Dilated Saliency U-Net for White Matter Hyperintensities Segmentation using Irregularity Age Map

Yunhee Jeong [1], **Muhammad Febrian Rachmadi** [1,2], **Maria del C. Valdés-Hernández** [2,*], **and Taku Komura** [1]

[1] *School of Informatics, University of Edinburgh, Edinburgh, UK*
[2] *Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK*

Correspondence*:
Maria del C. Valdés-Hernández
M.Valdes-Hernan@ed.ac.uk

ABSTRACT

White matter hyperintensities (WMH) appear as regions of abnormally high signal intensity on T2-weighted magnetic resonance image (MRI) sequences. In particular, WMH have been noteworthy in age-related neuroscience for being a crucial biomarker for Alzheimer's disease and brain aging processes. However, the automatic WMH segmentation is challenging because of the variable intensity range, size and shape. U-Net tackled this problem through the dense prediction and showed competitive performances on not only WMH segmentation/detection but also on varied image segmentation tasks, but it still accompanies a high complexity of the network architecture. In this study, we propose to use Saliency U-Net architecture and irregularity age map(IAM) to decrease the U-Net complexity without a performance loss. We trained Saliency U-Net using both T2-FLAIR MRI sequence and IAM. Since IAM guides where irregularities, in which WMH is possibly included, exist on the MRI slice, Saliency U-Net performs better than the original U-Net trained only using T2-FLAIR. The better performance was achieved with fewer parameters and shorter training time. Moreover, the application of dilated convolution enhanced Saliency U-Net to recognise the shape of large WMH more accurately by learning multi-context on MRI slices. This network named Dilated Saliency U-Net improved Dice coefficient score to 0.5588 which is the best score among our experimental models, and recorded a relatively good sensitivity of 0.4747 with the shortest train time and the least number of parameters. In conclusion, based on the experimental results, incorporating IAM through Dilated Saliency U-Net resulted an appropriate approach for WMH segmentation.

Keywords: White matter hyperintensities, Irregularity age map, Saliency U-Net, Dilation, MRI, Segmentation

1 INTRODUCTION

White matter hyperintensities(WMH) are commonly identified as signal abnormalities with intensities higher than other normal regions on T2-FLAIR magnetic resonance imaging(MRI) sequence. WMH have clinical importance in the study and monitoring of Alzheimer's disease (AD) and dementia progression(Gootjes et al., 2004). Higher volume of WMH has been found in brains of AD patients compared to age-matched controls, and the degree of WMH has been reported more severe for senile onset AD patients than presenile onset AD patients (Scheltens et al., 1992). Furthermore, WMH volume generally increases with the advance of age (Raz et al., 2012; Jagust et al., 2008). Due to their clinical importance, various machine learning approaches have been implemented for the automatic WMH segmentation (Admiraal-Behloul et al., 2005; Bowles et al., 2017).

LOTS-IM algorithm is an unsupervised algorithm for detecting tissue irregularities, that successfully has been applied for segmenting WMH regions on brain T2-FLAIR images (Rachmadi et al., 2019). Without any ground-truth segmentation, this algorithm produces a map which describes how much each voxel is irregular compared with an overall area. This map is usually called 'irregularity map' (IM) or 'irregularity age map (IAM)'. The concept of this map was firstly suggested in the field of computer graphics to calculate

pixel-wise age values indicating how weathered/damaged each pixel is compared to the overall texture pattern of an image (Bellini et al., 2016). Rachmadi et al. then proposed similar approach to calculate the irregularity level of WMH with respect to the ”normal” tissue in T2-FLAIR brain MRI (Rachmadi et al., 2017, 2018b). As WMH highlight irregular intensities on each T2-FLAIR slice, IAM can be also used for WMH segmentation. Although performing better than some conventional machine learning algorithms, LOTS-IM still underperforms compared to the state-of-the-art deep neural networks. This is mainly because IAM essentially indicates irregular regions, including artefacts and some grey matter regions, in addition to WMH. However, considering IAM depicts quite accurate appearances of irregularities and can be produced without training process, we propose to use IAM as an auxiliary guidance map of WMH location for WMH segmentation.

Recently, the introduction of deep neural networks, the state-of-art machine learning approach, has remarkably increased performances of image segmentation and object detection tasks. Deep neural networks outperformed conventional machine learning approaches in bio-medical imaging tasks as well as general image processing. For example, Ciresan et al. built a pixel-wise classification by deep neural networks to identify neuronal membrane on electron microscope (EM) images (Ciresan et al., 2012). In another study, Ronneberger et al. proposed a new deep neural network architecture called U-Net with a purpose of segmenting neuronal structures on EM images (Ronneberger et al., 2015).

In medical images’ segmentation tasks, U-Net architecture and its modified versions have been massively popular due to the end-to-end segmentation architecture and high performance. For instance, U-Net based fully convolutional network was proposed to automatically detect and segment brain tumor in multi-modal MRI data (Dong et al., 2017). A 3D U-Net for segmenting the kidney structure in volumetric images produced a good quality 3D segmentation results (Çiçek et al., 2016). UResNet, which is a combination of U-Net and residual network, was proposed to differentiate between WMH and brain lesions which disturb clear WMH segmentation (Guerrero et al., 2018). Zhang et al. trained a randomly initialised U-Net for WMH segmentation and improved the segmentation accuracy post-processing the results (Zhang et al., 2018b).

While there have been many previous studies showing that U-Net performs well on image segmentation, it has one shortcoming that is long training time due to its high complexity (Briot et al., 2018; Zhang et al., 2018a). To ameliorate this problem, Karargyros et al. suggested the application of regional maps as an additional input, for segmenting anomalies on CT images, and named their architecture Saliency U-Net (Karargyros and Syeda-Mahmood, 2018). They pointed out that extraction of relevant features from images unnecessarily demands very complex deep neural network architectures. Thus, despite neural networks architecture with large number of layers being able to extract more appropriate features from raw image data, it often accompanies a long training time and causes overfitting. Saliency U-Net has regional maps and raw CT images as input, and separately learns features from each data. The additional features from regional maps add spatial information to the U-Net, which successful delineates anomalies better than the original U-Net with less number of parameters (Karargyros and Syeda-Mahmood, 2018).

Another way to improve the segmentation performance of deep neural network architectures is through the recognition of the multi-scale context image information. Multi-scale learning is important particularly for detection/segmentation of objects with variable sizes and shapes. A dilated convolution layer was proposed to make deep neural networks learn multi-scale context better (Yu et al., 2017). Using dilated convolution layers, an architecture can learn larger receptive fields without significant increase in the number of parameters. previous studies have reported improvements using dilated convolution layers in medical image processing tasks (Lopez and Ventura, 2017; Moeskops et al., 2017).

In this paper, we propose to use IAM as an additional input data to train a U-Net neural network architecture for WMH segmentation. LOTS-IM can easily produce IAM without manually marked WMH ground-truth data and the need for training. U-Net architecture is selected as a base model for our experiments as it showed the best performance on learning IAM (Rachmadi et al., 2018a). To address the incorporation of IAM to U-Net for WMH segmentation, we propose the use of Saliency U-Net but feed-forwarding IAM as regional map to the Saliency U-Net. We also propose Dilated Saliency U-Net which is a combination of Saliency U-Net and dilated convolution to learn multi-scale context from both T2-FLAIR and IAM data. We compared the original U-Net’s performance with performances of Saliency U-Net and Dilated Saliency U-Net on performing WMH segmentation.

Consequently, the contributions of our work can be summarised as follows:

- Proposing the use of IAM as an auxiliary input for WMH segmentation. Feeding T2-FLAIR and IAM into neural networks complements each other by managing challenging cases of WMH segmentation, especially for small WMH.
- Integration of Saliency U-Net and dilated convolution for WMH segmentation which showed improvements on more detailed boundary delineation of large WMH. It also attained the best Dice coefficient score compared to our other experimental models.

