IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 23, 2019, accepted June 4, 2019, date of publication June 14, 2019, date of current version July 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923218

Liver Tumor Segmentation Based on Multi-Scale Candidate Generation and Fractal Residual Network

ZHIQI BAI[1], HUIYAN JIANG[1], SIQI LI[1], AND YU-DONG YAO[2], (Fellow, IEEE)
[1]Software College, Northeastern University, Shenyang 110819, China
[2]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

Corresponding author: Huiyan Jiang (hyjiang@mail.neu.edu.cn)

This work was supported by the National Natural Science Foundation of China, under Grant 61872075.

ABSTRACT Liver cancer is one of the most common cancers. Liver tumor segmentation is one of the most important steps in treating liver cancer. Accurate tumor segmentation on computed tomography (CT) images is a challenging task due to the variation of the tumor's shape, size, and location. To this end, this paper proposes a liver tumor segmentation method on CT volumes using multi-scale candidate generation method (MCG), 3D fractal residual network (3D FRN), and active contour model (ACM) in a coarse-to-fine manner. First, livers are segmented using 3D U-Net and then MCG is performed on these liver regions for obtaining tumor candidates (all superpixel blocks). Second, 3D FRN is proposed to further determine tumor regions, which is considered as coarse segmentation results. Finally, the ACM is used for tumor segmentation refinement. The proposed 3D MCG-FRN + ACM is trained using the 110 cases in the LiTS dataset and evaluated on a public liver tumor dataset of the 3DIRCADb with dice per case of 0.67. The experimentations and comparisons demonstrate the performance advantage of the 3D MCG-FRN + ACM compared to other segmentation methods.

INDEX TERMS Fractal residual network, multi-scale candidate generation method, active contour model, liver tumor segmentation, CT volume.

I. INTRODUCTION

According to the World Health Organization (WHO) reports, liver cancer is one of the most common cancers in the world and is a main cause of death in all cancers. In 2012, 745,000 patients died of liver cancer worldwide [1], hepatic cell carcinoma (HCC) accounts for about 80% of all primary liver cancers and most patients with chronic liver disease have HCC. Detection of HCC at an early stage can greatly improve the cure rate of patients.

Computed tomography (CT), featured by its high spatial resolution and fast scanning speed, plays a significant role in liver cancer detection and diagnosis. The primary treatment methods include surgical resection, interventional therapy, locoregional ablation, etc. These treatment methods need the detail information of tumors, such as the size, shape, and location before therapy in order to develop a fine treatment program [2]. In routine clinical practices, the segmentation of liver cancer can be done manually by radiologists with good expertise and experience. However, this is a time-consuming task requiring the radiologist to search through a 3D CT scan which may include hundreds of slices and multiple lesions. At the same time, automatic liver tumor segmentation is a difficult task due to different image acquisition protocols, various contrast-agents, and varying levels of contrast enhancements. In addition, dissimilar scanner resolutions lead to unpredictable intensity, and many different types of lesions, especially tumor sub-types, can occur in livers. Thus, these different types of tumors with varying contrast levels (hyper-/hypo-intense tumors) create obstacles for automated tumor segmentation [3]. In recent decades, with the development of computer-aided diagnosis (CAD) [4], several methods based on machine learning for automatic liver tumor segmentation on CT images have been developed which include traditional machine learning methods and deep learning methods. For traditional machine learning methods, Smeets *et al.* [5] proposed a combining level set method with supervised pixel classification for liver tumor segmentation.

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

Vorontsov *et al.* [6] used support vector machines (SVM) with a texture based deformable surface model to refine segmentation results. Shimizu *et al.* [7] utilized adaboost on texture features for liver lesion extraction. Conze *et al.* [8] proposed scale-adaptive supervoxel-based random forests to segment liver tumors. Hong *et al.* [9] adopted fuzzy C-means clustering to automatically segment liver tumor. The classifier used 16×16 image patches on individual CT slices. Billelo *et al.* [10] utilized intensity-based histogram and liver contour refinement to segment liver tumors, followed by an SVM to sift out false alarms. Masuda *et al.* [11] described a method to segment liver tumors by enhancing the CT scan contrast levels and expectation maximization of the posterior marginal. This super-voxel method, called EM/MPM, produced candidates that were then filtered using shape and location information.

With the development of deep learning, many methods have been developed to segment liver tumors through convolutional neural networks. Li *et al.* [12] designed a hybrid densely connected U-Net to effectively probe hierarchical intra-slice features for liver and tumor segmentation. Jiang *et al.* [13] proposed a cascaded deployment of AHCNet, which combined soft and hard attention mechanism and long and short skip connections, producing competitive results for liver tumor segmentation in CT volumes. Chen *et al.* [14] trained a neural network classifier on the liver voxels with a custom feature descriptor. Ben-Cohen *et al.* [15] used a fully convolutional network (FCN) for liver segmentation and tumor detection. Christ *et al.* [16] combined a FCN with dense 3D conditional random fields to segment liver tumors. Yuan. *et al.* [17] segmented liver and tumor by two hierarchical convolutional-deconvolutional neural networks(DCNNs) and used histogram equalization to enhance segmented liver region. Then, the enhanced region was sent to the third DCNN as additional information for liver tumor segmentation. Vorontsov *et al.* [18] proposed a simple single-stage model that was trained end-to-end and there were two FCNs in this model. The first FCN was used to segment liver regions and the features, which were extracted by the first FCN, were fed into the second FCN to segment liver tumors. Lei *et al* [19] gradually learned and inferred the boundaries of both the liver and the liver lesions through cascaded deep resnet with multi-scale fusion. The first resnet was used to generate the probability map of liver and tumor, and the second can predict the final result. Li *et al.* [20] automatically segmented CT liver images through CNNs and then post-processed them through random forests. Bellver *et al.* [21] proposed a cascade network for liver tumor segmentation. First, they used deep retinal image understanding (DRIU) to segment the liver, and then used a detector to segment liver tumors.

These deep learning methods are image-level segmentation which result in unsatisfactory sensitivity for liver tumor segmentation. The sensitivity of segmentation can be effectively improved if block-level segmentation is adopted. Nowadays, a sliding window is the mainstream method for candidate regions selection of liver tumor. However, using a sliding window to obtain the liver tumor candidates will divide one tumor into multiple regions, which is unfavorable for the segmentation of liver tumor. To solve this problem, a tumor candidate generation method based on superpixel segmentation is proposed in this paper, which ensures the integrity of tumor information while decreasing the calculation complexity by reducing the number of tumor candidates. Nowadays, the superpixel segmentation and classification method is widely utilized for segmentation tasks in other fields of disease recognition. For example, Li *et al.* [22] segmented pancreas cancer based on the simple linear iterative clustering. Bechar *et al.* [23] proposed a semi-supervised glaucoma screening approach including applying superpixel method and incorporating prior knowledge of the optic cup and disc. Then they trained the Co-forest classifier by these superpixels. Chu *et al.* [24] developed a multiple stage method for breast mass detection. They segmented mass candidates through the simple linear iterative clustering (SLIC) method. Then they prescreened suspicious regions through superpixel classification. Finally, level set was utilized to optimize lesion contour refinement. Bejnordi *et al.* [25] implemented automatic detection of regions of interest in whole slide histopathological images through generating and classifying superpixels at multiple resolution.

