Cross-Domain Few-Shot Segmentation of Histopathology Images with Mamba

Joanna Szolomicka[0000−0001−5581−2952] and Urszula Markowska-Kaczmar[0000−0001−7606−3057]

Department of Artificial Intelligence, Wrocław University of Science and Technology
Wyb. Wyspianskiego 27, Wroclaw, Poland

Abstract. Accurate segmentation of histopathology images is critical for cancer diagnosis, characterization, and treatment planning. However, manual annotation at the cellular level is labor-intensive and requires expert knowledge, while variability across tissue types, staining protocols, and scanning devices limits the generalization of conventional segmentation models. In this work, we propose a few-shot learning framework for cross-domain histopathology segmentation, leveraging the selective state-space model *Mamba* to efficiently capture long-range dependencies in densely packed cellular images. Our approach enables the model to generalize to new tissue types or laboratory domains using only a few annotated examples. Additionally, we incorporate explainable AI techniques to provide interpretability of segmentation predictions, supporting clinical trust. We validate our method on the PanNuke dataset, providing a few-shot segmentation split, and demonstrate that our Mamba-based model outperforms existing few-shot and baseline methods. This work advances practical and efficient automated histopathology segmentation under limited supervision and domain variability.

Keywords: few-shot · histopathology image segmentation · Mamba.

1 Introduction

Segmentation of medical images is a cornerstone of modern diagnostic practice, enabling automated quantification and analysis of anatomical structures. However, obtaining accurate segmentation masks is expensive and time-consuming, as it requires expert annotation. Among medical imaging modalities, histopathology is particularly important for cancer detection, characterization, and treatment planning. These images reveal tissue morphology at the cellular level and provide critical diagnostic information.

Histopathology segmentation is inherently challenging. Tissue appearance varies across cancer types (e.g., breast, colon, cervix), cellular composition, and laboratory staining or scanning protocols. Moreover, images contain large numbers of densely packed cells with subtle morphological differences, making boundary delineation difficult. Since pixel-level annotation requires specialized expertise, large fully annotated datasets are rarely available.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

A key challenge is cross-domain generalization. Models trained on one tissue type or laboratory setting often fail to generalize to others due to differences in color distribution, texture, and cellular morphology. Such domain shifts limit the robustness of automated systems in real clinical environments.

Few-shot learning offers a promising solution by enabling models to adapt to new tissue types or domains using only a small number of annotated examples. While few-shot segmentation has shown success in natural images and medical modalities such as MRI or CT—where images typically contain a small number of large structures—histopathology images present additional difficulty due to dense cellular arrangements, high intra-class variability, and fine-grained details.

Recent works have introduced the selective state-space model *Mamba* [4, 19], which captures long-range dependencies with linear computational complexity, offering a scalable alternative to attention-based architectures. In this work, we integrate Mamba into a few-shot learning framework tailored for histopathology segmentation. Our contributions are threefold:

1. We propose a Mamba-based few-shot segmentation model for histopathology images that outperforms existing approaches.
2. We incorporate explainable AI (XAI) techniques to enhance interpretability and clinical transparency.
3. We introduce a few-shot segmentation split of the PanNuke dataset [2, 3] to support benchmarking in cross-domain histopathology segmentation. We additionally evaluate our model on the MoNuSeg [9] and TNBC Nuclei Segmentation [13] datasets collected from different laboratories with varying staining characteristics, demonstrating generalization beyond the source dataset.

By addressing the dual challenges of limited supervision and cross-domain variability, our work aims to advance practical and reliable automated segmentation tools for histopathology, reducing the annotation burden on experts while supporting clinical decision-making.

2 Related works

Mamba Segmentation models can be divided into two categories: conventional and few-shot approaches. In this work, we explore methods that utilize Mamba, a state-space model capable of capturing long-range dependencies with linear computational complexity. This capability is particularly important in histopathology image analysis, as the appearance of a nucleus often depends on the surrounding cellular context. Mamba provides an alternative to attention-based models, as it has proven effective in multiple computer vision tasks [19, 11] while requiring significantly lower computational resources. In [10], the authors propose SwinUMamba, a medical image segmentation model that leverages an ImageNet-pretrained backbone and outperforms CNNs, ViTs, and non-pretrained Mamba models. MambaVesselNet++ [18] combines CNNs for low-level feature extraction with Mamba for long-range modeling in a U-shaped architecture across multiple modalities. Vision Mamba has also been applied to

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

histopathology, outperforming ViTs in classification and emulating pathologists' cancer detection [12].

Few-shot segmentation Few-shot methods allow models to generalize to new tissue types and staining with minimal annotations [16]. HMNet [17] integrates Mamba modules into a few-shot prototypical network for natural image segmentation. Few-shot segmentation of cells in histopathology images has received limited attention, as most existing few-shot methods focus either on segmentation of large structures in MRI or CT scans [1, 15] or on classification tasks in histopathology [6, 5, 14]. To address this gap, we propose a Mamba-based few-shot learning framework for histopathology cell segmentation that includes an explainability component.

