Segmentation of Kidney Tumors on Non-Contrast CT Images Using Protuberance Detection Network

Taro Hatsutani[✉], Akimichi Ichinose, Keigo Nakamura, and Yoshiro Kitamura

Medical Systems Research and Development Center,
FUJIFILM Corporation, Tokyo, Japan
taro.hatsutani@fujifilm.com

Abstract. Many renal cancers are incidentally found on non-contrast CT (NCCT) images. On contrast-enhanced CT (CECT) images, most kidney tumors, especially renal cancers, have different intensity values compared to normal tissues. However, on NCCT images, some tumors called isodensity tumors, have similar intensity values to the surrounding normal tissues, and can only be detected through a change in organ shape. Several deep learning methods which segment kidney tumors from CECT images have been proposed and showed promising results. However, these methods fail to capture such changes in organ shape on NCCT images. In this paper, we present a novel framework, which can explicitly capture protruded regions in kidneys to enable a better segmentation of kidney tumors. We created a synthetic mask dataset that simulates a protuberance, and trained a segmentation network to separate the protruded regions from the normal kidney regions. To achieve the segmentation of whole tumors, our framework consists of three networks. The first network is a conventional semantic segmentation network which extracts a kidney region mask and an initial tumor region mask. The second network, which we name protuberance detection network, identifies the protruded regions from the kidney region mask. Given the initial tumor region mask and the protruded region mask, the last network fuses them and predicts the final kidney tumor mask accurately. The proposed method was evaluated on a publicly available KiTS19 dataset, which contains 108 NCCT images, and showed that our method achieved a higher dice score of 0.615 (+0.097) and sensitivity of 0.721 (+0.103) compared to 3D-UNet. To the best of our knowledge, this is the first deep learning method that is specifically designed for kidney tumor segmentation on NCCT images.

Keywords: Renal Cancer · Tumor Segmentation · Non-Contrast CT

1 Introduction

Over 430,000 new cases of renal cancer were reported in 2020 in the world [1] and this number is expected to rise [22]. When the tumor size is large (greater

than 7 cm) often the whole kidney is removed, however, when the tumor size is small (less than 4 cm), partial nephrectomy is the preferred treatment [20] as it could preserve kidney's function. Thus, early detection of kidney tumors can help to improve patient's prognosis. However, early-stage renal cancers are usually asymptomatic, therefore they are often incidentally found during other examinations [19], which includes non-contrast CT (NCCT) scans.

Segmentation of kidney tumors on NCCT images adds challenges compared to contrast-enhanced CT (CECT) images, due to low contrast and lack of multiphase images. On CECT images, the kidney tumors have different intensity values compared to the normal tissues. There are several works that demonstrated successful segmentation of kidney tumors with high precision [13,21]. However, on NCCT images, as shown in Fig. 1b, some tumors called isodensity tumors, have similar intensity values to the surrounding normal tissues. To detect such tumors, one must compare the kidney shape with tumors to the kidney shape without the tumors so that one can recognize regions with protuberance.

3D U-Net [3] is the go-to network for segmenting kidney tumors on CECT images. However, convolutional neural networks (CNNs) are biased towards texture features [5]. Therefore, without any intervention, they may fail to capture the protuberance caused by isodensity tumors on NCCT images.

In this work, we present a novel framework that is capable of capturing the protuberances in the kidneys. Our goal is to segment kidney tumors including isodensity types on NCCT images. To achieve this goal, we create a synthetic dataset, which has separate annotations for normal kidneys and protruded regions, and train a segmentation network to separate the protruded regions from the normal kidney regions. In order to segment whole tumors, our framework consists of three networks. The first is a base network, which extracts kidneys and an initial tumor region masks. The second protuberance detection network receives the kidney region mask as its input and predicts a protruded region mask. The last fusion network receives the initial tumor mask and the protruded region mask to predict a final tumor mask. This proposed framework enables a better segmentation of isodensity tumors and boosts the performance of segmentation of kidney tumors on NCCT images. The contribution of this work is summarized as follows:

1. Present a pioneering work for segmentation of kidney tumors on NCCT images.
2. Propose a novel framework that explicitly captures protuberances in a kidney to enable a better segmentation of tumors including isodensity types on NCCT images. This framework can be extended to other organs (e.g. adrenal gland, liver, pancreas).
3. Verify that the proposed framework achieves a higher dice score compared to the standard 3D U-Net using a publicly available dataset.
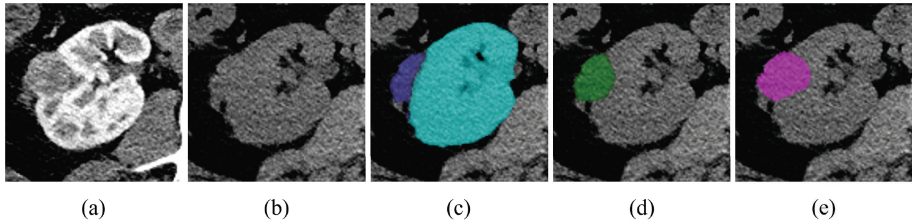
Fig. 1. Example CECT and NCCT images. Some kidney tumors have similar intensity values to its surrounding tissues. a) CECT image. b) NCCT image. c) Output of the protuberance detection network. d) Output of our model. e) Ground truth mask.

2 Related Work

The release of two public CT image datasets with kidney and tumor masks from the 2019/2021 Kidney and Kidney Tumor Segmentation challenge [8] (KiTS19, KiTS21) attracted researchers to develop various methods for segmentation.

Looking at the top 3 teams from each challenge [6,11,13,17,21], all teams utilized 3D U-Net [3] or V-Net [16], which bears a similar architecture. The winner of KiTS19 [13] added residual blocks [7] to 3D U-Net and predicted kidney and tumor regions directly. However, the paper notes that modifying the architecture resulted in only slight improvement. The other 5 teams took a similar approach to nnU-Net's coarse-to-fine cascaded network [12], where it predicts from a low-resolution image in the first stage and then predicts kidneys and tumors from a high-resolution image in the second stage. Thus, although other attempts were made, using 3D U-Net is the go-to method for predicting kidneys and tumors. In our work, we also make use of 3D U-Net, but using this network alone fails to learn some isodensity tumors. To overcome this issue, we developed a framework that specifically incorporates protuberances in kidneys, allowing for an effective segmentation of tumors on NCCT images.

In terms of focusing on protruded regions in kidneys, our work is close to [14,15]. [14] developed a computer-aided diagnosis system to detect exophytic kidney tumors on NCCT images using belief propagation and manifold diffusion to search for protuberances. An exophytic tumor is located on the outer surface of the kidney that creates a protrusion. While this method demonstrated high sensitivity (95%), its false positives per patient remained high (15 false positives per patient). In our work, we will not only segment protruded tumors but also other tumors as well.

3 Proposed Method

To capture the protuberances in kidneys, we specifically train a protuberance detection network, which receives a kidney region mask as an input and separates protruded regions from it. This enables us to extract a part of tumors that forms protuberance, but our goal is segmenting all visible kidney tumors on NCCT images. Thus, we make use of three networks as shown in Fig. 2.
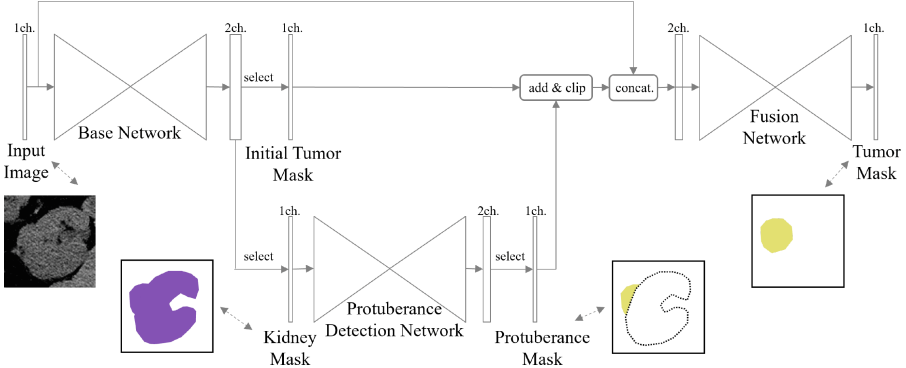
Fig. 2. Overview of our framework.

