KITS19: Kidney and kidney tumor segmentation in CT scans using a 3D U-Net based network with additional tasks

Gabriel E. Humpire-Mamani

[1] Radboud University Medical Center, Nijmegen, The Netherlands
g.humpiremamani@radboudumc.nl

Abstract. We propose a single network to segment kidneys and kidney tumors. We enforce the segmentation task by adding an additional task to the network. The network classifies whether a patch contains a kidney tumor or not. This step, helps to improve the confidence of the segmentation network. No additional annotations are needed for this task.

Keywords: Segmentation, Kidney, Kidney tumors, CT, Deep learning.

1 Introduction:

Kidney cancer is the 12th most mortal cancer in the world [1] with 14700 estimated deaths for 2019 and approximately 73820 new kidney & renal pelvis cancer cases in 2019 worldwide. In the past 25 years, the trend of deaths caused by kidney and renal pelvis cancer remained steady, although, the number of new cases keeps raising up [1]. In similar challenges (LiTS and medical decathlon) participants obtained high performance using cascade networks [3, 4], which consist of two networks. One for organ localization (rough segmentation) and one for organ and its tumor segmentation (fine segmentation). The preferred network used was 3D U-Net [2].

In this paper, we present a method that replaces cascade networks and uses a single network allowing backpropagation.

2 Automatic segmentation method:

2.1 Input format:

The annotations provided by the KITS19 challenge have three classes: background, kidneys, and kidney tumors. For our own convenience, we define one additional input formats. We join the kidneys and kidney tumors class as a single class (see Figure 1a) named as input format 1. We name the annotations provided by KITS19 as input format 2 (see Figure 1b).

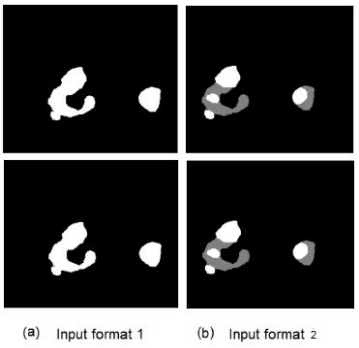

Figure 1: Additional input format used during training

2.2 Pre-processing:

All CT scans and annotations were resampled to 1*1*1mm (for fine segmentation format 3) and 4*4*4mm (for rough segmentation format 1) resolutions. Scans and annotations were resampled using cubic and nearest neighbor interpolation respectively. Hounsfield Units outside of the range [-500,400] were clipped.

2.3 Segmentation network

We propose an end-to-end method for kidney and kidney tumor segmentation. We schematized the proposal in Figure 2. The segmentation task produces two segmentation outputs:

1. 4x4x4 mm resolution using format1 for rough kidney localization.
2. 1x1x1 mm resolution using format 3 for fine segmentation.

Both inputs share the same center of gravity to match up segmentations of different resolutions, i.e., the center of gravity of the 1x1x1mm input is the same center of gravity in the 4x4x4mm input.

The 4x4x4mm input provides a larger context of the CT scan to the network to roughly get the localization of the kidneys. The output of the 4mm network defines regions of interest where the 1mm network will fine segment the kidneys and kidney tumors.

2.3.1 Proposed network: Single cascade network

The disadvantage of cascade networks is that they are not end-to-end solutions, where gradients cannot be backpropagated from the second to the first CNN. In this study, we propose replacing cascade networks by a single network to segment the kidneys and kidney tumors. Our single network is based on the idea of cascade networks (two independent networks) to join two networks to allow backpropagation. Additionally, we define an extra classification task to the segmentation network to enforce the kidney tumor segmentation.

This network takes 3D U-Net as backbone, the full architecture is depicted in Figure 2. The low resolution sub-network is a 3D U-Net (with 16 filters instead of 32) to perform low resolution segmentation by taking input patches of 4*4*4mm resolution as input. This first part of the network segments the kidney and the kidney tumors as a single class (format 1). The high resolution segmentation sub-network uses a 3D U-Net with 32 filters as the original implementation. The output of the low resolution sub-network is upsampled 4 times and padded with zeros to match up and mask out the second input image in 1*1*1mm resolution. The masked out input serves as additional input to the fine segmentation network creating a branch of filters, which will be passed through the skipping connection of the fine segmentation network.

2.3.2 Extra classification task

The extra classification task classifies whether a patch contains a kidney tumor. The feature maps located at the bottom of the 3D U-Net high resolution sub-network (see Figure 2) are taken to extend it using 32 filters of 3x3x3 convolutions, followed by 16 filters of 3x3x3 convolutions. Later, these feature maps are converted to a fully-connected layer of 1024, 256, and 256 neurons. In the meantime, the output of the high resolution segmentation sub-network is taken to extract only the kidney tumor channel, flatten it, and sum up the predictions of the kidney tumor channel (volume of the kidney tumor prediction in the patch). This value is concatenated to the 256 neurons of the classification task. This enforces that the classification task is directly linked to the segmentation task, having now 129 neurons. This branch is followed by 64 and 32 neurons.

