

Shape-supervised network for cross-domain nuclei instance segmentation

Sibei Chen

Hubei University of Technology

Zhiwei Ye

Hubei University of Technology

Taiyu Fan

Huazhong University of Science and Technology

Liye Mei

Hubei University of Technology

Junan Long

Tongji University

Die Luo

`luodie@hbut.edu.cn`

Hubei University of Technology

Research Article

Keywords: Shape-supervised, Cross-domain, Instance segmentation

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10940289/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Shape-supervised network for cross-domain nuclei instance segmentation

Sibei Chen¹, Zhiwei Ye², Taiyu Fan³, Liye Mei², Junan Long⁴,
Die Luo^{2*}

¹Detroit Green Technology Institute, Hubei University of Technology,
Wuhan, China.

²School of Computer Science, Hubei University of Technology, Wuhan,
China.

³Britton Chance Center for Biomedical Photonics, Wuhan National
Laboratory for Optoelectronics and MoE Key Laboratory for
Biomedical Photonics, School of Engineering Sciences, Huazhong
University of Science and Technology, Wuhan, China.

⁴College of Computer Science and Technology, Tongji University,
Shanghai, China.

*Corresponding author(s). E-mail(s): luodie@hbut.edu.cn;

Contributing authors: Anniechen0719@163.com; hgczyzw@hbut.edu.cn;
M202273565@hust.edu.cn; meiliye@hbut.edu.cn;
junanlong@tongji.edu.cn;

Abstract

Instance segmentation of nuclei in histopathological images is essential for the quantification and analysis of pathological features, which plays a critical role in accurate cancer diagnosis. Current deep learning techniques often aim to improve the detection of nucleus borders pixel by pixel, but this approach can miss key structural details that help identify each nucleus later. To address this issue, we propose a shape-supervised network named SSNet, which effectively integrates semantic segmentation with instance recognition by incorporating a nucleus shape model. Specifically, we employ a sparse Gaussian mixture model (GMM) to identify distinct nucleus regions and subsequently utilize the discrepancy between these regions and the expected oval shapes as an additional shape loss to guide the training of UNet. This shape-supervised approach not only restricts the segmentation network’s output but also makes its predictions

more trustworthy. These enhancements facilitate accurate instance segmentation of nuclei across cross-domain datasets. We evaluated our method on three publicly available tissue imaging datasets as well as private datasets derived from liquid-based cytology samples. Experimental results demonstrate that our method performs comparably to or outperforms existing approaches. All codes and trained models are available at <https://github.com/ldrunning/SSNet>.

Keywords: Shape-supervised, Cross-domain, Instance segmentation

1 Introduction

Cell state estimation and type identification have become an important part of digital pathology and can be helpful for rapidly assessing whether the sample is tumor, infected, or necrosis (Gurcan et al. 2009; Hou et al. 2019). This rapid assessment is usually achieved by nuclei segmentation which provides the quantitative nuclear morphological features, including shape, density, texture, and intensity. Advances in microscopic optical imaging and sample preparation techniques allow for observing numerous types of cell imaging datasets with high-resolution and large-scale and hence vastly promote the development of digital pathology, which produces massive images of cell morphology (Chang et al. 2012; Filipczuk et al. 2013; Mercan et al. 2017). Manual nuclei segmentation on these images is time-intensive and laborious. Accurate and robust automatic segmentation methods are needed to alleviate the situation (Jing et al. 2016).

Nuclei instance segmentation is a challenging task, mainly due to the difficulties of identifying densely packed nuclei and the situation that the generalization problem of the segmentation model still needs to be solved further (Mahbod et al. 2021; Kumar et al. 2017; Verma et al. 2021; Vu et al. 2019). Numerous nuclei instance segmentation methods have proposed effective strategies for addressing these two challenges. Several algorithms construct novel end-to-end deep networks that are directly trained on labeled instances (Chen et al. 2018; Xie et al. 2020; Bolya et al. 2019; Wang et al. 2020b). These models are capable of separating densely packed nuclei due to the powerful learning capabilities of deep neural networks. Furthermore, specific boundary and shape information has been incorporated into certain studies to further refine the segmentation results generated by deep networks (Han et al. 2023; Wang et al. 2022b). Several approaches employ graph neural networks to convert original images into graph-structured data that encapsulates rich semantic information. Subsequently, spatial-aware graph neural networks are applied to extract features from both the spatial and structural characteristics of the datasets, ultimately enabling precise semantic segmentation (Giraldo et al. 2022; Dong et al. 2024). Although these approaches achieve promising performance on specialized datasets, their segmentation accuracy tends to decline significantly when applied to other types of datasets. Some methods have been proposed to tackle the segmentation generalization issue. In studies (Chen et al. 2023; Sharma et al. 2022; Zou et al. 2022), the designed networks identify the nuclei regions and the related information, such as distance gradients, rather than directly

identify individual instances. By combining the distance gradients, instance segmentation can be automatically generated from the nuclei regions. Because the network is less complex, this segmentation network can be trained on multi-domain datasets and provides reliable instance segmentation. Another approach aims to enhance the throughput of instance annotation and train deep networks on large-scale datasets (Stringer et al. 2021; Greenwald et al. 2022). Although notable progress has been achieved, there remains a need to strike a balance between segmentation efficiency and network accuracy.

Here, we introduce a shape-supervised network (SSNet) specifically designed to improve the accuracy of nucleus segmentation. We assessed our approach using both public and private datasets, comparing its instance segmentation performance against several leading-edge methods. Additionally, we demonstrated the benefits of our method in cross-domain segmentation scenarios. The contributions of our proposed method are threefold:

We propose a novel segmentation pipeline that integrates semantic segmentation, instance identification, and the supervision of the semantic network with identified instances. This pipeline stands apart from existing instance segmentation approaches, such as end-to-end networks and cascades of semantic segmentation and instance identification.

We offer an alternative strategy for instance segmentation, where the nuclei region, as output by the network, is approximated by a series of elliptical regions, each corresponding to a nucleus region.

We enhance cross-domain instance segmentation by employing a parameter model to constrain the segmentation network’s output.

2 Related Works

2.1 Nuclei Instance Segmentation

Nuclei instance segmentation methods can be grouped into two types: end-to-end methods and two-stage methods. The end-to-end methods train the deep segmentation network and directly output each nuclei region (Chen et al. 2017; Graham et al. 2019; He et al. 2021). The two-stage methods design a convolution network to segment nuclei regions and identify individual nuclei regions from the segmentation results (Naylor et al. 2018; Wan et al. 2020). In the DCAN network (Chen et al. 2017), two branch outputs are combined to obtain accurate instance segmentation results. One branch generates the semantic segmentation image of the target, and the other outputs the precise contour of the instance object. Hover-Net (Graham et al. 2019) outputs horizontal and vertical distances of nuclear pixels to their centers of mass, which are subsequently leveraged to separate clustered nuclei. CDNet (He et al. 2021) adds a branch network of centroid estimation based on Hover-Net, using the centripetal direction feature to divide overlapping nuclei. DIST (Naylor et al. 2018) separates the touching nuclei by formulating the segmentation problem as a regression task of the distance map. ASPPU-Net (Wan et al. 2020) identifies the overlapping nuclei by detecting concave points in foreground segmentation image. These identifications are independent of the semantic segmentation of the deep network and can’t guide the

semantic network training. Our pipeline builds the relationship between the identification of individual nuclei and the semantic segmentation network, which reduces difficulties in dividing each nuclei region from the output of the segmentation network.

2.2 Instance Segmentation with Prior Information

Integrating prior information into the deep networks can boost the segmentation performance. CIA-Net (Zhou et al. 2019) considers the dependence between the nuclei and their contours and introduces a contour-aware informative aggregation module in the network. Star-DIST (Schmidt et al. 2018) locates the nuclei through a priori star-convex polygon rather than the bounding box. ESE-Seg (Xu et al. 2019) adopts a novel top-down instance segmentation framework based on explicit shape encoding to achieve real-time instance segmentation of the target. Yan (Xu et al. 2017) proposes shape preservation loss and simultaneously facilitates pixel-wise gland segmentation and boundary detection training. The above prior information is artificially summarized and cannot fully describe the nuclei shape. Our method employs a parametric shape model to provide explicit supervision for the nuclei segmentation network. During training, the model minimizes the discrepancy between the predicted nuclei shapes and the ground-truth shapes defined by the parametric model.

2.3 Cross-Domain Instance Segmentation

In general, oversized data training and cross-domain transfer learning are two common ways to improve the generation ability of the segmentation network. Typically, Cellpose (Stringer et al. 2021) and TissueNet (Greenwald et al. 2022) are trained on massive cell datasets and can accurately segment cells from various image types without further training or parameter adjustment of the model. Cross-domain transfer learning is used to reduce the diversity of the network input data (Liu et al. 2020; Reis and Turk 2023; Hollandi et al. 2020). Multi-domain data is usually converted into the target domain data and then input to the deep network to obtain the instance nuclei segmentation. In PDAM (Liu et al. 2020), the domain discriminator bridges the domain gap at the context level, realizing cross-domain panoramic segmentation through adaptive integration of semantic and instance features. Hatice (Reis and Turk 2023) collects a multi-category large colon cancer dataset MedCLNet and has conducted transfer learning on it, which improves the robustness and segmentation accuracy of the model. nucleAIzer (Hollandi et al. 2020) can automatically match unseen and unlabeled data with a nuclei-style model using the image-style transfer to generate augmented training samples, making the nuclei segmentation relatively simple and labor free. Our method uses the prior shape information suitable for most nuclei to constrain the network output and realizes the cross-domain nuclear instance segmentation.

