

MACHINE LEARNING-BASED ARTERIAL SPIN LABELING  
PERFUSION MRI SIGNAL PROCESSING

---

A Dissertation  
Submitted to the  
Temple University Graduate Board

---

In Partial Fulfillment  
of the Requirements for the Degree  
DOCTOR OF PHILOSOPHY

---

by  
Danfeng Xie  
May 2020

Examining Committee Members:

Dr. Li Bai, Advisory Chair, Department of Electrical and Computer Engineering

Dr. Ze Wang, Co-advisor, Department of Diagnostic Radiology and Nuclear  
Medicine, University of Maryland School of Medicine

Dr. Albert Kim, Department of Electrical and Computer Engineering

Dr. Xiaonan Lu, Department of Electrical and Computer Engineering

Dr. Bo Ji, External Reader, Department of Computer and Information Sciences

## ABSTRACT

Arterial spin labeling (ASL) perfusion Magnetic Resonance Imaging (MRI) is a noninvasive technique for measuring quantitative cerebral blood flow (CBF) but subject to an inherently low signal-to-noise-ratio (SNR), resulting in a big challenge for data processing. Traditional post-processing methods have been proposed to reduce artifacts, suppress non-local noise, and remove outliers. However, these methods are based on either implicit or explicit models of the data, which may not be accurate and may change across subjects. Deep learning (DL) is an emerging machine learning technique that can learn a transform function from acquired data without using any explicit hypothesis about that function. Such flexibility may be particularly beneficial for ASL denoising. In this dissertation, three different machine learning-based methods are proposed to improve the image quality of ASL MRI: 1) a learning-from-noise method, which does not require noise-free references for DL training, was proposed for DL-based ASL denoising and BOLD-to-ASL prediction; 2) a novel deep learning neural network that combines dilated convolution and wide activation residual blocks was proposed to improve the image quality of ASL CBF while reducing ASL acquisition time; 3) a prior-guided and slice-wise adaptive outlier cleaning algorithm was developed for ASL MRI.

In the first part of this dissertation, a learning-from-noise method is proposed for DL-based method for ASL denoising. The proposed learning-from-noise method shows that DL-based ASL denoising models can be trained using only noisy image pairs, without any deliberate post-processing for obtaining the quasi-noise-free reference during the training process. This learning-from-noise method can also be applied to DL-based ASL perfusion

prediction from BOLD fMRI as ASL references are extremely noisy in this BOLD-to-ASL prediction. Experimental results demonstrate that this learning-from-noise method can reliably denoise ASL MRI and predict ASL perfusion from BOLD fMRI, result in improved signal-to-noise-ratio (SNR) of ASL MRI. Moreover, by using this method, more training data can be generated, as it requires fewer samples to generate quasi-noise-free references, which is particularly useful when ASL CBF data are limited.

In the second part of this dissertation, we propose a novel deep learning neural network, i.e., Dilated Wide Activation Network (DWAN), that is optimized for ASL denoising. Our method presents two novelties: first, we incorporated the wide activation residual blocks with a dilated convolution neural network to achieve improved denoising performance in term of several quantitative and qualitative measurements; second, we evaluated our proposed model given different inputs and references to show that our denoising model can be generalized to input with different levels of SNR and yields images with better quality than other methods.

In the final part of this dissertation, a prior-guided and slice-wise adaptive outlier cleaning (PAOCSL) method is proposed to improve the original Adaptive Outlier Cleaning (AOC) method. Prior information guided reference CBF maps are used to avoid bias from extreme outliers in the early iterations of outlier cleaning, ensuring correct identification of the true outliers. Slice-wise outlier rejection is adapted to reserve slices with CBF values in the reasonable range even they are within the outlier volumes. Experimental results show that the proposed outlier cleaning method improves both CBF quantification quality and CBF measurement stability.

To my parents,  
Shouhong and Chunmei,  
and my girlfriend,  
Noheli

## ACKNOWLEDGEMENTS

It has been a great honor for me to spend several years in the Department of Electrical and Computer Engineering at Temple University. Its members will always remain dear to me. First of all, I would like to express my sincere gratitude to my advisor, Dr. Li Bai, who not only picked me at the begging of my academic career but also guide me and support me through my Ph.D. career. I have greatly benefited from his wisdom and support over these past years. Besides, I would like to express my sincere gratitude to my co-advisor Dr. Ze Wang, who greatly influenced my academic career and leads me into the field of functional magnetic resonance imaging. His research guidance and support are invaluable to me.

Special thanks to my Ph.D. committee members: Dr. Albert Kim, Dr. Xiaonan Lu and Dr. Bo Ji. They give me considerate suggestions to my final dissertation. I would also express my thanks to all my colleagues and lab mates: Dr. Lei Zhang, Dr. Qiangguo Ren, Dr. Ning Gong, Yiran Li, Hanlu Yang, Fang Yang and Dr. Yunfeng Hong who inspire me during my whole Ph.D. career.

I want to thank Prof. Li Peng, Zhe Wang, Heguang Liu, Bingwei Zeng from Jiangnan University and Dr. Shizeng Wang and Dr. Jianwei Yang from Chinese Academy of Sciences, who lead me to the gate of research and inspire me to pursue Ph.D. career. Their friendship and assistance meant a lot to me.

Finally, my deepest appreciation is dedicated to my parents Shouhong and Chunmei, and my girlfriend Noheli. Their love, support and encouragement have always been important to me.

I would like to acknowledge the support offered by NIH/NIA grant: 1 R01 AG060054-01A1 and the Temple University Dissertation Completion grant.

# TABLE OF CONTENT

	Page
ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	v
LIST OF FIGURES .....	x
LIST OF TABLES .....	xiv
CHAPTER	
1 INTRODUCTION .....	1
1.1 Major Challenges .....	5
1.2 Research Objectives .....	6
1.2.1 A Learning-From-Noise Method for ASL Denoising And Prediction .....	6
1.2.2 Dilated Wide Activation Network-based ASL Denoising.....	7
1.2.3 Prior-Guided and Slice-Wise Adaptive Outlier Cleaning For ASL...	7
1.3 Contributions.....	8
1.4 Organization of this Dissertation .....	8
2 BACKGROUND AND RELATED WORKS .....	10
2.1 Functional MRI.....	10
2.2 Deep Learning.....	13
2.3 Related Works.....	18
2.3.1 ASL Denoising.....	18
2.3.2 Outlier Cleaning.....	20
3 A LEARNING-FROM-NOISE DEEP LEARNING METHOD FOR ARTERIAL SPIN LABELING MRI DENOISING AND PREDICTION .....	21
3.1 Introduction.....	22
3.2 Methods.....	23

3.3 Application on ASL denoising.....	27
3.3.1 Materials and Experiment Setup.....	28
3.3.2 Experimental Results .....	32
3.4 Application on ASL perfusion prediction from BOLD fMRI .....	37
3.4.1 Materials and Experiment Setup.....	39
3.4.2 Experimental Results .....	42
3.5 Discussion and Conclusion .....	45
<b>4 DENOISING ARTERIAL SPIN LABELING PERFUSION MRI WITH DILATED WIDE ACTIVATION NETWORK .....</b>	<b>47</b>
4.1 Problem Formulation .....	47
4.2 Materials and Methods.....	48
4.2.1 Network Architecture.....	48
4.2.2 Image Acquisition and Preprocessing.....	51
4.2.3 Data Preparation and Model Training.....	53
4.2.4 Effects of Different Model Configurations.....	54
4.2.5 Evaluation Metrics .....	54
4.3. Experimental Results .....	56
4.3.1 Results of ASLDN For Projecting meanCBF-10 to meanCBF-40..	56
4.3.2 Effects of Different Noise Levels on the Model Trained with Meancbf-10_nsm .....	61
4.3.3 Effects of Model Configurations, Residual Learning, and Loss Function .....	63
4.4. Discussion and Conclusion .....	66
<b>5 PRIOR-GUIDED SLICE-WISE ADAPTIVE OUTLIER CLEANING FOR ARTERIAL SPIN LABELING PERFUSION MRI.....</b>	<b>69</b>
5.1 Introduction.....	69
5.2 Materials and Methods.....	70
5.2.1. Proposed Method .....	70
5.1.2 Image Acquisition.....	73
5.2 Experimental Results .....	74
5.3 Discussion and Conclusion .....	78

6 CONCLUSION AND FUTURE WORK ..... 81

    6.1 Conclusion ..... 81

    6.2 Future Work ..... 82

BIBLIOGRAPHY ..... 84

## LIST OF FIGURES

Figure 1.1: A demonstration of (A) a single MRI scan; (B) a single BOLD fMRI scan; and (C) an CBF map generated from ASL MRI scan. They are obtained at the same axial slice from one representative subject. An MRI scan only shows anatomical structure information while BOLD fMRI and ASL CBF show the metabolic activity within the anatomical structure. MRI has higher spatial resolution but lower temporal resolution than BOLD fMRI and ASL MRI. ASL CBF offers quantitative measurement, in which the intensity of a voxel is in a physical unit of ml/100g/min. BOLD fMRI has higher spatial and temporal resolution than ASL MRI but is a relative measurement.....	3
Figure 2.1: Illustration of hemodynamic process. When neurons are in resting states, the relative level of oxyhemoglobin is low in venules. When neurons are in a stimulated state, the vessel size, blood flow and the relative level of oxyhemoglobin in venules increases. HbO <sub>2</sub> = oxyhemoglobin, dHb = deoxyhemoglobin.....	11
Figure 2.2: An example of the ASL acquisition process. In an ASL MRI scan, multiple pairs of label and control images were acquired. Perfusion-weighted image is obtained by pair-wise subtraction and is subsequently quantified to CBF map in a unit of ml/100 g/min. By averaging a series of CBF maps, moderate noises were removed in the mean CBF map. ....	13
Figure 2.3: An example of a convolution operation. ....	15
Figure 2.4: The example of the pooling layer.....	17
Figure 3.1: An illustration of DL-based ASL denoising (ASLDN) and ASLDN-LFN. ASLDN using quasi-noise-free CBF images as training references whereas ASLDN-LFN only use noisy image pairs to train the DL model. ....	27
Figure 3.2: Schematic illustration of the architecture of our proposed DWAN network. The first layer consists of 3×3×32 convolutional filters for the input image. Then the output of the first layer was fed to the both local pathway and global pathway. Each pathway contains 4 consecutive wide activation residual blocks. Each wide activation residual block contains two convolutional layers (3×3×128 and 3×3×32) and one activation function layer. The 3×3×128 convolutional layers in the global pathway were dilated convolutional layers with a dilation rates of 2, 4, 8, 16, respectively. The output of the local pathway and global pathway were concatenated and fed to another 3×3×1 convolutional filter. The 3×3×1 convolutional layer was attached to the end to get the predicted output image with additional input from the input image with 3×3×1 convolution. (a×b×c indicates the property of convolution. a×b is the kernel size of one filter and c is the number of the filters). ....	30
Figure 3.3: Mean CBF images of a representative subject. The rows from top to bottom are: A. mean CBF maps generated from 10 L/C pairs (input to ASLDN-LFN); B. mean CBF maps from all 40 L/C pairs with smoothing and outlier cleaning (pseudo-groundtruth); C.	

output of ASLDN; and D. output of ASLDN-LFN. Only 5 axial slices were shown in each row. ....	33
Figure 3.4: The notched box plot of the SNR (left) and GM/WM contrast (right) from 60 test subjects' CBF maps with different processing methods.....	34
Figure 3.5: Correlation coefficient maps of ASLDN (top) and ASLDN-LFN (bottom). Only 5 axial slices were shown. Correlation coefficients less than 0.3 were thresholded to be 0.....	35
Figure 3.6: Mean CBF maps of a representative subject (Only 3 axial slices were shown). From left to right: input to ASLDN-LFN (column A); pseudo gold standard (column B); outputs of ASLDN-LFN trained with L2 loss (column C); and outputs of ASLDN-LFN trained with L1 loss (column D) .....	36
Figure 3.7: Illustration of dual-echo sequence and BOA-Net training process.....	40
Figure 3.8: From top to bottom: BOLD fMRI and ASL CBF produced by different methods. From left to right: slice 7, 8, 9, 10, 11. ....	43
Figure 3.9: The notched box plot of the SNR (left) and correlation coefficient maps between genuine mean CBF and output of BOA-Net (right). Original_nsm and original_sm represent the genuine mean CBF maps from non-smoothed and smoothed ASL data. BOA-Net_nsm and BOA-Net_sm represent mean CBF maps from outputs of BOA-Net_nsm and BOA-Net_sm. The correlation coefficient maps between genuine mean CBF and output of BOA-Net_sm is shown in the top row. The correlation coefficient maps between genuine mean CBF and output of BOA-Net_nsm is shown in the bottom row. Only 2 axial slices were shown. Correlation coefficients less than 0.3 were thresholded to be 0. ....	43
Figure 3.10: Two representative slices of the mean CBF maps produced by different processing methods. The three columns on the left side are mean CBF maps of outputs of U-Net_nsm, DilatedNet_nsm, and DWAN_nsm respectively. The three columns on the right side are mean CBF maps of outputs of U-Net_sm, DilatedNet_sm, and DWAN_sm. ....	45
Figure 4.1: Schematic illustration of a regular residual block (left) and a wide activation residual block (right).....	50
Figure 4.2: Schematic illustration of a series of dilated convolutions. A series of dilation convolution supports exponential expansion of the receptive field without loss of resolution or coverage. On the left: a 3×3 convolution filter without dilation produce a receptive field of 3×3 as each element (blue dot) is a 1×1 convolution. On the middle: a 2-dilated 3×3 convolution filter in the following convolution layer produce a receptive field of 7×7 as each element (blue dot) is a result of a 3×3 convolution from the previous layer. On the right, a 4-dilated 3×3 convolution filter in the next convolution layer produce a receptive field of 15×15 as each element (blue dot) is a result of a 7×7 convolution from the previous layer. The size of convolution filter (3×3) in each layer is identical but the receptive field grows exponentially. ....	50

Figure 4.3: 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 the first layer was fed to both local pathway and global pathway. Each pathway contains 4 consecutive wide activation residual blocks. Each wide activation residual block contains 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 the global pathway were dilated convolutional layers with a dilation rate of 2, 4, 8, 16, respectively. The output of the 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).

..... 52

Figure 4.4: 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. Image display window was 0-200 ml/100g/min..... 57

Figure 4.5: CBF image profiles are taken from the 50th row (marked as red line) of the 54th axial slice from one representative subject. .... 58

Figure 4.6: Bland-Altman plots of different methods obtained in the GM area from one representative subject. Y-axis shows the differences in CBF values between the surrogate GT and the compared method and X-axis shows the mean CBF values of the two. The unit for x and y axes are in ml/100g/min. Solid grey lines indicate mean difference. Dashed red lines at top and bottom correspond to upper and lower margins of 95% limits of agreement. blue solid lines are linear regression lines. .... 58

Figure 4.7: Comparison of Radiologic score between different methods over 60 subjects' CBF maps. Radiologic scores are displayed in four different colors. The vertical axis indicates the number of subjects of each radiologic score..... 60

Figure 4.8: 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). The image display window was 0-200 ml/100g/min. .... 61

Figure 4.9: 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..... 62

Figure 4.10: 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. .... 63

Figure 4.11: Top: performance of DWAN with a 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 residual learning. DWAN\_NoRes has the same architecture as DWAN except residual learning..... 64

Figure 4.12: Left, training and validation loss of DWAN\_sm. Right, training and validation loss of DWAN\_nsm..... 66

Figure 5.1: Flowchart of the prior-guided AOC algorithm. pmCBF means pseudo mean CBF map; SL means the maximum number of slices; slnei means the neighboring slices around the current slice sl;  $\beta$  is an empirically defined constant; CC means correlation coefficient; corr means correlation; WM/GM means white matter/grey matter; std means standard deviation. .... 72

Figure 5.2: ASL CBF maps of a representative subject processed with different methods: A) NAOC, B) SCORE, C) PAOC, D) PAOCSL. Image display window was 0-60 ml/100g/min. Green ovals mark the places with improved CBF image contrasts in PAOCSL. .... 74

Figure 5.3: Boxplots of SNR of different outlier cleaning approaches. Red bar indicates median of the SNR of all subjects (n=19). PAOCSL showed significantly higher SNR than other methods ( $p < 0.005$  for all possible two-sample t-tests). .... 75

Figure 5.4: Number of outlier slices removed at different timepoint for one representative subject from the ADNI database..... 76

Figure 5.5: ICC maps of CBF maps obtained with different methods: A) no AOC, B) SCORE, C) PAOC, D) PAOCSL. The threshold was 0.5 and the colorbar indicates the display window of ICC. Red oval and red box indicate places with higher CBF measurement test-retest stability after applying PAOCSL. .... 77

Figure 5.6: 3D ASL CBF maps of a young healthy subject processed with: A) NAOC, B) SCORE, C) PAOC, D) PAOCSL. Image display window was 0-80 ml/100g/min. Green ovals mark the places with improved CBF image contrasts in PAOCSL in terms of better image homogeneity and less signal loss in inferior frontal regions. .... 78

## LIST OF TABLES

Table 3.1: The average PSNR and SSIM of mean CBF maps produced by different CNN architectures in different training schemes. ....	35
Table 3.2: The average PSNR and SSIM from different CNN architectures used in BOA-Net_sm and BOA-Net_nsm .....	44
Table 4.1: The average PSNR and SSIM over 60 test subjects. All the models were trained with meanCBF-40_sm as reference images.....	59
Table 4.2: The average PSNR and SSIM over 60 test subjects. All the models were trained with meanCBF-40_nsm as reference images.....	59
Table 4.3: The average MAE and CCC over 60 test subjects. ....	60
Table 4.4: 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 the limitation of GPU memory and causing unstable training of DWAN.....	65

# CHAPTER 1

## INTRODUCTION

Analysis of human brain activity has drawn enormous attention from both clinical and scientific research. Many technologies have been invented to analyze brain activities, including electroencephalogram (EEG), positron emission tomography (PET), and functional Magnetic Resonance Imaging (fMRI). With the help of these technologies, brain activity can be detected and analyzed based on different signal processing techniques: EEG monitors brain activity by measuring voltage changes from ionic current produced by neurons inside the brain. EEG has the highest temporal resolution, but the lowest spatial resolution compared with other methods. It is difficult for EEG to locate the signal source as it can only measure electrical activity at the surface of the brain, while not able to locate the signal inside the brain. Positron-emission tomography (PET) measures brain activity by detecting high radioactivity areas of the brain using a radioactive tracer. PET has higher spatial resolution than EEG, but it requires a radioactive tracer to measure CBF, which can cause safety concerns to healthy subjects. Functional Magnetic Resonance Imaging (fMRI) detects brain activity by measuring cerebral blood flow (CBF) or metabolism using strong magnetic field and radiofrequency pulses. FMRI is safe, non-invasive, and offers the highest spatial resolution compared to other technologies. Since the invention of fMRI, it has become one of the most popular tools in human brain mapping research [1] [2]. Table 1.1 shows a summary of the spatial and temporal resolution of each technology.

Table 1.1. the summary of the spatial and temporal resolution of EEG, PET and fMRI.

Technology	EEG	PET	fMRI
Temporal Resolution	10 <sup>-3</sup> to 1 sec	0.37 to 5 sec	1 to 3 sec
Spatial Resolution	1-5 cm	0.5 to 7.5 cm	3 to 5 mm

FMRI is simply MRI scanning that reflects tissue function rather than tissue structure. The signal induced by tissue function can be attributed to metabolism or blood flow. Typically, such a signal is only a few percent of overall signal intensity. Acquiring such a signal reliably in high resolution and a short amount of time is the major goal of current fMRI techniques. Currently, there are two type of fMRI techniques: (1) Blood-Oxygenation-Level-Dependent (BOLD) fMRI [1] and (2) Arterial Spin Labeling (ASL) perfusion MRI [2]. BOLD fMRI offers higher spatial and temporal resolution than ASL perfusion MRI but is only a relative measurement and fluctuates artifactually over space and time. Different anatomical parameters (e.g. vessel size and orientation), physiological parameters (e.g. oxygenation level and blood volume), and magnetic resonance-related parameters (e.g. magnetic field and echo time) can affect BOLD signals [3]. Thus, BOLD signal is not comparable across different brain regions and subjects, let alone between healthy subjects and patients. Besides, it is also not suitable for longitudinal studies due to large variations of BOLD signal changes. The most common practice for BOLD fMRI study is to compare BOLD signals across different tasks in the same brain regions.

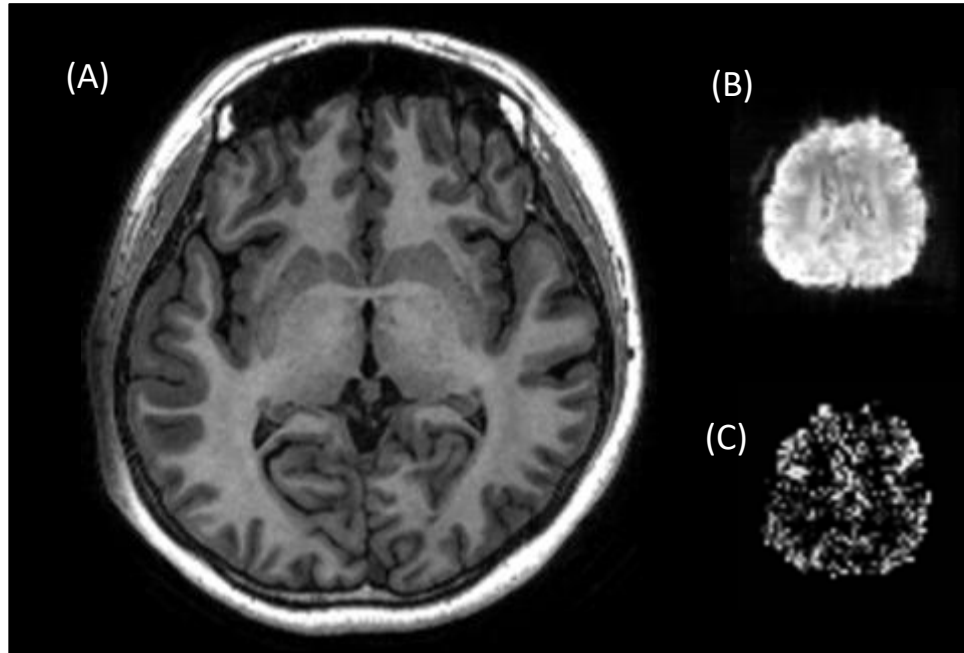


Figure 1.1: A demonstration of (A) a single MRI scan; (B) a single BOLD fMRI scan; and (C) an CBF map generated from ASL MRI scan. They are obtained at the same axial slice from one representative subject. An MRI scan only shows anatomical structure information while BOLD fMRI and ASL CBF show the metabolic activity within the anatomical structure. MRI has higher spatial resolution but lower temporal resolution than BOLD fMRI and ASL MRI. ASL CBF offers quantitative measurement, in which the intensity of a voxel is in a physical unit of ml/100g/min. BOLD fMRI has higher spatial and temporal resolution than ASL MRI but is a relative measurement.

In contrast, the disadvantages of BOLD fMRI can be overcome using ASL perfusion MRI techniques. ASL perfusion MRI can quantitatively measure the Cerebral Blood Flow (CBF) in a physical unit of ml/100g/min. The quantitative nature of ASL MRI makes it independent to imaging parameters and magnetic field. Therefore, ASL perfusion fMRI is very useful for cross-sectional and longitudinal studies [4] [5]. In addition, ASL MRI measures signals from the capillary bed, which is potentially more accurate for localizing brain functional activation than BOLD fMRI as BOLD signal is mainly contributed by oxygen level change in venous vessels rather than the neural activation site [4] [5]. Despite its advantages, ASL perfusion MRI is less popular because it has lower

signal-to-noise ratio (SNR), poorer temporal resolution and is more difficult to implement than BOLD fMRI [4]. Among the limitations of ASL perfusion MRI, the low SNR is the most critical one. In order to improve the SNR and get reliable perfusion measurements, repeated ASL scans are required. By averaging numerous repeated ASL scans, noises are suppressed, and signals are enhanced. However, repeated ASL scans result in impractical long scanning time and increased susceptibility to motion artifacts [6] [7].

