

Fully Automatic Segmentation of Kidney and Tumor Based on Cascaded U-Nets

Yu Li

College of Science, Nanjing University of Science and Technology, Nanjing 210000,
P.R.China
liyu910118@sina.com

Abstract. Fully automatic segmentation of kidney and its lesions is an important step to obtain accurate clinical diagnosis and computer aided decision support system. In this paper, a method of automatic segmentation of kidney and renal tumor in CT abdominal images using cascade 3D U-Net convolutional neural network (3D cU-Nets) is presented. We trained and cascaded two 3D U-Nets for the joint segmentation of kidney and renal tumor. In the first step, we trained a 3D U-Net to segment kidney as the ROI input for the second 3D u-net. The second 3D U-Net only segmented the lesion from the renal ROI predicted in step 1.

Keywords: Kidney · renal tumor · segmentation · 3D U-Net · Deep Learning;

1 Introduction

With the development of CT, magnetic resonance imaging and other imaging equipment technology, the role of medical images in clinical diagnosis and treatment is becoming more and more significant. How to make a perfect surgical plan according to the assessment of the anatomical structure and functional reserve of viscera and the judgment of the invasion range of target lesions is always the pursuit of precision surgery in the industry. This refers to image segmentation in the medical field. Medical image segmentation is a basic problem in medical image analysis. It can recognize the boundaries of objects, such as organs in images or abnormal areas (such as tumors). The obtained segmentation results allow it to be used for shape analysis, detection of volume changes, and the development of accurate radiotherapy plans.

Kidney tumor is a common clinical disease in urology, which can be divided into benign and malignant types. Among them, malignant tumors accounted for the majority, about Ninety-five percent. Generally, B-ultrasound or CT examination can basically determine whether the kidney tumor is benign or malignant, but to determine the specific type of tumor can only rely on the kidney biopsy method (referred to as renal biopsy).

Tracing back to the development history of medical image segmentation technology, its development can be summarized into three stages: manual segmentation, semi-automatic segmentation and automatic segmentation. However, man-

ual segmentation and semi-automatic segmentation are both subjective, operator dependent and time-consuming. Due to different judgment criteria and subjective opinions on the pathological tissues of patients in medical images, the judgment results of different doctors at different moments may be more or less different from the real value. In recent years, with the development of image processing and analysis technology, people try to provide objective reference and decision support for doctors through computer, so as to improve the limitations of doctors' empirical interpretation as much as possible. Therefore, computer-aided Detection/Diagnosis (CAD) based on medical imaging has been gradually developed [1,2]. However, some difficulties in the automatic segmentation of kidney and lesion area remain unsolved, such as low contrast between kidney and lesion area, different types of contrast level, abnormality, size and different amount of lesion area of organ tissue.

Nevertheless, various methods have been developed to segment the kidney and renal lesion areas in CT images. In 2009, Shim [3] et al. adopted the comprehensive method of regional growth and graph cutting theory to realize the segmentation of renal CT images. In order to manage and monitor renal tumor tissues in CT images perfectly, Linguraru [4,5] et al. proposed a computer-aided clinical diagnostic tool to estimate and classify them. Spiegel [6] et al. used active shape model to segment kidney and adopted non-rigid registration to solve the problem of point correspondence between training samples. However, due to their low contrast between heterogeneous processing speed and robustness and real CT data, these methods are not widely used in clinical practice. Therefore, Deep Learning methods have been developed to overcome these shortcomings.

Deep convolutional neural network CNN has gained new attention from the scientific community in solving computer vision tasks such as target recognition, classification and segmentation [7,8], which are often more competitive than existing methods. Literature [9] developed a semantic image segmentation method based on convolutional neural network FCN, and achieved remarkable results in the natural image segmentation competition [10].

In the past, CNN was usually used for classification, while biomedical images focused more on segmentation and positioning. And when dealing with large, high-resolution 3D data, a common problem is that the input to the deep CNNs is either cropped or downsampled due to the limited memory capacity of the computing device. These operations lead to a decrease in the resolution of the input data and an increase in the class imbalance, thus reducing the performance of the segmentation algorithm. Therefore, an improved medical image segmentation method based on CNN and FCNs [11,12] have also been developed. In literature [13], Milletari et al. has introduced V-Net as well as an unique and well performed objective function to do volumetric segmentation from MRI. A successful adaptation of U-Net from 2D to a 3D network has been also shown in literature [13,14]. Compared with the most advanced mathematical segmentation method, it has strong competitiveness. Most importantly, the 3D U-Net method [13] has been proved to be very robust to different image shapes, which

prompted us to apply it to the automatic segmentation of kidney and tumor in CT images.

