Biomedical Signal Processing and Control 112 (2026) 108711

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

MACA-Net: Multi-aperture curvature aware network for instance-nuclei segmentation

Siyavash Shabani [a], Sahar A Mohammed [a], Muhammad Sohaib [a], Bahram Parvin [a,b,c,*]

[a] Department of Electrical and Biomedical Engineering, University of Nevada, Reno (UNR), USA
[b] Pennington Cancer Institute, USA
[c] Department of Microbiology and Immunology, (UNR), USA

ARTICLE INFO

Keywords:
Nuclear segmentation
Vision transformers
Principal curvature
Computational histopathology

ABSTRACT

Nuclei instance segmentation is one of the most challenging tasks and is considered the first step in automated pathology. The challenges stem from technical biological variations, and high cellular density that lead adjacent nuclei to form perceptual boundaries. This paper demonstrates that a multi-aperture representation encoded by the fusion of Swin Transformers and Convolutional blocks improves nuclei segmentation. The loss function is augmented with the curvature and centroid consistency terms between the growth truth and the prediction to preserve morphometric fidelity and localization. These terms are used to panelize for the loss of shape localization (e.g., a mid-level attribute) and mismatches in low and high-frequency boundary events (e.g., a low-level attribute). The proposed model is evaluated on three publicly available datasets: PanNuke, MoNuSeg, and CPM17, reporting improved Dice and binary Panoptic Quality (PQ) scores. For example, the PQ scores for PanNuke, MoNuSeg, and CPM17 are 0.6888 ± 0.032, 0.634 ± 0.003, and 0.716 ± 0.002, respectively. The code is located at https://github.com/Siyavashshabani/MACA-Net.

1. Introduction

Analysis of histology images, stained with hematoxylin and eosin (H&E) dyes, is the first step in diagnostics, where morphometry and organization of nuclei play an important role. As a result, robust nuclei segmentation has been of significant interest in computer-aided pathology. This area is partly driven by the projected shortage of pathologists and partly by the emerging applications of large-scale image-based data. Robust segmentation of nuclei enables applications that include, but are not limited to computing the frequency of mitotic cells [9] and profiling nuclear morphometry [14,15] (e.g., aneuploidy, pleomorphism, vesicular phenotype), and cellular organization within the Tumor Microenvironment. For example, nuclear morphology plays an essential role in the oncogenic program due to alterations in the genetic and epigenetics of cancer cells, and studies have shown that poor prognosis in breast cancers is associated with increased nuclear area and altered shape [16]. Another emerging application is in precision medicine based on the large-scale availability of libraries of low-cost histology sections, which will become more prevalent with the rapid proliferation of whole slide imaging, cloud computing, and population studies. In this context, robust nuclei segmentation plays an important role since it is site-specific, interpretive, and explainable. Robust nuclei segmentation in H&E-stained images is hindered by technical (e.g., sample thickness and staining) and biological (e.g., cell state, morphometric diversity, diseased state) variations. The latter can also be augmented by complexities associated with high cellular density, forming perceptual boundaries between adjacent nuclei. Methods of nuclei segmentation have benefited from advances in the applications of machine learning in computer vision and image analysis. Some of these advances are summarized below, which are then placed in context for the segmentation of nuclei.

CNN models have proven effective in many image analysis tasks; however, they are limited to local receptive fields and may struggle to capture long-range spatial relationships. However, recent advances in Transformer-based models in the field of NLP [19] provide the global context in the form of Vision Transformers (ViTs[20]). These architectures are based on the self-attention mechanism [19], enabling the model to integrate long-range dependencies. Vision Transformers have been used in image classification [23], object detection [25,26], and segmentation [29,30]. ViTs have been quite influential in medical image analysis [26,27,29,31–33]. For instance, Swin-U-Net [27] introduces a novel U-Net architecture by incorporating transformer blocks in the

* Corresponding author at: Department of Electrical and Biomedical Engineering, University of Nevada, Reno (UNR), USA.
E-mail address: bparvin@unr.edu (B. Parvin).

https://doi.org/10.1016/j.bspc.2025.108711
Received 27 January 2025; Received in revised form 14 July 2025; Accepted 18 September 2025
Available online 2 October 2025

encoder and decoder branches, marking the first generation of fully convolutional-free U-Net frameworks. In general, the above advances are either in terms of a unique model architecture or customized loss functions [34].

Building on previous research, we propose a model that is faster and has a smaller footprint than prior art, as shown in Fig. 1. This model architecture is shown in Fig. 2, which also utilizes a novel loss function that captures local and mid-level discrepancies. The proposed backbone utilizes multi-aperture Swin Transformers and convolutional blocks, whose outputs are integrated with fusion blocks. The benefit of the multi-aperture transformer is that, unlike the multi-resolution representation, the details of the original image content are preserved. Hence, by maintaining the original resolution, the chance of adjacent nuclei in very close proximity is reduced.

One novel aspect of the proposed loss function is the application of principal curvature. The boundary curvature captures bends and folds along the contour. A related measurement is based on computing the eigenvalues (e.g., principal curvature) of the Hessian matrix calculated from the distance transform in which the contour resides in a 2D surface, as shown in Fig. 3. This related measurement has the benefit of vectorization for high-performance computation. Penalizing for the discrepancies in principal curvatures can (i) promote the formation of perceptual boundaries between adjacent nuclei, (ii) localize the boundary overlap between predicted and ground truth, and (iii) attenuate the formation of localized high-frequency artifacts. The proposed approach is applied to multiple datasets and compared with prior ones to show improved performance. These datasets include CPM17 [35], MoNuSeg [36], and PanNuke [37].

In summary, our contributions are:

- A backbone with four parallel Swin Transformer modules that utilizes the multi-aperture representation for preserving the original image's details. In turn, each Swin Transformer is fused with a convolutional block that captures local features in a high-dimensional space.
- A custom loss function based on the principal curvatures is used to enforce more accurate localization and topological consistency while preserving high-frequency events.
- An improved performance of nuclei segmentation tested on standardized public datasets, followed by extensive ablation studies.

The net results of these innovations have enabled a platform with a reduced number of parameters while improving performance in terms of FLOPs.

The organization of this paper is as follows. Section 2 summarizes the related research. Section 3 outlines the detailed methodologies. Section 4 summarizes the results and the comparative performance with the prior research. Section 5 outlines the details of the ablation studies. Finally, Section 6 provides a discussion and suggestions for future efforts.

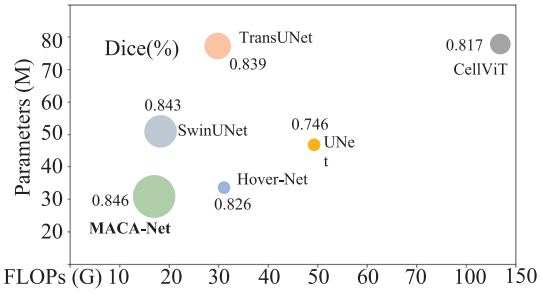

Fig. 1. With 31 million parameters with a computational cost of 16.5 GFLOPs, MACA-Net has a lower footprint and computational cost than prior research.

2. Related works

A complete review of techniques in nuclear segmentation is beyond the scope of this paper, and the topic has been extensively surveyed in recent literature [38,39]. Below, we summarize some of the most relevant works in context.

