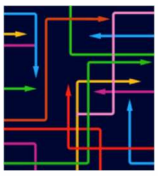
Artificial Intelligence in Oncology

Available online at: www.aioncology.org



Original Article

Cascaded U-net for Kidney and Tumor Segmentation from CT volumes

Asha K Kumaraswamy, Chandrashekar M. Patil*

Vidyavardhaka College of Engineering, Department of Electronics and Communication, Mysuru, 570002, India

ARTICLE INFO

Article history:
Received 04 August 2020
Revised 02 September 2020
Accepted 05 September 2020
Published 09 September 2020

Keywords:
Contrast enhanced computed tomography,
2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS19),
UNet++

ABSTRACT

Contrast-enhanced Computed Tomography (CT) imaging is most useful tool in diagnosing and locating the kidney lesions. An automated kidney and tumor segmentation are very helpful because it can provide the precise information about the location and size of lesions which can be used in quantitative analysis of the tumor. Semantic segmentation of kidney is very challenging as it requires large dataset for training and its morphological heterogeneity makes it a difficult problem. The 2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS19) was a competition held in conjunction with the 2019 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) has publicly released a 210 cross sectional CT images with kidney tumors along with corresponding semantic segmentation masks. In this work we proposed a novel two stage 2D segmentation method to automatically segment kidney and tumor using the combination of UNet++ and squeeze and excite approach. The proposed network is trained in keras framework. Our method achieves a dice score of 0.98 and 0.965 with kidney and tumor respectively on training data and the results demonstrates the accuracy of our proposed method. Proposed method was able to segment kidney and tumor from abdominal CT images which can provide the exact location and size of the tumor. This information can also be used to analyze treatment response.

Introduction

Uncontrolled growth of abnormal cells in the body is termed as Cancer. Cells in nearly any part of the body can become cancer and can spread to other areas. Kidney cancer starts in kidney and can spread to other neighboring organs. It happens when healthy cells in one or both the kidneys grow out of control and form lump called tumor. Renal cell carcinoma (RCC) is the most common type of kidney cancer. About 9 out of 10 kidney cancers are renal cell carcinomas. Renal cell cancer often stays within the kidney, but it may spread to other parts of the body, most often the bones, lungs, or brain. Possible risk factors of kidney cancer are smoking, high blood pressure, obesity.

The American Cancer Society's most recent estimates for kidney cancer in the United States for 2020 are, about 73,750 new cases of kidney cancer (45,520 in men and 28,230 in women) will occur and about 14,830 people (9,860 men and 4,970 women) will die from this disease [1]. Removing all or part of the kidney with surgery is called a nephrectomy. It is the most common treatment for kidney cancer. It has three general options: partial nephrectomy (PN), radical nephrectomy (RN), or ablation. Removing only part of the kidney affected by tumor is PN, which requires earlier tumor detection. PN is typically less invasive and cause limited renal function impairment, thus it is preferred when feasible. Whereas RN refers to removal of complete kidney. In ablation cancerous cell are killed through the application of radio waves, heat or cold. It is used mostly in the people with small tumors. Proper understanding of the location and the extent of the kidney damage can help to diagnose, treat and to find the treatment response [2]. Contrast-enhanced CT scan will help to identify any tumors in the kidneys, provide information about tumor size, shape and position and will also provide information about whether cancer has spread to other organs and tissues. A contrast medium is a special dye that provides better detail on the images taken by CT scans [3]. This dye is injected into a patient's vein. Renal cysts will not take up the contrast medium, but renal tumor will. This helps to differentiate tumors from cysts.

Automatic segmentation and detection of kidney tumors helps radiologist to diagnose a greater number of patients in less time. Semantic

* Corresponding author: Asha K Kumaraswamy
E-mail address: ashakk06@gmail.com

segmentation of kidney is very challenging as it requires large number of datasets for training and lot of efforts for annotation. Goal of the 2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS19) [4] was to correctly identify the kidney and cancerous tissue (tumors). Multiphase CT images from the patients who underwent partial or radical nephrectomy for one or more kidney tumors at the University of Minnesota Medical Center between 2010 and 2018 along with their segmentations are supplied in the NIFTI format. Total 300 patient's data are available. 210 contain both imaging and segmentation labels and are used for training.

