CR-U-Net: Cascaded U-Net with Residual Mapping for Liver Segmentation in CT Images*

Yiwei Liu, Na Qi, Qing Zhu and Weiran Li
Faculty of Information Technology , Beijing University of Technology, Beijing, China
lyw1001@emails.bjut.edu.cn, {qina, ccgszq, liweiran}@bjut.edu.cn

***Abstract*—Abdominal computed tomography (CT) is a common modality to detect liver lesions. Liver segmentation in CT scan is important for diagnosis and analysis of liver lesions. However, the accuracy of existing liver segmentation methods is slightly insufficient. In this paper, we propose a liver segmentation architecture named CR-U-Net, which is composed of cascade U-Net combined with residual mapping. We make use of the MDice loss function for training in CR-U-Net, and the second-level of cascade network is deeper than the first-level to extract more detailed image features. Morphological algorithms are utilized as an intermediate-processing step to improve the segmentation accuracy. In addition, we evaluate our proposed CR-U-Net on liver segmentation task under the dataset provided by the 2017 ISBI LiTS Challenge. The experimental result demonstrates that our proposed CR-U-Net can outperform the state-of-the-art methods in term of the performance measures, such as Dice score, VOE, and so on.**

***Index Terms*—Liver segmentation, Residual mapping, U-Net, CT image, Convolutional neural networks**

I. Introduction

Liver is the largest substantive organ of human body [1]. It maintains important life activities such as detoxification and metabolism. Computed tomography (CT) is a common modality to detect liver lesions [2]. In these cases, liver segmentation in CT images is necessary in the diagnosis of liver lesions. It is important not only for the formulation of treatment plan, but also for the evaluation of the follow-up treatment effect [3]. Due to the manual segmentation is error-prone and time-consuming, automatic liver segmentation methods is extensively studied.

The existing automatic liver segmentation methods are divided into two types: image-based segmentation and statistical model-based segmentation. Thresholding [4], region growing [5] and graph cut [6] are some commonly used image-based segmentation methods, which directly segment images by the gray level, texture and gradient. Most of them have low robustness and easy to under or over segmentation. Therefore, in recent years, these methods are rarely applied to liver segmentation alone, but are usually used as a post-processing for other methods. The statistical model-based segmentation methods include traditional machine learning methods and deep learning methods, which usually achieve better results by binary classifying pixels with supervised learning than those image-based segmentation methods. However, traditional machine learning methods need to extract handcrafted image features, such as SVM [7] and Adaboost [8], which is not efficient. In contrast, deep learning methods extract image features automatically. Recent researches have witnessed the success of Convolutional Neural Networks (CNNs) [9] to solve the computer vision tasks, such as image classification, segmentation and object detection. Segmentation is a highly relevant task in medical image analysis [10], which can be applied to perform the computer assisted diagnosis, interventions and extraction of quantitative indices from images. Thus, in this paper, we focus on the liver segmentation in CT images based on CNNs.

Thanks to the 2017 ISBI LiTS Challenge for providing the CT dataset.

FCN [11] and U-Net [12] are both fully convolution network, which are successfully used in liver segmentation. Christ *et al.* [13] use cascade FCN of 2 levels for liver and liver tumor segmentation. The first level is used to segment the liver, and the second level is used to segment the liver tumor. The 3D conditions random filed (CRF) is the post-processing method. Liu *et al.* [14] propose IU-Net to improve the U-Net. They directly apply the max pooling layers to the deconvolution layers in order to reduce the loss of image features. And finally, IU-Net makes better result than original U-Net.