Hover-Net [1] is one of the first successful systems that integrates a CNN encoder block with multiple convolutional decoder blocks for simultaneous segmentation and classification. They demonstrate that by predicting horizontal and vertical distance maps, a better segmentation of a clump of cells is achieved. Koohbanani et al. [40] utilize a spatially-aware network (SpaNet) to capture spatial information by predicting the pixel-wise segmentation and the centroid of nuclei. Naylor et al. [7] frame nuclei segmentation as a regression problem by leveraging a CNN framework complemented by the distance map. Schmidt et al. [41] extend the U-Net backbone with additional layers for simultaneous pixel-level segmentation and localization. They suggested that localization can benefit from an improved shape representation based on "star-convex polygons" which encode a radial distance map. Chen et al. [11] proposed sampling a point set from each nucleus instead of a single point (e.g., centroid) to compute distance information, hence improving contextual representation. In summary, the current state of the art overcomes the problem of dense or overlapping nuclei by incorporating a representation of the distance map within the CNN framework. MA-Net [42] proposed an end-to-end architecture by incorporating an attention gate in the decoder block at multiple resolutions. CellViT [5] incorporates the vanilla ViT as an encoder within a U-Net-shaped framework for segmentation and classification, connecting the model's bottleneck to two MLP layers specifically for classification purposes. CellT-Net [43] proposes a novel framework based on dual Swin Transformers to segment and detect cells simultaneously in fluorescent microscopy images. NuHTC [44] introduces a novel WSPN block, inspired by the watershed algorithm, which is integrated into the decoder module. The backbone architecture is based on the Swin Transformer, aiming to enhance performance in both segmentation and classification tasks. Finally, in another recent manuscript, the authors leveraged the inherent heterogeneity of nuclei and their domains and proposed unsupervised domain adaptation. This is based on a two-stage disentanglement framework for nuclei instance segmentation under the Open Compound Domain Adaptation (OCDA) setting [45]. The authors define the source and target domains of labelled and unlabeled images, respectively, where the target is heterogeneous.

We hypothesize that multi-aperture representation better preserves the original image content, and the fusion of convolution and transformation blocks provides a more effective method for integrating local and global attributes. Finally, a curvature-aware loss function enhances topological consistency between the ground truth and prediction.

3. Methods

This section summarizes the details of the datasets, the model architecture, implementation, and evaluation metrics. The architecture consists of five modules: Swin-transformer blocks, convolutional blocks, fusion blocks, decoder blocks, and post-processing blocks for creating instance maps. It is tightly coupled with a unique loss function based on principal curvature. This section concludes with the implementation details and evaluation metrics. To compare the performance of the method with the published literature, we restricted the comparison to those studies that use (a) the same datasets, and (b) a common, well-defined stratification of training and validation partitions. This is often specified in distinct folders that training and validation datasets resides.

3.1. Datasets

Three standardized and independent datasets are used for comparative evaluation. The size of the images varies between each dataset. For

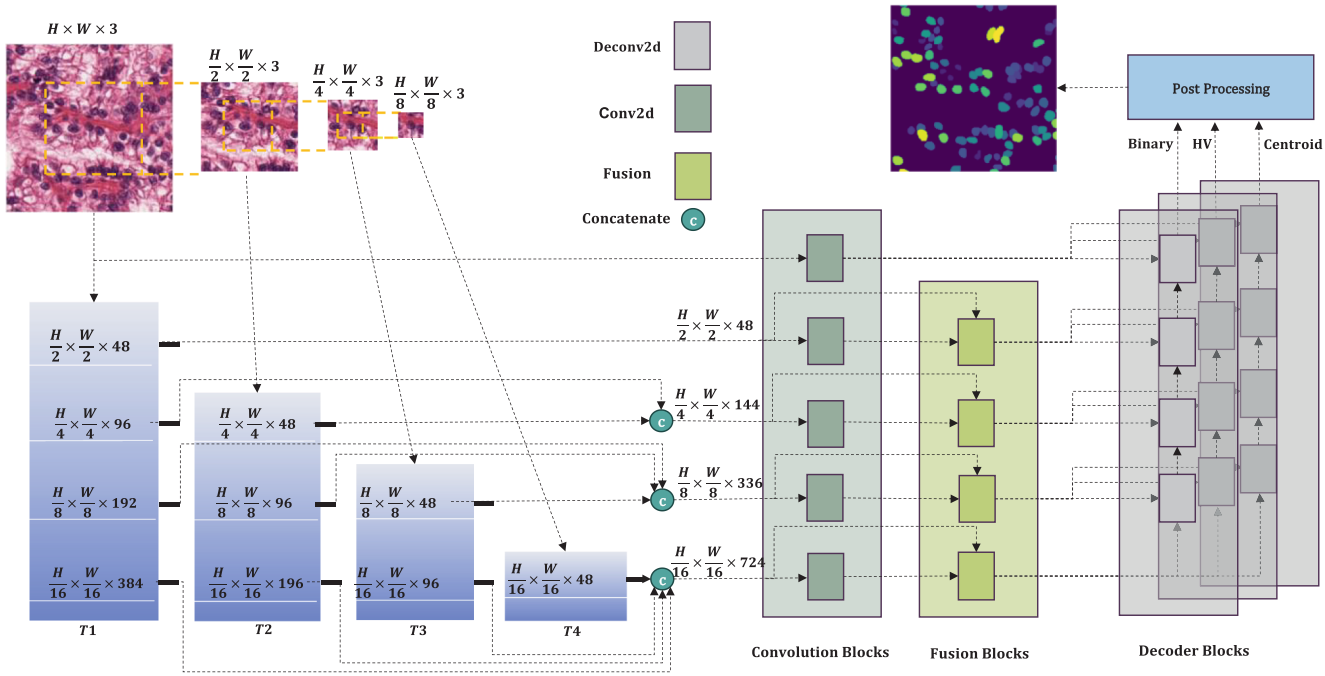

Fig. 2. Overview of MACA-Net: (a) The Proposed framework couples four Swin Transformers and their corresponding convolutional blocks via fusion blocks.