Related Work: In the recent years with advances in artificial intelligence field there have been various models built to segment cancerous tumors from contrast enhanced CT scan images of kidneys.

Some authors make use of an image-based deep learning framework to distinguish three major subtypes of renal cell carcinoma (clear cell, papillary, and chromophobe) using images acquired with computed tomography (CT) [5]. These images were acquired at three phases (phase 1, before injection of the contrast agent; phase 2, 1 min after the injection; phase 3, 5 min after the injection). After image acquisition, rectangular region of interest (ROI) in each phase image was marked by radiologists. A deep learning neural network was trained to classify the subtypes of renal cell carcinoma, using the drawn ROIs as inputs and the biopsy results as labels.

Fabian Isensee *et al.* proposed apply a 3D U-Net to the KiTS19 and attempt to improve upon it by augmenting it with residual and pre-activation residual blocks [6]. Some author used a multi-stage semantic segmentation pipeline for kidney and tumor segmentation from 3D CT images based on 3D U-Net architecture [7].

State of The Art: The U-Net based segmentation architecture is arguably most widely used method for semantic segmentation and in the same line 3D U-Net is more often used for medical image segmentation problems because it performs better than plane 2D method [8, 9]. It's not surprising that even in the KiTS19 challenge, top performing methods are based out 3D U-Nets or variations of 3D methods [10, 11]. The main reasons behind this very minor difference in accuracy of segmentation results are due to different preprocessing techniques, number of layers in the network, activations used and post processing techniques. The main challenge with 3D U-Net is memory, while training. It's hard to fit full 3D image of higher resolution, for example in case of KiTS19 challenge typical CT images have the resolution of 512x512x612. In order to solve this issue, typical techniques used are resizing the original image to lower resolution. Resizing the complete image lower resolution will result in loss of information, which might reduce the accuracy of the result particularly if the area of interest is very small like tumors. Patch based method is a widely used approach where small 3D patch is extracted from both original and ground truth images for training and for prediction of these small patches are combined to generate the high-resolution image [12-14]. This method suffers from outliers because of cropping. One way to solve this problem is overlapped cropping while predicting and taking maximum or average between the overlapping areas. Prediction time in patch-based method is high as it involves prediction several small patches for one image. UNet++is designed for biomedical image segmentation, it uses the Dense block ideas from DenseNet to improve U-Net [15].

Methods and Materials

Data Gathering

As mentioned in introduction, the data acquired was from the GitHub account managed by KiTS19 Challenge. The images and masks were obtained both in 'nii.gz' extension i.e. NIFTI - GNU Zipped Neuroimaging Informatics Technology Initiative-1 Format.

Data Preprocessing

As a first step of preprocessing, intensity is clipped to the range [-30, 300] HU. Our first stage processes the 2D images to identify course kidney region, for this we converted all volumetric CT images to axial orientation and resampled to a new dimension of 128x128 keeping all slices. Intensity values are then normalized to zero mean and unit standard deviation. Data augmentation was performed by randomly rotating ±15° with respect to X, Y and Z axes, translating along X and Y directions and by applying random scaling. In the second stage the identified kidney region is cropped and resized to 256x256 dimension by keeping all slices which has kidney. With this we were able to maintain the high resolution which helped in segmenting tumor accurately.

Methodology

Many proposed methods try to solve this segmentation problem using single neural network. We propose a two-state modified U-Nets to improve the final segmentation. Our proposed method inspired from the conventional algorithms which are usually cascade of different algorithms, where each algorithm outputs are fed in to the next one to obtain result [16, 17].

