Lighter U-Net for Segmenting White Matter Hyperintensities in MR Images

Jun Zhuang
Indiana University-Purdue University
Indianapolis
Indianapolis, IN, USA
junz@iu.edu

Mingchen Gao
University at Buffalo
Buffalo, NY, USA
mgao8@buffalo.edu

Mohammad AI Hasan
Indiana University-Purdue University
Indianapolis
Indianapolis, IN, USA
alhasan@iupui.edu

ABSTRACT

White matter hyperintensities (WMH) is one of main consequences of small vessel diseases. Automated WMH segmentation techniques play an important role in clinical research and practice. U-Net has been demonstrated to yield the best precise segmentation results so far. However, sometimes it losses more detailed information as network goes deeper. In addition, it usually depends on data augmentation or a large number of filters. Large filters increase the complexity of model, which may be an obstacle for real-time segmentation on cloud computing. To solve these two issues, a new architecture, Lighter U-Net is proposed to reinforce feature use, to reduce the number of parameters as well as to retain sufficient receptive fields without losing resolution. The extensive experiments suggest that the proposed network achieves comparable performance as the state-of-the-art methods by only using 17% parameters of standard U-Net.

KEYWORDS

U-Net, DenseNet, White matter hyperintensities, Cloud computing

1 INTRODUCTION

White matter hyper-intensities (WMH), also known as leukoaraiosis or white matter lesions, refer to high intensity areas on T1 weighted or fluid-attenuated inversion recovery (FLAIR) human brain magnetic resonance imaging (MRI) scans, which are considered as lesions produced largely by demyelination and axonal loss. They are usually seen in healthy elderly people but also related to some neurological disorders and psychiatric illnesses. High quality of WMH segmentation is high desirable, but the manual segmentation is tedious, time-consuming and suffers from large inter and inter-observers variations. Automated WMH segmentation techniques in real-time will play a more and more important role in future clinical research and practice. As early as 2001, unsupervised learning method was implemented to detect sclerosis lesions by model outlier detection [24]. Approximately three years later, supervised learning approach was developed to segment different size of white matter lesions [1]. In 2008, Dyrby et al. employed neural networks to segment WMH [6]. Numerous automatic brain tissue segmentation algorithms have been developed after that [3, 7, 9, 15], but due to the huge heterogeneity and various brain abnormalities, WMH segmentation is still a challenging problem compared to general brain tissue segmentation [21].

Fully convolutional network (FCN) has achieved great success in semantic segmentation [20]. Especially in the field of biomedical image segmentation, a kind of FCN, so-called "U-Net" [23], is wildly used in recent two years. Compared to FCN and SegNet [2], U-Net has more elegant architecture which can yield more precise segmentations while using smaller training dataset. There are many deep learning based segmentation algorithms proposed for WMH segmentation [8, 14].

There is a trend that convolutional networks are getting deeper and deeper. Nevertheless, U-Net will lose more resolution as network goes deeper. In order to gain more precise performance, U-Net sometimes depends on data augmentation or larger numbers of filters instead of deeper network. Employing larger numbers of filters may make the network become more and more redundant. However, this kind of redundant may be a obstacle for real-time segmentation on cloud computing. The growing popularity of mobile terminals indicates that the network should be much lighter and less computation. Numerical researches attempt at model compression of image or video [11, 17–19].

The main purpose of this study is to solve these two challenges. One solution to resolution loss is using dilated convolution [5, 25]. Dilated convolution can retain sufficient receptive fields without losing resolution even if network goes deeper. Inspired by DenseNet [13], on the other hand, the study fuses dense connection with U-Net. This integration encourages feature reuse, making it possible to design light-weight architectures by reducing the number of filters.

Derived from those two points mentioned above, this paper proposes an advanced architecture, Lighter U-Net, which combines dense connection and dilated convolution. The advantage of Lighter U-Net can be concluded to "3R" Principles, which can reinforce

feature reuse, reduce the number of parameters and retain lager receptive fields.

