Kidney and Kidney-tumor Segmentation Using Cascaded V-Nets

Mohammad Arafat Hussain

BiSICL, University of British Columbia, Vancouver, BC, Canada
arafat@ece.ubc.ca

Abstract. Kidney cancer is the seventh most common cancer worldwide, accounting for an estimated 140,000 global deaths annually. Kidney segmentation in volumetric medical images plays an important role in clinical diagnosis, radiotherapy planning, interventional guidance and patient follow-ups however, to our knowledge, there is no automatic kidney-tumor segmentation method present in the literature. In this paper, we address the challenge of simultaneous semantic segmentation of kidney and tumor by adopting a cascaded V-Net framework. The first V-Net in our pipeline produces a region of interest around the probable location of the kidney and tumor, which facilitates the removal of the unwanted region in the CT volume. The second sets of V-Nets are trained separately for the kidney and tumor, which produces the kidney and tumor masks respectively. The final segmentation is achieved by combining the kidney and tumor mask together. Our method is trained and validated on 190 and 20 patients scans, respectively, accesses from 2019 Kidney Tumor Segmentation Challenge database. We achieved a validation accuracy in terms of the Sørensen Dice coefficient of about 97%.

1 Introduction

Kidney cancer is the 7th most common cancer in men and 10th most common cancer in women [1] accounting for an estimated 140,000 global deaths annually [2]. The natural growth pattern varies across kidney cancers, which has led to the development of different prognostic models for the assessment of patient-wise risk [3]. Kidney segmentation in medical images plays an important role in clinical diagnosis, radiotherapy planning, interventional guidance and patient follow-ups [4] however, to our knowledge, there is no automatic kidney-tumor segmentation method present for in the 3D volumetric medical images (e.g. computed tomography (CT), magnetic resonance (MR), etc.) .

Prior to the wide range of use of various machine-learning approaches for kidney segmentation, a number of traditional methods were proposed in the literature that made use of image thresholding, graph cuts, level sets, active contours, multi-atlas image registration, template deformation, etc. For example, Yan et al. [5] proposed a simple intensity thresholding-based method, which is often inaccurate and was limited to 2D. Other intensity-based methods have used graph cuts [6] and active contours/level sets [7]. But these methods are sensitive to the choice of parameters [8], which often need to be tweaked for different images. In addition, the graph cuts [6] and level sets-based [7] methods are prone to leaking through weak anatomical boundaries in the image, and often require considerable computation [8]. The methods proposed by Lin et al. [9] and Yang et al. [10] rely extensively on prior knowledge of kidney shapes. However, building a realistic model of kidney shape variability and balancing the influence of the model on the resulting segmentation are non-trivial tasks.

To overcome the aforementioned limitations of the traditional methods, a number of kidney segmentation methods have been proposed based on manual feature-engineering-based supervised learning. Cuingnet et al. [11] used a classification forest to generate a kidney spatial probability map and then deformed a ellipsoidal template to approximate the probability map and generate the segmentation. Due to this restrictive template-based approach, it is likely to fail for kidneys having abnormal shape due to disease progression and/or internal tumors. Therefore, crucially, [11] did not include the truncated kidneys (16% of their data) in their evaluation. Even then, their proposed method did not correctly detect/segment about 20% of left and 20% of right, and failed for another 10% left and 10% right kidneys of their evaluation data set. Glocker et al. [12] used a joint classification-regression forest scheme to segment different abdominal organs, but their approach suffers from leaking, especially for kidneys, as evident in their results.

Avoiding complex manual feature engineering, supervised deep learning using convolutional neural networks (CNN) have exploded in popularity for automatic feature learning, classification, as well as localization and dense labelling. Thong et al. [4] showed promising kidney segmentation performance using CNN, however, it was designed only for 2D contrast-enhanced CT slices. To facilitate 3D segmentation of organs including kidney, a number 3D neural network approaches have been proposed recently. For example, Kekeya et al. [13] used judgment assisted probabilistic atlas to generate probability map for eight different organ locations and then trained eight 3D U-Nets [14] per organ for segmentation. Chen et al. [15] and Roth et al. [16,**?**] used two 3D U-Nets [14] in a cascaded fashion for organ segmentation, where the first 3D U-Net produces the region-of-interest (ROI) to reduce the search space, and the second 3D U-Net produces the organ segmentation. Gibson et al. [17] proposed a variant of the V-Net [18], namely DenseVNet, for multi organ segmentation including left kidney. These 3D segmentation approaches showed promise in kidney segmentation, however, not tested on the kidney tumor segmentation task.

