Cross-Domain Knowledge Separation and Positive Transmission for Noisy Domain Incremental Learning

Kunlun Xu[*]
Wangxuan Institute of Computer Technology
Peking University
xkl@stu.pku.edu.cn

Zhengyuan Cai[*]
Wangxuan Institute of Computer Technology
Peking University
caizhengyuan@stu2023.jnu.edu.cn

Jiahuan Zhou [†]
Wangxuan Institute of Computer Technology
Peking University
jiahuanzhou@pku.edu.cn

Abstract

Domain Incremental Learning (DIL) aims to continuously adapt to new domains while retaining the knowledge of previous domains. Existing DIL methods are generally built upon an idealized assumption of clean supervision. However, real-world data streams are often affected by label noise, giving rise to the more challenging Noisy Domain Incremental Learning (N-DIL) scenario. Under this setting, not only is intra-domain knowledge acquisition hindered, but inter-domain knowledge conflicts are also exacerbated, which amplifies catastrophic forgetting. To address these challenges, we propose a novel Cross-Domain Knowledge Separation and Positive Transmission (ST-Prompt) framework. Specifically, to mitigate inter-domain knowledge conflicts, a Cross-domain Prompt Knowledge Contrastive Isolator is developed to enhance domain-wise knowledge separation, thereby mitigating the cross-domain knowledge interference during inference. Furthermore, to improve intra-domain knowledge acquisition, a Synergistic Knowledge Transmission scheme is introduced, which extracts reliable knowledge from previous domains to facilitate noisy data learning in the current domain. These two components are mutually reinforcing, jointly promoting effective cross-domain knowledge learning and utilization. Extensive experiments on diverse benchmarks demonstrate that ST-Prompt achieves state-of-the-art performance. Our code is available at https://github.com/PKU-OV3-LAB/NeurIPS2026-ST-Prompt.

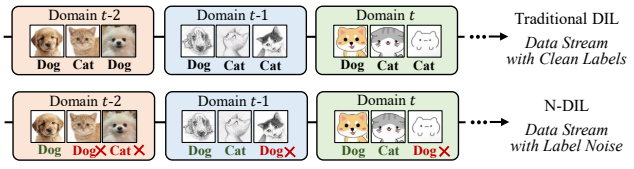
1 Introduction

Domain incremental learning (DIL) aims to adapt a model continuously to a sequential stream of distinct domains [1; 2]. Its central challenge is catastrophic forgetting, where learning new domain representations inadvertently overwrites previously acquired knowledge [3; 4]. As shown in Fig. 1, existing DIL works [5; 6; 7] are typically built upon an idealized assumption of clean supervision, where each sample is associated with accurate and reliable labels. However, in real-world scenarios, data streams are often corrupted by label noise, arising from unreliable automated data collection (e.g., web scraping) and human annotation errors [8; 9; 10].

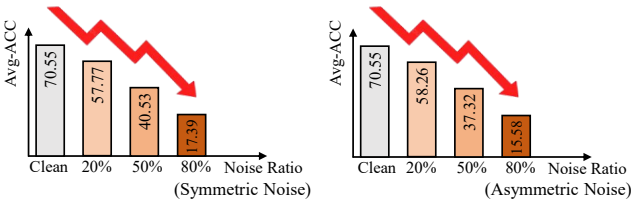
[*]Equal Contribution
[†]Corresponding author

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

This practical yet challenging scenario, which we introduce as Noisy Domain Incremental Learning (N-DIL), precipitates unprecedented dual challenges between intra-domain noise robustness and inter-domain anti-forgetting ability. Specifically, as shown in Figure 1 (b), the label noise within a domain inevitably introduces erroneous knowledge, degrading the intra-domain performance. Furthermore, although the prompt-based methods have shown advanced anti-forgetting capacity in traditional DIL, they suffer severe forgetting in N-DIL. As shown in Fig. 2, due to the blurred inter-domain boundary, the domain-wise instance-prompt mismatch frequently occurs during inference. Since the knowledge interrupted by label noise is less generalizable for other-domain samples, the mismatch leads to extreme catastrophic forgetting [11; 12; 13].



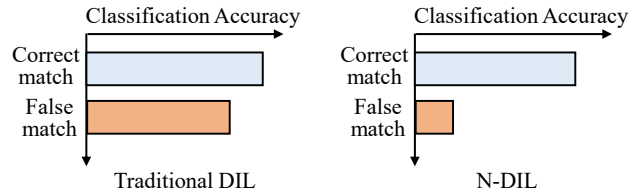
(a) Comparison of traditional DIL and N-DIL



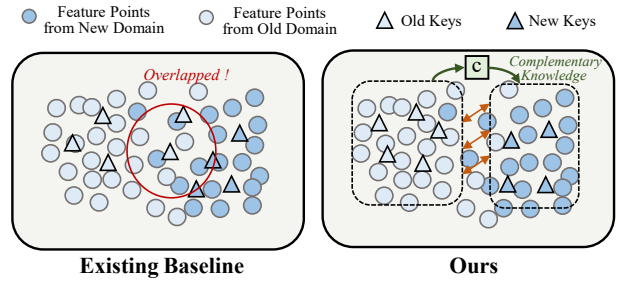
(b) Influence of data label noise

Figure 1: (a) Compared to the traditional DIL task, the N-DIL task investigates continuous data streams plagued by noisy labels. (b) As the label noise ratio increases, the performance of the existing method suffers severe degradation.

To overcome the above challenges, we introduce a novel Cross-Domain Knowledge **S**eparation and Positive **T**ransmission (**ST-Prompt**) approach. Our key idea lies in improving inter-domain knowledge separation to mitigate the influence of inter-domain conflict and exploiting the inter-domain complementary correct knowledge to facilitate correct knowledge acquisition in subsequent domains. Specifically, ST-Prompt comprises a Cross-domain Prompt Knowledge Contrastive Isolator mechanism (PKCI) and a Synergistic Knowledge Transmission scheme (SKT). PKCI enhances domain-wise knowledge separation to mitigate cross-domain knowledge interference during inference, thereby mitigating the influence of inter-domain knowledge conflicts. SKT extracts reliable knowledge from previous domains to facilitate learning in the current domain, thereby improving intra-domain knowledge acquisition. These two components are mutually complementary, jointly promoting effective cross-domain knowledge learning and utilization. Comprehensive experiments on diverse benchmarks demonstrate that ST-Prompt exhibits increasing advantages in severe noisy conditions.



(a) Influence of instance-prompt matching on Classification Accuracy



(b) Comparison of existing baseline and our approach

Figure 2: (a) Under the label noise scenario, the learned knowledge exhibits exacerbated inter-domain conflicts. (b) State-of-the-art baselines suffer from catastrophic prompt mismatch due to blurred domain boundaries and overlapping key spaces, whereas our approach achieves cross-domain knowledge separation.

In summary, the contributions of this work are threefold: (1) We present a pioneer investigation into the practical and challenging Noisy DIL task and introduce a ST-Prompt approach to overcome exacerbated catastrophic forgetting and limited acquisition capacity simultaneously. (2) We develop a Cross-domain Prompt Knowledge Contrastive Isolator to achieve strict domain-level knowledge separation. Besides, a Synergistic Knowledge Transmission scheme is introduced to improve noisy

new data by integrating historical clean knowledge. (3) Extensive experiments conducted on three cross-domain DIL benchmarks demonstrate that our ST-Prompt significantly outperforms all existing methods that combine DIL approach with LNL techniques.

2 Related Work

2.1 Domain Incremental Learning

Domain Incremental Learning (DIL) aims to continuously acquire knowledge from non-stationary data streams across varying domains while preserving historical representations [14; 15]. In this setting, the category label space remains identical across domains, but the underlying data distributions—such as visual styles, sensor modalities, or physical environments—exhibit significant shifts. These severe distribution shifts exacerbate catastrophic forgetting [16], which remains the fundamental challenge in continuous learning [3; 17; 18; 19; 20]. Existing DIL methods typically rely on data replay or parameter regularization to mitigate this forgetting. Replay-based approaches rehearse stored historical exemplars[1; 21; 22], which effectively maintains past knowledge but introduces critical privacy risks and prohibitive storage overheads in scalable scenarios[23]. Conversely, regularization-based methods penalize updates to critical historical weights [24; 25; 26; 27]. However, these rigid constraints often severely restrict the model's plasticity, hindering its capacity to adapt to new domains[28]. Recently, prompt-based methods [29; 6; 30], particularly when built on a pre-trained Vision Transformer (ViT) backbone, have emerged as a highly effective paradigm for DIL[5; 7]. These architectures maintain a pool of lightweight, learnable, and domain-specific prompts. By conditionally matching and composing prompts from different domains during inference, they facilitate the transfer of shared knowledge across domains to mitigate forgetting.

