Instance-Aware Diffusion Model for Gland Segmentation in Colon Histology Images

Mengxue Sun, Wenhui Huang(✉), and Yuanjie Zheng

School of Information Science and Engineering, Shandong Normal University, Jinan, China
whhuang.sdu@gmail.com

Abstract. In pathological image analysis, determination of gland morphology in histology images of the colon is essential to determine the grade of colon cancer. However, manual segmentation of glands is extremely challenging and there is a need to develop automatic methods for segmenting gland instances. Recently, due to the powerful noise-to-image denoising pipeline, the diffusion model has become one of the hot spots in computer vision research and has been explored in the field of image segmentation. In this paper, we propose an instance segmentation method based on the diffusion model that can perform automatic gland instance segmentation. Firstly, we model the instance segmentation process for colon histology images as a denoising process based on the diffusion model. Secondly, to recover details lost during denoising, we use Instance Aware Filters and multi-scale Mask Branch to construct global mask instead of predicting only local masks. Thirdly, to improve the distinction between the object and the background, we apply Conditional Encoding to enhance the intermediate features with the original image encoding. To objectively validate the proposed method, we compared state-of-the-art deep learning model on the 2015 MICCAI Gland Segmentation challenge (GlaS) dataset and the Colorectal Adenocarcinoma Gland (CRAG) dataset. The experimental results show that our method improves the accuracy of segmentation and proves the efficacy of the method.

Keywords: Gland segmentation · Diffusion model · Colon histology images

1 Introduction

Colorectal cancer is a prevalent form of cancer characterized by colorectal adenocarcinoma, which develops in the colon or rectum's inner lining and exhibits glandular structures [5]. These glands play a critical role in protein and carbohydrate secretion across various organ systems. Histological examinations using Hematoxylin and Eosin staining are commonly conducted by pathologists to evaluate the differentiation of colorectal adenocarcinoma [15]. The extent of

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 662–672, 2023.
https://doi.org/10.1007/978-3-031-43987-2_64

gland formation is a crucial factor in determining tumor grade and differentiation. Accurate segmentation of glandular instances on histological images is essential for evaluating glandular morphology and assessing colorectal adenocarcinoma malignancy. However, manual annotation of glandular instances is a time-consuming and expertise-demanding process. Hence, automated methods for glandular instance segmentation hold significant value in clinical practice.

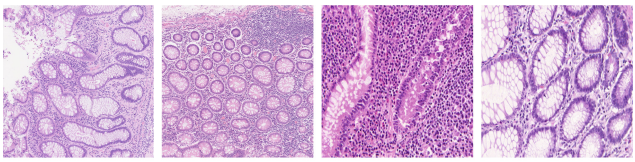

Fig. 1. (a–b) Example images from the CRAG dataset. (c–d) Example images from the GlaS dataset.

Automated segmentation has been explored using deep learning techniques [21,33], including U-Net [17], FCN [13], Siamese network [10,11] and their variations for semantic segmentation [31]. There are also methods that combine information bottleneck for detection and segmentation [23]. Additionally, two-stage instance segmentation methods like Mask R-CNN [7] and BlendMask [3] have been utilized, combining object detection and segmentation sub-networks. However, these methods may face difficulties in capturing different cell shapes and distinguishing tightly positioned gland boundaries. Limitations arise from image scaling and cropping, leading to information loss or distortion, resulting in ineffective boundary recognition and over-/under-segmentation. To overcome these limitations, we aim to perform gland instance segmentation to accurately identify the target location and prevent misclassification of background tissue.

Recently, diffusion model [9] has gained popularity as efficient generative models [16]. In the task of image synthesis, diffusion model has evolved to achieve state-of-the-art performance in terms of quality and mode coverage compared with GAN [32]. Furthermore, diffusion model has been applied to various other tasks [18]. DiffusionDet [4] treats the object detection task as a generative task on the bounding box space in images to handle projection detection. Several studies have explored the feasibility of using diffusion model in image segmentation [26]. These methods generate segmentation maps from noisy images and demonstrate better representation of segmentation details compared to previous deep learning methods.