3 Methods

The shape-supervised network contains three modules: segmentation network, sparse GMM clustering, and ellipse region matching. Figure 1 presents a schematic illustration of the architecture of the shape supervision network. The segmentation network is a UNet (Ronneberger et al. 2015) structure that extracts nuclei regions from images. Sparse GMM (Arellano and Dahyot 2016; Zhao et al. 2021) clustering identifies individual nuclei from the extracted regions. Each identified nuclei shape matches the elliptical region by ellipse fitting. The matching degree is regarded as an additional network loss function to supervise the segmentation network, which approaches the segmented region by a series of elliptical regions as much as possible. Here, we use a shape model to constrain the network output, and a relatively complex way to calculate the loss function. These considerations are inspired by the work (Raissi et al. 2020) that use the partial equation to constrain the network output and the work (Wang et al. 2020a) in which loss function are calculated by solving an optimal problem.

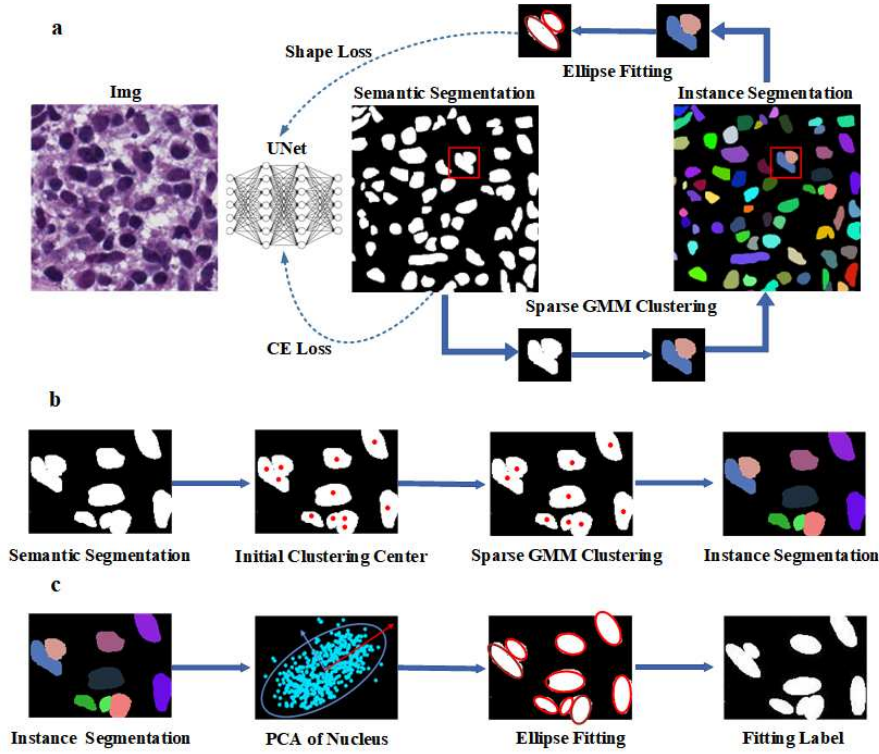


Fig. 1 The scheme of the shape-supervised network. (a) The shape-supervised network includes segmentation network, sparse GMM clustering, and elliptical region matching. CE loss refers to the difference between nuclear foreground segmentation and ground truth. The shape loss refers to the matching degree between identified nuclei shape and the elliptical shape. (b) The procedure to identify the isolated nuclei from the segmented region. (c) The procedure to calculate the shape loss.

3.1 Semantic segmentation network

The semantic segmentation network is designed to identify the nuclei regions from images. Its structure is based on UNet with encoding and decoding stages (Ronneberger et al. 2015; Wang et al. 2022a; Raza et al. 2019). In the encoding stage, five convolution layers and five pooling (down-sampling) layers are included. Correspondingly, the decoding stage has five convolution layers and five up-sampling layers to ensure that each phase of the network has the same size to the encoding stage. To supervise the semantic segmentation network, a combination of binary cross entropy (CE) loss and shape loss is used.

$$L_{total} = Loss_{CE}(y, \hat{y}) + \lambda Loss_{shape}(\hat{y}) \quad (1)$$

where $y \in \{0, 1\}$ is the label, $\hat{y} \in [0, 1]$ is the prediction probability and λ is set to 0.2. Shape loss is generated by a combination of sparse GMM clustering and ellipse fitting. Minimizing shape loss indicates that the segmented region (nuclei regions) can be approximated by a series of elliptical regions, with each ellipse corresponding to a nucleus region, thereby achieving instance segmentation of nuclei.

3.2 Sparse GMM clustering

The Gaussian Mixture Model (GMM) is recognized for its strong representational capacity, allowing for relatively accurate modeling of image content; in other words, pixel values can be approximated using a linear combination of multiple Gaussian functions (Arellano and Dahyot 2016; Zhao et al. 2021). In this study, the proposed sparse GMM method employs a clustering strategy to partition multiple connected cell nucleus regions detected by the network into several distinct single nucleus regions. Typically, connected regions composed of multiple nuclei exhibit numerous concave boundary points.

Following processing with sparse GMM, each resulting region displays an approximately elliptical shape and predominantly retains a convex structure. Individual nucleus identification comprises three main stages: connected region extraction, seed selection to provide initial conditions for the model, and sparse GMM clustering. Sparse GMM clustering is applied to identify individual nucleus regions from the results of image segmentation, specifically the output generated by the semantic segmentation network.

The region growing method is used to extract connected regions, in which the positions of all pixels form a point set denoted as P . For simplicity, P also refers to the extracted connected region. For each $p_i \in P$, we calculate the minimal distance between its position and the boundary points of the connected region denoted by E .

$$d(p_i) = \min_{p \in E} \|p_i - p\| \quad (2)$$

By this calculation, the connected region is transformed into a surface function $\{p_i, d(p_i)\}$. The initial seeds are regarded as the local maximum points of the surface

function and can be calculated as below:

$$c_k = \arg \max_{p \in \delta(c_k)} d(p) \quad (3)$$

Here, $\delta(c_k)$ represents the 8-point neighborhood of c_k . The number of points in P and the initial seeds are denoted by N and K respectively.

After obtaining the initial seeds in connected region P , sparse GMM clustering is employed for dividing P into several sub-regions. Each sub-region corresponds to the region of a single nuclei. Sparse GMM is modeled as:

$$\ln P \left(\{(\pi_k, \mu_k, \Sigma_k)\}_{k=1, \dots, K} \right) = \sum_{i=1}^N \ln \sum_{k=1}^K \pi_k \phi(p_i | \mu_k, \Sigma_k) - \lambda \sum_{k=1}^K \text{tr}(\Sigma_k) \quad (4)$$

$$\sum_{k=1}^K \pi_k = 1 \quad (5)$$

$$0 \leq \pi_k \leq 1 \quad (6)$$

Here, $\phi(\cdot | \mu_k, \Sigma_k)$ represents a Gaussian probability density function with mean μ_k and covariance matrix Σ_k . $\text{tr}(\cdot)$ represents the trace of a matrix and λ is a trade-off parameter that controls the sparsity of $\text{tr}(\Sigma_k)$. Dividing connected region P into K sub-regions can be achieved by maximizing the sparse GMM model, shown as:

$$(\pi_k^*, \mu_k^*, \Sigma_k^*)_{k=1, \dots, K} = \arg \max \ln P \left(\{(\pi_k, \mu_k, \Sigma_k)\}_{k=1, \dots, K} \right) \quad (7)$$

We employ the EM algorithm to solve the above optimization problem. In the EM iteration, the initial values of π_k , μ_k , and Σ_k are set to $\frac{1}{K}$, c_k , and 2×2 identity matrix, respectively. For each $p_i \in P$, its posterior probability can be calculated as below:

$$\gamma_{i,k} = \begin{cases} \frac{\pi_k \phi(p_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \phi(p_i | \mu_j, \Sigma_j)} & \text{if } \text{tr}(\Sigma_k) > \text{thre} \\ 0 & \text{if } \text{tr}(\Sigma_k) \leq \text{thre} \end{cases} \quad (8)$$

Then, p_i is assigned the class that corresponds to its maximum posterior probability. *thre* is set to 0.25, which is determined by experimental experience. The above procedure for dividing the connected region P is also expressed by pseudo codes in Algorithm 1.

