2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

DeepQSMSeg: A Deep Learning-based Sub-cortical Nucleus Segmentation Tool for Quantitative Susceptibility Mapping*

Yonghang Guan[1], Xiaojun Guan[2], Jingjing Xu[2], Hongjiang Wei[3], Xiaojun Xu[2※], Yuyao Zhang[1,4※]

Abstract— **Deep brain nuclei are closely related to the pathogenesis of neurodegenerative diseases. Automatic segmentation for brain nuclei plays a significant role in aging and disease-related assessment. Quantitative susceptibility mapping (QSM), as a novel MRI imaging technique, attracts increasing attention in deep gray matter (DGM) nuclei-related research and diagnosis. This paper proposes DeepQSMSeg, a deep learning-based end-to-end tool, to segment five pairs of DGM structures from QSM images. The proposed model is based on a 3D encoder-decoder fully convolutional neural network. For concentrating network on the target regions, spatial and channel attention modules are adopted in both encoder and decoder stages. Dice loss is combined with focal loss to alleviate the imbalance of ROI classes. The result shows that our method can segment DGM structures from QSM images precisely, rapidly and reliably. Comparing with ground truth, the average Dice coefficient for all ROIs in the test dataset achieved 0.872±0.053, and Hausdorff distance was 2.644±2.917 mm. Finally, an age-related susceptibility development model was used to confirm the reliability of DeepQSMSeg in aging and disease-related studies.**

Clinical Relevance—**Accurate and automatic segmentation tool for sub-cortical regions in QSM can significantly alleviate the pressure of radiologists. It can also accelerate the progress of related research and clinical translation.**

I. Introduction

Quantifying tissue magnetic susceptibility using MRI provides a noninvasive method to measure brain tissue components related to iron. This is attributed to the Quantitative Susceptibility Mapping (QSM) imaging technique which is sensitive to the spatial variations of molecular and cellular components that exhibit different magnetic susceptibility characteristics. In the human brain, iron is primarily located in the deep gray matter (DGM) nuclei [1]. DGMs are crucially involved in brain learning, memory and cognition functions [2]. Excessive iron overload in DGM nuclei has been reported in numerous neurodegenerative diseases such as Parkinson's disease (PD), Alzheimer's disease (AD) and multiple sclerosis (MS) [3], [4]. Thus, QSM increasingly attracts attention in the researches that are focusing on iron distributions in DGMs. Segmentation for DGM nuclei in QSM is crucial for brain iron-related studies. However, existing DGM nucleus segmentation based on QSM is rare. Most reported works are based on manual or atlas-based segmentation.

Very recently, a few segmentation methods for QSM in DGM nuclei are reported, which can be summarized into two categories: atlas-based methods and learning-based methods. In documentation [5], atlas-based segmentation pipelines were used to segment brain DGM nuclei. These methods mainly involved template generation, subject registration and label propagation. This kind of method requires a series of nontrivial and computationally expensive operations. The performance is highly sensitive with data variation (e.g., contrast difference and scanner parameter settings) between atlas and subjects. The authors in the study [6] proposed a 2-D fully convolutional network to segment DGM structures from QSM. The 2-D network cannot capture the inter-slices spatial correlation, which is crucial in pixel-level volumetric image segmentation. Besides, it only involved manually selected slices containing DGM structures for training and testing. This segmentation is thus not fully automatic. Chai et al. proposed a volumetric double branch encoder-decoder model that used paired T1 and QSM data as input [7]. However, it still cannot handle the foreground imbalance problem, limiting the increase in segmentation accuracy. Therefore, an automatic, accurate, robust and efficient segmentation method for QSM is in urgent need to depict DGM structures.