At last, we restate our contribution here. This study proposes a new architecture, called Lighter U-Net, by integrating those two advanced structures with U-Net. The straightforward and light-weight architecture are demonstrated in WMH segmentation challenge with leading performance. The model achieved dice similarity coefficient (DSC) 93.42% in training set, 82.53% in validation set and 79.28% in testing set without data augmentation. Moreover, compared to the standard U-Net model, the model saves running time by 26% and also reduces the size of weight by 83%.

2 METHODS

2.1 Preliminaries

In general, U-Net consists of two symmetrical components, encoder (contracting on left side) and decoder (expanding on right side) with corresponding concatenations, which are designed to compensate the feature loss through max-pooling. To alleviate the shortcomings of U-Net mentioned above, we propose two novel modifications to U-Net, dense connection and dilated convolution. Dense connection is inspired by DenseNet [13]. Each layer uses all preceding layers as input, and its feature map will also be used as inputs to all subsequent layers. The dense connections attenuate the vanishing-gradient problem by promoting the efficiency of information flow.

Dilated convolution retains a large receptive field and avoids resolution loss from pooling (both upsampling and downsampling) without increasing the number of learned parameters. This operation can relieve resolution loss in U-Net architecture. This is particular important when segmenting delicate WMH structures while high level information is also helpful. The idea of dense connection and dilated convolution has been explored in many segmentation problems, such as multi-organ abdominal CT segmentation [10] and MRI Glioma segmentation [4].

2.2 Lighter U-Net

Our model is illustrated in Fig.1. Lighter U-Net derives from U-Net, which captures more features and thus yields precise segmentations by adding back previous corresponding layers. According to our model, those corresponding layers could be more than one. Assume that Lighter U-Net has n layers. Theoretically, n could be an infinite positive even number since U-Net usually has symmetrical layers. Practically, $n \geqslant 4$, i.e. $n := \{x \in Z^+ : 2x + 2\}$. Let's define the number of those corresponding layers in dense connection as r, where $r \in Z^+$. Based on this theory, dense connection in layer (n-a) is no more than r, where $a \in [1, \frac{n}{2} - 1]$. Let's define the dense connection space as $dc(i) := \{i \in [1, r] : a + r - i\}$. This space includes elements no more than i. The constrain comes from the bottom layer in the network, where $dc(i) \in [a, \frac{n}{2} - 1]$. Thus, Lighter U-Net can be defined as $F := \phi(n, r, f_n)$, where $\{f_n\}$ represents the sequence of the number of filters. Qualitatively speaking, Larger n leads to deeper network, which may make worse performance. Larger r causes the network more dense, which will definitely increase computation and occupy too much memory. Because of small dataset, this study uses $\phi(10, 2, f_n)$ to achieve best result.

Specifically, this network has 10 layers. The 5th layer implements dilated convolution, whose dilated rate is equal to 2. In the rest

Table 1: Descriptions of MICCAI WMH Challenge dataset, which contains 60 patients from 3 different scanners.

Dataset	Scanner	Images Size	No. Patients
Utrecht	3T Philips Achieva	240x240	20
Singapore	3T Siemens TrioTim	252x232	20
GE3T	3T GE Signa HDxt	132x256	20

