

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Development of Deep Learning Methods for Magnetic Resonance Phase Imaging of Neurological Disease

Permalink

<https://escholarship.org/uc/item/0fp8k57g>

Author

Chen, Yicheng

Publication Date

2019

Peer reviewed|Thesis/dissertation

Development of Deep Learning Methods for Magnetic Resonance Phase Imaging of
Neurological Disease

by
Yicheng Chen


DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY


in
Bioengineering

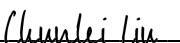
in the
GRADUATE DIVISION

of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:

F1E4E52A4E3D4D8... Janine Lupo
Chair

DocuSigned by:

Peder Larson

DocuSigned by:

147ED61F87EA4BD... Chunlei Liu

Committee Members

Acknowledgement

There are many people I would like to thank for guiding and helping me throughout these years. First and foremost, I would like to thank my academic advisor, Dr. Janine Lupo. I first met Dr. Lupo in the summer of 2016. At that time, my interest in MR phase imaging coincided with her vision and idea of future directions and I was lucky to join her lab as her first Ph.D. student. Dr. Lupo is the most gentle and patient person I have ever met. She is always ready to explain concepts to me and willing to discuss project ideas with me. Her constructive advice and comments helped sharpen my academic writing and presentation skills. Her detail-oriented attitude in scientific research influenced me a lot. I am also very thankful to her for supporting me to pursue deep learning as one of the major topics for this dissertation.

I would like to thank Dr. Peder Larson. Dr. Larson taught me my first MRI course and is a member of my qualifying exam and dissertation committee. I am always amazed by his broad knowledge in almost every field of MRI. His advice on deep learning helped me improve the projects in my dissertation.

I also want to thank Dr. Michael Lustig and Dr. Chunlei Liu from UC Berkeley. I learned a lot of technical details from Dr. Lustig's MRI course. His dedicated attitude and passion for MRI reconstruction always inspired me. Dr. Liu discussed many techniques and ideas about QSM with me and helped me gain a better big picture of this field. He is also a member of my qualifying exam and dissertation committee. I also appreciate the guidance from former members of his lab, Dr. Hongjiang Wei and Dr. Yuyao Zhang, on QSM processing code.

I would like to thank Dr. Duan Xu for being the chair of my qualifying exam committee. He provided invaluable guidance for my preparation and made sure I was on the right track.

I would also like to thank the late Dr. Sarah Nelson. Although I didn't have the chance to directly work with her, her enthusiasm and spirit influenced the entire brain lab and greatly inspired me.

I am grateful for help from the colleagues in the Lupo lab. It was a pleasure to collaborate with Dr. Melanie Morrison on the CMB detection project. The GUI for CMB detection she developed enabled my automatic CMB detection project and the follow-up research. Thanks to Sivakami Avadiappan for labeling tons of CMB and collaboration on data processing. Thanks to Dr. Maryam Vareth for bringing me into the collaboration group in BIDS for brain tumor segmentation. Special thanks to Angela Jakary for overcoming so many 7T scanner failure and conducting so many tedious MRI scans with me, especially those SWAN QSM with multiple orientations. Without her, my deep learning models would have nothing to learn from.

I also appreciate the help from my friends in the Surbeck Lab: thank Xucheng Zhu, Dr. Zihan Zhu and Dr. Shuyu Tang for all the lunchtime discussions on all aspects of MRI, thank Dr. Yiran Chen for collaboration in many projects and helps on writing this dissertation, thank Dr. Peng Cao and Andrew Leynes for exchanging fancy (or crazy) deep learning ideas with me, thank Dr. Wenwen Jiang for helping me get familiar with the UCSF research environment, and thank Dr. Hong Shang for suggestions for my career decisions.

Last but not least, I would like to thank my wife Dan Shen for all the support she gave me over the past few years. It was her that comforted me during the hard times and encouraged me to work hard to finish this dissertation.

Yicheng Chen

December 2019

Contributions

Chapter 3 is adapted from the following publication:

Chen, Y., Villanueva-Meyer, J. E., Morrison, M. A., & Lupo, J. M. (2019). Toward automatic detection of radiation-induced cerebral microbleeds using a 3D deep residual network. *Journal of digital imaging*, 32(5), 766-772.

<https://www.ncbi.nlm.nih.gov/pubmed/30511280>

Chapter 6 is adapted from the following publication:

Chen, Y., Jakary, A., Avadiappan, S., Hess, C. P., & Lupo, J. M. (2019). QSMGAN: Improved Quantitative Susceptibility Mapping using 3D Generative Adversarial Networks with Increased Receptive Field. *NeuroImage*, 116389.

<https://www.ncbi.nlm.nih.gov/pubmed/31760151>

Abstract

Development of Deep Learning Methods for Magnetic Resonance Phase Imaging of Neurological Disease

Yicheng Chen

Magnetic resonance imaging (MRI) is a high-resolution, non-invasive medical imaging modality that is widely used in human brain. In recent years, susceptibility weighted imaging (SWI) and quantitative susceptibility mapping (QSM) have been proposed to utilize MR phase signal to generate contrast from tissue magnetic susceptibility and even quantify the property. On the other hand, deep learning, especially deep convolutional neural networks (DCNNs), have achieved state-of-the-art performances in numerous computer vision tasks and gained significant attention in the field of medical imaging in the recent years. This dissertation combined the idea of deep learning with the two MR phase imaging methods.

To combined deep learning with SWI, we designed and trained a 3D deep residual network that can distinguish false positive detected candidates from cerebral microbleeds (CMBs) and built an automatic CMB detection pipeline with high performance. We further confirmed the generalizability of this deep learning-based pipeline using multiple datasets with different scan parameters and pathologies and provided lessons for application and generalization of generic deep learning based medical imaging methods.

To combine deep learning with QSM, we developed a 3D U-Net based network that learned to perform dipole inversion from gold standard QSM acquired from data with multiple orientations.

The model was further improved with an adversarial training strategy and achieved significantly lower reconstruction error than traditional QSM algorithms. In addition, we also performed various background removal and dipole inversion algorithms on both brain tumor patients and healthy volunteers to study and compare their performance. The results can provide guidance on future applications of QSM in different clinical scenarios.

Table of Contents

Chapter 1. Introduction	1
Chapter 2. Background	4
2.1. Basics of magnetic resonance imaging	4
2.1.1. Magnetic resonance physics and signals	4
2.1.2. Imaging principles and parallel imaging	6
2.2. Using phase MRI to image differences in magnetic susceptibility.....	9
2.2.1. Susceptibility weighted imaging	9
2.2.2. Foundations of quantitative susceptibility mapping	10
2.2.3. QSM processing methods.....	12
2.3. Clinical applications of magnetic susceptibility imaging.....	14
2.3.1. Clinical applications of SWI.....	14
2.3.2. Clinical applications of QSM	16
2.4. Deep convolutional neural networks	18
2.4.1. Machine learning	18
2.4.2. Neural networks and deep learning.....	20
2.4.3. Examples of Deep learning in MRI.....	28
Chapter 3. Automatic detection of cerebral microbleeds using 3D deep convolutional neural networks	29
3.1. Introduction.....	29
3.2. Methods.....	31

3.2.1.	Subjects and Image Acquisition	31
3.2.2.	SWI Processing	31
3.2.3.	Identification of Candidate CMBs & Labeling of False Positives	32
3.2.4.	3D Deep Residual Network.....	33
3.2.5.	Implementation	34
3.3.	Results.....	35
3.4.	Discussion and conclusions	41
Chapter 4.	Generalization and analysis of cerebral microbleed detection algorithm	44
4.1.	Introduction.....	44
4.2.	Methods.....	45
4.2.1.	DCNN-based algorithm.....	45
4.2.2.	Subjects and data acquisition	46
4.2.3.	Data processing and labeling	49
4.2.4.	Training and evaluation.....	49
4.3.	Results.....	52
4.3.1.	Direct application of pre-trained models.....	52
4.3.2.	The effects of fine-tuning and training using combined RT dataset	53
4.3.3.	Application of the algorithm to CCM.....	54
4.4.	Discussion and conclusions	55
4.4.1.	Generalization to 3T RT dataset	55
4.4.2.	Generalization to CCM.....	55

4.4.3.	Limitations	56
4.4.4.	Conclusion.....	57
Chapter 5.	Comparison of Quantitative Susceptibility Mapping Methods at 3T and 7T	58
5.1.	Introduction.....	58
5.2.	Methods.....	60
5.2.1.	Subjects and Image Acquisitions	60
5.2.2.	Image Reconstruction and Preprocessing.....	61
5.2.3.	QSM processing	62
5.2.4.	Image Analysis and Comparison.....	65
5.3.	Results.....	68
5.3.1.	Noise level	68
5.3.2.	White matter homogeneity	69
5.3.3.	Vein contrast.....	70
5.3.4.	CMB Contrast	72
5.3.5.	Whole-brain QSM metrics	73
5.3.6.	Basal Ganglia ROIs	74
5.4.	Discussion and conclusions	76
Chapter 6.	Improved Quantitative Susceptibility Mapping using 3D Generative Adversarial Networks	81
6.1.	Introduction.....	81
6.2.	Methods.....	82

6.2.1.	Theory of QSM dipole inversion and GANs.....	82
6.2.2.	QSMGAN framework	84
6.2.3.	Subjects and data acquisition	85
6.2.4.	QSM data processing and dataset preparation	86
6.2.5.	Training and validation.....	87
6.2.6.	Evaluation metrics	90
6.3.	Results.....	91
6.3.1.	Baseline 3D U-Net.....	91
6.3.2.	Effectiveness of QSMGAN.....	94
6.3.3.	Comparison with non-learning-based methods	96
6.3.4.	Application of networks in patients with radiation-induced CMBs	97
6.4.	Discussion and conclusions	99
Chapter 7.	Conclusions and future directions	103
7.1.	Conclusions	103
7.2.	Future directions	104
References	105

List of Figures

Figure 2.1 Excitation of magnetization.	5
Figure 2.2 1D signal encoding.....	7
Figure 2.3 2D signal encoding.....	7
Figure 2.4 GRAPPA.....	8
Figure 2.5 Processing pipeline of SWI.	9
Figure 2.6 A typical QSM processing pipeline.	12
Figure 2.7 Zero cone region in QSM inverse problem.	14
Figure 2.8 Cerebral microbleeds on SWI.....	15
Figure 2.9 Depictions of deep brain nuclei on various MRI methods.	17
Figure 2.10 A 2-layer feedforward neural network.....	21
Figure 2.11 Sigmoid function.....	21
Figure 2.12 ReLU function.	22
Figure 2.13 2D convolution.	23
Figure 2.14 2D pooling.....	23
Figure 2.15 LeNet-5.....	24
Figure 2.16 AlexNet	24
Figure 2.17 Skip connection.....	25
Figure 2.18 U-Net.	26
Figure 2.19 Mechanism of GAN.	27
Figure 2.20 DCGAN.....	27

Figure 3.1 CMBs on SWI.	33
Figure 3.2 Processing pipeline and network architecture.	34
Figure 3.3 AUC comparison of different models and data augmentations.	37
Figure 3.4 Likelihood scores of representative candidates from the test set overlaid on SWI.	39
Figure 3.5 Example of classification results for representative CMBs in the test subjects.....	40
Figure 3.6 The likelihood score predicted by the network compared to a neuroradiologist’s scoring of all candidates.....	41
Figure 4.1 Appearance of CMBs on SWI axial slices.....	44
Figure 4.2 Pipeline of automatic CMB detection algorithm.	46
Figure 4.3 Training strategies for the deep convolutional neural network using different datasets and weight initializations.	50
Figure 4.4 Comparison of ROC curves of different models.....	52
Figure 4.5 Examples of false positive candidates in 3T RT test subjects.....	54
Figure 5.1 QSM processing pipeline adopted in this study.	59
Figure 5.2 Definition of metrics used for brain tumor patients with microbleeds.....	66
Figure 5.3 Segmentation of basal ganglia ROIs of a healthy volunteer acquired by applying the QSM atlas.	68
Figure 5.4 Boxplots of noise level at 3T and 7T.....	69
Figure 5.5 Boxplots of white matter homogeneity.....	70
Figure 5.6 Boxplots of vein contrast.	71
Figure 5.7 Visual comparison of a vein of different algorithms at 3T and 7T.....	71

Figure 5.8 Boxplots of CMB contrast.....	72
Figure 5.9 Visual comparison of CMBs of different algorithms at 3T and 7T.	73
Figure 5.10 Visual comparison of QSM dipole inversion algorithms of a healthy volunteer (subject #5).	74
Figure 5.11 Comparison of mean susceptibility within basal ganglia ROIs using different dipole inversion algorithms.	75
Figure 5.12 Scatter plot and regression plot.	76
Figure 6.1 QSMGAN network architecture.	85
Figure 6.2 QSM data processing pipeline.....	87
Figure 6.3 Demonstration of the relationship between receptive field and input/output patch size.	88
Figure 6.4 Comparison of raw QSM and tanh transformed QSM.	89
Figure 6.5 Comparison of reconstructed QSM using 3D U-Net with different input/output patch sizes.	93
Figure 6.6 Comparison of QSM of test subject 1.....	95
Figure 6.7 Comparison of QSM of test subject 2.	96
Figure 6.8 Comparison of median CMB susceptibilities measured from different QSM algorithms.	98
Figure 6.9 QSM of two patients with brain tumors who had developed cerebral microbleeds due to prior radiation therapy.	99

List of Tables

Table 3.1 Number of patches used in all dataset splits.....	33
Table 3.2 Confusion matrix of the classification results of the test set.....	38
Table 4.1 Comparison of performance of the 3D-DRNN on different datasets.	48
Table 4.2 Comparison of performance of the 3D-DRNN on different datasets.	51
Table 5.1 Summary of selected QSM algorithms.	64
Table 5.2 Comparison of whole-brain metrics from 2016 QSM challenge of different dipole inversion algorithms.....	74
Table 6.1 Test set performance of U-Net baseline with different input and output patch sizes. .	92
Table 6.2 Test set performance of U-Net baseline, QSMGAN and non-learning-based algorithms.	93

Chapter 1. Introduction

Magnetic resonance imaging (MRI) is a powerful non-invasive medical imaging modality that is widely used due to its flexibility, non-radioactivity, high spatial resolution and tunable contrast for soft tissue. MRI has been used in various clinical applications, especially diseases related to the human brain, including different types of cancers, neurodegenerative diseases, ischemic stroke, and many others.

Currently, most clinical MRI applications acquire information and generate contrast from the difference in signal magnitude. However, the MRI signal is inherently complex-valued, and the phase of the signal also carries useful information that can be utilized to perform quantitative imaging. Among various sources that lead to differences in MR phase images, magnetic susceptibility is the property gaining most interest due to its clinical relevance and potential to be a biomarker.

Susceptibility weighted imaging (SWI) is a MR phase imaging technique that utilizes the $T2^*$ phase by filtering it in order to further weight the magnitude image to highlight susceptibility differences. SWI has been widely used in clinical applications, especially neurovascular diseases. However, one major limitation of SWI is that it is qualitative. Quantitative susceptibility mapping (QSM) is a recent method to quantify the distribution of magnetic susceptibility from the MR phase signal.

In recent years, deep learning is an emerging branch of machine learning developed to tackle computer vision problems such as image classification, image segmentation, and object detection. Deep learning has also gained much attention in the field of MRI due to its performance and great potential. It has been shown that deep learning could facilitate clinical MRI in various tasks, from signal acquisition, to image reconstruction, to image post-processing and assisted diagnosis.

In this dissertation, we investigated the techniques of susceptibility imaging using MR, developed deep learning methods combined with these techniques, and analyzed them on clinical applications:

Chapter 2 introduced the background information for this dissertation. First, the basic principles of MRI are described, then the methods of SWI and QSM and their clinical applications are introduced. Finally, this chapter includes an overview of machine learning and the development of deep learning.

Chapter 3 presents the development of a novel 3D deep convolutional neural network (DCNN) for automatic detection of radiotherapy-induced cerebral microbleeds (CMBs) based on SWI images. The technique significantly reduced the number of false positives and nearly automates the CMB detection and segmentation process.

Chapter 4 investigates the automatic CMB detection method developed in Chapter 3 in further detail by applying it on different datasets to test its generalizability. We found that the model could be well generalized to data with different scan parameters and pathology without severe overfitting. This project could also inform generic medical imaging tasks using deep learning on the development and optimization of deep neural networks.

Chapter 5 compares different background removal and dipole inversion algorithms for QSM processing and evaluated the methods on volunteers and patients with radiation-induced CMBs using various metrics, including noise, white matter homogeneity, vein and CMB contrast, whole-brain image metrics and basal ganglia ROI metrics. This chapter could provide guidance for future application of QSM in different pathologies.

Chapter 6 develops a learning-based QSM dipole inversion algorithm by training a generative adversarial network (GAN) with 3D U-Net architecture. This new method supersedes the traditional dipole inversion algorithms in accuracy, robustness and speed.

Chapter 7 summarizes the work presented in this dissertation and outlines the future directions of related research topics.

Chapter 2. Background

2.1. Basics of magnetic resonance imaging

2.1.1. Magnetic resonance physics and signals

Magnetic resonance imaging or MRI, is a tomographic imaging technique that produces images of internal physical and chemical characteristics of an object from externally measured nuclear magnetic resonance signals. [1] Although a rigorous and accurate description of the basic MR physics requires quantum mechanics, the classical treatment is sufficient to describe the macroscopic behaviors in most cases, including the scope of this dissertation. Therefore, we will introduce the classical description of MR and MR physics in this section.

Atoms with odd numbers of protons/neutrons (for example the hydrogen ^1H in water, most abundant in biological specimens) have spin angular momentum. [2] In a strong static magnetic field \vec{B}_0 (usually 0.5~7 Tesla in modern medical MRI scanners), spins tend to align with \vec{B}_0 and generate a net magnetization. At a certain frequency called Larmor frequency, f , the nuclear spins exhibit resonance. The Larmor frequency relates to the applied magnetic field by the following equation:

$$f = \frac{\gamma}{2\pi} B \quad (2.1)$$

where γ is gyromagnetic ratio and for ^1H atoms, $\gamma/2\pi = 42.58\text{MHz/T}$.

Applying a radiofrequency magnetic field \vec{B}_1 at the Larmor frequency calculated using Equation 2.1 will excite the spin/magnetization. Following the excitation, the magnetization will precess about the direction of the applied field and create a transverse component M_{xy} of the magnetization as Figure 2.1 demonstrates. According to Faraday's law of induction, the oscillation results in a signal called a free induction decay (FID) and represents the basic MR signal to be recorded.

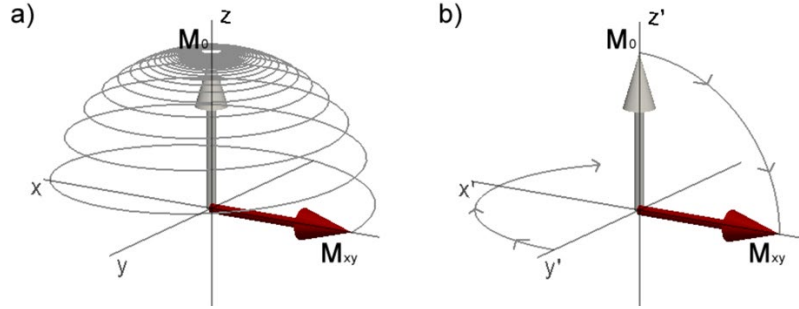


Figure 2.1 Excitation of magnetization.
 (a) in the lab frame. (b) in the rotation frame of Larmor frequency.

Following the excitation, the transverse component of the magnetization decays away and the longitudinal component returns to its thermal equilibrium. Classically, these behaviors can be described by:

- Longitudinal relaxation

$$\frac{dM_z}{dt} = -\frac{M_z - M_0}{T_1} \quad (2.2)$$

- Transverse relaxation

$$\frac{dM_{xy}}{dt} = -\frac{M_{xy}}{T_2} \quad (2.3)$$

Where M_0 is the equilibrium magnetization, T_1 is spin-lattice time constant, and T_2 is the spin-spin time constant. [3] These time constants vary among different tissue components and are used to generate image contrast in MRI.

More generically, the dynamics of nuclear magnetization is described by the Bloch equation: [2]

$$\frac{d\vec{M}}{dt} = \vec{M} \times \gamma \vec{B} - \frac{M_x \vec{i} + M_y \vec{j}}{T_2} - \frac{(M_z - M_0) \vec{k}}{T_1} \quad (2.4)$$

where \vec{i} , \vec{j} and \vec{k} are unit vectors in the x , y and z directions.

A compact form of the general solution for the transverse component at position \vec{r} and time t for an inhomogeneous object and nonuniform field $\Delta\omega(\vec{r}, \tau)$ is:

$$M(\vec{r}, t) = M^0(\vec{r})e^{-t/T2(\vec{r})}e^{-i\omega_0 t} \exp\left(-i \int_0^t \Delta\omega(\vec{r}, \tau) d\tau\right) \quad (2.5)$$

The total signal received is the sum of all magnetization in the imaging field of view (FOV):

$$s_r(t) = \int_x \int_y \int_z M(x, y, z, t) dx dy dz \quad (2.6)$$

With the following simplifications and application of time-varying gradient magnetic fields $\vec{G}(\tau)$ that:

- 1) ignore the relaxation term $e^{-t/T2(\vec{r})}$,
- 2) assume a plane centered at $z = z_0$ of thickness Δz and let $m(x, y) =$

$$\int_{z_0 - \Delta z/2}^{z_0 + \Delta z/2} M_0(x, y, z) dz,$$

- 3) drop the $e^{-i\omega_0 t}$ by demodulation,

the signal equation 2.6 can be simplified as:

$$s(t) = \int_x \int_y m(x, y) e^{-i2\pi[k_x(t)x + k_y(t)y]} dx dy \quad (2.7)$$

where $k_x(t) = \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau$ and $k_y(t) = \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau$.

2.1.2. Imaging principles and parallel imaging

The simplified signal equation 2.7 has the form of a 2D Fourier transform of $m(x, y)$. Because k_x and k_y are integrals of time-varying gradient magnetic fields \vec{G} , by applying a certain gradient field, we can acquire the signal to cover the Fourier transform space (or k-space) of the image. A most commonly used signal encoding method is demonstrated in Figure 2.2. By applying the G_x on the middle, we can acquire the signal s that corresponds to the k-space data shown on the right.

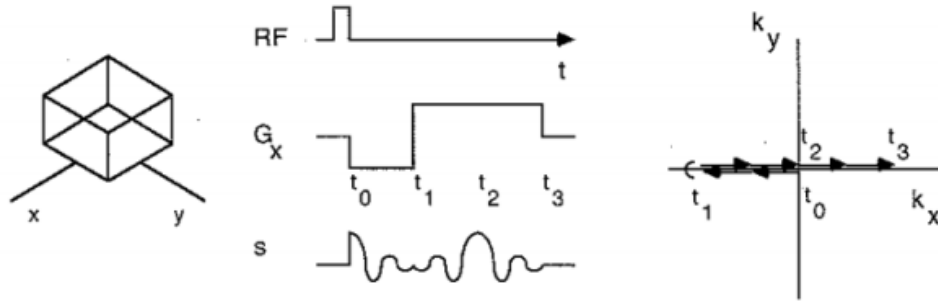


Figure 2.2 1D signal encoding.

Left: an example object to be imaged. Middle: the imaging sequence. Right: k-space trajectory.

Applying different gradient fields in the y direction as shown in Figure 2.3 traverses different positions in k -space. Once we acquired enough k -space data points according to the Nyquist sampling theorem [4], we can reconstruct the image/object by inverse 2D Fourier transform.

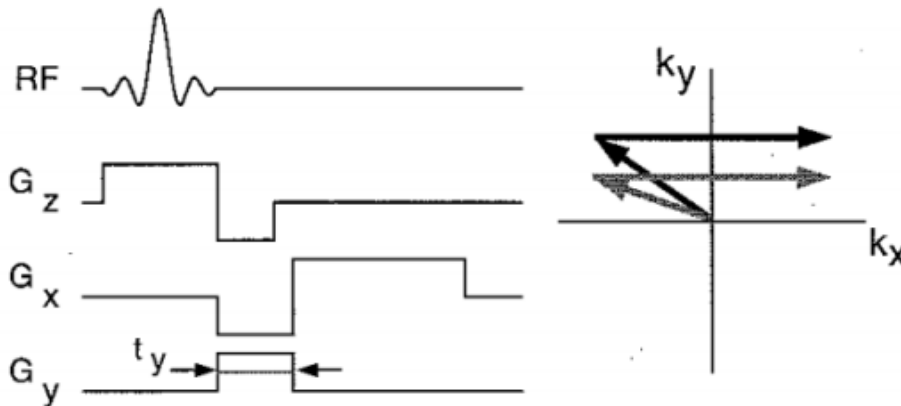


Figure 2.3 2D signal encoding.

Left: 2D imaging sequence. Right: k-space trajectories.

Although the above is a simple example of MRI signal encoding in the 2D scenario, many other k -space trajectories have been developed such as radial [5] and spiral [6] trajectories, that can also be easily extended to 3D by adding a gradient field along the z -direction.

Despite the flexibility of MRI signal encoding, it is also time consuming because of the phase-encoding step [1], which applies a certain G_y to sample one line of k-space according to the Nyquist sampling theorem as shown in the right panel of Figure 2.3. A method to accelerate this scan time is parallel imaging [7], [8], which acquires the signal using multiple receiver coils (sometimes called phased-array) at different spatial locations and “unwraps” the aliasing due to reduced phase-encoding steps by incorporating the spatial sensitivity information of the coils. Nowadays, a widely implemented and commonly used parallel imaging method in clinical MRI to accelerate the image acquisition is called Generalized autocalibrating partially parallel acquisitions or GRAPPA [7]. In GRAPPA, lines through the center part of the k-space are fully sampled and constitute the autocalibration signal (ACS) region. These fully sampled k-space data are used to calculate weighting factors for each coil. The weighting factors reflect how each coil distorts or displaces spatial frequencies within the full field-of-view k-space. In the outer region of k-space, missing or skipped k-space lines are estimated using the calibrated weighting factors combined with local known data for each small region (known as a block or kernel). Figure 2.4 illustrates how the missing k-space data points are filled.

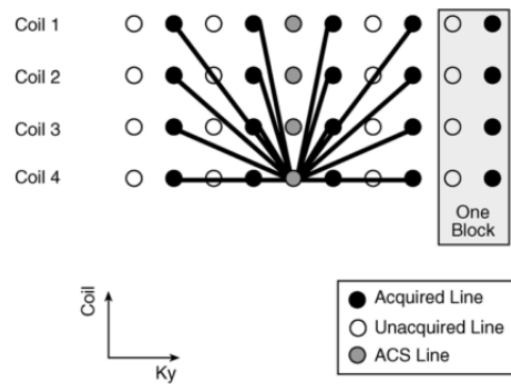
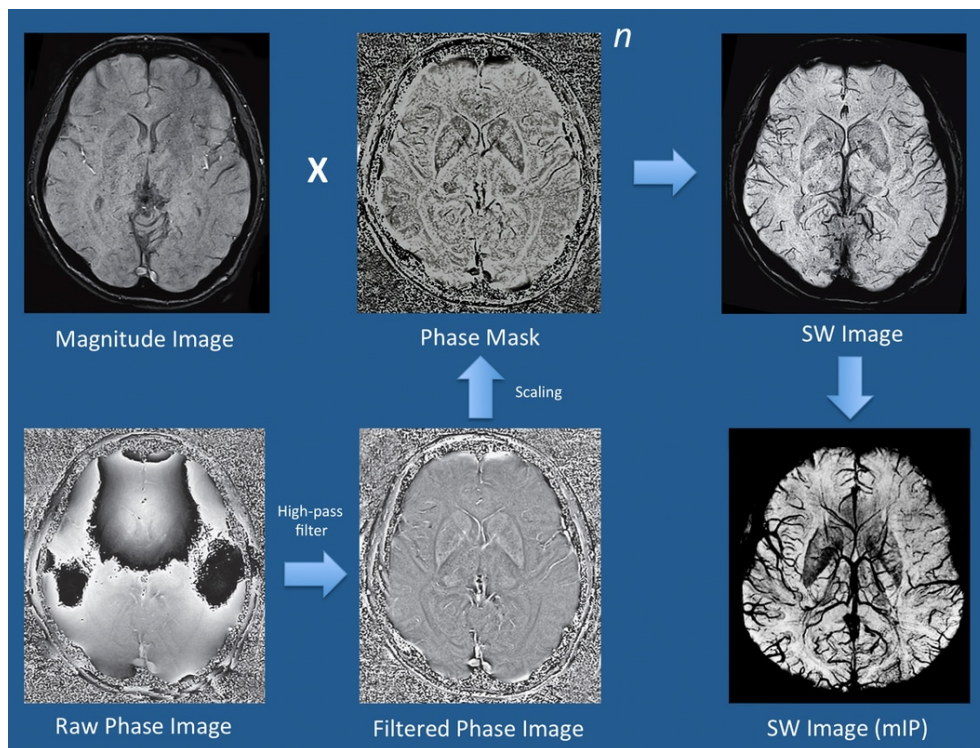


Figure 2.4 GRAPPA.

2.2. Using phase MRI to image differences in magnetic susceptibility

2.2.1. Susceptibility weighted imaging

Magnetic susceptibility is an intrinsic physical property that measures how much a material or tissue will become magnetized in an applied magnetic field. When being imaged in MRI, tissues themselves will generate an induced magnetic field in addition to the original field and thus cause field inhomogeneity and affect the surrounding tissue. [2] This field inhomogeneity leads to dephasing and generates T2*-weighted contrast on magnitude images when using gradient echo sequences in MRI. It can be also combined with the phase information to create useful contrast by a technique called susceptibility-weighted imaging or SWI. [9]–[11]



*Figure 2.5 Processing pipeline of SWI.
(adapted from <http://mriquestions.com/making-an-sw-image.html>)*

A typical processing pipeline to reconstruct SWI from gradient-echo sequences is illustrated in Figure 2.5. The raw phase image as well as the magnitude image are acquired. The raw phase

is then high-pass filtered to remove the low frequency phase fluctuations and phase wraps. Next, a phase mask is calculated that scales data from the filtered phase image over a 0-1 range to enhance tissues with negative susceptibilities. This phase mask is then element-wisely multiplied with the magnitude image for several times until the desired mix of phase information is incorporated. [10] In order to better visualize vascular structures, a minimum intensity projection (mIP) is typically performed on the 3D SWI image volume. [12] The SWI processing can be extended to multi-echo GRE sequences by applying filters that are adjusted according to the TE on phase images and then taking a weighted combination of the SWI images from each individual echo to obtain the final SWI [13].

2.2.2. Foundations of quantitative susceptibility mapping

Although SWI utilizes the phase induced by magnetic susceptibility and generates high contrast with different tissue susceptibilities, it is inherently only a qualitative imaging method. Recently, quantitative susceptibility mapping (QSM) [14] is an emerging MR technique that can quantify the intrinsic magnetic susceptibility from gradient-echo phases and provide useful information for various clinical applications.

