Limited One-time Sampling Irregularity Age Map (LOTS-IAM): Automatic Unsupervised Detection of Brain White Matter Abnormalities in Structural Magnetic Resonance Images

Muhammad Febrian Rachmadi[a,b], Maria del C. Valdés-Hernández[b], Hongwei Li[c], Ricardo Enrique Guerrero Moreno[d], Jianguo Zhang[c], Daniel Rueckert[d], Taku Komura[a]

[a] *School of Informatics, University of Edinburgh, Edinburgh, UK*
[b] *Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK*
[c] *Computing, School of Science and Engineering, University of Dundee, Dundee, UK*
[d] *Department of Computing, Imperial College London, London, UK*

Abstract

We propose a novel unsupervised approach of detecting and segmenting white matter abnormalities, using limited one-time sampling irregularity age map (LOTS-IAM). LOTS-IAM is a fully automatic unsupervised approach to extract brain tissue irregularities in magnetic resonance images (MRI) (*i.e.*, T2-FLAIR white matter hyperintensities (WMH)). In this study, the limited one-time sampling scheme is proposed and implemented on GPU. We compared the performance of LOTS-IAM in detecting and segmenting WMH with various methods, including *state-of-the-art* unsupervised WMH segmentation of Lesion Growth Algorithm from public toolbox Lesion Segmentation Toolbox (LST-LGA) and *state-of-the-art* supervised WMH segmentation of convolutional neural network (CNN) based methods. Based on our experiments, LOTS-IAM outperformed LST-LGA, the *state-of-the-art* unsupervised WMH segmentation method, both in performance and processing speed. Furthermore, our proposed method also outperformed conventional supervised machine learning algorithms of support vector machine (SVM) and random forest (RF), and supervised deep neural networks algorithms of deep Boltzmann machine (DBM) and convolution encoder network (CEN).

Keywords: white matter hyperintensities, lesion segmentation, unsupervised segmentation, dementia, MRI, voxel-based irregularity age map

1. Introduction

White matter hyperintensities (WMH) are common brain abnormalities found in brain magnetic resonance images (MRI) from patients with dementia/Alzheimer's Disease and other brain pathologies such as stroke and multiple sclerosis. WMH can be easily seen in T2-Fluid Attenuation Inversion Recovery (FLAIR) MRI as they appear brighter than the normal brain tissues. It is believed that WMH are associated with the progression of dementia (Wardlaw et al., 2013) and other comorbidities. Hence, not surprisingly, there have been many studies on methods for detecting or segmenting WMH automatically.

Supervised machine learning algorithms such as support vector machine (SVM), random forest (RF) (Ithapu et al., 2014) and deep learning convolutional neural network schemes, e.g. DeepMedic (Kamnitsas et al., 2017; Rachmadi et al., 2017a), uNet (Ronneberger et al., 2015; Li et al., 2018) and uResNet (Guerrero et al., 2018) have emerged as the *state-of-the-art* machine learning algorithms for automatic WMH segmentation. However, all supervised methods are highly dependent on manual labels produced by experts (*i.e.*, physicians) for training process. Furthermore, the quality of the label itself is dependent on and varies according to expert's skill and opinion, which rises questions about reproducibility in different sets of data. These intra/inter-observer inconsistencies usually are quantified and reported, but this does not solve the problem.

Unsupervised machine learning algorithms which do not need manual labels to work can eliminate the

aforementioned dependency. Methods such as Lesion Growth Algorithm from Lesion Segmentation Tool toolbox (LST-LGA) (Schmidt et al., 2012) and Lesion-TOADS (Shiee et al., 2010) have been developed, tested in many studies and publicly available for unsupervised WMH segmentation. Unfortunately, their performance is very limited compared to that from the supervised ones (Ithapu et al., 2014; Rachmadi et al., 2017a).

An unsupervised method named irregularity age map (IAM) (Rachmadi et al., 2017b) and its faster version one-time sampling IAM (OTS-IAM) (Rachmadi et al., 2018c) have been recently proposed and reported to work better than LST-LGA, which is still the most commonly used method and the *state-of-the-art* for unsupervised WMH segmentation. This study completes the development of this IAM method by evaluating all its parameters and using other metrics to analyse its quality and applicability.

In summary, the main contributions of this study are:

1. Proposing a new approach named limited one-time sampling IAM (LOTS-IAM) which is faster than IAM (*i.e.*, the original scheme) and OTS-IAM.
2. A comprehensive analysis and evaluation of LOTS-IAM, including (1) comparison between IAM, OTS-IAM and LOTS-IAM, (2) analysis of trade off trade off between speed and quality in LOTS-IAM, (3) analysis of LOTS-IAM's parameters and their performances, (4) analysis of LOTS-IAM's random sampling stability and (5) analysis of LOTS-IAM's performance on different size/volume of WMH.

2. Irregularity Age Map

The irregularity age map (IAM) approach for WMH assessment on brain MRI was proposed in our previous work (Rachmadi et al., 2017b). It is based on a previous work in computer graphics (Bellini et al., 2016) to detect aged or wandered regions in texture images. The term "age map" is used to name the 2D array with values between 0 and 1 that denote irregularities in textures dubbed as *age values*. The closer the value to 1, the more probable the image pixel/voxel is to belong to a group of clusters with different texture from that considered as the "norm". The age map can be calculated using, instead, structural MRI, to detect abnormal regions within normal tissue. For this process, four steps are necessary: 1) preparation of the regions of interest where the algorithm will work (e.g. brain tissue mask), 2) patch generation, 3) age value calculation and 4) final age map generation. These four steps are visualised in Fig. 1 and described in the rest of this section. Note that steps 2 to 4 are executed slice by slice (*i.e.* in 2D).

2.1. Brain tissue mask generation

For brain MRI scans, the brain tissue mask is necessary to exclude non-brain tissues not needed in the calculation of IAM and which can represent "irregularities" *per se*; for example skull, cerebrospinal fluid, veins and meninges. We want to compare and identify brain tissues within other brain tissues, not skull or other parts of non-brain tissues. For this purpose we use two binary masks: intracranial volume (ICV) and cerebrospinal fluid (CSF) masks, the latter containing also pial elements like veins and meninges. In our experiments, the ICV mask was generated by using optiBET (Lutkenhoff et al., 2014) while the CSF mask was generated by using an in-house algorithm developed by The University of Edinburgh (Valdés Hernández et al., 2015). However, several tools that produce accurate output exist and can be used for this purpose (*e.g.* bricBET[1], freesurfer[2]). The pre-processing step before computing LOTS-IAM only involves the generation of these two masks as per in the original IAM and in OTS-IAM (Rachmadi et al., 2017b, 2018b). This study also uses the normal appearing white matter (NAWM) mask to exclude brain non-white matter area, as per OTS-IAM (Rachmadi et al., 2018b). NAWM masks were generated using the FSL-FLIRT tool (Jenkinson et al., 2002), but can also be generated using in-house tools or freesurfer, for example.

2.2. Patch generation

Patch generation generates two sets of patches; non-overlapping grid-patches called *source patches* and randomly-sampled patches called *target patches*, which can geometrically overlap to each others. Each of *source patches* will be compared with all *target patches* using a distance function to calculate each *source patch*'s irregularity level. The

[1] `https://sourceforge.net/projects/bric1936/files/MATLAB_R2015a_to_R2017b/BRIClib/`

