Enhancement of Perivascular Spaces Using a Very Deep 3D Dense Network

Euijin Jung[1], Xiaopeng Zong[2], Weili Lin[2], Dinggang Shen[2(✉)],
and Sang Hyun Park[1(✉)]

[1] Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science
and Technology, Daegu, South Korea
shpark13135@dgist.ac.kr
[2] Department of Radiology, BRIC, University of North Carolina, Chapel Hill, USA
dinggang_shen@med.unc.edu

Abstract. Perivascular spaces (PVS) in the human brain are related to various brain diseases or functions, but it is difficult to quantify them in a magnetic resonance (MR) image due to their thin and blurry appearance. In this paper, we introduce a deep learning based method which can enhance a MR image to better visualize the PVS. To accurately predict the enhanced image, we propose a very deep 3D convolutional neural network which contains densely connected networks with skip connections. The densely connected networks can utilize rich contextual information derived from low level to high level features and effectively alleviate the gradient vanishing problem caused by the deep layers. The proposed method is evaluated on seventeen 7T MR images by a twofold cross validation. The experiments show that our proposed network is more effective to enhance the PVS than the previous deep learning based methods using less layers.

Keywords: Perivascular spaces · Enhancement
Deep convolutional neural network · Densely connected network
Skip connections

1 Introduction

Perivascular spaces (PVS) are thin fluid-filled spaces in the human brain. Recently, studies have shown that increasing the PVS number and thickening the PVS are associated with brain diseases [1]. Also, it is revealed that the PVS enlargement is related to cognitive abilities of healthy elderly men [2]. To demonstrate these hypotheses, it is necessary to quantify the relationship between the thickness, length, distribution of PVS and the brain diseases or functions.

However, the PVS are not clearly visible in magnetic resonance (MR) images acquired by traditional 1.5T, 3T or even by 7T MR scanners. Accordingly, Bouvy *et al.* [3] and Zong *et al.* [4] proposed novel acquisition parameters of 7T MR scanner that make the PVS more visible. However, it is difficult to find the

parameters which can improve only the PVS while reducing the noisy in background. Thus, distinguishing small PVS is still difficult although several methods have been proposed to segment the PVS from MR images [5,6].

Accordingly, instead of carefully looking for a certain specific parameter of MR scanner, several studies have been proposed to enhance the PVS by using image processing methods after the MR images are acquired. For example, Uchiyama *et al.* [7] used the white top hat transform to highlight the tubular structures and proved that this enhancement is effective to detect the PVS. Hou *et al.* [8] proposed a method which improves the intensity of thin tubular structures using a nonlinear mapping function in Haar domain, and then removes noisy in background by using the block matching filtering. Although these methods help to extract the PVS by enhancing the intensity of PVS, they require heuristic parameter tuning such as controlling the filter size or defining the parameters of nonlinear mapping function according to the image.

In this paper, we propose an end to end PVS enhancement method which does not require the heuristic parameter tuning and the additional processing steps for distinguishing the PVS from noisy. Specifically, we suggest a very deep 3D neural network consisting of 39 convolution layers which are densely connected by skip connections. The proposed network using the dense skip connections effectively improves the prediction accuracy by utilizing rich contextual information derived from low level to high level features and alleviating the gradient vanishing problem. The prediction accuracy of our proposed network was evaluated on seventeen 7T MR images. Experimental results show that our deep network is more effective to enhance the PVS than the state-of-the-art deep learning based image enhancement methods.

1.1   Related Works

Deep learning based methods have achieved the best performance for the super resolution problem which converts a low resolution image into a high resolution image. For example, Dong *et al.* [9] proposed a method using three convolution layers and achieved better prediction results than the previous methods using sparse coding and regression. After that, several studies using deeper network [10,11] have been proposed to utilize higher level contextual features. Specifically, Kim *et al.* [10] proposed a recursive neural network to reflect a large contextual information without additional weight parameters and Tong *et al.* [11] proposed a network using densely connected blocks with skip connections to reflect the various levels of features for the prediction.

