Original contribution

Denoising arterial spin labeling perfusion MRI with deep machine learning

Danfeng Xie[a], Yiran Li[a], Hanlu Yang[a], Li Bai[a], Tianyao Wang[c], Fuqing Zhou[d], Lei Zhang[b], Ze Wang[b],*

[a] Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA, USA
[b] Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD, USA
[c] Department of Radiology, the Fifth People's Hospital of Shanghai, Fudan University, Shanghai, China
[d] Department of Radiology, the First Affiliated Hospital, Nanchang University, Nanchang, China

A R T I C L E   I N F O

A B S T R A C T

Purpose: Arterial spin labeling (ASL) perfusion MRI is a noninvasive technique for measuring cerebral blood flow (CBF) in a quantitative manner. A technical challenge in ASL MRI is data processing because of the inherently low signal-to-noise-ratio (SNR). Deep learning (DL) is an emerging machine learning technique that can learn a nonlinear transform from acquired data without using any explicit hypothesis. Such a high flexibility may be particularly beneficial for ASL denoising. In this paper, we proposed and validated a DL-based ASL MRI denoising algorithm (DL-ASL).

Methods: The DL-ASL network was constructed using convolutional neural networks (CNNs) with dilated convolution and wide activation residual blocks to explicitly take the inter-voxel correlations into account, and preserve spatial resolution of input image during model learning.

Results: DL-ASL substantially improved the quality of ASL CBF in terms of SNR. Based on retrospective analyses, DL-ASL showed a high potential of reducing 75% of the original acquisition time without sacrificing CBF measurement quality.

Conclusion: DL-ASL achieved improved denoising performance for ASL MRI as compared with current routine methods in terms of higher PSNR, SSIM and Radiologic scores. With the help of DL-ASL, much fewer repetitions may be prescribed in ASL MRI, resulting in a great reduction of the total acquisition time.

1. Introduction

Arterial spin labeling (ASL) perfusion MRI is a non-invasive technique for measuring cerebral blood flow (CBF) [1,2]. In ASL, arterial blood water is labeled with radio-frequency (RF) pulses in locations proximal to the tissue of interest. A perfusion-weighted MR image is acquired after the labeled spins reach the imaging place and perfuse into brain tissue. To remove the background MR signal, a control image is also acquired using the same ASL sequence and acquisition timing but with phase modulations to the labeling pulses (by alternating phase of adjacent pulses to be either 0 or 180 degree in the pseudo-continuous ASL) so that arterial spins can approximately stay unaffected. Perfusion signal is subsequently determined by pair-wise subtraction of the spin labeled image (L image) and the spin untagged image (the control image or C image) which is then converted into the quantitative CBF in a unit of ml/100 g/min [3]. Limited by the longitudinal relaxation rate (T1) of blood water, labeling efficiency, and the post-labeling delay, the labeled blood signal is small, resulting in a low SNR [4]. Many pairs of

L/C images are often acquired to improve SNR of the mean perfusion map. Because the total scan time is often around 3–6 min, only 10–50 L/C pairs can be acquired, resulting in a modest SNR improvement by averaging across the limited number of measurements. A number of post-processing methods have been proposed to reduce different types of artifacts in ASL data, including ASL MRI specific motion correction [5], physiological noise correction [6], and spatial noise reduction [7–9]. Advanced methods have also been published to suppress non-local noise [10], spatio-temporal noise [11–14], and outliers [15–18]. These methods are based on either implicit or explicit models about the data, which may not be accurate and may change across subjects.

Deep learning (DL) is a subtype of machine learning (ML) algorithms [19] which has made widespread impact on nearly every research field it has been applied (from image classification [20,21], video recognition [22–24], voice recognition/generation [25–28], medical image processing [29,30], to AlphaGo [31], AlphaGo Zero [32], etc.). The concept of DL can be traced back to early 1980s [33],

but only became practical until the advent of fast general-purpose graphics processors in late 2000s [20,34]. DL is now dominating nearly every field it has reached such as classification, computer vision, auditory processing, information generation, and translational research [21,29,30,35,36]. Using a hierarchical multiple layers (deep) of "neurons" (the processing units) with a greedy layer-wise training, DL can reliably learn any nonlinear function from the sampled data [35,37,6,19]. In natural image denoising, DL-based denoising has been reported by a few papers [38–42]. The most popular DL-based denoising approach is based on convolutional neural networks (CNNs) [21]. CNNs learn a hierarchy of features by a series of convolution, feature pooling, and non-linear activation operations, presenting a high flexibility and capability for learning distributions or manifold of images [42]. Encouraged by the outstanding performance as listed above, DL has been introduced into many medical imaging processing fields, including image segmentations [29,30], image reconstructions [43,44], image synthesis [45,46] etc.

The purpose of this study was to test DL for ASL MRI denoising. For the simplicity of description, we dubbed our DL-based ASL denoising method as DL-ASL thereafter. One journal paper [47], one conference paper [48], and a conference abstract [49] have been published in ASL denoising focusing on different scopes of work. Kim et al. [47] published the first paper on this research topic. Their denoising CNNs consist of two parallel pathways to integrate the multi-scale contextual information. As an initial study, the model was trained with a small dataset and the CNN architecture adapted therein was originally designed for image segmentations [50,51], which may not be optimal for denoising. Ulas et al. [48] trained a deep learning model with a customized loss function based on the Buxton Kinetic model [52] but with a simple CNN architecture. Gong et al. [49] proposed a technique first using multi-lateral guided filter to pre-process input data, generating denoised ASL with different smoothing levels. Then they combined a stack of multi-contrast images as input to train a deep learning network for final CBF denoising. However, the multi-lateral filter is a local filter which cannot incorporate global information for denoising. Generating a stack of multi-contrast image could be time-consuming. While encouraging, these studies were all based on a small sample size and used a standard CNN not specifically optimized for denoising.

By contrast, our paper presents two novelties: first, we incorporated wide activation residual blocks [53] with a Dilated Convolution Neural Network (DilatedNet) [54] to achieve improved denoising performance in term of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) index; second, we evaluated our proposed model given different inputs and references to show that DL-ASL denoising model can be generalized to input with different levels of SNR and yielded images with better quality than other methods.

The rest of this paper is organized as follows. In Section 2, we present our proposed DL-ASL architecture in details. Section 3 demonstrates the experiment of our methods on clinical ASL datasets. Section 4 discusses the main contributions and results of this work. In Section 5, we conclude the paper with several remarks.

2. Materials and methods

2.1. Problem formulation

Traditional denoising methods often rely on establishing an explicit model for either signal or noise, so the two components can be easily separated from their mixtures. By contrast, DL-based denoising methods learn the denoising model directly from the noise contaminated data. Denote the noisy image by x_i and its reference (noise-free or less noisy version) by y_i, where $i = 1…N$, N is the total number of training samples. A parametric regression model f_Θ, typically a convolutional neural network (CNN), can be built to learn the mapping $f_\Theta(x_i) \rightarrow y_i$ by minimizing the following loss function:

$$\sum_i L(f_\Theta(x_i), y_i),$$

where Θ are the parameters of CNN and are adjusted through the training process.

