

Research Article

MHA-Net: A Multibranch Hybrid Attention Network for Medical Image Segmentation

Meifang Zhang,¹ Qi Sun ,² Fanggang Cai ,³ and Changcai Yang²

¹Department of Health Management, Fujian Health College, Fuzhou 350101, China

²College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China

³Department of Vascular Surgery, the First Affiliated Hospital, Fujian Medical University, Fuzhou 350005, China

Correspondence should be addressed to Qi Sun; sunqi9696@126.com and Fanggang Cai; chfafu@qq.com

Received 15 June 2022; Revised 11 September 2022; Accepted 13 September 2022; Published 6 October 2022

Academic Editor: Luminita Moraru

Copyright © 2022 Meifang Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The robust segmentation of organs from the medical image is the key technique in medical image analysis for disease diagnosis. U-Net is a robust structure for medical image segmentation. However, U-Net adopts consecutive downsampling encoders to capture multiscale features, resulting in the loss of contextual information and insufficient recovery of high-level semantic features. In this paper, we present a new multibranch hybrid attention network (MHA-Net) to capture more contextual information and high-level semantic features. The main idea of our proposed MHA-Net is to use the multibranch hybrid attention feature decoder to recover more high-level semantic features. The lightweight pyramid split attention (PSA) module is used to connect the encoder and decoder subnetwork to obtain a richer multiscale feature map. We compare the proposed MHA-Net to state-of-art approaches on the DRIVE dataset, the fluoroscopic roentgenographic stereophotogrammetric analysis X-ray dataset, and the polyp dataset. The experimental results on different modal images reveal that our proposed MHA-Net provides better segmentation results than other segmentation approaches.

1. Introduction

The robust organ segmentation from medical images is essential for medical image analysis. Its critical task is to segment objects of interest (e.g., lesions or organs) in medical images, providing important significance and value for accurate identification, rational diagnosis, prediction, prevention of diseases, etc. However, medical image segmentation is still a significantly tricky task because of the poor quality of medical images with a low signal-to-clutter ratio. In addition, organs of interest are often buried in complex backgrounds with muscles, blood vessels, bones, etc.

Many traditional medical image segmentation approaches have been proposed in the past decades. Massoptier and Casciaro [1] used the level set method to compute a smoother liver surface and then adopted automatic classification to detect the hepatic lesions. A semiautomatic level set method, which includes the spiral-scanning approach and the statistical pixel classification method, was proposed by Smeets et al. [2]

for liver tumors combining segmentation. A multiconcavity modeling vessel segmentation method, which combines both line-shape concavity measure and locally normalized concavity measure, was proposed by Lam et al. [3]. Azzopardi et al. [4] proposed trainable COSFIRE filters for retinal vessel image segmentation. The common drawback of the traditional supervised learning image segmentation algorithms is that they depend on accurate and complex feature extraction. The modeling and selection of these features require not only professional knowledge but also difficulty to obtain, which makes the quality of the feature model directly impact image segmentation and classification, thus affecting the versatility and accuracy of the algorithm.

With the development of machine learning in image and video analyses, feature automatic learning methods using convolutional neural networks (CNN) have become a viable method for medical image segmentation. Unlike classic pixel classification methods that typically utilize hand-crafted image features, CNN methods learn image features and solve

hand-crafting problems. The U-Net [5] is one of the well-known medical image segmentation network structures, showing superior performance in neuronal structure segmentation and cellular segmentation. The U-Net that fuses feature maps from different stages by hopping connections is an encoder-decoder model. The spatial dimensionality of the feature maps is progressively reduced by using the encoder. Then, more high-level semantic features can be obtained. The decoder is used to find the details and spatial dimensions of the medical image. Wang et al. [6] built an efficient dual encoder U-Net (DEU-Net) to improve the pixel-to-pixel segmentation performance of retinal vessels. Alom et al. [7] developed a novel recurrent residual U-Net (R2U-Net), which has shown good performance in many biomedical image applications. Wu et al. [8] presented a Vessel-Net, which combines the perceptual and residual models for medical image segmentation. Yang et al. [9] proposed a retinal vessel segmentation model (MSFFU-Net) incorporating the multiscale features. Samuel and Veeramalai [10] proposed a two-stage vessel segmentation method to selectively learn the appropriate blood vessels. Xu et al. [11] developed a dual-context network for segmenting medical images aggregating multiscale and global contexts. In this network, the global context is recovered using the global context transformer consisting of a new adaptive context fusion module which is proposed to capture the global context. Lachinov et al. [12] proposed projective skip connections between an encoder and a decoder to address medical image segmentation problems in the subset of dimensions. These CNN-based approaches have achieved good performances. However, the consecutive downsampling encoders of FCN-based U-Net and its varieties result in the loss of contextual information and insufficient recovery of high-level semantic features during upsampling, which is not conducive to medical image segmentation. Khan et al. [13] developed a deep hybrid network (DH-Net), which combines DenseNet with U-Net, to classify the land cover in satellite images. DenseNet builds the multiple-scale feature extraction. Khan et al. [14] proposed an encoder-decoder network (EE-Net) to detect the building footprints in satellite images. These two methods have achieved good results in remote sensing image segmentation, but the effectiveness of medical image segmentation needs to be further verified.