layers from one to nine, batch normalization and pre-activation, LeakReLU, are applied before first convolutional layer. As mentioned in the technical report of DenseNet [22], batch normalization can provide a unique scale and bias to previous input while pre-activation can reduce the error significantly. In addition, those layers have two convolutional layers with the same number of filters. Kernel size is 3x3. The number of filters from layer one to four is [16, 32, 64, 128]. Symmetrically, The number of filters from layer six to nine is [128, 64, 32, 16]. The 10th layer is a output layer with sigmoid activation function. From layer one to four, 2D MaxPooling layer is used as downsampling. From layer six to nine, deconvolution is implemented to up-sampling with 2x2 kernel size and strides. Its number of filters reduces 50% compared to the number of filters in previous convolutional layer. Both convolution and deconvolution use "same" padding and "He Normal" initializer [12]. Compared to standard U-Net, most importantly, our model uses dense connection in layer seven to nine. For example, both layer one and two are concatenated back to layer nine after deconvolution. Other layers have the same pattern correspondingly.
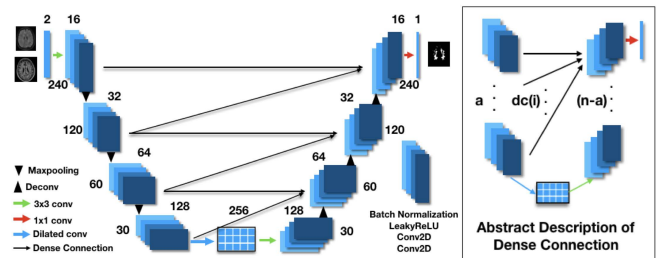

Figure 1: The Architecture of Lighter U-Net and The Abstract Description of Dense Connection.

3 EXPERIMENTS

3.1 Dataset

This dataset was acquired from three different hospitals in Netherlands and Singapore. Each of them contains 20 patients. For each patient, T1-weighted image, FLAIR image and manual label of the WMH regions were provided. An example of the image was illustrated in Figure 2. Details of dataset were described in Table 1. Five-fold validation was used for training and testing. In each fold, 80% of data were used to train the model while the rest 20% were used as testing set. Training images were further separated as 80% for training and 20% for validation. Sample slides of training/Validation/Testing were 1669/418/522.
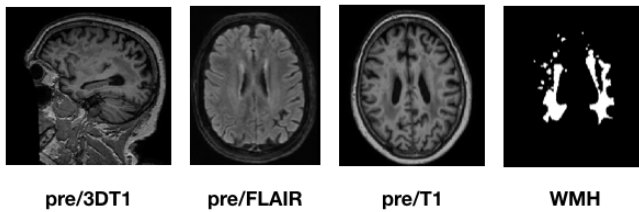
Figure 2: Left most image is the 3D T1 image (with mask removing the face); T1-weighted and FLAIR are used as two input channels; Right most image is the WMH manual label.

3.2   Data Preprocessing

T1 modality presents clear contrast among white matter, gray matter and cerebrospinal fluid. FLAIR is wildly used in brain imaging for highlighting the hyper-intense lesions. The combination of T1-weighted and FLAIR has been demonstrated to provide affable information on WMH detection and also significantly to increase the performance [6, 16]. We followed this suggestion and combined these two images into one two-channel image. Each patient's images were removed 1/8 along z-axis on both edges. After that, dataset was randomly split into training set and testing set on patientâĂŹs level. Due to large variation in image sizes using different scanners, we preprocessed the images to standard size 240 × 240 by padding or cropping. Lastly, different set of data were stacked into different arrays correspondingly. Note that here didn't use data augmentation.

3.3   Training

This experiment was conducted in the platform of Amazon web service Nvidia Tesla K80 with the instance, p2.xlarge. This instance used 4 virtual CPUs, 61 GB memory. The network was trained by Adam. Initial learning rate was set to be 1e-4 with step decay in the future epochs. Decay protocol might be flexible. Here the learning rate declined by 50% after 50 epochs and then dropped to 0.1 of initial rate after 80 epochs. The batch size was empirically set as 32, as batch size larger than 32 would trap the value of loss function in a very small number at early epoch and not update in the rest of training. As mentioned in Equation 1, given ground-truth y_g and predicted label y_p, DSC were selected to evaluate the performance as measuring the overlap between y_g and y_p in percentage was most suitable here. Negative DSC was tested to be the best loss function. Dropout gave worse results and wasn't utilized here. The model converged after 100 epochs' training.

$$DSC = \frac{2|y_g \cap y_p|}{(|y_g| + |y_p|)} \quad (1)$$

3.4   Five-Fold Cross-Validation

Table 2 shows the descriptive statistics of five-fold cross-validation. Our results substantially outperformed the baseline standard U-Net. Our model was also quite stable with a small standard deviation.

