Two-phase Framework for Automatic Kidney and Kidney Tumor Segmentation

Hao Wei*, Qin Wang*, Weibing Zhao*,
Minqing Zhang, Kun Yuan, and Zhen Li†

The Chinese University of HongKong (Shenzhen), China
Shenzhen Reaserch Institute of Big Data, China

Abstract. Precise segmentation of kidney and kidney tumor is essential for computer aided diagnosis. Considering the diverse shape, low contrast with surrounding tissues and small tumor volumes, it's still challenging to segment kidney and kidney tumor accurately. To solve this problem, we proposed a two-phase framework for automatic segmentation of kidney and kidney tumor. In the first phase, the approximate localization of kidney and kidney tumor is predicted by a coarse segmentation network to overcome GPU memory limitation. Taking the coarse prediction from first phase as input, the region of kidney and tumor is cropped and trained in an enhanced two-stage network to achieve a fine-grained segmentation result in the second phase. Besides, our network prediction flows into multiple post-processing steps to remove false positive such as cyst by taking medical prior knowledge into consideration. Data argumentation through registration and ensemble models are used to further enhance performance.

Keywords: Computer aided diagnosis · Automatic Kidney and kidney tumor segmentation · Deep learning · CT images.

1 Introduction

Accurate segmentation of kidney and kidney tumor can largely influence the effect of the following computer-aided treatment of renal cell carcinoma. Precise segmentation of kidney tumor can assist tumor resection surgery in the early stage of renal cell carcinoma, after which patients can heal themselves instead of removing the whole kidney. However, routine manual delineation for kidney tumor is inefficient and inaccurate due to heavy annotation workload and manual labeling error. Thus, automatic semantic segmentation of kidney and kidney tumor is essential and attracts wide attentions in medical image analysis.

But there are lots of difficulties in automatic segmenting kidney and kidney tumor in CT images. Location, morphology and volumetric size of tumors vary a lot between different patients. And in some confused cases, the size of tumor is

* Equal contribution
† Corresponding author

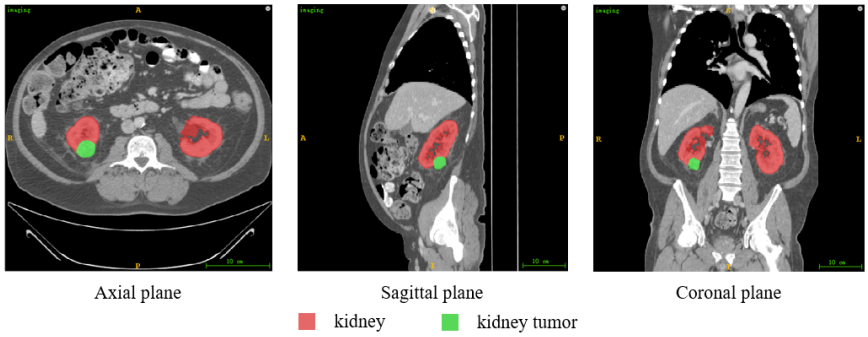

Fig. 1. A segmentation result from three views

much smaller than kidney. Although all the CT images are contrast-enhanced, the contrast and density of kidney tumors and surrounding kidney tissues are still similar. And the intensity of CT images varies a lot due to different dose of contrast media or different scanning period. Also, false positives can be easily caused by tissues, such as cysts, which resemble kidney tumors.

Deep learning based methods can achieve state-of-the-art performance on automatic segmentation of organs and lesions in medical images. Medical image segmentation can be mainly categorized into 2D slice fashion or 3D volumetric fashion. 2D segmentation like U-Net [4] cannot utilize 3D context information. So 3D segmentation like V-Net [3] or 3D U-Net [1] were proposed. But due to the limitation of GPU memory, 3D segmentation network can only take CT images which have been down-sampled to low resolution as input.

