

Segmentation of Kidney and Renal Tumor in CT Scans Using Convolutional Networks

Shaofeng Yuan¹, Feng Yang², Yujiao Tang², Yanyan Xing², Liyun Zhang²

¹ Shanghai United Imaging Healthcare Co. Ltd., Shanghai, China
shaofeng.yuan.smu@gmail.com
shaofeng.yuan@united-imaging.com

² Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University, Guangzhou, China
yangf@smu.edu.cn

Abstract. Accurate segmentation of kidney and renal tumor in CT images is a prerequisite step in surgery planning. However, this task remains a challenge. In this report, we use convolutional networks (ConvNet) to automatically segment kidney and renal tumor. Specifically, we adopt a 2D ConvNet to select a range of slices to be segmented in the inference phase for accelerating segmentation, while a 3D ConvNet is trained to segment regions of interest in the above narrow range. In localization phase, CT images from several publicly available datasets were used for learning localizer. This localizer aims to filter out slices impossible containing kidney and renal tumor, and it was fine-tuned from AlexNet pre-trained on ImageNet. In segmentation phase, a simple U-net with large patch size ($160 \times 160 \times 80$) was trained to delineate contours of kidney and renal tumor. In the 2019 MICCAI Kidney Tumor Segmentation (KiTS19) Challenge, 5-fold cross-validation was performed on the training set. 168 (80%) CT scans were used for training and remaining 42 (20%) cases were used for validation. The resulting average Dice similarity coefficients are **0.9662** and **0.7905** for kidney and renal tumor, respectively.

Keywords: Localization, Segmentation, Convolutional Networks, AlexNet, U-net.

1 Introduction

There were more than 400,000 kidney cancer diagnoses worldwide in 2018 resulting in more than 175,000 deaths [1]. Accurate segmentation of kidney tissues and renal tumors in CT scans is challenging due to the following reasons: (1) large inhomogeneity of the kidney; (2) high anatomical variations between patients both kidney and tumor; (3) similar intensities and blurred boundaries of adjacent organs such as spleen and liver; (4) varying intensities due to contrast agent and capture time [2].

Many researchers have proposed a number of methods for (semi-) automatic segmentation of kidney or kidney tumor from CT images. In recent years, deep learning based approaches have been widely used in medical image analysis [3-5]. Among

them, U-net and V-net are the most popular ones. However, Yang et al. [6] proposed 3D FCN with pyramid pooling module for kidney and renal tumor segmentation, and obtained considerable improvement. Yu et al. [7] presented Crossbar-Net and cross-bar patch sampling strategy devoting to 2D kidney tumor segmentation. In this report, we use simple 3D U-net with large patch size ($160 \times 160 \times 80$) to segment kidney and renal tumor. In order to accelerating the segmentation process in inference phase, we adopt slice-wise classification along the z-axis based on 2D ConvNet for determine a narrow range containing regions of interest.

2 Methods

2.1 Abdomen Localization Using 2D ConvNet

Besides dedicated abdominal scans, the KiTS dataset contains some whole-body CT scans, which may be quite large and often varies a lot (as shown in Fig.1). If we can select a limited range along the z-axis, computation in the inference phase of segmentation will be saved. A kidney is bounded by the abdomen, whatever it is healthy or pathological. Following [8], we perform slice-wise classification by assigning a cross-section image to one of three classes: above abdomen, among abdomen and below abdomen.

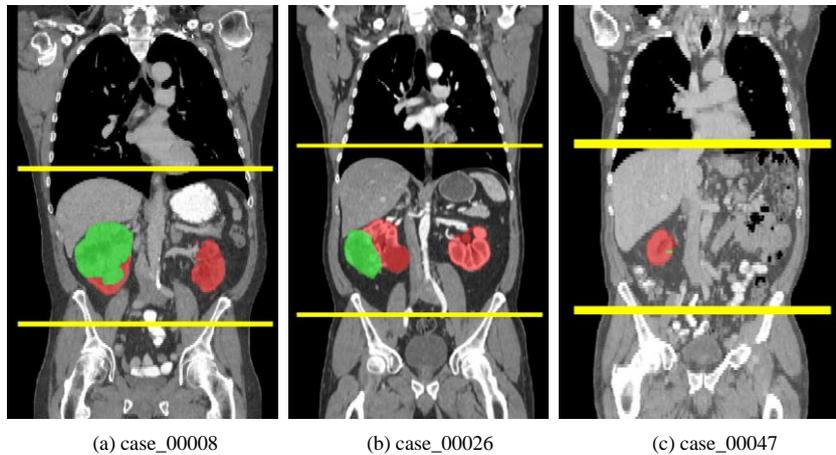


Fig. 1. Three whole-body CT scans from the KiTS training dataset. Red and green colors denote ground-truths of kidney and tumor. Yellow color denotes prediction of abdomen localizer.