In this paper, we propose a new method for gland instance segmentation based on the diffusion model. (1) Our method utilizes a diffusion model to perform denoising and tackle the task of gland instance segmentation in histology images. The noise boxes are generated from Gaussian noise, and the predicted ground truth (GT) boxes and segmentation masks are performed during the diffusion process. (2) To improve segmentation, we use instance-aware techniques

to recover lost details during denoising. This includes employing a filter and a multi-scale Mask Branch to create a global mask and refine finer segmentation details. (3) To enhance object-background differentiation, we utilize Conditional Encoding to augment intermediate features with the original image encoding. This method effectively integrates the abundant information from the original image, thereby enhancing the distinction between the objects and the surrounding background. Our proposed method was trained and tested on the 2015 MICCAI Gland Segmentation (GlaS) Challenge dataset [20] and Colorectal Adenocarcinoma Gland (CRAG) dataset [6] (as shown in Fig. 1), and the experiment results demonstrate the efficacy of the method.

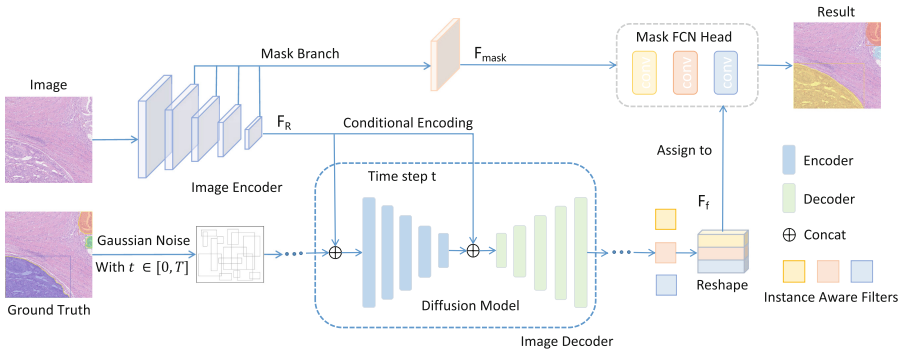

Fig. 2. The network architecture diagram of our model. The Image Encoder consists of a backbone that extracts multi-scale features from the input image. The Image Decoder based on a diffusion model incorporates the original image features as conditions to enhance the intermediate features. To preserve multi-scale information, we introduce a Mask Branch that operates on F_{mask}. By applying convolutions with weights assigned from filters to F_{mask}, we obtain instance masks.

2 Method

In this section, we present the architecture of our proposed method, which includes an Image Encoder, an Image Decoder, and a Mask Branch. The network structure is shown in Fig. 2.

2.1 Image Encoder

We propose to perform subsequent operations on the features of the original image, so we use an Image Encoder for advanced feature extraction. The Image Encoder takes the original image as input and we use a convolutional neural network such as ResNet [8] for feature extraction and a Feaure Pyramid Network

(FPN) [12] is used to generate a multi-scale feature map for ResNet backbone following.

The input image is $\mathbf{x}$ and the output is a high-level feature F_R.

$$F_R = \mathcal{F}(\mathbf{x}) \tag{1}$$

where $\mathcal{F}$ is the ResNet. The Image Encoder operates only once and uses the F_R as condition to progressively refine and generate predictions from the noisy boxes.

2.2 Image Decoder

We designed our model based on the diffusion model [30], which typically uses two Markov chains divided into two phases: a forward diffusion process and a reverse denoising process. The components of diffusion model are a learning reverse process called $p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t)$ that creates samples by converting noise into samples from $q(\mathbf{z}_0)$ and a forward diffusion process called $q(\mathbf{z}_t|\mathbf{z}_{t-1})$ that gradually corrupts data from some target distribution into a normal distribution. The forward diffusion process is defined as:

$$q(\mathbf{z}_t|\mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t, \sqrt{1-\beta_t}\mathbf{z}_{t-1}, \beta_t\mathbf{I}) \tag{2}$$

A variance schedule $\beta_t \in (0,1)$, $t \in \{1,...,T\}$ determines the amount of noise that is introduced at each stage. Alternatively, we can obtain a sample of $\mathbf{z}_t$ from direct $\mathbf{z}_0$ as follows:

$$\mathbf{z}_t = \sqrt{\bar{\alpha}_t}\mathbf{z}_0 + (1-\bar{\alpha}_t)\boldsymbol{\epsilon} \tag{3}$$

where $\bar{\alpha}_t = \prod_{s=0}^{t}(1-\beta_s)$, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$.

Our Image Decoder is based on diffusion model, which can be viewed as a noise-to-GT denoising process. In this setting, the data samples consist of a set of bounding boxes represented as $\mathbf{z}_0$, where $\mathbf{z}_0$ is a set of N boxes.

