A Multi-stage Noise Suppression Network for Segmenting Polyp Images Containing Noise Interference

Mianduan Lin, Kaoru Hirota, Yaping Dai and Shuai Shao*

Beijing Institute of Technology, Beijing 100081, P. R. China
3220210896@bit.edu.cn; hirota@bit.edu.cn; daiyaping@bit.edu.cn; shaoshuai@bit.edu.cn

Abstract. Unlike other medical images, polyp images usually contain a lot of noise interference, which reduces the accuracy of polyp segmentation. To solve the problem of polyp images containing a large amount of noise interference, a Multi-stage Noise Suppression Network (MNSNet) that integrates Transformer and CNN is proposed. Firstly, for the problem that low-level polyp features contain a lot of background noise interference, the Polyp Background Noise Suppression (PBNS) module is constructed based on the self-attention to improve the anti-background noise ability of MNSNet in the feature extraction stage, which in turn improves the network's performance in polyp segmentation. Secondly, to address the lack of anti-interference ability of the semantic fusion method in the existing polyp segmentation network, the Polyp Dynamic Noise Suppression (PDNS) module is constructed based on the dynamic kernel method to improve the adaptability of MNSNet to complex and variable noise interference in the polyp images during the semantic fusion stage, thereby improving the network's polyp segmentation accuracy. Experiment results show that the MNSNet has best performance compare with five methods (SANet, SSFormer, PPFormer, TransFuse and Meta-Polyp), under five benchmark polyp segmentation datasets (the Kvasir dataset, the CVC-ClinicDB dataset, the CVC-ColonDB dataset, the CVC-T dataset and the ETIS dataset). In particular, compared with the Meta-Polyp, MNSNet improves mDice and mIoU by 2.2% and 2.0% on the ETIS dataset.

Keywords: Deep Learning, Attention Mechanism, Dynamic Kernel, Noise Suppression, Polyp Image Segmentation

1 Introduction

Early colonoscopy helps physicians respond to colorectal polyps before they develop into colorectal cancer (CRC), which reduces the incidence of CRC [1]. Polyp detection based on colonoscopy is highly dependent on the physician's experience and has a rate of missed detections [2]. Based on deep learning theory and computer technologies, it is of great academic significance and application value to study automatic and accurate polyp segmentation methods to assist doctors in detecting polyps.

The automatic segmentation of polyps has gradually evolved from traditional methods based on manually designed features to modern methods based on deep learning. SANet [3] uses shallow attention blocks to fuse different levels of polyp features, and uses a probability correction strategy to alleviate the problem of polyp pixel imbalance, thereby improving the effectiveness of the model's polyp segmentation. SSFormer [4] uses a progressive local decoder to solve the attention dispersion problem of the Transformer encoder network, thereby improving the adaptability of the Transformer structure to polyp segmentation tasks. PPFormer [5] uses predictive graph guidance algorithms to focus on difficult to segment areas in the image, which enhances the model's perception of polyp boundaries. TransFuse [6] uses two branches, CNN and Transformer, to extract low-level detail information and improve the efficiency of global context modeling, which enhances the robustness of the algorithm. Meta-Poly [7] introduced multi-scale up-sampling blocks in the decoder to improve the algorithm's sensitivity to detailed textures, thereby further improving the accuracy of polyp segmentation. The above representative studies have overlooked a problem: unlike other medical images, polyp images often contain a large amount of noise interference, such as intestinal impurities, specular reflection, intestinal peristalsis, etc. The noise interference in polyp images will affect the perception of polyp areas by polyp segmentation methods and reduce the accuracy of polyp segmentation methods. However, existing research rarely considers the issue of noise interference in polyp images during both feature extraction and semantic fusion stages.

In order to solve the problem of polyp images containing a large amount of noise interference, a Multi-stage Noise Suppression Network (MNSNet) integrating Transformer and CNN is proposed to improve the accuracy of polyp segmentation. MNSNet uses parallel CNN branches and Transformer branches to extract polyp features separately, improving the network's ability to obtain global and local information, thereby enhancing the robustness of the network. In the stage of polyp feature extraction, a Polyp Background Noise Suppression (PBNS) module is designed based on self-attention mechanism. The PBNS module is used to increase the attention of MNSNet to the polyp target area and improve the network's anti-interference ability against background noise, thereby improving the polyp segmentation effect of the network. In the semantic fusion stage of polyps, a Polyp Dynamic Noise Suppression (PDNS) module is designed based on the dynamic kernel method. The PDNS module dynamically generates a variable convolution kernel based on input data, which enables MNSNet to flexibly handle complex and variable polyp noise interference and improve its adaptability to different noise interferences, thereby improving the polyp segmentation accuracy of MNSNet. To evaluate the performance of the MNSNet for polyp segmentation, comparison experiments were conducted with five methods on five benchmark polyp segmentation datasets. The experiment results show the effectiveness and superiority of MNSNet.

