Efficacy and Utility of Image Quality Metrics in Magnetic Resonance Image Reconstruction

by

Allister William Mason

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2019

For my grandparents, Marion and Donald, my parents, Carolyn and Bill, and brother, Layton. Thank you for your never-ending support.

Table of Contents

List of Tables

List of Figures

Abstract

Quantitative measurement of MRI image quality is a difficult, but important problem that can have direct and indirect benefits to MRI research. Objective image quality metrics (IQMs), such as root-mean-squared-error (RMSE) and structural similarity (SSIM), are routinely employed in a research context to evaluate MRI image quality. Here, work is presented that investigates the efficacy of 10 common IQMs including RMSE and SSIM at measuring MRI image quality as perceived by radiologists. While some metrics demonstrated high correlation with the radiologists' scores, RMSE and SSIM were not among the top performing metrics. SSIM did not perform statistically superior to RMSE. In a second study, an IQM-based heuristic is developed to determine an optimal temporal footprint in dynamic-contrast-enhanced (DCE) MRI for pharmacokinetic parameter measurement, specifically the volume transfer constant K^{trans}. In simulations, the heuristic was able to consistently recover K^{trans} within 10%. This technique may allow for selection of temporal footprint in DCE MRI at an individual level.

List of Abbreviations and Symbols Used

$\vec{M}$	Net magnetization vector
$\vec{\mu}$	Magnetic moment
V	Volume
$\vec{B}$	Net magnetic field
$\vec{B}_0$	Main magnetic field from scanner
$\vec{r}$	Spatial position
t	Time
M_0	Equilibrium magnetization strength
γ	Gyromagnetic ratio
$\vec{M}_{xy}$	Transverse magnetization
T_1	Longitudinal relaxation time
T_2	Transverse relaxation time
α	Flip angle
ω	Precessional frequency
ϕ	Spin phase
$\vec{G}$	Gradient field
s	MRI signal
$\vec{k}$	Spatial frequency vector
ρ	Effective proton density
A	Sensing matrix
λ	Regularization weight
Ψ	Sparsifying transform
K^{trans}	Volume transfer constant
v_p	Blood plasma volume per unit volume of tissue
v_e	EES volume per unit volume of tissue
k_{ep}	Efflux rate constant from EES to plasma
C_t	Contrast agent time course in tissue
C_p	Contrast agent time course in blood plasma
C_b	Contrast agent time course in whole blood
r_1	Longitudinal relaxivity

AIF	Arterial Input Function
ACS	Auto Calibration Signal
BART	Berkeley Advanced Reconstruction Toolbox
CIRCUS	CIRcular Cartesian UnderSampling
CS	Compressed Sensing
DCE	Dynamic Contrast Enhanced
DISCO	DIfferential Subsampling with Cartesian Ordering
DWI	Diffusion Weighted Imaging
EES	Extravascular Extracellular Space
FSIM	Feature SIMilarity index
GMSD	Gradient Magnitude Similarity Deviation
GRAPPA	Golden Angle Radial Sparse Parallel
HDRVDP	High Dynamic Range Visual Difference Predictor
IQM	Image Quality Metric
IWSSIM	Information Weighted SSIM
MRI	Magnetic Resonance Imaging
MSSSIM	Multi Scale SSIM
NQM	Noise Quality Metric
nufft	Non-Uniform Fast Fourier Transform
PI	Parallel Imaging
PI-RADS	Prostate Imaging – Reporting and Data System
PK	Pharmacokinetic
PSNR	Peak Signal to Noise Ratio
RF	Radiofrequency
RMSE	Root Mean Squared Error
SENSE	Sensitivity Encoding
SNR	Signal to Noise Ratio
SPGR	Spoiled Gradient
SROCC	Spearman Rank Order Correlation Coefficient
SSIM	Structural Similarity Index
TR	Repetition Time
VIF	Visual Information Fidelity

Acknowledgements

I would like to start by acknowledging all the support provided by my co-supervisors Steven Beyea and James Rioux across a range of topics from a technical to personal level. Your guidance has been much appreciated. Thanks also to my other committee members Sharon Clarke and Dirk Arnold for providing valuable feedback throughout this work.

There are many members of the NSHA and Dalhousie community without whom this work would also not be possible. Thank you to the faculty and staff of the Medical Physics and Physics departments, the friendly faces of BIOTIC, the MR technologists, and the radiologists who contributed to this work. Working in this community has been an honour and a privilege. I'd also like to acknowledge Nathan Murtha who completed much of the groundwork for the work presented here.

Thanks to the other grad students in the Medical Physics program who understand the difficulties of this program like few others and who made the day-to-day of this program as enjoyable as possible. Specifically, my classmates Miriam, Mary, Anna, Matthew, and Jenny. We survived some trying times together. Finally, I'd like to acknowledge my family and my friends in Halifax, Sydney, and back home for providing many much-needed escapes during my studies.

Chapter 1: Introduction

1.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a sophisticated diagnostic imaging technique used in modern medicine. MR images are generated from the signal detected from the magnetic moment of an atomic nucleus. In diagnostic MRI, this nucleus of choice is the single proton nucleus of the hydrogen atom found primarily in water and fat. Due to the large water content in soft tissue, MRI is particularly advantageous for imaging of soft tissues such as brain, breast, and other organs in the abdomen. Another advantage of MRI over other medical imaging techniques such as X-ray, computed tomography, or positron emission tomography is that it does not use ionizing radiation. MRI is also unique among medical imaging technologies in that it uses radiation of a wavelength that is much larger than the resolution of the resulting image. This can be accomplished because the raw data collected for an MR image are not directly sensed in the image domain, but in the spatial frequency/Fourier domain. Fourier components are discretely sampled and an image is then generated by inverse Fourier transform or other advanced reconstruction techniques. Unfortunately, this discrete sampling of Fourier components is a time-consuming process, resulting in long MRI scan times.

A common source of image contrast in MRI is the different relaxation rates of the hydrogen nucleus in different tissues. Properties such as tissue density, phase, and chemical composition will affect the rate at which the local signal decays resulting in different local signal intensity. It was the apparent difference in relaxation rates between tumors and healthy tissues that originally motivated the use of MRI as a diagnostic tool [1]. Since this time, MRI has evolved to be more than just a device for anatomical imaging. Functional

MRI (fMRI) can measure neural activity based on difference of blood flow in different areas [2]. Diffusion weighted imaging (DWI) can generate maps of the apparent diffusion coefficient of water within a tissue [3]. Magnetic resonance spectroscopy is a technique that provides information about the presence of metabolites in a tissue [4]. Of particular importance to this thesis is a technique called *dynamic contrast enhanced (DCE) MRI* [5], in which image contrast is affected by the permeability and perfusion in the microvasculature of a tissue after the intravenous injection of a contrast agent.

1.2 Dynamic Contrast Enhanced MRI

DCE MRI is a dynamic imaging technique, which means that a series of images of the same anatomical region are collected sequentially over a period of time, typically between five to ten minutes. DCE MRI seeks to track the kinetics of a contrast agent as is passes through the circulatory system and tissues. During a clinical scan, the contrast agent is injected intravenously and travels through the blood stream. The contrast agent can also escape from blood vessels into the extravascular extracellular tissue space, so the local contrast agent concentration will depend both on local blood perfusion and blood vessel permeability [5]. The contrast agent is usually gadolinium based, which is a paramagnetic material. This alters local relaxation rates leading to local signal enhancement. DCE MRI scans can be used in many locations in the body such as the brain [6], heart [7], breast [8], kidney [9], and prostate [5]. In the prostate, it is useful for detecting subtle lesions. However, due to significant overlap of behavior between healthy and cancerous tissue it can only be used for diagnosis if the suspicious region shows abnormality in the corresponding T2 weighted images or DWI scan [10].

Cancerous tissues can display earlier enhancement in DCE scans than healthy tissue, so a radiologist can scroll through the image time series and look for areas of early local enhancement. This visual assessment can be complemented by semi-quantitative and quantitative information by extracting parameters from the *signal enhancement curve* in a pixel or region of interest. Semi-quantitative parameters include parameters such as the slope, time-to-peak, or the area under the signal enhancement curve (Figure 1.1). While the relative values of these parameters can be useful, the absolute values are meaningless due to sensitivities to imaging site dependent features such as magnetic field strength, imaging protocol, or amount of contrast agent injected [11]. These sensitivities can cause difficulties with semi-quantitative parameters when trying to determine a diagnostic threshold or when comparing the progress of a disease in a patient over time. Quantitative DCE parameters define physiologically meaningful tissue properties that should be independent of these features by using an appropriate *pharmacokinetic* (PK) model. Common PK parameters include the volume transfer constant, K^{trans}, the percent volume of extravascular extracellular space, v_e, and the percent volume of blood plasma, v_p. Studies have shown

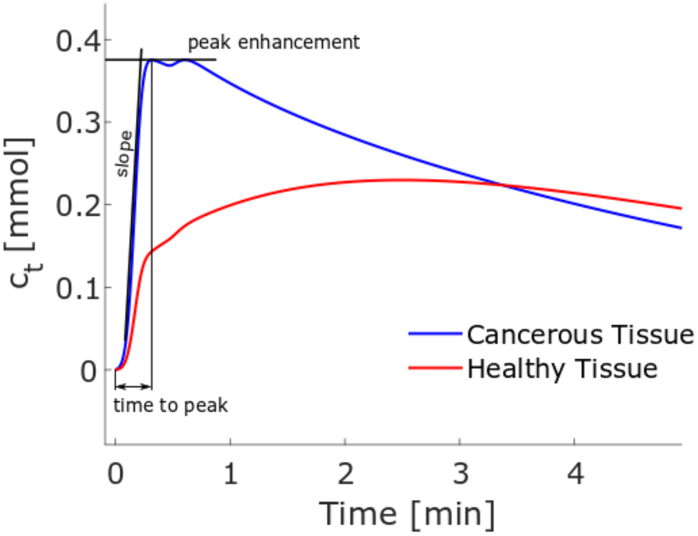


Figure 1.1: Time courses of contrast agent concentration (derived from signal enhancement curve) in tissues showing representative behaviour for cancerous and healthy tissues. Semi-quantitative parameters are also shown.

that these parameters have the potential to serve as biomarkers to differentiate between healthy and cancerous tissues [8, 12, 13] as well as track therapy progression [14, 15], but to date there is still too much variability in their measurements to warrant widespread clinical adoption [16]. A precise and accurate technique for measurement of these PK parameters could be clinically beneficial.

One of the fundamental difficulties in DCE MRI is the trade-off between temporal resolution (i.e. the time between each image of the time series) and the quality of the images. Collecting more data for each image will result in higher image quality, but at the expense of temporal resolution. This trade-off can result in loss of the visualization of the early enhancement of a tumor or inaccurate PK parameter mapping. Golden angle-based radial imaging techniques such as GRASP [17] allow for multiple reconstructions with arbitrary temporal resolution chosen retrospectively, but a technique for making the choice of temporal resolution still does not exist. Faster imaging techniques may help mitigate this problem. Recent advances in MR imaging speeds via techniques such as parallel imaging, compressed sensing, and machine learning are helping to generate high image quality DCE scans with higher temporal resolution.

1.3 Accelerated MR Imaging

MRI has always suffered from slow imaging speeds. Early research looked at ways of increasing the speed of data collection. This included techniques such as stronger magnetic field gradients/faster slew rates and more advanced pulse sequences. However, there are physical, physiological, and engineering limitations to this kind of approach. For instance, magnetic relaxation rates limit sequence repetition times, tissue specific absorption rate and peripheral nerve stimulation limits the strength of magnetic fields and gradient fields, and imaging pulse sequences can only be made so efficient. This fundamentally limits the possible sampling speed of MRI. Due to these limitations much focus has been placed on advanced image reconstruction techniques that can generate high quality images with less data. It is common practice to sample only a subset of the available Fourier components of the MR image, a technique known as *undersampling,* resulting in scan time *acceleration*. However, this can violate the Nyquist criteria, which will lead to aliasing and a degradation in image quality. Many modern approaches then take advantage of the enforced and/or inherent redundancy in MRI data to attempt to de-alias the undersampled image.

An acceleration technique that takes advantage of enforced data redundancy is *parallel imaging* (PI). PI makes use of multiple receiver coils and the inherent spatial sensitivities of each coil to localize the MR signal. In this approach, each coil collects undersampled data in parallel resulting in an aliased image for each coil. Each coil detects a signal from the same image, but due to the different spatial locations of the coils they will exhibit different spatial sensitivities to the signal of the image. By exploiting the differing coil sensitivities, the multiple aliased images, and using advanced reconstruction algorithms (i.e. more than just a Fourier transform), it is possible to recover an unaliased image. These approaches can take place in the image domain [18] or in the Fourier-space domain [19]. In theory, PI can achieve an undersampling factor as high as the number of coils used in the scan [18], but in practice undersampling rarely goes past a factor of two.

An approach that utilizes the inherent redundancy of MRI data is *compressed sensing* (CS) [20]. CS MRI exploits the sparsity exhibited in MR images to recover high quality images from an undersampled set of data. Here, *sparsity* means that in some transform domain, the majority of the signal will be contained in a relatively small number of components, meaning that in this transform domain a few components will have a high intensity and the majority will be zero or close to zero (Figure 1.2). CS requires pseudorandom sampling of the raw MR data (see Section 2.1.3). This type of sampling results in noise-like aliasing artifacts from which the sparse components can be separated

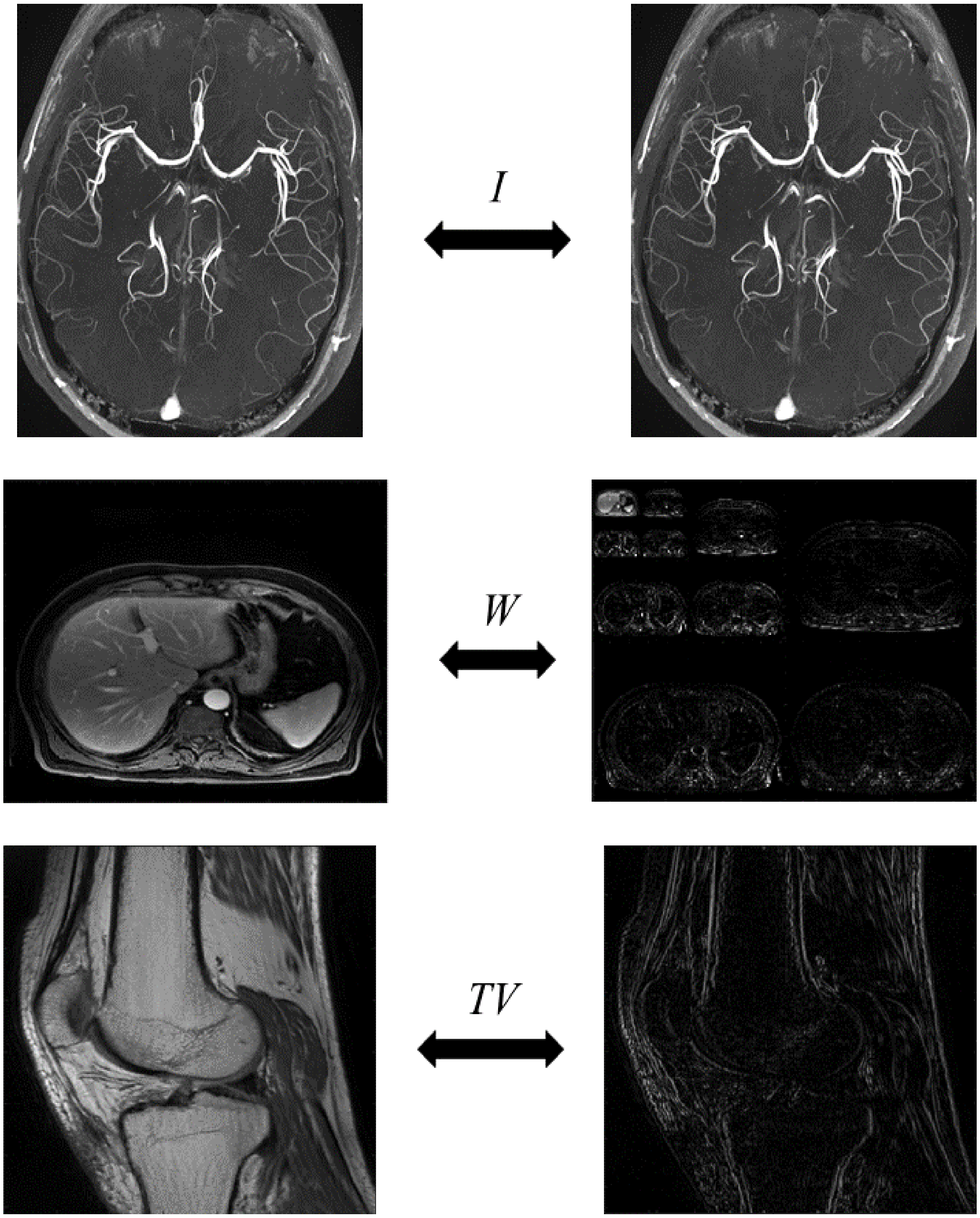


Figure 1.2: Examples of sparsity in different domains. (Top) Sparsity in the image domain (identity transform, *I*). (Middle) Sparsity in the wavelet domain (wavelet transform, *W*). (Bottom) Sparsity in the total variation domain (total variation transform, *TV*).

to recover an un-aliased image. CS can achieve undersampling factors up to an order of about 10 in a research setting [21], however undersampling factors would likely be much lower in a clinical setting. PI and CS can also be used simultaneously to achieve even faster imaging speeds [22].

In recent years, increased computing power and improved algorithms have allowed for the adoption of advanced machine learning based reconstruction algorithms [23]. For MRI, it is common to think of machine learning on the diagnostic side of the imaging workflow (e.g. detecting a tumor), but machine learning can also be a powerful tool for image reconstruction. Some machine learning approaches learn patterns or features common in MR images from a previous data set and use this information to prospectively reconstruct images. In a sense, this is similar to how CS works as CS leverages the prior knowledge that the reconstructed image is sparse in some transform domain. Since the machine learning approaches are less restricted in what prior information it uses, there is possibility for further acceleration of the MR scan [25-28]. Common approaches for development of machine learning techniques typically involve learning to de-alias an undersampled image by training a neural network on a set of fully sampled and retrospectively undersampled data [29-31]. Machine learning has even been used to estimate PK parameters maps directly from k-space [30]. One of the advantages of machine learning approaches is the short image reconstruction time compared to iterative approaches like CS. It is possible to reconstruct images of a similar or better quality than conventional approaches but within a fraction of the reconstruction time [31].

All the accelerated imaging techniques add much value to MRI, especially for dynamic imaging techniques such as DCE. However, the goal of MR research is not just to have fast imaging techniques. The techniques must also maintain the high quality required for diagnosis. Therefore, it is imperative that image quality be closely monitored as these techniques develop.

1.4 Image Quality in MRI

The purpose of an MR image is to convey diagnostic information to a doctor, typically a radiologist. In many cases, diagnostic information means the visual presence or absence of an abnormality such as a lesion or hemorrhage. The quality of the MR image is then equivalent to the ability of the image to convey this information. For example, if the image is very noisy, a radiologist would not be able to say whether a lesion is present or absent, so the image is of low quality. Over the course of their extensive training and practice, radiologists learn what features of the image will best allow it to be used for diagnosis. These are features such as signal to noise ratio (SNR), sharpness, or imaging artifacts, for example. This knowledge can cause radiologists to have differing opinions of image quality compared to a non-expert [32] who may consider image quality only in the abstract, as they would consider the quality of a natural image, for example. With this in consideration there are numerous ways to quantify image quality. Since MR images are acquired for diagnosis the gold-standard would be a task-based measure such as a visibility rating of a lesion/anatomical feature or diagnostic accuracy when a particular degradation is applied. This approach can be difficult to implement in practice due to variation in lesion visibility, so radiologists rating of overall image quality is widely accepted as a surrogate.

However, this too can be difficult to implement in practice on a large-scale due to time limitations or concerns of inter/intra-rater reliability issues. The subjective nature of scoring by radiologists also makes it difficult to quantitatively compare results between studies. This encourages the use of alternative techniques, such as objective *image quality metrics* (IQMs), that can provide a consistent measure of image quality [33] in a rapid and automatic manner. Objective IQMs can broadly be broken into three categories based on the availability of a gold-standard reference image [34]. The three categories are:

1. Full-reference IQMs: a reference is known and the score for the IQM is calculated from the similarity between the test image and the reference image.
2. Reduced-reference IQMs: a reference is partially known and a score is calculated from what information is known and the test image.
3. No-reference IQMs: no reference is known, so the score is calculated solely from the test image.

Many MR imaging studies adopt the framework of retrospectively undersampling or adding artifacts to a known image and attempt to recover a high-quality image. Since a reference image is known, full-reference IQMs are commonly used and so will be the focus of this thesis. Common full-reference IQMs include the root mean squared error (RMSE), peak signal to noise ratio (PSNR), and the structural similarity index (SSIM) [34].

There have been some recent advances with the use of machine learning techniques for the assessment of image quality. Most of these approaches are designed with a binary output either for artifact detection [35,36] or to classify an image as diagnostic/non-diagnostic [37,38]. These approaches could be useful for on-the-fly assessment of scan quality during a clinical scan as they could catch non-diagnostic scans before the patient leaves thus reducing patient recall. For assessing new imaging techniques, a scale with a varying degree of scores is more useful as it could measure superior diagnostic quality for a new imaging technique compared to an existing technique. A study by Kustner *et al*. used a deep neural network to achieve a 93.7% accuracy for scoring image quality on a 1-5 scale as determined by a radiologist [38]. Despite this progress there remains no widespread consensus for using a particular machine learning based technique for assessing image quality. For this, IQMs continue to play a large role, thus motivating the need for their continued study.

When considering these non-radiologist methods to measure image quality it is important to remember how image quality was originally defined in this thesis. Image quality is defined by a radiologist’s opinion in relation to how well the image can be used for diagnosis. Most objective IQMs were calculated to correlate with the opinions of non-experts on natural images so there is no guarantee of their performance for medical images in a clinical setting. This highlights an important gap of knowledge in the literature.

1.5 Contributions of this Thesis

This thesis examines how to quantify MR image quality and explores how these quantitative measures can be used in MRI techniques, specifically DCE MRI. The results focus on MR image reconstruction. This thesis is composed of two independent studies that are broken into their own chapters. The first study examines the problem of determining which objective IQMs have the highest correlation with radiologists’ opinions of image quality. This is an important problem to study because IQMs such as RMSE and SSIM are commonly used to measure image quality in MRI studies, however they have never been validated as an accurate measure of MR image quality. In this study, the correlation between these IQMs and radiologists' rating of image quality is calculated along with other common IQMs. We hypothesize that RMSE and SSIM will not demonstrate the highest correlations with the radiologists’ rating of image quality. This is motivated by the more sophisticated algorithms of other IQMs and the results from natural-image studies. The results of this thesis may serve as a valuable reference when choosing an IQM in future MR image acquisition and reconstruction studies.

The second study applies IQMs to DCE MRI reconstruction with the goal of accurate PK parameter measurement. We hypothesize IQMs can be used to determine a balance between temporal resolution and image quality in DCE MRI to maximize PK parameter measurement accuracy. This is based on the results of previous studies that found a correlation between IQM scores and accuracy of model parameters in DCE MRI simulations [39] and further development in golden angle based sampling techniques [17]. In this study, a technique for PK modelling based on the IQM score of the reconstructed images is presented. In a clinical setting, this technique would allow for patient-specific optimization of temporal resolution in DCE MRI.

The thesis is outlined as follows. The theory and technical background of methods used throughout the two studies are presented in Chapter 2. The full details of the two studies are presented in Chapters 3 and 4 respectively. These chapters are relatively free-standing, and each contain their own Motivation, Methods, Results, Discussions, and Conclusions sections. Finally, Chapter 5 provides a summary and conclusions of the thesis.

Chapter 2: Theory and Techniques

2.1 MRI Data Acquisition

2.1.1 MRI Physics

The source of the MRI signal is the proton found in the nucleus of hydrogen atoms, which are predominantly found in water and fat in the body. External magnetic fields are used to manipulate the states of these protons to encode spatial information in the resulting signal that gives rise to the image. Fundamentally this is a quantum mechanical phenomenon, but given the large number of hydrogen atoms in a human body we may adapt a classical interpretation. Protons are a spin-1/2 particle and so possess a magnetic moment $\vec{\mu}$. Under normal conditions the spins of the protons in the body are randomly aligned and the net magnetization is zero. However, when placed in a strong magnetic field, such as an MRI machine, the spins will align resulting in a non-zero net magnetization $\vec{M}$ that is parallel to the applied magnetic field:

$$\vec{M} = \frac{1}{V}\sum_{i} \vec{\mu}_i\,, \tag{2.1}$$

where V is the volume containing the i magnetic moments.

It is conventional to define the z-axis as parallel to the main magnetic field $\vec{B}_0$, such that $\vec{B}_0 = B_0\hat{z}$. The initial magnetization at a point $\vec{r}$ is then $\vec{M}(\vec{r}, t = 0) = M_0(\vec{r})\hat{z}$, where $M_0(\vec{r})$ is the magnitude of the net magnetization at this location. This magnitude depends on factors such as the local proton density, temperature, and magnetic field strength. An arbitrary external magnetic field, $\vec{B}$, will induce a torque on the magnetization. This provides the 3D equations of motion for $\vec{M}$ known as the Bloch equations:

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \times \vec{B}, \tag{2.2}$$

where γ is the gyromagnetic ratio (42.6 MHz/T for a proton).