[2] `https://surfer.nmr.mgh.harvard.edu/`

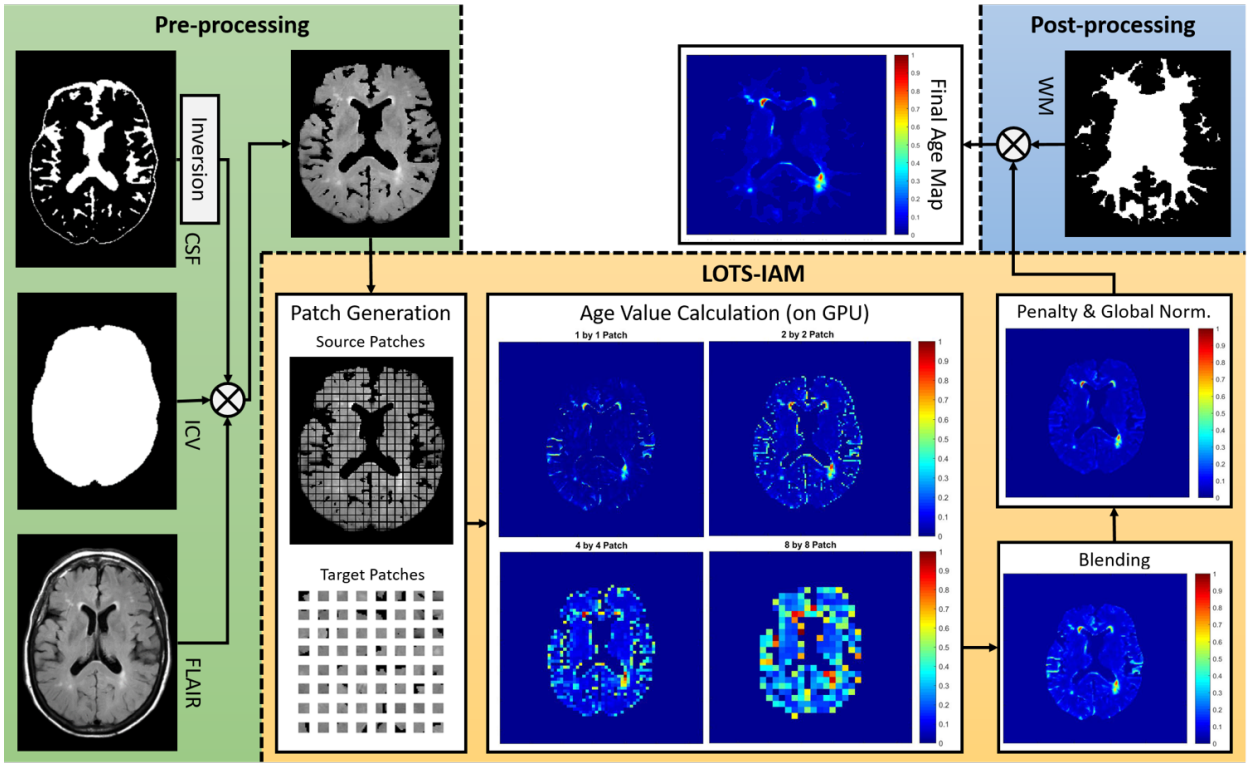

Figure 1: Flow of the proposed LOTS-IAM-GPU. **1) Pre-processing**: brain tissue-only T2-FLAIR MRI 2D slices are generated from the original T2-FLAIR MRI and its corresponding brain masks (*i.e.*, intracranial volume (ICV) and cerebrospinal fluid combined with pial regions (CSF)). **2) LOTS-IAM**: the brain tissue-only T2-FLAIR MRI slice is processed through the LOTS-IAM algorithm on GPU. **3) Post-processing**: final age map of the corresponding input MRI slice is produced after a post-processing step, which is optional.

rationale behind this is: if we successfully sample *target patches* mostly from normal brain tissues and calculate distance values between *source patch* and all *target patches*, then irregular textures located within the *source patch* will produce high absolute distance values for the respective *source patch*. In this study, we use hierarchical subsets of four different sizes of source/target patches which are 1×1, 2×2, 4×4 and 8×8. Unlike in the original study on natural images where all possible target patches are used to produce the age map (Bellini et al., 2016), we use a set of randomly sampled target patches to accelerate the computation.

2.3. Age value calculation

Age value calculation is the core computation of the IAM where a distance value called *age value* is computed by using the function defined below. Let $\mathbf{s}$ be a source patch and $\mathbf{t}$ a target patch, the *age value* of the two patches d is:

$$d = \alpha \cdot |\max(\mathbf{s} - \mathbf{t})| + (1 - \alpha) \cdot |\text{mean}(\mathbf{s} - \mathbf{t})| \quad (1)$$

where $\alpha = 0.5$ in this study. Both maximum and mean values of the subtracted patches are used to include maximum and average differences between source and target patches in calculation. Please note that source/target patches are matrices in the size of either 1×1, 2×2, 4×4 or 8×8. Also, please note that each *source patch* will be computed against a set of *target patches*, so each source patch has a set of age values. To get the final age value for one source patch, the corresponding set (*i.e.* to that source patch) of age values is sorted in ascending order and then the mean of the first 100 age values is calculated. The rationale is simple: the mean of the first 100 age values produced by an irregular source patch is still comparably higher than the mean of the first 100 age values produced by a normal source patches. All final age values from all source patches are then normalised to real values between 0 to 1 to create the age map for one MRI slice. Examples of age maps generated by using four different sizes of source/target patches are shown in Fig. 1

2.4. Final age map generation

The final age map generation consists of three sub-steps, which are *blending four age maps from age value calculation*, *penalty* and *global normalisation*. *Blending of four age maps* is performed by using the following formulation:

$$bAM = \alpha \cdot AM_1 + \beta \cdot AM_2 + \gamma \cdot AM_4 + \delta \cdot AM_8 \quad (2)$$

where $\alpha + \beta + \gamma + \delta$ is equal to 1 and AM_1, AM_2, AM_4 and AM_8 are age maps from 1×1, 2×2, 4×4 and 8×8 source/target patches. In this study, $\alpha = 0.65$, $\beta = 0.2$, $\gamma = 0.1$ and $\delta = 0.05$ as weight blending parameters. Before the blending, age maps resulted from different size of source/target patches are up-sampled to fit the original size of the MRI slice and then smoothed by using Gaussian filter. The blended age map is then *penalised* using formulation below:

$$p_o = p_i \times v_i \quad (3)$$

where p_i is voxel from the blended age map, v_i is voxel from the original MRI and p_o is the penalised voxel. Lastly, all age maps from different MRI slices are normalised together to produce 0 to 1 probability values of each voxel to be an "irregularity" with respect to the normal brain tissue. We name this normalisation procedure *global normalisation*. Visualisations of age value calculation, blending, penalty and global normalisation are shown in Fig. 1.

Some important notes on the computation of the IAM are: 1) source and target patches need to have the same size within the hierarchical framework, 2) the centre of source/target patches need to be inside ICV and outside CSF masks at the same time to be included in age value calculation and 3) the slice which does not provide any source patch is skipped to accelerate computation (i.e where no brain tissue is observed).

3. Limited one-time sampling IAM (LOTS-IAM)

While the original IAM has been reported to work well for WMH segmentation, its computation takes long time because it performs one sampling process for each source patch, selecting different target patches per source patch. For clarity, we dubbed this scheme as *multiple-time sampling* (MTS) scheme. MTS scheme is performed in the original IAM to satisfy the condition that target patches should not be too close to the source patch (*i.e.*, location based condition). MTS scheme makes every source patch to have its own set of target patches, so extra time to do sampling for each source patch is unavoidable.

To accelerate the overall IAM's computation, we propose one-time sampling (OTS) scheme for IAM where target patches are randomly sampled only once for each MRI slice, hence abandoning the location based condition of the MTS. In other words, age values of all source patches from one slice will be computed against one (i.e. the same) set of target patches. We call this combination of OTS and IAM one-time sampling IAM (OTS-IAM). The OTS-IAM was proposed in our previous study (Rachmadi et al., 2018c).

In this study, we propose to limit the number of target patches to accelerate the overall IAM's computation. The original IAM and OTS-IAM, which run on CPUs, use an undefined large random number of target patches which could range from 10% to 75% of all possible target patches, depending on the size of the brain tissue in an MRI slice. We name our new scheme limited one-time sampling (LOTS) IAM or LOTS-IAM. The LOTS scheme enables us to implement IAM on GPU to accelerate IAM's computation even more.

We tested 6 different numbers of target patch samples which are 2048, 1024, 512, 256, 128 and 64. Because it is possible for LOTS-IAM to use less than 100 target patch samples, we also modified the number of samples to be used to calculate the mean for age values. For LOTS-IAM, the first 128, 128, 64, 32, 32 and 16 age values are used to calculate the mean of age values for 2048, 1024, 512, 256, 128 and 64 number of target patch samples respectively. Limited number of samples in power of two eases GPU implementation, especially GPU memory allocation, which is the case for LOTS-IAM.

4. MRI Data, Other Machine Learning Algorithms and Experiment Setup

A set of 60 T2-Fluid Attenuation Inversion Recovery (T2-FLAIR) MRI data from 20 subjects from the ADNI database was used for evaluation. Each subject had three scans obtained in three consecutive years. All of them were selected randomly and blind to any clinical, imaging or demographic information. All T2-FLAIR MRI sequences have the same dimension of $256 \times 256 \times 35$. Full data acquisition information can be looked at in our previous study (Rachmadi et al., 2018b).Ground truth was produced semi-automatically by an expert in medical image analysis using the region-growing algorithm in the Object Extractor tool of Analyze^{TM} software guided by the co-registered T1- and T2-weighted sequences. For more details of the dataset, please see (Rachmadi et al., 2017a) and data-share url[3] to access the dataset.

