Kidney Tumor Segmentation

Hosein Beheshtifard
Computer Science
University of Calgary
Calgary, Canada
hosein.beheshtifard@ucalgary.ca

ABSTRACT

Kidney cancer is one of the most common malignant tumors in the world. Manual segmentation of kidney tumors is time-consuming and subjective, as it relies on the expertise of radiologists and surgeons to interpret the imaging data. Therefore, there is a critical need for a more efficient and precise method for segmenting kidney tumors from CT images. Deep learning models have shown promise in medical image analysis, and can potentially provide objective and accurate segmentation of kidney tumors. In this project, the aim is to segment kidneys and tumors using 2D slices of 3D CT images, which is easier to implement on cheaper hardware and can be utilized by individuals worldwide. we developed a U-Net base model with ResNet34 encoders. We trained and tested our model using the KITS21 dataset, which included 300 patients who underwent partial or radical nephrectomy for suspected renal malignancy. Our model achieved an average dice score of 0.961 for kidney segmentation and 0.811 for tumor segmentation. This performance is comparable to the top winners in the KITS19 and KITS21 challenges, who used 3D segmentation with heavy computations.

KEYWORDS

Medical Imaging, Kidney Tumors, Semantic Segmentation, Deep Learning, U-Net model

1 Introduction

Kidney cancer is the 13th most common cancer worldwide, accounting for 2.4% of all cancers, with more than 330,000 new cases diagnosed yearly, and its incidence is still increasing [1]. The most common type of kidney cancer is renal cell carcinoma, which accounts for about 90% of all cases [2]. The treatment of renal cell carcinoma often involves surgical removal of the tumor, which requires accurate preoperative planning to ensure complete removal of the tumor while minimizing damage to the surrounding healthy tissue. Therefore, accurate segmentation of the tumor and surrounding anatomy from CT images is crucial for effective treatment and better patient outcomes.

Manual segmentation of kidney tumors is time-consuming and subjective, as it relies on the expertise of radiologists and surgeons to interpret the imaging data. Automated segmentation methods have the potential to overcome these limitations and provide objective and standardized measurements of tumor size, shape, and appearance. Over the years, different methods have been proposed to improve the accuracy and efficiency of automated kidney tumor segmentation.

One of the earliest approaches to kidney tumor segmentation was region growing, which involves selecting a seed point within the tumor and growing a region around it that satisfies certain criteria, such as intensity homogeneity. While this method is simple and fast, it is highly dependent on the choice of seed point and can lead to oversegmentation or undersegmentation of the tumor.

Another commonly used method is thresholding, which involves selecting a threshold value for the image intensity that separates the tumor from the surrounding tissue. This method is fast and easy to implement, but it is highly dependent on the choice of threshold value and can lead to inaccurate segmentation when the tumor and surrounding tissue have similar intensity values.

More recently, machine learning algorithms, such as convolutional neural networks (CNNs), have been developed to improve the accuracy and efficiency of kidney tumor segmentation. CNNs are a type of deep learning algorithm that can learn features from raw input data, such as CT images, and use them to make accurate predictions. CNN-based segmentation methods have shown promising results in various medical imaging applications, including kidney tumor segmentation.

In this project, we aim to use a U-Net base model to segment tumors existing in the kidney. Since analyzing 3D CT images requires heavy computations and expensive GPUs, we chose to implement a 2D segmentation using 2D slices of volumetric images.

2 Related Work

The 2019 Kidney Tumor Segmentation Challenge (KiTS19) was organized in parallel with the 2019 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) with the aim of addressing the challenges of automated segmentation and promoting progress in this area. The competition provided a training set consisting of 210 cross-sectional CT images with kidney tumors, along with corresponding semantic segmentation masks, which were made publicly available. A total of 106 teams from five continents utilized this data to develop automated systems capable of predicting the true segmentation masks for a private test set of 90 CT images. These predictions were evaluated and ranked based on their average Sørensen-Dice coefficient for the kidney and tumor across all 90 cases. The winning team utilized a 3D U-Net model and achieved a Dice coefficient of 0.974 for the kidney and 0.851 for the tumor [3].

The KITS21 challenge is the latest version of the Kidney and Kidney Tumor Segmentation challenge, building upon the success of the KITS19 challenge. As with its predecessor, KITS21 was held in conjunction with the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) and aimed to address the challenge of automatic segmentation of kidneys and tumors in CT images. The challenge provided a training set of 300 CT scans, including 210 scans from the KITS19 challenge and an additional 90 scans with corresponding semantic segmentation masks for the kidney and tumor regions. Participants were tasked with developing deep learning models capable of accurately predicting the segmentation masks for a private test set of 100 CT scans. The performance of the teams was assessed using the Sørensen-Dice coefficient, a common metric for evaluating segmentation accuracy. The winning team achieved an impressive Dice coefficient of 0.975 for kidney segmentation and 0.86 for tumor segmentation by using a coarse-to-fine framework which is based on the nnU-Net.

