RECOVER AND PROTECT: MANAGING INTERFERENCE IN CONTINUAL TEST-TIME ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Continual Test-Time Adaptation (CTTA) requires learning from changing, unlabeled domains while retaining knowledge for recurring conditions. However, adaptation updates can both transfer useful knowledge and interfere with previous learning. We propose Interference Recovery and Protection (IRP), a geometric framework that proposes a novel decomposition of shared and domain-specific knowledge. These components allow precise identification and mitigation of harmful interference for each domain while preserving plasticity. IRP unifies two complementary mechanisms: *recovery* adaptively corrects accumulated cross-domain interference to rapidly restore a model suited to the returning domain, while *protection* minimally adjusts optimizer updates that oppose the direction previously learned for that domain, limiting self-training error accumulation. Domain shifts and recurring domains are identified online through reconstruction losses and gradient signatures, without domain labels or annotated boundaries. Results on CTTA benchmarks show that IRP reaches state of the art results over multiple adaptation rounds, demonstrating the benefit of targeted interference management for knowledge transfer and retention.

1 INTRODUCTION

Deep neural networks deployed in real-world environments encounter distribution shifts as conditions such as weather, illumination, and sensor quality change, often degrading the performance of models trained on a fixed source distribution (Hendrycks & Dietterich, 2019). Continual Test-Time Adaptation (CTTA) addresses this problem by adapting a pretrained model online to a stream of unlabeled data from changing target domains, without access to the source training set (Wang et al., 2022). Over prolonged deployment, previously encountered conditions may recur, making adaptation to the current domain only part of the objective (Park et al., 2025). Effective CTTA must therefore retain knowledge useful for returning domains while remaining plastic enough to learn from new ones and exploit knowledge shared across them (Liu et al., 2024b).

A fundamental challenge in Continual Test-Time Adaptation (CTTA) is that adapting a model to an incoming domain inevitably perturbs previously learned representations—a phenomenon commonly known as *catastrophic forgetting* (McCloskey & Cohen, 1989). Yet, novel updates can both transfer knowledge across domains and interfere with previously acquired knowledge, effects that standard gradient-based adaptation does not explicitly distinguish. Prior work (Liu et al., 2024b) demonstrates the benefit of disentangling domain-specific and domain-invariant knowledge. For these reasons, we propose a geometric view novel to CTTA, that allows decomposing the adapted model into *transferable* and *domain-specific* components, and, consequently, isolating harmful *interference* in a grounded way.

Building on this geometric view, we propose **Interference Recovery and Protection (IRP)**, a CTTA strategy that explicitly prevents interference on domain-specific components while preserving plasticity. We characterize *interference* relative to a past domain through the projection of a parameter displacement onto that domain's specific component. This decomposition is grounded in a well-established result in continual learning: parameter updates that are orthogonal to a domain's direction in weight space cause less forgetting (Farajtabar et al., 2020; Saha et al., 2021). IRP act against both *cross-domain interference* and *error accumulation interference*. For cross-domain interference, when a domain reappears during adaptation, IRP dynamically corrects the model along

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.

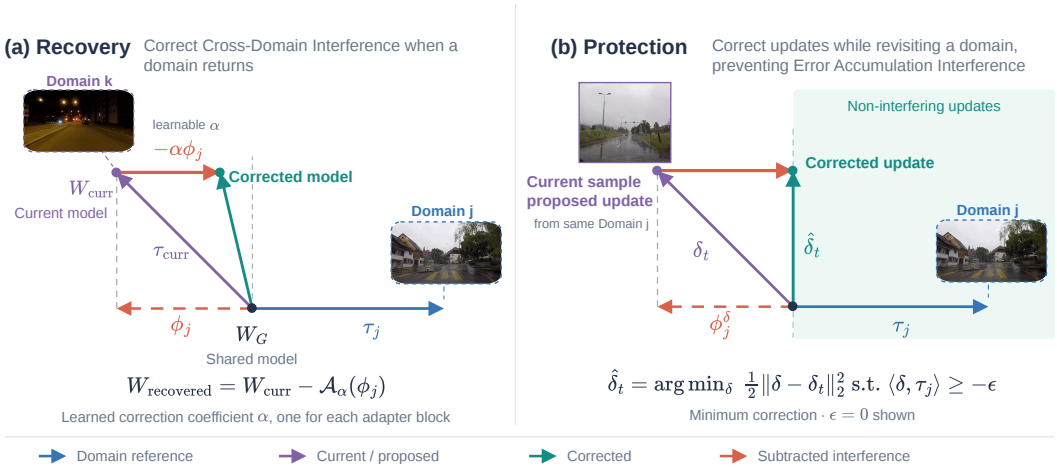


Figure 1: (a) IRP enables recovery by quickly restoring the previous model checkpoint for a reoccurring domain. (b) Protection prevents interference by dampening interfering updates, limiting the accumulation of self-training errors while preserving plasticity.

the direction indicated by interference, restoring the model to a state geometrically consistent with that domain's requirements (see Fig. 1a). We call this operation *recovery*. Recovery speeds up and improves adaptation to the reoccurring domain. For error accumulation interference, when a domain is revisited, IRP *protects* the previous domain-specific direction learned during its previous occurrence from proposed optimizer displacement which could oppose it. We minimally correct only interfering updates, limiting the accumulation of self-training errors while preserving plasticity (see Fig. 1b).

Our contributions are threefold: (1) a geometric decomposition of CTTA model parameters into domain-general and domain-specific components, grounding interference geometrically; (2) IRP, which unifies recovery, correcting accumulated cross-domain interference to rapidly restore a model suited to a returning domain, with protection, minimally adjusting optimizer updates that oppose a domain's previously learned direction; and (3) an online mechanism for detecting domain shifts and identifying recurring domains, without domain labels or boundaries. On CTTA segmentation and classification benchmarks, IRP achieves state-of-the-art results that improve over multiple adaptation rounds.

2 Related Work

Continual test-time adaptation. Test-time adaptation learns from unlabeled test data, commonly through entropy minimization (Wang et al., 2021) or teacher–student consistency (Wang et al., 2022). Continually evolving distributions introduce a stability–plasticity challenge closely related to continual learning, where parameter regularization (Zenke et al., 2017), knowledge distillation (Li & Hoiem, 2017), and episodic memories (Lopez-Paz & Ranzato, 2017) preserve previously acquired capabilities. CTTA must additionally contend with unreliable supervision and unannounced distribution changes. Stabilization strategies therefore include averaged predictions and stochastic restoration (Wang et al., 2022), adaptive selective resets (Lim et al., 2026), and reconstruction-based supervision (Liu et al., 2024a; Park et al., 2025). Beyond stabilizing updates, explicitly organizing domain knowledge can support both transfer and retention. ViDA (Liu et al., 2024b) separates shared and domain-specific representations through adapters of different ranks, while BECoTTA (Lee et al., 2024) combines low-rank experts through domain-adaptive routing. When domains recur, previously acquired knowledge can be reused through an adaptive prompt coreset (Zhang et al., 2025) or a reservoir of domain-specialized models (Vray et al., 2026). Such reuse must also account for compatibility between historical and incoming knowledge, a concern addressed by KFF through prompt fusion and fission (Zhou et al., 2026). IRP approaches this problem through checkpoint-derived domain directions, which guide recovery of the current model while allowing it to build on intervening adaptation.

Geometric interference control. Projection-based continual learning translates knowledge preservation into geometric constraints. GEM (Lopez-Paz & Ranzato, 2017) constrains updates using gradients from episodic memories, whereas GPM (Saha et al., 2021) projects gradients onto the complement of subspaces associated with previous tasks. IRP shares the principle of correcting interfering movement but uses a checkpoint-derived direction for the revisited domain. Its one-sided constraint retains compatible optimizer displacements and minimally corrects conflicting ones without replaying past examples, complementing recovery with selective protection during readaptation.

3 SETTING

CTTA aims to adapt a model pretrained on source data $\mathcal{D}_s = (\mathcal{X}_S, \mathcal{Y}_S)$ to a sequence of unlabeled data distributions at test-time $\{\mathcal{X}_{T_1}, \mathcal{X}_{T_2}, \ldots, \mathcal{X}_{T_n}\}$, where n is the number of unseen domains. Three main challenges constrain CTTA scenarios: (1) source data $\mathcal{D}_s$ is not available; (2) each instance at test time can only be encountered once, prohibiting multi-epoch learning and needing efficient use of that instance; (3) boundaries and identities of domains are not signaled to the model; (4) test time data is naturally unlabeled.