During the scan, the net magnetization along the z-axis is much less than the $\vec{B}_0$ field and so is impossible to detect. For a signal to be detected, $\vec{M}$ must be tipped into the transverse x, y-plane where the transverse components can be detected. This is accomplished by a short radiofrequency (RF) pulse along the x or y directions that induces a torque that rotates $\vec{M}$ into the transverse plane. This is also referred to as *excitation* due to the equivalent phenomenon in the quantum mechanical interpretation. After the RF pulse, all magnetic fields are along the $\hat{z}$ direction. Due to the symmetry in the x, y-plane, it is convenient to combine the x and y components of $\vec{M}$, M_x and M_y, into a single transverse component:

$$M_{xy} = M_x + iM_y, \tag{2.3}$$

After $\vec{M}$ is disturbed from equilibrium it will begin to precess about the applied magnetic field, similar to a gyroscope. While this occurs, thermal relaxation will also affect $\vec{M}$. The z-component of $\vec{M}$ will relax to its equilibrium value M_0 according to a relaxation time constant T_1. The transverse components will decay independently due to loss of phase coherence of the underlying magnetic moments due to local field fluctuations. This decay occurs with a relaxation time constant T_2. Including these effects, equation 2.2 becomes:

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \times \vec{B} + \frac{M_0 - M_z}{T_1}\hat{z} - \frac{\left(M_x\hat{x} + M_y\hat{y}\right)}{T_2}, \tag{2.4}$$

where $\vec{B} = B(\vec{r})\hat{z}$.

Without loss of generality we can define the initial conditions after RF pulse excitation as:

$$M_z(r, 0) = M_0(\vec{r})\cos(\alpha), \tag{2.5}$$

$$M_{xy}(\vec{r}, 0) = M_0(\vec{r})\sin(\alpha), \tag{2.6}$$

where α is how far from the z-axis $\vec{M}$ is tipped, known as the flip angle. The coupled first order different in equations of 2.3 can then be solved to find the equations of motion for $\vec{M}$:

$$M_z(\vec{r}, t) = M_0(\vec{r})\left(1 - \exp\left(-\frac{t}{T_1(\vec{r})}\right)\right) + M_z(\vec{r}, 0)\exp\left(-\frac{t}{T_1(\vec{r})}\right), \tag{2.7}$$

$$M_{xy}(\vec{r}, t) = M_{xy}(\vec{r}, 0)\exp\left(\frac{-t}{T_2(\vec{r})}\right)\exp(-i\omega(\vec{r})t), \tag{2.8}$$

where $\omega(\vec{r}) = \gamma B(\vec{r})$ is the *Larmor frequency* of the magnetic moments in the $\vec{B}$-field. This relationship between magnetic field strength and Larmor frequency is foundational to MR imaging. When a spin is placed within the main $\vec{B}_0$-field, its Larmor frequency is typically denoted as: $\omega_0 = \gamma B_0$.

The main $\vec{B}_0$ field can be modified by adding additional linear gradient fields, $\vec{G}(\vec{r})$. The gradient can lie along any direction, but the field is always along the z-axis. For example, a linear gradient along the x-axis would result in the field:

$$\vec{B}(x) = (B_0 + G_x x)\hat{z}, \tag{2.9}$$

where G_x is the strength of the gradient in units of mT/m. The resulting local precessional frequency would be (returning to 3D):

$$\omega = \gamma\left(B_0 + \vec{r}\ \cdot \vec{G}\right) = \omega_0 + \gamma\vec{r} \cdot \vec{G}. \tag{2.10}$$

At this point it is convenient to adopt a rotating frame coordinate system. In this system the transverse plane rotates about the z-axis at ω_0. The result of this is that all spins rotating at this frequency will appear stationary and the gradient field only will be what determines the precessional frequency of the spin:

$$\omega = \gamma\vec{r} \cdot \vec{G} \tag{2.11}$$

The effect of a spatially varying gradient field is that phase ϕ will accrue in spins based on the local gradient field strength:

$$\phi(r,t) = \int_0^t \omega(t')dt' = \gamma\int_0^t \vec{r} \cdot \vec{G}(t')dt' = \gamma t\vec{r} \cdot \vec{G}, \tag{2.12}$$

for a time independent gradient $\vec{G}$.

The accumulation of phase is crucial to the acquisition of the MRI signal. By carefully controlling the amount of phase with the gradient field strength and duration, we can encode spatial variations in the local magnetization that are either in or out of phase with the spatial distribution of spins in the patient being scanned. This will allow us to detect a signal that will provide spatial information about the underlying anatomy.

2.1.2 Signal Detection and K-space

As discussed above, the signal is detected from the transverse magnetization, M_{xy}. Specifically, as seen in equation 2.8, the transverse magnetization rotates in the transverse plane about the z-axis. This results in a changing magnetic field which induces current in the nearby receiver coils according to Faraday's Law. Since the resulting current is proportional to the derivative of the magnetic flux in the coil, the signal detected from location $\vec{r}$ will be proportional to the magnitude of the transverse magnetization and the Larmor frequency:

$$s_{\vec{r}}(t) \propto \omega_0 \left| M_{xy}(\vec{r}, t) \right|. \tag{2.13}$$

where the subscript $\vec{r}$ indicates that s originates from point $\vec{r}$.

The total signal is the vector sum of contribution of the signal from the transverse magnetization at all points in the volume being imaged. Combining equations 2.8 and 2.13 and integrating, we find the following 3D signal equation:

$$s(t) = \eta\omega_0 \int M_{xy}(\vec{r}, 0) \mathrm{e}^{-\frac{\mathrm{t}}{\mathrm{T}_2(\vec{r})}} \mathrm{e}^{-i\phi(\vec{r},t)} d\vec{r}\,, \tag{2.14}$$

where η is the constant of proportionality that incorporates information such as coil sensitivity and electronic gain. This equation can be simplified by first ignoring the effects of T_2 relaxation. This is true if the sampling time $t \ll T_2$. Second, the terms η, ω_0, and $M_{xy}(\vec{r}, 0)$ are combined into an effective proton density term, $\rho(\vec{r})$. This gives the equation:

$$s(t) = \int \rho(\vec{r}) \mathrm{e}^{-i\phi(\vec{r},t)} d\vec{r}. \tag{2.15}$$

We can further simplify this equation by defining:

$$\vec{k}(t) = \frac{\gamma}{2\pi} \int_0^t \vec{G}(t') dt'\,, \tag{2.16}$$

such that equation 2.12 becomes:

$$\phi(\vec{r},t) = 2\pi\vec{r}\cdot\vec{k}(t). \tag{2.17}$$

Finally, we can then write equation 2.15 as:

$$s(\vec{k}) = \int \rho(\vec{r})\mathrm{e}^{-2\pi i\vec{r}\cdot\vec{k}}d\vec{r}. \tag{2.18}$$

This is the imaging equation for MRI and it demonstrates that the acquired signal $s(\vec{k})$ is the Fourier transform of the effective proton density $\rho(\vec{r})$, which is the target image. So, in order to recover the image $\rho(\vec{r})$, one must simply collect $s(\vec{k})$ and then take the inverse Fourier transformation:

$$\rho(\vec{r}) = \int s(\vec{k})\mathrm{e}^{2\pi i\vec{r}\cdot\vec{k}}d\vec{k}. \tag{2.19}$$

The space containing all the vectors $\vec{k}$ is called k-space. The vector $\vec{k}$ is the *spatial frequency vector* and has units of inverse length. This means $s(\vec{k})$ is a measure of the "amount" of the spatial frequency vector $\vec{k}$ in the image $\rho(\vec{r})$. From equation 2.16, we can see that the location of the point in k-space that we are sampling at time t is controlled by the area under the gradient curve. For a constant gradient magnitude, this is thus the product of the gradient magnitude and the time for which it is applied. The spatial frequency is encoded into the signal by the phase that is accumulated from the gradient. This can be visualized by the example shown in Figure 2.1. When the spatial frequency of $\vec{k}$ constitutes a significant component of $\rho(\vec{r})$ the signal $s(\vec{k})$ will be large. Otherwise, the signal will be smaller. The signal near the centre of k-space will contain information about the low spatial

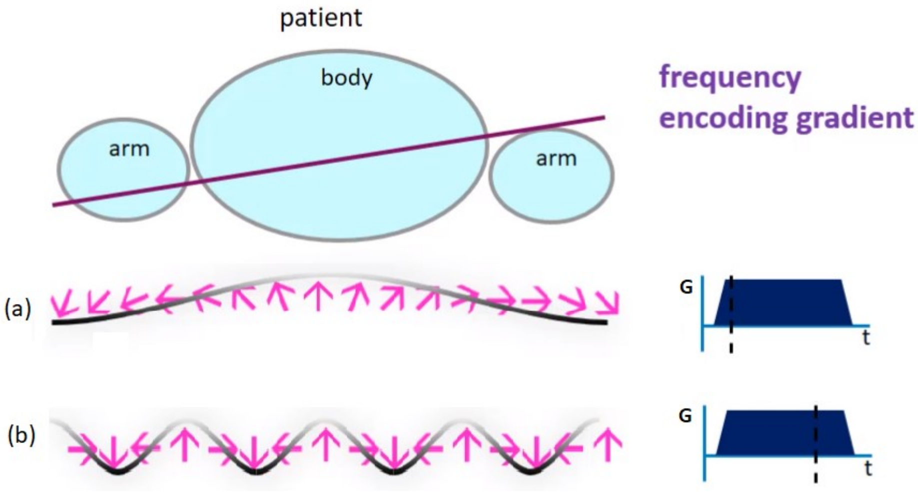


Figure 2.1: Visual representation of how gradient area encodes spatial frequencies into the MRI signal. A linear gradient of magnitude G is applied laterally across the patient. This causes the spins of the protons (shown in pink arrows) to accumulate phase depending on the strength of the gradient field at their location and the duration of the applied gradient. (a) The gradient is applied for a short period of time. This causes little phase to accumulate in the spins and so corresponds to a low spatial frequency. (b) The gradient field is left on for longer allowing more phase to accumulate in the spins. This allows for higher spatial frequency components to be sampled. Figure adopted from David Higgins educational session at ISMRM 2018.

frequencies contained in the image. This provides details about the overall image contrast and larger features in the image. The periphery of k-space contains the higher spatial frequencies in the image, which provides information about edges and fine details. As can be seen in Figure 2.2, much of the energy in k-space is concentrated near the center in the low spatial frequencies. Thus, when sampling k-space, it is important that the central components be prioritized over the components in the periphery.

So far, $\vec{r}$ and $\vec{k}$ have been treated as continuous variables. However, in practice an MRI image is a discrete array of pixels and k-space can only be sampled in non-zero intervals. Thus, k-space is also a discrete array. So, the collection of MRI data is equivalent to traversing through k-space by controlling the area under the gradient curve (Figure 2.2). To represent the discrete Fourier transform between the image and k-space, let us assume a 3D image of dimensions $N_x \times N_y \times N_z$. with resolution $\Delta x \times \Delta y \times \Delta z.$ Let us assume

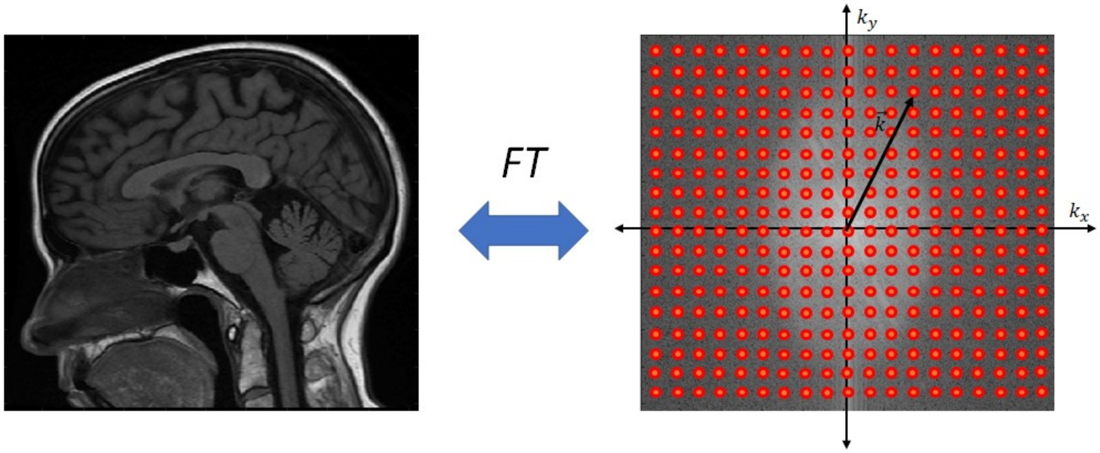


Figure 2.2: Example demonstrating the relationship between an image (left) and its k-space (right). The dots on the k-space are to emphasize the fact that it is sampled at discrete locations. To sample at the location shown by vector $\vec{k}$, a gradient $\vec{G}(t)$ would need to be applied for a time t such that $\vec{k} = \frac{\gamma}{2\pi}\int_0^t \vec{G}(t')dt'$.

fixed Cartesian sampling in k-space such that the resolution is $\Delta k_x \times \Delta k_y \times \Delta k_z$. The resulting discrete Fourier transform pairs are:

$$\rho(x,y,z) = \frac{1}{N_x N_y N_z}\sum_{n_x=0}^{N_x-1}\sum_{n_y=0}^{N_y-1}\sum_{n_z=0}^{N_z-1} s\left(n_x\Delta k_x, n_y\Delta k_y, n_z\Delta k_z\right)\mathrm{e}^{2\pi i\left(n_x\Delta k_x x + n_y\Delta k_y y + n_z\Delta k_z z\right)}, \qquad (2.20)$$

$$s\left(k_x,k_y,k_z\right) = \sum_{n_x=0}^{N_x-1}\sum_{n_y=0}^{N_y-1}\sum_{n_z=0}^{N_z-1} \rho\left(n_x\Delta x, n_y\Delta y, n_z\Delta z\right)\mathrm{e}^{-2\pi i\left(n_x k_x\Delta x + n_y k_y\Delta y + n_z k_z\Delta z\right)}. \qquad (2.21)$$

The factor out front of equation 2.20 is required to preserve the total energy in the system.

2.1.3 K-Space Sampling Patterns

The choice of k-space trajectory is an important decision to make prior to an MRI scan. This trajectory is commonly called the *sampling pattern*. The choice of sampling pattern is determined by the application of the scan and the type of reconstruction to be used. As shown in Figure 2.1, while a constant gradient is applied the spatial frequency being sampled along this direction will continuously increase. This allows for rapid sampling of the k-space vectors along this direction. This gradient is referred to as the *frequency*

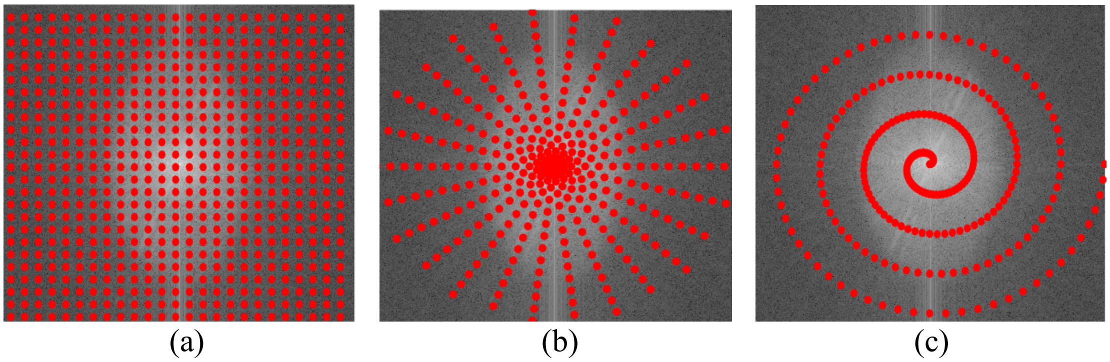

(a) (b) (c)

Figure 2.3: Examples of common 2-dimensional sampling patterns. (a) Cartesian with uniform sampling density, (b) Radial, and (c) Spiral.

encoding/read out gradient. If one wishes to collect data in 2D (as is required to generate a conventional MR image), phase must be pre-encoded in a perpendicular direction before the frequency encoding gradient is applied. This second gradient, which is applied for a fixed time and magnitude prior to readout, is called the *phase encoding gradient*. This combination of phase and frequency encoding gradients allows one to sample a full 2D plane in k-space. If one wished to sample in 3D, a second phase encoding gradient is usually required. Broadly speaking, 2D sampling patterns can be either Cartesian or non-Cartesian. As the name suggests, Cartesian sampling samples k-space points that lie on the Cartesian x, y grids, usually at fixed intervals, but not necessarily (Figure 2.3a). Non-Cartesian sampling patterns encompass all other sampling patterns. Common non-Cartesian sampling patterns include radial and spiral sampling patterns (Figure 2.3 b and c). These sampling patterns offer some advantages over Cartesian sampling. For example, since a radial sampling pattern samples the centre of k-space where the low spatial frequencies are located during each frequency encoding line, it is typically more rigorous to motion artifacts. Spiral allows for rapid coverage of k-space since the whole pattern is a single frequency encoding line. However, these non-Cartesian sampling patterns typically require more advanced reconstruction algorithms and precise gradient fields. Three-dimensional sampling allows for additional flexibility in sampling patterns. It is common practice to perform phase encoding along two dimensions and then frequency encoding along the third. Common three dimensional sampling patterns include Poisson disc, DISCO [40], or stack-of-stars. Stack-of-stars sampling is an example of a 3D sampling schema that requires only a single phase encoding gradient.

2.1.4 Golden Angle Radial Sampling with Sliding Window Reconstruction

Radial sampling is achieved by repeatedly rotating a line (called a *ray*) that passes through the centre of k-space while incrementing the orientation of the ray by a fixed angle $\Delta\varphi$ each time. For many applications it is advantageous to set $\Delta\varphi$ to be the golden angle (137.5 degrees/2.40 radians). Under this condition, each successive ray will be guaranteed to bisect the largest gap in k-space. To demonstrate this, Figure 2.4 shows the acquisition of the first 9 rays under two scenarios: $\Delta\varphi = \pi/9$ and $\Delta\varphi = 2.40$ radians. For the non-golden angle based sampling pattern, the gaps in k-space coverage are non-uniform until the sampling is complete, whereas golden angle sampling ensures consistent sampling uniformity regardless of the number of rays used [41]. Note that k-space sampling uniformity is not the same as k-space coverage. The number of rays collected determines

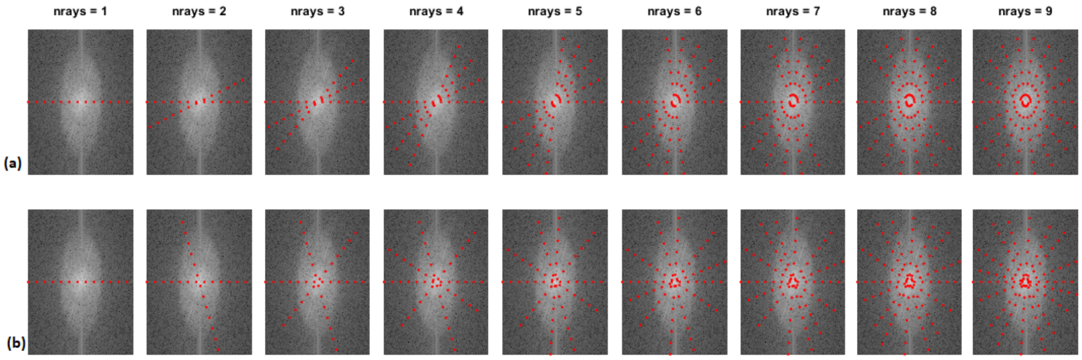


Figure 2.4 Examples of data acquisition for radial sampling patterns for (a) $\Delta\varphi = \pi/9$ and (b) $\Delta\varphi = 2.40$ radians (the golden angle).

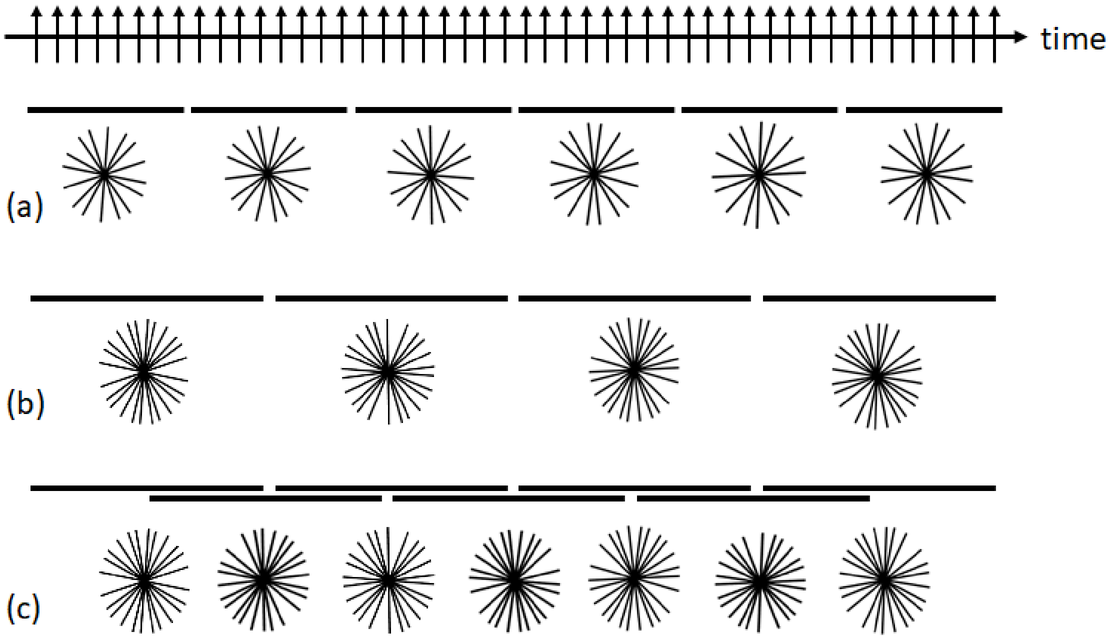


Figure 2.5: Visualization of how rays sampled with golden angle separation can be grouped during sliding window reconstruction. The vertical arrows at the top of the figure represent the sequential collection of rays through time. (a) Rays are grouped in eight rays per phase. (b) Rays are grouped in 12 rays per phase. (c) Rays are grouped in 12 rays per phase with 50% overlap.

the amount of k-space coverage and $\Delta\varphi$ determines the uniformity of sampling. A 3D extension of this imaging scheme is called *stack-of-stars*, which samples a ray in the k_x, k_y-plane at a fixed angle for a specified range of k_z before rotating the angle of the ray and repeating.

Since the angle of the first ray is arbitrary, when performing golden angle sampling any subset of rays may be grouped together and uniform coverage of k-space is still guaranteed. This allows one to use a *sliding window reconstruction* when reconstructing images of a dynamic data set [41]. This technique allows for images to be reconstructed with arbitrary window length and temporal position (Figure 2.5). When imaging a dynamic process, as in DCE MRI, the window length determines the *temporal footprint* of the reconstructed images. The temporal footprint is the period of time during which an image "occurs". All temporal dynamics that occur in this time period are averaged. This is a similar concept to temporal resolution, which is the smallest period of time over which a time-dependent behavior can be resolved. However, due to the ambiguity in measuring temporal resolution, especially with a sliding window reconstruction where the temporal footprint and frame rate can be de-coupled, temporal footprint is used to define the measure of temporal fidelity of a scan throughout this thesis. In a DCE scan, the resulting image from a group of rays is called a *frame*, similar to a frame in a movie. *Frame rate* refers to the number of images generated in the DCE scan per unit time. Since the temporal footprint is equivalent to the number of rays combined in an image, it also specifies the amount of k-space coverage and thus image quality. Therefore, there is an inverse relationship between image quality and temporal footprint.

The power of the sliding window reconstruction is that the temporal footprint does not need to be specified before the scan since an arbitrary number of rays can be combined and uniform k-space coverage is still guaranteed due to the golden angle sampling pattern. Further, with sliding window reconstruction, it is also possible to have *overlap* of the windows. This is controlled by the percent of overlap between adjacent frames. Overlap provides additional frame rate without changing the temporal footprint and thus the image quality (Figure 2.5c). For example, using a 50% overlap of two would double the frame rate, while the temporal footprint would remain the same.

2.2 Image Reconstruction Techniques

When considering image reconstruction techniques, it is often convenient to consider the image as a one-dimensional vector instead of a multi-dimensional array. Let us then define the image as $\vec{x} \in \mathbb{R}^N$, where $N = N_x \times N_y \times N_z$. The corresponding fully sampled k-space would then be $\vec{y} \in \mathbb{C}^N$, such that $y = \mathrm{FT}\{x\}$. If we were then to sample N components of k-space (while satisfying the Nyquist criteria) the process of image reconstruction would be trivial as the image could be recovered by simply Fourier transforming the k-space data, a linear transformation. However, it is often impractical to fully sample k-space due to reasons such as high temporal resolution requirements, concerns of motion artifacts, or increasing patient throughput. Thus, we adopt a technique called undersampling where only a subset of points in k-space are collected. The degree of undersampling is controlled by the undersampling factor, which is defined as the ratio of points in the fully sampled k-space array to the number of points actually sampled. The undersampling factor can also be called the *acceleration factor* because an undersampling factor of R will reduce the sampling requirements by a factor of R, thus accelerating the scan time by the same factor.

