3D Kidneys and Kidney Tumor Semantic Segmentation using Boundary-Aware Networks

Andriy Myronenko[1] and Ali Hatamizadeh[1,2]

[1] NVIDIA, Santa Clara, CA, USA
[2] Computer Science Department, University of California, Los Angeles, CA, USA
{amyronenko,ahatamizadeh}@nvidia.com

Abstract. Automated segmentation of kidneys and kidney tumors is an important step in quantifying the tumor's morphometrical details to monitor the progression of the disease and accurately compare decisions regarding the kidney tumor treatment. Manual delineation techniques are often tedious, error-prone and require expert knowledge for creating unambiguous representation of kidneys and kidney tumors segmentation. In this work, we propose an end-to-end boundary aware fully Convolutional Neural Networks (CNNs) for reliable kidney and kidney tumor semantic segmentation from arterial phase abdominal 3D CT scans. We propose a segmentation network consisting of an encoder-decoder architecture that specifically accounts for organ and tumor edge information by devising a dedicated boundary branch supervised by edge-aware loss terms. We have evaluated our model on 2019 MICCAI KiTS Kidney Tumor Segmentation Challenge dataset and our method has achieved dice scores of 0.9742 and 0.8103 for kidney and tumor repetitively and an overall composite dice score of 0.8923.

Keywords: Abdominal CT · Kidneys · Tumor · Segmentation Deep Learning · Convolutional Neural Networks

1 Introduction

Kidney cancer accounted for nearly 175,000 deaths worldwide in 2018 [1], and it is projected that 14,770 deaths will occur due to the disease in 2019 in the US [10]. Current kidney tumor treatment planning include Radical Nephrectomy (RN) and Partial Nephrectomy (PN). In RN, both the tumor and the affected kidney are removed whereas in PN the tumor is removed but kidneys are saved [11]. Although RNs were historically prevalent as a standard treatment procedure for kidney tumors, new capabilities for earlier detection of the tumors as well as advancements in surgery has made PNs a viable treatment approach [6].

Automated segmentation of kidneys and kidney tumors assists physicians to obtain accurate morphometrical details of the tumor in an efficient and reliable manner as the manual delineation process is often tedious and error-prone. The decision for kidney tumor treatment plan can be made by leveraging such important tumor's morphometrical information. Recently, deep learning approaches for
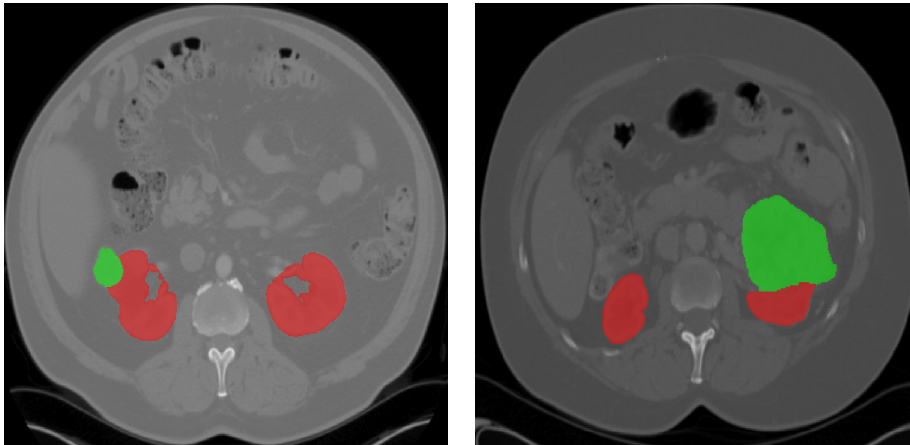
Fig. 1. Example of an axial slice of 3D CT images of two patients in KiTS dataset. Red color indicates kidneys, green color indicates tumor region.

semantic image segmentation have demonstrated prominent results in medical image analysis for various applications [8,9,3,2]. The powerful non-linear feature extraction capabilities of CNNs along with the effectiveness of the encoder-decoder architectures have made it possible to employ CNNs for challenging segmentation tasks.