The neural network $f_\theta(\mathbf{z}_t, t)$ is trained to predict $\mathbf{z}_0$ from the $\mathbf{z}_t$ based on the corresponding image $\mathbf{x}$. In addition, to achieve complementary information by integrating the segmentation information from $\mathbf{z}_t$ into the original image encoding, we introduce Conditional Encoding, which uses the encoding features of the current step to enhance its intermediate features.

$$f_\theta(\mathbf{z}_t, F_R, t) = D((Concat(E(\mathbf{z}_t, F_R), F_R), t), t) \tag{4}$$

$$\mathbf{z}_0 = f(\cdots(f(\mathbf{z}_{T-m}, F_R, T-m)) \tag{5}$$

where D represent the decoder, E represent the encoder and $m \in \{1,...,T\}$. We use Instance Aware Filters (IAF) during iterative sampling, which allows sharing parameters between steps.

$$F_f^t = IAF(f_\theta(\mathbf{z}_t, F_R, t), t) \tag{6}$$

where F_f^t is the output feature of the filter.

2.3 Mask Branch

We have also utilized dynamic mask head [22] to predict masks in our study. In this stage, we use the Mask Branch to fuse the different scale information of the FPN and output the mask feature F_{mask}. The diffusion process decodes RoI features into local masks, and multi-scale features can be supplemented with more detailed information for predicting global masks to compensate for the detail lost in the diffusion process, and we believe that instance masks require a larger perceptual domain because of the higher demands on instance edges. Specifically, the instance mask can be generated by convolving the feature map F_{mask} from the Mask Branch and F_f^t from the IAF, which is calculated as follows:

$$\mathbf{s} = MFH(F_{mask}, F_f^t) \tag{7}$$

where the predicted instance mask is denoted by $\mathbf{s} \in R^{H \times W}$. The Mask FCN Head, denoted by MFH, is comprised of three 1×1 convolutional layers.

We enhance our loss function by incorporating two components, L_d and L_s, and utilize the γ parameter to optimize the balance between these two losses.

$$L = L_d + \gamma L_s(\mathbf{s}, \mathbf{s}^{GT}) \tag{8}$$

where the L_s in our model represents the measure of overlap between the predicted instance mask and the ground truth $\mathbf{s}^{GT}$ [14], and the L_d is the loss of DiffusionDet. The optimal value for the parameter γ is usually determined based on achieving the best overall performance on the validation set. In this work, we chose $\gamma = 5$ to balance these two losses.

3 Experiments and Results

We presented the segmentation results of our model compared to the ground truth in Fig. 3, and provided both qualitative and quantitative evaluations that validate the effectiveness of our proposed network for gland instance segmentation.

Data and Evaluation Metrics: We evaluated the effectiveness of the proposed model on two datasets: the GlaS dataset and the CRAG dataset. The GlaS dataset comprises 85 training and 80 testing images, divided into 60 images in Test A and 20 images in Test B. The CRAG dataset consists of 173 training and 40 testing images. We have adopted Vahadane method for stain normalization [1]. Furthermore, to enhance the training dataset and mitigate the risk of overfitting, we employed random combinations of image flipping, translation, Gaussian blur, brightness variation, and other augmentation techniques.

We assessed the segmentation results using three metrics from the GlaS Challenge: (1) Object F1, which measures the accuracy of detecting individual glands,

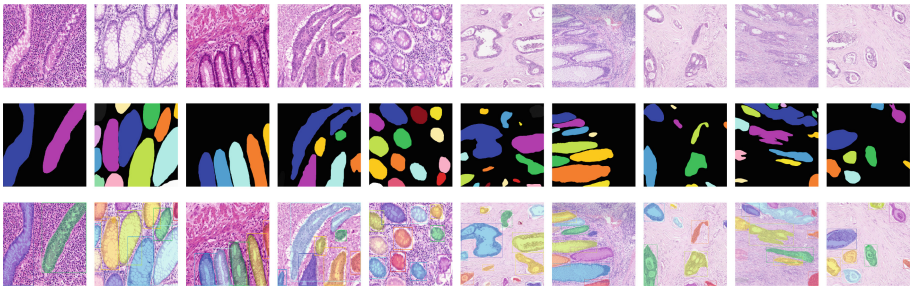

Fig. 3. The instance segmentation results on the GlaS dataset and CRAG dataset. From top to bottom: the original images, the ground truth, and the segmentation results produced by our method.

