



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

FAKULTÄT FÜR
ELEKTROTECHNIK UND
INFORMATIONSTECHNIK

**Incorporation of Prior Knowledge
into Dynamic MRI Reconstruction**

Dissertation

zur Erlangung des akademische Grades

Doktoringenieur

(Dr.-Ing.)

von M.Eng. Chompunuch Sarasaen

geb. am 15. Juli 1988 in Bangkok, Thailand

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg

Gutachter

Prof. Dr. rer. nat. Georg Rose (Otto-von-Guericke-Universität Magdeburg)

Prof. Dr. rer. nat. habil. Oliver Speck (Otto-von-Guericke-Universität Magdeburg)

Promotionskolloquium am Januar 5, 2024

Everyone is a genius. But if you judge a fish on its ability to climb a tree, it will live its whole life believing that it is stupid.

- NOT ALBERT EINSTEIN

Abstract

Magnetic Resonance Imaging (MRI) is a non-ionizing modality providing superior soft tissue contrast and high-resolution cross-sectional images, which is beneficial for image guidance in minimally invasive interventions. Dynamic MRI is a fast MR imaging technique used for assessing physiological changes during interventions. Nevertheless, to obtain high temporal resolution for dynamic MRI, the compromise between spatial resolution and temporal resolution, known as a spatio-temporal trade-off, is unavoidable. State-of-the-art fast imaging employs various methods. However, most available methods utilize iterative computations, making them unsuitable for interventions. Along with the advanced deep-learning techniques, super-resolution (SR) algorithms are promising for restoring high-resolution images with a fast inference speed. Moreover, prior knowledge, which refers to the information available before the current study, such as high-resolution planning scans or temporal redundancy in dynamic images, could help achieve MR high-resolution images but are typically neglected.

Therefore, this thesis presents a method to alleviate the trade-off by incorporating prior knowledge with SR into dynamic MRI reconstruction to generate high spatial and high temporal resolution with less-than-complete data. The study investigates the incorporation of prior information into two deep learning based SR frameworks: Fine-tuned SR DynMRI and DDoS-UNet DynMRI. The Fine-tuned SR DynMRI makes use of a high-resolution planning scan as prior-knowledge for fine-tuning the patched-based SR reconstruction. Furthermore, the DDoS-UNet DynMRI adds additional temporal information to the SR network by employing dual-channel training of static and dynamic abdominal MR images at different time-points. The performance of both proposed methods was evaluated with different in-plane undersampled levels. The investigated results show that the Fine-tuned SR DynMRI approach achieve the average SSIM value of the highest undersampling (6.25% of the k-space) before and after fine-tuning, which are 0.939 ± 0.008 and 0.957 ± 0.006 , respectively. The DDoS-UNet DynMRI approach shows the ability to reconstruct even higher undersampling of 4% of the k- space with the average SSIM of 0.951 ± 0.017 . Finally, this thesis demonstrates the utilization of prior knowledge for reconstructing dynamic MRI using the proposed SR frameworks, which can tackle the spatio-temporal trade-off in MRI.

Zusammenfassung

Magnetresonanztomographie (MRT) ist eine nichtionisierende Modalität, die einen überlegenen Weichteilkontrast und hochauflösende Schnittbilder liefert, was für die Bildführung bei minimal-invasiven Eingriffen von Vorteil ist. Dynamische MRT ist eine schnelle MR-Bildgebungstechnik, die zur Beurteilung physiologischer Veränderungen während interventioneller Eingriffe verwendet wird. Um eine hohe zeitliche Auflösung in dynamischen MRT-Aufnahmen zu erhalten, ist der Zielkonflikt zwischen räumlicher Auflösung und zeitlicher Auflösung, der als räumlich-zeitlicher Kompromiss bekannt ist, unvermeidlich. Die moderne schnelle Bildgebung umfasst verschiedene Methoden, von denen die meisten jedoch iterative Berechnungen verwenden, was sie daher für Interventionen ungeeignet macht. Zusammen mit fortschrittlichen Deep-Learning-Techniken sind Superauflösungsalgorithmen (SR) vielversprechend für die Rekonstruktion hochauflösender Bilder mit schneller Inferenzgeschwindigkeit. Zudem kann Vorwissen wie etwa hochauflösende Planungsscans oder zeitliche Redundanz in dynamischen Bildern in der Rekonstruktion berücksichtigt werden.

Ziel dieser Dissertation ist, den Zielkonflikt zu mildern, indem Vorwissen zusammen mit SR in die dynamische MRT-Rekonstruktion integriert wird, um eine hohe räumliche und hohe zeitliche Auflösung mit unter-abgetasteten Daten zu erzeugen. Die Arbeit untersucht die Einbeziehung von Vorinformationen in zwei auf Deep Learning basierende SR-Frameworks: Fine-tuned SR DynMRI und DDoS-UNet DynMRI. Das Fine-tuned SR DynMRI nutzt einen hochauflösenden Planungsscan als Vorwissen zur Feintuning der Patch-basierten SR-Rekonstruktion. Darüber hinaus zielt das DDoS-UNet DynMRI darauf ab, die zusätzlichen zeitlichen Informationen in das SR-Netzwerk einzubeziehen, indem es ein zweikanaliges Training von statischen und dynamischen Bildern der verschiedenen Zeitpunkte einsetzt. Die Leistung beider vorgeschlagenen Verfahren wurde mit unterschiedlichen dem Feintuning bewertet. Die untersuchten Ergebnisse zeigten, dass der Fine-tuned SR DynMRI-Ansatz den durchschnittlichen SSIM-Wert der höchsten Unterabtastung (6,25% des k-Raums) vor und nach der Feinabstimmung von $0,939 \pm 0,008$ bzw. $0,957 \pm 0,006$ erreichte. Der DDoS-UNet DynMRI-Ansatz zeigte die Fähigkeit, noch höheres Undersampling von 4% des k-Raums mit der durchschnittlichen SSIM von $0,951 \pm 0,017$ zu rekonstruieren. Zusammenfassend demonstriert diese Dissertation die Nutzung von Vorwissen zur Rekonstruktion dynamischer MRT unter Verwendung der vorgeschlagenen SR-Frameworks, welches eine Erhöhung der räumlichen und zeitlichen Auflösung erlaubt.

Contents

Abstract	iv
Zusammenfassung	vi
List of Acronyms	xi
List of Figures	xiii
List of Tables	xviii
1 Introduction	1
1.1 Motivation	2
1.2 Scientific Contributions	3
1.3 Thesis Outline	4
2 Background	7
2.1 Principles of Magnetic Resonance Imaging	7
2.1.1 Proton Nuclei	7
2.1.2 Resonance and Relaxation	9
2.1.3 MR Image contrast: Tissue Characterization	11
2.1.4 MR Hardware Components	14
2.1.5 Data Acquisition Techniques, k-space, and Image Reconstruction . .	15
2.1.6 Data Acquisition Technique	16
2.1.7 k-space	16
2.1.8 Image Reconstruction	17
2.1.9 Spatial Encoding and Gradient Echo Pulse Sequence	19
2.1.10 Image Quality Overview	19
2.2 Principles of Neural Networks	22
2.2.1 Modeling the Artificial Neurons	22
2.2.2 A Perceptron	23
2.2.3 Multi-layer Perceptrons	24
2.2.4 Forward Propagation	24
2.2.5 Backpropagation	25
2.2.6 Loss Function	25

2.2.7	Gradient Descent	25
2.2.8	Stochastic Gradient Descent	25
2.2.9	Convolutional Neural Networks	27
2.2.10	Feature Detection	28
2.2.11	Classification	30
2.2.12	Deep Learning and Applications	30
2.2.13	Types of Learning Paradigms	32
2.3	Dynamic MRI and State-of-the-art	33
2.3.1	Dynamic MRI	33
2.3.2	Common Imaging Sequence in Abdomen	35
2.3.3	Spatio-Temporal Trade-off	36
2.3.4	State-of-the-Art Dynamic MRI Reconstruction	37
3	Fine-tuned SR DynMRI	43
3.1	Synopsis	43
3.2	Related Work	44
3.3	Methodology	46
3.3.1	Super-resolution Reconstruction	46
3.3.2	Problem Statement of Fine-tuning	47
3.3.3	Experimental Datasets	47
3.3.4	Undersampled Data Generation	49
3.3.5	Fine-tuned SR DynMRI Framework and Network Architecture	51
3.3.6	Implementation Details	54
3.3.7	Image Reconstruction Evaluation	54
3.4	Results	54
3.5	Discussion	61
3.6	Conclusions	62
4	DDoS-UNet DynMRI	64
4.1	Synopsis	64
4.2	Related Work	65
4.3	Methodology	66
4.3.1	Experimental Dataset	66
4.3.2	Dynamic Data Generation	68
4.3.3	Undersampled Data Generation	68
4.3.4	DDoS-UNet DynMRI Framework	69
4.3.5	Network Architecture	72
4.3.6	Implementation Details	73
4.3.7	Image Reconstruction Evaluation	74
4.4	Results	74

4.5	Discussion	84
4.6	Conclusions	85
5	Summary and Outlook	88
5.1	Summary	88
5.2	Outlook	90
	Appendix A Additional Data	93
	Appendix B Image Quality Assessment	96
	Appendix C GitHub Repository and Colab Notebook	98
	Bibliography	100
	List of Publications	123

List of Acronyms

ADC analog-to-digital converter

AI Artificial Intelligence

ANN Artificial Neural Networks

B0 static magnetic field ***B0***

CNNs Convolutional Neural Networks

CNR contrast-to-noise ratio

CS Compressed Sensing

CSF cerebrospinal fluid

CT Computed Tomography

DCE-MRI dynamic contrast-enhanced magnetic resonance imaging

DDoS-UNet DynMRI Improved Deep Learning Based Super-resolution
Reconstruction of Dynamic MRI by Incorporating Temporal Information in
Dynamic Dual-channel UNet

DL Deep Learning

DNNs Deep Neural Networks

FID free-induction-decay

FLASH Fast Low Angle Short

FOV FOV Field of View

FSE fast spin echo

Fine-tuned SR DynMRI Improved Deep Learning Based Super-resolution
Reconstruction of Dynamic MRI by Fine-tuning with a Subject-specific Scan

GRAPPA GeneRalized Autocalibrating Partial Parallel Acquisition

GRE Gradient Recalled Echo

- iMRI** interventional MRI
- LAVA- Flex** Liver Acquisition with Volume Acceleration Flex
- MRCP** Magnetic Resonance Cholangiopancreatography
- MRI** Magnetic Resonance Imaging
- NMR** Nuclear Magnetic Resonance
- PBSR** Patch-based super-resolution
- PI** Parallel imaging
- RF** radio frequency
- SGD** Stochastic gradient descent
- SNR** signal-to-noise ratio
- SPGR** Spoiled Gradient Recalled Echo
- SR** Super-Resolution
- SR-MRI** Super-resolution MRI
- T1** longitudinal relaxation T_1
- T1-FFE** T1-weighted Fast Field Echo
- T2** transversal relaxation T_2
- TE** echo time
- THRIVE** T1-weighted High-resolution Isotropic Volume Examination
- TR** repetition time
- VIBE** Volumetric Interpolated Breath-hold Examination

List of Figures

1.1	An outline of the thesis.	5
2.1	A graphic of anatomical planes: (a) transverse, (b) coronal and (c) sagittal planes.	8
2.2	A cross-section abdominal dissection [26] and a MRI; $T1$ -weighted image. For a MRI, different internal organs are illustrated in different shades of gray. As can be seen in the image, liver is brighter than kidneys because of the higher signal intensity. Abdominal fat is the most brightest and air is the most darkest in the image.	8
2.3	Graphic description of Proton alignment. (a) Normal state (b) With a presence of external magnetic field \mathbf{B}_0 , and (c) Energy differences (Zeeman diagram) [27].	9
2.4	Graphical abstract of $T1$ relaxation. The figure portrays the $T1$ magnetization recovery after 90° pulse applying at the equilibrium. The magnetization \mathbf{M}_z is initially dropped to zero and is then recovered to the equilibrium exponentially over time. The time constant of magnetization recovery to 63% of the original value \mathbf{M}_0 denotes as $T1$	12
2.5	Graphical abstract of $T2$ relaxation time. The figure shows the FID signal decaying after 90° pulse applying at the equilibrium. Then, the magnetization decays to 37% of the original value \mathbf{M}_0 and continue decaying over time. The time constant of magnetization decaying to 37% \mathbf{M}_0 denotes as $T2$. $T2^*$ time is shorter than $T2$ time due to the effect of magnetic field inhomogeneity.	12
2.6	An example of MR image contrasts. From left to right: $T1w$, $T2w$ and PD images or brain MRIs [32].	13
2.7	$T1$ relaxation time of the different tissue.	13
2.8	$T2$ relaxation time of the different tissue.	14
2.9	An illustration of the main parts of MR hardware components. Left: MRI machine outside, and right: inside components of system	14
2.10	The production of MRI includes the step of image acquisition and image reconstruction through k-space.	16

2.11	An illustration of each k-space data as a complex number in real and imaginary planes or magnitude and phase components as an alternative. Note that $Mag = \sqrt{Re^2 + Im^2}$ and $\Phi = \tan^{-1}[Im/Re]$	17
2.12	k-space and the corresponding reconstructed images. (a) Fully-sampled k-space, (b) Only low spatial frequency of the k-space data is acquired, as a results, the reconstructed image has contrast information but no details, and (c) Only high spatial frequency of the k-space data is acquired, the reconstructed image depicts only the edge of anatomical structures [25]. . .	18
2.13	A diagram of gradient echo pulse sequence.	20
2.14	An example of good and poor MR image quality. A good quality of MRI is in the acceptable range which has a good SNR. In contrast, a poor quality of MRI as a result of low basic resolution (blurry image) or too high basic resolution (grainy image).	21
2.15	An illustration of physical structures of neurons. Upper: Biological neuron and lower: Artificial neuron.	23
2.16	Multi-layer Perceptrons.	24
2.17	A graphical representation of how gradient descent works [49].	26
2.18	An illustration of stochastic gradient descent with global and local minima errors.	26
2.19	The original duck-and-rabbit illusion from Fliegende Blätter 1982, captioned "Welche Tiere gleichen einander am meisten?" [53]	27
2.20	A simple example of feature representations of a cat.	28
2.21	An example of CNNs [54]. The CNN architecture is built with several neural network layers, feature learning, and classification parts.	28
2.22	An example of a 2D convolution operation [49].	29
2.23	Flattening step.	30
2.24	An example of cat or dog classification task after full connection layer. . .	31
2.25	The divisions of artificial intelligence, machine learning and deep learning.	31
2.26	A comparison of the lesion visibility on MRI to ultrasound and CT, in different wash-in and wash-out phases. a-d: ultrasound, e-i: MRI and j-m: CT images. The arrows were marked the liver lesion at the same locations, together with (h) the specimen after operation [92].	34
2.27	An example of MR-guided puncture in interventional radiology.	34
2.28	An example of gradient echo images based on the Dixon method: in-phase, oppose-phase (out-phase), water-only and fat-only images.	36
2.29	A pictorial illustration of nearest neighbor interpolation from 2x2 matrix to 4x4 matrix. This interpolation method is the simplest interpolation that upsample the missing data from the neighbor's pixel values without adding new values.	39

2.30	A graphical illustration of SRCNN architecture, consisting only three parts: patch extraction, nonlinear mapping and reconstructions [152].	40
3.1	CHAOS dataset T1-dual images. (a) T1w in-phase image and (b) T1w opposed-phase image.	48
3.2	Fine-tuning data. An example of subject-specific high-resolution image (breath-hold) for different slices.	49
3.3	Graphical representation of masks and the corresponding reconstructed images. The undersampled (us) masks were generated by taking only 25%, 10% and 6.25% of the k-space center. The data points in black denote the sampled point, and the white dots denote the undersampled data.	51
3.4	Method Overview. The main steps in Fine-tuned SR DynMRI Framework: Main training, fine-tuning and inference.	52
3.5	The proposed network architecture based on 3D U-Net combined with perceptual loss network. The super-resolution training was performed using 3D patches (24^3 size) as the input. The size of patch is provided on the left side to each block (feature map) of each layer. The perceptual loss network consists of three levels of the contraction path that was pre-trained on MRA.	52
3.6	A qualitative comparison of low-resolution dynamic images (undersampling 25%, 10% and 6.25% of k-space center) on the same slice. Images at lower resolutions (scaled up), trilinear interpolated input, super-resolution results of main training (SR Results Main Training), Fine-tuned SR DynMRI, and their ground-truth images are shown from left to right.	56
3.7	Another qualitative comparison using SSIM Maps of the lowest resolution input investigated in this work (6.25% of k-space) with the reconstructed results of Fine-tuned SR DynMRI over different time-points. The corresponding SSIM outputs are shown on the top right of the image.	58
3.8	An example from the reconstructed results compared against its ground-truth of low-resolution images from 6.25% of k-space. Upper figure from left to right: ground-truth, trilinear interpolation (input of the network), SR main training, and Fine-tuned SR DynMRI. Lower figures show the yellow and red and ROIs of reconstructed results and the corresponding difference images from ground-truth of (a and b) trilinear interpolation, (c and d) SR main training, and (e and f) Fine-tuned SR DynMRI.	59

3.9	Line plot showing the mean and 95% confidence interval of the resultant SSIM and PSNR over the different time-points at each subject. The upper row shows the SSIM values and the lower row shows the PSNR values. The blue, orange, green, red and lilac colors represent the reconstruction results of trilinear interpolation, area interpolation, zero-padding (sinc interpolation), SR main training, and SR after fine-tuning respectively.	60
4.1	Flowchart of dynamic data generation. In this work, $n = 24$ and the total number of TP is 25 time-points.	68
4.2	Method overview of the training in DDoS-UNet DynMRI framework. Initially, random elastic deformation is applied to the CHAOS dataset (fully sampled) to generate the artificial CHAOS dynamic dataset. Then the CHAOS dynamic dataset was undersampled to generate the final training dataset. Then the model is trained by providing a low-resolution (undersampled) current time-point ($LRTP_n$) along with the high-resolution (fully sampled) previous time-point ($H RTP_{n-1}$) as input, and the output is compared against the ground-truth high-resolution current time-point ($H RTP_n$).	70
4.3	Method overview of the inference in DDoS-UNet DynMRI framework. 3D static subject-specific planning scan (fully sampled) is supplied as the high-resolution prior image (HR_{Prior}) along with the first low-resolution (undersampled) time-point (LR_{TP0}) of the 3D dynamic dataset are supplied as input to the trained DDoS-UNet model, and the model super-resolves LR_{TP0} to obtain SR_{TP0} . This initial phase is called the "Antipasto" phase. SR_{TP0} is then supplied as input together with the next low-resolution time-point LR_{TP1} to the same trained DDoS-UNet model to obtain SR_{TP1} . This process is continued repeatedly until all the time-points of the low-resolution (undersampled) 3D dynamic are super-resolved by supplying pairs of $SR_{TP_{n-1}}$ and LR_{TP_n} to obtain each of the SR_{TP_n}	71
4.4	Network architecture of the modified 3D UNet with perceptual loss network. Unlike the patch-based super-resolution (see figure 3.5), the training was performed using the whole 3D volumes of I_{Prior} and I_{LR} as dual inputs to the network.	73
4.5	Comparative results of low-resolution data (undersampled 10%, 6.25% and 4% of k-space) of the same slice. From left to right: low-resolution images (scaled-up), zero-padding, interpolated input (Trilinear interpolation), SR result of the UNet baseline trained with CHAOS dynamic (UNet CHAOS Dynamic), DDoS-UNet DynMRI and its corresponding ground-truth images.	77

4.6	An example comparison of the low-resolution images of the 4% of k-space with reconstructed results of the DDoS-UNet DynMRI, compared against the high-resolution ground-truth using SSIM maps. Four different time-points are shown here, including the Antipasto phase (initial time point TP_0). The corresponding SSIM outputs are displayed on the top right of the image.	78
4.7	An example from the reconstructed results compared against its ground-truth (GT) of low-resolution images from 4% of k-space. Upper figure from left to right: ground-truth, trilinear interpolation (interpolated input), SR result of the UNet baseline trained with CHAOS dynamic (UNet CHAOS Dynamic) and the DDoS-UNet DynMRI. Lower figures show the yellow and green and ROIs of reconstructed results and the corresponding difference images from GT of (a and b) trilinear interpolation, (c and d) SR result of UNet CHAOS Dynamic, and (e and f) DDoS-UNet DynMRI.	79
4.8	Comparison of baseline methods against the proposed DDoS-UNet DynMRI of all subjects for different undersampling levels. The box plots show SSIM (left column) and PSNR (right column) values of (a) 10% of the k-space (b) 6.25% of the k-space, and (c) 4% of the k-space.	80
4.9	Subject-wise line plots show the average SSIM (left column) and PSNR (right column) values of all time-points for different undersampling levels (a) 10% of the k-space (b) 6.25% of the k-space, and (c) 4% of the k-space. The plots also portrays mean and 95% confidence interval of the resultant values over the different time-points, excluding the initial one from the Antipasto phase. The red, orange, blue and green lines represent the reconstruction results of DDoS-UNet DynMRI, UNet trained with CHAOS Dynamic dataset, zero-padding (sinc interpolation), and trilinear interpolation, respectively.	81
4.10	Line plots show the average SSIM (left column) and PSNR (right column) values over time-points, of all subjects for different undersampling levels (a) 10% of the k-space (b) 6.25% of the k-space, and (c) 4% of the k-space. The red, orange, blue and green lines represent the reconstruction results of DDoS-UNet DynMRI, UNet trained with CHAOS Dynamic dataset, zero-padding (sinc interpolation), and trilinear interpolation, respectively. .	82
4.11	A qualitative results, compared the previously-proposed method [70] (Fine-tuned SR DynMRI) against the proposed DDoS-UNet DynMRI. The reconstructed results obtained from undersampled 4% of the k-space center. .	84

A.1	Box plots of SSIM and PSNR of SR result of UNet baselines trained with CHAOS dataset compared with CHAOS dynamic dataset while super-resolving of 4% of the k-space center.	94
A.2	A comparison results of UNet baselines trained with CHAOS dataset, compared against UNet baselines trained with CHAOS dynamic dataset and the proposed DDoS-UNet DynMRI. The results are reconstructed from 4% of the k-space center.	95

List of Tables

3.1	The acquisition protocols of the experimental dataset including CHAOS dataset of 40 subjects (total 80 volumes of in- and opposed-phase images), and subject-wise 3D dynamic scans of three subjects for ten time-points (TPs). Note that the dynamic and static scans were acquired using the same sequence at a different scan session but the latter was acquired for one TP.	50
3.2	Effective resolutions and estimated acquisition times (per time-point) for different levels of undersampling (25%, 10% and 6.25% of the k-space). . .	50
3.3	Quantitative results of the proposed framework compared with baseline methods in terms of SSIM, PSNR and standard deviation of the difference images with ground-truth (average values \pm SD). The p-values $<$ 0.001 for all subjects, resolution-wise and for all resolution with subject-wise. . . .	57
4.1	The acquisition protocol of the experimented datasets: CHAOS dataset of 40 subjects and subject-wise 3D dynamic scans of five subjects for 25 time-points (TPs). It should be noted that the static scans were acquired using the same sequence as the dynamic scans, but only for one TP, acquired in a different session.	67
4.2	Effective resolutions and estimated acquisition times (per time-point) of the dynamic and static datasets after performing different levels of artificial undersampling.	69
4.3	The resultant values of SSIM, PSNR, and NRMSE (average values \pm SD). The table shows the comparison results for all different resolutions. The p-values were less than 0.0001 for all scenarios.	83

1 Introduction

“I have not failed. I’ve just found 10,000 ways that won’t work.”

– Thomas Alva Edison

Magnetic Resonance Imaging (MRI) is an application of Nuclear Magnetic Resonance (NMR) which has several clinical applications. Unlike other medical imaging modalities, MRI has the advantages of multi-contrast imaging, non-ionizing radiation, and being non-invasive, making it a preferred imaging modality. MR-guided interventions, such as liver biopsy or radiofrequency ablation (RFA), allow physicians to examine internal organs with high spatial resolution in real-time or near real-time. In fact, the advance of live-guided intervention has evolved over the past few decades. The accuracy of its diagnostic or therapeutic use specifically depends on the image resolution during each examination, thus, achieving a clear visualization between target lesions and adjacent tissues is crucial.

The dynamic imaging technique is essential for interventions. It is typically obtained by acquiring the k-space data (in the frequency domain) and reconstructing a sequence of images over time, which obey the Nyquist theorem [1]. Acceleration of the imaging process results in less data being acquired. A trade-off between high spatial resolution and temporal resolution is the consequence. The commonly used technique to accelerate the data acquisition of MR images is compressed sensing (CS) [2–4]. CS technique utilizes the concept of sparsity by using fewer data to reconstruct the fully sampled image. Typically, it is assumed that the image is sparse in a transform domain; nonetheless, the corresponding iterative reconstruction is relatively slow and cannot be applied to real-time applications such as intervention MRI. Parallel imaging (PI) [5, 6] techniques, such as generalized simultaneous acquisition of spatial harmonics (SMASH) [7], sensitivity encoding (SENSE) [8], or generalized autocalibrating partial parallel acquisitions (GRAPPA) [9] are also promising techniques used to speed up image acquisition. PI technique can reduce the acquisition time by combining additional spatial information from multiple receiver coils. However, regarding the limitation of devices and accessibility during interventions, PI may not be suitable for this application.

Deep learning based super-resolution reconstruction has been shown to be a valuable tool for super-resolution of MRI [10, 11], including for dynamic MRI [12, 13]. Moreover,

some literature [14–16] have investigated the use of temporal information incorporation and reported the possibility to improve the reconstruction quality of dynamic MRI. To strengthen the quality of super-resolved images, prior-knowledge has been included in super-resolution methods [17, 18]. Moreover, the advantage of multi-channel learning allows the neural network to learn more from feature extractions by using multiple types of data [19]. Harnessing different information in combined training, multi-channel training has been used across numerous applications, including image recognition [20], speech recognition [21, 22], and natural language processing [23]. Therefore, the focus of this thesis is the reconstruction of high spatial and temporal resolution images for sparse or undersampled (less-than-complete) data tackling the spatio-temporal trade-off in MRI. The incomplete reconstruction problem is dealt with by the inclusion of prior-knowledge about the object into the proposed frameworks. The application to abdominal MRI is presented in this work. The motivations behind the research investigation are described in section 1.1, the scientific contributions are explained in section 1.2, and the thesis outline is located in section 1.3.

1.1 Motivation

MR data acquisition is relatively slow in nature. Typically, the k-space data (in the Fourier domain) is spatially encoded line by line in the readout direction for all phase coding directions. This process results in a long scan-time and has to be compromised with image spatial resolution. Thus, the primary motivation of this research is to develop approaches for alleviating the trade-off between spatial and temporal resolution in dynamic MRI by incorporating prior-knowledge. The examinations are designed to answer the following research questions: *How can we mitigate this trade-off? How can undersampled (less-than-complete) data be used or how fast can we speed up the image acquisition? Can any prior-knowledge be incorporated into this approach?* In this work, the potential for incorporating prior-knowledge into deep learning based super-resolution and its application in abdominal dynamic MRI is explored. With the fast inference speed of deep learning based super-resolution predictions, the proposed approaches are hypothesized to be extended for application in real-time or near real-time interventions. Prior-knowledge was derived from two different sources: The **Fine-tuned SR DynMRI** (which makes use of a high-resolution planning scan) and the **DDoS-UNet DynMRI** (which harnesses the dual static high-resolution image and the low-resolution dynamic scan) approaches. The contributions of these approaches are summarized in subsections 1.2, and the details of each method are discussed in chapter 3 and 4 accordingly.

1.2 Scientific Contributions

The contributions of this research to address the spatio-temporal trade-off are reducing the required scan-time (highly undersampled data), while maximizing the spatial and temporal resolution. The two approaches are as follows:

Fine-tuned SR DynMRI: Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Fine-tuning with a Subject-specific Scan

This approach tackles the spatio-temporal trade-off of dynamic MRI by fine-tuning the model parameters with a static high-resolution scan as prior-knowledge. This framework could enhance the spatial resolution of super-resolution reconstruction (from main training), while reducing the required scan-time per volume efficiently. The key contributions of this work are outlined below:

- The proposed Fine-tuned SR DynMRI framework shows that the super-resolution reconstruction of the dynamic MRI could be improved by fine-tuning with a subject-specific scan for only one epoch. Moreover, the proposed method demonstrates promising results, even when reconstructing from the 6.25% of the k-space center.
- A 3D modified U-Net combined with the perceptual loss was utilized, with the help of a perceptual loss network pre-trained on MRI, in order to promote realistic results close to human visual perception.
- Two approaches deliver promising results for overcoming the lack of a large training dataset in dynamic MRI. The first approach utilizes a pre-trained network on a benchmark dataset and follows by fine-tuning using a subject-specific planning scan. The second approach uses a patch-based super-resolution technique to increase the size of training data by splitting volumetric data into smaller patches.

DDoS-UNet DynMRI: Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Incorporating Temporal Information in a Dynamic Dual-channel UNet

This method presents the incorporation of both spatial information from a high-resolution planning scan, and also temporal redundancy of dynamic MRI as prior-knowledge. The DDoS-UNet DynMRI approach extends the single-image super-resolution by super-resolving the outputs recursively, which does not reconstruct each time-point separately. The main contributions of this study are summarized as follows:

- The proposed DDoS-UNet DynMRI framework makes use of dual-channel (static and dynamic) inputs with deep learning based super-resolution, which can mitigate

the spatio-temporal issue.

- The reconstructed results of DDoS-UNet DynMRI framework, even super-resolving from highly undersampling 4% of the k-space center, indicate the high similarity to the ground-truth qualitatively and quantitatively.
- The incorporation of temporal information could improve the quality of reconstructed images better than methods that include only spatial information. Moreover, the temporal information inherent in dynamic images has the benefit of overcoming the accumulation error caused by continuously reconstructing the super-resolved images.

1.3 Thesis Outline

The thesis is divided into the following chapters: **chapter 1** - introduction; **chapter 2** - scientific background of MRI, the neural network, dynamic MRI and reviews of current technological state of dynamic MRI; **chapter 3 and 4** are the contribution works based on the author's publications; and **chapter 5** provides a summary of the thesis and recommendations for future research. Figure 1.1 provides an outline of the thesis.

In **chapter 2**, **section 2.1** describes nuclear magnetic resonance (NMR) phenomenon; MR image contrast; MR hardware components; MR image production (including data acquisition techniques; k-space; and image reconstruction); along with an overview of MR image quality. **Section 2.2** outlines the concept of neural networks, including the neural network terminologies, basic structures of convolutional neural networks (CNNs), and applications of deep learning. In **section 2.3**, an overview of dynamic MRI in interventional radiology, common imaging sequences in abdominal imaging, the explanation of the spatio-temporal trade-off, and state-of-the-art dynamic MRI reconstruction are presented in this chapter. **Chapter 3** and **chapter 4** include the contribution frameworks using prior-knowledge. Both sections include a synopsis, background, methodology, results, discussion, and conclusions. Lastly, a summary and outlook of the key contributions are summarized in **chapter 5**.

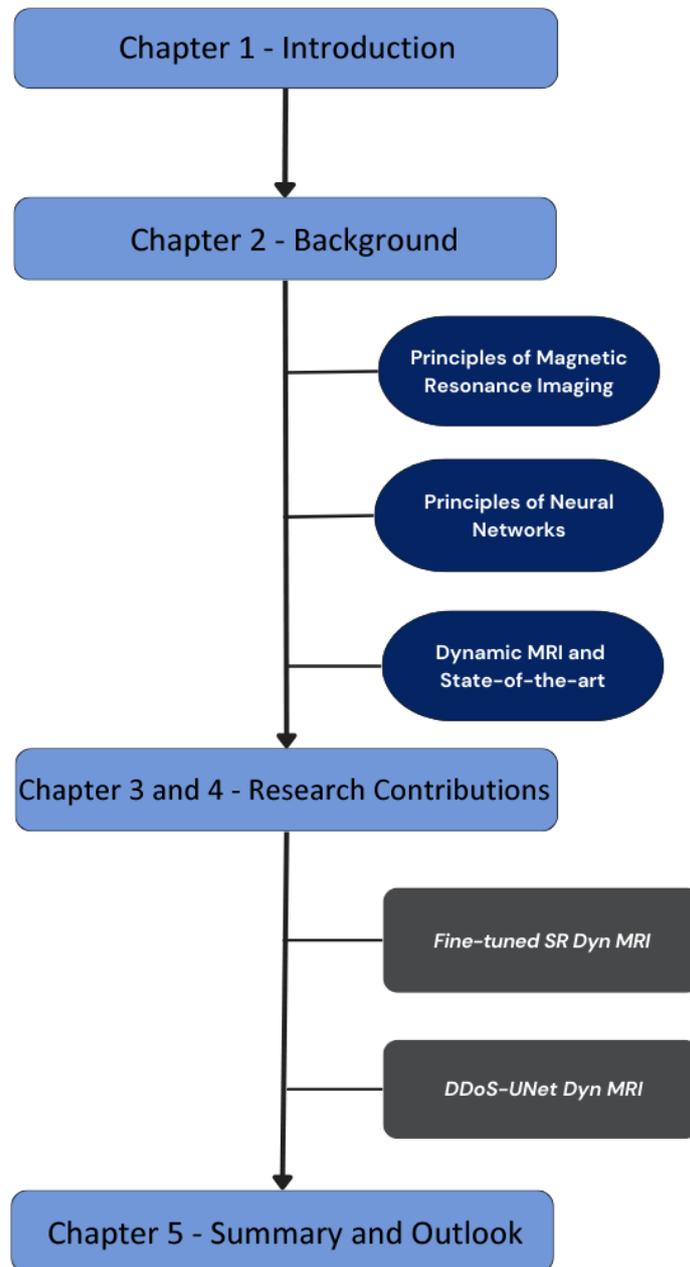


Figure 1.1: An outline of the thesis.

2 Background

“To change the world, we need to combine Ancient wisdom with new technologies.”

– Paulo Coelho

This chapter explains the background regarding the principle of Magnetic Resonance Imaging (MRI), neural networks, dynamic MRI and its state-of-the-art.

2.1 Principles of Magnetic Resonance Imaging

MRI was developed in the 1970s, and applications of MRI have been used in clinical practice for around 40 years. MR images have advantages over other imaging modalities because they are non-invasive, provide high-quality cross-sectional images, and portray superb soft tissue contrast without ionizing radiation. The cross-sectional image is also called a tomographic image, a term derived from the Greek word "*Tomos*", which means a cut [24]. In radiology, anatomical planes describing the orientation of the cross-section are commonly referred to as transverse or axial (dividing upper and lower parts), coronal (splitting front and back), and sagittal (separating left and right) planes, as depicted in figure 2.1. When it comes to MRI, internal structures are reconstructed in different shades of gray based on the signal intensities of body tissue or fluid such as water and fat [25]. Figure 2.2 compares cross-sectional images in the transverse plane of human abdominal dissection and a MRI. The grayscale of MR images can be represented differently depending on the MR sequences.

2.1.1 Proton Nuclei

Nuclear magnetic resonance (NMR) was early defined by Bloch and later by Purcell and Pound [25]. NMR imaging can be called imaging of protons based on the interaction between hydrogen nuclei in the body and the strong magnetic field of the MR scanner. The human body is composed of approximately 70 percent of water, which means we have abundant hydrogen nuclei in the body. The hydrogen atom (1H) has only one proton in each atom. A single proton can produce a small NMR signal and generally aligns randomly with the other nuclei in the normal state. The net magnetization in this state is therefore equal to zero.

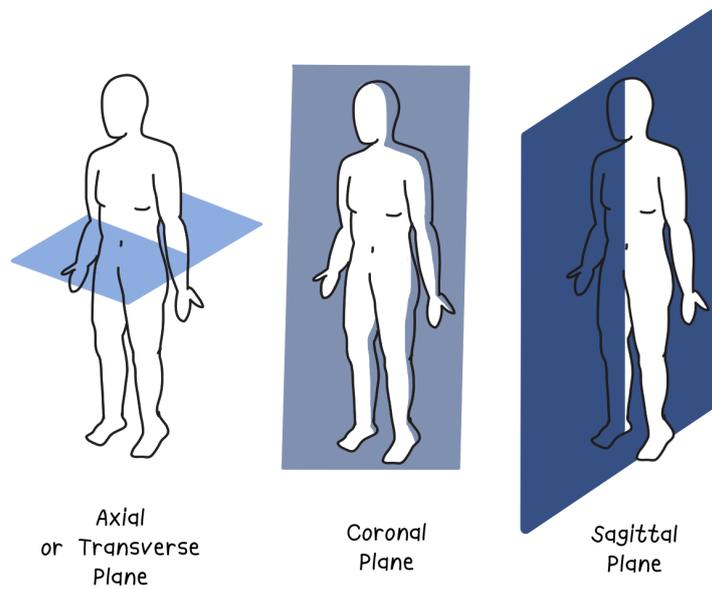


Figure 2.1: A graphic of anatomical planes: (a) transverse, (b) coronal and (c) sagittal planes.