2.2 Label Noise Learning

Label Noise Learning (LNL) has been extensively investigated in recent years to construct robust models when the training data contains corrupted annotations [31; 8; 32]. Existing LNL methods have evolved into several mainstream paradigms. Some category relies on robust loss functions, which aim to alleviate confirmation bias by assigning lower weights to instances with large losses or gradients [33; 34]. Some focuses on sample selection and semi-supervised learning [9; 35]. For instance, Gaussian Mixture Models (GMM) were used to divide data into clean and noisy subsets, subsequently treating the noisy branch as unlabeled data for consistency regularization [10]. Alternatively, label correction and pseudo-labeling strategies attempt to dynamically rectify erroneous annotations based on iterative model predictions [36]. More recently, unsupervised representation and contrastive learning have been introduced to extract robust features without relying heavily on corrupted semantic labels [37; 38]. However, existing studies have revealed an inherent phenomenon: noise samples near the decision boundary often exhibit similar performance in prediction scores or regularization effects to clean samples, which makes standard selection and correction inaccurate [39]. When these LNL techniques are directly applied to continuous learning tasks, the randomly fused knowledge between different learning stages will exacerbate semantic confusion. The inherent noise not only misleads the acquisition of new tasks but also irreversibly contaminates the shared representation space and continuously accumulates incorrect information, making the catastrophic forgetting problem still very important.

2.3 Noisy Continual Learning

Despite the success of recent Domain Incremental Learning (DIL) paradigms, they implicitly rely on the idealized assumption of perfectly annotated data streams. In real-world deployments, inevitable label noise precipitates a severe stability-plasticity dilemma within the Noisy Domain Incremental Learning setting [12]. To mitigate catastrophic forgetting and adapt to new domains, DIL models demand high plasticity. However, this aggressive fitting forces models to overfit label noise, resulting in the undesirable memorization of erroneous knowledge. Conversely, standard Label Noise Learning techniques overly restrict plasticity, causing under-fitting on new tasks and indirectly exacerbating historical forgetting. Thus, a naïve integration of DIL and LNL fails to balance intra-domain noise robustness with inter-domain anti-forgetting capabilities. As sequential learning progresses, the blurring of inter-domain boundaries corrupts the shared semantic representation space [40], leading to a continuous accumulation of errors. Consequently, simultaneously preventing the catastrophic

forgetting of correct knowledge and the acquisition of erroneous knowledge remains a critical challenge in prompt-based DIL. While recent works explore noisy continual learning, these methods are predominantly replay-based; we therefore exclude them to ensure fair comparisons within our strictly replay-free paradigm.

3 The Proposed Method

3.1 Problem Definition and Formulation

In Noisy Domain Incremental Learning (N-DIL), a model learns from a sequential stream of T domains with label noise, $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_T\}$. At step t, only the current domain is accessible: $\mathcal{D}_t = \{(x_{t,i}, \tilde{y}_{t,i})\}_{i=1}^{N_t}$. Let $\tilde{y}_{t,i} \in \mathcal{C}$ be the observed noisy label and $y_{t,i} \in \mathcal{C}$ the latent clean label, with overall noise ratio η. The corruption process is:

$$\mathbb{P}(\tilde{y}_{t,i} = j \mid y_{t,i} = k) = \begin{cases} 1 - \eta, & j = k, \\ \eta_{kj}, & j \neq k, \end{cases} \tag{1}$$

where η_{kj} is the probability of mislabeling class k as class j, and $\sum_{j \neq k} \eta_{kj} = \eta$. Depending on the deterministic distribution of η_{kj}, the label corruption is categorized into Symmetric Noise, Asymmetric Noise or Instance-dependent Noise. According to the basic setting of DIL, all domains share the same label space $\mathcal{C}$ but have shifted visual distributions. Past domains $\{\mathcal{D}_1, \ldots, \mathcal{D}_{t-1}\}$ are unavailable when learning $\mathcal{D}_t$. We denote by M_t the model after learning the first t domains. In our prompt-based setting, the model maintains a reusable prompt pool $\mathcal{P}_t = \{(k_m, p_m)\}_{m=1}^{M_t}$ with key–prompt pairs. For evaluation, each training domain $\mathcal{D}_t$ is noisy, while each test set $\mathcal{D}_t^{\text{test}}$ is strict clean with correct labels.

3.2 ST-Prompt Approach Overview

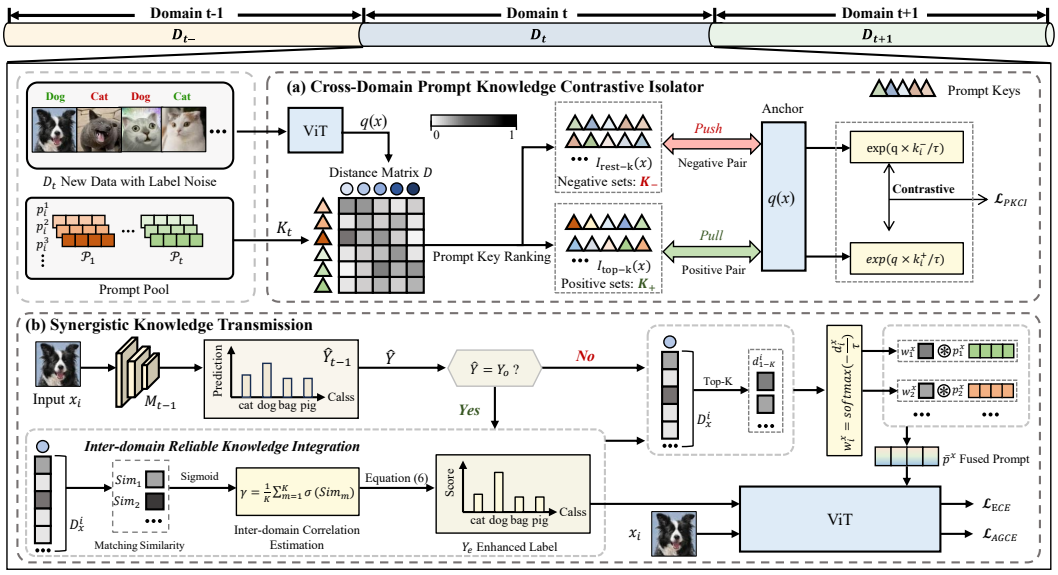


Figure 3: The illustration of our ST-Prompt approach

As illustrated in Figure 3, we propose the **ST-Prompt** framework which achieves an explicit key-level separation and prompt-level positive reusable knowledge transmission to strictly insulate cross-domain knowledge while enabling secure knowledge transmission to inherit credible historical experiences. Specifically, the framework comprises two sequential components: (1) a **Cross-Domain Prompt Knowledge Contrastive Isolator** that matches the queries of images with a global prompt pool via similarity. Then pull in the top-k relevant queries and keys, and exclude the irrelevant ones to construct a non-overlapping knowledge space; and (2) a **Synergistic Knowledge Transmission** module that dynamically routes optimization pathways according to the label predicted from the

previous domain model, reserving knowledge transmission exclusively for clean samples while insulating against noise. To theoretically validate the effectiveness of our proposed approach, we provide a formal proof based on a simplified statistical model in Appendix A.

3.3 Cross-Domain Prompt Knowledge Contrastive Isolator

In our DIL framework, ST-Prompt maintains a continuous prompt pool $\mathcal{P} = \{(k_m, p_m)\}_{m=1}^{M}$ of M key-prompt pairs. For an input image $x \in \mathcal{D}_t$ from the current domain, we extract its query feature $q(x) \in \mathbb{R}^d$ (d is the embedding dimension) using a frozen pre-trained Vision Transformer (ViT) as the backbone, i.e., $q(x) = \text{ViT}(x)$.

At the first iteration of each domain, we initialize the current prompt pool P_t (with its key set denoted as $\mathcal{K}_t$) using historical pools $\bigcup_{i=1}^{t} \mathcal{P}_i$ via similarity matching and greedy search. During training, given a batch of B queries, we construct a cosine similarity matrix $C \in \mathbb{R}^{B \times M}$ between the queries and all keys $k_j \in \mathcal{K}_t$. Each element $C_{i,j}$ is calculated as:

$$C_{i,j} = \text{sim}(q(x_i), k_j) = \frac{q(x_i)^\top k_j}{\|q(x_i)\| \|k_j\|}. \tag{2}$$

Based on C, we select the top-K keys with the highest similarity scores, storing their indices as $\mathcal{I}_{\text{top-}K}(x_i)$. The corresponding distance matrix is obtained by $D_{i,j} = 1 - C_{i,j}$.

For a given query $q(x_i)$, we define the selected top-K keys as the positive set $\mathcal{K}_i^+ = \{k_j \mid j \in \mathcal{I}_{\text{top-}K}(x_i)\}$, and the remaining keys as the negative set $\mathcal{K}_i^- = \mathcal{K}_t \setminus \mathcal{K}_i^+$, treating the frozen query $q(x_i)$ as an invariant anchor. We formulate the matching probability distribution over the key space using a temperature-scaled softmax function. Specifically, the probability of the query being correctly matched to a specific positive key $k^+ \in \mathcal{K}_i^+$ against all available candidates is defined as:

$$p(k^+ \mid x_i) = \frac{\exp\left(\text{sim}(q(x_i), k^+)/\tau\right)}{\exp\left(\text{sim}(q(x_i), k^+)/\tau\right) + \sum_{k^- \in \mathcal{K}_i^-} \exp\left(\text{sim}(q(x_i), k^-)/\tau\right)}, \tag{3}$$

where τ is the temperature scalar. The model simultaneously pulls the positive keys towards the query while explicitly repelling all negative keys by maximizing this probability. The overall PKCI loss is:

$$\mathcal{L}_{PKCI} = -\frac{1}{B \cdot K} \sum_{i=1}^{B} \sum_{k^+ \in \mathcal{K}_i^+} \log p(k^+ \mid x_i). \tag{4}$$

where B is the batch size, K is the number of selected positive prompts per query. By minimizing $\mathcal{L}_{PKCI}$, the currently active keys are forced to explore novel representation sub-spaces explicitly repelled by the historical negative anchors. This mechanism constructs a precise and non-overlapping knowledge space, ensuring that cross-domain knowledge is strictly separated at the key-level.