Undersampling allows for accelerated scans, however if the sampling density is too low (i.e. below the Nyquist criterion of twice the highest frequency in the image) aliasing artifacts that will degrade the quality of the image will occur if one tries to reconstruct the image with a simple Fourier transform. Aliasing artifacts occur when signals from higher spatial frequencies are erroneously attributed to lower spatial frequencies. The degree of aliasing artifacts is controlled by the undersampling factor, but the characteristics of the aliasing artifacts are dependent on the choice of sampling pattern. Sampling patterns can be considered to be either coherent or incoherent. Coherent sampling patterns have equispaced k-space undersampling and will lead to coherent artifacts (e.g. phase wrap around) whereas incoherent random undersampling will lead to incoherent artifacts that appear noise-like (Figure 2.6). Due to the non-Cartesian design of radial sampling, the

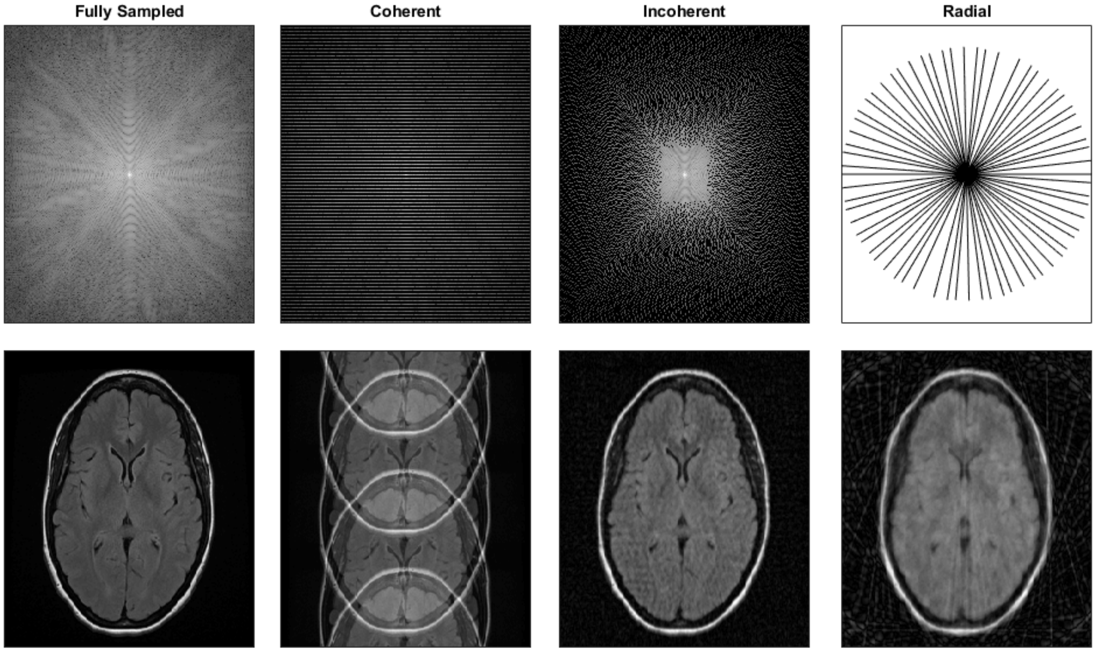


Figure 2.6: Examples of how different sampling patterns affect images reconstructed with Fourier transform. The top-row shows the zero-filled k-space for a retrospectively undersampled data set with the corresponding image reconstructed with a Fourier transform on the bottom row. Coherent undersampling was done by skipping two of every three lines along the vertical direction. Incoherent sampling was performed with the CIRCUS [71] sampling pattern. Since radial is non-Cartesian the sampling pattern is shown instead of the k-space.

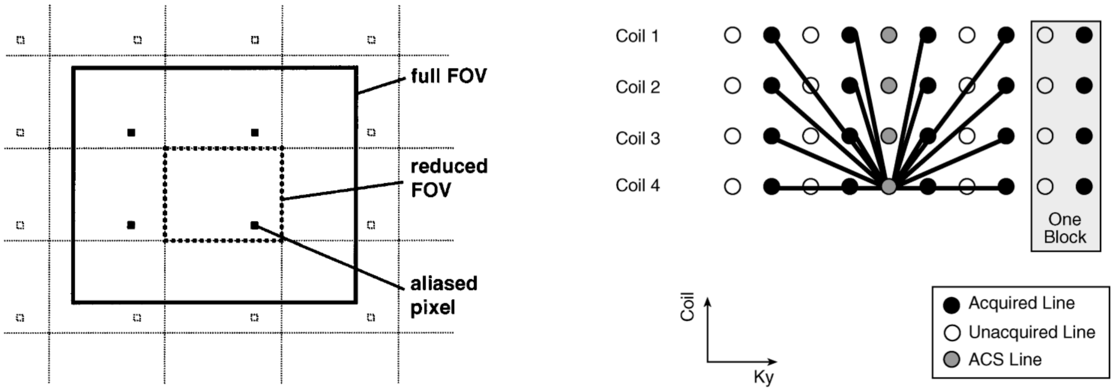
streaking artifacts characteristic to this sampling pattern will also appear to a large degree incoherent. In order to reduce the presence of aliasing artifacts from undersampled k-space data, advanced non-linear reconstruction techniques are required.

2.2.1 Parallel Imaging

So far, it has been assumed that the signal is detected by a single receiver coil. However, multiple coils can be used and they each detect a signal in parallel. Since the coils are located in different locations in space, they will have different spatial sensitivities to regions in the body i.e. a coil will be more sensitive to a signal that originates near it than a signal that originates from further away. These differences in coil sensitivities provide additional spatial information within the MR signal. This additional information from the coil sensitivities profile can compensate for collecting fewer phase encoding lines, which would result in an acceleration of scan time. The multiple receiver coils can also compensate for the loss of SNR that results from undersampling.

Parallel imaging works by collecting undersampled k-space data for each coil and then using the information inherent in the spatial sensitivities of each coil to remove the aliasing artifacts. PI works best with coherent artifacts where phase encoding lines of k-space are systemically skipped. As discussed in the Introduction, the problem of removing these artifacts can be broken into two categories: image domain approaches (e.g. SENSE) where the phase wrapping artifacts are unfolded or frequency domain approaches (e.g. GRAPPA) where the missing phase encoding lines are filled in before the Fourier transform is applied.

SENSE-based parallel imaging techniques are performed in the image domain after the aliased images are reconstructed from the undersampled data. The first step in this process is to generate the coil sensitivity maps. This is done by collecting a set of fully sampled low resolution images prior to the collection of the accelerated data. Then, the k-space data is acquired at full resolution with a specified undersampling factor along the phase encoding direction. Due to the low sampling density in this direction wrap around artifacts will occur where the signal of one voxel is erroneously mapped to another voxel as seen in Figure 2.7a. The process of unwrapping the aliased image requires the generation of an unfolding matrix U that can only be calculated when the number of coils is greater than the number of voxels superimposed by aliasing. This puts a theoretical limit on the undersampling factor at the number of coils used during imaging. For full details on how to calculate U, please refer to the original SENSE paper [18].



(a) (b)

Figure 2.7: Visualizations of parallel imaging techniques. (a) Image-domain based reconstruction. Underdamping reduces the FOV, resulting in aliasing artifacts where pixels are mapped to incorrect locations. Reconstruction amounts to unfolding the reduced FOV. Figure from Pruessmann *et al*. [18]. (b) Frequency-domain based reconstruction. Missing k-space data is estimated as a weighted sum of neighbouring values. ACS stands for auto-calibration signal. Figure from Griswold *et al*. [19].

The data acquisition during GRAPPA-like techniques is similar to SENSE in that phase encoding lines are also skipped, which would result in phase wrapping artifacts. These techniques work by estimating the missing lines in k-space before the Fourier transform such that aliased images are never generated. During data acquisition, a fully sampled auto-calibration signal (ACS) is collected in the center of k-space. From the ACS and its neighbouring points, a set of weights is learned to interpolate k-space points from a linear combination of these neighbouring points (Figure 2.7b). The weights use data collected by all the coils. This set of weights is then used to fill in all the missing lines and unaliased images are then generated for each coil, which are then combined to form the final image by taking the sum of squares of the coil image. Again, for details on how to calculate the weights to find the missing k-space points please refer to the original GRAPPA paper [19].

2.2.2 Compressed Sensing

A compressible image/signal is one where it is possible to remove a significant portion of the information from an image and the image will look largely the same. This is accomplished by removing many unnecessary components of the image in some transform domain (e.g. in the wavelet domain for the JPEG2000 compression algorithm). This allows the digital image to be stored in a format requiring much less computer memory. MR images are compressible. This is possible because MR images exhibit transform sparsity i.e. they have a sparse representation in some transform domain. By definition, compressibility is similar to sparsity. Whereas sparsity means that most components in a signal are zero, compressible means that most components are of such a low intensity relative to a small number of relatively high intensity signals that they can be approximated as zero. With this in mind, I will refer to both compressible and sparse signals simply as sparse.

CS was motivated by the fact that, if an MR signal is sparse such that it can be represented by a fraction of the components of the signal in some transform domain, why collect all the other data during acquisition since this will prolong imaging time? CS reconstruction is akin to recovering the few high intensity components from highly undersampled k-space data. This is accomplished by incorporating the prior information about the sparsity of the image into the reconstruction algorithm. This theory was an important development in MRI as this technique allows one to undersample based on the redundancy of information in MRI data as opposed to the enforced redundancy such as that resulting from the use of multiple receive coils. This resulted in a paradigm shift in MRI research towards reconstruction of images from highly undersampled data.

There are three conditions that are required for successful CS reconstruction of MRI data:

1. The recovered image must be sparse in some transform domain.

2. Data must be sampled via a pseudorandom sampling trajectory.
3. A nonlinear reconstruction technique which can enforce data consistency and sparsity must be used.

Common examples of transform sparsity were shown in Figure 1.2. However, in this thesis only wavelet sparsity is explored. Wavelets are functions of zero net-area that represent oscillations of finite duration. Similar to a Fourier transform, which represents an image in a frequency basis, a wavelet transform represents an image in the wavelet basis. The requirement for transform sparsity in CS is related to the requirement for incoherent sampling. As was seen in Figure 2.6, incoherent sampling leads to incoherent and diffuse artifacts that appear noise-like. The artifacts will appear similarly noise-like in the transform domain. Since the sparse representation of the image has a low number of high intensity components, they can easily be separated from the noise-like artifacts. An example of this is shown in Figure 2.8. Here a sum of sinusoidal functions is presented. While not sparse in the time domain, sinusoidal functions are very sparse in the frequency domain. In this example, three scenarios are presented: (1) sampling above the Nyquist criteria, (2) uniformly sampling sub-Nyquist, and (3) non-uniformly sampling sub-Nyquist. In the first scenario, the sparse components of the signal are easily recoverable in the Fourier domain. In the second scenario, the low sampling density leads to coherent aliasing artifacts that are indistinguishable from the true signal in the Fourier domain. In the third scenario, the same number of samples are distributed non-uniformly across the time-domain signal, the aliasing artifacts are diffuse, noise-like, and of a much lower magnitude than the true frequency components. It would thus be possible to recover the true signal from this highly undersampled data with an appropriate reconstruction algorithm.

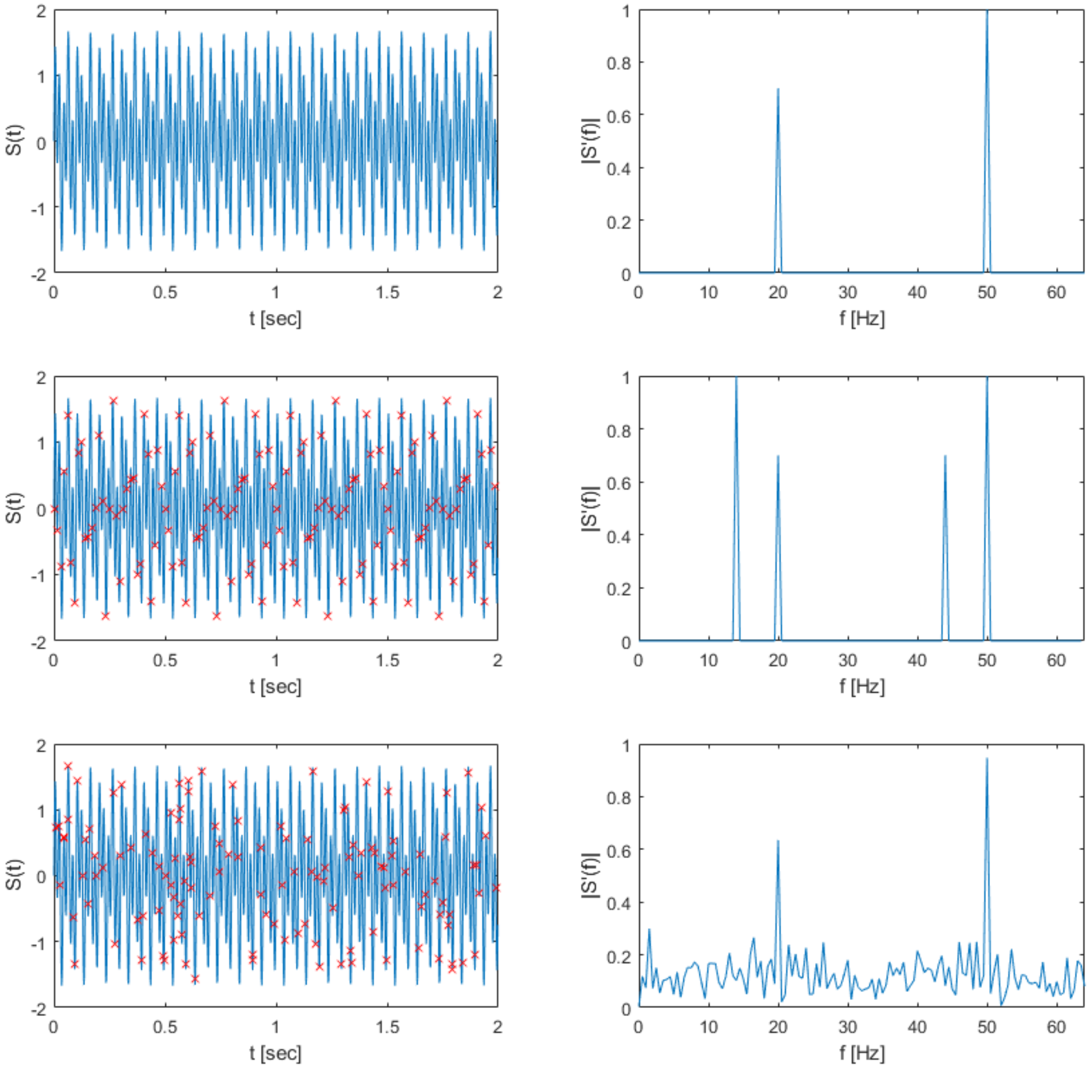


Figure 2.8: Example of different sampling rates and patterns of a signal composed of the sum of two sinusoids (20Hz and 50Hz). The signal in the time domain is shown in the column on the left and the Fourier transform is on the right. (Top) Sampling rate of 512Hz. The signal is perfectly represented in the Fourier domain. (Middle) Uniform sampling at 64Hz (128 samples total). The red markers indicate the sampling points on the original signal. Due to the low sampling rate, the true signal in the Fourier domain is indistinguishable from the aliasing artifacts. (Bottom) Non-uniform sampling of the same number of points (128). The non-uniform sampling leads to non-uniform artifacts in the Fourier domain that appear diffuse and noise like. The true components are easily distinguishable.

As shown in the Figure 2.8, a simple Fourier transform on the undersampled data will lead to an aliased image. Even with non-uniform undersampling, this will be a noisy image of low quality. CS reconstruction works by enforcing sparsity in the transform domain. The Fourier transform at the bottom of Figure 2.8 is not sparse even though we know that the true signal should be (See the Fourier transform at the top of Figure 2.8). By incorporating the prior knowledge that an MR image must be sparse in an appropriate transform domain into the reconstruction algorithm, we can suppress these aliasing artifacts. The reason incoherent sampling is required is because otherwise the aliasing artifacts would be indistinguishable from the true signal. However, it is not enough to just promote sparsity, as it is trivial to generate an arbitrarily sparse signal. Consistency between the raw data and the reconstructed data must also be ensured. Thus, CS reconstruction is a balancing act between data consistency and promoting sparsity. In symbols, the image $\vec{x}$ reconstructed from the undersampled k-space data $\vec{y}$ is the image that minimizes the following equation:

$$\underset{\vec{x}}{\operatorname{argmin}} \, ||A\vec{x} - \vec{y}||_2^2 + \lambda||\Psi\vec{x}||_1 \tag{2.22}$$

where A is the sensing matrix that incorporated information such as the Fourier transform, sampling pattern, and coil sensitivities, Ψ is the sparsifying transform, λ is the regularization weight that controls the trade-off between data consistency and sparsity, and $||\cdot||_1$ is the ℓ_1-norm. Image reconstruction is thus is an unconstrained convex optimization problem, which can be solved by iterative methods such as soft thresholding or non-linear conjugate gradients.

2.2.3 Berkeley Advanced Reconstruction Toolbox (BART)

The Berkeley Advanced Reconstruction Toolbox (BART) [42] is an open source image reconstruction library that is a valuable research tool for MR image reconstruction. This toolbox has the capabilities for multi-dimensional array manipulation, Fourier and wavelet analysis, and iterative reconstruction algorithms such as PI and CS. The toolbox is built in C but allows for MATLAB (MathWorks, MA, USA) interfacing. It also has the capabilities of handling non-Cartesian data, which is extremely valuable for dealing with the radial data in Chapter 4. The `pics` (parallel imaging and compressed sensing) command was used for all CS reconstruction in this thesis. This command solves equation 2.22 for an input k-space $\vec{y}$ with specified sparsifying transform and regularization weight. Coil sensitivity maps may also be provided to incorporate PI into the reconstruction, however all images in this thesis were reconstructed with just a single coil. The `nufft` (non-uniform fast Fourier transform) was also used to generate the stack-of-stars data in Chapter 4. BART version 3.01 was used for all cases.

2.3 T1 Relaxation and Contrast Agents

The k-space of an image is typically not all collected in a single continuous acquisition. This is infeasible because, as the data is being collected, the signal is decaying due to T_1 and T_2 relaxation. During a single scan, it is common to use multiple RF pulses to repeatedly excite the magnetization until all the data is collected. The time between RF pulses is called the repetition time or TR. If TR $\gg T_1$ in a region, the longitudinal magnetization will fully return to its equilibrium value, M_0. During the next RF pulse, all of M_0 can be tipped into the transverse plane to generate a signal. However, if TR $\approx T_1$ in a region, then the longitudinal magnetization will not be fully recovered before the next RF pulse. This means less magnetization is accessible to be tipped into the transverse plane, so the signal will be smaller. The amount of magnetization that is recovered between each TR will depend upon T_1 thus generating contrast between regions in the image based on variations of T_1.

T_1 relaxation, also known as *longitudinal relaxation*, is the time constant that defines how quickly longitudinal magnetization that is knocked into the transverse plane will return to its equilibrium value as described by equation 2.7. For a short TR, a tissue with a shorter T_1 value will recover more longitudinal magnetization between excitations, so it will appear brighter than a tissue with a longer T_1 value. However, if a long TR is used, then both tissues can fully recover their longitudinal magnetization so no additional contrast would be included in the image. Enhancing tissue contrast based on T_1 value by using short TR is called T_1 weighted imaging.

Another name for T_1 relaxation is *spin-lattice relaxation*, which provides more insight into the mechanics driving T_1 relaxation. To understand this mechanism, we need to adopt a quantum mechanical framework. A proton is a spin ½ particle, meaning it can be in one of two energy states: a high energy excited state and a low energy relaxed state. These energy states are separated by an energy gap $E = \hbar\omega = \hbar\gamma B_0$, where $\hbar$ is Planck's reduced constant. The relaxed energy state is when the magnetization is parallel to the external field. The excited state occurs when the RF pulse is applied and the magnetization gets knocked into the transverse plane. The misalignment between the spins and the magnetic field adds energy to the system. This is why we say RF pulse excitation. A spin in the high energy state will not spontaneously release energy to fall to the low energy state within any reasonable amount of time. There must be a mechanism that facilitates this transition. In clinical MRI, this mechanism is time varying dipole-dipole interactions between the proton and neighbouring molecules (called the *lattice*) due to thermal vibrations. This allows energy transfer between the excited proton and the lattice, which results in relaxation and recovery of the longitudinal magnetization. The degree of coupling between the spins and lattice depends on factors of the lattice such as the size and state of the constituent molecules. If the thermal vibrations of the neighbouring molecules occur at a similar frequency to the precessional frequency of the spin then there will be more efficient energy transfer between the spins and the lattice. Higher energy transfer leads to faster relaxation and thus a shorter T_1 time. These properties will vary between tissue leading to different T_1 times.

Tissue relaxation times can be affected by the injection of contrast agent materials. In MRI, contrast agents are typically gadolinium based. Gadolinium is a strongly paramagnetic metal. This facilitates energy transfer between the spins and the lattice, thus reducing the local T_1 relaxation times. This causes regions where the contrast agent is present to appear bright in T_1 weighted images.

2.4 Quantitative DCE MRI

2.4.1 DCE MRI

As discussed above, a contrast agent shortens the local T_1 relaxation time leading to local signal enhancement. DCE MRI is a dynamic imaging sequence that visualizes the movement of contrast agent as it travels through the circulatory system. This is achieved by generating a series of T_1 weighted images with a high temporal resolution (~10-15 sec). Each image in the series is called a frame. Historically, the k-space data for each frame is collected independently. Since the temporal dynamics of the contrast agent occurs on the order of tens of seconds, a high temporal resolution is required.

Spoiled gradient echo (SPGR) pulse sequences with a short TR (3-5msec) and a low flip angle (12°) are commonly used. This allows for rapid data acquisition, helping achieve high temporal resolution. Undersampling can also be used to further increase temporal resolution. DCE MRI is actually very well suited for CS reconstruction due to the redundancy of information in the temporal dimension since images only change slightly between frames [43]. So, while the data for each frame can be collected independently, reconstructing the full series at once can lead to improved image quality. View sharing, where sections of k-space are shared between frames, is also commonly used as with the DISCO sampling pattern [40]. Although not routinely used yet clinically, golden angle stack-of-stars sampling pattern with sliding window reconstruction has shown potential [17,44].

2.4.2 Pharmacokinetics

Pharmacokinetics (PK) is the study of motion of a substance after it enters a living organism. For DCE MRI, this refers to the behavior of the contrast agent. The contrast agent is injected in bolus form intravenously. The contrast agent can then travel systemically throughout the body in the blood vessels. The time-dependent contrast agent concentration in the blood vessel is an important measurement for performing PK modelling in a tissue. The contrast agent concentration time-course of an artery that provides blood for a tissue of interest is called the *arterial input function* (AIF).

The contrast agent can also leak out of the vessels into nearby tissues through the capillary walls. The rate at which this occurs plays an important role in determining the intensity of the signal in a tissue during a DCE scan. The dynamics of the contrast agent in a tissue is typically modelled with a two-compartment model: the vessel being one compartment and the *extravascular extracellular space* (EES) being the other. The EES is the space in the tissue between the cells that make up the tissue. The contrast agent diffuses into and out of the EES at a rate proportional to the difference in concentration between the two compartments. This constant of proportionality is called the *volume transfer constant* K^{trans}, and reflects tissue and vessel properties such as blood perfusion rate,

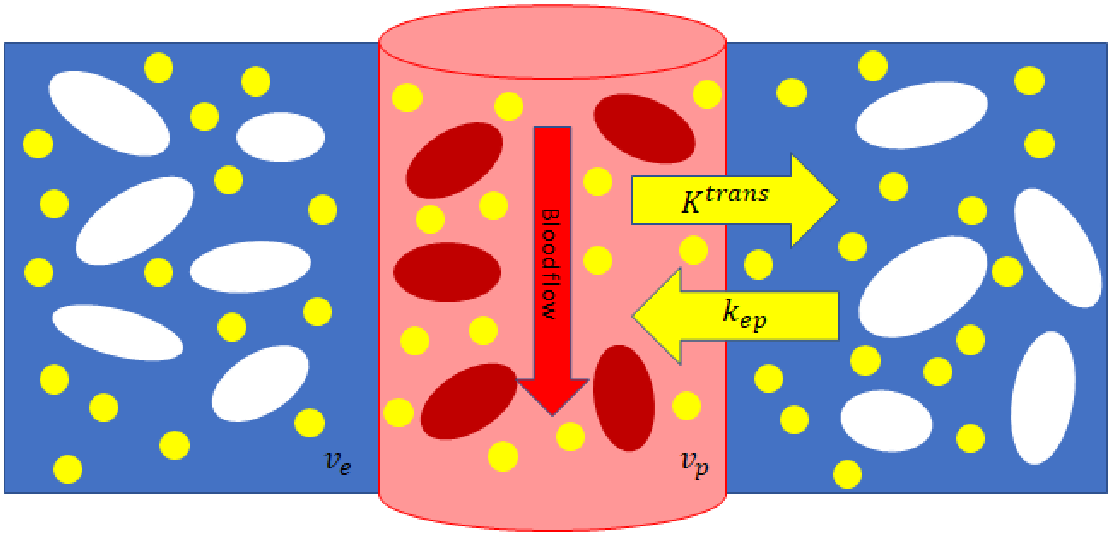


Figure 2.9: Visualization of PK parameters. In this diagram the blue background represents the extravascular extracellular space, the red cylinder is a blood vessel, the white ovals are tissue cells, the red ovals are blood cells and the yellow circles are the molecules of the contrast agent. The PK parameters K^{trans}, k_{ep}, v_e, and v_p are described in the text.

capillary permeability, and capillary surface area and has units of min^{-1}. Other important PK parameters included the *blood plasma volume per unit volume of tissue* v_p, the *EES volume per unit volume of tissue* $v_{e,}$ and the *efflux rate constant from EES to plasma* k_{ep}, with units of min^{-1} [45]. These parameters are shown visually in Figure 2.9. The PK parameters define the time course of the contrast agent concentration in the tissue $C_t(t)$.