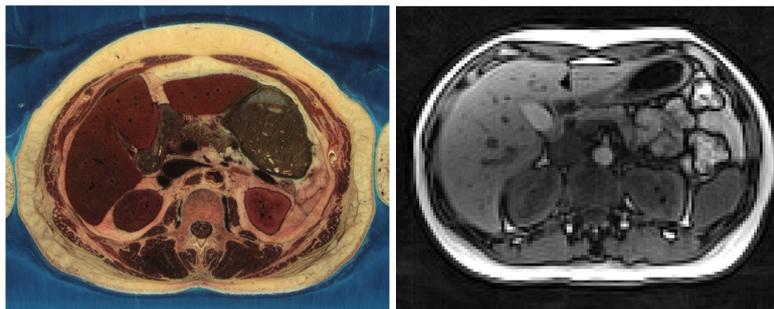


Figure 2.2: A cross-section abdominal dissection [26] and a MRI; T_1 -weighted image. For a MRI, different internal organs are illustrated in different shades of gray. As can be seen in the image, liver is brighter than kidneys because of the higher signal intensity. Abdominal fat is the most brightest and air is the most darkest in the image.

When the body tissue is exposed to an external magnetic field \mathbf{B}_0 along the z-axis, the net magnetization of protons is no longer zero. The proton energy levels will be split into high and low energy states. This interaction is called the Zeeman effect. The protons will be aligned or anti-aligned to the external magnetic field corresponding to its energy. The protons are aligned to \mathbf{B}_0 in the lower energy state and align the opposite direction of \mathbf{B}_0 in the higher energy state. Figure 2.3 shows the alignment of proton nuclei in the absence of an external magnetic field \mathbf{B}_0 , in the presence of \mathbf{B}_0 , and the energy differences according to Zeeman effect in the \mathbf{B}_0 field. The energy differences between the higher and lower energy states are given by equation 2.1.

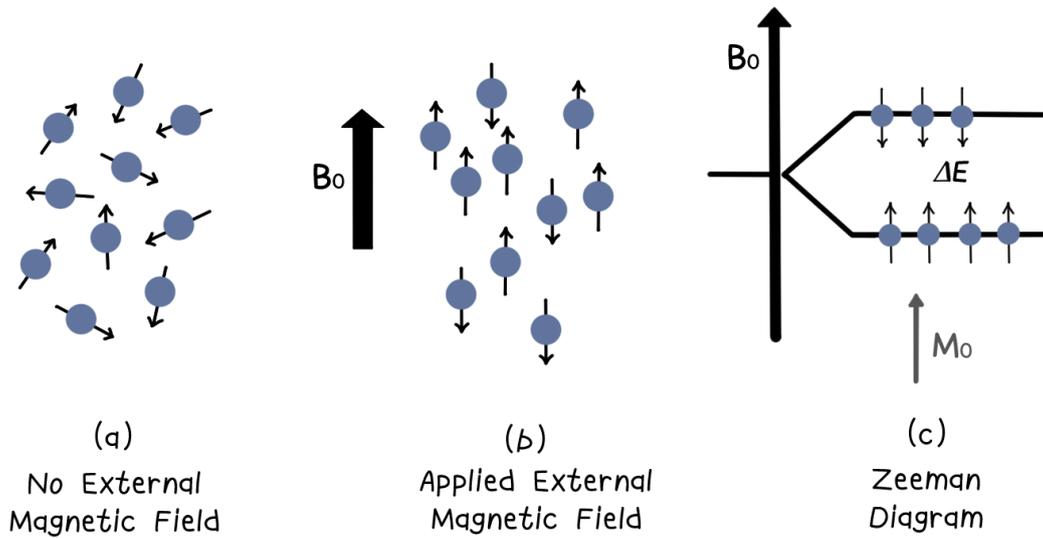


Figure 2.3: Graphic description of Proton alignment. (a) Normal state (b) With a presence of external magnetic field \mathbf{B}_0 , and (c) Energy differences (Zeeman diagram) [27].

$$\Delta E = \gamma \cdot \hbar \cdot B_0 = h \cdot v, \quad (2.1)$$

where v is the frequency of the magnetic field and \hbar is the Planck constant. The energy difference is in proportional to the magnetic field \mathbf{B}_0 .

2.1.2 Resonance and Relaxation

Since most protons align parallel to static magnetic field \mathbf{B}_0 , the net magnetization \mathbf{M} (macroscopic) has the same direction as \mathbf{B}_0 . Each proton has intrinsic magnetic momentum, which behaves like a tiny bar magnet. A proton will spin or process around the \mathbf{B}_0 axis at a specific rate according to the different tissues. This precession frequency is known as Larmor frequency (ω_0) [28] which can be formulated in equation 2.2.

$$\omega_0 = \gamma B_0, \quad (2.2)$$

where \mathbf{B}_0 is the static magnetic field and γ is called gyromagnetic ratio, which is proportional to magnetic field strength ($\gamma = 42.58$ MHz/T - close to radio frequency range). As typical magnetic fields used in the clinical range from 1.5T to 3T, the Larmor frequencies would approximately be 63.87 MHz and 127.74 MHz, respectively. The resonance is an important phenomenon in the MR Imaging process. Only the energy at the specific frequency (0) can stimulate (excitation) the protons to spin up (align the \mathbf{B}_0) and down (anti-align the \mathbf{B}_0). The protons then absorb the energy and re-transmit (relaxation) after some time. The frequency for the energy absorption interaction is known as resonant frequency, or in NMR phenomena, it is Larmor frequency.

Normally, the net magnetization of proton spins in the static magnetic field is at equilibrium and cannot be detected. Therefore, the radio frequency (RF) pulse, or \mathbf{B}_1 , is applied to excite the protons perpendicular to \mathbf{B}_0 (z-axis) to image these protons. In this way, the net magnetization of the proton flips to the x-y axis (in-phase). The protons start precessing around the z-axis, absorbing the energy, and then moving to a higher energy state. The applied RF pulse is known as RF excitation or 90° pulse.

When switching off the RF field, the protons will spin to the original direction of \mathbf{B}_0 to the equilibrium state (dephase). The MR signal is then produced and can be detected using the receiver coils. Owing to the analog nature of the MR signal, it is necessary to use an analog-to-digital converter (ADC) to digitize the signal to form the MR image.

The dephasing or decaying of MR signal occurs because of the interaction among protons and the protons with the environment. This phenomenon was introduced by Felix Bloch in 1946 [29]. Bloch explained about the NMR induction phenomena that the decays of NMR signal arise from the interactions among nuclei and the nuclei with their environment. The process of net magnetization \mathbf{M} returning to the equilibrium is referred to as relaxation, consisting of longitudinal \mathbf{M}_z and transverse components \mathbf{M}_{xy} .

The longitudinal magnetization is caused by the interaction between proton spins and the environment. The recovery time is referred to as spin-lattice relaxation time $T1$, which can be explained by [30]:

$$\mathbf{M}_z(\tau) = \mathbf{M}_0(1 - e^{(-\tau/T1)}), \quad (2.3)$$

where τ is the time after the 90° pulse and $T1$ means the time required time for \mathbf{M}_z to recover from 0 to about 63% of their original value \mathbf{M}_0 . Figure 2.4 illustrates proton spins

in the rotating frame and the recovery graph of longitudinal magnetization during the $T1$ relaxation process.

The transverse magnetization resulted from the interaction between each proton spin. The decay time is referred to as spin-spin relaxation time $T2$. Proton spins in the rotating frame and the decay curve of transverse magnetization during the $T2$ relaxation process are shown in figure 2.5. In reality, transverse magnetization decays faster than expected due to magnetic field inhomogeneity [31]. These inhomogeneities may result from intrinsic defects in the magnet itself or susceptibility-induced field distortions produced by the tissue or other materials placed within the field. The actual MR signal is also called the free-induction-decay (FID) signal, where the actual decay time $T2^*$ is expressed by:

$$1/T2^* = 1/T2 + 1/T2_M + 1/T2_{MS}, \quad (2.4)$$

Which $T2_M$ denotes the relaxation time due to the magnetic field inhomogeneity and $T2_{MS}$ indicates the relaxation time due to the magnetic susceptibility. The ideal $T2$ is considered a natural $T2$, and $T2^*$ is always shorter than or equal to $T2$. Thus, the the transverse magnetization according to time constant $T2^*$ can be described by:

$$\mathbf{M}_{xy}(t) = \mathbf{M}_{xy}(0)e^{(-t/T2^*)}, \quad (2.5)$$

where $\mathbf{M}_{xy}(0)$ is the transverse magnetization directly after the 90 pulse and $T2$ represents the required time for \mathbf{M}_{xy} to decay to about 37% of their original maximum value \mathbf{M}_0 .

For different body tissues, the recovery and decay times are different. For example, a proton of fat recovers to an equilibrium state faster than cerebrospinal fluid (CSF), the fluid surrounding the brain and spinal cord; thus, the fat appears to be brighter than CSF in the $T1w$ image. This characterization of different tissues allows the manipulation of the MR signal for representing the MR contrasts, which will be explained in section 2.1.3.

2.1.3 MR Image contrast: Tissue Characterization

One of the advantages of MRI is the ability to manipulate the signal from the characteristics of the tissues by adjusting the specific imaging parameters. There are three main types of contrast: $T1$ -weighted, $T2$ -weighted and proton density. The key to generating the different contrasts depends on the time between detecting the proton signals (echo) after 90° pulse, so-called echo time (TE); and the time between the successive FR pulses (repetition time; TR). TheThe same soft tissue is represented in different shades of gray in the different MRI contrasts, as seen in figure 2.6. In clinical routine, it is common to

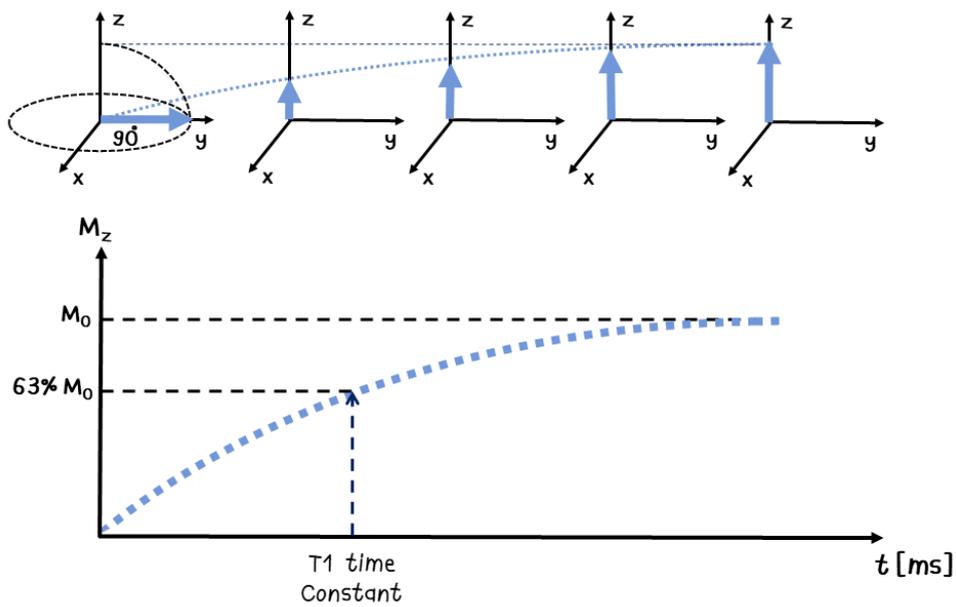


Figure 2.4: Graphical abstract of T_1 relaxation. The figure portrays the T_1 magnetization recovery after 90° pulse applying at the equilibrium. The magnetization M_z is initially dropped to zero and is then recovered to the equilibrium exponentially over time. The time constant of magnetization recovery to 63% of the original value M_0 denotes as T_1 .

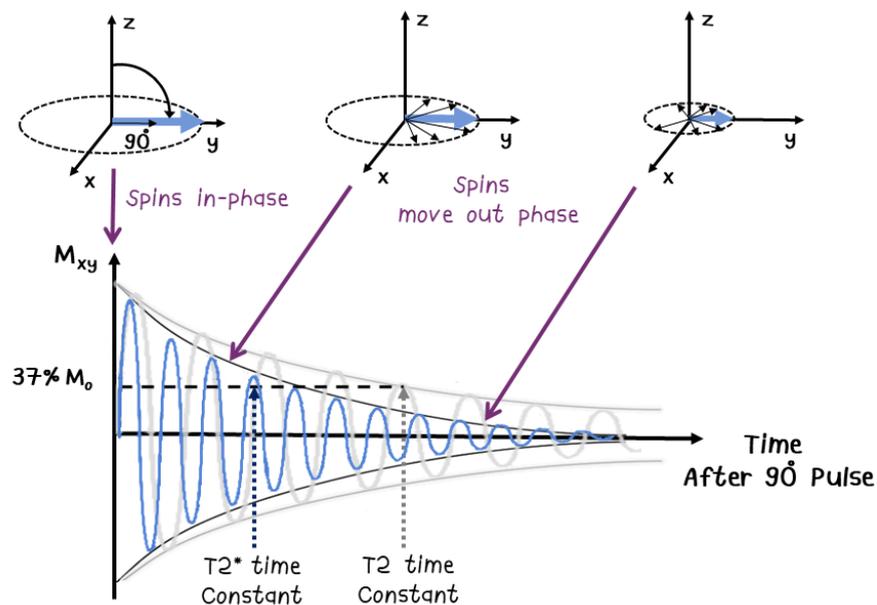


Figure 2.5: Graphical abstract of T_2 relaxation time. The figure shows the FID signal decaying after 90° pulse applying at the equilibrium. Then, the magnetization decays to 37% of the original value M_0 and continue decaying over time. The time constant of magnetization decaying to $37\%M_0$ denotes as T_2 . T_2^* time is shorter than T_2 time due to the effect of magnetic field inhomogeneity.

have mixed MRI contrast.

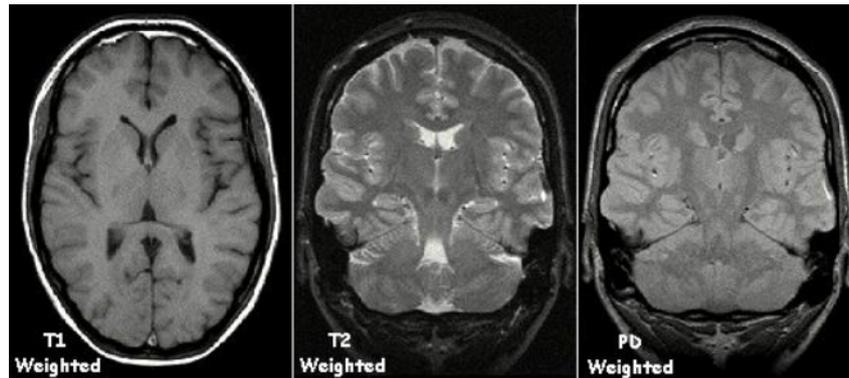


Figure 2.6: An example of MR image contrasts. From left to right: $T1w$, $T2w$ and PD images or brain MRIs [32].

$T1$ -weighting

A $T1$ -weighted ($T1w$) image provides excellent tissue contrast for differentiating various lesions and has been primarily used for diverse types of examinations. $T1w$ image is emphasized by short TR and short TE . Figure 2.7 shows various $T1$ relaxation times of different tissues.

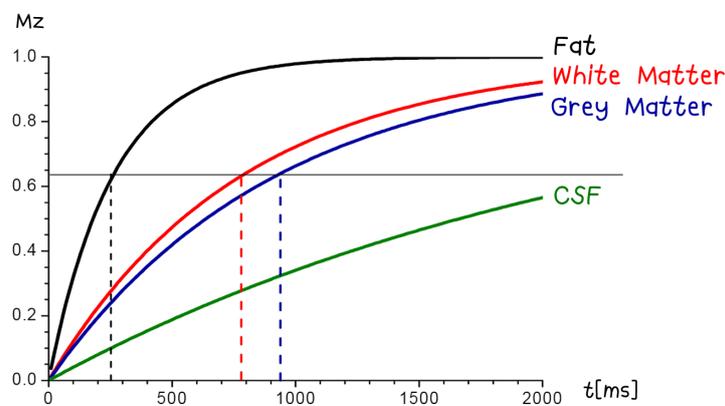
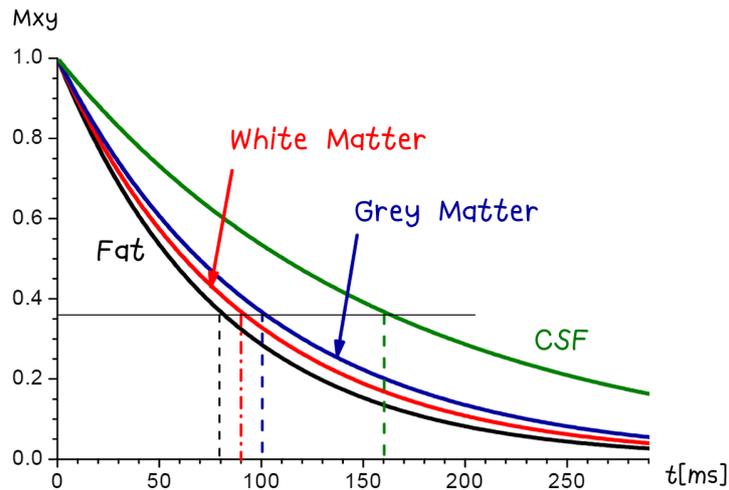


Figure 2.7: $T1$ relaxation time of the different tissue.

$T2$ -weighting

The second basic contrast is called a $T2$ -weighted ($T2w$) image, which is based on transverse magnetization decay. A $T2w$ image, in comparison to $T1w$, is emphasized by long TR and long TE . Figure 2.8 depicts the various $T2$ relaxation times of different tissues.

Figure 2.8: T_2 relaxation time of the different tissue.

Proton density

A proton density image (PD) refers to the concentration or density of protons in the tissue. The PD -weighted (PDw) image is produced by minimizing the effect of T_1 and T_2 , where it has long TR and short TE . There is a strong signal in PDw over T_1w and T_2w , representing a strong appearance of soft tissue.

2.1.4 MR Hardware Components

The hardware components of an MRI scanner consist of a computer processor system, a magnetic system, a gradient system, a radio frequency system, and a console for acquiring the signal and adjusting the scan parameters. Figure 2.9 illustrates the outside and inside of a MR scanner.

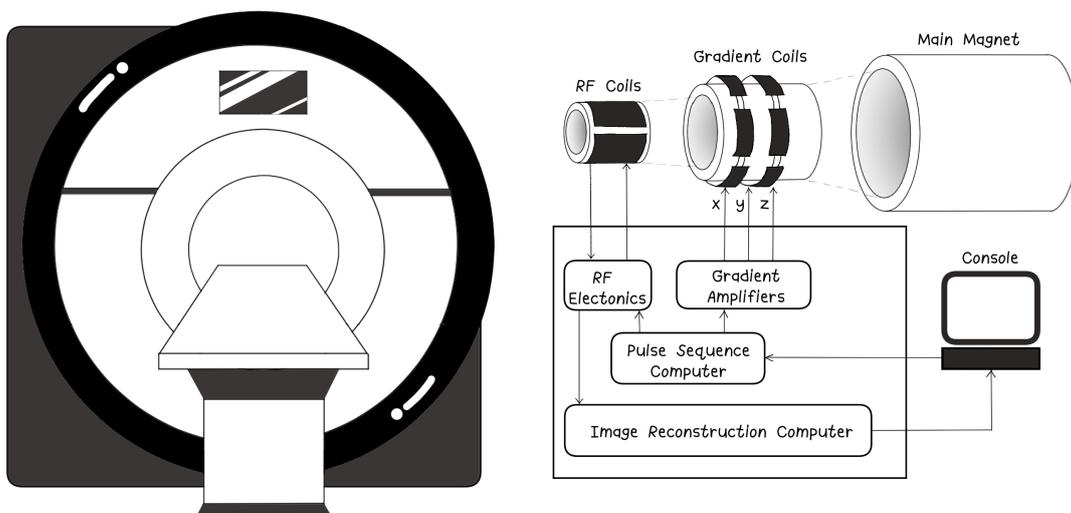


Figure 2.9: An illustration of the main parts of MR hardware components. Left: MRI machine outside, and right: inside components of system

To prevent unnecessary frequency interfering with the imaging process, the MR room is designed to block external signals with conductive material (known as a Faraday cage). The three main parts of magnetic fields are the static magnetic field \mathbf{B}_0 , a gradient system; G_x, G_y, G_z , and a radio frequency system.

Static Magnetic Field: $\approx (T)$

The static magnetic field \mathbf{B}_0 points in the z-direction (longitudinal plane) as a reference for the direction of tissue magnetization. The net magnetization and resonance frequency depend on the \mathbf{B}_0 field strength. The higher field strengths enable the higher signal-to-noise ratio (SNR). The 1.5T to 3T MRI scans are most widely available, while MRIs reaching up to 7T are for clinical examinations [33].

Gradient System (G_x, G_y, G_z): $\approx mTm^{-1}$

The gradient system consists of G_x, G_y , and G_z coils for producing gradients in the three orthogonal directions: x, y, and z. The gradient coils are used to spatially localize the object's signal.

Radio Frequency System: $\approx \mu T$

The radio frequency (RF) system consists of transmitter, coil, and receiver. The transmitter is used to produce the desired RF pulse, or \mathbf{B}_1 , (in terms of frequencies, bandwidths, amplitudes, and phases) at the Larmor frequency. This \mathbf{B}_1 is generated from the coil and is responsible for the excitation of proton nuclei. The RF coil is also used for receiving the produced MR signal.

2.1.5 Data Acquisition Techniques, k-space, and Image Reconstruction

The production of MRI is achieved by data acquisition and image reconstruction. The data acquisition commences by acquiring the proton signal intensity of the object. To spatially localize the signal, slice selection is the first step, followed by frequency encoding and phase encoding steps. The raw data matrix obtained from the MR scanner is known as k-space, which can be reconstructed to image space using Fourier transform. Figure 2.10 summarizes the main two steps of MR image production: data acquisition and image reconstruction through the k-space.

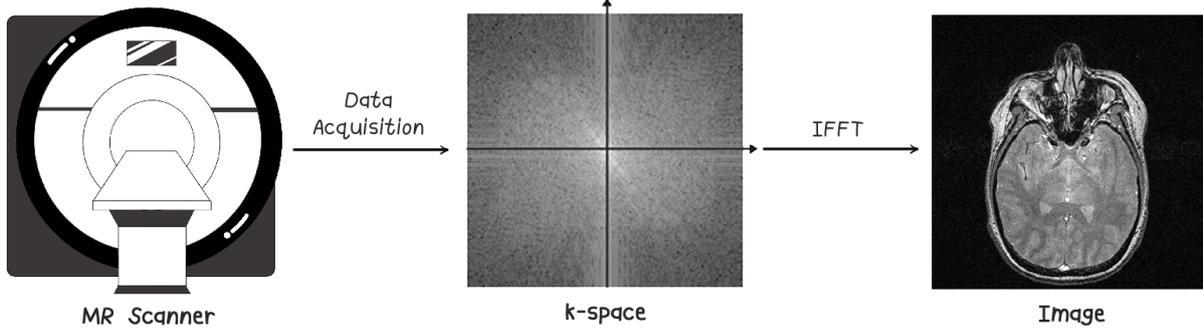


Figure 2.10: The production of MRI includes the step of image acquisition and image reconstruction through k-space.

2.1.6 Data Acquisition Technique

To acquire the data of the entire volume, the MR scanner needs to acquire many slices by looping multiple iterations of pulse sequences. Data acquisition techniques of these loops are categorized into 2D and 3D-volume acquisitions. During 2D-multislice acquisition technique, the data is acquired line by line for each slice and continuously acquired for the entire tissue volume. Thus, the total scan time is relevant to the number of acquired line data (N_{AQ} , repetition time), and the number of phase encoding steps (N_{PE}). Each phase encoding step acquires one echo. The total scan time is [30]:

$$\text{Scan time}_{2D \text{ Multislice}} = N_{AQ} \times TR \times N_{PE}. \quad (2.6)$$

For 3D volume acquisition, it is advantageous to have small voxels with thin and continuous slices, which is good for vascular imaging applications. Although 3D imaging provides higher SNR than multi-slice, the scan time may take longer. Basically, 3D volume acquisition technique is a double phase encoding technique, where the second phase encoding gradient is applied in the slice selection direction to divide the volume into individual slices. So, the number of divided partitions (N_{PART}) affects the number of slices and scan time. The total scan time is given as:

$$\text{Scan time}_{3D \text{ volume}} = N_{AQ} \times TR \times N_{PE} \times N_{PART}. \quad (2.7)$$

2.1.7 k-space

The concept of k-space was introduced in 1983 [34]. The letter "k", originated in honor of Heinrich Kayser, represented a wavenumber in physics [35]. For MRI, k-space data is simply described as the array of data in the frequency domain. The k-space data obtained from the MR signal is stored as complex data with real and imaginary components (see

figure 2.11). The MR signal can also be represented as magnitude (Mag) and phase (Φ) alternatively. For clinical routine, magnitude images are generally used for diagnosis, while phase-images are used only for particular exams such as a CSF flow study.

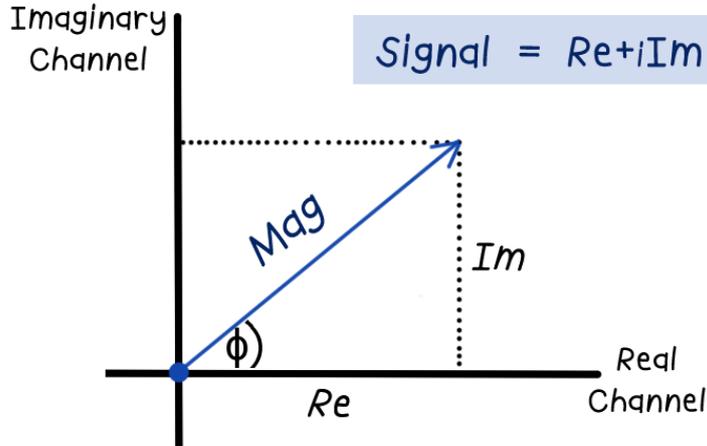


Figure 2.11: An illustration of each k-space data as a complex number in real and imaginary planes or magnitude and phase components as an alternative. Note that $Mag = \sqrt{Re^2 + Im^2}$ and $\Phi = \tan^{-1}[Im/Re]$.

The traditional method of MR image acquisition is performed by collecting k-space data line by line in a Cartesian grid. The k-space data is treated as a k_x and k_y grid of data points, in which each k_x refers to the data in frequency (readout) direction and k_y represents to the data in phase encoding direction. The transformation of the data from the frequency domain to the image domain can be performed by Fourier transform (FT). Therefore, the MR image is actually a frequency map of proton signal obtained from a specific location of the tissue. And vice versa, the image data can be mapped into k-space data using inverse Fourier transform.

In addition, each point of the k-space data does not map directly to the point in image data; however, it is contributed to a specific spatial frequency. The central part of the k-space contains a strong signal and low frequency, giving contrast information and general shapes of the object, whereas the outer part contains high spatial frequency providing edge information and contours of the object. Figure 2.12 illustrates the k-space data and the corresponding reconstructed images for collecting different parts of the k-space. It can be observed in figure 2.12(b) that the reconstructed image has the general information about contrast but loses the details when collecting only the k-space center.

2.1.8 Image Reconstruction

Given the raw data matrix as a function of time $S(t)$, the MR signal can be described as:

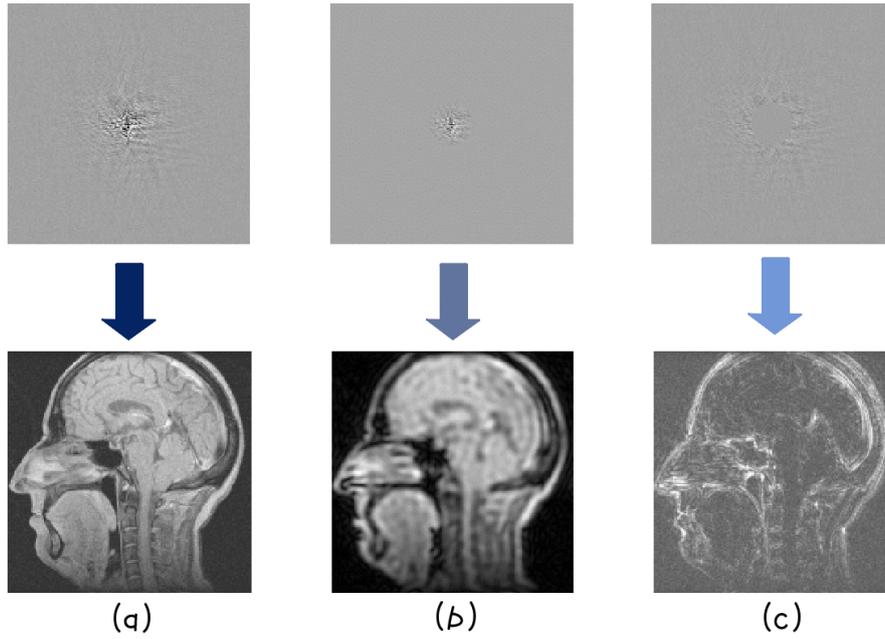


Figure 2.12: k-space and the corresponding reconstructed images. (a) Fully-sampled k-space, (b) Only low spatial frequency of the k-space data is acquired, as a results, the reconstructed image has contrast information but no details, and (c) Only high spatial frequency of the k-space data is acquired, the reconstructed image depicts only the edge of anatomical structures [25].

$$S(t) \propto \int \int M(r) e^{-2\pi i k(t) \cdot r} dr, \quad (2.8)$$

denoted that $S(t)$ is the raw data at the particular time t , $M(r)$ is the net magnetization at the position r , and $e^{-2\pi i k(t) \cdot r}$ is the Fourier basis. It can be seen that the MR signal equation is the Fourier transform (FT) of the spin density of the object. For a given k-space data point $k(k_x, k_y)$, equation 2.8 can be written for 2D imaging as:

$$\begin{aligned} S(k_x, k_y) &\propto \int_x \int_y \rho(x, y) e^{-2\pi i (k_x x + k_y y)} dx dy \\ &\propto FT\{\rho(x, y)\}, \end{aligned} \quad (2.9)$$

where $S(k_x, k_y)$ is the MR signal at $k(k_x, k_y)$, $\rho(x, y)$ is the spin density composed from \mathbf{M}_0 and \mathbf{M}_{xy} of the object at the position k_x and k_y , $e^{-2\pi i (k_x x + k_y y)}$ is the Fourier basis, and $FT\{\rho(x, y)\}$ is the Fourier operation of the object signal. From 2D extended to 3D imaging, the k-space data is encoded step-wise for each combination of k_x , k_y , and k_z which can be modeled as:

$$\begin{aligned}
 S(k_x, k_y, k_z) &\propto \int_x \int_y \int_z \rho(x, y, z) e^{-2\pi i(k_x x + k_y y + k_z z)} dx dy dz \\
 &\propto FT\{\rho(x, y, z)\}.
 \end{aligned}
 \tag{2.10}$$

Therefore, the proton density of the object can be reconstructed in the reverse direction by performing the inverse Fourier transform FT^{-1} on the detected signal.

$$FT^{-1}[S(k_x, k_y, k_z)] \propto \rho(x, y, z).
 \tag{2.11}$$

2.1.9 Spatial Encoding and Gradient Echo Pulse Sequence

To localize and encode the spatial information of the object, the three main mechanisms are processed, namely slice selection, phase encoding, and frequency encoding. These mechanisms can be defined to manipulate the FID or MR signal using a pulse sequence. A pulse sequence consisting of RF pulse and gradients is repeated many times during the scan. The basic pulse sequence, namely gradient echo or gradient-recalled echo (GRE), is explained briefly in this section. A GRE-based pulse sequence such as spoiled gradient echo is commonly employed for fast imaging as the low flip angle of RF pulse allows the reduction of repetition time (TR). The steps of GRE technique are shown in figure 2.13 and can be summarized as follow:

1. The RF pulse with the flip angle α excites a selected slice.
2. Phase encoding with opposite polarity is performed.
3. Frequency (readout) gradient recalls the echo. The time between RF pulse and echo is called echo time (TE).
4. The gradient is spoiled or de-phased by the remain transverse magnetization.
5. The next RF pulse is repeated at the interval time of repetition time (TR : time between the sequential excitation pulses to a given slice).

2.1.10 Image Quality Overview

Since this thesis is relevant to the trade-off between spatial and temporal resolution, the overview of image quality, including the signal-to-noise ratio (SNR), trade-off, and spatial and temporal resolution, is given in this subsection.

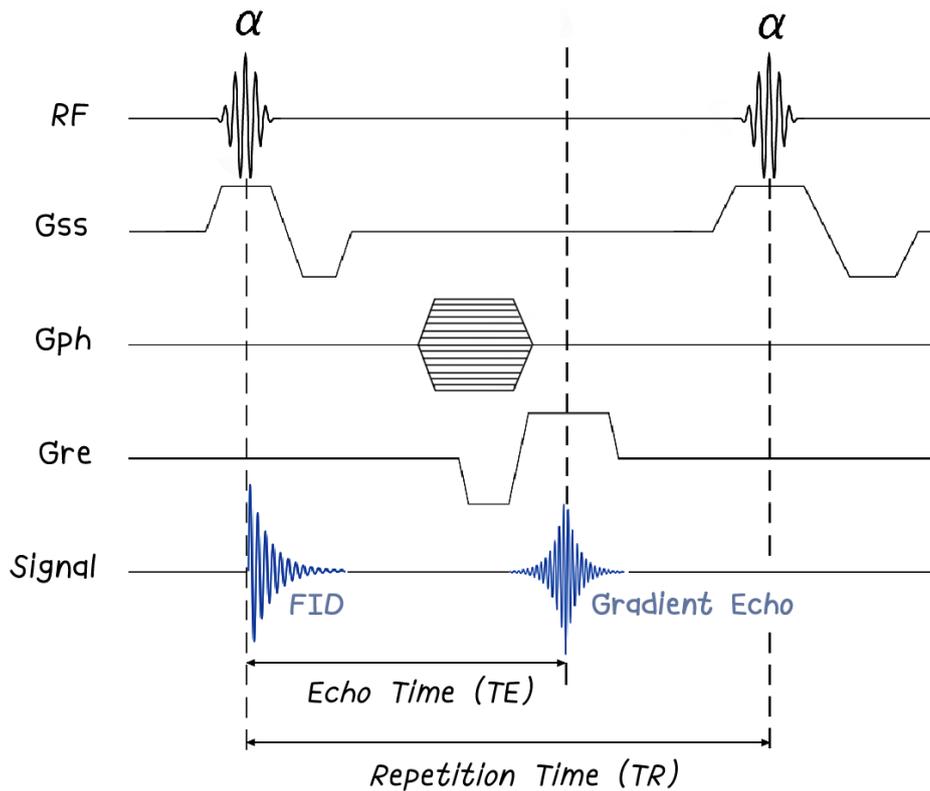


Figure 2.13: A diagram of gradient echo pulse sequence.

SNR and the trade-off

The image quality of MRI can be influenced by several factors such as image resolution, contrast-to-noise ratio (CNR), artifacts, and signal-to-noise ratio (SNR). The SNR is an essential parameter for MRI. It corresponds to the level of the signal and presented noise, and there are many factors that can contribute to the SNR level including magnetic field strength, field inhomogeneity, receive coil sensitivity, voxel size or tissue characteristic, and the selected pulse sequences. The desired MR image should typically have an adequate spatial resolution, an acceptable signal-difference-to-noise ratio, and a reasonable scan time [30]. These parameters are compromised by one another, which is called a trade-off in MRI.

An optimization parameter is varied in different areas of examination. For instance, when an abdominal MRI is done using a breath-hold technique, a short scan time is preferable to reduce motion artifacts even though spatial resolution might be reduced. For dynamic imaging, temporal resolution is significant and needs to be optimized with spatial resolution. Inadequate image resolution might lead to inaccurate diagnosis of small lesions. Figure 2.14 shows an example of good and poor MR image quality.

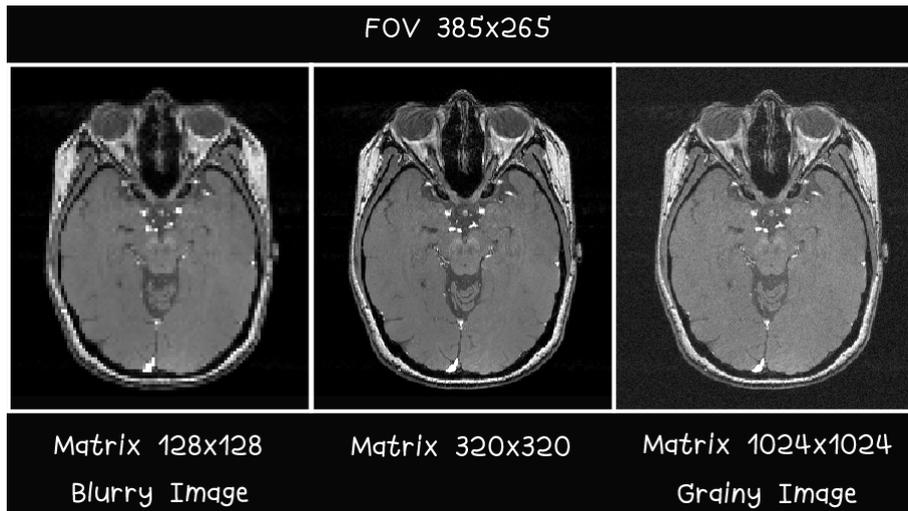


Figure 2.14: An example of good and poor MR image quality. A good quality of MRI is in the acceptable range which has a good SNR. In contrast, a poor quality of MRI as a result of low basic resolution (blurry image) or too high basic resolution (grainy image).

