STRUCTURED SEMANTIC CONSISTENCY FOR UNSUPERVISED CROSS-DOMAIN NUCLEI SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Domain shifts in staining, tissue composition, and annotation conventions limit the transferability of nuclei instance segmentation models across histopathology datasets. Prediction consistency alone cannot ensure correct cell-type assignments, making semantic alignment a central challenge in unsupervised domain adaptation. To address this problem, we propose a structured semantic consistency framework that transfers a multi-head nuclei segmentation network from a labeled source domain to an unlabeled target domain. The framework combines source foreground preservation with masked target-view consistency, aligns reliable target features with supervised source-class prototypes, and regularizes inter-class relationships using fixed text embeddings. A class-support constraint is applied only when source and target annotation spaces differ. Training follows a two-stage strategy consisting of joint source–target pre-adaptation and subsequent supervised source refinement. Experiments cover internal domain transfers within PanNuke and cross-dataset adaptation from PanNuke to CoNSeP and Lizard. On CoNSeP and Lizard, class-aware panoptic quality rises from 0.1059 to 0.2335 and from 0.0733 to 0.1127. These results demonstrate the effectiveness of combining prediction consistency, prototype alignment, and semantic relational constraints for cross-domain nuclei instance segmentation without target-domain annotations.

1 INTRODUCTION

Nuclei instance segmentation provides a cellular description of histopathology by identifying individual nuclei and assigning their types. Accurate predictions require both instance separation and semantic discrimination. A model trained on one collection encounters changes in staining, tissue composition, and annotation conventions when applied to another. Unsupervised domain adaptation (UDA) addresses this setting using labeled source images and unlabeled target images, enabling the transfer of source annotations to a new collection.

Prediction consistency provides a useful learning signal for target images: a nucleus should retain its identity when its appearance or surrounding context changes. Semantic alignment complements this signal by specifying which class a stable prediction represents. In nuclei segmentation, these two requirements operate at different resolutions. Foreground agreement preserves nuclear occupancy, instance geometry separates adjacent cells, and type discrimination organizes cells into meaningful categories. Cross-dataset transfer further requires an explicit correspondence between the source label space and the target annotation convention.

We propose structured semantic consistency, a framework that connects target-view agreement to supervised source-class references. Source foreground preservation maintains nuclear features under appearance perturbations. Masked target-view consistency transfers predictions from a complete image to an occluded view. Source prototypes anchor reliable target features to labeled classes, while fixed text embeddings organize relationships among class centers. A dataset-specific support constraint restricts pseudo-labels to the predefined target annotation space. The framework applies these constraints during joint source–target pre-adaptation and then refines the segmentation network with source supervision.

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.

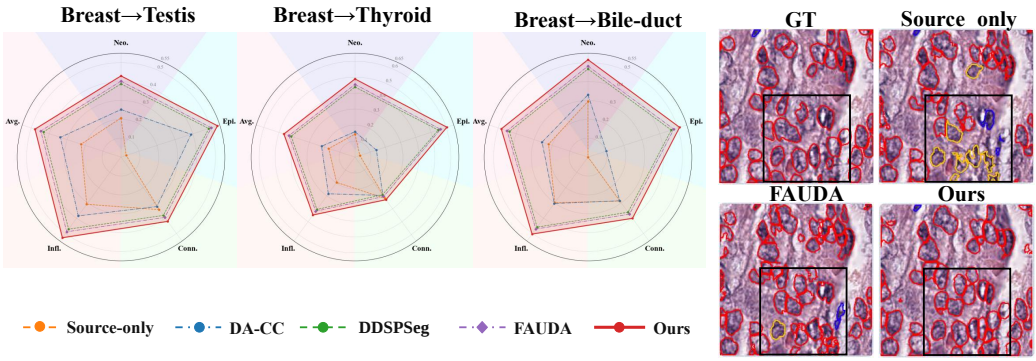


Figure 1: Cross-organ adaptation overview. Left: radar plots of classwise and average PQ for Breast-to-Testis, Breast-to-Thyroid, and Breast-to-Bile-duct transfer, comparing Source-only, DA-CC, DDSPSeg, FAUDA, and our method. Right: GT, Source-only, FAUDA, and our predictions on a representative example. The red curves denote our method; Tables 1–3 provide the numerical comparisons.

Figure 1 gives an overview of the reported cross-organ results. Our contributions are threefold: (i) we formulate an adaptation objective that couples foreground preservation and masked prediction consistency with binary and type prototypes; (ii) we integrate fixed text-derived class relations and conditionally apply target class support under label-space mismatch; and (iii) we evaluate the framework on internal PanNuke transfers and external adaptation to CoNSeP and Lizard. The reported class-aware PQ increases from 0.1059 to 0.2335 on CoNSeP and from 0.0733 to 0.1127 on Lizard. The method uses a shared multi-head segmentation network and retains its inference procedure.

2 Related Work

Nuclei segmentation and domain adaptation. Nuclei instance segmentation has progressed from shape-aware models such as StarDist (Schmidt et al., 2018) and generalist cellular segmentation such as Cellpose (Stringer et al., 2021) to histology-specific multi-task and transformer systems including HoVer-Net, CellViT, and NuLite (Graham et al., 2019; Hörst et al., 2024; Tommasino et al., 2026); PanNuke and Lizard provide diverse nuclei benchmarks (Gamper et al., 2019; Graham et al., 2021). In semantic UDA, output-space and entropy alignment (Tsai et al., 2018; Vu et al., 2019), cross-domain mixing and prototype-based self-training (Tranheden et al., 2021; Zhang et al., 2021), and architecture-aware adaptation (Hoyer et al., 2022a;b) establish complementary routes to target transfer. For cellular recognition, category-aware alignment and contextual correspondence directly address cross-domain nuclei analysis (Li et al., 2022; Fan et al., 2024).

