WP-UNet: Weight Pruning U-Net with Depth-wise Separable Convolutions for Semantic Segmentation of Kidney Tumours

P Kiran Rao (✉ kiranraocse@gmail.com)

MS Ramaiah University of Applied Sciences

Dr Subarna Chatterjee

MS Ramaiah University of Applied Sciences

Dr. M Sreedhar Sharma

Govt General Hospital, Kurnool

Research Article

Keywords: Kidney, Kidney Tumour Segmentation, Depth-wise Separable Convolution, Pruning, WP-UNet

Posted Date: May 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-526418/v1>

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

[Read Full License](#)

Abstract

Background

Accurate semantic segmentation of kidney tumours in computed tomography (CT) images is difficult because tumours feature varied forms and, occasionally, look alike. The KiTs19 challenge sets the groundwork for future advances in kidney tumour segmentation.

Methods

We present WP-UNet, a deep network model that is lightweight with a small scale; it involves few parameters with a quick assumption time and a low floating-point computational complexity.

Results

We trained and evaluated the model with CT images from 300 patients. The findings implied the dominance of our method on the training Dice score (0.98) for the kidney tumour region. The proposed model only uses 1,297,441 parameters and 7.2e FLOPS, three times lower than those for other network models.

Conclusions

The results confirm that the proposed architecture is smaller than that of U-Net, involves less computational complexity, and yields good accuracy, indicating its potential applicability in kidney tumour imaging.

1. Introduction

American Cancer Society has reported on the prevalence of kidney cancer in both men and women. Overall, the lifetime risk to develop kidney cancer is approximately 1/48 and 1/83 for men and women, respectively. The types of kidney cancer in this study were of an advanced stage. Kidney cancers are generally this advanced because the kidneys are situated deep inside the body and are not physically perceived on a physical inspection. Several imaging methods are currently in use to track the growth of kidney tumours. This method has become increasingly popular because it can selectively extract diseased tissues and retain additional stable tissue. This approach was successful in treating small kidney masses. After the precise evaluation of the kidney tumour, details such as the kidney, tumour structure, and others can be collected. In a recent study (Hesamian et al., 2019), it was impossible to derive the essential details from computed tomography (CT) or magnetic resonance imaging scans. Kidney tumours vary in colour, form, and scale, and have a similar appearance to their parenchyma and other

nearby tissues. Given the segmentation of the kidney(Kanishka Sharma, 2017) tumour area, segmenting kidney tumours is extremely difficult.

Currently, there is an increased need to deploy deep learning solutions on mobile handheld devices (Hooman Vaseli, 2019), embedded systems (Karakanis al., 2020), or machines with minimal resources. An important reason why convolutional neural networks (CNNs) are challenging to train is because they are over-parameterised (Denil, 2013), and they typically require greater computational power and storage space for training and inference. Deep learning researchers have claimed many ‘pruning’ strategies or quantising learned parameters on broad image datasets (LeCun et al., 1990; Alvarez and Salzmann, 2017; Han et al., 2016). Others have concentrated on teaching compact models (Howard et al., 2017; Zhan et al., 2017; Qinet et al., 2018) from scratch by factorising regular convolution layers into depth-wise separable convolution layers for cheaper computations.

Although CNNs have achieved the best results in functional implementations, robustness and accuracy remain challenging. Ronneberger et al. (2015) proposed a tool called U-Net for automated medical image segmentation to solve these issues. The U-Net synthesises vital information by reducing the cost function in the first half of the network and generates an image in the second half. Inspired by the U-Net model, we approached the current challenge of kidney tumour segmentation by proposing a WP-U-Net model. We implemented weight pruning of the U-Net with a depth-wise separable convolution architecture, and thus it refines even tiny regions in the output tumour picture. The system precisely separates the tumour regions of the kidney and offers established quantification and qualitative validity.

2. Related Works

Several computer-aided diagnosis models and artificial neural networks have been developed to classify and segment renal tumours using CT scans. Lingararu et al. (2011) published a computer-aided method which was used to examine a collection of brain CT scans of 43 patients. In this system, tumours were robustly segmented with approximately 80% overlap. The methodology studied morphological variations between various types of lesions. Lee et al. (2017) developed a computer program capable of detecting and identifying small renal masses in CT images. Their tests yielded a specific signal-to-noise ratio of 99.63%.

Shah et al. (2017) presented a segmentation approach using machine learning. Yang et al. (2014) created a system to automatically segment CT images of the kidney based on multi-atlas registration. First, they recorded a low-resolution image with a series of higher-resolution images to create a patient-registered image. Next, the kidney tissues were segmented and aligned to achieve the final segmented production.

