Coarse-to-fine Kidney Segmentation Framework Incorporating with Abnormal Detection and Correction

Yue Zhang[1,2], Jiong Wu[1,3], Yifan Chen[4], Ed X. Wu[2], and Xiaoying Tang[1]

[1] Southern University of Science and Technology, Shenzhen 518055, China
`tangxy@sustech.edu.cn`
[2] The University of Hong Kong, Hong Kong 999077, China
[3] Sun Yat-sen University, Guangzhou 510006, China
[4] University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract. In this work, we formulate this segmentation problem into two sub-task: 1) kidney segmentation 2) tumor segmentation. In the first task, three 2D CNN are used to separate the kidney(including tumor)region with background (two class segmentation). In the second task, one 3D CNN and one two channel 2D CNN are used to separate detect and refine tumor within a relative smaller image region.

Keywords: Prostate Segmentation · Deep learning · Coarse-to-fine · Convexity-preserving.

1 Kidney segmentation

Let an CT image be $\mathbf{X}$ and the corresponding ground truth segmentation be $\mathbf{Y}$ where $y_i = 1$ indicates a foreground voxel (including kidney and tumor). Both image size and voxel size are usually varied for clinical dataset. Since spatially inconsistent data might not be ideal for machine learning applications, we firstly interpolated all the dataset to the same voxel size $d \times h \times w$ but the corresponding image size $D \times H \times W$ still varied. Previous coarse-fo-fine work [?] inspires us to make use of a coarsely predicted segmentation mask to constrain the input region with the same image size and the same pixel size. In other words, we use one CNN (coarse segmentation model $\mathbb{C}$) to find the rough location of the kidney and then use another CNN (fine segmentation model $\mathbb{F}$) to localize the kidney more accurately. We also design a CNN (abnormal correction model $\mathbb{A}$) to correct the ill segmentation in coarse stage. $\mathbb{C}$, $\mathbb{A}$ and $\mathbb{F}$ have the same network architecture and the architecture used in this work is U-net [4].

1.1 Training

Firstly, we divide each 3D volume $\mathbf{X}$, $\mathbf{Y}$ into a set of 2D slices $\{\mathbf{x}, \mathbf{y}\}$ with image size $H \times W$ and pixel size $h \times w$. The 2D slices are typically isotropic in terms of both size and resolution, i.e., $H = W$ and $h = w$. Otherwise, resizing and

padding can be easily adopted to make the 2D slices isotropic. For 2D slices of different subjects, either the image size or the pixel size can be different.

To begin our C2F segmentation pipeline, we conduct two-fold preprocessing. On the one hand, we directly resize all $\{\mathbf{x}, \mathbf{y}\}$ to be of the same image size $H^{\mathbb{C}} \times W^{\mathbb{C}}$ (this transformation is denoted as $\mathcal{R}_W$) and obtain 2D image and ground truth pairs $\{\mathbf{x}^{\mathbb{C}}, \mathbf{y}^{\mathbb{C}}\}$, which are then fed into a coarse segmentation model $\mathbb{C} : \mathbf{p}^{\mathbb{C}} = f(\mathbf{W}^{\mathbb{C}}, \mathbf{x}^{\mathbb{C}})$, where $\mathbf{W}^{\mathbb{C}}$ denotes the model parameters and $\mathbf{p}^{\mathbb{C}}$ is the predicted mask at the coarse stage. On the other hand, we crop images with the same pixel size ($w \times h$ mm^2) to obtain image patches of the same image size $W^{\mathbb{F}} \times H^{\mathbb{F}}$ surrounding a single kidney (this transformation is denotes as $\mathcal{C}^{\mathbb{F}}$). The processed images and labels $\{\mathbf{x}^{\mathbb{F}}, \mathbf{y}^{\mathbb{F}}\}$ are then fed into a fine segmentation model $\mathbb{F} : \mathbf{p}^{\mathbb{F}} = f(\mathbf{W}^{\mathbb{F}}, \mathbf{x}^{\mathbb{F}})$. To train a abnormal correction model $\mathbb{A}$, the similar cropping operation $\mathcal{C}^{\mathbb{A}}$ was used along sagittal plane to get a image patches with the same image size ($D^{\mathbb{A}} \times H^{\mathbb{A}}$) as well the same pixel size ($d \times h$ mm^2).

