LANGUAGE-AWARE COMPLEMENTARY DECOMPOSITION FOR EXEMPLAR-FREE DOMAIN-INCREMENTAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Exemplar-free domain-incremental learning (DIL) requires models to continually adapt to new domains while retaining previously acquired knowledge without storing historical images. Although parameter isolation mitigates forgetting by preserving domain-specific knowledge in separate lightweight modules, effective inference still requires reliably identifying which knowledge should be invoked when domain identities are unavailable at test time. This is particularly challenging because class semantics and domain variations are entangled in visual representations, causing the evidence useful for recognition and domain selection to interfere with each other. Motivated by the insight that language semantics provide a natural reference for organizing these two types of information, we propose **LaCoD**, a **Language-Aware Complementary Decomposition** framework for exemplar-free DIL. LaCoD constructs a semantic subspace from frozen CLIP text embeddings and decomposes visual features into a language-aligned semantic component and a complementary residual. The residual retains domain-sensitive cues with reduced class-semantic interference and supports domain routing, which in turn selects suitable domain-specific predictors. The semantic component preserves domain-agnostic class evidence to complement the routed prediction. In this way, LaCoD assigns complementary roles to the two components, enabling more reliable knowledge invocation and robust classification. Extensive experiments on four DIL benchmarks demonstrate the effectiveness of LaCoD, consistently outperforming prior methods.

1 INTRODUCTION

Continual learning (CL) aims to acquire new knowledge from sequentially arriving data while retaining previously learned knowledge (Rolnick et al., 2019; Wang et al., 2024b; Zhou et al., 2024). In contrast to classical class incremental learning, which expands the label space over time, domain incremental learning (DIL) focuses on continually learning from newly emerging domains while retaining knowledge of earlier ones, so that a single model can recognize the shared categories across all observed domains (Hsu et al., 2018; van de Ven et al., 2022). However, distribution shifts among these domains can substantially impair the model's discriminative and generalization capabilities across domains (Zhou et al., 2025). As more domains accumulate in the learning sequence, repeated adaptation to new distributions can intensify interference with knowledge acquired from previous domains. Consequently, catastrophic forgetting remains a critical challenge in DIL.

Existing DIL methods primarily mitigate catastrophic forgetting through two strategies: experience replay and parameter isolation. Experience-replay methods retain and revisit representative samples from earlier domains to stabilize previous knowledge during subsequent learning (Lopez-Paz & Ranzato, 2017; Buzzega et al., 2020). However, storing and repeatedly processing historical data incurs additional memory and computation costs and may be incompatible with data-retention restrictions in privacy-sensitive applications such as medical imaging (Price & Cohen, 2019). These limitations make parameter isolation an attractive alternative. Parameter-isolation methods preserve newly acquired domain knowledge in dedicated lightweight modules, such as prompts, adapters, and low-rank updates, while keeping previously learned parameters unchanged (Wang et al., 2022c; Houlsby et al., 2019; Hu et al., 2022). However, parameter isolation alone does not address how

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.

domain-adapted visual features should be converted into effective class predictions. Representative CLIP-based methods still derive class predictions from image–text similarity (Wang et al., 2022a; Nicolas et al., 2024), which may limit their classification performance across domains. Moreover, parameter isolation relies on accurate parameter selection at inference. Recognizing the impact of parameter selection accuracy on pretrained-model-based DIL, existing methods have explored several domain-prediction strategies, including k-nearest neighbors (Wang et al., 2022a), nearest mean classification (Song et al., 2024), and patch shuffle selection (Wang et al., 2025a). However, the accuracy of these domain-selection strategies tends to deteriorate as the number of learned domains increases. Additionally, the original visual features used for domain selection encode both category semantics and domain-related variations (Choi et al., 2026), making domain routing less reliable.

These limitations motivate us to look beyond existing parameter-isolation pipelines and explore a new way to address the DIL classification problem. CLIP aligns visual and textual representations in a shared feature space, allowing text features to serve as semantic references for exposing complementary visual information through feature subtraction. InstantStyle follows this principle by subtracting content-text features from CLIP image representations to extract residual information for style transfer (Wang et al., 2024a). Similarly, XIL uses the residual between image and class-text features to represent domain information for cross-domain knowledge transfer (Choi et al., 2026). However, in exemplar-free DIL, our objective is to continually learn across multiple target domains without forgetting the knowledge acquired from previous domains. Thus, how to effectively exploit category-related and domain-related information in visual representations remains an open question.

