3D Cascaded U-Net with a Squeeze-and-Exicitation Block for Semantic Segmentation on Kidney and Renal Cell Carcinoma in Abdonimal Volumetric CT

Sungwon Ham[1], Sungchul Kim[1], Jihye Yun[1], Yongjin Chang[2], Donghoon Yu[2]
Hyungi Seo[2] , MyungKi Yoon[3], Yoon Soo Kyung[4], Namkug Kim[1]

[1]Department of Convergence Medicine, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
{swlove920,sungchul7039,dool0120,namkugkim}@gmail.com

[2]CorelineSoft, Co., Ltd., Sung-Myung B/D 5F, World Cup buk-ro 6-gil 49, Mapo-Gu, Seoul 03991, South Korea
{yongjin.chang,donghoon.yu, hyungi.seo}@corelinesoft.com

[3] College of Medicine, Hallym University, Chuncheon, Republic of Korea
{theyoonicon}@gmail.com

[4]Department of Health Screening and Promotion Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul , Republic of Korea
{mdyunsoo}@gmail.com

Abstract. Segmentation is a fundamental process in medical image analysis. Recently, convolutional neural networks (CNNs) has allowed for automatic segmentation; however, segmentaiton of complex organs and diseases including the kidney or renal cell carcinoma (RCC) remains a different task due to limited data and labor-intensive labeling work. The purpose of this study is to segment kideny and RCC in CT using cascaded 3D U-Net with a squeeze-and-excitation (SE) block using a cascaded method. 210 kidneys and their RCC in abdominal CT images were used as training and validation sets. The Dice similarity coefficients (DSCs) of kidney and RCC in test set were 0.963 and 0.734 respectively. The cascaded semantic segmentation can potentially reduce segmentation efforts and increase the efficiency in clinical workflow.

Keywords: cascaeded 3D U-net; deep learning (DL); kidney; renal cell carcinoma (RCC); semantic segmentation;

1   Introduction

Segmentation of medical images in computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound is one of the most important and essential

parts in medical image analysis [1-3]. It has many applications in medical fields ranging from the diagnosis of disease to quantitative imaging for diagnosis and treatment planning [4,5]. The segmentation, which is a fundamental process to all other techniques, is not only a difficult task due to the diverse qualities of medical images, but also labor-intensive and time-consuming work [6,7]. Especially, segmentation of complex abdominal organ structures, for example, renal cell carcinoma (RCC) and the kidney is much more difficult due to anatomical variabilities.

There have been many studies of segmentation with traditional techniques [8-10]. Recently, deep convolutional neural networks (CNNs) have rapidly developed as a powerful tool in anatomy delineation. Among the segmentation architectures, U-Net with a contracting path and a symmetric expanding path is one of the most successful networks in medical image segmentation [11]. In spite of the wide usage of U-Net have been proposed to enhance relationship with features by introducing skip connection bridge between the encoding layer and a corresponding the decoding layer [12-13], these methods are insufficient to capture a non linear-relationship with features. To solve the problem, we proposed low cost but effective method to combine "Squeeze and Excitation" (SE) blocks into these recent semantic segmentation network.

In this paper, we proposed cascaded 3D U-Net with SE blocks to segment the kidney and RCC in abdominal CT.

2   Materials and Methods

Our cascaded approach consists of a network that predicts the ROI region in the entire image (detection module) and a network that predicts detailed labels within the ROI image (segmentation module). In a detection module, the entire image of the folded in half were trained as one label merged with all labels. On the contrary, in a segmentation module, the all labels in the ROI images were simultaneously trained.
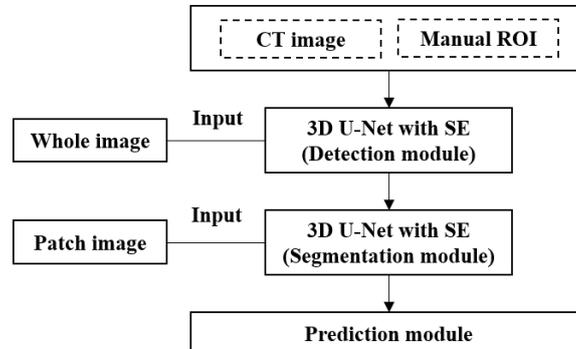
Figure 1: Overall procedure of our cascaded semantic segmentation method.

2.1   Data acquisitions

The abdominal CT images as well as their ground truth labels were provided in an anonymized NIFTI format with image matrix information (num_slices, height, width). Here, num_slices correspond to the number of axial images with increased slice number from superior to inferior. In all cases, the patient was supine position during CT scanning. If there were multiple CT series of a patient, that with the smaller slice thickness was chosen. Slice thicknesses range from 1mm to 5mm.