Our proposed method consists of two neural networks. Stage 1 is a localization network, where given 3D volume is resized to 128x128 in x and y plane and we keep the slice numbers as it is. In this step we are not worried about losing information due to resizing. The sole purpose of this method is to extract the kidney region from the CT whole body or chest scan. To achieve this, we have used 2D U-Net. The output of the network is a kidney region as binary mask. In this step we are not segmenting tumors as the image resolution is low (**Figure 1C, D**). Once we predict for all the slices, we can extract two largest connected components represents two kidneys. Some subjects have only one kidney in that case the size of the second connected component should be at least 30% of the largest component otherwise its neglected. This method is not only fast because its 2D U-Net but also helps to remove not relevant regions as shown in **Figure 1C, D**. Overall using this step we were able to remove 4/5th of the pixels

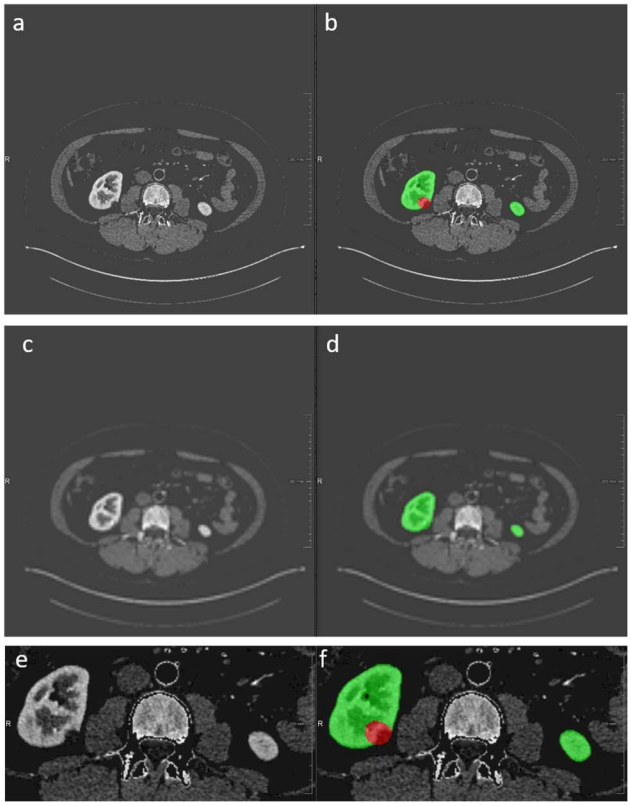


Fig. 1. A) Original CT image with resolution 512x512. **B)** Ground kidney (Green) and tumor (red) overlaid on original image. **C, D)** An example of 2D axial slice of low-resolution(128x128) preprocessed 3D CT images used as input to stage1. The kidney and tumor are considered as a single class and it is shown in green. **E, F)** An example of 2D axial slice of cropped kidney region used for training in stage 2. The kidney is shown in green and tumor is shown in red.

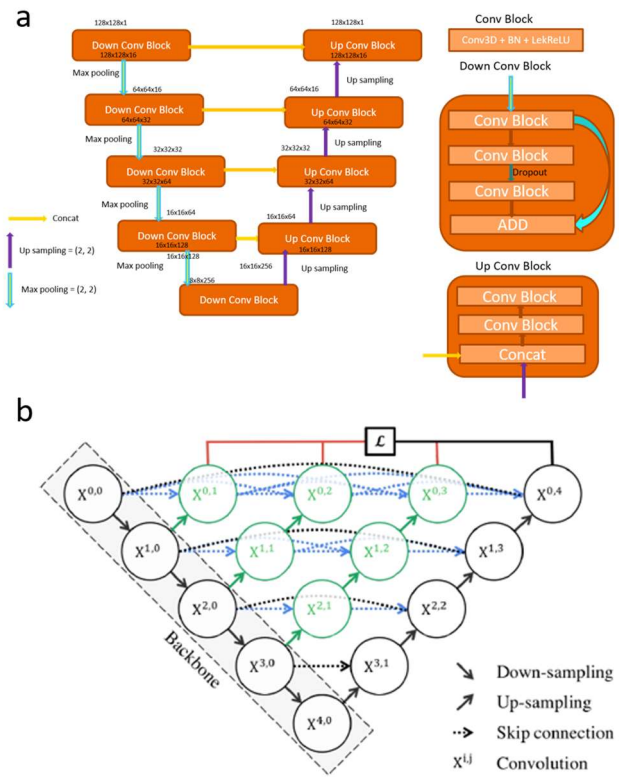


Fig. 2. A) U-Net used in first stage. **B)** UNet++ architecture with skip connection and deep supervision used in [15]