In this study, we explore a new multibranch hybrid attention network (MHA-Net) for medical image segmentation that can recover more high-level semantic features. The proposed MHA-Net mainly consists of an encoder module with the pretrained ResNet, a lightweight pyramid split attention (PSA) [15], and a novel multibranch hybrid attention (MHA) feature decoder module. The U-Net lacks the ability to integrate the multiscale context due to consecutive convolution and pooling operations in the decoders. We use PSA as a bridge to connect the encoder and decoder to recover more multiscale spatial information by infusing four groups with a multiscale pyramid convolution structure. The proposed MHA block could recover more high-level semantic information by adopting concatenation and summation operations with the corresponding feature maps. The contributions of this paper can be listed as follows.

Firstly, we propose a novel multibranch hybrid attention feature decoder to recover more high-level semantic features.

Secondly, we propose an MHA-Net for medical image segmentation, which contains a feature encoder, a lightweight PSA connecting the encoder and the decoder, and our proposed multibranch hybrid attention feature decoder.

Finally, the experimental results on the DRIVE dataset, the fluoroscopic roentgenographic stereophotogrammetric analysis (FRSA) X-ray dataset, and the polyp dataset compared with the standard segmentation methods show that our proposed MHA-Net is better.

2. MHA-Net

Figure 1 illustrates the schematic of the proposed MHA-Net. It contains a feature encoder block, a feature decoder block, and a multiscale PSA module that connects the encoder and decoder subnetworks. Next, we describe each part of the proposed MHA-Net in details.

2.1. Feature Encoder Block. The encoder block in U-Net consists of two 3×3 convolutions with a ReLU and a 2×2 max pooling. In this paper, following the Ce-Net [16], a pretrained ResNet-34 [17] is used as the feature encoder block. The pretrained weight parameters on the ImageNet dataset are used as parameter initialization.

2.2. PSA. To obtain more multiscale spatial information and high-level semantic information from the medical image, the lightweight PSA module [15] is used as a bridge to connect the encoder and bottom of decoder subnetworks. The feature map of the encoder is split into four groups, and each group has 128 channels. Following [15], the convolution kernel parameters are set to be 3, 5, 7, and 8. And the group convolution parameters are set to be 1, 4, 8, and 16. A multiscale pyramid convolution structure, which can more accurately merge adjacent scales of context features, is adopted to integrate the feature of different-scale feature maps on each channel. The channel attention weight of the multiscale feature map is obtained by using the SEWeight module [18]. Then, the cross-dimensional interactions can be established. The attention weights of the correlated channels, which establish the remote channel dependencies, are recalculated by using the Softmax operation. Finally, the multiscale feature maps are multiplied with the corrected attention vectors to extract a richer multiscale feature map passed to the feature decoder block.

