Palimpsest: Reconciling the CISS Trilemma for Incremental Nuclei Segmentation

Jiajia Li[1], **Huisi Wu**[1*]

[1]College of Computer Science and Software Engineering, Shenzhen University
Shenzhen, China
2400101028@mails.szu.edu.cn, hswu@szu.edu.cn

Abstract

Adapting computational pathology models to evolving clinical diagnostics via Class-Incremental Semantic Segmentation (CISS) is critical. However, this task imposes a unique CISS Trilemma: a simultaneous failure to preserve the intricate tissue background (stability), distinguish morphologically similar new nuclei (plasticity), and maintain a constant model size (scalability), all under a strict exemplar-free constraint. To resolve this, we introduce *Palimpsest*, a novel framework that systematically decouples these conflicting demands. *Palimpsest* integrates three synergistic mechanisms: a Parameter-Conserving Synthesis (PCS) module merges lightweight adapters to ensure scalability; a novel Similarity-Aware Centroid Recalibration (SCR) module executes differentiated recalibration to counteract non-uniform foreground drift, securing plasticity; and an Adaptive Residual Shading (ARS) module performs logit-space decoupling to preserve background integrity, ensuring stability. Extensive experiments on two histopathology datasets demonstrate that *Palimpsest* significantly outperforms state-of-the-art methods, achieving a superior stability-plasticity balance, particularly in challenging long-term incremental scenarios.

Code — https://github.com/jj-sterne/Palimpsest

Introduction

The precise segmentation of nuclei in histopathological images is a fundamental task in computational pathology, providing critical quantitative evidence for cancer diagnosis, grading, and prognostic assessment (Xing and Yang 2016; Hou et al. 2017; Sirinukunwattana et al. 2017). While deep learning models show remarkable success in this task (LeCun, Bengio, and Hinton 2015; Graham et al. 2019; Lal et al. 2021; Wu et al. 2022; Wang, Wu, and Qin 2024), their rigid nature limits their usage in clinical settings, where new pathological entities are continually emerging (Yang et al. 2022; Hou et al. 2017). This reality necessitates a shift towards a lifelong learning paradigm (McCloskey and Cohen 1989), for which Class-Incremental Semantic Segmentation (CISS) offers a promising solution.

However, applying CISS to nuclei segmentation exposes a fundamental and acute challenge, which we formulate as the Histopathology CISS Trilemma (Figure 1). A model must simultaneously satisfy three conflicting objectives: (1) *stability*, the ability to retain knowledge not only of past foreground classes but also of the complex, information-rich tissue background, (2) *plasticity*, the capacity to efficiently learn new, often morphologically similar, classes while maintaining a constant computational footprint for clinical scalability, and (3) *exemplar-free learning* constraint, a non-negotiable mandate in medicine prohibiting the reuse of past patient data.

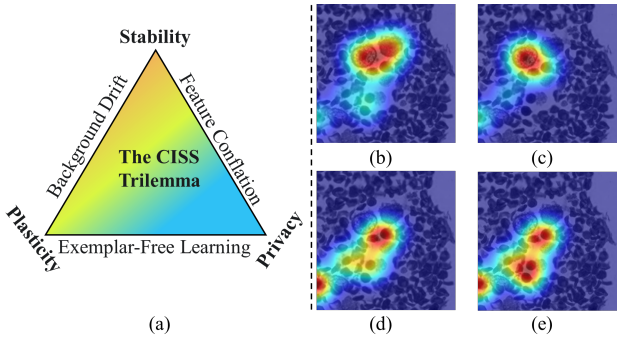

Figure 1: The CISS Trilemma in Computational Pathology. (a) The fundamental trade-off between Stability, Plasticity, and Scalability under an exemplar-free constraint. (b-e) Illustrative failure modes: compared to the (b) ground-truth, models may exhibit (c) catastrophic forgetting, (d) poor plasticity against the background, or (e) confusion between similar classes.

Existing attempts to resolve this trilemma are trapped by their methodological choices, particularly when facing the complexities of histopathology. The dominant knowledge distillation-based approaches (Li and Hoiem 2017a; Cermelli et al. 2020; Zhang et al. 2022) enforce a monolithic regularization that, while preserving old knowledge, inherently strangles the fine-grained plasticity needed to distinguish morphologically similar nuclei and fails to prevent corruption of the intricate tissue background. Architectural expansion methods (Wang et al. 2022; Lu et al. 2024; Gao et al. 2022), which isolate parameters to guarantee stability, introduce an unsustainable, linear growth in model complexity, a critical barrier for resource-constrained clinical de-

*Corresponding Author

ployment. Even the most advanced exemplar-free strategies fall short in this specific context. While some employ proxy data from background regions or generative models (Cha et al. 2021; Mei et al. 2024; Zhuang et al. 2024), they provide only indirect supervision, insufficient for carving out new classes from the information-rich tissue landscape. Others that attempt to directly counteract representation drift often do so with a uniform correction (Fukuda, Kera, and Kawamoto 2025), failing to recognize that drift is non-uniform and disproportionately impacts classes with high morphological similarity. Thus, the field remains trapped, lacking a holistic strategy that can navigate the specific demands of the CISS Trilemma.

To resolve this, we introduce *Palimpsest*, whose novelty lies in a principled architecture that systematically decouples these demands through causal synergy. To ensure scalability, a Parameter-Conserving Synthesis (PCS) module first merges lightweight adapters. This process, however, inevitably induces feature drift, which our pathology-aware Similarity-Aware Centroid Recalibration (SCR) module is then specifically designed to correct. Concurrently, an Adaptive Residual Shading (ARS) module preserves background integrity via logit-space decoupling, thus isolating background stability from foreground plasticity. Our main contributions can be summarized as follows.

