exp(\tau \langle \boldsymbol{\mu}_b^I, \boldsymbol{\mu}_c^I \rangle)}{\sum_{i \in \mathcal{Y}_0} \exp(\tau \langle \boldsymbol{\mu}_i^I, \boldsymbol{\mu}_c^I \rangle)}$ measures prototype similarity.

4.1.2. Inter-Modal Calibration

We fuse calibrated textual and visual prototypes:

$$\boldsymbol{\mu}_c^{\text{base}} = \beta \tilde{\boldsymbol{\mu}}_c^T + (1 - \beta)\tilde{\boldsymbol{\mu}}_c^I \tag{9}$$

Classification scores use cosine similarity:

$$s_c^{\text{calib}}(\mathbf{x}) = \frac{f(\mathbf{x})^\top \boldsymbol{\mu}_c^{\text{base}}}{\|f(\mathbf{x})\|_2 \cdot \|\boldsymbol{\mu}_c^{\text{base}}\|_2} \tag{10}$$

4.1.3. Enhanced Metrics

Anisotropic Covariance. We maintain a global covariance matrix evolving across tasks:

$$\tilde{\boldsymbol{\Sigma}}_G^t = \frac{|\mathcal{Y}_{t-1}|}{|\mathcal{Y}_t|} \tilde{\boldsymbol{\Sigma}}_G^{t-1} + \left(1 - \frac{|\mathcal{Y}_{t-1}|}{|\mathcal{Y}_t|}\right) \tilde{\boldsymbol{\Sigma}}^t \tag{11}$$

where $\tilde{\boldsymbol{\Sigma}}^t = \boldsymbol{\Sigma}^t + \frac{\gamma_{\text{reg}}}{d}\text{tr}(\boldsymbol{\Sigma}^t)\mathbf{I}_d$. For base classes:

$$s_c^{\text{cov}}(\mathbf{x}) = -\frac{1}{d}(f(\mathbf{x}) - \tilde{\boldsymbol{\mu}}_c^I)^\top (\tilde{\boldsymbol{\Sigma}}_G^t)^{-1} (f(\mathbf{x}) - \tilde{\boldsymbol{\mu}}_c^I) \tag{12}$$

Cross-Modal Nearest Neighbor. For novel classes with unreliable covariance:

$$s_c^{\text{nn}}(\mathbf{x}) = \max_{j \in [1, n_c]} \left\{ \frac{f(\mathbf{x})^\top g(\mathbf{t}_{c,j})}{\|f(\mathbf{x})\|_2 \cdot \|g(\mathbf{t}_{c,j})\|_2} \right\} \tag{13}$$

Masked Ensemble. Final prediction combines metrics based on class type:

$$p_c(\mathbf{x}) = \begin{cases} \alpha \cdot p_c^{\text{calib}} + (1 - \alpha) \cdot p_c^{\text{cov}} & \text{if } c \in \mathcal{Y}_0 \\ \alpha \cdot p_c^{\text{calib}} + (1 - \alpha) \cdot p_c^{\text{nn}} & \text{if } c \notin \mathcal{Y}_0 \end{cases} \tag{14}$$

where $p_c^{(\cdot)} = \text{softmax}(s_c^{(\cdot)})$.

4.2. Meta-VPDA: Meta-learned Visual Prompt for Domain Adaptation

CLIP must adapt to domain-specific characteristics (sketch styles, painting textures) without forgetting. We introduce Meta-VPDA, a two-stage meta-learning framework that learns a generalizable prompt initialization and specializes it for each domain.