Spatial and temporal resolution

Image spatial resolution in general refers to the ability to differentiate between two adjacent objects, which can be estimated in several ways such as the number of dots, lines, or pixels. Spatial resolution in MRI is determined by pixel size in a specific field of view [36]. Pixel size can be calculated as follows:

$$\text{Pixel size} = \text{Field of view (FOV)} \div \text{Matrix size}, \quad (2.12)$$

e.g., for a given matrix size of 256×256 mm in FOV of 256 mm, the pixel size is equal to 1 mm. A high number of pixels results in a high spatial resolution of the image. As MRI is often obtained as a three-dimension image, so, the image resolution is represented as voxel size (mm/pixel) from pixel size and the slice thickness.

For any MRI examination, temporal resolution is also a parameter that is commonly taken into consideration with spatial resolution. Temporal resolution refers to how often the set of images is acquired in a specific time frame, which can be determined using this formula [37]:

$$\text{Temporal resolution (cine)} = \text{NVS} \times \text{TR}, \quad (2.13)$$

where NVS denotes the number of views per segment and TR is the minimum TR from pulse sequence.

2.2 Principles of Neural Networks

2.2.1 Modeling the Artificial Neurons

The principle of modeling an artificial neural unit from a biological neuron is meant for a better understanding of a conceptual model, which is derived from a simplified function of known physical properties, so as to build a computational model, which is more accurate in terms of mathematics [38].

In fact, a neuron can be simply considered as a small unit of computation. The first formal approaches of computing models studied during the 1930s to 1940s, known as the five models of computations, comprise the mathematical model [39–42], the logical operation model [43], the computer model from Z1 to ENIAC von Neumann architecture [44], cellular automata [45], and the biological model [46, 47]. The last approach was the first conception of artificial neurons, proposed in 1943 based on the McCulloch and Pitts model. Since then, the amount of research dealing with neural networks has been increasing dramatically, opening up a diverse range of research questions.

The scientific questions inspired by human neurons are relevant to synaptic, or connection, efficiency and postsynaptic cell properties regarding anatomy and physiology. A neuron, or nerve cell, is a complex system that can be more complicated than computers. The three main classes of neurons based on their functions can be categorized into neurons, motor neurons, and interneurons [48]. Although each of those types of neurons function differently in the body, the general structures are similar. Each neuron consists of multiple branches of dendrites, a cell body (soma), and only one axon. As the nervous system is a huge network, several neurons are connected via synapses to conduct nerve impulses. This is where information, electrical and chemical, is transmitted.

To understand how artificial neurons work, a simplified explanation of a biological neuron (2.15 - upper) compared with an artificial neuron (2.15 - lower) is listed in the following points:

- Dendrites - Input of the network
- A cell body (soma) - Network node
- An axon - Output of the network
- Synapses - Interconnections with weights

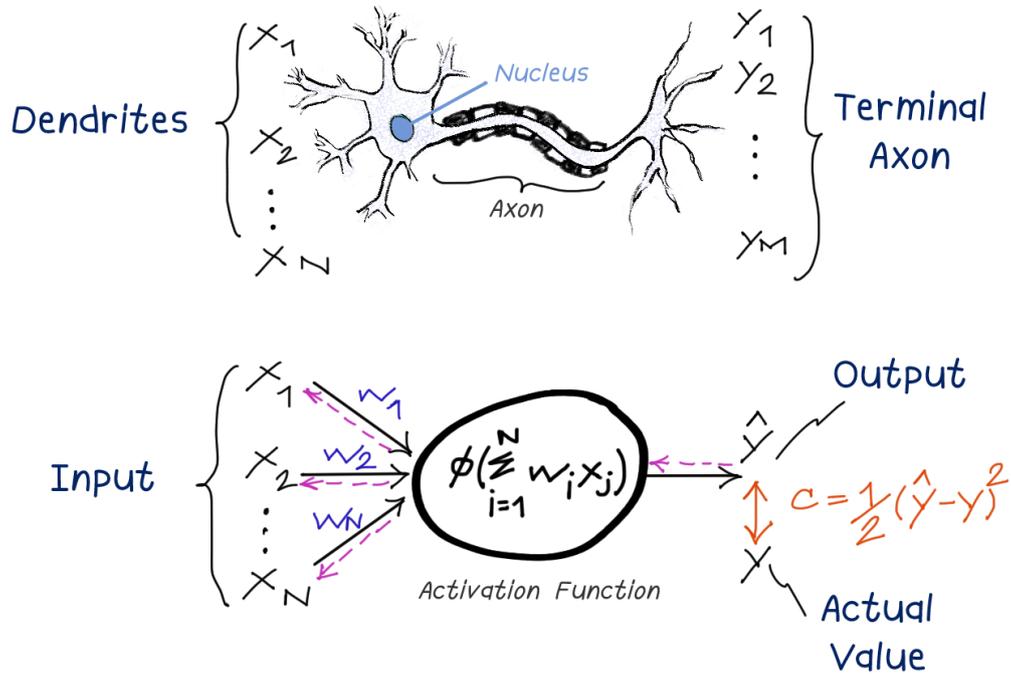


Figure 2.15: An illustration of physical structures of neurons. Upper: Biological neuron and lower: Artificial neuron.

2.2.2 A Perceptron

A single layer of artificial neuron is called a perceptron. It is comprised of inputs (x_1, \dots, x_n) , weights (w_1, \dots, w_n) and bias b , a net summation $(\sum_{i=1}^n)$, and an activation function (ϕ) . The learning process of the neural networks is associated with synapses that are connected to the cell body. It is operated by adjusting weights at these contact points. These weighted inputs are then accumulated and passed through the activation function. The net inputs can be calculated as:

$$\begin{aligned}
 y_{in} &= \phi(x_1 \cdot w_1 + x_2 \cdot w_2 + \dots + x_n \cdot w_n) + b \\
 &= \phi\left(\sum_{i=1}^n x_i \cdot w_n + b\right)
 \end{aligned} \tag{2.14}$$

To minimize the error function c , the obtained output (\hat{y}) is compared with the actual value (y) as described in equation 2.15.

$$c = 1/2(\hat{y} - y)^2 \tag{2.15}$$

2.2.3 Multi-layer Perceptrons

Multi-layer perceptrons (MLPs) or feed-forward neural networks are a stack of perceptrons that have multiple hidden layers. The diagram illustrates an MLP structure. Each neuron from the input layers goes into hidden layers to make decisions for the output layers. These neural layers are connected to perform forward and backward operations (backpropagation).

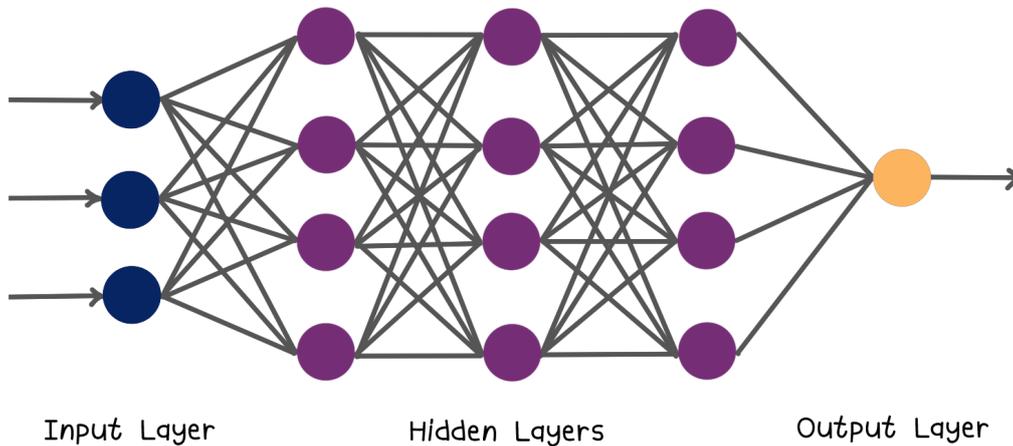


Figure 2.16: Multi-layer Perceptrons.

The learning process of a neural network can be defined into two main processes: forward and backpropagation. The first step in forward propagation is initializing parameters (weights and biases) randomly and feeding input data into the network. Then, the predicted output will be compared with the expected value to calculate the error or loss. The next step is to process the backpropagation and update the parameters based on the loss with the help of the gradient descent algorithm. These steps will then be iterated until the loss is minimized and good predictions are achieved. The explanations of important terms for each step will be clarified in the following subsections.

2.2.4 Forward Propagation

Forward propagation, or forward pass, is the process of propagating the information from the input to the output layer. These neurons are connected to the next layers through weights, which are the numerical values assigning their importance to the inputs. These weights are initialized randomly and multiplied to the inputs, then summed up and sent as inputs to the next adjacent layers in the hidden layers. The weighted summation is passed through an activation function which determines the contribution of each neuron to the next layers. The obtained output is simply a probability from 0 to 1.

2.2.5 Backpropagation

Backpropagation, or backward propagation, is a method to compute the gradient of the loss, which corresponds to their weights, and propagate the information in the opposite direction of forward propagation. The term gradient is the first derivative of a function at a given point. Backpropagation starts by calculating the loss, then the weights are adjusted in the reverse direction with the optimizer corresponding to the measured loss. The reason why the neural network can learn by itself is that the backpropagation enables the minimum of the loss function and properly weighs every iteration of the learning process.

2.2.6 Loss Function

The solution to the learning process is to iterate all steps involving forward propagation and backpropagation until the loss, or error, of the objective function is minimized [49]. To estimate the error of the actual output and prediction, several lost functions are available. For example, mean square error (MSE or L1) is commonly employed as a lost function by summation of squaring the error before averaging it.

2.2.7 Gradient Descent

Gradient Descent is a mathematical algorithm for minimizing the error function. The method of gradient descent is a more realistic error function that can tackle the issue of not converting to the global minimum [49]. The graphical illustration of the gradient descent algorithms is shown in figure 2.17. The equation of minimization or maximization of a function F_x can be defined as:

$$x^* = \operatorname{argmin} F(x) \tag{2.16}$$

2.2.8 Stochastic Gradient Descent

Stochastic gradient descent (SGD) is an extension of the gradient descent optimization algorithm used for overcoming the local minima error, as illustrated in figure 2.18. SGD estimates the gradient randomly by using a small set of samples. As this approach approximates the data after each example of the training, it is faster than algorithms that need to load the whole data into the memory.

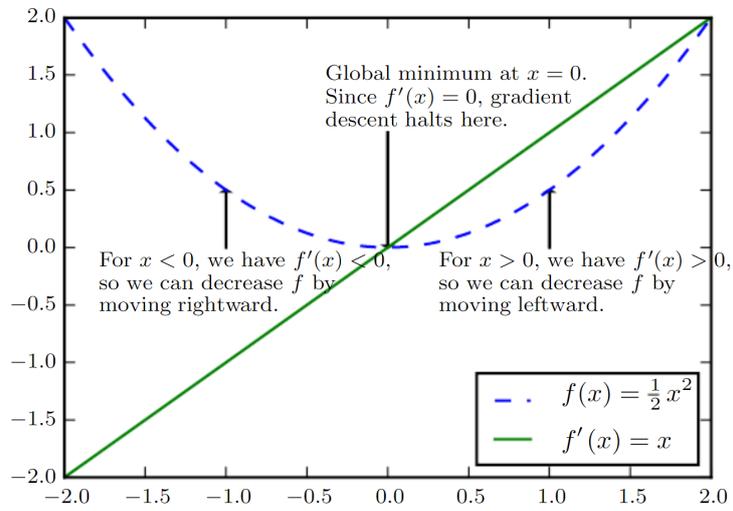


Figure 2.17: A graphical representation of how gradient descent works [49].

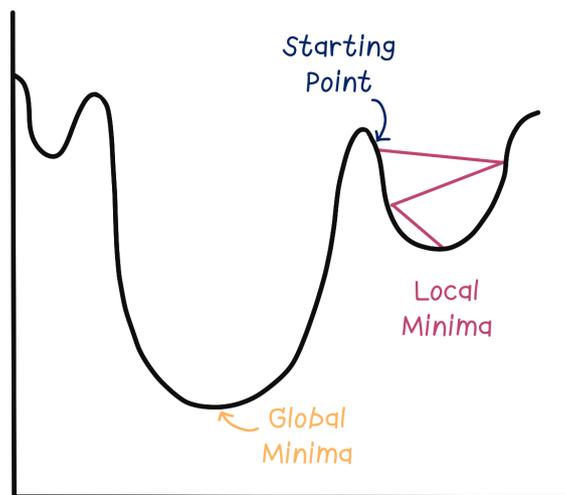


Figure 2.18: An illustration of stochastic gradient descent with global and local minima errors.

2.2.9 Convolutional Neural Networks

In neuroscience, there are studies investigating the response of the human brain to perceptual ambiguity from ambiguous figures [50–52]. The research results showed that the interpretation of the brain depends on the a certain feature. The neurons in human nervous system are only stimulated in the receptive field in visual system. The ambiguous figure, such as figure 2.19, can be interpreted in multiple ways with respect to the region of the visual field, which in this case it can be defined as both rabbit and duck.



Figure 2.19: The original duck-and-rabbit illusion from Fliegende Blätter 1882, captioned "Welche Thiere gleichen einander am meisten?" [53]

Similar to how a human classifies an object, the process of convolutional neural networks (Convolutional Neural Networkss (CNNss)) data responds to its receptive field or so-called feature maps. CNNss is nothing but a class of Artificial Neural Networkss (ANNs) where each unit in each layer is structured to a grid-like topology called feature maps [49]. Given an example of a classic classification problem using neural networks, the feature representations can be ears, mouth or tail (see figure 2.20) that use to identify whether the image is a cat or not.

[!htbp]

CNNs widely utilized for the applications such as image classification, hand written text recognition and natural language processing. This class of neural networks employs convolution as a replacement for matrix multiplication in at least one of their layers. The convolution is a special type of linear operation that can be formulated as:

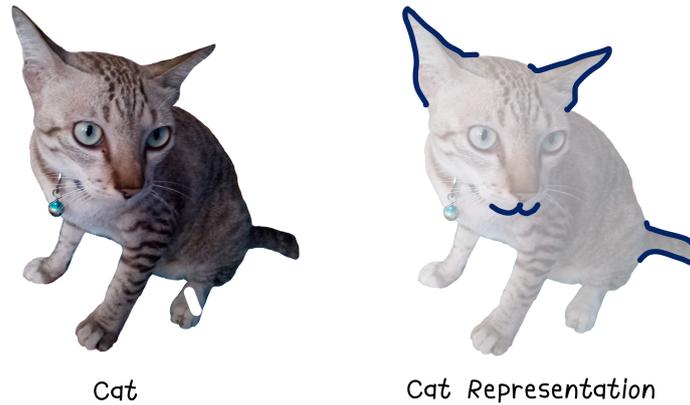


Figure 2.20: A simple example of feature representations of a cat.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad (2.17)$$

Whereby the input g is convolved with a kernel f . The convolutional function is a combination of two functions where each one can modify the other. The convolutional neural network was introduced by Lecun [53] in 1998. The architecture of CNNs generally comprises of two parts: feature detection, including convolutional layers and pooling, and classification layers involving flattening and full connection steps, see figure ??.

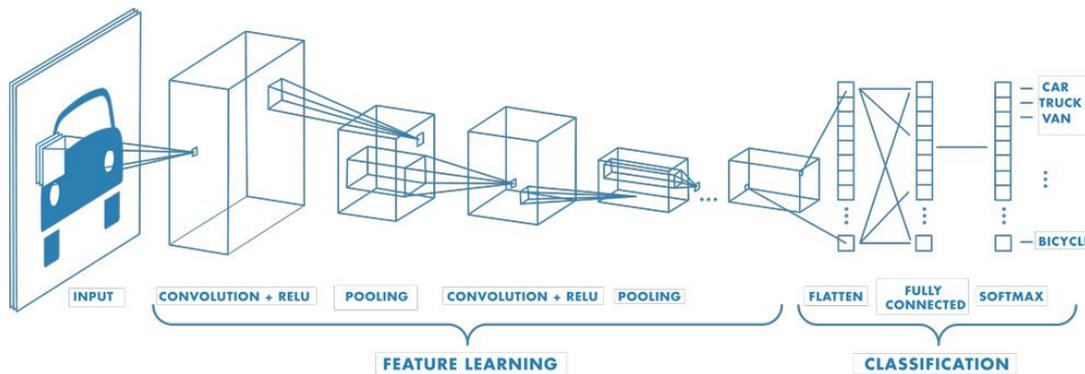


Figure 2.21: An example of CNNs [54]. The CNN architecture is built with several neural network layers, feature learning, and classification parts.

2.2.10 Feature Detection

This part is the main building block of the CNNs, performing convolution and pooling operations. The convolution is calculated on the input by a dot product between the filter (or kernel) and the receptive field (a region of patches in the input). Figure 2.22 displays an example of a simple 2D convolution. As one can see in the figure, the input is initially divided into smaller patches. Then a filter or kernel sliding convolves around each patch one by one until the whole input is complete. The results of this convolution are called

activation maps or feature maps.

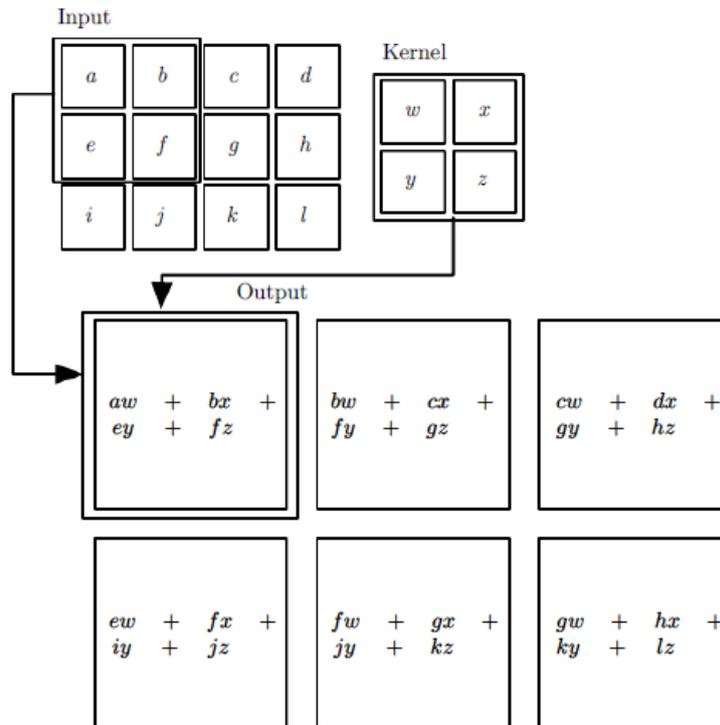


Figure 2.22: An example of a 2D convolution operation [49].

An activation function is then applied to the convolution output. The activation function serves as a non-linearity introduction enabling the model to learn more complex representations in the data. The most frequently used activation functions are Sigmoid, Tanh, Softmax, and Rectified Linear Unit (ReLU).

Stride is the hyperparameter to determine the step the convolution filter moves from one to another. The typical size of stride is one. Specifying the proper stride size allows the CNNs to learn to capture the important features of the input. However, increasing the stride size results in patch overlapping which in turn increases the computational complexity.

Furthermore, after convolving the input, the dimension of the output will be smaller than the input and the convolution operation could lead to the loss of information at the image's corner. To prevent this, padding is an essential process to consider. Padding in CNNs is simply the process of adding the zero values to the outer part of the input so that the activation map will not shrink.

The last step of feature detection is the pooling function. Normally, a pooling layer is added in between each convolution layer. The pooling layer is used for consolidating the feature extracted by the filters, to reduce the size of feature representations and computations before generating the output. The well-known pooling function is called maximum pooling,

or max pooling, which considers the maximum value in each feature map.

2.2.11 Classification

The components of classification combine flattening and full connection functions as the last steps in CNNs. After finishing the earlier steps, multiple pooled feature map are produced. Then, flattening is employed to convert the high dimensional pooled feature maps (2D or 3D) into a one-dimensional vector, as can be seen in figure ???. In this vector of input, the features or attributes were extracted and encoded which make it able to classify the tasks.

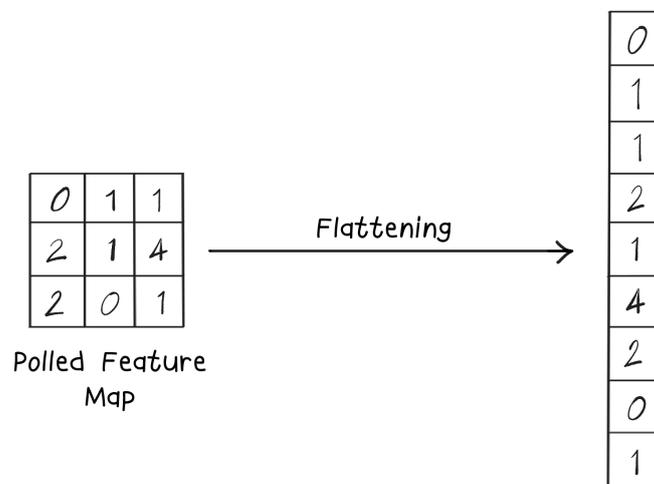


Figure 2.23: Flattening step.

Following flattening, a long vector of input data will be passed through the full connection step for further processing. Basically, there are at least three layers in the full connection - input, fully-connected and output layers. At the end, the prediction from convolutional neural networks will be given. Figure 2.24 shows an example of cat or dog classification with an input image. The given image was predicted to be a dog and cat classes with probability of 0.95 and 0.05 respectively, based on the feature information such as ears or nose detected from the image.

2.2.12 Deep Learning and Applications

Deep learning (DL) was introduced by [55] through deep belief networks for fast learning algorithm in 2006. DL model is derived from ANN, comprising multiple layers of non-linear information for extracting high levels of features [56]. Basically, deep learning is a form of machine learning, which has turned into a sub-field of artificial intelligence (see the divisions illustration in figure 2.25). Machine learning is more specific to a particular domain as

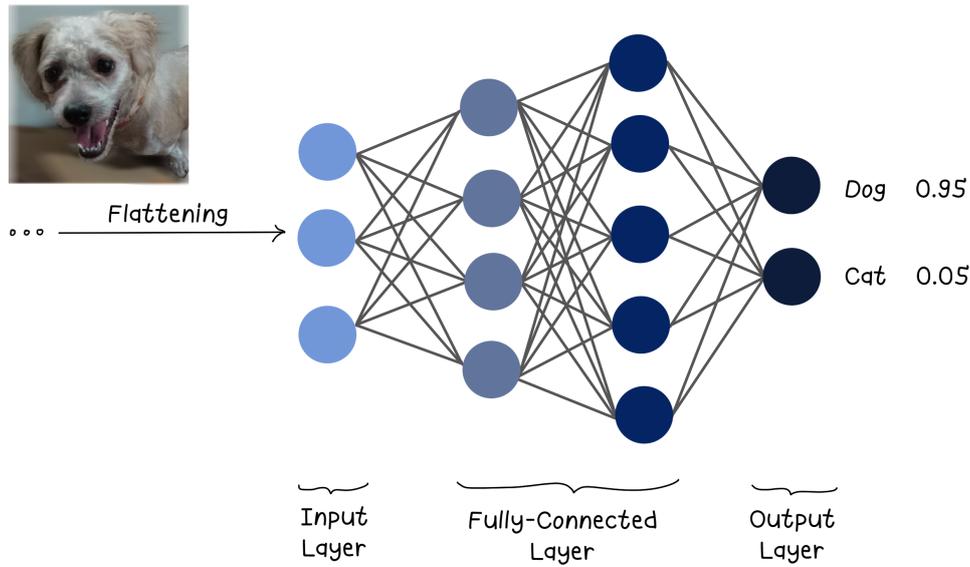


Figure 2.24: An example of cat or dog classification task after full connection layer.

it requires complex features defined by humans, whereas deep learning involves feature representations directly from the data, making it flexible for adapting to different domains.

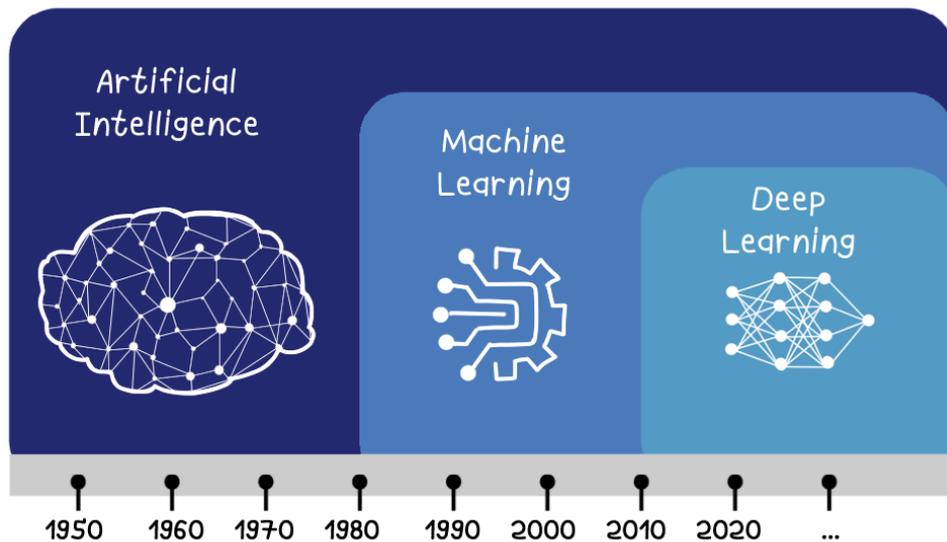


Figure 2.25: The divisions of artificial intelligence, machine learning and deep learning.

Deep learning is not a new concept but has a long history since the 1940s. Due to the development of hardware, algorithmic advances, and increasing datasets, deep learning has gained extreme popularity after the 2010s and has been applied to a variety of fields, from industry to research (such as computer vision [57–60], natural language processing [61–65], bioinformatics [66, 67], and medical imaging [10, 68–71]).

Particularly in the medical imaging research, deep learning has become the driven-

mainstream approach for clinical support owing to the increasing of the data in this field and the advantage of high accelerated reconstruction speed. Applications of deep learning have been broadly utilized in various types of medical images. For instance, a hybrid deep neural network was developed to classify the immunofluorescence image of Human Epithelial-2 (HEp-2) cells for autoimmune disease detection [72]. Based on chest x-ray images, the classification and interpretability of the recent global pandemic COVID-19 (SAR-CoV-2) [73] were also investigated using deep learning methods [69, 74, 75]. For CT images, deep neural networks was used for automatic segmentation for radiation therapy [68, 76]. Furthermore, deep learning has also been widely applied to MRI reconstruction with the high performance achievement [10–12, 77, 78].

2.2.13 Types of Learning Paradigms

Traditional deep learning based method solves the problem by learning featured transforms from the training data (labeled data) to predict the transform for the test data (unseen data). This is known as supervised learning with respect to influence of training labeled data to the test set in its learning paradigm [79, 80]. The other learning paradigms to extract featured transforms without the help of labeled data (unsupervised learning [81–83]) or combining data between labeled and unlabeled data (semi-supervised learning [84, 85]) can address the issue of insufficient labeled data. However, the non-supervised learning paradigms cannot assure the successful of any data distributions [86] and this learning paradigm could lead to deterioration the model performance [87]. Thus, to cope with unavailable or insufficient labeled data problem with supervised learning, the techniques such as transfer learning and data augmentation have been proven to be useful for this scenarios []. The applications of both transfer learning and data augmentation were also applied in this research and will be explained in chapter 3 and 4, accordingly.

In conclusion, the principle of the neural networks is summarized in this chapter. The understanding of the model parameters and network architecture benefit for choosing the proper network with a specific application. There is notwithstanding no best network architecture. Due to the constantly advancing in the field, a number of deep neural network architectures have been developed exponentially. A point of focusing in this thesis is to utilize prior knowledge to dynamic MRI with the help of fast inference speed of deep learning. Therefore, an overview of dynamic MR imaging is introduced in chapter ?? and a review of deep learning based approaches, particular with super-resolution for medical imaging, will be described in section 2.3.4.

2.3 Dynamic MRI and State-of-the-art

2.3.1 Dynamic MRI

Dynamic MRI is a valuable technique used in a wide range of applications in diagnosis, therapy and medical research. Dynamic MRI does not only acquire the data at one specific time-point but entails multiple image acquisitions and reconstructions. This is in contrast to regular MRI, which images the internal organs at one specific time. In addition, a regular diagnosed MR image can be considered as static, representing a photo; whereas a dynamic MR image is like a video recording of the internal organs. Some examples of common dynamic MRI procedures include Blood oxygenation level dependent (BOLD) - used for monitoring brain activity from oxygen concentration level [25], dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) - a collection of multiple breath hold images acquired for sequential images from multiple blood flow phases; arterial, venous, and delayed phases [88], dynamic chemical exchange saturation transfer (CEST) - detecting molecular changes based on chemical metabolic properties [89].

In interventional radiology (IR), which is an advanced procedure used for minimally invasive operations and image guidance for diagnosis and treatment [90], most common procedures includes angiography (arterial and venous), catheters insertions, percutaneous tumor treatment and needle biopsy. Due to these procedures demanding fast visualization speed of the target along with IR devices, such as needle, the imaging modalities typically used in interventions is fluoroscopy or C-arm Cone-beam CT (CBCT) - which are mobile imaging units able for real-time or near real-time interventions (e.g., 30 frames per second (fps) in available commercial C-arm systems [91]).

Nevertheless, fluoroscopy and other CT-like modality use ionizing radiation, which is harmful to patients and medical professionals. On the other hand, MRI is radiation-free ionizing, which offer an excellent soft tissue contrast. MR images are remarkable for specific lesions such as hepatocellular carcinoma (HCC) - a common type of liver cancer. MRI can be considered a main tool for IR operations in the case that other imaging modalities have poorer detection. Figure 2.26 is an example showing the superb lesion visibility of MRI compared to ultrasound and CT.

Among common usages of MRI includes liver, kidney, intestine, prostate, one fascinating use for MRI guidance is for liver lesions, where it is used to treat hepatic cancers or to classify liver metastases [93]. An illustration of interventional MRI (iMRI) is portrayed in figure 2.27. The procedures performed during dynamic imaging, such as MR-guided liver biopsy or MR thermometry, are highly dependent on the speed of real-time or near real-time supervision. Therefore, these procedures have certain demands for high spatial resolution and imaging speed. Nevertheless, a high spatial resolution image can be obtained

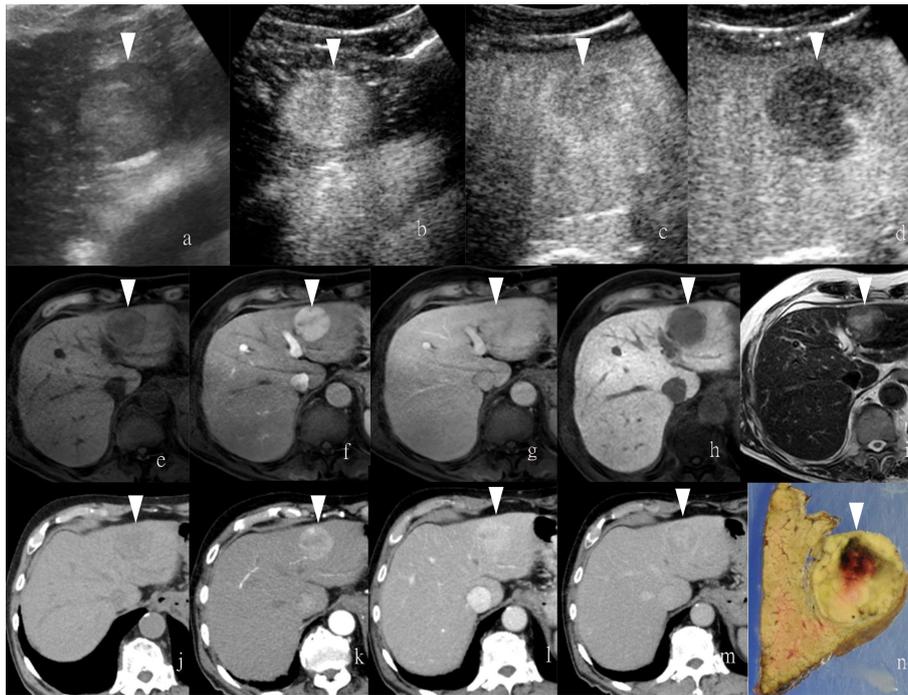


Figure 2.26: A comparison of the lesion visibility on MRI to ultrasound and CT, in different wash-in and wash-out phases. a-d: ultrasound, e-i: MRI and j-m: CT images. The arrows were marked the liver lesion at the same locations, together with (h) the specimen after operation [92].

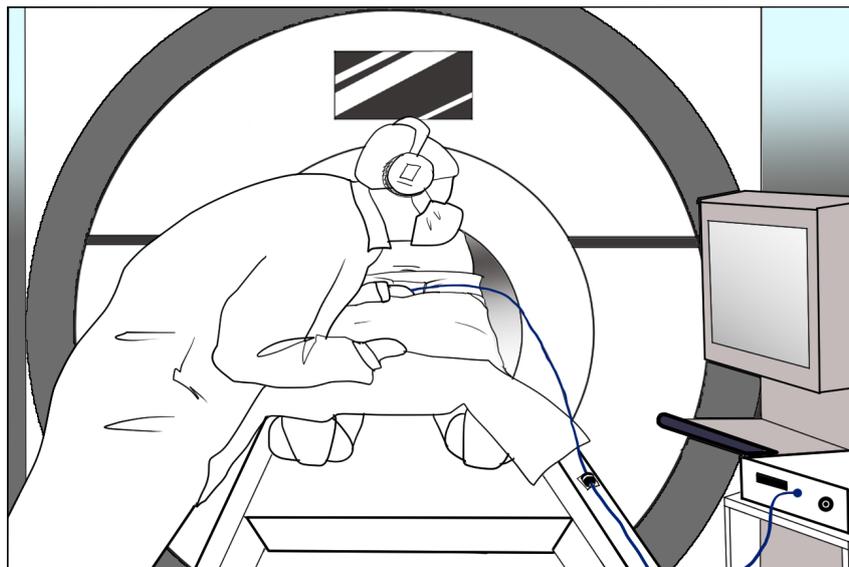


Figure 2.27: An example of MR-guided puncture in interventional radiology.

at the expense of temporal resolution.

2.3.2 Common Imaging Sequence in Abdomen

There is a need for particular imaging sequences for dynamic MRI. Several imaging sequences, such as T1-weighted fast spin echo (FSE), T2 FSE, contrast-enhanced T1W sequences or magnetic resonance cholangiopancreatography (MRCP), are commonly used for abdominal imaging in clinical. The reader can explore more in [27, 90, 94]. This section mentions only the techniques that are relevant to this thesis.

Fast T1w Imaging

This popular imaging technique is gradient echo based, in which the short TE minimizes the magnetic susceptibility effects. Example sequences of this fast T1-weighted imaging are Spoiled Gradient Recalled Echo (SPGR), T1-weighted Fast Field Echo (T1-FFE), Fast Low Angle Shot (FLASH), Volumetric Interpolated Breath-hold Examination (VIBE), and T1-weighted High-resolution Isotropic volume Examination (THRIVE). Their names correspond to specific MR vendors. Moreover, fat saturation is usually used in association with partial flip angle (less than 90°) technique to suppress the fat signal for a better visualization.

Dixon Method

A particular technique for fat and water differentiation in abdomen is called Dixon method [95]. Originally, a set of spin echo images were acquired for this method, while various imaging sequences such as gradient echo or steady state free precession can be used nowadays. The Dixon method is based on the fact that protons precess slightly slower in fat than in the water, resulting in the difference in resonance frequency, known as chemical shift. Practically, the in-phase and the oppose-phase (out-phase) images are acquired in the same acquisition and co-registered. Given S_{ip} and S_{opp} are the signal intensity of the in-phase and oppose-phase images respectively, the in-phase and oppose-phase can be defined by:

In-phase:

$$S_{ip} = S_w + S_f, \quad (2.18)$$

Oppose-phase:

$$S_{opp} = S_w - S_f, \quad (2.19)$$

where S_w and S_f denote the signal contributions of water and fat, respectively. Additionally,

water-only and fat-only images can be reconstructed by averaging the combination and differences of fat and water signals from each point. Figure 2.28 illustrates an example of gradient echo images based on the Dixon method.