In this work, we address the challenge of simultaneous semantic segmentation of kidney and kidney-tumor by using a cascaded V-Net architecture. The first V-Net, namely ROI-V-Net produces a ROI around the probable location of the kidney and tumor, while the second sets of V-Nets, namely Kidney-V-Net and Tumor-V-Net, are trained in parallel on the ROI data to produce the kidney and tumor segmentation separately. These kidney and tumor segmentations are then joined together by comparing the probabilities associated with each voxel to be a kidney or a tumor or a background.

2 Materials and Methods

2.1 Data

We used CT scans of 300 patients from the 2019 Kidney Tumor Segmentation Challenge database [19]. These patients underwent radical nephrectomy or partial nephrectomy at the University of Minnesota between 2010 and mid-2018 to excise a renal tumor. Ground truth semantic segmentations for arterial phase abdominal CT scans of 300 unique kidney cancer patients were performed by medical students under the supervision of an expert radiologist. Out of 300 patient scans, we used 190 and 20 scans in model training and validation, respectively. The remaining 90 scans are used for objective model evaluation. Although the voxel spacing among the datasets were variable, the challenge database made available of an uniformly interpolated version of the same datasets, where the voxel spacing was set to 3mm, 0.78162497mm and 0.78162497mm in the axial, coronal and sagittal directions, respectively. In this work, we used these interpolated scans.

2.2 Semantic Segmentation of Kidney and Tumor

Kidney ROI Generation Using ROI-V-Net: We use a V-Net (Fig. 2) architecture, namely ROI-V-Net, to roughly segment the kidney and tumor semantically in order to produce a ROI around the kidney and tumor. In this way, we can discard the unwanted region in the CT volume. The input to this V-Net is a 128×128×128 voxels 1-channel 3D patch. The V-Net has 4 levels and it uses 1, 2, 3, and 3 number of convolution operations in the compression side (i.e. left side), respectively and 3, 3, 2, and 1 numbers of convolutions in the decompression side (i.e. right side). The output layer produces 2-channel 3D predictions of size 128×128×128 voxels, one for background and one for the kidney and tumor. These 2-channel predictions are fed to a softmax operator. We use parametric rectified linear unit (PReLU) activation throughout the network.

Once the ROI-V-Net is trained, we used this model on the test data to generate the kidney+tumor predictions, i.e. inseperate kidney+tumor mask. Note that at this phase, we do not distinguish between the kidney and tumor. We divide the prediction volume in half along the coronal direction in order to separate the left and right kidneys first. Then we projected the 3D prediction into a 2D distribution map by adding the mask along the axial and coronal directions. We used a median filtering of window sizes 10×25 pixels and 25×25 pixels on the coronal and axial projections, respectively. Then from the mid point (x_m, y_m, z_m) of the kidney and tumor distribution, we select ROIs from the left and right half CT scans (Fig. 1) of dimension $[max(1, x_m - 127), min(x_m + 128, x_f)] \times [max(1, y_m - 127), min(y_m + 128, \lfloor (y_f/2) \rfloor)] \times [max(1, z_m - 127), min(z_m + 128, \lfloor (z_f/2) \rfloor)]$, where x_m, y_m, and z_m are the dimension of the interrogated CT scan, and $max(a, b)$ and $min(a, b)$ take the maximum and minimum between a and b, respectively. If there is no kidney present or the ROI-V-Net fails to detect any kidney mask, then we consider the corresponding medially divided whole CT scan as the ROI. We also record the ROI locations inside the actual image volume for later use.

