

WP-UNet: Weight Pruning U-Net with Depthwise Separable Convolutions for Semantic Segmentation of Kidney Tumors

Patike Kiran Rao (✉ kiranraocse@gpcet.ac.in)

M S Ramaiah University of Applied Sciences

Dr Subarna Chatterjee

M S Ramaiah University of Applied Sciences

Research Article

Keywords: UNet, WP-UNet, Convolutional, Depthwise Separable Convolutions, Normalization, CT scan Image

Posted Date: January 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-140504/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background:

The major challenge in medical imaging is to achieve high accuracy output during semantic image segmentation tasks in biomedical imaging while having fewer computational operations and faster inference. It is necessary in medical modalities such as kidney tumor CT scan activities, to assist radiologists. Several previous studies have carried out a complex deep network that requires high computational resources. However, a deep network on semantic segmentation of kidney tumor CT scans with fewer flops and parameters has not yet been evaluated.

Methods:

This research paper presents a novel network model called Weight Pruning U-Net (WP-UNet) which is extremely fast, compact, and computationally efficient to address this problem with kidney tumor CT scan images as an application.

Results

We apply the proposed deep network model on the kidney tumor CT scan image dataset on computational devices with limited resources for computing. We build a CNN model with minimum parameters inspired by the commonly adapted U-Net architecture of the deep convolution neural network model for CT scan image analysis by making use of a depthwise separable convolution functional layer in the entire network model. We proposed weight pruning with the depthwise separable and batch normalized UNet model to reach the expected performance and reduce the loss in the process. WP-UNet has 3 major benefits,- : (a) a lightweight model with a smaller size (b) fewer parameters, and (c) a faster assumption time with a less than floating point calculation with computational complexity (FLOPs). WP-UNet was tested on the KiTs challenge Biomedical CT Scan imaging Dataset for kidney tumor semantic segmentation (KiTs), and the results showed that comparable and often better results were obtained by the WP-UNet model compared to the existing state-of-the-art models.

Conclusions:

Unlike previous assumptions, our findings indicate that the architecture proposed is smaller than U-Net and demands 3x less computational complexity while retaining respectable accuracy results. Moreover it affects kidney tumor medical image analysis and their practical application.

1. Introduction

With more than 400,000 new cases in 2020, kidney cancer [25] accounts for 2.4% of cancers worldwide [1], with significant differences in incidence rates depending on geography, race, sex and time [2]. The increasing use of medical imaging has had two consequences in the last two decades: the size of the

renal tumor diagnosed has steadily decreased [3], and the number of localized renal masses detected incidentally has increased [4].

In computer assisted diagnostic systems that are helpful in allowing clinical diagnosis of kidney tumors, CT scan image segmentation [20] plays a central role. Segmenting the region of interest, such as an organ or tumor volume, the location of the region of interest has been an increasing demand for accurate, fast and cost effective automated processing medical image analysis equipment, such as computer tomography (CT scan). Automated processes can not only save time and costs, but also reduce the reliance on manual labor (radiographer) and human error.

There has recently been a growing need to incorporate deep learning solutions on devices equipped with low computational complexity, such as handheld computers or other low budget computing machines. The major problem of complex deep network models is that they are more parameterized which requires substantial computational equipment for testing. In-depth learning of several methods, including pruning [14][27] or quantification of weights has been suggested by researchers. Pretrained models on large medical modality datasets [20]. Others have concentrated on lightweight preparation. By factorizing regular convolution layers into depthwise separable convolution layers [21] from scratch [7-11] for less parameterized computational layers.

A similar methodology using these lightweight designs is introduced in this article, also known as mobile network architectures with the objective of improving the U-Net model with fewer parameters [23] that require less disk capacity, fewer numerical specifications, and faster inference. However, in terms of accuracy, separable convolutions are considered to have degraded efficiency relative to layers of regular convolution. Therefore, weight pruning [30] coupled with batch normalization [22] is implemented on each input layer's weights to restore the lack of precision. This architecture is fresh It is known as WP-UNet. WP-UNet performance is measured on the KiTs challenge dataset.

2. Related Work

In this section, we describe in detail the literature works that inspired our study.

2.1 Depthwise Separable Convolutions

For two reasons, deeply wise separable convolutions have recently become common in biomedical and other models:

- They have fewer parameters and are thus less likely to overfit than "regular" convolution layers.
- They also require fewer operations to compute with fewer parameters, and are thus cheaper and quicker.

The difference between the number of filters in normal convolutions and depthwise separable convolutions[21] is as follows:

- **Spatial dimension (S)** - Width and height, with square inputs assumed,
- **Filter layer (F)** - Width and height of the filter, square filter believed,
- **Input Channel (InC)** - Input channel count,
- **Output Channel (OutC)** - Number of channels of output.

In a regular convolution(Fig. 1) there are $F \times F \times InC \times OutC$ parameters, because every filter is 3D and there is one such filter per output channel.

In depthwise separable convolution[21](Fig. 2)there are $F \times F \times InC$ parameters for depthwise convolution [21] and $InC \times OutC$ parameters for the mixing part. The sum of these two parameters is less than the regular convolutions.

Using depthwise convolutions[21] some of the deep network models are able to reduce the computational process by 8 or 9 times compared to standard convolutions.

2.2 Normalization

In deep network models, normalization techniques (Fig. 3) are used to decrease our network model's training time by a huge factor. The major benefit of, normalizing is that it normalizes each feature so that they maintain the contribution of every feature, as some features have higher numerical values than others. It also reduce internal covariate shift. It makes the optimization faster because normalization does not allow weights to explode all over the place and restricts them to a certain range. Many researchers have proposed different normalization techniques to optimize network models. The most commonly used techniques are batch, switchable and group normalizations [22].

Over the years, batch normalization [24] (BN) has become a commonly accepted process and has proven to be very successful in many deep learning tasks. To normalize its functions during activations, BN [24] makes use of the mean and variance computed within a mini batch of results. To have zero mean and unit deviation, BN [24] standardizes activations. BN [24]'s key benefits include facilitating quicker convergence in fewer iterations of instruction, and having a degree of regularization, thereby reducing the error of generalization. Because of device constraints, BN [24] does not work effectively. Therefore, group normalization (GN) [22] was implemented as a layer that separates channels into groups, calculates the mean and standard deviation over these channel groups into groups and calculates the mean and standard deviation over these channel groups during training over each case. Batch measurements are not exploited by GN. This helps it to do well with smaller micro batch sizes than BN [24].

