Original Article

Weakly Supervised Teacher–Student Framework with Progressive Pseudo-mask Refinement for Gland Segmentation

Hikmat Khan[*] , Wei Chen and Muhammad Khalid Khan Niazi

Department of Pathology, College of Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, USA

Abstract

Background and objectives: Colorectal cancer histopathological grading relies on the accurate segmentation of glandular structures. Current deep learning–based methods depend heavily on large-scale pixel-level annotations that are labor-intensive and not amenable to clinical practice. Weakly supervised semantic segmentation offers a promising alternative; yet, existing class activation map–based weakly supervised semantic segmentation approaches often produce incomplete, low-quality pseudo-masks that overemphasize discriminative regions and fail to provide reliable supervision for unannotated glandular structures, limiting their suitability for dense histopathology segmentation under sparse supervision. We propose a novel weakly supervised teacher–student framework that leverages sparse pathologists' annotations and an Exponential Moving Average–stabilized teacher network to generate refined pseudo-masks. ***Methods:*** Our framework integrates confidence-based filtering, adaptive fusion of teacher predictions with limited ground truth, and curriculum-guided refinement, enabling the student network to progressively delineate and accurately segment unannotated glandular regions. We validated our framework on an institutional colorectal cancer cohort from The Ohio State University Wexner Medical Center, consisting of 60 hematoxylin and eosin-stained whole-slide images from independent patients with varying degrees of gland differentiation, as well as on public benchmarks including the Gland Segmentation dataset (derived from stage T3–T4 colorectal adenocarcinomas), TCGA-COAD, TCGA-READ, and SPIDER. ***Results:*** The proposed framework achieved strong performance on the institutional dataset despite limited annotations. On the Gland Segmentation dataset, it demonstrated competitive performance compared to both weakly and fully supervised approaches, achieving a mean Intersection over Union of 80.10% ± 1.52 and a mean Dice coefficient of 89.10% ± 2.10. Moreover, cross-cohort evaluations showed robust generalization on TCGA-COAD and TCGA-READ without requiring additional annotations, while reduced performance on SPIDER reflected pronounced domain shift. ***Conclusions:*** Our framework provides an annotation-efficient and generalizable paradigm for accurate gland segmentation in colorectal histopathology, offering a practical pathway toward significantly reducing annotation burdens while preserving high segmentation fidelity.

Introduction

Colorectal cancer (CRC) is a leading cause of global cancer mortality, with epidemiological projections indicating a 60% increase in incidence by 2030, resulting in approximately 2.2 million new cases and 1.1 million fatalities annually.[1] Histopathological assessment of tissue architecture using hematoxylin and eosin (H&E)–stained sections is the gold standard for evaluating glandular structures in various types of adenocarcinomas, including CRC,[2] breast cancer,[3] prostate cancer,[4] and endometrial adenocarcinoma.[5] Histopathologic grading relies heavily on the degree of gland formation: well- and moderately differentiated tumors are classified as low grade, retaining largely intact glandular architecture, whereas poorly differentiated tumors display markedly complex, abortive, or absent glandular formation and are associated with worse survival outcomes.[6] Consequently, accurate gland segmentation in digitized whole-slide images (WSIs) is crucial for quantifying and characterizing glandular morphology, which directly informs tumor grading and risk stratification of colorectal and other types of adenocarcinomas.[3–5]

Fully supervised deep learning methods have set the benchmark for gland segmentation in histopathological images. Early work by DCAN established a multi-task learning framework that simultaneously segments glands and their contours to delineate benign, malignant, and closely apposed glands, and subsequent advances have incorporated domain-specific inductive biases,[7] such as Gabor-based encoders and topology-aware networks.[8,9] To address residual segmentation errors, especially in apposed glands, Xie *et al.*[10] introduced a Deep Segmentation-Emendation model, which

[*]**Correspondence to:** Hikmat Khan, Department of Pathology, College of Medicine, The Ohio State University Wexner Medical Center, Pelotonia Research Center, 2nd Floor, 2281 Kenny Road, Suite 450, Columbus, OH 43210, USA. ORCID: https://orcid.org/0009-0008-2550-1991. Tel: +1-856-405-1073, E-mail: Hikmat.khan@osumc.edu

employs a dedicated emendation network to predict and correct inconsistencies in initial segmentation masks. While fully supervised methods demonstrate impressive performance, their effectiveness is intrinsically contingent upon the availability of large-scale, pixel-level annotated datasets—a major bottleneck in clinical practice due to the significant time and expertise required from pathologists.

This annotation burden has spurred interest in weakly supervised semantic segmentation (WSSS), which substantially reduces the demand for dense labels by leveraging weaker forms of supervision such as image-level labels or sparse annotations, thereby reducing annotation time by approximately sixty-fold.[11,12] The predominant WSSS approach uses classification networks to generate class activation maps (CAMs) as initial pseudo-labels for training segmentation models.[13–22] However, CAMs have inherent limitations; they tend to activate only the most discriminative regions of an object, resulting in pseudo-masks with ambiguous boundaries, noise, discontinuity, and structural fragmentation.[17] Various CAM refinement techniques have been developed to address these limitations, including SEAM, which enforces spatial consistency,[23] and AMR, which uses complementary activation branches to enhance under-activated regions.[24] Similarly, Kweon *et al*.[20] adopted an adversarial strategy, using an image reconstructor to force the classifier to generate more complete activation maps by minimizing inter-segment inferability. In medical imaging, domain-specific modifications such as C-CAM and MLPS have also been proposed.[11,12,25]

Nevertheless, a significant challenge persists in learning from the noisy, CAM-generated pseudo-masks in the subsequent segmentation stage. Although recent frameworks such as ARML attempt to address both CAM refinement and noisy label learning in histopathology,[21] general WSSS methods still often underperform on gland segmentation due to the high morphological similarity between gland types and the critical need for precise instance boundaries. Therefore, there remains a significant methodological gap for a WSSS framework specifically designed to address the challenges of gland segmentation—namely, to generate high-quality, complete pseudo-masks from sparse annotations that can reliably guide the training of a dense segmentation model.

To bridge this gap, we introduced a novel weakly supervised teacher–student framework with progressive pseudo-mask refinement for multi-class gland segmentation in colorectal histopathology. The framework integrates an Exponential Moving Average (EMA)–stabilized teacher network with confidence-based filtering, curriculum-guided loss weighting, and adaptive pixel-wise fusion of sparse expert annotations to progressively discover and segment previously unannotated glandular structures.

Our contributions are threefold:
1. We introduce a pixel-wise pseudo-label fusion strategy that preserves pathologist-provided sparse annotations while leveraging EMA-stabilized teacher predictions to supervise unlabeled regions during self-training.
2. We propose a curriculum-driven refinement mechanism that combines cosine-decayed confidence thresholding with dynamic loss weighting, enabling progressive expansion of supervision from high-confidence gland regions to previously unannotated and ambiguous regions. This approach explicitly addresses annotation sparsity in the dense and morphologically complex setting of glandular histopathology.
3. We perform a comprehensive, clinically grounded multi-cohort evaluation reflecting real-world variability. The framework is validated on (i) an institutional dataset with sparse annotations, (ii) the fully annotated public

Gland Segmentation (GlaS) benchmark,[26] and (iii) three external cohorts—The Cancer Genome Atlas (TCGA) Colon Adenocarcinoma (COAD), Rectum Adenocarcinoma (READ), and SPIDER[27]—to assess cross-domain generalization. This multi-tiered evaluation demonstrates competitive performance relative to fully supervised methods and systematically characterizes robustness and failure modes under substantial domain shift, providing actionable insights for clinical translation.

Materials and methods

Study design and problem formulation

The proposed framework leverages the nnUNet backbone for robust semantic segmentation and comprises two identical networks: a student model (θ_S) trained via gradient descent using a supervised segmentation loss and a consistency regularization term, and a teacher model (θ_T) updated exclusively through an EMA of the student parameters, providing stable pseudo-labels that guide student learning. Formally, let $x \in R^{H \times W \times 3}$ denote an input image, and $y \in \{0,1,…,C\}^{H \times W}$ the corresponding pixel-level labels in the segmentation mask, where $C = 4$ represents the total number of classes (background stroma, benign glands, malignant glands, and poorly differentiated clusters/glands). The goal is to learn a function $f_\theta (x)$ that outputs pixel-wise class probabilities $p_\theta (x) \in [0,1]^{H \times W \times C}$ to segment both annotated and unannotated glandular structures at the pixel level.