3 Problem to be Solved

We consider the problem of cross-domain few-shot segmentation in histopathology image analysis. Let the source domain consist of n datasets

$$S = \{S_1, S_2, \ldots, S_n\},$$

where each dataset $S_i = \{(x_j, y_j)\}_{j=1}^{|S_i|}$ contains microscopy images x_j and their corresponding segmentation masks y_j. These datasets represent well-annotated source domains.

The target domain $T = \{(x_j, y_j)\}_{j=1}^{|T|}$ is drawn from a different data distribution and contains only K annotated samples per class. The goal is to learn a segmentation function capable of transferring knowledge from the source domains to the target domain under limited supervision.

During episodic training, a query image x_q and a support set:

$$\mathcal{S} = \{(x_s^k, y_s^k)\}_{k=1}^{K}$$

are sampled from the source domain. The model learns to predict the segmentation mask of the query image conditioned on the support set:

$$\hat{y}_q = f(x_q \mid \mathcal{S}).$$

During inference, the model is provided with K annotated support examples from the target domain for previously unseen or sparsely annotated classes. The objective is to accurately segment query images from the target domain using only this limited supervision.

This formulation enables transfer of segmentation knowledge from well-annotated microscopy datasets to new biomedical domains with minimal labeling effort, addressing annotation scarcity and domain variability in histopathology. The main challenge lies in modeling fine-grained cellular boundaries under significant domain shift and extreme supervision constraints, where both morphological variability and small object size limit the effectiveness of conventional segmentation models.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

4 Proposed Method

In this work, we introduce *SwinUHMnet* (Fig. 1), a hybrid architecture tailored for few-shot histopathology image segmentation. The model addresses two central biomedical challenges: (i) reliable support-conditioned adaptation under limited supervision and (ii) multiscale modeling of small, densely packed cellular structures.

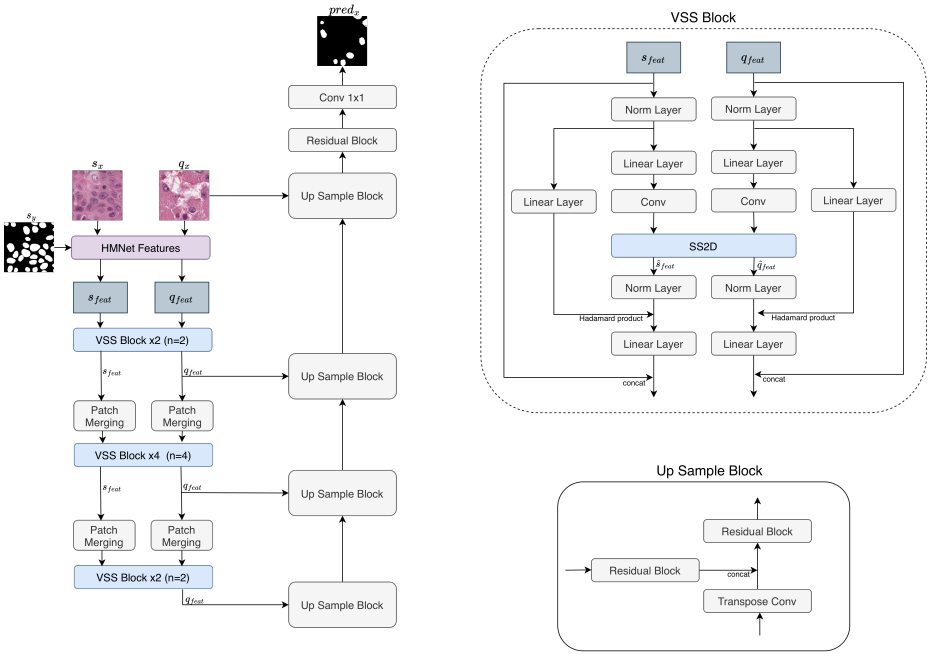

Fig. 1: Architecture of the few-shot SwinUHMnet model. The input to the model consists of query image q_x, support image s_x and support mask s_y. The output of the model is predicted query segmentation mask $pred_x$.

The proposed model follows a support-conditioned encoder–decoder architecture, as illustrated in Fig. 1. The network processes a support image s_x with its corresponding mask s_y, together with a query image q_x, and predicts the segmentation mask for the query. The HMnet Features module extracts backbone representations from both support and query images.

Using the support mask, foreground-aware prototype features are constructed and integrated with support feature maps. In parallel, query features are augmented with support-derived prototype and prior correlation information. As a result, the outputs of this stage are support-conditioned feature maps s_{feat} and q_{feat}. Then, hierarchical encoding is performed by the Mamba structure [4].

