Expert Systems With Applications 126 (2019) 54–63

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Liver CT sequence segmentation based with improved U-Net and graph cut

Zhe Liu[a,*], Yu-Qing Song[a], Victor S. Sheng[b,*], Liangmin Wang[a], Rui Jiang[a], Xiaolin Zhang[a], Deqi Yuan[c]

[a] School of Computer Science and Telecommunication, Jiangsu University, Zhenjiang, Jiangsu, PR China
[b] Department of Computer Science, University of Central Arkansas, Conway, Arkansas, USA
[c] Zhenjiang First People's Hospital Branch, Zhenjiang, Jiangsu, PR China

ARTICLE INFO

Article history:
Received 22 September 2018
Revised 19 January 2019
Accepted 20 January 2019
Available online 27 January 2019

Keywords:
Liver segmentation
Deep learning
Graph cut
U-Net
GIU-Net

ABSTRACT

Liver segmentation has always been the focus of researchers because it plays an important role in medical diagnosis. However, under the condition of low contrast between a liver and surrounding organs and tissues, CT image noise and the large difference between the liver shapes of patients, existing liver image segmentation algorithms are difficult to obtain satisfactory results. To improve this situation, we propose a liver CT sequesnce image segmentation algorithm GIU-Net, which combines an improved U-Net neural network model with graph cutting. Specifically, we initially segment a liver from a liver CT sequence using an improved U-Net and obtain the probability distribution map of the liver regions. Secondly, the sequence segmentation start slice is selected, and then the context information of the liver sequence images and the liver probability distribution map are used to construct a graph cut energy function. Finally, the segmentation is done by minimizing the graph cut energy function. Our experimental results show that GIU-Net has a good performance when segmenting liver sequence images in terms of segmentation accuracy and robustness.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Liver cancer is one of the six most common cancers in the world. It is the leading cause of cancer death. Segmentation of liver images is very important for liver disease diagnosis, functional assessment and treatment. However, the liver segmentation is still a challenging task in the field of medical image processing because of the complexity of the liver anatomy, the low contrast and the morbid condition of its adjacent organs. Computed tomography (CT) is now widely used in the noninvasive diagnosis of liver diseases. However, in CT images, the shape and the size of livers vary from person to person because of different slice positions, which make liver automatic segmentation difficult.

The methods of liver segmentation can be divided into three categories: manual, semi-automatic, and fully automatic (Gotra et al., 2017; Heimann, Meinzer, and Wolf, 2007). Manual segmentation relies entirely on doctors. The manual outline of a desired organ on each slice is highly demanding and laborious, and results are directly affected by doctors' professional knowledge and physical conditions. However, only these segmentation results can be used as ground truths for other algorithm evaluations.

Compared with manual segmentation, semi-automatic segmentation methods have less dependence on humans. However, when using semi-automatic algorithms to segment livers, it is still necessary to provide seed points or seed regions for background and foreground respectively, such as region growth (Lu, Wu, Ren, Zhang, and Li, 2014; Zhang, Xiong, and Kuang, 2015), active contour modelling (Han, Liu, Meng, and Wang, 2015; Ji, He, and Yang, 2014; Peng, Wang, and Kong, 2014), graph cut (Afifi and Nakaguchi, 2012; Chen et al., 2012; M. Liao et al., 2016).

Fully automatic segmentation methods can segment livers with few human interventions. Convolution neural network is currently the most widely used method. FCN (Full Convolution Neural Network) (Long, Shelhamer, and Darrell, 2015) and its variants are used for both 2D segmentation (in short 2D-FCN) (Ben-Cohen, Diamant, Klang, Amitai, and Greenspan, 2016; Chen, Cao, Cao, Yang, and Zhang, 2018; Christ et al., 2017; Kaluva, Khened, Kori, and Krishnamurthi, 2018) and 3D segmentation (in short 3D-FCN) (Li et al., 2018; Lu, Wu, Hu, Peng, and Kong, 2017; Rafiei et al., 2018). The 2D-FCN takes each CT image as an input to obtain the seg-

* Corresponding authors.
E-mail addresses: 1000004088@ujs.edu.cn (Z. Liu), yqsong@ujs.edu.com (Y.-Q. Song), ssheng@uca.edu (V.S. Sheng), wanglm@ujs.edu.cn (L. Wang), 2211708011@stmail.ujs.edu.cn (R. Jiang), 2211608038@stmail.ujs.edu.cn (X. Zhang), 1627377803@qq.com (D. Yuan).

https://doi.org/10.1016/j.eswa.2019.01.055
0957-4174/© 2019 Elsevier Ltd. All rights reserved.

mentation. 3D-FCN is limited by computation ability and memory problems. In most cases, 3D-FCN uses three adjacent slices to form a three-channel image as input rather than the entire sequence images of a patient (Li et al., 2018). In this condition, spatial information along the third dimension of medical images is considered. However, it still is only limited to three adjacent slices, rather than the entire CT sequence. As for cropping images, it is not a good way, because the liver of different people at different ages can locate at a different area of their abdomen. Some people's liver even crosses their whole abdomen. Cropping images may loss some areas of a liver. On the other hand, the sagittal and the coronal plane are greatly affected by the slice thickness of CT images. If an image dataset doesn't contain different slice thicknesses (such as 1 mm, 2.5 mm, 5 mm), the generalization ability of a model will be poor on images whose slice thickness is different from the thickness of training images. Although, 2D convolution neural networks may leverage the spatial information along the third dimension, it is not affected by the slice thickness and can get the similar accuracy of 3D network without too much memory.

In this paper, we propose a new framework named IU-Net, and further propose a novel method GIU-Net (based on IU-Net) to improve and optimize the liver segmentation for CT images on a basis of the size of input, output and so on. A probability distribution map of a liver region is obtained by pre-segmentation with an improved U-Net. After that, a starting slice of a liver sequence is selected, and then the context information of the liver sequence and the liver probability distribution map are used to construct a graph cut energy function. Finally, segmentation is done by minimizing the graph cut energy function.

The main contributions of our works:

1. We proposed a new framework named IU-Net. We have increased the depth of the U-Net to get more advanced semantic features which can help get better segmentation results. When come to skip-connections, we do not copy the activation features after convolutions, we just copy the pooling layer's features, so we can reduce the loss of information.
2. We use graph cut to refine the segmentation results of IU-Net, which makes the segmentation result's boundary smoother and remove some extra part.

The remaining of this paper is organized as follows. In Section 2, a brief review of graph cut algorithms is presented. Section 3 explains our proposed method GIU-Net. Section 4 describes our experiments on evaluating the performance of GIU-Net by comparing with state-of-the-art methods. Conclusions are drawn in Section 5.

2. Graph cut

Graph cut is a supervised interactive segmentation approach, which is first proposed by Boykov and Funka-Lea (2006), Boykov and Kolmogorov (2004), Boykov and Veksler (2006), Boykov, Veksler, and Zabih (2001). It transforms the image segmentation problem into solving a minimized energy function. It has been widely applied in the field of computer vision because it has high segmentation efficiency and can find the globally optimal segmentation of N-dimensional images.