Assume that the i-th subject's CBF image x_i is a summation of the latent noise-free CBF image y_i and unknown noise n_i: $x_i = y_i + n_i$. DL-ASL is to learn an implicit transform which can predict y_i from x_i, so that $f_\Theta(x_i) \rightarrow y_i$. In DL-ASL, we used mean CBF map obtained without using any explicit denoising as the input and the image denoised with the current state-of-the-art methods (see below) as the reference y_i. To avoid potential overfitting due to the high similarity between x_i and y_i, DL-ASL is configured to learn their difference n_i following the residual learning strategy [41,42,53,54]. Correspondingly, the transform to learn became $T(x_i) = y_i - x_i = - n_i$. During testing or new image denoising, the output of DL-ASL was added up to the input to generate the final denoised CBF map y_i. Mean Square Error (MSE) and Mean Absolute Error (MAE) are standard loss function used for training. According to [42,53], MAE is better for image restoration tasks because MSE tends to give over-smoothed results. Thus, to control the training accuracy, MAE between input and reference images was used as our loss function:

$$L(\Theta) = \frac{1}{N} \sum_{i=1}^{N} |y_i + T(x_i; \Theta) - x_i|$$
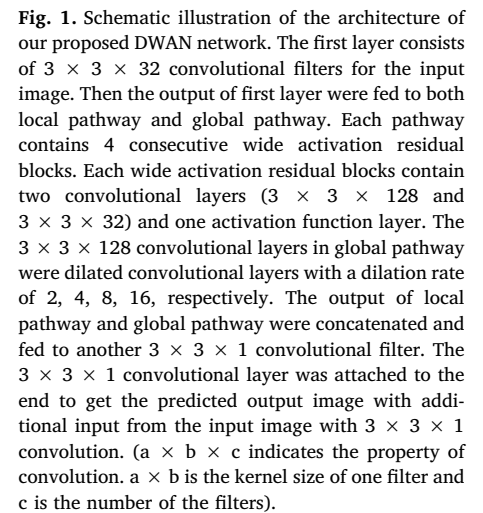
2.2. Network architecture

Fig. 1 illustrates the new CNN structure that was used to build our DL-ASL network. Overall it was a residual learning CNN [55] but with dilated filters [54] replacing the regular non-dilated filters at each layer and with wide activation residual blocks [53] replacing the regular residual blocks [55]. DilatedNet was used due to the capability of extracting both local and global contextual features [54]. Similar to [47], we used two pathways: a local path for extracting the local features and a global path for preserving global data patterns. Wide activation residual blocks are new in the literature but were shown to be able to expand data features and pass more information through the network [53]. As this combination was new, we dubbed it as Dilated Wide Activation Network (DWAN) thereafter.

2.2.1. Wide activation residual block

A regular residual block for low-level vision tasks (e.g., image denoising and image super-resolution) has a convolution layer with a ReLU activation layer, followed by another convolution layer. Each convolution layer has the same number of filters inside a regular residual block. The direct connection from input to output represents the residual learning [55], a key element of modern CNNs for the purpose of improving accuracy and training speed [41,42,53,55]. Since the ReLU layer inside the residual block may cause information loss before being passed to next layer [53], the so-called wide activation residual block uses more filters to extract more features in the layer preceding the ReLU layer, which can subsequently pass more information to the next block [53]. For denoising, this modification would preserve more high resolution information to reduce the blurring effects as supported by the results in image super-resolution [53] and image restoration [56]. In this paper, the number of filters of the convolution layer before the ReLU activation layer was expanded by a factor of 4. After the ReLU activation layer, the following convolution layer shrinks the number of feature maps back to input size.

2.2.2. Residual learning and batch normalization

DL-ASL was configured to learn the residuals between the input and the reference. The residual learning was implemented through a 3 × 3 convolution [53] from the input layer to the output layer. Though Kim et al. [47] and Gong et al. [46] used Batch Normalization (BN) [57] layer in their network, recent researches [53,58–60] found that BN

Fig. 1. Schematic illustration of the architecture of our proposed DWAN network. The first layer consists of $3 \times 3 \times 32$ convolutional filters for the input image. Then the output of first layer were fed to both local pathway and global pathway. Each pathway contains 4 consecutive wide activation residual blocks. Each wide activation residual blocks contain two convolutional layers ($3 \times 3 \times 128$ and $3 \times 3 \times 32$) and one activation function layer. The $3 \times 3 \times 128$ convolutional layers in global pathway were dilated convolutional layers with a dilation rate of 2, 4, 8, 16, respectively. The output of local pathway and global pathway were concatenated and fed to another $3 \times 3 \times 1$ convolutional filter. The $3 \times 3 \times 1$ convolutional layer was attached to the end to get the predicted output image with additional input from the input image with $3 \times 3 \times 1$ convolution. ($a \times b \times c$ indicates the property of convolution. $a \times b$ is the kernel size of one filter and c is the number of the filters).

undermined the accuracy for image super-resolution tasks. We conducted experiments and found out that BN did not improve PSNR and SSIM in our case either. Thus, we discarded BN layers.

2.3. Image acquisition and preprocessing

ASL data were pooled from 280 subjects (normal healthy subjects with age from 23 to 47, 110 females, 170 males) in local database. The data were acquired in a Siemens Trio 3 T scanner using a pseudo-continuous ASL sequence [61,62] (40 control/labeled image pairs with labeling time = 1.5 s, post-labeling delay = 1.5 s, FOV = 22 cm, matrix = 64×64, TR/TE = 4000/11 ms, 20 slices with a thickness of 5 mm plus 1 mm gap). ASLtbx [63] was used to preprocess ASL images using the following updated procedures: 1) ASL-specific motion correction method was applied to the raw ASL images (C/L images) to correct systematic label/control labeling induced spurious motions [5]; 2) the average of all 40C/L image pairs was calculated and used as a template for registering the ASL C/L images to the high-resolution T1 image. Registration was performed with SPM12 (Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ucl.ac.uk/spm); 3) simple regression was used to regress out residual motions, mean CSF signal and global signal; an isotropic Gaussian kernel with full-width-half-maximum = 3 mm was used to smooth ASL C/L images subsequently; 4) adjacent C and L images were subtracted using simple subtraction to generate perfusion-weighted images which were then converted into quantitative CBF using the same method as in [63]. M0 is approximated by the control image in each label/control image pair and M0 calibration is performed at each voxel separately using the

value at the corresponding voxel location of the control image. Outlier CBF image timepoints were identified and removed using the priors-guided slice-wise adaptive outlier cleaning algorithm [17,18]; 5) each subject's structural MRI was spatially normalized to the Montreal Neurologic Institute (MNI) standard brain using SPM12. The same transform was then applied to the CBF image series; 6) mean CBF maps were calculated from different number of CBF images and were named as meanCBF-n where n indicates the number of CBF images to be averaged. meanCBF-40 was the one calculated from the full ASL scan. The same procedures were repeated after removing the spatial smoothing step and the corresponding mean CBF maps were named by an affix "nsm".

2.4. Data preparation and model training

To maximally demonstrate the benefit of DL-ASL, we chose meanCBF-10_nsm (mean CBF map of the first 10 ASL C/L pairs without being spatially smoothed during preprocessing) as the input to the DL-ASL models. As there is no groundtruth (GT) to ASL CBF maps, meanCBF-40 – mean CBF map of the entire 40C/L image pairs, was used as a surrogate GT (the reference) for training the DL-ASL models. The same model was also trained for projecting the meanCBF-10_nsm to meanCBF-40_nsm and name appended by the suffix 'nsm' was added to the name of each model to mark the difference of both input and the reference during model training.

200 subjects' CBF maps were randomly selected as the training set. 20 subjects' CBF maps were used as validation samples during training. The remaining 60 subjects were used as the test set. For each subject, we extracted axial slices every 3 slices from slice 36 to slice 60 of 3D CBF maps in the MNI space, resulting a training set containing $200 \times 9 = 1800$ 2D CBF image slices. Each image slice had 91×109 pixels.

U-Net [64], a popular CNN structure widely used in medical imaging, was implemented as a comparison to our DWAN-based DL-ASL. Because U-Net has pooling and upsampling layers that requires the image size equal to the power of 2, all the images were resized to 128×128 during the training and testing of U-Net, by centering images and padding 0 s on each side. U-Net, U-Net with BN (U-Net_BN) [45], DilatedNet and DilatedNet with BN (DilatedNet_BN) [47] were implemented as additional comparisons. To test the denoising capability of DL-ASL for inputs with different noise levels, we applied the same model (trained by meanCBF-10_nsm) to meanCBF-15, 20, 25, 30, 35 and 40 datasets (mean CBF maps obtained from 15, 20, 30, 35, and 40 L/C images).

All networks were implemented using the Keras platform [65]. The network was trained using Adaptive moment estimation (ADAM) algorithm [66] with basic learning rate of 0.001. All the models were trained with batches, each containing 64 training samples. DWAN was trained for 10,000 iterations and DWAN_nsm was trained for 5000 iterations. Tensorflow [67] was used as the backend of Keras to train all the models. All experiments were performed on a PC with Intel(R) Core (TM) i7-5820k CPU @3.30GHz and a Nvidia GeForce Titan Xp GPU.