This results in having to adapt to a continuous sequence of data, requiring the model to be robust to unknown domain shifts. Importantly, the same domain *can reappear again in the test sequence*, necessitating the model not to discard old knowledge. This property is crucial for real-world scenarios, where it is unrealistic to assume that past domains cannot appear again in the future.

Motivating the Importance of Transfer in CTTA. While in most continual learning scenarios avoiding forgetting is the main challenge, we argue that in CTTA encouraging knowledge transfer between domains is critical, especially because they share a certain degree of semantic correlation. We show in Table 1 that a simple oracle-driven baseline adapting on each domain separately (*Oracle Separate Domains*) yields inferior results compared to the typical sequential training that characterizes CTTA, evidencing the benefit of parameter sharing that enables positive transfer between domains. However, interference is still present: the sequential baseline without interference management is surpassed by our IRP method, which actively prevents it.

Table 1: Average accuracy across adaptation rounds on Cityscapes-to-ACDC. All Avg is computed over all ten rounds.

Method	R1	R4	R7	R10	All Avg ↑
Source	56.72	56.72	56.72	56.72	56.72
Oracle Separate Domains	58.06	57.98	58.04	58.13	58.10
Sequential Training	60.60	62.90	62.12	61.44	62.20
IRP	**60.63**	**63.26**	**62.75**	**62.35**	**62.70**

4 METHOD

In this Section we present our method IRP that is able to geometrically disentangle domain-specific from domain-general knowledge, while managing interference by means of *recovery* and *protection*.

4.1 LEARNING LOSS

A common learning strategy in CTTA is to employ a weight-averaged *teacher* model to generate pseudo-labels for the unlabeled target data (Wang et al., 2022). Both student f_θ and teacher $f_{\theta'}$ models are initialized with the source model parameter θ_s. At test time, as the model encounters a stream of target inputs x_t, the teacher produces pseudo-labels $\hat{y}'_t = f_{\theta'_t}(x_t)$, which supervise the student via a cross-entropy loss $\mathcal{L}_{\text{CE}}$ between the student's own prediction $\hat{y}_t$ and $\hat{y}'_t$. After each student update, the teacher weights are refined through an exponential moving average (EMA) of the student's parameters: $\theta'_{t+1} = \alpha \cdot \theta'_t + (1-\alpha) \cdot \theta_{t+1}$, where α controls the influence of the student on the teacher update.

Beyond pseudo-label supervision, a growing body of CTTA methods (Liu et al., 2024a; Park et al., 2025) augments this framework with a self-supervised *Masked Image Modeling* (MIM) loss (Xie et al., 2022; He et al., 2022), leveraging the rich visual signal present in unlabeled images. In this paradigm, a portion of the input patches is randomly masked before being fed to the encoder, and an auxiliary reconstruction decoder g_ψ is trained to recover the original intensities of the masked

regions:

$$\mathcal{L}_{\text{rec}} = \frac{1}{|\mathcal{M}|} \left\| \left(x - g_\psi(f_\theta(\tilde{x}))\right) \odot \mathcal{M} \right\|_1 , \tag{1}$$

where $\tilde{x} = x \odot (1 - \mathcal{M}) + e \odot \mathcal{M}$ denotes the masked input, $\mathcal{M}$ is a binary patch mask, e are learnable mask tokens, and $\odot$ is element-wise multiplication. The total adaptation objective at each time step t is then:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_r \cdot \mathcal{L}_{\text{rec}}, \tag{2}$$

where λ_r balances the two terms. We adopt this combined mean-teacher and MIM framework as our base adaptation strategy, and build our IRP strategy on top of it.

4.2 Domain-specific component and Interference

IRP keeps track of transferable and domain-specific weight components and updates them continually during CTTA Given a sequence of checkpoints of our model after adapting on each domain $\theta_1, \ldots, \theta_n$, we define the *common transfer model* θ_G as the average of $\theta_1, \ldots, \theta_n$:

$$\theta_G = \frac{1}{n} \sum_{i=1}^{n} \theta_i \, . \tag{3}$$

This definition is motivated by parameter-space model merging, where averaging checkpoints derived from a common initialization can consolidate knowledge shared across different fine-tuning trajectories into a single model (Wortsman et al., 2022; Ilharco et al., 2023). We consequently interpret θ_G as an estimate of the domain-general model that holds common knoledge accumulated throughout adaptation. Each checkpoint can be decomposed as $\theta_i = \theta_G + \tau_i$ where τ_i represents the *domain-specific* component of domain i, that captures domain-specific knowledge, Displacement τ_i is beneficial for domain i, but may interfere with, and potentially degrade, performance on other domains.

The decomposition above isolates the domain-specific directions along which interference can arise while separating out shared knowledge. Analogously to task vectors (Ilharco et al., 2023), $\tau_j = \theta_j - \theta_G$ points from the common model towards the checkpoint adapted to domain j. We interpret it as an approximate descent direction for the domain loss $\mathcal{L}_j$ (Wei et al., 2025). For a generic parameter displacement Δ, a first-order expansion therefore gives:

$$\mathcal{L}_j(\theta + \Delta) - \mathcal{L}_j(\theta) \approx \langle \nabla \mathcal{L}_j(\theta), \Delta \rangle \ . \tag{4}$$

Consequently, $\langle \tau_j, \Delta \rangle < 0$ predicts an increase in the loss on domain j. We isolate the component responsible for this interaction by projecting Δ onto τ_j:

$$\phi_\Delta^j = \text{proj}_{\tau_j}(\Delta) = \frac{\langle \Delta, \tau_j \rangle}{\|\tau_j\|^2} \tau_j \, . \tag{5}$$

Subtracting this projection removes the conflicting component. Since

$$\left\langle \tau_j, \Delta - \phi_\Delta^j \right\rangle = \langle \tau_j, \Delta \rangle - \frac{\langle \Delta, \tau_j \rangle}{|\tau_j|^2} |\tau_j|^2 = 0, \tag{6}$$

the assumed alignment between τ_j and the negative loss gradient yields

$$\mathcal{L}_j\big(\theta + \Delta - \phi_\Delta^j\big) - \mathcal{L}_j(\theta) \ \approx \left\langle \nabla \mathcal{L}_j(\theta), \Delta - \phi_\Delta^j \right\rangle \approx 0. \tag{7}$$

Thus, the correction counteracts the predicted loss increase while preserving all components orthogonal to τ_j, consistent with gradient-projection methods (Saha et al., 2021) under this local approximation.

4.3 Managing Interference

Building on the decomposition introduced in the previous section, IRP manages interference on additional parameters added to the backbone. Following prior CTTA approaches (Park et al., 2025; Tian et al., 2026), we augment each backbone MLP block with a bottleneck MLP adapter, using AdaptFormer (Chen et al., 2022) architecture. Unlike the rest of the network which is pretrained on the source domain, these adapters are randomly initialized, thus having more capacity for adaptation to incoming domains.

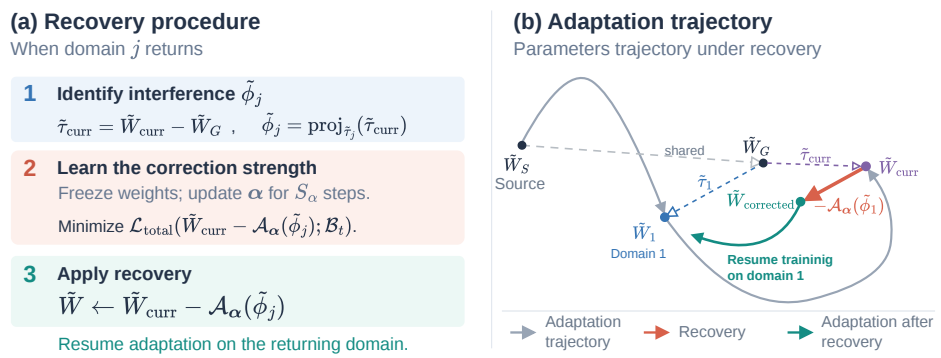


Figure 2: Recovery algorithmical procedure (a) and visualized adaptation trajectory in parameter space for recovery (b).