Actually, graph cut is a semi-supervised segmentation algorithm, which needs human interactions. Specifically, in the process of segmentation, a user must select some pixels as foreground and background seed points, which will guide following segmentation. It is generally used to segment the foreground from the background. Two vertices (a source point and a sink point) represented by S and T, respectively are added to the weighted undirected graph discussed above. These two points are fully connected with all the other vertices. There are two kinds of vertices and two

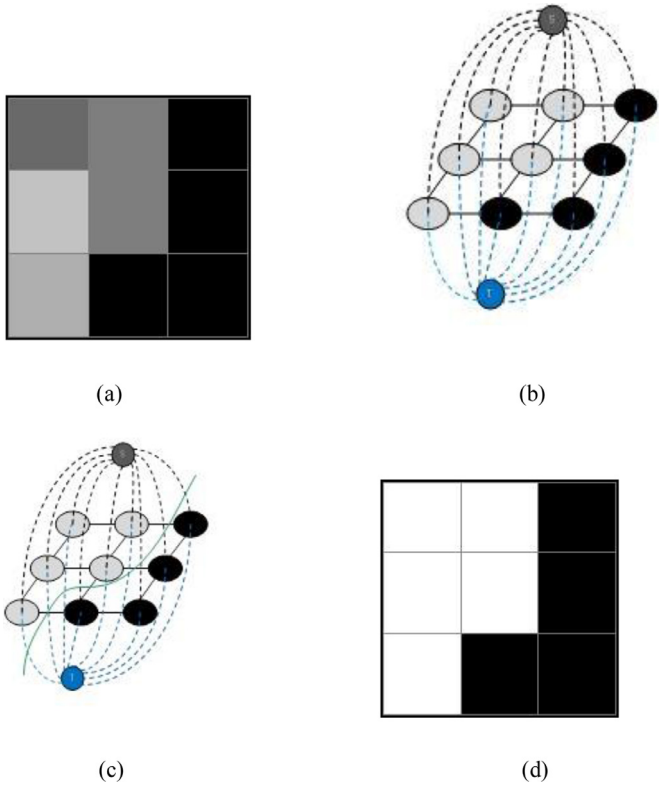


Fig. 1. A diagram of segmentation based on graph cut.

kinds of edges in graph cut. The first kind of vertices correspond to an image mapping to a graph in graph theory, and the second kind is the source point S and the sink point T. The first kind of edges called n-links is interconnections between vertices, and represents the similarity of adjacent pixels. The second kind of edges called t-links is the edges of all vertices connecting to the source point S and the sink point T when the image is mapped to graphs and represents the similarity between the image pixel points and the front and background points. Hence, the graph constructed by graph cut is greatly different from the ordinary one. The segmentation process of graph cut is shown in Fig. 1, where solid lines are n-links, dotted lines are t-links, and the source and sink points have been marked with S and T. The target of graph cut is to find a secant line, which makes the weights of t-link and n-link have the minimum summation. After segmentation, the graph is divided into two parts. One part is connected to the source S and the other part is only connected with the sink T. Thus, the corresponding image segmentation is completed. In Fig. 1(c), the green line is a secant, and Fig. (d) represents the final segmentation result.

According to the human interaction information and the feature information of an image itself, graph cut constructs an energy function and obtains the optimal segmentation result by minimizing the energy function. Ford and Fulkerson have proved that the maximum flow problem solving the network is the same as the minimum cut problem (Dantzig and Fulkerson, 2003). The final segmentation can be accomplished after the minimum cut of the weighted graph is obtained using a maximum flow algorithm.

Graph cut, which takes the region and the boundary information of images into account, has the advantages of global optimization, high segmentation accuracy and strong robustness. It can achieve great satisfactory results when facing more complex images after designing appropriate regional and boundary energy functions for different application scenes and carrying out multiple interaction artificially.

3. Our proposed method GIU-Net

In this section, we will explain our liver sequence CT image segmentation solution GIU-Net, which combines an improved U-Net and a graph cutting algorithm, to solve the problem of low contrast between a liver and its surrounding organs and the problem of the large difference among individual livers in CT image segmentation. In the first place, a probability distribution map of a liver region is obtained by performing image pre-segmentation with an improved U-Net (IU-Net), and then a graph cut energy function is constructed by integrating graph cut with the liver region probability distribution map and the context information of the sequence images. Finally, the segmentation will be completed by minimizing the graph cut energy function.

Five specific steps of our solution GIU-Net are as follows:

1. Using Hounsfield unit (HU) scale and denoising preprocesses an original data set.
2. Select some of the preprocessed data set as a training set to train an improved U-Net model (i.e., IU-Net). We will explain IU-Net in details later.
3. Perform an initial segmentation on the test set to get the probability distribution map of the liver regions with the trained IU-Net model.
4. Select a slice with the largest liver region in the liver CT sequence as an initial slice, and then construct the graph cut energy function of the initial slice by using the output probability distribution graph obtained from IU-Net and Formula (1), which will be defined in Section 3.2, and finally complete the initial liver slice segmentation according to the maximum flow minimum cut algorithm.
5. Starting from the initial slice, construct the graph cut energy function according to the sequence image context information and the output probability distribution graph of IU-Net, perform segmentation from the initial slice to both ends until it is arrived at the beginning slice and the end slice of its sequence.

The whole procedure of our GIU-Net is shown in Fig. 2:

3.1. An improved U-Net (IU-Net)

Let us first explain the improved U-Net (IU-Net). U-Net was first proposed and applied to cell image segmentation by Ronneberger, Fischer, and Brox (2015). It is a kind of Full Convolution Neural Network. The major difference between U-Net and FCN is that its network up-sampling process replicates the feature maps of its corresponding down-sampling process (i.e., pooling) through skip-connection operations, and combines with the corresponding down-sampling process together (called as a concat layer) to include the image pixel context information to the up-sampling convolution process. It can achieve a higher segmentation accuracy on a smaller training set. It is called U-Net due to the elegant U type formed by the symmetrical structure of the down-sampling and the up-sampling in the network. Fig. 3 shows an improved U-Net (IU-Net) model, which is modified from the original U-Net and used in our later experiments.

Fig. 3 shows that the IU-Net framework consists of 15 convolution layer groups, 7 pool (i.e., down-sampling) layers and 7 up-sampling layers. Each convolution block contains two convolution layers. Both convolution layers and up-sampling layers use ReLU as the activation function. The dimensions of each layer is $1 \times 64 \times 512 \times 512$ (batch-size $\times$ channel number $\times$ image height $\times$ image width) in the first convolution layer group, $1 \times 128 \times 256 \times 256$ in the second group, $1 \times 128 \times 128 \times 128$ in the third group, $1 \times 256 \times 64 \times 64$ in the fourth group,$1 \times 256 \times 32 \times 32$ in the fifth group, $1 \times 512 \times 16 \times 16$ in the sixth group, $1 \times 512 \times 8 \times 8$ in the seventh group and $1 \times 1024 \times 4 \times 4$ in the eighth group. The convolution kernel size of all convolution layers is 3×3, and both the stride and the pad is 1. In up-sampling, the convolution layer dimension, the convolution kernel size, the stride and the pad are kept the same as the settings in the front corresponding layers. In the pooling layer, the maximum pooling is adopted, in which the convolution kernel size is 2×2, the stride is 2 and the pad is 0.

