Application of U-shaped Convolutional Neural Network Based on Attention Mechanism in Liver CT Image Segmentation

Chen Li, Wei Chen, Xin Luo, Mingfei Wu, Xiaogang Jia, Yusong Tan, and Zhiying Wang

The School of Computer Science, National University of Defense Technology, Changsha 410073
lichen14@nudt.edu.cn

Abstract. Automatic segmentation of livers from medical images is a difficult task because of liver's uneven shapes. Recently, deep convolutional neural networks have become popular in this field because the deep learning method can learn hierarchical semantic information. In this paper, we propose to apply a deep learning model Attention U-Net to liver medical image segmentation. The proposed model uses classic UNet as basic architecture and combines the attention mechanism. Attention U-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task. Besides, Attention U-Net is an end-to-end image-based approach for liver segmentation and does not rely on any pre-trained models or common post-processing techniques. As a consequence, there is no need for the clipping of the region of interest(ROI) and the locating of target object in the network. The proposed model is evaluated on the public dataset from ISBI2019 Challenge for CT liver segmentation. Experiments demonstrate that Attention U-Net achieves an IoU gain of 5.15%,a Dice gain of 3.51%, a accuracy gain of 5.78% and a recall gain of 4.81% over UNet.

Keywords: Attention; UNet; Liver Segmentation;Deep Learning

1 Introduction

Liver cancer is the common internal malignancies in the world and the leading causes of cancer death, which pose a huge threat to human health. Identifying the position of liver from medical images is a preparation step in diagnosing liver lesions and plays an indispensable role in disease treatment. However, segmentation of liver from medical images is very challenging because of the liver's uneven shape and the low contrast between the tissues in human body.

Researches on solutions to liver segmentation have been proposed. Generally, liver segmentation methods can categorized in to 2 classes: (1)manual and semi-automatic segmentation, (2)automatic segmentation. Manual segmentation largely relies on experts with advanced technical skills to perform such tasks. As a consequence, the quality of the segmentation relies heavily on the judgment of experts, which is time-consuming and not repeatable. These factors make manual segmentation impractical. Meanwhile, semi-automatic segmentation still requires manual intervention, which can lead to biases and errors. Because densely labeling massive medical images by manual or semi-manual methods is a tedious and high error rate task, extensive research has been conducted on automatic medical image segmentation. With the advent of convolutional neural network (CNN) and its rapid application in medical image segmentation, it has been used in many automatic medical image analysis tasks such as cardiac MRI image segmentation [1] and cancerous pulmonary nodule segmentation[2]. However, this CNN-based method usually divides the segmentation task into two steps: localization and segmentation. The extra localization step will increase the amount of model parameters and bring extra time consumption. The accuracy of model segmentation also depends heavily on the first step positioning accuracy.

In order to accelerate and facilitate diagnosis, and assist doctors in removing diseased areas such as tumors, it is necessary to develop a reliable automated solution to accurately segment organs from medical images. Inspired by attention mechanism, we use an attention-based U-shaped convolutional neural network called Attention U-Net for liver segmentation, designed to efficiently outline the liver from CT images without localization step. Attention U-Net integrates the UNet encoder-decoder architecture and Attention Gate learning mechanism, which can accelerate inference process and improve segmentation performance. The contributions of our work are summarized as follows:

1) Attention U-Net is introduced for liver CT image segmentation.
2) The experiments are conducted on public dataset show that Attention Gate has the ability to focus on specific parts of the whole image.
3) Attention U-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task.
4) Attention U-Net has superior performance on liver segmentation task compared to equivalent UNet model.

The structure of this paper is as follows: Section 2 will briefly review related segmentation methods. Section 3 will describe in detail our segmentation framework. Then we will introduce the settings of the experiment in Section 4. After that, the experimental results will be displayed and analyzed. Finally, the conclusions are given in Section 5.

2 Related Works

2.1 U-shaped convolutional neural networks

The state-of-art image segmentation models are variants of encoder-decoder architectures such as UNet[3] and Fully Convolutional Network (FCN)[4]. These models have a common similarity: skip connection. Skip connection combines the abstract feature information restored from the decoder and the surface feature information extracted from the encoder in the channel dimension. In complex segmentation tasks, skip connection can effectively retain the feature information of each level and generate segmentation results with low-level details.