Table 2: Five-Fold Cross-Validation for DSC.

DSC mean±std	Training Set	Validation Set	Testing Test
Lighter U-Net	93.42% ± 1.39%	82.53% ± 1.25%	79.28% ± 0.57%
Standard U-Net	79.29% ± 1.44%	74.98% ± 1.50%	71.03% ± 1.25%

Table 3: DSC on different schemes for our model. P1(2): Position 1(2).

Evaluation	Preprocessing		Batch Normalization	
	Padding	Cropping	P1	P2
DSC	79.07%	72.20%	78.47%	77.03%

Fig.3 presents the differences between ground truth and predicted labels, baseline U-Net model and Lighter U-Net. The yellow arrows point the location where predictions are missing. Compared to baseline U-Net model, visually our model is more similar to the ground truth.
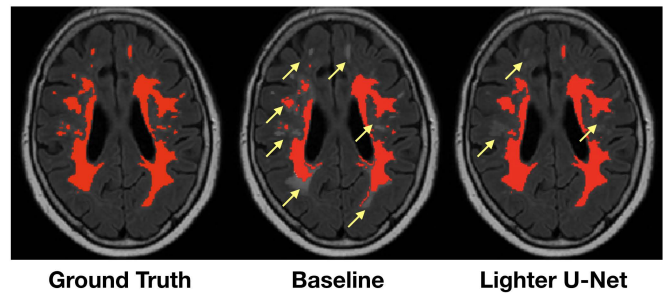

Figure 3: Ground truth, predictions from the baseline standard U-Net and Lighter U-Net, respectively.

3.5   Scheme Design

Table 3 reports scheme design of our architecture. U-Net is a symmetrical network. Due to dense concatenation, the size of input is related to the model's complexity. In this study, all inputs were resized to 240 by 240. Before resizing, the preprocessing has two possible options, padding and cropping. The result revealed that padding got much higher DSC than cropping. Thus, rest experiments used padding before resizing images.

Moreover, the position of batch normalization (BN) could affect the performance. In our model, two 2D convolutions (Conv2D) were operated in each layer. BN was attempted to execute in two positions: 1) In front of two Conv2Ds; 2) Between two Conv2Ds. This attempt illustrated that BN in position one attained better result.

3.6   Ablation Study

In this experiment, we demonstrated the effectiveness of the two major modifications, dense connections and dilated convolution. A significant advantage of utilizing dense connections was that it could reduce the number of parameters by encouraging the feature

Table 4: Comparison between Lighter U-Net and standard U-Net. DC: Dense Connection; DilConv: Dilated Convolution; 50%PM: Retain 50% parameters.

Model	Testing DSC	Avg. Time	Weight
Baseline U-Net	71.03%	92s	93.3Mb
DilConv+DC+100%PM	81.21%	141s	63.1Mb
DC+50%PM	77.38%	72s	24.6Mb
DilConv+DC+50%PM	79.28%	70s	16Mb

reuse. We investigated the contribution of dense connection and its influence on filter numbers. As shown in Table 4, the model utilizing dense connection, dilated convolution as well as the same number of filers as the standard U-Net achieved the best DSC, 81.21%. However, that model was very slow to be trained and tested. Our proposed Lighter U-Net used only half number of filters. It sacrificed testing DSC by about 2%, but cut the running time by half and reduced the model weight significantly.