2.3. Proposed Feature Decoder Block. In U-Net, the decoder block contains an upsampling operation, a concatenation operation, and two convolutions with a ReLU. The concatenation operation uses the skip connection to capture some context from the encoder. However, unlike the U-Net, the proposed decoder consists of our multibranch hybrid attention block, which can recover more high-level semantic features. A novel feature decoder, which contains a multibranch hybrid attention (MHA) block, a transposed convolution, and two 3×3 convolutions, is proposed. Our proposed MHA block contains a transposed convolution, a concatenation with the correspondingly cropped feature maps, two 3×3 convolutions, channel attention, and a

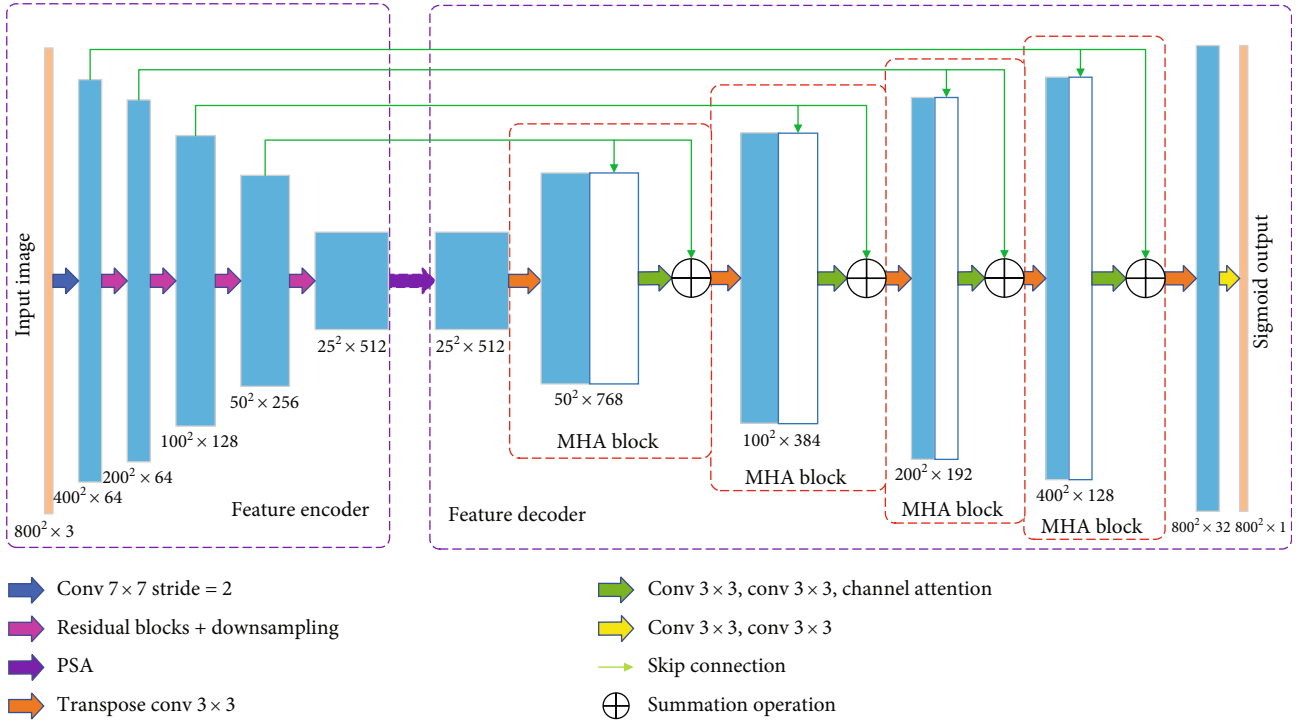


FIGURE 1: The proposed MHA-Net with a $800 \times 800 \times 3$ input data size. The proposed MHA-Net contains a feature encoder block, the proposed MHA decoder block, and a PSA module that connects the encoder and decoder subnetworks.

summation with the corresponding cropped feature maps consecutively. The architecture of the multibranch hybrid feature decoder with an $800 \times 800 \times 3$ input data size is shown in Figure 2. The $25 \times 25 \times 512$ feature map is input into the proposed feature decoder. We use a transposed convolution to upsample the feature map and obtain a feature map with a $50 \times 50 \times 512$ size. Then, this feature map and its corresponding feature map from the feature encoder are concatenated and a feature map with a $50 \times 50 \times 768$ size can be obtained. Two 3×3 convolutions and channel attention are used to selectively weigh each channel's significance, and a feature map with a $50 \times 50 \times 256$ size is obtained. After that, we make a summation operation between it and its corresponding feature map from the feature encoder. We obtain a feature map with a $100 \times 100 \times 384$ size. After the other three multibranch hybrid attention blocks, we obtain a feature map with a $800 \times 800 \times 32$ size. Finally, the segmentation result is obtained using two 3×3 convolution operations.