Data used in this study were obtained from the ADNI (Mueller et al., 2005) public database[4]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found here[5].

We compare performances of LOTS-IAM with other machine learning algorithms that are commonly used for WMH segmentation; namely the original IAM, One-time Sampling IAM (OTS-IAM), Lesion Growth Algorithm from Lesion Segmentation Tool (LST-LGA), support vector machine (SVM), random forest (RF), deep Boltzmann machine (DBM), convolutional encoder network (CEN), patch-based 2D CNN with global spatial information (2D patch-CNN-GSI), patch-uResNet and patch-uNet. LST-LGA (Schmidt et al., 2012) is the current *state-of-the-art* for unsupervised hyperintensities segmentation. SVM and RF are machine learning algorithms commonly used for WMH segmentation in several studies (Rachmadi et al., 2017a), and they are used in this study as representations of supervised conventional machine learning algorithms. On the other hand, DBM, CNN and U-Net based methods represent supervised deep learning algorithms which are commonly used in recent years for WMH segmentation. For clarity, we do not further elaborate in the implementation of these algorithms. All experiments' setup (*i.e.*, training/testing and algorithm's configurations) for SVM, RF, DBM and CEN algorithms are described in detail in (Rachmadi et al., 2017a).

[3] `http://hdl.handle.net/10283/2214`

[4] `http://adni.loni.usc.edu/`

[5] `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

Whereas, experiments' setup for the 2D Patch-CNN-GSI, patch-uResNet and patch-uNet follow previous studies of (Rachmadi et al., 2018b), (Guerrero et al., 2018) and (Li et al., 2018) respectively.

Dice similarity coefficient (DSC) (Dice, 1945), which measures similarity between ground truth and automatic segmentation results, is used here as the primary metric for comparison between algorithms. Higher DSC score means better performance, and the DSC score itself can be computed as follow:

$$DSC = \frac{2 \times TP}{FP + 2 \times TP + FN} \quad (4)$$

where TP is true positive, FP is false positive and FN is false negative.

Additional metrics positive predictive value (PPV), specificity (SPC) and true positive rate (TPR) are also calculated. Non-parametric Spearman's correlation coefficient (Myers et al., 2010) is used to compute correlation between WMH volume produced by each automatic method and visual ratings of WMH. Visual ratings of WMH are commonly used in clinical studies to describe and analyse severity of white matter disease (Scheltens et al., 1993). Correlation between visual ratings and volume of WMH is known to be high (Hernández et al., 2013). In this study, Fazekas's (Fazekas et al., 1987) and Longstreth's visual rating scales (Longstreth et al., 1996) are used for non-parametric evaluation of each automatic method. This non-parametric test is similar as in previous study (Rachmadi et al., 2017a).

5. Experiments and Results

5.1. General Experiment Results

Table 1 shows the overall results of the performance of all methods that we have tested. Please note that the original IAM is listed as IAM-CPU.

From Table 1, we can see that all IAM configurations (*i.e.*, IAM-CPU, OTS-IAM-CPU and LOTS-IAM-GPU methods) outperformed LST-LGA in mean DSC, positive predictive value (PPV), specificity (SPC) and true positive rate (TPR) metrics. Furthermore, we also can see that performances of IAM/OTS-IAM/LOTS-IAM not only outperformed LST-LGA but also some other supervised machine learning algorithms (*i.e.*, SVM, RF and DBM). Moreover, some LOTS-IAM-GPU implementations also successfully outperformed CEN in four evaluated metrics and performed better than CNN based algorithms (*i.e.*, CEN, Patch-uResNet, Patch-uNet and 2D Patch-CNN-GSI) in either PPV, SPC or TPR metrics. The best value for each evaluation metric is written in bold letters.

Evaluation metric values listed in Table 1 are extracted by using optimum threshold value of each algorithm. In this study, the optimum threshold value is based on DSC metric. We produced DSC performance curves for each algorithm similar to Figure 2. Figure 2 itself shows (mean) DSC performance curves for LST-LGA, SVM, RF, DBM, CEN, patch-uResNet, patch-uNet, 2D patch-CNN-GSI and LOTS-IAM-GPU-512s64m for all possible threshold values to their probabilistic output.

A visual example showing the performance of IAM and other segmentation methods such as 2D Patch-CNN (2D-DeepMedic), Patch-uNet, LST-LGA and Patch-uResNet are compared in Figure 4 (raw) and Figure 5 (cut off). Figure 6 shows how LOTS-IAM could potentially be applied to characterise abnormalities using other MRI sequences, such as T1-weighted (T1W).

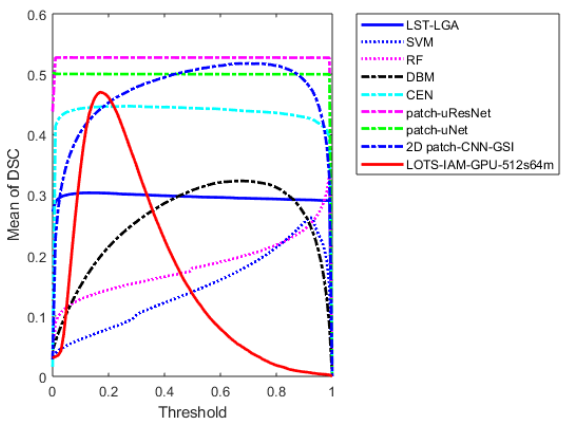

Figure 2: Mean of dice similarity coefficient (DSC) score for LST-LGA, SVM, RF, DBM, CEN, patch-uResNet, patch-uNet, 2D patch-CNN-GSI and LOTS-IAM-GPU-512s64m in respect to all possible threshold values. LOTS-IAM is represented by LOTS-IAM-GPU-512s64m.

5.2. IAM vs. OTS-IAM vs. LOTS-IAM

One-time sampling (OTS) and limited one-time sampling (LOTS) not only successfully accelerated IAM's computation but also improved IAM's performance, as shown in Table 1. Implementation of IAM on GPU successfully accelerated IAM's processing speed by 17 to 435 times with respect to the

Table 1: Algorithm's information and experiment results based on the Dice similarity coefficient (DSC), positive predictive value (PPV), specificity (SPC) and true positive rate (TPR) for each algorithm evaluated (the best value is written in **bold**). **Explanation of abbreviations**: "SPV/UNSPV" for supervised/unsupervised, "Deep Net." for deep neural networks algorithm, "Y/N" for Yes/No, "T2F/T1W" for T2-FLAIR/T1-weighted, "#MTPS" for maximum number of target patches, "#meTPS" for number of target patches used for calculating mean of age value, "TRSH" for optimum threshold and "Training/Testing" for training/testing time. Given "speed increase" is relative to IAM-CPU.

No	Method	SPV/ UNSPV	Deep Net.	Input Modality	#MTPS/ #meTPS	TRSH	DSC		PPV (Mean)	SPC (Mean)	TPR (Mean)	Training (min)	Testing (min)	Speed increase
							Mean	Std						
1	LST-LGA	UNSPV	N	T2F-T1W	-	0.134	0.3037	0.1658	0.3158	0.9946	0.3625	-	0.67	-
2	SVM	SPV	N	T2F-T1W	-	0.925	0.2630	0.1498	0.0474	0.9869	0.1259	26	1.38	-
3	RF	SPV	N	T2F-T1W	-	0.995	0.3633	0.1843	0.0482	0.9860	0.1320	37	0.68	-
4	DBM	SPV	Y	T2F	-	0.687	0.3235	0.1345	0.0642	0.9955	0.0542	1,341	0.28	-
5	CEN	SPV	Y	T2F	-	0.284	0.4308	0.1582	0.5255	0.9975	0.4815	152	0.08	-
6	Patch-uResNet	SPV	Y	T2F	-	0.200	**0.5277**	**0.1729**	0.5899	0.9970	**0.5968**	215	0.08	-
7	Patch-uNet	SPV	Y	T2F	-	0.200	0.5030	0.1487	0.6480	0.9985	0.4886	211	0.08	-
8	2D Patch-CNN-GSI	SPV	Y	T2F	-	0.801	0.5225	0.1690	0.5950	0.9985	0.5276	392	0.45	-
9	IAM-CPU	UNSPV	N	T2F	75%/100	0.179	0.3930	0.1732	**0.7001**	**0.9993**	0.3757	-	217.18	-
10	OTS-IAM-CPU	UNSPV	N	T2F	75%/100	0.164	0.4297	0.1734	0.6994	0.9992	0.3827	-	173.50	1.26
11	LOTS-IAM-GPU-2048s128m	UNSPV	N	T2F	2,048/128	0.159	0.4346	0.1764	0.6904	0.9991	0.3915	-	12.43	17.52
12	LOTS-IAM-GPU-1024s128m	UNSPV	N	T2F	1,024/128	0.179	0.4432	0.1784	0.6259	0.9986	0.4448	-	4.00	54.45
13	LOTS-IAM-GPU-512s64m	UNSPV	N	T2F	512/64	0.179	0.4437	0.1813	0.6093	0.9983	0.4551	-	1.42	153.73
14	LOTS-IAM-GPU-256s32m	UNSPV	N	T2F	256/32	0.194	0.4433	0.1785	0.5867	0.9981	0.4730	-	0.75	290.38
15	LOTS-IAM-GPU-128s32m	UNSPV	N	T2F	128/32	0.213	0.4373	0.1862	0.4909	0.9959	0.5481	-	0.60	362.97
16	LOTS-IAM-GPU-64s16m	UNSPV	N	T2F	64/16	0.228	0.4284	0.1854	0.4370	0.9939	0.5849	-	0.50	435.57