- We propose *Palimpsest*, the first decoupled framework to systematically resolve the Histopathology CISS Trilemma, by assigning dedicated mechanisms to the conflicting demands of plasticity, stability, and scalability in a clinical context.
- We introduce three synergistic components tailored for pathology: a SCR to counteract non-uniform foreground drift, an ARS to preserve intricate background integrity via logit-space decoupling, and a PCS to ensure a constant model footprint.
- We establish a new state-of-the-art on challenging pathology benchmarks, demonstrating that *Palimpsest* achieves a superior stability-plasticity balance, particularly in realistic, long-term incremental scenarios.

Related Work

Knowledge Distillation and the Stability-Plasticity Trade-off

Knowledge distillation (KD) forms the bedrock of CISS, aiming to preserve past knowledge by regularizing model outputs or features (Li and Hoiem 2017a,b; Michieli and Zanuttigh 2021). This monolithic regularization, however, creates an intrinsic stability-plasticity trade-off. By penalizing any deviation from the old model, it inherently suppresses the plasticity required to learn fine-grained features of new, morphologically similar pathological classes. More advanced strategies (Zhang et al. 2022; Upschulte et al. 2023), attempt to mitigate this by introducing learnable modules to explicitly align feature spaces. Nevertheless, these methods are still confined by the core distillation paradigm, failing to address the unique pathological challenge of carving out new classes from a complex tissue background. This necessitates our decoupled approach, where a dedicated mechanism, ARS, preserves background integrity in logit-space without stifling foreground learning.

Architectural Expansion and the Scalability Challenge

Architectural expansion methods, including parameter-isolation and more recent Parameter-Efficient Fine-Tuning (PEFT) techniques (Hu et al. 2022; Wang et al. 2025), prevent forgetting by allocating new parameters for new tasks. While effective, this leads to an unsustainable $O(N)$ growth in model complexity, a disqualifying drawback for resource-constrained clinical deployments. Even forward-looking proposals based on hierarchical prompting or automatic expansion (Douillard et al. 2022; Ye and Bors 2023) are bound to this parameter-accumulating model. Inspired by advances in model merging (Xiao et al. 2023), our PCS mechanism operationalizes this concept for CISS, achieving constant $O(1)$ complexity by systematically fusing adapter weights, thus resolving the scalability barrier for clinical deployment.

Exemplar-Free Recalibration and Representation Drift

The exemplar-free constraint, critical for medical data privacy, introduces the final challenge of representation drift, especially after model updates like our PCS. Existing exemplar-free strategies, which rely on indirect supervision from background regions or generative data, are ill-equipped to correct for this update-induced drift (Cha et al. 2021; Zhuang et al. 2024; Yuan and Zhao 2024). Even recent prototypical methods (Wei et al. 2023; Asadi et al. 2023), while tracking class representations, often lack a dedicated mechanism to explicitly compensate for feature space shifts caused by model consolidation. Those that do attempt compensation typically assume a uniform drift, applying a single global correction to all past classes (Fukuda, Kera, and Kawamoto 2025). This simplistic assumption is flawed in histopathology, where the introduction of a new class disproportionately affects its morphologically similar neighbors. The proposed SCR is a differentiated recalibration, modulating the compensation for each past class based on its similarity to new ones, thus offering a more precise solution to non-uniform representation drift.

Methodology

Problem Formulation

We address Class-Incremental Semantic Segmentation (CISS) in computational pathology, a setting dictated by clinical realities. The learning process unfolds over a sequence of T tasks, $\mathcal{T} = \{\mathcal{T}_1, \ldots, \mathcal{T}_T\}$, where each task $\mathcal{T}_t$ introduces a disjoint set of new pathological entities, denoted as classes $\mathcal{C}_t$. For any two tasks $\mathcal{T}_t$ and $\mathcal{T}_j$ with $t \neq j$, their class sets are disjoint, *i.e.*, $\mathcal{C}_t \cap \mathcal{C}_j = \emptyset$.

During task $\mathcal{T}_t$, the model f is trained on a dataset $\mathcal{D}_t = \{(x, y)\}$, where x is a histopathological image. Crucially, the pixel-level annotation y only provides explicit labels for the new classes $\mathcal{C}_t$. All previously learned classes,

