A U-Net based multi-scale feature extraction for liver tumour segmentation in CT images

Ming Gong[1], John Soraghan[1], Gaetano Di Caterina[1], Derek Grose[2]

[1]Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
[2]Beatson West of Scotland Cancer Centre, Glasgow, UK

Abstract. A new method for automatic liver tumour segmentation from computed tomography (CT) scans based on deep neural network is presented. Two cascaded deep convolutional neural networks are used to segment the CT image of the abdominal cavity. The first U-net is used for coarse segmentation to obtain the approximate position of the liver and tumour. Using this as a prediction the original image is cropped to reduce its size in order to increase the segmentation accuracy. The second modified U-net is employed for accurate segmentation of the actual liver tumours. Residual modules and dense connections are added to U-net to help the network train faster while producing more accurate results. In addition, multi-dimensional information fusion is introduced to make the network more comprehensive in restoring information. The Liver Tumour Segmentation (LiTs) dataset is used to evaluate the relative segmentation performance obtaining an average dice score of 0.665 based our method.

Keywords: CT data, Liver tumour, U-Net, multi-scale feature fusion

1 Introduction

In recent years, the incidence and mortality of liver cancer have maintained a worldwide upward trend. Proper liver image segmentation can yield accurate liver volume, which has an important impact on liver surgery and patient's postoperative recovery evaluation. However, due to highly variable shape, close proximity to other organs and diverse pathologies, the liver tissue in the abdominal CT image will have the same in-tensity value as the adjacent organs like stomach, heart, pancreas, kidney [1]; and furthermore its shape will be deformed. In clinical applications, liver segmentation is usually done by experienced radiologists. This remains an extremely time-consuming task that can reach 90 minutes per patient, with a degree of inter and intra segmentation errors. Recently threshold segmentation regional growth algorithm, active contour model and level set methods have also been applied to the segmentation of liver tumours [2]. However, the above methods all have disadvantages of taking significant time. Consequently liver segmentation is a challenging task in the field of medical image processing.

In the past few years, Deep Learning has become the main research direction of computer vision problems. Due to the achievements of the Convolutional Neural

adfa, p. 1, 2011.

Network in image classification, a growing number of researchers choose to use deep neural networks to solve a series of image processing tasks, such as image recognition, object detection, and segmentation [3].

A convolutional neural network for end-to-end and pixel-to-pixel semantic segmentation [4] led to a new direction for image segmentation. In particular, U-net is a popular modified segmentation network based on fully convolutional neural network that is widely used in biomedical image segmentation, such as pulmonary nodule, sclerotic metastases, lymph node and colonic polyp [5]. Benefitting from the symmetrical encoding-decoding structure and the skip-connection of feature map, U-net can extract features at different levels and obtain high-resolution prediction. Due to the lighter network structure of U-net and the high accuracy performance of medical image segmentation, U-net has become the baseline for many medical image segmentation challenges.

A symmetric encoder-decoder architecture called SegNet with pooling indices to segment liver tumour was presented in [6]. Two cascaded U-net separate the liver and the tumour independently was presented in [7], and the tumour segmentation is based on the results of liver segmentation. They subsequently used a 3D-CRF to post- process the segmentation results to improve the accuracy of the segmentation. A joint cascaded U-Net is used to segment liver and tumour simultaneously was proposed in [8] . They cascaded the predicted image of liver and the original scan to segment the tumour. The above methods all use a U-Net similar structure for tumor segmentation, but they are not sufficient for feature extraction in different dimensions. Moreover, the cascaded network will lead to huge network parameters, excessive training time and over-fitting problems. The method we propose will improve these problems.

The rest of this article is organized as follows, the second part introduces the new tumour segmentation including cascade segmentation method and a modified U-Net. The third part shows the performance of this method on public data sets. The last part summarizes the full paper.