Datasets

We conducted experiments on the in-house The Ohio State University Wexner Medical Center (OSUWMC) dataset containing limited pathologist annotations, as well as on the publicly available GlaS dataset with high-quality pixel-level annotations to demonstrate the broad applicability of the framework.[26] Additionally, three external publicly available CRC histopathology datasets, TCGA-COAD, TCGA-READ, and SPIDER,[27] were used to qualitatively assess the generalizability of the proposed framework on external cohorts where ground-truth (GT) annotations are not available.

OSUWMC in-house dataset: We used an in-house CRC histology dataset collected at OSUWMC, consisting of 60 H&E-stained WSIs from independent patients with histologically confirmed colorectal adenocarcinoma. All WSIs were retrospectively acquired from surgical resection specimens, annotated by two pathology residents using sparse pixel-level labels, and scanned at 40× magnification. The dataset is WSIs only; no patient-level clinical or demographic metadata (e.g., age, sex, tumor stage, grade, or treatment history) were collected or were available, as this study focused exclusively on technical development of weakly supervised segmentation algorithms rather than clinical outcome prediction. The annotations include four tissue categories: benign glands, malignant glands (better-formed tumor glands with obvious lumina), poorly differentiated clusters/glands (encompassing tumor buds, poorly differentiated clusters, and poorly formed tumor glands with absent or minimal lumina), and background stroma. The cohort captures a broad range of glandular morphologies, including well-formed glands, irregular malignant glands, and poorly differentiated structures, reflecting real-world histopathologic variability. WSIs were scanned at 40× magnification, and 512 × 512-pixel patches were extracted at 5× magnification for model development. In total, 74,179 patches were generated and split into 63,191 training, 5,460 validation, and 5,528 test patches. Approximate class prevalence at the patch level
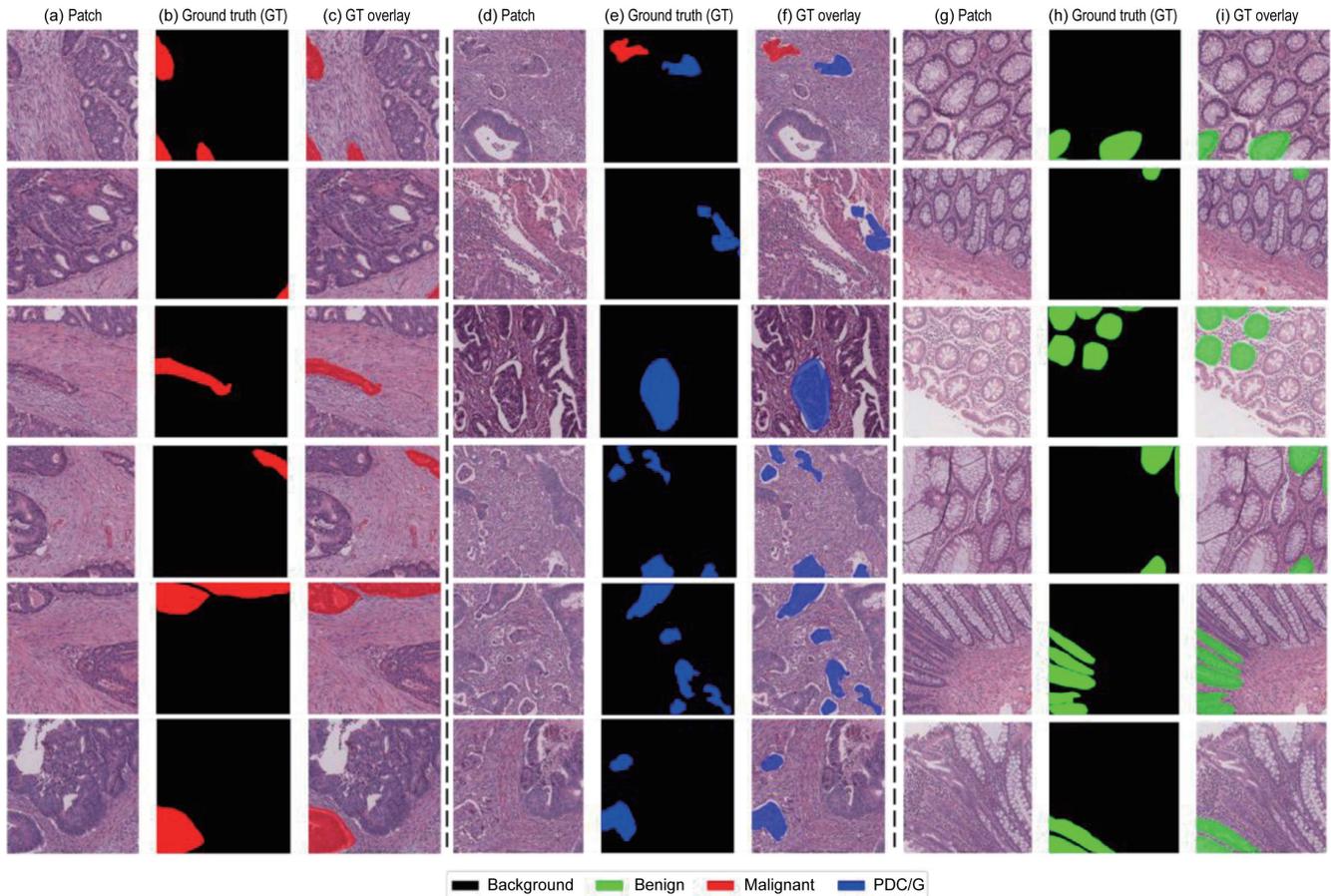
■ Background ■ Benign ■ Malignant ■ PDC/G

Fig. 1. Representative samples from the in-house The Ohio State University Wexner Medical Center (OSUWMC) dataset, illustrating sparse annotations provided by pathologists for three key gland classes: benign glands, malignant glands, and poorly differentiated clusters/glands (PDC/G). For each class, the column of triplets shows: (i) the original histopathology image, (ii) the corresponding sparse ground truth mask from the two experts, and (iii) an overlay of the annotations on the input image. Only a few sparsely annotated glandular structures are present within each patch, leaving substantial regions unlabeled, which significantly increases the difficulty of learning accurate segmentation models under weak supervision. Color coding for all segmentation masks is as follows: red represents malignant glands, green represents benign glands, blue indicates poorly differentiated clusters or glands (PDC/G), and black denotes stroma (best viewed in color).

was ~45% benign glands, ~35% malignant glands, ~15% background stroma, and ~5% PDC/G, with stratified sampling used to preserve class proportions across splits. Figure 1 shows representative samples from our in-house dataset, which contains sparse annotations for background stroma, benign glands, malignant glands, and poorly differentiated clusters/glands. Notably, most patches contained both annotated and unannotated glands, posing a significant challenge for accurate segmentation under weak supervision.

GlaS Dataset: We subsequently conducted experiments using the GlaS dataset,[26] a publicly available histological image collection released as part of the MICCAI 2015 Gland Segmentation Challenge. The dataset comprises 165 H&E-stained images extracted from 16 colorectal tissue sections, each obtained from a different patient diagnosed with stage T3 or T4 colorectal adenocarcinoma. All cases correspond to advanced-stage disease, and no earlier-stage tumors are included in the cohort.[26] Per the official GlaS challenge protocol, patient-level demographic information (e.g., age, sex, exact TNM substage) is not provided with the dataset and is not required for the benchmark segmentation task, which is defined strictly at the image and pixel level. The images were scanned at 20× magnification with a native