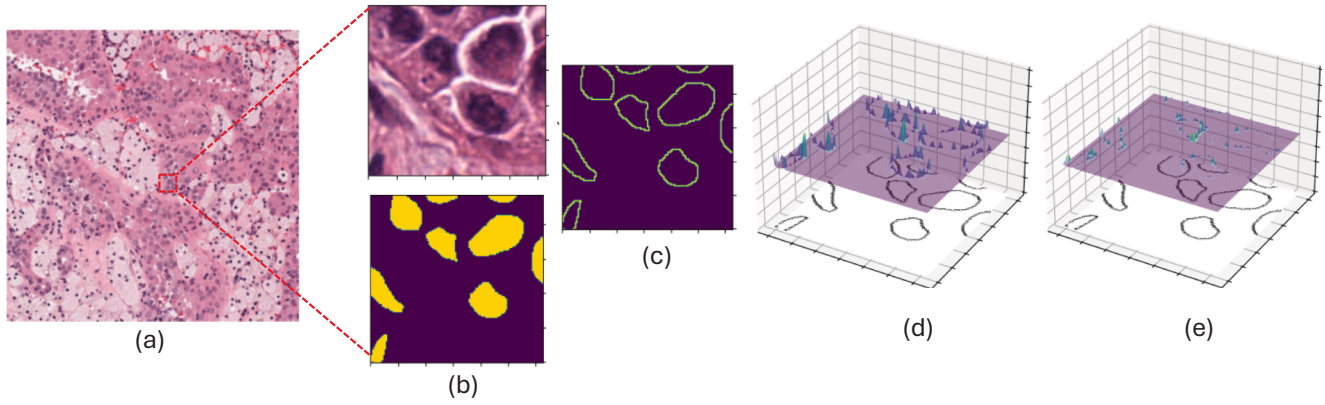

Fig. 3. The largest principal curvature captures sharp turns (e.g., rate of change in the tangent) along the boundary of each nucleus. (a) an original H&E image from the MoNuSeg dataset, (b) a small patch and its binary mask, (c) the extracted boundary of the mask patch, and (d-e) the eigenvalues of the Hessian matrix at the boundary of nuclei.

the purpose of training, images are patchified with 256-by-256 pixels and 50 % overlap between patches.

MoNuSeg [36] (Multi-Organ Nuclei Segmentation) dataset is one of the most extensive collections of manually annotated nuclei. It comprises 30H&E-stained histopathology images, each measuring 1000 × 1000 pixels, sourced from seven distinct organs, and includes 21,623 individually annotated nuclei. For an equitable comparison, we utilize the identical training and testing splits outlined in[37], which were also employed during the MoNuSeg Grand Challenge 2017.

CPM17 [35] is included in the *Computational Precision Medicine Digital Pathology Challenge*. It includes 64H&E-stained histopathology images, with a size of 500 × 500 pixels, featuring 7,570 annotated nuclear boundaries. In line with the original challenge specifications [35], the dataset is divided into two sets, each containing 32 images for training and testing.

PanNuke [37] contains 189,744 annotated nuclei from 7904 images of size 256-by-256 pixels. These images originate from 19 tissue sections (e.g., stomach, breast, and pancreatic) with five phenotypes (e.g., neoplastic, epithelial, inflammatory, connective, and dead cells). This dataset exhibits substantial imbalance, especially in underrepresenting dead nuclei, as evidenced by statistical analysis. PanNuke is recognized as one of the most demanding datasets for simultaneous segmentation and classification. Supplementary Fig.1 illustrates the number of extracted patches for three datasets, for training and testing the proposed model.

3.2. Model architecture

Fig. 2 shows the proposed architecture for multi-aperture representation, where the outputs of four Swin Transformers (Swin T1, Swin T2, Swin T3, Swin T4) are fused with convolutional modules using fusion blocks (Fig. 4). Each Swin Transformer captures global contextual relationships. In contrast, 2D convolutional blocks extract local features from this representation in a higher-dimensional space. Next, the outputs of the fusion blocks were connected to two decoder branches, each containing four deconvolutional blocks for reconstructing the binary mask and horizontal and vertical distance transforms. The details of these components and the loss function are as follows:

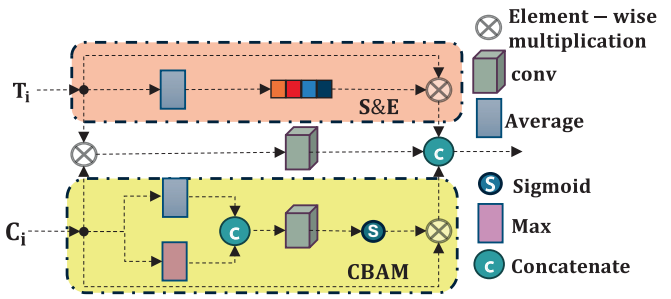

Fig. 4. The fusion block consists of a squeeze and excitation block and a CBAM block.

3.3. Swin transformers

The input data of the model is a RGB patch which is extracted from huge H&E images $x \in R^{H\times W\times 3}$ This 2D patch is then processed by four Swin Transformer blocks [27], each with a different aperture, as shown in Fig. 2. The extracted patch $[H, W, 3]$ serves as the input for the first Transformer block. Then, the center of the input for each block is selected iteratively as the input for the next block, i.e., the center of the first block at $\left[\frac{H}{2}, \frac{W}{2}\right]$ is selected as the input for the second block, the center of the input of the second block at $\left[\frac{H}{4}, \frac{W}{4}\right]$ is selected as input for the next block, etc.

3.4. Convolutional blocks

The model architecture of Fig. 2 consists of five convolutional blocks, which receive either the raw images or concatenated feature maps from the outputs of two Swin Transformers at two consecutive apertures. These blocks are designed to extract local feature maps from either the input image or a representation of it in a higher-dimensional space.

3.5. Fusion blocks

Fusion blocks, as shown in Fig. 4, integrate the global and local spatial features. Each fusion block concatenates the outputs of a Transformer and corresponding Convolutional blocks through a series of non-linear operations that include a squeeze and excitation (SE) module [46], element-wise multiplication, and Convolutional Block Attention Module (CBAM) [47]. Each module accentuates a specific aspect of the outputs of the input data, i.e., transformers and convolutions. For example, the SE block prioritizes the significance of channel-wise information, simple elementwise multiplication (e.g., Hadamard dot product) modulates the importance of features, and CBAM weights spatial similarities.

3.6. Decoder blocks

The decoder blocks have three branches for predicting the mask, centroids of nuclei, and HV distance maps. Each branch comprises four blocks containing convolutional and deconvolutional layers, which are essential for reconstructing the output maps. The simultaneous prediction of a binary mask, centroid map, and the HV distance map infers a more constrained and stable solution.

3.7. Post processing block

For each patch, the final probability map corresponding to the ground truth undergoes a series of morphometric operations. These include smoothing, hole filling, and watershed to produce an instance map. These patches are then aggregated to form the entire image.

3.8. Principal curvature loss

One of the novelties of the proposed method is the inclusion of principal curvature in the loss function, and its details are as follows.

During the training process, the boundaries of predicted and ground truths are extracted. These are defined as the set of all pixels at the interface between nuclei and the background. Following boundary extraction, the signed distance transforms (SDT) are computed from both ground truth and predicted masks, as defined below:

$$D(p) = \begin{cases} -min_{q\in Boundary}\|p-q\|, If\ p\ is\ inside \\ min_{q\in Boundary}\|p-q\|, If\ p\ is\ outside \end{cases} \quad (1)$$

Next, the Hessian matrix is computed at each boundary point , p_i, where Hessian is a second-order differential operator:

$$H(p_i) = \begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial x \partial y} \\ \frac{\partial^2 D}{\partial y \partial x} & \frac{\partial^2 D}{\partial y^2} \end{bmatrix} \quad (2)$$

Subsequently, the eigenvalues λ_1 and λ_2 of the Hessian matrix, at each boundary point, are calculated and ordered. In other words, principal curvatures at non-boundary points are set to zero. These eigenvalues correspond to the principal curvatures that capture the change in the tangent or how sharply the boundary turns. An example of principal curvature computation is shown in Fig. 3.

