

Received December 23, 2019, accepted March 12, 2020, date of publication March 23, 2020, date of current version April 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982543

Colorectal Tumor Segmentation of CT Scans Based on a Convolutional Neural Network With an Attention Mechanism

YUN PEI^{ID}¹, LIN MU², YU FU², KAN HE², HONG LI^{ID}¹, SHUXU GUO¹, XIAOMING LIU¹, MINGYANG LI¹, HUIMAO ZHANG², AND XUEYAN LI^{ID}^{1,3}

¹State Key Laboratory on Integrated Optoelectronics, College of Electronic Science and Engineering, Jilin University, Changchun 130012, China

²Department of Radiology, The First Hospital of Jilin University, Changchun 130021, China

³Peng Cheng Laboratory, Shenzhen 518000, China

Corresponding authors: Huimao Zhang (huimaozhanglinda@163.com) and Xueyan Li (leexy@jlu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFE0200700, in part by the National Natural Science Foundation of China under Grant 61627820, in part by the National Natural Science Foundation of China under Grant 61934003, in part by the Natural Science Foundation of Jilin Province under Grant 20180101038JC, in part by the National Natural Science Foundation of China under Grant 81871406, in part by the Foundation of Jilin Provincial Department of Finance under Grant 2018SCZWSZX-02, in part by the Jilin Province Science and Technology Department Science and Technology Innovation Talents Cultivation Program under Grant 20180519008JH, and in part by the Foundation of Jilin Provincial Department of Finance (Establishment of standardized database for colorectal cancer and exploration of new diagnosis and treatment model based on big data analysis).

ABSTRACT Due to the irregularity of colorectal tumor contours, it is a challenging task for clinicians to segment colorectal tumors manually in CT scans. To solve this problem, a novel algorithm based on a convolutional neural network with an attention mechanism is proposed to automatically achieve tumor segmentation. The proposed network consists of three major modules: an encoder module, which is fed CT scans to attain the feature map; a dual attention module, which includes a channel attention module and a position attention module to obtain more contextual information in the deep layer of the network; and a decoder module, which restores the feature map to the original size of the input images. We used 1131 CT slices of colorectal tumors to train and test the proposed network. Compared with U-Net and CE-Net, the Dice coefficient increased by 1.46% and 0.66% respectively, for our model. The comprehensive results show that the proposed network performs more effectively in colorectal tumor segmentation than the other methods.

INDEX TERMS Colorectal tumor, image segmentation, deep learning, attention mechanism.

I. INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer in the world and is one of the malignancies with the highest incidence in Western countries. Approximately 41% of all colorectal cancers occur in the proximal colon, with approximately 22% involving the distal colon and 28% involving the rectum [1]. Approximately 20% of individuals who are diagnosed with CRC have metastatic disease on presentation. Colonoscopy remains the best choice to diagnose colorectal cancer. Prior to any treatment, CT imaging of the chest, abdomen and pelvis with contrast is needed for staging the patient's CRC. Staging is commonly performed using the TNM - classification system, which is based on the depth of tumor invasion (T), lymph node involvement (N) and

metastatic spread (M), and it is strongly associated with the 5-year survival rate [2]. For CRC patients who are diagnosed at the localized stage, the 5-year relative survival rate is approximately 90%. Once the cancer has spread outside the colon or rectum, the 5-year relative survival rate decreases dramatically [3].

CT colonography has an advantage in diagnosis, characterization (differentiation between mucinous and nonmucinous tumors), staging (depth of tumor spread, extramural vascular invasion, and the presence of malignant lymph nodes and distant metastasis), surgical planning (circumferential resection margin and sphincteric involvement in rectal cancer) [4]. CT plays an important role on determining tumor depth in colorectal cancer [5]. Lymph node metastases from colorectal cancer often present with calcifications. CT is superior for detecting calcified metastases. Three-dimensional CT to visualize the vascular anatomy facilitates laparoscopic

The associate editor coordinating the review of this manuscript and approving it for publication was Jeonghwan Gwak .

surgery [6], [7]. CT colonography provides important information for the preoperative assessment and follow up after therapy on CRC patients.

The amount of medical imaging data is increasing rapidly, so it is impossible to segment these data by hand. The precise automated segmentation of colorectal tumors from CT scans is a challenging task for several reasons. In clinical routines, segmentation of CT images of colon rectal tumors is usually undertaken by radiologists on a slice-by-slice basis, which is tedious, time consuming and very expensive, and suffers from intra- and interobserver variability [8]. Moreover, the density distributions of the colorectal tumors in the CT scans of different patients may be different, and the size and shape of the tumors also vary considerably across patients [7]. In addition, cleaning and filling the intestinal tract before a CT scan is hardly uniform. Compared to dividing parenchymal organs such as the liver, the segmentation of colorectal cancer is a challenge for doctors or other technicians. Thus, efficient and reliable methods are required.

Several studies have been performed in colorectal tumor segmentation using different deep learning modules, and the results of segmentation accuracy are acceptable [9]–[11]. However, there are some concerns that the false positive rate is high, and the segmentation edge is rough. The main reason is that the segmentation networks lose much feature information due to max pooling in the encoder module. In this paper, we adopt an attention mechanism with a position attention model and channel attention model to learn more long-range contextual information. The proposed network fuses sufficient feature information from the deep layer so that the network can concentrate on the core of the colorectal tumor.