3.4 Synergistic Knowledge Transmission

To determine the credibility of the observed original label Y_o, we employ the frozen historical model from the previous domain M_{t-1} to generate a pseudo-label $\widehat{Y}_{t-1} = \arg\max M_{t-1}(x)$. The training optimization is subsequently divided into two distinct complementary branches based on the judgement $\widehat{Y}_{t-1} = Y_o$.

Clean Branch with Knowledge Transmission: If the pseudo-label matches the hard label ($\widehat{Y}_{t-1} = Y_o$), the sample is deemed credible. To securely inherit historical knowledge and prevent overfitting to the current domain, these samples are routed to a specialized *Inter-domain Reliable Knowledge Integration* module. Specifically, a dynamic inter-domain correlation estimation score $\gamma \in [0, 1]$ is computed by aggregating the sigmoid-activated similarities $sim_{i,m}$ of the top-K keys:

$$\gamma = \frac{1}{K} \sum_{m=1}^{K} \sigma(sim_{i,m}). \tag{5}$$

A higher γ indicates greater similarity between the sample query and the selected prompt keys, and places more weight on the observed label. For samples on which the historical prediction agrees with

the observed label, we construct a dynamically smoothed target:

$$\mathbf{Y}_e = \gamma\, \mathbf{e}_{Y_o} + (1-\gamma)\, \mathbf{u}, \qquad \mathbf{u} = \frac{1}{C}\mathbf{1}_\mathbf{c}, \tag{6}$$

where $\mathbf{e}_{Y_o}$ is the one-hot vector of the observed label and C is the number of classes. Then, to necessarily exploit the retrieved prompts $\mathcal{P}_{\text{top-}K} = \{p_j \mid j \in \mathcal{I}_{\text{top-}K}(x)\}$ against label noise, we dynamically aggregate them into a fused prompt $\bar{p}^x$. As illustrated in Figure 3, this is achieved via a distance-weighted softmax function:

$$w_j^x = \text{Softmax}\left(-\frac{D_{i,j}}{\tau_{fuse}}\right), \quad \bar{p}^x = \sum_{j \in \mathcal{I}_{\text{top-}K}(x)} w_j^x \cdot p_j, \tag{7}$$

where $D_{i,j}$ denotes the distance between the query q and the j-th key, and τ_{fuse} is the temperature scaling factor. The fused prompt $\bar{p}^x$ is then prepended to the ViT input sequence to produce the model's prediction.

The model is optimized subsequently using the Enhanced Cross-Entropy loss:

$$\mathcal{L}_{E-CE} = -\sum_c Y_e^{(c)} \log P(y=c|x), \tag{8}$$

Here, $\mathbf{Y}_e$ provides a similarity-based smoothed target for samples whose historical predictions agree with their observed labels. The resulting loss discourages overconfident fitting to those labels while preserving their class supervision; historical prompt knowledge contributes through the fused prompt in the forward pass. "'

Noisy Branch: If a mismatch occurs ($\widehat{Y}_{t-1} \neq Y_o$), the given label is highly suspicious. For these samples, the *Inter-domain Reliable Knowledge Integration* module is bypassed. To truncate the misleading gradients while still preserving the robust feature representation of the uniformly fused prompt $\bar{p}^x$, the sample is routed to the Adaptive Generalized Cross Entropy loss branch:

$$\mathcal{L}_{AGCE} = \frac{1 - P_{y_o}^q}{q}, \tag{9}$$

where P_{y_o} is the predicted probability of the given observed label, and $q \in (0, 1]$ controls the gradient truncation intensity. This branch effectively insulates the backbone and prompts from noise contamination.

By unifying the classification pathways for credible and suspicious samples, we obtain the branch-conditional classification loss $\mathcal{L}_{cls}$. For an input sample x with hard label Y_o and pseudo-label $\widehat{Y}_{t-1}$, it is defined as:

$$\mathcal{L}_{cls} = \begin{cases} \mathcal{L}_{E-CE}(x, Y_e), & \text{if } \widehat{Y}_{t-1} = Y_o, \\ \mathcal{L}_{AGCE}(x, Y_o), & \text{if } \widehat{Y}_{t-1} \neq Y_o \end{cases} \tag{10}$$

During training, when learning the initial domain, the model is optimized with the ViT using the overall loss

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_{pw}\mathcal{L}_{pw} + \lambda_{PKCI}\mathcal{L}_{PKCI}, \tag{11}$$

where $\mathcal{L}_{cls}$ is the classification loss, $\mathcal{L}_{pw}$ is the prompt matching penalty, and $\mathcal{L}_{PKCI}$ is the cross-domain prompt knowledge contrastive isolator loss. When training on domain t ($t > 1$), the new prompt slots are first initialized by greedy historical prompt search. Then, the same optimization pipeline is applied.

4 Experiment

4.1 Benchmark

In this paper, following the existing Domain Incremental Learning and Label Noise Learning works, a Noisy DIL benchmark is proposed as below.

Datasets: We evaluate our method on three benchmark datasets: DomainNet, ImageNet-R and ImageNet-C. Detailed information about each dataset is provided in the appendix D.

Table 1: The Avg-ACC and Avg-F performance comparison under varying symmetric noise ratios across three Noisy DIL benchmarks.

r	Type	Method	Publication	ImageNet-R		ImageNet-C		DomainNet		Average	
				$\mathcal{A}_T$ (↑)	$\mathcal{F}_T$ (↓)	$\mathcal{A}_T$ (↑)	$\mathcal{F}_T$ (↓)	$\mathcal{A}_T$ (↑)	$\mathcal{F}_T$ (↓)	$\mathcal{A}_T$ (↑)	$\mathcal{F}_T$ (↓)
20%	DIL	CODA-Prompt [7]	*CVPR 2023*	40.29	12.31	51.42	5.10	55.18	11.43	48.96	9.61
		Dualprompt [6]	*ECCV 2022*	50.17	10.94	61.92	13.45	55.11	10.18	55.73	11.52
		SOYO [41]	*CVPR 2025*	37.14	**0.64**	67.58	**0.10**	58.02	**1.02**	54.25	**0.59**
		MOS [42]	*AAAI 2025*	27.59	9.62	39.47	4.15	42.37	1.46	36.48	5.08
		KA-Prompt [15]	*ICML 2025*	57.77	4.04	81.79	4.79	60.02	6.08	66.53	4.97
	N-DIL	*KA-Prompt*									
		+ ELR [13]	*NeurIPS 2020*	61.16	4.23	75.48	10.80	50.04	23.05	62.23	12.69
		+ DivideMix [10]	*ICLR 2020*	65.15	2.26	84.51	4.25	54.65	18.51	68.10	8.34
		+ AGCE [43]	*TPAMI 2023*	55.81	0.78	83.84	0.60	60.08	6.34	66.58	2.57
		+ SOP [44]	*ICML 2022*	57.40	5.45	80.30	7.52	60.00	6.43	65.90	6.47
		+ JoAPR [45]	*CVPR 2024*	56.44	4.32	75.38	12.28	61.82	5.14	64.55	7.25
		+ NLPrompt [46]	*CVPR 2025*	61.48	3.19	83.31	6.51	**63.53**	4.53	69.44	4.74
		ST-Prompt(Ours)	*This paper*	**70.57**	2.32	**89.01**	1.07	62.68	5.28	**74.09**	2.89
50%	DIL	CODA-Prompt [7]	*CVPR 2023*	29.09	12.45	45.75	4.09	53.33	8.71	42.72	8.42
		Dualprompt [6]	*ECCV 2022*	34.86	12.10	58.64	7.61	53.41	7.33	48.97	9.01
		SOYO [41]	*CVPR 2025*	23.25	**0.05**	50.41	**0.13**	50.19	**0.39**	41.28	**0.19**
		MOS [42]	*AAAI 2025*	26.35	8.00	38.49	4.15	41.12	1.43	35.32	4.53
		KA-Prompt [15]	*ICML 2025*	40.53	5.14	71.74	4.47	57.18	5.59	56.48	5.07
	N-DIL	*KA-Prompt*									
		+ ELR [13]	*NeurIPS 2020*	44.20	5.41	64.13	9.33	47.24	21.24	51.86	11.99
		+ DivideMix [10]	*ICLR 2020*	52.78	3.00	75.44	6.08	53.69	16.62	60.64	8.57
		+ AGCE [43]	*TPAMI 2023*	46.09	0.42	75.74	0.31	56.92	6.24	59.58	2.32
		+ SOP [44]	*ICML 2022*	39.55	5.64	71.95	6.54	56.88	6.29	56.13	6.16
		+ JoAPR [45]	*CVPR 2024*	48.56	3.22	76.32	5.03	59.29	5.85	61.39	4.70
		+ NLPrompt [46]	*CVPR 2025*	47.49	4.63	80.66	2.20	60.42	5.17	62.86	4.00
		ST-Prompt(Ours)	*This paper*	**65.91**	2.78	**84.86**	1.50	**61.94**	4.42	**70.90**	2.90
80%	DIL	CODA-Prompt [7]	*CVPR 2023*	14.71	5.79	32.30	2.51	49.25	5.76	32.09	4.69
		Dualprompt [6]	*ECCV 2022*	16.26	5.93	39.89	2.45	47.86	5.24	34.67	4.54
		SOYO [41]	*CVPR 2025*	6.90	0.46	14.35	0.14	32.52	**0.27**	17.92	**0.29**
		MOS [42]	*AAAI 2025*	25.44	7.56	36.23	5.04	41.25	1.72	34.31	4.77
		KA-Prompt [15]	*ICML 2025*	17.39	3.96	37.67	4.71	47.09	5.30	34.05	4.66
	N-DIL	*KA-Prompt*									
		+ ELR [13]	*NeurIPS 2020*	19.29	4.46	25.07	6.48	33.60	19.18	25.99	10.04
		+ DivideMix [10]	*ICLR 2020*	17.54	4.10	52.90	2.65	47.84	13.96	39.43	6.90
		+ AGCE [43]	*TPAMI 2023*	24.80	**0.37**	59.91	**0.08**	46.85	5.95	43.85	2.13
		+ SOP [44]	*ICML 2022*	17.42	4.12	48.08	2.96	46.92	5.85	37.47	4.31
		+ JoAPR [45]	*CVPR 2024*	22.09	5.73	60.72	3.72	6.90	15.23	29.90	8.23
		+ NLPrompt [46]	*CVPR 2025*	18.61	6.50	61.51	5.75	54.75	4.71	44.96	5.65
		ST-Prompt(Ours)	*This paper*	**39.60**	0.62	**68.32**	0.98	**56.60**	3.40	**54.84**	1.67