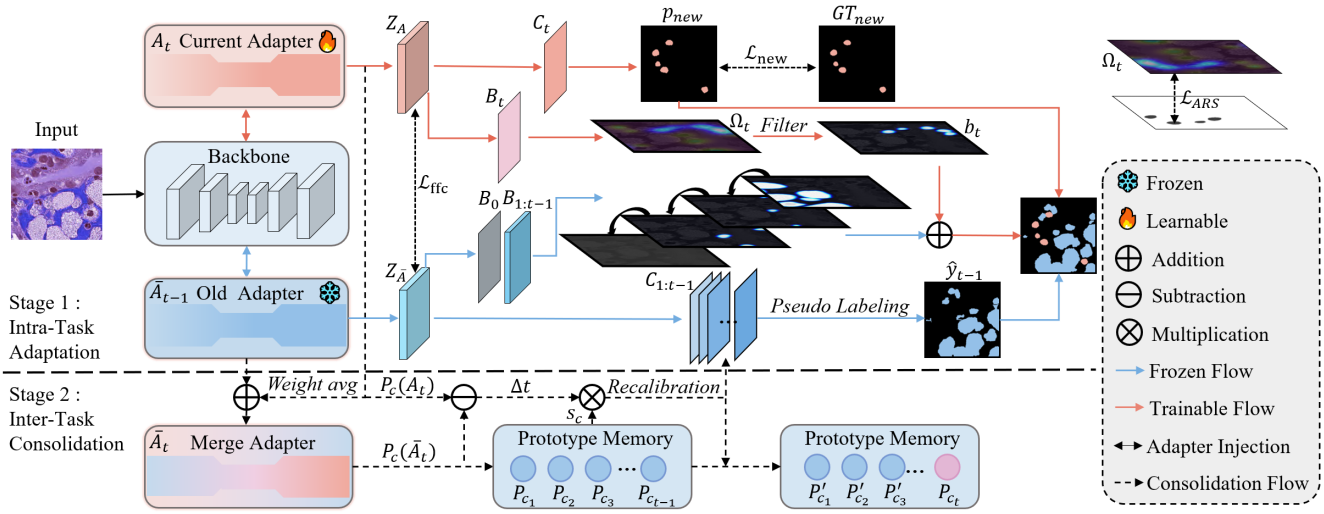

Figure 2: Overview of the Palimpsest framework, which decouples learning into two stages. Stage 1 (Intra-Task Adaptation): A new adapter (A_t) learns new classes while an Adaptive Residual Shading (ARS) module (B_t) preserves background integrity. Stage 2 (Inter-Task Consolidation): Adapters are merged via Parameter-Conserving Synthesis (PCS) for scalability. The resulting feature drift is then corrected by a Similarity-Aware Centroid Recalibration (SCR) module.

$\mathcal{C}_{1:t-1} = \bigcup_{i=1}^{t-1} \mathcal{C}_i$, are absorbed into a single, complex background class c_{bg}. This protocol simulates the clinical workflow where pathologists annotate only the current diagnostic markers, treating past findings as part of the general tissue landscape.

Upon completion of task $\mathcal{T}_t$, the updated model f_t must accurately segment all classes learned to date, $\mathcal{C}_{1:t}$. This must be achieved under two non-negotiable, real-world constraints: a strict exemplar-free paradigm, forbidding access to any past data $\{\mathcal{D}_1, \ldots, \mathcal{D}_{t-1}\}$ to protect patient privacy, and a constant model footprint, precluding any growth in parameters or computational overhead for clinical deployment.

Framework Overview

Our framework, *Palimpsest*, systematically resolves the CISS Trilemma by decoupling the incremental learning process into two synergistic stages for each task $\mathcal{T}_t$: Intra-Task Adaptation and Inter-Task Consolidation. Each stage deploys a specialized mechanism to surgically address a distinct facet of the trilemma.

The first stage, Intra-Task Adaptation, confronts the dual challenge of plasticity and background stability, as illustrated in Figure 2. With the main feature extractor Φ frozen, we introduce two lightweight modules. To ensure plasticity, a task-specific adapter, A_t, is trained to capture the nuanced features of the new pathological classes $\mathcal{C}_t$. Concurrently, to preserve background stability, an ARS module, B_t, operates directly on the logit space. It learns a precise, negative residual that "carves out" new classes from the complex tissue background, leaving the core representation of stable, non-pathological tissues untouched.

The second stage, as shown in Figure 3, Inter-Task Consolidation, addresses scalability and long-term foreground stability in an offline, exemplar-free manner . To achieve scalability, a PCS mechanism merges the newly trained adapter A_t into the single, unified adapter $\bar{A}_t$, guaranteeing a constant $O(1)$ model footprint. This merge, however, inevitably induces a non-uniform representation drift that threatens foreground stability. To counteract this, a SCR module is immediately executed. It leverages only current-task data to estimate the drift and performs a differentiated recalibration of all past class prototypes, robustly restoring foreground knowledge without violating the exemplar-free constraint.

Intra-Task Adaptation: Balancing Plasticity and Stability

During Intra-Task Adaptation, the model must learn new classes $\mathcal{C}_t$ from $\mathcal{D}_t$ while preserving the established knowledge base. This is achieved by freezing the main feature extractor Φ and introducing specialized, lightweight modules to manage the plasticity-stability trade-off. The training is governed by a composite loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{new}} + \lambda_{\text{ARS}}\mathcal{L}_{\text{ARS}} + \lambda_{\text{FFC}}\mathcal{L}_{\text{FFC}}, \tag{1}$$

where λ_{ARS} and λ_{FFC} are balancing hyperparameters.

Plasticity via Task-Specific Adapters. To learn new classes without catastrophic forgetting, we inject a parameter-efficient adapter A_t into each layer of the frozen extractor Φ. Parameterized by θ_t, the adapter A_t refines an intermediate feature map $z \in \mathbb{R}^{H\times W\times d}$ via a bottleneck structure ($r \ll d$): $z' = z + A_t(z)$, where $A_t(z) = \sigma(zW_{\text{down}})W_{\text{up}}$. The resulting feature representation $\Phi_{A_t}(x)$ is then fed to a new classifier for $\mathcal{C}_t$. This learning is driven by a focused cross-entropy loss, $\mathcal{L}_{\text{new}}$, applied only to the pixels Ω_t of new classes:

$$\mathcal{L}_{\text{new}} = -\frac{1}{|\Omega_t|} \sum_{i\in\Omega_t, c\in\mathcal{C}_t} y_{i,c} \log(p_{i,c}). \tag{2}$$

Adaptive Residual Shading. To prevent new classes from corrupting the complex background representation (e.g., stroma), we introduce the ARS module B_t. Instead of altering the shared feature space, B_t operates in the logit space, learning a negative residual that "carves out" the new classes. The final background logit $\mu_{b,t}(x)$ is a dynamic composition of a frozen base logit $\mu_{b,0}$ and all subsequent corrective residuals:

$$\mu_{b,t}(x) = \mu_{b,0}(\Phi(x)) + \sum_{j=1}^{t} B_j(\Phi(x)). \tag{3}$$

To maintain strict $O(1)$ inference complexity, these task-specific residuals $\{B_j\}$ are cumulatively merged into a single module offline. The optimization of B_t is guided by a margin-based loss, which compels a strong negative signal on new class pixels Ω_t and a near-zero signal elsewhere:

$$\mathcal{L}_{\text{ARS}} = \frac{1}{|\Omega_t|} \sum_{i \in \Omega_t} \max(0, B_t(\Phi(x_i)) + m), \tag{4}$$

where $m > 0$ is a margin. This additive architecture ensures that the core background knowledge is never overwritten.

