2019 Kidney Tumor Segmentation Challenge Method Manuscript

MengLei Jiao, Hong Liu

Beijing Key Laboratory of Mobile Computing and Pervasive Device Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Abstract. This paper framework in detail for KiTS19, which is the 2019 Kidney Tumor Segmentation Challenge. We adopt two model ResUNetSM and DeepLabV3 plus to segment kidney and tumor respectively. Firstly, we propose a model ResUNetSM to segment kidney, which uses ResNet for encoder, and adopts SELayer and MobileBlock for decoder. ResUNetSM also adopts ASPP and skip-connect structure. To segment tumor region, we adopt DeepLabV3 plus and segment tumor in the 3D ROI region from above kidney segmentation results to reduce noise. Finally, we use 3DCRF and 3D connected component analysis as post-processing to improve the final segmentation results. Our framework gets the 96.31% mean dice for kidney and 81.64% mean dice for tumor on validation set.

Keywords: CT · semantic segmentation · lession · deep learning · CNN.

1 Introduction

The goal of KiTS19 is to accelerate the development of reliable kidney and kidney tumor semantic segmentation method. KiTS challenge provide the ground truth of semantic segmentation for arterial phase abdominal CT scans from 300 unique kidney cancer patients. Among these data, 210 patient data is released for model training and validation, and the remaining 90 is used for objective model evaluation.

Tumor segmentation is a very challenging problem due to significant variations in location, size, shape, intensity, texture, and the number of occurrences of tumor across different patients. To tackle these difficulties, many segmentation methods have been proposed, including intensity thresholding, region growing, and deformable models. Recently, fully convolutional neural networks (FCNs) have achieved great success. 2D FCNs include UNet architecture [1], the multi-channel FCN [2], and the FCN based on VGG-16 [3]. 3D FCNs replace 2D convolutions by 3D convolutions with volumetric data input [4]. There is also 2.5D structure, which use only a few adjacent slices [5]. The combination of 2D FCNs and 3D FCNs also achieved good results [6].

This paper focuses on kidney and tumor segmentation.Figure 1 shows our proposed framework, which uses two models. Firstly, we propose ResUNetSM for kidney segmentation which combining advantage of ResNet[7], DeepLabV3

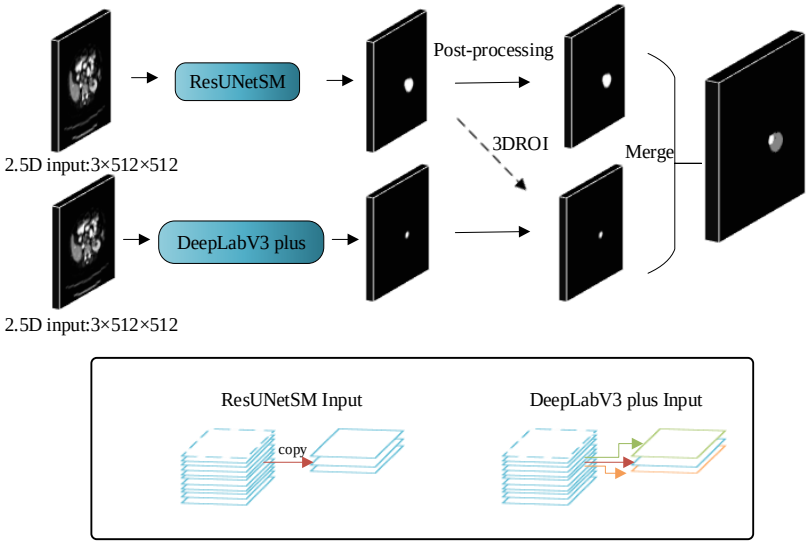

Fig. 1. proposed framework for kidney segmentation and tumor segmentation.

plus[8], MobileNetv2[9] and SENet[10], The ResUNetSM for kidney segmentation is 2D. Secondly, we only use DeepLabV3 plus for tumor segmentation. The DeepLabV3 plus for tumor segmentation is 2.5D, which uses several adjacent axial slices as input to the model. These two models are independent when training. During inference phase, we first use ResUNetSM for kidney segmentation and then appropriately expand the 3D region. Then, we only use CT image in the 3D ROI for tumor segmentation. Thirdly, we use dense 3DCRF and 3D connected component analysis as post processing to improve the tumor segmentation result. Finally, we combine the result of kidney segmentation and tumor segmentation, in figure.1