In this paper, we propose a method for liver tumor segmentation. First, 3D U-Net [26] is used to segment liver regions. Then, liver regions are segmented to tumor candidates by multi-scale candidate generation. Tumor candidates are consequently classified and fused for the purpose of segmentation. This method can increase the proportion of liver tumor information compared with the liver information in candidate regions. In order to improve the classification accuracy and sensitivity of the network for liver tumor segmentation, a new network structure, namely, fractal-residual structure, was proposed.

The contributions of this paper are as follows.

- (1) This paper proposes a method combining the multi-scale superpixel segmentation method and multiple neighborhood information to generate liver tumor candidates for segmentation, which can involve more complete tumor information for candidate regions. This increases the network's classification sensitivity to liver tumor details and reduces computational complexity caused by redundant information while increasing the amount of effective information.
- (2) It is well known that residual structure and fractal structure are both very efficient network structures. We introduce the idea of the fractal structure into the residual structure to obtain the fractal residual structure, which increases the network width while keeping the depth of network in order to extract more modal characteristics and increase the generalization ability of the network.

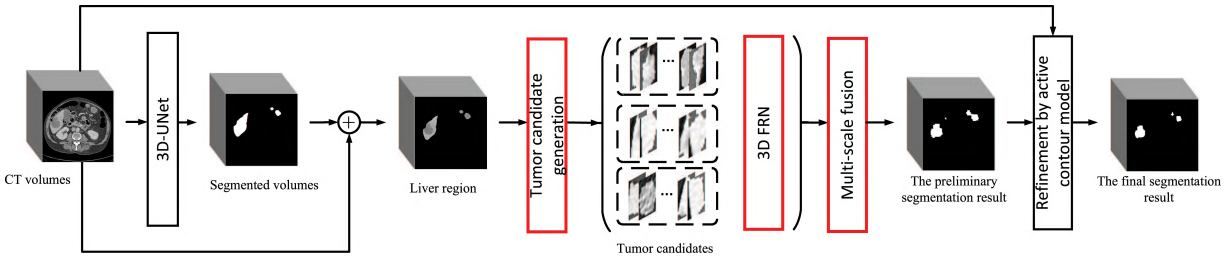


FIGURE 1. The flowchart of the proposed liver tumor segmentation method. The red box part (tumor candidate generation, 3D FRN, and multi-scale fusion) highlights the contributions of this paper.

- (3) Auxiliary classifiers are introduced in the proposed network. Specifically, we use a joint loss function to enable the auxiliary classifier to assist the main classifier. The method takes full account of the features proposed in each block and uses them to jointly supervise the main classifier. Each auxiliary classifier is considered as a back propagation gradient signal added to the network.

II. METHOD

The flowchart of proposed liver tumor segmentation method, which includes (1) liver region segmentation; (2) liver tumor candidate generation; (3) tumor candidate classification, and (4) liver tumor boundary refinement using an active contour model [27], is shown in Figure 1.

The specific steps are as follows.

Step 1: We segment liver regions from CT volumes through 3D U-Net [26].

Step 2: The multi-scale candidate generation method, which combines multi-scale superpixels segmentation method based on linear spectral clustering (LSC) [28] and multiple neighborhood information, is proposed for segmenting tumor candidates within liver regions.

Step 3: 3D FRN combining the fractal structure [29] and the residual structure [30] is proposed for classifying the liver tumor candidates.

Step 4: The active contour model [27] is introduced to refine the boundary of the liver tumor and the high probability region in the network output is used as the initial contour.

A. TUMOR CANDIDATE GENERATION METHOD IN THE LIVER REGION

Because the size of liver tumors in the CT image varies, accurate liver tumor segmentation remains a challenge. To this end, the region of background needs to be narrowed by segmenting the liver region, which can effectively improve the segmentation efficiency of tumors. In this paper, we use the 3D U-Net for liver segmentation before liver tumor segmentation. Although the image after liver segmentation reduces a large number of non-interested regions, the tumor region of interest is still too small to be segmented in liver regions. To solve this problem, we decided to cut each segmented liver image into tumor candidates and then classify the tumor candidates to obtain the segmentation results. Nowadays, the effective candidate generation method is to use a sliding window to generate candidate regions pixel by pixel, which is a type of pixel-level classification. Although the neighborhood information of each pixel can be fully considered in this method, it also causes a series of problems, such as, large amount of redundant information and significant requirement of calculation. Also, some scattered points are caused by false positives, which can be misleading to doctors. Therefore, for the above problems, we propose a tumor candidate generation method (MCG) for dividing liver regions into tumor candidates. The method includes multi-scale superpixel method and multiple neighborhood information.

1) MULTI-SCALE SUPERPIXEL METHOD

The superpixel method, linear spectral clustering (LSC), which maps images to high-dimensional spaces to find relationships between pixels, putting the same type of pixels into the same candidates. LSC can generate good candidates with prior knowledge which benefits tumor segmentation. Therefore, LSC is used to generate tumor candidates.

The traditional superpixel classification method is to directly classify superpixels as classification units and class them by extracting superpixel features. However, it has been experimentally found that the single-scale segmentation results heavily depend on the superpixel segmentation results. If the segmentation result of a single-scale superpixel segmentation method cannot segment the tumor boundary accurately, it will also affect the final segmentation result. The single-scale superpixel segmentation method has low tolerance to false positive and is based on a single-scale candidate region. Consequently, only limited local information is used, which results in a higher false positive rate and lower sensitivity. Therefore, a multi-scale superpixel method is chosen as the method to generate tumor candidates.

For $I_k \in R^{c\times c}$, the kth slice of input image I which only includes the liver region, we calculate superpixel results of s scales. The formula for the single-scale superpixel result is as follows.

$$F_n(I_k) = LSC(I_k, n) \tag{1}$$

where $LSC(\cdot)$ stands for linear spectral clustering operation. Parameter n represents the number of superpixels we segment I_k by LSC in one scale. $F_n(I_k)$ is the superpixel mask of I_k and lth superpixel region in $F_n(I_k)$ is $B_{(k,l)}$. Some examples of superpixel segmentation are shown in Figure 2.

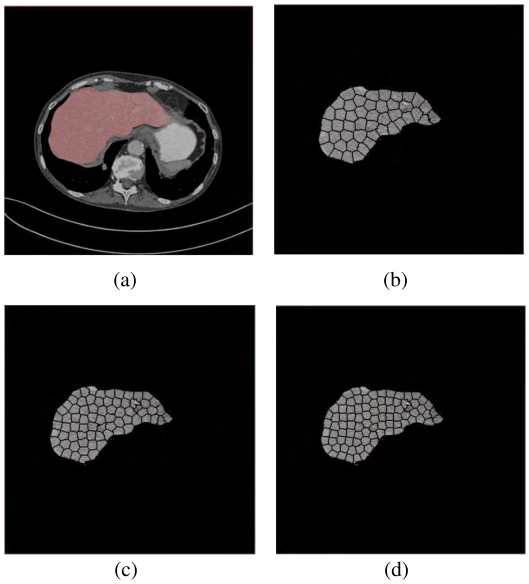


FIGURE 2. Examples of three-scale superpixel segmentation results. (a) Input CT image with the red region indicating the liver region. (b) Superpixel segmentation result for scale with $n = 1000$. (c) Superpixel segmentation result for scale with $n = 1500$. (d) Superpixel segmentation result for scale with $n = 2000$.