Building on this analysis, we propose **LaCoD**, a **La**nguage-Aware **Co**mplementary **D**ecomposition framework for exemplar-free domain-incremental learning. As shown in Fig. 1, LaCoD consists of three coordinated modules: language-guided feature decomposition (LFD), residual-guided domain-specific prediction (RDP), and semantic projection (SP). First, LFD constructs a semantic subspace from frozen CLIP class-text embeddings and decomposes each visual feature into a language-aligned semantic component that preserves class-level evidence and a complementary residual that suppresses dominant language-aligned semantics while exposing complementary variations. These representations provide specialized evidence for the subsequent modules. Second, RDP employs residual-enhanced domain routing to identify suitable domain-specific knowledge for each test sample. The residual provides domain-sensitive evidence with reduced interference from class semantics, thereby enabling the router to identify relevant domains more effectively. For each incoming domain, the module further learns a domain-specific prompt–classifier pair to encode the corresponding knowledge. The visual prompt adapts the frozen image encoder to the current domain. Compared with fixed text-prototype classifiers, the learnable classifiers provide greater flexibility in capturing the discriminative characteristics of individual domains. At inference time, the selected prompt–classifier pairs produce class predictions, which are combined according to their routing scores to obtain the routed prediction. Third, SP derives domain-independent class logits from the semantic component and adaptively combines them with the routed prediction. In contrast to methods that commit entirely to the selected domain-specific predictor, this module supplies additional class evidence when routing is ambiguous, thereby enabling more effective class prediction. In summary, our contributions are summarized as follows:

- We introduce a divide-and-conquer perspective that decomposes visual representations into language-semantic and domain-sensitive components, allowing them to play distinct yet complementary roles in class prediction and domain routing for exemplar-free DIL.
- We propose the LaCoD framework to coordinate the two components through residual-guided domain-specific prediction and semantic projection, enabling reliable domain routing and robust classification in exemplar-free DIL.
- Extensive experiments demonstrate that LaCoD achieves the best performance on DomainNet, ImageNet-R, ImageNet-C, and ImageNet-Mix among the compared methods.

2 Related Work

Domain-Incremental Learning. Continual learning (CL) aims to learn from a non-stationary data stream while mitigating catastrophic forgetting of previously acquired knowledge (Parisi et al., 2019; De Lange et al., 2022). Existing CL methods are commonly grouped into regularization-based approaches that constrain updates important to previous tasks (Li & Hoiem, 2016; Kirkpatrick et al.,

2017), replay-based approaches that rehearse stored or generated samples from earlier tasks (Lopez-Paz & Ranzato, 2017; Rolnick et al., 2019), and parameter-isolation approaches that reduce interference through task-specific masks or parameter subsets (Mallya & Lazebnik, 2018; Serrà et al., 2018). This work focuses on domain-incremental learning (DIL), where domains with different input distributions arrive sequentially while sharing a fixed label space (van de Ven et al., 2022). In DIL, replay methods that retain historical images conflict with the exemplar-free setting, while strong regularization can restrict adaptation under substantial domain shifts. Parameter-isolation strategies alleviate interference by introducing domain-specific components, but they can make task-agnostic inference challenging when the domain identity is unavailable at test time (Wang et al., 2022a; 2025a). Consequently, exemplar-free DIL requires models to learn new domains, preserve previously acquired knowledge, and classify samples from all observed domains without access to historical images or test domain identities.

Domain Identification and Routing in DIL. Domain identification is necessary in task-agnostic DIL when different domains use separate prompts or parameters. Existing routing methods mainly rely on prototype matching, learnable routing, or multi-candidate selection. S-Prompts (Wang et al., 2022a) matches test features to stored domain centroids, while PINA (Wang et al., 2025a) constructs prototypes from patch-shuffled images to suppress class-dependent cues. Learnable approaches such as L2P (Wang et al., 2022c), DualPrompt (Wang et al., 2022b), and SOYO (Wang et al., 2025b) infer task- or domain-specific components through learned keys or routing networks, whereas CODA-Prompt (Smith et al., 2023) and MoP-CLIP (Nicolas et al., 2024) combine multiple candidates using input-dependent weights. Despite different mechanisms, these methods largely derive routing signals from visual representations where class semantics and domain variations remain entangled.

Subspace Learning and Representation Decomposition. Subspace-based methods organize model updates or representations to reduce interference and improve transfer. GPM (Saha et al., 2021) and PGP (Qiao et al., 2024) constrain gradient updates using SVD-derived subspaces, whereas O-LoRA (Wang et al., 2023) and LoRanPAC (Peng et al., 2025) reduce interference through orthogonal task updates or truncated feature spaces. In vision–language models, TPPT (Lu et al., 2025) uses textual anchors for continual prompt tuning, while SSP (Zhu et al., 2024), COLA (Zhu et al., 2025), and HVTE (Song et al., 2025) exploit visual–textual projection or alignment to improve cross-modal representations. In contrast, LaCoD uses language-guided representation decomposition to address the distinct requirements of exemplar-free DIL.

3 METHOD

3.1 PROBLEM FORMULATION

We study exemplar-free domain-incremental learning (DIL), where a sequence of visual domains $\{\mathcal{D}_t\}_{t=1}^{T}$ arrives sequentially and all domains share the same label space $\mathcal{Y}$. At stage t, the learner receives training samples: $\mathcal{D}_t = \{(x_i^t, y_i^t)\}_{i=1}^{n_t}$, where $y_i^t \in \mathcal{Y}$. The domain identity is available during training but unavailable during inference. After learning domains $\mathcal{D}_{1:t}$, the model must classify samples from any observed domain without knowing their domain labels. Following the exemplar-free setting, historical images cannot be stored.