3.3 Elliptical region matching

Sparse GMM clustering generates the isolated nuclei regions from the semantic segmentation of an image. Specifically, if a single nucleus region is convex, its boundary is relatively close to that of an ellipse. In this case, applying principal component analysis to this region can derive an ellipse region. The image of the nucleus region is similar to the derived ellipse region, and the matching degree between these two regions is high. Conversely, if the single nucleus region has a non-convex shape, the ellipse

Algorithm 1 Individual Nuclei Identification

Require: Surface function of connected regions $\{p_i, d(p_i)\}$

Require: Initial seeds c_k as local maxima of the surface function

Require: Minimum parameters' adjustment $\Delta_0 \leftarrow 1e^{-7}$

Ensure: Label of point which maximizes the posterior probability

```
1: Parameters initialization:
2:  $\pi_k^0 \leftarrow 1/K$ 
3:  $\mu_k^0 \leftarrow c_k$ 
4:  $\Sigma_k^0 \leftarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ 
5: while  $\Delta^{(n)} > \Delta_0$  do
6:   for  $k \leftarrow 1$  to  $K$  do
7:     E-step
8:     for  $i \leftarrow 1$  to  $N$  do
9:       Calculate posterior probability  $\gamma_{i,k}$  of each point belonging to each
       component:
10:      if  $\text{tr}(\Sigma_k) > \text{thre}$  then
11:         $\gamma_{i,k} \leftarrow \frac{\pi_k \varphi(p_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \varphi(p_i | \mu_j, \Sigma_j)}$ 
12:      else
13:         $\gamma_{i,k} \leftarrow 0$ 
14:      end if
15:    end for
16:    M-step
17:    Adjust parameters to minimize log-likelihood:
18:     $\ln P(\{\pi_k, \mu_k, \Sigma_k\}_{k=1, \dots, K}) = \sum_{i=1}^N \ln \sum_{k=1}^K \pi_k \phi(p_i | \mu_k, \Sigma_k) - \lambda \sum_{k=1}^K \text{tr}(\Sigma_k)$ 
19:    Update parameters:
20:     $\mu_k^{(n+1)} \leftarrow \frac{\sum_{i=1}^N \gamma_{i,k}^{(n)} p_i}{\sum_{i=1}^N \gamma_{i,k}^{(n)}}$ 
21:     $\Sigma_k^{(n+1)} \leftarrow \frac{\sum_{i=1}^N \gamma_{i,k}^{(n)} (p_i - \mu_k^{(n)})(p_i - \mu_k^{(n)})^T}{\sum_{i=1}^N \gamma_{i,k}^{(n)}} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 
22:     $\pi_k^{(n+1)} \leftarrow \frac{\sum_{i=1}^N \gamma_{i,k}^{(n)}}{N}$ 
23:  end for
24:  Calculate adjustment of all parameters  $\Delta^{(n+1)}$ 
25: end while
26: Assign each point according to posterior probability
```

region derived through principal component analysis will inevitably be convex (Zhu et al. 2025; Hu et al. 2025). In this situation, the matching degree between these two regions is low. The matching degree is the similarity between the isolated nuclei segmentation region and the matched ellipse. Let X denote the positions of points in an isolated nuclei region, and μ, C denote the mean and the covariance matrix of X . The elliptical region of S is defined as:

$$S = \{\mathbf{x} \mid (\mathbf{x} - \mu)^\top C^{-1}(\mathbf{x} - \mu) \leq 2\} \quad (9)$$

The ellipse matching degree is calculated as: Here, $\text{num}()$ represents the total number of points in a point set. Note that the covariance matrix C is a positive-definite matrix and thus S is an elliptical region. We assume the segmented image includes L isolated nuclei regions. The shape loss can be calculated below:

$$\text{Loss}_{\text{shape}} = \frac{L - \sum_{i=1}^L m_{x_i}}{L} \quad (10)$$

In fact, by sparse GMM clustering and ellipse matching, the segmented image, i.e., the output of the semantic segmentation network, can generate its corresponding image that includes a series of oval-shaped regions. Shape loss can be regarded as the difference between the segmented image and its generated image. When isolated nuclei regions are oval-shaped, their matching degrees are close to 1, and thus shape loss approaches 0. Shape loss can be back-propagated to supervise the semantic segmentation network.

4 Experiments

4.1 Datasets and evaluation metrics

We have verified the effectiveness of the method on the public datasets and private datasets. The public datasets include MoNuSeg (Kumar et al. 2017), CoNSeP (Graham et al. 2019), and CPM17 (Vu et al. 2019). The private cervical glandular nuclei datasets are provided by Tongji Medical College, Wuhan National Laboratory for Optoelectronics, Hubei Provincial Women and Children’s Hospital, and Wuhan Union Medical College Hospital. Correspondingly, these four groups of datasets are denoted by TJ, WN, HB, and UN. Both public datasets are multi-organ nuclear instance segmentation datasets, which are currently one of the largest manually labeled cell instance segmentation databases. MoNuSeg is taken from 9 different organs and has 44 images. Of these, 30 are training sets, and 14 are test sets. CPM17 is taken from four different organs and has 64 images, 32 of which are used as training sets and 32 as test sets. The CoNSeP dataset contains only one type of tissue but covers six cell categories, with a total of 41 images, of which 27 are used for the training set and 14 for the test set. In TJ datasets, a total of 705 images are labeled and cross-verified by two pathologists, of which 500 are training sets, and 205 are test sets. A total of 6908 isolated nuclei are labeled. The other three private datasets are used as test sets, containing 205 images. Some details in the public and private datasets are listed in Table 1.

Table 1 The description of testing datasets

Datasets	Number of images	Labelled nuclei	Magnification	Number of organ types	Size
MoNuSeg	44	28,846	40	9	1000×1000
CPM-17	64	7,570	40, 20	4	600×600
CoNSeP	41	24139	40	1	1000×1000
TJ	705	6,908	40	1	512×512
WN	205	1,594	40	1	512×512
HB	205	1,742	40	1	512×512
UN	205	1,686	40	1	512×512

To verify the effectiveness of this method, we adopted three evaluation metrics commonly used in public datasets (Mahbod et al. 2021; Kumar et al. 2017; Verma et al. 2021; Vu et al. 2019), the Dice coefficient (Dice), Aggregated Jaccard Index (AJI), and Panoptic Quality (PQ), to evaluate this method.

4.2 Implementation and training details

We conduct all experiments on PyTorch 1.6.0 with Nvidia GeForce GTX 1080Ti. We use mini-batch stochastic gradient descent (SGD) with a momentum 0.9, and weight decay 5e-5. The batch size is set to 4, and the training epoch is set to 1200. The initial learning rate is 0.001, and it is decayed by cosine annealing with a period of 1200. The total number of parameters of this network is 31.25M. With GPU acceleration enabled during the instance recognition process, the entire network can achieve a response within seconds for images with a resolution of 1000×1000.

During the training process, the training data are augmented to improve the stability of the model. First, we randomly scale the input images; the scaling range is between 0.5 and 2. After that, randomly select one of the Gaussian blur and the Median blur to change the texture of the images. Then, we randomly select 50% of the images to perform a horizontal flip and 50% of them to perform a vertical flip. Finally, the images are normalized. Throughout the network optimization process, the network first generates semantic segmentation results and calculates the cross-entropy loss based on them; then, instance recognition is performed based on these semantic segmentation results to obtain the elliptical constraint loss. Both loss functions are simultaneously introduced into the network for joint optimization. Among them, the weight of the cross-entropy loss is relatively higher, so in the early stage of network training, the convergence of the model is mainly dominated by the cross-entropy loss; as the training process progresses, the influence of the elliptical constraint loss gradually increases and eventually plays a key role in the convergence stage.

4.3 Ablation experiments

We performed the ablation experiments on the public datasets MoNuSeg. In the testing set of MoNuSeg, the images are included from seven different kinds of tissue slices.

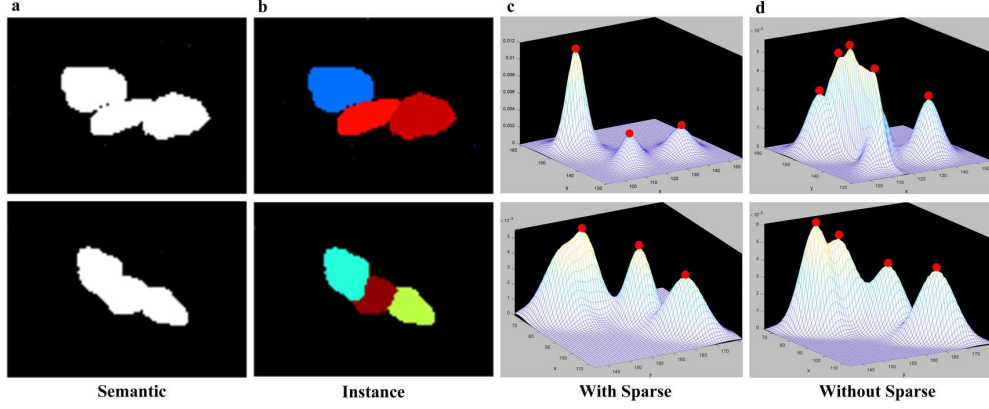


Fig. 2 The performance comparison of GMM with and without sparse constraint on identifying individual nuclei. (a) The segmented region of several touched nuclei. (b) The individual nuclei are identified with sparse GMM and labeled with different colors. (c) The probability density map is generated by sparse GMM, and the red dot represents the cluster center. (d) The description is similar to (c) for GMM.

4.3.1 The influence of sparsity constraints on GMM clustering

To verify that the sparsity constraint in GMM clustering enhances clustering accuracy, we conducted two sets of comparative experiments. In one set, sparse GMM clustering was applied, enabling the elimination of some initial seed points during the clustering process. In the other set, standard GMM clustering without sparsity constraints was used, where the number of clusters obtained is equal to the number of initial seed points. The part of method outlines a method for determining initial seed points. Due to the irregular shapes of nuclei, particularly their adhesion and overlap, the outer contours of connected domains often exhibit numerous corners, resulting in an overestimation of initial seed points relative to the actual number of nuclei. However, by incorporating sparsity constraints into GMM clustering, these constraints facilitate the merging of certain initial seed points, thereby aligning the final clustering count more closely with the true number of isolated nuclei.