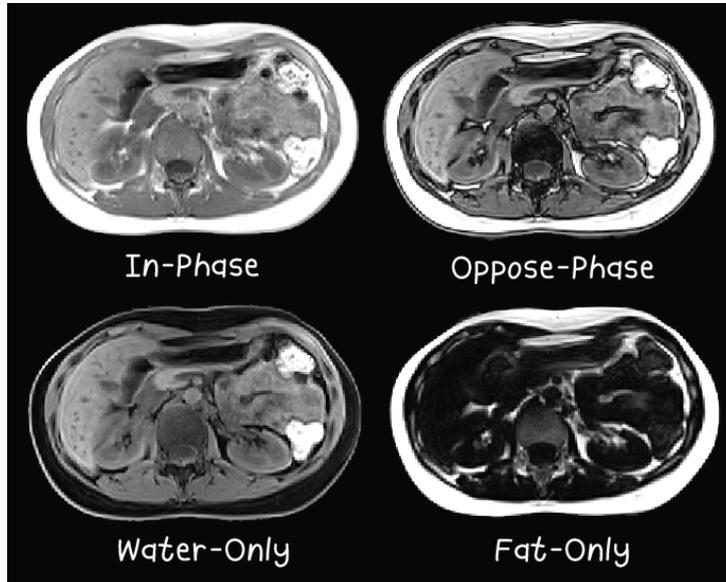


Figure 2.28: An example of gradient echo images based on the Dixon method: in-phase, oppose-phase (out-phase), water-only and fat-only images.

2.3.3 Spatio-Temporal Trade-off

Although MRI has the advantage of non-ionizing radiation and offering a unique tissue contrast, major challenges arise from its inherently slow acquisition. Furthermore, the long scan-time makes MR images susceptible to motion artifacts. Concerning the time constraint, the potential data to be acquired is restricted, leading to a trade-off between spatial and temporal resolution.

The optimization of the acquisition technique is necessary for capturing the kinetic property of the target lesions or organs. While the temporal resolution is crucial for fast imaging, sufficient SNR also needs to be maintained. The inadequate spatio-temporal resolution would compromise the image quality and could lead to poor diagnosis or treatment quality. Therefore, it is preferable to have a high temporal resolution for real-time applications [93].

Currently, general methods to deal with this trade-off include partial Fourier reconstruction, view-sharing technique, compressed sensing, and parallel imaging, all of which are considered state-of-the-art dynamic MRI and will be described in the following section.

2.3.4 State-of-the-Art Dynamic MRI Reconstruction

Partial Fourier Reconstruction

Partial Fourier reconstruction, also known-as half-Fourier reconstruction or $\frac{1}{2}$ NEX, only samples half of the k-space and then generate the other half during image reconstruction process. Examples of partial Fourier algorithms include phase-conjugate synthesis [96], Margosian [97], Homodyne [98], Cuppen [99], and projection onto convex set (POCS) [100]. Though half-Fourier imaging techniques take the advantage of not complicated and straightforward argument, the acceleration is limited and the SNR is compromised from reducing the scan-time while maintaining spatial resolution.

View-sharing Technique

View-sharing (VS) technique make use of repeated imaging data in multiple time frames by sharing a part of k-space over temporal phases. It is the commonly used techniques for DCE-MRI. The examples of VS techniques are Time-Resolved Imaging of Contrast KineticS (TRICKS, GE) [101], Time-resolved angiography With Stochastic Trajectories (TWIST, SIEMENS) [102], and differential subsampling with Cartesian ordering (DISCO) [103], for Cartesian sampling and k-Space weighted image contrast (KWIC) [104] for radial acquisition.

Nevertheless, the drawback of view-sharing is that it is prone to temporal blurring from motion across the shared phases. To deal with this issue, compressed sensing (CS) can eliminate the need for VS by recovering missing k-space data from pseudo-random undersampling, reducing temporal blurring while maintaining spatial resolution.

Compressed Sensing

Compressed sensing (CS) is a well-known technique that utilizes a sparsity or compressibility concept for reconstructing a signal sampled below Nyquist rate [4,105]. CS plays a vital role and emerges in several applications such as signal processing [106–108], acoustics [109–111] and video processing [112–114]. Compressed sensing MRI (CS-MRI) can substantially reduce the acquisition time and has been applied in various MR applications, including dynamic imaging [3,115–119]. The main components to recover the missing k-space from undersampled data (violating Shannon-Nyquist theorem [120,121]) are sparse transform (e.g., wavelet or Fourier), incoherent measurement, and non-linear reconstruction. The reconstructed image can be obtained by solving the constrained optimization in equation 2.20 [122].

$$\min_X = \frac{1}{2} \|y - Ax\|^2 + \lambda \|\Psi x\|_1, \quad (2.20)$$

where y denotes the measurement data from MR scanner, A refers to the imaging model, x is the reconstructed image, λ is the regularization parameter, and Ψ is an operator that transforms Ψx to be sparse.

Furthermore, by using random sampling, CS could overcome the blurry image issue raised by sharing the data over temporal phases in VS. However, CS has some restrictions that should be considered. For example, the efficiency of compressed sensing depends on the selected sparsifying transform and iterative regularization, the iterative computing requires an expensive computation, and the optimal parameters tuning for each case is time-consuming. Therefore, these reasons could hamper the workflow in clinical interventions.

Parallel Imaging

Parallel imaging (PI) is another technique to shorten the acquisition time by using phased array coils, which have multiple receiver channels, to reconstruct the missing part of k-space data [9].

The signal near the receiver coil has a strong signal and will decrease for faraway locations. The image formation can be performed via image space or frequency domain (k-space). The general equation of parallel imaging in MRI can be given by:

$$FT(f c_l) = F * C_l, \quad (2.21)$$

where FT is Fourier transform operation, f is the object, c_l is the coil sensitivity, and F and C_l are the parameter represented in the k-space for the object and coil sensitivity for each individual coils respectively. The examples of parallel imaging techniques are sensitivity encoding (SENSE) [8], generalized simultaneous acquisition of spatial harmonics (SMASH) [7], generalized autocalibrating partial parallel acquisitions (GRAPPA) [9], CAIPIRINHA [123, 124], k-t BLAST and k-t SENSE [125], k-t GRAPPA [126] and SPEAR [127]. Additionally, to further obtain a higher acceleration, parallel imaging is combined with compressed sensing such as CS-SENSE [128–130].

Although parallel imaging can reduce the scan-time significantly by using multiple coils, the imaging speed has to compensate for the signal-to-noise ratio (SNR). For example, given the acceleration factor - the ratio of the amount of data for fully sampled to the amount of acquired data - of 2 (i.e. every other line of k-space is acquired), the SNR is reduced by $\sqrt{2}$ or about 40%. However, with the limitation of devices and accessibility

during interventions - that require space for interventional equipment to perform delicate surgical procedures - PI may not be optimal for interventions.

Super-resolution with Deep Learning

Super-resolution (SR) a signal processing technique in computer vision broadly used in several different applications, such as face recognition [131,132], satellite imaging [133], and medical imaging [134,135]. SR allows the restoration of a high-resolution image from the corresponding low-resolution image. Such resolution enhancement techniques were developed to address the physical limitation of imaging sensors [136,137]. Traditionally, interpolation methods, such as the nearest neighbor or known as round interpolation, are used to upsample the pixel data by approximating the pixel values from its neighbors. A pictorial illustration in figure 2.29 shows an example of nearest neighbor interpolation.

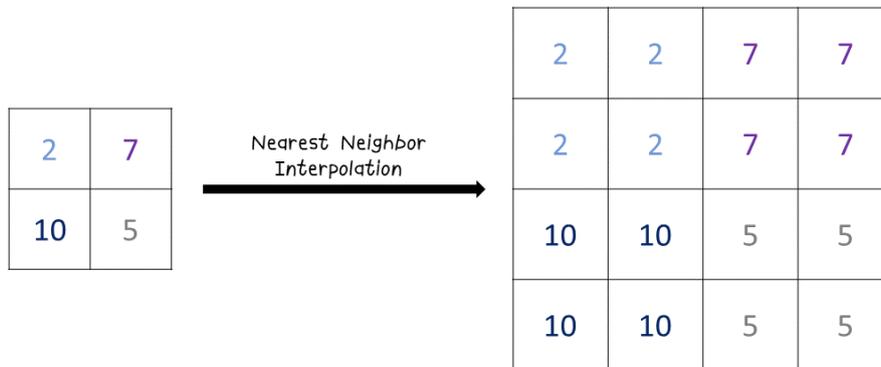


Figure 2.29: A pictorial illustration of nearest neighbor interpolation from 2x2 matrix to 4x4 matrix. This interpolation method is the simplest interpolation that upsample the missing data from the neighbor's pixel values without adding new values.

In the last few years, deep learning methods has revolutionized the research community with the advantage of fast learning and inference. Super-resolution algorithms using deep learning have been developed and successfully applied in a variety of fields [138–148], including medical imaging such as MRIs [149–151]. The well-known SRCNN framework [152], presented in figure 2.30), is originally the first few works on applying convolutional neural networks in super-resolution reconstruction. On the contrary to traditional interpolation based methods, deep learning based methods could overcoming blurring and noisy image issues that often appear from performing image interpolation [153–156].

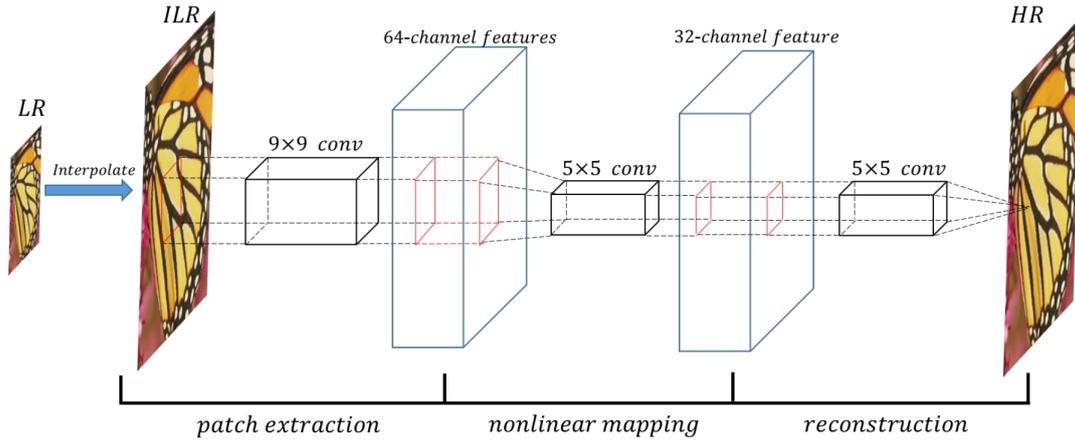


Figure 2.30: A graphical illustration of SRCNN architecture, consisting only three parts: patch extraction, nonlinear mapping and reconstructions [152].

Challenges of Super-resolution in MRI

Existing super-resolution techniques can be classified into single image super-resolution (SISR) and video super-resolution (VSR). Several studies have demonstrated the ability of deep learning based super-resolution to improve the time-series images (e.g., dynamic imaging) by incorporating temporal information [14–16]. Nevertheless, such a VSR approach, which could exploit the temporal information from a sequence of images to enhance the spatial resolution and frame rate [157–159], still has a gap in applying it to MRI applications, particular dynamic MRI. Moreover, one of the main challenges for deep learning based method lies in the need for a large training dataset, and available dynamic MRI datasets are limited. Such artificial data generations [160–162] and patch-based methods [163–166] are commonly used for addressing the training data limitation issues. Our studies also applied these techniques to cope with this challenge; the methodological details are described in chapters 5 and 6.

Loss Functions in SR reconstruction

The SR image reconstruction is generally considered as ill-posed problem according to the ill-conditioned of blur operator and the lack of low-resolution images. To solve this inverse ill-posed problem, typical SR models are trained to super-resolve the missing image information by optimizing the loss between super-resolved image (the reconstructed high-resolution image) and its ground-truth. The loss function is crucial and can affect the network performance. However, the commonly loss function used in neural network for image processing is based on pixel-based losses, such as L1 (absolute loss) [] or L2 (squared loss) [], which has poor correlation to human perception [167]. The challenge in selecting the loss function for medical image application is to choose the loss metric that can generate diagnostic quality results. Therefore, a number of research on investigating loss functions has increasing, particularly for loss in terms of human perspective [167–170].

In this dissertation, the combination of perceptual loss with a modified U-Net architecture has been investigated and will be explained in section 3.3.5 in chapter 3. The similar architecture of this network and perceptual loss is also described in section 4.3.5 in chapter 4.

3 Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Fine-tuning with a Subject-specific Scan

“Discoveries are made by pursuing possibilities suggested by existing knowledge.”

– Michael Polanyi

This chapter is based on the following published journal:

- **Sarasaen, C.**, Chatterjee, S., Breitkopf, M., Rose, G., Nürnberger, A. and Speck, O. "Fine-tuning deep learning model parameters for improved super-resolution of dynamic MRI with prior-knowledge." *Artificial Intelligence in Medicine*, 121, 2021: 102196 [70].

The preliminary results for the following publications were presented at conferences:

- **Sarasaen, C.**, Chatterjee, S., Nürnberger, A., and Speck, O. "Super resolution of dynamic mri using deep learning, enhanced by prior-knowledge." *In 37th Annual Scientific Meeting Congress of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), 33(Supplement 1): S03.04, S28-S29*. Springer, 10 2020 [147].
- **Sarasaen, C.**, Chatterjee, S., Saad, F., Breitkopf, M., Nürnberger, A., and Speck, O. "Fine-tuning deep learning model parameters for improved super-resolution of dynamic mri with prior-knowledge." In Proc. The Annual Meeting of The International Society for Magnetic Resonance in Medicine (ISMRM), 5 2021 [148].

3.1 Synopsis

Deep learning based super-resolution (SR) demonstrates the ability to enhance spatial image resolution from relatively low-resolution images. SR can be useful for dynamic MRI reconstruction, which suffers from a spatial and temporal trade-off. Moreover, before dynamic scans for interventional procedures, a set of static high-resolution planning scans is typically acquired. This planning scan is utilized in this work as prior knowledge

by incorporating it into the proposed Fine-tuned SR DynMRI framework to tackle the spatio-temporal trade-off. A 3D UNet model with a perceptual loss network was used for main training on a benchmark dataset and fine-tuned with a subject-specific planning scan for only one epoch. The inference step was tested with 3D dynamic scans ($n=3$), acquired with different imaging parameters. The performance of the framework was evaluated using low-resolution dynamic images obtained from different in-plane undersamplings (25%, 10% and 6.25% of the k-space center). Results show that SR main training reconstructs the images sharper than the baseline methods. Furthermore, the Fine-tuned SR DynMRI (after fine-tuning the SR main training) achieved better performance surpassing the SR main training, even for relatively low-resolution images. The SSIM values of low-resolution images from 6.25% of the k-space, which is the lowest resolution investigated in this study, before and after fine-tuning were 0.939 ± 0.008 and 0.957 ± 0.006 (average value \pm SD), respectively. Therefore, the proposed Fine-tuned SR DynMRI framework illustrated the potential to improve the SR reconstruction of dynamic MRI, and might be beneficial for fast imaging applications such as interventions.

3.2 Related Work

Spatial and temporal trade-off in for deep learning based-approaches

In MR-guided interventions, such as liver biopsy, the dynamic imaging technique is crucial for monitoring the dynamic movements of target organs. However, as discussed above in 2.3.3, the challenge imposed by spatial and temporal resolution trade-off is derived from the inherently slow speed of MR image acquisition. To achieve a high temporal resolution, the amount of data to be acquired has to be reduced, which could lead to the loss of spatial resolution. Although several techniques dealing with this spatio-temporal trade-off [3, 4, 119, 125] are available, their reconstruction speed could still create a hindrance for real-time or near real-time imaging. With the fast inference of deep neural networks in recent works [140, 147, 152, 158, 159], the reduction of reconstruction time, while restoring image resolution, could be expected from deep learning based super-resolution.

Patch-based super-resolution

Recent research indicates that deep learning-based approaches are valuable tools for MR image reconstruction, including for dynamic MRI [10–12, 78]. However, most deep learning based techniques have challenges such as a requirement for large amount of training datasets. Moreover, it is crucial to keep in mind that when using a training set that is different from the test set, poor quality results can arise [171, 172]. The problem of small datasets in deep learning can be addressed using different methods, including data

augmentation and synthetic data generation [160–162].

The problem with these methods is that they involve manipulating the data artificially in order to enlarge the dataset’s size. The small dataset problem can also be addressed through patch-based training by dividing data into smaller patches. Therefore, the patch-based technique allows the increasing of the number of samples in the dataset effectively without modifying the data artificially. The patch-based super-resolution (PBSR) algorithms [163–166, 173–176] are able to map the function of a given pair of patches between low-resolution image and the corresponding high-resolution image without modifying the data artificially. Thus, PBSR is able to handling the issue of lacking abdominal MRI for training in this study [149, 150, 177].

U-Net and perceptual loss

After the U-Net model [178] for medical image segmentation in the 1980s was introduced, it was employed in many applications, and shown to be a valuable tool in solving inverse problems [179–182]. Moreover, the U-Net based models were developed for super-resolution (SR) reconstruction [143–145, 183–185]. Furthermore, for training a deep learning model, the choice of a loss function is important for reconstruction error calculation between the predicted and the ground-truth image. It is common to use pixel-based loss functions for SR. However, in terms of perceptual quality, these functions generate over-smooth results, because they are incapable of recovering the high-frequency information [168, 186–188]. Since the human visual system can recognize details and specifically identify differences in structures, a perceptual loss function that is known to be related to human perception, has been applied in deep learning tasks, such as in style transfer and super-resolution applications [168, 189, 190] to achieve high perceptual images. Additionally, some research has demonstrated the efficiency of perceptual loss in reconstructing MRI as well [170, 191, 192]. Hence, this work combines the 3D U-Net and perceptual loss for dynamic MRI to gain high quality results perceptually.

Transfer learning and prior knowledge

In deep learning, the transfer learning process involves re-proposing a pre-trained model with fine-tuning [193]. Transfer learning uses the pre-trained network weights learning from one task to train the network for a new task. In this way, the network is fine-tuned and ready for another similar task. Transfer learning has been broadly applied in computer vision [194–198] and natural language processing [199–202] to address the issue of insufficient training data [203, 204]. It has been shown that fine-tuning significantly improves the neural network performance and allows it to converge in fewer epochs with a small dataset [151, 205, 206]. Prior knowledge refers to information that is accessible or

was available prior to a given task. For example, there is accessible MR data obtained prior to a specific examination, such as static high-resolution planning scans or so-called preinterventional scans [90]. Therefore, the prior knowledge from a subject-specific planning scan is incorporated into transfer learning. This research intends to improve spatial image resolution on the super-resolved images by fine-tuning the deep learning model parameters.

3.3 Methodology

In this research, the proposed Fine-tuned SR DynMRI framework is applied for reconstructing the abdominal dynamic MRI. The problem statement of super-resolution reconstruction and fine-tuning are discussed in this section.

3.3.1 Super-resolution Reconstruction

With regards to a simple form of single image super-resolution (SISR), a low-resolution and a corresponding high-resolution image can be denoted as I_{LR} and I_{HR} . The reconstructed high-resolution image \hat{I}_{HR} , or the so-called super-resolved image, can be recovered from a super-resolution reconstruction using the following equation [207]:

$$\hat{I}_{HR} = F(I_{LR}; \theta), \quad (3.1)$$

where F denotes the super-resolution model that maps the image counterparts, and θ denotes the parameters of F . For the patch-based super-resolution, I_{LR} can be obtained by modifying the equation 3.1 with:

$$I_{LR} = I_{LR_M} \times N, \quad (3.2)$$

where I_{LR_M} is a patch of low-resolution image and N is the number of patches used in the reconstruction model. The problem of super-resolution reconstruction is an ill-posed problem, thus, a network model aims to solve the loss function:

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}(\hat{I}_{HR}, I_{HR}) + \lambda R(\theta), \quad (3.3)$$

where the former term is the data term, and the latter is the optimization term. $\mathcal{L}(\hat{I}_{HR}, I_{HR})$ denotes the loss function between the approximated HR image \hat{I}_{HR} and ground-truth

image I_{HR} , $R(\theta)$ is a regularization term and λ denotes the trade-off parameter.

3.3.2 Problem Statement of Fine-tuning

The SR model in this study is not only trained with the training data but also fine-tuned with a static high-resolution image as prior-knowledge. This strategy attempts to transfer the knowledge of the learned weights and fine-tune the trained model through shared parameters.

Let X_a is the feature space of Dixon images from main training domain A and X_b is the feature space of T1w image from 3D dynamic MRI domain B , which will be fine-tuned. $P(x_a)$ and $P(x_b)$ are marginal probability distribution of domain A and B according. As the two domains are similar, it can be assumed [205] that

$$X_a \approx X_b \quad \text{but} \quad P(x_a) \neq P(x_b). \quad (3.4)$$

Therefore, fine-tuning meant to fill the gap between the two domains by optimizing the distributions of both marginal probabilities:

$$P(x_a) = P(x_b). \quad (3.5)$$

3.3.3 Experimental Datasets

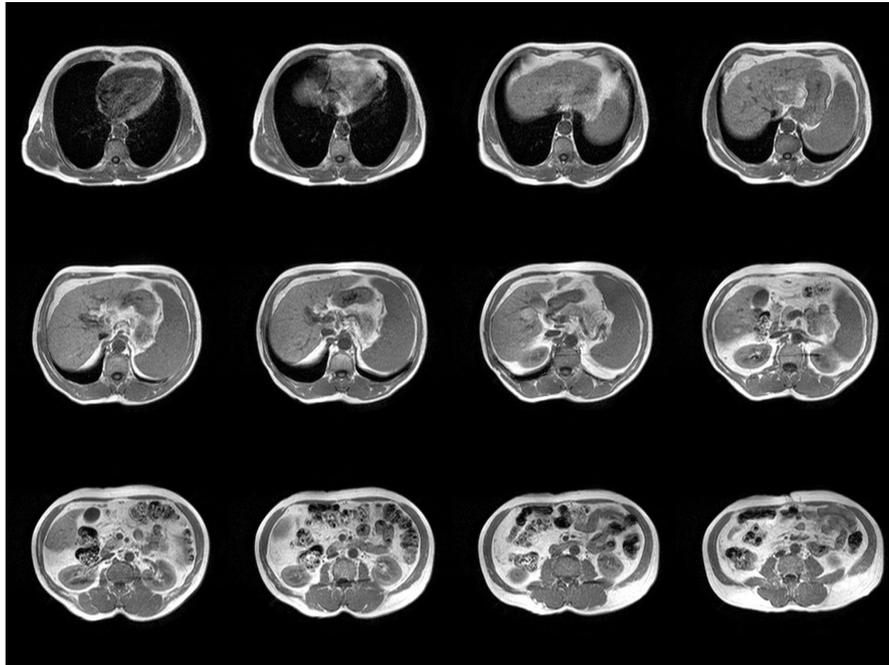
In this work, three sets of abdominal MRI were employed for main training, fine-tuning and inference. The MR volumes were used as high-resolution ground-truth images I_{HR} . The generation of their corresponding low-resolution images I_{LR} was performed by artificially undersampled in-plane for different undersampling levels, see section 3.3.4.

Main Training Data

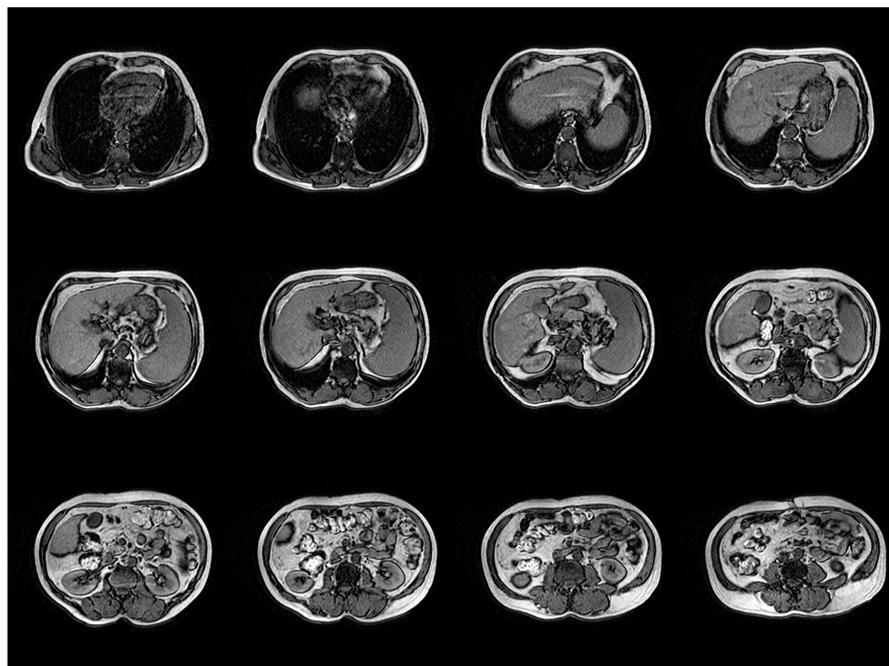
The CHAOS dataset [208], consisting of T1-Dual images for 40 subjects, in- and opposed-phase images for each subject, was used for the main training (see figure 3.1). Accordingly, the total 80 volumes were separated with a 70:30 ratio for training and validation data.

Fine-tuning Data

Three healthy subjects were scanned on a 3T MRI (Siemens Magnetom Skyra). The high-resolution 3D static images during breath-hold (for one time-point) were acquired



(a) T1w In-phase image



(b) T1w Opposed phase image

Figure 3.1: CHAOS dataset T1-dual images. (a) T1w in-phase image and (b) T1w opposed-phase image.

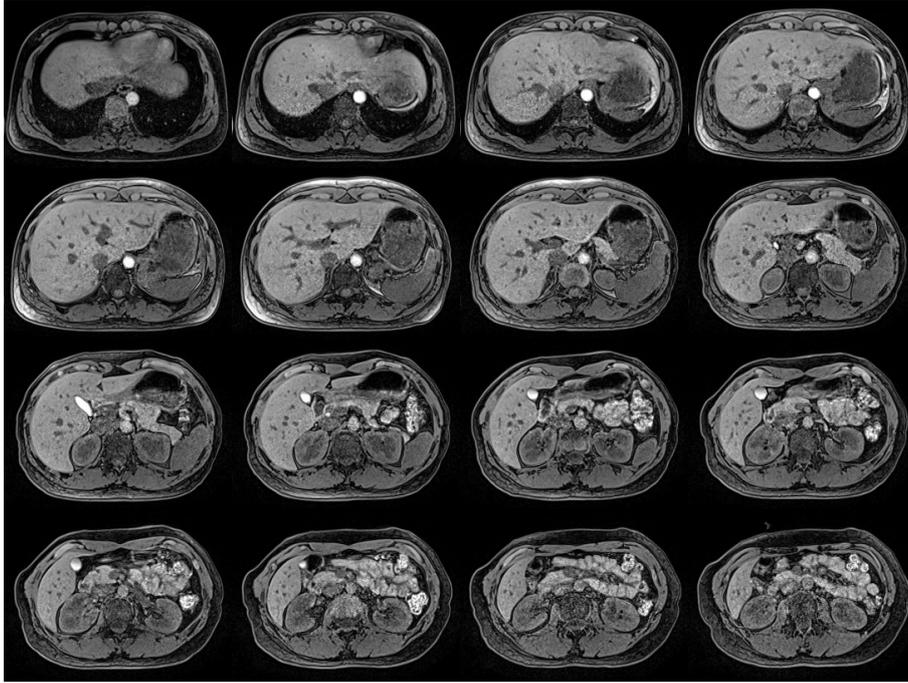


Figure 3.2: Fine-tuning data. An example of subject-specific high-resolution image (breath-hold) for different slices.

using the T1w FLASH sequence to fine-tune the trained model subject-specifically. Figure 3.2 illustrates an example of fine-tuning data.

Inference Data

The model was evaluated using 3D dynamic images of three subjects to simulate the low-resolution dynamic images in the section 3.3.4. For each subject, the dynamic acquisitions were acquired during free-breathing for ten time-points (TPs) and were handled as an individual 3D Volume for feeding to the network. Although all subjects were acquired using the same T1w FLASH sequence, some imaging parameters, such as FOV, iPAT [209] or fat suppression, were adjusted differently for each subject to test the network’s generalization. It is worth mentioning that the dynamic and static scans were obtained at a different scan session, similar to a planning scan in actual clinical practice. The acquisition protocols of CHAOS, 3D subject-wise dynamic and static datasets can be found in Table 3.1.

3.3.4 Undersampled Data Generation

All datasets, the CHAOS dataset (for main training), the static 3D scans (for fine-tuning), as well as the dynamic 3D scans (for inference), were artificially undersampled in-plane for three different levels. These low-resolution images were considered as the input to the network, and its high-resolution images were used as ground-truth images for comparison.

Table 3.1: The acquisition protocols of the experimental dataset including CHAOS dataset of 40 subjects (total 80 volumes of in- and opposed-phase images), and subject-wise 3D dynamic scans of three subjects for ten time-points (TPs). Note that the dynamic and static scans were acquired using the same sequence at a different scan session but the latter was acquired for one TP.

	CHAOS (40 Subjects)	Subject 1	Subject 2	Subject 3
Sequence	T1 Dual In-Phase & Opposed-Phase	T1w Flash 3D	T1w Flash 3D	T1w Flash 3D
Resolution	1.44 x 1.44 x 5 - 2.03 x 2.03 x 8 mm^3	1.09 x 1.09 x 4 mm^3	1.09 x 1.09 x 4 mm^3	1.36 x 1.36 x 4 mm^3
FOV x, y, z	315 x 315 x 240 - 520 x 520 x 280 mm^3	280 x 210 x 160 mm^3	280 x 210 x 160 mm^3	350 x 262 x 176 mm^3
Encoding matrix	256 x 256 x 26 - 400 x 400 x 50	256 x 192 x 40	256 x 192 x 40	256 x 192 x 44
Phase/Slice oversampling	N/A	10/0 %	10/0 %	10/0 %
TR/TE	110.17 - 255.54 ms / 4.60 - 4.64 ms (In-Phase) 2.30 ms (Opposed-Phase)	2.34/0.93 ms	2.34/0.93 ms	2.23/0.93 ms
Flip angle	80°	8°	8°	8°
Bandwidth	N/A	975 Hz/Px	975 Hz/Px	975 Hz/Px
iPAT (GRAPPA factor)	None	2	None	None
Phase/Slice partial Fourier	N/A	Off/Off	Off/Off	Off/Off
Phase/Slice resolution	N/A	75/65 %	75/65 %	50/64 %
Fat suppression	N/A	None	On	On
Time per TP	N/A	5.53 sec	11.76 sec	8.01 sec

Table 3.2: Effective resolutions and estimated acquisition times (per time-point) for different levels of undersampling (25%, 10% and 6.25% of the k-space).

	Subject 1		Subject 2		Subject 3	
	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)
High-resolution (Ground-truth)	1.09 x 1.09 x 4	4.81	1.09 x 1.09 x 4	9.61	1.36 x 1.36 x 4	6.62
25% of k-space	2.19 x 2.19 x 4	1.22	2.19 x 2.19 x 4	2.43	2.73 x 2.73 x 4	1.65
10% of k-space	3.50 x 3.50 x 4	0.47	3.50 x 3.50 x 4	0.94	4.38 x 4.38 x 4	0.66
6.25% of k-space	4.38 x 4.38 x 4	0.28	4.38 x 4.38 x 4	0.56	5.47 x 5.47 x 4	0.42

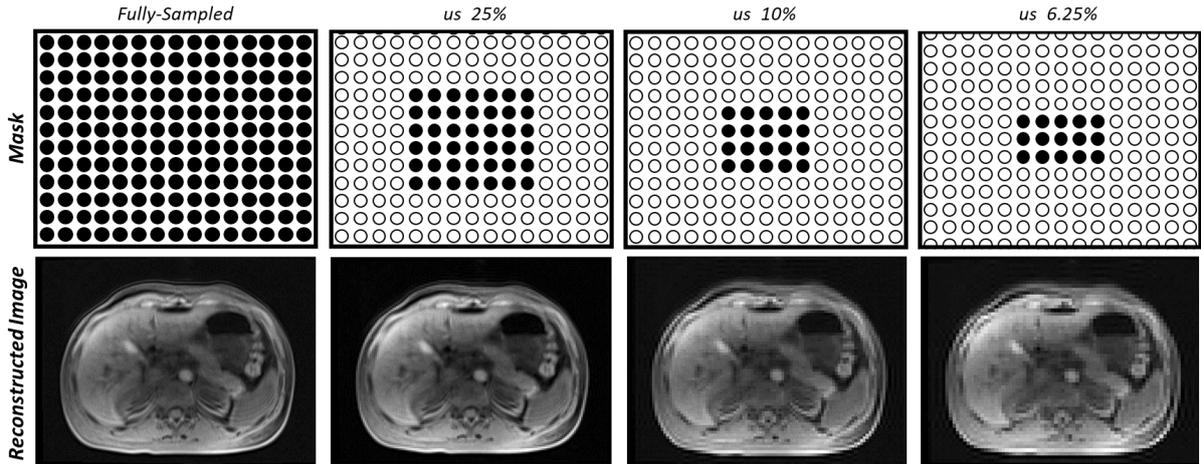


Figure 3.3: Graphical representation of masks and the corresponding reconstructed images. The undersampled (us) masks were generated by taking only 25%, 10% and 6.25% of the k-space center. The data points in black denote the sampled point, and the white dots denote the undersampled data.

A representation of the undersampling masks and the reconstructed images can be seen in figure 3.3. A low-resolution data set has been generated by undersampling the k-space center without zero padding and taking 25%, 10% and 6.25% of it, but the data is not padded in the phase-encoding and read-out directions. Considering the amount of data used for the SR reconstruction, this could theoretically equals to an acceleration factor (2.3.4) of 4, 9 and 16, respectively, depending on the amount of data used.

Table 3.2 provides the effective resolutions and estimated acquisition times of different undersampled levels based on their corresponding high-resolution images. The estimation of the scan times can be calculated as [30]:

$$\text{Scan time}_{3D \text{ volume}} = N_{AQ} \times TR \times N_{PE} \times N_{PART}, \quad (3.6)$$

which N_{AQ} is the number of acquired line data, TR is the repetition time, N_{PE} is the number of phase encoding lines - which equals to the matrix dimension in this direction, and N_{PART} is the number of divided partitions (see section 2.1.6 for further explanations).

3.3.5 Fine-tuned SR DynMRI Framework and Network Architecture

Figure 3.4 portrays the framework overview, which consists of main training, fine-tuning and inference. The network architecture in this work is shown in figure 3.5. The network is based on a 3D version of U-Net [147,178,210] with a combination of perceptual loss network [211]. To reconstruct a super-resolved image \hat{I}_{HR} , the network takes a low-resolution image patch I_{LRM} as an input. It is worth mentioning that the input and output needs to

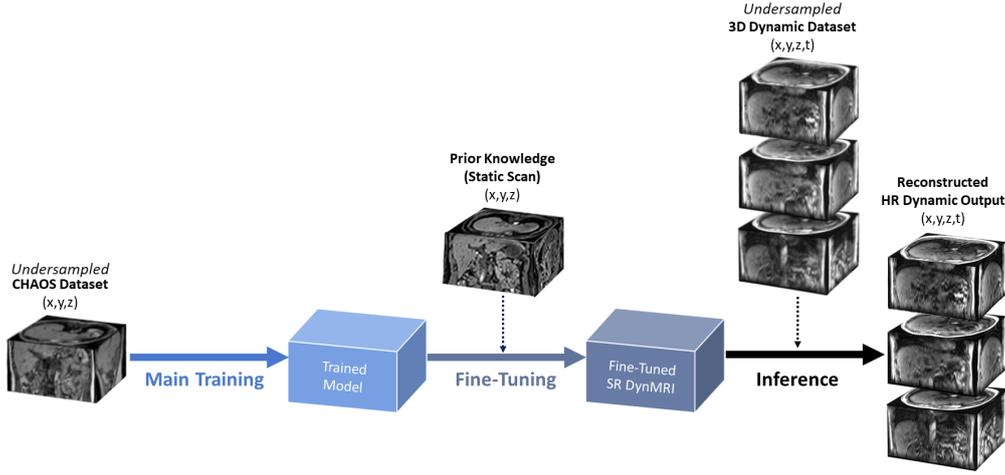


Figure 3.4: Method Overview. The main steps in Fine-tuned SR DynMRI Framework: Main training, fine-tuning and inference.