3.2 LANGUAGE-GUIDED FEATURE DECOMPOSITION

Effective task-agnostic DIL requires visual evidence to serve two competing goals: preserving class semantics for recognition and exposing domain-dependent variation for selecting domain-specific knowledge. Using a single visual feature for both goals may allow strong class semantics to obscure domain-sensitive cues, whereas simply removing semantic information would discard evidence useful for classification. This motivates a complementary decomposition that preserves both sources of information while making their distinct roles explicit.

We construct a shared semantic reference from the complete class vocabulary. Given class names $\{c_i\}_{i=1}^{C}$, we encode the template "a photo of a $\{class_name\}$" using the frozen CLIP text encoder and obtain class-text anchors $A = [a_1, \ldots, a_C]^\top \in \mathbb{R}^{C \times D}$, where C is the number of classes, D is the embedding dimension, and $a_i \in \mathbb{R}^D$ is the text embedding of the i-th class.

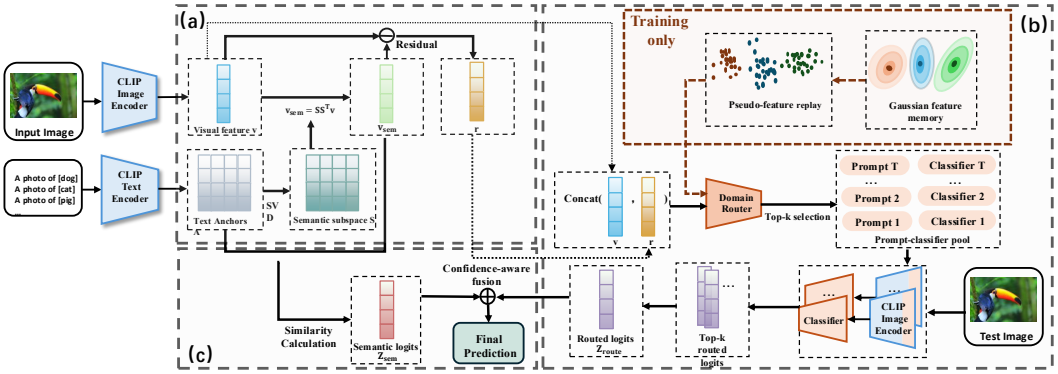


Figure 1: Overview of LaCoD. (a) Language-guided feature decomposition uses frozen CLIP text anchors to construct a semantic subspace and separates each visual representation into a language-aligned semantic component and a complementary residual. (b) Training and inference follow different paths. During sequential training, current-domain routing features are used to estimate class-conditional Gaussian statistics. The router is then trained on pseudo-features sampled from the memories of all observed domains, including the current domain. Only the current domain's prompt–classifier pair is optimized, while all previously learned pairs remain frozen. During task-agnostic inference, no domain identity is provided: the residual-enhanced router selects the top-k domain-specific prompt–classifier pairs, whose logits are aggregated using their routing weights. (c) Semantic projection derives domain-independent category logits from the semantic component and fuses them with the routed logits according to routing confidence.

To focus the semantic subspace on variations across classes, we normalize and center the embeddings as $\bar{A} = \mathrm{Norm}(A) - \mu^\top$, with $\mu = \frac{1}{C}\sum_{i=1}^{C} \mathrm{Norm}(a_i)$. As shown in Fig. 1(a), we then apply SVD, $\bar{A} = U\Sigma V^\top$, where the columns of V represent orthogonal semantic directions in the CLIP language space and the singular values $\{\lambda_i\}$ in Σ indicate their importance.

Instead of manually specifying the semantic dimension, we adaptively select the number of semantic directions according to the cumulative singular value energy. Specifically, the rank k is determined by:

$$k = \min\left\{ m : \frac{\sum_{i=1}^{m} \lambda_i^2}{\sum_{i=1}^{D} \lambda_i^2} \geq \eta \right\}, \tag{1}$$

where η is the retained energy ratio and is set to 0.95 in our implementation. The resulting basis $S = V_{[:,1:k]}$ spans the dominant class-semantic variation captured by the text anchors.

Given a visual feature $v \in \mathbb{R}^D$ extracted from the frozen CLIP image encoder, we project it onto the semantic subspace:

$$v_{\mathrm{sem}} = SS^\top v, \tag{2}$$

where v_{sem} represents the component of the visual feature aligned with class-level language semantics. The complementary residual is obtained by removing this component:

$$r = v - v_{\mathrm{sem}} = (I - SS^\top)v. \tag{3}$$

Since $S^\top S = I$, we have $S^\top r = 0$ and consequently $v_{\mathrm{sem}}^\top r = 0$. Thus, $v = v_{\mathrm{sem}} + r$ forms an orthogonal decomposition into a semantic component and a complementary residual.

