Journal of Computers Vol. 35 No. 6, December 2024, pp. 45-54
doi: 10.53106/199115992024123506004

Research on Image Segmentation Method Based on Multi-Scale Feature Fusion and Dual Attention

Zhihong Wang[1], Chaoying Wang[2*], Jianxin Li[2], Tianxiang Wu[2], Jiajun Li[2], Hongxing Huang[2], and Lai Jiang[2]

[1] School of Artificial Intelligence, DongGuan Polytechnic, Dongguan, Guangdong, 523808 China
359783250@qq.com

[2] School of Electronic Information, Dongguan Polytechnic, Dongguan, 523808, China
591460133@qq.com, 279149042@qq.com, 1585596231@qq.com, 3345340890@qq.com, 2960632949@qq.com, 1225643914@qq.com

Received 24 March 2024; Revised 14 May 2024; Accepted 3 July 2024

Abstract. This paper proposes a multi-scale attention fusion mechanism for automatic gland segmentation based on the colorectal adenocarcinoma cell dataset, aiming to address the issue of unclear cell adhesion and confusion with the background in colorectal adenocarcinoma cell instance segmentation. Deep learning methods are employed for top-down gland cell instance segmentation. The neural network features a novel spatial-pyramid dual-path attention module that not only integrates multi-dimensional feature map spatial information but also enriches feature space through cross-dimensional feature information interaction. With the assistance of the new fusion module, it can perceive higher resolution features effectively fuse multi-scale features, leading to higher segmentation accuracy, stronger robustness, and generalization. It demonstrates excellent performance on the GlaS and CRAG datasets.

Keywords: colorectal gland cells, deep learning, instance segmentation, attention mechanism

1 Introduction

With the modern pace of life accelerating, unhealthy diets, lifestyles, and environmental pollution have had a certain impact on human health, especially leading to the development of colorectal adenocarcinoma, which is the third most common type of cancer globally. Early observation of colorectal adenoma tissue slice images and statistical analysis based on the morphology of individual cells can help detect the degree of colorectal adenoma lesions at an early stage, playing a crucial role in formulating treatment plans for colorectal adenocarcinoma. Adenocarcinoma cell instance segmentation is a necessary step for pathologists to quantitatively analyze the malignancy of adenocarcinoma. Currently used segmentation methods include manual visual inspection and computerized pathological image analysis. Generally, the accuracy of manual visual inspection is higher than that of computerized methods, but manual visual inspection requires a large amount of labor. Due to differences in experience among different doctors, and the same doctor's state may also vary at different stages.

Accurately segmenting glands in histopathological images is a prerequisite for conducting pathology image-assisted diagnosis and holds significant importance. There are many segmentation methods currently available, including those based on traditional machine learning and deep neural networks. Traditional image processing methods work well for segmenting benign glands but perform poorly for malignant glands with irregular shapes. Additionally, traditional methods rely too heavily on manual features and prior knowledge. The irregularity of malignant glands increases the difficulty of gland instance segmentation, making it challenging for traditional natural image processing methods to meet the requirements of current gland instance segmentation tasks. Deep learning is the latest trend in machine learning and artificial intelligence research. Researchers are combining deep learning models with the characteristics of colorectal adenocarcinoma cells to study and gradually improve segmentation algorithms, which can effectively capture boundary information in images, thereby enhancing segmentation accuracy.

The success of deep learning segmentation methods is attributed to the powerful feature learning and ex-

* Corresponding Author

traction capabilities of deep neural networks, which can perform pixel-level image classification and solve semantic-level image segmentation problems. Deep learning methods can automatically extract features from images and perform classification and recognition in an unsupervised manner, making them widely applicable in image processing. As a result, research on automatic gland segmentation using deep learning methods is gradually gaining attention. In 2016, Chen et al. proposed an efficient deep contour-aware network (DCAN) that utilizes multi-level contextual feature representation in an end-to-end manner to achieve precise gland segmentation, capable of simultaneously outputting probabilities of glands and gland boundaries [1]. In 2017, Xu et al. introduced a multi-channel convolutional neural network and designed a sub-network that efficiently integrates outputs from multiple channels [2]. This method not only incorporates target detection and edge detection for each gland with spatial constraints but also enhances instance segmentation performance by fusing contour information with positional information through two-channel integration.