3. Experimental and Discussion

To verify the proposed MHA-Net performance for medical image segmentation, we test it on the DRIVE dataset [19] and the FRSA X-ray dataset [20]. The DRIVE dataset contains 20 trained and 20 tested retinal images with a $584 \times 565 \times 3$ size. The FRSA X-ray dataset includes 76 trained and 26 tested images with a $800 \times 800 \times 3$ size. To evaluate the generalization ability of the proposed MHA-Net, we adopt the polyp dataset, which contains Kvasir [21], ClinicDB [22], and ColonDB [23] datasets. The Kvasir, ClinicDB, and ColonDB datasets contain 1000, 612, and 380

images, respectively. Following [24], we use 90% of Kvasir and ClinicDB data as a training set, the remaining 10% as a validation set, and the ColonDB dataset as a test set. The size of all polyp data is set to $352 \times 352 \times 3$. In this paper, we use the binary crossentropy as the loss function. We use the batch size of 4 for the DRIVE dataset, 4 for the FRSA X-ray dataset, and 8 for the polyp dataset. We adopt the Adam optimizer with initial learning rate 0.0001 to optimize all models. We set the epoch to 300 for the DRIVE dataset and 150 for the FRSA X-ray dataset and polyp dataset. The parameter size of the proposed MHA-Net is 34.79 M.

3.1. Experimental with DRIVE Data. The proposed MHA-Net is compared to U-Net [5], JSPL-Net [25], Ce-Net [16], IterNet [26], AGAC-Net [27], DAP [28], and VSSC-Net [10]. In this paper, three common performance measures, area under the receiver operating characteristic curve (AUC), sensitivity (Se), and accuracy (Ac), are adopted as standards for our evaluation in retinal image segmentation.

Table 1 presents the values of the quality of AUC, Se, and Ac for each medical image segmentation method. Table 1 shows that our proposed MHA-Net provides the best segmentation performances in light of AUC, Se, and Ac. For example, the AUC score of our proposed MHA-Net is 0.9864 and is over 0.9834, 0.9752, 0.9831, 0.9813, 0.9847, 0.9788, and 0.9789, given by the U-Net, R2U-Net, Ce-Net, Ce-Net, IterNet, AGAC-Net, DAP, and VSSC-Net methods. We can find similar results for Se and Ac scores.

Figure 2 shows segmentation results by different approaches on DRIVE images. Column 1 in Figure 2 shows two original retinal images. Columns 2–4 in Figure 2 show