Consistency and semantic preservation. Medical UDA has combined image- and feature-level alignment (Chen et al., 2019), self-ensembling (Perone et al., 2019), and disentangled cross-modality representations (Yang et al., 2019). More generally, Mean Teacher and FixMatch show how stable targets, perturbation consistency, confidence filtering, and pseudo-labeling can exploit unlabeled data (Tarvainen & Valpola, 2017; Sohn et al., 2020). DDSP preserves semantics under distribution disruption (Zheng et al., 2024), MIC transfers complete-view predictions to masked target views (Hoyer et al., 2023), and selective entropy constraints provide additional target supervision (Feng et al., 2023). Our formulation couples these principles to nuclear occupancy and type prediction while using source prototypes as class-specific references.

Prototype and text-derived semantic references. Prototype-guided adaptation can denoise pseudo-labels and organize target structure (Zhang et al., 2021), while CLIP supplies transferable visual–language representations (Radford et al., 2021). DenseCLIP extends language guidance to dense prediction (Rao et al., 2022), and CoPT uses domain-agnostic text embeddings for adaptive segmentation (Mata et al., 2024). We match relations among source visual class centers to relations

among fixed class-name embeddings, complementing pixel-to-prototype alignment; the target support operator is activated only when the target annotation space is a strict subset of the unified label space.

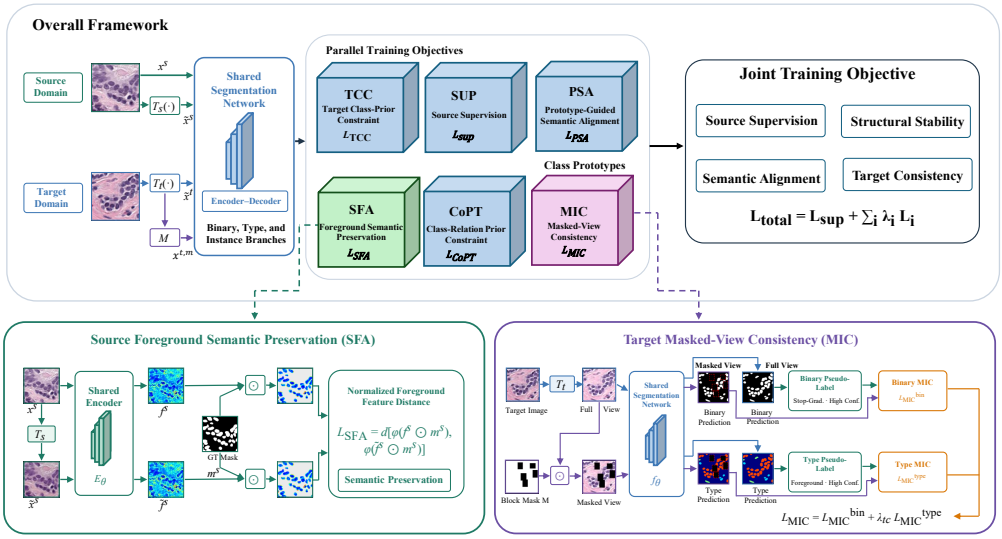


Figure 2: Pipeline of structured semantic consistency. Source and target images share a segmentation network. Source foreground agreement (SFA), masked image consistency (MIC), prototype semantic alignment (PSA), and text-derived class relations (CoPT) complement source supervision; target class support (TCC) is conditional on label-space mismatch. The lower panels expand SFA and MIC.

3 Structured Semantic Consistency

3.1 Learning setting and segmentation network

Let $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and $\mathcal{D}_t = \{x_j^t\}_{j=1}^{N_t}$ denote labeled source and unlabeled target datasets. Source labels specify instances and nuclear types. The shared network f_θ follows the NuLite family with a FastViT-based encoder (Vasu et al., 2023; Tommasino et al., 2026). It produces features $h_\theta(x)$, binary logits z^b, type logits z^c, and horizontal/vertical maps for instance separation. For branch $r \in \{b, c\}$, $p^r = \mathrm{softmax}(z^r)$ supplies probabilities for pseudo-label selection, while cross-entropy consumes logits. Inference retains the inherited reconstruction procedure, combining foreground and geometric predictions before assigning an instance type.

3.2 Source supervision and foreground preservation

Appearance transformations produce $\widetilde{x}^s = T_s(x^s)$ and $\widetilde{x}^t = T_t(x^t)$ while preserving nuclear geometry. Binary segmentation, type classification, and horizontal/vertical regression define $\mathcal{L}_{\mathrm{sup}}$. Both source views are supervised through $\mathcal{L}_{\mathrm{src}} = \mathcal{L}_{\mathrm{sup}}(x^s, y^s) + \mathcal{L}_{\mathrm{sup}}(\widetilde{x}^s, y^s)$. With source foreground mask m^s resized to feature resolution and normalization $\nu(v) = v/(\|v\|_2 + \epsilon)$, source foreground agreement is

$$\mathcal{L}_{\mathrm{sfa}} = \frac{\sum_u m_u^s \|\nu(h_\theta(x^s)_u) - \nu(h_\theta(\widetilde{x}^s)_u)\|_2^2}{\sum_u m_u^s + \epsilon}. \tag{1}$$

This constraint preserves annotated nuclear features under appearance changes and complements the target-domain objectives in Figure 2. Source labels provide both the supervised task targets and the spatial mask for feature agreement. Averaging over foreground positions concentrates the penalty on nuclei, while feature normalization emphasizes representational direction. This source reference remains available throughout joint adaptation, independently of target pseudo-label confidence. The supervised losses retain the association between each nuclear region, its type, and its

horizontal/vertical instance geometry. Feature agreement adds an appearance-invariance constraint at the same annotated locations, connecting the source representation to the segmentation task during adaptation.