original IAM-CPU. However, it is worth stressing that this increase in processing speed was not only due to the use of GPU instead of CPU, but also depending on the number of target patch samples used in the IAM's computation. One of the GPU implementations of LOTS-IAM (*i.e.*, LOTS-IAM-GPU-64s16m) ran faster than LST-LGA. Note that the testing time listed in Table 1 excludes registrations and the generation of other brain masks used either in pre-processing or post-processing steps. The increase in speed achieved by the GPU implementation of IAM shows the effectiveness of the LOTS implementation for IAM's computation and performance.

5.3. Evaluation of Speed vs. Quality in LOTS-IAM

The biggest achievement of this work is the increase in processing speed achieved by the implementation of LOTS-IAM on GPU, compared to the original IAM and OTS-IAM. The first iteration of IAM can only be run on CPU because it uses multiple-time sampling (MTS). OTS-IAM samples patches only once, but still uses a high number of target patches (*i.e.*, 2,048 samples) to compute the age map. In this study, we show that using a limited number of target patches leads to not only faster computation but also better quality of WMH segmentation in some cases based on mean of DSC. Figure 3 illustrates the relation between speed and quality of the output (mean of DSC) produced by IAM, OTS-IAM and all configurations of LOTS-IAM. Please note that Figure 3 is extracted from Table 1. On the other hand, LOTS-IAM-GPU using more target patches produced better PPV and SPC evaluation metrics than LOTS-IAM-GPU using less target patches. This case is then reversed in TPR metric where using less target patches is better than using more target patches.

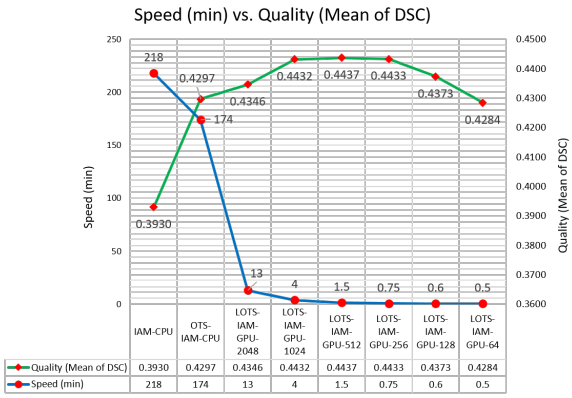

Figure 3: Speed (min) versus quality (Mean of DSC) of different settings of IAM methods (extracted from Table 1). By implementing IAM on GPU and limiting number of target patch samples, computation time and result's quality are successfully improved.

5.4. Analysis of IAM's Blending Weights

In this experiment, different sets of blending weights in IAM's computation were evaluated. As previously discussed, the only parameters that IAM has are four weights used to blend the four age maps hierarchically produced for the generation of the final age map. We tested the 7 different sets of IAM's blending weights listed in Table 2. The first 3 sets blend all four age maps while the other 4 only use

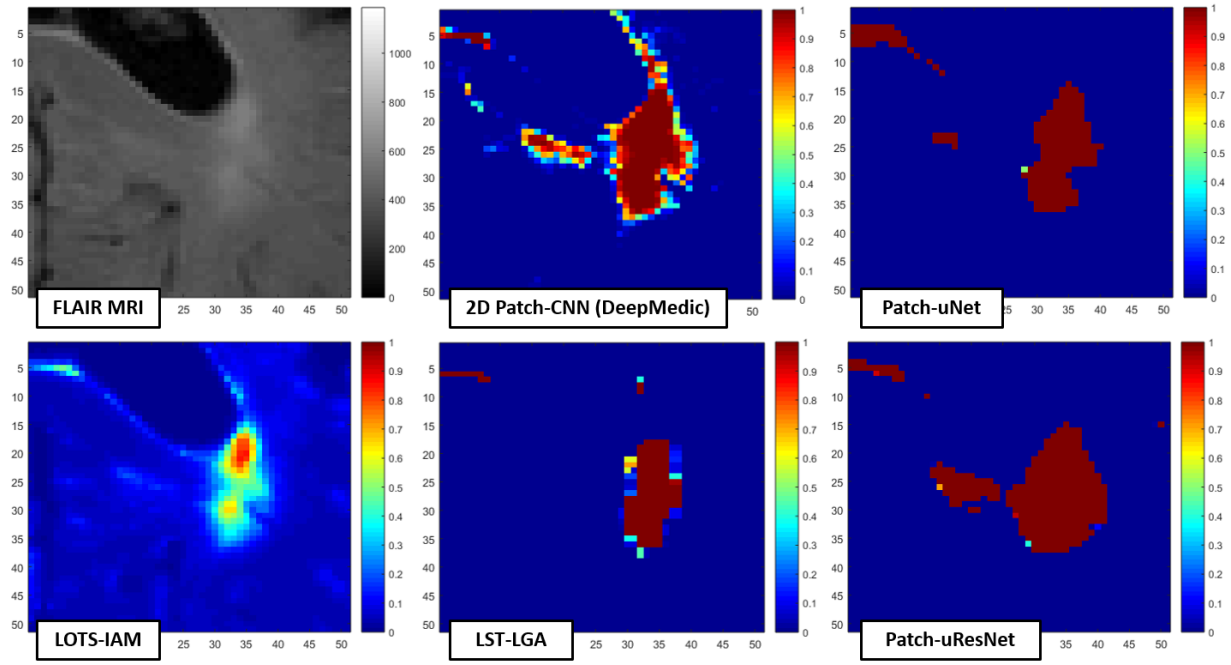

Figure 4: Visualisation of **raw** probabilistic values of WMH produced by LOTS-IAM compared to other methods, which are the 2D Patch-CNN (2D-DeepMedic), Patch-uNet, LST-LGA, and Patch-uResNet.

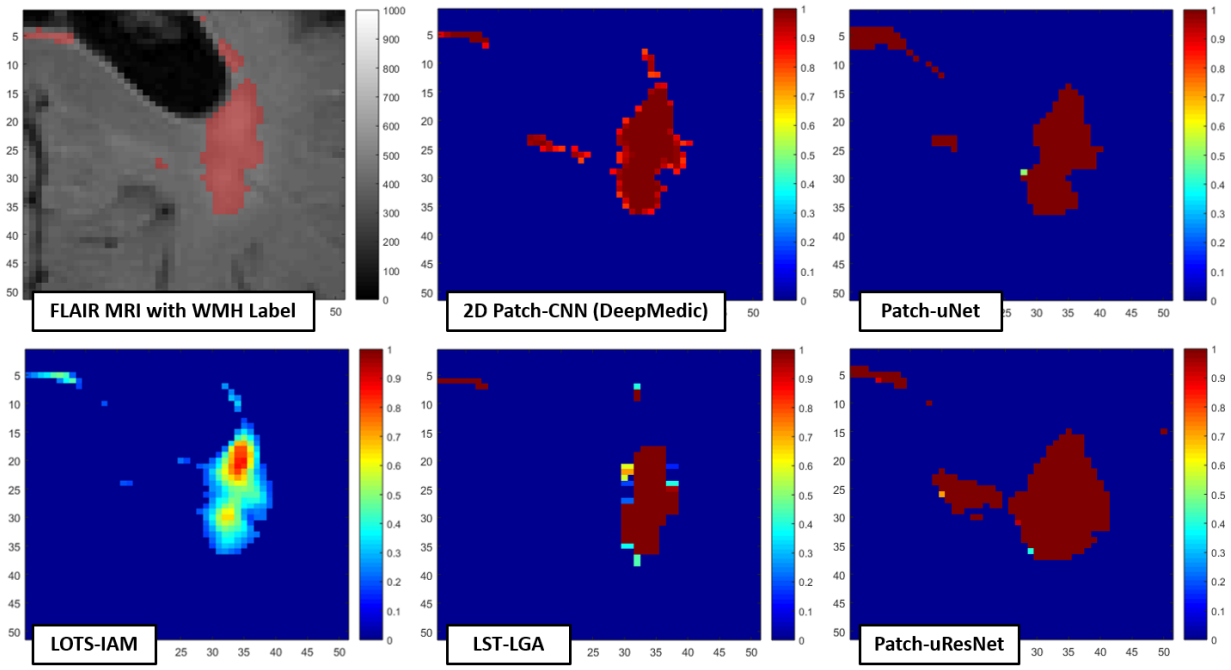

Figure 5: Visualisation of **thresholded** probabilistic values of WMH produced by LOTS-IAM compare to other methods, which are the 2D Patch-CNN (2D-DeepMedic), Patch-uNet (uNet based deep neural networks), LST-LGA, and Patch-uResNet. The probabilistic values are cut off by their respective optimum threshold values.