Yan et al. proposed an innovative loss function composed of a shape similarity loss and weighted pixel cross-entropy to obtain well-shaped glands and effectively separate adherent glands [3]. The overall network structure is similar to DCAN, as it is a dual-task network capable of simultaneously outputting probability maps for glands and boundaries. This loss function reduces model overfitting and improves training and execution speed, making the model more stable and faster to converge. Graham et al. introduced the Minimal Information Loss Dilated Network, which utilizes dilated convolutions and spatial pyramid pooling to incorporate high-resolution information in the middle of the network [4]. At the end of the network, random transformation sampling is used to reduce model uncertainty, thereby enhancing the accuracy of segmentation.

In 2017, Yang et al. proposed an efficient training data method that combines active learning framework with Fully Convolutional Network (FCN) [5]. This method reduces annotation workload and model uncertainty by searching for the most effective annotated areas and integrating the similarity and uncertainty information provided by FCN. In comparison to previous methods, the active learning-based approach only requires half of the training data to achieve highly competitive results. In 2020, two efficient gland segmentation methods emerged. Ding et al. introduced a multi-scale fully convolutional network that integrates dilated convolutions and residual structures to extract features with different receptive fields corresponding to objects of different sizes [6]. Additionally, Wen et al. presented a Gabor-based Cascade Squeeze Bi-Attention Network [7]. The Gabor-based encoder module in this network not only captures rich image texture features but also enhances the interpretability of convolutional kernels.

2 Network Design

2.1 Algorithm Description

Traditional gland cell instance segmentation methods typically start with pixel-level classification, followed by a combination of object detection and semantic segmentation for instance segmentation. This approach makes it difficult to perceive objects simultaneously during the segmentation process, leading to potential errors in instance segmentation. To address the issues of mis-segmentation and lack of segmentation precision commonly encountered in existing gland cell instance segmentation methods, this paper adopts an end-to-end approach for two-stage gland cell instance segmentation [8]. Building upon the Mask R-CNN model, this method proposes a multi-scale attention fusion gland cell instance segmentation network, as illustrated in Fig. 1. It enables end-to-end learning to simultaneously output the positions and categories of candidate boxes as well as the masks for target instance segmentation.

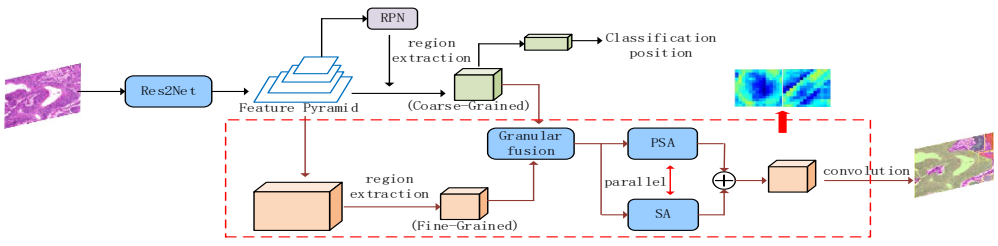

Fig. 1. The network structure diagram of this method

2.2 Dual Attention

The dual-path refinement attention module proposed in this paper models semantic correlations in both spatial and channel dimensions by attaching two modules: the spatial attention (GSA) branch and the pyramid compression attention (PSA) branch.

The PSA mechanism can learn multi-scale feature representations from input images and recalibrate channel attention weights at different scales, thereby fully extracting multi-dimensional spatial information of feature maps and realizing cross-dimensional feature interaction. The PSA mechanism proposes a multi-scale feature map extraction strategy based on the Squeeze-and-Excitation Networks (SENet) and the overall structural diagram is shown in Fig. 2. It combines feature information at corresponding scales to form a feature pyramid, aiming to achieve multi-scale feature fusion, enrich feature space, and generate better pixel-level attention. The most important module for multi-scale feature extraction in the PSA module is the SPC, which is mainly implemented in the following steps. First, the input feature map is grouped, and the Squeeze and Concat (SPC) module is used to partition the channels of the feature map to obtain multi-scale feature maps in terms of channels. Next, the SEWeight module extracts the target features of the multi-scale feature maps to obtain spatial-level visual attention vectors. Then Softmax operation is applied to recalibrate the weights of the attention vectors for the corresponding channels. Finally, the calibrated attention vectors are element-wise multiplied with the feature map to generate the output.