In the presence of an applied magnetic field \vec{H} , an isotropic, non-ferromagnetic material with susceptibility χ will generate a magnetization $\vec{M} = \chi \cdot \vec{H}$. The distribution of $\vec{M}(\vec{r})$ causes the field variation: [15]

$$\Delta \vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \int \int \int \nabla \times \left[\frac{\vec{M}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \right] d^3 r' \quad (2.8)$$

Assuming that the main field \vec{B}_0 is oriented in the z direction and using the vector identity, the above equation can be simplified to:

$$\Delta B_z(\vec{r}) \approx \frac{\mu_0}{4\pi} \int \int \int \left[\frac{3M_z(\vec{r}') (z - z')}{|\vec{r} - \vec{r}'|^5} (z - z') - \frac{M_z(\vec{r}')}{|\vec{r} - \vec{r}'|^3} \right] d^3\vec{r}' \quad (2.9)$$

This equation can be written as a convolution between $M_z(\vec{r})$ and the z-component of the point-dipole response $d_z(\vec{r}) = \frac{1}{4\pi} \frac{3 \cos^2(\theta) - 1}{|\vec{r}|^3}$ as:

$$\Delta B_z(\vec{r}) = \mu_0 \int M_z(\vec{r}') d_z(\vec{r} - \vec{r}') d^3\vec{r}' \quad (2.10)$$

In $d_z(\vec{r})$, θ is the angle between the z direction and \vec{r} . With a condition that $\chi \ll 1$, $M_z(\vec{r}) \approx \chi(\vec{r})B_0/\mu_0$, the above equation can be written in the following form:

$$\Delta B_z(\vec{r}) = B_0 \int \chi(\vec{r}') d_z(\vec{r} - \vec{r}') d^3\vec{r}' \quad (2.11)$$

which in the frequency domain takes the form:

$$\Delta B_z(\vec{k}) = [\chi(\vec{k}) \cdot d_z(\vec{k})] B_0 \quad (2.12)$$

where

$$d_z(\vec{k}) = \frac{1}{3} - \frac{k_z^2}{|\vec{k}|^2} \quad (2.13)$$

In frequency domain, the convolution becomes a more computationally efficient element-wise multiplication.

From the principles of MRI we know that the measured local frequency $f(\vec{r}) = \frac{\gamma}{2\pi} [b_0 + \Delta B_z(\vec{r})]$.

In GRE sequences, the frequency variation between the local resonance frequency and the scanner's reference frequency caused by $\Delta B_z(\vec{r})$ accumulates as phase. Thus, by measuring the phase we can invert the above equations to map the susceptibility distribution $\chi(\vec{r})$.

The above representation is a simple phase-susceptibility model where the magnetization is treated as a magnetic dipole. Recent studies have revealed more complex properties of magnetization and susceptibility, leading to the development of susceptibility tensor imaging

(STI) [16]. STI is beyond the scope of this dissertation, and a detailed explanation could be found in [15].

2.2.3. QSM processing methods

Although the link between tissue phase and intrinsic susceptibility seems clear and simple in equation 2.12, in practice, careful processing is required to correctly solve for the susceptibility distribution from measured total phase. A typical QSM pipeline is illustrated in Figure 2.6.

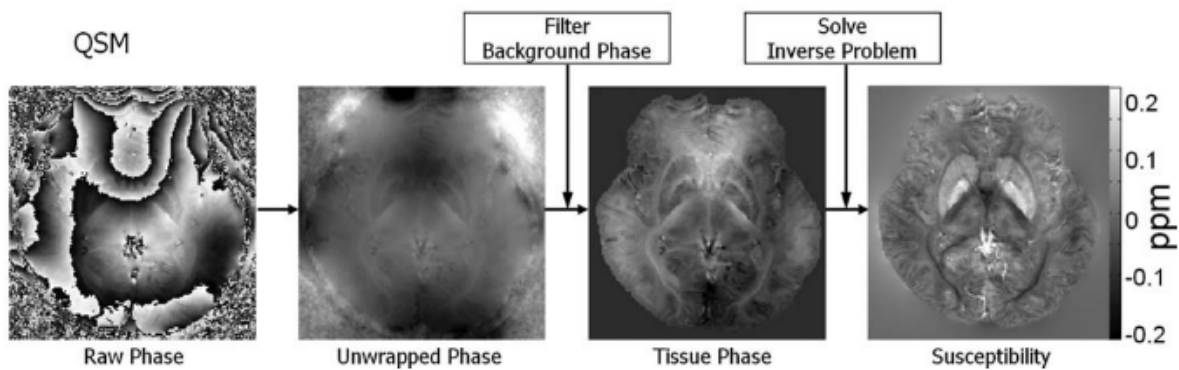


Figure 2.6 A typical QSM processing pipeline. (Adapted from [9])

The raw phase acquired from a GRE sequence is within the range of $[-\pi, \pi]$; any phase signal outside this range gets wrapped and 2π phase jumps occur as the first image in Figure 2.6 demonstrates. Therefore, usually we first need to remove the phase aliasing, or “unwrap” the raw phase to restore the full-range phase map. The phase unwrapping step can be performed in either the spatial domain or frequency domain. In the frequency domain, path following, minimum norm solving, or filtering methods [17] have been developed. In the spatial domain, a Laplacian-based phase unwrapping has been proposed by [18] and it is more frequently used in QSM due to its numerical simplicity and robustness to noise.

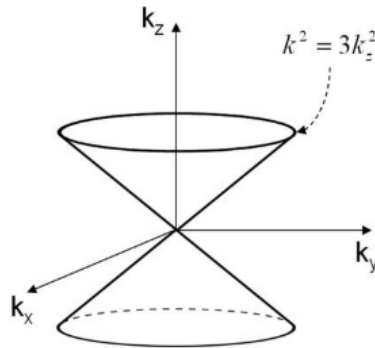
The next step in QSM is usually called background field removal. The magnetic field perturbation in a given region-of-interest (ROI), e.g. the brain, not only originates from the susceptibility sources inside of the ROI, but also includes added susceptibilities outside of the ROI such as the air-tissue interface in the sinuses, the human torso, main field inhomogeneities, and shimming fields. These sources outside of the ROI (also known as background field) are much larger than the tissue susceptibilities we are interested in, confounding the solution of QSM. Therefore, the background field is usually removed before solving the linear equation 2.12. Various approaches have been developed to separate the background field from the total field. [19] A common solution to this problem currently involves solving the Laplacian equation using the spherical mean value property [20] of the background field or by assuming a boundary value [21]. Chapter 5 introduces some of these methods and compares their performances.

The final step of QSM is to solve the linear equation in 2.12, also known as dipole inversion.

The major challenge in directly inverting the equation to solve for χ is that when $3k_z^2 \approx |\vec{k}|^2$, $d_z(\vec{k})$ is close to zero, making the linear equation an ill-posed inverse problem. In the frequency domain, the zero coefficients locate on a double-cone region as Figure 2.7 demonstrates. A theoretically feasible solution to this ill-posed issue is to acquire data N times where the subject has different orientations relative to the main field and fill in the missing data near or at the zero cone. The computational solution for this multi-orientation scheme called Calculation of Susceptibility Through Multiple Orientation Sampling (COSMOS) was proposed in [22].

However, it is rarely used in practice because it requires $N \geq 3$ repeated acquisition of the same sequence and changing the subject's orientation between scans, greatly prolonging the total scan time. Therefore, modern QSM dipole inversion algorithms are usually based on a single acquisition and approach the ill-posed problem by supplementing additional information through regularization [15]. A simple and efficient idea is to threshold the dipole kernel $d_z(\vec{k})$ and replace the coefficients close to zero. This method is called thresholded k-space division (TKD)

[23]. Morphological-enabled dipole inversion (MEDI) [24] incorporates prior information from the magnitude image, assuming the anatomical structure have similar edges. More recent algorithms are introduced and compared in Chapter 5.



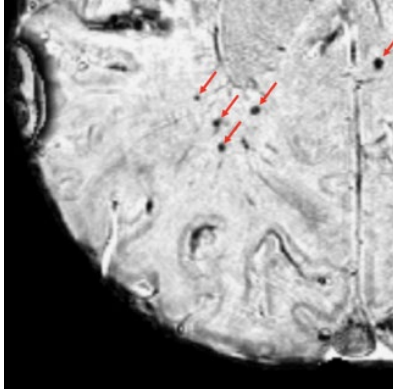
*Figure 2.7 Zero cone region in QSM inverse problem.
(Adapted from [9])*

2.3. Clinical applications of magnetic susceptibility imaging

2.3.1. Clinical applications of SWI

The unique contrast of SWI has played an important role in many clinical applications, especially neurovascular diseases including ischemic stroke, cerebral amyloid angiopathy, cerebral cavernous malformations [25] and cerebral microbleeds (CMB) [10]. CMBs are defined as small foci of chronic blood products in normal or near normal brain tissue. These small lesions are prevalent in various diseases such as stroke, neurodegenerative disorders, traumatic brain injury, and irradiated brain tumors and can serve as biomarkers of complications in these pathologies.

Due to its sensitivity to tissue of different susceptibilities and high spatial resolution, SWI has been shown to be much more sensitive than standard GRE sequence in detecting cerebral microbleeds. On SWI, CMBs usually appear as round hypointense lesions with high contrast with normal brain tissues, as Figure 2.8 illustrates.



*Figure 2.8 Cerebral microbleeds on SWI.
CMBs highlighted by red arrows.*

In order to understand the impact of CMBs on various diseases, large-scale detection and quantification of the lesion burden is an essential step. [26] Traditionally, this process was conducted manually by trained radiologists by visually examining the SWI images. However, despite the high contrast SWI provides on CMBs, visual inspection of CMBs is difficult and labor-intensive due to their small sizes and wide distribution across the entire brain. Another disadvantage of manual labeling is that it requires domain-related expertise. As a result, various computer-aided detection methods have been developed to automate the CMB detection problem. [27] assumed the Gaussian distribution of the background normal tissue in a small region and proposed a method of intensity-based local statistical thresholding that detects CMBs as outliers. The false positive detections are then further removed by a support vector machine based on manually designed features. [28] adapted the fast radial symmetry transform (FRST) that could enhance objects with spherical geometry and improved detection accuracy. [29] proposed a fully-automated segmentation algorithm based on 2D FRST and combined it with a user-guided CMB classification tool to further reduce false positive detections, achieving a 5-fold reduction in the user's time spent on evaluating CMB burden.

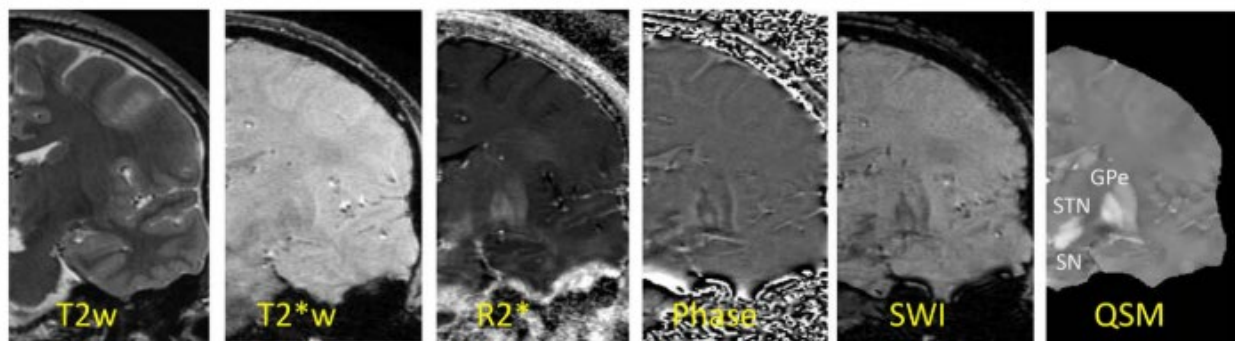
In recent years, deep convolutional neural networks (DCNNs) have been applied to CMB detection [30]–[32] and achieved better performances than traditional methods and models based on image processing and hand-crafted features. Chapter 3 of this dissertation describes one approach to this problem. Details of DCNN and deep learning are introduced in Chapter 2.4.

2.3.2. Clinical applications of QSM

A wide range of physiologically important molecules could change the tissue magnetic susceptibility and be quantified using QSM. Among them, iron is the most common component investigated through QSM [33]. Two major sources that contribute to iron content changes are iron deposition and blood components.

Iron deposition is thought to be linked with many neurodegenerative diseases such as Alzheimer's disease (AD) [34], Parkinson's disease (PD) [35] and Huntington's disease (HD) [36]. Although it is unknown whether the excessive iron accumulation is the cause or the consequence of the disease, monitoring the spatial deposition and temporal changes of iron could help understand the pathology of these diseases [34]. In AD, increased iron milieu has been suggested to facilitate the proliferation and perpetuation of β -amyloid [37], a protein thought to be toxic and influences the development of AD. [34] used QSM to compared the susceptibility in deep brain nuclei in a cross-sectional study and found that AD patients have significantly higher susceptibility in putamen, posterior grey and white matter regions. In PD, it is known that iron deposition occurs when dopaminergic neurons in substantia nigra (SN) dies, which associate with the disease. [35] found an increase in susceptibility of the pars compacta in PD patients compared with healthy controls, and a rostral-caudal gradient in susceptibility in both patients and controls. [38] demonstrated that QSM has a higher sensitivity and specificity

than R2 and R2* MRI in differentiating the PD patients and healthy controls. In HD, elevated iron deposition has been found in the striatum, which is believed to activate free radicals and cause neuronal loss [39]. Detection of iron increase could enable earlier diagnosis of HD. [36] demonstrated that QSM could detect increased iron levels in caudate nucleus, putamen and globus pallidus in premanifest HD patients and the measured susceptibility is correlated with genetic burden scores. [40] also found that the iron deposition measured by QSM is significantly associated with HD severity.



*Figure 2.9 Depictions of deep brain nuclei on various MRI methods.
(Adapted from <https://www.ncbi.nlm.nih.gov/pubmed/23674786>)*

Although T2*-weighted MRI and SWI are sensitive to intracerebral hemorrhage, it suffers from blooming artifacts that will enlarge the volume of the lesion and highly dependent on the imaging parameters. QSM has become a more accurate technique to quantify both the magnetic susceptibility and volume of brain hemorrhages [41].(Figure 2.9) Since calcification is diamagnetic and hemosiderin is paramagnetic [42], QSM could easily differentiate them. [43] demonstrated that QSM achieved significantly better detection performance than GRE phase image in diagnosis of hemorrhage or calcium. [44] applied QSM to detection and measurement of cerebral microbleeds.

2.4. Deep convolutional neural networks

2.4.1. Machine learning

Computers can solve many problems by algorithms. An algorithm is defined as a sequence of instructions to be carried out to transform the input to output. [45] However, for some tasks we do not have an algorithm, or it is very difficult to come up with an algorithm that works well – for example, to distinguish spam emails from legitimate emails [46], or to detect and classify brain tumors on MR images. [47] In these cases, the technique of machine learning has been proved to be powerful. Machine learning uses the theory of statistics to build mathematical models that could “learn” from examples (data). [45] In other words, we train the computer/machine to automatically extract the algorithm for the task. A more formal definition of machine learning is described in [48]:

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

As an example, consider building a deep convolutional neural network (a modern deep learning model, see following section for details) to distinguish pictures of cats and dogs (task T) and measure the performance P by the percentage of pictures that are correctly classified, or classification accuracy. We can feed the network with thousands of cat and dog images (experience E) and then train the model to classify these example pictures using proper methods. After exposure to enough samples, we will find that the model can classify the pictures with higher accuracy.

Common types of machine learning algorithms include supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning refers to the approach of building a model with a set of data containing both the inputs X and the correct outputs or labels Y . [49] The algorithm learns a mapping from the input to the output. Assume the model is defined up to a set of parameters θ :

$$y = g(x|\theta) \quad (2.14)$$

where $g(\cdot)$ is the model function, regression (output continuous values), or classification (output discrete values). Supervised learning algorithms optimize the parameters θ such that the approximation error of the model is minimized. That is, our estimates are as close as possible to the correct values given by the training dataset.

On the other hand, in unsupervised learning, no correct outputs or labels Y are given in the dataset and we only have the input data. The aim of unsupervised learning is to find the regularities or unknown patterns in the input data. It is also known as self-organization and allows modeling probability densities of given inputs. [50] Principle component and clustering are two main methods used in unsupervised learning. Generative adversarial network (GAN) is a modern approach developed for unsupervised learning based on neural networks and game theory. [51] We will introduce the principles of GAN in the following section.

Reinforcement learning (RL) applies to the cases where the desired output is a sequence of actions in an environment. The target of RL algorithms is to learn a policy that can conduct the sequence of correct actions to reach the goal or optimize some notion of cumulative reward. [52] Although RL is beyond the scope of this dissertation, it is gaining research interest in the field of MRI. [53], [54]

The central component of a machine learning algorithm is the model. Various types of models have been developed and used for machine learning system, some of the useful categories are:

- Support vector machines (SVM) [55]
- Decision trees [56]

- Neural networks [57]
- K-means clustering [45]

2.4.2. Neural networks and deep learning

2.1.1.1. Feedforward neural networks

In this section, we will take an in-depth review in (artificial) neural networks (NN) because it is the essential backbone of deep learning [58], a modern branch of machine learning that is extensively utilized in this dissertation.

The human brain is an excellent information processing device in many domains such as vision, speech recognition, and learning. [45] If we can understand how the brain performs these functions, we can formulate the solutions as computer algorithms and implement them on machines. This idea inspired the first artificial neural networks models. An NN contains a collection of connected units called artificial neurons, which are simplified models of neurons in the biological brain. Each connection mimics the synapses in the biological brain and transmits a signal to other neurons. The neuron that receives these signals processes them by computation and then signals other neurons. The most commonly used NN model architecture in computer vision is a feedforward neural network, in which connections between the nodes do not form a cycle. A simple 2-layer feedforward neural network is shown in Figure 2.10. The red layer on the left receiving input data is the input layer, the green layer on the right producing the final result is the output layer. The blue layer in between conducts intermediate processing and is called a hidden layer. A network with more than one hidden layer is usually called a deep neural network and is an important category of models in deep learning. [59]

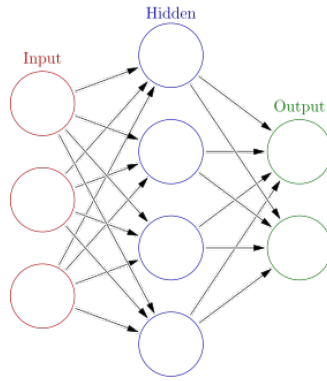


Figure 2.10 A 2-layer feedforward neural network.

Linear combination is commonly used in NN neurons. Mathematically, the neural network in Figure 2.10 with linear combination can be represented as:

$$y_k(x, w) = \sigma \left(\sum_{j=0}^M w_{kj}^{(2)} h \left(\sum_{i=1}^D w_{ji}^{(1)} x_i \right) \right) \quad (2.15)$$

Where x, y are input and output of the network, w is the set of linear combination weights, D, M are number of neurons in the hidden layer and output layer respectively, $\sigma(\cdot)$ and $h(\cdot)$ are nonlinear functions called activation functions that enable the network to compute nontrivial (nonlinear) problems. [60] Some commonly used activation functions are:

- Sigmoid function: $s(x) = \frac{1}{1+e^{-x}}$
- Rectified linear unit (ReLU): $s(x) = \max(0, x)$

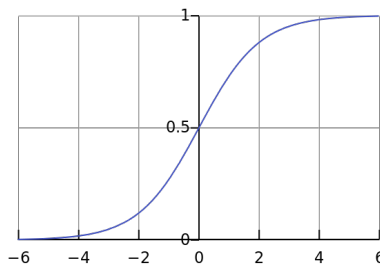


Figure 2.11 Sigmoid function.

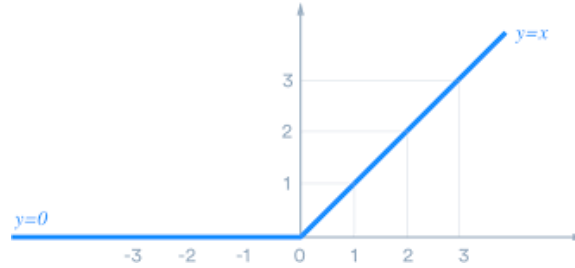


Figure 2.12 ReLU function.

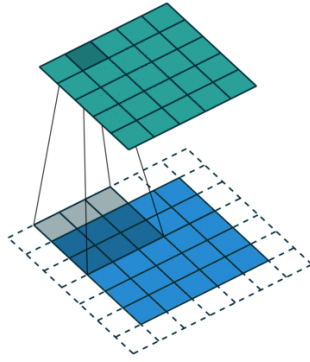
2.1.1.2. Convolutional neural networks

When applied to computer vision tasks that involve understanding images, a few limitations of traditional feedforward NN emerges: 1) An image usually contains thousands even millions of pixels. Directly feeding the pixel values to a fully connected NN (each neuron is connected with all neurons in the previous layer) dramatically increases the number of weights w of the network and increases the risk of overfitting. [61] 2) The feed forward network could not capture the spatial dependencies and shift invariance in images.

Convolutional Neural Network (CNN) [62] is a class of deep neural networks that was developed to mitigate the above problems in image-related learning tasks. The name “convolutional” indicates that the network employs a mathematical operation called a convolution. CNNs are simply neural networks that use convolutions in place of general linear combinations in at least one of their layers. [59]

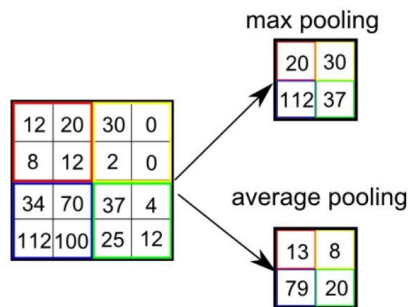
Convolution is a specialized kind of linear operation defined as follows. Consider the discrete convolution between a two-dimensional image I and a two-dimensional matrix K (also known as the kernel); the operation can be represented using the following equation (visually illustrated in Figure 2.13):

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \#(2.16)$$



*Figure 2.13 2D convolution.
Blue squares: convolution input, 3x3 gray blocks: convolutional kernel; green blocks: convolution output*

Another critical component of CNN is pooling layers. Pooling layers reduce the spatial dimensions of the input data by combining outputs of neurons clusters at one layer into a single neuron in the next layer. [63] There are two basic types of pooling layers: local pooling combines small clusters, typically 2x2 blocks. Global pooling acts on all the neurons of a layer and aggregate them into a single output. Figure 2.14 demonstrates two commonly used pooling: max pooling and average pooling.



*Figure 2.14 2D pooling.
Max pooling takes the maximum value inside the cluster, and average pooling calculates the average value.*

A pioneering CNN is LeNet-5 proposed by [62] in 1998. It contains 7 layers and achieved state-of-the-art performance on recognizing 32x32 digital images of hand-written digits (Figure 2.15).

layer learns its weights. This simple modification of a CNN architecture enables the ability to train much deeper CNNs and significantly improves the model capacity. Since the emergence of ResNet, the idea of the skip connection has been incorporated into various CNN architectures and has achieved state-of-the-art performances in numerous domains.

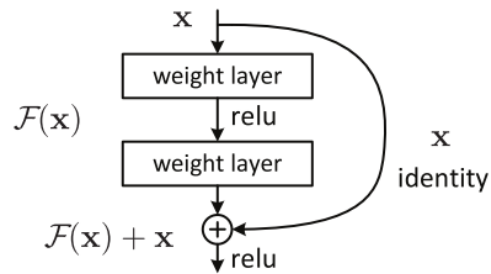


Figure 2.17 Skip connection.

In the domain of image segmentation and image reconstruction, especially biomedical imaging-related tasks, where the output is an image of the same size of the input image instead of a class label, the U-Net is the most useful backbone network architecture. [66] Based on the fully convolutional network [67], a U-Net consists of three basic sections: the contraction path, the bottleneck, and the expansion path. The architecture looks like a 'U' as Figure 2.18 shows that justifies its name.

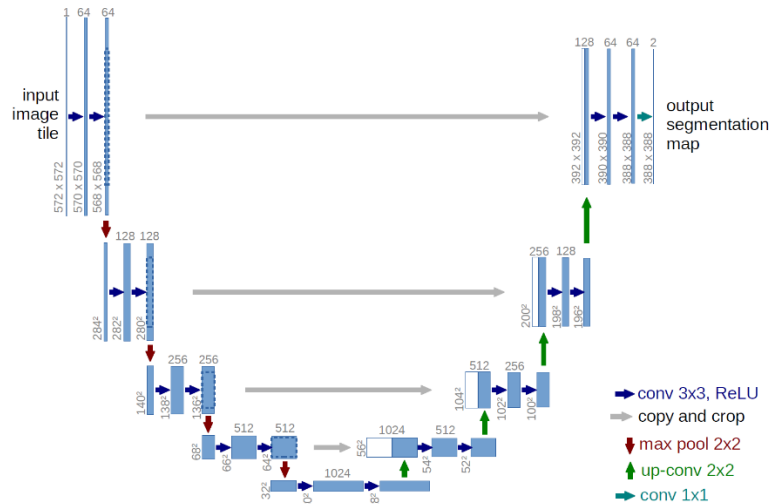


Figure 2.18 U-Net.

The key to U-Net’s capability of predicting a high resolution output lies in the expansion path. The convolutional layers not only take an input from the upsampled features from the previous level, but also combine the feature maps from the corresponding layer in the contraction path, indicated by the gray horizontal arrows in Figure 2.18. This ensures that the features learned while contracting the image will be used to reconstruct the output, and therefore the network gains back the high-resolution detail of the input image.

2.1.1.3. Generative adversarial networks

A Generative Adversarial Network (GAN) is a class of machine learning systems proposed by [51]. In a GAN, two neural networks (the generator and discriminator) contest with each other in a game. The aim of the generator (G) is to learn to generate new data with the same statistics as the training set that can ‘fool’ the discriminator, while the aim of the discriminator (D) is to distinguish the data generated by G from the true training data (Figure 2.19).

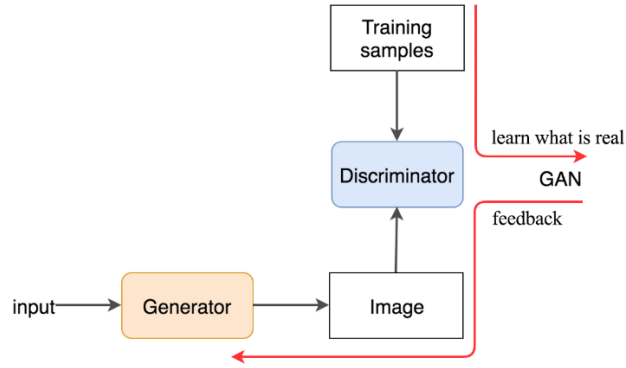


Figure 2.19 Mechanism of GAN.

Mathematically, GAN is often defined as a minimax game with cross-entropy loss objective:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(z)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (2.17)$$

where x is the real sample from the training set and z is the random input for the generator.

Although GANs were originally invented for unsupervised learning, they have also proved useful in semi-supervised learning [68] and supervised learning [69].

Combining with deep convolutional neural networks, a GAN can achieve state-of-the-art performance in image-related tasks. DCGAN [70] is one of the most popular architectures in image generation. Figure 2.20 illustrates the generator architecture of DCGAN.

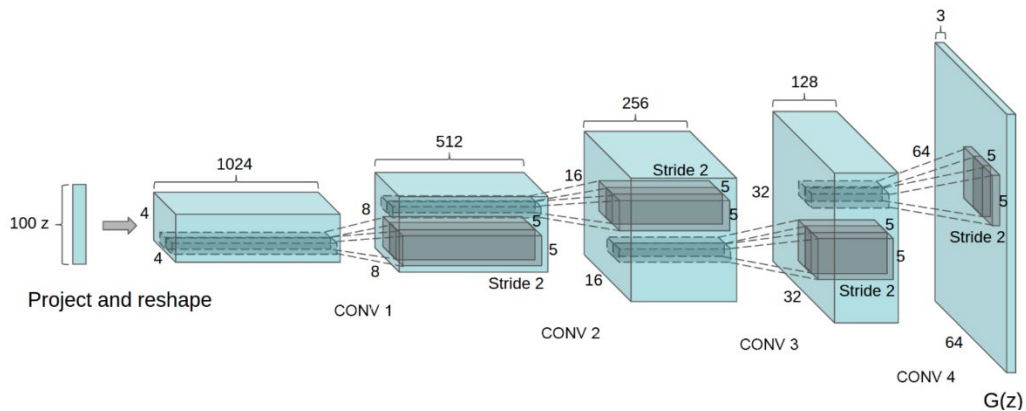


Figure 2.20 DCGAN.

Despite the success of GAN, it is also notoriously difficult to train. Models may never converge, and mode collapse is common. Wasserstein GAN (WGAN) [71] is an approach to facilitate the training of GAN by using a Wasserstein distance metric that has a smoother gradient everywhere and the loss function reflects the image quality, which is more desirable when monitoring the training process. However, in WGAN, the 1-Lipschitz constraint needs to be enforced:

$$|f(x_1) - f(x_2)| \leq |x_1 - x_2| \quad (2.18)$$

Although this condition is ensured by weight clipping in the original WGAN, it can still cause problems in convergence and mode collapse. WGAN with gradient penalty (WGAN-GP) [72] is an approach to enforce the constraint by adding a gradient penalty to the loss function. It further enhances training stability and has been widely adapted in recent GAN related research.

2.4.3. Examples of Deep learning in MRI

The success of deep convolutional neural networks and deep learning in the field of computer vision inspired the research of these state-of-the-art techniques combining with MRI [73]. Many MRI-related tasks such as MR-based diagnosis, MRI segmentation, MR image reconstruction, can be regarded as specific subdomains of generic computer vision tasks and therefore suitable for CNNs. AUTOMAP [74] is a network with fully connected layers acting on k-space data followed by convolutional layers to formulate MRI image reconstruction as a data-driven learning task. In [75], the authors designed an approach to synthesize abnormal MRI images with brain tumors by training a GAN based on pix2pix [69]. QSMnet is a three-dimensional CNN based on U-Net that could solve the ill-posed phase-susceptibility dipole inversion from single orientation data and reconstruct high quality quantitative susceptibility maps. Many MRI segmentation could also be improved by deep learning, such as brain tumor segmentation [76], kidney segmentation [77], and ischemic lesion segmentation in DWI [78].

Chapter 3. Automatic detection of cerebral microbleeds using 3D deep convolutional neural networks

3.1. Introduction

As introduced in Chapter 2.3.1, Cerebral microbleeds (CMBs) are small chronic brain hemorrhages that are prevalent in various diseases such as cerebral amyloid angiopathy [79], stroke [80], neurodegenerative disorders [81], traumatic brain injury [82] as well as following radiation therapy for brain and head and neck tumors [83]. CMBs have been found to serve as biomarkers of complications in these various pathologies. As understanding the role of CMBs is critical, new methods to rapidly detect and quantify CMBs could help shed light on their evolution and correlation with disease status.