Recovery of Cross-Domain Interference. We want the MLP adapters, whose parameters are denoted by $\tilde{W} \subset \theta$, to remain highly plastic and specialize to the current domain. We analogously decompose their parameters into a transferable component $\tilde{W}_G$ and domain-specific components $\tilde{\tau}_1, \ldots, \tilde{\tau}_n$. When a previously encountered domain j returns, IRP performs *recovery*: it adaptively corrects accumulated cross-domain interference in the adapters to rapidly restore a model state suited to the incoming domain. This targeted correction preserves transferable knowledge acquired across domains and provides a favorable starting point for subsequent adaptation, accelerating readaptation.

Let $\tilde{W}_{\text{curr}}$ be the adapter parameters at the beginning of the new adaptation period and let $\tilde{\tau}_{\text{curr}} = \tilde{W}_{\text{curr}} - \tilde{W}_G$. We first compute, in the concatenated adapter-parameter space, the component of the current displacement parallel to the stored direction of domain j, i.e. interference towards j:

$$\tilde{\phi}_j = \text{proj}_{\tilde{\tau}_j}(\tilde{\tau}_{\text{curr}}) = \frac{\langle \tilde{\tau}_{\text{curr}}, \tilde{\tau}_j \rangle}{\|\tilde{\tau}_j\|_2^2} \tilde{\tau}_j. \tag{8}$$

IRP learns one adaptive coefficient α_b that regulates interference subtraction for each transformer block, allowing the strength of the interference correction to vary across network depth. Let $\mathcal{B}$ denote the set of transformer blocks containing adapters and let $\tilde{\phi}_j^{(b)}$ denote the portion of $\tilde{\phi}_j$ corresponding to all adapter parameters in block b. The block-wise scaled correction is defined as:

$$\mathcal{A}_{\boldsymbol{\alpha}}(\tilde{\phi}_j) = \text{concat}_{b \in \mathcal{B}} \left(\alpha_b \, \tilde{\phi}_j^{(b)} \right). \tag{9}$$

Thus, each coefficient α_b jointly regulates the interference subtraction applied to all adapter weights and biases within transformer block b.

At each domain revisit, the model parameters are held fixed and the persistent coefficients $\boldsymbol{\alpha}$ are updated for S_α auxiliary steps using the same current-batch adaptation objective:

$$\boldsymbol{\alpha}^* = \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{B}|}} \mathcal{L}_{\text{total}} \left(\tilde{W}_{\text{curr}} - \mathcal{A}_{\boldsymbol{\alpha}}(\tilde{\phi}_j); \mathcal{B}_t \right). \tag{10}$$

The optimization is evaluated through temporary corrected parameters, after which the resulting correction is committed once:

$$\tilde{W}_{\text{corrected}} = \tilde{W}_{\text{curr}} - \mathcal{A}_{\boldsymbol{\alpha}^*}(\tilde{\phi}_j). \tag{11}$$

The coefficients are maintained across the test stream rather than reinitialized at every revisit. Consequently, $\alpha_b = 0$ suppresses the correction for a parameter group, a positive value subtracts the projected component with adaptive intensity, and a negative value reverses its effect when this better minimizes the current adaptation loss. This data-driven correction avoids imposing the same amount of interference subtraction throughout the adapters and provides a parameter configuration from which adaptation to the revisited domain can resume efficiently.

Protection against Error Accumulation Interference. IRP also to prevents the phenomenon known as *error accumulation* (Wang et al., 2022) that characterizes teacher–student consistency

self-training losses. This can be prevented by forcing the updates for the current domain not to interfere with the past state of that domain. Again, we have $\tilde{\tau}_j = \tilde{W}_j - \tilde{W}_G$ as the domain-specific direction for domain j. We want to protect the past $\tilde{\tau}_j$ from new updates when j reappears during adaptation.

Rather than projecting the raw gradient, we predict the parameter displacement δ_t that Adam would apply at step t, accounting for its learning rate, moments, and weight decay. A displacement satisfying $\langle \delta_t, \tilde{\tau}_j \rangle < 0$ opposes the direction previously learned for domain j and therefore constitutes harmful interference. We obtain the closest admissible displacement by solving

$$\widehat{\delta}_t = \arg\min_{\delta} \frac{1}{2} |\delta - \delta_t|_2^2 \quad \text{s.t.} \quad \langle \delta, \tilde{\tau}_j \rangle \geq -\epsilon, \tag{12}$$

where $\epsilon \geq 0$ permits limited movement against $\tilde{\tau}_j$, preserving plasticity as the domain continues to evolve. Since only the direction of the currently revisited domain is protected, the associated dual problem contains a single non-negative multiplier and admits the closed-form solution:

$$\lambda_t^* = \max\left(0, -\frac{\langle \delta_t, \tilde{\tau}_j \rangle + \epsilon}{|\tilde{\tau}_j|_2^2}\right), \qquad \widehat{\delta}_t = \delta_t + \lambda_t^* \tilde{\tau}_j. \tag{13}$$

Thus, non-interfering updates are retained unchanged, whereas violating updates receive the minimum correction required to reach the admissible half-space. In practice, Adam first applies δ_t, after which $\lambda_t^* \tilde{\tau}_j$ is added to the selected backbone parameters. Protecting only the active domain direction has the aim to avoid error accumulation over reiterating errors on the same revisited domain.

The constrained update can be viewed as the Euclidean projection of the proposed displacement δ_t onto the half-space $\delta : \langle \delta, \tilde{\tau}_j \rangle \geq -\epsilon$. If δ_t already lies in this admissible region, it is left unchanged; otherwise, it is shifted only along the protected direction $\tilde{\tau}_j$ by the minimum amount required to reach the boundary $\langle \widehat{\delta}_t, \tilde{\tau}_j \rangle = -\epsilon$. Thus, all components orthogonal to $\tilde{\tau}_j$ are preserved.

4.4 Domain Shift Detection and Domain Assignment

IRP must determine when the input distribution changes and whether the incoming domain has been encountered before, without domain labels or annotated boundaries. We infer both online using reconstruction losses and gradient signatures evaluated on the frozen source model.

Domain Shift Detection by Reconstruction Loss. We monitor the source model's reconstruction loss $\mathcal{L}_{\text{rec}}^s$; using the source model as a fixed reference provides a domain-agnostic change signal. We employ an Exponential Moving Average band detector on the denoised loss: a shift is detected when the loss moving average loss leaves an adaptive band tracking its recent mean and variability. A warmup period after each detection allows these statistics to stabilize before monitoring resumes. More details in Appendix A.1. Empirical analysis in Appendix C.

Domain Assignment by Gradient Signatures. After a detected shift, we average adapter gradients over a short observation window to obtain a *gradient signature* $\bar{g}_*^{\tilde{W}} \in \mathbb{R}^{|\tilde{W}|}$, defined as the average gradient on adapters $\tilde{W}$ with respect to the full adaptation loss $\mathcal{L}_{\text{total}}^s$ on the source model. Evaluating signatures at fixed source parameters makes them comparable across encounters without accumulated parameter drift. We compare $\bar{g}_*^{\tilde{W}}$ with gradient signatures $\bar{g}_k^{\tilde{W}}$ of previously observed domains $k \in \{1, \ldots, K\}$ using cosine similarity: a sufficiently strong match identifies a returning domain; otherwise, a new domain is registered. For returning domains, the assignment activates recovery and protection using the corresponding stored domain-specific component. More technical details are provided in Appendix A.2. Empirical analysis in Appendix C.

5 Experiments

We evaluate IRP on semantic segmentation and image classification test time benchmarks under recurring distribution shifts. Our experiments examine whether interference management improves adaptation over repeated domain encounters and how recovery and protection interact.