(2) Object Dice, which evaluates the volume-based accuracy of gland segmentation, and (3) Object Hausdorff, which assesses the shape similarity between the segmentation result and the ground truth. We assigned each method three ranking numbers based on these metrics and computed their sum to determine the final ranking for each method's overall performance.

Implementation Details: In our experiments, we choose the ResNet-50 with FPN as the backbone in the proposed method. The backbone is pretrained on ImageNet. Image decoder, Mask Branch and Mask FCN Head are trained end-to-end. We trained on the GlaS and CRAG datasets in a Python 3.8.3 environment on Ubuntu 18.04, using PyTorch 1.10 and CUDA 11.4. During training, we utilized an SGD optimizer with a learning rate of 2.5×10^{-5} and the weight decay as 10^{-4}. We set diffusion timesteps $T = 1000$ and chose a linear schedule from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$. Training was performed on A100 GPU with a batch size of 2.

Results on the GlaS Challenge Dataset: We conducted experiments to evaluate the performance of our proposed model by comparing it with the DSE model [27], the DMCN [28], the DCAN [2], the SPL-Net [29], the DoubleU-Net [24], the MILD-Net [6], the GCSBA-Net [25], and the MPCNN [19]. Table 1 provides an overview of the average performance of these models.

Our proposed model demonstrated a enhancement in performance, surpassing the second-best method on both Test A and Test B datasets. Specifically, on Test A, we observed an improvement of 0.006, 0.01, and 1.793 in Object F1, Object Dice, and Object Hausdorf. Similarly, on Test B, resulting in an improvement of 0.022, 0.014 and 3.694 in Object F1, Object Dice, and Object Hausdorf, respectively. Although Test B presented a more challenging task due to the presence of complex morphology in the images, our proposed model demonstrated accurate segmentation in all cases. The experimental results highlighted the effectiveness of our approach in improving the accuracy of gland instance segmentation.

Results on the CRAG Dataset: The proposed model was additionally evaluated on the CRAG dataset by comparing it against the GCSBA-Net, DoubleU-Net, DSE model, MILD-Net, and DCAN. The average performance of these models is shown in Table 2. Our experimental results demonstrate that our proposed method achieves superior performance, with improvements of 0.017, 0.012, and 4.026 for Object F1, Object Dice, and Object Hausdorff, respectively, compared to the second-best method. These results demonstrate the effectiveness of our method in segmenting different datasets.

Ablation Studies: Our network utilizes the Mask Branch and Conditional Encoding to enhance performance and segmentation quality. Ablation studies on the GlaS and CRAG datasets confirm the effectiveness of these modules (Table 3). The Mask Branch is responsible for multi-scale feature extraction and fusion with the backbone network, as well as refining the Image Decoder's output. Without the Mask Branch, direct usage of original image features lacks multi-scale information and results in less accurate segmentation. Conditional Encoding is employed to establish a connection between input image features

Table 1. The experimental results on GlaS Challenge dataset. The S represents the score, R represents the rank and Rank Sum refers to the sum of rank for each evaluation metric.

Models	Object F1				Object Dice				Object Hausdorff (pixels)				Rank Sum
	Test A		Test B		Test A		Test B		Test A		Test B		
	S	R	S	R	S	R	S	R	S	R	S	R	
Proposed	0.941	1	0.893	1	0.939	1	0.889	1	26.042	1	72.351	1	6
DoubleU-Net [24]	0.935	2	0.871	2	0.929	2	0.875	2	27.835	2	76.045	2	12
DSE model [27]	0.926	3	0.862	3	0.927	3	0.871	3	31.209	3	80.509	3	18
GCSBA-Net [25]	0.916	5	0.832	6	0.914	4	0.834	6	41.49	4	102.88	4	29
MILD-Net [6]	0.914	6	0.844	4	0.913	5	0.836	5	41.540	5	105.890	5	30
SPL-Net [29]	0.924	4	0.844	4	0.902	7	0.840	4	49.881	8	106.075	6	33
DMCN [28]	0.893	8	0.843	5	0.908	6	0.833	7	44.129	6	116.821	7	39
DCAN [2]	0.912	7	0.716	7	0.897	8	0.781	9	45.418	7	160.347	9	47
MPCNN [19]	0.891	9	0.703	8	0.882	9	0.786	8	57.413	9	145.575	8	51

Table 2. The experimental results on the CRAG dataset.