Several convolutional networks are tried to fine-tune to perform the slice-wise classification. We found the performance of AlexNet is better than VGGNet-16 with

batch normalization, ResNet-50 and DenseNet-121. A compelling feature of AlexNet is that at inference all z-axial slices can be stacked in an input tensor, and then fed into the trained network. The 2D input image along the z-axis is first resized to 256×256 , and then cropped randomly to 224×224 . We performed random horizontal flipping and random rotating for data augmentation.

Training images in abdomen localization are from the LiTS Challenge, the LUNA16 (only subset0) and the MALBCV datasets. We used images from KiTS dataset as validation set. All annotations are labeled by authors using ITK-SNAP. The annotations include a small blue disk set at the bottom of the heart as the upper limit of the abdomen, and a small yellow disk set at the top of the pubic symphysis as the lower limit of the abdomen.

2.2 Kidney and Renal Tumor Segmentation Using 3D ConvNet

3D U-net [4] with five times pooling along the x-, y-axis and four times pooling along the z-axis and large patch size ($160 \times 160 \times 80$) was trained on the KiTS training set. Following [9], instance normalization layers are added to deep 3D U-net and all ReLU nonlinearities are replaced with leaky ReLU ones. The loss function in deep 3D U-net is the combination of Dice loss and cross entropy loss for Handling input and output imbalance problem.

2.3 Implementation Details

In abdomen localization, batch size of 1024 was used to fine-tune AlexNet pre-trained on ImageNet. SGD optimizer with initial learning rate of 0.008, momentum of 0.9, and epoch of 50 was used to update weights of network. In epoch 30, learning rate was halved. Weighted cross entropy loss was used, and weights of 3-class are 1.000, 0.514 and 1.042, respectively. In kidney and renal tumor segmentation, due to large patch size, a small batch size of 2 was used to train 3D U-net from scratch. Adam optimizer with initial learning rate of 0.0003 and epoch of 500 was used. Weights of combining Dice loss and cross entropy loss are set 1. All steps have been implemented using PyTorch 0.4.1 and performed on a workstation with two GPUs of TITAN Xp.

3 Results

3.1 Datasets

The experimental data in abdomen localization is collected from the LiTS Challenge¹, the LUNA16² (only subset0) and the MALBCV³ datasets. Total 340 (201+89+50) cases are used to training abdomen localizer. The 210 cases from the KiTS Challenge⁴ [10] are used as validation set. All these CT volumes are first re-sampled to $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$, and then all axial images are resized to 256×256 pixels. For training

¹<https://competitions.codalab.org/competitions/17094>

²<https://luna16.grand-challenge.org>

³<https://www.synapse.org/#!Synapse:syn3193805/wiki>

⁴<https://kits19.grand-challenge.org>

set, Class **lower** (lower the abdomen), Class **middle** and Class **upper** have 19,082, 37,131 and 18,320 images. For validation set, three classes have 12,800, 34,123 and 4,198 images. Obviously, in theory, 33% of images from the KiTS Challenge will be filtered out.

The experimental data in segmentation is collected from the KiTS Challenge. There 210 cases in the training set, and 90 cases in the test set. We performed 5-fold cross-validation on the training set for evaluating the proposed **loc+seg** method. The segmentation results are obtained using ensemble of these 5 models.

3.2 Evaluation metrics

The rate of filtering out (FOR) non-ROI slices in a CT volume is used for evaluating abdomen localization. The Dice similarity coefficients of kidney and tumor are used for evaluating ROI segmentation and comparing U-net with several variants of V-net.

3.3 Localization accuracy

We found 28% of slices from 210 cases in the training set and 32% of slices from 90 cases in the test set were filtered out. The localizer has two benefits, 1) filtering out non-ROI slices accelerating the next segmentation process, 2) filtering out non-ROI slices reducing false positive voxels due to misclassification of other organs or tissues similar to kidney and renal tumor.

3.4 Segmentation accuracy

Table 1 shows the Dice metrics of 5-fold cross-validation and the averaged Dice. The Dice score of kidney is very high, indicating the 3D U-net accurately delineates contours of kidney. Although evaluations performed on different datasets, 3D U-net is better than 3D FCN with pyramid pooling module [6] with 3.5 points. However, as for tumor segmentation, the method proposed in [6] surpasses 3D U-net. Tumor segmentation task need an elaborated ConvNet with more discriminative capacity.

Table 2 shows the comparison of simple 3D U-net and several variants of V-net on

Table 1. Five-fold cross-validation performance on the training set.