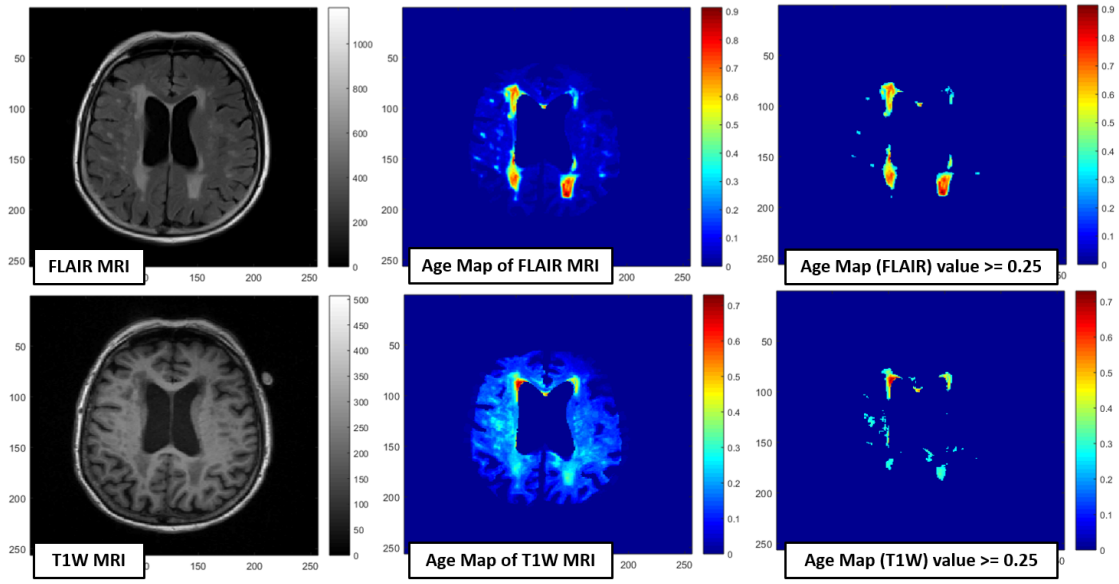

Figure 6: Example of how LOTS-IAM could be performed on both FLAIR and T1-weighted (T1W). Please note that the current version of LOTS-IAM is highly optimised for FLAIR modality of MRI, not T1W.

one of the age maps. The effect of different sets of blending weights is illustrated in Figure 7.

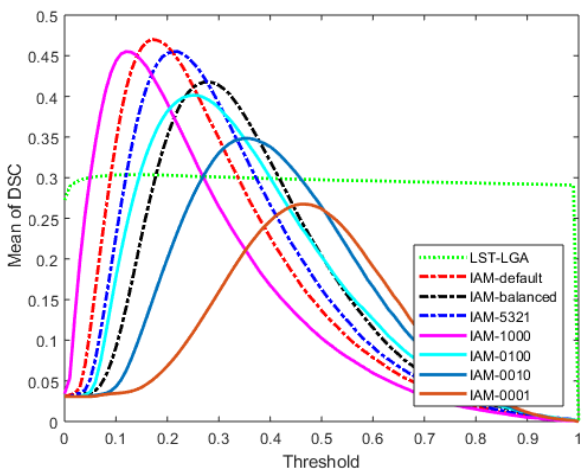

Figure 7: Curves of mean dice similarity coefficient (DSC) produced by using different settings of weights for blending different age maps. LOTS-IAM used in this experiment is LOTS-IAM-GPU-512s64m. All settings of blending Weights are listed in Table 2.

From Figure 7, we can see that blending all four age maps improves IAM's performance and works better than using only one of the four available age maps. Based on the results listed in Table 2, blending weights of 0.65, 0.20, 0.10 and 0.05 for age maps produced from 1×1, 2×2, 4×4 and 8×8 source/target patches respectively gives the best DSC score of 0.4434. As this combination produced the best DSC score in this evaluation, we made this set the default set for IAM computation. Coincidentally, this set of blending weights has been used from the start of IAM's development and also used in the first paper of IAM (Rachmadi et al., 2017b).

The fact that combining the age maps of different patch sizes performs the best proves that it is needed to consider not only the colour of the individual pixels but also the local distribution of the pixel colours to correctly label WMH lesions. Still, it appears that the individual pixel colour is a strong feature for classification as the best performing blending weights are 0.65, 0.2, 0.1 and 0.05.

5.5. Analysis of IAM's Distance Function's α-Parameter

IAM's distance function (Equation 1) has the α-parameter which controls blending of maximum difference and mean difference between source and target patches. Mean difference is used where $\alpha = 0$ whereas maximum difference is used where $\alpha = 1$. The reason behind the use of α-parameter is that we might want to calculate distance value based on mean difference only, maximum difference only or combination of mean and maximum differences.

Table 2: Mean and standard deviation of DSC produced by using different settings of weights for blending different age maps. Plots corresponding to settings listed in this table can be seen in Figure 7. LOTS-IAM tested in this experiment is LOTS-IAM-GPU-512s64m.

Name	Blending Weights				TRSH	DSC	
	1x1	2x2	4x4	8x8		mean	std
LST-LGA	-	-	-	-	0.134	0.2936	0.1658
IAM-default	0.65	0.20	0.10	0.05	0.179	0.4434	0.1499
IAM-balanced	0.25	0.25	0.25	0.25	0.287	0.3957	0.1362
IAM-5321	0.50	0.30	0.20	0.10	0.228	0.4308	0.1469
IAM-1000	1	0	0	0	0.139	0.4253	0.1417
IAM-0100	0	1	0	0	0.267	0.3597	0.1262
IAM-0010	0	0	1	0	0.376	0.3085	0.1085
IAM-0001	0	0	0	1	0.475	0.2476	0.0839

Based on the results depicted in Figure 8 and Table 3, we can see that combining/averaging mean and maximum differences using $\alpha = 0.5$ produced the best mean DSC though the performance differences are minimum. Coincidentally, $\alpha = 0.5$ has been used from the start of IAM's development and also used in the first paper of IAM (Rachmadi et al., 2017b).

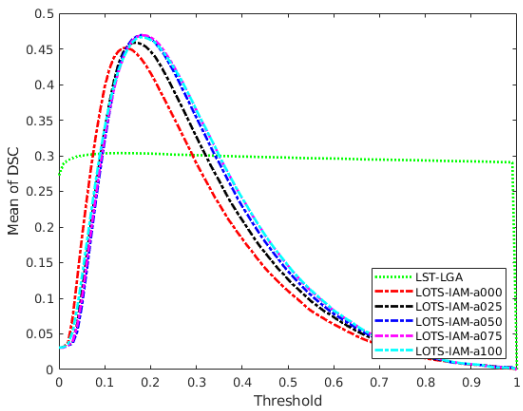

Figure 8: Curves of mean dice similarity coefficient (DSC) produced by using different values of α-parameter (see Table 3). LOTS-IAM used in this experiment is LOTS-IAM-GPU-512s64m.

5.6. Analysis of IAM's Random Sampling Scheme

To automatically detect FLAIR's WMH without any expert supervision, IAM works on the assumption that normal brain tissue is predominant compared with the extent of abnormalities. Due to this assumption, random sampling is used in the computation of IAM to choose the target patches. However, it raises an important question on the stability of IAM's performance to produce the same level of

Table 3: Mean and standard deviation values of DSC produced by using different values of α-parameter. Corresponding plots are depicted in Figure 8. LOTS-IAM used in this experiment is LOTS-IAM-GPU-512s64m.