2 MATERIALS AND METHODS

2.1 Dataset

MRI can produce different types of images to display the normal tissues and different types of abnormalities. It is desirable to choose suitable image types considering the properties of bio-markers or diseases targeted in the segmentation task. T2-weighted is one of the MRI sequences that emphasises fluids as bright intensities. The bright intensity of fluids makes WMH difficult to identify beacuse their intensities are also bright on T2-weighted. T2-fluid attenuated inversion recovery (T2-FLAIR) is generated by removing cerebrospinal fluid (CSF) signal from the T2-weighted sequence. This process makes more distinctive contrast between WMH and other brain tissues, so we have chosen T2-FLAIR as the main source of image data for our experiments.

We obtained T2-FLAIR MRI sequences from the public dataset *the Alzheimer's Disease Neuroimaging Initiative* (ADNI)[1] which was initially launched by Mueller et al. (Mueller et al., 2005). This study has mainly aimed to examine combinations of biomarkers, MRI sequences, positron emission tomography (PET) and clinical-neuropsychological assessments in order to diagnose the progression of mild cognitive impairment (MCI) and early AD. From the whole ADNI database, we randomly selected 60 MRI scans collected for three consecutive years from 20 subjects with different degrees of cognitive impairment. Each MRI scan has dimensions of $256 \times 256 \times 35$. We describe how train and test dataset are composed in Section 2.8.

2.2 Irregularity Age Map(IAM)

As described in Section 1, the concept of IAM was firstly proposed with the development of the LOTS-IM algorithm and its application to the task of WMH segmentation (Rachmadi et al., 2017, 2019, 2018b). This algorithm was inspired by the concept of 'age map' proposed by Bellini and colleagues while calculating the level of weathering or damage of pixels compared to the overall texture pattern on natural images (Bellini et al., 2016). Rachmadi et al. adopted this algorithm to compute the degree of irregularity in brain tissue from T2-FLAIR MRI.

In detail, the steps of LOTS-IM algorithm are as follows: source and target patches are extracted from given MRI slices with four different sizes (1×1, 2×2, 4×4 and 8×8). All fragments of grid which consists of $n \times n$ sized patches are regarded as source patches, while target patches are picked at random locations. Then, the difference between each source patch and one target patch on the same slice is calculated by Eq 1;

$$Difference = \theta \cdot |max(s-t)| + (1-\theta) \cdot |mean(s-t)| \tag{1}$$

where s and t mean a source patch and a target patch each, also θ was set to 0.5. For one each patch, only 100 largest difference values are averaged as an age value of the patch. After that, four age maps of different patch sizes are made of normalised age values between 0 and 1. Each age map is up-sampled into the original image size and then smoothed by Gaussian filter. The final age map is produced by blending these four age maps like Eqn 2.

$$Final\ age\ map = \alpha \cdot AM_1 + \beta \cdot AM_2 + \gamma \cdot AM_4 + \delta \cdot AM_8 \tag{2}$$

where AM_x means the age map of $x \times x$ sized patches and $\alpha + \beta + \gamma + \delta = 1$. Finally, the final age map is penalised by multiplying the original T2-FLAIR image slice to reflect high intensity of WMH, and globally normalised from 0 to 1 over all slices of one brain. The overall steps are demonstrated in Figure 1.

[1] http://adni.loni.usc.edu/

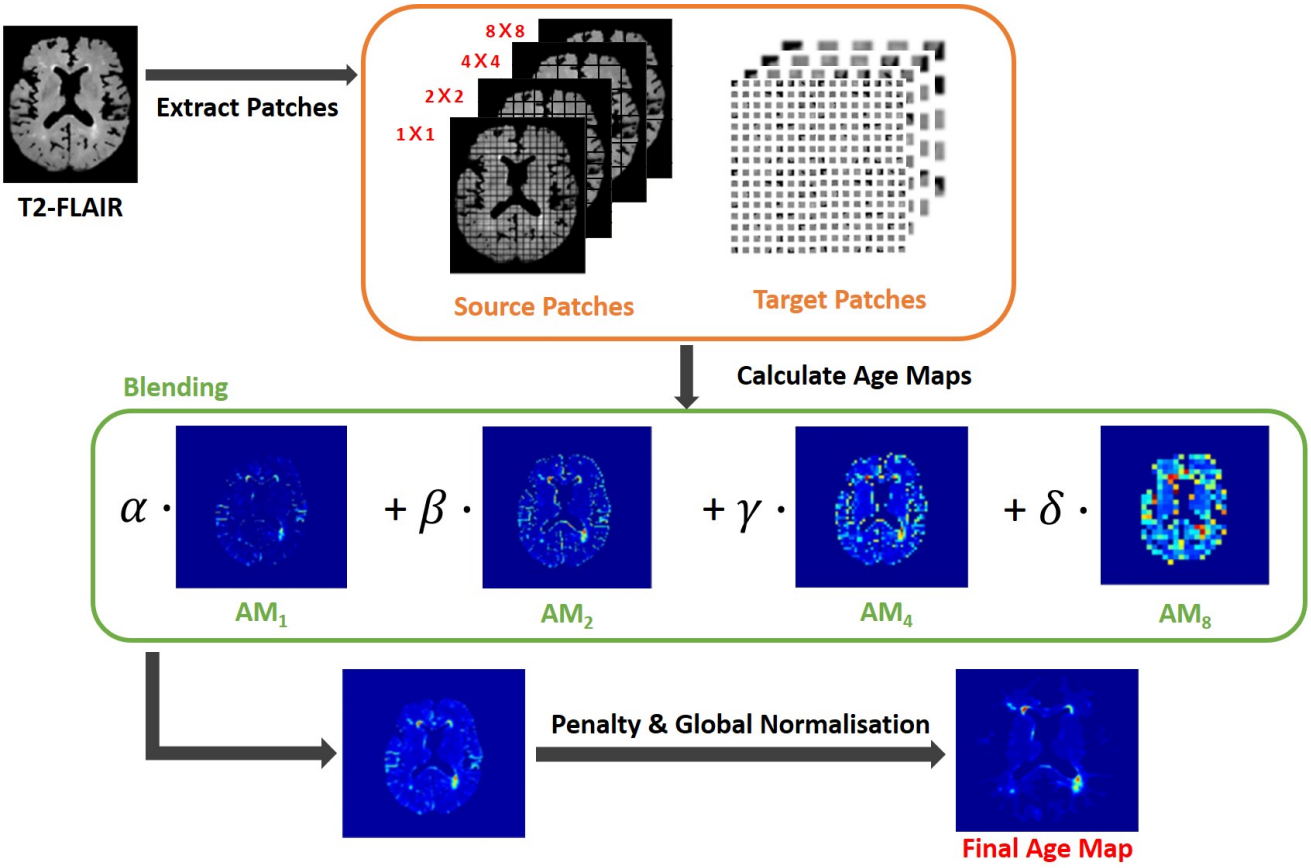

Figure 1. Flow chart illustrating the LOTS-IM algorithm proposed by Rachmadi and colleagues (Rachmadi et al., 2019) applied to WMH segmentation. This study uses the final map generated by this algorithm, and refers to it as "IAM data".

Though regarded as WMH segmentation map in the original paper, IAM essentially calculates the probability of each voxel to constitute an irregularity of the "normal" tissue. This irregular pattern includes not only WMH but more features such as artefacts, T2-FLAIR hyperintensities of other nature, as well as sections of the cortex that could be hyperintense. To compensate these flaws and take advantage of its usefulness, we developed a new scheme that uses IAM as an auxiliary guidance map for training deep neural networks rather than producing final WMH segmentation results. LOTS-IM algorithm, as an unsupervised learning method, generates IAM fast without the need of any manual process to create ground-truth data or training.