2) MULTIPLE NEIGHBORHOOD INFORMATION

Although using the multi-scale superpixel segmentation method can generate good tumor candidates, these tumor candidates, which contain only single slice image information and ignore background information, are not sufficient for effective segmentation. Therefore, a novel approach (MCG) which includes both context information and multi-scale information is developed to increase multiple neighborhood information in superpixel results of each scale. Let function $\mathscr{T}$ represent the transformation from the liver region volume data to the input of the proposed network and $\mathscr{T}^{-1}$ stand for the transformation from the output of the proposed network to the segmentation results:

$$X_{in} = \mathscr{T}(I_k, F_n(I_k)), \quad X_{in} \in R^{m\times i\times L\times L\times s} \tag{2}$$

$$Y = \mathscr{T}^{-1}(X_{out}, F_n(I_k)), \quad Y \in R^{m\times 512\times 512} \tag{3}$$

where m is the number of slices, $\mathscr{T}^{-1}$ is the inverse operation of $\mathscr{T}$, X_{in} is the input of the 3D fractal residual network (3D FRN) and X_{out} is the output of the 3D FRN. $\mathscr{T}$ includes a series of operations and the detailed process of $\mathscr{T}$ is as follows.

First, we use LSC to generate $F_n(I_k)$ of I_k and calculate the area $S_{(k,l)}$ of $B_{(k,l)}$. Second, if $B_{(k,l)}$ is not all 0, the center point $P_{(k,l)}$ of $B_{(k,l)}$ is used as the center of the square tumor candidates. Third, the size of the tumor candidates $L(k,l,i)$ is determined according to $S_{(n,k)}$. There are i kinds of size we choose, and each size is calculated as follows,

$$L(k,l,i) = \begin{cases} \left\lfloor \sqrt{S_{(k,l)}} \right\rfloor, & if\ i = 1 \\ L(k,l,1) + (i-1)\sqrt{L(k,l,1)}, & otherwise \end{cases} \tag{4}$$

Then, i square areas, which take $P_{(k,l)}$ as center point, $L(k,l,i)$ as side length, are taken from I_k and resized to the mean size among all candidates. They are then stacked into an i channel tumor candidate.

Finally, the context information is also utilized to introduce more sufficient neighborhood information. For an image I_k, I_{k-1} and I_{k+1} are involved to generate 3-channel tumor candidates. The class of a tumor candidate can be predicted through the 3D FRN and the prediction result is also regarded as the class of the corresponding superpixel that generates the candidate. For a specific superpixel scale of I_k, the classification results of all superpixels are combined based on $F_n(I_k)$ to form the image-level result of this superpixel scale. The detailed generation process is shown in Figure 3. We obtain s binary image-level results corresponding to s superpixel scale, which means, for each pixel, there are s prediction results $[P_1, P_2, \ldots, P_s]$. $\min([P_1, P_2, \ldots, P_s])$ is the finally segmentation result of this pixel.

B. 3D FRACTAL RESIDUAL STRUCTURE

It is well known that liver tumor segmentation is considered as a challenging task. First, the morphology of liver tumor varies widely in patients, such as size, shape, location and numbers. Second, some tumor boundaries are fuzzy, which is similar to blood vessels and surrounding tissues in morphology. Therefore, we propose a 3D fractal residual (FR) structure inspired by the idea of the fractal structure. The original fractal network, due to the random discarding mechanism, has improved the generalization ability of the network, but has also led to the discarding of many effective features. In order to increase the generalization ability of the network as well as acquire more features of different resolutions, we add the shortcut connection in the FR structure. By means of building a deep network, the FR structure enlarges the width of the network, expands the dimension of the features extracted by the network, realizes the reuse of the features, and greatly improves the ability of the network to classify tumor candidates. FR structure can be expanded iteratively to an i-level structure, which is expressed as follows,

$$M_i = \begin{cases} res, & if\ i = 1 \\ 2M_{i-1} \bigoplus M_1 \bigoplus z, & otherwise \end{cases} \tag{5}$$

where res represents residual structure. There are two types of residual structure as shown in Figure 4, We use residual structure (b) instead of residual structure (a) in order to reduce parameters. M_i represents the i-level FR structure, which is the adjacent structure of M_{i-1}. In order to illustrate the FR structure in detail, the structure of M_3 is presented in Figure 5 as an example. z is the input and $\oplus$ is a join operation. The

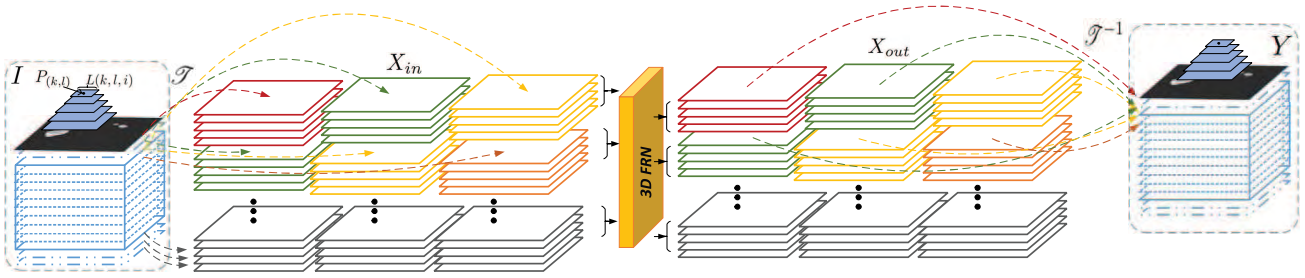


FIGURE 3. The detailed generate process $\mathcal{T}$ and multi-scale fusion process $\mathcal{T}^{-1}$.

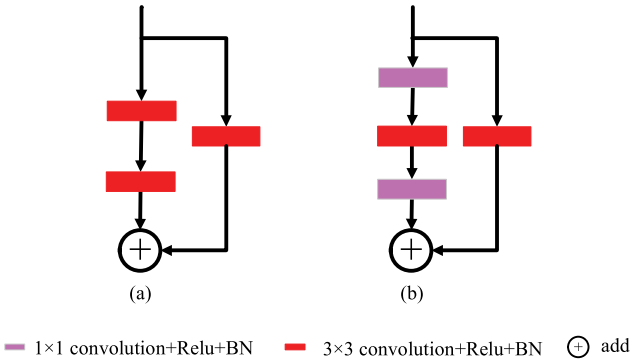


FIGURE 4. Two types of residual structure. (a) is the convolutional residual block with two-layer learning unit and (b) is the convolutional residual block with three-layer learning unit.

joint operation is to merge three globs into a single output glob. The three globs are obtained by residual block, M_i, and the original input. These globs represent three types of features in visual levels, which are extracted to increase the classification efficiency of network. In addition, because of the existence of random dropping paths in the join operation, the back propagation process randomly enhances each path, thereby, improving the overall performance of the network. Then, each path of the FR structure is a single resnet in the process of back propagation. resnet introduces the input of each layer of the network to the later layers by introducing a characteristic shortcut connection, thus making the gradient flow uninterrupted, allowing parameters to be updated in depth in the network, and making the convergence speed faster in the course of training. The FR structure, which combines the advantages of fractal structure and residual structure, reduces overfitting and improves the accuracy of classification with a continuous improvement in learning ability.

