Triplanar Convolutional Neural Network for Automatic Liver and Tumor Image Segmentation
Zhenggang Wang and Guanling Wang*

College of Electrical Engineering, Anhui Polytechnic University, Wuhu, 241000, China

Abstract

The automatic image segmentation of liver and liver tumors is important in the diagnosis and treatment of hepatocellular carcinoma. A novel triplanar fully convolutional neural network (FCN) composed of three 2D convolutional neural networks (CNNs) is proposed to handle the issue. It performs segmentation through the transverse plane, coronal plane, and sagittal plane and can effectively use multi-dimensional features for 3D segmentation. Then, a cascaded structure is used to balance the positive and negative samples for segmentation of the tumor. The experimental results are obtained through data analysis and tested on the 3DIRCADb. They show that our method outperforms the existing methods and achieves a volume overlap error of 6.7% and 3.6% on the liver and tumors respectively.

Keywords: liver; liver tumors; automatic image segmentation; triplanar FCN; data analysis

(Submitted on September 10, 2018; Revised on October 16, 2018; Accepted on November 20, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

1. Introduction

Hepatocellular carcinoma is a common cancer that is responsible for many deaths each year [1]. Computed tomography (CT) image diagnosis can help doctors accurately evaluate the liver cancer and develop a treatment plan. In the general diagnosis process, the liver condition and liver lesions are determined by a radiologist according to the CT image slice by slice. This is highly time-consuming and also produces different results among different doctors. Therefore, the automatic segmentation of liver and liver tumors is of great value to clinical diagnosis.

Based on image processing technology, traditional liver segmentation mainly relies on some superficial features of the image, such as grayscale, statistical structure, and texture, to segment the liver contour. These superficial features, which can be obtained directly from the image or by a manually designed extraction operator, are less robust, less representative, and susceptible to noise interference [2]. It has been proven by practice that the abstract, deep features are often more representative. Deep learning technology can mine the deep abstract features of data from a large amount of data [3].

The liver and other neighboring organs have low intensity, so it is very challenging to automatically delineate the liver from CT images. As shown in Figure 1, radiologists usually use an injection protocol to enhance CT scans to clearly observe the tumor, and this may increase the image noise in the liver area. Liver tumor segmentation is a more challenging task compared to liver segmentation. First, liver tumors have different sizes, shapes, positions, and numbers in a patient, which are obstacles to automatic segmentation, as depicted in Figure 1. Second, there exist no clear boundaries in some tumors, which limits the performance of edge-based methods. Third, the accuracy of CT scans has a high degree of difference (the voxel range of 0.45mm to 6.0mm), further increasing the challenge of automatic segmentation.

To solve these difficulties, researchers need to effectively extract the image features. In recent years, the effect of FCNs on medical image segmentation has been very impressive. As far as the liver and liver tumor segmentation are concerned, many researchers use the deep learning methods, mainly divided into two categories: (1) 2D FCNs, such as U-Net architecture [4], multichannel FCN [5], and FCN on the basis of VGG-16 [6], (2) 3D FCNs, in which the 3D convolutions...
replace the 2D convolutions [7].

![Example of contrast-enhanced CT slices](image-url)

Figure 1. Examples of contrast-enhanced CT slices

However, 2D FCNs ignore the spatial information along the Z-axis, which might result in limited segmentation performance, or some adjacent slices are captured from a 3D image into a 2D FCN [5-6]. Despite the utilization of adjacent slices, it is still not enough to make full use of the spatial correlation of all third dimensions. This case may result in the degradation of segmentation accuracy. Compared with 2D FCNs, 3D FCNs require a large amount of computational resources (such as memory and computation time) [8]. 3D convolution, which consumes many computational resources, also hinders its application in training large-scale data sets. Faced with the above problems, we propose a novel 3D perspective based FCN, train FCNs in the three orthogonal planes of the 3D image to extract the features of the 2D image effectively, and then integrate the three models into the 3D spatial information. At the same time, we use a cascaded architecture to train the liver and liver tumor models and use the weighted cross entropy loss function to effectively handle the problem of sample balance. The proposed method achieves a better performance overall than the existing methods.