In the proposed work, we devised a deep learning-based sub-cortical nucleus segmentation tool for QSM, named DeepQSMSeg. Our segmentation backbone is a one-stage 3D encoder-decoder fully convolutional network (FCN). Most of the target structures are of small size, which makes the foreground and background voxels imbalanced. We thus adopted attention modules into the network and supervised the training process using both dice loss and focal loss [8]. Inspired by [9], we combine voxel attention and channel attention into both encoder and decoder stages in our network to focus on the target structures. Due to the anisotropic nature of our dataset, anisotropic convolution is adopted to reduce the model parameters and accelerate the model inference. For testing, one QSM volume was segmented in 2.600±0.018s. 5-fold cross validation was performed to evaluate the method's performance. Average Dice score coefficient (DSC) for DGM structures achieved 0.872±0.053 and Hausdorff distance was 2.644±2.917 mm. We then computed average susceptibility values in each DGM ROI using segmentation masks and fitted

*Researuch funded by the National Nature Science Foundations of China (Grant No. 62071299 and Grant No. 82001767).

[1] Yonghang Guan and Yuyao Zhang are with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China.

[2] Jingjing Xu, Xiaojun Guan and Xiaojun Xu are with the Department of Radiology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China.

[3] Hongjiang Wei is with the Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China.

[4]Yuyao Zhang is also with Shanghai Engineering Research Center of Intelligent Vision and Imaging, ShanghaiTech University, Shanghai, China.

※Yuyao Zhang (E-mail: zhangyy8@shanghaitech.edu.cn) and Xiaojun Xu (E-mail: xxjmailbox@zju.edu.cn) are corresponding authors.

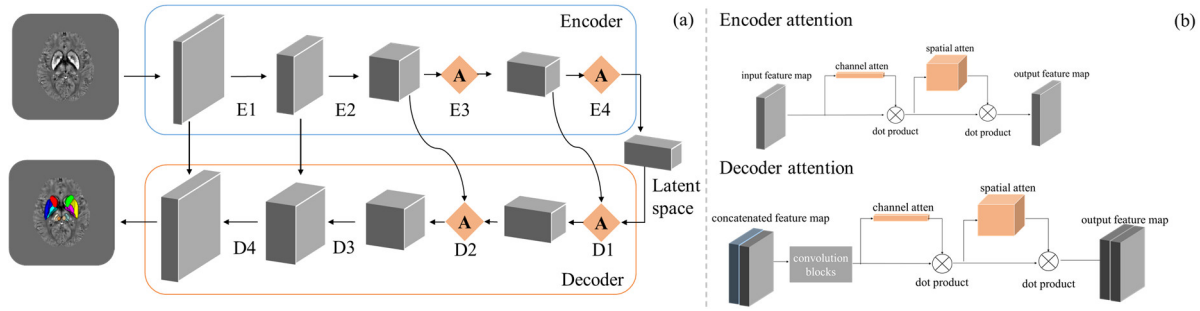

Figure 1. The proposed DeepQSMSeg model architecture. (a) Left part is the entire network architecture. E1 to E4 represent four feature extraction modules and D1 to D4 represent four feature reconstruction modules. Four attention modules are inserted between the last two encoder stages and the first two decoder stages. (b) Right part is attention module architecture in encoder and decoder stage that consists of concatenated channel-wise attention and voxel-wise attention.

the QSM values in each ROI using the susceptibility age development model. The curves fitted using ground truth and DeepQSMSeg were quite close and showed perfect consistency with previous age development study [10], which validates the feasibility of DeepQSMSeg in clinical and research usage. DeepQSMSeg can be easily embedded into diagnosis procedures, thus alleviating the pressure of radiologists and accelerating clinical translation progress.

II. Material and Methods

A. Dataset construction

Data acquisition. Total 631 subjects, including 338 normal controls (NC) and 293 PD patients, were involved in data acquisition. All participants were scanned using a multi-echo 3D gradient echo (GRE) sequence: repetition time = 33.7 ms; first echo time/spacing/eighth echo time = 4.556 ms/3.648 ms/ 30.092 ms; flip angle = 20°; field of view =240×240 mm^2; matrix = 416×384; slice thickness 2 mm; slice gap = 0 mm.