3 Dataset

The KiTS21 cohort includes patients who underwent partial or radical nephrectomy for suspected renal malignancy between 2010 and 2020 at either an M Health Fairview or Cleveland Clinic medical center. A retrospective review of these cases was conducted to identify all patients who had undergone a contrast-enhanced preoperative CT scan that includes the entirety of all kidneys.

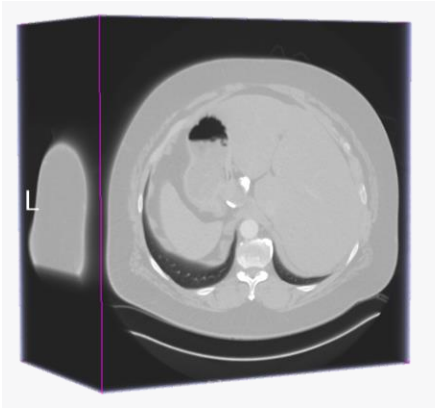

Figure 1: A patient's original abdominal CT image

Each case's most recent corticomedullary preoperative scan was independently segmented three times for each instance of Kidney, Tumor, and Cyst.

The annotation process involves placing 3D bounding-boxes around each region of interest, annotating axial slices within these bounding boxes with "guidance pin annotations," and reviewing the annotations by experts and trainees. Laypeople then perform "contour annotations" based on the approved guidance, and the annotations are post-processed to generate segmentations [3].

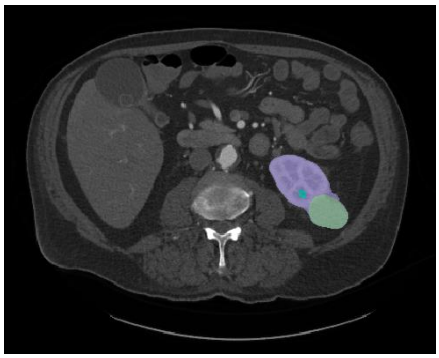

Figure 2: An example of a segmented axial slice from KITS21 dataset

The dataset for this project comprises data for 300 patients, along with segmentation masks for both kidney and tumor regions. While cyst data was not utilized in this study, their similarity to tumors can potentially impact the accuracy of tumor segmentation predictions.

4 Methodology

This segmentation process will be performed in two phases. In the first phase, we will segment the kidney from the CT scan images, and in the second phase, we will segment the tumors existing in the kidney.

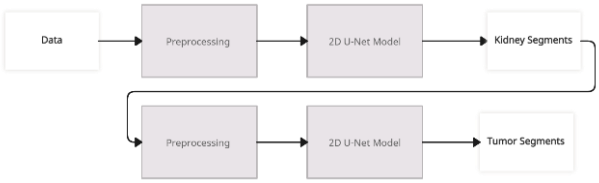

Figure 3: Project pipeline

For both segmentation phases, we will employ U-Net based models that will be optimized for performance. In order to speed up the segmentation process and reduce the computational burden, we will design a 2D segmentation model.

4.1 Preprocessing

The first step is to convert the 3D images and their related masks into 2D slices.

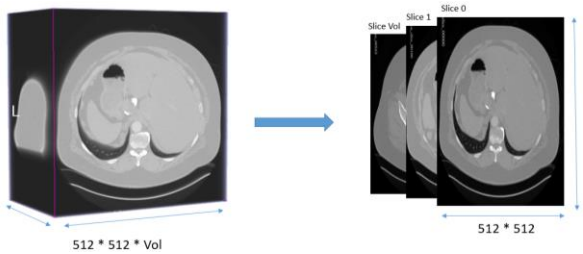

Figure 4: Convert 3D CT image to 2D slices

It is worth noting that during the first phase of the segmentation, we ignored tumors and cysts as we were only interested in segmenting the kidneys. In the second phase, we solely used tumor masks.

In the second step, we divided the data into three sets: train, validation, and test, with proportions of 80%, 10%, and 10%, respectively. So we have 240 patient data for training, 30 for validation, and 30 cases for testing that we did not feed the model with during both phases of training.

As part of our preprocessing, we changed the intensities based on a range of 0 to 200, which is an approximate range for the kidneys and some soft tissues. Any intensities greater than the maximum value were set to the maximum, and any intensities lower than the minimum value were set to the minimum. This allowed us to disregard other organs with different intensity ranges such as bones or metals. This transformation was only applied during the first phase since we only dealt with kidney segments during the second phase. Additionally, the data was normalized using the training dataset's mean and standard deviation.

