International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 5, Number 1, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v5n1.20
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Colorectal Cancer Segmentation Algorithm Based on Deep Learning

Jing Zhao, Xiaodan Wang, Chenxi He

College of Computer Science and Artificial Intelligence, Southwest Minzu University, Chengdu, 610225, China

ABSTRACT

Colorectal cancer screening is important for colorectal cancer prevention and early colorectal cancer diagnosis. To address the problems of polyp color, shape, size and blurring of edges, which are common in medical images of colorectal polyps, the UNet Colorectal Cancer Segmentation Algorithm Based on Efficient Downsampling and Joint Attention Mechanism Through CIRKD (EJ-UNet-C) is proposed. Efficient Downsampling and Joint Attention Mechanism Through CIRKD, EJ-UNet).The UNet algorithm improves the downsampling part in the encoder part, which reduces the loss of information generated in the downsampling process, and obtains the E-UNet; by incorporating the joint attention module, the By adding the joint attention module, the extraction ability of polyp edges is further improved, and EJ-UNet is obtained; at the same time, EJ-UNet is used as the basic student network, and the knowledge distillation mechanism is introduced to use Deeplabv3 as the instructor's network for guided learning. The experimental results show that the optimized network EJ-UNet-C has an IOU of 0.763, which is 8.4 percentage points higher than the IOU of 0.679 of the basic network UNet. And the distillation-optimized model has a small number of parameters and high accuracy, and the overall performance of the network is excellent, which provides a reference for the research of establishing a clinical lightweight colorectal cancer image segmentation model.

KEYWORDS

Colorectal polyp; Segmentation; Attention mechanism; Deep learning; Downsampling

1. INTRODUCTION

With the development of medical imaging technology, medical image processing plays an increasingly important role in computer-aided diagnosis. The colorectum is an important organ in the human digestive system, and its main functions include absorption of water and inorganic salts, storage of feces, excretion of feces, and participation in intestinal immunity. The health of the colorectum is critical to the functioning of the digestive system.

Colorectal cancer, the second leading cause of cancer-related deaths globally, is on the rise in terms of morbidity and mortality with the improvement of living standards and problems related to dietary habits [1]. Colorectal cancer mainly develops from polyps, so early detection and treatment can effectively control the disease. However, detection of polyps requires exhaustive observation by specialists, and the highly variable appearance of polyps present and their similarity to the intestinal texture make polyp segmentation a rather difficult task.

In order to solve the problems of uneven distribution of polyp size and location, complex boundary information and variable shapes, this paper proposes an improved colorectal cancer segmentation algorithm EJ-UNet-C based on UNet [2].The details are as follows:

1) Change the downsampling in the base network to hybrid downsampling module

That is, the Efficient hybrid downsampling biock (EHD) proposed in this paper aims to reduce the information loss caused by traditional downsampling.

2) Adoption of joint attention mechanism

A multi-scale joint attention mechanism based on Coordinate Attention [3] (MJCA), which is improved in this paper, aims to enhance the characterization of image detail features, as well as to improve the model's ability to characterize polyp edge information. as well as to enhance the model's ability to extract polyp edge information.

3) Adoption of knowledge distillation

Deeplabv3 is used as a teacher network and CIRKD (Cross-Image Relational Knowledge Distillation) [4] is used to improve the model performance without increasing the number of parameters.

2. RELATED WORK

In order to improve the accuracy and efficiency of polyp segmentation, researchers have developed various Deep Learning (DL) architectures using different techniques to solve this complex task. Tajbakhsh, N. et al. analyzed whether fine-tuning pre-trained convolutional neural networks could achieve better performance in medical image segmentation and applied it to the segmentation of polyp images in colonoscopy, verifying the effectiveness of fine-tuning in small dataset scenarios [12]. Ribeiro, E. used three convolutional layers and two pooling layers to extract the features of the image and segmented the colon polyp image by the extracted features [13]. Full Convolutional Network and U-Net are the two main approaches in image segmentation task. Akbari, M. et al [5] used Full Convolutional Network (FCN) for polyp segmentation. Sun, X. et al [6] proposed a FCN-based polyp segmentation framework in which dilated convolution is introduced to learn high-level semantic features without degrading the resolution. In addition, UNet-based approaches with an encoder-decoder architecture have shown good segmentation performance. Two variants of the UNet architecture, ResUNet++ [7] and UNet++ [8], have been developed for polyp segmentation and have achieved satisfactory performance. However, the above approaches tend to focus on preserving only geometric features in order to reduce computational cost, while neglecting some valuable semantic features. To overcome this problem, some works have introduced critical point layer (CPL) [9] or constructed content-adaptive downsampling techniques [10] to improve segmentation performance. In addition Li, H. integrated a lightweight spatial and channel attention module to adaptively refine features, which both also help to accurately segment polyps [11]. While these methods can achieve accurate segmentation results, their performance may be less robust when confronted with a variety of polyp features.