The first base network is responsible for predicting kidney and tumor region masks. Our architecture is based on 3D U-Net, which has an encoder-decoder style architecture, with few modifications. To reduce the required size of GPU memory, we only use the encoder that has only 16 channels at the first resolution, but instead we make the architecture deeper by having 1 strided convolution and 4 max-pooling layers. In the decoder, we replace the up-convolution layers with a bilinear up-sampling layer and a convolution layer. In addition, by only having a single convolution layer instead of two in the original architecture at each resolution, we keep the decoder relatively small. Throughout this paper, we refer this architecture as our 3D U-Net.

The second protuberance detection network is the same as the base network except it starts from 8 channels instead of 16. We train this network using synthetic datasets. The details of the dataset and training procedures are described in Sect. 3.2.

The last fusion network combines the outputs from the base network and the protuberance detection network and makes the final tumor prediction. In detail, we perform a summation of the initial tumor mask and the protruded region mask, and then concatenate the result with the input image. This is the input of the last fusion network, which also has the same architecture as the base network with an exception of having two input channels. This fusion network do not just combine the outputs but also is responsible for removing false positives from the base network and the protuberance detection network.

Our combined three network is fully differentiable, however, to train efficiently, we train the model in 3 steps.

3.1 Step1: Training Base Network

In the first step, we train the base network, which is a standard segmentation network, to extract kidney and tumor masks from the images. We use a sigmoid function for the last layer. And as a loss function, we use the dice loss [16] and the cross-entropy loss equally.

3.2   Step2: Training Protuberance Detection Network

In the second step, we train the protuberance detection network alone to separate protruded regions from the normal kidney masks. Here, we only use the cross-entropy loss and label smoothing with a smoothing factor of $\epsilon = 0.01$.

Synthetic Dataset. To enable a segmentation of protruded regions only, a separate annotation of each region is usually required. However, annotating such areas is time-consuming and preparing a large number of data is challenging. Alternatively, we create a synthetic dataset that mimics a kidney with protrusions. The synthetic dataset is created through the following steps:

1. Randomly sample a kidney mask without protuberance and a tumor mask.
2. Apply random rotation and scaling to the tumor mask.
3. Randomly insert the tumor mask into the kidney mask.
4. If both of the following conditions are met, append to the dataset.

$$\frac{\sum_i k_i t_i}{\sum_i k_i} < 0.3, \tag{1}$$

$$\frac{\sum_i k_i t_i}{\sum_i t_i} < 0.95, \tag{2}$$

where k_i is a voxel value (0 or 1) in the kidney mask and t_i is a voxel value in the tumor mask. Equation 1 ensures that only up to 30% of the kidney is covered with a tumor. Equation 2 ensures that not all tumors are covered by the kidney (at least 5% of the tumor is protruded from the kidney).

3.3   Step3: End-to-End Training with Fusion Network

In the final step, we train the complete network jointly. Although our network is fully differentiable, since there is no separate annotation for protruded regions other from the synthetic dataset, we freeze the parameters in protuberance detection network.

The output of the protuberance detection network will likely have more false positives than the base network since it has no access to the input image. Thus, when the output of the protuberance detection network is concatenated with the output of the base network, the fusion network can easily reduce the loss by ignoring the protuberance detection network's output, which is suboptimal. To avoid this issue, we perform summation not concatenation to avoid the model from ignoring all output from the protuberance detection network. We then clip the value of the mask to the range of 0 and 1. As a result, the input to the fusion network has two channels. The first channel is the input image, and the second channel is the result of summation of the initial tumor mask and the protruded region mask. We concatenate the input image so that the last network can remove false positives from the predicted masks as well as predicting the missing tumor regions from the protuberance detection network.

We use the dice loss and the cross-entropy loss as loss functions for the fusion network. We also keep the loss functions in the base network for predicting kidneys and tumors. The loss function for tumors in the base network acts like an intermediate supervision. Our network shares some similarities with the stacked hourglass network [18] where the network consists of multiple U-Net like hourglass modules and has intermediate supervision at the end of each hourglass module. By having multiple modules in this manner, the network can fix the initial mistakes in early modules and corrects in later modules.

4 Experiments

No prior work exists that uses NCCT images from KiTS19 [8,9]. Thus, we first created our baseline model and compared the performance with existing methods on CECT images. This allows us to ensure that our baseline model and training procedure is appropriate. We then trained the model using NCCT images and compared with our proposed method.