2.3 U-Net

Since U-Net architecture was firstly presented (Ronneberger et al., 2015), various image segmentation studies have been done based on its architecture due to its competitive performance regardless of the targeted object types. Different to the natural image segmentation, bio-medical image segmentation involves a more challenging circumstance where lack of data for training process is a common problem. U-Net deals with this challenge with dense prediction of the input image using up-sampling layers that produce equal-sized input and output. This approach was drew by fully convolutional networks (Long et al., 2015).

U-Net is comprised of two parts, the encoding part where feature maps are down-sampled by max-pooling layers and the decoding part where the reduced size of feature maps are up-sampled to the original size. It retains the localisation accuracy with contracting path which concatenates feature maps stored in the encoding part to the decoding part. These kept high resolution features help to restore details of localisation removed by max-pooling layer, when the feature maps are up-sampled in the decoding part. The architecture is depicted in Figure 3 (a).

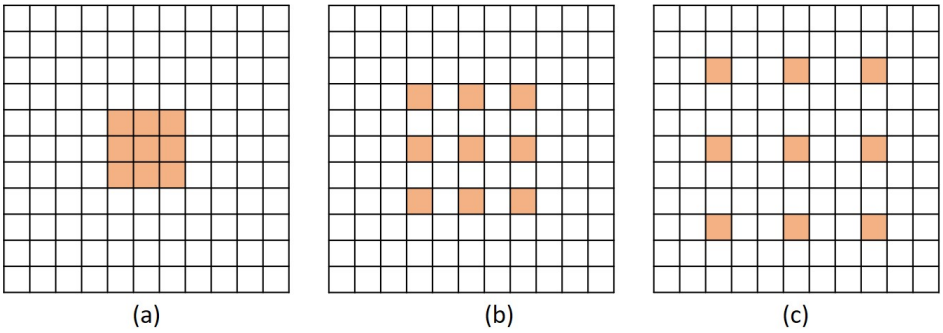

Figure 2. Examples of dilated convolution filter with 3×3 size. (a) Dilation factor = 1, (b) Dilation factor = 2 and (c) Dilation factor = 3.

A drawback of U-Net is its large number of parameters. To restore the high resolution localisation, it should increase the number of feature channels in the decoding part. Train time and memory usage is proportional to the number of parameters. So training U-Net is constrained to consume a lot of time and memory. Moreover, the complex neural network often induces overfitting problem.

2.4 Saliency U-Net

Saliency U-Net was first introduced to detect anomalies in medical images using a combination of raw (medical) images and simple regional maps (Karargyros and Syeda-Mahmood, 2018). Saliency U-Net performed better than U-Net while using less number of parameters. Architecture with less number of parameters is preferable as it is easier and faster to be trained. Karargyros and Syeda-Mahmood showed that convolution layers are not needed to extract more relevant features from raw images if auxiliary information of regional map is given as input. The Saliency U-Net architecture has two branches of layers in the encoding part (Figure 3 (b)). Each branch extracts features from raw image and regional map independently, and the extracted features are fused before the decoding part.

Segmentation results from Saliency U-Net showed more precise localisation and better performance than the original U-Net, which contained a larger number of convolution layers. Therefore, we proposed to use Saliency U-Net to take T2-FLAIR as raw input image and IAM as regional map for WMH segmentation.

2.5 Dilated Convolution

One common issue for image segmentation via deep neural networks is the use of a pooling layer to capture global contextual information by reducing the size of the feature maps. While the pooling layer is useful to get rid of some redundancies in feature maps, lower size of feature maps after pooling layer also causes loses in some its original detailed information, decreasing the segmentation performance where the targeted regions are not spatially prevalent(Hamaguchi et al., 2018; Yu et al., 2017).

Dilated convolution solved this problem by calculating a convolution over a larger region without reducing the resolution (Yu and Koltun, 2015). Dilated convolution layer enlarges a receptive field including k skips between each input pixel. k is called dilation factor. In numerical form, a dilated convolution layer with a dilation factor k and a $n \times n$ filter is expressed as follows:

$$F(r,c) = \sum_{x=-n}^{x=n} \sum_{y=-n}^{y=n} W(x,y) I(r+ki, c+kj) \tag{3}$$

Figure 2 (a)-(c) show examples of dilated convolution filters with dilation factor 1 to 3.

The additional advantage of dilated convolution is to widen the receptive field without increasing the number of parameters. Large receptive fields learn the global context by covering a wider area over the input feature map, but brings a memory leak and time consumption out for a growing number of parameters. Dilation can expand the receptive field of the convolution layer as much as skipped pixels without extra parameters. For instance, as shown in Figure 2 (a) and (c), the filter with dilation factor 3 has 7×7 sized receptive field, while the filter with dilation factor 1 has 3×3 sized receptive field.

In this study, we propose the incorporation of dilated convolution to Saliency U-Net for WMH segmentation. Since the size of WMH is variable, it is necessary to recognise different sizes of spatial context for more accurate delineation of WMH. We believe that dilated convolutions can manage varying size of WMH from different sizes of receptive field.

2.6 Our Experimental Models

We examined three different models in order to apply IAM as an auxiliary input data to learning WMH segmentation. The original U-Net model was trained using input data with two channels, T2-FLAIR and IAM each. As mentioned in Section 2.3, U-Net architecture has encoding and decoding parts. In the encoding part, input images or feature maps are down-sampled by max-pooling layers to obtain relevant features for WMH segmentation. Then, in the decoding part, reduced feature maps are up-sampled again by up-sampling layers to acquire the original size of final segmentation map. Max-pooling and Up-sampling layers are followed by two CONV blocks (yellow blocks in Figure 3). The CONV block contains a convolution layer, an activation layer and a batch normalisation layer. Batch normalisation allows to train neural networks with less careful initialisation and higher learning rate by performing normalisation every batch (Ioffe and Szegedy, 2015). All activation layers except the last one are ReLU (Nair and Hinton, 2010), but the last activation layer calculates the categorical cross-entropy to yield a probability map for each label.

On the other hand, we trained Saliency U-Net and Dilated Saliency U-Net by feed forwarding both T2-FLAIR and IAM separately. In this way, we assume that IAM works as a simple regional map which provides localisation information of WMH rather than just being a different image channel. While U-Net architecture has one branch of the encoding part, Saliency U-Net encoding part consists of two branches to learn raw images and regional maps individually. Furthermore, we applied dilation factors of 1, 2, 4 and 2 to the first four convolution layers of Saliency U-Net to form the Dilated Saliency U-Net. The architectures of U-Net, Saliency U-Net and Dilated Saliency U-Net can be seen in Figure 3.

Performance of these models are compared to each other in Section 3 and we additionally conducted experiments on the original U-Net model trained only with T2-FLAIR and only with IAM in order to see how using both T2-FLAIR and IAM as input affect to learning WMH segmentation. Our five experimental models are listed in Table 1.

2.7 Preprocessing

In machine learning, data preprocessing alters a given data into a comparable range. It is especially important when we deal with MRI data whose intensity is not in a fixed range. The difference of intensity is caused by the various MRI acquisition protocols, scanner models, calibration settings, etc.(Shah et al., 2011).

For this reason, we standardised the intensity of our train and test data by normalising them. Voxels only included in brain tissue were normalised. Non-brain tissue voxels whose intensity value is zero constitute the majority in each MRI slice. Thus normalising intensities from all voxels together can bias whole intensity values close to zero and reduce the effect of normalisation especially on brain tissue voxels. Brain tissue voxels were filtered using CSF and the intracranial volume (ICV) masks as follows:

$$Brain\,Tissue\,Region \;=\; MRI\;scan \;\cap\; (\,\neg\, CSF \;\cap\; ICV\,) \tag{4}$$

For the normalisation, we altered the brain tissue voxels on each slice into a distribution with zero-mean and unit variance by subtracting the mean value from eachvoxel value and dividing the result by the standard deviation.