Various researchers have also experimented with the segmentation of renal tumours using deep learning. Thong et al. (2016) used an online patch-wise convolutional kernel to classify the central voxel in 2D patches. Then, the ConvNet analysed the CT scan data of each kidney tumour slice. Skalski et al. (2016) demonstrated an efficient hybrid level-set approach with elliptical-form restrictions for kidney

segmentation. The RUSBoost algorithm and decision trees were used to differentiate between kidney and tumour structures, serving as a solution to class imbalance and the need for defining additional voxels. Their model achieved an average precision of 92.1%. Wang et al. (2018) defined a CNN-based model for kidney segmentation. They proposed a CNN-based segmentation scheme that integrates the bounding box information. They also improved the CNN model by fine-tuning the model for each picture.

Network prototypes. Deep neural networks are superior in their capacity and ability to be generalised. Deep models that learn entirely from data produce excellent results for many tasks when compared with humans. They enhance the plot depth. Researchers have achieved further advances in neural networks. The use of skip links in deep neural networks makes them more trainable to perform tasks such as deep learning. U-Net was initially planned to resolve image segmentation, but others such as VGGNet and ResNet were designed for deep classification (Linguraru et al., 2011) supervision to further enhance segmentation. Network pruning has been widely studied to compress the CNN models (He et al., 2017, 2018). In early work, network pruning proved to be a valid way to reduce network complexity and overfitting (LeCun et al., 1989; Hanson and Pratt, 1989; Hassibi et al., 1993; Strom, 1997). Recently, Han et al. (2015) pruned state-of-the-art CNN models with no accuracy loss.

3. Proposed Method

In this section, we propose the WP-UNet model and describe the modified objective function.

3.1 Image Pre-processing

All CT images were resized to 256×256 pixels in the training set and separated by 255 pixels to normalise the pixel values from 0 to 1.

3.2 Dataset

The KiTS challenge dataset for kidney tumour disease segmentation was used to assess the performance of WP-UNet. The KiTS dataset (Helleret et al., 2019) consists of 210 high-contrast CT scans collected in the preoperative arterial process. They were chosen from a cohort of subjects who underwent partial or radical nephrectomy (Kutikov et al., 2009) for one or more kidney tumours at the University of Minnesota Medical Center and were eligible for inclusion between 2010 and 2018. The volumes included are characterised by different plane resolutions ranging from 0.437 to 1.04 mm, with slice thicknesses ranging from 0.5 mm to 5.0 mm in each case.

The dataset also provides the ground-truth mask of healthy kidney tissue and healthy tumours (Figure 1) for each case. 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 Helleret al. (2019). The KiTs challenge dataset is provided with shape (number of slices, height, width) in the standard NIFTI format.

3.3 WP-UNet Model

Figure 2 shows the detailed architecture of the proposed WP-UNet model. The network has the properties of the encoder and decoder structure of the vanilla U-Net (Shenet al., 2015). As suggested byLiu et al. (2018), first, the input image is passed into the standard convolution layer; subsequently, it is passed to the encoder part of the WP-UNet block. Here, to improve the model's generalisation capacity, a depth-wise separable convolutional layer is used, which helps the network select the features related to translation invariance with fewer parameters (Karakanis 2020) than the standard convolution layer.

WP-UNet encoding is composed of the following four blocks:

- Block1: A standard convolution layer, filters, a ReLu activation function, and a batch normalisation layer.
- Blocks 2 and 3: One WP-UNet block and a max-pooling layer. A WP-UNet block comprises two depth-wise separable convolution layers along with filters, two activation layers, and one BN layer (Figure 3).
- Block4: One WP-UNet block, a dropout layer to introduce regularisation (Kehl, 2017), and a max-pooling layer. All depth-wise (3×3) convolution layers were weight-standardised.

Up-sampling is performed in the decoder section, which is used to combine depth-wise separable convolutions and WP-UNet blocks as shown in Figure 2. It also consists of five blocks:

- Block1: A depth-wise separable convolution layer with its features concatenated with the dropout layer from Block4 of the encoding path.
- Blocks 2, 3, and 4: WP-UNet block and depth-wise separable layer concatenated with corresponding blocks from the encoding path.
- Block 5: Two WP-UNet blocks and two depth-wise separable layers, with the last one as the final prediction layer (Figure 2).

To improve the model performance and reduce the number of floating-point operations, we added network pruning (Liu et al., 2019) to the proposed architecture, as shown in Figure 4. The output of the network pruning (Han, 2016) WP-UNet model includes the kidney region, tumour region, and background, as shown in Figure 5.