As illustrated in Figure 2, for regions with adhered cell nuclei (Figure 2a), the number of clusters identified using sparse GMM clustering (Figure 2c) precisely matches the actual number of cell nuclei (Figure 2b). In contrast, standard GMM clustering without sparsity constraints leads to a higher cluster count due to inaccuracies in the initial seed points (Figure 2d). In addition, to provide a more rigorous quantitative evaluation of the differences between the two clustering methods, a comparative experiment was conducted using the MoNuSeg test dataset. The number of instance nuclei detected by each method was recorded, and the results are summarized in Figure 3. The findings indicate that, across tissue slices from seven different sources, the number of nuclei identified by sparse GMM clustering is more consistent with the ground truth compared to that obtained using standard GMM clustering. This suggests that the

incorporation of sparsity constraints effectively reduces over-segmentation by merging incorrectly partitioned clusters, thereby demonstrating the effectiveness of the proposed approach.

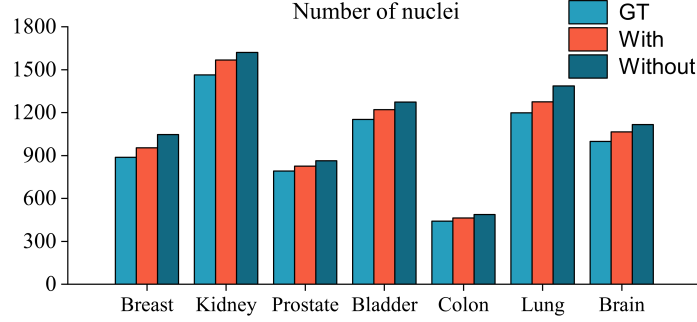


Fig. 3 The number of instances of nuclei segmented under sparse and non-sparse conditions.

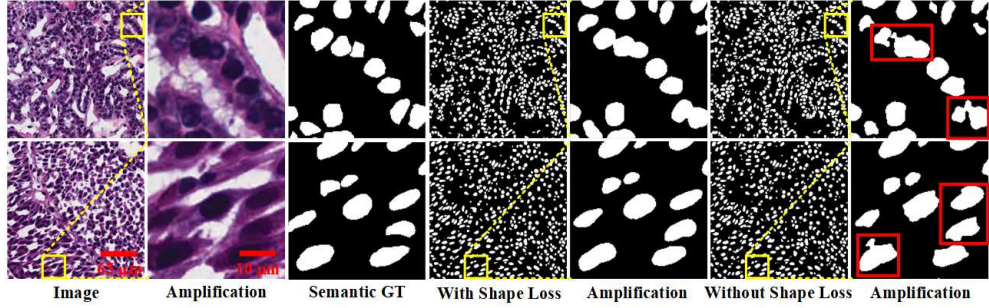


Fig. 4 A comparison of semantic segmentation results under conditions with and without shape constraints is presented. From left to right in the figure are the original image, the locally magnified region, the semantic segmentation label for the local area, the semantic segmentation result with shape supervision, a close-up of the shape-supervised segmentation result, the semantic segmentation result without shape supervision, and a close-up of the segmentation result without shape supervision. The red box in the seventh column highlights the non-smooth regions of the foreground segmentation.

4.3.2 The influence of shape supervision on semantic segmentation

We carried out a control experiment to assess the impact of shape supervision on the model’s semantic segmentation performance. In this series of experiments, we compared the semantic segmentation outcomes achieved with and without shape supervision in our segmentation pipeline. Shape supervision is invalid by setting Lambda to zeros in Equation (1). As shown in the comparison between the fifth and seventh columns of Figure 4, both segmentation networks effectively segmented the nuclei foreground, accurately capturing all nuclei regions without omissions or misclassifications. The network that included shape supervision delivered smoother

foreground segmentation for nuclei clusters. More precisely, isolated nuclei were segmented into shapes that closely resembled ellipses, and clustered nuclei demonstrated smooth boundaries free from significant concavities or protrusions. Conversely, the network without shape supervision produced foreground contours marked by irregular sharp angles and indentations.

In Table 2, we quantified the results of this comparative experiment. When evaluating tissue data from seven types of organs in the MoNuSeg dataset, it was found that, compared to the network supervised solely by cross-entropy, the network supervised by both shape and cross-entropy could significantly enhance segmentation accuracy. This improvement in segmentation accuracy is attributed to the introduction of a prior shape supervision mode, which reduces the likelihood of abrupt changes at the boundaries of nuclear regions.

Table 2 The semantic segmentation accuracy and ellipse matching degree are derived from the segmentation networks with shape supervision and without shape constraints, respectively. GT represents the true average elliptical matching degree of the cell nucleus.

Organ	Dice (%)		GT	Ellipse matching degree (%)	
	With	Without		With	Without
Breast	85.71	83.63	88.69	84.59	83.15
Kidney	85.33	84.21	85.14	84.71	74.29
Prostate	81.71	81.64	85.13	82.13	71.39
Bladder	88.05	87.15	88.75	88.43	85.36
Colon	83.06	82.18	84.74	83.27	80.71
Lung	83.61	81.67	88.15	87.70	75.59
Brain	82.39	81.51	87.24	83.01	80.21

Furthermore, we compared the degree of matching between the nucleus instance segmentation results obtained with and without shape supervision and elliptical regions. The results indicate that, despite varying degrees of deformation observed in the nuclei of cancerous cells, the average elliptical matching degree remains consistently above 85%, which suggests the elliptical matching degree can be used as a quantitative feature of the cell nucleus. Moreover, the nuclear regions segmented by the shape-supervised network exhibit a superior degree of elliptical alignment and are closer to the ground truth values. Essentially, shape supervision reduces the variability in the network’s outputs, ensuring that the shapes of the individually segmented nuclear regions closely approximate the ground truth, which is modeled as elliptical regions.

4.3.3 The influence of shape supervision on instance segmentation

We conducted a control experiment to evaluate the contribution of shape supervision to instance segmentation in the model. As confirmed by the ablation study in the previous section, the semantic segmentation network with shape supervision produces more rounded foregrounds for cell nuclei, with smooth contours free of irregular protrusions or depressions. In contrast, the network without shape supervision generates foregrounds with irregularities, leading to noticeable errors in instance segmentation. As illustrated in the comparison between the fifth and seventh columns of Figure 5, sparse GMM without shape supervision erroneously splits one nucleus into two regions

and incorrectly merges two distinct nuclei into one. Shape supervision effectively mitigates these issues, aligning the results more closely with the ground truth shown in the third column of the figure.

We also calculated the instance segmentation accuracy for both experimental groups on the test set. As shown in Table 3, compared with the unsupervised segmentation method, the instance segmentation accuracy for the seven types of tissue nuclei improved after incorporating shape supervision. Specifically, the closer the actual shape of the nuclei approximates an ellipse, the more significant the improvement in instance segmentation accuracy. Among the seven tissue types, the highest improvement reached 4%, with an average improvement of 2%. These results demonstrate that shape supervision is a critical component for achieving precise instance segmentation.

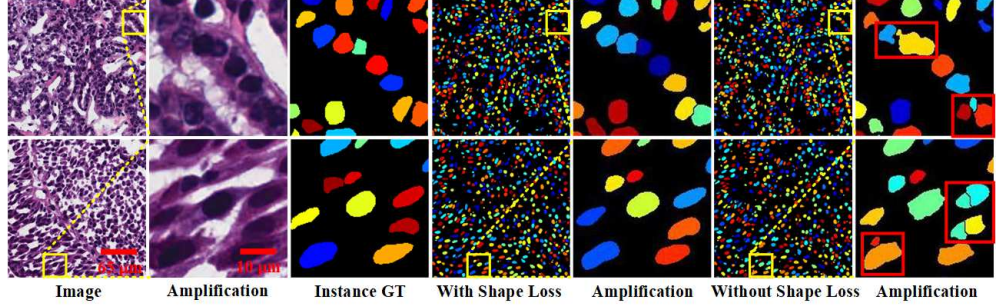


Fig. 5 A comparison of instance segmentation results under with and without shape supervision. The small images in the figure are similar to those in Figure 5, showing the instance segmentation results under the two conditions. The red box in the seventh column of the figure highlights the regions where instance recognition errors occurred.

Table 3 The instance segmentation accuracy is generated by networks with and without shape supervision.

Organ	AJI(%)	
	With	Without
Breast	63.71	61.16
Kidney	67.93	63.47
Prostate	69.28	68.97
Bladder	70.39	66.86
Colon	61.94	60.67
Lung	68.09	65.24
Brain	65.76	63.21

4.3.4 The influence of different shape supervision losses on segmentation results

To evaluate the impact of shape loss on the model, we incorporated two types of shape supervision losses during the training process. The first involves directly adding Dice

shape supervision loss to the semantic segmentation network, while the second employs elliptical matching shape loss that is back-propagated following nuclei instance recognition. As illustrated in the comparison between the fifth and seventh columns of Figure 6, both segmentation networks successfully segmented the foreground of the nuclei. However, the network with Dice loss exhibits mis-segmentation by incorrectly identifying part of the background as nuclei. While the segmented foreground lacks sharp corners or pits, its overall contour remains irregular and less smooth. In contrast, the network with elliptical matching loss produces more accurate segmentations, where isolated nuclei are closer to an elliptical shape, and the edges of adherent nuclei clusters are significantly smoother.