The most commonly used medical imaging modality to detect CMBs is susceptibility weighted magnetic resonance imaging (SWI). SWI is highly sensitive to paramagnetic tissues such as hemosiderin, which is abundant in most CMBs, and therefore can provide high contrast between normal brain parenchyma and veins or vascular injury such as CMBs [84]. To understand the impact of CMBs in various diseases, large scale detection, segmentation, and quantification of CMB burden is necessary, the most challenging aspect of which is the efficient and accurate detection of CMBs on SWI. Typically, CMB detection requires not only intense human labor but domain-related expertise, making the job time-consuming and laborious, which inevitably affects the detection accuracy and performance. As a result, much effort has been devoted to developing computer vision algorithms to automatically aid in the detection of CMBs. [27], [28], [30], [85] Although these recent advances have improved automatic or semi-automatic detection of CMBs, all existing methods suffer from low specificity with a large number of false positives (FPs) that ultimately reduces their value and widespread adoption in both the clinical and research settings.

Since AlexNet [63] won the ImageNet Large Scale Visual Recognition Challenge [64] in 2012 by achieving a 40% lower image classification error rate than traditional computer vision methods, deep convolutional neural networks (DCNNs) have demonstrated their superiority in a wide range of computer vision tasks such as semantic segmentation, object detection, and natural image classification. Although the idea of DCNNs was first proposed more than 30 years ago, they have only recently achieved huge successes due to the availability of parallel computing hardware such as GPUs and large datasets. With the emergence of advanced network structures and techniques such as residual unit [65], DenseNet [86], batch normalization [87], and dropout [63], DCNNs can reach performance superior to humans in some tasks.

The goal of this study was to develop a fully automatic pipeline for the identification and labeling of CMBs by combining our previously developed base detection method [88] that used a series of traditional computer vision algorithms with a novel 3D deep residual neural network [65] architecture to reduce the FPs that remain after initial CMB detection and improve specificity. This pipeline uses 3D SWI images as the input to the initial detection algorithm in order to identify the position of potential CMB candidates. These candidates are then passed to a trained 3D deep residual network to both remove definitive CMB mimics and assign a likelihood score for each CMB included in the final detection result. Although this pipeline was trained on 7T SWI images from patients diagnosed with radiation-induced CMBs after receiving treatment for glioma, our framework is flexible for other imaging acquisitions and readily generalizable to other diseases that result in the formation of CMBs.

3.2. Methods

3.2.1. Subjects and Image Acquisition

Seventy-three patients with gliomas were recruited for the study. All patients had previously undergone radiation therapy with a maximum dose ranging from 50 to 60 Gy and had confirmed radiation-induced CMBs. Thirty-one patients were scanned with a 4-echo 3D TOF-SWI sequence (TE=2.4/12/14.3/20.3ms, TR=40ms, FA=25°, image resolution=0.5x0.5x1mm, axial plane matrix size=512x512) [89], while 49 patients were scanned with a standard SWI sequence (3D-SPGR sequence with flow compensation along the readout direction, TE/TR=16/50ms, FA=20°, image resolution=0.5x0.5x2mm, axial matrix size=512x512). Serial imaging was performed on 12 patients, resulting in a total of 91 scans. Patients were scanned using either an 8- or 32-channel phased-array coil on a 7T scanner (GE Healthcare Technologies, Milwaukee, WI, USA). A GRAPPA-based parallel image acquisition was implemented with an acceleration factor of 3 and 16 auto-calibration lines [90].

3.2.2. SWI Processing

The raw k-space data were transferred to a Linux workstation and all image processing was performed using in-house software developed using Matlab 2015b (MathWorks Inc., Natick, MA, USA). The following steps were performed to obtain SWI images from the multi-echo sequence: (1) Auto-calibrating Reconstruction for Cartesian sampling (ARC) algorithm was applied to restore the missing k-space lines of each channel [90]. (2) Magnitude images from each individual channel were combined using the root sum of squares and the skull was stripped using FMRIB Software Library (FSL) Brain Extraction Tool (BET) [91]. (3) The complex data of each coil of echo 2-4 were homodyne filtered with Hanning filter sizes of 72, 88 and 104 for the 2nd, 3rd and 4th echo respectively. [89] (4) The resulting high pass filtered phase images from

echo 2-4 were averaged to produce a mean phase image and used to construct a negative phase mask by scaling the phase values to between 0 to 1. (5) The final composite SWI image was calculated by multiplying the mean magnitude image of three echoes with the phase mask 4 times. SWI images were reconstructed from the single-echo scans using the same pipeline but without the multi-echo averaging and the filter size of 96 was selected empirically [83].

3.2.3. Identification of Candidate CMBs & Labeling of False Positives

An example of multiple CMBs is shown in Figure 3.1. A computer-aided detection method based on traditional image processing techniques previously developed in our group [88] was adopted to propose candidates for the following neural networks. The method first performs a 2D fast radial symmetry transform on the entire input SWI image volume slice by slice. Candidate voxels will then go through a series of processing and filtering such as vessel mask screening, 3D region growing, and 2D geometric feature extraction (area, circularity, number of spanned slices, centroid shift distance). The output is a set of voxels that satisfy the predetermined thresholds as CMB candidates. Although this initial algorithm has been shown to detect 86% of all CMBs, many structural mimics or FPs are also incorrectly identified as CMBs. In order to achieve a higher sensitivity, we reduced the number of missing CMBs by lowering the threshold in this algorithm and subsequently applied an interactive graphical user interface (GUI) where the user was asked a series of questions surrounding potential CMBs in order to individually label all FPs identified by the original automated approach. The coordinates of the candidates were then used to extract 3D patches of size 16x16x8 to use as the input of the deep neural networks. An experienced research scientist with several years' experience in identifying microbleeds was asked to use this system to examine all candidates and generate labels used as ground truth after prior guidance by a neuroradiologist. The entire dataset contained 19,762 candidates, 2,835 of which were true CMBs, as detailed in Table 3.1.

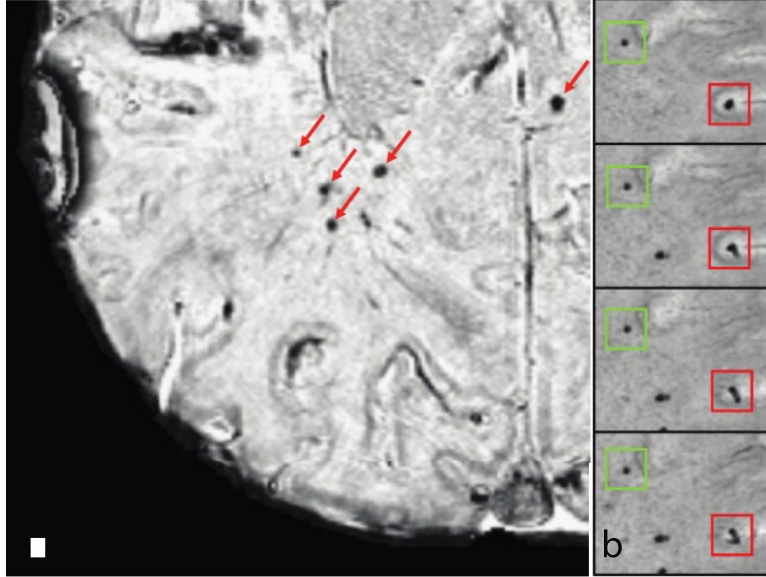


Figure 3.1 CMBs on SWI.

(a) An example SWI slice showing multiple cerebral microbleeds (red arrows). (b) Serial slices in the axial direction demonstrating the difference in 3D structure between a CMB (green box) and a vertical vein (red box) with CMB-like appearance on one slice.

Table 3.1 Number of patches used in all dataset splits. Numbers in parentheses are the number of subjects in each dataset

	CMB	FP	Total
Train (54)	2,243	14,669	16,912
Validation (7)	215	1,023	1,238
Test (12)	377	1,235	1,612
Total	2,835	16,927	19,762

3.2.4. 3D Deep Residual Network

The network we proposed to refine the detection result is the patch-based 3D deep residual network described in Figure 3.2. The network takes a three-dimensional patch of the SWI image centered at the coordinate of the candidate discovered by the CMB candidate identification step as input and outputs a likelihood score of the candidate being real CMBs. The network contains a total of 12 3D residual blocks [65] at 3 different resolution levels connected by 2x2x2 max

pooling layers. After the residual blocks, a global average pooling layer was used to integrate the global information of each channel and the output was fed to a series of fully connected layers for binary classification. ReLU [92] was used as an activation function for all layers except for the last one, where a sigmoid was used to generate a likelihood score that ranged from 0 to 1. Binary cross entropy was selected as loss function of the network. For final classification of candidate patches, we selected a relatively low threshold of 0.1 for CMB/FP decision as a tradeoff for high detection sensitivity. The total number of parameters of the network is about 244,000. Figure 3.2 shows the data pipeline of the CMB candidate labeler and the detailed architecture of the 3D deep residual network.

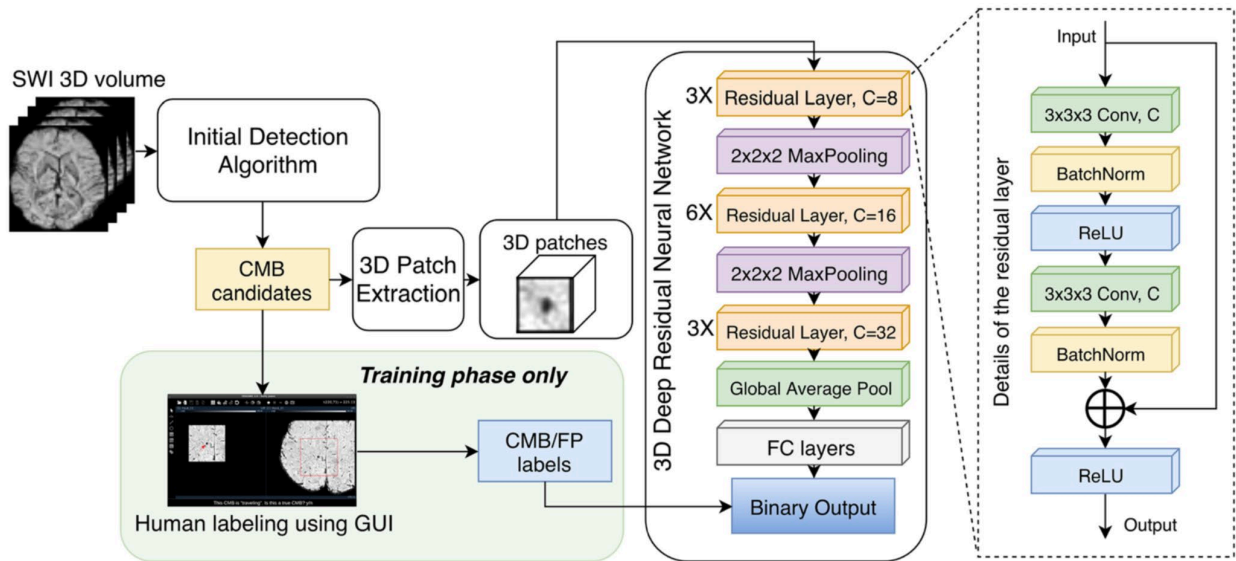


Figure 3.2 Processing pipeline and network architecture. Human labeling of CMB candidates was performed during the training phase to obtain the input and output pairs for this supervised network. During the test phase, all candidates were fed into the network for false positive reduction.

3.2.5. Implementation

Our CNN was implemented using Keras 2.1 [93] with Tensorflow 1.3 [94] backend. The computation was accelerated using an Nvidia Titan Xp GPU with 12GB memory. The “Adam”

algorithm [95] with a learning rate of $1e-4$, $\beta_1=0.9$, and $\beta_2=0.999$ was used for parameter updating. The network was trained for 200,000 iterations with a batch size of 16. Among the 49 patients with a single-echo acquisition, 7 were randomly selected as a validation set, 12 as a test set, while the remaining were used for training. The multi-echo scans were all included in the training set to enlarge the dataset. In the validation and test sets, only the most recent scan was used if the patient had multiple scans. Because our dataset was relatively small compared to modern deep learning tasks, we improved the generalizability of the network by implementing the following data augmentation techniques during training: A) rotating the input patch around the axial axis by a random degree; B) shifting the input patch by 1 voxel in the axial plane; and C) flipping the patch. This greatly extended the capacity of the training set. Imbalance between classes during training was accounted for by weighting the network loss by the proportion of CMBs to FPs. The model with lowest validation loss was selected as the model for testing.

3.3. Results

The benefits of using a 3D-patch based deep residual network with data augmentation over a simple CNN for our application are shown in Figure 3.3. The performance of the network as characterized by the AUC score of the validation set improved with each of the three proposed data augmentation strategies. Although random rotation of patches had the most significant effect the combination all augmentation techniques significantly outperformed applying each separately. The addition of Gaussian noise or random constant patches as forms of data augmentation did not affect the network performance.

In the 12 test patients, the 3D deep residual network successfully classified 90.1% of the candidate patches. Three hundred fifty seven out of 377 candidate patches or 94.7% of true CMBs were correctly identified by the network, while the number of FPs were reduced by 89.1%

(1096 out of 1235) compared to prior studies. The average precision of the network on test patients was 72%, substantially higher than all previously published methods.

Table 3.2 shows the confusion matrix of the classification by the network. On average, the number of FPs per patient were reduced significantly from 103 to only 11.6, and only 1.7 true CMBs per patient were missed by the network. The entire detection pipeline successfully detected 90% of CMBs with only an average of 11.6 remaining FPs per scan within 2 minutes (Intel i7-6700K CPU with 16GB RAM and 12GB GPU installed), providing a practical automatic detection algorithm for clinical and research applications.

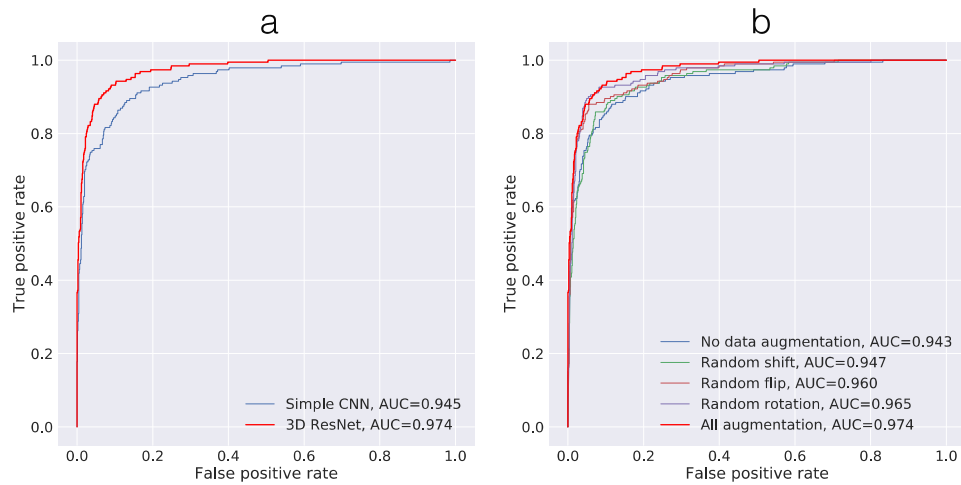


Figure 3.3 AUC comparison of different models and data augmentations.
(a) AUC scores of a simple CNN model compared to our proposed 3D deep residual model, both trained with the same configuration and data augmentations. (b) AUC scores of the 3D deep residual model trained with different data augmentation schemes. Combining all augmentation schemes provided the best performance.

Table 3.2 Confusion matrix of the classification results of the test set. The network removed over 88% of FPs and only missed 5% of CMBs. The numbers in parentheses refer to the average number of candidates per test subject.

	Predicted CMB	Predicted FP
Actual CMB	TP=357(29.8)	FN=20(1.7)
Actual FP	FP=139(11.6)	TN=1,096(91.3)

Since we used a sigmoid function in the last fully connected layer in the network as the activation function, the prediction of the network for each candidate CMB is a number between 0 to 1 and can be regarded as the likelihood of the candidate being a true CMB. Figure 3.4 shows likelihood values for representative true CMBs (panel (a) of Figure 3.4) and the difference between a true and FP CMB (panel (b) of Figure 3.4), overlaid on the corresponding SWI images. In this example the FP CMB was mis-classified as a true CMB due to its likelihood score of .4 even though it is a FP. This is because we set our classification threshold to be .1 in order to maximize sensitivity. If the classification threshold was set to 0.5 as in common classification tasks, it would have been correctly classified, however, the overall sensitivity of our network would have been reduced. The advantage of this approach over binary classification is that it could be used as a soft detection algorithm that guides the users to quickly locate the potential CMBs in patients and distinct thresholds can easily be applied for different populations and applications.

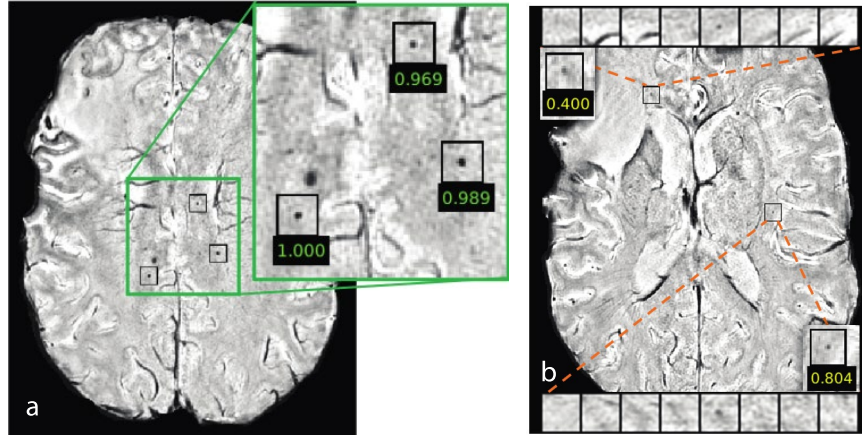


Figure 3.4 Likelihood scores of representative candidates from the test set overlaid on SWI. (a) shows the three correctly identified CMBs by the network with high likelihood scores close to one, while (b) demonstrates two remaining FPs with lower likelihood scores. Since the classification threshold was empirically selected to be 0.1 instead of 0.5 as in most tasks to maintain high sensitivity, the candidate with likelihood score of .4, which would have been correctly classified if we use the usual 0.5 threshold, was mistakenly classified as CMB by the network.

Figure 3.5 demonstrates the network classification results on some representative CMB candidates from the test patients. The missing CMBs (false negatives of the network) were less clear visually than the correctly identified ones and had more complex structure that likely interfered with the classification. In general, the false negatives tended to be located on the upper or lower edges of the entire image volume (example 4, 5 and 8 in Figure 3.5 ‘Failed-CMB’), resulting in incomplete patches that likely negatively affect the decision process. We also observed that the remaining FPs (‘Failed FP’ set: # 1, 4 and 8 in Figure 3.5) might actually be CMBs but were mistakenly labeled as FPs by the human rater, and the network correctly labels them as CMBs. To further verify this hypothesis, we compared the likelihood scores predicted by the network with a neuroradiologist’s (J.E.V.) scoring of the candidates from two randomly chosen patients. Figure 3.6 shows the scatter plot of likelihood scores v.s. neuroradiologist’s ratings (normalized to 0~1 range) for the two patients, with higher scores representing a higher likelihood of the candidate being a true CMB. For both patients, the

neuroradiologist score was significantly correlated with the network likelihood score ($r=.85,.84$ for patient 1,2; $p<0.0001$). These results demonstrate the potential of our network to correct minor errors in the dataset labeled by raters.

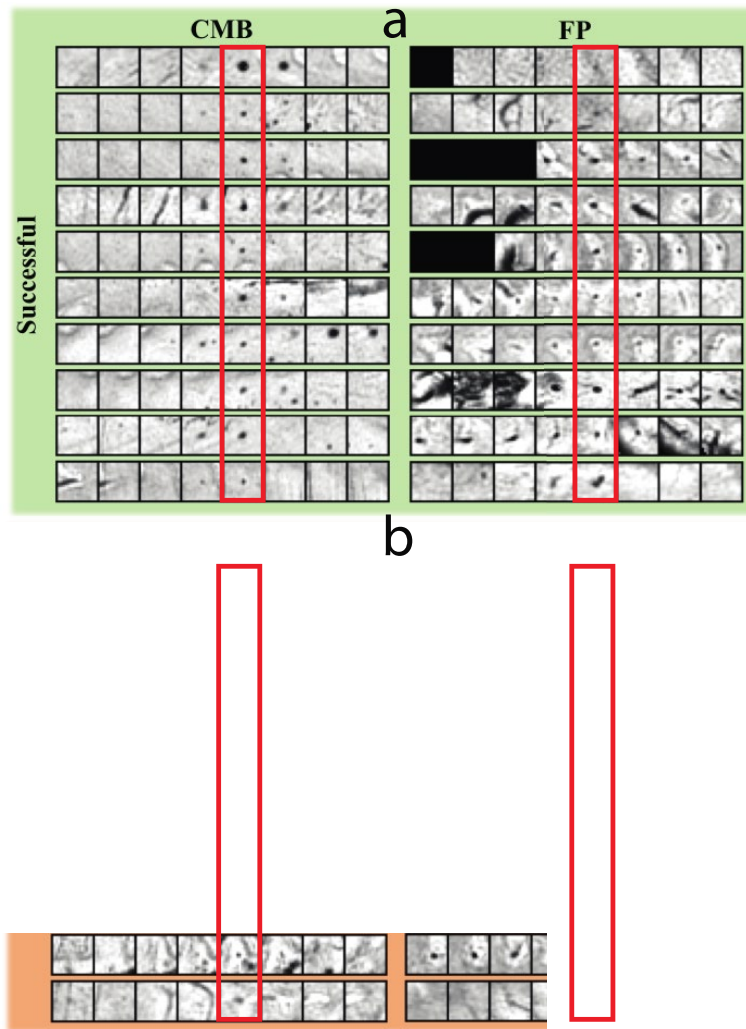


Figure 3.5 Example of classification results for representative CMBs in the test subjects. Each small square shows a slice of a patch centered on a candidate CMB, and each row contains eight consecutive slices in the z-direction for a given 3D patch. Red rectangles highlighted the centered candidates of each patch. The examples shown in the green box on the left were correctly classified as CMBs and false positives, while the orange box on the right illustrates examples where errors are made by the network. ‘Failed-CMBs’ are true CMBs classified as FPs, while ‘Failed-FP’ are false positives that were incorrectly identified as CMBs by the network.

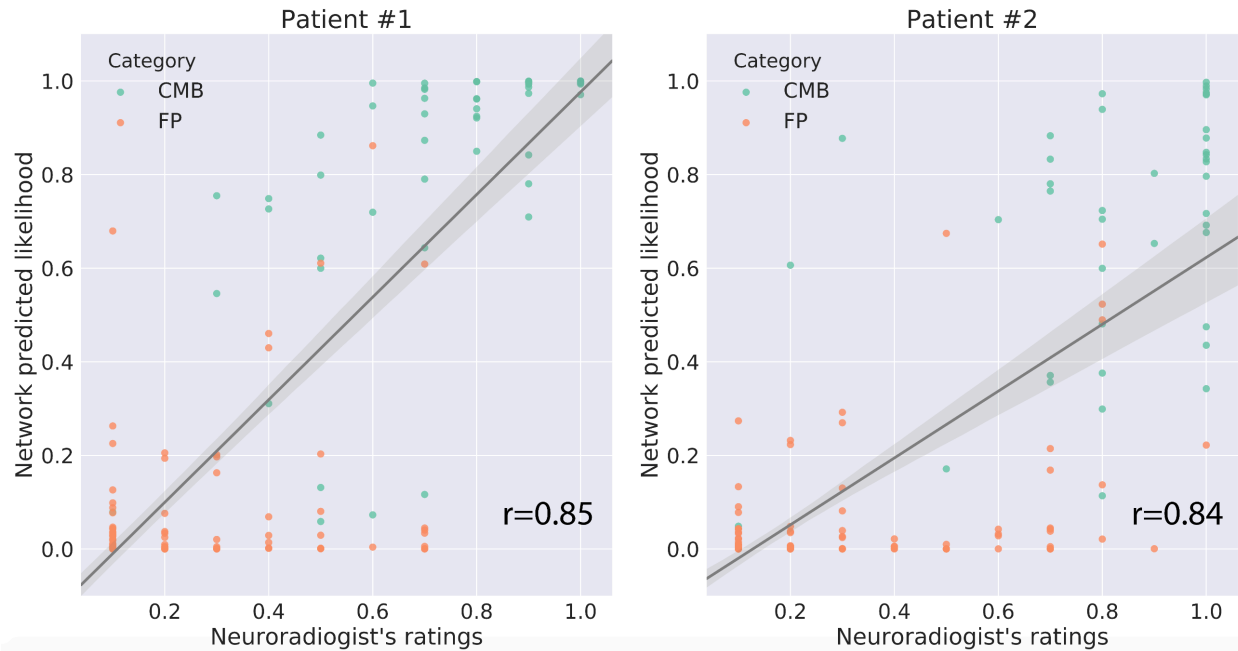


Figure 3.6 The likelihood score predicted by the network compared to a neuroradiologist's scoring of all candidates. Two panels show all of the candidate CMBs from two randomly selected patients with >20 and <60 CMBs.

3.4. Discussion and conclusions

We have demonstrated the capability of using a deep residual CNN architecture for the detection and classification of CMBs, small chronic brain hemorrhages, from their mimics on SWI images. The network was designed to both remove FPs and refine CMB detection by adopting several modern features of CNNs in addition to special designs in order to achieve significantly improved performance. The 3D patch input facilitated the learning of 3D features necessary to classify CMBs and structural mimics as demonstrated in panel (b) of Figure 3.1, where 2D information from a single slice was not enough to distinguish true CMB and veins vertical to the image plane. A deep residual network architecture was selected for this task because of its superior performance in classification, detection tasks, and patch-based image segmentation by enabling training of deeper architectures where the layers learn residual

functions with reference to the layer inputs rather than learning unreferenced functions. This framework eases the training of CNNs by simplifying a network's optimization while requiring a similar number of parameters or weights. By building a deeper network with the help of residual connections, we were able to observe significant improvements in classification performance. To demonstrate the strength of incorporating residual units in our network, the 3D patch classification network used in Dou, et al. [30] was also implemented for comparison. Both networks were trained using the same dataset and configuration. Panel (a) of Figure 3.3 demonstrated that in the validation dataset, we observed a 0.029 increase in AUC score with our 3D residual network. In order to balance the size of the dataset with the task complexity, we limited the capacity of our network to 244,000 trainable parameters to avoid overfitting. Although this number is considered small in comparison to modern architectures based on large scale image datasets such as VGG16 [96], Inception [97], and ResNet101 [65], we believed that it is sufficient for the task of distinguishing true CMBs and FPs because good convergence on training set was observed.

Although the proposed 3D residual DCNN model has the potential to largely refine the detection of CMBs, approximately 10 FPs still remained for each patient. While this number is well within the range of human counting error and less relevant clinically for patients with over 100 CMBs, patients with only a handful of CMBs would still require manual removal of FPs, though at a considerable time savings. Although adding Gaussian noise as a form of data augmentation during training did not help improve the performance of our network, one approach to further reduce the number of FPs would be to apply a denoising filter or other image pre-processing steps to improve SWI image quality before inputting these images into the network. While the individual layers of our network were carefully optimized, new CNN model architectures that have since come to fruition might be able to further reduce the remaining FPs. Our current network was constructed based on data from 73 patients with radiation-induced CMBs that

resulted in less than 20,000 candidate patches. Compared to other modern deep learning-based computer vision tasks, the size of our dataset is still relatively small. Including additional patients with more CMBs that are caused by other diseases might also facilitate the learning of additional features that the network can use to distinguish CMBs and FPs, as well as improve its generalizability.

Although the goal of this study was to use DCNNs to focus on the most challenging part of CMB detection and segmentation, ultimately, we want to implement an end-to-end learning-based approach for complete detection and segmentation that does not rely on a base detection algorithm whose parameters were empirically defined, or further image processing for automatic segmentation once the CMB centers are identified by our network. Incorporating these more advanced approaches [98] that have recently achieved success in other applications into our pipeline, holds promise for full automation of the entire detection and segmentation process. As in many applications of DCNNs, our 3D deep residual network approach can accurately detect CMBs, but the network itself lacks transparency. In order for DCNNs to be routinely adopted by clinicians, future exploration of their interpretation is necessary in order to provide a more concrete explanation for the rationale behind different network design strategies and increase confidence in their results.

In conclusion, we have successfully implemented a 3D patch-based deep residual network that was specifically tailored to differentiate true CMBs from their mimics on 7T SWI images. The 3D residual network was able to achieve 90% sensitivity overall on the 12 patients and reduced the number of FPs to only 11.6 per scan, suggesting that this new approach could greatly facilitate research of various diseases that present with CMBs and potentially increase the benefit of their evaluation in the clinic.

Chapter 4. Generalization and analysis of cerebral microbleed detection algorithm

4.1. Introduction

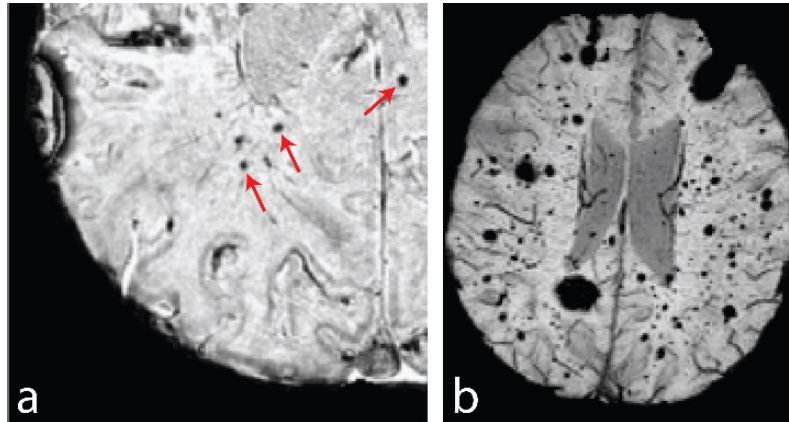


Figure 4.1 Appearance of CMBs on SWI axial slices.

a) A brain tumor patient who developed CMBs after radiation therapy. CMBs are highlighted by red arrows. b) A CCM patient with numerous CMBs and larger hemorrhages.

In the previous chapter, we developed a 3D patch-based deep residual network that could differentiate true CMBs from their mimics on 7T SWI images. The 3D residual network was able to achieve very high detection sensitivity without few false positives per patient on test set. However, the model was trained and evaluated on a dataset of a single disease with similar image quality. Since DCNNs are susceptible to overfitting, especially in the case of medical imaging where the training dataset size is usually limited [99], the ability of the models to be transferred to unseen pathologies with different imaging strategies is unknown. Therefore, in this study, we aim to investigate the performance and generalizability of various strategies for adapting the previously developed DCNN-based CMB detection model that was originally trained on 7T SWI datasets from brain tumor patients who developed radiation-induced CMBs

to images comprised of different scan parameters such as magnetic field strength and image resolution, and an unseen pathology: patients with CCM. (Figure 4.1)

4.2. Methods

4.2.1. DCNN-based algorithm

We investigated the method and model we developed in Chapter 3, which is an automatic CMB detection and segmentation pipeline that uses a 3D SWI scan and is based on a fast radial symmetry transform for initial detection [88] for the initial detection step of all potential candidate CMBs, which are then segmented automatically using automated in house software [29]. A DCNN [31] was then utilized to distinguish remaining false positive (FP) mimics from true microbleeds and generate a binary mask of the final detected microbleeds. The detailed processing pipeline is shown in Figure 4.2. This present study, focused on evaluating the generalizability of the DCNN, a 3D deep residual neural network (3D-DRNN), in reducing the false positives generated by our detection algorithm on a more diverse dataset through establishing a systematic framework for which to assess different strategies for training and testing on unseen data.