Models	Object F1		Object Dice		Object Hausdorff (pixels)		Rank Sum
	S	R	S	R	S	R	
Proposed	0.853	1	0.906	1	113.224	1	3
GCSBA-Net [25]	0.836	2	0.894	2	146.77	4	8
DoubleU-Net [24]	0.835	3	0.890	3	117.25	2	8
DSE model [27]	0.835	3	0.889	4	120.127	3	10
MILD-Net [6]	0.825	4	0.875	5	160.140	5	14
DCAN [2]	0.736	5	0.794	6	218.76	6	17

and the diffusion model. Performing reverse diffusion without any reference condition can introduce numerous errors and require multiple iterations to achieve the desired outcome. When employing Mask Branch, our approach resulted in an improvement of 0.082, 0.09, 0.07 in Object F1, and 0.07, 0.078, 0.07 in Object Dice, while Object Hausdorff decreased by 10.29, 11.11, 24.47 on GlaS Test A, GlaS Test B, and CRAG, respectively. Similarly, by utilizing Conditional Encoding, we observed an improvement of 0.048, 0.034, 0.052 in Object F1, and 0.026, 0.042, 0.057 in Object Dice, while Object Hausdorff decreased by 6.771, 8.115, 12.141 on GlaS Test A, GlaS Test B, and CRAG, respectively.

Table 3. The ablation study results on the CRAG and GlaS datasets demonstrate the impact of different modules on performance. The Mask Branch module contributes to multi-scale feature extraction, while the Conditional Encoding module establishes the connection between input image features and the diffusion model.

Dataset	Model		Object F1	Object Dice	Object Hausdorff (pixels)
	Mask Branch	Conditional Encoding			
GlaS Test A	✓	✓	0.941±0.041	0.939±0.057	26.042±1.688
	✓		0.893±0.105	0.913±0.092	32.813±2.731
		✓	0.859±0.098	0.869±0.107	36.332±1.460
GlaS Test B	✓	✓	0.893±0.046	0.889±0.072	72.351±1.271
	✓		0.859±0.061	0.847±0.072	80.466±4.604
		✓	0.803±0.077	0.811±0.082	83.461±3.704
CRAG	✓	✓	0.853±0.045	0.906±0.066	113.224±2.464
	✓		0.801±0.062	0.849±0.071	125.365±3.599
		✓	0.783±0.073	0.836±0.083	137.694±5.934

4 Conclusion and Discussion

In this paper, we propose a diffusion model based method for gland instance segmentation. By considering instance segmentation as a denoising process based on diffusion model. Our model contains three main parts: Image Encoder, Image Decoder, and Mask Branch. By utilizing a diffusion model with Conditional Encoding for denoising, we are able to improve the precision of instance localization while compensating for the missing details in the diffusion model. By incorporating multi-scale information fusion, our approach results in more accurate segmentation outcomes. Experimental results on the GlaS dataset and CRAG dataset show that our method surpasses state-of-the-art approach, demonstrating its effectiveness.

Although our method demonstrates excellent performance in gland instance segmentation, challenges arise in certain scenarios characterized by irregular shapes, flattening, and overlapping. In such cases, our network tends to classify multiple small targets with unclear boundaries as a single object, indicating limitations in segmentation accuracy when dealing with high aggregation or overlap. This limitation may stem from the difficulty in accurately distinguishing fine details between instances and the incorrect identification of boundaries.

To address these limitations, future work will focus on improving segmentation performance in challenging scenarios by specifically targeting three identified limitations: (1) Incorporate random noise during training to reduce reliance on bounding box information for denoising; (2) Explore more efficient methods for cross-step denoising in the diffusion model to improve processing time without compromising segmentation accuracy; and (3) Develop a more effective Conditional Encoding method to provide accurate instance context for noise filtering in discriminative tasks like nuclear segmentation.

Acknowledgments. This work is supported by funds from the National Natural Science Foundation of China (62003196, 62076249, 62072289, and 62073201) and the Provincial Natural Science Foundation of Shandong Province of China (ZR2020QF032).