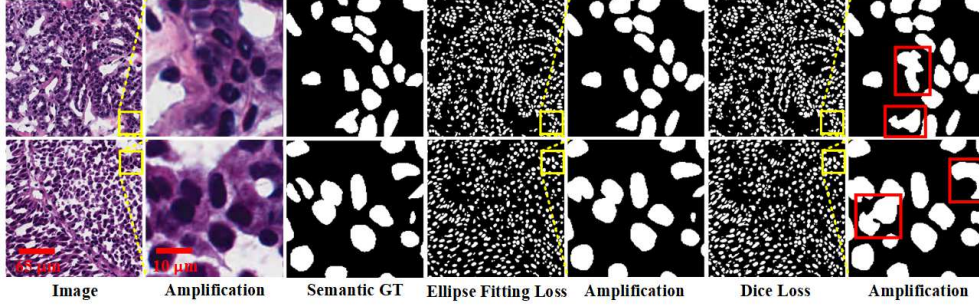


Fig. 6 A comparison of semantic segmentation results under two shape losses: elliptical matching loss and Dice loss. From left to right in the figure are the original image, the local magnified image, the semantic segmentation label, the local image semantic segmentation label, the semantic segmentation result under elliptical matching loss, the local image of the semantic segmentation result under elliptical matching loss, the semantic segmentation result under Dice loss, and the local image of the semantic segmentation result under Dice loss. The red box in the seventh column of the figure highlights the regions where foreground segmentation errors occur.

This comparative experiment also evaluated the instance segmentation results under two different shape losses. As shown in the seventh column of Figure 7, the segmentation network with elliptical matching loss achieved accurate and smooth nuclei semantic segmentation, enabling precise identification of isolated nuclei during the instance recognition process. The segmentation results were consistent with the true labels presented in the third column of the figure. In contrast, the segmentation network with Dice loss exhibited inaccurate foreground segmentation of the nuclei, leading to the erroneous splitting of one nucleus into two separate regions and the incorrect merging of two nuclei into one.

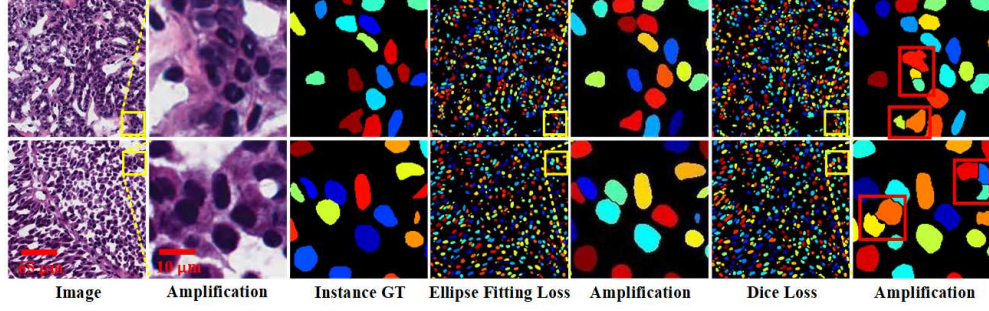


Fig. 7 A comparison of instance segmentation results under the elliptical matching loss and Dice loss. The small images in the figure are similar to those in Figure 6, showing the instance segmentation results under the two conditions. The red box in the seventh column of the figure highlights the regions where instance recognition errors occurred.

Finally, the semantic and instance segmentation accuracies of the two experimental groups were evaluated on the test set. As shown in Table 4, compared to the results obtained using Dice loss, the use of elliptical matching loss led to consistent improvements in both semantic and instance segmentation accuracies across the seven tissue types. Specifically, the average semantic segmentation accuracy improved by 1%, while the average instance segmentation accuracy increased by 2%. Notably, the enhancement in instance segmentation accuracy was more pronounced than that in semantic segmentation accuracy. A comparison of Tables 3 and 4 reveals that the segmentation network employing Dice loss achieved only a marginal improvement in semantic segmentation accuracy compared to the network without shape supervision, with little change observed in instance segmentation accuracy. These findings indicate that the segmentation network incorporating elliptical matching loss effectively integrates semantic segmentation and instance recognition, allowing these two tasks to mutually reinforce each other and thereby achieve superior overall segmentation performance.

Table 4 Semantic and instance segmentation accuracy under two shape loss conditions: elliptical matching loss and Dice loss

Organ	Dice(%)		AJI(%)	
	Ellipse Matching Loss	Dice Loss	Ellipse Matching Loss	Dice Loss
Breast	85.71	83.86	63.71	61.19
Kidney	85.33	84.74	67.93	63.53
Prostate	81.71	81.21	69.28	68.65
Bladder	88.05	87.42	70.39	66.93
Colon	83.06	82.14	61.94	60.53
Lung	83.61	81.93	68.09	66.11
Brain	82.39	81.88	65.76	63.37

4.3.5 Comparisons with state-of-the-art methods

We compared our method with the state-of-the-art methods on the MoNuSeg, CoNSeP, and CPM17 datasets. In this experiment, we compared not only end-to-end instance segmentation methods such as Mask-RCNN(He et al. 2017), DCAN(Chen et al. 2017), CIANet(Zhou et al. 2019), Hover-Net(Graham et al. 2019), StarDist(Schmidt et al. 2018), TopoSeg(He et al. 2023), CellViT++(Hörst et al. 2025), SEINET(Zhang et al. 2025), CDNet(He et al. 2021), and Meta-MTL(Han et al. 2022), but also typical two-stage segmentation methods such as Fiji(Schindelin et al. 2012), DIST(Naylor et al. 2018), and CellPose2(Pachitariu and Stringer 2022).

As presented in Table 6, our method attained segmentation performance of 84.28% Dice, 66.62% AJI, and 64.11% PQ on the MoNuSeg dataset; 86.89% Dice, 74.32% AJI, and 71.62% PQ on the CPM17 dataset; and 82.36% Dice, 55.63% AJI, and 52.07% PQ on the CoNSeP dataset. Compared to end-to-end deep learning approaches, our method demonstrates superior performance in cell nuclei instance segmentation over recently published methods such as TopoSeg and StarDist. Specifically, these competing methods require complex training labels, including instance-level annotations for each nuclei region and feature maps representing the centripetal direction of centroids, whereas our approach only requires semantic segmentation labels for the nuclei. Furthermore, when benchmarked against the classical segmentation method Mask-RCNN, our method exhibits a significant improvement of approximately 10% in segmentation accuracy. As illustrated in the seventh column of Figure 8, Mask-RCNN tends to produce numerous missed detections for dense small targets, while our method successfully segments these instances, as shown in the fifth column. Additionally, our method was compared with the state-of-the-art model CellPose2. On the MoNuSeg and CoNSeP datasets, our method exhibits semantic and instance segmentation accuracy that is approximately 1% lower than that of CellPose2. However, it outperforms CellPose2 on the CPM17 dataset. Considering the overall performance across these public datasets, our method achieves segmentation performance comparable to CellPose2, thereby demonstrating its competitiveness.

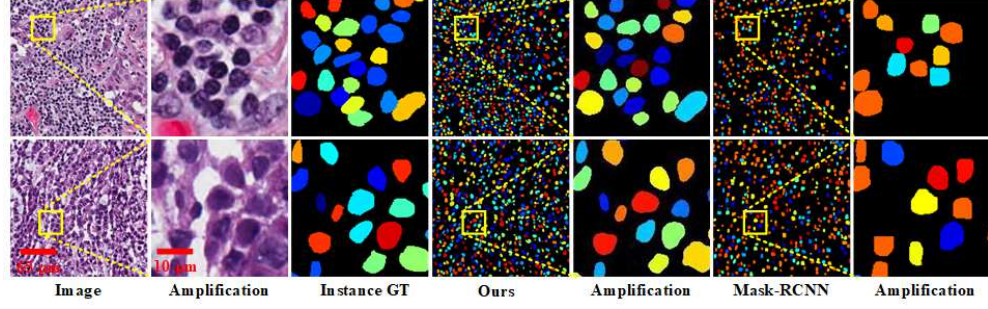


Fig. 8 The comparison of instance segmentation results between the shape-supervised instance segmentation method and Mask-RCNN on MoNuSeg. From left to right in the figure are the original image, the locally magnified image, the instance segmentation label for the local region, the instance segmentation result of our proposed method, the local view of the segmentation result from our method, the instance segmentation result of Mask-RCNN, and the local view of the segmentation result from Mask-RCNN. Notably, in the seventh column, Mask-RCNN exhibits a significant number of missed detections on densely distributed small targets.

Table 5 The comparison of test results between shape-supervised instance segmentation methods and mainstream approaches on public datasets.