In the context of a deep segmentation networks, the Dice loss $\mathcal{L}(\mathbf{p}, \mathbf{y})$ is optimized with respects to $\mathbf{W}$ via gradient back-propagation. The objective function is

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}(\mathbf{p}_n, \mathbf{y}_n), \tag{1}$$

where N denotes the total number of samples, $\mathbf{W}^*$ denotes the optimal weights obtained from the training procedure. After the training process, $\mathbf{W}^{*\mathbb{C}}$, $\mathbf{W}^{*\mathbb{A}}$ and $\mathbf{W}^{*\mathbb{F}}$ are saved.

1.2   Testing

The overall testing flow chart is shown in Fig. 1 and the voxel spacing of input images has been normalized to be $d \times h \times w$ like training stage. The image transformation operations $\mathcal{R}_W$ and cropping operation $\mathcal{C}^{\mathbb{F}}$ and $\mathcal{C}^{\mathbb{A}}$ are similar to that used in training stage. The only difference is that the cropping operation refer the ground truth in training stage but refer the coarse segmentation in testing stage. $\mathcal{R}_W^{-1}, \mathcal{C}^{-\mathbb{F}}$ and $\mathcal{C}^{-\mathbb{A}}$ respectively denote their inverse transformation operations.

Coarse-to-fine prediction At the coarse segmentation stage, we resize all 2D CT slices $\mathbf{x}$ at the axial view to be of image size $H^{\mathbb{C}} \times W^{\mathbb{C}}$ and predict the coarse segmentation results $\mathbf{S}^{\mathbb{C}}$ by

$$\begin{aligned}
\mathbf{p}^{\mathbb{C}} &= f(\mathbf{W}^{\mathbb{C}*}, \mathcal{R}_W(\mathbf{x})) \\
\mathbf{S}^{\mathbb{C}} &= \{\mathcal{R}_W^{-1}(\mathbf{p}^{\mathbb{C}})\}.
\end{aligned} \tag{2}$$

where $\{\cdot\}$ denote composing all axial sliced segmentation results $\mathbf{p}^{\mathbb{C}}$ of one 3D coarse segmentation results $\mathbf{S}^{\mathbb{C}}$ for each 3D CT image $\mathbf{X}$.

And then $\mathbf{S}^{\mathbb{C}}$ will be judged according to the criterion defined in 1.2. Under different circumstances, the mask $\mathbf{M}$ used to guided fine segmentation stage is
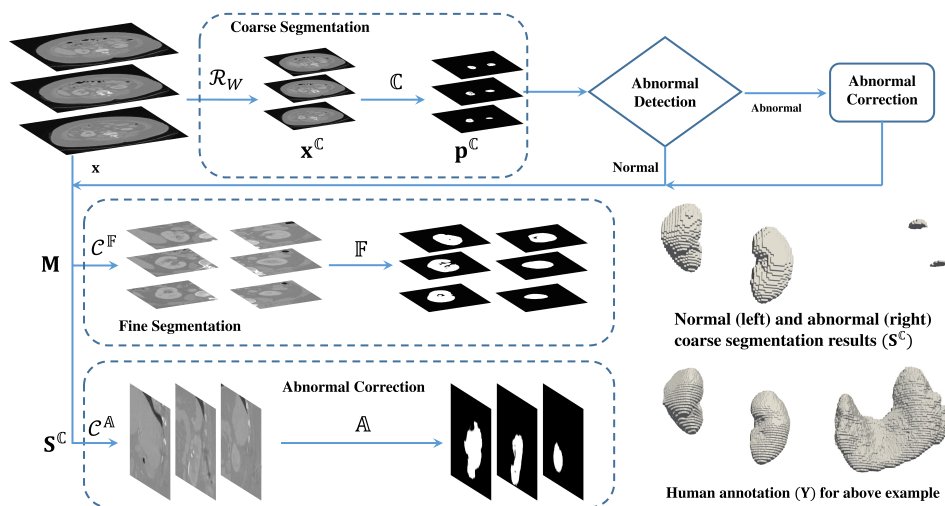
Fig. 1. Flow chart of the kidney segmentation

defined by.

