TUM

Advanced Reconstruction Techniques in Perfusion Magnetic Resonance Imaging

Cagdas Ulas

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Julien Gagneur

Prüfende der Dissertation:
1. Prof. Dr. Björn H. Menze
2. Prof. Dr. Matthias Günther

Die Dissertation wurde am 07.05.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 26.09.2021 angenommen.

Abstract

Magnetic resonance imaging (MRI) has been one of the most valuable diagnostic imaging techniques due to its capability of non-invasively producing highly detailed anatomical and physiological information of the human body. Perfusion MRI, as being one of the most crucial and promising dynamic magnetic resonance imaging (dMRI) modality, enables the quantification of perfusion-related parameters via exogenous or endogenous contrast agents (CAs). These techniques can be effectively used in assessing stroke, brain tumors, and treatment of patients with neurodegenerative diseases. Nevertheless, due to the vital necessity of tracking the rapid kinetics of a CA uptake following its administration, conventional perfusion MRI techniques suffer from limited spatial and temporal resolution, and often produce images with considerably low signal to noise ratio (SNR). To this end, perfusion MRI can significantly benefit from acceleration of the acquisition process.

In this thesis, we essentially focus on developing novel image and parameter reconstruction methods to accelerate perfusion MRI. To put it more concretely, we aim at reconstructing high quality images and perfusion parameter maps given the undersampled and/or corrupted perfusion magnetic resonance (MR) image-time series. To be able to achieve this, we substantially leverage recent methodological advances in image reconstruction and enhancement, which largely span compressed sensing, convex optimization, and deep learning fields. Specifically, we have demonstrated superior outcomes with scientific contributions in four areas of perfusion MRI: (1) compressed sensing based reconstruction of dynamic MRI, (2) robust reconstruction of perfusion MR image sequences using local and nonlocal spatio-temporal regularizers, (3) deep learning based direct reconstruction of pharmacokinetic parameters from undersampled dynamic contrast enhanced (DCE) MRI sequences, (4) denoising perfusion-weighted images of arterial spin labeled MRI based on deep fully convolutional neural networks (FCNs) together with residual learning strategy.

Zusammenfassung

Die Magnetresonanztomographie (MRT) ist eine der wertvollsten diagnostischen Bildgebungsverfahren aufgrund seiner Fähigkeit, nicht-invasiv sehr detaillierte anatomische und physiologische Informationen über den menschlichen Körper zu erhalten. Perfusion MRT, als eine der wichtigsten und vielversprechendsten dynamischen Magnetresonanz Resonanz-Imaging Modalität, ermöglicht die Quantifizierung von perfusionsbezogenen Parametern mittels exogener oder endogener Kontrastmittel. Diese Techniken können effektiv bei der Beurteilung von Schlaganfall, Hirntumoren und Behandlung von Patienten mit neurodegenerativen Erkrankungen eingesetzt werden. Dennoch, aufgrund der Notwendigkeit, die schnelle Kinetik der Kontrastmittelaufnahme nach der Verabreichung zu verfolgen, leiden konventionelle Perfusions-MRT-Techniken unter einer begrenzten räumlichen und zeitlichen Auflösung und erzeugen oft Bilder mit einem sehr niedrigem Signal-Rausch-Verhältnis. Aus diesem Grund kann die Perfusions-MRT erheblich von einer Beschleunigung des Aufnahmeprozesses profitieren.

In dieser Arbeit konzentrieren wir uns im Wesentlichen auf die Entwicklung neuartiger Bild- und Parameter Rekonstruktionsmethoden zur Beschleunigung der Perfusions-MRT. Um es konkreter zu formulieren, wir zielen darauf ab, qualitativ hochwertige Bilder und Perfusionsparameter-Karten zu rekonstruieren der unterabgetasteten und/oder korrumpierten Perfusions-Magnetresonanz (MR) Bildzeitreihen. Um dies erreichen zu können, nutzen wir im Wesentlichen die jüngsten methodische Fortschritte bei der Bildrekonstruktion und -verbesserung, die sich weitgehend auf die Bereiche Compressed Sensing, Convex Optimization und Deep Learning erstrecken. Speziell haben wir überlegene Ergebnisse mit wissenschaftlichen Beiträge in vier Bereichen der Perfusions-MRT: (1) Compressed Sensing-basierte Rekonstruktion von dynamischen MRT, (2) robuste Rekonstruktion von Perfusions-MR Bildsequenzen unter Verwendung lokaler und nicht-lokaler räumlich-zeitlicher Regularisierer, (3) Deep Learning basierte direkte Rekonstruktion von pharmakokinetischen Parame-

tern aus unterabgetasteten dynamischen kontrastverstärkten MRT-Sequenzen, (4) Entrauschen Perfusions-gewichteter Bilder von arterieller Spin-Label-MRT basierend auf tiefen fully Convolutional Neural Networks (FCNs) zusammen mit einer Residual-Learning-Strategie.

Acknowledgements

First of all, I would like to give my sincere gratitude to my doctoral advisor, Bjoern Menze, who offered me to pursue a doctoral degree in his research lab starting in 2015. I am grateful for his advice and continuous guidance, providing a flexible research environment and consistently showing his trust against me. I am glad that I was able to repay your trust with my work, and you have been a huge source of inspiration for me. I am also thankful to Marion Menzel for providing insightful feedback at the early stage of my thesis project and for reviewing this thesis.

Next, I would like to acknowledge the support of all my co-authors. Their inputs, in the form of data collection, sharing technical discussion and feedback, reviewing manuscripts, were always highly appreciated. I would like to give my special thanks to Christine Preibisch, who have been very supportive, provided stimulating discussions and fruithful feedback when reviewing the manuscripts, as well as to Stephan Kaczmarz who spent his time and effort for helping me to acquire MRI data and providing valuable feedback on manuscript submissions.

I would like to acknowledge the great reasearch collaboration with Center for Clinical Brain Sciences at the University of Edinburgh. Special thanks to Ian Marshall, Mike Davies and Michael Thrippleton for their great help and support during my 3-months secondment in Edinburgh. Thanks to our effective collaboration, we have published a few cool works together which constitutes the backbone of this thesis. I am very grateful for all your help.

Furthermore, I would like to thank my colleagues and friends at the image-based biomedical modeling (IBBM) group and GE Global Research, particularly Markus Rempfler, Dhritiman Das, Jana Lipkova, Anjany Sekuboyina, Pedro Gomez, Miguel Molina, Giles Tetteh, Yu Zhao, Xin Liu, Esther Alberts, Eduardo Coello, Bran Li, Ivan Ezhov and Carolin Pirkl. Thanks a lot for creating a wonderful working environment as well as for shared memories and activities. I will never forget the day when some of you joined my wedding celebration in Munich. I was so glad to have you joining me in one of the most

happiest day of my life!

Finally, to my wife Müzeyyen, for her unconditional support and immense help when integrating the new life together with her in a new country, and for her patience and understanding whenever I have been busy and loaded with my doctoral work. And, to my lively daughter Yare, who partly took a role on the late completion of this thesis due to her sweet naughtiness. You deserve everything in this world! This thesis is for you...
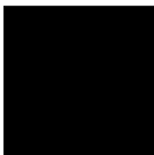
Contents

List of Figures

Acronyms

DWT . . . . discrete wavelet transform

EPI . . . . echo-planar imaging

FCN . . . . fully convolutional neural network

GAN . . . . generative adversarial network

GE . . . . gradient-echo

GM . . . . gray matter

LASSO . . . . least absolute shrinkage and selection operator

MCA . . . . middle cerebral artery

ML . . . . machine learning

MR . . . . magnetic resonance

MRI . . . . magnetic resonance imaging

MTT . . . . mean transit time

NLLS . . . . nonlinear least squares

NMR . . . . nuclear magnetic resonance

NN . . . . neural network

PET . . . . positron emission tomography

PI . . . . parallel imaging

PVE . . . . partial volume effect

PWI . . . . perfusion-weighted imaging

ReLU . . . . rectified-linear unit

RF . . . . random forest

RF . . . . radiofrequency

RIP . . . . restricted isometry property

ROI . . . . region of interest

SE . . . . spin-echo

SGD . . . . stochastic gradient descent

SNR . . . . signal to noise ratio

SVM . . . . support vector machine

TIC . . . . time intensity curve

TV . . . . total variation

WM . . . . white matter

1

Introduction

Magnetic resonance imaging (MRI) is a non-invasive and non-ionising imaging technique and generates cross-sectional images of the body. These images are subdivided in smaller units called *voxels*, which are the 3-dimensional version of a pixel. MRI scans apply a strong magnetic field to the subject, then measure the signal emitted by hydrogen nuclei in response to changes in this magnetic field resulting from nuclear magnetic resonance (NMR) [1, 2]. These emitted signals can – subject to certain assumptions – be used to reconstruct a per-voxel map of tissue properties. Images obtained with magnetic resonance (MR) are a reflection of the response of different tissues to a controlled stimulus [3]. This is conceptually different from other imaging techniques that measure intrinsic material properties, such as computed tomography (CT) scans which provide a measure of attenuation coefficient [4]. MRI is today an indispensable tool for medical diagnosis and research. Its success can largely be originated by the detailed soft tissue contrast that it can generate and by the great flexibility enabled due to its acquisition mechanism. A brief illustration of the diversity of MR images is shown in Fig. 1.1, ranging from structural to functional and dynamic imaging.

Dynamic magnetic resonance imaging (dMRI) is a very effective type of MR imaging technique that enables the visualization and analysis of anatomical and functional changes of internal body structures through time, resulting in a spatio-temporal signal [5]. In dMRI, the main objective is to characterise the anatomies in motion through quantitative analysis. Although MRI is a non-invasive, non-ionizing technology and provides an unmatched quality in soft tissue contrast, physical and physiological limitations on scanning speed makes this an inherently slow process [6, 7]. Beside that, specifically in dMRI, there is a trade-off between the spatial and temporal resolution. The main reason is that the physics of MRI scanners impose a sequential sampling and the physiology and motion of body organs limit the speed at which it can be performed,

1

(a) Structural brain   (b) Functional brain   (c) Cardiac cine   (d) Perfusion brain
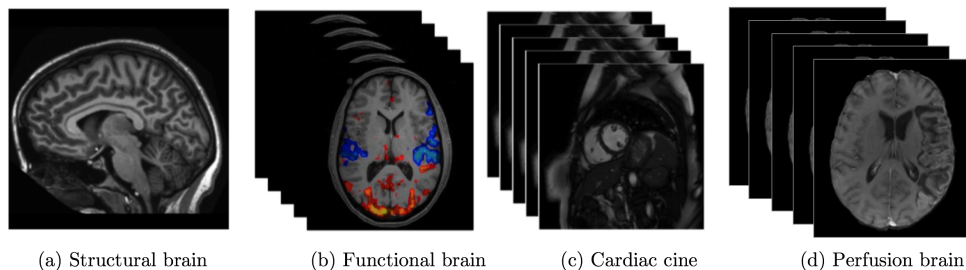
Figure 1.1: Examples of different MR image modalities. Structural MR for brain (a) is a static scan, functional brain MR (b), cardiac cine (c) and perfusion brain (d) are dynamic modalities. (Image adapted from [12])

hence satisfying the Nyquist criterion often entails prohibitive scan durations or unfavourable balancing of spatial and temporal resolution [8]. In dMRI a typical clinical examination can last up to 45 minutes of image acquisition. With increasing acquisition times, these techniques become more susceptible to motion-related artifacts, including voluntary (e.g. head movement) and involuntary (e.g. cardiac pulsations, breathing, brain pulsations) motion [9]. Furthermore, quantitative MRI techniques are generally constrained to sacrifice signal to noise ratio (SNR) or spatial resolution in favor of faster measurements to achieve feasible scan duration for clinical usability [10, 11].

One of the most popular dMRI methods is known as perfusion-weighted imaging (PWI), also called perfusion MRI. Medical diagnosis and research extensively exploit PWI techniques to estimate the blood flow and blood volume through examination of the spatio-temporal changes of the signal intensities following the injection of a blood bolus via exogenous paramagnetic tracers [13]. In neuroimaging, these techniques have become widespread clinical tools in the diagnosis of stroke [14, 15] – for the assessment of the tissue at risk –, and the treatment of patients with cerebrovascular diseases [16, 17]. One of the major obstacles in the clinical use of perfusion imaging is the immense need to track the rapid kinetics of contrast agent (tracer) uptake for precise perfusion quantification [18, 19]. Moreover, the short scanning time available for each frame often results in limited spatial and temporal resolution, and poor SNR images. In order to improve the spatial or temporal resolution, one widely used approach is to accelerate the acquisition of each frame through the undersampling of k-space by acquiring only a subset of k-space lines [20]. Acquiring fewer k-space samples than those dictated by the Nyquist criterion accelerates the process considerably, but exhibits aliasing artifacts in image space [8]. In the context of perfusion MRI, accurate reconstruction of the

complete temporal perfusion signal with its peak and the high dynamic range becomes an even more challenging task. Therefore, dealiasing of MR images and denoising of temporal perfusion signals from undersampled acquisitions is an important milestone for the robust acceleration of PWI. This topic has long been an active research objective of the field [21, 22, 23, 24], and is the main goal of this thesis.

The rest of this chapter is structured as follows: In Section 1.1, we provide a brief introduction to MRI principles and give a review of existing approaches for acceleration of MRI in Section 1.2. Section 1.3 presents a brief description of available acquisition techniques for perfusion MRI. Subsequently, we summarize our main contributions in Section 1.4 and outline the remaining sections of the thesis in Section 1.5.

1.1 Magnetic resonance imaging principles

MRI is a non-invasive and non-ionising medical imaging modality largely popularised by an unmatched soft tissue contrast. Images obtained with MR are a reflection of the response of different tissues to a controlled stimulus. Essentially, the generation of MR images can be summarized with the following three successive steps: (i) The magnetic pulses used to create a transverse magnetisation should be designed, (ii) during the acquisition stage the signal emitted is sampled by a receiver coil in the scanner, (iii) the captured signal is modeled and reconstructed to generate an image.

In the following sections we provide a summary of the NMR phenomenon that is at the core of MRI, and explain how it can be utilized to create images of the body. The presentation of MRI given in this part mainly follows the classical description available in [25, 26].

1.1.1 Nuclear magnetic resonance

The fundamental physical phenomenon enabling MRI is known as NMR, which refers to the exchange of energy between atoms and a magnetic field rotating at a resonant frequency [1]. To introduce this concept more concretely, we should highlight a sequence of three atomic phases composing the NMR experiment, which are *spin polarisation*, *excitation and relaxation*, and lastly *Bloch equation*, formally summarising these physical events.

Spin polarisation

MRI operates by measuring the net magnetisation of hydrogen atoms ^1H abundant in a biological specimen. ^1H – an atom with an odd number of protons – possesses angular momentum called *spin*. Under the influence of a homogeneous magnetic field $\mathbf{B}_0$, two phenomena occur to the spins of hydrogen atoms: (i) The spins align with $\mathbf{B}_0$ in parallel or anti-parallel direction, which results in a net magnetisation $\mathbf{M}$ of the body in the direction of $\mathbf{B}_0$, (ii) the polarised spins exhibit resonance at Larmor frequency w,

$$w = \gamma B \tag{1.1}$$

where γ is the gyromagnetic ratio and B is the surrounding magnetic field strength. γ is nucleus dependent and ^1H atomic nuclei has $\gamma = 2.68 \times 10^8$ rad/s/T. In MRI, we represent the main static magnetic field as $\mathbf{B}_0 = (0, 0, B)$, which by convention is pointing towards z-direction (longitudinal axis). At equilibrium state, the net magnetisation is denoted $\mathbf{M}_0 = (0, 0, M_0)$. In a typical high-field MRI device, the main homogeneous magnetic field $\mathbf{B}_0$ is generated by a cylindrical superconducting magnet.

Excitation and relaxation

The Larmor frequency (w) is the resonance frequency at which energy from an external magnetic induction field can be absorbed by the nuclear spin system. In the NMR domain, this field is produced by a radiofrequency (RF) pulse $\mathbf{B}_1$ which is perpendicular to the static B_0. If enough energy is deposited in the system, the spins will come into phase and some low energy spins will jump to a high energy state, macroscopically tilting the net magnetisation onto the x-y plane [25]. As a result, longitudinal magnetisation M_z decreases and the net magnetisation $\mathbf{M}$ includes a net transverse component $\mathbf{M}_{xy}$ spinning at the Larmor frequency as illustrated in Fig. 1.2(c), this step is called *excitation*.

After excitation, net magnetisation returns to the equilibrium state through a process called *relaxation*. In relaxation, RF energy absorbed during excitation by some spins to adopt a high energy state is liberated as they return back to a low energy state, leading to a gradual decay of transverse magnetisation $\mathbf{M}_{xy}$ and the recovery of the original longitudinal magnetisation M_z as depicted in Fig. 1.2(d). The liberated energy constitutes the emitted signal which can be sampled and used for generating images [27].

Two effects govern the transition from the excitation state to equilibrium over a period of time τ. First, longitudinal relaxation is caused by energy exchange between the spins and the surrounding lattice, falling back to thermal

(a) Thermal equilibrium

(b) Spin polarisation

(c) Excitation

(d) Relaxation
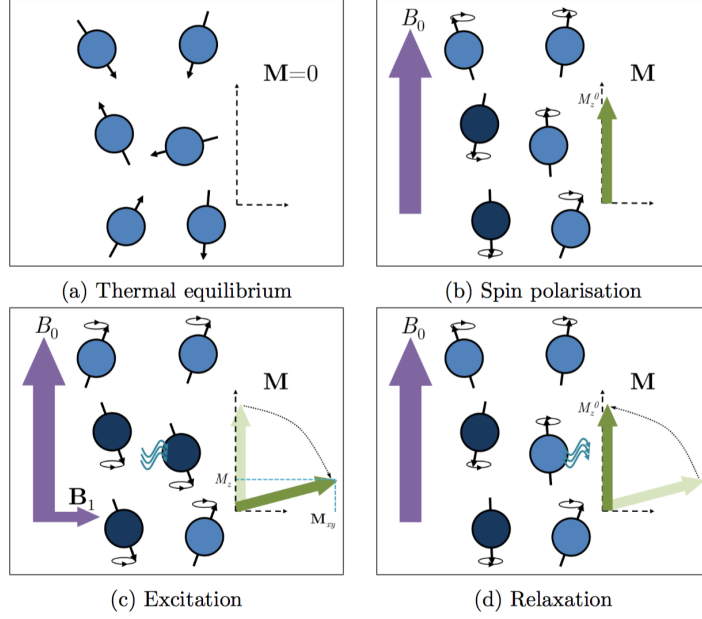
Figure 1.2: Different atomic nuclei phases of NMR. Low-energy state nuclei are represented in light blue and high-energy state nuclei are represented in dark blue. After spin polarisation with the B_0 field, the system of spins can be excited resulting in the transverse magnetisation $\mathbf{M}_{xy}$. Following to the excitation, the bulk magnetisation relaxes and returns a state where it again aligns with the B_0 field. (Image courtersy of [12])

equilibrium. It is described by an exponential curve characterised by the spin-lattice relaxation time T_1,

$$M_z(t+\tau) = M_z^0 - \left(M_Z^0 - M_z(t)\right) e^{-\frac{\tau}{T_1}}, \qquad (1.2)$$

where T_1 is a tissue specific constant which is usually longer at higher magnetic field strengths. Similarly, transverse relaxation is obtained from spins dephasing in the absence of the RF signal formulated by,

$$\mathbf{M}_{xy}(t+\tau) = \mathbf{M}_{xy}(t) e^{-\frac{\tau}{T_2}}. \qquad (1.3)$$

This decay is also governed by an exponential curve with the spin-spin relaxation time T_2, which is also tissue dependent. One should note that T_1 and T_2 are tissue specific, given that this makes different tissues react differently to the

same stimulus, becoming a mechanism not only to resolve them but also to generate images with various contrast patterns [25].

Bloch equation

The temporal behaviour of the net magnetisation of a nuclear spin system was formalised in 1946 by Felix Bloch [28], known as the *Bloch equation*,

$$\frac{d\mathbf{M}(t)}{dt} = (\mathbf{M}(t) \times \gamma\mathbf{B}_1(t)) - \frac{M_z(t) - M_z^0}{T_1} - \frac{\mathbf{M}_{xy}(t)}{T_2}. \tag{1.4}$$

This equation synthesises the excitation reaction of a net magnetisation $\mathbf{M}$ initially in equilibrium to an external electromagnetic radiation $\mathbf{B}_1(t)$ and the subsequent relaxation phase governed by tissue specific T_1 and T_2 constants.

1.1.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is a technique which exploits the NMR signal produced by the transverse magnetisation of atomic nuclei to generate images. The emitted signal is a rotating entity dependent on space and time, and hence can be described by a complex-valued variable,

$$m(\mathbf{r}, t) = m(\mathbf{r})e^{i\phi(\mathbf{r}, t)}, \tag{1.5}$$

where $\mathbf{r}$ refers to (x, y, z) space and $m(\mathbf{r})$ is the signal of interest [25]. Essential components of MR imaging are *gradient encoding* which enables the spatial characterisation of the received signal, and the notion of *k-space* which relates the sampled signal to the magnetisation image.

Gradient encoding

The RF pulse induces all nuclei in the body precessing at the resonant frequency to emit a signal, regardless of their spatial location within the body. A typical MRI scanner contains encoding gradient coils which are capable of generating constant gradients that are superposed to the B_0 field creating spatially linear varying precessing frequencies [3]. Given the Eq. 1.1, a gradient $\mathbf{G}(t)$ will cause a frequency deviation from the Larmor frequency equivalent to,

$$\Delta f(\mathbf{r}, t) = \frac{\gamma}{2\pi}\mathbf{G}(t) \cdot \mathbf{r}. \tag{1.6}$$

where $\mathbf{G}(t) = [G_x(t), G_y(t), G_z(t)]$ and linearly dependent on the spatial position. There are generally three mechanisms for spatial localisation using encoding gradients: selective excitation, phase encoding and frequency encoding.

Selective excitation. The first magnetic gradient G_z is used along the z direction to select the slice to image by limiting the RF excitation to the slice of interest. This selection is performed by imposing a dependance between the static magnetic field and the spatial localization of the MR signal in the z direction,

$$B(z) = B_0 + G_z z. \tag{1.7}$$

Applying a RF pulse B_1 with frequency f_0 and bandwidth $2\Delta f$ will only excite the spins in the selected slice $[z_0 - \Delta z, z_0 + \Delta z]$. The gradient strength and timing thus controls the localization and size of the slice, e.g. a strong gradient or narrow bandwidth will result in a thinner slice.

Phase encoding. A second magnetic field gradient G_y is then applied along the y direction after the RF pulse excitation, and hence the precession frequency of the spins in a selected slice is affected for a short term. As a result, spins at various spatial position will precess at different frequencies. Once the magnetic gradient G_y is switched off, the spins in the selected slice precess at the original frequency but the phase $\Delta\phi$ changes linearly along the y axis. This step is referred to as the *phase encoding* and can be mathematically formulated by the following equation,

$$\Delta\phi(y) = e^{-i\gamma G_y \tau_y y}, \tag{1.8}$$

where γ is the gyromagnetic ratio, and τ_y denotes the duration of the gradient G_y along the y dimension.

Frequency encoding. Lastly, a third magnetic field gradient G_x is applied along the x direction. This gradient is applied directly after the slice selection gradient and RF pulse excitation, and enables the spins to precess at different frequencies, linearly depending on their position along the x axis. The use of this gradient to resolve the last encoding dimension is referred to as the *frequency encoding* step. This step can be formulated as,

$$\Delta\phi(x) = e^{-i\gamma G_x \tau_x x}, \tag{1.9}$$

where the spins phase change at time τ_x along x direction.

Signal equation

The phase component of the spatially encoded signal $m(\mathbf{r}, t)$ can be included in the description of the emitted signal, and after neglecting the carrier Larmor frequency, this turns into

$$m(\mathbf{r}, t) = m(\mathbf{r})e^{i\Delta\phi(\mathbf{r}, t)} = m(\mathbf{r})e^{i2\pi\mathbf{k}(t)\cdot\mathbf{r}}, \tag{1.10}$$

where $\mathbf{k}(t) = \frac{\gamma}{2\pi}\int_0^t \mathbf{G}_\tau d\tau$, and we use $\Delta\phi(\mathbf{r},t) = 2\pi\mathbf{k}(t)\cdot\mathbf{r}$ to generalize encoding in the y and x dimensions, as expressed in Eq. 1.8 and Eq. 1.9.

Given the above formulation per spatial location, the signal received is the aggregate response from nuclei at all spatial locations which have been subject to excitation [26]. Assuming that a single body coil is used in MRI acquisition, the received signal simplifies to

$$\hat{m}(t) = \int_{-\infty}^{+\infty} m(\mathbf{r},t)d\mathbf{r} = \int_{-\infty}^{+\infty} m(\mathbf{r})e^{-i2\pi\mathbf{k}(t)\cdot\mathbf{r}}. \tag{1.11}$$

We should remark that above expression (Eq. 1.11) does not take into account a few factors such as the decay of transverse magnetisation, B_0 inhomogeneity and changes of the object in time.

K-space

The direct mapping observed between precessing frequencies and space locations through encoding gradients is identical to a *Fourier* relationship. Therefore, samples can be interpreted as acquired in the spatial-frequency domain, also called *k-space*, where each sample is a sum of magnetisation from atoms in the entire volume weighted by the Fourier kernel $e^{-i2\pi\mathbf{k}(t)\cdot\mathbf{r}}$ [25]. The coordinates of a sample in *k-space* are given as $\mathbf{k}(t) = (k_x(t), k_y(t), k_z(t))$ and directly controlled through the gradients $\mathbf{G}(t)$. The use of $G_x(t)$ and $G_y(t)$ gradients enables traversing the x-y plane of k-space. Provided that gradient pulse durations relate to k-space locations, the received signal can be rewritten as

$$m(\mathbf{k}) = \int_{-\infty}^{+\infty} m(\mathbf{r})e^{-i2\pi\mathbf{k}(t)\cdot\mathbf{r}}. \tag{1.12}$$

The main goal of the conventional MRI process is to produce the image $m(\mathbf{r})$ by acquiring a sufficient set of signals $\{m(\mathbf{k})\}_{k\in\mathbb{R}^3}$ in k-space and perform inverse Fourier transform.

An appropriate coordination of gradients leads to a scan protocol traversing locations of k-space and acquiring various samples. This protocol, commonly known as a *trajectory*, plays an important role on the appearance of the reconstructed image [26]. One of the widely used trajectory type is the Cartesian trajectory, which traverses a rectangular k-space acquiring equidistant samples as shown in Fig. 1.3. Cartesian sampling poses a very simple reconstruction stage, which simply consists of a discrete Fourier transform (DFT), but has the limitation of requiring many excitation pulses to traverse a full k-space.
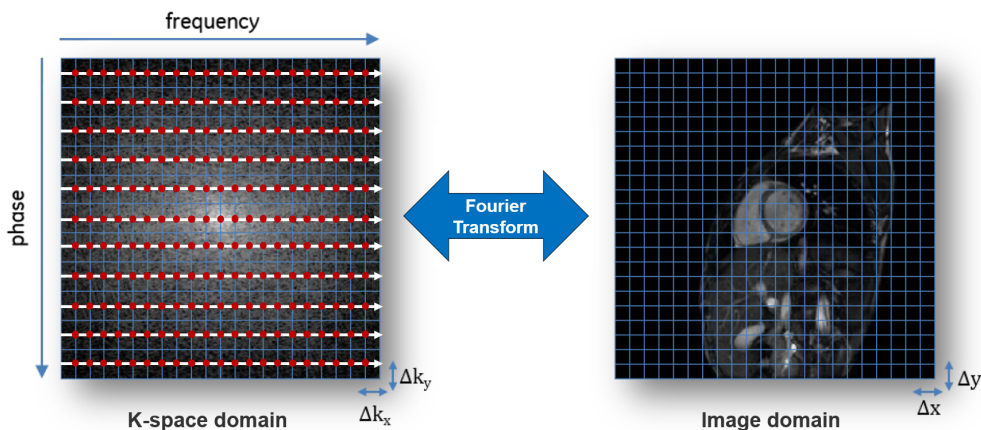
Figure 1.3: Data in k-space collected on a Cartesian grid (left) for cardiac MR imaging. The Fourier transform is used to transform the k-space samples to an image (right). The spacing between adjacent samples in k-space (Δk_x and Δk_y) is inversely proportional to the resolution in image space (Δx and Δy). (Image adapted from [29])

1.2 Acceleration of MRI acquisitions

Magnetic resonance imaging (MRI) is often very time-consuming process due to the fact that k-space samples need to be collected sequentially. The acceleration of MRI acquisition is desirable for a number of reasons. One of the main potential benefits of faster scans is an increased comfort for patients. Ideally the patient needs to remain static within the scan duration, but the patients who are severely ill or especially at a young age may not tolerate long scanning procedures. Faster scans can also have a great impact on the quality of images specifically for dynamic MRI because they can reduce the image distortion introduced by motion artefacts [12]. Moreover, accelerating MR acquisition may bring considerable benefits on decreasing the high costs of MR scanning sessions and enabling to scan higher number of patients while eliminating long waiting lists.

Until recently many solutions have been proposed for accelerating MRI acquisitions in the field. One of the earliest solutions was to develop pulse sequence designs that can traverse k-space with fewer RF excitations through non-cartesian sampling trajectories including radial [30], spiral [31] and variable density [32]. These trajectories naturally favor denser sampling for low frequencies and sparser at higher frequencies, and shown to be robust against motion. Moreover, as it was proposed by fast spin echo strategies [33], appropriate

(a) Full   (b) Low-frequency   (c) Regular   (d) Random
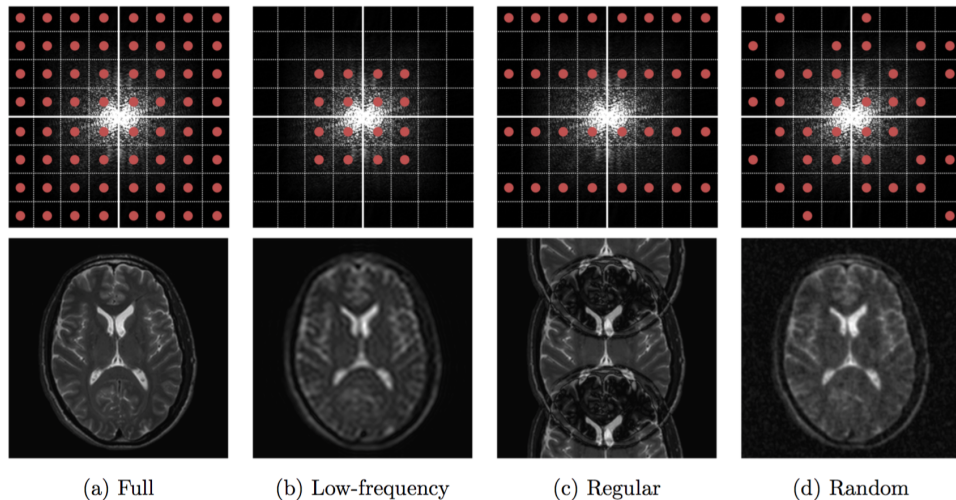
Figure 1.4: Various undersampling strategies in k-space (top) and the aliasing forms they result in image domain (bottom). (Image courtesy of [12])

design of pulse sequences may allow to collect multiple separately encoded readouts per RF pulse.

A different approach on acceleration of MRI was considered by the methods that assume a particular trajectory but undersample k-space by skipping the acquisition of several samples at selected locations [34, 20]. Given a binary mask $\mathbf{M} \in \mathbb{R}^{N \times N}$, with entries $M(x, y) = 1$ if k-space sample at (x, y) location has been acquired, the undersampled k-space data is formulated as

$$\hat{\mathbf{m}} = \mathbf{MFm}, \tag{1.13}$$

where $\mathbf{F}$ denotes the Fourier operator. The undersampling of k-space intended to be sampled at the Nyquist rate leads to aliasing in image space, which can be observed in different forms as shown in Fig. 1.4. Sampling only low frequencies results in worse image resolution, regular undersampling creates coherent aliasing and random sampling leads to random aliasing.

In order to tackle the problem arised by k-space undersampling, researchers initially proposed to create sampling redundancy through additional hardware. In particular, the use of multiple receiver coils linearly increases the number of samples for a particular image which theoretically allows maintaining a sufficient amount of information even after k-space undersampling. This method is known as parallel imaging (PI) [35]. In PI, data is acquired using N_c receiver coils, where each receiver coil is more sensitive to signals generated

in the proximity of that coil. Mathematically, the coil sensitivity maps can be expressed by spatially varying weights $\mathbf{S}_i \in \mathbb{C}^{N \times N}$ for the ith coil, and the image produced by the coil can be expressed as $\mathbf{S}_i\mathbf{m}$. Then, for each receiver coil the measurement k-space data satisfies the following formulation,

$$\hat{\mathbf{m}}_i = \mathbf{E}_i\mathbf{m} + \epsilon, \qquad (1.14)$$

where $\mathbf{E}_i$ is the encoding operator comprising the ith coil sensitivity weighting and the Fourier sampling, $\mathbf{m}$ is the image to recover, $\hat{\mathbf{m}}$ is the acquired k-space data and ϵ is the measurement error. The image $\mathbf{m}$ is typically reconstructed by solving the following large linear system

$$\hat{m}_k = \sum_{j=1}^{N_c} E_{k,j} m_j, \quad \forall k \in \mathbb{K}. \qquad (1.15)$$

Several PI techniques have been proposed for the reconstruction of $\mathbf{m}$ given $\hat{\mathbf{m}}$. These methods can be mainly subdivided into those that applied on the aliased image, i.e. SENSE-type reconstruction [36] and those that reconstruct the missing k-space data, i.e. GRAPPA-type reconstruction [37]. These two PI methods are routinely employed on clinical MRI scanners today.

An effective way for accelerating MRI acquisitions which have appeared more recently is the use of *inherent data redundancy*. The intuition behind this approach is that the more information about the object to be imaged is known or assumed *a priori*, the fewer samples are needed for its reconstruction. From this perspective, compressed sensing (CS) [20, 38] based methods have received a lot of attention by the research community in the last decade. The fundamental observation used for CS reconstructions is that images or sequences of images can be sparsely represented in the image or a transform domain. This effectively enables that missing k-space data can be recovered with the *a priori* assumption that the reconstructed image is compressible. Early work on CS-MRI primarily focused on applying predefined universal sparsifying transforms, such as the DFT, discrete cosine transform (DCT), total variation (TV), or discrete wavelet transform (DWT), and developing efficient numerical algorithms to solve nonlinear optimization problems. Later, data-driven sparsifying transforms (i.e. dictionary learning) have gained much attention in CS-MRI due to their ability to express local features of reconstructed images more accurately compared to pre-defined universal transforms [39, 40].

Dynamic MR sequences usually provide redundant information in both spatial and temporal domains, which allows the reduction of acquisition time by applying sparsifying transforms in both domains. One of the first methods

11

proposed under the CS-MRI framework for dynamic imaging seek for sparsity in the x-f support, which exploits the Fourier transform of the image data along the temporal dimension. An example of successful methods enforcing x-f support sparsity is k-t FOCUSS [41]. The earlier method k-t SPARSE [42, 34] additionally sparsifies data across space using a wavelet transform, which is known to provide approximately sparse representations for the vast majority of natural signals [43, 44]. More recently, researchers have proposed sophisticated CS-based reconstruction methods that exploit both spatial and temporal redundancies of the entire dataset, such as spatio-temporal total variation [45, 46], dictionary learning [40, 47, 48], patch-based regularization [49], low-rank approximation and sparsity [45, 50, 51], structured sparsity [52, 53], and manifold learning [54, 55].

1.3 Perfusion MRI

In MRI domain there are two major approaches to measure the tissue perfusion. The first type of methods is mainly based on application of an exogenous, intravascular, nondiffusible contrast agent – usually a gadolinium-based CA – that emphasizes either the susceptibility effects of the contrast agent on the signal echo, known as dynamic susceptibility contrast (DSC) MR perfusion or the relaxivity effects of the contrast agent on the signal echo, namely dynamic contrast enhanced (DCE) MR perfusion. The second type of perfusion methods involve application of an endogenous contrast agent using magnetically labeled arterial blood water as a diffusible flow tracer, and this technique is known as arterial spin labelling (ASL) MR perfusion. Below we provide a brief description of each of these perfusion techniques, focusing on the details related to acquisition protocols and quantification models used in each of this technique.

1.3.1 Dynamic susceptibility contrast

Dynamic susceptibility contrast (DSC) MRI is one of the well established imaging techniques that can be used to measure blood flow and other related hemodynamic parameters. It involves the intravenous injection of a paramagnetic MR contrast agent (CA) and relies on measuring and modeling the induced changes in T_2 and $T_2^\star$ relaxation, introduced by the induced susceptibility effect [13]. These changes are amplified when the CA remains compartmentalized. Due to this amplified effect related to compartmentalization, DSC-MRI is an acquisition technique that has primarily been used to assess cerebral perfusion: the blood vessels in the brain are surrounded by the so-called blood-brain barrier (BBB), which, when intact, makes the

gadolinium-based CAs commonly used for DSC-MRI behave effectively as an intravascular contrast bolus [56].

Given that DSC-MRI exploits the transient changes in T_2 and $T_2^\star$, the most commonly used MRI sequences in clinical investigations are based on a gradient-echo (GE) or a spin-echo (SE) echo-planar imaging (EPI) acquisition. While the SE-based sequence has low sensitivity to large vessels and, thus, is more suited for measuring microvascular perfusion, i.e. , at the capillary level, the much larger susceptibility effect in GE-based sequences makes them the sequence of choice in practice. Nevertheless, this comes at the expense of a significant macrovascular artifacts, where cerebral blood flow (CBF) is greatly overestimated in voxels containing partial volume effect (PVE) with large arteries and veins. Some of the major advantages of DSC-MRI lie in its simple acquisition method, its short acquisition time (approximately 1-2 mins), relative high SNR compared with other perfusion imaging methods (e.g. CT perfusion, arterial spin labelling (ASL), positron emission tomography (PET)), and high variety of hemodynamic information or parameters that can be obtained from the same dataset [57].

In DSC, a tracer kinetic model is commonly applied for perfusion quantification. The major perfusion-related parameters CBF, cerebral blood volume (CBV), and mean transit time (MTT) are estimated from tracer kinetic models based on the use of the central volume theorem [58]. Using either SE or GE pulse sequences, the MRI signal intensity following contrast agent administration can be determined from the changes in the transverse relaxation rates as,

$$S_t(t) = S_t(0) \cdot e^{-T_{\mathrm{E}} \cdot \Delta R_2^\star(t)}, \tag{1.16}$$

where T_{E} is the echo time of the acquisition, $S_t(t)$ is the post-contrast injection signal intensity, $S_t(0)$ is the pre-contrast signal intensity. The signal time-intensity curves of each voxel can be directly used to estimate the CA concentration $C_t(t)$ from the change in the transverse relaxation rate ($\Delta R_2^\star$) as follows

$$C_t(t) = -\frac{1}{r_2 \cdot T_{\mathrm{E}}} \cdot \ln\left(\frac{S_t(t)}{S_t(0)}\right). \tag{1.17}$$

The hemodynamic parameters CBV, CBF and MTT can be quantified as

$$\mathrm{CBV} = \frac{\int_{-\infty}^{+\infty} C_t(\tau)d\tau}{\int_{-\infty}^{+\infty} C_a(\tau)d\tau}, \tag{1.18}$$

$$C_t(t) = \mathrm{CBF} \cdot C_a(t) \otimes R(t) = \mathrm{CBF} \cdot \int_{-\infty}^{t} C_a(\tau) \cdot R(t-\tau)d\tau, \tag{1.19}$$

13

$$\text{MTT}(t) = \frac{\int_{-\infty}^{+\infty} \tau \cdot C_t(\tau) d\tau}{\int_{-\infty}^{+\infty} C_t(\tau) d\tau}, \tag{1.20}$$

where $C_a(t)$ is the arterial concentration of the CA also called arterial input function (AIF), $R(t)$ is the tissue residue function which describes the fraction of the CA remaining in the tissue at time t following an instantaneous bolus administration. By definition, at time $t = 0$, all the contrast agent (CA) stays in the tissue, i.e. $R(0) = 1$. For a nondiffusible tracer, all the tracer leaves the tissue after a sufficiently long time, $R(\infty) = 0$. According to Eq. (1.18-1.20), CBV is obtained as the ratio of the areas under the tissue concentration–time curve and the arterial concentration–time curve. CBF can be estimated via a deconvolution of the tissue concentration–time curve with the AIF. The deconvolution process can be complex and requires an accurate measurement of the AIF. Finally, MTT can be derived from the first moment of the tissue concentration–time curve.

In practice, a single estimate of the AIF for the whole brain (the so-called global AIF) is obtained from measuring the changes in $R_2^\star$ in and around a medium-size artery, such as the M1 or M2 segment of the middle cerebral artery (MCA) at 1.5T and the M2 or M3 segments at 3T [19]. The measurement of the AIF can be essentially subject to a number of major error sources for DSC quantification, including PVE, bolus delay and dispersion, nonlinearities, truncation effect and voxel-shift signal artifacts [18, 19].

1.3.2 Dynamic contrast enhanced

DCE MRI perfusion technique typically consists of intravenous injection of a CA followed by the repeated acquisition of T_1-weighted images, providing measurements of signal enhancement as a function of time [16]. DCE-MRI combined with tracer kinetic modelling is widely used for assessing permeability and perfusion in brain tumours [59, 60] and several neurological disorders that lead to disruption and breakdown of blood-brain barrier (BBB) [61, 62]. In this technique, a set of baseline images are acquired without contrast enhancement, followed by a series of images acquired over time – usually over a few minutes – during and after the arrival of the CA in the tissue of interest. The acquired signal is then used to generate so-called time intensity curve (TIC) for the tissue which contains an induced variation correlating to the local CA concentration in the tissue upon to the arrival of paramagnetic particle. Through the quantitative analysis on these curves, several physiological properties that are

related to the microvascular blood flow, such as vessel permeability, tissue volume fractions and vessel surface area product, can be derived [63].

Following measurements are typically required in DCE-MRI to be able to determine relevant physiological parameters [64]:

- Measurement of a map of pre-contrast (before gadolinium injection) T1 values ($T_1(0)$ map) for the calculation of CA concentrations,

- Acquisition of T_1-weighted image prior and following CA introduction at a reasonably high temporal resolution to be able characterize the underlying kinetics of the CA entry and exit out of the tissue,

- Determination of the AIF which is usually based on the estimation of a smooth CA concentration-time curve (averaged over voxels in a region of interest (ROI)) in the blood plasma of a feeding artery.

The quantitative analysis are mainly based on fitting an appropriate tracer kinetic modeling which provides a link between the tissue CA concentrations and physiological or so-called *pharmacokinetic parameters*. The fitting is commonly done voxelwise manner and performed using nonlinear least squares (NLLS) methods. Pharmacokinetic – or tracer kinetic – modeling was first introduced for the analysis of DCE-MRI in the early 1990s by Brix et al. [65] and Tofts et al. [66], followed by a consensus paper on the notations [67]. Since then, several more complex compartment models have been proposed. We refer the interested readers to [68, 69] for a more detailed review of different pharmacokinetic models available in literature.

1.3.3 Arterial spin labeling

Arterial spin labelling (ASL) is a powerful MRI approach to measure the CBF. ASL is completely non-invasive and uses magnetically labeled blood water as an endogenous tracer [70, 71]. To this end, it is highly suitable for repetitive follow-ups and allows for absolute quantification of CBF. Promising research results have shown how ASL can be used in the diagnosis and analysis of stroke, tumours, dementia and paediatric medicine, in addition to many other pathologies and areas [72, 73]. Perfusion imaging techniques have been an active area of research for decades, but there are many practical barriers to the widespread use of perfusion imaging. However, due to the fact that ASL does not require the injection of an exogenous CA, it has the potential to cope with these limitations and possibly even to bring perfusion imaging to common clinical practice, as well as expanding its scope in research.

Despite all these benefits, ASL imaging suffers from a few limitations: ASL has an inherently low SNR with typically small amount of blood flowing into the brain over the timescale on which ASL images are acquired [74, 75]. Therefore, typically 20–50 image-pairs (so-called repetitions) with low in-plane resolutions (3–4 mm) and through plane resolutions (4–8 mm) need to be acquired to achieve sufficient SNR [76]. Furthermore, ASL is highly prone to corruption by subject motion, scanner coil instability, boundary artifacts, and PVE which hinder obtaining accurate signals that reflect the underlying perfusion, hence leading to significant quantification errors especially in the gray matter (GM) and white matter (WM) [77, 78].

The basic principles of image acquisition in ASL can be summarised as follows: A magnetically labelled bolus is created upstream to the tissue of interest by inverting the magnetization inside the bolus. After a delay time during which the blood can travel to the tissue and the labeled molecules transit from the capillary bed to the pool of tissue water, a so-called *label image* is acquired. Since only 3-5% of the tissue water stems from blood water and relaxation further decreases the concentration of labelled blood, the perfusion signal makes up only 1-2% of the overall signal. Therefore, an additional *control image* without any prior labeling is acquired. Ideally, the signal of the static tissue is identical in the control and label images and the difference between them arises from the perfusion process. Subtracting the label from the control image will in theory reveal signal originating from blood-flow and perfusion, hence the difference signals are directly proportional to the CBF.

An absolute quantitative perfusion maps can be obtained using the General Kinetic Model developed by Buxton et al. [79]. This model is a type of single-compartment model which assumes that the labeled water diffuses freely – no outflow – throughout the tissue voxel. Assuming that ΔM denotes the subtraction between labeled and control images, also known as *perfusion-weighted image*, the relationship between ΔM and the CBF depends on the proton density of the tissue, the relaxation time T_1 of the labeled tissue and blood, as well as their difference. It also depends on the transit time between the labeling area and the volume of interest. According to the single-compartment Buxton model, the difference in magnetization between labeled and control images is described by the following expression,

$$\Delta M = 2 \cdot M_a(0) \cdot \text{CBF} \cdot \int_0^t c(\tau) r(t-\tau) m(t-\tau) d\tau, \qquad (1.21)$$

where $M_a(0)$ is the magnetization at equilibrium in an arterial blood voxel, $c(t)$ is the fractionated arterial input function, $r(t-\tau)$ is the output of the labeled

protons from the voxel and $m(t - \tau)$ is the decaying effect of longitudinal relaxation.

In addition to the CBF, other perfusion parameters can be estimated with ASL, provided that specific methods of acquisition or different quantification models are used. It is possible to quantify the arterial transit time (ATT) and the volume of arterial blood with a multi-TI sequence [80]. However, the main disadvantage of using multiple inversion times in ASL is that it can considerably increase the scan duration to acquire sufficient data. Moreover, the model fit required to estimate the perfusion parameters become more complex than simple division, as in the single-TI case. Simultaneous inference of multiple model parameters can be efficiently solved via nonlinear least squares (NLLS) [81] and more advanced Bayesian inference methods [82].

For the interested readers we refer the consensus paper [75] which provides a detailed explanation of recommended implementations of ASL for clinical applications. For more detailed description of various compartmental models for perfusion quantification in ASL we refer to [83].

1.4 Summary of Contributions

This thesis essentially spans the context of perfusion MR image and parameter reconstruction from subsampled, corrupted and noisy acquisitions. We have targeted to tackle the aforementioned challenges associated with perfusion MRI acquisition and parameter quantification. This encapsulates all three existing perfusion MRI techniques and we have explored advanced compressed sensing (CS) and machine learning (ML)/deep learning (DL) based approaches to solve these problems.

In the following, we provide a brief summary of the contributions of this thesis covering mainly four different topics.

Dynamic MRI reconstruction by enforcing local and global regularities

In this work, we introduce a new spatio-temporal reconstruction approach for the fast reconstruction of dMRI data from undersampled k-space measurements. We have mainly aimed at addressing the problem of accurately reconstructing the fully-sampled image series and removing aliasing artifacts. Inspired by the dynamic total variation (dTV) sparsity inducing norm which was proposed for real-time reconstruction, we propose a novel reconstruction scheme that iteratively enforces not only the local (spatial) regularity in every single frame but also the global (spatio-temporal) regularity of a full sequence. To this

end, we introduce a reconstruction model which jointly uses dTV sparsity and nuclear norm penalties, exploiting both the sparsity of inter-frame differences and the low-rank structure of the dynamic MR sequences in the full spatio-temporal space. We validate our approach on both *in-vivo* 3D cardiac MRI and DSC-MRI brain perfusion image series in comparison to state-of-the-art methods. This work was published as a peer-reviewed conference paper [84] which will be provided in Chapter 3.

Accelerated reconstruction of perfusion MRI using local and nonlocal constraints

This work specifically tackles the problem of recovering perfusion MR image sequences from undersampled k-space data. One of the major obstacles in the clinical use of perfusion imaging is the need to track the rapid kinetics of contrast agent (tracer) uptake for accurate perfusion quantification. In perfusion imaging, undersampling the k-space by acquiring a small subset of k-space lines not only leads to aliasing artifacts in image space but also significantly corrupts and degrades the true temporal perfusion signal with its peak and high dynamics. To circumvent these limitations, we propose a novel reconstruction model which integrates two fundamentally different constraints: (i) we enforce a pixel-wise local sparsity constraint on the temporal differences that limits the overall dynamic of the perfusion time series, (ii) we enforce a patch-wise similarity constraint on the spatio-temporal neighborhoods of whole MR sequence, which provides smooth spatial image regions with less temporal blurring especially when there is significant inter-frame motion and noise. Results have been demonstrated on brain perfusion datasets as well as on a publicly available dataset of *in-vivo* breath-hold cardiac perfusion. This work was published as a peer-reviewed workshop paper [85] and we present it in Chapter 4. An extension of this work, which additionally addresses the perfusion parameter reconstruction problem from the undersampled DSC and DCE MRI series, was made available as an arXiv manuscript [86] and we provide it in Appendix A.