To improve the SNR and reduce the number of repeated ASL scans, a number of post-processing methods have been proposed, including ASL MRI specific motion correction [8], physiological noise correction [9], and spatial noise reduction [10] [11] [12]. Advanced methods have also been published to suppress non-local noise [13] and spatio-temporal noise [14] [6] [15] [16]. Though these methods improved SNR of ASL MRI, they are based on either implicit or explicit models about the data, which may not be accurate and may change across subjects.

Machine learning, especially one of its branches, Deep Learning (DL), is now dominating nearly every field it has reached such as image classification, computer vision, auditory processing, information generation, and translational research [21,29,30,35,36]. Deep learning is an emerging machine learning technique that can learn a transform function from acquired data without using any explicit hypothesis about that function. Such flexibility is particularly beneficial for ASL perfusion MRI signal processing. Encouraged by the outstanding performance as listed above, machine learning and deep learning have been introduced into ASL signal processing, including adaptive outlier cleaning [17] [18] [19] [20], ASL denoising [21] [22] [23], and ASL quantification [15]. However, current

machine learning-based ASL signal processing has several limitations that are discussed in the next section.

### 1.1 Major Challenges

One major limit for DL-based ASL denoising is that there is no noise-free groundtruth image as training reference. When training a Deep Learning Neural Network (DLNN) for image denoising, noise-free images were required as reference images for training. There are abundant noise-free images in natural image denoising domain, but for ASL MRI, no noise-free images can be obtained as ASL MRI scans have intrinsically low SNR. By averaging of repeated ASL MRI scans and deliberate post-processing, quasi-noise-free reference can be obtained for training. However, quasi-noise-free reference is still of low SNR and may affect the performance of trained DLNN, as it may introduce artifacts during training. Besides, obtaining quasi-noise-free reference is time-consuming and reduces the number of available training samples significantly.

In addition, current network architectures for DL-based ASL MRI denoising methods, such as residual network [22] and Dilated Network [21], were not optimized for image denoising. These networks were originally proposed for high-level vision tasks such as image segmentation and classification [24] [25] [26]. Applying these architectures directly to low-level vision tasks such as denoising and super-resolution can be suboptimal [27] [28]. Furthermore, current studies [22] [21] are only trained on the limited size of *in vivo* dataset with limited validation. The fidelity and generalization ability of DL-based ASL denoising on *in vivo* dataset is yet examined thoroughly.

One of the other challenges in ASL perfusion MRI signal processing is in ASL adaptive outlier cleaning. During the pairwise subtraction process to extract CBF map from

ASL MRI scan, extreme outliers may be introduced due to low signal-to-noise-ratio (SNR) and unavoidable head motions. Thus, removing outliers is a very important step in the post-processing pipeline for ASL MRI. Current state-of-art Adaptive outlier cleaning (AOC) methods [18] has two issues: one is that current AOC methods discard whole outlier volumes (3D images) without considering whether they contain non-outlier slices (2D images) or not. Discarding the entire volume reduces the sample size and the SNR for non-outlier voxels. The other is that current AOC methods did not consider prior information (such as anatomical structure information) about CBF distribution.

## 1.2 Research Objectives

The main focus of this dissertation was to develop machine learning-based methods to overcome the major challenges mentioned in section 1.1, improving SNR of ASL MRI and reducing ASL MRI acquisition time. We proposed three methods to address these challenges, respectively. First, we proposed a learning-from-noise method to address the lack of noise-free reference images problem. Second, we proposed a novel deep learning neural network that is optimized for ASL image denoising. Third, a prior-guided adaptive outlier cleaning algorithm was proposed for ASL MRI post-processing. In what follows, we give a brief overview of these proposed methods.

### 1.2.1 A Learning-From-Noise Method for ASL Denoising And Prediction

In Chapter 3, a learning-from-noise method is proposed to show that DL-based ASL denoising (ASLDN) models can be trained using only noisy image pairs. The proposed method does not require any quasi-noise-free reference during the training process. We dub this new method as ASLDN-LFN (learning-from-noise). This method can be applied

to denoising ASL images directly. Experimental results show that ASLDN-LFN can reliably denoise ASL images and achieve improved SNR than ASLDN. Moreover, by using ASLDN-LFN, more training data can be generated as it requires less ASL MRI scans to generate reference mean ASL CBF maps, which is particularly useful when ASL CBF data are limited.

In addition, ASLDN-LFN can also be applied to predict ASL CBF maps directly from BOLD fMRI. BOLD fMRI offers higher spatial and temporal resolution than ASL MRI while ASL MRI offers quantitative measurement. It would be ideal if we can extract quantitative ASL CBF maps and obtain high SNR and temporal resolution from BOLD fMRI. However, predicting ASL CBF is a challenging task as a single ASL CBF image has very low SNR, which makes it unsuitable to be used as a training reference. ASLDN-LFN shows it is feasible to use a DL model to achieve this goal as the training process of DL model converge to mean or median of ASL CBF maps.

### 1.2.2 Dilated Wide Activation Network-based ASL Denoising

In chapter 4, we proposed a novel DL model that optimized for ASL denoising. Our proposed method presents two novelties. first, we incorporated wide activation residual blocks [28] with a Dilated Convolution Neural Network (DilatedNet) [25] to achieve improved denoising performance in term of several quantitative and qualitative measurements; second, we evaluated our proposed model given different inputs and references to show that ASLDN denoising model can be generalized to input with different levels of SNR and yielded images with better quality than other methods.

### 1.2.3 Prior-Guided and Slice-Wise Adaptive Outlier Cleaning For ASL

In chapter 5, a prior-guided and slice-wise AOC method was proposed to address the limitations of current AOC methods. First, we adopt anatomical structure information as prior information and use it to guide the outliers cleaning method in early iterations. Second, we reject outliers slice-wise instead of volume-wise, which saves non-outlier slices from the outlier volume and improves the SNR for non-outlier voxels. Compared with current outlier cleaning methods, the proposed method showed both improved CBF quantification quality and CBF measurement stability.

### 1.3 Contributions

Contributions have been made in the following aspects:

- This is the first known attempt to use the learning-from-noise method for DL-based ASL denoising and prediction, which does not require any noise-free reference images, improving SNR and increasing training data samples.
- This is the first study to extract quantitative CBF from BOLD fMRI using deep learning, improving the SNR and temporal resolution of ASL CBF.
- A new CNN architecture that is optimized for image denoising was developed. The method achieves better ASL image quality than current DL-based ASL denoising and reduces ASL acquisition time by 75%. We tested generalizability of our model on a large *in vivo* dataset with different input noise levels.
- A prior-guided and slice-wise adaptive outlier cleaning method was proposed to improve CBF quantification quality.

### 1.4 Organization of this Dissertation

The rest of this dissertation is organized as follows: chapter 2 introduces the necessary background of fMRI and deep learning and reviews related works in machine learning-based ASL signal processing. The proposed learning-from-noise ASL signal processing method and its applications are described in Chapter 3. Chapter 4 presents a new DL model that is optimized for ASL denoising and extensive validation on large *in vivo* dataset. A prior-guided slice-wise adaptive outlier cleaning method is proposed in Chapter 5. Conclusions and future works are discussed in Chapter 6.

## **CHAPTER 2**

### **BACKGROUND AND RELATED WORKS**

In this chapter, a brief overview of fMRI and deep learning is provided. In addition, we review related works in ASL MRI signal processing.

#### 2.1 Functional MRI

Both MRI and fMRI are primarily measuring signals from protons (H) on water molecules (H<sub>2</sub>O). Due to different imaging techniques and parameters, MRI shows the anatomical structure while functional MRI (fMRI) shows the tissue function. FMRI signal in each voxel is an accumulation of signals from extravascular tissue, intravascular arterial blood, intravascular venous blood, and cerebrospinal fluid. Their relative contributions to the fMRI signal are weighted by different imaging techniques and parameters. Typically, the BOLD signal is 0.5 - 5% of the total signal intensity while the ASL perfusion signal is about one percent of the total signal intensity [4] [5][16].

fMRI detects tissue function using either Blood-oxygenation-level-dependent (BOLD) contrast or Arterial Spin Labeling (ASL) perfusion contrast. The BOLD fMRI was first invented by Seiji Ogawa et al. [1] in 1990. This technique detects tissue function activation by measuring the oxygenation level in venules [29], i.e., the relative level of oxyhemoglobin and deoxyhemoglobin in venules (As illustrated in Figure 2.1). When neurons are in resting states, the relative level of oxyhemoglobin is low in venules. When neurons are in stimulated states, the relative level of oxyhemoglobin in venules increases. The difference in magnetic properties of oxyhemoglobin and deoxyhemoglobin can be detected by BOLD fMRI, producing an effective map of active and inactive areas of a brain.

BOLD fMRI has higher spatiotemporal resolution and is easier to implement, making it a more popular fMRI technique than ASL perfusion MRI [4]. However, BOLD fMRI is only a relative measurement as it is sensitive to physiological parameters, anatomical parameters, and magnetic resonance-related parameters, which might not be suitable for cross-sectional study or longitudinal study. Besides, there is also a concern about BOLD fMRI localization accuracy. The goal of fMRI is to measure intravascular signals from capillary beds as capillary beds are close to neuronal activation sites. But BOLD signals are mainly from venules, which might affect localization accuracy.

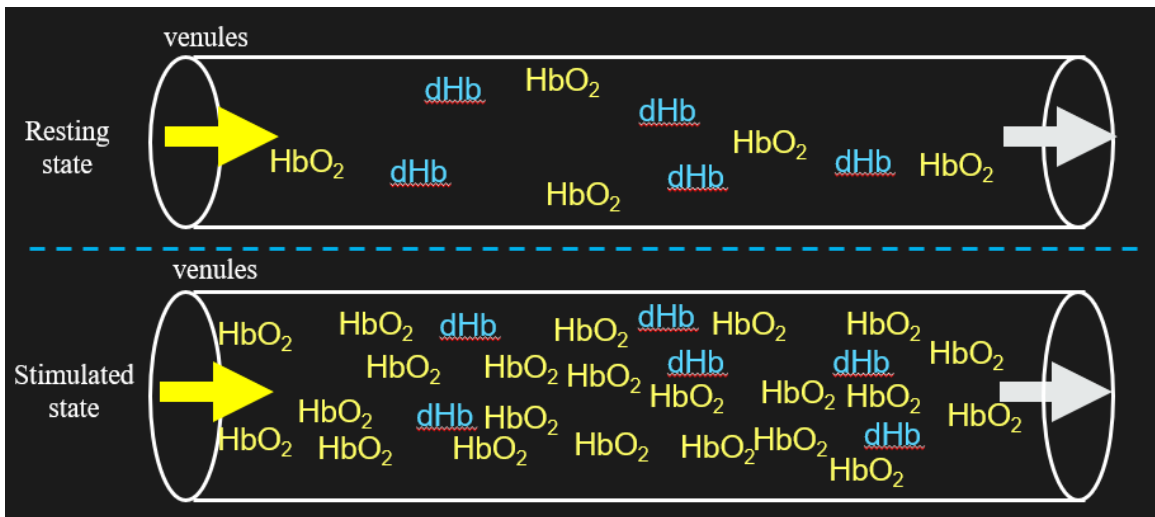


Figure 2.1: Illustration of hemodynamic process. When neurons are in resting states, the relative level of oxyhemoglobin is low in venules. When neurons are in a stimulated state, the vessel size, blood flow and the relative level of oxyhemoglobin in venules increases. HbO<sub>2</sub> = oxyhemoglobin, dHb = deoxyhemoglobin.

Arterial spin labeling (ASL) measures cerebral blood flow (CBF) quantitatively by magnetically labeling blood water as it flows throughout the brain [2] [30]. ASL CBF is a quantitative measurement that has a physical meaning, which is defined as an amount of blood moving into capillaries within 100g tissue per minute (unit ml /100 g / min]). Based

on the hemodynamic response, when the neural activity in certain areas increases, the CBF in that area increases. Thus, a change in CBF is a good indicator of neural activity.

The process of ASL perfusion MRI scan is as follows: (1) arterial blood water is labeled with radiofrequency (RF) pulses in locations proximal to the tissue of interest; (2) a spin labeled image is acquired after the labeled blood water reaches the imaging plane and perfuse into brain tissue. It not only has CBF perfusion-weighted signal but also background tissue signals; (3) To remove the background tissue signal, a control image without perfusion-weighted signal is also acquired using the same ASL technique but without labeling of flowing blood; (4) A perfusion-weighted image is subsequently determined by pair-wise subtraction of the spin labeled image (the label image or L image) and the spin untagged image (the control image or C image); (5) then, the perfusion-weighted image is converted into the quantitative CBF map in a unit of ml/100 g/min [7].

Limited by the longitudinal relaxation rate ( $T_1$ ) of blood water, labeling efficiency, and the post-labeling delay, the labeled blood signal is about one percent of total signal intensity, resulting in a very low SNR [31]. Thus, many pairs of L/C images are often acquired to improve the SNR of the mean perfusion map (as illustrated in Figure 2.2). Because the total scan time is often around 3-5 mins, only 10-50 L/C pairs can be acquired, resulting in a modest SNR improvement by averaging across the limited number of measurements. It should be noted that even for the best research subjects, head motion will occur during the scan (e.g. due to swallowing) [3]. Head motion will significantly affect the pair-wise subtraction results, introducing excess outliers and noises to ASL CBF. Thus, shorter scanning time is preferred.

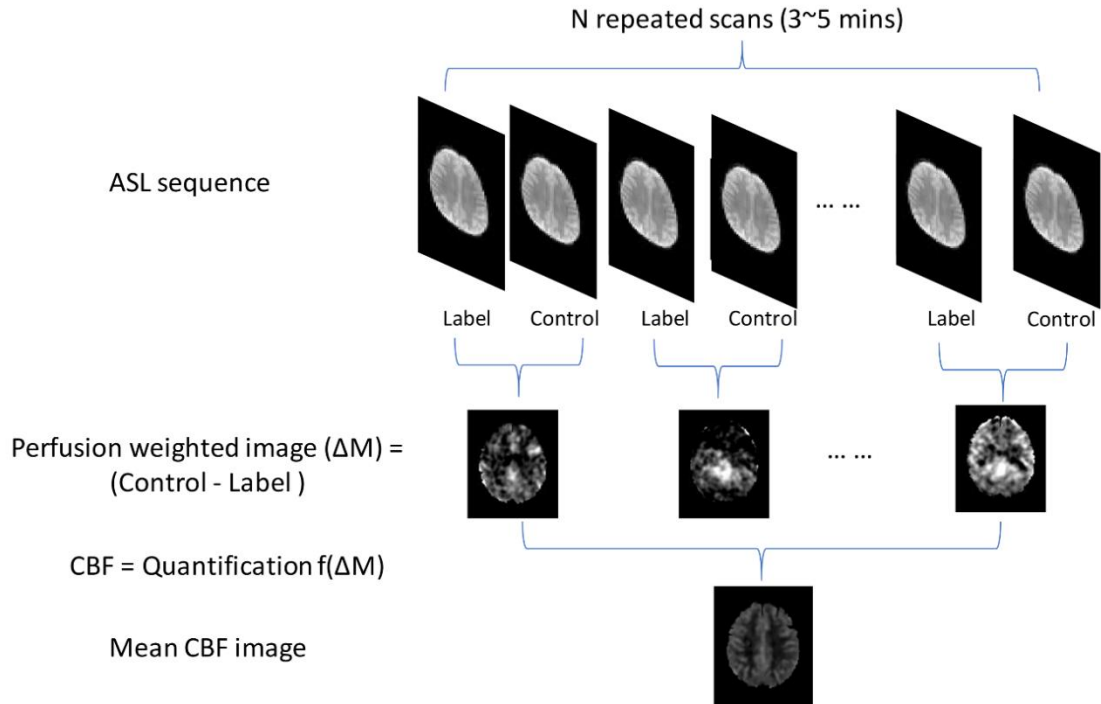


Figure 2.2: An example of the ASL acquisition process. In an ASL MRI scan, multiple pairs of label and control images were acquired. Perfusion-weighted image is obtained by pair-wise subtraction and is subsequently quantified to CBF map in a unit of ml/100 g/min. By averaging a series of CBF maps, moderate noises were removed in the mean CBF map.

## 2.2 Deep Learning

Deep learning (DL) is a subtype of machine learning (ML) algorithms [32] which has made widespread impact on nearly every research field it has been applied (from image classification [33] [34], video recognition [35] [36] [37], voice recognition/generation [38] [39] [40] [41], medical image processing [42] [43] [44], to AlphaGo [45], AlphaGo Zero [46], etc.). The concept of DL can be traced back to the early 1980s [47], but only became practical until the advent of fast general-purpose graphics processors in the late 2000s [33] [48]. DL is now dominating nearly every field it has reached such as classification, computer vision, auditory processing, information generation, and translational research [34] [42] [43] [49] [50] [51]. Using 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 [49] [52] [9] [32]. Encouraged by the outstanding performance as listed above, DL has been introduced into many medical imaging processing fields, including image segmentation [43] [42], image reconstructions [53] [54], image synthesis [55] [56] [57] [58], etc.

The most popular DL based denoising approach is based on convolutional neural networks (CNNs) [34]. CNNs learn a hierarchy of features by a series of convolution, feature pooling, and non-linear activation operations, presenting high flexibility and capability for learning distributions or manifold of images [59]. CNN achieves tremendous success in both high-level computer vision tasks such as image classification [34] [36], object detection [26] and low-level computer vision tasks such as image denoising and image super-resolution [28]. A typical CNN for high-level computer vision tasks contains four types of layers: convolution layers, activation layer, sub-sampling layers, and fully connected layers. Convolution layers are used to extract features while the activation layer provides non-linearity to the network. Sub-sampling layers are used to reduce the size of the input and extract dominant features. Fully connected layers are used for high-level reasoning, which is necessary for classification tasks. An introduction to each type of layer is provided in the following paragraphs.

Input					
3	15	64	22	55	62
92	213	7	32	145	34
17	178	86	33	12	21
231	87	48	5	23	234
59	56	55	45	3	218
82	97	94	33	238	44

Convolution kernel		
1	1	1
1	0	2
1	0	1

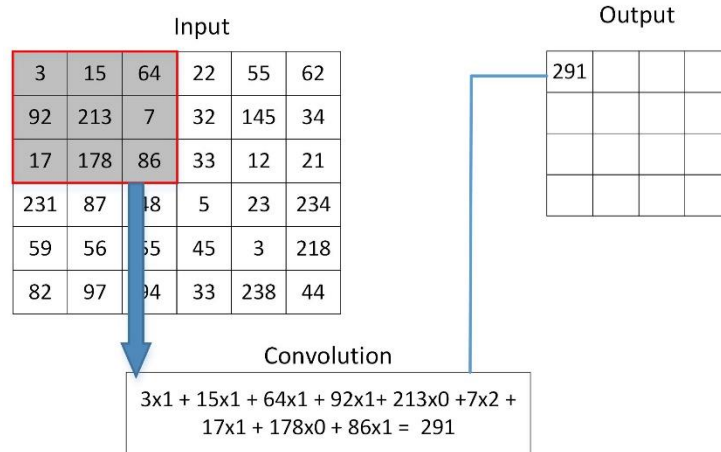


Figure 2.3: An example of a convolution operation.

**Convolution Layer:** The convolution layer takes the convolution of the input image with the convolution kernel and generates the output. As Figure 2.3 shows, in the convolution layer, the top left matrix is the input matrix, which can be regarded as pixel values of a digital image, and the top right matrix is a convolution kernel. The convolution kernel is called filter and the output is called filter response or feature map. Each time a block of pixels is convolved with a filter and generates value in the filter response. Convolution kernels are learnable weights that are updated by the backpropagation algorithm during training.

**Activation Layer:** the output of the convolution layer is further processed by the activation layer, which provides nonlinearity to the neural network. The activation function in the activation layer can be sigmoid, hyperbolic tangent (tanh), and Rectified Linear Unit (ReLU), etc. The Sigmoid function is defined as  $f(x) = \frac{1}{1+e^{-x}}$ , which bounds the output to (-1,1). Tanh is a variation of the sigmoid function. Sigmoid and tanh have two disadvantages when used as activation functions: 1) they both have exponential operations which is of high computational complexity 2) gradients of these functions tend to zero after

many layers of backpropagation, introducing vanishing gradient problems. These disadvantages make training a neural network difficult.

Rectified Linear Unit (ReLU) is an activation function that was proposed to overcome the disadvantages mentioned above. ReLU is formulated as  $f(x) = \max(0, x)$ . It has very low computational complexity while provides nonlinearity to a neural network. Besides, the gradient of ReLU is also non-trivial when the input  $x$  is larger than zero. Thus, ReLU not only provides nonlinearity and low computational complexity to a neural network but also eliminates vanishing gradient problems. It has become the most common activation function used in current CNN architectures.

One problem of ReLU is that the gradient will become zero when the input  $x$  is smaller than zero, which cannot be updated by back-propagation. This is the so-called dead ReLU problem. To solve this problem, variations of ReLU, such as Leaky ReLU and Parametric ReLU (PReLU) [60], were proposed. Leaky ReLU is defined as  $f(x) = \max(\alpha x, x)$ , where  $\alpha$  is a small constant value (typically 0.01). PReLU is also formulated as  $f(x) = \max(\alpha x, x)$ , while  $\alpha$  is a parameter updated by back-propagation. These variations allow a small gradient when input  $x$  negative, prevent dead ReLU problem.

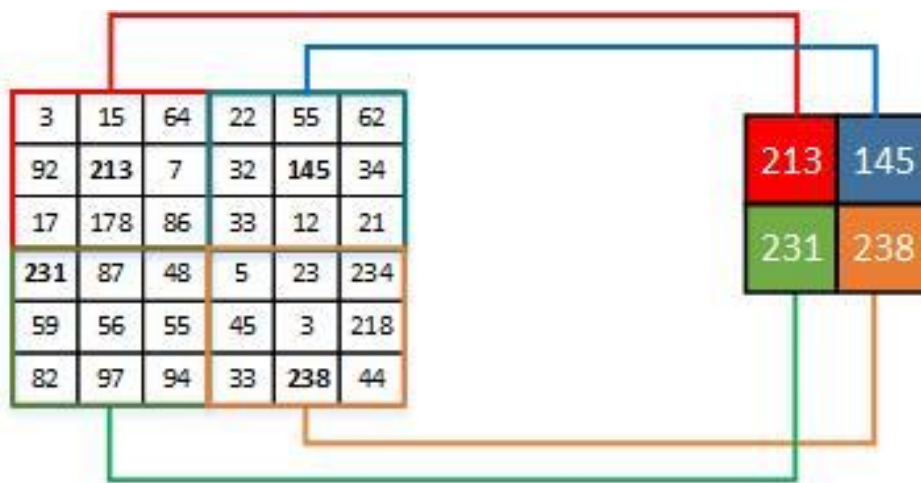


Figure 2.4: The example of the pooling layer.

**Pooling layer:** The pooling layer, or sub-sampling layer, is an important layer to a convolutional neural network. The most used method for the pooling layer in image processing tasks is max-pooling. The max-pooling method is shown in Figure 2.4. The image is divided into blocks and the maximum value of each block is the corresponding pixel value of the output image. The reason to use the pooling layer is as follows: First, the pooling layer decreases the input matrix size, which reduces the computational power for processing the data. Second, a pooling layer extracts dominant features, making CNN invariant to small translation and rotation among the input pattern.

**Fully Connected Layer:** When a CNN is trained for high-level vision tasks such as image classification or object detection, fully connected layers are stacked after convolution layers and pooling layers for high-level reasoning. Fully connected layers make the neural network feed forward into vectors with a predefined length. By doing so, we could fit the vector into certain categories for classification tasks or take it as a representation vector for further processing. A fully connected layer has connections to all activations in the previous layer. Each of their outputs is defined as  $z_j = \mathbf{w}_j^T \mathbf{x} + b$ , where  $\mathbf{x}$  is the input vector,  $b$  is the bias term, and  $\mathbf{w}_j$  is the learnable weight vector corresponding to  $j$ -th neuron and is updated by backpropagation.

