Segmentation of kidney tumor by multi-resolution VB-nets

Guangrui Mu[1,2], Zhiyong Lin[3], Miaofei Han[1], Guang Yao[1], Yaozong Gao[1]

1 Shanghai United Imaging Intelligence Inc., Shanghai, China

2 Biomedical Engineering Department, Southern Medical University, Guangzhou, China

3 Radiology Department, Peking University First Hospital, Beijing, China

Abstract. Accurate segmentation of kidney tumors can assist doctors to diagnose diseases, and to improve treatment planning, which is highly demanded in the clinical practice. In this work, we propose multi-resolution 3D V-Net networks to automatically segment kidney and renal tumor in computed tomography (CT) images. Specifically, we adopt two resolutions and propose a customized V-Net model called VB-Net for both resolutions. The VB-Net model in the coarse resolution can robustly localize the organs, while the VB-Net model in the fine resolution can accurately refine the boundary of each organ or lesion. We experiment in the KiTS19 challenge, which shows promising performance.

Keywords: Kidney Tumor, Segmentation, Multi-resolution

1 Introduction

Segmentation of kidney and renal tumor in contrast-enhanced computed tomography (CT), which is widely used in clinics for disease diagnosing and treatment planning. The tedious manual segmentation of organs from CT images is very time-consuming, and there is also some inconsistency in the segmentation between different senior doctors. For some lesions, the segmentation is especially challenging: shape and position vary greatly across patients; the lesion boundaries in CT images are of low contrast, and sometimes can be absent.

In recent years, deep learning based methods have been widely used in medical image segmentation [1-2]. Among them, U-Net [3] and V-Net are the most popular ones. V-Net was proposed to replace 2D convolutions in U-Net with 3D convolutions and also incorporate the idea of the residual networks into U-Net. By doing so, V-Net encourages much smoother gradient flow, thus easier in optimization and convergence. In this work, we developed a customized V-Net called VB-Net to segment organs and target tumors. At the same time, different from previous patch-based methods that use sliding windows during application, we achieved 3D fine-grained segmentation results in a much shorter time using 3D fully convolutional network. Validated on this challenge dataset, the proposed VB-Net shows promising results in accuracy, speed and robustness.

2 Method

2.1 Data Preprocessing

Firstly, we truncated the image intensity values of all images to the range of [-200, 500] HU to remove the irrelevant details. Then, truncated intensity values are normalized into the range of [-1, 1]. As the in-plane resolution varies and z-resolutions are not uniform in CT scans, we resample images into the same isotropic resolution. During training, we randomly sample 96×96×96 crops from images, and use them as network input to reduce the GPU memory consumption.

Sometimes it can be difficult to distinguish between cysts and tumors in appearance. A professional doctor assists us to manually annotate the renal cysts as an additional category, and we use this class in our multi-task learning to improve the separation between cysts and tumors. This strategy is verified to be effective in our experiments.

2.2 VB-Net for Accurate Organ Segmentation

V-Net, proposed by Milletari [4], was initially used to segment the prostate by training an end-to-end fully convolutional network on MRI. It is composed of two paths, the left contraction path is used to extract high-level context information by convolutions and down-samplings. The right expanding path uses skip connections to fuse high-level context information with fine-grained local information for precise boundary localization. By means of introducing residual function and skip connection, V-Net show better segmentation accuracy compared with many classical CNNs.

Our VB-Net replaces the conventional convolutional layers inside V-Net with the bottleneck structure. Due to the use of bottle-neck structure, we named the architecture as VB-Net (B stands for bottle-neck). The bottleneck structure consists of three convolutional layers. The first convolutional layer applies a 1×1×1 convolutional kernel to reduce the channels of feature maps. The second convolutional layer performs a spatial convolution with the same kernel size as the conventional convolutional layer. The last convolutional layer applies a 1 ×1×1 convolution kernel to increase the channels of feature maps back to the original size. By performing spatial convolutions on the feature maps with reduced channels, there are two benefits: 1) the model size is largely reduced, e.g., from V-Net (250 MB) to VB-Net (8.8 MB); 2) the inference time is also reduced. With a small model size of VB-Net, it becomes easy to deploy the segmentation network either to cloud or to the mobile applications.

2.3 Multi-resolution strategy

As 3D medical images (e.g., CT, MR) are often large in size， passing the whole 3D image volume into networks will consume a lot of GPU memory, hence increasing the chances of segmentation failure. One solution is to resample the image volume into a lower resolution for segmentation, however, the image details will be lost in this way and the segmentation boundary will be zigzag. Another commonly used strategy is dividing the whole image volume into overlapping sub-volumes using a sliding window. However, this strategy is very time-consuming and not practical in industry deployment.