II. RELATED WORK

An increasing number of segmentation algorithms for medical images adopt the deep learning method to accurately locate the area of interest. In reference [12], an improved convolutional neural network (CNN), called a fully convolutional network (FCN), is proposed to achieve image segmentation at the pixel level. In an FCN, the fully connected layer in a CNN is replaced by a deconvolution layer to recover the original size of the input image. Based on the FCN, Huan *et al.* [13], Wang *et al.* [14], and Zhang *et al.* [15] fused the multiple supervised mechanism to the network to improve the accuracy in segmenting the osteosarcomas and rectal tumors. Sun *et al.* [16] chose the improved FCN to segment multiphase contrast-enhanced CT images. In reference [17], U-Net consists of an encoder model and a decoder model. The encoder model extracts the feature map, and the decoder model recovers the original shape of the input image. In contrast to an FCN, U-Net applies the skip connection between the encoder model and decoder model to restore more feature information. Based on U-Net, Li *et al.* [18] converted the max pooling layers into convolution layers to reduce the information loss during downsampling. In reference [19], a context encoder network was proposed to capture high-level features in medical images. U-Net, as a classic segmentation network,

can effectively extract the feature map of input images. However, part of the information from the original input images is lost due to the pooling operation in the encoder module. A context model is adopted to solve this problem. The context model includes two blocks: a novel dense atrous convolution block (DAC) and a residual multi kernel pooling (RMP) block. The DAC block uses an atrous convolution to exchange the conventional convolution in the Inception-ResNet module. The RMP block uses four different-sized pooling kernels to reduce the scale of the parameters. The algorithm proposed in this paper outperforms other methods in optic disk segmentation and lung segmentation tasks. Recently, attention modules have gradually been used more often in the computer vision field to improve the accuracy of classification and detection. In reference [20], a position attention module and channel attention module are proposed for semantic segmentation. Convolution operations lead to the local receptive field. Therefore, the proposed attention module can capture long-range contextual information. The network uses ResNet as the baseline to extract features and adopts the dual attention module to calculate the feature map from the encoder module. The module achieves new state-of-the-art results in the Cityscapes dataset. In reference [21], a novel criss-cross attention module is proposed, which is similar to the dual attention network. They both use the attention module to capture contextual information. The criss-cross network (CCNet) achieves superior performance on the detection dataset MSCOCO. In reference [22], a novel generative adversarial network is proposed to adjust the segmentation results from U-Net. The original CT scans are fed into U-Net, and then, the probabilistic maps are labeled by a module called a label assignment generative adversarial network (LAGAN), which consists of a generator model and a discriminator model.

III. METHODS

The proposed neural network is consisted of three parts: encoder module, dual attention module and decoder module as shown in Fig. 1. We adopt modified U-Net architecture as the baseline to achieve extracting feature map and recovering the original size of input image. At the same time, we integrate a dual attention module into the baseline.

A. ENCODER MODULE

In the U-Net architecture, there are four convolutional blocks in the encoder part. Each block includes a convolutional layer, a batch normalization layer to avoid the vanishing gradient problem and a ReLU layer to improve the performance of the network. The convolutional layer uses a 3×3 convolution core in 1 step to convolute the image, from which the output is zero padded. Following the convolutional block, the max pooling layer, which uses a 2×2 pooling core in 2 steps, is used to reduce the scale of the feature parameters. However, downsampling causes the information of the input image to be lost. Thus, we deleted the fourth max pooling layer to retain more feature information to accurately segment the profile of the lesion.