2 Methods

2.1 Data Preprocessing

The data set used [9] contains scanned images of the patient's entire abdominal cavity. We only select slices containing liver tissue for training and testing and save them locally for batch processing. We perform denoising, smoothing and contrast enhancement of the data set to increase the final prediction accuracy. In order to reduce the serious class imbalance problem caused by relatively small size tumours in the prediction problem the original image is divided into 4 slices with size 256*256. Slices containing the tumour according to the tumour label are selected. In addition, we also specially cut a scan that only contains the liver for training. Data augmentation is employed using rotation, flip and zoom for original CT scan to obtain more variant tumour shapes.

2.2 Methodology

Our proposed segmentation workflow is shown in Fig.1. The workflow consists of three major steps. The first step includes data selection and data preprocessing for network. In a second step， using a cascade of a U-Net followed by a modified U-Net to jointly segment the liver and tumor, where the tumor segmentation is based on the seg-mentation result of the liver. In the final step, the predicted tumor lesions need to be restored to the original image size according to the boundary of the cropping.

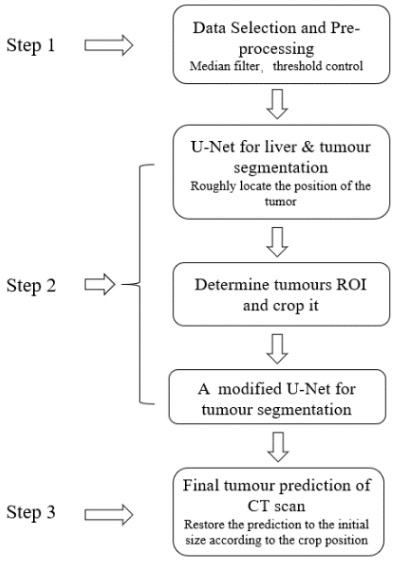

Fig. 1. A flow diagram of the proposed U-Net Cascade algorithm

A classic U-Net mainly composes a down-sampling path, up-sampling path, and skip connection. The down-sampling path is the same as an ordinary deep convolutional network, which usually includes convolution operation, pooling operation, and drop-out to obtain features at different levels. The purpose of up-sampling is to restore the extracted high-level features to the size of the input image. This process is mainly done by deconvolution or un-pooling. Using skip connection to introduce feature information on the corresponding scale into the up-sampling process, providing multi-scale and multi-level information for later image segmentation. At the end of the network, the sigmoid function is used to classify the feature map between 0-1 to obtain the prediction distribution prediction map.

An original U-net is used for roughly segmenting liver and tumour to ensure the approximate location of them. The purpose of segmenting the liver is to exclude those tumour predictions out of liver organ. We use a U-Net that includes four down-sampling and up-sampling. The number of convolution kernels is increased from 64 to 512 in a double increment and the size of the convolution kernel is set to 3*3. Use same padding to ensure that the final prediction result matches the input image.

Using the prediction results of the first U-net, we crop the input image to obtain a smaller input image to increase the ratio of foreground to background, which can improve the class imbalance problem. Through the four boundary points of the tumour, we can calculate the centre position of the tumour distribution and crop a 256*256 image with this centre point. The boundary of the tumour will be recorded and used to restore the prediction result to the original size at the end (512*512).