The paper is organized as follows. In Section 2, the details of proposed network are described. In Section 3, the experiment results are presented. Lastly, a brief conclusion is given in Section 4.

2 Multi-stage Noise Suppression Network

The overall structure of the proposed MNSNet is shown in Fig. 1. The network mainly includes two branches: the Transformer branch and the CNN branch. The Transformer branch of MNSNet uses the cascaded partial decoder CPD [8] to aggregate the high-level polyp features f_i (i =1,2,3) extracted by the PVT backbone network to generate a global feature map S_g rich in global semantic information. The CNN branch of MNSNet uses the second, third and fourth level polyp features extracted by the ResNet backbone network to generate a polyp feature map to retain more local spatial detail information in the polyp image. Specifically, in the feature extraction stage, the CNN branch of MNSNet adopts the classic structure of UNet. The two proposed PBNS modules are used to process the polyp features f_i (i =4,5,6) and obtain a local feature map S_l rich in spatial detail information and reduce the influence of background noise interference. The local feature map S_{lSE} is processed by the PSE module [9] to obtain the enhanced local feature S_l. In the semantic fusion stage, the proposed PDNS module is used to fuse the global feature map S_g and the enhanced local feature S_{lSE} to obtain the final feature map S_o and improve the adaptability of the network to different polyp noises. The global feature map S_g, local feature map S_l and final feature map S_o are directly compared with the polyp segmentation true label GT to calculate the loss. The prediction result map Prediction is obtained by processing S_o with the Sigmoid function. The design of each component of MNSNet is introduced in detail below.

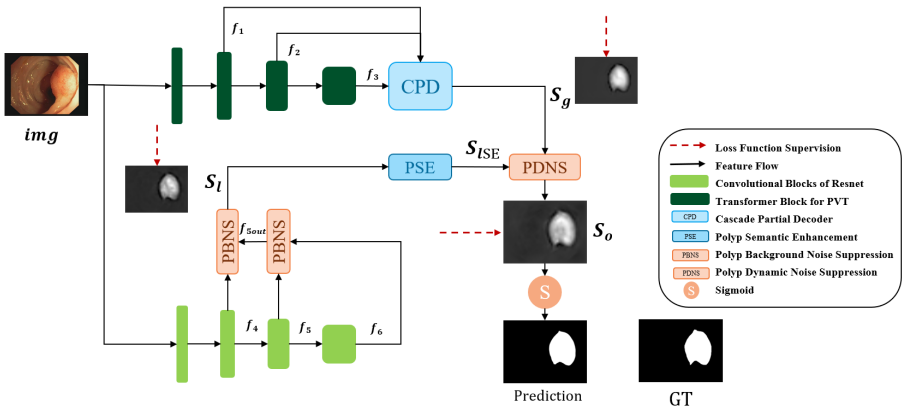

Fig. 1. Overview of the proposed MNSNet.

2.1 Polyp Background Noise Suppression Module

In order to retain more local detail information of polyp images, the MNSNet uses the second, third and fourth level polyp features in the ResNet50 backbone network to generate local feature maps. The low-level polyp features have a larger resolution and contain more spatial detail information, but also contain a lot of background noise

interference. In order to solve the problem of polyp background noise interference in the CNN branch of MNSNet, a PBNS module is designed based on the self-attention mechanism. The PBNS module activates the semantic information of the polyp target area in the network feature extraction stage, thereby improving the ability of MNSNet to resist background noise interference. The specific structure of PBNS is shown in Fig. 2.

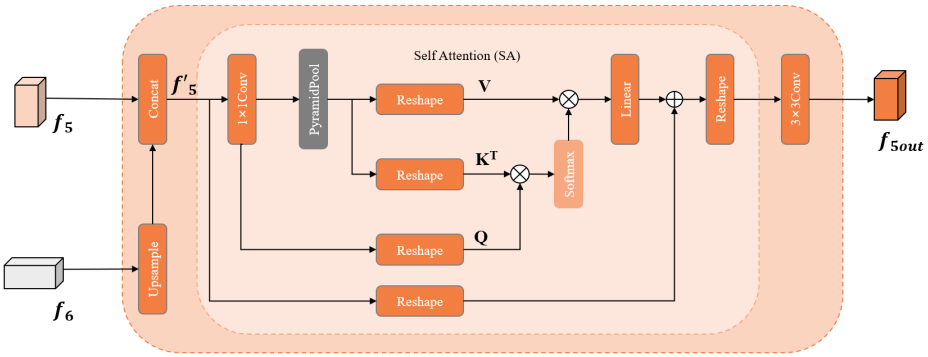

Fig. 2. Structure of the PBNS module.