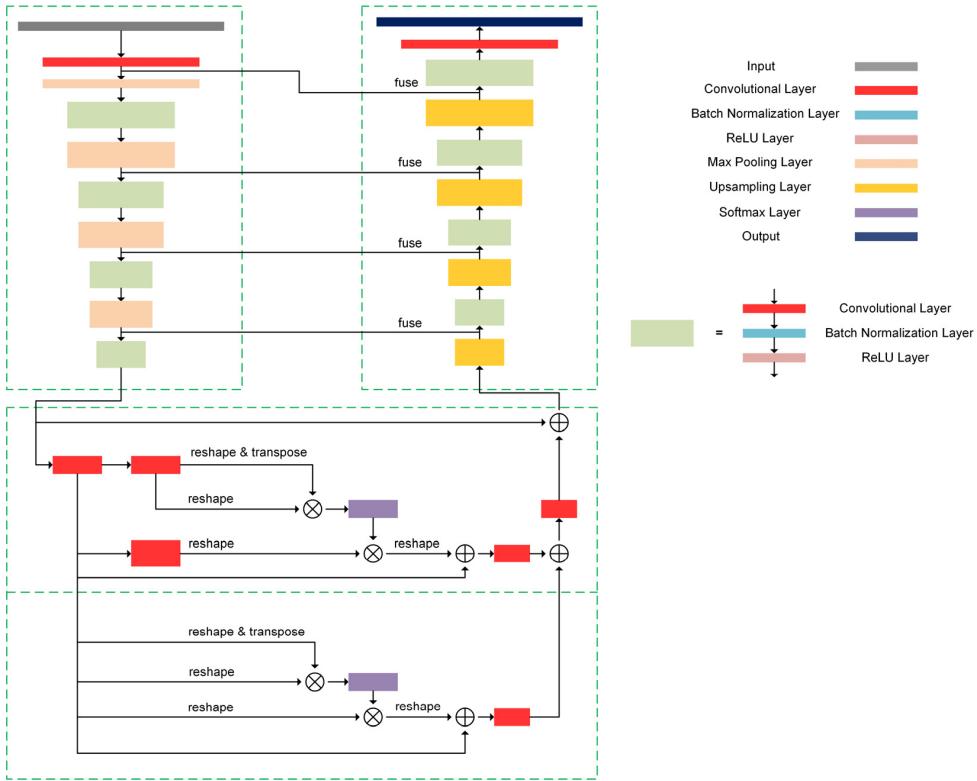


FIGURE 1. Proposed network structure.

B. DUAL ATTENTION MODULE

The feature map in the top layer is enriched with the abstract semantic information. We propose a structure that applies the shortcut connection in the dual attention module, which comprises two attention modules, to generate long-range contextual information simultaneously from pixel positions and feature map channels.

The position attention module, as shown in Fig. 2(a), given the output feature map from the encoder module, the local feature map is fed into a 3×3 convolution core to obtain $A \in \Gamma^{C \times H \times W}$, and then we adopt three branches to process A . In the first branch, A is fed into 1×1 convolution core to reduce the number of channels by a quarter. Then, we reshape it to $\Gamma^{C \times N}$, where $N = H \times W$, and transpose it to obtain $B \in \Gamma^{N \times C}$. Similarly, in the second branch, we feed A into 1×1 convolution core and reshape it to obtain $Q \in \Gamma^{C \times N}$, where $N = H \times W$. Then, we perform matrix multiplication using B and Q to obtain $D \in \Gamma^{N \times N}$:

$$D_{ab} = \sum_{i=1}^C B_{ai} \times Q_{ib}, \quad a, b = 1, 2, 3 \dots N \quad (1)$$

The output result is calculated by the softmax function to generate $D' \in \Gamma^{N \times N}$:

$$D'_{ab} = \frac{e^{D_{ab}}}{\sum_{b=1}^N e^{D_{ab}}} \quad (2)$$

In the third branch, A is reshaped into $A' \in \Gamma^{C \times N}$, and then matrix multiplication is performed again, this time with

the transpose of D' :

$$P_{ab} = \alpha \sum_{i=1}^N A'_{ai} D'_{bi} + U_{ab}, \quad a = 1, 2 \dots C, b = 1, 2 \dots N \quad (3)$$

Finally, we use the parameter α to adjust the position attention feature and connect it to the original feature map from the encoder module.

In the channel attention module, as shown in Fig. 2(b), we reshape U' into $\Gamma^{C \times N}$, and transpose it to obtain $E \in \Gamma^{N \times C}$, and then directly reshape U' into $F, F' \in \Gamma^{C \times N}$. Then we perform a matrix multiplication using F and E to obtain G :

$$G_{ab} = \sum_{i=1}^N F_{ai} \times E_{ib}, \quad a, b = 1, 2, 3 \dots C \quad (4)$$

We apply the softmax function to calculate the G :

$$G'_{ab} = \frac{e^{G_{ab}}}{\sum_{b=1}^C e^{G_{ab}}} \quad (5)$$

After that, we matrix multiply the transpose of G' and F' . We choose β to adjust the channel attention feature and connect it to U' .

$$P'_{ab} = \beta \sum_{i=1}^C G'_{ia} F'_{ib} + U'_{ab}, \quad a = 1, 2 \dots C, b = 1, 2 \dots N \quad (6)$$