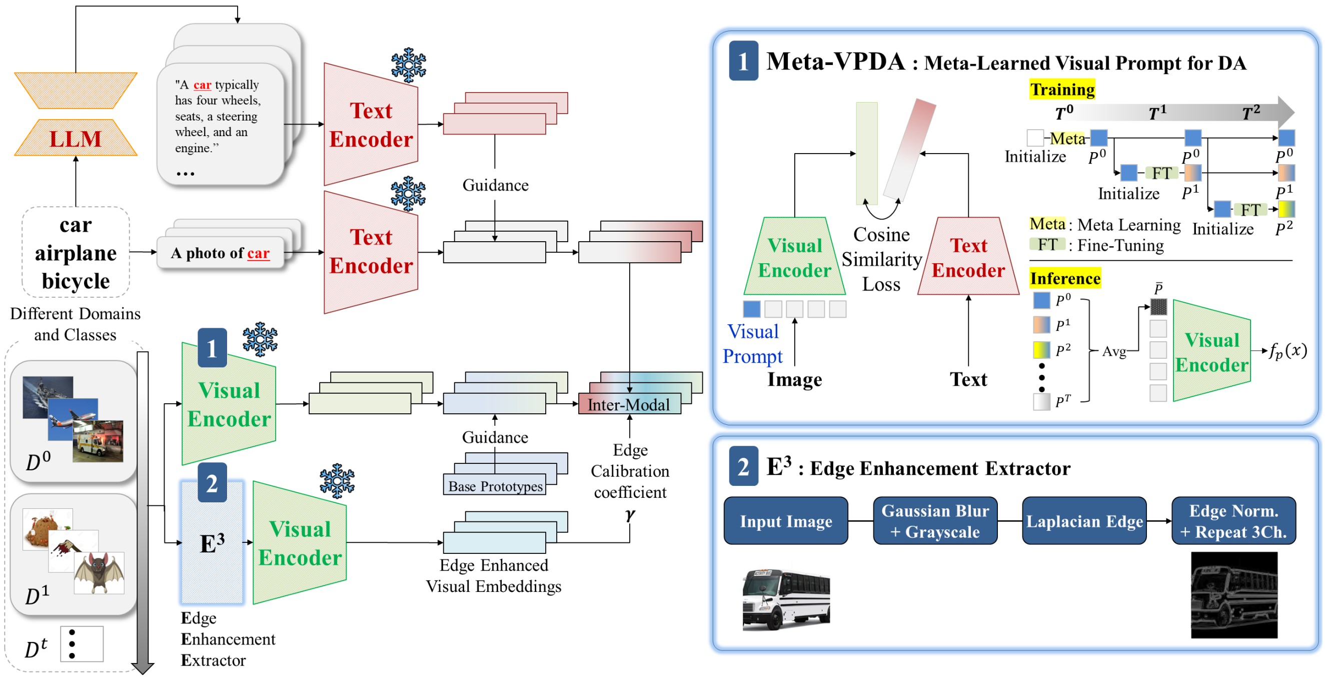


Figure 3. **Overview of the PRISM framework**. A Novel framework for DG- FSCIL that combines: (1) bi-level modality calibration built upon CLIP, (2) Meta-VPDA (Meta-learned Visual Prompt for Domain Adaptation), and (3) E3 (Edge Enhancement Extractor) for domain-invariant features.

4.2.1. Visual Prompt Architecture

We augment CLIP's Vision Transformer with K learnable prompt tokens $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_K] \in \mathbb{R}^{K \times d}$. Given patch tokens $\{\mathbf{z}_i\}_{i=1}^N$ and class token $\mathbf{z}_{\text{cls}}$ with positional embeddings $\{\mathbf{e}_i\}_{i=0}^N$, prompts are injected after positional embeddings:

$$\mathbf{Z}_0 = [\mathbf{z}_{\text{cls}} + \mathbf{e}_0, \mathbf{p}_1, \ldots, \mathbf{p}_K, \mathbf{z}_1 + \mathbf{e}_1, \ldots, \mathbf{z}_N + \mathbf{e}_N] \quad (15)$$

Prompts receive no positional embeddings, making them position-agnostic modulators. The sequence processes through transformer layers, extracting features from the class token:

$$f(\mathbf{x}; \mathbf{P}) = \text{Proj}(\text{Transformer}_L(\mathbf{Z}_0)^{[\text{cls}]}) \quad (16)$$

4.2.2. Stage 1: Episode-Based Meta-Learning

On the base task (Task 0, 240 classes), we perform episodic meta-learning. Each episode samples classes into:

- Support set $\mathcal{S}$: 28 classes, 5 shots per class
- Query set $\mathcal{Q}$: 7 classes, 5 shots per class

The meta-objective maximizes image-text alignment after adaptation:

$$\min_{\mathbf{P}_0} \mathbb{E}_{\mathcal{S},\mathcal{Q}} \left[\mathcal{L}_{\text{query}}(\mathbf{P}_0 - \alpha \nabla_{\mathbf{P}_0} \mathcal{L}_{\text{support}}(\mathbf{P}_0; \mathcal{S}); \mathcal{Q})\right] \quad (17)$$

where $\mathcal{L}(\mathbf{P}; \mathcal{D}) = -\frac{1}{|\mathcal{D}|} \sum_{(\mathbf{x},y) \in \mathcal{D}} \frac{f(\mathbf{x};\mathbf{P})^\top g(y)}{\|f(\mathbf{x};\mathbf{P})\|_2 \cdot \|g(y)\|_2}$, $f(\mathbf{x}; \mathbf{P})$ is the prompt-adapted image embedding, and $g(y)$ is the text embedding for class y.