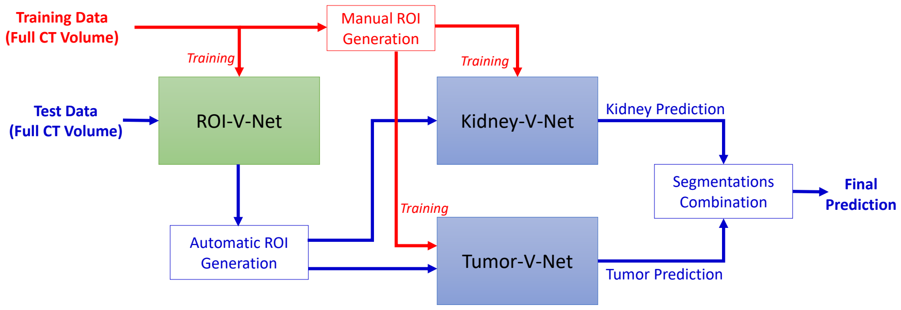

Fig. 1. The schematic diagram of our cascaded V-Net architecture for the semantic segmentation of kidney and tumor.

Semantic Segmentation of Kidney and Tumor: We use another two V-Nets (Fig. 2) of similar architecture as in ROI-V-Net, namely Kidney-V-Net and Tumor-V-Net, to segment the kidney and tumor, respectively. These two V-Nets are trained using manually generated ROI around the kidney and tumor. The input to both these V-Nets is a 128×128×128 voxels 1-channel 3D patch. These V-Nets also have 4 levels having 1, 2, 3, and 3 number of convolution operations in the compression side (i.e. left side), respectively and 3, 3, 2, and 1 numbers of convolutions in the decompression side (i.e. right side). The output layer produces 2-channel 3D predictions of size 128×128×128 voxels, one for background and one for the kidney (by Kidney-V-Net) or tumor (by Tumor-V-Net). These 2-channel predictions are fed to a softmax operator. Here also, we use parametric rectified linear unit (PReLU) activation throughout the network.

After the Kidney-V-Net and Tumor-V-Net are trained, we used these models on the test data ROI, produced by the ROI-V-Net, to generate the kidney and tumor predictions, respectively. Then we overlay the tumor mask in the kidney mask. Finally, we use the previously recorded ROI location information to reconstruct the prediction map of size equal to the interrogated image volume, with value 0 for background, 1 for kidney, and 2 for tumor.

Training: We trained our networks by minimizing the Sørensen Dice loss between the ground truth and predicted labels defined as:

$$L = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N p_i}, \tag{1}$$

where N is the total number of voxels in an interrogated image volume, p_i is the predicted binary voxel value, and g_i is the ground truth binary voxel value. The

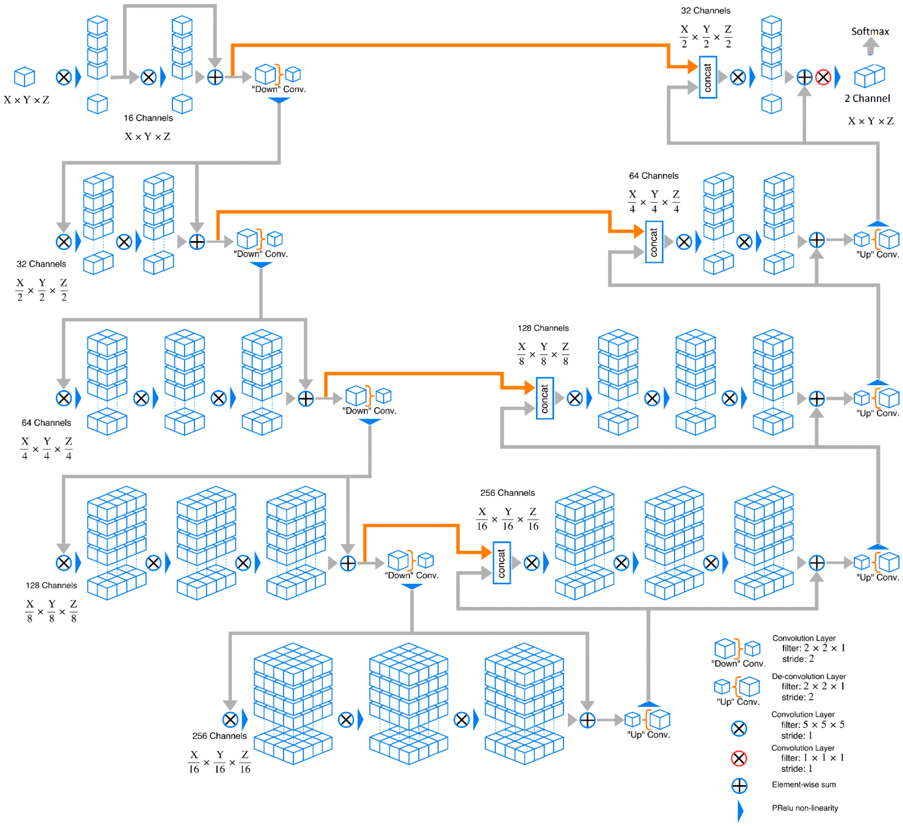

Fig. 2. The V-Net architecture used in this work. Image credit [20].