Semantic Preservation and Reduced Class Bias. To characterize the roles of the two components, suppose a visual feature can be approximately written as $v = s_y + \delta + \epsilon$, where s_y denotes the class-semantic component associated with label y, δ denotes domain-dependent variation, and ϵ

represents other residual factors. If the dominant class-semantic component is largely contained in the language-guided subspace, i.e., $s_y \in \mathrm{span}(S)$, then

$$v_{\mathrm{sem}} = SS^\top v = s_y + SS^\top \delta + SS^\top \epsilon. \tag{4}$$

When the overlap between domain-dependent variation and the semantic subspace is limited, $SS^\top \delta$ remains relatively small, so v_{sem} preserves the dominant class-semantic component while suppressing variations outside the language-aligned subspace. This provides a principled basis for using v_{sem} as domain-independent semantic evidence for classification.

Conversely, consider two samples x_i and x_j from the same domain but different classes, with $v_i = v^i_{\mathrm{sem}} + r_i$, $v_j = v^j_{\mathrm{sem}} + r_j$. Because the semantic and residual components lie in orthogonal subspaces, $(v^i_{\mathrm{sem}} - v^j_{\mathrm{sem}})^\top (r_i - r_j) = 0$, and therefore

$$\|v_i - v_j\|^2 = \|v^i_{\mathrm{sem}} - v^j_{\mathrm{sem}}\|^2 + \|r_i - r_j\|^2. \tag{5}$$

The first term includes the variation induced by differences in class semantics, even when the two samples originate from the same domain. The residual suppresses this language-aligned variation and provides complementary evidence with reduced class-semantic interference.

3.3 Residual-Guided Domain-Specific Prediction

Residual-Enhanced Domain Routing. Reliable exemplar-free routing requires domain-sensitive evidence while preserving knowledge of previous domains without storing images. Combining v with r retains complete visual information and adds cues with reduced class-semantic interference; replaying pseudo-features from compact class-conditional Gaussian statistics maintains discrimination among previous domains. Specifically, as shown in Fig. 1(b), we concatenate the original visual feature v with the residual r:

$$h = \mathrm{Norm}([\mathrm{Norm}(v), \mathrm{Norm}(r)]) \in \mathbb{R}^{2D}. \tag{6}$$

The domain router uses a lightweight MLP g_ϕ to map the combined representation to logits over all observed domains:

$$q = g_\phi(h) \in \mathbb{R}^t, \tag{7}$$

where t is the number of domains observed so far. To preserve historical routing knowledge without storing images, we use class-conditional Gaussian feature replay. For each observed class in the current domain, we estimate a Gaussian mixture from its routing features and retain the component means, diagonal variances, and mixture weights. After incorporating the current-domain statistics, we sample pseudo-features from the distributions of all observed domains, including the current domain. Specifically, for class c observed in domain $j \le t$, we draw

$$\tilde{h} \sim \sum_{m=1}^{M_{j,c}} \pi_{j,c,m}\, \mathcal{N}\big(\mu_{j,c,m}, \mathrm{diag}(\sigma^2_{j,c,m})\big)\,, \tag{8}$$

where $M_{j,c}$ is the number of components, $\pi_{j,c,m}$ are within-class mixture weights summing to one, and $\mu_{j,c,m}$ and $\sigma^2_{j,c,m}$ are the component means and diagonal variances. The sampled feature–domain pairs form $\tilde{\mathcal{H}}_t$. The router is then optimized using

$$\mathcal{L}_{\mathrm{route}} = \frac{1}{|\tilde{\mathcal{H}}_t|} \sum_{(\tilde{h}, j) \in \tilde{\mathcal{H}}_t} \mathrm{CE}(g_\phi(\tilde{h}), j), \tag{9}$$

where $\tilde{h}$ is a sampled routing feature and j is its domain label. The retained statistics support routing across previously observed domains during subsequent learning stages.

Domain-Specific Predictor Learning. Domain shifts require predictors that can adapt their visual representations and class boundaries to each domain. We assign an independent visual prompt P^v_t to each incoming domain t while keeping the CLIP image encoder $\mathcal{E}_v$ frozen. Given an image x^t_i from domain t, the corresponding prompt is injected into the visual encoder to obtain a representation adapted to domain t:

$$u^t_i = \mathrm{Norm}\big(\mathcal{E}_v(x^t_i; P^v_t)\big)\,. \tag{10}$$

To learn class discrimination boundaries for each domain, we equip its prompted visual representation with a learnable **Linear Classifier** f_t. The classifier learns class discrimination boundaries directly from the prompted visual representations and produces

$$\ell_i^t = f_t(u_i^t) = W_t u_i^t + b_t, \tag{11}$$

where $W_t \in \mathbb{R}^{C\times D}$ and $b_t \in \mathbb{R}^C$ denote the weight and bias of the classifier for domain t, respectively. The current visual prompt and linear classification head are jointly optimized using

$$\mathcal{L}_{\text{cls}}^t = \frac{1}{n_t}\sum_{i=1}^{n_t} \text{CE}(\ell_i^t, y_i^t). \tag{12}$$

During training, the prompts and linear classification heads of previous domains remain frozen, forming a growing pool of prompt–classifier pairs for the routing mechanism.