We implement gradient-based meta-learning:

1. **Inner loop**: Adapt $\mathbf{P}_0$ on $\mathcal{S}$ via $T = 10$ gradient steps with learning rate $\alpha = 0.01$
2. **Outer loop**: Evaluate on $\mathcal{Q}$ and update meta-prompt with learning rate $\beta = 0.0001$

After 100 episodes, we obtain meta-prompt $\mathbf{P}_0$.

4.2.3. Stage 2: Task-Specific Fine-Tuning

For each domain-specific task $t \in \{0, 1, 2, 3\}$:

- **Task 0**: $\mathbf{P}_0^* = \mathbf{P}_0$ (no fine-tuning)
- **Tasks 1-3**: Initialize $\mathbf{P}_t^{(0)} = \mathbf{P}_0$ and fine-tune on 5-shot data for 10 epochs:

$$\mathbf{P}_t^* = \arg\min_{\mathbf{P}_t} \mathcal{L}(\mathbf{P}_t; \mathcal{D}_t^{\text{train}}) \quad \text{s.t.} \quad \mathbf{P}_t^{(0)} = \mathbf{P}_0 \quad (18)$$

Task-specific prompts $\{\mathbf{P}_t^*\}_{t=0}^3$ are stored in a prompt pool.

4.2.4. Inference with Averaged Prompts

At test time, we average all task-specific prompts from the $T + 1$ training domains:

$$\bar{\mathbf{P}} = \frac{1}{T+1} \sum_{t=0}^{T} \mathbf{P}_t^* \quad (19)$$

This averaged prompt provides balanced domain adaptation across all sessions. The adapted feature extraction becomes:

$$f_{\text{prompt}}(\mathbf{x}) = f(\mathbf{x}; \bar{\mathbf{P}}) \quad (20)$$

4.3. E^3: Edge Enhancement Extractor

Domain shift primarily affects low-level features (texture, color) while structural information (edges, shapes) remains consistent. E^3 extracts domain-invariant structural features to complement prompt-adapted CLIP features.

4.3.1. Laplacian-of-Gaussian Edge Detection

Given input $\mathbf{x} \in \mathbb{R}^{H\times W\times 3}$, E^3 applies fixed preprocessing:

1. **Gaussian smoothing**: We apply a 2D Gaussian filter with kernel size 5 and standard deviation σ to suppress noise while preserving structural edges:

$$G_\sigma(u,v) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{u^2+v^2}{2\sigma^2}\right) \tag{21}$$

   where (u,v) are spatial coordinates relative to the kernel center. The smoothed image is obtained via convolution:

$$\mathbf{x}_{\text{blur}} = G_\sigma * \mathbf{x} \tag{22}$$

   Larger σ values remove fine-grained texture variations while retaining dominant object boundaries, yielding more domain-invariant representations.

2. **Grayscale conversion**: $\mathbf{x}_{\text{gray}} = 0.299R + 0.587G + 0.114B$
3. **Laplacian edge detection**: Apply the discrete Laplacian kernel to detect edges:

$$\mathbf{L} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{23}$$

4. **Normalization**: The edge response is normalized to [0, 1] and replicated across RGB channels:

$$\mathbf{x}_{\text{edge}} = \frac{|\mathbf{L} * \mathbf{x}_{\text{gray}}|}{\max(|\mathbf{L} * \mathbf{x}_{\text{gray}}|)} \cdot \mathbf{1}_{1\times 1\times 3} \tag{24}$$

4.3.2. Domain-Invariant Feature Extraction

Edge images are encoded through CLIP's visual encoder *without prompts*:

$$\mathbf{f}_{\text{edge}}(\mathbf{x}) = \frac{f(\mathbf{x}_{\text{edge}};\emptyset)}{\|f(\mathbf{x}_{\text{edge}};\emptyset)\|_2} \tag{25}$$

where $\emptyset$ indicates no prompt, preserving domain-invariance. We construct edge prototypes:

$$\boldsymbol{\mu}_c^{\text{edge}} = \frac{1}{|\mathcal{D}_c|}\sum_{\mathbf{x}\in\mathcal{D}_c} \mathbf{f}_{\text{edge}}(\mathbf{x}) \tag{26}$$

and apply the same calibration as visual prototypes for novel classes.