	Kidney	Renal tumor
Fold 1	0.9710	0.7810
Fold 2	0.9681	0.8135
Fold 3	0.9533	0.7557
Fold 4	0.9674	0.8038
Fold 5	0.9711	0.7985
Average	0.9662	0.7905

kidney segmentation. NestedV-net is a nested V-net architecture, similar to UNet++ [11]. DualAttV-net is V-net with position and channel dual attention modules [12]. AttGatedV-net is V-net with attention gate mechanism, similar to attention gated networks [13]. DualAttGatedV-net is V-net with dual attention module and attention gate module. Simple 3D U-net surpasses V-net and variants of V-net because of these complex networks may be more difficult to be tuned and optimized.

Fig.2 shows three examples of kidney and renal tumor segmentation with different tumor size. In general, 3D U-net accurately delineates contours of kidney. However, in details, there is still room for improvement.

Table 2. Comparison of 3D U-net, V-net and variants of V-net. Dice scores computed on the last 10 cases from training set.

	3D U-net ¹	3D U-net ²	3D V-net	
Kidney	0.9742	0.9661	0.9598	
	NestedV-net	DualAttV-net	AttGatedV-net	DualAttGatedV-net
Kidney	0.8966	0.9448	0.9564	0.9382

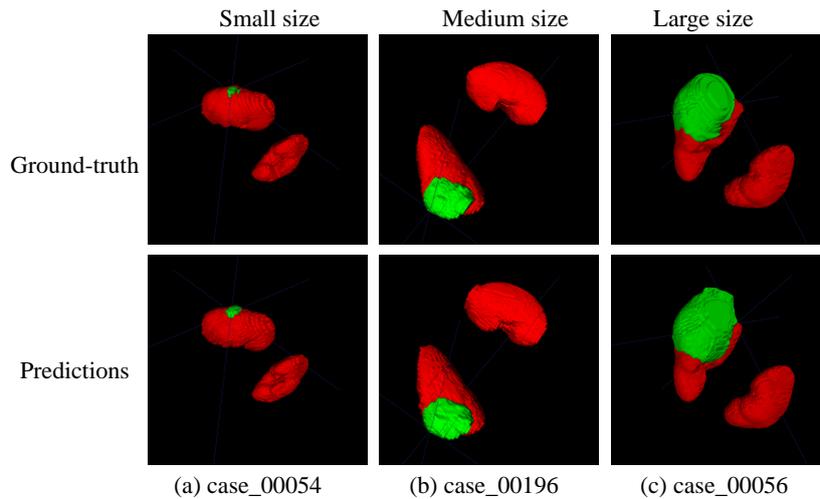


Fig. 2. Three examples of kidney and tumor segmentation with different tumor sizes from the KiTS training dataset. The first row lists ground-truths of kidney and tumor, and the second row shows predictions from 3D U-net.

4 Conclusions

In conclusion, an abdomen localizer and a kidney tumor segmentor are used to process data from the KiTS Challenge. The initial experimental results on KiTS 2019

training set show the superiority of our simple but effective method for segmentation of kidney and renal tumor. In the future, we pay more attention to improve the segmentation performance of tumor.

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. Jin, C., Shi, F., Xiang, D., Jiang, X., Zhang, B., Wang, X., Zhu, W., Gao, E., Chen, X.: 3D fast automatic segmentation of kidney based on modified AAM and random forest. *IEEE transactions on medical imaging* 35(6), 1395–1407 (2016).
3. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciampi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sanchez, C.I.: A survey on deep learning in medical image analysis. *Medical image analysis* 42, 60–88 (2017).
4. 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).
5. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *Proceedings of 3DV 2016*, pp. 565–571 (2016).
6. Yang, G., Li, G., Pan, T., Kong, Y., Wu, J., Shu, H., Luo, L., Dillenseger, J.-L., Coatrieux, J.-L., Tang, L., Zhu, X.: Automatic segmentation of kidney and renal tumor in CT images based on 3D fully convolutional neural network with pyramid pooling module. In: *ICPR 2018*, pp. 3790–3795 (2018).
7. 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* 28(8), 4060–4074 (2019).
8. Zheng, Y., Liu, D., Georgescu, B., Xu, D., Comaniciu, D.: Deep learning based automatic segmentation of pathological kidney in CT: Local versus global image context. In: *Deep learning and convolutional neural networks for medical image computing*, pp. 241–255 (2017).
9. Isensee, F., Petersen, J., Kohl, S.A.A., Jager, P.F., Maier-Hein, K.H.: nnU-Net: Breaking the spell on successful medical image segmentation. *arXiv: 1904.08128* (2019).
10. Heller, N., Sathianathan, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., Dean, J., Tradewell, M., Shah, A., Tejpaul, R., Edgeton, Z., Peterson, M., Raza S., Regmi, S., Papanikolopoulos, N., Weight, C.: The KiTS challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentation, and surgical outcomes. *arXiv: 1904.00445* (2019).
11. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: A nested U-net architecture for medical image segmentation. *arXiv: 1807.10165* (2018).
12. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H.: Dual attention network for scene segmentation. *arXiv: 1809.02983* (2018).
13. 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).