In this paper, we apply the deep neural networks, mainly have been applied to the super resolution of 2D images, to the enhancement of PVS in 3D MR images. The PVS are thin and oriented at different angles in three dimensions, and thus it is difficult to distinguish the PVS from noisy in a 2D image. In addition, since the difference between a MR image and its enhanced MR image is relatively larger (see Fig. 2) than that between the low resolution image and the high resolution image in super resolution, sophisticated contextual features need to be learned. Therefore, we design a very deep 3D network including six

dense blocks and dense skip connections to reduce the feature redundancy and utilize the rich contextual information in three dimensions. Although several 3D networks [12–14] recently have been proposed for the super resolution of MR images, those models use shallow structures while our model includes six dense blocks and skip connections between them. The closest model to our proposed network is the network proposed by Tong *et al.* [11], but our model consists of 3D layers and there are some differences in the structure such as not using a deconvolution layer. To the best of our knowledge, this is the first work to use the deep learning based method for the PVS enhancement.

2   Method

We introduce a deep learning based method which generates an enhanced 7T MR image from a 7T MR image. Learning a deep network that maps the whole 3D MR image is infeasible due to memory limitations. Thus, if an image is given, we sample 3D patches at a regular interval, and then perform the prediction in each patch using a deep 3D convolutional neural network, and finally generate the whole enhanced image by merging the predictions on the 3D patches. Since the predictions near the boundary of patch may not be accurate, the predictions on the central region are collected to generate the whole enhanced image. The sampling interval is determined so that the prediction is obtained in every voxel.

In the training step, we sample the 3D patches from 7T MR images and those from their enhanced 7T MR images in a training set, and then learn the deep 3D convolutional neural network which learns the relationship between patches. The proposed network consists of an initial convolution layer for learning low level features, several dense blocks for learning middle level to high level features, a bottleneck layer for reducing the number of feature maps, and a prediction layer for generating the enhanced 3D patch. Figure 1 shows the proposed network and detailed descriptions follow in the subsections.
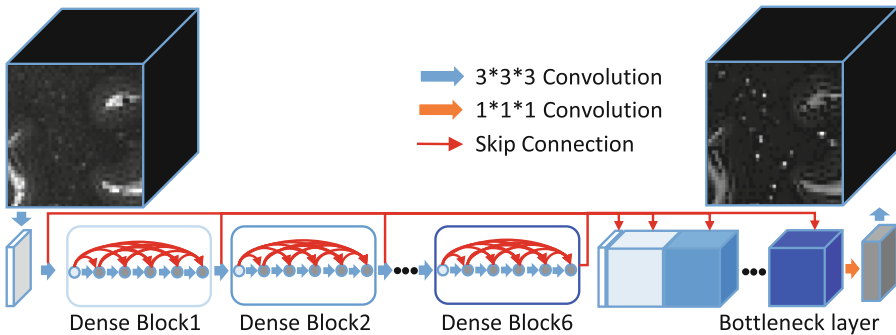
Fig. 1. The proposed deep 3D convolutional neural network for PVS enhancement.

2.1   Densely Connected Deep Neural Network

The proposed network learns the relationship between the patch X sampled from a 7T MR image and the patch Y from its enhanced 7T MR image. The relevance is parameterized by weights $\mathbf{w} = [w_1, ..., w_N]$ and residuals $\mathbf{b} = [b_1, ..., b_N]$ between layers where N is the number of convolution layers, and X is transformed into $P(X, \mathbf{w}, \mathbf{b})$ by those parameters. In training, the parameters $\mathbf{w}$ and $\mathbf{b}$ are updated by an optimizer so that the mean squared error between $P(X, \mathbf{w}, \mathbf{b})$ and Y is minimized.

The proposed network consists of 39 convolution layers ($N = 39$). First, the input patch X is passed through a convolution layer and then six dense blocks where each dense block consists of 6 convolution layers to produce low level to high level feature maps. Specifically, 8 kernels with a size $3 \times 3 \times 3$ is used for the convolution layers and a rectified linear unit (ReLU) layer is connected for nonlinear mapping behind each convolution layer.