In this work, we adopt a multi-resolution strategy. Specifically, two VB-Nets are trained separately on different image resolutions. In the coarse resolution (resampled to 6 mm), we train a VB-Net to roughly localize the volume of interest (VOI) for the whole kidney. In the fine resolution (resampled to 1 mm), we train VB-Net to accurately delineate the kidney and tumor boundary within the detected VOI. And in the inference stage, we test the image globally as shown in **Fig. 1**.

2.4 Training Procedure

The training images are resampled to isotropic resolutions and normalized first. In the coarse resolution, we resample images to 6 mm isotropic spacing. We resample images to 1 mm isotropic spacing without any mask dilation. After resampling, 3D sub-image volumes of size 96×96×96 are randomly samples as training crops. In the coarse resolution, we randomly sample sub-volumes from the entire image domain. In the fine resolution, we randomly sample sub-volumes only in the area indicated by the ground-truth mask. In this way, the fine-resolution network will focus more on the organ boundary than the coarse-resolution network. For each sampled image crop, the corresponding mask crop is extracted as the ground-truth mask, which is used as the network prediction target. With pairs of image and mask crops we independently train segmentation networks for coarse-resolution and fine-resolution segmentation, respectively.

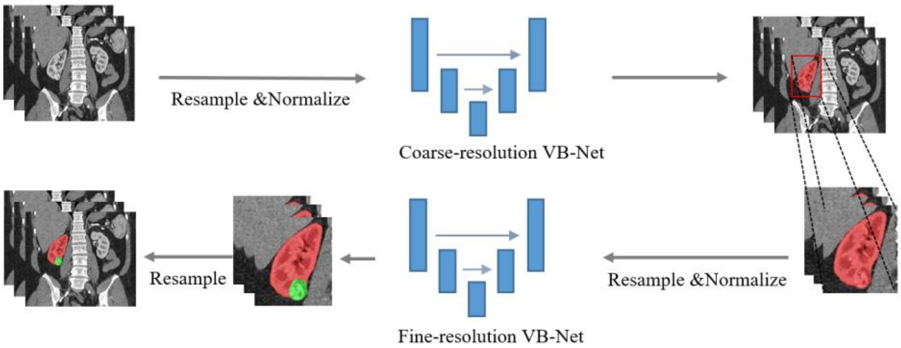

Fig. 1. Inference flow of our propose method.

2.5 Loss Function

In our experiment, we adopt a generalized dice loss function that focuses only on foreground voxels disregarding how many the background voxels in the whole image. The mathematical formulation is given below:

$$\mathrm{D} = \frac{1}{C}\sum_{c=1}^{C}\frac{2\sum_i^N p_c(i)g_c(i)}{\sum_i^N p_c^2(i) + \sum_i^N g_c^2(i)}$$

where the inner summation runs over the N voxels in the image domain, C represents the number of class labels, $p_c(i)$ is the probability of class c at voxel i predicted by the network, $g_c(i) \in \{0,1\}$ is the binary label indicating whether the label of voxel i is class c.

3 Data and Result

3.1 Dataset

There are 210 and 90 abdominal CT scans for training and testing in the KiTS Challenge dataset, respectively. Images in the training set have 512×512 pixels in-plane size with spatial resolution varying from 0.438 mm to 1.04 mm, and the number of slices varies from 29 to 1059 with a slice thickness between 0.5 mm and 5 mm. We randomly split the given 210 training CT volumes into 180 for training and 30 for validation, and evaluate the segmentation accuracy using the Dice score.

3.2 Result

We validated our method on 30 CT scans of the KiTS Challenge, the performance is shown in **Table 1**. To improve performance further, we remove the isolated small segments out of kidney by picking the largest connected component.

Table 1. The experimental results in validation data using our method.

	Kidney		Tumor	
	Dice	Range	Dice	Range
VB-Net	0.974±0.014	0.921~0.990	0.789±0.227	0.817~0.177

4 Discussion

In conclusion, we propose a multi-resolution VB-Net framework to segment kidney and renal tumor. The multi-resolution strategy reduces the GPU memory cost while maintains a high segmentation accuracy especially for kidney, demonstrating the potential for automating kidney segmentation in disease diagnosing and treatment planning. At the same time, we have observed that adding the cyst type during training can greatly enhance the discrimination of the model to distinguish between cysts and tumors. However, for tumors or cysts with uneven densities, our model tends to be misled. In order to solve this problem, we optimized the post-processing algorithm to make corrections to these mis-segmented areas based on the spatial relationship between the cyst and tumor, and their average HU value. The results show that our method is effective in both kidney and kidney tumor segmentation.

References

[1]. Havaei, Mohammad, et al. "Brain tumor segmentation with deep neural networks." Medical image analysis 35 (2017): 18-31.

[2]. Hu, Peijun, et al. "Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets." International journal of computer assisted radiology and surgery 12.3 (2017): 399-411.

[3]. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[4]. Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016.