Due to the difficulty in collecting medical image data, it is impossible to obtain sufficient mass training data in biomedical image analysis tasks. But since UNet[3] proposed in 2015, it has been widely used in the field of image segmentation due to its simple but effective structural characteristics. The network structure is shown in Fig.1 below. UNet is composed of two parts, the concatenation path on the left and the expansive path on the right. It uses a U-shaped frame of a fully convolutional neural network to make the expansion path and the contraction path approximately symmetrical. UNet can get the feature context information from each level in the contraction path and highlight the position of the foreground target object. The restored feature information is obtained in the expansion path, and then the features extracted on multiple dimensions are merged through skip connections, combined shallow features and deep features.

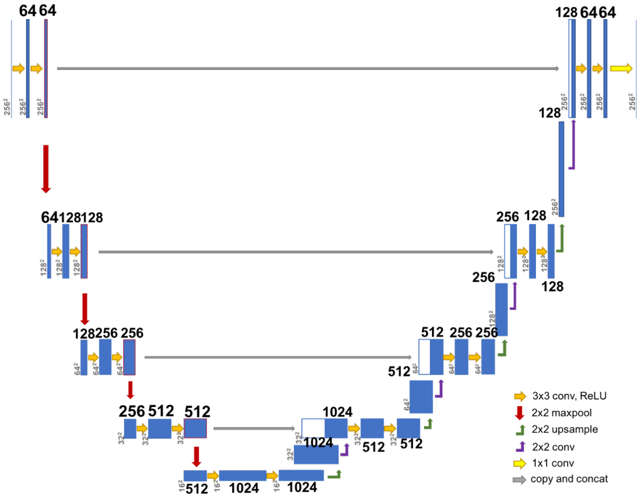

Fig. 1: UNet Architecture

2.2 Attention Mechanism

Most current segmentation frameworks simplify the segmentation task into two steps: object position and segmentation. Although UNet is well-represented and widely-applied in medical image segmentation such as cardiac MRI[5], abdominal CT[6] segmentation, pulmonary CT nodule detection[2] and previous UNet-based segmentation networks still rely on multi-level cascaded CNNs to extracts the region of interest (ROI), especially when target objects show large patient-to-patient differences such as shape and size. However, this multi-level cascaded CNN method largely relies on previous object positioning accuracy and requires more model parameters and higher computing power. In order to solve this problem, we refer to the method proposed by Oktay[7] to add a simple but effective Attention Gate mechanism to the network architecture.

Attention mechanism firstly emerged in the field of natural language processing(NLP) and quickly gained dominance. It was Nonlocal[8] proposed by He Kaiming's team that first introduced the attention mechanism to the field of computer vision. Since then, a series of research booms have focused on attention mechanisms. The attention mechanism was firstly introduced in the semantic segmentation in[9], which combined the shared network with the attention mechanism and achieved superior performance. Later in 2018, many high-impact articles were published such as DANet[10], OCNet[11], CCNet[12], and the Local Relation Net[13] in 2019. Recently, attention mechanism has gradually been applied to medical image segmentation.

Inspired by UNet[3], Oktay[9] proposed a model combining attention mechanism and UNet. The attention mechanism focuses the network learning attention on distinguishing the relevant positions of the object through the Attention Gate. The inputs of the Attention Gate are the upsampling feature in the expansion path and the corresponding feature in the contraction path. The former is used as a gating signal to enhance the learning of the task area of interest, and suppress the tissue area unrelated to the segmentation task. Therefore, the Attention Gate can increase the effectiveness of skip connections in return. Next, the S-shaped activation function sigmoid is selected to train the convergence of parameters in the Attention Gate. The output is the multiplication result of the input feature and the attention coefficient α. Another advantage of Attention Gate is that it can use standard back-propagation to update the training parameters in the gate without using the sample-based update method in[14].

3 Methodology

The proposed Attention U-Net uses UNet as the basic network framework, and an encoder and a decoder are symmetrically arranged on both sides of the network. The structure parameter configuration of Attention U-Net is listed in Table 1.