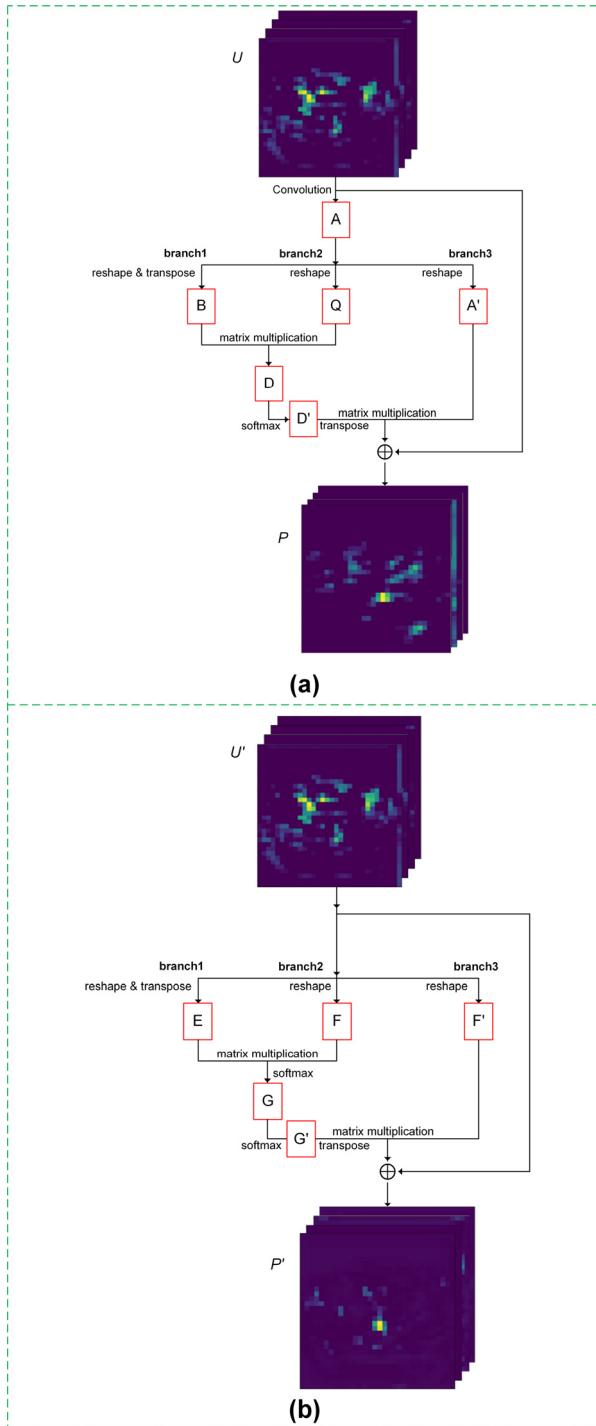


FIGURE 2. (a): Structure of the position attention module structure. (b): Structure of the channel attention module.

We propose a shortcut connection in the dual attention module. The position attention module, channel attention module and local feature map are combined to obtain more abundant semantic information. The position attention feature and the channel attention feature are fed into the 1×1 convolution core and are connected with the local feature map from the encoder output.

C. DECODER MODULE

In the decoder module, the proposed network uses four upsampling blocks to recover the size of the input image. The skip connection between the encoder module and the decoder module can effectively restore details lost due to pooling and convolution. In the proposed network, we choose an upsampling operation to increase the size of the feature map with linear interpolation. Each upsampling block includes an upsampling layer, a convolution layer with a 3×3 convolution core, a batch normalization layer and a ReLU layer.

D. LOSS FUNCTION

Segmenting medical images is considered a pixel-classification problem. The pixels can be divided into foreground pixels, which consist of the tumor, and the background, which belongs to other tissues. For traditional classification problems, the cross entropy is the most common loss function used to train models. However, when the number of positive objects and negative objects is imbalanced, the cross entropy cannot guide the network to learn the tumor features. In the image segmentation task, we focus on the similarity between the ground truth label and the predicted result. Therefore, we choose the Dice coefficient loss function, which can describe the overlap between two images to adjust the parameters in the backpropagation period.

$$\text{loss} = 1 - \frac{2 \times |P \cap G|}{|P| + |G|} \quad (7)$$

The Dice coefficient loss can guide our proposed network to predict the pixels correctly.

IV. EXPERIMENTS

A. EXPERIMENTAL DATA

We selected 158 patients (109 males and 49 females, age from 28-year-old to 84-year-old) from the First Hospital of Jilin University. These patients who were received from January 2016 to October 2018, were diagnosed with colorectal cancer by pathology and all the CT scans were taken before the patients were treated. The CT scans were produced under the equipment from three companies (Philips, GE Medical System and Siemens) and all the data were Dicom format. The scanning images were reconstructed in a 512×512 pixels matrix, and the average voxel size is $0.779 \times 0.779 \times 5\text{mm}^3$.

The ground truth was marked by two experienced radiologists (with 6 years' experience and 8 years' experience). They separately marked region of interest (ROI) - colorectal tumor areas as the label. The chief physician of Department of Radiology (with 25 years' experience) checked the results and confirmed the ground truth.

The whole dataset includes 1131 images. The dataset is divided into training dataset and testing dataset. The amount of images which are used to train network is 904 and the amount of images which are used to test network is 227.

B. EXPERIMENTAL SETTINGS

The network is trained and tested on a Windows 10 system with NVIDIA GeForce RTX 2080 graphics cards, which has

16 GB memory. The processor is an Intel (R) Core (TM) i7-9700K CPU. The public PyTorch 0.4.1 (<https://pytorch.org/>) platform is used as the basic implementation platform for the proposed network. Python 3.6 is used as the programming language. During training, we adopt Adam as the network's optimizer to guide the adjustment of the network parameters. The learning rate is set to 0.001, and the batch size is 1.

C. EVALUATION INDEXES

To evaluate the performance of our proposed algorithm, five indexes are used in this paper.

1) DICE COEFFICIENT

The Dice coefficient calculates the overlap between the ground truth and the predicted result.

$$Dice = \frac{2 \times |P \cap G|}{|P| + |G|} \quad (8)$$

where P represents the predicted result; G represents the ground truth; $|\cdot|$ represents an algorithm that measures the same area of two matrixes.

