

RESEARCH ARTICLE

SERU: A cascaded SE-ResNeXT U-Net for kidney and tumor segmentation

Xiuzhen Xie¹ | Lei Li²  | Sheng Lian² | Shaohao Chen⁴ | Zhiming Luo³

¹College of Mathematics and Information Engineering, Longyan University, Longyan, China

²Artificial Intelligence Department, Xiamen University, Xiamen, China

³Post-doctoral Mobile Station of Information and Communication Engineering, Xiamen University, Xiamen, China

⁴The First Affiliated Hospital of Fujian Medical University, Fuzhou, China

Correspondence

Zhiming Luo, Post-doctoral Mobile Station of Information and Communication Engineering, Xiamen University, Xiamen, Fujian 361005, China.

Email: zhiming.luo@xmu.edu.cn

Funding information

China Postdoctoral Science Foundation, Grant/Award Number: No.2019M652257; Fujian educational research programs for young and middle-aged teachers, Grant/Award Number: Grant number: JT180187; National Natural Science Foundation of China, Grant/Award Numbers: No.61571188, No.61572409, No.61806172, No.61876159, No.U1705286

Summary

According to statistics, kidney cancer is one of the most deadly cancer. An early and accurate diagnosis can significantly increase the cure rate. Accurate segmentation of kidney tumors in CT images plays an important role in kidney cancer diagnosis. However, it is a challenging task due to many different aspects, such as low contrast, irregular motion, diverse shapes, and sizes. For solving this issue, we proposed a SE-ResNeXT U-Net (SERU) model in this study, which takes the advantages of SE-Net, ResNeXT and U-Net. Besides, we implement our model in a coarse-to-fine manner to utilize the information of context and key slices from the left and right kidney. We train and test our method on the KiTS19 Challenge. Experimental results demonstrate that our model can achieve promising results.

KEYWORDS

kidney, SE-ResNeXT, segmentation, tumor, U-Net

1 | INTRODUCTION

According to statistics, more than 400 000 kidney cancers were diagnosed worldwide in 2018, leading to 17 500 deaths.¹ Kidney cancer is one of the most common lethal tumors in the urinary system, and the incidence rate in coastal areas of China has increased significantly in recent years. A computed tomography (CT) scan is among the best choices for the workup of kidney tumors, including kidney cancer. The fine segmentation is helpful to the quantitative and qualitative study of renal cancer.

However, accurate segmentation of kidney tumors in CT images is a challenging task due to many different aspects, such as low contrast, irregular motion, diverse shapes, and sizes. Moreover, most of the tumors show a similar appearance in regular CT scans with surrounding tissues. It is worth mentioning that in most CT cases, the kidney region accounts for a small proportion of the entire CT image, let alone the tumor area. For example, in the Kidney Tumor Segmentation Challenge (KiTS19) challenge's dataset, the kidney region pixels only take up about 0.69% within the whole image. The tumor region pixels take an even much smaller ratio of 0.17%. This leads to a severe class imbalance problem.^{2,3}

For solving this challenging task, we proposed SE-ResNeXT U-Net (SERU) model, which combines the advantages of SE-Net,⁴ ResNeXT,⁵ and U-Net.⁶ For utilizing the information of context information and the key slices, we implement our model in a two-stage coarse-to-fine manner. In

the first stage, we train a model to perform a coarse segmentation in the whole image. Then, we find the key slice in the left and right kidney and obtain key patches by finding max contours. In the second stage, we refine the segmentation by taking the obtained patches as input. In this study, the refinement network has the same architecture as the first stage. After refining segmentation, for fully utilizing the global and local information, we integrate the prediction from the first stage and the second stage to get the final segmentation.

We train, validate, and test our method on the KiTS19 Challenge,¹ which contains 300 kidney tumor cases with clinical context, CT semantic segmentation, and surgical outcomes. Following the protocol in this dataset, we randomly selected 210 (70%) as the training set and the rest 90 (30%) cases as the testing set. A more detailed introduction of the KiTS19 dataset can be referred to KiTS19 Challenge Homepage* and KiTS GitHub page †.

To evaluate the performance, we compare the proposed SERU model with the two most widely used medical image segmentation methods, that is, U-Net and FCN. We adopt kidney + tumor Dice, tumor Dice, and average Dice as the evaluation metrics. Besides, we conduct an ablation study to analyze the effectiveness of each component in our proposed SERU.

Experimental results show that the performance of the first stage can surpass the baseline U-Net and FCN model, and the second stage can further improve the accuracy. These results demonstrate the superiority of our proposed model, as well as the advantages of our coarse-to-fine architecture.