2.3 Weight Pruning

The weight pruning model aims to cause sparsity in different relation metrics in a deep neural network, thus reducing the number of parameters in the model that are not valued at 0. Recent researchers (Han et al., 2015a; Narang et al., 2017) plant deep networks at the expense of just a small loss of precision and accomplish a substantial reduction in the size of the model. Prune models achieve up to 3x decreases in the number of nonzero parameters with limited loss of precision across a wide variety of neural network architectures (deep CNNs, stacked LSTM and seq2seq LSTM models).

2.4 Fully Convolution Networks(FCNs)

The most basic concept behind FCNs [18] is that they only consist of locally connected layers without completely connected or thick layers (dropout, convolutions, activation, pooling). This helps to decrease the time and the number of parameters used for computation. It also implies that, regardless of the input image size, an FCN can operate. Usually, FCN's [18] are made up of:

- **Encoding Path:** The model collects the contextual information on the input image on this path and interprets it.
- **Decoding Path:** Precise location or creation of segmentation maps in the encoding path from the derived context.
- **Skip connections/Bottlenecks:** Integrates encoding and decoding route information by summing up function charts.

2.5 U-Net

UNet[19] has evolved from the CNN for medical image modality analysis. U-Net is contraction, expansion and the bottleneck layers that merge knowledge from the contraction and expansive paths do so by concatenating the function maps as summarized in the FCN[18] deep network architecture. The key distinction between U-Net and FCN[18] is that four sections, each consisting of two unpadded 3x3 CNNs with a ReLU activation layer and a 2x2 max-pooling layer, comprise the direction of the U-encoding Net. After each sampling stage, the number of feature channels tends to double, but due to max-pooling, the size of the feature maps is limited. 2x2-up sampling of 3x3 regular convolutions is used in the direction of decoding. A concatenation of characteristics from the respective layers in the encoding direction is followed by each convolution. This helps move the localization knowledge that is retrieved from the contraction to the expansive route during downsampling.

3. Materials And Methods

In this section, we outline the proposed technique, describe the WP-UNet architecture and conduct experiments.

3.1 Weighted Pruning (WP) with Depthwise Separable Convolutions

WP-UNet has been proposed to be implemented on standard convolutions. It was recommended that WP-UNet be added to the regular convolutions. In this work, to minimize the number of parameters and necessary computations in the U-Net model, the regular convolution layers are replaced with depthwise separable layers[21]. WP is added to the U-Net's usable layers. Therefore, WP-UNet achieves a higher and smoother failure curve during training and helps increase model accuracy.

3.2 WP-UNet (Proposed Architecture)

With a few changes, WP-UNet(Fig. 5) follows a similar architecture to U-Net. The other convolution layers are constructed of separable convolutions, except for the first convolution layer, which has a regular convolution. Five blocks are made up of the pruning [27] of encoding layers. The design of the WP-UNet architecture is as follows:

- **Block 1:** Initial block with a regular convolution layer, ReLU activation function and normalization batch
- **Blocks 2, 3 and 4:** These blocks of two depthwise separable convolution layers [21], two activation layers and one normalization layer are composed of a WP-Unet block(Fig. 4).
- **Block 5:** A separable final depthwise layer[21] with a dropout layer[17]

The upsampling of the decoding path is performed with a scale of two to restore the size of the segmentation map. The WP-decoding UNet's direction is made up of a mixture of standard convolution blocks and WP-UNet blocks. It also consists of the same number of network layer blocks.

- **Block1:** A separable convolution layer in depth with its characteristics concatenated with the dropout layer [17] from the encoding direction block4.
- **Blocks 2, 3, and 4:** a WP-UNet block and a separable depthwise layer [21] concatenated from the encoding direction with matching blocks
- **Block 5:** Two WP-UNet blocks with the last block one as the final layer and two depthwise separable layers.

3.3 Configuration

The training was based on Keras with a TensorFlow backend as a Google Colab deep learning framework enabled with an NVidia GPU such as T4(12 GB memory) with a high-memory VM.

3.4 Datasets(KiTs Challenge Dataset)

The KiTs challenge datasets for kidney tumor disease segmentation are the datasets used to assess the performance of WP-UNet. Proposed deep network model applied on the KiTs dataset [5]. It consists of 210 high contrast CT scans of patients, collected in the preoperative arterial process and chosen from a cohort of subjects who underwent partial or radical nephrectomy [26] for one or more kidney tumors at the University Of Minnesotal Medical Center and were applicants for inclusion in this database between 2010 and 2018. The volumes included are characterized by different plane resolutions ranging from 0.437 to 1.04 mm, with slice thicknesses ranging from a minimum of 0.5 mm to a minimum of 5.0 mm in each case.

The dataset also provides the ground-truth mask of both healthy kidney tissue and healthy tumors (Fig. 6) for each case included. Under the guidance of experienced radiologists, a group of medical students manually generated sample labels with only CT scan image axial projections. A detailed description of the segmentation strategy for the ground truth is described in [5]. The KiTs challenge dataset is provided with shape (num slices, height, width) in the standard NIFTI format.

Figure 6 Sample CT scan imaging and ground truth labels from the Kidney, and Kidney Tumor Segmentation (KiTs) Datasets.

3.5 Data Preprocessing

Initially, the resolution of the images of the KiTs challenge dataset stacks was originally 512 x 512, but because of technical limitations, it was resized to 256 x 256. To reduce disk capacity, the data stack is accessible in the standard NIFTI format, which is converted into tfrecords. Owing to the small number of training images available data augmentation techniques have been used. A smaller number of images could lead to a concept known as overfitting, where a trained model performs on training data very well but on new test data performs poorly. Horizontal flip, zoom range, height and width adjustment range were used in these enhancement techniques. After improvement, the number of images of the box stacks dataset grew to 120. The resolution of the images was also decreased in the Kit data set (512 x 512). Center cropping and data normalization were also used to ensure 0 mean and unit variance and the original 3D slices were converted into 2D slices with separable and ReLU convolution layers for training and testing of UNet[19]. For the training of 44175 images and 17030 image verification, the suggested WP-UNet with ReLU activation function is used.