At stage t, the router retains $K_t = \min(K, t)$ candidate domains as $\mathcal{K}(x) = \text{TopK}(q, K_t)$, with $K = 2$. Their routing weights are computed using a softmax:

$$\alpha_j = \frac{\exp(q_j/\tau)}{\sum_{j'\in\mathcal{K}(x)} \exp(q_{j'}/\tau)}, \quad j \in \mathcal{K}(x), \tag{13}$$

where τ controls the sharpness of the routing distribution. Each candidate domain j produces classification logits $\ell_j(x) = f_j(\text{Norm}(\mathcal{E}_v(x; P_j^v)))$ using its corresponding prompt–classifier pair. The candidate predictions are then aggregated according to the routing weights:

$$z_{\text{route}} = \sum_{j\in\mathcal{K}(x)} \alpha_j \ell_j(x). \tag{14}$$

3.4 Semantic Projection

We observe that domain routing cannot be perfectly accurate, and an incorrect or ambiguous selection can undermine the routed prediction. We therefore introduce a semantic projection branch that derives category evidence directly from v_{sem} without relying on the selected domain-specific predictors. When routing confidence is high, prediction is dominated by the routed branch; when confidence is low, the domain-independent semantic branch contributes more. Specifically, the semantic projection branch computes category logits from the similarity between the semantic component and the class text features(as shown in Fig. 1(c)):

$$z_{\text{sem}} = s \cdot \text{Norm}(v_{\text{sem}})\,\text{Norm}(A)^\top, \tag{15}$$

where s denotes the logit scale of CLIP. To determine the contribution of semantic projection, we measure the routing confidence using the difference between the two largest routing weights. For $K_t \geq 2$, sorting the routing weights in descending order gives the confidence

$$\gamma = \alpha_{(1)} - \alpha_{(2)}. \tag{16}$$

When $K_t = 1$, we set $\gamma = 1$ and use only the routed branch. Otherwise, larger γ indicates more certain routing. The final classification logits combine the routed prediction and semantic projection according to the routing confidence:

$$z = \gamma\frac{z_{\text{route}}}{\tau_{\text{route}}} + (1-\gamma)\frac{z_{\text{sem}}}{\tau_{\text{sem}}}, \tag{17}$$

where τ_{route} and τ_{sem} are the temperature parameters for the routed and semantic logits, respectively. Finally, the predicted label is obtained directly from the fused logits as $\hat{y} = \arg\max_j z_j$.

4 Experiments

4.1 Experimental Setup

Datasets. Following KA-Prompt (Xu et al., 2025), we adopt its benchmark construction and fixed domain sequences for four DIL benchmarks. DomainNet (Peng et al., 2019) contains 586,575 images from 345 categories across six visual domains. ImageNet-R (Hendrycks et al., 2021) contains

Table 1: Comparison with competing methods on four DIL benchmarks.

	DomainNet		ImageNet-R		ImageNet-C		ImageNet-Mix	
Method	AT ↑	AA ↑	AT ↑	AA ↑	AT ↑	AA ↑	AT ↑	AA ↑
L2P (Wang et al., 2022c)	54.58	50.48	67.48	64.26	49.87	49.87	53.24	52.13
DualPrompt (Wang et al., 2022b)	54.56	50.53	70.92	66.70	62.68	62.68	64.21	62.37
S-Prompts (Wang et al., 2022a)	67.34	63.22	66.94	64.83	68.07	68.07	65.32	62.89
PINA (Wang et al., 2025a)	70.97	72.46	73.07	68.40	74.51	74.51	73.27	71.87
DUCT (Zhou et al., 2025)	68.23	66.70	69.18	62.89	71.04	71.04	69.16	61.24
C-Prompt (Liu et al., 2024)	58.79	57.11	70.50	66.23	68.36	68.36	67.97	66.02
KA-Prompt (Xu et al., 2025)	64.33	63.21	71.12	67.35	74.25	74.25	70.16	68.82
LaCoD (Ours)	**74.32**	**74.76**	**82.42**	**79.44**	**76.62**	**76.62**	**79.40**	**76.45**

30,000 images from 200 categories across 15 rendition domains, using the released training/test splits with an approximate 7:3 ratio. ImageNet-C (Hendrycks & Dietterich, 2019) uses the same 200 categories across 15 corruption domains at severity level 5, with 7,000 training and 3,000 test images per domain. ImageNet-Mix (Liu et al., 2024) interleaves the ImageNet-C and ImageNet-R domain sequences, starting with ImageNet-C, to form 30 domains over the shared 200-category label space. Each domain constitutes one incremental stage.

Evaluation Metrics. We report average task accuracy (AT) and average accuracy (AA). AT averages the accuracy of each domain with equal weight and therefore reflects performance balance across domains. AA measures accuracy over all test samples, naturally weighting each domain by its number of samples. We additionally report routing accuracy ($\mathrm{Acc}_{\mathrm{route}}$), defined as the proportion of test samples assigned to the correct domain. Detailed formulations are provided in Appendix F.