2. Related Work

CT technology can effectively solve the shortcomings of common X-ray imaging blur and poor tissue resolution. It can clearly visualize the structure of soft tissues and solve the imaging problem where certain parts are difficult for common X-rays to shoot. At the same time, the technological advancement of bulb detectors and the popularization of low-dose technology have greatly improved the accuracy and safety of clinical diagnosis. Because of the high resolution of CT imaging and the obvious imaging of soft tissues, the pathological organs can be highlighted. Furthermore, the costs are moderate. Therefore, CT technology is widely used in the diagnosis of liver diseases. As a metabolically active organ in the human body, the liver is the largest organ and the largest digestive gland in the human body and participates in a variety of physiological activities. Therefore, it is of great importance. It is vital to accurately diagnose liver disease as it is also very dangerous. Abdominal CT imaging is a common method for detecting abdominal organ lesions or soft tissue lesions, along with quantitative analysis of disease progression and three-dimensional visualization modeling.

Accurate and reliable segmentation of liver contour from abdominal CT images is the first step in the early diagnosis of liver disease. Estimation and three-dimensional modeling of the liver size and disease are also very critical steps. The segmentation results have a direct impact on follow-up work. In practical clinical applications, liver contours are manually segmented from CT images by physicians with relevant practical experience and expertise. However, this process is very time-consuming and energy-consuming. Moreover, different segmentation results may be achieved due to subjective factors, such as different physicians' experience and knowledge. Therefore, in order to reduce the workload of doctors, improve work efficiency, and obtain more objective and accurate segmentation results, it is necessary to introduce computer-aided diagnosis technology to help physicians segment liver CT images.

Over the past few decades, many algorithms, such as threshold based approaches, regional growth methods, deformable model based methods, and machine learning based approaches, have been proposed. Approaches based on the threshold classify the object and background in accordance with whether the intensity of an image pixel is higher than a certain threshold. Regional growth methods are also common in liver and lesion segmentation tasks. Researchers pay attention to the level set method on the advantages of numerical computations of the curve and the surface. Many machine learning based approaches, dealing with segmentation of liver tumors, have also been devised.

CNN has achieved great improvement in object recognition. Likewise, many researchers have proposed various methods based on FCNs for segmentation of liver and liver lesions. For example, an FCN solution [6] to liver segmentation and metastases has been proposed. At the same time, [5] designed an approach for segmenting liver tumors. The probability
3D based CNN algorithms [8-9] have also recently emerged, with the aim of extracting more discriminative volumetric features. For instance, [9] devised a 3D network for dual paths that uses dense inference techniques in image segments to overcome the computational burden. The 2D U-Net is expanded by 3D U-Net into 3D variation, consisting of an encoding path for extracting the features of the input image. The same resolution applied in skipping connections between layers was established by the encoding and decoding paths. V-Net is used to solve the imbalance of training data. VoxResNet was inspired by the idea of a 2D depth residual network [10] to construct a very deep 3D network.

3. The Proposed Algorithm

The overview of the network is shown in Figure 2. Slices are cropped according to three different planes of the CT volume, which are fed into three different FCNs. Then, the results of liver segmentation are combined. After that, the intensity of the image outside the liver is set to zero, and slices are cropped according to three different planes, which are fed into three different FCNs as well. Finally, the results of tumor segmentation are combined.

The architecture of FCN is shown in Figure 3. This is a cascaded-structure network that combines the liver segmentation model with the tumor segmentation model, and each model is fused together by an FCN of different 3D perspectives. The orange arrows present convolution with non-linear activation, the gray arrows show concatenation, the red arrows indicate down-sampling, and the green arrows indicate up-sampling. There is a total of four layers. The initial number of first layers in the feature map is 64, and the number is expanded to double in size layer by layer. The details of each part will be described in detail below.

The preprocessing of CT images mainly involves the following steps. The pipeline of data preprocessing is shown in Figure 4.

First, different hospitals may use different CT devices, and different CT devices will produce images in different orientations, including the fact that the liver is on the left or right, and the body is supine or prone. To reduce the data complexity, it is necessary to rotate the CT image up and down or left and right to a uniform orientation. In this paper, the body is regarded as supine with the liver on the left side as the standard orientation of the image. Second, in order to exclude
the effects of the existence of other organs or structures around the liver, the Hounsfield unit value of the CT image is set between -100 and 400 [11] to enhance the contrast between the normal liver, the liver tumor, and other abdominal tissues. Finally, to reduce the presence of noise in the data annotation, the area marked with the liver that is less than 128 pixels and the area marked with the liver tumor that is less than 32 pixels are removed.

The latter part of a classic CNN, such as [12] and [2], is usually a fully connected layer. Although it can effectively use the pixel spatial information, the calculation complexity is very large. [13] used the deconvolution layer instead of the fully connected layer and proposed a FCN. The first half of the FCN uses the convolution layer to obtain images, and the second half amplifies the feature map to match the input image.

