PROMPT-FREE TASK-AGNOSTIC CONTINUAL LEARNING ACROSS HETEROGENEOUS BINARY SEGMENTATION TASKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Continual segmentation is commonly studied under changes in semantic classes or image domains while the underlying prediction objective remains relatively stable. We study a more heterogeneous continual setting in which sequential binary segmentation tasks may differ not only in image distribution but also in the semantic definition of foreground, while retaining the same binary prediction interface. We further consider prompt-free, task-agnostic inference: task boundaries are known during sequential training, but neither task identity nor interactive prompts are available at test time. Using the same shared-path continual learner, a domain-only stream retains 0.809 AA-FG-IoU with 0.118 forgetting, whereas a heterogeneous stream that changes both domain and foreground semantics drops to 0.532 AA-FG-IoU with 0.310 forgetting. Motivated by this degradation, we construct a continual segmentation framework with a frozen DINOv2-B/14 backbone, parameter-isolated segmentation experts, and an adapted distribution-replay selector that activates one expert for each unlabeled test image. On the COD→Polyp→USOD→TOD stream, Hard Top-1 inference reaches 0.806 AA-mIoU with 92.7% task-selection accuracy, compared with 0.814 AA-mIoU under Task-ID Oracle selection, leaving a routing gap of 0.009. These results support a decomposition of heterogeneous continual segmentation into two complementary requirements: preserving task-specific segmentation capabilities and reliably accessing the appropriate capability at inference.

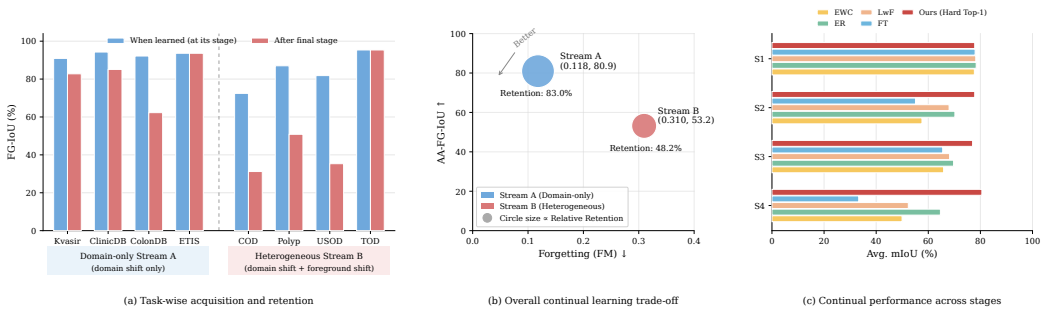
1 INTRODUCTION

Continual image segmentation aims to acquire new dense-prediction capabilities from sequentially arriving data while preserving previously learned knowledge. Existing continual segmentation settings have largely studied changes within a relatively stable prediction formulation, including evolving semantic classes, background-label shift, and image-domain variation Cermelli et al. (2020); Douillard et al. (2021); Gong et al. (2024); Yang et al. (2024). Recent work has broadened this scope toward foundation-model adaptation and more diverse segmentation task collections Lin et al. (2025); Wang et al. (2026b); Li et al. (2026). However, a continual stream may change not only how the input looks, but also what should be regarded as foreground. In this case, the model must sequentially acquire multiple segmentation functions that share the same binary output interface but correspond to different foreground semantics.

We use *heterogeneous continual binary segmentation* descriptively to denote this setting: both the input distribution and the semantic definition of foreground may change across stages, while every task remains a binary segmentation problem. We additionally consider prompt-free, task-agnostic inference. Task boundaries and the current task identity are available during sequential training, but an incoming test image is not accompanied by its task identity, point prompt, box prompt, or text prompt. Our primary stream follows COD→Polyp→USOD→TOD, spanning camouflaged objects, polyp lesions, underwater salient objects, and transparent objects.

A central question is whether continual behavior under ordinary domain variation remains representative when the foreground definition also changes. Figure 1 provides an empirical comparison using the same shared-path continual-learning framework under two four-stage streams. In the domain-only

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.



(a) Task-wise acquisition and retention (b) Overall continual learning trade-off (c) Continual performance across stages

Figure 1: Motivation for heterogeneous continual binary segmentation. (a) Task-wise acquisition and retention under domain-only and heterogeneous streams. (b) Final AA-FG-IoU versus forgetting (FM), with bubble size proportional to relative retention. (c) Stage-wise average mIoU on the heterogeneous COD→Polyp→USOD→TOD stream for shared-path baselines and our Hard Top-1 system.

Kvasir→ClinicDB→ColonDB→ETIS stream, the foreground remains "polyp" while the dataset domain changes. The learner retains 0.809 final AA-FG-IoU, with 0.118 forgetting and 0.830 relative retention. Under the heterogeneous COD→Polyp→USOD→TOD stream, where both domain and foreground semantics change, final AA-FG-IoU drops to 0.532, forgetting rises to 0.310, and relative retention decreases to 0.482. Figure 1(a) further shows that this difference is driven by substantially stronger degradation of previously learned tasks.