Taking the background noise suppression operation of polyp features f_5 and f_6 as an example, the PBNS module can be expressed as:

$$f_{5out} = C\{SA[Cat(f_5, Up(f_6))]\}, \tag{1}$$

where $Up(\cdot)$ is an up-sampling operation, $Cat(\cdot)$ is the feature concatenation operation in the channel dimension, $SA(\cdot)$ is the background noise suppression operation based on self-attention, and $C(\cdot)$ is the semantic refinement operation of the polyp target area, which sequentially includes 3×3 convolutional layer, batch normalization layer and ReLU. The PBNS module uses up-sampling and feature concatenation operations to fuse polyp features of different levels and obtain f_5', which retains the original information of polyp features of different levels.

The background noise suppression operation $SA(\cdot)$ is the key to the PBNS module, and the self-attention calculation formula for this operation is:

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\times\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)\cdot\boldsymbol{V}, \tag{2}$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{K}$ are queries, keys, and values matrices respectively, and d_{k} is the dimension of each attention head. Different from the general Transformer structure, the background noise suppression operation uses pyramid pooling (PyramidPool) and reshape operations to generate the keys matrix $\boldsymbol{K}$ and the values matrix $\boldsymbol{V}$, and uses

the reshape operation to change the dimension of the input feature to generate the query matrix $\boldsymbol{Q}$. The introduction of pyramid pooling allows the background noise suppression operation to use smaller-scale global features to calculate the standard attention, which on the one hand enhances the multi-scale perception and global perception capabilities of the PBNS module, and on the other hand effectively reduces the amount of calculation. The background noise suppression operation uses pyramid pooling and reorganization to achieve self-attention modeling, and then improves the attention of low-level polyp features to the polyp target area through self-attention. Therefore, the PBNS module can activate the effective information of the polyp target area, thereby reducing the adverse effects of background noise interference on the polyp image segmentation network.

The PBNS module uses self-attention to activate the semantic information of the polyp target area and improve the feature contrast between the polyp target area and the background area, which reduces the influence of noise interference in the background area of the polyp image and further improves the polyp segmentation accuracy of MNSNet.

2.2 Polyp Dynamic Noise Suppression Module

The existing polyp image segmentation network that integrates Transformer and CNN usually uses a fixed kernel method to fuse polyp semantics extracted from different structures. These fusion methods lack anti-interference ability and cannot suppress complex and changeable polyp image noise. In order to solve the problem of insufficient anti-interference ability of polyp semantic fusion methods, a PDNS module is designed based on the dynamic kernel method to improve the adaptability of the polyp image segmentation network to different polyp image noise interference. The specific structure of the PDNS module is shown in Fig. 3.

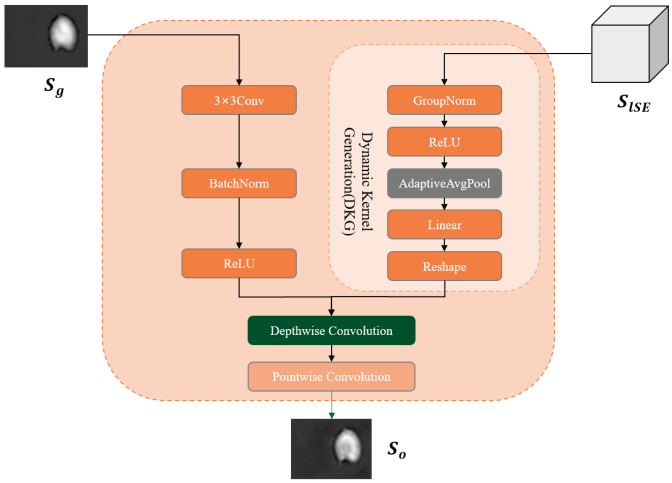

Fig. 4. Structure of the PDNS module.

In the semantic fusion stage of MNSNet, the PDNS module uses dynamic kernel generation operation to process local feature S_{lSE} to generate dynamic convolution kernels, and uses depthwise separable convolution to achieve deep fusion of polyp semantics and obtain the final feature map S_o. The PDNS module can be expressed as:

$$f_{5out} = C\{SA[Cat(f_5, Up(f_6))]\}, \tag{3}$$

where S_{lSE} is the local feature extracted and enhanced by the CNN branch, S_g is the global feature map extracted by the Transformer branch, $DSC(\cdot)$ is the depthwise separable convolution, $DKG(\cdot)$ is the dynamic kernel generation operation, and $C(\cdot)$ is the global semantic expansion operation, which sequentially includes 3×3 convolution layer, batch normalization Batch Norm layer, ReLU.