Lastly, we added some random augmentations such as random rotation, random flip, random crop, and random Gaussian noise. The aim of these transformations is to increase the variability of the data to avoid overfitting.

4.2 Model Architecture

The chosen model architecture is a U-Net base model. The encoder path is based on a pre-trained ResNet34 that extracts feature maps at different scales. The decoder path generates accurate segmentation masks by combining information from different scales. The U-Net architecture consists of 23 convolutional layers in the encoder and decoder, excluding the final 1x1 convolutional layer, with skip connections between corresponding encoder and decoder blocks. The encoder path gradually reduces the spatial resolution of the input image and increases the number of feature channels. The decoder path gradually upsamples the feature maps to the original input size to produce a segmentation map. Each block in the encoder consists of two convolutional layers followed by a max-pooling layer, while each block in the decoder comprises an upsampling layer and a convolutional layer. The number of feature channels is doubled after each downsampling step and halved after each upsampling step, allowing the network to learn increasingly abstract representations of the input image and gradually refine the segmentation map. The skip connections concatenate feature maps from the corresponding encoder block with the upsampled feature maps in the decoder, which helps the network capture fine-grained details while maintaining information.

We used pre-trained weights on imageNet for our final model. Although we have only one input channel, it resulted in faster convergence.

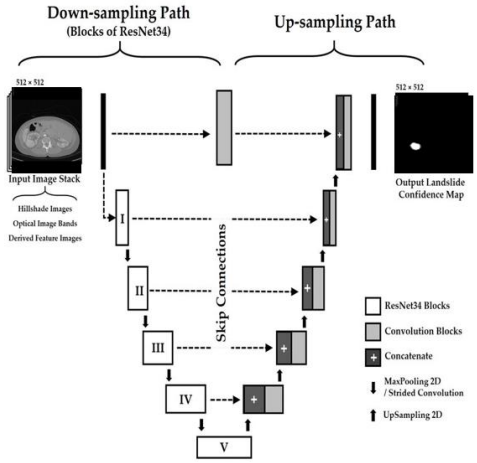

Figure 5: U-net architecture with ResNet34 blocks in the down-sampling path[4].

4.3 Model Training

For the loss function, we used BCEWithLogitsLoss, which is a commonly used loss function in segmentation tasks due to its ability to handle binary classification problems and its ease of use in combination with sigmoid activation functions. Furthermore, we utilized Adam as the optimizer with a learning rate of 0.001.

For training, we created dataloaders with a batch size of 128, and trained our model per 100 epochs. However, we only saved the best model regarding validation loss to avoid overfitting.

We followed the same approach for both phases of the segmentation. During the first phase, we predicted the kidney mask and then multiplied it by the actual CT image to segment the kidneys. Then, in the second phase, we used the kidney segments to locate tumors.

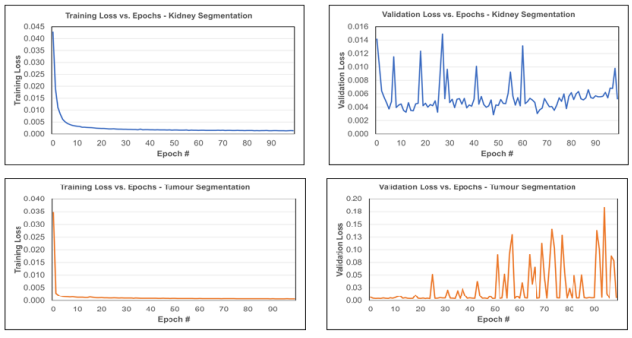

Figure 6: Training and Validation Loss in Kidney Segmentation (Phase 1) and Tumor Segmentation (Phase 2).

4.3 Postprocessing

One postprocessing technique we employed involved disregarding areas of predicted tumors that were outside of the kidneys.

However, this approach did not yield a significant improvement in the results.

Finally, we construct a 3D tumour segment by combining the 2D segments.

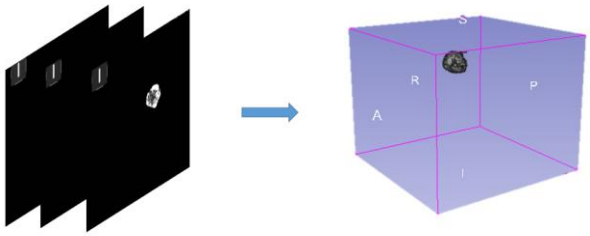

Figure 7: Convert 2D segment slices to 3D image.