QSM reconstruction. QSM image reconstruction was completed using Susceptibility Tensor Imaging Suite V3.0 software package to compute the magnetic susceptibility from GRE phase images. The raw phase was unwrapped using Laplacian-based phase unwrapping method and the normalized background phase was removed with the V_SHARP method. The final susceptibility map was calculated by STAR-QSM algorithm [11] to acquire the final QSM volume.

Training and testing dataset generation. We chose five pairs of DGM structures (caudate nucleus (CN), putamen (PUT), globus pallidus (GP), substantia nigra (SN) and red nucleus (RN) in both left and right hemispheres) that are most concerned in QSM related research works as segmentation targets. The manual segmentation labels were annotated by experienced radiologists and double-checked, which were used as the ground truth (GT) in both training and testing processes. All images and labels were resampled to 256×256×64. The entire dataset includes 631 pairs of QSM images and ROI labels. The dataset was divided into five independent parts for training and evaluating segmentation accuracy with 5-fold cross validation.

B. Network architecture

The proposed network architecture is shown in Fig.1 (a). The backbone consists of an encoder and a decoder. The encoder aims to extract semantic features into latent space from QSM images and the decoder will reconstruct the segmentation map using latent features. The encoder consists of an input module and four feature extraction modules. Each module includes one to three volumetric anisotropic convolutions blocks (5×5×3) and adopts residual structure to reuse features and achieve fast convergence. In between modules, stride two convolutions (2×2×2) are used to halve the spatial size and double the feature channels. The decoder includes four feature reconstruction modules and one output module, which are symmetric to the encoder. The basic convolution blocks are similar and transpose convolutions are used to double the spatial size and decrease feature channels between modules. We adopted skip connections from encoder to decoder to achieve high-resolution feature maps with strong semantic representation. The network's output, which is activated by soft-max functions, consists of 11 channels (10 ROI structures and one background class) and the resolution is the same as the input volume.

Between the last two encoder stages and the first two decoder stages, we insert attention modules to capture the small target structures' semantic features. The attention module is adapted from CBAM [9], which can be divided into encoder form and decoder form. For encoder attention, attention modules take each stage's output directly as input. For decoder attention, the combined feature maps from encoder and decoder are first processed by convolution blocks and then used as input of attention modules. The attention module is illustrated as Fig.1 (b). It can be considered as a cascade of a channel-wise attention module and a spatial-wise attention module. Channel-wise attention exploits the inter-channel relationship while spatial-wise attention can concentrate on spatial features. For channel-wise attention, global adaptive pooling (both max-pooling and average-pooling) is used to extract the channel-wise feature. Then the channel-wise attention weight is computed via multilayer perceptron (MLP). For spatial-wise attention, average and maximum values along the channel axis are calculated and 1×1×1 convolution is adopted to calculate the spatial-wise attention weight.

C. Loss function

Dice loss combined with voxel-wise focal loss is adopted to alleviate the high imbalance of classes and focus on the poorly classified voxels [8]. Dice loss is a region-based loss derived from Dice similarity coefficient (DSC) [12]. The mathematical formula of dice loss is as follows:

$$Dice\_loss = 1 - \frac{2\sum p_n g_n + \epsilon}{\sum(p_n + g_n) + \epsilon} \quad (1)$$

Where p_n and g_n are respectively the predicted results and ground truth voxels at the same position, and ϵ is the smooth constant.

Although dice loss can eliminate the class imbalance class problem, it is still not robust enough for the severe imbalance situation (such as SN and RN in Fig.2). In our segmentation tasks, the target structures are of small size compared with the whole brain volume. Thus, it tends to overwhelm by the background voxels. Focal loss was first proposed to tackle the classification problem in one-stage object detection [13]. It can distinguish hard and easy examples and make the training process focus on the hard examples. Focal loss can be easily modified to voxel level, which can be written as:

$$Focal\_loss = mean(g_n(1 - p_n)^2 \log(p_n + \epsilon)) \quad (2)$$

The definition of p_n, g_n and ϵ are the same as (1).

