Deeply supervised network for white matter hyperintensities segmentation with transfer learning

Yilei Wu[1] YILEI.WU@U.NUS.EDU
Fang Ji[1] JI.FANG@NUS.EDU.SG
Yao Feng Chong[1] CHONG.YAOFENG@U.NUS.EDU
Christopher Li-Hsian Chen[1] PHCCCLH@NUS.EDU.SG
Juan Helen Zhou[1] HELEN.ZHOU@NUS.EDU.SG
[1] *Yong Loo Lin School of Medicine, National University of Singapore, Singapore*

Editors: Under Review for MIDL 2022

Abstract

White matter hyperintensities (WMH) are brain white matter lesions commonly found in the elderly. Due to its association with cerebrovascular and neurodegenerative diseases, quantifying WMH volume is critical for many neurological applications. Previous segmentation approaches using 2D U-Net potentially omit the learning of 3D spatial contextual information. This paper proposes a deeply supervised 3D U-Net-like network with transfer learning to perform WMH segmentation in fluid attenuation inversion recovery (FLAIR) magnetic resonance images (MRI). We leveraged a pretrained network constructed by predicting brain age from structural MRIs. The proposed method achieved a Dice score of 82.3 on the MICCAI WMH Challenge training dataset and 75.3 on another independent testing dataset, outperforming other state-of-the-art methods.

Keywords: white matter hyperintensities, transfer learning, deep supervision

1. Introduction

The prevalence of white matter hyperintensities (WMH) increases with age, which indicates a higher risk of stroke and dementia. WMH segmentation remains challenging due to high inter-subject variability in anatomy, contrast, and spatial properties as well as differences in MR acquisition. Here, we proposed a 3D transfer learning based WMH segmentation approach by using a 3D network pretrained to predict the brain age of 14,503 brain MRIs from the UK Biobank database (Peng et al., 2021). To encourage the model to learn discriminative features across different levels, we employed deep supervision at multiple resolutions (Dou et al., 2016).

2. Methods

The proposed network (Figure 1) is a modified version of 3D U-Net (Çiçek et al., 2016), which consists of an encoder-decoder structure with skip-connection. The first five blocks from the pretrained network (Peng et al., 2021) were used as the encoder. Each block consists of a 3D conv layer, a Batch Normalization (BN) layer, and a ReLU layer followed by a max-pooling layer. Correspondingly, the decoder has five upsampling blocks. Each block consists of a transposed 3D conv layer followed by a squeeze-and-excitation (SE) block

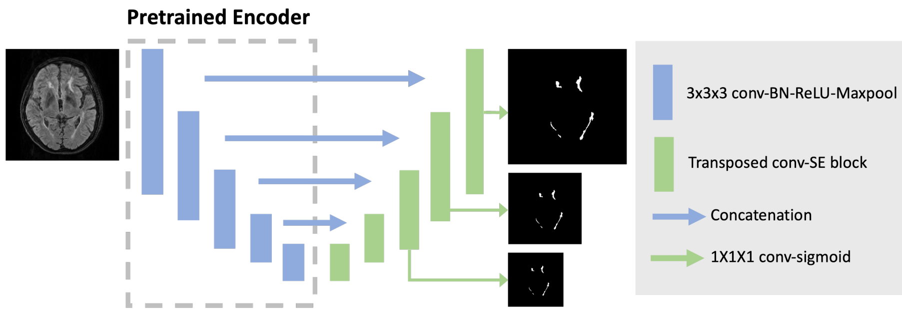

Figure 1: Illustration of the proposed network architecture.

(Hu et al., 2018). Compared with vanilla ResNet, SE blocks facilitate feature recalibration by reweighting the feature maps channel-wisely through its learnable channel-wise attention.

We implemented deep supervision by connecting output layers to decoders at multiple stages with different resolutions. Specifically, our model added output layers at the third and fourth upsampling blocks, thus generating outputs with $\frac{1}{4}$ and $\frac{1}{2}$ of the spatial resolution of the final output respectively. We then downsampled the corresponding labels to calculate the auxiliary loss at each resolution.

For data augmentation, we applied on-the-fly data augmentation with brightness change (0.2), contrast change (0.25), mirroring (0.5), resampling (0.25), and rotation (0.5); numbers in parentheses refer to the probability per sample of being transformed during the training.