base learning rate was set to 0.001 and weight decay was set to 0.01. We used the stochastic gradient descent optimizer with momentum set to 0.9. Before training, we normalized each dataset using statistical normalization with $\sigma = 0.25$. We also used resampling of datasets using linear interpolation so that each voxel has a dimension of 0.45mm×0.45mm×0.45mm. We also added random noise in the training data for augmentation and used random cropping to generate 3D image patch during training. The training batch size per iteration was set to 1. The V-Net is implemented in Tensorflow, which is adopted from this repository [21]. Training was performed on a workstation with Intel 4.0 GHz Core-i7 processor, an Nvidia GeForce Titan Xp GPU with 12 GB of VRAM, and 32 GB of RAM.

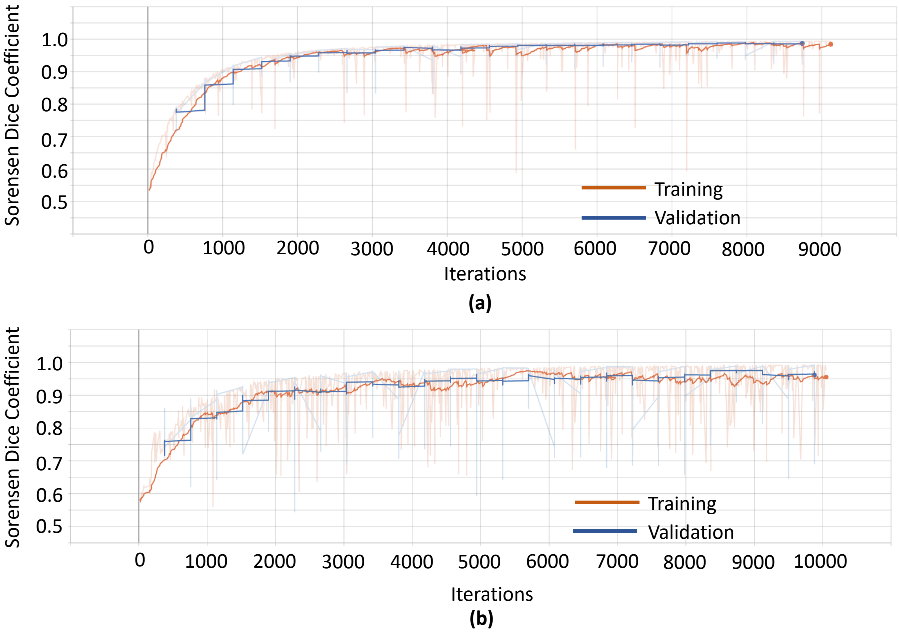

Fig. 3. Graphs showing Sørensen Dice coefficient values vs. training iterations for (a) Kidney-V-Net and (b) Tumor-V-Net.

3 Results

In Fig.3, we show the Sørensen Dice coefficients with respect to the training iterations. We see in Fig.3(a) for the Kidney-V-Net that after about 8,500 iterations (i.e. ∼47 epochs), the Sørensen Dice coefficient for validation data was about 0.9650. We also see Fig.3(b) for the Tumor-V-Net that around 10,000 iterations (i.e. ∼56 epochs), the Sørensen Dice coefficient for validation data was about 0.9860. The average Sørensen Dice coefficient combining the Kidney and Tumor segmentation for 20 validation patient cases is about 0.9755.

4 Conclusions

In this work, we proposed a cascaded V-Nets framework for simultaneous semantic segmentation of kidney and tumor. Our first V-Net in the pipeline produced a ROI around the probable location of the kidney and tumor, facilitating the removal of the unwanted region in the CT volume, while the second sets of V-Nets produced the kidney and tumor segmentations. Our validation results in terms of the Sørensen Dice coefficient of about 97% shows robust segmentation performance.

