IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 April 2023, accepted 9 May 2023, date of publication 15 May 2023, date of current version 19 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275966

RESEARCH ARTICLE

HarDNet and Dual-Code Attention Mechanism Based Model for Medical Images Segmentation

TONGPING SHEN[1,2], (Member, IEEE), FANGLIANG HUANG[1], (Member, IEEE), AND HUANQING XU[1,3], (Member, IEEE)

[1]School of Information Engineering, Anhui University of Chinese Medicine, Hefei 230012, China
[2]Graduate School, Angeles University Foundation, Angeles 2009, Philippines
[3]School of Electrical and Information Engineering, Tianjin University, Tianjin 300000, China

Corresponding author: Tongping Shen (shentp2010@ahtcm.edu.cn)

This work was supported in part by the Excellent Young Talents in Anhui Universities Project gxyq2022026, in part by the Anhui Province Quality Engineering Project 2021jyxm0801, and in part by the Science Foundation of Universities in Anhui Province under Grant 2022AH050428.

ABSTRACT During the formation of medical images, they are easily disturbed by factors such as acquisition devices and tissue backgrounds, causing problems such as blurred image backgrounds and difficulty in differentiation. In this paper, we combine the HarDNet module and the multi-coding attention mechanism module to optimize the two stages of encoding and decoding to improve the model segmentation performance. In the encoding stage, the HarDNet module extracts medical image feature information to improve the segmentation network operation speed. In the decoding stage, the multi-coding attention module is used to extract both the position feature information and channel feature information of the image to improve the model segmentation effect. Finally, to improve the segmentation accuracy of small targets, the use of Cross Entropy and Dice combination function is proposed as the loss function of this algorithm. The algorithm has experimented on three different types of medical datasets, Kvasir-SEG, ISIC2018, and COVID-19CT. The values of JS were 0.7189, 0.7702, 0.9895, ACC were 0.8964, 0.9491, 0.9965, SENS were 0.7634, 0.8204, 0.9976, PRE were 0.9214, 0.9504, 0.9931. The experimental results showed that the model proposed in this paper achieved excellent segmentation results in all the above evaluation indexes, which can effectively assist doctors to diagnose related diseases quickly and improve the speed of diagnosis and patients' quality of life.

INDEX TERMS Attention module, medical images, segmentation, deep learning.

I. INTRODUCTION

The medical image can reflect anatomical structures or functional tissues in the human body. Commonly used imaging techniques include CT, MRI, X-ray, etc. [1]. The task of medical image segmentation is mainly to segment medical images into several regions of similarity or difference by automatic or semi-automatic methods. The segmented images are provided to doctors for different tasks such as lesion location determination, symptom determination, tissue and organ localization, depiction of anatomical structures, and treatment planning [2]. Doctors often need to use medical image segmentation technology to facilitate detailed analysis of these areas so that the accuracy and reliability of diagnosis can be effectively improved [3].

Early medical image segmentation algorithms mainly used manually formulated rules for segmenting medical images. These methods mainly segment images based on physical features such as medical images' shape, angle, and edge structure.

Manickavasagam and Selvan proposed a gradient-driven active contour algorithm, which uses normalization and gray co-occurrence matrix to extract nodule shape, and finally uses a support vector machine algorithm to detect and classify pulmonary nodules [4]. Bruntha et al. proposed an image segmentation algorithm with edge-free active contours. By pre-

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

processing, segmenting, and detecting lung nodules in the CT image, the segmentation accuracy reached 91.5% [5]. Kurmi and Chaurasia [6] Proposed a new technique based on local features for histopathological image segmentation. Savic et al. divided the image into different partial regions and then segmented the lung nodule image using the K-means algorithm [7].

Xie et al. [8] proposed a residual network for the segmentation of liver images, which uses conditional random fields to optimize the liver edge five boundaries and remove interfering pixels to improve segmentation accuracy. Zhang et al. [9] introduced a feature fusion strategy to obtain multi-scale feature information on cancer cell location during breast cancer segmentation.

To sum up, the traditional segmentation method has a fast segmentation speed. But traditional segmentation method has great uncertainty for the segmentation results of different case images. This method of relying on manual methods for segmentation is costly and time-consuming, and the accuracy of segmentation markers cannot be guaranteed.

With the widespread use of deep learning techniques in the field of medical image segmentation, a series of research results have been achieved. The deep learning algorithm has been greatly improved in the accuracy of segmentation and the degree of automation of the algorithm. Convolutional neural network (CNN) has also been widely used in medical image segmentation tasks.

Baldeon-Calisto and Lai-Yuen [10] designed a ConvUNeXt structure for image segmentation tasks. Bilal et al. [11] proposed a neural network that automatically optimizes the segmentation network size. Bilal et al. [12] used an improved gray wolf optimization (IGWO) and CNN for the diagnosis of diabetic retinopathy. Wang et al. used the DCNN model to validate and evaluate two publicly available polyp datasets [13]. Feng et al. [14] proposed a novel contextual pyramid fusion network where the network can exploit and fuse rich contextual information.

Brandao et al. [15] proposed FCN networks to recognize and segment polyp images. Ronneberger et al. [16] proposed a U-Net network with a symmetrical encoding and decoding structure based on FCN. The network has since been widely used in the field of medical imaging. Bilal et al. [17] used the U-Net network to detect and classify diabetic retinopathy. Cao et al. [18] proposed the Swin-Unet segmentation network, which can adequately learn multi-dimensional information about images. Jiang et al. [19] combined UNet++, attention mechanism, and jump connection to improve the semantic segmentation accuracy of medical images. Belh et al. [20] extracted more feature information from breast tumor images based on the U-Net network by extending the residual convolution module and hybrid attention loss function for the image segmentation aspect.

In recent years, many researchers have added attention mechanism modules to medical image segmentation networks to improve the segmentation effect.

Valanarasu et al. [21] added a control module to the attention mechanism to improve medical image segmentation. Shen and Li [22] added an attention mechanism to a semi-supervised medical segmentation network to complete the segmentation task successfully. Li et al. [23] proposed an alternative converter-based segmentation framework, which compared to existing methods, obtains the highest segmentation accuracy by incorporating an attention mechanism and exhibits good generalization across domains.

Researchers have introduced the Transformer module to the field of medical image processing to improve the segmentation accuracy of medical images.

Zhang et al. [24] proposed a novel parallel branching for the segmentation network, combining both Transformer and CNN structures to improve the image segmentation network's ability. Gulzar et al. [25] proposed a TransUNet image segmentation network for dermatological image segmentation. Li et al. [26] applied the Transformer mechanism on the encoder and decoder respectively, focusing on capturing various global feature dimensions and long-term dependencies of the feature maps. Wu et al. [27] integrate an additional Transformer branch in the encoder, effectively capturing the remote global contextual information.

The low contrast of disease image edges and insufficient segmentation accuracy in medical image segmentation can easily cause missed diagnoses and misdiagnoses. Therefore, how constructing a larger sensory field for contextual modeling to achieve the extraction of feature information without losing spatial resolution as much as possible has been the focus of research in image segmentation.

Inspired by the above research, this paper proposes a dual-path attention network capable of extracting both spatial and location information of input features to extract rich contextual feature information of medical images. The backbone network uses the HarDNet68 module, which first extends the low-dimensional compressed representation of the input to higher dimensions and uses a multi-scale convolutional layer to extract the image feature information. The decoding part fuses the low-level information in HarDNet68 with the high-level information after the dual-path attention module, and finally up-samples the feature map to the size of the input image using the automatic learning capability of transposed convolution.

The main innovations of this paper are as follows:

(1) To reduce the total number of model parameters and improve the model running speed, this paper uses the HarDNet68 module as the backbone network to extract medical image feature information. The HarDNet68 module can improve the operation speed, the segmentation effect, and the accuracy of medical images.

(2) Extracting the low-level picture features of medical images, transferring the feature information to the position attention and the channel attention mechanism module, respectively, and finally performing concatenation operations on the tensor output from the two attention modules to obtain a feature map containing more information.

(3) To better constrain and guide the model, we sum and combine the Cross Entropy loss and Dice loss functions, and then find the mean value as the final loss to achieve accurate segmentation of medical images.

II. THE PROPOSED ARCHITECTURE

The general architecture of the medical image segmentation network proposed in this paper is similar to that of the classical segmentation U-Net network. The network is designed with an encoder-decoder structure as the basic architecture. The encoder uses HarDNet68 as the backbone network to extract image feature information and introduces a dual-code attention mechanism module to fuse the image feature information. The decoder takes the received image feature information, performs operations such as upsampling and image transposition, and finally obtains the final segmentation result by the Sigmoid function.

In the coding stage, we use the HarDNet68 network as the backbone structure for medical image feature information extraction. The HarDNet68 structure can reduce the model's parameters and improve the model training speed. The feature information is transmitted to the position attention and the channel attention mechanism module by extracting the low-level picture features of medical images through the convolution operation. The input image feature information in the location attention module is generated into an attention matrix. Vector operations are performed between the attention matrix and the initial features. Finally, element summation operations are performed between the result matrix and the initial features after the above operations. The spatial location information reflecting the long-range global context is obtained. The computational principle in the channel attention module is very similar to the mechanism of the location attention model, where the channel attention matrix is first computed in the channel dimension. Finally, the tensor output from the two attention modules is concatenated to obtain a feature map containing more information.

In the decoding stage, some of the outputs of the backbone's low-level features are fused with the high-level features that pass through the two-path attention module. A series of transpose convolution and up-sampling operations are applied to make the two types of features capable of concatenation operations. The architecture of the model proposed in this paper is shown in Fig.1.

A. HARDNET

To achieve the goal of increasing the computational speed of the network while reducing memory occupancy and power consumption, in 2019, Chao et al. [28] proposed a Harmonic DenseNet network structure (HarDNet) based on the DenseNet network. The HarDNet network structure is widely used in the fields of target recognition and image segmentation, as shown in Fig.2.

HarDNet consists of multiple Batch Normalization (BN), ReLU activation function, and convolutional structure, and the structure diagram is shown in Fig.2. Where k in the network is the initial growth rate of the l layer, the number of channels in this layer is k • m^n, n is the maximum positive integer that satisfies l is divisible by 2^n, m is the low-dimensional compression factor, and the l layer is connected to the l - 2^n layer, where l is divisible by 2^n, l - 2^n $\geqslant$ 0, and n is a non-negative integer.

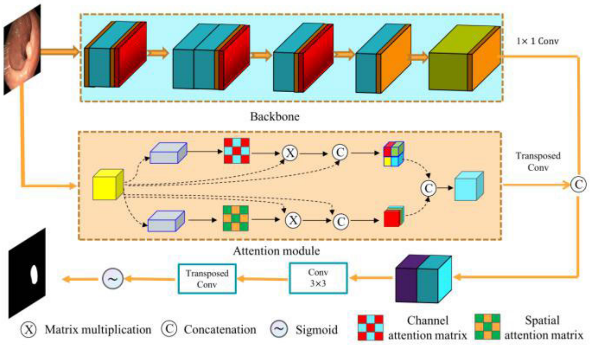


FIGURE 1. The proposed architecture.

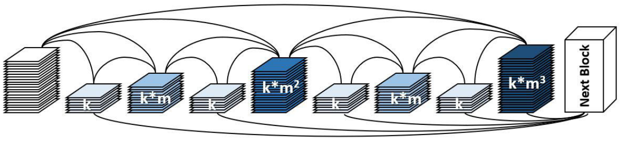


FIGURE 2. HarDNet network structure [28].

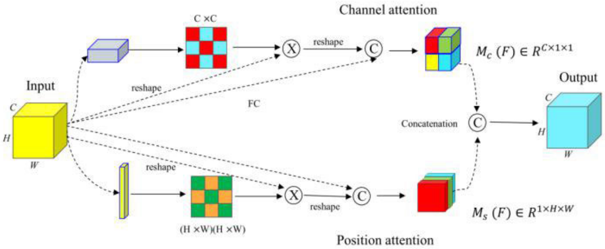


FIGURE 3. Dual-code attention mechanism network structure.

In this paper, the HarDNet68 network structure is used, and each model is combined in the Conv, BN, and ReLU order. At the same time, the global dense connection is removed, and the max pool is used for down-sampling.

B. ATTENTION MECHANISM

We know that the image pixels at different locations have different roles for the whole image, so when performing image segmentation, we assign different weight information to the input image of the model in terms of both location and channel. The structure of the integrated attention mechanism model proposed in this paper is shown in Fig. 3

1) LOCATION ATTENTION

The local feature in the encoding stage of the location attention part consists of various image pixels, and the pixel relationships at different locations are connected in a certain way. Traditional medical image segmentation networks are unable to obtain and express the relationships between the contexts of local features, so the location attention module is proposed and used to capture the spatial dependence in the feature map on each other [29]. For some special image regions, the features of the regions are updated by weighted summation, and the feature similarity between the corresponding two-pixel locations determines the weights. Therefore, the positional relationships between image pixels can contribute to each other to form the local feature information of the image and further enrich the feature information extracted in the coding stage. Suppose the local image features are represented as $A \in R^{C\times H\times W}$, and two new local features B and C are generated by the convolution operation, where $\{B, C\} \in R^{C\times H\times W}$. The newly generated local features B and C are reshaped as $R^{C\times N}$, where $N = H \times W$, represents the number of pixel points. The local features B and C are input to the softmax layer after matrix multiplication operation to get the location attention features, and the computational process is represented as follows.

$$S_{ij} = \frac{\exp(B_i \cdot C_j)}{\sum_{i=1}^{N} \exp(B_i \cdot C_j)} \tag{1}$$

where S_{ij} indicates the influence of position i in a pixel point on position j. B_i, and C_j denotes the two new feature mappings produced by the local features through the convolution layer. If the features of two-pixel points are similar, it means that there is a strong connection between the pixel points.

After the convolution operation of the local feature A, a new feature map $D \in R^{C\times H\times W}$ is generated and then reshaped as the tensor $R^{C\times N}$. After getting the new tensor, the position attention feature S and the new feature map D are transposed operation, matrix multiplication is performed, and the result of the operation is reshaped as a feature map $R^{C\times H\times W}$, and finally multiplied by the learning parameters θ_1. The obtained position attention map and the original input image feature map A are summed operations to get the final output $E \in R^{C\times H\times W}$, expressed as follows.

$$E_j = \theta_1 \sum_{i=1}^{N} \left(S_{ij} \cdot D_i\right) + A_j \tag{2}$$

The values of θ_1 are obtained by random initialization. A_j is the local feature map of the original input, and D_i is the local feature A input to the convolution layer to generate a new feature mapping. After the image segmentation network is continuously trained, more weight data will be gradually learned. The output feature information E possesses a contextual location relationship. During the network operation, pixels with similar semantic features can be aggregated into one class, and pixels with different semantic features can be aggregated into another class according to the contextual location relationship.

2) CHANNEL ATTENTION

The channel feature module allows mapping the dependencies between different channels for special meaningful feature representation. Unlike the location attention module, using the middlemost feature mapping is not necessary. It computes the channel attention map directly in the original image features and updates the feature mapping of each channel by weighted summation of each channel feature mapping. It is to compute the channel attention maps X $\in R^{C\times C}$ directly on the original image features A $\in R^{C\times H\times W}$. Firstly, the feature map A is transformed by the reshaped operation as $R^{C\times N}$, followed by matrix multiplication operation after transposing the feature maps A and A. It is input to the softmax layer to obtain the location attention feature, which is represented as follows.

$$X_{ij} = \frac{\exp(A_i \cdot A_j)}{\sum_{i=1}^{C} \exp(A_i \cdot A_j)} \tag{3}$$

X_{ij} indicates the influence of the ith layer feature map channel on the j layer feature map channel. A_i is the local feature map of the original input.

After transposing the spatial attention feature X and the original feature A, the matrix multiplication is performed. The operation's result is re-reshaped into a feature map and finally multiplied by the learning parameters. Finally, the summation operation of the elements is executed on feature map A. The result is finally output, and the process is represented as follows.

$$E_j = \theta_2 \sum_{i=1}^{C} \left(X_{ij} \cdot A_i\right) + A_j \tag{4}$$

The values of θ_2, which are obtained by random initialization, A_j is the local feature map of the original input. After continuously training the image segmentation network, it will gradually learn more weighted data. According to the above equation, the result E of image local feature output is a weighted sum operation of channel feature information and original feature information in each region, kind of making a long-term semantic dependency between the formation and image feature mapping on the channel, which can increase the recognizability in the image feature map and further improve the image segmentation accuracy and quality.

C. LOSS FUNCTION

The loss function is mainly used to measure the inconsistency between the predicted and actual values an is used to measure the performance of the network model. The smaller the loss function is, the better the performance of the model is indicated. However, the volume occupied by the lesion objects studied by medical image segmentation processing is small, and the direct use of the Cross Entropy loss function is not effective.

Therefore, in this paper, the loss function Cross Entropy [16] and the Dice [30] loss function are combined as the loss function of the model. The combined loss function combines the advantages of the two loss functions to better achieve the segmentation performance of the network. The cross-entropy loss function evaluates the loss incurred when classifying pixel points in the image segmentation process and the smaller the value, the better the segmentation model.

$$\mathcal{L}_{\text{Celoss}} = \frac{1}{N}\sum_{i=1} - \sum_{c=1}^{C} y_i \lg(p_i) \tag{5}$$

where C is the label and y_i refers to whether it is category i. If it is that category, $y_i = 1$; otherwise $y_i = 0$. p_i is the result of the model prediction.

$\mathcal{L}_{\text{Diceloss}}$ is the loss function of Dice loss, which is often used in the semantic segmentation of medical images. Dice loss is mainly used to evaluate the degree of similarity between two samples, and its larger value means that the two samples are more similar, and the value range is [0,1]. The calculation formula is shown in Eq. (6). |X∩Y| denotes the intersection of the actual medical image pixel points and the pixel points predicted by the segmentation network, and |X| and |Y| denote the actual medical image pixel points and the pixel points predicted by the segmentation network, respectively.

$$\mathcal{L}_{\text{Diceloss}} = 1 - \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{6}$$

$\mathcal{L}_{\text{Total}}$ function is a combination of the Cross Entropy loss and the Dice loss advantages, and the calculation formula is shown in Eq. (7).

$$\mathcal{L}_{\text{Total}} = \frac{\mathcal{L}_{\text{Diceloss}} + \mathcal{L}_{\text{Celoss}}}{2} \tag{7}$$

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASETS AND PREPROCESSING

In this paper, we perform validation and comparative analysis on ISIC2018, COVID-19 CT, and Kvasir-SEG datasets.

The Kvasir-SEG[1] dataset, collected and labeled by endoscopy specialists at Oslo University Hospital in Norway, contains 1,000 images of polyps and their corresponding labels, and we divided the training set, validation set and test set according to the ratio of 80%, 10%, and 10%. Images vary in size, from 332*487 to 1920*1072 and also the size of polyps that appear in the images vary in size and shape. The image size was resized to 256*256 according to the need of model input. The training samples include the original images, as well as corresponding target binary images containing cancer or non-cancer lesions. We used data augmentation on the training set data to increase the number of samples. The data enhancement method we adopted mainly consists of rotate, crop, color transform, flip, and other operations. Among, the Gaussian blurring kernel size is (4,4), the values of HSV in color transform are (0.015, 0.7, 0.4), random rotation is 30 degrees, the center cropping size is (170,170), Gaussian blurring kernel size is (3,3), and the value of flip is the left and right flip.

The ISIC2018[2] dataset is published by International Skin Imaging Collaboration in 2018, the dataset is mainly about skin disease feature segmentation, detection, and classification and contains 2594 images in total. We divide the training set, validation set and test set according to the ratio of 80%, 10%, and 10%. The original size of each image is 700*900, and the image size is resized to 256*256 according to the need of model input. The training samples include the original images, as well as corresponding target binary images containing cancer or non-cancer lesions. We used data augmentation on the training set data to increase the number of samples. The data enhancement method we adopted mainly consists of rotate, crop, color transform, flip, and other operations. Among, the Gaussian blurring kernel size is (4,4), the values of HSV in color transform are (0.015, 0.7, 0.4), random rotation is 30 degrees, the center cropping size is (170,170), Gaussian blurring kernel size is (3,3), and the value of flip is the left and right flip.

The COVID-19 CT[3] scans dataset contains CT scans of 20 patients diagnosed with COVID-19 and expert segmentation of the lungs and infections. On average, there are 162 images per category, with image sizes of 630*630, 512*512, and 401*630. The images were resized to 256*256 according to the needs of the model input. The training samples include the original images, as well as corresponding target binary images containing cancer or non-cancer lesions. We used data augmentation on the training set data to increase the number of samples. The data enhancement method we adopted mainly consists of rotate, crop, color transform, flip, and other operations. Among, the Gaussian blurring kernel size is (4,4), the values of HSV in color transform are (0.015, 0.7, 0.4), random rotation is 30 degrees, the center cropping size is (170,170), Gaussian blurring kernel size is (3,3), and the value of flip is the left and right flip. We divide the training set, validation set and test set according to the ratio of 80%, 10%, and 10%.

B. TRAINING AND MEASUREMENT METRICS

To facilitate model comparison and analysis, the same runtime platform is used for all three medical datasets, and the model training parameters are the same. The system software is Windows 10 Professional, the deep learning platform is PyTorch 1.6, and the processor is Intel i5-12400F. The batch size is set to 6 and the total number of training is 200 during the training of the network models. The Loss function is a combination of the Cross Entropy loss and the Dice loss advantages. The model optimization algorithm uses the Adam, and the learning rate is set to 1e-4.

To objectively evaluate the model segmentation effect, we used several evaluation metrics such as JS, SE, SP, AUC, F1-Score, ACC, PRE, and PRC, and the formulas of some of the evaluation metrics are shown below.

$$\text{JS} = \frac{TP}{TP + FP + FN} \tag{8}$$

$$\text{sens} = \frac{TP}{TP + FN} \tag{9}$$

[1] Kvasir-SEG: https://datasets.simula.no/kvasir-seg

[2] ISIC2018: https://www.kaggle.com/datasets/xxc025/isic2018

[3] COVID-19 CT: https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans

TABLE 1. Segmentation results of different network structures on the Kvasir-SEG.

Model	AUC	PRC	JS	ACC	SENS	SPE	PRE	Params
Unet [16]	0.8009	0.7735	0.6159	0.8567	0.6726	0.8223	0.7976	19M
PraNet [31]	0.8233	0.8231	0.6783	0.8823	0.7145	0.8675	0.8265	32M
Unet++ [32]	0.8365	0.8305	0.6891	0.8892	0.7156	0.8605	0.8743	**26M**
Attention-Unet [33]	0.8233	0.8267	0.6790	0.8843	0.7145	0.8634	0.8673	45M
UACANet [34]	0.8134	0.8188	0.6509	0.8719	0.6834	0.8603	0.8887	69M
Ours	**0.8469**	**0.8561**	**0.7189**	**0.8946**	**0.7634**	**0.8656**	**0.9214**	33M

TABLE 2. Segmentation results of different network structures on the ISIC2018.

MODEL	AUC	PRC	JS	ACC	SENS	SPE	PRE	PARAMS
Unet [16]	0.8557	0.8235	0.6648	0.9085	0.7262	0.9452	0.8115	19M
PraNet [31]	0.8813	0.8739	0.7394	0.9399	0.7703	0.9799	0.9408	32M
Unet++ [32]	0.8770	0.8807	0.7324	0.9352	0.7690	0.9850	0.9391	**26M**
Attention-Unet [33]	0.8741	0.8750	0.7239	0.9327	0.7652	0.9829	0.9306	45M
UACANET [34]	0.8902	0.8798	0.7376	0.9338	0.7907	0.9881	0.9499	69M
Ours	**0.9003**	**0.8979**	**0.7702**	**0.9491**	**0.8204**	**0.9891**	**0.9504**	33M

TABLE 3. Segmentation results of different network structures on the COVID-19 CT.

MODEL	AUC	PRC	JS	ACC	SENS	SPE	PRE	PARAMS
Unet [16]	0.9678	0.9237	0.8478	0.9578	0.9669	0.9487	0.8574	19M
PraNet [31]	0.9876	0.9878	0.9799	0.9939	0.9889	0.9922	0.9834	32M
Unet++ [32]	0.9913	0.9884	0.9740	0.9937	0.9866	0.9960	0.9871	**26M**
Attention-Unet [33]	0.9905	0.9891	0.9747	0.9939	0.9900	**0.9970**	0.9904	45M
UACANET [34]	0.9869	0.9894	0.9654	0.9943	0.9954	0.9935	0.9899	69M
Ours	**0.9954**	**0.9973**	**0.9895**	**0.9965**	**0.9976**	0.9963	**0.9931**	33M

$$\text{spe} = \frac{TN}{TN + FP} \tag{10}$$

where TP is the image pixel that is correctly segmented, TN is the image pixel that is incorrectly segmented, FP is the background pixel that is incorrectly segmented as an image pixel, and FN is the image pixel that is incorrectly segmented as a background pixel.

C. ANALYSIS OF EXPERIMENTAL RESULTS

1) COMPARATIVE ANALYSIS OF DIFFERENT ALGORITHM MODELS

The segmentation model proposed in this paper was validated on three different types of medical datasets, Kvasir-SEG, ISIC2018 and COVID-19 CT, and the results were compared and analyzed with the results of Unet, PraNet, UNet++, Attention-Unet, PraNet, and UACANet networks. The results of the analysis are shown in Tab.1 to Tab.3.

From Tab.1, for the Kvasir-SEG dataset, the AUC, PRC, JS, ACC, SENS, SPE, and PRE of the U-Net network are 0.8009, 0.7735, 0.6159, 0.8567, 0.6726, 0.8223 and 0.7276 respectively. Compared with U-Net, our model proposed in this paper increased by 4.6%, 8.26%, 10.3%, 3.79%, 9.08%, 4.33%, and 12.38% on AUC, PRC, JS, ACC, SENS, SPE and PRE respectively.

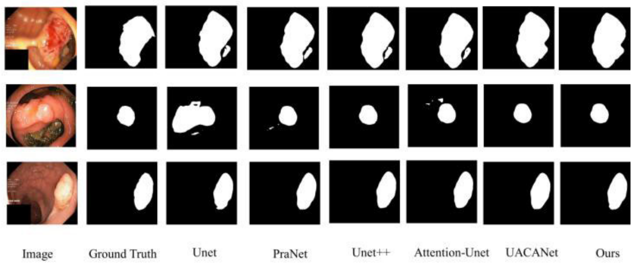


FIGURE 4. Model segmentation results in the Kvasir-SEG dataset.

From Tab.2, for the ISIC2018 data set, the AUC, PRC, JS, ACC, SENS, SPE, and PRE of our model are 0.9003, 0.8979, 0.7702, 0.9491, 0.8204, 0.9891 and 0.9504 respectively. Compared with U-Net, our model proposed in this paper increased by 4.46%, 7.44%, 10.54%, 4.06%, 9.42%, 4.39%, and 13.89% on AUC, PRC, JS, ACC, SENS, SPE and PRE respectively.

From Tab.3, for the COVID-19 CT data set, the AUC, PRC, JS, ACC, SENS, SPE, and PRE of the U-Net network are 0.9678, 0.9237, 0.8478, 0.9578, 0.9669, 0.9487 and 0.8574 respectively. Compared with U-Net, our model proposed in this paper increased by 2.76%, 7.36%, 14.17%, 3.87%, 3.07%, 4.76%, and 13.57% on AUC, PRC, JS, ACC, SENS, SPE, and PRE respectively.

Our model proposed has achieved excellent segmentation results in several evaluation indexes on the three datasets.

Segmentation of objects of interest from medical images, such as disease parts from human tissues, and quantitative measurement and analysis by relevant evaluation metrics to help doctors with diagnosis and treatment.

The segmentation experiments are carried out on Kvasir-SEG, ISIC2018, and COVID-19 CT datasets and compared with the Unet, PraNet, UNet++, Attention-Unet, PraNet, and UACANet network. The paper runs the six comparison networks in the same experimental environment, and their visual effects are shown in Fig.4 to Fig.6.

Fig.4 shows the segmentation effect of various networks on the Kvasir-SEG dataset. The third and eighth columns are the resulting diagram of six network segmentations. From Figure 4, the proposed algorithm can completely distinguish lesion regions with blurred boundaries, while other algorithms have some omissions for segmentation targets with blurred boundaries. From the visualization results, it can be concluded that it can well overcome the problem of similar color polyps and backgrounds, detect polyp tissue with different shapes, sizes, and colors, and divide the region and boundary more clearly and accurately without missing phenomena.

Fig.5 shows the segmentation effect of various networks on the ISIC2018 data set. The third and eighth columns are the resulting diagram of six network segmentations. From Figure 5, several networks can segment the edge information

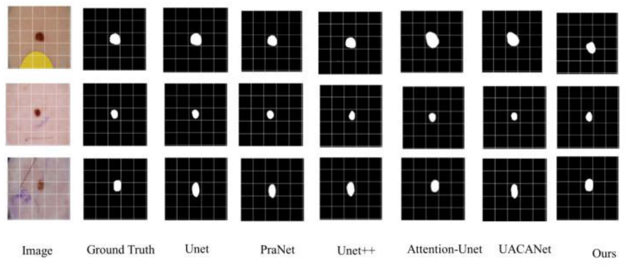


FIGURE 5. Model segmentation results in the ISIC2018 dataset.

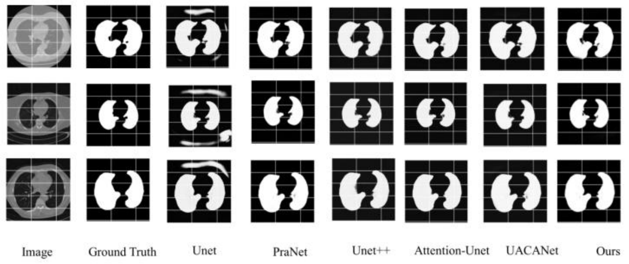


FIGURE 6. Model segmentation results in the COVID-19 CT dataset.

of skin cancer images, but our proposed model is better than other networks in dealing with the edge part. The segmentation boundary is clearer, the structure is relatively complete, and it achieved the best segmentation performance.

Fig.6 shows the segmentation effect of various networks on the COVID-19 CT dataset. The third and eighth columns are the resulting diagram of six network segmentations. From Figure 6, the U-Net has learned too many redundant features, and there are always obvious noise points; several other networks also have good segmentation performance on the segmentation boundary, but it pays too much attention to the image boundary, thus ignoring the internal features of the image. However, our model proposed in this paper retains more image details, and the segmentation results are consistent with the standard segmented images.

We also compare the accuracy and loss of the different models on the three datasets, as shown in Fig.7 a to Fig.7 c.

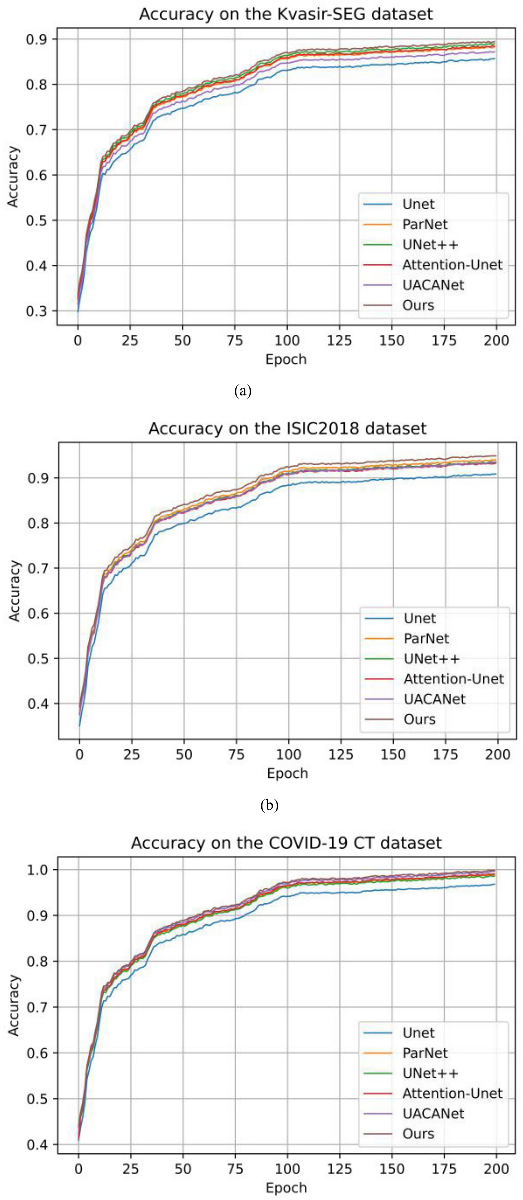
The comparative analysis of the three figures shows that the model proposed in our paper converges fast on the three datasets. Finally, the model is almost converged and achieved a high accuracy rate.

2) ABLATION EXPERIMENTS

To explore the effect of the HarDNet module and Attention module on segmentation performance in this paper, ablation experiments were conducted by controlling different modules on the COVID-19 CT dataset. The experimental results are shown in Tab. 4.

The experimental results show that both the HarDNet module and the Attention module can improve the segmentation performance to some extent, and the best segmentation results are obtained by integrating the two modules.

We also compare the proposed method in this paper with several recently proposed segmentation methods for analysis.



(a)

(b)

(c)

FIGURE 7. a. Accuracy of the Kvasir-SEG dataset, b. Accuracy of the ISIC2018 dataset, c. Accuracy of the COVID-19 CT dataset.

TABLE 4. Segmentation results of different network structures on the COVID-19 CT.

Model	AUC	PRC	JS	ACC	SENS	SPE	PRE
Add HarDNet module	0.9578	0.9669	0.9578	0.9678	0.9734	0.9509	0.9644
Add Attention module	0.9613	0.9799	0.9681	0.9773	0.9827	0.9726	0.9792
Ours	**0.9954**	**0.9973**	**0.9895**	**0.9965**	**0.9976**	**0.9963**	**0.9931**

Tab.5 shows the performance of different segmentation methods on Kvasir-SEG and ISIC2018 datasets. On the ISIC2018 dataset, the model proposed in this paper achieves more satisfactory results on three metrics, ACC, SENS, and SPE. On the Kvasir-SEG dataset, the model proposed in this paper

TABLE 5. Comparisons against existing approaches on the Kvasir-SEG and ISIC2018.

Datasets	Methods	ACC	SENS	SPE
Kvasir-SEG	GLIA-Net[35]	0.8110	0.7529	**0.9594**
	MSRF-Net[36]	-	**0.9402**	0.9310
	HarDNet-MESG[37]	**0.9691**	0.9230	0.9073
	Ours	0.8946	0.7634	0.8656
ISIC2018	DAGAN[38]	0.9324	0.9072	0.9588
	CKNet[39]	0.9492	0.9055	0.9701
	FAT-Net[27]	**0.9578**	**0.9100**	0.9699
	Ours	0.9491	0.8204	**0.9891**

failed to achieve better results on three evaluation metrics, which is an area for future improvement.

IV. DISCUSSION AND LIMITATION

Based on the classical U-Net network structure analysis, some major modifications have been made in this paper. Tab.2 to Tab.4 show the evaluation results of network models such as Unet, PraNet, UNet++, Attention-Unet, and UACANet on Kvasir-SEG, ISIC2018, and COVID-19CT datasets, respectively. Fig.4 to Fig.6 are the results of segmentation experiments on Kvasir-SEG, ISIC2018, and COVID-19CT datasets by network models such as Unet, PraNet, UNet++, Attention-Unet, and UACANet, respectively. Analyzing the above results, the proposed network has higher accuracy and fine segmentation results.

We mainly modify the original U-net network in two aspects. In the coding stage, we use the HarDNet68 network as the backbone structure for medical image feature information extraction. The HarDNet68 structure can reduce the model's parameters and improve the model training speed. The feature information is transmitted to the position and channel attention mechanism modules, respectively. In the decoding stage, some of the outputs of the backbone's low-level features are fused with the high-level features that pass through the dual-path attention mechanism module. To enable the two types of features to perform concatenation operations, transpose convolution, and up-sampling operations are applied to the image feature output information at different stages to obtain a predicted output consistent with the input size finally.

The model proposed in this paper has achieved some performance in medical image segmentation but has some limitations.

First, numerous free medical image datasets exist on the Internet, covering various parts of human tissues and organs. However, we selected only three medical image datasets for model validation and did not further expand the number of medical image datasets. Although we tried to select different types of medical datasets for validation, the variety and number of datasets were small to measure the model's generalization ability. Second, we uniformly set the image size to 256*256 during the training of the network model and did not verify the effect of different sizes of images on the performance of the segmentation network. In future research work, we will focus on these issues.

V. CONCLUSION

Our model adopts a similar U-Net backbone structure and includes two parts: encoding and decoding. In the coding stage, we use HarDNet68network as the backbone structure and use four null space convolutional pooling pyramids for medical image feature information extraction. HarDNet68 structure can greatly reduce the model's parameters and improve the model training speed. The feature information is transmitted to the position and channel attention mechanism modules, respectively. In the decoding stage, some of the outputs of the backbone's low-level features are fused with the high-level features that pass through the dual-path attention mechanism module. A series of transpose convolution and up-sampling operations are applied to enable the two types of features to perform concatenation operations. The algorithm is compared and analyzed on several medical image datasets, such as Kvasir-SEG, ISIC2018, and COVID-19CT datasets. Compared with other improved U-Net segmentation networks, the proposed network structure improves the evaluation metrics of ACC, SENS, and SPE, and achieves better segmentation results in general. We will further expand the datasets and extend the method to the 3D medical dataset and accurate segmentation of other diseases in the future.

REFERENCES

[1] J. N. Chen, Y. Y. Lu, Q. H. Yu, X. D. Luo, and E. Adeli, "Trans Ronneberger: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

[2] E. Vorontsov, A. Tang, C. Pal, and S. Kadoury, "Liver lesion segmentation informed by joint liver segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Washington, DC, USA, Apr. 2018, pp. 1332–1335.

[3] C. Oksuz, O. Urhan, and M. K. Gullu, "Ensemble-LungMaskNet: Automated lung segmentation using ensembled deep encoders," in *Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, Kocaeli, Turkey, Aug. 2021, pp. 1–8.

[4] R. Manickavasagam and S. Selvan, "GACM based segmentation method for lung nodule detection and classification of stages using CT images," in *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol. (ICIICT)*, Chennai, India, Apr. 2019, pp. 1–5.

[5] P. M. Bruntha, S. I. A. Pandian, and P. Mohan, "Active contour model (without edges) based pulmonary nodule detection in low dose CT images," in *Proc. 2nd Int. Conf. Signal Process. Commun. (ICSPC)*, Coimbatore, India, Mar. 2019, pp. 222–225.

[6] Y. Kurmi and V. Chaurasia, "Multifeature-based medical image segmentation," *IET Image Process.*, vol. 12, no. 8, pp. 1491–1498, Aug. 2018.

[7] M. Savic, Y. Ma, G. Ramponi, W. Du, and Y. Peng, "Lung nodule segmentation with a region-based fast marching method," *Sensors*, vol. 21, no. 5, p. 1908, Mar. 2021.

[8] X. Xie, W. Zhang, H. Wang, L. Li, Z. Feng, Z. Wang, Z. Wang, and X. Pan, "Dynamic adaptive residual network for liver CT image segmentation," *Comput. Electr. Eng.*, vol. 91, May 2021, Art. no. 107024.

[9] X. Zhang, X. Zhu, K. Tang, Y. Zhao, Z. Lu, and Q. Feng, "DDTNet: A dense dual-task network for tumor-infiltrating lymphocyte detection and segmentation in histopathological images of breast cancer," *Med. Image Anal.*, vol. 78, May 2022, Art. no. 102415.

[10] M. Baldeon-Calisto and S. K. Lai-Yuen, "AdaResU-Net: Multiobjective adaptive convolutional neural network for medical image segmentation," *Neurocomputing*, vol. 392, pp. 325–340, Jun. 2020.

[11] M. Bilal, G. M. Sun, S. Mazhar, and A. Imran, "Improved grey wolf optimization-based feature selection and classification using CNN for diabetic retinopathy detection," in *Evolutionary Computing and Mobile Sustainable Networks*. Bengaluru, India: Springer, 2022, pp. 1–14.

[12] A. Bilal, G. Sun, and S. Mazhar, ''Diabetic retinopathy detection using weighted filters and classification using CNN,'' in *Proc. Int. Conf. Intell. Technol. (CONIT)*, Hubli, India, Jun. 2021, pp. 1–6.

[13] L. S. Wang, Y. Q. Qian, and Y. X. Hu, ''IDDF2018-ABS-0259 segmentation of intestinal polyps via a deep learning algorithm,'' in *Proc. Int. Digestive Disease Forum (IDDF)*, Hong Kong, 2018, pp. 83–84.

[14] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, ''CPFNet: Context pyramid fusion network for medical image segmentation,'' *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.

[15] P. Brandao, E. Mazomenos, G. Ciuti, R. Caliò, F. Bianchi, A. Menciassi, P. Dario, A. Koulaouzidis, A. Arezzo, and D. Stoyanov, ''Fully convolutional neural networks for polyp segmentation in colonoscopy,'' in *Proc. SPIE*, Mar. 2017, pp. 101–107.

[16] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241.

[17] A. Bilal, G. Sun, S. Mazhar, A. Imran, and J. Latif, ''A transfer learning and U-Net-based automatic detection of diabetic retinopathy from fundus images,'' *Comput. Methods Biomechanics Biomed. Eng., Imag. Visualizat.*, vol. 10, no. 6, pp. 663–674, Nov. 2022.

[18] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, ''Swin-Unet: Unet-like pure transformer for medical image segmentation,'' 2021, *arXiv:2105.05537*.

[19] H. Jiang, T. Shi, Z. Bai, and L. Huang, ''AHCNet: An application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes,'' *IEEE Access*, vol. 7, pp. 24898–24909, 2019.

[20] K. B. Soulami, N. Kaabouch, M. N. Saidi, and A. Tamtaoui, ''Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation,'' *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102481.

[21] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and M. Vishal Patel, ''Medical transformer: Gated axial-attention for medical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Strasbourg, France, 2021, pp. 36–46.

[22] T. Shen and X. Li, ''Automatic polyp image segmentation and cancer prediction based on deep learning,'' *Frontiers Oncol.*, vol. 12, Jan. 2023, Art. no. 1087438, doi: 10.3389/fonc.2022.1087438.

[23] S. Li, X. Sui, X. Luo, X. Xu, Y. Liu, and R. Goh, ''Medical image segmentation using squeeze-and-expansion transformers,'' 2021, *arXiv:2105.09511*.

[24] Y. D. Zhang, H. Y. Liu, and Q. Hu, ''Transfuse: Fusing transformers and CNNs for medical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Strasbourg, France, 2021, pp. 14–24.

[25] Y. Gulzar and S. A. Khan, ''Skin lesion segmentation based on vision transformers and convolutional neural networks—A comparative study,'' *Appl. Sci.*, vol. 12, no. 12, p. 5990, Jun. 2022.

[26] Y. Li, J. Yang, J. Ni, A. Elazab, and J. Wu, ''TA-Net: Triple attention network for medical image segmentation,'' *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104836.

[27] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and Z. Wen, ''FAT-Net: Feature adaptive transformers for automated skin lesion segmentation,'' *Med. Image Anal.*, vol. 76, Feb. 2022, Art. no. 102327.

[28] P. Chao, C.-Y. Kao, Y. Ruan, C.-H. Huang, and Y.-L. Lin, ''HarDNet: A low memory traffic network,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 3552–3561.

[29] T. P. Shen and H. Q. Xu, ''Facial expression recognition based on multichannel attention residual network,'' *CMES-Comput. Model. Eng. Sci.*, vol. 135, no. 1, pp. 539–560, 2023.

[30] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, ''Medical image segmentation using deep learning: A survey,'' *IET Image Process.*, vol. 16, no. 5, pp. 1243–1267, Apr. 2022.

[31] D. P. Fan, G. P. Ji, T. Zhou, G. Chen, H. Z. Fu, J. Shen, and L. Shao, ''PraNet: Parallel reverse attention network for polyp segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Lima, Peru, 2020, pp. 263–273.

[32] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. M. Liang, ''UNet++: A nested U-Net architecture for medical image segmentation,'' in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain: Springer, 2018, pp. 3–11.

[33] Y. Sun, F. Bi, Y. Gao, L. Chen, and S. Feng, ''A multi-attention UNet for semantic segmentation in remote sensing images,'' *Symmetry*, vol. 14, no. 5, p. 906, Apr. 2022.

[34] T. Kim, H. Lee, and D. Kim, ''UACANet: Uncertainty augmented context attention for polyp segmentation,'' in *Proc. 29th ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2021, pp. 2167–2175.

[35] L. L. Li, X. Bian, G. L. Wang, and H. R. Wang, ''GLIA-NET: Deep learning based polyp segmentation network,'' *Comput. Eng.*, vol. 26, no. 5, pp. 1–12, 2022.

[36] A. Srivastava, D. Jha, S. Chanda, U. Pal, H. Johansen, D. Johansen, M. Riegler, S. Ali, and P. Halvorsen, ''MSRF-Net: A multi-scale residual fusion network for biomedical image segmentation,'' *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2252–2263, May 2022.

[37] 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,'' 2021, *arXiv:2101.07172*.

[38] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, Y. Xu, J. Qin, S. Chen, T. Wang, and S. Wang, ''Skin lesion segmentation via generative adversarial networks with dual discriminators,'' *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101716.

[39] Q. Jin, H. Cui, C. Sun, Z. Meng, and R. Su, ''Cascade knowledge diffusion network for skin lesion diagnosis and segmentation,'' *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106881.

TONGPING SHEN (Member, IEEE) received the B.S. degree in information management and information systems from the Anhui University of Chinese Medicine, Hefei, China, in 2007, and the M.S. degree in intelligence from Anhui University, Hefei, in 2010. He is currently pursuing the Ph.D. degree in information technology with Angeles University Foundation, Angel, Philippines.

He is currently an Associate Professor with the School of Pharmaceutical Information Engineering, Anhui University of Chinese Medicine. He mainly researches on Chinese medicine informatization, publishes more than 30 papers, authorizes four invention patents, and presides over several scientific research projects.

FANGLIANG HUANG (Member, IEEE) received the B.S. degree in instructional technology from Anqing Normal University, Anqing, China, in 2008, and the M.S. degree in computer application technology from Shanghai Ocean University, Shanghai, China, in 2011. He is currently pursuing the Ph.D. degree in computer science with National University, Manila, Philippines.

He is currently an Associate Professor with the School of Pharmaceutical Information Engineering, Anhui University of Chinese Medicine. He mainly researches on Chinese medicine informatization, publishes more than 15 papers, authorizes one invention patents, and presides over several scientific research projects.

HUANQING XU (Member, IEEE) received the Ph.D. degree from Tianjin University.

He is currently a Teacher of medical information engineering with the Anhui University of Chinese Medicine. His research interests include medical image segmentation, machine learning, and deep learning.