3.6 Optimization

The Adam optimization algorithm [16] has been used to train the network model with a learning rate range from 0,0001 to 0,00001 on the KiTs CT scan image dataset. Losses in the training were based on the KiTs dataset's binary cross-entropy loss. The loss was a weighted sum of the loss of a negative dice and of the binary algorithms for the KiTsdataset.

3.7 Performance Metrics

The key performance metrics used in measuring WP-UNet performance on the CT scan dataset are explained in detail in this section.

• Accuracy (AC)

In the formula given below, accuracy measures the percentage of correct predictions and is given by:

$$AC = \frac{TP+TN}{TP+TN+FP+FN}$$

where TP = Predicted Positive and Its True, TN =Predicted Negative and Its True, FP = Predicted Positive and Its False, FN = Predicted Negative and Its False

• Mean Intersection Over Union (Mean IOU)

Mean IoU [28] is a popular evaluation method for semantically segmented images that first determines the IOU for each semanticized class and then determines the average over classes. The mean IOU shall be described as:

$$\text{Mean IOU} = \frac{TP}{TP+FP+FN}$$

• Flops

Floating point FLOPs are essentially a calculation of the number of multiplications and additions of the floating point number to be performed by the processor of the computation device. A neural network in progress requires such floating point operation calculations to estimate the complexity of the proposed model.

4. Results

In this section, experiments are performed to calculate the computational requirements of WP-UNet, its inference speed, and the efficiency of segmentation (Fig. 9,10) on the specified datasets.

4.1 Ablation Study

To test the efficiency of the proposed model and to help the final design decisions made in this report, a detailed ablation study is conducted. There are three separate changes to the design of the architecture:

- **U-Net**-Original architecture of U-Net with Batch normalization
- **U-Net (Depthwise + BN)**-U-Net with batch normalization architecture
- **WP-UNet (Weight Prune + Depthwise + BN)** - WP-UNet was proposed based on the pruning of the practical UNet[19] architecture consisting of strongly separable convolutions with batch normalization.

Alongside the original U-Net architecture, the proposed WP-UNet based on BN [24] in Table 1, the performance of these modifications is recorded.

4.2 Results on KiTs Dataset

U-Net and WP-UNet are trained from scratch for only 10 epochs and their mean dice scores are related to their inference velocity on validation data. The choice of a smaller number of epochs is due to the potential of the Adam [16] optimization algorithm to reach a minimum efficiently and each epoch runs 1014 iterations over the KiTs dataset. A higher loss and average Dice coefficient are obtained by WP-UNet(Fig. 7) relative to the U-Net(Fig. 8) model. Its inference rate is also faster on a single GPU device.

Table 1
Comparison of results between WP-UNet and ther models

Model	Training Loss	Training Accuracy	Mean IoU
U-Net	0.5601	97.87	0.435
U-Net (Depthwise + BN)	0.4439	93.62	0.362
WP-UNet (Weight Pruning + Depthwise + BN)	0.066	98.43	0.428

Table 2
Computational Comparison between WP-UNet and other Models

Model	# Parameters	#Flops
U-Net	5,680,353	62.4e
U-Net (Depthwise + BN)	2,601,921	7.8e
WP-UNet (Weight Pruning + Depthwise + BN)	1,297,441	7.2e

Conclusion

To help in disease detection, therapy, and general research the segmentation of biomedical images is an important first phase in distinguishing tissues in image scans. To help avoid complications that may occur due to late detection, early diagnosis is necessary. However, with the availability of large-scale biomedical evidence, the workload has also expanded for neurologists, radiologists, and other field specialists. Several deep learning architectures have been proposed to help provide faster, precise and timely detections, and several have experienced considerable success in these tasks. One such model that is widely agreed upon by CT scanning image segmentation researchers is the U-Net architecture.

Portable devices have recently been enabled with computing capacities that were only imaginable for large machines in the past. Deep learning implementations, however, require much greater computing. This makes the development of deep learning systems on mobile or embedded computers very difficult. For example, the U-Net architecture needs more than 62 M FLOPs and a storage capacity of over 370 megabytes (Mb), which are very high requirements. In comparison, little attention has been paid to the implementation of deep learning approaches in biomedical imaging fields on resource-constrained systems.

For the segmentation of CT scan image data on devices with small computational budgets, weighted pruning UNet (WP-UNet)(Fig. 9 & 10) was presented in this review. Separable convolutions are used by the WP-UNet architecture. Our findings indicate that the architecture proposed is smaller than U-Net and demands 3x less computational complexity (Table 2) while retaining respectable accuracy results. This means that WP-UNet can be implemented on embedded computers and devices with limited computing power on any handheld computer. It is shown that WP-UNet performs impressively on CT scan images based on the findings of the studies carried out on the kidney tumor CT scan image dataset of the KiTs challenge for kidney and tumor segmentation.

The robustness of WP-UNet is also shown during test results to do considerably better on smaller CT scan tumor segmentations than the initial UNet architecture and it may be extended to other tasks, such as CT scan lung cancer diagnosis, detection of breast cancer, and many other related biomedical applications.

In the future, the authors plan to continue working on the design of deep architectures that require much less computing and concentrate on embedded systems, thereby yielding further test results.

Abbreviations

WP – Weight Pruning

CT – Computer Tomography

BN – Batch Normalization

FLOPs – Floating Point Operations

KiTs 19 – KiTs 19 World Challenge Dataset

NIFTI – Neuroimaging Informatics Technology Initiative

Declarations

1. Ethics approval and consent to participate

Not applicable

2. Consent for publication

Not applicable

3. Availability of data and materials

The datasets generated and/or analysed during the current study are available in the KiTs 19 repository, [Data - KiTs19 - Grand Challenge \(grand-challenge.org\)](https://grand-challenge.org/data-kiTs19)(Link: [Data - KiTs19 - Grand Challenge \(grand-challenge.org\)](https://grand-challenge.org/data-kiTs19))

4. Competing interests (same as provided on the submission system)

The authors declare no competing interests as defined by journal or other interests that might be perceived to influence the results and/or discussion reported in this paper.

5. Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

6. Authors' contributions (same as provided on the submission system)

Patike analyzed and interpreted the dataset regarding kidney tumor disease using WP-UNet. Subarna validated the model based on the number of parameters and flops.

7. Acknowledgements

The datasets used for the analysis in this manuscript were obtained from KiTS 19 . We gratefully acknowledge the contribution of the people and organizations involved in the cancer image archive initiative as participants, organizers, or funders.

8. Authors' information (optional)

Dr. Subarna Chatterjee joined the Faculty of Engineering & Technology at M S Ramaiah University of Applied Sciences, Bangalore. Her research areas are computer graphic, soft computing, programming with C, VLSI computer architecture, and organization. She got Inspire fellowship from DST Govt. of India and selected for RF for SMSI, IIT Kharagpur (Sponsored by Texas Instruments).

P Kiran Rao obtained the Degree of Master of Technology in CSE from JNTU Aanantapuramu. Presently he is pursuing a doctoral degree in the faculty of engineering from M S Ramaiah University of Applied Sciences, Bangalore. Moreover, he is working as an assistant professor in the CSE department for G Pullaiah College of Engineering & Technology, Kurnool. His research interests include cloud computing, deep learning, and medical image processing.

9. Conflicts of Interest:

The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal (2018) 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), pp. 394–424.
2. Scelo and T. L. Larose (2018) Epidemiology and risk factors for kidney cancer. *Journal of Clinical Oncology* 36 (36), pp. 3574–3581.
3. M. Nguyen, I. S. Gill, and L. M. Ellison (2006) The evolving presentation of renal carcinoma in the united states: trends from the surveillance, epidemiology, and end results program. *The Journal of urology* 176 (6), pp. 2397–2400.
4. Sun, F. Abdollah, M. Bianchi, Q. Trinh, C. Jeldres, R. Thuret, Z. Tian, S. F. Shariat, F. Montorsi, P. Perrotte, et al. (2012) Treatment management of small renal masses in the 21st century: a paradigm shift. *Annals of surgical oncology* 19 (7), pp. 2380–2387.
5. Heller, N. Sathianathan, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, et al. (2019) The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445*.
6. P Kiran Rao, Dr Subarna Chatterjee et al. (2020) Diagnosis of Kidney Renal Cell Tumor through Clinical data mining and CT scan image processing: A Survey <https://doi.org/10.26452/ijrps.v11i1.1778>. pp. 13-24.
7. Sifre, L. Rigid-Motion Scattering for Image Classification. Ph.D. Thesis, Ecole Polytechnique, Palaiseau, France, 2014. CMAP Rigid-Motion Scattering For Image Classification. Available online: https://www.di.ens.fr/data/publications/papers/phd_sifre.pdf (accessed on 17 February 2020).
8. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
9. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: E_cient Convolutional Neural Networks for Mobile Vision Applications. Available online: <https://arxiv.org/abs/1704.04861> (accessed on 18 February 2020).
10. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shu_eNet: An Extremely E_cient Convolutional Neural Network for Mobile Devices. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018*, Salt Lake City, AL, USA, 18–22 June 2018; pp. 6848–6856.
11. Qin, Z.; Zhang, Z.; Chen, X.; Peng, Y. FD-MobileNet: Improved MobileNet with a Fast Downsampling Strategy. 2018. Available online: <https://ieeexplore.ieee.org/abstract/document/8451355> (accessed on 17 February 2020).
12. Denil, M.; Shakibi, B.; Dinh, L.; Ranzato, M.; de Freitas, N. Predicting Parameters in Deep Learning. Available online: <https://papers.nips.cc/paper/5025-predicting-parameters-in-deep-learning.pdf> (accessed on 18 February 2020).

13. LeCun, Y.; Denker, J.S.; Solla, S.A. Optimal Brain Damage. *Adv. Neural Inf. Process. Syst.* 1990, 2, 598–605.
14. Sharan Narang, Gregory F. Diamos, Shubho Sengupta, and Erich Elsen. Exploring sparsity in recurrent neural networks. *CoRR*, abs/1704.05119, 2017.
15. Song Han, Huizi Mao, and William J. Dally. Deep compression: Compressing deep neural network with pruning, trained quantization and Huffman coding. *CoRR*, abs/1510.00149, 2015a.
16. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 17 February 2020).
17. Srivastava, N.; Hinton, A.; Sutskever, K.I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 2017, 15, 1929–1958.
18. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 640–651. [CrossRef]
19. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Form. Asp. Compon. Softw.* 2015, 9351, 234–241.
20. Yao, W.; Zeng, Z.; Lian, C.; Tang, H. Pixel-wise regression using U-Net and its application on pansharpening. *Neurocomputing* 2018, 312, 364–371. [CrossRef]
21. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
22. Wu, Y.; He, K. Group Normalization. *Formal Asp. Compon. Softw.* 2018, 11217 LNCS, 3–19.
23. Alvarez, J.M.; Salzmann, M. Compression-aware Training of Deep Networks 2017. Available online: <http://papers.nips.cc/paper/6687-compression-aware-training-of-deep-networks> (accessed on 17 February 2020).
24. Daejin Jung, Wonkyung Jung, and Byeongho Kim, Sunjung Lee, Wonjong Rhee, Jung Ho Ahn, *Restructuring Batch Normalization to Accelerate CNN Training*, 2018
25. “Kidney Cancer Statistics.” World Cancer Research Fund, 12 Sept. 2018, wcrf.org/dietandcancer/cancer-trends/kidney-cancer-statistics.
26. Kutikov, Alexander, and Robert G. Uzzo. “The RENAL nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth.” *The Journal of urology* 182.3 (2009): 844-853.
27. Hassibi, B.; Stork, D.G. Second Order Derivatives for Network Pruning: Optimal Brain Surgeon. Available online: [https://authors.library.caltech.edu/54983/3/647-second-order-derivatives-for-networkpruning-optimal-brain-surgeon\(1\).pdf](https://authors.library.caltech.edu/54983/3/647-second-order-derivatives-for-networkpruning-optimal-brain-surgeon(1).pdf) (accessed on 15 February 2020).
28. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation December 2016 DOI: [1007/978-3-319-50835-1_22](https://doi.org/10.1007/978-3-319-50835-1_22)
29. Michael H. Zhu, Suyog Gupta “To prune, or not to prune: exploring the efficacy of pruning for model compression” arxiv.org/pdf/1710.01878[stat.ML] : 2017