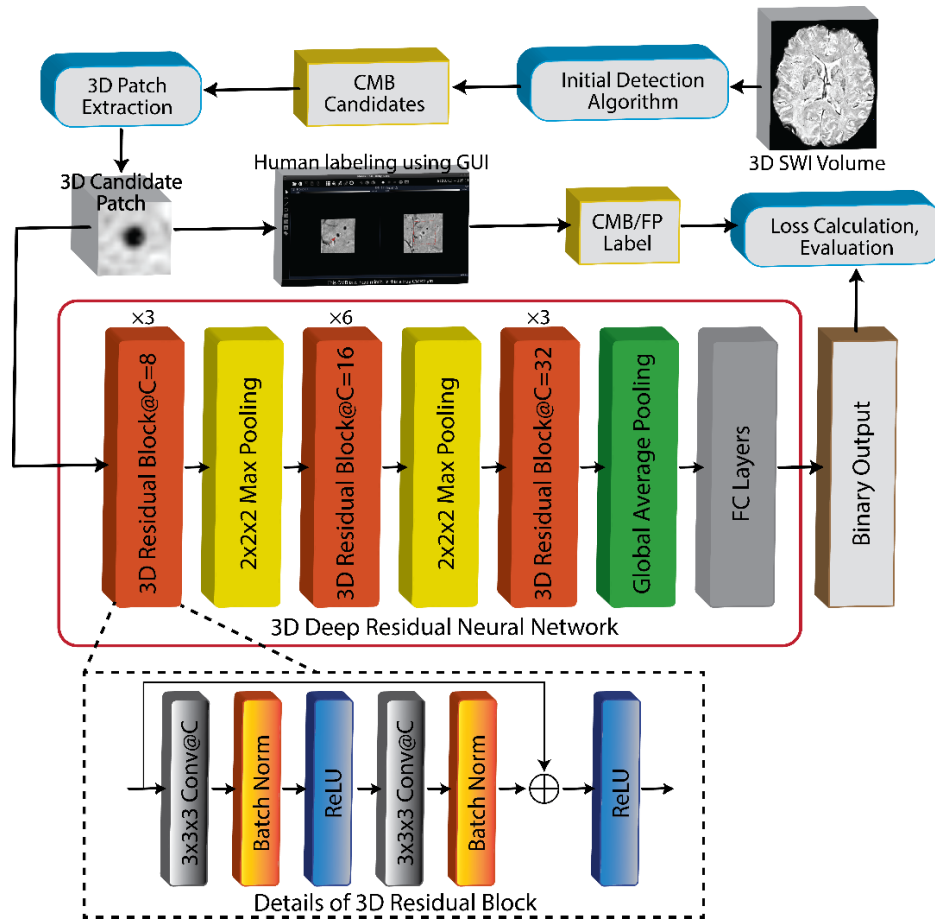


Figure 4.2 Pipeline of automatic CMB detection algorithm. Red box highlights the 3D deep residual neural network we investigated in this study. Black dashed box illustrates the detailed architecture of 3D residual block in the neural network, and “C” represents number of channels of the convolutional layer. The dataset was labeled by experienced staff using a GUI.

4.2.2. Subjects and data acquisition

We acquired data from three groups of patients with different pathologies and/or MRI scan parameters:

- (1) 80 brain tumor patients with radiation-induced CMBs were recruited and scanned using a 7T MRI scanner (GE Healthcare Technologies, Milwaukee, WI, USA). 49 of these patients were scanned with a standard 3D gradient-echo flow-compensated SWI sequence (resolution=0.5x0.5x2mm, in-plane matrix size=512x512, TE/TR=16/50ms,

FA=20°, 8 receiver channels) and 31 patients were scanned with a 4-echo 3D simultaneous time-of-flight MR angiography and SWI sequence (resolution=0.5x0.5x1mm, axial matrix size=512x512, TE=2.4/12/14.3/20.3ms, TR=40ms, FA=25°, 32 receiver channels). 12 patients underwent serial imaging, resulting in a total of 122 scans. All of the scans were accelerated using a GRAPPA-based parallel imaging with an acceleration factor of 3 and 16 auto-calibration lines [90]. This dataset will be referred to as “7T RT”.

(2) Another 97 brain tumor patients with radiation-induced CMBs were scanned using a 3T MRI scanner (GE Healthcare Technologies, Milwaukee, WI, USA) with a similar 3D gradient-echo SWI sequence as used in 7T RT cohort (resolution=0.5x0.5x2mm, axial matrix size=512x512, TE/TR=25/56ms, FA=20°, 8 receiver channels). This dataset will be referred to as “3T RT”.

(3) 48 patients with CCM, that had multiple CMBs in addition to larger hemorrhagic lesions, were scanned using a 3T MRI scanner (Siemens, Erlangen, Germany) with a standard SWI sequence (resolution=1x1x1.5mm, in-plane matrix size=256x192, TE/TR=20/28ms). 18 of these patients were scanned serially, resulting in a total of 66 scans. This dataset will be referred to as “3T CCM”.

Table 4.1 Comparison of performance of the 3D-DRNN on different datasets. The AUC scores shown are mean and standard deviation of AUC scores of all 5-fold networks in the test set.

Dataset	7T RT	3T RT	3T CCM
#Subjects	80	97	48
#Scans (train+val/test)	122 (110/12)	97 (61/36)	66 (45/21)
Image Resolution(mm)	0.5x0.5x1 or 2	0.5x0.5x2	1x1x1.5
#CMB (#/scan)	3824 (31.3)	1626 (16.8)	3912 (59.3)
#FP (#/scan)	24085 (197.4)	12163 (125.4)	5651 (85.6)
#Total	27909	13789	9563
#CMB/#FP	0.159	0.133	0.694

Note: # is short for “number of”. CMB: cerebral microbleeds, FP: false positives/mimics, val: validation.

4.2.3. Data processing and labeling

The k-space data of the 7T RT and 3T RT datasets were transferred to a Linux workstation and SWI images were reconstructed using an in-house software developed using Matlab 2015b (MathWorks, Inc., Natick, MA, USA) following the steps described in [31]. For the 3T CCM dataset, the SWI images were transferred from the scanner for further processing and analysis. All of the SWI images were fed into the automatic CMB detection algorithm to generate potential CMB candidates (11). These candidates were then divided into four subsets and labeled by four experienced researchers respectively using the false positive (FP) labeling GUI developed in our laboratory [29] (see Figure 4.2 for illustration of processing pipeline). These CMB/FP binary labels were regarded as ground truth labels for evaluation and analysis. Summary data (i.e. number of subjects, image acquisition scheme, number of CMB candidates) for the three datasets (i.e. 7T RT, 3T RT, 3T CCM) are listed in Table 4.1.

4.2.4. Training and evaluation

To test and compare the performance of the automatic CMB detection algorithm, we set aside 12 patients from the 7T RT dataset (370 CMBs and 1203 FPs), 16 patients from the 3T RT dataset (362 CMBs and 2367 FPs) and 21 patients from the 3T CCM dataset (1512 CMBs and 2076 FPs) as the test set. If serial scanning was performed on a patient, the latest time point was included in the test set. The remaining datasets were used as training and validation datasets with 5-fold cross-validation for the 3D-DRNN for each of the different training strategies demonstrated in Figure 4.3. In this study, “train from scratch” refers to initializing the 3D-DRNN with random weights and training the model until its convergence, while “fine-tune” indicates transferring of weights from a previously trained 3D-DRNN to the new model and training the loaded model with another dataset using a smaller learning rate. We used an Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) to train all of the 3D-DRNNs for 200,000 iterations with a batch size of 16.

The initial learning rates were set to 10^{-4} for the model trained from scratch and 10^{-5} for the fine-tuned model and gradually reduced to 10^{-6} when validation loss plateaued. Area-under-curve (AUC) of the receiver operating characteristics (ROC), classification sensitivity, precision, and mean number of remaining FPs/scan were calculated to evaluate model performance. During the testing phase, CMB likelihood values were generated for all candidates by averaging the result from each 5-fold-model. A threshold of 0.1 was selected to maximize detection sensitivity in all datasets except for the 3T CCM dataset where 0.1 and 0.3 were tested and compared.

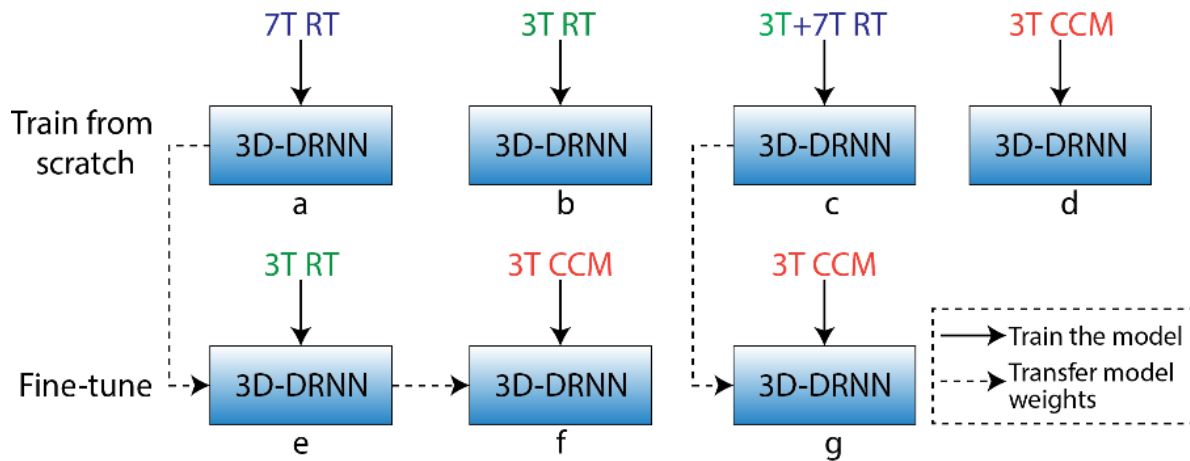


Figure 4.3 Training strategies for the deep convolutional neural network using different datasets and weight initializations.

Top Row: train the model from scratch (with random weight initialization). Bottom Row: first transfer the weights from another trained model and fine-tune the model using a new dataset.

Table 4.2 Comparison of performance of the 3D-DRNN on different datasets. The AUC scores shown are mean and standard deviation of AUC scores of all 5-fold networks in the test set.

#	Trained on	Test on	Sensitivity	Precision	AUC score (5-fold)	Remaining FPs/scan
1	7T RT	7T RT	96.2%	68.5%	0.971±0.003	13.7
2	7T RT + 3T RT	7T RT	96.2%	73.6%	0.972±0.001	10.7
3	7T RT	3T RT	96.1%	38.7%	0.964±0.001	20.7
4	3T RT (finetune #1)	3T RT	94.5%	51.5%	0.969±0.001	12.0
5	3T RT (train from scratch)	3T RT	93.6%	44.6%	0.958±0.005	15.8
6	7T RT + 3T RT	3T RT	95.7%	46.0%	0.969±0.002	15.2
7	7T RT	3T CCM	85.7%	62.8%	0.820±0.015	36.8
8	3T RT (finetune #1)	3T CCM	84.7%	66.5%	0.825±0.010	30.8
9	3T RT (train from scratch)	3T CCM	86.2%	62.5%	0.779±0.026	37.3
10	7T RT + 3T RT	3T CCM	88.0%	64.6%	0.843±0.009	34.7
11	3T CCM (finetuned #10)	3T CCM	97.8%	59.0%	0.888±0.004	48.9
12	3T CCM (finetuned #8)	3T CCM	98.0%	58.5%	0.888±0.004	50.1
13	3T CCM (train from scratch)	3T CCM	97.9%	58.3%	0.887±0.004	50.3
14	3T CCM (finetuned #10)	3T CCM	92.3%	67.9%	0.888±0.004	31.5
15	3T CCM (finetuned #8)	3T CCM	92.9%	67.5%	0.888±0.004	32.2
16	3T CCM (train from scratch)	3T CCM	92.9%	68.0%	0.887±0.004	31.5

4.3. Results

Table 4.2 compares the performance of the 3D-DRNN on different datasets among the various training strategies. The AUC scores shown are the mean and standard deviation of AUC scores of all 5-fold networks in the test set. Figure 4.4 illustrates the ROC curves for each training strategy on the three test datasets.

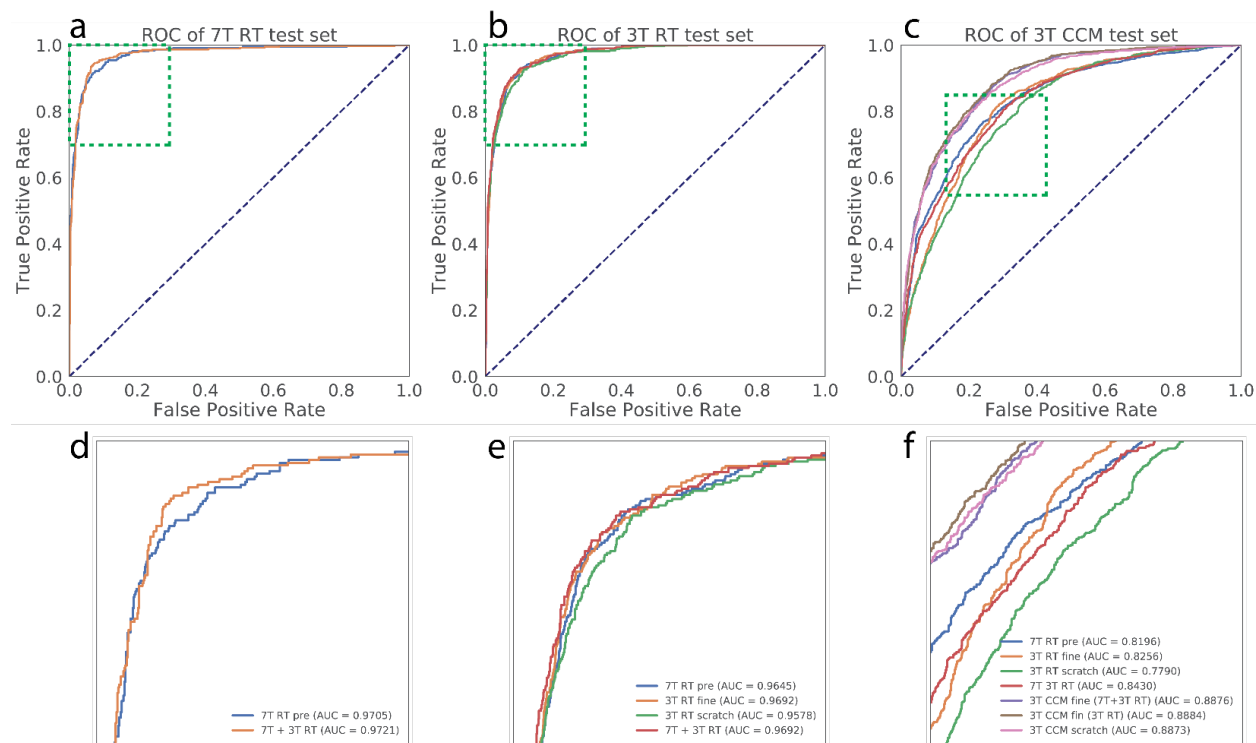


Figure 4.4 Comparison of ROC curves of different models.

a) 7T RT dataset, b) 3T RT dataset, and c) 3T CCM dataset. d-f) demonstrate zoomed in views of a-c).

4.3.1. Direct application of pre-trained models

Row 3 in Table 4.2 shows the performance of the 7T RT trained model directly applied to the 3T RT test set. Compared to row 5 where the model was trained using only 3T RT data, direct application of the pre-trained model on the 3T RT achieved even higher AUC (0.964 vs. 0.958)

and sensitivity (96.1% vs 93.6%). However, despite this higher model performance, there were more remaining FPs per scan (lower precision) when using the same classification threshold.

4.3.2. The effects of fine-tuning and training using combined RT dataset

Fine-tuning the model that was trained using the 7T RT dataset with 3T RT dataset (Table 4.2, row 4) further improved the performance of direct application in row 3: AUC score of 0.969 vs. 0.964, remaining FPs/scan 12.0 vs. 20.7. Figure 4.5 shows several FP candidates in the 3T RT test dataset that were falsely predicted as CMBs using the 7T RT trained model but were successfully removed by fine-tuning the model using 3T RT data. The fine-tuned model also achieved similar AUC score as the model trained using the 7T and 3T RT dataset together (Table 4.2, row 6). Training a new model using only the 3T RT data had the worst performance, even when compared to directly applying the 7T RT model, while both the fine-tuned model and 3T + 7T RT trained model provided the best classification power on the 3T RT dataset. (b) and (e) of Figure 4.4 compares the of ROC curves of each model strategy applied on the 3T RT dataset. Even on the 7T RT test dataset, the 3T + 7T RT jointly trained model had a slightly higher AUC than the original 7T RT trained dataset, with the remaining FPs/scan further reduced to only 10.7/scan, a 21.8% reduction. (a) and (d) of Figure 4.4 shows the comparison of ROC curves on the 7T RT test dataset.









Candidate Patches	CMB Likelihood	
	Pretrain	Finetune
	0.539	→ 0.061
	0.368	→ 0.038
	0.342	→ 0.095
	0.307	→ 0.096
	0.223	→ 0.012
	0.174	→ 0.022
	0.168	→ 0.014
	0.122	→ 0.038

Figure 4.5 Examples of false positive candidates in 3T RT test subjects. Examples of false positive candidates in 3T RT test subjects that were falsely classified as CMBs by the model pre-trained on the 7T RT dataset (predicted likelihood shown in red) and corrected by fine-tuning the model using 3T RT training set (predicted likelihood shown in green). Each row on the left shows consecutive axial slices of a candidate.

4.3.3. Application of the algorithm to CCM

Rows 7 to 10 of Table 4.2 list the performance of models applied to the 3T CCM test dataset without including the CCM dataset during the training phase (using only the 3T or 7T RT dataset). Among the four models, the model trained with combined 3T and 7T RT dataset achieved the highest AUC score (0.843). The 7T RT trained model and fine-tuned model provided slightly lower AUC score (0.820 and 0.825) than the jointly trained model, while the model trained using only 3T RT dataset has a much lower AUC score (0.779). Rows 11 to 13 show the comparison of models trained using CCM dataset. Both training from scratch using only CCM (row 13) and fine-tuning previously RT trained models (row 11 and 12) resulted in a 0.044-0.045 AUC score improvement. We observed that in these three models, the sensitivities

were relatively high (~98%), but the remaining FPs were much higher than in the RT dataset. This might be due to the much higher CMB/FP ratio (see Table 4.1) and improper selection of threshold of 0.1 for classification. With an increased classification threshold of 0.3 for 3T CCM test data (rows 14-16), the remaining FPs/scan was lowered to ~32/scan without sacrificing too much detection sensitivity (still higher than 92%). Although the AUC score did not change for each pair because the threshold does not affect the likelihood predicted by the model, a better balance between sensitivity and precision was achieved. (c) and (f) of Figure 4.4 shows the comparison of ROC curves for the 3T CCM test dataset.

4.4. Discussion and conclusions

4.4.1. Generalization to 3T RT dataset

Although the original 3D-DRNN automatic CMB detection algorithm was developed and tested using 7T data only, direct application of this 7T RT trained model to a 3T RT dataset with different scan parameters preserved the high performance. The good generalizability of the method to different acquisitions was likely because a similar MR sequence was used to acquire the 3T dataset as most of the 7T dataset (which the model had already seen during the training stage), despite the fact that the 3T RT dataset was acquired using a lower magnetic field strength scanner. Directly applying the 7T RT model achieved even better results than training a separate model using 3T RT dataset, suggesting that the contrast and intensity distributions of the two datasets are relatively quite similar despite variations in parameters.

4.4.2. Generalization to CCM

The generalization of the automatic detection algorithm to CCM could be regarded as a relatively difficult task compared to the 3T RT dataset because: 1) the pathology of CCM-CMBs

is substantially different from radiation-induced CMBs, with various-sized larger lesions co-existing with CMBs (see Figure 4.1 for comparison of appearance); 2) the image acquisition parameters are vastly different from RT subjects (0.5mm vs. 1mm in-plane resolution, TE/TR, scanner manufacturers); 3) the RT datasets underwent the same offline SWI processing pipeline while for the CCM patients SWI was performed on the scanner; and 4) a much higher CMB/FP ratio (0.694 vs. 0.133 existed) in the 3T CCM dataset. Still, the model trained using 7T and 3T RT dataset combined without seeing CCM data could still perform relatively well with 88% detection sensitivity and 64.6% precision. This provides evidence that the model has the ability to learn the inherent features of CMBs and FPs and distinguish them when different underlying diseases are present and under a wide variety of imaging conditions. Although we have shown that the model is still applicable if a CCM training set is not available, model re-training and threshold tuning is still required to improve the performance in this population. In the case of 3T CCM, this approach elevated the AUC score by ~ 0.045 , or 5% improvement and provided a better balance between detection sensitivity and precision. By comparing row 11 to 13 in Table 4.2, we also found that pre-training the network with RT dataset for CCM does not improve the performance in 3T CCM test dataset. This is likely due to the fact that the 3T CCM dataset was acquired using very different parameters from the 7T and 3T RT dataset, and suggests that the optimal model parameters for the 3T CCM dataset are less likely to be learned from the model that performed well on RT dataset.

4.4.3. Limitations

There are a few potential limitations of this study. Although the size of the CMB patch-based dataset we acquired and used is relatively large compared to other medical imaging related deep learning investigations, it is still limited compared to typical large-scale computer vision tasks such as ImageNet [100]. Although we validated the generalizability of the DCNN by

measuring and comparing the macroscopic metrics such as AUC score, sensitivity, precision, the interpretability of the CMB detection algorithm, or how the network makes decision on certain candidates, remains unclear. This has been an important issue under active research for deep learning methods, especially in the field of medical imaging where clinical use of a tool requires more interpretation for wide-scale adoption. To further probe the DCNN used for automatic CMB detection and confirm its performance in a larger variety of conditions, more advanced approaches to analyze deep neural networks are still needed.

4.4.4. Conclusion

In conclusion, we evaluated the DCNN-based automatic CMB detection algorithm that was developed for brain tumor patients with radiation-induced CMBs scanned at 7T by generalizing the model to other scan environments, imaging parameters, and a previously unseen pathology. We demonstrated that the deep convolutional neural network could learn the intrinsic features of the target lesions without severe overfitting and be applicable to various conditions when properly trained. The framework provided can also inform on the development and optimization of deep neural network-based methods for other detection and object recognition tasks in medical imaging.

Chapter 5. Comparison of Quantitative Susceptibility

Mapping Methods at 3T and 7T

5.1. Introduction

As introduced in Chapter 2.2, Despite the encouraging results obtained using QSM, the methodology itself suffers from several limitations that are still the topic of ongoing investigation and limit its widespread adoption in the clinic. Most pipelines for generating QSM images are comprised of 2 main processes: background field removal and dipole inversion. The first step in generating a uniform QSM image of brain tissue involves the accurate removal of the background phase comes from susceptibility sources outside the brain, such as the environment and air-tissue interfaces [9]. Since in most cases the background phase is 1 to 2 order of magnitude stronger than the tissue phase, a robust background removal algorithm is required to extract tissue phase from the total phase signal [101]–[104]. The second step typically involves employing a method to efficiently perform dipole inversion whereby the magnetic field perturbation of any given voxel is a superposition of all dipole fields generated by all voxels. This step is the most computationally intensive because it is calculated by a point-wise multiplication in frequency space followed by solving the inverse problem ($\Delta B_z(\mathbf{k}) = B_0(\frac{1}{3} - \frac{k_z^2}{|\mathbf{k}|^2})\chi(\mathbf{k})$) to estimate the susceptibility field, where ΔB_z is the local field perturbation, B_0 is the main magnetic field, χ represents the tissue susceptibility, \mathbf{k} is the frequency space vector and k_z is the z-component. When $k_z^2/|\mathbf{k}|^2 \approx 1/3$, which is commonly referred as the “zero-cone region”, the bracket term on the right-hand side becomes close to zero, causing either missing measurements or noise amplification when solving the inverse problem. One approach to avoid this inherently ill-posed problem is to acquire phase data with 3 or more different B_z direction and fill in the missing data near the zero-cone. [105] Although this method, known as COSMOS, is considered the gold-standard for producing highly accurate artifact-free, it clinically impractical

because of the prolonged scan times, uncomfortable head positioning, and computational requirements. As a result, modern QSM methods typically rely on either 1) dipole inversion algorithms to solve the ill-posed inversion problem after background field removal [104], [106]–[108], or 2) algorithms that integrate the two steps to reduce the error propagation between consecutive processing steps [109], [110].

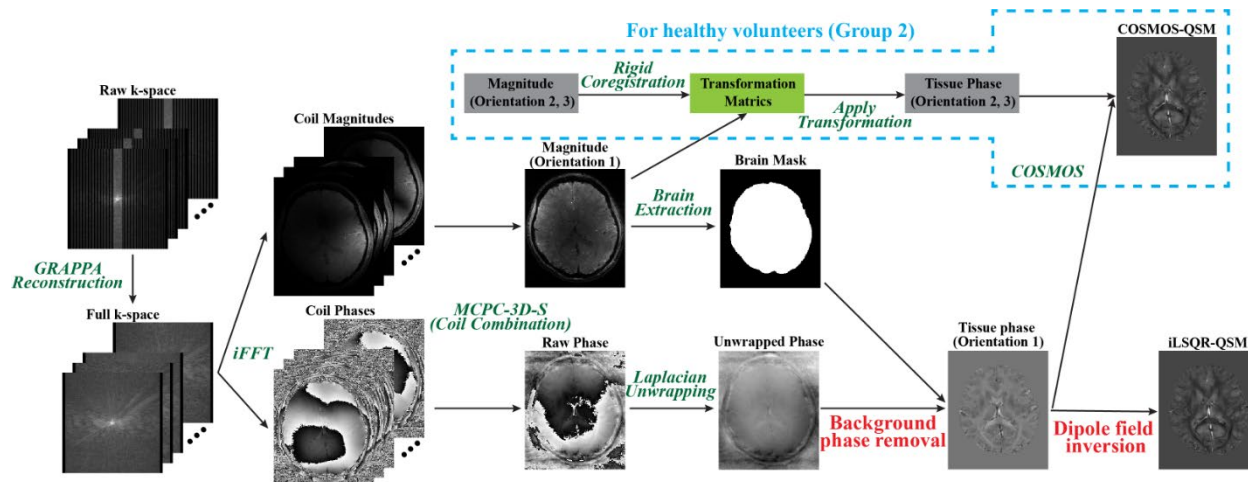


Figure 5.1 QSM processing pipeline adopted in this study. Blue dashed box only applied on group 2 subjects with multiple orientation scans to reconstruct COSMOS QSM.

Despite its promise as quantitative biomarker in the diagnosis and monitoring numerous pathologies, the technical challenges faced by background removal and dipole inversion in QSM, the lengthy computation times, like of widespread algorithm availability, and need for fine-tuning of parameters for a given acquisition and set up, have limited their adoption in clinical practice. Although a variety of algorithms and approaches have been proposed to tackle these challenges, the accuracy of these methods and the resulting QSM image quality have not been quantitatively assessed in a clinical setting. The goal of this study was to evaluate the effects of applying different background field removal and dipole inversion algorithms on noise characteristics, image uniformity, and structural contrast for CMB quantification of QSM images acquired at 3T and 7T field strengths. To accomplish this, we selected eight widely used

background phase removal and dipole field inversion algorithms for QSM and applied them to 11 patients with CMBs and 8 volunteers with ground truth QSM reference calculated using multiple-orientation scans.

5.2. Methods

5.2.1. Subjects and Image Acquisitions

Two groups of subjects were recruited for this study. The first consisted of eleven (11) patients with high grade gliomas who received radiation therapy between 2 and 15 years before imaging and developed cerebral microbleeds. This group of patients were scanned on both 3T and 7T MRI scanners (GE Healthcare Technologies, Milwaukee, WI, USA) on the same day less than 30 minutes apart. The second group consisted of eight (8) healthy volunteers (average age=28, M/F=3/5) scanned only on a 7T MRI scanner. This study was approved by our Committee of Human Research, and written informed consent was obtained from all subjects. The image acquisition parameters for the two groups were as follows:

- 1) Group 1: High resolution T2*-weighted spoiled gradient (SPGR) echo sequence with 3D flow-compensation was performed using whole-body 3T and 7T MRI scanners with 8-channel phased array coils. TE/TR was 28/46ms at 3T and 16/50ms at 7T. A two-fold(3T)/three-fold(7T) generalized auto-calibrating partial parallel acquisition (GRAPPA) parallel imaging acquisition with 16 auto-calibrating lines was implemented to keep the total acquisition within 7min. A 24cm FOV, 0.5×0.5×2mm resolution and flip angle of 20° were used at both field strength.
- 2) Group 2: 3D multi-echo gradient-recalled sequence (4 echoes, TE= 6/9.5/13/16.5ms, TR=50ms, FA=20°, bandwidth=50kHz, 0.8mm isotropic resolution, FOV=24x24x15cm) was performed using a 32-channel phase-array coil on the 7T MRI scanner. The

sequence was repeated three times on each volunteer with different head orientations (normal position, head tilted forward and tilted left) to acquire data for COSMOS reconstruction. GRAPPA-based parallel imaging [111] with an acceleration factor of 3 and 16 auto-calibration lines were also adopted to reduce the scan time of each orientation to about 17 minutes.

5.2.2. Image Reconstruction and Preprocessing

The raw complex k-space data were transferred from the scanner off-line to a Linux workstation, where image reconstruction was performed using an in-house program based on MATLAB 2015b (MathWorks, Natick, MA). The reconstruction pipeline for the single-echo SPGR and multi-echo sequences are demonstrated in Figure 5.1. Missing phase-encoding lines were filled in using the auto-calibrating reconstruction for Cartesian sampling (ARC) method for each individual coil [111] and then a channel-wise inverse Fourier transform was applied to obtain the coil magnitude and phase images. Coil images were combined using the MCPC-3D-S method [112] to obtain robust magnitude and raw phase images for each echo. For the multi-echo sequence used for group 2, this process was performed individually on each echo and repeated orientation scan. Skull stripping and brain mask extraction were performed on the magnitude image using FMRIB Software Library (FSL) Brain Extraction Tool (BET) [91]. For the multi-echo sequence, BET was applied on all four echoes and the final brain mask was generated by calculating the intersection of all the masks. Since the raw phases acquired from k-space data is constrained to $-\pi$ - $+\pi$ range, a 3D Laplacian phase unwrapping method [113] was applied on the phase image and both the unwrapped phase image and Laplacian image were saved for subsequent QSM processing.