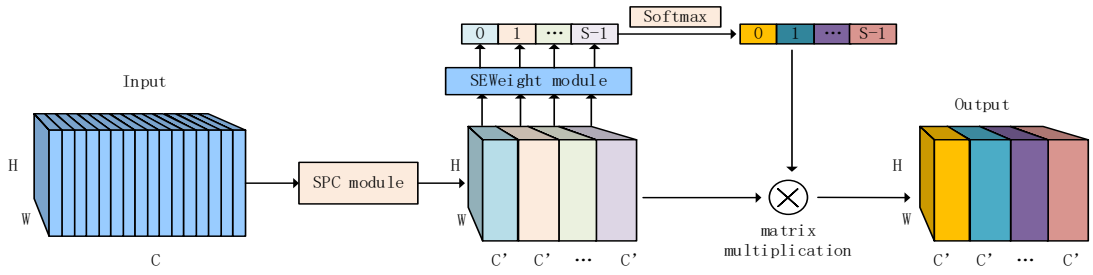

Fig. 2. The overall structure of pyramid compressed attention mechanism

(1) SPC module

The SPP module can perform multi-scale pooling operations on the input feature map while keeping the network input size unchanged [9]. Its basic structure is shown in Fig. 3. Each branch feature map of F_i has an input channel number of C and an output channel number of C' =C/S, where i=0,1,…,S-1 and S=4. Each branch at different scales independently learns the spatial information of the input feature map without interference and interacts information when concatenated for integration. To avoid a significant increase in computational complexity of convolutional kernels, the group convolution method is introduced.

$$G = 2^{\frac{K-1}{2}} \tag{1}$$

$$F_i = Conv(k_i \times k_i \times k_i, G_i)(X) \tag{2}$$

Among them, K represents the size of convolution kernels at different scales, and G represents the size of the group. $X = \mathbb{R}^{C\times D\times H\times W}$ represents input feature maps, and $F_i = \mathbb{R}^{C\times D\times H\times W}$ represents branch feature maps of different scales.

$$k_i = 2\times(i+1)+1 \tag{3}$$

$$G_i = 2^{\frac{k_i-1}{2}} \tag{4}$$

In the context, k_i represents the size of the *i-th* convolutional kernel, and G_i represents the size of the *i-th* group. At that time, $k_0 = 3$, and G_0 has the default value of 1. The number of channels in the input feature map for the classification experiments in this paper is C=160, and the number of channels for each branch of the SPC module, denoted as C', is 40. Since C' must be divisible by G_i, $k_3 = k_2 = 7$, $G_3 = G_2 = 8$ is set.

$$F = Cat([F_0, F_1, ..., F_{S-1}]) \tag{5}$$

Among them, $F \in \mathbb{R}^{C \times D \times H \times W}$ represents the complete multi-scale feature map obtained by integrating each branch feature map F_i through cascading.