spatial resolution of 0.465 μm/pixel, and most have an original size of 775 × 522 pixels. Each image is accompanied by instance-level segmentation ground truth, providing precise delineation of glandular boundaries. Within each image, both benign and malignant glandular structures are present, reflecting the heterogeneous histologic architecture typical of advanced colorectal adenocarcinoma. The dataset is divided into 85 training images (37 benign and 48 malignant) and 80 test images (37 benign and 43 malignant). This benign/malignant distribution at the image level is reported in accordance with established benchmark practice and provides sufficient characterization for the segmentation task.[26] To ensure consistency with prior work and standardize input resolution, all images were resized to 512 × 512 pixels. For model development, the training set was further partitioned into 70 images for training (~82.4% of the training set) and 15 for validation (~17.6% of the training set) using a stratified sampling strategy to preserve the benign–malignant class balance, while the 80 test images were used exclusively for final performance evaluation. Because pathological stage is fixed (T3–T4) across the dataset, no stage-based stratification was required during training or evaluation. A key challenge posed by GlaS is the substantial inter-subject
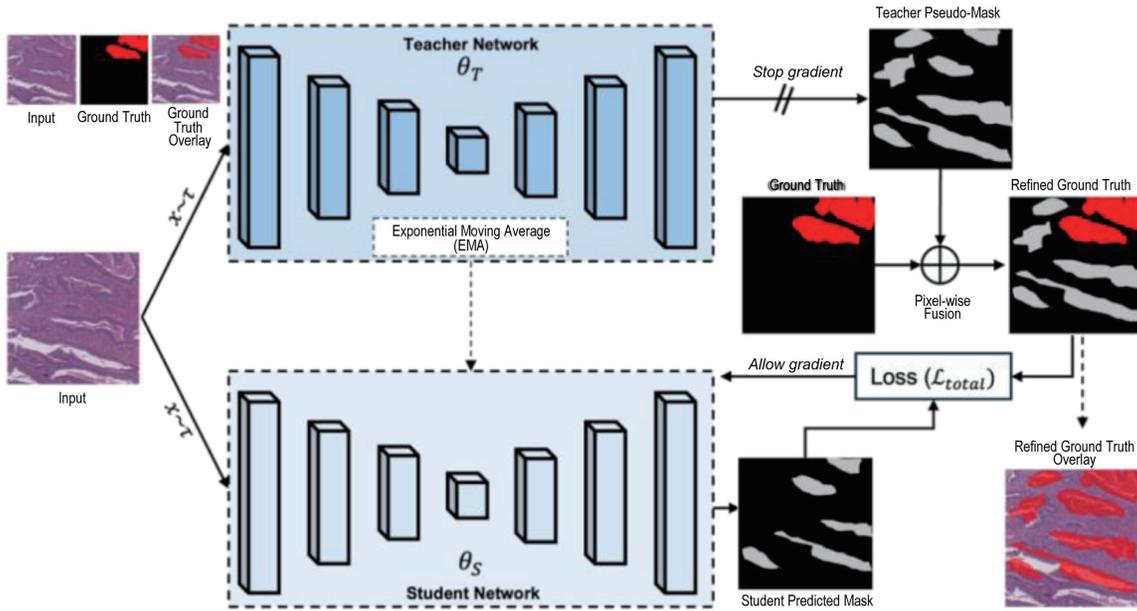
Fig. 2. Schematic overview of the proposed teacher–student self-training framework for weakly supervised multi-class gland segmentation. The teacher network (θ_T), stabilized by an Exponential Moving Average (EMA), generates initial pseudo-masks. The initial pseudo-masks are then refined via confidence-based filtering and adaptively fused with sparse ground-truth annotations to produce high-quality supervision for training the student network (θ_S). The student's parameters are then used to update the teacher via EMA. This iterative process, governed by a total loss (L_{total}), enables progressive discovery and segmentation of unannotated glandular structures (best viewed in color).

variability in staining characteristics and tissue morphology, arising from differences in laboratory processing, which makes the dataset a rigorous benchmark for gland segmentation algorithms.

Two-phase training protocol

Figure 2 illustrates a schematic overview of the proposed framework, which comprises two phases: a supervised warm-up phase and a teacher–student co-training phase.

Phase 1: Supervised warm-up: During the warm-up phase, the teacher network remains inactive, and the student network is trained solely on the available sparse annotations. This strategy ensures the student learns robust and meaningful representations that are essential for subsequent pseudo-label generation. The student network is optimized using a supervised loss ($L_{\text{supervised}}$), defined as a weighted combination of Dice loss (L_{dice}) and categorical cross-entropy loss (L_{cce}) as follows:

$$L_{\text{supervised}} = L_{dice} + L_{cce} \qquad (1)$$

here, L_{dice} maximizes the overlap between predicted and ground truth masks, while L_{cce} evaluates the pixel-wise classification accuracy across the $C = 4$ classes. Formally, these losses are defined as follows:

$$L_{dice} = 1 - \frac{2 \sum_i y_{i,c} \cdot \hat{y}_{i,c}}{\sum_i y_{i,c} + \sum_i \hat{y}_{i,c}} \qquad (2)$$

$$L_{cce} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C=4} y_{i,c} \log \hat{y}_{i,c} \qquad (3)$$

where N denotes the total number of pixels, $y_{i,c} \in \{0,1\}$ indicates whether pixel i belongs to class c, and $\hat{y}_{i,c}$ represent the predicted probability of class c at pixel i. The warm-up phase

typically spans 20% to 25% of the total epochs, providing a stable initialization for the teacher network.

Phase 2: Teacher–Student co-training: Upon completion of the warm-up phase, the teacher is initialized with the student's parameters, i.e., $\theta_T \leftarrow \theta_S$. Subsequently, the student network (θ_S) is optimized via gradient descent, while the teacher network (θ_T) is updated using an EMA of the student's weights. Formally, the teacher parameters are updated as follows:

$$\theta_T \leftarrow \beta\theta_T + (1-\beta)\beta_S \qquad (4)$$

where the EMA decay coefficient β is set to 0.999 to ensure temporally smooth teacher updates and to suppress short-term fluctuations in the student model. A high decay value is particularly important in weakly supervised dense segmentation settings, as it stabilizes pseudo-label generation and mitigates confirmation bias arising from noisy early predictions. This choice is consistent with prior teacher–student and Mean Teacher frameworks, which commonly adopt decay values in the range of 0.99–0.999 for segmentation tasks.

This update strategy yields temporally smooth teacher predictions, enhancing pseudo-label stability and mitigating confirmation bias induced by noisy CAMs. During this phase, the student is trained using a hybrid loss that integrates supervised learning on labeled data with consistency regularization provided by the teacher as follows:

$$L_{total} = \alpha(t) \, L_{supervised} + (1-\alpha(t)) \, L_{consistency} \qquad (5)$$

here, $a(t)$ is a dynamic, epoch-dependent weighting factor that governs the trade-off between the supervised and consistency losses. We employ a cosine-decaying schedule for $a(t)$ to gradually shift emphasis from GT supervision to teacher-guided consistency. After warm-up, $a(t)$ is initialized at 0.9, placing 90% reliance on supervised loss, and decays

to 0.01 by the end of training, progressively increasing reliance on teacher-generated pseudo-labels. The smooth cosine decay prevents abrupt transitions, reduces early over-reliance on noisy pseudo-labels, and enables stable late-stage refinement.

Teacher-generated pseudo-mask: The consistency term in Eq. (5) encourages the student to align with the teacher's segmentation predictions on both labeled and unlabeled pixels. To ensure the reliability of the teacher-generated pseudo-labels, we employ a confidence-based filtering mechanism that suppresses low-confidence or ambiguous pseudo-labels, particularly during the early phase of training. Formally, the confidence mask is defined as follows:

$$m(x) = 1 \left[\max \left(\sigma \left(f_{\theta_T}(x) \right) \right) > \tau_{confidence}(t) \right] \quad (6)$$

where $\sigma(.)$ is the softmax, $1[.]$ is the indicator function, and $\tau_{confidence}(t)$ is a cosine-decaying threshold that monotonically decreases from 0.95 to 0.25 over the course of training. The high initial threshold restricts supervision to only the most confident teacher predictions when the teacher model is still stabilizing, while the gradual relaxation allows progressively more ambiguous regions—such as gland boundaries and poorly differentiated structures—to be incorporated as training proceeds. This curriculum-guided design enables stable expansion of pseudo-label coverage while minimizing noise propagation.

These bounds are empirically selected to incorporate a curriculum learning strategy that emphasizes high-confidence teacher pixel-level supervision in early training and gradually incorporates other teacher-generated pseudo-labels for unlabeled regions as the teacher stabilizes. To maximally leverage sparse annotations, the teacher-generated pseudo-labels are fused with GT labels using a pixel-wise integration strategy defined as follows:

$$m(x) = \begin{cases} GT(m(x)), & if \ GT(x) > 0 \\ m(x), & otherwise. \end{cases} \quad (7)$$

This formulation ensures that pathologist-provided annotations are preserved exactly in labeled regions, while teacher-generated pixel-level pseudo-masks supervise the unlabeled regions. This fusion strategy is employed only after the teacher model has reached sufficient stability. The consistency loss is defined as follows:

$$L_{consistency} = \left\| \sigma \left(f_{\theta_S}(x) \right) - m(x) \right\|^2 \quad (8)$$

where $\sigma(.)$ denotes the softmax function, converting the student network output logits into per-pixel class probabilities. We employ logit-level mean squared error for consistency regularization, which empirically stabilizes training and reduces sensitivity to early-stage noise in the pseudo-labels.