Histopathological differences in perfusion and permeability are what makes DCE a valuable tool for diagnosis. During growth, a tumor can perform angiogenesis where it creates new blood vessels. These blood vessels can have higher permeability than healthy tissue resulting in elevated K^{trans} values. This results in rapid transfer of the contrast agent to the tissue and will manifest as early signal enhancement in the DCE scan.

2.4.3 Tofts Model

To determine the PK parameters in a tissue, C_t must be fit to a PK model such as the Tofts model [6], which will be the formalism used in this thesis. The driving force of uptake of the contrast agent into the tissue is the difference in concentration between the blood stream and the tissue. When considering this concentration gradient, we need to consider the concentration in blood plasma, $C_p(t) = C_b(t)/v_p$ where C_b is the concentration in whole blood, since the contrast agent does not enter blood cells. Similarly, in the tissue, the contrast agent is not known to enter the tissue cells, so the important concentration for contrast agent dynamics is the concentration in the EES, $C_e(t) = C_t(t)/v_e$.

If we then assume linear intercompartmental flow, isodirectional flow rate across the capillary, and neglect contributions of C_p to C_t (accurate for $v_p \ll 1$) we get the following differential equation:

$$\frac{dC_t}{dt} = K^{trans}\left(C_p - \frac{C_t}{v_e}\right). \qquad (2.23)$$

This shows in symbols the premise that the rate of contrast agent uptake into/out of the tissue is proportional to the difference in concentration between the plasma and the EES, and the constant of proportionality is K^{trans}. This simple first order differential equation can be solved to find:

$$C_t(t) = K^{trans}\int_0^t C_p(t')\mathrm{e}^{-\frac{K^{trans}}{v_e}(t'-t)}dt'. \qquad (2.24)$$

If the assumption $v_p \ll 1$ does not hold (i.e. if the tissue is highly vascular), then an additional term my be added to account for the contribution of the blood plasma concentration:

$$C_t(t) = v_p C_p(t) \, + \, K^{trans}\int_0^t C_p(t')\mathrm{e}^{-\frac{K^{trans}}{v_e}(t'-t)}dt'. \qquad (2.25)$$

This is commonly referred to as the *extended Tofts model*.

2.4.4 MRI Signal Conversion

During a DCE scan, we do not measure $C_t(t)$ directly. We measure a signal $S(t)$ in arbitrary units, which must be converted to C_t before PK modelling can occur. In this framework, S is the intensity of a voxel and t is the effective time the frame occurred. The first step to doing this is understanding how C_t affects T_1. As discussed, the contrast agent decreases the local T_1 relaxation time. Thus, since C_t is a function of time, T_1 will also be. It is common to assume a linear relationship between the relaxation rates and C_t:

$$\frac{1}{T_1(t)} = r_1 C_t(t) + \frac{1}{T_{10}}, \tag{2.26}$$

where r_1 is the *longitudinal relaxivity* (units: $\text{min}^{-1}\text{sec}^{-1}$) is a measure of how much the contrast agent affects the local T_1 and T_{10} is the initial T_1 when no contrast agent is present. This means an additional scan must be performed before DCE to find this value. So, we now need to find a relationship between T_1 and S.

As mentioned above, it is common practice to use a short TR value in DCE scans. When this occurs, M_z may not fully recover to M_0 resulting in saturation. However, if M_z returns to the same value after each RF pulse in a sequence of pulses the magnetization is said to be in a *steady state*. To connect this to S, lets assume that just before an RF pulse is applied the longitudinal magnetization has a value of M_{zA}, which may or may not be equal to M_0. An RF pulse is applied for a flip angle α such that the new value of M_z is:

$$M_{zB} = M_{zA} \cos \alpha \, . \tag{2.27}$$

The M_z will relax according to equation 2.7, such that the magnetization at a time TR later will be:

$$M_{zC} = M_{zA} \cos\alpha \exp\left(-\frac{TR}{T_1}\right) + M_0\left(1 - \exp\left(-\frac{TR}{T_1}\right)\right), \tag{2.28}$$

A steady state magnetization will occur with $M_{zC} = M_{zA}$. Plugging this into equation (2.28) and solving for M_{zA} results in:

$$M_{zA} = M_0 \frac{1 - \exp\left(-\frac{TR}{T_1}\right)}{1 - \cos(\alpha)\exp\left(-\frac{TR}{T_1}\right)}, \tag{2.29}$$

The signal is proportional to the transverse magnetization which is equal to $\sin\alpha\, M_z$, ignoring T_2 effects (equation 2.13). Thus, if we add a constant of proportionality k we find the *spoiled gradient equation*:

$$S(t) = kM_0 \sin\alpha \frac{1 - \exp\left(-\frac{TR}{T_1(t)}\right)}{1 - \cos(\alpha)\exp\left(-\frac{TR}{T_1(t)}\right)}, \tag{2.30}$$

Combining equation 2.26 and 2.30, one can convert between the tissue concentration of the contrast agent and the resulting MR signal.

2.5 Image Quality Metrics

IQMs are computer algorithms that allow a machine to calculate a score for an image that corresponds to the quality of that image. Quality, in this sense, is defined individually by each IQM. For example, quality may be defined by the average error between pixels of a reference and degraded image or by how by how much shared information there is between the two images. For full-reference IQMs, the image quality score is calculated for a degraded image Y relative to a reference image X. This section will summarize the 10 full reference IQMs used in this thesis.

2.5.1 Root Mean Squared Error (RMSE)

The RMSE is the simplest of all full-reference IQMs. It is the square root of the sum of squares of the pixel-by-pixel difference between the reference and the degraded images. The equation for the RMSE for images X and Y with N pixels is:

$$RMSE(X,Y) = \sqrt{\frac{\sum_{i=1}^{N}\left(X(i) - Y(i)\right)^2}{N}}. \qquad (2.31)$$

Where $X(i)$ is the value of the i^{th} pixel in image X.

2.5.2 Peak Signal to Noise Ratio (PSNR)

The PSNR is a transformation of RMSE based on the interpretation of the degraded image as the sum of the reference image and an error signal. When one subtracts the reference image from the degraded image, like when calculating the RMSE, what remains is the error signal. The PSNR is the ratio of the power of the maximum signal in the reference image to the average signal in the error (i.e. RMSE), given by the equation:

$$PSNR(X,Y) = 20\log\left(\frac{\max(X)}{RMSE(X,Y)}\right). \qquad (2.32)$$

PSNR is a built-in MATLAB function.

2.5.3 Structural Similarity Index (SSIM)

The RMSE and PSNR incorporate no information about how the human visual system (HVS) interprets an image and so are generally accepted to not correlate well with subjective scoring of image quality. The SSIM [34] was formed with the hypothesis that the HVS was adapted to extract structural information, such as edges, from an image. So a loss of structural information will correspond to a loss of image quality as perceived by a human. The SSIM is also normalized for local variations in luminance and contrast. The comparison of luminance, contrast, and structure between X and Y are given by the equations:

$$l(X,Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_y^2 + C_1}, \tag{2.33}$$

$$c(X,Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_Y^2 + \sigma_Y^2 + C_2}, \tag{2.34}$$

$$s(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}, \tag{2.35}$$

respectively, where μ is the mean of the signal, σ is the standard deviation, σ_{XY} is the covariance between X and Y, and C_1, C_2, and C_3 account for instabilities when the local luminance or contrast is close to zero. SSIM is the weighted product of l, c, and s such that the local SSIM for pixel i is:

$$SSIM(X,Y) = [l(X,Y)]^{\alpha}[c(X,Y)]^{\beta}[s(X,Y)]^{\gamma}$$

where α, β and γ control the relative importance of each component. Each component can all be weighted individually, but it is common to weight them all equally and set $C_3 = C_2/2$, resulting in the equation:

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \tag{2.36}$$

The SSIM is calculated on local 8x8 squares with Gaussian weighting and then averaged to generate a single score for the image. SSIM is a built in MATLAB function.

2.5.4 Multi-scale SSIM (MSSSIM)

MSSSIM [46] extends the SSIM to multiple scales. In this context, scale is related to the size of features and how visible they are at different resolutions and viewing distances. This can be an important factor to consider in image quality assessment since the HVS is not uniformly sensitive to all spatial frequencies. The MSSSIM operates on up to five scales. For each scale, the images are down-sampled by a factor of two and passed through a low-pass filter. The structure and contrast are compared at each scale, but the luminance is only compared at the final scale M. The final score is then the weighted product of each of these components:

$$MSSSIM(X,Y) = [l_M(X,Y)]^{\alpha_M} \prod_{j=1}^{M} \left[c_j(X,Y)\right]^{\beta_j} \left[s_j(X,Y)\right]^{\gamma_j}, \qquad (2.37)$$

The weights α_j, β_j, and γ_j can all be set individually, but again it is common practice to keep the weights equal for each scale. The authors of this metric determined the default weights by measuring the visibility of degradations at different scales. Scales for degradations that were determined to be more visible were then weighted higher. Normalizing the weights to unity resulted in the following default weights: $\beta_1 = \gamma_1 = 0.0448, \beta_2 = \gamma_2 = 0.2856, \beta_3 = \gamma_3 = 0.3001, \beta_4 = \gamma_4 = 0.2363,$ and $\alpha_5 = \beta_5 = \gamma_5 = 0.1333.$ MATLAB code for MSSSIM can be found at https://www.ece.uwaterloo.ca/~z70wang/publications/msssim.html.

2.5.5 Information Weighted SSIM (IWSSIM)

In all the methods presented to far, a quality score is calculated locally and then averaged uniformly across the whole image, a process called error pooling to generate a single scalar value. IWSSIM [47] was developed with the hypothesis the error should not be pooled uniformly, but should be weighted by the local information content i.e. regions where the HVS is drawn towards, such as edges or regions of fine texture. Each pixel is then assigned a weight w_i proportional to the amount of information that it contains. For information on how these weights are calculated please refer to the IWSSIM reference [47]. Aside from the weighting, IWSSIM is calculated the same as MSSSIM:

$$ISSSIM(X,Y) = [l_M(X,Y)]^{\delta_M} \prod_{j=1}^{M} \frac{\sum_{i=1}^{N} w_{ji} \left[c_j(X_i, Y_i) s_j(X_i, Y_i)\right]^{\delta_j}}{\sum_{i=1}^{N} w_{ji}}. \qquad (2.38)$$

The factor δ_j is used for simplicity since $\alpha_j = \beta_j = \gamma_j$. The additional $\sum_i$ is only implied in equation 2.37 since all the local calculations are averaged uniformly. There is no summation on l_M because no weights are applied on the M^{th} scale. Code for IWSSIM can be found at https://ece.uwaterloo.ca/~z70wang/research/iwssim/.

2.5.6 Gradient Magnitude Similarity Deviation (GMSD)

The GMSD [48] is an attempt to develop an IQM that correlates well with subjective opinion of image quality, but is also computationally efficient. GMSD is the standard deviation of the gradient magnitude similarity (GMS) map. The GMS map of X and Y is:

$$GMS(\vec{x}) = \frac{2m_X(\vec{x})m_Y(\vec{x}) + c}{m_X^2(\vec{x}) + m_Y^2(\vec{x}) + c}, \qquad (2.39)$$

where $\vec{x}$ is the index of the pixel, c is a stability constant, similar to C_1 and C_2 in SSIM, and m is the gradient magnitude maps of X and Y. The gradient maps are generated by convolving a horizontal and vertical Prewitt filter, h, with each image:

$$m_X = \sqrt{(X * h_x)^2 + \left(X * h_y\right)^2}, \qquad (2.40)$$

where $*$ is the convolution operator. The gradient map for Y is similarly defined. MATLAB code for GMSD can be found at http://www.comp.polyu.edu.hk/~cslzhang/IQA/GMSD/GMSD.htm.

2.5.7 Feature Similarity (FSIM)

Similar to SSIM, FSIM [49] measures local structure, however it does so by measuring a property called the phase congruency (PC). The PC is a dimensionless feature that is high for regions where the underlying Fourier components are in phase. Since PC is dimensionless, it is invariant to changes in intensity and scale, which makes it attractive for image quality measurements.

If we consider the Fourier expansion of a 1-D signal:

$$F(x) = \sum_n A_n \cos(n\omega x + \phi_n), \tag{2.41}$$

the phase congruency of the signal can be calculated by solving the following equation:

$$PC(x) = \max_{\theta \in [0,2\pi]} \frac{\sum_n A_n \cos(n\omega x + \phi_n - \theta)}{\sum_n A_n}, \tag{2.42}$$

although more sophisticated techniques are typically used in practice [50].

The PC map for X and Y can be calculated for each pixel (indexed by vector $\vec{x}$) to generate the PC similarity map:

$$PCS(\vec{x}) = \frac{2PC_X(\vec{x})PC_Y(\vec{x}) + c}{PC_X^2(\vec{x}) + PC_Y^2(\vec{x})^2 + c}, \tag{2.43}$$

where again c is a small constant to add stability. Similarly, a GMS is calculated the same as 2.39. A net similarity map is generated from the weighted product of the PCS and GMS:

$$S_L(\vec{x}) = [PCS(\vec{x})]^{\alpha}[GMS(\vec{x})]^{\beta}. \tag{2.44}$$

The weights α and β can be chosen independently, but again it is common practice to set them both to unity.

The map $S_L(\vec{x})$ is a measure of FSIM across the whole image. A weighted average is used to calculate a single scalar value, similar to IWSSIM. The weights are generated from the maximum value of PC_X and PC_y at each location, i.e. $w(\vec{x}) = \max(PC_X(\vec{x}), PC_Y(\vec{x}))$. So, the final formula for FSIM is:

$$FSIM = \frac{\sum_{\vec{x}} w(\vec{x}) S_L(\vec{x})}{\sum_{\vec{x}} w(\vec{x})}, \tag{2.45}$$

MATLAB code for FSIM can be found at https://www4.comp.polyu.edu.hk/~cslzhang/IQA/FSIM/FSIM.htm.

2.5.8 Noise Quality Metrics (NQM)

The authors of NQM [51] start from the framework that the degraded image is generated from the reference image degraded by linear frequency distortions and additive noise, however the two degradations are treated as independent by the HVS. This is a similar interpretation to the design of PSNR, but PSNR works under the assumption that all degradation can be modelled by noise whereas NQM separates the two components. NQM is calculated from the equation:

$$NQM = 10 \log\left(\frac{\sum_{\vec{x}} O_s^2(\vec{x})}{\sum_{\vec{x}}\big(O_s(\vec{x}) - I_s(\vec{x})\big)^2}\right), \tag{2.46}$$

where O_s and I_s are the simulated versions of X and Y. Full details on how O_s and I_s are calculated can be found in the NQM paper [51]. NQM was implemented using the MeTriX MuX MATLAB package: https://github.com/sattarab/image-quality-tools.

2.5.9 Visual Information Fidelity (VIF)

All IQMs presented so far defined the quality of the degraded image as the similarity/fidelity to the reference image. VIF [52] takes a different approach where it is hypothesized that a degradation in image quality is due to a loss of information. In this model, the degraded image is modelled as the reference after it has passed through a distortion channel (Figure 2.10). VIF is then related to how much information that is present in the reference image can be extracted from the degraded image.

The reference image is modelled in the wavelet domain using Gaussian scale mixtures based on natural scene statistics [53]. The distortion channel is modelled by common natural image degradations such as blurring, additive noise, and global/local contrast changes. The HVS is essentially modelled as an additional distortion channel that limits the available information. These distortions are modelled entirely as zero mean Gaussian noise. The VIF is defined as the ratio of mutual information, I, between the reference image and the output of the HVS channel when the distortion channel is present to when it is not present in the wavelet domain:

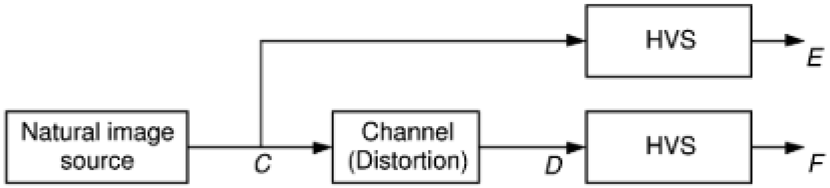


Figure 2.10: Framework for VIF calculations [52]. A natural image C is passed through a distortion channel to generate image D. Both images would be viewed by the HVS as images E and F.

$$VIF = \frac{\sum_{j \in \text{subbands}} I\left(C^j, F^j | s^j\right)}{\sum_{j \in \text{subbands}} I\left(C^j, E^j | s^j\right)}, \tag{2.47}$$

where subbands are the subbands of the wavelet decomposition, s are the natural scene statistic model parameters, and $C, F,$ and E are as defined in Figure 2.10. MATLAB code for VIF is also found in the MeTriX MuX package.

2.5.10 High Dynamic Range Visual Difference Predictor (HDRVDP)

HDRVDP [54] can be considered a “bottom up” approach to image quality assessment. Whereas “top down” approaches like SSIM work from assumptions about the image and how that will affect visual quality, a bottom up approach works from assumptions about the HVS such as the optical transfer function, light scattering in eyes, and intensity sensitivity. HDRVDP works as both a visual difference predictor (i.e. probability of detecting a difference between two images) and a measure of image quality. Although the metric was designed for high dynamic range natural images, it has been used to evaluate the quality of medical images [55]. MATLAB code for HDRVDP can be found at http://hdrvdp.sourceforge.net/wiki/.

Chapter 3: Correlating IQMs and Radiologists' Scores of Image Quality

Contents of this chapter are based on the manuscript submitted to IEEE Transactions on Medial Imaging: "Comparison of Objective Image Quality Metrics to Expert Radiologists' Scoring of Diagnostic Quality of MR Images" by Allister Mason et al. © 2019 IEEE.

3.1 Study Motivation

As discussed in Chapter 1, the purpose of an MR image is to convey diagnostic information to a radiologist. This means the quality of the image can be defined as how well it conveys this information [56]. Since, in a clinical setting, it is typically radiologists who assess these images, the radiologists' ratings of image quality can be considered an appropriate measure of diagnostic image quality. However, applying this standard on a large scale is often infeasible due to large image library sizes, limited time availability of expert radiologists, and the inherent variability in a subjective scoring technique. Objective IQMs provide an alternative to manual subjective scoring by allowing image quality to be calculated by a computer. However, the relationship between radiologists' rating of medical image quality and IQM scores is not well explored.

The RMSE and, more recently, the SSIM [34] are two of the most popular full-reference IQMs used in the MRI literature for validating new image acquisition [57] and reconstruction techniques [58], including machine learning algorithms [59–62]. This may in part be driven by the fact that they are widely available and implemented by default within existing environments such as MATLAB. SSIM has also been implemented as an automated measure of MR image quality directly into new techniques. For instance, Hansen *et al.* [63] used SSIM to estimate singular-value thresholds to denoise C-13 data, and Akasaka *et al.* [64] used SSIM to guide the choice of regularization weight in compressed sensing reconstruction.

Since the goal in these studies is to generate images of high diagnostic value, an implicit assumption when using RMSE/SSIM is that they will correlate well with how a radiologist would perceive image quality. Along these lines, the metrics should correlate with radiologists' rating of diagnostic image quality for a variety of image contrasts and degradations that can affect the diagnostic quality of an MR image such as noise or motion. Moreover, many other objective IQMs besides RMSE/SSIM have been developed in the image processing literature, and there may exist a more appropriate choice of IQM for assessment of MR image quality. To our knowledge, the efficacy of RMSE and SSIM have not been previously studied in this specific manner.

Some previous studies that have attempted to quantify performance of common full reference objective IQMs for MR images have used non-expert raters [65,66]. For many non-medical IQM studies (for example, rating the quality of a television picture) the use of non-expert raters is sufficient because the quality is to be optimized for the target audience [67], which is usually a non-expert. However, this is not the case for medical images because medical images are designed to be viewed by expert radiologists. Through their specialized training, radiologists learn to evaluate images from a unique clinical perspective and so may have different opinions of image quality compared to non-expert raters [32]. A previous study by Renieblas *et al*. used expert raters, but only studied IQMs from the SSIM family [68]. Our work is extended to a more diverse group of IQMs.

An understanding of the relationship of different objective IQMs for MR images and diagnostic image quality as perceived by a radiologist is important because, as discussed above, these metrics are increasingly used to validate and design new imaging techniques.

Therefore, it is imperative that the choice of metric reflect the goal: a high-quality diagnostic image for radiologist assessment. This study investigates correlations between RMSE, SSIM, and eight other common objective IQMs with radiologists' assessments of diagnostic MR image quality. We hypothesize that the RMSE and SSIM will not demonstrate the highest correlation with radiologists' ratings of diagnostic quality of MR images. High performing IQMs should have consistently high correlations across image type, anatomical regions, and degradation types.

3.2 Methods and Materials

3.2.1 Generation of Image Library

Reference images were selected from Picture Archiving and Communication System of the QEII Health Sciences Centre in Halifax, NS with Research Ethics Board approval. All images were anonymized before being transferred to a research server. The need for patient consent was waived. These MR images – clinically indicated and previously interpreted diagnostically as being negative for clinically relevant pathology - were chosen by a Royal College certified radiologist to have high signal to noise ratio, no visible artifacts, and no visible malignancy. The decision to use images void of pathology was made pointedly, given that intraindividual variation in pathology leads to differences in lesion conspicuity, whereas for this study the only desired differences were those introduced by the degradation. Nine reference images were selected from the abdomen and from the brain each. All reference images were acquired with either a GE 3T MR 750 Discovery or 1.5T Signa HDxt scanner. Of the nine abdominal images, three each were of the liver (post-contrast axial T1 LAVA-Flex), pancreas (pre-contrast axial T1 LAVA-Flex), and prostate (axial T2 PROPELLER). For the brain images, three axial T2 FLAIR, three axial T2

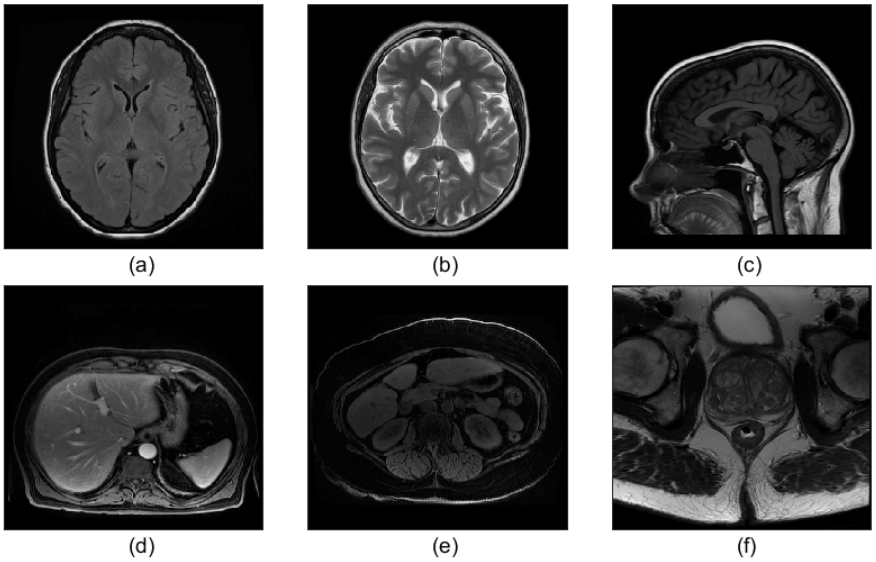


Figure 3.1: Representative set of reference images. (a) T2 FLAIR, (b) T2 PROPELLER, (c) T1 FLAIR, (d) T1 LAVA-FLEX (post contrast), (e) T1 LAVA-FLEX (pre-contrast), (f) T2 PROPELLER. © 2019 IEEE.

PROPELLER, and three sagittal T1 FLAIR images were used (Figure 3.1). All reference images were of size 512x512, except for two of the axial T2 FLAIR images, which were 256x256. The reference images were originally stored as 16-bit integers with varying dynamic range. The differences in dynamic range complicated the objective image quality assessment calculations. To mitigate this problem, the images were first converted to 32-bit floating point and normalized to unit intensity by linear contrast stretching.

Six types of degradations were individually applied to each image: white Gaussian noise, Gaussian blurring, Rician noise, undersampling of k-space data, wavelet compression, and motion artifacts. All degradation techniques were retrospectively applied in varying strengths to each reference image, with the exception of motion, which was only applied to the brain reference images. Retrospectively degrading the images allowed for controlled and consistent degradation of the images. Further, a ground truth reference image is available with this approach, which is how full reference IQMs are used in practice. The degradation techniques were chosen for their commonality in MR images and

Table 3.1: Description of the six degradation types used. The control parameter controls the degree of each degradation. Degradation strengths for parameters were chosen at random between the minimum/maximum values. © 2019 IEEE.