Even though WMH segmentation can be regarded as the binary classification of voxels, we made the ground-truth data with three labels. Considering the intensity distributions, voxels would be divided into two main groups, brain tissue and non-brain. Brain tissue voxel intensity distribution, in WMH segmentation task, will have two sub-distributions comprised of WMH and non-WMH voxel intensities each. Therefore, we made ground-truth data marking three different regions of non-brain, non-WMH brain tissue and WMH. Figure 4 shows the example of a T2-FLAIR slice, the same slice after preprocessing and normalisation, and the ground-truth slice.

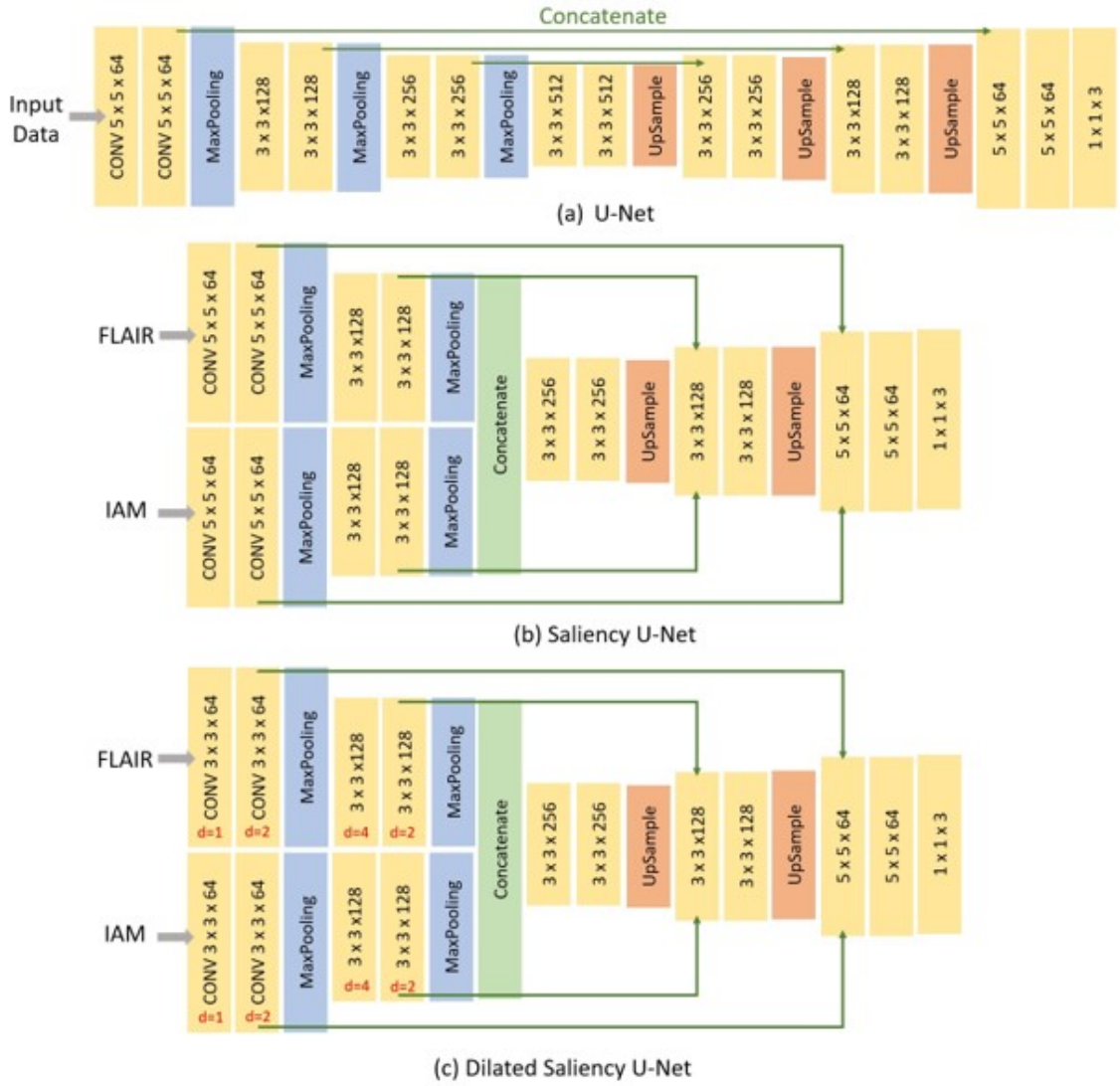

Figure 3. Architecture of three different networks used in this study. (a) the original U-Net, (b) Saliency U-Net and (c) Dilated Saliency U-Net. Three numbers of CONV block (yellow block) represents $filter\ size \times filter\ size \times filter\ channels$. For the Dilated Saliency U-Net model, red numbers mean a dilation factor for the convolution layer in each CONV block.

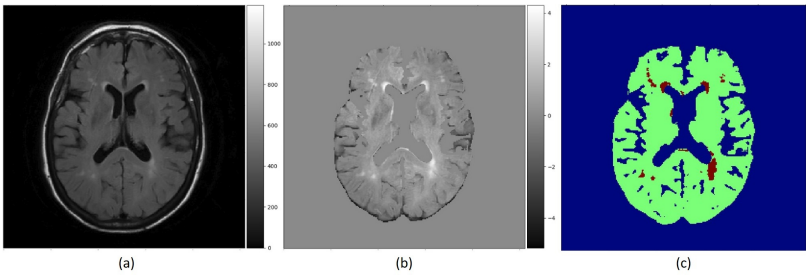

Figure 4. (a) Raw T2-FLAIR image, (b) T2-FLAIR input after preprocessing and normalisation, (c) Ground truth data with three labels. Blue region is non-brain area, green region is non-WMH brain tissues and red region is WMH.

2.8 Training and Testing Setup

For training, 30 MRI scans of the ADNI dataset described in Section 2.1 were randomly selected. These 30 MRI scans were collected from 10 subjects for three consecutive years. We trained our networks with image patches generated from these MRI scans, not slices, because of two reasons. First, if we train our models using slice images, the amount of training data is only $35 \times 30 = 1050$ slices, which is not ideal for training a deep neural networks architecture. Instead, by extracting 64×64 sized patches from whole image slices, we could have 30,000 patches for training data. Second, the ratio of WMH regions on one image slice is normally small which might cause bias towards the non-WMH regions distribution. In a smaller size of patch, the ratio of WMH regions to the non-WMH regions is more balanced.

For testing, we used the rest 30 scans of ADNI dataset, which are not used during training. These scans were also obtained from another 10 subjects for three consecutive years. The testing dataset was comprised of slice images without patch extraction. Slice image data is necessary to analyse the results from our models based on the distributions or volumes of WMH. Our testing dataset holds 1050 of 256×256 image slices in total as each scan contains 35 slices.

All experimental models were trained under equal configuration. We set learning rate to $1e^{-5}$ and batch size to 16. As an optimisation method, we selected the Adam optimisation algorithm (Kingma and Ba, 2014), although the original U-Net scheme used the stochastic gradient descent (SGD) optimiser. This is because the Adam optimiser can handle sparse gradients. It is highly possible that our training data produce sparse gradients as non-brain voxels, which are the majority, have zero intensity. We applied the Adam optimiser accordingly, considering this data property.

3 RESULTS

In this section, we present how experiments were processed, and analyse and compare the experimental results.

3.1 Evaluation Metrics

We use sensitivity and Dice similarity coefficient (DSC) to evaluate the tested models. Sensitivity measures the rate of true positive per whole postive voxels as below:

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

where TP means true positive, and FN means false negative. Whereas, DSC is a statistic method to compare the similarity between two samples including discrete values (Dice, 1945). It is useful to evaluate the segmentation result of one object which is grouping discrete labels of voxels or pixels. In particular, DSC is one of the most common evaluation metric in the medical image segmentation. The formula is as follow:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{6}$$

where TP and FN are equal to Equation 5 and FP means false positive. DSC is interpreted as the overlapping ratio to the whole area of prediction and target object, while sensitivity measures the correctly predicted region of target object. The difference between two metrics can be caused by false positive rate. If the prediction includes not only true positive but also wrong segmentation results (false positive), DSC score can decrease despite high sensitivity.

