



Three Uses of One Neural Network: Automatic Segmentation of Kidney Tumor and Cysts Based on 3D U-Net

Yi Lv^{1,2} and Junchen Wang^{2,3} (✉)

¹ North China Research Institute of Electro-Optics, Beijing 100015, China

² School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China

wangjunchen@buaa.edu.cn

³ Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing, China

Abstract. Medical image processing plays an increasingly important role in clinical diagnosis and treatment. Using the results of kidney CT image segmentation for three-dimensional reconstruction is an intuitive and accurate method for diagnosis. In this paper, we propose a three-step automatic segmentation method for kidney, tumors and cysts, including roughly segmenting the kidney and tumor from low-resolution CT, locating each kidney and fine segmenting the kidney, and finally extracting the tumor and cyst from the segmented kidney. The results show that the average dice of our method for kidney, tumor and cysts is about 0.93, 0.57, 0.73.

Keywords: Medical image segmentation · Deep learning · Neural network

1 Introduction

Segmentation and reconstruction of CT or MRI medical images is the main source of navigation data, i.e. anatomical structure of tissues and organs [1–4]. At present, the most commonly used method of medical image segmentation is still manual segmentation, which takes a long time and depends on the operator's experience and skills [5]. In recent years, breakthroughs have been made in the research of neural networks [6–9]. The deep learning technology based on neural networks can achieve fast segmentation, and effectively solve the problem of low accuracy and long time-consuming image segmentation [10]. In the field of medical image segmentation, the breakthrough of in-depth learning began with the introduction of Full Convolutional Neural Network (FCN), and another breakthrough of neural network architecture U-Net made it possible to achieve high-precision automatic segmentation of medical images. Long Jonathan et al. [11] proposed Fully Convolutional Networks structure in 2015. In the same year, Olaf Ronneberger et al. [12] proposed the U-Net network structure. U-Net is a semantic

segmentation network based on FCN, which is suitable for medical image segmentation. With the proposal of 3D convolutional neural networks such as 3D U-Net [13] and V-Net [14], the segmentation accuracy of some organs has reached a milestone. For example, in the MICCAI challenge 2019 kits19 competition, the accuracy of 3D U-Net in the task of kidney segmentation is very close to that of human, but the required time to complete a segmentation is far less than that of manual segmentation. The deep learning-based methods not only surpass the traditional algorithms, but also approach the accuracy of manual segmentation.

In this paper, we propose a three-step automatic segmentation method for kidney, tumors and cysts, including roughly segmenting the kidney and tumor from low-resolution CT, locating each kidney and fine segmenting the kidney, and finally extracting the tumor and cyst from the segmented kidney.

2 Methods

2.1 Network Architecture

The network architecture (3D U-Net) is illustrated in Fig. 1. We choose 3D U-Net as the neural network for the all three steps. 3D U-net includes an encoding path and a decoding path, each of which has four resolution levels. Each layer of the encoding path contains two $3 \times 3 \times 3$ convolution, each followed by a ReLu layer, followed by a $2 \times 2 \times 2$ Maximum pool layer with step size of 2 in each direction of 2. In the decoding path, each layer contains a 2 with a step size of $2 \times 2 \times 2$, followed by two $3 \times 3 \times 3$, each followed by a RuLu layer.

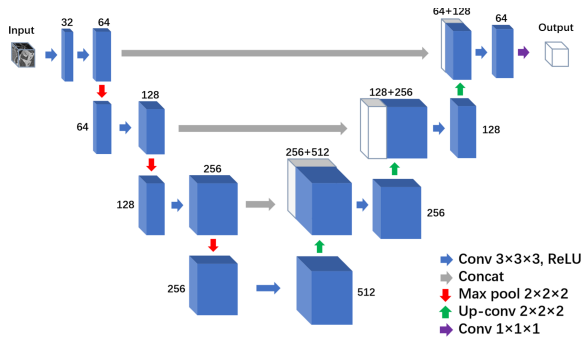


Fig. 1. The network architecture. Blue cuboids represent feature maps. The number of channels is denoted next to the feature map. (Color figure online)

2.2 Segmentation from Low-Resolution CT

Considering that the complete CT sequence is too large for the GPU, we scaled the spacing of all CT sequences to 3.41, 1.71 and 1.71. Then we cut the size of the sequence to $128 * 248 * 248$ under the condition of constant spacing. Finally, we uniformly scale all the sequence sizes to $96 * 128 * 128$ for the training of neural network. Then we use the generated data for training of the 3D U-Net for the first step.

