



An Ensemble of 2.5D ResUnet Based Models for Segmentation of Kidney and Masses

Cancan Chen¹ and Rongguo Zhang^{1,2}(✉)

¹ Infervision Advanced Research Institute, Beijing, China

² Academy for Multidisciplinary Studies, Capital Normal University, Beijing, China
zrongguo@cnu.edu.cn

Abstract. The automatic segmentation of kidney, kidney tumor and kidney cyst on Computed Tomography (CT) scans is a challenging task due to the indistinct lesion boundaries and fuzzy texture. Considering the large range and unbalanced distribution of CT scans' thickness, 2.5D ResUnet is adopted to build an efficient coarse-to-fine semantic segmentation framework in this work. A set of 489 CT scans are used for training and validation, and an independent never-before-used CT scans for testing. Finally, we demonstrate the effectiveness of our proposed method. The dice values on test set are 0.954, 0.792, 0.691, the surface dice values are 0.897, 0.591, 0.541 for kidney, tumor and cyst, respectively. The average inference time of each CT scan is 20.65s and the max GPU memory is 3525 MB. The results suggest that a better trade-off between model performance and efficiency.

Keywords: Coarse-to-fine · Semantic-segmentation · ResUnet · KiTS23

1 Introduction

In recent years, over 430,000 people are diagnosed with kidney cancer and roughly 180,000 deaths are caused by kidney cancer annually [10]. Kidney tumors are found in an even larger number each year, and in most circumstances, it's not currently possible to radiographically determine whether a given tumor is malignant or benign [1]. Computer tomography (CT) scans is an import clinical tool to diagnose and detect kidney tumors. Surgery is the most common treatment option. Radiologists and surgeons are also dedicated to study kidney tumors on CT scans to design optimal treatment schedule by annotating the kidney and its masses manually. However, the manual annotation is a repetitive heavy laborious work and always subjective and varied from the different radiologists. Considering this, automatic segmentation of kidney and kidney tumors is a promising tool for alleviating these clinical problems.

Based on the 2019 and 2021 Kidney Tumor Segmentation Challenge [3,4], KiTS23 features an expanded training set (489 cases) with a fresh never-before-used test set (110 cases), and aims to serve a stronger benchmark and develop

the best automatic semantic segmentation system for kidney tumors. Besides, hardware (GPU, CPU, etc.) about average inference time of each case are also real factors in clinical application scenes, so it is important to balance the performance and efficiency of the automatic semantic segmentation system.

In this paper, based on the original ResUnet [2], we propose an efficient coarse-to-fine semantic segmentation framework to automatically segment kidneys and tumors. In the coarse segmentation stage, the whole CT images are re-sampled to $128 \times 128 \times 128$ as the input. In the fine segmentation stage, we firstly obtain regions of interest (ROIs) for the kidney on the whole CT images based on the coarse segmentation mask, and according to this, randomly crop cubes along z-axis, which are re-sampled to $48 \times 224 \times 384$ as the input. Besides, a cascaded model, consisting of the kidney segmentation model and the kidney-tumor-cyst segmentation model, is applied on the second stage.

The main contributions of this work are summarized as follows:

- We propose a coarse-to-fine semantic segmentation framework, which can effectively segment kidney, kidney tumor and kidney cyst from the abdominal CT images.
- We firstly conduct a statistical analysis on the spacing resolution of all CT images, especially the thickness distribution at the z-axis, which sparks the major design ideas about the random cropping method, patch size and 2.5D ResUnet structure on the fine segmentation stage.
- We evaluate our proposed framework by 5-fold cross validation on Kits23 data set.

2 Methods

Semantic segmentation of organs and lesions is a common task for medical image analysis. There are already numerous accurate and efficient algorithms for medical image segmentation, such as U-Net [9], ResUNet [2], nnU-Net [5], et al. Based on the natural properties of Kits23 CT images and the strong baseline [3], we develop a whole-volume-based coarse-to-fine framework as follows, which consists of coarse segmentation, fine kidney segmentation (two-classification task of kidney and others) and fine tumor-mass segmentation (three-classification task of tumor, cyst and other kidney regions).

2.1 Preprocessing

Our proposed method includes the following preprocessing steps:

- Cropping strategy:

In the coarse segmentation stage, the input is the whole volumes. In the fine segmentation stage, the kidney ROIs are firstly cropped from the whole volumes based on the coarse segmentation mask, and after that, we randomly crop 3D cubes from the kidney ROIs only along z-axis to ensure 2D kidney scans' structural integrity.