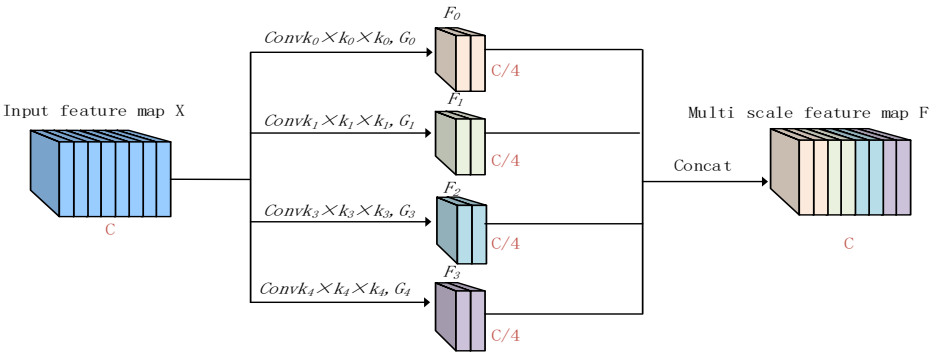

Fig. 3. Basic structure of SPC module

(2) SEWeight module

The basic structure of the SEWeight module [10] is shown in Fig. 4, which is derived from the excitation part of SENet. It can allocate weights to channels after channel interaction, ultimately obtaining feature maps with channel attention, effectively improving model accuracy. The SEWeight module mainly consists of Global Average Pooling (GAP) and Excitation. Through GAP, the SEWeight module uses the average value of each feature map as the overall representation of that feature to obtain global information for each channel. The excitation module learns specific sampling activation for each channel through two fully connected layers, controlling the excitation of each channel. The main function of this module is to obtain feature channel weights. Firstly, each channel feature map is computed into a scalar as the weight of the feature map through global pooling. Then, it is multiplied and added channel-wise to the previous feature to recalibrate the original features of the channels, integrating the feature information from each channel. This process results in obtaining more refined feature maps with richer multi-scale feature information. After multi-scale processing, each branch feature map F_i extracts the channel attention of the corresponding scale feature map to obtain its weight vector $Z_i \in \mathbb{R}^{C' \times 1 \times 1 \times 1}$.

$$Z_i = SE\ Weight(F_i) \tag{6}$$

$$g_C = \frac{1}{D \times H \times W} \sum_{i=1}^{D} \sum_{j=1}^{H} \sum_{n=1}^{W} x_C(i, j, n) \tag{7}$$

$$w_C = \sigma(W_1 \delta(W_0(g_C))) \tag{8}$$

g_C represents the GAP operator, cw represents the attention weight of the c-th channel in the SEWeight module. $W_0 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and are fully connected layers, δ denotes the ReLU operation, and σ represents

the activation function. Fully connected layers efficiently fuse feature information between channels and fully interact high-dimensional and low-dimensional information across channels. The activation function σ allocates weights to channels after information interaction, enabling the model to acquire rich feature information. The multi-scale channel attention vector *att* integrates branch feature map vectors across dimensions in a concatenated manner, facilitating information interaction of attention weights.

$$att_i = Soft\max(Z_i) = \frac{\exp(Z_i)}{\Sigma_{i=0}^{S-1}\exp(Z_i)} \tag{9}$$

$$att = att_0 \oplus att_1 \oplus \cdots \oplus att_{S-1} \tag{10}$$

Where $\oplus$ represents the concatenation operator. To ensure that the sum of channel attention weights of branch feature maps is 1, Softmax is applied to recalibrate the attention weight vector Z_i. Finally, multiplying the multi-scale channel attention weights att with the multi-scale feature map F forms the pyramid compressed attention feature map Y.

$$Y = F \odot att \tag{11}$$

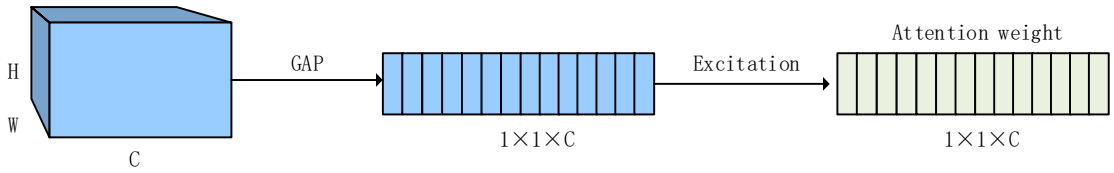

Fig. 4. Basic structure of SEWeight module

2.3 Loss Function Design

This article uses two Dice loss functions [11] to optimize segmentation models for different enhancement domains, and ultimately guides the optimization segmentation of the entire network model by summarizing the two losses. The Dice loss function is defined as follows:

$$L_{dice} = 1 - \frac{2|A \cap B|}{|A| + |B|} \tag{12}$$

Among them, |A| and |B| represent the truth map and prediction map of the segmented labels. The final loss function can be defined as:

$$Loss = L_{dice}(BN(X^S), Y^S) + L_{dice}BN(X^d), Y^d \tag{13}$$

In formula 13, $BN(X^S)$ represents a batch normalization layer for processing data with similar sources, which can enable the network to converge faster during training; Y^S represents a batch normalization layer for processing data with similar sources, which can enable the network to converge faster during training; represents the prediction result of the source similar domain, and similarly, $BN(X^d)$ and Y^d represent the batch normalization layer for processing source dissimilar data and the prediction result of the source dissimilar domain, respectively.

3 Experimental Content

3.1 Experimental Environment

The experimental training environment in this study was developed on a desktop machine running Ubuntu 18.04, with CUDA version 10.2, Python version 3.9, using the MMDetection 2.17 framework, and utilizing an NVIDIA GeForce GTX 2080 Ti GPU. The Mask R-CNN [12] was used as the baseline in the experiment, with all hyper-parameters, except for specific modules, being the same as those implemented in the Mask R-CNN algorithm in the MMDetection framework [13].