3.3 Masked target-view consistency

A block mask M produces $x^{t,m} = \widetilde{x}^t \odot M$. The complete view supplies detached references $q^b = \mathrm{sg}(p^b(\widetilde{x}^t))$ and $q^c = \mathrm{sg}(\Pi_{\mathcal{V}_t}(p^c(\widetilde{x}^t)))$, where $\Pi_{\mathcal{V}_t}$ is the support operator below. Both views share network parameters. Define $\widehat{y}^r_u = \arg\max_k q^r_u(k)$, $w^b_u = \mathbf{1}[\max_k q^b_u(k) > \tau_b]$, and $w^c_u = \mathbf{1}[\max_k q^c_u(k) > \tau_c]\mathbf{1}[\widehat{y}^b_u = 1]$. Masked consistency is

$$\mathcal{L}^r_{\mathrm{mic}} = \frac{\sum_u w^r_u \,\mathrm{CE}(z^r(x^{t,m})_u, \widehat{y}^r_u)}{\sum_u w^r_u + \epsilon}, \qquad r \in \{b, c\}. \tag{2}$$

Type consistency supervises high-confidence foreground locations using pseudo-labels from the complete target view. PSA aligns confidence-selected target features with class prototypes constructed from labeled source features. Binary consistency preserves occupancy under partial observation; foreground-restricted type consistency transfers accepted class predictions. An empty confidence mask contributes zero loss. The stop-gradient reference fixes the complete-view class assignment within a training step, while the masked prediction receives the consistency gradient. Dividing by the number of accepted positions makes the objective an average over selected supervision. The binary foreground decision also selects the locations used for type consistency, linking class learning to nuclear occupancy.

3.4 Prototype alignment and semantic relations

Source class masks $a^s_{u,k}$ define normalized prototypes $\mu^r_k = \nu(\sum_u a^s_{u,k} h_\theta(x^s)_u / (\sum_u a^s_{u,k} + \epsilon))$. Let $\mathcal{K}^r_s$ contain classes present in the source batch and Ω_r contain confident target positions with a corresponding source prototype. For temperature $T_r > 0$, prototype semantic alignment is

$$\mathcal{L}^r_{\mathrm{psa}} = -\frac{1}{|\Omega_r| + \epsilon} \sum_{u \in \Omega_r} \log \frac{\exp(\langle \nu(h_\theta(\widetilde{x}^t)_u), \mu^r_{\widehat{y}^r_u} \rangle / T_r)}{\sum_{k \in \mathcal{K}^r_s} \exp(\langle \nu(h_\theta(\widetilde{x}^t)_u), \mu^r_k \rangle / T_r)}. \tag{3}$$

An empty Ω_r contributes zero loss. Each target feature is contrasted with all available source-class prototypes, with its pseudo-label selecting the positive class. Thus, the reference comes from source annotations and the assignment comes from the target prediction. Binary prototypes separate foreground and background, while type prototypes organize the finer nuclear categories.

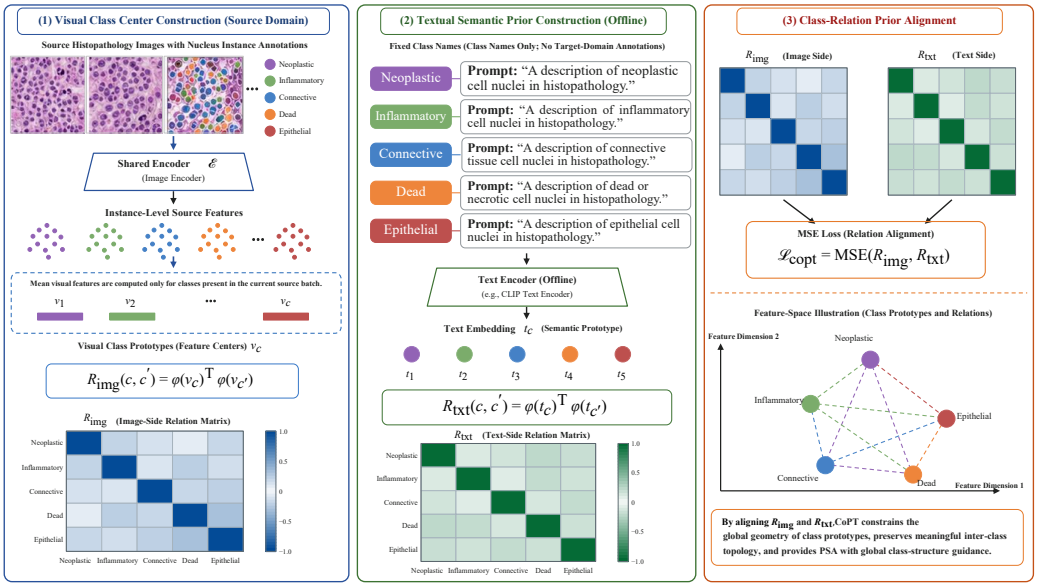


Figure 3: Details of structured semantic alignment.