2 Dataset and preprocessing

There are 300 CT scans in KiTS19[11], and 210 scans are released for training and 90 scans for test. In this paper, we divide the 210 CT scans into train set and validation set, There are 195 CT scans for training and 15 CT scans for validation. We use the raw CT scans in this work. For all CT scans, the slice thickness ranges from 1mm to 5mm, and the size is 512x512 pixels. But the number of slices in each scan differs greatly and varies between 29 and 1059. There are 15856 slices contains kidney, 5696 slices contains tumor and 29068 slices contains other tissues.

To make the original CT image clearer, we truncated the image Hounsfield values of all scans to the range of [-512,512] to ignore irrelevant image details.

3 Method

We propose a framework for kidney and tumor segmentation. We use proposed ResUNetSM model for kidney segmentation and DeeplabV3 plus model for tumor segmentation. In our training stage, kidney segmentation training and tumor segmentation training are independent. After finishing the training of two model, we first use ResUNetSM to get the result of kidney segmentation which may contains tumor regions. Then we get a 3D bounding box which contains kidney, but we can't guarantee that all kidney and tumor in this bounding box, so we expand it. After that, we only use CT slices which are contained in the bounding box for tumor segmentation by DeepLabV3 plus model. Finally, we use 3DCRF and 3D connected component analysis method to optimize the result of tumor segmentation and merge kidney and tumor segmentation.

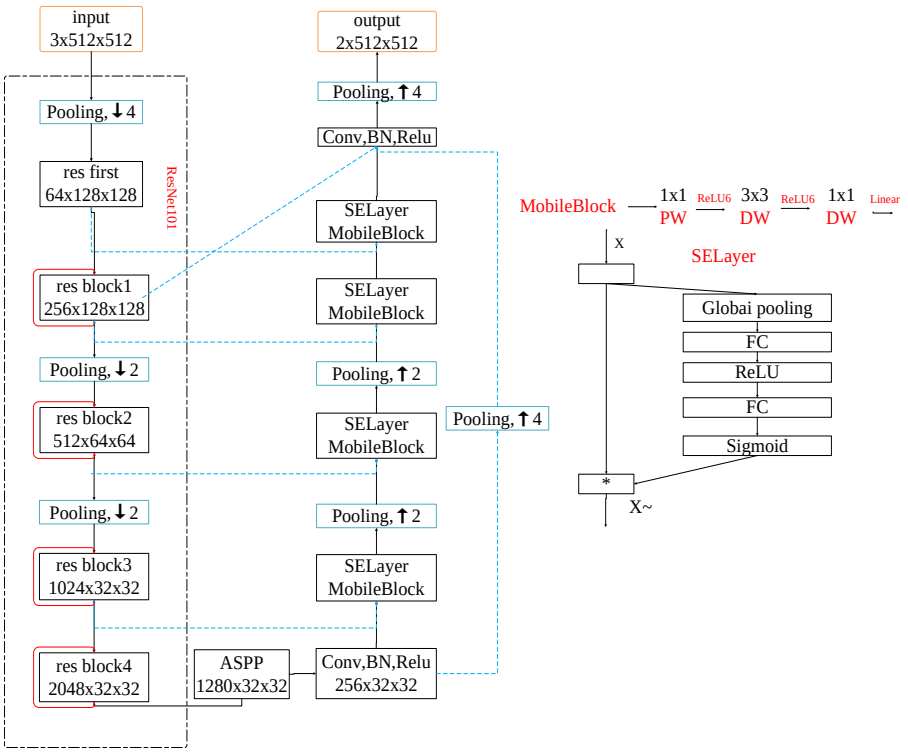

Fig. 2. ResUNetSM model architecture for kidney segmentation. The shape of input is 3x512x512: repeat stacking one slice to expand the channel. The red lines indicate the short-range residue connections and the blue and green lines indicate the long range concatenation connections

3.1 ResUNetSM for Kidney Segmentation

The proposed ResUNetSM model is shown as Fig. 2, which consists of encoder and decoder parts. For encoder part, it is based on ResNet101 and consists of one enter layer and four blocks. Each block has a short-range residual connection. For decoder part, we combine the SELayer and MobileBlock in each block as Fig. 2 shows. The SELayer[10] can model the interdependencies between feature channels, which is a substructure and can be embedded in other structures. The MobileBlock[9] is a lightweight mobile network structure based on inverted residual structure, which can significantly reduce model parameters and remain similar accuracy. After encoding, we add the ASPP layer to get multi-scale information. In addition, some feature information may be lost during the decoding process. So we add long range concatenation connection between encoder and decoder, which connects the first block feature from encoding and the last block feature from decoding. This model can combine information from various scales.