4.3.3. Multi-Level Feature Fusion

We integrate edge features at two levels:

Feature-level fusion combines prompt-adapted and edge features:

$$\mathbf{f}_{\text{test}}(\mathbf{x}) = (1-\gamma)f_{\text{prompt}}(\mathbf{x}) + \gamma\mathbf{f}_{\text{edge}}(\mathbf{x}) \tag{27}$$

Prototype-level fusion incorporates edge information:

$$\boldsymbol{\mu}_c^{\text{final}} = (1-\gamma)\boldsymbol{\mu}_c^{\text{base}} + \gamma\tilde{\boldsymbol{\mu}}_c^{\text{edge}} \tag{28}$$

Final calibrated score:

$$s_c^{\text{calib}}(\mathbf{x}) = \frac{\mathbf{f}_{\text{test}}(\mathbf{x})^\top \boldsymbol{\mu}_c^{\text{final}}}{\|\mathbf{f}_{\text{test}}(\mathbf{x})\|_2 \cdot \|\boldsymbol{\mu}_c^{\text{final}}\|_2} \tag{29}$$

Unlike adaptive edge extractors, E^3 uses fixed hyperparameters to prevent overfitting. Edge features remain domain-agnostic by bypassing prompts during encoding.

4.4. Complete Pipeline

Training: (1) Meta-learning: 100 episodes on Task 0 → meta-prompt $\mathbf{P}_0$ (2) Fine-tuning: Task 0 uses $\mathbf{P}_0$ directly; Tasks 1-3 fine-tune for 10 epochs → prompt pool $\{\mathbf{P}_t^*\}_{t=0}^3$ (3) Prototypes: Build $\{\boldsymbol{\mu}_c^{\text{final}}\}$ via calibration + edge fusion

Inference: (1) Load averaged prompt $\bar{\mathbf{P}} = \frac{1}{4}\sum_{t=0}^3 \mathbf{P}_t^*$ (2) Extract fused features $\mathbf{f}_{\text{test}}(\mathbf{x}) = (1-\gamma)f(\mathbf{x};\bar{\mathbf{P}}) + \gamma\mathbf{f}_{\text{edge}}(\mathbf{x})$ (3) Compute scores: s_c^{calib}, s_c^{cov} (base), s_c^{nn} (novel) (4) Predict: $\hat{y} = \arg\max_c p_c(\mathbf{x})$ using masked ensemble

5. Experiments

5.1. Experimental Setup

Datasets. We evaluate on DomainNet, a large-scale domain shift benchmark with 345 classes across 6 domains. Following the DG-FSCIL protocol, we use: (1) **Source domains** for training: Real (Task 0, 240 base classes), Infograph (Task 1, 35 classes), Painting (Task 2, 35 classes), Sketch (Task 3, 35 classes), all with 5-shot incremental learning. (2) **Target domains** for testing: Clipart and Quickdraw (unseen during training).

Implementation Details. We use CLIP ViT-B/16 as the backbone. For meta-learning, we run 100 episodes with 28 support and 7 query classes per episode, using 5-shot per class. Inner loop uses 10 gradient steps with $\alpha = 0.01$, outer loop uses $\beta = 0.0001$. For task-specific fine-tuning (Tasks 1-3), we train for 10 epochs with batch size 64.

Hyperparameters: intra-modal calibration $\lambda_T = 0.5$, $\lambda_I = 0.1$, temperature $\tau = 16$; inter-modal calibration β via validation; covariance regularization $\gamma_{\text{reg}} = 1.0$ (base), $\gamma_{\text{reg}} = 5.0$ (incremental); ensemble weight $\alpha = 0.6$. For E^3, we use Gaussian blur $\sigma = 5.0$ and edge fusion weight $\gamma = 0.4$ (selected via hyperparameter search in Section 5.3).

Evaluation Metrics. We report per-class accuracy on target domains (Clipart and Quickdraw) after training on all source domains (Tasks 0-3). Mean accuracy across classes and target domain average are also reported.

Table 1. Accuracy (%) on DomainNet target domains (Clipart and Quickdraw). All methods are trained on source domains (Real, Infograph, Painting, Sketch) and tested on unseen target domains. Bold indicates best performance per column.