To organize relationships among type centers, let v_k be a source visual class mean and t_k its fixed text embedding. A frozen CLIP encoder (Radford et al., 2021) encodes fixed class-specific histopathology descriptions, as illustrated in Figure 3. Define $R(V)_{kl} = \phi(v_k)^\top \phi(v_l)$, where ϕ centers each row and applies ℓ_2 normalization. With the same class order in text matrix T, the CoPT-inspired relation objective is

$$\mathcal{L}_{\text{copt}} = |\mathcal{K}_s^c|^{-2} \|R(V) - R(T)\|_F^2. \tag{4}$$

Only source classes present in the batch contribute; fewer than two classes give zero loss. Matching relation matrices compares the same class pairs even when visual and text features have different dimensions. PSA associates individual target locations with source classes, whereas the relation term organizes the class centers themselves. Their shared class indexing connects local feature alignment to a fixed semantic reference.

3.5 Target support and two-stage optimization

Let $\mathcal{C}$ be the unified type set and $\mathcal{V}_t \subseteq \mathcal{C}$ the allowed target annotation set, including background. Before type pseudo-label selection, apply

$$[\Pi_{\mathcal{V}_t}(p)]_c = \frac{\mathbf{1}[c \in \mathcal{V}_t] p_c}{\sum_{k \in \mathcal{V}_t} p_k + \epsilon}, \qquad \mathcal{L}_{\text{tcc}} = \frac{1}{|\Omega|} \sum_{u \in \Omega} \sum_{c \notin \mathcal{V}_t} p_u^c(c), \tag{5}$$

TCC is applied only when $\mathcal{V}_t \subsetneq \mathcal{C}$. It suppresses forbidden classes and renormalizes the admissible set. For PanNuke-to-Lizard, the admissible classes are Inflammatory, Connective, and Epithelial. Internal PanNuke transfers and PanNuke-to-CoNSeP use $\mathcal{V}_t = \mathcal{C}$, so $\Pi_{\mathcal{V}_t}$ is the identity and $\mathcal{L}_{\text{tcc}} = 0$. Joint pre-adaptation minimizes

$$\begin{aligned} \mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{src}} + \lambda_s \mathcal{L}_{\text{sfa}} + \lambda_m (\mathcal{L}_{\text{mic}}^b + \lambda_{tc} \mathcal{L}_{\text{mic}}^c) \\ + \lambda_{pb} \mathcal{L}_{\text{psa}}^b + \lambda_{pc} \mathcal{L}_{\text{psa}}^c + \lambda_c \mathcal{L}_{\text{copt}} + \lambda_t \mathcal{L}_{\text{tcc}} + \lambda_e \mathcal{L}_{\text{ent}}, \end{aligned} \tag{6}$$

where $\mathcal{L}_{\text{ent}} = -|\Omega|^{-1} \sum_{u,k} p_u^b(k) \log p_u^b(k)$ is optional. After two pre-adaptation epochs, source refinement optimizes $\mathcal{L}_{\text{sup}}$. The protocol uses source-validation loss for checkpoint selection and reserves target annotations for final evaluation. Target images therefore inform the initial adaptation stage, followed by source supervision of foreground, type, and instance geometry.

At each pre-adaptation step, source labels supervise both source views and define the masks for foreground agreement and prototype construction. The complete target view supplies reference predictions and features for prototype alignment; the masked view supplies the consistency predictions. Class-support filtering precedes type pseudo-label selection only when $\mathcal{V}_t \subsetneq \mathcal{C}$. During refinement, target-dependent losses are removed and source supervision continues to optimize the segmentation network.

4 Experiments

4.1 Datasets and evaluation protocol

We evaluate internal PanNuke transfers in multi-source/multi-target and Breast-centered settings (Gamper et al., 2019), and external transfer from PanNuke to CoNSeP (Graham et al., 2019) and Lizard (Graham et al., 2021). All inputs use 256×256 patches with stride 256 and consistent instance reindexing.

Label space. The unified space contains Neoplastic, Inflammatory, Connective, Dead, Epithelial, and background. Lizard neutrophil, lymphocyte, plasma, and eosinophil annotations map to Inflammatory, while connective tissue and epithelial annotations map to Connective and Epithelial. CoNSeP malignant and normal epithelial labels map to Neoplastic and Epithelial; inflammatory maps to Inflammatory; fibroblast, muscle, and endothelial labels merge into Connective; miscellaneous maps to Dead. These mappings define annotation conventions rather than biological equivalence.

Training and target access. The three PanNuke folds are pooled before partitioning the source data into training and validation sets. Joint pre-adaptation lasts two epochs with synchronized source

and target batches of sizes 4 and 8, respectively. Source-only supervised refinement follows, with model selection based on source validation. Target annotations are reserved for final evaluation. Target images include those subsequently evaluated; their labels are reserved for evaluation. AdamW uses learning rate 3×10^{-4} and weight decay 10^{-4} with mixed precision on an NVIDIA RTX 3090.

Metrics and comparisons. Binary PQ evaluates instance segmentation without type distinctions, whereas mPQ averages class-specific PQ values. Detection F1 assesses instance detection, and class-wise F-scores characterize classification-aware detection. Classes absent from the mapped target label space are excluded from the corresponding class averages. Figure 4 and Tables 1–3 pair cross-organ visual comparisons with the complete classwise results.