Foreground Stability via Feature Consistency. While the adapter A_t focuses on new classes, its optimization can inadvertently perturb features of old classes. To counteract this, a Foreground Feature Consistency loss, $\mathcal{L}_{\text{FFC}}$, aligns the feature representations. Using pseudo-labels $\bar{y}_{t-1}$ generated by the previous model f_{t-1} to identify old-class pixels Ω_{old}, $\mathcal{L}_{\text{FFC}}$ minimizes the L2 distance between their current and previous feature maps:

$$\mathcal{L}_{\text{FFC}} = \frac{1}{|\Omega_{\text{old}}|} \sum_{i \in \Omega_{\text{old}}} \|\Phi_{A_t}(x_i) - \Phi_{\bar{A}_{t-1}}(x_i)\|_2^2. \tag{5}$$

Here, Φ_{A_t} denotes the extractor augmented with the *current* adapter A_t, while $\Phi_{\bar{A}_{t-1}}$ is the extractor with the *frozen, unified* adapter from the prior consolidation step. For $t = 1$, this term is omitted. This feature-level constraint directly preserves foreground knowledge without resorting to conventional logit-based distillation.

Inter-Task Consolidation: Synthesis and Recalibration

Following the adaptation stage, *Palimpsest* executes an offline, exemplar-free consolidation process. This two-step mechanism guarantees model scalability and restores long-term foreground stability.

Parameter-Conserving Synthesis. To resolve the scalability dilemma, PCS merges the newly trained adapter A_t into the unified adapter $\bar{A}_{t-1}$, precluding the linear growth in model size that is clinically impractical. The parameters of the new unified adapter, $\bar{\theta}_t$, are updated via a weighted average:

$$\bar{\theta}_t = (1 - \alpha_t)\bar{\theta}_{t-1} + \alpha_t \theta_t, \tag{6}$$

where $\alpha_t = 1/t$ is a progressively decreasing weight. This fusion strategy, initialized with $\bar{\theta}_1 = \theta_1$, guarantees a constant $O(1)$ model footprint. However, this parameter fusion inevitably perturbs the feature space, causing a non-uniform representation drift that degrades the recognition

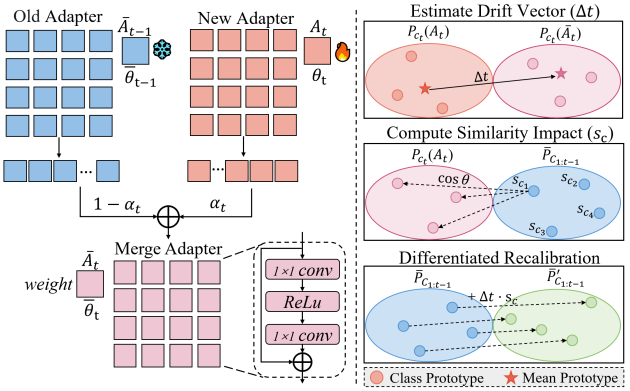

Figure 3: The Inter-Task Consolidation stage for achieving scalability and long-term stability. PCS (left) merges the new adapter's weights into the main model. To correct the inevitable feature drift caused by this merge, the SCR (right) executes a three-step, exemplar-free recalibration.

of past classes, particularly those morphologically similar to the new ones.

Similarity-Aware Centroid Recalibration. To counteract the drift induced by PCS, we introduce the SCR module. This exemplar-free mechanism performs a precise, differentiated recalibration of past class prototypes in three steps, as illustrated in Figure 3.

Drift Estimation: First, as an efficient first-order approximation of the dominant drift, we estimate a single global vector Δ_t using the current task's validation data D'_t. It measures the average feature displacement of new class prototypes before (A_t) and after ($\bar{A}_t$) the PCS merge:

$$\Delta_t = \frac{1}{|\mathcal{C}_t|} \sum_{c \in \mathcal{C}_t} (P_c(\bar{A}_t) - P_c(A_t)), \tag{7}$$

where $P_c(A) = \mathbb{E}_{x,y \sim \mathcal{D}'_t | c \in y}[\Phi_A(x)]$ is the prototype for class c under adapter A.

Impact Modulation: Next, SCR computes a modulation factor $s_{c'}$ for each past class $c' \in \mathcal{C}_{1:t-1}$. This factor quantifies the drift's likely impact based on the feature-space similarity between old and new classes. A higher similarity implies a stronger impact.

$$s_{c'} = \frac{1}{|\mathcal{C}_t|} \sum_{c \in \mathcal{C}_t} \text{sim}(\bar{P}_{c'}, P_c(A_t)), \tag{8}$$

where $\bar{P}_{c'}$ is the stored prototype of the old class c' from the previous step, $P_c(A_t)$ is the prototype of the new class before drift, and $\text{sim}(\cdot, \cdot)$ is the cosine similarity.