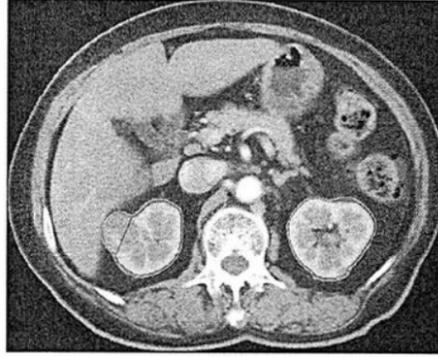


Fig. 1. right nephropathy

The CT images shown in Fig. 1 illustrates the irregularity of tumor shape, the low contrast with normal tissues, and the similarity of gray scale with surrounding tissues and organs. Therefore, tumor segmentation in CT images is a challenging challenge internationally.

In this work, we present an approach to automatically segment the kidney and its lesions in low-contrast non-uniform CT images. First of all, we performed the training and application of convolutional network 3D u-net on renal CT images for the first time, demonstrating the adaptability of challenging segmentation for renal lesions. Secondly, it is recommended to use cascade convolutional neural network (3D cU-Net) [15] on CT slices to segment the kidney and renal tumors in sequence, so as to achieve a higher segmentation accuracy and significantly improve the segmentation quality while maintaining a lower computational cost and memory consumption.

2 Method

In the next section, we represent the volume of the 3D image as I , the total number of voxels as N , and the possible tag set as $L = \{0, 1, \dots, l\}$. For each individual voxel i , we define a variable $x_i \in L$ that represents the assigned label. Given image I , the probability that voxel i belongs to tag k is described by $P(x_i = k | I)$ and will be modeled by 3D U-Net. In our special study, we used $L = \{0, 1, 2\}$ as background, kidney and renal tumor, respectively.

2.1 Data preprocessing

Preprocessing is done in a piecemeal manner. Firstly, the data in the data group is repositioned to get a new data group that is redirected, and the data size

is a 4 by 4 matrix. The new data group data is then repositioned to the standard view. Next, the unit value window of Hounsfield was set within the range of $[-160,240]$ to exclude irrelevant organs and objects, and then the CT image was initially segmented through threshold segmentation. The obtained training data set is used to train two cascade 3D U-Nets based on U-Net architecture [16].

2.2 Cascade convolutional neural network (3D CU-Net)

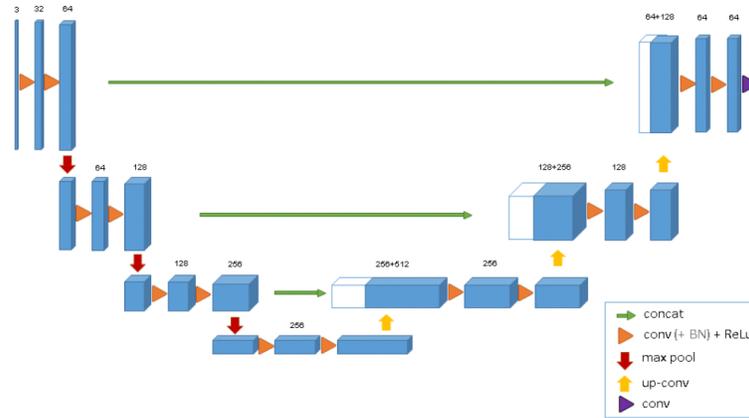


Fig. 2. 3D U-Net

Figure 2 illustrates the network architecture. Like the standard U-Net, it has a compression and expansion path that can be viewed as an encoder on the left and a decoder on the right. The encoder has four submodules. In the compression path, each submodule contains two $3 \times 3 \times 3$ convolutional layers, each convolutional layer has a modified linear unit (ReLU), and each submodule has a lower sampling layer realized by Max pool, in which the maximum pool is $2 \times 2 \times 2$, and the step size of each dimension is 2. In the extended path, each layer is composed of Max pool of $2 \times 2 \times 2$, with the step size of 2 for each dimension, and then there are 2 convolution layers of $3 \times 3 \times 3$, and each convolution layer is ReLU. In the last layer, $1 \times 1 \times 1$ convolution reduces the number of output channels to the number of labels. In figure 2, the left side is the down-sampling process, the right side is the up-sampling process, the green arrow in the middle connects the feature map with the skip layer, which can fuse the positioning information of the shallow layer with the classification and determination information of the pixels at the high level, so as to get better results.

In our method, we trained a 3D U-Net neural network to segment the kidney in the abdominal section to obtain an image of the kidney (step 1). Another 3D U-Net neural network was used to segment the kidney tumor to further obtain

the kidney tumor (step 2). The kidney segmented in step 1 was cropped and re-sampled to the required input size of the cascade u-net in step 2.

The motivation behind the cascading approach is that 3D U-Nets and other forms of CNNs learn to provide hierarchical representations of the data. Stacked convolution filter layers are tailored to the desired classification in a data-driven manner, rather than being designed manual characteristics to separate the different organizational types. By cascading two 3D U-Nets, we ensured that the convolutional neural network in step 1 learned from the overall abdominal CT scan filters specifically for the detection and segmentation of kidneys, while in step 2, 3D U-Net arranged a set of filters for the segmentation of tumors in the kidneys.