Figures

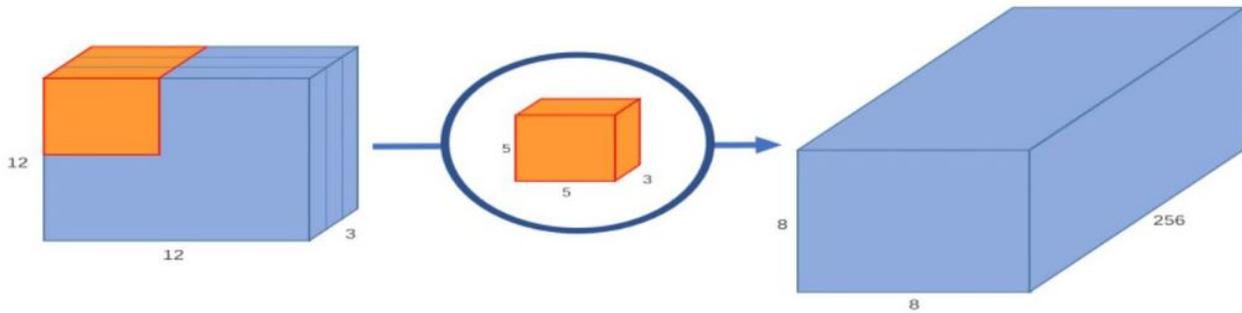


Figure 1

Normal Convolution Layers

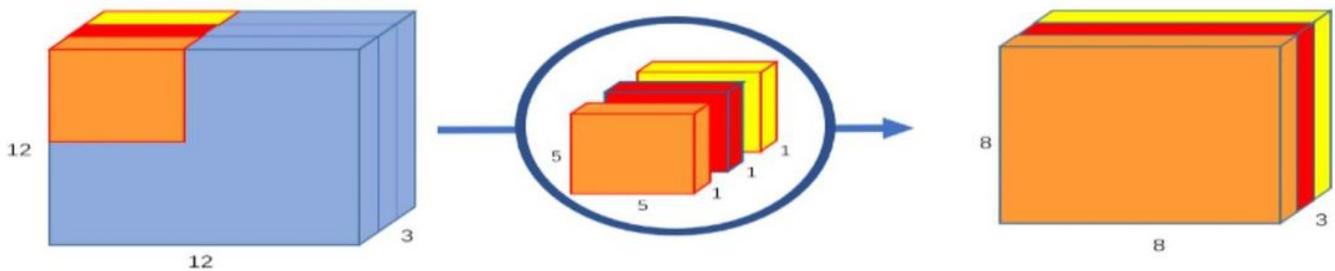


Figure 2

Depthwise convolution uses 3 kernels to transform a 12x12x3 image to an 8x8x3 image