In addition, the dilated convolution enlarges the receptive field without losing resolution or increasing computation. It is a powerful scheme compared to deconvolution. As shown in Table 4, the performance had increased approximately 2% after conducting the dilated convolution. The size of weight had an extra decline from 24.6 Mb to 16 Mb. This decline made the model lighter and portable. Dilated convolutional layer is a crucial component in the bottom of Lighter U-Net.

Table 4 summarizes the ablation study of Lighter U-Net. The dense connections enables good performance even with fewer filters. Compared to the standard U-Net, only half number of filters were used. This made our model much lighter, and faster to compute.

3.7 Comparison among Different Models

The competition of WMH Segmentation Challenge at MICCAI 2017, aimed at comparing different approaches for automatic segmentation of WMH. In this competition, the team "sysu_media" employed ensemble techniques and won the first place. This ensemble approach integrated n U-Net with random initialization and with shuffled batch. Multiple predicted labels were averaged and then transformed into a binary label under an empirical threshold. The paper regards this approach as a benchmark [16]. The result of competition is presented in Table 5. Our model achieved comparable performance as "sysu_media". Note that our preprocessing didn't use data augmentation, such as flipping or rotation. Indeed, preprocessing played an important role in the process of training. One observation proved this point. Three teams implemented the standard U-Net while only "sysu_media" achieved the best DSC. Another observation, on the other hand, illustrated that deeper networks did not guarantee better performance. Sometimes the more layers, the lower DSC. This observation enlightened us to make our model more lighter.

4 CONCLUSION AND FUTURE WORK

This study proposes a new architecture, Lighter U-Net, for WMH segmentation. Dense connection and dilated convolution make three main contributions: 1. Reinforce feature reuse; 2. Reduce

Table 5: Comparison of DSC among different models. Our model generates comparable results but is much smaller and faster to compute.

Team Name	Approach	DSC (Testing)
Our model	Lighter U-Net	0.79
sysu_media	Ensemble U-Net	0.80
cian	Multi-Dimensional GRU	0.78
nlp_logix	4-layer CNN	0.77
nic-vicorob	10-layer CNN	0.77
k2	U-Net	0.77
nih_cidi2	U-Net	0.75
lrde	Deep FCN	0.73
misp	3D 18-layer CNN	0.72
knight	Voxel-Wise Logistic Regression	0.70

the number of parameters and running time; 3. Retain sufficient receptive fields without losing resolution. The experiment shows that our model gains state-of-the-art performance as well as yields high quality prediction. However, this model has two shortcomings. Firstly, dense concatenation will occupy extra memory. What's more, this network is still too large and can be further compressed in order to achieve faster segmentation in real-time computing. Those are the plans for further improvement.

REFERENCES

[1] Petronella Anbeek, Koen L Vincken, Matthias JP van Osch, Robertus HC Bisschops, and Jeroen van der Grond. 2004. Automatic segmentation of different-sized white matter lesions by voxel probability estimation. *Medical image analysis* 8, 3 (2004), 205–215.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39, 12 (2017), 2481–2495.

[3] Maria Eugenia Caligiuri, Paolo Perrotta, Antonio Augimeri, Federico Rocca, Aldo Quattrone, and Andrea Cherubini. 2015. Automatic detection of white matter hyperintensities in healthy aging and pathology using magnetic resonance imaging: A review. *Neuroinformatics* 13, 3 (2015), 261–276.

[4] Lele Chen, Yue Wu, Adora M DSouza, Anas Z Abidin, Axel Wismüller, and Chenliang Xu. 2018. MRI tumor segmentation with densely connected 3D CNN. In *Medical Imaging 2018: Image Processing*, Vol. 10574. International Society for Optics and Photonics, 105741F.

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2016. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv preprint arXiv:1606.00915* (2016).

[6] Tim B Dyrby, Egill Rostrup, William FC Baaré, Elisabeth CW van Straaten, Frederik Barkhof, Hugo Vrenken, Stefan Ropele, Reinhold Schmidt, Timo Erkinjuntti, Lars-Olof Wahlund, et al. 2008. Segmentation of age-related white matter changes in a clinical multi-center study. *Neuroimage* 41, 2 (2008), 335–345.