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

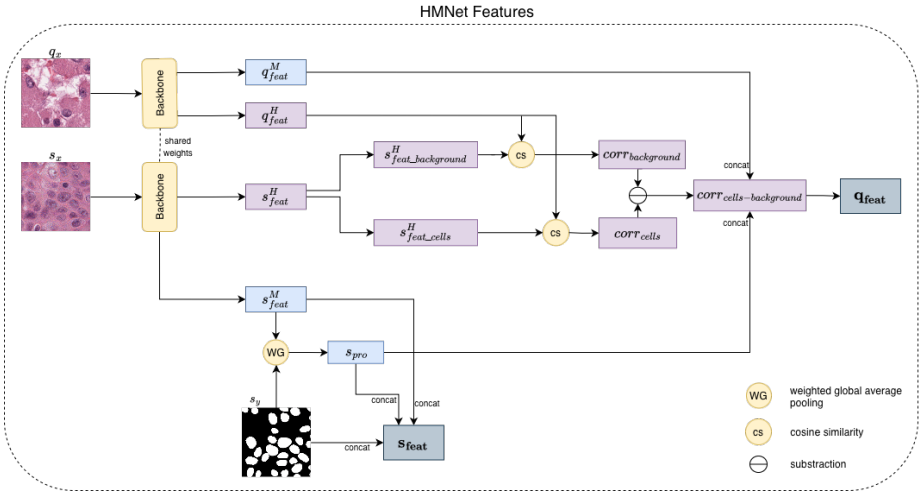

Fig. 2: The HMnet Features module is the first component of SwinUHMnet model. It takes the query image q_x, the support image s_x and the support segmentation mask s_y as inputs and returns the support features s_{feat} and the query features q_{feat}.

Mamba is a state-space model (SSM) that enables efficient modeling of long-range dependencies with linear computational complexity. The encoder consists of three hierarchical stages composed of Visual State Space (VSS) blocks arranged in a 2-4-2 configuration. At each stage, both support and query features are processed in parallel.

Within each VSS block, features are first normalized and linearly projected. Convolutional transformations are then applied before the Selective Scan 2D (SS2D) module performs directional state-space scanning, enabling spatial information propagation across the entire feature map. The Selective Scan 2D mechanism, presented in Fig. 3 extends one-dimensional state-space modeling to two-dimensional feature maps. Given an input feature map, SS2D unfolds it into directional sequences along four scanning paths: left-to-right, right-to-left, top-to-bottom, and bottom-to-top. Each sequence is processed independently by a state-space operator, and the outputs are aggregated to reconstruct a refined two-dimensional representation.

This directional scanning enables global spatial information propagation without explicit pairwise similarity computation. Such long-range modeling is particularly beneficial in cell segmentation, where small cellular structures are spatially distributed and contextual relationships between distant regions improve prediction consistency. Element-wise modulation and linear refinement complete the block operation.

Between stages, Patch Merging reduces spatial resolution while increasing channel dimensionality, forming multi-scale representations. This hierarchical

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

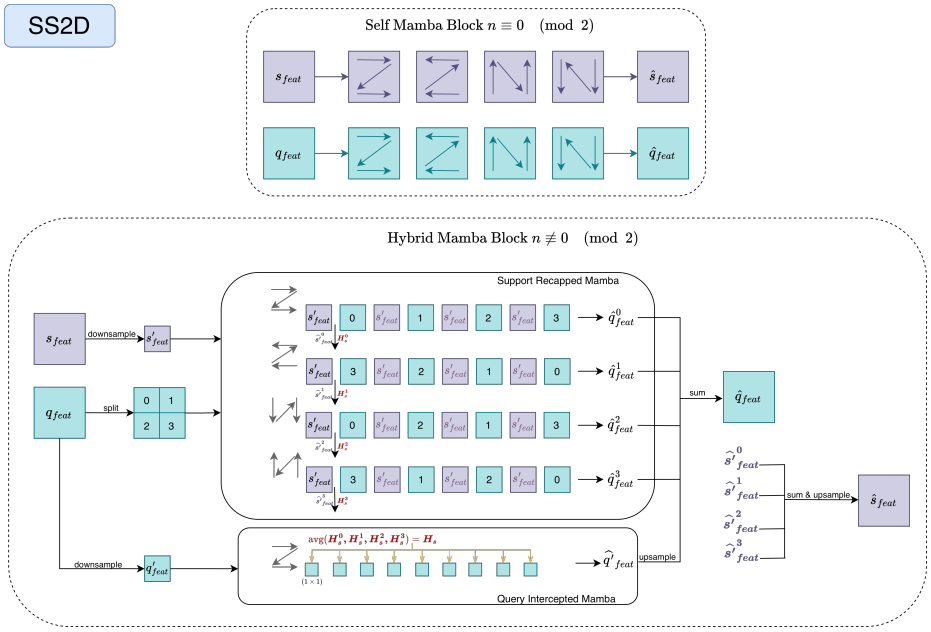

Fig. 3: The SS2D function [17] is a main component of the VSS blocks. In even-indexed VSS blocks, SS2D operates as a Self-Mamba Block, while in odd-indexed VSS blocks, it is implemented as a Hybrid Mamba Block.

processing progressively enlarges the receptive field while maintaining support – query conditioning throughout the encoder.