The comparison suggests that repeatedly adapting a single segmentation pathway becomes increasingly fragile when both appearance and foreground semantics change. Figure 1(c) further contrasts the stage-wise behavior of representative shared-path baselines with our isolated-expert system on the heterogeneous stream. Rather than forcing one parameter set to reconcile different foreground definitions, we preserve task-specific segmentation capabilities in separate parameter-isolated experts built on a frozen DINOv2-B/14 backbone Oquab et al. (2024). Once an expert is learned, its parameters and the shared backbone remain fixed, preventing subsequent stages from directly overwriting the pathway responsible for that segmentation function.

Parameter isolation, however, changes the deployment problem. If multiple segmentation capabilities are preserved independently, the model must determine which capability should process an unlabeled test image. We therefore adapt the distribution-replay parameter-selection paradigm of SOYO Wang et al. (2025) to the segmentation expert bank. Intermediate and deep DINOv2 representations form the routing representation, while compact historical feature distributions are used to generate pseudo-features for sequential selector updates without retaining historical raw images or masks. At inference, Hard Top-1 selection activates a single expert. This design separates two sources of continual performance: the segmentation capability preserved inside the expert bank and the loss introduced when that capability must be accessed without task identity.

Experiments support this decomposition. On the four-stage heterogeneous stream, Hard Top-1 inference reaches 0.806 AA-mIoU with 92.7% task-selection accuracy, while the same expert bank evaluated with ground-truth task identity reaches 0.814 AA-mIoU. The resulting 0.009 oracle-to-routed gap indicates that most of the available task-specific segmentation capability is recovered by automatic expert selection.

Contributions. Our contributions are threefold:

- We study prompt-free, task-agnostic continual learning across heterogeneous binary segmentation tasks, where both image distribution and foreground definition may change across stages, and empirically contrast this behavior with a domain-only stream.
- We construct a frozen-foundation-model framework with parameter-isolated segmentation experts and a distribution-replay selector, preserving task-specific dense-prediction pathways while enabling automatic expert access without task identity or interactive prompts.

- We evaluate capability preservation and task-agnostic access separately through routed performance, task-selection accuracy, and Task-ID Oracle diagnostics, making the remaining routing-induced degradation directly measurable.

2 Related Work

2.1 Continual Segmentation under Changing Labels, Domains, and Objectives

Continual semantic segmentation primarily studies sequential changes within a common semantic-segmentation formulation. Representative methods address evolving class spaces, background shift, and catastrophic forgetting, including MiB Cermelli et al. (2020), PLOP Douillard et al. (2021), and CoMasTRe Gong et al. (2024). Recent work broadens continual segmentation beyond class-incremental learning. CoSAM studies foundation-model adaptation across multiple image domains using domain-specific adapters Yang et al. (2024), while CUE considers continual universal segmentation across diverse segmentation tasks Lin et al. (2025). C^{2}GR further considers streams in which image appearance and segmentation objectives may change concurrently Li et al. (2026).

Our setting focuses on a complementary case in which all stages retain a binary prediction interface, but both the input distribution and the semantic definition of foreground may change. We further require task-agnostic, prompt-free inference after sequential training.

2.2 Modular Adaptation and Task-Agnostic Selection

Parameter-efficient continual adaptation often freezes pretrained representations and assigns task- or domain-specific modules to reduce direct interference. MoDA, SAMCL, and CA-SAM follow this modular principle and additionally require automatic module selection at inference Yang et al. (2024); Wang et al. (2026b;a). Task-agnostic retrieval has also been explored in continual classification through prompt selection, including L2P Wang et al. (2022b), DualPrompt Wang et al. (2022a), and CODA-Prompt Smith et al. (2023).

SOYO explicitly studies parameter selection for parameter-isolated continual learning using multi-level feature fusion, distribution compression, and historical feature replay Wang et al. (2025). Our selector follows this distribution-replay paradigm rather than treating these primitives as new. We adapt it to a DINOv2 segmentation expert bank and evaluate automatic selection jointly with the capability preserved by the expert bank through routed performance and Task-ID Oracle diagnostics.

3 Method

3.1 Problem Formulation

We consider a continual stream of T binary segmentation tasks $\{\mathcal{T}_t\}_{t=1}^{T}$. At stage t, the model receives a task-specific training set $\mathcal{D}_t = \{(\mathbf{x}_n^t, \mathbf{y}_n^t)\}_{n=1}^{N_t}$, where $\mathbf{x}_n^t \in \mathbb{R}^{3\times H\times W}$ is an input image and $\mathbf{y}_n^t \in \{0,1\}^{H\times W}$ is its binary foreground mask. Across stages, both the input distribution and the semantic definition of foreground may change, while the prediction interface remains binary.