2.3 Fine Segmentation of Kidney

According to the results of rough segmentation in the previous step, we first locate the position of two (or one) kidneys. Secondly, we crop the CT image with the region located and scale the size to $80 * 128 * 128$. Then, we set the kidney, tumor and cyst to a unique label as the training data for the second step. Finally, we conduct the data augmentation including random translation, scaling, and trained the second 3D U-Net with the generated data.

2.4 Segmentation of Tumor and Cysts

To prevent the network from identifying areas outside the kidney as tumors or cysts, only the region of kidney segmented in the above step is used as the input for the network in step 3. Tumor and cysts were labeled with 1 and 2, respectively. After segmentation, we scale the segmented results to the original size as the final output.

2.5 Training Protocols

All the algorithms were implemented using Pytorch1.2 with Python3.7 and ran on a workstation with a AMD 5800X CPU, 32G memory and a NVIDIA RTX8000 GPU. The training protocols of the proposed method is shown in Table 1.

Table 1. Training protocols

Data augmentation methods	Scaling, rotations, brightness, contrast, gamma
Initialization of the network	Kaiming normal initialization
Batch size	4
Total epochs	50
Loss function	Dice loss and weighted cross entropy

3 Results

As the accuracy metrics, the Dice similarity coefficient (DSC), average symmetric surface distance (ASSD [mm]) and surface distance deviation (SDD [mm]) between the predicted mask and the ground truth mask were employed. Assume A and B are two masks, these metrics are given by (1), (2) and (3), where $S(A)$ and $S(B)$ are the surface points of A and B, respectively, and $d(a, S(B))$ is the minimum Euclidian distance between the point a and the points on the surface $S(B)$. We perform 5-fold cross-validation on 300 data sets, but due to limited time, only one-fold has finished at the time of submission. The result is shown in Table 2. Figure 2 shows the results with voxel-based rendering

from three examples in the evaluation dataset. Figure 3 shows the segmentation results on CT slices.

$$\text{DSC} = \frac{2(A \cap B)}{A + B} \quad (1)$$

$$\text{ASSD} = \frac{1}{|S(A)| + |S(B)|} \times \left(\sum_{a \in S(A)} d(a, S(B)) + \sum_{b \in S(B)} d(b, S(A)) \right) \quad (2)$$

$$\text{SSD} = \sqrt{\frac{1}{|S(A)| + |S(B)|} \times \left(\sum_{a \in S(A)} (d(a, S(B)) - \text{ASSD})^2 + \sum_{b \in S(B)} (d(b, S(A)) - \text{ASSD})^2 \right)} \quad (3)$$

Table 2. Dice comparison on three structures

	Dice	ASSD (mm)	SSD (mm)
Kidney	0.93	1.24	2.45
Tumor	0.57	8.24	10.86
Cysts	0.73	6.70	6.90

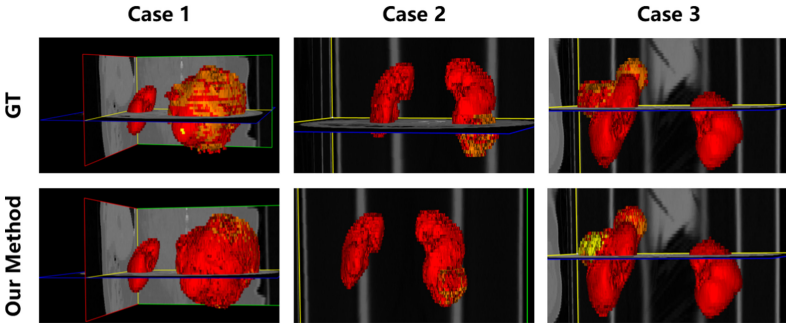


Fig. 2. Segmentation results with voxel-based rendering from three examples in the evaluation dataset. For each example, the ground truth and the segmentation results are given for visual comparison. (yellow: cysts, red: kidney, brown: tumor) (Color figure online)