Acknowledgement: We thank NVIDIA Corporation for supporting our research through their GPU Grant Program by donating the GeForce Titan Xp.

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2016. CA: a cancer journal for clinicians **66**(1) (2016) 7–30
2. Ding, J., Xing, Z., Jiang, Z., Chen, J., Pan, L., Qiu, J., Xing, W.: CT-based radiomic model predicts high grade of clear cell renal cell carcinoma. European journal of radiology **103** (2018) 51–56
3. Escudier, B., Porta, C., Schmidinger, M., Rioux-Leclercq, N., Bex, A., Khoo, V., Gruenvald, V., Horwich, A.: Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Annals of Oncology **27**(suppl_5) (2016) v58–v68
4. Thong, W., Kadoury, S., Piché, N., Pal, C.J.: Convolutional networks for kidney segmentation in contrast-enhanced ct scans. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **6**(3) (2018) 277–282
5. Yan, G., Wang, B.: An automatic kidney segmentation from abdominal ct images. In: 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems. Volume 1., IEEE (2010) 280–284
6. Li, X., Chen, X., Yao, J., Zhang, X., Tian, J.: Renal cortex segmentation using optimal surface search with novel graph construction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2011) 387–394
7. Zhang, Y., Matuszewski, B.J., Shark, L.K., Moore, C.J.: Medical image segmentation using new hybrid level-set method. In: 2008 fifth international conference biomedical visualization: information visualization in medical and biomedical informatics, IEEE (2008) 71–76
8. Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S.: Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. Medical image analysis **30** (2016) 120–129
9. Lin, D.T., Lei, C.C., Hung, S.W.: Computer-aided kidney segmentation on abdominal ct images. IEEE transactions on information technology in biomedicine **10**(1) (2006) 59–65
10. Yang, G., Gu, J., Chen, Y., Liu, W., Tang, L., Shu, H., Toumoulin, C.: Automatic kidney segmentation in ct images based on multi-atlas image registration. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE (2014) 5538–5541
11. Cuingnet, R., Prevost, R., Lesage, D., Cohen, L.D., Mory, B., Ardon, R.: Automatic detection and segmentation of kidneys in 3d ct images using random forests. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2012) 66–74
12. Glocker, B., Pauly, O., Konukoglu, E., Criminisi, A.: Joint classification-regression forests for spatially structured multi-object segmentation. In: European conference on computer vision, Springer (2012) 870–881
13. Kakeya, H., Okada, T., Oshiro, Y.: 3d u-japa-net: Mixture of convolutional networks for abdominal multi-organ ct segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2018) 426–433

14. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention, Springer (2016) 424–432
15. Chen, S., Roth, H., Dorn, S., May, M., Cavallaro, A., Lell, M.M., Kachelrieß, M., Oda, H., Mori, K., Maier, A.: Towards automatic abdominal multi-organ segmentation in dual energy ct using cascaded 3d fully convolutional network. arXiv preprint arXiv:1710.05379 (2017)
16. Roth, H.R., Oda, H., Zhou, X., Shimizu, N., Yang, Y., Hayashi, Y., Oda, M., Fujiwara, M., Misawa, K., Mori, K.: An application of cascaded 3d fully convolutional networks for medical image segmentation. Computerized Medical Imaging and Graphics **66** (2018) 90–99
17. Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B., Pereira, S.P., Clarkson, M.J., Barratt, D.C.: Automatic multi-organ segmentation on abdominal ct with dense v-networks. IEEE transactions on medical imaging **37**(8) (2018) 1822–1834
18. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), IEEE (2016) 565–571
19. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., Dean, J., Tradewell, M., Shah, A., Tejpaul, R., Edgerton, Z., Peterson, M., Raza, S., Regmi, S., Papanikolopoulos, N., Weight, C.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes (2019)
20. Monteiro, M.: Vnet-tensorflow: Tensorflow implementation of the V-Net architecture for medical imaging segmentation. https://github.com/MiguelMonteiro/VNet-Tensorflow (2018)
21. Ko, J.K.: Implementation of V-Net in tensorflow for medical image segmentation. https://github.com/jackyko1991/vnet-tensorflow (2018)