In each dense block, as proposed by Huang *et al.* [15], the feature maps generated in previous layers are concatenated and pass through a convolution layer to generate new feature maps. The new feature maps are also concatenated to the previous feature maps and then pass through the next convolution layer. Thus, the number of feature maps linearly increased by the number of kernel. Since we use six convolution layers with 8 kernels, the number of feature maps increased by 8 in six times and the dense block generates 48 feature maps. The concatenation of the feature maps not only reduces the number of parameters but also alleviates the vanishing gradient problem. Finally, the 8 feature maps generated from the last layer are used as the input of the next dense block.

After passing through all six dense blocks, the prediction can be performed by using the feature maps from the 6th dense block. However, in this way, the low level and middle level features extracted by the initial layer and the initial dense blocks are rarely reflected in the prediction. Thus, to use all levels of information for the prediction, we use skip connections between the following layer and the initial convolution layer and six dense blocks. Specifically, 8 feature maps obtained from the initial convolution layer and all 288 ($= 48 \times 6$) feature maps from six dense blocks are connected to the following layer in the network.

Connecting all these feature maps to the prediction layer for predicting a single channel output at once (*i.e.*, 296 to 1) is computationally inefficient and hard to keep the model compactness. Therefore, a $1 \times 1 \times 1$ convolution layer with 16 kernels is utilized as the bottleneck layer between the 6th dense block and the prediction layer to reduce the number of feature maps. Finally, the 16 feature maps generated from the bottleneck layer are passed through the prediction layer to predict the final output (*i.e.*, 296 to 16, and then 16 to 1). With through the bottleneck layer, prediction can be more accurate and efficient, since this layer use all feature map from low to high levels and reduce the number of feature map in computationally efficient way.

2.2   Implementation Details

Most PVS are located in the white matter and the non-brain region is large in a MR image. Thus, it is inefficient to sample the training patches in the whole image. We extracted the brain region by using the brain extraction tool [16] and then sampled 3D patches which contain a part of brain region for training. The patch size was determined as $60 \times 60 \times 60$ by considering the receptive field of our network. In testing, we similarly extracted the brain region using [16], and then estimated the enhanced image by performing the prediction on $60 \times 60 \times 60$ 3D patches containing the brain region and merging them.

Regarding the proposed network, the weights $\mathbf{w}$ were initialized by the method proposed in [17] and the biases $\mathbf{b}$ were initialized to 0. ReLU was used for the activation function and the batch size was set as 5. The Adam optimizer was used to minimize the mean squared error between $P(X, \mathbf{w}, \mathbf{b})$ and Y. The learning rate was initially set as 0.0001 and then decreased by 2×10^{-7} for each epoch. The experiment was ended up to 500 epochs. The method was implemented using Tensorflow and all training and testing were performed on a workstation with NVIDIA Titan XP GPU.

3   Experimental Results

3.1   Evaluation Setting

Seventeen 7T MR images were used for the experiment. For training and validation, we made those enhancement images by using the Hou *et al.*'s method [8]. The enhanced images were used for computing the mean square error in training, while used for evaluating the prediction accuracy in testing. We divided the images into two subsets and then performed a two-fold cross validation.

The prediction accuracy was measured by PSNR and SSIM between the predicted images and the enhanced images. The PSNR and SSIM were measured in the white matter as well as in the whole brain region since most PVS were in the white matter. The white matter was extracted by an brain tissue segmentation method [18].

To demonstrate the superiority of the proposed network (DCNN6+SC+B) using the six dense blocks, skip connections (SC), and bottleneck layer (B), we compared this with SRCNN [9] using three convolution layers with the kernel sizes 9, 5, and 5 and DCNN [13] using only one dense block for the prediction. To demonstrate the effect of skip connections between the dense blocks and the bottleneck layer, we provide the results obtained by the deep networks without the skip connections and the bottleneck layer (DCNN6 and DCNN6+SC). In addition, to demonstrate the effect of network depth related to the number of parameters and the size of receptive field, we provide the results obtained by using the proposed networks with two and four dense blocks (DCNN2+SC+B and DCNN4+SC+B, respectively) instead of six dense blocks.

For a fair comparison, we modified 2D SRCNN [9], which was proposed for the image super resolution problem, to the 3D network to address the PVS

enhancement problem. Also, we modified the kernel size and the number of layers of DCNN [13], which was proposed for the super resolution of a brain MR image, to be comparable with our network.