Task boundaries and the current task identity are available during sequential training. Our method does not retain historical raw images or masks. After learning stage t, it maintains an expert bank $\mathcal{E}_t = \{E_1, \ldots, E_t\}$ in which previously learned experts remain fixed. At inference, an input may originate from any previously observed task, but its task identity is unknown and no point, box, or text prompt is provided. A selector therefore predicts an expert index $\hat{t}$ and the selected expert produces the binary mask. The problem thus has two complementary requirements: preserving task-specific segmentation capability across stages and accessing the appropriate capability without task identity at test time.

3.2 Overall Framework

Figure 2 summarizes the framework. At stage t, a frozen DINOv2-B/14 backbone supplies multi-level representations to a new task-specific segmentation expert E_t. Only the current expert is optimized;

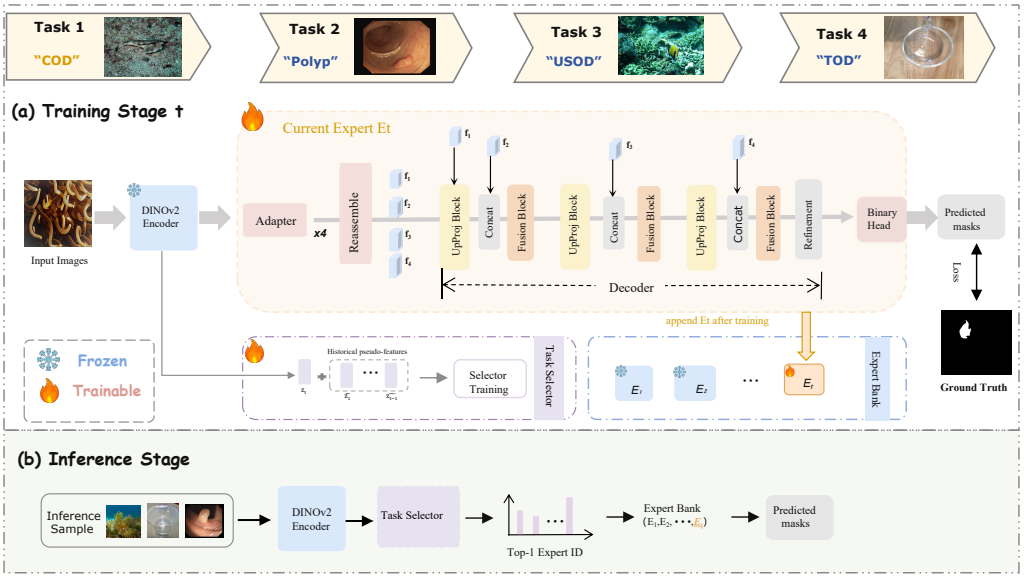


Figure 2: Overall framework. (a) At stage t, a new segmentation expert is learned while the frozen backbone and previous experts remain fixed; the selector is subsequently updated using current routing features and historical pseudo-features. (b) Prompt-free Hard Top-1 inference. Routing and segmentation reuse the same frozen DINOv2 weights but are executed as separate forward passes in the current implementation.

the shared backbone and all previously learned experts remain frozen. Once expert training is complete, E_t is frozen and added to the expert bank.

The task selector is then updated using real routing features from the current task together with pseudo-features sampled from compressed routing distributions of previous tasks. After the selector update, the current-task routing features are compressed and added to the historical routing memory. This memory is used only to maintain the selector and is separate from the segmentation expert bank.

At inference, no task identity or interactive prompt is available. A first frozen-DINOv2 forward extracts routing features and predicts a Hard Top-1 expert index. A second forward through the same frozen DINOv2 weights provides the multi-level patch features required by the selected expert, which then produces the binary mask. Thus, the current implementation uses separate routing and segmentation forwards, while executing only one task-specific expert. No historical replay or multi-expert ensemble is used at test time.

3.3 Parameter-Isolated Segmentation Experts

Frozen multi-level representations. We use a pretrained DINOv2-B/14 Oquab et al. (2024) as the shared visual backbone and freeze it throughout the continual stream. For an input image, we extract patch-token representations from hidden-state indices 2, 5, 8, and 11 after removing the CLS and optional register tokens. Let $\mathbf{X}_\ell \in \mathbb{R}^{N\times 768}$ denote the patch-token representation at level $\ell \in \{2, 5, 8, 11\}$.

For each incoming task $\mathcal{T}_t$, we instantiate one residual bottleneck adapter per selected feature level. The adapted representation is

$$\widetilde{\mathbf{X}}_{t,\ell} = \mathbf{X}_\ell + \mathcal{U}_{t,\ell}\left(\mathrm{GELU}\left(\mathcal{D}_{t,\ell}(\mathbf{X}_\ell)\right)\right), \tag{1}$$

where $\mathcal{D}_{t,\ell} : \mathbb{R}^{768} \to \mathbb{R}^{128}$ and $\mathcal{U}_{t,\ell} : \mathbb{R}^{128} \to \mathbb{R}^{768}$ are task-specific down- and up-projections. The residual connection retains the frozen representation while allowing lightweight task-specific adaptation.