In the process of convolution, the feature map of the pool layer is reused. U-Net replicates the features of the pooling layer after two convolutions and the ReLU function activation, while IU-Net directly copies the characteristics of the pooling layer. Therefore, the IU-Net information is lost less. Besides, IU-Net is deeper than U-Net, so it can get more advanced semantic features. In the training process, the loss is calculated by the SoftmaxWithLoss layer and the segmentation accuracy rate is observed in the user-defined Dice layer. The structure and the depth of the network layer are determined by the experience gained after many experiments. The reason for the convolution layer to use the 3×3 convolution kernel is to maintain the segmentation accuracy when minimize the complexity of the neural network (Szegedy, Vanhoucke, Ioffe, Shlens, and Wojna, 2016), meanwhile, to solve the problem of image resolution changes after the convolution with the pad being set to 1.

3.2. Graph cut energy function construction

IU-Net finally outputs two probability distribution maps by a softmax classifier, which indicates the probabilities of a pixelI_{ij} belonging to the foreground and background, denoted as p_{ij} and q_{ij} respectively (i.e.,$p_{ij} + q_{ij} = 1$). Using the fixed threshold 0.5 to distinguish foreground and background directly is a general solution at present. With the foreground and background predictions, we can obtain a rough segmentation result. In particular, it is inaccurate to get the segmentation result using a threshold in boundary blurred regions (Dou et al., 2016). IU-Net may produce a low quality segmentation when dealing with more complicated slices of livers. However, please note that the segmentation objects of our method GIU-Net is the CT image sequence of a liver. The location and the size of the liver have no obvious changes in the CT image sequence. Therefore, graph cut as a post-processing method can make segmentation more precise and stable by adding the context information of the CT liver image sequence to IU-Net.

GIU-Net uses a graph cut algorithm to perform the segmentation by globally minimizing the constructed energy function to solve the problems discussed above. In order to integrate the sequence information of a liver, it is necessary to ensure a high segmentation accuracy for the first image slice in the sequence. Based on our simple observation: the larger a liver region in an image, the more efficient result can be obtained by GIU-Net. Therefore, GIU-Net selects the image with the largest liver area as the Initial segmentation image. From the selected initial slice to both ends, GIU-Net carries out the segmentation.

The graph cut energy function of the initial slice is shown in Formula (1).

$$E(f) = \lambda R(f_p) + B(f) \tag{1}$$

In this function, $R(f_p)$ represents the regional term and $B(f)$ represents the boundary term. λ represents the equilibrium parameter of the regional term, comparing to the boundary term.

The regional term of the graph cut energy function here is mainly used to smooth the contour of a liver region and has little effect on the segmentation. We are inspired by the literature (Song et al., 2013) to define the regional term energy function $R(f_p)$ as Formula (2), because it produces great results on dealing with the problems of the blurred boundary of lung tumors and the un-

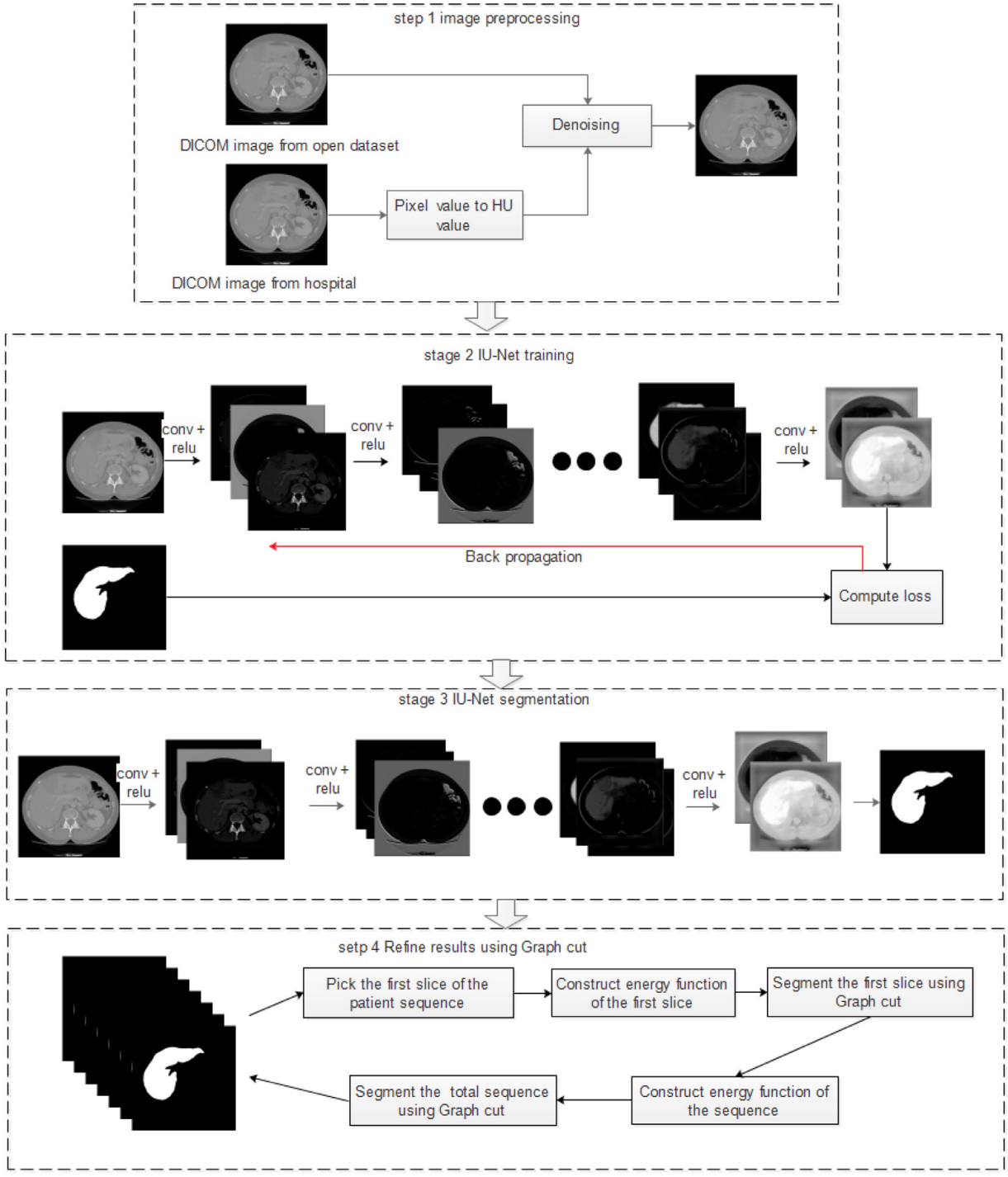


Fig. 2. The flow chart of our proposed GIU-Net.