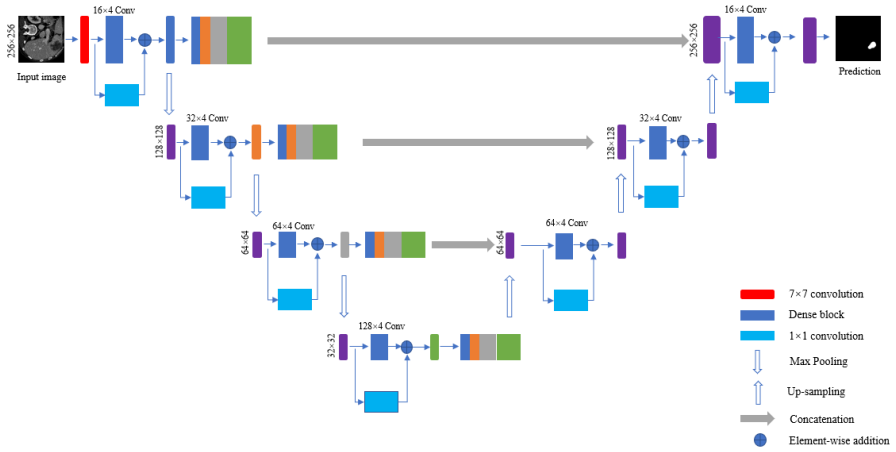

Fig.2. The architecture of proposed neural network for the second segmentation task

As illustrated in Fig.2, the fine segmentation process comprises a modified U-Net structure. The initial U-net designed 4 or 5 pooling layers for extracting deeper features. However excessive pooling process will inevitably lead to the loss of information, which is non negligible in small target segmentation. The size of earlier liver tumours is usually less than 5cm, and it only accounts for approximately 5%-8% of the entire CT image. After several times of down-sampling, this important tumour information will be reduced or even disappear completely. We have experimentally determined that using three pooling layers allows the network to obtain sufficient features, while retaining enough detailed information.

The input image first passes through a 7×7 convolutional layer to obtain some low-level semantic information with a larger receptive field. Then use a convolution block to generate the feature map of this layer. The pooling operation is used to halve the dimension of the feature map, and convolution block is used repeatedly to obtain a higher-level feature map. We used pooling operation three times and obtained four levels of semantic features, such as the blue, orange, gray, and green blocks in Fig. 2. These different levels of semantic information are fused together in their respective level. This part will be explained in the later part. Use the feature maps of these information fusions to perform feature restoration with skip connection.

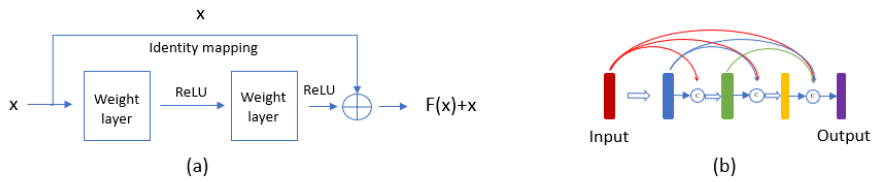

Fig. 3. (a) represents a basic residual connection, (b) represents a dense block

Fig. 3 shows the structure of the residual module and the dense module, respectively. In each feature extraction block, we use dense module, respectively. In each feature extraction block, we use dense module [10] to increase the depth of the network and reduce the risk of gradient vanish. Each feature layer is cascaded with previous features to form a new input, and all feature layers are cascaded to form the final output. Dense connection emphasizes the repeated use of features, and the features of the previous layer will be used as part of the input in the subsequent feature extraction. This allows the network to learn more features with limited layers. Furthermore, we use the residual connection throughout the whole network [11]. This process is realized by adding the identity mapping of the input and the output. Because residual learning is easier than learning the original features directly. The mapping after introducing the residual is more sensitive to the change of the output, which is conducive to the spread of the network.

As illustrated in Fig. 2 the features obtained at four different scales are reused in up-sampling to obtain sufficient information through skip connections, because deconvolution is used to expand the size of the feature map and cannot create information. How-ever, the transmission of information between corresponding scales is not sufficient, the shallow features will not be used to restore deep features. In order to make full use of the features in each scale, we concatenate these feature maps to ensure the richness of information in each scale. Due to the different sizes of feature maps between different dimensions, we need to adjust them to the same size through maximum pooling and up-sampling. In the last step, we need to restore the prediction map（256×256） to its original size（512×512）. This part is completed by adding 0 around the prediction map, according to the boundary when the image is cropped.