3.2 The Effects of IAM as an Auxiliary Input Data

Table 1 shows overall performances on our five experimental models. Adoption of IAM as an auxiliary input data for U-Net (i.e., U-Net(F+I)) improved sensitivity to 0.4902 but decreased the DSC score. On the other hand, Saliency U-Net(F+I) improved DSC score to 0.5535 while Dilated Saliency U-Net(F+I) achieved the best DSC score of 0.5588. For sensitivity, Dilated Saliency U-Net(F+I) yielded the second best rate after U-Net trained with T2-FLAIR and IAM (i.e., U-Net(F+I)). From these results, We can see that three models trained with T2-FLAIR and IAM particularly grew sensitivity performance.

Regarding the number of parameters, Saliency and Dilated Saliency U-Net include much less parameters than the other three U-Net models. As shown in Table 1, Saliency and Dilated Saliency U-Net has smaller amount of parameters and shorter training time than the original U-Net while having better if not similar performance on performing WMH segmentation. Consequently, this result presents that Dilated Saliency U-Net with T2-FLAIR and IAM as an auxiliary data is better than U-Net for WMH segmentation task.

Model	DSC	Sensitivity	Training Time	# Parameters
U-Net(FLAIR)	0.5440	0.4594	1h 52m 55s	7,859,715
U-Net(IAM)	0.5274	0.4179	1h 53m 52s	7,859,715
U-Net(F+I)	0.5281	**0.4902**	1h 24m 22s	7,861,315
Saliency U-Net(F+I)	*0.5535*	0.4730	1h 30m 1s	2,756,803
Dilated Saliency U-Net(F+I)	**0.5588**	*0.4747*	1h 4m 18s	2,623,683

Table 1. Dice Similarity Coefficient (DSC), train time and the number of parameters for our five experimental models. These results were yielded on whole train and test dataset described in Section 2.8. Value in bold is the highest score and in italic is the second highest one. In the bracket behind model names, input data type is specified. 'FLAIR' is equivalent to T2-FLAIR and 'F+I' refers to taking both T2-FLAIR and IAM as input.

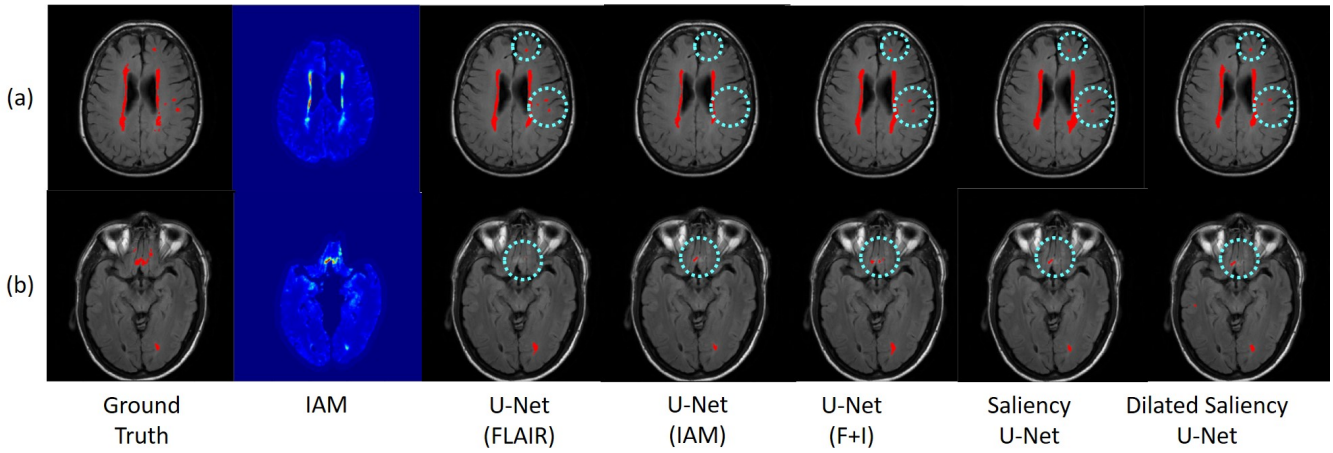

Figure 5. Visualisation of WMH segmentation results by our experimental models. Cyan circles are indicating WMH detected only by one model of either U-Net(FLAIR) or U-Net(IAM). Row (a) shows the case of WMH distinguishable exclusively on the T2-FLAIR image, while row (b) is the case that IAM highlights WMH more clearly. By training networks using T2-FLAIR and IAM, it is available to detect both cases of WMH region.

Figure 5 visualises WMH segmentation results by our experimental models. In most cases, training the network using both data could compensate each other in respect to detecting tricky WMH regions. Depending on the contrast/size of WMH or the quality of IAM, there are some cases of WMH which are distinguishable on IAM but unclear on T2-FLAIR and vice versa. For example, if WMH region is spatially too small, it is hard to differentiate on T2-FLAIR but more observable on IAM where WMH regions and normal brain tissue regions have better contrast due to normalised values of IAM. By contrast, when there exist other irregular patterns such as extremely low intensity from brain wrinkles around WMH, T2-FLAIR can indicate the WMH region more clearly than IAM. In Figure 5 Row (a), U-Net(FLAIR) produces better WMH segmentation result than U-Net(IAM) due to the poor quality of IAM. On the other hand, U-Net(FLAIR) could not detect WMH regions well caused by unclear intensity contrast of WMH regions on T2-FLAIR while U-Net(IAM) could segment this WMH regions supported by IAM marking this region as anomalies (Figure 5 Row (b)). Furthermore, incorporating both T2-FLAIR and IAM together for input data produced better WMH segmentation in general (5th-7th columns of Figure 5).

3.3 WMH Volume analysis

Group	Range of WMH volume (mm^3)	# Scans
Large	$10,000 \leq$ WMH Vol	6
Medium	$4,000 \leq$ WMH Vol $< 10,000$	10
Small	$1 \leq$ WMH vol $< 4,000$	14

Table 2. Criteria sorting MRI scans according to WMH voxel volume. '# Scans' means the number of included MRI scans. Most of scans are included in Small and Medium groups.

In this experiment, we evaluate our models based on the WMH volumes of the MRI scan (i.e., WMH burden) to examine the influence of WMH burden on the performance of WMH segmentation. WMH burden of each MRI scan is calculated by multiplying the number of WMH voxels and the unit voxel volume of the scan. We grouped MRI scans into three groups according to the range of WMH volume. Table 2 shows the range of WMH volume which become criteria of the groups and the number of scans included in each group. Figure 6 (a) shows that MRI scans of each group are obviously divided without any ambiguously classified scan.

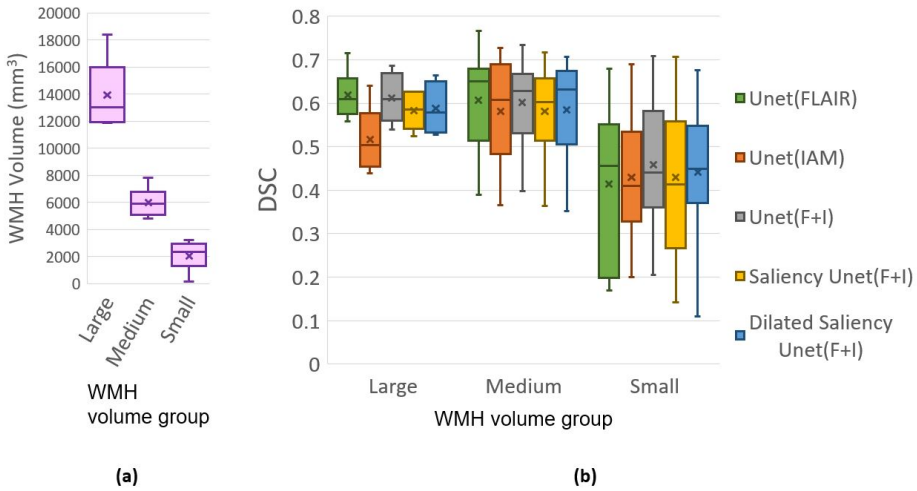

Figure 6. (a) Distributions of data (MRI scan) grouped together based on WMH volume. (b) DSC distributions yielded by tested five models based on WMH volume. 'x' and bar at the middle of box indicate mean and median each. Bottom and top of each box mean the first and third quartile.