which are not part of kidney and surrounding to it. Which is really helps our Stage 2 neural network to focus on the kidney region **Figure 1E, F**.

State 2 is a 2D UNet++ architecture to locate and segment the kidney and tumor from CT images. The KiTS19 training data has minimum of 29 slices and maximum of 1059 slices. So, resizing slices to a fixed number may result in loss of information. To overcome this problem, we selected 2D approach by keeping all slices even in stage 2. Input to this neural network is a 256x256 2D cropped kidney region from State 1. We have used stage 1 to locate kidney region, as resizing the complete image 256x256 leads to lower resolution and may not be effective to segment the tumors. Our first stage removes the areas other than kidney, so resizing the resulting image from first stage to 256x256 can maintain the high resolution. Second stage focus on the high-resolution kidney region. Since the tumor area is usually located inside the kidney and sometime extend outwards from the organ, using the 2D net we can better segment the tumor. Most of the present neural networks perform better for segmenting the kidney but variations occur in tumor segmentation. Our work presents that using two state neural network we can better segment tumor which is the most important part for disease analysis and treatment planning.

Network Architecture

Stage 1: It is 2D U-Net (shown in **Figure 2A**), for course kidney segmentation. Input is of dimension 128x128 (**Figure 1A** shows an example 2D slice used in stage 1 training). After identifying the course kidney region, the input image is converted back to its original dimension and a 3D bounding box is calculated for the segmented kidney region to crop the kidney area. We removed isolated points by considering two largest connected components. After finding the bounding box of the kidney, we apply 30 pixels offset around the 3D box. This is to avoid some