Direct reconstruction of pharmacokinetic parameters of DCE MRI using deep learning

This study presents the first work in the field which leverages the machine learning algorithms – specifically deep learning – to directly estimate pharmacokinetic (PK) parameters from undersampled DCE-MRI sequences. Through the direct estimation we eliminate the image reconstruction stage which is

conventionally applied as the first step and usually takes high-computation time considering the entire 4D volume. Motivated by the recent advances of deep learning in medical imaging, we formulate the direct inference as a non-linear mapping problem between the corrupted image-time series and sparse, topologically less complex *residual* PK maps, where the mapping is learned via deep convolutional neural networks (CNNs). Experiments on clinical DCE brain datasets demonstrate the efficacy of the proposed approach in terms of fidelity of PK parameter reconstruction and significantly faster parameter inference compared to existing state-of-the-art iterative reconstruction method. This contribution has been published as a peer-reviewed conference paper [87] and given in Chapter 5.

Following the above contribution, we demonstrate how a DL based model can replace the conventional pipeline of pharmacokinetic parameter quantification in DCE-MRI, allowing us to directly infer the parameters from the acquired DCE image time series. We propose a new parameter inference framework using deep CNNs which enables more robust and faster inference than traditional NLLS based model fitting approaches. This study is validated on *in-vivo* DCE datasets acquired from clinically evident mild ischaemic stroke patients. The trained CNN model is able to yield PK parameters which can better discriminate different brain tissues, including stroke regions. The results also demonstrate that the CNN model generalizes well to new cases even if a subject specific AIF is not available for the new data under the assumption that the data has been acquired with similar acquisition protocol and parameters. This work has been published as a peer-reviewed journal paper [88] and made available in Chapter 6.

Denoising ASL MRI using kinetic model incorporated loss and deep residual learning

The last contribution of the thesis targets to improve the perfusion-weighted image quality obtained from a subset of all available control/spin-labeled (C/L) pairwise subtractions in ASL imaging. As previously mentioned, ASL suffers from inherently low SNR which requires repeated measurements of C/L pairs in conventional acquisitions, leading to clinically prolonged scanning times. To alleviate this limitation, this work proposes a novel denoising scheme based on a deep fully convolutional neural networks (FCNs) together with *residual learning* strategy to improve the perfusion-weighted image quality obtained by using a lower number of subtracted C/L pairs. Our main contribution to this domain is that we incorporate the Buxton kinetic model – or CBF signal model – in the loss function during training which enables the network to

19

produce high quality images while simultaneously enforcing the CBF estimates to be as close as reference values. We demonstrate the superior performance of our method by validations on synthetic and clinical single-TI ASL datasets. We have published this work as a peer-reviewed conference paper [89] and the manuscript is provided in Chapter 7.

1.5 Thesis overview

This publication-based dissertation is structured as follows. Chapter 2 covers the most relevant methodology to the presented publications, starting with compressed sensing and image reconstruction with prior constraints, followed by a brief summary of machine learning and deep convolutional neural networks.

Chapter 3 to 7 are composed of five publications [84, 85, 87, 88, 89] in their original published versions. They have been published as journal articles or in peer-reviewed conference & workshop proceedings, and thus are all self-contained. Each of these chapters starts with a summary page, containing the full citation of the original publication, the abstract part of the published manuscript and main contributions of the owner of this thesis.

Chapter 8 provides a detailed discussion of the presented methods as well as possible directions for future work.

Appendix A contains an unpublished manuscript on quantitative perfusion MRI reconstruction which were submitted to arXiv [86]. Appendix B presents additional experimental results from related works which are not directly discussed here, but supplement the scope of the thesis. Finally, a complete list of authored and co-authored publications that have been published during the time period of this doctoral thesis can be found in Appendix C.

2

Methodology

This work draws from compressed sensing (CS), MR image reconstruction with prior constraints, machine learning (ML) and convolutional neural networks (CNNs). This chapter aims at giving a concise summary of key concepts and notation used throughout the thesis, but it is not intended to serve as a representative overview of the most important concepts of each subject. For a more complete and in-depth description of CS, we refer the interested readers to [90, 91, 92]. For a thorough review on MR image reconstruction with prior constraints or regularization, we refer to [20, 93, 94], and to [95, 96, 97, 98, 99] for topics related to ML and CNNs for deep learning (DL).

2.1 Compressed sensing

Sparsity representation has been vastly studied and recognized as a powerful tool in many signal and image processing problems, such as image enhancement, image restoration, image segmentation, and signal compression with the recent advancements in compressed sensing (CS). CS is based on the principle that a measured signal can be accurately recovered from far fewer samples than required by Nyquist–Shannon sampling theorem. To make it more concrete, CS aims to perfectly recover the original signal $\mathbf{x} \in \mathbb{C}^N$ from a set of linear measurements $\mathbf{y} \in \mathbb{C}^M$ which are related as follows:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{2.1}$$

where $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ is a sensing operator matrix whose rows contain sensing functions. These functions are mostly obtained by the physics of the acquisition mechanism, for instance, Fourier functions are sensing functions specific to MR sampling. Sub-Nyquist sampling is possible if $\mathbf{\Phi}$ can produce a set of measurements satisfying $M \ll N$. Eq. 2.1 corresponds to an under-determined system of linear equations meaning that for a given $\mathbf{y}$ there is an infinite

number solutions for $\mathbf{x}$. CS proves that this problem is nevertheless solvable and tractable under the following three assumptions:

1. *Signal sparsity:* The signal can be sparsely represented in an appropriate transform basis,

2. *Sampling incoherence:* The measured samples are randomly chosen, i.e. incoherent measurements in the signal or transform domain,

3. *Nonlinear reconstruction:* The CS problem imposes the use of a nonlinear reconstruction.

A very important property of the above assumptions is the *notion of sparsity* in some basis, which corresponds to the number of nonzero (or significant) coefficients in a specific transform domain. This also assumes the property that natural signals can be represented by a linear combination of basis functions. More concretely, given a signal $\mathbf{x} \in \mathbb{C}^N$ we can express it as a coding vector $\gamma \in \mathbb{C}^K$ in a basis or frame $\boldsymbol{\Psi} \in \mathbb{C}^{N \times K}$, with normalized columns $\psi_1, \psi_2, \ldots, \psi_K$, such that $\|\mathbf{x} - \boldsymbol{\Psi}\gamma\|_p = \epsilon$ ensures a very small error for an approximate representation of the signal.

The second assumption given above states that as undersampling leads to aliasing of data points, the behavior of the aliasing artifacts must be incoherent (noise-like) in the transform domain. When the undersampling is not applied in randomized manner, it is impossible to distinguish between signal and its aliases. The measurement matrix $\boldsymbol{\Phi}$ must map two different signals into two different sets of measurements. Hence all column sub-matrices of $\boldsymbol{\Phi}$ must be well contained. For a given constant δ_s and for all k-sparse vectors x_1 and x_2, the measurement matrix $\boldsymbol{\Phi}$ must satisfy the following condition:

$$(1 - \delta_s) \|x_1 - x_2\|_2^2 \leqslant \|\boldsymbol{\Phi}x_1 - \boldsymbol{\Phi}x_2\|_2^2 \leqslant (1 + \delta_s) \|x_1 - x_2\|_2^2. \qquad (2.2)$$

This property is called restricted isometry property (RIP) and the constant δ_s is called restricted isometry constant. The property signifies that all pairwise distances must be well preserved in measurement matrix $\boldsymbol{\Phi}$. It has been proven that many types of random matrices such as independent and identically distributed Gaussian measurement matrix satisfies the RIP.

Once the above mentioned two conditions are met, a nonlinear reconstruction of the signal $\mathbf{x}$ should be performed to enforce the sparsity of the signal in transform domain and the consistency of the reconstruction with the acquired measurements $\mathbf{y}$. This turns into solving the following optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\arg\min} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \\ \text{s.t.} \quad & \|\boldsymbol{\Psi}\mathbf{x}\|_0 \leq k, \end{aligned} \qquad (2.3)$$

where $\mathbf{\Psi}$ is a sparsifying transform of $\mathbf{x}$ and $\|\cdot\|_0$ is an ℓ_0 norm which counts the number of non-zero entries and k is the maximum number of coefficients that are allowed in the sparse representation. $\mathbf{A}$ is known as *encoding matrix* which encodes the combined effect of the sparsifying transform and pseudo-random sampling pattern. In practice, solving for such sparse solution is an NP-hard problem, as selecting k non-zero entries is inherently combinatorial. Fortunately, it turns out that the convex relaxation of Eq. 2.3 robustly converges to the same sparse solution as reformulated below:

$$\operatorname*{argmin}_{\gamma} \left\| \mathbf{A}\mathbf{\Psi}^H \gamma - \mathbf{y} \right\|_2^2 + \lambda \|\gamma\|_1, \tag{2.4}$$

where $\mathbf{x} = \mathbf{\Psi}^H \gamma$ and γ is the corresponding sparse representation of $\mathbf{x}$. The Eq. 2.4 is known as least absolute shrinkage and selection operator (LASSO) regression. The attractiveness of CS is that there is a theoretical guarantee on the number of measurements M required to perfectly reconstruct $\mathbf{x}$ of the dimensionality N that is k-sparse in $\mathbf{\Psi}$. Choosing the appropriate transformation basis $\mathbf{\Psi}$ is an important step in CS reconstruction, and can provide strong prior information for image reconstruction and effective sparse representation.

The analysis of CS almost directly translates to the case of MRI reconstruction problem. As the acquisition time is proportional to the number of samples required in k-space, CS provides a great potential for accelerating MR image reconstruction. Nevertheless, we should note that MRI acquisition process is generally not purely random, and searching for a sampling pattern that makes the acquired k-space completely incoherent in a sparsity transform is a challenging task. Only pseudo-random sampling can be achieved in clinical practice.

2.2 Regularized MR image reconstruction

Let $\mathbf{x} \in \mathbb{C}^N$ denote a discrete, complex-valued MR image to be reconstructed, represented as a vector with $N = N_x \times N_y \times N_z$ where N_x, N_y, and N_z are the width, height and depth of the image, respectively. The image can be 2D $(N_z = 1)$ or 3D, corresponding to a MR volume. Let $\mathbf{y} \in \mathbb{C}^M$ represent all the k-space measurements flattened into a vectorised format. Then, the acquired discrete set of samples in k-space can be expressed as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{2.5}$$

where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is called a forward model, or a Fourier encoding matrix, and $\mathbf{e}$ is a zero-mean complex Gaussian noise with a specific noise variance,

i.e. $\mathbf{e}_i \sim \mathcal{N}(0, \sigma)$. For instance, for 2D Cartesian sampling mentioned in the previous sections, $\mathbf{A}$ represents the sampled Fourier coefficients. In general, $\mathbf{A}$ is generic and it may represent Cartesian or nonuniform data acquisition, as well as incorporating multiple weighted measurements for the case of parallel imaging (PI) techniques. As given in Eq. 2.5, obtaining a quantitatively accurate estimate of $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$ is proven to be a challenging task. The possible reasons of this can be listed as following: i) large size of the problem, ii) natural imperfection of the coils sensitivities, iii) reduced amount of acquired lines due to the undersampling, iv) acquisition noise which corrupts the acquired data. These various factors make the MR reconstruction model an *ill-posed* problem meaning that not all of the well-posedness conditions listed below are satisfied:

1. *Existence of the solution:* For every $\mathbf{y}$ there exists a $\mathbf{x}$ for which $\mathbf{y} = \mathbf{A}\mathbf{x}$.

2. *Uniqueness of the solution:* The solution or estimated $\mathbf{x}$ is unique.

3. *The solution depends smoothly on the data*: Small perturbations in $\mathbf{y}$ do not result in large perturbations in $\mathbf{x}$

In order to overcome the ill-posedness of the problem, MR image reconstruction is typically cast as a regularized optimization problem formulated as follows:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \tag{2.6}$$

where reconstructed image $\mathbf{x}$ is found by balancing the trade-off between data consistency ($\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 < \epsilon$) and reduction of measurement artifacts. The regularization term $\mathcal{R}(\mathbf{x})$ is incorporated as an additional *constraint* on the reconstructed image, trading-off between reduction of noise and preservation of image structures and edges. The parameter λ controls the degree of regularization in the reconstructed image and is usually adapted to the noise level. The performance of a specific reconstruction technique thus depends on the good choice of regularizer $\mathcal{R}$ as well as a robust optimization method. In particular, incorporating sparsity and statistical properties of images into the reconstruction problem have shown great potential to enhance image quality and sharpness. The regularization terms that have been extensively applied for MRI reconstruction problem in literature involves wavelet, ℓ_0 and ℓ_1 sparsity, low-rank regularization, total variation – due to the inherent piecewise smooth structure of MR images –, and patch-based regularization. For a detailed description of various types of regularizers or penalty terms used in CS-based MRI reconstruction we refer to [20, 94].

After a particular sparsity and regularization term is chosen, advanced and robust convex optimization techniques are essentially used to solve the optimization problem in Eq. 2.6. These techniques enable to iteratively reconstruct the final MR image, hence known as iterative reconstruction algorithms. Depending on the complexity and number of the chosen sparsity and regularization terms, researchers have employed various optimization methods to solve the iterative reconstruction problem, including nonlinear conjugate gradient [100], alternating direction method of multipliers (ADMM) [101], alternating minimization algorithms [102], iterative shrinkage-thresholding algorithms [103, 104], and composite splitting based algorithms such as proximal forward-backward splitting [105]. For a comprehensive review of the several key optimization algorithms used in MR reconstruction domain we refer the readers to [106].

2.3 Machine learning

Machine learning (ML) allows to statistically learn complex functions or socalled *mappings* by building on knowledge from a set of observations. As opposed to the rule-based systems, which are mainly based on defined heuristics, a ML system can learn directly from data and subsequently act as a predictive model. Depending on the availability and use of label data in the observations, machine learning algorithms can be divided into three different learning categories:

- *Supervised learning:* It refers to ML algorithms where the *ground truth* is available in the training set, i.e. class labels or continuous outputs. The ground truth information is supposed to be reliable to govern the learning process.

- *Unsupervised learning:* Unlike the supervised learning, unsupervised learning does not require any ground truth for the training. This type of learning essentially aims to learn the intrinsic structure of the manifold.

- *Semi-supervised learning:* This learning approach contains the ML algorithms where only very small amount of labeled data are used together with a relatively large amount of unlabeled data. The main objective is to improve the supervised learning by leveraging large amount of unlabeled data during training stage especially when there is limited access to the labeled data.

Among these three learning strategies, supervised learning has been the most commonly used strategy on the real-world prediction problems. In a supervised learning framework, we can build on a training set $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ involving available ground truth or labels to learn a function f as follows:

$$\mathbf{y} = f(\mathbf{x}; \mathbf{w}), \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathcal{D} \tag{2.7}$$

where $\mathbf{w}$ are the function parameters or known as *weights* which can be learned. The learned function mainly maps an input $\mathbf{x}$ to an output $\mathbf{y}$ based on input-output pairs from training set, and estimates an approximation of the ground truth such that $f(\mathbf{x}; \mathbf{w}) = \hat{\mathbf{y}} \approx \mathbf{y}$.

Loss function. The parameters of mapping function f are optimized during *learning* or *training* stage using the labeled training set according to a predefined loss function. For supervised learning, the general formula of the loss function $\mathcal{L}$ for a given ground truth and predicted labels or output,

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \min_{f} \sum_{i=1}^{N} L\left(\mathbf{y_i}, f\left(\mathbf{x_i}; \mathbf{w}\right)\right), \tag{2.8}$$

where $L(\cdot)$ is the chosen loss function that describes the cost of misclasifying the ground truth label which may be either discrete or continuous. Prior knowledge, i.e. domain-specific knowledge, can be incorporated into the loss function – as a soft constraint – to find an optimal or sub-optimal solution. The regularized loss function is formulated using Lagrangian multiplier as

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \min_{f} \sum_{i=1}^{N} \underbrace{L\left(\mathbf{y_i}, f\left(\mathbf{x_i}; \mathbf{w}\right)\right)}_{\text{Prediction Error}} + \underbrace{\lambda \cdot R(f)}_{\text{Regularization}}, \tag{2.9}$$

where λ is the regularization parameter to control the effect of the prior information.

Learning. The learning procedure of supervised learning tries to determine the weights $\hat{\mathbf{w}}$ satisfying the minimum loss function value $\mathcal{L}$ given the labeled training set $\mathcal{D} = \{(\mathbf{x}_d, \mathbf{y}_d)\}_{d=1}^{n_d}$ as follows:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \mathcal{L}\left(\mathcal{D}, \mathbf{w}\right). \tag{2.10}$$

In ML domain, this is also referred to as *training*, *fitting* or *parameter inference* specificially in the context of probabilistic ML models. The loss function $\mathcal{L}\left(\mathcal{D}, \mathbf{w}\right)$ usually does not have closed-form solution, and hence finding a global minimum of the loss function is usually hard. Many algorithms rely on

approximate or heuristic optimization schemes and do not thus guarantee to find the exact optimum of Eq. 2.10. One of the most prominent examples are deep neural networks (NNs), where optimization can be done with stochastic gradient descent (SGD). The learned weights $\hat{\mathbf{w}}$ may correspond to the different forms with respect to both the type of classification problem and used ML algorithm. For instance, if we use support vector machine (SVM) classifier for a binary classification problem, a combination of $\hat{\mathbf{w}}$ represents a set of hyperplanes in a high dimensional input space, which minimizes the loss and separates the samples of two classes such that the distance between them is as large as possible.

In the scope of a traditional ML pipeline, we are not only interested in how well the learned model fits the training data $\mathcal{D}$, but also essentially how well it generalizes on *unseen* data, which denotes the set of samples that were not made available during the learning stage. To this end, within the ML pipeline it is a common practice to split the available dataset $\mathcal{D}$ into a training and validation set, where the former is used to learn the weights $\hat{\mathbf{w}}$ and the latter is used to validate the performance of the learned model on unseen data. Extensively used resampling methods for validating the model's performance are cross-validation and bootstrapping. One of the common problems in ML is known as *bias-variance trade-off*. Bias measures the difference between the average prediction of our model and the correct value which we are trying to predict. Variance is defined as deviation of predictions, in simple it tells us how much the estimate of the learned model will change if different training data is used. In learning stage, one should usually need to avoid the *underfitting* and *overfitting* problems to improve the generalization capability of the model. In general, high bias results in underfitting of our model where we miss some important and relevant relations between the feature and target outputs. On the other hand, high variance is likely caused by overfitting of the model in which we learn the random noise in the training data rather than the intended outputs, therefore the predictions on the different datasets exhibit high variance.

Inference. Once the weights of the model $\hat{\mathbf{w}}$ have been learned, we can predict the associated target $\mathbf{y}$ of an unseen test sample or observation $\tilde{\mathbf{x}}$ as given in Eq. 2.7, i.e. $\hat{\mathbf{y}} = f(\tilde{\mathbf{x}}; \hat{\mathbf{w}})$. This step is also called *testing* or *prediction*. If $\mathbf{y}$ is categorical, the problem is known as a *classification* problem and if it is continuous, we name it as *regression* problem. Unlike the learning step, the inference phase is usually fast to evaluate as the model weights are already available. For instance, for linear classifiers, this step often involves just an inner product and a thresholding operation applied on the given test sample $\tilde{\mathbf{x}}$.

Majority of ML algorithms exploit the concepts from several domains, including pattern recognition, statistical learning theory, mathematical optimization and data mining. Some of the examples for widely-used learning algorithms are SVM [107], the random forest (RF) [108], the neural network (NN) [109, 110] and its more recent variants that employ the concepts of deep learning [99, 111]. Over the last few years ML techniques have witnessed a tremendous amount of attention in a broad range of fields. Especially in medical image analysis, ML methods have been extensively used in a wide range of tasks [112, 113] such as segmentation of brain tumors, stroke lesion and multiple body organs, image reconstruction and enhancement, image registration, disease diagnosis or prognosis, and so on.

Within the scope of this thesis, we have mainly exploited ML based models for parameter reconstruction and image enhancement (specifically denoising) tasks, each of which was formulated as a supervised regression problem.

2.4   Convolutional neural networks

Convolutional neural networks (CNNs) are a class of neural networks with properties of using spatially structured neurons, local connectivity and weight sharing. These networks have been shown to be extremely effective for numerous applications in computer vision in the recent years [114, 115, 116]. CNN essentially extracts features locally, which can build complex image representations while maintaining spatial correspondences. The building blocks of CNNs are convolutional layers and fully connected layers, pooling layers, nonlinearity layers, and normalisation layers. In the following we will describe each of these layers and provide a brief summary of how the training via *backpropagation* is employed in the CNN framework.

2.4.1   CNN layers

Convolutional layer.   The most basic building block of a CNN is the convolutional layer. In a convolutional layer, each node is connected only locally to its input layer, in which a set of shared *kernels* or *filters* are convolved with the input to create the next layer. The layer multiplication can be performed via a convolution operation,

$$W^T x = w \otimes x, \tag{2.11}$$

where W has a form of *Toeplitz matrix*. More formally, for an input tensor $\mathbf{x} \in \mathbb{R}^{N_x \times N_y \times N_c}$, e.g. a 2D image with N_c channels, and a weight $\mathbf{w} \in$

$\mathbb{R}^{k_x \times k_y \times N_c \times N_{c'}}$ and bias $\mathbf{b} \in \mathbb{R}^{N_{c'}}$, convolution layer f is defined as,

$$[f(\mathbf{x}; \mathbf{w}, \mathbf{b})]_{l,m,n} = \sum_{i=l-k_x/2}^{l+k_x/2} \sum_{j=m-k_y/2}^{m+k_y/2} \sum_{k=0}^{N_{c'}} \mathbf{w}_{i,j,k,n} \mathbf{x}_{i,j,k} + \mathbf{b}_n. \qquad (2.12)$$

Weights in convolution layers are often called convolutional kernels. The hyper-parameters of convolutional layers typically include kernel width (k_x, k_y), number of features $N_{c'}$, stride and dilation factors. In comparison to fully connected layer in feed-forward networks, convolutional layer has much less parameters as the weights are shared across the image spatially. This enables the network to extract local features and build complex representations before they are aggregated for further analyses.

Pooling layer. A pooling layer is used to locally aggregate the statistics of the intermediate features within a certain pooling window. If a pooling window size is $p_x \times p_y$, a max-pooling layer on $\mathbf{x}$ is defined as,

$$\mathbf{x}_{k,n,m}^{(l)} = \max_{\substack{i \in [k-p_x/2,...,k+p_x/2] \\ j \in [n-p_y/2,...,n+p_y/2]}} \mathbf{x}_{i,j,m}^{(l-1)}. \qquad (2.13)$$

Average pooling is another commonly used operation which instead computes the average value in pooling region. As the layer aggregates the local values, CNNs involving pooling layers are less sensitive to local perturbations. This attribute is often known as *translational invariance*. Another key benefit of a pooling layer is to reduce the spatial dimension of input tensor. The stride of the operation is often matched by the pooling window size. This enables that the output of pooling is down-scaled by a factor proportional to the pooling window size. Once the representation is down-scaled, the network can learn a representation which aggregates the local information and effectively build a complex hierarchical representation.

Fully connected layer. Following several convolutional and pooling layers, the high-level feature extraction in the CNNs is usually performed via fully connected (FC) layers. The output feature maps of the final convolution or pooling layer is typically flattened, i.e. transformed into a one-dimensional (1D) vector, and connected to one or more FC layers, also known as *dense* layers, in which every input neuron is connected to every output neuron by unique learnable weights. Their activation can thus be calculated via a matrix multiplication followed by a bias offset. The output vector of FC layers can be either fed forward into a certain number of categories for classification task or used as a feature vector for further processing.
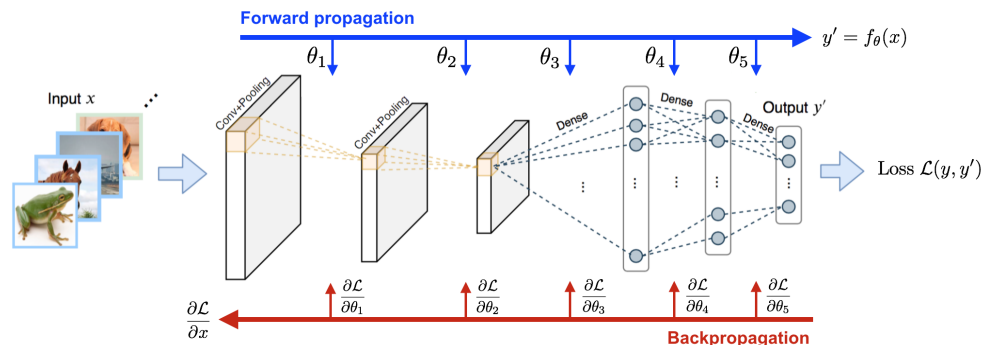
Figure 2.1: The schematic illustration of how an exemplary convolutional neural network based architecture is trained via backpropagation for an image classification task. (Image adapted from [117])

Nonlinearity layer. Although the convolutional and fully connected layers standalone can be used to build up the relation between input and output, their intrinsic linear nature diminishes their utility. Thus, nonlinear activations are an important component of neural networks as they enable learning arbitrarily complex mappings. A popular choice of a nonlinearity is rectified-linear unit (ReLU), which is defined element-wise as follows:

$$\mathrm{ReLU}(x) = \left\{ \begin{array}{ll} x & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{array} \right. \tag{2.14}$$

ReLU is a nonlinear activation function which solves *gradient vanishing problem* of other nonlinear functions like *sigmoid* and *tanh*, enabling training of extremely deep networks. In practice, most activation functions are applied right after the linear operation to induce variance in neuron activations.

Normalization layer. As observed in other ML algorithms, neural networks (NNs) can be very sensitive to numerical problems as the introduced operations are unbounded in theory. This results in unstable training where parameter updates can either vanish towards 0 or explode to high values, effectively resulting in dead neurons or constantly oversatured outputs. To this end, it is a common practice to employ batch or group normalization during the training phase of the CNNs to stabilize the training. Moreover, convolution bias can be eliminated when applying a post-activation normalization after each layer since each intermediate input is whitened.

30

2.4.2 Network training

During *training*, the input $\mathbf{x}$ of a training sample $(\mathbf{x}, \mathbf{y})$ is fed through the network with pre-initialized parameters and results in a prediction $\hat{\mathbf{y}}$. The prediction error can be determined by a loss function defined as follows:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{i=1}^{n} \phi\left(\mathbf{y}_i, \hat{\mathbf{y}}_i\right) = \sum_{i=1}^{n} \phi\left(\mathbf{y}_i, f\left(\mathbf{x}_i; \mathbf{w}, \mathbf{b}\right)\right), \tag{2.15}$$

where $\phi(\cdot)$ computes the distance of $\hat{\mathbf{y}}$ to $\mathbf{y}$, e.g. ℓ_2 norm. Given the loss function, the network weights are updated via *backpropagation* so that the network output becomes closer to the target output for the next training sample.

For any given differentiable loss function $\mathcal{L}$ evaluated with network output $f(\mathbf{x})$ and expected output $\mathbf{y}$, the gradient $\frac{\partial \mathcal{L}(f(\mathbf{x}), \mathbf{y})}{\partial \theta}$ allows us to optimize the model parameters θ. Let us assume that our network consists of multiple convolution and activation blocks of the following form,

$$f_i\left(\mathbf{x}_i\right) = \sigma_i\left(\mathbf{x}_i * \mathbf{w}_i + \mathbf{b}_i\right), \tag{2.16}$$

where $\sigma(\cdot)$ represents the nonlinear activation function. To efficiently calculate a gradient update, as depicted in Fig. 2.1, we can leverage the layered structure by working backwards from the output and computing the partial derivatives along the way. Starting from the last layer, we compute its gradient as follows:

$$\frac{\partial \mathcal{L}(f(\mathbf{x}), \mathbf{y})}{\partial \theta_n} = \mathcal{L}'\left(f_n\left(\mathbf{x}_n\right), \mathbf{y}\right) \cdot \sigma_n'\left(\mathbf{x}_n * \mathbf{w}_n + \mathbf{b}_n\right) \cdot \left(\mathbf{x}_n * \mathbf{w}_n + \mathbf{b}_n\right)'. \tag{2.17}$$

Following the chain rule, the changes in the network are propagated up to the first layer to assemble the full gradient and perform a descent step. The established way to train neural networks relies on first-order optimization variants such as stochastic gradient descent (SGD) with momentum [118] or adaptive variants like ADAM [119]. However, training deep networks efficiently and precisely is still ongoing research problem as it is challenging to entirely capture the nonlinear and highly nonconvex behavior.

During *testing*, a new unseen sample of $\tilde{\mathbf{x}}$ is fed forward through the network with learned weights and the final layer produces the prediction of the test sample.

3

Spatio-temporal MRI Reconstruction by Enforcing Local and Global Regularities

Abstract: In this paper, we propose a new spatio-temporal reconstruction scheme for the fast reconstruction of dynamic magnetic resonance imaging (dMRI) data from undersampled k-space measurements. To utilize both spatial and temporal redundancy in dMRI sequences, our method investigates the potential benefits of enforcing local spatial sparsity constraints on the difference to a reference image for each frame and additionally exploiting the low-rank property of global spatio-temporal signal via nuclear norm (NN) minimization. We present here an iterative algorithm that solves the convex optimization problem in an alternating fashion. The proposed method is tested on in-vivo 3D cardiac MRI and dynamic susceptibility contrast (DSC)-MRI brain perfusion datasets. In comparison to two state-of-the-art methods, numerical experiments demonstrate the superior performance of our method in terms of reconstruction accuracy.

Contributions of thesis author: Design and implementation of the proposed reconstruction algorithm, experimental design, data analysis, interpretation, composition and revision of manuscript.

SPATIO-TEMPORAL MRI RECONSTRUCTION BY ENFORCING LOCAL AND GLOBAL REGULARITY VIA DYNAMIC TOTAL VARIATION AND NUCLEAR NORM MINIMIZATION

Cagdas Ulas [†¶⋆]   *Pedro A. Gómez* [†¶]   *Jonathan I. Sperl* [¶]   *Christine Preibisch* [‡]   *Bjoern H. Menze* [†§]

[†] Department of Computer Science, TU München, Munich, Germany
[§] Institute for Advanced Study, TU München, Munich, Germany
[‡] Neuroradiology, Klinikum rechts der Isar der TU München, Munich, Germany
[¶] GE Global Research, Munich, Germany

ABSTRACT

In this paper, we propose a new spatio-temporal reconstruction scheme for the fast reconstruction of dynamic magnetic resonance imaging (dMRI) data from undersampled k-space measurements. To utilize both spatial and temporal redundancy in dMRI sequences, our method investigates the potential benefits of enforcing local spatial sparsity constraints on the difference to a reference image for each frame and additionally exploiting the low-rank property of global spatio-temporal signal via nuclear norm (NN) minimization. We present here an iterative algorithm that solves the convex optimization problem in an alternating fashion. The proposed method is tested on in-vivo 3D cardiac MRI and dynamic susceptibility contrast (DSC)-MRI brain perfusion datasets. In comparison to two state-of-the-art methods, numerical experiments demonstrate the superior performance of our method in terms of reconstruction accuracy.

Index Terms— compressed sensing, dynamic MR imaging, low-rank approximation, total variation, nuclear norm

1. INTRODUCTION

Dynamic magnetic resonance imaging (dMRI) is an important medical imaging technique that enables the visualization of anatomical and functional changes of internal body structures through time, resulting in a spatio-temporal signal. Although MRI is a non-invasive, non-ionizing technology and provides an unmatched quality in soft tissue contrast, physical and physiological limitations on scanning speed makes this an inherently slow process [1]. Besides, there is a trade-off between the spatial and temporal resolution. The reason is that acquiring fewer k-space samples than those dictated by the Nyquist criterion accelerates the process significantly, but exhibits aliasing artifacts in image space. Fortunately, dynamic MR sequences usually provide redundant information

in both spatial and temporal domains, which allows the reduction of acquisition time by using compressed sensing (CS) approaches [2, 3]. More recently, CS theory has been applied to MRI enabling highly accurate reconstructions from fewer k-space measurements depending on the assumption of sparsity of the reconstructed data under some transform domain [4].

In recent years, researchers have proposed sophisticated CS-based reconstruction methods that exploit both spatial and temporal redundancies of the entire dataset, such as spatio-temporal total variation [5], dictionary learning [6], and low-rank approximation and sparsity [5, 7]. In general, dynamic MR images are temporally correlated due to slow changes of the same organ(s) through the whole image sequence, and such high correlation in the temporal domain has been successfully investigated based on a sparsity constraint in the temporal domain for dMRI reconstruction [6]. As an extension of the conventional spatial total variation (TV), a new sparsity inducing norm called dynamic total variation (dTV) [8] has been recently introduced to utilize both spatial and temporal correlations in online reconstruction.

In this paper, we make an attempt to integrate two fundamentally different approaches for CS-based reconstruction: we enforce local coherences at the pixel-level via dynamic total variation (dTV) and global regularity in the full spatio-temporal domain via a nuclear norm (NN) minimization constraint. We present the dTV/NN optimization in a joint formal framework which allows us to rely on an iterative minimization algorithm. The joint minimization problem is solved iteratively by utilizing an alternating minimization strategy. The proposed method is validated on two different dynamic MR sequences with comparisons to state-of-the-art methods.

Our main contributions can be summarized as follows: We propose a novel reconstruction scheme that iteratively enforces not only the local (spatial) regularity in every single frame but also the global (spatio-temporal) regularity of a full sequence. To this end, we introduce a reconstruction model that is jointly using dTV sparsity and nuclear norm penalties, exploiting both the sparsity of inter-frame differences and the low-rank structure of the dynamic MR sequences in the full

spatio-temporal space. Our approach also employs, for the first time, the dTV sparsity inducing norm in an offline reconstruction scheme.

2. METHODS

2.1. Problem formulation

Here, we denote $\mathbf{X}^{3D}$ as a dMRI sequence to be represented as a spatio-temporal 3D volume of size $P = N \times N \times T$, i.e., the images are of size $N \times N$ and T is the total number of frames in the sequence. Let X_t denote the MR image matrix at the tth frame, y_t is the k-space data for the tth frame and $\mathbb{T} = \{1, 2, ..., T\}$ is the set of frame number indices. The main objective here is to reconstruct all X_t's, $t \in \mathbb{T}$, from the collected k-space measurements y_t's. The MR imaging equation for each frame is formulated as

$$y_t = \mathcal{F}_t x_t + \eta \qquad (1)$$

where $\mathcal{F}_t$ denotes the undersampling 2D Fourier operator for frame t, i.e., $\mathcal{F}_t = R_t \mathcal{F}_{2D}$, where $R_t \in \mathbb{R}^{m \times N}$, $m \ll N$, is the undersampling mask to acquire only a subset Ω of k-space, x_t denotes the MR image vector formed by row/column concatenation of X_t and $\eta \in \mathbb{C}^m$ is additive Gaussian noise in k-space. We stack the data for all the frames of the MR sequence as columns and denote them as follows: $Y = [y_1|y_2|...|y_T]$, $X = [x_1|x_2|...|x_T]$, and $\mathcal{F}_u = \text{diag}\{\mathcal{F}_1, \mathcal{F}_2, .., \mathcal{F}_T\}$.

We propose solving the following optimization problem for the reconstruction of dMRI sequences:

$$\min_X \quad \nu_1 \|X\|_* + \nu_2 (dTV(X, \bar{x}))$$
$$\text{s.t.} \quad \|\mathcal{F}_u X - Y\|_2^2 \leq \epsilon \qquad (2)$$

where ν_1 and ν_2 are respective regularization parameters for the two terms, and $\|X\|_*$ denotes the nuclear norm of X and is calculated as

$$\|X\|_* = \sum_i \sigma_i(X)$$

where $\sigma_i(X)$ represents the ith singular value of X. For an image x_t with N^2 pixels, $dTV(X, \bar{x})$ can be defined as

$$dTV(X, \bar{x}) = \sum_{t \in \mathbb{T}} \sum_{n=1}^{N^2} \sqrt{\left(\nabla_x (x_t - \bar{x})_n\right)^2 + \left(\nabla_y (x_t - \bar{x})_n\right)^2}$$

where $\bar{x}$ is the reference image calculated by averaging all the frames in the sequence, ∇_x and ∇_y represent the finite-difference matrices along the x and y dimensions respectively.

Let us introduce new variables $z_t = x_t - \bar{x}$ and $b_t = y_t - \mathcal{F}_t \bar{x}$, then the problem (2) can be reformulated as follows:

$$\min_{X,z} \quad \nu_1 \|X\|_* + \nu_2 \sum_{t \in \mathbb{T}} \|z_t\|_{TV}$$
$$\text{s.t.} \quad \begin{cases} \|\mathcal{F}_u X - Y\|_2^2 \leq \epsilon \\ \|\mathcal{F}_t z_t - b_t\|_2^2 \leq \epsilon, \ \forall t \end{cases} \qquad (3)$$

where $z = [z_1, ..., z_T]$ and $\|z_t\|_{TV} = \|[D_1 z_t, D_2 z_t]\|_{2,1}$, where D_1 and D_2 are two $N^2 \times N^2$ first order finite difference matrices in vertical and horizontal directions, and $\ell_{2,1}$ norm is the summation of the ℓ_2 norm of each row, $[\mathbf{a_1}, \mathbf{a_2}]$ denotes concatenating two vectors $\mathbf{a_1}$ and $\mathbf{a_2}$ horizontally.

2.2. Image reconstruction algorithm

The optimization problem (3) is convex and we choose to split it into two simpler subproblems that can be efficiently solved with greedy algorithms. Alternating the solution of these two subproblems iteratively will give an approximate solution to Eq. (3). In this approach, an approximate generic solution is refined towards a better solution.

- Subproblem 1 : *Enforcing local (spatial) regularity*

$$\min_{z_t} \quad \frac{1}{2} \|\mathcal{F}_t z_t - b_t\|_2^2 + \nu_2 \|z_t\|_{TV}, \ \forall t \qquad (4)$$

For each frame x_t in the sequence, we solve the optimization problem (4) to reconstruct each frame individually given a reference image $\bar{x}$. This guarantees that the sum of TV norms in Eq. (3) is also minimized. The problem (4) can be efficiently solved by the fast iteratively reweighted least squares (FIRLS) algorithm [9] based on preconditioned conjugate gradient. This algorithm provides fast convergence and low computational cost by designing a new preconditioner which can be accurately approximated using the properties of the Fourier transform and diagonally dominant structure of the $\mathcal{F}_t^H \mathcal{F}_t$ matrix, where H denotes the conjugate transpose. We refer the reader to [9] for more details on FIRLS.

- Subproblem 2 : *Enforcing spatio-temporal regularity*

$$\min_X \quad \frac{1}{2} \|\mathcal{F}_u X - Y\|_2^2 + \nu_1 \|X\|_* \qquad (5)$$

The spatio-temporal signal representation of a dMRI sequence can be arranged as a 2D matrix of X, where each column represents a vectorized image frame. Due to the repetitive structure of the dMRI sequence between consecutive frames, and the resulting high correlation between each column of X, this matrix can be generally approximated to be low-rank, i.e., X has only a few significant singular values.

By exploiting the low-rank property of X, we can solve a low-rank matrix recovery problem using convex nuclear norm as a prior. In this way, we pose low-rank matrix recovery as a nuclear norm regularized linear least squares problem as stated in (5). This problem can be solved iteratively through an accelerated proximal gradient (APG) algorithm [10]. The algorithm provides a computationally efficient way of recovering low-rank matrices iteratively and consists of two main steps: a first order update and a proximal projection of the penalty that is solved via the singular value thresholding operator, i.e., $S_\alpha(\mathbf{G}) = \mathbf{U} \text{diag}\{(\mathbf{\Sigma} - \alpha)_+\} \mathbf{V}^H$, where $\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}$ are obtained from singular value decomposition of $\mathbf{G}$.

Our proposed scheme follows an iterative refinement of an

307

Algorithm 1 NNMdTV reconstruction

1: **Input:** $Y, \nu_1, \nu_2, \mathcal{F}_u, IterNo$
2: **Output:** X
3: **Initialize:** $X = X_0 = \mathcal{F}_u^H Y, \bar{x} = \bar{x}_0 = mean(\mathcal{F}_u^H Y)$
4: **for** $i = 1$ to $IterNo$ **do**
5:    **for each** $t \in \{1, 2, ..., T\}$ **do**
6:      $\hat{z}_t \leftarrow \arg\min_{z_t} \quad \frac{1}{2}\|\mathcal{F}_t z_t - b_t\|_2^2 + \nu_2\|z_t\|_{TV}$
7:      $x_t \leftarrow \hat{z}_t + \bar{x}_i$
8:    **end for**
9:    Form updated $X_i = [x_1|x_2|...|x_T]$
10:   $X_i \leftarrow \arg\min_{X} \quad \frac{1}{2}\|\mathcal{F}_u X - Y\|_2^2 + \nu_1\|X\|_*$
11:   $\bar{x}_i \leftarrow mean(X_i)$
12: **end for**

---

initial solution. First, we start with zero-filled sequence and iteratively improve the previous reconstruction by first solving the Subproblem 1 for each frame and refining this solution by solving the Subproblem 2 as a following step. Second, in each iteration we update the reference image that is used for solving Subproblem 1, providing a better reference image given as the input to the problem (4), thus yielding more accurate reconstructions. Throughout the paper we will simply term our proposed method as NNMdTV. Algorithm 1 summarizes the steps of the NNMdTV algorithm.

3. EXPERIMENTS AND RESULTS

3.1. Experimental setup

We evaluate our method on two different types of dynamic MR sequences. We use an in-vivo breath-hold cardiac perfusion sequence [8] of size $192 \times 192 \times 40$ and a dynamic susceptibility contrast (DSC)-MRI brain perfusion sequence of size $128 \times 128 \times 60$ with normalized intensities. Both sequences are artificially corrupted by multiplying its corresponding k-space representation with a binary undersampling mask and subsequently adding complex Gaussian white noise with a standard deviation σ. A radial sampling mask is used to simulate undersampling. The same undersampling mask is used for all frames in our experiments.

3.2. Evaluation

For quantitative evaluation, we adopt the Peak Signal-to-Noise Ratio (PSNR) as the metric in our experiments. We compare our method with two state-of-the-art methods: k-t SLR [5] and dynamic total variation (dTV) [8]. The codes of dTV and k-t SLR reconstruction methods are downloaded from each author's website and for k-t SLR we use the default parameter settings in the package for all experiments. For dTV reconstruction, we use the first frame as the reference frame with $1/4$ sampling rate and $1/6$ sampling rate for the remaining frames. The sampling rate for all frames is also

set to $1/6$ for NNMdTV and k-t SLR. To ensure fair comparison, the parameters settings used in dTV reconstruction are also used in our NNMdTV method for all experiments. For the NNMdTV method, we set $\nu_1 = 5 \times 10^{-8}$, $\nu_2 = 0.001$ and $IterNo = 5$ for both sequences. The noise standard deviation is set to $\sigma = 10^{-5}$ for all reconstruction methods.

3.3. Experimental results

In Figs. 1 and 2, we present qualitative results for the DSC brain and cardiac perfusion datasets respectively. Fig. 1 shows the temporal profile of the DSC brain data along a fixed row. From the error maps (see Fig. 1(d, f, h)), it is clearly visible that NNMdTV reconstructs better result compared to other two methods. The red arrows in Fig. 1(h) show the regions where the reconstruction is improved with NNMdTV. A frame of the reconstructed cardiac sequence is shown in Fig. 2. Visible artifacts can clearly be seen on the images reconstructed by k-t SLR. In contrast, compared to the dTV, the reconstruction result of NNMdTV is more similar to the fully-sampled frame, and less noisy (see Fig. 2(h)).

Quantitative results of different methods on two perfusion datasets are shown in Fig. 3. From the figure, we can clearly observe that the proposed NNMdTV achieves the highest PSNR for each iteration and for all frames of the sequences. The graphs at the top of Fig. 3 mainly validate the fact that iteratively updated mean reference image in NNMdTV enables better reconstruction accuracy.

4. CONCLUSION

In this paper, we have proposed a new CS-based reconstruction model for dynamic MRI based on the joint minimization of local differences in each frame and global differences in the full spatio-temporal space and developed an iterative reconstruction algorithm to solve this minimization problem.
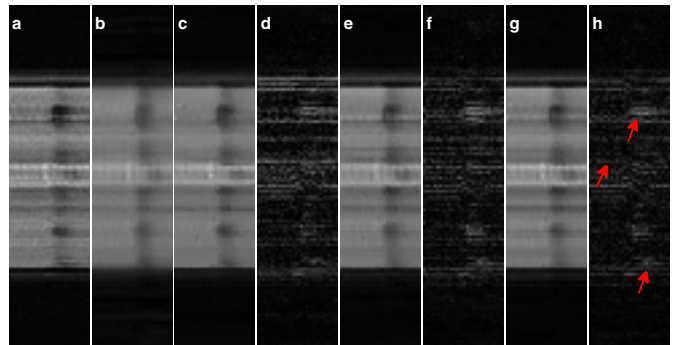
Fig. 1. Temporal profile of row 75 in the original DSC brain dataset (a), its undersampled by 6 zero-filled version (b), and reconstructions using k-t SLR (c), dTV (e), and NNMdTV (g) with their respective errors multiplied by 3 (d, f, h).

Fig. 3. PSNR comparisons of different reconstruction methods. Cardiac dataset (left), DSC brain dataset (right).

Fig. 2. Visual comparison of a fully sampled frame of cardiac dataset (a), its undersampled by 6 zero-filled version (b), and reconstructions using k-t SLR (c), dTV (e), and NNMdTV (g) with their respective errors magnified by 4 (d, f, h).

Experiments on two different perfusion datasets have demonstrated the effectiveness of our method over the state-of-the-art. Future work will aim at extending our method with the use of patch-wise redundancies of spatio-temporal neighborhoods in adjacent frames and making it more robust to noisy scenarios and large inter-frame motion.

5. REFERENCES

[1] P. A. Gomez et al., "Learning a spatiotemporal dictionary for magnetic resonance fingerprinting with compressed sensing," in *1st Int. Work. on Patch-based Techniques in Medical Imaging, MICCAI*, October 2015.

[2] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[3] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, no. 2, pp. 365–373, 2008.

[4] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, 2011.

[5] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. on Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.

[6] J. Caballero, A.N. Price, D. Rueckert, and J. Hajnal, "Dictionary learning and time sparsity for dynamic MR data reconstruction," *IEEE Trans. on Med. Imag.*, vol. 33, no. 4, pp. 979–994, April 2014.

[7] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled (k,t)-space data with joint partial separability and sparsity constraints," *IEEE Trans. on Med. Imag.*, vol. 31, no. 9, pp. 1809–1820, September 2012.

[8] C. Chen, Y. Li, L. Axel, and J. Huang, "Real time dynamic MRI with dynamic total variation," in *Proc. 17th Int. Conf. MICCAI*, 2014, LNCS, pp. 138–145.

[9] C. Chen, J. Huang, L. He, and H. Li, "Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014, pp. 2713–2720.

[10] K. C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimazition*, vol. 6, pp. 615–640, 2010.

309

4

Robust Reconstruction of Accelerated Perfusion MRI Using Local and Nonlocal Constraints

This work has been published as **peer-reviewed workshop paper**.

Abstract: Dynamic perfusion magnetic resonance (MR) imaging is a commonly used imaging technique that allows to measure the tissue perfusion in an organ of interest via assessment of various hemodynamic parameters such as blood flow, blood volume, and mean transit time. In this paper, we tackle the problem of recovering perfusion MR images from undersampled k-space data. We propose a novel reconstruction model that jointly penalizes spatial (local) incoherence on temporal differences obtained based on a reference image and the patch-wise (nonlocal) dissimilarities between spatio-temporal neighborhoods of MR sequence. Furthermore, we introduce an efficient iterative algorithm based on a proximal-splitting scheme that solves the joint minimization problem with fast convergence. We evaluate our method on

dynamic susceptibility contrast (DSC)-MRI brain perfusion datasets as well as on a publicly available dataset of in-vivo breath-hold cardiac perfusion. Our proposed method demonstrates superior reconstruction performance over the state-of-the-art methods and yields highly accurate estimation of perfusion time profiles, which is very essential for the precise quantification of clinically relevant perfusion parameters.

Contributions of thesis author: Design and implementation of the proposed reconstruction algorithm, computational experiments and interpretation, composition and revision of manuscript.