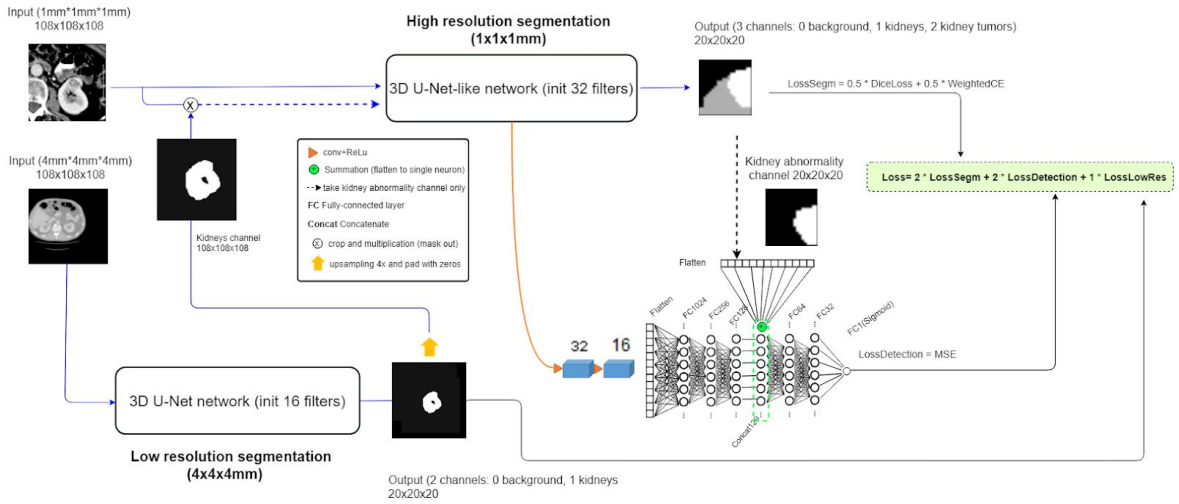

Figure 2. Architecture of the network.

2.3.3 Training

The low resolution sub-network receives 4x4x4mm of 108x108x108 voxels and the high resolution sub-network receives 1x1x1mm of 108x108x108 voxels. The output of the full architecture returns three outputs: low resolution 20x20x20 of 4mm, high resolution 20x20x20 of 1mm, and the classification task. The network is trained using Adam optimization function with learning rate 1e-4. The learning rate is multiplied by 0.9 after every epoch.

During training, the patches have 50% overlapping in all directions. The patches were sampled to balance the number of positive and negative samples. We consider a positive sample, a patch that contains at least a voxel of the kidney or kidney tumor class.

We trained on 80% of the training set and used the remaining 20% for validation. The network stopped training when the performance in the validation set did not improve for 10 epochs.

2.3.4 Loss

The output of the classification task uses Sigmoid as activation function and mean squared error (MSE) as loss. The loss of the high resolution segmentation sub-network uses 0.2 * dice loss + 0.8 weighted-crossentropy. The low resolution segmentation sub-network uses crossentropy as loss. The final loss of the network is defined as:

$$Loss = 2 * LossHighResolution + 2 * LossClassification + 1 * LossLowResolution$$

70% Top-k is applied to the loss to apply online voxel hard mining.

2.4 Data augmentation

To generalize the network and prevent overfitting, we use multiple techniques of data augmentation. Elastic deformation is used for the segmentation task using a grid of 10x10x10 and B-Spline interpolation. 3D scaling ± 0.05, 3D rotation ± 5°, gaussian noise, and HU variation ± 50.

2.5 Post-processing

At inference time, the patches are not overlapped. The predictions are stitched together to compose a full 3D prediction of the CT scan. The output of the high resolution sub-network are post-processed. All the output channels are thresholded at 0.5 to get binary channels. We take the largest component of the left and the right side of the kidney channel; We get the connected-components of the kidney tumor channel and discard the components that are not connected to the kidney regions. This guarantees that the final result will return only kidney tumors that are in the kidney area. Additionally, the output of the low resolution is upsampled to 1x1x1 resolution and dilated (10 iterations) to mask out the final output.

2.6 Hardware and software details

This network was designed using Keras and Tensorflow as backend. A GPU GTX1080ti was used to train the network.

3 Experiments

Empirical experiments showed that higher performance is achieved when the low resolution segmentation sub-network is trained first until reaching 0.90 dice on its task. Once the low resolution sub-network reached this performance, the full architecture is trained. To prevent the low resolution sub-network to get affected by the random initialization of the full architecture, we freeze the low resolution sub-network and left the last three blocks of filters as trainable for fine-tuning. The network was trained for 28 epochs in total

4 Conclusions

In this study, we presented a single network able to segment kidneys and kidney tumors. The low resolution segmentation sub-network provides a large context and defines candidates of regions of interest that may contain the kidneys. This helps the high resolution sub-network to focus only in the kidney area. Additionally, we added an extra task to the network, a classification task that indicate whether the patch contains a kidney tumor.

References

1. Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. "Cancer statistics, 2019." *CA: a cancer journal for clinicians*69.1 (2019): 7-34.
2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G. & Wells, W. (eds.)Med Image Comput Comput Assist Interv, 424–432 (Springer International Publishing, Cham, 2016). URL https://doi.org/10.1007/978-3-319-46723-8_49. DOI 10.1007/978-3-319-46723-8_49.
3. Isensee, F., Petersen, J., Kohl, S. A. A., Jäger, P. F. & Maier-Hein, K. H. nnU-Net: Breaking the spell on successful medical image segmentation. arXiv:1904.08128(2019).http://arxiv.org/abs/1904.08128v1.
4. Chlebus, G.et al. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. Nat Sci Rep, 15497 (2018). DOI 10.1038/s41598-018-33860-7.