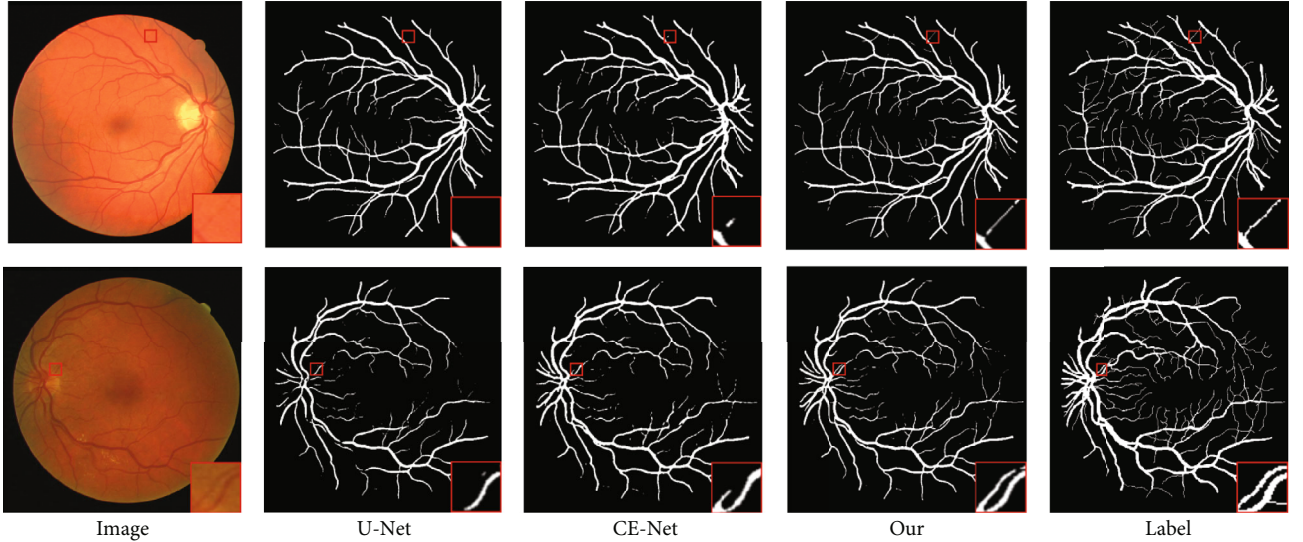


FIGURE 2: Retinal images and segmentation results provided by different approaches.

TABLE 1: AUC, Se, and Ac Values on the DRIVE dataset by difference methods.

Method	Year	AUC	Se	Ac
U-Net [5]	2015	0.9834	0.8059	0.9627
JSPL-Net [25]	2018	0.9752	0.7653	0.9542
Ce-Net [16]	2019	0.9831	0.8330	0.9542
IterNet [26]	2020	0.9813	0.7791	0.9574
AGAC-Net [27]	2020	0.9847	0.7941	0.9558
DAP [28]	2021	0.9788	0.8227	0.9545
VSSC-Net [10]	2021	0.9789	0.7827	0.9627
Our approach	2022	0.9864	0.8412	0.9636

the corresponding segmentation results provided by U-Net, Ce-Net, and our proposed MHA-Net, respectively. Column 5 in Figure 2 show the ground truth. Carefully observing Figure 2, we can find that the segmentation results given by our proposed MHA-Net are superior to that provided by U-Net and Ce-Net. For example, our proposed MHA-Net provides a little better segmentation accuracy, as shown in the red circled region in Figure 2.

3.2. Experimental with FRSA X-Ray Data. The proposed MHA-Net is compared to U-Net [5], Ce-Net [16], RSAN [29], and SA-UNet [30]. In this section, three common performance measures, Dice coefficient (DC), intersection ratio (IR), and Ac, are used as standards for our evaluation of X-ray image segmentation:

$$\begin{aligned}
 DC &= \frac{2 \times TP}{(TP + FP) + (TP + FN)}, \\
 IR &= \frac{TP}{TP + FN + FP}.
 \end{aligned}
 \tag{1}$$

TABLE 2: DC, IR, and Ac values of difference segmentation methods on the FRSA X-ray dataset.

Method	Year	DC	IR	Ac
U-Net [5]	2015	0.9264	0.8678	0.9980
Ce-Net [16]	2019	0.9623	0.9276	0.9990
RSAN [29]	2020	0.9389	0.8848	0.9985
SA-UNet [30]	2021	0.9204	0.8525	0.9980
Our approach	2022	0.9645	0.9325	0.9992

Table 2 gives the values of the quality of DC, IR, and Ac on the FRSA X-ray images for different segmentation methods. Table 2 shows that our proposed MHA-Net provided the best segmentation performances in terms of DC, IR, and Ac scores. The DC score of our proposed MHA-Net is 0.9645 and is over 0.9264, 0.9623, 0.9389, and 0.9204, respectively, obtained by the U-Net, Ce-Net, RSAN, and SA-UNet methods. We can find similar results for IR and Ac scores.

Column 1 in Figure 3 shows two original stent images. Columns 2–6 of Figure 2 show the stent segmentation results by U-Net, Ce-Net, RSAN, our proposed MHA-Net, and the ground truth. The segmentation effect of our MHA-Net is better than that given by U-Net, Ce-Net, and RSAN. For example, our proposed MHA-Net provides a little better segmentation accuracy, as shown in the red-circled region in Figure 3.

