A double cascaded framework based on 3D SEAU-Net for kidney and kidney tumor Segmentation

Jianhong Cheng[1,3], Jin Liu[1], Liangliang Liu[1], Yi Pan[2], and Jianxin Wang[1*]

[1] School of Computer Science and Engineering, Central South University, Changsha, 410083, China
`jxwang@mail.csu.edu.cn`
[2] Department of Computer Science, Georgia State University, Atlanta, GA 30302-3994, USA
[3] Institute of Guizhou Aerospace Measuring and Testing Technology, Guiyang, 550009, China

Abstract. Accurate segmentation of kidney and kidney tumor from CT-volumes is vital to many clinical endpoints, such as differential diagnosis, prognosis and radiation therapy planning. While manual segmentation is subjective and time-consuming, fully automated extraction is quite imperative and challenging due to intrinsic heterogeneity of tumor structures. To address this problem, we propose a double cascaded framework based on 3D SEAU-Net to hierarchically and successively segment the subregions of the target. This double cascaded framework is used to decompose the complex task of multi-class segmentation into two simpler binary segmentation tasks. That is to say, the region of interest (ROI) including kidney and kidney tumor is trained and extracted in the first step, and the pre-trained weights are used as the initial weights of the network that is to segment the kidney tumor in second step. Our proposed network, 3D SEAU-Net, integrates residual network, dilated convolution, squeeze-and-excitation network and attention mechanism to improve segmentation performance. To speed training and improve network generalization, we take advantage of transfer learning (i.e., weight transfer) in the whole training phase. Meanwhile, we use 3D fully connected conditional random field to refine the result in post-processing phase. Eventually, our proposed segmentation method is evaluated on KiTS 2019 dataset and experimental results achieves mean dice scores 93.51% for the whole kidney and tumor, 92.42% for kidney and 74.34% for tumor on the training data.

Keywords: First keyword · Second keyword · Another keyword.

1 Introduction

Kidney cancer is a relatively common cancer in men and women. According to global cancer statistics in 2018, there are over 400,000 new cases each year, suffering from kidney cancer [1]. Surgery is the most common treatment option at present. Due to the diversity of kidney and kidney tumor morphology, there is a growing interest in studying the relationships between the tumor morphology and surgical outcomes [14, 4], as well as in developing state-of-the-art treatment techniques [15]. Automatic semantic segmentation based on deep learning algorithms is a promising tool to address these issues, however the heterogeneity of morphology makes it a challenging problem.

Recently, deep learning based methods, especially convolution neural networks (CNNs), have exhibited state-of-the-art performance in semantic segmentation [2, 5, 19] and are also widely applied for segmenting lesions or regions of interests (ROIs) in medical images [17, 12, 11]. In the past several years, U-Net based methods have made significant achievements in the medical image segmentations domain. In 2015, the U-Net won the championship with a great advantage in multiple challenge tasks in ISBI, such as Glioblastoma-astrocytoma U373 cells segmentation, HeLa Cells Tracking, etc. And since then, U-Net based architecture has become a benchmark for medical image segmentations and has achieved better performance on other public challenge tasks. For instance, Ronneberger et al. further extended the previous U-Net from two-dimensional to three-dimensional structure and used the proposed 3D U-Net to segment the Xenopus kidney in 2016 [3]. In the same year, a similar network named V-Net was proposed by Milletari et al. and was used to segment prostate volumes [16]. Besides, Isensee et al. improved the 3D U-Net to segment glioma with multi-subregion architectures and achieved competitive results in both BraTS 2017 and BraTS 2018 challenges [9, 10]. For kidney and kidney tumor segmentation, there are only a handful of research works in recent years due to the challenges of data collection sets annotation. One early study showed that Yang et al. presented an kidney segmentation method by using multi-atlas image registration [23]. Although this method can segment the kidney with a high precision in terms of Dice similarity coefficient (DSC), it takes a long time to infer the result on each image. Another relevant work is that Taha et al. also inspired by U-Net architecture and proposed a Kid-Net for kidney vessels(i.e., artery, vein and ureter) segmentation from CT-volumes [20]. Besides, Yu et al. developed a Crossbar-Net for kidney tumor segmentation and presented a new sampling method named crossbar patches [24]. They trained two Crossbar-Nets via a cascaded training strategy and the final result is generated by a majority vote of all sub-models.

