

# Two Stream Auto-encoder Decoder Network for Kidney and Tumor Segmentation

Pouria Moradi<sup>1</sup>, Reza Azad<sup>1</sup>, and Maryam Asadi-Aghbolaghi<sup>2</sup> \*

<sup>1</sup> Sharif University of Technology, Tehran, Iran

<sup>2</sup> Institute for Research in Fundamental Sciences (IPM), Tehran, Iran  
pu.moradi, rezazad68@gmail.com  
masadi@ipm.ir

**Abstract.** In this competition, we apply two-streams auto-encoder decoder structure for learning kidney and kidneys tumor segmentation. To do so, first, we extract axial layers of the tissues along with their segmentation mask from the 3D volume. These axial layers are then clipped using Hanford distance between +512 to -512 to eliminate non-object of interest. These axial layers are then normalized to form the 2D grayscale images. For each of these normalized images, we generate kidney and kidney tumor masks to train two-stream deep networks. The two-streams deep model learns kidney and tumor masks separately and they generate final mask by concatenating the generated masks. We utilize BCDU-net (extended version of U-Net model) as a deep auto-encoder decoder model for segmentation. We utilize 70% of the Kits19 as the training set and the rest of data as the validation set. Experimental results demonstrate that the proposed structure achieves state-of-the-art performance in the segmentation of kidney and tumor region.

**Keywords:** deep learning · kidney tumor segmentation · U-net · BDCU-net.

## 1 Introduction

Deep learning-based network have achieved outstanding result in image segmentation like in other fields of research in computer vision, e.g. classification. Medical image segmentation is becoming very important in medical image processing, particularly in modeling patient-specific anatomy. The surgical procedures are highly depends on the medical decision made before the surgery. For instance, in kidney surgical treatments, segmentation of kidney and the clinical tumor size is vital. Fully-connected convolutional neural networks (FCN) were from the first deep models that have been proposed for semantic segmentation. U-Net [1] architecture is one of the most successful networks introduced for medical image segmentation. Many extension of U-Net, like Residual U-Net (RU-Net) [2] have been proposed to improve the performance the original network.

---

\* All three authors contributed equally.

In this work, we utilize BCDU-Net and an extension version of U-Net for segmentation kidney and kidneys tumor. We propose a pre-processing procedure to make the input data ready for the deep networks. To do that, the input 3D volume is projected into different axial layers. These axial layers are passed through the Hanford distance and then normalized. The final 2D images are employed as the input data to the deep models. We train two-stream networks for the segmentation of kidney and kidneys tumor. The output segmentation of all axial layers of one 3D volume are combined again to produce the 3D volume mask. We achieve 97% of accuracy for the kidney segmentation and 50% of accuracy for kidneys tumor segmentation.

## 2 Proposed Approach

General overview of the implemented method for kidney and kidneys tumor segmentation task is depicted in Figure 1. For each input 3D volume, we have two segmentation masks, one for kidney and the other for kidneys tumor. We follow the pre-processing, segmentation and post-processing steps to segment kidney and tumor regions.

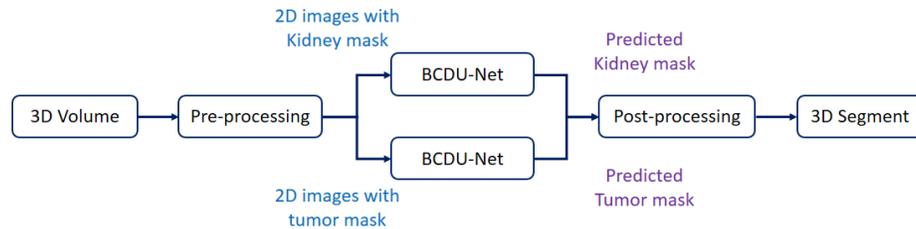


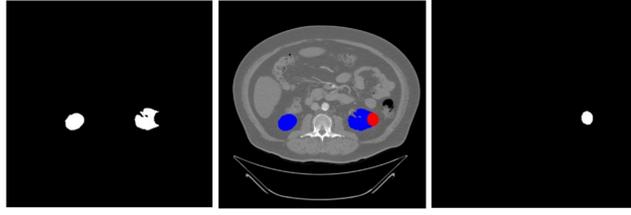
Fig. 1. Two stream network for kidney and tumor segmentation.

### 2.1 Pre-processing

We extract different axial layers of the tissue by projecting the input 3D volume into 2d image plane with different depth. We then need to remove non-object of interest to have only kidney and kidneys tumor. To do that, we employ Hanford distance between range of  $[-512, +512]$  (which is obtained experimentally), after which only kidney and kidneys tumor are remained on images. The 2D images extracted from the previous step are then normalized to produce grayscale images. For each 2D image, we produce two masks of kidney and kidneys tumor. Figure 2 shows an example of the pre-processing step.