2.5. Effects of different model configurations

An experiment was conducted to assess the performance of DL-ASL containing different number of wide activation residual blocks. A grid search experiment was conducted to find the best combination of expansion rate E and the number of input filter M for each wide activation residual block. Noted that the regular residual block [55,58] is essentially a special case of wide activation residual block [53] when the expansion rate E is 1. Thus, by setting expansion rate as 1, we also compared the DilatedNet that contains regular residual blocks with DWAN. Furthermore, additional experiments were performed to assess the effects of residual learning and the MAE/MSE loss function used in training.

2.6. Evaluation metrics

We used Peak Signal to Noise Ratio (PSNR), Structural Similarity index (SSIM), Mean Absolute Error (MAE), Lin's Concordance Correlation Coefficient (CCC) [68] and Radiologic score as our evaluation metrics. PSNR, SSIM, MAE, and CCC were calculated from the denoised CBF maps and the surrogate GT to evaluate performance of different methods. PSNR assess the quality of image denoising quantitatively while SSIM measures the nonlocal structure similarity. MAE and CCC were used to assess the accuracy of predicted CBF values. The Bland-Altman plot and image profiles of CBF map were used to access the fidelity of predicted CBF values qualitatively. CBF image profiles were taken from the 50th row of the 54th axial slice. The Bland-Altman plot was drawn using the CBF values in the GM area of one representative subject's 50th axial slice. CBF image quality was qualitatively assessed by two radiologists (T.W. and F.Z.). The quality score ranges from 1 to 4 and the value 1/2/3/4 means severe noise/moderate noise that disturbs evaluation/mild noise that not affects evaluation/clear perfusion map, respectively. Randomly selected 10 subjects' CBF images processed by the 4 different methods: UNet, DilatedNet, DWAN, non-DL method were provided to the two radiologists independently without disclosing which method was used to process which image.

2.6.1. PSNR
PSNR is defined as:

$$PSNR = 20 * \log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right)$$

where MAX is the maximum possible pixel value of the image and MSE is the mean square error between the denoised image (the output of DL-ASL) and the surrogate GT (meanCBF-40_sm). Higher PSNR means less voxel-wise differences.

2.6.2.
The SSIM is defined as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$, and σ_{xy} are the mean, variance, and covariance of two patches (x, y) at the same location of the output image and the surrogate GT, respectively. c_1 and c_2 are constants used to avoid division by zero. Higher SSIM means more structure similarity. For the networks trained with meanCBF-40_nsm (the mean CBF map calculated from the non-smoothed ASL C/L images), PSNR and SSIM were calculated based on the difference between the network output and the meanCBF-40_sm as our goal of DL-ASL was to provide better quality in relative to the mean CBF map processed with current standard steps.

2.6.3.
The Lin's CCC is formulated as:

$$\rho_c = \frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}$$

where μ_x, μ_y, σ_x and σ_y are the mean and variance of the output image and the surrogate GT, respectively. σ_{xy} is the covariance between the output image and the surrogate GT. The CCC accesses the image fidelity to the surrogate GT, in which a value of 1 represents a perfect agreement and a value of 0 represents no agreement.

3. Results

3.1. Results of DL-ASL for projecting meanCBF-10 to meanCBF-40

Fig. 2 shows a representative subject's mean CBF maps produced by different denoising methods. As compared with meanCBF-10_nsm
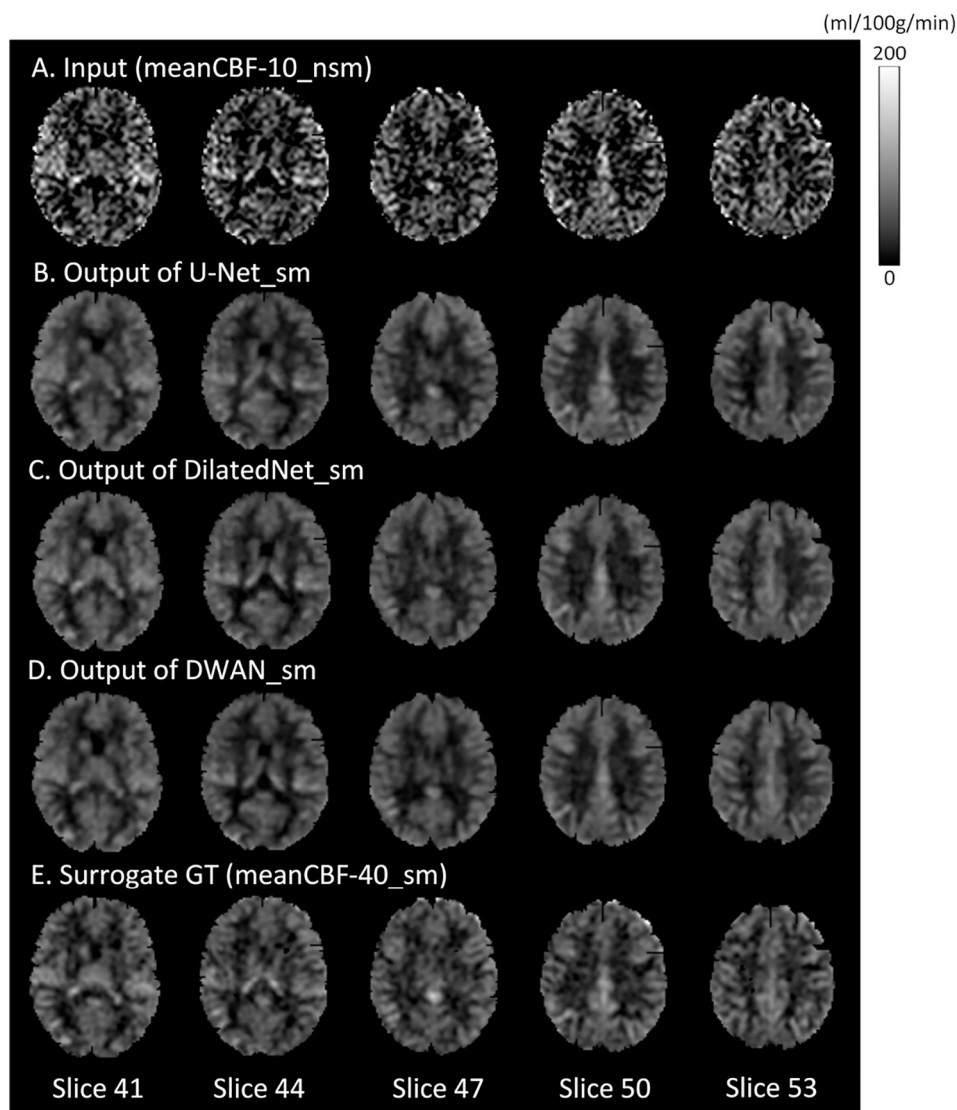
Fig. 2. Mean CBF images (only 5 axial slices were shown) from a representative subject with different processing methods. From top to bottom: meanCBF-10_nsm as input, the output of U-Net_sm, the output of DilatedNet_sm, the output of proposed DWAN_sm and the surrogate GT (meanCBF-40_sm) as training reference.
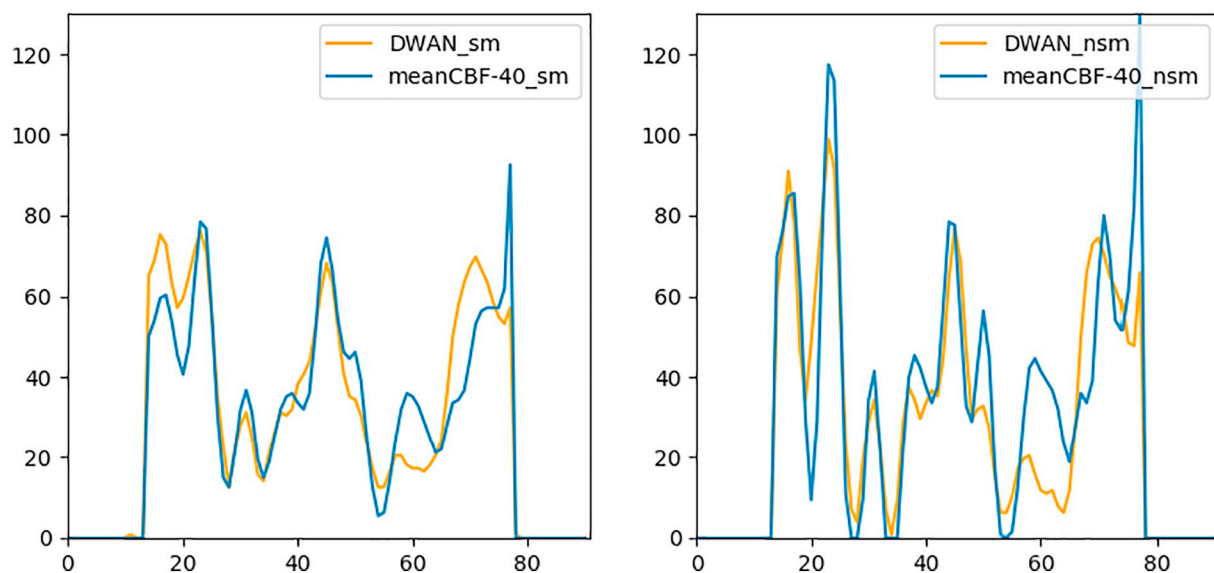