C. 3D FRN WITH MULTI-CLASSIFIER

The 3D RFN is composed of b blocks which combine global information with detail information. Each block is a M_i structure. Then, the 3D FRN includes $b \times 4 \times \left(2^i - 1\right)$ layers. In this paper, we build a 3D RFN with 112 layers ($b = 4$, $i = 3$) which includes 3D convolution layer, 3D maxpooling layer and predict layer. The structure of 3D FRN is shown in Figure 6. In [31], it is shown that for liver tumor analysis, the network stacking a large number of small convolutional kernels provides the same sensitivity field effect as the network with a large convolution kernel, but small convolution kernel reduces the computational cost of training network. To this end, the shape of all convolution kernels in 3D FRN are $3 \times 3 \times 3$.

In order to increase the classification ability of the network, some auxiliary classifiers are added after the output of each block. This method can effectively alleviate the problem of gradient disappearance and assist the training process with direct supervision on the hidden layers.

Although the 3D FRN and the fractal network use the similar structure, there are major differences. First, 3D FRN combines with shortcut connection and randomly dropout paths and we extend the 3D FRN by replacing a single layer with a larger fractal residual block, so that the features of any layer can be transferred to any subsequent layer or discarded. In this way, the loss of a large amount of detailed information due to the high level of hierarchy (for example, fast R-CNN [32] and faster R-CNN [33] both use the features of the last convolution layer for target detection) can be avoided. Furthermore, this greatly increases the variety of features extracted by 3D FRN, either the detailed information of the lower level or the semantic features of the higher level. Second, the back propagation mode of the 3D FRN is different from the fractal network. The 3D FRN can reach a much deeper extent than the fractal network because of the existence of multiple shortcut connections during the back propagation process. Regardless if all paths are reserved or only one path is reserved, the shortcut connection can be used to ensure that the gradient can be passed back, ensuring the efficiency of back propagation and speeding up.

D. LOSS FUNCTION

In order to ensure that the multiple auxiliary classifiers can effectively supervise the training of the entire network, we use a joint loss function. The loss function consists of three parts. The first two parts are the loss of the main classifier and the loss of the auxiliary classifier and, for preventing the network from overfitting, we add a regularization term to the loss function.

$$\mathcal{L}(\mathcal{X}; W, w^m) = \mathcal{L}(\mathcal{X}; W) + \sum_{m=1}^{M} \lambda_m \mathcal{L}_m(\mathcal{X}; W, w^m) + \mu \|W\|_2 \tag{6}$$

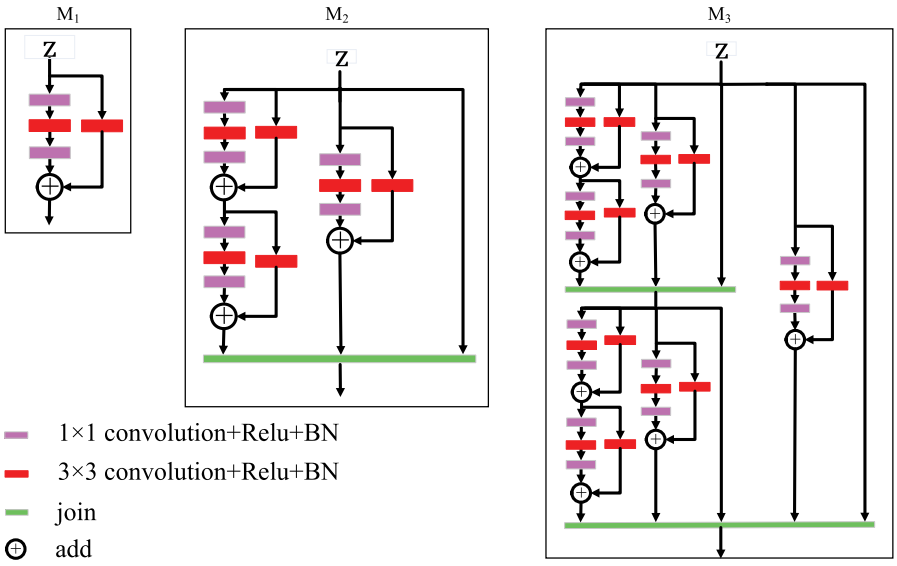


FIGURE 5. FR structure with 3 levels, M_3 is composed by M_2 and M_2 is composed by M_1. In order to explain the structure of M_3 more clearly, we also exhibit the structure of M_1 and M_2. From left to right: M_1, M_2 and M_3 represent FR structures with 1 level, 2 levels and 3 levels, respectively.

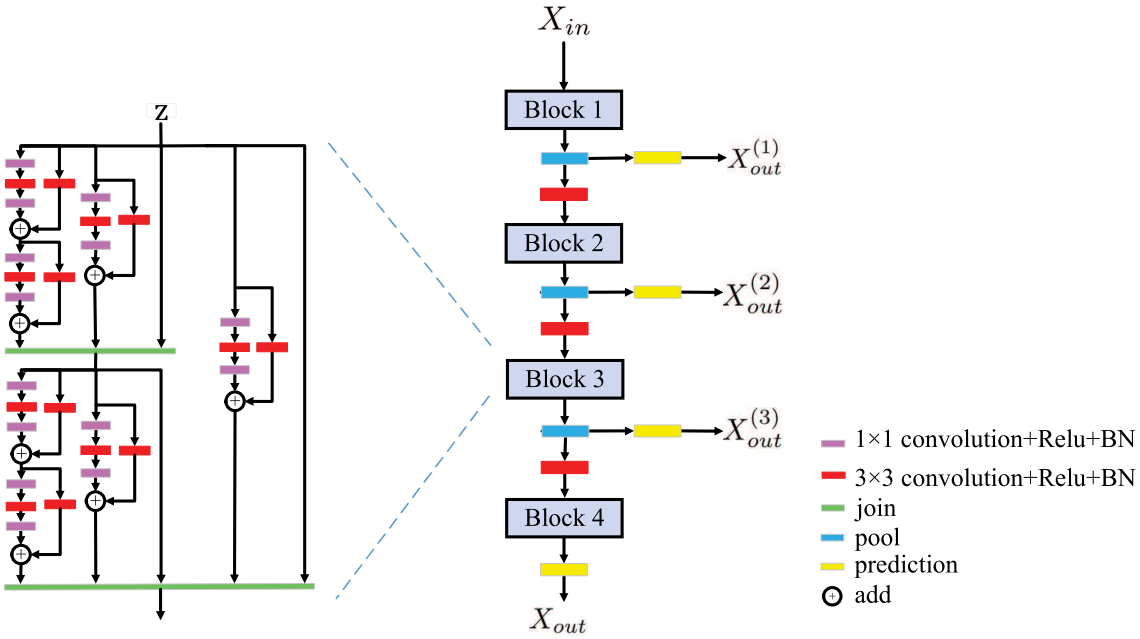


FIGURE 6. The structure of 3D FRN which contains four blocks, each of which is a 3-level FR structure. The $X_{out}^{(1)}$, $X_{out}^{(2)}$ and $X_{out}^{(3)}$ is the output of each auxiliary classifiers.

where W and w_m represent the weight of a main classifier and m auxiliary classifiers respectively and $\mathcal{X}$ is the training sample. The main classifier's cross-entropy loss function $\mathcal{L}(\mathcal{X}; W)$ and the auxiliary classifiers' cross-entropy loss function $\mathcal{L}_m(\mathcal{X}; W, w^m)$ are calculated as follows.

$$\mathcal{L}(\mathcal{X}; W) = \sum_{x_i \in \mathcal{X}} -y_i \log \hat{y}_i(x_i; W) \tag{7}$$

$$\mathcal{L}_m(\mathcal{X}; W, w^m) = \sum_{x_i \in \mathcal{X}} -y_i \log \hat{y}_i(x_i; W, w^m) \tag{8}$$

where $\hat{y}_i$ is the probability of x_i belong to each class and y_i is the label of x_i.