even distribution of the luminance.

$$R_{p_{ij}\in P}(f_{p_{ij}}) = \begin{cases} 1 - \dfrac{1}{1+\exp\left(-\frac{p_{ij}-\mu}{\gamma}\right)}, p_{ij} = \text{“}obj\text{”} \\ \dfrac{1}{1+\exp\left(-\frac{p_{ij}-\mu}{\gamma}\right)}, p_{ij} = \text{“}bkg\text{”} \end{cases} \tag{2}$$

where p_{ij} represents the probability of each pixel point in the probability graph P output by IU-Net. The parameters μ and γ control the central point and the curvature, which are set as 0.4 and 0.05 respectively in our later experiments.

The boundary term $B(f)$ is defined using a classical boundary function (Boykov et al., 2001) as follows.

$$B(f_{\{p,q\}}) \propto \exp\left(-\frac{(p-q)^2}{2\sigma^2}\right) \cdot \frac{1}{dist(p,q)} \tag{3}$$

where $p$and q represent two adjacent pixels in the probability distribution map. As we mentioned before, the location and the size of a liver in adjacent slices don't change much, these types of context information can be used to rebuild the sequence energy function. Therefore, we have a novel energy function, which is shown

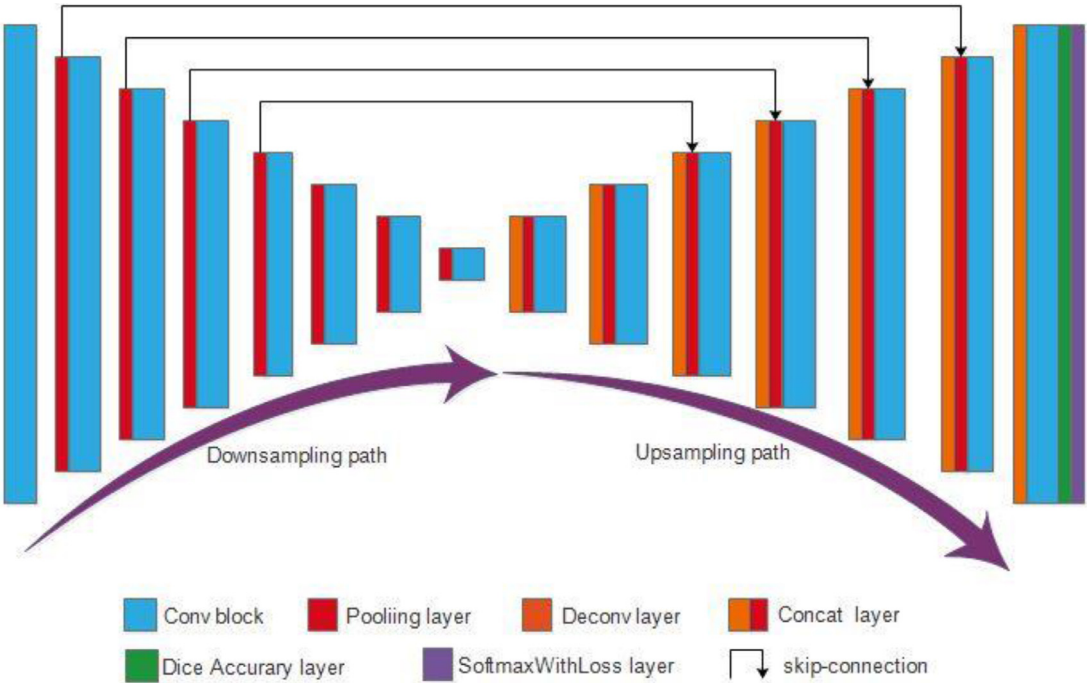


Fig. 3. The architecture of IU-Net.

as Formula (4).

$$E(f) = \lambda(R(f_P) + R(f_I)) \cdot F(f) + B(f) \tag{4}$$

where, $R(f_P)$ and $R(f_I)$ represent the probability regional term function and the section energy function of a slice image, respectively. The probability regional term function $R(f_P)$ can be defined by the probability distribution map information output from IU-Net, and the section energy function of the slice $R(f_I)$ can be defined by its gray information. To improve the performance of segmentation, avoiding over-segmentation or under-segmentation of small liver region of image slices, the gray information of slices should be added into the definition of the regional term in Formula (1), which is defined as follows.

$$R(f_I) = \begin{cases} 1 - \exp\left(\dfrac{(I_i - \bar{I}_R)^2}{\sigma_R^2}\right), I_i = \text{“}obj\text{”} \\ \exp\left(-\dfrac{(I_i - \bar{I}_R)^2}{\sigma_R^2}\right), I_i = \text{“}bkj\text{”} \end{cases} \tag{5}$$

Since the location of a liver has been restricted by the position constraint function, the energy function of the newly added slice area term is designed according to the Gauss mixture model. $\bar{I}_R$ and σ_R represent the brightness mean and the standard deviation of all the pixels in the liver priori area of the CT slices respectively.

$F(f)$ represents a position constraint function mainly established based on a priori condition that the liver position of adjacent slices will not change significantly, which can be defined as Formula (6).

$$F(f_i) = \begin{cases} 1 - \exp(-\mu D(I_i)), I_i = \text{“}obj\text{”} \\ \exp(-\mu D(I_i)), I_i = \text{“}bkg\text{”} \end{cases} \tag{6}$$

where $D(I_i)$ is the shortest Euclidean distance between pixel I_i and the edge of the prior region of a liver, which is mapped from adjacent slices of a liver. In order to reflect the difference of the liver shapes between adjacent slices, a direct mapping area is corroded by an ellipse with a long axis 3 and a short axis 1. For each pixel I_i in the liver priori area, $D(I_i)$ is 0.μ plays a role of controlling the search range of a liver region, which would cause liver over-segmentation or under-segmentation when too large or too small. Hence, the value of μ needs to be determined by multiple experiments based on the prior work proposed in Liao et al. (2016), which showed the value of μ could be any value inside the interval [0.02, 0.05]. Therefore, we set μ=0.05 in our later experiments.

4. Experiments

In this section, we will conduct experiments to evaluate our solution GIU-Net by comparing with existing solutions in terms of popular evaluation criteria and evaluation indexes, and analyze our experimental results. Specifically, we will compare GIU-Net and IU-Net with FCN-8 s, U-Net, three 2D-FCN methods (Ben-Cohen et al., 2016; Christ et al., 2017; Kaluva et al., 2018), and three 3D CNN methods (Li et al., 2018; Lu et al., 2017; Rafiei et al., 2018).