References

1. Anand, D., Ramakrishnan, G., Sethi, A.: Fast gpu-enabled color normalization for digital pathology. In: 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 219–224. IEEE (2019)
2. Chen, H., Qi, X., Yu, L., Heng, P.A.: Dcan: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2487–2496 (2016)
3. Chen, H., Sun, K., Tian, Z., Shen, C., Huang, Y., Yan, Y.: Blendmask: top-down meets bottom-up for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8573–8581 (2020)
4. Chen, S., Sun, P., Song, Y., Luo, P.: Diffusiondet: diffusion model for object detection. arXiv preprint arXiv:2211.09788 (2022)
5. Fleming, M., Ravula, S., Tatishchev, S.F., Wang, H.L.: Colorectal carcinoma: pathologic aspects. J. Gastrointesti. Oncol. **3**(3), 153 (2012)
6. Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.A., et al.: Mild-net: minimal information loss dilated network for gland instance segmentation in colon histology images. Med. Image Anal. **52**, 199–211 (2019)
7. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. **33**, 6840–6851 (2020)
10. Huang, W., Gu, J., Duan, P., Hou, S., Zheng, Y.: Exploiting probabilistic siamese visual tracking with a conditional variational autoencoder. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 14213–14219. IEEE (2021)
11. Huang, W., Gu, J., Ma, X., Li, Y.: End-to-end multitask siamese network with residual hierarchical attention for real-time object tracking. Appl. Intell. **50**, 1908–1921 (2020)
12. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)

13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
15. Niazi, M.K.K., Parwani, A.V., Gurcan, M.N.: Digital pathology and artificial intelligence. Lancet Oncol. **20**(5), e253–e261 (2019)
16. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Sanchez, P., Kascenas, A., Liu, X., O'Neil, A.Q., Tsaftaris, S.A.: What is healthy? generative counterfactual diffusion for lesion localization. In: Deep Generative Models: Second MICCAI Workshop, DGM4MICCAI 2022, Held in Conjunction with MICCAI 2022, Singapore, 22 September 2022, Proceedings, pp. 34–44. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-18576-2_4
19. Sirinukunwattana, K., et al.: Gland segmentation in colon histology images: the glas challenge contest. Med. Image Anal. **35**, 489–502 (2017)
20. Sirinukunwattana, K., Snead, D.R., Rajpoot, N.M.: A stochastic polygons model for glandular structures in colon histology images. IEEE Trans. Med. Imaging **34**(11), 2366–2378 (2015)
21. Song, J., Zheng, Y., Xu, C., Zou, Z., Ding, G., Huang, W.: Improving the classification ability of network utilizing fusion technique in contrast-enhanced spectral mammography. Med. Phys. **49**(2), 966–977 (2022)
22. Tian, Z., Shen, C., Chen, H.: Conditional convolutions for instance segmentation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 282–298. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58452-8_17
23. Wang, J., et al.: Information bottleneck-based interpretable multitask network for breast cancer classification and segmentation. Med. Image Anal. **83**, 102687 (2023)
24. Wang, P., Chung, A.C.S.: DoubleU-net: colorectal cancer diagnosis and gland instance segmentation with text-guided feature control. In: Bartoli, A., Fusiello, A. (eds.) ECCV 2020. LNCS, vol. 12535, pp. 338–354. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66415-2_22
25. Wen, Z., Feng, R., Liu, J., Li, Y., Ying, S.: Gcsba-net: gabor-based and cascade squeeze bi-attention network for gland segmentation. IEEE J. Biomed. Health Inf. **25**(4), 1185–1196 (2020)
26. Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., Cattin, P.C.: Diffusion models for implicit image segmentation ensembles. In: International Conference on Medical Imaging with Deep Learning, pp. 1336–1348. PMLR (2022)
27. Xie, Y., Lu, H., Zhang, J., Shen, C., Xia, Y.: Deep segmentation-emendation model for gland instance segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 469–477. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_52
28. Xu, Y., Li, Y., Wang, Y., Liu, M., Fan, Y., et al.: Gland instance segmentation using deep multichannel neural networks. IEEE Trans. Biomed. Eng. **64**(12), 2901–2912 (2017)

29. Yan, Z., Yang, X., Cheng, K.-T.T.: A deep model with shape-preserving loss for gland instance segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 138–146. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_16
30. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., et al.: Diffusion models: a comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796 (2022)
31. Zhang, Z., Tian, C., Bai, H.X., Jiao, Z., Tian, X.: Discriminative error prediction network for semi-supervised colon gland segmentation. Med. Image Anal. **79**, 102458 (2022)
32. Zheng, Y., et al.: Symreg-gan: symmetric image registration with generative adversarial networks. IEEE Trans. Pattern Anal. Mach. Intell. **44**(9), 5631–5646 (2021)
33. Zheng, Y., et al.: Image matting with deep gaussian process. IEEE Trans. Neural Netw. Learn. Syst. (2022)