$DKG(\cdot)$ is the dynamic kernel generation operation, which includes group normalization operation (GroupNorm), ReLU, adaptive average pooling (AdaptiveAvgPool), Linear and Reshape in sequence. Where the group normalization operation groups and normalizes the input polyp local features S_{lSE}, helping the dynamic kernel generation operation to better understand the distribution and relationship of the input features; the ReLU activation function introduces nonlinear fitting capabilities, allowing the dynamic kernel generation operation to handle complex noise interference information; adaptive average pooling introduces global perception capabilities to the dynamic kernel generation operation, and generates the initial dynamic convolution kernel at the same time; Linear linearly transforms and projects the initial dynamic convolution kernel from the channel dimension, helping the dynamic kernel generation operation to achieve more refined feature representation; the Reshape operation adjusts the dimension of the initial dynamic convolution kernel so that it can be used as a dynamic convolution kernel to perform convolution operations on the expanded global feature map. The dynamic kernel generation operation enables the PDNS module to dynamically adjust the convolution kernel according to the noise information of the input data during the inference stage of the algorithm network, which improves the pertinence and flexibility of the PDNS module in handling different polyp image noise interference information, thereby improving the adaptability of the polyp image segmentation algorithm to complex and changeable polyp image noise interference.

Depthwise separable convolution $DSC(\cdot)$ sequentially includes depthwise convolution and pointwise convolution. Depthwise convolution performs independent convolution operations on each channel of the input feature, which can fuse the spatial information of different features. Pointwise convolution performs convolution operations on each pixel of the output feature of Depthwise convolution, which can integrate the information of the channel dimension. Depthwise separable convolution fuses the global feature map S_g and the local feature S_{lSE} from the spatial and channel dimensions, realizing the multi-dimensional fusion of polyp semantics. The introduction of Depthwise convolution further improves the flexibility and adaptability of the semantic fusion method, enabling the PDNS module to effectively handle the complex and changeable noise interference in polyp images.

The PDNS module dynamically adjusts the convolution kernel according to the noise interference information contained in the input features and uses depthwise separable convolution to achieve multi-dimensional semantic fusion, which improves the adaptability of MNSNet to different polyp noise interferences and improves the network's polyp segmentation performance.

2.3 Loss Function

The loss function is defined as：

$$\mathcal{L} = \mathcal{L}_{\mathrm{IoU}}^{\mathrm{w}} + \mathcal{L}_{\mathrm{BCE}}^{\mathrm{w}}, \tag{4}$$

where $\mathcal{L}_{\mathrm{IoU}}^{\mathrm{w}}$ denotes the weighted IoU loss [10] and $\mathcal{L}_{\mathrm{BCE}}^{\mathrm{w}}$ denotes the weighted binary cross entropy (BCE) loss [11]. Unlike the widely used standard IoU loss and standard BCE loss, weighted IoU loss and weighted BCE loss pay more attention to pixels with greater segmentation difficulty, which helps MNSNet increase its attention to pixels with noise interference, thereby improving the accuracy of network segmentation of polyp images with noise interference. Specifically, the loss is calculated for the local feature map $\boldsymbol{S}_l$ extracted by the CNN branch, the global feature map $\boldsymbol{S}_g$ extracted by the Transformer branch and the final feature map $\boldsymbol{S}_o$. The loss function can be expressed as:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}(\mathrm{GT}, \boldsymbol{S}_g) + \mathcal{L}(\mathrm{GT}, \boldsymbol{S}_l) + \mathcal{L}(\mathrm{GT}, \boldsymbol{S}_o), \tag{5}$$

where GT is the true label for polyp segmentation.

3 Comparative Experiments of Polyp Segmentation

In this section, the proposed MNSNet is compared with five representative polyp segmentation methods in experimental analysis.

3.1 Datasets and Implementation Details

The polyp segmentation performance of MNSNet is tested on five benchmark polyp segmentation datasets including Kvasir [12], CVC-ClinicDB [13], CVC-ColonDB [14], CVC-T [15] and ETIS [16]. The training set and test set are divided in the same way as SANet [3]. The training set consists of 900 images in Kvasir and 550 images in CVC-ClinicDB (1450 images in total). The test set consists of 100 images in Kvasir that are not involved in training, 62 images in CVC-ClinicDB that are not involved in training, 60 images in CVT-T, 380 images in CVC-ColonDB, and 196 images in ETIS (798 images in total).

MNSNet is implemented based on the PyTorch framework and trained on a single 3090 GPU for 50 epochs with mini-batch size 16. The resolution of all input images is uniformly resized to 352×352 and image scaling is used for data enhancement. Adam as the optimizer is applied with the learning rate 1e−4 during training.

3.2 Evaluation Metrics

Six evaluation metrics is used to evaluate the polyp segmentation performance of CSE-Net, including mean Dice (mDice), mean IoU (mIoU), mean absolute error (MAE), weighted F-measure (F_β^w), S-measure (S_α) and E-measure (E_φ^{m}). The lower value is better for the MAE and the higher is better for others.