Label Noise Generation: To comprehensively simulate real-world annotation unreliability in DIL, we evaluate our models under two standard noise generation protocols. **(1) Symmetric Noise:[33]** uniformly flips a certain percentage (20%, 50%, 80%) of training labels to other random classes. This setting effectively models general, unconstrained crowd-sourcing errors. Given its broad applicability and prevalence in diverse data collection scenarios, symmetric noise serves as our primary benchmark in the main text. The selected ratios span from mild (20%) to extreme (80%) corruption, allowing us to rigorously stress-test the robustness of the models. **(2) Asymmetric Noise:[31]** models structured, class-dependent human mistakes, such as confusing visually similar categories. Following established practices, we approximate this via a cyclic flip strategy, where a percentage of labels are deterministically flipped to the next adjacent class. Since asymmetric noise targets a more specific, fine-grained corruption scenario, its comprehensive evaluation results are deferred to the Appendix.

Evaluation Metrics: Following standard practices in Domain Incremental Learning (DIL) under noisy environments, we employ two primary metrics to evaluate the performance of our proposed

method and the baselines: **Average Accuracy** (***Avg-ACC***, denoted as $\mathcal{A}_T$) and **Average Forgetting** (***Avg-F***, denoted as $\mathcal{F}_T$). Detailed information on the evaluation metrics can be found in the appendix.

Implementation Details: We employ KA-Prompt as our baseline by default and follow its prompt configurations. The Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) is adopted to train the model. The batch size is fixed at 128 across all benchmarks. The learning rates are empirically set to 0.007 for ImageNet-R, 0.0006 for DomainNet and 0.005 for ImageNet-C. The default training epochs of ImageNet-C, ImageNet-R are set to 5, and DomainNet is 10. The training images are resized to 224×224 by default, whereas a resolution of 230×230 is applied for ImageNet-C. For the essential hyperparameters of our method, q for the Adaptive Generalized Cross Entropy loss and λ_{pkci} for the contrastive loss are set to 0.55 and 0.1, respectively. The temperature parameter for the contrastive isolator is set to 0.07, while the temperature for prompt weight reallocation is fixed at 0.01. All experiments are conducted on a single NVIDIA GeForce RTX 4090 GPU.

Compared Methods: To comprehensively evaluate the proposed approach, we benchmark it against recent incremental learning methods, specifically CODA-Prompt [7], DualPrompt [6], SOYO [41], MOS [42], and the state-of-the-art KA-Prompt [15]. Here, to mitigate potential data privacy risks associated with storing sensitive historical data, we exclusively select state-of-the-art methods that do not rely on data rehearsal as our baselines for fair comparison. Given the varying symmetric noise ratios inherent in our setting, we construct robust baselines by integrating KA-Prompt with established Label Noise Learning techniques, including ELR [13], DivideMix [10], AGCE [43] , SOP [44], JoAPR [45], and NLPrompt [46]. All evaluated methods are implemented using their official codebases and utilize a ViT-B/16 backbone pre-trained on ImageNet-21k.

4.2 Comparison Results with State-of-the-Art

The comparison results across three Noisy DIL datasets under varying symmetric noise ratios, along with the average performance, are presented in Table 1.

Compared with existing DIL methods: Our ST-Prompt framework significantly outperforms the state-of-the-art DIL method, KA-Prompt. Specifically, at a 20% noise ratio, ST-Prompt yields absolute improvements of **12.80%, 7.22%, and 2.66%** on the ImageNet-R, ImageNet-C and DomainNet datasets, respectively. When the noise ratio escalates to 50%, our method exceeds the baseline by margins of **25.38%, 13.12%, and 4.76%** across the three datasets. Under severe 80% noise, we deliver substantial gains of **22.21% , 30.65%, 9.51%** on ImageNet-R, ImageNet-C,and DomainNet. Averaged across all three datasets, ST-Prompt achieves a comprehensive performance ($\mathcal{A}_T$) gain of **7.56%, 14.42%, 20.79%** on 20%,50%,80% label noise, thoroughly surpassing the existing SOTA KA-Prompt. Notably, the performance margin between ST-Prompt and existing baselines widens significantly as the noise ratio escalates, demonstrating that the superiority of our method becomes even more pronounced under extreme label noise conditions.

Compared with existing DIL methods with LNL techniques: Furthermore, our approach exhibits clear superiority when compared to KA-Prompt integrated with the state-of-the-art LNL technique NLPrompt. Under the 20% noise setting, ST-Prompt secures notable gains of **9.09%** and **5.70%** on ImageNet-R,ImageNet-C ,while maintaining competitive performance on DomainNet. As the label corruption increases to 50%, our method registers consistent absolute improvements of **18.42%, 4.20%, and 1.52%** respectively. In the extreme 80% noise scenario, the performance advantage becomes even more pronounced, with ST-Prompt outperforming the baseline by **20.99%**, **6.81%**, and **1.85%** across the corresponding datasets. Overall, our method consistently elevates the average performance ($\mathcal{A}_T$) by **4.65%, 8.04%, and 9.88%** under the 20%, 50%, and 80% noise settings respectively, surpassing the advanced KA-Prompt+NLPrompt paradigm.

Overall, ST-Prompt consistently surpasses the strongest baselines across all datasets, maintaining high accuracy while keeping the average forgetting rate at a low level. These results underscore the robustness of ST-Prompt to severe label noise and domain shifts, which can be attributed to strict knowledge separation and positive transmission of reliable historical knowledge during training.

While the regularization-based SOYO achieves the lowest Average Forgetting ($\mathcal{F}_T$), its rigid constraints severely restrict plasticity, yielding exceptionally poor accuracy ($\mathcal{A}_T$). Evaluated strictly within the highly plastic prompt-based DIL paradigm, however, ST-Prompt attains the best anti-

forgetting performance. By reducing forgetting by absolute margins of **2.08%**, **2.17%**, and **2.99%** across 20%, 50%, and 80% noise ratios compared to the SOTA KA-Prompt, we demonstrate the effectiveness of our PKCI and SKT modules in safely isolating and synergistically transferring historical knowledge.

To further analyze the cross-domain learning process under severe label noise, we visualize the Avg-ACC performance of all compared methods on ImageNet-C with 80% symmetric label noise. Figure 4 illustrates the performance trends throughout the domain incremental learning process. The results show that ST-Prompt achieves performance comparable to existing methods in the initial domain, while exhibiting consistent advantages from the second domain onward. These improvements can be attributed to the proposed Cross-Domain Prompt Knowledge Contrastive Isolator, which alleviates inter-domain knowledge conflicts, and the Synergistic Knowledge Transmission scheme, which enhances new-domain knowledge acquisition. Additional visualizations under different noise settings and datasets are provided in Appendix E.