Feature reassembly and decoding. The four adapted token sequences are reshaped into spatial feature maps and reassembled into a four-level pyramid at approximately $1/4$, $1/8$, $1/16$, and $1/32$ of the input resolution. A task-specific coarse-to-fine decoder progressively upsamples deeper features, fuses them with the corresponding shallower features, and produces a $1/4$-resolution representation. A residual refinement block with squeeze-and-excitation recalibration Hu et al. (2018) is applied before the binary prediction head. Exact channel widths and layer configurations are provided in the appendix.

Binary prediction and expert isolation. The binary head upsamples the refined feature map to the input resolution and produces mask logits $\mathbf{s}_t$. The current expert is optimized with binary cross-entropy and soft Dice loss,

$$\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{BCE}}(\mathbf{s}_t, \mathbf{y}) + 1 - \frac{2\sum_p \sigma(\mathbf{s}_{t,p})\mathbf{y}_p + \epsilon}{\sum_p \sigma(\mathbf{s}_{t,p}) + \sum_p \mathbf{y}_p + \epsilon}, \tag{2}$$

where $\sigma(\cdot)$ is the sigmoid function, p indexes mask elements, and ϵ is a numerical-stability constant. During stage t, only the newly instantiated expert is updated; the backbone and all earlier experts remain frozen. Consequently, later stages do not directly overwrite the parameters responsible for previously acquired segmentation functions.

3.4 Task-Agnostic Distribution-Replay Selector

Parameter isolation prevents direct interference between segmentation experts but creates a capability-access problem at inference. We instantiate the distribution-replay parameter-selection paradigm of SOYO Wang et al. (2025) on the DINOv2 expert bank. Specifically, we follow its multi-level feature-fusion and historical distribution-replay principles, while using DINOv2 routing representations and a growing task classifier for the segmentation setting.

Multi-level routing representation. Figure 3 shows the selector. For an input image $\mathbf{x}$, we extract the CLS representations from DINOv2 hidden-state index 5 and the final hidden state, $\mathbf{c}^{\text{mid}}, \mathbf{c}^{\text{final}} \in \mathbb{R}^{768}$. Their concatenation is ℓ_2-normalized to form the 1536-D routing vector

$$\mathbf{z} = \text{Norm}\left([\mathbf{c}^{\text{mid}}; \mathbf{c}^{\text{final}}]\right). \tag{3}$$

Let $\mathbf{z}^{\text{mid}}, \mathbf{z}^{\text{final}} \in \mathbb{R}^{768}$ denote the corresponding two halves. Following the MDFN fusion used in SOYO Wang et al. (2025), we compute

$$\mathbf{h} = \mathbf{z}^{\text{final}} + g_1(\mathbf{z}^{\text{mid}}) + g_2(\mathbf{z}^{\text{final}}), \tag{4}$$

where g_1 and g_2 are lightweight two-layer MLPs.

The fused feature and each task-classification weight are normalized before scaled cosine classification:

$$\ell_j = s\,\bar{\mathbf{h}}^\top \bar{\mathbf{w}}_j, \qquad p_j = \frac{\exp(\ell_j)}{\sum_{r=1}^{t}\exp(\ell_r)}, \quad j = 1, \ldots, t, \tag{5}$$

where $\bar{\mathbf{h}} = \mathbf{h}/\|\mathbf{h}\|_2$, $\bar{\mathbf{w}}_j = \mathbf{w}_j/\|\mathbf{w}_j\|_2$, and s is a fixed scale. The classifier expands only when a new task is observed.

Compressed distribution replay. Training the selector only on the current task would bias the classifier toward recently observed tasks. Following SOYO's feature-compression and resampling principle Wang et al. (2025), we therefore compress historical routing vectors rather than retaining historical images, masks, or the full routing-feature set.

For each completed task $\mathcal{T}_i$, PCA projects its normalized 1536-D routing vectors to a 128-D latent space, where a diagonal Dirichlet-process Bayesian Gaussian mixture model (DP-BGMM) is fitted. Let

$$q_i(\mathbf{u}) = \sum_{k=1}^{K_i} \pi_{i,k}\mathcal{N}\left(\mathbf{u}; \boldsymbol{\mu}_{i,k}, \text{diag}(\boldsymbol{\nu}_{i,k})\right) \tag{6}$$

denote the compressed distribution for task i. Historical pseudo-features are generated by

$$\mathbf{u} \sim q_i(\mathbf{u}), \qquad \mathbf{r}'_i = \text{Norm}\left(\text{PCA}_i^{-1}(\mathbf{u})\right). \tag{7}$$