Ben-Cohen et al. (2016) used three adjacent CT image slices to form a three-channel image service as the input of FCN-8 s to segment livers. Christ et al. (2017) used two cascaded U-Nets to segment livers ant liver tumors. Kaluva et al. (2018) used 2D-Densely connected convolution neural networks to finish liver and liver tumor segmentation. Lu et al. (2017) used 3D convolution neural networks to do coarse liver segmentation first, and then used graph cut to refine their results. Li et al. (2018) proposed a method namely H-DenseUNet, which combines 2D DenseUNet with 3D DenseUNet to segment livers and liver tumors. Rafiei et al. (2018) also used U-Net-like neural network to segment livers, but they used 3D convolution in its down-sample path and 3D convolution in its up-sample path. To make quick comparisons, the segmentation results of the three 2D-FCNs and the three 3D-CNNs are taken from their corresponding article and shown in Table 1.

4.1. Experimental data and environment

One data set used in our experiment is codalab, which is the liver segmentation competition data set and contains 131 sequences. Codalab has three phase ground truths of the liver CT

Table 1
Quantitative comparison among different methods.

Methods	DSC (%)	VOE (%)	RVD (%)	ASD (mm)	RMSD (mm)	MSD (mm)
FCN-8s	89.36 ± 3.45	18.66 ± 5.04	−1.25 ± 10.93	6.19 ± 2.32	12.49 ± 6.16	52.48 ± 12.42
U-Net	82.76 ± 4.52	25.24 ± 6.07	−2.56 ± 11.04	10.25 ± 3.12	16.72 ± 6.18	82.78 ± 42.14
2D-FCN2 (Ben-Cohen et al., 2016)	89.0	–	–	–	–	–
2D-FCN1 (Christ et al., 2017)	93.1	12.8	−3.3	2.3	–	46.7
2D-dense-FCN (Kaluva et al., 2018)	92.3	15.0	−0.8	6.465	9.682	45.928
3D-FCN (Lu et al., 2017)	–	9.36 ± 3.34	0.97 ± 3.26	1.89 ± 1.08	4.15 ± 3.16	33.14 ± 16.36
H-DenseUNet[a] (Li et al., 2018)	96.5	7.4	1.8	1.450	–	27.118
3D U-Net (Rafiei et al., 2018)	92.8	–	–	–	–	–
IU-Net	93.75 ± 2.65	9.49 ± 4.33	0.64 ± 7.18	3.22 ± 2.48	5.49 ± 3.46	30.45 ± 23.39
GIU-Net	95.05 ± 1.32	7.07 ± 2.37	−1.22 ± 4.51	1.80 ± 0.71	3.64 ± 1.44	20.25 ± 5.34

Note: As we mentioned before, we directly took the experimental results from the corresponding article of the six comparison methods with citations (from row 4 to 9). Missing values and missing standard deviations are because they are not available in the original article.
[a] Except the DCS value, its other results are obtained from https://competitions.codalab.org/competitions/17094#results.

images. The CT format is enhanced CT, and the resolution is 512 × 512. Another data set is from Riverside hospital with corresponding liver ground truths. The input of both IU-Net and U-Net is the original images of each data set. FCN-8s uses its original frame structure, so the size of training and testing images is cut to 500 × 500.

In our experiments, we used Ubuntu 16.04 with i7 6700 K (CPU) NVIDIA GT × 1080 (GPU), 32 G memory. We used Caffe and matlab2015a to implement corresponding algorithms. There are 111 sequences from codalab extracted as the training set to train models and the rest are used as the testing set. The Voxel size interval of each CT image in the testing set of codalab is 0.64–0.84 mm.

4.2. Evaluation indexes

Dice's Similarity Coefficient (DSC) (Dice, 1945; Heimann et al., 2009) is used to evaluate the performance of different segmentation methods in our experiments. DSC calculates the space coincidence degree between segmentation results and ground truths. Besides DSC, we also have the experimental results on five segmentation evaluation indexes provided by MICCAI, such as Volumetric Overlap Error (VOE), Relative Volume Difference (RVD), Average Symmetric Surface Distance (ASD), Root Mean Square Symmetric Surface Distance (RMSD) and Maximum Symmetric Surface Distance (MSD) (Liao et al., 2016). The definitions of DSC and the five segmentation evaluation indexes are shown as follows.

DSC is defined as follows.

$$DSC(U_1, U_2) = \frac{2|U_1 \cap U_2|}{|U_1| + |U_2|} \tag{7}$$

where U_1 and U_2 represent the segmentation results and the corresponding ground truth respectively. The higher the value of DSC, the better the segmentation effect is. The ideal DSC value is 1.

The five segmentation evaluation indexes provided by MICCAI are defined as follows.

$$VOE(U_1, U_2) = 100 \times \left(1 - \frac{|U_1 \cap U_2|}{|U_1 \cup U_2|}\right) \tag{8}$$

$$RVD(U_1, U_2) = 100 \times \left(\frac{|U_1| - |U_2|}{|U_2|}\right) \tag{9}$$

$$ASD(U_1, U_2) = \frac{1}{|S(U_1)| + |S(U_2)|}\left(\sum_{s_{U_1} \in S(U_1)} d(s_{U_1}, S(U_2)) + \sum_{s_{U_2} \in S(U_2)} d(s_{U_2}, S(U_1))\right) \tag{10}$$

$$RMSD(U_1, U_2) = \sqrt{\frac{1}{|S(U_1)|+|S(U_2)|}} \times \sqrt{\sum_{s_{U_1} \in S(U_1)} d^2(s_{U_1}, S(U_2)) + \sum_{s_{U_2} \in S(U_2)} d^2(s_{U_2}, S(U_1))} \tag{11}$$

$$MSD(U_1, U_2) = \max\{\max_{s_{U_1} \in S(U_1)} d(s_{U_1}, S(U_2)), \max_{s_{U_2} \in S(U_2)} d(s_{U_2}, S(U_1))\} \tag{12}$$

Again, U_1 andU_2represent the segmentation results and the corresponding ground truth respectively. $S(U_1)$ and $S(U_2)$ represent the outline of the segmented liver region and of the corresponding ground truth liver region respectively. $d(v, S(U_1))$ represents the shortest distance between any pixel v and $S(U_1)$. That is, $d(v, S(U_1)) = \min_{s_{U_1} \in S(U_1)} ||v - s_{U_1}||$, where $||\bullet||$ is an Euclidean distance. Different from DSC, the lower the four segmentation indexes (VOE, ASD, RMSD, and MSD) defined above, the smaller the value of the index, the better the segmentation result is. Since the index RVD could be negative (under-segmentation), the absolute value of RVD shows the segmentation quality. The smaller the absolute value of RSD, the better the segmentation result is. In general, the five segmentation indexes are 0 for the perfect segmentation.

4.3. Parameter setting

The main parameters in GIU-Net, IU-Net, U-Net, and FCN-8 s are set as follows in our experiments.