4.1 Datasets and Preprocessing

We used a dataset from KiTS19 [8] which contains both CECT and NCCT images. For CECT images, there are 210 images for training and validation and, 90 images for testing. For NCCT images, there are 108 images, which are different series of the 210 images. The ground truth masks are only available for the 210 CECT images. Thus, we transfer the masks to NCCT images. This is achieved by extracting kidney masks and adjusting the height of each kidney. The ground truth mask contains a kidney label and a kidney tumor label. Cysts are not annotated separately and included in the kidney label on this dataset. The data can be downloaded from The Cancer Imaging Archive (TCIA) [4,9].

The images were first clipped to the intensity value range of $[-90, 210]$ and normalized from -1 to 1. The voxel spacings were normalized to 1 mm. During the training, the images were randomly cropped to a patch size of $128 \times 128 \times 128$ voxels. We applied random rotation, random scaling and random noise addition as data augmentation.

During the *Step*2 phase of the training, where we used the synthetic dataset, we created 10,000 masks using the method from Sect. 3.2. We applied some augmentations during training to input masks to simulate the incoming inputs from the base network. The output of the base network is not binarized to keep gradient from flowing, so the values are in the range [0, 1] and the edge of kidneys are usually smooth. Therefore, we applied gaussian blurring, gaussian noise addition and intensity value shifting.

Table 1. Dice performance of existing method and our baseline model. Evaluated using CECT images from KiTS19. Composite dice is an average dice between kidney and tumor dice. The results were obtained by submitting our predicted masks to the Grand Challenge page.

Method	Composite Dice	Kidney Dice	Tumor Dice
Isensee and Maier-Hein [13]	0.9123	0.9737	0.8509
Our baseline model	0.8832	0.9728	0.7935

4.2   Training Details and Evaluation Metrics

Our model was trained using SGD with a 0.9 momentum and a weight decay of $1e-7$. We employed a learning rate scheduler, which we warm-up linearly from 0.0001 to 0.1 during the first 30% (for $Step1$ and $Step3$) or 10% (for $Step2$) of the total training steps and decreased following the cosine decay learning rate. A mini-batch size of 8, 16 and 4 were used, and trained for 250k, 100k and 100k steps during $Step1$ to 3 respectively. We conducted our experiments using JAX (v.0.4.1) [2] and Haiku (v.0.0.9) [10]. We trained the model using a single NVIDIA RTX A5000 GPU.

For the experiment on CECT images, we used the dice score as our evaluation metrics following the same formula from KiTS19. For the experiment on NCCT images, we also evaluated the sensitivity and false positives per image (FPs/image). We calculated as true positive when the predicted mask has the dice score greater than 0.5, otherwise we calculated as false negative. On the other hand, false positives were counted when the predicted mask did not overlap with any ground truth masks.

5   Results

5.1   Performance on CECT Images

To show that our model is properly tuned, we compare our baseline model with an existing method using CECT images. As can be seen from Table 1, our model showed comparable scores to the winner of KiTS19 challenge. We used this baseline model as our base network for the experiments on NCCT images.

5.2   Performance on NCCT Images

Table 2 shows our experimental results and ablation studies on NCCT images. The proposed method (Table 2-bottom) outperformed the baseline model (Table 2-top). The ablation studies show that adding each component (CECT images and the protuberance detection network) resulted in an increase in the performance. While adding CECT images contributed the most for the increase in tumor dice and sensitivity, adding the protuberance detection network further pushed the performance. However, the false positives per image (FPs/image)

Table 2. Result of our proposed method on NCCT images from KiTS19. The values are average values of a five-fold cross-validation.

Protuberance Detection Network	with CECT images	Tumor Dice	Sensitivity	FPs/Image
✗	✗	0.518	0.618	**0.283**
✗	✓	0.585	0.686	0.340
✓	✓	**0.615**	**0.721**	0.421
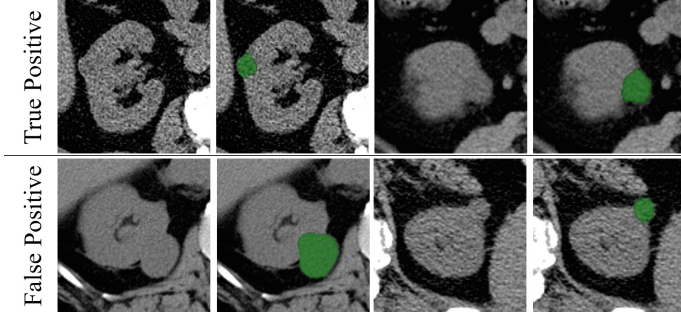
Fig. 3. Output examples from our model. The first row shows true positive results and the second row shows false positive results.

increased from 0.283 to 0.421. The protuberance detection network cannot distinguish the protrusions that were caused by tumors or cysts, so the output from this network has many FPs at this stage. Thus, the fusion network has to eliminate cysts by looking again the input image, however, it may have failed to eliminate some cysts (Fig. 3 second row).