5.2.3. QSM processing

5.2.3.1. Background field removal algorithms

The background removal and dipole inversion algorithms selected for comparison in this study are summarized in Table 5.1. From the variety of background phase removal algorithms, we selected the following four commonly used and readily available methods for evaluation and comparison:

- 1) **PDF**: Projection onto Dipole Fields [101], included in MEDI toolbox provided by Cornell MRI Research Lab (<http://pre.weill.cornell.edu/mri/pages/qsm.html>) and applied on unwrapped phase images.
- 2) **RESHARP**: Regularization-Enabled Sophisticated Harmonic Artifact Reduction for Phase data [102], implemented in Matlab 2015b according to the published paper and applied on unwrapped phase images. We chose the radius of the spherical convolution kernel as 6 voxels.
- 3) **iHARPERELLA**: Improved HARmonic (background) Phase REmoval using the LAplacian operator [103], included in STISuite Matlab toolbox provided by UC Berkeley (<https://people.eecs.berkeley.edu/~chunlei.liu/software.html>) and applied on Laplacian phase images obtained using 3D Laplacian phase unwrapping.
- 4) **VSHARP**: Sophisticated Harmonic Artifact Reduction for Phase data with varying spherical kernel [104], included in STISuite Matlab toolbox and applied on unwrapped phase images.

To acquire the final susceptibility maps, tissue phase images obtained using the above algorithms were further processing using iLSQR dipole inversion algorithm (see below) for comparison. For the multi-echo sequence performed on group 2 volunteers, background removal algorithms were applied on each echo individually. The resulting tissue phase images

were then divided by their corresponding echo time and averaged to get a single tissue phase image for dipole inversion.

5.2.3.2. Phase-susceptibility dipole inversion

To compare the effects of different dipole inversion algorithms, three methods were selected, processed, and analyzed after performing background field removal with VSHARP:

- 1) **MEDI**: Morphology-Enabled Dipole Inversion [114], included in MEDI toolbox and required magnitude image, phase image and brain mask. We selected the regularization parameter of data fidelity to be $\lambda = 2000$.
- 2) **CSC**: Compressed Sensing Compensated QSM method [104], implemented in Matlab 2015b according to the published paper and required phase image and brain mask. We selected the k-space threshold as 0.0875, regularization parameters of the total variation term $\lambda_T = 0.001$ and the wavelet term $\lambda_W = 0.01$.
- 3) **iLSQR**: a method for estimating and removing streaking artifacts in QSM using improved LSQR algorithm [107], included in STISuite toolbox and required phase image and brain mask.

5.2.3.3. Integrated QSM algorithm

We also included a QSM algorithm known as QSIP, or Quantitative Susceptibility mapping by Inversion of a Perturbation field model [110] that performs background removal and dipole inversion in an integrated way by updating the calculation of background field during dipole inversion. The code was developed in Matlab 2012b and provided by the author. For our analyses, we regarded QSIP as a dipole inversion method and compared it with the other three dipole inversion algorithms mentioned in the previous section.

5.2.3.4. COSMOS calculation

For the subjects in group 2, COSMOS QSM images (Liu et al., 2009) were calculated as the ground truth susceptibility map. Magnitude images acquired at 3 different head orientations were co-registered with FSL FLIRT [91] and the resulting transformations applied to the corresponding phase images from each orientation. The dipole field inversion was solved using the algorithm proposed in Liu et al., 2009 after background field removal with VSHARP.

Table 5.1 Summary of selected QSM algorithms.

The computation speed was measured on single echo sequence from group 1 patients using a work station with an AMD Opteron 6380 CPU (single core used) and 256GB memory. Approximate numbers were listed.

Background removal			
Name	Input	Speed	Summary
PDF	U, B	500s	Projects the total field onto a subspace spanned by background fields.
RESHARP	U, B	200s	Uses Tikhonov regularization to promote a harmonic internal field with small norm.
iHARPERELLA	L, B	50s	Uses the inverse Laplacian kernel that to project the solution of harmonic equation onto the subspace spanned by all external fields.
VSHARP	U, B	20s	Uses a combination of small and large sphere kernels when applying the SHARP property.
Dipole inversion			
MEDI	M, T, B	2,000s	Inverts the dipole convolution with regularization with anatomic image.
CSC	T, B	8,000s	Promotes image sparsity in the wavelet domain when inverting the dipole convolution.
iLSQR	T, B	100s	Uses an iterative approach to estimate the streaking artifact from ill-conditioned k-space regions only.
Integrated			
QSIP	U, B	6,000s	Inverts a perturbation model that relates phase to susceptibility in spatial domain with a tissue/air susceptibility atlas to estimate field inhomogeneity.

5.2.4. Image Analysis and Comparison

All the reconstructed QSM were zero-mean normalized by subtracting the mean susceptibility inside the brain mask before further analysis and comparison. The following 2 sections describe the metrics used for comparing techniques for each group. Figure 5.2 illustrates visually the definition of these quantified metrics used for comparison.

5.2.4.1. Group 1 (brain tumor patients with CMBs)

- 1) **Noise level and white matter homogeneity.** 10 square ROIs of 150-250 voxels with relatively homogeneous susceptibility values in normal-appearing white matter were manually selected on each patient as in Figure 5.2A. Noise level was defined as the standard deviation of susceptibility values inside each ROI. White matter homogeneity was defined as the standard deviation of the mean susceptibility values of each ROI on each patient and is a proxy measure of residual background phase and low frequency artifacts.
- 2) **Vein contrast.** To measure the QSM reconstruction performance on veins, which have high susceptibility values due to the abundance of deoxyhemoglobin, we drew line profiles through transverse sections of veins on axial maximum intensity projected (8mm) susceptibility maps in order to quantify vein contrast as height divided by full width half height (FWHH) of the line profile (Figure 5.2B).
- 3) **Microbleed contrast.** Radiation-induced CMBs from each patient were segmented on reconstructed susceptibility weighted images (SWI) using in-house software [29], [31]. Each resulting CMB mask was eroded by 1 voxel in the axial plane to remove the blooming artifact present on SWI. The periphery of a microbleed was calculated as the original mask subtracting the eroded mask. CMB contrast (C) was then calculated as the

difference between mean susceptibilities of the CMB and its immediate periphery ($C = \chi_{\text{CMB}} - \chi_{\text{periphery}}$).

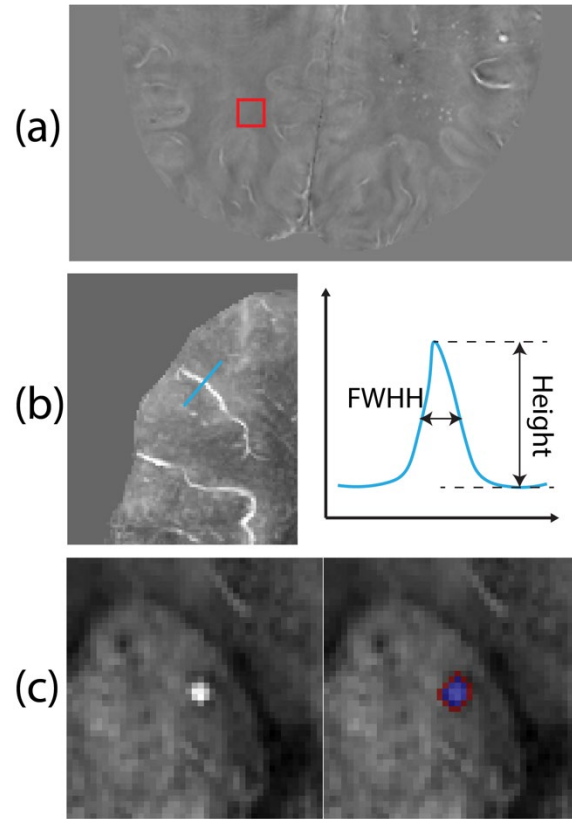


Figure 5.2 Definition of metrics used for brain tumor patients with microbleeds. (a) An example square ROI drawn on white matter to measure noise level and white matter homogeneity. (b) An example line profile perpendicular to a vein on maximum intensity projected QSM and the definition of line profile height and FWHH is shown on the right. (c) Left: a CMB on iLSQR-VSHARP QSM. Right: blue area defines the CMB mask and the red area defines its peripheral region.

5.2.4.2. Group 2 (healthy volunteers)

1) Whole-brain susceptibility. To numerically compare the QSM reconstruction accuracy of the whole brain, we adopted three metrics used in the 2016 QSM reconstruction challenge [115] described as follows:

- **Root mean squared error (RMSE)**, which measures the relative residual error of the reconstructed QSM as:

$$\text{RMSE} = 100 \times \frac{\|\chi - \hat{\chi}\|_2}{\|\chi\|_2} \quad (5.1)$$

where χ represents the ground truth susceptibility map and $\hat{\chi}$ represents the reconstructed susceptibility map.

- **High-frequency error norm (HFEN).** This metric described by [116] estimates the fidelity of reconstructed QSM at high spatial frequencies. It is computed by applying a LoG (Laplacian of a Gaussian) filter of the reconstructed and reference QSM volumes and calculating the L2 norm of their difference normalized by the norm of the filtered reference.
- **Structural similarity index (SSIM).** Described in Simoncelli et al., 2004, this metric quantifies the “visual” similarity between the reconstructed QSM and the reference by combining three similarity components (luminance, contrast, structural).

2) Basal ganglia ROI susceptibility. Since QSM is commonly used to investigate iron deposition in deep gray matter, we compared the mean susceptibility of different QSM methods on the following five ROIs (with left and right ROIs measured separately): red nucleus (RN), substantia nigra (SN), caudate nucleus (CN), putamen (PU), globus pallidus (GP). The ROIs were defined by warping a QSM atlas and its predefined whole brain segmentation [118] to individual COSMOS QSM using FSL non-linear registration tool FNIRL [91]. Figure 5.3 shows the resulting ROI segmentations of a healthy volunteer. Linear regression was also conducted on mean susceptibility values of basal ganglia ROIs between different dipole inversion algorithms and COSMOS mapping.

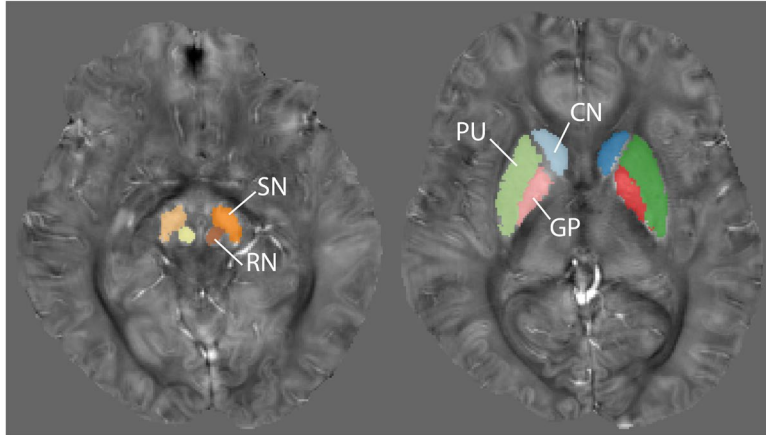


Figure 5.3 Segmentation of basal ganglia ROIs of a healthy volunteer acquired by applying the QSM atlas.

5.2.4.3. Statistical methods

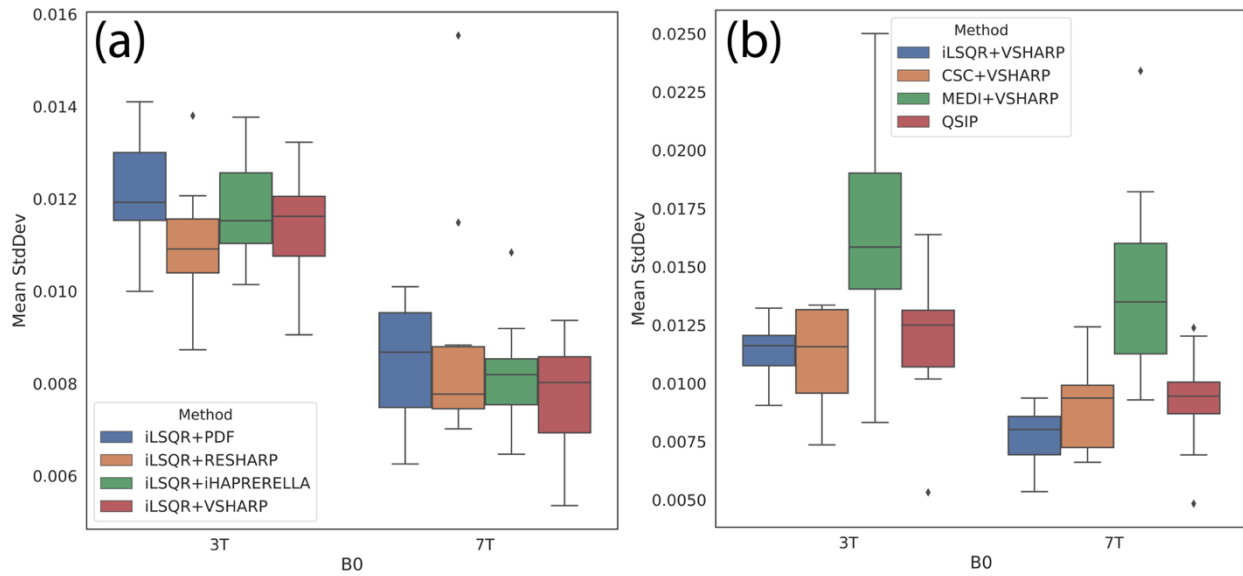
Kruskal-Wallis test and Wilcoxon signed rank tests were used to identify statistically significant differences pairwise among methods. Mann-Whitney (Wilcoxon rank-sum) tests were used to test for significant differences between samples of unequal sizes. A Bonferroni correction was adopted to reduce type I error (false positive results) from multiple comparisons.

5.3. Results

5.3.1. Noise level

Among the four background removal algorithms being compared, PDF had higher noise level than RESHARP and VSHARP ($p=0.006$ and 0.004) at 3T, while at 7T, VSHARP had lower noise level than PDF and iHARPERELLA (Figure 5.4a). Among the four dipole inversion algorithms, MEDI had significantly higher noise levels than the other three algorithms at both field strengths ($p<0.01$; Figure 5.4b). At 7T, iLSQR generated QSM images with the lowest noise ($p<0.05$). As expected, all background field removal and dipole inversion algorithms

achieved a lower noise level at 7T than 3T (all $p < 0.01$), corresponding to the higher SNR available at 7T.



*Figure 5.4 Boxplots of noise level at 3T and 7T.
(a) Background removal algorithms. (b) Dipole inversion algorithms.*

5.3.2. White matter homogeneity

At both field strengths, VSHARP and RESHARP had the highest white matter homogeneity (lowest standard deviation of mean susceptibility of the 10 ROIs within each patient) followed by PDF and iHARPERELLA (Figure 5.5a). Among the dipole inversion algorithms, QSIP and iLSQR had significantly improved white matter homogeneity compared to CSC and MEDI at 3T ($p < 0.05$; Figure 5.5b). Although similar trends were observed at 7T, QSIP had the most homogenous images compared to the other techniques, while MEDI had statistically significant worse white matter inhomogeneity compared to the other 3 techniques (both $p < 0.05$). No statistically significant differences were observed between 3T and 7T for all background removal and dipole inversion algorithms.

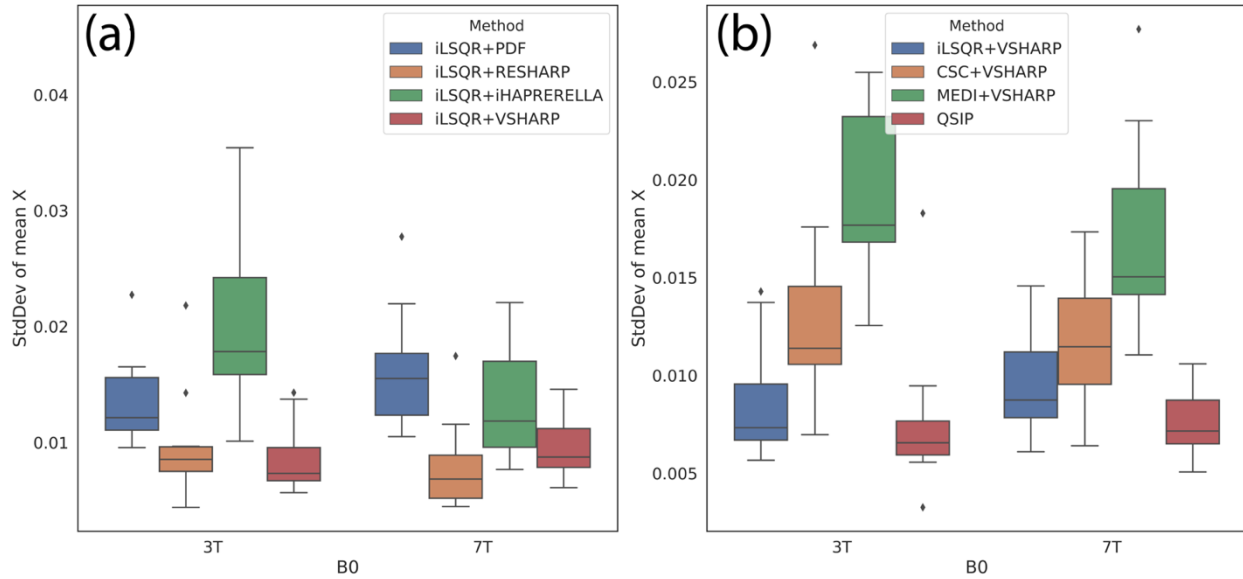


Figure 5.5 Boxplots of white matter homogeneity. (a) Background removal algorithms. (b) Dipole inversion algorithms.

5.3.3. Vein contrast

Figure 5.6 shows the boxplots of vein contrast of all evaluated methods. VSHARP provided higher vein contrast than all other three background removal algorithms at both field strengths ($p < 0.05$). Of the four dipole inversion algorithms, iLSQR and QSIP had higher vein contrast than CSC and MEDI at 3T ($p < 0.0001$ for iLSQR vs CSC, $p < 0.003$ for iLSQR vs MEDI, $p = 0.03$ for QSIP vs MEDI). At 7T, QSIP had higher vein contrast than all the other dipole inversion algorithms ($p < 0.05$), with iLSQR also having significantly higher vein contrast than CSC and MEDI ($p = 0.0008$). When algorithms were compared between field strengths, only MEDI has significantly elevated vein contrast at 7T compared to 3T ($p = 0.001$). Figure 5.7 provides a visual comparison of a vein processed by different algorithms at both field strength. Note that the images at 3T and 7T were not coregistered in order to avoid alterations in image quality caused by interpolation during alignment, so the orientation of the vein varies slightly between field strengths.

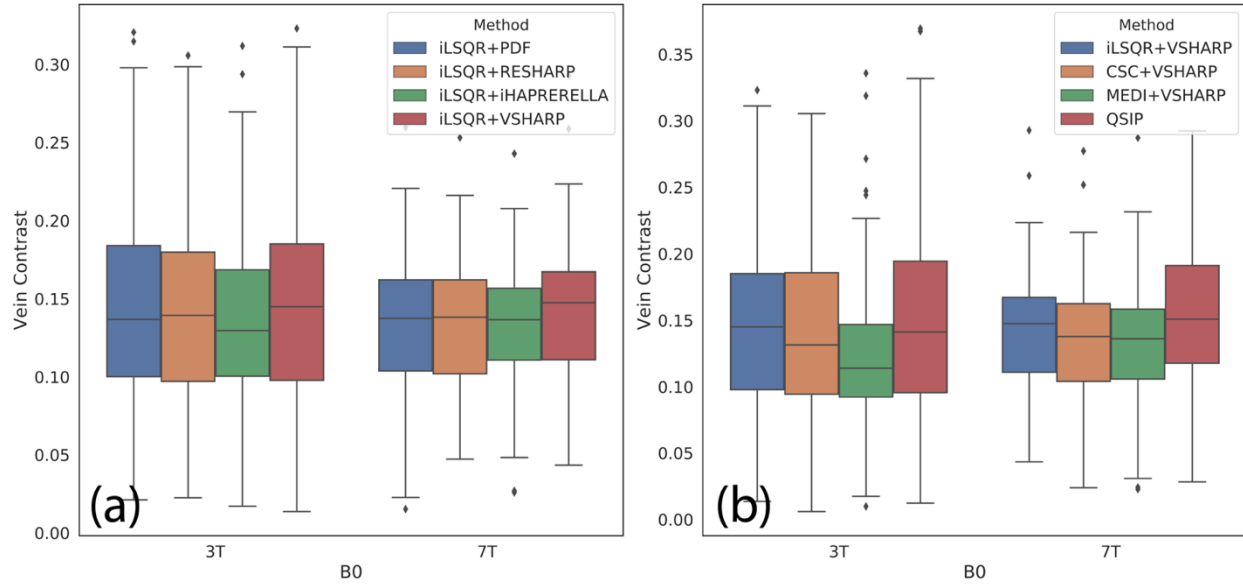


Figure 5.6 Boxplots of vein contrast.
 (a) Background removal algorithms. (b) Dipole inversion algorithms.

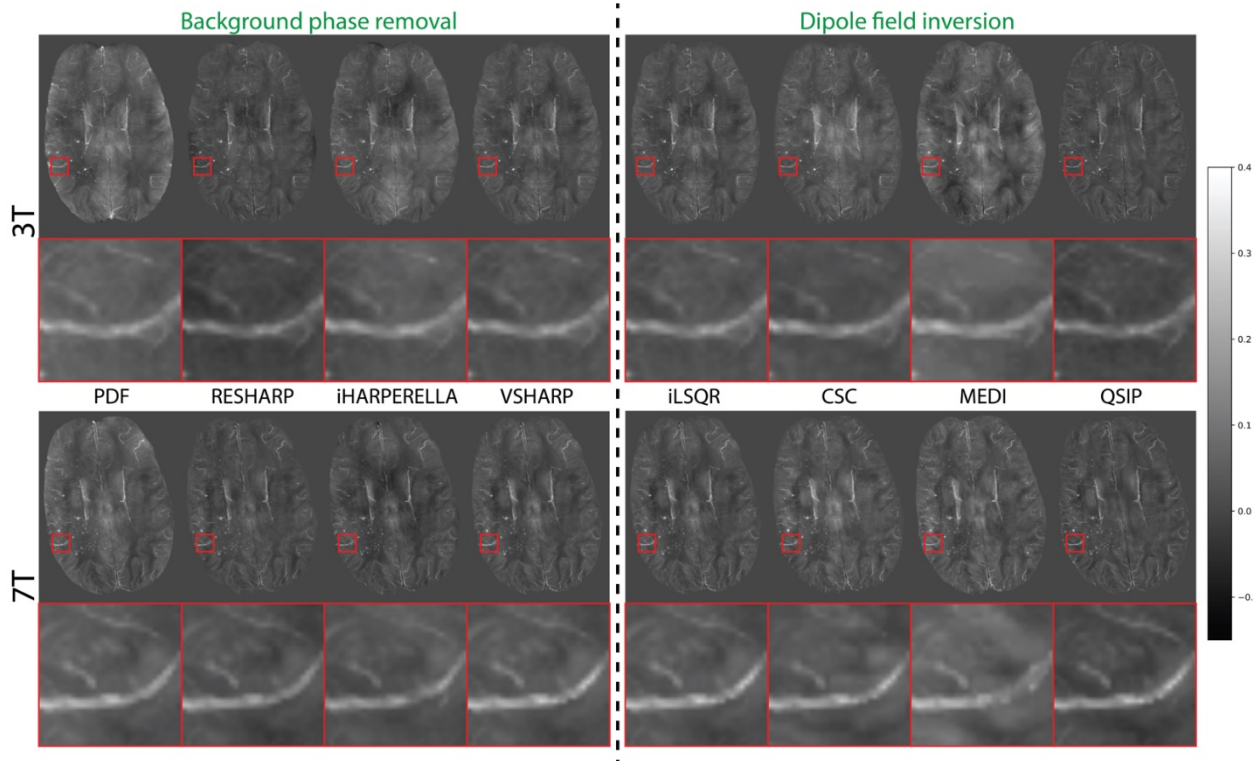


Figure 5.7 Visual comparison of a vein of different algorithms at 3T and 7T.

Top left: background removal algorithms at 3T. Top right: dipole inversion algorithms at 3T.
 Bottom left: background removal algorithms at 7T. Bottom right: dipole inversion algorithms at 7T.

5.3.4. CMB Contrast

No significant differences were observed in CMB contrast among all of the background field removal algorithms (Kruskal-Wallis $p=0.52$ and 0.22 for 3T and 7T respectively, boxplots in Figure 5.8a) Among the dipole inversion algorithms, QSIP had significantly higher CMB contrast than other methods at both field strengths ($p < 0.01$, Figure 5.8b). No significant differences were found between field strengths for all background removal and dipole inversion algorithms ($p > 0.3$) Figure 5.9 demonstrates a visual comparison of three CMBs computed using different background field removal algorithms at 3T and 7T. Similar to the measurement and display of vein contrast, the orientations of CMBs shown in Figure 5.9 are displayed without co-registration between field strengths.

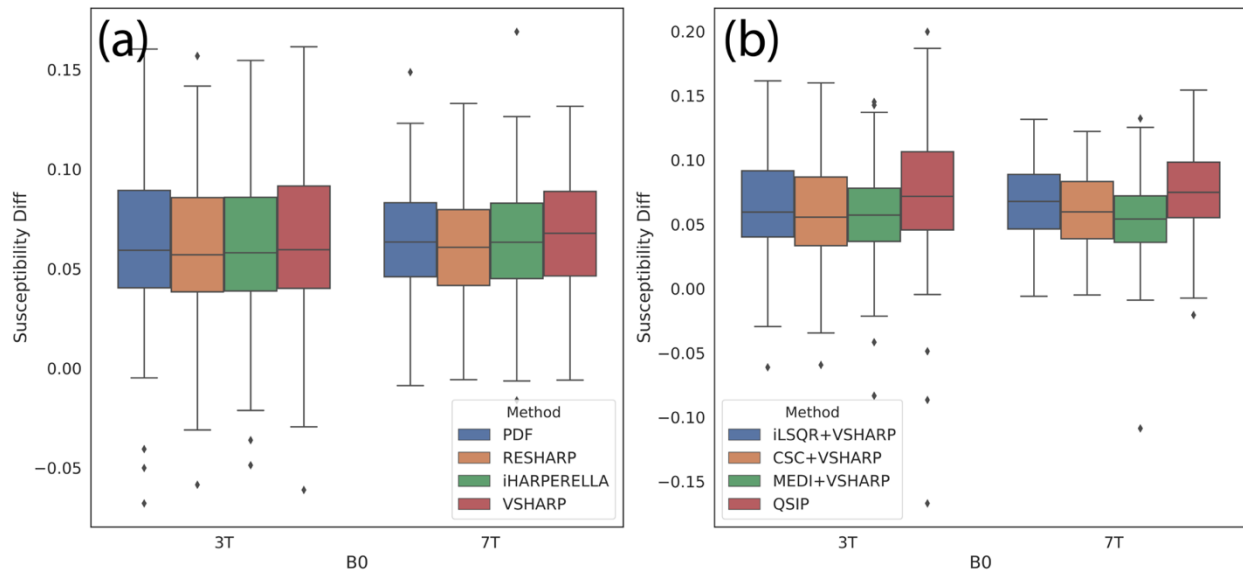


Figure 5.8 Boxplots of CMB contrast.
 (a) Background removal algorithms. (b) Dipole inversion algorithms.

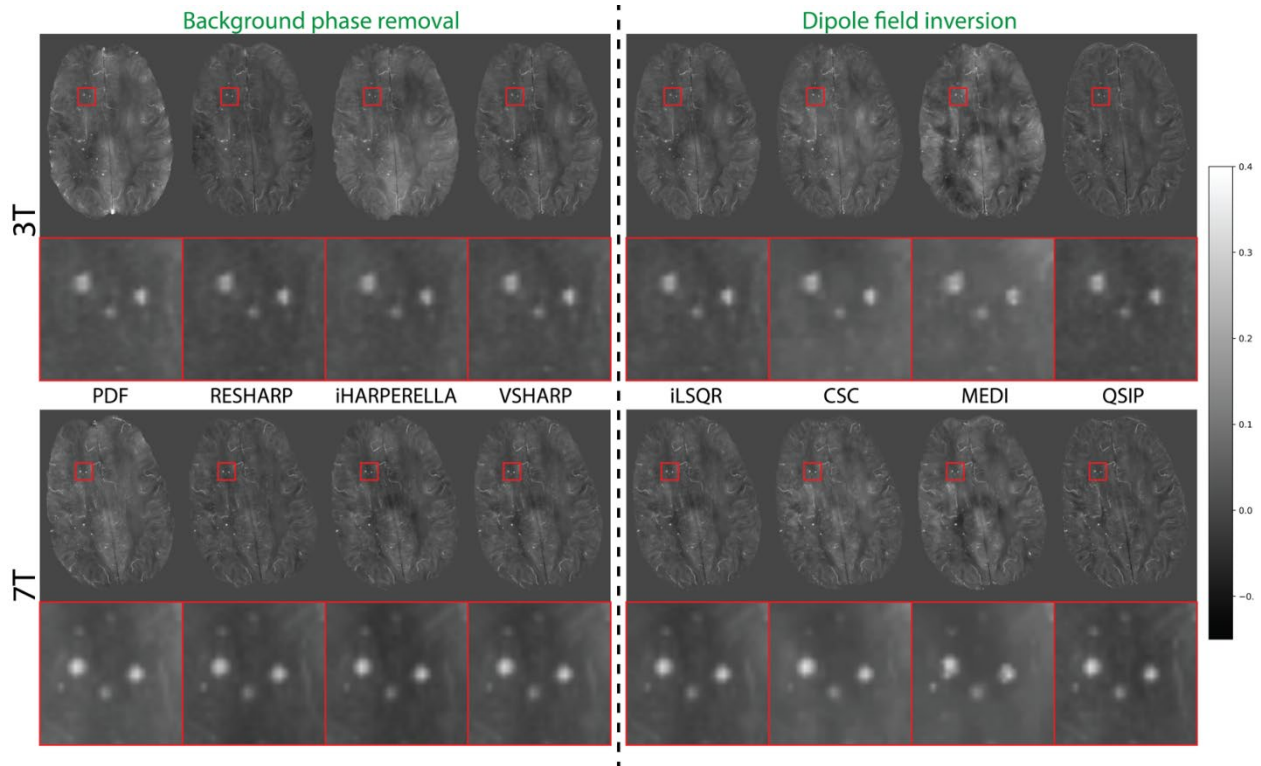


Figure 5.9 Visual comparison of CMBs of different algorithms at 3T and 7T. Top left: background removal algorithms at 3T. Top right: dipole inversion algorithms at 3T. Bottom left: background removal algorithms at 7T. Bottom right: dipole inversion algorithms at 7T.