Table 2: CTTA results for Cityscapes-to-ACDC. We report mIoU (↑) for each weather condition over multiple rounds.

	rounds →																
	1				4				7				10				**All↑**
Method	Fog	Night	Rain	Snow	Fog	Night	Rain	Snow	Fog	Night	Rain	Snow	Fog	Night	Rain	Snow	**Mean**
Source	69.1	40.3	59.7	57.8	69.1	40.3	59.7	57.8	69.1	40.3	59.7	57.8	69.1	40.3	59.7	57.8	56.7
BN Stats Adapt (Schneider et al., 2020)	62.3	38.0	54.6	53.0	62.3	38.0	54.6	53.0	62.3	38.0	54.6	53.0	62.3	38.0	54.6	53.0	52.0
TENT (Wang et al., 2021)	69.0	40.2	60.1	57.3	66.5	36.3	58.7	55.0	64.2	32.8	55.3	50.9	61.8	29.8	51.9	47.8	52.3
CoTTA (Wang et al., 2022)	70.9	41.2	62.4	59.7	70.9	41.0	62.7	59.7	70.9	41.0	62.8	59.7	70.9	41.0	62.8	59.7	58.6
EcoTTA (Song et al., 2023)	68.5	35.8	62.1	57.4	68.1	35.3	62.3	57.3	67.2	34.2	62.0	56.9	66.4	33.2	61.3	56.3	55.2
SVDP (Yang et al., 2024)	71.0	43.2	64.8	62.0	**71.8**	43.9	66.6	62.3	71.9	43.6	66.5	62.3	71.6	43.5	66.6	62.0	61.0
ViDA (Liu et al., 2024b)	**71.6**	43.2	**66.0**	63.4	70.9	44.0	66.0	**63.2**	**72.3**	44.8	66.5	62.9	**72.2**	45.2	66.5	**62.9**	61.6
BECoTTA (Lee et al., 2024)	72.0	41.2	63.4	60.3	72.3	41.4	64.0	61.3	72.7	40.7	64.6	61.0	72.5	40.8	64.8	60.3	59.6
+ReservoirTTA (Vray et al., 2026)	72.7	41.5	63.3	61.0	72.7	41.4	64.6	61.2	72.4	41.5	64.7	61.1	72.4	41.4	64.9	61.2	59.9
Hybrid-TTA (Park et al., 2025)	70.3	44.2	64.3	63.1	71.8	48.7	66.4	63.7	70.3	48.4	65.6	62.8	68.7	48.1	65.2	62.2	61.9
IRP (*ours*)	70.2	**44.6**	64.3	**63.5**	71.7	**49.2**	**67.8**	**64.4**	71.4	**49.0**	**67.2**	**63.4**	70.8	**49.1**	**66.7**	**62.9**	**62.7**

5.1 Experimental Setup

Benchmarks and evaluation. For semantic segmentation, we evaluate Cityscapes-to-ACDC (Cordts et al., 2016; Sakaridis et al., 2021) using ten rounds of the sequence Fog→Night→Rain→Snow, following the recurring-domain protocol of prior CTTA work (Wang et al., 2022; Park et al., 2025). We report mean intersection-over-union (mIoU), with per-domain results at rounds 1, 4, 7, and 10 and an overall mean across all ten rounds. For classification, we evaluate CIFAR10-C, CIFAR100-C and ImageNet-C (Krizhevsky et al., 2009; Deng et al., 2009; Hendrycks & Dietterich, 2019) over five adaptation rounds, and report the mean classification error within each round. Lower error and higher mIoU indicate better performance. Adaptation uses unlabeled target inputs without source data, and IRP infers domain boundaries and identities online.

Comparisons and implementation. We compare IRP against the methods reported in Tables 2 and 3, including entropy minimization, teacher–student adaptation, and approaches that organize domain knowledge through adapters or prompts. We adopt ViT-base (Dosovitskiy et al., 2021) as the backbone in the classification experiments, pretrained on the respective non-corrupted dataset. Input images are resized to 224x224. For segmentation CTTA, we adopt the pre-trained Segformer-B5 model (Xie et al., 2021) on Cityscapes Cordts et al. (2016) as the source model. We down-sample the input size from 1920x1080 to 960x540 for target domain data, following other works (Liu et al., 2024b; Park et al., 2025). We employ Adam optimizer (Kingma & Ba, 2015) and set specific learning rate values for each benchmark: 6e-05 for CIFAR10-C and ImageNet-C, 3e-05 for CIFAR100-C, and 7.5e-5 for ACDC. See Appendix D for more details on hyperparameters and implementation.

5.2 Main Results

CTTA Semantic Segmentation. Table 2 reports mIoU for ACDC adaptation. IRP achieves the highest overall mIoU, 62.7, improving over *source* by 6.0 points and over the listed Hybrid-TTA and ViDA results by 0.8 and 1.1 points, respectively. Coherent with its *recovery* and *protection* components, IRP advantage is most apparent after repeated adaptation: at round 10, IRP exceeds the second-best method Hybrid-TTA results in all four weather conditions, by 2.1 points on Fog, 1.0 on Night, 1.5 on Rain, and 0.7 on Snow. Night improves from 44.6 in the first round to 49.1 in the tenth, while Rain improves from 64.3 to 66.7. Nevertheless, IRP does not dominate every condition, as competing methods retain higher scores on Fog. We attribute this to Fog being the easiest domain, as reflected by its consistently higher performance across methods. Without explicit interference management, self-training may preferentially specialize the model toward this already well-learned condition, potentially at the expense of other domains. IRP instead corrects parameter displacement according to cross-domain interference, which may slightly reduce specialization on Fog while yielding a better trade-off across recurring conditions. This interpretation is consistent with the stronger aggregate performance of IRP and its gains on the more challenging Night, Rain, and Snow domains. Overall, the aggregate and later-round results support IRP's effectiveness for adaptation to recurring conditions.

CTTA Image classification. Similarly to the previous benchmark, IRP shows a consistent benefit from repeated domain exposure for classification tasks in Table 3. From the first to the fifth round, IRP reduces error from 13.5 to 9.5 on CIFAR10-C, from 25.4 to 16.2 on CIFAR100-C, and from

Table 3: CTTA results for CIFAR10-C, CIFAR100-C and ImageNet-C classification benchamark. We report average error (↓) across five adaptation rounds.

Method	CIFAR10-C					CIFAR100-C					ImageNet-C				
	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5
Source	28.2	28.2	28.2	28.2	28.2	35.4	35.4	35.4	35.4	35.4	55.8	55.8	55.8	55.8	55.8
TENT-continual (Wang et al., 2021)	24.1	21.2	20.4	20.1	19.8	32.1	33.2	52.2	86.7	98.9	51.0	47.0	45.9	45.2	45.0
CoTTA (Wang et al., 2022)	23.5	22.8	22.8	22.4	23.4	32.1	31.4	31.7	31.5	31.5	55.3	54.4	54.0	53.9	53.9
ViDA (Liu et al., 2024b)	18.5	19.7	18.9	18.5	20.2	26.8	28.4	34.1	62.3	88.0	43.6	42.5	42.3	42.3	42.4
CMAE (Liu et al., 2024a)	12.6	10.9	10.4	10.5	10.4	26.4	–	–	–	–	42.5	–	–	–	–
DPCore (Zhang et al., 2025)	15.4	–	–	–	–	25.1	–	–	–	–	39.9	40.5	40.8	41.2	41.5
KFF (Zhou et al., 2026)	**12.3**	12.5	12.5	12.5	12.6	**22.4**	21.5	21.4	21.2	21.0	**36.0**	**34.8**	34.6	34.6	34.6
IRP (*ours*)	13.5	**10.0**	**9.3**	**9.4**	**9.5**	25.4	**17.6**	**16.2**	**16.0**	**16.1**	44.6	36.5	**33.8**	**33.1**	**32.7**

Note. For CMAE and DPCore, "–" indicates unavailable multi-round results due to limitations of the released code; R1 values are taken from the original papers.

44.5 to 32.8 on ImageNet-C. Relative to second-best KFF, IRP starts with higher error on all three datasets but overtakes it in round 2 on CIFAR10-C and CIFAR100-C, and in round 3 on ImageNet-C. By round 5, the respective error reductions are 3.1, 4.8, and 1.8 percentage points. On CIFAR10-C, IRP also outperforms the reported CMAE results from round 2 onward. Most importantly, IRP is the only method to show consistent, significant improvement across rounds, with relative gains from R1 to R5 of 29.6% on CIFAR10-C, 36.2% on CIFAR100-C, and 26.2% on ImageNet-C. All competing methods remain comparatively stable across repeated adaptation, with some even collapsing. These gains are achieved without excessive computational overhead: as shown in Fig. 3, IRP completes five ImageNet-C rounds in 4083 s, faster than CoTTA, ViDA, DPCore, and KFF. This efficiency largely stems from not requiring multiple test-time augmentations, avoiding a major source of computational cost in several CTTA methods. These results confirm IRP's ability to preserve model knowledge through its *recovery* and *protection* components, while remaining plastic to cross-domain knowledge transfer. Nonetheless, IRP's performance is subpar in the first round, likely because its adapters are trained from scratch rather than initialized from source-domain pretraining (Liu et al., 2024b). We report full per-domain errors for round 5 in Appendix E.