2 | RELATED WORK

In recent years, along with the progress of computing power, we have witnessed great development in many fields.⁷⁻¹¹ In the past few years, deep learning-based methods, especially convolutional neural networks, achieved great success in medical image analyzing. It has become the first choice for various medical image analyzing applications,¹² including medical image segmentation,^{6,13,14} detection and classification of abnormality,¹⁵⁻¹⁷ medical image retrieval,¹⁸⁻²⁰ etc. Specifically, in the task of medical images segmentation using deep learning, many promising approaches emerged in recent years. In Reference 21, FCN model is proposed to solve the problem of low resolution after convolutional neural networks (CNN's) pooling layers in the task of semantic segmentation. Then U-Net⁶ improved FCN through a series of skip-connection and U-shaped architecture to maintain the information from lower and higher layers. U-Net achieved gratifying results on medical image segmentation tasks, and many variants of U-Net are proposed for different subtasks of medical image segmentation, including brain tumor segmentation,²² pancreas localization,²³ iris segmentation,¹⁴ etc. Cicek et al²⁴ extends the original two-dimensional (2D) U-Net to a three-dimensional (3D) version by replacing all 2D operations with their 3D counterparts. This work is used for semi-automatically and fully automatically segmenting the animal kidney volumetric given a sparse annotation, and achieve good performance. Also, some researchers focus on the deep model itself, and their eyes are not limited to a single medical image segmentation task. Isensee et al²⁵ proposed nnU-Net, a robust and self-adapting framework that uses only naive U-Net model and can automatically adjust hyperparameters to a new dataset, is proposed. The nnU-Net performs very well in a competition consisting of a series of medical image segmentation tasks, called Medical Segmentation Decathlon challenge.

Specifically, in the field of kidney-related analysis, many different deep learning-based methods have been proposed.²⁶⁻²⁸ Sharma et al²⁸ proposed a deep learning-based approach for segmenting polycystic kidneys and for total kidney volume computation in CT images. In this method, a basic encoder-decoder model following the architecture of VGG-16, with batch normalization for solving the internal covariance shift problem. Researchers²⁶ from the Memorial Sloan Kettering Cancer Center compared the performance of manual features and "deep" features on renal tissue slice images for renal cancer detection, and obtained 89% accuracy of disease detection on the tissue microarray dataset by using AlexNet. Hussian et al²⁷ proposed an approach that rearranges the 3D images into an extended 2D image collage, and train a deep-learning-based model to detect kidney slice containing renal cell carcinoma with 98% accuracy. Yin et al²⁹ represented an end-to-end framework, which integrated boundary distance regression and pixel classification networks, to cope with kidney segmentation on ultrasound images. Overall, given that kidney tumor segmentation is a challenging task with severe sample imbalance problems, few approaches attempt to solve this issue. In this paper, we attempt to solve this problem and achieve higher segmentation accuracy.

3 | METHODOLOGY

In this section, we will describe the details of our proposed cascaded SE-ResNeXT U-Net (SERU) for kidney and tumor segmentation in CT images. Firstly, we introduce the preprocessing step used to enhance the contrast of CT images in Section 3.1. Then, the detailed network architecture of SERU is depicted in Section 3.2. After, we present the cascaded training and testing scheme used for improving the segmentation accuracy in Section 3.3.

* <https://kits19.grand-challenge.org/>

† <https://github.com/neheller/kits19>

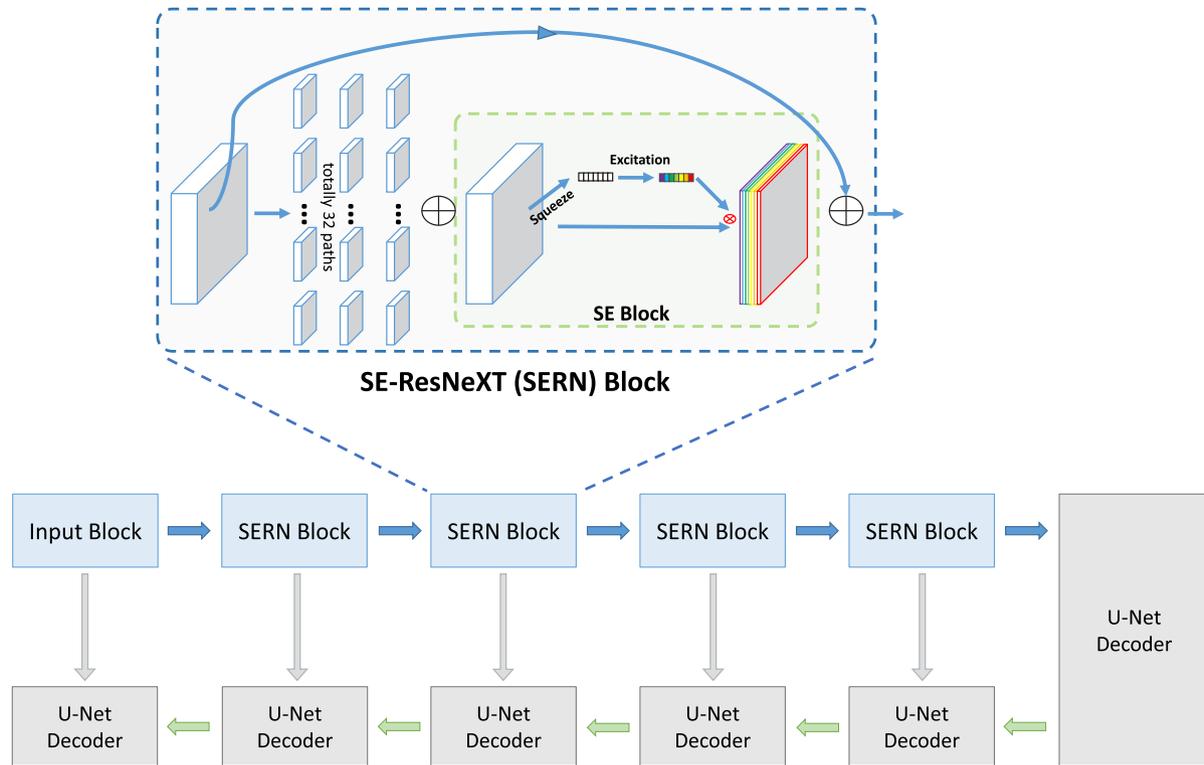


FIGURE 1 The framework of the proposed SE-ResNeXT U-Net (SERU) model

3.1 | Preprocessing

For better extract effective features on deep model, we first normalize the given CT images by converting the original Hounsfield Unit value to grayscale value ranged from 0.0 to 1.0. We then apply contrast limited adaptive histogram equalization (CLAHE) operation³⁰ as preprocessing step for increasing CT images' contrast, and alleviating the learning difficulty of CNNs. The CLAHE is an improvement of adaptive histogram equalization, which limits the contrast of the image and reduces the noise. In addition, we adopt random horizontal and vertical flip as data augment.