In this paper, we propose a boundary-aware fully Convolutional Neural Networks for end-to-end and reliable semantic segmentation of kidneys and kidney tumor by encoding the information of edges in a dedicated stream that is supervised by edge-aware losses. We have trained and tested our model on 2019 KiTS Kidney Tumor Segmentation Challenge and results demonstrate the effectiveness of our proposed framework.

2 Related Work

Traditionally, various techniques such as deformable models, GrabCuts, region growing and atlas-based methods have been applied to the problem of kidney segmentation. In recent years, researchers have attempted to leverage the power of deep learning and CNNs to build segmentation frameworks that are more automated and less dependant on incorporation of prior shape statistics. Thong et al. [12] proposed a 2D patch-based approach for kidney segmentation in contrast-enhanced CT scans by leveraging a modified ConvNet. Jackson et al. [7] developed a framework for detection and segmentation and of kidneys in non-contrast CT images by utilizing a 3D U-Net. Yang et al. [14] proposed a method for kidney and renal tumor segmentation in CT angiography image by a modified residual FCN that is equipped with a pyramid pooling module. Furthermore, Yin et al. [15] employed a cascaded approach for segmentation of kidneys with renal cell carcinoma by training a CNN that predicts a bounding box around kidney and a
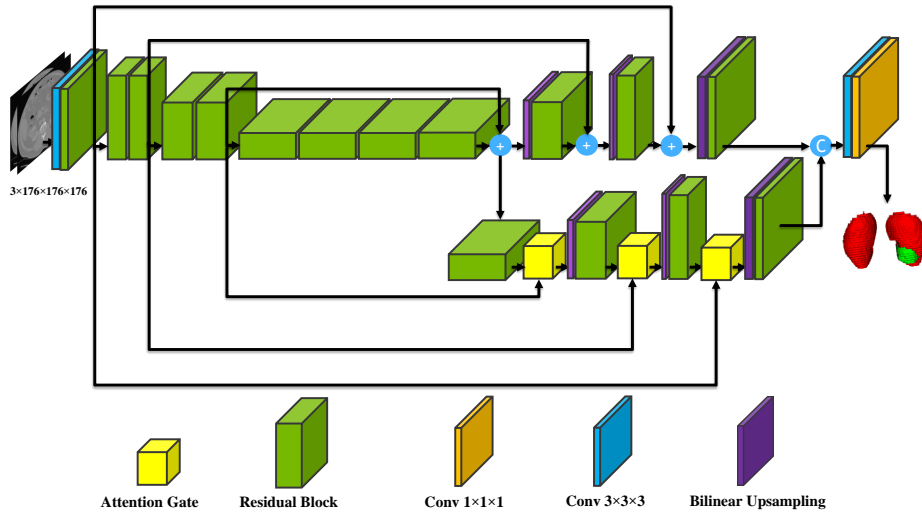
Fig. 2. Our proposed CNN architecture.

subsequent CNN that segments the kidneys. Recently, Xia et al. [13] proposed a two-stage approach for segmentation of kidney and space-occupying lesion area by using SCNN and ResNet for image retrieval and SIFT-flow and MRF for smoothing and pixel matching.

3   Methods

3.1   Framework Architecture

As illustrated in Figure 2, our network consists of the main segmentation branch and the additional boundary stream that processes the feature maps at the boundary level [4]. The main branch follows [9] an asymmetric encoder-decoder structure. The input to the encoder is a 176x176x176 crop which is initially fed into a 3x3x3 convolution with 16 filters. Feature maps are then extracted at each resolution by feeding them into a residual block [5] followed by a strided 3x3x3 convolution (for downsizing and doubling of feature dimension). The bottom of the encoder entails four consecutive residual blocks that are connected to the decoder. The extracted feature maps in the decoder are upsampled using bilinear interpolation and added with feature maps from the encoder. The output of the decoder is concatenated with the output of the boundary and fed into a 1x1x1 convolution with 2 channels where channel-wise sigmoid activation $\sigma(X) = \frac{1}{1+\exp(-X)}$ determines the probability of each voxel belonging to kidneys and tumor or only tumor classes.