3. UNET-BASED SEGMENTATION OF COLORECTAL CANCER

The performance of UNet is very superior on most medical segmentation tasks, but there are still some shortcomings. Considering the problem of information loss in UNet maximum pooling downsampling, a combined downsampling module is planned to be obtained to carry out certain information complementation and reduce information loss through simultaneous downsampling of different branches. On the basis of the modeling study, a joint attention module is designed to improve the model's ability to deal with multi-scale polyps as well as edge preservation and extraction by introducing a multi-scale convolutional attention module. In addition, knowledge distillation is incorporated to further improve the performance of the model. The model proposed in this paper is referred to as EJ-UNet-C, and the overall structure is shown in Fig. 1.

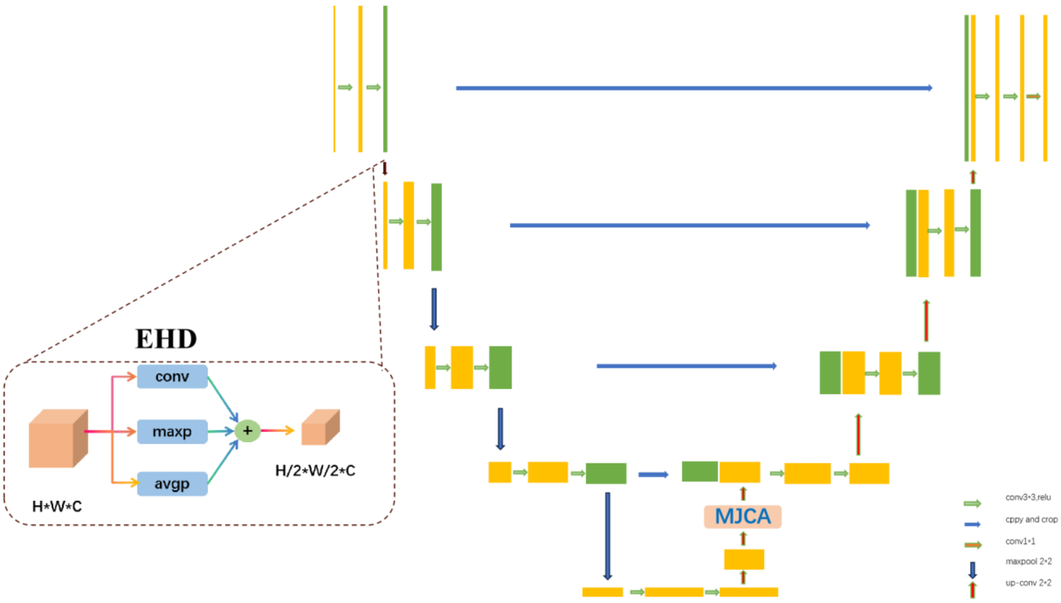

Figure 1. EJ-UNet

3.1. EJ-UNet

Maximum pooling downsampling used in UNet networks works by retaining the maximum value in the pooling window as the salient feature of the feature map, with the aim of reducing the resolution of the feature map, expanding the sensory field for subsequent feature extraction, and making the feature extraction progressively converge to the high-level semantics. However, this leads to the existence of information loss, which is unable to retain the useful information of the medical image to a great extent and transfer it to the later network, which may reduce the network's ability to perceive the local details of the image, in order to improve the situation, this paper proposes the lower Efficient Hybrid Downsampling module (EHD).