Method	Clipart					Quickdraw					Target Average
	Class 1	Class 2	Class 3	Class 4	Mean	Class 1	Class 2	Class 3	Class 4	Mean	
BiMC (Baseline)	65.87	**42.31**	**67.01**	**78.03**	64.57	6.54	0.57	6.07	16.82	6.93	35.75
BiMC + VPDA	64.13	39.15	64.13	77.95	62.75	5.3	0.12	3.47	10.87	5.15	33.95
BiMC + Meta-VPDA	68.85	31.87	65.43	77.31	65.24	9.30	0.47	7.31	12.42	8.52	36.88
BiMC + E^3	69.50	38.55	60.41	74.07	**65.70**	11.30	**1.15**	**9.35**	**23.23**	11.28	**38.49**
PRISM (Ours)	**70.78**	23.06	58.50	77.09	64.99	**12.81**	0.41	8.54	16.58	**11.50**	38.24

Table 2. Hyperparameter analysis on Clipart (Mean Accuracy %). Rows: edge fusion weight γ, Columns: Gaussian blur σ. Bold indicates best overall configuration, underline indicates our selected configuration ($\gamma = 0.4$, $\sigma = 5.0$).

$\gamma \backslash \sigma$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.2	64.26	65.19	65.51	65.58	65.68	65.71	65.72	65.73	65.73	**65.74**
0.4	61.99	64.05	64.65	64.72	64.87	64.99	65.00	64.98	64.97	64.99
0.5	59.73	62.52	63.23	63.32	63.47	63.52	63.56	63.56	63.59	63.59
0.6	56.26	60.09	61.06	60.81	60.95	60.95	60.92	60.91	60.92	60.93
0.8	44.07	47.20	47.69	42.30	41.69	41.44	41.34	41.31	41.34	41.35

Table 3. Hyperparameter analysis on Quickdraw (Mean Accuracy %). Rows: edge fusion weight γ, Columns: Gaussian blur σ. Bold indicates best overall configuration, underline indicates our selected configuration ($\gamma = 0.4$, $\sigma = 5.0$).

$\gamma \backslash \sigma$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.2	8.86	9.50	9.52	9.72	9.83	9.87	9.92	9.94	9.95	9.96
0.4	9.38	10.73	10.65	11.01	11.25	11.35	11.41	11.46	11.49	11.50
0.5	9.55	11.21	11.05	11.37	11.59	11.71	11.75	11.76	**11.78**	**11.78**
0.6	9.61	11.42	11.21	11.22	11.29	11.30	11.32	11.33	11.32	11.33
0.8	8.70	9.79	9.35	7.37	7.11	6.85	6.58	6.40	6.30	6.24

5.2. Main Results

Table 1 presents the main results on DomainNet target domains. BiMC + E^3 achieves the highest overall target average (38.49%) with strong performance on Clipart (65.70%).

Component Analysis. We first compare prompt-based methods. BiMC + VPDA (direct task-specific fine-tuning without meta-learning) underperforms the baseline (33.95% vs 35.75%), showing that naive prompt tuning fails under domain shift. In contrast, BiMC + Meta-VPDA (meta-learned initialization followed by task-specific fine-tuning) improves target average by +1.13% (36.88%), demonstrating that episodic meta-learning provides crucial domain-robust initialization. Adding E^3 alone (BiMC + E^3) provides the largest gain of +2.74% (38.49%), showing that domain-invariant structural features are particularly effective. The full PRISM model (38.24%) excels on base classes (Class 1: 70.78%/12.81%), showing complementary benefits of edge features and meta-learned prompts for different aspects of domain generalization.

Domain-Specific Trade-offs. On Clipart, BiMC + E^3 achieves the highest mean accuracy (65.70%), with the baseline surprisingly performing best on Classes 2-4, suggesting that for moderately challenging domain shifts, simpler approaches may suffice for incremental classes. However, PRISM demonstrates superior base class generalization with the highest Class 1 accuracy (70.78%). On Quickdraw, the more challenging domain with extreme domain shift, PRISM and BiMC + E^3 significantly outperform the baseline on most classes (e.g., Class 3: 8.54% and 9.35% vs 6.07%, Class 4: 16.58% and 23.23% vs 16.82%), highlighting the critical importance of domain-invariant features and adaptive prompts.

Class-wise Analysis. PRISM excels on Class 1 (base session) across both domains (Clipart: 70.78%, Quickdraw: 12.81%), indicating strong base class generalization enabled by meta-learned prompts. BiMC + E^3 shows more balanced performance across incremental classes, particularly on Quickdraw Class 4 (23.23% vs PRISM's 16.58%). Performance degrades on later incremental classes for both methods (e.g., Clipart Class 2: 23.06% for PRISM, 38.55% for BiMC + E^3), suggesting room for improvement in few-shot incremental scenarios with severe domain shift.