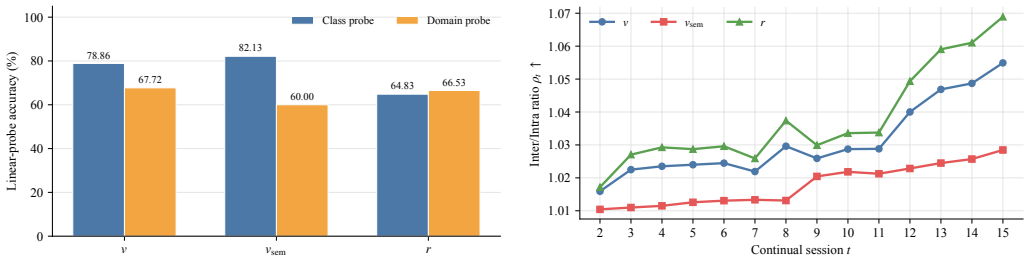
Implementation Details. We use CLIP ViT-B/16 (Radford et al., 2021) as the backbone and keep both its image and text encoders frozen. At each stage, we learn a domain-specific prompt comprising 16 context tokens together with its classifier. For the SVD decomposition, the rank k is selected as the smallest rank that preserves at least 95% of the cumulative energy of the text anchors. The residual-guided router is a two-layer MLP with a hidden dimension of 512 and dropout of 0.1. Domain prompts and classifiers are trained for 50 epochs using AdamW with a learning rate of 0.01, weight decay of 10^{-4}, a batch size of 128, and cosine decay with 10 warm-up iterations; the router is trained for 20 epochs with a learning rate of 10^{-3}. We use a routed-logit temperature of 0.5. For DomainNet, ImageNet-R, ImageNet-C, and ImageNet-Mix, the router temperatures are set to 4.0, 6.5, 10.0, and 6.5, while the semantic-logit temperatures are 2.0, 2.2, 8.0, and 2.2, respectively. At inference, the router retrieves the top-2 domain heads. Sensitivity analyses for the three inference temperatures are provided in Appendix G. We estimate class–domain routing distributions by K-means clustering with at most five components, capped by the available sample count, and sample ten pseudo-features per observed class in each domain. Component weights are proportional to cluster sizes; diagonal variances are floored at 10^{-4}, including for singleton clusters, and absent classes are omitted. Samples are drawn directly in the concatenated routing-feature space without further projection or normalization. The router is initialized anew at each stage. All experiments are conducted on a single NVIDIA RTX 4090 GPU.

4.2 Experimental Results

Main Results. Table 1 compares LaCoD with seven competing methods reimplemented under the same training and evaluation protocol. LaCoD consistently achieves the best performance across all four DIL benchmarks in terms of both AT and AA. The gains are particularly pronounced on ImageNet-R and ImageNet-Mix, where LaCoD improves AT by 9.35 and 6.13 percentage points over the strongest competing methods, respectively, together with substantial improvements in AA. Consistent gains are also observed on DomainNet and ImageNet-C, demonstrating that the performance advantage generalizes across different types of domain shifts.

Table 2: Representation analysis on ImageNet-R. Distances use separately L2-normalized representations and shared sample pairs (Appendix F).

Representation	Class probe ↑	Domain probe ↑	Intra-domain distance ↓	Inter/Intra ratio ↑
v	78.86	67.72	0.9113	1.0547
v_{sem}	**82.13**	60.00	1.2150	1.0279
r	64.83	66.53	**0.8233**	**1.0688**
$[v, r]$	—	**69.11**	—	—



(a) (b)

Figure 2: Representation characteristics on ImageNet-R. (a) Class- and domain-probe accuracy. (b) Inter-/intra-domain distance ratio over continual sessions; the residual maintains the highest ratio.

Table 3: Domain-routing accuracy ($\text{Acc}_{\text{route}}$, %) on four DIL benchmarks.

Method	DomainNet	ImageNet-R	ImageNet-C	ImageNet-Mix
KNN (Wang et al., 2022a)	84.33	49.63	73.45	72.54
NMC (Rebuffi et al., 2017)	84.99	56.78	78.36	70.07
PSS (Wang et al., 2025a)	85.02	52.34	73.42	63.34
SOYO (Wang et al., 2025b)	85.95	57.11	94.70	81.35
RDR (Ours)	**88.17**	**63.90**	**95.60**	**90.10**

Representation Analysis. We analyze ImageNet-R representations using class and domain linear probes together with intra-domain inter-class distance and the inter-/intra-domain distance ratio, where a higher ratio indicates better domain separability (definitions are provided in Appendix F; dashes in Table 2 denote unevaluated metrics). Table 2 and Fig. 2 validate the complementary roles predicted by our analysis: v_{sem} raises class-probe accuracy from 78.86% to 82.13%, whereas r suppresses class-related information while preserving domain cues, achieving the lowest intra-domain distance and the highest inter-/intra-domain ratio throughout the continual sequence. Combining both cues in $[v, r]$ further yields the highest domain-probe accuracy of 69.11%, supporting their coordinated use for classification and domain routing.