Table 1. Mean PSNR (dB) and SSIM scores between the predictions and the enhanced images, and the training time for each method. The scores were measured in the white matter (WM) and in the brain region (Brain), respectively. SC represents the skip connections, B represents the bottleneck layer, and bold indicates the highest score.

	PSNR-WM	PSNR-Brain	SSIM-WM	SSIM-Brain	Time (hour)
SRCNN [9]	36.373	30.957	0.962	0.924	4.5
DCNN [13]	37.647	31.951	0.971	0.942	2.5
DCNN6	38.518	32.825	0.975	0.950	12
DCNN6+SC	38.636	32.918	0.975	**0.951**	14.5
DCNN2+SC+B	38.040	32.386	0.973	0.947	4
DCNN4+SC+B	38.420	32.718	0.975	0.949	8
DCNN6+SC+B	**38.739**	**33.015**	**0.976**	**0.951**	12.5

3.2   Result

Table 1 shows the mean PSNR and SSIM measured from the results obtained by the proposed method and the comparison methods, and the computational times for training. The result obtained by SRCNN was the worst since the small number of hidden layers could not produce the high level features useful for prediction. DCNN achieved better performances than SRCNN with less computations. The deeper network and the skip connections between convolution layers helped to use relatively high level features while reducing the number of parameters. Likewise, DCNN6 composed of approximately six times more layers achieved much better results since the deeper network could learn the higher level features on a large receptive field which could not be considered in DCNN.

The method using the dense skip connections (DCNN6+SC) further improved the performance by predicting the enhanced image with the low level to high level features together on a large receptive field. Using the bottleneck layer also helped to improve the performance slightly while reducing the computation (DCNN6+SC+B). According to the results obtained by DCNN2+SC+B, DCNN4+SC+B, and DCNN6+SC+B, we could confirm that the performance was improved as the depth of network deepened.

Figure 2 shows the qualitative results obtained by SRCNN, DCNN, and the proposed method. SRCNN or DCNN improved the PVS, but noises near the PVS were not suppressed effectively. On the other hand, the prediction results obtained by our proposed method were very similar to the enhanced images.
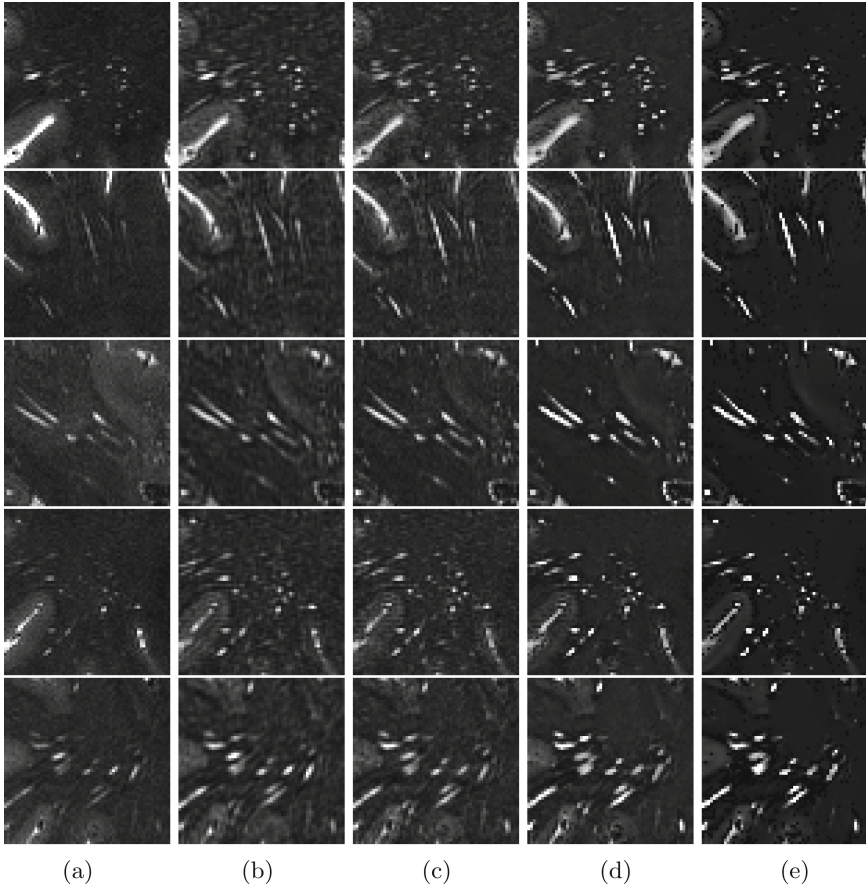
Fig. 2. Visual comparison between the proposed method and the comparison methods on several local regions. (a) Regions in original images, (b) the results obtained by SRCNN [9], (c) the results by DCNN [13], (d) the results by our proposed method (DCNN6+SC+B), and (e) regions in the enhanced images.