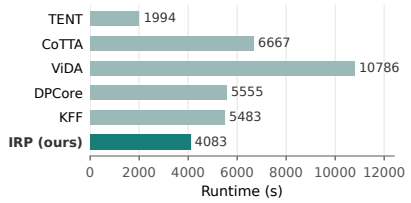


Figure 3: Adaptation time on 5 rounds of ImageNet-C in seconds (s).

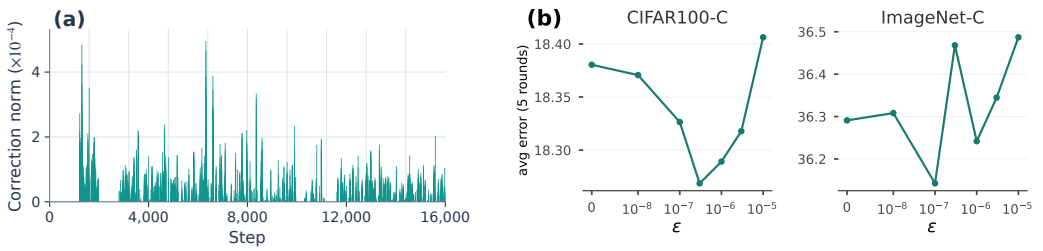
6 Recovery and Protection Analysis

Ablation of IRP components. Table 4 shows that the complete method achieves the best performance, improving over the variant without recovery and protection by 0.50 mIoU. Recovery alone improves adaptation, while protection provides a further gain when combined with recovery but slightly degrades performance when used alone. This interaction is consistent with their complementary roles: recovery adjusts the starting point for a returning domain, whereas protection constrains subsequent conflicting updates. Without recovery, this constraint may hinder updates that would otherwise help restore the model to the current domain. Finally, the full method's advantage over removing both components increases from 0.03 points in round 1 to 0.91 in round 10, supporting the benefit of their combination over repeated domain encounters.

Table 4: Component ablation on Cityscapes-to-ACDC. All Avg is computed over all ten rounds.

Method	R1	R4	R7	R10	All Avg ↑
Source	56.72	56.72	56.72	56.72	56.72
IRP w/o recovery & protection	60.60	62.90	62.12	61.44	62.20
IRP w/o recovery	60.44	62.62	62.17	61.33	62.09
IRP w/o protection	**60.64**	62.89	62.42	61.81	62.32
IRP	60.63	**63.26**	**62.75**	**62.35**	**62.70**

Learning the correction strength. Table 5 compares learned α coefficients with fixed $\alpha = 1$ and random coefficients extracted from a gaussian distribution with $\mu = 0$ and $\sigma = 2$. Learning α exceeds fixed subtraction by 0.53 points in mIoU and the reported random-coefficient variant by 1.75.



(a) (b)

Figure 4: (a) Activation of *protection* component of IRP during adaptation on Cityscapes→ACDC. (b) Ablation of ϵ coefficient regulating tolerance to interfering updates in *protection*.

Random performs the worst, as introducing random coefficients cause the loss of any geometric significance of the *recovery* mechanism. With coefficient set to 1, instead, we still have high-level geometric guarantees, but losing the adaptability given by per-block learnable α, instead of having a common direction for all parameters with the same scale. Overall, these results justify adapting the correction strength in a granular way to the current objective instead of prescribing a fixed subtraction. More analysis on α are provided in Appendix F.

Table 5: Ablation on α recovery hyperparameters on Cityscapes-to-ACDC. All Avg is computed over all ten rounds.

Method	R1	R4	R7	R10	All Avg ↑
Source	56.72	56.72	56.72	56.72	56.72
IRP w/ random α	60.61	62.76	62.12	61.80	60.95
IRP w/ fixed $\alpha = 1$	60.46	62.69	62.19	61.57	62.17
IRP w/ lernable α	**60.63**	**63.26**	**62.75**	**62.35**	**62.70**

Protection during adaptation. Figure 4(a) shows interference corrections throughout the stream, interspersed with intervals of little or no correction. Their varying magnitude is consistent with selective intervention when the proposed displacement conflicts with the stored domain direction. We can observe a few spikes of correction occurring across adaptation, casued by problematic updates that mostly oppose that domain's τ. Together with the component ablation, this shows that protection is active and beneficial within the complete method.

ϵ Ablation Figure 4(b) ablates the tolerance ϵ, which controls how much a proposed update may oppose the protected domain direction before being corrected. On both CIFAR100-C and ImageNet-C, average error follows a similar U-shaped trend: performance improves as ϵ increases from 0, reaching a minimum around $\epsilon = 10^{-7}$. Strict $\epsilon = 0$ enforces non-interference exactly, but overly constrains the optimizer, slightly penalizing plasticity. Conversely, large ϵ values admit more conflicting updates, weakening protection. Overall, the method remains robust to this hyperparameter: error variation are relatively small, indicating that IRP does not require precise tuning of ϵ.

7 Conclusion

We presented Interference Recovery and Protection (IRP), a geometric CTTA framework that decomposes model parameters into domain-general and domain-specific components to explicitly identify and manage interference. IRP combines recovery, which corrects accumulated cross-domain interference to restore a model suited to a returning domain, with protection, which minimally dampens updates that conflict with a domain's previously learned direction to limit error accumulation. Both operate online via reconstruction-based shift detection and gradient-signature domain assignment, requiring no domain labels or boundaries. On segmentation (Cityscapes-to-ACDC) and classification (CIFAR10/100-C, ImageNet-C) benchmarks, IRP achieves state-of-the-art results, with gains growing over successive rounds as domains recur. Ablations confirm recovery and protection are complementary and that learned, per-block correction strength outperforms fixed or random alternatives. Limitations include reliance on sufficiently reliable domain recurrence identification, on checkpoint-derived directions as an approximation of domain-specific knowledge, and on conventional learning losses. While these design choices enable fully online interference management without domain labels or replay, future work will explore online estimation of the transferable component, richer representations of domain-specific subspaces, interference-aware losses and extending IRP beyond adapter-based backbones.

AI USE DISCLOSURE

We used generative AI tools solely to assist with text polishing, formatting, and rewording, and with code and software implementation. All scientific ideas, methodological decisions, and interpretations of results are those of the authors. We reviewed and edited all AI-assisted text and checked AI-assisted code for correctness. We take full responsibility for the final content of this work, including all AI-assisted text and code.

REFERENCES

Shoufa Chen, Chongjian Ge, Zhan Tong, Jiangliu Wang, Yibing Song, Jue Wang, and Ping Luo. Adaptformer: Adapting vision transformers for scalable visual recognition. *Advances in neural information processing systems*, 35:16664–16678, 2022.

Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. IEEE, 2009.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Mehrdad Farajtabar, Navid Azizan, Alex Mott, and Ang Li. Orthogonal gradient descent for continual learning. In *International conference on artificial intelligence and statistics*, pp. 3762–3773. PMLR, 2020.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 15979–15988. IEEE, 2022.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=HJz6tiCqYm.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=6t0Kwf8-jrj.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1412.6980.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Daeun Lee, Jaehong Yoon, and Sung Ju Hwang. BECoTTA: Input-dependent online blending of experts for continual test-time adaptation. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 27072–27093. PMLR, 21–27 Jul 2024.

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

Taejun Lim, Joong-Won Hwang, and Kibok Lee. When and where to reset matters for long-term test-time adaptation. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 74988–75015, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/79bdd6fe3f012befcc459ad13de65d13-Paper-Conference.pdf.

Jiaming Liu, Ran Xu, Senqiao Yang, Renrui Zhang, Qizhe Zhang, Zehui Chen, Yandong Guo, and Shanghang Zhang. Continual-mae: Adaptive distribution masked autoencoders for continual test-time adaptation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28653–28663. IEEE, 2024a.

