Renal Tumor Segmentation in CT using Cascade U-Net with 2.5D approach

XueJian HE[1], Ping Shun Leung[2], Lu Wang
[1] jimhe@astri.org, [2] simonleung@astri.org,
The Hong Kong Applied Science and Technology Research Institute Company Limited

Abstract. We propose Cascade U-Net with 2.5D approach to segment kidney and tumor from 3D CT image. We use standard U-Net to generate segmentations per each volume slide (2D image). 4 prediction volumes are generated per different magnification and slice direction. Then, consolidate the volumes to formulate the final prediction volume. Per experiment on the KiTS19 dataset, we get a 12% raise in dice coefficient when compare with single U-Net prediction.

1 INTRODUCTION

There are more than 400,000 new cases of kidney cancer each year. Surgery is the most common treatment. Semantic segmentation is a promising tool for advanced surgical planning and showing how tumor morphology relates to surgical outcomes. This paper utilizes multiple standard U-Net to segment kidney and kidney tumor from CT 3-D images. Then, consolidate the network outcomes to refine the segmentation results. Finally, submit the results to the KiTS19 challenge.

2 METHODOLOGY

2.1 Work flow

U-Net is the commonly used CNN in the medical image segmentation task. The network consists of encoder and decoder part. Per KiTs19 challenge, the segmentation task includes 3 categories: background, kidney, tumor. We use two U-Net to predict from inferior to superior (z-direction) and from posterior to anterior (y-direction) respectively. Then, remove false positive and false negative with reference to the common part and different part of the 2 results. Based on the more accurate segmentation, we extract small regions-of-interest (ROI) to refine the tumor segmentation via the same approach.

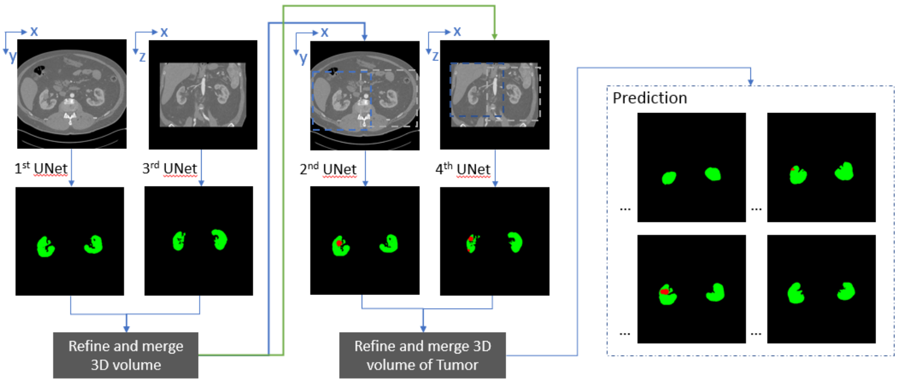

Fig. 1. Work flow of Cascade U-Net

2.2 Data preparation

First, clip the volume data per the Hounsfield unit values range [-512, 512]. Values outside this range are irrelevant to the kidney/tumor information, thus noise. The volume is sliced into 2D images for training. The 2D images fall into 4 categories:

A. empty (no segmentation)
B. kidney only
C. kidney and tumor
D. tumor only

Consider class balance issue, the images are selected for training. Almost all B, C, D class images are selected. B, C class quantity are comparable. D class quantity are the least. Only select limited A class images to avoid biases during training.

Additional rescale is done for generating cross section images along y-direction. The slice thicknesses (along z-direction) range from 1mm to 5mm. The organs in cross section image show a large shape difference. The scale of cross section images is normalized per the slice thickness.

2.3 Network Architecture

Standard 2D U-Net is used. Input dimension is (256, 256). The output dimension is (256, 256, 3) which consist of background (channel 0), kidney (channel 1) and tumor (channel 2).

2.4 Cascade U-Net

It involves 2 steps.

First, look up the kidney and tumor area based on the whole cross-section area. The location and surrounding provide important information if there exist kidney/tumor and where the kidney locates. This step focus on the finding the correct kidney/tumor location.

Second, extract ROIs based on the kidney/tumor segmentation. This step focus on refining the boundary of tumor segmentation.