Encoder is a typical framework of a convolutional neural network and consists of downsampling modules. Each downsampling module firstly consists of two consecutive 3x3 convolutional layers. The convolutional layer processes the local information layer by layer to obtain the image feature x_l^c of the c-th channel extracted by the l-th layer. After each convolution layer, a rectified linear function (ReLU) is used as the activation function $\sigma(x_l^c) = max(0, x_l^c)$. The end of the downsampling module is a 2x2 max pooling for downsampling and doubling the number of feature channels. In summary, the feature information x_l^c can be

Table 1: Attention U-Net network structure parameter configuration

Encoder	**Output Size**	**Decoder**	**Skip Connection**	**Output Size**
Input	256ˆ2*1	Up1		32ˆ2*256
Conv1	256ˆ2*32	Attention Gate 1	[Up1,Conv4]	32ˆ2*512
Pooling	128ˆ2*32	Conv6		32ˆ2*256
Conv2	128ˆ2*64	Up2		64ˆ2*128
Pooling	64ˆ2*64	Attention Gate 2	[Up2,Conv3]	64ˆ2*256
Conv3	64ˆ2*128	Conv7		64ˆ2*128
Pooling	32ˆ2*128	Up3		128ˆ2*64
Conv4	32ˆ2*256	Attention Gate 3	[Up3,Conv2]	128ˆ2*128
Pooling	16ˆ2*256	Conv8		128ˆ2*64
Conv5	16ˆ2*512	Up4		256ˆ2*32
		Attention Gate 4	[Up4,Conv4]	256ˆ2*64
		Conv9		256ˆ2*32
		Conv10		256ˆ2*1

expressed as:

$$x_l^c = \sigma\left(\sum_{c\prime \in F_l} x_{l-1}^{c\prime} * K^{c\prime,c}\right),$$

Where c represents the channel, F_l represents the number of convolution kernels of the l-th layer, $K^{c\prime,c}$ represents the $c\prime$-th convolution kernel, $*$ represents the convolution operation, and σ represents the activation function ReLU.

Decoder consists of upsampling modules that upsamples the features to restore image information. Each upsampling module firstly uses 2x2 deconvolution to halve the feature size and the number of channels, then connects the compressed features after concatenation, and finally performs 3x3 convolutions twice. In the last layer, a 1x1 convolution is used to map each feature vector to the required number of classes.

The structure of the Attention U-Net network is shown in Fig.2. Attention U-Net is designed for liver segmentation and does not rely on any pre-trained models or common post-processing techniques. As a consequence, there is no need for the clipping of the region of interest(ROI) and the locating of target object at the beginning. Compared with the traditional classic UNet, Attention U-Net optimizes the previous Conv block from “Conv 3x3, ReLU” to “Conv 3x3, Batch Normalization, ReLU” . We want to accelerate the model’s convergence speed and alleviate the vanishing gradient problem in the deep network to a certain extent through this optimization. As a consequence, it makes training the deep network model easier and more stable.