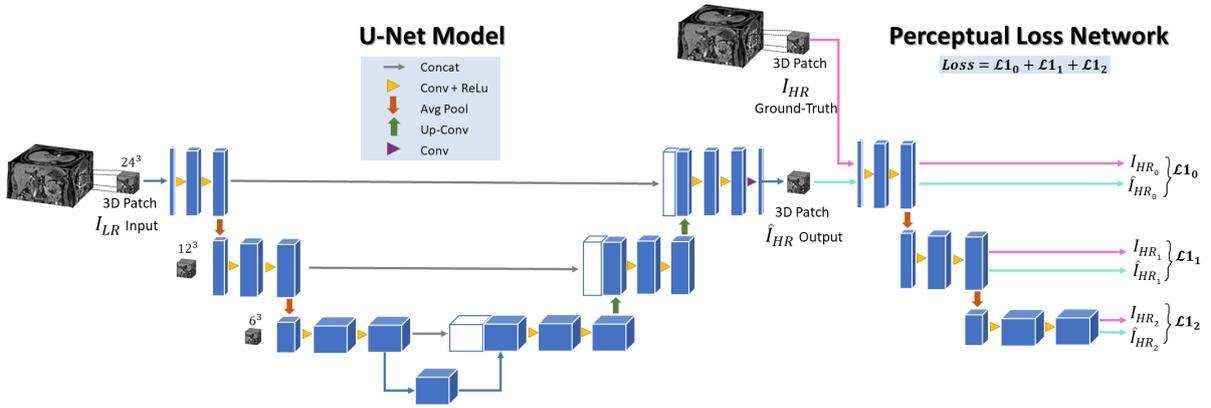


Figure 3.5: The proposed network architecture based on 3D U-Net combined with perceptual loss network. The super-resolution training was performed using 3D patches (24^3 size) as the input. The size of patch is provided on the left side to each block (feature map) of each layer. The perceptual loss network consists of three levels of the contraction path that was pre-trained on MRA.

have the same size before feeding to a U-Net model. Therefore, trilinear interpolation was applied on a low-resolution image patch before feeding it to the network.

Basically, U-Net architecture contains two main parts concatenated by skip connections at each equivalent level. The first path (on the left) is a contracting path, which encodes the features and down-samples the input size by half. The contracting path consists of two convolutions, each followed by a rectified linear unit (ReLU) activation function and a max pooling. The second path (on the right) is an expanding path, which decodes the features from the network and up-samples the input size by a factor of 2. The expanding path consists of two up convolutions with ReLU activation functions. Only at the last layer was a convolution without an activation function used to reduce the number of feature channels and generate the final output. In this work, the modified 3D U-Net has three

levels of feature contraction and expansion, and average pooling operations were used. Instead of using convolutional layers in the original U-Net architecture, the convolutional transpose layers were applied in the expanding path.

To super-resolve image with patch-based techniques, this framework was implemented using 3D patches of the volumes. CHAOS patches were used for the main training, with a 3D patch size of 24^3 for slice dimensions and strides of 6 for other dimensions. Following that, the network was fine-tuned using a static 3D scan of the same subject from an earlier session, see figure 3.4. Furthermore, a patch size of 24 and a stride of one were applied for fine-tuning and evaluation. When inference is performed using patch-based training, there is a possibility to have patching artifacts. Therefore, a stride of one is set in the inference stage, and the overlapped parts are averaged after reconstruction to remove these artifacts.

Perceptual loss is a term of error calculation between output and ground-truth that focuses on perceptual differences. In this work, the perceptual loss was determined for calculating the loss during training and fine-tuning. The first three blocks of a frozen pre-trained U-Net MSS model, on 7T MR angiography (MRA) for vessel segmentation [211], was used as the perceptual loss network (PLN) for extracting tiny feature abstractions from the final super-resolved output of the model and the ground-truth images. According to [170], perceptual loss networks such as VGG-16 can also be trained on three-channel RGB (ImageNet dataset - non medical images), even for medical imaging tasks. This work used this pre-trained network as it was previously trained on single-channel medical images. Although these MRA images were acquired for the brain, not the abdomen, and have different contrasts from the data experimented on in this research, they had better usage potential than the one trained on three-channel RGB images. L1 loss (mean absolute error) was computed to compare the extracted features from the network’s output and the ground-truth images. Finally, after adding up the losses for each feature, the total losses were back propagated. This perceptual loss L between a ground-truth image $\mathbf{y}_{s,t}$ and the corresponding predicted image $\hat{\mathbf{y}}_{s,t}$ can be formulated as:

$$L_{\mathbf{y}_{s,t},\hat{\mathbf{y}}_{s,t}} = \sum_{b=1}^B \sum_{f=1}^{F_b} |f_{\mathbf{y}_{s,t}} - f_{\hat{\mathbf{y}}_{s,t}}|, \quad (3.7)$$

where B is the number of blocks of the PLN to be used for loss calculation, F_b is the number of features block b can generate (depending upon the network architecture), $f_{\mathbf{y}_t}$ is a particular feature generated from the ground-truth \mathbf{y}_t , and $f_{\hat{\mathbf{y}}_t}$ is the corresponding feature generated from the prediction $\hat{\mathbf{y}}_t$.

3.3.6 Implementation Details

A 3D patch from the CHAOS dataset was used to train the model (main training) with a patch size of 24^3 and a stride of 6 for the slice dimension and 12 for the other dimensions. The network was subsequently fine-tuned using a 3D static scan for each subject, see figure 3.4. A patch size of 24^3 and a stride of 1 were used for fine-tuning and evaluation. The implementation using PyTorch was done using Nvidia Tesla V100 GPUs. During the training, Adam optimizer [212] was used, the learning rate was 0.0001, and the predefined epochs was 200. After the main training converged, the fine-tuning was performed on only one epoch using a subject-specific scan and a smaller learning rate of 0.000001. In addition to being high-resolution, the subject-specific static scans have the same resolution, contrast, and volume coverage as high-resolution dynamic scans. In order to keep the simulation similar to a real-life scenario and to keep inference speed high, the static and dynamic scanning were not co-registered. The code of this implementation is publicly available on GitHub ¹. The tutorial of the code is accessible on Google Colab Notebook (see appendix C)).

3.3.7 Image Reconstruction Evaluation

For the purpose of evaluating the quality of reconstructed images relative to ground-truth HR images, two of the most widely used metrics were chosen - the structural similarity index (SSIM) and the peak signal-to-noise ratio (PSNR). The standard deviation of the different images (diff SD) was also calculated. The high SSIM and PSNR values with minimal diff SD indicate the high performance of the reconstruction quality. The formulas of evaluation metrics used in this study can be found in appendix B. Finally, a statistical test was calculated using the paired t-test and the Wilcoxon signed-rank test.

3.4 Results

The performance and limitation of the proposed framework were validated using three levels of low-resolution images obtained from different undersampled levels (25%, 10% and 6.25% of the k-space center). The improvement of the reconstructed dynamic MRI was evaluated by comparing SR results before and after fine-tuning. The evaluation was performed on the dynamic abdominal dataset of three subjects, consisting of 10 TPs for each subject. The proposed method was compared against the baseline methods, including traditional trilinear interpolation, area interpolation (based on adaptive average pooling, as implemented in PyTorch), and Fourier interpolation (zero-padded k-space, or so-called sinc interpolation), which is a widely used technique in MRI.

¹<https://github.com/soumickmj/FTSuperResDynMRI>

The results show that the proposed Fine-tuned SR DynMRI improved both qualitative and quantitative data, even for reconstructing the relatively low-resolution image - taking only 6.25% of the k-space. A qualitative comparison of the SR results after fine-tuning against the baseline methods for all undersampling levels is illustrated in Figure 3.6. Another qualitative result of the proposed method was evaluated by comparing the SSIM Maps of the low-resolution input of the lowest resolution investigated in this work (6.25% of the k-space) with super-resolution (SR) results after fine-tuning. The SSIM maps were calculated against the high-resolution ground-truth - the respective SSIM outputs can be found on top right of the image in Figure 3.7. These SSIM maps show in which part of the image the difference between the images is higher or lower locally. Large values of local SSIM values appear as bright pixels in the local SSIM map, portraying the better quality according to the given measurement [213]. Moreover, to provide a more precise visualization, Figure 3.8 portrays example results from one subject in two regions of interest (ROIs). The reconstructed SR results of main training, Fine-tuned SR DynMRI, and trilinear interpolation were compared against its corresponding ground-truth. It was also found that the undersampling artifacts presented in the SR results of main training were mitigated in the SR results after fine-tuning, resulting in edge enhancement even for small structures.

For the purpose of quantitative analysis, Table 3.3 shows the comparison of the proposed framework with the baseline methods using SSIM, PSNR and standard deviation of the difference images with ground-truth (average values \pm SD) of all subjects for all time-points. Figure 3.9 shows the distribution of the resultant metrics over the different time-points at each subject. The error bands denote a 95% confidence interval of resultant values over time-points. The first row depicts the SSIM values, and the second shows the PSNR values. The columns are the results of 25%, 10%, and 6.25% of the k-space, respectively. These plots show that the proposed method (SR after fine-tuning) noticeably outperforms all of the baseline methods investigated in this study.

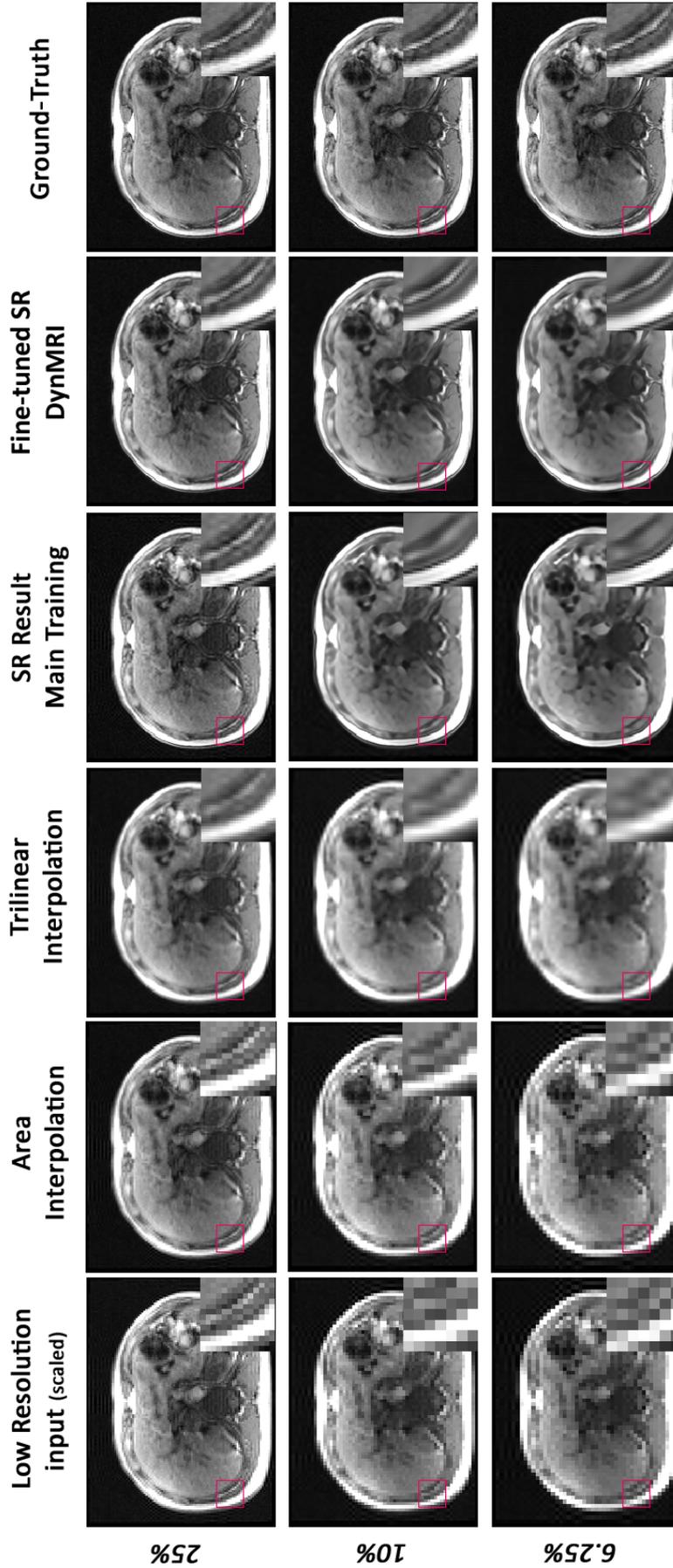


Figure 3.6: A qualitative comparison of low-resolution dynamic images (undersampling 25%, 10% and 6.25% of k-space center) on the same slice. Images at lower resolutions (scaled up), trilinear interpolated input, super-resolution results of main training (SR Results Main Training), Fine-tuned SR DynMRI, and their ground-truth images are shown from left to right.

Table 3.3: Quantitative results of the proposed framework compared with baseline methods in terms of SSIM, PSNR and standard deviation of the difference images with ground-truth (average values \pm SD). The p-values < 0.001 for all subjects, resolution-wise and for all resolution with subject-wise.

Data	25% of k-space		
	SSIM	PSNR	diff SD
Area Interpolation	0.911 \pm 0.011	29.721 \pm 1.948	0.068 \pm 0.011
Trilinear Interpolation	0.964 \pm 0.005	37.680 \pm 1.770	0.013 \pm 0.002
Zero-padded	0.977 \pm 0.013	37.980 \pm 4.078	0.064 \pm 0.011
SR Main Training	0.986 \pm 0.007	42.781 \pm 2.424	0.009 \pm 0.002
<i>Fine-tuned SR DynMRI</i>	<i>0.993\pm0.004</i>	<i>45.706\pm2.169</i>	<i>0.005\pm0.002</i>

Data	10% of k-space		
	SSIM	PSNR	diff SD
Area Interpolation	0.814 \pm 0.018	26.250 \pm 1.867	0.071 \pm 0.013
Trilinear Interpolation	0.906 \pm 0.007	33.148 \pm 1.780	0.022 \pm 0.004
Zero-padded	0.926 \pm 0.009	31.844 \pm 2.260	0.067 \pm 0.013
SR Main Training	0.961 \pm 0.009	36.710 \pm 1.086	0.014 \pm 0.002
<i>Fine-tuned SR DynMRI</i>	<i>0.973\pm0.005</i>	<i>39.433\pm2.144</i>	<i>0.007\pm0.001</i>

Data	6.25% of k-space		
	SSIM	PSNR	diff SD
Area Interpolation	0.723 \pm 0.031	24.092 \pm 1.964	0.080 \pm 0.013
Trilinear Interpolation	0.872 \pm 0.011	31.504 \pm 1.786	0.026 \pm 0.005
Zero-padded	0.888 \pm 0.012	29.803 \pm 2.147	0.069 \pm 0.015
SR Main Training	0.939 \pm 0.008	35.377 \pm 1.653	0.0174 \pm 0.003
<i>Fine-tuned SR DynMRI</i>	<i>0.957\pm0.006</i>	<i>37.306\pm2.357</i>	<i>0.014\pm0.004</i>

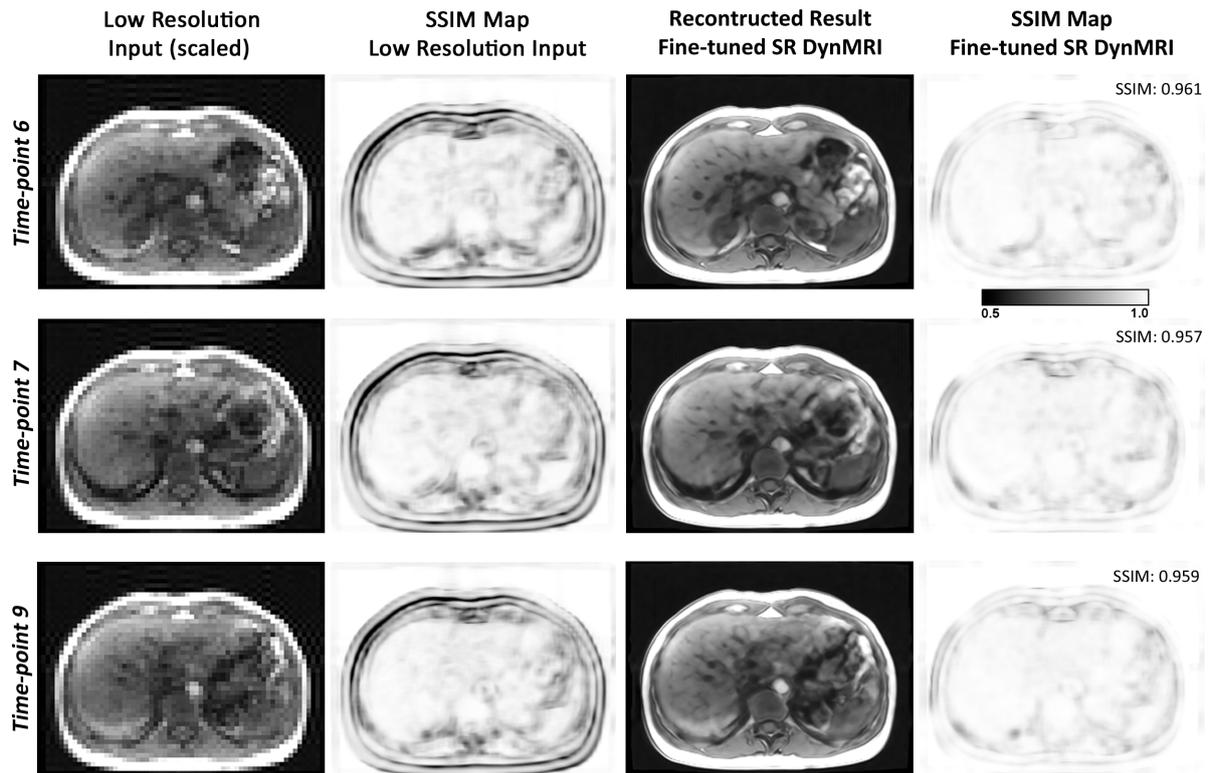


Figure 3.7: Another qualitative comparison using SSIM Maps of the lowest resolution input investigated in this work (6.25% of k-space) with the reconstructed results of Fine-tuned SR DynMRI over different time-points. The corresponding SSIM outputs are shown on the top right of the image.

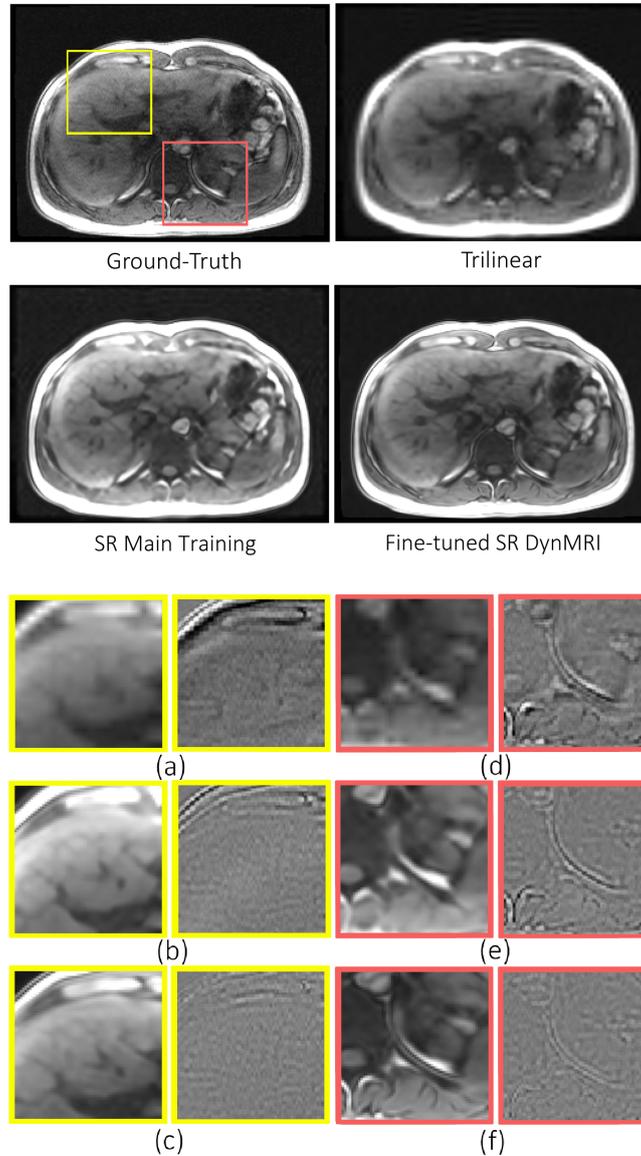
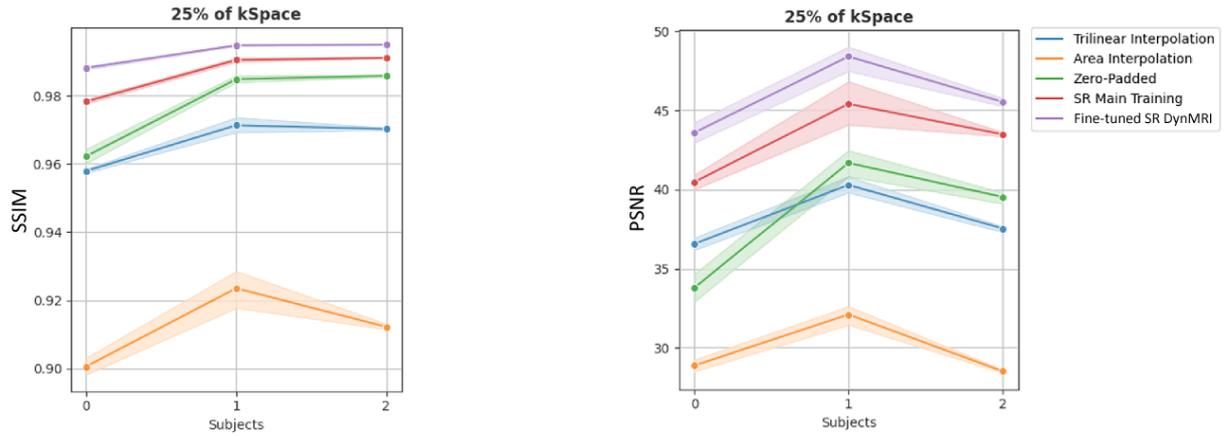
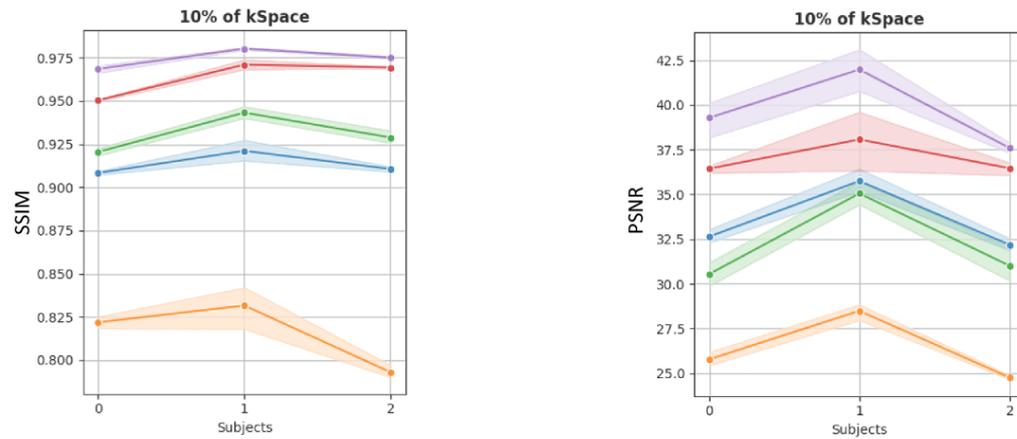


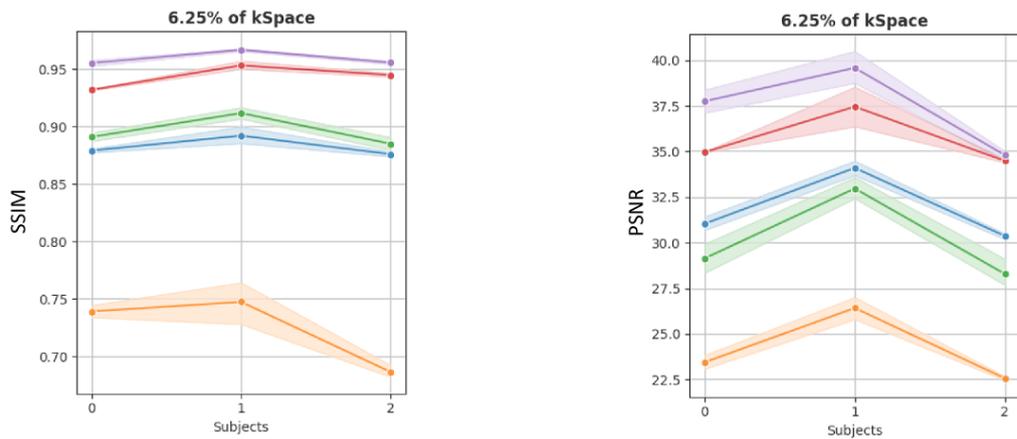
Figure 3.8: An example from the reconstructed results compared against its ground-truth of low-resolution images from 6.25% of k-space. Upper figure from left to right: ground-truth, trilinear interpolation (input of the network), SR main training, and Fine-tuned SR DynMRI. Lower figures show the yellow and red and ROIs of reconstructed results and the corresponding difference images from ground-truth of (a and b) trilinear interpolation, (c and d) SR main training, and (e and f) Fine-tuned SR DynMRI.



(a) 25% of the k-space



(b) 10% of the k-space



(c) 6.25% of the k-space

Figure 3.9: Line plot showing the mean and 95% confidence interval of the resultant SSIM and PSNR over the different time-points at each subject. The upper row shows the SSIM values and the lower row shows the PSNR values. The blue, orange, green, red and lilac colors represent the reconstruction results of trilinear interpolation, area interpolation, zero-padding (sinc interpolation), SR main training, and SR after fine-tuning respectively.

3.5 Discussion

It can be seen that after fine-tuning the model using a static planning scan, the overall results improved from SR main training with sharper images and better edge enhancement. The statistical significance of the model's performance in terms of SSIM was estimated using the paired t-test and the Wilcoxon signed-rank test. It was found that the improvement of the proposed Fine-tuned SR DynMRI was statistically significant, with p-values less than 0.001, for every investigated scenario (for all subjects with different resolutions and all resolutions with different subjects).

In this investigation, the acquisition time of high-resolution 3D "pseudo"-dynamic reference data was ten seconds without parallel acquisition, and five seconds with GRAPPA factor two (see Table 3.2). These might be insufficient for real-time applications, and may induce blurring as a result of the subject's breathing. This study, therefore, demonstrates the possibility of acquiring such volume in less than half a second with minimum loss of spatial information.

Each subject's fine-tuning procedure took approximately eight hours using the mentioned setup in section 3.3.6. Each time-point was super-resolved in a matter of seconds. Reducing the patch overlap (stride) can further shorten the time needed for fine-tuning and inference, but doing so may degrade the quality of the resulting super-resolved images. It can also be demonstrated that the network can generate findings that are very close to the ground truth (SSIM of 0.957) even while super-resolving 6.25% of k-space, which can accelerate acquisition times by 16 times. This work may be expanded to be utilized for real-time or near real-time MRI during interventions by combining this rapid acquisition speed with the fast inference speed of the proposed method.

It is important to take note that the static planning scans and the actual dynamic scans during intervention are typically acquired using different sequences, with the planning scans typically having a higher resolution than the dynamic scans. Static and dynamic scans in this study were conducted with the same sequence, but at different resolutions and positions (mimicking different scanning sessions in the actual scenarios). A volumetric interpolated breath-hold examination (VIBE) sequence was used as a planning scan for fine-tuning a subject-specific in an additional investigation. This resulted in a reconstructed image with a 0.032 lower SSIM than from fine-tuning with the same sequence for undersampled 6.25 percent of the k-space. However, an additional study would be essential before determining whether this is a drawback of the current approach.

Moreover, this research investigates only undersampled masks from the k-space center, resulting in a loss of resolution without explicit image artifacts. Future research may include variable density undersampling or GRAPPA-like uniform undersampling of higher

spatial frequencies. In addition, clinical interventional radiology usually involves devices, such as a needle or catheter, which are not included in the planning scan. Therefore, the proposed framework could be conducted by evaluating these devices further.

3.6 Conclusions

The proposed Fine-tuned SR DynMRI framework demonstrates the capability to reconstruct the dynamic MRI from relatively low-resolution images. Deep learning-based super-resolution reconstruction results can be noticeably improved by fine-tuning with a subject-specific planning scan for only one epoch. Although the network was trained on the benchmark dataset using MRI sequences that differed from the test set of dynamic MRI, the SR results in this study (both main training and fine-tuning) showed a high similarity to the high-resolution ground-truth images. In brief, this work indicates that the proposed framework could reduce spatio-temporal trade-off by enhancing the spatial image resolution without slowing the acquisition speed. The rapid inference speed of the deep learning network could make this framework viable for real-time dynamic acquisitions such as interventional MRI.

4 Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Incorporating Temporal Information in a Dynamic Dual-channel UNet

“Nothing we learn in this world is ever wasted.”

– Eleanor Roosevelt

This chapter is based on the following publications (as the first author and the co-first author):

- **Sarasaen, C.**, Chatterjee, S., Nürnberger, A., and Speck, O. "DDoS: Dynamic dual-channel u-net for improving deep learning based super-resolution of abdominal dynamic mri." In 38th Annual Scientific Meeting Congress of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), 34(Supplement 1): S6.O3., S44. Springer, 10 2021 [214].
- Chatterjee, S., **Sarasaen, C.**, Rose, G., Andreas Nürnberger, and Speck, O. "DDoS-UNet: Incorporating temporal information using dynamic dual-channel unet for enhancing super-resolution of dynamic mri." In Medical Imaging with Deep Learning (MIDL), Zürich, Switzerland, 7 2022 [71].
- Chatterjee, S., **Sarasaen, C.**, Rose, G., Andreas Nürnberger, and Speck, O. "DDoS-UNet: Incorporating temporal information using dynamic dual-channel unet for enhancing super-resolution of dynamic mri." arXiv preprint arXiv:2202.05355, 2022 [215].

4.1 Synopsis

From the previous work [70], deep learning based super-resolution (SR) has shown promising results in dealing with sparse data. Moreover, the available temporal information of dynamic MRI has not been exploited in this prior work. The proposed DDoS-UNet DynMRI (**D**ynamic **D**ual-channel training of **S**uper-Resolution **U**-Net of **D**ynamic MRI) framework attempts to learn both spatial and temporal resolution relationship. A dual-channel input consisting of a static image from the available planning scan and dynamic

images was used to super-resolve the first time point with deep learning based super-resolution and U-Net model. After that, it continues by feeding the super-resolved of the first time point along with the dynamic low-resolution image of the next following time point. Generally, the high-resolution images were super-resolved from each time-point separately, treating them as individual volumes. In this approach, the reconstructed results were super-resolved recursively, and the network was fed with the high-resolution static scans as the starting time-point. This work obtained the average SSIM of 0.951 ± 0.017 while reconstructing the highest undersampling level of 4% of the k-space center, which could theoretically result in an acceleration factor of 25.

4.2 Related Work

Deep learning based super-resolution and utilization of temporal information

Several deep learning based super-resolution models have been proposed drastically and achieved remarkable results for improving low-resolution images [10, 11, 188, 216, 217]. The SR techniques of restoring a single high-resolution image from a single low-resolution counterpart are called single image super-resolution (SISR). In contrast to SISR, video super-resolution (VSR) exploits the temporal information in a sequence of images to enhance the spatial resolution and frame rate [157–159]. The temporal information has been investigated in many works of literature and demonstrated the potential for improving the image quality of dynamic MRI reconstruction [14–16]. Previous work [70] attempting to super-resolve 3D dynamic MRI treats each time-point as a single 3D volume and then super-resolves them individually (SISR technique). Nonetheless, the inherent relationship between the different time-points of the dynamic MRI is not utilized yet, which might be helpful for improving the super-resolution performance. Dynamic MRI can also be considered a three-dimensional (3D) video. For super-resolving a two-dimensional (2D) video, recurrent networks [218–220] are commonly employed, which utilize the aforementioned relationship [221–223]. However, these networks are typically more computationally expensive during training, making them difficult to employ for a 3D volumetric scenario.

Several deep learning based super-resolution (SR) models have been proposed, which achieved remarkable results for improving low-resolution images [10, 11, 188, 216, 217]. The SR techniques of restoring a single high-resolution image from a single low-resolution counterpart are called single image super-resolution (SISR). In contrast to SISR, video super-resolution (VSR) exploits the temporal information in a sequence of images to enhance the spatial resolution and frame rate [157–159]. The temporal information has been investigated in many research papers, and demonstrated the potential for improving the image quality of dynamic MRI reconstruction [14–16]. Previous work [70] attempting

to super-resolve 3D dynamic MRI treats each time-point as a single 3D volume and then super-resolves them individually (SISR technique). Nonetheless, the inherent relationship between the different time-points of the dynamic MRI is not utilized yet, which might be helpful for improving the super-resolution performance. Recurrent neural networks (RNNs) [218–220] are a class of artificial neural networks with closed-loop connections, allowing the use of the previous information from the input sequences. RNNs are often used for temporal-related tasks, including VSR [221–223]. However, RNNs require more computationally expensive during training than traditional deep learning models. This restriction makes them difficult to employ for a 3D volumetric scenario such as the 3D dynamic MRI.

Prior-knowledge and multi-channel technique

Prior-knowledge is shown to be helpful for an inverse problem such as super-resolution [17, 70, 224]. The prior-information has been investigated and incorporated into a multi-channel technique to enhance the super-resolution reconstruction [18, 225]. A multi-channel network allows better feature extractions when learning with multiple types of channels [19]. Multi-channel training has been used across numerous applications, including image recognition [20], speech recognition [21, 22], audio classification [226], and natural language processing [23]. Thus, prior knowledge of a pair of high- and low-resolution, acquired from different scan sessions, has been incorporated as a dual-channel input to reconstruct super-resolved images in this work. Furthermore, the temporal relationship inherited between time-points has been hypothesized to support the network in learning both temporal and spatial resolution using the proposed DDoS-UNet DynMRI framework.

4.3 Methodology

In this section, the proposed DDoS-UNet DynMRI framework is introduced. The detail of datasets and pre-processing steps are included. Then, network architecture and perceptual loss network, along with the model implementations are discussed.

4.3.1 Experimental Dataset

For training, the proposed framework was trained on the artificial dynamic data generated from the benchmark CHAOS dataset [227]. This dataset contains T1-Dual images for 40 subjects (in- and opposed phases for each subject). The details of dynamic data generation is described in section 4.3.2. A ratio of 70:30 was used for dividing training and validation sets.

For inference, 3D dynamic images of five healthy subjects were scanned on a 3T MRI (Siemens Magnetom Skyra), and each subject has 25 time-points (TPs). Since DDoS-UNet DynMRI requires a dual-channel input, a high-resolution 3D static scan (breath-hold) of each subject was acquired using the same acquisition protocols, but in different sessions, miming a planning scan in the actual clinical workflow. The acquisition parameters of the data used in this work are listed in Table 4.1. It is worth mentioning that these 3D dynamic scans can be considered as "pseudo-dynamic" scans as they are not available in real-life scenarios. Moreover, to evaluate the reconstruction performance of the model, the 3D dynamic scans were artificially undersampled for different levels, as discussed in section 4.3.3.

Table 4.1: The acquisition protocol of the experimented datasets: CHAOS dataset of 40 subjects and subject-wise 3D dynamic scans of five subjects for 25 time-points (TPs). It should be noted that the static scans were acquired using the same sequence as the dynamic scans, but only for one TP, acquired in a different session.

	CHAOS (40 Subjects)	Protocol 1 (2 Subjects)	Protocol 2 (1 Subject)	Protocol 3 (1 Subject)	Protocol 4 (1 Subject)
Sequence	T1 Dual In-Phase & Opposed-Phase	T1w Flash 3D	T1w Flash 3D	T1w Flash 3D	T1w Flash 3D
Resolution	1.44 x 1.44 x 5 - 2.03 x 2.03 x 8 mm ³	0.90 x 0.90 x 4 mm ³	0.90 x 0.90 x 4 mm ³	0.90 x 0.90 x 4 mm ³	1.00 x 1.00 x 4 mm ³
FOV x, y, z	315 x 315 x 240 - 520 x 520 x 280 mm ³	300 x 225 x 176 mm ³	350 x 262 x 176 mm ³	350 x 262 x 192 mm ³	350 x 262 x 176 mm ³
Encoding matrix	256 x 256 x 26 - 400 x 400 x 50	320 x 240 x 44	384 x 288 x 44	384 x 288 x 48	352 x 264 x 44
Phase/Slice oversampling	N/A	10/0 %	10/0 %	10/0 %	10/0 %
TR	110.17 - 255.54 ms	2.37 ms	2.40 ms	2.40 ms	2.31 ms
TE	4.60 - 4.64 ms (In-Phase) 2.30 ms (Opposed-Phase)	1.00 ms	1.02 ms	1.02 ms	0.97 ms
Flip angle	80°	8°	8°	8°	8°
Bandwidth	N/A	920 Hz/Px	930 Hz/Px	930 Hz/Px	950 Hz/Px
GRAPPA factor	None	None	None	None	None
Phase/Slice partial Fourier	N/A	Off/Off	Off/Off	Off/Off	Off/Off
Phase/Slice resolution	N/A	50/64 %	50/64 %	50/64 %	50/64 %
Fat saturation	N/A	On	On	On	On
Time per TP	N/A	10.52 sec	12.80 sec	13.96 sec	11.36 sec

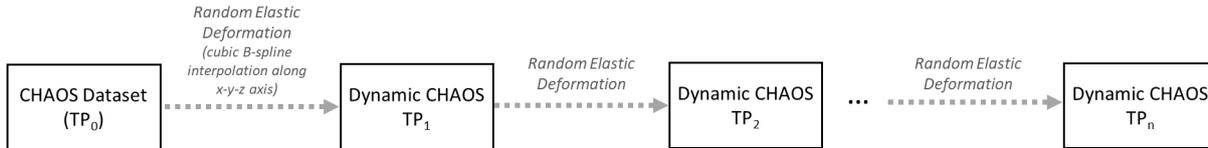


Figure 4.1: Flowchart of dynamic data generation. In this work, $n = 24$ and the total number of TP is 25 time-points.