4 Conclusion

We have proposed a novel PVS enhancement method using a deep dense network with skip connections. We have demonstrated that the deep learning techniques usually used for the super resolution problem can be used for the PVS enhancement problem. The proposed method does not require empirical parameter tuning and additional processing such as denoising. The proposed deep network has outperformed the state-of-the-art deep learning networks and it has been proved that using various levels of features is helpful to improve the prediction accuracy. In the future, we will perform several experiments to prove how the proposed method can help in PVS segmentation and quantitative analysis.

References

1. Zhu, Y.C., et al.: Severity of dilated Virchow-Robin spaces is associated with age, blood pressure, and MRI markers of small vessel disease: a population-based study. Stroke **41**(11), 2483–2490 (2010)
2. Maclullich, A.M., et al.: Enlarged perivascular spaces are associated with cognitive function in healthy elderly men. J. Neurol. Neurosurg. Psychiatry **75**(11), 1519–1523 (2004)
3. Bouvy, W.H., et al.: Visualization of perivascular spaces and perforating arteries with 7T magnetic resonance imaging. Invest. Radiol. **49**(5), 307–313 (2014)
4. Zong, X., et al.: Visualization of perivascular spaces in the human brain at 7T: sequence optimization and morphology characterization. NeuroImage **125**, 895–902 (2016)
5. Park, S.H., et al.: Segmentation of perivascular spaces in 7T MR images using auto-context model with orientation-normalized features. NeuroImage **134**, 223–235 (2016)
6. Zhang, J., et al.: Structured learning for 3D perivascular spaces segmentation using vascular features. IEEE Trans. Biomed. Eng. **64**(12), 2803–2812 (2017)
7. Uchiyama, Y., et al.: Computer-aided diagnosis scheme for classification of lacunar infarcts and enlarged Virchow-Robin spaces in brain MR images. In: Conference Proceedings of IEEE Engineering in Medicine and Biology Society (2008)
8. Hou, Y., et al.: Enhancement of perivascular spaces in 7T MR image using Haar transform of non-local cubes and block-matching filtering. Sci. Rep. **7**, 8569 (2017)
9. Dong, C., et al.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
10. Kim, J., et al.: Deeply-recursive convolutional network for image super-resolution. In: Computer Vision and Pattern Recognition (2016)
11. Tong, T., et al.: Image super-resolution using dense skip connections. In: International Conference on Computer Vision (2017)
12. Pham, C.H., et al.: Brain MRI super-resolution using deep 3D convolutional networks. In: International Symposium on Biomedical Imaging (2017)
13. Chen, Y., et al.: Brain MRI super resolution using 3D deep densely connected neural networks. In: International Symposium on Biomedical Imaging (2018)
14. Shi, J., et al.: MR image super-resolution via wide residual networks with fixed skip connection. IEEE J. Biomed. Health Inf., 2168–2194 (2018)
15. Huang, G., et al.: Densely connected convolutional networks. In: Computer Vision and Pattern Recognition (2017)
16. Smith, S.: Fast robust automated brain extraction. Hum. Brain Mapp. **17**(3), 143–155 (2002)
17. He, K., et al.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Computer Vision and Pattern Recognition (2015)
18. Zhang, Y., et al.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging **20**(1), 45–57 (2001)