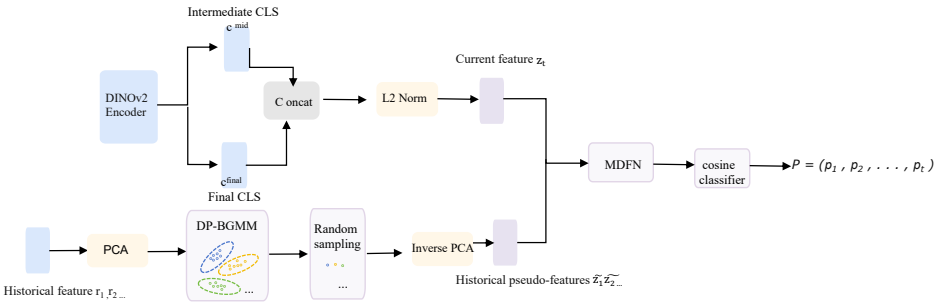


Figure 3: Distribution-replay task selector. Online routing uses intermediate and final DINOv2 CLS features, multi-level fusion, and a cosine classifier. PCA and DP-BGMM are used only to compress historical routing features and generate pseudo-features for later selector updates; they are not part of the online routing path.

The replayed vector $\mathbf{r}'_i$ is then processed by the same fusion and cosine-classification pathway as a real routing vector. Only PCA parameters and mixture statistics are retained for previous tasks; detailed mixture hyperparameters are given in the experimental setup and appendix.

Sequential selector update and inference. After training the current segmentation expert, the selector is updated with real routing vectors from the current task and balanced pseudo-features sampled for previous tasks. The fusion modules and cosine classifier are optimized with cross-entropy over the observed task identities. The current-task routing vectors are compressed only after this update and then added to the historical routing memory.

At inference, PCA, DP-BGMM fitting, and pseudo-feature replay are not used. The selector computes $\hat{t} = \arg\max_j p_j$, and only $E_{\hat{t}}$ is executed for mask prediction. The compressed historical distributions therefore support continual selector learning but do not participate in online expert selection.

4 Experiments

4.1 Experimental Setup

Benchmarks and continual protocol. Our primary benchmark is the heterogeneous four-stage stream COD→Polyp→USOD→TOD, where both image appearance and foreground semantics change across stages. COD contains CAMO Le et al. (2019), COD10K Fan et al. (2020), and NC4K Lv et al. (2021); Polyp is constructed from Kvasir-SEG Jha et al. (2020), CVC-ClinicDB, CVC-300/EndoScene Vazquez et al. (2017), and ETIS-LaribPolypDB Silva et al. (2014); USOD contains MAS3K Li et al. (2020), SUIM Islam et al. (2020a), and the SVAM-associated benchmark Islam et al. (2020b); and TOD uses Trans10K-v2 Xie et al. (2021). All annotations are converted to binary foreground masks.

For the motivation study in Fig. 1, we additionally compare two four-stage stream structures using the same representative shared-path continual learner and the same evaluation definitions. The domain-only stream, Kvasir→ClinicDB→ColonDB→ETIS, keeps the foreground definition fixed as polyp while changing the dataset domain. The heterogeneous stream is the primary COD→Polyp→USOD→TOD sequence and changes both image distribution and foreground semantics. We report task-wise foreground IoU when each task is learned and after the final stage, together with final average performance, forgetting, and relative retention.

Task boundaries and the current task identity are available during training. At inference, task identity is unavailable and no point, box, or text prompt is provided. Unless otherwise stated, our method retains no historical raw images or masks; the selector retains only compressed routing-feature distributions. We compare against sequential fine-tuning, LwF Li & Hoiem (2016), EWC Kirkpatrick et al. (2017), and Experience Replay, together with offline Joint Training and a Task-ID Oracle diagnostic. The generic continual baselines use the same frozen backbone and a single segmentation pathway corresponding to one expert in our framework.

Table 1: Continual learning on COD→Polyp→USOD→TOD. Joint is offline; Oracle is diagnostic.

Method	AA-mIoU ↑	AA-mF1 ↑	AA-mMAE ↓	FM-mIoU ↓	Raw Replay
Sequential FT	0.333	0.377	0.125	0.481	No
LwF Li & Hoiem (2016)	0.523	0.602	0.120	0.279	No
EWC Kirkpatrick et al. (2017)	0.499	0.576	0.131	0.307	No
ER	0.647	0.717	0.051	0.160	40 pairs
Ours (Hard Top-1)	0.806	0.861	0.031	0.008	No[†]
Joint Training[*]	0.805	0.857	0.026	–	–
Task-ID Oracle[‡]	0.814	0.870	0.029	–	–

[†]No raw replay; compressed routing statistics only. [*]Offline pooled training. [‡]Ground-truth task ID, diagnostic only.

Implementation details. All methods use DINOv2-B/14 Oquab et al. (2024) as the frozen shared backbone. Images are resized to 364×364. Segmentation experts are trained for at most 50 epochs with SGD and selected by validation mIoU. The selector is optimized with AdamW for 20 epochs and uses the 1536-D intermediate/deep DINOv2 routing representation, PCA compression to 128 dimensions, and a diagonal DP-BGMM with at most 10 active components for historical feature replay. Additional preprocessing and optimization details are provided in the appendix.