Currently, residual mapping [15] is used in combination with image segmentation architectures, which is an effective way to prevent overfitting and meanwhile to improve accuracy. Milletari *et al.* [10] combine residual learning with U-Net to construct V-Net for 3D image segmentation. Han [16] proposes a 2.5D cascaded U-Net combined with residual mapping (DCNN) for the segmentation of liver and liver tumor in CT images. The input of DCNN is one central slice and several adjacent axial slices of CT data. The output is a 2D segmentation map corresponding to the input data. The method of Han wins the 1st place in the 2017 ISBI LiTS Challenge. Vorontsov *et al.* [18] employ residual blocks similar to ResNet [15] combined with 2 FCN architectures for liver and liver tumor segmentation, which gets the 2nd place in the 2017 ISBI LiTS Challenge. The 2 FCN architectures are trained together end-to-end with the same input data. The FCN 1 is trained to segment the liver and the FCN 2 is trained to segment the tumor, meanwhile the output of FCN 1 is passed as an additional input to FCN 2.

Although these existing methods achieve good performance for the liver segmentation, it still need to be improved in order to make the segmentation more effectively. In this paper, CR-U-Net is proposed refer to the work of Han [16] to improve the segmentation accuracy. We adopt a cascaded network with different depths, and we add an intermediate-processing step into the cascaded network to restore the mis-segmented pixels of the positive sample. This paper is organized as follows. Section II introduces the architecture and loss function of CR-U-Net. Section III presents the details and the segmentation results of our experment. CR-U-Net is evaluated on a specific dataset through liver segmentation performance measures and visual results. Finally, the paper is concluded in the Section IV.

II. The Proposed CR-U-Net

A. CR-U-Net

The architecture and segmentation process of CR-U-Net is illustrated in Fig. 1. Residual mapping is combined with U-Net to constitute R-U-Net, which contains down-sampling paths and up-sampling paths that are also utilized in the U-Net. Here, image features are extracted in down-sampling path and located in up-sampling path. R-U-Net is composed of several unit blocks, whose structures are shown in Fig. 2. Three convolution layers with the filter of size 3×3 and one max-pooling layer with the kernel of size 2×2 are taken as a down-sampling unit block. One deconvolution layer and three convolution

978-1-7281-3723-0/19/$31.00 ©2019 IEEE

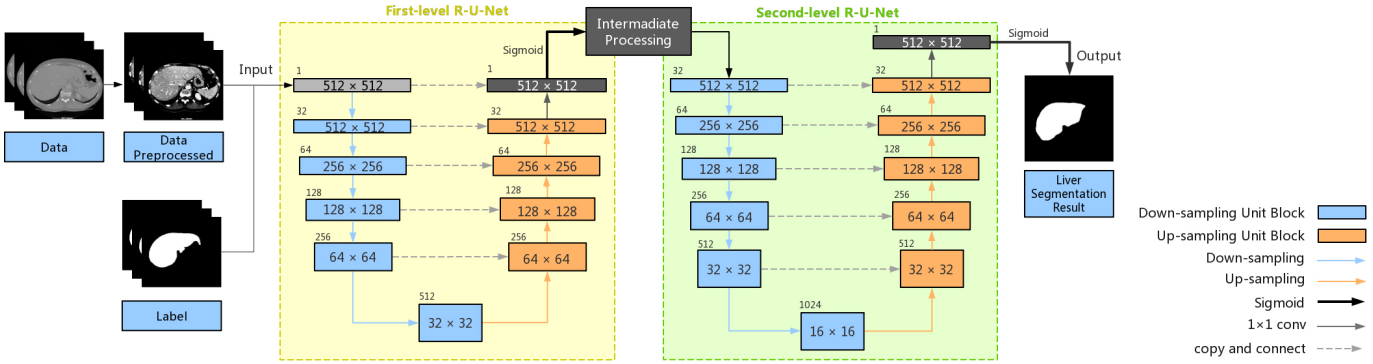


Fig. 1. The architecture and segmentation process of CR-U-Net. Residual mapping is combined with U-Net to constitute a R-U-Net. CR-U-Net is composed of cascaded R-U-Net and an intermediate-processing step. The intermediate-processing restores the positive sample pixels that are mis-segmented by the first-level R-U-Net. The second-level R-U-Net is utilized to extract more features. The pre-processed CT slice images and segmentation label images are as input. The liver segmentation prediction map is as the output.