3.3 Comparative Experiments of Polyp Segmentation

MNSNet was compared with five representative polyp segmentation methods, namely SANet [3], SSFormer [4], PPFormer [5], TransFuse [6] and Meta-Polyp [7]. SANet is a polyp segmentation method based on CNN, SSFormer and PPFormer are polyp segmentation methods based on Transformer, and TransFuse and Meta-Polyp are polyp segmentation methods that integrate Transformer and CNN.

As shown in Table 1, the majority of the metrics of MNSNet are better than the remaining five representative polyp segmentation methods on four benchmark data sets. Compared to the suboptimal method Meta-Polyp, MNSNet improves mDice and F_β^w by 1.0% and 1.4% on the CVC-ClinicDB dataset, and improves mDice and S_α by 2.2% and 1.5% on the CVC- ColonDB dataset. The PBNS module designed based on the self-attention mechanism can learn the relationship between different regions in the polyp image and increase the attention of MNSNet to the polyp target area, thereby reducing the adverse effects of background noise on MNSNet. In addition, the design of the PDNS module based on the dynamic kernel method can improve the pertinence and flexibility of the semantic fusion operation of MNSNet, which improves the adaptability of MNSNet to the complex and changeable noise interference in polyp images. The experiments show that the polyp segmentation accuracy of MNSNet is better than the other five representative methods and validate the effectiveness and superiority of MNSNet.

Table 1. Quantitative comparison results of MNSNet with 5 representative polyp segmentation methods

Dataset	Method	mDice	mIoU	F_β^w	S_α	E_φ^{m}	MAE
Kvasir	SANet	0.904	0.847	0.892	0.915	0.953	0.028
	SSFormer	0.917	0.864	0.910	0.925	0.956	0.023
	PPFormer	0.912	0.86	0.896	0.923	0.961	0.027
	TransFuse	0.915	0.86	0.906	0.919	0.956	0.023
	Meta-Polyp	0.916	0.869	0.907	0.926	0.959	**0.022**
	MNSNet(ours)	**0.921**	**0.871**	**0.910**	**0.928**	**0.962**	0.024
ClinicDB	SANet	0.916	0.859	0.909	0.939	0.976	0.012
	SSFormer	0.908	0.857	0.902	0.935	0.958	0.011
	PPFormer	0.919	0.876	0.915	0.943	0.969	0.008
	TransFuse	0.917	0.873	0.924	0.937	0.955	0.007
	Meta-Polyp	0.925	0.880	0.915	0.940	0.971	0.010
	MNSNet(ours)	**0.935**	**0.889**	**0.929**	**0.946**	**0.979**	**0.007**
CVC-T	SANet	0.888	0.815	0.859	0.928	0.972	0.008
	SSFormer	0.887	0.821	0.863	0.929	0.947	0.010
	PPFormer	0.872	0.800	0.842	0.919	0.958	0.011
	TransFuse	0. 870	0. 797	0. 844	0. 916	0. 943	0. 010
	Meta-Polyp	0.898	0.833	0.884	**0.935**	**0.973**	0.008
	MNSNet(ours)	**0.903**	**0.836**	**0.879**	0.932	0.965	**0.007**
ColonDB	SANet	0.753	0.670	0.726	0.837	0.878	0.043
	SSFormer	0.772	0.697	0.766	0.844	0.878	0.036
	PPFormer	0.791	0.709	0.769	0.850	0.907	0.033
	TransFuse	0.790	0.710	0.756	0.858	0.908	0.033
	Meta-Polyp	0.808	0.727	0.785	0.865	0.905	0.031
	MNSNet(ours)	**0.810**	**0.731**	**0.790**	**0.868**	**0.910**	**0.030**
ETIS	SANet	0.750	0.654	0.685	0.849	0.881	0.015
	SSFormer	0.767	0.698	0.736	0.863	0.857	0.016
	PPFormer	0.774	0.687	0.722	0.859	0.912	0.017
	TransFuse	0.748	0.657	0.695	0.85	0.835	0.018
	Meta-Polyp	0.772	0.692	0.738	0.854	0.877	0.022
	MNSNet(ours)	**0.794**	**0.712**	**0.747**	**0.869**	**0.887**	**0.013**