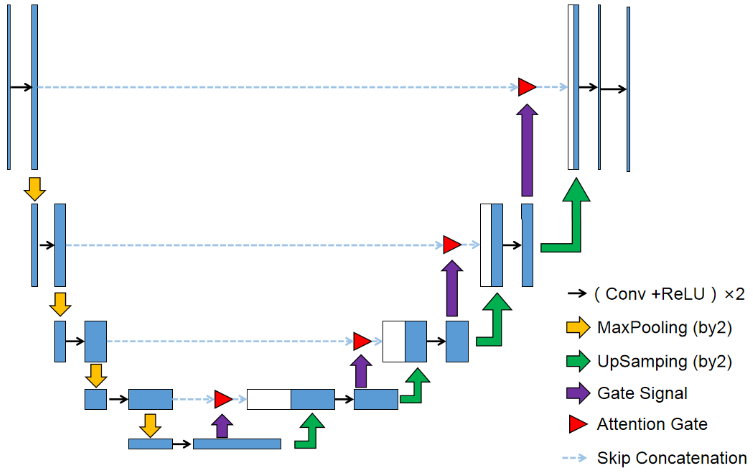

Fig. 2: Attention U-Net architecture

4 Experiments and results

4.1 Experimental setup

In this paper, we use the public dataset CHAOS from the 2019 IEEE International Symposium on Biomedical Imaging (ISBI) to train and test the network. This database contains DICOM images of 20 different patients with a resolution of 512x512. CT images were obtained from the patient's upper abdomen after injection of the contrast agent. The data were collected by three instruments: Philips SecuraCT with 16 detectors, Philips Mx8000 CT with 64 detectors, and Toshiba AquilionOne with 320 detectors. All three instruments are equipped with the spiral CT option. The xy pitch is between 0.7-0.8 mm. Images in the CHAOS dataset is shown in Fig.3.

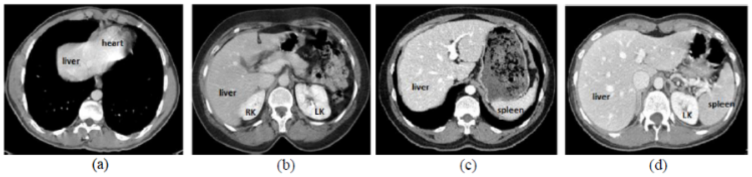

Fig. 3: Example image from the CHAOS dataset

The average number of slices in per patient sub-dataset is 90, the minimum number of slices is 79, and the maximum number of slices is 264. The total 2342 images are randomly divided into a train set and a test set. The train set includes

2303 slices and the test set includes 39 slices. The experiment is based on 64-bit Ubuntu operating system and NVIDIA GTX 1080Ti GPU, and the deep learning framework used is Pytorch. The entire network uses Adam's method to optimize parameters, using a varying learning rate, which is initially 0.03 and decreases with epoch. Dice similarity coefficient is selected as loss function for back propagation of errors, which is defined as follows:

$$Dice\left(Y,\overline{Y}\right)=\frac{2Y\overline{Y}}{Y+\overline{Y}}\ ,$$

Where $\overline{Y}$ is the real result and Y is the segmentation result.

4.2 Experimental results and analysis

We use the Attention U-Net model to segment CT images of the liver to verify the effectiveness of the proposed method. Dice similarity coefficient, IoU coefficient, precision, recall are used as performance indicators to evaluate the performance of liver CT image segmentation. The larger the values of these four indicators, the larger the overlapping area between the segmentation result and the real result, the higher the similarity, and the greater the accuracy of the segmentation. The results are shown in the following Table 2.

Table 2: Segmentation results of methods on the test dataset

Network	**IoU**	**Dice**	**Precision**	**Recall**
UNet[3]	0.8350	0.8697	0.8699	0.8822
Attention U-Net	**0.8780**	**0.9002**	**0.9202**	**0.9246**

It is known from Table 2 that under the evaluation of four indicators, the U-shaped convolutional neural network after introducing the attention mechanism has greatly improved the liver CT image segmentation performance, of which the IoU ratio increased by 5.15%, the Dice coefficient increased by 3.51%, and the accuracy increased 5.78%, recall rate increased 4.81%.

By comparing the segmentation results of our model and UNet in Fig.4, we can intuitively conclude that the Attention U-Net can better restore the characteristic structure of the liver, and more fully capture the detailed information that U-Net ignores.

By comparing the differences between the segmentation results and the actual segmentation results, we can conclude that our experimental results proved that the attention mechanism can effectively solve the problem of cascaded CNN