3 Experiments

The data set used for training and testing is the Liver Tumour Segmentation Challenge (LiTs), which comes from clinical sites around the world [9]. The data set contains CT scans of 130 patients. All scans have been provided in nii format with an axial size of 512*512. For each patient, it contains hundreds of cross-sectional scans of the abdominal cavity and marked images provided by professional radiologists. The new algorithms were implemented in Anaconda with Python, running on PC with

32G RAM, 3.8GHz AMD Ryzen7 3800X 8-core CPU, and a NVIDIA RTX2080 GPU with 8GB memory.

The 130 patients provided are not all liver cancer patients. After simplifying the data set, we only retained CT scan images containing tumors, for a total of 73 patients. Considering the difference in tumour size, we require the ratio of the number of large tumours and small tumours to be 1:1 when selecting the data to ensure the generalization of the network. So, we select 60 patients for training and testing. The training set and test set contained 40 patients and 20 patients, respectively.

This work is implemented using Keras based on the TensorFlow backend. The first liver and tumour segmentation U-Net is trained for 60 epochs using Adam optimizer with learning rate 1e-4. For the second tumour segmentation network, we used data augmentation to improve the robustness of the network, including rotation (0.2 range), shift on horizontal (0.05) and vertical (0.05) direction, shear (0.05), zoom (0.05), and flip (horizontal). Similarly, we still use the Adam optimizer for the second network, but the learning rate is set to 1e-5 with 150 epochs.

Table 1. Dice score for five different tumour segmentation method

Author	Method	Dice per case
Dey, Hong [12]	Hybrid Cascaded Neural Network	0.681
He et al. [13]	Multi-scale Attention U-Net	0.596
Bellver et al. [14]	Fully Convolutional Network with bounding box sampling	0.590
Vorontsov et al. [8]	Joint U-Net	0.661
Proposed Method	Proposed U-Net with multi-scale feature fusion	0.665

Table 1 shows the dice scores of five different tumour segmentation methods. Our method ranks second, followed by the hybrid cascade neural network. We got a 0.665 average dice score based our method, which is better than most segmentation network. Compared with cascaded network, our network structure is more concise and lighter because of fewer network layers. And use the residual structure to reduce overfitting. Compared with the attention mechanism, it is more convenient to use multi-layer information fusion because there is no need to select each feature map, including the addition, multiplication, and extension of the feature map. Our result is not as good as hybrid cascade U-Net. It may be because we did not use 3d CNN and thus lost a certain amount of information in the vertical space. After all, there is a connection between slices, and the features extracted from a series of slices are much richer than the features obtained from an image. However, the computational cost is also huge.

Fig. 4 shows some visualization results of tumour segmentation from three patients. Our method has good performance for large, medium, and small tumours. Ordinary U-Net has a good segmentation ability for large and medium-sized liver tumours, we mainly discuss the segmentation performance of small tumours. Generally speaking, the network is very insensitive to small targets due to the serious class imbalance problem, such as Fig.4(c). Changing the loss function like dice loss or adding attention mechanism is the method to solve the class imbalance problem. However, the effects of these methods are not as obvious as directly reducing the size of the

target area. After our test, the accuracy of fine segmentation using the cropped image of the target area is about 12% higher than segmenting original scan directly.