The loss term in principal curvature (L_{PCurv}) represents the dissimilarities between the ground truth and prediction:

$$L_{PCurv} = \frac{\sum_{i,j\in B}\left|\lambda_1^{i,j}GT - \lambda_1^{i,j}P\right|}{\sum_{i,j\in B}\lambda_1^{i,j}GT} + \frac{\sum_{i,j\in B}\left|\lambda_2^{i,j}GT - \lambda_2^{i,j}P\right|}{\sum_{i,j\in B}\lambda_2^{i,j}GT} \quad (3)$$

where i and j are boundary points ($i,j \in \{B_{GT}, B_p\}$) with $\lambda_1^{i,j}GT$ and $\lambda_1^{i,j}P$ being the largest eigenvalues at that location. This loss function is not symmetrical and is designed to penalize discrepancies from the ground truth. In Fig. 7c and Fig. 7d, we investigated the effect of this proposed loss term on the performance of MACA-Net.

3.9. Loss function

The augmented loss function incorporates four consistencies that correspond to the losses in (i) binary mask (L_{Bin}), (ii) HV distance maps (L_{HV}), centroids of nuclei (L_{Cntr}) (e.g. localization error based on Euclidean distance), and (iii) principal curvatures (L_{PCurv}). Binary and HV distance loss maps have been used in previous literature [5]. Hence, we focus on principal curvature. as follows:

$$Loss = \alpha L_{Bin} + \beta L_{HV} + \gamma L_{Cntr} + \delta L_{PCurv} \quad (4)$$

where α, β, γ and δ are hyperparameters. L_{bin} combines dice and cross-entropy loss:

$$L_{bin} = \epsilon L_{Dice} + \zeta L_{CE} \quad (5)$$

L_{HV} consists of horizontal and vertical loss:

$$L_{HV} = \eta L_{MSE}^{H,V} + \theta L^{H,V}{}_{MSGE} \quad (6)$$

where L_{MSE} *and* L_{MSGE} are Mean Square Error (MSE) and mean squared gradient errors (MSGE) of the horizontal and vertical distance transform map, respectively. This notation was also used in CellViT [5].

To compute L_{Cntr}, we generate binary centroid maps from the predicted and ground-truth nuclei masks and apply the primary Dice loss:

$$L_{cntr} = 1 - \frac{2\sum_{i,j}P_{cntrij}GT_{cntrij}}{\sum_{i,j}P_{cntrij} + \sum_{i,j}GT_{cntrij} + \epsilon} \quad (7)$$

where P_{cntr} and GT_{cntr} are the predicted and ground-truth binary centroid maps, and ϵ is a small constant to avoid division by zero.

Finally, we have L_{HV}, which represents the loss function for the centroids of cells, utilizing the MSE (Mean Squared Error) loss function. The hyperparameter tuning details for the components of our proposed loss function are provided in Supplementary Table 1.

3.10. Implement details

Our model was implemented in PyTorch 2.3.0. For data augmentation, we used the MONAI library for limited affine transformation (e.g., flipping, rotation), contrast modulation, limited scaling, and added noise. Experiments were conducted on a Linux Cluster Server with 8 NVIDIA RTX 3080 GPUs, each with 12 GB of memory. The MoNuSeg and CPM17 datasets each include training and test sets. We divided the training sets into training and validation subsets, allocating 80 % for training and 20 % for validation, respectively. Although the image dimensions in the PanNuke dataset are compatible with our model's input, we generate 256-by-256 patches with 50 % overlap during preprocessing to ensure consistency and enhance learning. Following patch extraction, a series of augmentation techniques are applied to improve model generalization. Details of the associated hyperparameters are provided in Supplementary Table 2. The best-performing model from the training process was saved for final evaluation. Finally, a five-fold cross-validation is used for reporting and statistical analysis.

3.11. Evaluation metrics

We employed two evaluation metrics to measure the overall segmentation performance: the average Dice coefficient (Dice) and Panoptic Quality (PQ). The Dice and PQ scores measure pixel-level and object-level consistencies, where the PQ score is defined as follows:

$$PQ = \frac{|TP|}{|TP| + 0.5|FP| + 0.5|FN|} \times \frac{\sum_{(y,\hat{y})\in TP} IoU(y,\hat{y})}{|TP|} \tag{8}$$

Here, TP represents True Positives, the correctly identified cells; FP stands for False Positives, which are the nuclei incorrectly identified as being present; and FN denotes False Negatives, which are the nuclei that are present but were not detected. y and $\hat{y}$ correspond to the ground truth and prediction, respectively. Finally $IoU(y, \hat{y})$ denoting the intersection-over-union[48].

4. Results

This section summarizes the evaluation of the performance of our proposed MACA-Net and ablation studies. In reporting the performance, we followed the same protocol used in the past for a more robust comparative analysis.

4.1. Performance on the MoNuSeg and CPM17 datasets

MACA-Net, as shown in Table 1, has an improved Dice score of 0.846 ± 0.002 and.889 ± 0.003 on MoNuSeg and CPM17 datasets, respectively. The PQ scores have also been improved on the same datasets. We also conducted a qualitative visualization analysis on the MoNuSeg and CPM17 datasets. As shown in Fig. 5, our proposed framework effectively separates nuclear instance cells across various images. For a more precise comparison, we have highlighted the differences in distinguishing nuclear pixels from the background and segmenting clustered instances with red insets in Fig. 5. These highlights facilitate direct comparison between our model and other state-of-the-art models such as CellViT, CDNet, and Hover-Net.

Table 1
Performance comparisons on MoNuSeg and CPM17 datasets. *Although the results of CellViT were not recorded in the published manuscript, we ran their model on both datasets and recorded the Dice and PQ scores. The results are reported with five folds cross validation.

Method	Par	MoNuSeg		CPM17	
		Dice↑	PQ↑	Dice↑	PQ↑
U-Net[2]	38.5 M	0.746	0.581	0.757	0.625
MaskRCNN[4]	44.0 M	0.760	0.509	0.850	0.674
DIST[7]	–	0.789	0.504	0.826	0.504
DCAN[10]	–	0.792	0.492	0.828	0.545
CIA-Net[12]	46.0 M	0.818	–	0.841	–
PFF-Net[15,17]	20.8 M	0.809	0.587	–	–
CellViT[5]*	42.0 M	0.817	0.619	0.879	0.685
Hover-Net[1]	34.0 M	0.826	0.597	0.869	0.697
CDNet[8]	36.0 M	0.831	–	0.880	–
ToPoSeg[21]	–	–	0.625	–	0.705
UN-SAM[22]	308.0 M	0.841	0.624	–	–
UDTransNet[24]	33.9 M	0.790	–	–	–
SwinUNet[27]	41.3 M	0.843	**0.675**	0.860	0.601
TransUNet[28]	66.8 M	0.839	0.661	0.845	0.606
MACA-Net	**31.0 M**	**0.846**	0.634	**0.889**	**0.716**

4.2. Performance on the PanNuke dataset

Table 2 compares the performance of the proposed model against prior research on the PanNuke dataset for 19 diverse tissues, with MACA-Net producing an improved PQ score of 0.688 ± 0.032 with three-fold cross-validation. On average, MACA-Net performs better than the state-of-the-art studies CellViT and PointNu-Net using a one-tail paired-*t*-test with p-values of 0.01 and 0.04, respectively. Moreover, out of 19 tissue types, we observed notable improvement in 14 tissue types: Adrenal, Bile Duct, breast, Colon, Esophagus, Liver, Lung, Ovarian, Pancreatic, Skin, Stomach, Testis, Thyroid, and Uterus. Finally, the proposed model has only 31 M parameters, which is lower than that of prior research, as per Table 1. Fig. 6 illustrates the quality of the MACA-Net on sections for each of the 19 tissue types.