Baselines

To benchmark the efficacy of our proposed framework, we evaluated its performance against a comprehensive set of existing methods, including thirteen WSSS and eleven fully supervised segmentation approaches.[28–30] The WSSS baselines include SEAM,[23] ReCAM,[19] AMR,[24] MLPS,[12] OEEM,[31] AME-CAM,[32] HAMIL,[33] CBFNet MPFP,[29,34] Adv-CAM,[35] SC-CAM, and MAA.[13,28,29] The fully supervised baselines consist of widely used architectures: UNet,[36] Seg-Net,[37] MedT,[38] TransUNet,[39] Attention Unet,[40] UNet++,[41] KiU-Net,[42] ResU-Net++,[43] DA-TransUNet,[39] TransAttUNet,[44] and EWASwin UNet.[30] To ensure a fair comparison, we adhered to the experimental protocols and key hyperparameters (e.g., patch size) specified in the respective original baseline publications.[28–30]

Implementation details

All experiments were conducted using PyTorch 1.13.1 with CUDA 11.7 on Python 3.10. Training was performed on NVIDIA A100 GPUs. To ensure reproducibility, the random seed was fixed at 42 across all libraries (Python, NumPy, PyTorch, and CUDA), and deterministic algorithms were enforced. However, to quantify statistical variability, we performed five independent training runs with different random seeds for all experiments and report the mean ± standard deviation across these runs. The models were trained using the AdamW optimizer with an initial learning rate of 0.01 and a weight decay of 0.001.[45] A cosine annealing schedule was employed to decay the learning rate to a minimum of 0.00001. We used a batch size of 16 and an input patch resolution of 512×512 pixels. To stabilize training, gradient clipping was applied with a maximum norm of 1.0. To enhance generalization, we utilized a comprehensive data augmentation strategy, including random discrete rotations (0°,90°,180°,270°), horizontal flipping (P = 0.5), hue–saturation–value jittering, Gaussian noise, and Gaussian blur, followed by standard ImageNet normalization.[46] The maximum training duration was set to 250 epochs, with an early stopping mechanism triggered to prevent overfitting if validation performance did not improve for 50 consecutive epochs.

Evaluation metrics

We employed two widely adopted metrics in gland segmentation[47]: mean Intersection over Union (mIoU) and mean Dice coefficient (mDice). Both metrics are derived from pixel-level classification outcomes, where each pixel is categorized as true positive (TP), false positive (FP), or false negative (FN) with respect to the GT annotation. The mIoU measures the overlap between the predicted and GT gland regions and is defined as:

$$\text{mIoU} = \frac{TP}{TP + FP + FN}.$$

The mDice evaluates the similarity between the predicted mask and the ground truth and is formulated as:

$$\text{mDice} = \frac{2 \times TP}{2 \times TP + FP + FN}.$$

Both metrics are normalized to the range [0,1], where a value of 1 indicates perfect alignment between the prediction and the ground truth, and 0 implies no overlap. Higher scores correspond to superior segmentation accuracy and better boundary delineation.

Results

To validate the efficacy of our proposed framework, we assessed the proposed framework using the public GlaS dataset with dense annotations and an in-house OSUWMC cohort to evaluate performance under sparse-label conditions. Generalization beyond the training domain was examined by applying the model trained on the OSUWMC cohort to the TCGA-COAD, TCGA-READ, and SPIDER datasets. As GT annotations are unavailable for these external cohorts, evaluation was limited to qualitative analysis. The proposed framework was compared against a broad range of state-of-the-art approaches, including weakly supervised methods (summarized in Table 1[12,13,19,23,24,28,29,31–35]) and fully supervised architectures (summarized in Table 2[30,36–44]).

Table 1. Comparison with weakly supervised gland segmentation methods on the GlaS dataset

Method	Year	mIoU (%)	mDice (%)
SEAM[23]	2020	71.36 ± 0.49	79.59 ± 4.88
ReCAM[19]	2022	56.31 ± 2.53	–
AMR[24]	2022	72.83 ± 0.37	–
MLPS[12]	2022	73.60 ± 0.16	–
OEEM[31]	2022	76.48 ± 0.10	83.40 ± 5.36
AME-CAM[32]	2023	74.09 ± 0.13	–
HAMIL[33]	2023	77.37 ± 0.73	–
CBFNet[34]	2024	76.30 ± 0.26	–
MPFP[29]	2025	80.44 ± 0.05	–
Adv-CAM[35]	2021	68.54 ± 3.36	81.33 ± 5.26
SC-CAM[13]	2020	71.52 ± 3.50	83.40 ± 5.36
MAA[28]	2025	81.99 ± 2.26	90.10 ± 3.31
Ours	–	80.10 ± 1.52	89.10 ± 2.10

Metrics are mean Intersection over Union (mIoU) and mean Dice coefficient (mDice), reported as mean ± standard deviation across five independent training runs with different random seeds. Our framework achieves competitive performance compared with current state-of-the-art approaches.

Table 2. Comparison with fully supervised gland segmentation methods on the GlaS dataset

Method	Year	mIoU (%)	mDice (%)
UNet[36]	2015	64.8	77.6
Seg-Net[37]	2017	66.0	78.6
MedT[38]	2021	69.6	81.0
TransUNet[39]	2021	70.1	81.5
AttentionUNet[40]	2018	70.1	81.6
UNet++[41]	2018	70.2	81.9
KiU-Net[42]	2020	72.8	83.3
ResUNet++[43]	2019	73.8	84.1
DA-TransUNet[39]	2024	75.6	85.3
TransAttUNet[44]	2023	77.7	86.7
EWASwin UNet[30]	2025	81.5	89.4
Ours	–	80.1	89.1

Our framework achieves competitive performance among current state-of-the-art approaches, including classical CNN-based models, transformer-enhanced architectures, and recent hybrid methods, as measured by mean Intersection over Union (mIoU) and mean Dice coefficient (mDice). Although designed for limited-annotation scenarios, the performance of our framework improves with the availability of higher-quality annotations.

Performance against weakly supervised methods

As Table 1 summarizes, our framework achieves competitive state-of-the-art performance on the GlaS benchmark,[28,29] achieving an mIoU of 80.10% and an mDice of 89.10%, while our mIoU is slightly below that of the leading MAA method. The proposed framework demonstrates markedly superior training stability, evidenced by a lower variance (±1.52 mIoU, ±2.10 mDice) compared to MAA (±2.26 mIoU, ±3.31 mDice). The high consistency and lower variance are critical prerequisites for clinical translation and underscore the robustness of our pseudo-label refinement strategy.

Performance against fully supervised methods

As Table 2 summarizes, our framework achieves competi-

tive performance compared to fully supervised state-of-the-art methods on GlaS.[30] Specifically, our framework attains 0.801 mIoU and 0.891 mDice, which are on par with the top-performing supervised baseline, EWASwin UNet (0.815 mIoU).[30] Moreover, our framework surpasses traditional architectures such as UNet++ and ResUNet++ (mIoU ~0.70–0.74) as well as other advanced models such as TransAttUNet (0.777 mIoU).[41,43,44] These quantitative findings are corroborated by the qualitative comparisons shown in Figure 3 (with additional examples in Fig. 4), which demonstrate the model's ability to generate precise segmentation masks.[30] Notably, these results underscore the capability of our weakly supervised framework to effectively leverage sparse annotations to obtain performance competitive with leading fully
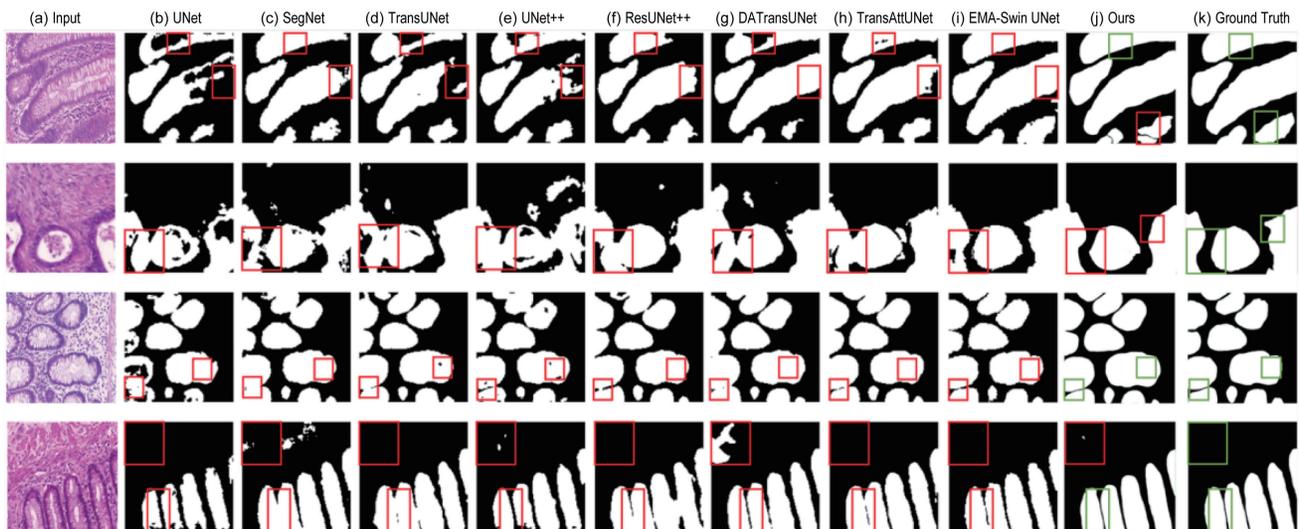