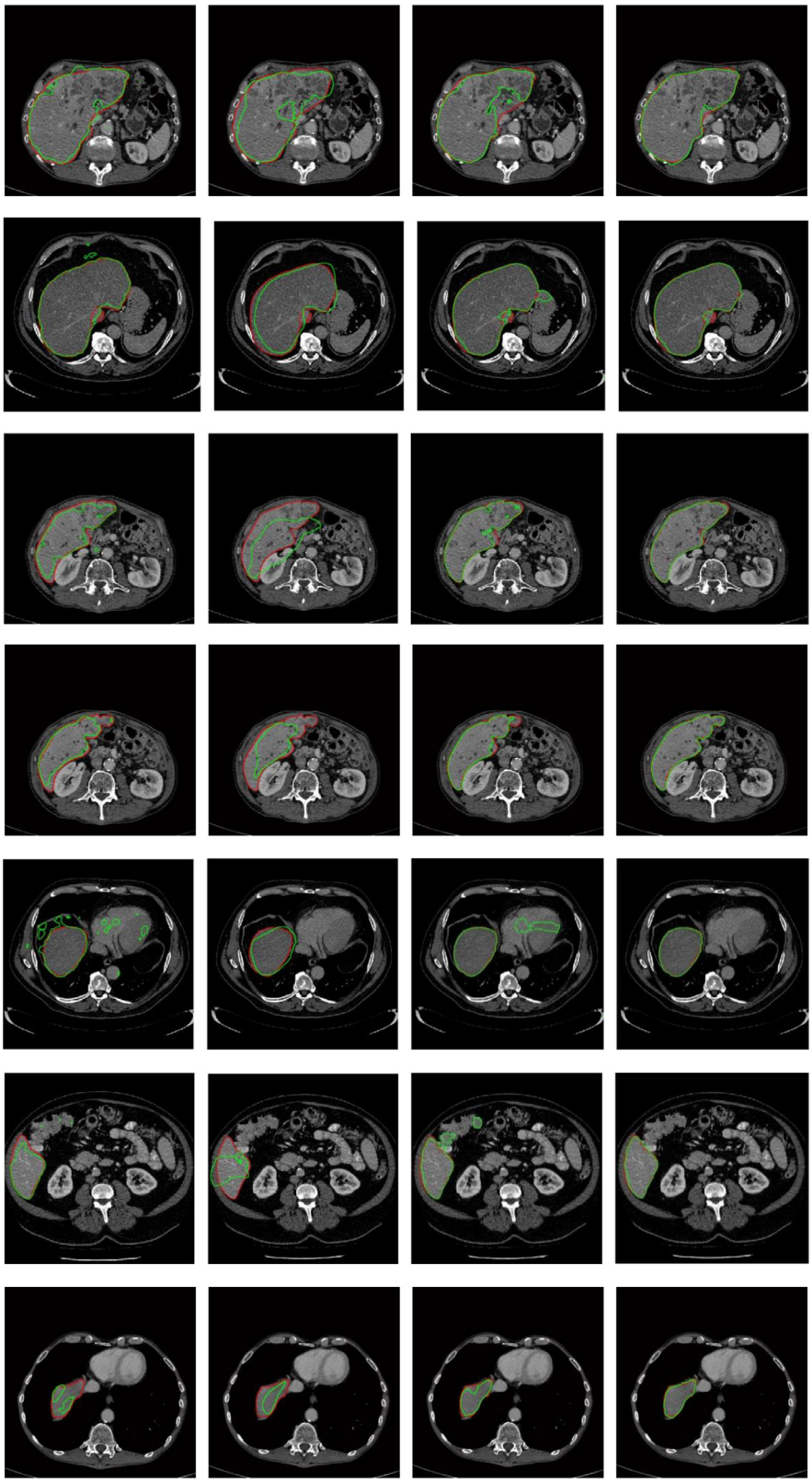
"Xavier" (Glorot and Bengio, 2010) is used as the initialization method, which is used in GIU-Net, IU-Net and U-Net. The initial learning rates of GIU-Net, IU-Net and U-Net are set as 1e-4, and their attenuation is set as the same 0.0005. The value of gamma is set as 0.1. The Attenuation strategy and the changing formula of learning rate are 'step' and ir=base_ir×0.1^(floor(interation/stepsize)), respectively. The values of batch_size is set as 1, and the training epoch is set as 100.

In the comparison experiment, the learning rate of FCN-8 s is set as 10e-8 (the default value) as the initial one with fine-tuning in our training process. The training epoch is set as 50.

4.4. Experiment analysis

To validate the effectiveness, our proposed solution GIU-Net has been compared with FCN-8s (Long et al., 2015), U-Net (Ronneberger et al., 2015) and our IU-Net. The settings of these networks are discussed in the above subsection, and the evaluation indexes are DSC,VOE,RVD,ASD,RMSD and MSD.

Fig. 4 shows the comparison segmentation results obtained from four algorithms (i.e., FCN-8s, U-Net, IU-Net, and GIU-Net)



(a) FCN-8s (b) U-Net (c) IU-Net (d) GIU-Net

Fig. 4. Segmentation results of the four algorithms (FCN-8 s, U-Net, IU-Net, and GIU-Net).

based on three groups (i.e., larger, medium, small liver region groups) of classic slices. The red outline shows the ground truth while the green line shows the segmentation result.