We applied z-score normalization to all images and padded them to 160*192*160 to accommodate the pretrained encoder. The compound loss of dice loss and binary cross-entropy loss was used as the loss function. During training, we used a stochastic gradient descent optimizer with Nesterov momentum and a learning rate of 0.01. The implementation is publicly available at `https://github.com/yilei-wu/wmh_seg_public`.

The proposed method was validated on the MICCAI WMH Challenge (MWC) (Kuijf et al., 2019) training dataset (N = 60 from three sites: Utrecht, Amsterdam, and Singapore) and another independent dataset from Memory Aging and Cognition Centre (MACC), National University of Singapore (N = 58). For the MACC dataset, our clinical experts performed manual delineation of the WMH.

	MICCAI Challenge					MACC				
Methods	Dice	HD95	AVD	Recall	Precision	Dice	HD95	AVD	Recall	Precision
Li et. al.	0.806	5.524	20.401	0.846	0.788	0.605	14.445	36.950	0.549	0.736
nnUNet2D	0.797	5.357	23.571	**0.856**	0.764	0.730	9.254	32.197	0.718	**0.795**
nnUNet3D	0.796	5.409	21.258	0.844	0.772	0.645	12.205	36.030	0.639	0.723
Proposed	**0.823**	**4.866**	**12.619**	0.835	**0.820**	**0.753**	**8.196**	**26.772**	**0.788**	0.752

Table 1: Results on MWC and MACC

3. Results

We compared the performance of our approach with the approach by Li and his colleagues (Li et al., 2018) and nnUNet for both 2D and 3D configurations (Isensee et al., 2021). Li et al. is the winning solution of the MWC based on an ensembled 2D U-Net. nnUNet is a self-adaptive medical image segmentation framework that demonstrates strong performance on many medical image segmentation tasks. We applied standard voxel-wise metrics (e.g. Dice, Hausdorff distance (HD), Average Volume Distance(AVD), Recall, Precision) for the performance evaluation.

Five-fold cross-validation was performed within each of the two datasets (Table 1). The proposed method achieved a competitive performance on both datasets. Importantly, to evaluate the generalizability to unseen data, we applied the models trained with the MWC dataset on the independent testing MACC dataset (Table 2). Our proposed approach achieved the best performance on most metrics.

To further investigate the effectiveness of transfer learning (TL), deep supervision (DS), and data augmentation (DA), an ablation study was carried out on the MWC dataset (Table 3). All three strategies enhanced the segmentation performance.

	Dice	HD95	AVD	Recall	Precision
Li et. al.	0.636	17.391	43.339	**0.678**	0.650
nnUNet2D	0.617	15.204	46.080	0.534	**0.893**
nnUNet3D	0.466	25.732	58.425	0.414	0.843
Proposed	**0.682**	**13.207**	**43.251**	0.659	0.791

Table 2: Cross-dataset prediction on MACC

TL	DS	DA	Dice	HD95	AVD	Recall	Precision
			0.776	9.633	21.461	0.810	0.764
✓			0.784	7.657	20.501	0.807	0.784
✓	✓		0.789	7.258	18.128	0.808	0.790
✓		✓	0.814	5.005	14.744	0.832	0.817
✓	✓	✓	0.823	4.866	12.619	0.835	0.820

Table 3: Ablation study on MWC

4. Conclusion

The proposed 3D deeply supervised network achieved competitive performance for WMH segmentation. We showed that transfer learning from UKBiobank, deep supervision, and highly optimized data augmentation significantly improved the segmentation performance.

References

Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *MICCAI*, pages 424–432. Springer, 2016.

Qi Dou, Hao Chen, Yueming Jin, Lequan Yu, Jing Qin, and Pheng-Ann Heng. 3d deeply supervised network for automatic liver segmentation from ct volumes. In *MICCAI*, pages 149–157. Springer, 2016.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

Hugo J Kuijf, J Matthijs Biesbroek, Jeroen De Bresser, Rutger Heinen, Simon Andermatt, Mariana Bento, Matt Berseth, Mikhail Belyaev, M Jorge Cardoso, Adria Casamitjana, et al. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. *IEEE transactions on medical imaging*, 38(11):2556–2568, 2019.

Hongwei Li, Gongfa Jiang, Jianguo Zhang, Ruixuan Wang, Zhaolei Wang, Wei-Shi Zheng, and Bjoern Menze. Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images. *NeuroImage*, 183:650–665, 2018.

Han Peng, Weikang Gong, Christian F Beckmann, Andrea Vedaldi, and Stephen M Smith. Accurate brain age prediction with lightweight deep neural networks. *Medical image analysis*, 68:101871, 2021.