3.2 | Model architecture

To deal with the challenges in kidney and tumor segmentation in CT scans, we proposed a SERU model, which combines the advantages of SE-Net,⁴ ResNeXT,⁵ and U-Net.⁶ The SE-Net won the last ImageNet competition Image Classification mission with great advantage, which shows the effectiveness of such architecture in computer vision tasks. The ResNeXT is a combination of ResNet³¹ and Inception Net,³² which won the ILSVRC 2016. The U-Net is the most widely used CNN model in medical image segmentation, due to it is effective for processing biological structured information and does not require a very large amount of training data.

The framework of the proposed SERU is depicted in Figure 1. Overall, the proposed SERU has a U-Net-like architecture with an encoder, a decoder and skip connections. However, in the contracting path (encoder) of SERU model, we adopt *SE-ResNeXT50_32x4d* blocks pretrained with ImageNet[‡]. While in the expanding path (decoder), we use the original blocks as in U-Net⁶ model.

SE-Net. The SE-Net terms the "Squeeze-and-Excitation"(SE) block, mainly consist of the squeeze operation and the excitation operation. The SE block first performs a squeeze operation on the features to generate a channel descriptor, which represents a distribution of channel-wise feature responses. We adopt a global average pooling layer as a channel descriptor to generate channel-wise statistics, which is calculated by average each channel. Then, an excitation operation is adapted to learn the sample-specific activations for each channel. Considering the model complexity, we use two fully connected (FC) layers to implement the excitation. The parameters of FC are $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r}} \times C$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$, respectively, with reduction ratio r . Finally, the initial features are reweighted by the learned activations to produce the output.

[‡]<https://github.com/Cadene/pretrained-models.pytorch>

$$\text{out} = \text{FC}(z, \mathbf{W}) = \mathbf{W} \times z, \quad (1)$$

ResNeXt. Inspired by VGG/ResNets and Inception, the ResNeXt uses a stack of residual blocks with the same topology and adopts a split-transform-merge structure in each block (i)Split: The feature is sliced as a low-dimensional embedding. (ii) Transform: The split data is transformed by the group convolution layer. (iii) Merge: The transformations are finally merged by addition. In the SE-ResNeXt50_32x4d block, the input goes through a split-transform-merge structure, the grouped convolutions with 32 groups, for extracting features. Then the data is fed into an SE-Net module, the result of which is added with initial input by a skip-connection.

3.3 | The cascaded training scheme

Inspired by our previous work on retinal vessel segmentation,³³ 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. Then we select the key slices of the kidneys based on the coarse segmentation, and generate key patches for refinement. The overall cascaded implementation pipeline of our proposed SERU model is displayed in Figure 2.

In the first stage, we train the SERU model with preprocessed training data, and produce coarse segmentation results. For one case with a series of CT slides, after coarse segmentation results generation, we regard the predicted kidney and tumor region as one meta class. Then we find out the key slides by checking which slide contains the maximum contour of the left kidney and right kidney, respectively. In addition, we extract the max bounding-boxes from the contours of left and right kidney for each case, and treat them as the region of interest for the second refinement stage.