3.2 Experimental Evaluation Criteria

The evaluation metrics used in this study include mAP, as well as AP50, AP75, APS, APM, and APL. Here, L/S/M respectively represent Large/Small/Middle, indicating the evaluation performance on large objects, small objects, and medium-sized objects. We conducted detailed comparisons of various segmentation methods using different experimental parameters to evaluate their segmentation performance.

In instance segmentation tasks, Intersection over Union (IoU) is used to define the loss function for bounding box regression. The idea is to calculate the ratio between the intersection area and the union area of the predicted box and the ground truth box, indicating the correlation between the two sets of data, i.e.:

$$IoU = \frac{M_P \bigcap M_{GT}}{M_P \bigcup M_{GT}} \tag{14}$$

IoU is used to evaluate the difference between predicted values and ground truth values. The magnitude of the value indicates the level of similarity and accuracy of the prediction. An IOU value of 1 signifies complete agreement between predicted and actual pixels. At a certain threshold, the model classifies results with an IOU value greater than the threshold as positive samples and those with an IOU value less than the threshold as negative samples. In the evaluation method used in the experiment, when averaging, thresholds ranging from 0.95 to 0.5 are utilized with a step size of 0.05. Calculations are performed within each interval to evaluate the predicted results.

TP, FP, and FN stand for True Positive, False Positive, and False Negative respectively. They are calculated based on the predictions compared to the ground truth annotations. TP represents the number of positive samples that are classified correctly by the algorithm, FP is the number of negative samples that are classified incorrectly by the algorithm, and FN is the number of positive samples that are classified incorrectly by the algorithm. These variables can be used to further calculate Precision and Recall, with the following formulas:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

3.3 Experimental Results and Analysis

This method has been implemented on the GlaS and CRAG datasets to evaluate various commonly used image segmentation techniques in recent years, comparing them with our network using multiple evaluation metrics. The compared methods include baseline methods such as SOLO [14], YOLACT [15], Mask R-CNN, Mask Scoring R-CNN, Refinemask [16], and Point Rend. The experimental results are compared and evaluated as shown in Table 1. Compared to the baseline method Mask R-CNN, our method demonstrates superior performance, achieving higher quality and more accurate masks.

Table 1. Comparison of performance using different instance segmentation methods on the GlaS dataset

Moddel	Segm mAP	Bbox mAP	AP_S	AP_m	AP_L	AP_{50}	AP_{75}
SOLO	42.4	0	3.1	60.1	59.8	70.4	66.9
YOLACT	57.5	53.4	1.8	34.2	50.8	80.5	45.3
MS R-CNN	64.1	62.1	6.6	63.1	66.9	84.4	73.8
Refinemask	65.2	61.9	9.2	65.5	69.2	83.8	76.0
PointRend	66.6	63.4	7.2	64.7	71.5	85.4	78.1
Mask R-CNN	64.2	61.3	**10.1**	64.8	65.8	82.8	73.5
Ours	**67.3**	**65.3**	6.1	**65.5**	**72.1**	**85.6**	**78.6**

By applying different segmentation methods on the GlaS dataset for image segmentation, the performance differences of each segmentation method can be deduced through comparisons using various evaluation metrics. The comparative data in the Table 1 shows that the accuracy of the proposed method on the Segmentation mAP metric reaches 67.3, which is 3.0 points higher than that of Mask R-CNN; furthermore, the accuracy on the Bounding Box mAP metric reaches 65.3, which is 4.0 points higher than the Mask R-CNN method.

Table 2. Comparison of performance using different instance segmentation methods on the CRAG dataset

Model	Segm mAP	Bbox mAP	AP_S	AP_m	AP_l	AP_{50}	AP_{75}
SOLO	27.1	0	0	11.2	31.8	55.6	24.3
YOLACT	45.9	47.1	0	33.4	50.1	73.5	49.2
Refinemask	60.7	61.5	30.5	**49.2**	64.4	85.4	68.5
PointRend	57.9	58.8	**33.2**	47.2	61.3	84.1	65.6
Mask R-CNN	58.3	60.7	27.2	46.7	61.9	84.6	66.6
Ours	**61.5**	**64.2**	31.4	48.9	65.6	**86.8**	69.7

By applying different segmentation methods to conduct image segmentation on the CRAG dataset, the performance gaps of each segmentation method can be inferred through comparisons using various evaluation metrics. The comparative data in Table 2 shows that the accuracy of the proposed method on the Segmentation mAP metric reaches 61.5, which is 3.2 points higher than that of Mask R-CNN; furthermore, the accuracy on the Bounding Box mAP metric reaches 64.2, which is 3.5 points higher than the Mask R-CNN method.