E. ACTIVE CONTOUR MODEL

The preliminary result can be experimentally obtained through aforesaid methods as shown in Figure 7. However, the data generated based on MCG leads to superfluous information. Consequently, in the boundary of the tumor, many candidate regions contain only part of tumors because the border of candidates overlap with tumor edges, but they tend to be classified as tumor regions due to the generalization

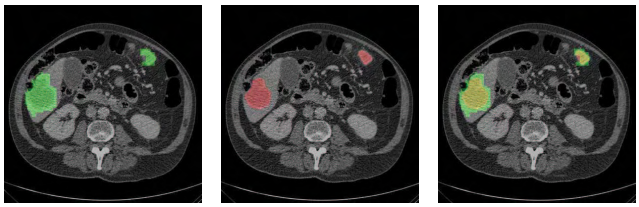


FIGURE 7. The preliminary result. (a). The green regions are the predict of liver tumor by 3D MCG-FRN. (b). The red regions are the ground truth and (c). The yellow regions denote correctly tumor segmentation we predict while the green ones for tumor false positive pixels and the red ones for tumor false negative pixels.

process of 3D FRN. As a result, the final segmentation result will be larger than the real tumor. In addition, even though multi-scale superpixel can narrow the burr boundaries to some extent, the smoothness of segmentation boundary still needs improving since the final result is a superpixel-level result. Therefore, the active contour model [27] is applied as a simple post-processing procedure to finetune the obtained boundary since 3D FRN could provide an excellent initial boundary. In this way, a secondary correction step will be performed on candidate regions. Meanwhile, the burr boundary problem can be effectively solved.

III. EXPERIMENTS AND RESULTS

A. DATASET

Experiment dataset is the LiTS dataset, involving 131 patients' contrast enhanced 3D abdominal CT scans which include 118 cases of tumor patients and 13 cases of normal patients. The number of slices per patient varies greatly. The dataset contains 58638 slices in total which is collected from six clinical sites by different scanners and protocols, with very different in-plane resolution from 0.55mm to 1.0mm and section spacing from 0.45mm to 6.0mm. The dataset covers dramatic changes in population, contrast, scan range, pathology and field of vision (FOV). Most CT scans are pathological, including tumors of different sizes, metastases and cysts. It is worth noting that the 3DIRCADb dataset[1] is a subset of the LiTS dataset.[2] In order to evaluate our model fairly, we use 3DIRCADb dataset as the testing data and the remaining data in LiTS dataset is considered as the training data.

B. PREPROCESSING

To facilitate experimental training, we clip the pixels outside the range (-150, 250) of the original images to the minimum/maximum value to improve the contrast between the liver and surrounding organs and tissues, therefore, excluding organs that are not of interest. After that, the image is normalized.

Since smaller proportion of the number of slices containing tumors in the dataset, in order to enhance the classification effect of the network, the dataset needs to be expanded. The data expansion is in addition to a series of geometric transformations, such as random cropping, flipping, shifting, scaling and tilting, and also using Elastic Distortions, which enables the network to learn deformation invariance, to expand the training data, to prevent over-fitting, and further enhance the ability of generalization. We use a random displacement vector to generate a smooth deformation on a coarse 3×3 grid. The displacement is sampled from a Gaussian distribution of 10-pixel standard deviation and then the displacement per pixel is calculated using bicubic interpolation. The expanded image is shown in Figure 8.

C. TRAINING PROCESS

To validate the method we proposed, we trained on the LiTS dataset and tested on the 3DIRCADb dataset. The 3DIRCADb dataset is the subset of the LiTS dataset, which is publicly available and provides a more accurate label. We use 3DIRCADb as the testing data and the other 110 cases in LiTS as the training and validation data (90% for training and 10% for validation). The model was implemented using Keras package [34].

The experimental process is divided into two stages. The first stage is the process of dividing the image into tumor candidates. First, the 3D U-Net is used to find the mask of the liver on CT image, which is used to reduce the background regions. Then, the image of liver region is segmented by MCG. The number of superpixel scale is 3 and the number of superpixel n of each scale is 1000, 1500, 2000, respectively. The stretch ratio of each scale is 0.2, and the final size of candidates is 32×32. In order to balance the scale between positive cases and negative cases in training dataset, the number of liver tumor candidates is limited to the same as the number of background candidates during the training process.

The second stage is to classify tumor candidates, 3D FRN depth 112 ($b = 4, i = 3$) is used for classification, with initial learning rate 0.002. When the loss of validation data does not reduce during three consecutive epochs, the learning rate decays to one tenth of the original. We train the network for 400 epochs with SGD optimizer. The momentum of SGD is 0.9 and the batch size is 20. From the first to the fourth block, we use 128, 256, 256, 512, 3×3 convolutions respectively. At the end of each block, we add a classifier. There are total four classifiers, including a main classifier and three auxiliary classifiers and the weights of auxiliary classifiers are $\lambda_m = (0.3, 0.3, 0.3)$. All tumor candidates are integrated into the original image size according to the superpixel mask that was originally divided and then the three-scale images are combined. The minimum probability value is selected as the final segmentation result. Finally, refine the preliminary result through ACM to optimize the boundary of the segmentation results.

All computations were performed on an Intel ®CoreTM i7-7700K CPU@3.60GHz, 8 Cores, 32 GB RAM running on Ubuntu 16.04 LTS and NVIDIA GeForce GTX 1080ti GPU with 3584 cores and 11GB memory.

[1] https://www.ircad.fr/research/3dircadb/
[2] https://competitions.codalab.org/competitions/17094

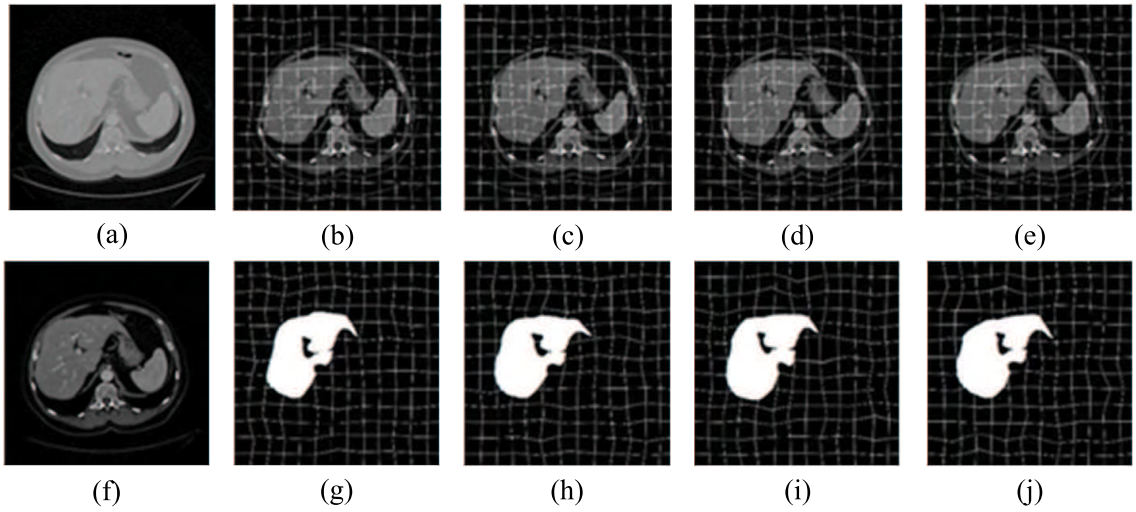


FIGURE 8. The sample of Elastic Distortions. (a) The original CT image. (b)–(e) cropped CT image to the elastic distortion sample. (f) The image after the Hu value. (g)–(j) cropped liver ground truth to the elastic distortion sample.