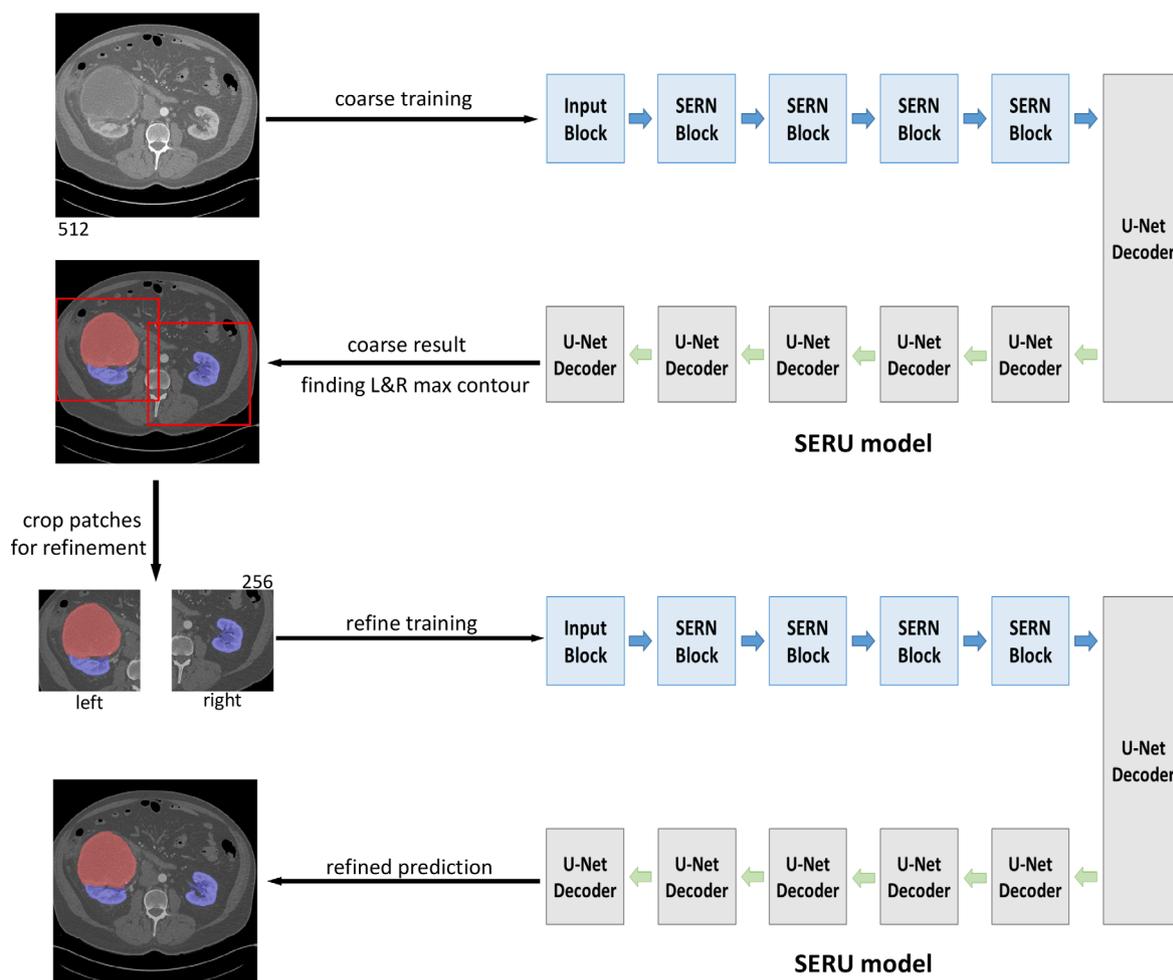


FIGURE 2 The overall implementation pipeline of our proposed SE-ResNeXT U-Net (SERU) model in cascaded manner

In the second stage, we enlarge the bounding-boxes to the size 256×256 , and propagate the enlarge bounding-box of the key slice to all other slides within the same case for finer segmentation. Then we adopt the left and right 256×256 refine patches generated by last step for refine training, and use the same SERU model with different input scale.

After getting the segmentation results from the first and second stage, we fuse them by a linear combination with different weights, which goes as

$$\text{predictions} = \alpha * \text{coarse} + (1 - \alpha) * \text{refined}, \quad (2)$$

where α is the ratio weighting the coarse results and refined results when fusing.

4 | EXPERIMENTS

4.1 | Dataset

We evaluated our method on the kidney dataset provided by KiTS19.¹ The data is collected from 300 underwent partial or radical nephrectomy patients between 2010 and 2018 at the University of Minnesota Medical Center. We can access the abdominal CT images and corresponding voxel-wise ground truth labels. The dataset contains 300 cases, of which 210 cases for training and 90 for testing. In our experiments, we randomly split the 210 cases into 168 cases as the training set and the remaining 42 cases as the validation set. The axial 2D data is utilized in this study, and the size of each slice is 512×512 .

4.2 | Implementation

We implement our model with PyTorch. The Adam optimizer is adopted to train our model with an initial learning rate of 1×10^{-4} , and we decrease the learning rate every five epochs with a rate of 0.1. We initialize the weights of the contracting path in the SERU by using the pre-trained weights on ImageNet provided by this GitHub repository⁵. The proposed model is trained and tested following the pipeline discussed in Section 3.3 on a machine with Intel i7-7700K CPU and an NVIDIA 1080Ti GPU. We train our model for 30 epochs. In addition, we adopt random horizontal and vertical flipping for data augment. The fusion weighting ratio α is set to be 0.4 in this study.

4.3 | 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. It represents the ratio of the intersecting area between the prediction results and the ground truth label to their total area.

The DSC can be formulated as

$$DSC = \frac{2TP}{2TP + FP + FN}, \quad (3)$$

where TP, FP, and FN stand for true positive, false positive, and false negative, respectively. Specially, two metrics have been calculated in which **kidney+tumor** and **tumor** are regarded as foreground tissues, respectively.

In testing stage, the overall ranking score of the 90 cases in the testing set can be calculated as

$$S = \frac{1}{90} \sum_{i=0}^{89} \frac{1}{2} \left(\frac{2 * n_{t,tp}^{(i)}}{2 * n_{t,tp}^{(i)} + n_{t,fp}^{(i)} + n_{t,fn}^{(i)}} + \frac{2 * n_{k,tp}^{(i)}}{2 * n_{k,tp}^{(i)} + n_{k,fp}^{(i)} + n_{k,fn}^{(i)}} \right), \quad (4)$$

where k represents kidney or tumor, and t represents tumor only.

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

5 | EXPERIMENTAL RESULTS

5.1 | Overall Performance

In this study, we evaluate our model on the validation set, and adopt kidney+tumor DSC, tumor DSC, and average DSC as evaluation metrics. As can be seen from TABLE 1, our SERU model can obtain the kidney+tumor DSC of 96.77%, the tumor DSC of 74.32%, and the average DSC of 85.55%.