Only the encoded query representation is propagated to the decoder. The decoder is composed of sequential Up Sample Blocks, each consisting of a transpose convolution for spatial upsampling followed by residual refinement. Skip connections from corresponding encoder stages are concatenated with upsampled features to recover fine-grained spatial details.

After the final upsampling stage, a residual block and a 1×1 convolution layer produce the final segmentation mask for the query image.

Finally, a convolutional decoder with structured skip connections reconstructs a high-resolution segmentation map, preserving fine-grained cellular boundaries critical in histopathology.

The following subsections describe each component of the framework in detail.

4.1 HMNet Features

The HMNet Features module constructs support-conditioned representations for both support and query images. It takes as input the query image q_x, the support image s_x, and the corresponding support mask s_y.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

A shared pretrained backbone $B(\cdot)$ extracts multi-scale features:

$$q_{feat} = B(q_x), \quad s_{feat} = B(s_x).$$

Support Prototype Construction. Low-level support features s^M_{feat} are combined with the support mask s_y to compute a foreground prototype using weighted global average pooling:

$$s_{pro} = GAP(s^M_{feat} \odot s_y),$$

where $\odot$ denotes element-wise multiplication.

The final support representation forwarded to the encoder is defined as:

$$s_{feat} = Concat(s^M_{feat}, s_{pro}, s_y).$$

Prior Correlation Maps. High-level support features s^H_{feat} are used to construct foreground and background prototypes:

$$s^H_{cells} = GAP(s^H_{feat} \odot s_y), \quad s^H_{background} = GAP(s^H_{feat} \odot (1 - s_y)).$$

Spatial correlation maps are computed via cosine similarity between these prototypes and the high-level query features:

$$\begin{aligned} corr_{cells} &= \cos(s^H_{cells}, q^H_{feat}), \\ corr_{background} &= \cos(s^H_{background}, q^H_{feat}), \\ corr &= corr_{cells} - corr_{background}. \end{aligned}$$

The final query representation is constructed as:

$$q_{feat} = Concat(q^H_{feat}, s_{pro}, corr),$$

where *corr* provides a spatial prior indicating the likelihood of support-class presence in the query image.

4.2 Mamba Encoder

The Mamba encoder consists of three hierarchical stages containing 2, 4, and 2 Visual State Space (VSS) blocks, respectively. Both support and query features are processed in parallel at each stage. Between stages, Patch Merging reduces spatial resolution while increasing channel dimensionality, forming a multi-scale hierarchy.

Visual State Space Block. Each VSS block contains normalization, linear projection, convolutional transformation, and the SS2D state-space operator. The SS2D module performs directional spatial scanning to propagate contextual information across the feature map.

Self-Mamba Blocks. In even-indexed VSS blocks, support and query features are processed independently using SS2D. This operation refines representations while preserving their respective structures.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

Hybrid Mamba Blocks. In odd-indexed VSS blocks, a Hybrid Mamba design is employed, consisting of Support Recapped Mamba (SRM) and Query Intercepted Mamba (QIM). In SRM, support features are spatially downsampled, while query features are partitioned into patches. Directional scanning is performed jointly, allowing support information to guide query feature refinement. To mitigate support information decay during sequential scanning, support features are periodically reintroduced into the hidden state.

In QIM, query features are scanned using hidden states initialized from support-enhanced representations. This mechanism suppresses excessive query self-reinforcement and encourages stronger incorporation of support-derived signals.

The outputs from directional scans are aggregated to produce enhanced support and query representations. By alternating self and hybrid blocks across stages, the encoder progressively strengthens support-conditioned adaptation at multiple scales.

4.3 Datasets and Experimental Setup

PanNuke. Experiments were conducted on the PanNuke dataset [2, 3], which contains histopathology images with instance-level cell segmentation annotations across 19 tissue types and five cell categories: neoplastic, non-neoplastic epithelial, inflammatory, connective, and dead. All images have a spatial resolution of 256×256 pixels. Following the official tissue partitioning, the training set includes images from bile duct, esophagus, head and neck, kidney, liver, lung, ovary, pancreas, prostate, testis, stomach, thyroid, and uterus. The validation set contains images from skin tissue, while the test set consists of images from adrenal gland, bladder, breast, cervix, and colon.