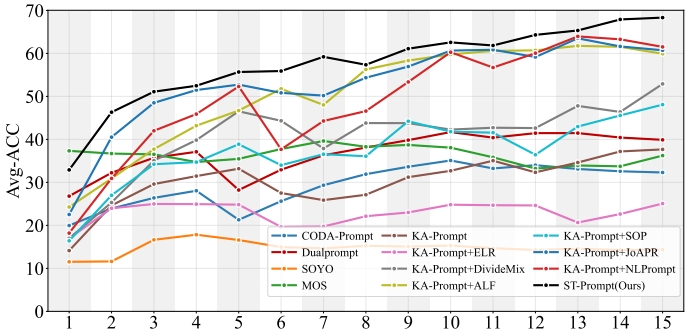


Figure 4: The seen domain performance tendency along the domain incremental learning process

4.3 Ablation Studies

Analysis on the proposed modules: In Figure 5, we conduct ablation studies on the two key components of our framework, namely the Cross-Domain Prompt Knowledge Contrastive Isolator (PKCI) and the Synergistic Knowledge Transmission (SKT) scheme, under the challenging setting of 80% symmetric label noise. Here, AGCE+ denotes the KA-Prompt+AGCE baseline [43], which serves as the reference setting in our component analysis.

The results in Figure 5 show that introducing the PKCI module alone improves the baseline performance from 24.80% to 32.42%. These gains demonstrate that explicit knowledge separation effectively reduces the interference caused by noisy cross-domain knowledge. In contrast, incorporating only the SKT module further boosts the performance to 37.73%, verifying the effectiveness of the proposed synergistic knowledge transmission mechanism in facilitating robust knowledge acquisition.

Furthermore, when PKCI and SKT are jointly employed, the average accuracy is further improved to 39.60%, indicating that the two components are mutually complementary and reinforcing. Overall, these results demonstrate that PKCI primarily enhances historical knowledge isolation, while SKT strengthens synergistic robust knowledge utilization during training. Their integration further exploits the complementary advantages of knowledge separation and positive knowledge transfer, leading to the best overall performance.

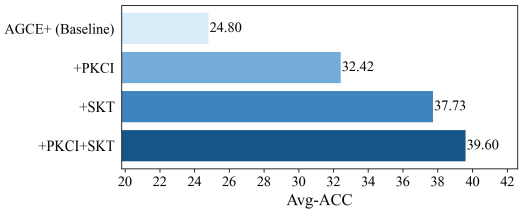


Figure 5: Ablation on the model components

Analysis on the hyper-parameters: In Figure 6, we evaluate the performance of our method under different values of the hyper-parameters λ and τ under 80% severe symmetric label noise. The parameter λ controls the weight of the PKCI loss, balancing cross-

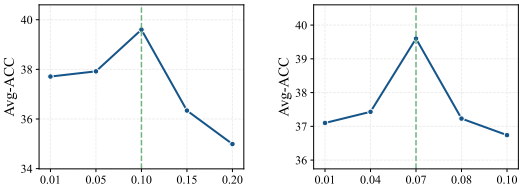


Figure 6: Ablation on λ and τ

domain prompt separation and current-task optimization. The parameter τ is the temperature in the PKCI contrastive loss, which determines the sharpness of similarity distribution between queries and prompt keys. Based on the empirical results, the best performance is achieved at $\lambda = 0.1$ and $\tau = 0.07$, which are adopted as the default configuration in all experiments.

5 Conclusion and Limitation

In this paper, we investigate the practical yet challenging Noisy Domain Incremental Learning (N-DIL) task and propose a novel Cross-Domain Knowledge Separation and Positive Transmission (ST-Prompt) framework. Based on the observation that label noise not only hinders intra-domain knowledge acquisition but also exacerbates inter-domain knowledge conflicts to amplify catastrophic forgetting, we develop a mutually reinforcing paradigm featuring two key components. Specifically, to tackle blurred inter-domain boundaries, the Cross-domain Prompt Knowledge Contrastive Isolator enhances cross-domain knowledge separation, thereby mitigating image-prompt mismatching and knowledge interference during inference. Furthermore, to improve intra-domain knowledge acquisition, the Synergistic Knowledge Transmission scheme extracts reliable knowledge from previous domains to facilitate noisy data learning in the current domain. Extensive experimental results on three N-DIL benchmarks demonstrate that ST-Prompt effectively overcomes these dual challenges, achieving state-of-the-art performance especially under severe noisy conditions.

Limitation: Compared with prompt-based baselines such as KA-Prompt, which directly optimize prompts, our method introduces approximately a 10.5% increase in training time and less than 1% additional memory overhead. Nevertheless, across three noise ratios and three diverse datasets, ST-Prompt achieves relative performance improvements ranging from 11.3% to 29.7% under low-noise settings, and up to 76.4% under extreme noise scenarios. These results demonstrate that our method achieves a favorable trade-off between computational overhead and overall model performance.

Acknowledgement

This work was a research achievement of National Engineering Research Center of New Electronic Publishing Technologies, and was supported by the National Natural Science Foundation of China (62376011) and the National Key R&D Program of China (2024YFA1410000).

References

[1] K. Jeeveswaran, E. Arani, and B. Zonooz, "Gradual divergence for seamless adaptation: A novel domain incremental learning method," in *ICML*, 2024.

[2] C. Lamers, R. Vidal, N. Belbachir, N. Van Stein, T. Bäeck, and P. Giampouras, "Clustering-based domain-incremental learning," in *ICCV*, 2023, pp. 3384–3392.

[3] Q. Chen, C. Shui, L. Han, and M. Marchand, "On the stability-plasticity dilemma in continual meta-learning: Theory and algorithm," *NeurIPS*, vol. 36, pp. 27 414–27 468, 2023.

[4] G. Oren and L. Wolf, "In defense of the learning without forgetting for task incremental learning," in *ICCV*, 2021, pp. 2209–2218.

[5] Y. Wang, Z. Huang, and X. Hong, "S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning," *NeurIPS*, vol. 35, pp. 5682–5695, 2022.

[6] Z. Wang, Z. Zhang, S. Ebrahimi, R. Sun, H. Zhang, C.-Y. Lee, X. Ren, G. Su, V. Perot, J. Dy *et al.*, "Dualprompt: Complementary prompting for rehearsal-free continual learning," in *ECCV*. Springer, 2022, pp. 631–648.

[7] J. S. Smith, L. Karlinsky, V. Gutta, P. Cascante-Bonilla, D. Kim, A. Arbelle, R. Panda, R. Feris, and Z. Kira, "Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning," in *CVPR*, 2023, pp. 11 909–11 919.

[8] H. Song, M. Kim, D. Park, Y. Shin, and J.-G. Lee, "Learning from noisy labels with deep neural networks: A survey," *CSVT*, vol. 34, no. 11, pp. 8135–8153, 2022.

[9] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," *NeurIPS*, vol. 31, 2018.

[10] J. Li, R. Socher, and S. C. Hoi, "Dividemix: Learning with noisy labels as semi-supervised learning," in *ICLR*, 2020.

[11] K. Xu, H. Zhang, Y. Li, Y. Peng, and J. Zhou, "Mitigate catastrophic remembering via continual knowledge purification for noisy lifelong person re-identification," in *ACM MM*, 2024, pp. 5790–5799.

[12] N. Karim, U. Khalid, A. Esmaeili, and N. Rahnavard, "Cnll: A semi-supervised approach for continual noisy label learning," in *CVPR*, 2022, pp. 3878–3888.

[13] S. Liu, J. Niles-Weed, N. Razavian, and C. Fernandez-Granda, "Early-learning regularization prevents memorization of noisy labels," *NeurIPS*, vol. 33, pp. 20 331–20 342, 2020.

[14] D.-W. Zhou, Z.-W. Cai, H.-J. Ye, D.-C. Zhan, and Z. Liu, "Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need," *IJCV*, vol. 133, no. 3, pp. 1012–1032, 2025.

[15] K. Xu, X. Zou, G. Hua, and J. Zhou, "Componential prompt-knowledge alignment for domain incremental learning," in *ICML*, 2025.

[16] K. Xu, L. Ma, J. Li, X. Tong, X. Liu, Y. Guo, and J. Zhou, "Score: Subdistribution-aware collaborative knowledge reinforcing for cloth-hybrid lifelong person re-identification," in *ECCV*. Springer, 2026, pp. 93–110.

[17] K. Xu, H. Cheng, J. Li, X. Zou, and J. Zhou, "Vision-language attribute disentanglement and reinforcement for lifelong person re-identification," in *CVPR*, 2026, pp. 40 397–40 406.

[18] H. Kang and D.-W. Choi, "Recall-oriented continual learning with generative adversarial meta-model," in *AAAI*, vol. 38, no. 12, 2024, pp. 13 040–13 048.

[19] Q. Li, K. Xu, Y. Peng, and J. Zhou, "Exemplar-free lifelong person re-identification via prompt-guided adaptive knowledge consolidation," *IJCV*, vol. 132, no. 11, pp. 4850–4865, 2024.

[20] K. Xu, Y. Zhang, W. Qiang, and J. Zhou, "Hyper-llava: Hyperbolic uncertainty-aware modality-balanced routing for multimodal continual instruction tuning," in *ICML*, 2026.

[21] H. Shi and H. Wang, "A unified approach to domain incremental learning with memory: Theory and algorithm," *NeurIPS*, vol. 36, pp. 15 027–15 059, 2023.

[22] J. Xie, S. Yan, and X. He, "General incremental learning with domain-aware categorical representations," in *CVPR*, 2022, pp. 14 351–14 360.

[23] K. Xu, X. Zou, and J. Zhou, "Lstkc: Long short-term knowledge consolidation for lifelong person re-identification," in *AAAI*, vol. 38, no. 14, 2024, pp. 16 202–16 210.