In this paper, we propose a cascaded two-phase pipeline for accurate kidney and tumor segmentation. Specifically, to fully utilize the global information with limited GPU memory, we propose to first localize kidney and kidney tumor area through a coarse segmentation network in the first phase. Then we train another 2-stage segmentation network by taking the cropped approximate location as Region of interest (ROI) from the first phase as input. An example of the segmentation results is presented in Fig.1. In summary, the main contributions of our paper are listed as follows.

- In the first phase, coarse localization of kidney and kidney tumor is predicted by a segmentation network. A fixed size of region of interest will be cropped based on the coarse localization.
- In the second phase, cropped ROI is trained in a two-stage coarse-to-fine network to generate a more precise segmentation result.

2 Data Preprocessing

2.1 Data set

Training data consists of 210 patients, and test data consists of 90 patients [2]. Slice thicknesses of raw data range from 1mm to 5mm, while we only train and test on dataset uniformly interpolated to 3mm as we didn't achieve obvious performance gain by taking different thickness as input.

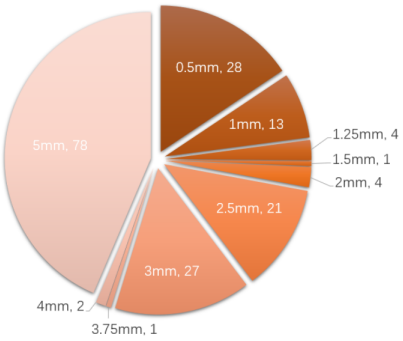

Fig. 2. Slice thickness distribution

2.2 Intensity clipping

Since intensity values over the whole CT image have a wide range, the intensity value is clipped to $[-250, 250]$ which preserves most of the intensity values within kidney and kidney tumor according to the introduction of dataset [2].

2.3 Data augmentation

Registration between CT images are utilized to augment training data. We collect the top 20 worst training data in term of Dice metric, then NiftyReg [5] is used to non-linearly align 10 data randomly selected from the rest 190 data with each of these 20 hard cases respectively. Thus, the training dataset can be increased by 200 CT images with annotations, which tremendously improves the scale of the training dataset representing worse cases. Registration can make the network more stable and converge faster.

3 Architecture

3.1 Overview

Our method consists of two phases. In the first phase, we can get a coarse localization of kidney and kidney tumor by training a V-Net in patch-fashion.

Then in the second phase, the area of kidney and kidney tumor can be cropped and trained in a two-stage segmentation network to generate a more accurate segmentation result. We have done experiments on the original dataset consisting of various slice thickness, whose result indicates that multi-thickness input doesn't make large difference to the segmentation accuracy. Thus, we choose an uniform slice $3mm$ provided by official website. Thus each patient's ROI input is a $384 \times 240 \times 80$ volume.

3.2 Coarse segmentation of kidney and kidney tumor

Due to the limitation of GPU memory, the whole CT images are too large to be fed into deep neural networks. So the original CT images are down-sampled and input to V-Net to localize kidney and kidney tumor coarsely. A bounding box with size $384 \times 240 \times 80$ can contain kidneys and kidney tumors in all training cases. So regions of interest (ROI) in test CT images can be cropped based on the size of bounding box.