No	Name	α	DSC	
			mean	std
1	LOTS-IAM-a000	0.00	0.4513	0.1498
2	LOTS-IAM-a025	0.25	0.4588	0.1533
3	LOTS-IAM-a050	0.50	0.4694	0.1587
4	LOTS-IAM-a075	0.75	0.4690	0.1600
5	LOTS-IAM-a100	1.00	0.4665	0.1597

Table 4: Mean and standard deviation values of DSC distributions for each IAM's setting depicted in Figure 9. Each of IAM's setting is tested on a random MRI data 10 times.

No	Method	TRSH	DSC	
			mean	std
1	LOTS-IAM-GPU-2048s128m	0.15	0.5681	0.0041
2	LOTS-IAM-GPU-1024s128m	0.20	0.5901	0.0018
3	LOTS-IAM-GPU-512s64m	0.20	0.5922	0.0033
4	LOTS-IAM-GPU-256s32m	0.20	0.5925	0.0075
5	LOTS-IAM-GPU-128s32m	0.25	0.5848	0.0092
6	LOTS-IAM-GPU-64s16m	0.25	0.5852	0.0141

results for one exact MRI data, especially using different number of target patches.

In this experiment, we randomly choose one MRI data out of the 60 MRI data that we have, and ran LOTS-IAM-GPU multiple times (*i.e.*, 10 times in this study) using different number of target patch samples. Each result was then compared to the ground truth, grouped together and plotted as boxplots (Figure 9) and listed in Table 4.

Figure 9 and Table 4 show that the deviation of IAM's computation for one MRI data is small in all different settings of LOTS-IAM-GPU. However, it is true that by adding number of target patches in IAM's patch generation alienates this deviation as shown in Figure 9, where the boxplots produced by LOTS-IAM-GPU-2048s128m are smaller than the ones produced by LOTS-IAM-GPU-64s16m (see Table 4 for quantification).

5.7. WMH Burden Scalability Test

In this experiment, all methods were tested and evaluated to see their performances on doing WMH segmentation in different volumes of WMH (*i.e.*, WMH burden). DSC metric is still used in this experiment, but the dataset is grouped into three different groups based on the WMH burden of each patient. The groups are listed in Table 5 while the results can be seen in Figure 10 and Table 6. Please note that IAM is represented by LOTS-IAM-GPU-

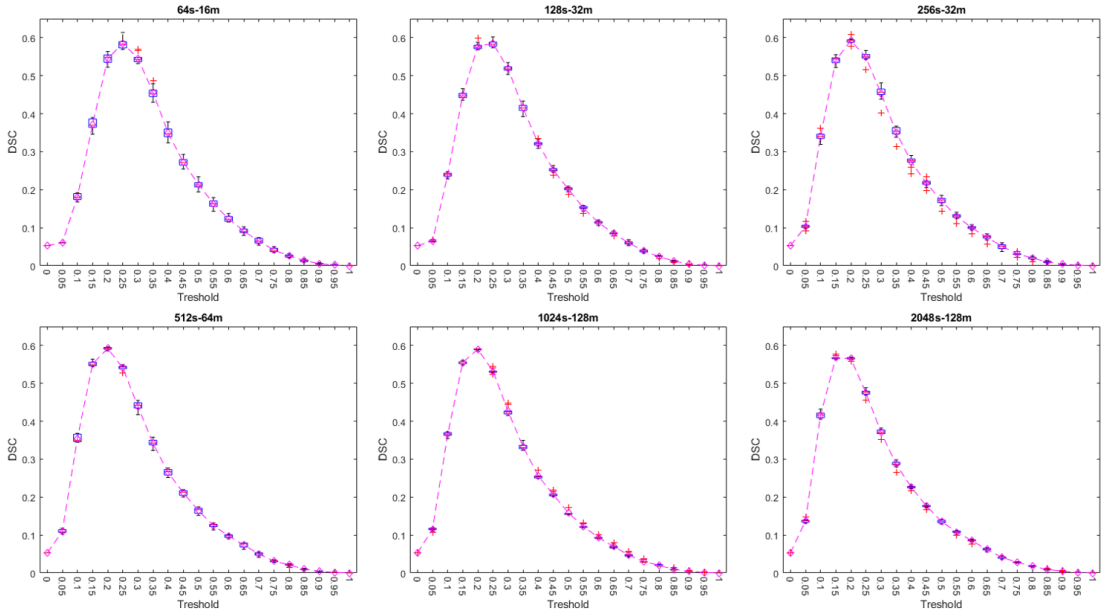

Figure 9: DSC distributions of 1 random MRI scan tested 10 times by using LOTS-IAM-GPU with different numbers of target patch samples. **From upper-left to upper-right:** LOTS-IAM-GPU using 64, 128 and 256 target patches respectively. **From lower-left to lower-right:** LOTS-IAM-GPU using 512, 1024 and 2048 target patches respectively. Each box represents distribution of DSC produced on respective threshold value while ashed magenta line represents the mean of DSC on respective threshold value.

512s64m as it is the best performer amongst the IAM methods (see Table 1).

Table 5: Three groups of MRI data based on WMH volume.

No.	Group	WMH Vol. (mm^3)	# MRI Data
1	Small	WMH $\leq$ 4500	27
2	Medium	4500 $<$ WMH $\leq$ 13000	25
3	Large	WMH $>$ 13000	8

From Figure 10, it can be appreciated that LOTS-IAM-GPU-512s64m performed better than LST-LGA in this experiment outperforming LST-LGA's performances distribution in all groups. LOTS-IAM-GPU-512s64m also performed better than the conventional supervised machine learning algorithms (i.e. SVM and RF) in Small and Medium groups. Whereas, LOTS-IAM-GPU-512s64m's performance was at the level, if not better, than the supervised deep neural networks algorithms DBM and CEN. However, LOTS-IAM-GPU-512s64m still could not beat the *state-of-the-art* supervised deep neural networks 2D patch-CNN in any group.

To make this observation clearer, Table 6 lists the mean and standard deviation values that correspond to the box-plot shown in Figure 10. From both Figure 10 and Table 6 it can be observed that the deviation of IAM's performances in Small WMH burden is still high compared to the other methods evaluated. However, IAM's performance is more stable in Medium and Large WMH burdens.

Table 6: Mean and standard deviation values of dice similarity coefficient (DSC) score's distribution for all methods tested in this study in respect to WMH burden of each patient (see Table 5). Note that LOTS-IAM-GPU-512s64m is listed as LIG-512s64m in this table.

Method	**TRSH**	**DSC - Small**		**DSC - Medium**		**DSC - Large**	
		Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**
LST-LGA	0.138	0.2335	0.1785	0.3524	0.1208	0.4645	0.1399
SVM	0.925	0.1792	0.0958	0.3360	0.1284	0.4966	0.0377
RF	0.995	0.2512	0.1298	0.4150	0.1662	0.6055	0.0559
DBM	0.687	0.3127	0.1432	0.3442	0.1350	0.4014	0.1474
CEN	0.284	0.4359	0.1802	0.4474	0.1485	0.4896	0.1122
Patch-uResNet	0.200	0.5007	0.2064	0.5403	0.1432	0.6064	0.0579
Patch-uNet	0.200	0.4872	0.1596	0.5079	0.1697	0.5447	0.0574
2D Patch-CNN	0.801	0.5230	0.1722	0.5118	0.1340	0.6053	0.0341
LIG-512s64m	0.179	0.4682	0.2278	0.4660	0.1331	0.4940	0.0932

5.8. Longitudinal Test

In this experiment, we evaluate spatial agreement between the produced results in three consecutive years. For each subject, we aligned Year-2 (Y2) and Year-3 (Y3) MRI and derived data to

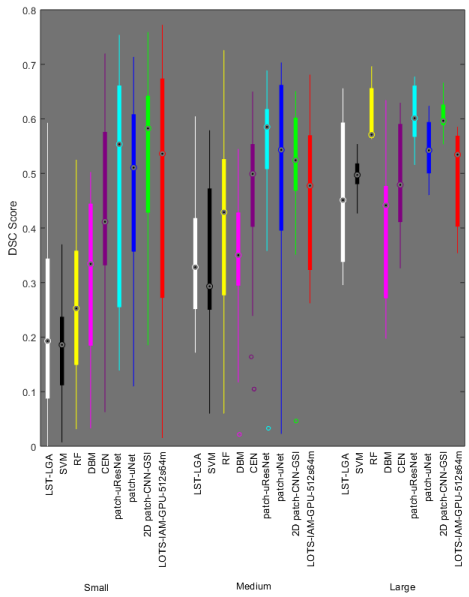

Figure 10: Distributions of dice similarity coefficient (DSC) score for all methods tested in this study in respect to WMH burden of each patient (see Table 5).