[24] N. Asadi, M. Davari, S. Mudur, R. Aljundi, and E. Belilovsky, "Prototype-sample relation distillation: towards replay-free continual learning," in *ICML*. PMLR, 2023, pp. 1093–1106.

[25] J. Bonato, F. Pelosin, L. Sabetta, and A. Nicolosi, "Mind: Multi-task incremental network distillation," in *AAAI*, vol. 38, no. 10, 2024, pp. 11 105–11 113.

[26] Y. Li, W. Xu, H. Wang, Y. Qi, J. Guo, and R. Li, "Personalized federated domain-incremental learning based on adaptive knowledge matching," in *ECCV*. Springer, 2024, pp. 127–144.

[27] Y. Wang, Z. Ma, Z. Huang, Y. Wang, Z. Su, and X. Hong, "Isolation and impartial aggregation: A paradigm of incremental learning without interference," in *AAAI*, vol. 37, no. 8, 2023, pp. 10 209–10 217.

[28] K. Xu, X. Zou, Y. Peng, and J. Zhou, "Distribution-aware knowledge prototyping for non-exemplar lifelong person re-identification," in *CVPR*, 2024, pp. 16 604–16 613.

[29] Z. Wang, Z. Zhang, C.-Y. Lee, H. Zhang, R. Sun, X. Ren, G. Su, V. Perot, J. Dy, and T. Pfister, "Learning to prompt for continual learning," in *CVPR*, 2022, pp. 139–149.

[30] C. Zhang, K. Xu, Z. Liu, Y. Peng, and J. Zhou, "Scap: Transductive test-time adaptation via supportive clique-based attribute prompting," in *CVPR*, 2025, pp. 30 032–30 041.

[31] G. Patrini, A. Rozza, A. Krishna Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: A loss correction approach," in *CVPR*, 2017, pp. 1944–1952.

[32] K. Xu, J. Li, Y. Guo, Y. Peng, and J. Zhou, "Mitigate catastrophic remembering via continual self-paced dual-knowledge purification for noisy lifelong person re-identification," *IEEE TPAMI*, 2026.

[33] A. Ghosh, H. Kumar, and P. S. Sastry, "Robust loss functions under label noise for deep neural networks," in *AAAI*, vol. 31, no. 1, 2017.

[34] X. Zhou, X. Liu, J. Jiang, X. Gao, and X. Ji, "Asymmetric loss functions for learning with noisy labels," in *ICML*. PMLR, 2021, pp. 12 846–12 856.

[35] X. Yu, B. Han, J. Yao, G. Niu, I. Tsang, and M. Sugiyama, "How does disagreement help generalization against label corruption?" in *ICML*. PMLR, 2019, pp. 7164–7173.

[36] W. Toner and A. Storkey, "Label noise: Correcting the forward-correction," *arXiv preprint arXiv:2307.13100*, 2023.

[37] Y. Xue, K. Whitecross, and B. Mirzasoleiman, "Investigating why contrastive learning benefits robustness against label noise," in *ICML*. PMLR, 2022, pp. 24 851–24 871.

[38] S. Li, X. Xia, S. Ge, and T. Liu, "Selective-supervised contrastive learning with noisy labels," in *CVPR*, 2022, pp. 316–325.

[39] Y. Lu, Y. Zhang, B. Han, Y.-m. Cheung, and H. Wang, "Label-noise learning with intrinsically long-tailed data," in *ICCV*, 2023, pp. 1369–1378.

[40] H. Li, J. Gu, J. Song, A. Zhang, and L. Gao, "Debiased orthogonal boundary-driven efficient noise mitigation," in *ACL*, 2026, pp. 3643–3662.

[41] Q. Wang, X. Song, Y. He, J. Han, C. Ding, X. Gao, and Y. Gong, "Boosting domain incremental learning: Selecting the optimal parameters is all you need," in *CVPR*, 2025, pp. 4839–4849.

[42] H.-L. Sun, D.-W. Zhou, H. Zhao, L. Gan, D.-C. Zhan, and H.-J. Ye, "Mos: Model surgery for pre-trained model-based class-incremental learning," in *AAAI*, vol. 39, no. 19, 2025, pp. 20 699–20 707.

[43] X. Zhou, X. Liu, D. Zhai, J. Jiang, and X. Ji, "Asymmetric loss functions for noise-tolerant learning: Theory and applications," *IEEE TPAMI*, vol. 45, no. 7, pp. 8094–8109, 2023.

[44] S. Liu, Z. Zhu, Q. Qu, and C. You, "Robust training under label noise by over-parameterization," in *ICML*. PMLR, 2022, pp. 14 153–14 172.

[45] Y. Guo and X. Gu, "Joapr: Cleaning the lens of prompt learning for vision-language models," in *CVPR*, 2024, pp. 28 695–28 705.

[46] B. Pan, Q. Li, X. Tang, W. Huang, Z. Fang, F. Liu, J. Wang, J. Yu, and Y. Shi, "Nlprompt: Noise-label prompt learning for vision-language models," in *CVPR*, 2025, pp. 19 963–19 973.

[47] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *ICCV*, 2019, pp. 1406–1415.

[48] D. Hendrycks, S. Basart, N. Mu, S. Kadavath, F. Wang, E. Dorundo, R. Desai, T. Zhu, S. Parajuli, M. Guo *et al.*, "The many faces of robustness: A critical analysis of out-of-distribution generalization," in *ICCV*, 2021, pp. 8340–8349.

[49] D. Hendrycks and T. G. Dietterich, "Benchmarking neural network robustness to common corruptions and surface variations," in *ICLR*, 2019.

A Theoretical Justification

In this section, we theoretically demonstrate how our Cross-Domain Prompt Knowledge Contrastive Isolator (PKCI) bounds the probability of catastrophic cross-domain prompt mismatch from a statistical perspective, thereby insulating historical knowledge from current domain noise.

Assumption 1 (Statistical Distribution of Queries and Keys). Let $\text{sim}(q, k)$ denote the cosine similarity between a query q and a key k. For a new domain sample $q \in \mathcal{D}_t$, its similarity to the optimal task-specific new key k_{new} is assumed to follow a Gaussian distribution $\text{sim}(q, k_{new}) \sim \mathcal{N}(\mu_{new}, \sigma^2)$. Similarly, its similarity to an arbitrary historical key $k_{old} \in \mathcal{K}_{old}$ follows $\text{sim}(q, k_{old}) \sim \mathcal{N}(\mu_{old}, \sigma^2)$.

A catastrophic cross-domain mismatch occurs when a new domain query incorrectly retrieves a historical key instead of a new key, i.e., $\text{sim}(q, k_{old}) > \text{sim}(q, k_{new})$. The probability of this event is defined as the mismatch probability $\mathbb{P}(\text{Mismatch})$.

Lemma 1 (Mismatch Probability without PKCI). Without the explicit constraint of PKCI, the learning of new keys is unguided relative to historical keys. Consequently, the distribution means μ_{new} and μ_{old} may severely overlap near the domain boundaries ($\mu_{new} \approx \mu_{old}$).

Let $\Delta = \text{sim}(q, k_{old}) - \text{sim}(q, k_{new})$. Since both similarities are independent Gaussians, their difference follows $\Delta \sim \mathcal{N}(\mu_{old} - \mu_{new}, 2\sigma^2)$. The baseline mismatch probability is evaluated by integrating the probability density function of Δ over the region $\Delta > 0$:

$$\mathbb{P}_{base}(\text{Mismatch}) = \mathbb{P}(\Delta > 0) = \int_0^\infty \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left(-\frac{\big(x - (\mu_{old} - \mu_{new})\big)^2}{4\sigma^2}\right) dx. \tag{12}$$

By expressing this integral in terms of the standard Gauss error function $\text{erf}(\cdot)$, we obtain:

$$\mathbb{P}_{base}(\text{Mismatch}) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\mu_{new} - \mu_{old}}{2\sigma}\right)\right]. \tag{13}$$

When domains overlap ($\mu_{new} \approx \mu_{old}$), $\text{erf}(0) = 0$, yielding $\mathbb{P}_{base}(\text{Mismatch}) \to 50\%$. This indicates a massive probability of catastrophic cross-domain interference, exposing historical prompts to severe label noise.

Theorem 1 (Mismatch Probability Reduction with PKCI). By explicitly treating k_{old} as negative anchors in the InfoNCE contrastive objective, PKCI forcibly pushes μ_{old} away from q, while the prompt weight loss $\mathcal{L}_{pw}$ pulls μ_{new} closer. This joint optimization effectively enforces a strict geometric margin $M > 0$, guaranteeing that the distance between the distribution means satisfies $\mu'_{new} - \mu'_{old} \geq M \gg 0$. The mismatch probability under our architecture thus becomes:

$$\mathbb{P}_{PKCI}(\text{Mismatch}) \leq \frac{1}{2}\left[1 - \text{erf}\left(\frac{M}{2\sigma}\right)\right]. \tag{14}$$

Since $\text{erf}(x)$ monotonically approaches 1 as x increases, enforcing a large margin M via PKCI yields:

$$\mathbb{P}_{PKCI}(\text{Mismatch}) \ll \mathbb{P}_{base}(\text{Mismatch}). \tag{15}$$

Interpretation. Theorem 1 provides a straightforward statistical proof of our framework's robustness. Without PKCI, unguided keys overlap across domain boundaries, resulting in a high probability of new noisy samples corrupting historical prompts. By employing PKCI, we enforce a strict distributional margin M. This exponentially reduces the mismatch probability, physically locking the historical knowledge in a safe subspace insulated from current domain noise.