3.4 Loss Function

In this study, the Adam optimiser (Kingma and Ba, 2014) is applied, which correctly updates the network weights by iteration in the training data. Adam makes an average in the first and second moments of gradients to adapt the learning rate parameter. Sabarinathan et al. (2019) proposed that the loss function be the sum of the categorical cross-entropy Dice loss channel one (C0) and Dice loss channel two (C1), as defined in Eq. (1).

$$Loss = L + DiceL(C0) + DiceLoss(C1) \quad (1)$$

$$Dice\ Loss = 1 - \frac{2 \sum_{n=k}^{i=0} Y_i P_i + \epsilon}{\sum_{n=k}^{i=0} Y_i + \sum_{n=k}^{i=0} P_i + \epsilon} \quad (2)$$

$$L = - \sum_{j=0}^M \sum_{i=0}^N Y_{ij} \log P_{ij} \quad (3)$$

where L is the cross-entropy loss. In Eq. (2), y_i and p_i are the ground truth and predicted segmented images, respectively. Moreover, to ensure the loss function stability, the coefficient ϵ is used.

3.5. Performance Metrics

The key performance metrics used to measure the WP-UNet performance on the CT scan dataset are explained in this subsection.

Accuracy (AC):

Accuracy measures the percentage of correct predictions, and is given as,

$$AC = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4)$$

where TP = correctly predicted positive, TN = correctly predicted negative, FP = incorrectly predicted positive, FN = incorrectly predicted negative.

Mean Intersection over Union (Mean IOU):

The mean IOU (Hassibi and Stork, 1993) is a popular evaluation method for semantically segmented images that first determines the IOU for each semantic class and then determines the average over classes. The mean IOU is expressed as follows:

$$\text{Mean IOU} = \frac{TP}{TP + FP + FN} \quad (5)$$

FLOPs:

FLOPs essentially calculates the number of multiplications and additions of floating-point numbers to be performed by the computation device's processor. A neural network in progress requires floating-point operation calculations to estimate the complexity of the proposed model.

4. Experiment And Results

4.1 Training

The proposed network was trained with two outputs, namely the kidney and kidney tumour regions. The weight updates were performed using the Adam optimiser with a learning rate of 0.001. The batch size was set to 16, and the total number of epochs was set to a hundred. 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 virtual machine.

4.2 Results

The standard Dice score is considered an evaluation metric for the performance of the proposed WP-UNet model. We employed 35,865 and 10,158 images as training and validation images, respectively, in our experiments. Table 1 shows the segmentation results of the proposed WP-UNet model for the training and validation images. From the table, we observe that during training, the proposed method achieves a training accuracy of 0.98 for the tumour region. Similarly, the computational resource usage of our network is listed in Table 2. Based on the experimental results, we perceive the power of network pruning in the proposed network. Because network pruning is added to the proposed architecture, the total number of flops and parameters is three times smaller than the typical UNet architecture.

Table 1: Comparison of results between WP-UNet and other models

Model	Training Loss	Training Accuracy	Mean IOU
U-Net	0.5601	97.87	0.435
U-Net (Depth-wise + BN)	0.4439	93.62	0.362
WP-UNet (Network Pruning + Depth-wise + 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 (Depth-wise + BN)	2,601,921	7.8e
WP-UNet (Network Pruning + Depth-wise + BN)	1,297,441	7.2e

In Figure 6, the qualitative effects of the KiTs19 dataset on the proposed WP-UNet model are shown. We used the provided input images and ground-truth reality images to perform the experiments. The segmented performance image is depicted in Figure 5. The red-coloured area is the kidney region in the output picture, and the green-coloured part is the kidney tumour. Numerous structures outside the tumour and kidney areas were neglected for simplicity. The final segmented output closely matches the ground-truth image from the quantitative results, which demonstrates the usefulness of the proposed WP-UNet.

5. Conclusion

Medical image segmentation is an important preliminary step in the identification of kidney organ structure and tumour tissues in CT image scans to aid in illness diagnosis, treatment, and general analysis. Early diagnosis is necessary to help in preventing complications that may arise due to late detections. However, with the increasing availability of large biomedical data, the workload on nephrologists, radiologists, and other experts in the field has also increased. To help provide easier, accurate, and timely detections, several deep learning methods have been proposed, most of which have proven to be successful. The U-Net architecture is one such model that is widely accepted among researchers for biomedical image segmentation tasks.