**Up-sampling layer:** The up-sampling layer is often used to recover resolution from max-pooling and other image down-sampling layers. The up-sampling layer is also used for image super-resolution tasks to improve the resolution of input images. The most common up-sampling layers are the deconvolution layer and sub-pixel convolution layers.

Deconvolution, or transposed convolution, can be regarded as a product between each input pixel and a filter element-wisely with stride  $r$  and add up the resulting output windows.

**Batch Normalization:** Batch Normalization (BN) [61] is a strategy to accelerate the training process and improve the training accuracy. BN was designed to prevent internal covariance shift due to mini-batch stochastic gradient descent (SGD) which changes the distributions of internal non-linearity inputs during training. BN is motivated by the fact that data whitening process improves performance. First, BN normalizes the output of the previous layer (Conv or ReLU) with zero mean and unit variance within a batch of training images; Second, BN optimally shifts and scales these normalized activations.

## 2.3 Related Works

In this section, current machine learning-based ASL denoising methods are reviewed in section 2.3.1. and related outlier cleaning methods are reviewed in section 2.3.2.

### 2.3.1 ASL Denoising

Advanced methods have also been published to suppress non-local noise [13] and spatio-temporal noise [14] [6] [15] [16]. Zhu et al. [49] used Robust principal component analysis to denoise the ASL CBF maps. The CBF images series were decomposed into two parts: a low-rank component which captures perfusion patterns; and a sparse component which captures spatially incoherent spiky variations. The sparse component was regarded as noise and was subsequently discarded. Wang [15] proposed to use Support Vector Machine (SVM) to suppress spatiotemporal noise during the CBF quantification process (dubbed as SVMASLQ). SVM was used to separate the label and control images of ASL MRI and

then the perfusion-weighted image was subsequently extracted from the multivariate SVM classifier. Zhu et al. [16] further improved SVMASLQ by using a patch-wise classification kernel to separate local signal and noise variations. Image patches centered at each voxel were extracted from both the labeled images and controlled images, and then input into SVMASLQ to find the surrogate perfusion map. Though these methods improved SNR of ASL MRI, they are based on either implicit or explicit models about the data, which may not be accurate and may change across subjects.

By contrast, DL-based denoising methods learn the denoising model directly from the noise-contaminated data. One journal paper [21], one conference paper [22], and a conference abstract [23] have been published in ASL denoising using DL. Kim et al. [21] published the first paper on this research topic. Their denoising CNNs consist of two parallel pathways to integrate the multiscale 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 segmentation [62] [63], which may not be optimal for denoising. Ulas et al. [22] trained a deep learning model with a customized loss function based on the Buxton Kinectic model [64], but with a simple CNN architecture. Gong et al. [23] proposed a technique first using a multi-lateral guided filter to post-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 that cannot incorporate global information for denoising. Generating a stack of multi-contrast images could be time-consuming. While encouraging, these studies were all based on small sample sizes and used a standard CNN not specifically optimized for denoising.

### 2.3.2 Outlier Cleaning

A series of ASL images are usually obtained to generate the mean perfusion map so that final CBF quantification can have higher SNR. However, a small number of outlier images can significantly affect the mean CBF map, causing signal loss or hyper-intensity areas. Various outlier cleaning methods have been proposed to identify and remove outliers. Wang et al. [65] identify outliers based on amplitude and successive differences of head motions as well as the mean and standard deviation of the whole brain CBF time series. Tan et al. [66] used the mean and standard deviation of each CBF volume to determine outlier volumes. Wang et al. [17] proposed an adaptive outlier cleaning algorithm (AOC) based on the correlation between grey matter and CBF. Maumet et al. [67] estimated perfusion-weighted maps using Huber's M-estimator that is robust to outliers. A common issue of these methods is that the reference CBF map used for identifying outliers is the intermediate mean of the remaining CBF images, which may be initially contaminated by the outliers and will favor outliers and reject non-outliers. Another issue is that they discard the entire outlier volume even if the volume contains slices that are non-outlier. A third concern is that the rejection criterion is purely distance (or correlation) based. The method by Tan et al. [66] can reject outlier slices but the method is still subject to other concerns as listed above. Dolui et al. [18] improved AOC based on an empirical assumption of high correlations between outliers and the reference mean, but left the volume-wise rejection issue alone.

## **CHAPTER 3**

### **A LEARNING-FROM-NOISE DEEP LEARNING METHOD FOR ARTERIAL SPIN LABELING MRI DENOISING AND PREDICTION**

One major limit for DL-based ASL denoising is that there are no noise-free groundtruth images as training references. When training a DL model for image denoising, noise-free images were required as reference images for training. There are abundant noise-free images in the natural image denoising domain, but there is no noise-free image that can be obtained for ASL MRI as ASL scans have intrinsically low SNR. By averaging of repeated ASL scans and deliberate post-processing, quasi-noise-free reference can be obtained for training. However, quasi-noise-free reference is still of low SNR and may affect the performance of trained DL model, as it may introduce artifacts during the signal averaging and post-processing steps. Besides, obtaining quasi-noise-free reference is time-consuming and reduces the number of available training samples significantly.

In this chapter, a learning-from-noise method is proposed for DL-based ASL signal processing (dubbed as ASLDN-LFN). The proposed learning-from-noise method shows that DL-based ASL denoising models can be trained using only noisy image pairs, without any deliberate post-processing for getting the quasi-noise-free reference during the training process. Specifically, we show that this learning-from-noise training process converges exactly with the signal averaging process of ASL CBF maps. In other words, it is not necessary to have a noise-free reference for ASLDN-LFN. By using this method, we can skip the step of obtaining quasi-noise-free reference and obtaining equal or even better performance than ASLDN. Moreover, more training data can be generated as ASLDN-

LFN does not need extra samples to generate quasi-noise-free reference, which is particularly useful when ASL CBF data are limited.

Besides, this learning-from-noise method can also be applied to DL-based ASL CBF prediction from BOLD fMRI. ASL is quantitative, insensitive to low-frequency drift but has lower signal-to-noise-ratio (SNR) and lower temporal resolution than BOLD. Currently, there still lacks a way to fuse the benefits provided by both, i.e., to quantify cerebral blood flow (CBF) like ASL MRI but with high SNR as in BOLD fMRI. The main challenge for using DL-based methods to predict ASL CBF from BOLD fMRI is that the CBF from a single scan of ASL MRI is too noisy to be used as a reference in this BOLD-to-ASL prediction. However, our learning-from-noise method shows that it is possible to achieve BOLD-to-ASL prediction using extremely noisy ASL CBF maps as training references. Experimental results demonstrated that this learning-from-noise method can reliably predict ASL CBF from BOLD fMRI.

### 3.1 Introduction

DL-based denoising methods typically use an end-to-end training scheme to train the DL model. A noisy image that we want to denoise is fed to the input end and the desired noise-free image is fed to the output end as the training reference. By using stochastic gradient-based optimization [68, 69], the parameters of the DL model are automatically adjusted so that given the input noisy image, the DL model can predict an output image that is as similar as possible to the noise-free training reference. The similarity between output image and the reference image is determined by a loss function (typically mean square error or mean absolute error). Essentially, the DL-based denoising is a regression task.

Formally, denote the noisy image by  $\hat{x}_i$  and its reference (noise-free or less noisy version) by  $y_i$ , where  $i=1\dots N$ ,  $N$  is the total number of training samples. A parametric regression DL model  $f_\theta$ , typically a convolutional neural network (CNN), can be built to learn the mapping  $f_\theta(\hat{x}_i) \rightarrow y_i$  by minimizing the following loss function:

$$\arg \min_{\theta} \sum_i L(f_\theta(\hat{x}_i), y_i) \quad (3.1)$$

, where  $\theta$  are the parameters of CNN and are adjusted through the training process under the loss function  $L$ .

Several groups have used DL in ASL MRI denoising [5, 17]. Different from other denoising applications, DL-based ASL denoising networks (ASLDN) do not have noise-free training references. Accordingly, its denoising performance might be uplimited by the reference image SNR. However, the potential uplimit doesn't seem to exist, as several studies [70] [21] [22] showed that ASLDN could produce CBF images with even higher SNR than the reference. This apparent learning-from-noise capability could provide a versatile ASLDN without deliberate post-processing for getting the quasi-noise-free reference. Lehtinen et al. [71] has recently been explicitly investigating this learning-from-noise capacity in natural image denoising. Inspired by their work in natural image denoising, we want to formally introduce this learning-from-noise capability of DL on ASL MRI signal processing and demonstrate that this learning-from-noise capability is particularly beneficial for ASL MRI where data are corrupted and limited.

### 3.2 Methods

When estimating a true value  $y$  from a set of unreliable observations  $(x_1, x_2, \dots, x_N)$ , A common strategy to estimate the true value  $y$  is to find a value  $\hat{y}$  that has the smallest average deviation from all the unreliable observations according to some loss function  $L$ .

This strategy can be formulated as:

$$\arg \min_{\hat{y}} \frac{1}{n} \sum_{i=1}^n L(x_i, \hat{y}). \quad (3.2)$$

When the loss function is  $L_2$  loss, i.e.,  $L(x, \hat{y}) = (x - \hat{y})^2$ , the optimal value to minimize the loss function can be derived as follows:

$$\begin{aligned} \frac{\partial L}{\partial x} &= \frac{2}{N} \sum_{i=1}^N (x_i - \hat{y}) = 0 \\ \hat{y} &= \frac{1}{N} \sum_{i=1}^N x_i = E(\mathbf{x}) \end{aligned} \quad (3.3)$$

Thus, the optimal value  $\hat{y}$  to minimize the loss function  $L$  is the expectation of all observations  $(x_1, x_2, \dots, x_N)$ . Optimal value can be derived for the  $L_1$  loss  $L(x, \hat{y}) = |x - \hat{y}|$  using the same method:

$$\frac{\partial L}{\partial x} = \frac{1}{N} \sum_{i=1}^N \text{sign}(x_i - \hat{y}) = 0, \quad (3.4)$$

where

$$\text{sign}(x_i - \hat{y}) = \begin{cases} -1, & x_i < \hat{y} \\ 0, & x_i = \hat{y} \\ 1, & x_i > \hat{y} \end{cases}. \quad (3.5)$$

Thus, the optimal value to minimize  $L_I$  loss is the median of observations  $(x_1, x_2, \dots, x_N)$ . When the observations are corrupted by Gaussian noise, i.e., a set of noisy observation  $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N)$  that  $\hat{x}_i = x_i + n_i$ , where  $x_i$  is the observation without noise and  $n_i$  is an independent sample draw from a zero-mean gaussian distribution ( $n_i \sim N(0, \sigma)$ ). When minimizing the loss function:

$$\arg \min_{\hat{y}} \frac{1}{n} \sum_{i=1}^n L(\hat{x}_i, \hat{y}). \quad (3.6)$$

For L2 loss  $L(\hat{x}, \hat{y}) = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - \hat{y})^2$ , the optimal value  $\hat{y}$  can be derived as follows:

$$\begin{aligned} \frac{\partial L}{\partial \hat{x}} &= \frac{2}{N} \sum_{i=1}^N (\hat{x}_i - \hat{y}) = 0 \\ \hat{y} &= \frac{1}{N} \sum_{i=1}^N \hat{x}_i = \frac{1}{N} \sum_{i=1}^N (x_i + n_i) \\ &= \frac{1}{N} \sum_{i=1}^N x_i + \frac{1}{N} \sum_{i=1}^N n_i = E(\mathbf{x}) + E(\mathbf{n}) = E(\mathbf{x}) \end{aligned} \quad (3.7)$$

Therefore, when the observations were corrupted by zero-mean gaussian noise, the optimal value to minimize the loss function is the mean of the latent clean observations  $(x_1, x_2, \dots, x_N)$ , given infinite data samples. The more generalized class of this deviation-minimizing type estimators are called the M-estimators [72].

This point estimation procedure can be generalized to training a neural network. Given a set of input-reference pairs  $(x_i, y_i)$  and a neural network function  $f_{\theta}(x)$  parameterized by  $\theta$ , the training procedure of the neural network is:

$$\arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n L(f_{\theta}(x_i), y_i). \quad (3.8)$$

If the network  $f_{\theta}$  does not depend on input data, and merely outputs a learned scalar, the whole training procedure reduces to the same point estimation procedure of (3.2) at every training sample. Thus, training a neural network using  $L_2$  loss is to find the expectation of references  $y_i$ .

When the references  $\hat{y}_i$  are drawn from a corrupted distribution of the latent clean reference  $y_i$  such that  $E\{\hat{y}_i\} = y_i$ , and are combined with corrupted inputs from the equation (3.1), the learning process of neural network is equal to minimize the loss function:

$$\arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n L(f_{\theta}(\hat{x}_i), \hat{y}_i). \quad (3.9)$$

When given infinite training samples and the loss function is  $L_2$ , the solution is  $f_{\theta}(\hat{x}_i) = E\{\hat{y}_i\} = y_i$ . When the training samples are finite and mutually uncorrelated, the expected squared difference between  $E\{\hat{y}_i\}$  and  $E\{y_i\}$  is equal to  $\frac{1}{N^2} \text{Var}(\sum_{i=0}^N y_i)$  according to [71]. Thus, the variance of minimizing  $L_2$  loss given finite data is the mean variance of the corruptions divided by the number of samples  $N$ . When the number of training data increases, the error tends to zero. The above derivation assumes scalar data. When data are images,  $N$  is the total number of pixels in the images, i.e., *the number of images*  $\times$  *the number of pixels per image*.

According to the above derivation, we do not need clean images as references when training neural network for image denoising. Only noisy image pairs are needed for DL-based image denoising. This is particularly useful when no absolute clean reference images are available in ASL denoising.

### 3.3 Application on ASL denoising

In this section, we proposed a DL-based learning-from-noise method for ASL denoising, which does not require any deliberate post-processing for getting the quasi-noise-free reference during the training process. We dubbed this new method ASLDN-LFN. Figure 3.1 shows the difference between ASLDN and ASLDN-LFN. ASLDN requires quasi-noise-free image as training reference while ASLDN-LFN only use noisy image as training reference.

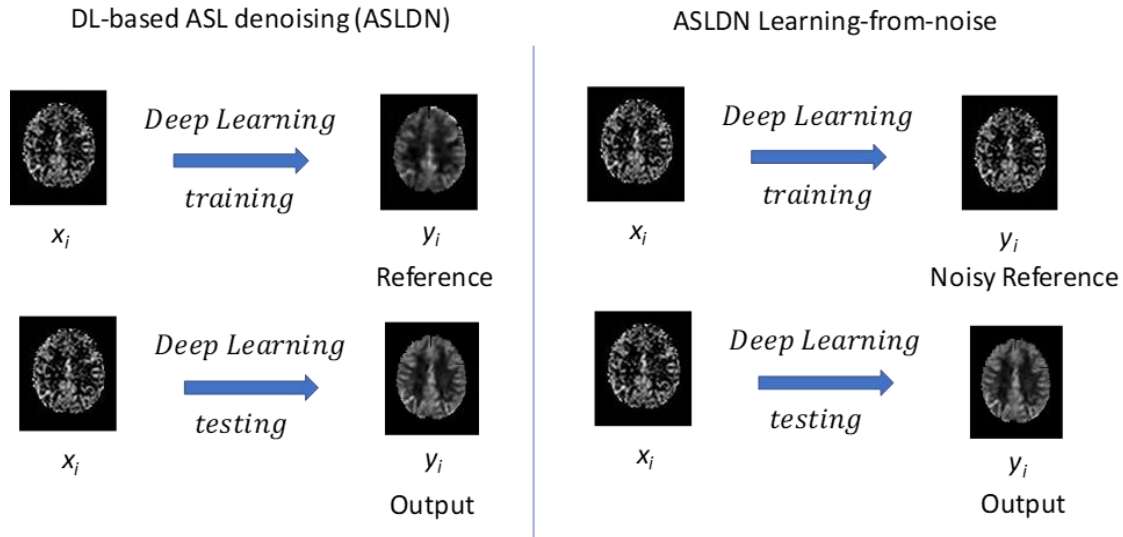


Figure 3.1: An illustration of DL-based ASL denoising (ASLDN) and ASLDN-LFN. ASLDN using quasi-noise-free CBF images as training references whereas ASLDN-LFN only use noisy image pairs to train the DL model.

Similar to [71], the assumption of ASLDN-LFN is that both the noisy reference  $\hat{y}_i$  and the noisy input CBF maps  $\hat{x}_i$  are drawn from the same data distribution. When minimizing the

$L_2$  loss function  $\frac{1}{n} \sum_{i=1}^n (f_{\theta}(\hat{x}_i) - \hat{y})^2$ , the CNN regressor  $f_{\theta}(\hat{x}_i)$  is to find the optimum at the arithmetic mean of the observations, i.e.  $f_{\theta}(\hat{x}_i) = E\{\hat{y}_i\} = y_i$ , given enough training samples. This process converges exactly with the method of averaging one subject's all CBF maps to generate a quasi-noise-free mean CBF map as a pseudo-groundtruth for training in ASLDN. However, considering ASL CBFs have excessive outliers, training with  $L_1$  loss is preferred as training with  $L_1$  loss is to find the median of the observations and the median is more robust to outliers than mean. We also conducted experiments to compare the effects of training with  $L_1$  loss versus training with  $L_2$  loss.

### 3.3.1 Materials and Experiment Setup

ASLDN-LFN using the Dilated Wide Activation Network (DWAN) that was proposed in chapter 4. As shown in Figure 3.2, DWAN has two pathways. The difference between the local pathway and global pathway is that the first convolution layer of the 4 wide activation residual blocks in the global pathway used a dilation rate of 2, 4, 8 and 16 respectively. The local pathway extracts the local features and the global pathway uses dilation convolutions to reserve global data patterns. Furthermore, the wide activation residual blocks in DWAN are able to expand data features and pass more information through the network, improving performances for low-level computer vision tasks without additional parameters and computation [28] [73]. By combining the two-pathway structure and the wide activation residual block, this new CNN structure (DWAN) improves the denoising performance in ASLDN-LFN.

ASL data were pooled from 280 subjects in a local database. The data were acquired with a pseudo-continuous ASL sequence (40 control/labeled image pairs with labeling time = 1.5 sec, post-labeling delay = 1.5 sec, Field of View (FOV)=22×22 cm<sup>2</sup>, matrix=64×64,

Repetition Time (TR)/ Echo Time (TE) = 4000/11 ms, 20 slices with a thickness of 5 mm plus 1 mm gap).

ASLtbx [65] 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 [8]; 2) the average of all 40 C/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 = 3mm 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 [65].  $M_0$  is approximated by the control image in each label/control image pair and  $M_0$  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

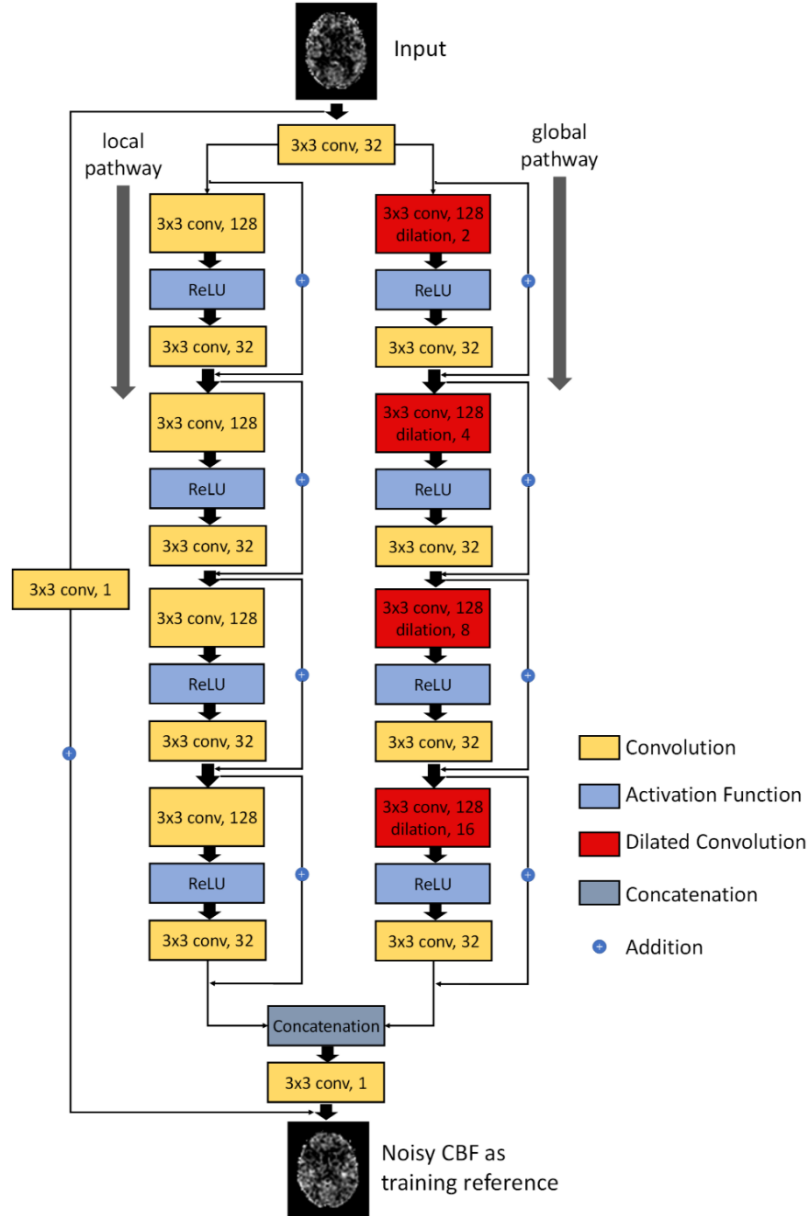


Figure 3.2: Schematic illustration of the architecture of our proposed DWAN network. The first layer consists of  $3 \times 3 \times 32$  convolutional filters for the input image. Then the output of the first layer was fed to the both local pathway and global pathway. Each pathway contains 4 consecutive wide activation residual blocks. Each wide activation residual block contains 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 the global pathway were dilated convolutional layers with a dilation rates of 2, 4, 8, 16, respectively. The output of the 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).

using the prior-guided slice-wise adaptive outlier cleaning algorithm [19] [20]; 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.

CBF image slices from 200 subjects were used as the training dataset. CBF images from 20 different subjects were used for validation. The remaining 60 subjects were used as the testing set. Input to ASLDN-LFN was the axial slice. All CBF maps were spatially normalized into the Montreal Neurological Institute (MNI) space. Every 3 slices from the 35th to the 59th axial slices were extracted from each of the 3D CBF maps. The 40 ASL CBF images of each subject were divided into 4 time segments, each with 10 successively acquired images. The mean maps of the 1st segment and the 2nd segment were taken as the input and the corresponding reference for DL model training. Another set of input-reference image pairs was obtained from the mean CBF maps of the 3rd and the 4th segment. During model testing, the mean CBF image slices of the first 10 L/C pairs (in the first time segment) were used as the input.

Due to intrinsic low SNR of ASL MRI, the input and reference CBF maps were already contaminated with severe noise (As Figure 3.3 A. shows). Therefore, no additional artificial noises were added to the input and reference CBF maps. Mean CBF maps of the entire 40 L/C image pairs with Gaussian smoothing (FWHM = 3mm) and state-of-art outlier cleaning [20] were used as pseudo-groundtruth  $y_i$ . Compared with the previous method ASLDN [74] using pseudo-groundtruth  $y_i$  as training references, the proposed ASLDN-LFN only used noisy data  $\hat{y}_i$  as the training reference. U-Net [24] and DilatedNet [21], two popular CNN structures widely used in medical imaging, were implemented as a

comparison to our DWAN-based ASLDN-FLN. Additional experiments were conducted to compare the effects of the different loss functions (L1 and L2) on denoising performance.

We used Keras [75] and Tensorflow [76] platforms to implement all the DL algorithms. Network training was through the adaptive moment estimation (ADAM) algorithm with a learning rate of 0.001 and a batch size of 64. All experiments were performed on a PC with Intel(R) Core(TM) i7-5820k CPU @3.30GHz and an Nvidia GeForce Titan Xp GPU.