In this work, we aim to segment kidney and kidney tumor from CT-volumes and our contributions are summed up the following three-fold. First, we propose a double cascaded framework based on 3D CNNs to hierarchically and successively segment the subregions of the target. We decompose the complex task of multi-class segmentation into two simpler binary segmentation tasks, and make full use of the hierarchical structure of target to reduce false positives via the proposed cascaded framework. Second, we extend the works of literatures [9, 8] and develop a 3D convolutional neural network, namely 3D SEAU-Net, to segment each binary segmentation task. The 3D SEAU-Net embeds residual network, dilated convolution, squeeze-and-excitation network and attention mechanism to improve segmentation performance. Third, we take advantage of transfer learning to speed training and improve network generalization in the whole training phase, and use 3D fully connected conditional random field (CRF) [13] to refine the result in post-processing phase.

2 Methods

2.1 KiTS19 data

The whole CT imaging dataset presented by the KiTS19 challenge was collected from multi-centers between 2010 and 2018. And the manual annotations were preformed on a distributed system by medical students under the supervision of senior clinical chair [7]. After thresholding, hilum filling and interpolation processing, there are 210 samples with different size as training set and 90 samples as testing set. In order to extract as much valuable semantic information as possible, we first crop each sample to a smaller size under retaining foreground size, and then resize it $128 \times 128 \times 128$ via nearest neighbour interpolation. Before feeding the images into the deep learning network, each sample is normalized by the following equation:

$$V_m^* = \frac{V_m - \frac{1}{N}\sum V_m}{\sqrt{\frac{1}{N}\sum (V_m - \frac{1}{N}\sum V_m)^2}}, \tag{1}$$

where V_m indicates the voxels value of sample m, N is the number of sample and V_m^* indicates the voxels value of normalized sample m. This method is referred to as zero-mean normalization and widely used for preprocessing in deep learning.

2.2 Double Cascaded Framework

Our proposed cascaded framework is presented in Fig. 1. We utilize two identical networks to hierarchically and successively segment the subregions of the target and each network performs an end-to-end binary segmentation task. The first network segments the whole target including kidney and kidney tumor of each patient. Then the model weights of the first network trained is obtained and is used as the initial weights of the second network that is to segment the kidney tumor. The inputs of both networks are the same images except for labels. Particularly speaking, the input label of the first network including kidney and kidney tumor is regarded as the identical target and converted to one foreground before feeding images into network. Similarly, the input label of the second network only includes kidney tumor and the label including kidney is served as background. The segmentation results of both networks are post-processed via a 3D fully connected conditional random field (CRF) [13] which can effectively remove false positives and be widely used for post-processing on biomedical 3D volumes [12, 11]. Furthermore, the final segmentation result is obtained by fusing the two post-processed results of which the second one is taken as the mask of the first one.

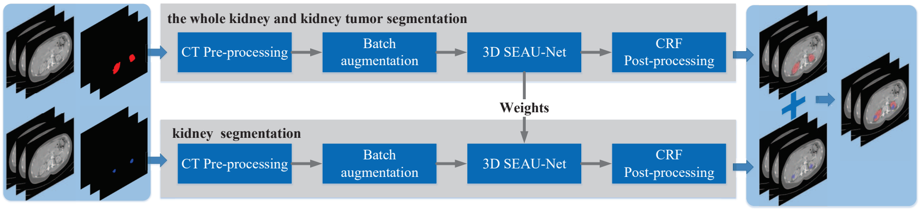

Fig. 1. Our proposed double cascaded framework for kidney and kidney tumor Segmentation.

Baseline Network For a 3D convolutional neural network, it is necessary to take the balance between model complexity and memory consumption into account. Generally speaking, the deeper or wider a CNN model, the performance of the model becomes more better. However, due to memory limitations, it may not be the best way to increase performance [22]. Thus, some researches have focused on improving the performance of the model via some mechanisms, such as residual block [6], squeeze-and-excitation network (SENet) [8] and attention-aware [21, 18, 5]. Based on that, we attempt to find a good compromise between the depth and width. Afterwards, we develop a 3D convolutional neural network, namely 3D SEAU-Net shown in Fig. 2, for our each binary segmentation task. Our proposed network is inspired by a typical 2D U-Net architecture [17] and we go through minor modifications that the max-pooling operation is replaced by a convolutional block with stride of 2 in each down-sampling and the dilation rate is accordingly increased in all convolution layers of network. In addition, we extend the previous 2D SENet [8] and 2D residual attention mechanism [21] into 3D structure, and embed them to improve representation ability of our baseline network. One important modification in our architecture is that