In order to qualitatively evaluate the segmentation effect of MNSNet, 4 images from test sets are used as examples to conduct comparative experiments with 5 representative methods. The experimental results are shown in Fig. 4. The experiment shows that compared with the other 5 representative polyp segmentation methods, MNSNet has

the best polyp segmentation effect. Specifically, Figure (a) in (Ⅰ) contains multiple white intestinal impurities, Figures (b) to (f) in (Ⅰ) show that the other methods misjudge some white impurities as polyp tissues, and Figure (f) in (Ⅰ) shows that MNSNet accurately distinguishes polyp tissues from intestinal impurities. Figure (a) in (Ⅱ) shows that the upper part of the long polyp is dark and the lower part has reflective interference. Figures (b) to (f) in (Ⅱ) show that only MNSNet can completely segment the polyp lesion area, and the segmentation effect of the other methods is poor. Figure (a) in (Ⅲ) has serious intestinal impurities and light spots in the background area. Figures (b) to (f) in (III) show that the other methods are affected by the noise interference in the background area, misjudging the impurities and light spots in the background area as polyp tissues, and Figure (f) in (III) shows that MNSNet can correctly perceive the complete polyp lesion area. Figure (a) in (IV) shows that the edge of the polyp has a low contrast with normal tissue and is covered with a large amount of mucus. Figures (b) to (f) in (II) show that only MNSNet has a good segmentation effect, and the other methods cannot accurately determine the polyp boundary while locating the polyp. In the polyp segmentation process, MNSNet uses dual branches to extract polyp features at different levels and scales in parallel, uses the PBNS module to increase the network's attention to the polyp target area, reduces the interference of background noise on the network, and uses the PSDF module to improve the flexibility and adaptability of the semantic fusion method and improve the adaptability of MNSNet to noise interference in polyp images. From the qualitative comparison experimental results, it can be seen that the polyp segmentation effect of MNSNet is better than the other five representative polyp segmentation methods, and it can better complete the polyp segmentation task under noise interference conditions.

3.4 Ablation Study

In order to verify the effectiveness of the constructed PBNS module and PDNS module, ablation comparison experiments were conducted on five datasets. The MNSNet without the PBNS module is recorded as "MNSNet-PBNS", and the MNSNet without the PDNS module is recorded as "MNSNet-PDNS".

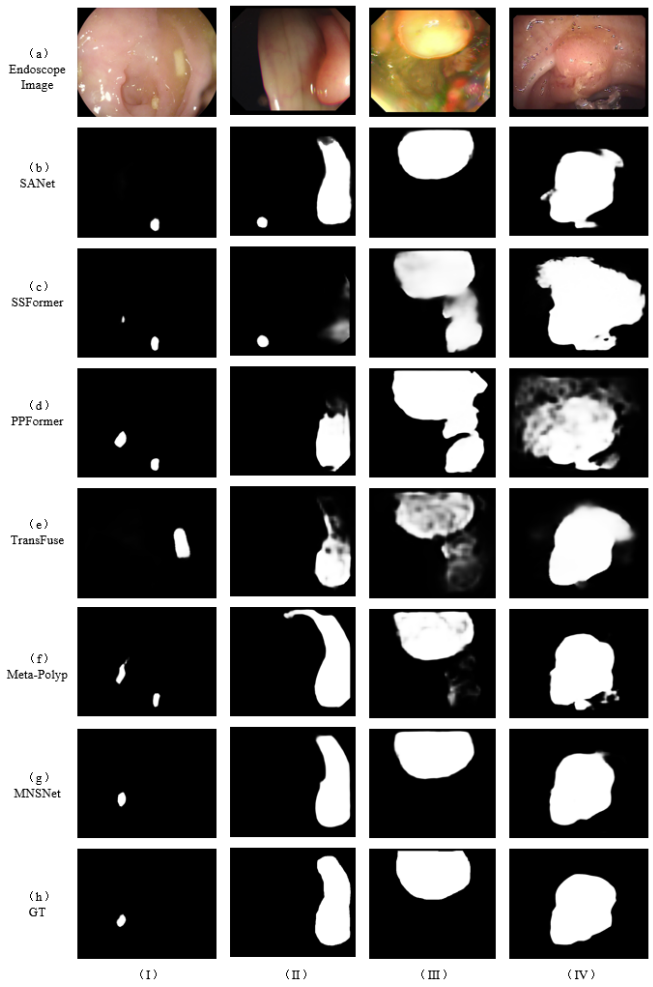

Fig. 5. Qualitative comparison results of MNSNet with 5 representative polyp segmentation methods.

As can be seen from Table 2., in the five data sets, most of the indicators of the complete MNSNet are better than those of the MNSNet without the PBNS module and the MNSNet without the PDNS module. Specifically, in the CVC-ClinicDB data set, compared with the MNSNet without the PBNS module, the complete MNSNet has improved the mIoU indicator and the F_{β}^{w} indicator by 2.4% and 2.6%; In the CVC-T data set, compared with the MNSNet without the PDNS module, the complete MNSNet has improved the mIoU indicator and the E_{φ}^{m} indicator by 3.6% and 2.5%. The PBNS module uses self-attention to improve the feature contrast between the polyp target area and the background area, so that MNSNet pays more attention to the polyp target area,

thereby suppressing the interference of background noise. The PDNS module adaptively adjusts the parameters of the convolution kernel according to different input data, improving the flexibility and adaptability of the fusion strategy, which enables MNSNet to effectively cope with complex and changeable noise interference. Experiments show that the PBNS module and the PDNS module improve the polyp segmentation accuracy of MNSNet and are necessary.