We used Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index (SSIM) to quantitatively compare the performance of DWAN with U-Net and DilatedNet. When computing PSNR and SSIM, pseudo-groundtruth (mean CBF from entire 40 L/C pairs) was used as groundtruth. SNR and Grey Matter/White Matter (GM/WM) contrast was calculated to measure the image quality of ASL CBF. The SNR was calculated by using the mean signal of a grey matter (GM) region-of-interest (ROI) divided by the standard deviation of a white matter (WM) ROI in slice 50. The GM/WM contrast was calculated as the mean value of GM masked area divided by the mean value of WM masked area.

The Correlation coefficient between the DL-produced CBF values and pseudo-groundtruth was calculated to measure the similarity of the DL-produced CBF values to those processed with non-DL methods. This process was performed at each voxel for ASLDN and ASLDN-FLN separately. The correlation coefficient maps were thresholded by  $r > 0.3$  for the purpose of comparison and display.

### 3.3.2 Experimental Results

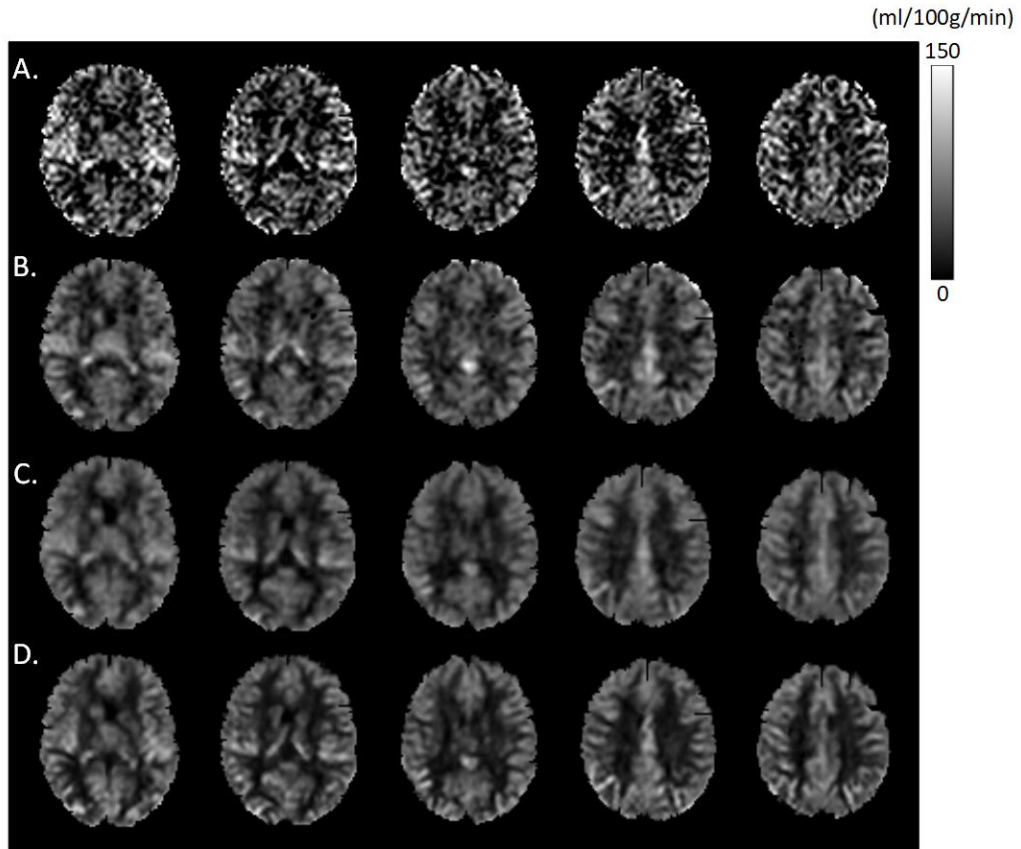


Figure 3.3: Mean CBF images of a representative subject. The rows from top to bottom are: A. mean CBF maps generated from 10 L/C pairs (input to ASLDN-LFN); B. mean CBF maps from all 40 L/C pairs with smoothing and outlier cleaning (pseudo-groundtruth); C. output of ASLDN; and D. output of ASLDN-LFN. Only 5 axial slices were shown in each row.

Figure 3.3 shows the mean CBF maps produced by different algorithms. Compared to pseudo-groundtruth (Figure 3.3.B.) and the output of ASLDN (Figure 3.3.C.), the CBF maps produced by ASLDN-LFN (Figure 3.3.D.) showed substantially improved quality in terms of suppressed noise and better perfusion contrast between tissues. Moreover, ASLDN-LFN recovered CBF signals in the air-brain boundaries and reduced partial volume effects.

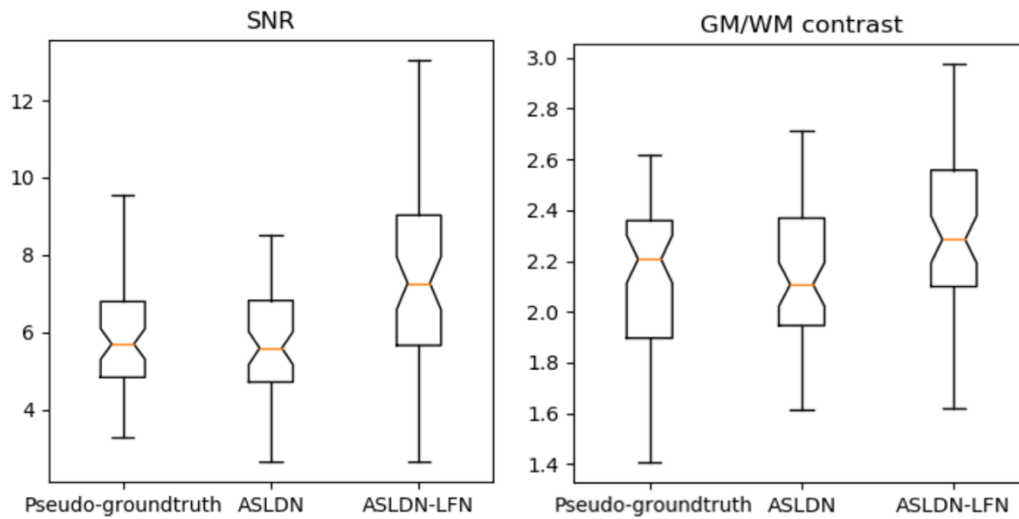


Figure 3.4: The notched box plot of the SNR (left) and GM/WM contrast (right) from 60 test subjects' CBF maps with different processing methods.

Figure 3.4 shows the notched box plot of the SNR and GM/WM contrast from 60 test subjects' mean CBF maps processed with different methods. The average SNR of pseudo-groundtruth, the output of ASLDN and the output of ASLDN-LFN were 5.87, 6.36 and 8.06 respectively. The average GM/WM contrast of pseudo-groundtruth, the output of ASLDN and the output of ASLDN-LFN were 2.14, 2.15 and 2.32. ASLDN-LFN improved SNR by 26.7% and improved GM/WM contrast by 7.9% compared with ASLDN.

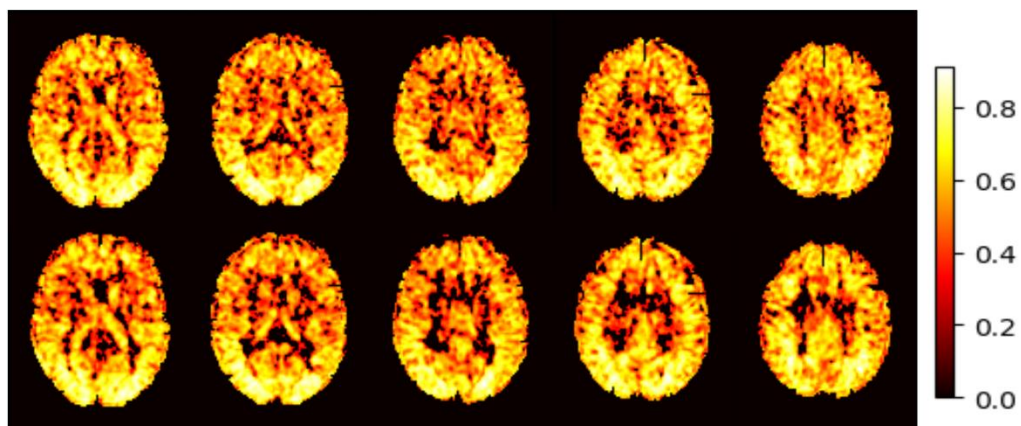


Figure 3.5: Correlation coefficient maps of ASLDN (top) and ASLDN-LFN (bottom). Only 5 axial slices were shown. Correlation coefficients less than 0.3 were thresholded to be 0.

Figure 3.5 shows the correlation coefficient maps of ASLDN and ASLDN-LFN. Correlation coefficient at each voxel was calculated between the pseudo-groundtruth and network output. Outputs of ASLDN and ASLDN-LFN strongly correlated to the pseudo-groundtruth, proving that both networks can preserve individual subjects' CBF patterns while suppressing noise. Output of ASLDN-LFN showed less correlation to input in WM because ASLDN-LFN removed more noises in WM than ASLDN.

Table 3.1: The average PSNR and SSIM of mean CBF maps produced by different CNN architectures in different training schemes.

	<b>ASLDN</b>			<b>ASLDN Learning-from-noise</b>		
<b>Model</b>	U-Net	DilatedNet	DWAN	U-Net	DilatedNet	DWAN
<b>PSNR</b>	24.53	24.92	<b>25.26</b>	24.84	25.06	<b>25.28</b>
<b>SSIM</b>	0.796	0.793	<b>0.803</b>	0.798	0.797	<b>0.803</b>

Table 3.1 lists the PSNR and SSIM performance of ASLDN [74] and the proposed ASLDN-LFN with or without using the DWAN network structure. ASLDN-LFN showed higher PSNR and SSIM than previous ASLDN. Using DWAN in ASLDN and ASLDN-LFN provided higher PSNR and SSIM than without.

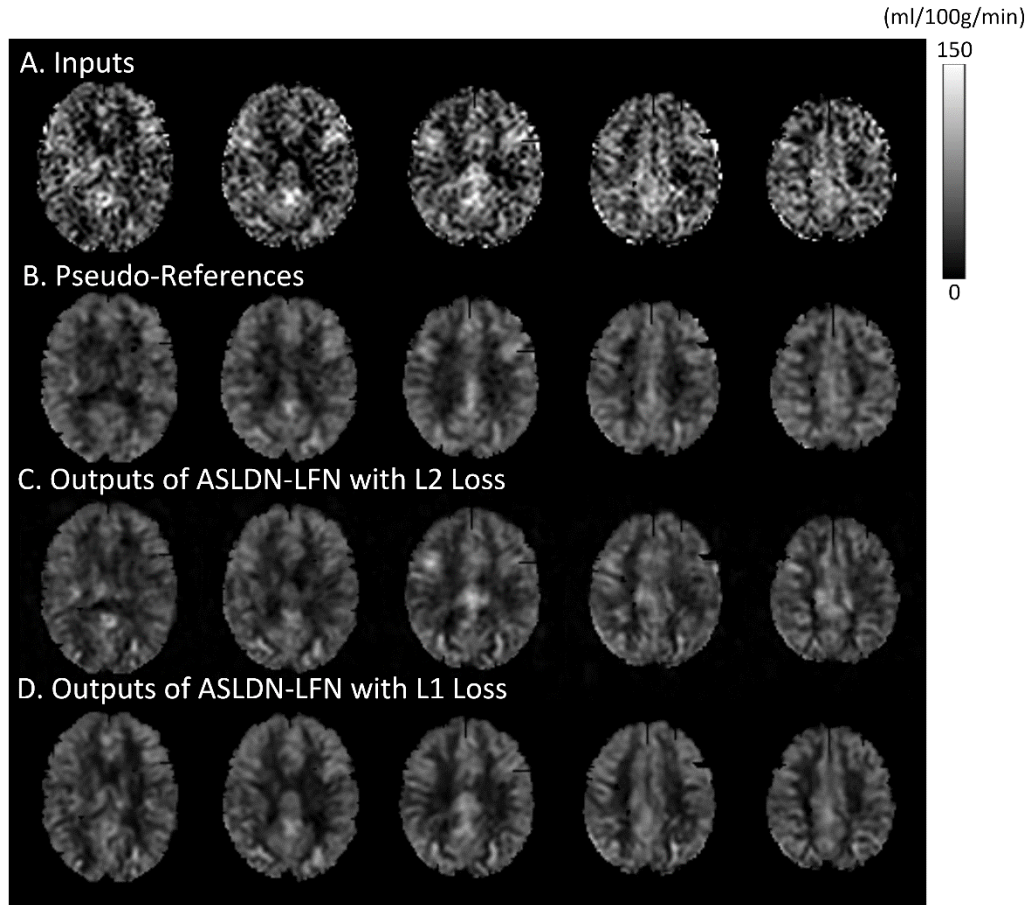


Figure 3.6: Mean CBF maps of a representative subject (Only 3 axial slices were shown). From left to right: input to ASLDN-LFN (column A); pseudo gold standard (column B); outputs of ASLDN-LFN trained with L2 loss (column C); and outputs of ASLDN-LFN trained with L1 loss (column D)

Figure 3.6 shows the result of ASLDN-LFN that was trained with different loss functions. When input mean CBF maps contained large amounts of outliers, ASLDN-LFN trained with L2 loss was affected, resulting in deteriorated perfusion in grey matter area. ASLDN-LFN trained with L1 loss, in contrast, remains unaffected due to its robustness to outliers. PSNR and SSIM are used to quantitatively measure the denoising performance of ASLDN-LFN trained with L1 loss and L2 loss. PSNR and SSIM were 24.40 and 0.677 when ASLDN-LFN was trained with L2 loss, whereas PSNR and SSIM were 25.28 and 0.803 when ASLDN-LFN was trained with L1 loss.

### 3.4 Application on ASL perfusion prediction from BOLD fMRI

BOLD fMRI is more widely used fMRI technique than ASL. It offers high temporal and spatial resolution, but only provides relative values, is sensitive to low-frequency drift, and suffers from the susceptibility gradient-induced artifacts. By contrast, ASL MRI measures cerebral blood flow (CBF) in a physical unit of ml/100 g/min. The quantitative nature of ASL MRI makes it insensitive to low-frequency drift. Measuring signal from the capillary bed, ASL MRI is potentially more accurate for localizing functional activation than BOLD fMRI which is often mainly contributed by oxygen level change in venous vessels rather than the activation site. However, ASL MRI has lower signal-to-noise-ratio (SNR), lower temporal resolution, and has only seen increasing visibility in recent years. An important but still open question is how to fuse the benefits provided by the two complementary functional imaging modalities. Since many finished and ongoing large size fMRI projects only had or have BOLD fMRI, a related question is whether we can reliably extract CBF signal from BOLD fMRI. Solving both questions requires understanding the elusive relationship between these two imaging modalities.

Due to the limitation of the longitudinal relaxation rate ( $T_1$ ) of blood water and the post-labeling transmit process, a small portion of tissue water can be labeled, resulting in a low SNR [31]. Thus, ASL often acquires many pairs of L/C images to improve the SNR of the mean perfusion map. Practically, 10-50 L/C pairs are allowed in a typical 3-6 min scan, which can only provide minor to moderate SNR improvement by averaging across the limited number of measurements. The interleaved labeling and non-labeling procedure reduce the temporal resolution of ASL MRI by half compared to the regular dynamic MR imaging. The relatively long labeling and post-labeling delay time before data acquisition

further reduce the temporal resolution of ASL MRI. Ideally, these drawbacks can be avoided instantly if CBF can be extracted from BOLD fMRI. From a technical point of view, ASL MRI can be acquired with many different imaging sequences. That is why the gradient-echo weighted BOLD imaging sequence is still widely used to acquire ASL MRI data. It is then theoretically reasonable to hypothesize that CBF can be extracted from the BOLD fMRI. The challenge is then to find an appropriate model for the unknown BOLD-CBF relationship.

A canonical BOLD-CBF model has been proposed in [77], but requires data acquired under gas-challenging. The underlying assumption of no change of cerebral metabolic rate of oxygen by gas-challenging may also be inaccurate. Without extra experiments, there isn't an analytic way to extract quantitative CBF from BOLD fMRI. Alternatively, a learning-based approach might be able to solve this problem. Over the years, machine learning, especially deep machine learning has been increasingly used to achieve astonishing success for modeling various highly complex data relationships [49].

Deep learning (DL) is motivated by the hierarchical learning in the visual system [47]. The most widely used deep neural networks consist of multiple layers of receptive field constrained local filters which are trained layer by layer by error backpropagations [34] and are often called convolutional neural networks (CNN). The local feature extraction, hierarchical abstraction, step-wise backpropagation of CNN and the introduction of several training strategies such as weight drop-out, batch-normalization, skip connection, and residual learning, etc., make CNN highly flexible and capable for modeling any nonlinear function buried in a large data. Because medical imaging processing is often hindered by some unknown nonlinear processes or transforms, DL may provide a potentially versatile

tool for medical imaging processing as increasingly demonstrated in a variety of applications, including image segmentation [42] and image reconstructions [53], etc. Specific to ASL MRI, DL has been adopted to improve SNR of ASL CBF maps [21] [74]. Most related to this study is that Xie et al. [78] piloted a pairwise label to control image prediction using CNN. Since the ASL MRI used in the so-called super-ASL network was acquired with the gradient-echo weighted BOLD fMRI sequence, it suggests the feasibility of directly extracting CBF from BOLD fMRI.

In this section, we want to build and validate a learning-from-noise DL-based BOLD-ASL relationship learning model to predict the CBF signal directly from BOLD fMRI. We dubbed the network as the BOA-Net. Different from the super-ASL work, we used current ASL MRI and BOLD fMRI acquired with a dual-echo ASL MRI sequence [79] so the network doesn't need to consider the physiological difference or signal drift-induced difference between the BOLD fMRI and ASL CBF.

### 3.4.1 Materials and Experiment Setup

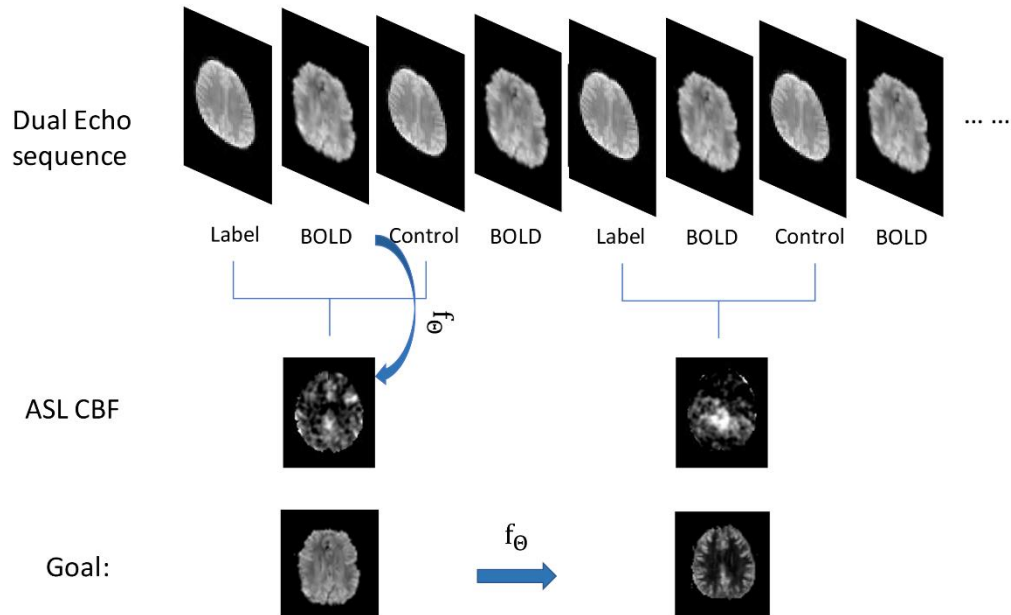


Figure 3.7: Illustration of dual-echo sequence and BOA-Net training process.

Figure 3.7 illustrates the dual-echo sequence and BOA-Net training process. Denote the CBF image generated by  $i$ -th L/C pair by  $\hat{y}_i$  and the BOLD image (the 2nd echo) after the  $i$ -th C image by  $\hat{x}_i$ . Given same brain structure and transitory acquisition time, we want to build a parametric regression model  $f_\theta$  that learns the mapping  $f_\theta(\hat{x}_i) = \hat{y}_i$ , where  $i = 1, 2, \dots, N$ , and  $N$  is the total number of one subject's CBF maps and  $y_i$  is the latent clean version of ASL CBF of  $\hat{y}_i$ .  $\theta$  are the parameters of the model and are adjusted through the training process. The model, typically a CNN, can be learned by minimizing the loss function:  $\sum_{i=1}^n L(f_\theta(\hat{x}_i) - \hat{y}_i)$ , where the loss function can be either the mean square error or mean absolute error between input and reference.

As we do not have gold standard CBF maps as the training references, using the low SNR ASL CBF images as the training references may result in an inaccurate BOA-Net. However, the learning-from-noise method showed that the inaccurate model training concern due to the use of noisy reference is not necessarily true. Therefore, we proposed a noisy reference-based BOA-Net. Instead of using the L2 norm as the loss function, we chose the L1 norm to reduce sensitivity to outliers which are common to ASL MRI [20].

We use DWAN, the same network architecture as shown in Figure 3.3, for BOA-Net. The two pathways of DWAN were used to extract both local and global contextual features. The wide activation residual blocks were adapted to expand data features and pass more information through the network [28]. ASL and BOLD fMRI data were acquired with the dual-echo ASL sequence [79] from 50 young healthy subjects at Hangzhou Normal University with signed informed written consent forms. The experiment and the form were applied by local IRB. Imaging parameters were: labeling time/delay time/TR/TE1/TE2 =

1.4s/1.5s/4.5s/11.7msec/68msec, 90 acquisitions (90 BOLD images and 45 C/L image pairs), FOV=22 cm, matrix=64×64, 16 slices with a thickness of 5 mm plus 1 mm gap. We used ASLtbx [65] to preprocess ASL images with the procedures in [20].

The BOA-Net was trained with data from 23 subjects' CBF maps (input and reference). 4 different subjects were used for validation samples. The remaining 23 subjects' CBF maps were used as test samples. For each subject, we extracted slices from 7 to 11 of 3D ASL CBF maps. The number of total 2D CBF maps extracted for training and validation were  $27 \times 5 \times 45 = 6075$ . The 2D CBF maps were  $64 \times 64$  pixels. U-Net [55] and DilatedNet [21], two popular CNN structure widely used in medical imaging, were implemented as a comparison to our DWAN-based BOA-Net.

We also compare the effects of training with smoothing CBFs versus non-smoothing CBFs. The CBFs that were generated from the L/C pairs with Gaussian smoothing were called smoothing CBFs. The CBFs that were generated from the L/C pairs without Gaussian smoothing were called non-smoothing CBFs. the suffix 'sm' and 'nsm' were added to the name of each model to represent that the model was trained using the smoothed or non-smoothed CBFs, respectively. We use Peak signal-to-noise ratio (PSNR) and structure similarity index (SSIM) to quantitatively compare the performance of DWAN with U-Net and DilatedNet. When calculating PSNR and SSIM, all the predicted results were compared with genuine mean CBF maps from smoothed ASL data.

All networks were implemented using the Keras and Tensorflow platform. The network was trained using the adaptive moment estimation (ADAM) algorithm with a basic learning rate of 0.001. All the models were trained with batches, each containing 64

training samples. All experiments were performed on a PC with Intel(R) Core(TM) i7-5820k CPU @3.30GHz and an Nvidia GeForce Titan Xp GPU.

We used SNR to measure the image quality of ASL CBF. The SNR was calculated by using the mean signal of a grey matter (GM) region-of-interest (ROI) divided by the standard deviation of a white matter (WM) ROI in slice 9. The similarity of mean CBF from the outputs of BOA-Net to genuine mean CBF maps from ASL data was evaluated by the correlation coefficient between the CBF values of all testing subjects (n=23). This process was performed at each voxel for BOA-Net\_sm and BOA-Net\_nsm separately. The correlation coefficient maps were thresholded by  $r > 0.3$  for comparison and display.