Robust Reconstruction of Accelerated Perfusion MRI Using Local and Nonlocal Constraints

Cagdas Ulas[1,3]($^($ $^)$), Pedro A. Gómez[1,3], Felix Krahmer[2], Jonathan I. Sperl[3], Marion I. Menzel[3], and Bjoern H. Menze[1,4]

[1] Computer Science, Technische Universität München, Munich, Germany
cagdas.ulas@tum.de
[2] Applied Numerical Analysis, Technische Universität München, Munich, Germany
[3] GE Global Research, Munich, Germany
[4] Institute for Advanced Study, Technische Universität München, Munich, Germany

Abstract. Dynamic perfusion magnetic resonance (MR) imaging is a commonly used imaging technique that allows to measure the tissue perfusion in an organ of interest via assessment of various hemodynamic parameters such as blood flow, blood volume, and mean transit time. In this paper, we tackle the problem of recovering perfusion MR images from undersampled k-space data. We propose a novel reconstruction model that jointly penalizes spatial (local) incoherence on temporal differences obtained based on a reference image and the patchwise (nonlocal) dissimilarities between spatio-temporal neighborhoods of MR sequence. Furthermore, we introduce an efficient iterative algorithm based on a proximal-splitting scheme that solves the joint minimization problem with fast convergence. We evaluate our method on dynamic susceptibility contrast (DSC)-MRI brain perfusion datasets as well as on a publicly available dataset of in-vivo breath-hold cardiac perfusion. Our proposed method demonstrates superior reconstruction performance over the state-of-the-art methods and yields highly accurate estimation of perfusion time profiles, which is very essential for the precise quantification of clinically relevant perfusion parameters.

1 Introduction

Medical diagnosis and research heavily employ perfusion-weighted magnetic resonance imaging (MRI) techniques to estimate the blood flow and volume through examination of the spatio-temporal changes of the signal intensities following the injection of a blood bolus via exogenous paramagnetic tracers. In neuroimaging, these techniques have become widespread clinical tools in the diagnosis of stroke – for the assessment of the tissue at risk –, and the treatment of patients with cerebrovascular disease. One of the major obstacles in the clinical use of perfusion imaging is the need to track the rapid kinetics of contrast agent (tracer) uptake for accurate perfusion quantification [6]. Moreover, the short scanning

time available for each frame often results in limited spatial and temporal resolution, or poor signal-to-noise ratio (SNR) images. In order to improve the spatial or temporal resolution, one widely used approach is to accelerate the acquisition of each frame through the undersampling of k-space by acquiring only a subset of k-space lines [3,15]. However, the undersampling often results in aliasing artifacts in image space and in the context of perfusion MRI, accurate reconstruction of the complete temporal perfusion signal with its peak and high dynamics becomes an even more challenging task.

In recent years, various approaches have been proposed to solve the reconstruction problem in related dynamic imaging tasks, considering, such as piecewise smoothness in the spatial domain [17], high correlation and sparsity in the temporal domain [3,4,10], sparse representations of local image regions via learned dictionaries [3] and low-rank property of MR sequences in the full spatiotemporal space [10,14,17]. Although these methods allow highly accurate reconstructions from fewer k-space data, the main drawback is that their performance is very sensitive to motion and rapid intensity changes occurring over the duration of image acquisition as encountered in perfusion MRI. In addition, these methods often result in over smooth and blurry image regions that are lacking finer details when the acquired data are contaminated with high noise.

In this paper, we integrate two fundamentally different approaches that both increase the robustness of the reconstruction for perfusion MRI: (i) we enforce pixel-wise local sparsity constraint on the temporal differences that limits the overall dynamic of the perfusion time series, (ii) we enforce patch-wise similarity constraint on the spatio-temporal neighborhoods of whole MR sequence, which provides smooth spatial image regions with less temporal blurring especially when there is significant inter-frame motion and noise. We present the main optimization problem in a joint formal framework and introduce a new proximal splitting strategy that benefits from the weighted-average of proximals – thus, overcome a key limitation of the widely used Fast Composite Splitting Algorithm (FCSA) [9] –, and efficiently solves the joint minimization problem with fast convergence. The proposed method is validated on different types of MR perfusion datasets in comparison with the state-of-the-art methods and extensive experiments demonstrate the superior performance of our method in terms of reconstruction accuracy and accurate estimation of perfusion time profiles from undersampled k-space data even when being presented with high noise levels.

Contributions. Our main contributions are four-fold: (1) We present a new reconstruction scheme which cannot only produce high-quality spatial images for dynamic MRI but also enable to reconstruct the complete temporal signal dynamics for perfusion MRI from undersampled k-space data (Sect. 2). (2) We introduce an efficient proximal-splitting algorithm based on a generalized forward-backward splitting scheme [13]. This algorithm provides fast convergence and can be easily applied to various medical image applications that consider optimization problems where the objective function is the sum of several convex regularization terms (Sect. 3). (3) We demonstrate the efficiency and effectiveness of our method by comparing with state-of-the-art techniques on clinical

datasets (Sect. 4). (4) Our proposed reconstruction model can enable accurate quantification of clinically useful perfusion parameters while accelerating the acquisition through the use of fewer k-space samples.

2 Formulation

Throughout the paper we consider the reconstruction problem only on 2D+t data (i.e., on a single slice followed over time), however the idea presented here can also easily be extended to 3D+t data. We assume that $X^{3D} \in \mathbb{C}^{M \times N \times T}$ is a 2D perfusion image series represented as a spatio-temporal 3D volume. Let $x_t \in \mathbb{C}^{M \times N}$ denote one perfusion MR frame at time t with $M \times N$ pixels, y_t is the corresponding undersampled k-space measurements of x_t, and $\mathbb{T} = \{1, 2, ..., T\}$ is the set of frame number indices in the sequence. The main goal is to recover all x_t's from the collected k-space measurements y_t's. The observation model between x_t and y_t can be mathematically formulated as

$$y_t = R_t(\mathcal{F}_{2D}x_t + \eta) \tag{1}$$

where R_t denotes the undersampling mask to acquire only a subset of k-space, $\mathcal{F}_{2D}$ is the 2D Fourier Transform operator and η is additive Gaussian noise in k-space. We also denote the partial 2D Fourier operator for frame t as $\mathcal{F}_t = R_t\mathcal{F}_{2D}$, and stack the $\mathcal{F}_t$ for all frames of the sequence as $\mathcal{F}_u = \text{diag}\{\mathcal{F}_1, \mathcal{F}_2, .., \mathcal{F}_T\}$.

We propose solving the following optimization problem for the reconstruction of perfusion MR sequences:

$$\hat{X} = \arg\min_X \left\{ \frac{1}{2}\|\mathcal{F}_u X - Y\|_2^2 + \lambda_1 \mathcal{G}_1(X) + \lambda_2 \mathcal{G}_2(X) \right\} \tag{2}$$

where $X \in \mathbb{C}^{MN \times T}$ denotes the whole perfusion MRI sequence and $Y \in \mathbb{C}^{MN \times T}$ represents the collection of all the k-space measurements. λ_1 and λ_2 are the tuning parameters for two regularization terms.

Local ($\mathcal{G}_1$) regularizer: The first regularization term in (2) penalizes the sum of spatial finite differences on the difference images calculated based on a reference for every image frame $x_t \in \mathbb{C}^{M \times N}$, and this term is named as dynamic total variation (TV) [4] and for the whole MR sequence, it can be defined as

$$\mathcal{G}_1(X) = \sum_{t \in \mathbb{T}} \sum_{n=1}^{M \times N} \sqrt{\left(\nabla_x(x_t - \bar{x})_n\right)^2 + \left(\nabla_y(x_t - \bar{x})_n\right)^2} \tag{3}$$

where $\bar{x}$ is the reference image computed by averaging all the frames in MR sequence, ∇_x and ∇_y represent the finite-difference operators along the x and y dimensions, respectively. The intuition behind using dynamic TV over standard TV is that it is better adjusted to the variation in time, and this regularizer serves as a penalty on the overall dynamic of the temporal perfusion signal.

Nonlocal ($\mathcal{G}_2$) regularizer: The second regularization term in (2) penalizes the weighted sum of ℓ_2 norm distances between spatio-temporal neighborhoods (patches) of MR sequence, and this penalty term can be specified by [16]

$$\mathcal{G}_2(X) = \sum_{(\mathrm{p}_x,\mathrm{p}_y,\mathrm{p}_t)\in\Omega} \sum_{(\mathrm{q}_x,\mathrm{q}_y,\mathrm{q}_t)\in\mathcal{N}_\mathrm{p}} w(\mathrm{p},\mathrm{q})\|P_\mathrm{p}(\mathrm{X}^{3D}) - P_\mathrm{q}(\mathrm{X}^{3D})\|_2^2 \tag{4}$$

where $\mathrm{p} = (\mathrm{p}_x,\mathrm{p}_y,\mathrm{p}_t)$ and $\mathrm{q} = (\mathrm{q}_x,\mathrm{q}_y,\mathrm{q}_t)$ are two voxels, and the voxel of interest is $\mathrm{p} \in \Omega$, where $\Omega = [0,M] \times [0,N] \times [0,T]$. The term $P_\mathrm{p}(\mathrm{X}^{3D})$ denotes a spatio-temporal 3D patch of the MR sequence centered at voxel p. We represent $\mathcal{N}_\mathrm{p}$ as a 3D search window around voxel p, and the size of the patch should be smaller than the size of the search window. We simply denote N_p and N_w as the size of a patch and search window, respectively. The weights $w(\mathrm{p},\mathrm{q})$ are determined based on ℓ_2 norm distance between the patches and calculated as

$$w(\mathrm{p},\mathrm{q}) = e^{-\frac{\|P_\mathrm{p}(\mathrm{X}^{3D})-P_\mathrm{q}(\mathrm{X}^{3D})\|_2^2}{h^2}} \tag{5}$$

where h is a smoothing parameter controlling the decay of the exponential function. The use of exponential weighting ensures that a voxel which is more similar to the voxel of interest in terms of Euclidean distance receives higher weight.

This regularizer is capable of exploiting the similarities between patch pairs in adjacent frames and it can enforce smooth solutions in the spatio-temporal neighbourhoods of MR sequence even when there is significant inter-frame motion and high noise introduced during acquisition.

3   Algorithm

To efficiently solve the primal problem (2), we propose to apply a proximal-splitting framework to this problem. Before describing our proximal-splitting based algorithm, we should first give the definition of a proximal map.

Proximal map: Given a continuous convex function $g(x)$ and a scalar $\rho > 0$, the proximal operator associated to convex function g can be defined as [9]

$$prox_\rho(g)(z) := \underset{x\in\mathcal{H}}{\arg\min}\left\{\frac{1}{2\rho}\|x-z\|_2^2 + g(x)\right\} \tag{6}$$

Now we can reformulate the problem (2) as the following denoising problem

$$\hat{X} = \underset{X}{\arg\min}\left\{\frac{1}{2}\|X-X_g\|_2^2 + \rho\lambda_1\mathcal{G}_1(X) + \rho\lambda_2\mathcal{G}_2(X)\right\} \tag{7}$$

Since each of the regularization term in the cost function (2) is convex, the problem (7) can be represented as the proximal map of the sum of two regularization terms as described in Fast Composite Splitting Algorithm (FCSA) [9]

$$\hat{X} = prox_\rho(\lambda_1\mathcal{G}_1 + \lambda_2\mathcal{G}_2)(X_g) \tag{8}$$

The problem (7) admits to a unique solution as given in (8). However, the proximity operator of the sum of two functions is usually intractable. To compute it iteratively, one can adopt an efficient proximal-splitting method to this problem. Proximal-splitting methods are first-order iterative algorithms that solve relatively large-scale optimization problems with several nonsmooth penalties. They operate by splitting the convex objective function to minimize and generating individual subproblems which are evaluated easily via proximal operators.

To solve our main problem in (7), we split the objective function into two individual subproblems that we term $\mathcal{G}_1$-subproblem and $\mathcal{G}_2$-subproblem.

$\mathcal{G}_1$-**subproblem:** The proximal map for this subproblem can be defined as

$$X_{\mathcal{G}_1} = prox_\rho(\lambda_1 \mathcal{G}_1)(X_g) = \arg\min_X \left\{ \frac{1}{2\rho}\|X - X_g\|_2^2 + \lambda_1 \mathcal{G}_1(X) \right\} \quad (9)$$

In order to solve the subproblem (9), we first reformulate it by introducing new variables $d_t = x_t - \bar{x}$ and $d_g^t = X_g^t - \bar{x}$, in this way the problem turns into

$$\hat{d} = \arg\min_d \left\{ \sum_{t\in\mathbb{T}} \left(\frac{1}{2\rho}\|d_t - d_g^t\|_2^2 + \lambda_1\|d_t\|_{TV} \right) \right\} \quad (10)$$

where $d = \{d_1, ..., d_T\}$ and $\|d_t\|_{TV} = \|[Q_1 d_t, Q_2 d_t]\|_{2,1}$, where Q_1 and Q_2 are two $MN \times MN$ first order finite difference matrices in vertical and horizontal directions, and $\ell_{2,1}$ norm is the sum of the ℓ_2 norm of each row of given matrix.

Given a reference image $\bar{x}$, the cost function in (10) can be minimized individually for every frame x_t of MR sequence. This guarantees that the sum of the costs in (10) is also minimized. The cost function can be efficiently minimized by using the fast iteratively reweighted least squares (FIRLS) algorithm [5] based on preconditioned conjugate gradient method. This algorithm enables fast convergence and low computational cost by adopting a new preconditioner which can be accurately approximated using the diagonally dominant structure of the matrix $\mathcal{F}_t^H \mathcal{F}_t$, where H is the conjugate transpose. Once the problem (10) is solved, the estimated solution for problem (9) can be calculated as

$$\hat{X}_{\mathcal{G}_1} = \left[\hat{d}_1 + \bar{x}, \hat{d}_2 + \bar{x}, ..., \hat{d}_T + \bar{x} \right] \quad (11)$$

$\mathcal{G}_2$-**subproblem:** The proximal map for $\mathcal{G}_2$ subproblem can be specified by

$$X_{\mathcal{G}_2} = prox_\rho(\lambda_2 \mathcal{G}_2)(X_g) = \arg\min_X \left\{ \frac{1}{2\rho}\|X - X_g\|_2^2 + \lambda_2 \mathcal{G}_2(X) \right\} \quad (12)$$

The problem (12) can be solved using a two-step alternating minimization scheme in an iterative projections onto convex sets (POCS) framework [11]. In each iteration, the first step involves the projection of image estimate onto the data fidelity term via a steepest descend update and the second step performs the minimization of the neighborhood penalty term on the projected data. The minimization of the penalty function in (12) is equivalent to applying non-local

means (NLM) filter [2] to the projected images. This is mathematically derived in [12] with a single assumption that only one sub-iteration of the penalty term is performed with constant and predetermined weights. The mathematical formulation of a NLM filter is given as [12]

$$\hat{X}(\mathrm{p}_x, \mathrm{p}_y, \mathrm{p}_t) = \frac{\sum_{(\mathrm{q}_x, \mathrm{q}_y, \mathrm{q}_t) \in \mathcal{N}_\mathrm{p}} w(\mathrm{p}, \mathrm{q}) X(\mathrm{q}_x, \mathrm{q}_y, \mathrm{q}_t)}{\sum_{(\mathrm{q}_x, \mathrm{q}_y, \mathrm{q}_t) \in \mathcal{N}_\mathrm{p}} w(\mathrm{p}, \mathrm{q})} \tag{13}$$

We have now iterative solvers for each subproblem $\mathcal{G}_1$ and $\mathcal{G}_2$. In this work, we have developed an efficient algorithm by adopting a generalized forward-backward splitting (GFBS) framework [13] that minimizes the sum of multiple convex functions. Basically, FCSA and GFBS are operator-splitting algorithms and they both use forward-backward schemes. The main difference between GFBS and FCSA is that GFBS enables the use of weighted-average of the outputs of individual proximal mappings for finitely many convex functions, whereas FCSA only applies simple averaging. The weighted-average of the outputs of proximals may practically yield better results depending on the effectiveness of each penalty (regularization) term employed in various applications.

We further accelerate the convergence of the algorithm with an additional acceleration step similar to the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [1]. This step adaptively increases the value of step size parameter (α_k) through iterations and make it sufficiently close to 1. Our proposed reconstruction algorithm is outlined in Algorithm 1. The most computationally expensive step of our algorithm is solving each proximal map. Fortunately, the computation of proximal maps can be done in parallel since there is no dependency between the inputs of proximity operators. All the other steps involve adding and multiplying vectors or scalars, and are thus very cheap in terms of computational complexity. The GFBS method has been shown to converge when $\gamma < 2/L$ if the convex function $f = \frac{1}{2}\|X - X_g\|_2^2$ has a Lipschitz continuous gradient with constant L [13]. We refer the readers to original GFBS paper [13] for details concerning the proof of the convergence of the algorithm.

4   Experiments

Experimental Setup: We evaluate our method on two different types of perfusion MRI datasets. We use three DSC-MRI brain perfusion sequences ($128 \times 128 \times 60$) and one in-vivo breath-hold cardiac perfusion sequence[1] ($192 \times 192 \times 40$) from [4] with normalized intensities. All the perfusion datasets used in the experiments are acquired with full-sampling and the fully-sampled sequences are artificially corrupted by multiplying its corresponding k-space representation with a binary undersampling mask and subsequently adding Gaussian white noise. To simulate undersampling, we retrospectively apply a time-varying variable density Cartesian mask in our experiments (see Fig. 1). The sampling ratio is set to 1/4 for brain sequences and 1/6 for cardiac sequence.

---

[1] Available at: http://web.engr.illinois.edu/~cchen156/SSMRI.html.

Algorithm 1. Proposed algorithm

Input: Undersampled k-space data Y, $\mathcal{F}_u$, λ_1, λ_2

Initialize: $z_1^0 = z_2^0 = \mathcal{F}_u^H Y$, w_1, w_2, $X^0 = \sum_{i=1}^2 w_i z_i^0$, $\alpha_0 = 0.5$, $\gamma = 1$, $k = 0$

while *stopping criteria not met* **do**

> $X_g = X^k - \gamma \mathcal{F}_u^H (\mathcal{F}_u X^k - Y)$;
>
> $z_1^{k+1} = z_1^k + \alpha_k (prox_{\frac{\gamma}{w_1}}(2\lambda_1 \mathcal{G}_1)(X^k + X_g - z_1^k) - X^k)$;
>
> $z_2^{k+1} = z_2^k + \alpha_k (prox_{\frac{\gamma}{w_2}}(2\lambda_2 \mathcal{G}_2)(X^k + X_g - z_2^k) - X^k)$;
>
> $X^{k+1} = w_1 z_1^{k+1} + w_2 z_2^{k+1}$;
>
> $\alpha_{k+1} = 1 + 2(\alpha_k - 1)/(1 + \sqrt{1 + 4(\alpha_k)^2})$;
>
> $k \leftarrow k + 1$;

end

Output: Reconstructed image data X

We compare our method with three state-of-the-art reconstruction methods: the dynamic total variation (DTV) [4], (k,t)-space via low-rank plus sparse prior (kt-RPCA) [14], and fast total variation and nuclear norm regularization (FTVNNR) [17]. To ensure fair comparison, similar to the experiments presented in [3], we empirically fine-tune the optimal regularization parameters for all methods individually for each dataset and depending on noise level. For our proposed method, we specifically set $\lambda_2 = 0.25$ for all noise levels and set $\lambda_1 = 0.025$ for relatively high level noise and $\lambda_1 = 0.001$ for low noise levels. We test the following noise levels and report the results: $\sigma = \{10^{-1}, 5 \times 10^{-2}, 10^{-2}, 5 \times 10^{-3}, 10^{-3}\}$. For the proposed method, we set $N_w = 7 \times 7 \times 7$, $N_p = 5 \times 5 \times 5$, and $w_1 = w_2 = 0.5$ for all sequences. We consider using small cubic neighborhoods for N_w and N_p since large neighborhoods drastically increase the computation time. To reduce the computational burden, we also employ an optimized blockwise version of the non-local means filter that was proposed by Coupé et al. [7] for 3D medical data. We adopt the Peak Signal-to-Noise Ratio (PSNR) as the metric for quantitative evaluation. Instead of directly calculating PSNR on a whole image or 3D sequence, we employ a region-based analysis by calculating the PSNR on randomly selected 100 image blocks (50×50) in 2D frames. This allows us to test for statistical differences using paired t-test when comparing different methods.

Results: Figures 1 and 2 demonstrate a single reconstructed frame of the first and third brain perfusion dataset, respectively, and the estimation of perfusion time profiles averaged over voxels inside a small region of interest. The results in Fig. 1 reveal that kt-RPCA and FTVNNR show quite similar performances, and the DTV yields both better reconstruction and estimation of perfusion signal compared to these two methods. Compared with all three methods, our proposed method can achieve the best reconstruction and very accurate estimation of perfusion time profiles even when the k-space measurements are contaminated with a relatively high level noise ($\sigma = 5 \times 10^{-2}$). The reconstruction results of our method are also statistically significant (p-value $< 10^{-5}$) when compared with all other methods. Moreover, both kt-RPCA and FTVNNR result in over spatial
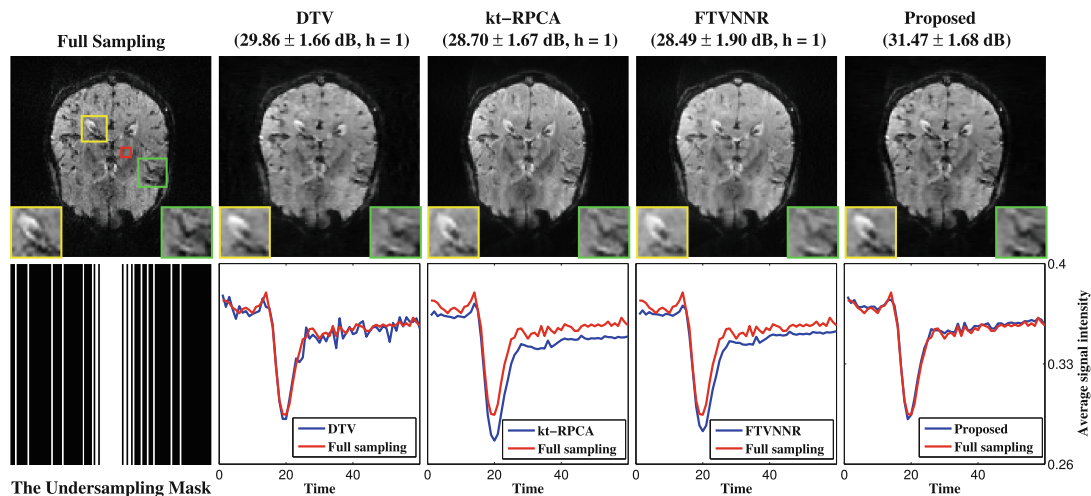
Fig. 1. (Top) Results (mean±std, h-value) of the 22nd frame of the first brain dataset and close-up views of two regions of interest (yellow and green square). h = 1 specifies the statistical significance between the results of proposed and compared method, (Bottom) An exemplary undersampling mask and for each method, estimation of perfusion time profiles averaged over the voxels inside the red square shown in top-left figure. The standard deviation of added Gaussian noise is $\sigma = 5 \times 10^{-2}$. Our method achieves both the best frame-based reconstruction and the most accurate estimation of peaks and temporal pattern of perfusion signal. (Color figure online)
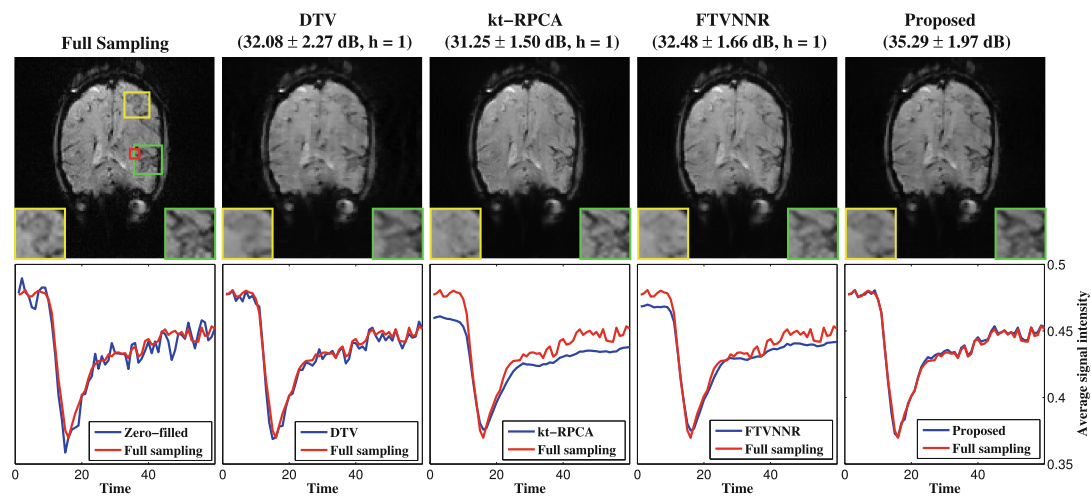
Fig. 2. (Top) Results (mean±std, h-value) of the 15th frame of the third brain dataset and close-up views of two regions of interest (yellow and green square), (Bottom) For each method, estimation of perfusion time profiles averaged over the voxels inside the red square shown in top-left figure. The standard deviation of added Gaussian noise is $\sigma = 10^{-3}$. Our method again achieves both the best frame-based reconstruction and the most accurate estimation of peaks and temporal pattern of perfusion signal. (Color figure online)

smoothing (see close-up views in Fig. 1) and along time as well, which can be clearly seen from smoothening of the perfusion peaks in the third-fourth column of Fig. 1. In contrast, the proposed method reconstructs a perfusion pattern that is in good agreement with the pattern of the fully sampled data (see Fig. 1 bottom fifth column), and produces less blurry image regions and sharper edges. The perfusion time profiles obtained from the third dataset (see Fig. 2 bottom plots) also demonstrate the success of our proposed method. Considering the spatial outputs, when looking at details in close-up views of Fig. 2, the reconstructions obtained by kt-RPCA and FTVNNR are more blurry and thus lacking some finer details in yellow region, whereas the reconstruction obtained by proposed method involves more finer information in yellow region and provides sharper edges in green region.
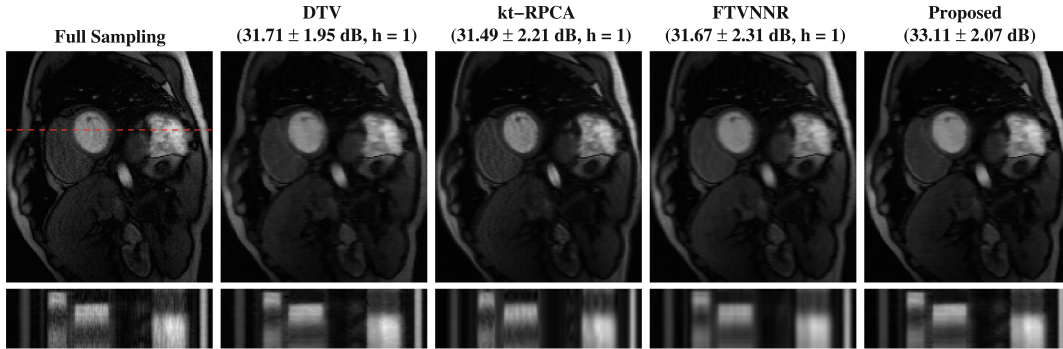
Fig. 3. (Top) Results of the 18th frame of the cardiac dataset with added noise $\sigma = 10^{-2}$, (Bottom) Temporal cross sections by the red dashed line. All methods can produce high quality spatial frames, however, our method yields less noisy and blurry temporal profiles, and the aliasing artifacts are also mostly removed. (Color figure online)

We also validate our method on a cardiac perfusion data from [4] and the results are presented in Fig. 3. All methods here can produce high quality images, however, when looking temporal cross sections at bottom, it can be observed that our method gives less noisy and with lower aliasing artifacts reconstruction on myocardium surface while FTVNNR provides more blurry result. The reason is that our method can utilize both local consistency in temporal differences and nonlocal similarities between spatio-temporal neighborhoods of MR sequence while the FTVNNR does not explicitly exploit sparsity in temporal domain.

Quantitative results of different reconstruction methods on both brain and cardiac perfusion datasets are shown in Fig. 4. Note that the NLM only solves the $\mathcal{G}_2$-subproblem of Sect. 3. From the figure, one can clearly see that our proposed method achieves the highest mean PSNR for all noise levels applied. The running time of all methods on the brain and cardiac datasets is provided in Table 1. Compared with the other three methods, our method needs the highest amount of processing time. However, due to its faster convergence, our method
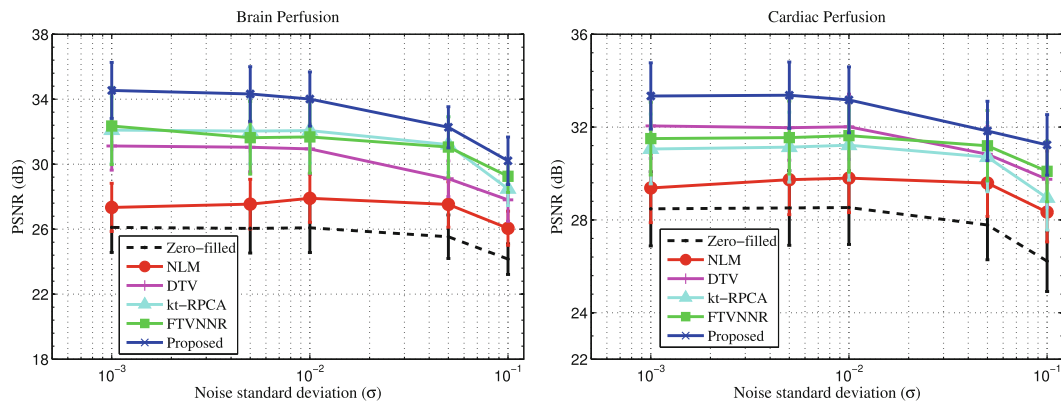
Fig. 4. PSNR results versus noise std (σ) for (left) Brain, (right) Cardiac datasets. Our joint local and nonlocal regularization based method performs the best.

can achieve the best reconstruction accuracy within the first 3–4 iterations on average, which approximately takes 4.5 min for cardiac dataset on a desktop with Intel Xeon CPU E3-1226 v3 Processor.

Table 1. The time cost of different reconstruction methods.

Time (Seconds)	DTV	kt-RPCA	FTVNNR	Proposed
Brain ($128 \times 128 \times 60$)	54.5	194.5	74.3	304.6
Cardiac ($192 \times 192 \times 40$)	46.2	263.9	81.7	278.1

5   Conclusion

We have presented a robust reconstruction model for perfusion MRI, which is based on a joint regularization of pixel-wise and patch-wise spatio-temporal constraints. Numerical experiments validate the efficiency of our method over the state-of-the-art methods in terms of reconstruction accuracy and estimation of perfusion time profiles in varying noise conditions. We also introduce an iterative algorithm that efficiently solves convex optimization problems with mixtures of regularizers. Our algorithm provides fast convergence and can be easily extended to other medical image applications, in particular denoising and super-resolution. The proposed method can be also extended to parallel MR imaging [8] and be applied to multi-coil data. Future work will aim at expanding our work with the fitting of pharmacokinetic models and quantitative analysis of perfusion parameters on 3D+t brain perfusion data using partial k-space measurements.

References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. **2**(1), 183–202 (2009)
2. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 60–65 (2005)
3. Caballero, J., Price, A.N., Rueckert, D., Hajnal, J.: Dictionary learning and time sparsity for dynamic MR data reconstruction. IEEE Trans. Med. Imag. **33**(4), 979–994 (2014)
4. Chen, C., Li, Y., Axel, L., Huang, J.: Real time dynamic MRI with dynamic total variation. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 138–145. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10404-1_18
5. Chen, C., et al.: Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2713–2720 (2014)
6. Conturo, T.E., et al.: Arterial input functions for dynamic susceptibility contrast MRI: requirements and signal options. J. Mag. Reson. Imag. **22**, 697–703 (2005)
7. Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans. Med. Imag. **27**(4), 425–441 (2008)
8. Deshmane, A., Gulani, V., Griswold, M.A., Seiberlich, N.: Parallel MR imaging. J. Mag. Reson. Imag. **36**(1), 55–72 (2012)
9. Huang, J., Zhang, S., Metaxas, D.N.: Efficient MR image reconstruction for compressed MR imaging. Med. Imag. Anal. **15**(5), 670–679 (2011)
10. Lingala, S.G., Hu, Y., DiBella, E., Jacob, M.: Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. IEEE Trans. Med. Imag. **30**(5), 1042–1054 (2011)
11. Marks, R.J.: Alternating projections onto convex sets. In: Jansson, P.A. (ed.) Deconvolution of Images and Spectra, 2nd edn. Academic Press, Orlando (1996)
12. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the nonlocal-means to super-resolution reconstruction. IEEE Trans. Imag. Proc. **18**(1), 36–51 (2009)
13. Raguet, H., Fadili, J., Peyré, G.: A generalized forward-backward splitting. SIAM J. Imag. Sci. **6**(3), 1199–1226 (2013)
14. Trémoulhéac, B., Dikaios, N., Atkinson, D., Arridge, S.R.: Dynamic MR image reconstruction-separation from undersampled (k-t)-space via low-rank plus sparse prior. IEEE Trans. Med. Imag. **33**(8), 1689–1701 (2014)
15. Ulas, C., Gómez, P., Sperl, J.I., Preibisch, C., Menze, B.H.: Spatio-temporal MRI reconstruction by enforcing local and global regularity via dynamic total variation and nuclear norm minimization. In: IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 306–309 (2016)
16. Yang, Z., Jacob, M.: Nonlocal regularization of inverse problems: a unified variational framework. IEEE Trans. Imag. Proc. **22**(8), 3192–3203 (2013)
17. Yao, J., Xu, Z., Huang, X., Huang, J.: Accelerated dynamic MRI reconstruction with total variation and nuclear norm regularization. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9350, pp. 635–642. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24571-3_76

Direct Estimation of Pharmacokinetic Parameters from DCE-MRI using Deep CNN with Forward Physical Model Loss

Abstract: Dynamic contrast enhanced (DCE) MRI is an evolving imaging technique that provides a quantitative measure of pharmacokinetic (PK) parameters in body tissues, in which series of T_1-weighted images are collected following the administration of a paramagnetic contrast agent. Unfortunately, in many applications, conventional clinical DCE-MRI suffers from low spatiotemporal resolution and insufficient volume coverage. In this paper, we propose a novel deep learning based approach to directly estimate the PK parameters from undersampled DCE-MRI data. Specifically, we design a custom

loss function where we incorporate a forward physical model that relates the PK parameters to corrupted image-time series obtained due to subsampling in k-space. This allows the network to directly exploit the knowledge of true contrast agent kinetics in the training phase, and hence provide more accurate restoration of PK parameters. Experiments on clinical brain DCE datasets demonstrate the efficacy of our approach in terms of fidelity of PK parameter reconstruction and significantly faster parameter inference compared to a model-based iterative reconstruction method.

Contributions of thesis author: Algorithm design and implementation, computational experiments, data analysis and interpretation, composition and revision of manuscript.

Direct Estimation of Pharmacokinetic Parameters from DCE-MRI Using Deep CNN with Forward Physical Model Loss

Cagdas Ulas[1]([ ]), Giles Tetteh[1], Michael J. Thrippleton[2], Paul A. Armitage[4],
Stephen D. Makin[2], Joanna M. Wardlaw[2], Mike E. Davies[3],
and Bjoern H. Menze[1]

[1] Department of Computer Science, Technische Universität München,
Munich, Germany
cagdas.ulas@tum.de
[2] Department of Neuroimaging Sciences, University of Edinburgh, Edinburgh, UK
[3] Institute for Digital Communications, University of Edinburgh, Edinburgh, UK
[4] Department of Cardiovascular Sciences, University of Sheffield, Sheffield, UK

Abstract. Dynamic contrast-enhanced (DCE) MRI is an evolving imaging technique that provides a quantitative measure of pharmacokinetic (PK) parameters in body tissues, in which series of T_1-weighted images are collected following the administration of a paramagnetic contrast agent. Unfortunately, in many applications, conventional clinical DCE-MRI suffers from low spatiotemporal resolution and insufficient volume coverage. In this paper, we propose a novel deep learning based approach to directly estimate the PK parameters from undersampled DCE-MRI data. Specifically, we design a custom loss function where we incorporate a forward physical model that relates the PK parameters to corrupted image-time series obtained due to subsampling in k-space. This allows the network to directly exploit the knowledge of true contrast agent kinetics in the training phase, and hence provide more accurate restoration of PK parameters. Experiments on clinical brain DCE datasets demonstrate the efficacy of our approach in terms of fidelity of PK parameter reconstruction and significantly faster parameter inference compared to a model-based iterative reconstruction method.

1 Introduction

Dynamic contrast-enhanced (DCE) MRI involves the administration of a T_1-shortening Gadolinium-based contrast agent (CA), followed by the acquisition of successive T_1-weighted images as the contrast bolus enters and subsequently leaves the organ [9]. In DCE-MRI, changes in CA concentration are derived from changes in signal intensity over time, then regressed to estimate pharmacokinetic (PK) parameters related to vascular permeability and tissue perfusion [6]. Since perfusion and permeability are typically affected in the presence of vascular and

cellular irregularities, DCE imaging has been considered as a promising tool for clinical diagnostics of brain tumours, multiple sclerosis lesions, and neurological disorders where disruption of blood-brain barrier (BBB) occurs [4,7].

Despite its effectiveness in quantitative assessment of microvascular properties, conventional DCE-MRI is challenged by suboptimal image acquisition that severely restricts the spatiotemporal resolution and volume coverage [2,3]. The shortest possible scanning time often leads to limited spatial resolution hampering detection of small image features and accurate tumor boundaries. Low temporal resolution hinders accurate fitting of PK parameters. Furthermore, volume coverage is usually inadequate to cover the known pathology, for instance in the case multiple metastatic lesions [3]. Facing such severe constraints, DCE imaging can significantly benefit from undersampled acquisitions.

So far, existing works in [2,6,11] have proposed compressed sensing and parallel imaging based reconstruction schemes to accelerate DCE-MRI acquisitions, mainly targeting to achieve better spatial resolution and volume coverage while retaining the same temporal resolution. These methods are referred to as indirect methods [3] because they are based on the reconstruction of dynamic DCE image series first, followed by a separate step for fitting the PK parameters on a voxel-by-voxel level using a tracer kinetic model [9]. More recently, a model-based direct reconstruction model [3] has been proposed to directly estimate PK parameters from undersampled (k, t) space data. The direct reconstruction method generally poses the estimation of PK maps as an error minimization problem. This approach has been shown to produce superior PK parameter maps and allows for higher acceleration compared to indirect methods. However, the main drawback of this method is that parameter reconstruction of an entire volume requires considerably high computation time.

Motivated by the recent advances of deep learning in medical imaging, in this paper, we present a novel deep learning based approach to directly estimate PK parameters from undersampled DCE-MRI data. First, our proposed network takes the corrupted image-time series as input and *residual* parameter maps, which represent deviations from a kinetic model fitting on fully-sampled image-time series, as output, and aims at learning a nonlinear mapping between them. Our motivation for learning the *residual* PK maps is based on the observation that residual maps are more sparse and topologically less complex compared to target parameter maps. Second, we propose the *forward physical model loss*, a custom loss function in which we exploit the physical relation between true contrast agent kinetics and measured time-resolved DCE signals when training our network. Third, we validate our method experimentally on human *in vivo* brain DCE-MRI dataset. We demonstrate the superior performance of our method in terms of parameter reconstruction accuracy and significantly faster estimation of parameters during testing, taking approximately 1.5 s on an entire 3D test volume. To the best of our knowledge, we present the first work leveraging the machine learning algorithms – specifically deep learning – to directly estimate PK parameters from undersampled DCE-MRI time-series.

2 Methods

We treat the parameter inference from undersampled data in DCE imaging as a mapping problem between the corrupted intensity-time series and *residual* parameter maps where the underlying mapping is learned using deep convolutional neural networks (CNNs). We provide a summary of general tracer kinetic models applied in DCE-MRI in Sect. 2.1, formulate the forward physical model relating the PK parameters to undersampled data in Sect. 2.2, finally describe our proposed deep learning methodology for PK parameter inference in Sect. 2.3.
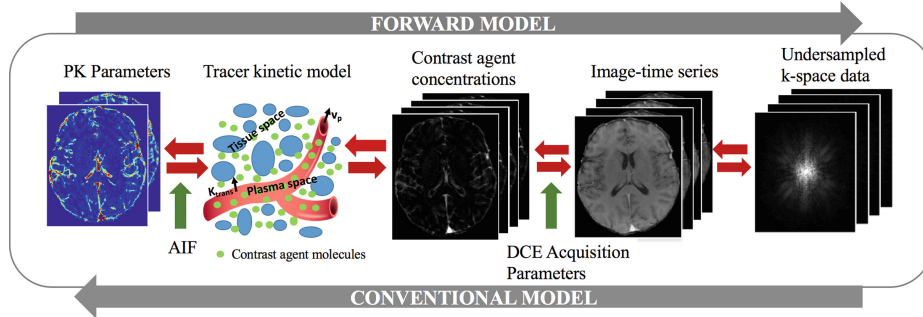
Fig. 1. Computational steps in the forward model and the conventional pipeline of PK parameter estimation in DCE-MRI.

2.1 Tracer Kinetic Modeling in DCE-MRI

Tracer kinetic modeling aims at providing a link between the tissue signal enhancement and the physiological or so-called pharmacokinetic parameters, including the fractional plasma volume (v_{p}), the fractional interstitial volume (v_{e}), and the volume transfer rate (K^{trans}) at which contrast agent (CA) is delivered to the extravascular extracellular space (EES). One of the well-established tracer kinetic models is known as Patlak model [8]. This model describes a highly perfused two compartment tissue, ignoring backflux from the EES into the blood plasma compartment. The CA concentration in the tissues is determined by,

$$C(\mathbf{r}, t) = v_p(\mathbf{r})C_p(t) + K^{\mathrm{trans}}(\mathbf{r}) \int_0^t C_p(\tau)d\tau, \tag{1}$$

where $\mathbf{r} \in (x, y, z)$ represent image domain spatial coordinates, $C(\mathbf{r}, t)$ is the CA concentration over time, and $C_p(t)$ denotes the arterial input function (AIF) which is usually measured from voxels in a feeding artery.

In this work, we specifically employ the Patlak model for tracer pharmacokinetic modeling and estimation of ground truth tissue parameters. This model is a perfect match for our DCE dataset because it is often applied when the temporal resolution is too low to measure the cerebral blood flow, and it has been commonly used to measure the BBB leakage with DCE-MRI in acute brain stroke and dementia [4,9]. An attractive feature of Patlak model is that the model equation in (1) can be linearized and fitted using linear least squares which has a closed-form solution, hence parameter estimation is fast [9].

2.2   Forward Physical Model: From PK Parameters to Undersampled Data

Figure 1 depicts the conventional and forward model approaches relating the PK parameter estimation to undersampled or fully-sampled k-space data, and vice versa. For direct estimation of PK parameters from the measured k-space data, as proposed in [1,3], a forward model can be formulated by inverting the steps in the conventional model as follows:

1. Given the sets of PK parameter pairs $(K^{\mathrm{trans}}(\mathbf{r}), v_{\mathrm{p}}(\mathbf{r}))$ and arterial input function $C_p(t)$, CA concentration curves over time $C(\mathbf{r}, t)$ are estimated using the Patlak model equation in (1).
2. Dynamic DCE image series $S(\mathbf{r}, t)$ are converted to $C(\mathbf{r}, t)$ through the steady-state spoiled gradient echo (SGPR) signal equation [3], given by

$$S(\mathbf{r}, t) = \frac{M_0(\mathbf{r})\sin\alpha(1 - e^{-(K+L)})}{1 - \cos\alpha e^{-(K+L)}} + \left(S(\mathbf{r}, 0) - \frac{M_0(\mathbf{r})\sin\alpha(1 - e^{-K})}{1 - \cos\alpha e^{-K}}\right) \quad (2)$$

where $K = T_{\mathrm{R}}/T_{10}(\mathbf{r})$, $L = r_1 C(\mathbf{r}, t)T_{\mathrm{R}}$, T_{R} is the repetition time, α is the flip angle, r_1 is the contrast agent relaxivity taken as 4.2 s^{-1}mM^{-1}, $S(\mathbf{r}, 0)$ is the baseline (pre-contrast) image intensity, and $T_{10}(\mathbf{r})$ and $M_0(\mathbf{r})$ are respectively the T_1 relaxation and equilibrium longitudinal magnetization that are calculated from a pre-contrast T_1 mapping acquisition.
3. The undersampled raw (k, t)-space data $S(\mathbf{k}, t)$ can be related to $S(\mathbf{r}, t)$ for a single-coil data by an undersampling fast Fourier transform (FFT), F_u,

$$S(\mathbf{k}, t) = F_u S(\mathbf{r}, t), \quad (3)$$

where $\mathbf{k} \in (k_x, k_y, k_z)$ represents k-space coordinates.
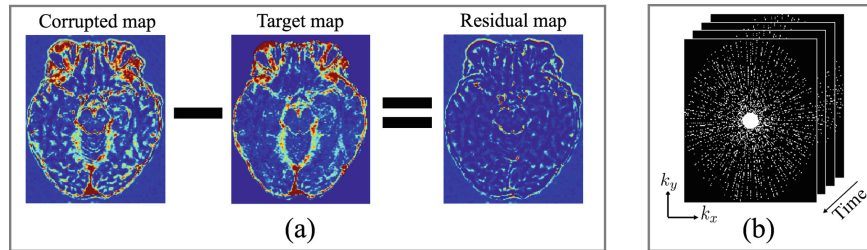
Fig. 2. (a) The relation between a corrupted (θ_u), target (θ_t) and residual (θ_r) PK maps, (b) Exemplary golden-angle sampling scheme in the k_x-k_y plane through time.

By simply integrating the three computation steps in (1–3), we can form a single function f_m modeling the signal evolution in (k-t) space given the PK maps $\theta = \{K^{\mathrm{trans}}(\mathbf{r}), v_{\mathrm{p}}(\mathbf{r})\}$, as $S(\mathbf{k}, t) = f_m(\theta; \boldsymbol{\xi})$, where $\boldsymbol{\xi}$ denotes all the predetermined acquisition parameters as mentioned above.

Given the undersampled (k,t)-space data $S(\mathbf{k}, t)$, the corrupted image series $S_u(\mathbf{r}, t)$ can be obtained by applying IFFT to $S(\mathbf{k}, t)$, i.e. $S_u(\mathbf{r}, t) = F_u^{\mathsf{T}} S(\mathbf{k}, t)$. We further define a new function $\tilde{\boldsymbol{f}}_m$ that integrates only the first two computation steps (1–2) to compute the dynamic DCE image series. We will incorporate $\tilde{\boldsymbol{f}}_m$ in our custom loss function that will be explained in the following section.

2.3 PK Parameter Inference via Forward Physical Model Loss

Formulation. We hypothesize that a direct inversion between corrupted PK parameter maps θ_u and $S_u(\mathbf{r}, t)$ is available through forward model, i.e., $S_u(\mathbf{r}, t) = \tilde{\boldsymbol{f}}_m(\theta_u)$. However, this cannot provide yet sufficiently accurate estimate of target parameter maps θ_t obtained from fully-sampled data $S(\mathbf{r}, t)$. To this end, we estimate a correction or residual map θ_r from the available signal $S_u(\mathbf{r}, t)$ satisfying $\theta_r = \theta_u - \theta_t$. As shown in Fig. 2-(a), we observe that *residual* PK maps involve more sparse representations and exhibit spatially less varying structures inside the brain. The task of learning a residual mapping was shown to be much easier and effective than the original mapping [10]. Following the same approach, we adopt the residual learning strategy using deep CNNs. Our CNN is trained to learn a mapping between $S_u(\mathbf{r}, t)$ and θ_r to output an estimate of residual maps $\tilde{\theta}_r$; $\tilde{\theta}_r = \mathcal{R}(S_u(\mathbf{r}, t)|\mathbf{W})$, where $\mathcal{R}$ represents the forward mapping of the CNN parameterised by $\mathbf{W}$. The final parameter estimate is obtained via $\tilde{\theta}_t = \theta_u - \tilde{\theta}_r$.

Loss Function. We simultaneously seek the signal belonging to the corrected model estimates to be sufficiently close to true signal, i.e., $\tilde{\boldsymbol{f}}_m(\tilde{\theta}_t) \approx S(\mathbf{r}, t)$. Therefore, we design a custom loss function which requires solving the forward model in every iteration of the network training. We refer the resulting loss as *forward physical model loss*. Given a set of training samples $\mathcal{D}$ of input-output pairs $(S_u(\mathbf{r}, t), \theta_r)$, we train a CNN model that minimizes the following loss,

$$\mathcal{L}(\mathbf{W}) = \sum_{(S_u(\mathbf{r},t), \theta_r) \in \mathcal{D}} \lambda \|\theta_r - \tilde{\theta}_r\|_2^2 \; + \; (1 - \lambda)\|S(\mathbf{r}, t) - \tilde{\boldsymbol{f}}_m(\theta_u - \tilde{\theta}_r; \boldsymbol{\xi})\|_2^2, \quad (4)$$

where λ is a regularization parameter balancing the trade-off between the fidelity of the parameter and signal reconstruction. We emphasize that the second term in (4) allows the network to intrinsically exploit the underlying contrast agent kinetics in training phase.