Table 2. Quantitative comparison results of ablation experiments with the PBNS module and PDNS module.

Dataset	Settings	mDice	mIoU	F_β^w	S_α	E_φ^{m}	MAE
Kvasir	MNSNet-PBNS	0.917	0.868	0.905	**0.928**	0.958	0.025
	MNSNet-PDNS	0.906	0.856	0.897	0.921	0.950	0.028
	MNSNet	**0.921**	**0.871**	**0.910**	0.928	**0.962**	**0.024**
ClinicDB	MNSNet-PBNS	0.913	0.865	0.903	0.939	0.960	0.012
	MNSNet-PDNS	0.928	0.886	0.925	0.946	0.975	0.009
	MNSNet	**0.935**	**0.889**	**0.929**	**0.946**	**0.979**	**0.007**
CVC-T	MNSNet-PBNS	0.893	0.827	0.869	**0.933**	0.955	0.008
	MNSNet-PDNS	0.868	0.800	0.838	0.919	0.940	0.01
	MNSNet	**0.903**	**0.836**	**0.879**	0.932	**0.965**	**0.007**
ColonDB	MNSNet-PBNS	0.796	0.712	0.774	0.858	0.902	**0.029**
	MNSNet-PDNS	0.793	0.716	0.770	0.857	0.892	0.031
	MNSNet	**0.810**	**0.731**	**0.790**	**0.868**	**0.910**	0.030
ETIS	MNSNet-PBNS	0.777	0.698	0.724	0.861	0.861	0.018
	MNSNet-PDNS	0.784	0.708	0.73	0.869	0.858	0.015
	MNSNet	**0.794**	**0.712**	**0.747**	**0.869**	**0.887**	**0.013**

In order to qualitatively evaluate the contribution of the PBNS and PDNS modules, 4 images from test sets are used as examples for ablation experiments, and the experimental results are shown in Fig. 6. The experiment shows that the MNSNet that contains both the PBNS module and the PBSF module has the best polyp segmentation effect. Specifically, in Figure (a) of (I), there is a bubble in the upper left corner of the polyp tissue that interferes with the endoscopic imaging. Figures (b) and (c) of (I) show that the MNSNet without the PBNS module or the PDNS module misjudges the bubble as polyp tissue. Figure (d) of (I) shows that the MNSNet method with the PBNS module and the PDNS module added can accurately distinguish between polyp tissue and bubble interference. The left half of Figure (a) of (II) is dark and there is an obvious light spot interference. Figures (b), (c) and (d) of (II) show that only the complete MNSNet can identify the light spot interference and correctly perceive the polyp area. Figure (a) in (III) has mirror reflection interference in the lower left corner and multiple white intestinal impurities in the right half. Figures (b) and (c) in (II) show that MNSNet without the PBNS module or PDNS module cannot handle these interferences correctly. Figure (d) in (III) shows that the complete MNSNet can correctly locate the polyp and segment the complete polyp tissue. Figure (a) in (IV) has blurred intestinal peristalsis

imaging in the left half. Figures (b) and (c) in (IV) show that MNSNet without the PBNS module or PDNS module misjudges the blurred area as polyp tissue. Figure (d) in (IV) shows that the complete MNSNet can segment the polyp area well. In the feature extraction stage, the PBNS module helps the CNN branch focus on the polyp target area based on the self-attention mechanism, reduces the sensitivity of MNSNet to background noise, and improves the robustness of MNSNet. In the semantic fusion stage, the PDNS module enriches the information representation of the model based on the dynamic kernel, improves the flexibility of the fusion method, and enables MNSNet to effectively handle the complex and changeable noise interference in the polyp image. From the ablation experiment results, it can be seen that the PBNS module and the PDNS module improve the segmentation effect of MNSNet and are necessary.

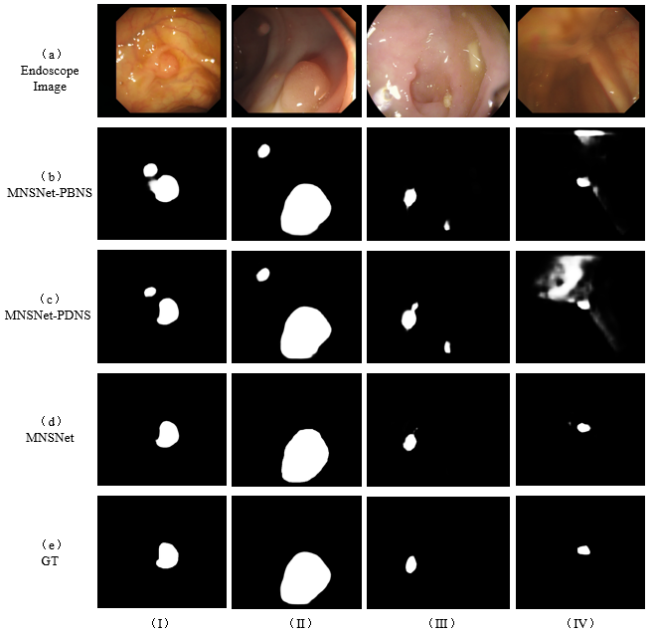

Fig. 7. Qualitative comparison results of ablation experiments with the PBNS module and PDNS module.