Figure 6 (b) plots DSC scores yielded with MRI scans of different WMH volume groups by our five experimental models. This result was averaged over DSC scores of each MRI scan, not of slice batches which are used for overall performance evaluation in Section 3.2. Thus, the two evaluations could have a slightly different tendency. Also, we make clear that larger WMH volume does not exactly correspond to that the MRI scan contains one single large WMH. Scans of the Large group might have small but many WMH regions rather than one large WMH region.

All models tested in this study showed high median values of DSC scores in the Medium group, which is better than the other groups. In the Large group, U-Net(FLAIR) and U-Net(F+I) models performed similarly well, but U-Net(IAM) noticeably performed worse. Furthermore, U-Net(FLAIR) was the best model in the Medium group and U-Net(F+I) model could be regarded as the best model for the Small group. Although the median value of U-Net(FLAIR) and U-Net(F+I) are similar in the Small group, U-Net(F+I) showed higher average of DSC score than the others. Regarding Saliency and Dilated Saliency U-Net models, the application of dilation factors showed the improvement on DSC score in the Medium and Small groups with better distribution of the results. Mean, median and standard deviation (std.) values of DSC score distribution in each group are shown in Table 3

Model	DSC Mean.			DSC Median.			DSC std.		
	Large	Medium	Small	Large	Medium	Small	Large	Medium	Small
U-Net(FLAIR)	0.6184	0.6070	0.4147	0.6987	0.6499	0.4559	0.0524	0.1076	0.1746
U-Net(IAM)	0.5168	0.5817	0.4294	0.5036	0.6080	0.4106	0.0668	0.1111	0.1455
U-Net(F+I)	0.6124	0.6025	0.4580	0.6092	0.6276	0.4400	0.0548	0.0931	0.1460
Saliency U-Net(F+I)	0.5824	0.5812	0.4299	0.5853	0.6023	0.4134	0.0377	0.0956	0.1687
DSU-Net_1224	0.5722	0.5929	0.4003	0.5814	0.6286	0.3876	0.0592	0.0965	0.1733
DSU-Net_4221	0.5711	0.5768	0.4253	0.5776	0.6152	0.4250	0.0574	0.1097	0.1640
DSU-Net_1242	0.5882	0.5852	0.4407	0.5782	0.6320	0.4498	0.0536	0.1069	0.1558

Table 3. Mean, median and standard deviation values of DSC score distribution from our experimental models according to different WMH volume groups. DSU-Net_1242 was used for Dilated Saliency U-Net model evaluated in Section 3.3.

3.4 U-Net vs. Saliency U-Net

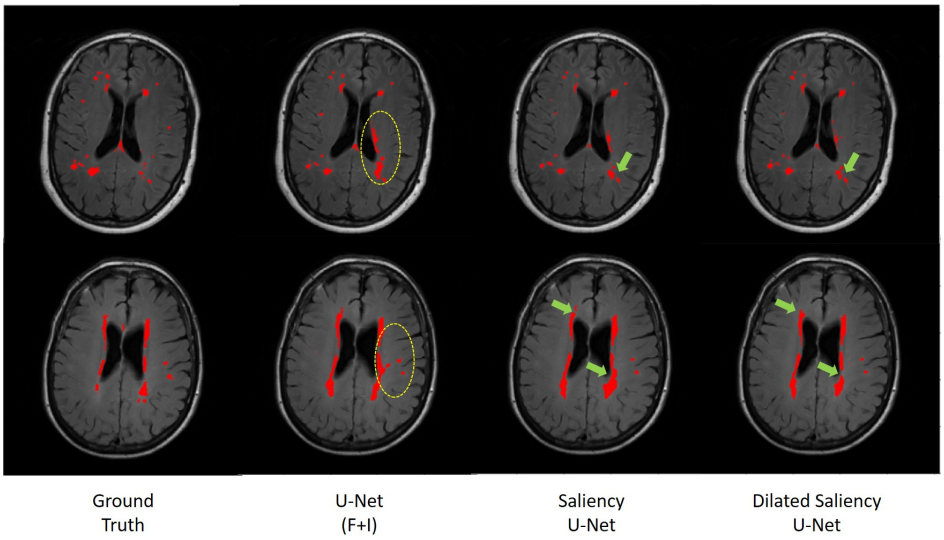

Figure 7. Comparison of WMH segmentation results from U-Net(F+I), Saliency U-Net and Dilated Saliency U-Net. Yellow circles are false positive results by U-Net(F+I). These false positive results are eliminated in the results from Saliency and Dilated Saliency U-Net. Green arrows are pointing that more detailed shape can be segmented by Dilated Saliency U-Net.

In order to figure out the effectiveness of Saliency U-Net architecture, we compare the original U-Net and Saliency U-Net models in which both of them are trained with T2-FLAIR and IAM. As shown in Table 1, Saliency U-Net yielded higher DSC score than U-Net(F+I) despite high sensitivity value of U-Net(F+I). This can be interpreted as U-Net(F+I) creates more false positives or false negatives. Figure 7 shows that Saliency U-Net successfully gets rid of some false positive segmentation from the U-Net(F+I) result.

We also investigated the change in the performance of Saliency U-Net's in relation to Saliency U-Net's complexity when the number of convolution layers increased/decreased. DSC score, training time and model complexity (i.e., the number of parameters) are compared in Figure 8. The rule of changing the complexity of Saliency U-Net is 2 CONV blocks are attached/detached at both ends and connect/disconnect them with skip connection. However, since the encoder part is a two-branch architecture, 6 CONV blocks are included at once when increasing the complexity (i.e., 4 CONV blocks are added to the encoder part and 2 CONV block are added to the decoder part). Similar approach is done when decreasing the complexity where 4 CONV blocks and 2 CONV block are dropped from the encoder and decoder respectively. For clarity, our original Saliency U-Net model (i.e., evaluated in Table 1 of Section 3.2) contains 14 CONV blocks and each CONV block holds one convolution layer as shown in Figure 3.