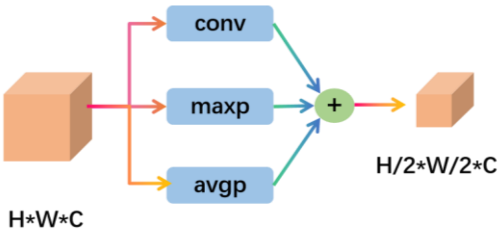

Figure 2. EHD

As shown in Figure 2, the three branches in the hybrid downsampling module all have downsampling functions, in which the first branch uses a 3*3 convolution operation with a step size of 2, the second branch uses a maximum pooling operation with a step size of 2, and the third branch uses an average pooling operation with a step size of 2, followed by summing up the feature matrices of the three branches by elements. In essence, the maximum pooling branch retains the largest features within the sampling step range, which is conducive to highlighting the saliency information of the features; the average pooling calculates the average of the pixel values within the sampling window, which to some extent retains more original information; and the convolutional downsampling branch downsamples the image for filtering by convolutional operation, which is conducive to retaining the detailed features within the step range. This design allows the three branches to retain different levels of information during downsampling, and the fusion can form a certain complementary information, thus predicting the pixel-level semantic segmentation more accurately.

Then the channel attention module is introduced after sampling on the UNet, which allows the network to focus its attention on the key features acquired, and then amplify the needed features, weaken the unwanted features, suppress the redundant features, and improve the segmentation effect of the network model.

The attention mechanism enhances the model's perceptual ability and segmentation accuracy by weighting the input features, focusing the model's attention on the critical parts of the feature map and suppressing unimportant regions, thus enhancing the model's perceptual ability and segmentation accuracy.CoordAttention breaks down the channel attention into two 1D feature encoding processes that aggregate features along different directions. The resulting feature maps are then encoded separately to form a pair of direction-aware and location-sensitive feature maps, which can be applied complementarily to the input feature map to enhance the representation of the target of interest. Since contextual information is crucial for understanding the positional information of the target, and the acquisition of contextual information requires a large receptive field, compared with standard convolution, null convolution can efficiently expand the receptive field, so four 3×3 convolutions with expansion rates of (1, 2, 5, 9) are designed. The specific structure is shown in Figure 4 below.

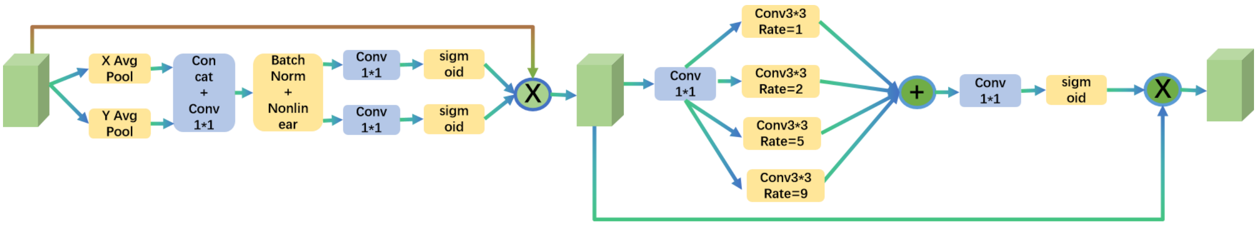

Figure 3. MJCA

3.2. EJ-UNet-C

We used the improved network EJ-UNet as the student network, chose deeplabv3 as the teacher model, and used CIRKD (Cross-Image Relational Knowledge Distillation) knowledge distillation to enhance the performance of the student model.

$$L_{CIRKD} = L_{task} + L_{kd} + \alpha L_{batch\ p2p} + \beta L_{memory\ p2p} + \gamma L_{memory\ p2r} \# \quad (1)$$

$$L_{task} = \frac{1}{H*W} \sum_{h=1}^{H} \sum_{w=1}^{W} CE(\sigma(Z_{h,w}), y_{h,w}) \# \quad (2)$$

The loss function of CIRKD consists of a total of 5 parts, i.e., L_{task} is the traditional segmentation task loss that uses cross-entropy to train each pixel and its ground-truth label. CE denotes the cross-entropy loss, $\sigma(\cdot)$ denotes the softmax function, and $y_{h,w}$ denotes the true label of the (h, w) th pixel.

$$L_{kd} = \frac{1}{H*W} \sum_{h=1}^{H} \sum_{w=1}^{W} KL(\sigma(\frac{Z_{h,w}^{s}}{T}) \Big\| \sigma(\frac{Z_{h,w}^{t}}{T})) \# \quad (3)$$

L_{kd} is Pixel-wise Class ProbabilityDistillation. Inspired by Hinton's KD [14], the class probability distribution of each pixel from student to teacher is aligned. $\sigma(\frac{Z_{h,w}^{s}}{T})$ and $\sigma(\frac{Z_{h,w}^{t}}{T})$ denote the soft class probability of the (h, w) th pixel for student and teacher, respectively. KL denotes the Kullback-Leibler scatter, and T is the temperature (T = 1).