Domain Routing Analysis. Table 3 compares our residual-enhanced domain router (RDR) with representative domain-selection methods using $\text{Acc}_{\text{route}}$. RDR achieves the highest routing accuracy on all four benchmarks, outperforming the strongest competing selector by 2.22, 6.79, 0.90, and 8.75 percentage points on DomainNet, ImageNet-R, ImageNet-C, and ImageNet-Mix, respectively. These results demonstrate that RDR provides more reliable domain-discriminative evidence across diverse domain shifts.

Semantic Projection Analysis. We examine whether the semantic pathway specifically mitigates routing failures by partitioning ImageNet-R test samples according to the correctness of the top-1 domain prediction. As shown in Table 4, semantic projection improves accuracy from 61.70% to 66.12% on incorrectly routed samples, corresponding to a gain of 4.41 percentage points. In

Table 4: Classification accuracy (%) under correct and incorrect routing on ImageNet-R.

Prediction	Correct routing	Incorrect routing	Overall
Routed branch	86.52	61.70	77.56
+ Semantic projection	**86.97**	**66.12**	**79.44**
Gain (pp)	+0.45	**+4.42**	**+1.88**
Samples	5,135	2,901	8,036

Table 5: Component and feature-input ablations on ImageNet-R and ImageNet-C. RDR(v) replaces the default routing input $[v, r]$ with v, while SP(v) replaces the default semantic input v_{sem} with v.

	ImageNet-R		ImageNet-C	
Variant	AT ↑	AA ↑	AT ↑	AA ↑
Baseline	66.94	64.83	68.07	68.07
+ LC	78.46	73.72	71.05	71.05
+ LC + RDR(v)	80.72	77.16	76.21	76.21
+ LC + RDR	81.08	77.56	76.48	76.48
+ LC + SP	78.98	74.21	72.88	72.88
+ LC + RDR + SP(v)	81.90	78.23	76.35	76.35
+ LC + RDR + SP (LaCoD)	**82.42**	**79.44**	**76.62**	**76.62**

contrast, the gain is 0.45 points for correctly routed samples, while the overall accuracy improves by 1.88 points. This indicates that the semantic pathway is particularly effective when domain selection is unreliable, supporting its role as a domain-independent branch rather than merely a generic ensemble component.

Ablation Study. Table 5 evaluates each component on ImageNet-R and ImageNet-C, using the original S-Prompts as the baseline while retaining its domain-specific visual prompts and training procedure. We first add LC to form the prompt–classifier pairs used by the subsequent modules; its clear gains confirm the benefit of explicitly learning domain-specific class boundaries. On this common foundation, RDR and SP further improve performance by exploiting domain-sensitive and class-semantic cues, respectively. The feature-input variants also validate the language-guided decomposition: RDR performs best with the residual-enhanced representation $[v, r]$, whereas SP benefits most from the semantic component v_{sem}. This demonstrates that the decomposition provides specialized and complementary representations for routing and classification, with their complete combination achieving the best results on both benchmarks.

5 Conclusion

Domain-incremental learning requires models to continually adapt to changing data distributions within a shared label space, while appearance variations of the same categories across domains make reliable category recognition more challenging. Thus, we propose a novel framework LaCoD to address the typical catastrophic forgetting problem in domain-incremental learning. Unlike existing methods that operate directly on complete visual features, LaCoD explores language semantics as a reference to decompose each visual feature into orthogonal and complementary semantic and residual components for improving class recognition and domain discrimination. Residual-guided domain-specific prediction enables effective domain selection and domain-specific knowledge learning. Semantic projection extracts domain-independent category evidence from the semantic component to complement the main prediction branch and improve classification robustness. Extensive experiments on four domain-incremental learning benchmarks demonstrate the effectiveness and superiority of LaCoD.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing, literature retrieval, brainstorming of research ideas and experimental considerations, and drafting selected parts of the manuscript. All methodological decisions, experiments, results, and conclusions were independently reviewed and verified by the authors, who take full responsibility for the content of this paper.

REFERENCES

Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. In *Advances in Neural Information Processing Systems*, volume 33, pp. 15920–15930. Curran Associates, Inc., 2020.

Heayoun Choi, Hyundong Jin, and Eunwoo Kim. Xil: Cross-expanding incremental learning. In *International Conference on Learning Representations*, volume 2026, pp. 18799–18814, 2026.

Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3366–3385, 2022.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019.

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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8320–8329, 2021.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2790–2799. PMLR, 2019.

Yen-Chang Hsu, Yen-Cheng Liu, Anita Ramasamy, and Zsolt Kira. Re-evaluating continual learning scenarios: A categorization and case for strong baselines. In *NeurIPS Continual Learning Workshop*, 2018.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

Zhizhong Li and Derek Hoiem. Learning without forgetting. In *European Conference on Computer Vision*, volume 9908 of *Lecture Notes in Computer Science*, pp. 614–629. Springer, 2016.