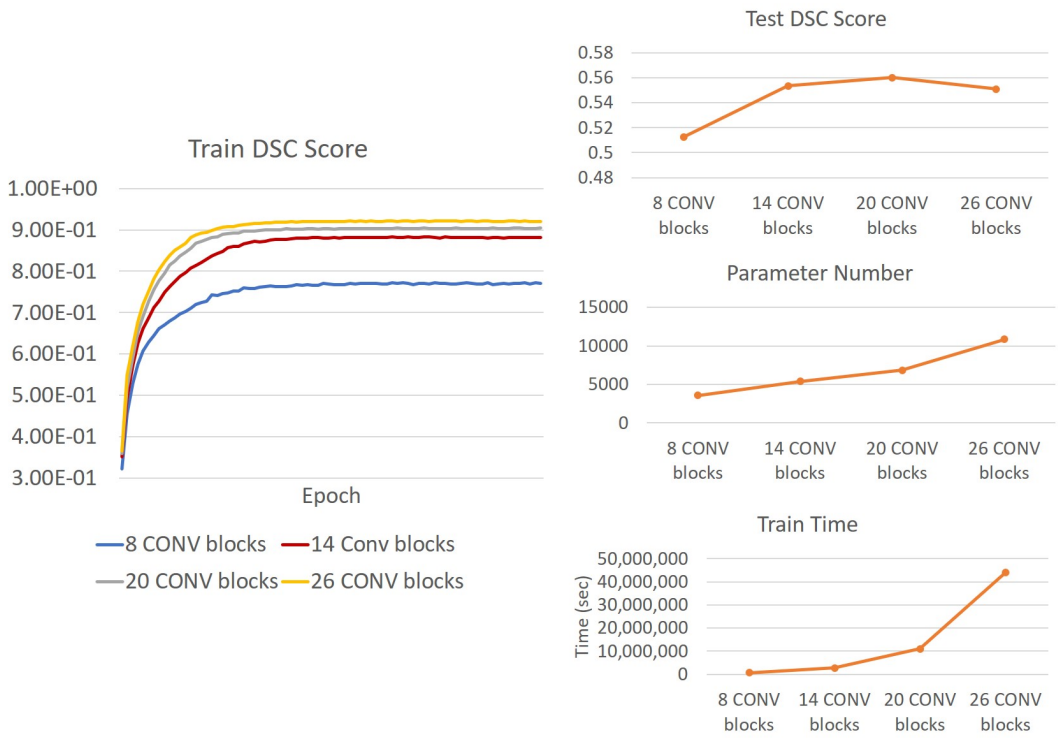

Figure 8. (Right) Trends of DSC score, training time and number of parameters of Saliency U-Net when more convolution layers are changed. It is also shown Saliency U-Net with 26 CONV blocks performance in testing **(upper Right)** and training **(Left)** decreases due to overfitting.

As shown in Figure 8, adding more CONV blocks means increasing both number of parameters and training time significantly. Furthermore, using too many CONV blocks (i.e., Saliency U-Net with 26 CONV blocks) decreased the DSC score due to overfitting problem.

3.5 Exploration of Dilated Saliency U-Net Architecture

In this experiment, we applied different dilation factors in Dilated Saliency U-Net which captures multi-context information on image slices without having to change the number of parameters. As shown in Figure 7, we can see that visualisation results from Saliency U-Net still involve poor boundary delineation specifically for large WMH regions. Furthermore, we also can see in Figure 7 that dilated convolutions help Saliency U-Net to detect more detailed shape of WMH regions. Because of that, it is important to know the influence of different dilated convolution configuration in Dilated Saliency U-Net for WMH segmentation.

In order to find the most appropriate configuration of dilation factors, we compared different sequences of dilation factors. Figure 3 (c) shows the base Dilated Saliency U-Net architecture used in this experiment. Only four dilation factors in the encoding part were altered while the rest of the parameters for the training schemes stayed the same. Yu and Koltun noted to use a fixed filter size for all dilated convolution layers but exponential dilated factors (e.g. $2^0,\ 2^1,\ 2^2$ ...) (Yu and Koltun, 2015). Therefore, we assessed 'increasing', 'decreasing' and 'increasing & decreasing' dilation factor sequences with factor numbers of 1, 2, 2, 4 and fixed filter size of 3×3. Details of these configurations are presented in Table 4. From this table, we can appreciate that despite DSU-Net_4221 performed best in DSC score metric (0.5622), it recorded the lowest sensitivity score. The best sensitivity metric was produced by DSU-Net_1242 (0.4747), but it did not outperform DSU-Net_4221 in DSC score metric.

Additionally, we investigated the influence of dilation factors in DSC score performance per WMH volume of MRI scans. Evaluation was conducted on three groups of the test dataset described in Table 2. Figure 9 shows that DSU-Net_1242 outperformed other models in every group. The report of mean, median and standard deviation of DSC score distribution in each group can be seen in Table 3.

Model	Encoder	DSC	Sensitivity
DSU-Net_1224 (Increasing)	CONV $3 \times 3 \times 64$, d = 1	0.5304	0.4395
	CONV $3 \times 3 \times 64$, d = 2		
	MaxPooling		
	CONV $3 \times 3 \times 128$, d = 2		
	CONV $3 \times 3 \times 128$, d = 4		
	MaxPooling		
DSU-Net_4221 (Decreasing)	CONV $3 \times 3 \times 64$, d = 4	**0.5622**	0.4381
	CONV $3 \times 3 \times 64$, d = 2		
	MaxPooling		
	CONV $3 \times 3 \times 128$, d = 2		
	CONV $3 \times 3 \times 128$, d = 1		
	MaxPooling		
DSU-Net_1242 (Increasing & Decreasing)	CONV $3 \times 3 \times 64$, d = 1	0.5588	**0.4747**
	CONV $3 \times 3 \times 64$, d = 2		
	MaxPooling		
	CONV $3 \times 3 \times 128$, d = 4		
	CONV $3 \times 3 \times 128$, d = 2		
	MaxPooling		

Table 4. Encoder architecture of Dilated Saliency U-Net with different dilation factors and their performances. Model name *DSU-Net_abcd* refers to Dilated Saliency U-Net model with dilation factors *a*, *b*, *c*, *d* in order from the first to the fourth convolutional layers, and its trend of dilation factor pattern is specified in the bracket. These dilation factors are applied on convolution layers in the encoding part (i.e., before concatenating T2-FLAIR and IAM feature maps) of CONV blocks, which consists of convolution, ReLU, and batch normalisation layers. Three numbers in the CONV block stans for 'filter size $\times$ filter size $\times$ filter number' and 'd' means a dilation factor.

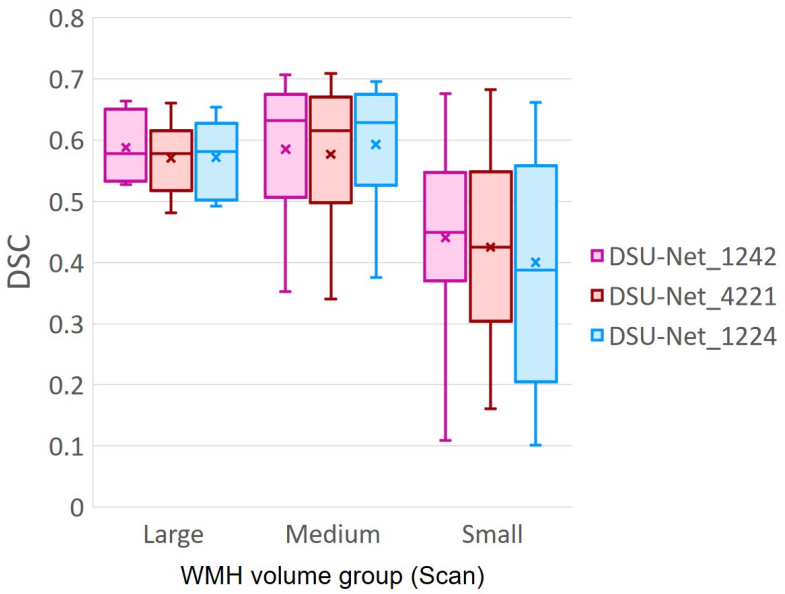

Figure 9. DSC score of three groups based on WMH volume in MRI scans. The group information is described in Table 2. 'x' and bar at the middle of box indicate mean and median each. Bottom and top of each box means the first and third quartile.

4 DISCUSSION

In this study, we explored the use of IAM as an auxiliary data to train deep neural networks for WMH segmentation. IAM produces a probability map of each voxel to be considered a textural irregularity

compared to other voxels considered "normal" (Rachmadi et al., 2019). While incorporating IAM as an auxiliary input data, we compared three deep neural network architectures to find the best architecture for the task, namely U-Net, Saliency U-Net and Dilated Saliency U-Net. It has been suggested that Saliency U-Net is adequate to learn medical image segmentation task with both raw and pre-segmented regional map (Karargyros and Syeda-Mahmood, 2018). The original U-Net did not improve DSC score despite using both T2-FLAIR and IAM as input, but the DSC score from Saliency U-Net was superior to that from the original U-Net trained only with T2-FLAIR. This is because Saliency U-Net is able to learn the joint encoding of two different distributions: i.e. from T2-FLAIR and IAM. Saliency U-Net generated better results than U-Net despite having less parameters. We also found that Saliency U-Net had lower false positive rate compared to U-Net.