*Few-Shot Splits.*We propose a few-shot protocol in which each query–support pair shares at least one common cell category. To ensure robustness, five independent dataset splits were created. For each split, 70% of the images from the selected tissue types are used as query images, while the remaining 30% form the candidate support pool. During episodic training, support examples are sampled from this pool such that at least one cell category overlaps with the query image. This setting ensures meaningful support conditioning while preserving cross-image variability.

Data Augmentation. During training, data augmentations are applied to the query images and their corresponding masks, including random rescaling, horizontal flipping, rotation, cropping, and Gaussian noise perturbations.

Cross-Dataset Evaluation. To evaluate cross-dataset generalization, the trained models are additionally tested on MoNuSeg [9] and TNBC Nuclei Segmentation [13] datasets. MoNuSeg contains 30 histopathology images from seven different organs, each with a resolution of 1000×1000 pixels. For MoNuSeg, support images are randomly selected from the same organ as the query image. The TNBC dataset consists of 50 breast cancer histopathology images with a resolution of 512×512 pixels. Since all samples originate from breast cancer tissue, support images are randomly selected without organ constraints.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

4.4 Implementation Details

We trained SwinUHMNet on the PanNuke dataset and compared its performance with HMNet and SwinUMamba. All models were trained from scratch on the same training splits. Each experiment was conducted across five independent data splits. Due to incompatibility of the publicly available pretrained checkpoint, SwinUMamba was trained from scratch in our setting. We denote this variant as SwinUMamba*.

Backbone Networks. For fair comparison, ResNet101 [7] was used as the backbone for all models unless otherwise stated. Additionally, for SwinUHMNet we evaluated a variant with a ViT-small-patch8 backbone pretrained on pathology images using DINO [8]. The patch size was set to 8.

Model Configuration. To mitigate overfitting in SwinUHMNet, we employed a lightweight encoder configuration with VSS blocks arranged as $(2, 4, 2)$ and a maximum hidden dimension of 384. Mamba parameters were set to $d_{state} = 8$ and $expand = 1.5$. In contrast, SwinUMamba* uses a deeper configuration with VSS blocks $(2, 2, 9, 2)$, a maximum hidden dimension of 768, and Mamba parameters $d_{state} = 16$ and $expand = 2$. HMNet and SwinUHMNet were trained in a one-shot setting ($K = 1$), following the few-shot protocol described earlier.

Loss function. SwinUHMnet was trained using binary cross-entropy and dice loss.

Hardware. All experiments were performed on an NVIDIA A100 GPU.

Data. For evaluation on the Monuseg and TNBC datasets, images were divided into 256×256 tiles with a 64-pixel overlap, which were then fed into the model. The resulting segmentation masks were merged by taking the maximum value in the overlapping regions.

5 Results

5.1 Evaluation Protocol

All models were evaluated in a one-shot setting ($K = 1$). For each query image, a single support image was sampled according to the predefined few-shot split protocol.

Segmentation performance was measured using pixel-wise Intersection over Union (IoU) and F1-score. Metrics were computed independently for each tissue type and averaged across five data splits. Each split was used to train and evaluate an independent model. Unless otherwise stated, results are reported as mean ± standard deviation.

On the PanNuke dataset, evaluation was performed on unseen tissue types. For MoNuSeg and TNBC, models trained on PanNuke were directly evaluated without fine-tuning to assess cross-dataset generalization.

5.2 Quantitative analysis

Table 1 reports the mean IoU and F1 scores on the five PanNuke splits, averaged across splits. For each split, a separate model was trained and evaluated. We com-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

pare SwinUMamba*, HMNet, and two variants of SwinUHMNet with ResNet-101 (SwinUHMNet$_{\mathrm{R}}$) and ViT-small-patch8 (SwinUHMNet$_{\mathrm{V}}$) backbones.

SwinUHMNet consistently outperforms HMNet across all tissue types and achieves slightly better results than the non-few-shot SwinUMamba*, while maintaining competitive parameter efficiency. Although SwinUMamba has the smallest number of parameters, SwinUHMNet$_{\mathrm{R}}$ provides improved interpretability through support prototypes and prior correlation maps, which is particularly important in medical image analysis. We further compare ResNet-101 (pretrained on natural images) and ViT-small (pretrained on pathology images) as feature extractors. Both achieve similar segmentation performance; however, SwinUHMNet$_{\mathrm{R}}$ uses significantly fewer parameters. Despite domain mismatch, ResNet-101 effectively captures multi-scale local features, which are crucial for densely structured histopathology images. In contrast, ViTs focus more on global context, leading to comparable but not superior results. The improved performance of SwinUHMNet over HMNet can be attributed to its higher output resolution enabled by the upsampling decoder. Low within-tissue standard deviation indicates stable performance, while higher variability across tissues reflects differences in segmentation difficulty.