Zichen Liu, Yuxin Peng, and Jiahuan Zhou. Compositional prompting for anti-forgetting in domain incremental learning. *International Journal of Computer Vision*, 132(12):5783–5800, 2024.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Haodong Lu, Xinyu Zhang, Kristen Moore, Jason Xue, Lina Yao, Anton van den Hengel, and Dong Gong. Continual learning on CLIP via incremental prompt tuning with intrinsic textual anchors. *Transactions on Machine Learning Research*, 2025.

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*, pp. 7765–7773, 2018.

Julien Nicolas, Florent Chiaroni, Imtiaz Ziko, Ola Ahmad, Christian Desrosiers, and Jose Dolz. MoP-CLIP: A mixture of prompt-tuned CLIP models for domain incremental learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1751–1761, 2024.

German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019.

Liangzu Peng, Juan Elenter, Joshua Agterberg, Alejandro Ribeiro, and René Vidal. LoRanPAC: Low-rank random features and pre-trained models for bridging theory and practice in continual learning. In *International Conference on Learning Representations*, 2025.

Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1406–1415, 2019.

W. Nicholson Price, II and I. Glenn Cohen. Privacy in the age of medical big data. *Nature Medicine*, 25(1):37–43, 2019.

Jingyang Qiao, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yong Peng, and Yuan Xie. Prompt gradient projection for continual learning. In *International Conference on Learning Representations*, volume 2024, pp. 16294–16315, 2024.

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 *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021.

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*, pp. 5533–5542, 2017.

David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Gobinda Saha, Isha Garg, and Kaushik Roy. Gradient projection memory for continual learning. In *International Conference on Learning Representations*, 2021.

Joan Serrà, Dídac Surís, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4548–4557. PMLR, 2018.

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*, pp. 11909–11919, 2023.

Xiang Song, Yuhang He, Songlin Dong, and Yihong Gong. Non-exemplar domain incremental object detection via learning domain bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 15056–15065, 2024.

Yeji Song, Jimyeong Kim, Wonhark Park, Wonsik Shin, Wonjong Rhee, and Nojun Kwak. Harmonizing visual and textual embeddings for zero-shot text-to-image customization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20549–20557, 2025. doi: 10.1609/aaai.v39i19.34264.

Gido M. van de Ven, Tinne Tuytelaars, and Andreas S. Tolias. Three types of incremental learning. *Nature Machine Intelligence*, 4(12):1185–1197, 2022.

Haofan Wang, Matteo Spinelli, Qixun Wang, Xu Bai, Zekui Qin, and Anthony Chen. InstantStyle: Free lunch towards style-preserving in text-to-image generation. *arXiv preprint arXiv:2404.02733*, 2024a.

Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: Theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5362–5383, 2024b.

Qiang Wang, Yuhang He, Songlin Dong, Xinyuan Gao, Shaokun Wang, and Yihong Gong. Non-exemplar domain incremental learning via cross-domain concept integration. In *European Conference on Computer Vision*, volume 15107 of *Lecture Notes in Computer Science*, pp. 144–162. Springer, 2025a.

Qiang Wang, Xiang Song, Yuhang He, Jizhou Han, Chenhao Ding, Xinyuan Gao, and Yihong Gong. Boosting domain incremental learning: Selecting the optimal parameters is all you need. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4839–4849, 2025b.

Xiao Wang, Tianze Chen, Qiming Ge, Han Xia, Rong Bao, Rui Zheng, Qi Zhang, Tao Gui, and Xuanjing Huang. Orthogonal subspace learning for language model continual learning. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 10658–10671. Association for Computational Linguistics, 2023.

Yabin Wang, Zhiwu Huang, and Xiaopeng Hong. S-Prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5682–5695, 2022a.

Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. DualPrompt: Complementary prompting for rehearsal-free continual learning. In *European Conference on Computer Vision*, volume 13686 of *Lecture Notes in Computer Science*, pp. 631–648. Springer, 2022b.

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*, pp. 139–149, 2022c.

Kunlun Xu, Xu Zou, Gang Hua, and Jiahuan Zhou. Componential prompt-knowledge alignment for domain incremental learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 70032–70046. PMLR, 2025.

Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Class-incremental learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9851–9873, 2024.

Da-Wei Zhou, Zi-Wen Cai, Han-Jia Ye, Lijun Zhang, and De-Chuan Zhan. Dual consolidation for pre-trained model-based domain-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20547–20557, 2025.

Xingyu Zhu, Beier Zhu, Yi Tan, Shuo Wang, Yanbin Hao, and Hanwang Zhang. Selective vision-language subspace projection for few-shot CLIP. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 3848–3857. Association for Computing Machinery, 2024.

Xingyu Zhu, Beier Zhu, Shuo Wang, Kesen Zhao, and Hanwang Zhang. Enhancing CLIP robustness via cross-modality alignment. In *Advances in Neural Information Processing Systems*, volume 38, pp. 17553–17576. Curran Associates, Inc., 2025.