layers with the filter of size 3×3 are taken as a up-sampling unit block. The strides of all layers are 2. Each convolution layer contains the sub-layers of weight, convolution, batch normalization, ReLU function and Dropout. Each deconvolution layer should be connected to the 3rd convolution layer with the down-sampling unit block of the same dimension. Moreover, each unit block contains a residual mapping structure. He *et al.* argue that the order of layers in residual mapping has a great influence on the accuracy [18]. In our proposed CR-U-Net, we utilize the residual structure proposed in [18] as the residual mapping after batch normalization, as shown in Fig.2.

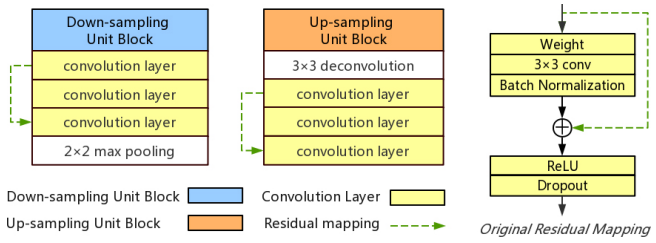


Fig. 2. Unit block of down-sampling and up-sampling, and residual mapping of CR-U-Net.

CR-U-Net is composed of two-levels of cascaded R-U-Net. The first-level R-U-Net is composed of 4 down-sampling unit blocks and 4 up-sampling unit blocks, which are total 35 layers. The input size is 512×512. The output of this level is transformed into the same dimension as the input data by convolution layer with the filter of size 1×1. The intermediate-processing is implemented after Sigmoid function of the first-level R-U-Net, which contains an opening operation with radius 5 and a closing operation with radius 5. Morphological technique is commonly used for the boundary refinement. However, making it as a post-processing method may lose some ROI (positive sample) pixels. In CR-U-Net, we take a morphological technique as an intermediate-processing unit. Some pixels are mis-segmented in the training of the first-level R-U-Net. Through the intermediate-processing, the missing ROI pixels will be restored and the tiny interfering pixels of the background will be removed. These pixels are hardly to be mis-segmented again in the next training. The closing operation with radius 5 is adopted to fill the under-segmentation part inside the liver area, and the opening operation with radius 5 is used to eliminate the interfering pixels outside the liver area.

After this, the output data is employed as a new input data to the second-level R-U-Net with the same size as the first-level. There are 5 down-sampling unit blocks and 5 up-sampling unit blocks in the second-level R-U-Net, which are total 43 layers. More image features can be extracted in this level. Finally, the predicted liver segmentation map of CR-U-Net is output after a 1×1 convolution layer and Sigmoid function of the last up-sampling.

B. Loss Function

The loss function has a significant impact on the performance of CNN. In medical image segmentation, ROI only covers a small region, which leads that the loss function draftly fall into a local minimum when training, and thus makes the segmentation deviate greatly [10]. Milletari *et al.* propose the Dice loss function in the 3D image segmentation task [10]. This loss function is effective in reducing the segmentation deviation caused by the imbalance of ROI region and background. The Dice loss function proposed in [10] is formulated as follows,

$$Dice\ loss = 1 - \frac{2|PG|}{|P|^2 + |G|^2} \tag{1}$$

where P is the predicted binary partition volume and G is the standard binary partition volume. The predictive performance of the model is judged by measuring the proportion of overlapping parts of P and G.

However, because of the false positive (FP) and false negative (FN) detections have the same weight in the Dice loss function, the small ROI can not be accurately predicted. So, when a small positive sample is mis-segmented, the loss value will ascend rapidly. Therefore, it is necessary to increase the weight of FN for Dice loss function in order to improve the recognition ability of the network for small ROI pixels. Tversky coefficient [19] is a generalized Dice score, which allows the weight of FN and FP to be adjusted flexibly through hyperparameters. In this paper, we formulate a multiplicative Dice loss function, denoted as a MDice loss, by multiplying the Dice score with the Tversky coefficient for our proposed CR-U-Net.