5. Ablation study

Ablation studies include the effect of different loss functions, the number of Transformers, fusion blocks, and data size on the MoNuSeg and CPM17 datasets.

5.1. Effect of transformers and fusion blocks

Figs. 7a and 4b summarize the impact of Transformers and fusion blocks on model performance for MoNuSeg and CPM17 datasets. The multi-aperture Transformers and fusion blocks improved the proposed framework's performance.

5.2. Quantitative effect of principal curvature

The loss function of Eq. (1) consists of four terms: $\boldsymbol{L_{bin}}$, $\boldsymbol{L_{HV}}$, $\boldsymbol{L_{Cntr}}$ and $\boldsymbol{L_{PCurv}}$, which are deviation from (i) binary ground truth, (ii) horizontal and vertical distance transform map, (iii) centroid of nuclei, and (iv) principal curvature. In this section, as detailed in Fig. 7c and Fig. 7d, we empirically analyzed the effectiveness of each term on the performance of our MACA-Net framework. Fig. 7c and Fig. 7d illustrate our findings, which indicate that every loss function contributes to a certain degree to the overall performance as measured by the Dice and PQ scores. For example, horizontal and vertical distance losses or curvature- contribute considerably to overall performance. On the other hand, contributions from the centroid are marginal. Finally, Supplementary Table 1 indicates the result of hyperparameter tuning for each of the four loss functions.

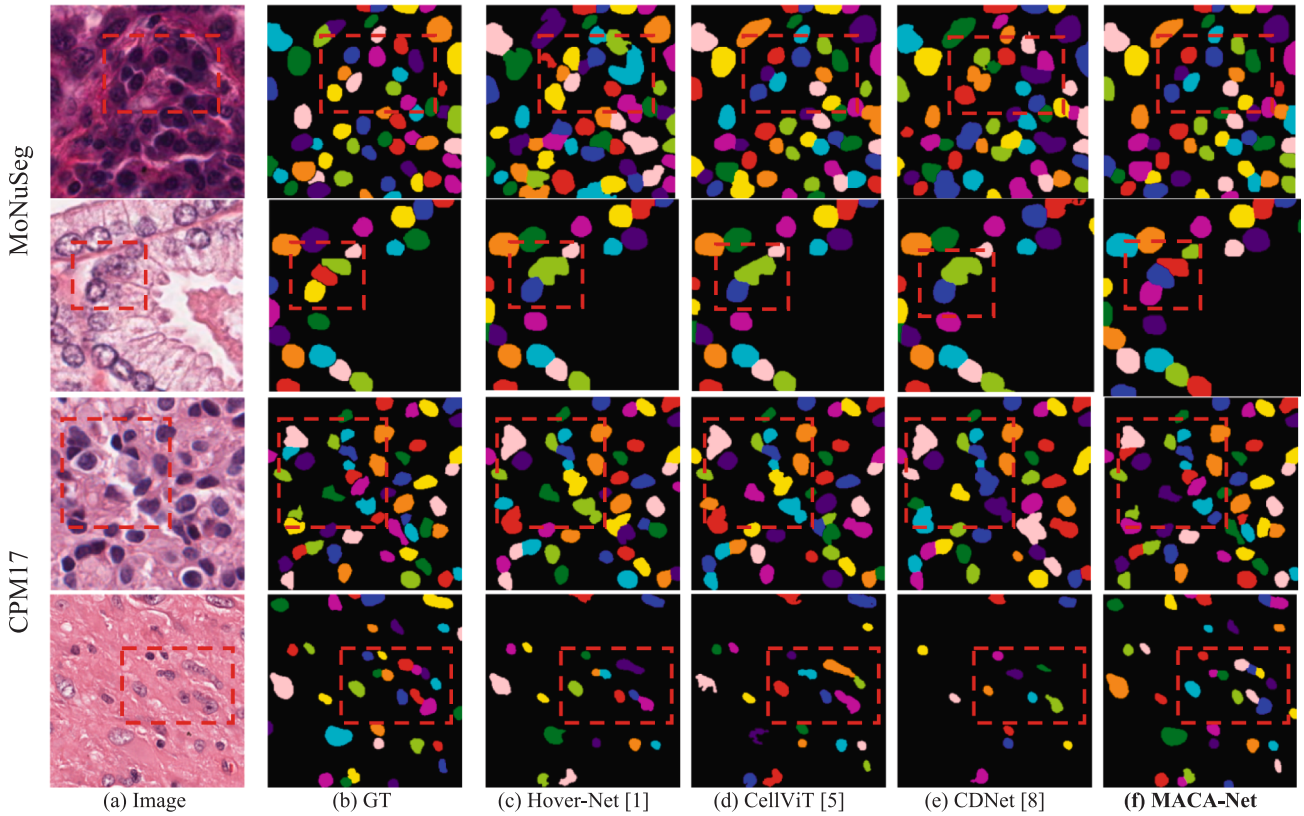

Fig. 5. Improved performance of the MACA-Net is visualized on representative images from MoNuSeg and CPM17 and compared with the prior art. (a) Original images; (b) Ground Truth (GT); (c-f) Superior performance of MACA-Net (f) is shown when compared with Hover-Net, CellViT, and CDNet. The red rectangles provide context for qualitative comparison.

Table 2
MACA-Net has an improved PQ score on the PanNuke multi-tissue dataset with P-Value < 0.05 using a one-tailed *t*-test.

Tissue	Hover-Net[1]	TSFD-Net[3]	STARDIST[6]	CPP-Net[11]	Micro-Net[13]	CellViT[5]	PointNu-Net[18]	MACA-Net
Adrenal	0.696	0.690	0.697	0.703	0.644	0.708	0.713	**0.726**
Bile Duct	0.669	0.628	0.669	0.673	0.623	0.678	0.681	**0.686**
Bladder	0.703	0.677	0.698	0.705	0.648	0.706	**0.722**	0.682
Breast	0.647	0.624	0.666	0.671	0.602	0.674	0.670	**0.689**
Cervix	0.665	0.656	0.669	0.688	0.610	0.687	**0.689**	0.664
Colon	0.557	0.537	0.577	0.588	0.497	0.592	0.594	**0.599**
Esophagus	0.642	0.630	0.665	0.675	0.601	0.668	0.676	**0.698**
Head & Neck	0.633	0.627	0.643	0.646	0.524	0.654	**0.654**	0.642
Kidney	0.683	0.682	0.699	0.700	0.632	**0.709**	0.691	0.695
Liver	0.724	0.667	0.723	0.727	0.666	0.732	0.731	**0.740**
Lung	0.630	0.594	0.636	0.636	0.558	0.642	0.635	**0.659**
Ovarian	0.630	0.643	0.666	0.679	0.601	0.672	0.686	**0.708**
Pancreatic	0.649	0.624	0.660	0.674	0.607	0.665	0.679	**0.697**
Prostate	0.661	0.640	0.674	0.690	0.604	0.682	**0.685**	0.676
Skin	0.623	0.607	0.628	0.619	0.581	0.656	0.649	**0.660**
Stomach	0.688	0.652	0.694	0.704	0.629	0.702	0.701	**0.714**
Testis	0.689	0.643	0.686	0.700	0.630	0.695	0.705	**0.709**
Thyroid	0.698	0.669	0.696	0.709	0.655	0.715	0.707	**0.733**
Uterus	0.639	0.620	0.659	0.662	0.582	0.662	0.663	**0.692**
Average	0.659	0.637	0.669	0.676	0.605	0.679	0.680	**0.688**
STD	0.037	0.034	–	–	0.033	0.030	–	0.032