3.2 DeepLabV3 plus for tumor Segmentation

The DeeplabV3 plus is used for tumor segmentation, which is also ResNet101. During training, we set the different learning rate for different resolution.

3.3 Loss function and Optimizer

We adapted the dice loss in our method. which is usually used for natural image and medical image segmentation[12]. The function of dice loss is:

$$loss = \frac{2 \times pred \times target}{pred + target}$$

We adapted Adam optimizer for kidney segmentation and Stochatic Gradient Decent(SGD) optimizer for tumor segmentation. Adam optimizer converges quickly and SGD optimizer is very stable.

3.4 Implementation details

In this section, we will introduce our method in detail. During training, the ResUNetSM and DeepLabV3 plus are independent with the same input resolution 3x512x512. Because the raw slice is 512x512 pixels, so we change the input of ResUNetSM to 3x512x512 by repeating three times of the same slice. And we change the input of DeepLabV3 plus to 3x512x512 by concatenating three neighboring slices, and predicting the middle one. During inference, we first used ResUNetSM to get the result of kidney segmentation, then find the first slice and last slice with kidney on z-axis, find the first column and last column with kidney on x-axis, find the first row and last row with kidney on y-axis. Then we use the boundary information to get the 3D bounding box which only contains kidney in this CT sequence. For kidney tumor should be located in kidney regions. While we can't make sure that 3D bounding box is complete, so we expand

the 3D bounding box on x-axis, y-axis and z-axis. We send the processed CT image to the DeepLabV3 plus model, and get the result of tumor segmentation. Then, we perform tumor segmentation in this with 3D bounding box. Because tumor region is small, the result may coarse and have much noise, so we further use 3DCRF[13] and 3D connected component analysis as the post-processing to improve segmentation result. Finally, we combine the result of kidney and tumor segmentation. The pipeline of inference phase is in Fig. 1.

3.5 Experiment details

Our models are implemented using the public popular frame PyTorch. We adopt ResNet101 pre-train model on ImageNet. Both models use Dice loss as loss function. During training, ResUNetSM uses adam optimizer, and DeepLabV3 plus uses SGD optimizer. Each model is trained for 100 epochs. The initial learning rate is 0.002 for ResUNetSM model and 0.007 for DeepLabV3 plus model. The learning rate declines according to the following formula:

$$lr = base\_lr * (1 - \frac{epoch}{100})^{0.9}$$

Because the tumor sample is small, we adopt oversampling method when training DeepLabV3 plus. For every sample, we give weight ratio of kidney and tumor is $0.3 : 0.7$ with 30% probability to get the kidney sample and 70% probability to get the tumor sample.

Training each model took about two days using two NVIDIA 1080Ti GPU with 12GB memory. Applying the model take about 0.02 second to generate the segmentation result for each slice. The total processing time for final segmentation thus depends on the image resolution and the number of slices for each scan, which is ranged from 7 seconds to 211 seconds for the KiTS19 test data.

We use 3DCRF for post-processing, the parameter of 3DCRF is very important. It has 9 parameters to optimize. The parameters comes from [1]. and can be optimized by BOBYQA algorithm. A set of parameters may perform well in one case but poorly in another. So we manually found some parameters in our experiment which can improve the result of each case in validation set as shown in Table. 1.