L_{task} and L_{kd} deal only with pixel-level predictions, ignoring the semantic relationships between pixels. Inspired by memory-based comparative learning [15-18], pixel embeddings of other images are retrieved from the current small batch data or online storage. Thus adding $L_{batch\ p2p}$, $L_{memory\ p2p}$, $L_{memory\ p2r}$.

4. EXPERIMENTATION

This study is based on Kvasir-SEG [19] dataset. The experiments were performed using Pytorch, operating system Windows 11, based on NVIDIA GeForce RTX 3050 GPU. In this paper, the model was trained by setting the size of the input image to 256×256 and epoch to 100. To evaluate the performance of the segmentation algorithm, Flops, number of parameters and IOUs were used as metrics.

4.1. Comparative EXPERIMENTS

In the study, different segmentation algorithms including FCN, unet++, MultiResUNet, Pranet and other algorithms are compared based on Kvasir-SEG dataset, and the results of the comparison experiments are shown in Table 1.

FCN converts the fully connected layer of traditional convolutional neural network into a convolutional layer, which can handle images of any size. In this experiment, the Flops is 165.96GFLOPs, the number of parameters is 85.92M, and the IOU is 0.752, and the segmentation accuracy of gastrointestinal polyp images is not as good as that of EJ-UNet-C due to the simple structure. unet++ is improved based on Unet, and introduces the special connection and fusion features. The fusion feature of MultiResUNet is improved based on Unet and introduces special connection features, the Flops is 139.61GFLOPs, the number of parameters is 9.16M, and the IOU is only 0.650, which makes the segmentation accuracy lower than EJ-UNet-C due to the inaccurate distinction of the features when processing polyp image. MultiResUNet improves the performance by multi-resolution feature fusion, with the lowest number of parameters (7.25M), 75.02GFLOPs in Flops, but only 0.560 in IOU, which is the worst segmentation effect due to the limited learning ability.

In contrast, our EJ-UNet-C network has 41.73M parameters and 308.46GFLOPs, which does not have an absolute advantage in computational complexity and model size, but with the downsampling module's effective capture of multi-scale information, the MJCA attentional mechanism's precise focus on the key regions, and the optimization of knowledge distillation for the model learning process, we can achieve the best results in the segmentation of gastrointestinal polyp images. gastrointestinal polyp images, it can locate and segment the polyps more accurately, and the IOU reaches 0.782 with the best segmentation effect. This shows that our study is effective.

Table 1. Comparison with existing methods and our EJC-UNet-C on Kvasir datasets. (optimal results are shown in bold font.)

Network	Flops (GFLOPs)	Param (M)	Iou
FCN [5]	165.96	85.92	0.752
unet++ [8]	139.61	9.16	0.650
MultiResUNet [21]	75.02	7.25	0.560
Pranet [20]	23.59	29.22	0.712
EJC-UNet-C	308.46	41.73	0.782

4.2. Ablation Experiments

In order to validate the effectiveness of each component of the model, ablation experiments were conducted based on the Kvasir-SEG dataset, which was gradually improved from the original UNet structure to the EJ-UNet-C model, and four groups of experiments were set up.

The first group is based on the original UNet structure. In the second group, the downsampling module is used to replace the downsampling part of the UNet to form the E-UNet network, which can effectively capture the multi-scale image information, reduce the computation and memory consumption, and improve the perception of the polyp boundary details, and the IOU is improved from 0.679 to 0.683. In the third group, the EJ-UNet network is obtained by adding the improved

MJCA attentional mechanism on the basis of the E-UNet. MJCA allows the model to focus on the polyp region, suppresses background interference, enhances the polyp feature expression by weighting different channels and spatial locations of the features, makes the target localization more accurate, and the segmentation results more complete and continuous, and the IOU is significantly improved to 0.726. The fourth group uses knowledge distillation to obtain the EJ-UNet-C network based on the EJ-UNet. Knowledge distillation transfers the knowledge from the teacher model to the student model, and EJ-UNet-C optimizes the learning process to achieve more accurate segmentation in complex gastrointestinal polyp images, with an IOU of 0.782.

In terms of Flops, number of parameters and IOU metrics, each improvement of the algorithm positively affects the model performance, enhancing target localization, segmentation accuracy, completeness and continuity of segmentation results.