Degradation Type	Control Parameter	Minimum value	Maximum value
White Noise	Standard deviation of Gaussian	0.03	0.1
Gaussian Blur	Standard deviation of Gaussian kernel (pixels)	1	4
Motion	Percentage of frame shifted	0	10
Rician Noise	Standard deviation of Gaussian noise applied to k-space	0.02	0.05
Undersampling	Undersampling factor	2	20
Wavelet Compression	Threshold level	0.01	0.25

use in other similar imaging studies [69]. For each degradation type, a control parameter was determined that controlled the strength of the degradation. Each degradation was applied to each reference at four different strengths. This yielded a total image library of 414 images including the reference images. Table 3.1 provides a summary of degradation methods and control parameter ranges, and a representative set of the degradations is shown in Figure 3.2.

The Gaussian white noise and Rician noise images were generated by adding noise of a Gaussian distribution either directly to the image, or to the real and imaginary components of the Fourier transforms of the image, respectively. Gaussian blurred images were generated by applying a 2D Gaussian smoothing kernel of specified standard deviation σ, which defines the strength of the degradation. Again, the standard deviation of the Gaussian distribution controlled the strength of these degradations. To implement motion artifacts, the 2D image was repeatedly Fourier transformed as it was translated horizontally across the frame to simulate motion during k-space acquisition, similar to Braun *et al*. [70]. After each Fourier transform, four lines of k-space were stored in a separate array. Once

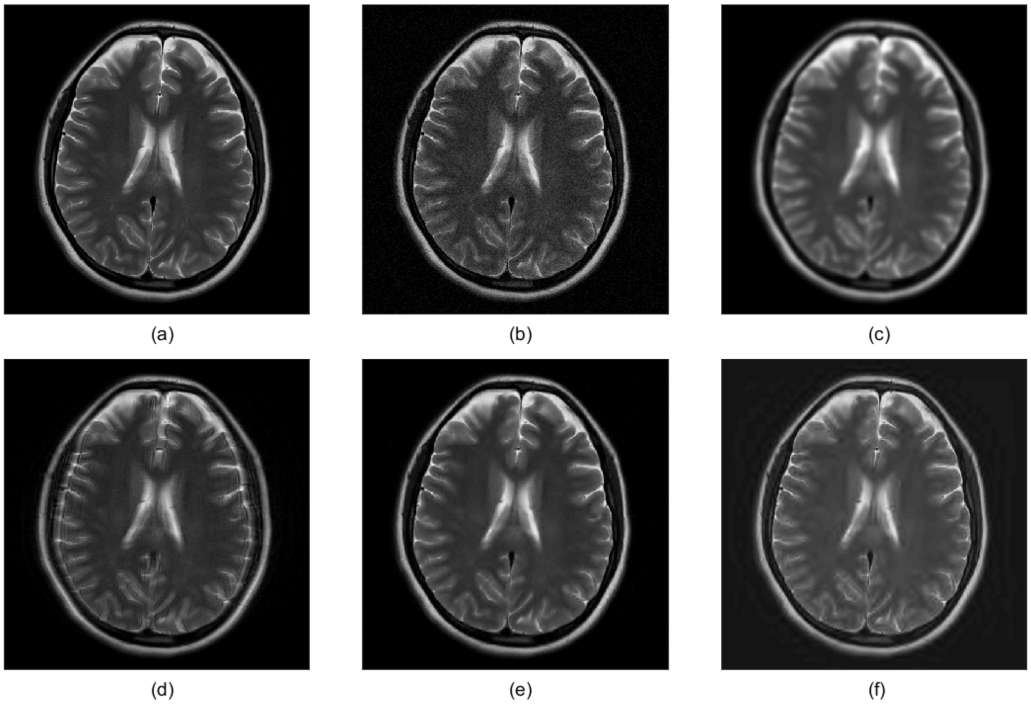


Figure 3.2: Five degradations applied to a 512x512 T2 PROPELLER image of the brain. (a) Reference image. (b) White noise (σ =0.04). (c) Gaussian blur (σ = 3). (d) Motion (percent shift = 6). (e) Undersampling (R = 8) with CS reconstruction. (f) Wavelet compression (threshold = 0.2). © 2019 IEEE.

the k-space was filled, it was Fourier transformed back to the image domain. Undersampling was introduced by retrospectively removing components from the Fourier transform of the images. While any pattern of under-sampling could have been arbitrarily used (e.g., Poisson disc, radial, etc.) in this work we used the CIRcular Cartesian UnderSampling pattern (CIRCUS) [71]. Undersampled images were reconstructed with the BART toolbox [42] wavelet regularization with a regularization weight of 0.01. The undersampling factor controlled the strength of this degradation. Finally, wavelet compressed images were generated by applying a global threshold of a specified value to the wavelet transform of the reference image. Wavelet transforms were generated for four levels with *sym8* type wavelets.

3.2.2 Objective IQMs

Ten full-reference objective IQMs were included in this study: RMSE, PSNR, SSIM, MSSSIM, IWSSIM, GMSD, FSIM, NQM, VIF and HDRVDP v2. Other IQMs have been proposed in the literature, but these metrics were chosen for their prevalence and performance in other imaging studies as well as ease of implementation for diagnostic images. A brief description of each of these metrics is given in Section 2.5.

3.2.3 Radiologist Image Quality Assessment

Three body radiologists and two neuro radiologists were involved in the study. The radiologists scored only images from their subspecialties, using a 1-5 Likert scale. The radiologists were asked to rate the diagnostic quality of the images with respect to delineation of relevant anatomy and ability to confidently exclude a lesion on a 5-point scale as follows: excellent diagnostic quality (5), good diagnostic quality (4), fair diagnostic quality (3), poor diagnostic quality (2), and non-diagnostic (1). This scale was calibrated by consensus of radiologists in each subspecialty based on a training set to span the full range of quality of the images. The training set consisted of three reference MR images (different from those in the testing set) and degraded images generated by applying each degradation technique to each reference at two representative strengths. All judgments of quality were made in their opinion as diagnostic radiologists (e.g. their ability to discriminate relevant tissues, their confidence in using the image to detect, or in this case not detect, pathology, etc.). All subsequent scoring was performed individually. Agreement between radiologists scores were calculated using Cohen's kappa with quadratic weighting [72].

3.2.4 Data Analysis

The scores for the radiologists were not evaluated in their raw form. To account for potential differences in quality of each reference image, raw scores were converted first to a difference score:

$$D_{mn} = s_{m,ref} - s_{mn}, \tag{3.1}$$

where D_{mn} is the difference score for the m^{th} radiologist on the n^{th} degraded image, $s_{m,ref}$ is the raw score of the m^{th} radiologist for the reference image corresponding to the n^{th} degraded image, and s_{mn} is the raw score of the m^{th} radiologist on the n^{th} degraded image. These scores were then converted to a z-score to account for differences in mean and standard deviation for each radiologist:

$$z_{mn} = \frac{(D_{mn} - \mu_m)}{\sigma_m} \tag{3.2}$$

where μ_m and σ_m are the mean and standard deviation of the difference scores of the m^{th} radiologist. This converts all the scores for each radiologist to a zero mean, unit standard deviation distribution. The z-scores from each radiologist were then averaged and rescaled from 0-100.

The Spearman rank order correlation coefficient (SROCC) was calculated between the transformed radiologist scores and each of the IQM scores. The SROCC is equivalent to a linear correlation coefficient on the rank order of the data. A higher SROCC would then correspond to a better performing IQM. This metric was used because of the nonlinear relationship between subjective scores and objective IQM scores [73] (visible in Figure 3.3).

Correlations were calculated under three scenarios: when scores were divided by individual radiologists, by image type, and by degradation type. For the first division, SROCCs were calculated between an individual radiologist’s scores and the IQM scores of the images they scored. This division also includes a “combined” group, which averages

Table 3.2: Information about each group of images in the testing set for each of the three subdivisions used for analysis. N is the number of images in each group. For groups in which some images received different numbers of ratings, the number in parenthesis in the ratings per image column is the number of images rated by the accompanying number of radiologists, i.e. 36 images degraded by white noise were rated by 3 radiologists and the other 36 were only rated by 2 radiologists. © 2019 IEEE.

By radiologist			By degradation type			By image type		
Image group	N	Ratings per image	Image group	N	Ratings per image	Image group	N	Ratings per image
Combined	414	3(189) or 2(225)	White Noise	72	3(36) or 2(36)	T1 Flair	75	2
Body Rad 1	189	1	Gaussian Blur	72	3(36) or 2(36)	T1 GRE (post)	63	3
Body Rad 2	189	1	Motion	36	2	T1 GRE (pre)	63	3
Body Rad 3	189	1	Rician Noise	72	3(36) or 2(36)	T2 Flair	75	2
Neuro Rad 1	225	1	Undersample	72	3(36) or 2(36)	T2 PROP (brain)	75	2
Neuro Rad 2	225	1	Wavelet	72	3(36) or 2(36)	T2 PROP (body)	63	3

all radiologists' scores for each image in the study. The image type division presents SROCCs for each group of reference image. The radiologists' scores for images in each group are averaged and the SROCC is calculated with the corresponding IQM scores. Finally, images are grouped by degradation type where all images degraded by a particular technique are grouped and the SROCC is calculated. The sizes of each group are described in Table 3.2.

A variance-based hypothesis test was performed to measure statistical significance in the difference in the performance of the IQMs. First, a non-linear regression was performed on the IQM scores according to the equation:

$$Q_p = \beta_1\left(\frac{1}{2} - \frac{1}{1 + \exp\left(\beta_2(Q - \beta_3)\right)}\right) + \beta_4 Q + \beta_5, \tag{3.3}$$

where Q are the original IQM scores and β are the model parameters [73]. The residuals between the IQM scores after the regression and the radiologists scores were calculated and Gaussianity was confirmed by measuring the kurtosis of the residuals. Residuals with a kurtosis between 2 and 4 were taken to be Gaussian (97% were found to be Gaussian). To test for statistical differences in the variance of residuals an F-test of equality of variances was performed, with the null hypothesis being that the residuals of two IQMs come from distributions of equal variance (with 95% confidence). Since each IQM was compared to all nine other IQMs, a Benjamini-Hochberg correction for false discovery rate controlling was performed [74]. An IQM performed statistically better than another IQM if the variance of its residuals was statistically less than the variance of the residuals of the other IQMs.

3.2.5 IQM Calculation Times

The time required by various IQM algorithms to calculate a quality score is also of interest to researchers looking to adopt these metrics. To measure this, we repeatedly calculated all IQMs for all body images (512x512, N = 189). Timing calculations were performed in MATLAB 2017b running on a 24 CPU Linux research server (Intel X5650, 2.67GHz).

3.3 Results

Figure 3.3 shows the radiologists' scores versus each IQM for the "combined" subgroup, which is the combination of all images and all radiologist scores. The data is sub-divided by degradation type. The IQMs are ordered by decreasing average SROCC when the data is broken up by each radiologist. This order is kept throughout all results for consistency. As shown in Figure 3.3, all IQM scores demonstrated a trend to improve (decrease for RMSE and GMSD, increase for all others) as radiologists' image quality scores increase. However, a varying strength of this trend is seen among IQMs. The sensitivities of each of the IQMs to different degradation types can also be clearly seen. The weighted kappa

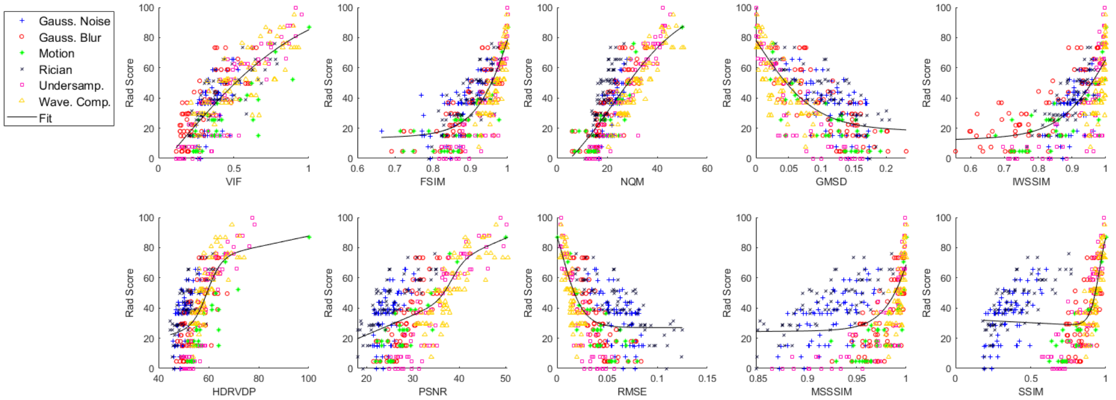


Figure 3.3: Relationship between subjective radiologist score and IQMs for the full image library (414 images). Images are sorted by degradation type. The fit is calculated on all images using the non-linear regression in equation 3.3. © 2019 IEEE.

Table 3.3: Weighted Cohen's kappa for scores between radiologists rating the same set of images. © 2019 IEEE.

	Body Rad 1.	Body Rad. 2	Body Rad. 3	Neuro Rad. 1	Neuro Rad. 2
Body Rad. 1	1.00	0.850	0.655	--	--
Body Rad. 2	0.850	1.00	0.780	--	--
Body Rad. 3	0.655	0.780	1.00	--	--
Neuro Rad. 1	--	--	--	1.00	0.615
Neuro Rad. 1	--	--	--	0.615	1.00

between radiologists who rated the same set of images is presented in Table 3.3. The distribution of radiologists scores is shown in Figure 3.4.

The SROCC of each IQM with each radiologist's scores are shown in Figure 3.5. This ordering of the IQMs show the decreasing correlations of the IQMs with the radiologist scores in each subgroup. Note that RMSE and SSIM are among the metrics with the lowest SROCC. Overall, VIF had the highest SROCC values. Results of the hypothesis testing on the variance in the residuals for each IQM in each subgroup are shown in Table 3.4. The sorted IQMs show that for many of the radiologists in the study, VIF, FSIM, and NQM perform statistically better than the other IQMs included in the study. SSIM did not perform

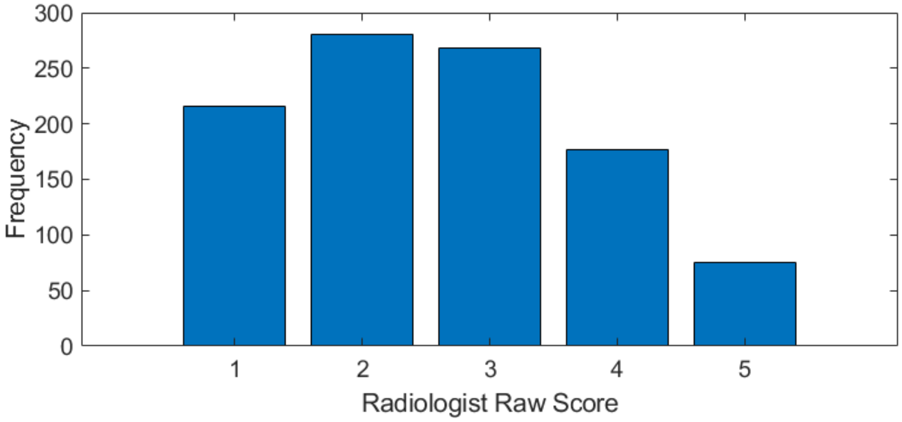


Figure 3.4: Histogram showing the distribution of radiologist scores.

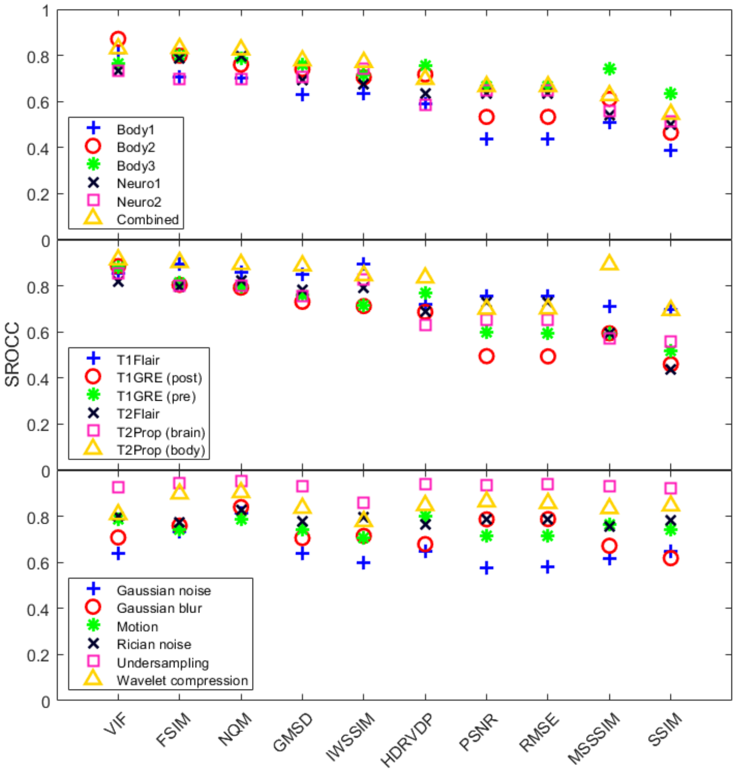


Figure 3.5: Spearman rank order correlation coefficient for each IQM when data is divided by radiologist (top), reference type (middle), or degradation type (bottom). © 2019 IEEE.

statistically better than any another IQM including RMSE. When dividing data by reference image type, similar results were obtained (Figure 3.5). The same metrics demonstrated higher correlation with radiologists' scores. The statistical differences in performance based on the variance in the residuals of IQMs also demonstrated a similar

Table 3.4: Statistical significance in residuals of IQM scores after regression and radiologists scores (significance level = 0.05 with Benjamini-Hochberg correction for multiple comparisons) when data is broken up by radiologist. A ‘1’ means the IQM performed statistically better than the IQM of the column. A ‘0’ means it was statistically worse. A ‘-’ means no significant difference. The order of the sub-elements is: combined, body radiologist 1, body radiologist 2, body radiologist 3, neuroradiologist 1, and neuroradiologist 2. © 2019 IEEE.

	VIF	FSIM	NQM	GMSD	IWSSIM	HDRVDP	PSNR	RMSE	MSSSIM	SSIM
VIF	------	-11---	--1---	111---	111---	111---	111---	111---	111-11	111---
FSIM	--0---	------	------	1-----	1-1-1-	1-----	1-11--	1111--	111-1-	1-11--
NQM	--0---	------	------	1---1-	1-1-1-	1---1-	1-111-	11111-	111-1-	1-111-
GMSD	000---	------	0---0-	------	------	------	------	--1---	1-1--1	1-----
IWSSIM	000---	0-0-0-	0-0-0-	------	------	------	------	------	------	------
HDRVDP	0-0---	0-----	0---0-	------	------	------	--1---	--1---	1-1---	--1---
PSNR	000---	0-0---	0-000-	--0---	------	--0---	------	------	1----1	------
RMSE	000---	0000--	00000-	--0---	------	--0---	------	------	1-----	------
MSSSIM	000-00	000-00	000-00	0-0--0	0----0	000---	0----0	0----0	------	------
SSIM	000---	0-00--	0-000-	------	------	--0---	------	------	------	------

Table 3.5: Statistical significance in residuals of IQM scores after regression and radiologists scores (significance level = 0.05 with Benjamini-Hochberg correction for multiple comparisons) when data is broken up by reference image type. A description of the table format is provided in Table 3.4. The order of the sub-elements is: T1 Flair (brain), T1 LAVA-Flex (post-contrast, liver), T1 LAVA-Flex (pre-contrast, pancreas), T2 Flair (brain), T2 Propeller (brain), T2 Propeller (prostate). © 2019 IEEE.

	VIF	FSIM	NQM	GMSD	IWSSIM	HDRVDP	PSNR	RMSE	MSSSIM	SSIM
VIF	------	------	-1----	-1----	-11--1	-1--11	-11--1	-11--1	11111-	-11-1
FSIM	------	------	------	------	------	------	-----1	-----1	1-111-	---1-
NQM	------	------	------	------	--1---	------	--1--1	--1--1	--111-	--11-
GMSD	------	------	------	------	------	------	------	------	----1-	-----
IWSSIM	-0---0	------	------	------	------	------	------	------	------	-----
HDRVDP	----0-	------	------	------	------	------	------	------	------	-----
PSNR	-00--0	-----0	--0--0	-----0	------	------	------	------	----1-	-----
RMSE	-00--0	-----0	--0--0	-----0	------	------	------	------	------	-----
MSSSIM	00000-	0-000-	0-000-	0---0-	0---0-	--0---	----0-	----0-	------	----0
SSIM	-000-0	---0-0	--00-0	------	------	------	------	------	------	-----

behaviour (Table 3.5). The fewer significant differences is likely due to fewer images in each subgroup. When dividing the data by degradation type, the variation in SROCC between IQMs is less clear (Figure 3.5). This lack of variation is shown by the small number of statistically different performance of the IQMs (Table 3.6). NQM appears to

Table 3.6: Statistical significance in residuals of IQM scores after regression and radiologists scores (significance level = 0.05 with Benjamini-Hochberg correction for multiple comparisons) when data is broken up by degradation type. A description of the table format is provided in Table 3.4. The order of the sub-elements is: Gaussian noise, Gaussian blur, motion, Rician noise, undersampling, and wavelet compression. © 2019 IEEE.

	VIF	FSIM	NQM	GMSD	IWSSIM	HDRVDP	PSNR	RMSE	MSSSIM	SSIM
VIF	------	------	------	------	------	------	------	------	------	------
FSIM	------	------	------	------	----11	------	------	------	-----1	------
NQM	1---11	------	------	1---1-	1---11	1-----	1-----	1-----	1---11	1-----
GMSD	------	------	------	------	------	------	------	------	------	------
IWSSIM	----0-	----00	----00	----0-	------	----0-	----0-	----0-	------	----0-
HDRVDP	------	------	------	------	----1-	------	------	------	------	------
PSNR	------	------	------	------	----1-	------	------	------	------	------
RMSE	------	------	0-----	------	----1-	------	------	------	------	------
MSSSIM	------	-----0	----00	------	------	------	------	------	------	------
SSIM	------	------	------	------	------	------	------	------	------	------

Table 3.7: Time for calculations in seconds of each metric for all body images (resolution: 512x512, N = 189). © 2019 IEEE.

	VIF	FSIM	NQM	GMSD	IWSSIM	HDRVDP	PSNR	RMSE	MSSSIM	SSIM
mean	241.38	31.11	39.21	1.77	42.01	403.37	0.32	0.23	14.98	6.07
Standard error	6.52	0.16	0.91	0.03	0.36	5.39	3E-3	7E-3	0.41	0.03

perform particularly well for images degraded by noise as it has a statistically better performance than all other IQMs except FSIM for these images. IWSSIM performed poorly for images degraded by undersampling artifacts, showing statistically larger residuals compared to all other IQMs for these images. The hypothesis tables are not perfectly symmetric due to the due to the Benjamini-Hochberg corrections for multiple comparisons.

The calculation times of the IQMs are shown in Table 3.7. The simple and rapid algorithms of RMSE, PSNR, and GMSD demonstrate short calculation times (all less than 2 seconds). SSIM and MSSSIM have slightly longer calculation times (less than 20 seconds) and FSIM, NQM, and IWSSIM are longer still (less than 45 seconds). The VIF

(241±6seconds) and HDRVDP (403±5seconds) IQMs are shown to have the longest calculation times of the metrics in this study.

3.4 Discussion

The results of this study may have important implications for researchers who are developing MR image acquisition and reconstruction techniques and using objective IQMs to test, validate and/or optimize their techniques. Recently, SSIM has gained popularity as a complement to RMSE as the IQM of choice in imaging studies, with the underlying assumption that it provides a more accurate measure of image quality. However, the results of this study demonstrate that, in the retrospectively degraded images used, SSIM does not show a significantly stronger correlation with radiologist opinion of diagnostic image quality than RMSE, and that there are other objective IQMs that perform better. This does not imply that previous studies that use RMSE or SSIM are invalid, since RMSE and SSIM were still seen to correlate with radiologists' scores, but that there exist other metrics that may provide a more accurate measure of diagnostic image quality.

When considering the trends in Figure 3.3 and the bottom of Figure 3.5, it appears as if the factor that most affects IQM performance is how uniformly the IQM quantifies the quality of images with different degradations. For instance, in Figure 3.3 VIF shows substantial overlap of all degradation techniques, but as one progresses through the metrics, the distributions of images degraded with different degradation techniques in the Rad Score-IQM plane become much more distinguishable. One can clearly discern the distributions of different degradation techniques for IQMs such as PSNR or RMSE. In the extreme case of SSIM, a bimodal distribution appears between the noise and other degradations. As seen in Figure 3.5, when the images are divided by degradation type, each

IQM has a similar average SROCC with the radiologists score. It is only when the different degradations are grouped together that the clear differences in performance of IQMs arise. This is important to notice because, as discussed in the Motivation of this Chapter, an IQM should correlate with radiologists' opinion over a range of degradations. This also shows how choice of degradation types and strengths can affect the results of studies of this nature.

After normalizing each radiologist's score and combining scores across all images, we found that VIF exhibited the highest SROCC of all the metrics evaluated in this study. These results suggest that VIF provides the most accurate surrogate measure of subjective image quality scores of a radiologist of the IQMs included in this study. VIF is unique among IQMs in this study in that it generates a quality score based on shared information between the reference image and the distorted image, instead of generating a score from an algorithm based on some definition of signal fidelity. In VIF, the information in the reference image is calculated from natural scene statistics. The distorted image is modelled as the reference image passed through a distortion channel. The VIF is found from the information remaining in the degraded image from the reference image. It should be noted that VIF is designed with natural scene statistics, not medical image statistics, indicating an area of potential future research.