4.2 CROSS-ORGAN COMPARISON

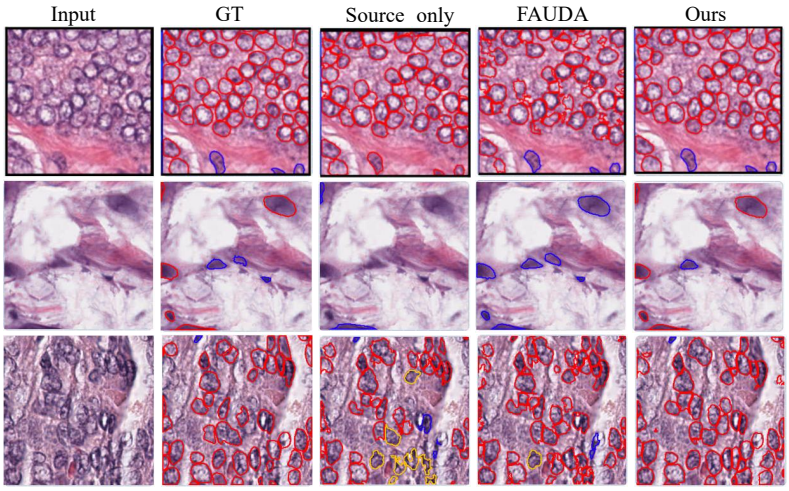


Figure 4: Internal-transfer examples: Input, GT, Source-only, FAUDA, and our method. Rows show densely packed, weakly stained, and type-ambiguous nuclei.

Table 1: Breast → Testis: classwise PQ and the recorded F score / F1.

	PQ score ↑					F score / F1 ↑				
Method	Neo.	Epi.	Conn.	Infl.	Avg.	Neo.	Epi.	Conn.	Infl.	Avg.
Source-only	0.207	0.029	0.344	0.309	0.222	0.428	0.070	0.529	0.607	0.409
DA-RCNN	0.235	0.179	0.323	0.340	0.269	0.527	0.357	0.579	0.460	0.481
BAFA	0.236	0.145	0.311	0.332	0.256	0.535	0.215	0.510	0.572	0.458
MGA	0.243	0.150	0.289	0.266	0.237	0.540	0.302	0.556	0.358	0.439
HT	0.214	0.207	0.306	0.313	0.260	0.518	0.373	0.585	0.579	0.514
DA-CC	0.252	0.390	0.324	0.385	0.338	**0.596**	0.521	**0.594**	0.645	0.589
DDSPSeg	0.388	0.489	0.382	0.472	0.432	0.541	0.715	0.546	0.615	0.604
FAUDA	0.402	0.503	0.395	0.489	0.447	0.553	0.730	0.558	0.629	0.618
Ours	**0.430**	**0.537**	**0.421**	**0.528**	**0.479**	0.580	**0.762**	0.584	**0.660**	**0.647**

Table 2: Breast → Thyroid: classwise PQ and the recorded F score / F1.

	PQ score ↑					F score / F1 ↑				
Method	Neo.	Epi.	Conn.	Infl.	Avg.	Neo.	Epi.	Conn.	Infl.	Avg.
Source-only	0.144	0.032	0.320	0.196	0.173	0.311	0.036	0.445	0.368	0.290
DA-RCNN	0.095	0.138	0.237	0.271	0.185	0.224	0.303	0.381	0.417	0.331
BAFA	0.127	0.103	0.302	0.264	0.199	0.293	0.228	0.477	0.434	0.358
MGA	0.121	0.045	0.282	0.257	0.176	0.306	0.116	0.448	0.413	0.321
HT	0.137	0.084	0.262	0.177	0.165	0.391	0.264	0.401	0.330	0.347
DA-CC	0.159	0.141	0.297	0.284	0.220	0.460	0.359	**0.482**	0.452	0.438
DDSPSeg	0.439	0.548	0.292	0.405	0.421	0.631	0.668	0.428	0.444	0.543
FAUDA	0.455	0.563	0.305	0.419	0.435	0.647	0.684	0.441	0.458	0.558
Ours	**0.489**	**0.605**	**0.330**	**0.448**	**0.468**	**0.674**	**0.713**	0.465	**0.486**	**0.584**

Table 3: Breast → Bile-duct: classwise PQ and the recorded F score / F1.

	PQ score ↑					F score / F1 ↑				
Method	Neo.	Epi.	Conn.	Infl.	Avg.	Neo.	Epi.	Conn.	Infl.	Avg.
Source-only	0.295	0.000	0.288	0.299	0.220	0.553	0.000	0.498	0.522	0.393
DA-RCNN	0.282	0.011	0.273	0.270	0.209	0.548	0.024	0.466	0.568	0.401
BAFA	0.297	0.015	0.257	0.301	0.218	0.558	0.049	0.465	0.590	0.416
MGA	0.294	0.000	0.263	0.295	0.212	0.535	0.000	0.473	0.526	0.384
HT	0.309	0.032	0.295	0.283	0.230	0.572	0.056	0.501	0.595	0.431
DA-CC	0.330	0.102	0.286	0.304	0.256	0.627	0.203	0.509	**0.615**	0.488
DDSPSeg	0.467	0.464	0.361	0.459	0.438	0.632	0.605	0.486	0.570	0.573
FAUDA	0.482	0.478	0.374	0.471	0.451	0.647	0.619	0.499	0.584	0.587
Ours	**0.516**	**0.510**	**0.401**	**0.504**	**0.483**	**0.672**	**0.644**	**0.523**	0.609	**0.612**

4.3 CROSS-DATASET ADAPTATION

We use a cumulative ablation sequence to analyze the adaptation components. Because TCC is active only when the target label support is a strict subset of the unified class space, the complete cumulative sequence (Source Only → SFA → MIC → PSA → CoPT → TCC) is reported on PanNuke→Lizard. PanNuke→CoNSeP retains the full class support, so TCC is inactive there and no TCC-specific cumulative step is reported.