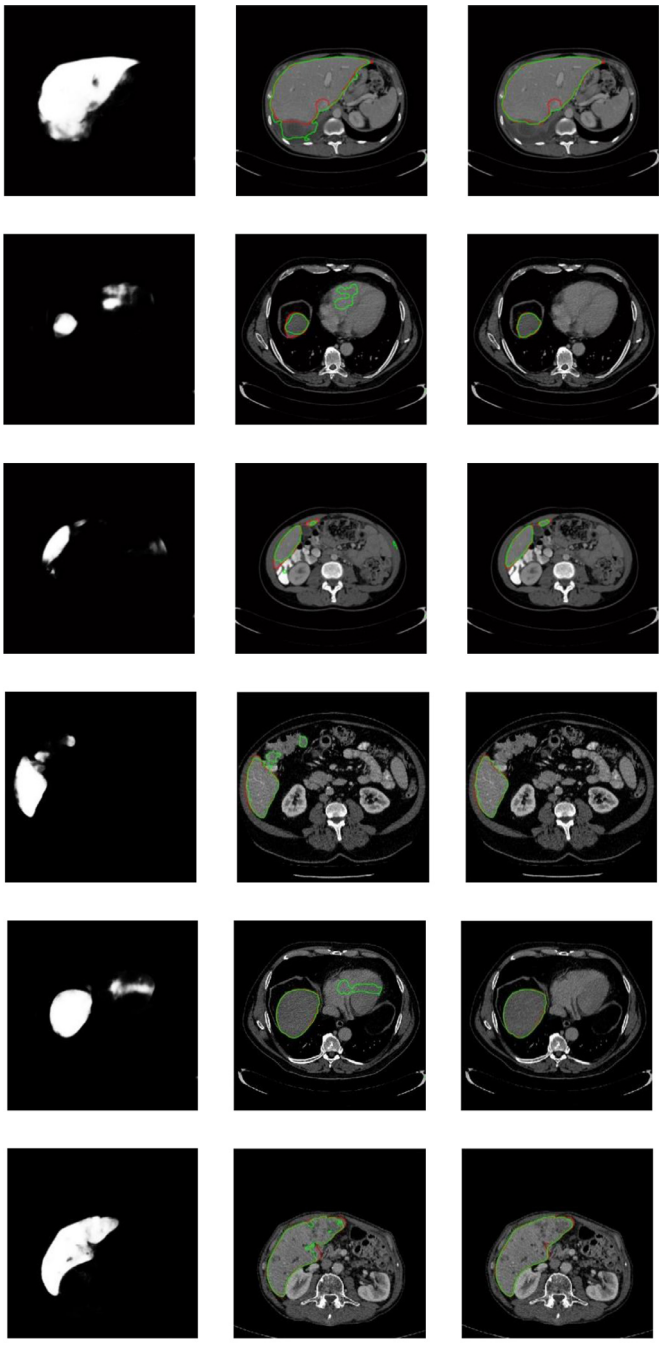
From Fig. 4, we can see that when dealing with large and middle livers, all models can segment livers out of images, but FCN-8s, U-Net, IU-Net have some degree of under-segmentation or over-segmentation. Some of their segmented livers do not have complete boundaries; some have holes; and some even have extra parts that do not belong to livers. However, our method GIU-Net can get more precise boundaries, and its segmented livers do not have extra holes. When dealing with images with small liver regions, FCN-8s often performs poorly on the liver boundary treatment with significant under-segmentation. It can only segment small part of liver region. U-Net can separate the approximate location of the liver area, but cannot get the precise boundary. IU-Net performs much better than FCN-8s on segmenting small liver regions of image slices, while its treatment of pathological slices and liver boundaries is still poor. By comparison, our proposed solution GIU-Net, which uses a graph cut algorithm to deal with the liver probability function output by IU-Net, makes up the deficiency of IU-Net in processing the liver boundary and adds sequence information to the sequence segmentation. Therefore, it performs much better than IU-Net in dealing with pathological slices. The mean and mean square deviation of segmentation results of the four algorithms are shown in Table 1, including three 2D-FCN methods (Ben-Cohen et al., 2016; Christ et al., 2017; Kaluva et al., 2018), and three 3D CNN methods (Li et al., 2018; Lu et al., 2017; Rafiei et al., 2018). The unit of VOE, RVD and DSC is%, and the unit of ASD, RMSD and MSD is mm.

Table 1 shows that our GIU-Net (95.05%) performs significantly much better than IU-Net, U-Net, FCN-8s and all other methods, except H-DenseUNet (96.5%) in terms of DSC and ASD. It is also the best among all comparison methods in terms of all three indicators (i.e., VOE, RMSD, and MSD). IU-Net performs the best in terms of RVD. Although our GIU-Net is not better than IU-Net in terms of RVD, it is more stable than IU-Net. Besides, it is obvious that GIU-Net significantly improves the performance of IU-Net in terms of other five indicators (i.e., DSC, VOE, ASD, RMSD, and MSD).

Our IU-Net significantly improves the performance of U-Net, and performs much better than FCN-8s in terms of all six indicators. Both U-Net and FCN-8s have a larger value in terms of ASD, RMSD and MSD. These values indicate that the segmentation results of both U-Net and FCN-8s have a great distance from the ground truth of liver regions.

Table 1 also shows that the experimental results of U-Net and FCN-8s have very high variances. Their larger variance shows that both U-Net and FCN-8s are not very stable in the segmentation. This is because their segmentation results are not further refined, their output boundary is relatively rough, and there exists mis-segmentation or over-segmentation.

These comparisons show that the sequence context information is very import in CT sequence images. Most 2D CNN models, such as U-Net, FCN, and 2D-dense-FCN, always focus on the accuracy of individual images and ignore the spatial relationship of the images. But, our proposed solution GIU-Net solves this issue. It corrects the mis-segmentation by combining the sequence context information. As for 3D CNN models, such as 3D-FCN, 3D-UNet and H-DenseUNet, those models are limited by GPU memory, so they crop the input images and use several adjacent slices as the input. For example, H-DenseUNet uses the image size $224 \times 224 \times 12$. It crops images from 512×512 to 224×224, so that some images may lost some liver in CT sequence, because the abdominal area of different patients is different. Besides, the location of livers can be also affected by age. GIU-Net uses original images during the whole stage, including the refinement stage. Furthermore, it also uses the whole sequence of images.



(a) probability graph (b) IU-Net (c) GIU-Net

Fig. 5. Segmentation results between IU-Net and our proposed GIU-Net.

We further conducted experiments to investigate the impact of using the sequence information and the graph cut function in GIU-Net by comparing with IU-Net. We selected six classical liver slices to illustrate the effect of using sequence information and the graph cut function to optimize the IU-Net output results. Note that these slices contain normal and pathological liver images. Our comparison results are shown in Fig. 5. Again, the red outline presents the ground truth while the green line shows the segmentation results.

From Fig. 5 we can see that IU-Net misuses the adjacent gray scale approximate region of livers as lesion areas when processing the liver slices and misjudges when treating the small

slices of liver regions, which results in mis-segmentation or under-segmentation. Under these circumstances, it is difficult to achieve satisfactory segmentation results using the threshold only with the means of probability distribution maps. Our proposed solution GIU-Net corrects the mis-segmentation by combining the sequence context information. As long as the segmentation accuracy of the first slice is guaranteed, the segmentation precision and robustness of the slices in the sequence can be improved to a certain extent, which is the advantage of GIU-Net, comparing to IU-Net.

5. Conclusions

In the process of liver segmentation, it is necessary to improve the low contrast between a liver and its peripheral organs and the large difference between liver individuals in CT images. Based U-Net, we proposed a novel solution GIU-Net. Specifically, we first improved U-Net by directly coping the characteristics of the pooling layer, instead replicating the features of the pooling layer after two convolutions and the ReLU activation function in U-Net. Then we used IU-Net to segment a liver and obtain a probability distribution map. We then constructed a graph cut energy function with the probability distribution map obtained with the context information of the sequence of images. Finally, we obtained final segmentation results by minimizing the graph cut energy function through the maximum flow minimum cut algorithm. Our proposed solution GIU-Net combines the advantages of IU-Net and graph cut, and improves the accuracy and robustness of liver segmentation with the context information of liver sequences. Although GIU-Net is proposed for liver segmentation, it is very interesting to study its performance on other segmentation tasks (e.g., the automatic segmentation of images containing petroglyphs (Deufemia, Paolino, and de Lumley, 2012; Zeppelzauer et al., 2016)) and further improve it for these segmentation tasks with adjustments.