According to [8], dice loss combined with voxel-wise focal loss is adopted in our model to concentrate on the hard examples and tackle the class imbalance problem. The total loss is computed as:

$$Dice\_Focal\_loss = Dice\_loss + \lambda Focal\_loss \quad (3)$$

Where λ is the hyperparameter to adjust the weight between dice loss and focal loss.

D. Training and evaluating

Transformation and augmentation were performed on the fly. All volumes were normalized and cropped to patches with a fixed size of 128×128×64 around the center. For training, random spatial transformations were adopted and assigned random parameters within a reasonable range. All images from normal controls and Parkinson's disease patients were used as the input of CNN indiscriminately.

The proposed model was implemented using PyTorch 1.5. The network was trained by Adam optimizer. The initial learning rate is 0.01 and decayed to 1/10 in the 30th, 75th and 300th epoch. The total number of epochs was set to 800. The trade-off coefficient λ in focal loss was set to 0.5.

To evaluate the segmentation accuracy, we calculated the DSC and Hausdorff distance (HD) between the manual segmentation and predicted masks. DSC can evaluate the region similarity while Hausdorff distance can measure the segmentation boundary difference. DSC is similar to (1) but without the smooth constant. Hausdorff distance is defined as:

$$H(G, P) = \max\{h(G, P), h(P, G)\} \quad (4)$$

Where G and P are mask boundaries of ground truth and prediction. h(G, P) is calculated by:

$$h(G, P) = \max_{g_n \in G} \min_{p_n \in P} ||g_n - p_n|| \quad (5)$$

E. Susceptibility development model with age

Segmentation for DGM structures in QSM is commonly used as a preprocessing step for iron-related neuroimage studies. To further prove the feasibility of using DeepQSMSeg to extract susceptibility values in DGM structures, we computed average susceptibility values for each ROI in NC subjects of the test dataset using ground truth and predicted masks, respectively. Then we used two sets (GT & predicted) of susceptibility values to fit the susceptibility development model with age. According to [10], the susceptibility development model with age is as:

$$\chi = \alpha\left(1 - e^{-\beta \times age}\right) + \gamma \quad (6)$$

Where χ denotes the average susceptibility values in a specific ROI, age denotes the subject's age, α, β and γ denote the ROI specific parameters needed to be fitted. The susceptibility development model is fitted using the nonlinear least square method.

III. Results

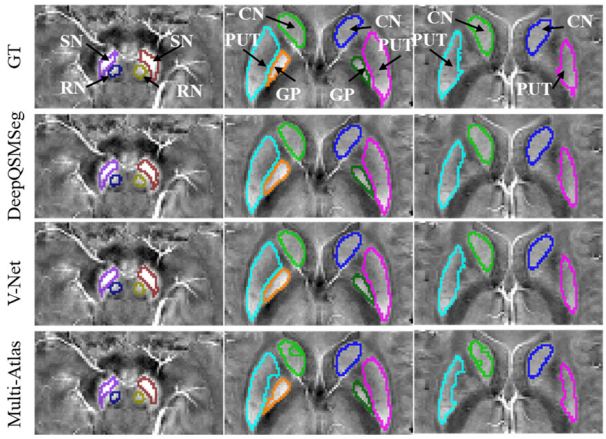

Figure. 2. Qualitative comparison between ground truth and predicted segmentation using different methods. The first row is ground truth, illustrating different ROI names and masks. The other three rows represent segmentation results by different methods.