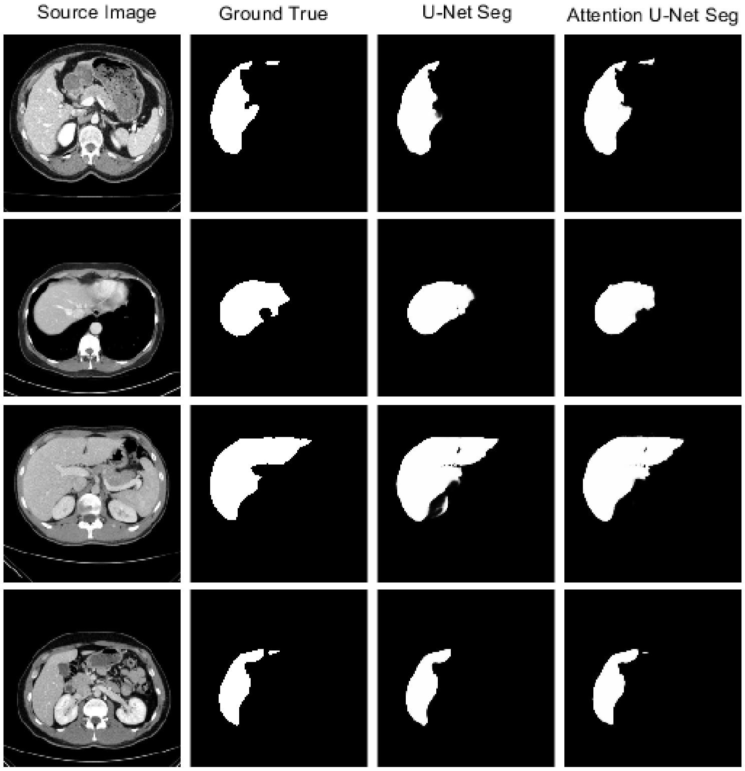

Fig. 4: Liver segmentation results of UNet model and Attention U-Net model and true segmentation results

method in Section2.2. The main innovations in Attention U-Net are: Attention Gate mechanism is added and the features extracted at different levels can be merged with a focused selection. The features extracted by the encoder is propagated to the decoder through the Attention Gate before skip connection, so that more targeted hierarchical context information can be extracted; after receiving the features, the decoder restores the features in a bottom-up manner. In the expansion path, the accuracy and stability of the network are enhanced.

5 Conclusions

In this paper, a U-shaped convolutional neural network based on attention mechanism is proposed and applied to liver CT image segmentation. Our experimental results proved that Attention U-Net enhances the learning of task regions and increases the effectiveness of skip connections. Attention Gate merges the features selected by different levels in the encoder with a task-related selection into the decoder. Attention U-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task. As a consequence, there is no need to trim the ROI and locate the target object in our model. In addition, the method is generalized and modular in the field of medical image analysis, so it can also be applied to image classification and regression problems. We conclude that introduced Attention mechanism can enhance the accuracy and scalability of network, and Attention U-Net used in this paper can accurately segment human tissues or organs.

References

1. Giacomo Tarroni. Wenjia Bai, Matthew Sinclair, “Human-level CMR image analysis with deep fully convolutional networks,” *CoRR*, vol. abs/1710.09289, 2017.
2. Fangzhou Liao, Liang Ming, Li Zhe, Xiaolin Hu, and Sen Song, “Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, 2017.
3. Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” .
4. Jonathan Long, Evan Shelhamer, and Trevor Darrell, “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2014.
5. Mahendra Khened, Varghese Alex, and Ganapathy Krishnamurthi, “Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers,” *Medical Image Analysis*.
6. Holger R. Roth, Le Lu, Nathan Lay, Adam P. Harrison, Amal Farag, Andrew Sohn, and Ronald M. Summers, “Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation,” *CoRR*, vol. abs/1702.00045, 2017.
7. Ozan Oktay, Jo Schlemper, Loic Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert, “Attention u-net: Learning where to look for the pancreas,” 04 2018.
8. Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, “Non-local neural networks,” .
9. Jo Schlemper, Ozan Oktay, Liang Chen, Jacqueline Matthew, Caroline L. Knight, Bernhard Kainz, Ben Glocker, and Daniel Rueckert, “Attention-gated networks for improving ultrasound scan plane detection,” *CoRR*, vol. abs/1804.05338, 2018.
10. Jun Fu, Jing Liu, Haijie Tian, Zhiwei Fang, and Hanqing Lu, “Dual attention network for scene segmentation,” .
11. Yuhui Yuan and Jingdong Wang, “Ocnet: Object context network for scene parsing,” .
12. Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu, “Ccnet: Criss-cross attention for semantic segmentation,” .
13. Han Hu, Zheng Zhang, Zhenda Xie, and Stephen Lin, “Local relation networks for image recognition,” .
14. Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu, “Recurrent models of visual attention,” *CoRR*, vol. abs/1406.6247, 2014.