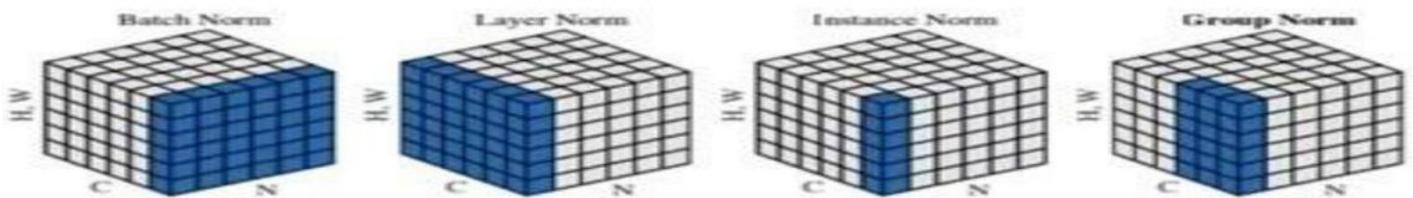


Figure 3

A Visual comparisons of various normalization methods

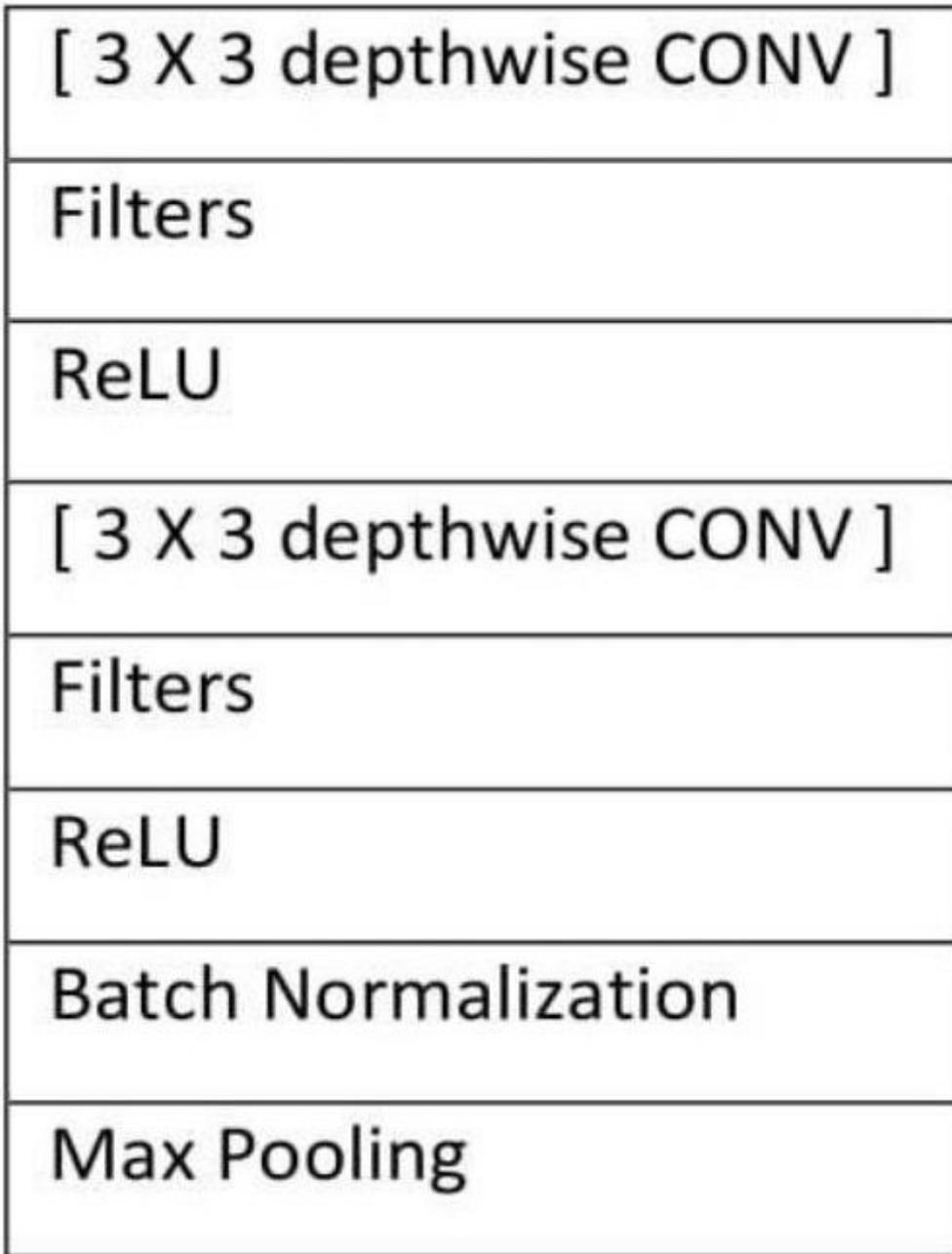


Figure 4

WP-UNet Block

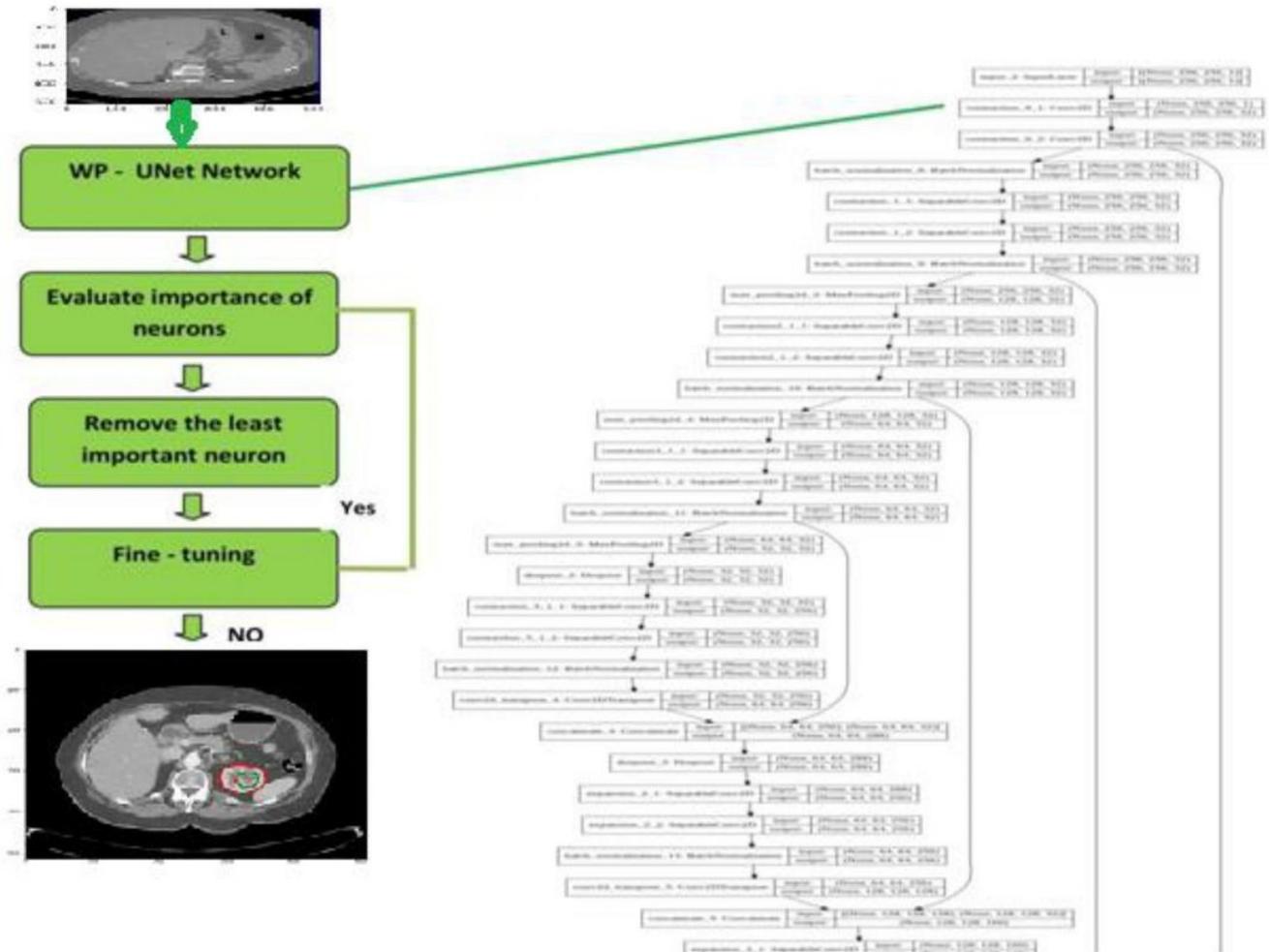