All datasets from the 2019 MICCAI KiTS (Kidney Tumor Segmentation Challenge) were enrolled in this study. Manual segmentation labels of KiTS dataset were delineated by medical students under the supervision of a clinical chair, Dr. Christopher Weight. The annotation procedure was as follows: The first and last slices containing kidney, its renal hilum or tumor tissue were recorded. For each selected slice, the annotator drew one or more contours such that all tissue depicted inside the contours was kidney, tumor, or fat as kidney contours. In addition, the annotator then drew a new series of contours such that the only tissue depicted in their interiors was tumor or not kidney as tumor contours. For detail, please refer to MICCAI KiTS website (https://kits19.grand-challenge.org/).

2.2   Cascaded 3D U-Net with a squeeze-and-exicitation block

3D U-net was used as a back-bone network architecture. Down- and up-samplings were performed three times as shown in Fig 2. The filter size was set to 32 multiples, SE and residual blocks were applied except for the beginning and end convolution layers, due to the limitation of GPU memory. To solve the imbalance problem between the background and the object, the sigmoid function was used in the prediction layer so that each label except the background was trained independently. In addition, thresholded LeakyReLu, which empirically determined for more efficient training, was applied after sigmoid function.
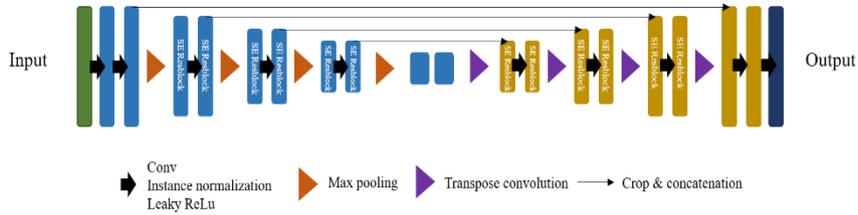
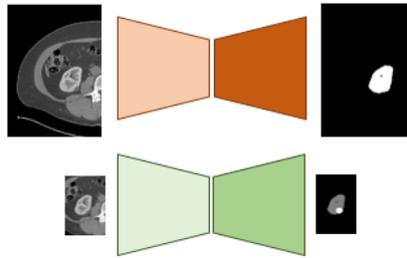Figure 2: 3D U-Net architecture with a squeeze-and-excitation block.

Figure 3: Cascaded 3D U-Net for kidney and renal cell carcinoma segmentation in abdominal CT.

2.3   Pre-processing and post-processing

The preprocessing for resizing and matching histogram of image density were performed. Considering the limitation of GPU memory, the entire images were resized. Because CT has a physical meaning of the pixel value, pixel value normalization was applied differently and CT pixel value were changed to increase the contrast for region of interests (ROI) with the five kinds of specific windowing levels and widths in Table 1. In the final prediction module, segmentation results of ROI images with these 5 models were ensembled by hard voting. The majority voting order was as follows: cancer, kidney and background.

Table 1: The pre-processing and training parameters of the five kinds of models.

	Windowing Level	Intensity normalization	Augmentation	Loss function
Model 1	[-300 600]	[0 1]	Horizontally flip	Average dice coefficient
Model 2	[-100 400]	[0 1]	Horizontally flip & rotation [-10 10]	Average dice coefficient
Model 3	[-100 400]	[0 1]	Horizontally flip & rotation [-5 5],	Average dice coefficient
Model 4	[-100 400]	[0 1]	Horizontally flip & rotation [-5 5]	Average dice coefficient + Focal loss
Model 5	[-100 400]	Mean 80.96 Std 58.84	Horizontally flip & rotation [-5 5]	Average dice coefficient

2.5   Statistical Evaluations

The model was implemented in Keras and Adam optimizers with learning rates of 0.0001. Gaussian noise was added to each input images and the most of loss functions were used as the Dice similarity coefficient (DSC). The DSC is evaluated as the equation (1). VGT and VCNN were defined as the volumes of ground truth and CNN segmentation result..

$$\text{DSC}\,(V_{GT}, V_{CNN}) = \frac{2|V_{GT} \cap V_{CNN}|}{|V_{GT}| + |V_{CNN}|} \quad (1)$$