To better demonstrate the effectiveness of our proposed SERU model, we compared with two widely used segmentation models, that is, U-Net⁶ and FCNs.²¹ As shown in TABLE 1, our SERU model can achieve better performance in kidney DSC, tumor DSC and average DSC. Especially, the Average DSC of SERU is 85.55% which is around 15% higher than the U-Net and fully convolutional network (FCN). This demonstrates the effectiveness of the backbone network of our model.

Further, we visualize the segmentation results of U-Net, different version of FCN, and our proposed SERU model in FIGURE 3. As can be seen, our method can obtain much accurate segmentation which with less false positive prediction. In the first row, the FCN fails to identify both kidney and tumor region. In the third row where there are no kidney/tumor region, the U-Net model and FCN model generate false positive prediction.

5.2 | Ablation study

5.2.1 | The influence of each component

Our proposed SERU model mainly takes the advantages from the U-Net, SENet, and ResNeXT. In this section, we conduct a ablation study to analyze the influence of each component, and reported the results in Table 2. Firstly, we can observe that combining the unet with the senet or resnext has a very slight effects for the Kidney DSC, while can give around 6% improvement of the Tumor DSC. However, when combining the unet with both senet and resnext, we can obtain a significant boost of the Tumor DSC and Average DSC.

5.2.2 | The effectiveness of two-stage training scheme

In the first stage, our method performs well in segmenting kidneys, while worse in tumors with a dice score of 96.87%. Then, we can obtain the bounding-box for kidney region and crop the patches from original images as the input of refine model. Through the cascaded refinement, the

Methods	Kidney DSC	Tumor DSC	Average DSC
U-Net	94.34%	43.98%	69.16%
FCN	94.38%	48.39%	71.38%
FCN8s	94.03%	52.49%	73.26%
FCN16s	93.33%	53.00%	73.17%
FCN32s	91.56%	35.94%	63.75%
Ours	96.77%	74.32%	85.55%

TABLE 1 Results compares with different methods

Abbreviation: DSC, Dice similarity coefficient; FCN, fully convolutional network.

TABLE 2 The segmentation accuracy under different model combinations at the first stage and the second stages

Methods	First stage			Second stage		
	Kidney DSC	Tumor DSC	Avg DSC	Kidney DSC	Tumor DSC	Avg DSC
unet	93.46%	39.03%	66.25%	95.39%	49.77%	72.58%
unet+senet	93.89%	45.63%	69.76%	94.58%	49.98%	72.28%
unet+resnext	93.30%	44.82%	69.06%	95.54%	57.97%	76.76%
unet+senet+resnext	96.87%	71.57%	84.22%	96.77%	74.32%	85.55%

Abbreviation: DSC, Dice similarity coefficient.