Fig. 3. Results on the GlaS dataset. In each row, the leftmost image shows the original H&E patch, followed by segmentation results from baseline methods, with the rightmost column displaying the ground truth annotations (best viewed in color). EMA, Exponential Moving Average; H&E, hematoxylin and eosin.
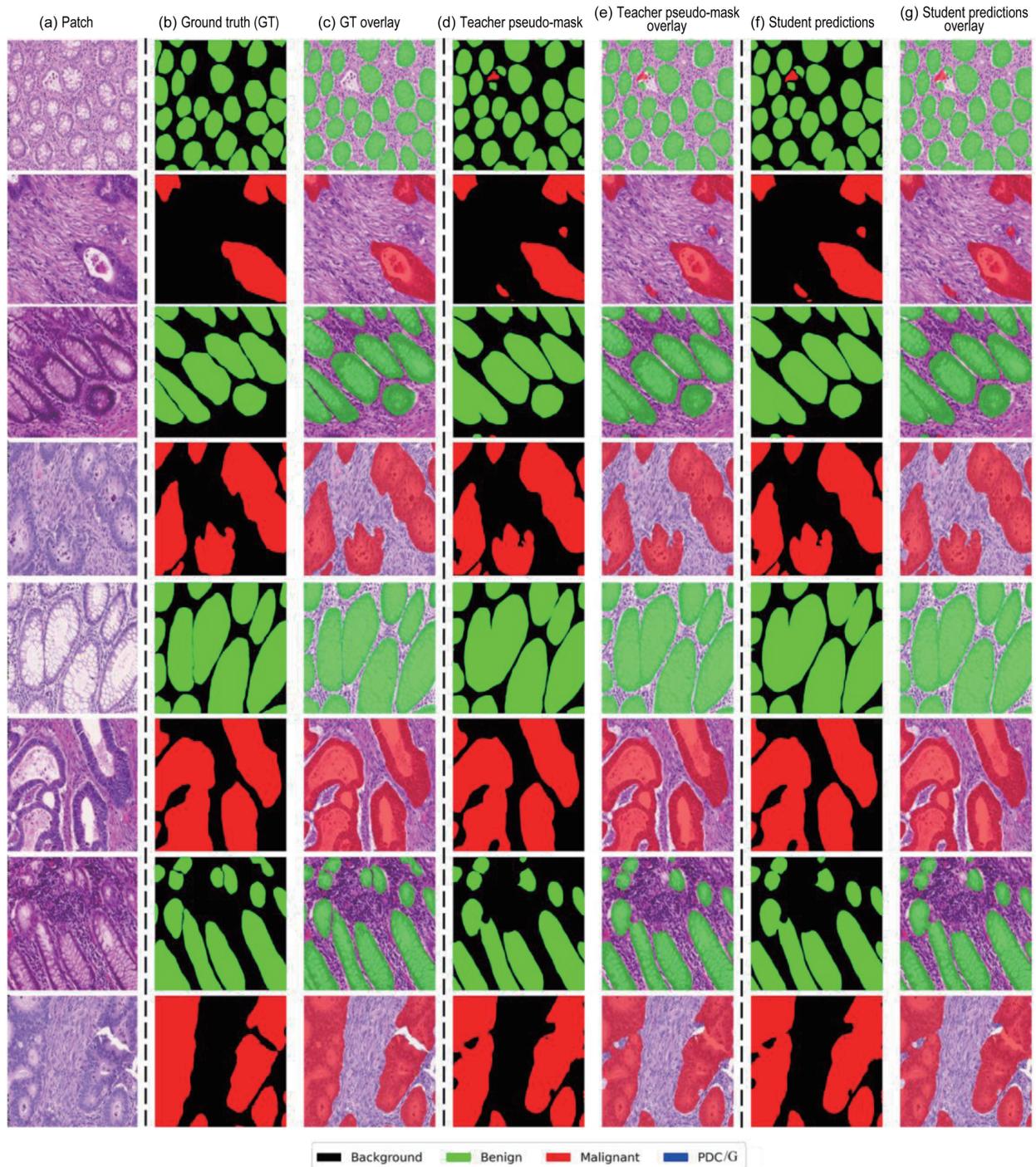
Fig. 4. Qualitative results on the GlaS test set. For each sample: (a) original H&E image; (b) dense ground truth mask; (c) overlay of ground truth on the input; (d) pseudo-mask predicted by the teacher model; (e) overlay of the teacher's prediction; (f) final segmentation mask predicted by the student model; and (g) overlay of the student's prediction. Color coding: red = malignant glands, green = benign glands, black = background stroma (best viewed in color). GT, ground-truth; H&E, hematoxylin and eosin; PDC/G, poorly differentiated clusters/glands.

supervised methods.

Results on the OSUWMC dataset and out-of-domain generalization to TCGA-COAD, TCGA-READ, and SPIDER

Figure 5 illustrates the qualitative performance of our framework on the in-house OSUWMC dataset. These visualizations demonstrate how the stabilized teacher network effectively guides the student model via pseudo-masks, enabling the discovery and precise segmentation of unannotated gland structures using only limited supervision. To evaluate robustness and clinical transferability, we performed whole-slide in-
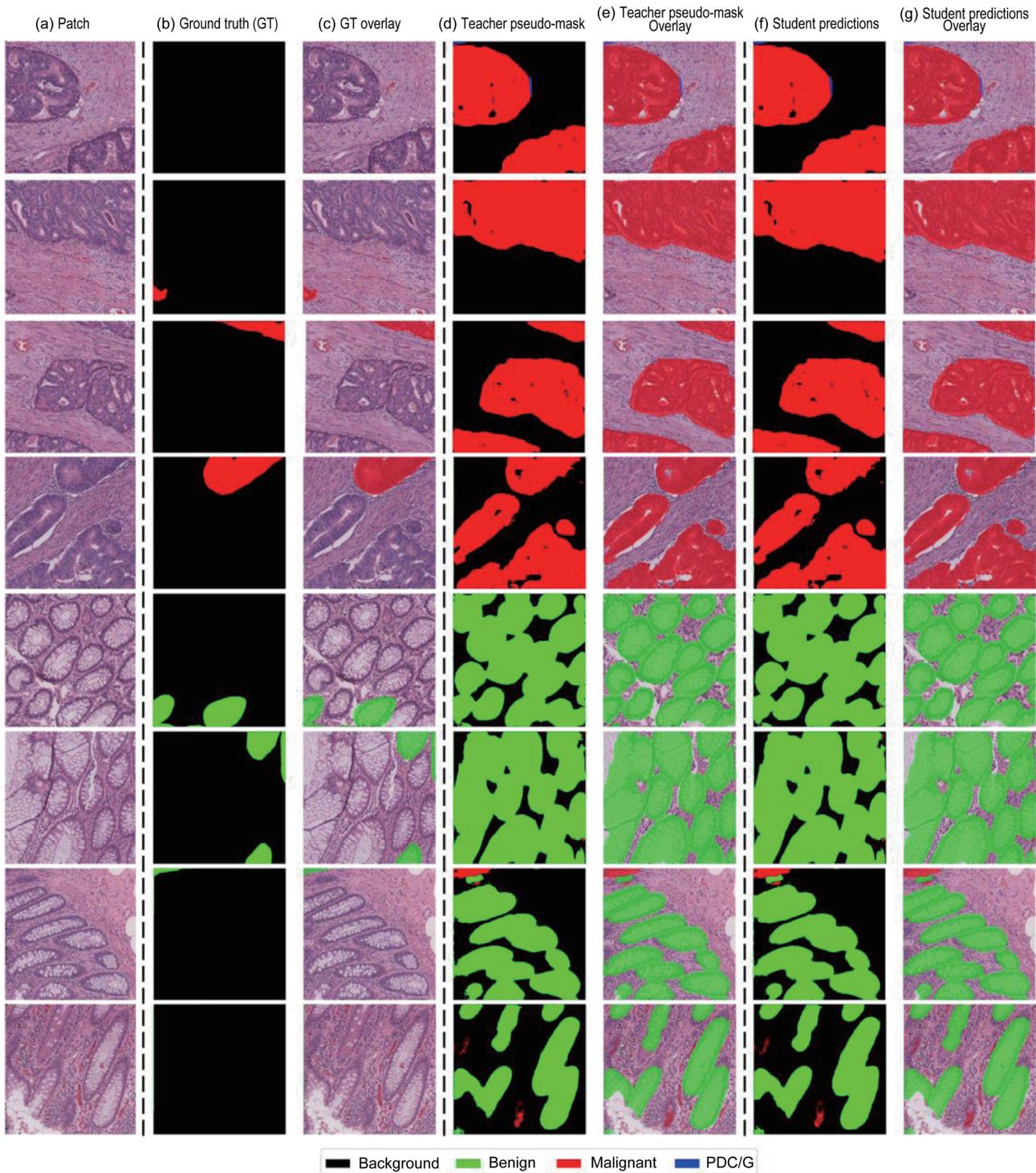
Fig. 5. Qualitative segmentation results on the in-house Ohio State University Wexner Medical Center (OSUWMC) dataset. For each representative sample: (a) input H&E image; (b) sparse ground-truth annotations provided by two pathologists; (c) annotation overlay on the input image; (d) pseudo-mask generated by the teacher model; (e) teacher pseudo-mask overlay; (f) final prediction from the student model; and (g) student prediction overlay. Color coding: red = malignant glands, green = benign glands, blue = poorly differentiated clusters/glands (PDC/G), black = background stroma. This visualization highlights the framework's ability to perform gland segmentation under sparse or limited-annotation supervision (best viewed in color). H&E, hematoxylin and eosin.