3.2   Boundary Stream

The purpose of the boundary stream is to highlight the edge information of the feature maps extracted in the main encoder by leveraging an additional attention-driven decoder. The attention gates in every resolution of the boundary stream process the feature maps that are learned in the main encoder as well as the output of the previous attention gates. For the first attention gate, we first concatenate the output of the encoder with its previous resolution and feed it into a residual block. In the attention gates, each input is first fed into a 3x3x3 convolutional layer with matching number of feature maps and then fused together, followed by ReLU. The output of the ReLU is fed into a 1x1x1 convolution layer followed by sigmoid function σ to obtain the attention map. Consecutively, an element-wise multiplication between the boundary stream feature maps and the computed attention map results in the output of the attention gates.

3.3   Loss Functions

We use a dice loss function on the predicted outputs of the main stream as well as the boundary stream. The dice loss is as follows [8]:

$$L_{Dice} = 1 - \frac{2 * \sum y_{true} * y_{pred}}{\sum y_{true}^2 + \sum y_{pred}^2 + \epsilon} \tag{1}$$

Where y_{pred}, y_{true} denote the voxel-wise semantic predictions of the main stream and their corresponding labels, ϵ is a small constant to avoid division by zero and summation is carried over the total number of voxels.

Additionally, we add a weighted Binary Cross Entropy (BCE) loss to the boundary stream loss in order to deal with the imbalanced number of boundary and non-boundary voxels:

$$\begin{aligned} L_{BCE} = &-\beta \sum_{j \in y_+} log P(y_{pred,j} = 1|x;\theta) \\ &- (1-\beta) \sum_{j \in y_-} log P(y_{pred,j} = 0|x;\theta) \end{aligned} \tag{2}$$

Where x, θ, y_- and y_+ denote the 3D input image, CNN parameters, edge and non-edge voxel sets respectively. β is the ratio of non-edge pixels over the entire number of voxels and $P(y_{pred,j})$ denotes the probability of the predicated class at voxel j.

The total loss function that is minimized during training is computed by taking the average of losses for tumor-only and foreground class predictions.

4 Implementation Details and Dataset

KiTS 2019 dataset : Kidney Tumor Segmentation Challenge (KiTS 2019) provides data of multi-phase 3D CTs, voxel-wise ground truth labels, and comprehensive clinical outcomes for 300 patients who underwent nephrectomy for kidney tumors between 2010 to 2018 at University of Minnesota [6]. 210 patients were randomly selected for the training set and the remaining 90 patients were left as a testing set. The annotation was performed in the transverse plane with regular subsampling of series in the longitudinal direction with roughly 50 annotated slices depicting the Kidney for each patient. The labels for excluded slices were computed by using a contour interpolation algorithm [6]. Figure 1 illustrates 2D axial view of the example images from two patients in the training set of KiTS 2019.

Data processing : We normalized the CT data to [-1, 1] range by dividing the intensity values by 1000 and clipping the values that fall outside this range. For training, images were re-sampled to 1x1x1mm isotropic resolution and re-sampled back to their original resolution after the inference. The re-sampled output size of the images was on average 512x512 in axial plane and $400 - 800$ along the inf-sup direction.

Implementation details :We have implemented our method in Pytorch[3]. Since the re-sampled CT image were often large, we used a 176x176x176 crop during training. The cropping region was centered on the kidney tumor label (with probability 0.8), on any foreground (with probability 0.1) and on background (with probability 0.1). We found it important to sample more frequently from the tumor region. The model was trained on 8 NVIDIA Tesla V100 16GB GPUs (DGX-1 server). We used a batch size of 8 and the Adam optimization algorithm with the initial learning rate of $\alpha_0 = 5e - 5$ that was further decreased according to $\alpha = \alpha_0 * (1 - e/N_e)^{0.9}$ [9], where e and N_e denote the current epoch counter total number of epochs (300 in our case). During inference, we have leveraged test time augmentation (TTA) and have used an ensemble of 5 models to further improve the results.