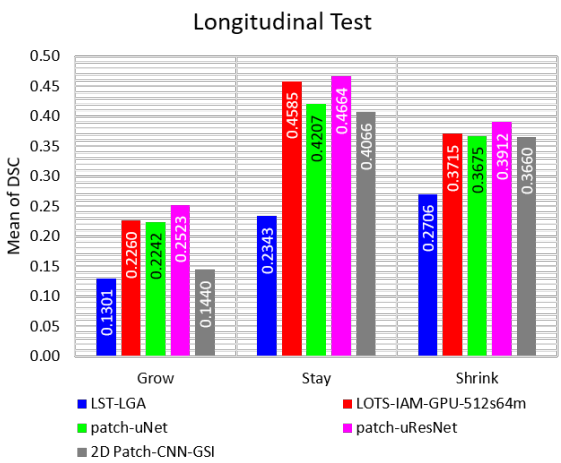

Figure 11: Quality of spatial agreement (Mean of DSC) of the produced results in longitudinal test. Longitudinal test is done to see the performance of tested methods in longitudinal dataset of MRI (see Table 7 for full report).

Table 7: Mean and standard deviation values produced in longitudinal tes (see Table 5). LOTS-IAM-GPU-512s64m is listed as LIG-512s64m in this table.

Method	Dice Similarity Coefficient (DSC)					
	Grow		Stay		Shrink	
	Mean	Std	Mean	Std	Mean	Std
LST-LGA	0.1301	0.0350	0.2343	0.0199	0.2706	0.0058
LIG-512s64m	0.2260	0.0084	0.4585	0.0104	0.3715	0.0018
Patch-uNet	0.2242	0.0125	0.4207	0.0125	0.3675	0.0242
Patch-uResNet	**0.2523**	0.0199	**0.4664**	0.0211	**0.3912**	0.0044
2D Patch-CNN	0.1440	0.0228	0.4066	0.0298	0.3660	0.0129

the Year-1 (Y1), subtracted the aligned WMH labels of the baseline/previous year from the follow-up year(s)(*i.e.*, Y2-Y1, Y3-Y2, and Y3-Y1), and then labelled each voxel as *'grow'* if it has value above zero after subtraction, with *'shrink'* if it has value below zero after subtraction, and with *'stay'* if it has value of zero after subtraction and one before subtraction. This way, we can see whether the method captures the progression of WMH across time (*i.e.*, longitudinally). An example of the output from this experiment is shown in Figure 12 where sections of the original FLAIR MRI, LOTS-IAM, LST-LGA and uNet across three respective years (Y1, Y2 and Y3) for a subject are depicted.

Figure 11 summarises the results listed in Table 7 for all methods (*i.e.*, LST-LGA, LOTS-IAM-GPU-512s64m, Patch-uNet, Patch-uResNet and 2D Patch-CNN-GSI). We can see that LOTS-IAM-GPU-512s64m outperforms LST-LGA and competes with deep neural networks methods of Patch-uNet, Patch-uResNet and 2D Patch-CNN-GSI, where LOTS-IAM-GPU-512s64m is the second best performer after Patch-uResNet in this longitudinal evaluation. This, again, confirms that the LOTS-IAM shows comparable performance with the *state-of-the-art* deep learning convolutional neural network methods.

5.9. Correlation with Visual Rating

In this experiment, we want to see how close IAM's results correlate with visual ratings of WMH, specifically Fazekas's visual ratings (Fazekas et al., 1987) and Longstreth's visual ratings (Longstreth et al., 1996).

The correlation was calculated by using Spearman's correlation. The correlation coefficients calculated were: 1) between the total Fazekas's rating (*i.e.*, the sum of periventricular white matter hyperintensities (PVWMH) and deep white matter hyperintensities (DWMH)) and manual/automatic WMH volumes and 2) between Longstreth's rating and manual/automatic WMH volumes. The results are listed in Table 8.

Table 8 shows that, although not much better, all LOTS-IAM-GPU methods highly correlate with visual rating clinical scores. Despite LST-LGA's output having lower value of mean DSC (see Table 1) compared to other methods, it still highly correlates with visual ratings. LOTS-IAM-GPU implementations have high values of both mean DSC and correlation with visual ratings.

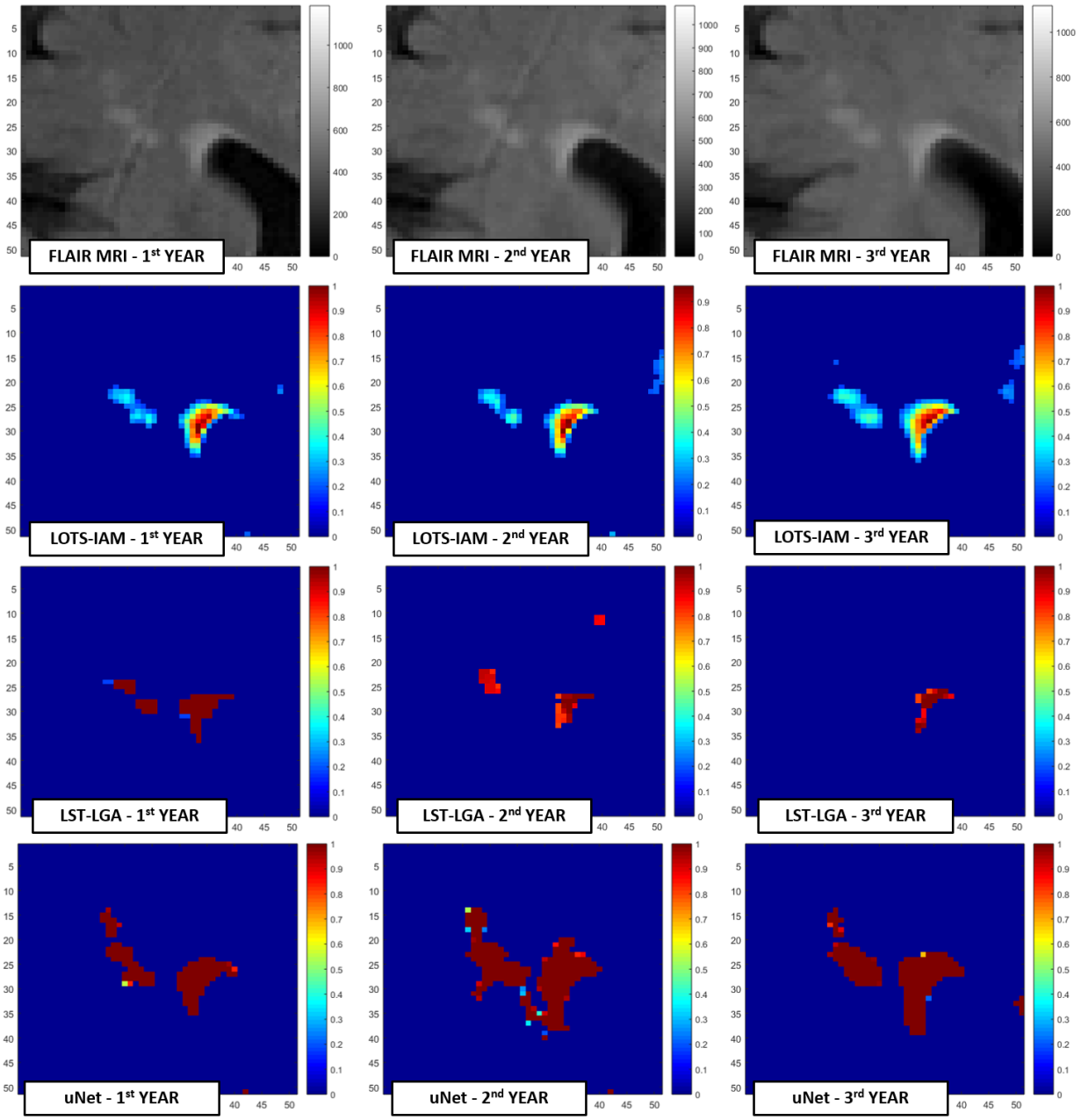

Figure 12: White matter hyperintensities (WMH) information presented by FLAIR MRI, LOTS-IAM, LST-LGA and uNet in 3 (three) consecutive years. The $2^{nd}/3^{rd}$ year of FLAIR MR images are co-registered first to the 1st year of FLAIR MRI and then processed by LOTS-IAM, LST-LGA and uNet. LOTS-IAM produces richer information of the WMH progression than LST-LGA/uNet as age maps of LOTS-IAM preserve the underlying variance of WMH's intensity, giving a good perspective on how WMH grow over time.