In this study, weight pruning UNet (WP-UNet) was proposed for the segmentation of kidney tumour data with limited computational resources. The WP-UNet architecture makes use of depth-wise separable convolutions (Figure 2) and pruning to reduce the parameters and floating-point operations. Moreover, the WP-UNet deep learning method exhibits a faster inference speed than that of the UNet method.

Our findings indicated that the proposed WP-UNet architecture yielded a satisfactory accuracy. Our system obtained a Dice score of 0.9799 and 0.9599 for the preparation and validation sets, respectively. The proposed WP-UNet model achieved the best segmentation outcomes in terms of the Dice score and usage of computational resources. Additionally, WP-UNet is shown to have a faster inference speed on test data and is beneficial for situations where rapid and accurate segmentation results are required.

List Of Abbreviations

WP – Weight pruning

CT – Computed tomography

BN – Batch normalisation

FLOPs – Floating-point operations

KiTs19 – KiTs19 World Challenge Dataset

NIFTI – Neuroimaging Informatics Technology Initiative

References

- Alvarez, J.M. and Salzmann, M. (2017) *Compression-aware Training of Deep Networks*. [online]<http://papers.nips.cc/paper/6687-compression-aware-training-of-deep-networks> (Accessed 17 February 2020).
- American Cancer Society. [online] <https://www.cancer.org/cancer/kidney-cancer/detection-diagnosis-staging/detection.html> (Accessed 17 December 2019).
- Denil, M., Shakibi, B., Dinh, L., Ranzato M.C. and Freitas, N. (2013) *Predicting Parameters in Deep Learning*. [online]<https://papers.nips.cc/paper/5025-predicting-parameters-in-deep-learning.pdf> (accessed on 18 February 2020).
- Han, S., Mao, H. and Dally, W.J. (2016) 'Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding', *International Conference on Learning Representations (ICLR)*.<https://arxiv.org/abs/1510.00149>
- Han, S., Mao, H. and Dally, W.J. (2016) 'Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization, and Human Coding'. [online]<https://arxiv.org/abs/1510.00149> (accessed on 17 February 2020).
- Hassibi, B. and Stork, D.G. (1993) *Second Order Derivatives for Network Pruning: Optimal Brain Surgeon*. [online][https://authors.library.caltech.edu/54983/3/647-second-order-derivatives-for-network-pruning-optimal-brain-surgeon\(1\).pdf](https://authors.library.caltech.edu/54983/3/647-second-order-derivatives-for-network-pruning-optimal-brain-surgeon(1).pdf) (Accessed 15 February 2020).
- He, Y., Zhang, X. and Sun, J. (2017) 'Channel Pruning for Accelerating Very Deep Neural Networks', *2017 IEEE International Conference on Computer Vision (ICCV)* pp. 1398-1406, doi: 10.1109/ICCV.2017.155.
- He, Y., Lin J., Liu Z., Wang, H., Li L.J. and Han, S. (2018) 'AMC: AutoML for model compression and acceleration on mobile devices', *in Proceedings of the European Conference on Computer Vision (ECCV)*,

pp.784–800.

Heller, N., Sathianathan, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., Dean, J., Tradewell, M., Shah, A., Tejpal, R., Edgerton, Z., Peterson, M., Raza, S., Regmi, S., Papanikolopoulos, N. and Weight, C. (2019) *The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes*, arXiv:1904.00445.

Hesamian, M.H., Jia, W., He, X. and Kennedy, P. (2019) 'Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges', *Journal of Digit Imaging*, Vol. 32, pp.582.
DOI:10.1007/s10278-019-00227-x

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang W., Weyand, T., Andreetto, M. and Adam, H. (2017) *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*.
[online]<https://arxiv.org/abs/1704.04861> (accessed on 18 February 2020).

Karakanis, S. and Leontidis, G. (2020) 'Lightweight deep learning models for detecting COVID-19 from chest X-ray images', *Computers in Biology and Medicine*, Vol. 130, pp.104181.
DOI:10.1016/j.compbiomed.2020.104181. Epub ahead of print. PMID: 33360271; PMCID: PMC7831681.

Kehl, W., Tombari, F., Ilic, S. and Navab, N. (2017) 'Real-Time 3D Model Tracking in Color and Depth on a Single CPU Core', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, pp. 465–473.

Kingma, D.P. and Ba, J. (2014) 'Adam: A method for stochastic optimization', arXiv:1412.6980.

Kutikov, A. and Uzzo, R.G. (2009) 'The renal nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth'. *The Journal of Urology*, Vol. 182. No. 3, pp.844–853.