Network Architecture. Figure 3 illustrates our network architecture. The network takes a 4D image-time series as input, where time frames are stacked as input channels. The first convolutional layer applies 3D filters to each channel individually to extract low-level temporal features which are aggregated over frames via learned filter weights to produce a single output per voxel. Following the first layer, inspired by the work on brain segmentation [5], our network consists of parallel dual pathways to efficiently capture multi-scale information. The

local pathway at the top focuses on extracting details from the local vicinity while the global pathway at the bottom is designed to incorporate more contextual global information. The global pathway consists of 4 dilated convolutional layers with dilation factors of $2, 4, 8, 16$, implying increased receptive field sizes. The filter size of each convolutional layer including dilated convolutions is $3 \times 3 \times 3$, and the rectified linear units (ReLU) activation is applied after each convolution. Local and global pathways are then concatenated to form a multi-scale feature set. Following this, 2 fully-connected layers are used to determine the best possible feature combination that can accurately map the input to output of the network. Finally, the last layer outputs the estimated residual maps.

3   Experiments and Results

Datasets. We perform experiments on fully-sampled DCE-MRI datasets acquired from three mild ischaemic stroke patients. DCE image series were acquired using a 1.5 T clinical scanner with a 3D T1W spoiled gradient echo sequence (TR/TE = 8.24/3.1 ms, flip angle = 12°, FOV = 24×24 cm, matrix = 256×192, slice thickness = 4 mm, 73 s temporal resolution, 21 dynamics). An intravenous bolus injection of 0.1 mmol/kg of gadoterate meglumine (Gd-DOTA) was administered simultaneously. The total acquisition time for DCE-MRI was approximately 24 minutes. Two pre-contrast acquisitions were carried out at flip angles of 2° and 12° to calculate pre-contrast longitudinal relaxation times.
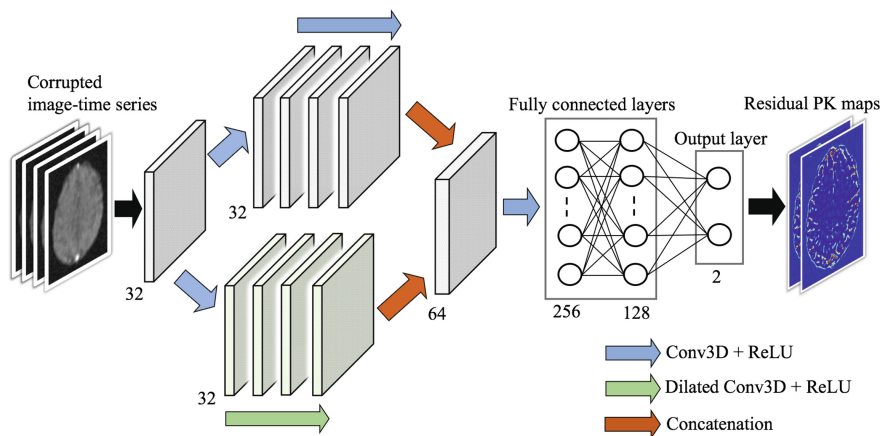
Fig. 3. The network architecture used for the estimation of residual PK maps. The number of filters and output nodes are provided at the bottom of each layer.

Preprocessing. Undersampling was retrospectively applied to the fully-sampled data in the k_x-k_y plane using a randomized golden-angle sampling pattern [12] over time (see Fig. 2-(b)) with a 10-fold undersampling factor. The pre-contrast first frame was fully sampled. Due to the low temporal resolution of our data, we estimated subject-specific vascular input functions (VIFs) extracted by averaging a few voxels located on the superior sagittal sinus where the inflow

artefact was reduced compared to a feeding artery [4]. Data augmentation was employed by applying rigid transformations on image slices. We generated random 2D+t undersampling masks to be applied on the images of different orientations. This allows the network to learn diverse patterns of aliasing artifacts. All the subject's data required for network training/testing were divided into non-overlapping 3D blocks of size $52 \times 52 \times 33$, resulting in 64 blocks per subject.

Experimental Setup. All experiments were performed in a leave-one-subject-out fashion. The networks were trained using the Adam optimizer with a learning rate of 10^{-3} (using a decay rate of 10^{-4}) for 300 epochs and mini-batch size of 4. To demonstrate the advantage of the proposed method, we compare it with the state-of-the-art model-based iterative parameter reconstruction method using the MATLAB implementation provided by the authors [3]. We use the concordance correlation coefficient (CCC) and structured similarity metric (SSIM) metrics to quantitatively assess the PK parameter reconstruction, and peak signal-to-noise ratio (PSNR) metric to assess the image reconstruction. Experiments were run on a NVIDIA GeForce Titan Xp GPU with 12 GB RAM.

Results. Figure 4 shows the qualitative PK parameter reconstructions obtained from different methods using 10-fold undersampling. The results indicate that CNN-$\lambda = 0.5$ incorporating two loss terms simultaneously produces better maps and considerably higher SSIM score calculated with respect to fully-sampled PK maps. The model-based iterative reconstruction yields the PK maps where the artifacts caused by undersampling are still observable. In Fig. 5 we present the exemplary reconstructed images obtained by applying the operation $\tilde{\boldsymbol{f}}_m$ to the estimated PK maps. All the reconstruction approaches result in high quality images, however, the model-based reconstruction can better preserve the finer details. Unfortunately, our fully-sampled data suffer from Gibbs artifacts appearing as multiple parallel lines throughout the image. As marked by white arrows, our CNN method can significantly suppress these artifacts whereas they still appear in the image obtained by model-based iterative reconstruction. Finally, Fig. 6 demonstrates the quantitative results of parameter estimation and image reconstruction. The highest CCC and SSIM values for parameter estimation are achieved by our CNN model when both loss terms are incorporated with $\lambda = 0.3$ and $\lambda = 0.5$, yielding an average score of 0.88 and 0.92, respectively. The difference is statistically significant for both CCC ($p = 0.017$) and SSIM ($p = 0.0086$) when compared against model-based reconstruction. The model-based reconstruction performs the highest PSNR for image reconstruction, where it is followed by the proposed CNN with $\lambda = 0.3$. The difference between them is statistically significant with $p \ll 0.05$. The PSNR also shows a decreasing trend with increasing λ as expected.

We emphasize that the parameter inference of our method on a 3D test volume takes around 1.5 s while the model-based method requires around 95 min to reconstruct the same volume, enabling $\approx 4 \times 10^3$ faster computation.
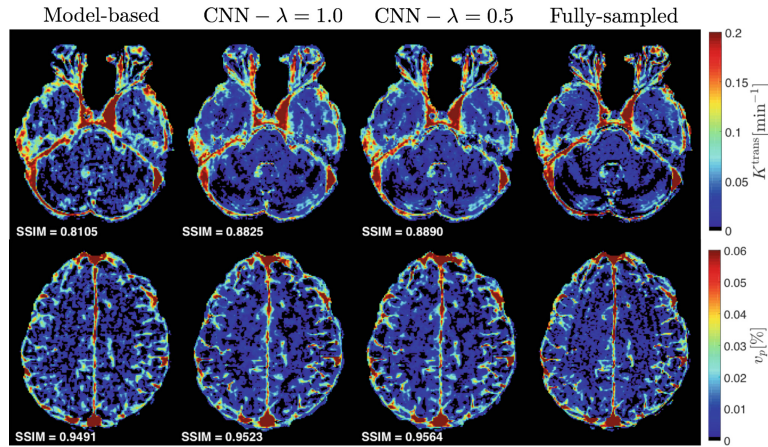
Fig. 4. Reconstructed PK parameter maps of two exemplary slices of a test subject with a 10-fold undersampling. Brain masks are applied to estimated maps. Our CNN model incorporating both loss terms ($\lambda = 0.5$) achieves the best paramater estimates. The resulting SSIM values are provided at the bottom-left corner of each map.
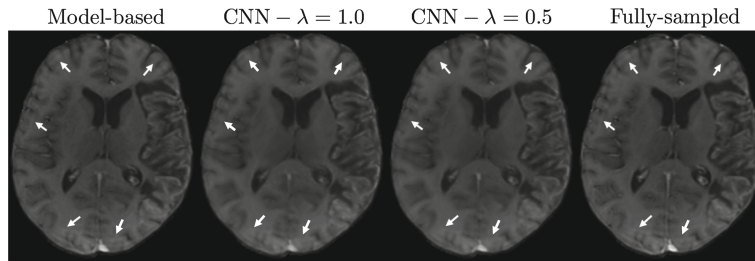
Fig. 5. Visual comparison of the image reconstruction results of an examplary DCE slice. White arrows indicate a few regions where the Gibbs artifacts are observable. Our CNN model with both $\lambda = 0.5$ and 1.0 can significantly suppress the artifacts appearing in fully-sampled image and model-based reconstruction as well.
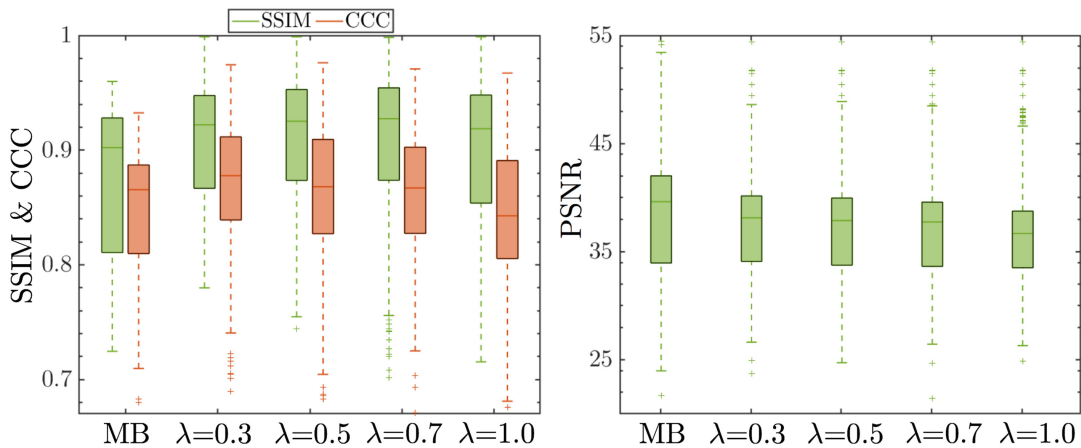
Fig. 6. Parameter estimation (SSIM & CCC) and image reconstruction (PSNR) performances calculated on all test slices for model-based (MB) reconstruction method and our proposed CNN model with different λ settings.

4   Conclusion

We present a novel deep learning based framework for direct estimation of PK parameter maps from undersampled DCE image-time series. Specifically, we design a *forward physical model loss* function through which we exploit the physical model relating the contrast agent kinetics to the time-resolved DCE signals. Moreover, we utilize the residual learning strategy in our problem formulation. The experiments demonstrate that our proposed method can outperform the state-of-the-art model-based reconstruction method, and allow almost instantaneous inference of the PK parameters in the clinical workflow of DCE-MRI.

References

1. Fang, R., et al.: Direct estimation of permeability maps for low-dose CT perfusion. In: IEEE ISBI, pp. 739–742, April 2016
2. Guo, Y., et al.: High-resolution whole-brain DCE-MRI using constrained reconstruction. Med. Phys. **43**(5), 2013–2023 (2016)
3. Guo, Y., et al.: Direct estimation of tracer-kinetic parameter maps from highly undersampled brain dynamic contrast enhanced MRI. MRM **78**(4), 1566–1578 (2017)
4. Heye, A.K., et al.: Tracer kinetic modelling for DCE-MRI quantification of subtle bloodbrain barrier permeability. NeuroImage **125**, 446–455 (2016)
5. Kamnitsas, K., et al.: Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
6. Lebel, R.M., et al.: Highly accelerated dynamic contrast enhanced imaging. MRM **71**(2), 635–644 (2014)
7. O'Connor, J.P.B., et al.: Dynamic contrast-enhanced MRI in clinical trials of anti-vascular therapies. Nat. Rev. Clin. Oncol. **9**(3), 167–77 (2012)
8. Patlak, C.S., et al.: Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. J. Cereb. Blood Flow Metab. **3**(1), 1–7 (1983)
9. Sourbron, S.P., Buckley, D.L.: Classic models for dynamic contrast-enhanced MRI. NMR Biomed. **26**(8), 1004–1027 (2013)
10. Zhang, K., et al.: Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017)
11. Zhang, T., et al.: Fast pediatric 3d free-breathing abdominal dynamic contrast enhanced MRI with high spatiotemporal resolution. JMRI **41**(2), 460–473 (2015)
12. Zhu, Y., et al.: GOCART: GOlden-angle CArtesian randomized time-resolved 3D MRI. Magn. Reson. Imag. **34**(7), 940–950 (2016)

CNNs for Pharmacokinetic Parameter Inference in Stroke DCE-MRI

C. Ulas, D. Das, M. J. Thrippleton, M. d. C. Valdés Hernández, P. A. Armitage, S. D. Makin, J. M. Wardlaw, and B. H. Menze. "Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI." in: *Frontiers in Neurology* 9 (2019), p. 1147. DOI: 10.3389/fneur.2018.01147

Abstract: *Background and Purpose:* The T1-weighted dynamic contrast enhanced (DCE)-MRI is an imaging technique that provides a quantitative measure of pharmacokinetic (PK) parameters characterizing microvasculature of tissues. For the present study, we propose a new machine learning (ML) based approach to directly estimate the PK parameters from the acquired DCE-MRI image-time series that is both more robust and faster than conventional model fitting.

Materials and Methods: We specifically utilize deep convolutional neural networks (CNNs) to learn the mapping between the image-time series and corresponding PK parameters. DCE-MRI datasets acquired from 15 patients with clinically evident mild ischaemic stroke were used in the experiments. Training and testing were carried out based on leave-one-patient-out cross-validation. The parameter estimates obtained by the proposed CNN model were compared against the two tracer kinetic models: (1) Patlak model, (2)

Extended Tofts model, where the estimation of model parameters is done via voxelwise linear and nonlinear least squares (NLLS) fitting respectively.

Results: The trained CNN model is able to yield PK parameters which can better discriminate different brain tissues, including stroke regions. The results also demonstrate that the model generalizes well to new cases even if a subject specific arterial input function (AIF) is not available for the new data.

Conclusion: A ML-based model can be used for direct inference of the PK parameters from DCE image series. This method may allow fast and robust parameter inference in population DCE studies. Parameter inference on a 3D volume-time series takes only a few seconds on a GPU machine, which is significantly faster compared to conventional nonlinear least squares (NLLS) fitting.

Contributions of thesis author: Development and implementation of the methodology, experimental design, interpretation of the results, composition and revision of manuscript.

Check for updates

Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI

Cagdas Ulas[1]*, Dhritiman Das[1,2], Michael J. Thrippleton[3], Maria del C. Valdés Hernández[3], Paul A. Armitage[4], Stephen D. Makin[3], Joanna M. Wardlaw[3] and Bjoern H. Menze[1,5]

[1] Department of Computer Science, Technische Universität München, Munich, Germany, [2] GE Global Research, Munich, Germany, [3] Department of Neuroimaging Sciences, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom, [4] Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, United Kingdom, [5] Institute of Advanced Study, Technische Universität München, Munich, Germany

Background and Purpose: The T1-weighted dynamic contrast enhanced (DCE)-MRI is an imaging technique that provides a quantitative measure of pharmacokinetic (PK) parameters characterizing microvasculature of tissues. For the present study, we propose a new machine learning (ML) based approach to directly estimate the PK parameters from the acquired DCE-MRI image-time series that is both more robust and faster than conventional model fitting.

Materials and Methods: We specifically utilize deep convolutional neural networks (CNNs) to learn the mapping between the image-time series and corresponding PK parameters. DCE-MRI datasets acquired from 15 patients with clinically evident mild ischaemic stroke were used in the experiments. Training and testing were carried out based on leave-one-patient-out cross-validation. The parameter estimates obtained by the proposed CNN model were compared against the two tracer kinetic models: (1) Patlak model, (2) Extended Tofts model, where the estimation of model parameters is done via voxelwise linear and nonlinear least squares fitting respectively.

Results: The trained CNN model is able to yield PK parameters which can better discriminate different brain tissues, including stroke regions. The results also demonstrate that the model generalizes well to new cases even if a subject specific arterial input function (AIF) is not available for the new data.

Conclusion: A ML-based model can be used for direct inference of the PK parameters from DCE image series. This method may allow fast and robust parameter inference in population DCE studies. Parameter inference on a 3D volume-time series takes only a few seconds on a GPU machine, which is significantly faster compared to conventional non-linear least squares fitting.

Keywords: dynamic contrast enhanced MRI, pharmacokinetic parameter inference, convolutional neural networks, ischaemic stroke, tracer kinetic modeling, contrast agent concentration, loss function

1. INTRODUCTION

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is an effective dynamic imaging technique that can be used to study microvascular structure *in vivo* by tracking the diffusion of a paramagnetic contrast agent such as gadopentate dimeglumine (Gd-DTPA) over time (1). By collecting a series of T_1-weighted MR images at intervals of a few seconds, the uptake and washout of the administered contrast agent can be observed in the imaged tissue, resulting in characteristic intensity-time curves across different tissues (2). Vascular and cellular regularities in human body usually have a strong impact on the local vascular perfusion and permeability. To this end, DCE imaging has been used as a promising tool for clinical diagnostics of brain tumors, multiple sclerosis lesions, and several neurological disorders that lead to disruption and breakdown of blood-brain barrier (BBB) (3–6). In DCE-MRI, changes in contrast agent concentration are determined from changes in signal intensity over time, and then regressed through the use of tracer kinetic (TK) models to estimate pharmacokinetic (PK) parameters which characterizes the vascular permeability and tissue perfusion (7, 8).

One of the key limitations of TK modeling methods is that they are simply based on the fitting of voxelwise PK parameters to contrast agent concentration-time curves (9). The fitting is usually performed using a nonlinear least squares (NLS) approach. However, the acquired voxelwise concentration-time curves are generally very noisy and involve only a small number of sampling points, hence the model fitting may yield parameter estimates with large variance as well as considerable bias (see **Figure 1** for an exemplary representation of this limitation). Moreover, an iterative NLS solver may converge to erroneous solutions since the NLS objective is not convex and can have multiple local minima (10). Another major drawback is that the voxelwise model fitting is computationally demanding considering the thousands of voxels in a single MR slice (11). More sophisticated approaches (10, 12) were also proposed based on Bayesian theory of statistical inference of the DCE parameters for the fitting of nonlinear models. Unlike the standard NLS regression, these approaches exploit the spatial information of the neighboring voxels and provide reduce variability of parameters in local homogeneous regions. However, the bottleneck is their drastically increased computation time, usually taking hours for the estimation of parameters on a single DCE scan.

Machine learning (ML) methods have been extensively used in the medical imaging community for several tasks (13) such as parameter estimation, disease classification, segmentation, so on. Recently, a random forest regression based method (14) was proposed to estimate accurate spectral parameters in MR spectroscopy. Deep learning methods (15–17) have recently gained large popularity and achieved predominantly state-of-the-art results in the medical imaging field including various image-to-image translation tasks (18–20). A deep neural network based approach for perfusion parameter estimation (21) was first proposed for dynamic susceptibility contrast (DSC) MRI without requiring a standard deconvolution process.

To alleviate the aforementioned limitations in DCE-MRI, we present a direct and fast PK parameter estimation method which introduces several concepts from machine learning. Our proposed approach can directly infer the PK parameters from the observed signal intensity over time. In order to achieve this, we first train a deep convolutional neural network (CNN) to learn the underlying mapping – or relation – between intensity image-time series and PK parameters using a large training data consisting of millions of voxels taken from the brain DCE dataset. In our method, the target PK parameters used in training step can be either estimated by any existing tracer kinetic models, or can be defined with reference values depending on a specific biomarker or disease that has been built on one specific type of model. Our method can intrinsically provide the following advantages over the conventional model fitting based parameter estimation approaches:

- The proposed method can directly estimate the corresponding physiological perfusion parameters when only observed signal intensities over time given, which eliminates several intermediate computation steps of the conventional pipeline as illustrated in **Figure 2**.
- Our method serves as a high-level parameter estimation model such that we can train a network from which we expect to yield parameter estimates as close as the target values that are obtained using any optimization approach, e.g., standard NLS fitting, regularized Bayesian estimation methods, etc.
- Due to its strong generalization ability, this method shows increased robustness to signal noise and outliers, and it can significantly mitigate the effect of irregularity and discontinuity problem which is quite apparent in the parameter maps estimated by conventional NLS fitting.
- The parameter estimates obtained by the proposed approach yields improved statistically significant differences between different tissue types, which can ultimately allow better discrimination of normal and pathological regions in stroke analysis.
- Compared to conventional fitting methods, the PK parameter inference with our ML based approach is computationally faster, taking only seconds on an entire 3D DCE-MRI volume.

2. MATERIALS AND METHODS
2.1. Dataset and Preprocessing
2.1.1. Patients
Fifteen patients were recruited for this study. The patient cohort presents first clinically evident mild (i.e., expected to be non-disabling) ischaemic stroke from the local stroke service. The patients were over 18 years old and had a definite diagnosis of ischaemic stroke. They were able to consent themselves, had an MRI scan at diagnosis and were medically stable enough to return for a DCE-MRI scan at between 1 and 3 months post-stroke and a follow-up after 1 year. All patients underwent clinical assessment by a stroke physician, diagnostic MR imaging and cognitive testing at presentation. An expert panel of stroke physicians and neuro-radiologists assessed each case in order to confirm the diagnosis of ischaemic stroke and classify the ischaemic stroke
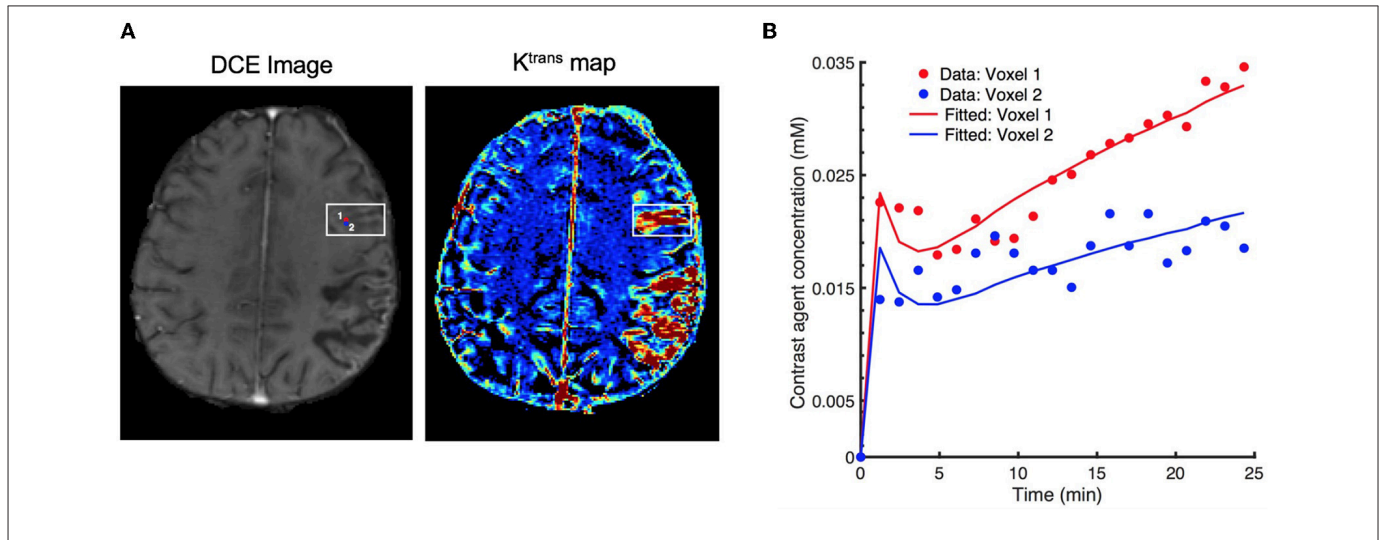
FIGURE 1 | Effect of signal noise on the resulting parameters with conventional fitting models. **(A)** an examplary DCE image (left) displaying two neighboring voxels (marked by red and blue circles) in the stroke region, and the corresponding Ktrans maps (right), **(B)** resulting fitted contrast agent concentration curves for these two voxels using Extended Tofts model. Although the neighboring voxels are spatially very close to each other (only 1-pixel away), the observed concentration data are different due to the excessive signal noise. Eventually, there is a substantial difference in the fitted concentration curves and parameter values ($K^{trans} = 6.18 \times 10^{-3}$ min^{-1} for voxel 1, and $K^{trans} = 2.48 \times 10^{-3}$ min^{-1} for voxel 2).
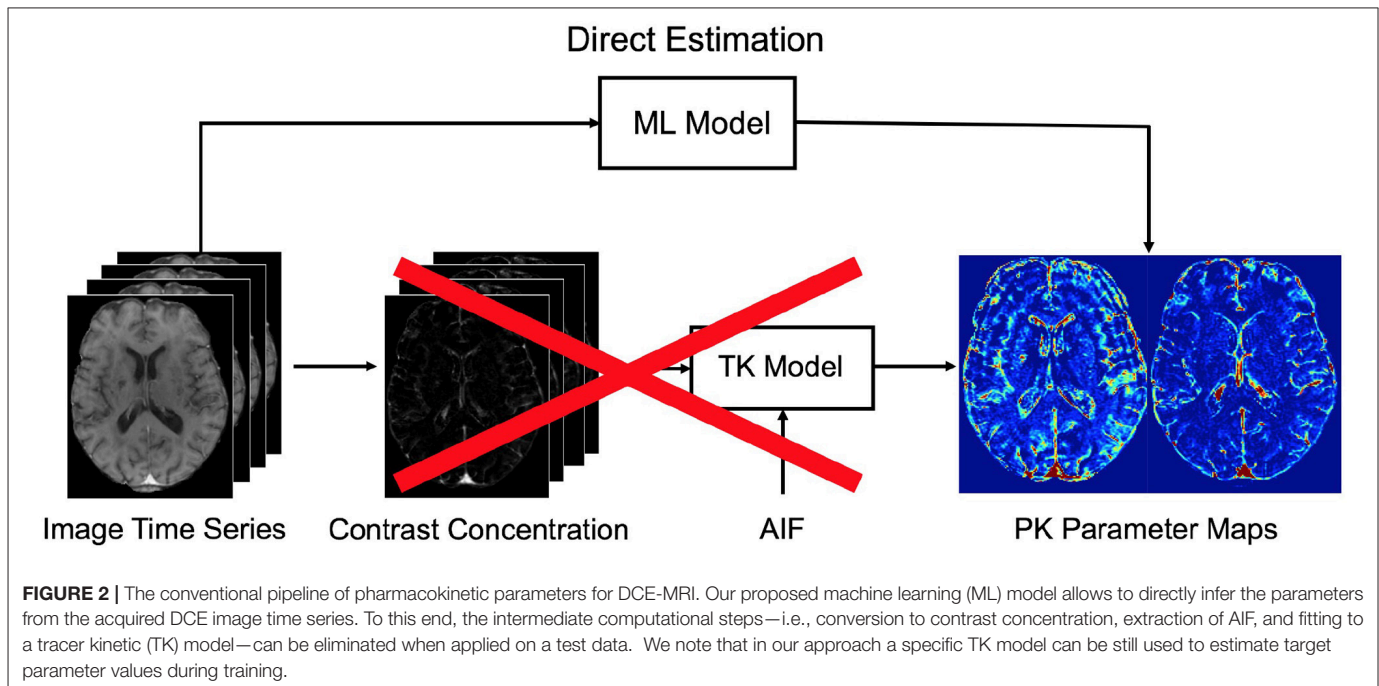
FIGURE 2 | The conventional pipeline of pharmacokinetic parameters for DCE-MRI. Our proposed machine learning (ML) model allows to directly infer the parameters from the acquired DCE image time series. To this end, the intermediate computational steps—i.e., conversion to contrast concentration, extraction of AIF, and fitting to a tracer kinetic (TK) model—can be eliminated when applied on a test data. We note that in our approach a specific TK model can be still used to estimate target parameter values during training.

subtype. DCE-MRI was performed a minimum of 1 month after the stroke in order to avoid acute effects of the stroke on the local BBB (22). This study was approved by the Lothian Ethics of Medical Research Committee (REC 09/81101/54) and the NHS Lothian R + D Office (2009/W/NEU/14), and all patients gave written informed consent.

2.1.2. MRI Acquisition
MR imaging was performed on a 1.5 T MRI scanner (Signa HDxt, General Electric (GE), Milwaukee, WI) using an 8-channel

phased-array coil. Structural MR images for diagnostic purpose were acquired at first including axial T2-weighted (T2W; TR/TE = 6000/90 ms, FoV = 240 × 240 mm, acquisition matrix = 384 × 384, 1.5 averages, 28 × 5 mm slices, 1 mm slice gap), and axial fluid-attenuated inversion recovery (FLAIR; TR/TE/TI = 9000/153/2200 ms, FoV= 240 × 240 mm, acquisition matrix = 384 × 224, 28 × 5 mm slices, 1 mm slice gap).

DCE image series were acquired using a 3D T1W spoiled gradient echo sequence (TR/TE = 8.24/3.1 ms, flip angle = 12°, FoV = 240 × 240 mm, acquisition matrix = 256 × 192, slice

FIGURE 3 | A representative MRI image and corresponding tissue segmentations. FLAIR image (left) and tissue masks superimposed on FLAIR image (right). NAWM, normal-appearing white matter; WMH, white matter hyperintensities; DGM, deep gray matter; RSL, recent stroke lesion.
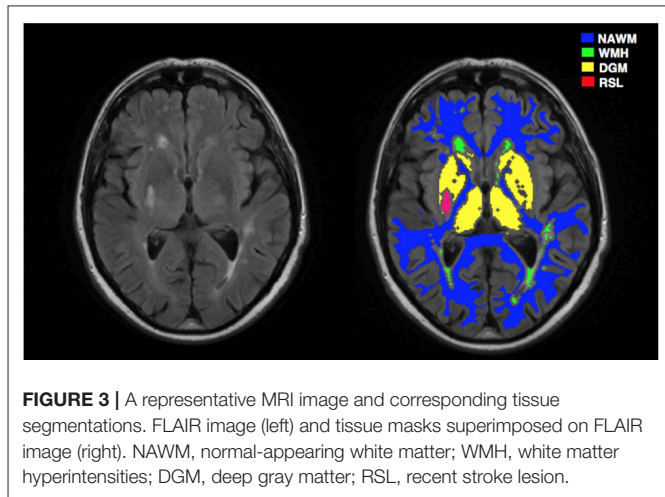
thickness = 4 mm, 42 slices). Two pre-contrast acquisitions were carried out at flip angles of 2° and 12° to calculate pre-contrast longitudinal relaxation times (T_{10}). An intravenous bolus injection of 0.1 mmol/kg of gadoterate meglumine (Gd-DOTA, Dotarem, Guerbet, France) was administered simultaneously with the start of 20 acquisitions with 12° flip angle and a temporal resolution of 73 seconds. The total acquisition time for DCE-MRI was approximately 24 minutes.

2.1.3. Image Processing

For image preprocessing, we mainly followed the steps described in Heye et al. (22). First, all structural and DCE MR images were coregistered to the 12° pre-contrast image using rigid-body registration to correct for bulk patient movement. All small vessel features were determined according to agreed STRIVE standards (23). We employed a multispectral MRI data fusion and minimum variance quantization method (24) for the segmentation of white matter hyperintensities (WMH) and normal-appearing white matter (NAWM), and the resulting masks were manually refined. We used the "Region of Interest" tool of Analyze 11.0$^{\text{TM}}$ (AnalyzeDirect, KS) to semi-automatically outline the old stroke lesions and recent stroke lesion (RSL) boundaries separately. Stroke lesion masks were checked for precision by a neuroradiologist; all other tissue masks were checked visually for accuracy and manually edited by an expert if necessary. Moreover, subcortical/deep gray matter (DGM) masks were generated automatically using a software pipeline as described in Heye et al. (22). In order to minimize any residual contamination of the DGM, the resulting mask was eroded by one voxel. **Figure 3** depicts a representative FLAIR image and corresponding tissue segmentation.

2.2. DCE-MRI Analysis

Data collected at multiple flip angles were first used to calculate the T_{10} map based on the variable flip angle method proposed in Brookes et al. (25), given by

$$\frac{1}{T_{10}} = \frac{1}{T_R} \ln\left(\frac{S_R \sin\alpha_b \cos\alpha_a - \sin\alpha_a \cos\alpha_b}{S_R \sin\alpha_b - \sin\alpha_a}\right), \qquad (1)$$

where $S_R = S_a/S_b$ with S_a and S_b denoting the signal intensities of the two pre-contrast acquisitions with flip angles $\alpha_a = 2°$ and $\alpha_b = 12°$, and T_R is the repetition time.

Dynamic DCE image series $S(t)$ are converted to contrast agent concentration $C_t(t)$ through the steady-state spoiled gradient echo (SGPR) signal equation (26),

$$S(t) = \frac{M_0 \sin\alpha_b (1 - e^{-(K+L)})}{1 - \cos\alpha_b e^{-(K+L)}} + \left(S(0) - \frac{M_0 \sin\alpha_b (1 - e^{-K})}{1 - \cos\alpha_b e^{-K}}\right), \tag{2}$$

where $K = T_R/T_{10}$, $L = r_1 C_t(t) T_R$, r_1 is the contrast agent relaxivity taken as 4.2 $\text{s}^{-1}\text{mM}^{-1}$, $S(0)$ is the baseline (pre-contrast) image intensity, and T_{10} and M_0 are respectively the T_1 relaxation and equilibrium longitudinal magnetization that are calculated from a pre-contrast T_1 mapping acquisition.

For each subject, we extracted a vascular input function (VIF) from a region located on the superior sagittal sinus (SS) because partial volume effects and inflow artifact were reduced at this location compared to obtaining the arterial input function (AIF) from a feeding artery (22); the delay between arterial and venous responses is expected to be very small compared with the temporal resolution of our acquired data. Instead of selecting only a single voxel, we determined a 3×3 patch inside the SS region and estimated the VIF by averaging the time-signal intensities over the voxels within the patch. This enabled us to obtain more smooth variations in the DCE-MRI time course. We converted the whole-blood concentration $C_b(t)$ measured in the SS to plasma concentration using the formula $C_p(t) = C_b(t)/(1 - \text{Hct})$ where Hct is the blood hematocrit measured in large arteries and assumed to be Hct = 0.45 as previously used in literature (22, 26, 27).

2.2.1. Tracer Kinetic Models

Tracer kinetic modeling (28) is applied in DCE-MRI to provide a link between the contrast agent concentration and the physiological or so-called pharmacokinetic parameters, including the fractional plasma volume (v_p), the fractional interstitial volume (v_e), the volume transfer rate (K^{trans}) at which contrast agent (CA) is delivered to the extravascular extracellular space (EES) from plasma space.

In this study, we fitted the following two models to the tissue concentration curves $C_t(t)$: (i) the extended Tofts model, (ii) the Patlak model. A schematic overview of the two models and their relationship is illustrated in **Figure 4**.

The extended Tofts (eTofts) model (29) mainly describes a highly perfused ($F_p = \infty$) two- compartment tissue model considering bidirectional transport between the blood plasma and EES. The concentration of contrast agent in the tissue is determined by,

$$C_t(t) = v_p C_p(t) + K^{\text{trans}} \int_0^t C_p(\tau) e^{-k_{\text{ep}}(t-\tau)} d\tau, \qquad (3)$$

FIGURE 4 | Illustration of two tracer kinetic models: Extended Tofts (left) and Patlak (right) model. Target parameters of DCE-MRI modeling are the contrast agent transfer rate from plasma space to tissue space K^{trans}, the fractional plasma volume v_p, the fractional interstitial volume v_e, and transfer constant from the tissue space to the blood plasma k_{ep}. Patlak model is related to Extended Tofts model through the assumption $C_p >> C_e$ such that the backflux from the EES into the plasma space is negligible.

where $k_{ep} = K^{trans}/v_e$ represents the transfer constant from the EES back to the blood plasma. For the fitting of eTofts model, we used limited-memory Broyden-Fletcher Goldfarb-Shannon (l-BFGS) method for nonlinear minimization of the sum of squared residuals. The algorithm was run till convergence for a maximum of 30 iterations.

The Patlak Model (30) can be considered as a special case of the eTofts model, where the backflux from the EES into the blood plasma compartment is negligible. To this end, this model only allows measurement of the two parameters K^{trans} and v_p given by,

$$C_t(t) = v_p C_p(t) + K^{trans} \int_0^t C_p(\tau) d\tau, \qquad (4)$$

An attractive feature of Patlak model is that the model equation in (4) is linear and model parameters can be fitted using linear least squares which has a closed-form solution, hence parameter estimation is fast (9).

2.3. Deep Learning for Pharmacokinetic Parameter Estimation

In this study, we consider the PK parameter inference in DCE-MRI as a mapping problem between intensity image-time series and parameter maps where the underlying mapping can be efficiently learned using deep CNNs. The proposed CNN aims at learning data-driven features with the use of convolutional feature filters to effectively detect the local spatio-temporal characteristics of the DCE time series. The extracted spatio-temporal features are desired to represent the underlying relation between the input and output of the network as much as possible.

Specifically, our CNN is trained to learn a mapping between $S(t)$ and θ to output an estimate of PK maps $\tilde{\theta}$; $\tilde{\theta} = \mathbf{f}(S(t)|\mathbf{w})$, where $\mathbf{f}$ denotes the forward mapping of the CNN with the learned set of filter weights $\mathbf{w}$. We note that set of parameters are represented by $\theta = \{K^{trans}, v_p\}$ for Patlak model and $\theta = \{K^{trans}, k_{ep}, v_p\}$ for eTofts model.

2.3.1. Loss Function

To learn the network weights ($\mathbf{w}$) during training, we need to define an objective function (or loss function) to be minimized. In addition to the standard mean squared error (MSE) loss between the true PK parameter values θ and the estimated values $\tilde{\theta}$ which enforces high fidelity in parameter reconstruction, we simultaneously seek the fitted contrast agent concentrations of the PK parameters to be sufficiently close to the observed concentrations, $C_t(t)$. To this end, we formulate a new loss function which jointly incorporates these two loss criteria. Given a large number of training samples $\mathcal{D}$ of input-output pairs $(S(t), \theta)$, we train a CNN model that minimizes the following loss,

$$\mathcal{L}(\mathbf{w}) = \sum_{(S(t),\theta) \in \mathcal{D}} \left(\|\theta - \tilde{\theta}\|_2^2 \ + \ \|C_t(t) - f_{tk}(\tilde{\theta})\|_2^2 \right), \quad (5)$$

where f_{tk} is the tracer kinetic model equation of either eTofts or Patlak model as formulated by Equation (3) or Equation (4), respectively.

2.3.2. Network Architecture

We illustrate the network structure used in this study in **Figure 5**. The network takes DCE image-time series as input with a patch size of $24 \times 24 \times 21$, where time frames are stacked as input channels. The first convolutional layer applies 2D filters to each channel individually to extract low-level temporal features which are aggregated over frames via learned filter weights to produce a single output per voxel. Inspired by the work on brain segmentation (31) and denoising in arterial spin labeling (32), our network consists of parallel dual pathways to efficiently capture multi-scale information after the first layer. The local pathway focuses on extracting details from the local image regions while the global pathway is designed to incorporate more contextual global information. The global pathway consists of 3 dilated convolutional layers with dilation factors of 2, 4, and 8, indicating increased receptive field sizes. Zero-padding is applied before every convolution operation to keep the spatial

dimensions of the output equal to the input. The filter size of each convolutional layer including dilated convolutions is chosen as 4×4. The rectified linear units (ReLU) activation function ($f(x) = \max(0, x)$) is applied after each convolution to introduce non-linearity into the mapping. Local and global pathways are then concatenated to form a multi-scale feature set. Following this, two fully-connected layers of 256 and 128 hidden nodes are used to determine the best possible feature combination that can accurately map the input to output of the network. Finally, the last fully-connected layer outputs the parameter estimates of a patch size $24 \times 24 \times n$, where n is the number of kinetic model parameters. We emphasize that as our proposed network was structured to estimate outputs for every single voxel of the input patch, it is essential to keep the spatial dimensions of the input and output same throughout the network. Therefore, in our network we can consider a fully-connected (FCN) layer as a convolutional (CONV) layer with 1×1 convolutions.

2.3.3. Network Training
Among all the follow-up scans we only selected one DCE-MRI scan per subject in our experiments. All these scans were acquired at between 1-3 months post-stroke. For each patients data, we neglected the first and last 5 image slices due to insufficient brain coverage. Among the remaining slices of each patient we randomly selected 20 slices to be considered in analysis. We note that these are the central 20 slices that contain most of the brain regions in overall. Following to this, each 2D DCE image slice was divided into overlapping patches of size 24×24 voxels with step size of 6 voxels. This resulted in a collection of approximately 12,000 patches for every patients data. We applied the same procedure on contrast agent concentration data and target parameter maps required for network training.

All experiments were performed in a leave-one-subject-out fashion, i.e., 30 different networks were trained in total based on both Patlak and eTofts model parameters. Randomly chosen 10,000 overlapping patches of each subject were split into training (80%) and validation (20%) sets. The networks were trained using the Adam optimizer with a learning rate of 10^{-3} and a decay rate of 10^{-4} for maximum number of 200 epochs and a mini-batch size of 1000 patches. Early stopping was applied to prevent poor generalization performance when the validation loss stopped improving within consecutive 15 epochs. In **Figure 6** we provide two exemplary plots depicting the changes in training and validation loss over epochs for CNN trained on Patlak and eTofts models. Both losses show a decreasing trend and converge to a minimum. We implemented our code using Keras library with TensorFlow (33) backend, and experiments were run on a NVIDIA GeForce Titan Xp GPU with 12 GB RAM.

2.3.4. Testing
Once the network is trained and network parameters are learned, DCE image-time series data of a test subject can be fed into the network to directly predict the PK parameters. Since the predictions are processed in a patch-wise manner, all overlapping 16 predictions of a neighborhood are averaged to obtain a final value for every individual voxel.

3. RESULTS

3.1. Comparison of Pharmacokinetic Maps
We compare the qualitative PK parameter maps obtained by Patlak model fitting, eTofts model fitting and CNN model trained by either Patlak or eTofts model. **Figures 7B,C** shows PK parameter maps of an exemplary slice of a patients data. In overall, the parameter maps by CNN model looks very similar with the Patlak model fitting. However, the CNN model produces higher estimates of K^{trans} in especially small RSL region as marked on the DCE image in **Figure 7A**. Moreover, the RSL region is more distinctive and can be discriminated well with respect to other tissues in both the parameter maps of CNN model. For numerical evaluation of output parameter maps, we used two evaluation metrics calculated within the entire brain region: Structural similarity index (SSIM) and normalized root mean square error (nRMSE). These values were calculated by considering the output maps of Patlak model as reference, shown in **Figures 7B,C**. For K^{trans}, we obtain a high SSIM of 0.991 and a low nRMSE of 0.0144. For v_p, SSIM is calculated as 0.973 and nRMSE is 0.0168.

Figures 8B,C demonstrates PK parameter maps of an exemplary slice of an another patients data fitted by eTofts model. The parameter estimates significantly match each other (for CNN and eTofts) in many of the tissue regions except NAWM as depicted on the DCE image in **Figure 8A**. As shown in **Figure 8C**, CNN model yields lower v_p values in comparison to eTofts model in NAWM. Hence, the discrimination of the NAWM with respect to WMH is more prominent. Quantitatively, when compared against the parameter maps obtained by eTofts model, CNN maps yield a SSIM score of 0.998 and 0.961 for K^{trans} and v_p, respectively, while nRMSE is 0.0073 and 0.0156.

3.2. Fitting to the Observed Concentration -Time Series
We evaluate the accuracy of the fitting to the observed concentration-time series data. The fitted contrast agent concentration-time series were estimated via (3) and (4) by using the parameter estimates of Patlak, eTofts, and CNN models separately.

Table 1 demonstrates the quantitative comparison of the fitting to the observed contrast agent concentration time series data for different models in terms of nRMSE and SSIM. The metric values were calculated for every 2D slice of a subject's volume, and statistical values (mean ± std) were obtained using all 15 subject's data. The results indicate that standard Patlak and eTofts model can fit the data better compared to the CNN model trained with these models separately. However, the difference is not substantial that CNN model still achieves high accuracy with less than an average %2 fitting error.

Figures 9A,B shows the fitting of contrast concentration (in mM) for the NAWM and RSL regions in a single patient data. In general, the CNN model trained by either Patlak or eTofts model parameters can fit the data similarly

FIGURE 5 | Illustration of the deep learning architecture used for the estimation of PK parameter maps given the DCE image patch-time series as input. All convolutional layers except concatenation layer learn 32 filters whereas concatenation layer learns 64 filters. Every convolutional layer involves a filter size of 4×4. ReLU is used as a non-linear activation function after each convolutional and fully connected layer. The size of the outputs from each layer operation [e.g., input, convolution and full connection (FCN)] are also displayed at the bottom of each layer.

FIGURE 6 | Training and validation loss over epochs obtained by training a CNN model using **(A)** Patlak and **(B)** eTofts model parameters. Gradual decrease in the loss indicates the efficiency of the network for learning useful representations related to the underlying mapping between the input and output.

well when compared with Patlak and eTofts model. An interesting observation in **Figure 9B** is that the eTofts model does not fit the observed data well whereas the fitting obtained by CNN model trained on eTofts parameters is more accurate.

3.3. Statistical Analysis of PK Parameter Estimation
We perform statistical analysis of the parameter estimates on different tissues. A comparison between tissue types is shown in **Figure 10**. We assessed the statistical significance of the
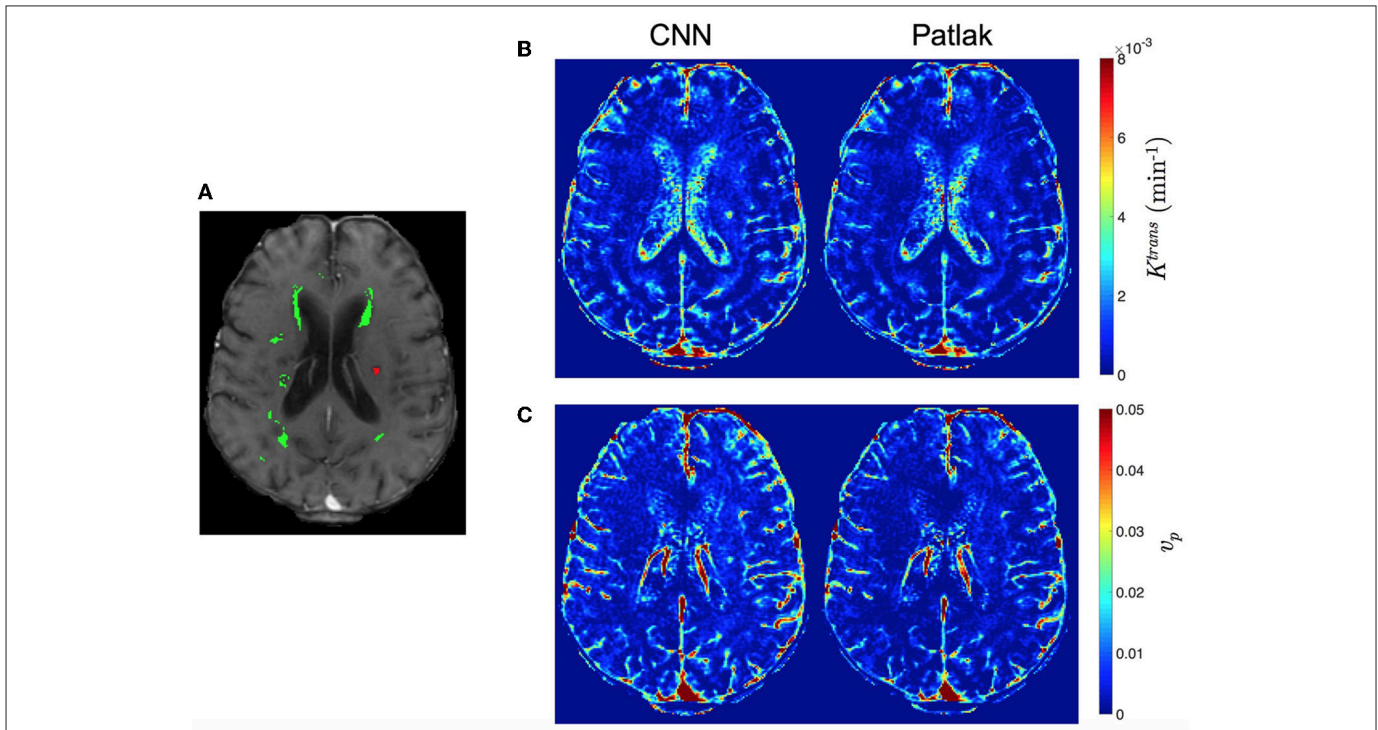
FIGURE 7 | Comparison of qualitative PK parameter maps from a slice of a stroke patient data. **(A)** a DCE image slice on which the tissue masks are superimposed (WMH: green, RSL: red), **(B)** K^{trans} and **(C)** v_p parameter maps obtained by CNN model and Patlak fitting.
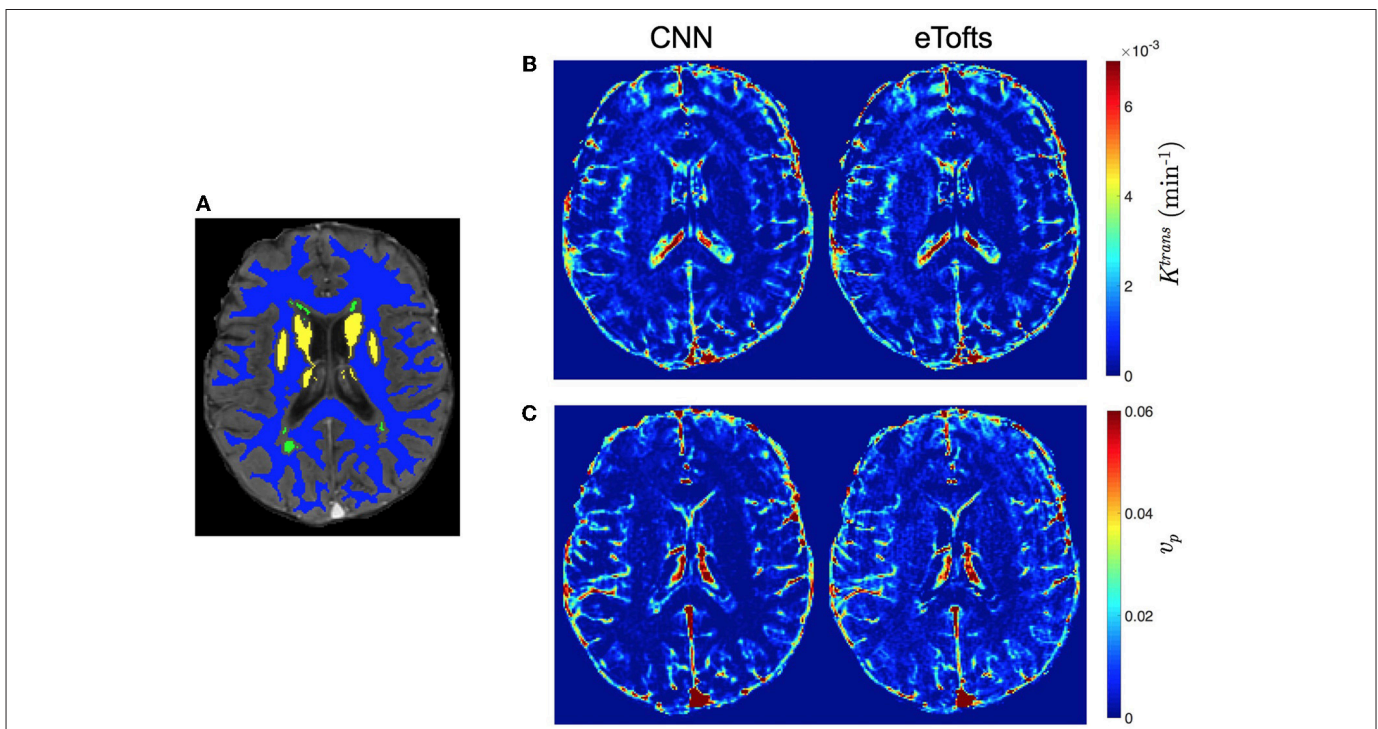
FIGURE 8 | Comparison of qualitative PK parameter maps from a slice of a patient data with white matter hyperintensities. **(A)** a DCE image slice on which the tissue masks are superimposed (NAWM: blue, WMH: green, DGM: yellow), **(B)** K^{trans} and **(C)** v_p parameter maps obtained by CNN model and eTofts fitting. We remark that WMH represents the WM tissue associated with increased risk of dementia and cognitive decline.

