

Paper Reading No.10

A journey to MICCAI KiTS19 challenge (Introduction of our
proposed solution and top5 methods)

Sheng Lian

July 2019

1 To start

Recently, our team participated in the 2019 MICCAI Kidney Tumor Segmentation Challenge (KiTS19 challenge). The detailed introduction of KiTS19 can be referred to [KiTS19 Challenge Homepage](https://kits19.grand-challenge.org/)¹ and [KiTS GitHub page](https://github.com/neheller/kits19)².

We proposed a cascade SE-ResNeXT U-Net for solving the problem. In July 15, the testing imaging has just released. We submitted our prediction results and the manuscript then. And today, we got the [leader board](#). We ranked #22 in 106 valid teams.

It's not a very good ranking, but we have harvest a lot. So in this time's PPR, I think it is necessary to record it. In the following, I will briefly introduce the methods we proposed and the top five methods on the list.

¹<https://kits19.grand-challenge.org/>

²<https://github.com/neheller/kits19>

2 KiTS19’s brief intro

The KiTS19 Challenge contains 300 Kidney Tumor cases with clinical context, CT semantic segmentations, and surgical outcomes. Among them, 210 (70%) cases were selected at random as the training set, and 90 (30%) cases were selected as testing set. The detailed introduction of KiTS19 dataset can be referred to KiTS19 Challenge Homepage ³ and KiTS GitHub page ⁴.

The Srensen-Dice similarity coefficient (DSC) was used as a statistical validation metric to evaluate the performance of both the *kidney+tumor* segmentation results and *tumor* segmentation results. The DSC can be formulated as

$$DSC = \frac{2TP}{2TP + FP + FN}, \quad (1)$$

In testing stage, we have 90 cases. The overall ranking score can be calculated as

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

Where k representing kidney or tumor, while t representing tumor only.

3 Our proposed method

Our solution towards the KiTS19 is summarized in [this manuscript](#). It should be noted that this is not a formal paper, but a rough introduction.

³<https://kits19.grand-challenge.org/>

⁴<https://github.com/neheller/kits19>

For the KiTS19 challenge, we proposed **SE-ResNeXT U-Net (SERU)** model, which combines the advantages of SE-Net, ResNeXT and U-Net. The detailed framework of the proposed SERU model is displayed in Fig 1.

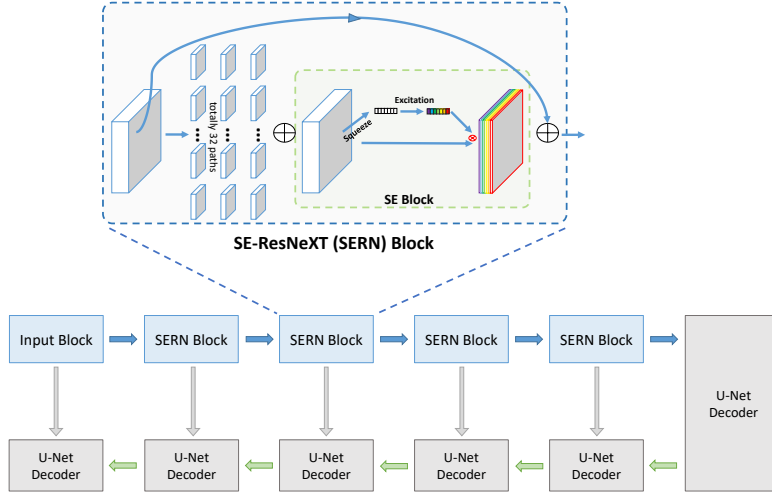


Figure 1: The framework of the proposed SE-ResNeXT U-Net (SERU) model.

For utilizing context information and key slices' information, we implement our model in a coarse-to-fine manner. In this way, we can first obtain coarse segmentation results and then utilize the information given in the coarse results to generate key patches for refine. The overall implementation pipeline of our proposed SERU model in cascaded manner is displayed in Fig 2.