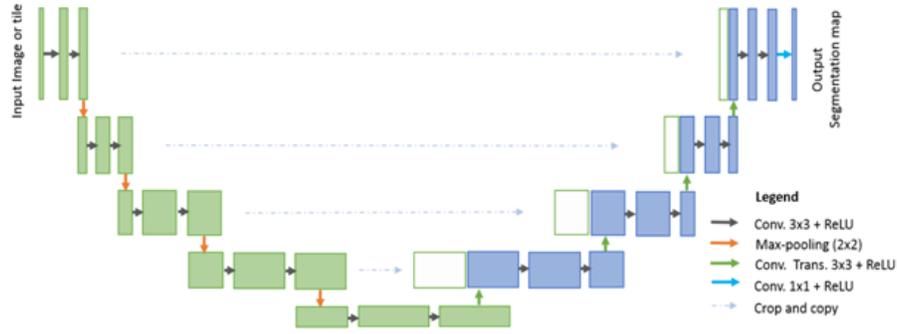
### 2.2 Segmentation

In this step, we utilize BCDU-Net [3] for both kidney and tumor segmentation separately. We create two segmentation masks for each input image, one as the



**Fig. 2.** Sample of 2D image along with kidney and tumor masks. Blue region shows the kidney mask and the red region shows the tumor mask.

kidney segmentation and the other for kidneys tumors. We then train two networks, i.e. BCDU-Net for kidney and kidneys tumor segmentation. The structure of the U-net is shown in Figure 3.



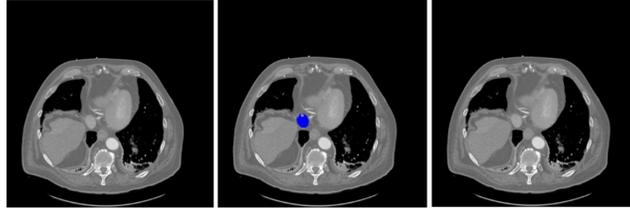
**Fig. 3.** U-net model [1].

In the standard U-Net, the features are simply copied from the encoding path and concatenated with the features from the previous layer in the decoding path in the skip connections. The features from the contracting path have higher resolution while the features from expanding path have higher semantic information. In BCDU-Net, bi-directional convolutional LSTM is utilized in the skip connection layer to combine these two kinds of features in a more complex way to improve the performance of the original U-Net. Furthermore, it applies densely connected convolutions to capture collective knowledge.

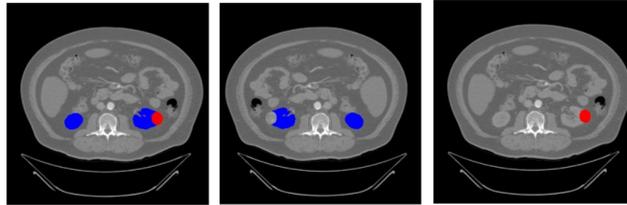
### 2.3 Post-processing

The output of two stream network contains some noisy segmentation region. In this step we apply voxel based connected component algorithm to remove non-object of interest. To do so, we first find connected components with 8 connectivity in voxel level and then we remove those objects that has less connectivity.

The post-processing step is more effective in eliminating noisy detections. Sample effect of post-processing step for removing false detection region is shown in Figure 4.



**Fig. 4.** Effect of post-processing step for removing false segmentation, left (ground truth), middle (predicted mask), right(post-processing).



**Fig. 5.** Sample of segmentation results, left (2D image with segmentation parts) middle(detected kidney regions) right(detected tumor regions).

### 3 Experimental Result

The data consists of 210 input 3D volumes, among which we utilize 150 3D volumes for training and 60 ones for validation. When we project the an input 3D volume into different axial layers, some of these layers contain no segmentation information about the kidney and kidneys tumor, therefore both kidney and kidneys tumor masks have no pixel values more than zero. If we train the network with these input data, it will over fit on these black masks. Therefore, we remove these layers from training phase. In test step, we extract the output segmentation of the network for the axial layers of the input 3D volumes. We then produce a 3D volume output segmentation by combining all the output together to produce the final output segmentation of the input data. We obtained 97% of accuracy for the kidney segmentation and 50% of accuracy for the kidneys tumor segmentation

in the validation set. Sample of segmentation result on 2D image is shown in Figure 5. <sup>3</sup>

## References

1. Ronneberger, Olaf, 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, Cham, 2015.
2. Alom, Md Zahangir, et al. "Recurrent residual convolutional neural network based on u-net (R2U-net) for medical image segmentation." arXiv preprint arXiv:1802.06955 (2018).
3. R. Azad, et al, Bi-Directional ConvLSTM U-Net with Densely Connected Convolutions, submitted to the ICCV Workshop (2019).

---

<sup>3</sup> Code: Code for this completion and also the result of BCDU-net on several medical image segmentation tasks will be released after ICCV workshop review process on the link <https://github.com/rezazad68/LSTM-U-net>.