Fig. 5 displays the segmentation testing on the GlaS dataset using the method employed in this paper and the baseline image segmentation method Mask R-CNN. A comparison of the visual results indicates a significant improvement in the segmentation method's performance. Glandular cells with significant morphological variations can be accurately identified, attributed to the reduction of background interference by the attention mechanism. Firstly, compared to the Mask R-CNN segmentation method, our approach introduces a fine-grained fusion module that integrates fine-grained discriminative features and coarse-grained diverse features, as shown in rows 1 and 2 of the figure. Secondly, our method utilizes a dual-head attention module that combines spatial attention and pyramid attention, integrating spatial features from different scales to obtain more accurate and reliable spatial information, reducing background interference, as illustrated in rows 3 and 4 of the figure.

Through the visual results of the experiments, it is visually evident that the improved image segmentation method proposed in this paper can achieve the best Segm APL metric, demonstrating excellent performance. This is because the glandular cell dataset exhibits significant variability in cell characteristics, with a majority being large-sized cells. Therefore, our method maps the highest resolution size to high-resolution feature maps, allowing for the refinement of feature maps for large-sized cells with high-resolution information.

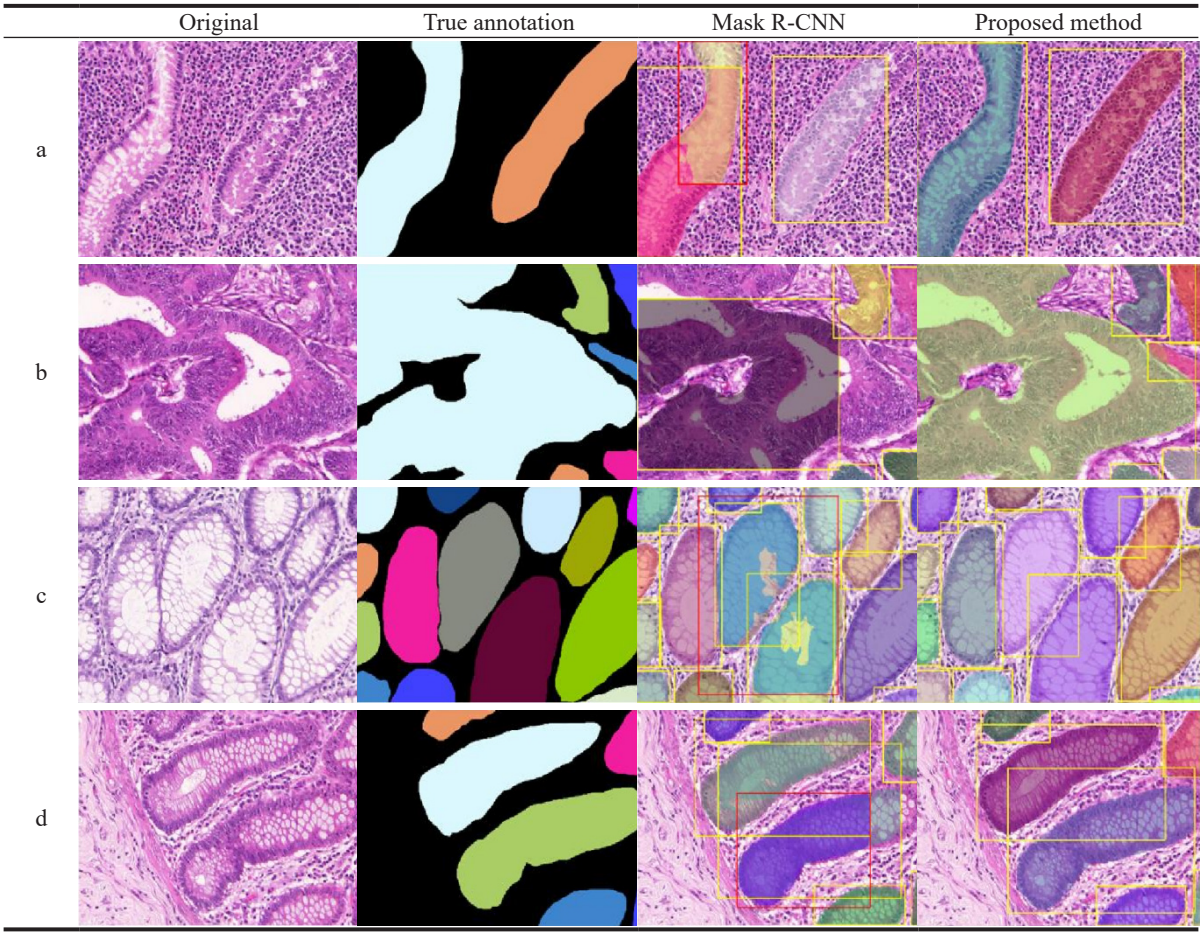

Fig. 5. Visual comparison of results between our method and Mask R-CNN on GlaS dataset

3.4 Ablation Experiment

When conducting segmentation experiments with different methods to compare their segmentation performance, we also conducted both overall experiments and ablation experiments on the GlaS dataset to validate the performance of each module of the segmentation model proposed in this paper. The results of the ablation experiments are presented in Table 3 for comparison. The data in Table 3 demonstrates that on the GlaS dataset, the integration of the fine-grained fusion module improved segmentation performance by 1.4. Subsequently, the incorporation of the dual-head attention mechanism further enhanced segmentation performance by 1.1, and finally, the integration of the Res2Net module resulted in an additional performance improvement of 1.7 on top of the existing baseline.

Table 3. Ablation experiments conducted in our network

Basic network	Granular fusion	Dual attention	Res2Net	Bbox mAP	Segm mAP
√				61.2	63.2
√	√			61.8	64.6
√	√	√		62.4	65.7
√	√	√	√	65.4	67.4

4 Conclusion