6   Conclusion

In this paper, we proposed a novel framework for kidney tumor segmentation on NCCT images. To cope with isodensity tumors, which have similar intensity values to their surrounding tissues, we created a synthetic dataset to train a network that extracts protuberance from the kidney masks. We combined this network with the base network and fusion network. We evaluated our method using the publicly available KiTS19 dataset, and showed that the proposed method can achieve a higher sensitivity than existing approach. Our framework is not limited to kidney tumors but can also be extended to other organs (e.g., adrenal gland, liver, pancreas).

References

1. Estimated number of new cases in 2020, world, both sexes, all ages (excl. nmsc). https://gco.iarc.fr/today/online-analysis-table/. Accessed 27 Feb 2023

2. Bradbury, J., et al.: JAX: composable transformations of Python+NumPy programs (2018). http://github.com/google/jax

3. 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.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49

4. Clark, K., et al.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging **26**(6), 1045–1057 (2013)

5. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.: Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In: Proceedings of International Conference on Learning Representations (2019)

6. Golts, A., Khapun, D., Shats, D., Shoshan, Y., Gilboa-Solomon, F.: An ensemble of 3D U-net based models for segmentation of kidney and masses in CT scans. In: Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Papanikolopoulos, N., Weight, C. (eds.) KiTS 2021. LNCS, vol. 13168, pp. 103–115. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-98385-7_14

7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

8. Heller, N., et al.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the kits19 challenge. Med. Image Anal. **67**, 101821 (2021)

9. Heller, N., et al.: C4KC kits challenge kidney tumor segmentation dataset (2019)

10. Hennigan, T., Cai, T., Norman, T., Babuschkin, I.: Haiku: Sonnet for JAX (2020). http://github.com/deepmind/dm-haiku

11. Hou, X., Chunmei, X., Li, F., Yang, N.: Cascaded semantic segmentation for kidney and tumor. Submissions to the 2019 Kidney Tumor Segmentation Challenge: KiTS19 (2019)

12. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2020)

13. Isensee, F., Maier-Hein, K.: An attempt at beating the 3D U-NET. Submissions to the 2019 Kidney Tumor Segmentation Challenge: KiTS19 (2019)

14. Liu, J., Wang, S., Linguraru, M.G., Yao, J., Summers, R.M.: Computer-aided detection of exophytic renal lesions on non-contrast CT images. Med. Image Anal. **19**(1), 15–29 (2015)

15. Liu, J., Wang, S., Yao, J., Linguraru, M.G., Summers, R.M.: Manifold diffusion for exophytic kidney lesion detection on non-contrast CT images. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8149, pp. 340–347. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40811-3_43

16. 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), pp. 565–571. IEEE (2016)

17. Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y.: Segmentation of kidney tumor by multi-resolution VB-Nets. Submissions to the 2019 Kidney Tumor Segmentation Challenge: KiTS19 (2019)

18. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 483–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_29

19. Pinsky, P.F., et al.: Incidental renal tumours on low-dose CT lung cancer screening exams. J. Med. Screen. **24**(2), 104–109 (2017)

20. Touijer, K., et al.: The expanding role of partial nephrectomy: a critical analysis of indications, results, and complications. Eur. Urol. **57**(2), 214–222 (2010)

21. Zhao, Z., Chen, H., Wang, L.: A coarse-to-fine framework for the 2021 kidney and kidney tumor segmentation challenge. In: Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Papanikolopoulos, N., Weight, C. (eds.) KiTS 2021. LNCS, vol. 13168, pp. 53–58. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-98385-7_8

22. Znaor, A., Lortet-Tieulent, J., Laversanne, M., Jemal, A., Bray, F.: International variations and trends in renal cell carcinoma incidence and mortality. Eur. Urol. **67**(3), 519–530 (2015)