FSIM and NQM also consistently demonstrated high correlations with the radiologist scores. Indeed, NQM had performed statistically better than VIF for images degraded with Gaussian noise, undersampling, or wavelet compression. This is consistent with other similar studies of MR images [65,66]; however, a key distinction between these studies and ours is that we correlated with the image quality scores of expert radiologists instead of non-experts. For Renieblas *et al*. [68], who did use expert raters, moderate agreement in

SROCC was found from IQMs used in both our study and theirs (SSIM: 0.54 versus 0.44; MSSSIM: 0.63 versus 0.60). Differences are likely due to differences in degradation techniques/strengths, as well as variability in the subjective scoring by experts.

For all IQMs, there is significant variation in the radiologists scores for a particular value of the IQM (Figure 3.5). For example, in images that had an averaged scaled radiologist score of 50, the VIF ranged from 0.30-0.68. Similar trends can be seen in the results of Chow *et al.* [66]. This highlights the difficulties in using IQMs as a replacement for manual scoring of image quality, since there will always be variability in the manual scoring and the IQMs and radiologists may have different sensitivities or preferences for different forms of degradation. As machine learning algorithms advance, it is possible they may be able to learn this sensitivity and preferences in ways objective IQMs cannot.

There are some aspects of our study which may limit the generalizability of our results. First, we limited the scope of this study to ten objective IQMs. Since many IQMs exist beyond those studied here, including all of them would not be feasible. It is possible that a metric not included in this study could exist or be developed that demonstrates a higher correlation with radiologists' scoring of image quality than any of the IQMs we evaluated. We also limited our choice of IQM to full-reference IQMs because these are most commonly used for the validation of imaging techniques. Full-reference IQMs allowed us to use retrospectively degraded images, which adds more control to the study, but may add artificiality to the degraded images and limit the generalizability of the results in practice. A similar study with no-reference IQMs may also be considered, particularly for techniques that wish to assess the diagnostic quality of MR images on the scanner as part of a built-in quality assurance system. The present study only included data from brain images and body images, but presented both independently and together, which allows for the interpretation of the data for specific applications. However, our results may not be generalizable to other MRI systems, anatomical regions, or even different MRI sequences. Finally, it should be noted that the scoring of diagnostic quality in a “clinically normal” MR image is strongly related to but not necessarily equivalent to scoring of an image containing pathology. Our current work focused on clinically normal images – a critical first step in determining if IQMs developed for non-experts rating natural images would also correlate with radiologists rating diagnostic quality of MR images – which will inform future task-based studies examining whether these same IQMs correlation with other measures such as lesion conspicuity scores in images containing pathology.

3.5 Conclusions

We measured the correlations between 10 full-reference objective IQMs and the subjective image quality score of five subspecialty radiologists. When considering images divided by reference location or combining all images in the study, SSIM and RMSE demonstrated statistically worse performances than other metrics evaluated, suggesting that SSIM and RMSE are potentially not ideal surrogate measures of MR image quality as determined by radiologist evaluation. The VIF, FSIM, and NQM demonstrated the highest correlation with radiologists’ opinions of MR image quality. However, these metrics come at the cost of longer calculation times, which may influence their use in future research. Differences in the performances of the IQMs were also largely lost when images are divided by degradation type. Both the IQM SROCC values and calculation times presented in this study should be considered in future imaging studies applying an objective IQM to assess the quality of an MR image, for example in studies evaluating novel image reconstruction algorithms. These data also highlight the importance of the ongoing development of techniques for automatic and objective assessment of image quality.

Chapter 4: An IQM-Based Heuristic for Accurate Pharmacokinetic Parameter Measurement in DCE MRI

4.1 Motivation

The PI-RADS v2 steering committee recommends the temporal resolution of DCE MRI be at least 10 seconds and preferably less than seven seconds [10]. Technically this is not difficult to achieve, however without advanced image reconstruction techniques the quality of the reconstructed image will be unacceptable. The choice of temporal resolution will not just affect the visual assessment of the DCE scan, it can introduce bias into the recovered PK parameters [75]. Standard DCE imaging techniques acquire data at a fixed temporal resolution that cannot be altered once it is acquired. This is an impersonal technique that does not allow for any optimization of temporal resolution on a patient-by-patient basis.

Golden angle radial sampling with a sliding window reconstruction allows for multiple reconstructions of the same data set with arbitrary temporal resolution. This technique allows for a high degree of flexibility during reconstruction because the different temporal resolution reconstructions could be optimized for different purposes. For instance, one could be optimized for visibility of the early signal enhancement in the image series, one could be optimized for input into a machine learning algorithm, and one could be optimized for accurate PK modelling. This chapter focuses on the third scenario: optimizing the temporal resolution of DCE MRI scan for accurate PK modelling.

The work presented in this chapter extends previous work done in our group that looked at the relationship between image quality and accuracy of recovered function parameters (e.g. amplitude) of simulated sinusoidal and exponential functions [39]. That work utilized simulations in a numerical phantom in MATLAB that could simulate a DCE scan while certain features of the phantom evolved by the exponential and sinusoidal functions. Using a golden angle-based sampling pattern, the DCE data could be reconstructed with arbitrary temporal footprint[1]. Since the ground truth function parameters were known exactly, it was possible to measure the effect of temporal footprint on the error of the recovered parameters. Further, since a ground truth of reference image was known, it was possible to study the relationship between the accuracy of the recovered function parameters and the image quality as defined by full-reference objective IQMs. A correlation was found between the recovered function parameters and the IQMs tested in the study (RMSE and SSIM).

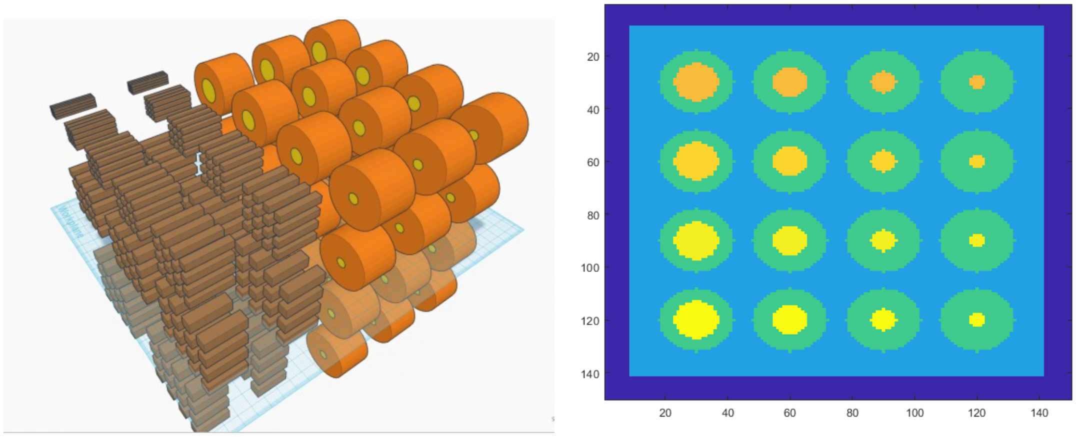
The current work extends these simulations to model the more physiologically relevant Tofts model using golden angle radial sampling with sliding window reconstruction. Ultimately, the goal of this work is to develop a technique that allows one to automatically chose a temporal footprint that maximizes PK parameter modelling accuracy on a patient-by-patient basis. Based on the correlation between IQMs and the function parameter accuracy of the previous study, we hypothesize that IQMs can be used to automatically choose a temporal footprint in DCE MRI that optimally balances the trade-off between temporal footprint and image quality to maximize PK parameter accuracy.

4.2 Methods

The simulations in this study make use of a numerical phantom originally created by Murtha *et al*. [39] (Figure 4.1) that runs in MATLAB. The matrix size of the phantom is

---

[1] As discussed in Section 2.1.4, the term temporal footprint is used in place of temporal resolution due to the clearer definition when using a sliding window reconstruction technique.



(a) (b)

Figure 4.1: Visualization of numerical phantom used in simulations. (a) Three-dimensional rendering of the phantom. (b) Cross-sectional view through a plane of cylinders.

defined by the user. While the cross-sectional dimensions can independently be any size, the number of slices is fixed to 32. The phantom contains groups of rectangular prisms and concentric cylinders, however only the concentric cylinders are used in these experiments. The cylinders are grouped together into objects called *features*. The feature contains all relevant information for the simulations including the indexed location of the feature, the signal enhancement curve, and its imaging parameters. For the concentric cylinders in each slice there are five features: all the outer cylinders in the slice (shown in orange in Figure 4.1a) form a single feature and all the inner cylinders (shown in yellow in Figure 4.1a) of each row form an individual feature. All results in this chapter are shown for a 150x150 cross-sectional phantom. For this size, the outer cylinders were 12 pixels in radius, and the inner cylinders are 7, 5, 4, and 2 pixels in radius.

The outer cylinder feature and the top three inner cylinder features of each slice was defined to evolve according to the extended Tofts model (equation 2.25), repeated here:

$$C_t(t) = v_p C_p(t) + K^{trans} \int_0^t C_p(t') \exp\left(-\frac{K^{trans}}{v_e}(t - t')\right) dt' , \qquad (4.1)$$

where C_p is the contrast agent concentration in blood plasma, and K^{trans}, v_p, and k_{ep} are the PK parameters that are used to fit the model (See Section 2.4.3). The AIF is equivalent to $C_p(t)$ after correcting for the hematocrit, which was equal to 0.4 in these simulations. In these simulations the bottom row of cylinders of each slice simulated an AIF that was calculated from the empirical model created by Parker *et al.* [76]. The simulated signal $S(t)$ in each feature is calculated from $C_t(t)$ using equations 2.30 and 2.26. In these simulations $(T_{10})_{blood} = 1660$ msec, $(T_{10})_{tissue} = 1500$ msec, $r_1 = 3$ $\mathrm{s}^{-1}\mathrm{mmol}^{-1}$, TR = 5 msec, $\alpha = 12°$, and $k = M_0 = 1$. If a voxel was simulating the AIF, $(T_{10})_{blood}$ is used for calculating $S(t)$, otherwise $(T_{10})_{tissue}$ is used. For features simulating tissue, v_e and v_p were fixed at 0.2 and 0.02 respectively. For the outer cylinders, K^{trans} was set to 0.1/min always. The remaining nine rows of inner cylinders had K^{trans} values between 0.1/min and 0.5/min in steps of 0.05/min.

The k-space for the simulated DCE data was generated using a golden angle stack-of-stars sampling pattern with a single channel of uniform spatial sensitivity. One ray in k-space was collected per TR. To collect each ray, the phantom was Fourier transformed by using the `nufft` command in BART. The signal intensities for the phantom were then calculated a time TR later and the process was repeated for the duration of the simulated scan. All simulations created the equivalent of 120 seconds of DCE data.

Images were reconstructed from this data using a sliding window reconstruction, where the rays per frame determines the temporal footprint, which is the amount of time over which this section of rays was collected. Unless otherwise stated, all images were reconstructed with no overlap. Each group of rays was reconstructed independently to create each frame in the time series of images. For these experiments, images were reconstructed with between 7 and 69 rays per frame in increments of two. The equivalent temporal footprints are between 1.12sec and 11.03sec in steps of 0.32sec. The timestamp for each image in the series was the average of the times at which each of the constituent rays for that image were collected. All images were reconstructed using the `pics` command of the BART toolbox. All images were reconstructed with wavelet regularization with a regularization weight of 0.05.

PK parameter recovery from the reconstructed images was performed on a voxel-by-voxel basis. The value of $S(t)$ from a voxel is recovered from the intensity of that voxel in the time series of reconstructed images. First, $S(t)$ is converted back to $C_t(t)$ by inverting the SPGR equation. PK parameters are extracted from the recovered $C_t(t)$ data by fitting it to equation 4.1 via a least squares fitting algorithm [77]. This was done for two choices of AIF since there is still no widely accepted technique for finding the AIF in practice. The first was using a "population AIF", where the Parker AIF was recalculated for the time points of each image in the series. The second was a "simulated AIF" where the AIF used for fitting was also recovered from the simulated data by extracting the signal enhancement curve for the bottom row of cylinders, which were defined to evolve according to the Parker AIF.

Two image quality metrics were calculated for the reconstructed images in these experiments: the RMSE and the SSIM [34]. These IQMs were calculated under two cases of reference image. The first reference case used the known values of the numerical phantom as reference. For each time in the series of reconstructed images, the true value of the phantom was recalculated at that time. The IQMs were calculated between the phantom at that time and the corresponding frame then averaged over all frames in the series. However, this approach is not realistic in practice because a known ground truth does not exist in clinical MRI. The second case generated the reference from the simulated DCE data. This was accomplished by combining all rays collected during the final 30 seconds of the simulated scan. This was equivalent to 188 rays. This section of the scan was chosen because the temporal dynamics of the contrast agent are much slower in this region. The IQMs were then calculated for each frame during this time interval then averaged.

Additionally, similar simulations were performed with a phantom generated from a 3D clinical data set. The base 3D image was created by combining all rays from a 3D stack-of-stars DCE scan of the prostate into a single frame. A five-slice thick feature was defined in the center of prostate. Like the numerical phantom, this feature was a pair of concentric cylinders (Figure 4.2). The outer shell and each slice of the inner cylinder were defined individually such that they could each have their own independent signal enhancement curves. As with the numerical phantom, v_e and v_p were set to 0.2 and 0.02 in all cases, respectively. The value of K^{trans} for the outer cylinder was 0.05/min. For the inner

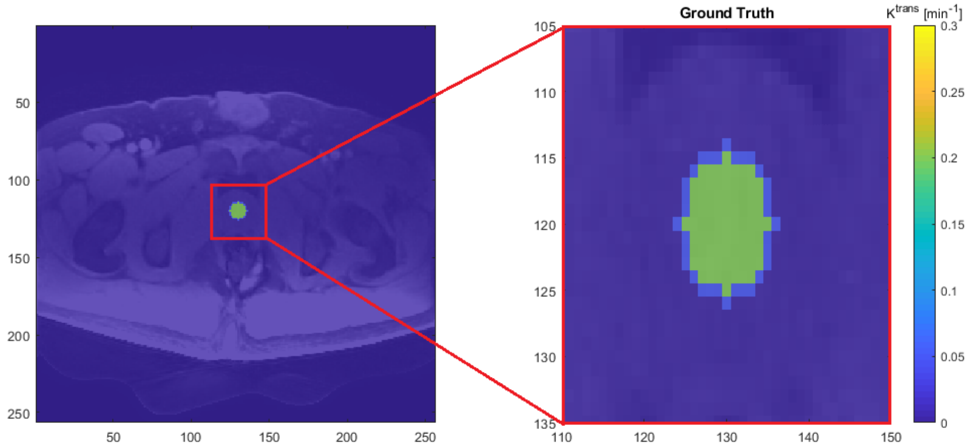


Figure 4.2: Cross-sectional view of the clinical data set used for the synthetic dynamic clinical simulations. This figure also shows the K^{trans} maps of the artificial tumour. The outer radius = 7 pixels, inner radius = 6 pixels.

cylinder in each slice the value of K^{trans} increased from 0.1/min to 0.5/min in steps of 0.1/min. The value of M_0 for each voxel was defined such that the intensity with zero contrast agent would give the intensity of the voxel in the initial static image. A cylindrical region (radius = 2 pixels) in a nearby blood vessel was defined to simulate the AIF. Beyond this, the framework of these simulations was then the same as that of the numerical phantom.

4.3 Results

4.3.1 Numerical Phantom

A representative set of images for reconstructions over the tested range of temporal footprints is shown in Figure 4.3. The images shown are the frame in each respective reconstruction closest to 20 sec after the start of the simulated scan, which is approximately when the peak of the AIF occurs. The contrast between the increased signal due to the higher K^{trans} values of the inner cylinders in the top three rows is visible. The increased

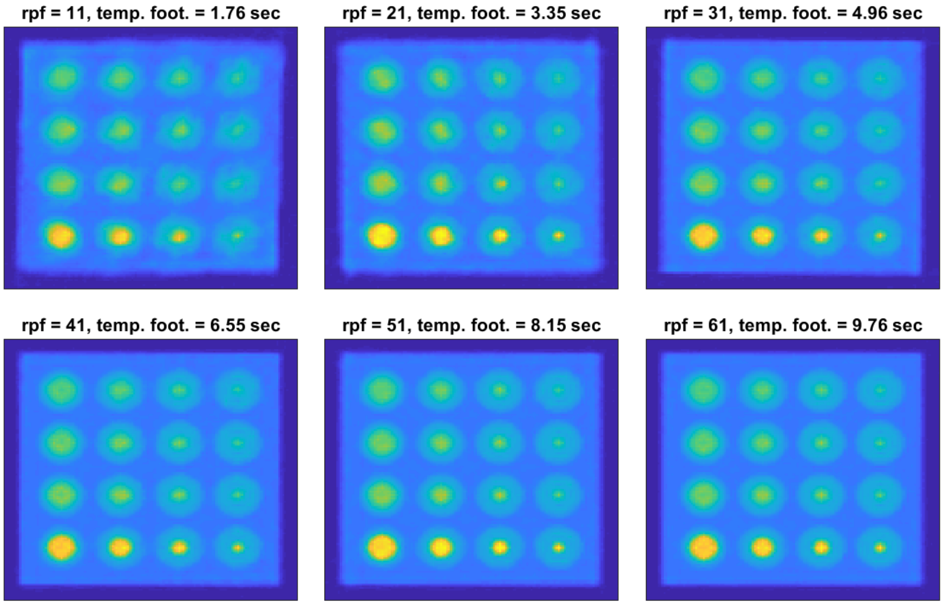


Figure 4.3: Representative set of images reconstructed from 150x150 numerical phantom (rpf = rays per frame).

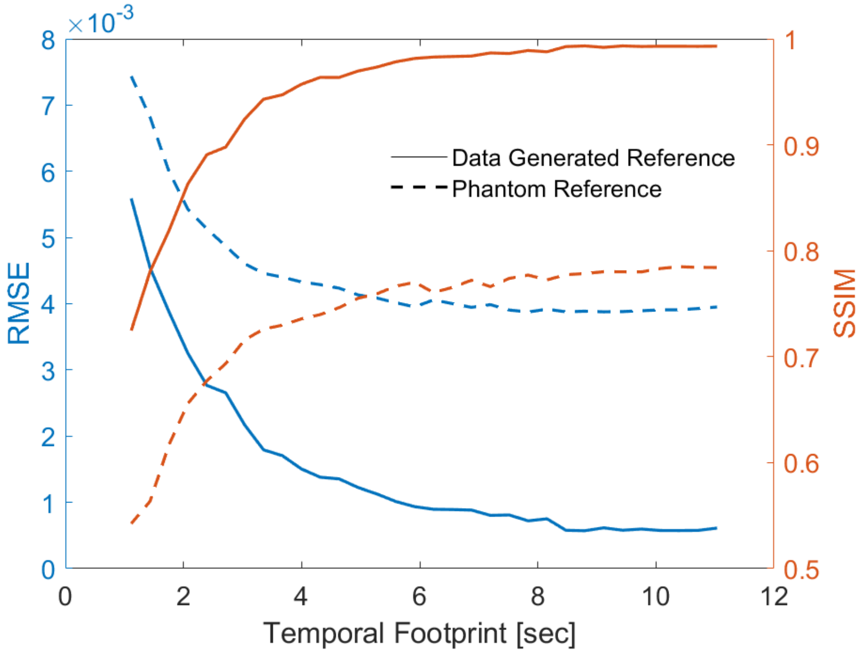
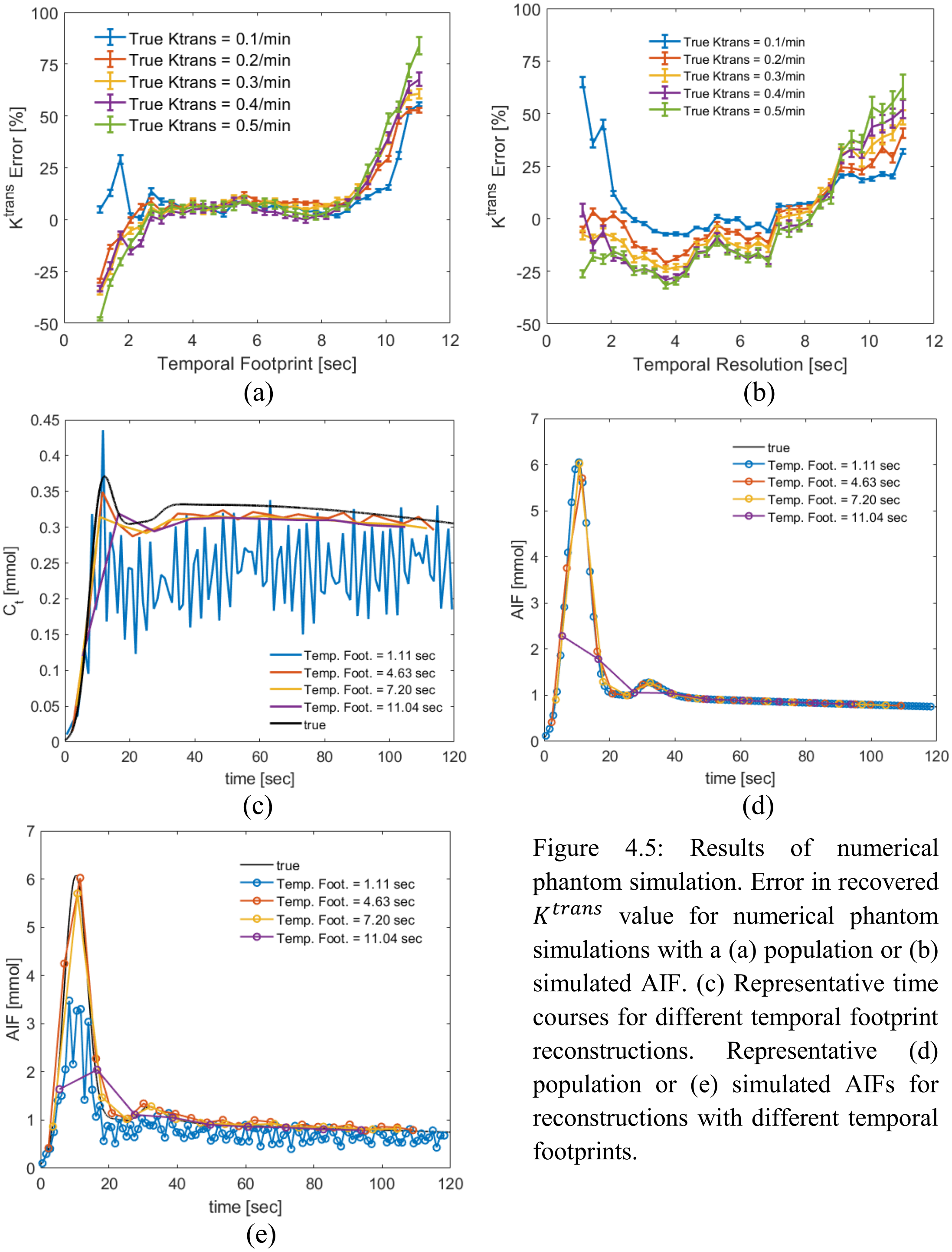


Figure 4.4: The average RMSE and SSIM of image series generated with different temporal footprints for both a data generated reference and the numerical phantom as reference.

signal intensity of the AIF is also visible in the bottom row. Note that even for the images with a large temporal footprint, the sharp edges of the numerical phantom are still blurred. Beyond a temporal footprint of approximately 8 seconds (50 rays per frame), the values of the IQMs begin to reach a plateau (Figure 4.4). This is observed for both the IQMs tested and both cases of reference.

A systemic bias was observed in the error of the recovered K^{trans} values depending on choice of temporal footprint size (Figure 4.5). The trends are consistent across all K^{trans} values tested. The systemic bias when PK parameters are recovered with the population AIF (Figure 4.5a) can be broken into three regions. At small temporal footprints, the value of K^{trans} is systemically underestimated. At large temporal footprints, the value of K^{trans} is systemically overestimated. However, in between these two extremes there is a region



(a) (b) (c) (d) (e)

Figure 4.5: Results of numerical phantom simulation. Error in recovered K^{trans} value for numerical phantom simulations with a (a) population or (b) simulated AIF. (c) Representative time courses for different temporal footprint reconstructions. Representative (d) population or (e) simulated AIFs for reconstructions with different temporal footprints.

(between a temporal footprint of approximately 4 and 8 seconds) where the recovered K^{trans} value is stable with error consistently less than 10%. Similar trends are observed for the K^{trans} values recovered with the simulated AIF (Figure 4.5b). However, the two data sets differ at small temporal footprints (less than ~4 sec).