Table 1: IoU and F1-score results of the predicted segmentation masks, averaged across five splits of the PanNuke test dataset

	IoU / F1			
shots	-	1	1	1
number of parameters	65,799,220	80,586,696	83,088,899	104,611,715
model	SwinUMamba*	HMNet	SwinUHMnet	SwinUHMnet
backbone	ResNet101	ResNet101	ResNet101	ViT-small-patch8
adrenal gland	0.62 ± 0.01 / 0.77 ± 0.01	0.57 ± 0.00 / 0.72 ± 0.0	0.63 ± 0.01 / 0.77 ± 0.01	0.63 ± 0.01 / 0.78 ± 0.00
bladder	0.75 ± 0.01 / 0.86 ± 0.01	0.66 ± 0.01 / 0.80 ± 0.01	0.75 ± 0.01 / 0.86 ± 0.01	0.75 ± 0.01 / 0.86 ± 0.01
breast	0.62 ± 0.01 / 0.76 ± 0.01	0.57 ± 0.01 / 0.73 ± 0.01	0.64 ± 0.00 / 0.78 ± 0.00	0.64 ± 0.00 / 0.78 ± 0.00
cervix	0.65 ± 0.02 / 0.79 ± 0.02	0.59 ± 0.02 / 0.74 ± 0.01	0.70 ± 0.01 / 0.82 ± 0.00	0.71 ± 0.01 / 0.83 ± 0.00
colon	0.56 ± 0.01 / 0.72 ± 0.01	0.52 ± 0.01 / 0.69 ± 0.01	0.60 ± 0.01 / 0.75 ± 0.00	0.60 ± 0.01 / 0.75 ± 0.00
all	0.64 ± 0.06 / 0.78 ± 0.05	0.58 ± 0.05 / 0.74 ± 0.04	0.66 ± 0.05 / 0.80 ± 0.04	0.67 ± 0.05 / 0.80 ± 0.04

MoNuSeg & TNBC Table 2 presents IoU and F1 scores on seven MoNuSeg tissues and breast tissue from TNBC. SwinUHMNet outperforms SwinUMamba* and HMNet on most MoNuSeg tissues. For example, on stomach tissue, SwinUHMNet$_{\mathrm{R}}$ improves IoU by 0.13 over SwinUMamba* and shows greater stability than HMNet. On TNBC, all models achieve comparable performance. Overall,

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

Table 2: IoU and F1-score results for the predicted segmentation masks on the Monuseg and TNBC datasets

		IoU / F1			
	shots	-	1	1	1
	model	SwinUMamba*	HMNet	$\text{SwinUHMnet}_\text{R}$	$\text{SwinUHMnet}_\text{V}$
	backbone	ResNet101	ResNet101	ResNet101	ViT-small-patch8
Monuseg	stomach	0.40 ± 0.10 / 0.56 ± 0.11	0.52 ± 0.11 / 0.67 ± 0.10	0.53 ± 0.07 / 0.69 ± 0.06	0.37 ± 0.04 / 0.53 ± 0.04
	colon	0.59 ± 0.01 / 0.74 ± 0.01	0.55 ± 0.01 / 0.71 ± 0.01	0.61 ± 0.01 / 0.76 ± 0.01	0.62 ± 0.00 / 0.76 ± 0.00
	bladder	0.69 ± 0.01 / 0.82 ± 0.01	0.64 ± 0.01 / 0.78 ± 0.01	0.71 ± 0.01 / 0.83 ± 0.00	0.68 ± 0.01 / 0.81 ± 0.01
	kidney	0.62 ± 0.03 / 0.77 ± 0.03	0.56 ± 0.01 / 0.72 ± 0.01	0.58 ± 0.02 / 0.74 ± 0.01	0.62 ± 0.01 / 0.77 ± 0.01
	breast	0.56 ± 0.02 / 0.71 ± 0.02	0.55 ± 0.02 / 0.71 ± 0.01	0.58 ± 0.01 / 0.73 ± 0.01	0.57 ± 0.01 / 0.73 ± 0.01
	liver	0.62 ± 0.01 / 0.77 ± 0.01	0.59 ± 0.01 / 0.74 ± 0.01	0.64 ± 0.01 / 0.78 ± 0.00	0.64 ± 0.01 / 0.78 ± 0.01
	prostate	0.68 ± 0.00 / 0.81 ± 0.00	0.64 ± 0.01 / 0.78 ± 0.00	0.69 ± 0.00 / 0.82 ± 0.00	0.68 ± 0.00 / 0.81 ± 0.00
	all	0.60 ± 0.02 / 0.75 ± 0.02	0.58 ± 0.02 / 0.73 ± 0.02	0.62 ± 0.00 / 0.77 ± 0.00	0.61 ± 0.00 / 0.76 ± 0.00
TNBC	all	0.54 ± 0.03 / 0.70 ± 0.02	0.55 ± 0.02 / 0.71 ± 0.02	0.55 ± 0.02 / 0.71 ± 0.01	0.55 ± 0.01 / 0.71 ± 0.01

SwinUHMNet demonstrates consistent generalization across diverse histological structures.