3.3. Experimental with the Crossdataset. The proposed MHA-Net is compared to U-Net [5], UNet++ [31], Ce-Net [16], PraNet [24], and HarDNet-MSEG [32]. In this section, eight common performance measures, Recall(Rec), Specificity(Spec), Precision(Prec), DC, IoU_poly(IoUp), IoU_

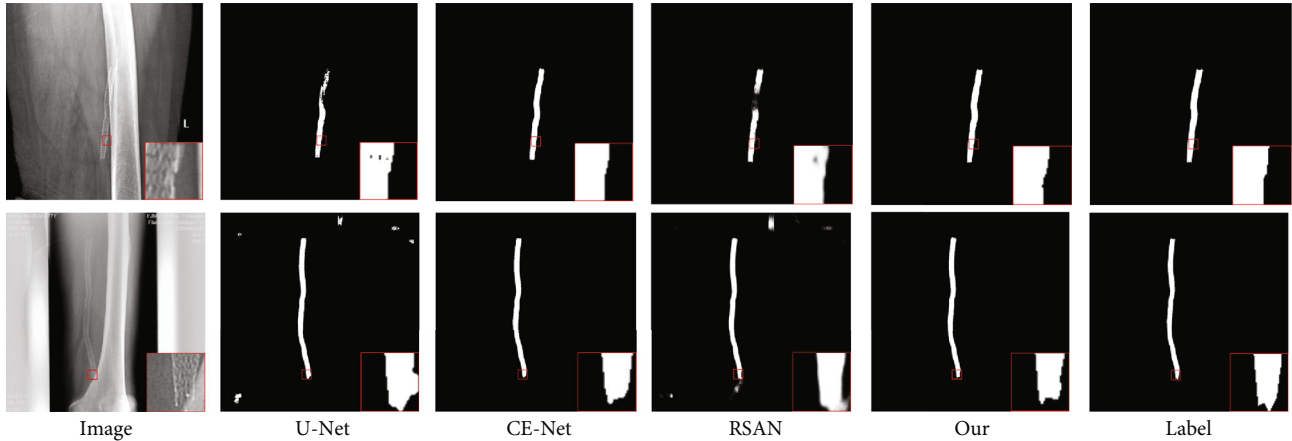


FIGURE 3: FRSA X-ray images and segmentation results provided by different approaches.

TABLE 3: Eight standard performance measure values of different segmentation methods on the polyp dataset.

Method	Rec	Spec	Prec	DC	IoUp	IoUb	mIoU	Ac
U-Net [5]	0.6530	0.9940	0.8760	0.6481	0.5762	0.9564	0.7663	0.9588
UNet++ [31]	0.7705	0.9565	0.6375	0.6385	0.5401	0.9460	0.7431	0.9481
Ce-Net [16]	0.8129	0.9701	0.8167	0.7569	0.6873	0.9471	0.8172	0.9504
PraNet [24]	0.7820	0.9862	0.8110	0.7435	0.6644	0.9607	0.8126	0.9635
HarDNet-MSEG [32]	0.5886	0.9988	0.9255	0.6490	0.5736	0.9563	0.7649	0.9578
Our approach	0.8000	0.9905	0.8618	0.7692	0.6982	0.9640	0.8311	0.9656

bg(IoUb), IoU_mean(mIoU), and Ac, are used as standards for our evaluation in polyp image segmentation:

$$\begin{aligned}
 \text{Rec} &= \frac{TP}{TP + FN}, \\
 \text{Spec} &= \frac{TN}{TN + FP}, \\
 \text{Prec} &= \frac{TP}{TP + FP}, \\
 \text{IoUp} &= \frac{\text{IoUp}}{TP + FP + FN}, \\
 \text{IoUb} &= \frac{TN}{TN + FP + FN}, \\
 \text{mIoU} &= \frac{\text{IoUp} + \text{IoUb}}{2.0}.
 \end{aligned} \tag{2}$$

We verify the model's generalization ability on Kvasir [21], ClinicDB [22], and ColonDB [23]. Table 3 presents the values of quality of Rec, Spec, Prec, DC, IoUp, IoUb, mIoU, and Ac on the polyp images for different segmentation methods. Table 3 shows that the proposed MHA-Net provided the best segmentation performances in terms of scores of Rec, DC, IoUp, IoUb, mIoU, and Ac. For Spec and Prec metrics, HarDNet-MSEG achieves the best results, followed closely by U-Net and the proposed MHA-Net.

TABLE 4: Ablation experiments on the femoropopliteal stent dataset.

Baseline	PSA	SE	Sum	DC	IR	Ac
✓				0.9581	0.9203	0.9990
✓	✓			0.9624	0.9282	0.9991
✓		✓		0.9614	0.9262	0.9991
✓	✓	✓	✓	0.9645	0.9325	0.9992

Table 3 shows that the proposed MHA-Net has good model generalization ability.