Table 1. 3dcrf parameters

name	value
pos_x_std	3
pos_y_std	3
pos_z_std	3
bilateral_x_std	3
bilateral_y_std	3
bilateral_z_std	3
bilateral_intensity_std	3
pos_w	3
bilateral_w	1

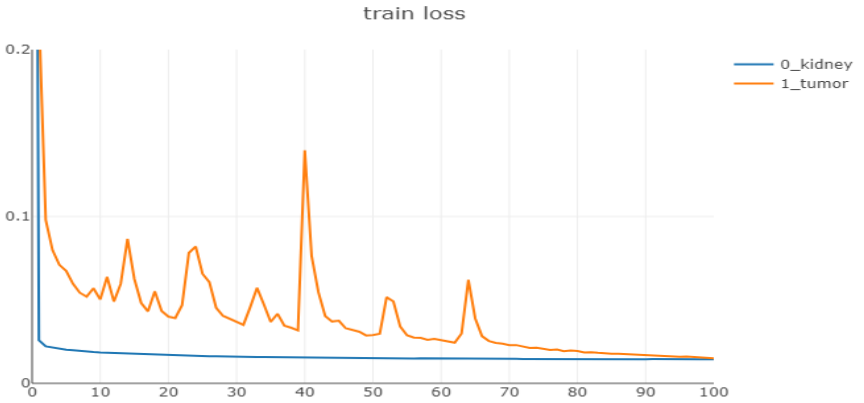

Fig. 3. Training loss lines with the blue one as kidney loss and yellow one as tumor loss.

4 Result on validation set

Table 2. Result on validation set

method	kidney dice	tumor dice	kidney-tumor dice
ResUNetSM_DeepLabV3+	0.9392	0.7424	0.8408
ResUNetSM_DeepLabV3+_3DCRF	0.9412	0.7542	0.8477
ResUNetSM_DeepLabV3+_3DCCA	0.9630	0.8134	0.8882
ResUNetSM_DeepLabV3+_3DCCA_3DCRF	0.9631	0.8164	0.8897

The loss lines during training are shown as Fig. 3. During inference stage, when use 3DCRF and 3D connect component analysis(3DCCA), the result on validation set is shown as Table. 2. The result shows the base dice value for kidney is 0.9392, for tumor is 0.7424. The result is improved to 0.9412 and 0.7542 after 3DCRF. The result is improved to 0.9630 and 0.8134 after 3D connect component analysis. On the basis of the previous, the result is improved to 0.9631 and 0.8164 after 3DCRF. So, the post-processing is also important for medical image segmentation. It can remove a lot of misidentification, and improve recognition accuracy. However, this is just the result on the validation set, there may be some bias. We still can't know the results on the test set before the leaderboard is announced.

References

1. Patrick Ferdinand Christ, Mohamed Ezzeldin A. Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D'Anastasi, Wieland H. Sommer, Seyed-Ahmad Ahmadi, and Bjoern H. Menze. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3d conditional random fields. *CoRR*, abs/1610.02177, 2016.
2. Changjian Sun, Shuxu Guo, Huimao Zhang, Jing Li, Meimei Chen, Shuzhi Ma, Lanyi Jin, Xiaoming Liu, Xueyan Li, and Xiaohua Qian. Automatic segmentation of liver tumors from multiphase contrast-enhanced ct images based on fcns. *Artificial intelligence in medicine*, 83:58–66, 2017.
3. Avi Ben-Cohen, Idit Diamant, Eyal Klang, Michal Amitai, and Hayit Greenspan. Fully convolutional network for liver segmentation and lesions detection. In *Deep learning and data labeling for medical applications*, pages 77–85. Springer, 2016.
4. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *CoRR*, abs/1606.04797, 2016.
5. Christoph Angermann, Markus Haltmeier, Ruth Steiger, Sergiy Pereverzyev Jr., and Elke Ruth Gizewski. Projection-based 2.5d u-net architecture for fast volumetric segmentation. *CoRR*, abs/1902.00347, 2019.
6. Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018.
7. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
8. Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611, 2018.
9. Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381, 2018.
10. Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. *CoRR*, abs/1709.01507, 2017.

11. Nicholas Heller, Niranjan Sathianathen, Arveen Kalapara, Edward Walczak, Keenan Moore, Heather Kaluzniak, Joel Rosenberg, Paul Blake, Zachary Rengel, Makinna Oestreich, Joshua Dean, Michael Tradewell, Aneri Shah, Resha Tejpaul, Zachary Edgerton, Matthew Peterson, Shaneabbas Raza, Subodh Regmi, Nikolaos Papanikolopoulos, and Christopher Weight. The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes, 2019.
12. Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sébastien Ourselin, and M. Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *CoRR*, abs/1707.03237, 2017.
13. Patrick Ferdinand Christ, Mohamed Ezzeldin A Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D'Anastasi, et al. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 415–423. Springer, 2016.