Credit author statement

Drs. Zhe Liu, Yu-Qing Song, and Victor S. Sheng worked together to come out the conceptualization and methodologies for this research. Mr. Rui Jiang and Mr. Xiaolin Zhang implemented the solutions, validated and visualized their experimental results. Drs. Liangmin Wang and Deqi Yuan provided resources and supervisions. All authors together prepared the original draft and revised this manuscript multiple times.

Acknowledgment

The authors would like to thank Radiologists of the Medical Imaging department of Affiliated Hospital of Jiangsu University. This work was supported by the National Natural Science Foundation of China (61772242, 61728205, 61572239, 61402204); Research Fund for Advanced Talents of Jiangsu University (14JDG141); Science and Technology Project of Zhenjiang City (SH20140110); Special Software Development Foundation of Zhenjiang City (201322); Science and Technology Support Foundation of Zhenjiang City (Industrial) (GY2014013).

References

Afifi, A., & Nakaguchi, T. (2012, October). Liver segmentation approach using graph cuts and iteratively estimated shape and intensity constrains. In *Proceedings of the international conference on medical image computing and computer-assisted intervention* (pp. 395–403). Springer, Berlin, Heidelberg.

Ben-Cohen, A., Diamant, I., Klang, E., Amitai, M., & Greenspan, H. (2016). Fully convolutional network for liver segmentation and lesions detection. In *Proceedings of the deep learning and data labeling for medical applications* (pp. 77–85). Springer, Cham.

Boykov, Y., & Funka-Lea, G. (2006). Graph cuts and efficient ND image segmentation. *International Journal of Computer Vision, 70*(2), 109–131.

Boykov, Y., & Kolmogorov, V. (2004). An experimental comparison of min–cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(9), 1124–1137.

Boykov, Y., & Veksler, O. (2006). Graph cuts in vision and graphics: Theories and applications. In *Handbook of mathematical models in computer vision* (pp. 79–96). Boston, MA: Springer.

Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(11), 1222–1239.

Chen, X., Udupa, J. K., Bagci, U., Zhuge, Y., & Yao, J. (2012). Medical image segmentation by combining graph cuts and oriented active appearance models. *IEEE Transactions on Image Processing, 21*(4), 2035–2046.

Chen, Y., Cao, Z., Cao, C., Yang, J., & Zhang, J. (2018, June). A modified U-Net for brain MR image segmentation. In *Proceedings of the international conference on cloud computing and security* (pp. 233–242). Springer, Cham.

Christ, P. F., Ettlinger, F., Grün, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S., et al. (2017). Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv:1702.05970.

Dantzig, G., & Fulkerson, D. R. (2003). On the max flow min cut theorem of networks. *Linear Inequalities and Related Systems, 38*, 225–231.

Deufemia, V., Paolino, L., & de Lumley, H. (2012, November). Petroglyph recognition using self-organizing maps and fuzzy visual language parsing. In *Proceedings of the IEEE 24th international conference on tools with artificial intelligence (ICTAI): 1* (pp. 852–859). IEEE.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology, 26*(3), 297–302.

Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., & Heng, P. A. (2016). 3D deeply supervised network for automatic liver segmentation from CT volumes. In *Proceedings of the* international conference on medical image computing and computer-assisted intervention (pp. 149–157). Springer, Cham.

Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

Gotra, A., Sivakumaran, L., Chartrand, G., Vu, K. N., Vandenbroucke-Menu, F., Kauffmann, C., et al. (2017). Liver segmentation: indications, techniques and future directions. *Insights Imaging, 8*(4), 377–392. doi:10.1007/s13244-017-0558-1.

Han, M., Liu, J., Meng, J., & Wang, Z. (2015). Local energy information combined with improved signed distance regularization term for image target segmentation algorithm. *Journal of Electronics & Information Technology, 37*(9), 2047–2054.

Heimann, T., Meinzer, H., & Wolf, I. (2007). A statistical deformable model for the segmentation of liver CT volumes using extended training data. In *Proceedings of the MICCAI Work* (pp. 161–166).

Heimann, T., Van Ginneken, B., Styner, M. A., Arzhaeva, Y., Aurich, V., Bauer, C., et al. (2009). Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Transactions on Medical Imaging, 28*(8), 1251–1265.

Ji, H., He, J., & Yang, X. (2014). Three-dimensional CT liver image segmentation based on hierarchical contextual active contour. *Journal of Biomedical Engineering, 31*(2), 405–412.

Kaluva, K. C., Khened, M., Kori, A., & Krishnamurthi, G. (2018). 2D-densely connected convolution neural networks for automatic liver and tumor segmentation. arXiv:1802.02182.

Liao, M., Zhao, Y., Zeng, Y., Huang, Z., & Zou, B. J. (2016a). Liver segmentation from abdominal CT volumes based on graph cuts and border marching. *Journal of Electronics & Information Technology, 38*(6), 1152–1156.

Liao, M., Zhao, Y. Q., Wang, W., Zeng, Y. Z., Yang, Q., Shih, F. Y., et al. (2016b). Efficient liver segmentation in CT images based on graph cuts and bottleneck detection. *Physica Medica, 32*(11), 1383–1396.

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C., & Heng, P. (2018). H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Transactions on Medical Imaging, 37*(12), 2663–2674.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).

Lu, F., Wu, F., Hu, P., Peng, Z., & Kong, D. (2017). Automatic 3D liver location and segmentation via convolutional neural network and graph cut. *International Journal of Computer Assisted Radiology and Surgery, 12*, 171–182.

Lu, X., Wu, J., Ren, X., Zhang, B., & Li, Y. (2014). The study and application of the improved region growing algorithm for liver segmentation. *Optik-International Journal for Light and Electron Optics, 125*(9), 2142–2147.

Peng, J., Wang, Y., & Kong, D. (2014). Liver segmentation with constrained convex variational model. *Pattern Recognition Letters, 43*, 81–88.

Rafiei, S., Nasr-Esfahani, E., Soroushmehr, S. M., Karimi, N., Samavi, S., & Najarian, K. (2018). Liver segmentation in CT images using three dimensional to two dimensional fully connected network. arXiv:1802.07800.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer, Cham.

Song, Q., Bai, J., Han, D., Bhatia, S., Sun, W., Rockey, W., et al. (2013). Optimal co-segmentation of tumor in PET-CT images with context information. *IEEE Transactions on Medical Imaging, 32*(9), 1685–1697.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).

Zeppelzauer, M., Poier, G., Seidl, M., Reinbacher, C., Schulter, S., Breiteneder, C., et al. (2016). Interactive 3D segmentation of rock-art by enhanced depth maps and gradient preserving regularization. *Journal on Computing and Cultural Heritage (JOCCH), 9*(4), 19:1–19:30.

Zhang, X. Q., Xiong, B., & Kuang, G. Y. (2015). A ship target discrimination method based on change detection in SAR imagery. *Journal of Electronics & Information Technology, 37*(1), 63–70.