4.3.2 Dynamic Data Generation

Due to the lack of large training datasets of dynamic MRI, in this work, the dynamic MRI dataset was created artificially using random elastic deformation of TorchIO [228], applied to the volumes from the CHAOS dataset. The deformation fields were created randomly using five control points, maximum displacements along x-y-z dimensions of 5-20-20 mm, and two locked borders. The first deformation field was interpolated to the coarse grid of control points of the CHAOS data using cubic B-spline interpolation. The original CHAOS volume is referred to as TP_0 , and the generated volume is TP_1 . To imitate the breathing pattern that moves continuously, the next adjacent time-point TP_2 was generated by applying the deformation field on TP_1 . Then, repeatedly applying to the current time-point TP_n to create the next time-points TP_{n+1} until having totally 25 time-points ($TP_0 - TP_{24}$). The flowchart of dynamic data generation is shown in figure 4.1.

The quality of the artificial dynamic MRI dataset was verified by manually inspecting the data (e.g., no image folding) and setting up the similarity metric of less than 80% of the previous time-points. Although the dynamic dataset is not generated from the human breathing motion, it contains different motion patterns as the deformation fields were created randomly. Therefore, this should be adequate for the network to learn the generic relationship between time-points.

4.3.3 Undersampled Data Generation

The artificial dynamic MRI dataset (for training) and 3D dynamic images (for testing) were undersampled in-plane for three different levels. These low-resolution images were fed to the network as one of dual-channel input, and its high-resolution images served as the corresponding ground-truth images. The undersampling masks were generated by taking only 10%, 6.25%, and 4% of the k-space center without zero-padding, resulting in MR acceleration factors (2.3.4) of 3, 4, and 5 when considering only the phase-encoding direction. However, concerning the actual amount of data used for reconstruction (reducing both phase-encoding and read-out directions), this could be considered theoretically as

Table 4.2: Effective resolutions and estimated acquisition times (per time-point) of the dynamic and static datasets after performing different levels of artificial under-sampling.

	Protocol 1		Protocol 2		Protocol 3		Protocol 4	
	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)
High-Resolution (Ground-truth)	0.90 x 0.90 x 4	8.76	0.90 x 0.90 x 4	10.68	0.90 x 0.90 x 4	11.76	1.00 x 1.00 x 4	9.38
10% of k-space	2.70 x 2.70 x 4	0.88	2.70 x 2.70 x 4	1.07	2.70 x 2.70 x 4	1.18	3.00 x 3.00 x 4	0.94
6.25% of k-space	3.60 x 3.60 x 4	0.55	3.60 x 3.60 x 4	0.67	3.60 x 3.60 x 4	0.74	4.00 x 4.00 x 4	0.59
4% of k-space	4.50 x 4.50 x 4	0.35	4.47 x 4.47 x 4	0.43	4.47 x 4.47 x 4	0.47	4.99 x 4.99 x 4	0.38

acceleration factors of 10, 16, and 25, respectively. The effective resolutions and estimated acquisition times of different undersampled levels for each of the dynamic test datasets and listed in Table 4.2. The estimation of the scan times can be calculated as follow [30]:

$$\text{Scan time}_{3D \text{ volume}} = N_{AQ} \times TR \times N_{PE} \times N_{PART}, \quad (4.1)$$

which N_{AQ} denotes the number of acquired line data within the given number of phase-encoding lines N_{PE} and the number of divided partitions N_{PART} , and TR is the repetition time.

4.3.4 DDoS-UNet DynMRI Framework

The framework **Dynamic Dual-channel of Super-resolution UNet** (DDoS-UNet) is proposed for dynamic MRI reconstruction, and is termed as DDoS-UNet DynMRI. The method overview of training and inference steps are summarized in figure 4.2 and 4.3, respectively. The network architecture is based on a 3D-UNet, combined with a perceptual loss network, similar to [70]. The main difference of the DDoS-UNet DynMRI framework is that the dual-channel input consists of a pair of $HR - LR$ images from different time-points. Since the inputs are obtained from different time-points and abdominal organs changes during breathing, the same patch volume of both inputs might not have the same locations. Therefore, super-resolution reconstruction was carried out using the entire 3D volumes rather than the 3D patch-based super-resolution.

Let the first channel of the dual-channel input be a high-resolution static scan obtained from the previous session $HR_{TP_{n-1}}$ and the second channel be a patient-specific low-resolution dynamic MRI of the current time-point LR_{TP_n} , the reconstructed image of the current time-point $\hat{H}R_{TP_n}$ can be recovered (super-resolved) using the following equation:

$$\hat{H}R_{TP_n} = F(LR_{TP_n}; \hat{H}R_{TP_{n-1}}; \theta), \quad (4.2)$$

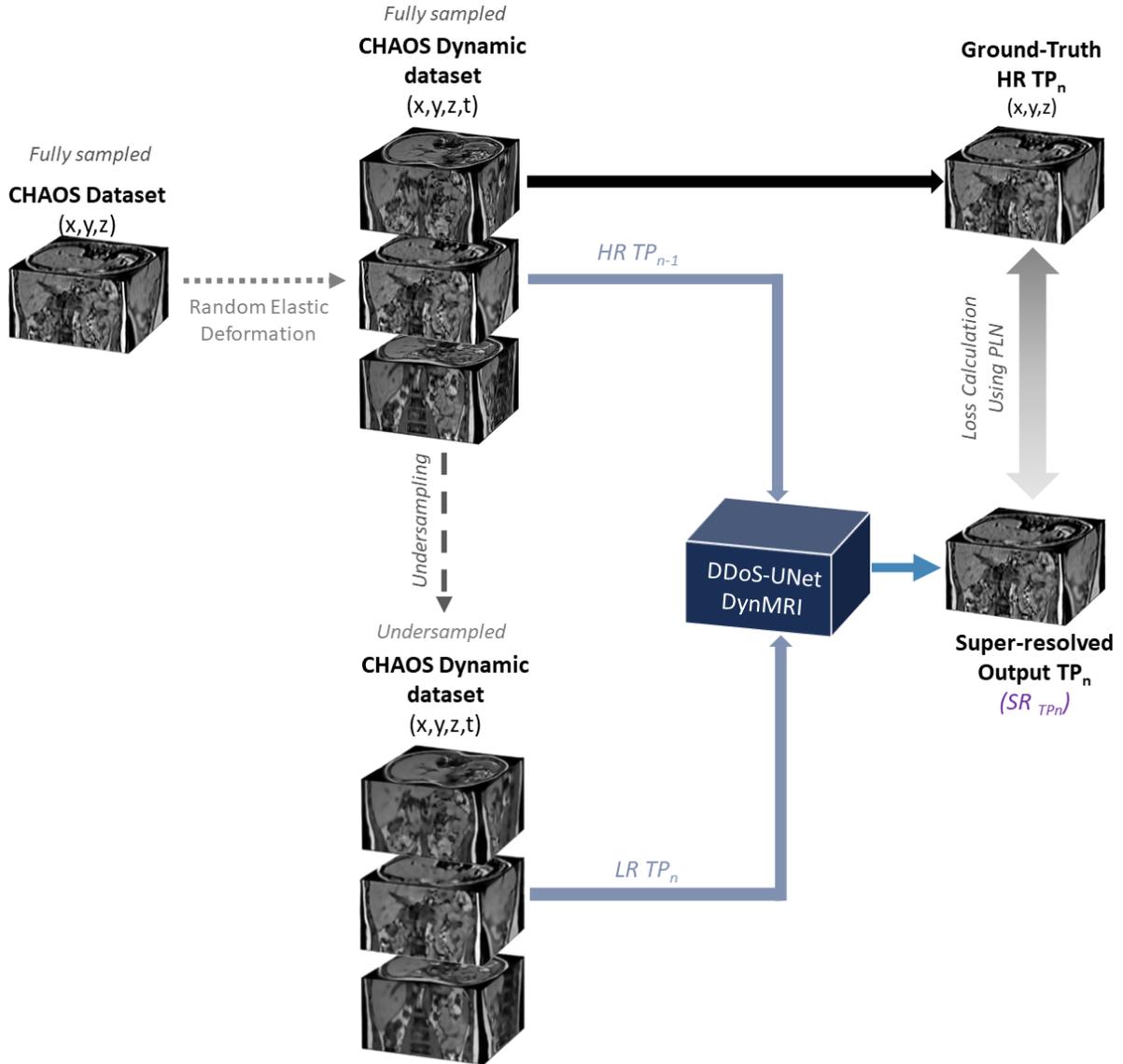


Figure 4.2: Method overview of the training in DDoS-UNet DynMRI framework. Initially, random elastic deformation is applied to the CHAOS dataset (fully sampled) to generate the artificial CHAOS dynamic dataset. Then the CHAOS dynamic dataset was undersampled to generate the final training dataset. Then the model is trained by providing a low-resolution (undersampled) current time-point ($LR TP_n$) along with the high-resolution (fully sampled) previous time-point ($HR TP_{n-1}$) as input, and the output is compared against the ground-truth high-resolution current time-point ($HR TP_n$).

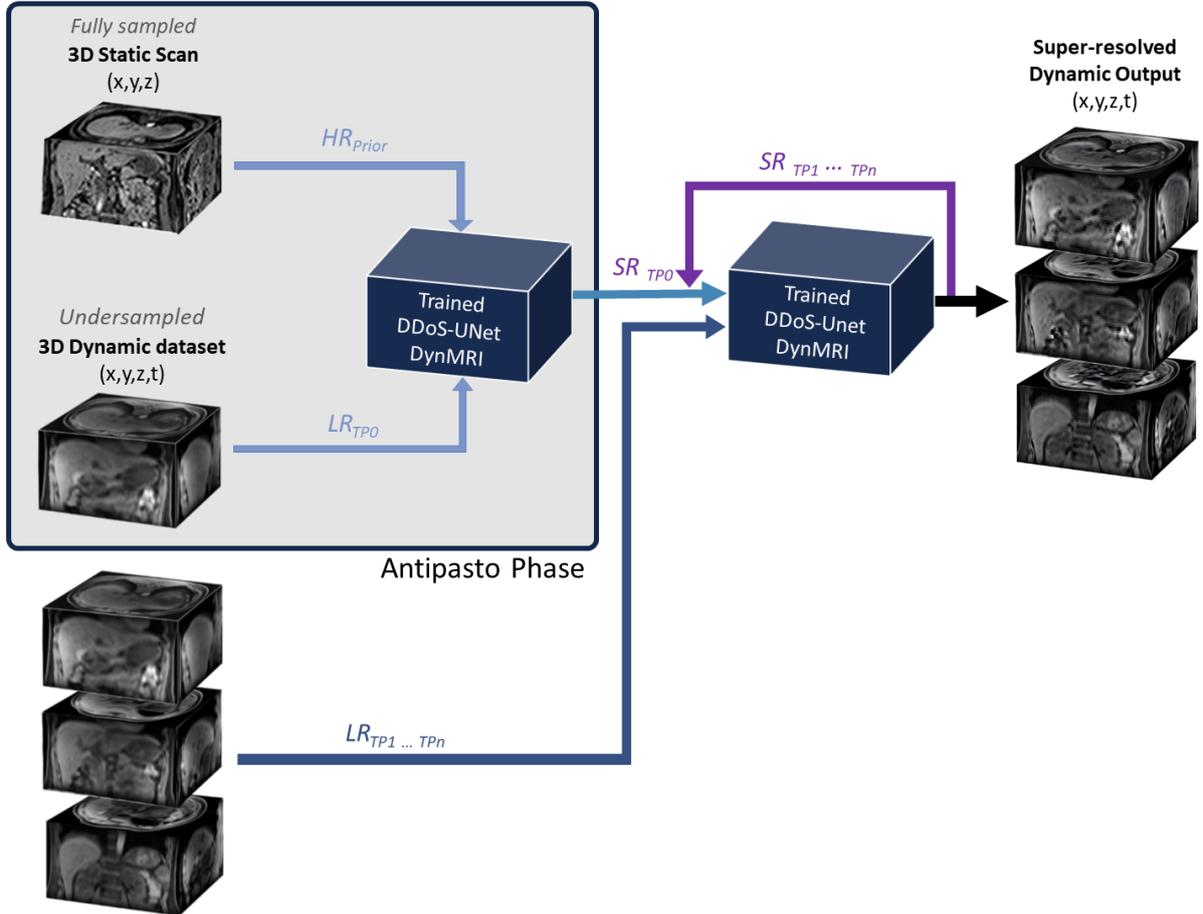


Figure 4.3: Method overview of the inference in DDoS-UNet DynMRI framework. 3D static subject-specific planning scan (fully sampled) is supplied as the high-resolution prior image (HR_{prior}) along with the first low-resolution (undersampled) time-point (LR_{TP0}) of the 3D dynamic dataset are supplied as input to the trained DDoS-UNet model, and the model super-resolves LR_{TP0} to obtain SR_{TP0} . This initial phase is called the "Antipasto" phase. SR_{TP0} is then supplied as input together with the next low-resolution time-point LR_{TP1} to the same trained DDoS-UNet model to obtain SR_{TP1} . This process is continued repeatedly until all the time-points of the low-resolution (undersampled) 3D dynamic are super-resolved by supplying pairs of $SR_{TP_{n-1}}$ and LR_{TP_n} to obtain each of the SR_{TP_n} .

This equation is modified from equation 3.1, where F denotes the super-resolution model, mapping the given pair of high- and low-resolution images, and θ is the parameters of F . The proposed framework tries to solve the super-resolution reconstruction using deep-learning, by optimizing the loss-function \mathcal{L} between the reconstructed image $\hat{H}R$ and its high-resolution ground-truth HR :

$$\hat{\theta} = \arg \min_{\theta} \sum_{s=1}^S \sum_{TP=1}^{TP_n} \mathcal{L}(HR_{s,TP_{n-1}}, HR_{s,TP_n}) + R(\theta) \quad (4.3)$$

Where the first term represents the data term and the second term $R(\theta)$ represents a regularization term, when TP_n is the number of time-point, and S is the number of subjects in the dataset.

For the framework mechanism, it is assumed that the network can learn the spatial relationship between $HR - LR$ and the temporal relationship between different time-points TP_n . Hence, let the DDoS-UNet DynMRI mechanism be Ψ and the set of network parameters be $(\hat{\theta})$, the hypothesis of the proposed framework can be modeled as:

$$\Psi(\hat{\theta}) = F_1(LR_{TP_n}; \hat{H}R_{TP_n}; \theta_1) + F_2(\hat{H}R_{TP_{n-1}}; \hat{H}R_{TP(n)}; \theta_2), \quad (4.4)$$

where F_1 denotes the spatial-relationship operator between LR_{TP_n} and $\hat{H}R_{TP_n}$. θ_1 are their set of the first parameters. While F_2 denotes the temporal-relationship operator between $\hat{H}R_{TP_{n-1}}$ and $\hat{H}R_{TP_n}$. θ_2 are their set of second parameters.

4.3.5 Network Architecture

The network architecture is based on a 3D UNet-based model combined with a perceptual loss network, as illustrated in figure 4.4. It was used to solve a super-resolution problem in previous work by the same author [70] but was extended with two modifications. First, a dual-channel input was supplied to the network rather than one, and upsampling operators were used instead of upconvolutions in the expanding path.

The proposed network has three levels, including two main paths: contracting (encodes the features) and expanding (decodes the generated features). The contracting and expanding paths are connected with skip connections to concatenate the input and output at the same level. The contracting path comprises of two convolutions, each block is followed by a rectified linear unit (ReLU) activation function and an average pooling. The expanding path consists of two trilinear-upsamplings and convolutions, and each block is followed by a convolution with ReLU activation function. The final step has only a convolution to combine all the features and provide the final output.

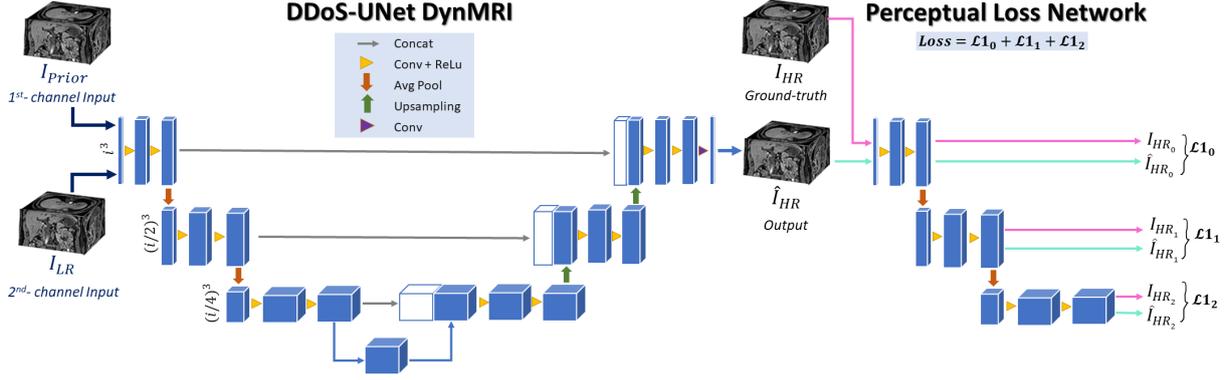


Figure 4.4: Network architecture of the modified 3D UNet with perceptual loss network. Unlike the patch-based super-resolution (see figure 3.5), the training was performed using the whole 3D volumes of I_{prior} and I_{LR} as dual inputs to the network.

Unlike the earlier work (patch-based) that applied convolutional transpose, this work uses the upsampling operation, as it did not generate checkerboard artifacts [229] when super-resolving the entire volume. Moreover, before supplying the inputs to the network, the size of low-resolution inputs was increased using trilinear interpolation to be the same image size as its high-resolution ground-truth. This additional step was derived from the fact that UNet-like architecture demands the same input and output matrix size.

The perceptual loss was employed for loss calculation, same as [70]. The perceptual loss network (PLN) was derived from the first three blocks of a frozen pre-trained U-Net MSS model (for vessel segmentation of MR angiography at 7T) [211]. This PLN intended to extract the small feature abstractions at different levels between the super-resolved output of the model and its corresponding high-resolution ground-truth. These extracted features were also compared against each other using L1 loss (mean absolute error), and the obtained losses, which were accumulated together before propagating the total losses back. The formulation of PLN can be found in equation 3.7.

4.3.6 Implementation Details

The proposed model was trained with 3D volumes from the artificially created dynamic version of a publicly available benchmark dataset, as shown in figure 4.2. The overview of inference step is shown in figure 4.3. The starting phase, referred here to as "Antipasto", is the preliminary phase prior to the main reconstruction - by supplying the high-resolution patient-specific static scan as a prior image on the first channel of the network (as $\hat{H}R_{TP(n-1)}$ is not yet available), and by supplying $LR_{TP(0)}$ on the second channel of the network. After the network super-resolves $LR_{TP(0)}$ to $\hat{H}R_{TP(0)}$, the next time-point, $\hat{H}R_{TP(0)}$ and $LR_{TP(1)}$ are supplied as input to the network and obtain the next super-resolved image $\hat{H}R_{TP(1)}$.

This process continues until low-resolution images of all time-points are super-resolved.

The training and inference were performed using Nvidia Tesla V100 GPUs. The gradient updates were estimated using a batch size of one for each iteration. This small batch size promotes the accuracy of the error gradient computation [230]. The perceptual loss network [211] was utilized for loss calculation and minimized using Adam optimizer. The predefined epoch was 100 epochs and a learning of 0.0001. The implementation code using Pytorch is publicly available ¹. The tutorial of this code is available on Google Colab Notebook (see appendix C).

It should be noted that the static scan has the same resolution, contrast, and volume coverage as the high-resolution ground-truth dynamic scan. However, to keep the testing environment similar to the real-life scenario and to keep a fast speed of inference, the static and dynamic datasets were not co-registered, as registration is typically time-consuming.

4.3.7 Image Reconstruction Evaluation

The quality of super-resolution reconstruction was evaluated using the structural similarity index (SSIM) [187], the peak signal-to-noise ratio (PSNR) [231], and the normalized root mean square error (NRMSE). The SSIM metric measures the perceptual difference between the output of the proposed method and its ground-truth, by comparing the luminance, contract, and structure terms between two given images. The formulas used in this work are described in appendix B. The quality of the reconstructed results was also evaluated statistically using PSNR and NRMSE. These metrics estimate the pixels individually, based on the mean-square error (MSE) between two images. Lastly, the Mann–Whitney U test was selected for a statistical test to compare the proposed method against the other baseline methods.

4.4 Results

The performance and limitation of the DDoS-UNet DynMRI framework were evaluated using different low-resolution images (simulated by taking 10%, 6.25%, and 4% of the k-space center). These reconstructed results were compared against the low-resolution input, traditional trilinear interpolation, and Fourier interpolated input (zero-padded k-space). In addition, the results were also compared against baseline deep learning models using the UNet model trained on the artificial dynamic CHAOS. Although the baseline UNet was trained on the same training dataset as the proposed DDoS-UNet DynMRI framework, this UNet received only one input, unlike dual-channel input in the

¹<https://github.com/soumickmj/DDoS>

proposed framework. Finally, the model was tested using the dynamic abdominal MRI of five subjects, containing 25 TPs for each subject. The DDoS-UNet DynMRI framework started inferring the data since the Antipasto phase by using the dual-channel input of the subject-specific high-resolution planning scan (HR_{prior}) and the first time-point (TP_0) of dynamic low-resolution. Subsequently, the network continued inferring the data using the inputs of the previous super-resolved image SR_{TP0} and the current dynamic low-resolution LR_{TP1} . This mechanism was processed to super-resolve the current TP until TP_n . The inference details of the method can be found in section 4.3.4.

For qualitative comparison, Figure 4.5 portrays the results of different methods for all experimented undersampling (10%, 6.25% and 4% of the k-space center). As seen in the figure, the results of the DDoS-UNet DynMRI illustrate the potential to reconstruct the images similar to its corresponding ground-truth, while achieving better edge delineation compared to the other baseline methods. Moreover, Figure 4.6 shows the SSIM Maps of the low-resolution inputs and the results of the DDoS-UNet DynMRI, calculated against the high-resolution ground-truth. The corresponding SSIM outputs can be found on the top right of the image. SSIM values typically range from zero to one; thus, the higher values result in brighter pixels in the local SSIM Map. The reconstructed results during the Antipasto phase appeared to be inferior to the other time-points - although it was a considerable improvement compared to the low-resolution input. In addition, Figure 4.7 illustrates the reconstructed results from the undersampled 4% of the k-space center (the lowest-resolution experimented here) in two regions of interest (ROIs), along with the subtracted image from the ground-truth. The proposed DDoS-UNet DynMRI framework managed to reconstruct the results with better details of the delicate structures, such as hepatic vessels in the liver. Although the deep learning based SR using the UNet model generated better qualitative results than the traditional trilinear interpolation method, it failed to preserve the anatomical topology, unlike the DDoS-UNet DynMRI, which could maintain the organ structures better.

The quantitative analysis can be found in Table 4.3. For all undersampling levels, the proposed DDoS-UNet DynMRI method outperformed the baseline UNet model in terms of evaluation metrics (SSIM, PSNR, NRMSE), with statistical significance (p-values less than 0.001). Moreover, Figure 4.9 portrays the subject-wise line plot of mean values of SSIM and PSNR over a time-point, not including the Antipasto phase. The error bands present the 95% confidence interval of the resultant metrics over time-points. As can be seen in the figure, the proposed framework demonstrated an improvement through the higher undersampling levels. Finally, Figure 4.10 portrays the TP-wise line plot of the average SSIM for all subjects over the different time-points. The results obtained during the Antipasto phase TP_0 - the initial phase of reconstruction - yielded the low value of resultant SSIM. However, after this phase, the proposed DDoS-UNet DynMRI method

achieved the resultant SSIM values higher than the baseline methods and continued to reconstruct the results steadily over time-points. During the observation over time-points, the mean values of SSIM of each subject ranged from 0.988 to 0.975, 0.980 to 0.960, and 0.970 to 0.945 for 10%, 6.25%, and 4% of k-space, respectively. Therefore, these quantitative results demonstrate the proposed method's effectiveness in reconstructing the images from different subjects with dissimilar anatomical structures, and were acquired using different acquisition protocols.

Moreover, the proposed method was also compared to the previously-proposed fine-tuning based super-resolution of dynamic MRI [70], referred to here as "Fine-tuned SR DynMRI". Figure 4.11 shows a qualitative comparison of these methods to the proposed DDoS-UNet DynMRI while super-resolving 4% of the center k-space. Although the Fine-tuned SR DynMRI approach could restore the information from highly undersampled input, the result is very smooth and fails to recover the details of anatomy and fine structures compared to the DDoS-UNet DynMRI. However, the Fine-tuned SR DynMRI works with patches - making it suitable for GPUs with lesser memories (e.g. 12GB Nvidia GeForce RTX 2080 TI) than the ones required for DDoS-UNet DynMRI (e.g. 32GB Nvidia Tesla V100). But, working with patches also increases the processing time. During inference, DDoS-UNet DynMRI took (on average of all subjects) 0.36 seconds for each time-point (9 secs for 25 TPs). On the other hand, Fine-tuned SR DynMRI took 2 minutes per time-point (50 minutes for 25 TPs - average over all five subjects), while working with a batch-size of 1900 on the same GPU. If the same inference is performed using an inferior GPU (Nvidia GeForce RTX 2080 TI) with a batch-size of 96 (similar to training and fine-tuning stages), the inference time increases to X for each time-point. This makes DDoS-UNet DynMRI better suitable for real-time or near real-time applications than the Fine-tuned SR DynMRI approach. Moreover, the time required for inferring one time-point using Fine-tuned SR DynMRI depends on the matrix size of the volumes, as a larger matrix would result in more patches if the same patch size and strides are used. If Fine-tuned SR DynMRI also works with the whole volume, this difference in inference time can be resolved. But in that case, there will be only one forward pass using one static volume for fine-tuning (currently it uses all the possible patches) - making it unsuitable. Moreover, the Fine-tuned SR DynMRI approach requires an additional step after training the model - the step of fine-tuning using subject-specific static scans. Depending upon the available resources, the fine-tuning can take 8-10 hours - which can be avoided using DDoS-UNet DynMRI as it does not require this step.

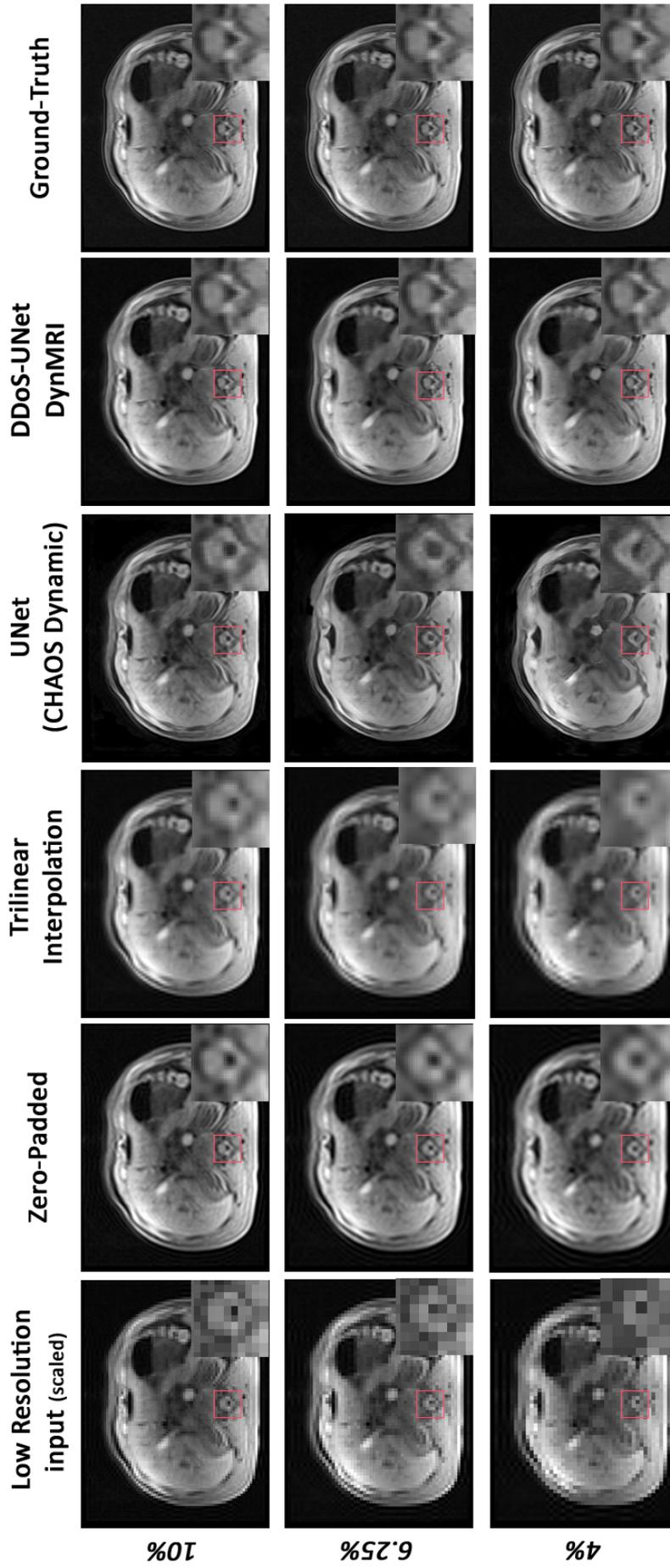


Figure 4.5: Comparative results of low-resolution data (undersampled 10%, 6.25% and 4% of k-space) of the same slice. From left to right: low-resolution images (scaled-up), zero-padding, interpolated input (Trilinear interpolation), SR result of the UNet baseline trained with CHAOS dynamic (UNet CHAOS Dynamic), DDoS-UNet DynMRI and its corresponding ground-truth images.

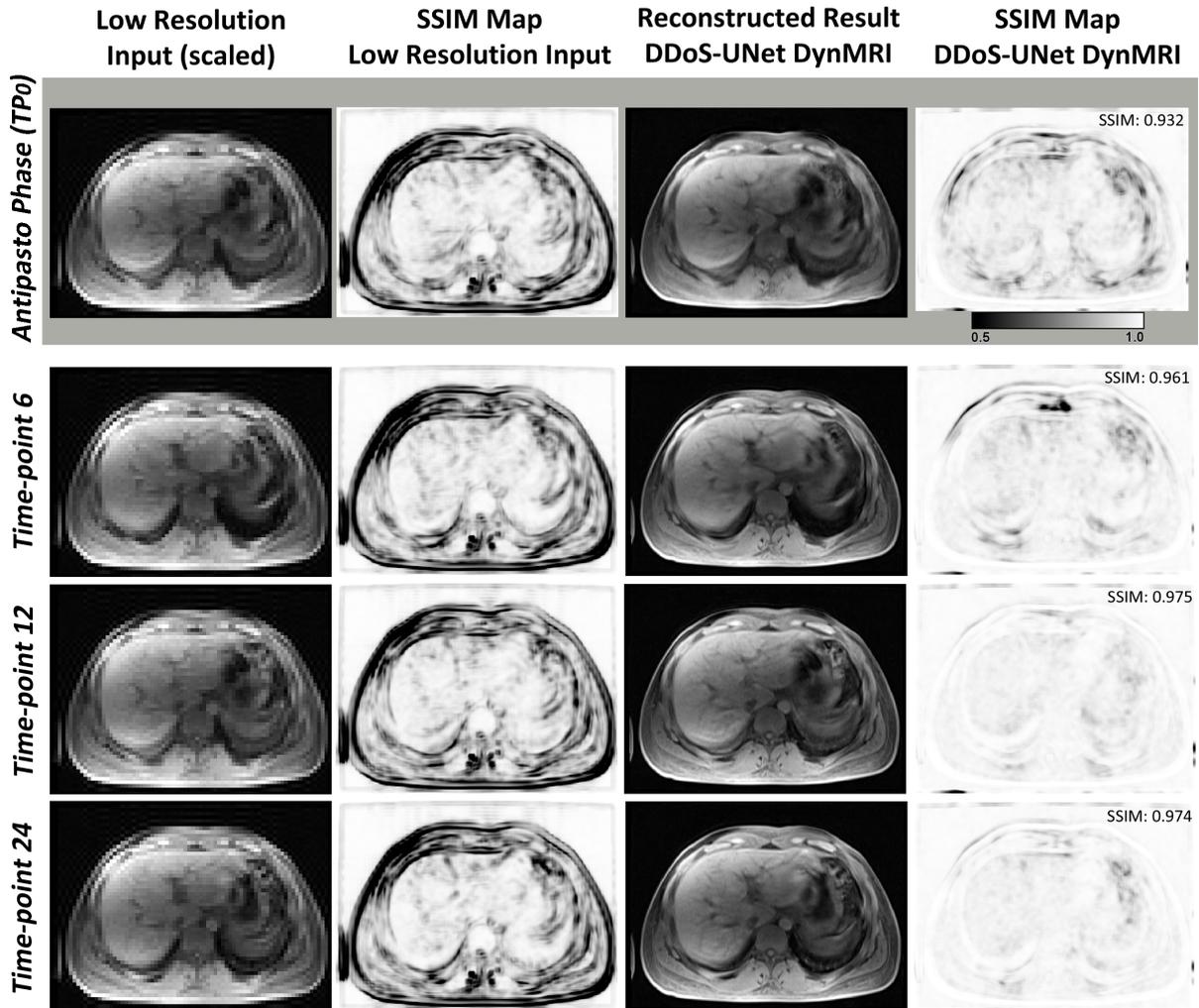


Figure 4.6: An example comparison of the low-resolution images of the 4% of k-space with reconstructed results of the DDoS-UNet DynMRI, compared against the high-resolution ground-truth using SSIM maps. Four different time-points are shown here, including the Antipasto phase (initial time point TP_0). The corresponding SSIM outputs are displayed on the top right of the image.