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **68**(6), 394–424 (2018)
2. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)
3. Çiçek, Ö., 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 (2016)
4. Ficarra, V., Novara, G., Secco, S., Macchi, V., Porzionato, A., De Caro, R., Artibani, W.: Preoperative aspects and dimensions used for an anatomical (padua) classification of renal tumours in patients who are candidates for nephron-sparing surgery. European urology **56**(5), 786–793 (2009)

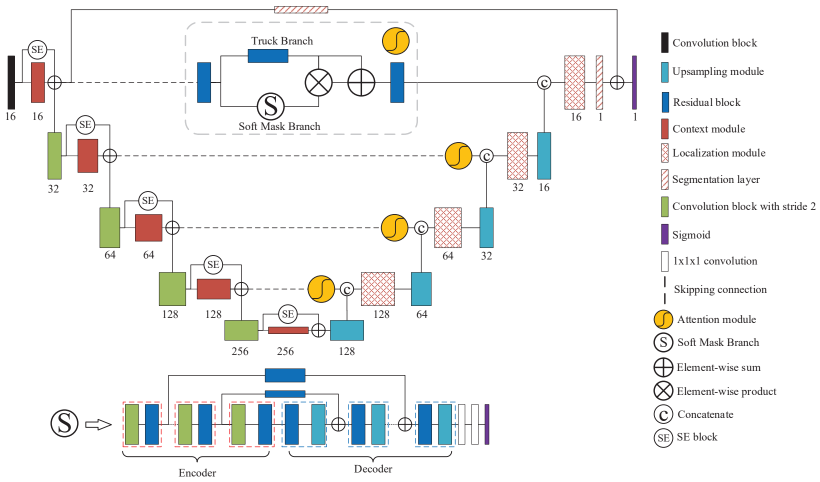

Fig. 2. Our framework. The atrous convolution is applied to all the convolution operations whose dilated rate increases proportionally with depth levels of network. The number in Figure denotes the number of feature channels.

5. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H.: Dual attention network for scene segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3146–3154 (2019)
6. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision. pp. 630–645. Springer (2016)
7. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019)
8. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
9. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: International MICCAI Brainlesion Workshop. pp. 287–297. Springer (2017)
10. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: No new-net. In: International MICCAI Brainlesion Workshop. pp. 234–244. Springer (2018)
11. Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B.: Multi-scale 3d convolutional neural networks for lesion segmentation in brain mri. Ischemic stroke lesion segmentation **13**, 46 (2015)
12. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. Medical image analysis **36**, 61–78 (2017)
13. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: Advances in neural information processing systems. pp. 109–117 (2011)

14. Kutikov, A., Uzzo, R.G.: The renal nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. The Journal of urology **182**(3), 844–853 (2009)
15. Li, J., Lo, P., Taha, A., Wu, H., Zhao, T.: Segmentation of renal structures for image-guided surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 454–462. Springer (2018)
16. 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)
17. 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. 234–241. Springer (2015)
18. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis **53**, 197–207 (2019)
19. Sinha, A., Dolz, J.: Multi-scale guided attention for medical image segmentation. arXiv preprint arXiv:1906.02849 (2019)
20. Taha, A., Lo, P., Li, J., Zhao, T.: Kid-net: convolution networks for kidney vessels segmentation from ct-volumes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 463–471. Springer (2018)
21. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3156–3164 (2017)
22. Wu, Z., Shen, C., Van Den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognition **90**, 119–133 (2019)
23. Yang, G., Gu, J., Chen, Y., Liu, W., Tang, L., Shu, H., Toumoulin, C.: Automatic kidney segmentation in ct images based on multi-atlas image registration. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5538–5541. IEEE (2014)
24. Yu, Q., Shi, Y., Sun, J., Gao, Y., Zhu, J., Dai, Y.: Crossbar-net: A novel convolutional neural network for kidney tumor segmentation in ct images. IEEE Transactions on Image Processing (2019)