FIGURE 9. The comparison between segmentation results obtained by proposed method and the ground truth in 3D. (a) The green regions is the segmentation by 3D MCG-FRN. (b) The red regions is the ground truth.

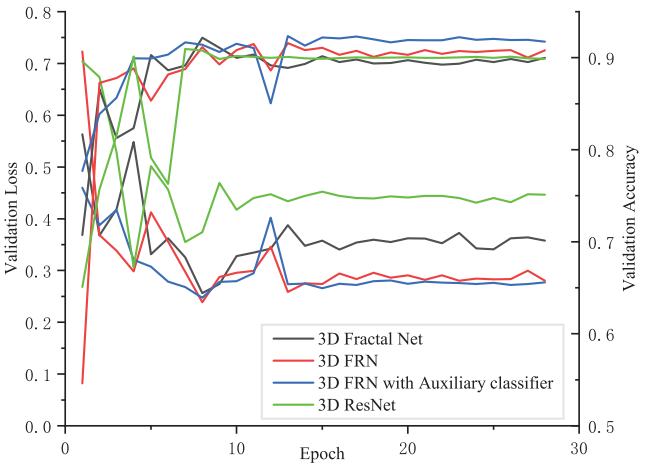


FIGURE 10. The training process of classify candidates with different model.

D. RESULTS

1) EVALUATION METRICS

This paper evaluates the segmentation results using the following metrics, dice per case, dice global, volumetric overlap error (VOE), relative volume difference (RVD), average symmetric surface distance (ASD) and the maximum surface distance (MSD). The smaller the value of the last four evaluation metrics, the better the segmentation results.

2) LIVER TUMOR SEGMENTATION RESULTS

This section presents the segmentation performance of the proposed method and the comparison with other methods, We compare our method with some recent segmentation methods [17], [35]–[37] on 3DIRCADb dataset. We found that our method still outperforms 3D U-Net [35] and resnet [37], with 16% and 7% improvement on dice per case for tumor segmentation.

IV. DISCUSSIONS

Liver tumor segmentation is an important prerequisite for effective treatment of liver disease. The method we propose is based on MCG, and the advantage is to find liver tumor regions accurately. Different from other candidate generation methods, our method has two important characteristics with respect to the proposed MCG. First, the context information and multi-scale information are involved in tumor candidates while reducing the number of tumor candidates and increasing computational efficiency. Second, the final segmentation results do not directly dependent on the results of superpixel segmentation. All tumor regions can be accurately included in one of the multi-scale superpixels or a combination of multiple scales, which facilitates later tumor segmentation. After that, we proposed 3D FRN. The new structure contains the residual structure and fractal structure, which increases the width and depth of the network and extracts more modal features, which effectively increases the accuracy of classification and improves the performance of the segmentation.

Figure 9 shows the 3D visualization result of the tumor through our proposed method. It can be seen that our results are approximately consistent with the label. However, there

TABLE 1. Segmentation results by ablation study of our methods on the 3Dircadb dataset. Statistically significant results are highlighted in bold font. (The CG means the MCG with single scale superpixel.)

Method	Dice per case	Dice global	VOE	RVD	ASD(mm)	MSD(mm)
3D CG-ResNet + ACM	0.599	0.687	0.464	0.385	4.708	12.201
3D MCG-ResNet + ACM	0.631	0.737	0.263	0.201	**2.055**	8.239
3D CG-Fractal Net + ACM	0.583	0.663	0.414	0.209	5.695	9.695
3D MCG-Fractal Net + ACM	0.616	0.669	**0.194**	-0.241	3.882	**5.474**
3D CG-FRN +ACM	0.615	0.692	0.537	0.593	7.005	12.146
3D MCG-FRN	0.572	0.654	0.638	0.829	7.791	13.408
3D MCG-FRN + ACM	**0.674**	**0.764**	0.324	**0.194**	4.408	7.113

TABLE 2. Our method's segmentation results compare with some other method on the 3Dircadb dataset. Statistically significant results are highlighted in bold font.

Method	Dice per case	VOE	RVD	ASD(mm)	MSD(mm)
Weight U-Net [35]	0.51	0.625	0.380	11.11	16.71
Yuan *et al.* [17]	0.56	0.378	0.228	**1.151**	**6.269**
Christ *et al.* [36]	0.56	-	-	-	-
ResNet [37]	0.60	0.564	-0.41	6.36	11.69
3D MCG-FRN + ACM	**0.67**	**0.324**	**0.194**	4.408	7.113

are some burrs on the boundary and the tumor surface is not smooth. Figure 12 shows the segmentation results for several patients with different tumor sizes. Most were segmented well (the first four rows) and there are few false positives. The segmentation result boundary is closely matched with the label boundary. But, in some cases, such as the fifth row, our method divides two tumors which close to each other into one tumor regions. For very small tumors (sixth row), our method has some miss detection.

In order to test the performance of the network, we compared the training process between the 3D fractal network, the 3D FRN, the 3D ResNet, and the 3D FRN with auxiliary classifier. As shown in Figure 10, it can be observed that the 3D FRN with auxiliary classifier achieves the lowest loss and the highest accuracy of validation dataset. 3D FRN performs worse than 3D FRN with auxiliary classifier, but better than 3D fractal network and 3D ResNet. This shows that the 3D FRN is better than 3D fractal network and 3D ResNet, at the same time, the auxiliary classifier has some improvement over FRN.

Figure 11 presents the visualization performance comparison between 3D MCG-FRN + ACM and 3D MCG-Fractal Net + ACM. It is seen that 3D MCG-FRN + ACM is more accurate than 3D MCG-Fractal Net + ACM, and 3D MCG-FRN + ACM is more sensitive than 3D MCG-Fractal Net + ACM.

Table 1 presents our performance of liver tumor segmentation results by ablation study of our methods on the 3Dircadb dataset. The 3D MCG-FRN based with active contour method has the best segmentation performance on dice and VOE. FRN performs better than fractal net and resnet from the comparison between the result of 3D CG-Fractal Net + ACM, 3D CG-FRN + ACM and 3D CG-ResNet + ACM. At the same time, According to the comparison of the results between 3D CG-Fractal Net + ACM and 3D MCG-Fractal Net +