Evaluation metrics. For binary segmentation, we report mIoU, mF1, and mMAE. Let $a_{i,j}$ denote the mIoU on task $\mathcal{T}_j$ after training through stage i. Following the continual-segmentation protocol used in SAMCL Wang et al. (2026b), we compute

$$\mathrm{AA} = \frac{1}{T}\sum_{j=1}^{T} a_{T,j}, \qquad \mathrm{FM} = \frac{1}{T}\sum_{j=1}^{T} \left(a_{j,j} - a_{T,j}\right).$$

AA summarizes final performance across observed tasks, while FM measures degradation from the performance obtained when each task was learned to its final-stage performance. For the motivation comparison, we additionally report relative retention over previously learned tasks,

$$\mathrm{RR} = \frac{1}{T-1}\sum_{j=1}^{T-1} \frac{a_{T,j}}{a_{j,j}}.$$

For task-agnostic expert selection, we report task-selection Top-1 accuracy (ST-Top1), Task-ID Oracle AA-mIoU, and the oracle-to-routed gap Δ_{route}.

4.2 Main Results

Domain-only versus heterogeneous continual learning. Using the same continual learner, Fig. 1 shows that the domain-only Kvasir→ClinicDB→ColonDB→ETIS stream attains 0.809 AA-FG-IoU, 0.118 forgetting, and 0.830 relative retention, whereas the heterogeneous COD→Polyp→USOD→TOD stream yields 0.532, 0.310, and 0.482, respectively. The larger old-task degradation in Fig. 1(a) motivates studying domain and foreground changes jointly.

Comparison on the heterogeneous stream. Table 1 shows AA-mIoU values of 0.333, 0.523, and 0.499 for FT, LwF, and EWC, and 0.647 for ER with 40 replay pairs. Our Hard Top-1 system reaches 0.806 AA-mIoU, 0.861 AA-mF1, and 0.031 AA-mMAE without raw replay, essentially matching the 0.805 offline Joint Training reference. Task-ID Oracle reaches 0.814 AA-mIoU, leaving only a 0.009 routing gap, which separates strong preserved expert capability from the comparatively small loss due to automatic expert access.

4.3 Task-Agnostic Expert Selection

All selectors are evaluated on the same completed expert bank without retraining segmentation experts. We compare NMC, cosine KNN, a linear classifier, key matching, a SAMCL-style MLP, and

Table 2: Task-agnostic expert selection on the frozen expert bank; storage excludes backbone and segmentation experts.

Selector	ST-Top1 ↑	AA-mIoU ↑	Δ_{route} ↓	Storage (MB) ↓
NMC	0.885	0.794	0.020	**0.023**
Cosine KNN	0.905	0.798	0.016	3.525
Linear	0.910	0.801	0.013	3.548
Key-based	0.907	0.800	0.014	5.167
SAMCL-style MLP Wang et al. (2026b)	0.907	0.800	0.014	7.167
Distribution Replay (Ours)	**0.927**	**0.806**	**0.009**	6.048
Task-ID Oracle*	1.000	0.814	0.000	–

*Diagnostic only. "SAMCL-style MLP" is a selector-only adaptation under the same routing representation.

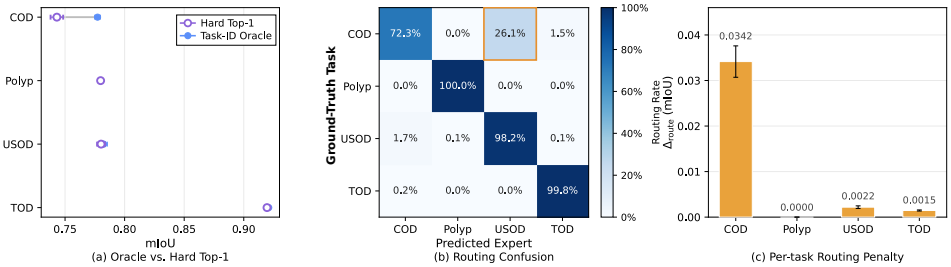


Figure 4: Task-agnostic routing diagnostics: (a) Oracle vs. Hard Top-1 mIoU, (b) normalized routing confusion, and (c) per-task routing penalty.