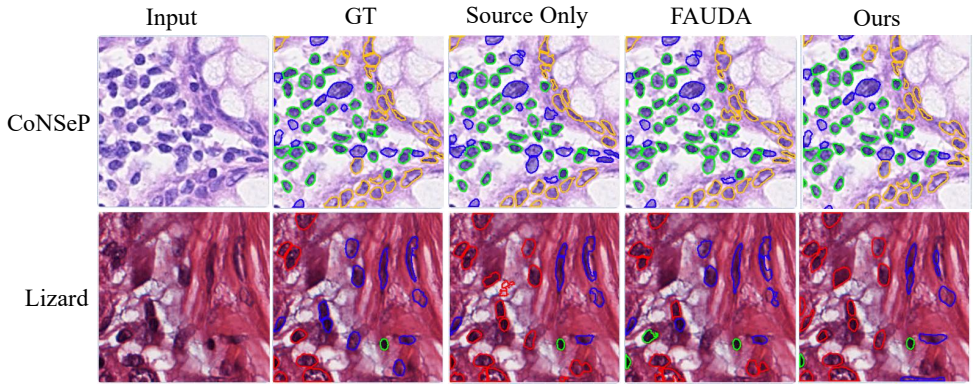


Figure 5: External-transfer examples on CoNSeP and Lizard. Columns compare Input, GT, Source-only, FAUDA, and our method. The Lizard cumulative ablation and the external-target summary appear in Tables 4 and 5, respectively.

Table 4: Cumulative adaptation on PanNuke → Lizard; TCC is active only in the final step.

Variant	bPQ ↑	mPQ ↑	F1 ↑
A: Source Only	0.2182	0.0733	0.6418
B: + SFA	0.2239	0.0808	0.6461
C: + MIC	0.2294	0.0936	0.6514
D: + PSA	0.2276	0.0909	0.6488
E: + CoPT	0.2328	0.1038	0.6549
F: + TCC	**0.2369**	**0.1127**	**0.6598**

Table 5: External-target summary. Deltas are relative to Source Only.

Target	bPQ	ΔbPQ	mPQ	ΔmPQ	F1	ΔF1	Precision	Recall
CoNSeP	0.4176	0.0253	0.2335	0.1276	0.7341	0.0179	0.8248	0.6616
Lizard	0.2369	0.0187	0.1127	0.0394	0.6598	0.0180	0.9012	0.5302

The final reported configuration improves mPQ from 0.1059 to 0.2335 on CoNSeP and from 0.0733 to 0.1127 on Lizard. The respective absolute gains are 0.1276 and 0.0394, alongside bPQ gains of 0.0253 and 0.0187. On CoNSeP, TCC is inactive because the target support equals the full unified class space; accordingly, the CoNSeP endpoint is not interpreted as a TCC gain.

On Lizard, the cumulative sequence shows mPQ values of 0.0808 after SFA, 0.0936 after MIC, 0.0909 after PSA, 0.1038 after CoPT, and 0.1127 after activating TCC. We therefore use Lizard for the TCC-specific cumulative analysis, where the target annotation support is a strict subset of the unified label space.

Table 5 further separates detection precision and recall. Lizard yields precision 0.9012 and recall 0.5302, while CoNSeP yields 0.8248 and 0.6616. These operating points identify missed nuclei as a remaining component of the external-transfer error. Figure 5 complements the aggregate scores with the original visual comparisons.

4.4 Component analysis

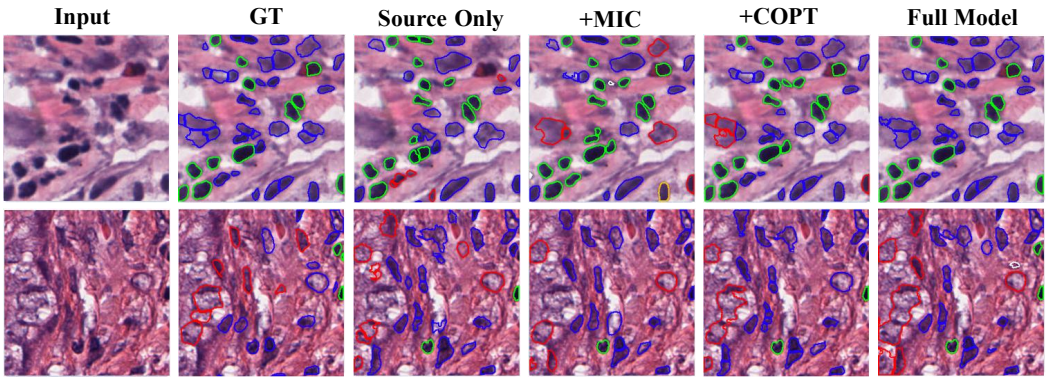


Figure 6: Qualitative component comparison. Columns show Input, GT, Source Only, +MIC, +CoPT, and Full Model on the same target examples.

