

# SERU: cascaded SE-ResNeXT U-Net for kidney and tumor segmentation on KITS2019

Lei Li<sup>1</sup>, Sheng Lian<sup>1</sup>[0000-0003-2967-3041], and Zhiming Luo<sup>2</sup>

<sup>1</sup> Cognitive Science Department, Xiamen University, Xiamen, China  
lancerlian@stu.xmu.edu.cn

<sup>2</sup> Postdoc Center of Information and Communication Engineering, Xiamen University, Xiamen, China. zhiming.luo@xmu.edu.cn

**Abstract.** Accurate segmentation of kidney tumor in CT images is a challenging task. For solving this, we proposed **SE-ResNeXT U-Net** (SERU) model, which combines the advantages of SE-Net, ResNeXT and U-Net. For utilizing context information and key slices' information, we implement our model in a coarse-to-fine manner. We find left and right kidney's key slice respectively, and obtain key patches for refine training. We train and test our method on the KiTS19 Challenge. The predictions on kidney segmentation and tumor segmentation by our model show promising results.

**Keywords:** kidney · tumor · SE-ResNeXT · U-Net · segmentation.

## 1 Introduction

Kidney cancer is one of the most common urinary system lethal tumors, and the incidence rate in China's coastal areas has increased significantly in recent years. A Computed Tomography (CT) scan is among the best choices for workup of kidney tumor, including kidney cancer. However, accurate segmentation of kidney tumor in CT images is a challenging task due to the low contrast, irregular motion, diverse shapes and sizes. Furthermore, most of tumors appear a similar appearance in regular CT scans with surrounding tissues.

For solving this challenging task, we proposed **SE-ResNeXT U-Net** (SERU) model, which combines the advantages of SE-Net [2], ResNeXT [6] and U-Net [5]. For utilizing context information and key slices' information, we implement our model in a coarse-to-fine manner. We find left and right kidney's key slice respectively, and obtain key patches for refine training. The prediction on kidney segmentation and tumor segmentation show promising results.

We train, validate and test our method on the KiTS19 Challenge [1], which contains 300 Kidney Tumor cases with clinical context, CT semantic segmentations, and surgical outcomes. Among them, 210 (70%) cases were selected at random as the training set, and 90 (30%) cases were selected as testing set. The

detailed introduction of KiTS19 dataset can be referred to KiTS19 Challenge Homepage<sup>3</sup> and KiTS GitHub page<sup>4</sup>.

## 2 Methodology

### 2.1 Pre-processing:

We apply contrast limited adaptive histogram equalization (CLAHE) operation [4] as pre-processing step for increasing CT images' contrast, and alleviating the learning difficulty of CNNs. Also, we adopt random flip and random scaling for data augmentation steps.

### 2.2 Model architecture:

For solving the challenging task of kidney and tumor segmentation in CT imaging, we proposed **SE-ResNeXT U-Net** (SERU), which combines the advantages of SE-Net [2], ResNeXT [6] and U-Net [5]. Additionally, SE-Net won the last ImageNet competition Image Classification mission with great advantage, which shows the effectiveness of such architecture in computer vision tasks. While U-Net is one of the most promising model in the field of medical imaging.

In the contracting path of SERU model, we adopt *SE-ResNeXt50-32x4d* blocks pre-trained with ImageNet<sup>5</sup>. While in the expanding path, we use original U-Net [5] model's blocks.

### 2.3 The cascaded training scheme

Inspired by our previous work on retinal vessel segmentation [3], we implemented our model with a cascaded manner. In this way, we can first obtain coarse segmentation results and then utilize the information given in the coarse results to generate key patches for refine. The overall implementation pipeline of our proposed SERU model in cascaded manner is displayed in Fig 2.