TABLE 1 | nRMSE (%) and SSIM statistics (mean ± std) obtained from concentration-time series data fitting. The SSIM value can vary between −1 and 1, where 1 indicates perfect similarity.

	Models			
Metric	Patlak	eTofts	CNN: Patlak trained	CNN: eTofts trained
nRMSE (%)	1.1200 ± 0.5225	1.0575 ± 0.5744	1.6398 ± 0.6878	1.7360 ± 0.7408
SSIM	0.9812 ± 0.0141	0.9835 ± 0.0127	0.9750 ± 0.0138	0.9719 ± 0.0162

differences using the paired Wilcoxon signed rank test. For Patlak and eTofts model, all differences between tissue types were significant ($p < 0.001$) except for K^{trans} in DGM and WMH, and v_p in WMH and RSL. For CNN model trained on Patlak model parameters, all differences of K^{trans} between tissue types were significant including the difference between WMH and DGM ($p = 3.4 \times 10^{-4}$). The difference between WMH and RSL for v_p is again statistically significant with $p = 1.6 \times 10^{-5}$. The CNN model trained on Patlak generally tends to overestimate the K^{trans} and v_p parameters compared to either Patlak or eTofts model. The difference between them are significant with $p < 0.001$, and this is valid for all tissue types except DGM ($p = 0.021$ for K^{trans}). On the other hand, the CNN model trained with eTofts parameters yield underestimated K^{trans} and overestimated v_p values when compared against either Patlak or eTofts model. The underestimation of K^{trans} by CNN is statistically significant for all tissue types except WMH ($p = 0.317$). The overestimation of v_p by CNN is significant for all tissue types ($p < 0.001$).

Figure 11 depicts the Bland-Altman plots of K^{trans} values in three different tissues (DGM, WMH, RSL) obtained from a patient's data. As can be observed in **Figure 11A**, when compared against the Patlak model, CNN model trained with Patlak tends to slightly underestimate the K^{trans} in DGM and overestimate the values in WMH and RSL. **Figure 11B** indicates that K^{trans} are underestimated by CNN trained with eTofts in DGM and RSL. The values in WMH highly match with Patlak fitting showing no systematic difference. In general, the results in Bland-Altman plots agree with the statistical results as shown in **Figure 10**, meaning that systematic differences are observable between the estimates of CNN and model fitting although concordance correlation coefficients (CCCs) indicate a strong agreement.
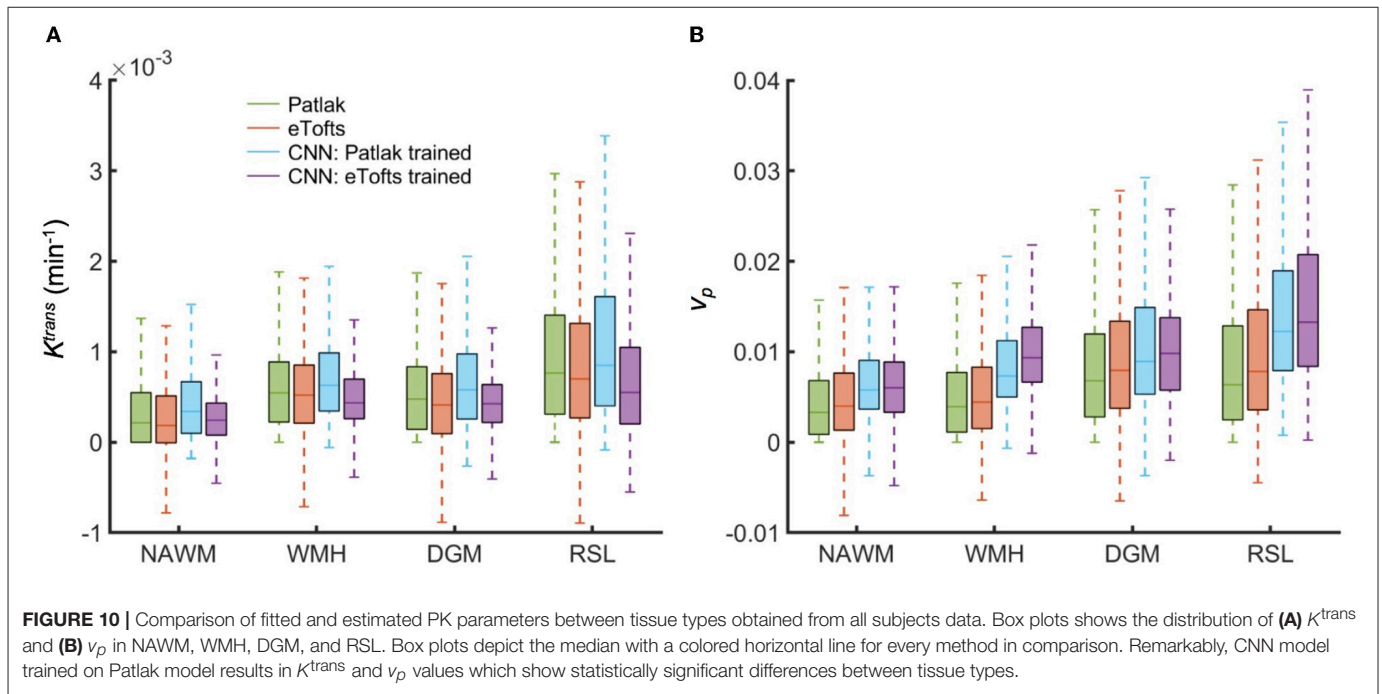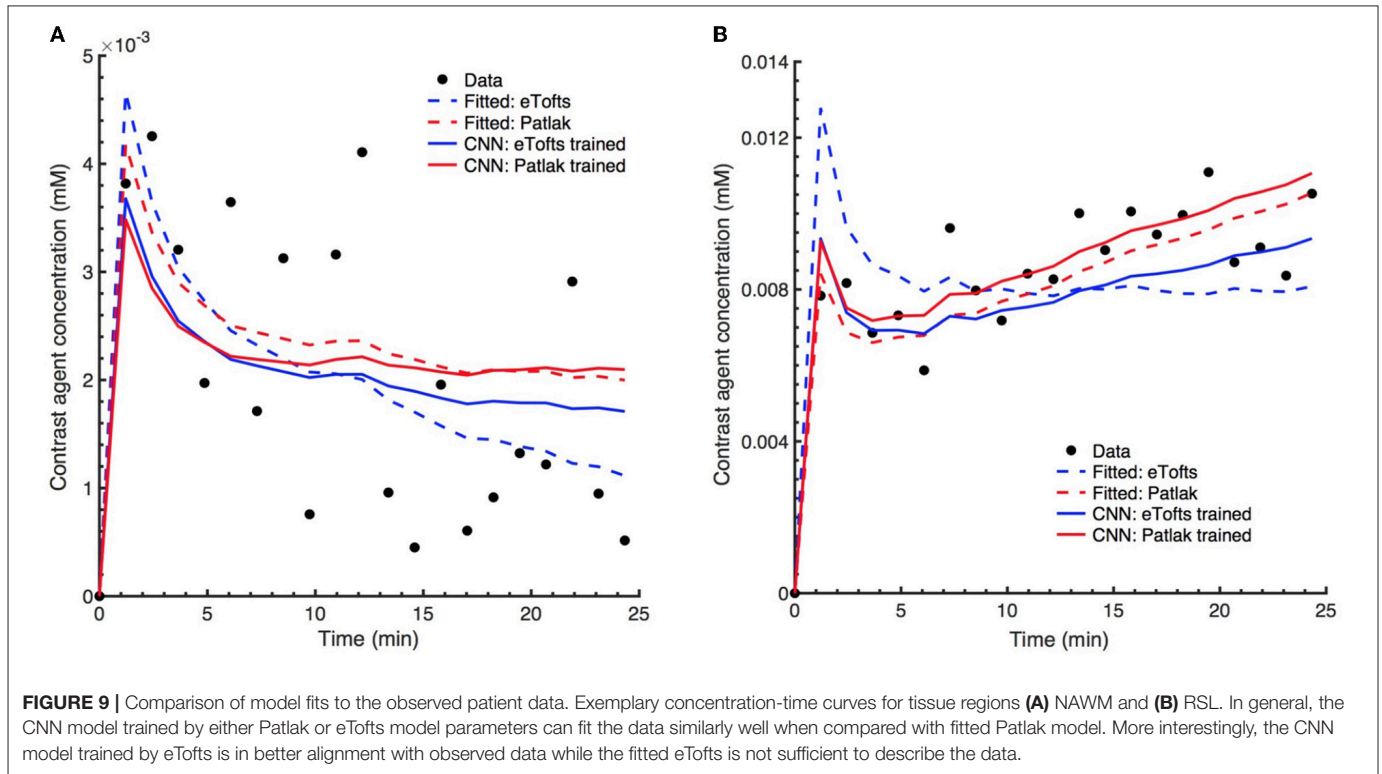
4. DISCUSSION

The results of this study show that a CNN based ML model can yield PK parameter estimates that are comparable to traditional model fitting. As depicted in **Figures 7, 8**, the qualitative parameter maps estimated by CNN models match highly with the ones obtained by conventional TK model fitting methods. Moreover, ML based models can enable better discrimination of different brain tissues. As can be seen in **Figure 7**, small stroke lesion is more visible with higher K^{trans} values assigned to this region. In addition to this, the discontinuities of parameter values arising especially at highly perfused regions (i.e., vessels) can be mitigated by CNN model, and more smoother local areas are produced in these regions as shown in **Figures 7, 8**.

Statistical analysis in **Figure 10** indicate that significant differences between tissue types can be achieved by CNN model whereas both Patlak and eTofts model fail in quantitatively differentiating some of the tissues pairwise including WMH-DGM. Especially higher K^{trans} values are generally assigned to stroke regions i.e., RSL, allowing better discrimination of these areas against non-stroke regions. To this end, the proposed ML model can be an appropriate parameter inference model for quantification of subtle BBB disruption where measuring low-level BBB permeability is vital in several diseases, including cerebral small vessel disease, lacunar stroke and vascular dementia. Another interesting observation is that the plasma volume v_p values estimated by CNN model in WMH are considerably greater than in normal-appearing WM areas. This may result in improved identification of the hyperintensity areas from the surrounding normal appearing WM tissue. WM hyperintensities are usually regarded as surrogates of small vessel disease and frequently seen in elderly people (34).

The major advantage of ML based model is that the parameter inference of a voxel belonging to a specific tissue type is performed by taking into account many other training samples, or voxels, of the same tissue type. Therefore, if the signal time series of a target voxel is subject to high noise, it is likely that a parameter value associated with the voxels that show similar signal trends and located in the same tissue type can be assigned to the target voxel. One example relevant to this observation can be seen in **Figure 9B**, where the fitted concentration time curves are provided for a ROI inside the RSL region of a patients data. Here, the eTofts model does not provide a good fit to the measured signal and the fitted concentration-time curve describes more a vascular region (i.e., blood vessel). However, the fit of the CNN model trained with eTofts model parameters can produce significantly better fit to the observed data, and the fit resembles more an RSL region, which is highly similar with the fits by Patlak and CNN model trained by Patlak model parameters. These findings reveal better generalization ability of ML models (35) which can extract and learn important tissue specific features from a large cohort of training examples. However, it should be noted that the correction of misfit of concentration time curves in **Figure 9B** does not point out an unique feature of our CNN based approach, but rather shows a specific case. The avoidance of a misfit with the CNN network primarily depends upon the model and optimization approach on which the network is trained.

Another observation from **Figure 11** also signifies the tendency of CNN model to produce parameter estimates close to a mean value of parameter distribution learned from

FIGURE 9 | Comparison of model fits to the observed patient data. Exemplary concentration-time curves for tissue regions **(A)** NAWM and **(B)** RSL. In general, the CNN model trained by either Patlak or eTofts model parameters can fit the data similarly well when compared with fitted Patlak model. More interestingly, the CNN model trained by eTofts is in better alignment with observed data while the fitted eTofts is not sufficient to describe the data.

FIGURE 10 | Comparison of fitted and estimated PK parameters between tissue types obtained from all subjects data. Box plots shows the distribution of **(A)** K^{trans} and **(B)** v_p in NAWM, WMH, DGM, and RSL. Box plots depict the median with a colored horizontal line for every method in comparison. Remarkably, CNN model trained on Patlak model results in K^{trans} and v_p values which show statistically significant differences between tissue types.

many training voxels within in a specific tissue. Here, when compared to the standard Patlak model parameters, we observe overestimated values in especially WMH and RSL region where the K^{trans} usually has higher values. The overestimation in some of the voxels within these tissues is presumably caused by the relatively lower parameter values estimated by

Patlak model due to significant signal noise and fitting to the local minima. In this regard, systematic differences between CNN model estimates and standard NLS fitting are inevitable because the parameter estimates by NLS fitting is not optimal and usually produces a parameter distribution from a high range of values within the voxels of a specific tissue, as it

FIGURE 11 | Bland-Altman plots of permeability K^{trans} parameters in different tissues, DGM (left), WMH (middle), RSL (right). Difference (y-axis) in K^{trans} values between CNN model and **(A)** Patlak, and, **(B)** eTofts fitting is plotted against the mean values of the two (x-axis). Solid gray line indicates mean difference (mdiff). Top and bottom gray dashed lines correspond to upper and lower margins of 95% limits of agreement estimated by mdiff $\pm 1.96 \times$ SD, SD = standard deviation. Units for horizontal and vertical axes are in min^{-1}. The computed Lin's concordance correlation coefficient (CCC) values are displayed at top-right corner of each plot.

can be seen in **Figure 10**. We anticipate that more accurate evaluation of systematic differences can be obtained using the synthetic DCE dataset where the ground truth parameters are known.

As mentioned before, one of the key advantages of our method is its utility to avoid intermediate computation steps of parameter inference in DCE-MRI by replacing it with a direct inference model. Although we use two existing TK models to estimate the reference parameters, based on the specific DCE application, one can use different TK models in literature (9) to infer the PK parameters to be used during training of the CNN network. If available, the network can be also trained using ground truth parameter values. In addition to this, as previously done in Banerji et al. (36) and Bosca and Jackson (37), synthetic DCE phantom data can be generated by simulating the signal equation

and TK model equations with the PK parameters estimated from real patient's data, and a CNN model can be trained based on the synthetic data and corresponding parameter maps. With this approach, more realistic synthetic DCE datasets can be generated by taking into account the acquisition noise and motion artifacts. The generated synthetic datasets may be utilized to train a network which can be later tested on *in vivo* DCE dataset to obtain less noise-sensitive parameter estimates.

In conventional DCE-MRI analysis pipeline, subject-specific AIF extraction from a ROI of a feeding artery is one of the essential steps for the estimation of kinetic parameters (28, 38). In this study, we demonstrate that CNN based ML model can estimate PK parameters by no need of subject-specific AIF of the test subject without introducing any significant bias in the parameter estimation. Although this can be seen as one of the

FIGURE 12 | Examplary subject specific vascular input functions (VIFs) extracted from averaging a few voxels located on the superior sagittal sinus (SS). Although the VIFs appear to have similar shapes, the peaks and steady state signal can have varying magnitude (contrast agent concentration).
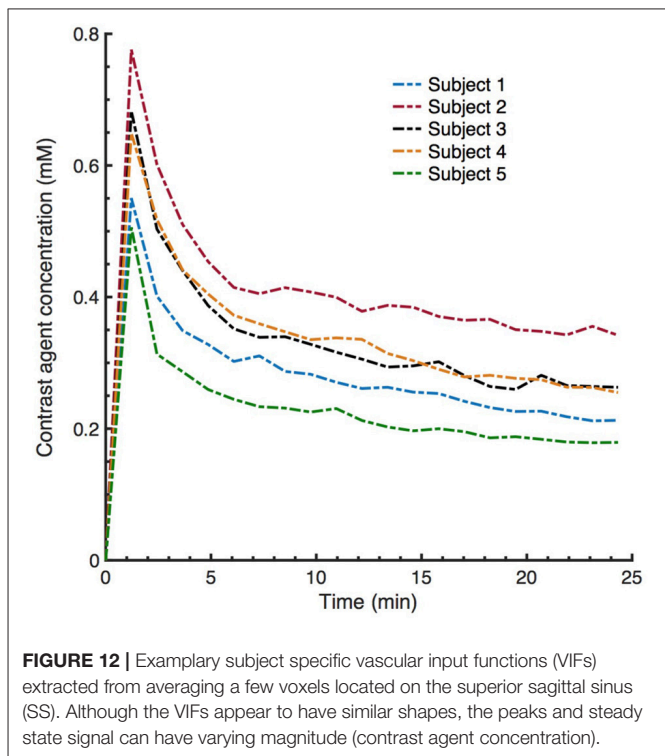
benefits of our model, we should remark that the data used in this work is a part of a population study where the temporal resolution and other parameters related to DCE acquisition and contrast injection are fixed in all subjects. However, as can be clearly seen in **Figure 12**, the subject-specific AIFs of our dataset usually have varying magnitudes of the peak and steady-state signal even though the time point where the signal reaches the peak is similar for all subjects. The signal pattern of the AIF curves are directly related to signal time intensities through Equation (2), hence the trained network can intrinsically learn the relation between the AIF and target parameters via the mapping between the input and output of the network and the designed loss function which takes into account the underlying TK model through its equation. On the other hand, the performance of proposed model on a mixed data—ideally involving DCE image series acquired with different acquisition parameters and protocols—can be subject to further investigation. For parameter estimation with a model trained on a mixed data, we anticipate that a bi-CNN input model similar to as proposed for DSC-MRI (21) might be a good approach to avoid bias and error in parameter estimation. In that setting, the DCE image time series and other acquisition parameters— including AIF—of both training and test subjects can be given to the network as two separate inputs.

We emphasize that our CNN model is not trained on a entire brain basis, but on individual time series. Out of the 15 patient datasets we extract more than 160 million training samples, i.e., number of total voxels in the training dataset. Moreover, our network architecture is not very deep and we demonstrate that this huge number of training samples is sufficient to train a network that generalizes well, where the inability to

generate reproducible results is not an issue. Nevertheless, a wider sampling of pathological cases and MRI artifacts in training data is highly desirable and is one of the major direction for improvements of the proposed approach. The proposed model can—even should—be updated accordingly when applied to a larger pool of patient datasets. In general, based on the literature in ML, we anticipate that CNN-based ML models perform better when there is a high correlation and similarity between the training and test data. The dataset used in this study for both training and testing involve voxels from different type of tissues, e.g., healthy and pathological tissues, containing a good mixture of different tissue characteristics. There is a high similarity between the temporal profiles of training and test image patches, hence, the performance of CNN is very stable and robust. However, a poor generalization issue may usually arise in a scenario that the training data only consists of healthy tissue voxels whereas the unseen test data with pathological tissues is tested using the trained model. In this scenario, since the model is not trained with sufficient number of pathological samples, it is quite likely that the CNN model shows a poor performance on these test data comprised of non-healthy tissues. In principle, in order to obtain a stable CNN model, it is necessary to constitute a training data pool according to the demands or expectation from such a prediction model in our specific clinical applications. For instance, if we aim to discriminate well the acute/post-acute stroke regions, our training data should contain high number of voxels from both stroke and non-stroke regions.

Nevertheless, we should discuss the several limitations of this study. First, although ML based methods can have strong generalization ability, the bias is also inevitable when tested on an unseen data because the model is always trained using other subject's data without any access to test data. Second, the performance of our method may be improved depending on the input patch size and filter size of the network. Moreover, we only considered 2D convolution operations, however, 3D convolutions may produce better results when more spatial context information are extracted. Third, further investigation on synthetic data is required to perform accurate assessment of error and bias when the ground truth parameter values are known. Lastly, our current approach is sensitive to variation in acquisition parameters, especially temporal resolution, i.e., number of time points in DCE data. One feasible solution to the variations in temporal resolution across multiple datasets is to apply interpolation on time. In practice, we may interpolate all training data acquired with various temporal resolutions to a common temporal resolution so that a test data with completely different temporal resolution can be also fed into the trained network to produce parameter estimates.

In conclusion, this study shows that a ML based direct inference approach can estimate PK parameters that are comparable to the conventional model fitting in DCE-MRI. Our results, based on a sample of mild ischaemic stroke patients, demonstrate the efficiency of CNN model to enable better discrimination of brain tissue types. Specifically, our proposed ML based method has the potential to improve the current quantitative analysis of DCE-MRI studies due to its increased robustness to noise. Significant difference of

permeability parameters between stroke and non-stroke regions may ultimately effect the stroke medical decision process. Finally, parameter inference of the proposed model on a 3D brain volume is considerably faster than the standard NLS fitting, demonstrating the applicability of such models in clinical practice. Considering such faster computation time, clinical experts may perform parameter inference using various TK models in parallel to benefit from making more detailed analysis between different models.

AUTHOR CONTRIBUTIONS

CU study concept, data analysis, experimental design, writing of manuscript. DD experimental design, writing of manuscript. MT study concept, experimental design. MV image preprocessing. PA and SM data collection. JW funding. BM study concept, writing of manuscript, funding. All authors contributed to reviewing and editing the final manuscript.

REFERENCES

1. O'Connor JP, Jackson A, Parker GJ, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. *Nat Rev Clin Oncol.* (2012) 9:167–77. doi: 10.1038/nrclinonc.2012.2

2. Yankeelov TE, Gore JC. Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. *Curr Med Imaging Rev.* (2007) 3:91–107. doi: 10.2174/157340507780619179

3. Collins DJ, Padhani AR. Dynamic magnetic resonance imaging of tumor perfusion. *IEEE Eng Med Biol Magazine* (2004) 23:65–83. doi: 10.1109/MEMB.2004.1360410

4. Farrall AJ, Wardlaw JM. Blood–brain barrier: ageing and microvascular disease-systematic review and meta-analysis. *Neurobiol Aging* (2009) 30:337–52. doi: 10.1016/j.neurobiolaging.2007.07.015

5. Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. *Lancet Neurol.* (2013) 12:483–97. doi: 10.1109/S1474-4422(13)70060-7

6. Heye AK, Culling RD, del C Valds Hernndez M, Thrippleton MJ, Wardlaw JM. Assessment of blood–brain barrier disruption using dynamic contrast-enhanced MRI. A systematic review. *Neuroimage* (2014) 6:262–74. doi: 10.1016/j.nicl.2014.09.002

7. Parker GJM, Buckley DL. Tracer Kinetic Modelling for T1-Weighted DCE-MRI. In: Jackson A, Buckley DL, Parker GJM, editors. *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology.* Berlin;Heidelberg: Springer (2005). pp. 81–92.

8. Lebel RM, Jones J, Ferre JC, Law M, Nayak KS. Highly accelerated dynamic contrast enhanced imaging. *Magn Reson Med.* (2014) 71:635–44. doi: 10.1002/mrm.24710

9. Sourbron SP, Buckley DL. Classic models for dynamic contrast-enhanced MRI. *NMR Biomed.* (2013) 26:1004–27. doi: 10.1002/nbm.2940

10. Kelm BM, Menze BH, Nix O, Zechmann CM, Hamprecht FA. Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge. *IEEE Trans Med Imaging* (2009) 28:1534–47. doi: 10.1109/TMI.2009.2019957

11. Debus C, Floca R, Nrenberg D, Abdollahi A, Ingrisch M. Impact of fitting algorithms on errors of parameter estimates in dynamic contrast-enhanced MRI. *Phys Med Biol.* (2017) 62:9322. doi: 10.1088/1361-6560/aa8989

12. Schmid VJ, Whitcher B, Padhani AR, Taylor NJ, Yang G. Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Trans Med Imaging* (2006) 25:1627–36. doi: 10.1109/TMI.2006.884210

13. Wang S, Summers RM. Machine learning and radiology. *Med Image Anal.* (2012) 16:933–51. doi: 10.1016/j.media.2012.02.005

14. Das D, Coello E, Schulte RF, Menze BH. Quantification of metabolites in magnetic resonance spectroscopic imaging using machine learning. In: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017-20th International Conference.* Quebec City, QC (2017) pp. 462–470.

15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* (2015) 521:436–44. doi: 10.1038/nature14539

16. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems* (Lake Tahoe, NV) (2012). pp. 1097–105.

17. Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. In: *Advances in Neural Information Processing Systems* (Lake Tahoe, NV) (2013). pp. 2553–61.

18. Golkov V, Dosovitskiy A, Sperl JI, Menzel MI, Czisch M, Smann P, et al. q-Space deep learning: twelve-fold shorter and model-free diffusion MRI scans. *IEEE Trans Med Imaging* (2016) 35:1344–51. doi: 10.1109/TMI.2016.2551324

19. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: deep De-Aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging* (2018) 37:1310–21. doi: 10.1109/TMI.2018.2858779

20. Benou A, Veksler R, Friedman A, Raviv TR. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Med Image Anal.* (2017) 42(Suppl. C):145–59. doi: 10.1016/j.media.2017.07.006

21. Ho KC, Scalzo F, Sarma KV, El-Saden S, Arnold CW. A temporal deep learning approach for MR perfusion parameter estimation in stroke. In: *23rd International Conference on Pattern Recognition (ICPR)* (Cancun) (2016). pp. 1315–20.

22. Heye AK, Thrippleton MJ, Armitage PA, Valdés Hernández MDC, Makin SD, Glatz A, et al. Tracer kinetic modelling for DCE-MRI quantification of subtle blood–brain barrier permeability. *Neuroimage* (2016) 125:446–55. doi: 10.1016/j.neuroimage.2015.10.018

23. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* (2013) 12:822–38. doi: 10.1016/S1474-4422(13)70124-8

24. Valdés Hernández MdC, Ferguson KJ, Chappell FM, Wardlaw JM. New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images. *Eur Radiol.* (2010) 20:1684–91. doi: 10.1007/s00330-010-1718-6

25. Brookes JA, Redpath TW, Gilbert FJ, Murray AD, Staff RT. Accuracy of T1 measurement in dynamic contrast-enhanced breast MRI using two- and three-dimensional variable flip angle

fast low-angle shot. *J Magn Reson Imaging*. (1999) 9:163–71. doi: 10.1002/(SICI)1522-2586(199902)9:2<163::AID-JMRI3>3.0.CO;2-L

26. Guo Y, Lingala SG, Zhu Y, Lebel RM, Nayak KS. Direct estimation of tracer-kinetic parameter maps from highly undersampled brain dynamic contrast enhanced MRI. *Magn Reson Med.* (2017) 78:1566-78. doi: 10.1002/mrm.26540

27. Calamante F, Ahlgren A, van Osch MJ, Knutsson L. A novel approach to measure local cerebral haematocrit using MRI. *J Cereb Blood Flow Metabolism* (2016) 36:768–80. doi: 10.1177/0271678X15606143

28. Sourbron SP, Buckley DL. Tracer kinetic modelling in MRI: estimating perfusion and capillary permeability. *Phys Med Biol.* (2012) 57:R1–33. doi: 10.1088/0031-9155/57/2/R1

29. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging* (1999) 10:223–32. doi: 10.1002/(SICI)1522-2586(199909)10:3<223::AID-JMRI2>3.0.CO;2-S

30. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J Cereb Blood Flow Metab.* (1983) 3:1–7. doi: 10.1038/jcbfm.1983.1

31. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal.* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004

32. Kim KH, Choi SH, Park SH. Improving arterial spin labeling by using deep learning. *Radiology* (2018) 287:658–66. doi: 10.1148/radiol.2017171154

33. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* (2015). Available online at: https://www.tensorflow.org/

34. Wardlaw JM, Valdés Hernández MC, Muñoz-Maniega S. What are white matter hyperintensities made of?: relevance to vascular cognitive impairment. *J Am Heart Assoc.* (2015) 4:e001140. doi: 10.1161/JAHA.114.001140

35. Bishop CM. *Pattern Recognition and Machine Learning.* New York, NY: Springer (2006).

36. Banerji A, Caunce A, Watson Y, Rose C, GA B, Parker G. A flexible software phantom for generating realistic dynamic contrast-enhanced MR images of abdominal tumours. In: *Proceedings of International Society of Magnetic Resonance Medicine* (Toronto, ON) (2008). p. 493.

37. Bosca RJ, Jackson EF. Creating an anthropomorphic digital MR phantoman extensible tool for comparing and evaluating quantitative imaging algorithms. *Phys Med Biol.* (2016) 61:974. doi: 10.1088/0031-9155/61/2/974

38. Cheng HM. T1 measurement of flowing blood and arterial input function determination for quantitative 3D T1-weighted DCE-MRI. *J Magn Reson Imaging* (2007) 25:1073–8. doi: 10.1002/jmri.20898

DeepASL: Denoising Arterial Spin Labeled MRI via Deep Residual Learning

Abstract: Arterial spin labelling (ASL) allows to quantify the cerebral blood flow (CBF) by magnetic labeling of the arterial blood water. ASL is increasingly used in clinical studies due to its noninvasiveness, repeatability and benefits in quantification. However, ASL suffers from an inherently low signal to noise ratio (SNR) requiring repeated measurements of control/spin-labeled (C/L) pairs to achieve a reasonable image quality, which in return increases motion sensitivity. This leads to clinically prolonged scanning times increasing the risk of motion artifacts. Thus, there is an immense need of advanced imaging and processing techniques in ASL. In this paper, we propose a novel deep learning based approach to improve the perfusion-weighted image quality obtained from a subset of all available pairwise C/L subtractions. Specifically, we train a deep fully convolutional neural network (FCN) to learn a mapping from noisy perfusion-weighted image and its subtraction (residual) from the clean

image. Additionally, we incorporate the CBF estimation model in the loss function during training, which enables the network to produce high quality images while simultaneously enforcing the CBF estimates to be as close as reference CBF values. Extensive experiments on synthetic and clinical ASL datasets demonstrate the effectiveness of our method in terms of improved ASL image quality, accurate CBF parameter estimation and considerably small computation time during testing.

Contributions of thesis author: Algorithm design and implementation, computational experiments, data analysis and interpretation, composition and revision of manuscript.

DeepASL: Kinetic Model Incorporated Loss for Denoising Arterial Spin Labeled MRI via Deep Residual Learning

Cagdas Ulas[1]([ ]), Giles Tetteh[1], Stephan Kaczmarz[2], Christine Preibisch[2], and Bjoern H. Menze[1]

[1] Department of Computer Science,
Technische Universität München, Munich, Germany
cagdas.ulas@tum.de
[2] Department of Neuroradiology,
Technische Universität München, Munich, Germany

Abstract. Arterial spin labeling (ASL) allows to quantify the cerebral blood flow (CBF) by magnetic labeling of the arterial blood water. ASL is increasingly used in clinical studies due to its noninvasiveness, repeatability and benefits in quantification. However, ASL suffers from an inherently low-signal-to-noise ratio (SNR) requiring repeated measurements of control/spin-labeled (C/L) pairs to achieve a reasonable image quality, which in return increases motion sensitivity. This leads to clinically prolonged scanning times increasing the risk of motion artifacts. Thus, there is an immense need of advanced imaging and processing techniques in ASL. In this paper, we propose a novel deep learning based approach to improve the perfusion-weighted image quality obtained from a subset of all available pairwise C/L subtractions. Specifically, we train a deep fully convolutional network (FCN) to learn a mapping from noisy perfusion-weighted image and its subtraction (residual) from the clean image. Additionally, we incorporate the CBF estimation model in the loss function during training, which enables the network to produce high quality images while simultaneously enforcing the CBF estimates to be as close as reference CBF values. Extensive experiments on synthetic and clinical ASL datasets demonstrate the effectiveness of our method in terms of improved ASL image quality, accurate CBF parameter estimation and considerably small computation time during testing.

1 Introduction

Arterial spin labeling (ASL) is a promising MRI technique that allows quantitative measurement of cerebral blood flow (CBF) in the brain and other body organs. ASL-based CBF shows a great promise as a biomarker for many neurological diseases such as stroke and dementia, where perfusion is impaired, and thereby the blood flow alterations need to be investigated [2]. ASL has been increasingly used in clinical studies since it is completely non-invasive and uses

magnetically labeled blood water as an endogenous tracer where the tagging is done through inversion radio-frequency (RF) pulses [2,12]. In ASL, a perfusion-weighted image is obtained by subtracting a label image from a control image in which no inversion pulse is applied. The difference reflects the perfusion, which can be quantified via appropriate modelling [2,11].

Despite its advantages, ASL significantly suffers from several limitations including the low signal-to-noise ratio (SNR), poor temporal resolution and volume coverage in conventional acquisitions [5]. Among these limitations, the low SNR is the most critical one, necessitating numerous repetitions to achieve accurate perfusion measurements. However, this leads to impractical long scanning time especially in multiple inversion time (multi-TI) ASL acquisitions with increased susceptibility to motion artifacts [2,9,12].

To alleviate this limitation, several groups have proposed spatial and spatio-temporal denoising techniques, for instance denoising in the wavelet domain [3], denoising in the image domain using adaptive filtering [13], non-local means filtering combined with wavelet filtering [10], spatio-temporal low-rank total variation [5], and spatio-temporal total generalized variation [12]. Just recently, a deep learning based ASL denoising method [9] has been shown to produce compelling results. All of these methods primarily consider improving the quality of noisy perfusion-weighted images, followed by CBF parameter estimation as a separate step although accurate quantification of CBF is the main objective in ASL imaging.

In this paper, unlike the previous deep learning work [9] which is only data driven, we follow a mixed modeling approach in our denoising scheme. In particular, we demonstrate the benefit of incorporating a formal representation of the underlying process – a CBF signal model – as a prior knowledge in our deep learning model. We propose a novel deep learning based framework to improve the perfusion-weighted image quality obtained by using a lower number of subtracted control/label pairs. First, as our main contribution, we design a custom loss function where we incorporate the Buxton kinetic model [4] for CBF estimation as a separate loss term, and utilize it when training our network. Second, we specifically train a deep fully-convolutional neural network (CNN) adopting the residual learning strategy [7]. Third, we use the images from various noise levels to train a single CNN model. Therefore, the trained model can be utilized to denoise a test perfusion-weighted image without estimating its noise level. Finally, we demonstrate the superior performance of our method by validations using synthetic and clinical ASL datasets. Our proposed method may facilitate scan time reduction, making ASL more applicable in clinical scan protocols.

2   Methods

2.1   Arterial Spin Labeling

In ASL, arterial blood water is employed as an endogenous diffusible tracer by inverting the magnetization of inflowing arterial blood in the neck area by using RF pulses. After a delay for allowing the labeled blood to perfuse into

the brain, label and control images are repeatedly acquired with and without tagging respectively [2,11]. The signal difference between control and label images is proportional to the underlying perfusion [2]. The difference images are known as perfusion-weighted images (ΔM), and can be directly used to fit a kinetic model. For CBF quantification in a single inversion-time (TI) ASL, the single-compartment kinetic model (so-called Buxton model [4]) is generally used. According to this model, the CBF in ml/100g/min can be calculated in every individual voxel for pseudo-continuous ASL (pCASL) acquisitions as follows,

$$f(\Delta M) = \text{CBF} = \frac{6000 \cdot \beta \cdot \Delta M \cdot e^{\frac{PLD}{T_{1b}}}}{2 \cdot \alpha \cdot T_{1b} \cdot \text{SI}_{\text{PD}} \cdot \left(1 - e^{-\frac{\tau}{T_{1b}}}\right)}, \tag{1}$$

where β is the brain-blood partition coefficient, T_{1b} is the longitudinal relaxation time of blood, α is the labeling efficiency, τ is the label duration, PLD is the post-label delay, and SI_{PD} is the proton density weighted image [2].

2.2   Deep Residual Learning for ASL Denoising

Formulation. Our proposed CNN model adopts the residual learning formulation [7,8]. It is assumed that the task of learning a residual mapping is much easier and more efficient than original unreferenced mapping [14]. With the utilization of a residual learning strategy, extremely deep CNN can be trained and superior results have been achieved for object detection [7] and image denoising [14] tasks.

The input of our CNN model is a noisy perfusion-weighted image ΔM_n that is obtained by averaging a small number of pairwise C/L subtractions. We denote a complete perfusion-weighted image as ΔM_c estimated by averaging all available C/L subtractions. We can relate the noisy and complete perfusion-weighted image as $\Delta M_n = \Delta M_c + N$, where N denotes the noise image which degrades the quality of the complete image. Following the residual learning strategy, our CNN model aims to learn a mapping between ΔM_n and N to produce an estimate of the residual image $\tilde{N}$; $\tilde{N} = \mathcal{R}(\Delta M_n | \Theta)$, where $\mathcal{R}$ corresponds to the forward mapping of the CNN parameterised by trained network weights Θ. The final estimate of the complete image is obtained by $\Delta \tilde{M}_c = \Delta M_n - \tilde{N}$.

Loss Function Design. In this work, we design a custom loss function to simultaneously control the quality of the denoised image and the fidelity of CBF estimates with respect to reference CBF values. Concretely, given a set of training samples $\mathcal{D}$ of input-target pairs $(\Delta M_n, N)$, a CNN model is trained to learn the residual mapping $\mathcal{R}$ for accurate estimation of complete image by minimizing the following cost function,

$$\mathcal{L}(\Theta) = \sum_{(\Delta M_n, N) \in \mathcal{D}} \lambda \|N - \tilde{N}\|_2^2 \; + \; (1 - \lambda)\|f_t - f(\Delta M_n - \tilde{N}; \xi)\|_2^2, \tag{2}$$

where λ is regularization parameter controlling the trade-off between the fidelity of the residual image and CBF parameter estimates, f_t represents the reference

CBF values corresponding to an input ΔM_n, and ξ denotes all predetermined variables as given in (1). We emphasize that the second term of our loss function (2) explicitly enforces the consistency of CBF estimates with respect to reference CBF values, computed from the complete perfusion-weighted image through the use of the Buxton kinetic model. This integrates the image denoising and CBF parameter estimation steps into a single pipeline allowing the network to generate better estimates of perfusion-weighted images by reducing noise and artifacts.

Network Architecture. Figure 1 depicts the architecture of our network. The network takes 2D noisy gray image patches as input and residual image patches as output. Our network consists of eight consecutive 2D convolutional layers followed by parametric rectified linear units (PReLU) activation. Although ReLU activation has been reported to achieve good performance in denoising tasks [9,14], we empirically obtained better results on our ASL dataset using PReLU in which negative activation is allowed through a small non-zero coefficient that can be adaptively learned during training [6]. The number of filters in every convolutional layer is set to 48 with a filter size of 3×3. Following eight consecutive layers, we apply one last convolutional layer without any activation function. The last layer only includes one convolutional filter, and its output is considered as the estimated residual image patch.
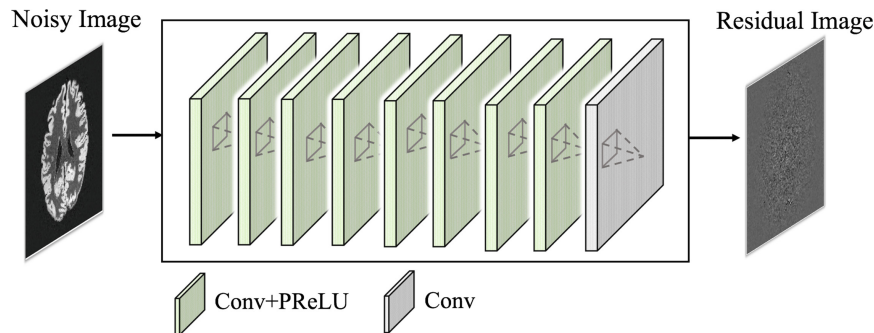
Fig. 1. The architecture of the proposed network used for the estimation of a residual image from the noisy perfusion-weighted image given as input.

Training. Training was performed using 18000 noisy and residual patch pairs of size 40×40. The network was trained using the Adam optimizer with a learning rate of 10^{-4} for 200 epochs and mini-batch size of 500. We trained a single CNN model for denoising the noisy input images from different noise levels. Inference on test data was also performed in a patch-wise manner.

3   Experiments and Results

Datasets. Pseudo-continuous ASL (pCASL) images were acquired from 5 healthy subjects on a 3T MR scanner with a 2D EPI readout using the following acquisition parameters (TR/TE = 5000/14.6 ms, flip angle = 90°, voxel

size $= 2.7 \times 2.7 \times 5$ mm^3, matrix size $= 128 \times 128$, 17 slices, labeling duration (τ) $= 1800$ ms, post-label delay (PLD) $= 1600$ ms). 30 C/L pairs and one SI$_{PD}$ image were acquired for each subject.

Additionally, high resolution synthetic ASL image datasets were generated for each real subject based on the acquired SI$_{PD}$ and coregistered white-matter (WM) and grey-matter (GM) partial volume content maps. To create a ground-truth CBF map, we assigned the CBF values of 20 and 65 mL/100g/min to the WM and GM voxels respectively, as reported in [12]. To generate synthetic data with a realistic noise level, the standard deviation over 30 repetitions was estimated from the acquired C/L images for each voxel. We subsequently added Gaussian noise with estimated standard deviation to each voxel of the synthetic images. This step was repeated 100 times to create a synthetic data per subject containing 100 C/L pairs. For synthetic data, we set $\tau = 1600$ ms and $PLD = 2200$ ms. All the other constant variables in (1) were fixed based on the recommended values for pCASL given in [2].

Fig. 2. Visual comparison of denoising results (top) and resulting CBF maps (bottom) on an examplary synthetic data using 20% of 100 pairwise subtractions. Corresponding PSNR and RMSE values calculated with respect to references are also displayed at top-left corner of each image estimate. The proposed method can yield the best results both qualitatively and quantitatively.

Data Preprocessing. Prior to training the network, the standard preprocessing steps (motion correction, co-registration, Gaussian smoothing with 4 mm kernel size) [2] were applied on C/L pairs using our in-house toolbox implementation for ASL analysis. The top and bottom slices of each subject were removed from the analysis due to excessive noise caused by motion correction.

Data augmentation was applied on every 2D image slices using rigid transformations. After augmentation, every image was divided into non-overlapping 2D patches of size 40×40, leading to 5440 patches per subject. This process was repeated for input, target, and other variables required for network training.

For each subject, we consider four different noise levels obtained by averaging randomly selected 20%, 40%, 60% and 80% of all available C/L repetitions, all of which were used during training and also tested on the trained network.

Experimental Setup. All experiments were performed using the leave-one-subject-out fashion. The synthetic and in-vivo models were trained and tested separately. In order to show the benefit of our proposed method, we compare it with the recent deep learning based denoising method [9] for ASL. Throughout the paper we refer to this method as *Dilated Conv.* For this network we use exactly same dilation rates and number of filters as suggested in the paper, and evaluate it using mean-squared-error (MSE) loss during training. We employ the peak signal-to-noise ratio (PSNR) to quantitatively assess the quality of image denoising, and the root-mean-squared error (RMSE) and Lin's concordance correlation coefficient (CCC) to assess the accuracy of CBF parameter estimation. We run the experiments on a NVIDIA GeForce Titan Xp GPU, and our code was implemented using Keras library with TensorFlow [1] backend.
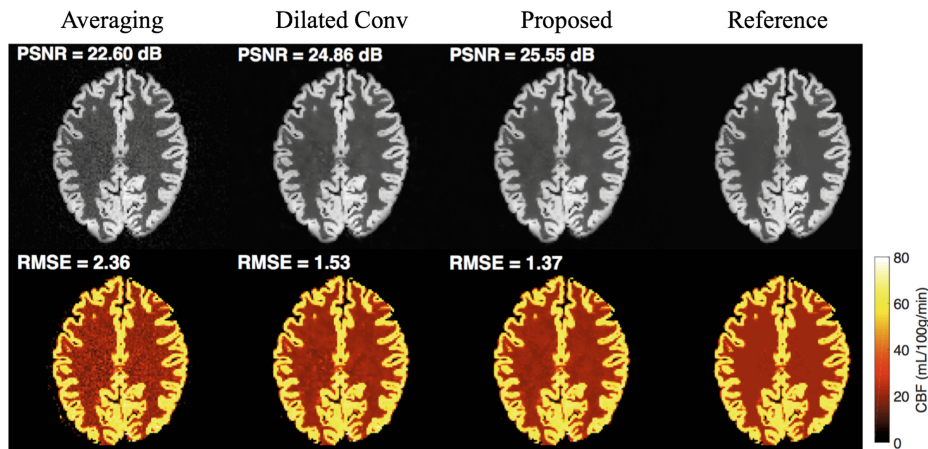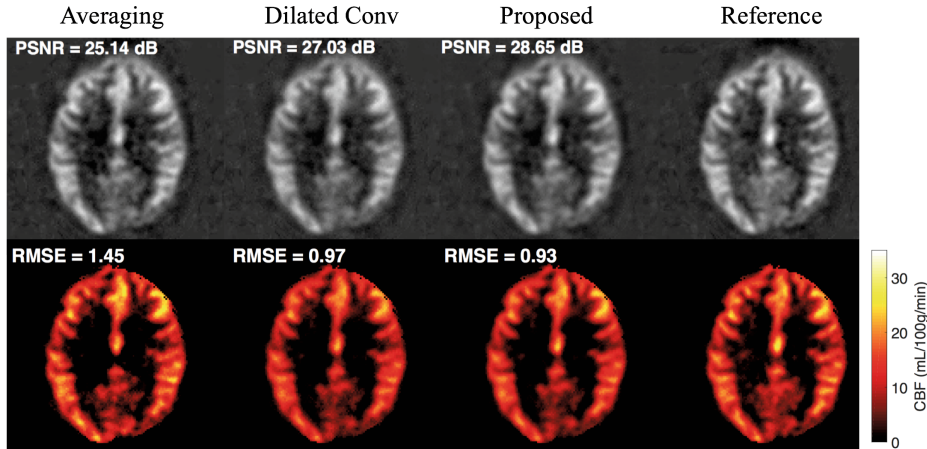

Fig. 3. Visual comparison of denoising results (top) and resulting CBF maps (bottom) on an examplary real data using 40% of 30 pairwise subtractions. Although the estimated images qualitatively look similar, the quantitative metrics calculated inside the brain demonstrates the better performance of the proposed method.

Results. Figure 2 demonstrates the denoised images and corresponding CBF maps of an exemplary slice of a synthetic dataset. Here, only 20% of 100 synthetic C/L subtractions were used. Our proposed model produces the highest quality perfusion-weighted images where noise inside the brain is significantly removed compared to conventional averaging. The resulting CBF map of our proposed method is also closer to the reference CBF map yielding the lowest RMSE score.