Jiaming Liu, Senqiao Yang, Peidong Jia, Renrui Zhang, Ming Lu, Yandong Guo, Wei Xue, and Shanghang Zhang. Vida: Homeostatic visual domain adapter for continual test time adaptation. In *International conference on learning representations*, volume 2024, pp. 48396–48417, 2024b.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pp. 109–165. Elsevier, 1989.

Hyewon Park, Hyejin Park, Jueun Ko, and Dongbo Min. Hybrid-tta: Continual test-time adaptation via dynamic domain shift detection. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2877–2886. IEEE, 2025.

Gobinda Saha, Isha Garg, and Kaushik Roy. Gradient projection memory for continual learning. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=3AOj0RCNC2.

Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10745–10755. IEEE, 2021.

Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. *Advances in neural information processing systems*, 33:11539–11551, 2020.

Junha Song, Jungsoo Lee, In So Kweon, and Sungha Choi. Ecotta: Memory-efficient continual test-time adaptation via self-distilled regularization. In *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11920–11929. IEEE, 2023.

Yuntong Tian, Kang Li, Tianyang He, Liang Wan, Pheng-Ann Heng, and Wei Feng. Dual domain-attribute learning framework with asynchronous adapters for continual test-time adaptation. *IEEE Transactions on Image Processing*, 2026.

Guillaume Vray, Devavrat Tomar, Xufeng Gao, Jean-Philippe Thiran, Evan Shelhamer, and Behzad Bozorgtabar. Reservoirtta: Prolonged test-time adaptation for evolving and recurring domains. *Advances in Neural Information Processing Systems*, 38:63357–63403, 2026.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7191–7201. IEEE, 2022.

Yongxian Wei, Anke Tang, Li Shen, Zixuan Hu, Chun Yuan, and Xiaochun Cao. Modeling multi-task model merging as adaptive projective gradient descent. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=EqoKRSR5Pa.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International conference on machine learning*, pp. 23965–23998. Pmlr, 2022.

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021.

Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: A simple framework for masked image modeling. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 9643–9653. IEEE, 2022.

Senqiao Yang, Jiarui Wu, Jiaming Liu, Xiaoqi Li, Qizhe Zhang, Mingjie Pan, Yulu Gan, Zehui Chen, and Shanghang Zhang. Exploring sparse visual prompt for domain adaptive dense prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16334–16342, 2024.

Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International conference on machine learning*, pp. 3987–3995. Pmlr, 2017.

Yunbei Zhang, Akshay Mehra, Shuaicheng Niu, and Jihun Hamm. DPCore: Dynamic prompt coreset for continual test-time adaptation. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 75757–75778. PMLR, 13–19 Jul 2025.

Jiahuan Zhou, Chao Zhu, Zhenyu Cui, Zichen Liu, Xu Zou, and Gang Hua. Class-aware domain knowledge fusion and fission for continual test-time adaptation. *Advances in Neural Information Processing Systems*, 38:55735–55759, 2026.

A Domain Shift Detection and Assignment Details

A.1 Reconstruction-Based Shift Detection.

Let $\mathcal{L}^s_{\mathrm{rec},t}$ denote the reconstruction loss in Eq. equation 1, evaluated on the current input using the frozen source model f_{θ_s}. We denoise this loss with an Exponential Moving Average (EMA) to obtain S_t, in order to eliminate misleading instantaneous fluctuations in the loss value. Subsequently, we maintain EMA estimates of both mean μ_t and standard deviation σ_t of S_t that are used to determine the EMA band. A shift is detected at time t^* when

$$S_{t^*} \notin \left[\mu_{t^*-1} - c \cdot \sigma_{t^*-1}, \mu_{t^*-1} + c \cdot \sigma_{t^*-1}\right], \tag{14}$$

where c is a multiplicative constant for the standard deviation. This two-sided criterion captures both increases and decreases in reconstruction loss. After a detection, the detector is temporarily disabled during a warmup period 50 steps to prevent repeated triggers while its statistics stabilize.

A.2 Gradient Signature-Based Domain Assignment.

For each known domain k, we maintain the average gradient relative to adapters parameters only:

$$\bar{g}_k^{\tilde{W}} = \frac{1}{|T_k|} \sum_{t \in T_k} \nabla_{\tilde{W}} \mathcal{L}_{\mathrm{total},t}|_{\theta=\theta_s} \ , \tag{15}$$

where T_k contains the steps assigned to domain k. The loss follows Eq. equation 2, with pseudo-targets supplied by the EMA teacher $f_{\theta'}$. Gradients are evaluated at fixed source parameters, rather than at the evolving adapted model: in this way, the gradient of the adaptation loss evaluated at $\tilde{W}_s$ provides a first-order estimate of the direction in which a domain pulls the model, which remains comparable across domain encounters and unconfounded by the adapter's accumulated drift.

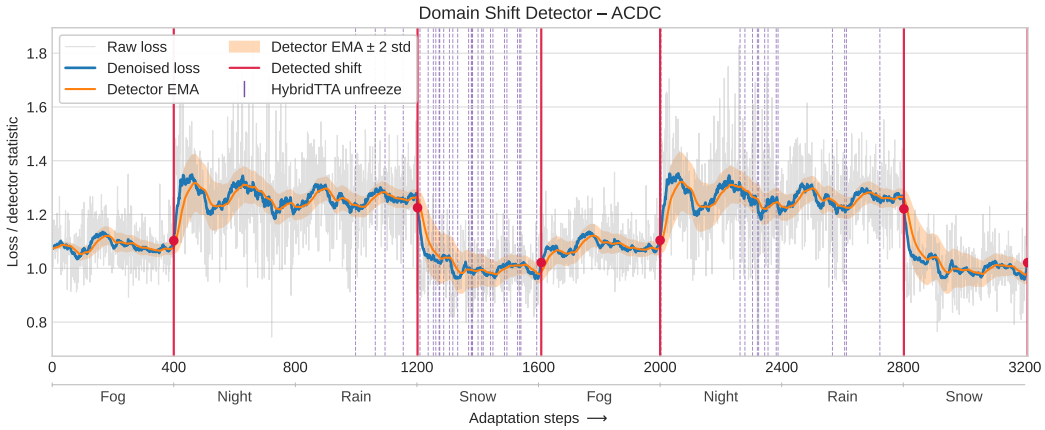


Figure 5: IRP domain shift detector performance on ACDC adaptation, focus on the first two rounds. Comparison with Hybrid-TTA unfreeze domain-shift criterion.

Following a detected shift, we collect gradients over a few steps and average them into a probe signature $\bar{g}_*^{\tilde{W}}$. No interference correction is applied during this observation window. Domain assignment is then determined by cosine similarity S_c against all stored signatures of seen domains. A minimum threshold $\beta_{\text{novel}} \in (0, 1)$ on cosine similarity discerns if the probe signature is assigned to an existing domain, or a novel one:

$$\text{assign}(k^*) = \begin{cases} k^* = \arg\max_{k \in \{1,\dots,K\}} S_c\left(\bar{g}_k^{\tilde{W}}, \bar{g}_*^{\tilde{W}}\right), & \text{if } \max_k S_c\left(\bar{g}_k^{\tilde{W}}, \bar{g}_*^{\tilde{W}}\right) \geq \beta_{\text{novel}} \\ K \leftarrow K + 1, & \text{otherwise} \end{cases} \tag{16}$$

In the first case, the model is assigned to the existing domain k^*, IRP applies the adaptive recovery in Eq. equation 11 and activates protection along its stored domain-specific direction.. At the next detected shift, the adapter checkpoint $\tilde{W}_{k^*}$ and accumulated gradient signature of the departing domain are stored or updated. In the second case, the current data is declared a novel domain and a new adapter domain checkpoint $\tilde{W}_{k+1}$ and respective gradient signature are saved once a new shift is detected.

B Analysis on Domain Shift Detector

Detection on ACDC. Figure 5 shows the EMA-band detector correctly flags most domain boundaries in the Fog→Night→Rain→Snow sequence, with the band visibly widening during noisier segments (Night, Rain) and narrowing once the statistic stabilizes. One transition is consistently missed: Night→Rain, likely because Night's already-elevated, high-variance loss leaves a band wide enough to absorb the subsequent change. For reference, we overlay Hybrid-TTA's (Park et al., 2025) unfreeze triggers, which fire densely within the Rain and Snow segments rather than at exact boundaries proving to be compatible with IRP mechanism which require clean boundaries.