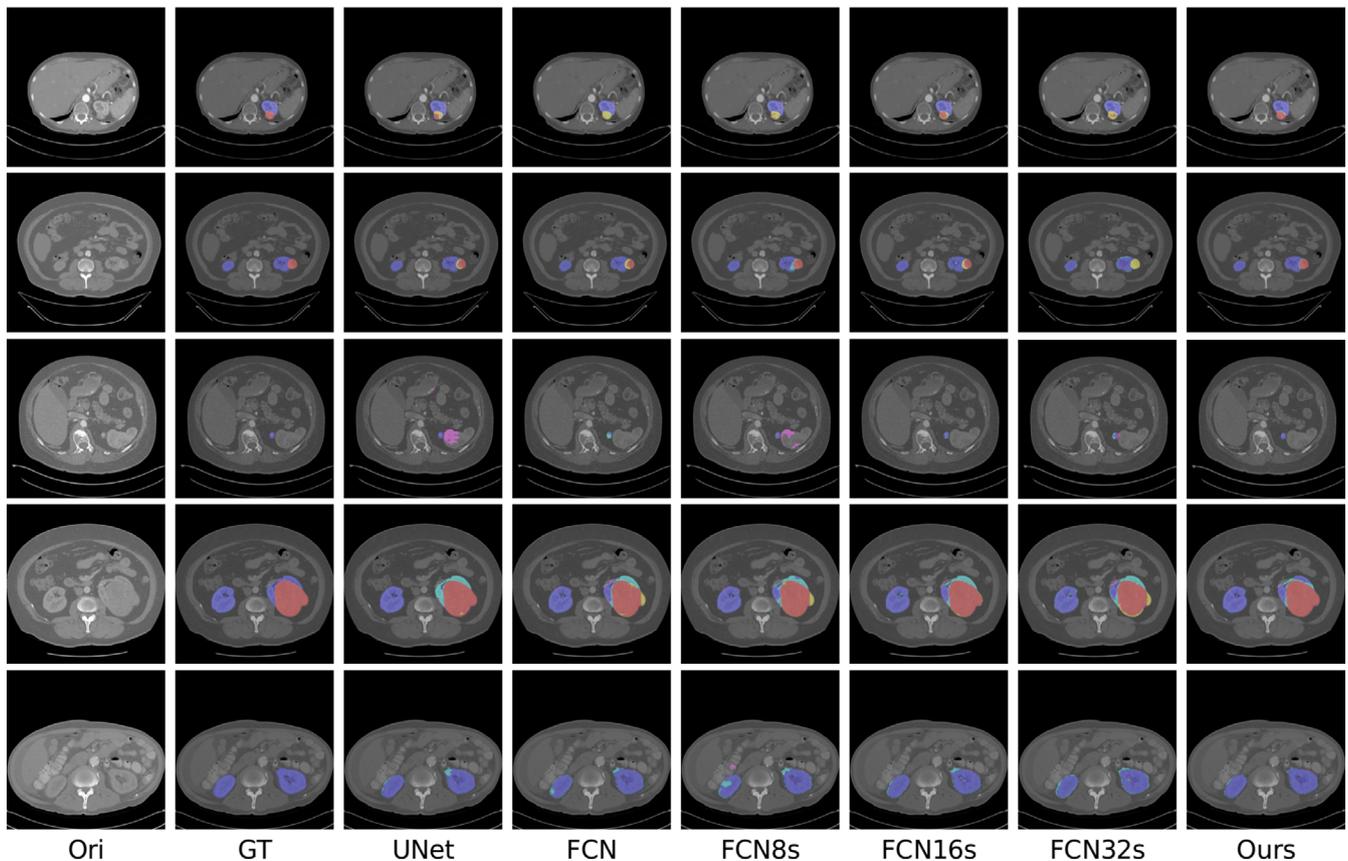


FIGURE 3 Results with different methods. Masks with colors are the ground truth and predictions results. **Ori**: Original images. **GT**: Ground truth of images. The mask with blue color indicates the kidney and red is for tumor. **Other Methods**: Difference between ground truth and prediction results of U-Net, FCN, FCN8s, FCN16s, FCN32s and ours. The blue mask represents prediction of kidney, red for tumor, the pink is for false kidney prediction, the yellow is ground truth of tumor not predicted, and the sky-blue is ground truth of kidney not predicted

dice of tumor can be improved from 71.57% to 74.32%, which indicates the second stage will focus on learning more discriminative features for distinguishing tumor from the kidney.

In Figure 4, we show several prediction results between two stages. We can see that the results of the first stage are incomplete or incorrect. But after applying the refine model, our model can correct those errors and then improve the performance.

5.2.3 | The impact of preprocessing and data augmentation

During the training process, we adopt the CLAHE preprocessing to enhance the contrast of the CT scans, as well as the horizon and vertical flipping for augmentation. In this section, we conduct experiments to evaluate the impact of the CLAHE and data augmentation step.

As shown in Table 3, for the unet, unet+senet, and unet+resnext model, the performance of the models trained with CLAHE are very similar with the models without using the CLAHE preprocessing step. On the other hand, for our proposed SERU which contains the unet, senet, and resnext, we can observe that there are a significant boost of the Tumor DSC (70.61% vs 74.32%) and the Average DSC (83.57% vs 85.55%). These results demonstrate that the CLAHE preprocessing is beneficial for improving the segmentation accuracy of proposed SERU.

In Table 4, we reported the segmentation results between the model trained with and without the data augmentation. From the Table 4, we can observe that the data augmentation step has a significant impact of increase the Tumor DSC and the Average DSC of our proposed SERU model.

5.2.4 | Parameter sensitivity analysis

As introduced in Section 3.3, after getting the coarse and fine segmentation from the first stage and the second stage, we then integrate them by a linear combination controlled the weight α . In this section, we analyze the sensitivity of α . As can be seen from Table 5, we can find that our model