Fig. 3. CBF image profiles taken from the 50th row of the 54th axial slice from one representative subject.
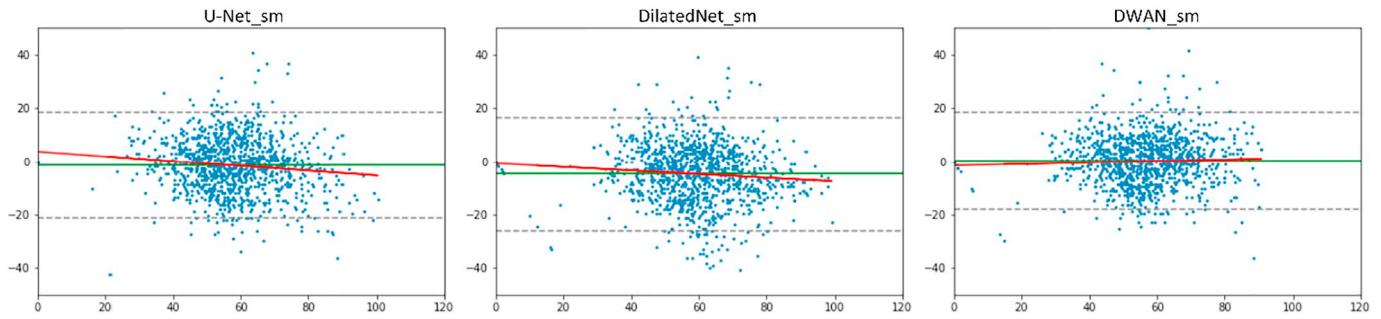
Fig. 4. Bland-Altman plots of different methods obtained in the GM area from one representative subject. Y-axis shows the differences between the surrogate GT and compared method in CBF values. X-axis shows the mean CBF values of the two. The unit for x and y axes are in ml/100 g/min. Solid green lines indicate mean difference. Dashed grey lines at top and bottom indicates upper and lower margins of 95% limits of agreement. Red solid lines are linear regression lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
The average PSNR and SSIM over 60 test subjects. All the models were trained with meanCBF-40_sm as reference images.

Model	U-Net_sm	U-Net_BN_sm	DilatedNet_sm	DilatedNet_BN_sm	DWAN_sm
PSNR	25.87	25.63	26.18	26.13	26.25
SSIM	0.813	0.806	0.819	0.817	0.822

Table 2
The average PSNR and SSIM over 60 test subjects. All the models were trained with meanCBF-40_nsm as reference images.

Model	U-Net_nsm	U-Net_BN_nsm	DilatedNet_nsm	DilatedNet_BN_nsm	DWAN_nsm
PSNR	24.53	24.16	24.78	24.60	24.81
SSIM	0.796	0.787	0.802	0.798	0.803

Table 3
The average MAE and CCC over 60 test subjects.

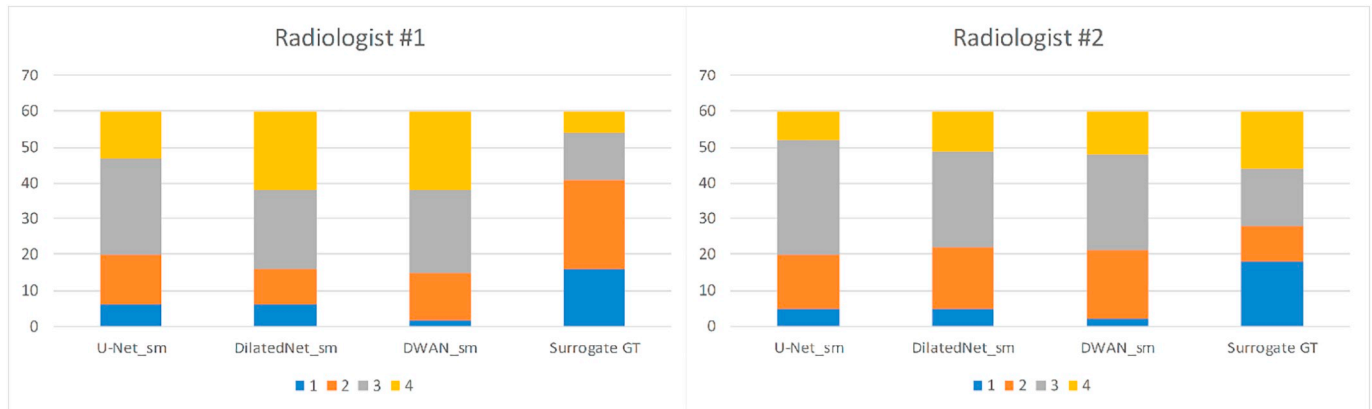|	U-Net_sm	DilatedNet_sm	DWAN_sm	U-Net_nsm	DilatedNet_nsm	DWAN_nsm
MAE	6.017	5.936	5.926	6.285	6.289	6.281
CCC	0.896	0.898	0.900	0.854	0.857	0.859

Fig. 5. Comparison of Radiologic score between different methods over 60 subjects' CBF maps. Radiologic scores are displayed with four different colors. Vertical axis indicates the number of subjects of each radiologic score.

(Fig. 2A, the input to DL-ASL), all DL methods (Fig. 2B–D) produced much better image quality in terms of improved tissue perfusion signal, suppressed noise (especially in white matter), better perfusion contrast between grey matter and white matter. The DL methods even showed CBF map quality improvement as compared with the surrogate GT (Fig. 2E, meanCBF-40_sm). DWAN_sm (Fig. 2D) showed the best quality as compared against other DL methods. Fig. 3 shows the image profile of CBF images processed with different methods. Fig. 4 shows the