$$\mathbf{M} = \begin{cases} \mathbf{S}^{\mathbb{C}}, & \text{Normal} \\ \{\mathcal{C}^{-\mathbb{A}}(f(\mathbf{W}^{\mathbb{A}*}, \mathcal{C}^{\mathbb{A}}(\mathbf{x}, \mathbf{S}^{\mathbb{C}})))\}, & \text{Abnormal} \end{cases} \tag{3}$$

For normal case, $\mathbf{S}^{\mathbb{C}}$ was directly set as $\mathbf{M}$. For abnormal case, $\mathbf{S}^{\mathbb{C}}$ was used to decide the centroid in sagittal plane and the cropping operation $\mathcal{C}^{\mathbb{A}}$ was used to get image patches with size $D^{\mathbb{A}} \times H^{\mathbb{A}}$. $\mathcal{C}^{-\mathbb{A}}$ was used to pad the predication to original size using 0.

At the fine segmentation stage, we crop the resultant images to obtain image patches of image size $H^{\mathbb{F}} \times W^{\mathbb{F}}$ according to the separated centroid in axial plane decided by $\mathbf{M}$. Two kidneys are predicted separately in fine segmentation stage by

$$\mathbf{S}^{\mathbb{F}} = \{\mathcal{C}^{-\mathbb{F}}(f(\mathbf{W}^{\mathbb{F}*}, \mathcal{C}^{\mathbb{F}}(\mathbf{x}, \mathbf{M})))\} \tag{4}$$

Abnormal detection A typical drawback of coarse-to-fine strategy is that the performance of the fine model depends on that of the coarse model. For example, most of people have two kidneys, however there are a few only have one kidney. As shown in Fig. 1, the abnormal coarse segmentation result is fail to detect the whole kidney.

For these abnormal cases, we design an automatic abnormal detection method based on component analysis and correct it using a CNN [1]. Specifically, we first extract all sub structures of the coarse segmentation result $\mathbf{S}^{\mathbb{C}}$ using connected component analysis [5]. The voxel number of each sub structures can be counted.

---

[1] This abnormal correction can also be designed as an iterative model. But based on our experience, one iteration is enough for most of cases.

Given that kidney volumes were 202 ± 36 ml for men and 154 ± 33 ml for women [6] and we have normalized all image to the same voxel size at the beginning, we can set a threshold voxel number TH_{vn} to count the kidney number N_{kidney}. Based on N_{kidney} detected in $\mathbf{S}^{\mathbb{C}}$, we define a discriminate criterion like

$$\begin{cases} N_{kidney} = 2, & \text{Normal} \\ else, & \text{Abnormal} \end{cases} \tag{5}$$

1.3   Implementation Details

The normalized spacing size $d \times h \times w$ was set as $3 \times 0.7816 \times 0.7816$ mm^3 as the public dataset has the interpolated version with this voxel spacing. The normalized image size in three CNN models $(\mathbb{C}, \mathbb{A}, \mathbb{F})$ are set as $H^{\mathbb{C}} = W^{\mathbb{C}} = 128$, $H^{\mathbb{A}} = 256$, $D^{\mathbb{A}} = 64$, and $H^{\mathbb{F}} = W^{\mathbb{F}} = 160$ empirically. The threshold voxel number TH_{vn} was set as 10000, which is roughly equal to 18 ml based on previous setting voxel spacing. We set this value relative lower to the real kidney volume to reduce the false negative rate and this value is larger enough to remove the noisy small lesions and judge the predicted kidney number. The hardware and software we used is the same as that introduced in [7]

2   Kidney tumor detection and segmentation

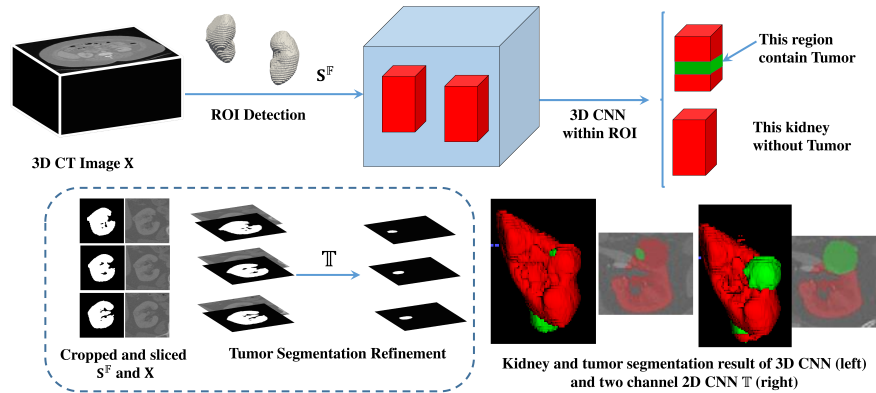As shown in Fig. 2, the inputs are sampled within the bounding box decided by fine segmentation results.