Given the scale complexity of colorectal gland images, the diverse morphological variations of cancer cells, and the existing issues in segmentation methods for colorectal gland cells, which require strong manual intervention and exhibit slightly lower accuracy, this paper proposes a more novel and effective spatial-pyramid dual-path attention module. Building upon this, a new gland cell instance segmentation network is constructed. This method enhances the Mask R-CNN framework by incorporating new functional modules and adopts a top-down approach for gland cell instance segmentation. The addition of these new modules allows the segmentation network to perceive higher-resolution mask-refined features. Through corresponding comparative experiments, it can be observed that the F1 score of this method has reached the best value of 0.9241, demonstrating superior performance compared to commonly used methods for glandular cell instance segmentation. The results of the ablation experiments conducted on the publicly available GlaS dataset demonstrate the effectiveness of the added modules in glandular cell segmentation. Furthermore, the comparison with representative methods in the field further validates the competitiveness of the proposed approach in this paper.

5 Acknowledgement

This paper is supported by The 2023 Dongguan Science and Technology Commissioner Project: research-based design for microphone product design in the context of “Her economy”, project (NO. 20231800500542), the Dongguan Science and Technology Ombudsman Project in 2022 (NO. 20221800500732), Special Fund for Guangdong Province’s Science and Technology Innovation Strategy (pdjh2024b663), Projects in the Science and Technology Situation of Dongguan City in 2023 (NO. 2023PZ08), 2022 Dongguan Vocational and Technical College Intelligent Terminal and Intelligent Manufacturing Special Project (NO. ZXD202201), 2023 Dongguan Vocational and Technical College Intelligent Terminal and Intelligent Manufacturing Special Project (NO. ZXD202315); Special fund for electronic information engineering technology specialty group of national double high program of Dongguang Polytechnic (NO. ZXD202303), Dongguan Science and Technology Ombudsman Project in 2023 (NO. 20231800500282), Dongguan Science and Technology of Social Development Program (NO. 20231800903592 & 20221800905842 & 20211800904472), Special Fund for Guangdong Province’s Science and Technology Innovation Strategy (NO. pdjh2023b1020), Special fund for electronic information engineering technology specialty group of national double high program of Dongguang Polytechnic (NO. ZXB202203, NO. ZXD202204, NO. ZXC202201), Phase III Supply and Demand Docking Employment Education Project of the Ministry of Education, Project(NO. 2023122759602), Special project of Guangdong Teaching Management Society of Higher Universities- -Research and Practice of online and offline Mixed Teaching Resources Development based on SPOC/MOOC —— takes Python Data Analysis as an example, 2024 Quality Engineering Project of Dongguan Vocational and Technical College: Curriculum Ideological and Political Demonstration Course, Data Processing and Analysis, the Special for key fields of colleges and universities in Guangdong Province under Grant (NO. 2024ZDZX1091 & NO. 2024ZDZX1092 & NO. 2024ZDZX1093), Dongguan Vocational and Technical College 2024 School-level Quality Engineering Project, Project Name: Data Processing and Analysis (NO. KCSZ202404).