After coarse segmentation results generation, we regard the predicted kidney and tumor region as one meta class, and find out max contour of left kidney and right kidney respectively. Then we find the center point of the max contour on both sides, and cut 256*256 rectangles in left and right kidney regions for refine training. We adopt the left and right 256*256 refine patches generated by last step for refine training, and the same SERU model

with different input scale. For utilizing the advantages of our coarse and refine results, we fuse them by applying different weights, which goes as

$$predictions = \alpha * coarse + (1 - \alpha) * refined \quad (3)$$

Here, α is the ratio weighting the coarse results and refined results when fusing. We find that when α is 0.4, we can obtain best results.

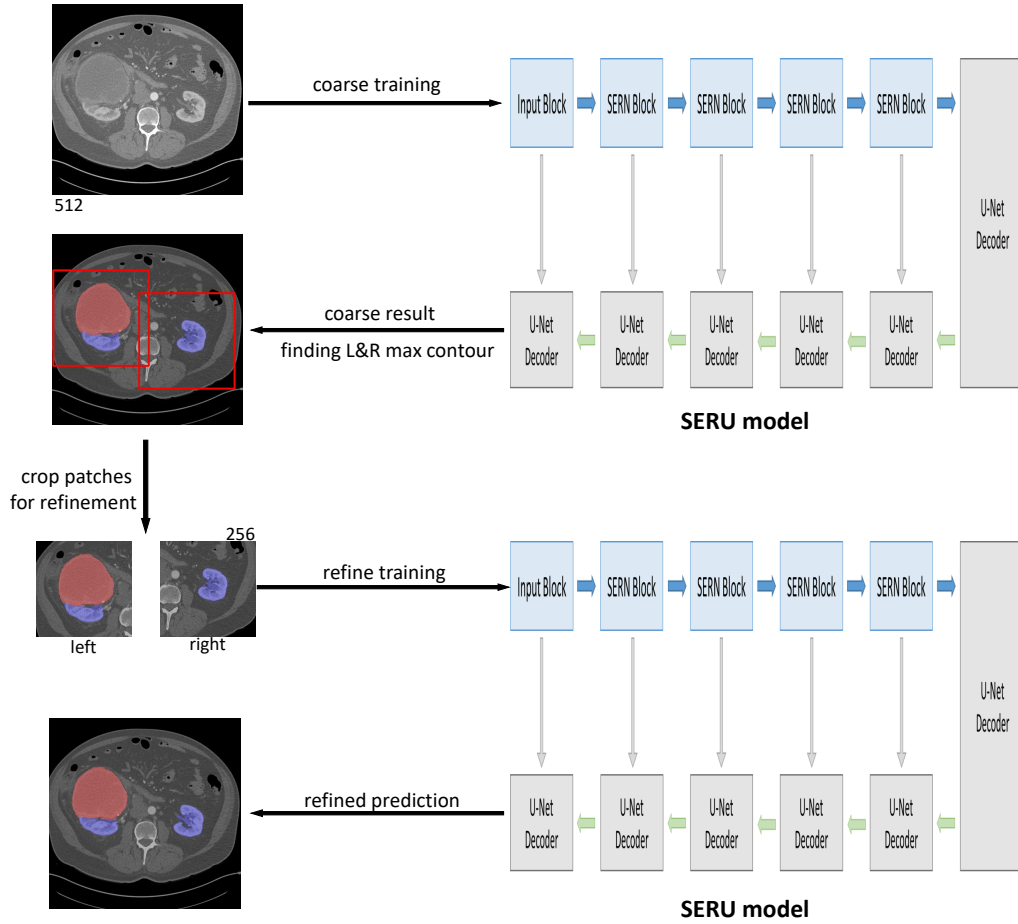


Figure 2: The overall implementation pipeline of our proposed SERU model in cascaded manner.

We validate our model on validation set, and achieve kidney+tumor DSC of 96.78%, and tumor DSC of 74.32%. The overall ranking score on our validate set is 85.55% While in the test set, we got the results of 96.91%, 73.74% and 85.33%.