### 3.4.2 Experimental Results

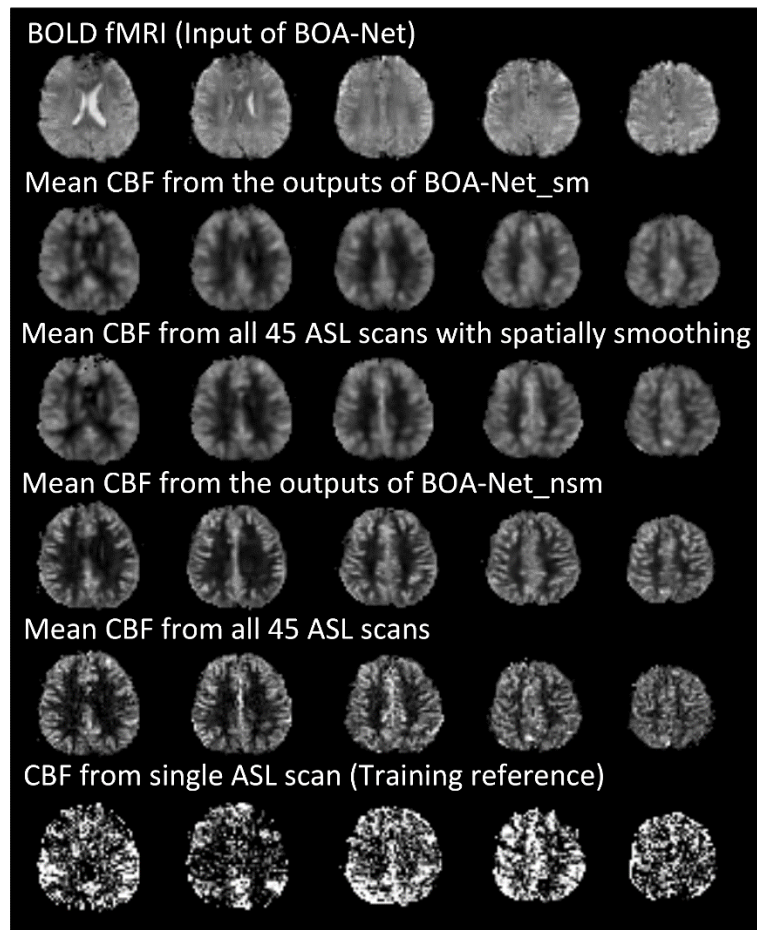


Figure 3.8: From top to bottom: BOLD fMRI and ASL CBF produced by different methods. From left to right: slice 7, 8, 9, 10, 11.

Figure 3.8 shows the results of BOLD-based CBF prediction for one representative subject. As compared to the genuine mean CBF map from the acquired ASL MRI, the CBF map produced by BOA-Net showed substantially improved quality in terms of suppressed noise and better perfusion contrast between tissues. Moreover, BOA-Net recovered CBF signals in the air-brain boundaries. Signal loss in the genuine mean CBF in the prefrontal region was caused by the signal loss in BOLD images.

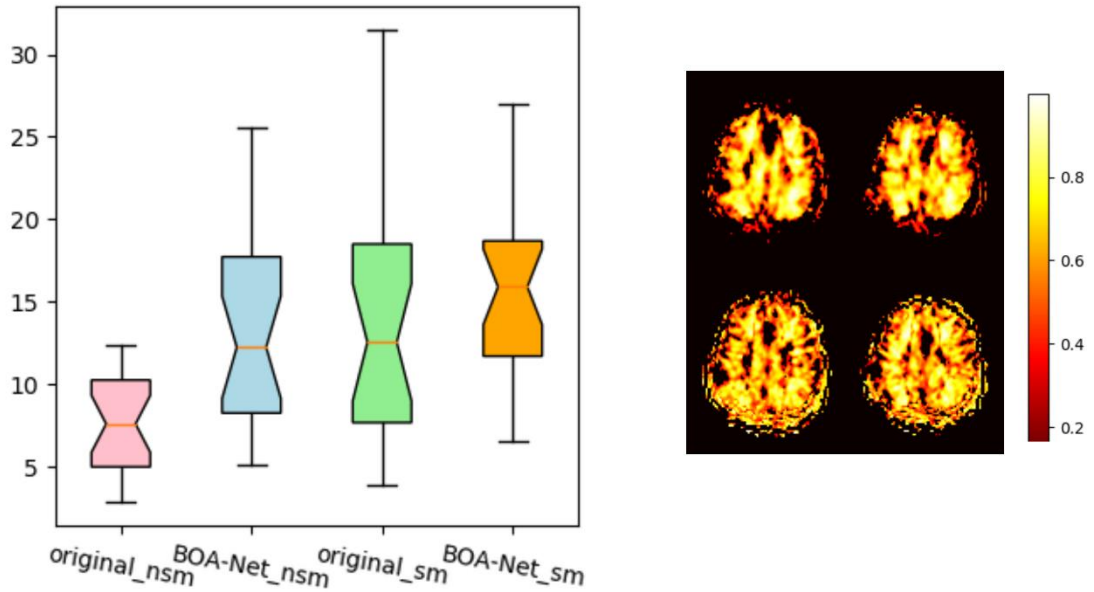


Figure 3.9: The notched box plot of the SNR (left) and correlation coefficient maps between genuine mean CBF and output of BOA-Net (right). Original\_nsm and original\_sm represent the genuine mean CBF maps from non-smoothed and smoothed ASL data. BOA-Net\_nsm and BOA-Net\_sm represent mean CBF maps from outputs of BOA-Net\_nsm and BOA-Net\_sm. The correlation coefficient maps between genuine mean CBF and output of BOA-Net\_sm is shown in the top row. The correlation coefficient maps between genuine mean CBF and output of BOA-Net\_nsm is shown in the bottom row. Only 2 axial slices were shown. Correlation coefficients less than 0.3 were thresholded to be 0.

Figure 3.9 shows the box plot of the SNR and spatial correlation of BOA-Net\_sm and BOA-Net\_nsm. The average SNR of genuine mean CBF maps from non-smoothed

and smoothed ASL data were 6.96 and 12.64 respectively. The average SNR of mean CBF maps from outputs of BOA-Net\_nsm and BOA-Net\_sm were 12.26 and 15.11. BOA-Net\_sm improved SNR by 19.54% compared with mean CBF maps of smoothed ASL while BOA-Net\_nsm achieved a 76.15% SNR improvement compared with the mean CBF maps of non-smoothed ASL. Correlation coefficient at each voxel was calculated between the genuine mean CBF map and network output. Figure 3.9 shows outputs of BOA-Net\_sm and BOA-Net\_nsm strongly correlated to the genuine mean CBF, proving that both networks can predict individual subjects' CBF patterns correctly.

Table 3.2: The average PSNR and SSIM from different CNN architectures used in BOA-Net\_sm and BOA-Net\_nsm

	<b>BOA-Net_sm</b>			<b>BOA-Net_nsm</b>		
<b>Model</b>	U-Net	DilatedNet	DWAN	U-Net	DilatedNet	DWAN
<b>PSNR</b>	24.15	24.10	<b>25.58</b>	21.88	21.51	<b>22.49</b>
<b>SSIM</b>	0.884	0.892	<b>0.893</b>	0.856	0.841	<b>0.865</b>

Table 3.2 shows the PSNR and SSIM of mean CBF maps predicted from different models. DWAN achieved the highest PSNR and SSIM in both BOA-Net\_sm and BOA-Net\_nsm categories. Figure 3.10 demonstrates the visual comparison of mean CBF maps predicted from BOLD fMRI using different CNN architectures. DWAN suppressed more noises than DilatedNet while recovered more details than U-Net. Moreover, DWAN\_nsm has better perfusion contrast than DWAN\_sm while DWAN\_sm recover more signals in air-brain boundaries.

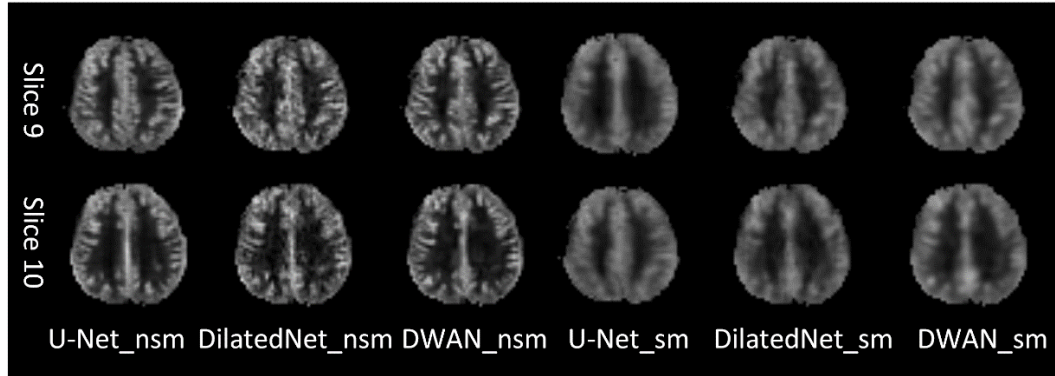


Figure 3.10: Two representative slices of the mean CBF maps produced by different processing methods. The three columns on the left side are mean CBF maps of outputs of U-Net\_nsm, DilatedNet\_nsm, and DWAN\_nsm respectively. The three columns on the right side are mean CBF maps of outputs of U-Net\_sm, DilatedNet\_sm, and DWAN\_sm. The image display window is 0-100 ml/100g/min.

### 3.5 Discussion and Conclusion

In this chapter, we proposed to apply the learning-from-noise method on DL-based ASL denoising models (ASLDN-LFN) and ASL perfusion prediction from BOLD fMRI (BOA-Net). In section 3.3, we showed that DL-based ASL denoising models can be trained using only noisy image pairs. The experimental results demonstrated that ASLDN-LFN can reliably denoise ASL CBF and even achieved improved image quality than ASLDN in terms of SNR and GM/WM contrast. Moreover, by using ASLDN-LFN, more training data can be generated as it requires less L/C pairs to generate reference mean CBF maps, which is particularly useful when ASL CBF data are limited.

In section 3.4, we showed that BOA-Net can reliably extract ASL perfusion maps from BOLD fMRI. This study represents the first effort to extract quantitative CBF from BOLD fMRI. Comparing with genuine mean CBF from ASL data, the BOA-Net can provide CBF measurement with higher SNR and higher temporal resolution, both inherited from BOLD fMRI (higher SNR is also contributed by DL denoising). For the existing

dataset without ASL MRI acquired, this provides a unique opportunity to generate a new functional imaging modality. For future studies, it offers an opportunity to avoid ASL MRI scans though that will need more evaluations, especially in diseased populations. Even if an ASL MRI scan is still needed, its scan time can be substantially shortened, and the reduced SNR can be compensated by CBF estimated from BOA-Net.

## CHAPTER 4

# DENOISING ARTERIAL SPIN LABELING PERFUSION MRI WITH DILATED WIDE ACTIVATION NETWORK

Current network architecture for DL-based ASL MRI denoising methods, such as residual network [22] and Dilated Network [21], were not optimized for image denoising. These networks were originally proposed for high-level vision tasks such as image segmentation and classification [24] [25] [26]. Applying these architectures directly to low-level vision tasks such as denoising and super-resolution can be suboptimal [27] [28]. In this chapter, we proposed a novel DL-based method for ASL MRI denoising (dubbed ASLDN). Our ASLDN presents two novelties: first, we proposed a DL model that is optimized for ASL denoising. Incorporating wide activation residual blocks [28] with a Dilated Convolution Neural Network (DilatedNet) [25], our dilated wide activation network (DWAN) achieved improved denoising performance both quantitatively and qualitatively while reduced ASL acquisition time. Second, we trained and evaluated our proposed model on large *in vivo* dataset, showing that our model can be generalized to inputs with different levels of SNR and yields images with better quality than other methods.

### 4.1 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. 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  $n_i$ :  $x_i = y_i + n_i$ . ASLDN is to learn an

implicit transform which can predict  $y_i$  from  $x_i$ , so that  $f_{\theta}(x_i) \rightarrow y_i$ . In ASLDN, we used mean CBF map obtained without using any explicit denoising as the input and the image denoised with the current state-of-art outlier cleaning methods [20] as the reference  $y_i$ . To avoid potential overfitting due to the high similarity between  $x_i$  and  $y_i$ , ASLDN is configured to learn their difference  $n_i$  following the residual learning strategy [80] [59] [28] [25]. Correspondingly, the transform to learn became  $g(x_i) = y_i - x_i = -n_i$ . During testing or new image denoising, the output of ASLDN 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. As MAE is more robust to outliers, the loss function for ASLDN can be defined as:

$$L(\theta) = \frac{1}{2N} \sum_{i=1}^N |y_i + g(y_i; \theta) - x_i|. \quad (4.1)$$

## 4.2 Materials and Methods

### 4.2.1 Network Architecture

We propose a Dilated Wide Activation Network (DWAN) that has optimized architecture for low-level vision tasks such as image denoising and super-resolution. As DWAN combines wide activation residual block, dilated convolution, and residual learning, we first introduce each module separately and then overview the whole architecture of the DWAN.

**Wide Activation Residual Block:** as shown in Figure 4.1, a regular residual block has a convolution layer with a ReLU activation layer, followed by another convolution layer. Each convolution layer has the same number of convolution filters inside a regular

residual block. The direct connection from input to output represents the residual learning [26], a key element of modern CNNs to improve accuracy and training speed [80] [59] [28] [26]. Since the ReLU layer inside the residual block may cause information loss before being passed to the next layer [28], the so-called wide activation residual block uses more convolution filters to extract more features in the layer preceding the ReLU layer, which can subsequently pass more information to the next block [28]. For denoising, this modification would preserve more high-resolution information to reduce the blurring effects as supported by the results in image super-resolution [28] and image restoration [73].

There are two hyper-parameters that define the wide activation residual block: the number of base convolution filters  $M$  and the expansion rate  $E$ . The number of convolution filters before the ReLU activation layer is  $E \times M$ . After the ReLU activation layer, the following convolution layer shrinks the number of convolution filters back to  $M$ . In comparison, a regular residual block has  $E \times M$  filters in the both convolution layers. According to [28] and our experiments, wide activation residual blocks not only reduce the complexity of the whole CNN architecture but also stabilize the training process. A CNN architecture with regular residual blocks often cause unstable training process in denoising and super-resolution tasks when the hyper-parameter  $E \times M$  is large.

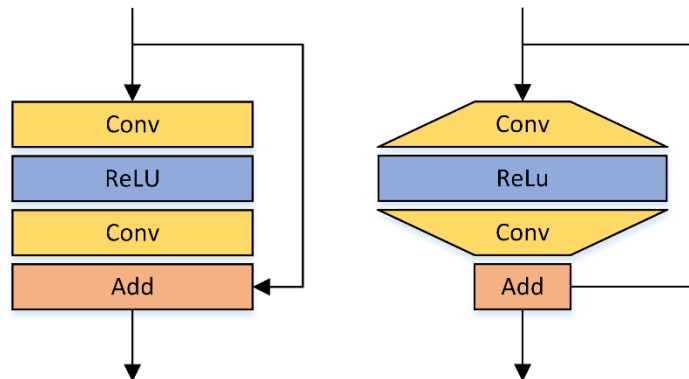


Figure 4.1: Schematic illustration of a regular residual block (left) and a wide activation residual block (right).

**Dilated Convolution:** Figure 4.2 describes the dilated convolution in detail. The main motivation of using dilated convolutions is that dilated convolutions support expanding receptive fields exponentially without losing resolution or coverage. We use dilated convolution to extract global features more effectively.

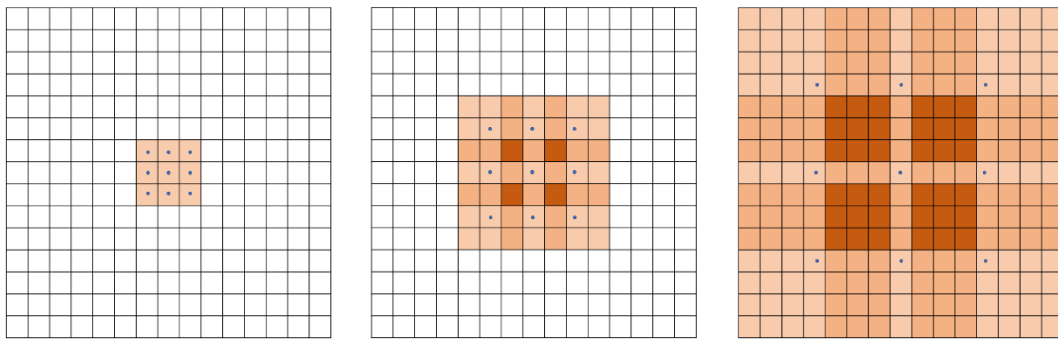


Figure 4.2: Schematic illustration of a series of dilated convolutions. A series of dilation convolution supports exponential expansion of the receptive field without loss of resolution or coverage. On the left: a  $3 \times 3$  convolution filter without dilation produce a receptive field of  $3 \times 3$  as each element (blue dot) is a  $1 \times 1$  convolution. On the middle: a 2-dilated  $3 \times 3$  convolution filter in the following convolution layer produce a receptive field of  $7 \times 7$  as each element (blue dot) is a result of a  $3 \times 3$  convolution from the previous layer. On the right, a 4-dilated  $3 \times 3$  convolution filter in the next convolution layer produce a receptive field of  $15 \times 15$  as each element (blue dot) is a result of a  $7 \times 7$  convolution from the previous layer. The size of convolution filter ( $3 \times 3$ ) in each layer is identical but the receptive field grows exponentially.

**Residual Learning and Batch Normalization:** ASLDN is configured to learn the residuals between the input and the reference. The residual learning is implemented through a  $3 \times 3$  convolution from the input layer to the output layer. Though Kim et al. [21] and Gong et al. [56] used Batch Normalization (BN) [61] layer in their network, recent studies [53] [27] [81] [82] found that BN 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.

**Dilated Wide Activation Network:** Figure 4.3 describe the proposed DWAN that is used to build our ASLDN network in detail. As shown in Figure 4.3, DWAN has a local pathway and a global pathway. The difference between the local pathway and global pathway is that the first convolution layer of the 4 wide activation residual blocks in the global pathway used a dilation rate of 2, 4, 8 and 16 respectively. The local pathway extracts the local features and the global pathway uses dilation convolutions to reserve global data patterns. Furthermore, the wide activation residual blocks in DWAN are able to expand data features and pass more information through the network, improving performances for low-level computer vision tasks without additional parameters and computation [28] [73].

#### 4.2.2 Image Acquisition and Preprocessing

ASL data were pooled from 280 subjects (normal healthy subjects with age from 23-47) in a local database. The data were acquired in a Siemens Trio 3T scanner using a pseudo-continuous ASL sequence [83] [84] (40 control/labeled image pairs with labeling time = 1.5 sec, post-labeling delay = 1.5 sec, FOV=22 cm, matrix=64×64, TR/TE=4000/11ms, 20 slices with a thickness of 5 mm plus 1 mm gap).

ASL images were preprocessed using the procedures in section 3.3.1. Mean CBF maps were calculated from a 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”.

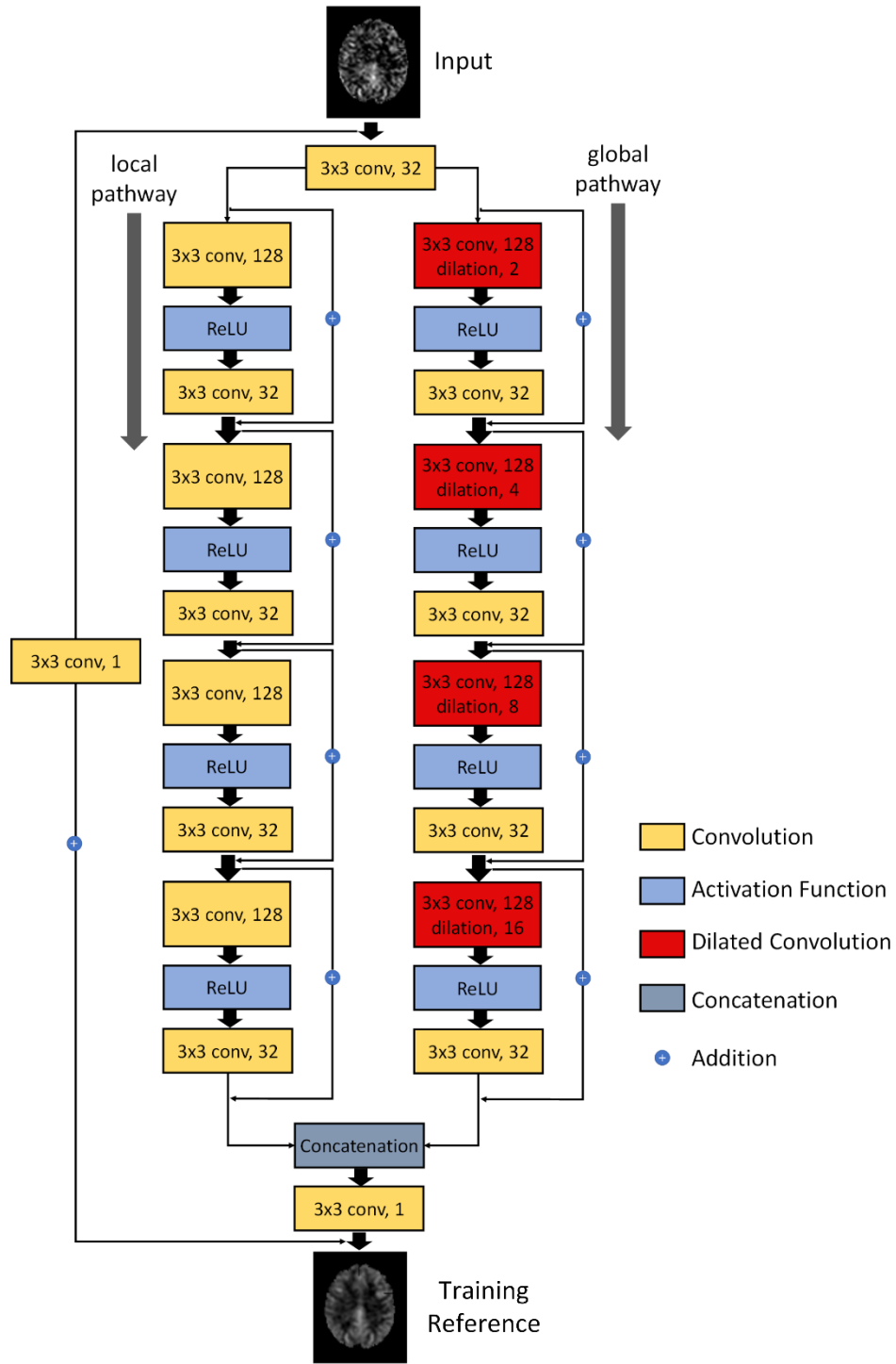


Figure 4.3: Schematic illustration of the architecture of our proposed DWAN network. The first layer consists of  $3 \times 3 \times 32$  convolutional filters for the input image. Then the output of

the first layer was fed to both local pathway and global pathway. Each pathway contains 4 consecutive wide activation residual blocks. Each wide activation residual block contains 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 the global pathway were dilated convolutional layers with a dilation rate of 2, 4, 8, 16, respectively. The output of the 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).

#### 4.2.3 Data Preparation and Model Training

To maximally demonstrate the benefit of ASLDN, 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 ASLDN models. As there is no groundtruth (GT) to ASL CBF maps, meanCBF-40 -- mean CBF map of the entire 40 C/L image pairs, was used as a surrogate GT (the reference) for training the ASLDN models. The same model was also trained for projecting the meanCBF-10\_nsm to meanCBF-40\_nsm and the 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 one out of every three axial 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  $91 \times 109$  pixels.

U-Net [24], a popular CNN structure widely used in medical imaging, was implemented as a comparison to our DWAN-based ASLDN. Because U-Net has pooling and upsampling layers that require the image size equal to the power of 2, all the images were resized to  $128 \times 128$  during the training and testing of U-Net, by centering images

and padding 0s on each side. U-Net, U-Net with BN (U-Net\_BN) [55], DilatedNet and DilatedNet with BN (DilatedNet\_BN) [21] were implemented as additional comparisons. To test the denoising capability of ASLDN 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 [75]. The network was trained using the Adaptive moment estimation (ADAM) algorithm [68] with a basic learning rate of 0.001. All the models were trained with batches, each containing 64 training samples. DWAN was trained for 10000 iterations and DWAN\_nsm was trained for 5000 iterations. Tensorflow [76] 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 an Nvidia GeForce Titan Xp GPU.