A key step in training the network is to properly balance classes according to the pixel-level frequency of each class in the data. It was observed that without class balance, training networks to segment small structures, such as lesions, was not possible because of the high class imbalance. Therefore, we introduce loss function in the network, cross entropy + Dice loss,

$$L = -\left(\sum_{i=1}^N \hat{P}_i \log P_i + (1 - \hat{P}_i) \log(1 - P_i) + \left(\frac{2 \sum_{i=1}^N P_i \hat{P}_i}{\sum_{i=1}^N P_i^2 + \hat{P}_i^2}\right)\right), \quad (1)$$

Where P_i represents the probability that voxel i belongs to the foreground, and \hat{P}_i represents the true value.

For the segmentation of the image, we carry out post-processing, the method is

- 1) remove the area with volume less than $2W$;
- 2) if there are outliers, the treatment is that the two kidneys should have similar center points in the axis direction, remove the outlier which is further away from them.

3 Results and Discussion

We applied our model to labeling data set about the abdominal CT image. The dataset included 300 abdominal CT images from a random sample at the University of Minnesota Medical Center. All patients undergoing partial or radical nephrectomy for one or more renal tumors may be included in this database. We did not find that preoperative arterial abdominal CT was excluded. We used 210 images as the training set to train the network, and the remaining 90 images as the test set to apply the network to segment renal tumors. And good segmentation results are obtained (Fig.3).

4 Conclusions

In this paper, abdominal CT images were analyzed to segment the kidney and then the tumor from the kidney. To explore a precise method for kidney tumor segmentation, so as to reduce the pain of renal biopsy in patients with kidney

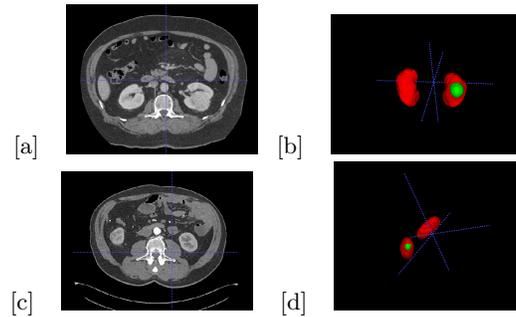


Fig. 3. pics

tumor and play a certain guiding role in subsequent surgical treatment, radiation or drug therapy. The experimental results are satisfactory. The dataset included 300 abdominal CT images from a random sample at the University of Minnesota Medical Center. All patients undergoing partial or radical nephrectomy for one or more renal tumors may be included in this database. We did not find that preoperative arterial abdominal CT was excluded. We used 210 images as the training set to train the network, and the remaining 90 images as the test set to apply the network to segment renal tumors. And good segmentation results are obtained.

References

1. Doi, K., MacMahon, H., Katsuragawa, S., et al.: Computer-aided diagnosis in radiology: potential and pitfalls. *European journal of Radiology* **31**(2), 97–109 (1999)
2. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics* **31**(4–5), 198–211 (2007)
3. Shim, H., Chang, S., Tao, C., et al.: Semiautomated segmentation of kidney from high-resolution multidetector computed tomography images using a graph-cuts technique. *Journal of computer assisted tomography*, **33**(6), 893–901 (2009)
4. Linguraru, M.G., Yao, J., Gautam, R., et al.: Renal tumor quantification and classification in contrast-enhanced abdominal CT. *Pattern Recognition* **42**(6), 1149–1161 (2009)
5. Linguraru, M.G., Wang, S., Shah, F., et al.: Automated noninvasive classification of renal cancer on multiphase CT. *Medical Physics* **38**(10), 5738–5746 (2011)
6. Spiegel, M., Hahn, D. A., Daum, V., et al.: Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. *Computerized Medical Imaging and Graphics* **33**(1), 29–39 (2009)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *NIPS*. pp. 1097–1105 (2012)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *CVPR*. pp. 3431–3440 (2015)
9. Chen, L.C., et al.: Semantic image segmentation with deep convolutional nets and fully connected crfs. *Computer Science* (4), 357–361 (2014)

10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI, vol. 9351, pp. 234–241 (2015)
11. Wang, J., MacKenzie, J.D., Ramachandran, R., Chen, D.Z.: Detection of glands and villi by collaboration of domain knowledge and deep learning. In: MICCAI, pp. 20–27 (2015)
12. Milletari, F., Navab, N., Ahmadi, S. A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 IEEE (2016)
13. ?i?ek, ?, Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention, pp. 424–432 Springer, Cham (2016)
14. Chen, H., Dou, Q., Yu, L., Heng, P. A. (2016).: Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. arXiv preprint arXiv:1608.05895. (2016)
15. Isensee, F., Petersen, J., Kohl, S. A., J?ger, P. F., Maier-Hein, K. H.: nnU-Net: Breaking the Spell on Successful Medical Image Segmentation. arXiv preprint arXiv:1904.08128. (2019)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 2015: 234–241. Springer, Cham (2015)