ference on three external cohorts, including TCGA-COAD, TCGA-READ, and SPIDER. Despite significant inter-institutional variations in staining protocols and scanner characteristics, our model maintained consistent qualitative performance (see Fig. 6), successfully identifying benign glands, malignant glands, and poorly differentiated clusters/glands on TCGA-
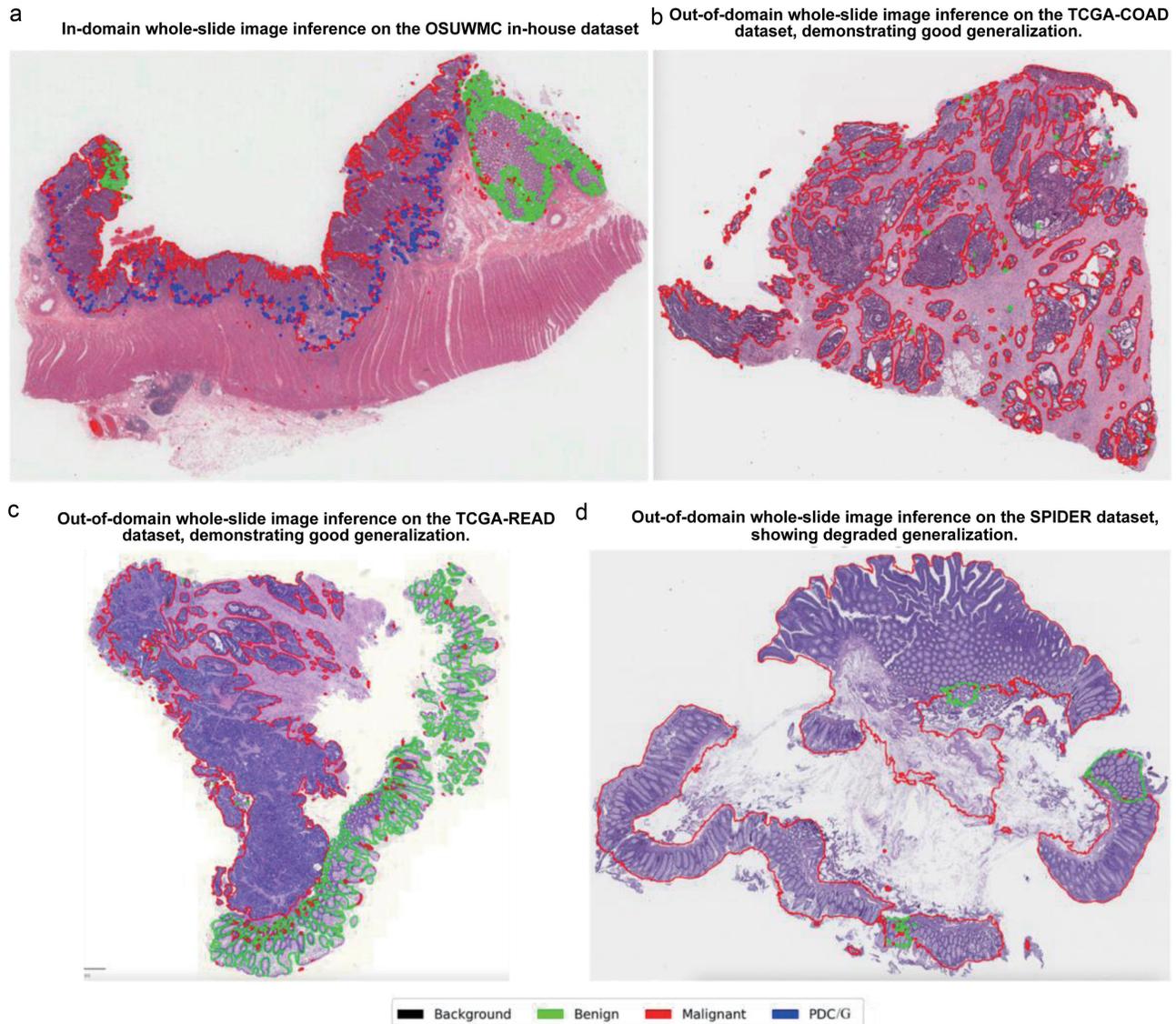
Fig. 6. Whole-slide–level qualitative assessment across (a) in-house OSUWMC, (b) TCGA-COAD (The Cancer Genome Atlas Colon Adenocarcinoma), (c) TCGA-READ (Rectal Adenocarcinoma), and (d) SPIDER, illustrating cross-domain generalization. Color coding: red = malignant glands, green = benign glands, blue = poorly differentiated clusters/glands (PDC/G), black = background stroma. Our framework performs robustly on OSUWMC with limited ground-truth annotation and on TCGA-COAD and TCGA-READ without any ground-truth annotation. Performance degradation on SPIDER—also evaluated without any annotation—is attributed to lower image quality, staining heterogeneity, and substantial domain shift (best viewed in color).

COAD and TCGA-READ. In contrast, on the SPIDER dataset, we observed notable qualitative performance degradation, characterized by fragmented gland boundaries, increased FPs in stromal regions, and reduced sensitivity to poorly differentiated glandular structures. Quantitative evaluation was not performed, as pixel-level gland annotations are not available for these datasets. The observed performance degradation is therefore reported qualitatively and is attributed to severe domain shift, including lower image quality, pronounced staining heterogeneity, and higher morphological variability inherent to that specific cohort, highlighting the challenges of cross-domain generalization in histopathology.[48,49]

Discussion

We developed and validated a novel weakly supervised teach-er–student framework for multi-class gland segmentation in CRC histopathology. The core of our approach lies in an EMA-stabilized teacher network, which employs confidence-based filtering and an adaptive fusion strategy to iteratively refine pseudo-masks, thereby guiding the student network with increasingly reliable supervision. Notably, the competitive results on the GlaS benchmark indicate that access to high-quality annotations allows our framework to further narrow the performance gap with fully supervised methods. In clinical settings, where dense, pixel-level annotation remains a major bottleneck, our framework offers practical benefits by substantially reducing annotation requirements while maintaining strong segmentation performance. Furthermore, its ability to generalize to TCGA-COAD and TCGA-READ without additional fine-tuning underscores its potential for multi-center application, where variations in staining and scanning

protocols are common.