4.3.2 Heuristic for Finding Optimal Temporal Footprint

The relationship between the recovered K^{trans} value and the RMSE/SSIM score of each of the reconstructed images of the numerical phantom is shown in Figure 4.6. Although the value of the temporal footprint is not shown explicitly in this figure, as the temporal footprint increases the IQMs move monotonically towards their optimal value (i.e. increasing for SSIM and decreasing for RMSE). For the small temporal footprint/highly undersampled images the IQMs are far from their optimal values. As the temporal footprint increases, the IQMs move towards their optimal values and the recovered K^{trans} value initially becomes more accurate as well. At the large temporal footprints, the overestimation of the K^{trans} values are again clear, but the IQMs stop changing. Clearly discernible in these plots is a vertical wall where the IQM does not pass. This corresponds to the temporal footprints where the image quality is “saturated” as discussed with Figure 4.4. Importantly, the base of the wall is located at the region of temporal footprints for

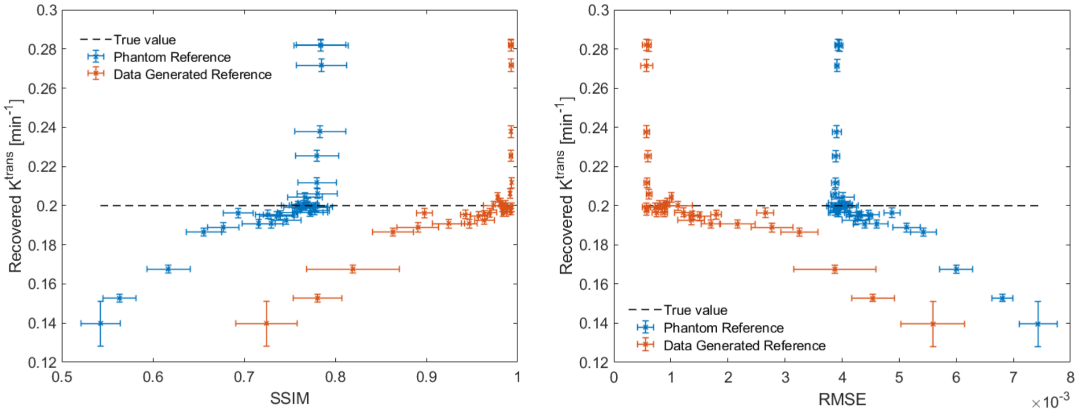


Figure 4.6: Relationship between recovered K^{trans} value and SSIM (left)/RMSE (right). Each marker corresponds to a reconstruction with a different temporal footprint. Shown here are the recovered values for the feature with $K^{trans} = 0.2$/min.

which the K^{trans} could be recovered consistently within 10% of the true values for the feature.

Thus, a technique to find the temporal footprint that corresponds to the base of this wall would allow one to locate a temporal footprint that can accurately recover the K^{trans} values. A heuristic for consistently finding this temporal footprint can be described in the following steps:

1. Begin at a sufficiently large temporal footprint where K^{trans} is overestimated (approximately 10 sec).
2. Reconstruct images at this temporal footprint and calculate the IQM of the reconstructed image series.
3. Decrease the temporal footprint. Reconstruct the images at this new temporal footprint and recalculate the IQM.
4. Repeat step 3 until the IQM of the reconstructed image changes significantly.

The choice of how to quantify a significant change in the IQM can be difficult to determine in a rigorous way, but we have found empirically that a threshold of the initial IQM score less twice the standard deviation in the initial reconstruction has been sufficient.

This heuristic was tested on the same simulated k-space data that was used to generate the results in Figures 4.3 – 4.6. The initial reconstruction was performed at a temporal footprint of 11 sec and each subsequent reconstruction was performed by decreasing the temporal footprint by 0.5 sec at each iteration. Once the IQM dropped below the threshold, the algorithm would stop. The results are shown in Figure 4.7, which uses SSIM as the figure of merit. As can be seen in Figure 4.7a, the SSIM initially remains flat. On the eighth iteration (corresponding to a temporal footprint of 7.5 sec), the SSIM drops below the

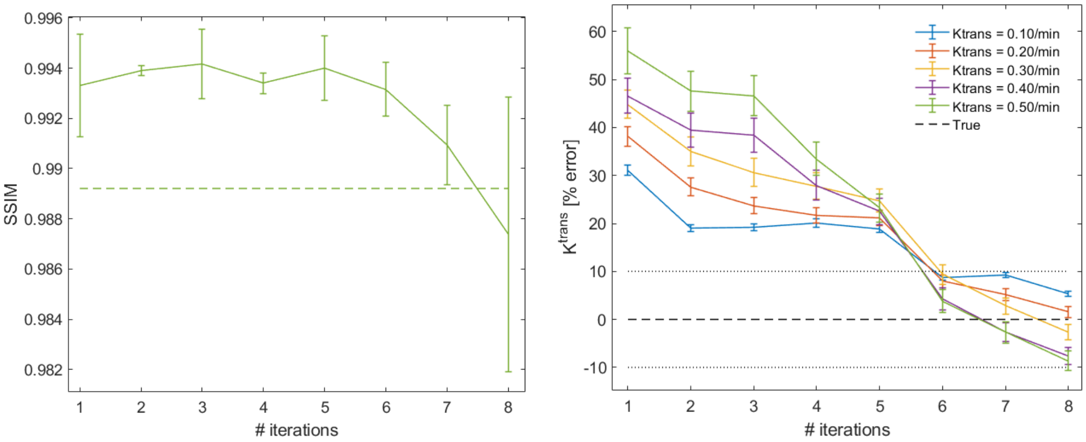


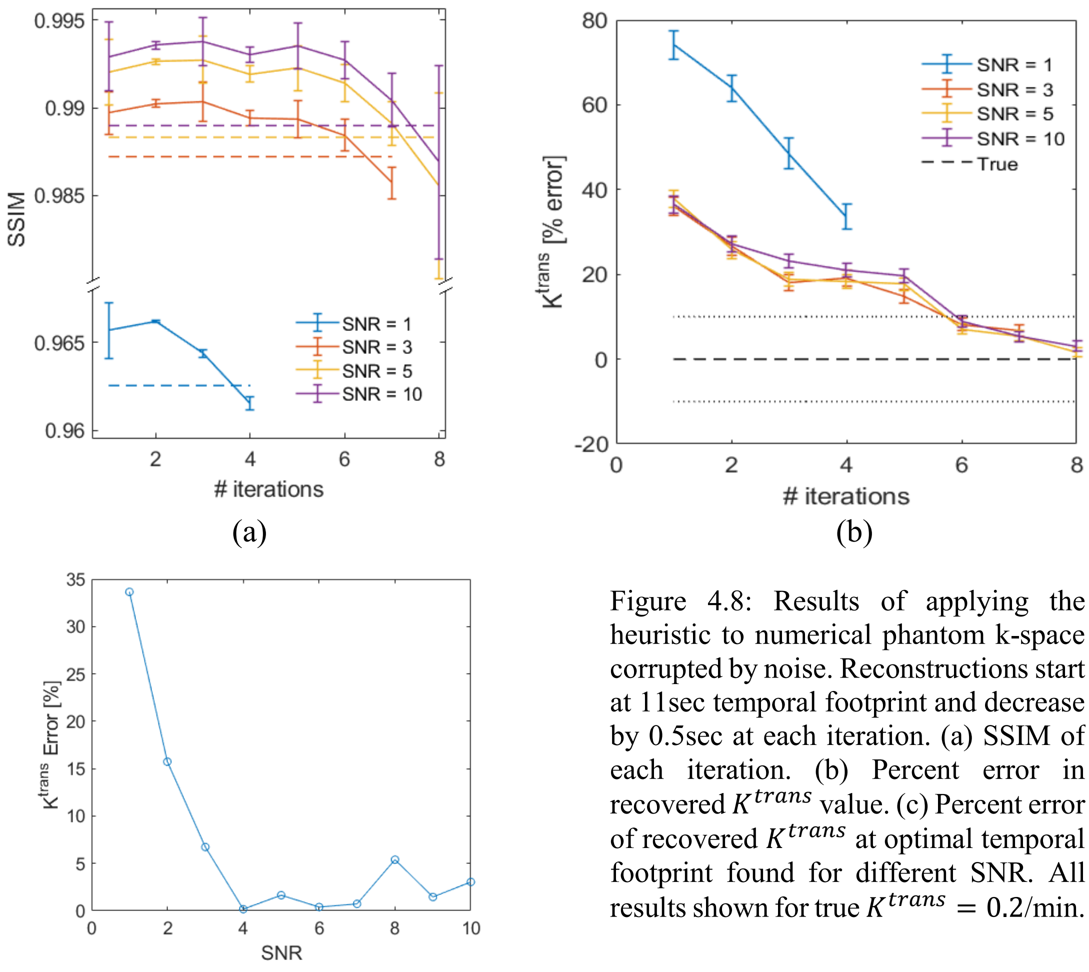
Figure 4.7: Results of applying the heuristic to find an appropriate temporal footprint for accurate K^{trans} recovery with the numerical phantom. The first iteration has a temporal footprint of 11 sec and all subsequent iterations decrease the temporal footprint by 0.5sec. The dashed line on the SSIM plot (left) is the cut-off for the algorithm. A simulated AIF was used to calculate the K^{trans} values.

defined threshold and the algorithm stops. The corresponding K^{trans} values of each iteration are shown in Figure 4.7b. The algorithm was able to recover the K^{trans} values of each of the features in the simulation to within 10% of their true values. Results are only shown using SSIM as the figure of merit, but the same results can be achieved using RMSE.

The algorithm does require sufficient SNR to work properly (see Figure 4.8). For low SNR, the IQM will fall below the target threshold even at a large temporal footprint, leading to overestimation of the K^{trans} value. However, this is unlikely to be an issue at clinically relevant SNR values (SNR > 5).

4.3.3 Effect of Overlapping Frames

The results of using a 50% overlap during sliding window reconstruction are shown in Figure 4.9. For the population AIF recovered K^{trans} values, the main difference between the 50% overlap data and no overlap is that the overestimation at large temporal footprints is much less drastic for the 50% overlap data (Figure 4.9a). However, this is not the case when the simulated AIF is used (Figure 4.9b), where a large positive bias is still observed.



(a)

(b)

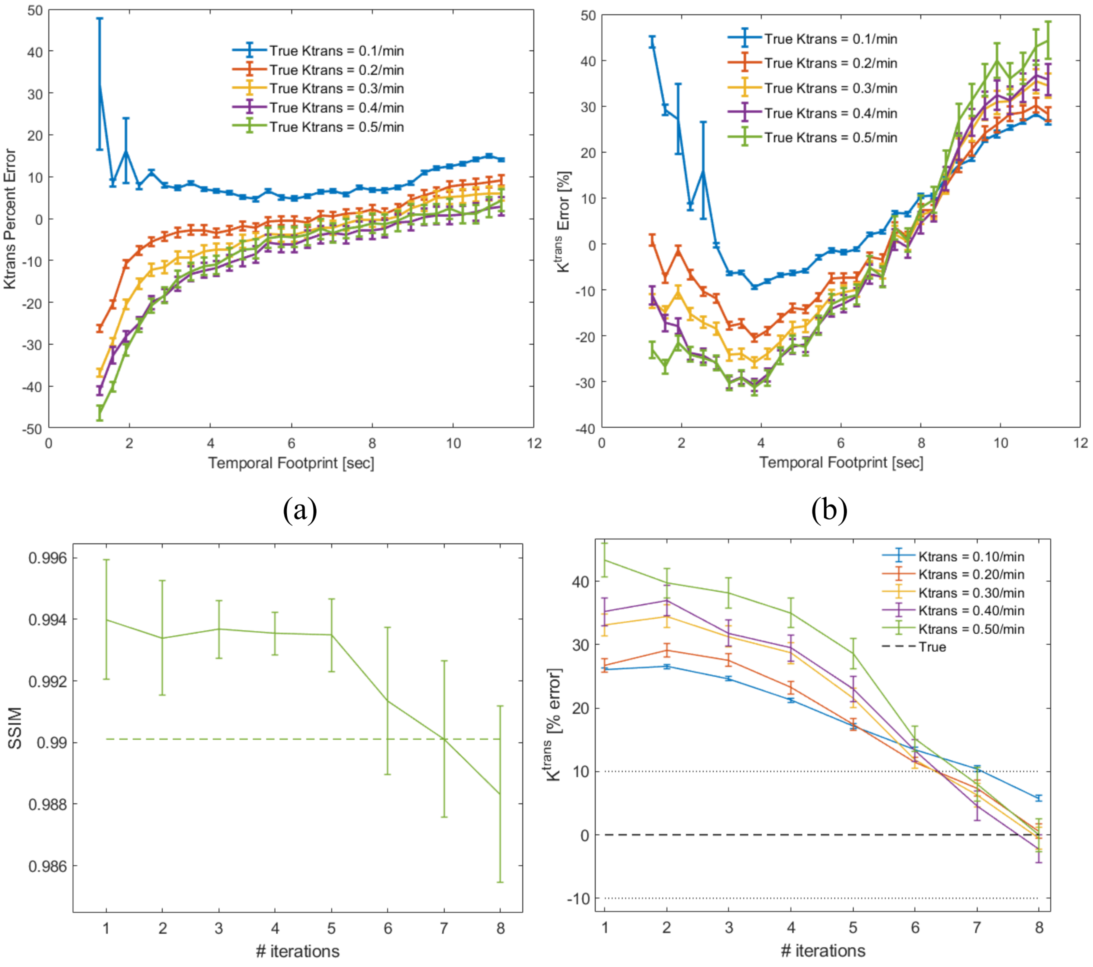
(c)

Figure 4.8: Results of applying the heuristic to numerical phantom k-space corrupted by noise. Reconstructions start at 11sec temporal footprint and decrease by 0.5sec at each iteration. (a) SSIM of each iteration. (b) Percent error in recovered K^{trans} value. (c) Percent error of recovered K^{trans} at optimal temporal footprint found for different SNR. All results shown for true $K^{trans} = 0.2$/min.

When applying the heuristic to the 50% overlap data using the simulated AIF, all K^{trans} values are recovered within 10% of their true values (Figure 4.9d). Compared to the noiseless no overlap data, the average error was slightly less in the 50% overlap data (average absolute error 1.83% vs 5.15%).

4.3.4 *Dynamic Synthetic Clinical Data*

The results are also consistent in the synthetic dynamic clinical data. The relationship between the K^{trans} error and the temporal footprint for this data set is shown in Figure 4.10. For the K^{trans} values recovered with the population AIF, the same three regions are



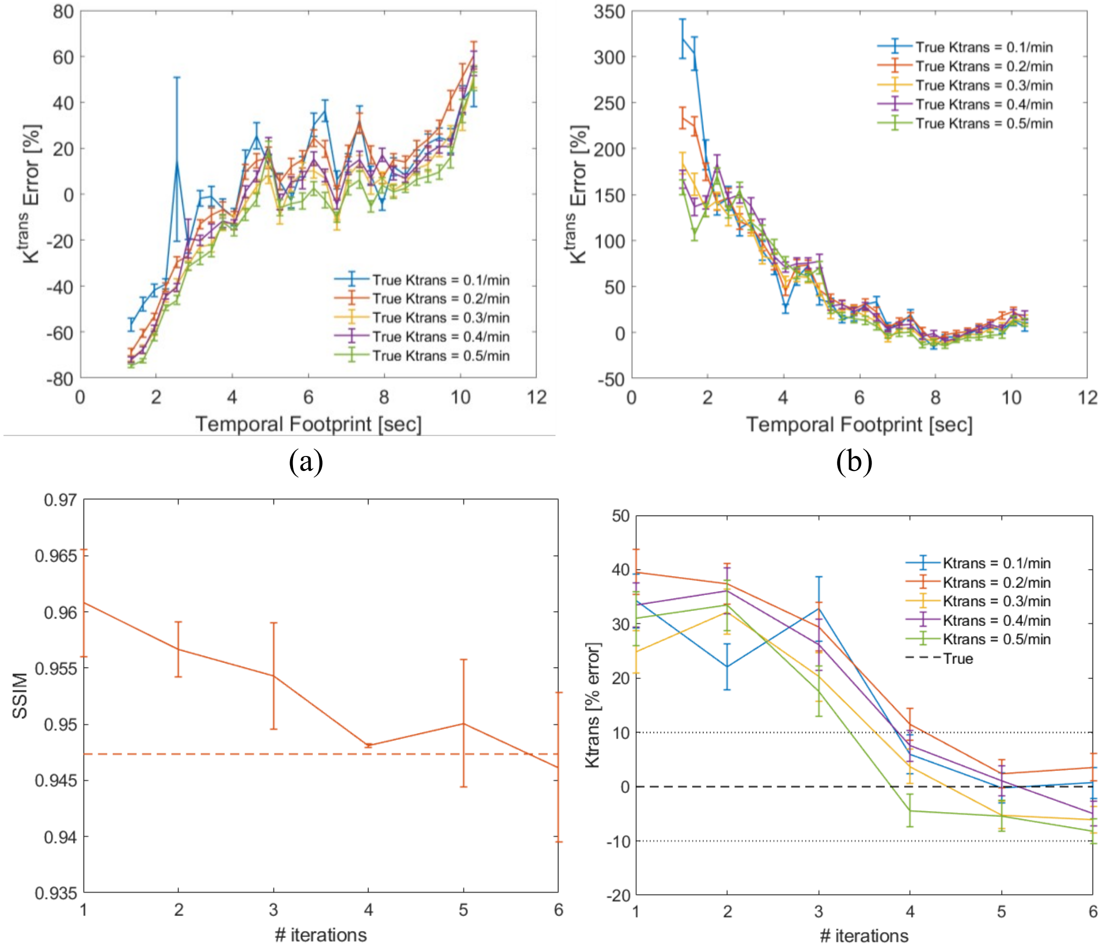
(a) (b)

(c) (d)

Figure 4.9: Effect of temporal resolution on K^{trans} error for numerical phantom data reconstructed with 50% overlapping of frames for curve fitting with (a) a population AIF and (b) a simulated AIF. The results of applying the heuristic to the data is shown in (c) and (d). Reconstructions start at 11sec temporal footprint and decrease by 0.5sec at each iteration. The simulated AIF data was used to calculate the K^{trans} values in (d).

visible as in the numerical phantom: negative bias at a small temporal footprint, positive bias at the large temporal footprint, and a region of stability in the middle. However, different from the numerical phantom data, there is significant positive bias in the K^{trans} at low temporal footprint when the simulated AIF is used for curve fitting.

Applying the heuristic to the synthetic clinical data set also achieves successful results (Figure 4.10 c and d). The algorithm ended after 6 iterations and again all the tested K^{trans} values were recovered within 10% of their true values. This corresponds to a temporal



(a) (b)

(c) (d)

Figure 4.10: Percent error in recovered K^{trans} value for the synthetic dynamic clinical data set reconstructed with (a) population AIF and (b) simulated AIF. Results of applying the heuristic to this data is shown in (c) and (d). The first iteration has a temporal footprint of 11sec and all subsequent iterations decrease the temporal footprint by 0.5sec. The dashed line on the SSIM plot (right) is the cut-off for the algorithm.

footprint of 8.5 seconds. However, when looking at the SSIM of each iteration (Figure 4.10c), no clear plateau is seen as was the case with the numerical phantom simulations. The threshold of twice the standard deviation less the initial SSIM still found an appropriate temporal footprint that recovered the K^{trans} values of the features within 10% of their true values.

4.4 Discussion

In this work, an IQM-based heuristic for choosing a temporal footprint to maximize PK parameter accuracy in DCE MRI is presented. The heuristic is based on the observation that when using golden angle stack-of-stars sampling with sliding window reconstruction, the image quality as measured by IQMs reaches a plateau for temporal footprints above ~8 sec depending on the image being reconstructed. It was found that at the temporal footprints just below this range a stable region occurred where the error in the recovered K^{trans} values is minimized. This led to the development of the heuristic that allows for tailoring of temporal footprint on a patient-by-patient basis by choosing a temporal footprint that is just below where the image quality plateau occurs.

4.4.1 Image Quality Measurements

The reason for the IQM plateau (Figure 4.4) is likely is because the marginal increase in k-space coverage tends towards zeros as more rays are combined. Without significant increase in k-space coverage there is no improvement in image quality and adding any additional rays only serves to increase the temporal footprint. It was also seen that the plateau of the IQMs calculated with the phantom reference is not as close to the optimal IQM value as when using the data-generated reference (optimal SSIM = 1; optimal RMSE = 0). The reason for this is because the phantom reference has artificially sharp edges that the reconstructed images cannot replicate, regardless of how many rays are combined into a single image. Since the data generated reference is itself an imperfect reconstruction of the numerical phantom, the reconstructed images can achieve IQMs much closer in magnitude to their optimal values for this case of reference.

4.4.2 Bias in Recovered K^{trans} Values

The bias in the K^{trans} values with respect to temporal footprint can be broken into three regions in general: small temporal footprints (less than ~4 sec), large temporal footprints (greater than ~8 sec), and the region in between. The recovered K^{trans} was systemically overestimated at large temporal footprints, but for a range of temporal footprints below this recovered K^{trans} error was minimized. This was true for both phantoms, overlap factors, and types of AIF. The reason K^{trans} is overestimated at large temporal footprints is likely because of a loss of scale of the AIF due to a combination of low sampling rate and temporal smearing, a phenomenon akin to blurring but through the temporal dimension (Figure 4.5d,e). The decreased values of the AIF are compensated by an increased value of K^{trans}. This effect is minimized for the 50% overlap data with a population AIF. The overlap doubles the sampling rate of the AIF, which decreases its loss of scale. However, even with the increased sampling rate, a large positive bias is observed when the simulated AIF is used in calculations. While the increased overlap increases the sampling rate, the temporal footprint is still the same size. This means that the simulated AIF is still susceptible to temporal smearing, where the signal is averaged across a large period of time. The full width half maximum of the initial peak of the AIF is approximately 7.9 seconds, so when the temporal footprint begins to exceed this width, the effect of temporal smearing will cause the intensity of the AIF to be decreased, resulting in a positive bias of the recovered K^{trans}.

Under all scenarios when the population AIF is used, the K^{trans} is systemically underestimated at small temporal footprints. This is likely due to a loss of scale of C_t due to signal washout (Figure 4.5c). Since C_t has a decreased intensity but the AIF does not since it is recalculated, K^{trans} is underestimated during curve fitting to compensate. However, results at small temporal footprints are not as consistent when the simulated AIF is used. For the numerical phantom, there was minimum bias at low temporal footprints when the simulated AIF was used. A possible explanation is that since the simulated AIF comes from the same images as C_t, any artifacts or degradations, such as signal washout, in the recovered C_t are also observed in the simulated AIF, which is recovered from the same data. For example, the AIF for the reconstruction with a temporal footprint of 1.11 sec (Figure 4.5e) shows a similar decrease in signal as the corresponding C_t. The effect is balanced out and the systemic underestimation of K^{trans} is not observed. The overestimation of the feature with K^{trans}= 0.1/min for this data (Figure 4.5b and Figure 4.9b) is potentially due to signal wash in, due to the low signal intensity of that feature. The synthetic dynamic clinical data saw significant positive bias in K^{trans} for low temporal footprints. This is likely a combination of two factors that are different than in the numerical phantom simulations. First, the feature in which the AIF is defined is smaller than it was in the numerical phantom, which leads to more signal washout and thus a more significant loss of scale of the AIF intensity. This smaller AIF was required due to the size of the vessel in the base image. Second, the size of the feature in which the AIF is defined and the feature in which the Tofts model time course is defined are not the same. Thus, the equivalent effects from the undersampling are not present in both features as was the case with the numerical phantom.

The region of stability was fairly large (~4 sec) in the numerical phantom data. This would suggest that finding a temporal footprint that minimizes K^{trans} error is not a difficult task. However, the stable region is not as broad in the synthetic dynamic clinical data

(Figure 4.10a). This motivated the need for the heuristic. This narrowing of the stable region is likely due to the increased complexity of the image, particularly in the region of the prostate where simulated features were defined. Due to the differences in M_0 values, each voxel had a different signal enhancement curve even for the same PK parameters, which could lead to the larger error. This was not the case in the numerical phantom where $M_0 = 1$ for all voxels.

4.4.3 The Heuristic

For the numerical phantom and the synthetic dynamic clinical data tested the optimal temporal footprints found with the heuristic was 7.5 sec and 8.5 sec. The larger temporal footprint of the clinical dataset likely represents the increased heterogeneity of the image. Due to the relatively broad stable region, there are multiple temporal footprints that would have led to equally accurate K^{trans} recovery. However, since the algorithm is designed to find the temporal footprint that is closest to the IQM wall, it chooses the temporal footprint in the accurate K^{trans} recovery region that maximize image quality as defined by the chosen IQM.

For both scenarios, the algorithm found a temporal footprint that is above the 7 seconds recommended by PI-RADS [10]. However, this is not necessarily a contradiction with the PI-RAD recommendations because these recommendations are interested in a different outcome than what this heuristic is optimized for. The heuristic presented here is designed for accurate PK parameter modelling, but the recommendations for PI-RADS is for visibility of early signal enhancement in cancerous tissue. This actually demonstrates one of the advantages of the golden angle stack-of-stars sampling with sliding window reconstruction technique over traditional fixed temporal footprint approaches: multiple reconstructions are possible. One can perform one reconstruction at a temporal footprint that is optimized for PK parameter fidelity and another that allows for the visibility of early signal enhancement.

There is another aspect of this research that seemingly contradicts another recommendation, one that was made in Chapter 3 of this thesis. In Chapter 3, it was found that RMSE/SSIM did not demonstrate the highest correlation with radiologists scoring of diagnostic image quality and are therefore not the most appropriate objective metrics for the scoring of MR image quality. The reason that RMSE and SSIM were able to be used for the work presented in the current Chapter is that, similar to the PI-RADS discussion, the goal was not to optimize image quality for radiologists' visual assessment, but to optimize it for PK parameter modelling. Since the purpose of the images is different, the results from Chapter 3 are not transferable to this chapter. The only requirement is that the IQM must have a finite optimal value. For example, the optimal value of PSNR is infinite, so it would not be a suitable IQM for this heuristic. Without a finite optimal value, the wall at this value could not occur and the heuristic would not work because there would always be a change in IQM for different temporal footprints.