Bland-Altman plots of CBF values in GM area obtained from different processing methods using one representative subject. The plots demonstrate that our proposed method achieved better accuracy in CBF values with smaller bias (green solid line) and variance (difference between dashed grey lines). The linear regression line (solid red) in U-Net_sm and DilatedNet_sm shows a systematic underestimation error while this error is reduced by DWAN_sm as its regression line is closer to the $y = 0$ line. Note that outlier voxels existed in all the methods due
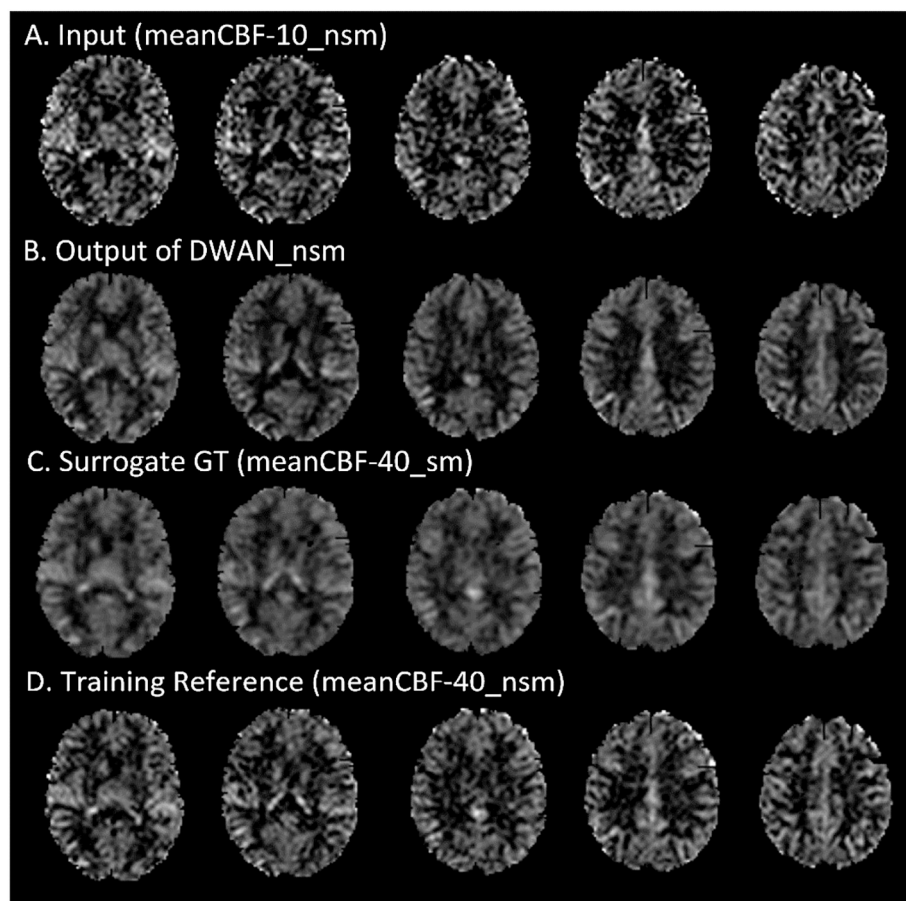
Fig. 6. ASL CBF images with meanCBF-10_nsm as input. From top to bottom: the input (meanCBF-10_nsm), the output of DWAN_nsm, the surrogate GT (meanCBF-40_sm) for comparison and the training reference (meanCBF-40_nsm).
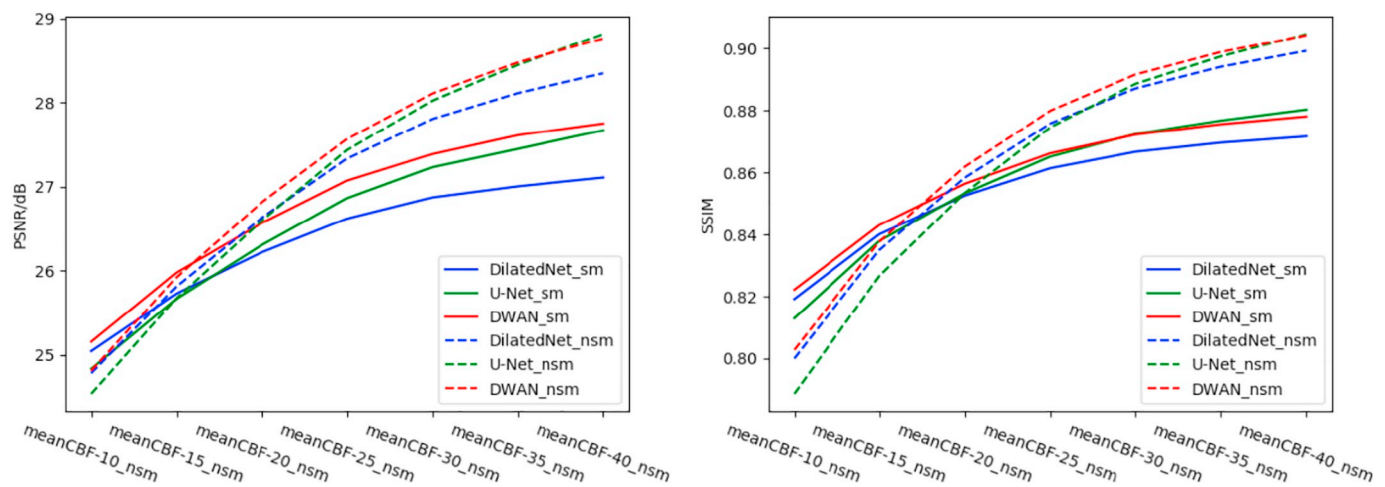

Fig. 7. PSNR and SSIM of U-Net_sm, U-Net_nsm, DilatedNet_sm, DilatedNet_nsm, DWAN_sm and DWAN_nsm over different mean CBF test datasets.

to excessive noises and artifacts from L/C subtractions.

Tables 1 and 2 list the performance indices of different methods. As compared with other methods, DWAN_sm and DWAN_nsm presented the better performance indices (PSNR and SSIM) ($p < .05$). Our experiments also confirmed that adding BN layers did not improve denoising performance as described in [53,58,59,60]. Table 3 list the MAE and CCC of different methods. DWAN_sm and DWAN_nsm has lower MAE and higher CCC than other methods.

Fig. 5 shows the radiologic scores of ASL CBF maps processed by different methods. DWAN_sm, DilatedNet_sm, and U-Net_sm had an