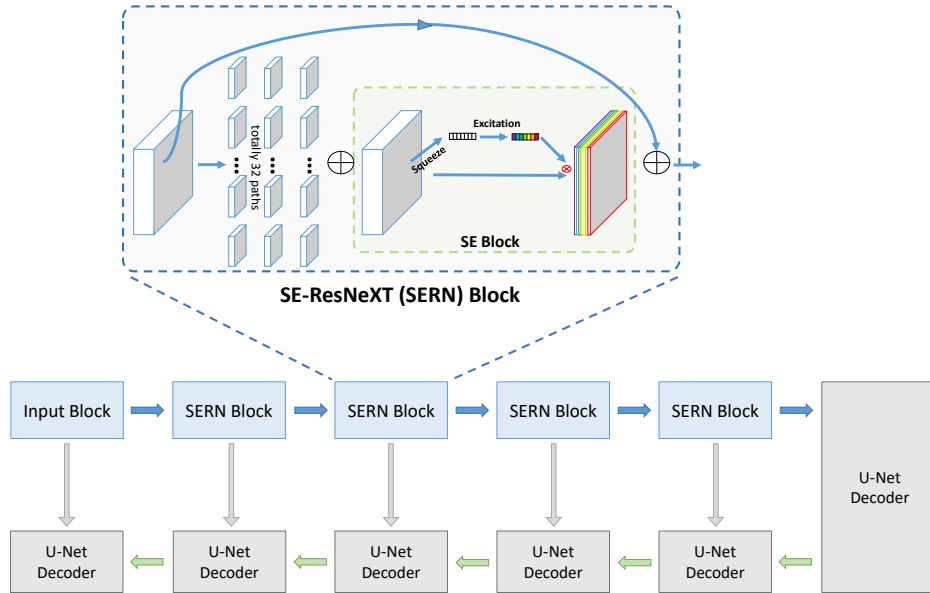
**Coarse segmentation** We train our model in a cascaded manner. For the first part, we train the SERU model with pre-processed training data, and generated coarse segmentation results.

**Refine patch generation** After coarse segmentation results generation, we regard the predicted kidney and tumor region as one meta class, and find out max contour of left kidney and right kidney respectively. Then we find the center point of the max contour on both sides, and cut 256\*256 rectangles in left and right kidney regions for refine training.

<sup>3</sup> <https://kits19.grand-challenge.org/>

<sup>4</sup> <https://github.com/neheller/kits19>

<sup>5</sup> <https://github.com/Cadene/pretrained-models.pytorch>



**Fig. 1.** The framework of the proposed SE-ResNeXT U-Net (SERU) model.

**Refine training** We adopt the left and right 256\*256 refine patches generated by last step for refine training, and the same SERU model with different input scale.

**Results fusion** For utilizing the advantages of our coarse and refine results, we fuse them by applying different weights, which goes as

$$predictions = \alpha * coarse + (1 - \alpha) * refined \quad (1)$$

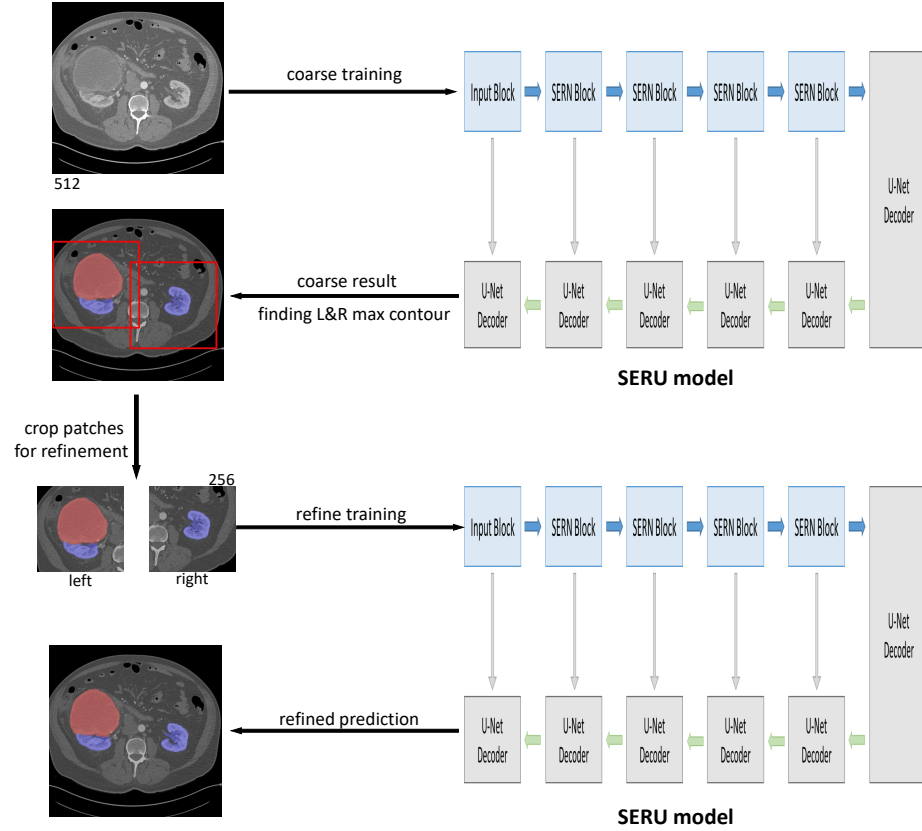
Here,  $\alpha$  is the ratio weighting the coarse results and refined results when fusing. We find that when  $\alpha$  is 0.4, we can obtain best results.