References

[1] H. Chen, X. Qi, L. Yu, P.A. Heng, DCAN: deep contour-aware networks for accurate gland segmentation, in: Proc. IEEE Transactions on Computer Vision and Pattern Recognition, 2016.

[2] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, E.I.C. Chang, Gland instance segmentation using deep multichannel neural networks, IEEE Transactions on Biomedical Engineering 64(12)(2017) 2901-2912.

[3] Z. Yan, X. Yang, K.T.T. Cheng, A deep model with shape-preserving loss for gland instance segmentation, in: Proc. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, 2018.

[4] S. Graham, H. Chen, J. Gamper, Q. Dou, P.A. Heng, D. Snead, Y.W. Tsang, N. Rajpoot, MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images, Medical Image Analysis 52(2019) 199-211.

[5] L. Yang, Y. Zhang, J. Chen, S. Zhang, D.Z. Chen, Suggestive annotation: A deep active learning framework for biomedical image segmentation, in: Proc. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017,

2017.
[6] H. Ding, Z. Pan, Q. Cen, Y. Li, S. Chen, Multi-scale fully convolutional network for gland segmentation using three-class classification, Neurocomputing 380(2020) 150-161.
[7] Z. Wen, R. Feng, J. Liu, Y. Li, S. Ying, GCSBA-Net: Gabor-based and Cascade Squeeze Bi-Attention Network for Gland Segmentation, IEEE Journal of Biomedical and Health Informatics 25(4)(2021) 1185-1196.
[8] B. Li, S. Liu, F. Wu, G. Li, M. Zhong, X. Guan, RT-Unet: An advanced network based on residual network and transformer for medical image segmentation, International Journal of Intelligent Systems 37(11)(2002) 8565-8582.
[9] Y. Ma, Y. Zhang, L. Chen, Q. Jiang, B. Wei, Dual attention fusion UNet for COVID-19 lesion segmentation from CT images, Journal of X-Ray Science and Technology 31(4)(2023) 713-729.
[10] L. Zhang, K. Zhang, H.W. Pan, SUNet++: A Deep Network with Channel Attention for Small-Scale Object Segmentation on 3D Medical Images, Tsinghua Science and Technology 28(4)(2023) 628-638.
[11] Y.L. Yang, S. Dasmahapatra, S. Mahmoodi, ADS_UNet: A nested UNet for histopathology image segmentation, Expert Systems with Applications 226(2023) 120128.
[12] X. Wang, Z. Li, Y. Huang, Y. Jiao, Multimodal medical image segmentation using multi-scale context-aware network, Neurocomputing 486(2022) 135-146.
[13] L. Xiao, Z. Pan, X. Du, W. Chen, W. Qu, Y. Bai, T. Xu, Weighted skip-connection feature fusion: A method for augmenting UAV oriented rice panicle image segmentation, Computers and Electronics in Agriculture 207(2023) 107754.
[14] Y.S. Zhou, L.L. Chen, T. Wang, S.Z. Xu, Brain tumor MR image segmentation method based on double attention mechanism and iterative aggregation U-Net, Journal of South-central Minzu University (Natural Science Edition) 42(3) (2023) 373-381.
[15] J. Han, Y. Wang, H. Gong, Fundus Retinal Vessels Image Segmentation Method Based on Improved U-Net, IRBM 43(6)(2022) 628-639.
[16] T. Huang, Y. Liu, Research on the magnetic resonance imaging brain tumor segmentation algorithm based on DO-UNet, International Journal of Imaging Systems and Technology 33(1)(2023) 143-157.