#### 4.2.4 Effects of Different Model Configurations

We conducted an experiment to assess the performance of ASLDNs that contain a different number of wide activation residual blocks. A grid search experiment was also conducted to find the best combination of expansion rate  $E$  and the number of input filters  $M$  for each wide activation residual block. Noted that the regular residual block [26] [27] is essentially a special case of wide activation residual block [28] when the expansion rate  $E$  is 1. Thus, by setting the 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 L1/L2 loss function used in training.

#### 4.2.5 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 (Tianyao Wang and Fuqing Zhou). 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.

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 ASLDN) and the surrogate GT (meanCBF-40\_sm). Higher PSNR means less voxel-wise differences.

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  $\sigma_{xy}$  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 ASLDN was to provide better quality in relative to the mean CBF map processed with current standard steps.

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  $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ , and  $\sigma_y$  are the mean and variance of the output image and the surrogate GT, respectively.  $\sigma_{xy}$  is the covariance between the output image and the surrogate GT. The CCC assesses 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.

### 4.3. Experimental Results

#### 4.3.1 Results of ASLDN For Projecting meanCBF-10 to meanCBF-40

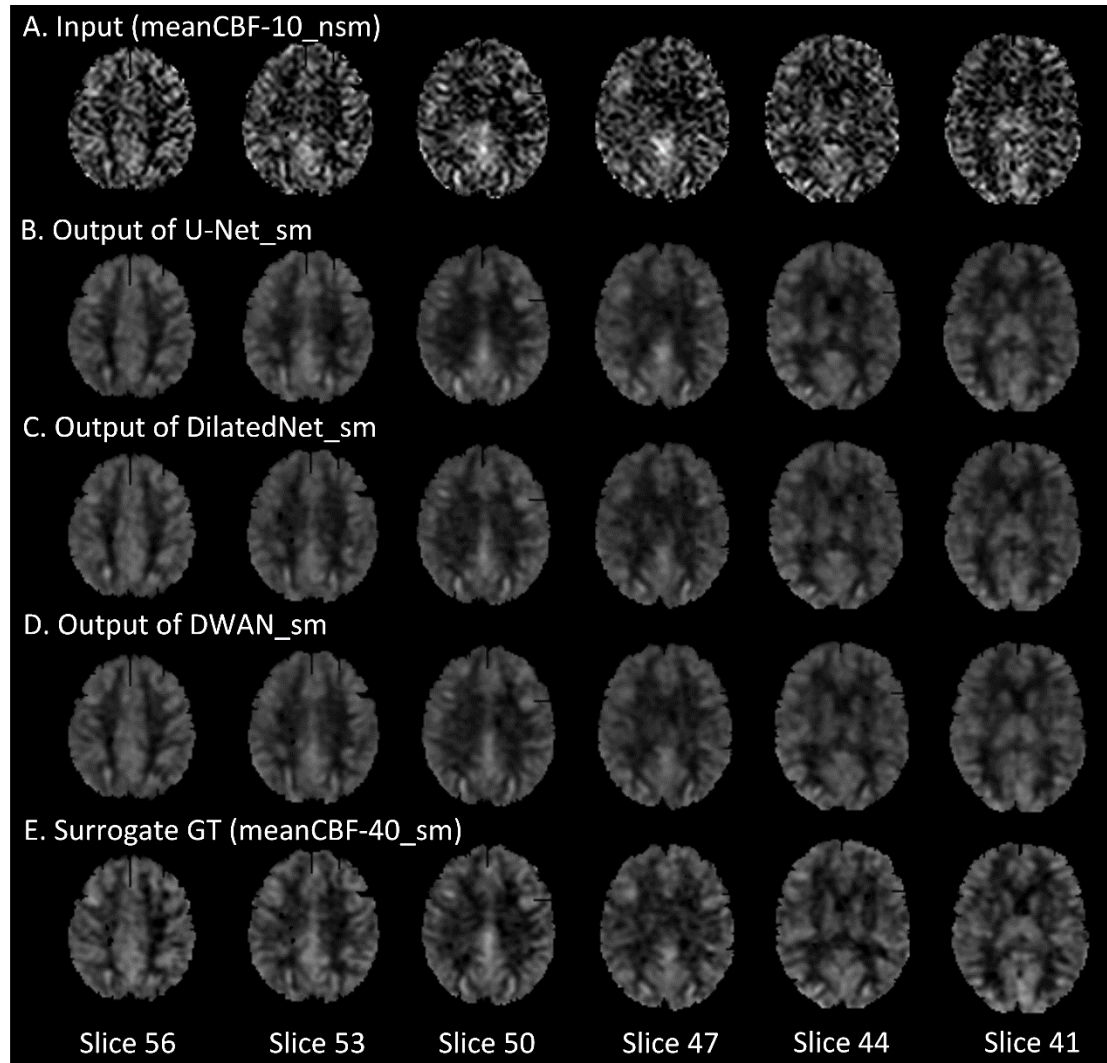


Figure 4.4: Mean CBF images (only 5 axial slices were shown) from a representative subject with different processing methods. From top to bottom: meanCBF-10\_nsm as input, the output of U-Net\_sm, the output of DilatedNet\_sm, the output of proposed DWAN\_sm and the surrogate GT (meanCBF-40\_sm) as training reference. Image display window was 0-200 ml/100g/min.

Figure 4.4 shows a representative subject's mean CBF maps produced by different denoising methods. As compared with meanCBF-10\_nsm (Figure 4.4 A, the input to ASLDN), all DL methods (Figure 4.4 B-D) produced much better image quality in terms of improved tissue perfusion signal, suppressed noise (especially in white matter), and better perfusion contrast between grey matter and white matter. The DL methods even

showed CBF map quality improvement as compared with the surrogate GT (Figure 4.4 E, meanCBF-40\_sm). DWAN\_sm (Figure 4.4 D) showed the best quality as compared against other DL methods.

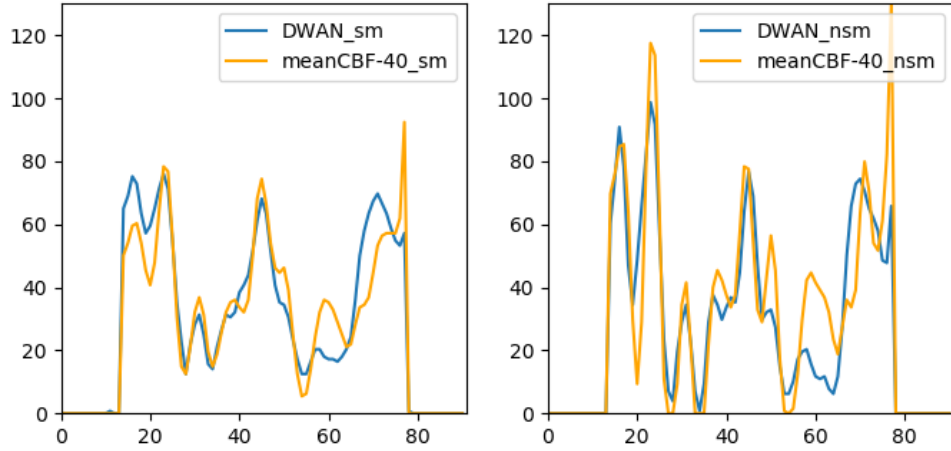


Figure 4.5: CBF image profiles are taken from the 50th row (marked as red line) of the 54th axial slice from one representative subject.

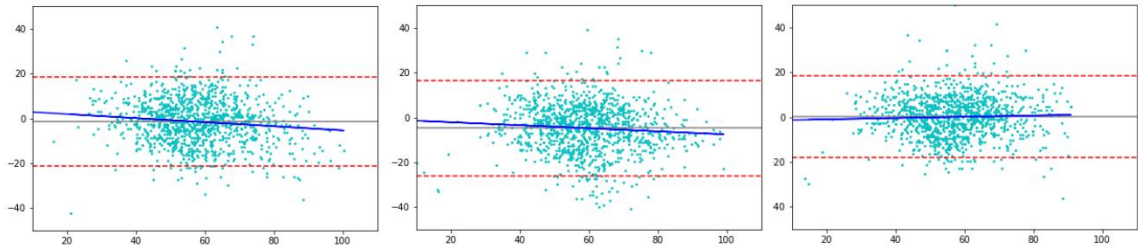


Figure 4.6: Bland-Altman plots of different methods obtained in the GM area from one representative subject. Y-axis shows the differences in CBF values between the surrogate GT and the compared method and X-axis shows the mean CBF values of the two. The unit for x and y axes are in ml/100g/min. Solid grey lines indicate mean difference. Dashed red lines at top and bottom correspond to upper and lower margins of 95% limits of agreement. blue solid lines are linear regression lines.

Figure 4.5 shows the image profile of CBF images processed with different methods. DL-denoised CBF maps didn't show observable tissue boundary damages, indicating the preservation of structural details and spatial resolution of DL-based

denoising. Figure 4.6 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 (grey solid line) and variance (difference between dashed red lines). The linear regression line (solid blue) 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 to excessive noises and artifacts from L/C subtractions..

Table 4.1: 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	<b>26.25</b>
SSIM	0.813	0.806	0.819	0.817	<b>0.822</b>

Table 4.2: The average PSNR and SSIM over 60 test subjects. All the models were trained with meanCBF-40\_nsm as reference images.

Model	U-Net	U-Net_BN	DilatedNet	DilatedNet_BN	DWAN
PSNR	24.53	24.16	24.78	24.60	<b>24.81</b>
SSIM	0.796	0.787	0.802	0.798	<b>0.803</b>

Table 4.1 and

Table 4.2 list the performance indices of different methods. As compared with other methods, DWAN\_sm and DWAN\_nsm presented better performance indices (PSNR and SSIM) (paired t-test,  $P < 0.05$ ). Our experiments also confirmed that adding BN layers did

not improve denoising performance as described in [53] [27] [81] [82] . Table 4.3 list the MAE and CCC of different methods. DWAN\_sm and DWAN\_nsm has lower MAE and higher CCC than other methods.

Table 4.3: The average MAE and CCC over 60 test subjects.

Model	U-Net_sm	DilatedNet_sm	DWAN_sm
MAE	6.017	5.936	<b>5.926</b>
CCC	0.896	0.898	<b>0.900</b>
Model	U-Net_nsm	DilatedNet_nsm	DWAN_nsm
MAE	6.285	6.289	<b>6.281</b>
CCC	0.854	0.857	<b>0.859</b>

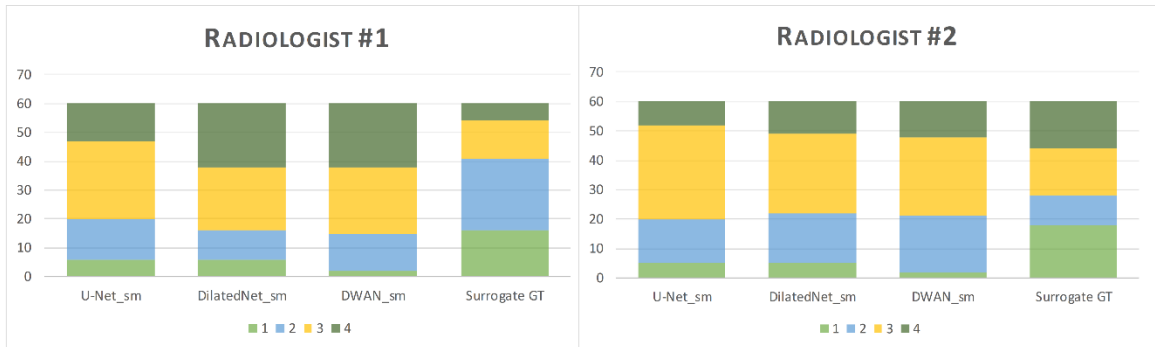


Figure 4.7: Comparison of Radiologic score between different methods over 60 subjects' CBF maps. Radiologic scores are displayed in four different colors. The vertical axis indicates the number of subjects of each radiologic score.

Figure 4.7 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 when rated by radiologist #1 (Tianyao Wang), and an average score of 2.81, 2.73 and 2.72 when rated 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 figure

12 shows, DWAN\_sm consistently achieves better denoising results than other methods qualitatively. Figure 4.8 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.

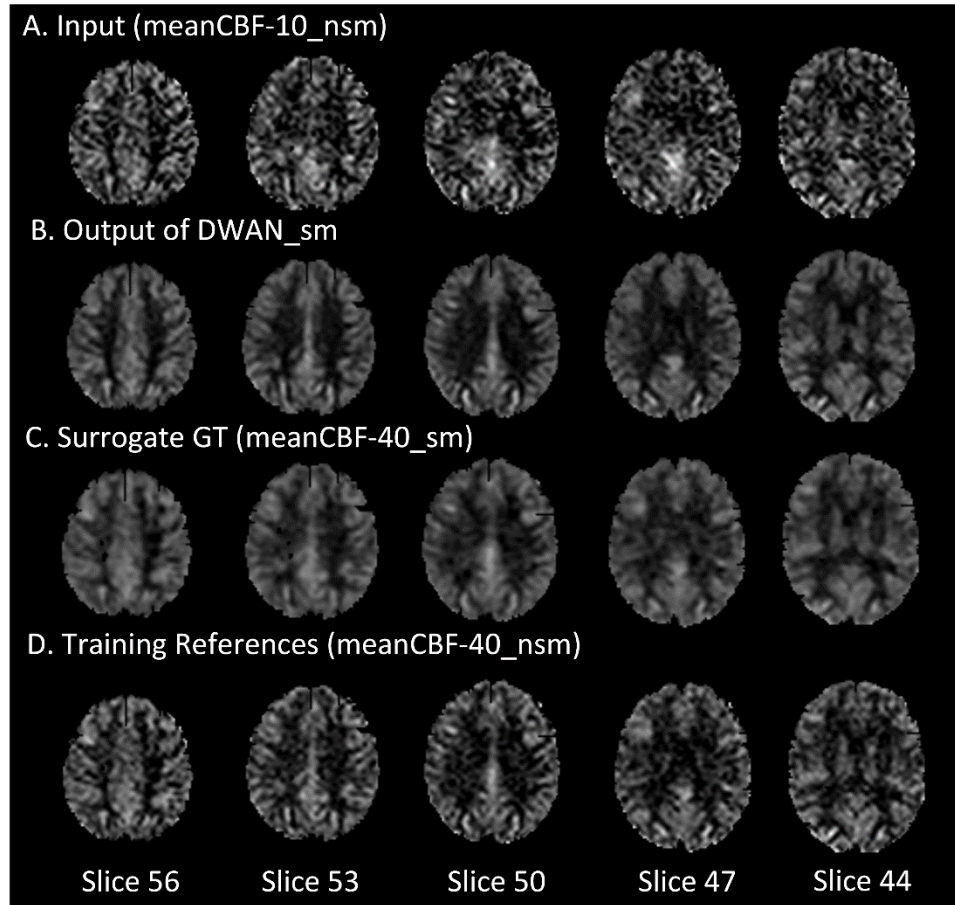


Figure 4.8: ASL CBF images with meanCBF-10\_nsm as input. From top to bottom: the input (meanCBF-10\_nsm), the output of DWAN\_nsm, the surrogate GT (meanCBF-40\_sm) for comparison and the training reference (meanCBF-40\_nsm). The image display window was 0-200 ml/100g/min.

#### 4.3.2 Effects of Different Noise Levels 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.

Figure 4.10 shows the quantitative results in terms of PSNR and SSIM. When the input is meanCBF-10\_nsm to meanCBF-30\_nsm, DWAN\_sm showed higher PSNR and SSIM than U-Net\_sm and DilatedNet\_sm. When the input is meanCBF-35\_nsm or meanCBF-40\_nsm, DWAN\_nsm had higher PSNR and SSIM than all other models.

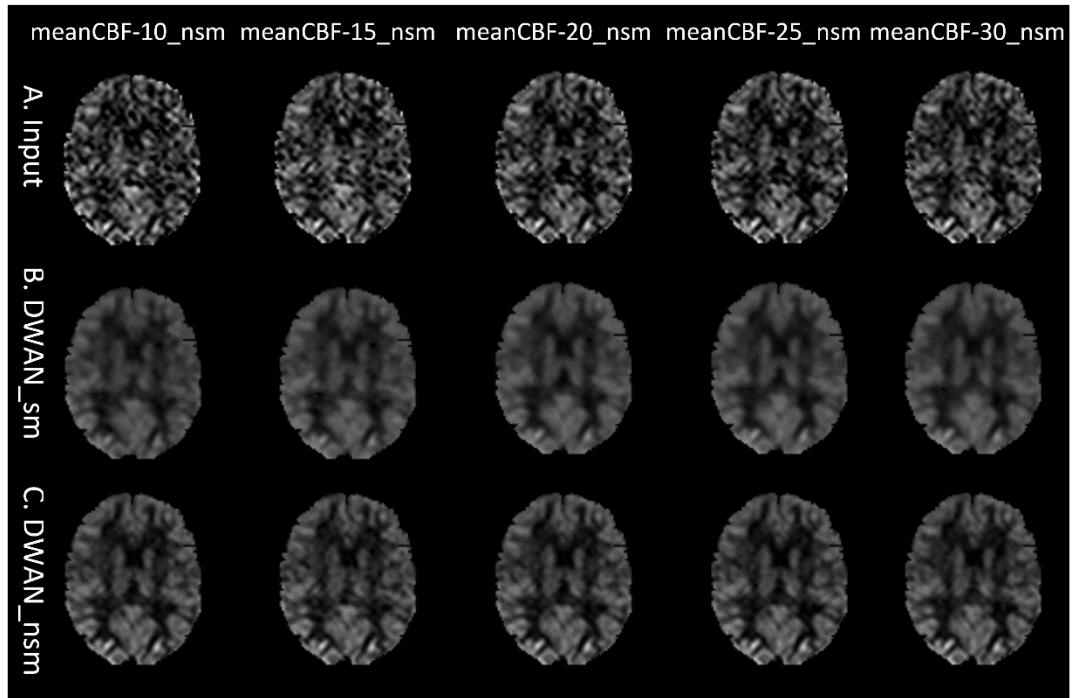


Figure 4.9: The same slice of a representative subject's mean CBF images (only 1 axial slice was shown) processed using the same DWAN\_sm or DWAN\_nsm model trained with the (meanCBF-10\_nsm, meanCBF-40\_sm) or (meanCBF-10\_nsm, meanCBF-40\_nsm) image pairs, respectively. A) input images to the model, 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.

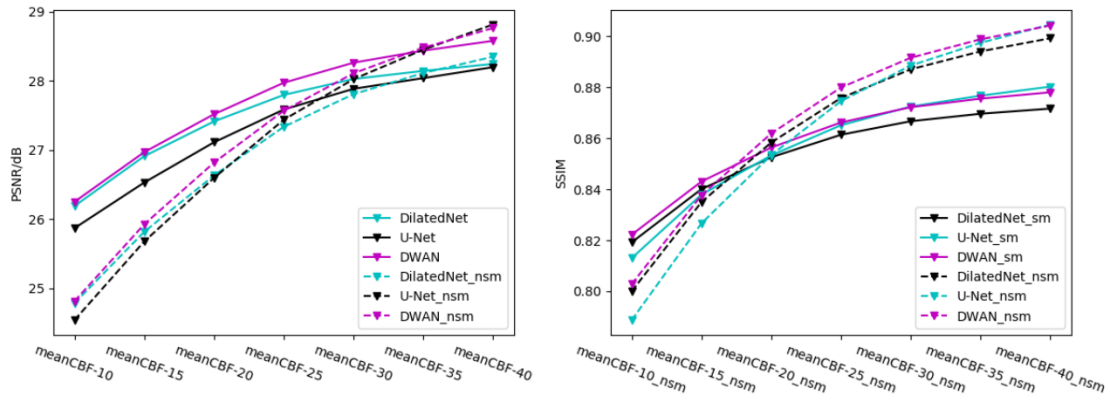


Figure 4.10: PSNR and SSIM of U-Net\_sm, U-Net\_nsm, DilatedNet\_sm, DilatedNet\_nsm, DWAN\_sm and DWAN\_nsm over different mean CBF test datasets.

Figure 4.9 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 Figure 4.10.

#### 4.3.3 Effects of Model Configurations, Residual Learning, and Loss Function

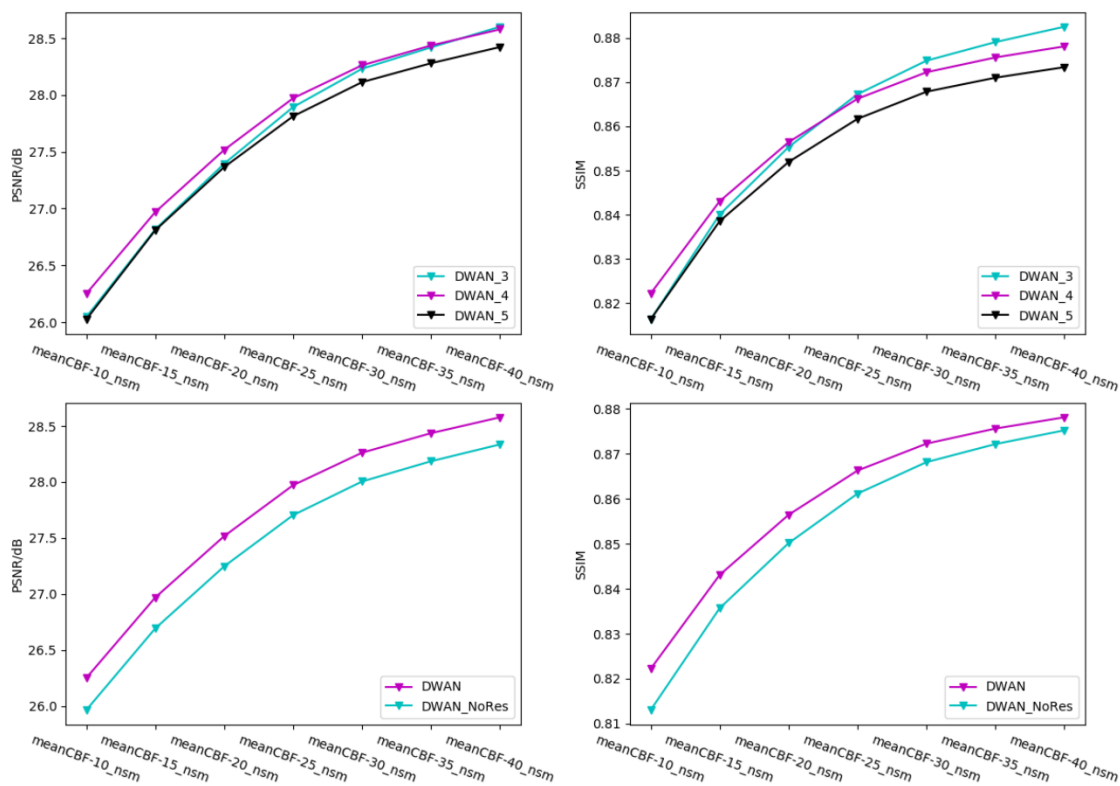


Figure 4.11: Top: performance of DWAN with a different number of wide activation residual blocks. The number in the suffix of the model represents the number of wide activation residual blocks in each pathway. Bottom: performance of DWAN with and without residual learning. DWAN\_NoRes has the same architecture as DWAN except residual learning.