B Algorithms

The overall processes of our key phases PKCI and SKT are shown in Alg. 1 and Alg. 2, respectively.

Algorithm 1 Cross-Domain Prompt Knowledge Contrastive Isolator (PKCI)

Input: Current domain data $D_t = \{(x_i, y_i)\}_{i=1}^{N_t}$, Prompt pool $\mathcal{P} = \mathcal{P}_{old} \cup \mathcal{P}_{new}$, Pre-trained ViT θ.
Output: Contrastive isolation loss $\mathcal{L}_{PKCI}$

for each $(x_i, y_i) \in D_t$ **do**

 # Query Extraction and Prompt Matching
Extract query feature $q(x_i) = \text{ViT}_\theta(x_i)$;
Compute cosine similarity matrix $C_{i,j} = \frac{q(x_i)^\top k_j}{\|q(x_i)\|\|k_j\|}$ for $k_j \in \mathcal{P}$, Eq.2;
Obtain top-K key indices $\mathcal{I}_{\text{top-}K}(x_i)$ by keeping row-wise maximum values of C;

 # Non-overlapping Knowledge Space Construction
Construct positive key set $\mathcal{K}_i^+ = \{k_j \mid j \in \mathcal{I}_{\text{top-}K}(x_i)\}$;
Construct negative key set $\mathcal{K}_i^- = \mathcal{P} \setminus \mathcal{K}_i^+$;
Detach gradients for all historical keys $k_{old} \in \mathcal{P}_{old}$;

 # Contrastive Probability Calculation
Compute temperature-scaled similarity $s(x_i, k) = \exp\left(\frac{q(x_i)^\top k}{\tau \cdot \|q(x_i)\|\|k\|}\right)$, Eq.4;
Calculate matching probability $p(k^+ \mid x_i) = \frac{s(x_i, k^+)}{s(x_i, k^+) + \sum_{k^- \in \mathcal{K}_i^-} s(x_i, k^-)}$, Eq.3;

end for

Compute overall contrastive isolation loss $\mathcal{L}_{PKCI} = -\frac{1}{N_t \cdot K} \sum_{i=1}^{N_t} \sum_{k^+ \in \mathcal{K}_i^+} \log p(k^+ \mid x_i)$, Eq.4;

Return $\mathcal{L}_{PKCI}$

Algorithm 2 Synergistic Knowledge Transmission (SKT)

Input: Current domain noisy data $D_t = \{(x_i, Y_{o,i})\}_{i=1}^{N_t}$, Old model M_{t-1}, Prompt pool $\mathcal{P}$.
Output: Branch-conditional Classification Loss $\mathcal{L}_{cls}$

Initialize $\mathcal{L}_{cls} = 0$;
for each $(x_i, Y_{o,i}) \in D_t$ **do**

 # Synergistic Knowledge Judgment
Obtain pseudo-label from historical model $\widehat{Y}_{t-1,i} = \arg\max M_{t-1}(x_i)$;
Retrieve top-K prompts $\mathcal{P}_{\text{top-}K}$ and similarity scores $sim_{i,m} = 1 - D_{i,m}$;
Compute distance-weighted prompt distribution $w_j^{x_i} = \text{Softmax}(-D_{i,j}/\tau_{fuse})$, Eq.7;
Obtain fused prompt $\bar{p}^{x_i} = \sum_{j \in \mathcal{I}_{\text{top-}K}(x_i)} w_j^{x_i} \cdot p_j$, Eq.7;
Generate prediction probability $P(y|x_i)$ via ViT conditioned on $\bar{p}^{x_i}$;

if $\widehat{Y}_{t-1,i} == Y_{o,i}$ **then**
 # Clean Branch: Positive Knowledge Transmission (f_T)
Calculate a dynamic inter-domain correlation estimation score $\gamma = \frac{1}{K} \sum_{m=1}^{K} \sigma(S_{i,m})$, Eq.5;
Blend hard labels with uniform distribution $\mathbf{Y}_e = \gamma\, \mathbf{e}_{Y_o} + (1 - \gamma)\, \mathbf{u}$, Eq.6;
Calculate knowledge inheritance loss $\mathcal{L}_{cls}^{(i)} = \mathcal{L}_{E-CE}(x_i, Y_e)$;
else
 # Noisy Branch: Misleading Gradient Truncation (f_M)
Calculate adaptive generalized cross-entropy loss $\mathcal{L}_{cls}^{(i)} = \frac{1 - P(y=Y_{o,i}|x_i)^q}{q}$, Eq.9;
end if
Update $\mathcal{L}_{cls} \leftarrow \mathcal{L}_{cls} + \mathcal{L}_{cls}^{(i)}$;
end for

Compute average classification loss $\mathcal{L}_{cls} = \mathcal{L}_{cls}/N_t$, Eq.10;

Return $\mathcal{L}_{cls}$

C Experimental Result with Asymmetric Label Noise

Table 2: The Avg-ACC and Avg-F performance comparison under 20%,50% asymmetric noise across two Noisy DIL benchmarks.

r	Type	Method	Publication	ImageNet-R		ImageNet-C		Average	
				$\mathcal{A}_T$ (↑)	$\mathcal{F}_T$ (↓)	$\mathcal{A}_T$ (↑)	$\mathcal{F}_T$ (↓)	$\mathcal{A}_T$ (↑)	$\mathcal{F}_T$ (↓)
20%	DIL	MOS	*AAAI 2025*	27.59	9.62	39.49	3.93	33.54	6.78
	DIL	KA-Prompt	*ICML 2025*	57.40	4.94	77.36	3.85	67.38	4.40
	N-DIL	KA-Prompt + NLPrompt	*CVPR 2025*	60.78	3.78	80.05	3.40	70.42	3.59
	N-DIL	**ST-Prompt(Ours)**	*This paper*	**69.13**	**2.39**	**86.68**	**0.93**	**77.91**	**1.66**
50%	DIL	MOS	*AAAI 2025*	26.73	8.32	37.97	3.82	32.35	6.07
	DIL	KA-Prompt	*ICML 2025*	37.32	3.24	42.51	3.24	39.92	3.24
	N-DIL	KA-Prompt + NLPrompt	*CVPR 2025*	39.00	3.06	42.94	2.16	40.97	2.61
	N-DIL	**ST-Prompt(Ours)**	*This paper*	**44.21**	**1.69**	**44.38**	**0.11**	**44.30**	**0.90**

As shown in Tab. 2, we evaluate ST-Prompt under asymmetric label noise, which poses a significantly greater challenge than symmetric noise due to its deterministic class-conditional mislabeling patterns. Here, two widely adopted asymmetric noise ratios, namely 20% and 50%, are considered.

Under the 20% asymmetric noise setting, ST-Prompt consistently outperforms all compared baselines, achieving an average accuracy ($\mathcal{A}_T$) of 77.91% across datasets while reducing the average forgetting ($\mathcal{F}_T$) to only 1.66%. When the noise ratio further increases to 50%, the performance of MOS and KA-Prompt deteriorates substantially due to the severe semantic confusion introduced by asymmetric label flips. In contrast, ST-Prompt still maintains a competitive $\mathcal{A}_T$ of 44.30% together with the lowest $\mathcal{F}_T$ of 0.90%.

These results demonstrate that the proposed PKCI effectively isolates misleading semantic interference caused by asymmetric noise, while SKT facilitates robust knowledge transfer even when a large portion of deterministic class relationships are corrupted.

D Datasets Information

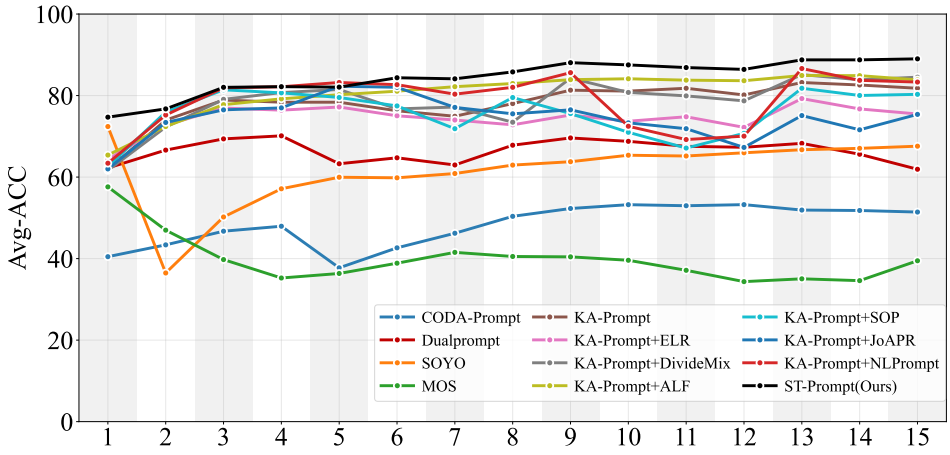
Following previous works [15], we investigate the noisy domain incremental learning problem under three cross-domain benchmarks:

- **DomainNet**[47]: A cross-domain benchmark comprising 586,575 images across 345 classes. It spans six distinct visual domains (e.g., *Real, Quickdraw, Sketch*) for rigorously testing domain generalization.
- **ImageNet-R[48]**: Consists of 30,000 images across 200 categories. It is designed to assess model performance on 15 stylistic renditions (e.g., *art, cartoon, graffiti*).
- **ImageNet-C[49]**: Shares the 200 classes of ImageNet-R but evaluates robustness using 15 types of synthetic corruptions. These corruptions are grouped into *Noise, Weather, Blur,* and *Digital* to simulate real-world distribution shifts.