Dilated convolution can learn spatially multi-context by expanding the receptive field without increasing the number of parameters. We added dilation factors to the convolution layers in the encoding block of Saliency U-Net to improve WMH segmentation, especially due to the high variability in the WMH size. This new model is named 'Dilated Saliency U-Net'. Dilated convolution improved both DSC score and sensitivity with shorter amount of training time. Dilated Saliency U-Net also yielded more accurate results in the presence of large WMH volumes and worked well in Medium and Small WMH volume MRI data groups which are more challenging. We identified that dilated convolution is effective when dilation factors are increased and decreased sequentially.

To our knowledge, this is the first attempt of successfully combining dilation, saliency and U-Net. We could reduce the complexity of a deep neural network architecture while increasing its performance through the integrated techniques and the use of IAM. Due to the trade-off between performance and training time, which is proportional to the model complexity, it is crucial to develop less complex CNN architectures without decreasing their performance.

Anomaly detection in the medical imaging field has been broadly studied (Quellec et al., 2016; Schlegl et al., 2017). One of its difficulties relies on the inconsistent shape and intensity of these anomalies. IAM helped the CNN scheme to overcome this problem by providing the localisation and morphological information of irregular regions. We believe it is possible to generate IAM from different modalities of medical images. Thus, the application of IAM is highly expandable to detect different imaging bio-markers involving abnormal intensity value in other diseases.

DATA AVAILABILITY STATEMENT

The code that corresponds with the experiments described and analysed in this manuscript can be found in `https://github.com/hanyangii/DilatedSaliencyUNet`.

CONFLICT OF INTEREST STATEMENT

All authors (Y.J., M.F.R., M.C.V.H., and T.K.) declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Y.J., M.F.R., M.C.V.H., and T.K. conceived and presented the idea. Y.J. and M.F.R. planned the experiments. Y.J. carried out the experiments. All authors provided critical feedback and analysis, and edited the manuscript.

FUNDING AND ACKNOWLEDGEMENTS

Funds from the Indonesia Endowment Fund for Education (LPDP) of Ministry of Finance, Republic of Indonesia (M.F.R.) and Row Fogo Charitable Trust (Grant No. BRO-D.FID3668413)(M.C.V.H.) are gratefully acknowledged. This project is partially funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) through the International Partnership Award BB/P025315/1 to M.C.V.H.

This study uses data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC.; Johnson and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

Admiraal-Behloul, F., Van Den Heuvel, D., Olofsen, H., van Osch, M. J., van der Grond, J., Van Buchem, M., et al. (2005). Fully automatic segmentation of white matter hyperintensities in mr images of the elderly. *Neuroimage* 28, 607–617

Bellini, R., Kleiman, Y., and Cohen-Or, D. (2016). Time-varying weathering in texture space. *ACM Transactions on Graphics (TOG)* 35, 141

Bowles, C., Qin, C., Guerrero, R., Gunn, R., Hammers, A., Dickie, D. A., et al. (2017). Brain lesion segmentation through image synthesis and outlier detection. *NeuroImage: Clinical* 16, 643–658

Briot, A., AI, G., Creteil, V., Viswanath, P., and Yogamani, S. (2018). Analysis of efficient cnn design techniques for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 663–672

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), 424–432

Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*. 2843–2851

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302

Dong, H., Yang, G., Liu, F., Mo, Y., and Guo, Y. (2017). Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *Annual Conference on Medical Image Understanding and Analysis* (Springer), 506–517

Gootjes, L., Teipel, S., Zebuhr, Y., Schwarz, R., Leinsinger, G., Scheltens, P., et al. (2004). Regional distribution of white matter hyperintensities in vascular dementia, alzheimer's disease and healthy aging. *Dementia and geriatric cognitive disorders* 18, 180–188

Guerrero, R., Qin, C., Oktay, O., Bowles, C., Chen, L., Joules, R., et al. (2018). White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. *NeuroImage: Clinical* 17, 918–934

Hamaguchi, R., Fujita, A., Nemoto, K., Imaizumi, T., and Hikosaka, S. (2018). Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE), 1442–1450

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*

Jagust, W. J., Zheng, L., Harvey, D. J., Mack, W. J., Vinters, H. V., Weiner, M. W., et al. (2008). Neuropathological basis of magnetic resonance images in aging and dementia. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 63, 72–80

Karargyros, A. and Syeda-Mahmood, T. (2018). Saliency U-Net: A regional saliency map-driven hybrid deep learning network for anomaly segmentation. In *Medical Imaging 2018: Computer-Aided Diagnosis*.

vol. 10575 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 105751T. doi:10.1117/12.2293976

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

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440

Lopez, M. M. and Ventura, J. (2017). Dilated convolutions for brain tumor segmentation in mri scans. In *International MICCAI Brainlesion Workshop* (Springer), 253–262

Moeskops, P., Veta, M., Lafarge, M. W., Eppenhof, K. A., and Pluim, J. P. (2017). Adversarial training and dilated convolutions for brain mri segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Springer). 56–64

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., et al. (2005). Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni). *Alzheimer's & Dementia* 1, 55–66

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 807–814

Quellec, G., Lamard, M., Cozic, M., Coatrieux, G., and Cazuguel, G. (2016). Multiple-instance learning for anomaly detection in digital mammography. *Ieee transactions on medical imaging* 35, 1604–1614

Rachmadi, M. F., del C. Valdés Hernández, M., and Komura, T. (2018a). Transfer learning for task adaptation of brain lesion assessment and prediction of brain abnormalities progression/regression using irregularity age map in brain MRI. In *PRedictive Intelligence in MEdicine - First International Workshop, PRIME 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings*. 85–93. doi:10.1007/978-3-030-00320-3_11

Rachmadi, M. F., Valdés Hernández, M., Li, H., Guerrero, R., Meijboom, R., Wiseman, S., et al. (2019). Limited one-time sampling irregularity map (lots-im): Automatic unsupervised quantitative assessment of white matter hyperintensities in structural brain magnetic resonance images. *bioRxiv* doi:10.1101/334292

Rachmadi, M. F., Valdés-Hernández, M. d. C., and Komura, T. (2017). Voxel-based irregularity age map (iam) for brain's white matter hyperintensities in mri. In *Advanced Computer Science and Information Systems (ICACSIS), 2017 International Conference on* (IEEE), 321–326

Rachmadi, M. F., Valdés-Hernández, M. d. C., and Komura, T. (2018b). Automatic irregular texture detection in brain mri without human supervision. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, eds. A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (Cham: Springer International Publishing), 506–513

Raz, N., Yang, Y., Dahle, C. L., and Land, S. (2012). Volume of white matter hyperintensities in healthy adults: contribution of age, vascular risk factors, and inflammation-related genetic variants. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1822, 361–369

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR* abs/1505.04597

Scheltens, P., Barkhof, F., Valk, J., Algra, P., HOOP, R. G. V. D., Nauta, J., et al. (1992). White matter lesions on magnetic resonance imaging in clinically diagnosed alzheimer's disease: evidence for heterogeneity. *Brain* 115, 735–748

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging* (Springer), 146–157

Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D. L., Collins, D. L., et al. (2011). Evaluating intensity normalization on mris of human brain with multiple sclerosis. *Medical image analysis* 15, 267–282

Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*

Yu, F., Koltun, V., and Funkhouser, T. (2017). Dilated residual networks. In *Computer Vision and Pattern Recognition*. vol. 1, 2

Zhang, C., Luo, W., and Urtasun, R. (2018a). Efficient convolutions for real-time semantic segmentation of 3d point clouds. In *2018 International Conference on 3D Vision (3DV)* (IEEE), 399–408

Zhang, Y., Chen, W., Chen, Y., and Tang, X. (2018b). A post-processing method to improve the white matter hyperintensity segmentation accuracy for randomly-initialized u-net. *arXiv preprint arXiv:1807.10600*