5.3.5. Whole-brain QSM metrics

RMSE, HFEN, and SSIM of the four dipole inversion algorithms compared to COSMOS are listed in Table 5.2. iLSQR had significantly lower RMSE and HFEN compared to other methods ($p < 0.01$) except for QSIP. For HFEN, iLSQR and QSIP were significantly reduced compared to CSC and MEDI ($p < 0.01$), indicating less deviation from COSMOS for these algorithms. MEDI achieved the highest SSIM to COSMOS among all the dipole inversion algorithms ($p < 0.01$), followed by QSIP and iLSQR. Figure 5.10 visually compares all the dipole inversion algorithms against COSMOS QSM (with VSHARP for background field removal). Although all algorithms achieved relatively low residual streaking artifacts, they have distinct appearances due to different regularization approaches.

Table 5.2 Comparison of whole-brain metrics from 2016 QSM challenge of different dipole inversion algorithms.

The best algorithm under each metric is highlighted in bold.

Method	RMSE	HFEN	SSIM
iLSQR	83.56±5.49	75.54±5.76	0.863±0.027
MEDI	101.43±5.81	97.08±7.49	0.892±0.017
CSC	93.85±4.92	92.34±6.06	0.848±0.025
QSIP	96.49±4.68	78.92±6.34	0.869±0.011

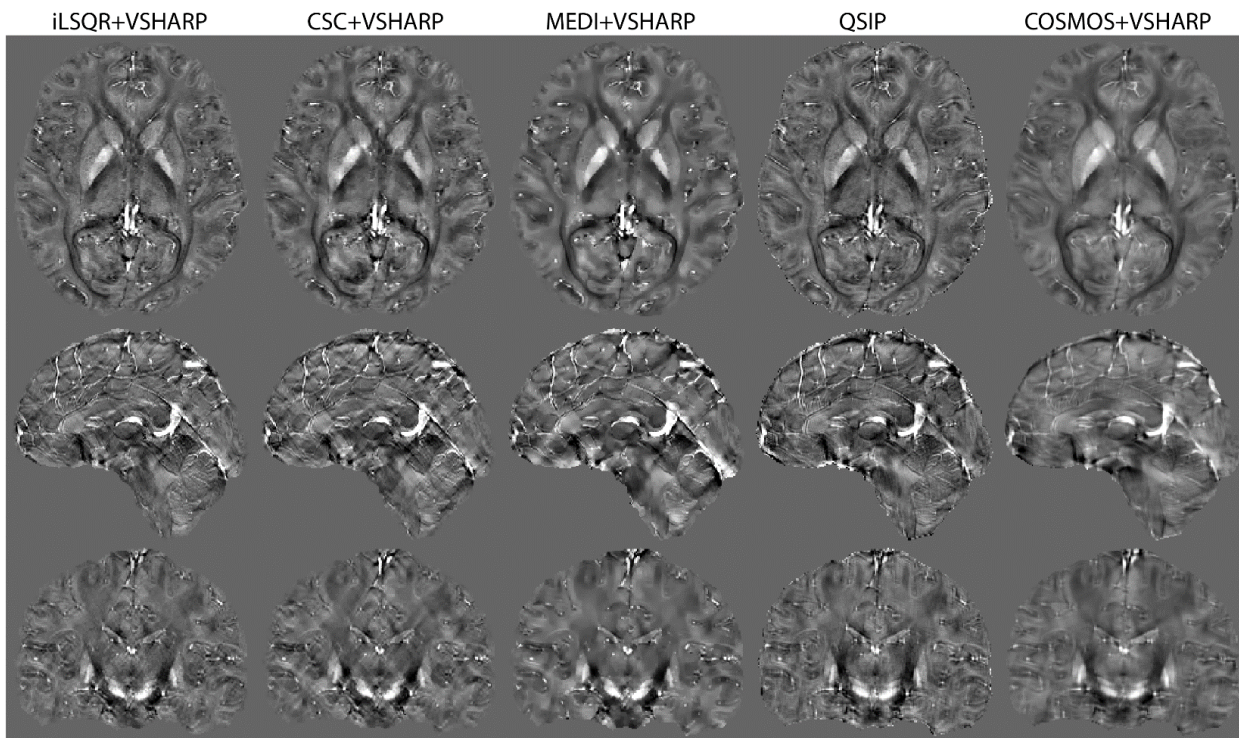


Figure 5.10 Visual comparison of QSM dipole inversion algorithms of a healthy volunteer (subject #5).

5.3.6. Basal Ganglia ROIs

Figure 5.11 shows the boxplot of mean susceptibilities within different basal ganglia ROIs for the different dipole inversion algorithms. All algorithms overestimated susceptibility in the left substantia nigra and underestimated susceptibility values in the striatum (caudate nucleus,

putamen, and globus pallidus), especially QSIP. Figure 5.10 row 1 shows an axial slice of the striatum confirming these findings. Figure 5.12 displays scatter plots of susceptibility values within each ROI for each dipole inversion algorithm compared to COSMOS. The ideal algorithm would achieve a regression line close to identity with low residual error (or a coefficient of determination R close to 1). iLSQR achieved a slope closest to 1 and the highest correlation coefficient, while QSIP had a more dispersed scatter plot and lower R^2 than the other three methods. However, because COSMOS used the same VSHARP algorithm for background field removal as iLSQR, CSC, and MEDI while QSIP integrated background removal into its processing, the lower performance might in part be explained by differences in background field removal rather than dipole inversion approaches.

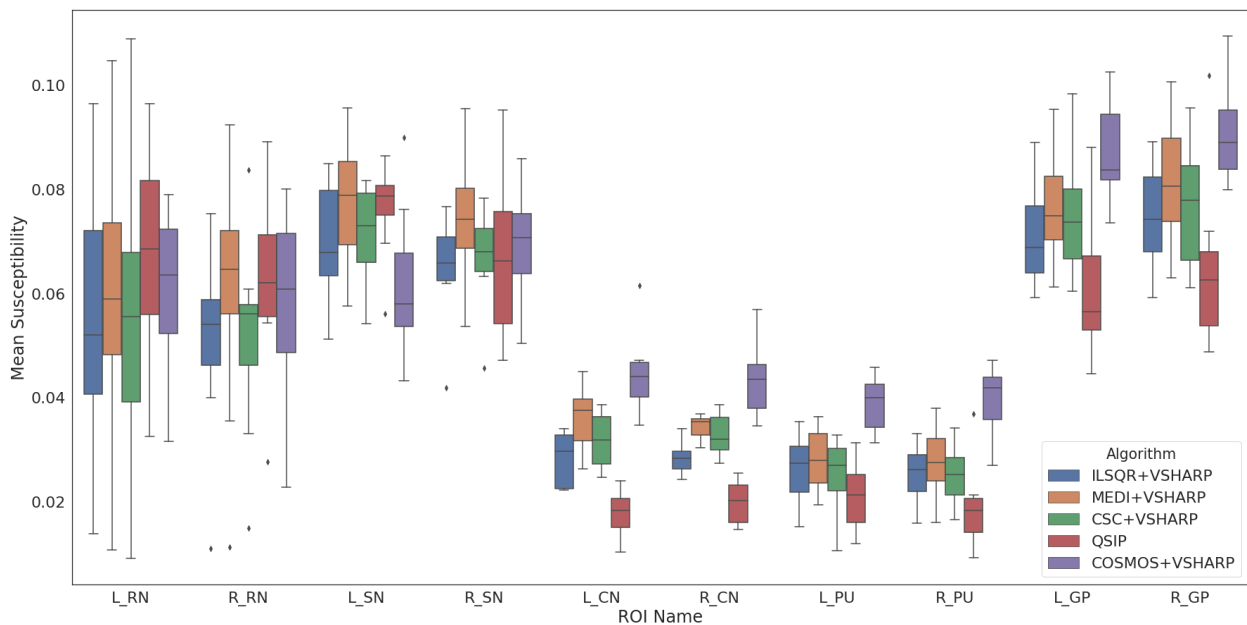


Figure 5.11 Comparison of mean susceptibility within basal ganglia ROIs using different dipole inversion algorithms.

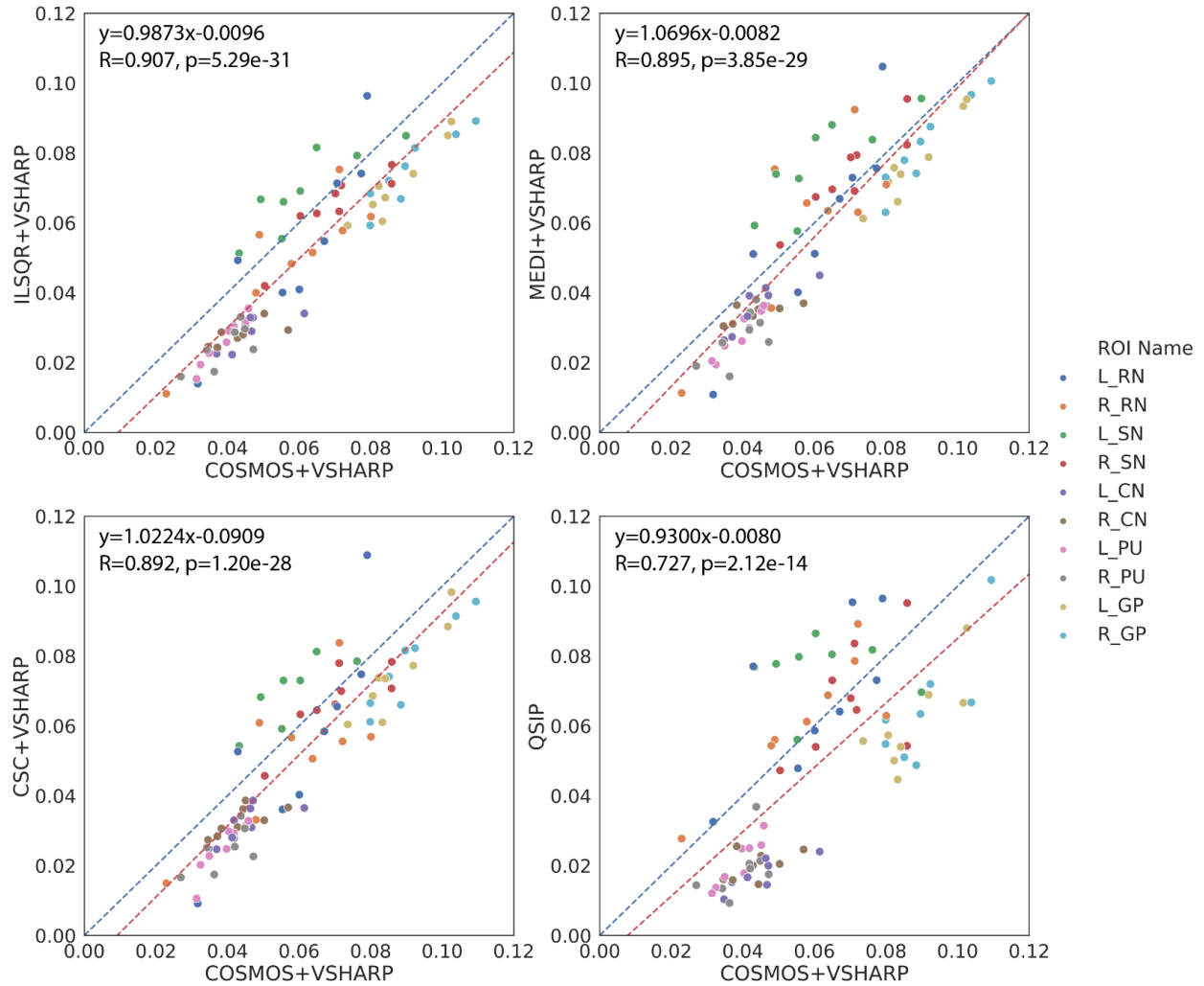


Figure 5.12 Scatter plot and regression plot. Blue: identity, red: linear regression.

5.4. Discussion and conclusions

The results of this study highlight the importance of carefully selecting a QSM processing strategy based on the anatomy and question of interest. Susceptibility values can vary based on the method selected for both background field removal and dipole inversion and depending on whether the goal is to visualize vascular injury, accurately quantify susceptibility or iron deposition within a given region or take on a more global whole brain approach for analysis, a different strategy may be preferred.

To simplify the comparison process, we selected iLSQR for all background field removal methods and VSHARP for all dipole inversion methods, except for QSIP, which is an integrated algorithm. These two methods were selected as base algorithms for three main reasons: 1) better visual quality of reconstructed QSM 2) better performance in many numerical metrics we compared, and 3) relatively high computational efficiency as shown in Table 5.1. Two different groups of subjects were scanned with different parameters in this study in order to cover the majority of use cases of QSM. Group 1 was comprised of brain tumor patients who developed radiation-induced CMBs, an ideal example of abnormal susceptibility. This group of patients were scanned with higher axial-plane resolution sequence to maximize the detection ability of small brain lesions and cerebrovascular structures such as veins. Group 2, on the other hand, consisted of healthy volunteers that were scanned using a multi-echo sequence with isotropic resolution designed for the investigation of iron deposition within the basal ganglia, where research of most neurodegenerative diseases involving QSM have been heavily studied. As the spatial resolution was not as critical as in patients with CMB, we could reduce the axial-plane resolution to match the z-direction resolution to achieve an isotropic configuration. This isotropic setting could also facilitate the processing of COSMOS QSM, which requires the co-registration of scans conducted at different head orientations.

In patients with CMBs, we observed that QSM at 7T has a lower noise level than at 3T. Besides the intrinsic benefit of higher signal-to-noise ratio inherent with increasing field strength, the sequence used at 7T could have potentially induced more phase accumulation than that employed at 3T. At 7T, $TE=16\text{ms}$, so $B_0 \times TE = 112$, and at 3T, $TE=28\text{ms}$, so $B_0 \times TE = 84$. In this study, we defined a metric called white matter homogeneity by calculating the standard deviation of mean susceptibility measured in ten different square ROIs drawn on pure white matter under the assumption that the resulting QSM of white matter should be relatively homogeneous if the processing pipeline is robust. A high standard deviation or low homogeneity

likely indicates either more residual background phase components or other low spatial frequency artifacts. The fact that QSIP images were the most homogeneous regardless of field strength, suggests that imperfect background field removal may be the largest contributor to the inhomogeneity present in the other dipole inversion methods.

The rationale for comparing vein and CMB contrast among algorithms was to 1) examine the residual low spatial frequency artifacts, and 2) validate the algorithms on tissue and lesions with relatively high susceptibility values. QSIP produced the highest vein and CMB contrast at both field strengths and may be the optimal method for applications requiring the segmentation of these structures. The lower vein contrast of MEDI might be due to 1) higher brain tissue susceptibility and 2) blurrier vein structure caused by regularization. Although we favored higher contrast in this comparison, our results cannot be translated to higher accuracy of reconstructed QSM because ground truth susceptibility maps or COSMOS scans were not obtained.

Although QSIP outperformed the other dipole inversion algorithms on the majority of metrics, it resulted in values that were the most different from the gold standard COSMOS images. This was demonstrated by susceptibility values within the striatal regions being significantly lower than those derived from COSMOS, as well as weaker correlation coefficients observed across all regions. More inferiorly in the substantia nigra and red nuclei, however, QSIP resulted in susceptibility values that were the most similar to those quantified by COSMOS. iLSQR, on the other hand, consistently resulted in the most similar susceptibility values to COSMOS across all basal ganglia analyses, making it the preferred method for local absolute quantification of susceptibility within smaller structures.

There are a few additional limitations of this study. Although we conducted a relatively coarse grid-search of parameters and selected the ones that achieved the best visual correspondence with the other algorithms for algorithms that required regularization parameter tuning such as CSC and MEDI, better fine-tuning and optimization of these methods may achieve improved

QSM reconstruction. For our group of healthy volunteers, we chose COSMOS as the ground truth susceptibility map for comparison. This method ignores the susceptibility anisotropy in some tissues such as myelin in white matter fiber tracts. Although Langkammer et al [115] have suggested using the last diagonal component (χ_{33}) in susceptibility tensor theory [16], our data acquisition scheme was limited to three orientations, making the calculation of χ_{33} impossible. But because basal ganglia susceptibility is relatively isotropic, COSMOS should still be a valid image of ground truth of susceptibility in our striatal ROI analyses. Future studies could perform a data acquisition that enables the reconstruction of χ_{33} and use it as the ground truth for comparison on other brain tissue with higher susceptibility anisotropy. Finally, the complete processing pipeline of QSM consists of other pre-processing steps before background field removal, such as multi-coil phase combination and phase unwrapping, that could also affect the resulting QSM image quality and quantification accuracy [119] were not investigated in this paper. Since QSM is recently gaining traction in other organs such as liver and kidney, a similar evaluation should be performed in these organs.

This work evaluated and compared a variety of commonly adopted algorithms for background field removal and dipole inversion of QSM. When applied to patients with CMBs scanned at both 3T and 7T, we found that 7T MRI could provide QSM images with lower noise than 3T MRI. QSIP and VSHARP + iLSQR achieved the highest white matter homogeneity and vein contrast, with QSIP also providing the highest CMB contrast. Compared to ground truth COSMOS QSM images, iLSQR and QSIP had the lowest RMSE and HFEN, while MEDI achieved best SSIM. Although variations were observed among methods throughout the entire brain, overall good correlations between susceptibility values of dipole inversion algorithms and the COSMOS reference were observed in the basal ganglia ROIs, with VSHARP + iLSQR achieving the most similar susceptibility values to COSMOS. This work suggests that selection of QSM method can not only influence the quality of maps obtained but could potentially affect the quantification of

iron or CMB detection in studies that rely on precise quantification of a spatial distribution of susceptibility values.

Chapter 6. Improved Quantitative Susceptibility Mapping using 3D Generative Adversarial Networks

6.1. Introduction

The previous chapter investigated and compared various background removal and dipole inversion algorithms proposed for QSM. From the metric comparison in RMSE, HFEN, and SSIM of the four dipole inversion algorithms compared to COSMOS are listed in Table 5.2. iLSQR had significantly lower RMSE and HFEN compared to other methods ($p < 0.01$) except for QSIP. For HFEN, iLSQR and QSIP were significantly reduced compared to CSC and MEDI ($p < 0.01$), indicating less deviation from COSMOS for these algorithms. MEDI achieved the highest SSIM to COSMOS among all the dipole inversion algorithms ($p < 0.01$), followed by QSIP and iLSQR. Figure 5.10 visually compares all the dipole inversion algorithms against COSMOS QSM (with VSHARP for background field removal). Although all algorithms achieved relatively low residual streaking artifacts, they have distinct appearances due to different regularization approaches.

Table 5.2 and visual comparison shown in Figure 5.10, we can see that although different dipole inversion algorithms try to tackle the ill-posed inverse problem using different regularization, an optimal QSM dipole inversion algorithm is still in need for more accurate QSM reconstruction.

Recently, Deep Convolutional Neural Networks (DCNNs) have shown great potential in computer vision tasks such as image classification [65], semantic segmentation [67] and object detection [98]. Among various deep neural network architectures, U-Net [66] has become the most popular backbone for many medical image-related problems [120]–[122] due to its effectiveness and universality. Bollmann et al. [123] and Yoon et al. [108] adopted the U-Net structure and extended it to 3D to solve the dipole inversion problem of QSM by training the network to learn the inversion using patches of various sizes as the input. Since its inception in

2014, Generative Adversarial Networks (GANs) [51] have been incorporated into CNNs to further improve performance of segmentation, classification, and especially contrast generation tasks [70], [71], [124]–[127] by combining a generator that is trained to generate more realistic and accurate images with a discriminator that is trained to distinguish the real from the generated images. This idea of adversarial learning has recently been extended to applications in medical imaging [128]–[130]. The goals of this study were to for the first time: 1) incorporate the physical principles of the dipole inversion model that describes the susceptibility-phase relationship into the training of a deep neural network to generate QSM images and 2) harness the power of adversarial learning in this new application. We achieved these aims by: 1) modifying the structure of the 3D U-Net proposed by Bollmann et al. [123] and Yoon et al. [108] to incorporate an increased receptive field of the input phase image patches in conjunction with a cropping of resulting output in order to emulate the dipole physics within the structure of the model; and 2) by utilizing a GAN to regularize the model training process and further improve the accuracy of QSM dipole inversion.

6.2. Methods

6.2.1. Theory of QSM dipole inversion and GANs

As described in Chapter 2.2.2, assume that the susceptibility-induced magnetization is regarded as a magnetic dipole and the orientation of the main magnetic field B_0 is defined as the z-axis in the imaging Cartesian coordinate, the magnetic field perturbation and susceptibility distribution is related by a convolution, which can be efficiently calculated by a point-wise multiplication in frequency space [41].

$$\Delta B_z(\mathbf{k}) = B_0 \left(\frac{1}{3} - \frac{k_z^2}{|\mathbf{k}|^2} \right) \chi(\mathbf{k}) \quad (6.1)$$

Where ΔB_z is the local field perturbation, B_0 is the main magnetic field, χ represents the tissue susceptibility, \mathbf{k} is the frequency space vector and k_z is the z-component. In practice, we measure ΔB_z by phase variation and solve the inverse problem for the susceptibility distribution χ . However, notice that when $k_z^2/|\mathbf{k}|^2 \approx 1/3$, the bracket term on the right-hand side becomes close to zero, which causes missing measurements or noise amplification when solving the inverse problem, making it ill-posed.

Assume $y(\mathbf{d}) = \Delta B_z(\mathbf{d})$ is the acquired tissue phase of the subject and $x(\mathbf{d})$ is the susceptibility map of the subject we want to solve in the ill-posed phase-susceptibility dipole inversion problem, and function f represents the relationship between them, then we can simplify equation 6.1 with:

$$y = f(x) \tag{6.2}$$

To solve the dipole inversion problem, we are finding a function h that gives:

$$\tilde{x} = h(y) \tag{6.3}$$

where \tilde{x} is an estimate of the true susceptibility map x . The idea of GANs is to define a game between two competing components (networks): the discriminator (D) and the generator (G). G takes an input and generates a sample that D receives and tries to distinguish from a real sample. The goal of G is to “fool” D by generating more realistic samples. In this case, we use G as the function h :

$$\tilde{x} = G(y) \tag{6.4}$$

The adversarial game between G and D is a minimax objective:

$$\min_G \max_D \mathbb{E}_{x \sim \mathbb{P}_q} [\log D(x)] + \mathbb{E}_{y \sim \mathbb{P}_t} [\log(1 - D(G(y)))] \tag{6.5}$$

where \mathbb{P}_q is the distribution of true susceptibility maps and \mathbb{P}_t is the distribution of tissue phases. To stabilize the training process, we adopt the method of Wasserstein GAN (WGAN) [71], and the value function for WGAN is:

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim \mathbb{P}_q} [D(x)] - \mathbb{E}_{y \sim \mathbb{P}_t} [D(G(y))] \quad (6.6)$$

where \mathcal{D} is the set of 1-Lipschitz functions, which can be enforced by adding a gradient penalty (GP) term to the value function [72]:

$$\lambda_{gp} \mathbb{E}_{y \sim \mathbb{P}_t} [(\|\nabla D(G(y))\|_2 - 1)^2] \quad (6.7)$$

where λ_{gp} is a parameter that controls the weight of the gradient penalty. Since the goal for G in this task is to recover/reconstruct QSM from a certain input tissue phase, we also included an L1 loss as content loss in the objective function of G:

$$\min_G \lambda_c \|x - G(y)\|_1 + \lambda_{adv} L_{adv} \quad (6.8)$$

where L_{adv} is the adversarial loss indicated in equation 6.6.

6.2.2. QSMGAN framework

We designed a 3D U-Net architecture similar to [123] and [108] as the generator part of the QSMGAN framework as shown in Figure 6.1. In each U-Net block, there are two 3x3x3 Conv3d-BatchNorm-LeakyReLU (negative slope of 0.2) layers, where Conv3d is the commonly used 3D convolution layer, the BatchNorm (batch normalization) accelerates and stabilizes the optimization and the LeakyReLU facilitates the training of the GAN. 3D average pooling was used to down-sample the image patch as proposed in the classic U-Net architecture, while 3D transpose convolution was applied to restore the resolution in the up-sampling path while incorporating the low frequency information back in the model. At the end of the generator, we applied a cropping layer to focus the training on only the center part of the patch. For the discriminator part of the QSMGAN, we designed a 3D patch-based convolutional neural network where each block of the network is composed of a 3D convolution (4x4x4 kernel size and stride 2) and a LeakyReLU (negative slope of 0.2). The four blocks in the network lower the input

patch to 1/16 of the original size and the 3D convolution layer at the end converts the resulting patch to a binary output corresponding to the prediction of real and fake QSM patches.

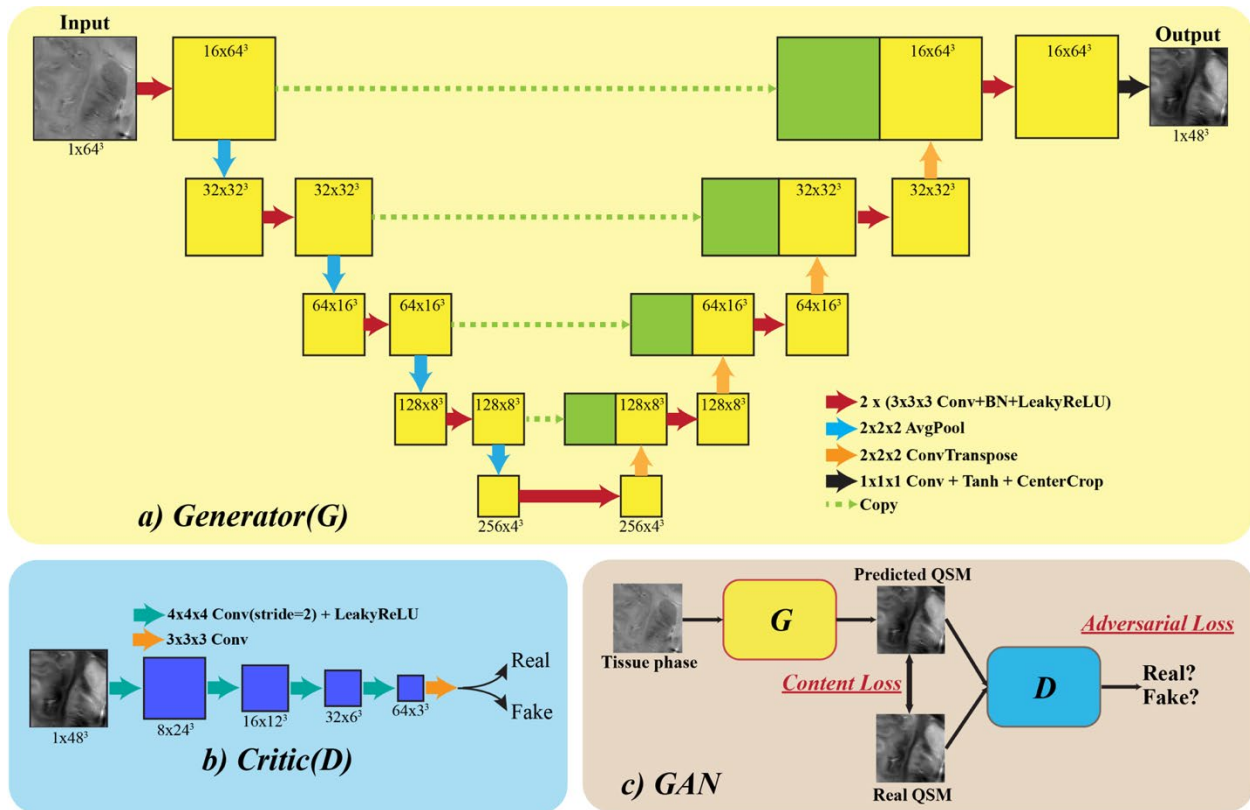


Figure 6.1 QSMGAN network architecture.

a) The generator part of the GAN, which adopts a 3D U-Net with center cropping as a building block. b) The discriminator (“critic” in WGAN-GP) is constructed using 3D convolution with stride=2 to reduce image size. c) The overall GAN structure combines the generator and discriminator, where G is trained to generate more realistic and accurate QSM to fool D and D is trained to distinguish real and generated(fake) QSM.

6.2.3. Subjects and data acquisition

Eight healthy volunteers (average age 28, M/F=3/5) were recruited for this study as the training and validation dataset for QSMGAN. All volunteers were scanned with a 3D multi-echo gradient-recalled sequence (4 echoes, TE= 6/9.5/13/16.5ms, TR=50ms, FA=20°, bandwidth=50kHz, 0.8mm isotropic resolution, FOV=24x24x15cm) using a 32-channel phase-array coil on a 7T MRI scanner (GE Healthcare Technologies, Milwaukee, WI, USA). The sequence was repeated

three times on each volunteer with different head orientations (normal position, tilted forward and tilted left) to acquire data for COSMOS reconstruction. GRAPPA-based parallel imaging [111] with an acceleration factor of 3 and 16 auto-calibration lines were also adopted to reduce the scan time of each orientation to about 17 minutes.

To evaluate the generalization ability of our networks, we used a cohort of 12 patients with brain tumors who had developed CMBs years after being treated with radiation therapy. This type of vascular injury was an ideal pathology to test the generalizability of our network because they can both be extremely small in size and difficult to detect, and have very high susceptibility values compared to normal brain tissue due to deposits of hemosiderin. These patients were scanned using the same 7T QSM protocol as the healthy volunteer subjects except the slice thickness was 1.0mm. Only one orientation scan was performed on each patient. After the GRAPPA reconstruction, the image volumes were resampled to 0.8mm isotropic resolution to match the input of the deep learning models.

6.2.4. QSM data processing and dataset preparation

The raw k-space data were retrieved from the scanner and processed on a Linux workstation using in-house software developed in Matlab 2015b (Mathworks Inc., Natick, MA, USA). The following processing steps (summarized in Figure 6.2) were performed to obtain the tissue phase maps required for input to the QSMGAN and the calculation of the gold standard COSMOS-QSM which was used as the learning target of the QSMGAN: 1) GRAPPA reconstruction was applied to interpolate the missing k-space lines due to parallel imaging acceleration and channel-wise inverse Fourier transform was applied to obtain the coil magnitude and phase images; 2) coil images were combined to obtain robust echo magnitude and phase images using the MCPC-3D-S method [112]; 3) raw phase was unwrapped using a Laplacian-based algorithm [113]; 4) FSL BET [91] was applied on magnitude images from all

echoes to obtain a composite brain mask from the intersection of each individual echo mask; 5) V-SHARP [104] was used to remove the background field phase to get the tissue phase map; 6) images from different orientations were co-registered using magnitude images with FSL FLIRT [91]; 7) the dipole field inversion was solved using the COSMOS algorithm [22]. In addition, TKD [23], MEDI [24] and iLSQR [107] QSM maps were also reconstructed from single orientation data for evaluation and comparison. A threshold of 0.15 was selected for the TKD algorithm, and $\lambda=2000$ was used in MEDI. The reconstructed single orientation tissue phase maps from the patient data were 1) used to compute iLSQR QSM and 2) fed into both the 3D U-Net and QSMGAN networks to generate COSMOS-like QSM.