3.4. Ablation Study. To evaluate the effectiveness of individual components in the proposed MHA-Net, we performed ablation experiments on the femoral-popliteal stent dataset. We refer to the network without PSA, channel attention (SE), and summation (Sum) operation as our baseline. Table 4 shows that comparing the baseline with the PSA module can improve the DC, IR, and Ac indicators, showing that the proposed PSA module captures rich multiscale features. Combining the channel attention to compare the baseline can also improve the DC, IR, and Ac metrics. Channel attention values useful features while suppressing ones not important to the current task. Combining the PSA module, the SE module, and the Sum operation, our network can obtain the best performance and Sum operation can obtain more advanced semantic information.

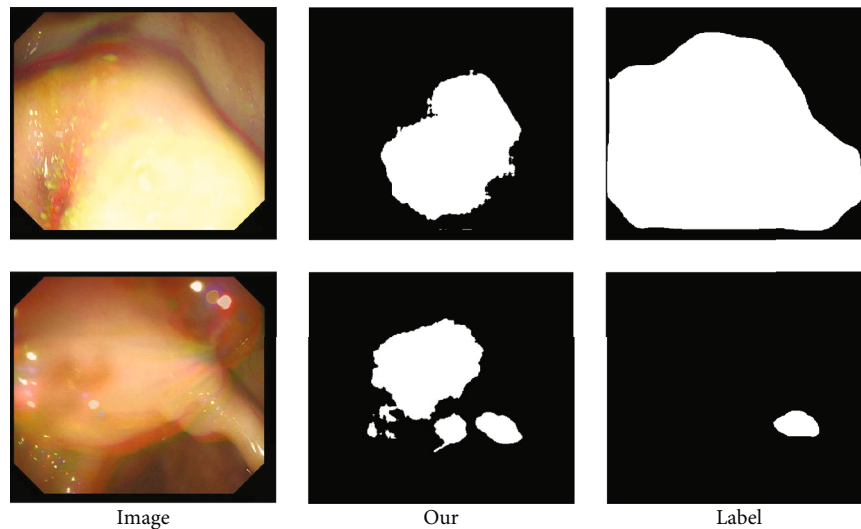


FIGURE 4: Examples show that our model cannot fully segment medical images.

3.5. Failure Cases. The proposed MHA-Net does not successfully address all cases of medical segmentation problems. Figure 4 shows the segmentation results of the ColonDB dataset. The proposed MHA-Net does not segment polyps very well, as shown in Figure 4. In polyp data, the size and shape of the lesion vary greatly. Moreover, the boundary of the lesion area is often blurred due to the low contrast between the lesion area and the surrounding area. This problem is not well addressed using our proposed model. However, this issue also affects most other state-of-the-art image segmentation methods.

4. Conclusions

In this study, we propose the multibranch hybrid attention feature decoder block and present an MHA-Net for medical image segmentation. Experimental results on DRIVE dataset, FRSA X-ray dataset, and polyp dataset show that our MHA-Net is superior to other segmentation approaches, including U-Net [5], JSPL-Net [25], Ce-Net [16], IterNet [26], AGAC-Net [27], DAP [28], VSSC Net [10], UNet++ [31], PraNet [24], and HarDNet-MSEG [32]. The multibranch hybrid attention feature decoder of the proposed MHA-Net can recover more high-level semantic features.

Data Availability

The segmentation results for the proposed method data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China with Grant 62171130 and in part by the Fujian Province Health Education Joint Research Project with Grant 2019WJ28.

References

- [1] L. Massotier and S. Casciaro, "A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from ct scans," *European Radiology*, vol. 18, no. 8, pp. 1658–1665, 2008.
- [2] D. Smeets, D. Loeckx, B. Stijnen, B. D. Dobbelaer, D. Vandermeulen, and P. Suetens, "Semi-automatic level set segmentation of liver tumors combining a spiral-scanning technique with supervised fuzzy pixel classification," *Medical Image Analysis*, vol. 14, no. 1, pp. 13–20, 2010.
- [3] B. S. Y. Lam, Y. Gao, and A. W.-C. Liew, "General retinal vessel segmentation using regularization-based multiconcavity modeling," *IEEE Transactions on Medical Imaging*, vol. 29, no. 7, pp. 1369–1381, 2010.
- [4] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical Image Analysis*, vol. 19, no. 1, pp. 46–57, 2015.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [6] B. Wang, S. Qiu, and H. He, "Dual encoding u-net for retinal vessel segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 84–92, Springer, 2019.
- [7] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," 2018, <https://arxiv.org/abs/1802.06955>.