Detection on CIFAR10-C. Figure 6 shows the detector reliably catches boundaries with large reconstruction-loss jumps (e.g., *impulse_noise*, *snow*, *frost*, *brightness*), but misses several within the blur family (*defocus/glass/motion/zoom_blur*) and among the digital corruptions family, where the loss stays essentially flat across the labeled boundary. This reflects the detection signal rather than a method failure: reconstruction loss on the frozen source model does not distinguish corruptions belonging to the same group that are perceptually or statistically similar, even if benchmark labels do. Grouping such corruptions as one continuing domain is benign and coherent with IRP goals, since recovery and protection only need to counteract interference between genuinely dissimilar domains.

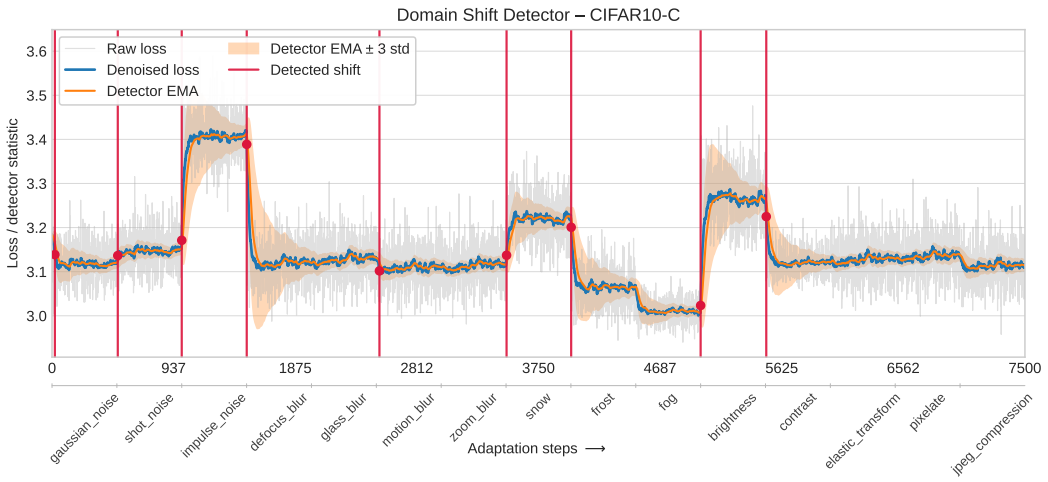


Figure 6: IRP domain shift detector performance on CIFAR10-C adaptation, focus on the first round.

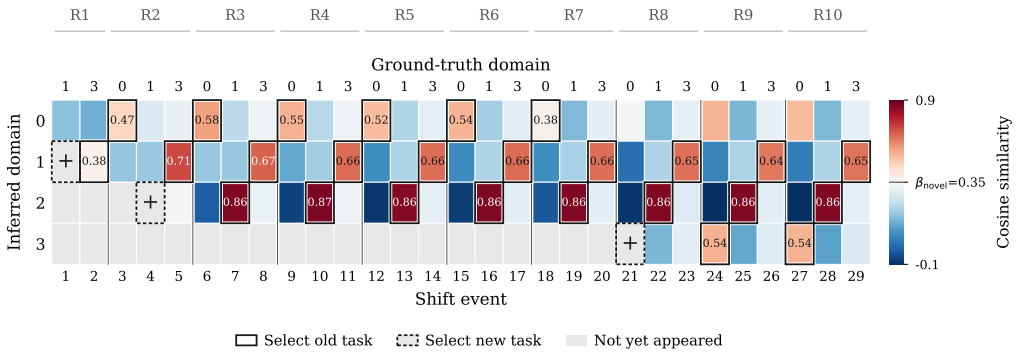


Figure 7: Domain assignment based on gradient domain signature similarities on Cityscapes-to-ACDC adaptation.

C Analysis on Domain Assigner

Assignment on ACDC. Figure 7 shows that once a domain's gradient signature is established, subsequent recurrences are assigned to it with cosine similarities comfortably above $\beta_{\text{novel}} = 0.35$ (e.g., $\sim$0.86–0.87 for one domain, $\sim$0.65–0.71 for another), and these similarities remain stable across rounds rather than drifting, indicating that signatures stay representative of a domain over many revisits. New domains are registered only once, at their true first occurrence ('+' markers), and every subsequent encounter is correctly matched to an existing signature (solid borders) rather than mistakenly spawning a new one. One assignment is a close call: the second shift event,registers at 0.38, only marginally above threshold, This can be explained by the missed Night→Rain detection noted in Fig. 5, which causes to have a single, less clear signature for the combination of the two domains. Gradient signatures captured over merged or noisier observation windows are consequently less cleanly separated from others.

Assignment on CIFAR10-C. Figure 8 shows a similar pattern with a much higher threshold ($\beta_{\text{novel}} = 0.90$), reflecting more homogeneous domains encountered in the CTTA stream. Correct re-matches score 0.93–0.99, well clear of the boundary, and new domains are introduced one at a time ('+') rather than all in the first round, implying that learning process needs multiple rounds to isolate visually similar corruptions. This is coherent with IRP performance reported in Table 3: representations improve and beat competitors only after the first round of adaptation.

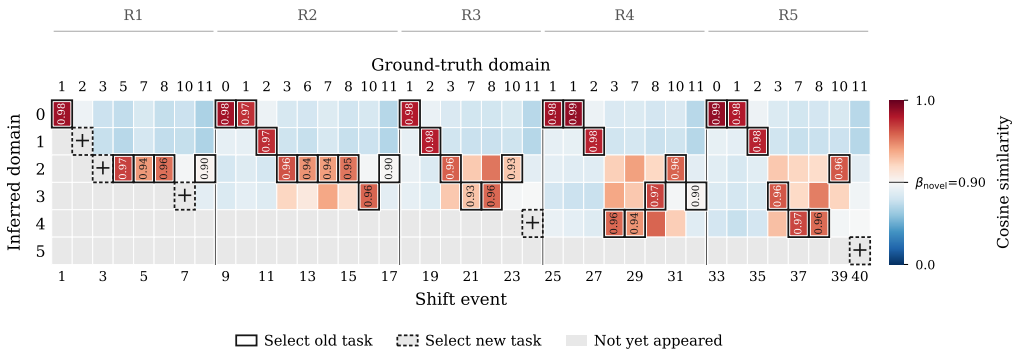


Figure 8: Domain assignment based on gradient domain signature similarities on CIFAR10-C adaptation.

D Further Implementation Details

For all experiments, we set the weight of the reconstruction loss $\lambda_r = 0.3$, following Park et al. (2025). For each *recovery* operation, the set of coefficients $\boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{B}|}$ is learned in $S_\alpha = 20$ steps over the latest minibatch, using the adaptation loss $\mathcal{L}_{\text{total}}$ as the objective and with $\text{lr} = 0.1$. Our method does not employ multiple augmentations. Regarding the domain shift detector, we update the EMA estimates of mean μ_t, standard deviation σ_t and the denoising EMA with discount factors of 0.95, 0.99 and 0.95 respectively for all IRP experiments; the standard deviation coefficient c from Equation 14 is always set to 3. Instead, for the domain assigner, we set a different similarity threshold for each benchmark: 0.35 for ACDC, 0.90 for CIFAR10-C, 0.70 for CIFAR100-C, 0.65 for ImageNet-C. For all runs we average the gradient over 10 steps to calculate the probe gradient signature $\bar{g}_*^{\tilde{W}}$ used to infer the current domain. We run all experiments on a single NVIDIA A100 GPU with either 40 or 80 GB of VRAM.

E CTTA Classification per-Domain Errors

Tables 6, 7, and 8 report error rates at round 5 (the final round) of continual adaptation on CIFAR10-C, CIFAR100-C, and ImageNet-C, respectively, allowing us to assess robustness to long-horizon error accumulation rather than single-round performance.

Across all three benchmarks, IRP achieves the lowest mean error at round 5, with 9.5% on CIFAR10-C, 16.1% on CIFAR100-C, and 32.7% on ImageNet-C, indicating that our method degrades the least as adaptation progresses.