Figure 4.11 shows the performance of DWAN<sub>sm</sub> with a different number of wide activation residual blocks. In general, DWAN<sub>sm</sub> with 4 wide activation residual blocks (DWAN<sub>4</sub>) yielded the highest PSNR and SSIM on meanCBF-10<sub>nsm</sub> to meanCBF-20<sub>nsm</sub>, whereas DWAN<sub>sm</sub> with 3 wide activation residual blocks (DWAN<sub>3</sub>) yielded better SSIM for meanCBF-30<sub>nsm</sub> to meanCBF-40<sub>nsm</sub>. Since our goal was to train a model that has the best denoising performance with low SNR input, we chose DWAN<sub>4</sub> in our ASLDN. Figure 4.11 also shows that DWAN<sub>sm</sub> with residual learning had better PSNR and SSIM performance than without residual learning. Table 4.4 shows the grid

search results of finding the best combination of expansion rate  $E$  and the number of input filters  $M$  for all wide activation residual blocks. We found that  $E=4$  and  $M=32$  achieved the best denoising performance.

Table 4.4: 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 the limitation of GPU memory and causing unstable training of DWAN.

$E \backslash M$	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

Figure 4.12 shows the training loss and validation loss of DWAN\_sm. Training loss was the mean absolute errors between output images and surrogate GT over the training dataset. Validation loss was the mean absolute errors between output images and surrogate GT over the validation dataset (the validation dataset was a small sample of data that was separated from the 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 [85] technique was used during training.

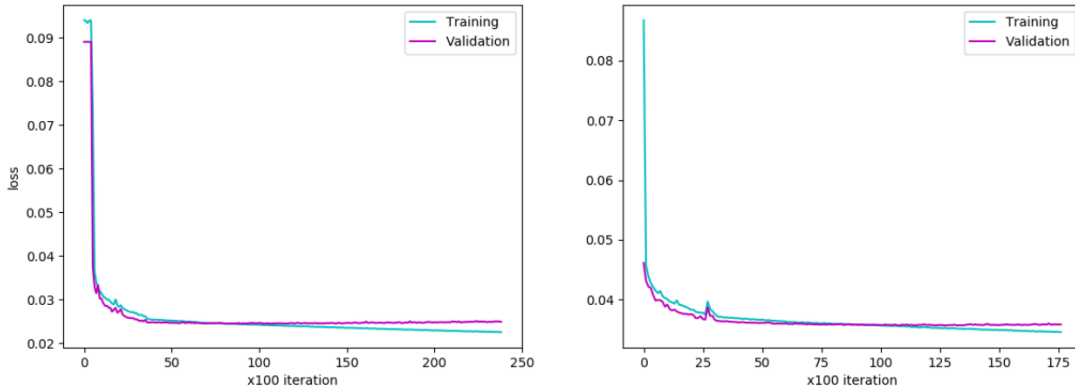


Figure 4.12: Left, training and validation loss of DWAN\_sm. Right, training and validation loss of DWAN\_nsm.

#### 4.4. Discussion and Conclusion

In this chapter, we 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 a 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 and SSIM) and qualitative measure (radiologic scores). Smoothing-based denoising inevitably introduces blurring, but interestingly the DL-based approaches didn't show observable image blurring as evidenced by the image profiles. 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 [86]. Traditional convolution often reduces spatial resolution. In a preliminary version of our published method [74] [70], we observed blurring effects after using an early version of CNNs in ASLDN. In our proposed method, the capability of preserving tissue boundary sharpness and spatial resolution of DWAN was mainly due to the use of two pathways enhanced with dilated filters and wide activation blocks [28].

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

It is worth to note that ASLDN was only trained with data from healthy subjects. Both signal and noise might have different distributions in diseased populations, which may cause 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 fully capable of suppressing noise or artifacts for patients, the model must be trained with additional data from patients.

In summary, we introduced DL into ASL MRI denoising and proposed a new denoising model DWAN in this chapter. ASLDN substantially improved the quality of ASL CBF, reducing acquisition time by 75% (as compared to the 5 mins scan). Besides, ASLDN showed the generalizability of denoising CBFs with different noise levels without model retraining.

## **CHAPTER 5**

### **PRIOR-GUIDED SLICE-WISE ADAPTIVE OUTLIER CLEANING FOR ARTERIAL SPIN LABELING PERFUSION MRI**

#### 5.1 Introduction

Due to the low SNR of a single ASL scan, we need to conduct repeated ASL scans to obtain a relative high-quality mean CBF map for data analysis. However, the mean CBF map is very sensitive to outliers that originated from head motion or other contaminating sources. During the signal averaging process, outliers with large positive and negative values can dominate the process and result in artifacts of bright spots or dark spots in the mean CBF map. Thus, removing outliers is a very important step for ASL MRI post-processing. Current state-of-art Adaptive outlier cleaning (AOC) methods [18] has two issues: one is that current AOC methods discard whole outlier volumes (3D images) without considering whether they contain non-outlier slices (2D images) or not. Discarding the entire volume reduces the sample size and the SNR for non-outlier voxels. The other is that if the reference CBF is already dominated by noise, outlier rejection criterion based on the low correlation with the reference CBF will actually reject the less noisy volumes but keep the noisier ones. Current AOC methods do not consider incorporating prior information (such as anatomical structure information) to identifying outliers in the early iterations.

In this chapter, we propose a better adaptive outlier cleaning method by adding prior guidance and a slice-wise outlier rejection. First, anatomical structure information was adapted as prior information to guide the outliers cleaning method in early iterations.

Second, we rejected outliers slice-wise instead of volume-wise, which saves non-outlier slices from the outlier volume and improves the SNR for non-outlier voxels. For ease of description, we dubbed the new AOC method as prior-guided AOC (PAOC).

## 5.2 Materials and Methods

### 5.2.1. Proposed Method

Figure 5.1 shows the flowchart of PAOCSL. The algorithm can be modified as the volume-wise PAOC after removing the slice-wise loop. The algorithm is detailed below:

**Identification of a potential outlier:** Because a CBF map follows the similar spatial distribution of a grey matter density map, a pseudo CBF map that combines structural information of GM and WM is introduced as the prior guided CBF references. Specifically, the pseudo CBF map is a weighted sum of the GM and WM probability maps from the segmentation of the structural images, i.e.  $\text{pseudo CBF} = \text{GM} \times \beta + \text{WM}$ . The weight  $\beta$  is a constant and is larger than 1 to ensure GM has a higher value than WM ( $\beta$  is 1.8 in this paper). We do not incorporate cerebrospinal fluid (CSF) in the pseudo CBF as CSF has no meaningful ASL signal. A potential outlier is identified when the slice (or volume) has least correlation with the pseudo CBF. In the next step, if the exclusion of the potential outlier from the averaging process resulted in reduced spatial variation within each tissue (GM/WM) type, the potential outlier will be identified as a true outlier and subsequently removed. It should be noted that usually the exclusion of a slice or volume from the averaging process should have increased the variation within each tissue type because of less suppression of random noise. But since outliers dominates the averaging process, removing the outliers will reduce the variations within each tissue type. The variations

within each tissue type were computed as the sum of standard deviation of CBF signals in each tissue type.

**Constraints for removing an outlier:** Because bulk head motions usually occur in a time long enough to acquire several slices, the most adjacent slices within the same volume should see similar outliers. If adjacent slices are processed independently, they may have quite different outlier timepoints removed, resulting in large discontinuity at certain places in the final cleaned image. Thus, we penalize variations of GM and WM include the neighboring slices, which will constrain the process of outlier removal to have similar outlier timepoints removed for adjacent slices. We empirically found that including 3 slices (the current slice, the superior slice and the inferior slice) provides a balance between performing outlier removal locally while obtaining a continuous resulting CBF map. Noted that larger number of neighboring slices will bring PAOCSL closer to a volume-wise PAOC.

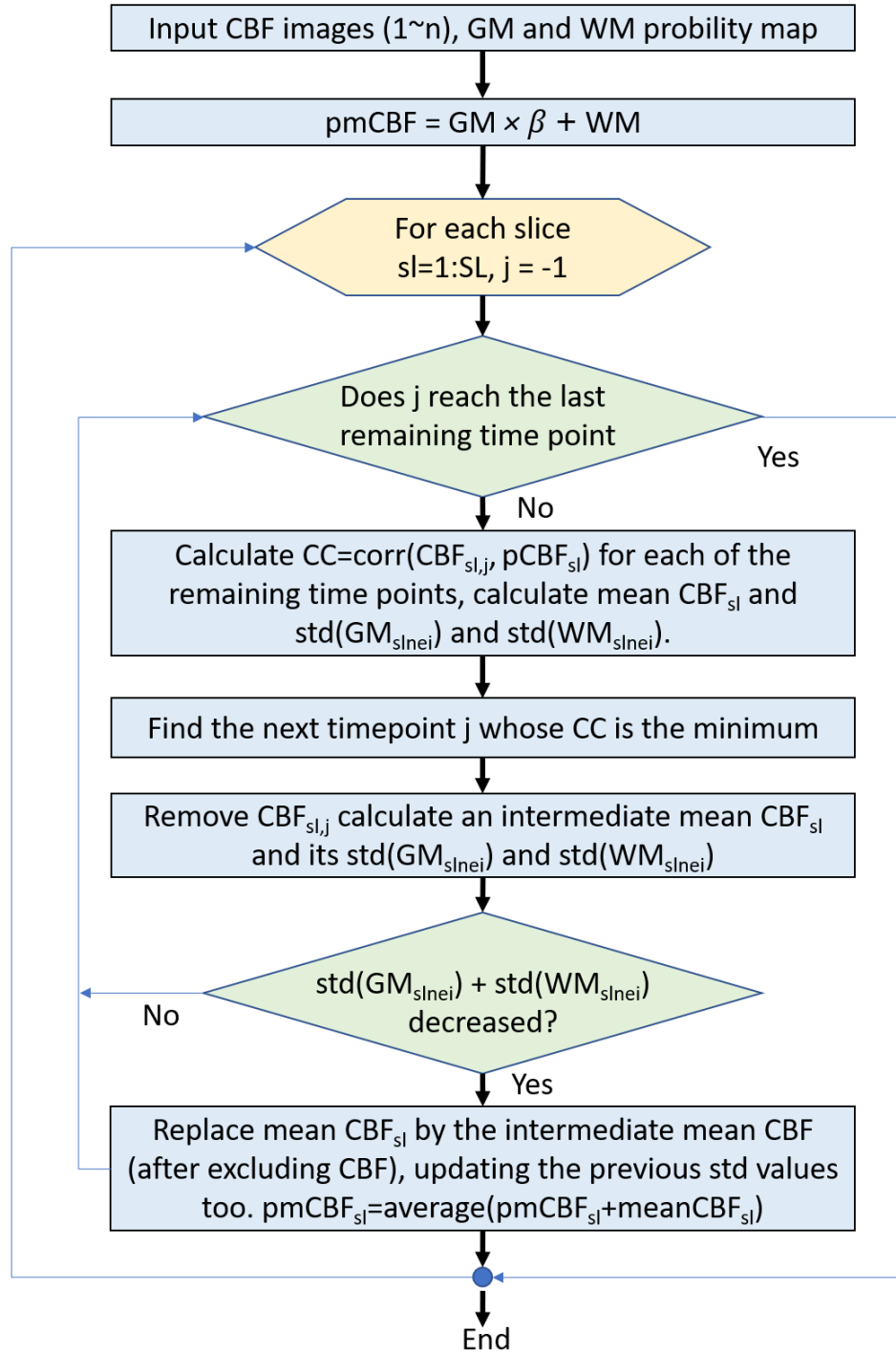


Figure 5.1: Flowchart of the prior-guided AOC algorithm. *pmCBF* means pseudo mean CBF map; *SL* means the maximum number of slices; *slnai* means the neighboring slices around the current slice *sl*;  $\beta$  is an empirically defined constant; *CC* means correlation coefficient; *corr* means correlation; WM/GM means white matter/grey matter; *std* means standard deviation.

### 5.1.2 Image Acquisition

Repeat ASL data were acquired twice with 3 months apart. The data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. This data included 19 EC [age:  $74.7 \pm 6.6$  yrs (mean  $\pm$  standard deviation), Mini-Mental Status Examination (MMSE):  $29.2 \pm 1.1$ , 9 males, 10 females]. The structural images were acquired using a 3D MPRAGE T1-weighted sequence with the following parameters: TR/TE/TI = 2300/2.98/900 ms, 176 sagittal slices, within plane FOV= $256 \times 240$  mm<sup>2</sup>, voxel size=  $1.1 \times 1.1 \times 1.2$  mm<sup>3</sup>, flip angle= $9^\circ$ , bandwidth=240 Hz/pix. ASL data were acquired using the Siemens product PICORE sequence [87], which is a pulsed ASL (PASL) sequence using the Q2TIPs [88] technique for defining the spin bolus. The acquisition parameters were: TR/TE=3400/12 ms, TI1/TI2=700/1900 ms, FOV=256 mm, 24 sequential 4 mm thick slices with a 25% gap between the adjacent slices, partial Fourier factor= 6/8, bandwidth = 2368 Hz/pix, imaging matrix= $64 \times 64$ . The first volume of the 105 ASL acquisitions was used as the M0 image.

ASL images were preprocessed using ASLtbx [65] with the procedures in section 3.3.1. GM/WM masks were then created by thresholding the GM/WM probability maps with a threshold of 0.4. After CBF calculations, the CBF time series were cleaned using the following cleaning algorithms: the Structural Correlation based Outlier Rejection (SCORE) algorithm [18] (the first improved adaptive outlier cleaning method), prior-guided adaptive outlier cleaning (PAOC) method applied to image volumes, and slice-wise PAOC (PAOCSL). Other methods were not included because they were compared with SCORE in [18].

The effects of outlier cleaning were evaluated with SNR and test-retest stability. The SNR of CBF maps was the mean GM CBF value divided by the standard deviation of WM CBF values. Mean CBF maps were projected into the MNI standard brain space. We used intraclass correlation coefficient (ICC) to measure the test-retest stability. ICC was calculated at each voxel of all subjects' repeat ASL CBF maps obtained by various outlier cleaning methods. PAOCSL may also benefit 3D ASL data. To verify this possibility, we applied PAOCSL to recently published 3D ASL data [89]. The 3D ASL data was acquired using our state-of-art background suppressed spiral readout 3D pseudo continuous ASL sequence with partition-wise parallel imaging acceleration [89] [90].

## 5.2 Experimental Results

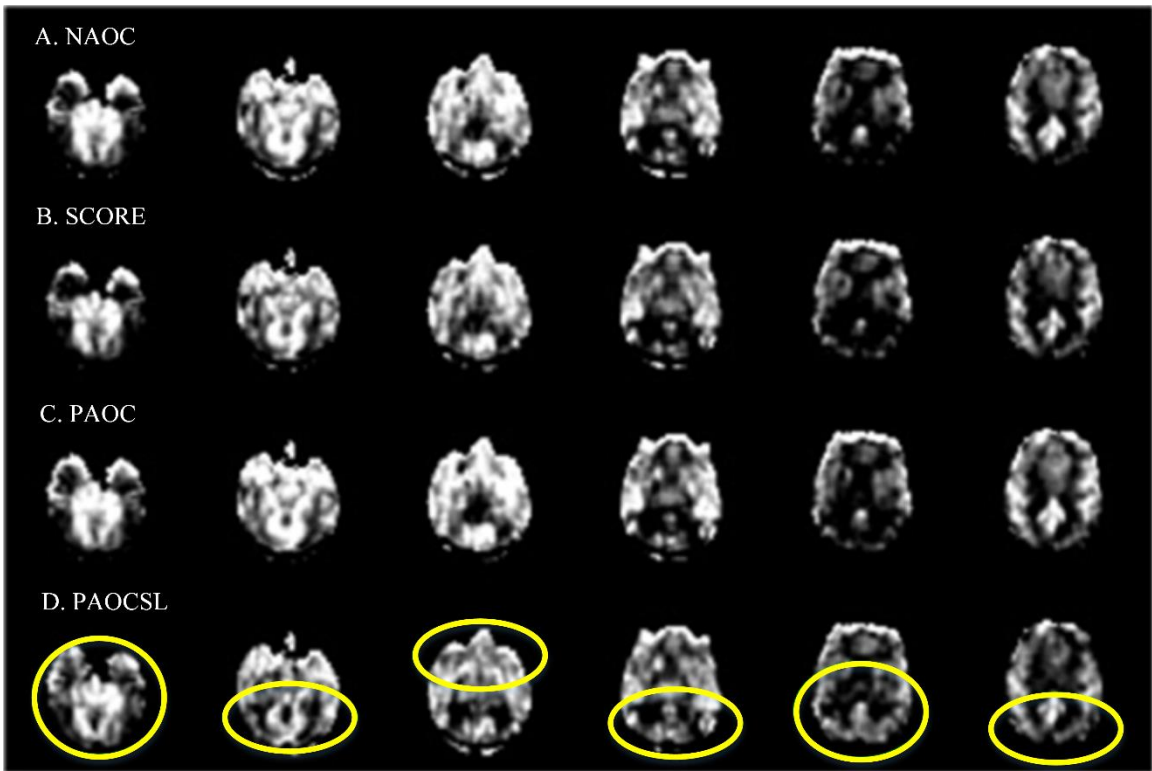


Figure 5.2: ASL CBF maps of a representative subject processed with different methods: A) NAOC, B) SCORE, C) PAOC, D) PAOCSL. Image display window was 0-60

ml/100g/min. Green ovals mark the places with improved CBF image contrasts in PAOCSL.

Figure 5.2 demonstrates results from one representative subject. NAOC means no adaptive outlier cleaning method used. Moving from NAOC to SCORE, PAOC, and then PAOCSL (Figure 5.2 A-D), the image quality of CBF gradually improved in places marked by the yellow ovals. PAOCSL recovered signal in the posterior part of the brain, and suppressed the artificial bright spot in the anterior part of the brain.

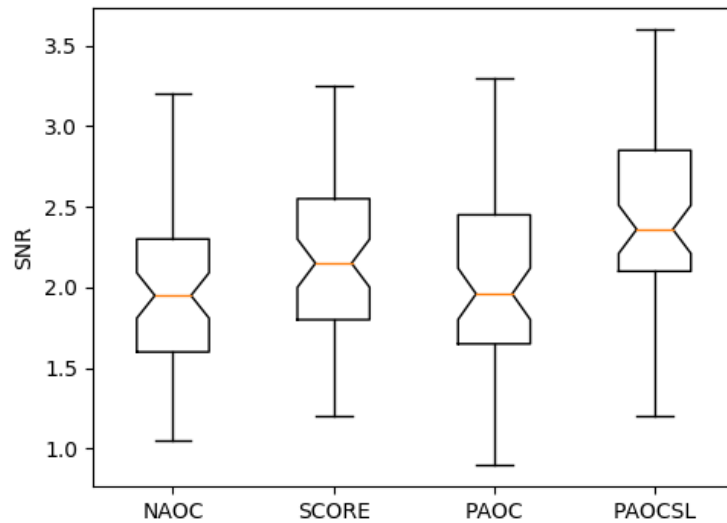


Figure 5.3: Boxplots of SNR of different outlier cleaning approaches. Red bar indicates median of the SNR of all subjects (n=19). PAOCSL showed significantly higher SNR than other methods ( $p < 0.005$  for all possible two-sample t-tests).

Figure 5.3 shows the SNR boxplot of 19 subjects' session one data. PAOCSL produced the highest SNR than NAOC, SCORE, and PAOC ( $p < 0.005$  for all possible two-sample t-tests). No significant SNR difference was found between other methods. SCORE and PAOC identified 12 and 14 outlier CBF images on average (containing all 24 slices), respectively. Across slices, PAOCSL identified 23 outlier timepoints, but only 5 timepoints on average had the entire CBF image volume removed. Figure 5.4 shows the number of

removed outlier slices versus time curve for a representative subject from ADNI. Even though most of timepoints presented outlier slices, only 11 timepoints had the whole 3D CBF volume removed as the outliers.

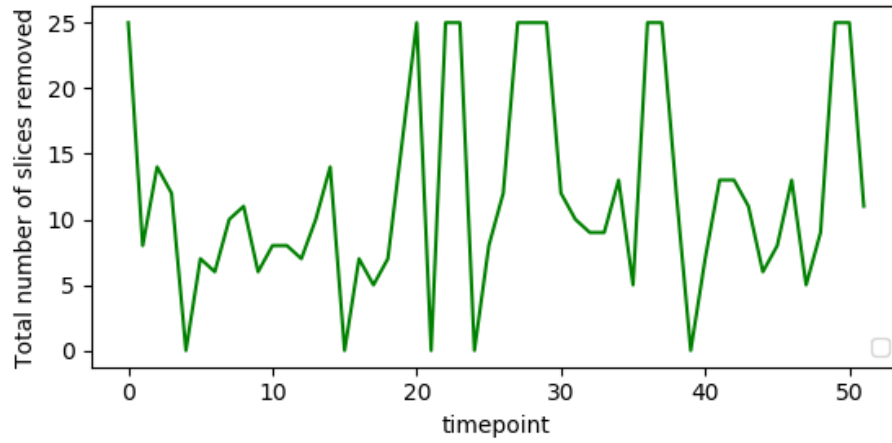


Figure 5.4: Number of outlier slices removed at different timepoint for one representative subject from the ADNI database

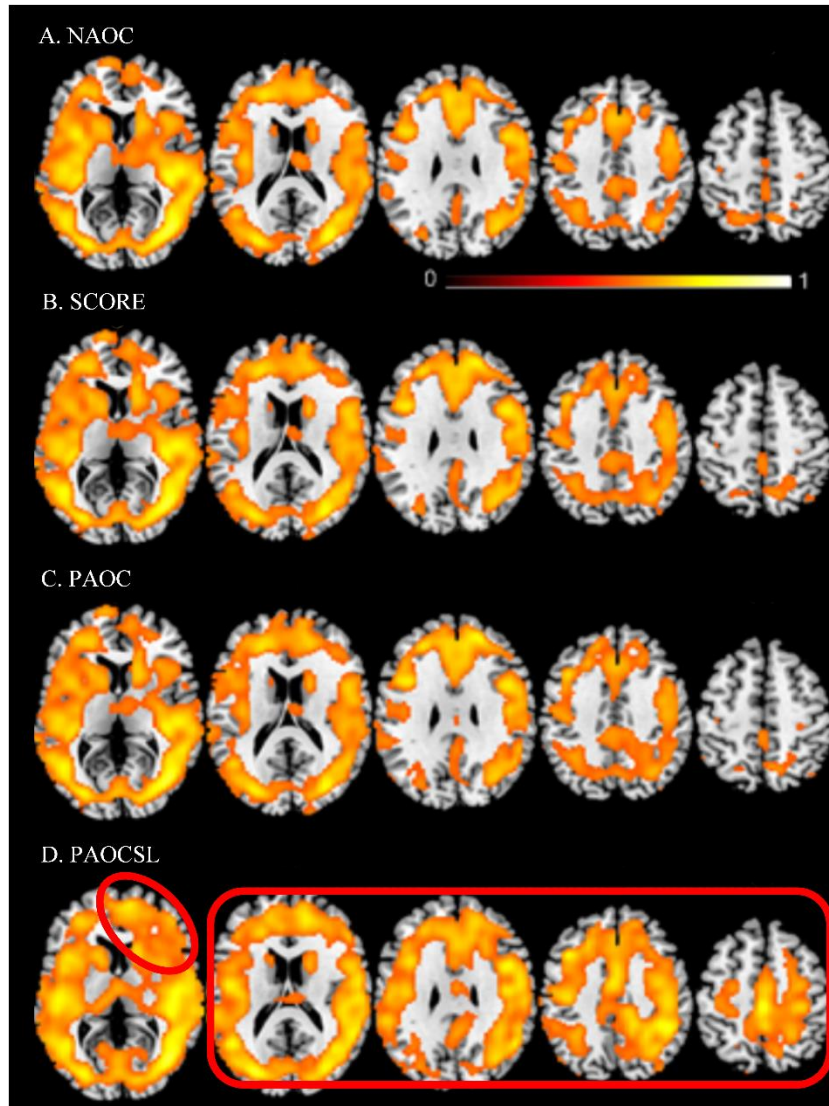


Figure 5.5: ICC maps of CBF maps obtained with different methods: From top to bottom: no AOC, SCORE, PAOC, and PAOCSL. The threshold was 0.5 and the colorbar indicates the display window of ICC. Red oval and red box indicate places with higher CBF measurement test-retest stability after applying PAOCSL.

Figure 5.5 illustrates the ICC analysis results. PAOC improved ICC in both the anterior and posterior part of the brain compared to NAOC. Though PAOC showed similar ICC patterns to those of SCORE, PAOCSL yielded better CBF ICC in a few brain regions as marked by the red oval and red box in Figure 5.5 D.