2) JACCARD INDEX

The Jaccard index is used to describe the similarity between two images and is calculated by the area common to both the ground truth and the predicted label and the area that is different from the ground truth and the predicted label.

$$J(P, G) = \frac{|P \cap G|}{|P| + |G| - |P \cap G|} \quad (9)$$

where $J(P, G)$ represents the Jaccard index for the predicted result and the ground truth.

3) RECALL

For the positive objects in the tumor area, some pixels are predicted to be negative objects. Recall is an evaluation index that is used to calculate the percentage of correctly predicted positive pixels out of all the true positive pixels.

$$R = \frac{TP}{TP + FN} \quad (10)$$

where TP represents the positive pixels that are correctly predicted; FN represents the positive pixels that are falsely predicted.

4) PRECISION

Unlike the recall, the precision is an index that describes the percentage of correctly predicted positive objects out of all the predicted positive objects. It can effectively show the prediction performance of the model.

$$P = \frac{TP}{TP + FP} \quad (11)$$

where FP represents the negative pixels that are incorrectly predicted.

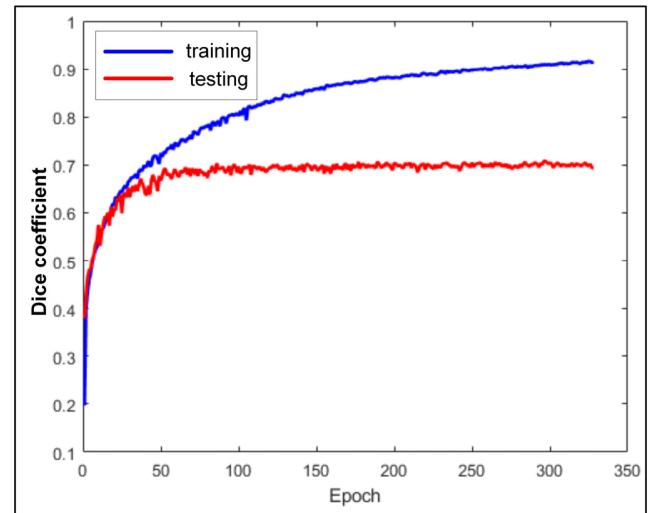


FIGURE 3. The line-charts of Dice coefficient.

TABLE 1. Comparison of segmentation results.

Model	Dice	Jaccard	Recall	Precision	F1
U-Net	0.6943	0.5622	0.7647	0.6702	0.6974
CE-Net	0.7023	0.5690	0.7511	0.6892	0.7085
Proposed	0.7089	0.5782	0.7885	0.6748	0.7089

5) F1-SCORE

The F1-score is a synthetic index that consists of the recall and precision. The recall and precision are contradictory for the same task. When the recall of a model is high, its precision may be low. Therefore, the F1 score synthetically describes the performance of a model.

$$F1 = \frac{2 \times P \times R}{P + R} \quad (12)$$

where P represents the precision; R presents the recall.

For the five evaluation indexes, the higher the index is, the better the segmentation performance is.

D. SEGMENTATION RESULTS

The curves of the proposed network are shown in Fig. 3. The Dice coefficient gradually increases and then maintains a steady state.

To verify the performance of the proposed network, we chose two image segmentation algorithms to compare with the proposed algorithm. In total, 227 images are fed into these networks to obtain the predicted results. Each predicted image is divided into two parts: positive pixels (the value of the pixel is 255) and negative pixels (the value of the pixel is 0). The overlap between the predicted image and ground truth is calculated. The five evaluation indexes are used to verify the effectiveness of the different networks on the same dataset.

As shown in Table 1, the proposed network has the best performance. Compared with U-Net, the Dice coefficient, Jaccard index, recall, precision and F1-score of the proposed network increase by 1.46%, 1.60%, 2.38%, 0.42% and 1.15%, respectively. Compared with CE-Net, the precision