Table 8: Non-parametric correlation using Spearman's correlation coefficient between WMH volume and Fazekas and Longstreth visual ratings.

	Visual Rating	**Fazekas (Total)**		**Longstreth**	
	Method	**Spearman's Corr. Val.**		**Spearman's Corr. Val.**	
		rho	**p**	**rho**	**p**
1	Manual label	0.7562	1.04×1^{-12}	0.7752	1.45×10^{-12}
1	LST-LGA	0.5718	3.38×6^{-12}	0.4813	1.50×10^{-4}
2	SVM	0.4062	0.0017	0.3602	0.0059
3	RF	0.2447	0.0666	0.2128	0.1119
4	DBM	0.2436	0.0679	0.1659	0.2174
5	CEN	0.2359	7.74×10^{-2}	0.3618	0.0057
1	Patch-uResNet	0.3602	5.90×10^{-3}	0.5171	3.80×10^{-5}
1	Patch-uNet	0.4618	2.99×10^{-4}	0.5140	4.33×10^{-5}
6	2D Patch-CNN-GSI	0.7054	9.01×10^{-10}	0.8664	3.19×10^{-18}
7	LIG-2048s128m	0.4727	2.05×10^{-4}	0.4579	3.42×10^{-4}
8	LIG-1024s128m	0.4892	1.13×10^{-4}	0.4849	1.32×10^{-4}
9	LIG-512s64m	0.5010	7.19×10^{-5}	0.5065	5.82×10^{-4}
10	LIG-256s64m	0.5009	7.22×10^{-5}	0.5085	5.37×10^{-4}
11	LIG-128s32m	0.4505	4.38×10^{-4}	0.4946	9.22×10^{-4}
12	LIG-64s16m	0.4393	6.30×10^{-4}	0.4858	1.28×10^{-4}

6. Conclusion and Future Work

The optimisation of IAM presented (LOTS-IAM-GPU) improves both performance and processing time with respect to previous versions of IAM. Despite not being a WMH segmentation method *per se*, it can be successfully applied for this purpose. Being unsupervised confers an additional value to this fully automatic method as it does not depend on expert-labelled data, and therefore is independent from any subjectivity and inconsistency from human experts, which usually influence supervised machine learning algorithms. Furthermore, our results show that LOTS-IAM also successfully outperformed some supervised deep neural networks algorithms which are DBM and CEN. Some improvements still could be done by adding or using different sets of brain tissues masks other than CSF and NAWM, for example cortical mask and cerebrum brain mask.

One major drawback of the original IAM is the long computation time that takes to process a single MRI data. LOTS-IAM-GPU successfully speeds up IAM's computation time by 17 to 435 times, not only owed to its implementation in GPU, but also to the use of a limited number of target patch samples. LOTS-IAM-GPU also outperforms LST-LGA, the current *state-of-the-art* method for unsupervised WMH segmentation, in both DSC metric and processing speed.

IAM could provide unsupervised labels for pre-training supervised deep neural networks algorithms and transfer learning, but this has not been explored too much. In our preliminary research, IAM is used in task adaptation transfer learning and predicting brain abnormalities progression/regression (Rachmadi et al., 2018a). Due to its nature, it can be hypothesised its applicability to segment brain lesions in CT scans or different brain pathologies. Further works should also explore its implementation on a multispectral approach that combines different MRI sequences. The implementation of LOTS-IAM-GPU is publicly available at https://github.com/febrianrachmadi/lots-iam-gpu.

Acknowledgements

The first author would like to thank Indonesia Endowment Fund for Education (LPDP) of Ministry of Finance, Republic of Indonesia, for funding his study at School of Informatics, the University of Edinburgh. Funds from Row Fogo Charitable Trust (Grant No. BRO-D.FID3668413)(MCVH) are also gratefully acknowledged. Data collection and sharing for this project was funded by 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, Alzheimers Association; Alzheimers 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 Alzheimers Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Labora-

tory for Neuro Imaging at the University of Southern California.

References

Bellini, R., Kleiman, Y., Cohen-Or, D., 2016. Time-varying weathering in texture space. ACM Transactions on Graphics (TOG) 35 (4), 141.

Dice, L. R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.

Fazekas, F., Chawluk, J. B., Alavi, A., Hurtig, H. I., Zimmerman, R. A., 1987. Mr signal abnormalities at 1.5 t in alzheimer's dementia and normal aging. American journal of roentgenology 149 (2), 351–356.

Guerrero, R., Qin, C., Oktay, O., Bowles, C., Chen, L., Joules, R., Wolz, R., Valdés-Hernández, M., Dickie, D., Wardlaw, J., et al., 2018. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. NeuroImage: Clinical 17, 918–934.

Hernández, M. d. C. V., Morris, Z., Dickie, D. A., Royle, N. A., Maniega, S. M., Aribisala, B. S., Bastin, M. E., Deary, I. J., Wardlaw, J. M., 2013. Close correlation between quantitative and qualitative assessments of white matter lesions. Neuroepidemiology 40 (1), 13–22.

Ithapu, V., Singh, V., Lindner, C., Austin, B. P., Hinrichs, C., Carlsson, C. M., Bendlin, B. B., Johnson, S. C., 2014. Extracting and summarizing white matter hyperintensities using supervised segmentation methods in alzheimer's disease risk and aging studies. Human brain mapping 35 (8), 4219–4235.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3d {CNN} with fully connected {CRF} for accurate brain lesion segmentation. Medical Image Analysis 36, 61 – 78.

Li, H., Jiang, G., Wang, R., Zhang, J., Wang, Z., Zheng, W.-S., Menze, B., 2018. Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images. arXiv preprint arXiv:1802.05203.

Longstreth, W., Manolio, T. A., Arnold, A., Burke, G. L., Bryan, N., Jungreis, C. A., Enright, P. L., O'Leary, D., Fried, L., Group, C. H. S. C. R., et al., 1996. Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people the cardiovascular health study. Stroke 27 (8), 1274–1282.

Lutkenhoff, E. S., Rosenberg, M., Chiang, J., Zhang, K., Pickard, J. D., Owen, A. M., Monti, M. M., 2014. Optimized brain extraction for pathological brains (optibet). PloS one 9 (12), e115551.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W., Beckett, L., 2005. The alzheimer's disease neuroimaging initiative. Neuroimaging Clinics of North America 15 (4), 869–877.

Myers, J. L., Well, A., Lorch, R. F., 2010. Research design and statistical analysis. Routledge.

Rachmadi, M. F., Hernández, M. V., 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. bioRxiv, 345033.

Rachmadi, M. F., Valdés-Hernández, M. d. C., Agan, M. L. F., Di Perri, C., Komura, T., Initiative, A. D. N., et al., 2018b. Segmentation of white matter hyperintensities using convolutional neural networks with global spatial information in routine clinical brain mri with none or mild vascular pathology. Computerized Medical Imaging and Graphics 66, 28–43.

Rachmadi, M. F., Valdés-Hernández, M. d. C., Agan, M. L. F., Komura, T., 2017a. Deep learning vs. conventional machine learning: Pilot study of wmh segmentation in brain mri with absence or mild vascular pathology. Journal of Imaging 3 (4), 66.

Rachmadi, M. F., Valdés-Hernández, M. d. C., Komura, T., 2017b. 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, pp. 321–326.

Rachmadi, M. F., Valdés-Hernández, M. d. C., Komura, T., 2018c. Automatic irregular texture detection in brain mri without human supervision. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 000–000.

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

Scheltens, P., Barkhof, F., Leys, D., Pruvo, J., Nauta, J., Vermersch, P., Steinling, M., Valk, J., 1993. A semiquantative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. Journal of the neurological sciences 114 (1), 7–12.

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., et al., 2012. An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. Neuroimage 59 (4), 3774–3783.

Shiee, N., Bazin, P.-L., Ozturk, A., Reich, D. S., Calabresi, P. A., Pham, D. L., 2010. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. NeuroImage 49 (2), 1524–1535.

Valdés Hernández, M. d. C., Armitage, P. A., Thrippleton, M. J., Chappell, F., Sandeman, E., Muñoz Maniega, S., Shuler, K., Wardlaw, J. M., 2015. Rationale, design and methodology of the image analysis protocol for studies of patients with cerebral small vessel disease and mild stroke. Brain and behavior 5 (12).

Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R. I., T O'Brien, J., Barkhof, F., Benavente, O. R., et al., 2013. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. The Lancet Neurology 12 (8), 822–838.