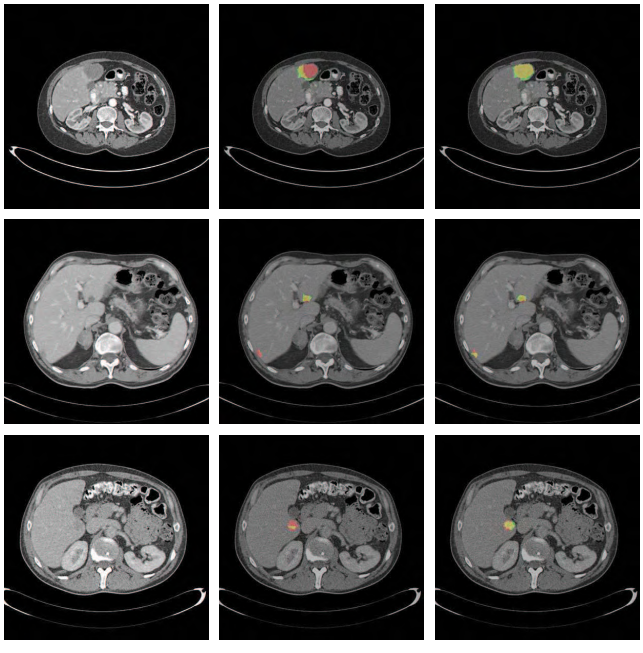


FIGURE 11. Tumor segmentation sample compare results, from left to right: Test image, the segmentation results by 3D MCG-Fractal Net + ACM and our results. Comparison between our result and 3D MCG-Fractal Net + ACM. For each comparison the following conventions were used: In red - ground truth, in green - out segmentation result, in yellow - overlapping regions between segmentation results and ground truth. (a) Test image. (b) The segmentation result by 3D MCG-Fractal Net+ACM. (c) The segmentation result by 3D MCG-FRN + ACM.

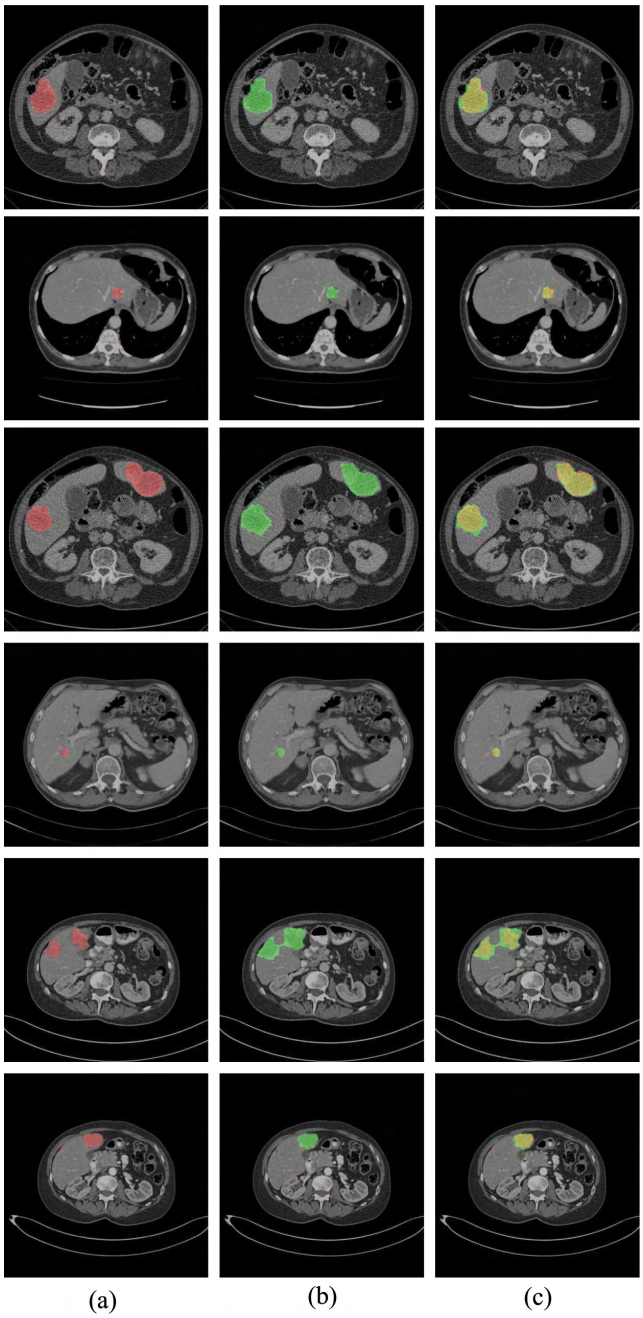


FIGURE 12. Tumor segmentation sample compare results. There are six samples, from left to right: Ground truth, the segmentation result by 3D MCG-FRN + ACM and overlap of them. Comparison between our result and ground truth. For each comparison the following conventions were used: In red - ground truth, in green - out segmentation result, in yellow - overlapping regions between our segmentation results and ground truth.

ACM and the comparison of the results of 3D CGFRN + ACM and 3D MCGFRN + ACM, we can see that using the multi-scale method is helpful for the final dice result. From the result of 3D MCG-FRN, we can see the importance of post processing.

The performance comparison of the proposed methods with other methods using 3DIRCADb dataset is shown in Table 2. For five evaluation indicators, 3D MCG-FRN + ACM achieves best results in terms of dice per case, dice global and RVD. As ACM is not very sensitive to fuzzy boundaries, 3D MCG-FRN + ACM performs not as good as [17] in terms of ASD and MSD.

V. CONCLUSION

In this paper, we propose a new method for liver tumor segmentation in CT images, including liver segmentation, multi-scale tumor candidate generation, tumor candidate classification, and the active contour model. First, we use 3D U-Net for liver segmentation and obtaining tumor candidates using MCG method. Then, with regards to candidate classification, we propose 3D FRN. Finally, for better segmentation results, active contour model is selected for post processing. We performed the segmentation tasks on 3DIRCADb dataset, the experiment results and comparisons with related work demonstrate that our proposed model can achieve a better segmentation performance.

There are some limitations in the methods proposed in this paper. For example, the boundaries of tumors cannot be accurately segmented and multiple adjacent tumors may be segmented into a single tumor region. In the future, we plan to address those issues using advanced algorithms, such as the 3D level set method or the 3D conditional random field method [16].

REFERENCES

[1] B. W. Stewart and C. P. Wild, *World Cancer Report 2014*. Geneva, Switzerland: WHO, 2014.

[2] P. Campadelli, E. Casiraghi, and A. Esposito, "Liver segmentation from computed tomography scans: A survey and a new algorithm," *Artif. Intell. Med.*, vol. 45, nos. 2–3, pp. 185–196, 2009.

[3] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," *Artif. Intell. Rev.*, vol. 50, no. 4, pp. 497–537, Mar. 2017.

[4] H. Fujita, Y. Uchiyama, T. Nakagawa, D. Fukuoka, Y. Hatanaka, T. Hara, G. N. Lee, Y. Hayashi, Y. Ikedo, X. Gao, and X. Zhou, "Computer-aided diagnosis: The emerging of three CAD systems induced by Japanese health care needs," *Comput. Methods Programs Biomed.*, vol. 92, no. 3, pp. 238–248, Dec. 2008.

[5] D. Smeets, D. Loeckx, B. Stijnen, B. De Dobbelaer, D. Vandermeulen, and P. Suetens, "Semi-automatic level set segmentation of liver tumors combining a spiral-scanning technique with supervised fuzzy pixel classification," *Med. Image Anal.*, vol. 14, no. 1, pp. 13–20, Feb. 2010.

[6] E. Vorontsov, N. Abi-Jaoudeh, and S. Kadoury, "Metastatic liver tumor segmentation using texture-based omni-directional deformable surface models," in *Abdominal Imaging. Computational and Clinical Applications*. Cham, Switzerland: Springer, 2014, pp. 74–83.

[7] A. Shimizu, T. Narihira, D. Furukawa, H. Kobatake, S. Nawano, and K. Shinozaki, "Ensemble segmentation using AdaBoost with application to liver lesion extraction from a CT volume," in *Proc. MICCAI Workshop 3D Segmentation Clinic, Grand Challenge II*, New York, NY, USA, Sep. 2008.