5.3. Effect of data size on the performance

The size of the training can have a direct impact on performance. While there is no benchmark for evaluating the effect of training size, we opted to investigate this factor. Fig. 7e and Fig. 7f illustrate the performance of MACA-Net on two independent datasets of MoNuSeg and CPM17 as a function of increasing the training data size. The training sets' total size is 815 and 3132 patches for CPM17 and MoNuSeg, respectively. Interestingly, with only 10 % of the data, a Dice score of 0.732 is achieved, which suggests that MACA-Net has a steep initial learning curve, a distinct property of convolutional networks. The next significant performance improvement is at 50 % of data for training, and then the learning curve plateaus.

5.4. Effect of the fusion block

The fusion block may assume alternative architectures; hence, a simulation study is needed. Presently, the fusion block applies Squeeze-

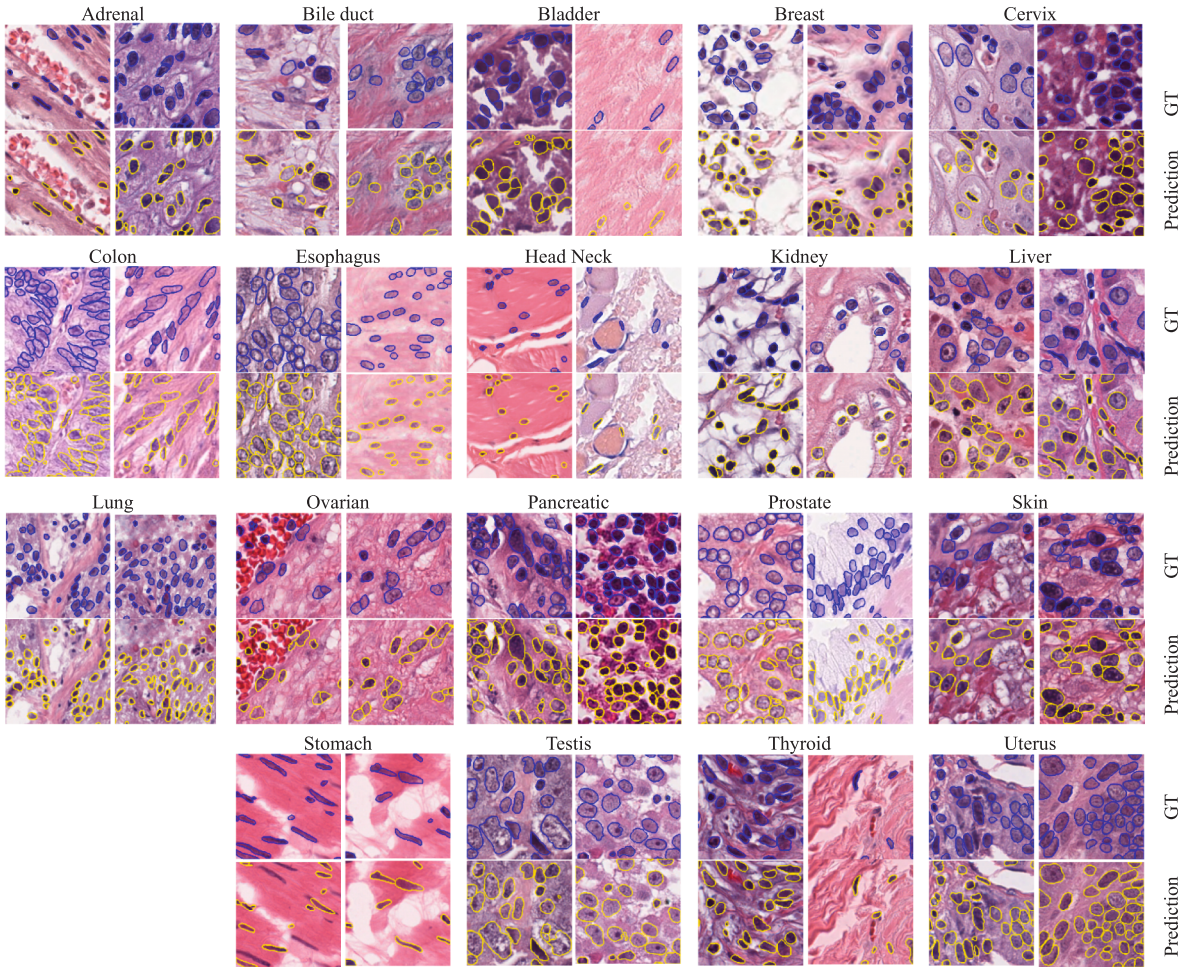

Fig. 6. Quality of MACA-Net nuclei segmentation on different tissue types from a section in the PanNuke dataset is illustrated against the ground truth (GT). The performance of MACA-Net for histology sections from 19 different organs is shown. In each case, the top and bottom rows correspond to the ground truth and prediction.

and-Excitation (S&E) and Convolutional Block Attention Module (CBAM) to the output of the Transformer and Convolutional blocks, respectively. The intuition is that transformers generate a very long feature sequence, and removing redundancies through an S&E block makes sense. On the other hand, convolutional blocks lack attention; hence, it is intuitive to complement them with an attention block. This intuition is complemented with the simulation results of Table 3 with three-datasets, indicating that the fusion block constructed with S&E and CBAM blocks performs the highest Dice and PQ scores.

6. Discussion and future research

A summary of the results, significance, and potential applications, and future efforts follows. From a methodological perspective, we have shown improved performance by (i) multi-aperture representation with an encoder backbone that fuses the outputs of Swin Transformers and convolutional blocks, and (ii) integration of curvature loss in the loss function. Rigorous ablation studies complement the study to demonstrate proof of concept. Additionally, we also learned that combining all three datasets further enhances the system performance, as shown in Supplementary Table 3, which suggests intrinsic diversity in each dataset. The proposed protocol has three significance. First, this is significant in clinical studies, where the differences between treatment and control are minor and necessitate the need for a more sensitive computational method. Second, the proposed protocol performs better in nuclei segmentation across multiple organ datasets, providing a more robust pipeline for preclinical studies where histology sections are collected from multiple organs. For example, a typical preclinical study includes the administration of a drug in a rodent model, followed by tissue collection from organs. Typically, several replicate animals are used to meet the required power requirements. As a result of multiple histology sections per organ and per animal, a large number of histology sections are collected that require automated analysis and summarization. Third, the potential of this system in computer-aided pathology cannot be overstated, especially given the shortage of pathologists, the increasing age of the workforce, and the increased volume of the workload [49]. The current limitations of the proposed protocols are fourfolds: (i) There are opportunities to improve the segmentation quality by increasing the PQ score. (ii) Segmentation needs to be complemented by classification, such as mitotic-, immune-, tumor versus stroma-cells. (iii) A finer stratification is needed in terms of pleomorphism and aneuploidy. And (iv) the system needs to be tested in the context of a preclinical or clinical trial for its efficacy. These will be some of our focus areas in the future.