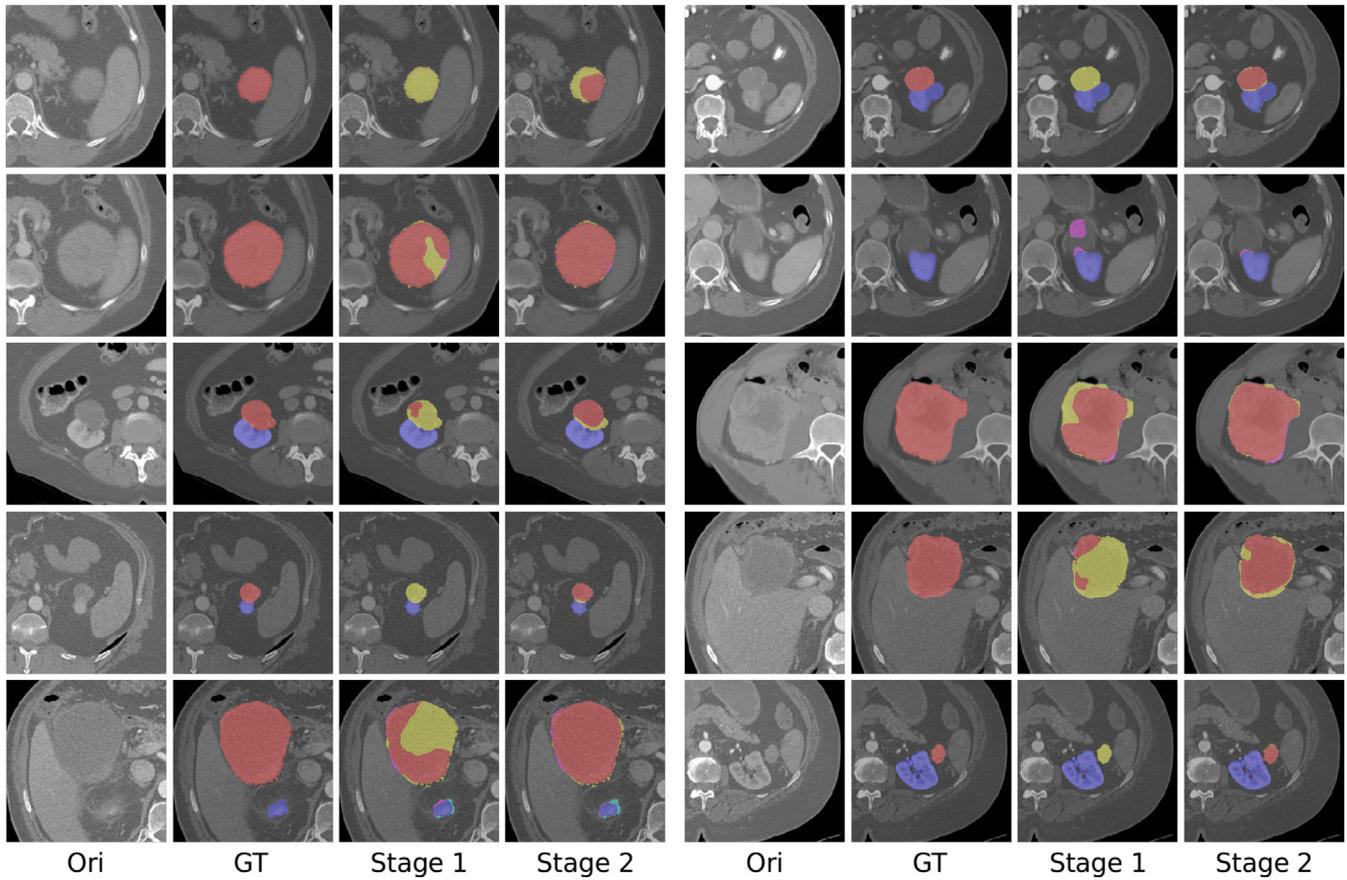


FIGURE 4 Results with different stages. **Ori**: Original images. **GT**: The ground truth of the images. **Stage 1**: Results of the first stage. **Stage 2**: Results of the second stage. The blue mask represents prediction of kidney, red for tumor, the pink is for false kidney prediction, the yellow is ground truth of tumor not predicted, and the sky-blue is ground truth of kidney not predicted

Methods	Kidney DSC	Tumor DSC	Avg DSC
unet (w/o CLAHE)	94.59%	50.12%	72.36%
unet+senet (w/o CLAHE)	95.31%	49.68%	72.49%
unet+resnext(w/o CLAHE)	95.45%	62.00%	78.72%
unet+senet+resnext (w/o CLAHE)	96.53%	70.61%	83.57%
unet	95.39%	49.77%	72.58%
unet+senet	94.58%	49.98%	72.28%
unet+resnext	95.54%	57.97%	76.76%
unet+senet+resnext	96.77%	74.32%	85.55%

TABLE 3 The segmentation results of our proposed model trained with/without the CLAHE preprocessing step

Abbreviation: DSC, Dice similarity coefficient.

achieves the best average DSC when α equals to 0.4. Compared with the result only from the first stage ($\alpha=1$) and the second stage ($\alpha=0$), there are a 2.12% and 0.28% increase of the average DSC, respectively. These results suggest that fusing the results from two stages is helpful for improving accuracy.

TABLE 4 The segmentation results of the different model trained with/without the data augmentation

Methods	Kidney DSC	Tumor DSC	Avg DSC
unet (w/o Data Augment)	95.67%	48.94%	72.31%
unet+senet (w/o Data Augment)	95.31%	49.68%	72.49%
unet+resnext (w/o Data Augment)	95.90%	55.91%	75.90%
unet+senet+resnext (w/o Data Augment)	95.45%	58.04%	76.75%
unet	95.39%	49.77%	72.58%
unet+senet	94.58%	49.98%	72.28%
unet+resnext	95.54%	57.97%	76.76%
unet+senet+resnext	96.77%	74.32%	85.55%

Abbreviation: DSC, Dice similarity coefficient.

TABLE 5 The performance of fusion the results from the first stage and second stage by using different α

Alpha	Kidney DSC	Tumor DSC	Avg DSC
0	96.58%	73.96%	85.27%
0.1	96.64%	74.15%	85.40%
0.2	96.69%	74.29%	85.49%
0.3	96.73%	74.32%	85.53%
0.4	96.77%	74.32%	85.55%
0.5	96.83%	73.85%	85.34%
0.6	96.87%	73.05%	84.96%
0.7	96.90%	72.53%	84.71%
0.8	96.90%	72.13%	84.51%
0.9	96.89%	71.82%	84.35%
1	96.40%	70.46%	83.43%

Abbreviation: DSC, Dice similarity coefficient.

6 | CONCLUSION

In this work, we propose a 2D kidney tumor segmentation model named cascaded SE-ResNeXT U-Net (SERU), which combines the advantages of SE-Net, ResNeXT, and U-Net. The model is implemented in a coarse-to-fine manner, which first performs a coarse segmentation. Then for fully utilizing the information from coarse results, we find the key slices with the largest left and right kidney region, respectively. The key slices' predicted kidney regions are utilized for generating kidney bounding-boxes, which serve as the input patches of the refinement model. After, we apply the same SERU model on the patches as mentioned above and produce predictions refined in patches. To fully utilizing the global and local information of given CT slides, we integrate the results of the coarse and refine stage by a weighting parameter α . We train and test our method on the KiTS19 Challenge. We got promising results among 2D-based methods that, compared with U-Net and FCN, our model achieves better results on both kidney DSC, tumor DSC, and the overall average DSC. For finding the best policy for integrating different stages' prediction, we implement an ablation study on the weighing parameter. Visualized comparisons and quantified results show not only the superiority of our backbone SERU network for extracting features of kidney CT images, but also the effectiveness of our coarse-to-fine segmentation procedure.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 61876159, 61806172, 61572409, U1705286 & 61571188); Fujian educational research programs for young and middle-aged teachers (Grant number: JT180187); the China Postdoctoral Science Foundation Grant (No. 2019M652257); Fujian Province 2011 Collaborative Innovation Center of TCM Health Management.