However, the performance drop on SPIDER highlights the well-known challenge of domain generalization in computational pathology.[48,49] While our method generalizes well to TCGA-COAD and TCGA-READ domains (similar domains), SPIDER represents a severe domain shift that likely requires explicit domain adaptation techniques. Future work will focus on incorporating advanced domain adaptation strategies to improve broader cross-institutional generalization.[48,49] Additionally, we plan to extend the framework to other adenocarcinoma types, such as prostate, breast, and lung cancers, where glandular segmentation is equally critical for diagnosis and grading. By further reducing reliance on manual annotations while maintaining high segmentation fidelity, our framework offers a scalable and practical pathway toward wider adoption of computational pathology tools in clinical workflows.

Limitations

Despite the promising performance of the proposed framework, several limitations merit consideration. First, the OSU-WMC dataset lacks patient-level clinical metadata, precluding clinicopathologic correlation analyses. While this does not affect the technical validity of pixel-wise segmentation evaluation, it limits assessment of downstream prognostic or clinical utility. Second, although results are reported as mean ± standard deviation across independent runs, additional statistical measures, e.g., confidence intervals, will be investigated in future work. Third, the performance degradation observed on the SPIDER dataset highlights the impact of severe domain shift; addressing this limitation will likely require explicit domain adaptation or stain normalization strategies. Finally, while the proposed framework substantially reduces annotation burden, it still relies on limited sparse expert annotations. Achieving fully annotation-free segmentation remains an open and important direction for future research.

Conclusions

In this study, we introduced a novel weakly supervised teacher–student framework for multi-class gland segmentation in CRC histopathology, specifically designed to overcome the critical bottleneck and demand for extensive pixel-level annotations. By leveraging an EMA-stabilized teacher network, our framework efficiently utilizes sparse annotations, progressively refining pseudo-labels through confidence-based filtering and adaptive GT fusion. Comprehensive evaluation demonstrates that our framework achieved performance competitive with state-of-the-art methods in both weakly and fully supervised settings. Furthermore, the model exhibits strong in-house performance and robust generalization to external cohorts, including TCGA-COAD and TCGA-READ. While performance limitations on SPIDER highlight challenges under extreme domain shift, overall this work establishes an annotation-efficient paradigm that directly addresses a fundamental impediment in computational pathology. By substantially reducing reliance on costly manual curation while maintaining high segmentation fidelity, the proposed framework offers a practical, translatable solution to accelerate the adoption of automated diagnostic tools in clinical workflows.

Acknowledgments

None.

Conflict of interest

The authors declare no competing interests.

Author contributions

Leadership, experimental design, data analysis, figure and table preparation, manuscript drafting (HK), funding acquisition, and writing—review and editing (WC, MKKN). All authors have approved the final version and publication of the manuscript.

Ethical statement

The use of the in-house OSUWMC dataset was approved by the Institutional Review Board of The Ohio State University Wexner Medical Center (IRB No. 2018C0098). Written informed consent was obtained from all patients or was waived by the IRB due to the retrospective nature of the study. Public datasets (TCGA and SPIDER) were used in compliance with their respective data usage agreements and ethical guidelines and do not require additional institutional approval. All procedures were performed in accordance with the ethical standards of the Declaration of Helsinki (as revised in 2024).

Data sharing statement

The in-house OSUWMC dataset used in this study is available upon reasonable request by contacting the corresponding author, Hikmat Khan, at Hikmat.Khan@osumc.edu. All code was implemented in Python using PyTorch as the primary deep-learning library. The complete pipeline for processing WSIs, as well as training and evaluating the deep-learning models, will be available at: https://github.com/hikmatkhan/gland-segmentation-teacher-student.

References

[1] Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 2019;14(2):89–103. doi:10.5114/pg.2018.81072, PMID:31616522.
[2] Kim BH, Kim JM, Kang GH, Chang HJ, Kang DW, Kim JH, *et al*. Standardized Pathology Report for Colorectal Cancer, 2nd Edition. J Pathol Transl Med 2020;54(1):1–19. doi:10.4132/jptm.2019.09.28, PMID:31722452.
[3] Rechsteiner A, Dietrich D, Varga Z. Prognostic relevance of mixed histological subtypes in invasive breast carcinoma: a retrospective analysis. J Cancer Res Clin Oncol 2023;149(8):4967–4978. doi:10.1007/s00432-022-04443-x, PMID:36310301.
[4] Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, *et al*. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. Eur Urol 2016;69(3):428–435. doi:10.1016/j.eururo.2015.06.046, PMID:26166626.
[5] Stolnicu S, Park KJ, Kiyokawa T, Oliva E, McCluggage WG, Soslow RA. Tumor Typing of Endocervical Adenocarcinoma: Contemporary Review and Recommendations From the International Society of Gynecological Pathologists. Int J Gynecol Pathol 2021;40(Suppl 1):S75–S91. doi:10.1097/PGP.0000000000000751, PMID:33570865.
[6] Fleming M, Ravula S, Tatishchev SF, Wang HL. Colorectal carcinoma: Pathologic aspects. J Gastrointest Oncol 2012;3(3):153–173. doi:10.3978/j.issn.2078-6891.2012.030, PMID:22943008.
[7] Chen H, Qi X, Yu L, Dou Q, Qin J, Heng PA. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Med Image Anal 2017;36:135–146. doi:10.1016/j.media.2016.11.004, PMID:

27898306.

[8] Wen Z, Feng R, Liu J, Li Y, Ying S. GCSBA-Net: Gabor-Based and Cascade Squeeze Bi-Attention Network for Gland Segmentation. IEEE J Biomed Health Inform 2021;25(4):1185–1196. doi:10.1109/JBHI.2020.3015844, PMID:32780703.

[9] Wang H, Xian M, Vakanski A. TA-Net: Topology-Aware Network for Gland Segmentation. IEEE Winter Conf Appl Comput Vis 2022;2022:3241–3249. doi:10.1109/wacv51458.2022.00330, PMID:35509894.

[10] Xie Y, Lu H, Zhang J, Shen C, Xia Y. Deep Segmentation-Emendation Model for Gland Instance Segmentation. In: Shen D, Liu T, Staib LH, Essert C, Yap PT, Peters TF, *et al* (eds). Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science, vol 11764. Cham: Springer; 2019:469–477. doi:10.1007/978-3-030-32239-7_52.

[11] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, *et al*. Microsoft COCO: Common Objects in Context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds). Computer Vision – ECCV 2014. Lecture Notes in Computer Science, vol 8693. Cham: Springer; 2014:740–755. doi:10.1007/978-3-319-10602-1_48.

[12] Han C, Lin J, Mai J, Wang Y, Zhang Q, Zhao B, *et al*. Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Med Image Anal 2022;80:102487. doi:10.1016/j.media.2022.102487, PMID:35671591.

[13] Chang YT, Wang Q, Hung WC, Piramuthu R, Tsai YH, Yang MH. Weakly-supervised semantic segmentation via sub-category exploration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. Piscataway (NJ): IEEE; 2020. p. 8988–8997. doi:10.1109/CVPR42600.2020.00901.

[14] Lee J, Yi J, Shin C, Yoon S. BBAM: Bounding box attribution map for weakly supervised semantic and instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 19–25; Nashville, TN, USA. Piscataway (NJ): IEEE; 2021. p. 2643–2651. doi:10.1109/CVPR46437.2021.00267.

[15] Zhang X, Zhu L, He H, Jin L, Lu Y. Scribble hides class: promoting scribble-based weakly-supervised semantic segmentation with its class label. Proceedings of the AAAI Conference on Artificial Intelligence 2024;38(7):7332–7340. doi:10.1609/aaai.v38i7.28563.

[16] Laradji I, Rodriguez P, Mañas O, Lensink K, Law M, Kurzman L, *et al*. A weakly supervised consistency-based learning method for COVID-19 segmentation in CT images. In: Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV); 2021 Jan 3–8; Waikoloa, HI, USA. Piscataway (NJ): IEEE; 2021. p. 2452–2461. doi:10.1109/WACV48630.2021.00250.

[17] Ahn J, Kwak S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018; Salt Lake City, UT, USA. p. 4981–4990. doi:10.1109/CVPR.2018.00523.

[18] Jiang PT, Han LH, Hou Q, Cheng MM, Wei Y. Online Attention Accumulation for Weakly Supervised Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell 2022;44(10):7062–7077. doi:10.1109/TPAMI.2021.3092573, PMID:34170821.