(a) Query image q_x (b) Query mask q_y (c) SwinUMamba* segmentation mask (d) $\text{SwinUHMNet}_\text{R}$ segmentation mask (e) $\text{SwinUHMNet}_\text{V}$ segmentation mask (f) HMNet segmentation mask

Fig. 4: Qualitative comparisons on colon tissue from PanNuke dataset.

5.3 Qualitative analysis

Fig. 4 presents example segmentation mask predictions for a query image of colon tissue from the PanNuke test dataset generated by SwinUMamba*, HMNet,

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

SwinUHMNet$_{\mathrm{V}}$, and SwinUHMNet$_{\mathrm{R}}$. Predictions from HMNet exhibit lower resolution compared to the other models, although the spaces between cells, as observed in the target mask, are generally preserved. SwinUMamba* predictions show more connected cells, with nuclei appearing more complete in the predicted masks. SwinUHMNet combines these strengths, producing higher-resolution predictions in which individual cells are easier to distinguish while maintaining overall nuclei structure.

5.4 Explainability of SwinUHMnet

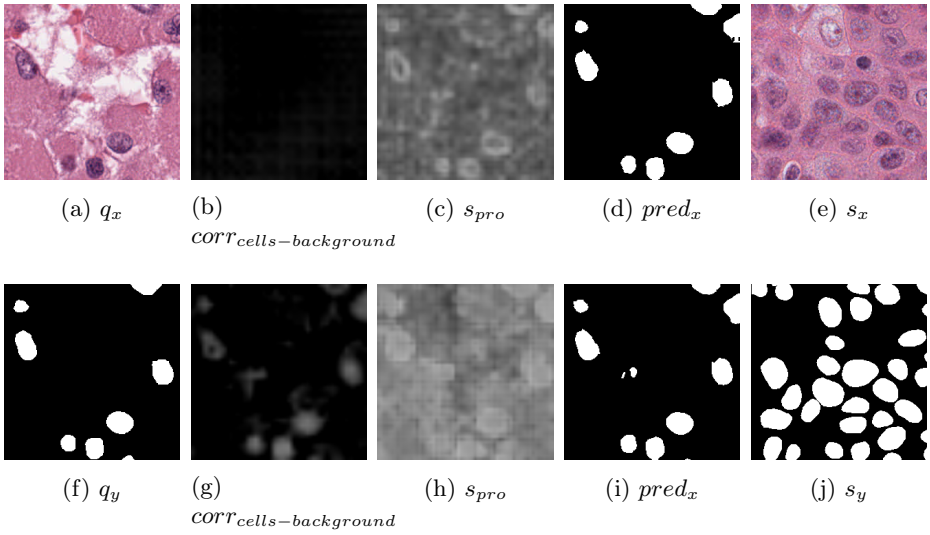

Fig. 5: Visualization of prototype activations and prior correlation on query image q_x from adrenal gland tissue: (a) query image q_x from adrenal gland with connective tissue cells (b) prior correlation map $corr_{cells-background}$ SwinUHMnet$_{\mathrm{R}}$ (c) heatmap of prototype s_{pro} activation on query features q^M_{feat} from SwinUHMnet$_{\mathrm{R}}$ (d) segmentation mask predicted by SwinUHMnet$_{\mathrm{R}}$ (e) support image s_x from cervix with connective tissue cells (f) query mask q_x (g) prior correlation map $corr_{cells-background}$ from SwinUHMnet$_{\mathrm{V}}$ (h) heatmap of prototype s_{pro} activation on query features q^M_{feat} from SwinUHMnet$_{\mathrm{V}}$ (i) segmentation mask predicted by SwinUHMnet$_{\mathrm{V}}$ (j) support mask s_y

Fig. 5 presents visualizations of the support prototype activation s_{pro} on query features q^M_{feat} and the prior correlation map $\mathrm{corr}_{\text{cells-background}}$ within the HMNetFeatures module. Support prototypes and prior correlation maps guide the model to focus on regions of the query image that resemble patterns observed in the support image.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

The heatmap of the support prototype s_{pro} highlights spatial locations in the query image that closely match the cell prototype extracted from the support image. For both SwinUHMNet$_{\text{R}}$ and SwinUHMNet$_{\text{V}}$, high-activation regions align well with true cell locations in the ground-truth mask. This improves interpretability, as it reveals not only where the model predicts cells, but also which regions contribute most to the decision.