Fig.2 illustrates the comparison result between ground truth and predicted segmentation using three methods. We compared DeepQSMSeg with V-Net and a popular multi-atlas method. V-Net was implemented according to [12]. Multi-atlas method used ten manually labeled QSM subjects as atlases. The registration and label fusion algorithm was implemented according to [14]. The three axial slices clearly show all the target DGM structures in the left and right hemispheres. It is shown that DeepQSMSeg achieved the best performance compared with other methods, especially in the small nuclei SN and RN.

The quantitative structural consistency of each ROI between ground truth and predicted segmentation was calculated. As shown in Table1, our method provides the overall best performance among all methods. The average DSC for all target structures is 0.872±0.053 and Hausdorff distance is 2.644±2.917mm. Besides, our method can perform QSM segmentation in near-real-time (2.600±0.018s per volume in test dataset), which is much faster than the atlas-based method (4131.300±79.923s).

The structure volumes (in mm^3) and average magnetic susceptibility values (in ppm) in each ROI were extracted based on DeepQSMSeg. As shown in Fig.3, the values extracted by our method are in general consistent and significantly correlated with those extracted using ground truth. The overall correlations of structure volumes and

susceptibility values between manual labels and our method in all target structures are 0.985 ($p<0.001$) and 0.991 ($p<0.001$), respectively.

TABLE I. DSC AND HD (IN MM) BETWEEN GROUND TRUTH AND PREDICTED SEGMENTATION OF DIFFERENT METHODS.

Methods	**DeepQSMSeg**		V-Net		Multi-Atlas	
ROIs	*DSC*	*HD*	*DSC*	*HD*	*DSC*	*HD*
CN	**0.856**	**3.01**	0.848	3.11	0.733	7.05
PUT	**0.882**	**3.02**	0.873	3.32	0.817	5.30
GP	**0.892**	2.88	0.887	**2.68**	0.860	5.14
SN	**0.843**	**2.61**	0.833	3.23	0.816	3.99
RN	**0.889**	**1.70**	0.880	1.76	0.845	2.19

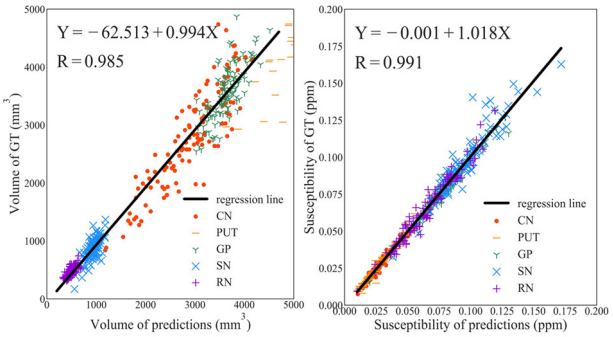

Figure 3. Linear correlation of ROI volumes (left) and average susceptibility (right) between ground truth and predictions.

Fig. 4 shows the fitted susceptibility development models in 5 ROIs using GT and prediction by DeepQSMSeg. We can see that the models fitted using ground truth and DeepQSMSeg are quite close. Besides, the fitted models are in high agreement with previous work [10].