4 No.1 solution

The No.1 solution comes from German Cancer Research Center. Their manuscript can be viewed through [THIS LINK](#).

Here, authors adopt 3D U-Net as backbone. They also add residual blocks and pre-activation residual blocks to their model. Nothing special, but it works well.

It is worth noting that they did some dataset modifications. They manually inspected the cases with the worst tumor dice and excluded some cases, and replaced some annotation. Their loss function is the sum of dice loss and cross entropy loss.

They achieved kidney+tumor DSC of 97.37%, and tumor DSC of 85.09%. The overall ranking score is 91.23%

5 No.2 solution

The No.2 solution comes from Pingan technology, which is very similar to ours. The manuscript can be viewed [HERE](#). The overall architecture is depicted in Fig 3.

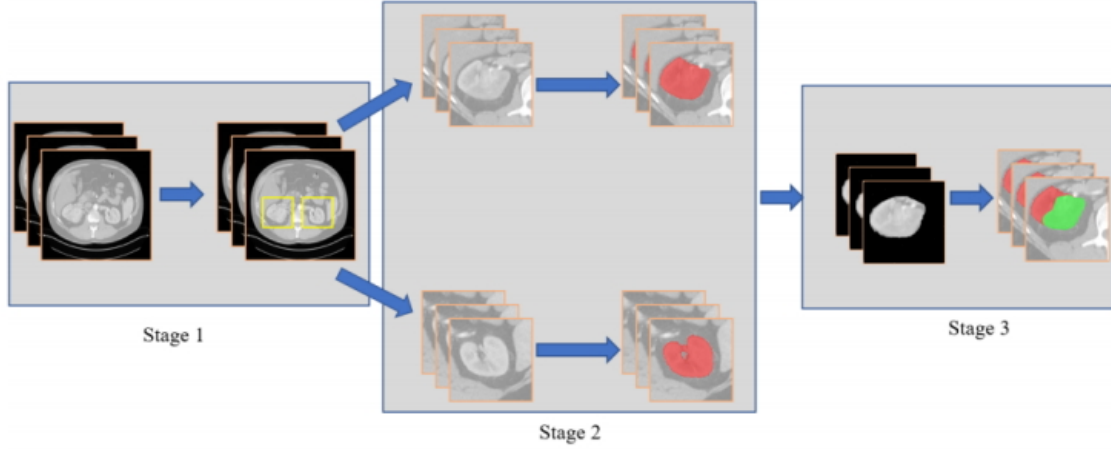


Figure 3: The architecture of No.2 solution.

This is more like a 3D version of our model, and the architecture is obviously still 3D U-Net. Their loss function is a combination of dice loss and cross entropy loss.

It is worth noting that authors mentioned they "*finally adopt some post-processing method to fill the holes inside the tumor and remove some false positives.*" But the post-processing steps are not described in the manuscript.

They achieved kidney+tumor DSC of 96.74%, and tumor DSC of 84.54%. The overall ranking score is 90.64%.

6 No.3 solution

The No.3 solution (see the manuscript [HERE](#)) is also similar to No.2 solution, whose architecture goes as Fig 4, a cascaded model. In this approach, authors adopt VB-Net. It is a modified version of V-Net, with bottleneck structure. The loss function is a generalized dice loss function that focuses only on foreground voxels disregarding how many the background voxels in the whole image, which goes as below:

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

They achieved kidney+tumor DSC of 97.29%, and tumor DSC of 83.21%.
The overall ranking score is 90.25%.