average score of 3.08, 3.00 and 2.78 by radiologist #1 (Tianyao Wang), and an average score of 2.81, 2.73 and 2.72 by radiologist #2 (Fuqing Zhou), respectively. The surrogate GT (meanCBF-40_sm) had an average score of 2.16 and 2.50 by the two radiologists. As Fig. 5 shows, DWAN_sm consistently achieves better denoising results than other methods qualitatively. Fig. 6 shows the corresponding results when using the meanCBF-40_nsm as the reference during model training. The outputs of DWAN_nsm are less blurring than those of DWAN_sm.
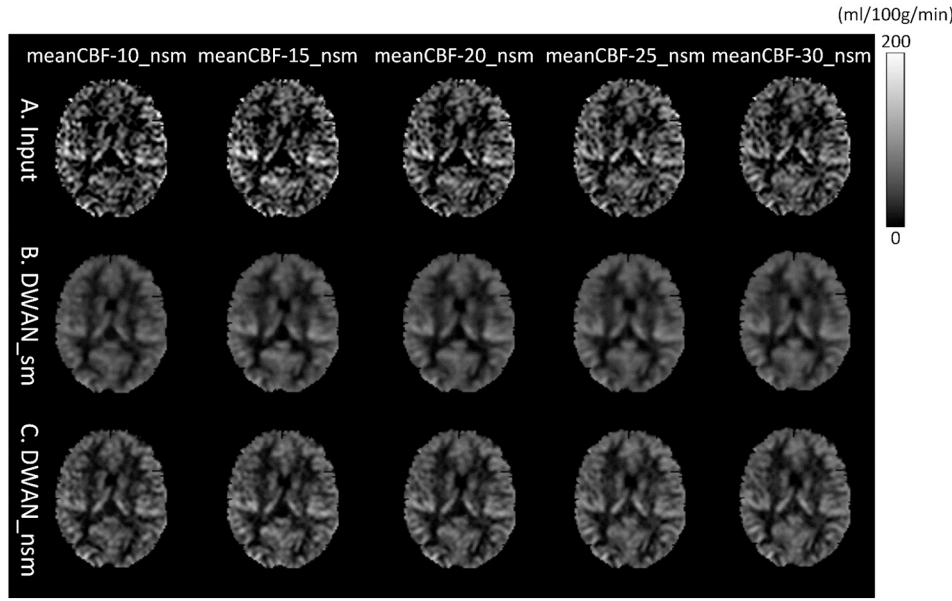
Fig. 8. The same slice of a representative subject's mean CBF images (only 1 axial slice was shown) processed using the same DWAN_sm or DWAN_nsm model trained with the (meanCBF-10_nsm, meanCBF-40_sm) or (meanCBF-10_nsm, meanCBF-40_nsm) image pairs, respectively. A) input images to the model, B) output images of DWAN_sm for the input shown in A, C) output images of DWAN_nsm for the input shown in A. The input to the model is meanCBF-10_nsm, meanCBF-15_nsm, meanCBF-20_nsm, meanCBF-25_nsm and meanCBF-30_nsm from the leftmost column to the rightmost column.
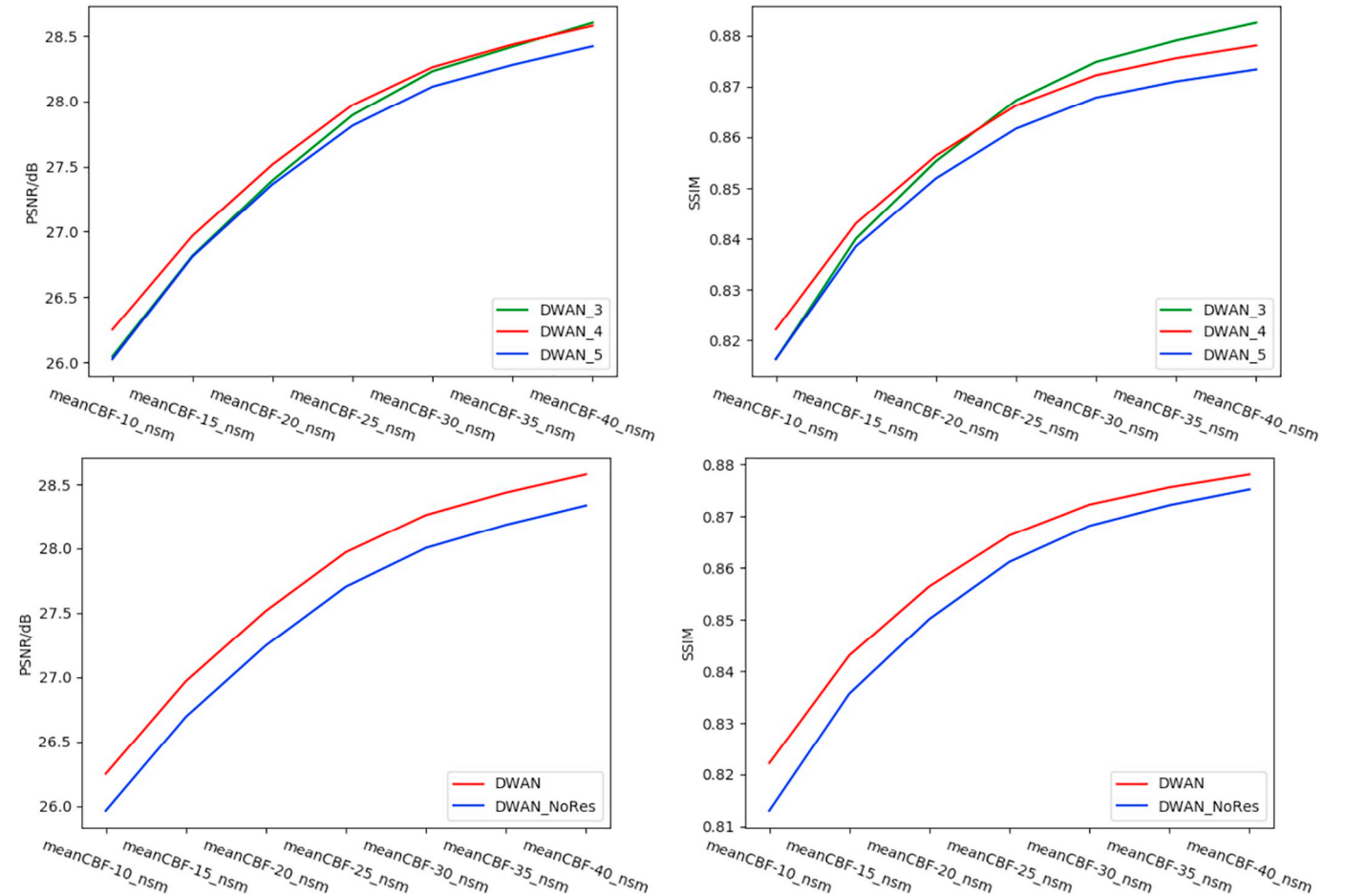
Fig. 9. Top: performance of DWAN with different number of wide activation residual blocks. The number in the suffix of the model represents the number of wide activation residual blocks in each pathway. Bottom: performance of DWAN with and without of residual learning. DWAN_NoRes has the same architecture as DWAN except residual learning.

3.2. Effects of different noise levels in input data on the model trained with meanCBF-10_nsm

Without any additional training, we directly applied the same model (trained by meanCBF-10_nsm) to different meanCBF-n_nsm (n = 10, 15, 20, 25, 30, 35, 40) datasets. Fig. 7 shows the quantitative results in terms of PSNR and SSIM. When input is meanCBF-10_nsm to meanCBF-30_nsm, DWAN_sm showed higher PSNR and SSIM than U-Net_sm and

Table 4
The PSNR of DWAN with different expansion rate *E* and different number of input filters *M*. *N/A* = Not Applicable due to the total number of filters E*M exceeding limitation of GPU memory and causing unstable training of DWAN.

M	E				
	1	2	4	6	8
16	26.00	25.83	26.02	25.81	26.11
32	26.07	26.12	26.26	26.12	N/A
64	26.14	26.15	N/A	N/A	N/A
128	25.63	N/A	N/A	N/A	N/A

DilatedNet_sm. When input is meanCBF-35_nsm or meanCBF-40_nsm, DWAN_nsm had higher PSNR and SSIM than all other models.

Fig. 8 shows the output of DWAN_sm and DWAN_nsm when the input was different meanCBF-n_nsm (n = 10, 15, 20, 25 and 30) from one representative subject. When the quality of input image increased, the quality of output image of DWAN_sm and DWAN_nsm increased. Besides, when input was meanCBF-25_nsm or meanCBF-30_nsm, DWAN_nsm produced better results with better grey matter/white matter contrast and image sharpness and less noise. Similar quantitative results were shown in the PSNR and SSIM plot in Fig. 7.

3.3. Effects of model configurations, residual learning, and loss function

Fig. 9 shows the performance of DWAN_sm with different number of wide activation residual blocks. In general, DWAN_sm with 4 wide activation residual blocks (DWAN_4) yielded the highest PSNR and SSIM on meanCBF-10_nsm to meanCBF-20_nsm, whereas DWAN_sm with 3 wide activation residual blocks (DWAN_3) yielded better SSIM for meanCBF-30_nsm to meanCBF-40_nsm. Since our goal was to train a model that has the best denoising performance with low SNR input, we chose DWAN_4 in our DL-ASL. Fig. 9 also shows that DWAN_sm with residual learning had better PSNR and SSIM performance than without residual learning. Table 4 shows the grid search results of finding the best combination of expansion rate *E* and the number of input filter *M* for all wide activation residual blocks. We found that E = 4 and M = 32 achieved the best denoising performance.

Fig. 10 shows the training loss and validation loss of DWAN_sm. Training loss was the MAE between output images and surrogate GT over the training dataset. Validation loss was the MAE between output images and surrogate GT over the validation dataset (the validation dataset was a small sample of data that was separated from training set, and was used to provide an unbiased evaluation of a model's performance during training). When the validation loss stopped decreasing

and the training loss was still decreasing, the model started to overfit the training dataset, which would make the model performance decrease on test datasets. To avoid the overfitting problem, the early stopping [69] technique was used during training.