The Dice score and Tversky coefficient are respectively formulated as follows,

$$D(P,G) = \frac{2|PG|}{|P| + |G|} = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i} \tag{2}$$

$$T(P,G) = \frac{|PG|}{|PG| + \alpha|P \setminus G| + \beta|G \setminus P|} = \frac{\sum_i^N p_i g_i}{\sum_i^N p_i g_i + \alpha \sum_i^N p_{\bar{i}} g_i + \beta \sum_i^N p_i g_{\bar{i}}} \tag{3}$$

where (2) is the Dice score, (3) is the Tversky coefficient. To avoid the loss gradient disappearing, we add a smooth paramater ϵ refer to [20] to our MDice loss, which is formulated as (4),

$$\begin{aligned} MDice\ loss &= 1 - D(P,G) \cdot T(P,G) \\ &= 1 - \frac{2\sum_i^N p_i g_i \cdot \sum_i^N p_i g_i + \epsilon}{(\sum_i^N p_i + \sum_i^N g_i)(\sum_i^N p_i g_i + \alpha \sum_i^N p_{\bar{i}} g_i + \beta \sum_i^N p_i g_{\bar{i}}) + \epsilon} \end{aligned} \tag{4}$$

where P is the segmentation predicted probability image, p_i is the probability of pixel i be a positive region, $p_{\bar{i}}$ is the probability of pixel i be a negative region. G is the ground truth binary label, g_i is the positive pixel of G, $g_{\bar{i}}$ is the negative pixel of G. N is the total number of pixels in each input data. Meanwhile, $FP = |P \backslash G|$, $FN = |G \backslash P|$. Hyperparameters α and β can be flexibly adjusted to change the weight of FP and FN, and further improve the accuracy of the segmentation. Both of these hyperparameters is range in [0,1]. We tested a number of sets of hyperparameters α and β. It was observed that the parameter of α = 0.5 and β = 0.7 will lead to the best training results. Therefore, all models with MDice loss were trained with these hyperparameters in our experiment.

III. Experimental Result

A. Dataset and Pre-processing

The experiments were conducted on the dataset provided in 2017 ISBI LiTS Challenge. The dataset contains abdominal CT data of 130 patients as training-set and abdominal CT data of 70 patients as test-set. The number of slices in CT data ranges from 75 to 987, and the size of each slice is 512×512 pixels. The liver segmentation ground truth is hand-drawn by experts.

As the pre-processing step, the intensity of HU was unified within the range of [-200,250] to remove the details of irrelevant organs and objects, and the contrast and sharpness of CT images were enhanced by the histogram equalization.

B. Training and Test

The experiments were conducted using the Geforce GTX1070 GPU with 8GB memory. CR-U-Net was constructed by using the Tensorflow framework [21] and trained by forward propagation with batch size of 2 and learning rate of 1^{-4}-1^{-10}, totaling 20,000 epochs. Xavier method [22] was adopted for weight initialization, and the bias was initialized to 0.10. In order to get a credible result, ten-fold cross-validation [23] was implemented.

The performance measures of CR-U-Net was evaluated by the quality measures proposed in the 2017 ISBI LiTS Challenge, including the Dice score (DICE), the volume overlap error (VOE), the relative volume difference (RVD), the average symmetric surface distance (ASSD) and the maximum surface distance (MSD) [13].

C. Result

In order to verify the performance of CR-U-Net, architectures of superior existing researches and all the segmentation networks generated during our experimental process were implemented. The quantitative results are reported in TABLE I, and the visual results are shown in Fig. 3. We rebuilt the architectures of Han [16] who won the first place in the ISBI LiTS challenge and the work of Liu *et al.* [14]. The existing segmentation models were trained using their own loss function. The architectures generated in our experimental process were trained using the MDice loss with $\epsilon = 1^{-5}$. To verify the validity of the MDice loss, CR-U-Net was trained with the Dice loss and the MDice loss independently for comparison.