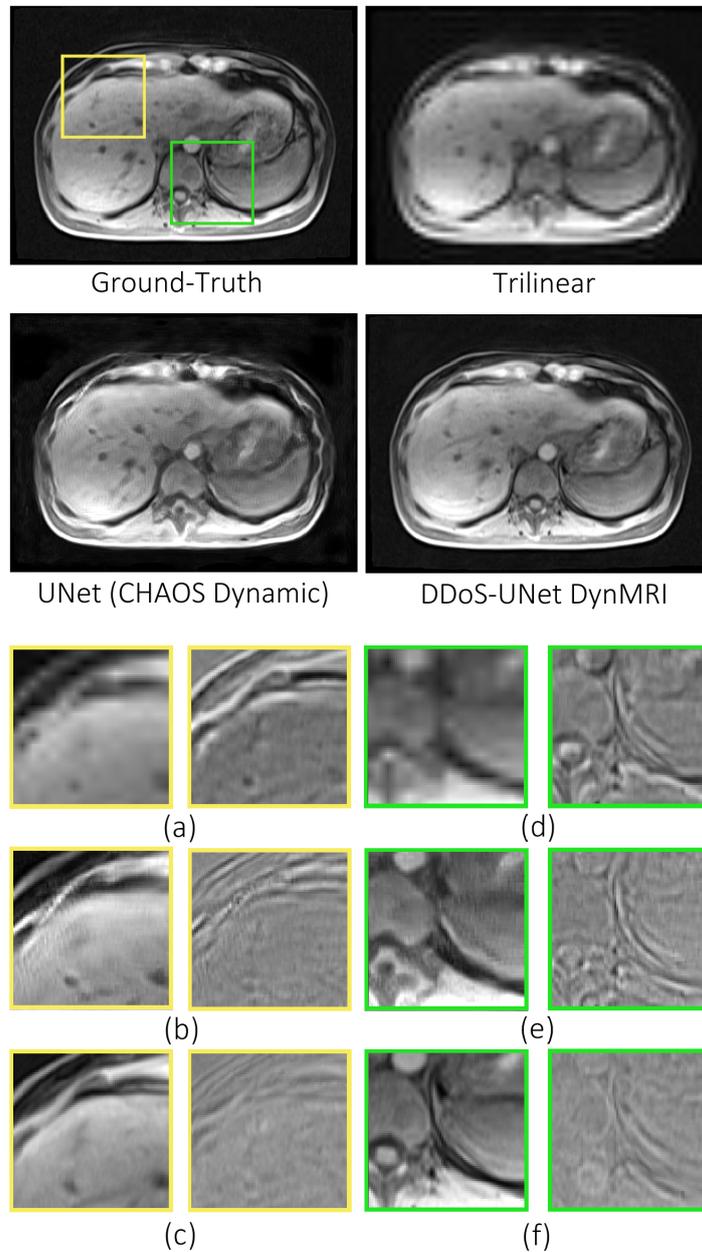
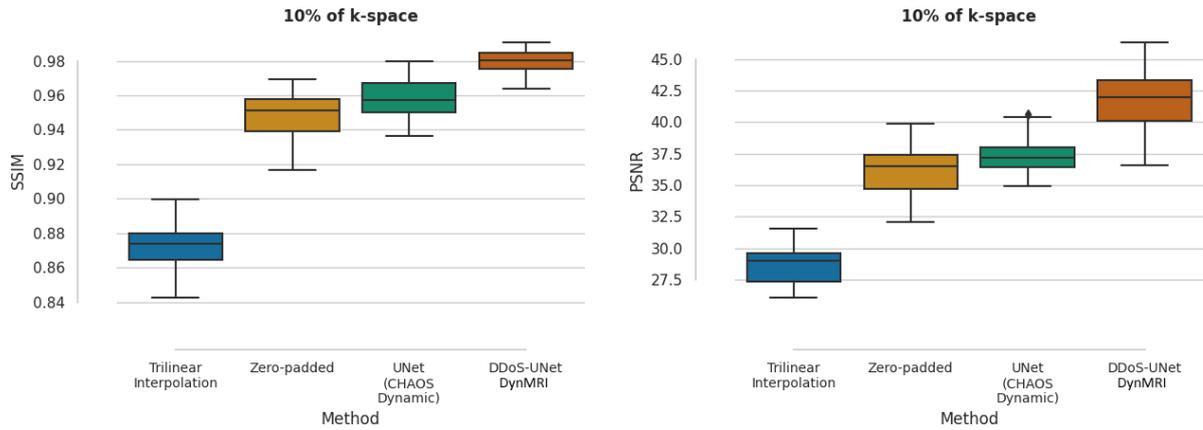
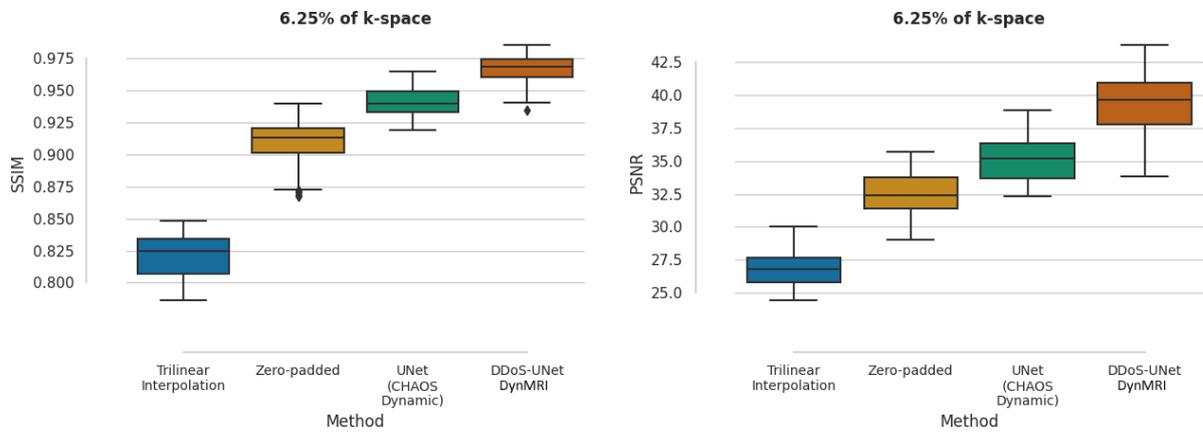


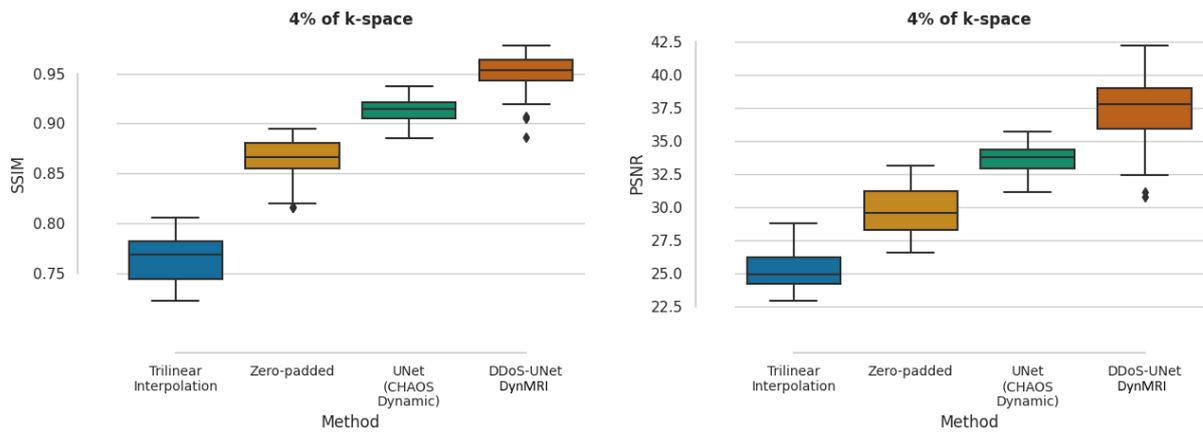
Figure 4.7: An example from the reconstructed results compared against its ground-truth (GT) of low-resolution images from 4% of k-space. Upper figure from left to right: ground-truth, trilinear interpolation (interpolated input), SR result of the UNet baseline trained with CHAOS dynamic (UNet CHAOS Dynamic) and the DDoS-UNet DynMRI. Lower figures show the yellow and green and ROIs of reconstructed results and the corresponding difference images from GT of (a and b) trilinear interpolation, (c and d) SR result of UNet CHAOS Dynamic, and (e and f) DDoS-UNet DynMRI.



(a) 10% of the k-space

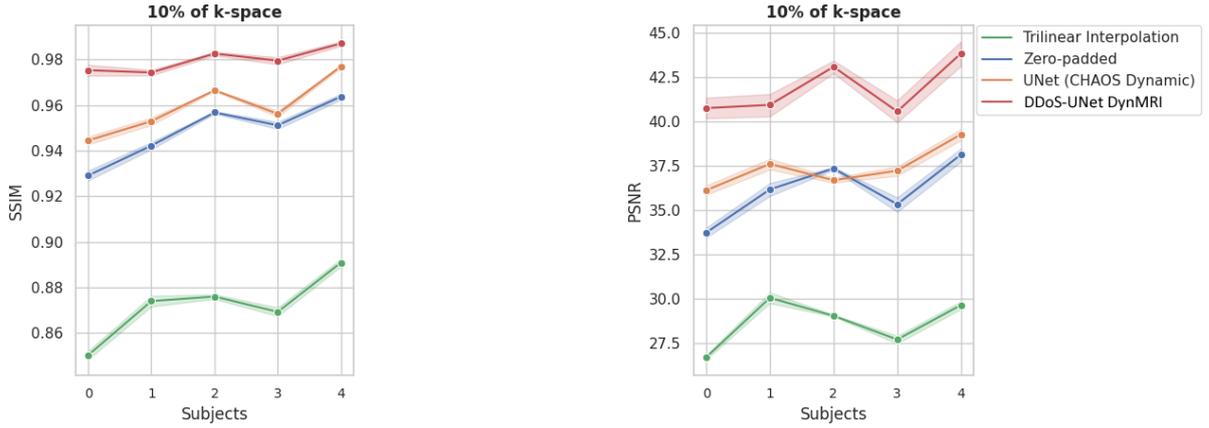


(b) 6.25% of the k-space

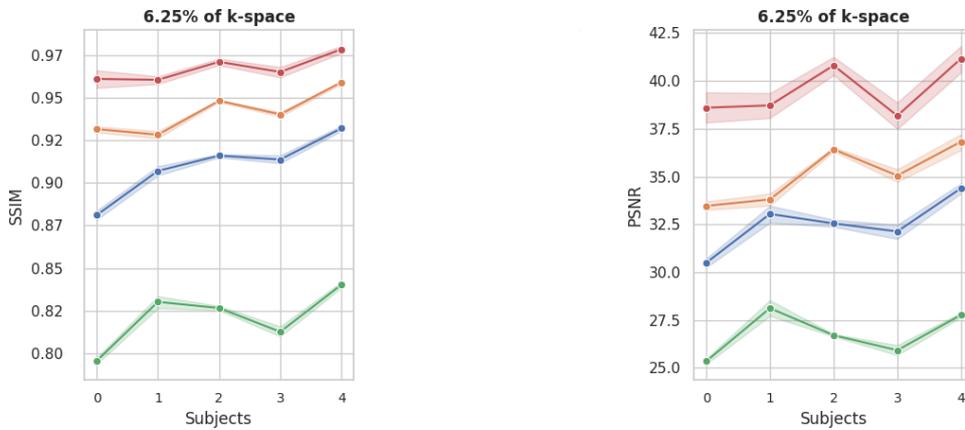


(c) 4% of the k-space

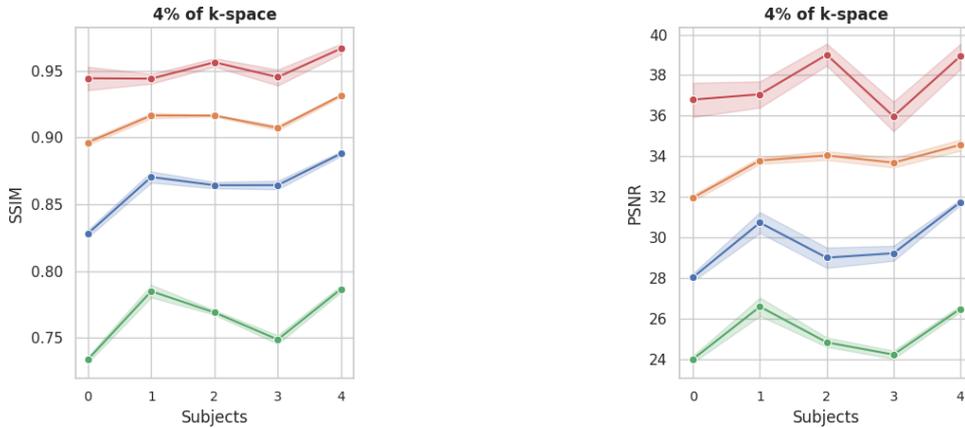
Figure 4.8: Comparison of baseline methods against the proposed DDoS-UNet DynMRI of all subjects for different undersampling levels. The box plots show SSIM (left column) and PSNR (right column) values of (a) 10% of the k-space (b) 6.25% of the k-space, and (c) 4% of the k-space.



(a) 10% of the k-space

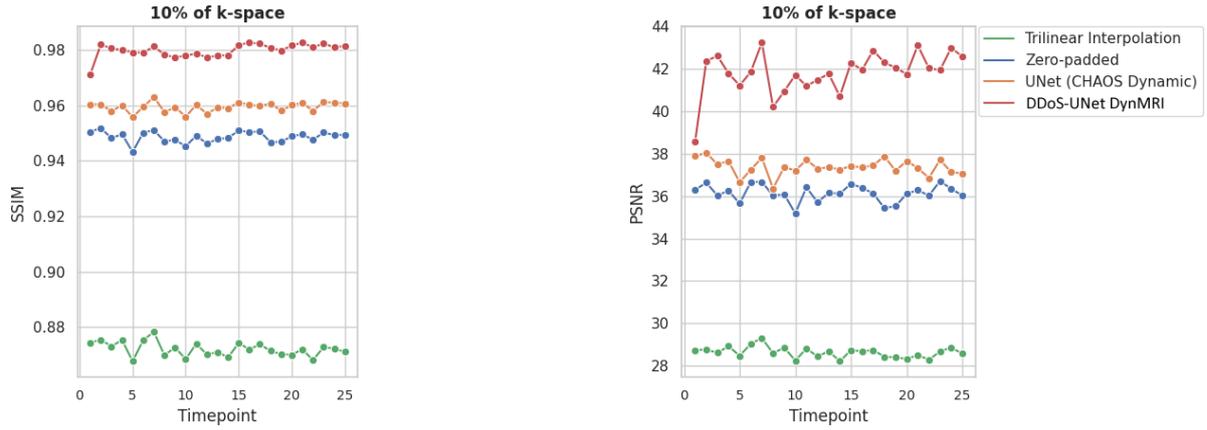


(b) 6.25% of the k-space

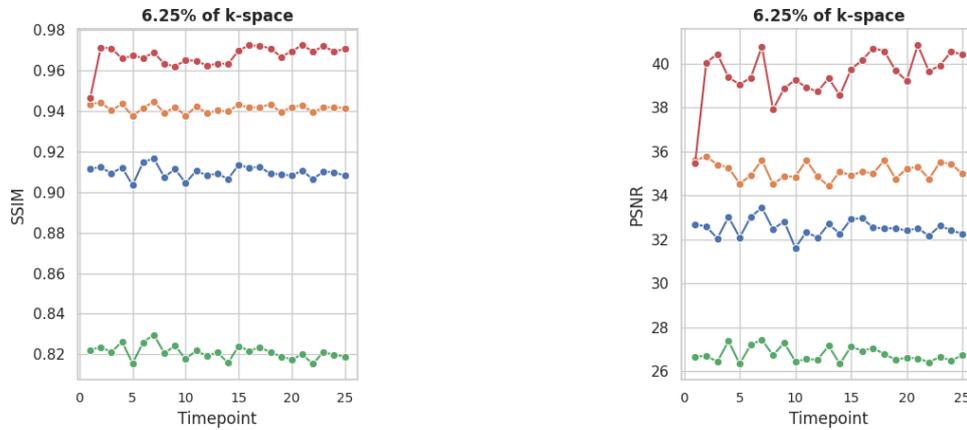


(c) 4% of the k-space

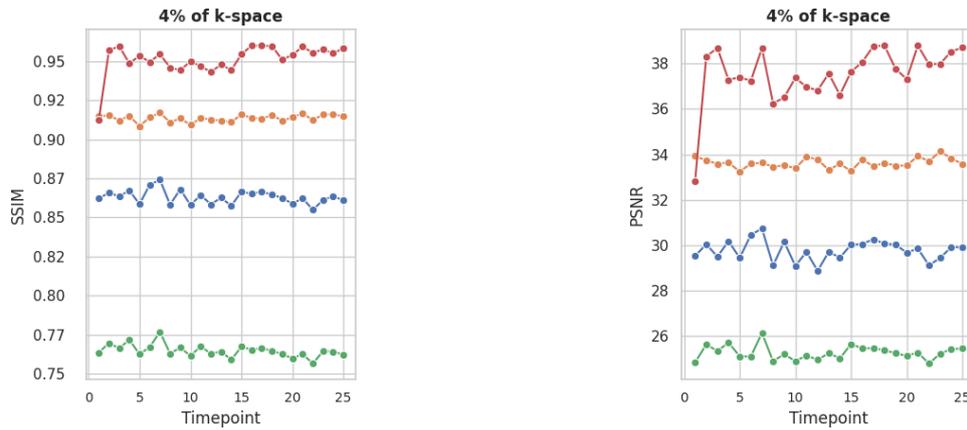
Figure 4.9: Subject-wise line plots show the average SSIM (left column) and PSNR (right column) values of all time-points for different undersampling levels (a) 10% of the k-space (b) 6.25% of the k-space, and (c) 4% of the k-space. The plots also portrays mean and 95% confidence interval of the resultant values over the different time-points, excluding the initial one from the Antipasto phase. The red, orange, blue and green lines represent the reconstruction results of DDoS-UNet DynMRI, UNet trained with CHAOS Dynamic dataset, zero-padding (sinc interpolation), and trilinear interpolation, respectively.



(a) 10% of the k-space



(b) 6.25% of the k-space



(c) 4% of the k-space

Figure 4.10: Line plots show the average SSIM (left column) and PSNR (right column) values over time-points, of all subjects for different undersampling levels (a) 10% of the k-space (b) 6.25% of the k-space, and (c) 4% of the k-space. The red, orange, blue and green lines represent the reconstruction results of DDoS-UNet DynMRI, UNet trained with CHAOS Dynamic dataset, zero-padding (sinc interpolation), and trilinear interpolation, respectively.

Table 4.3: The resultant values of SSIM, PSNR, and NRMSE (average values \pm SD). The table shows the comparison results for all different resolutions. The p-values were less than 0.0001 for all scenarios.

Data	10% of k-space		
	SSIM	PSNR	NRMSE
Trilinear Interpolation	0.872 \pm 0.014	28.631 \pm 1.364	0.192 \pm 0.023
Zero-padded	0.949 \pm 0.013	36.138 \pm 1.753	0.082 \pm 0.016
UNet (CHAOS Dynamic)	0.959 \pm 0.012	37.376 \pm 1.275	0.024 \pm 0.003
<i>DDoS-UNet DynMRI</i>	<i>0.980\pm0.006</i>	<i>41.824\pm2.070</i>	<i>0.014\pm0.003</i>
Data	6.25% of k-space		
	SSIM	PSNR	NRMSE
Trilinear Interpolation	0.821 \pm 0.017	26.770 \pm 1.226	0.238 \pm 0.024
Zero-padded	0.910 \pm 0.018	29.761 \pm 1.640	0.124 \pm 0.019
UNet (CHAOS Dynamic)	0.941 \pm 0.012	35.113 \pm 1.566	0.031 \pm 0.006
<i>DDoS-UNet DynMRI</i>	<i>0.967\pm0.011</i>	<i>39.494\pm2.121</i>	<i>0.019\pm0.005</i>
Data	4% of k-space		
	SSIM	PSNR	NRMSE
Trilinear Interpolation	0.765 \pm 0.022	25.248 \pm 1.298	0.283 \pm 0.025
Zero-padded	0.863 \pm 0.021	32.520 \pm 1.508	0.170 \pm 0.025
UNet (CHAOS Dynamic)	0.914 \pm 0.012	33.620 \pm 1.035	0.036 \pm 0.004
<i>DDoS-UNet DynMRI</i>	<i>0.951\pm0.017</i>	<i>37.557\pm2.179</i>	<i>0.024\pm0.006</i>

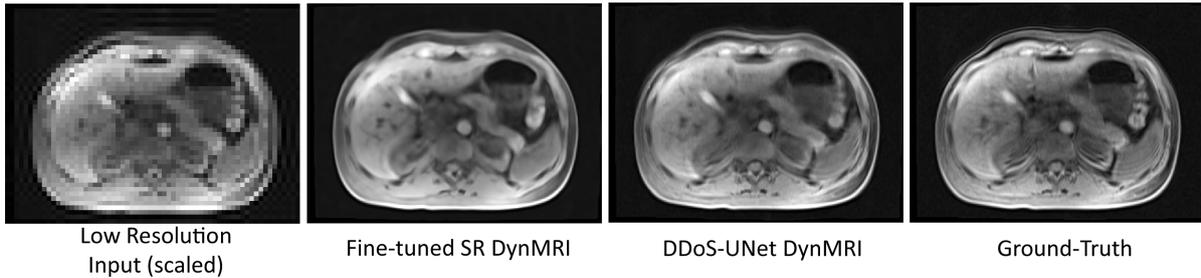


Figure 4.11: A qualitative results, compared the previously-proposed method [70] (Fine-tuned SR DynMRI) against the proposed DDoS-UNet DynMRI. The reconstructed results obtained from undersampled 4% of the k-space center.

4.5 Discussion

As shown in the qualitative and quantitative results, the proposed DDoS-UNet DynMRI framework produced more promising results than the baseline methods. The proposed method could improve the results further from deep learning based super-resolution using the 3D modified UNet model. For the highest undersampling investigated in this work (used only 4% of the k-space center), the proposed method demonstrated the average SSIM of 0.951 ± 0.017 . On the other hand, the baseline UNet model without incorporating the prior information of spatio-temporal relation resulted in the average SSIM of 0.916 ± 0.015 . In contrast, the proposed DDoS-UNet DynMRI recovered the details of small structures better than other baseline methods, which obtain only spatial information but not temporal information.

The fundamental difference between the proposed and the baseline methods is the temporal prior. The baseline UNet model only receives the low-resolution image as input; however, the DDoS-UNet DynMRI receives the super-resolved previous time-point (as the temporal prior) along with the current low-resolution image as input. Hence, the authors attribute the improvements observed with DDoS-UNet DynMRI to the addition of the prior image. The authors hypothesize that this prior image could help the network to super-resolve the current time-point as both time-points are temporally related. The results in this study support the hypothesis in equation 4.4.

Another observation relates to the Antipasto phase, referred to in Figure 4.3. In this phase, the reconstructed results were obtained with the help of a prior image from the high-resolution planning scan, while the remaining time-points were supported by a prior image from the super-resolved of an earlier time-point. Since the planning scan was acquired in a different session to the dynamic scan, the big differences of acquisition conditions, such as subject positioning or respiratory cycle, could result in a poor reconstruction during the Antipasto phase. Nonetheless, after super-resolving the initial time-point, the network could reconstruct the images with better performance consistency for rest of the

time-points. This evidence confirms the assumption of the proposed DDoS-UNet DynMRI framework, which is discussed in section 4.3.4.

In comparison with our earlier patch-based super-resolution [70], which treated each low-resolution time-point separately, DDoS-UNet DynMRI could exploit the temporal information in the dynamic MRI and be able to reconstruct the image from the higher undersampling levels (6.25% of the k-space in the previous work and 4% of the k-space data in this work). Furthermore, the fine-tuning done in the previous work requires it to be fine-tuned for every subject. Therefore, DDoS-UNet DynMRI eliminates this step and can work directly with different subjects without any additional required time. Nevertheless, the limitation of the proposed DDoS-UNet DynMRI is the computational complexity. The DDoS-UNet DynMRI model needs high performance computing to handle the 3D volumes, unlike the patch-based approach. Although the patch-based solution can be an alternative way to save memory, it is not possible for DDoS-UNet DynMRI, due to different time-points having different patches.

It can be noticed that the reconstructed results from undersampled 10% and 6.25% of the k-space of the UNet trained with CHAOS dynamic (deep learning based) and zero-padded (non-deep learning based) were not much different quantitatively for some subjects (see figure 4.9 (a) and (b)). This occurrence could arise from the quality of the dynamic training dataset. Since the original CHAOS volumetric image was interpolated by the displacement fields from one volume to another (in total 24 times), the images might be blurry and lose some anatomical details. Further improvement could be made by applying realistic human breathing motions in the future.

In addition, the reconstruction (inference) time of DDoS-UNet DynMRI was approximately 0.36 seconds for each time point (in total 9 seconds for 25 TPs) using an Nvidia Tesla V100 GPU. If the acquisition time (in Table 4.2) is taken into account, this method can acquire and reconstruct each time point within less than a minute (0.71 seconds for 4% of k-space with Protocol 1). This rapid scan time facilitates the application of near real-time or real-time MRI. Moreover, the potential of further shortening the acquisition time could be done by combining parallel imaging techniques such as GRAPPA (as presented in the previous work [70]) with the proposed DDoS-UNet DynMRI method.

4.6 Conclusions

The proposed DDoS-UNet DynMRI framework shows that incorporating the spatio-temporal information existing in time-series imaging could reconstruct the 3D dynamic MRI so as to be similar to the ground-truth, even for super-resolving from relatively low-resolution images. The results also show that the proposed framework improved

qualitatively and quantitatively compared to other methods. With the help of the dual-channel input, consisting of the temporal information in the time-series images and a subject-specific high-resolution planning scan, the proposed framework demonstrates the improvement qualitatively and quantitatively. Considering the highest undersampling achieved in this work, which is theoretically equivalent to 25 times faster in acquiring the image, such a fast speed could potentially be applied for real-time procedures, such as MRI-guided interventions.

5 Summary and Outlook

“There are no secrets to success. It is the result of preparation, hard work, and learning from failure.”

– Colin Powell

In this chapter, the details of each investigated phase in the dissertation are summarized. The contributions of this thesis lie in incorporating prior knowledge into dynamic MRI reconstruction. The objective to speed up the acquisition process was achieved using deep learning based super-resolution reconstruction. This approach could maximize the spatial information of the low-resolution dynamic image while reducing the overall scan-time using less-than-complete data. Lastly, potential future research is included in the final section.

5.1 Summary

The working steps during the course of research can be divided into the preliminary phase (subsection 5.1 and 5.1) for identifying the feasible approaches, and the further exploration of the prior-information phase (subsection 5.1). The work during phase I was commenced by using image registration without using any machine learning techniques. Thus, this part of the study is discussed briefly in this chapter, as it is not the main focus of fast dynamic MRI reconstruction. Afterward, deep learning based methods were examined to speed up the reconstruction process. The different types of prior information were incorporated in the proposed methods. At phase II, the proposed super-resolution reconstruction of dynamic MRI was improved by using fine-tuning the deep learning model parameters. Finally, the approach during phase III intended to improve the quality of the previous work with spatial and temporal information inherent in dynamic imaging through dual-channel training.

Phase I: Application of the breathing deformation model in image registration

The very first step toward the research goal, by using prior knowledge while using undersampled (less-than-complete) data for accelerating the image reconstruction, was commenced by applying breathing deformation model in image registration. The deformation model of different time-points was estimated from undersampled data (low-resolution images) using the B-Spline SyN deformable model and was applied to static high-resolution images. The extraction of breathing deformation was performed based on the assumption that breathing patterns are sparse and smooth. Although this preliminary study was not focused on the deep learning based techniques, the results of the Phase I approach could be utilized for highly undersampled data (taking only 0.5% of the k-space center) with high SSIM values - but at the expense of a long processing time during registration.

This work was published as a conference full-paper at the 41st International Engineering in Medicine and Biology Conference (EMBC 2019) [232] and the further extension of this study was accepted as a poster for the 4th Image-Guided Interventions Conference (IGIC 2019) [233].

Phase II: Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Fine-tuning with a Subject-specific Scan (Fine-tuned SR DynMRI)

Based on the time-consuming process of traditional registration, the focus of the project shifted towards the application of deep learning methods. In the second phase of the research, the reconstruction of high-resolution MRI from the corresponding low-resolution MRI using a patch-based super-resolution (SR) was proposed. A subject-specific planning scan was used for fine-tuning the SR reconstruction. The highest acceleration factors with the high evaluation matrices (SSIM and PSNR) obtained in this work were theoretically equal to a factor of 16.

The preliminary results of this work were accepted as an oral presentation at the 37th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB 2020) [147]. The subsequent results as a conference abstract were accepted at the 29th Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM 2021) [148]. The final results of this study have been accepted for publication in the Journal of Artificial Intelligence in Medicine [70].

Phase III: Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Incorporating Temporal Information in Dynamic Dual-channel UNet (DDoS-UNet DynMRI)

From phase II, deep learning based super-resolution (SR) has shown promising results in dealing with sparse data. However, the available temporal information of dynamic MRI has not been exploited in this prior work. Therefore, phase III extended the previous work by improving on the previous model [214] using the DDoS-UNet DynMRI. Incorporating temporal relationships inherently present in dynamic time series also helped to alleviate the spatio-temporal trade-off in MRI. The proposed method demonstrated more accurate results both numerically and visually compared with the other baseline methods., In addition the lowest-resolution image that the proposed framework could achieve with high accuracy was from taking only 4% of the k-space center.

The preliminary findings of this work were accepted as an oral presentation at the 38th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB) 2021, titled – “DDoS: Dynamic Dual-channel U-Net for improving deep learning based super-resolution of abdominal dynamic MRI” [214]. The additional experiment for motion-robust application was accepted as an abstract at the Joint Annual Meeting ISMRM-ESMRMB & ISMRT 31st Annual Meeting (ISMRM-ESMRMB 2022), titled - "Motion-robust dynamic abdominal MRI using kt grasp and dynamic dual-channel training of super-resolution U-Net (DDoS-UNet)" [234]. Eventually, the final results of this work were accepted as a short paper in MIDL 2022 [71] and the manuscript has been submitted to the Medical Image Analysis journal and is being reviewed [215].

5.2 Outlook

The proposed approaches in this thesis allow a relatively fast image reconstruction for dynamic MRI. Therefore, these approaches could be a candidate for near real-time applications such as interventions. Nonetheless, there are still some suggestions for further research as follows:

- Generating more realistic human breathing motions and examining more patients to improve the generalization and robustness of the frameworks.
- Experimenting with the framework with different undersampled k-space masks or trajectories (e.g., variable density Poisson) to test for higher acceleration factors.
- Simulating an interventional radiology device such as a needle or catheter into the undersampled images and investigating how the network handles the artifact, such as magnetic susceptibility, caused from the device.

- Extending the proposed frameworks to mitigate the spatio-temporal trade-off for other dynamics applications such as cardiac cine imaging.

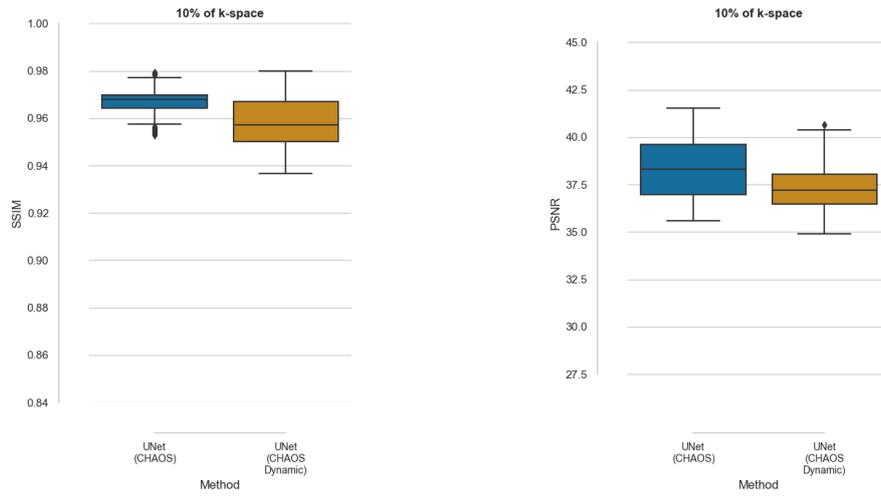
A Additional Data

Extended Results of Baseline UNets: Trained with Dynamic CHAOS Dataset compared with Trained with CHAOS Dataset (Non-dynamic)

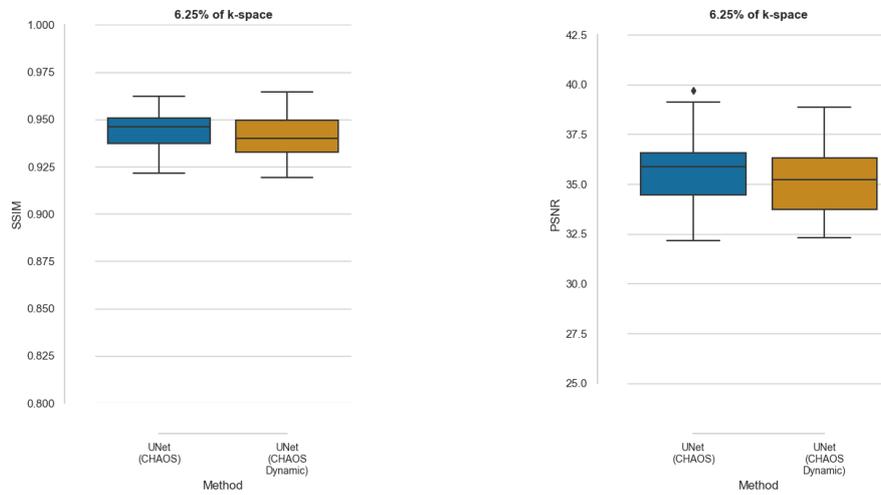
To validate the quality of CHAOS dynamic dataset on the baseline UNet models, which was artificially created from CHAOS dataset, the super-resolution (SR) results were compared by UNet trained with dynamic CHAOS and with CHAOS dataset (Non-dynamic). Figure A.1 portrays a quantitative results and A.2 shows a qualitative example results of one of the subject.

For visual comparison, the results of the UNet model trained on the CHAOS dynamic dataset could recover small structures and be less blurry than the UNet model trained on the original CHAOS dataset. However, for numerical comparison, the UNet model results that were trained on the original CHAOS dataset slightly outperformed quantitatively the UNet model trained on the CHAOS dynamic dataset, despite the latter having 25 times more volumes (24 artificially created time-points over the original one). This might be caused by the poor quality of the artificial dynamic MRI dataset. As the interpolation errors increased in the later time points from repeatedly applying the random elastic deformation to the original data and the previous time-points, the images of the later time points became blurry and less clear. Therefore, the quality of the dynamic MRI dataset should be taken care of to improve the performance of the baseline UNet model for training on the CHAOS dynamic dataset. Consequently, this can also improve the performance of the DDoS-UNet DynMRI framework.

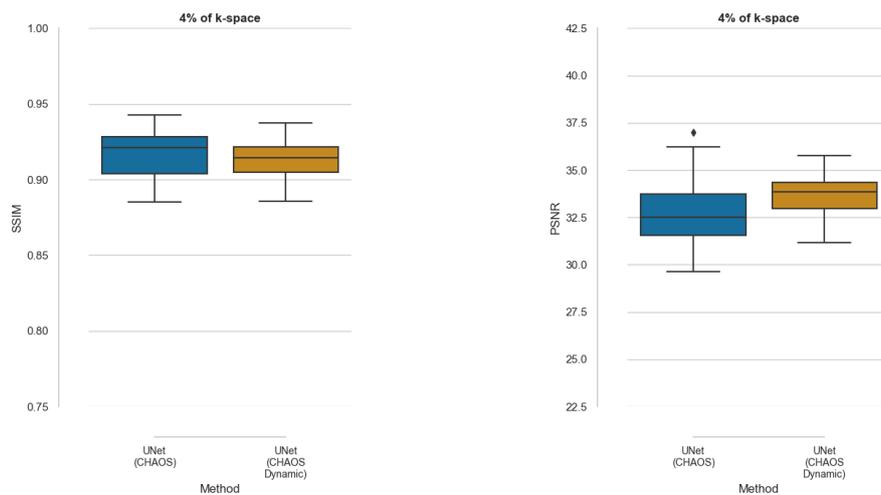
In addition, a dual-channel input network in the proposed framework manages the dual inputs by handling them evenly before merging them to the initial layer. While the other techniques for supplying more than one input, such as a dual-branch, can deal with the two branches of inputs differently (e.g., different assigning weights). This technique could be worth experimenting in the future work.



(a) 10% of the k-space



(b) 6.25% of the k-space



(c) 4% of the k-space

Figure A.1: Box plots of SSIM and PSNR of SR result of UNet baselines trained with CHAOS dataset compared with CHAOS dynamic dataset while super-resolving of 4% of the k-space center.

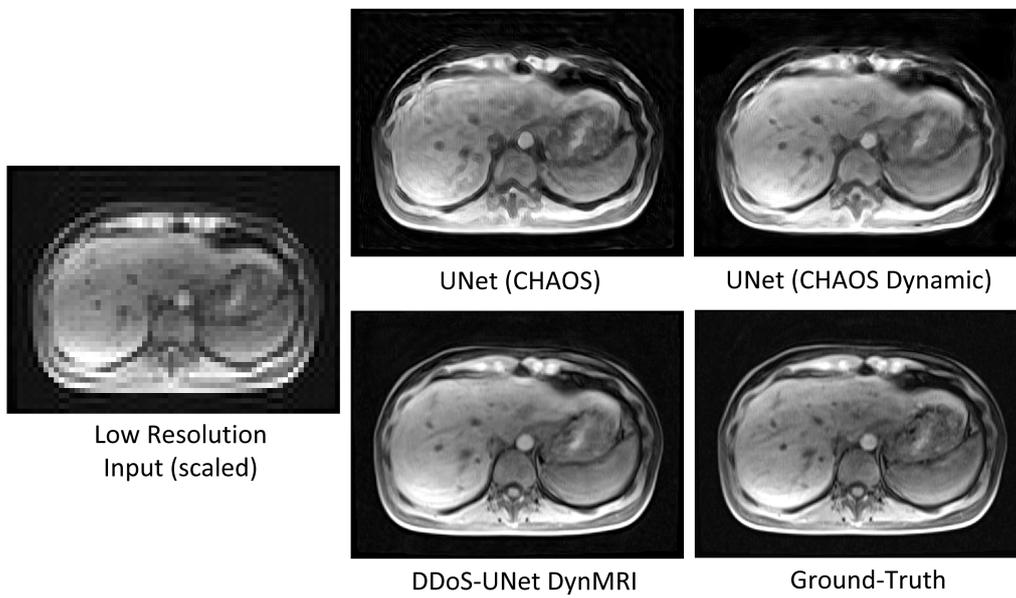


Figure A.2: A comparison results of UNet baselines trained with CHAOS dataset, compared against UNet baselines trained with CHAOS dynamic dataset and the proposed DDoS-UNet DynMRI. The results are reconstructed from 4% of the k-space center.