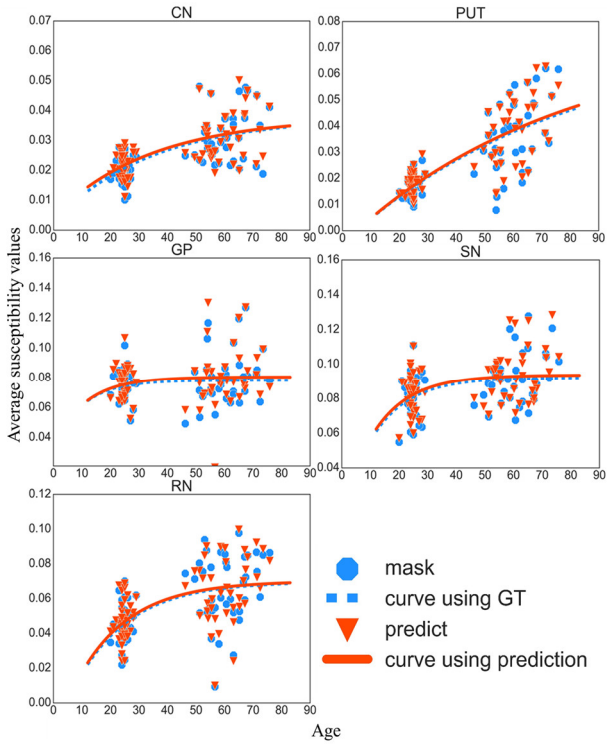

Figure 4. Susceptibility development model with age.

IV. CONCLUSIONS

In the present study, we propose an accurate, automatic and fast DGMs structures segmentation tool for QSM, named DeepQSMSeg. DeepQSMSeg is adaptive for the high anatomical variability and target imbalance and achieves similar performance as manual masks, thus proves its potential in research and clinical usage. It can be easily embedded into diagnosis procedures and accelerate the progress of clinical translation. Since there is still a lack of QSM data in clinical practice, in the future, we will focus on multi-modality DGMs segmentation through a weakly-supervised manner to better suit the clinical need.

V. COMPLIANCE WITH ETHICAL STANDARDS

The experimental procedures involving human subjects described in this paper were approved by the Ethics Committee of the second affiliated hospital of Zhejiang university school of medicine.

REFERENCES

[1] B. Hallgren *et al.*, "The effect of age on the non-haemin iron in the human brain" *Journal of Neurochemistry*, vol. 3, no. 1, pp. 41-51, 1958.

[2] R. H. Benedict *et al.*, "Memory impairment in multiple sclerosis: correlation with deep grey matter and mesial temporal atrophy," *J Neurol Neurosurg Psychiatry*, vol. 80, no. 2, pp. 201-6, Feb 2009.

[3] Y. Ge *et al.*, "Quantitative assessment of iron accumulation in the deep gray matter of multiple sclerosis by magnetic field correlation imaging," *AJNR Am J Neuroradiol*, vol. 28, no. 9, pp. 1639-44, Oct 2007.

[4] Q. Chen *et al.*, "Iron deposition in Parkinson's disease by quantitative susceptibility mapping," *BMC Neurosci*, vol. 20, no. 1, p. 23, May 22 2019.

[5] X. Li *et al.*, "Multi-atlas tool for automated segmentation of brain gray matter nuclei and quantification of their magnetic susceptibility," *Neuroimage*, vol. 191, pp. 337-349, 2019.

[6] S. S. Raj *et al.*, "A Deep Approach to Quantify Iron Accumulation in the DGM Structures of the Brain in Degenerative Parkinsonian Disorders Using Automated Segmentation Algorithm," in *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*. IEEE, 2019, pp. 1-8.

[7] C. Chai *et al.*, "Automated Segmentation of Brain Gray Matter Nuclei on Quantitative Susceptibility Mapping Using Deep Convolutional Neural Network," 2020.

[8] W. Zhu *et al.*, "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Medical physics*, vol. 46, no. 2, pp. 576-589, 2019.

[9] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.

[10] Y. Zhang, H. Wei, M. J. Cronin, N. He, F. Yan, and C. J. N. Liu, "Longitudinal atlas for normative human brain development and aging over the lifespan using quantitative susceptibility mapping," *Neuroimage*, vol. 171, pp. 176-189, 2018.

[11] H. Wei *et al.*, "Streaking artifact reduction for quantitative susceptibility mapping of sources with large dynamic range," *NMR in Biomedicine*, vol. 28, no. 10, pp. 1294-1303, 2015.

[12] F. Milletari *et al.*, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565-571.

[13] T.-Y. Lin *et al.*, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.

[14] H. Wang *et al.*, "Multi-atlas segmentation with joint label fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611-623, 2012.