It is reported in TABLE I that the performance of CR-U-Net with the MDice loss is the best in terms of DICE, VOE and other measures, which proves that CR-U-Net with the MDice loss is the most accurate. Meanwhile, as shown in Fig. 3, the visual results of CR-U-Net with the MDice loss are the most similar as the ground truth. By observing and comparing experimental results, we believe that the two improvements for cascaded networks: deepening the depth of the second-level network in the cascade network, adding the intermediate-processing in the cascade network, are both effective to improve the segmentation accuracy. However, it can be seen in Fig. 3(a)(9) and Fig. 3(b)(9) that a few tiny negative samples are under-segmented, due to the morphological algorithm or the unit block numbers of CR-U-Net architecture.

In addition, it is shown in TABLE I that the DICE of CR-U-Net with the MDice loss is significantly higher than that of CR-U-Net with the Dice loss. Meanwhile, it is obviously in Fig. 3 that the results of Fig. 3(a)(9) and Fig. 3(b)(9) are much better than those in Fig. 3(a)(8) and Fig. 3(b)(8). So, it can be inferred that the MDice loss is effective in increasing the liver segmentation accuracy. However, the CR-U-Net using the MDice loss as loss function requires to determine the optimal α and β, which takes some time. Therefore, in the future, we will adaptively determmine the hyperparameters and optimize the loss function to reduce the training time.

TABLE I
THE SEGMENTATION RESULTS OF CR-U-NET AND OTHER METHODS

Model	Performance Measures				
	***DICE* (%)**	**VOE**	**RVD**	***ASSD* (mm)**	***MSD* (mm)**
U-Net [12]	93.27	0.076	0.045	6.728	34.378
FCN [11]+CRF	94.35	0.088	0.082	5.917	32.322
Liu Z *et al.* [14]	94.80	-	-	-	-
Han X [16]	96.25	0.023	0.023	3.669	28.178
R-U-Net	95.76	0.042	0.043	4.473	28.815
CR-U-Net[a]	96.51	0.036	0.036	4.049	27.493
CR-U-Net[b]	96.52	0.039	0.039	3.910	27.830
CR-U-Net[c]	95.42	0.033	0.039	7.447	33.513
CR-U-Net[d]	**97.16**	**0.023**	**0.020**	**3.099**	**25.628**

[a]CR-U-Net with 2 levels network of the same depth.
[b]CR-U-Net without the intermediate-processing.
[c]CR-U-Net with the Dice loss.
[d]**CR-U-Net with the MDice loss.**

IV. Conclusion

In this paper, we propose a CR-U-Net for automatic liver segmentation in CT images, where CR-U-Net is cascading by two-levels of R-U-Net and an intermediate-processing unit. The R-U-Net is constituted by U-Net with residual mapping . We formulate a general MDice loss function for the network learning. Due to the deepened R-U-Net and the closing as well as the opening operation in the intermediate-processing unit, our proposed CR-U-Net can provide the best segmentation results. Experimental results on the test dataset demonstrate that our CR-U-Net with the MDice loss can present the best results compared with the state-of-the-art methods. In the future, we will apply our proposed CR-U-Net for other medical image datasets to further verify the effectiveness of our proposed network.