In Fig. 3 we present the qualitative results from a real subject's data using 40% of 30 C/L subtractions. Although the proposed method achieves the best PSNR and RMSE for perfusion-weighted image and CBF map respectively, the improvement against conventional averaging is less apparent compared to the
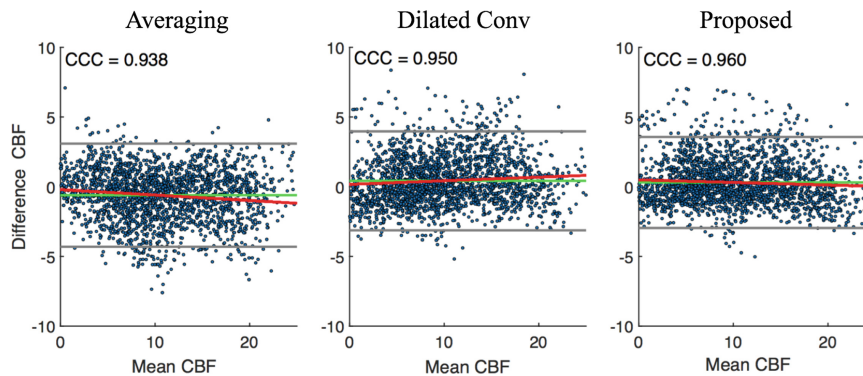
Fig. 4. Bland-Altman plots of different methods obtained in a grey-matter region of a real subject's data. Differences in CBF (y-axis) between the reference and compared method is plotted against the mean (x-axis) values of the two. The unit for horizontal and vertical axes are in ml/100g/min. Solid green lines indicate mean difference. Solid gray lines at top and bottom correspond to upper and lower margins of 95% limits of agreement. Linear regression lines are also shown with red solid lines. Corresponding CCC values are displayed at top-left corner of each plot.

synthetic data. The underlying reason is that as it can be clearly seen in Fig. 3, our reference perfusion-weighted images obtained by averaging all 30 C/L subtractions still suffer from significant noise and artifacts. Since we train our network using these images as target, the network cannot produce results that show better quality beyond the reference images. The Dilated Conv method also faces similar problem for real data. Figure 4 depicts the Bland-Altman plots of CBF values in GM tissue obtained from different methods using a real subject's data. The plots indicate that our proposed method can yield better fidelity of CBF estimation with smaller bias (green solid line) and variance (difference between solid grey lines). The linear regression line (solid red) fitted in the averaging method also shows a systematic underestimation error whereas this error is considerably reduced by the proposed method where the regression line is closer to a straight line, $y = 0$. Note that all three methods contain outlier voxels caused due to excessive noise and artifacts observable in most of the C/L subtractions.

We also quantitatively compare the predicted results in Table 1 in terms of PSNR, RMSE and CCC. Our proposed method outperforms other competing methods in all the metrics when either $\lambda = 0.2$ or $\lambda = 0.5$, which further demonstrates the advantage of the incorporation of CBF estimation model in denoising step. Taking into account data from all subjects, the differences between PR-$\lambda = 0.2$ and the Dilated Conv method on synthetic dataset are statistically significant with $p \ll 0.05$ for all metrics. The differences are also statistically significant on real dataset for PSNR and RMSE, but not significant for CCC with $p = 0.1388$. Finally, we emphasize that image denoising using our trained network takes approximately 5 ms on a single slice of matrix size 128×128.

Table 1. Quantitative evaluation in terms of $mean(std)$ obtained by different methods using all the subjects for synthetic and real datasets. The best performances are highlighted in bold font. All the metric values are calculated inside the brain region. Note that PR-$\lambda = x$ denotes our proposed method when λ value is set to x.

Method	Synthetic dataset			Real dataset		
	PSNR	RMSE	CCC	PSNR	RMSE	CCC
Averaging	20.3(4.81)	2.20(2.51)	0.88(0.06)	23.6(6.16)	1.49(0.80)	0.85(0.07)
Dilated Conv	25.2(5.09)	1.48(1.06)	0.93(0.05)	24.2(5.90)	1.41(0.72)	0.87(0.06)
PR-$\lambda = 0.2$	**28.0(3.82)**	**1.33(0.79)**	**0.95(0.04)**	**25.1(5.36)**	**1.37(0.65)**	**0.88(0.05)**
PR-$\lambda = 0.5$	26.9(4.23)	1.40(0.84)	**0.95(0.04)**	24.3(5.35)	1.38(0.67)	**0.88(0.06)**
PR-$\lambda = 0.7$	25.6(6.07)	1.51(1.76)	0.94(0.06)	24.0(5.39)	1.39(0.69)	**0.88(0.06)**
PR-$\lambda = 1.0$	25.3(5.62)	1.49(1.56)	0.93(0.05)	23.9(6.00)	1.42(0.70)	0.87(0.06)

4   Conclusion

We have proposed a novel deep learning based method for denoising ASL images. In particular, we utilize the Buxton kinetic model for CBF parameter estimation as a separate loss term where the agreement with reference CBF values is simultaneously enforced on the denoised perfusion-weighted images. Furthermore, we adopt the residual learning strategy on a deep FCN which is trained to learn a single model for denosing images from different noise levels. We have validated the efficacy of our method on synthetic and in-vivo pCASL datasets. Future work will aim at extending our work to perform denoising on multi-TI ASL data where the estimation of the arterial transit time (ATT) parameter can be also exploited in the loss function.

References

1. Abadi, M., et al.: TensorFlow: large-scale machine learning on heterogeneous systems (2015). https://www.tensorflow.org/, software available from tensorflow.org
2. Alsop, D.C.: Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. MRM **73**(1), 102–116 (2015)
3. Bibic, A., et al.: Denoising of arterial spin labeling data: wavelet-domain filtering compared with gaussian smoothing. MAGMA **23**(3), 125–137 (2010)
4. Buxton, R.B., et al.: A general kinetic model for quantitative perfusion imaging with arterial spin labeling. MRM **40**(3), 383–396 (1998)
5. Fang, R., et al.: A spatio-temporal low-rank total variation approach for denoising arterial spin labeling MRI data. In: IEEE ISBI, pp. 498–502, April 2015
6. He, K., et al.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: IEEE ICCV, pp. 1026–1034, December 2015

7. He, K., et al.: Deep residual learning for image recognition. In: IEEE CVPR, pp. 770–778, June 2016
8. Kiku, D., et al.: Residual interpolation for color image demosaicking. In: IEEE ICIP, pp. 2304–2308, September 2013
9. Kim, K.H., et al.: Improving arterial spin labeling by using deep learning. Radiology **287**(2), 658–666 (2018)
10. Liang, X.: Voxel-wise functional connectomics using arterial spin labeling functional magnetic resonance imaging: the role of denoising. Brain Connect. **5**(9), 543–53 (2015)
11. Owen, D., et al.: Anatomy-driven modelling of spatial correlation for regularisation of arterial spin labelling images. In: Descoteaux, M., et al. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 190–197. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_22
12. Spann, S.M., et al.: Spatio-temporal TGV denoising for ASL perfusion imaging. Neuroimage **157**, 81–96 (2017)
13. Wells, J.A., et al.: Reduction of errors in ASL cerebral perfusion and arterial transit time maps using image de-noising. MRM **64**(3), 715–724 (2010)
14. Zhang, K.: Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017)

8

Conclusions and Outlook

This dissertation essentially explored advanced techniques for image enhancement – involving both reconstruction and denoising – and parameter reconstruction problems for perfusion MR imaging. In this chapter we will present a brief summary and general discussion of each thesis contributions, specifically highlight their importance as well as their limitations, and will discuss possible directions for future research.

Accelerated dynamic MR data reconstruction

In Chapter 3 we have presented a new reconstruction approach for the fast reconstruction of dynamic MRI data from undersampled k-space measurements. By jointly integrating two fundamentally different constraints, which enforces not only local coherences at the pixel-level but also global correlation in the full spatio-temporal domain, our proposed approach can iteratively reconstruct the full dynamic MRI sequences. In comparison to the state-of-the-art iterative reconstruction techniques, our method demonstrates superior performance in terms of reconstruction accuracy and image quality when test on *in-vivo* 3D cardiac and brain MRI datasets.

Nevertheless, our approach exhibits a few limitations. First, despite working with real imaging data, the undersampling was applied in retrospective manner, where we simulated the undersampling from fully sampled data. Ideally, a proper evaluation of our reconstruction method should be done with true data distribution, i.e. from raw undersampled data which are prospectively sampled on the MR scanner. It is worth mentioning that several previous works [22, 120, 52] have shown that the reconstruction performance on prospectively undersampled data usually align well with retrospectively undersampled data unless there is a high amount of motion and imaging artifacts introduced during the acquisition of prospectively sampled data. Second, the computational time of reconstructing the entire 3D dynamic sequence is typically long, i.e. on the

order of minutes or sometimes hours depending on the data size and efficiency of solver of the optimization problem. In our case, it takes roughly 3 minutes to reconstruct a 3D dynamic sequence of size $128 \times 128 \times 10$. Another relevant weakness of our method is that it requires fine-tuning of the regularization parameters for each regularization term. Finding the optimal values of these parameters is not an easy task because it heavily depends on various factors, including the image data type, sampling trajectories, acceleration factor, etc. It is usually good practice to apply exhaustive grid search based techniques to find the optimal combinations of the hyper-parameters. Inappropriate settings of the regularization parameters generally yield poor reconstruction quality.

Over the last three years, deep learning (DL) based networks [121, 122, 123, 124] have been predominantly used for the dynamic MRI reconstruction task and have been shown to outperform the conventional CS-based dynamic MRI reconstruction techniques. The main advantages of DL based data-driven models over the CS-based iterative reconstruction techniques are as follows: (i) The reconstruction of full dynamic sequence is significantly faster once a network is trained, i.e. on the order of miliseconds or seconds. (ii) The intrinsic relation between the undersampled data and fully-sampled (high quality) images are directly learned from a large amount of training data, meaning that one does not need to find the optimal sparsifying transform or constraints for the given dynamic MR data. To this end, exploring the deep neural networks (NNs) on a prospectively undersampled dynamic MRI might be an interesting direction of future work. However, due to the intrinsic generalization property of deep NNs, one needs to further investigate the reconstruction quality around the specific pathological tissue regions.

Perfusion MR data reconstruction

We proposed a reconstruction model specifically designed for recovering perfusion MR image series from undersampled k-space data in Chapter 4. Appendix A additionally presents the asssesment of proposed model with the estimations of perfusion related parameters obtained from clinical DSC and DCE MRI brain perfusion sequences. The essence of our model is based on the observation that the rapid signal contrast variations over time – occurring due to attenuation related to the passage of the contrast agent (CA) – generally vary in different tissues, including blood vessels, normal tissues, tumor regions, etc, appearing as small areas inside the organ being imaged. One of the key findings here is that the quality of the reconstructed images highly depends on the resolution of the acquired images, the noise introduced during acquisition and the acceleration factor of undersampling. Chapter 4 shows that 4x to

6x acceleration can yield satisfactory reconstruction quality when applied on relatively low-resolution data, whereas pushing towards 20x acceleration still enables us to obtain perfect quality (highly similar with fully-sampled images) when the images have high-resolution, as presented in Appendix B. The parameter quantification was applied as a post-processing step on the reconstructed and fully-sampled image series data, therefore the quality of image reconstruction directly reflects on the accuracy of estimated parameter maps. To be more precise, up to 8-fold acceleration our model produces very similar parameter maps with respect to the reference maps, however for the higher acceleration rates we start observing oversmooth regions in different brain areas of the parameter maps. Furthermore, it is worth mentioning that as the proposed approach is based on an iterative reconstruction scheme, it inherently possesses the above-mentioned drawbacks of our dynamic MRI reconstruction model.

One potential direction of future research on this topic might be using the recurrent neural network based DL models to exploit the full temporal dynamics of the perfusion signal. Several approaches towards this direction [125, 126] have been recently applied on undersampled cardiac cine MRI data for motion estimation and compensation as well as reconstruction of full dynamic sequence. As the temporal variations in perfusion image series occur essentially at the local regions of the organ, it is benefical to train the network on small image patches instead of entire images. Furthermore, a novel loss function capturing the complete temporal perfusion signal information can be incorporated when training the network.

Direct reconstruction/inference of pharmacokinetic parameters of DCE MRI

In Chapter 5 we presented a deep learning based approach to directly estimate the pharmacokinetic (PK) parameters from undersampled DCE-MRI data. We treat the parameter inference from undersampled data in DCE imaging as a mapping problem between the corrupted image-time series and residual parameter maps where the underlying mapping is learned uing deep CNNs. One of the main contribution of this work is that by integrating the forward physical model into the loss function, which relates the PK parameters to the corrupted perfusion signal, our proposed model can allow the network to inrinsically exploit the underlying contrast agent kinetics in training phase, thus provide more accurate restoration of PK parameters. Results on 10-fold undersampled *in-vivo* brain DCE-MRI dataset demonstrates high similarity with the reference parameter maps obtained from fully-sampled data. The

key benefits of our approach are as follows. First, our method serves as a direct parameter reconstruction approach where the image reconstruction step is completely eliminated. Second, parameter inference at testing stage takes significantly less time, i.e. around 1-2 seconds on a 3D DCE volume, in comparison to state-of-the-art iterative model-based direct reconstruction method.

Following to the above-mentioned work, we further targeted to directly infer the PK parameters from the observed signal intensity over time, where no undersampling was applied on the original DCE image series. As provided in Chapter 6, we presented a comprehensive assessment of the quality of estimated parameters from our model on brain DCE datasets acquired from patients with clinically evident mild ischaemic stroke. The results demonstrated several advantages of our approach over the traditional model fitting based parameter estimation techniques, including (i) avoiding a few intermediate computation steps of quantification of PK parameters, (ii) increased robustness to signal noise and outliers, (iii) enabling statistically significant differences of parameter values between different tissue types.

Nevertheles, we should point out the several limitations of both of these works. First, our approach works only for the situation where both the training and test data were acquired with the same acquisition protocol and parameters, ideally acquired in a population study. Especially when the subject-specific AIFs vary significantly across different subjects or dataset, this should be handled appropriately, otherwise paramater inference on a *unseen* test data may provide quite poor results. Second, further investigation on synthetic data is required to perform more precise assessment of error and bias when the *ground-truth* parameter values are known. Third, due to the strong generalization ability of our DL model, we partly observe the problem of overestimated parameter values over the highly perfused brain tissues, e.g. in white matter hyperintensities or stroke lesion. The overestimation within these tissue regions is presumably caused by the relatively lower paramater values estimated by reference kinetic model fitting (e.g. Patlak model) due to significant signal noise and fitting to the local minima of the NLLS. However, these observations neccessiate a deep investigation of the parameter estimations within the critical regions of interest, performed by the expert radiologists. Until clinical trials are performed, the true assessment of the utility of our proposed method remain unclear.

Following to our first DL-based contributions [87, 88] to this field, several other works have been recently proposed mainly considering different network architectures [127, 128] and uncertainity estimation of parameter inference [129]. One of the interesting direction of future work related to this topic might be

designing a single DL network which can be trained and tested on a mixed dataset – ideally involving DCE image series acquired with different acquisition parameters and protocols – to investigate the possibility and efficacy of an acquisition protocol-agnostic network. Related to this idea, an emerging field of federated learning based approaches [130, 131] can be further exploited without centralising the mixed dataset if each data must reside in different sites to avoid data governance and privacy problems.

Denoising arterial spin labeled MRI

As presented in Chapter 7, we proposed a novel DL based framework for denoising the perfusion-weighted image quality obtained from a subset of available pairwise control/label image subtractions in ASL perfusion imaging. Our key contribution here is that we follow a mixed modeling approach in our denoising scheme by incorporating a formal representation of the CBF signal model as a prior knowledge in our network, more precisely, we use the Buxton kinetic model as a separate term of the loss function during training stage. This allows us to produce high quality perfusion-weighted images while simultaneously enforcing the network to estimate CBF maps that are highly similar with reference CBF maps. In comparison to the first DL-based work [132] appeared in this domain, we demonstrated the efficiency and superior performance of our proposed denoising model using both synthetic and *in-vivo* single-TI (or post-label delay) ASL datasets. Another main benefit of our work is that we showed how a single CNN model can be effectively learned to denoise perfusion-weighted images subject to different noise levels. However, as a main drawback of this work, our model was designed to primarily work on a single-TI setting, where all ASL images are acquired with the similar acquisition parameters, including a constant post-label delay, label duration and labeling efficiency. Extending this work to learn a parameter-agnostic DL network, which can effectively denoise perfusion-weighted images acquired with different acquisiton protocols and parameters, might be an interesting future work.

Following to our published work [89], a few other studies have been recently proposed on single-TI ASL denoising [133, 134], artifact suppression [135, 136] and super-resolution [137] using different DL network architectures. Nevertheless, until now researchers have not paid enough attention to develop denoising methods directly applied on multi-TI ASL data even though their utility and effectivenes are much more prominent when considering significantly high scan duration of a complete multi-TI acquisition session. Moreover, multi-TI acquisitions provide estimation of additional perfusion parameters, e.g. arterial transit

time (ATT), and more precise estimation of CBF. Therefore, as an interesting line of future work, we currently work on estimating high-quality and realistic perfusion parameters directly inferred from noisy perfusion-weighted images acquired at multiple post-label delays or inversion times, where we explore to adapt generative adversarial networks (GANs) framework.

Appendices

Accelerated Reconstruction of Quantitative Perfusion-weighted MRI

C. Ulas, C. Preibisch, J. Sperl, T. Pyka, J. Kalpathy-Cramer, and B. Menze. "Accelerated Reconstruction of Perfusion-Weighted MRI Enforcing Jointly Local and Nonlocal Spatio-temporal Constraints." In: *arXiv* 1708.07808 (2017)

Abstract: Perfusion-weighted magnetic resonance imaging (MRI) is an imaging technique that allows one to measure tissue perfusion in an organ of interest through the injection of an intravascular paramagnetic contrast agent (CA). Due to a preference for high temporal and spatial resolution in many applications, this modality could significantly benefit from accelerated data acquisitions. In this paper, we specifically address the problem of reconstructing perfusion MR image series from a subset of k-space data. Our proposed approach is motivated by the observation that temporal variations (dynamics) in perfusion imaging often exhibit correlation across different spatial scales. Hence, we propose a model that jointly penalizes the voxel-wise deviations in temporal gradient images obtained based on a baseline, and the patch-wise dissimilarities between the spatio-temporal neighborhoods of entire image sequence. We validate our method on dynamic susceptibility contrast (DSC)-MRI and dynamic contrast enhanced (DCE)-MRI brain perfusion datasets acquired from 10 tumor patients in total. We provide extensive analysis of

reconstruction performance and perfusion parameter estimation in comparison
to state-of-the-art reconstruction methods. Experimental results on clinical
datasets demonstrate that our reconstruction model can potentially achieve
up to 8-fold acceleration by enabling accurate estimation of perfusion param-
eters while preserving spatial image details and reconstructing the complete
perfusion time intensity curves (TICs).

Contributions of thesis author: Algorithm design and implementation,
computational experiments and interpretation, composition and revision of
manuscript.

Accelerated Reconstruction of Perfusion-Weighted MRI Enforcing Jointly Local and Nonlocal Spatio-temporal Constraints

Cagdas Ulas*, Christine Preibisch, Jonathan Sperl, Thomas Pyka, Jayashree Kalpathy-Cramer and Bjoern Menze

Abstract—Perfusion-weighted magnetic resonance imaging (MRI) is an imaging technique that allows one to measure tissue perfusion in an organ of interest through the injection of an intravascular paramagnetic contrast agent (CA). Due to a preference for high temporal and spatial resolution in many applications, this modality could significantly benefit from accelerated data acquisitions. In this paper, we specifically address the problem of reconstructing perfusion MR image series from a subset of k-space data. Our proposed approach is motivated by the observation that temporal variations (dynamics) in perfusion imaging often exhibit correlation across different spatial scales. Hence, we propose a model that jointly penalizes the voxel-wise deviations in temporal gradient images obtained based on a baseline, and the patch-wise dissimilarities between the spatio-temporal neighborhoods of entire image sequence. We validate our method on dynamic susceptibility contrast (DSC)-MRI and dynamic contrast-enhanced (DCE)-MRI brain perfusion datasets acquired from 10 tumor patients in total. We provide extensive analysis of reconstruction performance and perfusion parameter estimation in comparison to state-of-the-art reconstruction methods. Experimental results on clinical datasets demonstrate that our reconstruction model can potentially achieve up to 8-fold acceleration by enabling accurate estimation of perfusion parameters while preserving spatial image details and reconstructing the complete perfusion time-intensity curves (TICs).

Index Terms—Perfusion-weighted magnetic resonance imaging, reconstruction, tracer kinetic modeling, acceleration

I. INTRODUCTION

STUDYING blood flow and blood flow patterns is a major field in clinical radiology and diagnostics. Perfusion-weighted magnetic resonance imaging (MRI) provides a mean for assessing tissue perfusion and vascular permeability *in vivo* through examination of the spatio-temporal changes of signal intensities following the injection of an exogenous paramagnetic contrast agent (CA) [1]. These techniques have become valuable clinical tools since they play a crucial role, for instance, in the diagnosis of stroke, the determination of tissue(s) at risk of infarction, and the prediction of prognosis after treatments of patients with stroke and tumors [2]. Two of the most common methods used in perfusion-weighted imaging (PWI) are dynamic susceptibility contrast MRI (DSC-MRI) and dynamic contrast enhanced MRI (DCE-MRI).

C. Ulas* and B. Menze are with the Department of Computer Science at the Technical University of Munich, Germany (e-mail: cagdas.ulas@tum.de).
C. Preibisch and T. Pyka are with the Department of Neuroradiology at the Technical University of Munich, Germany.
C. Ulas* and J. Sperl are with GE Global Research, Garching, Germany.
J. Kalpathy-Cramer is with the Department of Radiology at Harvard Medical School and the Department of Neuroscience at Massachusetts General Hospital, Boston, USA.

Both techniques require intravenous bolus administration of gadolinium, followed by the acquisition of successive images as the contrast bolus enters and subsequently leaves the organ of interest. DSC-MRI relies on dynamic alterations of the T_2^* transverse relaxation times of tissues and it is employed to assess the hemodynamic status of tissues [3]. DCE-MRI on the other hand relies on changes of the T_1 longitudinal relaxation times of the tissues. DCE-MRI is widely used to interrogate the vascular characteristics of tumors in clinical settings [2].

Vast majority of clinical research on PWI have considered the problem of estimating accurate voxel-wise perfusion parameters which are generally obtained by fitting a tracer kinetic model to the observed time-intensity curves (TICs) [4]. An illustration displaying the major steps of kinetic parameter estimation in DSC-MRI is provided in supplementary material.

One of the major obstacles in the clinical use of PWI is the immense need of high temporal resolution to capture the rapid contrast changes of CA uptake for precise perfusion quantification [1]. Furthermore, the short scanning time available for each frame often leads to limited spatial resolution to detect small image features and accurate tumor boundaries, and low signal-to-noise ratio (SNR) to enable precise fitting of kinetic model parameters. Considering such severe constraints, PWI can benefit from subsampled acquisitions [5]. However, sub-Nyquist sampling often results in aliasing artifacts in image space and in the context of PWI, reconstruction of complete temporal signal with its peak and high dynamics (observable in blood vessels) constitutes even a more challenging problem.

Recently, various reconstruction approaches have been proposed in related dynamic imaging applications, based on, such as piece-wise smoothness in the spatial domain [6], [7], high correlation and sparsity in the temporal domain [7], [8], [9], [10], sparse representations of local image regions via learned dictionaries [10], [11] and low-rank property of MR sequences in the full spatio-temporal space [8], [12], [6]. However, there are only a few works dedicated directly to reconstruction problem in PWI, considering the constraint of the image frames based on a baseline (pre-contrast) image [13], the penalization of time curves with high temporal gradients [14], and the minimization of temporal finite-differences enforced together with multiple constraints on spatial domain [7], [15]. The main limitation of these methods is that they consider the temporal variations only in single voxel level and neglect the similarities and variations between the voxels located in a spatially close neighborhood. For this reason, their performance is very sensitive to rapid signal changes occurring over the duration of

image acquisition as encountered in PWI. These methods often produce blurry image regions and oversmooth reconstruction of TICs, and ultimately result in underestimated peak value of the perfusion signal, which substantially deteriorates the accuracy of estimated perfusion parameters. To this end, an optimal choice of the reconstruction model is essential for PWI targeting to recover the complete temporal pattern of the perfusion TICs while preserving the spatial quality of image series. We believe that a reconstruction model satisfying such conditions can be used to accelerate the acquisition of PWI and practically allow to acquire more samples in time domain and improve the volume coverage of the organ of interest.

In perfusion imaging, we observe rapid signal (contrast) variations over time due to the attenuation related to the passage of the contrast agent. Depending on the level of perfusion incurred by the tracer, these variations mainly differ in tissues, blood vessels, and tumor regions, etc., appearing as small local areas inside the organ to be imaged. Motivated by this observation, we propose a new reconstruction model specifically for PWI. Our model primarily integrates two fundamentally different data-driven constraints: (i) a voxel-wise local sparsity constraint on the temporal gradient images with respect to a baseline, limiting the overall dynamic of the perfusion time series, and (ii) a patch-wise similarity constraint on the spatio-temporal neighborhoods of the entire perfusion image series, providing smooth spatial regions with better alignment to temporal variations in small local areas represented with patches. We formulate the main optimization problem in a joint formal framework and introduce a new proximal splitting strategy [16] that benefits from the weighted-average of proximals, and can efficiently solve the joint minimization problem with fast convergence. The proposed method is validated on DSC and DCE-MRI perfusion datasets collected from brain tumor patients and compared with existing reconstruction methods. Extensive experiments demonstrate the efficiency of our method in terms of reconstruction performance and estimation of perfusion parameters from accelerated acquisitions. To the best of our knowledge, this is the first work to exploit the spatial and temporal variations jointly at different scales for the purpose of reconstruction of PWI, successfully applied on both DSC and DCE-MRI time series.

Preliminary results of this work presented at a conference [17] are herein extended by additional validation on DCE-MRI datasets collected from a clinical cohort of glioma patients, assessment of our method with estimated hemodynamic and pharmacokinetic parameters underlying perfusion, and experimental analysis of convergence properties of the proposed algorithm. The key contributions of this paper can be summarized as follows:

- We present a robust reconstruction method for PWI dynamic series. Our proposed model exploits the spatio-temporal variations jointly at single voxel and patch level.
- The proposed reconstruction model can enable accurate quantification of clinically useful perfusion parameters while attaining up to 8-fold acceleration through the use of only a subset of k-space measurements.
- We introduce a formal iterative algorithm to solve the minimization of the sum of convex regularizers based on

proximal-splitting applied to a reconstruction problem.

The remainder of this paper is organized as follows. Section II provides the detailed description of the proposed reconstruction model along with its formulation and the algorithm to solve the reconstruction problem. In Section III, we briefly describe the acquisition parameters of our clinical perfusion datasets and employed tracer kinetic models for data analysis. Section IV presents the experimental setup and the results of conducted experiments. After a general discussion we provide the concluding remarks in Section V.

II. RECONSTRUCTION MODEL

Our proposed reconstruction model jointly imposes two spatio-temporal constraints both on a voxel-wise (local) and patch-wise (nonlocal) level as illustrated in Fig. 1. In the following sections we provide the intuition behind using specifically these constraints and describe how to mathematically formulate the joint regularization problem along with the algorithm that efficiently solves the optimization.

A. Formulation

We remark that throughout the paper we describe our method on 2D + t data only for simplicity of the presentation. However, a generalization to 3D + t volumes is also straightforward. We assume that $X \in \mathbb{R}^{N_x \times N_y \times T}$ is a 2D perfusion MR image series (sequence) represented as spatio-temporal 3D data involving a total number of $N = N_x \times N_y \times T$ voxels. Let $x_t \in \mathbb{R}^{N_x \times N_y}$ denote a perfusion MR frame at time t, y_t is the acquired undersampled k-space measurement of x_t, and $\mathbb{T} = \{1, 2, ..., T\}$ is the set of frame number indices in the sequence. The main objective is to reconstruct all x_t's from the acquired k-space measurements y_t's. The physical model between x_t and y_t can be formulated as,

$$y_t = A_t(Fx_t + \eta), \quad (1)$$

where A_t represent the sampling matrix to acquire only a subset of k-space samples, F is the 2D Fourier Transform operator and η is additive Gaussian noise in k-space. We denote the partial 2D Fourier operator for frame t as $\mathcal{F}_t = A_t F$, and stack the $\mathcal{F}_t$'s for all frames of the sequence as $\mathcal{F}_u = \text{diag}\{\mathcal{F}_1, \mathcal{F}_2, .., \mathcal{F}_T\}$. The investigated problem in (1) is an ill-posed inverse problem [6], [10]. Regularization is often required to find a unique and stable solution to such problems.

We pose the joint regularization for the reconstruction of perfusion image series as the following optimization problem,

$$\hat{X} = \arg\min_X \frac{1}{2}\|\mathcal{F}_u X - Y\|_2^2 + \lambda_1 \mathcal{R}_L(X) + \lambda_2 \mathcal{R}_{NL}(X), \quad (2)$$

where X denotes the entire perfusion image series and Y represents the corresponding k-space measurements. The first term in (2) ensures data consistency, $\mathcal{R}_L$ and $\mathcal{R}_{NL}$ are two regularization terms imposed on reconstruction, λ_1 and λ_2 are the tuning parameters for these penalty terms.

Local ($\mathcal{R}_L$) regularizer: This regularizer penalizes the sum of voxel-wise intensity differences in temporal gradients calculated based on a reference for every frame x_t, defined as,

$$\mathcal{R}_L(X) = \sum_{t \in \mathbb{T}} \sum_{n=1}^{N_x \times N_y} \sqrt{(\nabla_x(x_t - \bar{x})_n)^2 + (\nabla_y(x_t - \bar{x})_n)^2},$$

where $\bar{x}$ is a reference (baseline) image, ∇_x and ∇_y represent finite-difference operators along x and y dimensions, respectively. This penalty is termed as dynamic total variation (DTV) and was orignally proposed for online reconstruction [9]. Here we employ DTV in offline manner, where all the frames are available. This term is better adjusted to the variations in time since it explicitly enforces temporal coherence by minimizing the difference with respect to a reference for every frame. Presumably, if there are deviations from the baseline intensities, they should be spatially homogeneous, i.e., a block of neighboring voxels should exhibit the same amount of deviation. Intuitively, this regularizer serves as a penalty on the large deviations from a baseline perfusion signal, enabling to smooth extreme local image regions.

Nonlocal ($\mathcal{R}_{\mathrm{NL}}$) regularizer: This regularizer penalizes the weighted sum of ℓ_2 norm distances between spatio-temporal neighborhoods (patches) of the image series and we use it as a fully 3D nonlocal scheme. The term is specified by [18],

$$\mathcal{R}_{\mathrm{NL}}(X) = \sum_{\mathbf{p}\in\Omega}\sum_{\mathbf{q}\in\mathcal{N}_{\mathbf{p}}} \varphi(\mathbf{p},\mathbf{q})\|P_{\mathbf{p}}(X) - P_{\mathbf{q}}(X)\|_2^2,$$

where $\mathbf{p} = (\mathrm{p}_x, \mathrm{p}_y, \mathrm{p}_t)$ and $\mathbf{q} = (\mathrm{q}_x, \mathrm{q}_y, \mathrm{q}_t)$ are two voxels, and the voxel of interest is $\mathbf{p} \in \Omega$, where $\Omega = [0, N_x] \times [0, N_y] \times [0, T]$. The term $P_{\mathbf{p}}(X)$ denotes a spatio-temporal 3D patch of the image sequence X, centered at voxel $\mathbf{p}$. We represent $\mathcal{N}_{\mathbf{p}}$ as a 3D search window around voxel $\mathbf{p}$. We simply denote N_p and N_w as the size of a patch and search window, respectively. The size of the patch must be smaller than the size of the search window. The weights $\varphi(\mathbf{p},\mathbf{q})$ are calculated based on Euclidean distance between the patches,

$$\varphi(\mathbf{p},\mathbf{q}) = \exp\left(-\|P_{\mathbf{p}}(X) - P_{\mathbf{q}}(X)\|_2^2/h^2\right), \quad (3)$$

where h is a parameter controlling the decay of the exponential function. The exponential weighting favors the similar patches in terms of Euclidean distance by assigning higher weight to their center voxels. Intuitively, this regularizer can exploit the similarities between patch pairs in adjacent frames and enforce smooth solutions in the spatio-temporal neighborhoods of the MR sequence even in the presence of significant inter-frame signal changes and high noise introduced during acquisition.

B. Algorithm

The algorithm solving the primal problem (2) is mainly based on a proximal-splitting framework. For the better understanding of our algorithm, we first start with the definition of a proximal map.

Proximal map: Given a continuous convex function $g(x)$ and a scalar $\rho > 0$, the proximal operator associated to convex function g can be defined as [19]

$$prox_\rho(g)(z) := \underset{x\in\mathcal{H}}{\arg\min}\ \frac{1}{2\rho}\|x - z\|_2^2 + g(x). \quad (4)$$

Concretely, this operator serves as an individual minimizer for the convex function g, which approximates a value close to a reference point z [20].

The reconstruction problem in (2) can be reformulated as the following denoising problem,

$$\hat{X} = \underset{X}{\arg\min}\ \frac{1}{2}\|X - Z\|_2^2 + \rho\lambda_1\mathcal{R}_{\mathrm{L}}(X) + \rho\lambda_2\mathcal{R}_{\mathrm{NL}}(X), \quad (5)$$
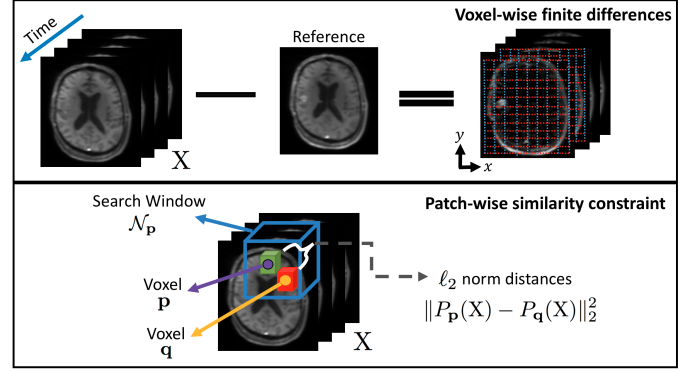

Fig. 1. Illustration of the two constraints enforced in our reconstruction model.

where $Z = \mathcal{F}_u^H Y$ and H is the conjugate transpose. Assuming that each of the regularization terms in the cost function (2) is convex, the above denoising problem can be represented as the proximal map of the sum of two regularization terms as described in [19],

$$\hat{X} = prox_\rho(\lambda_1\mathcal{R}_{\mathrm{L}} + \lambda_2\mathcal{R}_{\mathrm{NL}})(Z). \quad (6)$$

The problem (5) admits to a unique solution as provided in (6). Proximal-splitting methods can allow tractable solutions for the proximity operator of the sum of two convex functions. These methods are first-order iterative algorithms that solve relatively large-scale optimization problems with several nonsmooth penalties. They operate by splitting the main objective function into individual subproblems which can be easily evaluated via proximal operators [20].

To solve our main problem in (5), we therefore split the objective function into two individual subproblems that we term $\mathcal{R}_{\mathrm{L}}$-subproblem and $\mathcal{R}_{\mathrm{NL}}$-subproblem.

$\mathcal{R}_{\mathrm{L}}$-subproblem: The proximal map for this subproblem can be defined as,

$$prox_\rho(\lambda_1\mathcal{R}_{\mathrm{L}})(Z) = \underset{X}{\arg\min}\ \frac{1}{2\rho}\|X - Z\|_2^2 + \lambda_1\mathcal{R}_{\mathrm{L}}(X).$$

To efficiently solve this subproblem, we first reformulate it by introducing new variables $d_t = x_t - \bar{x}$, $Z_t = \mathcal{F}_t^H y_t$ and $d_g^t = Z_t - \bar{x}$, then the problem can be turned into

$$\hat{d} = \underset{\mathrm{d}}{\arg\min}\ \sum_{t\in\mathbb{T}}\left(\frac{1}{2\rho}\|d_t - d_g^t\|_2^2 + \lambda_1\|d_t\|_{TV}\right), \quad (7)$$

where $\mathrm{d} = \{d_1, ..., d_T\}$ and $\|d_t\|_{TV} = \|[Q_1 d_t, Q_2 d_t]\|_{2,1}$, where Q_1 and Q_2 are two $N_x N_y \times N_x N_y$ first order finite difference matrices in vertical and horizontal directions, $\ell_{2,1}$ norm is the sum of the ℓ_2 norm of each row of given matrix.

When a reference image $\bar{x}$ is given, the cost function in (7) can be minimized individually for every frame x_t [9]. This guarantees that the sum of the costs is also minimized. The minimization can be efficiently solved using the fast iteratively reweighted least squares (FIRLS) algorithm [21] based on preconditioned conjugate gradient method. This algorithm provides fast convergence and low computational cost by adopting a preconditioner approximated using diagonally dominant structure of the symmetric matrix $\mathcal{F}_t^H\mathcal{F}_t$. Once the problem (7) is solved, the final solution of $\mathcal{R}_{\mathrm{L}}$-subproblem is simply obtained by,

$$\hat{X}_{\mathcal{R}_{\mathrm{L}}} = \left[\hat{d}_1 + \bar{x}, \hat{d}_2 + \bar{x}, ..., \hat{d}_T + \bar{x}\right]. \quad (8)$$

Algorithm 1: Reconstruction algorithm

Input: Acquired k-space data Y, $\mathcal{F}_u$, λ_1, λ_2
Initialize: $z_1^0 = z_2^0 = \mathcal{F}_u^H Y$, w_1, w_2, $X^0 = \sum_{i=1}^{2} w_i z_i^0$, α_0, $\gamma = 1$, $k = 0$
while *stopping criteria not met* **do**

    $X_g = X^k - \gamma \mathcal{F}_u^H (\mathcal{F}_u X^k - Y)$
    $z_1^{k+1} = z_1^k + \alpha_k (prox_{\frac{\gamma}{w_1}}(2\lambda_1 \mathcal{R}_L)(X^k + X_g - z_1^k) - X^k)$
    $z_2^{k+1} = z_2^k + \alpha_k (prox_{\frac{\gamma}{w_2}}(2\lambda_2 \mathcal{R}_{NL})(X^k + X_g - z_2^k) - X^k)$
    $X^{k+1} = w_1 z_1^{k+1} + w_2 z_2^{k+1}$
    $\alpha_{k+1} = 1 + 2(\alpha_k - 1)/(1 + \sqrt{1 + 4(\alpha_k)^2})$
    $k \leftarrow k + 1$

end
Output: Reconstructed image sequence X

$\mathcal{R}_{NL}$-*subproblem*: The proximal map for this subproblem can be specified by,

$$prox_\rho(\lambda_2 \mathcal{R}_{NL})(Z) = \arg\min_X \frac{1}{2\rho}\|X - Z\|_2^2 + \lambda_2 \mathcal{R}_{NL}(X).$$

The nonlocal penalty function in this problem is nonconvex. However, it has been shown in [22] that the problem has a convex nature when the nonlocal regularization functional is assumed to be explicitly dependent on constant and pre-determined weights φ. The minimization problem here can be solved via a two-step alternating minimization scheme in an iterative projections onto convex sets (POCS) framework [23]. In each iteration, the first step projects the image estimate onto the data consistency term and the second step performs the minimization of the neighborhood penalty term on the projected data after re-estimating the weights from the current data estimate. The minimization of penalty term is equivalent to applying a non-local means (NLM) filter to the projected images [24]. The NLM filter is mathematically formulated as,

$$\hat{X}(p_x, p_y, p_t) = \frac{\sum_{(q_x, q_y, q_t) \in \mathcal{N}_\mathbf{P}} \varphi(\mathbf{p}, \mathbf{q}) X(q_x, q_y, q_t)}{\sum_{(q_x, q_y, q_t) \in \mathcal{N}_\mathbf{P}} \varphi(\mathbf{p}, \mathbf{q})}, \quad (9)$$

and essentially calculates a weighted average of closest patches in a search neighborhood and updates every voxel accordingly. To reduce the computational burden of searching closest patches, we employed an optimized blockwise version of NLM proposed by Coupé *et al.* [25].

Primal problem: After solving each subproblem[1], we adopt a generalized forward-backward splitting (GFBS) framework [16] that jointly minimizes the sum of convex functions as given in our primal problem (2). GFBS is an operator-splitting algorithm and uses a forward-backward scheme [20]. Our proposed reconstruction algorithm is outlined in Algorithm 1. The algorithm mainly involves the computation of proximals on the gradient projections in every iteration and then weighted averaging of the two resulting proximal maps with weights denoted as (w_1, w_2). We further accelerate the convergence of the algorithm with an additional acceleration step similar to the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [26]. This step adaptively updates the value of step size parameter (α_k) through iterations and make it sufficiently close to 1. The effect of adaptive α_k update on the convergence will be emprically demonstrated in Section IV-B1. The GFBS

method has been shown to converge when $\gamma < 2/L$ if the convex function $f = \frac{1}{2}\|X - X_g\|_2^2$ has a Lipschitz continuous gradient with constant L. We refer the readers to GFBS paper [16] for more details concerning the proof of convergence.

III. DATA ANALYSIS

A. Dynamic Susceptibility Contrast (DSC)

1) Data: Five glioma patients were imaged on a 3T MRI scanner using a 16-channel head neck coil. DSC-MRI image series were acquired using a 2D single-shot gradient-echo EPI sequence with parameters ($T_R = 1500$ ms, $T_E = 30$ ms, flip angle = $70°$, voxel size = $1.8 \times 1.8 \times 4$ mm^3, acquisition matrix = 128×128, 20 slices). A bolus of 15 ml Gd-DTPA (Magnevist, 0.5 mmol/ml) was injected 3 minutes after an initial first bolus of 7.5 ml with 4 ml/s injection rate. In total 60 frames were collected up to around 1.5 minutes.

2) Analysis: The signal time-intensity curves (TICs) of each voxel were directly used to estimate the CA concentration C_{raw} from the change in the gradient echo transverse relaxation rate, $\triangle R_2^*$ [4],

$$C_{raw}(t) \propto \triangle R_2^* = -(1/T_E) \log[S(t)/S(0)], \quad (10)$$

where $S(t)$ is the post-contrast injection signal intensity, $S(0)$ is the pre-contrast signal intensity, and T_E is echo time. Based on a well established tracer kinetic model in [3], the amount of contrast in the tissue is characterized by,

$$C_t(t) = \text{CBF} \cdot \int_0^t C_a(\tau) R(t - \tau) d\tau, \quad (11)$$

where $C_t(t)$ is the average CA concentration in a tissue voxel, CBF is the cerebral blood flow, $C_a(t)$ is the local CA concentration at the artery inlet, known as arterial input function (AIF), and $R(t)$ is the tissue residue function which measures the fraction of CA remaining in the given vascular network over time. The arterial input function $C_a(t)$ was determined over voxels in a small region extracted from branches of the middle cerebral artery. The noisy AIF signal was fit through a gamma-variate function to provide smooth concentration curves. Tissue residue functions $R(t)$ were obtained by deconvolving the tissue concentration time curves with the AIF using circulant truncated singular value decomposition [27]. The CBF was computed as the peak of the residue function, the CBV was determined as the area under the concentration time curves and the MTT was computed as the ratio of the CBV to CBF according to the central volume theorem [3].

B. Dynamic Contrast-Enhanced (DCE)

1) Data: The data from five different glioblastoma patients were used for evaluation of the methods. For each patient, two data sets were sequentially acquired on a 3T MRI scanner: one for constructing T_1 maps and one for the DCE-MRI analysis. Data for constructing T_1 maps were acquired using a 3D fast gradient echo multiple flip angle approach with parameters ($T_R = 6$ ms, $T_E = 2.32$ ms, flip angles of $\{2°, 5°, 10°, 15°, 20°, 30°\}$, voxel size = $2 \times 2 \times 2$ mm^3, acquisition matrix = 128×128, 20 slices). Dynamic DCE series were acquired with identical parameters but with a single flip angle of $10°$. A bolus of 0.1 mmol/kg of GD-DTPA (gadopentetic acid) was injected after 52 s. Initially,

---

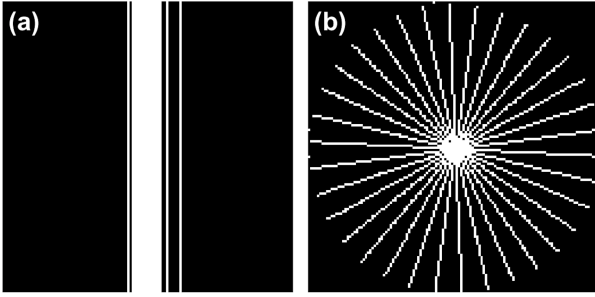[1]The details of algorithms solving each subproblem are given in supplementary material.

Fig. 2. Exemplary sampling patterns in (kx, ky) space corresponding to 8-fold acceleration: (a) Variable density Cartesian sampling (b) Radial Sampling.

Fig. 3. (a) RMSE versus iteration number for different (w_1, w_2) combinations obtained from proposed method, (b) PSNR versus iteration number for different reconstruction methods displaying the convergence of our algorithm depending on varying settings of step size (α) parameter.

a total of 250 frames were collected for up to 6 minutes. Each set of 5 frames were averaged before reconstruction, resulting 50 frames in total to constitute a dynamic sequence.

2) Analysis: Data collected at multiple flip angles for the T_1 map were fit using a non-linear least-squares fitting to the gradient echo signal intensity equation, given by [5]

$$S(t) = M \sin \alpha \frac{1 - \exp[-T_R/T_1(t)]}{1 - \cos \alpha \exp[-T_R/T_1(t)]}, \quad (12)$$

where M is the proton density, α is the flip angle and T_R is the repetition time. Here, we assume that $T_E \ll T_2^*$.

The increase in the relaxation rate can be further linearly related to the concentration of CA in the tissue, $C_t(t)$ [28],

$$1/T_1(t) = 1/T_1(0) + r_1 C_t(t), \quad (13)$$

where $T_1(0)$ is the tissue T_1 relaxation prior to the CA administration, $T_1(t)$ is the T_1 relaxation during and after injection, and r_1 is the CA relaxivity. We use the Patlak model [29] for tracer pharmacokinetic modeling and estimation of tissue perfusion parameters. This model describes a highly perfused two compartment tissue, ignoring backflux from the extracellular extravascular space (EES) into the blood plasma compartment. The CA concentration in the tissues is given by,

$$C_t(t) = v_p C_p(t) + K^{\text{trans}} \int_0^t C_p(\tau) d\tau, \quad (14)$$

where K^{trans} is the volume transfer rate at which CA is delivered to the EES, v_p is the plasma fraction, and $C_t(t)$ is the CA concentration in blood plasma. $C_p(t)$ denotes the AIF measured as in DSC analysis. The ROCKETSHIP toolbox [28] is used for T_1 map fitting and Patlak model implementation.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

All perfusion datasets described in Section III were acquired as fully sampled data. The min-max normalized fully-sampled dynamic sequences were retrospectively undersampled by multiplying its corresponding k-space data with a binary undersampling mask and subsequently adding complex additive white Gaussian (AWG) noise, as formulated in (1). The power of AWG noise was fixed to $\sigma^2 = 10^{-10}$ in all experiments. Undersampling was simulated with a time-varying variable density Cartesian sampling and Radial sampling (see Fig. 2). These sampling strategies were commonly used for dynamic MR applications [12], [9], [10]. To provide better evaluation of different methods, we considered increasing acceleration
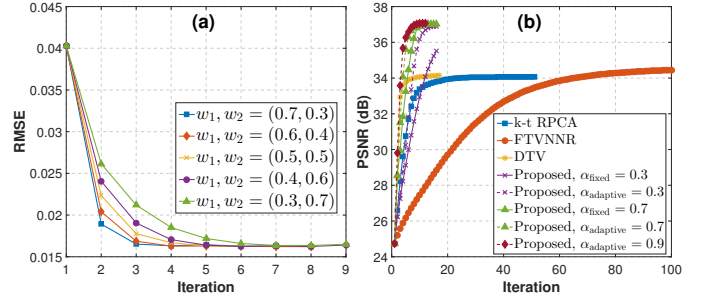
factors of R = $\{2x, 4x, 8x, 12x, 16x\}$ in the experiments. The reference image used in the $\mathcal{R}_L$-subproblem was initially taken as the direct inverse FFT (zero-filled) reconstruction of the first frame with 2-fold subsampling. Later on, it was updated as the average of all frames in the reconstructed sequence through iterations. Perfusion parameter estimation was employed after image reconstruction as a separate step.

We compare our method with three state-of-the-art dynamic reconstruction techniques: (k,t)-space via low-rank plus sparse prior (k-t RPCA) [12], dynamic total variation [9], fast total varitation and nuclear norm regularization (FTVNNR) [6]. To ensure fair comparison, as presented similarly in [10], we empirically fine-tuned the optimal regularization parameters for all three methods and individually for each dataset considering the suggested parameter space in the respective papers. The regularization parameters of our algorithm were set as $\lambda_1 = 0.001$ and $\lambda_2 = 0.25$. We also fixed $N_w = 7 \times 7 \times 7$ and $N_p = 5 \times 5 \times 5$ in all experiments. We considered using small cubic patches for N_w and N_p because larger patches drastically increase the computation time despite not improving the results substantially. The choice of proximal weights w_1, w_2 and step size α parameters of Algorithm 1 will be explained in Section IV-B1.