4. Discussion

This paper proposed a new CNN architecture, DWAN, and validated it in ASL CBF denoising. Dilated convolutions were combined with wide activation residual blocks to form the basic units of DWAN. Applied to mean CBF maps calculated from different number of L/C image pairs, DWAN demonstrated better denoising performance than other CNNs and the non-DL based approach though all DL-based methods outperformed the traditional approaches. The performance was evidenced by image appearance, quantitative measures (PSNR, SSIM, MAE and CCC) and qualitative measure (radiologic scores). We also demonstrated that the same DWAN model can be used to denoise CBF images with different noise levels without retraining the model.

The decrease of MAE and increase of PSNR, SSIM and CCC of DL-ASL can be explained by two features of the CNN-based DL network. First, the spatially incoherent noise is suppressed by the many hierarchical convolutional filters in CNN. Each filter is learned to have different weights, resulting in suppressing different aspects of noise while retaining the cross-subject and cross-regional perfusion signal features. By using a hierarchical structure, the remaining noise can be further suppressed in the upper level through another series focal convolution processes. Second, DL-ASL is constructed to learn the cross-subject cross-regional perfusion features, making it robust to noise which usually doesn't have common spatial patterns across subjects and regions [70]. A common concern in denoising is the loss of resolution or equivalently the increase of blurring as typical noise contains a substantial high-frequency band like the random noise. However, DL has been repeatedly shown to be able to increase image resolution in many super-resolution papers [72,59,58]. Our group has recently demonstrated this capability of DL in ASL using data acquired with different resolutions [73,74]. We also showed that DL can achieve higher SNR while increase ASL CBF resolution. But the dataset used in this paper didn't have high resolution ASL CBF maps and gold standard CBF images, making it difficult to evaluate the potential blurring effects. Nevertheless, the concern of blurring still stands and will need future work to be fully addressed.

We chose the central slices as the training data because of the relatively better SNR they provided than other slices. For 2D ASL MRI without inferior saturation, the inferior slices in 2D ASL often suffer from large flow effects (fresh blood entered the imaging slices) and
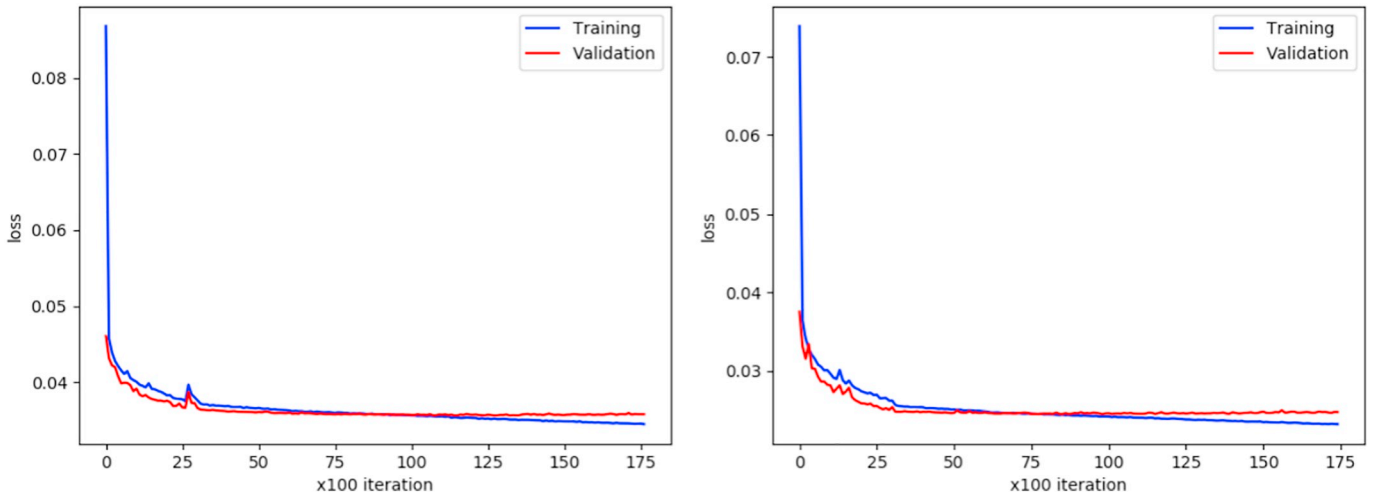

Fig. 10. Left, training and validation loss of DWAN_sm. Right, training and validation loss of DWAN_nsm.

should be avoided during training because current processing approaches do not have a good way to remove those artifacts, so they will be preserved in the reference image as well as in the model. The superior slices had the lowest SNR due to the long transit time. According to additional experiments, the superior slices can be included in training without affecting the denoising effects especially because of the "learning from noise" capability of DL [71].

Our evaluations converge to a major benefit: with the help of DWAN, we can shorten the acquisition time by 4 times to be 1 min 20 s without losing much CBF image quality as compared with the full-length acquisition, which was 40C/L pairs in this paper. Shortening acquisition time is highly valuable when subjects or patients have difficulty in keeping still in the scanner.

It is worth to note that DL-ASL was only trained with data from healthy subjects. Both signal and noise might have different distributions in diseased populations, which may cause a less efficient denoising for patients' data as it may not be able to suppress the extra noise in diseased populations. While that problem is universal to non-deep learning and deep learning-based methods, it might be less severe in DL as the model is trained to learn the signal or noise across many different people and distributed brain regions which themselves present large variability. On the contrary, traditional methods often rely on a pre-determined model which doesn't change based on the data and then is less adaptive to the data. To have a model full capable of suppressing noise or artifacts for patients, the model must be trained with additional data from patients.

5. Conclusion

In conclusion, we introduced DL into ASL MRI denoising and proposed a new denoising model DWAN. DL-ASL substantially improved the quality of ASL CBF, suggesting a potential of significantly reducing acquisition time. Besides, DL-ASL showed the generalizability to denoising CBFs with different noise levels without model retraining.

CRediT authorship contribution statement

Danfeng Xie: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft.**Yiran Li:**Validation, Formal analysis.**Hanlu Yang:**Writing - review & editing.**Li Bai:**Writing - review & editing, Supervision.**Tianyao Wang:**Validation.**Fuqing Zhou:**Validation.**Lei Zhang:**Writing - review & editing.**Ze Wang:**Writing - review & editing, Supervision, Project administration, Funding acquisition, Resources.

Acknowledgements

References

[1] Detre JA, Leigh JS, Williams DS, Koretsky AP. Perfusion imaging. Magn Reson Med 1992;23:37–45.

[2] Williams DS, Detre JA, Leigh JS, Koretsky AP. Magnetic resonance imaging of perfusion using spin inversion of arterial water. Proc Natl Acad Sci 1992;89:212–6.

[3] Alsop DC, Detre JA, Golay X, Günther M, Hendrikse J, Hernandez-Garcia L, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn Reson Med 2015;73:102–16.

[4] Wong EC. Potential and pitfalls of arterial spin labeling based perfusion imaging techniques for MRI. Functional MRI. Heidelberg: Springer; 1999. p. 63–9.

[5] Wang Z. Improving cerebral blood flow quantification for arterial spin labeled perfusion MRI by removing residual motion artifacts and global signal fluctuations. Magn Reson Imaging 2012;30:1409–15.

[6] Behzadi Y, Restom K, Liau J, Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 2007;37:90–101.

[7] Wang J, Wang Z, Aguirre GK, Detre JA. To smooth or not to smooth? ROC analysis of perfusion fMRI data. Magn Reson Imaging 2005;23:75–81.

[8] Bibic A, Knutsson L, Ståhlberg F, Wirestam R. Denoising of arterial spin labeling data: wavelet-domain filtering compared with Gaussian smoothing. Magn Reson Mater Phys, Biol Med 2010;23:125–37.

[9] Wells JA, Thomas DL, King MD, Connelly A, Lythgoe MF, Calamante F. Reduction of errors in ASL cerebral perfusion and arterial transit time maps using image denoising. Magn Reson Med 2010;64:715–24.

[10] Liang X, Connelly A, Calamante F. Voxel-wise functional connectomics using arterial spin labeling functional magnetic resonance imaging: the role of denoising. Brain Connect 2015;5:543–53.

[11] Zhu H, Zhang J, Wang Z. Arterial spin labeling perfusion MRI signal denoising using robust principal component analysis. J Neurosci Methods 2018;295:10–9.