$x^{l+1}$ and $x^l$ are set as the input and output of the $l^{th}$ layer, respectively. The output of the $(l+1)^{th}$ layer equals the input to the $l^{th}$ layer. The $j^{th}$ feature map of layer $l$ is denoted as $x_j^l$. The output is implemented in Formula (1).

$$x_j^l = \sigma \left( \sum_i a_{ij}^l * x_i^{l+1} + b_j^l \right) \quad (1)$$

Where $*$ denotes the convolution operation. $\sigma(\cdot)$ is a function that is nonlinear activation, e.g., Rectified Linear Units (ReLU) [14]. $a_{ij}^l$ is the kernel linking $i^{th}$ to $j^{th}$. Finally, the scalar $b_j^l$ is the parameter for the $j^{th}$ feature map output of the $l^{th}$ layer.

Pooling is used as the down-sampling layer, where the output of the $j^{th}$ feature map of layer $l$ is shown in Formula (2).

$$x_j^l = f \left( \beta_j^l * down \left( x_i^{l-1} \right) + b_j^l \right) \quad (2)$$

Where $\beta_j^l$ and $b_j^l$ are bias terms. $down(\cdot)$ is a down-sampling function that reduces the size of feature maps and enlarges the receptive field of convolution kernels. Deconvolution is used in the up-sampling layer, and it is similar to convolution except that it upscales the feature map before convolution. The output is given as Formula (3).

$$x_j^l = \sigma \left( \sum_i a_{ij}^l * up \left( x_i^{l-1} \right) + b_j^l \right) \quad (3)$$

Where $up(\cdot)$ is an up-sampling function.

Without the full connection layer, the FCN has no limit on the input image size and can classify pixels of the entire image at once. However, with an increase in convolution layers, the output results are relatively rough. [15] proposed U-Net to further improve the FCN. U-Net is divided into the encoding and decoding stage. The encoding stage can obtain the semantic image features. The decoding stage amplifies the feature map with convolution and deconvolution layers and integrates the feature maps of the coding and decoding stages to improve the accuracy of output. It achieves pixel-level segmentation of the entire image. The convolution layer will reduce the size of the image. For example, U-Net is fed with images in a size of $512 \times 512$, and the output is in a size of $388 \times 388$. Therefore, input needs to be padded with zero. In order to overcome this shortcoming, images are padded before each convolution layer, so the size of the image is constant.
during convolution. The network size of the input and output are exactly the same. The network is completely symmetrical and eliminates the restrictions on the input image size.

A cascaded CNN is used to achieve the liver and live tumor segmentation [11]. The two models are cascaded together. The prior model segments the liver from other tissues while the posterior model segments the tumor from the liver. In order to exclude the interference from other organs, the researchers use the results of the liver segmentation model to segment the image of the liver part of the initial image, because FCN does not have the constraint of input image size. In the segmentation of the liver tumor, all non-hepatic part are filled with 0 and then entered into the tumor segmentation model. The motivation of cascaded CNN is that CNN can learn a hierarchical representation from CT volumes. By cascading two CNN together, the researchers can make the prior model specific for separating the liver from other abdominal organs and the posterior model specific for separating lesions from normal liver tissue. The liver region from the liver model can also reduce false positives in the lesion model.

Previous works usually build 2D CNNs based on transverse planes [16-18]. Some models require an input of a few consecutive slices [19], which are called 2.5D CNN. They cannot fully use the spatial information in 3D CT images. Therefore, the researchers propose a triplanar FCN model. As shown in Figure 5, there are some examples of CT images in the transverse plane, coronal plane, and sagittal plane respectively. The liver and tumor exhibit different shapes and contours in different planes. In clinical routine, radiologists diagnose mainly based on the images in the transverse plane. They will refer to the images in the coronal plane and sagittal plane when it is necessary. Thus, three models are trained according to each plane. To leverage the spatial information, the results of the three models are integrated by majority voting. It is noted that the area of the liver in the abdomen is small as well as the area of the tumor in the liver. For class balance, the weighted entropy is used as the loss function, as calculated in Formula (4).

\[
L = -\frac{1}{n} \sum_{i=1}^{N} \omega_i \left[ y_i \log y_i + \left(1 - y_i\right) \log \left(1 - y_i\right) \right] \tag{4}
\]

Where \( y_i \) represents the probability that pixel \( i \) is part of the foreground and \( 1 - y_i \) denotes the ground truth.