Evaluation metrics : We have adopted the same three evaluation metrics as outlined by KiTS 2019 challenge. Kidneys dice denote the segmentation performance when considering both kidneys and tumors as the foreground whereas tumor dice considers everything except the tumor as background. Composite dice is simply the average of kidneys dice and tumor dice.

5 Results and Discussion

Preliminary : Table 1 represents the evaluation results of our model on our own dataset split. We divided the training set of KiTS 2019 dataset into our own

---

[3] http://pytorch.org/

Table 1. Preliminary dice results based on our own data split as well as 2 approximate scores provided by KiTS 2019 submission portal.

Model	Kidneys Dice	Tumor Dice	Composite Dice
Our split (single model)	0.957	0.821	0.889
Our split (TTA + ensemble)	**0.970**	**0.834**	**0.902**
Approximate score (single model)	0.955	0.736	0.845
Approximate score (TTA + ensemble)	**0.974**	**0.784**	**0.879**

subsets for training and validation and evaluated the performance of a single model as well as an ensemble of 5 models. Finally the KiTS 2019 submission provided two approximate scores based on a small subset of its validation dataset and this allowed us to list the approximate scores of the single model and the ensemble model. Evidently, the dice scores from evaluations on our own split were similar and consistent with the approximate scores provided by the submission portal(see Table 1). Figure 3 illustrates the segmentation visualizations of our method and their corresponding ground truth from two cases in the validation set of our own split.

KiTS 2019 Test Set : The evaluation results[4] of our model on the testing set of KiTS 2019 dataset is presented in Table 2. The kidneys dice is very similar to approximate scores obtained by the ensemble model that utilizes TTA while the tumor dice is %3.35 better than its counterpart. Moreover, our method ranks 9th overall in terms of the composite dice of kidneys and Tumor among 100 participants in KiTS 2019 challenge. Our model in particular performed better on kidneys segmentation task.

Table 2. The Evaluation results of our model on KiTS 2019 test set.

Kidneys Dice	Tumor Dice	Composite Dice
0.9742	0.8103	0.8923

6   Conclusion

In this work, we have proposed an end-to-end 3D framework for reliable and automated segmentation of kidneys and kidney tumors. Our network consists of a an encoder-decoder architecture equipped with a boundary stream that processes the edge information separately and is supervised by edge-aware losses. We have validated the effectiveness of our approach by training and testing our model on 2019 MICCAI KiTS Kidney Tumor Segmentation Challenge dataset. Our method has achieved dice scores of 0.9742 and 0.8103 for kidney and tumor
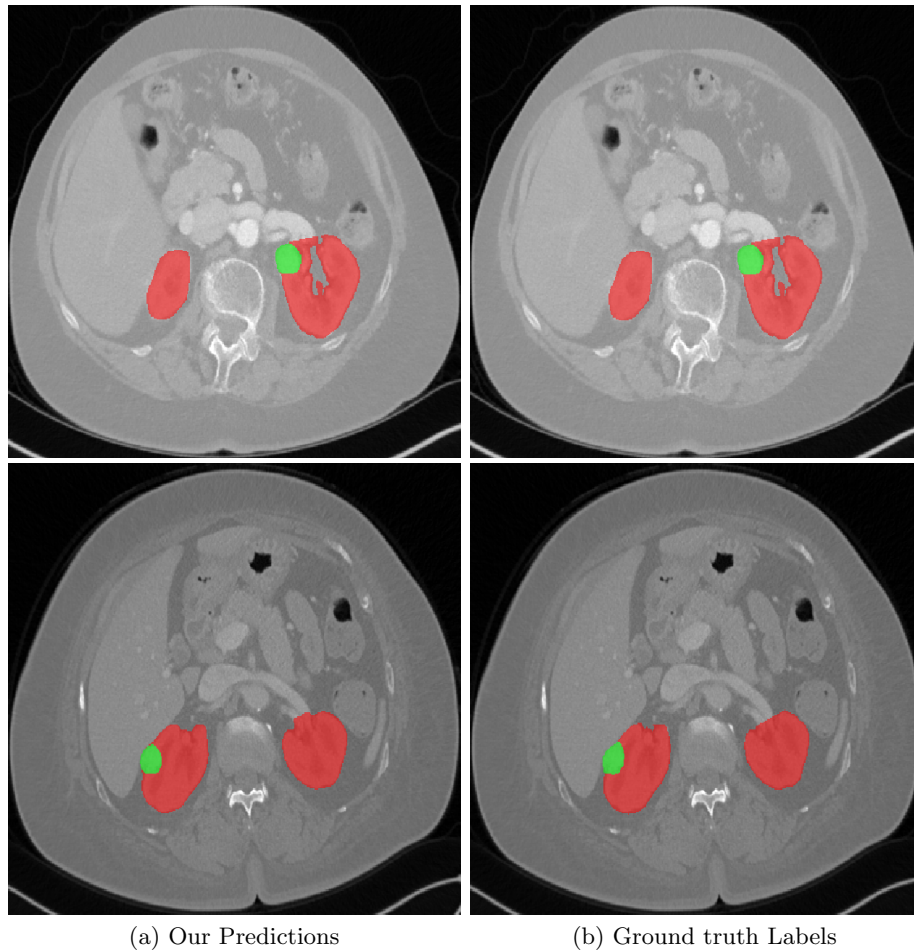
---