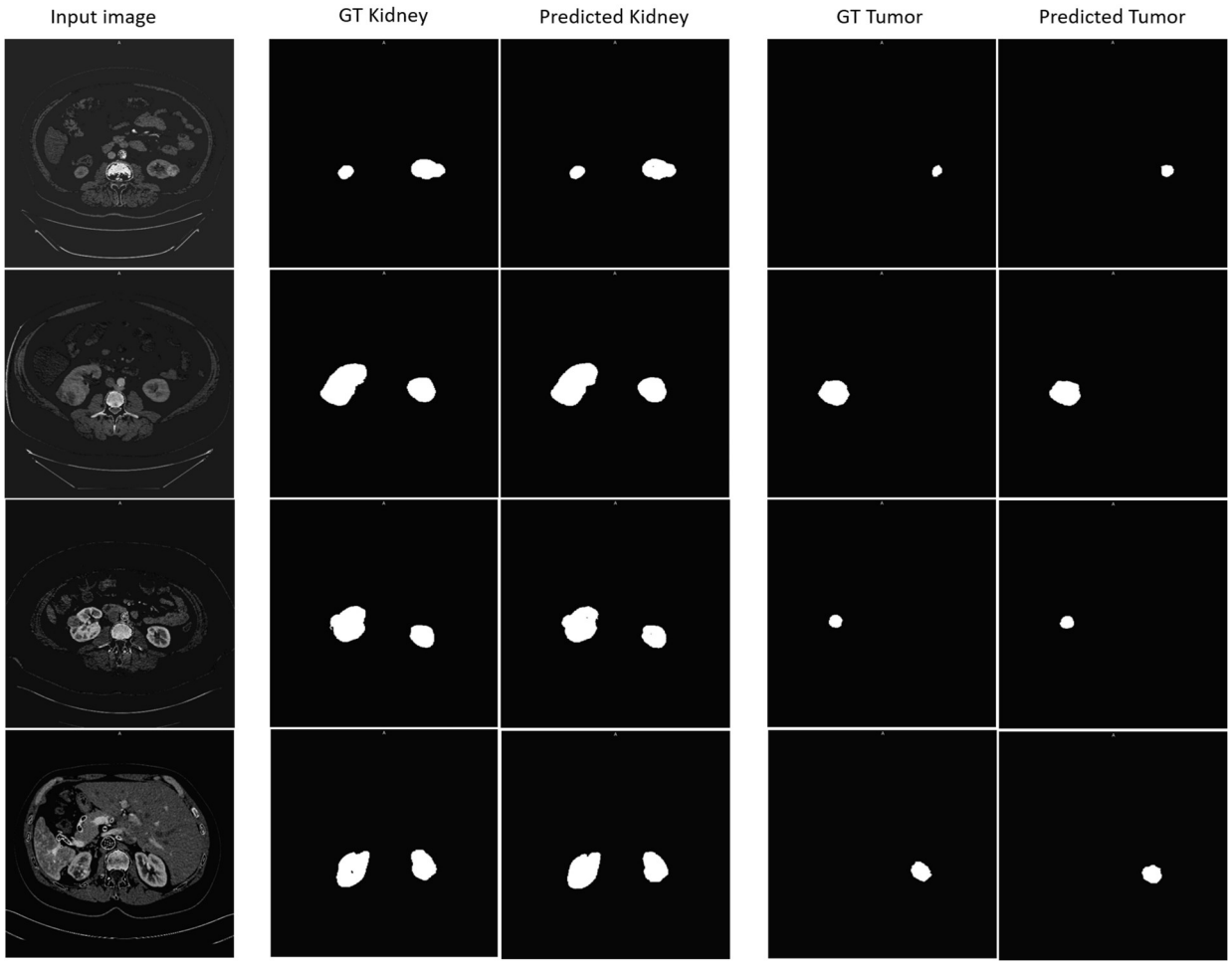


Fig. 3. Example of prediction results.

lesions may grow outside the kidney and our stage 1 network only segments the kidney not tumors.

Stage 2: UNet++ is used in stage2, UNet++ differs from the original U-Net in three ways: i. Having convolution layers on skip pathways, which bridges the semantic gap between encoder and decoder feature maps. ii. Having dense skip connections on skip pathways, which improves gradient flow. iii. Having deep supervision, which enables model pruning and improves or in the worst case achieves comparable performance to using only one loss layer. **Figure 2B** shows the architecture of UNet++ used in [15], we have used the UNet++ available in the Git repository [15] and made some modifications for our training requirements.

The high-resolution cropped kidney region is then resampled to a dimension of 256x256 by keeping all the slices which have kidney area (**Figure 1C** shows an example 2D slice used in stage 2 training). These images are trained to segment the proper kidney area and tumor using U-Net++.

We used Xavier initialization for weights in both the networks. Networks are implemented using Keras with Tensorflow back end and trained on DGX-1 with 2, 16GB Tesla V1 GPUs, batch size of 2, learning rate of 10-4. Dice coefficient between the network output and target mask was used as loss function. The networks are trained for 150 epochs. Dice coefficient (DSC) is a F1- oriented statistic used to gauge the similarity of two sets. Given two sets A and B, the dice coefficient between them is given as follows [18]:

Dice coefficient = $(2|A \cap B|)/(|A|+|B|)$

In our case, A is the set that contains of all positive examples predicted by a model, and B is the set of all golden positive examples in the dataset.

Dice coefficient = $2TP/(2TP+FN+FP) = F1$

TP, True Positive; FN, False Negative; FP, False Positive.

Results

Kidney dice scores were computed by treating both the actual kidney label as well as the tumor label as foreground and everything else as background. The dice computation of the tumors is done only on the tumor labels, like

what they used for the MICCAI challenge evaluation. **Table 1** shows the mean dice of Kidney and tumor of proposed method.