Table 2. Quantitative comparison of ablation experiments on Kvasir. (Optimal results are shown in bold font.)

Network	Flops (GFLOPs)	Param (M)	Iou
Unet	218.95	31.04	0.679
E-UNet	221.37	31.08	0.683
EJ-UNet	261.91	41.56	0.726
EJ-UNet-C	308.46	41.73	0.782

5. CONCLUSION

The EJ-UNet-C model based on semantic segmentation can effectively realize the complete semantic segmentation of polyp images by classifying images at the pixel level, which can excellently adapt to the diversity of colorectal polyp images in terms of scale and detailed features. In this study, the Kvasir-SEG dataset is used to train and test the model. The experimental results show that the algorithm successfully preserves the detailed features of the polyp region, and is highly sensitive to the segmentation of polyp edges, and can accurately segment the polyp region, which is very convenient for doctors to make accurate diagnosis. This suggests that the results of this research are very likely to provide strong support for doctors in the diagnosis and treatment of colorectal cancer.

ACKNOWLEDGMENTS

This project is supported by the Southwest Minzu University Graduate Innovative Scientific Research Project (Project No.YCYB2024067)

REFERENCES

[1] World Health Organization. (2023, July 11). Colorectal cancer. From https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

[3] Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13713-13722).

[4] Yang, C., Zhou, H., An, Z., Jiang, X., Xu, Y., & Zhang, Q. (2022). Cross-image relational knowledge distillation for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12319-12328).

[5] Akbari, M., Mohrekesh, M., Nasr-Esfahani, E., Soroushmehr, S. R., Karimi, N., Samavi, S., & Najarian, K. (2018, July). Polyp segmentation in colonoscopy images using fully convolutional network. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 69-72). IEEE.

[6] Sun, X., Zhang, P., Wang, D., Cao, Y., & Liu, B. (2019, December). Colorectal polyp segmentation by U-Net with dilation convolution. In 2019 18th IEEE international conference on machine learning and applications (ICMLA) (pp. 851-858). IEEE.

[7] Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., De Lange, T., Halvorsen, P., & Johansen, H. D. (2019, December). Resunet++: An advanced architecture for medical image segmentation. In 2019 IEEE international symposium on multimedia (ISM) (pp. 225-2255). IEEE.

[8] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging, 39(6), 1856-1867.

[9] Nezhadarya, E., Taghavi, E., Razani, R., Liu, B., & Luo, J. (2020). Adaptive hierarchical down-sampling for point cloud classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12956-12964).

[10] Marin, D., He, Z., Vajda, P., Chatterjee, P., Tsai, S., Yang, F., & Boykov, Y. (2019). Efficient segmentation: Learning downsampling near semantic boundaries. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2131-2141).

[11] Li, H., Qiu, K., Chen, L., Mei, X., Hong, L., & Tao, C. (2020). SCAttNet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images. IEEE Geoscience and Remote Sensing Letters, 18(5), 905-909.

[12] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging, 35(5), 1299-1312.

[13] Ribeiro, E., Uhl, A., & Häfner, M. (2016, June). Colonic polyp classification with convolutional neural networks. In 2016 IEEE 29th international symposium on computer-based medical systems (CBMS) (pp. 253-258). IEEE.

[14] Hinton, G. (2015). Distilling the Knowledge in a Neural Network. arXiv preprint arXiv:1503.02531.

[15] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.

[16] Tian, Y., Krishnan, D., & Isola, P. (2020). Contrastive multiview coding. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16 (pp. 776-794). Springer International Publishing.

[17] Wang, X., Zhang, H., Huang, W., & Scott, M. R. (2020). Cross-batch memory for embedding learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6388-6397).

[18] Wu, Z., Xiong, Y., Yu, S. X., & Lin, D. (2018). Unsupervised feature learning via non-parametric instance discrimination. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3733-3742).

[19] Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., De Lange, T., Johansen, D., & Johansen, H. D. (2020). Kvasir-seg: A segmented polyp dataset. In MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26 (pp. 451-462). Springer International Publishing.

[20] Fan, D. P., Ji, G. P., Zhou, T., Chen, G., Fu, H., Shen, J., & Shao, L. (2020, September). Pranet: Parallel reverse attention network for polyp segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 263-273). Cham: Springer International Publishing.

[21] Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural networks, 121, 74-87.