3   Experimental Setup

3.1   The evaluation for Dice similarity coefficient

The DSC values of approximate kidney and RCC segmentation were 0.963 and 0.734, repectively. The DSC of approximate average kiney and tumor was 0.849. The segmentation of parenchyma shows the higher DSC, while RCC segmentation presents relative lower DSC. The results indicate that the results of the cascade semantic segmentation were highly corresponds with those of gold standard by manual segmentation.
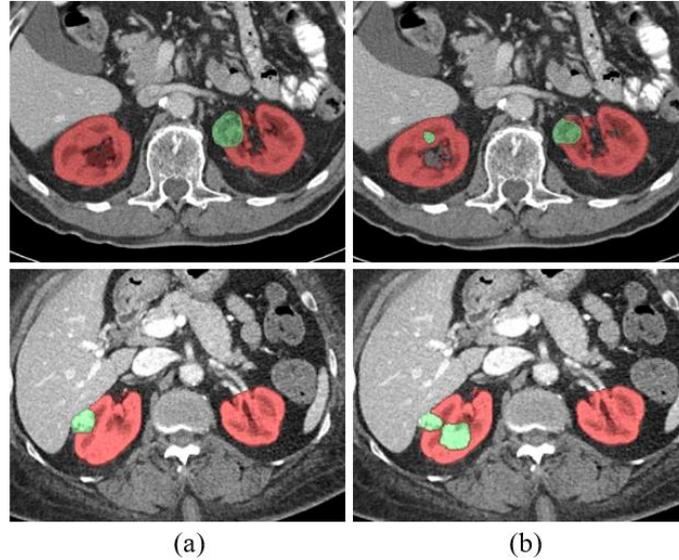
Figure 4: The results of kidney and RCC segmentation in two patients. (a) ground truth, (b) predicted result

4 Discussions and Conclusions

Segmentation in medical image analysis is a vital procedure for diagnosis and treatment planning. However, it is a tedious and labor-intensive task. The segmentation of anatomic structures like RCC or the kidney is much more difficult because they have various kinds of shapes, subtle differences within themselved and with neighborhood structures, which makes difficult to use conventional image processing methods including thresholds, region growing, mask editing, multiple slice editing, and Boolean operations. The semantic segmentation with CNN could be efficient in comparison with manual segmentation or conventional image processing methods and could be conducted much easier and faster than manual segmentation. The prediction result of parenchyma by our semantic segmentaiton showed more robust performance, while the segmentation result of relatively complex and subtle contrast organs like RCC was still challenging. This could be explained that the parenchyma has a relatively strong contrast to neighborhood tissues, large volume and little variations of shape and location in abdomin. On the other hand, the size and location of RCC are variable and the contrast of RCC to kidney parenchma is relatively subtle to cyst which was considered as one of kidney parenchymas. Further efforts to increase the number of training data with better network might be needed to increase the accuracy of sementic segmentation.

In conclusion, a cascaded semantic segmentation with CNN could be used for fully automated segmentation on kidney and RCC in volumetric CT images.

References

1. Chen, J. et al. Comparison of automatic and human segmentation of kidneys from CT images. International Journal of Radiation Oncology• Biology• Physics 54, 82 (2002).
2. Pham, D. L., Xu, C. & Prince, J. L. Current methods in medical image segmentation. Annual review of biomedical engineering 2, 315-337 (2000).
3. Noble, J. A. & Boukerroui, D. Ultrasound image segmentation: a survey. IEEE Transactions on medical imaging 25, 987-1010 (2006).
4. Nicolau, S., Soler, L., Mutter, D. & Marescaux, J. Augmented reality in laparoscopic surgical oncology. Surgical oncology 20, 189-201 (2011).
5. Xia, J. et al. Three-dimensional virtual-reality surgical planning and soft-tissue prediction for orthognathic surgery. IEEE transactions on information technology in biomedicine 5, 97-107 (2001).
6. Choi, J.-Y. et al. Analysis of errors in medical rapid prototyping models. International journal of oral and maxillofacial surgery 31, 23-32 (2002).
7. Tsagaan, B., Shimizu, A., Kobatake, H. & Miyakawa, K. in International Conference on Medical Image Computing and Computer-Assisted Intervention. 556-563 (Springer).
8. Pohle, R. & Toennies, K. D. in International Conference on Computer Analysis of Images and Patterns. 238-246 (Springer).
9. Pohle, R. & Toennies, K. D. in Medical Imaging 2001: Image Processing. 1337-1347 (International Society for Optics and Photonics).
10. 1Pohle, R. & Toennies, K. D. Self-learning model-based segmentation of medical images. Image Processing & Communications 7, 97-113 (2001).
11. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation, pages 234–241. Springer International Publishing, Cham, 2015.
12. S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 1175–1183, July 2017.
13. Guosheng Lin, Anton Milan, Chunhua Shen, and Ian D. Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. CoRR, abs/1611.06612, 2016.