CRediT authorship contribution statement

Siyavash Shabani: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Sahar A Mohammed:** Validation, Methodology. **Muhammad Sohaib:** Visualization. **Bahram Parvin:** Writing – review & editing, Writing – original draft, Investigation, Funding acquisition,

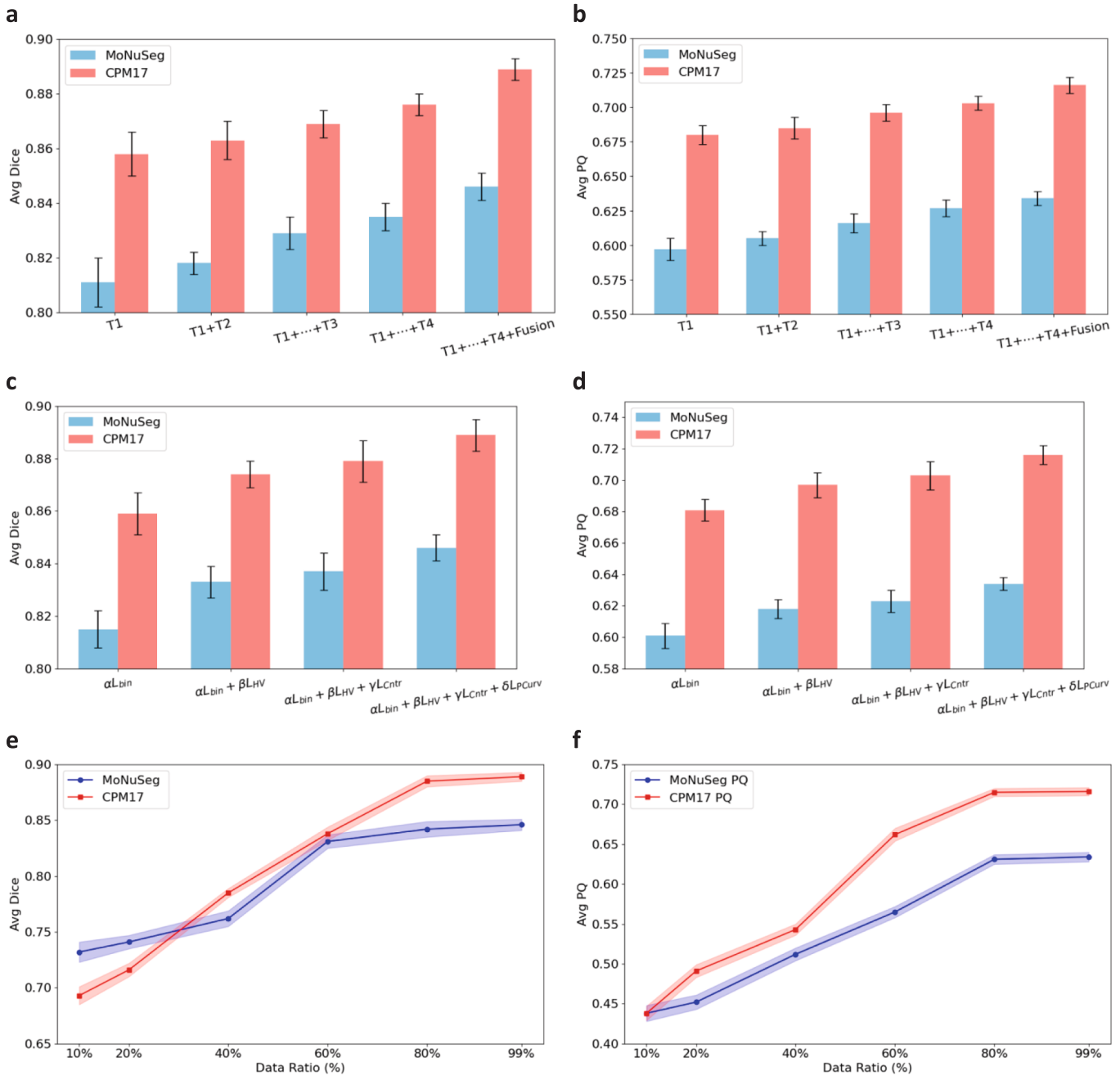

Fig. 7. (a, b) Ablation study of multi-aperture swin transformer shows improved performance with increasing apertures and adding the fusion blocks on the MoNuSeg and CPM17 datasets. (c, d) The inclusion of principal curvatures into the loss function improved performance on the MoNuSeg and cpm17 datasets. (e, f) MACA-Net can achieve an Average Dice score of approximately 0.8 with only 50% of the data, and the performance plateaus after 80% of the data is used.

Table 3
Ablation studies of the fusion block indicate that incorporating an S&E block for transformer feature maps and a CBAM block for convolution feature maps achieves the highest Dice and PQ scores on the MoNuSeg, CPM17, and PanNuke datasets (✓: Included, ×× : Excluded).

Tr(S&E)	Tr(CBAM)	Conv(S&E)	Conv(CBAM)	MoNuSeg		CPM17		PanNuke
				Dice↑	PQ↑	Dice↑	PQ↑	PQ↑
✓	×	✓	×	0.823	0.623	0.867	0.693	0.663
×	✓	×	✓	0.811	0.619	0.858	0.692	0.674
×	✓	✓	×	0.818	0.627	0.862	0.678	0.659
✓	×	×	✓	**0.846**	**0.634**	**0.889**	**0.716**	**0.688**

Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by a grant from NIH CA279408.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bspc.2025.108711.

Data availability

Public data are used.