The quality of reconstructions was quantitatively measured with the Root-Mean-Square Error (RMSE) and Peak Signal-to-Noise Ratio (PSNR) metrics. RMSE for a 3D reconstructed sequence involving in total N voxels is calculated as $\text{RMSE}(X_r) = \sqrt{\|X_f - X_r\|_2^2/N}$ and PSNR is computed as $\text{PSNR}(X_r) = 20 \log \left[1/\text{RMSE}(X_r)\right]$, where X_f denotes the fully sampled sequence. To ensure convergence, all reconstruction methods were stopped when a maximum number of iterations (varying depending on the method) was reached, or when $\left(\|X_r^{k+1} - X_r^k\|_2^2\right)/\|X_r^k\|_2^2 \leq 10^{-6}$, where k is the iteration number. For the purposes of evaluation, we treated the fully sampled data as ground truth. As a commonly used metric in quantitative imaging [27], [7] we adopted the Lin's Concordance Correlation Coefficients (CCCs) to quantitatively assess the agreement of estimated perfusion parameters with reference values obtained from ground truth.

B. Results

1) Convergence of Proposed Algorithm and Effects of Parameters: In this experiment, we investigate the effect of proximal weights (w_1, w_2) and step size (α) parameters on
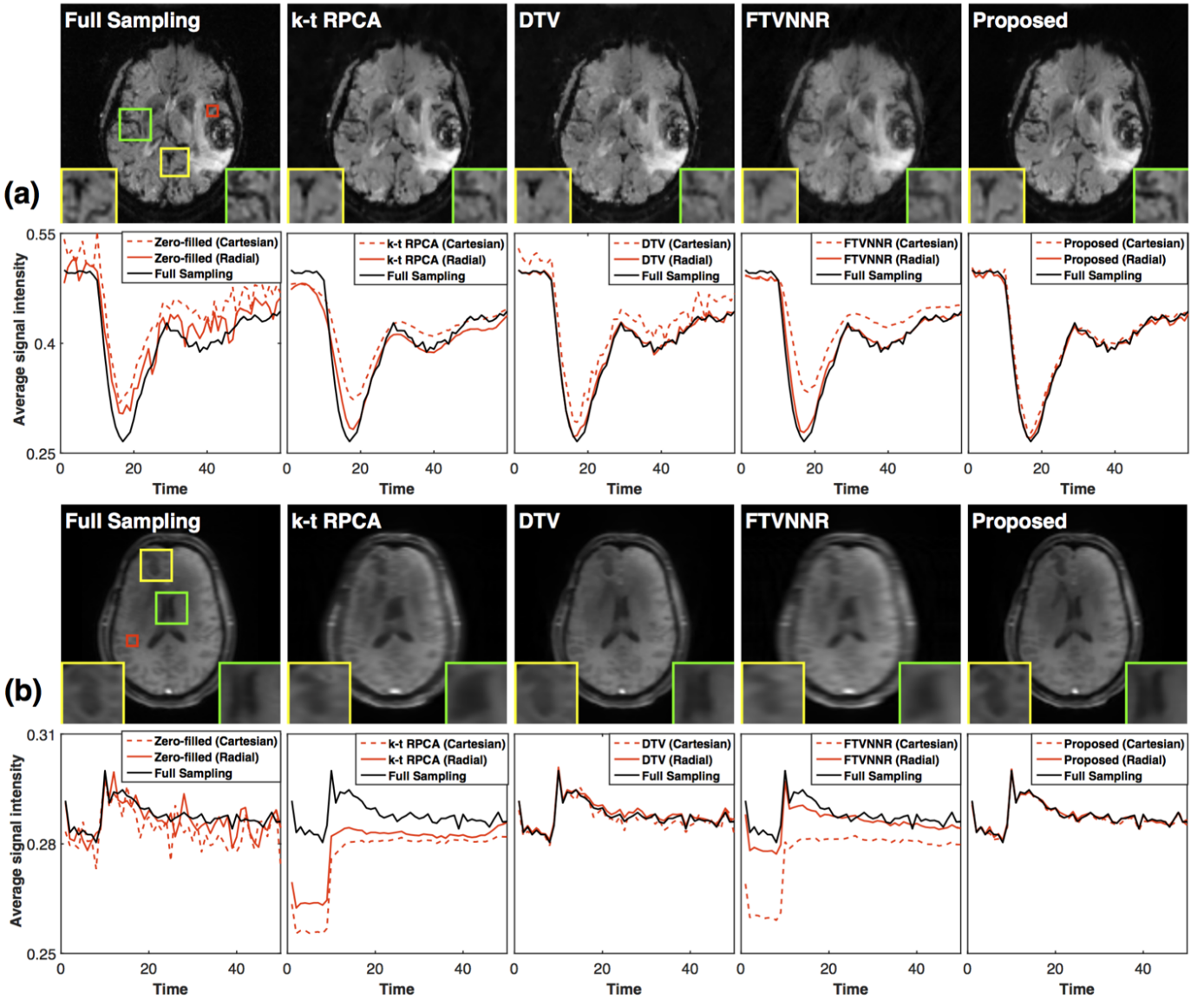
Fig. 4. Qualitative reconstruction results of a single frame and perfusion time intensity curves (TICs) obtained by each reconstruction algorithm. (a) DSC subject data. (b) DCE subject data. Single frame reconstructions are obtained using radial sampling with 12-fold acceleration for DSC data and 8-fold acceleration for DCE data. Close-up views of two regions of interest (yellow and green square) are also provided in every single frame reconstruction images. The TICs display the signal intensity over time averaged over the voxels inside the red square (corresponding to an arterial region) as shown in Full Sampling images, obtained from using Cartesian and Radial sampling schemes. Our proposed model results in high quality image frames as observed in close-up views. The reconstructed TICs also show strong alignment with the ones obtained from fully sampled data when undersampled by two different sampling schemes.

the convergence speed of our proposed algorithm. Fig. 3 shows the resulting RMSE and PSNR values versus iterations depending on various combinations of (w_1, w_2) pairs and varying settings of α. Fig. 3(a) reveals that the convergence rate of our proposed algorithm increases when the weight of the first proximal has a higher value. The highest convergence rate is achieved by the weight pair of $(0.7, 0.3)$. This implies that $\mathcal{R}_L$-proximal yields better intermediate reconstruction accuracy than the one obtained by $\mathcal{R}_{NL}$-proximal. Therefore, a higher weight given to $\mathcal{R}_L$-proximal will yield lower RMSE in a less number of iterations, indicating higher convergence rate. Fig. 3(a) also shows that all combinations of weights ultimately reach similar RMSE values, meaning that the final reconstructions are not affected by different proximal weight-

ing. This demonstrates the stability and robustness of the GFBS based splitting algorithm against proximal weighting.

The impact of various step size (α) parameter settings on reconstruction performance is displayed in Fig. 3(b). In this experiment, we set $w_1, w_2 = (0.5, 0.5)$ for all cases of α. α_{fixed} denotes a fixed α value while α_{adaptive} denotes adaptive α updated through iterations as given in Algorithm 1. The results in Fig. 3(b) reveal that our proposed adaptive setting of α increases the convergence speed of the algorithm. This is clearly visible when $\alpha_{\text{fixed}} = \alpha_{\text{adaptive}} = 0.3$. After 6th iteration, our proposed method reaches the highest PSNR value (with $\alpha_{\text{adaptive}} = 0.3$) when compared to other methods. Fig. 3(b) also demonstrates that the acceleration gain of convergence decreases when initial α_0 has a higher value, for instance
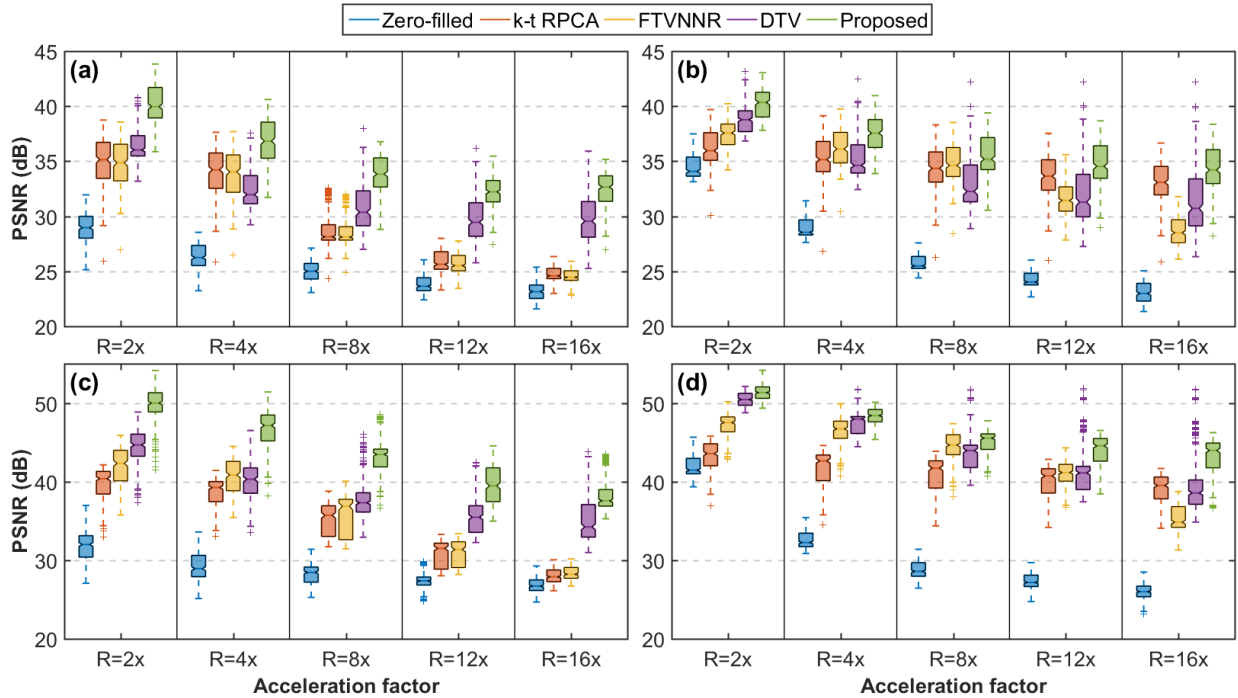
Fig. 5. Boxplots displaying the PSNR values of reconstruction methods with respect to increasing acceleration factors R. Reconstruction results obtained from all 5 DSC sequences with (a) Cartesian, and (b) Radial sampling; from all 5 DCE sequences with (c) Cartesian, and (d) Radial sampling. Each DSC sequence involves 60 frames and each DCE sequence involves 50 frames. PSNR values were calculated for every single frame individually. The proposed method performs the best average PSNR in all acceleration rates for both with DSC and DCE datasets.

$\alpha_{\text{fixed}} = \alpha_{\text{adaptive}} = 0.7$. The PSNR results obtained with $\alpha_{\text{adaptive}} = 0.9$ suggests that increasing step-size parameter of GFBS algorithm enables faster convergence. To this end, we set the initial $\alpha_0 = 0.9$ and use the adaptive setting scheme for α in the remaining experiments of the paper. This allows us to reach the highest reconstruction accuracy in the least number of iterations, thereby significantly reducing processing time.

2) Reconstruction Performance: This section presents the reconstruction results of all competing methods using variable density Cartesian and radial sampling schemes. Fig. 4 demonstrates a single reconstructed frame of one of the DSC and DCE brain perfusion datasets and estimated perfusion TICs averaged over voxels inside an arterial region. The acceleration factors here are 12 and 8 for DSC and DCE data, respectively. The results in Fig. 4 show that our proposed reconstruction algorithm can achieve the best spatial reconstruction and highly accurate estimation of TICs when compared to the other three methods. Considering the spatial results, when looking at details in close-up views of Fig. 4(a), DSC frame reconstructions obtained by k-t RPCA and our proposed method produce the best results compared to DTV and FTVNNR. DTV reconstructions are mostly lacking finer details whereas FTVNNR yields more blurry spatial regions. The TIC reconstructions in Fig. 4(a) indicate that radial sampling yields more accurate matching of full sampling TICs compared to variable density Cartesian sampling. Among all TIC results, DTV and our proposed method reconstruct perfusion signal patterns that are in good agreement with the pattern of the fully sampled data (see Fig. 4(a) bottom third and fifth column) when radial sampling is used for undersampling. DTV produces signifi-

cantly worse TICs with underestimated perfusion peaks while our method still yields very accurate matching of TICs with Cartesian sampling. As we also demonstrated in our previous work [17], k-t RPCA and FTVNNR estimate oversmooth TICs and reconstruct underestimated perfusion peaks with Cartesian sampling (see Fig. 4(a) bottom second and fourth column). However, radial sampling helps these two methods to improve their accuracy in TIC estimation, especially FTVNNR yields highly accurate matching of TICs with fully sampled data. These results evidence that radial sampling should be preferred over Cartesian sampling for the quantitative perfusion MRI in which the fidelity of TICs plays a vital role.

Considering the spatial results of DCE data in Fig. 4(b), DTV and our proposed method provide the best reconstructions. When looking at details in close-up views, it is visible that our method reconstructs finer details compared to DTV. FTVNNR again shows more blurry regions and thus lacking details. Unlike the results in DSC data, k-t RPCA yields spatial reconstructions with missing details and blurred edges in DCE data. In terms of TIC reconstruction, Fig. 4-(b) demonstrates that our method can achieve the most accurate TICs both with Cartesian and radial sampling, where DTV achieves the second best. However, when the observed signal dynamics are lower compared to DSC, both k-t RPCA and FTVNNR fail in the estimation of perfusion TICs in DCE. The peaks of the perfusion signal are underestimated and there is a large offset between the estimation and real value. This result reveals that k-t RPCA and FTVNNR are not robust against small temporal variations since these methods do not explicitly exploit sparsity in the temporal domain while our method exploits the temporal
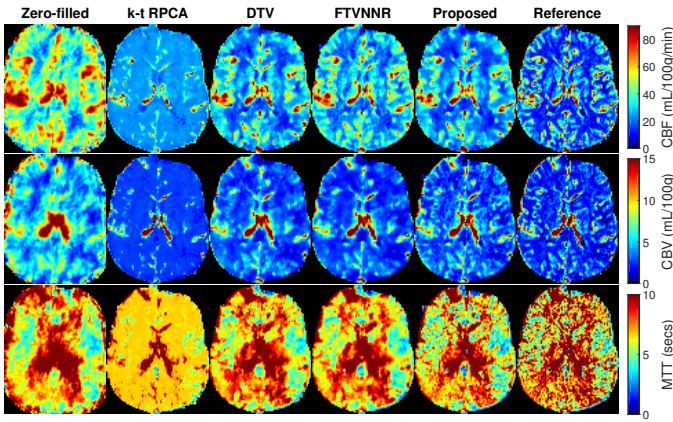
Fig. 6. Hemodynamic parameter maps (CBF, CBV, MTT) of a DSC subject for different methods with an 8-fold acceleration. This subject has a low-grade glioma which cannot be easily recognized.

variations in both voxel and patch-wise levels.

Fig. 5 presents the quantitative results (in terms of PSNR) of all reconstruction methods depending on increasing acceleration factors for both DSC and DCE sequences. As expected, PSNR values decrease with increasing acceleration due to the increase in missing k-space samples. Moreover, all reconstruction methods benefit from radial sampling because it can be seen that zero-filled reconstruction already gives an improved reconstruction compared to the one obtained by Cartesian sampling. The reason is that the resulting artifacts produced by radial sampling resemble noise compared to Cartesian undersampling [12]. Fig. 5 clearly shows that our proposed method performs the best PSNR in all acceleration rates for both with DSC and DCE datasets. It is also observed that the performance of k-t RPCA and FTVNNR is improved with radial undersampling while on average DTV performs worse with radial sampling especially in higher accelerations.

3) Perfusion Parameter Estimation: For the estimation of perfusion parameters, we only used the reconstruction results obtained by radial sampling due to its efficiency in both spatial and temporal reconstruction, as demonstrated in the previous section. Figs. 6-7 display qualitative results of estimated hemodynamic parameter maps from two DSC subjects. From Fig. 6,
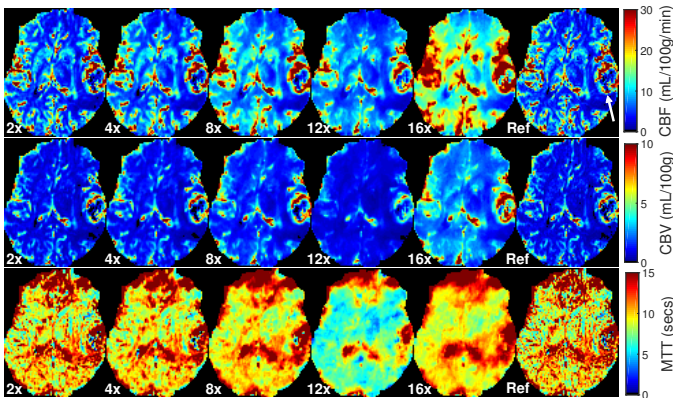

Fig. 7. Hemodynamic parameter maps (CBF, CBV, MTT) of a DSC subject resulting from our proposed method with respect to different acceleration factors and Reference (Ref) maps for comparison. White arrow in Reference CBF map indicates the tumor region.
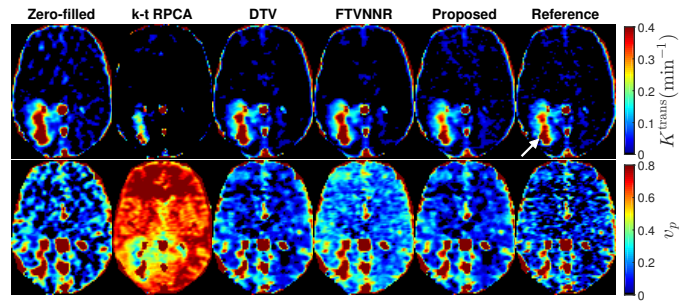

Fig. 8. Pharmacokinetic parameter maps (K^{trans}, v_p) of a DCE subject for different methods with a 12-fold acceleration. The tumor area is marked by a white arrow in the Reference K^{trans} map.

it can be seen that the proposed method results in perfusion maps where most of the tissue structures are preserved and appear sharper compared to FTVNNR and DTV with 8-fold acceleration. Compared to reference maps, CBF and CBV values are overestimated in some regions of white matter (WM) and gray matter (GM) while perfusion values in blood vessels (appear as small red areas) stay mostly accurate. The k-t RPCA method produces highly inaccurate (oversmooth) perfusion maps as expected by the mismatch of TICs shown in Fig. 4(a). Fig. 7 demonstrates that estimated hemodynamic parameter maps generated by our proposed method appear highly accurate up to 8-fold acceleration, however the maps start to deteriorate and involve oversmooth regions at higher acceleration rates. CBF assessment in tumor areas (at tumor boundary and core) also appear quite consistent with reference maps up to 12-fold acceleration.

Fig. 8 shows estimated pharmacokinetic parameter maps of a DCE subject with a 12-fold acceleration. The estimated maps indicate that significantly higher permeability (K^{trans}) can be observed in tumor tissues (white arrow). The proposed method produces parameter maps which show a strong match with reference maps at the tumor region and exhibit relatively oversmooth regions in WM and GM due to a high rate of acceleration (12-fold). The k-t RPCA especially produces highly overestimated plasma fraction v_p values and again reveals the fact that it is inadequate for PWI reconstruction. Due to the page limitation, we provide the qualitative results of an another DCE subject in supplementary material.

Fig. 9 presents Bland-Altman plots of the CBF and K^{trans} values obtained by our method with respect to increasing acceleration factors. The plots indicate that the fidelity of estimation is very high at lower acceleration rates where we observe very small differences between the estimation and ground truth. With increasing acceleration, the bias and variance of the differences generally become larger and subsequently CCC values diminish. This quantitative assessment coincides with the qualitative results shown in Fig. 7.

Finally, Table I reports the average CCCs of DSC and DCE perfusion parameter maps obtained from all methods with varying acceleration rates. Our reconstruction method yields the best CCCs for all parameters and accelerations. The k-t RPCA generally performs even worse than zero-filled reconstruction since it leads to oversmoothing of the temporal perfusion signal. The DTV and FTVNNR show relatively similar performance with our method at lower accelerations
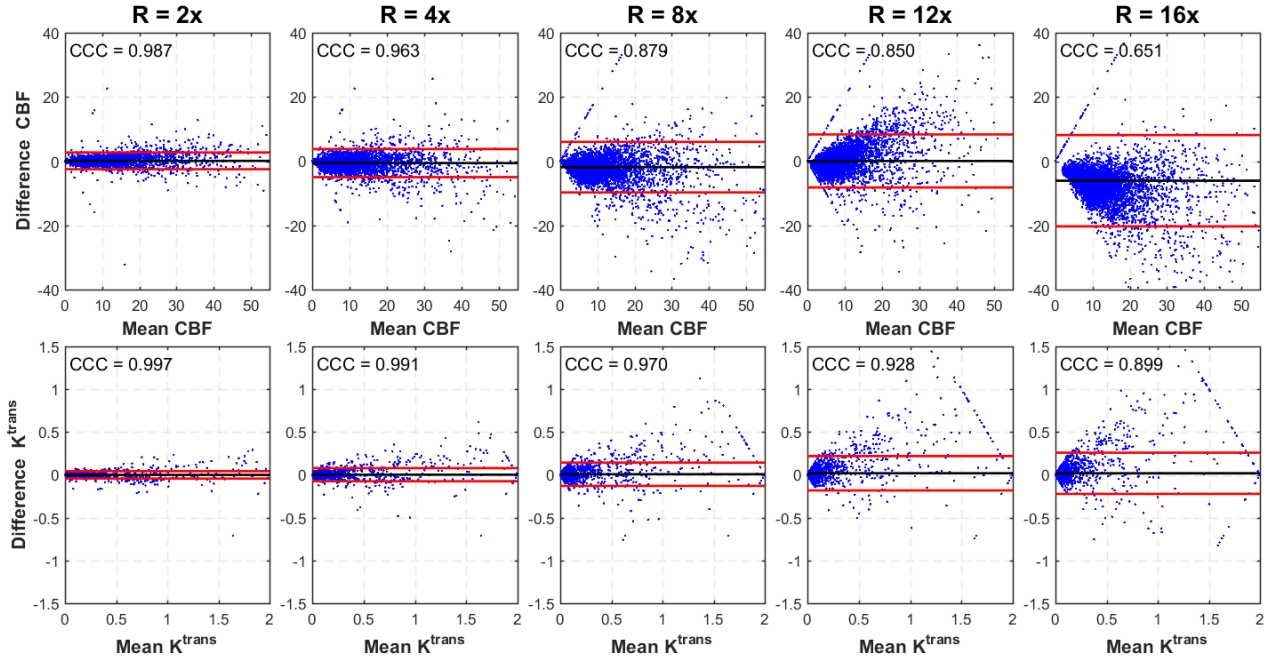
Fig. 9. Bland-Altman plots and 95% confidence intervals within two red lines for CBF (top) and K^{trans} (bottom) perfusion parameters of one DSC and DCE subject data depending on different acceleration factors, resulting from our proposed reconstruction method. Corrresponding CCCs are also provided at the top-left corner of each plot. For these subjects, CBF values can range from 0 to 71 mL/100g/min and K^{trans} values can range from 0 to 2 min^{-1}. As can be clearly seen in the plots, CCCs decrease with increasing acceleration, which coincides with the changes of bias and variance.

TABLE I
AVERAGE CCC VALUES OF THE DSC PERFUSION MAPS (CBF, CBV, MTT) AND DCE PERFUSION MAPS (K^{trans}, v_p) ESTIMATION OBTAINED FROM ALL THE SUBJECTS. CCCS TAKE VALUES IN $[0, 1]$ INTERVAL. THE BEST PERFORMANCE IS HIGHLIGHTED IN BOLD FONT.

		Acceleration Rate				
	Method	R=2x	R=4x	R=8x	R=12x	R=16x
DSC Parameters	Zero-filled	0.924, 0.888, 0.854	0.718, 0.684, 0.630	0.485, 0.479, 0.432	0.345, 0.362, 0.320	0.221, 0.205, 0.155
	k-t RPCA	0.839, 0.786, 0.767	0.889, 0.897, 0.796	0.812, 0.823, 0.728	0.733, 0.719, 0.654	0.634, 0.625, 0.644
	DTV	0.962, 0.957, 0.939	0.930, 0.930, 0.889	0.807, 0.804, 0.813	0.713, 0.691, 0.751	0.614, 0.628, 0.734
	FTVNNR	0.966, 0.941, 0.870	0.946, 0.937, 0.841	0.804, 0.784, 0.733	0.695, 0.689, 0.614	0.603, 0.600, 0.429
	Proposed	**0.968, 0.962, 0.947**	**0.958, 0.948, 0.897**	**0.887, 0.862, 0.821**	**0.741, 0.751, 0.759**	**0.637, 0.663, 0.739**
DCE Parameters	Zero-filled	0.987, 0.973	0.927, 0.857	0.877, 0.727	0.763, 0.677	0.741, 0.663
	k-t RPCA	0.790, 0.681	0.634, 0.370	0.596, 0.352	0.531, 0.312	0.498, 0.275
	DTV	0.990, 0.985	0.943, 0.941	0.847, 0.813	0.789, 0.736	0.764, 0.670
	FTVNNR	0.984, 0.903	0.952, 0.839	0.879, 0.694	0.848, 0.589	0.791, 0.546
	Proposed	**0.994, 0.991**	**0.973, 0.967**	**0.919, 0.845**	**0.889, 0.768**	**0.848, 0.686**

(up to 4-fold) and worse performance at higher accelerations.

4) Computation Time: The most computationally expensive step of our algorithm is solving each proximal map. We process the computation of proximal maps in parallel since proximal-splitting can allow it due to the independence between the inputs of proximity operators. Other steps of the algorithm involve adding and multiplying vectors or scalars, and are thus very cheap in terms of computational complexity.

All methods were ran using Matlab R2015b on a desktop computer with Intel Xeon CPU E3-1226 v3 Processor at 3.3 GHz and 32 GiB of memory. Table II provides the processing time (in seconds) of all methods on two types of datasets. Among all methods, our method requires the highest processing time. However, within a similar computation time (corresponding to $\simeq$ 3-4 iterations) as competing methods, our method can usually reach the best reconstruction accuracy.

Considering the longer processing times ($\simeq$ 15 minutes for DCE data analysis) for voxel-wise fitting of perfusion parameters, we believe that the slightly longer reconstruction time of our method can be negligible.

V. DISCUSSION AND CONCLUSION

In this paper, a new reconstruction model exploiting spatio-temporal variations jointly at multiple levels was proposed for the acceleration of PWI acquisitions. The proposed method was compared with existing state-of-the-art reconstruction methods and evaluated on clinical DSC and DCE-MRI patient datasets. Extensive experiments validated the effectiveness of our method in terms of improved spatial reconstructions, highly accurate matching of perfusion temporal signals, and more precise estimation of clinically relevant perfusion parameters. Experiments based on retrospective undersampling revealed that our reconstruction model can potentially enable up

TABLE II
COMPUTATION TIME OF DIFFERENT RECONSTRUCTION METHODS.

Time (secs)	k-t RPCA	DTV	FTVNNR	Proposed
DSC dataset	187.2	72.3	142.4	324.1
DCE dataset	165.6	61.8	106.3	268.7

to 8-fold acceleration on clinically feasible perfusion datasets. We also demonstrated that our method is very robust against incoherent artifacts caused by varying sampling patterns. Our recent work [30] showed that the proposed model can also achieve similar performances with Poisson-disc sampling.

The maximum acceleration achieved with our method can be further increased with the use of a high-spatial resolution data as applied in [7], [15] for DCE study. However, we remark that high spatial resolution is not so clinically realistic for PWI because high temporal resolution is vital to capture entire contrast dynamics for precise blood flow quantification.

As mentioned already, this paper considers the reconstruction on $2D + t$ data, i.e., on a single slice followed over time. However, our approach can be easily extended to $3D + t$ data. One should take into account time complexity because especially applying NLM filter on $4D$ patches can significantly increase the computation time due to exhaustive search of similar patches in larger windows. Hence, an optimized GPU implementation is necessary to perform it efficiently in $4D$.

We would like to emphasize that our regularization approach can also be adopted to many inverse problems in medical imaging with only a few modifications. Possible applications beyond perfusion imaging might be MR Spectroscopy [31], low-dose CT denoising [32], and MR super-resolution [33]. Moreover, our algorithm can provide an efficient way of solving such regularized inverse problems.

One of the limitation of our method is that it usually produces oversmooth reconstructions and therefore overestimated perfusion parameters when the acceleration factor is relatively higher ($\geq$12-fold). This observation is mostly valid for parenchyma voxels (i.e., WM and GM) where the signal drop or enhancement is very low compared to blood vessels. Iterative reconstruction algorithms usually tend to produce repeating structures and smooth the available information throughout neighboring image regions when there is a large amount of missing data due to the high undersampling. In order to obtain more accurate parameter estimates in highly undersampled data, deep neural networks [34] can be exploited to learn deeper spatio-temporal representations and similarities within the MR image series than conventional NLM can learn with a simple k-nearest neighbors approach. One idea might be to learn the noise-free temporal signal for every voxel within the spatially correlated image regions given the corresponding noisy signal using deep learning methods. Future lines of research will attempt to explore this idea.

REFERENCES

[1] T. E. Conturo, E. Akbudak, M. S. Kotys, M. L. Chen, S. J. Chun, R. M. Hsu, C. C. Sweeney, and J. Markham, "Arterial input functions for dynamic susceptibility contrast MRI: Requirements and signal options," *J. Magn. Reson. Imaging*, vol. 22, pp. 697–703, 2005.

[2] T. E. Yankeelov and J. C. Gore, "Dynamic contrast enhanced magnetic resonance imaging in oncology: Theory, data acquisition, analysis, and examples," *Curr. Med. Imaging Rev.*, vol. 3, no. 2, pp. 91–107, 2007.

[3] L. Østergaard, "Principles of cerebral perfusion imaging by bolus tracking," *J. Magn. Reson. Imaging*, vol. 22, no. 6, pp. 710–717, 2005.

[4] M. S. Shiroishi, G. Castellazzi, J. L. Boxerman, F. D'Amore, M. Essig, T. B. Nguyen, J. M. Provenzale, D. S. Enterline, N. Anzalone, A. Drfler *et al.*, "Principles of T2*-weighted dynamic susceptibility contrast MRI technique in brain tumor imaging," *J. Magn. Reson. Imaging*, vol. 41, no. 2, pp. 296–313, 2015.

[5] D. S. Smith, X. Li, J. V. Gambrell, L. R. Arlinghaus, C. C. Quarles, T. E. Yankeelov, and E. B. Welch, "Robustness of quantitative compressive sensing MRI: The effect of random undersampling patterns on derived parameters for DCE- and DSC-MRI," *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 504–511, 2012.

[6] J. Yao, Z. Xu, X. Huang, and J. Huang, "Accelerated dynamic MRI reconstruction with total variation and nuclear norm regularization," in *Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 635–642.

[7] R. M. Lebel, J. Jones, J.-C. Ferre, M. Law, and K. S. Nayak, "Highly accelerated dynamic contrast enhanced imaging," *Magn. Reson. Med.*, vol. 71, no. 2, pp. 635–644, 2014.

[8] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1042–1054, 2011.

[9] C. Chen, Y. Li, L. Axel, and J. Huang, "Real time dynamic MRI with dynamic total variation," in *Med. Image Comput. Comput. Assist. Interv.*, 2014, pp. 138–145.

[10] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity for dynamic MR data reconstruction," *IEEE Trans. Med. Imaging*, vol. 33, no. 4, pp. 979–994, 2014.

[11] C. Zhang, Z. Jin, H. Ye, and F. Liu, "Accelerating dynamic cardiac imaging based on a dual-dictionary learning algorithm," in *Proc. Int. Conf. Biomed. Eng. Inform.*, 2015, pp. 57–61.

[12] B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge, "Dynamic MR image reconstruction-separation from undersampled (k-t)-space via low-rank plus sparse prior," *IEEE Trans. Med. Imaging*, vol. 33, no. 8, pp. 1689–1701, 2014.

[13] D. Boschetto, P. D. Prima, M. Castellaro, A. Bertoldo, and E. Grisan, "Baseline constrained reconstruction of DSC-MRI tracer kinetics from sparse fourier data," in *Proc. IEEE Int. Symp. Biomed. Imaging*, April 2014, pp. 321–324.

[14] G. Adluru, S. P. Awate, T. Tasdizen, R. T. Whitaker, and E. V. DiBella, "Temporally constrained reconstruction of dynamic cardiac perfusion MRI," *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1027–1036, 2007.

[15] Y. Guo *et al.*, "High-resolution whole-brain DCE-MRI using constrained reconstruction: Prospective clinical evaluation in brain tumor patients," *Med. Phys.*, vol. 43, no. 5, pp. 2013–2023, 2016.

[16] H. Raguet, J. Fadili, and G. Peyré, "A generalized forward-backward splitting," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1199–1226, 2013.

[17] C. Ulas, P. A. Gómez, F. Krahmer, J. I. Sperl, M. I. Menzel, and B. H. Menze, "Robust reconstruction of accelerated perfusion MRI using local and nonlocal constraints," in *Reconstruction, Segmentation, and Analysis of Medical Images: First Int. Workshops, RAMBO 2016 and HVSMR 2016, in Conjunction with MICCAI*, 2017, pp. 37–47.

[18] Y. Q. Mohsin, S. G. Lingala, E. DiBella, and M. Jacob, "Accelerated dynamic MRI using patch regularization for implicit motion compensation," *Magn. Reson. Med.*, vol. 77, no. 3, pp. 1238–1248, 2017.

[19] J. Huang, S. Zhang, and D. Metaxas, "Efficient MR image reconstruction for compressed MR imaging." *Med. Image Anal.*, vol. 15, no. 5, pp. 670–679, 2011.

[20] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*. Springer New York, 2011, pp. 185–212.

[21] C. Chen, J. Huang, L. He, and H. Li, "Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2713–2720.

[22] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: A unified variational framework," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3192–3203, 2013.

[23] R. J. Marks, *Alternating projections onto convex sets*. In Deconvolution of images and spectra, 2nd ed. San Diego: Academic Press, 1996.

[24] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2005, pp. 60–65.

[25] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Trans. Med. Imaging*, vol. 27, no. 4, pp. 425–441, 2008.

[26] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[27] R. Fang, S. Zhang, T. Chen, and P. Sanelli, "Robust low-dose CT perfusion deconvolution via tensor total-variation regularization," *IEEE Trans. Med. Imaging*, vol. 34, no. 7, pp. 1533–1548, 2015.

[28] S. Barnes, T. Ng, N. Santa-Maria, A. Montagne, B. V. Zlokovic, and R. E. Jacobs, "ROCKETSHIP: a flexible and modular software tool for the planning, processing and analysis of dynamic MRI studies," *BMC Med. Imaging*, vol. 15, p. 19, 2015.

[29] C. S. Patlak, R. G. Blasberg, and J. D. Fenstermacher, "Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data," *J. Cereb. Blood Flow Metab.*, vol. 3, no. 1, pp. 1–7, 1983.

[30] C. Ulas, P. A. Gómez, J. I. Sperl, C. Preibisch, M. I. Menzel, A. Haase, and B. H. Menze, "A Robust Reconstruction Method for Quantitative Perfusion MRI: Application to Brain Dynamic Susceptibility Contrast (DSC) Imaging," in *Proc. Intl. Soc. Magn. Reson. Med.*, 2017.

[31] D. Das, E. Coello, R. F. Schulte, and B. H. Menze, "Spatially adaptive spectral denoising for MR spectroscopic imaging using frequency-phase non-local means," in *Med. Image Comput. Comput. Assist. Interv.*, 2016, pp. 596–604.

[32] Y. Tao, G.-H. Chen, T. A. Hacker, A. N. Raval, M. S. Van Lysel, and M. A. Speidel, "Low dose dynamic CT myocardial perfusion imaging using a statistical iterative reconstruction method," *Med. Phys.*, vol. 41, no. 7, 2014.

[33] F. Shi, J. Cheng, L. Wang, P. T. Yap, and D. Shen, "LRTV: MR image super-resolution with low-rank and total variation regularizations," *IEEE Trans. Med. Imaging*, vol. 34, no. 12, pp. 2459–2466, 2015.

[34] G. Wang, "A perspective on deep imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016.

VI. Supplementary Material

In this supplementary, we provide a few extra materials that have been referenced in the original manuscript.
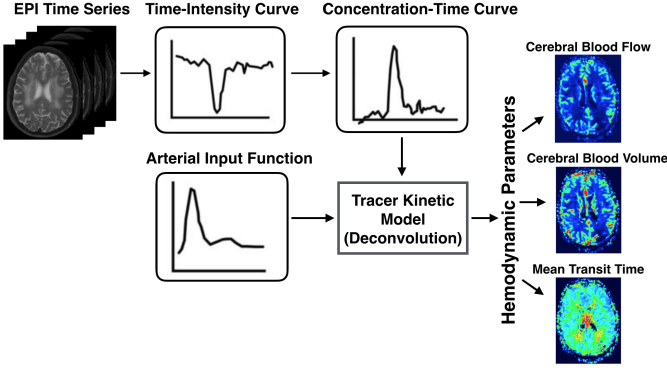
A. A Diagram on parameter estimation in DSC-MRI

Fig. 10. A diagram illustrating the steps of hemodynamic parameter estimation in dynamic susceptibility contrast (DSC) T_2^*-weighted perfusion MRI. First, time-intensity curves (TICs) for each voxel are converted to tracer tissue concentration-time courses based on a linear relationship between the signal drop in T_2^* and tissue concentration. Then, a tracer kinetic model based on the deconvolution of arterial input function and tissue concentrations is applied to determine perfusion parameters [4]. As the voxel-wise TICs are directly correlated to the amount of contrast medium in tissue, precise estimation of TICs is very crucial to accurately quantify the hemodynamic parameters. Especially under and over estimation of the peak of the TICs entirely reflect the resulting parameters obtained by a deconvolution process.

B. Algorithm for solving $\mathcal{R}_L$-subproblem

As stated in the manuscript, this subproblem can be solved efficiently using an accelerated iteratively reweighted least squares algorithm proposed for structured sparsity reconstruction [21]. The total-variation (TV) term in our problem (7) can be modeled as the group sparsity,

$$\|x\|_{TV} = \|x\|_{2,1} = \sum \|x_{g_i}\|_2, \;\; i = 1, 2, ..., m \quad (15)$$

where x_{g_i} denotes the component in the i-th group and m is the total number of groups. For instance, TV component in each dimension (e.g. $\nabla_x x$) constitutes a group. Following the notations given in (II-B), the problem in (7) is solved by iteratively updating the weight matrix W and the solution d_t for each frame individually [9]. W is a diagonal matrix with the i-th diagonal entry given as

$$W_i^k = 1/\sqrt{\left(\nabla_x d_{t_i}^k\right)^2 + \left(\nabla_y d_{t_i}^k\right)^2}, \;\; i = 1, 2, ..., N_x N_y \quad (16)$$

where k is the iteration number. d_t^{k+1} is updated by solving the following linear equation:

$$\left(\mathcal{F}_t^H \mathcal{F}_t + \lambda Q_1^H W^k Q_1 + \lambda Q_2^H W^k Q_2\right) d_t = \mathcal{F}_t^H b_t, \quad (17)$$

where $b_t = y_t - \mathcal{F}_t \bar{x}$. The close form solution of (17) is derived as $d_t^k = S^{-1} \mathcal{F}_t^H b_t$, where $S = \mathcal{F}_t^H \mathcal{F}_t + \lambda Q_1^H W^k Q_1 + \lambda Q_2^H W^k Q_2$ is the system matrix. However, the direct inversion of S is not computationally feasible. In [21], it is proposed to use a preconditioner P which is close to S and the inverse of S can be computed in a more computationally efficient way. The overall problem in (17) can be solved by a preconditioned conjugate gradient (PCG) method.

Algorithm 2: $\mathcal{R}_L$-subproblem

Input: $\mathcal{F}_u$, $\bar{x}$, Y, Initial estimate X^0, $\lambda = 2\lambda_1$
for $t \in \mathbb{T} = \{1, 2, ..., T\}$ **do**
    Initialize: $d_t^0 = x_t^0 - \bar{x}$, $b_t = y_t - \mathcal{F}_t \bar{x}$, $k = 0$
    while *stopping criteria not met* **do**
        Obtain W^k via (16) ;
        $S = \mathcal{F}_t^H \mathcal{F}_t + \lambda Q_1^H W^k Q_1 + \lambda Q_2^H W^k Q_2$;
        $P = sI + \lambda Q_1^H W^k Q_1 + \lambda Q_2^H W^k Q_2 = LU, P^{-1} = U^{-1} L^{-1}$;
        while *PCG stopping criteria not met* **do**
            Update d_t^{k+1} by PCG for $Sd_t = \mathcal{F}_u^H b_t$ with $P \approx LU$;
        end
        $k \leftarrow k + 1$;
    end
    $\hat{x}_t = d_t + \bar{x}$;
end
Output: $\hat{X}_{\mathcal{R}_L} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_t]$

The preconditioner P in our problem can be designed based on the following observation [9]: The symmetric matrix $A_t^H A_t$ is diagonal and therefore $\mathcal{F}_t^H \mathcal{F}_t = F^H A_t^H A_t F$ is diagonally dominant. Due to the properties of the Fourier transform, all the diagonal elements of $\mathcal{F}_t^H \mathcal{F}_t$ is equal to the mean of diagonal elements of $A_t^H A_t$, which is the undersampling factor denoted as s. A good approximation for $\mathcal{F}_t^H \mathcal{F}_t$ can be made with sI, where I is the identity matrix. The new preconditioner is finally defined as $P = sI + \lambda Q_1^H W^k Q_1 + \lambda Q_2^H W^k Q_2$. The new procoditioner P is a symmetric penta-diagonal matrix which does not have a closed form inverse. However, P is usually diaogonally dominant because the regularization parameter λ is often very small in our specific problem. Therefore, an incomplete LU decomposition can be applied to such matrix with $P \approx LU$, where L and U are a lower triangle matrix and an upper triangle matrix, respectively.

Provided all the details, the steps of algorithm solving $\mathcal{R}_L$-subproblem is outlined in Algorithm 2.

C. Algorithm for solving $\mathcal{R}_{NL}$-subproblem

The subproblem is solved via a simple two-step alternating minimization scheme. In every main iteration, the reconstructed image data estimate is first projected onto the data fidelity term and then NLM filtering is applied to the projected data in entire spatio-temporal (3D) space. The steps of the algorithm solving $\mathcal{R}_{NL}$-subproblem is provided below in Algorithm 3.

In this algorithm, NLM operation performs the NLM filtering on the projected data given the filter parameters. Basically, in each spatio-neighborhood window $\mathcal{N}_\mathbf{p}$, first the

---

Algorithm 3: $\mathcal{R}_{NL}$-subproblem

Input: $\mathcal{F}_u$, Y, Initial estimate X^0, $\alpha = 2\lambda_2$, N_w, N_p, σ
Initialize: $X_{est}^0 = X^0$, $h = 0.2\sigma$, $k = 0$
while *stopping criteria not met* **do**
    $X_{proj} = X_{est}^k + \mathcal{F}_u^H \left(Y - \mathcal{F}_u X_{est}^k\right)$;
    $X_{nlm} = NLM(X_{proj}, N_w, N_p, h)$;
    $X_{est}^{k+1} = X_{est}^k + \alpha(X_{nlm} - X_{est}^k)$;
    $k \leftarrow k + 1$;
end
Output: $\hat{X}_{\mathcal{R}_{NL}} = X_{est}$

---

weights between the centers of the patches are calculated via (3) and then these weights are directly used to update each voxel value via (9) taking into account the other voxels in $\mathcal{N}_\mathbf{p}$ surrounding the voxel $\mathbf{p}$. As the inter-pixel weights are re-estimated in every iteration, the estimated weights become more reliable when the quality of the reconstructed data X_{est} are improved through iterations.

We note that the classical NLM filter [24] was normally defined with a Gaussian-weighted Euclidean distance, $\|\cdot\|_{2,a}^2$, where a is the standard deviation of a Gaussian kernel. This kernel basically gives decaying weights to voxel differences away from the center of the patches. However, in this work, we used the classical Euclidean distance $\|\cdot\|_2^2$ to simplify the complexity of the problem and reduce the computational time as proposed in [25]. We anticipate that a Gaussian-weighted distance measure may slightly improve the quality of our reconstructions because assigning uniform weights to all the voxels inside a patch may lead to stronger smoothing in image regions involving especially large texture and more fine details.

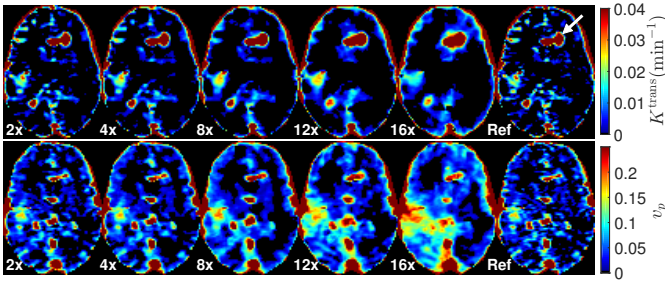## D. Additional Qualitative Parameter Maps of a DCE Subject

Fig. 11. Pharmacokinetic parameter maps (K^{trans}, v_p) of a DCE subject obtained by our proposed method with respect to increasing acceleration factors and Reference (Ref) maps for comparison. White arrow in Reference K^{trans} map marks the tumor region.

Fig. 11 shows the estimated pharmacokinetic parameter maps of our proposed reconstruction model with respect to increasing acceleration factors. The estimated maps look very similar up to 4-fold acceleration when compared to reference maps. However, starting from 8-fold acceleration, our method yields oversmooth image regions in different brain areas. Especially in higher acceleration factors ($\geq$12-fold), K^{trans} values are largely underestimated in WM and GM, and the maps contain expanded tumor regions (oversmoothed). On the other hand, v_p values are mostly overestimated in WM and GM due to oversmoothing, and we clearly observe expanded regions especially around a few blood vessels which are highly perfused and affect the neighboring brain areas as well.

Additional Experimental Results of Previous Works

This chapter mainly involves two sections:

1. Section B.1 contains a few experimental results which were obtained by running our proposed image reconstruction approach [85, 86] on high-resolution brain and cardiac DCE-MRI datasets.

2. Section B.2 provides several interesting plots of a published ISMRM abstract [138] which presents a random forest regression based ML method for direct estimation of pharmacokinetic parameters in DCE-MRI.

B.1   Image Reconstruction on High-resolution DCE-MRI

This part mainly presents the reconstruction results achieved by our reconstruction model [85, 86] using DCE-MRI brain and cardiac (rest and stress) perfusion datasets provided by the Brain Research Imaging Centre and Clinical Research Imaging Centre at The University of Edinburgh.

Brain dataset: Three patients with first clinically evident mild ischaemic stroke were imaged on a 1.5 T MRI scanner using an 8-channel phased-array head coil. DCE-MRI was performed at approximately 1 month after first presentation of stroke and consisted of a 3D T1W spoiled gradient echo sequence with TR/TE = 8.24/3.1 ms, 24×24 cm FOV, 256×192 acquisition matrix and 42×4 mm slices. Two pre-contrast acquisitions were carried out at flip angles of $2°$ and $12°$ to enable the calculation of pre-contrast longitudinal relaxation times ($T_1(0)$). An intravenous bolus injection of 0.1 mmol/kg of gadoterate meglumine (Gd-DOTA) was administered simultaneously with the

start of 20 acquisitions with 12° flip angle and a temporal resolution of 73 s, leading to a DCE-MRI duration of approximately 24 minutes [17].