[12] Spann SM, Kazimierski KS, Aigner CS, Kraiger M, Bredies K, Stollberger R. Spatio-temporal TGV denoising for ASL perfusion imaging. Neuroimage 2017;157:81–96.

[13] Wang Z. Support vector machine learning-based cerebral blood flow quantification for arterial spin labeling MRI. Hum Brain Mapp 2014;35:2869–75.

[14] Zhu H, He G, Wang Z. Patch-based local learning method for cerebral blood flow quantification with arterial spin-labeling MRI. Med Biol Eng Comput 2017;1–6.

[15] Wang Z, Das SR, Xie SX, Arnold SE, Detre JA, Wolk DA, et al. Arterial spin labeled MRI in prodromal Alzheimer's disease: a multi-site study. NeuroImage 2013;2:630–6.

[16] Dolui S, Wang Z, Shinohara RT, Wolk DA, Detre JA, Initiative ADN. Structural correlation-based outlier rejection (SCORE) algorithm for arterial spin labeling time series. J Magn Reson Imaging 2017;45:1786–97.

[17] Wang Z. Priors-guided adaptive outlier cleaning for arterial spin labeling perfusion MRI. ISMRM 2016. 2016.

[18] Li Y, Dolui S, Xie D-F, Wang Z, Initiative ADN, et al. Priors-guided slice-wise adaptive outlier cleaning for arterial spin labeling perfusion MRI. J Neurosci Methods 2018;307:248–53.

[19] Deng L, Yu D, et al. Deep learning: methods and applications. Foundations and Trends® in Signal Processing 2014;7:197–387.

[20] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. science 2006;313:504–7.

[21] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Proces Syst 2012:1097–105.

[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition arXiv preprint arXiv:1409.1556 2014.

[23] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition 2014:1725–32.

[24] Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[25] Graves A, Mohamed A-r, Hinton G. Speech recognition with deep recurrent neural networks. Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. 2013.

[26] Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, et al. Deep speech 2: end-to-end speech recognition in English and Mandarin. International conference on machine learning. 2016.

[27] Bahdanau D, Chorowski J, Serdyuk D, Brakel P, Bengio Y. End-to-end attention-based large vocabulary speech recognition. Acoustics, speech and signal processing (ICASSP), 2016 IEEE international conference on. 2016.

[28] Mohamed A, Dahl GE, Hinton G. Acoustic modeling using deep belief networks. IEEE Trans Audio Speech Lang Process 2012;20:14–22.

[29] Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221–48.

[30] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

[31] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. nature 2016;529:484–9.

[32] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, et al. Mastering chess and Shogi by self-play with a general reinforcement learning algorithm arXiv preprint arXiv:1712.01815 2017.

[33] Fukushima K. Neocognitron: a hierarchical neural network capable of visual pattern recognition. Neural Netw 1988;1:119–30.

[34] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 2006;18:1527–54.

[35] LeCun Y, Bengio Y, Hinton G. Deep learning. nature 2015;521:436.

[36] Xie D, Zhang L, Bai L. Deep learning in visual computing and signal processing. Applied Computational Intelligence and Soft Computing 2017;2017.

[37] Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. Adv Neural Inf Proces Syst 2007:153–60.

[38] Xie J, Xu L, Chen E. Image denoising and inpainting with deep neural networks. Advances in neural information processing systems. 2012.

[39] Burger HC, Schuler CJ, Harmeling S. Image denoising: Can plain neural networks compete with BM3D? Computer vision and pattern recognition (CVPR), 2012 IEEE conference on. 2012.

[40] Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. European conference on computer vision. 2014.

[41] Mao X-J, Shen C, Yang Y-B. Image restoration using convolutional auto-encoders with symmetric skip connections arXiv preprint arXiv:1606.08921 2016.

[42] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans Image Process 2017;26(7):3142–55.

[43] Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, et al. Accelerating magnetic resonance imaging via deep learning. 2016 IEEE 13th international symposium on biomedical imaging (ISBI). 2016.

[44] Han Y, Yoo J, Kim HH, Shin HJ, Sung K, Ye JC. Deep learning with domain adaptation for accelerated projection-reconstruction MR. Magn Reson Med 2018;80:1189–205.

[45] Xu J, Gong E, Pauly J, Zaharchuk G. 200x low-dose pet reconstruction using deep learning *arXiv preprint arXiv:1712.04119* 2017.

[46] Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging 2018;48:330–40.

[47] Kim KH, Choi SH, Park S-H. Improving arterial spin labeling by using deep learning. Radiology 2017:171154.

[48] Ulas C, Tetteh G, Kaczmarz S, Preibisch C, Menze BH. DeepASL: Kinetic model incorporated loss for Denoising arterial spin labeled MRI via deep residual learning. International conference on medical image computing and computer-assisted intervention. 2018.

[49] Gong E, Pauly J, Zaharchuk G. Boosting SNR and/or resolution of arterial spin label (ASL) imaging using multi-contrast approaches with multi-lateral guided filter and deep networks. *25th Annual Meeting of the International Society for Magnetic Resonance in Medicine*, Honolulu. 2017.

[50] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. Med Image Anal 2017;35:18–31.

[51] Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61–78.

[52] Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. Magn Reson Med 1998;40(3):383–96.

[53] Yu J, Fan Y, Yang J, Xu N, Wang Z, Wang X, et al. Wide activation for efficient and accurate image super-resolution *arXiv preprint arXiv:1808.08718* 2018.

[54] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions *arXiv preprint arXiv:1511.07122* 2015.

[55] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[56] Fan Y, Yu J, Huang TS. Wide-activated deep residual networks based restoration for BPG-compressed images. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops. 2018.

[57] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning. 2015.

[58] Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced deep residual networks for single image super-resolution. Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.

[59] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. Proc IEEE Conf Comput Vis Pattern Recognit 2018:2472–81.

[60] Fan Y, Shi H, Yu J, Liu D, Han W, Yu H, et al. Balanced two-stage residual networks for image super-resolution. Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.

[61] Dai W, Garcia D, Bazelaire CD, Alsop DC. Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields. Magn Reson Med 2008;60(6):1488–97.

[62] Wu W, Fernández-Seara M, Detre JA, Wehrli FW, Wang J. A theoretical and experimental investigation of the tagging efficiency of pseudocontinuous arterial spin labeling. Magn Reson Med 2007;58(5):1020–7.

[63] Wang Z, Aguirre GK, Rao H, Wang J, Fernández-Seara MA, Childress AR, et al. Empirical optimization of ASL data analysis using an ASL data processing toolbox: ASLtbx. Magn Reson Imaging 2008;26:261–9.

[64] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention. 2015. p. 234–41.

[65] Chollet F, et al. Keras. 2015.

[66] Kingma D, Ba J. Adam: A method for stochastic optimization *arXiv preprint arXiv:1412.6980* 2014.

[67] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous systems. 2015.

[68] Lawrence I-KL. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989:255–68.

[69] Prechelt L. Early stopping-but when? Neural networks: Tricks of the trade. Springer; 1998. p. 55–69.

[70] Ulyanov D, Vedaldi A, Lempitsky aV. Deep image prior. In proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[71] Xie D, Li Y, Yang H, Song D, Shang Y, Ge Q, et al. BOLD fMRI-based brain perfusion prediction using deep dilated wide activation networks. *In International Workshop on Machine Learning in Medical Imaging*, Shenzhen. 2019.

[72] Ouyang W, Aristov A, Lelek M, Hao X, Zimmer C. Deep learning massively accelerates super-resolution localization microscopy. Nat Biotechnol 2018;36(5):460.

[73] Liu Q, Shi J, Wang Z. Increasing arterial spin labeling perfusion image resolution using convolutional neural networks with residual-learning. *Proc of ISMRM*, Paris, France. 2018.

[74] Li Z, Liu Q, Li Y, Ge Q, Shang Y, Song D, et al. A two-stage multi-loss super-resolution network for arterial spin labeling magnetic resonance imaging. *Medical Image Computing and Computer Assisted Intervention*, Shenzhen, China. 2019.