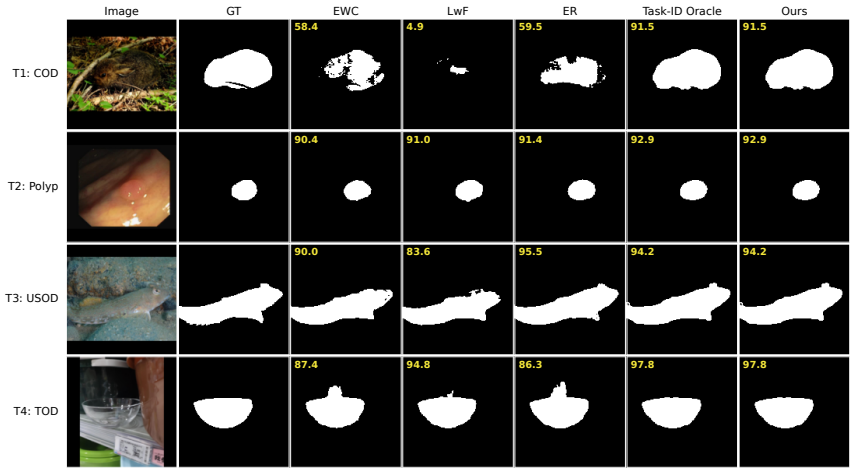


Figure 5: Qualitative segmentation after four stages. Shared-path baselines degrade unevenly; Oracle and Hard Top-1 recover stronger masks. Yellow values denote per-image IoU (%).

Distribution Replay; replay-based baselines use routing embeddings rather than raw images/masks, while ours stores compressed PCA/DP-BGMM statistics.

Table 2 shows that Distribution Replay reaches 0.927 ST-Top1 and 0.806 AA-mIoU with a 0.009 routing gap, versus 0.910 and 0.801 for the strongest AA-mIoU baseline (Linear). Its 6.048 MB selector footprint is larger than NMC or Linear, so the gain represents a routing-quality/storage trade-off rather than uniform efficiency dominance.

Figure 4 shows that residual routing error is concentrated in COD: 72.3% are routed correctly and 26.1% to USOD, whereas Polyp, USOD, and TOD achieve 100.0%, 98.2%, and 99.8% diagonal routing. The corresponding penalties are 0.0342, 0.0000, 0.0022, and 0.0015 mIoU.

Routing cases. Figure 6 connects the routing errors in Fig. 4 to the quality of the selected expert's mask. For the COD example routed to USOD, IoU falls from 79.7% under Task-ID Oracle selection to 44.9% under Hard Top-1. For the USOD example routed to COD, it falls from 95.2% to 23.2%. By contrast, correctly routed Polyp and TOD examples reproduce their Oracle masks at 92.9% and 97.8% IoU. These examples show the consequence of incorrect capability access, while the task-wise penalties in Fig. 4(c) summarize its effect across the respective evaluation sets. These fixed cases are diagnostic examples, not aggregate estimates: a large per-image loss can coexist with a small task-wide penalty when the corresponding misrouting is rare. Together with the near-Oracle result in Table 1, they distinguish errors in accessing a preserved expert from the expert's segmentation capability itself.

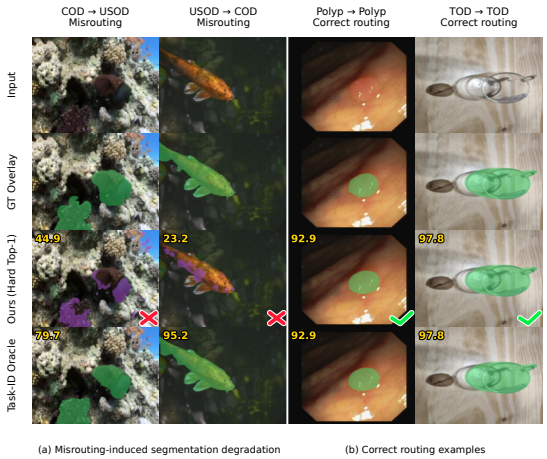


Figure 6: Verified routing cases. Wrong expert selection degrades the COD and USOD masks; correctly routed Polyp and TOD match the Oracle. Yellow values denote per-image IoU (%).

4.4 Ablation Studies

Routing representation and historical replay. Table 3 shows that Deep CLS (0.923 ST-Top1, 0.805 AA-mIoU) is more informative than Mid CLS (0.885, 0.791), while Mid+Deep slightly reduces the routing gap to 0.009 without improving ST-Top1. Retaining real historical embeddings gives the diagnostic upper reference of 0.943 ST-Top1, 0.811 AA-mIoU, and a 0.003 gap. Replacing that history with compressed distribution replay yields 0.927, 0.806, and 0.009, an approximately 0.005 AA-mIoU cost for compressed replay.

Table 3: Selector ablation on frozen experts.

Variant	ST-Top1 ↑	AA-mIoU ↑	Δ_{route} ↓
Deep CLS only	0.923	0.805	0.010
Mid CLS only	0.885	0.791	0.023
Mid + Deep CLS	0.922	0.806	0.009
Fusion + Real History	0.943	0.811	0.003
Full Distribution Replay	0.927	0.806	0.009
Task-ID Oracle	1.000	0.814	0.000

Table 4: Adapter ablation (AA-mIoU; M params./task).

Variant	Oracle	Hard	Params.
w/o Adapter	0.811	0.803	**11.370**
Full Expert	**0.814**	**0.806**	12.160

Residual adapters add 0.790M parameters per task (6.9%) for +0.003 Oracle and routed AA-mIoU.