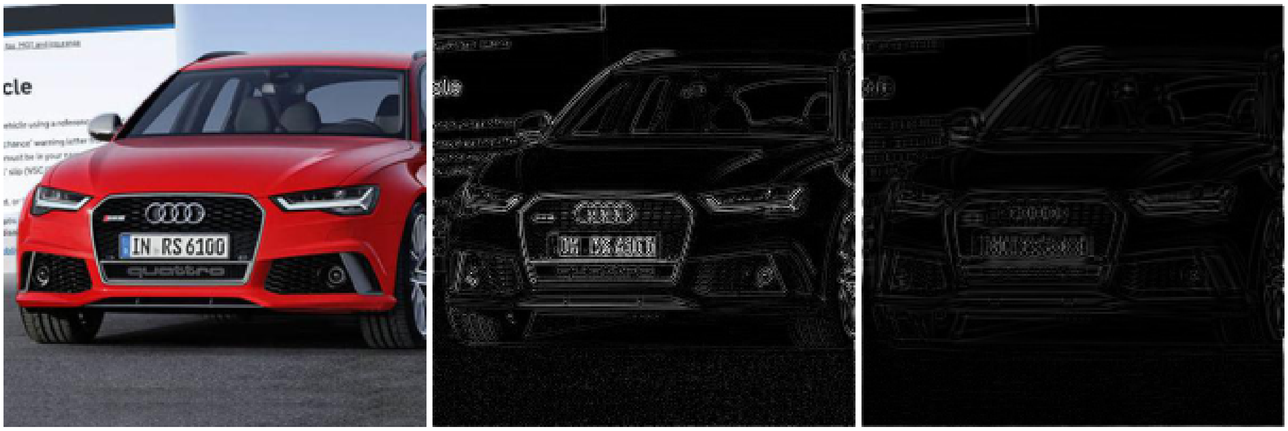


Figure 4. Effect of Gaussian blur σ on edge-enhanced representations in E^3. **Left**: original image. **Middle**: edge map with weak smoothing ($\sigma = 0.5$), preserving fine-grained texture but introducing noise. **Right**: edge map with strong smoothing ($\sigma = 5.0$), suppressing local texture while emphasizing stable structural contours.

5.3. Hyperparameter Analysis

We analyze the impact of E^3 hyperparameters: Gaussian blur σ and edge fusion weight γ. Tables 2 and 3 show mean accuracy on Clipart and Quickdraw across a comprehensive grid search.

Effect of σ (Gaussian Blur). For most γ values, increasing σ from 0.5 to 5.0 improves performance as stronger smoothing reduces noise while preserving structural edges. For $\gamma = 0.4$ on Clipart, performance increases from 61.99% ($\sigma = 0.5$) to 64.99% ($\sigma = 5.0$). On Quickdraw, the trend is similar, with performance improving from 9.38% ($\sigma = 0.5$) to 11.50% ($\sigma = 5.0$) for $\gamma = 0.4$.

Effect of γ (Edge Fusion Weight). The optimal γ varies by domain. On Clipart, low $\gamma = 0.2$ performs best (65.74% at $\sigma = 5.0$), suggesting that semantic features should dominate for less abstract domains. On Quickdraw, moderate $\gamma = 0.5$ achieves the highest accuracy (11.78% at $\sigma = 4.5 - 5.0$), indicating that stronger edge contribution helps on abstract sketches. We select $\gamma = 0.4$ as a balanced compromise that performs well on both domains without extreme specialization.

Selected Configuration. We choose $\gamma = 0.4, \sigma = 5.0$ as our default configuration. While not optimal for either domain individually (Clipart best: 65.74% at $\gamma = 0.2, \sigma = 5.0$; Quickdraw best: 11.78% at $\gamma = 0.5, \sigma = 4.5 - 5.0$), this configuration achieves strong performance on both (Clipart: 64.99%, Quickdraw: 11.50%), demonstrating robustness to domain shift. This trade-off prioritizes generalization over domain-specific optimization.

Stability Analysis. Low γ values (0.2-0.4) show stable performance across varying σ, while high γ (0.8) exhibits significant degradation on Clipart (41.31%-47.69%), confirming that over-reliance on edge features harms semantic understanding.