Differentiated Recalibration: Finally, SCR recalibrates each past prototype with surgical precision. The corrected prototypes $P'_{c'}$ are then directly imprinted as the new classifier weights for past classes $C_{1:t-1}$, an exemplar-free mechanism that surgically corrects decision boundaries without retraining. The update is defined as:

$$P'_{c'} = P_{c'} + s_{c'} \cdot \Delta_t, \quad \forall c' \in C_{1:t-1}. \tag{9}$$

Method	1-1 (4 steps)			2-1 (3 steps)			2-2 (2 steps)			3-1 (2 steps)		
	Old	New	Mean	Old	New	Mean	Old	New	Mean	Old	New	Mean
Finetune	*5.12*	*45.33*	*23.23*	*8.78*	*48.12*	*27.45*	*10.21*	*49.85*	*29.03*	*12.54*	*51.06*	*31.80*
LwF (Li and Hoiem 2017b)	29.25	51.18	43.21	32.33	54.05	43.69	39.61	57.12	48.87	42.17	64.34	54.23
MiB (Cermelli et al. 2020)	52.78	65.03	58.91	54.01	68.99	61.50	56.15	70.04	63.10	58.73	71.98	65.36
PLOP (Li and Hoiem 2017a)	54.11	64.21	59.16	55.98	70.83	62.41	58.03	71.95	64.99	60.29	73.17	66.73
EWF (Xiao et al. 2023)	63.55	68.14	65.85	60.21	75.61	67.92	64.92	73.03	69.48	69.04	75.81	70.93
BalConpas (Chen et al. 2024)	65.03	69.29	67.16	60.88	74.95	67.06	64.11	73.56	69.05	70.15	74.99	71.57
InSeg (Wang, Wu, and Qin 2024)	67.21	69.83	69.42	63.59	76.35	69.66	65.43	**76.90**	71.32	70.46	**77.81**	72.04
Palimpsest	**69.15**	**72.41**	**71.28**	**66.38**	**77.58**	**71.57**	**69.13**	76.59	**73.16**	**71.81**	77.75	**73.38**
Joint	*73.16*	*75.63*	*74.58*	*70.92*	*78.86*	*74.58*	*71.48*	*78.84*	*74.58*	*73.16*	*79.47*	*74.58*

Table 1: State-of-the-art comparison on the MoNuSAC dataset. All results are reported in mean Dice (%).

Method	1-1 (3 steps)			2-1 (2 steps)		
	Old	New	Mean	Old	New	Mean
LwF	39.11	43.34	41.46	48.23	53.27	50.82
MiB	67.02	67.06	67.07	70.99	61.96	67.93
PLOP	67.06	67.78	67.54	72.50	63.15	69.38
EWF	67.18	70.03	69.08	74.15	63.00	70.44
BalConpas	67.20	68.28	67.56	72.30	62.72	69.11
InSeg	66.64	**70.84**	69.44	74.29	**64.65**	71.08
Palimpsest	**68.53**	70.66	**70.12**	**75.59**	64.55	**71.84**

Table 2: State-of-the-art comparison on the CoNSeP dataset. All results are reported in mean Dice (%).

Method	PCS	ARS	SCR	Old	New	Mean
Baseline				15.82	**78.05**	46.94
+ PCS	✓			11.27	77.81	44.54
+ PCS + ARS	✓	✓		58.91	77.75	68.33
Palimpsest	✓	✓	✓	**66.38**	77.58	**71.57**

Table 3: Ablation study demonstrating the synergy of Palimpsest's components.

Experiments

Experimental Setup

Datasets. We validate our framework on two public, challenging histopathological datasets, strategically chosen to test distinct facets of the CISS problem. MoNuSAC (Verma et al. 2021), featuring four distinct cell types (Epithelial, Lymphocyte, Macrophage, Neutrophil) from multiple organs, serves as an ideal benchmark for multi-class catastrophic forgetting. CoNSeP (Graham et al. 2019), containing morphologically similar and densely packed nuclei from colorectal cancer tissues, presents a significant challenge for fine-grained distinction. We group its original classes into three classes (Epithelial, Spindle-shaped, and others) to construct a clear incremental task. All datasets are pre-processed and split following standard protocols.

Incremental Protocols. We adopt the realistic and challenging *overlapped* setting, which simulates the clinical reality of non-exhaustive annotation where training data for a new task may contain unlabeled instances of past or future classes, treating them as background. We design scenarios denoted as "N_{old}-N_{new}", where N_{old} and N_{new} are the number of initial and incrementally added classes, respectively. We focus on demanding settings like 3-1 and 2-1 for MoNuSAC, and 2-1 for CoNSeP. Given the high morphological similarity among nuclei types, these settings constitute a semantically challenging scenario, testing the model's fine-grained learning capacity more than sheer class quantity would.

Evaluation Metrics. To comprehensively assess performance, we report three standard metrics as percentages (%): Old, the average Dice Similarity Coefficient (DSC) over previously learned classes to measure stability; New, the average DSC over newly introduced classes to measure plasticity; and Mean, the average DSC over all classes seen so far to reflect overall performance. All results are averaged over 3 random seeds.

Implementation Details. Our framework is built upon a ResNet-101 (He et al. 2016) encoder, pre-trained on ImageNet (Deng et al. 2009), with a bottleneck dimension of $r = 16$ for all lightweight adapters. The key hyperparameters, the loss weight λ_{ARS} and the margin m for L_{ARS}, are set to 0.5 and 0.1, respectively. We employ the SGD optimizer (Robbins and Monro 1951) and train for 100 epochs per incremental step with a batch size of 12, utilizing a cosine annealing learning rate schedule. All experiments are implemented in PyTorch (Paszke et al. 2019) on NVIDIA 2080Ti GPUs. Further details on experimental setup and additional results are provided in the appendix.

Comparative Experiments

Quantitative Analysis. As presented in Table 1 and Table 2, *Palimpsest* demonstrates state-of-the-art performance, achieving a superior stability-plasticity balance across all incremental scenarios on both the MoNuSAC and CoNSeP datasets.