Table 6: Removal of shared adaptation components on the two external targets. TCC is analyzed separately on Lizard, where the target label support is restricted.

	PanNuke → CoNSeP			PanNuke → Lizard		
Configuration	bPQ ↑	mPQ ↑	F1 ↑	bPQ ↑	mPQ ↑	F1 ↑
Final configuration	**0.4176**	**0.2335**	**0.7341**	**0.2369**	**0.1127**	**0.6598**
w/o MIC	0.3989	0.1798	0.7186	0.2218	0.0724	0.6442
w/o PSA	0.4077	0.2167	0.7276	0.2323	0.1011	0.6549
w/o CoPT	0.4042	0.2039	0.7249	0.2287	0.0908	0.6508

Table 7: Independent module additions on PanNuke → Lizard.

Configuration	bPQ ↑	mPQ ↑	F1 ↑
Source Only	0.2182	0.0733	0.6418
Source Only + PSA	0.2219	0.0814	0.6431
Source Only + CoPT	0.2204	0.0768	0.6424
Source Only + TCC	0.2228	0.0831	0.6452
Full Model	**0.2369**	**0.1127**	**0.6598**

Figure 6 shows its corresponding visual progression. Removing MIC changes mPQ from 0.2335 to 0.1798 on CoNSeP and from 0.1127 to 0.0724 on Lizard; removing PSA or CoPT also reduces mPQ on both targets. On Lizard, adding PSA to SFA and MIC decreases mPQ from 0.0936 to 0.0909 (Table 4), whereas removing PSA from the full model decreases mPQ from 0.1127 to 0.1011 (Table

6), suggesting that its effect depends on the component configuration. TCC-specific evidence is reported only on Lizard: the cumulative sequence increases mPQ from 0.1038 to 0.1127 when TCC is activated, while the independent TCC addition reaches 0.0831 versus 0.0733 for Source Only.

4.5 INTERNAL TRANSFER ACROSS TISSUES

Table 8: Internal PanNuke adaptation in two transfer settings. TCC is inactive because source and target share the full type space.

	Multi-source / multi-target			Breast-centered		
Configuration	bPQ ↑	mPQ ↑	F1 ↑	bPQ ↑	mPQ ↑	F1 ↑
Source Only	0.6092	0.4502	0.7951	0.5884	0.4389	0.7903
Proposed Method	**0.6306**	**0.4734**	**0.8095**	**0.6095**	**0.4612**	**0.8055**

Table 9: Tissue-level transfer. Gains and averages follow the supplied evaluation record.

	mPQ ↑			bPQ ↑		
Tissue	Source Only	**Ours**	Gain ↑	Source Only	**Ours**	Gain ↑
Adrenal_gland	0.4407	**0.4860**	0.0453	0.6251	**0.6823**	0.0572
Bile_duct	0.4594	**0.4886**	0.0292	0.6172	**0.6655**	0.0483
Bladder	0.5285	**0.5657**	0.0372	0.6421	**0.6924**	0.0503
Cervix	0.4725	**0.4930**	0.0205	0.6233	**0.6618**	0.0385
Esophagus	0.4849	**0.5273**	0.0424	0.5862	**0.6469**	0.0607
HeadNeck	0.3824	**0.4403**	**0.0579**	0.5497	**0.6145**	**0.0648**
Liver	0.4685	**0.5031**	0.0346	0.6570	**0.7122**	0.0552
Ovarian	0.4749	**0.5091**	0.0342	0.5845	**0.6371**	0.0526
Testis	0.4570	**0.4929**	0.0359	0.6275	**0.6783**	0.0508
Thyroid	0.4157	**0.4465**	0.0308	0.6442	**0.6890**	0.0448
Uterus	0.3997	**0.4234**	0.0237	0.5712	**0.6120**	0.0408
Average	**0.4531**	**0.4887**	**0.0356**	**0.6116**	**0.6629**	**0.0513**

The reported method raises mPQ from 0.4502 to 0.4734 in multi-source/multi-target adaptation and from 0.4389 to 0.4612 in Breast-centered adaptation (Table 8). Since internal tasks share the full type space, TCC is inactive; this table is therefore presented as an endpoint transfer comparison rather than a TCC ablation.

The separate tissue-level record in Table 9 reports higher bPQ and mPQ in all eleven tissues. Average mPQ increases from 0.4531 to 0.4887, and bPQ from 0.6116 to 0.6629. These aggregates complement the overall internal evaluation with a tissue-level account of transfer.

5 CONCLUSION

We presented structured semantic consistency for unsupervised cross-domain nuclei instance segmentation. Source foreground preservation and masked target-view agreement are combined with source prototypes and text-derived class relations, with target annotation support applied only when source and target label spaces differ. Joint adaptation precedes supervised source refinement in a shared network. The reported internal and external results show improvements in binary and type-aware instance quality, supporting complementary structural and semantic references for cross-domain transfer.

AI USE STATEMENT

OpenAI Codex was used solely to assist with manuscript writing, including language polishing, manuscript organization, and LaTeX formatting. All scientific ideas, methodological designs, experiments, analyses, results, and conclusions were developed and verified by the authors. No AI tools were used for literature discovery, research ideation or execution, synthetic data generation, or mathematical proof.