5.4. Effect of Gaussian Blur on Edge Representation

Figure 4 visualizes the effect of different Gaussian blur strengths σ used in E^3. We show the original image and the corresponding edge-enhanced representations obtained with a weak blur ($\sigma = 0.5$) and a strong blur ($\sigma = 5.0$).

As shown in Figure 4, a small σ preserves high-frequency details, resulting in dense and noisy edge responses that are sensitive to domain-specific textures. In contrast, a larger σ removes local variations while retaining dominant object boundaries, yielding cleaner and more shape-oriented representations. This observation supports our empirical findings in Section 5.3, where larger σ values consistently improve cross-domain performance by promoting domain-invariant structural cues rather than texture-dependent features.

6. Conclusion

We introduced Domain Generalized Few-Shot Class-Incremental Learning (DG-FSCIL), a novel problem setting that requires models to incrementally learn novel classes under few-shot supervision while generalizing to unseen target domains. To address this challenge, we proposed PRISM, a novel framework combining bi-level modality calibration, Meta-VPDA for domain-adaptive visual prompts, and E³ for domain-invariant edge features. Extensive experiments on DomainNet validate our approach: Meta-VPDA improves target average by +1.13% and E³ provides +2.74% gains over baseline, with the full model excelling on base classes across both target domains. Our analysis reveals that balanced hyperparameter selection prioritizes cross-domain generalization over domain-specific optimization. Future work includes exploring learnable edge extraction and extending our framework to other continual learning scenarios with evolving distributions.

References

[1] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European conference on computer vision (ECCV)*, pages 139–154, 2018. 3

[2] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. In *Advances in Neural Information Processing Systems*, pages 998–1008, 2018. 3

[3] Shirsha Bose, Ankit Jha, Enrico Fini, Mainak Singha, Elisa Ricci, and Biplab Banerjee. Stylip: Multi-scale style-conditioned prompt learning for clip-based domain generalization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5530–5540, 2024. 3

[4] Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. *Advances in neural information processing systems*, 33:15920–15930, 2020. 3

[5] Hyuntak Cha, Jaeho Lee, and Jinwoo Shin. Co2l: Contrastive continual learning. In *Proceedings of the IEEE/CVF International conference on computer vision*, pages 9516–9525, 2021.

[6] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. In *International Conference on Learning Representations*, 2019.

[7] Arslan Chaudhry, Marcus Rohrbach, Mohamed Elhoseiny, Thalaiyasingam Ajanthan, Puneet K Dokania, Philip HS Torr, and Marc'Aurelio Ranzato. On tiny episodic memories in continual learning. *arXiv preprint arXiv:1902.10486*, 2019.

[8] Arslan Chaudhry, Albert Gordo, Puneet Dokania, Philip Torr, and David Lopez-Paz. Using hindsight to anchor past knowledge in continual learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 6993–7001, 2021. 3

[9] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 2, 4

[10] 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 *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 9881–9890, 2025. 3

[11] Marco D'Alessandro, Alberto Alonso, Enrique Calabrés, and Mikel Galar. Multimodal parameter-efficient few-shot class incremental learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3393–3403, 2023.

[12] Thang Doan, Sima Behpour, Xin Li, Wenbin He, Liang Gou, and Liu Ren. A streamlined approach to multimodal few-shot class incremental learning for fine-grained datasets. *arXiv preprint arXiv:2403.06295*, 2024. 3

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021. 3

[14] Robert M French. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135, 1999. 3

[15] Yaroslav Ganin, Natalia Ustinova, Hakan Ajakan, Pascal Germain, Hugo Larochelle, Francis Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. In *Journal of Machine Learning Research*, pages 1–35, 2016. 3

[16] 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*, pages 4075–4084, 2024. 3

[17] Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. *arXiv preprint arXiv:2107.01403*, 2021. 2, 3

[18] Tyler L Hayes, Nathan D Cahill, and Christopher Kanan. Memory efficient experience replay for streaming learning. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 9769–9776. IEEE, 2019. 3

[19] Menglin Jia et al. Visual prompt tuning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 2, 3

[20] Zixuan Ke, Bing Liu, and Xingchang Huang. Continual learning of a mixed sequence of similar and dissimilar tasks. *Advances in neural information processing systems*, 33:18493–18504, 2020. 3

[21] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 3

[22] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. *AAAI*, 2018. 3

[23] Haoliang Li, Sinno Jialin Wang, Shuo Pan, and Alex C Kot. Domain generalization with adversarial feature learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5400–5409, 2018. 3