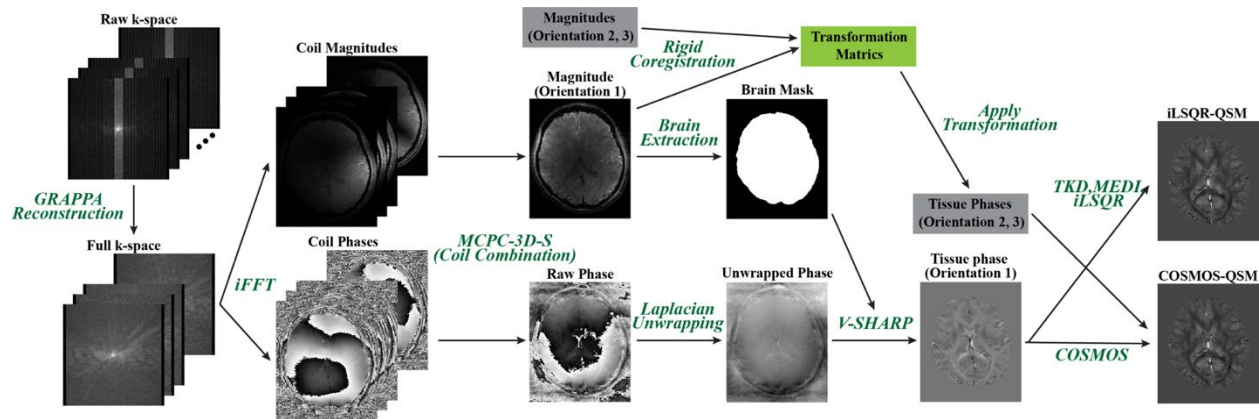


Figure 6.2 QSM data processing pipeline.

This figure shows processing of one scan orientation. Data from the other two orientations were processed similarly and introduced in the gray boxes in this figure to reconstruct the COSMOS-QSM.

6.2.5. Training and validation

The 8 healthy subjects were divided into 5 for training, 1 for validation, and 2 for testing. All three orientations were included in the dataset so the total number of scans in the training/validation/test set was 15/3/6. To build the training set, tissue phase and susceptibility patches were sampled by center coordinates with a gap of 8 voxels in all three spatial dimensions. Since background occupies most of the image volume, we sampled 90% patches

from inside the brain and only 10% from the background to increase the efficiency of the training. For validation and testing, the input tissue phase volume was divided into non-overlapping patches according to the output patch size and the susceptibility map was reconstructed patch-wise by feeding the input tissue phase patch into the trained network.

Figure 6.3 demonstrates the relationship between the receptive field and input/output patch size.

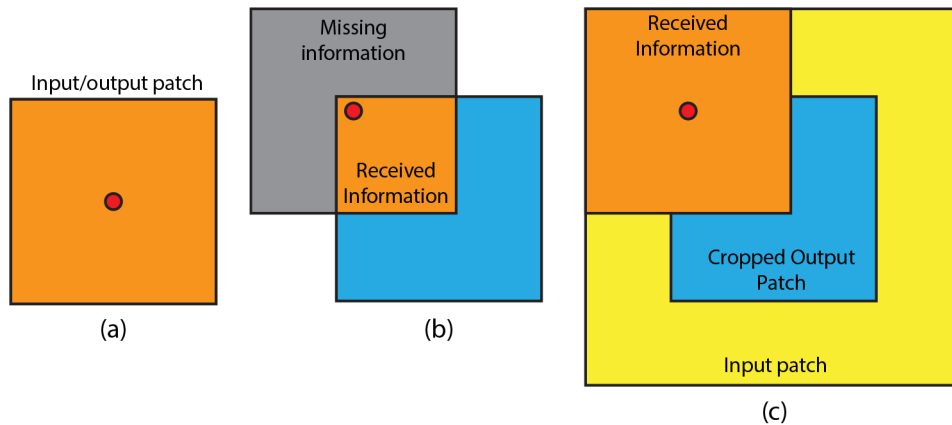


Figure 6.3 Demonstration of the relationship between receptive field and input/output patch size. a) Input patch size = output patch size, red dot represents voxels near the patch center. b) Input patch size = output patch size, voxels near the patch edge receive only information from the orange region. c) Input patch size > output patch size (with center cropping), voxels near the edge receive more information than in b).

To assist the neural network training, we multiplied the input phase by a scale factor of 100 and then transformed the output x by a scaled hyperbolic tangent operation to get the surrogate target \hat{x} :

$$\hat{x} = \tanh(10x) \quad (6.9)$$

This transform not only converts the range of the target susceptibility map to $[-1, 1]$, which aids in the network training, but also results in a more Gaussian distributed histogram, helping the network learn values in different ranges (Figure 6.4).

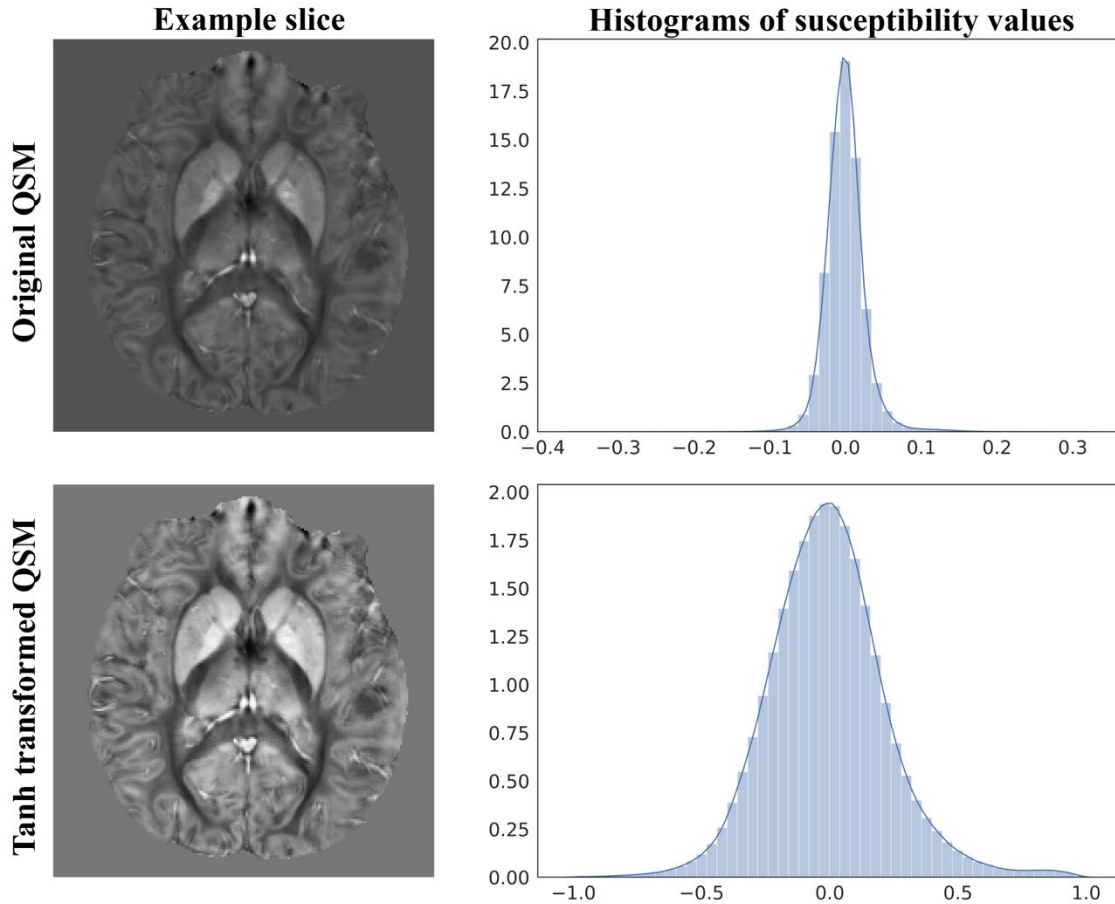


Figure 6.4 Comparison of raw QSM and tanh transformed QSM. An axial slice of the original QSM (top left) and its histogram (top right) compared to the tanh transformed QSM (bottom left) and its histogram (bottom right). We can see that the tanh transform distributed the susceptibility values more evenly between -1.0 and +1.0, resulting in better contrast and value ranges for the network training.

As the baseline network, we first trained the U-Net based generator separately with the pairs of input and output patch sizes listed in Table 6.1. To train the generator, an Adam optimizer with a learning rate of $1e-4$ was used and betas were set to (0.5, 0.999). The network was trained for 40,000 iterations with a batch size of 16 that was lowered to 8 for larger input patch sizes. L1 loss was used as the loss function for the baseline network.

To train the QSMGAN, we again started with the baseline network and then: 1) fixed the generator G and trained D for 20,000 iterations to ensure that D was well trained, as suggested

by Gulrajani et al. [72]; and 2) trained G and D together for 40,000 iterations. During each iteration, D (the critic) was updated 5 times with the gradient penalty $\lambda_{gp} = 100$. Adam optimizers were used for both G and D and the learning rate was lowered to 1e-5. To balance the content loss and adversarial loss, λ_c was set to 1 and λ_{adv} to 0.01.

6.2.6. Evaluation metrics

To evaluate the quality of the predicted QSM map reconstructed by the network (\tilde{x}), we calculated and compared the following metrics: 1) L1 error = $\|x - \tilde{x}\|_1$; 2) Peak Signal-to-Noise Ratio (PSNR) = $10 \log_{10} \left(\frac{R(x)}{MSE(x, \tilde{x})} \right)$, where $R()$ computes the voxel value range of the input image and $MSE()$ computes the mean squared error between the reconstructed image and the target image; 3) Normalized Mean Squared Error (NMSE) = $\frac{MSE(x, \tilde{x})}{\|x\|_2}$; 4) High-frequency error norm (HFEN); and 5) Structure similarity index (SSIM) as described in [115]. A Wilcoxon signed rank test was used to test for statistically significant differences in quality metrics between the optimized 3D U-Net and QSMGAN.

Radiation-induced CMBs from each patient were segmented on reconstructed susceptibility weighted images (SWI) using in-house software [29], [31]. The resulting CMB masks were eroded by 1 voxel in all directions to remove the blooming artifact present on SWI and then applied to the iLSQR QSM, 3D U-Net and QSMGAN maps in order to quantify the median CMB susceptibility from the 3 different QSM images. The number of CMBs were also counted for each patient using each of the 3 QSM maps by an experienced rater after blinded randomization of the images. A Kruskal-Wallis test was used to test for significant differences in median CMB susceptibility and CMB count among the 3 QSM methods and Bland-Altman Plots were used to visualize any discrepancies.

6.3. Results

6.3.1. Baseline 3D U-Net

We experimented with combinations of three different input patch sizes (32^3 , 48^3 , 64^3) and 5 output patch sizes (32^3 , 48^3 , 64^3 , 96^3 , 128^3 , with input > output) for the baseline 3D U-Net. Figure 6.5 demonstrates the qualitative effects of different input-output size pairs (shown on axial slices) while Table 6.2 compares the quantitative metrics (L1, PSNR, NMSE) used to evaluate the quality of the resulting QSM maps. When the input patch size was the same as the output patch size, the inversion error increased towards the edge of the patch, resulting in visible discontinuities in a grid-like pattern in the reconstructed QSM map. The higher L1 error, lower PSNR and higher NMSE supports this phenomenon quantitatively. When we increased the input patch size and applied center cropping at the end of the U-Net as shown in Figure 6.3, the patch edge artifact decreased and the metrics improved. Among the different combinations of patch sizes, the input patch size of 64^3 and the output patch size of 48^3 ($64 \rightarrow 48$) provided the best balance between sufficient accuracy of the U-Net dipole inversion and low computation burden/efficiency. Therefore, for the QSMGAN evaluation we used the $64 \rightarrow 48$ 3D U-Net as a basic building block.

Table 6.1 Test set performance of U-Net baseline with different input and output patch sizes.

3D U-Net Patch Size (input→output)	L1 error (1e-3)	PSNR	NMSE
32→32	1.490±0.184	42.25±1.01	0.302±0.056
48→32	1.403±0.204	43.07±1.22	0.252±0.063
64→32	1.316±0.230	43.39±1.37	0.237±0.072
96→32	1.319±0.216	43.38±1.32	0.237±0.068
48→48	1.424±0.195	42.58±1.13	0.281±0.061
64→48	1.309±0.210	43.53±1.31	0.229±0.065
96→48	1.310±0.212	43.37±1.28	0.237±0.067
128→48	1.311±0.215	43.40±1.31	0.236±0.068
64→64	1.389±0.211	42.87±1.21	0.264±0.063
96→64	1.316±0.207	43.46±1.28	0.233±0.066
128→64	1.322±0.211	43.32±1.27	0.240±0.067

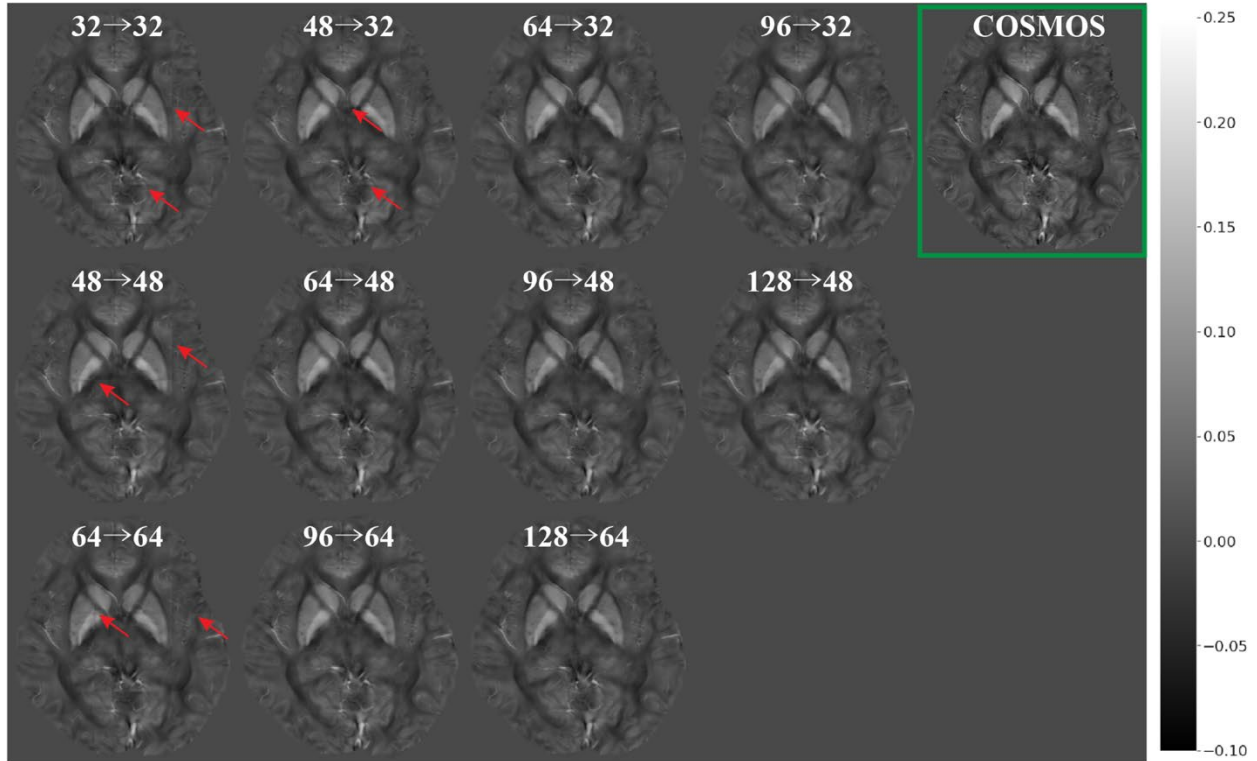


Figure 6.5 Comparison of reconstructed QSM using 3D U-Net with different input/output patch sizes.

The green box highlights the ground truth COSMOS QSM. Red arrows highlight the edge incontinuity artifacts.

Table 6.2 Test set performance of U-Net baseline, QSMGAN and non-learning-based algorithms.

Methods	L1 error (1e-3)	PSNR	NMSE	HFEN	SSIM
TKD	2.826±0.178	38.82±1.69	0.496±0.076	99.84±4.86	0.806±0.023
MEDI	2.909±0.194	41.24±1.71	0.539±0.059	100.99±5.02	0.912±0.027
iLSQR	2.193±0.227	42.03±1.45	0.410±0.088	74.40±7.15	0.896±0.025
3D U-Net 64-->48	1.309±0.210	43.53±1.31	0.229±0.065	48.45±8.30	0.944±0.018
QSMGAN 64-->48	1.199±0.215	44.16±1.42	0.200±0.065	45.68±8.53	0.952±0.018

6.3.2. Effectiveness of QSMGAN

Using the 64→48 3D U-Net as the generator, the metric-wise benefit of using QSMGAN over the 3D U-Net is shown by the quantitative metrics listed in Table 6.2. ($p=0.03$ for all metrics of 3D U-Net vs. QSMGAN) Column 4 and 5 in Figure 6.6 and Figure 6.7 demonstrates the visual comparison of reconstructed QSM of 3D U-Net and QSMGAN, where the adversarial training further improved the quality of the reconstructed QSM map by reducing both residual blurring and the remaining edge discontinuity artifacts from the relatively smaller input patch size, providing a more accurate and detailed mapping of susceptibility compared to the 3D U-Net baseline.

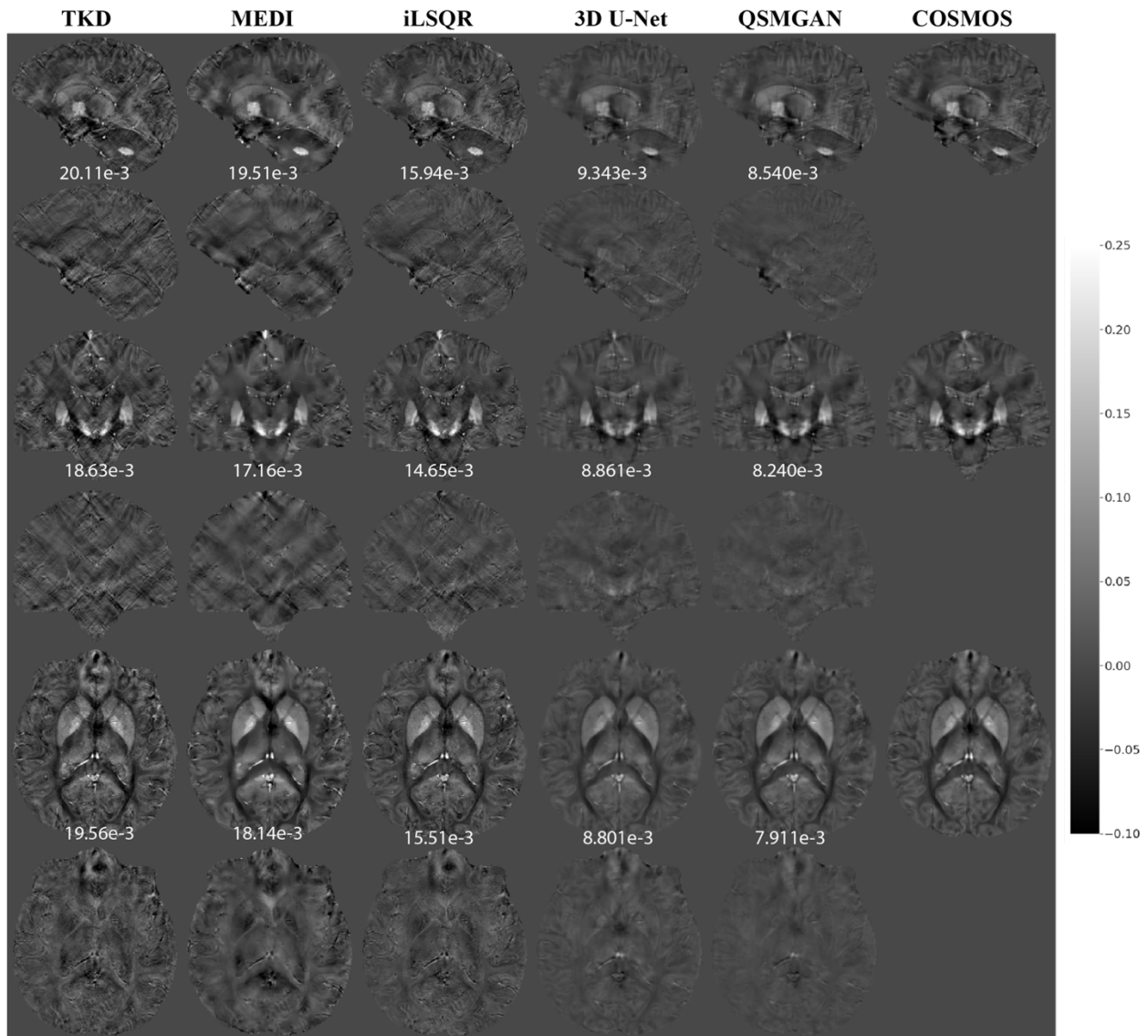


Figure 6.6 Comparison of QSM of test subject 1. Row 1,2: sagittal view and error map. Row 3,4: coronal view and error map. Row 5,6: axial view and error map. Numbers at bottom of each slice show the L1 error relative to COSMOS-QSM of the slice.

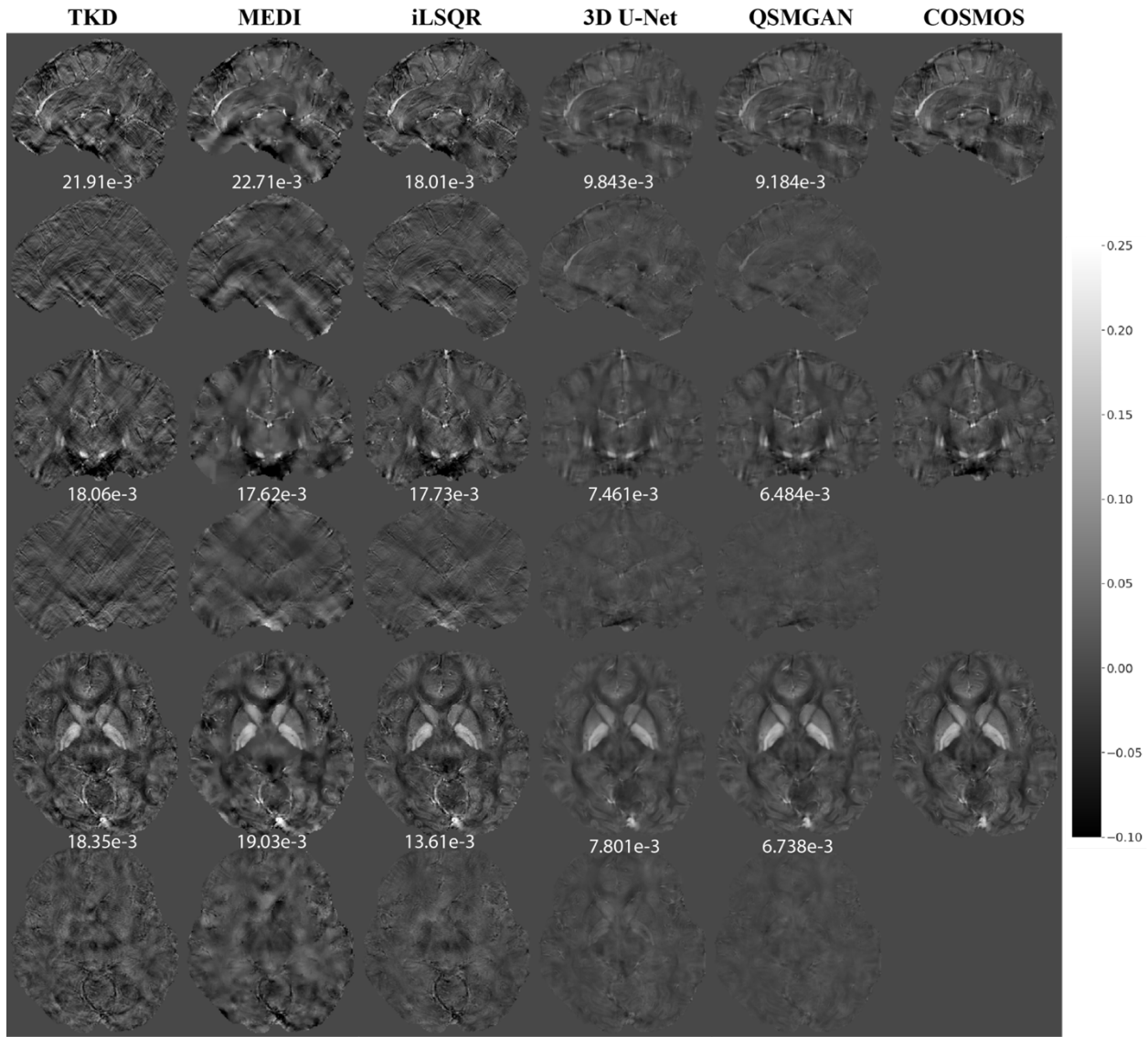


Figure 6.7 Comparison of QSM of test subject 2. Row 1,2: sagittal view and error map. Row 3,4: coronal view and error map. Row 5,6: axial view and error map. Numbers at bottom of each slice show the L1 error relative to COSMOS-QSM of the slice.

6.3.3. Comparison with non-learning-based methods

Compared to 3 common ‘non-learning-based’ QSM dipole inversion algorithms (TKD, MEDI and iLSQR), our QSMGAN approach had 42-59% reductions in NMSE and L1 error in the test datasets while increasing PSNR by 4-13% as shown in Table 6.2. Figure 6.6 and Figure 6.7

show examples of QSM slices from the two test subjects generated from our QSMGAN compared to non-learning-based algorithms. Although TKD had the lowest computational complexity, it also resulted in the most streaking artifacts. Despite its smooth appearance, MEDI was the least uniform with relatively high L1 error and inaccurate contrast of some fine structures such as vessels. It also required the longest computation time of all of the methods (about 2 hours on a regular desktop workstation). Although iLSQR QSM had lower L1 error than TKD and MEDI, it was visually noisier than all other methods. QSMGAN not only resulted in the best L1 error, PSNR, NMSE, HFEN, and SSIM, but achieved the most similar QSM map to COSMOS in only 2 seconds of reconstruction time per scan, the same order of time complexity as with the TKD method.

6.3.4. Application of networks in patients with radiation-induced CMBs

To evaluate the generalization ability of our networks, we tested our network in a cohort of 12 patients with brain tumors treated with prior radiation therapy. The median susceptibility values for each CMB and total number of CMBs per patient based on iLSQR, 3D U-Net and QSMGAN were not significantly different among methods (Kruskal-Wallis test $p=0.149$ and $p=0.936$, respectively; see Figure 6.8). This comparison demonstrates that the proposed QSMGAN could be well generalized to previously unseen pathology with extreme susceptibility values. Figure 6.9 demonstrates the robustness of QSMGAN to artifacts from imperfect preprocessing steps such as skull stripping and background phase removal as well as its ability to generate more uniform susceptibility maps. Patient 8 (row 1) suffered from poor brain extraction and background field removal that resulted in severe susceptibility artifacts from the air-tissue interface in the sinuses in the iLSQR QSM image. Although 3D U-Net partially alleviated this problem, QSMGAN provided the most uniform and highest quality susceptibility map with the least amount of residual artifacts. Patient 12 (row 2) had residual background phase that

obscured the detection of a microbleed (denoted by the red arrow) that was correctly visualized on both the deep learning-based QSM maps.

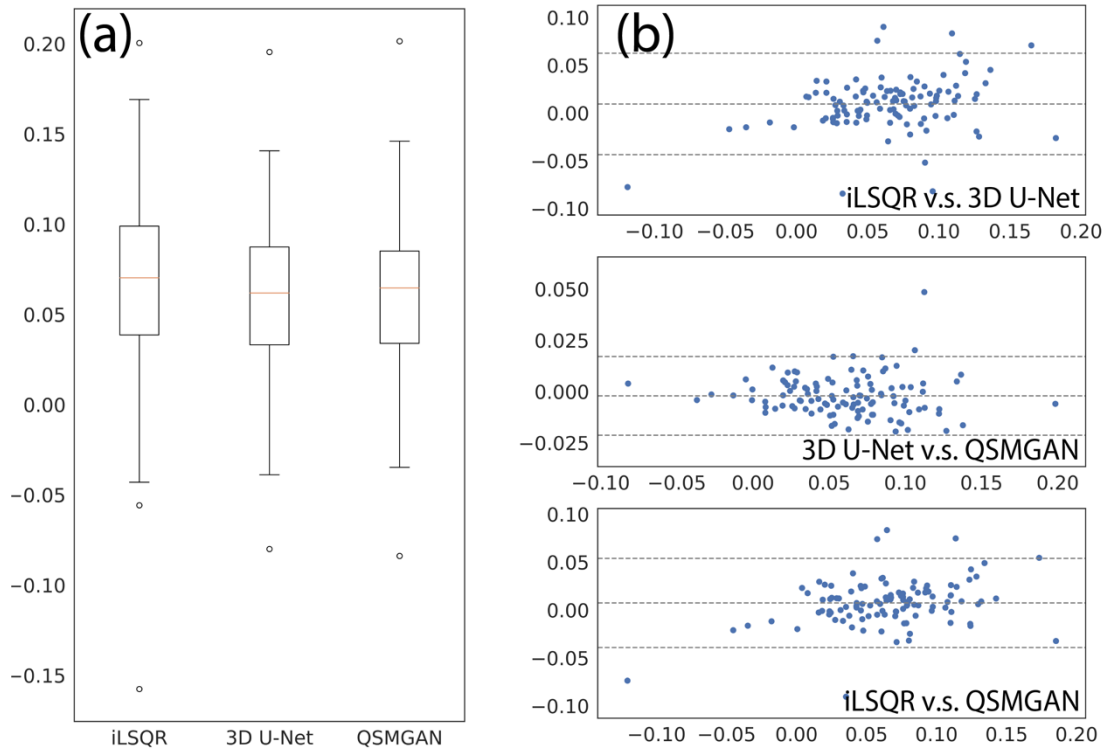


Figure 6.8 Comparison of median CMB susceptibilities measured from different QSM algorithms.

a) box plot of median CMB susceptibilities. B) Bland-Altman plots of algorithm pairs.