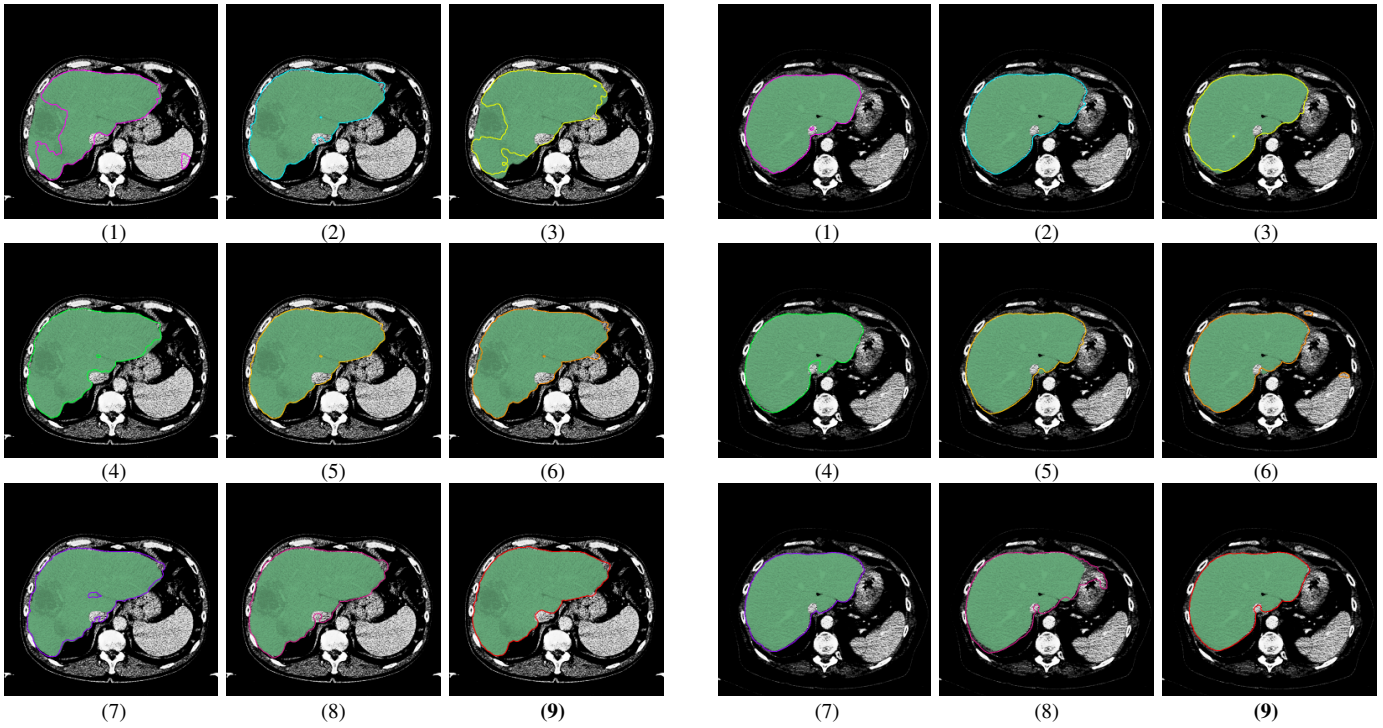
V. Acknowledgement

This work was supported by NSFC (No.61632006, 61672066, 61976011, 61906008, 61906009), Beijing Municipal Science and Technology Project (Z171100004417023), China Postdoctoral Science Foundation funded project (No. 2017M620558), Beijing Postdoctoral Research Foundation (No. 2017-ZZ-025) and the Chaoyang District Postdoctoral Research Foundation (No.2018ZZ-01-02).

References

[1] A.I. Ahmed, B.A.A. Alla, and M.E.M. Gar-Elnabi,"Quantitative Analysis and Characterization of Liver Diseases using Ultrasound Scans," International Journal of Science and Research, 2013, pp. 132–136.

[2] A. Ben-Cohen, I. Diamant, and E. Klang, "Fully Convolutional Network for Liver Segmentation and Lesions Detection," Deep Learning and Data Labeling for Medical Applications, Springer International Publishing, 2016, pp. 77–85.