ETHICS STATEMENT

This work studies nuclei segmentation using public histopathology benchmarks. The model is intended for research use. Benchmark segmentation performance does not establish clinical validity, and errors in instance separation or nuclear typing can affect downstream cellular statistics. Cross-dataset type mappings specify evaluation conventions rather than diagnostic equivalences.

REPRODUCIBILITY STATEMENT

The proposed method is specified in Section 3, including the network outputs, source supervision, masked target-view consistency, prototype alignment, text-derived relation regularization, conditional target-class support, and the complete training objective. Section 4 reports the datasets, cross-dataset label mappings, source/target access protocol, optimization schedule, hardware, and evaluation metrics used for the experiments. Target annotations are excluded from training and checkpoint selection and are used only for final evaluation; model selection relies on source-domain validation. All datasets used in this work are publicly available. To support independent reproduction, the authors will release the training and evaluation code, experiment configurations, class-mapping utilities, and the fixed text prompts/embeddings used by the method upon publication.

REFERENCES

Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 865–872, 2019. doi: 10.1609/aaai.v33i01.3301865.

Jianan Fan, Dongnan Liu, Canran Li, Hang Chang, Heng Huang, Filip Braet, Mei Chen, and Weidong Cai. Revisiting adaptive cellular recognition under domain shifts: A contextual correspondence view. In *Computer Vision—ECCV 2024*, pp. 275–292, 2024. doi: 10.1007/978-3-031-73464-9_17.

Wei Feng, Lie Ju, Lin Wang, Kaimin Song, Xin Zhao, and Zongyuan Ge. Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 623–631, 2023. doi: 10.1609/aaai.v37i1.25138.

Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benet, Ali Khuram, and Nasir Rajpoot. PanNuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification. In *Digital Pathology: European Congress on Digital Pathology*, pp. 11–19, 2019. doi: 10.1007/978-3-030-23937-4_2.

Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. HoVer-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019. doi: 10.1016/j.media.2019.101563.

Simon Graham, Mostafa Jahanifar, Ayesha Azam, Mohammed Nimir, Yee-Wah Tsang, Katherine Dodd, Emily Hero, Harvir Sahota, Atisha Tank, Ksenija Benes, Noorul Wahab, Fayyaz Minhas, Shan E Ahmed Raza, Hesham El Daly, Kishore Gopalakrishnan, David Snead, and Nasir Rajpoot. Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification. In

Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 684–693, 2021.

Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, and Jens Kleesiek. CellViT: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024. doi: 10.1016/j.media.2024.103143.

Lukas Hoyer, Dengxin Dai, and Luc Van Gool. DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9924–9935, 2022a.

Lukas Hoyer, Dengxin Dai, and Luc Van Gool. HRDA: Context-aware high-resolution domain-adaptive semantic segmentation. In *Computer Vision—ECCV 2022*, pp. 372–391, 2022b. doi: 10.1007/978-3-031-20056-4_22.

Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. MIC: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11721–11732, 2023.

Canran Li, Dongnan Liu, Haoran Li, Zheng Zhang, Guangming Lu, Xiaojun Chang, and Weidong Cai. Domain adaptive nuclei instance segmentation and classification via category-aware feature alignment and pseudo-labelling. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*, pp. 715–724, 2022. doi: 10.1007/978-3-031-16449-1_68.

Cristina Mata, Kanchana Ranasinghe, and Michael S. Ryoo. CoPT: Unsupervised domain adaptive segmentation using domain-agnostic text embeddings. In *Computer Vision—ECCV 2024*, pp. 424–440, 2024. doi: 10.1007/978-3-031-73033-7_24.

Christian S. Perone, Pedro Ballester, Rodrigo C. Barros, and Julien Cohen-Adad. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. *NeuroImage*, 194: 1–11, 2019. doi: 10.1016/j.neuroimage.2019.03.026.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021.

Yongming Rao, Wenliang Zhao, Guangyi Chen, Yansong Tang, Zheng Zhu, Guan Huang, Jie Zhou, and Jiwen Lu. DenseCLIP: Language-guided dense prediction with context-aware prompting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18082–18091, 2022.

Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*, pp. 265–273, 2018. doi: 10.1007/978-3-030-00934-2_30.

Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: A generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, 2021. doi: 10.1038/s41592-020-01018-x.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Cristian Tommasino, Cristiano Russo, and Antonio M. Rinaldi. NuLite—lightweight and fast model for nuclei instance segmentation and classification. *Biomedical Signal Processing and Control*, 114:109333, 2026. doi: 10.1016/j.bspc.2025.109333.

Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. DACS: Domain adaptation via cross-domain mixed sampling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1379–1389, 2021.

Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7472–7481, 2018.

Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel, and Anurag Ranjan. FastViT: A fast hybrid vision transformer using structural reparameterization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5785–5795, 2023.

Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Perez. ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2517–2526, 2019.

Junlin Yang, Nicha C. Dvornek, Fan Zhang, Julius Chapiro, MingDe Lin, and James S. Duncan. Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*, pp. 255–263, 2019. doi: 10.1007/978-3-030-32245-8_29.

Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12414–12424, 2021.

Boyun Zheng, Ranran Zhang, Songhui Diao, Jingke Zhu, Yixuan Yuan, Jing Cai, Liang Shao, Shuo Li, and Wenjian Qin. Dual domain distribution disruption with semantics preservation: Unsupervised domain adaptation for medical image segmentation. *Medical Image Analysis*, 97:103275, 2024. doi: 10.1016/j.media.2024.103275.