[7] Daniel Garcia-Lorenzo, Simon Francis, Sridar Narayanan, Douglas L Arnold, and D Louis Collins. 2013. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical image analysis* 17, 1 (2013), 1–18.

[8] Mohsen Ghafoorian, Nico Karssemeijer, Tom Heskes, Inge WM Uden, Clara I Sanchez, Geert Litjens, Frank-Erik Leeuw, Bram Ginneken, Elena Marchiori, and Bram Platel. 2017. Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. *Scientific Reports* 7, 1 (2017), 5110.

[9] Mohsen Ghafoorian, Nico Karssemeijer, Inge WM van Uden, Frank-Erik de Leeuw, Tom Heskes, Elena Marchiori, and Bram Platel. 2016. Automated detection of white matter hyperintensities of all sizes in cerebral small vessel disease. *Medical physics* 43, 12 (2016), 6246–6258.

[10] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian Davidson, Stephen P Pereira, Matthew J Clarkson, and Dean C Barratt. 2018. Automatic multi-organ segmentation on abdominal CT with dense v-networks. *IEEE Transactions on Medical Imaging* (2018).

[11] Salih Burak Gokturk, Carlo Tomasi, Bernd Girod, and Chris Beaulieu. 2001. Medical image compression based on region of interest, with application to

colon CT images. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 3. IEEE, 2453–2456.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.

[13] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. 2017. Densely Connected Convolutional Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708.

[14] Dakai Jin, Ziyue Xu, Adam P Harrison, and Daniel J Mollura. 2018. White matter hyperintensity segmentation from T1 and FLAIR images using fully convolutional neural networks enhanced with residual connections. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 1060–1064.

[15] Hugo J Kuijf, Pim Moeskops, Bob D de Vos, Willem H Bouvy, Jeroen de Bresser, Geert Jan Biessels, Max A Viergever, and Koen L Vincken. 2016. Supervised novelty detection in brain tissue classification with an application to white matter hyperintensities. In *Medical Imaging 2016: Image Processing*, Vol. 9784. International Society for Optics and Photonics, 978421.

[16] Hongwei Li, Gongfa Jiang, Ruixuan Wang, Jianguo Zhang, Zhaolei Wang, Wei-Shi Zheng, and Bjoern Menze. 2018. Fully Convolutional Network Ensembles for White Matter Hyperintensities Segmentation in MR Images. *arXiv preprint arXiv:1802.05203* (2018).

[17] Ying Liu and Joohee Kim. 2019. Variable block-size compressed sensing for depth map coding. *Multimedia Tools and Applications* (2019), 1–15.

[18] Ying Liu and Dimitris A Pados. 2016. Compressed-sensed-domain l 1-pca video surveillance. *IEEE Transactions on Multimedia* 18, 3 (2016), 351–363.

[19] Ying Liu, Krishna Rao Vijayanagar, and Joohee Kim. 2014. Quad-tree partitioned compressed sensing for depth map coding. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 870–874.

[20] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.

[21] Pim Moeskops, Jeroen de Bresser, Hugo J Kuijf, Adriënne M Mendrik, Geert Jan Biessels, Josien PW Pluim, and Ivana Išgum. 2018. Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. *NeuroImage: Clinical* 17 (2018), 251–262.

[22] Geoff Pleiss, Danlu Chen, Gao Huang, Tongcheng Li, Laurens van der Maaten, and Kilian Q Weinberger. 2017. Memory-efficient implementation of densenets. *arXiv preprint arXiv:1707.06990* (2017).

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

[24] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan Colchester, Paul Suetens, et al. 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE transactions on medical imaging* 20, 8 (2001), 677–688.

[25] Fisher Yu and Vladlen Koltun. 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122* (2015).