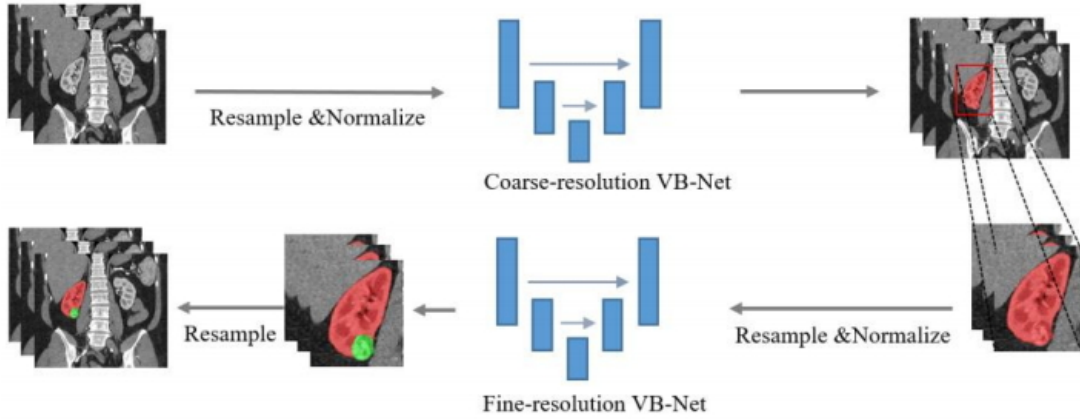


Figure 4: The architecture of No.3 solution.

7 No.4 solution

The manuscript of this solution can be viewed [HERE](#). It's another 3D UNet + two stage solution. They utilize nnUNet for backbone network training. For post-processing, they utilize the prior knowledge that no more than two kidneys exists in the abdomen, the segmentation results are followed by a connected component analysis to remove false positives.

They achieved kidney+tumor DSC of 97.42%, and tumor DSC of 83.06%.
The overall ranking score is 90.24%.

8 No.5 solution

The manuscript of this solution can be viewed [HERE](#).

This solution has some highlights. Although their backbone model is also 3D U-Net, they train a biased U-Net and compact loss for tumor segmentation. As we know, the main challenge for us is class imbalance. The segmentation of tumor is significantly worse than the accuracy of kidney. So authors train a U-Net to only segment kidney tumor, termed as biased U-Net.

Most kidney lesions have a ball shape in 3D or a nearly circle shape in 2D in the training sets. The commonly used loss functions (e.g., cross entropy, Dice loss) do not have shape constrain. Thus, authors propose the compact loss function to impose explicitly compact shape regularization. The shape compactness is defined as

$$C = \frac{P^2}{A}$$

where C is the shape compactness, P and A are the shape perimeter and area, respectively. When using g, s denote gound truth and predicted segmentation, the compact loss is defined as

$$L_C = \frac{(Length of s)^2}{Area of s} \approx \frac{(\int_{\Omega} \nabla s dx)^2}{\int_{\Omega} s dx} \quad (4)$$

To avoid the training stuck in trivial local minima, in practice, authors combine the proposed compact loss function with region-based active contour loss [?].

$$L_{AC} = Region_{in} + Region_{out} = \int_{\Omega} [s(1 - g)^2 + (1 - s)g^2] dx$$

Note that the compact loss is used to finetune and refine the segmentation generated by commonly used loss functions.

They achieved kidney+tumor DSC of 97.34%, and tumor DSC of 82.54%. The overall ranking score is 89.94%.

9 My thoughts

- Nearly all the top solutions utilized 3D-UNet based model. Maybe the main reason for our low metric value is that we didn't use 3D-based models (we have considered this solution, but it's so time consuming).
- Homogenization of model and implementation tends to be serious. e.g. many solutions choose five-fold cross validation. This is somewhat regrettable, but it illustrates the effectiveness of these implementations. We will consider it in future research.
- nnU-Net is mentioned many times, and is a powerful implementation of U-Net, according to the authors. But last month when I read the nnU-Net paper (see [HERE](#)), I didn't get the strong point of nnU-Net. So maybe I need to revisit that paper and look for a chance to try nnU-Net.
- Class imbalance is still a challenging issue for the task. Maybe a boundary based loss is worth trying.
- Post-processing steps according to the morphological prior knowledge, worth of research.