[19] Chen Z, Wang T, Wu X, Hua XS, Zhang H, Sun Q. Class re-activation maps for weakly-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. Piscataway (NJ): IEEE; 2022. p. 969–978. doi:10.1109/CVPR52688.2022.00104.

[20] Chen Z, Wang T, Wu X, Hua XS, Zhang H, Sun Q. Class re-activation maps for weakly-supervised semantic segmentation. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. Piscataway (NJ): IEEE; 2022. p. 959–968. doi:10.1109/CVPR52688.2022.00104.

[21] Feng S, Chen J, Liu Z, Liu W, Wang Z, Lan R, *et al*. Mining gold from the sand: weakly supervised histological tissue segmentation with activation relocalization and mutual learning. In: Linguraru MG, Dou Q, Feragen A, Giannarou S, Glocker B, Lekadir K, *et al* (eds). Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Lecture Notes in Computer Science, vol 15008. Cham: Springer; 2024:414–423. doi:10.1007/978-3-031-72111-3_39.

[22] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway (NJ): IEEE; 2016. p. 2921–2929. doi:10.1109/CVPR.2016.319.

[23] Wang Y, Zhang J, Kan M, Shan S, Chen X. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. Piscataway (NJ): IEEE; 2020. p. 12272–12281. doi:10.1109/CVPR42600.2020.01229.

[24] Qin J, Wu J, Xiao X, Li L, Wang X. Activation modulation and recalibration scheme for weakly supervised semantic segmentation. Proceedings of the AAAI Conference on Artificial Intelligence 2022;36(2):2117–2125. doi:10.1609/aaai.v36i2.20108.

[25] Chen Z, Tian Z, Zhu J, Li C, Du S. C-CAM: causal CAM for weakly supervised semantic segmentation on medical image. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. Piscataway (NJ): IEEE; 2022. p. 11666–11675. doi:10.1109/CVPR52688.2022.01138.

[26] Sirinukunwattana K, Pluim JPW, Chen H, Qi X, Heng PA, Guo YB, *et al*. Gland segmentation in colon histology images: The glas challenge contest. Med Image Anal 2017;35:489–502. doi:10.1016/j.media.2016.08.008, PMID:27614792.

[27] Nechaev D, Pchelnikov A, Ivanova E. SPIDER: A Comprehensive Multi-Organ Supervised Pathology Dataset and Baseline Models. arXiv 2025. doi:10.48550/arXiv.2503.02876.

[28] Liu Y, Lin M, Sang X, Bao G, Wu Y. Weakly Supervised Gland Segmentation Based on Hierarchical Attention Fusion and Pixel Affinity Learning. Bioengineering (Basel) 2025;12(9):992. doi:10.3390/bioengineering12090992, PMID:41007236.

[29] Feng S, Wang H, Han C, Liu Z, Zhang H, Lan R, Pan X. Weakly supervised gland segmentation with class semantic consistency and purified labels filtration. Proceedings of the AAAI Conference on Artificial Intelligence 2025;39(3):2987–2995. doi:10.1609/aaai.v39i3.32306.

[30] Cheng H, Yu C, Zhang Y, Li B, Huang W, Zhang C. Glandular tissue segmentation based on EMA-Swin UNet model. Authorea 2025. doi:10.22541/au.173641600.01007870/v1.

[31] Li Y, Yu Y, Zou Y, Xiang T, Li X. Online easy example mining for weakly-supervised gland segmentation from histology images. Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S (eds). Lecture Notes in Computer Science, vol 13434. Cham: Springer; 2022:619–628. doi:10.1007/978-3-031-16440-8_55.

[32] Chen YJ, Hu X, Shi Y, Ho TY. AME-CAM: attentive multiple-exit CAM for weakly supervised segmentation on MRI brain tumor. Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. In: Greenspan H, Madabhushi A, Mousavi P, Salcudean S, Duncan J, Syeda-Mahmood T, *et al* (eds). Lecture Notes in Computer Science, vol 14220. Cham: Springer; 2023:203–213. doi:10.1007/978-3-031-43907-0_17.

[33] Zhong L, Wang G, Liao X, Zhang S. HAMIL: High-Resolution Activation Maps and Interleaved Learning for Weakly Supervised Segmentation of Histopathological Images. IEEE Trans Med Imaging 2023;42(10):2912–2923. doi:10.1109/TMI.2023.3269798, PMID:37093729.

[34] Du W, Huo Y, Zhou R, Sun Y, Tang S, Zhao X, *et al*. Consistency label-activated region generating network for weakly supervised medical image segmentation. Comput Biol Med 2024;173:108380. doi:10.1016/j.compbiomed.2024.108380, PMID:38555701.

[35] Lee J, Kim E, Yoon S. Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA. Piscataway (NJ): IEEE; 2021. p. 4070–4078. doi:10.1109/CVPR46437.2021.00406.

[36] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. In: Navab N, Hornegger J, Wells W, Frangi A (eds). Lecture Notes in Computer Science, vol 9351. Cham: Springer; 2015:234–241. doi:10.1007/978-3-319-24574-4_28.

[37] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39(12):2481–2495. doi:10.1109/TPAMI.2016.2644615, PMID:28060704.

[38] Valanarasu JMJ, Oza P, Hacihaliloglu I, Patel VM. Medical Transformer: gated axial-attention for medical image segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. In: de Bruijne M, Cattin PC, Cotin S, Padoy N, Speidel S, Zheng Y, *et al* (eds). Lecture Notes in Computer Science, vol 12901. Cham: Springer; 2021:39–49. doi:10.1007/978-3-030-87193-2_4.

[39] Sun G, Pan Y, Kong W, Xu Z, Ma J, Racharak T, *et al*. DA-TransUNet: integrating spatial and channel dual attention with transformer U-net for medical image segmentation. Front Bioeng Biotechnol 2024;12:1398237. doi:10.3389/fbioe.2024.1398237, PMID:38827037.

[40] Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, *et al*. Attention U-Net: learning where to look for the pancreas. In: Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018); 2018 Jul 16–18; Amsterdam, The Netherlands.

[41] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018) 2018;11045:3–11. doi:10.1007/978-3-030-00889-5_1, PMID:32613207.

[42] Valanarasu JMJ, Sindagi VA, Hacihaliloglu I, Patel VM. KiU-Net: towards accurate segmentation of biomedical images using over-complete representations. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. In: Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, *et al* (eds). Lecture Notes in Computer Science, vol 12264. Cham: Springer; 2020:484–493. doi:10.1007/978-3-030-59719-1_36.

[43] Jha D, Smedsrud PH, Riegler MA, Johansen D, De Lange T, Halvorsen P. ResUNet++: an advanced architecture for medical image segmentation. In: Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM); 2019 Dec 9–11; San Diego, CA, USA. Piscataway (NJ): IEEE; 2019. p. 225C–2255. doi:10.1109/ISM46123.2019.00049.

[44] Chen B, Liu Y, Zhang Z, Lu G, Kong AWK. TransAttUnet: multi-level attention-guided U-Net with transformer for medical image segmentation. IEEE Trans Emerg Top Comput Intell 2024;8(1):55–68. doi:10.1109/TETCI.2023.3309626.

[45] Loshchilov I, Hutter F. Decoupled weight decay regularization. In: Proceedings of the International Conference on Learning Representations (ICLR 2019) [preprint]; 2019 May 6–9; New Orleans, LA, USA.

[46] Wang Z, Wang P, Liu K, Wang P, Fu Y, Lu CT. A comprehensive survey on data augmentation. IEEE Trans Knowl Data Eng 2026;38(1):47–66. doi:10.1109/TKDE.2025.3622600.

[47] Molina JM, Llerena JP, Usero L, Patricio MA. Advances in instance segmentation: technologies, metrics and applications in computer vision. Neurocomputing 2025;625:129584. doi:10.1016/j.neucom.2025.129584.

[48] Khan H, Zaidi SF, Shah PM, Balakrishnan K, Khan R, Waqas M, *et al*. MorphGen: Morphology-guided representation learning for robust single-domain generalization in histopathological cancer classification. arXiv 2025. doi:10.48550/arXiv.2509.00311.

[49] Jahanifar M, Raza M, Xu K, Vuong TTL, Jewsbury R, Shephard A, *et al*. Domain generalization in computational pathology: survey and guidelines. ACM Comput Surv 2025;57(11) 285:1–37. doi:10.1145/3724391.