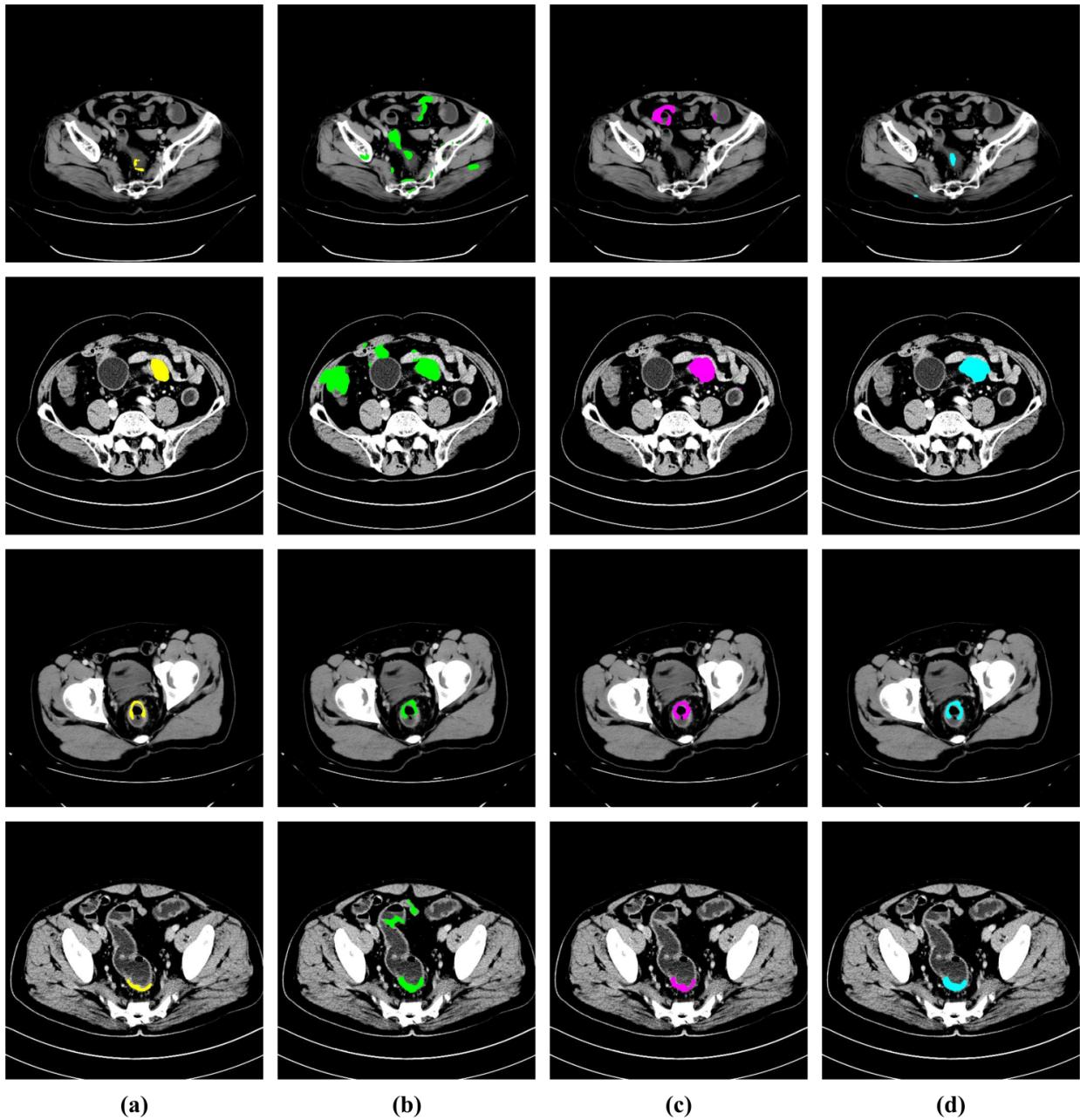


FIGURE 4. Four groups of colorectal segmentation on test dataset. (a): The ground truth of colorectal. (b): Colorectal tumor segmentation using the U-Net. (c): Colorectal tumor segmentation using the CE-Net. (d): Colorectal tumor segmentation using the proposed network.

of the proposed network is lower than that of CE-Net, but the recall of the proposed network is 2.38% higher than that of CE-Net. For the other indexes, the proposed network outperforms CE-Net.

V. DISCUSSION AND CONCLUSION

A. DISCUSSION

Because of the irregularity of colorectal tumor contour, it's difficult to segment these tumors. As shown in Fig. 4, when U-Net was used to segment the tumor, the module would predict many positive pixels wrongly. The module can't concentrate on the core of tumor because the similarity between

positive pixels and negative pixels is little. When CE-Net was used to segment the tumor, the edge contour to tumor wasn't accurate. Some detail information was ignored. The proposed network in this paper adopted the attention mechanism to reduce the false positive rate in prediction.

Compared with CE-Net and U-Net, the proposed network has the better performance because we could attain the long-range contextual information. Dual attention module is an effective attention module which can assist network to concentrate on the core of tumor. As shown in Fig. 5, input feature map from encoder module activates several parts including tumor core and other negative pixels. It causes the

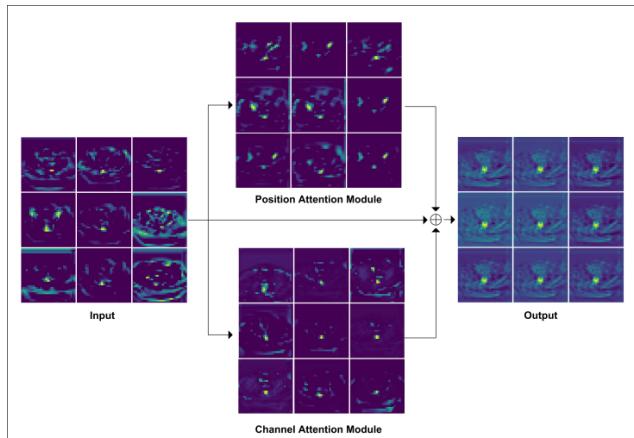


FIGURE 5. The feature map of the deep layers of the proposed network.

increase of false positive rate in the predicting result. When feature map is fed into dual attention module, the activated area is more centralized and correct. Compared with input feature map, the result from position attention module could decrease the area which doesn't belong to tumor. At the same time, the channel attention module could get the more precise tumor location. The activated area is more similarity to the tumor area in the ground truth. When the feature map from dual attention module and the original feature map from encoder module are fused, the ability that the neural network segments the colorectal tumor is obviously improved. Therefore, the dual attention module plays a significant role in extracting the feature.