E Seen-Domain Evaluation Curves

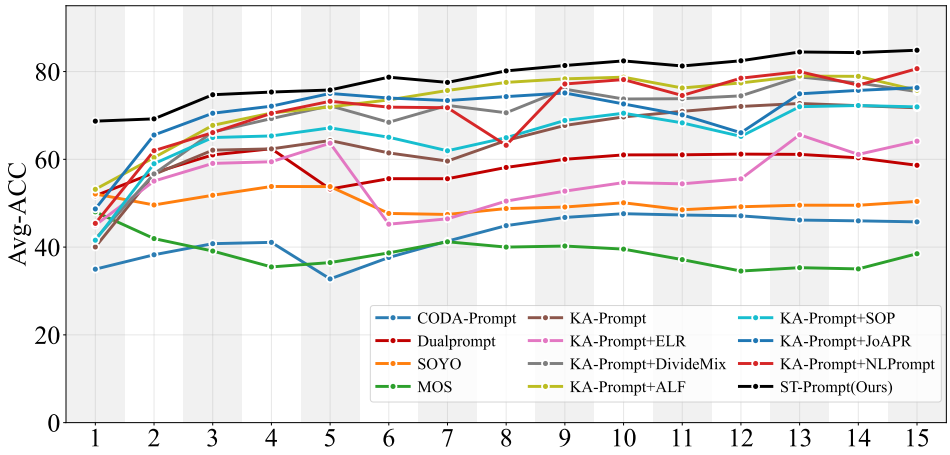
In addition to our main paper, we also provide the seen domain performance tendency on the other benchmarks (ImageNet-C and ImageNet-R) under 20% and 50% symmetric noise ratios, which are illustrated in Fig 7. The full results show that although our method may achieve comparable performance to existing methods in the first few domains, it consistently exhibits superior accuracy as more noisy domains are learned. This advantage becomes more evident under higher noise ratios, demonstrating that our knowledge separation and positive transmission designs effectively preserve

reliable historical knowledge while improving current-domain learning throughout the incremental process.



(a) ImageNet-C (20% Noise)

Analysis: Over the 15 domains of ImageNet-C at 20% noise, ST-Prompt establishes a clear advantage early on. The accuracy curve remains stable and elevated, whereas baselines experience significant fluctuations.



(b) ImageNet-C (50% Noise)

Analysis: For ImageNet-C with 50% noise, ST-Prompt consistently leads across almost all 15 domains. The model avoids catastrophic forgetting and maintains a robust average accuracy.

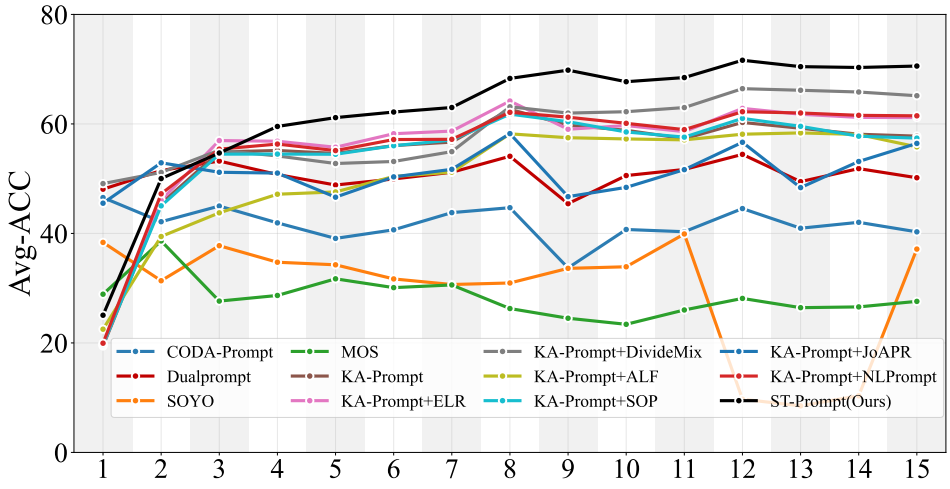
Figure 7: Seen-Domain Evaluation Curves under symmetric label noise on ImageNet-C.

F Details of Evaluation Metrics

F.1 Average Accuracy

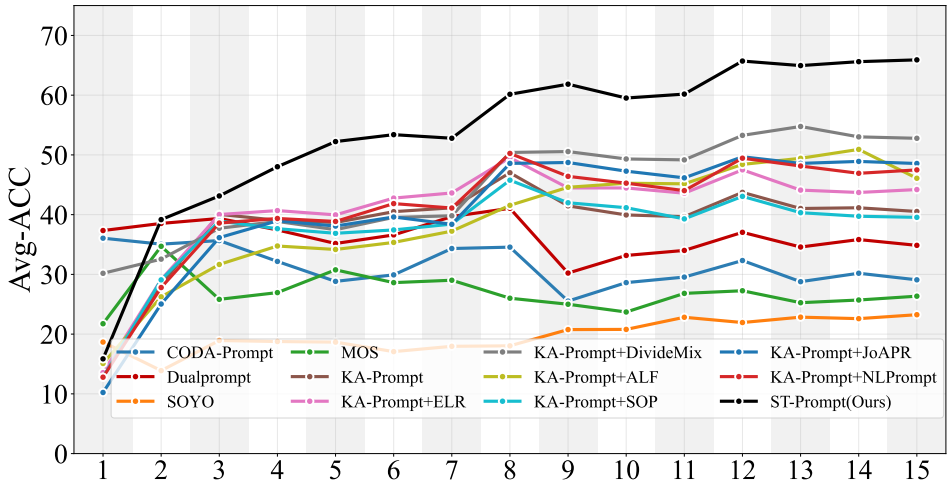
Given a DIL dataset comprising T sequentially learned domains, the Average Accuracy $\mathcal{A}_T$ measures the overall classification performance across all domains after the model has finished learning the final domain T. It is formally defined as:

$$\mathcal{A}_T = \frac{1}{T}\sum_{i=1}^{T} a_{T,i}, \tag{16}$$



(c) ImageNet-R (20% Noise)

Analysis: With 20% noise on ImageNet-R, ST-Prompt shows a consistent upward trend in seen domain accuracy. The accuracy of the incremental learning process expands over the entire 15 steps by an advantage over the performance of the state-of-the-art methods.



(d) ImageNet-R (50% Noise)

Analysis: With 50% noise on ImageNet-R, the inter-domain knowledge conflict of the baseline becomes more pronounced. ST-Prompt resists the accumulation of errors, achieving a significantly higher accuracy compared to the state-of-the-art methods.

Figure 7: Seen-Domain Evaluation Curves under symmetric label noise on ImageNet-R.

where $a_{T,i}$ represents the classification accuracy evaluated on the clean test set of the i-th domain, using the model parameters $\mathcal{M}_T$ obtained after training on the T-th domain. Higher $\mathcal{A}_T$ indicates superior overall knowledge retention and noise robustness.

F.2 Average Forgetting

To explicitly quantify the catastrophic forgetting caused by subsequent noisy domain learning, we compute the Average Forgetting $\mathcal{F}_T$, calculated as the performance degradation of each domain from its peak accuracy:

$$\mathcal{F}_T = \frac{1}{T-1} \sum_{i=1}^{T-1} \max_{l \in \{1,\dots,T-1\}} (a_{l,i} - a_{T,i}). \tag{17}$$

A lower $\mathcal{F}_T$ signifies a stronger capability to preserve previously acquired knowledge without being overwritten by incoming corrupted features.

Furthermore, to comprehensively assess the generalizability and stability of different models under various corruption scenarios, we report the **Average** performance (both Acc and F) across three challenging DIL benchmarks (i.e., ImageNet-R, ImageNet-C, and DomainNet) at distinct symmetric noise ratios (20%, 50%, and 80%).

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have analyzed the limitations of this method over computing and communication overheads.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We have provided a theoretical analysis of our distribution operation with a complete theoretical derivation proof.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper fully discloses all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We have attached our source code and data access links in the supplementary materials, with sufficient instructions to faithfully reproduce the main experimental results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We have specified all the training and test details necessary to understand the results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Error bars are not reported due to computational constraints. However, all experiments are conducted with consistent benchmark configurations and fixed random seeds, ensuring reproducibility. The significant and consistent improvements across multiple benchmarks substantiate the effectiveness of our method.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: For each experiment, the paper provides sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiment.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The Broader impacts has been discussed in the Appendix.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: They are properly credited and respected.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: This core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.