Fig. 2. Flow chart of the tumor segmentation.

2.1   3D CNN with Dense training

The tumor usually occupies a very small part of an CT volume. To balance the distribution of training samples from different segmentation class, N 3D patch pairs $\mathbf{x}$ and $\mathbf{y}$ are randomly selected from the $\mathbf{X}$ and $\mathbf{Y}$ via the sampling method in Dense training [2] in each epoch. In other words, the center of $N/3$ samples are in the background, $N/3$ samples are in the kidney region and the othesr are in the tumor. The architecture we used is 3D FCN with a long skip connection [3]. Let Θ be the network parameters (i.e., convolution weights and biases) of model $\mathbb{M}$ and $p(\mathbf{y}_i|\mathbf{x}_i;\Theta)$ be the predicted probability of image path $\mathbf{x}_i$. The negative log-likelihood loss can be formulated as:

$$\mathcal{L}(\Theta;\mathbf{x}_i,\mathbf{y}_i) = -\frac{1}{N}\sum_{i=1}^{N} log(p(\mathbf{y}_i|\mathbf{x}_i;\Theta)). \tag{6}$$

It is also know as the cross entropy loss in our binary segmentation setting.

2.2   Two channel 2D CNN

We find that patch-based 3D CNN segmentation results are usually discrete and have some false negative in term of tumor segmentation. As such, we trained a two channel 2D CNN $\mathbb{T}$ with CT image and kidney (including tumor) mask as two channel input and tumor as output. For this model, the training data was mixed by all slices that contain tumor and the same number of randomly sampled slices that didn't contain tumor.

2.3   Automatic abnormal correction

In the kidney segmentation stage, we also design a abnormal detection and correction method based on prior knowledge. The abnormal situation can be divided into two sides.

2.4   Residual tumor lesion out of ROI

To make the overall results robust, we pad a frame with size 30 around the $\mathbf{S}^{\mathbb{F}}$ and decided a bounding box for 3D CNN to detect tumor. However, there are a few tumor lesions detected out of $\mathbf{S}^{\mathbb{F}}$. We decide a criteria that keep the lesions if they are connected to $\mathbf{S}^{\mathbb{F}}$ and remove it otherwise.

2.5   No tummor detected in 3D CNN detection results

We have prior knowledge that one or more kidney tumors are existing for each subject [8]. But there are a few dataset that are judged as non-tumor (or just small number of pixels) by our 3D CNN, such as the testing cases 214, 220, 234 and 254,. In such situation, the two channel 2D CNN are used through whole tumors region to reduce false negative rate. Please be noted, the two channel 2D CNN are used within detected tumor region by 3D CNN to reduce false positive rate.

References

1. Zhu Z, et al.: A 3D coarse-to-fine framework for volumetric medical image segmentation. In: 2018 International Conference on 3D Vision (3DV), pp.682–690. IEEE,(2018)
2. Kamnitsas K, et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis, **36**, 61–78 (2017)
3. Wu, J., et al.: A multi-atlas guided 3D fully convolutional network for MRI-based subcortical segmentation. In: International Symposium on Biomedical Imaging (ISBI). pp. 705–708. IEEE (2018)
4. Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention.* Springer, 2015, pp. 234–241.
5. Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, pp. e453, 2014.
6. Benjamin Cheong, Raja Muthupillai, Mario F Rubin, and Scott D Flamm, "Normal values for renal length and volume as measured by magnetic resonance imaging," *Clinical journal of the American Society of Nephrology*, vol. 2, no. 1, pp. 38–45, 2007.
7. Yue Zhang, Jiong Wu, Chen Wanli, Chen Yifan, and Xiaoying Tang, "Prostate segmentation using z-net," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 11–14.
8. Nicholas Heller, Niranjan Sathianathen, Arveen Kalapara, Edward Walczak, Keenan Moore, Heather Kaluzniak, Joel Rosenberg, Paul Blake, Zachary Rengel, Makinna Oestreich, 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.