Methods	MoNuSeg			CPM17			CoNSEP		
	Dice(%)	AI(%)	PQ(%)	Dice(%)	AI(%)	PQ(%)	Dice(%)	AI(%)	PQ(%)
UNet(Ronneberger et al. 2015)	83.62	59.71	47.83	84.73	66.48	57.81	79.61	48.97	34.54
Fiji(Schindelin et al. 2012)	-	53.00	-	65.00	-	-	47.33	-	-
DCAN(Chen et al. 2017)	79.20	52.50	49.23	82.80	56.10	54.52	73.32	28.93	25.64
DIST(Naylor et al. 2018)	78.90	55.90	44.31	82.60	61.60	50.42	80.42	50.24	39.83
Mask-RCNN(He et al. 2017)	76.02	54.60	50.93	82.13	67.53	67.44	70.46	47.42	46.08
Micro-Net(Raza et al. 2019)	79.70	56.03	51.92	85.70	66.80	66.13	79.46	52.73	44.94
CIA-Net(Zhou et al. 2019)	81.80	62.01	57.72	84.16	66.48	-	-	-	-
Hover-Net(Graham et al. 2019)	83.21	63.21	60.15	87.70	73.10	69.73	85.31	<u>57.14</u>	<u>54.72</u>
CDNet(He et al. 2021)	83.16	63.31	60.57	88.01	73.26	70.64	82.24	54.83	48.87
TopoSeg(He et al. 2023)	-	64.30	62.52	-	73.30	70.51	-	-	-
StarDist(Schmidt et al. 2018)	81.87	65.43	-	85.43	73.21	-	79.78	55.63	-
Meta-MTL(Han et al. 2022)	62.60	61.53	57.03	86.52	67.51	64.33	-	-	-
CellViT++(Hörst et al. 2025)	83.52	64.89	62.17	86.77	73.72	70.67	<u>84.54</u>	57.74	54.86
SEINET(Zhang et al. 2025)	84.18	64.12	60.62	87.89	72.87	69.64	83.42	56.64	51.16
CellPose2(Pachitariu and Stringer 2022)	84.95	67.54	66.41	85.64	73.27	71.53	82.67	55.43	52.58
Ours	<u>84.28</u>	<u>66.62</u>	<u>64.11</u>	86.89	74.32	71.62	82.36	55.63	52.07

Among the three public datasets, our method outperforms most mainstream methods on the MoNuSeg and CPM17 datasets, and its performance is comparable to that of CellPose2. However, on the CoNSEP dataset, although our method performs better than general algorithms, its performance is still weaker than that of methods such as Hover-Net, CellViT++, and CellPose2. The main reason for this difference is that the CoNSEP dataset contains some severely deformed necrotic cell nuclei, whose shapes deviate from the typical elliptical structure, leading to mis-segmentation during the model’s segmentation process. Considering the overall performance across the three datasets, our method still demonstrates strong competitiveness.

A comparison was also conducted with mainstream methods using a private dataset of clustered cervical nuclei. Similar to the experiments on public datasets, this group

of experiments also benchmarked against mainstream deep learning algorithms. In the private dataset test, our method achieved the highest semantic and instance segmentation accuracy. As shown in Table 6, our method achieved 90.38% Dice, 78.24% AJI, and 76.14% PQ, surpassing Mask-RCNN (86.01% Dice, 76.53% AJI, and 74.12% PQ) and CellPose2 (87.26% Dice, 76.13% AJI, and 73.18% PQ).

Although the complexity of the shapes of the cell nuclei in the private dataset is lower than that in public datasets, the private cervical dataset contains a large number of densely distributed nuclei with low contrast between the nuclei and cytoplasm, making segmentation particularly challenging. Despite deformations in the clustered cervical nuclei, the shapes of isolated nuclei remain close to ellipses, which aligns well with the design principle of our method. As illustrated in the fourth and fifth columns of Figure 9, our method successfully separates these nuclei. In contrast, Mask-RCNN, a classical instance segmentation method, tends to miss some nuclei in dense regions and incorrectly merges two nuclei into one. This highlights the superiority of our method in dense nuclei instance segmentation. Upon comparing the test outcomes of the private dataset with those of the public dataset, it becomes evident that the performance of our proposed method rivals that of mainstream networks on the public dataset, while notably outperforming them on the private dataset. The primary rationale behind this is that the private dataset comprises cell pathological images. Despite the cancer cells undergoing a certain degree of deformation, their overall shape remains relatively elliptical, aligning closely with the prior information utilized by our method. Consequently, this allows our approach to fully leverage its strengths, yielding more precise segmentation results.

Table 6 The comparison between our method and mainstream methods on private datasets.

Methods	TJ		
	Dice(%)	AJI(%)	PQ(%)
UNet(Ronneberger et al. 2015)	87.21	64.43	59.14
DCAN(Chen et al. 2017)	84.56	64.54	57.57
ASPPU-Net(Wan et al. 2020)	88.97	71.22	60.15
Mask-RCNN(He et al. 2017)	86.01	76.53	74.12
Micro-Net(Han et al. 2022)	87.75	74.32	71.14
Hover-Net(Graham et al. 2019)	87.09	75.80	75.16
CDNet(He et al. 2021)	87.43	75.97	74.23
StarDist(Schmidt et al. 2018)	85.59	75.78	72.16
SEINET(Zhang et al. 2025)	87.63	76.42	74.15
CellPose2(Pachitariu and Stringer 2022)	87.26	76.13	73.18
Ours	90.38	78.24	76.14

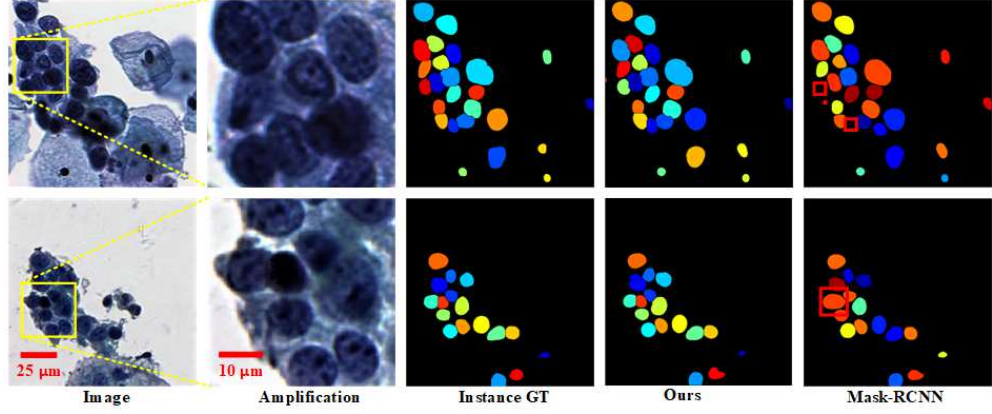


Fig. 9 The comparison of instance segmentation between the shape-supervised instance segmentation method and Mask-RCNN on the TJ dataset. From left to right in the figure are the original image, the magnified local image, the instance segmentation label, the instance segmentation result of our method, and the instance segmentation result of Mask-RCNN. The red box in the fifth column indicates the area where Mask-RCNN made a segmentation error.

4.3.6 Generalization performance of the shape-supervised network

This experiment was conducted to assess the generalization performance of the shape-supervised instance segmentation network. The MoNuSeg and TJ training datasets were combined into a cross-domain training dataset, and the networks were evaluated on this cross-domain dataset. In these two datasets, the nuclei in the MoNuSeg dataset are smaller in size but denser in distribution, while those in the TJ dataset are larger in size but more aggregated. Moreover, the cytoplasm in the MoNuSeg dataset is thicker and stained reddish, whereas in the TJ dataset it is sparser and stained bluish. These significant stylistic differences between the datasets impose stringent requirements on the generalization capability of the network. Five mainstream networks were selected for comparison with the shape-supervised instance segmentation network. The trained models were tested on the test sets of both datasets, and the results are summarized in Table 8. As shown in the table, our method achieved the highest semantic and instance segmentation accuracy, with 80.54% Dice and 63.66% AJI on the MoNuSeg dataset, and 88.38% Dice and 76.34% AJI on the TJ dataset. Compared with the separate training of the two datasets (Tables 6 and 7), the segmentation accuracy decreased slightly. CellPose2, a high generalization instance segmentation network, achieved segmentation accuracy comparable to ours. However, for other networks, cross-domain training resulted in a marked decrease in accuracy. For example, the high-precision segmentation network Hover-Net achieved only 57.15% AJI on the MoNuSeg dataset and 72.96% AJI on the TJ dataset.

Due to the morphological characteristics of nuclei and cytoplasm in the MoNuSeg dataset differing from those in the TJ dataset, the Hover-Net network failed to capture this distinction. As a result, on both test datasets, Hover-Net erroneously identified multiple aggregated nuclei as a single nuclei, as demonstrated in the seventh column

of Figure 10 and the fifth column of Figure 11. In contrast, our method successfully leveraged the prior shape information shared by all nuclei, thereby avoiding such misclassifications. Based on these findings, it can be concluded that incorporating prior shape constraints significantly enhances the accuracy and generalization capability of cross-domain segmentation.

Table 7 The comparison results between some mainstream methods and our method on cross-domain datasets.

Methods	MoNuSeg			TJ		
	Dice(%)	AJI(%)	PQ(%)	Dice(%)	AJI(%)	PQ(%)
DCAN(Chen et al. 2017)	73.20	48.93	45.68	78.11	60.62	52.16
Mask-RCNN(He et al. 2017)	72.00	52.09	48.66	83.32	73.97	73.15
Micro-Net(Raza et al. 2019)	75.50	51.23	47.43	84.93	71.35	69.53
Hover-Net(Graham et al. 2019)	75.63	57.15	55.78	85.75	72.96	72.11
SEINET(Zhang et al. 2025)	78.86	60.75	58.92	85.63	73.52	71.94
CellPose2(Pachitariu and Stringer 2022)	81.33	63.84	62.35	85.65	74.73	72.16
Ours	80.54	63.66	60.97	88.38	76.34	73.42

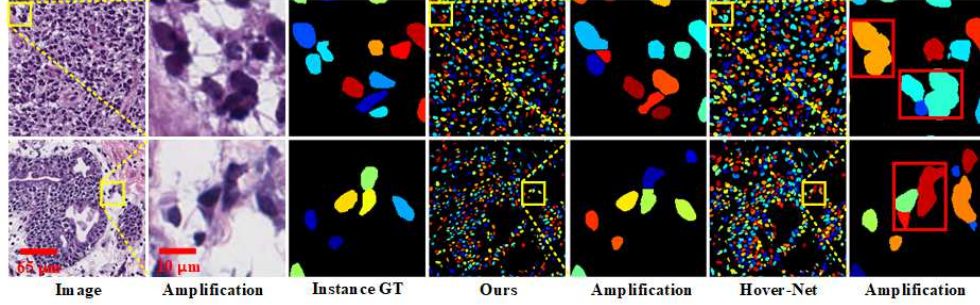


Fig. 10 The comparison of instance segmentation between shape-supervised instance segmentation method and Hover-Net on MoNuSeg. The small images in the figure are similar to those in Figure 8. The red boxes in the seventh column highlight the regions where Hover-Net incurred segmentation errors.