Figure 5

WP-UNet Architecture

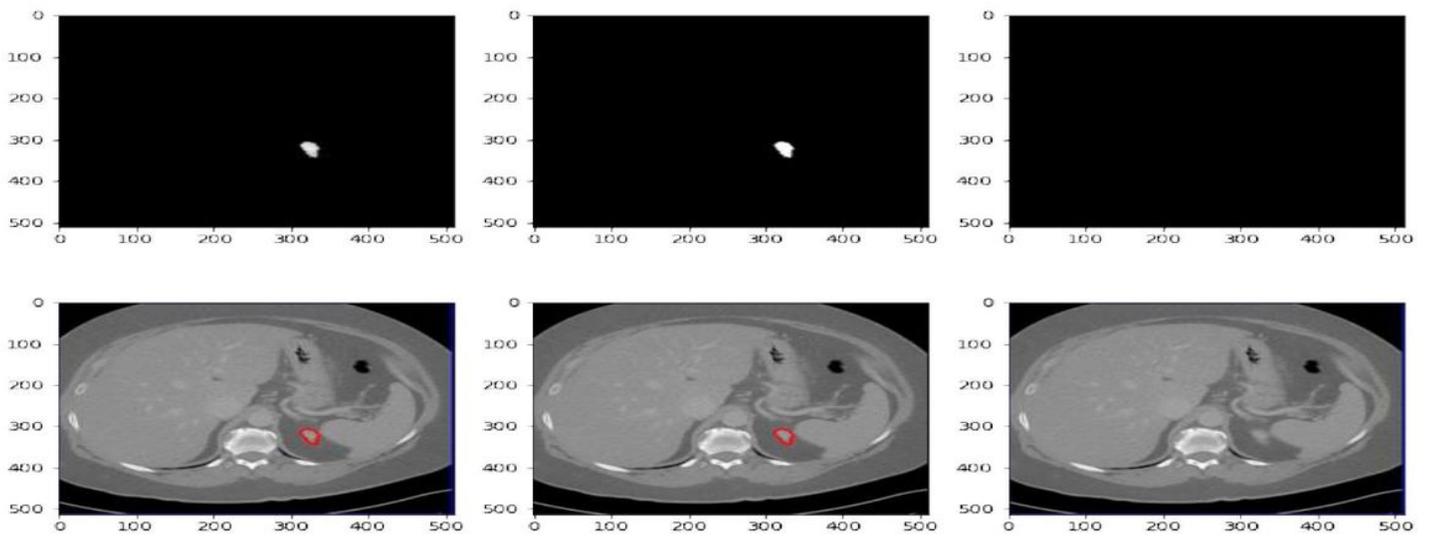


Figure 6

Sample CT scan imaging and ground truth labels from the Kidney, and Kidney Tumor Segmentation (KiTs) Datasets.