2.5 Weighted categorical cross entropy loss

The U-Net output 3 classes: background, kidney, tumor. Categorical cross entropy loss is used. To address the class un-balance problem, we add weight parameters to increase the loss caused by tumor.

Among the pixels of training images, the average percentage of background is 79.69~97.89%. The average percentage of kidney is 1.66~14.49%. The average percentage of tumor is 0.45~5.82%. The ratio of background: kidney: tumor ~= 28.3: 2.6: 1

The weight ratio of the cross-entropy loss is background: kidney: tumor = 1: 3: 6

2.6 Post-processing

There may exist false positive and/or false negative on each prediction. The predictions are stacked up the prediction plane to form volume. A coarse shape is observed when view the volume in orthogonal direction. The shape may even look stripped in some cases where the U-Net prediction is not accurate enough. By “referring to” the same prediction slice generated by the orthogonal direction, the coarse/stripped area are smoothen/reclaimed. In results, the shape of prediction looks better. Also, much false positive and false negative are removed. The accuracy can be improved.

First, intersect the 2 volumes. We get a volume which is common to prediction along z-, y-directions. Those volumes are considered as the segmentation with the highest confidence. We call it common volume. Then, rebuild the segmentation volume by the following steps:

For each slice prediction,

A. If the prediction area intersects the common volume, union the prediction area with common volume
B. Otherwise, drop the prediction area

3 Experiments

We validate the work flow design using U-Nets with different approach. Benchmark the results by dice coefficient, precision and recall. The CNN networks are implemented using Keras framework with Tensorflow backend.

Table 1. Performance on KiTS19 Test dataset with 10 test volume images.

U-Net descriptions:

a) U-Net1 denote step 1, along z-direction
b) U-Net3 denote step 1, along y-direction
c) U-Net2 denote step 2, along z-direction
d) U-Net4 denote step 2, along y-direction

Approach	Dice (%)	Precision (%)	Recall (%)
U-Net 1	0.6724	0.8746	0.7897
U-Net 1+2	0.7201	**0.8983**	0.8032
U-Net 1+3+2	0.7372	0.8839	0.8498
U-Net 1+3+2+4	**0.7924**	0.8078	**0.8841**

4 Results

Table 1 shows that Cascade U-Net with 2.5D approach yield the highest dice coefficient and recall but the lowest precision among the 4 approach. That imply the 4 U-Net approach output more false positive than the other 3 approach. Consolidation from 4 U-Net accumulates more false positive. In practice, doctors prefer false positive than false negative as reflected in the dice coefficient calculation.

5 Discussion

The slice thickness of CT images is not uniform among cases. Cross-section images along y-direction is lower quality than cross-section images along z-direction. We use prediction along y-direction as complement to prediction along z-direction.

Tumor detection is a difficult task. It can be very small. In the low-resolution CT images, the tumor can look very similar to kidneys or other organs. For cases with small tumor, a small shift on the tumor prediction can lower the dice coefficient a lot. However, current tumor refinement is not good enough. For some cases, the tumor boundary is clear enough as check visually. However, refined tumor prediction only covers partial area. It worth further investigate how to further improve the tumor boundary refinement.

We also try other form of U-Net such as Attention U-Net. However, the results do not show observable improvement. Those results are not shown in this manuscript for brevity. It also worth further study for the reason and study the latest U-Net design such as 3D U-Net.

6 Conclusion

Per our experiment, the cascade U-Net approach can reduce false positive. The 2.5D approach can reduce false negative but also increase false positive. That can be justified since we merge 2 prediction data into 1. The error of prediction is accumulative. By integrating the 2 approach, we can reduce the false negative area while maintain a reasonable false positive area. That should be an important consideration of this medical image segmentation task.

References

1. Ozan Oktay et al, "Attention U-Net: Learning Where to Look for the Pancreas", arXiv:1804.03999v3 [cs.CV] 20 May 2018
2. Nabila Abraham, Naimul Mefraz Khan, "A NOVEL FOCAL TVERSKY LOSS FUNCTION WITH IMPROVED ATTENTION U-NET FOR LESION SEGMENTATION", arXiv:1810.07842 [cs.CV] 18 Oct 2018
3. Patrick Ferdinand Christ et al., "Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields", arXiv:1610.02177v1 [cs.CV] 7 Oct 2016