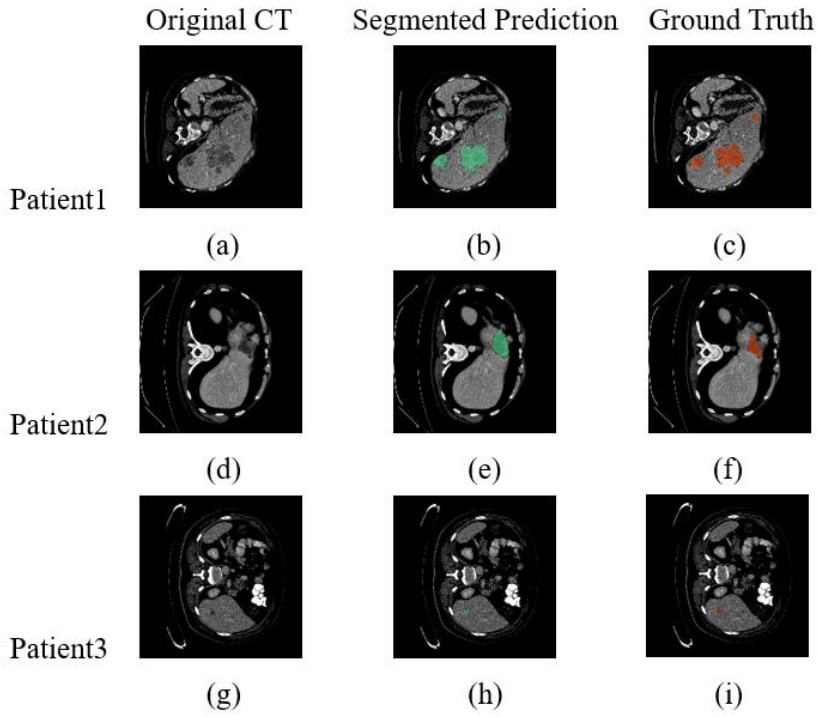

Fig. 4. Visualization result of three different sizes of tumours. The first column is original CT scans, the second column is our prediction result, and the third column is the ground truth.

4 Conclusion

In this paper, we presented a new method to segment liver tumours from CT scans, this method has good performance even for small tumours. We first use a U-net to lock the target area on the liver, and then used a second improved U-net to finely segment the tumour. Due to the reduced input image, the problem of class imbalance is im-proved. In addition, the use of residual structure and dense links makes information extraction more effective and sufficient. Multi-scale information fusion allows images to obtain more information when restoring features. This method can also be applied to CT image segmentation of other organs.

In the future, we will extend this method to a 3d convolutional network to enhance temporal features. It is also possible to introduce dilatated convolution while reducing the pooling layer to obtain high-level information.

Reference.

1. Sijens PE, Edens MA, Bakker SJ, Stolk RP. MRI-determined fat content of human liver, pancreas and kidney. World journal of gastroenterology: WJG. 2010;16(16):1993.
2. Gotra A, Sivakumaran L, Chartrand G, Vu K-N, Vandenbroucke-Menu F, Kauffmann C, et al. Liver segmentation: indications, techniques and future directions. Insights into imaging. 2017;8(4):377-92.
3. Shinde PP, Shah S, editors. A review of machine learning and deep learning applications. 2018 Fourth international conference on computing communication control and automation (ICCUBEA); 2018: IEEE.
4. Long J, Shelhamer E, Darrell T, editors. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.
5. Ben-Cohen A, Diamant I, Klang E, Amitai M, Greenspan H. Fully convolutional network for liver segmentation and lesions detection. Deep learning and data labeling for medical applications: Springer; 2016. p. 77-85.
6. Almotairi S, Kareem G, Aouf M, Almutairi B, Salem MA-M. Liver tumor segmentation in CT scans using modified SegNet. Sensors. 2020;20(5):1516.
7. Christ PF, Ettlinger F, Grün F, Elshaera MEA, Lipkova J, Schlecht S, et al. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv preprint arXiv:170205970. 2017.
8. Vorontsov E, Tang A, Pal C, Kadoury S, editors. Liver lesion segmentation informed by joint liver segmentation. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018: IEEE.
9. Bilic P, Christ PF, Vorontsov E, Chlebus G, Chen H, Dou Q, et al. The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:190104056. 2019.
10. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
11. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
12. Dey R, Hong Y, editors. Hybrid cascaded neural network for liver lesion segmentation. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020: IEEE.
13. He F, Zhang G, Yang H, Jiang Z, editors. Multi-scale Attention Module U-Net liver tumour segmentation method. Journal of Physics: Conference Series; 2020: IOP Publishing.
14. Bellver M, Maninis K-K, Pont-Tuset J, Giró-i-Nieto X, Torres J, Van Gool L. Detection-aided liver lesion segmentation using deep learning. arXiv preprint arXiv:171111069. 2017.