B Image Quality Assessment

The image quality assessment (IQA) is necessary to visualize image reconstruction quality. IQA can be categorized into full-reference, no-reference, and reduced reference. The full-reference IQA metrics, including structure similarity index (SSIM), peak signal-to-noise ratio (PSNR), and normalized root mean square error (NRMSE), were used in this thesis. The accuracy of the proposed approaches was achieved by comparing the reconstructed images to the ground truth using the mentioned metrics.

For perceptual quality assessment, the structure Similarity Index (SSIM) was selected as it can represent human perception. SSIM measures a group of pixels based on luminance, contrast, and structure terms between images x and y . The values of SSIM typically range between 0 and 1, and the value of 1 is perfectly identical. The formula of SSIM can be express as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (\text{B.1})$$

where μ_x and μ_y denotes the local means, σ_x and σ_y are standard deviations, and σ_{xy} are cross-covariance for images x and y respectively. By default, $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$, where L is the dynamic range of the pixel-values, $k_1 = 0.01$ and $k_2 = 0.03$.

Moreover, the model performance was measured statistically with PSNR and NRMSE. PSNR is a pixel-based metric, which calculates the pixel individually based on mean-square error (MSE) by using the equation:

$$PSNR = 10 \log_{10} \left(\frac{MAX_i^2}{MSE} \right), \quad (\text{B.2})$$

where MAX_I is the maximum signal value of the image. Normalized root mean square error (NRMSE) determines the variance between observed and predicted values, which can be calculated using:

$$NRMSE = \frac{RMSE * \sqrt{N}}{\|y\|}, \quad (\text{B.3})$$

where $\|\cdot\|$ represents the Frobenius norm, N is the number of data points, and y is the ground truth. Additionally, the mean-square error can be determined by:

$$MSE = \frac{1}{2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{B.4})$$

where y_i and \hat{y}_i are observed and predicted values, respectively. Lastly, the standard deviation (s), which used for measuring the differences between ground-truth and reconstructed images, is calculated using the expression:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (\text{B.5})$$

where N is the total number of samples, x_i and \bar{x} are individual value and the sample mean, respectively.

C GitHub Repository and Colab Notebook

The scripts of this work were written in Python, which can be found on GitHub and Google Colab Notebook. The implementation of the network was done using Pytorch. In addition, the scripts to visualize the results from both approaches can be found here <https://github.com/soumickmj/DDoS/visualisation>.

Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Fine-tuning with a Subject-specific Scan (Fine-tuned SR DynMRI)

Source

GitHub: <https://github.com/soumickmj/FTSuperResDynMRI> and

Hands-on Tutorial with Google Colab Notebook: <https://colab.research.google.com/drive/1DMF03PoUACzGy500NhaJLGvvUP3W1cYM?usp=sharing>

Improved Deep Learning Based Super-resolution Reconstruction of Dynamic MRI by Incorporating Temporal Information in Dynamic Dual-channel UNet (DDoS-UNet DynMRI)

Source

GitHub: <https://github.com/soumickmj/DDoS> and

Hands-on Tutorial with Google Colab Notebook: <https://colab.research.google.com/drive/1s0LzhpiiG0FxEpGcXCk79U4rcG1phkuE?usp=sharing>

Bibliography

- [1] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.
- [2] Michael Lustig, Juan M Santos, Jin-Hyung Lee, David L Donoho, and John M Pauly. Application of compressed sensing for rapid mr imaging. *SPARS,(Rennes, France)*, 2005.
- [3] Michael Lustig, Juan M Santos, David L Donoho, and John M Pauly. kt sparse: High frame rate dynamic mri exploiting spatio-temporal sparsity. In *Proceedings of the 13th annual meeting of ISMRM, Seattle*, volume 2420, 2006.
- [4] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.
- [5] Daniel K Sodickson and Charles A McKenzie. A generalized approach to parallel magnetic resonance imaging. *Medical physics*, 28(8):1629–1643, 2001.
- [6] Robin M Heidemann, Özkan Özsarlak, Paul M Parizel, Johan Michiels, Berthold Kiefer, Vladimir Jellus, Mathias Müller, Felix Breuer, Martin Blaimer, Mark A Griswold, et al. A brief review of parallel magnetic resonance imaging. *European radiology*, 13(10):2323–2337, 2003.
- [7] Daniel K Sodickson and Warren J Manning. Simultaneous acquisition of spatial harmonics (smash): fast imaging with radiofrequency coil arrays. *Magnetic resonance in medicine*, 38(4):591–603, 1997.
- [8] Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.
- [9] Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic Resonance in Medicine: An*

- Official Journal of the International Society for Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.
- [10] Kun Zeng, Hong Zheng, Congbo Cai, Yu Yang, Kaihua Zhang, and Zhong Chen. Simultaneous single-and multi-contrast super-resolution for brain mri images based on a convolutional neural network. *Computers in biology and medicine*, 99:133–141, 2018.
- [11] Xiuxiu He, Yang Lei, Yabo Fu, Hui Mao, Walter J Curran, Tian Liu, and Xiaofeng Yang. Super-resolution magnetic resonance imaging reconstruction using deep attention networks. In *Medical Imaging 2020: Image Processing*, volume 11313, page 113132J. International Society for Optics and Photonics, 2020.
- [12] Chen Qin, Jo Schlemper, Jose Caballero, Anthony N Price, Joseph V Hajnal, and Daniel Rueckert. Convolutional recurrent neural networks for dynamic mr image reconstruction. *IEEE transactions on medical imaging*, 38(1):280–290, 2018.
- [13] Qing Lyu, Hongming Shan, Yibin Xie, Alan C Kwan, Yuka Otaki, Keiichiro Kuronuma, Debiao Li, and Ge Wang. Cine cardiac mri motion artifact reduction using a recurrent neural network. *IEEE Transactions on Medical Imaging*, 40(8):2170–2181, 2021.
- [14] Julian Rasch, Ville Kolehmainen, Riikka Nivajärvi, Mikko Kettunen, Olli Gröhn, Martin Burger, and Eva-Maria Brinkmann. Dynamic mri reconstruction from undersampled data with an anatomical prescan. *Inverse problems*, 34(7):074001, 2018.
- [15] Andreas Kofler, Marc Dewey, Tobias Schaeffter, Christian Wald, and Christoph Kolbitsch. Spatio-temporal deep learning-based undersampling artefact reduction for 2d radial cine mri with limited training data. *IEEE transactions on medical imaging*, 39(3):703–717, 2019.
- [16] Thomas Küstner, Niccolo Fuin, Kerstin Hammernik, Aurelien Bustin, Haikun Qi, Reza Hajhosseiny, Pier Giorgio Masci, Radhouene Neji, Daniel Rueckert, René M Botnar, et al. Cinenet: deep learning-based 3d cardiac cine mri reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions. *Scientific reports*, 10(1):1–13, 2020.
- [17] C Andrew Segall, Aggelos K Katsaggelos, Rafael Molina, and Javier Mateos. Bayesian resolution enhancement of compressed video. *IEEE Transactions on image processing*, 13(7):898–911, 2004.
- [18] Stefanos P Belekos, Nikolaos P Galatsanos, and Aggelos K Katsaggelos. Maximum

- a posteriori video super-resolution using a new multichannel image prior. *IEEE Transactions on Image Processing*, 19(6):1451–1464, 2010.
- [19] Shoko Araki, Tomoki Hayashi, Marc Delcroix, Masakiyo Fujimoto, Kazuya Takeda, and Tomohiro Nakatani. Exploring multi-channel features for denoising-autoencoder-based speech enhancement. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 116–120. IEEE, 2015.
- [20] Pablo Barros, Sven Magg, Cornelius Weber, and Stefan Wermter. A multichannel convolutional neural network for hand posture recognition. In *International Conference on Artificial Neural Networks*, pages 403–410. Springer, 2014.
- [21] Zhong-Qiu Wang, Jonathan Le Roux, and John R Hershey. Multi-channel deep clustering: Discriminative spectral and spatial embeddings for speaker-independent speech separation. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2018.
- [22] Aditya Arie Nugraha, Antoine Liutkus, and Emmanuel Vincent. Multichannel audio source separation with deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(9):1652–1664, 2016.
- [23] Chang Xu, Weiran Huang, Hongwei Wang, Gang Wang, and Tie-Yan Liu. Modeling local dependence in natural language with multi-channel recurrent neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5525–5532, 2019.
- [24] Gengsheng Lawrence Zeng. *Medical image reconstruction: a conceptual tutorial*. Springer, 2010.
- [25] Donald W McRobbie, Elizabeth A Moore, Martin J Graves, and Martin R Prince. *MRI from Picture to Proton*. Cambridge university press, 2017.
- [26] Michael J Ackerman. The visible human project: From body to bits. *IEEE Pulse*, 8(4):39–41, 2017.
- [27] Peter Reimer, Paul M Parizel, James FM Meaney, and Falko A Stichnoth. *Clinical MR imaging*. Springer, 2010.
- [28] Joseph Larmor. Lxiii. on the theory of the magnetic influence on spectra; and on the radiation from moving ions. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44(271):503–512, 1897.
- [29] Felix Bloch and WWPM Hansen. W., packard, m. the nuclear induction experiment. *Physical review*, 70(7-8):474–485, 1946.

- [30] Mark A Brown and Richard C Semelka. *MRI: basic principles and applications*. John Wiley & Sons, 2011.
- [31] John P Ridgway. Cardiovascular magnetic resonance physics for clinicians: part i. *Journal of cardiovascular magnetic resonance*, 12(1):1–28, 2010.
- [32] Daphne Caligari Conti. Magnetic resonance imaging. 03 2016.
- [33] Mark E Ladd, Peter Bachert, Martin Meyerspeer, Ewald Moser, Armin M Nagel, David G Norris, Sebastian Schmitter, Oliver Speck, Sina Straub, and Moritz Zaiss. Pros and cons of ultra-high-field mri/mrs for human application. *Progress in nuclear magnetic resonance spectroscopy*, 109:1–50, 2018.
- [34] Donald B Twieg. The k-trajectory formulation of the nmr imaging process with applications in analysis and synthesis of imaging methods. *Medical physics*, 10(5):610–621, 1983.
- [35] Wikipedia. Heinrich Kayser — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Heinrich%20Kayser&oldid=1076330341>, 2022. [Online; accessed 19-March-2022].
- [36] Penelope J Allisy-Roberts and Jerry Williams. *Farr’s physics for medical imaging*. Elsevier Health Sciences, 2007.
- [37] American College of Radiology. Mri accreditation program clinical image quality guide. https://www.kansasradiationphysics.com/uploads/2/2/4/6/22462248/acr_mri_image_quality_guide__5.22.13.pdf.
- [38] Nicholas T Carnevale and Michael L Hines. *The NEURON book*. Cambridge University Press, 2006.
- [39] David Hilbert. Über das unendliche. *Mathematische Annalen*, 95(1):161–190, 1926.
- [40] D Hilbert and W Ackerman. Theoretische logik. *Julius Springer, Berlin*, 1928.
- [41] Stephen Cole Kleene. General recursive functions of natural numbers. *Mathematische annalen*, 112(1):727–742, 1936.
- [42] Alonzo Church. A note on the entscheidungsproblem. *The journal of symbolic logic*, 1(1):40–41, 1936.
- [43] Alan Mathison Turing et al. On computable numbers, with an application to the entscheidungsproblem. *J. of Math*, 58(345-363):5, 1936.
- [44] Raúl Rojas. *Neural networks: a systematic introduction*. Springer Science & Business Media, 2013.

- [45] John Von Neumann, Arthur W Burks, et al. Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1):3–14, 1966.
- [46] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [47] Norbert Wiener. *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 2019.
- [48] Richard S Snell. *Clinical neuroanatomy*. Lippincott Williams & Wilkins, 2010.
- [49] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [50] Gerald M Long and Thomas C Toppino. Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychological bulletin*, 130(5):748, 2004.
- [51] Jürgen Kornmeier and Michael Bach. Ambiguous figures—what happens in the brain when perception changes but not the stimulus. *Frontiers in human neuroscience*, 6:51, 2012.
- [52] Megan Wang, Daniel Arteaga, and Biyu J He. Brain mechanisms for simple perception and bistable perception. *Proceedings of the National Academy of Sciences*, 110(35):E3350–E3359, 2013.
- [53] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [54] Mathworks. CNN architecture. https://www.mathworks.com/discovery/convolutional-neural-network-matlab/_jcr_content/mainParsys3.html. Accessed: 2022-06-24.
- [55] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [56] Li Deng and Dong Yu. Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3–4):197–387, 2014.
- [57] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [58] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

- [59] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [60] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [61] Ossama Abdel-Hamid, Li Deng, and Dong Yu. Exploring convolutional neural network structures and optimization techniques for speech recognition. In *Interspeech*, volume 2013, pages 1173–5. Citeseer, 2013.
- [62] Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*, pages 1681–1691, 2015.
- [63] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR, 2015.
- [64] Yoav Goldberg. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57:345–420, 2016.
- [65] Jianpeng Cheng and Mirella Lapata. Neural summarization by extracting sentences and words. *arXiv preprint arXiv:1603.07252*, 2016.
- [66] Ritambhara Singh, Jack Lanchantin, Gabriel Robins, and Yanjun Qi. Deepchrome: deep-learning for predicting gene expression from histone modifications. *Bioinformatics*, 32(17):i639–i648, 2016.
- [67] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. Deep learning in bioinformatics. *Briefings in bioinformatics*, 18(5):851–869, 2017.
- [68] Wentao Zhu, Yufang Huang, Liang Zeng, Xuming Chen, Yong Liu, Zhen Qian, Nan Du, Wei Fan, and Xiaohui Xie. Anatomynet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics*, 46(2):576–589, 2019.
- [69] Soumick Chatterjee, Fatima Saad, Chompunuch Sarasaen, Suhita Ghosh, Rupali Khatun, Petia Radeva, Georg Rose, Sebastian Stober, Oliver Speck, and Andreas

- Nürnbergger. Exploration of interpretability techniques for deep covid-19 classification using chest x-ray images. *arXiv preprint arXiv:2006.02570*, 2020.
- [70] Chompunuch Sarasaen, Soumick Chatterjee, Mario Bretkopf, Georg Rose, Andreas Nürnbergger, and Oliver Speck. Fine-tuning deep learning model parameters for improved super-resolution of dynamic mri with prior-knowledge. *Artificial Intelligence in Medicine*, 121:102196, 2021.
- [71] Soumick Chatterjee, Chompunuch Sarasaen, Georg Rose, Andreas Nürnbergger, and Oliver Speck. DDoS-UNet: Incorporating temporal information using dynamic dual-channel unet for enhancing super-resolution of dynamic mri. In *Medical Imaging with Deep Learning (MIDL)*, Zürich, Switzerland, 7 2022.
- [72] Yuexiang Li and Linlin Shen. A deep residual inception network for hep-2 cell classification. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 12–20. Springer, 2017.
- [73] Yosra A Helmy, Mohamed Fawzy, Ahmed Elswad, Ahmed Sobieh, Scott P Kenney, and Awad A Shehata. The covid-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *Journal of clinical medicine*, 9(4):1225, 2020.
- [74] Zhong Qiu Lin Wang, Linda and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1):1–12, 2020.
- [75] Marwa A. Shouman Hemdan, Ezz El-Din and Mohamed Esmail Karar. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055*, 2020.
- [76] Hao Tang, Xuming Chen, Yang Liu, Zhipeng Lu, Junhua You, Mingzhou Yang, Shengyu Yao, Guoqi Zhao, Yi Xu, Tingfeng Chen, et al. Clinically applicable deep learning framework for organs at risk delineation in ct images. *Nature Machine Intelligence*, 1(10):480–491, 2019.
- [77] Jaeyeon Yoon, Enhao Gong, Itthi Chatnuntawech, Berkin Bilgic, Jingu Lee, Woojin Jung, Jingyu Ko, Hosan Jung, Kawin Setsompop, Greg Zaharchuk, et al. Quantitative susceptibility mapping using deep neural network: Qsmnet. *Neuroimage*, 179:199–206, 2018.
- [78] Qing Lyu, Hongming Shan, Yibin Xie, Debiao Li, and Ge Wang. Cine cardiac mri motion artifact reduction using a recurrent neural network. *arXiv preprint arXiv:2006.12700*, 2020.

- [79] Eric I Knudsen. Supervised learning in the brain. *Journal of Neuroscience*, 14(7):3985–3997, 1994.
- [80] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168, 2006.
- [81] Horace B Barlow. Unsupervised learning. *Neural computation*, 1(3):295–311, 1989.
- [82] Mario A. T. Figueiredo and Anil K. Jain. Unsupervised learning of finite mixture models. *IEEE Transactions on pattern analysis and machine intelligence*, 24(3):381–396, 2002.
- [83] M Emre Celebi and Kemal Aydin. *Unsupervised learning algorithms*. Springer, 2016.
- [84] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. 2005.
- [85] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.
- [86] Shai Ben-David, Tyler Lu, and Dávid Pál. Does unlabeled data provably help? worst-case analysis of the sample complexity of semi-supervised learning. In *COLT*, pages 33–44, 2008.
- [87] Fábio Gagliardi Cozman and Ira Cohen. Risks of semi-supervised learning: How unlabeled data can degrade performance of generative classifiers. *Semi-supervised learning*, 2006.
- [88] Nicholas C Gourtsoyiannis and Philip Aschoff. *Clinical MRI of the abdomen: why, how, when*. Springer, 2011.
- [89] B Wu, G Warnock, M Zaiss, C Lin, M Chen, Z Zhou, L Mu, D Nanz, R Tuura, and G Delso. An overview of cest mri for non-mr physicists. *EJNMMI physics*, 3(1):1–21, 2016.
- [90] Andreas H Mahnken, Jens Ricke, and Kai E Wilhelm. *CT-and MR-guided interventions in radiology*, volume 22. Springer, 2009.
- [91] Scott Tashman. Comments on,“validation of a non-invasive fluoroscopic imaging technique for the measurement of dynamic knee joint motion”. *Journal of biomechanics*, 41(15):3290, 2008.
- [92] Ying Zhang, Kazushi Numata, Yuewu Du, and Shin Maeda. Contrast agents for

- hepatocellular carcinoma imaging: Value and progression. *Frontiers in Oncology*, 12, 2022.
- [93] Regina GH Beets-Tan, Wim JG Oyen, and Vincenzo Valentini. *Imaging and Interventional Radiology for Radiation Oncology*. Springer.
- [94] Andrea S Kierans, Polytimi Leonardou, Faiq Shaikh, and Richard C Semelka. Body mr imaging: sequences we use and why. *Appl Radiol*, 38(5):7–12, 2009.
- [95] W Thomas Dixon. Simple proton spectroscopic imaging. *Radiology*, 153(1):189–194, 1984.
- [96] David A Feinberg, James D Hale, Jeffrey C Watts, Leon Kaufman, and Alexander Mark. Halving mr imaging time by conjugation: demonstration at 3.5 kg. *Radiology*, 161(2):527–531, 1986.
- [97] Paul Margosian, F Schmitt, and D Purdy. Faster mr imaging: imaging with half the data. *Health Care Instrum*, 1(6):195, 1986.
- [98] Douglas C Noll, Dwight G Nishimura, and Albert Macovski. Homodyne detection in magnetic resonance imaging. *IEEE transactions on medical imaging*, 10(2):154–163, 1991.
- [99] Jan Cuppen and Andre van Est. Reducing mr imaging time by one-sided reconstruction. *Magnetic Resonance Imaging*, 5(6):526–527, 1987.
- [100] E Mark Haacke, ED Lindskog, and Weili Lin. A fast, iterative, partial-fourier technique capable of local phase recovery. *Journal of Magnetic Resonance (1969)*, 92(1):126–145, 1991.
- [101] Frank R Korosec, Richard Frayne, Thomas M Grist, and Charles A Mistretta. Time-resolved contrast-enhanced 3d mr angiography. *Magnetic resonance in medicine*, 36(3):345–351, 1996.
- [102] Jürgen Hennig, Klaus Scheffler, Jörg Laubenberger, and Ralf Strecker. Time-resolved projection angiography after bolus injection of contrast agent. *Magnetic resonance in medicine*, 37(3):341–345, 1997.
- [103] Manojkumar Saranathan, Dan W Rettmann, Brian A Hargreaves, Sharon E Clarke, and Shreyas S Vasanawala. Differential subsampling with cartesian ordering (disco): a high spatio-temporal resolution dixon imaging sequence for multiphasic contrast enhanced abdominal imaging. *Journal of Magnetic Resonance Imaging*, 35(6):1484–1492, 2012.
- [104] Hee Kwon Song and Lawrence Dougherty. k-space weighted image contrast (kwic)

- for contrast manipulation in projection reconstruction mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 44(6):825–832, 2000.
- [105] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- [106] Shane F Cotter and Bhaskar D Rao. Sparse channel estimation via matching pursuit with application to equalization. *IEEE Transactions on Communications*, 50(3):374–377, 2002.
- [107] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51, 2008.
- [108] Mark A Davenport, Petros T Boufounos, Michael B Wakin, and Richard G Baraniuk. Signal processing with compressive measurements. *IEEE Journal of Selected topics in Signal processing*, 4(2):445–460, 2010.
- [109] Caglar Yardim, Peter Gerstoft, William S Hodgkiss, and James Traer. Compressive geoacoustic inversion using ambient noise. *The Journal of the Acoustical Society of America*, 135(3):1245–1255, 2014.
- [110] Michael Bianco and Peter Gerstoft. Compressive acoustic sound speed profile estimation. *The Journal of the Acoustical Society of America*, 139(3):EL90–EL94, 2016.
- [111] Youngmin Choo and Woojae Seong. Compressive sound speed profile inversion using beamforming results. *Remote Sensing*, 10(5):704, 2018.
- [112] Jae Young Park and Michael B Wakin. A multiscale framework for compressive sensing of video. In *2009 Picture Coding Symposium*, pages 1–4. IEEE, 2009.
- [113] Scott Pudlewski, Arvind Prasanna, and Tommaso Melodia. Compressed-sensing-enabled video streaming for wireless multimedia sensor networks. *IEEE Transactions on Mobile Computing*, 11(6):1060–1072, 2012.
- [114] Irena Orović, Seri Park, and Srdjan Stanković. Compressive sensing in video applications. In *2013 21st Telecommunications Forum Telfor (TELFOR)*, pages 745–748. IEEE, 2013.
- [115] Juan M Santos, Charles H Cunningham, Michael Lustig, Brian A Hargreaves, Bob S Hu, Dwight G Nishimura, and John M Pauly. Single breath-hold whole-heart mra using variable-density spirals at 3t. *Magnetic Resonance in Medicine: An*

- Official Journal of the International Society for Magnetic Resonance in Medicine*, 55(2):371–379, 2006.
- [116] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing mri. *IEEE signal processing magazine*, 25(2):72–82, 2008.
- [117] Urs Gamper, Peter Boesiger, and Sebastian Kozerke. Compressed sensing in dynamic mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 59(2):365–373, 2008.
- [118] Jim Ji and Tao Lang. Dynamic mri with compressed sensing imaging using temporal correlations. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1613–1616. IEEE, 2008.
- [119] Hong Jung, Kyunghyun Sung, Krishna S Nayak, Eung Yeop Kim, and Jong Chul Ye. k-t focuss: a general compressed sensing framework for high resolution dynamic mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 61(1):103–116, 2009.
- [120] C.E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, jan 1949.
- [121] Harry Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.
- [122] Jong Chul Ye. Compressed sensing mri: a review from signal processing perspective. *BMC Biomedical Engineering*, 1(1):1–17, 2019.
- [123] Felix A Breuer, Martin Blaimer, Robin M Heidemann, Matthias F Mueller, Mark A Griswold, and Peter M Jakob. Controlled aliasing in parallel imaging results in higher acceleration (caipirinha) for multi-slice imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 53(3):684–691, 2005.
- [124] Kawin Setsompop, J Cohen-Adad, J McNab, BA Gagoski, VJ Wedeen, and LL Wald. Improving snr per unit time in diffusion imaging using a blipped-caipirinha simultaneous multi-slice epi acquisition. In *Proceedings of the 18th Annual Meeting of ISMRM*, page 187, 2010.
- [125] Jeffrey Tsao, Peter Boesiger, and Klaas P Pruessmann. k-t blast and k-t sense: dynamic mri with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 50(5):1031–1042, 2003.

- [126] Feng Huang, James Akao, Sathya Vijayakumar, George R Duensing, and Mark Limkeman. k-t grappa: A k-space implementation for dynamic mri with high reduction factor. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 54(5):1172–1184, 2005.
- [127] Dan Xu, Kevin F King, and Zhi-Pei Liang. Improving k-t sense by adaptive regularization. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 57(5):918–930, 2007.
- [128] Jim X Ji, Chen Zhao, and Tao Lang. Compressed sensing parallel magnetic resonance imaging. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1671–1674. IEEE, 2008.
- [129] Dong Liang, Bo Liu, Jiunjie Wang, and Leslie Ying. Accelerating sense using compressed sensing. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 62(6):1574–1584, 2009.
- [130] JE Vranic, NM Cross, Y Wang, DS Hippe, E De Weerd, and M Mossa-Basha. Compressed sensing–sensitivity encoding (cs-sense) accelerated brain imaging: Reduced scan time without reduced image quality. *American Journal of Neuroradiology*, 40(1):92–98, 2019.
- [131] Marshall F Tappen and Ce Liu. A bayesian approach to alignment-based image hallucination. In *European conference on computer vision*, pages 236–249. Springer, 2012.
- [132] Xin Yu, Basura Fernando, Bernard Ghanem, Fatih Porikli, and Richard Hartley. Face super-resolution guided by facial component heatmaps. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 217–233, 2018.
- [133] Markus U Müller, Nikoo Ekhtiari, Rodrigo M Almeida, and Christoph Rieke. Super-resolution of multispectral satellite images using convolutional neural networks. *arXiv preprint arXiv:2002.00580*, 2020.
- [134] Jithin Saji Isaac and Ramesh Kulkarni. Super resolution techniques for medical image processing. In *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, pages 1–6. IEEE, 2015.
- [135] Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6070–6079, 2017.
- [136] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image

- reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.
- [137] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.
- [138] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.
- [139] Yu Zhu, Yanning Zhang, and Alan L Yuille. Single image super-resolution using deformable patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2917–2924, 2014.
- [140] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016.
- [141] Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *arXiv preprint arXiv:1701.05517*, 2017.
- [142] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017.
- [143] Chang Min Hyun, Hwa Pyung Kim, Sung Min Lee, Sungchul Lee, and Jin Keun Seo. Deep learning for undersampled mri reconstruction. *Physics in Medicine & Biology*, 63(13):135007, 2018.
- [144] Zohaib Iqbal, Dan Nguyen, Gilbert Hangel, Stanislav Motyka, Wolfgang Bogner, and Steve Jiang. Super-resolution 1h magnetic resonance spectroscopic imaging utilizing deep learning. *Frontiers in oncology*, 9, 2019.
- [145] Xiaodan Hu, Mohamed A Naiel, Alexander Wong, Mark Lamm, and Paul Fieguth. Runet: A robust unet architecture for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.
- [146] Qiong Ran, Xiaodong Xu, Shizhi Zhao, Wei Li, and Qian Du. Remote sensing images super-resolution with deep convolution networks. *Multimedia Tools and Applications*, 79(13):8985–9001, 2020.

- [147] Chompunuch Sarasaen, Soumick Chatterjee, Andreas Nürnberger, and Oliver Speck. Super resolution of dynamic mri using deep learning, enhanced by prior-knowledge. In *37th Annual Scientific Meeting Congress of the European Society for Magnetic Resonance in Medicine and Biology, 33(Supplement 1): S03.04, S28-S29*. Springer, 10 2020.
- [148] Chompunuch Sarasaen, Soumick Chatterjee, Fatima Saad, Mario Breitkopf, Andreas Nürnberger, and Oliver Speck. Fine-tuning deep learning model parameters for improved super-resolution of dynamic mri with prior-knowledge. In *Proc. The Annual Meeting of The International Society for Magnetic Resonance in Medicine (ISMRM)*, 5 2021.
- [149] Yi Tang and Ling Shao. Pairwise operator learning for patch-based single-image super-resolution. *IEEE Transactions on Image Processing*, 26(2):994–1003, 2016.
- [150] Debaleena Misra, Carlos Crispim-Junior, and Laure Tougne. Patch-based cnn evaluation for bark classification. In *European Conference on Computer Vision*, pages 197–212. Springer, 2020.
- [151] Qiyuan Tian, Berkin Bilgic, Qiuyun Fan, Chanon Ngamsombat, Natalia Zaretskaya, Nina E Fultz, Ned A Ohringer, Akshay S Chaudhari, Yuxin Hu, Thomas Witzel, et al. Improving in vivo human cerebral cortical surface reconstruction using data-driven super-resolution. *Cerebral Cortex*, 31(1):463–482, 2021.
- [152] Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang. Single-image super-resolution: A benchmark. In *European Conference on Computer Vision*, pages 372–386. Springer, 2014.
- [153] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, pages 2392–2399. IEEE, 2012.
- [154] Jonas De Vylder, Simon Donne, Hiep Luong, and Wilfried Philips. Image restoration using deep learning. In *25th Belgian-Dutch Conference on Machine Learning (Benelearn)*, 2016.
- [155] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*, 2016.
- [156] Toru Higaki, Yuko Nakamura, Fuminari Tatsugami, Takeshi Nakaura, and Kazuo Awai. Improvement of image quality at ct and mri using deep learning. *Japanese journal of radiology*, 37(1):73–80, 2019.

- [157] Takuma Yamaguchi, Hisato Fukuda, Ryo Furukawa, Hiroshi Kawasaki, and Peter Sturm. Video deblurring and super-resolution technique for multiple moving objects. In *Asian conference on computer vision*, pages 127–140. Springer, 2010.
- [158] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4778–4787, 2017.
- [159] Alice Lucas, Santiago Lopez-Tapia, Rafael Molina, and Aggelos K Katsaggelos. Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Transactions on Image Processing*, 28(7):3312–3327, 2019.
- [160] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.
- [161] Masitah Abdul Lateh, Azah Kamilah Muda, Zeratul Izzah Mohd Yusof, Noor Azilah Muda, and Mohd Sanusi Azmi. Handling a small dataset problem in prediction model by employ artificial data generation approach: A review. In *Journal of Physics: Conference Series*, volume 892, page 012016, 2017.
- [162] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.
- [163] Xinbo Gao, Kaibing Zhang, Dacheng Tao, and Xuelong Li. Image super-resolution with sparse neighbor embedding. *IEEE Transactions on Image Processing*, 21(7):3194–3205, 2012.
- [164] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*, pages 111–126. Springer, 2014.
- [165] Yu Zhang, Guorong Wu, Pew-Thian Yap, Qianjin Feng, Jun Lian, Wufan Chen, and Dinggang Shen. Reconstruction of super-resolution lung 4d-ct using patch-based sparse representation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 925–931. IEEE, 2012.
- [166] Pierrick Coupé, José V Manjón, Maxime Chamberland, Maxime Descoteaux, and Bassem Hiba. Collaborative patch-based super-resolution for diffusion-weighted images. *NeuroImage*, 83:245–261, 2013.
- [167] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image

- restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016.
- [168] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
- [169] Jinchang Xu, Yu Zhao, Yuan Dong, and Hongliang Bai. Fast and accurate image super-resolution using a combined loss. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 97–103, 2017.
- [170] Vahid Ghodrati, Jiaxin Shao, Mark Bydder, Ziwu Zhou, Wotao Yin, Kim-Lien Nguyen, Yingli Yang, and Peng Hu. Mr image reconstruction using deep learning: evaluation of network structure and loss functions. *Quantitative imaging in medicine and surgery*, 9(9):1516–1527, 2019.
- [171] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neuro-computing*, 312:135–153, 2018.
- [172] Garrett Wilson and Diane J Cook. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–46, 2020.
- [173] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2216–2223. IEEE, 2012.
- [174] José V Manjón, Pierrick Coupé, Antonio Buades, Vladimir Fonov, D Louis Collins, and Montserrat Robles. Non-local mri upsampling. *Medical image analysis*, 14(6):784–792, 2010.
- [175] François Rousseau, Alzheimer’s Disease Neuroimaging Initiative, et al. A non-local approach for image super-resolution using intermodality priors. *Medical image analysis*, 14(4):594–605, 2010.
- [176] Saurabh Jain, Diana M Sima, Faezeh Sanaei Nezhad, Gilbert Hangel, Wolfgang Bogner, Stephen Williams, Sabine Van Huffel, Frederik Maes, and Dirk Smeets. Patch-based super-resolution of mr spectroscopic images: application to multiple sclerosis. *Frontiers in neuroscience*, 11:13, 2017.
- [177] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Modeling the intra-class variability for liver lesion detection using a multi-class patch-based cnn. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 129–137. Springer, 2017.

- [178] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [179] Hoyeon Lee, Jongha Lee, Hyeongseok Kim, Byungchul Cho, and Seungryong Cho. Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2):109–119, 2018.
- [180] Zhun Wei and Xudong Chen. Deep-learning schemes for full-wave nonlinear inverse scattering problems. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4):1849–1860, 2018.
- [181] Zhengyang Lu and Ying Chen. Single image super resolution based on a modified u-net with mixed gradient loss. *arXiv e-prints*, pages arXiv–1911, 2019.
- [182] Feixiang Luo, Jun Wang, Jie Zeng, Lu Zhang, Boyu Zhang, Kuiwen Xu, and Xiling Luo. Cascaded complex u-net model to solve inverse scattering problems with phaseless-data in the complex domain. *IEEE Transactions on Antennas and Propagation*, 2021.
- [183] Younghyun Jo, Sejong Yang, and Seon Joo Kim. Investigating loss functions for extreme super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 424–425, 2020.
- [184] Navchetan Awasthi, Gaurav Jain, Sandeep Kumar Kalva, Manojit Pramanik, and Phaneendra K Yalavarthy. Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 67(12):2660–2673, 2020.
- [185] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021.
- [186] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.
- [187] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [188] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham,

- Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.
- [189] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
- [190] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018.
- [191] Guang Yang, Yunhua Cao, Xiaoyang Xing, and Min Wei. Perceptual loss based super-resolution reconstruction from single magnetic resonance imaging. In *International Conference on Artificial Intelligence and Security*, pages 411–424. Springer, 2019.
- [192] Daming Shen, Sushobhan Ghosh, Hassan Haji-Valizadeh, Ashitha Pathrose, Florian Schiffers, Daniel C Lee, Benjamin H Freed, Michael Markl, Oliver S Cossairt, Aggelos K Katsaggelos, et al. Rapid reconstruction of highly undersampled, non-cartesian real-time cine k-space data using a perceptual complex neural network (pcnn). *NMR in Biomedicine*, 34(1):e4405, 2021.
- [193] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*, volume 1. MIT press Massachusetts, USA:, 2017.
- [194] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27, 2014.
- [195] Kasthurirangan Gopalakrishnan, Siddhartha K Khaitan, Alok Choudhary, and Ankit Agrawal. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and building materials*, 157:322–330, 2017.
- [196] LI Xuhong, Yves Grandvalet, and Franck Davoine. Explicit inductive bias for transfer learning with convolutional networks. In *International Conference on Machine Learning*, pages 2825–2834. PMLR, 2018.
- [197] Zhilin Yang, Jake Zhao, Bhuwan Dhingra, Kaiming He, William W Cohen, Ruslan Salakhutdinov, and Yann LeCun. Glomo: Unsupervisedly learned relational graphs as transferable representations. *arXiv preprint arXiv:1806.05662*, 2018.

- [198] Simon Kornblith, Jonathon Shlens, and Quoc V Le. Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2661–2671, 2019.
- [199] Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. Transfer learning for low-resource neural machine translation. *arXiv preprint arXiv:1604.02201*, 2016.
- [200] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [201] Sebastian Ruder, Matthew E Peters, Swabha Swayamdipta, and Thomas Wolf. Transfer learning in natural language processing. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials*, pages 15–18, 2019.
- [202] Nelson F Liu, Matt Gardner, Yonatan Belinkov, Matthew E Peters, and Noah A Smith. Linguistic knowledge and transferability of contextual representations. *arXiv preprint arXiv:1903.08855*, 2019.
- [203] Wei Zhao. Research on the deep learning of the small sample data based on transfer learning. In *AIP Conference Proceedings*, volume 1864, page 020018. AIP Publishing LLC, 2017.
- [204] Young-Gon Kim, Sungchul Kim, Cristina Eunbee Cho, In Hye Song, Hee Jin Lee, Soomin Ahn, So Yeon Park, Gyungyub Gong, and Namkug Kim. Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections. *Scientific Reports*, 10(1):1–9, 2020.
- [205] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [206] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [207] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [208] A Emre Kavur, N Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan,

- et al. Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. *Medical Image Analysis*, page 101950, 2020.
- [209] O Dietrich, K Nikolaou, BJ Wintersperger, W Flatz, M Nittka, R Petsch, B Kiefer, and SO Schoenberg. ipat: applications for fast and cardiovascular mr imaging. *ELECTROMEDICA-ERLANGEN-*, 70(2):133–146, 2002.
- [210] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
- [211] Soumick Chatterjee, Kartik Prabhu, Mahantesh Pattadkal, Gerda Bortsova, Chompunuch Sarasaen, Florian Dubost, