The Visual Computer (2022) 38:749–762
https://doi.org/10.1007/s00371-021-02075-9

ORIGINAL ARTICLE

Contour-aware semantic segmentation network with spatial attention mechanism for medical image

Zhiming Cheng[1] · Aiping Qu[1,2] · Xiaofeng He[1]

Accepted: 25 January 2021 / Published online: 22 February 2021

Abstract
Medical image segmentation is a critical and important step for developing computer-aided system in clinical situations. It remains a complicated and challenging task due to the large variety of imaging modalities and different cases. Recently, Unet has become one of the most popular deep learning frameworks because of its accurate performance in biomedical image segmentation. In this paper, we propose a contour-aware semantic segmentation network, which is an extension of Unet, for medical image segmentation. The proposed method includes a semantic branch and a detail branch. The semantic branch focuses on extracting the semantic features from shallow and deep layers; the detail branch is used to enhance the contour information implied in the shallow layers. In order to improve the representation capability of the network, a MulBlock module is designed to extract semantic information with different receptive fields. Spatial attention module (CAM) is used to adaptively suppress the redundant features. In comparison with the state-of-the-art methods, our method achieves a remarkable performance on several public medical image segmentation challenges.

Keywords Medical image segmentation · Semantic segmentation · Neural network

1 Introduction

Image segmentation is one of the main research areas in medical image analysis, which attempt to assign the labels to each pixels and address the pixel-wise lesion recognition [9]. The morphological properties such as shapes, sizes and areas of segmentation outcomes usually provide significant cues for early manifestations of many malignant diseases. The techniques such as computed tomography (CT), magnetic resonance imaging (MRI), microscopy imaging and other imaging modalities, which could provide an intuitive and effective way to scan variant diseases, have been widely utilized in daily clinical diagnosis and treatment planning [36]. Segmentation of different focused objects in these images, for example, skin lesion segmentation in dermoscopy images [15], lung segmentation in CT images [27] and colorectal cancer segmentation in endoscopy images [31], is a fundamental step to extract relevant features accurately for developing computer-aided diagnosis systems (CAD), which could assist professional clinicians by reducing the time, cost and error of manual processing in clinical situation [3].

With the rapid development and wide popularization of medical imaging technologies, a large number of medical images are collected and could be used for analysis. It is emergent to develop automatic algorithms to efficiently and objectively analyze these medical images, with the aim of providing doctors with precise interpretation of diagnosis information contained in the images to have better treatment of a large amounts of patients [3]. However, automatic and accurate segmentation of lesion (tissue or organ) in medical images remains a challenging task. First, the morphological appearance of the focused objects have large variant among different individuals even for a same disease, which will increase the difficulty of segmentation. Figure 1 illustrates three examples for lung nodule, skin lesion and colorectal polyp. Second, the difference between the focus objects and background is unclear that it complicates the segmentation . In particularly, different tissues and organs are always included in the focused areas, which make the segmentation of these confusing boundaries more difficult. Third, texture

✉ Aiping Qu
qap@usc.edu.cn

[1] School of Computer, University of South China, Hengyang 421001, China

[2] Hunan Provincial Base for Scientific and Technological Innovation Cooperation, Hengyang 421001, China

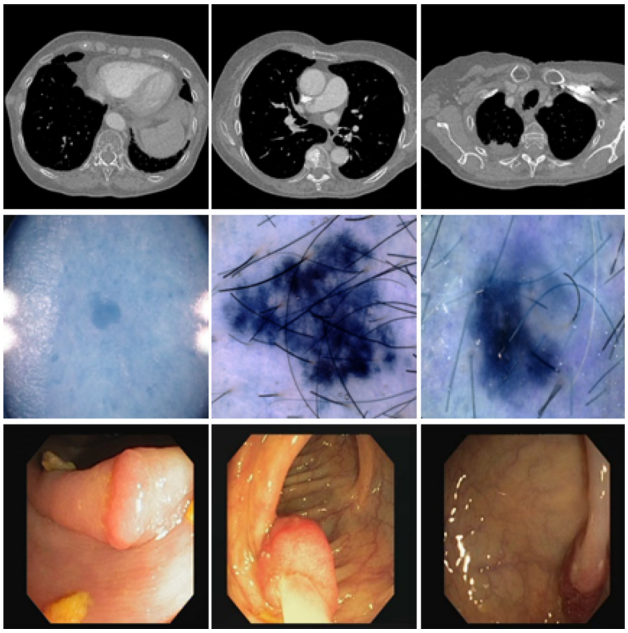


Fig. 1 Examples of several representative medical images. The first row indicates the lung nodule in CT images, the second row represents the skin lesion in the dermoscopy image, and the last row shows the colorectal polyp in endoscopy images

features, artifacts and imaging noise will also bring great challenge to segmentation.

In the past few decades, segmentation of medical images has received much attention; a large amount of accurate and automatic methods for this topic have been proposed [23,34,38]. The earlier methods are mainly based on traditional hand-crafted features [5,7,12,26,39,41]. According to the types of features, these methods can be roughly divided into three groups of: gray level based, texture based and level set atlas based. Although these methods obtained exciting performance at that time, they are unreliable in the real complex clinical situations, because they heavily depend on pre-processing, which is low robustness to image quality and artifacts. Due to the great success of deep learning (DL) in the field of computer vision, amount variants of DL methods are proposed and applied to medical image segmentation [23,33,37]. The representative Unet is the most popular selections and usually obtain good results [30]. The architecture of Unet consists of an encoding path to obtain context features and a decoding path that enables precise localization. It can be trained end to end for very few images. Although many variants of Unet have been proposed and widely used in medical image segmentation, they still suffer from inaccurate object boundaries and unsatisfactory results [1–3,25,32,35,40]. It is well known that the discriminative features heavily affect the segmentation performance. In order to accurately segment the focused objects, many researchers paid close attention to extract and aggregate the high-level context features and low-level fine details simultaneously.

In this manuscript, we propose a contour-aware semantic segmentation network for medical image segmentation, which is an extension of Unet and have two branches: semantic branch and detail branch. The semantic branch follows the classical encoding–decoding structure of Unet, which focuses on extracting semantic features from shallow and deep layers. The detail branch is designed to enhance the detailed contour information implied in the shallow layers. In addition, inspired by the densely connected convolution, we design a MultiBlock module to replace convolutional block of classical Unet in order to utilize different receptive fields. We also add a spatial attention module between the encoding and decoding path to suppress abundant features to improve the network's representation capability. The contributions of this work are summarized as follows:

- We propose a two-branch Convolutional Neural Network architecture containing a semantic branch and a detail branch, which aggregates the high-level semantic features and low-level fine details simultaneously.
- We design a MulBlock module which utilizes three paths with different receptive fields to extract semantic information from the input feature maps.
- In comparison with the state-of-the-art methods, the proposed method achieves a remarkable performance on four public medical image segmentation challenges.

The remainder of this manuscript is organized as follows: Sect. 2 reviews relevant works of medical image segmentation; Sect. 3 presents our proposed segmentation neural network; Sect. 4 presents the datasets and experimental results; Sect. 5 concludes this paper.

2 Related work

Semantic segmentation is one of the most crucial tasks in the field of medical imaging analysis. Prior literatures mainly utilize the traditional handcrafted features for semantic segmentation. With the fast development of deep learning, DL-based methods have achieved outstanding results and dominated this task. Among these methods, the convolutional neural networks (CNN) and fully convolutional network (FCN) are popular segmentation frameworks. In this section, we briefly review CNN-based and FCN-based methods for medical imaging segmentation.

2.1 CNN-based segmentation frameworks

To our knowledge, Ciresan et al. [13] first utilized a deep CNN to segment electron microscopy images. They clas-

sified each pixel of every slice through extracting a patch around the pixel by a sliding window. Because the sliding window method has plenty of overlap and redundant computation, this method is time inefficiency. Pereira et al. [29] utilized intensity normalization as pre-processing step and small kernels to improve deeper architecture of CNN, to reduce over-fitting for brain tumors segmentation in magnetic resonance images (MRI). In order to extract context information, Chen et al. [8] proposed a segmentation framework named Deeplab, which use the convolutional layers to replace all fully connected layers and atrous convolutional layers for increasing the feature resolution. Choudhury et al. [10] attempted to utilize the DeepLab network for the task of brain tumor segmentation in MRI. Based on residual learning, Li et al. [21] proposed a dense deconvolutional network for skin lesion segmentation, which combined global contextual information by dense deconvolutional layers, chained residual pooling and auxiliary supervision, to obtain multi-scale features.

2.2 FCN-based segmentation frameworks

The CNN-based methods use patch around the pixel to make a patch-wise prediction that they ignore the spatial information implied in the image when the convolutional features are fed into the fully connected (fc) layers [3]. In order to overcome this problem, Long et al. [24] proposed a fully convolutional network (FCN) which use convolutional and deconvolutional layers to replace all fc layers in CNN architecture. FCN is trained end to end and pixels to pixels for semantic segmentation. It is the most popular method utilized to segment medical and biomedical images. Christ et al. [11] proposed a liver segmentation method by cascading two FCNs, where the first FCN performs segmentation to predict the region of interest (ROI) of liver; the second FCN focuses on segmenting the liver lesions within the predicted ROIs of the first FCN. Zhou et al. [42] proposed a Focal FCN which applies the focal loss on a fully weighted FCN in medical image segmentation, with the aim of addressing the limited training data of small object by adding weights to the background and foreground loss.

Unet [30] has become one of most popular FCNs, which has been widely used in biomedical image segmentation since 2015. Unet utilizes an encoder-decoder architecture. The encoding path consists of several convolutions and pooling operation, and the decoding path utilizes up-sampling operation to restore the shape of original image and produce segmentation results. The shortcut connections between layers of equal resolution make Unet able to utilize the global location and context information at same time. In addition, it works well on limited training images [4]. Okatay et al. [28] proposed an attention Unet which employs a novel attention gate, to automatically focus on target structures of varying shapes and sizes. In order to utilize the strengths of different network. Altom et al. [1] proposed Recurrent Convolutional Neural Network (RUnet) and Recurrent Residual Convolutional Neural Network (R2U-Net) based on Unet for medical image segmentation. Azad et al. [3] proposed an extension of Unet , Bi-Directional ConvLSTM Unet with densely connected convolutions (BDCU), which combines the bi-directional ConvLSTM and dense convolution mechanism with Unet for medical image segmentation.

Recently, several researches focused on using the detail information implied in shallow feature maps to enhance the boundary information [18]. For example, Wang et al. [36] proposed a boundary-aware context neural network (BA-Net), which employs pyramid edge extraction module, mini multi-task learning module and interactive attention module, to capture context information for 2D medical image segmentation. Zhou et al. [43] first presented a nested Unet architecture named UNet++, which utilizes a series of nested and dense skip pathways to connect the encoder and decoder subnetworks. Then, they redesigned the skip connections in Unet++ by aggregating features of varying semantic scales in decoders and devised a pruning scheme to accelerate the inference speed of Unet++ [44].

3 Method

The flow chart of the proposed network is illustrated in Fig. 2. Our proposed model is an end-to-end trainable medical image segmentation network, which has two branches: semantic branch to obtain high-level semantic context information and detail branch to enhance low-level detail information. We introduce the details of our method below.

3.1 Semantic branch

The semantic branch is shown in the top part of Fig. 2. This branch has narrow channels and deep layers, with aim of capturing high-level semantic information of the image. The architecture of the semantic branch is an extension of Unet, which also uses the encoding–decoding structure. The difference is that we employ MultiBlock module and spatial attention module to replace the original convolutional filters in classical Unet. The encoding path consists of five steps. The first step contains a convolutional 1×1 layer and a MultiBlock module. Each of the next four steps is composed of a MultiBlock module. After MultiBlock module in each step, there is a 2×2 down-sampling layer for max pooling. The resolutions of the outputs of each layer in encoding path are 256×256, 128×128, 64×64, 32×32, 16×16, respectively. Then, the encoding path is followed by a spatial attention module to suppress the useless redundant features. The number of channels in each layer of the semantic branch

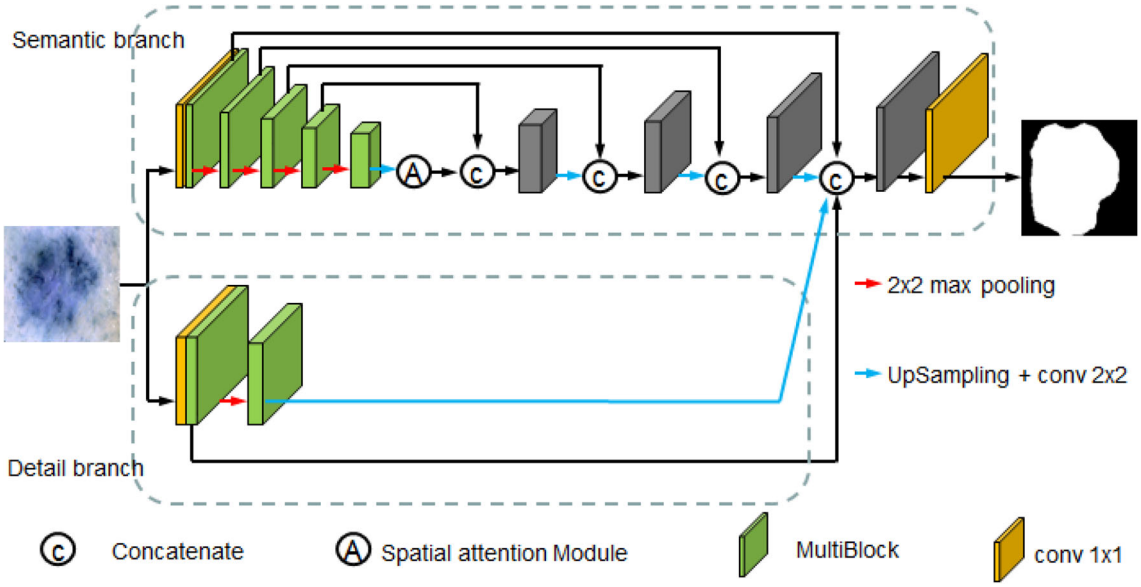


Fig. 2 Flowchart of the proposed method. The structure includes: (1) a detail branch with wide channels and shallow layers, used to capture the details of the underlying layer and generate high-resolution feature representations; (2) a semantic branch with narrow channels and deep layers used to get high-level semantic context information

Table 1 Number of channels in each layer of the semantic branch

Layer	Channels
Conv 1 × 1	32
MultiBlock1	64
MaxPooling1	64
MultiBlock2	128
MaxPooling2	128
MultiBlock3	256
MaxPooling3	256
MultiBlock4	512
MaxPooling4	512
MultiBlock5	1024
Spatial_attention	1024

is listed in Table 1. The decoding path has four steps. Each step concatenates the layer obtained by performing an up-sampling function over the output of the previous layer and the layer copied who has the same resolution from encoding path.

3.2 MultiBlock module

The details of MultiBlock module are illustrated in Fig. 3. Densely connected convolution [19] was proposed to mitigate the problem that the sequence of convolutional layers in original Unet might learn redundant features in the successive convolutions. It utilizes the idea of "collective knowledge" by allowing information flow

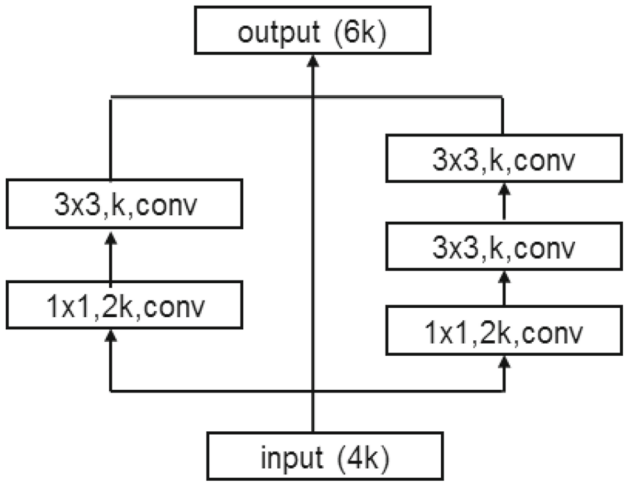


Fig. 3 MultiBlock module

through the network and reuse the feature maps. We employ the idea of densely connected convolution and propose a variant named MultiBlock in our network. Different from densely connected convolution, MultiBlock module reduces the channel of the original main branch to half for cutdown the size of the model. In addition, a new path with two 3 × 3 convolutional filters is added for expanding the receptive field of the module. Let us assume the number of input channels is $4k$, as shown in Fig. 3, the channels of left path is k and the channels of right path also is k. Then, they are concatenated with input maps, and we can get the output of MultiBlock with $6k$ channels.

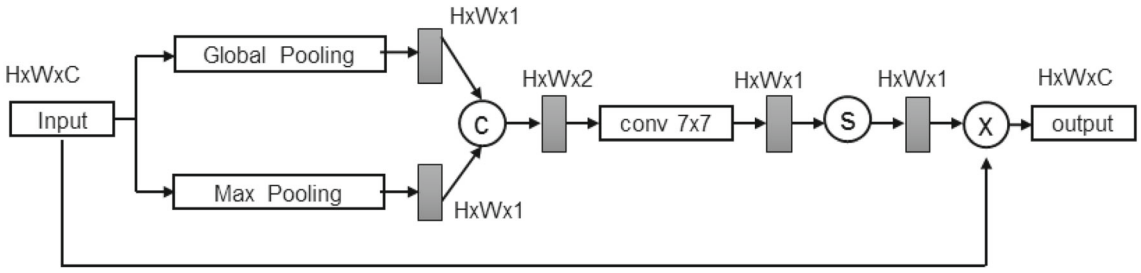


Fig. 4 Spatial attentive module. The size of each feature map is shown in $H \times W \times C$, where H, W, C indicate height, width and number of channels, respectively

3.3 Spatial attention module

Spatial attention module (SAM) was proposed to infer the attention map along the spatial dimension between features [17]. It is commonly used to perform adaptive feature refinement by multiplying the attention map with the input feature map for image segmentation [17]. As shown in Fig. 4, SAM firstly concatenates the feature maps which are obtained by performing average pooling and max pooling along the channel axis, to generate an efficient feature descriptor. Then, the concatenated feature descriptor is fed into a 7×7 convolutional layer and a sigmoid activation function to produce an attention map. Finally, the attention map is utilized to multiply the input image in order to generate the output feature map.

3.4 Detail branch

In medical images, the differences between the target objects and the background commonly are not obvious, especially that there exist amount jagged contours and tiny objects, the high-level semantic context information can improve the performance of larger structures segmentation, but it is easy to make mistakes when dealing with these boundary structures. It is well known that the shallow feature maps of a deep convolutional network contain abundant boundary information. In order to remarkably enhance the detail information, which implied in the shallow feature map, we design a small span shallow structure for detail branch. As shown in the bottom part of Fig. 2, we firstly pass the input image into a 1×1 convolutional layer followed by MultiBlock module to get the first layer feature map. After 2×2 max pooling and MultiBlock module, we get the second layer feature map. Finally, we send the first layer feature map and the feature map which is upsampled from the second layer feature map to concatenate with the semantic branch. The number of channels in each layer of the detail branch is listed in Table 2. It shows that the detail branch utilizes wide channels and shallow layers to enhance the detail information.

Table 2 Number of channels in each layer of the detail branch

Layer	Channels
Conv 1×1	64
MultiBlock1	128
MaxPooling	128
MultiBlock2	256

4 Experiments and results

4.1 Datasets

We employ four public datasets: The COVID-19 CT Segmentation dataset, CVC-ClinicDB dataset, ISIC2018 dataset and Lung segmentation dataset to verify the effectiveness of the proposed method. The four datasets are described in detail as follows:

4.1.1 The COVID-19 CT Segmentation dataset

The COVID-19 CT Segmentation dataset is the only one open-access CT segmentation dataset for the novel Coronavirus Disease 2019 (COVID-19) [16]. The dataset includes 100 axial CT images which are collected by the Italian Society of Medical and Interventional Radiology from different COVID-19 patients. The CT images are segmented by a radiologist with different labels to identify lung infections. We randomly split the dataset into a training set with 45 images, a validation set with 5 images and a testing set with the remaining 50 images. Since the dataset suffers from a small sample size, we utilize the same strategy described in [16] to augment the training dataset.

4.1.2 CVC-ClinicDB dataset

CVC-ClinicDB dataset [6] is a public fully annotated colonoscopy image dataset, which has been generated from frames of 29 different standard colonoscopy video sequences. The images in this dataset all have a polyp, and the total number of images is 612. Each image has manually annotated ground truth of polyp. The original resolution of images in

the dataset is 288 × 384. We randomly divide the dataset into three subsets: a training set with 414 images, a validation set with 85 images and a testing set with the rest 113 images.

4.1.3 ISIC 2018 dataset

The ISIC 2018 skin cancer segmentation dataset is published by the International Skin Imaging Collaboration (ISIC) and has become a major benchmark dataset to evaluate the performance of medical image algorithms [14]. The dataset consists of 2594 images with corresponding annotations of localizing lesions on skin images that containing melenoma. The original resolution of images in this dataset is 700 × 900. We use the same preprocessing strategy as [1] to process the input images. We also resize the input images to 256 × 256. We follow the same strategy described in [3] to split whole dataset into three subsets: a training set with 1815 images, a validation set with 259 images and a testing set with 520 images.

4.1.4 Lung segmentation dataset

Lung segmentation dataset is released at the Kaggle Data Science Bowl, with aim of developing algorithms that accurately determine when lesions in the lungs are cancerous for the Lung Nodule Analysis competition in 2017 [22]. The dataset contains 2D and 3D CT images with labels annotated by radiologists for lung segmentation. The resolution of each image in this dataset is 512 × 512. Since the CT image consists of not only the lung but also other tissues, it is worth to extract the mask of lung and ignore all other tissues. We use same preprocessing strategy described in [3] to extract the surrounding regions, and obtain the lung region which inside the surrounding regions. We train and test the model with extracted lung regions. We randomly split the images into three subsets: a training set with 571 images, a validation set with 143 images and a test set with 307 images.

4.2 Implementation details

We implement our method in python with Keras. We train and evaluate our method on a high performance computer with 35.4816 Tflops CPU and 18.8 Tflops GPU. The standard binary cross-entropy loss is used as training loss. Adam algorithm with initial learning rate of le-4 is used to optimize the model weights. We set the batch size of training with 50, 50, 100 and 50 for four datasets, respectively. We stop the training process when the validation loss does not decrease in 10 consecutive epochs.

4.3 Evaluation metrics

We utilize several common metrics to measure the experiment performance comparisons, such as F1 score (F-measure), accuracy, the area under ROC curve (AUC), sensitivity and specificity. We also use Frame Per Second (FPS), which is the number of images that can be processed per second, to measure the inference speed.

4.4 Ablation study

To verify the impacts of the detail branch, we compare the quantitative results of the proposed method with and without this mechanism. The detailed results are exhibited in Table 3. In Table 3, Ours_noDetail and Ours indicate the methods without and with the detail branch, respectively. From the results, we could see that the detail branch mechanism provides significant improvement for the COVID-19 and CVC-ClinicDB datasets, and slight improvement for the ISIC and Lung datasets.

In order to verify the proposed detail branch can be extended to other available networks except only limited with the semantic branch for contour-aware segmentation, we extend the detail branch to the Unet [30]. The detailed results are listed in Table 4. In Table 4, Unet_Detail means the method that combines the detail branch into the standard Unet architecture. From the results, we could see that the detail branch mechanism also has a significant improvement for the COVID-19 and CVC-ClinicDB datasets.

From the Ablation studies, we could see that the proposed detail branch could improve the segmentation results especial for the COVID-19 and CVC-ClinicDB datasets. Observing the images in the COVID-19 and CVC-CLinicDB datasets, we find the segmented objects have a high similar with the background. Several literatures consider these two segmentation tasks belong to camouflaged object detection. We attempt to explain this result, but it is difficult. We will verify whether this mechanism has good performance for such problems in our future work.

4.5 Results

In recent years, a large number of state-of-the-art algorithms focused on medical image segmentation are reported, such as Unet [30], Attention Unet [28], R2U-Net [1], BCDU [3], DeepLabv3+ [8] and Unet++ [44]. The experiment performance comparisons of above algorithms with our proposed method are presented in this section. For fairness, the codes of the comparison methods are all download from original websites. It should be noted out that for all four datasets, we use exactly the same network architecture of our method.

Firstly, we test the proposed method on the COVID-19 CT Segmentation dataset. The quantitative results are shown

Table 3 Effects of the detail branch with the semantic branch on four different datasets

Dataset	Method	F1-score	Sensitivity	Specificity	Accuracy	AUC
COVID-19	Ours_noDetail	70.54	62.84	96.86	91.08	79.85
	Ours	75.16	71.48	96.16	91.97	83.82
CVC-ClinicDB	Ours_noDetail	75.23	76.33	97.40	95.52	86.86
	Ours	79.98	77.70	98.38	96.53	88.04
ISIC	Ours_noDetail	85.72	85.47	94.46	91.90	89.96
	Ours	86.27	86.28	94.54	92.19	90.41
Lung	Ours_noDetail	98.50	98.66	99.66	99.48	99.16
	Ours	98.68	98.89	99.68	99.54	99.29

Table 4 Effects of the detail branch extended to Unet on four different datasets

Dataset	Method	F1-Score	Sensitivity	Specificity	Accuracy	AUC
COVID-19	Unet	55.85	44.54	96.94	88.03	70.74
	Unet_Detail	62.85	54.37	96.18	89.08	75.27
CVC-ClinicDB	Unet	71.24	63.03	98.63	95.46	80.83
	Unet_Detail	75.23	66.89	98.92	96.07	82.91
ISIC	Unet	85.07	80.65	96.44	91.95	88.54
	Unet_Detail	86.00	79.08	98.08	92.67	88.58
Lung	Unet	98.45	98.59	99.64	99.46	99.12
	Unet_Detail	98.59	99.11	99.59	99.51	99.35

in detail in Table 5. From the results, we could see that the proposed method achieves an exciting performance with 75.16 F1-Score, 91.97 Accuracy and 83.82 AUC, which are significant higher than other methods. To demonstrate the segmentation, we also display the detailed results for 7 random selected images in Fig. 5, which indicates the segmentation results are close to the ground truth with more clear boundaries.

Secondly, we test the proposed method on CVC-ClinicDB dataset. The quantitative results obtained by different methods and the proposed network are listed in Table 6. Table 6 shows that the proposed method achieves the best results except the metrics specificity. It obtains the highest F1-score, sensitivity, accuracy, AUC of 79.98, 77.70, 96.53 and 88.04, respectively, and specificity comparable. We also display the detailed results for 6 random selected images in Fig. 6. From Fig. 6, we can see that Unet has the problem of less segmentation and its produced contours are not smooth, Attention Unet is easy to produce over segmentation, and BCDU fails to segment targets with unclear contours. Our method also shows more precise and fine segmentation output than other methods.

Thirdly, we test the proposed method on ISIC2018 dataset. Table 7 shows the quantitative results achieved by different methods and the proposed network on ISIC dataset. As shown in Table 7, Unet++ achieves the best performance except the evaluation metric specificity. Our method obtains the F1-score, sensitivity, accuracy, AUC of 86.27, 86.28, 92.19, 90.41, respectively, which are slighter lower than Unet++. For clearly displaying the detailed results, we random select 6 images from the testing set and display the segmentation results in Fig. 7. From Fig. 7, we could see that images from ISIC2018 dataset exist unclear contours and heavy noisy information from background. Unet, Attention Unet and BCDU fail to well segment the contours of target, and R2U-Net produces serious over segmentation. Although the metrics of our method is lower than Unet++, the segmentation boundaries are more close to the ground truth than Unet++.

Finally, we test our method on the Lung segmentation dataset. The qualitative results of lung segmentation on testing set are reported in Table 8. From the results presented in Table 8, we can see that the proposed method achieves the best performance for most of evaluation metrics. It obtains the highest F1-score, specificity, Accuracy of 98.68, 99.68, 99.54, respectively, sensitivity and accuracy comparable. To demonstrate segmentation clearly, we also display the detailed results for 6 random selected images in Fig. 8. The results illustrate that Unet is not good at recognizing small target objects and controlling the overall shape of target objects. Attention Unet is prone to produce over segmentation. R2U-Net pays more attention to segmentation of large scale targets. BCDU lacks capability of recognizing contour details of tar-

Table 5 Performance comparison of the proposed network and the state-of-the-art methods on COVID-19 dataset

Method	F1-score	Sensitivity	Specificity	Accuracy	AUC
Unet	55.85	44.54	96.94	88.03	70.74
Attention Unet	57.80	46.80	96.90	88.39	71.85
BCDU	70.34	62.10	97.03	91.10	70.34
Unet++	64.36	52.44	**97.84**	90.13	75.14
DeeplabV3+	63.47	58.08	69.96	88.64	76.49
Ours	**75.16**	**74.18**	96.16	**91.97**	**83.82**

Bold values indicate the best result of each metric

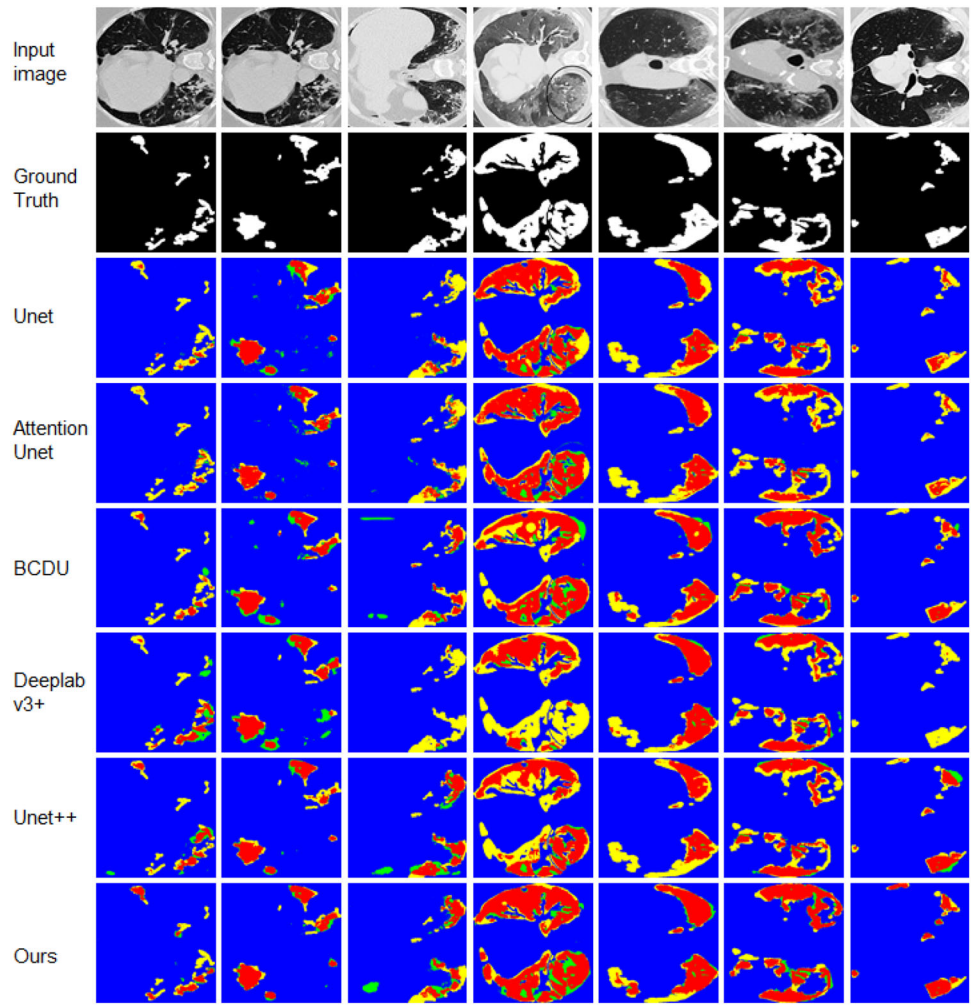


Fig. 5 Visual comparisons to different methods on COVID-19 dataset. Red = TP, blue = TN, yellow = FN, and green = FP

Table 6 Performance comparison of the proposed network and the state-of-the-art methods on CVC-ClinicDB dataset

Method	F1-score	Sensitivity	Specificity	Accuracy	AUC
Unet	71.24	63.03	**98.63**	95.46	80.83
Attention Unet	74.24	73.97	97.52	95.43	85.74
BCDU	71.12	64.17	98.41	95.36	81.28
Unet++	77.11	77.69	97.67	95.89	87.68
DeeplabV3+	73.75	68.26	98.35	95.67	83.30
Ours	**79.98**	**77.70**	98.38	**96.53**	**88.04**

Bold values indicate the best result of each metric

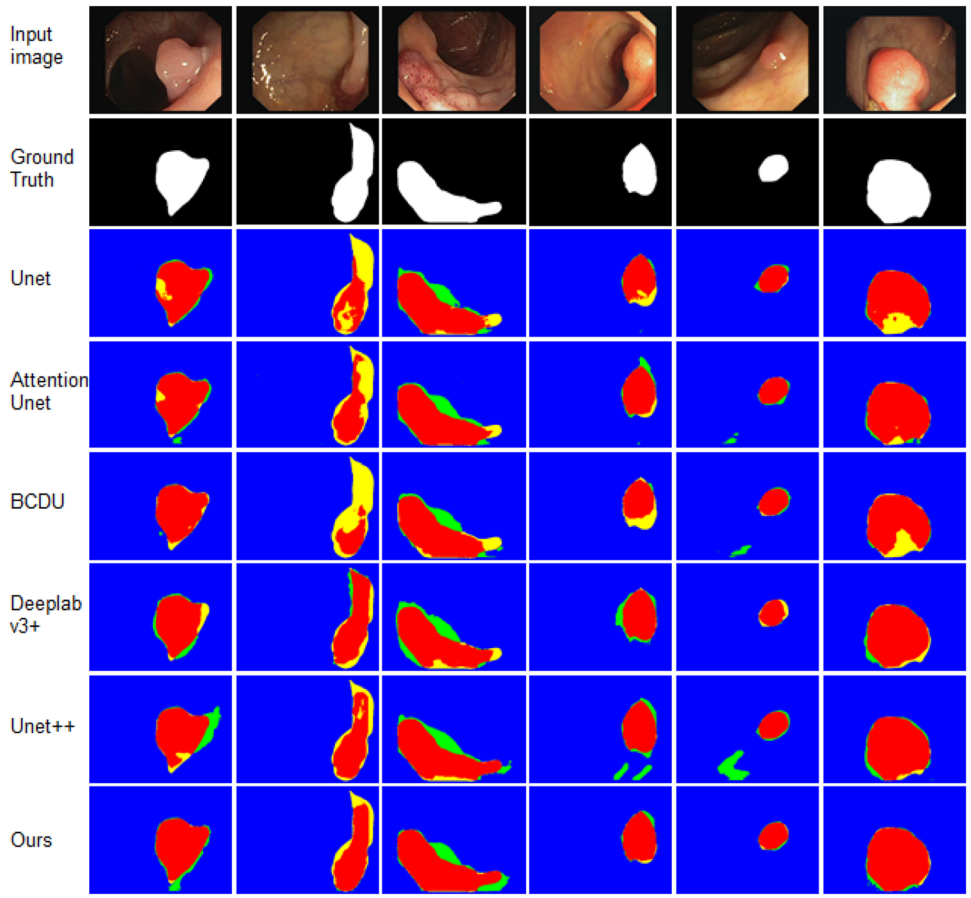


Fig. 6 Visual comparisons to different methods on CVC-ClinicDB dataset. Red = TP, blue = TN, yellow = FN, and green = FP

Table 7 Performance comparison of the proposed network and the state-of-the-art methods on ISIC 2018 dataset

Method	F1-score	Sensitivity	Specificity	Accuracy	AUC
Unet	85.07	80.65	96.44	91.95	88.54
R2U-Net	84.90	78.47	**97.46**	92.06	87.97
Attention Unet	84.97	79.57	96.93	91.99	88.25
BCDU	85.44	83.56	95.21	91.89	89.39
Unet++	**87.86**	83.35	**97.46**	**93.45**	**90.41**
DeeplabV3+	85.85	80.28	97.32	92.47	88.80
Ours	86.27	**86.28**	94.54	92.19	**90.41**

Bold values indicate the best result of each metric

get objects. Compared to these methods, our model exhibits remarkable performance in lung segmentation challenge.

4.6 Limitations and future work

We compare the total number of parameters contained in the proposed method with the above state-of-the-art methods. Table 9 lists the number of parameters of the above methods. From Table 9, we can see that parameters of the proposed method are 107.57M and it is less than other methods. We also compare the training speed (second per epoch) and the inference speed (FPS) of these methods. The results are also listed in Tables 10 and 11. We could see that Unet++ is converge faster, and Unet is faster than its extension methods in most situations. The results also show that the proposed method has less parameters, but its training speed and inference speed are slower than most methods. We attempt to explain this reason, but it is very difficult. We think it would be caused by the MultiBlock modules used in each step of encoding path. These MultiBlock modules add large convolutional filters to expand the receptive field; thus, they need more computations and increase the execution time. How to improve the speed and maintain the segmentation accuracy of this mechanism would be a significant problem.

It is well known that the segmentation performance of a deep neural network heavily relies on the characteristics of

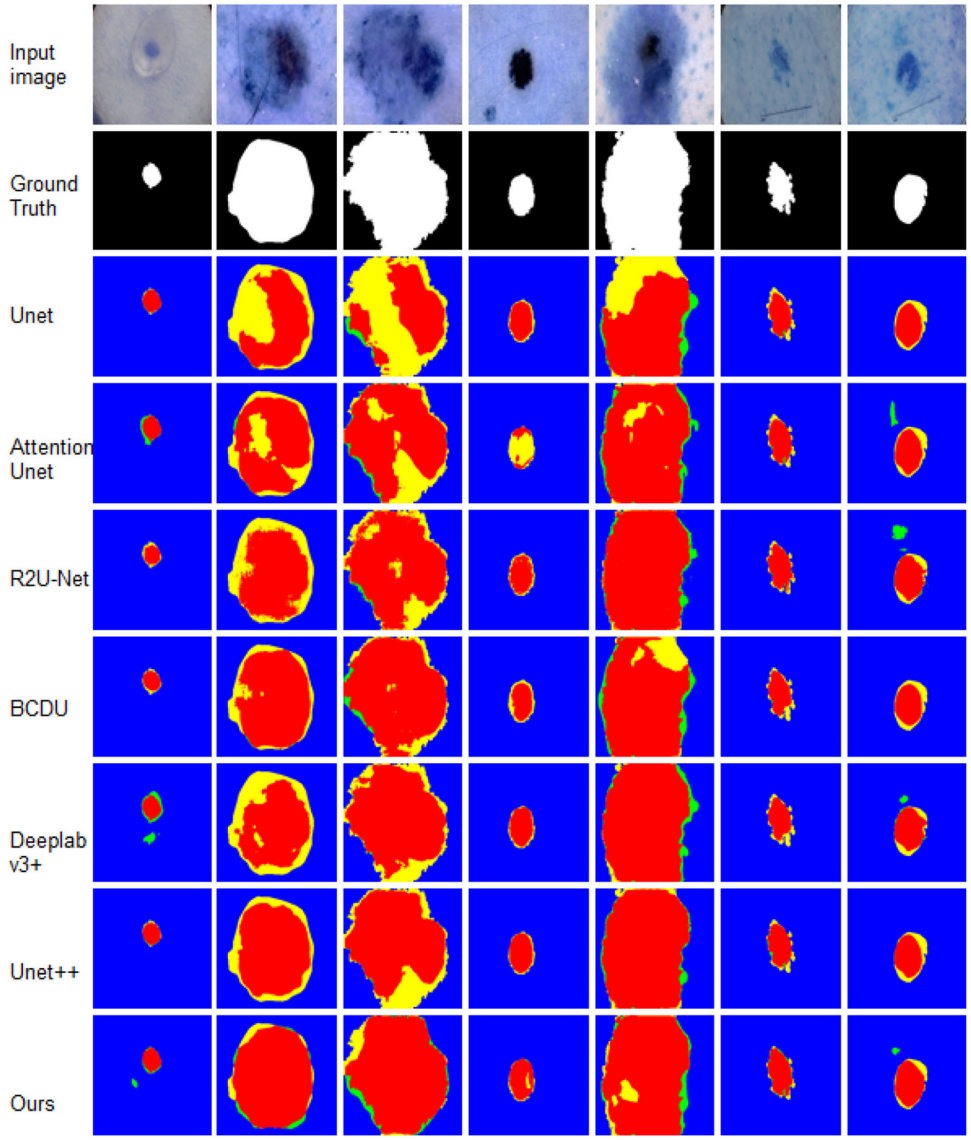


Fig. 7 Visual comparisons to different methods on ISIC 2018 dataset. Red = TP, blue = TN, yellow = FN, and green = FP

Table 8 Performance comparison of the proposed network and the state-of-the-art methods on Lung segmentation dataset

Method	F1-score	Sensitivity	Specificity	Accuracy	AUC
Unet	98.45	98.59	99.64	99.46	99.12
R2U-Net	98.33	99.38	99.42	99.42	**99.40**
Attention Unet	98.14	99.08	99.41	99.35	99.24
BCDU	98.43	99.02	99.55	99.45	99.28
Unet++	97.36	99.36	99.01	99.07	99.18
DeeplabV3+	94.40	**99.55**	97.63	97.96	98.59
Ours	**98.68**	98.89	**99.68**	**99.54**	99.29

Bold values indicate the best result of each metric

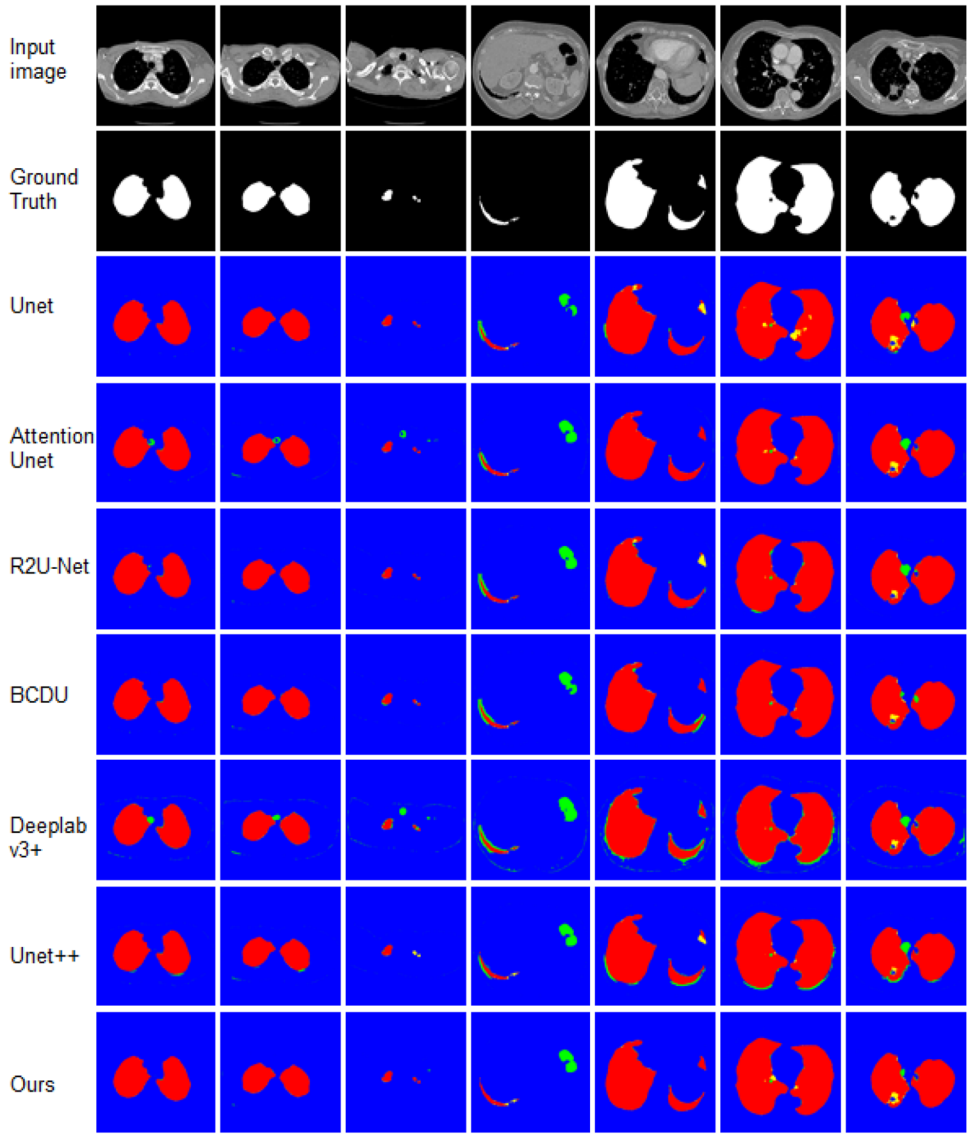


Fig. 8 Visual comparisons to different methods on Lung segmentation dataset. Red = TP, blue = TN, yellow = FN, and green = FP

the training sets. The experiments show that the presented method could be generalized across these four imaging modalities. However, although the boundaries between the segmented objects and background in these images are not obvious, the segmented objects are solitary without overlapping and adhesion. Segmentation of overlapping and adhesion objects is one of the most challenging problems which widely exist in medical image segmentation. We attempted to apply our method to segment nuclei on MoNuSeg dataset [20], which is a collection of Hematoxylin–Eosin (H&E) stained tissue images and exist large nuclear overlapping and adhesion. The results show that the proposed method fails to solve the problem of overlapping and adhesion as well as the comparable methods. The reasons would be that the lack of boundaries in the regions of overlapping and adhesion would affect the extraction of the contour information, which yields the method fails to recognize these objects. How to improve the capability of dealing with overlapping and adhesion would be one of our future works.

5 Conclusion

Medical image segmentation is a critical and important step for developing computer aided system in clinical situations. In this paper, we proposed an accurate algorithm for medical image segmentation which includes a semantic branch and a detail branch to extract the semantic and detail information, respectively. Inspired by the densely connected convolution, we design a MultiBlock module to replace convolutional block of classical Unet in order to utilize different receptive fields. We also add a spatial attention module between the encoding and decoding path to suppress abundant features to improve the network's representation capability. In

Table 9 Number of parameters of different methods

Method	Unet	R2U-Net	Attention Unet	BCDU	Unet++	DeeplabV3+	Ours
Parameters	355.39M	1.07G	365.56M	236.65M	122.69M	472.32M	**107.57M**

Bold values indicate the best result of each metric

Table 10 Training speed (second per epoch) of different methods on four datasets

Method	Unet	R2U-Net	Attention Unet	BCDU	Unet++	DeeplabV3+	Ours
COVID-19	213	–	241	323	**134**	436	210
CVC-clinicDB	60	–	**38**	39	**38**	65	58
Lung	154	182	96	149	**80**	116	135
ISIC	154	228	172	195	**120**	184	211

Bold values indicate the best result of each metric

Table 11 Inference speed (FPS) of different methods on four datasets

Method	Unet	R2U-Net	Attention Unet	BCDU	Unet++	DeeplabV3+	Ours
COVID-19	6.52	–	**6.69**	5.27	6.58	4.81	5.18
CVC-clinicDB	**10.86**	–	10.66	9.08	10.48	7.77	8.86
Lung	**7.85**	4.82	7.02	5.24	7.62	6.06	6.15
ISIC	46.26	26.92	42.91	33.23	36.09	**46.96**	33.94

Bold values indicate the best result of each metric

medical images, the differences between the target objects and the background commonly are not obvious, especially that there exist amount jagged contours and tiny objects; the high-level context information hardly well recognizes these objects. Therefore, we design a detail branch to enhance the detailed contour information implied in the shallow layers for improving the representation capability. In comparison with the state-of-the-art methods, our method achieves a remarkable performance on four public medical image segmentation challenges.

Acknowledgements This work has been partially supported by the National Natural Science Foundation of China (No. 61701218), the Natural Science Foundation of Hunan Province of China (No. 2018JJ3449), Education Commission of Hunan Province of China (No. 17B229, 18A253).

Compliance with ethical standards

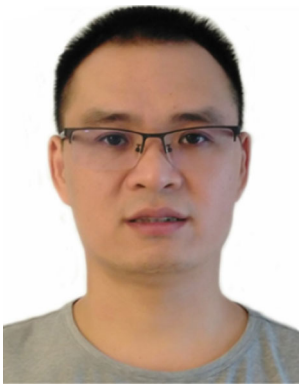
Conflict of interest The authors declare that they have no conflict of interest.

References

1. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)
2. Asadi-Aghbolaghi, M., Azad, R., Fathy, M., Escalera, S.: Multi-level context gating of embedded collective knowledge for medical image segmentation. arXiv preprint arXiv:2003.05056 (2020)
3. Azad, R., Asadi-Aghbolaghi, M., Fathy, M., Escalera, S.: Bi-directional convlstm u-net with densley connected convolutions. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 406–415 (2019)
4. Baldeon-Calisto, M., Lai-Yuen, S.K.: Adaresu-net: multiobjective adaptive convolutional neural network for medical image segmentation. Neurocomputing **392**, 325–340 (2020)
5. Bazin, P.L., Pham, D.L.: Homeomorphic brain image segmentation with topological and statistical atlases. Med. Image Anal. **12**(5), 616–625 (2008)
6. Bernal, J., Snchez, F.J., Fernndez-Esparrach, G., Gil, D., Rodrguez, C., Vilario, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput. Med. Imaging Graph. **43**, 99–111 (2015)
7. Carballido-Gamio, J., Belongie, S., Majumdar, S.: Normalized cuts in 3-D for spinal MRI segmentation. IEEE Trans. Med. Imaging **23**(1), 36–44 (2004)
8. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834 (2018)
9. Chen, X., Williams, B.M., Vallabhaneni, S.R., Czanner, G., Williams, R., Zheng, Y.: Learning active contour models for medical image segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11632–11640 (2019)
10. Choudhury, A.R., Vanguri, R., Jambawalikar, S.R., Kumar, P.: Segmentation of brain tumors using deeplabv3. In: International MICCAI Brainlesion Workshop, pp. 154–167 (2018)
11. Christ, P.F., Ettlinger, F., Grn, F., Elshaera, M.E.A.,Lipkova, J., Schlecht, S., Ahmaddy, F., Tatavarty, S.,Bickel, M., Bilic, P., Rempfler, M., Hofmann, F., Anastasi, M.D., Ahmadi, S.A., Kaissis, G., Holch, J., Sommer,W., Braren, R., Heinemann, V., Menze, B.: Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv preprint arXiv:1702.05970 (2017)

12. Chung, D.H., Sapiro, G.: Segmenting skin lesions with partial differential equations based image processing algorithms. In: Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), vol. 3, pp. 404–407 (2000)
13. Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. Adv. Neural Inf. Process. Syst. **25**, 2843–2851 (2012)
14. Codella, N.C.F., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S.W., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M.A., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1902.03368 (2019)
15. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J.M., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
16. Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: Inf-net: Automatic covid-19 lung infection segmentation from CT images. IEEE Trans Med Imaging (2020)
17. Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., Fan, C.: Sa-unet: Spatial attention u-net for retinal vessel segmentation. arXiv preprint arXiv:2004.03696 (2020)
18. Hatamizadeh, A., Terzopoulos, D., Myronenko, A.: End-to-end boundary aware networks for medical image segmentation. In: H. Suk, M. Liu, P. Yan, C. Lian (eds.) Machine Learning in Medical Imaging—10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings, Lecture Notes in Computer Science, vol. 11861, pp. 187–194. Springer (2019)
19. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017)
20. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans. Med. Imaging **36**(7), 1550–1560 (2017)
21. Li, H., He, X., Zhou, F., Yu, Z., Ni, D., Chen, S., Wang, T., Lei, B.: Dense deconvolutional network for skin lesion segmentation. IEEE J. Biomed. Health Informat. **23**(2), 527–537 (2019)
22. Liao, F., Liang, M., Li, Z., Hu, X., Song, S.: Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network. IEEE Trans. Neural Netw. **30**(11), 3484–3495 (2019)
23. Litjens, G.J.S., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Snchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
24. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015)
25. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 4th International Conference on 3D Vision (3DV), pp. 565–571 (2016)
26. Nguyen, H.T., Worring, M., van den Boomgaard, R.: Watersnakes: energy-driven watershed segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **25**(3), 330–342 (2003)
27. Nunzio, G.D., Tommasi, E., Agrusti, A., Cataldo, R., Mitri, I.D., Favetta, M., Maglio, S., Massafra, A., Quarta, M., Torsello, M., Zecca, I., Bellotti, R., Tangaro, S.S., Calvini, P., Camarlinghi, N., Falaschi, F., Cerello, P., Oliva, P.: Automatic lung segmentation in CT images with accurate handling of the Hilar region. J. Digit. Imaging **24**(1), 11–27 (2011)
28. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M.C.H., Heinrich, M.P., Misawa, K., Mori, K., McDonagh, S.G., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
29. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging **35**(5), 1240–1251 (2016)
30. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241 (2015)
31. Snchez-Gonzlez, A., Garca-Zapirain, B., Sierra-Sosa, D., Elmaghraby, A.: Automatized colon polyp segmentation via contour region analysis. Comput. Biol. Med. **100**, 152–164 (2018)
32. Sun, J., Darbehani, F., Zaidi, M., Wang, B.: Saunet: Shape attentive u-net for interpretable medical image segmentation. arXiv preprint arXiv:2001.07645 (2020)
33. Taghanaki, S.A., Abhishek, K., Cohen, J.P., Cohen-Adad, J., Hamarneh, G.: Deep semantic segmentation of natural and medical images: a review. Artif. Intell. Rev. 1–42 (2020)
34. Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)
35. Valindria, V.V., Lavdas, I., Cerrolaza, J., Aboagye, E.O., Glocker, B.: Small organ segmentation in whole-body MRI using a two-stage FCN and weighting schemes. arXiv preprint arXiv: 1804.03999 (2018)
36. Wang, R., Chen, S., Ji, C., Fan, J., Li, Y.: Boundary-aware context neural network for medical image segmentation. arXiv preprint arXiv:2005.00966 (2020)
37. Xi, P., Guan, H., Shu, C., Borgeat, L., Goubran, R.: An integrated approach for medical abnormality detection using deep patch convolutional neural networks. Vis. Comput. **36**(9), 1869–1882 (2020)
38. Xia, Y., Yang, D., Yu, Z., Liu, F., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Med. Image Anal. **65**, 101766 (2020)
39. Xie, J., Jiang, Y., tat Tsui, H.: Segmentation of kidney from ultrasound images based on texture and shape priors. IEEE Trans. Med. Imaging **24**(1), 45–57 (2005)
40. Yang, Y., Jia, W., Wu, B.: Simultaneous segmentation and correction model for color medical and natural images with intensity inhomogeneity. Vis. Comput. **36**(4), 717–731 (2020)
41. Yang, Y., Wang, R., Feng, C.: Level set formulation for automatic medical image segmentation based on fuzzy clustering. Signal Process. Image Commun. **87**, 115907 (2020)
42. Zhou, X.Y., Shen, M., Riga, C.V., Yang, G.Z., Lee, S.L.: Focal fcn: Towards small object segmentation with limited training data. arXiv:1711.01506 (2017)
43. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: a nested u-net architecture for medical image segmentation. In: Deep learning in Medical Image analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11 (2018)
44. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging **39**(6), 1856–1867 (2020)

Zhiming Cheng was born in Wuhan, Hubei Province, China, in 1994. He received his B.S. degree in electronic and information engineering from Wuhan University of Engineering Science. He is currently pursuing the M.S. degree in software engineering at University of South China. His research interest is medical image analysis.



Aiping Qu was born in Hunan Province, China, in 1982. He received B.S. and M.S. degrees in mathematics science from Hunan University, in 2005 and 2010, respectively, and the Ph.D degree in computer science from Wuhan University , in 2015. He is currently an associate professor of computer science in University of South China. His research interests include machine learning and medical image analysis.

Xiaofeng He was born in Hunan Province, China, in 1971. He received B.S. degrees in mathematics science from XiangTan University in 1995 and the M.S. degree in computer science from University of South China, in 2006. He is currently an associate professor of computer science in University of South China. His research interest is computer vision.