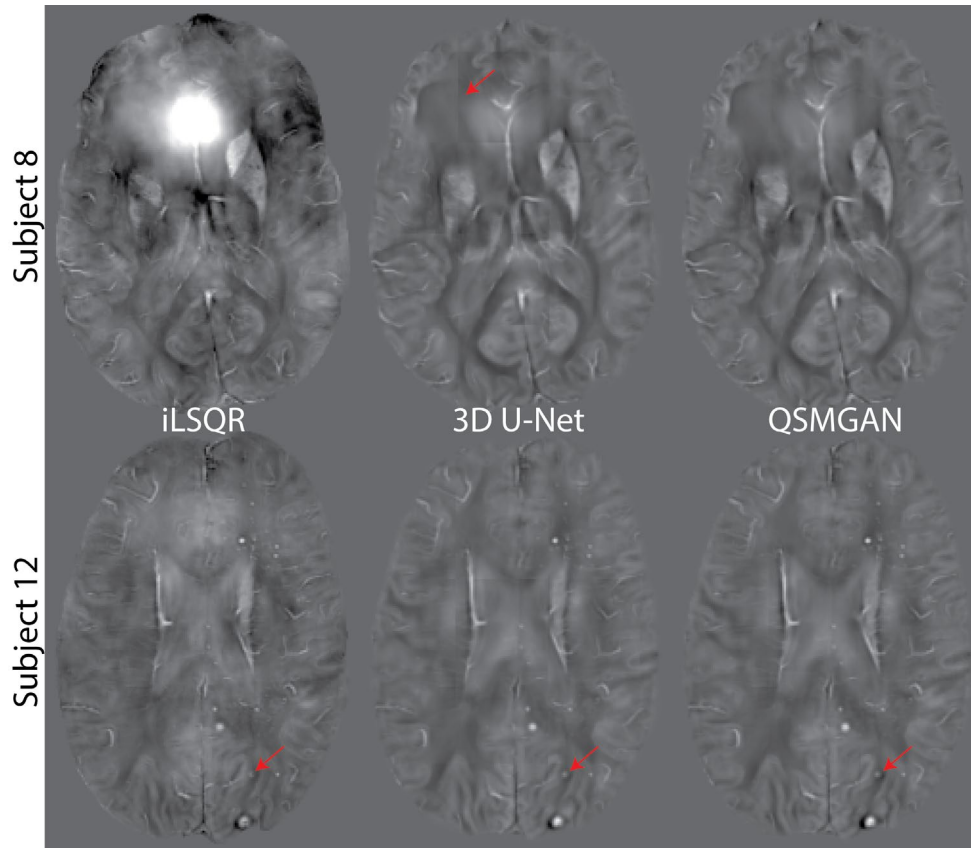


Figure 6.9 QSM of two patients with brain tumors who had developed cerebral microbleeds due to prior radiation therapy.

Subject 8 suffered from poor brain extraction and background field removal that resulted in severe susceptibility artifacts in the iLSQR QSM image. Both 3D U-Net and QSMGAN successfully removed the artifact but QSMGAN generated higher quality maps with less edge discontinuity artifacts, as highlighted by the red arrow. Subject 12 had residual background phase that obscured the detection of a microbleed (denoted by the red arrow) that was correctly visualized on both the deep learning-based QSM maps. The input-output size was 64-->48 for both subjects.

6.4. Discussion and conclusions

Although in theory the phase-susceptibility relationship in QSM is global, meaning the tissue phase is determined by the susceptibility of all locations in the imaging volume, we still adopted a patch-based deep learning approach similar to Yoon et al., 2018 for several reasons. Since the network is 3D, the patch-based method can significantly reduce the computation complexity and memory requirement compared to whole-volume based approaches, especially when

conducting high-resolution QSM. For example, if we needed to generate a full QSM volume with a $256 \times 256 \times 150$ matrix size using the entire volume as an input to the 3D U-Net architecture, even the most advanced GPU with 32GB of graphics memory would not be able to fit a single training sample. The patch-based method also converts one single scan into hundreds of input images, even before data augmentation. Since COSMOS requires a relatively long scan time and is cumbersome to conduct, training a more generalizable deep convolutional network is beneficial when only a limited amount of data is available. Because the phase is mostly determined by nearby susceptibilities due to the properties of the susceptibility-phase convolutional kernel, the patch-based approach yields a good approximation of the dipole inversion.

As Table 6.1 demonstrates, increasing the input patch size and applying center cropping at the end of the 3D U-Net significantly improved the quality of the reconstructed QSM maps. This can be intuitively described by Figure 6.3, where when the input patch size equaled the output patch size, an output voxel near the center of the patch (panel (a) of Figure 6.3) could receive information from the entire patch. However, a voxel near the edge of the output patch (panel (b) of Figure 6.3) would only receive information from the orange region and a large portion of the phase information from the gray region would be missing, reducing the ability of the network to accurately solve for the susceptibility. When we increased the input patch size (panel (c) of Figure 6.3) and cropped the output patch such that only the center of the patch was considered a valid QSM prediction, voxels near the edge of the patch regained phase input information thereby increasing the accuracy of the quantified susceptibility values.

Another observation from Table 6.1 is that the medium output patch size (48^3) achieved the best QSM reconstruction performance. The smaller patch size (32^3) performed worse because the output voxels received less information, introducing more error to the patch approximation of global convolution. Unexpectedly, the larger patch size (64^3) didn't provide any extra benefit to

the dipole inversion. This might be due to the fact that it introduced more variables into the computation process and increased the difficulty of training a good network for QSM reconstruction. In addition, for each output patch size, using excessively large input patches (such as 96→32) did not further reduce the error but slightly downgraded the QSM quality. This might be due to increased information far from the output patch interfering with the dipole inversion.

A disadvantage of using an excessively large input patch size is the dramatically increased computational complexity and GPU memory requirement. Note that the network is three-dimensional and the computational complexity and memory requirement of training the networks roughly increases with the input patch size by $O(n^3)$. The center cropping we applied to ensure a large enough receptive field, only exacerbated this problem, greatly reducing the efficiency of the prediction process. For example, if we increased the input patch size from 32^3 to 64^3 , the training/prediction time and memory became 8x as long and only 1/8 of the computed patches were utilized. Based on the observation that excessively large input patch sizes greatly increased the computational burden without improving the quality of the resulting QSM maps, we selected the 64→48 3D U-Net as the base network to integrate with the GAN.

The rationale for the GAN training, which included adding a discriminator or “critic”, was to guide the generator (or the 3D U-Net) to further refine its result so that it could not be distinguished from a real COSMOS QSM patch. Although it took a long time (48 hours) to train the QSMGAN, once the training was finished the discriminator was no longer needed. As a result, reconstruction or prediction of the QSM map for a new scan/subject from tissue phase only required one forward pass through the 3D U-Net for each input patch, thereby resulting in a computational complexity that is identical to the 3D U-Net baseline.

Although the QSMGAN was trained only on healthy volunteer data, when applying the network to patient data, it successfully recovered the previously unseen pathology of cerebral

microbleeds and assigned values similar to those obtained from iLSQR. This demonstrated that the networks avoided overfitting and managed to learn the underlying dipole convolution relationship between tissue phase and susceptibility sources. We also observed unforeseen robustness to imperfect preprocessing from QSMGAN. This was likely due to the fact that our QSMGAN was trained on carefully processed training data with little artifacts, so the generator would favor outputs with similar image quality and therefore tended to remove any abnormal susceptibility sources and remaining background phase components. Although our QSMGAN was trained on only brain images because QSM has been most widely utilized in the brain, the network can easily be trained using data from other organs of interest.

In this study, we implemented a 3D U-Net deep convolutional neural network approach to improve the dipole inversion problem in QSM reconstruction. To better approximate the global convolution property in the phase-susceptibility relationship through patch-based neural networks, we enlarged the input patch size and introduced center cropping to ensure an increased input receptive field for all neural network outputs. This cropping technique provided significantly lower edge discontinuity artifacts and higher accuracy. Including a generative adversarial network based on the WGAN-GP technique further improved the stability of training process, the image quality, and the accuracy of the susceptibility quantification. Compared to the other traditional non-learning dipole inversion algorithms such as TKD, MEDI and iLSQR, our proposed method could efficiently generate more accurate, COSMOS-like QSM maps from single-orientation, background-field-removed, tissue phase images. When tested on patients with radiation-induced CMBs, QSMGAN improved the robustness of the QSM reconstruction without sacrificing the sensitivity of CMB detection. Future directions include investigating the network's ability to generalize to other scan parameters (such as TE, TR, and image resolution) and evaluating the performance of QSMGAN in patients with different pathologies and in other organs to ultimately improve patient care.

Chapter 7. Conclusions and future directions

7.1. Conclusions

This dissertation investigated the techniques of susceptibility imaging using MR, developed deep learning methods combined with these techniques, and analyzed their utility in different neurological applications.

Chapter 3 and Chapter 4 focused on combining the SWI technique with deep convolutional neural networks to facilitate automatic detection and segmentation of CMBs. Chapter 3 proposed a deep residual neural network architecture for false positive reduction and demonstrated its effectiveness in a cohort of patients with brain tumors treated with radiotherapy. Chapter 4 further confirmed the generalizability of the model designed and trained in Chapter 3 by applying it to scans with different parameters and patients with CCM. Chapter 4 also tested and compared various model training strategies with different amounts of available data and optimization goals.

Chapter 5 and Chapter 6 focused on the quantitative MR phase imaging technique, QSM. Chapter 5 investigated multiple crucial algorithms for performing QSM, including background removal and dipole inversion algorithms. This chapter evaluated and compared these algorithms with clinically informed metrics in both volunteers and patients with brain tumors. Chapter 6 further focuses on improving the dipole inversion step for QSM with the aid of 3D DCNN and GAN. By training a WGAN with gradient penalty, the network was able to reconstruct COSMOS-grade QSM within seconds from single orientation scan data, greatly improving the accuracy, efficiency, and quality of QSM.

7.2. Future directions

The automatic CMB detection algorithm using deep learning presented and evaluated in Chapter 3 and Chapter 4 is still based upon a candidate proposal algorithm that depends on traditional image processing algorithms and the manual tuning of appropriate features and thresholds. A future direction of this project is to train an end-to-end CMB detection and segmentation network to eliminate the candidate proposal algorithm and maximize the benefit of representation learning brought by neural networks.

Chapter 5 compared different QSM algorithms in multiple dimensions. The next step will be applying the knowledge we learn from this study to different clinical and research scenarios and optimize the quality of the resulting QSM images, especially in cases where more extreme susceptibility values are observed.

The deep neural network, or QSMGAN, developed Chapter 6 only performs dipole inversion on tissue phase maps, which still requires multiple steps of careful image processing. Based on the large model capacity a DCNN could provide, a future direction worth exploring is to build and train a single-step end-to-end network that can reconstruct QSM from raw phase images, which has the potential to further reduce propagation error between multiple steps.

References

- [1] Z.-P. Liang and P. C. Lauterbur, *Principles of magnetic resonance imaging: a signal processing perspective*. SPIE Optical Engineering Press, 2000.
- [2] D. G. Nishimura, *Principles of magnetic resonance imaging*. Stanford University, 1996.
- [3] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI from Picture to Proton*. Cambridge university press, 2017.
- [4] A. Antoniou, *Digital signal processing*. McGraw-Hill, 2016.
- [5] K. T. Block, M. Uecker, and J. Frahm, “Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint,” *Magn. Reson. Med. An Off. J. Int. Soc. Magn. Reson. Med.*, vol. 57, no. 6, pp. 1086–1098, 2007.
- [6] M. Lustig, J. H. Lee, D. L. Donoho, and J. M. Pauly, “Faster imaging with randomly perturbed, under-sampled spirals and l1 reconstruction,” in *Proceedings of the 13th annual meeting of ISMRM, Miami Beach*, 2005, p. 685.
- [7] M. A. Griswold *et al.*, “Generalized autocalibrating partially parallel acquisitions (GRAPPA),” *Magn. Reson. Med. An Off. J. Int. Soc. Magn. Reson. Med.*, vol. 47, no. 6, pp. 1202–1210, 2002.
- [8] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “SENSE: sensitivity encoding for fast MRI,” *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, 1999.
- [9] C. Liu, W. Li, K. a Tong, K. W. Yeom, and S. Kuzminski, “Susceptibility-weighted imaging and quantitative susceptibility mapping in the brain.,” *J. Magn. Reson. imaging*, vol. 42, no. 1, pp. 23–41, 2015.
- [10] S. Liu *et al.*, “Susceptibility-weighted imaging: current status and future directions,” *NMR Biomed.*, vol. 30, no. 4, p. e3552, 2017.

- [11] E. M. Haacke, "Susceptibility weighted imaging," vol. 618, pp. 612–618, 2003.
- [12] E. M. Haacke, S. Mittal, Z. Wu, J. Neelavalli, and Y. C. N. Cheng, "Susceptibility-weighted imaging: Technical aspects and clinical applications, part 1," *Am. J. Neuroradiol.*, vol. 30, no. 1, pp. 19–30, 2009.
- [13] C. Denk and A. Rauscher, "Susceptibility weighted imaging with multiple echoes," *J. Magn. Reson. Imaging*, vol. 31, no. 1, pp. 185–191, 2010.
- [14] F. Schweser, A. Deistung, B. W. Lehr, and J. R. Reichenbach, "Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: An approach to in vivo brain iron metabolism?," *Neuroimage*, vol. 54, no. 4, pp. 2789–2807, 2011.
- [15] A. Deistung, F. Schweser, and J. R. Reichenbach, "Overview of quantitative susceptibility mapping," *NMR Biomed.*, no. December 2015, 2016.
- [16] C. Liu, "Susceptibility tensor imaging," *Magn. Reson. Med.*, vol. 63, no. 6, pp. 1471–1477, 2010.
- [17] S. D. Robinson, K. Bredies, D. Khabipova, B. Dymerska, J. P. Marques, and F. Schweser, "An illustrated comparison of processing methods for MR phase imaging and QSM: combining array coil signals and phase unwrapping," *NMR Biomed.*, vol. 30, no. 4, 2017.
- [18] M. a Schofield and Y. Zhu, "Fast phase unwrapping algorithm for interferometric applications.," *Opt. Lett.*, vol. 28, no. 14, pp. 1194–1196, 2003.
- [19] F. Schweser, S. D. Robinson, L. de Rochefort, W. Li, and K. Bredies, "An illustrated comparison of processing methods for phase MRI and QSM: Removal of background field contributions from sources outside the region of interest," *NMR Biomed.*, no. November 2015, 2016.

- [20] W. Li, A. V. Avram, B. Wu, X. Xiao, and C. Liu, "Integrated Laplacian-based phase unwrapping and background phase removal for quantitative susceptibility mapping," *NMR Biomed.*, vol. 27, no. 2, pp. 219–227, 2014.
- [21] D. Zhou, T. Liu, P. Spincemaille, and Y. Wang, "Background field removal by solving the Laplacian boundary value problem," *NMR Biomed.*, vol. 27, no. 3, pp. 312–319, 2014.
- [22] T. Liu, P. Spincemaille, L. De Rochefort, B. Kressler, and Y. Wang, "Calculation of susceptibility through multiple orientation sampling (COSMOS): a method for conditioning the inverse problem from measured magnetic field map to susceptibility source image in MRI," *Magn. Reson. Med. An Off. J. Int. Soc. Magn. Reson. Med.*, vol. 61, no. 1, pp. 196–204, 2009.
- [23] K. Shmueli, J. A. de Zwart, P. van Gelderen, T.-Q. Li, S. J. Dodd, and J. H. Duyn, "Magnetic susceptibility mapping of brain tissue in vivo using MRI phase data.," *Magn. Reson. Med.*, vol. 62, no. 6, pp. 1510–1522, 2009.
- [24] T. Liu *et al.*, "Morphology enabled dipole inversion (MEDI) from a single-angle acquisition: Comparison with COSMOS in human brain imaging," *Magn. Reson. Med.*, vol. 66, no. 3, pp. 777–783, 2011.
- [25] S. Mittal, Z. Wu, J. Neelavalli, and E. M. Haacke, "Susceptibility-weighted imaging: Technical aspects and clinical applications, part 2," *Am. J. Neuroradiol.*, vol. 30, no. 2, pp. 232–252, 2009.
- [26] M. Wahl, M. Anwar, C. Hess, S. M. Chang, and J. M. Lupo, "Relationship between radiation dose and microbleed formation in patients with malignant glioma," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 96, no. 2, p. E68, 2016.
- [27] S. R. S. Barnes, E. M. Haacke, M. Ayaz, A. S. Boikov, W. Kirsch, and D. Kido, "Semiautomated detection of cerebral microbleeds in magnetic resonance images,"

- Magn. Reson. Imaging*, vol. 29, no. 6, pp. 844–852, 2011.
- [28] H. J. Kuijf *et al.*, “Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform,” *Neuroimage*, vol. 59, no. 3, pp. 2266–2273, 2012.
- [29] M. A. Morrison *et al.*, “A user-guided tool for semi-automated cerebral microbleed detection and volume segmentation: Evaluating vascular injury and data labelling for machine learning,” *NeuroImage Clin.*, vol. 20, no. August, pp. 498–505, 2018.
- [30] Q. Dou *et al.*, “Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.
- [31] Y. Chen, J. E. Villanueva-Meyer, M. A. Morrison, and J. M. Lupo, “Toward Automatic Detection of Radiation-Induced Cerebral Microbleeds Using a 3D Deep Residual Network,” *J. Digit. Imaging*, 2018.
- [32] S. Liu *et al.*, “Cerebral microbleed detection using Susceptibility Weighted Imaging and deep learning,” *Neuroimage*, vol. 198, pp. 271–282, 2019.
- [33] Y. Wang *et al.*, “Clinical quantitative susceptibility mapping (QSM): Biometal imaging and its emerging roles in patient care,” *J. Magn. Reson. Imaging*, pp. 1–21, 2017.
- [34] J. Acosta-Cabronero, G. B. Williams, A. Cardenas-Blanco, R. J. Arnold, V. Lupson, and P. J. Nestor, “In vivo quantitative susceptibility mapping (QSM) in Alzheimer’s disease,” *PLoS One*, vol. 8, no. 11, 2013.
- [35] A. K. Lotfipour *et al.*, “High resolution magnetic susceptibility mapping of the substantia nigra in Parkinson’s disease,” *J. Magn. Reson. Imaging*, vol. 35, no. 1, pp. 48–55, 2012.
- [36] J. M. G. Van Bergen *et al.*, “Quantitative susceptibility mapping suggests altered brain iron in premanifest Huntington disease,” *Am. J. Neuroradiol.*, vol. 37, no. 5, pp. 789–796,

2016.

- [37] D. Schubert and M. Chevion, "The role of iron in beta amyloid toxicity," *Biochem. Biophys. Res. Commun.*, vol. 216, no. 2, pp. 702–707, 1995.
- [38] J. H. O. Barbosa *et al.*, "Quantifying brain iron deposition in patients with Parkinson's disease using quantitative susceptibility mapping, R2 and R2*," *Magn. Reson. Imaging*, vol. 33, no. 5, pp. 559–565, 2015.
- [39] S. Eskreis-Winkler *et al.*, "The clinical utility of QSM: disease diagnosis, medical management, and surgical planning," *NMR Biomed.*, vol. 30, no. 4, p. e3668, 2017.
- [40] A. C. L. Ng *et al.*, "Iron accumulation in the basal ganglia in Huntington's disease: cross-sectional data from the IMAGE-HD study," *J Neurol Neurosurg Psychiatry*, vol. 87, no. 5, pp. 545–549, 2016.
- [41] C. Liu, H. Wei, N. Gong, M. Cronin, R. Dibb, and K. Decker, "Quantitative Susceptibility Mapping: Contrast Mechanisms and Clinical Applications," *Tomography*, vol. 1, no. 1, pp. 3–17, 2015.
- [42] L. Pauling and C. D. Coryell, "The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin," *Proc. Natl. Acad. Sci.*, vol. 22, no. 4, pp. 210–216, 1936.
- [43] W. Chen *et al.*, "Intracranial calcifications and hemorrhages: characterization with quantitative susceptibility mapping.," *Radiology*, vol. 270, no. 2, pp. 496–505, 2014.
- [44] T. Liu, K. Surapaneni, M. Lou, L. Cheng, P. Spincemaille, and Y. Wang, "Cerebral microbleeds: burden assessment by using quantitative susceptibility mapping.," *Radiology*, vol. 262, no. 1, pp. 269–78, 2012.
- [45] E. Alpaydin, *Introduction to machine learning*. MIT press, 2009.

- [46] C. M. Bishop, *Pattern recognition and machine learning*. Springer Science+ Business Media, 2006.
- [47] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [48] T. M. Mitchell, *The discipline of machine learning*, vol. 9. Carnegie Mellon University, School of Computer Science, Machine Learning ..., 2006.
- [49] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited, 2016.
- [50] G. E. Hinton, T. J. Sejnowski, and T. A. Poggio, *Unsupervised learning: foundations of neural computation*. MIT press, 1999.
- [51] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [52] R. S. Sutton and A. G. Barto, *Introduction to reinforcement learning*, vol. 2, no. 4. MIT press Cambridge, 1998.
- [53] G. Maicas, G. Carneiro, A. P. Bradley, J. C. Nascimento, and I. Reid, "Deep reinforcement learning for active breast lesion detection from DCE-MRI," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 665–673.
- [54] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Trans. neural networks Learn. Syst.*, vol. 29, no. 6, pp. 2063–2079, 2018.
- [55] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines*,

regularization, optimization, and beyond. MIT press, 2001.

- [56] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
- [57] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer (Long. Beach. Calif.)*, vol. 29, no. 3, pp. 31–44, 1996.
- [58] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [59] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [60] S. Haykin, *Neural Networks and Learning Machines, 3/E*. Pearson Education India, 2010.
- [61] D. M. Hawkins, "The problem of overfitting," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, 2004.
- [62] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [63] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.
- [64] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.
- [65] 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, pp. 770–778.
- [66] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, May 2015.

- [67] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [68] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in neural information processing systems*, 2016, pp. 2234–2242.
- [69] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [70] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *arXiv Prepr. arXiv1511.06434*, pp. 1–16, 2015.
- [71] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv Prepr. arXiv1701.07875*, 2017.
- [72] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved Training of Wasserstein GANs," *arXiv Prepr. arXiv1704.00028*, 2017.
- [73] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Z. Med. Phys.*, vol. 29, no. 2, pp. 102–127, 2019.
- [74] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, p. 487, 2018.
- [75] H. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, and C. V Sep, "Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks."

- [76] M. Havaei *et al.*, “Brain tumor segmentation with deep neural networks,” *Med. Image Anal.*, vol. 35, pp. 18–31, 2017.
- [77] T. L. Kline *et al.*, “Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys,” *J. Digit. Imaging*, vol. 30, no. 4, pp. 442–448, 2017.
- [78] L. Chen, P. Bentley, and D. Rueckert, “Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks,” *NeuroImage Clin.*, vol. 15, pp. 633–643, 2017.
- [79] J. Linn *et al.*, “Prevalence of superficial siderosis in patients with cerebral amyloid angiopathy,” *Neurology*, vol. 74, no. 17, pp. 1346–1350, 2010.
- [80] H. Kato, M. Izumiyama, K. Izumiyama, A. Takahashi, and Y. Itoyama, “Silent cerebral microbleeds on T2*-weighted MRI: correlation with stroke subtype, stroke recurrence, and leukoaraiosis,” *Stroke*, vol. 33, no. 6, pp. 1536–1540, 2002.
- [81] H. Hanyu, Y. Tanaka, S. Shimizu, M. Takasaki, and K. Abe, “Cerebral microbleeds in Alzheimer’s disease,” *J. Neurol.*, vol. 250, no. 12, pp. 1496–1497, 2003.
- [82] K. M. Kinnunen *et al.*, “White matter damage and cognitive impairment after traumatic brain injury,” *Brain*, vol. 134, no. 2, pp. 449–463, 2010.
- [83] W. Bian, C. P. Hess, S. M. Chang, S. J. Nelson, and J. M. Lupo, “Susceptibility-weighted MR imaging of radiation therapy-induced cerebral microbleeds in patients with glioma: A comparison between 3T and 7T,” *Neuroradiology*, vol. 56, no. 2, pp. 91–96, 2014.
- [84] A. Charidimou, A. Krishnan, D. J. Werring, and H. Rolf Jäger, “Cerebral microbleeds: A guide to detection and clinical relevance in different disease settings,” *Neuroradiology*, vol. 55, no. 6, pp. 655–674, 2013.

- [85] T. L. A. van den Heuvel *et al.*, “Computer aided detection of brain micro-bleeds in traumatic brain injury,” in *Medical Imaging 2015: Computer-Aided Diagnosis*, 2015, vol. 9414, p. 94142F.
- [86] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, vol. 1, no. 2, p. 3.
- [87] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv Prepr. arXiv1502.03167*, 2015.
- [88] W. Bian, C. P. Hess, S. M. Chang, S. J. Nelson, and J. M. Lupo, “Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images,” *NeuroImage Clin.*, vol. 2, no. 1, pp. 282–290, 2013.
- [89] W. Bian *et al.*, “Simultaneous imaging of radiation-induced cerebral microbleeds, arteries and veins, using a multiple gradient echo sequence at 7 Tesla,” *J. Magn. Reson. Imaging*, vol. 42, no. 2, pp. 269–279, 2015.
- [90] J. M. Lupo *et al.*, “GRAPPA-based susceptibility-weighted imaging of normal volunteers and patients with brain tumor at 7 T,” *Magn. Reson. Imaging*, vol. 27, no. 4, pp. 480–488, 2009.
- [91] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, “FSL,” *Neuroimage*, vol. 62, no. 2, pp. 782–790, Aug. 2012.
- [92] G. E. Dahl, T. N. Sainath, and G. E. Hinton, “Improving deep neural networks for LVCSR using rectified linear units and dropout,” in *2013 IEEE international conference on acoustics, speech and signal processing*, 2013, pp. 8609–8613.
- [93] F. Chollet, “Keras.” 2015.

- [94] M. Abadi *et al.*, “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” *arXiv Prepr. arXiv1603.04467*, 2016.
- [95] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv Prepr. arXiv1412.6980*, 2014.
- [96] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv Prepr. arXiv1409.1556*, 2014.
- [97] C. Szegedy *et al.*, “Going Deeper with Convolutions,” *arXiv:1409.4842*, 2014.
- [98] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [99] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges,” *J. Digit. Imaging*, vol. 32, no. 4, pp. 582–596, Aug. 2019.
- [100] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, 2015.
- [101] T. Liu, I. Khalidov, L. De Rochefort, and P. Spincemaille, “A novel background field removal method for MRI using projection onto dipole fields (PDF),” no. March, pp. 1129–1136, 2011.
- [102] H. Sun and A. H. Wilman, “Background field removal using spherical mean value filtering and Tikhonov regularization,” *Magn. Reson. Med.*, vol. 71, no. 3, pp. 1151–1157, 2014.
- [103] W. Li, B. Wu, and C. Liu, “iHARPERELLA: an improved method for integrated 3D phase unwrapping and background phase removal,” *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 23, no. 1, p. 3313, 2015.

- [104] B. Wu, W. Li, A. Guidon, and C. Liu, "Whole brain susceptibility mapping using compressed sensing," *Magn. Reson. Med.*, vol. 67, no. 1, pp. 137–147, 2012.
- [105] T. Liu, P. Spincemaille, L. De Rochefort, B. Kressler, and Y. Wang, "Calculation of Susceptibility Through Multiple Magnetic Field Map to Susceptibility Source Image in," vol. 204, pp. 196–204, 2009.
- [106] J. Liu *et al.*, "Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map," *Neuroimage*, vol. 59, no. 3, pp. 2560–2568, 2012.
- [107] W. Li *et al.*, "A method for estimating and removing streaking artifacts in quantitative susceptibility mapping," *Neuroimage*, vol. 108, pp. 111–122, 2015.
- [108] J. Yoon *et al.*, "Quantitative susceptibility mapping using deep neural network: QSMnet," *Neuroimage*, vol. 179, pp. 199–206, Oct. 2018.
- [109] I. Chatnuntawech *et al.*, "Single-step quantitative susceptibility mapping with variational penalties," *NMR Biomed.*, no. October 2015, 2016.
- [110] C. Poynton, M. Jankinson, E. Adalsteinsson, E. Sullivan, A. Pfefferbaum, and W. Wells, "Quantitative Susceptibility Mapping by Inversion of a Perturbation Field Model: Correlation with Brain Iron in Normal Aging," *IEEE Trans. Med. Imaging*, vol. 34, no. 1, pp. 339–353, 2015.
- [111] P. J. Beatty *et al.*, "Design of k-space channel combination kernels and integration with parallel imaging," *Magn. Reson. Med.*, vol. 71, no. 6, pp. 2139–2154, 2014.
- [112] K. Eckstein *et al.*, "Computationally Efficient Combination of Multi-channel Phase Data From Multi-echo Acquisitions (ASPIRE)," *Magn. Reson. Med.*, vol. c, pp. 1–11, 2017.
- [113] W. Li, B. Wu, and C. Liu, "Quantitative susceptibility mapping of human brain reflects

- spatial variation in tissue composition,” *Neuroimage*, vol. 55, no. 4, pp. 1645–1656, 2011.
- [114] T. Liu *et al.*, “Morphology enabled dipole inversion (MEDI) from a single-angle acquisition: Comparison with COSMOS in human brain imaging,” *Magn. Reson. Med.*, vol. 66, no. 3, pp. 777–783, 2011.
- [115] C. Langkammer *et al.*, “Quantitative susceptibility mapping: Report from the 2016 reconstruction challenge,” *Magn. Reson. Med.*, vol. 79, no. 3, pp. 1661–1673, 2018.
- [116] S. Ravishankar and Y. Bresler, “MR image reconstruction from highly undersampled k-space data by dictionary learning,” *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.
- [117] E. P. Simoncelli, H. R. Sheikh, A. C. Bovik, and Z. Wang, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [118] Y. Zhang, H. Wei, M. J. Cronin, N. He, F. Yan, and C. Liu, “Longitudinal atlas for normative human brain development and aging over the lifespan using quantitative susceptibility mapping,” *Neuroimage*, vol. 171, no. October 2017, pp. 176–189, 2018.
- [119] M. J. Cronin, N. Wang, K. S. Decker, H. Wei, W.-Z. Zhu, and C. Liu, “Exploring the origins of TE-dependent QSM measurements in healthy tissue and cerebral microbleeds,” *Neuroimage*, vol. 149, no. October 2016, pp. 98–113, 2017.
- [120] J. Kleesiek *et al.*, “Deep MRI brain extraction: A 3D convolutional neural network for skull stripping,” *Neuroimage*, vol. 129, pp. 460–469, 2016.
- [121] J. Zbontar *et al.*, “fastMRI: An Open Dataset and Benchmarks for Accelerated MRI,” *arXiv Prepr. arXiv1811.08839*, pp. 1–29, 2018.
- [122] E. Gong, J. M. Pauly, M. Wintermark, and G. Zaharchuk, “Deep learning enables reduced

- gadolinium dose for contrast-enhanced brain MRI,” *J. Magn. Reson. Imaging*, vol. 48, no. 2, pp. 330–340, 2018.
- [123] S. Bollmann *et al.*, “DeepQSM - using deep learning to solve the dipole inversion for quantitative susceptibility mapping,” *Neuroimage*, vol. 195, no. March, pp. 373–383, 2019.
- [124] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 2242–2251, 2017.
- [125] D. Nie *et al.*, “Medical Image Synthesis with Deep Convolutional Adversarial Networks,” *IEEE Trans. Biomed. Eng.*, vol. 9294, no. c, pp. 1–11, 2018.
- [126] K. Hammernik *et al.*, “Learning a variational network for reconstruction of accelerated MRI data,” *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [127] G. Yang *et al.*, “DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction,” *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.
- [128] M. Mardani *et al.*, “Deep generative adversarial neural networks for compressive sensing MRI,” *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 167–179, 2019.
- [129] J. Zhu, G. Yang, and P. Lio, “Lesion focused super-resolution,” no. March, p. 56, 2019.
- [130] J. Schlemper *et al.*, “Stochastic deep compressive sensing for the reconstruction of diffusion tensor cardiac MRI,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11070 LNCS, pp. 295–303, 2018.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

DocuSigned by:
Yicheng Chen
50FF4C3FEB6D498... Author Signature

10/29/2019
Date