

Kidney Tumor Segmentation Using Dual Fully Convolutional Networks

Bing Han¹, Ye Yue², and Mengzhang Li³

¹ hanbingocean@126.com

² ianyueregister@gmail.com

³ 173963522@qq.com

Abstract. Kidney tumor segmentation is a hot topic in medical image analysis. In this work, we propose a concise but tricky method, dual Fully Convolutional Networks(FCN) to segment kidney and its surrounding tumor. First FCN is utilized to coarsely detect location of kidney and tumor, while second FCN is employed for refining results of first FCN. In the 2019 Kidney Tumor Segmentation Challenge(KiTS19), 160 patients' CT scans were used for training and our best result were obtained by this two-step FCN models trained with Focal loss.

Keywords: Tumor · Kidney · Segmentation · FCN.

1 Introduction

There are many cases of kidney cancer each year. Although the surgery is its common treatment, kidney cancer detection is difficult because of wide variety in kidney and kidney tumor morphology. Many difficulties lie in this domain, for example, it is very challenging for segmentation between different patients because each shape and position may change greatly.

In the past several years, deep learning methods are popular and widely used in many fields such as computer vision, natural language processing. In the doamin of medical image, deep leaning frameworks such as Fully Convolutional Networks(FCN), U-Net are widely used. In addition, preprocessing of raw data is also very important such as normalization of data, extraction of effective image domain. We developed a tricky method for 2D image treatment and utilizing those 2D image for a dual FCN framework. The result shows it performs well in kidney segmentation and works very robust in kidney tumor segmentation.

2 Method

2.1 Dual FCN for Accurate Segmentation

Fully Convolutional Networks(FCN)[2] was one of the most famous deep learning method for semantic segmentation. FCN learns to combine coarse and high level information with fine and low level information. By introducing upsampling and

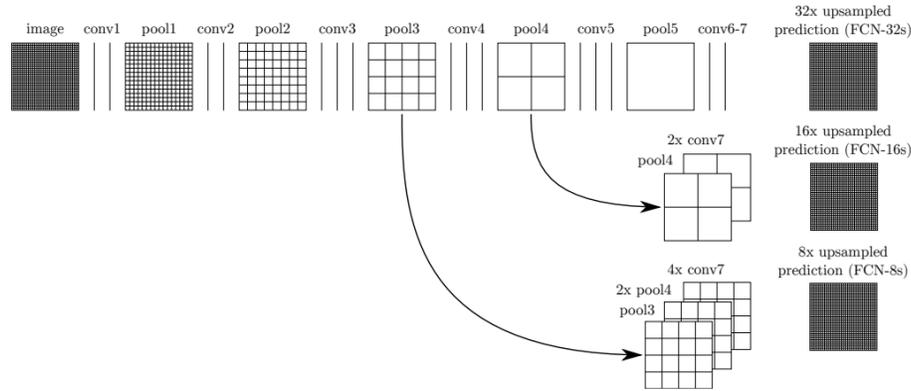


Fig. 1. Classical combination operation of downsampling and upsampling in FCN[2]

multilayer combinations, it's very effective for learning dense prediction information. The architecture of FCN in our model is shown in 1, this classical structure contains two parts. First part is downsampling, which contains convolution layer and pooling operation to extract useful coarse feature information. Second part is upsampling, which contains deconvolution layer and concatenates feature map of different level generated in the first part.

We adopt this classical framework for both coarse kidney detection and fine segmentation of kidney and tumor. There are many robust convolution architectures such as VGG nets[3] and Google nets[4]. To improve convergence speed, we introduced pretrained VGG-16 network's parameters which are publicly available from PyTorch website.

2.2 Two Step Segmentation

Due to the lack of computation resources and large size of many medical images, we use a multi-resolution strategy. Specifically, two step FCN models are trained separately on different tasks. In the coarse resolution task, we train a FCN to localize region of interests(ROI) for kidney and tumor. The result includes prediction of kidney and tumor which will be cropped as 192×192 pixel size for next step: in the fine resolution task, we train another FCN to accurately segment boundary of kidney and tumor.

2.3 Focal Loss Function

Loss function is very important for segmentation and some of them are very popular such as Dice loss, Cross Entropy loss and Focal loss. The Focal loss has many advantages over other algorithms in two-step tasks[1]. First, it enable

model focusing on samples which are difficult to classify thus addressing imbalance problem of foreground and background. The formulation is shown below:

$$FL_{p_t} = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

where γ is the modulating factor. When $\gamma = 0$, Focal loss is equivalent to Cross Entropy loss. When $\gamma > 0$, it reduces loss of samples which are easy to discriminate. Focal loss also uses α to balance the ratio of positive and negative samples. We set $\gamma = 2$ and $\alpha = 4$ in this task.

3 Numerical Experiments

3.1 Dataset

The experiment data are collected from KiTS19 Challenge training and testing dataset. The training and testing datasets include 210 and 90 patients respectively. The imaging and ground truth labels are provided in NIFTI format with *shape(num_slices, height, width)* where *num_slices* is viewed in axial direction.