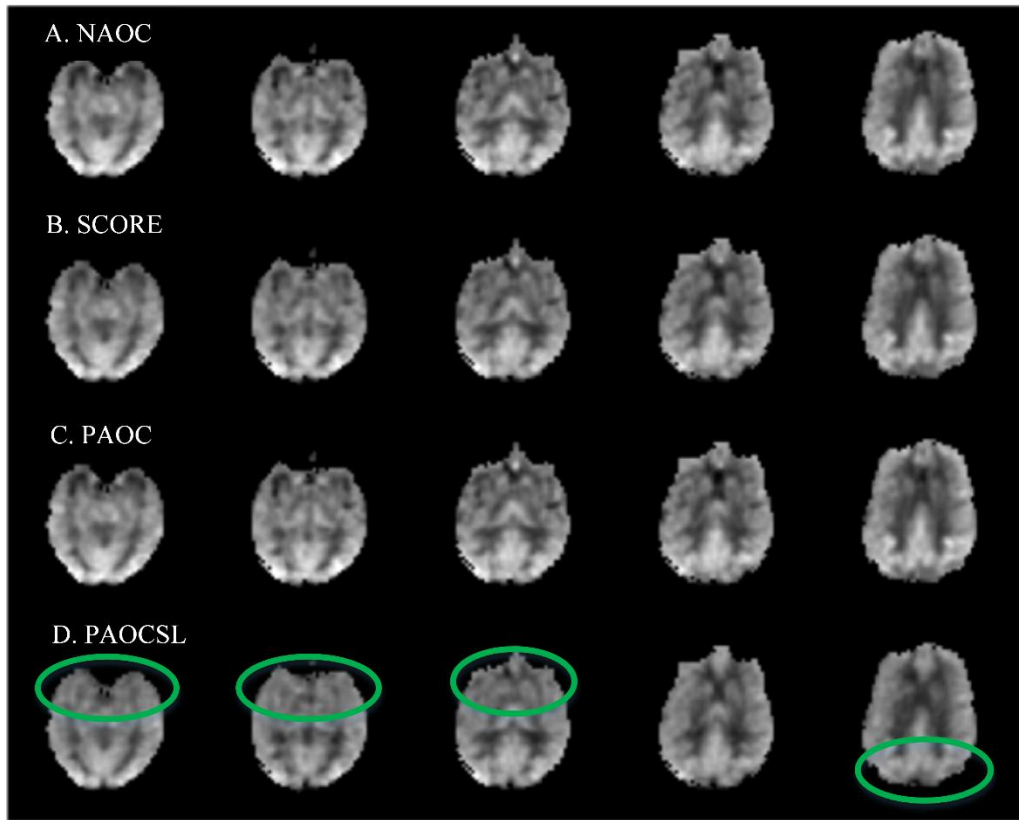


Figure 5.6: 3D ASL CBF maps of a young healthy subject processed with: A) NAOC, B) SCORE, C) PAOC, D) PAOCSL. Green ovals mark the places with improved CBF image contrasts in PAOCSL in terms of better image homogeneity and less signal loss in inferior frontal regions. Image display window was 0-80 ml/100g/min.

Figure 5.6 demonstrates the CBF map of a representative subject from the 3D ASL dataset. Quite similar CBF maps were produced by all the methods. Compared to non-outlier rejection, only the slice-wise outlier rejection achieved moderate improvement in areas as marked by green ovals.

### 5.3 Discussion and Conclusion

Outlier cleaning is an important step in ASL MRI data processing because of the low SNR and the limited number of samples in each scan. In this chapter, we showed that incorporating prior knowledge about CBF structural distributions into AOC and using a slice-wise outlier cleaning method enhanced both image quality and the test-retest stability

of ASL CBF. A prior guided reference CBF map can avoid bias from extreme outliers in the early iterations of outlier cleaning, ensuring a correct identification of the true outliers. Slice-wise outlier rejection can save slices with CBF values in the reasonable range even they are in the outlier volumes, or remove outlier slices even if they are in the non-outlier volumes. Our data demonstrated that this fine scale outlier cleaning method provided much better improvement of CBF image quality. Rejecting outliers reduces the number of samples and may reduce SNR if noises follows Gaussian or Rician distribution [91]. However, CBF image outliers usually show very different image distribution from the regular CBF image and subsequently induce large noise variations. Removing those outliers should help to reduce the noise variation and increase SNR of CBF. This variance reduction is in fact a mandatory criterion for rejecting the outliers from further image averaging process in SCORE, PAOC, and PAOCSL, which explains the SNR increase between these methods and the standard processing steps (NAOC). PAOCSL showed the highest SNR increase because noise reduction is executed for each slice separately. Similar to the extent of SNR improvement, volume-wise outlier cleaning (SCORE and PAOC) showed marginal ICC improvement. In contrast, PAOCSL produced high test-retest CBF stability in spatially more extended brain regions.

Although the reference CBF images used in this chapter were created empirically from the segmented GM and WM maps, they can be replaced with any proton density weighted images. Test-retest stability is widely used as a performance index in fMRI but higher test-retest stability may also show a result of suppressing true physiological variations. This however is much less of a concern because averaging out the physiological variations is unavoidable even without outlier rejection.

In conclusion, a prior-guided slice-wise adaptive outlier cleaning (PAOCSL) method was proposed in this chapter for ASL perfusion MRI. PAOCSL can markedly improve CBF quantification quality in terms of SNR and test-retest stability.

## **CHAPTER 6**

### **CONCLUSION AND FUTURE WORK**

#### 6.1 Conclusion

In this dissertation, we proposed machine learning-based methods for Arterial Spin Labeling MRI signal processing. This thesis has explored the effectiveness of using machine learning and deep learning to improve the image quality of ASL. Experimental results have demonstrated the efficacy of the techniques theoretically and practically. Especially, we have developed the following methods:

First, we developed a learning-from-noise method for DL-based for ASL denoising. The proposed learning-from-noise method shows that DL-based ASL denoising (ASLDN) models can be trained using only noisy image pairs, without any deliberate post-processing for getting the quasi-noise-free reference during the training process. This learning-from-noise method can also be applied to DL-based ASL perfusion prediction from BOLD fMRI as ASL reference are extremely noisy in this BOLD-to-ASL prediction. Experimental results demonstrated that this learning-from-noise method can reliably denoise ASL MRI and predict ASL perfusion from BOLD fMRI. Moreover, by using ASLDN-LFN, more training data can be generated as it requires fewer samples to generate quasi-noise-free reference, which is particularly useful when ASL CBF data are limited.

Second, we proposed a novel and optimized deep learning neural network, i.e., Dilated Wide Activation Network (DWAN), for ASL denoising. Our method present two novelties: first, we incorporated the wide activation residual blocks with a Dilated Convolution Neural Network to achieve improved denoising performance in term of

several quantitative and qualitative measurements; second, we evaluated our proposed model given different inputs and references to show that our denoising model can be generalized to inputs with different levels of SNR and yielded images with better quality than other methods.

Third, we designed a prior-guided and slice-wise adaptive outlier cleaning algorithm (PAOCSL) to improve the original Adaptive Outlier Cleaning (AOC) algorithm. A prior information guided reference CBF map is used to avoid bias from extreme outliers in the early iterations of outlier cleaning, ensuring correct identification of the true outliers. Slice-wise outlier rejection is adapted to reserve slices with CBF values in the reasonable range even they are within the outlier volumes, or remove outlier slices even if they are in the non-outlier volumes. Experimental results show that the proposed outlier cleaning method showed both improved CBF quantification quality and CBF measurement stability.

## 6.2 Future Work

Current DL-based ASL denoising methods are supervised learning, i.e., training with references. A future direction could be using semi-supervised or unsupervised learning for ASL denoising, that is, training an ASLDN with limited references or without any reference. This would be very beneficial as the number of training references is limited.

Besides, combining prior knowledge into ASLDN and ASLDN-LFN is another future direction. GM probability map provides useful prior information and could be incorporated into the training process of ASLDN as a regularization term. This could guide the training of ASLDN and possible improvement of ASLDN and ASLDN-LFN.

Furthermore, the prior-guided slice-wise adaptive outlier cleaning (PAOCSL) method can be used as a pre-processing step for ASLDN-LFN. PAOCSL can be used to remove outliers among noisy references, which could help ASLDN-LFN to learn the more accurate latent clean distribution. This method can also apply to BOA-Net where outliers would be even more abundant in noisy references.

## BIBLIOGRAPHY

- [1] S. Ogawa, Tso-Ming Lee, Alan R. Kay and a. D. W. Tank., "Brain magnetic resonance imaging with contrast dependent on blood oxygenation.," *proceedings of the National Academy of Sciences*, vol. 87, no. 24, pp. 9868-9872, 1990.
- [2] J. A. Detre, J. S. Leigh, D. S. Williams and A. P. Koretsky, "Perfusion imaging," *Magnetic resonance in medicine*, vol. 23, pp. 37-45, 1992.
- [3] R. A. Poldrack, J. A. Mumford and T. E. Nichols, *Handbook of functional MRI data analysis.*, Cambridge University Press, 2011.
- [4] J. A. Detre and J. Wang, "Technical aspects and utility of fMRI using BOLD and ASL," *Clinical Neurophysiology*, vol. 5, no. 113, pp. 621-634., 2002.
- [5] J. A. Detre and T. F. Floyd, "Functional MRI and its applications to the clinical neurosciences," *The neuroscientist*, vol. 7, no. 1, pp. 64-79, 2001.
- [6] S. M. Spann, K. S. Kazimierski, C. S. Aigner, M. Kraiger, K. Bredies and R. Stollberger, "Spatio-temporal TGV denoising for ASL perfusion imaging," *Neuroimage*, vol. 157, pp. 81-96, 2017.
- [7] D. C. Alsop, J. A. Detre, X. Golay, M. Günther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits and others, "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," *Magnetic resonance in medicine*, vol. 73, pp. 102-116, 2015.
- [8] Z. Wang, "Improving cerebral blood flow quantification for arterial spin labeled perfusion MRI by removing residual motion artifacts and global signal fluctuations," *Magnetic resonance imaging*, vol. 30, pp. 1409-1415, 2012.
- [9] Y. Behzadi, K. Restom, J. Liau and T. T. Liu, "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI," *Neuroimage*, vol. 37, pp. 90-101, 2007.
- [10] J. Wang, Z. Wang, G. K. Aguirre and J. A. Detre, "To smooth or not to smooth? ROC analysis of perfusion fMRI data," *Magnetic resonance imaging*, vol. 23, pp. 75-81, 2005.
- [11] A. Bibic, L. Knutsson, F. Ståhlberg and R. Wirestam, "Denoising of arterial spin labeling data: wavelet-domain filtering compared with Gaussian smoothing," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 23, pp. 125-137, 2010.

- [12] J. A. Wells, D. L. Thomas, M. D. King, A. Connelly, M. F. Lythgoe and F. Calamante, "Reduction of errors in ASL cerebral perfusion and arterial transit time maps using image de-noising," *Magnetic resonance in medicine*, vol. 64, pp. 715-724, 2010.
- [13] X. Liang, A. Connelly and F. Calamante, "Voxel-wise functional connectomics using arterial spin labeling functional magnetic resonance imaging: the role of denoising," *Brain connectivity*, vol. 5, pp. 543-553, 2015.
- [14] H. Zhu, J. Zhang and Z. Wang, "Arterial spin labeling perfusion MRI signal denoising using robust principal component analysis," *Journal of neuroscience methods*, vol. 295, pp. 10-19, 2018.
- [15] Z. Wang, "Support vector machine learning-based cerebral blood flow quantification for arterial spin labeling MRI," *Human brain mapping*, vol. 35, pp. 2869-2875, 2014.
- [16] H. Zhu, G. He and Z. Wang, "Patch-based local learning method for cerebral blood flow quantification with arterial spin-labeling MRI," *Medical & biological engineering & computing*, pp. 1-6, 2017.
- [17] Z. Wang, S. R. Das, S. X. Xie, S. E. Arnold, J. A. Detre, D. A. Wolk, A. D. N. Initiative and others, "Arterial spin labeled MRI in prodromal Alzheimer's disease: a multi-site study," *NeuroImage: Clinical*, vol. 2, pp. 630-636, 2013.
- [18] S. Dolui, Z. Wang, R. T. Shinohara, D. A. Wolk, J. A. Detre and A. D. N. Initiative, "Structural Correlation-based Outlier Rejection (SCORE) algorithm for arterial spin labeling time series," *Journal of Magnetic Resonance Imaging*, vol. 45, pp. 1786-1797, 2017.
- [19] Z. Wang, "Priors-guided adaptive outlier cleaning for arterial spin labeling perfusion MRI," in *ISMRM 2016*, 2016.
- [20] Y. Li, S. Dolui, D.-F. Xie, Z. Wang, A. D. N. Initiative and others, "Priors-guided slice-wise adaptive outlier cleaning for arterial spin labeling perfusion MRI," *Journal of neuroscience methods*, vol. 307, pp. 248-253, 2018.
- [21] K. H. Kim, S. H. Choi and S.-H. Park, "Improving Arterial Spin Labeling by Using Deep Learning," *Radiology*, p. 171154, 2017.
- [22] C. Ulas, G. Tetteh, S. Kaczmarz, C. Preibisch and B. H. Menze, "DeepASL: Kinetic Model Incorporated Loss for Denoising Arterial Spin Labeled MRI via Deep Residual Learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018.
- [23] E. Gong, J. Pauly and G. Zaharchuk, "Boosting SNR and/or resolution of arterial spin label (ASL) imaging using multi-contrast approaches with multi-lateral guided filter and deep networks," in *25th Annual Meeting of the International Society for Magnetic Resonance in Medicine*, Honolulu, 2017.

- [24] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *2015 pp. 234-241.*, In International Conference on Medical image computing and computer-assisted intervention,.
- [25] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [26] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [27] B. Lim, S. Son, H. Kim, S. Nah and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017.
- [28] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang and T. Huang, "Wide activation for efficient and accurate image super-resolution," *arXiv preprint arXiv:1808.08718*, 2018.
- [29] R. B. Buxton, Kâmil Uludağ, David J. Dubowitz and a. T. T. Liu., "Modeling the hemodynamic response to brain activation," *Neuroimage* , vol. 23 , pp. S220-S233, 2004.
- [30] D. S. Williams, J. A. Detre, J. S. Leigh and A. P. Koretsky, "Magnetic resonance imaging of perfusion using spin inversion of arterial water," *Proceedings of the National Academy of Sciences*, vol. 89, pp. 212-216, 1992.
- [31] E. C. Wong, "Potential and pitfalls of arterial spin labeling based perfusion imaging techniques for MRI," *Functional MRI. Heidelberg: Springer*, pp. 63-69, 1999.
- [32] L. Deng, D. Yu and others, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, pp. 197-387, 2014.
- [33] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, pp. 504-507, 2006.
- [34] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012.
- [35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [36] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014.
- [37] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.

- [38] A. Graves, A.-r. Mohamed and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on*, 2013.
- [39] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen and others, "Deep speech 2: End-to-end speech recognition in english and mandarin," in *International Conference on Machine Learning*, 2016.
- [40] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel and Y. Bengio, "End-to-end attention-based large vocabulary speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, 2016.
- [41] A.-r. Mohamed, G. E. Dahl and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 14-22, 2012.
- [42] D. Shen, G. Wu and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221-248, 2017.
- [43] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. Laak, B. Ginneken and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60-88, 2017.
- [44] D. Xie, Yiran Li, Hanlu Yang, Li Bai, Tianyao Wang, Fuqing Zhou, Lei Zhang and Z. Wang., " Denoising arterial spin labeling perfusion MRI with deep machine learning.," *Magnetic Resonance Imaging*, 2020.
- [45] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot and others, "Mastering the game of Go with deep neural networks and tree search," *nature*, vol. 529, pp. 484-489, 2016.
- [46] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel and others, "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," *arXiv preprint arXiv:1712.01815*, 2017.
- [47] K. Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," *Neural networks*, vol. 1, pp. 119-130, 1988.
- [48] G. E. Hinton, S. Osindero and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, pp. 1527-1554, 2006.
- [49] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *nature*, vol. 521, p. 436, 2015.
- [50] D. Xie, L. Zhang and L. Bai, "Deep learning in visual computing and signal processing," *Applied Computational Intelligence and Soft Computing*, vol. 2017, 2017.

- [51] D. Xie and L. Bai., "A hierarchical deep neural network for fault diagnosis on Tennessee-Eastman process.," in *IEEE 14th International Conference on Machine Learning and Applications*, 2015.
- [52] Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in neural information processing systems*, 2007.
- [53] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng and D. Liang, "Accelerating magnetic resonance imaging via deep learning," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016.
- [54] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," *Magnetic resonance in medicine*, vol. 80, pp. 1189-1205, 2018.
- [55] J. Xu, E. Gong, J. Pauly and G. Zaharchuk, "200x low-dose pet reconstruction using deep learning," *arXiv preprint arXiv:1712.04119*, 2017.
- [56] E. Gong, J. M. Pauly, M. Wintermark and G. Zaharchuk, "Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI," *Journal of Magnetic Resonance Imaging*, vol. 48, pp. 330-340, 2018.
- [57] Y. Li, D. Xie, A. Cember, R. P. R. Nanga, H. Yang, D. Kumar and H. Hariharan, "Accelerating GluCEST imaging using deep learning for B0 correction," *Magnetic Resonance in Medicine*, 2020.
- [58] D. Xie, Y. Li, H. Yang, D. Song, Y. Shang, Q. Ge, L. Bai and Z. Wang., ""BOLD fMRI-Based Brain Perfusion Prediction Using Deep Dilated Wide Activation Networks.," in *In International Workshop on Machine Learning in Medical Imaging*, 2019.
- [59] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, 2017.
- [60] K. He, Xiangyu Zhang, Shaoqing Ren and a. J. Sun., "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.," in *In Proceedings of the IEEE international conference on computer vision*, 2015.
- [61] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015.
- [62] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18-31, 2017.

- [63] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61-78, 2017.
- [64] R. B. Buxton, L. R. Frank, E. C. Wong, B. Siewert, S. Warach and R. R. Edelman, "A general kinetic model for quantitative perfusion imaging with arterial spin labeling," *Magnetic resonance in medicine*, vol. 40, no. 3, pp. 383-396., 1998.
- [65] Z. Wang, G. K. Aguirre, H. Rao, J. Wang, M. A. Fernández-Seara, A. R. Childress and J. A. Detre, "Empirical optimization of ASL data analysis using an ASL data processing toolbox: ASLtbx," *Magnetic resonance imaging*, vol. 26, pp. 261-269, 2008.
- [66] H. Tan, Joseph A. Maldjian, Jeffrey M. Pollock, Jonathan H. Burdette, Lucie Y. Yang, Andrew R. Deibler and a. R. A. Kraft., "A fast, effective filtering method for improving clinical pulsed arterial spin labeling MRI.," *Journal of Magnetic Resonance Imaging*, vol. 29, no. 5, pp. 1134-1139, 2009.
- [67] C. Maumet, Pierre Maurel, Jean-Christophe Ferré and a. C. Barillot., "Robust estimation of the cerebral blood flow in arterial spin labelling," *Magnetic resonance imaging*, vol. 32, no. 5, pp. 497-504, 2014.
- [68] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [69] S. Ruder, "An overview of gradient descent optimization algorithms," in *arXiv preprint arXiv:1609.04747*, 2016.
- [70] D. Xie, L. Bai and Z. Wang, "Denoising Arterial Spin Labeling Cerebral Blood Flow Images Using Deep Learning," *arXiv preprint arXiv:1801.09672*, 2018.
- [71] J. Lehtinen, Jacob Munkberg, J. Hasselgren, S. L. T. Karras, M. A. and a. T. Aila., "Noise2Noise: Learning Image Restoration without Clean Data.," in *In International Conference on Machine Learning*, 2018.
- [72] P. J. Huber, "Robust Estimation of a Location Parameter.," *The Annals of Mathematical Statistics*, pp. 73-101, 1964.
- [73] Y. Fan, J. Yu and T. S. Huang, "Wide-activated deep residual networks based restoration for BPG-compressed images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018.
- [74] D. Xie, L. Bai and Z. Wang, "Denoising Arterial Spin Labeling Cerebral Blood Flow Images Using Deep Learning," *ISMRM 2018*, 2018.
- [75] F. Chollet and others, *Keras*, 2015.

- [76] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015.
- [77] R. D. Hoge, Jeff Atkinson, Brad Gill, Gérard R. Crelier, Sean Marrett and a. G. B. Pike., "Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex," *Proceedings of the National Academy of Sciences*, vol. 96, no. 16, pp. 9403-9408, 1999.
- [78] D. Xie, Y. Li, L. Bai and Z. Wang, "Super-ASL: Improving SNR and Temporal Resolution of ASL MRI Using Deep Learning," in *ISMRM workshop on machine learning*, 2018.
- [79] D. D. Shin, T. T. Liu, E. C. Wong, A. Shankaranarayanan and a. Y. Jung., "Pseudocontinuous arterial spin labeling with optimized tagging efficiency," *Magnetic resonance in medicine*, vol. 68, no. 4, pp. 1135-1144, 2012.
- [80] X.-J. Mao, C. Shen and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," *arXiv preprint arXiv:1606.08921*, 2016.
- [81] Y. Zhang, Y. Tian, Y. Kong, B. Zhong and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [82] Y. Fan, H. Shi, J. Yu, D. Liu, W. Han, H. Yu, Z. Wang, X. Wang and T. S. Huang, "Balanced two-stage residual networks for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017.
- [83] W. Dai, D. Garcia, C. D. Bazelaire and D. C. Alsop, "Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields," *Magnetic Resonance in Medicine*, vol. 60, no. 6, pp. 1488-1497, 2008.
- [84] W. Wu, M. Fernández-Seara, J. A. Detre, F. W. Wehrli and J. Wang, "A theoretical and experimental investigation of the tagging efficiency of pseudocontinuous arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 58, no. 5, pp. 1020-1027, 2007.
- [85] L. Prechelt, "Early stopping-but when?," in *Neural Networks: Tricks of the trade*, Springer, 1998, pp. 55-69.
- [86] D. Ulyanov, A. Vedaldi and V. Lempitsky, "Deep image prior," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

- [87] E. C. Wong, Richard B. Buxton and a. L. R. Frank., "Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling," *NMR in Biomedicine*, vol. 10, no. 4-5, pp. 237-249, 1997.
- [88] W. Luh, E. C. Wong, P. A. B. and a. J. S. Hyde., "QUIPSS II with thin-slice T11 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling," *Magnetic resonance in medicine*, vol. 41, no. 6, pp. 1246-1254, 1999.
- [89] M. Vidorreta, Z. Wang, Y. V. Chang, D. A. Wolk, M. A. Fernandez-Seara and J. A. Detre, "Whole-brain background-suppressed pCASL MRI with 1D-accelerated 3D RARE Stack-Of-Spirals readout," *PloS one*, vol. 12, no. 8, 2017.
- [90] M. Vidorreta, E. Balteau, Z. Wang, E. D. Vita, M. A. Pastor, D. L. Thomas, J. A. Detre and M. A. Fernández-Seara, "Evaluation of segmented 3D acquisition schemes for whole-brain high-resolution arterial spin labeling at 3 T," *NMR in Biomedicine*, vol. 27, no. 11, pp. 1387-1396., 2014.
- [91] R. W. Brown, Y-C. Norman Cheng, E. Mark Haacke, Michael R. Thompson and a. R. Venkatesan., *Magnetic resonance imaging: physical principles and sequence design.*, John Wiley & Sons, 2014.
- [92] J. Xie, L. Xu and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in neural information processing systems*, 2012.
- [93] H. C. Burger, C. J. Schuler and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012.
- [94] C. Dong, C. C. Loy, K. He and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*, 2014.
- [95] N. Maleki, W. Dai and D. C. Alsop., "Optimization of background suppression for arterial spin labeling perfusion imaging," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 25, no. 2, pp. 127-133, 2012.
- [96] I.-K. L. Lawrence, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, pp. 255-268., 1989.