5 Results

5.1 Evaluation Metric

We used the dice coefficient metric to determine the overlap between the predicted segmentation mask and the ground truth mask. The Dice coefficient ranges from 0 to 1, where a value of 1 indicates a perfect overlap between the two sets, and a value of 0 indicates no overlap. is particularly useful for segmentation tasks because it takes into account both the true positive (TP) and false positive (FP) rates of the segmentation algorithm.

$$\text{Dice Score} = \frac{2\times(O\cap E)}{O\cup E} \quad (1)$$

Where O is the observed segmentation mask and E is the expected segmentation mask.

In the second phase, we only calculated the dice score for images that included kidneys.

5.2 Results on test set

We present the quantitative results in Table 1, which are based on the test set comprising 30 cases. Additionally, we showcase the visual outcome of our final predicted mask and the corresponding ground truth mask in Figure 9.

Initially, we utilized a simple U-Net model comprising of 5 layers, which yielded a dice score of 0.736 for the kidney segments. Since we did not expect to obtain a better dice score for tumor segmentation, we did not proceed with this model for the second phase.

Subsequently, we employed a proposed U-Net model with ResNet34 blocks, where the first run implemented a strict intensity range of [20, 45]. In the second iteration, we expanded this range to [0, 200] and augmented the dataset further by introducing random cropping and random Gaussian noise. These modifications resulted in a significant improvement, with the final dice score of 0.961 for kidney segmentation and 0.811 for tumor segmentation.

Phase	Average Dice Score		
	Simple U-Net with 5 Layers	Our model(first run)	Our model(after optimization)
Kidney Segmentation	0.736	0.932	0.961
Tumor Segmentation	-	0.793	0.811

Table 1: Dice scores for tumor and kidney segmentation in two iterations of model compared with baseline.

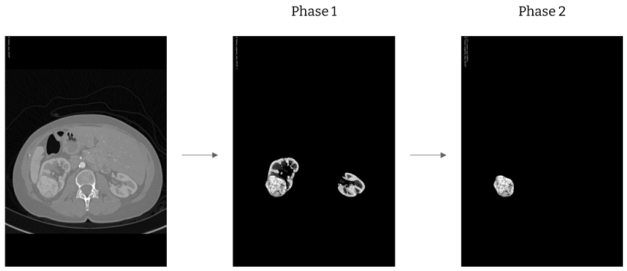

Figure 8: Progression of coarse-to-fine segmentation phases

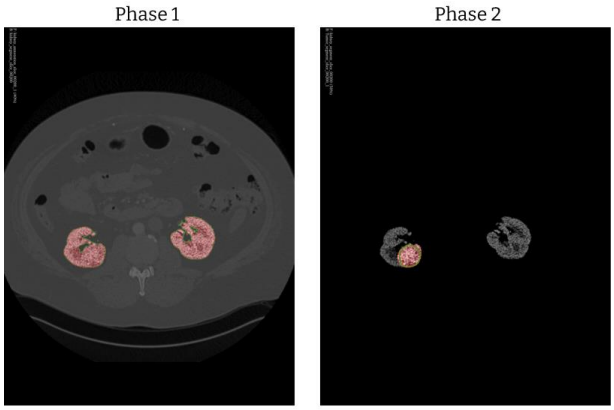

Figure 9: The predicted kidney and tumor mask shown in red, and the ground truth masks with green edges.

6 Discussion and Conclusion

In this study, we used a 2D U-Net model to segment the kidney and tumors from CT images. A basic U-Net architecture served as the baseline model, which was subsequently enhanced using ResNet34 encoders, transfer learning, and random augmentations. Regarding the simplicity of 2D segmentation, our result is promising when compared with top performers in KITS19 and KITS21. However, there is a limitation to focusing on 2D slices, as it is harder to recognize the shape and edges of a tumor. Generally, the Dice of the kidney is very high, but the segmentation of tumors demonstrated insufficient accuracy, highlighting the need for further improvement.

In the following works, we want to segment CT images from three anatomical planes: sagittal, coronal, and axial, and combine results at the end. One important following step is to synthesize new annotated data with the available dataset to create a larger, more diverse dataset that will enable us to attain even better results.

ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Farhad Maleki who guided me in this project.

REFERENCES

[1] Ghislaine Scelo and Tricia L Larose. Epidemiology and risk factors for kidney cancer. Journal of Clinical Oncology, 36(36):3574, 2018.
[2] Pandey J, Syed W. Renal Cancer.
[3] https://kits21.kits-challenge.org
[4] Prakash N, Manconi A, Loew S. Mapping landslides on EO data: Performance of deep learning models vs. traditional machine learning models. Remote Sensing. 2020 Jan 21;12(3):346.