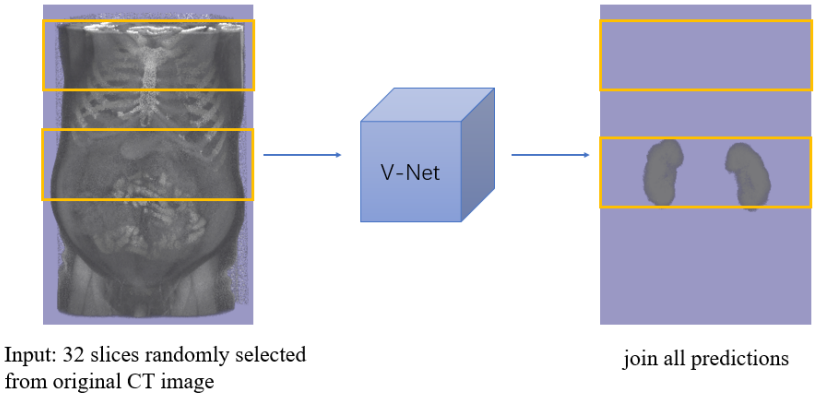

Fig. 3. one-stage v-net and coarse prediction result

3.3 Fine segmentation of kidney and kidney tumor

ROI CT images with fixed size ($384 \times 240 \times 80$) can be gotten from 3.2, which will be entered into a 2 stage coarse-to-fine framework to get a fine segmentation map. In stage one, ROI is entered into a V-Net to get a segmentation map which will serve as a global guidance for segmentation in stage 2. So we will concatenate the coarse segmentation result from stage 1 with the original ROI CT image and enter them into V-Net in stage 2 to produce a more precise segmentation result.

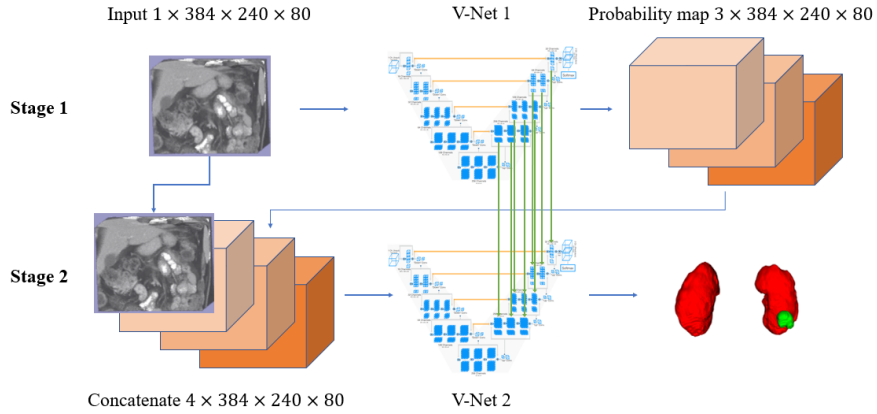

Fig. 4. Enhanced two-stage coarse-to-fine segmentation network pipeline, which consists of skip connections between decoder parts in V-Net1 and V-Net2.

feature map ensemble We add each feature map in the decoder path of V-Net 1 with the corresponding feature map in the decoder path of V-Net 2, which is represented as green connection in Fig.4

3.4 Training Loss

We have tried different training losses, such as Cross Entropy loss, Dice loss, focal loss, where CE loss outperforms the others. Cross Entropy loss is shown as follows,

$$\mathcal{L}(\mathcal{I}, \theta, \mathcal{G}) = -\sum_{i=1}^{N}\sum_{c=1}^{3} g_{i,c} \cdot \ln p_{i,c} \tag{1}$$

where $\mathcal{I}$ denotes the input CT image, θ is network parameters to be optimized, $\mathcal{G}$ denotes the ground truth. $p_{i,c}$ denotes the softmax output indicating the ith voxel being predicted as class c. And only when voxel i belongs to class c, $g_{i,c}$ equals to 1, otherwise 0.

4 Implementation details

Our framework is implemented by PyTorch - 1.0.0. The maximum training epochs is set to be 600. We use Adam as the optimizer with learning rate of $1e-4$ which decays at 200, 400 epoch. Our 2-stage model is trained on 4 NVIDIA TITAN Xp (Pascal) GPUs. Due to the limitation of GPU memory, we use a V-Net to segment region of interest coarsely. Then based on the coarse localization, a region of interest with size of $380 \times 240 \times 80$ is cropped from raw training data and then entered into the 2-stage network.