Cardiac dataset: All data were acquired using a 3T Verio system. Standard cardiac imaging planes and a short axis stack of left ventricular cine data were acquired using routine steady state free precession (TrueFISP) acquisitions. Stress imaging was performed by intravenous infusion of 140 μg/kg/min of adenosine (Adenoscan, Sanofi Aventis). Fifty dynamic perfusion images were obtained at diastole across three short-axis view slices: basal, mid-ventricular and apical slices according to the standard 16-segment heart model. Perfusion images were acquired using a turbo-fast low angle shot (FLASH) saturation recovery prepared single-shot gradient echo pulse sequence (TR/TE = 2.20/1.07 ms, flip angle = 12°, slice thickness = 8 mm, preparation pulse delay (PD) to central line of k-space = 100 ms, matrix size = 192 × 108 and FoV = 330 mm × 440 mm). With the application of GRAPPA (accelerator factor of 3) and partial Fourier acquisition of 0.75, each dynamic frame consisted of 48-phase encoded lines [139].
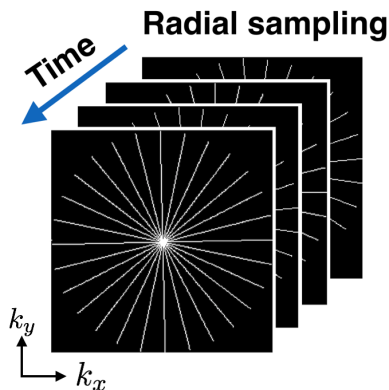

Figure B.1: Exemplary time varying radial sampling pattern in (k_x, k_y) space corresponding to 20-fold acceleration. Equi-angular spacing projections are used and incoherency in time is achieved by applying a random rotation between $[-30°, 30°]$ on the whole pattern across each acquisition frame. Note that the radial sampling is here directly approximated to the closest Cartesian trajectory, hence it is referred to as "pseudo-radial" [140].
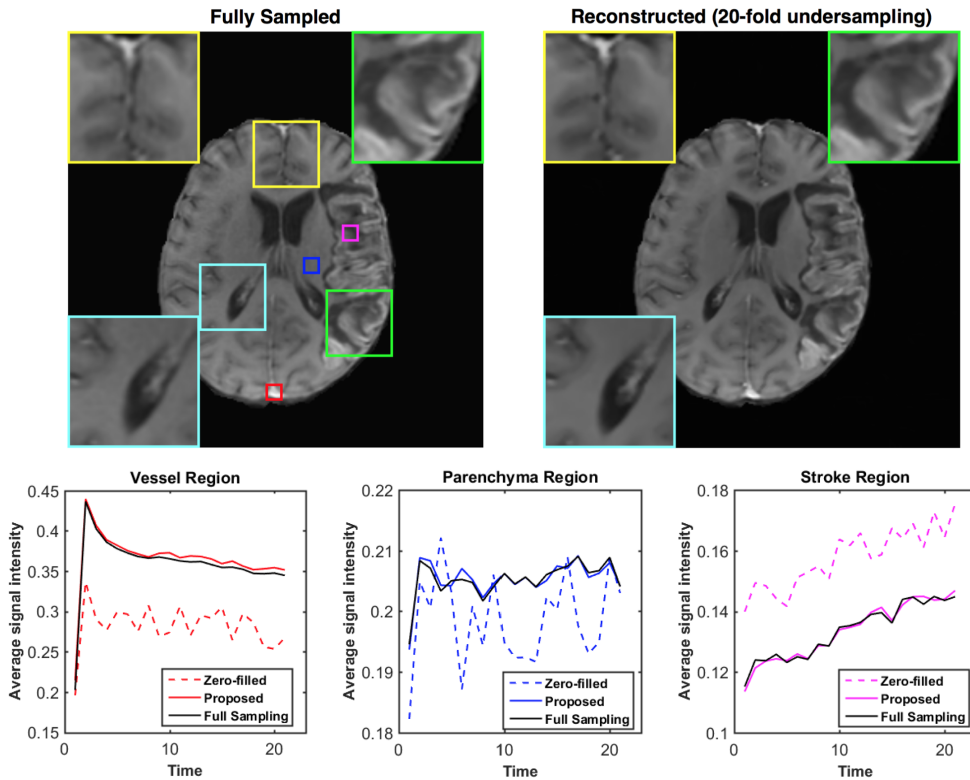
Figure B.2: **(Top)** Close-up views of three different regions of interest (yellow, green and cyan square) from a single frame of a fully sampled image and reconstructed version from 20-fold undersampling. Reconstructed image again provides sharper regions compared to fully sampled data with the preserved finer details. **(Bottom)** TICs displaying the signal intensity over time averaged over the voxels inside the small red, blue and magenta squares – corresponding to a vessel, parenchyma (white matter) and stroke region, respectively – as shown in *Fully Sampled* image. The proposed reconstruction model can achieve highly accurate matching of TICs in different brain regions despite a higher rate of undersampling.

B.2   Random Forest Regression for Pharmacokinetic Parameter Estimation

This study proposes a novel alternative approach to estimate pharmacokinetic (PK) parameters of dynamic contrast enhanced (DCE)-MRI. Our approach leverages machine learning field and mainly targets to automatically learn tem-
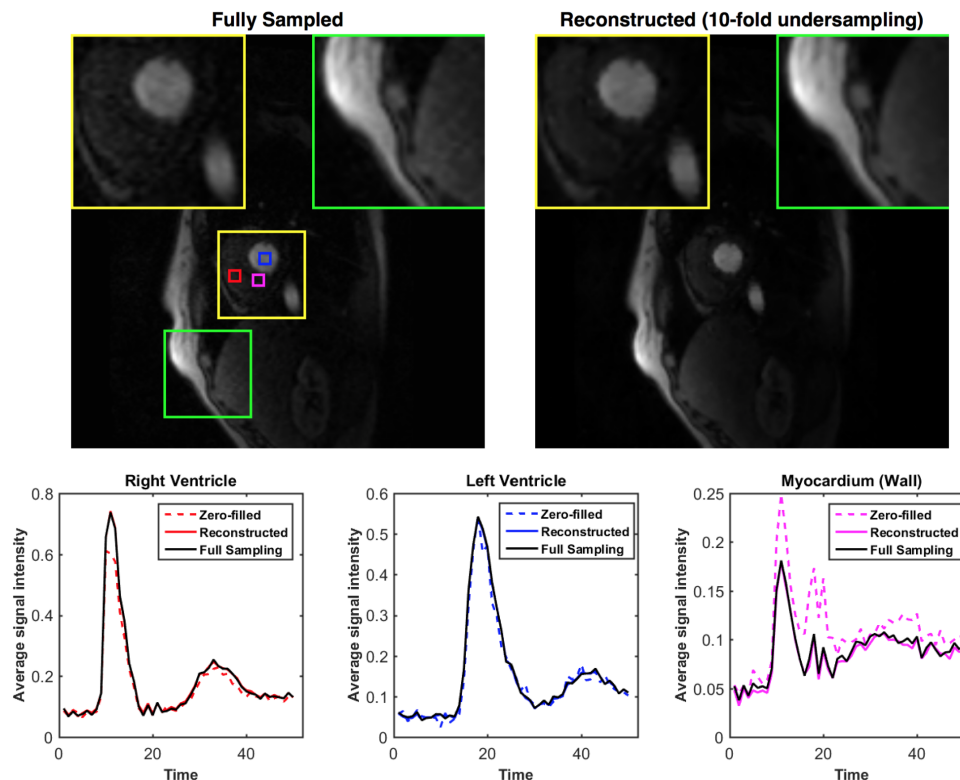
Figure B.3: Results on a basal slice taken from a cardiac rest perfusion acquisition. **(Top)** Close-up views of two different regions of interest (yellow and green square) from a single frame of a fully sampled image and reconstructed version from 10-fold undersampling. Reconstructed image reduces the noise in ventricles and provides sharper regions as displayed in green area. **(Bottom)** TICs displaying the signal intensity over time averaged over the voxels inside the small red, blue and magenta squares – corresponding to a right ventricle, left ventricle and myocardium wall, respectively – as shown in *Fully Sampled* image. The estimated time curves show a strong alignment with the ones obtained from fully sampled data. We remark that TICs obtained from zero-filled reconstruction of cardiac data can successfully capture the signal dynamics especially in ventricles from accelerated acquisitions while this is not observable in brain data.

poral patterns of the voxel-wise concentration-time curves (CTCs) from a large amount of training samples in order to make accurate parameter estimations. We consider the estimation of parameters as a regression problem and specifi-

cally use Random Forest (RF) regression. We demonstrate its potential and utility to improve the conventional model-fitting based quantitative analysis of DCE-MRI especially in various noise conditions, and validate our method on clinical brain stroke datasets.
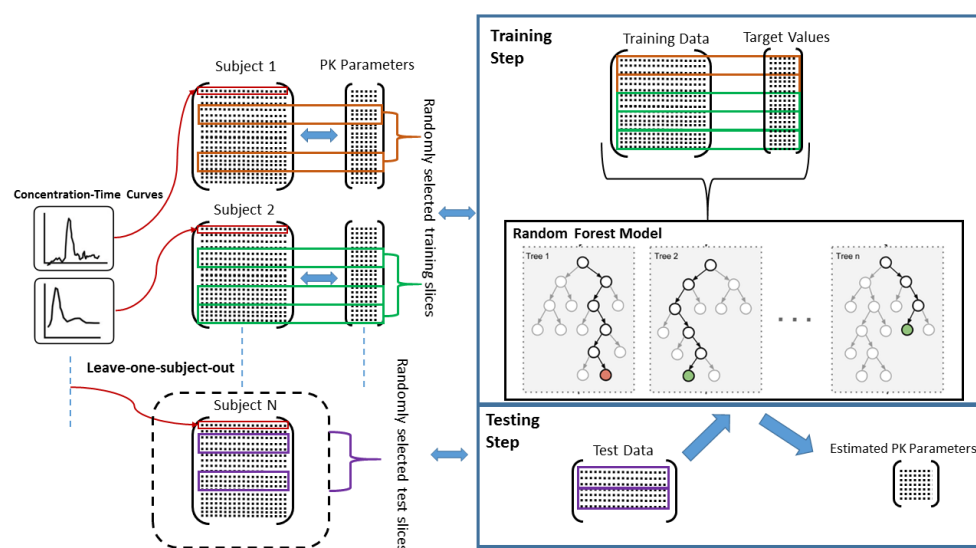

Figure B.4: A diagram illustrating the pipeline of the training and testing procedures of our method. Each subject's data is represented with a matrix where each row involves a concentration-time curve (CTC) per voxel. Training data is created from randomly selected 12 slices of held-in subject's data and test data is obtained from randomly selected 2 slices of held-out subject's data. A RF model is trained from the training data and its corresponding target values. Testing is then performed by giving the test data as input to the RF model which outputs the estimated PK parameters. This process is repeated 50 times (trials) for each subject to obtain unbiased estimation.

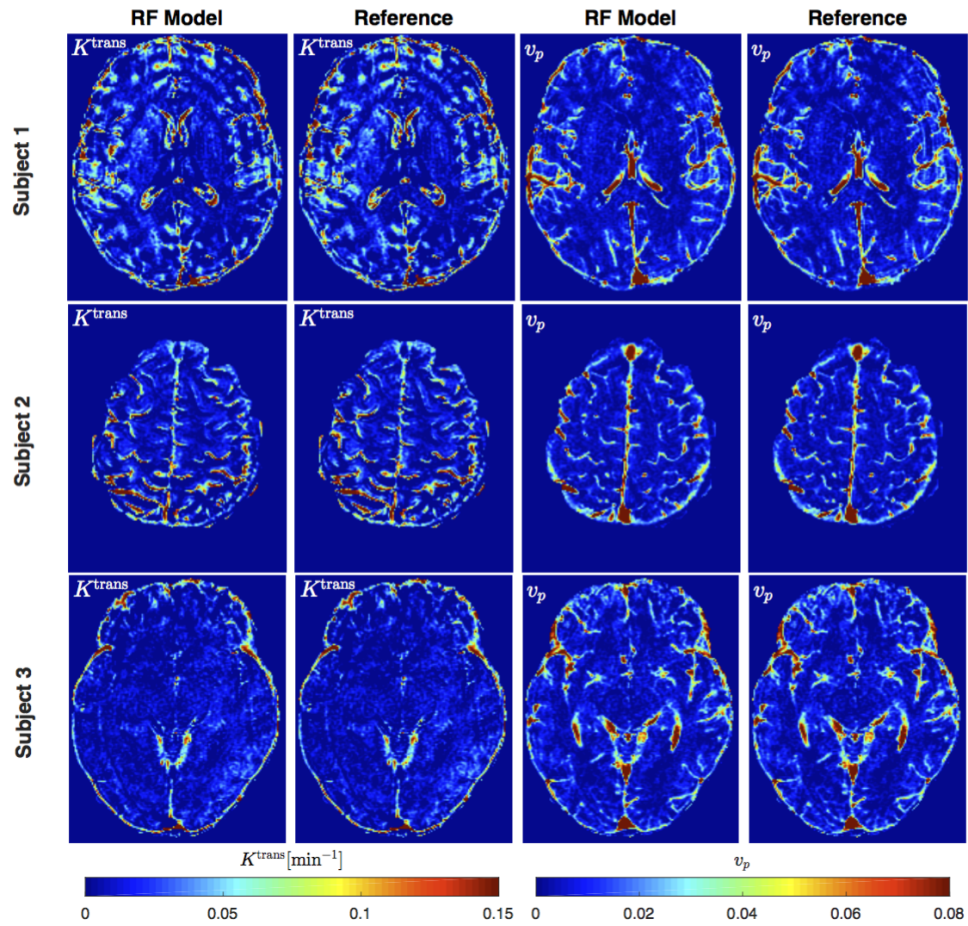Figure B.5: PK parameter maps estimated by RF model on noise-free data taken from each subject. The reference maps are also provided for comparison. In no-noise condition, the estimated maps by RF model are almost similar with reference maps. This asserts that it is possible to directly estimate accurate PK parameters of a subject using other subject's data with the use of a RF based machine learning approach.
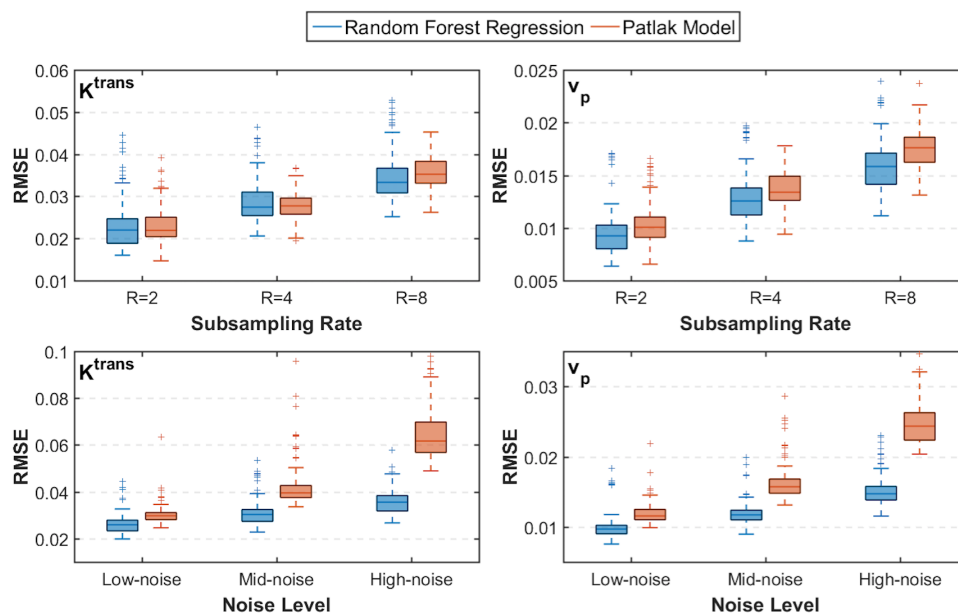
Figure B.6: Boxplots displaying the RMSE statistics of PK parameters – K^{trans}(left), v_p (right) – obtained from our RF regression model and Patlak model on noisy data. The top plots depict results for increasing subsampling factors whereas the bottom plots show results for increasing additive Gaussian noise levels. The RMSE statistics are reported from in total 150 trials of three subjects. The results demonstrate that RF model mostly produces lower median RMSE compared to Patlak model for both PK parameters. Another main conclusion is that RF model is more robust to increasing noise levels rather than subsampling rates.

C

List of Publications

The following authored and co-authored publications were written throughout the thesis period.

Journal Articles

- **C. Ulas**, D. Das, M. J. Thrippleton, M. d. C. Valdés Hernández, P. A. Armitage, S. D. Makin, J. M. Wardlaw, and B. H. Menze. "Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI." In: *Frontiers in Neurology* 9 (2019), p. 1147. DOI: 10.3389/fneur.2018.01147.

○ **C. Ulas**, C. Preibisch, J. Sperl, T. Pyka, J. Kalpathy-Cramer, and B. Menze. "Accelerated Reconstruction of Perfusion-Weighted MRI Enforcing Jointly Local and Nonlocal Spatio-temporal Constraints." In: *arXiv* 1708.07808 (2017).

Peer-reviewed Conference Proceedings

- **C. Ulas**, G. Tetteh, S. Kaczmarz, C. Preibisch, and B. H. Menze. "DeepASL: Kinetic Model Incorporated Loss for Denoising Arterial Spin Labeled MRI via Deep Residual Learning." In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Cham: Springer International Publishing, 2018, pp. 30–38. DOI: 10.1007/978-3-030-00928-1_4.

- **C. Ulas**, G. Tetteh, M. J. Thrippleton, P. A. Armitage, S. D. Makin, J. M. Wardlaw, M. E. Davies, and B. H. Menze. "Direct Estimation of Pharmacokinetic Parameters from DCE-MRI Using Deep CNN with Forward Physical Model Loss." In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Cham: Springer International Publishing, 2018, pp. 39–47. DOI: 10.1007/978-3-030-00928-1_5.

- P. A. Gómez, M. Molina-Romero, **C. Ulas**, G. Bounincontri, J. I. Sperl, D. K. Jones, M. I. Menzel, and B. H. Menze. "Simultaneous Parameter Mapping, Modality Synthesis, and Anatomical Labeling of the Brain with MR Fingerprinting." In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*. Ed. by S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells. Cham: Springer International Publishing, 2016, pp. 579–586. DOI: 10.1007/978-3-319-46726-9_67.

- **C. Ulas**, P. Gómez, J. I. Sperl, C. Preibisch, and B. H. Menze. "Spatio-temporal MRI reconstruction by enforcing local and global regularity via dynamic total variation and nuclear norm minimization." In: *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. 2016, pp. 306–309. DOI: 10.1109/ISBI.2016.7493270.

Peer-reviewed Workshop Proceedings

- P. A. Gómez, **C. Ulas**, J. I. Sperl, T. Sprenger, M. Molina-Romero, M. I. Menzel, and B. H. Menze. "Learning a Spatiotemporal Dictionary for Magnetic Resonance Fingerprinting with Compressed Sensing." In: *Patch-Based Techniques in Medical Imaging*. Ed. by G. Wu, P. Coupé, Y. Zhan, B. Munsell, and D. Rueckert. Cham: Springer International Publishing, 2015, pp. 112–119. DOI: 10.1007/978-3-319-28194-0_14.

- **C. Ulas**, P. A. Gómez, F. Krahmer, J. I. Sperl, M. I. Menzel, and B. H. Menze. "Robust Reconstruction of Accelerated Perfusion MRI Using Local and Nonlocal Constraints." In: *Reconstruction, Segmentation, and Analysis of Medical Images: First Int. Workshops, RAMBO 2016 and HVSMR 2016, in Conjunction with MICCAI*. 2017, pp. 37–47. DOI: 10.1007/978-3-319-52280-7_4.

Peer-reviewed Conference Abstracts

- **C. Ulas**, M. J. Thrippleton, I. Marshall, M. E. Davies, P. A. Armitage, S. D. Makin, J. M. Wardlaw, and B. H. Menze. "Estimation of Pharmacokinetic Parameters in Dynamic Contrast Enhanced MRI Via Random Forest Regression." In: *Proc Intl Soc Mag Reson Med.* 2018.

- **C. Ulas**, S. Kaczmarz, C. Preibisch, J. I. Sperl, M. I. Menzel, A. Haase, and B. H. Menze. "A Spatio-temporal Denoising Approach based on Total Variation Regularization for Arterial Spin Labeling." In: *Proc Intl Soc Mag Reson Med.* 2017.

- P. A. Gómez, G. Bounincontri, M. Molina-Romero, **C. Ulas**, J. I. Sperl, M. I. Menzel, and B. H. Menze. "3D Magnetic Resonance Fingerprinting with a Clustered Spatiotemporal Dictionary." In: *Proc Intl Soc Mag Reson Med.* 2016.

- **C. Ulas**, P. A. Gómez, J. I. Sperl, C. Preibisch, M. I. Menzel, A. Haase, and B. H. Menze. "A Robust Reconstruction Method for Quantitative Perfusion MRI: Application to Brain Dynamic Susceptibility Contrast (DSC) Imaging." In: *Proc. Intl. Soc. Magn. Reson. Med.* 2017.

Bibliography

[1] P. C. Lauterbur. "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance." In: 242.5394 (Mar. 1973), pp. 190–191.

[2] P. Mansfield and P. K. Grannell. "NMR 'diffraction' in solids?" In: *Journal of Physics C: Solid State Physics* 6.22 (Nov. 1973), pp. L422–L426.

[3] M. Vlaardingerbroek and J. den Boer. "Magnetic Resonance Imaging: Theory and Practice. 2nd ed." In: *Radiology* 216.2 (2000), pp. 554–554.

[4] D. T. Ginat and R. Gupta. "Advances in Computed Tomography Imaging Technology." In: *Annual Review of Biomedical Engineering* 16.1 (2014). PMID: 25014788, pp. 431–453.

[5] Y. Bresler. "Fast acquisition and sampling in MRI: Introduction to time-sequential sampling of spatio-temporal signals." In: *Proceedings IEEE International Symposium on Biomedical Imaging.* 2002, pp. 713–716.

[6] M. Lustig, D. Donoho, and J. Pauly. "Sparse MRI: The application of compressed sensing for rapid MR imaging." In: *Magn. Reson. Med.* 58.6 (2007), pp. 1182–1195.

[7] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. "Compressed Sensing MRI." In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 72–82.

[8] J. Caballero, A. Price, D. Rueckert, and J. Hajnal. "Dictionary learning and time sparsity for dynamic MR data reconstruction." In: *IEEE Trans. on Med. Imag.* 33.4 (April 2014), pp. 979–994.

[9] P. A. Gómez Damián. "Accelerating Quantitative Magnetic Resonance Imaging." Dissertation. München: Technische Universität München, 2017.

[10]   D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold. "Magnetic resonance fingerprinting." In: *Nature* 495.7440 (Mar. 2013), pp. 187–92.

[11]   P. A. Gómez, M. Molina-Romero, G. Buonincontri, M. I. Menzel, and B. H. Menze. "Designing contrasts for rapid, simultaneous parameter quantification and flow visualization with quantitative transient-state imaging." In: *Scientific Reports* 9.1 (June 2019), p. 8468.

[12]   J. Caballero. "Dictionaries for Fast and Informative Dynamic MRI Acquisition." PhD thesis. Imperial College London, 2015.

[13]   G.-H. Jahng, K.-L. Li, L. Ostergaard, and F. Calamante. "Perfusion Magnetic Resonance Imaging: A Comprehensive Update on Principles and Techniques." In: *Korean Journal of Radiology* 15.5 (Sept. 2014), pp. 554–577.

[14]   W. A. Copen, P. W. Schaefer, and O. Wu. "MR perfusion imaging in acute ischemic stroke." In: *Neuroimaging clinics of North America* 21.2 (May 2011). S1052-5149(11)00021-9[PII], pp. 259–x.

[15]   T. Ueda, W. T. Yuh, and T. Taoka. "Clinical application of perfusion and diffusion MR imaging in acute ischemic stroke." In: *Journal of Magnetic Resonance Imaging* 10.3 (1999), pp. 305–309.

[16]   J. P. B. O'Connor et al. "Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies." In: *Nat. Rev. Clin. Oncol.* 9.3 (Mar. 2012), pp. 167–77.

[17]   A. K. Heye et al. "Tracer kinetic modelling for DCE-MRI quantification of subtle blood–brain barrier permeability." In: *NeuroImage* 125 (2016), pp. 446–455.

[18]   T. E. Conturo, E. Akbudak, M. S. Kotys, M. L. Chen, S. J. Chun, R. M. Hsu, C. C. Sweeney, and J. Markham. "Arterial input functions for dynamic susceptibility contrast MRI: Requirements and signal options." In: *Journal of Magnetic Resonance Imaging* 22.6 (2005), pp. 697–703.

[19]   F. Calamante. "Arterial input function in perfusion MRI: A comprehensive review." In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 74 (2013), pp. 1–32.

[20]   J. C. Ye. "Compressed sensing MRI: a review from signal processing perspective." In: *BMC Biomedical Engineering* 1.1 (Mar. 2019), p. 8.

[21] D. Boschetto, P. D. Prima, M. Castellaro, A. Bertoldo, and E. Grisan. "Baseline constrained reconstruction of DSC-MRI tracer kinetics from sparse fourier data." In: *Proc. IEEE Int. Symp. Biomed. Imaging.* Apr. 2014, pp. 321–324.

[22] R. M. Lebel et al. "Highly accelerated dynamic contrast enhanced imaging." In: *MRM* 71.2 (2014), pp. 635–644.

[23] G. Adluru, S. P. Awate, T. Tasdizen, R. T. Whitaker, and E. V. DiBella. "Temporally constrained reconstruction of dynamic cardiac perfusion MRI." In: *Magn. Reson. Med.* 57.6 (2007), pp. 1027–1036.

[24] Y. Guo et al. "High-resolution whole-brain DCE-MRI using constrained reconstruction." In: *Med. Phys.* 43.5 (2016), pp. 2013–2023.

[25] Z.-P. Liang and P. C. Lauterbur. *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective.* 1st ed. Wiley-IEEE Press, 2000, p. 416.

[26] E. M. Haacke, ed. *Magnetic resonance imaging: physical principles and sequence design.* 1st ed. New York: Wiley, 1999.

[27] M. H. Levitt. *Spin dynamics: basics of nuclear magnetic resonance.* 2nd ed. OCLC: ocn141380283. Chichester, England ; Hoboken, NJ: John Wiley & Sons, 2008.

[28] F. Bloch. "Nuclear Induction." en. In: *Physical Review* 70.7-8 (Oct. 1946), pp. 460–474.

[29] A. Bustin. "Advanced Reconstruction Techniques in Free-Breathing Multi-Contrast High-Resolution Cardiac Magnetic Resonance Imaging." Dissertation. München: Technische Universität München, 2017.

[30] V. Rasche, R. W. D. Boer, D. Holz, and R. Proksa. "Continuous radial data acquisition for dynamic MRI." In: *Magnetic Resonance in Medicine* 34.5 (1995), pp. 754–761.

[31] B. M. Delattre, R. M. Heidemann, L. A. Crowe, J.-P. Vallée, and J.-N. Hyacinthe. "Spiral demystified." In: *Magnetic Resonance Imaging* 28.6 (2010), pp. 862–881.

[32] F. Knoll, C. Clason, C. Diwoky, and R. Stollberger. "Adapted random sampling patterns for accelerated MRI." In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 24.1 (Feb. 2011), pp. 43–50.

[33] M. Poustchi-Amin, S. A. Mirowitz, J. J. Brown, R. C. McKinstry, and T. Li. "Principles and Applications of Echo-planar Imaging: A Review for the General Radiologist." In: *RadioGraphics* 21.3 (2001). PMID: 11353123, pp. 767–779.

[34] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly. "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity." In: *Proc. 14th Annu. Meet. ISMRM.* 2006, p. 2420.

[35] A. Deshmane., V. Gulani, M. A. Griswold, and N. Seiberlich. "Parallel MR Imaging." In: *J. Mag. Reson. Imag.* 36.1 (2012), pp. 55–72.

[36] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. "SENSE: Sensitivity encoding for fast MRI." In: *Magnetic Resonance in Medicine* 42.5 (1999), pp. 952–962.

[37] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. "Generalized autocalibrating partially parallel acquisitions (GRAPPA)." In: *Magnetic Resonance in Medicine* 47.6 (2002), pp. 1202–1210.

[38] M. Lustig, D. Donoho, J. Santos, and J. Pauly. "Compressed sensing MRI." In: *Signal Processing Magazine, IEEE* 25.2 (2008), pp. 72–82.

[39] I. Tošić and P. Frossard. "Dictionary Learning." In: *IEEE Signal Processing Magazine* 28.2 (2011), pp. 27–38.

[40] J. Caballero, A. N. Price, D. Rueckert, and J. Hajnal. "Dictionary learning and time sparsity for dynamic MR data reconstruction." In: *IEEE Trans. Med. Imag.* 33.4 (2014), pp. 979–994.

[41] H. Jung, J. C. Ye, and E. Y. Kim. "Improved k-t BLAST and k-t SENSE using FOCUSS." In: *Physics in medicine and biology* 52.11 (June 2007), pp. 3201–26.

[42] J. Tsao, P. Boesiger, and K. P. Pruessmann. "k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations." In: *Magnetic Resonance in Medicine* 50.5 (2003), pp. 1031–1042.

[43] S. Mallat. *A wavelet tour of signal processing: the sparse way.* Academic press, 2008.

[44] B. Ophir, M. Lustig, and M. Elad. "Multi-Scale Dictionary Learning Using Wavelets." In: *IEEE Journal of Selected Topics in Signal Processing* 5.5 (2011), pp. 1014–1024.

[45] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob. "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR." In: *IEEE Trans. Med. Imag.* 30.5 (2011), pp. 1042–1054.

[46] C. Chen, Y. Li, L. Axel, and J. Huang. "Real time dynamic MRI with dynamic total variation." In: *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI).* Springer, 2014, pp. 138–145.

[47] S. Ravishankar and Y. Bresler. "MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning." In: *IEEE Transactions on Medical Imaging* 30.5 (2011), pp. 1028–1041.

[48] S. P. Awate and E. V. R. DiBella. "Spatiotemporal dictionary learning for undersampled dynamic MRI reconstruction via joint frame-based and dictionary-based sparsity." In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI).* 2012, pp. 318–321.

[49] Y. Q. Mohsin, S. G. Lingala, E. DiBella, and M. Jacob. "Accelerated dynamic MRI using patch regularization for implicit motion compensation." In: *Magn. Reson. Med.* 77.3 (2017), pp. 1238–1248.

[50] J. Yao et al. "Accelerated Dynamic MRI Reconstruction with Total Variation and Nuclear Norm Regularization." In: *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI).* Springer, 2015, pp. 635–642.

[51] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang. "Image reconstruction from highly undersampled (k,t)-space data with joint partial separability and sparsity constraints." In: *IEEE Trans. on Med. Imag.* 31.9 (September 2012), pp. 1809–1820.

[52] R. Otazo, E. Candès, and D. K. Sodickson. "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components." In: *Magnetic Resonance in Medicine* 73.3 (2015), pp. 1125–1136.

[53] C. Chen and J. Huang. "Compressive Sensing MRI with Wavelet Tree Sparsity." In: *Advances in Neural Information Processing Systems.* Ed. by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger. Vol. 25. Curran Associates, Inc., 2012, pp. 1115–1123.

[54] K. K. Bhatia, J. Caballero, A. N. Price, Y. Sun, J. V. Hajnal, and D. Rueckert. "Fast Reconstruction of Accelerated Dynamic MRI Using Manifold Kernel Regression." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi. Cham: Springer International Publishing, 2015, pp. 510–518.

[55] S. Poddar and M. Jacob. "Dynamic MRI Using SmooThness Regularization on Manifolds (SToRM)." In: *IEEE Transactions on Medical Imaging* 35.4 (2016), pp. 1106–1115.

[56] F. Calamante, D. Gadian, and A. Connelly. "Quantification of Perfusion Using Bolus Tracking Magnetic Resonance Imaging in Stroke." In: *Stroke* 33.4 (2002), pp. 1146–1151.

[57] F. Calamante. "Perfusion MRI using dynamic-susceptibility contrast MRI: quantification issues in patient studies." In: *Topics in magnetic resonance imaging : TMRI* 21.2 (Apr. 2010), pp. 75–85.

[58] L. Østergaard. "Principles of cerebral perfusion imaging by bolus tracking." In: *J. Magn. Reson. Imaging* 22.6 (2005), pp. 710–717.

[59] D. J. Collins and A. R. Padhani. "Dynamic magnetic resonance imaging of tumor perfusion." In: *IEEE Engineering in Medicine and Biology Magazine* 23.5 (Sept. 2004), pp. 65–83.

[60] J. Zhao, Z.-y. Yang, B.-n. Luo, J.-y. Yang, and J.-p. Chu. "Quantitative Evaluation of Diffusion and Dynamic Contrast-Enhanced MR in Tumor Parenchyma and Peritumoral Area for Distinction of Brain Tumors." In: *PLOS ONE* 10.9 (Sept. 2015), pp. 1–15.

[61] J. M. Wardlaw, C. Smith, and M. Dichgans. "Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging." In: *The Lancet Neurology* 12.5 (2018/04/29 2013), pp. 483–497.

[62] A. K. Heye, R. D. Culling, M. del C. Valdés Hernández, M. J. Thrippleton, and J. M. Wardlaw. "Assessment of blood–brain barrier disruption using dynamic contrast-enhanced MRI. A systematic review." In: *NeuroImage: Clinical* 6 (2014), pp. 262–274.

[63] G. J. M. Parker and D. L. Buckley. "Tracer Kinetic Modelling for T1-Weighted DCE-MRI." In: *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*. Ed. by A. Jackson, D. L. Buckley, and G. J. M. Parker. Springer Berlin Heidelberg, 2005, pp. 81–92.

[64] T. E. Yankeelov and J. C. Gore. "Dynamic Contrast Enhanced Magnetic Resonance Imaging in Oncology:Theory, Data Acquisition,Analysis, and Examples." In: *Current Medical Imaging Reviews* 3.2 (2007), pp. 91–107.

[65] G. Brix, W. Semmler, R. Port, L. R. Schad, G. Layer, W. J. Lorenz, et al. "Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging." In: *Journal of computer assisted tomography* 15.4 (1991), p. 621.

[66] P. S. Tofts and A. G. Kermode. "Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts." In: *Magnetic Resonance in Medicine* 17.2 (1991), pp. 357–367.

[67] P. S. Tofts, G. Brix, D. L. Buckley, J. L. Evelhoch, E. Henderson, M. V. Knopp, H. B. Larsson, T. Lee, N. A. Mayr, G. J. Parker, R. E. Port, J. Taylor, and R. M. Weisskoff. "Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted MRI of a diffusable tracer: Standardized quantities and symbols." In: *Journal of Magnetic Resonance Imaging* 10.3 (1999), pp. 223–232.

[68] S. P. Sourbron and D. L. Buckley. "Tracer kinetic modelling in MRI: estimating perfusion and capillary permeability." In: *Physics in medicine and biology* 57.2 (2012), R1–33.

[69] T. S. Koh, S. Bisdas, D. M. Koh, and C. H. Thng. "Fundamentals of tracer kinetics for dynamic contrast-enhanced MRI." In: *Journal of Magnetic Resonance Imaging* 34.6 (2011), pp. 1262–1276.

[70] D. S. Williams, J. A. Detre, J. S. Leigh, and A. P. Koretsky. "Magnetic resonance imaging of perfusion using spin inversion of arterial water." In: *Proceedings of the National Academy of Sciences of the United States of America* 89.1 (Jan. 1992). PMC48206[pmcid], pp. 212–216.

[71] J. A. Detre, J. S. Leigh, D. S. Williams, and A. P. Koretsky. "Perfusion imaging." In: *Magnetic Resonance in Medicine* 23.1 (1992), pp. 37–45.

[72] N. A. Telischak, J. A. Detre, and G. Zaharchuk. "Arterial spin labeling MRI: Clinical applications in the brain." In: *Journal of Magnetic Resonance Imaging* 41.5 (2015), pp. 1165–1180.

[73] J. M. Watts, C. T. Whitlow, and J. A. Maldjian. "Clinical applications of arterial spin labeling." In: *NMR in Biomedicine* 26.8 (2013), pp. 892–900.

[74]   J. Petr, J.-C. Ferre, J.-Y. Gauvrit, and C. Barillot. "Improving arterial spin labeling data by temporal filtering." In: *Medical Imaging 2010: Image Processing*. Ed. by B. M. Dawant and D. R. Haynor. Vol. 7623. International Society for Optics and Photonics. SPIE, 2010, pp. 1011–1019.

[75]   D. C. Alsop et al. "Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia." In: *MRM* 73.1 (2015), pp. 102–116.

[76]   S. M. Spann et al. "Spatio-temporal TGV denoising for ASL perfusion imaging." In: *Neuroimage* 157 (2017), pp. 81–96.

[77]   P. van Gelderen, J. de Zwart, and J. Duyn. "Pittfalls of MRI measurement of white matter perfusion based on arterial spin labeling." In: *Magnetic Resonance in Medicine* 59.4 (2008), pp. 788–795.

[78]   X. Liang et al. "Voxel-Wise Functional Connectomics Using Arterial Spin Labeling Functional Magnetic Resonance Imaging: The Role of Denoising." In: *Brain connectivity* 5 9 (2015), pp. 543–53.

[79]   R. B. Buxton et al. "A general kinetic model for quantitative perfusion imaging with arterial spin labeling." In: *MRM* 40.3 (1998), pp. 383–396.

[80]   A. Borogovac and I. Asllani. "Arterial Spin Labeling (ASL) fMRI: Advantages, Theoretical Constrains and Experimental Challenges in Neurosciences." In: *International Journal of Biomedical Imaging* 2012 (Feb. 2012), p. 818456.

[81]   A. Melbourne, N. Toussaint, D. Owen, I. Simpson, T. Anthopoulos, E. De Vita, D. Atkinson, and S. Ourselin. "NiftyFit: a Software Package for Multi-parametric Model-Fitting of 4D Magnetic Resonance Imaging Data." In: *Neuroinformatics* 14.3 (July 2016), pp. 319–337.

[82]   M. A. Chappell, A. R. Groves, B. Whitcher, and M. W. Woolrich. "Variational Bayesian Inference for a Nonlinear Forward Model." In: *IEEE Transactions on Signal Processing* 57.1 (2009), pp. 223–236.

[83]   L. M. Parkes. "Quantification of cerebral perfusion using arterial spin labeling: Two-compartment models." In: *Journal of Magnetic Resonance Imaging* 22.6 (2005), pp. 732–736.

[84] C. Ulas, P. Gómez, J. I. Sperl, C. Preibisch, and B. H. Menze. "Spatio-temporal MRI reconstruction by enforcing local and global regularity via dynamic total variation and nuclear norm minimization." In: *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. 2016, pp. 306–309.

[85] C. Ulas, P. A. Gómez, F. Krahmer, J. I. Sperl, M. I. Menzel, and B. H. Menze. "Robust Reconstruction of Accelerated Perfusion MRI Using Local and Nonlocal Constraints." In: *Reconstruction, Segmentation, and Analysis of Medical Images: First Int. Workshops, RAMBO 2016 and HVSMR 2016, in Conjunction with MICCAI*. 2017, pp. 37–47.

[86] C. Ulas, C. Preibisch, J. Sperl, T. Pyka, J. Kalpathy-Cramer, and B. Menze. "Accelerated Reconstruction of Perfusion-Weighted MRI Enforcing Jointly Local and Nonlocal Spatio-temporal Constraints." In: *arXiv* 1708.07808 (2017).

[87] C. Ulas, G. Tetteh, M. J. Thrippleton, P. A. Armitage, S. D. Makin, J. M. Wardlaw, M. E. Davies, and B. H. Menze. "Direct Estimation of Pharmacokinetic Parameters from DCE-MRI Using Deep CNN with Forward Physical Model Loss." In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Cham: Springer International Publishing, 2018, pp. 39–47.

[88] C. Ulas, D. Das, M. J. Thrippleton, M. d. C. Valdés Hernández, P. A. Armitage, S. D. Makin, J. M. Wardlaw, and B. H. Menze. "Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI." In: *Frontiers in Neurology* 9 (2019), p. 1147.

[89] C. Ulas, G. Tetteh, S. Kaczmarz, C. Preibisch, and B. H. Menze. "Deep-ASL: Kinetic Model Incorporated Loss for Denoising Arterial Spin Labeled MRI via Deep Residual Learning." In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Cham: Springer International Publishing, 2018, pp. 30–38.

[90] E. J. Candes, J. Romberg, and T. Tao. "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information." In: *IEEE Transactions on Information Theory* 52.2 (2006), pp. 489–509.

[91]   M. Rani, S. B. Dhok, and R. B. Deshmukh. "A Systematic Review of Compressive Sensing: Concepts, Implementations and Applications." In: *IEEE Access* 6 (2018), pp. 4875–4894.

[92]   R. E. Carrillo, A. B. Ramirez, G. R. Arce, K. E. Barner, and B. M. Sadler. "Robust compressive sensing of sparse signals: a review." In: *EURASIP Journal on Advances in Signal Processing* 2016.1 (Oct. 2016), p. 108.

[93]   J. A. Fessler. "Model-Based Image Reconstruction for MRI." In: *IEEE Signal Processing Magazine* 27.4 (2010), pp. 81–89.

[94]   M. Sandilya and S. Nirmala. "Compressed sensing trends in magnetic resonance imaging." In: *Engineering Science and Technology, an International Journal* 20.4 (2017), pp. 1342–1352.

[95]   C. M. Bishop. *Pattern recognition and machine learning.* Springer-Verlag New York, 2006.

[96]   T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction.* 2nd ed. Springer, 2009.

[97]   S. Marsland. *Machine Learning: An Algorithmic Perspective, Second Edition.* 2nd. Chapman & Hall/CRC, 2014.

[98]   A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems.* 2012, pp. 1097–1105.

[99]   I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning.* MIT press, 2016.

[100] W. Hager and H. Zhang. "A survey of nonlinear conjugate gradient methods." In: *Pacific journal of Optimization* 2.1 (2006), pp. 35–58.

[101] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. "Distributed optimization and statistical learning via the alternating direction method of multipliers." In: *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122.

[102] P. Tseng. "Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities." In: *SIAM Journal on Control and Optimization* 29.1 (1991), pp. 119–138.

[103] I. Daubechies, M. Defrise, and C. De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." In: *Communications on pure and applied mathematics* 57.11 (2004), pp. 1413–1457.

[104] A. Beck and M. Teboulle. "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems." In: *SIAM J. on Imag. Sci.* 2.1 (2009), pp. 183–202.

[105] P. L. Combettes and V. R. Wajs. "Signal Recovery by Proximal Forward-Backward Splitting." In: *Multiscale Modeling & Simulation* 4.4 (2005), pp. 1168–1200.

[106] J. A. Fessler. "Optimization Methods for Magnetic Resonance Image Reconstruction: Key Models and Optimization Algorithms." In: *IEEE Signal Processing Magazine* 37.1 (2020), pp. 33–40.

[107] C. Cortes and V. Vapnik. "Support-vector networks." In: *Machine learning* 20.3 (1995), pp. 273–297.

[108] L. Breiman. "Random forests." In: *Machine learning* 45.1 (2001), pp. 5–32.

[109] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386.

[110] G. Cybenko. "Approximation by superpositions of a sigmoidal function." In: *Mathematics of Control, Signals, and Systems (MCSS)* 2.4 (1989), pp. 303–314.

[111] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning." In: *Nature* 521.7553 (May 2015), pp. 436–444.

[112] S. Wang and R. M. Summers. "Machine learning and radiology." In: *Medical Image Analysis* 16.5 (2012), pp. 933–951.

[113] D. Shen, G. Wu, and H.-I. Suk. "Deep Learning in Medical Image Analysis." In: *Annual Review of Biomedical Engineering* 19.1 (2017). PMID: 28301734, pp. 221–248.

[114] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew. "Deep learning for visual understanding: A review." In: *Neurocomputing* 187 (2016). Recent Developments on Deep Big Vision, pp. 27–48.

[115] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis. "Deep Learning for Computer Vision: A Brief Review." In: *Computational Intelligence and Neuroscience* 2018 (Feb. 2018), p. 7068349.

[116] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi. "A survey of the recent architectures of deep convolutional neural networks." In: *Artificial Intelligence Review* 53.8 (Dec. 2020), pp. 5455–5516.

[117] J. Schlemper, J. Caballero, A. Aitken, and J. R. van Amersfoort. "Deep Hashing using Entropy Regularised Product Quantisation Network." In: *ArXiv* abs/1902.03876 (2019).

[118] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. "On the importance of initialization and momentum in deep learning." In: *Proceedings of the 30th International Conference on Machine Learning.* Ed. by S. Dasgupta and D. McAllester. Vol. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA: PMLR, 17–19 Jun 2013, pp. 1139–1147.

[119] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization." In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.* Ed. by Y. Bengio and Y. LeCun. 2015.

[120] M. Sarma, P. Hu, S. Rapacchi, D. Ennis, A. Thomas, P. Lee, P. Kupelian, and K. Sheng. "Accelerating dynamic magnetic resonance imaging (MRI) for lung tumor tracking based on low-rank decomposition in the spatial-temporal domain: a feasibility study based on simulation and preliminary prospective undersampled MRI." In: *International journal of radiation oncology, biology, physics* 88.3 (Mar. 2014). S0360-3016(13)03574-8[PII], pp. 723–731.

[121] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert. "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction." In: *IEEE Transactions on Medical Imaging* 37.2 (2018), pp. 491–503.

[122] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert. "Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction." In: *IEEE Transactions on Medical Imaging* 38.1 (2019), pp. 280–290.

[123] S. Biswas, H. K. Aggarwal, and M. Jacob. "Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM." In: *Magnetic Resonance in Medicine* 82.1 (2019), pp. 485–494.

[124] A. Bustin, N. Fuin, R. M. Botnar, and C. Prieto. "From Compressed-Sensing to Artificial Intelligence-Based Cardiac MRI Reconstruction." In: *Frontiers in Cardiovascular Medicine* 7 (2020), p. 17.

[125] G. Seegoolam, J. Schlemper, C. Qin, A. Price, J. Hajnal, and D. Rueckert. "Exploiting Motion for Deep Learning Reconstruction of Extremely-Undersampled Dynamic MRI." In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019.* Ed. by D. Shen, T. Liu,

T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan. Cham: Springer International Publishing, 2019, pp. 704–712.

[126] Q. Huang, Y. Xian, D. Yang, H. Qu, J. Yi, P. Wu, and D. N. Metaxas. "Dynamic MRI reconstruction with end-to-end motion-guided network." In: *Medical Image Analysis* 68 (2021), p. 101901.

[127] R. Meier, P. Lux, B. Med, S. Jung, U. Fischer, J. Gralla, M. Reyes, R. Wiest, R. McKinley, and J. Kaesmacher. "Neural Network–derived Perfusion Maps for the Assessment of Lesions in Patients with Acute Ischemic Stroke." In: *Radiology: Artificial Intelligence* 1.5 (2019), e190019.

[128] J. Zou, J. M. Balter, and Y. Cao. "Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network." In: *Medical Physics* 47.8 (2020), pp. 3447–3457.

[129] Y. Bliesener, J. Acharya, and K. S. Nayak. "Efficient DCE-MRI Parameter and Uncertainty Estimation Using a Neural Network." In: *IEEE Transactions on Medical Imaging* 39.5 (2020), pp. 1712–1723.

[130] Q. Yang, Y. Liu, T. Chen, and Y. Tong. "Federated Machine Learning: Concept and Applications." In: *ACM Trans. Intell. Syst. Technol.* 10.2 (Jan. 2019).

[131] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso. "The future of digital health with federated learning." In: *npj Digital Medicine* 3.1 (Sept. 2020), p. 119.

[132] K. H. Kim et al. "Improving Arterial Spin Labeling by Using Deep Learning." In: *Radiology* 287.2 (2018), pp. 658–666.

[133] K. Gong, P. Han, G. El Fakhri, C. Ma, and Q. Li. "Arterial spin labeling MR image denoising and reconstruction using unsupervised deep learning." In: *NMR in Biomedicine* n/a.n/a (Dec. 2019), e4224.

[134] D. Xie, Y. Li, H. Yang, L. Bai, T. Wang, F. Zhou, L. Zhang, and Z. Wang. "Denoising arterial spin labeling perfusion MRI with deep machine learning." In: *Magnetic Resonance Imaging* 68 (2020), pp. 95–105.

[135]   D. Owen, A. Melbourne, Z. Eaton-Rosen, D. L. Thomas, N. Marlow, J. Rohrer, and S. Ourselin. "Deep Convolutional Filtering for Spatio-Temporal Denoising and Artifact Removal in Arterial Spin Labelling MRI." In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Cham: Springer International Publishing, 2018, pp. 21–29.

[136]   P. W. Hales, J. Pfeuffer, and C. A. Clark. "Combined Denoising and Suppression of Transient Artifacts in Arterial Spin Labeling MRI Using Deep Learning." In: *Journal of Magnetic Resonance Imaging* 52.5 (2020), pp. 1413–1426.

[137]   J. Cui, K. Gong, P. Han, H. Liu, and Q. Li. "Super Resolution of Arterial Spin Labeling MR Imaging Using Unsupervised Multi-scale Generative Adversarial Network." In: *Machine Learning in Medical Imaging*. Ed. by M. Liu, P. Yan, C. Lian, and X. Cao. Cham: Springer International Publishing, 2020, pp. 50–59.

[138]   C. Ulas, M. J. Thrippleton, I. Marshall, M. E. Davies, P. A. Armitage, S. D. Makin, J. M. Wardlaw, and B. H. Menze. "Estimation of Pharmacokinetic Parameters in Dynamic Contrast Enhanced MRI Via Random Forest Regression." In: *Proc Intl Soc Mag Reson Med*. 2018.

[139]   G. Papanastasiou, M. C. Williams, L. E. Kershaw, M. R. Dweck, S. Alam, S. Mirsadraee, M. Connell, C. Gray, T. MacGillivray, D. E. Newby, and S. I. Semple. "Measurement of myocardial blood flow by cardiovascular magnetic resonance perfusion: comparison of distributed parameter and Fermi models with single and dual bolus." In: *Journal of Cardiovascular Magnetic Resonance* 17.1 (Feb. 2015), p. 17.

[140]   B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge. "Dynamic MR Image Reconstruction-Separation From Undersampled (k-t)-Space via Low-Rank Plus Sparse Prior." In: *IEEE Trans. Med. Imag.* 33.8 (2014), pp. 1689–1701.