[24] Xilai Li, Yingbo Zhou, Tianfu Wu, Richard Socher, and Caiming Xiong. Learn to grow: A continual structure learning framework for overcoming catastrophic forgetting. In *International conference on machine learning*, pages 3925–3934. PMLR, 2019. 3

[25] Yanan Li, Linpu He, Feng Lin, and Donghui Wang. Few-shot class-incremental learning via cross-modal alignment with feature replay. In *Chinese Conference on Pattern Recogni-*

tion and Computer Vision (PRCV), pages 19–33. Springer, 2024. 3

[26] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 3

[27] Yan-Shuo Liang and Wu-Jun Li. Inflora: Interference-free low-rank adaptation for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23638–23647, 2024. 3

[28] Noel Loo, Siddharth Swaroop, and Richard E Turner. Generalized variational continual learning. *arXiv preprint arXiv:2011.12328*, 2020. 3

[29] Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7765–7773, 2018. 3

[30] James L McClelland, Bruce L McNaughton, and Randall C O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419, 1995. 3

[31] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, pages 109–165. Elsevier, 1989. 3

[32] Mark D McDonnell, Dong Gong, Amin Parvaneh, Ehsan Abbasnejad, and Anton Van den Hengel. Ranpac: Random projections and pre-trained models for continual learning. *Advances in Neural Information Processing Systems*, 36:12022–12053, 2023. 3

[33] Keon-Hee Park, Kyungwoo Song, and Gyeong-Moon Park. Pre-trained vision and language transformers are few-shot incremental learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23881–23890, 2024. 3

[34] Quang Pham, Chenghao Liu, and Steven Hoi. Dualnet: Continual learning, fast and slow. *Advances in Neural Information Processing Systems*, 34:16131–16144, 2021. 3

[35] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 3, 4

[36] Alec Radford et al. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2021. 2

[37] Dushyant Rao, Francesco Visin, Andrei Rusu, Razvan Pascanu, Yee Whye Teh, and Raia Hadsell. Continual unsupervised representation learning. *Advances in neural information processing systems*, 32, 2019. 3

[38] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017. 3

[39] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016. 3

[40] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. In *International conference on machine learning*, pages 4548–4557. PMLR, 2018. 3

[41] James Seale Smith, Leonid Karlinsky, Vyshnavi Gutta, Paola Cascante-Bonilla, Donghyun Kim, Assaf Arbelle, Rameswar Panda, Rogerio Feris, and Zsolt Kira. Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11909–11919, 2023. 3

[42] Xiaoyu Tao et al. Few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[43] Xiran Wang, Jian Zhang, Lei Qi, and Yinghuan Shi. Balanced direction from multifarious choices: Arithmetic metalearngulrajani2021domainbeding for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. 3

[44] Yabin Wang, Zhiwu Huang, and Xiaopeng Hong. S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. *Advances in Neural Information Processing Systems*, 35:5682–5695, 2022. 3

[45] Zifeng Wang, Tong Jian, Kaushik Chowdhury, Yanzhi Wang, Jennifer Dy, and Stratis Ioannidis. Learn-prune-share for lifelong learning. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 641–650. IEEE, 2020. 3

[46] Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, et al. Dualprompt: Complementary prompting for rehearsal-free continual learning. In *European conference on computer vision*, pages 631–648. Springer, 2022. 3

[47] Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 139–149, 2022. 3

[48] Changsong Wen, Zelin Peng, Yu Huang, Xiaokang Yang, and Wei Shen. Domain generalization in clip via learning with diverse text prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. 3

[49] Yue Wu, Yinpeng Chen, Lijuan Wang, Yuancheng Ye, Zicheng Liu, Yandong Guo, and Yun Fu. Large scale incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 374–382, 2019. 3

[50] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. *arXiv preprint arXiv:1708.01547*, 2017. 3

[51] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International*

conference on machine learning, pages 3987–3995. PMLR, 2017. 3

[52] Tingting Zhao, Zifeng Wang, Aria Masoomi, and Jennifer Dy. Deep bayesian unsupervised lifelong learning. *Neural Networks*, 149:95–106, 2022. 3

[53] Da-Wei Zhou, Hai-Long Sun, Han-Jia Ye, and De-Chuan Zhan. Expandable subspace ensemble for pre-trained model-based class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23554–23564, 2024. 3

[54] Da-Wei Zhou, Zi-Wen Cai, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need. *International Journal of Computer Vision*, 133(3): 1012–1032, 2025. 3