### 3 Implementation

We implement our model in by PyTorch. The Adam optimizer is adopted to train our model with an initial learning rate of  $1 * 10^{-4}$ , and the learning rate drop every 5 steps with the rate of 0.1. We initialize the weights of the SERU's contracting path by using the pre-trained weights on ImageNet provided by this GitHub repository <sup>6</sup>.

The proposed model is trained and tested following the pipeline discussed in Section 2.3 on a machine with Intel i7-7700K CPU and an NVIDIA 1080Ti

<sup>6</sup> <https://github.com/Cadene/pretrained-models.pytorch>



**Fig. 2.** The overall implementation pipeline of our proposed SERU model in cascaded manner.

GPU. We train our model for 30 epochs. And the fusion weighting ratio  $\alpha$  is set to be 0.4.

## 4 Evaluation Metrics

The Sørensen-Dice similarity coefficient (DSC) was used as a statistical validation metric to evaluate the performance of both the *kidney+tumor* segmentation results and *tumor* segmentation results. The DSC can be formulated as

$$DSC = \frac{2TP}{2TP + FP + FN}, \quad (2)$$

using the definition of true positive (TP), false positive (FP), and false negative (FN). And *kidney+tumor* and *tumor* are regarded as foreground issues, respectively.

In testing stage, we have 90 cases. The overall ranking score can be calculated as

$$S = \frac{1}{90} \sum_{i=0}^{89} \frac{1}{2} \left( \frac{2 * n_{t,tp}^{(i)}}{2 * n_{t,tp}^{(i)} + n_{t,fp}^{(i)} + n_{t,fn}^{(i)}} + \frac{2 * n_{k,tp}^{(i)}}{2 * n_{k,tp}^{(i)} + n_{k,fp}^{(i)} + n_{k,fn}^{(i)}} \right), \quad (3)$$

Where  $k$  representing kidney or tumor, while  $t$  representing tumor only.

## 5 Results

We validate our model on validation set, and achieve kidney+tumor DSC of 96.78%, and tumor DSC of 74.32%. The overall ranking score on our validate set is 85.55% Our model's predictions on testing set is uploaded. We will add more visual results on the next version of the paper.

## References

1. 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)
2. 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)
3. Lian, S., Li, L., Lian, G., Xiao, X., Luo, Z., Li, S.: A global and local enhanced residual u-net for accurate retinal vessel segmentation. IEEE/ACM transactions on computational biology and bioinformatics (2019)
4. Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E.: Contrast-limited adaptive histogram equalization: speed and effectiveness. In: IEEE Conference on Visualization in Biomedical Computing. pp. 337–345 (1990)
5. 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)
6. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1492–1500 (2017)