References

[1] S. Graham, et al., Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images, Med. Image Anal. 58 (2019) 101563.
[2] H. Wu, N. Souedet, C. Jan, C. Clouchoux, T. Delzescaux, A general deep learning framework for neuron instance segmentation based on efficient UNet and morphological post-processing, Comput. Biol. Med. 150 (2022) 106180.
[3] T. Ilyas, Z.I. Mannan, A. Khan, S. Azam, H. Kim, F. De Boer, TSFD-Net: Tissue specific feature distillation network for nuclei segmentation and classification, Neural Netw. 151 (2022) 1–15.
[4] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
[5] F. Hörst, et al., Cellvit: Vision transformers for precise cell segmentation and classification, Med. Image Anal. 94 (2024) 103143.
[6] M. Stevens, A. Nanou, L.W. Terstappen, C. Driemel, N.H. Stoecklein, F.A. Coumans, StarDist image segmentation improves circulating tumor cell detection, Cancers 14 (12) (2022) 2916.
[7] P. Naylor, M. Laé, F. Reyal, T. Walter, Segmentation of nuclei in histopathology images by deep regression of the distance map, IEEE Trans. Med. Imaging 38 (2) (2018) 448–459.
[8] H. He, et al., Cdnet: Centripetal direction network for nuclear instance segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4026–4035.
[9] S. Nofallah, et al., Machine learning techniques for mitoses classification, Comput. Med. Imaging Graph. 87 (2021) 101832.
[10] H. Chen, X. Qi, L. Yu, P.-A. Heng, DCAN: deep contour-aware networks for accurate gland segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2487–2496.
[11] S. Chen, C. Ding, M. Liu, J. Cheng, D. Tao, CPP-net: Context-aware polygon proposal network for nucleus segmentation, IEEE Trans. Image Process. 32 (2023) 980–994.
[12] Y. Zhou, O.F. Onder, Q. Dou, E. Tsougenis, H. Chen, P.-A. Heng, Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation, in: Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26, Springer, 2019, pp. 682–693.
[13] S.E.A. Raza, et al., Micro-net: A unified model for segmentation of various objects in microscopy images, Med. Image Anal. 52 (2019) 160–173.
[14] I. Glahn, et al., Automated nuclear morphometry: A deep learning approach for prognostication in canine pulmonary carcinoma to enhance reproducibility, Vet. Sci. 11 (6) (2024) 278.
[15] K.J. Pienta, D.S. Coffey, Correlation of nuclear morphometry with progression of breast cancer, Cancer 68 (9) (1991) 2012–2016.
[16] J.P. Baak, H. Van Dop, P.H. Kurver, J. Hermans, The value of morphometry to classic prognosticators in breast cancer, Cancer 56 (2) (1985) 374–382.
[17] K. Mei, A. Jiang, J. Li, M. Wang, Progressive feature fusion network for realistic image dehazing, in: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14, Springer, 2019, pp. 203–215.
[18] K. Yao, K. Huang, J. Sun, A. Hussain, PointNu-Net: Keypoint-assisted convolutional neural network for simultaneous multi-tissue histology nuclei segmentation and classification, IEEE Trans. Emerging Top. Comput. Intell. 8 (1) (2023) 802–813.
[19] A. Vaswani, et al., Attention is all you need, Adv. Neural Inf. Proces. Syst. 30 (2017).
[20] A. Dosovitskiy, et al., An image is worth 16x16 words: Transformers for image recognition at scale. arXiv Preprint, 2020.
[21] H. He, et al., Toposeg: Topology-aware nuclear instance segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 21307–21316.
[22] Z. Chen, Q. Xu, X. Liu, Y. Yuan, Domain-adaptive self-prompt segmentation for universal nuclei images, Med. Image Anal. (2025) 103607.
[23] R.J. Chen, et al., Scaling vision transformers to gigapixel images via hierarchical self-supervised learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16144–16155.
[24] H. Wang, P. Cao, J. Yang, O. Zaiane, Narrowing the semantic gaps in u-net with learnable skip connections: The case of medical image segmentation, Neural Netw. 178 (2024) 106546.
[25] Z. Zhang, X. Lu, G. Cao, Y. Yang, L. Jiao, F. Liu, ViT-YOLO: Transformer-based YOLO for object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2799–2808.
[26] M. Heidari, et al., Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023.
[27] H. Cao, et al., Swin-unet: Unet-like pure transformer for medical image segmentation, in: European Conference on Computer Vision, Springer, 2022, pp. 205–218.
[28] J. Chen et al., "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.
[29] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, D. Zhang, Ds-transunet: Dual swin transformer u-net for medical image segmentation, IEEE Trans. Instrum. Meas. 71 (2022) 1–15.
[30] X. Huang, Z. Deng, D. Li, X. Yuan, Y. Fu, MISSFormer: An effective transformer for 2D medical image segmentation, IEEE Trans. Med. Imaging (2022).
[31] Q. Liu, C. Kaul, J. Wang, C. Anagnostopoulos, R. Murray-Smith, and F. Deligianni, "Optimizing vision transformers for medical image segmentation," in ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2023: IEEE, pp. 1-5.
[32] S.A. Mohammed, S. Shabani, M. Sohaib, C. Nicolescu, M. Helen, and B. Parvin, "Logsage: Log-Based Saliency for Guided Encoding in Robust Nuclei Segmentation of Immunofluorescence Histology Images," in *2025* IEEE 22nd International Symposium on Biomedical Imaging (ISBI), 2025: IEEE, pp. 1–5.
[33] S. Shabani, M. Sohaib, S.A. Mohamed, B. Parvin, "Coupled Swin Transformers and Multi-Apertures Network (CSTA-NET) Improves Medical Image Segmentation," in 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), 2025: IEEE, pp. 1-5.
[34] R. Azad, et al., Advances in medical image analysis with vision transformers: a comprehensive review, Med. Image Anal. (2023) 103000.
[35] Q.D. Vu, et al., Methods for segmentation and classification of digital microscopy tissue images, Front. Bioeng. Biotechnol. 7 (2019) 433738.
[36] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, A. Sethi, A dataset and a technique for generalized nuclear segmentation for computational pathology, IEEE Trans. Med. Imaging 36 (7) (2017) 1550–1560.
[37] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9995409&tag=1.
[38] J.D. Nunes, D. Montezuma, D. Oliveira, T. Pereira, J.S. Cardoso, A survey on cell nuclei instance segmentation and classification: Leveraging context and attention, Med. Image Anal. (2024) 103360.
[39] R. Hollandi, N. Moshkov, L. Paavolainen, E. Tasnadi, F. Piccinini, P. Horvath, Nucleus segmentation: Towards automated solutions, Trends Cell Biol. 32 (4) (2022) 295–310.
[40] N. Alemi Koohbanani, M. Jahanifar, A. Gooya, N. Rajpoot, "Nuclear instance segmentation using a proposal-free spatially aware deep learning framework," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22, 2019: Springer, pp. 622–630.
[41] U. Schmidt, M. Weigert, C. Broaddus, G. Myers, Cell detection with star-convex polygons, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11, Springer, 2018, pp. 265–273.
[42] Q. Pu, J. Tian, D. Wei, Q. Shu, M. Sun, L. Zhao, Multifunctional aggregation network of cell nuclei segmentation aiming histopathological diagnosis assistance: A new MA-Net construction, PLoS One 19 (9) (2024) e0308326.
[43] Z. Wan, et al., CellT-net: a composite transformer method for 2-D cell instance segmentation, IEEE J. Biomed. Health Inform. 28 (2) (2023) 730–741.
[44] B. Li, et al., NuHTC: A hybrid task cascade for nuclei instance segmentation and classification, Med. Image Anal. (2025) 103595.
[45] J. Fan, D. Liu, H. Chang, W. Cai, Learning to generalize over subpartitions for heterogeneity-aware domain adaptive nuclei segmentation, Int. J. Comput. Vis. 132 (8) (2024) 2861–2884.
[46] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[47] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[48] A. Kirillov, K. He, R. Girshick, C. Rother, P. Dollár, Panoptic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9404–9413.
[49] E. Walsh, N.M. Orsi, The current troubled state of the global pathology workforce: a concise review, Diagn. Pathol. 19 (1) (2024) 163.