The choice of whether to use frame overlap remains to be determined. Verma *et al.* [5] recommended as high a temporal resolution as possible for optimal curve fitting. While the results of applying the heuristic to the 50% overlap data resulted in smaller average absolute error across the K^{trans} values tested, it is not guaranteed that this will be the case in general. The reason for this is because the heuristic presented can only optimize temporal footprints over a fairly coarse step size of temporal footprints (0.5 sec). This, along with the noise that can be observed in the region of stability (Figure 4.5a) may cause the error to be greater than any additional accuracy provided by the additional data points from the overlapping frames. It could be possible to perform the heuristic with a smaller temporal footprint step size; however, there is a risk positively biasing the recovered K^{trans} values. Since the heuristic starts at a large temporal footprint where K^{trans} is overestimated, a smaller step size would cut off at a larger temporal footprint then a large step size and thus more of the positive bias would remain. A potentially useful approach would be to apply the heuristic to data reconstructed with no overlapping, then, once an appropriate temporal footprint is found, reconstruct the images at this temporal footprint with 50% overlap (or more). This could provide additional accuracy as well as reconstruction speed since only half as many images would need to be reconstructed during optimization.

The other PK parameters of the Tofts model have not been presented in the results here. This was decided in large part due to the popularity of K^{trans} among the PK parameters. It was also found during this work that the fitting algorithm was more sensitive to errors in K^{trans} than errors in v_e or v_p. This is likely due to the fact that K^{trans} directly affects both the scale and washout rate of C_t as seen in equation 4.1, whereas v_e and v_p depend mostly on one or the other. For instance, since v_p only affects the contribution of the AIF, which has rapid wash-in and wash-out rates, its effect to the overall signal is mostly present during the initial signal enhancement. More importantly, the small magnitude of v_p means it contributes less to the shape of the curve and minor deviations from the true scale of C_t can lead to large percent errors. The rate of contrast agent washout is controlled by the ratio of K^{trans} and v_e. This is the only contribution of v_e to C_t, so it is insensitive to changes to the initial signal enhancement, which is where a lot of the effects of the different temporal footprints manifest.

The SSIM values of the reconstructions at each iteration in Figure 4.7 vary over an exceedingly small range (approximately 0.01). This would correspond to an imperceptible difference in image quality to a human eye, however this change of image quality need not be detected by a human, only the IQM. IQMs are an appropriate tool for such small differences in image quality because, since many are designed for some definition of signal fidelity, they are monotonic relative to degradation strength on an individual reference image. In this study, there is an equivalence between temporal footprint and degradation strength because of the equivalence of temporal footprint and undersampling factor. Thus, one would expect to see a monotonic relationship between and IQM and temporal footprint, as was observed in Figure 4.4. Of course, since the degradations from radial undersampling are noise-like, the introduced degradations can experience constructive or destructive interference, so the relationship between temporal footprint and an objective IQM is not perfectly monotonic. The fluctuations in Figure 4.7 then amount to the "noise" in the SSIM values due to this phenomenon. The temporal footprint that this heuristic chooses then is simply the first temporal footprint where the SSIM is far enough away from the plateau that is can be distinguished from fluctuations due to this noise.

4.4.4 Future Works and Limitations

The clinical utility of this technique remains to be validated. The first step in accomplishing this would be exploring whether Figure 4.4 and Figure 4.5a and b are reproducible in clinical data. These two are required because in order for the heuristic to work, the image quality must reach a plateau at large temporal footprints and the K^{trans} must be overestimated due to low temporal resolution in C_t and the AIF. The next steps would be demonstrating the superiority of the K^{trans} maps generated from this technique compared to existing techniques. In this case, statistically superior would be defined as more accurate diagnoses when the radiologists had access to the DCE scan and associated K^{trans} maps from this technique than when provided traditional DCE data. Beyond this, K^{trans} maps can be a valuable input for deep learning based segmentation algorithms [78]. This is another potential application of this heuristic. However, it has also recently been shown by Lee *et al*. [79] that some information may be lost when the DCE images reduced to K^{trans} maps and that more accurate segmentation can be achieved using just DCE images alone compared to just K^{trans} maps alone, which calls into question the future of PK modelling of DCE data. However, more accurate segmentation was not found when other contrasts such as T2 weighted images and DWI were included.

While there are aspects of this study that required it to be done in simulations, such as the requirement to know the ground truth of K^{trans} exactly, this also results in some limitations to the potential applicability of the results. First, the imaging parameters such as flip angle, TR, and T_{10} were known exactly. This may have provided additional accuracy as these can be a source of error when converting between $S(t)$ and $C_t(t)$ [80]. Second, the Tofts model itself is only an approximation of the temporal dynamics of contrast agents in tissues. In these simulations, $C_t(t)$ was calculated with and then fit to the Tofts model. This removes any assumptions of the Tofts model such as linear intercompartmental flux of the contrast agent or time invariance of the PK parameters. This also adds additional accuracy to the simulations. Finally, even when using the simulated AIF, it and $C_t(t)$ are perfectly in sync, which is not realistic in practice, providing another potential source of accuracy.

4.5 Conclusions

The results presented in this chapter demonstrate a useful application of IQMs for golden angle stack-of-stars sampling with sliding window reconstruction MRI. The IQMs were calculated with a reference generated from the simulated data demonstrating that this technique could be applied in a clinical setting when no ground truth reference is known *a priori*. Using a heuristic based on the observation that the IQMs reach a plateau at large temporal footprints, the IQMs were used to guide the choice of temporal footprint in DCE-MRI to balance the requirements of temporal resolution and image quality to recover K^{trans} values within +/-10% of the true values of the simulated features. These results were achieved under various conditions including different simulated bases, with noise, and with multiple overlap factors. While successfully implemented in simulations, this technique has yet to be prospectively applied to clinical data and compared to standard clinical practice. This technique could provide a precise tool to allow for temporal footprint optimization in DCE MRI on an individual patient level.

Chapter 5: Conclusions

In this thesis, it was first explored how to properly quantify MR image quality for clinical utility and then a technique was developed and presented that use the quantitative measures of image quality in an advanced image reconstruction algorithm. This was performed in two separate studies. In the first study, the correlations between radiologists scores of MR image quality and 10 objective IQMs were measured. This was done with a specific focus on RMSE and SSIM, which are the two most popular metrics in the field of MRI. This was done because we hypothesized that other IQMs may demonstrate a higher correlation with the radiologists scores than these two IQMs and so may be a more appropriate metric when quantifying MR image quality in a research study. It was found that SSIM had the lowest SROCC with the radiologists score of the 10 IQMs included in the study when all images and radiologists scores were combined. RMSE tied for 7th highest SROCC. These results may have important implication in future MR reconstruction studies. SSIM is often used with or as opposed to RMSE because it was assumed that SSIM was a more accurate measure of image quality than RMSE, however the results presented here to not support that hypothesis. Another important observation made during this study was that it appeared as if the most important factor that affected an IQMs performance was how uniformly the IQM scored images of different degradation types that were deemed to be of equal quality by the radiologist. For instance, VIF, which had the highest SROCC in the combined data group, showed little variation in its score based on degradation type. However, with RMSE and SSIM, the scores of images of different degradation types are clearly discernible when plotted against radiologists score. Finally, it was found that some of the IQMs with the highest SROCCs (such as VIF, FSIM, and NQM) also had a longer computation time than RMSE and SSIM, which may be an issue for adoption if the IQM must be calculated for a large number of images. GMSD was found to offer a good trade-off between high correlation and low computational time, with the fourth highest SROCC and the third lowest computation time.

In the second study, IQMs were proposed as a tool that can address the issue of finding a balance in the trade-off between image quality and temporal resolution in DCE MRI. DCE scans with golden angle stack-of-stars sampling with sliding window reconstruction were performed in numerical simulations on a numerical phantom with known K^{trans} values. This allowed the relationships between properties of the reconstructed images such as K^{trans}, image quality, and temporal footprint size to be explored. A reference can be generated from the stack-of-stars data by reconstructing an image from a large number of rays. This allows for the use of full-reference IQMs like RMSE and SSIM. Within the stack-of-stars framework it was found that the IQMs reach a plateau for temporal footprints greater than approximately 8 seconds. It was also found that with both a population and simulated AIF, K^{trans} was systemically overestimated for large temporal footprints, but as the temporal footprint decreased there was a region of stability where the error of the recovered K^{trans} is minimized. The K^{trans} overestimation and the IQM plateau occurs over the same range of temporal footprints. In practice the ground truth K^{trans} is not known, but the IQMs can be calculated. Thus, a heuristic was developed for finding the temporal footprint before the IQM plateau, which is located in the region of stability for K^{trans} measurement. This heuristic was applied with the numerical phantom and synthetic dynamic clinical data and it was able to recover K^{trans} within 10% of its value for a range of physiologically relevant values. This technique potentially allows for optimization of temporal resolution in a DCE MRI scan on a patient-by-patient basis.

There are numerous avenues of research that could follow this work. For the first study, only images from the abdomen and brain were included. This means that the results may not be generalizable to other areas of the body such as musculoskeletal or cardiovascular images. Provided sufficient resources, expansion to these other areas would be useful for researchers working with these types of images. In this study, only full-reference IQMs were studied. While this was chosen because the study was focusing on how objective IQMs are used in MR reconstruction studies, no-reference IQMs could play an important role in clinical MRI, for example for the assessment of scan quality immediately following the scan. Correctly used, these could help reduce patient recall by detecting low quality scans while the patient is still in the scanner. A similar study to the one presented here with no-reference IQMs could help determine which of these metrics would be appropriate for this task. An important next step is determining the relationship between IQM score and lesion visibility for images presenting lesions. While there would be a relationship between diagnostic image quality as defined here and the visibility of a lesion, the two are not necessarily equivalent. A threshold for visibility (e.g. if SSIM > 0.9, lesion is visible 95% of time) would be a valuable diagnostic tool. For the second study, a natural next step is the expansion to clinical data. If the technique can be validated in a clinical setting it can be compared to standard of care DCE scans to test for statistical differences in diagnostic ability for a radiologist or a machine learning algorithm.

References

[1] R. Damadian, "Tumor Detection by Nuclear Magnetic Resonance," *Science.*, vol. 171, no. 3976, pp. 1151–1153, 1971.

[2] J. W. Belliveau *et al.*, "Functional Mapping of the Human Visual Cortex by Magnetic Resonance Imaging," *Science.*, vol. 254, pp. 716–719, 1991.

[3] L. Minati and W. P. Weglarz, "Physical Foundations, Resonance Imaging of Diffusion Magnetic Models, and Methods of the Brain: A Review," *Concepts Magn. Reson.*, vol. 30A, no. 5, pp. 278–307, 2007.

[4] D. E. Befroy and G. I. Shulman, "Magnetic resonance spectroscopy studies of human metabolism," *Diabetes*, vol. 60, no. 5, pp. 1361–1369, 2011.

[5] S. Verma *et al.*, "Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management," *Am. J. Roentgenol.*, vol. 198, no. 6, pp. 1277–1288, 2012.

[6] P. S. Tofts and A. G. Kermode, "Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts," *Magn. Reson. Med.*, vol. 17, no. 2, pp. 357–367, 1991.

[7] C. S. Restrepo and S. Tavakoli, "Contrast-Enhanced Cardiac Magnetic Resonance Imaging," *Magn. Reson. Imaging Clin. NA*, vol. 20, no. 4, pp. 739–760, 2012.

[8] L. Li *et al.*, "Parameters of Dynamic Contrast-Enhanced MRI as Imaging Markers for Angiogenesis and Proliferation in Human Breast Cancer," *Med. Sci. Monit.*, vol. 21, pp. 376–382, 2015.

[9] D. Wilson *et al.*, "Renal Cell Carcinoma Perfusion before and after Radiofrequency Ablation Measured with Dynamic Contrast Enhanced MRI: A Pilot Study," *Diagnostics*, vol. 8, no. 3, 2018.

[10] J. C. Weinreb *et al.*, "PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2," *Eur. Urol.*, vol. 69, no. 1, pp. 16–40, 2016.

[11] S. M. Noworolski, R. G. Henry, D. B. Vigneron, and J. Kurhanewicz, "Dynamic contrast-enhanced MRI in normal and abnormal prostate tissues as defined by biopsy, MRI, and 3D MRSI," *Magn. Reson. Med.*, vol. 53, no. 2, pp. 249–255, 2005.

[12] I. Ocak *et al.*, "Dynamic Contrast-Enhanced MRI of Prostate Cancer at 3 T : A Study of Pharmacokinetic Parameters," *Am. J. Roentgenol.*, vol. 189, no. October, pp. 192–201, 2007.

[13] R. H. . El Khouli, K. J. . Macura, I. R. . Kamel, M. A. . Jacobs, and D. A. . Bluemke, "3-T Dynamic Contrast-Enhanced MRI of the Breast: Pharmacokinetic Parameters Versus Conventional Kinetic Curve Analysis," *Am. J. Roentgenol.*, vol. 197, pp. 1498–1505, 2011.

[14] L. E. Kershaw, J. P. Logue, C. E. Hutchinson, N. W. Clarke, and D. L. Buckley, "Late tissue effects following radiotherapy and neoadjuvant hormone therapy of the prostate measured with quantitative magnetic resonance imaging," *Radiother. Oncol.*, vol. 88, no. 1, pp. 127–134, 2008.

[15] R. J. Hodgson, S. Connolly, T. Barnes, B. Eyes, R. S. D. Campbell, and R. Moots, "Pharmacokinetic modeling of dynamic contrast-enhanced MRI of the hand and wrist in rheumatoid arthritis and the response to anti-tumor necrosis factor-α therapy," *Magn. Reson. Med.*, vol. 58, no. 3, pp. 482–489, 2007.

[16] H. Kim, "Variability in Quantitative DCE-MRI : Sources and Solutions," *J. Nat. Sci.*, vol. 4, no. 1, p. e484, 2018.

[17] L. Feng *et al.*, "Golden-Angle Radial Sparse Parallel MRI : Combination of Compressed Sensing , Parallel Imaging , and Golden-Angle Radial Sampling for Fast and Flexible Dynamic Volumetric MRI," *Magn. Reson. Med.*, vol. 717, no. September 2013, pp. 707–717, 2014.

[18] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE : Sensitivity Encoding for Fast MRI," *Magn. Reson. Imaging*, vol. 42, pp. 952–962, 1999.

[19] M. A. Griswold *et al.*, "Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, no. 6, pp. 1202–1210, 2002.

[20] M. Lustig, D. Donoho, and J. M. . Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.

[21] O. N. Jaspan, R. Fleysher, and M. L. Lipton, "Compressed sensing MRI: A review of the clinical literature," *Br. J. Radiol.*, vol. 88, 2015.

[22] D. Liang, B. Liu, J. Wang, and L. Ying, "Accelerating SENSE using compressed sensing," *Magn. Reson. Med.*, vol. 62, no. 6, pp. 1574–1584, 2009.

[23] G. C. Kagadis, C. Kloukinas, K. Moore, J. Philbin, C. Alexakos, and P. G. Nagy, "Cloud computing in medical imaging," *Med. Phys.*, vol. 40, no. 7, 2013.

[24] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed Sensing using Generative Models," in *International Conference on Machine Learning*, 2017, p. 70.

[25] K. Hammernik, T. Klatzer, E. Kobler, D. K. . Recht, Michael P.; Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2018.

[26] Y. Yang, J. Sun, H. LI, and Z. Xu, "ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–18, 2018.

[27] 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,” *IEEE Trans. Med. Imaging*, vol. 37, no. 2, pp. 491–503, 2018.

[28] K. Kwon, D. Kim, and H. Park, “A parallel MR imaging method using multilayer perceptron:,” *Med. Phys.*, vol. 44, no. 12, pp. 6209–6224, 2017.

[29] S. Wang *et al.*, “Accelerating magnetic resonance imaging via deep learning,” in *IEEE Symposium on Biomedical Imaging*, 2016, vol. 13, pp. 514–517.

[30] C. Ulas *et al.*, “Direct estimation of pharmacokinetic parameters from DCE-MRI using deep CNN with forward physical model loss,” in *Medical Image Computing and Computer Assested Intervention*, 2018, vol. 21, pp. 39–47.

[31] M. Mardani *et al.*, “Deep generative adversarial neural networks for compressive sensing MRI,” *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 167–179, 2019.

[32] A. Keshavan, J. D. Yeatman, and A. Rokem, “Combining citizen science and deep learning to amplify expertise in neuroimaging,” *bioRxiv*, 2018.

[33] R. Kumar and M. Rattan, “Analysis Of Various Quality Metrics for Medical Image Processing,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 2, no. 11, pp. 137–144, 2012.

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[35] Y. Ding, S. Suffren, P. Bellec, and G. A. Lodygensky, “Supervised machine learning quality control for magnetic resonance artifacts in neonatal data sets,” *Hum. Brain Mapp.*, vol. 40, no. 4, pp. 1290–1297, 2019.

[36] M. S. Graham, I. Drobnjak, and H. Zhang, “A supervised learning approach for diffusion MRI quality control with minimal training data,” *Neuroimage*, vol. 178, no. May, pp. 668–676, 2018.

[37] A. Sreekumari *et al.*, “Deep learning based MR image diagnostic quality deduction to reduce patient recall,” in *Proc. Intl. Soc. Mag. Reson. Med.*, 2018, p. 2747.

[38] T. Küstner *et al.*, “A machine-learning framework for automatic reference-free quality assessment in MRI,” *Magn. Reson. Imaging*, vol. 53, no. November 2017, pp. 134–147, 2018.

[39] N. Murtha, J. Rioux, A. Marriott, C. Bowen, S. Clarke, and S. Beyea, “Simulation Reveals Evidence for Bias in Parameter Estimates for Compressed Sensing of Temporally Dynamic Systems,” in *Proc. Intl. Soc. Mag. Reson. Med.*, 2017, p. 3810.

[40] M. Saranathan, D. W. Rettmann, B. A. Hargreaves, S. E. Clarke, and S. S. Vasanawala, “DIfferential subsampling with cartesian ordering (DISCO): A high spatio-temporal resolution dixon imaging sequence for multiphasic contrast enhanced abdominal imaging,” *J. Magn. Reson. Imaging*, vol. 35, no. 6, pp. 1484–1492, 2012.

[41] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, “An Optimal Radial Profile Order baed on the Golden Ratio for Time-Resolved,” *IEEE Trans. Med. Imaging*, vol. 26, no. 1, pp. 68–76, 2007.

[42] M. Uecker *et al.*, “Berkeley Advanced Reconstruction Toolbox,” in *Proc. Intl. Soc. Mag. Reson. Med.*, 2015, vol. 23, no. 2486.

[43] S. G. Lingala and M. Jacob, “Blind Compressive Sensing Dynamic MRI,” *Med. Imaging, IEEE Trans.*, vol. 32, no. 6, pp. 1132–1145, 2013.

[44] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, “XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing,” *Magn. Reson. Med.*, vol. 75, no. 2, pp. 775–788, 2016.

[45] P. S. Tofts, “Modeling Tracer Kinetics in Dynamic Gd-DTPA MR Imaing,” *J. Magn. Reson. Imaging*, vol. 7, pp. 91–101, 1997.

[46] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-Scale Structural Similarity for Image Quality Assessment,” in *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems, and Computers*, 2003, vol. 2, pp. 9–13.

[47] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, 2011.

[48] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 668–695, 2014.

[49] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A Feature Similarity Index for Image Quality Assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.

[50] P. Kovesi, “Image Features From Phase Congruency Image,” in *University of Western Australia, Technical Report 95/4*, 1995.

[51] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, “Image quality assessment based on a degradation model,” *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, 2000.

[52] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, 2006.

[53] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, “On Advances in Statistical Modeling of Natural Images,” *J. Math. Imaging Vis.*, vol. 18, no. 1, pp. 17–33, 2003.

[54] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, “HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions,” *ACM Trans. Graph.*, vol. 30, no. 4, p. 1:14, 2011.

[55] B. Kim *et al.*, “Prediction of Perceptible Artifacts in JPEG2000 Compressed Abdomen CT Images Using a Perceptual Image Quality Metric,” *Acad. Radiol.*, vol. 15, no. 3, pp. 314–325, 2008.

[56] X. He and S. Park, “Model observers in medical imaging research,” *Theranostics*, vol. 3, no. 10, pp. 774–786, 2013.

[57] S. Kojima, Shinya; Shinohara, Hiroyuki; Hashimoto, Takeyuki; Suzuki, “Undersampling patterns in k-space for compressed sensing MRI using two-dimensional Cartesian sampling,” *Radiol. Phys. Technol.*, vol. 11, no. 3, pp. 303–319, 2018.

[58] H. Zheng *et al.*, “Multi-Contrast Brain MRI Image Super-Resolution With Gradient-Guided Edge Enhancement,” *IEEE Access*, vol. 6, pp. 57856–57867, 2018.

[59] V. Gözcü, Baran; Mahabadi, Rabeeh Karimi; Li, Yen Huan; Ilicak, Efe; Çukur, Tolga; Scarlett, Jonathan; Cevher, “Learning-Based Compressive MRI,” *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1394–1406, 2018.

[60] B. A. Duffy, “Retrospective correction of motion artifact affected structural MRI images using deep learning of simulated motion,” *Med. Imaging with Deep Learn.*, no. Midl 2018, pp. 1–8, 2018.

[61] H. Jeelani, J. Martin, F. Vasquez, M. Salerno, and D. S. Weller, “Image Quality Affects Deep Learning Reconstruction of MRI,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI*, 2018, pp. 357–360.

[62] A. S. Chaudhari *et al.*, “Super-resolution musculoskeletal MRI using deep learning,” *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2139–2154, 2018.

[63] R. B. Hansen *et al.*, “Multichannel Hyperpolarized 13C MRI in a Patient with Liver Metastases using Multi- slice EPI and an Alternating Projection Method for Denoising,” in *ISMRM-ESMRMB 2018*, 2018, pp. 3–6.

[64] T. Akasaka *et al.*, “Optimization of regularization parameters in compressed sensing of magnetic resonance angiography: Can statistical image metrics mimic radiologists’ perception?,” *PLoS One*, vol. 11, no. 1, pp. 1–14, 2016.

[65] H. Rajagopal, L. S. Chow, and R. Paramesran, “Subjective Versus Objective Assessment for Magnetic Resonance Images,” *ICCITE 2015 17th Int. Conf. Commun. Inf. Technol. Eng.*, vol. 9, no. 12, pp. 106–111, 2015.

[66] L. S. Chow, H. Rajagopal, and R. Paramesran, “Correlation between subjective and objective assessment of magnetic resonance (MR) images,” *Magn. Reson. Imaging*, vol. 34, no. 6, pp. 820–831, 2016.

[67] ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” 2009.

[68] G. P. Renieblas *et al.*, “Structural similarity index family for image quality assessment in radiological images assessment in radiological images,” *J. Med. Imaging*, vol. 4, no. 3, 2017.

[69] L. S. Chow and R. Paramesran, “Review of medical image quality assessment,” *Biomed. Signal Process. Control*, vol. 27, pp. 145–154, 2016.

[70] S. Braun, C. Xiao, B. Odry, B. Mailhe, and M. Nadar, “Motion Detection and Quality Assessment of MR images with Deep Convolutional DenseNets,” in *Proc. Intl. Soc. Mag. Reson. Med.*, 2018.

[71] J. Liu and D. Saloner, “Accelerated MRI with CIRcular Cartesian UnderSampling (CIRCUS): a variable density Cartesian sampling strategy for compressed sensing and parallel imaging,” *Quant. Imaging Med. Surg.*, vol. 4, no. 1, pp. 57–67, 2014.

[72] J. Cohen, “Weighted Kappa: Nominal Scale Agreement With Provision for Scaled Disagreement of Partial Credit,” *Psychol. Bull.*, vol. 70, no. 4, pp. 213–220, 1968.

[73] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A Statistical Evalution of Recent Full Reference Image Quality Assessment Algorithms,” *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3441–3452, 2006.

[74] Y. Benjamini and Y. Hochberg, “Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing,” *R. Stat. Soc.*, vol. 57, no. 1, pp. 289–300, 2019.

[75] S. G. Kim *et al.*, “Influence of Temporal Regularization and Radial Undersampling Factor on Compressed Sensing Reconstruction in Dynamic Contrast Enhanced MRI of the Breast,” *J. Magn. Reson. Imaging*, vol. 0, no. 0, pp. 1–9, 2015.

[76] G. Parker *et al.*, “Experimentally-Derived Functional Form for a Input Function for Dynamic Contrast-Enhanced MRI,” *Magn. Reson. Med.*, vol. 1000, no. July, pp. 993–1000, 2006.

[77] S. R. Barnes, T. S. C. 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, no. 1, 2015.

[78] S. Liu, H. Zheng, Y. Feng, and W. Li, “Prostate Cancer Diagnosis using Deep Learning with 3D Multiparametric MRI,” *arXiv*, pp. 3–6, 2017.

[79] P. Q. Lee *et al.*, “Model-free prostate cancer segmentation from dynamic contrast-enhanced MRI with recurrent convolutional networks : A feasibility study,” *Comput. Med. Imaging Graph.*, vol. 75, pp. 14–23, 2019.

[80] W. Ping, Y. Xue, X. Zhao, J. Yu, M. Rosen, and H. Kwon Song, “Effects of flip angle uncertainty and noise on the accuracy of DCE-MRI metrics : comparison between standard concentration-based and signal difference methods,” *Magn. Reson. Imaging*, vol. 33, no. 1, pp. 166–173, 2015.