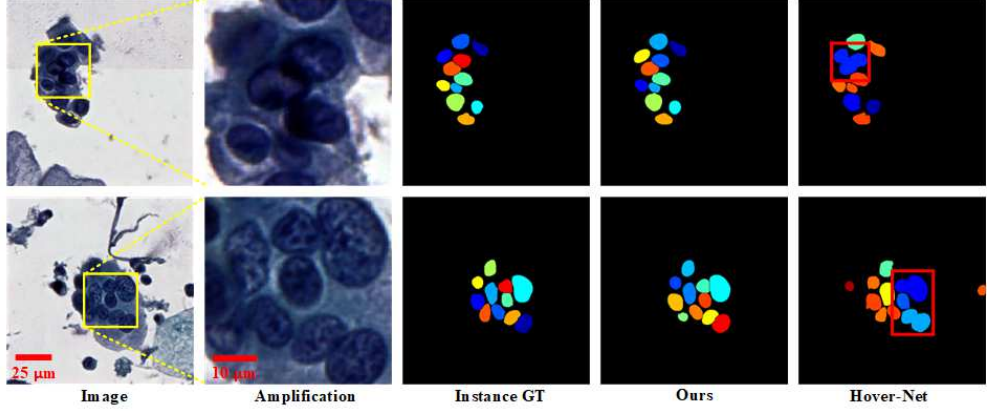


Fig. 11 The comparison of instance segmentation between shape-supervised instance segmentation method and Hover-Net on TJ. The small images in the figure are similar to those in Figure 9. The red boxes in the fifth column highlight the regions where Hover-Net incurred segmentation errors.

We further evaluated the generalization of our method on a variety of whole slide images with distinct characteristics. In this set of experiments, four datasets were selected from different institutions, namely TJ, WN, UN, and HB. These datasets are prepared with different methods, including the membrane, sediment, and natural sediments, which results in differences in image style. The images from the four test sets are illustrated in Figure 12a. Due to differences in data sources and preparation methods, significant style discrepancies exist among the four datasets. Specifically, notable variations are observed in features such as the color, contrast, and background brightness of the nuclei across the datasets. Figure 12b displays the numerical distribution of the datasets in the value channel of the HSV space, providing quantitative evidence of the marked differences in both the nuclei and image backgrounds. This model was trained exclusively on the TJ training set and evaluated on four test sets. The TJ dataset comprises 500 images for training and 205 images for testing, whereas the other three datasets consist solely of 205 test images each. As shown in Table 8, despite the significant style differences among the datasets, the segmentation accuracy of the model on the three untrained datasets closely matches that on the TJ dataset. Specifically, the error in semantic segmentation accuracy is less than 3%, and the error in instance segmentation accuracy is less than 2%. These results demonstrate that the trained segmentation model exhibits strong generalization capabilities when applied to untrained cervical data. This further confirms the robust generalization performance of our method.

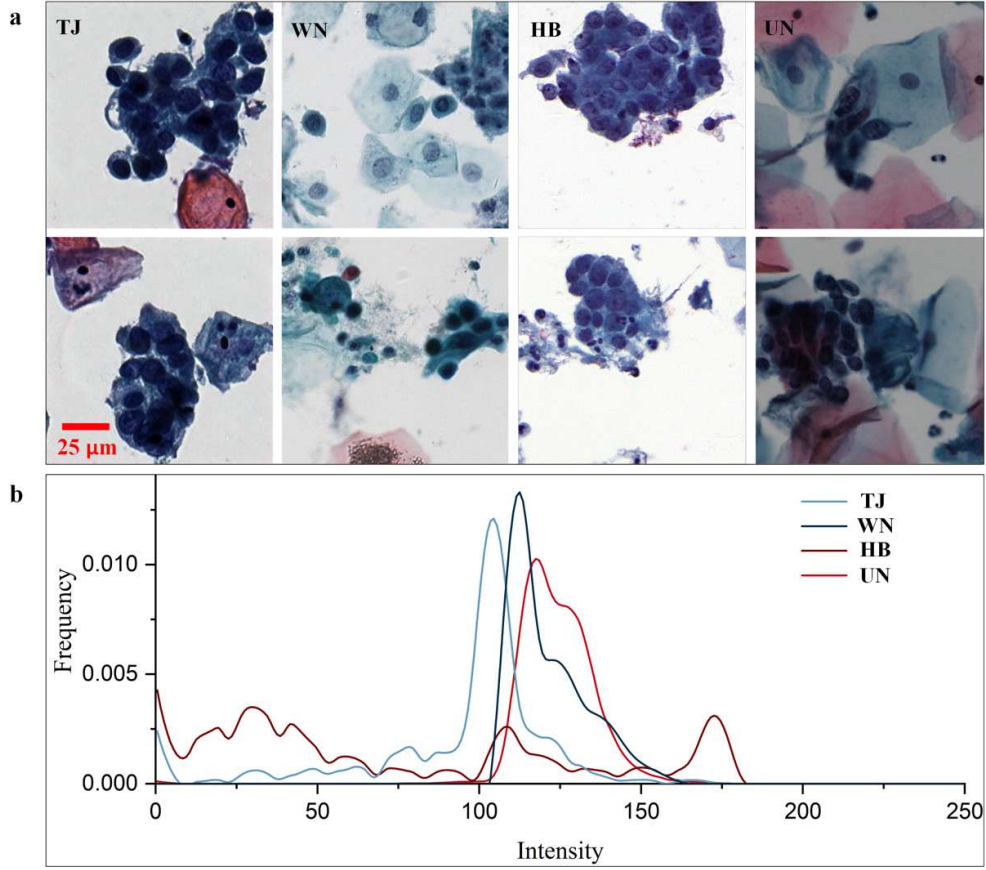


Fig. 12 Four different styles of cervical cell datasets. (a) A style diagram for the four datasets. (b) The numerical distribution of four datasets in the HSV spatial value channels (hue, saturation, value).

Table 8 Segmentation accuracy of our method on different datasets.

Datasets	Ours	
	Dice(%)	AJI(%)
TJ	90.38	78.24
WN	91.26	78.85
HB	89.67	77.13
UN	87.78	76.33

4.3.7 Test on whole slide images

To validate the segmentation capability of our method on cervical whole-slide images (WSIs), 100 WSIs were acquired from the Department of Pathology at Tongji Hospital for testing. From each WSI, 10 regions of interest (ROIs) were extracted, resulting in a test set comprising 1,000 ROIs (each sized 2500×2500 pixels). All cervical nuclei

in the test set were annotated, totaling 10,958 nuclear instances. Three instance segmentation networks, namely Mask-RCNN, Hover-Net, and CellPose2, which exhibited performance comparable to our method on the TJ dataset, were selected for comparison. All four instance segmentation networks were trained on the TJ dataset. The recall rate and precision of nuclear segmentation across the test set were calculated for each network. In Table 10, "returned nuclei" refers to the total number of cervical nuclei regions predicted by the networks, while "true nuclei" indicates the number of correctly identified cervical nuclei within these predictions. As shown in Table 10, the proposed method achieved the highest recall rate (98.86%) and precision (97.03%). Specifically, the recall rate was 9% higher than that of Mask-RCNN, and the precision was 4% higher than that of Hover-Net. Among the test results on 100 whole slide images, Mask-RCNN exhibited a significant number of missed detections, particularly for surface cells with smaller nuclei. In contrast, CellPose2 and Hover-Net produced a considerable number of false positives by misidentifying inflammatory cells or artifacts (e.g., black spots) as cervical nuclei. Compared to these three methods, our proposed method demonstrates clear advantages in terms of detection accuracy.

Table 9 Comparative results of four segmentation methods in nuclei classification accuracy on the test set.

Methods	Total cell nuclei	Returned cell nuclei	True cell nuclei	Precision	Recall
Ours	10958	11165	10833	97.03%	98.86%
Mask-RCNN(He et al. 2017)	10958	10225	9764	95.49%	89.11%
Hover-Net(Graham et al. 2019)	10958	10798	9995	92.56%	91.21%
CellPose2(Pachitariu and Stringer 2022)	10958	10895	10239	93.98%	93.44%

5 Conclusion and discussion

In this paper, we propose a nuclei instance segmentation method and introduce a novel processing pipeline. The pipeline comprises three key components: a segmentation network, instance identification, and shape constraint. Within the shape constraint module, shape parameters are derived from instance identification results, while simultaneously imposing constraints on the output of the segmentation network. This design ensures tight integration among the three components, enabling mutual reinforcement and forming a positive feedback loop. Furthermore, the incorporation of prior shape constraints provides strong supervision for the segmentation network, thereby enhancing segmentation accuracy on cross-domain datasets. These advantages are validated through comparative experiments conducted on multiple datasets.

Instance-encoding approaches represent a mainstream and widely adopted paradigm in cell segmentation. These methods typically involve generating distance or gradient fields via neural networks, followed by post-processing techniques to extract instance-level segmentation results. In contrast, our approach approximates the foreground regions produced by the network using a set of elliptical regions, with each ellipse corresponding to an individual cell. In this context, clustering serves merely as an auxiliary means to achieve this objective. This approach is conceptually aligned

with the Gauss Splating framework, where images can be represented using a set of Gaussian functions. Consequently, our method can incorporate strategies from Gauss Splating to further enhance both segmentation accuracy and computational efficiency.