4 Conclusion

MNSNet is proposed to improve the anti-noise interference ability of polyp segmentation methods, thereby improving the segmentation accuracy of polyp endoscopic images. Firstly, parallel CNN branches and Transformer branches are constructed to extract polyp features respectively, which improves the global and local information acquisition capabilities of the network, thereby improving the robustness of the network. Secondly, in the polyp feature extraction stage, a polyp background noise removal module is designed based on the self-attention mechanism, which increases the network's attention to the polyp target area and improves the network's anti-interference ability to

background noise. Finally, in the polyp semantic fusion stage, a polyp noise dynamic suppression module is designed based on the dynamic kernel method, and a variable convolution kernel is dynamically generated according to the input data noise information, which improves the flexibility of the semantic fusion method and thereby improves the network's adaptability to noise interference of various types of polyp images. Comparative experiments are conducted with 5 representative polyp segmentation methods under 5 benchmark polyp segmentation datasets to evaluate the polyp segmentation performance of MNSNet. Compared with the suboptimal method Meta-Polyp, MNSNet improves the mDice and mIoU indicators by 2.2% and 2.0% respectively under the ETIS dataset. Experimental results show that CSENet has the best polyp segmentation performance, combining effectiveness and superiority.

References

1. S. Johanna, A. Marko, L. Thomas, et al. Impact of AI-aided colonoscopy in clinical practice: a prospective randomised controlled trial. BMJ Open Gastroenterology, 2024, 11(01): 470-481.
2. Westerberg M, Holmberg L, Ekbom A, et al. The role of endoscopist adenoma detection rate in in sex differences in colonoscopy findings: Cross-sectional analysis of the SCREESCO randomized controlled trial[J]. Scandinavian Journal of Gastroenterology, 2023, 59(4): 503-511.
3. J. Wei, Y. Hu, R. Zhang, et al. Shallow attention network for polyp segmentation[C]. 2021 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2021: 699-708.
4. Wang, J., Huang, Q., Tang, F., et al. Stepwise Feature Fusion: Local Guides Global[C]. 2022 Medical Image Computing and Computer Assisted Intervention (MICCAI). 2022: 110–120.
5. Linghan Cai, Meijing Wu, Lijiang Chen, et al. Using Guided Self-Attention with Local Information for Polyp Segmentation[C]. 2022 Medical Image Computing and Computer Assisted Intervention (MICCAI). 2022, 13434: 629–638.
6. Zhang, Y., Liu, H., Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. 2022 Medical Image Computing and Computer Assisted Intervention (MICCAI). 2022, 12901: 14-24.
7. Q. Trinh. Meta-Polyp: A Baseline for Efficient Polyp Segmentation[C]. 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), 2023: 742-747.
8. Z. Wu, L. Su, Q. Huang. Cascaded partial decoder for fast and accurate salient object detection[C]. 2019 Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 3902-3911.
9. M. Lin, K. Hirota, Y. Dai, et al. A Cascaded Semantic Enhancement Network Based on Attention Mechanism for Blurred Small Polyp Segmentation[C]. 2023 42nd Chinese Control Conference (CCC), 2023: 8240-8245.
10. Qin, X., Zhang, Z., Huang, C., Gao, et al. Boundary-aware salient object detection[C]. 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 7479-7489.
11. Wei, J., Wang, S., Huang, Q. F3Net: Fusion, Feedback and Focus for Salient Object Detection[C]. 2020 AAAI Conference on Artificial Intelligence, 34(07): 12321-12328, 2020.

12. Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, et al. Kvasir-SEG: A Segmented Polyp Dataset[C]. 2020 International Conference on Multi-Media Modeling (MMM), 2020: 451–462.
13. Bernal, J., S´ anchez, F.J., et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized Medical Imaging and Graphics, 2015, 43(7): 99–111.
14. Tajbakhsh, N., Gurudu, S.R., Liang, J. Automated polyp detection in colonoscopy videos using shape and context information[J]. IEEE Transactions on Medical Imaging (TMI), 35(2): 630–644.
15. V´ azquez, D., Bernal, J., S´ anchez, et al. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. Journal of Healthcare Engineering, 2017: 531-543.
16. Silva, J., Histace, A., Romain, et al. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer[J]. International Journal of Computer Assisted Radiology and Surgery, 2014, 9(2): 283-293.