(a) Results 1 of different segmentation methods

(b) Results 2 of different segmentation methods

Fig. 3. Comparisons of CR-U-Net with other segmentation methods: (1) U-Net, (2) FCN + CRF, (3) The method of Liu Z *et al.*, (4) The method of Han X, (5) R-U-Net, (6) CR-U-Net with 2 levels network of the same depth, (7) CR-U-Net without the intermediate-processing, (8) CR-U-Net with the Dice loss, (9) CR-U-Net with the MDice loss. (a) and (b) are the segmentation result of different methods in 2 test images. (1)∼(4) are the result of the architectures of existing research. (5)∼(9) are the result of the segmentation network produced in the process of our experiment, which all trained using the MDice loss. The green area of each image is the ground truth. The results of different methods are represented with wireframes of different colors. It is observed that the result of CR-U-Net with the MDice loss is the best.

[3] A.M. Mharib, A.R. Ramli, and S. Mashohor, "Survey on liver CT image segmentation methods," Artificial Intelligence Review, vol. 37, pp. 83–95, February 2017.

[4] H. Jiang, and Q. Cheng, "Automatic 3D Segmentation of CT Images Based on Active Contour Models," 11th IEEE international conference on computer-aided design and computer graphics, 2009, pp. 540–543.

[5] X. Lu, J. Wu, and X. Ren, "The study and application of the improved region growing algorithm for liver segmentation," Optik - International Journal for Light and Electron Optics, vol. 125, pp. 2142–2147, September 2014.

[6] Y. Chen, Z. Wang, and J. Hu, "The domain knowledge-based graph-cut model for liver CT segmentation," Biomedical Signal Processing & Control, vol. 7, pp. 591–598, June 2012.

[7] J.S. Jin, and S.K. Chalup, "A Liver Segmentation Algorithm Based on Wavelets and Machine Learning," International Conference on Computational Intelligence and Natural Computing, 2009, pp. 122–125.

[8] X. Li, C. Huang, and F. Jia, "Automatic Liver Segmentation Using Statistical Prior Models and Free-form Deformation," International MICCAI Workshop on Medical Computer Vision, Springer, Cham, vol. 8848, 2014, pp. 181–188.

[9] Y. LÉcun, L. Bottou, and Y. Bengio, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, pp. 2278–2324, November 1998.

[10] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 3D Vision (3DV), 2016 Fourth International Conference on, IEEE, 2016, pp. 565–571.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2015, pp. 3431-3440.

[12] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, pp. 234–241.

[13] P.F. Christ, M.E.A. Elshaer, and F. Ettlinger, "Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields," International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016, pp.415–423.

[14] Z. Liu, X.L. Zhang, and Y.Q. Song, "Liver segmentation with improved U-Net and Morphsnakes algorithm," Journal of Image and Graphics, vol. 23, pp. 1254–1262, August 2018.

[15] K. He, X. Zhang, and S. Ren, "Deep Residual Learning for Image Recognition," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[16] X. Han, "Automatic Liver Lesion Segmentation Using A Deep Convolutional Neural Network Method," arXiv preprint arXiv:1704.07239, 2017.

[17] E. Vorontsov, A. Tang, and C. Pal, "Liver lesion segmentation informed by joint liver segmentation," arXiv preprint arXiv:1707.07734, 2017.

[18] K. He, X. Zhang, and S. Ren, "Identity Mappings in Deep Residual Networks," European Conference on Computer Vision (ECCV), Springer International Publishing, 2016, pp. 630–645.

[19] S.S.M Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," arXiv preprint arXiv:1706.05721, 2017.

[20] A. Nabila, and N.M. Khan, "A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation," arXiv preprint arXiv:1810.07842, 2018.

[21] M. Abadi, P. Barham, and J. Chen, "TensorFlow: a system for large-scale machine learning," OSDI, vol. 16, pp. 265-283, November 2016.

[22] X. Glorot, and Y Bengio, "Understanding the difficulty of training deep feedforward neural networks," Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), 2010, pp. 249–256.

[23] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification 2nd Edition, John Wiley and Sons, New York, 2001.