3.2 Preprocessing and Intensity Normalization

First of all, we transform format from NIFTI to png image in axial plane. To normalize pixel value in png image, we set minimum and maximum HU value -512 and 512 respectively. Value lower than minimum value or higher than maximum value will be set as 0 or 255 in R, G, B three bands. Value within minimum and maximum will be linearly normalized into the range from 0 to 255. For label, the kidney whose value is 1 in NIFTI file will be set as 255 for R bands and tumor whose value is 2 in NIFTI file will be set as 255 for B bands. The background will be set as 0 for all R, G, B bands.

3.3 Patch-wise Network Training

Firstly, we train a FCN model using PNG image from preprocessing step. After that, we localize the kidney and tumor in each slice based on the output of such model.

We locate core image slice which contains the most pixels of kidney and tumor. Those core images are treated with morphology operation: first with closing operation aiming at wiping out some trivial prediction area, and next with opening operation aiming at filling out majority prediction area. Then we count the foreground objects in core image. If the total number of foreground object are two, we record center of them as the center of left and right kidney center.

Then we find neighbor slice of core image which also include kidney and tumor region in coarse resolution result. Finally we crop core image and its neighbor slices with above centers and their size are 192×192 for left and right kidney patch.

These patches are sent to FCN models separately in the fine resolution tasks. We train the patch of left and right kidneys separately because of their surrounding morphology difference.

3.4 Results

We evaluated our models on 90 CT scans of KiTS19 contest. Source code is available at <https://github.com/YueIan/dl-kidney-segment> which is an end-to-end framework, i.e., it includes preprocessing, model training and testing, generation of NIFTI files output automatically without any manual operation.

Fine segmentation output will be used except that it can't get two objects in core image slice in ROI selection (like Case 296). In this case we will use coarse resolution output and resample them to original size. Finally, we will reset kidney and tumor pixel value as 1 and 2 respectively.

The score is shown below:

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

where k and t means kidney and tumor⁴ respectively.

Approximate Kidney Dice	0.935
Approximate Tumor Dice	0.535
Approximate Average Kidney-Tumor Dice	0.735

Table 1. Score of each segmentation tasks on KiTS19 online testing set.

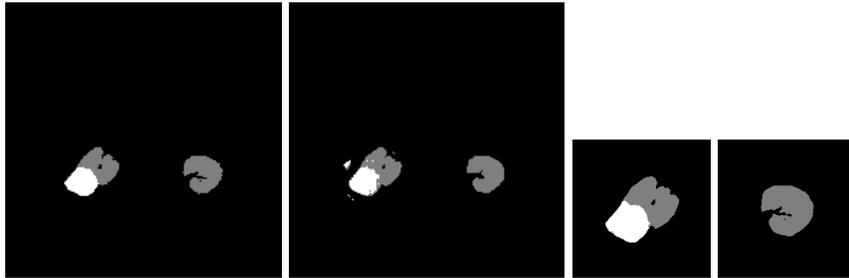


Fig. 2. Two-step segmentation results. Left image is label, middle image is coarse resolution result, right two images are fine resolution results.

⁴ <https://kits19.grand-challenge.org/rules/>

We show results at each step on Fig 2: at coarse resolution task, the left and right kidneys are located but intermediate results are rough. When treated with fine resolution task separately, more refined segmentation results of the right and left kidneys are generated.

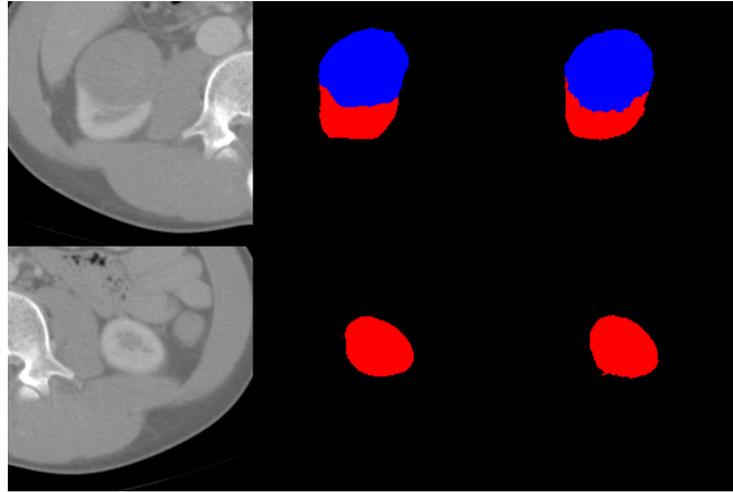


Fig. 3. CT image, kidney and tumor ground truth and their prediction result with dual FCN model.

Just as Fig 3 shows: our model performs well in many cases and strategy that training models separately on left and right kidney is reasonable because their surrounding morphology differences are significant. The prediction results' boundaries are not as smooth as ground truth, while they can generally capture majority of kidney and tumor.

3.5 Computation Cost and Running Time Analysis

We train our model on an Intel Xeon Gold 6128 CPU with 64 GB memory and a NVIDIA Quadro P2000 graphical card with 5GB GPU memory. The training time is about 30 minutes each epoch.

4 Conclusions

In summary, we introduce a dual FCN model to address a multi-resolution task. The experimental result shows although the framework we use is not the most advanced, the score especially kidney segmentation result meets industrial demand.

References

1. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **PP**(99), 2999–3007 (2017)
2. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **39**(4), 640–651 (2014)
3. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *Computer Science* (2014)
4. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions (2014)