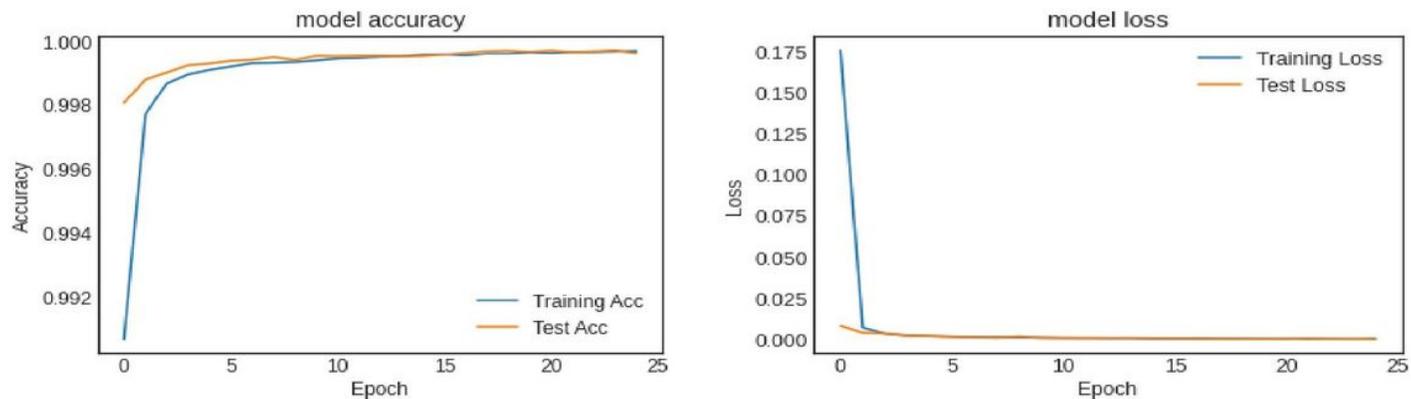


Figure 7

Standard U-Net architecture accuracy and Loss

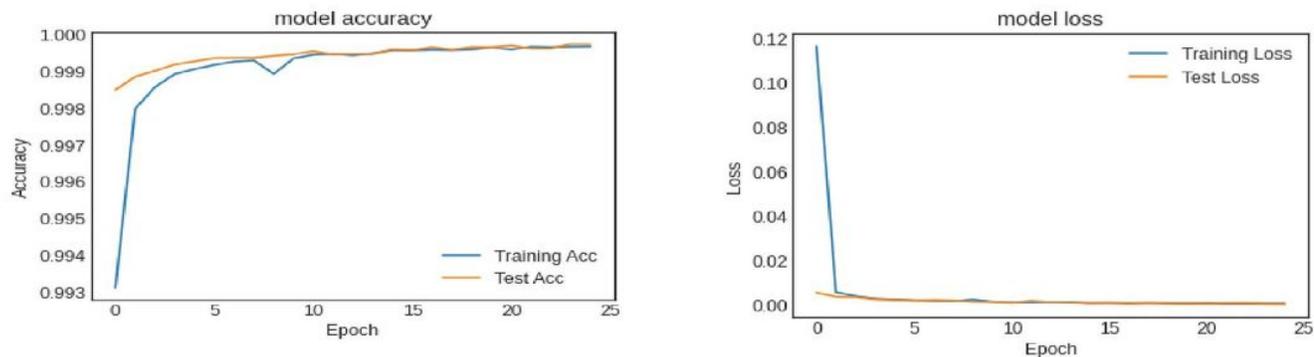


Figure 8

WP-UNet architecture accuracy and Loss

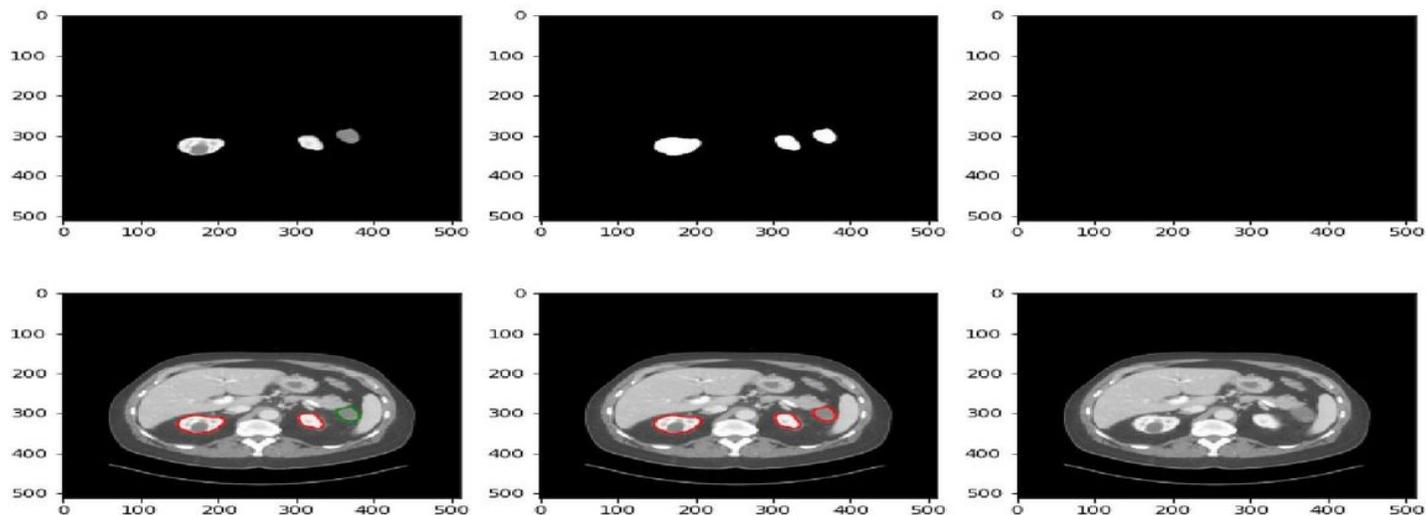


Figure 9

Sample Segmentation results on the KiTs Dataset from our test split. WP-UNet significantly performs better than U-Net on tumors

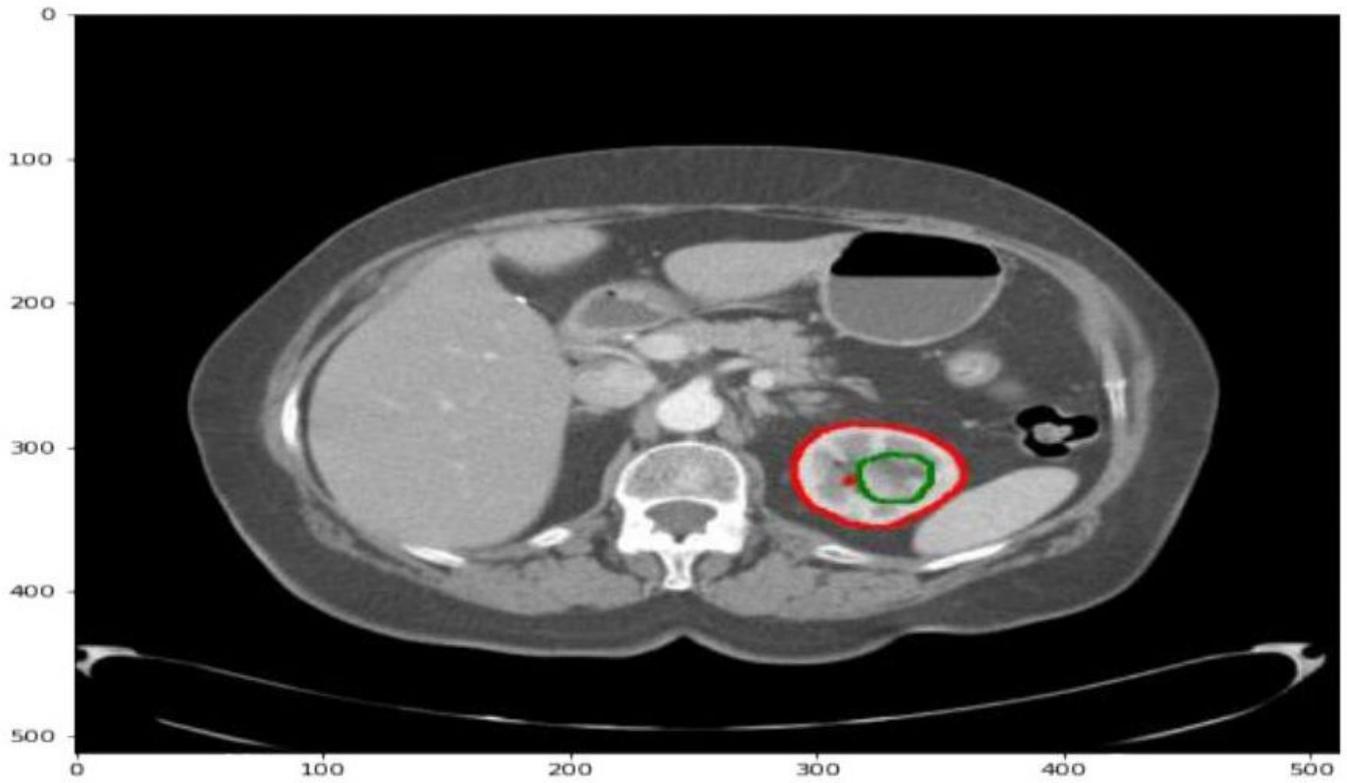


Figure 10

Sample Segmentation of Kidney and Tumor