4.1 Hyper-parameters

We apply our pipeline to segment the CT images which are pre-processed by data preparation. Following the 2-phase pipeline, we pre-process all CT images and feed them to phase 1. We train our model with ADAM optimizer by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The initial learning rate is initialized to 10^{-4} , which is reduced to half every 200 epoch. In our pipeline, all setups of convolutions are same as V-Net with 4 Down Conv stages and 4 Up Conv stages. The convolutions performed in each Down Conv stage use volumetric kernels with size $5 \times 5 \times 5$ voxels. As the data proceeds through different stages along the compression path, the resolution is reduced, and vice versa. Meanwhile, we zero-pad the boundaries of each bounding box to ensure the spatial size of kidney and tumor is fully wrapped by it.

5 inference and post-processing

5.1 Inference

Model ensemble (1) For the same model, we ensemble the prediction of the model loaded with top 5 best parameters in terms of dice on validation set. (2) We ensemble the prediction of different models. (3) We ensemble the prediction of a model trained on different train-valid dataset division.

5.2 Post-processing

To improve the segmented image, further processing is required which is performed in post processing step to eliminate false positive predictions and smooth the segmentation performance. We introduce largest connected region and intensity distribution histogram to extract effective features to differentiate between blurred kidney region and other organ regions. Extensive quantitative and qualitative evaluations on dataset illustrate the superiority of our post processing method with smooth and accurate results.

Largest connected region To eliminate the false positive predictions in liver and other organs, only the top 2 largest connected region is reserved. (Top 1 for cases with only one kidney).

Intensity distribution The intensity distribution histograms indicate that intensity values within tumor and cyst follow normal distribution approximately with different mean, as shown in Fig.5. The average intensity of cyst is around 0, while the average intensity of tumor is much greater than 20. So we can filter the false positive candidate regions whose average intensity is smaller than a threshold like 10.

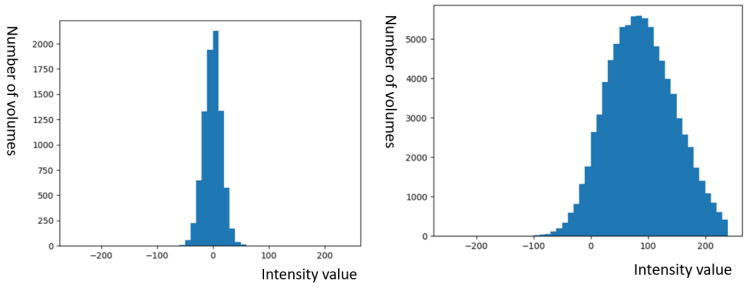

Fig. 5. Intensity value distribution within cyst (left) or tumor (right).

6 Results

First phase result In the first phase, the whole CT image is taken as input. Although the Dice for tumor is not satisfactory, the segmentation accuracy for kidney is good enough for a coarse localization of the region of kidney and kidney tumor in raw CT images.

Table 1. Dice of kidney and kidney tumor in the first phase.

Organ	Kidney	Tumor	Mean
Dice	0.9	0.19	0.545

Second phase result In the second phase, the region of interest cropped based on the rough prediction in phase 1 is entered into Phase 2. As shown in table 2, the segmentation result is improved stage by stage.

Table 2. Dice of kidney and kidney tumor in the second phase.

Stage \ Organ	Kidney	Tumor	Mean
stage 1 (Ensemble)	0.964	0.705	0.8495
stage 2 (Ensemble)	**0.968**	**0.75**	**0.859**

7 Conclusion

We introduce a novel pipeline, called two-phase model, that facilitates the design of, computational savings, deep convolutional networks for CT kidney data.

Kidney is located and segmented through these operations step by step. Although our architecture is inspired by the one proposed in [3], we developed it into two-phase that learns coarse segmentation for bounding box and coarse-to-fine segmentation. Besides, some medical prior knowledge is exploited as post-prepossessing to further improve our prediction. As empirically observed, we improve both results and convergence time to the state-of-the-art result for this competition.

References

1. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
2. 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)
3. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3d Vision (2016)
4. 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)
5. for Medical Image Computing at University College London, T.C.: Niftyreg Software. http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg, february 5, 2019