Figure 5. The examples of CT images in the transverse plane, coronal plane, and sagittal plane respectively

4. Experiments

This section first introduces the datasets and experiment metrics and then explains the parameter setting of the experiment. The effect of the method mentioned in this paper is verified and compared with the existing method.

The experiment employs the public 3DIRCADb database. 3DIRCADb consists of 20 intravenous 3D enhanced CT samples, including ten females and ten males, of which 75% suffer from liver tumors. The data comes from a number of European hospitals with different types of CT scanners. In this experiment, 15 samples with liver tumors are selected to train and evaluate our model and make a 2-fold cross validation.

The metrics in [1] are adopted by experiment section to assess the liver segmentation performance. \( X \) is set as the foreground in the ground truth and \( Y \) as the predicted foreground.

DICE is the main quality metric and is evaluated as Formula (5).
\[DICE(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}\] (5)

The \textit{DICE} score ranges from 0 to 1.

The volume overlap error (\textit{VOE}) is defined in Formula (6), and it is also known as Jaccard [19].

\[\text{VOE}(X, Y) = 1 - \frac{|X \cap Y|}{|X \cup Y|}\] (6)

The relative volume difference (\textit{RVD}) is defined in Formula (7).

\[\text{RVD}(X, Y) = \frac{|Y| - |X|}{|X|}\] (7)

The average symmetric surface distance (\textit{ASD}) is presented in Formula (8).

\[\text{ASD}(X, Y) = \frac{\sum_{p \in S(X)} d(p, S(Y)) + \sum_{q \in S(Y)} d(q, S(X))}{|S(X)| + |S(Y)|}\] (8)

The maximum surface distance (\textit{MSSD}) is defined in Formula (9).

\[\text{MSSD}(X, Y) = \max \left\{ \max_{p \in S(X)} d(p, S(Y)), \max_{q \in S(Y)} d(q, S(X)) \right\}\] (9)

The model is implemented with Tensorflow [17] without pre-trained weights. When calculating the loss, the background’s weight is set to 1 while the weight of the foreground is set to 3. Meanwhile, during the liver tumor segmentation, the background’s weight is set to 1 while the weight of the foreground is set to 6. The Adam optimizer is used to assess the model based on an original learning rate of 0.0005. This model uses the same hyper-parameter settings on each plane.

To obtain improved performance of our proposed triplanar FCN, the metrics with ordinary FCN are compared on the 3DIRCADb dataset. First, three base models are trained based on the transverse plane, coronal plane, and sagittal plane, respectively called FCN YZ, FCN XZ, and FCN XY. The results of liver and lesion segmentation are illustrated respectively in Tables 1 and 2. The model based on the transverse plane obtains the best performance among the three models for both the liver and tumor. Then, these models are integrated into a triplanar FCN model. The integrated model achieves better results, where the VOE for the liver is reduced from 8.9% to 6.7% while the VOE for the lesion is reduced from 16.3% to 13.3%. Figure 6 depicts the original CT scans (first column), ground truth, outputs of the three base models (last three columns), and triplanar FCN (third column). It shows that there are few diversities among the outputs of the three base models due to the spatial information of their own plane. By integrating these models, the triplanar FCN can introduce more spatial information and improve the segmentation accuracy.

### Table 1. Segmentation performance results of liver for different parameters

<table>
<thead>
<tr>
<th>Approach</th>
<th>VOE [%]</th>
<th>RVD [%]</th>
<th>ASD [mm]</th>
<th>MSSD [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN YZ</td>
<td>10.8</td>
<td>4.5</td>
<td>4.2</td>
<td>16.7</td>
</tr>
<tr>
<td>FCN XZ</td>
<td>11.5</td>
<td>5.7</td>
<td>3.4</td>
<td>20.4</td>
</tr>
<tr>
<td>FCN XY</td>
<td>8.9</td>
<td>-2.3</td>
<td>1.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Triplanar FCN</td>
<td>6.7</td>
<td>3.5</td>
<td>0.7</td>
<td>6.5</td>
</tr>
</tbody>
</table>