B. CONCLUSION

This paper proposes a novel segmentation model based on deep learning neural network. The framework of the proposed network includes encoder module, dual attention module which is used to capture the more abstract semantic information to make the network concentrate on the core location of tumor and decoder module. The position attention module and the channel attention module in the dual attention module can measure the impact between different positions and channels. Compared with the advanced segmentation networks, the proposed network has better performance in dice coefficient, recall F1-score and Jaccard index.

ACKNOWLEDGMENT

Yun Pei and Lin Mu contributed equally to this work.

REFERENCES

- [1] L. Cheng, C. Eng, L. Z. Nieman, A. S. Kapadia, and X. L. Du, "Trends in colorectal cancer incidence by anatomic site and disease stage in the united states from 1976 to 2005," *Amer. J. Clin. Oncol.*, vol. 34, no. 6, pp. 573–580, Dec. 2011.
- [2] Y. Yuan, "Prognostic and survival analysis of 837 chinese colorectal cancer patients," *World J. Gastroenterol.*, vol. 19, no. 17, pp. 2650–2659, 2013.
- [3] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. S. Meester, A. Barzi, and A. Jemal, "Colorectal cancer statistics, 2017," *CA, Cancer J. Clinicians*, vol. 67, no. 3, pp. 177–193, May 2017.
- [4] R. Glynn-Jones, L. Wyrwicz, and E. Tiret, "Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Ann. Oncol.*, vol. 28, no. 4, pp. 22–40, Jul. 2018.
- [5] N. Hoshino, T. Sakamoto, K. Hida, and Y. Sakai, "Diagnostic accuracy of computed tomography colonography for tumor depth in colorectal cancer: A systematic review and meta-analysis," *Surgical Oncol.*, vol. 30, pp. 126–130, Sep. 2019.
- [6] S. Kijima, "Preoperative evaluation of colorectal cancer using CT colonography, MRI, and PET/CT," *World J. Gastroenterol.*, vol. 20, no. 45, pp. 16964–16975, Dec. 2014.
- [7] P. P. Mainenti, A. Stanzone, S. Guarino, V. Romeo, L. Uggia, F. Romano, G. Storto, S. Maurea, and A. Brunetti, "Colorectal cancer: Parametric evaluation of morphological, functional and molecular tomographic imaging," *World J. Gastroenterol.*, vol. 25, no. 35, pp. 5233–5256, Sep. 2019.
- [8] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, Oct. 2017.
- [9] J. Jian, F. Xiong, W. Xia, R. Zhang, J. Gu, X. Wu, X. Meng, and X. Gao, "Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images," *Australas. Phys. Eng. Sci. Med.*, vol. 41, no. 2, pp. 393–401, Jun. 2018.
- [10] J. Wang, J. Lu, G. Qin, L. Shen, Y. Sun, H. Ying, Z. Zhang, and W. Hu, "Technical note: A deep learning-based autosegmentation of rectal tumors in MR images," *Med. Phys.*, vol. 45, no. 6, pp. 2560–2564, Jun. 2018.
- [11] A. Lahiri, K. Ayush, P. K. Biswas, and P. Mitra, "Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale microscopy images: Automated vessel segmentation in retinal fundus image as test case," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 794–800.
- [12] H. Dong, G. Yang, Y. Mo, Y. Guo, and F. Liu, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *Proc. Conf. Med. Image Understand. Anal.* Edinburgh, U.K.: Springer, Jul. 2017, pp. 506–517.
- [13] L. Huang, W. Xia, B. Zhang, B. Qiu, and X. Gao, "MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images," *Comput. Methods Programs Biomed.*, vol. 143, pp. 67–74, May 2017.
- [14] M. Wang, P. Xie, Z. Ran, J. Jian, R. Zhang, W. Xia, T. Yu, C. Ni, J. Gu, X. Gao, and X. Meng, "Full convolutional network based multiple side-output fusion architecture for the segmentation of rectal tumors in magnetic resonance images: A multi-vendor study," *Med. Phys.*, vol. 46, no. 6, pp. 2659–2668, Jun. 2019.
- [15] R. Zhang, L. Huang, W. Xia, B. Zhang, B. Qiu, and X. Gao, "Multiple supervised residual network for osteosarcoma segmentation in CT images," *Computerized Med. Imag. Graph.*, vol. 63, pp. 1–8, Jan. 2018.
- [16] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artif. Intell. Med.*, vol. 83, pp. 58–66, Nov. 2017.
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Munich, Germany: Springer, Oct. 2015, pp. 234–241.
- [18] Q. Li, S. Fan, and C. Chen, "An intelligent segmentation and diagnosis method for diabetic retinopathy based on improved U-NET network," *J. Med. Syst.*, vol. 43, no. 9, p. 304, Sep. 2019.
- [19] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
- [20] J. Fu, J. Liu, and H. Tian, "DANet: Dual attention network for scene segmentation," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2019, pp. 3141–3149.
- [21] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 603–612.
- [22] X. Liu, S. Guo, H. Zhang, K. He, S. Mu, Y. Guo, and X. Li, "Accurate colorectal tumor segmentation for CT scans based on the label assignment generative adversarial network," *Med. Phys.*, vol. 46, no. 8, pp. 3532–3542, Aug. 2019.