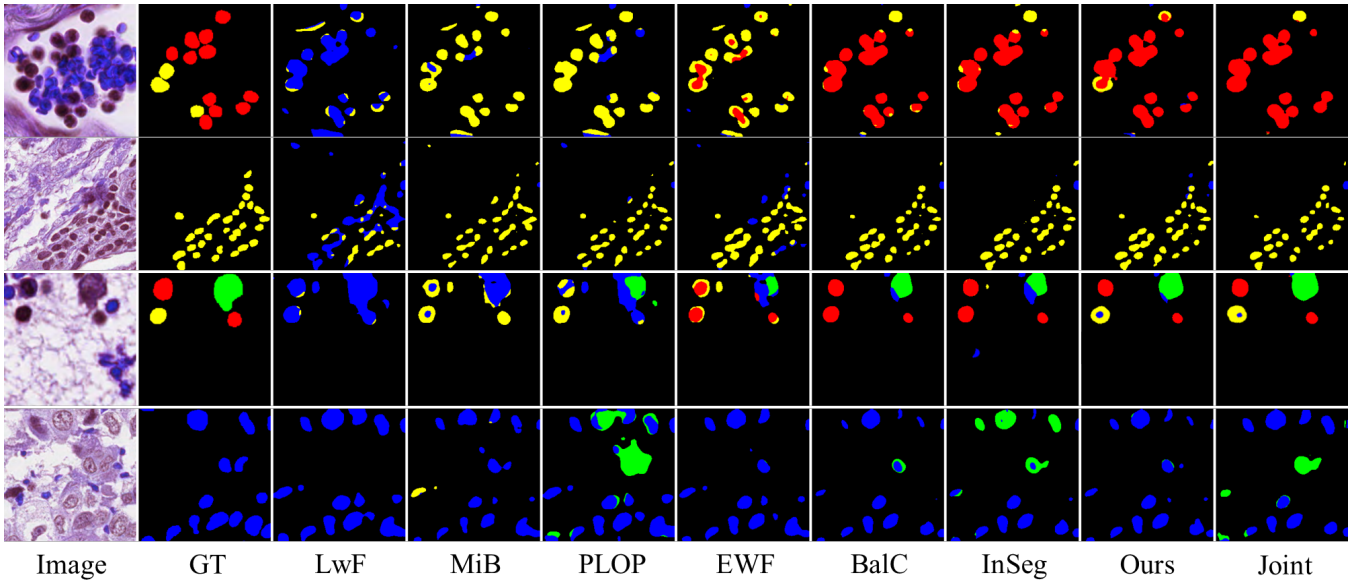

Figure 4: Qualitative segmentation results on the MoNuSAC dataset, visually demonstrating the performance under various incremental settings. Palimpsest consistently outperforms other methods in preserving past knowledge and learning new classes, closely approaching the quality of the Joint training upper bound.

$\mathcal{L}_{\text{new}}$	$\mathcal{L}_{\text{ARS}}$	$\mathcal{L}_{\text{FFC}}$	MoNuSAC		
			Old	New	Mean
✓			11.27	**77.81**	44.54
✓	✓		54.16	77.72	65.94
✓	✓	✓	**66.38**	77.58	**71.57**

Table 4: Ablation study of the loss components for Intra-Task Adaptation.

Strategy	Old	New	Mean
No Calibration & Prototypes	25.13	77.51	51.32
Random Recalibration	19.82	77.45	48.64
Uniform Recalibration	55.31	77.54	62.76
Naive Replay	*67.92*	*77.85*	*72.89*
Palimpsest (w/ SCR)	**66.38**	**77.58**	**71.57**

Table 5: Ablation study of foreground calibration strategies.

Qualitative and Methodological Analysis. The limitations of competing methods become evident when analyzed through the lens of the trilemma. Knowledge distillation-based are ensnared by the stability-plasticity trade-off. Their enforce a monolithic regularization, while preserving old knowledge, inherently strangles the plasticity required for fine-grained distinctions. Conversely, architectural-expansion methods directly sacrifice scalability. While they guarantee stability by isolating parameters, their linear $O(N)$ parameter growth is unsustainable for clinical deployment where a constant model footprint is non-negotiable. In stark contrast, *Palimpsest*'s decoupled design gracefully navigates this trilemma. As shown in Figure 4, it accurately segments both old and new nuclei, achieving high plasticity for new classes while robustly preserving both background and foreground knowledge, all within a constant-memory budget.

Ablation Studies

We conduct extensive ablation studies on the MoNuSAC 2-1 benchmark to dissect our framework and validate each design choice. All results are reported in Table 3, 4, and 5. **Effectiveness of Core Components.** As shown in Table 3, our components demonstrate critical synergy. The baseline, training a new adapter without consolidation, fails due to catastrophic forgetting. Applying only the PCS module to merge adapters for scalability further degrades old-class performance due to severe feature drift. Our decoupled stability mechanisms systematically resolve this. As shown in Figure 8, the ARS module restores background integrity, and the SCR module precisely corrects foreground drift. The t-SNE visualization in Figure 7 and the step-by-step qualitative results in Figure 5 visually corroborate this process, showing how SCR mitigates feature confusion and how each module progressively refines the segmentation.The trade-off in the PCS merging coefficient α_t is analyzed in Figure 6.

Analysis of Loss Components. The design of our loss functions for balancing stability and plasticity during intra-task adaptation.As shown in Table 4 , removing the foreground feature consistency loss ($\mathcal{L}_{\text{FFC}}$) leads to a drop in old-class performance, indicating its importance for stability. More critically, replacing our margin-based $\mathcal{L}_{\text{ARS}}$ with a standard

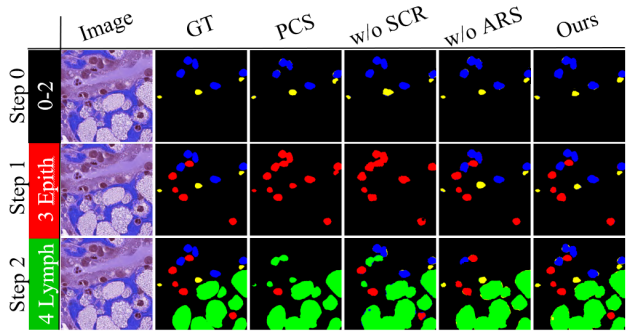

Figure 5: Visual ablation study demonstrating the distinct of Palimpsest's core components.