ORCID

Lei Li  <https://orcid.org/0000-0002-1500-7446>

REFERENCES

1. Heller N, Sathianathen N, Kalapara A, 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. Cai X, Niu Y, Geng S, et al. An under-sampled software defect prediction method based on hybrid multi-objective cuckoo search. *Concurr Comput Practice Exp*. 2019;32:e5478. <https://doi.org/10.1002/cpe.5478>.
3. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision; 2017:2980–2988.
4. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018:7132–7141.
5. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017:1492–1500.
6. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015:234–241.
7. Z C, Y C, X C, J C, J C. Optimal LEACH protocol with modified bat algorithm for big data sensing systems in Internet of Things. *J Parallel Distrib Comput*. 2019;132:217–229.
8. P W, F X, H L, Z C, L X, J C. A multi-objective DV-Hop localization algorithm based on NSGA-II in internet of things. *Mathematics*. 2019;7(2):184.
9. Cui Z, Zhang J, Wang Y, et al. A pigeon-inspired optimization algorithm for many-objective optimization problems. *Sci China Inf Sci*. 2019;62:070212.
10. Cui Z, Xue F, Cai X, Cao Y, Wang G-G, Chen J. Detection of malicious code variants based on deep learning. *IEEE Trans Ind Inform*. 2018;14(7):3187–3196.
11. Wang P, Huang J, Cui Z, Xie L, Chen J. A Gaussian error correction multi-objective positioning model with NSGA-II. *Concurr Comput Practice Exp*. 2020;32(5):e5464.
12. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
13. Man Y, Huang Y, Feng J, Li X, Wu F. Deep Q learning driven CT pancreas segmentation with geometry-aware U-net. *IEEE Trans Med Imag*. 2019;38:1971–1980.
14. Lian S, Luo Z, Zhong Z, Lin X, Su S, Li S. Attention guided U-Net for accurate iris segmentation. *J Visual Commun Image Rep*. 2018;56:296–304.
15. Yan K, Wang X, Lu L, Summers RM. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J Med Imag*. 2018;5(3):036501.
16. Haarbarger C, Baumgartner M, Truhn D, et al. Multi scale curriculum CNN for context-aware breast mri malignancy classification. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2019:495–503.
17. Wang Z, Meng Q, Wang S, Li Z, Bai Y, Wang D. Deep learning-based endoscopic image recognition for detection of early gastric cancer: a Chinese perspective. *Gastrointest Endosc*. 2018;88(1):198–199.
18. Anavi Y, Kogan I, Gelbart E, Geva O, Greenspan H. A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. Paper presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2015:2940–2943.
19. Anavi Y, Kogan I, Gelbart E, Geva O, Greenspan H. Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval. Paper presented at: Medical Imaging 2016: Computer-Aided Diagnosis. vol. 9785; 2016:978510.
20. Liu X, Tizhoosh HR, Kofman J. Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. Paper presented at: 2016 International Joint Conference on Neural Networks (IJCNN); 2016:2872–2878.
21. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015:3431–3440.
22. Mehta R, Sivaswamy J. M-net: a convolutional neural network for deep brain structure segmentation. Paper presented at: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017); 2017:437–440.
23. Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*; 2018.
24. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016:424–432.
25. Isensee F, Petersen J, Klein A, et al. nnu-net: self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*; 2018.
26. Schöffler PJ, Sarungbam J, Muhammad H, Reznik E, Tickoo S, Fuchs T. Mitochondria-based renal cell carcinoma subtyping: learning from deep vs. flat feature representations. Paper presented at: Machine Learning for Healthcare Conference; 2016:191–208.
27. Hussain MA, Amir-Khalili A, Hamarneh G, Abugharbieh R. Collage CNN for renal cell carcinoma detection from CT. Paper presented at: International Workshop on Machine Learning in Medical Imaging; 2017:229–237.
28. Sharma K, Rupprecht C, Caroli A, et al. Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. *Sci Rep*. 2017;7(1):1–10.
29. Yin S, Peng Q, Li H, et al. Subsequent boundary distance regression and pixelwise classification networks for automatic kidney segmentation in ultrasound images. *arXiv preprint arXiv:1811.04815*; 2018.
30. Pizer SM, Johnston RE, Ericksen JP, Yankaskas BC, Muller KE, Medical Image Display Research Group. Contrast-limited adaptive histogram equalization: speed and effectiveness. Paper presented at: Proceedings of the First Conference on Visualization in Biomedical Computing; May 22–25, 1990; Atlanta, GA:337.
31. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:770–778.
32. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015:1–9.

33. 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 Trans Comput Biol Bioinform*. 2019. <https://doi.org/10.1109/TCBB.2019.2917188>.

How to cite this article: Xie X, Li L, Lian S, Chen S, Luo Z. SERU: A cascaded SE-ResNeXT U-Net for kidney and tumor segmentation. *Concurrency Computat Pract Exper*. 2020;32:e5738. <https://doi.org/10.1002/cpe.5738>