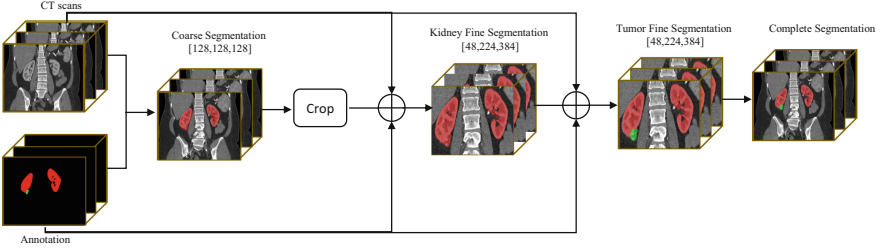


Fig. 1. An overview of our coarse-to-fine segmentation framework.

- Re-sampling method for anisotropic data:
The original images are re-sampled to $128 \times 128 \times 128$ for coarse segmentation. In the fine segmentation stage, if the shape of the cropped kidney ROI is $d \times w \times h$, it will be resampled to $d \times 224 \times 384$ (in this work, $d = 48$), i.e., no-re-sampling at z-axis direction, and re-sampling at x/y-axis direction due to the shape distribution of all kidneys.
- Intensity normalization method:
Images are clipped to range $[-200, 400]$ and normalized to range $[-1, 1]$.
- Others:
To improve the training and testing efficiency, mixed precision is adopted in the whole process of our framework working.

2.2 Proposed Method

Our proposed framework is shown in Fig. 1. The details of two stages are addressed as follows.

Coarse Segmentation. We firstly use a original ResUnet [2] to obtain the coarse segmentation mask of all kidneys, and the input size is $128 \times 128 \times 128$. The kidney tumor and masses are always located in the kidney region. Based on this, the kidney ROI of each CT image is cropped as the input of the next segmentation stage. This step reduces the computational cost of irrelevant information on this task and preserves all segmentation target.

Fine Segmentation. The fine segmentation consists of kidney fine segmentation and lesion fine segmentation. Notably, the thickness range of all CT scans is between 0.5 mm and 5 mm. To resolve the data heterogeneity, cropping or re-sampling should be used. Considering the framework efficiency, the kidney ROIs are re-sampled to the fixed size at x and y direction, and then, we crop the cubes from kidney ROIs only along z axis. That's to say, if the shape of the kidney ROI is $d \times w \times h$, it will be re-sampled to $d \times 224 \times 384$, and the re-cropped cube size for fine segmentation is $48 \times 223 \times 384$ in this work. Finally, we adopt 2.5D ResUnet as the segmentation backbone. Network architecture has

3 down-sample layers, 3 up-sample layers, and no down-sample at z direction for the high-performance and high-efficiency of our framework, which is shown in Fig. 2.

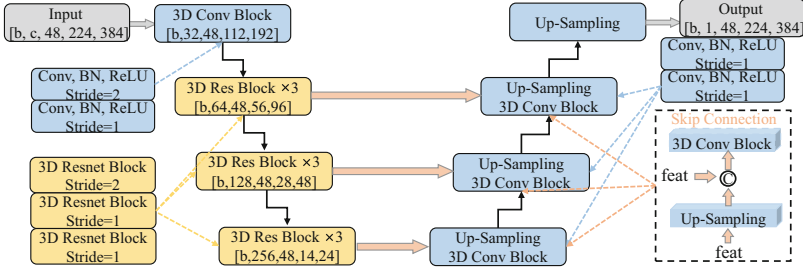


Fig. 2. Our proposed network architecture.

Loss Function. We use the summation of the weighted Dice loss and Cross-Entropy loss as the final compound loss function which has been proved to be robust in various medical image segmentation tasks [7].

Other Tricks. The mixup [12] and hard examples mining are adopted in the model training process, both of which significantly improve the ResUnet’s fitting capability.

2.3 Post-processing

In the inference process, the connected component analysis [11] is applied to avoid the influence of noise. Based on the natural attributes of the kidney and lesions, we choose connected component regions larger than 10000 pixels as the final segmentation results. Notably, we abandon the multi-models ensemble method for efficient inference. Our method consists of the coarse segmentation model, the kidney fine segmentation model (background, kidney) and the lesion fine segmentation model (kidney, cyst, tumor). The final result is the average of the two predictions for the original image and the mirror image along the z-axis.