Our method imposes stronger geometric constraints on the network output, which inherently limits the diversity of possible output configurations. As long as the nuclear regions do not significantly deviate from elliptical shapes, our method can accurately identify them. Experimental results on public datasets demonstrate that our approach achieves high segmentation accuracy across various image types. Moreover, we trained the model on data from a single hospital and successfully applied it to data from other institutions, indicating the method’s promising generalization capability.

It is also important to note that our clustering strategy primarily operates on the segmented foreground. For images with low contrast, as long as the foreground can be accurately identified by the segmentation network, precise instance segmentation results can still be obtained. This is due to the presence of an additional loss function dedicated to foreground identification, which works in conjunction with the shape loss to jointly supervise the training process.

Compared to other existing methods, our approach may encounter challenges in identifying cells with highly irregular shapes. Additionally, the segmentation network architecture is relatively simple and requires further refinement in future work.

Supplementary information. Not applicable.

Acknowledgements. Not applicable.

Declarations

Funding Not applicable.

Conflict of interest The authors declare no competing interests.

Ethics approval Please insert the ethics approval statement required by your study, if applicable.

Consent to participate Informed consent was obtained from all participants.

Consent for publication Not applicable.

Data availability The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability The code is available at <https://github.com/ldrunning/SSNet>.

Author contributions Sibe Chen: Methodology, Software, Validation, Writing—original draft. Taiyu Fan: Methodology, Software, Validation, Writing—review and editing. Liye Mei: Methodology, Validation, Writing—review and editing. Junan Long: Validation, Writing—review and editing. Zhiwei Ye: Validation, Writing—review and editing. Die Luo: Methodology, Validation, Writing—review and editing. All authors read and approved the final manuscript.

References

Arellano C, Dahyot R (2016) Robust ellipse detection with gaussian mixture models. *Pattern Recognition* 58:12–26

- Bolya D, Zhou C, Xiao F, et al (2019) Yolact: Real-time instance segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 9157–9166
- Chang H, Han J, Borowsky A, et al (2012) Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association. *IEEE transactions on medical imaging* 32(4):670–682
- Chen H, Qi X, Yu L, et al (2017) Dcan: Deep contour-aware networks for object instance segmentation from histology images. *Medical image analysis* 36:135–146
- Chen LC, Hermans A, Papandreou G, et al (2018) Masklab: Instance segmentation by refining object detection with semantic and direction features. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4013–4022
- Chen S, Ding C, Liu M, et al (2023) Cpp-net: Context-aware polygon proposal network for nucleus segmentation. *IEEE Transactions on Image Processing* 32:980–994
- Dong Y, Liu G, Yin Y, et al (2024) Superpixel-based hierarchical graph convolution for enhanced interpretability in semantic segmentation. In: 2024 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, pp 1615–1620
- Filipczuk P, Fevens T, Krzyżak A, et al (2013) Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. *IEEE transactions on medical imaging* 32(12):2169–2178
- Giraldo JH, Scarrica V, Staiano A, et al (2022) Hypergraph convolutional networks for weakly-supervised semantic segmentation. In: 2022 IEEE international conference on image processing (ICIP), IEEE, pp 16–20
- Graham S, Vu QD, Raza SEA, et al (2019) Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis* 58:101563
- Greenwald NF, Miller G, Moen E, et al (2022) Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology* 40(4):555–565
- Gurcan MN, Boucheron LE, Can A, et al (2009) Histopathological image analysis: A review. *IEEE reviews in biomedical engineering* 2:147–171
- Han C, Yao H, Zhao B, et al (2022) Meta multi-task nuclei segmentation with fewer training samples. *Medical Image Analysis* 80:102481
- Han Y, Lei Y, Shkolnikov V, et al (2023) An ensemble method with edge awareness for abnormally shaped nuclei segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4315–4325

- He H, Huang Z, Ding Y, et al (2021) Cdnet: Centripetal direction network for nuclear instance segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 4026–4035
- He H, Wang J, Wei P, et al (2023) Toposeg: Topology-aware nuclear instance segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 21307–21316
- He K, Gkioxari G, Dollár P, et al (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969
- Hollandi R, Szkalitsy A, Toth T, et al (2020) nucleaizer: a parameter-free deep learning framework for nucleus segmentation using image style transfer. *Cell systems* 10(5):453–458
- Hörst F, Rempe M, Becker H, et al (2025) Cellvit++: Energy-efficient and adaptive cell segmentation and classification using foundation models. *arXiv preprint arXiv:250105269*
- Hou L, Nguyen V, Kanevsky AB, et al (2019) Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern recognition* 86:188–200
- Hu J, Xia B, Chen B, et al (2025) Gaussiansr: High fidelity 2d gaussian splatting for arbitrary-scale image super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 3554–3562
- Jing J, Wan T, Cao J, et al (2016) An improved hybrid active contour model for nuclear segmentation on breast cancer histopathology. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE, pp 1155–1158
- Kumar N, Verma R, Sharma S, et al (2017) A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging* 36(7):1550–1560
- Liu D, Zhang D, Song Y, et al (2020) Pdam: A panoptic-level feature alignment framework for unsupervised domain adaptive instance segmentation in microscopy images. *IEEE Transactions on Medical Imaging* 40(1):154–165
- Mahbod A, Schaefer G, Bancher B, et al (2021) Cryonuseg: A dataset for nuclei instance segmentation of cryosectioned h&e-stained histological images. *Computers in biology and medicine* 132:104349
- Mercan C, Aksoy S, Mercan E, et al (2017) Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images. *IEEE transactions on medical imaging* 37(1):316–325

- Naylor P, Laé M, Reyat F, et al (2018) Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging* 38(2):448–459
- Pachitariu M, Stringer C (2022) Cellpose 2.0: how to train your own model. *Nature methods* 19(12):1634–1641
- Raissi M, Yazdani A, Karniadakis GE (2020) Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science* 367(6481):1026–1030
- Raza SEA, Cheung L, Shaban M, et al (2019) Micro-net: A unified model for segmentation of various objects in microscopy images. *Medical image analysis* 52:160–173
- Reis HC, Turk V (2023) Transfer learning approach and nucleus segmentation with medchnet colon cancer database. *Journal of Digital Imaging* 36(1):306–325
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp 234–241
- Schindelin J, Arganda-Carreras I, Frise E, et al (2012) Fiji: an open-source platform for biological-image analysis. *Nature methods* 9(7):676–682
- Schmidt U, Weigert M, Broaddus C, et al (2018) Cell detection with star-convex polygons. In: *International conference on medical image computing and computer-assisted intervention*, Springer, pp 265–273
- Sharma Y, Syed S, Brown DE (2022) Mani: Maximizing mutual information for nuclei cross-domain unsupervised segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp 345–355
- Stringer C, Wang T, Michaelos M, et al (2021) Cellpose: a generalist algorithm for cellular segmentation. *Nature methods* 18(1):100–106
- Verma R, Kumar N, Patil A, et al (2021) Monusac2020: A multi-organ nuclei segmentation and classification challenge. *IEEE Transactions on Medical Imaging* 40(12):3413–3423
- Vu QD, Graham S, Kurc T, et al (2019) Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology* 7:53
- Wan T, Zhao L, Feng H, et al (2020) Robust nuclei segmentation in histopathology using asppu-net and boundary refinement. *Neurocomputing* 408:144–156
- Wang B, Liu H, Samaras D, et al (2020a) Distribution matching for crowd counting. *Advances in neural information processing systems* 33:1595–1607

- Wang H, Wang Z, Wang J, et al (2022a) Ica-unet: An improved u-net network for brown adipose tissue segmentation. *Journal of Innovative Optical Health Sciences* 15(03):2250018
- Wang K, Zhang X, Zhang X, et al (2022b) Eanet: Iterative edge attention network for medical image segmentation. *Pattern Recognition* 127:108636
- Wang X, Kong T, Shen C, et al (2020b) Solo: Segmenting objects by locations. In: *European conference on computer vision*, Springer, pp 649–665
- Xie E, Sun P, Song X, et al (2020) Polarmask: Single shot instance segmentation with polar representation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 12193–12202
- Xu W, Wang H, Qi F, et al (2019) Explicit shape encoding for real-time instance segmentation. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp 5168–5177
- Xu Y, Li Y, Wang Y, et al (2017) Gland instance segmentation using deep multichannel neural networks. *IEEE Transactions on Biomedical Engineering* 64(12):2901–2912
- Zhang Y, Cai L, Wang Z, et al (2025) Seine: Structure encoding and interaction network for nuclei instance segmentation. *IEEE Journal of Biomedical and Health Informatics*
- Zhao M, Jia X, Fan L, et al (2021) Robust ellipse fitting using hierarchical gaussian mixture models. *IEEE Transactions on Image Processing* 30:3828–3843
- Zhou Y, Onder OF, Dou Q, et al (2019) Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation. In: *International conference on information processing in medical imaging*, Springer, pp 682–693
- Zhu L, Lin G, Chen J, et al (2025) Large images are gaussians: High-quality large image representation with levels of 2d gaussian splatting. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp 10977–10985
- Zou Y, Zhang Z, Li CL, et al (2022) Learning instance-specific adaptation for cross-domain segmentation. In: *European Conference on Computer Vision*, Springer, pp 459–476