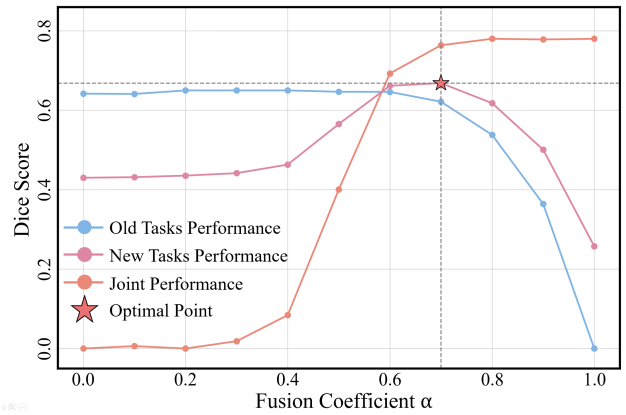

Figure 6: Analysis of the fusion coefficient α in PCS, illustrating the trade-off between Old Tasks and New Task.

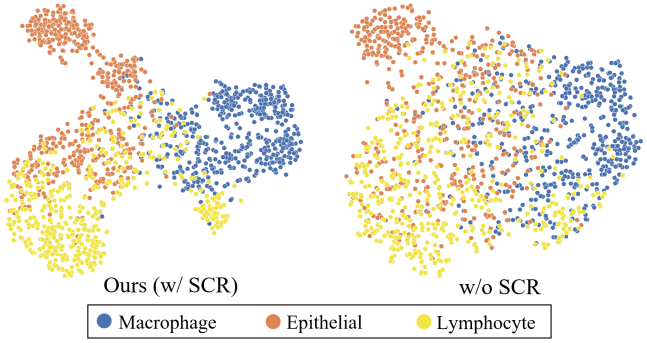

Figure 7: t-SNE visualization demonstrating that our SCR module significantly mitigates feature drift.

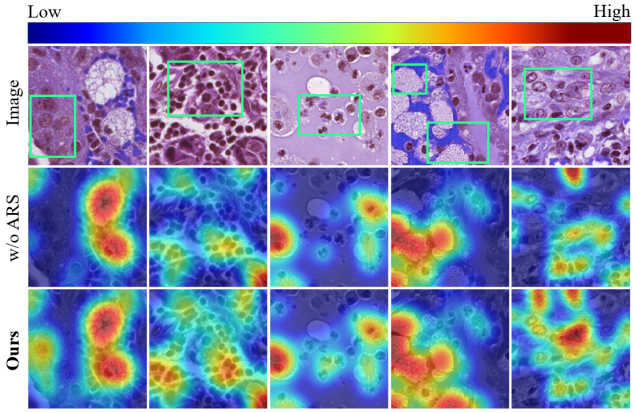

Figure 8: Comparison of activation maps from the model without ARS and with our ARS module.

BCE loss for background preservation causes a drastic performance collapse on old classes, confirming that our proposed logit-space shading is far more effective at preventing background corruption than naive distillation.

Analysis of Recalibration Strategies. To counteract the feature drift from PCS, a robust recalibration strategy is paramount. We compare our SCR with several alternatives in Table 5. Naive strategies like "No Calibration" or "Random Recalibration" result in poor old-task performance, confirming the necessity of a dedicated mechanism. A "Uniform Recalibration", which applies the same global drift vector to all past classes, offers a significant boost but is still suboptimal . This highlights the flawed assumption of uniform drift. In contrast, our SCR, which performs a differentiated recalibration based on class similarity, achieves the best performance, demonstrating the superiority of our non-uniform, similarity-aware approach.

Limitations and Discussions

Our *Palimpsest* framework's core novelty stems from its efficient, exemplar-free design, which hinges on two simplifying assumptions in the SCR module. First, it approximates the complex, class-specific feature drifts with a single global vector (Δ_t), an efficient yet potentially imprecise simplification when multiple new classes with diverse impacts are introduced. Second, the cosine similarity-based factor ($s_{c'}$) linearly maps feature similarity to drift magnitude, which may not capture all non-linear inter-class dynamics. Future work could address this by developing more sophisticated recalibration mechanisms.

Conclusion

In this paper, we presented *Palimpsest*, a novel framework that systematically resolves the CISS Trilemma in computational pathology by strategically decoupling the conflicting demands of incremental learning. Our solution comprises three synergistic yet independent mechanisms: a Parameter-Conserving Synthesis module to ensure constant-memory scalability; an Adaptive Residual Shading module to preserve complex background integrity via logit-space decoupling; and a Similarity-Aware Centroid Recalibration module to robustly counteract foreground feature drift. This entire framework operates under the strict, clinically-mandated exemplar-free constraint. Extensive experiments demonstrate that *Palimpsest* establishes a new state of the art, significantly outperforming existing methods, particularly in challenging long-term incremental scenarios.

Acknowledgments

This work was supported partly by National Natural Science Foundation of China (No. 62273241), Natural Science Foundation of Guangdong Province, China (No. 2024A1515011946), and the Shenzhen Research Foundation for Basic Research, China (No. JCYJ20250604181940054).

References

Asadi, N.; Davari, M.; Mudur, S.; Aljundi, R.; and Belilovsky, E. 2023. Prototype-sample relation distillation: towards replay-free continual learning. In *International conference on machine learning*, 1093–1106. PMLR.