3 Results

3.1 Dataset and Evaluation Measures

The KiTS23 organizer has publicly released an expanded training set, totally 489 cases, based on KiTs19 and KiTs21. The volumetric Dice coefficient and the Surface Dice are used to evaluate algorithms, and the following Hierarchical Evaluation Classes (HECs) will be used: Kidney + Tumor + Cyst, Tumor + Cyst and Tumor only.

3.2 Implementation Details

Environment Settings. The development environments and requirements are presented in Table 1.

Table 1. Development environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.06 LTS
CPU	Intel(R) Core(TM) i9-10900X CPU @ 3.70 GHz
RAM	96 GB
GPU (number and type)	Four NVIDIA RTX A4000 16G
CUDA version	11.5
Programming language	Python 3.7
Deep learning framework	Pytorch (Torch 1.7.1+cu110, torchvision 0.8.2)
Specific dependencies	
(Optional) Link to code	

Training Protocols. In our training process, we performed the following data augmentation with project MONAI [8]: 1) randomly crop the volumes with range $[0.6, 1.3]$; 2) add brightness, contrast and gamma augmentation on the volumes and lesions with range $[0.6, 1.5]$, respectively. 3) random elastic transform with prob=0.5 and with sigma from range 3 to 5 and magnitude from range 100 to 200; 4) clip volumes to range $[-1, 1]$. Details of our training protocols are shown in Table 2 and Table 3.

Table 2. Training protocols for coarse segmentation.

Network initialization	“he” normal initialization
Batch size	4
Patch size	$128 \times 128 \times 128$
Total epochs	300
Optimizer	ADAMW [6] ($weightdecay = 1e - 4$)
Initial learning rate (lr)	$1e-4$
Lr decay schedule	CosineAnnealing

Table 3. Training protocols for fine segmentation.

Network initialization	“he” normal initialization
Batch size	4
Patch size	$48 \times 224 \times 384$
Total epochs	600
Optimizer	ADAMW [6] ($weightdecay = 1e - 4$)
Initial learning rate (lr)	$1e-4$
Lr decay schedule	CosineAnnealing

3.3 Results on Cross Validation and Test Data

Originally, our proposed framework would be evaluated by 5-fold cross validation. However, we only train and evaluate model on fold-0 data set due to the time and computation resource constraints, and all scores are listed in Table 4. The average inference time on fold-0 validation (98 cases) and test set (110 cases) is 19.22s and 20.65s, respectively. The max GPU memory at inference step is 3525 MB.

Table 4. Results of our proposed method on fold-0 and test set.

Targets	Dice (Fold-0)	Surface Dice (Fold-0)	Dice (Test)	Surface Dice (Test)
Kidney	0.9661	0.9408	0.954	0.897
Cyst	0.8580	0.7365	0.792	0.591
Tumor	0.8591	0.7329	0.691	0.541

4 Conclusion

Based on 2.5D ResUnet, we propose a efficient coarse-to-fine framework for the automatic segmentation of kidney and masses. The experimental results indicate that our framework is effective, but the segmentation robustness of kidney tumors and cysts need further improvement. One possible reason is that the capability level of single model has a lower upper-limit for the hard segmentation task. Thus, the ensemble of multi-models is an alternative solution after balancing performance and efficiency.

References

1. De, L., Alberto, D., Ivan, P.: Imaging and screening of kidney cancer. *Radiol. Clin.* **55**, 1235–1250 (2017)
2. Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C.: ResUnet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote. Sens.* **162**, 94–114 (2020)
3. Heller, N., et al.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge. *Med. Image Anal.* **67**, 101821 (2021)
4. Heller, N., et al.: An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in CT imaging. *Proc. Am. Soc. Clin. Oncol.* **38**(6), 626 (2020)
5. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211 (2021)
6. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint [arXiv:1711.05101](https://arxiv.org/abs/1711.05101) (2017)
7. Ma, J., et al.: Loss odyssey in medical image segmentation. *Med. Image Anal.* **71**, 102035 (2021)
8. Nic, M., Wenqi, L., Richard, B., Yiheng, W., Behrooz, H.: MONAI. <https://github.com/Project-MONAI/MONAI>. [Version 0.8.1]
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
10. Sung, H., et al.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021)
11. William, S., Nico, K.: Odidev: CC3D. <https://github.com/seung-lab/connected-components-3d>. [Version 3.10.1]
12. Zhang, H., Cisse, M., Dauphin, N, Y., Lopez-Paz, D.: Mixup: beyond empirical risk minimization. arXiv preprint [arXiv:1710.09412](https://arxiv.org/abs/1710.09412) (2017)