### Table 2. Segmentation performance results of tumor for different parameters

<table>
<thead>
<tr>
<th>Approach</th>
<th>VOE [%]</th>
<th>RVD [%]</th>
<th>ASD [mm]</th>
<th>MSSD [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN YZ</td>
<td>22.8</td>
<td>-8.3</td>
<td>12.5</td>
<td>15.7</td>
</tr>
<tr>
<td>FCN XZ</td>
<td>25.7</td>
<td>13.8</td>
<td>11.4</td>
<td>19.9</td>
</tr>
<tr>
<td>FCN XY</td>
<td>16.3</td>
<td>-3.3</td>
<td>2.8</td>
<td>8.9</td>
</tr>
<tr>
<td>Triplanar FCN</td>
<td>13.3</td>
<td>6.3</td>
<td>1.8</td>
<td>7.5</td>
</tr>
</tbody>
</table>
Figure 6. Examples of segmentation results

For comparison with advanced methods, other methods’ results in the liver and tumor segmentation are also presented, such as the likelihood and local constraint (LLC) level set model [20], deformable graph cut [21], and other FCNs [5,11]. The results of segmentation performance of the liver and tumor are illustrated in Tables 3 and 4 respectively. They show that our method achieves good results on 3DIRCADb and excels compared to the other two algorithms in terms of the VOE, ASD, and MSSD on the liver and the VOE and ASD on tumors. The RVD value of the triplanar FCN is 3.5% on the liver, which indicates that it tends to be slightly over-segmented. The RVD and MSSD values of the triplanar FCN are 6.3% and 7.5 mm respectively on tumors. They are close to the values of the other two methods, which means the triplanar FCN has similar stability but produces an improved segmentation. This demonstrates the superiority of the triplanar FCN in dealing with the segmentation of the liver and tumors.

Table 3. Segmentation performance results of liver

<table>
<thead>
<tr>
<th>Approach</th>
<th>VOE [%]</th>
<th>RVD [%]</th>
<th>ASD [mm]</th>
<th>MSSD [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schreier [4]</td>
<td>10.7</td>
<td>-1.4</td>
<td>1.5</td>
<td>24.0</td>
</tr>
<tr>
<td>Li et al. [18]</td>
<td>9.2</td>
<td>-11.2</td>
<td>1.6</td>
<td>28.2</td>
</tr>
<tr>
<td>Triplanar FCN</td>
<td>6.7</td>
<td>3.5</td>
<td>0.7</td>
<td>6.5</td>
</tr>
</tbody>
</table>

Table 4. Segmentation performance results of tumor

<table>
<thead>
<tr>
<th>Approach</th>
<th>VOE [%]</th>
<th>RVD [%]</th>
<th>ASD [mm]</th>
<th>MSSD [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li et al. [11]</td>
<td>14.4</td>
<td>8.1</td>
<td>2.4</td>
<td>7.2</td>
</tr>
<tr>
<td>Zikic et al. [19]</td>
<td>15.6</td>
<td>5.8</td>
<td>2.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Triplanar FCN</td>
<td>13.3</td>
<td>6.3</td>
<td>1.8</td>
<td>7.5</td>
</tr>
</tbody>
</table>

5. Conclusion

Liver tumors are characteristic of high malignancy and rapid disease development. The mortality rate of liver tumors ranks second in the world while China is one of the highest incidence areas of liver cancer. CT has been extensively applied in diagnosis, especially for liver diseases, and it has become the preferred method for diagnosing liver diseases. The use of computer imaging technology, combined with medical imaging diagnostic technology for early diagnosis, three-dimensional modeling, and quantitative analysis of liver diseases, enables doctors to grasp sufficient data before surgery, make preoperative planning, improve the success rate of surgery, and formulate reasonable and effective treatment plans.

In this work, the researchers devise a triplanar FCN model solution for liver and tumors segmentation. By fusing the 2D FCN trained on the transverse plane, coronal plane, and sagittal plane, the spatial information of 3D enhanced CT images can be effectively utilized. It also can avoid the huge computational complexity of 3D convolution networks. In further works, the researchers of this paper would like to explore how they can integrate the three perspectives of the model more effectively, because it is still time-consuming to integrate three models. It could possibly be made more adaptive and trainable. The researchers will also apply the model to other types of segmentation.
Acknowledgements

This study was supported by the Program of Scientific Research of Anhui Province, China (No. TSKJ2017B26) and the Open Research Fund of Wanjiang Collaborative Innovation Center for High-end Manufacturing Equipment (No. GCKJ2018007).

References

8. S. Iqbal, M. U. Ghani, T. Saba, and A. Rehman, “Brain Tumor Segmentation in Multi-Spectral MRI Using Convolutional Neural Networks (CNN),” Microscopy Research and Technique, Vol. 81, No. 4, pp. 419-427, 2018