- [8] Y. Wu, Y. Xia, Y. Song et al., “Vessel-Net: retinal vessel segmentation under multi-path supervision,” in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2019*, pp. 264–272, Springer International Publishing, Cham, 2019.
- [9] D. Yang, G. Liu, M. Ren, B. Xu, and J. Wang, “A multi-scale feature fusion method based on u-net for retinal vessel segmentation,” *Entropy*, vol. 22, no. 8, p. 811, 2020.
- [10] P. M. Samuel and T. Veeramalai, “Vssc net: vessel specific skip chain convolutional network for blood vessel segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 198, article 105769, 2021.
- [11] R. Xu, C. Wang, S. Xu, W. Meng, and X. Zhang, “Dc-net: dual context network for 2d medical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 503–513, Springer, 2021.
- [12] D. Lachinov, P. Seeböck, J. Mai, F. Goldbach, U. Schmidt-Erfurth, and H. Bogunovic, “Projective skip-connections for segmentation along a subset of dimensions in retinal OCT,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 431–441, Springer, 2021.
- [13] S. D. Khan, L. Alarabi, and S. Basalamah, “Deep hybrid network for land cover semantic segmentation in high-spatial resolution satellite images,” *Information*, vol. 12, no. 6, 2021.
- [14] S. D. Khan, L. Alarabi, and S. Basalamah, “An encoder-decoder deep learning framework for building footprints extraction from aerial imagery,” *Arabian Journal for Science and Engineering*, pp. 1–22, 2022.
- [15] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, “EPSANet: an efficient pyramid split attention block on convolutional neural network,” <https://arxiv.org/abs/2105.14447>.
- [16] Z. Gu, J. Cheng, H. Fu et al., “Ce-net: context encoder network for 2d medical image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [18] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, Salt Lake City, 2018.
- [19] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [20] Q. Sun, F. Cai, and C. Yang, *Femoropopliteal Artery Stent Dataset*, IEEE Dataport, 2022.
- [21] D. Jha, P. H. Smedsrud, M. A. Riegler et al., “Kvasir-seg: a segmented polyp dataset,” in *International Conference on Multimedia Modeling*, pp. 451–462, Springer, 2020.
- [22] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, “Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians,” *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.
- [23] N. Tajbakhsh, S. R. Gurudu, and J. Liang, “Automated polyp detection in colonoscopy videos using shape and context information,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, 2016.
- [24] D.-P. Fan, G.-P. Ji, T. Zhou et al., “Pranet: parallel reverse attention network for polyp segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 12266, pp. 263–273, Springer, 2020.
- [25] Z. Yan, X. Yang, and K.-T. Cheng, “Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1912–1923, 2018.
- [26] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, “Iternet: retinal image segmentation utilizing structural redundancy in vessel networks,” in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 3656–3665, Snowmass Village, Colorado, USA, 2020.
- [27] Y. Lv, H. Ma, J. Li, and S. Liu, “Attention guided u-net with atrous convolution for accurate retinal vessels segmentation,” *IEEE Access*, vol. 8, pp. 32826–32839, 2020.
- [28] X. Sun, H. Fang, Y. Yang et al., “Robust retinal vessel segmentation from a data augmentation perspective,” in *International Workshop on Ophthalmic Medical Image Analysis*, pp. 189–198, Springer, 2021.
- [29] C. Guo, M. Szemenyei, Y. Yi, W. Zhou, and H. Bian, “Residual spatial attention network for retinal vessel segmentation,” in *International Conference on Neural Information Processing*, pp. 509–519, Springer, 2020.
- [30] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, “Sa-unet: Spatial attention u-net for retinal vessel segmentation,” in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 1236–1242, Milan, Italy, 2021.
- [31] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.
- [32] C.-H. Huang, H.-Y. Wu, and Y.-L. Lin, “HarDNet-MSEG: a simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps,” <https://arxiv.org/abs/2101.07172>.