5 Conclusion

We study prompt-free, task-agnostic continual learning across heterogeneous binary segmentation tasks. The heterogeneous stream degrades more strongly than the domain-only stream under the same shared-path learner. Parameter-isolated experts preserve task capability, while distribution-replay routing enables task-agnostic access: Hard Top-1 reaches 0.806 AA-mIoU versus 0.814 for Task-ID Oracle, a 0.009 gap. Future work may explore scalable expert sharing and routing.

AI USE STATEMENT

In this study, generative AI tools assisted with language polishing, while all experiments were conducted personally by the authors. The authors reviewed the AI-assisted content and take responsibility for the final version of the manuscript.

REFERENCES

Fabio Cermelli, Massimiliano Mancini, Samuel Rota Bulo, Elisa Ricci, and Barbara Caputo. Modeling the background for incremental learning in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9233–9242, 2020.

Arthur Douillard, Yifu Chen, Arnaud Dapogny, and Matthieu Cord. Plop: Learning without forgetting for continual semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4040–4050, 2021.

Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2777–2787, 2020.

Yizheng Gong, Siyue Yu, Xiaoyang Wang, and Jimin Xiao. Continual segmentation with disentangled objectness learning and class recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3848–3857, 2024.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, 2018.

Md Jahidul Islam, Chelsey Edge, Yuyang Xiao, Peigen Luo, Muntaqim Mehtaz, Christopher Morse, Sadman Sakib Enan, and Junaed Sattar. Semantic segmentation of underwater imagery: Dataset and benchmark. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1769–1776, 2020a. doi: 10.1109/IROS45743.2020.9340821.

Md Jahidul Islam, Ruobing Wang, and Junaed Sattar. Svam: Saliency-guided visual attention modeling by autonomous underwater robots. *arXiv preprint arXiv:2011.06252*, 2020b.

Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D. Johansen. Kvasir-seg: A segmented polyp dataset. In *MultiMedia Modeling*, pp. 451–462, 2020.

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

T.-N. Le, T. V. Nguyen, Z. Nie, M.-T. Tran, and A. Sugimoto. Anabranch network for camouflaged object segmentation. *Computer Vision and Image Understanding*, 184:45–56, 2019.

L. Li, E. Rigall, J. Dong, and G. Chen. Mas3k: An open dataset for marine animal segmentation. In *Benchmarking, Measuring, and Optimizing*, pp. 194–212, 2020. doi: 10.1007/978-3-030-71058-3_12.

W. Li, J. Zhang, G. Wang, J. Ning, Y. Chen, G. Yang, and L. Gu. C2gr: Coupled comprehensive generative replay for a continually learnable universal segmentation model. *arXiv preprint arXiv:2606.23473*, 2026.

Zhizhong Li and Derek Hoiem. Learning without forgetting. In *Computer Vision – ECCV 2016*, pp. 614–629, 2016.

Z. Lin, Z. Wang, and X. Wang. Towards continual universal segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 29417–29427, 2025.

Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan. Simultaneously localize, segment and rank the camouflaged objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11591–11601, 2021.

Maxime Oquab, Timothee Darcet, Theo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Juan Silva, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*, 9(2):283–293, 2014. doi: 10.1007/s11548-013-0926-3.

James Seale Smith, Leonid Karlinsky, Vyshnavi Gutta, Paola Cascante-Bonilla, Donghyun Kim, Assaf Arbelle, Rameswar Panda, Rogerio Feris, and Zsolt Kira. Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11909–11919, 2023.

David Vazquez, Jorge Bernal, F. Javier Sanchez, Gloria Fernandez-Esparrach, Antonio M. Lopez, Adriana Romero, Michal Drozdzal, and Aaron Courville. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering*, 2017:4037190, 2017.

J. Wang, W. Dai, H. Wang, S. Yang, H. Bi, and J. Sun. Continual alignment for sam: Rethinking foundation models for medical image segmentation in continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Findings*, pp. 7520–7529, 2026a.

Q. Wang, X. Song, Y. He, J. Han, C. Ding, X. Gao, and Y. Gong. Boosting domain incremental learning: Selecting the optimal parameters is all you need. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4839–4849, 2025.

Z. Wang, K. Ji, D. Wang, H. Zhang, and F. Cheng. Samcl: Empowering sam to continually learn from dynamic domains with extreme storage efficiency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26741–26750, 2026b. doi: 10.1609/aaai.v40i31.39884.

Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer G. Dy, and Tomas Pfister. Dualprompt: Complementary prompting for rehearsal-free continual learning. In *Computer Vision – ECCV 2022*, pp. 631–648, 2022a.

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 139–149, 2022b.

Enze Xie, Wenjia Wang, Wenhai Wang, Peize Sun, Hang Xu, Ding Liang, and Ping Luo. Segmenting transparent objects in the wild with transformer. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1194–1200, 2021.

J. Yang, Y. Wu, J. Cen, W. Huang, H. Wang, and J. Zhang. Continual learning for segment anything model adaptation. *arXiv preprint arXiv:2412.06418*, 2024.