Cermelli, F.; Mancini, M.; Bulo, S. R.; Ricci, E.; and Caputo, B. 2020. Modeling the background for incremental learning in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9233–9242.

Cha, S.; Yoo, Y.; Moon, T.; et al. 2021. Ssul: Semantic segmentation with unknown label for exemplar-based class-incremental learning. *Advances in neural information processing systems*, 34: 10919–10930.

Chen, J.; Cong, R.; Luo, Y.; Ip, H. H. S.; and Kwong, S. 2024. Strike a balance in continual panoptic segmentation. In *European Conference on Computer Vision*, 126–142. Springer.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Douillard, A.; Ramé, A.; Couairon, G.; and Cord, M. 2022. Dytox: Transformers for continual learning with dynamic token expansion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9285–9295.

Fukuda, T.; Kera, H.; and Kawamoto, K. 2025. Adapter merging with centroid prototype mapping for scalable class-incremental learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 4884–4893.

Gao, Q.; Luo, Z.; Klabjan, D.; and Zhang, F. 2022. Efficient architecture search for continual learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11): 8555–8565.

Graham, S.; Vu, Q. D.; Raza, S. E. A.; Azam, A.; Tsang, Y. W.; Kwak, J. T.; and Rajpoot, N. 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58: 101563.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hou, L.; Agarwal, A.; Samaras, D.; Kurc, T. M.; Gupta, R. R.; and Saltz, J. H. 2017. Unsupervised histopathology image synthesis. *arXiv preprint arXiv:1712.05021*.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W.; et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2): 3.

Lal, S.; Das, D.; Alabhya, K.; Kanfade, A.; Kumar, A.; and Kini, J. 2021. NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. *Computers in Biology and Medicine*, 128: 104075.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature*, 521(7553): 436–444.

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

Li, Z.; and Hoiem, D. 2017b. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 40(12): 2935–2947.

Lu, Y.; Zhang, S.; Cheng, D.; Xing, Y.; Wang, N.; Wang, P.; and Zhang, Y. 2024. Visual prompt tuning in null space for continual learning. *Advances in neural information processing systems*, 37: 7878–7901.

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

Mei, Y.; Yuan, L.; Han, D.-J.; Chan, K. S.; Brinton, C. G.; and Lan, T. 2024. Using Diffusion Models as Generative Replay in Continual Federated Learning–What will Happen? *arXiv preprint arXiv:2411.06618*.

Michieli, U.; and Zanuttigh, P. 2021. Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1114–1124.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Robbins, H.; and Monro, S. 1951. A stochastic approximation method. *The annals of mathematical statistics*, 400–407.

Sirinukunwattana, K.; Pluim, J. P.; Chen, H.; Qi, X.; Heng, P.-A.; Guo, Y. B.; Wang, L. Y.; Matuszewski, B. J.; Bruni, E.; Sanchez, U.; et al. 2017. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35: 489–502.

Upschulte, E.; Harmeling, S.; Amunts, K.; and Dickscheid, T. 2023. Uncertainty-aware contour proposal networks for cell segmentation in multi-modality high-resolution microscopy images. In *Competitions in Neural Information Processing Systems*, 1–12. PMLR.

Verma, R.; Kumar, N.; Patil, A.; Kurian, N. C.; Rane, S.; Graham, S.; Vu, Q. D.; Zwager, M.; Raza, S. E. A.; Rajpoot, N.; et al. 2021. MoNuSAC2020: A multi-organ nuclei segmentation and classification challenge. *IEEE Transactions on Medical Imaging*, 40(12): 3413–3423.

Wang, F.-Y.; Zhou, D.-W.; Ye, H.-J.; and Zhan, D.-C. 2022. Foster: Feature boosting and compression for class-incremental learning. In *European conference on computer vision*, 398–414. Springer.

Wang, H.; Lu, H.; Yao, L.; and Gong, D. 2025. Self-expansion of pre-trained models with mixture of adapters for continual learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 10087–10098.

Wang, H.; Wu, H.; and Qin, J. 2024. Incremental nuclei segmentation from histopathological images via future-class awareness and compatibility-inspired distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11408–11417.

Wei, Y.; Ye, J.; Huang, Z.; Zhang, J.; and Shan, H. 2023. Online prototype learning for online continual learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, 18764–18774.

Wu, H.; Wang, Z.; Song, Y.; Yang, L.; and Qin, J. 2022. Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11666–11675.

Xiao, J.-W.; Zhang, C.-B.; Feng, J.; Liu, X.; van de Weijer, J.; and Cheng, M.-M. 2023. Endpoints weight fusion for class incremental semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7204–7213.

Xing, F.; and Yang, L. 2016. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE reviews in biomedical engineering*, 9: 234–263.

Yang, G.; Fini, E.; Xu, D.; Rota, P.; Ding, M.; Nabi, M.; Alameda-Pineda, X.; and Ricci, E. 2022. Uncertainty-aware contrastive distillation for incremental semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2567–2581.

Ye, F.; and Bors, A. G. 2023. Self-evolved dynamic expansion model for task-free continual learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, 22102–22112.

Yuan, B.; and Zhao, D. 2024. A survey on continual semantic segmentation: Theory, challenge, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhang, C.-B.; Xiao, J.-W.; Liu, X.; Chen, Y.-C.; and Cheng, M.-M. 2022. Representation compensation networks for continual semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7053–7064.

Zhuang, H.; Chen, Y.; Fang, D.; He, R.; Tong, K.; Wei, H.; Zeng, Z.; and Chen, C. 2024. GACL: Exemplar-free generalized analytic continual learning. *Advances in Neural Information Processing Systems*, 37: 83024–83047.