Kidney Tumor Detection using Attention based U-Net

Prabod Rathnayaka[1], Vinoj Jayasundara[1], Rashmika Nawaratne[1], Daswin De Silva[1], Weranja Ranasinghe[1], and Damminda Alahakoon[1]

Research Centre for Data Analytics and Cognition, La Trobe University, Bundoora, Australia

Abstract. The advancement of deep learning techniques has provoked the potential of using Medical Image Analysis (MIA) for disease detection and prediction in numerous ways. This has been mostly useful in identifying tumours and abnormalities in many organs of the human body. Particularly in kidney diseases, the treatment options such as surgery have largely benefitted by the ability to detect tumours in early stages, thereby shifting towards more efficient methods including conservative nephron procedures. Therefore, to enable the early detection of kidney tumours, we propose a convolutional neural network based U-Net architecture which is able to detect tumours using an attention mechanism. The proposed architecture was evaluated using KiTS19 Challenge dataset that includes a collection of multi-phase CT imaging, segmentation masks, and comprehensive clinical outcomes for 300 patients who underwent nephrectomy for kidney tumours. The outcomes demonstrate the ability of the proposed architecture to distinguish images with tumours in the kidney and support early tumour detection.

Keywords: Kidney Tumour detection · Deep Learning · CNN.

1 Introduction

In the recent few years utilization of machine learning and deep learning techniques for the advancement of the Medical Image Analysis (MIA) has been proliferating. Many facets of the MIA has been using machine learning and deep learning techniques such as disease prediction in selected organs (e.g., brain, kidney, prostate, and spine), skin cancer detection, knee osteoarthritis diagnosis.

More than 400,000 new cases of kidney cancer surfaces each year. Surgery is considered as the most prevalent treatment for kidney cancer.Radical Nephrectomy (RN), was standard of care for kidney tumors, which is removal of both the tumor and the affected kidney. However advancements in surgery in conjunction with earlier tumor detection have precipitated a significant shift in kidney cancer treatment toward more conservative nephron sparing procedures, called Partial Nephrectomies (PNs). Therefore accurate and efficient methods of tumour identification are essential to decide between two treatment methodologies. Automatic semantic segmentation is a promising tool for these efforts.

KiTS 19 Challenge was proposed to accelerate the research and development of new nephrometric features to aid in prognosis and treatment planning for kidney tumors,and to enable the creation of reliable learning based kidney and kidney tumor semantic segmentation methods which will allow the features developed to be automated and applied at an unprecedented scale.

2 Proposed Architecture

We propose an attention based U-Net architecture [1] for kidney tumor detection. U-Net architecture is based on the concepts of Fully Convolutional Networks (FCNs) [2], systematically modified to achieve better performance as illustrated in Fig. 1. FCNs are the first end-to-end trainable semantic segmentation networks, which can produce spatial feature maps essential for dense prediction by means of convolutionalization, in addition to offering a substantial computational speedup.

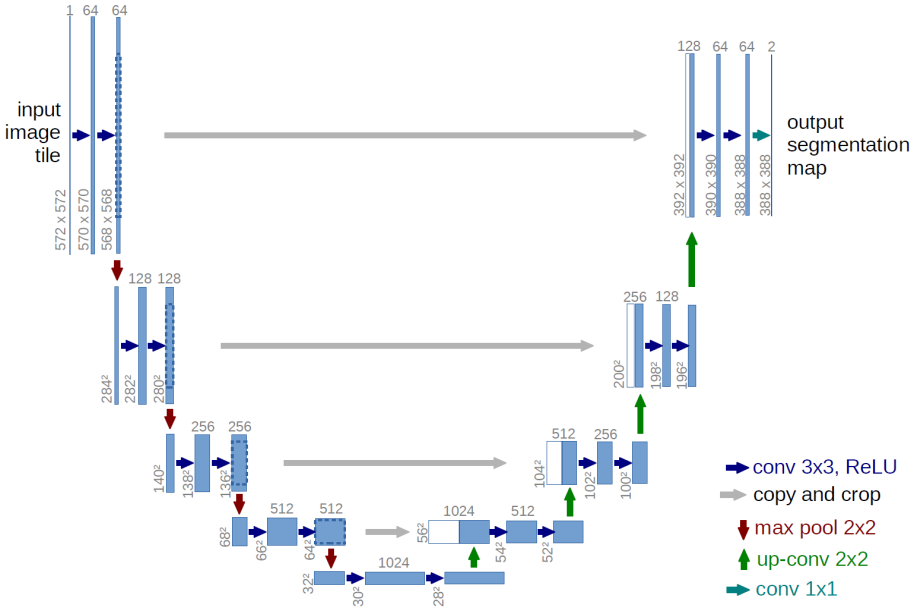

Fig. 1. U-Net Architecture

The intuition behind U-Net is to project a target image in to a latent space which learns nuanced feature mappings, and to reconstruct the corresponding mask back from the latent representation. Hence, as evident from Fig.1, the U-Net architecture consists of a downsampling path followed by a bottleneck section and an upsampling path, structured in a symmetric U-shape.