[8] P.-H. Conze, V. Noblet, F. Rousseau, F. Heitz, V. de Blasi, R. Memeo, and P. Pessaux, "Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 12, no. 2, pp. 223–233, Feb. 2017.

[9] J. S. Hong, T. Kaneko, R. Sekiguchi, K. H. Park, "Automatic liver tumor detection from CT," *IEICE Trans. Inf. Syst.*, vol. 84, no. 6, pp. 741–748, 2001.

[10] M. Bilello, S. B. Gokturk, T. Desser, S. Napel, R. B. Jeffrey, Jr., and C. F. Beaulieu, "Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT," *Med. Phys.*, vol. 31, no. 9, pp. 2584–2593, Sep. 2004.

[11] Y. Masuda, T. Tateyama, W. Xiong, J. Zhou, M. Wakamiya, S. Kanasaki, A. Furukawa, and Y. Wei, "Liver tumor detection in CT images by adaptive contrast enhancement and the EM/MPM algorithm," in *Proc. 18th IEEE Int. Conf. Image Process.* Brussels, Belgium, Feb. 2012, pp. 1421–1424. doi: 10.1109/ICIP.2011.6115708.

[12] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[13] H. Jiang, T. Shi, Z. Bai, and L. Huang, "AHCNet: An application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes," *IEEE Access*, vol. 7, pp. 24898–24909, 2019.

[14] E.-L. Chen, P.-C. Chung, C.-L. Chen, H.-M. Tsai, and C.-I. Chang, "An automatic diagnostic system for CT liver image classification," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 6, pp. 783–794, Jun. 1998.

[15] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan, "Fully convolutional network for liver segmentation and lesions detection," in *Deep Learning and Data Labeling for Medical Applications* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2016, pp. 77–85.

[16] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilicm, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W. H. Sommer, S.-A. Ahmadi, and B. H. Menze, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2016, pp. 415–423.

[17] Y. Yuan, "Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation," 2017, *arXiv:1710.04540*. [Online]. Available: https://arxiv.org/abs/1710.04540

[18] E. Vorontsov, A. Tang, C. Pal, and S. Kadoury, "Liver lesion segmentation informed by joint liver segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1332–1335.

[19] L. Bi, J. Kim, A. Kumar, and D. Feng, "Automatic liver lesion detection using cascaded deep residual networks," 2017, *arXiv:1704.02703*. [Online]. Available: https://arxiv.org/abs/1704.02703

[20] W. Li, F. Jia, and Q. Hu, "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks," *J. Comput. Commun.*, vol. 3, no. 11, pp. 146–151, 2015.

[21] M. Bellver, K.-K. Maninis, J. Pont-Tuset, X. Giro-I-Nieto, J. Torres, and L. Van Gool, "Detection-aided liver lesion segmentation using deep learning," 2017, *arXiv:1711.11069*. [Online]. Available: https://arxiv.org/abs/1711.11069

[22] S. Li, H. Jiang, Z. Wang, G. Zhang, and Y.-D. Yao, "An effective computer aided diagnosis model for pancreas cancer on PET/CT images," *Comput. Methods Programs Biomed.*, vol. 165, pp. 205–214, Oct. 2018.

[23] M. El Amine Bechar, N. Settouti, V. Barra, and M. A. Chikh, "Semi-supervised superpixel classification for medical images segmentation: Application to detection of glaucoma disease," *Multidimensional Syst. Signal Process.*, vol. 29, no. 3, pp. 979–998, Mar. 2017.

[24] J. Chu, H. Min, L. Liu, and W. Lu, "A novel computer aided breast mass detection scheme based on morphological enhancement and SLIC superpixel segmentation," *Med. Phys.*, vol. 42, no. 7, pp. 3859–3869, Jul. 2015.

[25] B. E. Bejnordi, G. Litjens, M. Hermsen, N. Karssemeijer, and J. A. W. M. van der Laak, "A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images," *Proc. SPIE*, vol. 9420, Mar. 2015, Art. no. 94200H.

[26] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2016, pp. 424–432.

[27] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.

[28] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1356–1363.

[29] G. Larsson, M. Maire, and G. Shakhnarovich, "FractalNet: Ultra-deep neural networks without residuals," 2016, *arXiv:1605.07648*. [Online]. Available: https://arxiv.org/abs/1605.07648

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[31] S. Karen, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[32] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*. [Online]. Available: https://arxiv.org/abs/1506.01497

[34] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fcholle/keras

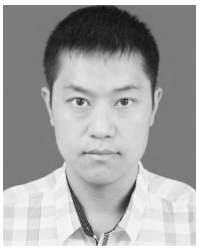
[35] G. Chlebus, H. Meine, J. H. Moltz, and A. Schenk, "Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering," 2017, *arXiv:1706.00842*. [Online]. Available: https://arxiv.org/abs/1706.00842

[36] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D. Anastasi, S.-A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. Menze "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," 2017, *arXiv:1702.05970*. [Online]. Available: https://arxiv.org/abs/1702.05970

[37] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," 2017, *arXiv:1704.07239*. [Online]. Available: https://arxiv.org/abs/1704.07239

ZHIQI BAI received the B.S. degree from the Department of Software Engineering, Northeastern University, China, in 2017, where he is currently a bachelor's degree with the Department of Software Engineering School. His research interests include medical image analysis and machine learning.

HUIYAN JIANG received the B.S. degree from the Department of Mathematics, Bohai University, China, in 1986, and the M.Sc. degree in computer application and the Ph.D. degree in control theory and control engineering from Northeastern University, China, in 2000 and 2009, respectively. From 2001 to 2002, she was a Visiting Scholar with Gifu University, Japan, and carried out the research of image processing and medical image computer-aided diagnosis (CAD) technology. She is currently a Professor and the Director of the Department of Digital Media Technology, Software College, Northeastern University, China; and the council member of the 3D Images Technology Association, China Society of Image and Graphics. Her main research interests focus on digital image processing and analysis, pattern recognition, 3D visualization, 3D video processing, artificial intelligence, and medical image computer-aided diagnosis (CAD).



SIQI LI received the B.S. degree from the Department of Mathematics and Software Sciences, Sichuan Normal University (SNU), China, in 2011, and the M.S. degree from the Department of Mathematics, Bohai University, China, in 2015. He is currently pursuing the Ph.D. degree with the Department of Software College, Northeastern University, China. His research interests include medical image detection, segmentation, classification, machine learning, and deep learning.

YU-DONG YAO received the B.Eng. and M.Eng. degrees from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree from Southeast University, Nanjing, in 1988, all in electrical engineering. From 1987 to 1988, he was a visiting student with Carleton University, Ottawa, Canada. From 1989 to 2000, he was with Carleton University, Spar Aerospace Ltd., Montreal, Canada, and Qualcomm Inc., San Diego, USA. Since 2000, he has been with the Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor and the Chair of the Department of Electrical and Computer Engineering. He holds one Chinese patent and 13 U.S. patents. His research interests include wireless communications, machine learning and deep learning techniques, and healthcare and medical applications. For his contributions to wireless communications systems, he was elected as a Fellow of the National Academy of Inventors (2015) and of the Canadian Academy of Engineering (2017). He has served as an Associate Editor for the IEEE COMMUNICATIONS LETTERS (2000–2008) and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2001–2006) and as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2001–2005).