On CIFAR10-C (Table 6), IRP and CMAE perform comparably, with IRP outperforming on the majority of corruptions ($10/15$), particularly structured distortions such as *glass*, *motion*, *zoom*, and *elastic_trans*, while CMAE remains stronger on noise-type corruptions (*Gaussian*, *shot*, *contrast*).

CIFAR100-C (Table 7) most clearly illustrates the instability of naive continual adaptation: TENT-continual collapses to near-random performance (98.9% mean error) by round 5, and ViDA similarly degrades substantially (88.0%), both underperforming even the unadapted source model. In contrast, IRP avoids this collapse entirely, outperforming the strongest stable baseline (KFF) by ~ 5 points and CoTTA by ~ 15 points, underscoring the importance of explicit mechanisms against error accumulation in long-horizon CTTA.

On ImageNet-C (Table 8), IRP's margin narrows but remains consistent, improving on KFF by ~ 2 points and on DPCore/ViDA by ~ 9–10 points. IRP is not uniformly best per-corruption – KFF is stronger on *frost*, *fog*, *brightness*, *contrast*, and *pixelate* – suggesting IRP's gains are concentrated on harder, blur/weather-type corruptions (*defocus*, *glass*, *motion*, *zoom*) rather than uniform across all shift types.

Table 6: Error rate (↓) for CIFAR10→CIFAR10-C under CTTA setting. We show error on each at round 5 of adaptation (last round).

Method	Gaussian	shot	impulse	defocus	glass	motion	zoom	snow	frost	fog	brightness	contrast	elastic_trans	pixelate	jpeg	Mean↓
Source	60.1	53.2	38.3	19.9	35.5	22.6	18.6	12.1	12.7	22.8	5.3	49.7	23.6	24.7	23.1	28.2
TENT-continual	51.1	46.5	20.0	11.1	27.1	13.0	10.2	8.4	8.9	13.0	4.3	21.9	15.0	27.6	19.1	19.8
CoTTA	59.5	52.0	35.1	17.4	32.5	17.4	12.7	9.5	9.4	14.0	3.9	24.5	18.3	25.6	18.9	23.4
ViDA	58.8	51.4	24.1	11.6	29.9	12.3	6.9	7.4	9.1	11.2	33.2	22.6	12.6	26.	15.5	20.2
CMAE	**17.9**	**17.4**	**8.4**	**7.4**	22.1	8.7	6.6	6.7	6.4	7.2	3.9	**6.8**	12.8	10.7	13.4	10.4
KFF	19.9	**17.4**	12.6	10.1	21.6	11.8	8.9	9.2	7.7	10.6	4.2	11.4	14.6	12.1	17.3	12.6
IRP (*ours*)	24.6	18.6	8.5	**7.4**	**12.1**	**6.9**	**5.5**	**5.7**	**5.0**	**7.1**	**3.6**	9.1	**8.8**	**8.6**	**11.5**	**9.5**

Table 7: Error rate (↓) for CIFAR100→CIFAR100-C under CTTA setting. We show error on each at round 5 of adaptation (last round).

Method	Gaussian	shot	impulse	defocus	glass	motion	zoom	snow	frost	fog	brightness	contrast	elastic_trans	pixelate	jpeg	Mean↓
Source	55.0	51.5	26.9	24.0	60.5	29.0	21.4	21.1	25.0	35.2	11.8	34.8	43.2	56.0	35.9	35.4
TENT-continual	98.8	98.9	98.8	98.1	99.0	98.8	98.6	98.9	99.0	99.0	98.9	98.9	99.0	99.0	99.0	98.9
CoTTA	50.7	45.7	23.2	24.1	52.2	27.0	19.4	19.6	20.9	29.7	12.0	30.3	37.4	48.1	32.4	31.5
ViDA	81.0	86.2	80.8	82.1	91.5	85.1	82.2	84.3	87.3	92.2	86.1	95.2	95.0	95.8	95.5	88.0
KFF	29.7	27.6	13.7	19.2	32.8	19.4	15.3	18.8	17.4	19.5	**10.7**	**13.2**	28.1	21.5	28.5	21.01
IRP (*ours*)	**18.0**	**16.9**	**12.7**	**14.7**	**17.1**	**14.7**	**14.4**	**14.9**	**13.9**	**19.2**	13.3	19.0	**15.9**	**14.5**	**21.9**	**16.1**

F MORE α ANALYSIS

Figure 9(a) reveals structured variation in the learned α coefficients across blocks and recovery events. Later stages predominantly exhibit positive coefficients, whereas earlier and intermediate blocks show more heterogeneous signs and magnitudes. This is consistent with lower layers encoding more transferable low-level features, for which strong recovery may be unnecessary, while higher-level representations can require more domain-specific correction. Occasional negative coefficients reflect the flexibility of the learned scaling to locally adjust the nominal geometric correction when this better matches the current adaptation objective. The normalized correction magnitude $|\alpha\phi|/|W|$ (see Fig. 9(b)) follows the magnitude of the learned coefficients, indicating that this block-wise modulation translates into meaningful parameter changes. Overall, the structured variation across depth supports learning block-specific correction strengths rather than applying uniform interference removal.

Table 8: Error rate (↓) for ImageNet→ImageNet-C under CTTA setting. We show error on each at round 5 of adaptation (last round).

Method	Gaussian	shot	impulse	defocus	glass	motion	zoom	snow	frost	fog	brightness	contrast	elastic_trans	pixelate	jpeg	Mean↓
Source	53.0	51.8	52.1	68.5	78.8	58.5	63.3	49.9	54.2	57.7	26.4	91.4	57.5	38.0	36.2	55.8
TENT-continual	47.9	45.8	46.8	51.9	59.1	46.4	52.9	40.4	44.9	41.1	24.2	54.5	51.3	33.4	33.5	45.0
CoTTA	49.7	48.2	49.1	64.8	74.9	57.0	61.2	46.5	50.1	58.6	26.5	96.0	55.3	36.1	33.9	53.9
ViDA	44.4	42.3	43.0	46.6	54.4	43.4	47.7	38.1	42.1	40.5	24.5	54.0	41.5	31.2	32.0	42.4
DPCore	42.8	40.2	39.7	53.5	50.3	49.1	48.4	38.6	36.6	41.7	23.7	48.8	43.7	32.1	32.5	41.5
KFF	40.0	37.1	37.5	44.5	46.4	39.4	39.5	32.1	**31.2**	**25.6**	**21.2**	**34.7**	35.9	**26.2**	**28.3**	34.6
IRP (*ours*)	**31.4**	**30.3**	**30.4**	**35.8**	**33.1**	**33.3**	**38.6**	**31.8**	32.9	35.6	28.5	40.7	**28.9**	28.6	30.4	**32.7**

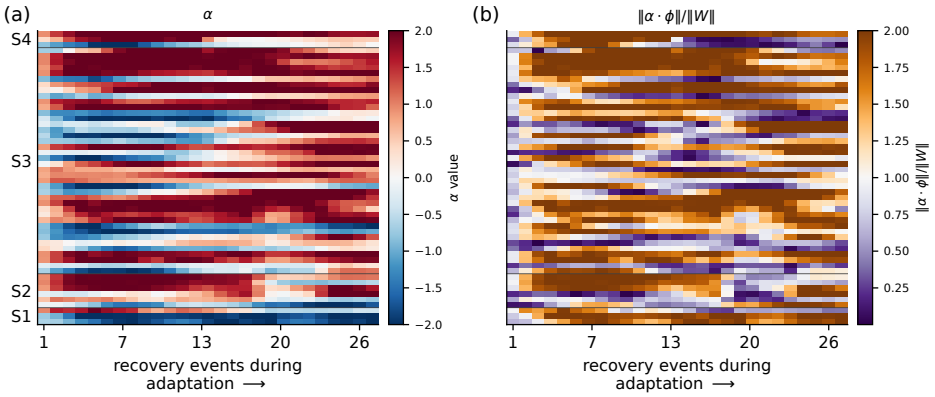


Figure 9: Evolution of the block-wise recovery coefficients α (a) and their corresponding normalized parameter correction (b)$\|\alpha\phi\|/\|W\|$ during ACDC adaptation. Each row denotes a transformer block, ordered by depth.