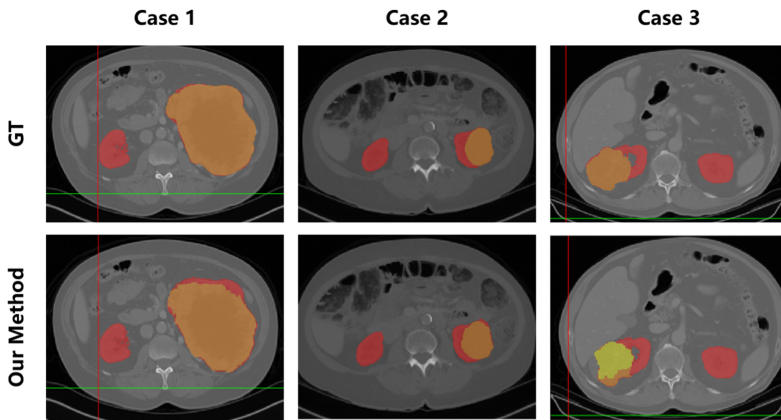


Fig. 3. Segmentation results on CT slices. The two rows are the ground truth and segmentation results of our method. (yellow: cysts, red: kidney, brown: tumor) (Color figure online)

4 Discussion and Conclusion

We propose a three-step automatic segmentation method for kidney, tumor and cysts based on 3D U-Net. The results show that the average dice of our method for kidney, tumor and cysts is about 0.93,0.57,0.73. According to the final result, the average Dice, Surface Dice and Tumor Dice are respectively 0.781, 0.618, 0.660. The results show that the accuracy of our method on kidney is better than that of tumor and cysts. The region of the kidney can be accurately identified, but the accuracy of tumor and cysts is not satisfactory. As limited by the competition time, the neural network requires more time to be fully trained. Future work will focus on promoting accuracy of our method on the tumors and cysts.

Acknowledgements. Junchen Wang was funded by National Key R&D Program of China (2017YFB1303004) and National Natural Science Foundation of China (61701014, 61911540075).

References

1. Couteaux, V., et al.: Kidney cortex segmentation in 2D CT with U-Nets ensemble aggregation. *Diagn. Interv. Imaging* **100**(4), 211–217 (2019)
2. Yang, Y., Jiang, H., Sun, Q.: A multiorgan segmentation model for CT volumes via full convolution-deconvolution network. *Biomed Res. Int.* **2017**, pp. 1–9 (2017)
3. Fu, Y., et al.: A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. *Med. Phys.* **45**(11), 5129–5137 (2018)
4. Zhao, C., Carass, A., Lee, J., He, Y., Prince, J.L.: Whole brain segmentation and labeling from CT using synthetic MR images. In: Wang, Q., Shi, Y., Suk, H.-I., Suzuki, K. (eds.) *MLMI 2017*. LNCS, vol. 10541, pp. 291–298. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67389-9_34

5. Li, J., Zhu, S.A., Bin, H.: Medical image segmentation techniques, *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi* **23**(4), 891–894 (2006)
6. Micheli-Tzanakou, E.: Artificial neural networks: an overview. *Netw.-Comput. Neural Syst.* **22**(1–4), 208–230 (2011)
7. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
8. Zhang, J., Zong, C.: Deep neural networks in machine translation: an overview. *IEEE Intell. Syst.* **30**(5), 16–25 (2015)
9. Zarandy, Á., Rekeczky, C., Szolgay, P., Chua, L.O.: Overview of CNN research: 25 years history and the current trends. In: 2015 IEEE International Symposium on Circuits and Systems, pp. 401–404. IEEE (2015)
10. Kim, D.Y., Park, J.W.: Computer-aided detection of kidney tumor on abdominal computed tomography scans. *Acta Radiol.* **45**(7), 791–795 (2004)
11. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(4), 640–651 (2017)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
14. Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of 2016 Fourth International Conference on 3D Vision, pp. 565–571. IEEE (2016)