LeCun, Y., Denker, J.S. and Solla, S.A. (1990) 'Optimal Brain Damage', *Advances in Neural Information Processing Systems*, Vol. 2, pp.598–605.

Lee, H.S., Hong, H. and Kim, J. (2017) 'Detection and segmentation of small renal masses in contrast-enhanced CT images using texture and context feature classification', *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, VIC, 2017, pp. 583–586. DOI:10.1109/ISBI.2017.7950588.

Linguraru, M.G., Wang, S., Shah, F., Gautam, R., Peterson, J., Linehan, W. M. and Summers, R.M. (2011) 'Automated non-invasive classification of renal cancer on multiphase CT', *Medical Physics*, Vol. 38, No. 10, pp.5738–5746. DOI:10.1118/1.3633898.

Liu, R., Lehman, J., Molino, P., Such, F.P., Frank, E., Sergeev, A. and Yosinski, J. (2018) 'An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution', arXiv:1807.03247, 2018.

Liu, Z., Sun, M., Zhou, T., Huang, G. and Darrell, T. (2019) 'Rethinking the value of network pruning', *International Conference on Learning Representations (ICLR)*.

Qin, Z., Zhang, Z., Checn, X., Wang, C. and Peng, Y. (2018) 'FD-MobileNet: Improved MobileNet with a Fast-Downsampling Strategy', *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp.1363-1367

Olaf, R., Fischer, P. and Brox, T. (2015) 'U-Net: Convolutional Networks for Biomedical Image Segmentation', In: Navab N., Hornegger J., Wells W., Frangi A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Sabarinathan, D., Beham, M.P. and Roomi, S.M.Md.M. (2019). 'Hyper Vision Net: Kidney Tumor Segmentation Using Coordinate Convolutional Layer and Attention Unit', in *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics*, pp.609–618. Springer, Singapore, 2019. DOI:10.1007/978-981-15-8697-2_57

Shah, B., Sawla, C., Bhanushali, S., Bhogale, P. (2017) 'Kidney Tumor Segmentation and Classification on Abdominal CT Scans', *International Journal of Computer Applications*, Vol. 164, No. 9, pp.1–5.

Sharma, K. (2017) 'Machine Learning Methods for Segmentation in Autosomal Dominant Polycystic Kidney Disease', PhD thesis, Technische Universität München, Munich, Germany.

Shen, W., Wang, X., Wang, Y., Bai, X. and Zhang A. (2015) 'Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.3982–3991.

Skalski, A., Jakubowski, J. and Drewniak, T. (2016) 'Kidney tumor segmentation and detection on Computed Tomography data', in *IEEE International Conference on Imaging Systems and Techniques (IST)*, Chania, 2016, pp.238–242. DOI:10.1109/IST.2016.7738230

Thong, W., Kadoury, S., Piche, N. and Pal, C.J. (2016) 'Convolutional networks for kidney segmentation in contrast-enhanced CT scans', *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. Vol. 6, No. 3, pp.277–282. DOI:10.1080/21681163.2016.1148636.

Vaseli, H., Liao, Z., Abdi, A.H., Girgis, H., Behnami, D., Luong, C., Dezaki, F.T., Dhungel, N., Rohling, R., Gin, K. and Abolmaesumi, P. (2019) 'Designing lightweight deep learning models for echocardiography view classification', in *SPIE Medical Imaging*, 2019, San Diego, California, United States.

Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S. and Vercauteren, T. (2018) 'Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning', *IEEE Transactions on Medical Imaging*, Vol. 37, No. 7, pp.1562–1573. <https://doi.org/10.1109/TMI.2018.2791721>

Yang G., Gu, J., Chen, Y., Liu, W., Tang, L., Shu, H. and Toumoulin, C. (2014) 'Automatic kidney segmentation in CT images based on multi-atlas image registration', in *36th Annual International Conference of the IEEE*

Zhang, X., Zhou, X., Lin, M., Sun, J. (2017) 'ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices', in *Proceedings IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, AL, USA, pp.6848–6856.

Figures

Figure 1

An example of CT scan images from the KiTs19 Challenge dataset

Figure 2

An overview of the detailed architecture of WP-UNet

3 x 3 Depthwise Conv
Filters
ReLU
3 x 3 Depthwise Conv
Filters
ReLU
Batch Normalization(BN)
MAX Pooling

Figure 3

Components of the WP-UNet block

Figure 4

WP-UNet network pruning

Figure 5

Sample kidney and tumourregions

Figure 6

Illustration of original input CT images and their respective kidney and tumour segmented output images