[4] http://results.kits-challenge.org/miccai2019/

(a) Our Predictions   (b) Ground truth Labels

Fig. 3. Visualization of (a) our model's predictions (b) ground truth labels

repetitively and an overall composite dice score of 0.8923 and ranks 9th overall in terms of composite dice among 100 participants of this challenge.

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **68**(6), 394–424 (2018)
2. Hatamizadeh, A., Hoogi, A., Sengupta, D., Lu, W., Wilcox, B., Rubin, D., Terzopoulos, D.: Deep active lesion segmentation. arXiv preprint arXiv:1908.06933 (2019)

3. Hatamizadeh, A., Hosseini, H., Liu, Z., Schwartz, S.D., Terzopoulos, D.: Deep dilated convolutional nets for the automatic segmentation of retinal vessels. arXiv preprint arXiv:1905.12120 (2019)
4. Hatamizadeh, A., Terzopoulos, D., Myronenko, A.: Boundary aware networks for medical image segmentation. arXiv preprint arXiv:1908.08071 (2019)
5. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European Conference on Computer Vision (ECCV) (2016)
6. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019)
7. Jackson, P., Hardcastle, N., Dawe, N., Kron, T., Hofman, M., Hicks, R.J.: Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy. Frontiers in oncology **8**, 215 (2018)
8. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision (3DV) (2016)
9. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: BrainLes, Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 311–320. LNCS, Springer (2018), https://arxiv.org/abs/1810.11654
10. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2019. CA: a cancer journal for clinicians **69**(1), 7–34 (2019)
11. Sun, M., Abdollah, F., Bianchi, M., Trinh, Q.D., Jeldres, C., Thuret, R., Tian, Z., Shariat, S.F., Montorsi, F., Perrotte, P., et al.: Treatment management of small renal masses in the 21st century: a paradigm shift. Annals of surgical oncology **19**(7), 2380–2387 (2012)
12. 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), 277–282 (2018)
13. Xia, K.j., Yin, H.s., Zhang, Y.d.: Deep semantic segmentation of kidney and space-occupying lesion area based on scnn and resnet models combined with sift-flow algorithm. Journal of medical systems **43**(1), 2 (2019)
14. Yang, G., Li, G., Pan, T., Kong, Y., Wu, J., Shu, H., Luo, L., Dillenseger, J.L., Coatrieux, J.L., Tang, L., et al.: Automatic segmentation of kidney and renal tumor in ct images based on 3d fully convolutional neural network with pyramid pooling module. In: 2018 24th International Conference on Pattern Recognition (ICPR). pp. 3790–3795. IEEE (2018)
15. Yin, K., Liu, C., Bardis, M., Martin, J., Liu, H., Ushinsky, A., Glavis-Bloom, J., Chantaduly, C., Chow, D.S., Houshyar, R., et al.: Deep learning segmentation of kidneys with renal cell carcinoma. (2019)