Heatmaps produced by SwinUHMNet$_{\text{V}}$ are smoother and less contrasted than those from SwinUHMNet$_{\text{R}}$, reflecting the broader receptive field of ViT embeddings. In contrast, ResNet features emphasize local edges and fine-grained textures. The prior correlation maps show higher contrast with the ViT backbone, as global context enhances separation between cells and background. With ResNet, stronger sensitivity to local textures may cause background structures to resemble cells, resulting in lower contrast in the correlation maps.

6 Discussion and Future Work

The results demonstrate that incorporating an upsampling decoder together with support-guided feature modulation improves cell segmentation accuracy and output resolution without increasing computational cost. In contrast, directly increasing the output resolution in HMNet resulted in excessive memory consumption, highlighting the efficiency of the proposed design.

Performance variability across tissue types and datasets indicates that segmentation difficulty depends on cellular density and morphological characteristics. The visualization of support prototype activations and prior correlation maps in SwinUHMNet enhances model interpretability by revealing regions that influence predictions, which is particularly important for clinical applications.

Future work will investigate multi-shot settings to assess whether additional support examples further improve segmentation accuracy. We also plan to explore more lightweight architectures to reduce computational cost while maintaining performance.

Acknowledgments. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH, CI TASK) for providing computer facilities and support within computational grant no. PLG/2025/018026).

References

1. Dissanayake, T., et al., G.: Few-shot learning for medical image segmentation: A review and comparative study. ACM Comput. Surv. **58**(1) (Sep 2025). https://doi.org/10.1145/3746224, https://doi.org/10.1145/3746224
2. Gamper, J., et al., K.: Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In: European Congress on Digital Pathology. pp. 11–19. Springer (2019)
3. Gamper, J., et al., K.: Pannuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778 (2020)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18

4. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces (2024), https://arxiv.org/abs/2312.00752
5. Guo, Z., Xiong, C., et al., M.: Focus: Knowledge-enhanced adaptive visual compression for few-shot whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15590–15600 (June 2025)
6. Han, M., Qu, L., et al., Y.: Mscpt: Few-shot whole slide image classification with multi-scale and context-focused prompt tuning. IEEE Transactions on Medical Imaging **44**(9), 3756–3769 (2025). https://doi.org/10.1109/TMI.2025.3564976
7. He, K., et al., X.Z.: Deep residual learning for image recognition (2015), https://arxiv.org/abs/1512.03385
8. Kang, M., Song, H.e.a.: Benchmarking self-supervised learning on diverse pathology datasets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3344–3354 (June 2023)
9. Kumar, N., et al., V.: A multi-organ nucleus segmentation challenge. IEEE Transactions on Medical Imaging **39**(5), 1380–1391 (2020). https://doi.org/10.1109/TMI.2019.2947628
10. Liu, J., Yang, H., et al., H.Y.Z.: Swin-umamba: Mamba-based unet with imagenet-based pretraining (2024), https://arxiv.org/abs/2402.03302
11. Liu, Y., Tian, Y., et al., Y.Z.: Vmamba: Visual state space model (2024), https://arxiv.org/abs/2401.10166
12. Nasiri-Sarvi, A., et al., T.: Vim4path: Self-supervised vision mamba for histopathology images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 6894–6903 (June 2024)
13. Naylor, P., Laé, M., et al., R.: Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE Transactions on Medical Imaging **38**, 1–1 (08 2018). https://doi.org/10.1109/TMI.2018.2865709
14. Ouahab, A., Ahmed, O.B.: Protomed: Prototypical networks with auxiliary regularization for few-shot medical image classification. Image and Vision Computing **154**, 105337 (2025). https://doi.org/https://doi.org/10.1016/j.imavis.2024.105337, https://www.sciencedirect.com/science/article/pii/S0262885624004426
15. Shi, W., et al., Z.: Segment anything model for few-shot medical image segmentation with domain tuning. Complex & Intelligent Systems **11**(1), 37 (2025)
16. Szolomicka, J., Markowska-Kaczmar, U.: Harnessing Few-Shot Learning Segmentation for Histopathology: A Comprehensive Practical Study, pp. 67–89. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-031-98149-4_4, https://doi.org/10.1007/978-3-031-98149-4_4
17. Xu, Q., Liu, X., et al., L.Z.: Hybrid mamba for few-shot segmentation (2024), https://arxiv.org/abs/2409.19613
18. Xu, Q., Chen, Y., Li, Y., Liu, Z.e.a.: Mambavesselnet++: A hybrid cnn-mamba architecture for medical image segmentation. ACM Trans. Multimedia Comput. Commun. Appl. (Aug 2025). https://doi.org/10.1145/3757324, https://doi.org/10.1145/3757324, just Accepted
19. Zhu, L., et al., B.L.: Vision mamba: Efficient visual representation learning with bidirectional state space model (2024), https://arxiv.org/abs/2401.09417

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_18