The downsampling path consists of several blocks. each with two convolution layers followed by a pooling layer. Subsequent to downsampling, the bottleneck portion of the network commences with two convolutional layers. The output of the bottleneck is then fed to an upsampling layer, and the result is channel-wise concatenated (skip-connected) with the convolutional output of the corresponding block in the downsampling path. Every upsampling block is skip-connected to the corresponding downsampling block accordingly. The excessively large number of feature maps used in the upsampling path facilitate the propagation of context information to the higher resolution layers.

2.1 Attention Mechanism

Recent work [3] suggests that the use of attention mechanisms with the U-Net architecture enhances prediction performance. Trainable attention mechanisms learn to choose a subset of the input features, while suppressing the rest, which are relevant to the task at hand. Hence, this aids the network to conveniently localize the target, kidney and tumour in our case.

We use attention gates integrated in to the U-Net architecture, which learns an attention coefficient per pixel of the input feature map. For the n^{th} feature map pixel of the l^{th} layer $x_{n,l}$, the resulting output after attention can be calculated by the element-wise multiplication $\kappa_{n,l} \cdot x_{n,l}$, where $\kappa_{n,l} \in [0, 1]$ is the trainable attention coefficient corresponding to $x_{n,l}$. Hence, the coefficient attenuates and prunes the input feature map pixels that are irrelevant to the task at hand.

2.2 Loss Function

We use the Sorensen-Dice loss [4] as the loss function for this implementation, as defined by,

$$Sorensen - Dice\ Loss = \frac{2 \sum_{n=1}^{N} p_n g_n}{\sum_{n=1}^{N} {p_n}^2 + \sum_{n=1}^{N} {g_n}^2} \tag{1}$$

where N is the total number of pixels, p_n is the n^{th} pixel of the predicted segmentation mask and g_n is the n^{th} pixel of the true segmentation mask.

2.3 Training Procedure

Due to the excessive class imbalance between the background class and the kidney/tumour classes, we downsampled the training examples containing only the background class to a randomly selected 2%, while preserving all the training examples containing either of the kidney and tumour classes.

For the training procedure, we used the Adam optimizer [5] with an initial learning rate of 0.001, which was reduced accordingly by monitoring the validation loss. To further account for the class imbalance, specifically the deficit of the tumour class, we used augmentation on the fly and model re-training. Image augmentation on the fly perturbs the training samples to add random horizontal shifts, flips and etc, such that each training sample in a given batch becomes unique across all the epochs, in an attempt to avoid the model overfitting to the dominant classes. Further, we re-train the model with the tumour sub-dataset to fine-tune the final segmentation layers, while freezing the shallow layers, to achieve fine-grained segmentation of the tumour class.

3 Experiments and Results

For the training process we set aside 20% of the cases for validation of the model. which is 42 cases for validation and 168 cases for training. Some of the model variants with results are shown in table 1

Model	**Train Kidney Dice**	**Train Tumor Dice**	**Val Kidney Dice**	**Val Tumor Dice**
UNet	0.89	0.80	0.81	0.37
UNet- Attention	0.96	0.90	**0.90**	**0.57**

Table 1. Model variants with Dice Score

4 Conclusion

In this paper, we proposed a convolutional neural network based U-Net architecture that is able to detect segments from medical imagery data using an attention mechanism. This enables early detection of tumours and abnormalities in many organs of the human body. The proposed model architecture is demonstrated and evaluated for kidney tumour detection use-case using KiTS19 challenge dataset, and the results validated the accuracy and robustness of the proposed model. Furthermore, we intend to extend the post-processing potential of the proposed model and modify the model using 3-dimensional convolutional neural network layers as our future work.

Bibliography

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[3] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[4] F. Milletari, N. Navab, and S.-A. Ahmadi, "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.

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.