Table 1. Mean dice of the proposed two-stage automatic segmentation method on KiTS19 training dataset.

	Kidney	Tumor
Training	0.98	0.965
Validation	0.97	0.93

We proposed a two-stage network, for coarse to precise segmentation of kidney and tumor on CT images automatically. The model in the first stage is trained on a low-resolution slice to robustly localize the kidney, while the network in the second stage is used to accurately identify each kidney and tumor. Dense network used in our second stage helps to segment kidney and tumor more accurately. Some of example results of our prediction is depicted in **Figure 3**.

Conclusion

In this paper we demonstrated our model and its performance on KiTS19 dataset. Our method was built based on UNet++ architecture and enhanced by squeeze and excite method in the first stage and connected-component based post processing. Proposed method achieved average training dice score of 0.98 and 0.965 with kidney and tumor respectively and on the validation set dice score of 0.97 and 0.93. The model can be improved by adding more training data and with different augmenting methods like applying elastic deformation. We were able to test our model with only Kits data as we were unable to get other data sources with available annotations.

REFERENCES

1. https://www.cancer.org/cancer/kidney-cancer/detection-diagnosis-staging.html by American cancer society (last accessed: 06 Sep. 2020).
2. Heller N, Sathianathen N, Kalapara A, Walczak E, Moore K, Kaluzniak H, Rosenberg J, Blake P, Rengel Z, estreich M, et al. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv:1904.00445
3. Bilic P, Christ PF, Vorontsov E, Chlebus G, Chen H, Dou Q, Fu CW, Han X, Heng PA, Hesser J, et al. (2019). The liver tumor segmentation benchmark (lits). arXiv:1901.04056.
4. https://kits19.grand-challenge.org/home/ (last accessed: 06 Sep. 2020).
5. Han S, Hwang SI, Lee HJ. (2019). The Classification of Renal Cancer in 3-Phase CT Images Using a Deep Learning Method. J Digit Imaging. 32: 638-643. doi: 10.1007/s10278-019-00230-2.
6. Isensee F, & Klaus H. Maier-Hein. (2019). An attempt at beating the 3D U-Net. arXiv:1908.02182.
7. Hou X, Xie C, Li F, Nan Y. (2019). Cascaded Semantic Segmentation for Kidney and Tumor. DOI: 10.24926/548719.002
8. Ronneberger O, Fischer P, Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597
9. Çiçek Ö, Abdulkadir A, Lienkamp S, Brox T, Ronneberger O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. arXiv:1606.06650
10. Ma J. (2019). Solution to the Kidney Tumor Segmentation Challenge 2019. DOI: 10.24926/548719.005
11. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. (2017). Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 35: 18-31. doi: 10.1016/j.media.2016.05.004.
12. de Brebisson A, & Giovanni M. (2015). Deep neural networks for anatomical brain segmentation. arXiv:1502.02445
13. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 36: 61-78. doi: 10.1016/j.media.2016.10.004.
14. Zhou Z, Md Mahfuzur RS, Nima T, Liang J. (2018). Unet++: A nested u-net architecture for medical image segmentation. arXiv:1807.10165
15. Schmitter D, Roche A, Marechal B, Ribes D, Abdulkadir A, Bach-Cuadra M, Daducci A, Granziera C, Kloppel S, Maeder P, Meuli R, et al. (2015). An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. Neuroimage Clin. 7: 7-17. doi: 10.1016/j.nicl.2014.11.001.
16. Fischl B. (2012). FreeSurfer. Neuroimage. 62: 774-781. doi: 10.1016/j.neuroimage.2012.01.021.
17. https://github.com/MrGiovanni/UNetPlusPlus (last accessed: 06 Sep. 2020)
18. Li X, Sun X, Meng Y, Liang J, Wu F, Li J. (2019). Dice Loss for Data-imbalanced NLP Tasks." arXiv:1911.02855.