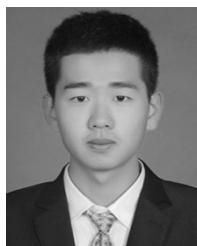
YUN PEI received the B.E. degree in electronic information engineering from Jilin University, Changchun, China, in 2019, where he is currently pursuing the M.E. degree in circuit and system. From 2017 to 2018, he worked as an Intern with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published the article on the 2018 IEEE Intelligent Vehicles Symposium. Since 2019, he has been an Intern with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China. His current research interests include the next-generation integrating driverless systems and the medical image analysis using deep learning.



SHUXU GUO received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Jilin University, Changchun, China, in 1982, 1989, and 2006, respectively. Since 2000, he has been a Professor with the State Key Laboratory on Integrated Optoelectronics, College of Electronic Science and Engineering, Jilin University, where he was involved in the reliability of semiconductor devices and the digital image processing and analysis.



XIAOMING LIU received the B.E. degree from Jilin University, Changchun, China, in 2016, where she is currently pursuing the Ph.D. degree in circuit and system. From 2019 to 2020, she won the Joint PH.D Scholarship from the Chinese Scholarship Council to visit the Hospital of Cologne University, German, for exchange. Her research interest mainly focuses on medical image analysis based on deep learning.



MINGYANG LI received the B.E. degree from Jilin University, Changchun, China, in 2017, where he is currently pursuing the M.E. degree in circuit and system. From 2013 to 2017, he mainly studied embedded hardware and software. From 2017 to 2020, he focused on medical imaging and artificial intelligence. Meantime, he has published three articles in total, which was published in *Quantitative Imaging in Medicine and Surgery*, *Physics in Medicine and Biology* as joint first authors, and the *Journal of China Clinical Medical Imaging* as the second author. His current research interests include radiomics and deep learning of medical images.



HUIMAO ZHANG received the B.S. degree in Japanese medicine from the Norman Bethune Health Science Center, Jilin University, Changchun, China, in 1994, and the M.S. and Ph.D. degrees in medical imaging and nuclear medicine from the School of Medicine, Jilin University, Changchun, in 2001 and 2006, respectively.

In 2012, she visited the Molecular Imaging Research Center, Stanford University. She has been a Chief Physician with the Department of Diagnostic Radiology, The First Hospital of Jilin University, Changchun. Her research interests mainly focus on body imaging, tumor molecular imaging diagnosis, and medical data analysis. From 2003 to 2004, she received the Ministry of Health Shikawa Medical Scholarship for visiting scholar qualification to visit the National Cancer Center in Japan for exchange.



XUEYAN LI received the B.S. degree in computer science and technology from Jilin University, Changchun, China, in 2003, the M.S. degree in computer science and technology from the Harbin Institute of Technology, Harbin, China, in 2005, and the Ph.D. degree in circuit and system from Jilin University, in 2008.

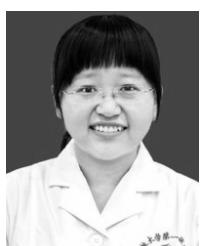
Since 2008, she has been a Lecturer with the College of Electronic Science and Engineering, Jilin University. Her research interest mainly focuses on medical image analysis based on deep learning. From 2013 to 2014, she worked as a Visiting Scholar with The State University of New York at Buffalo.



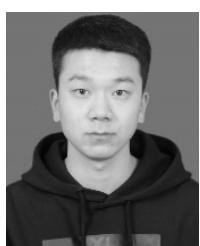
LIN MU received the B.S. degree in clinical medical course and the M.S. degree in medical imaging and nuclear medicine from the School of Medicine, Jilin University, Changchun, China, in 2010 and 2013, respectively. She currently works as an Attending Doctor with the Department of Diagnostic Radiology, The First Hospital of Jilin University, Changchun. Her research interests mainly focus on body imaging and medical data analysis.



YU FU received the B.S. and M.S. degrees from the School of Medicine, Jilin University, Changchun, China, in 2008 and 2011, respectively, where she is currently pursuing the Ph.D. degree. Since 2011, she has been an Attending Doctor with the Department of Diagnostic Radiology, The First Hospital of Jilin University, Changchun. Her research interests mainly focus on body imaging and medical data analysis.



KAN HE received the B.S. degree in clinical medical course and the M.S. degree in medical imaging and nuclear medicine from the Jilin University School of Medicine, Changchun, China, in 2007 and 2010, respectively, where she is currently pursuing the Ph.D. degree in medical imaging and nuclear medicine. Since 2010, she has been an Attending Doctor with the Department of Diagnostic Radiology, The First Hospital of Jilin University, Changchun. Her research interests mainly focus on body imaging and medical data analysis.



HONG LI is currently pursuing the bachelor's degree in electronic information engineering with Jilin University, China. His research interests mainly include image processing, computer vision, and machine learning.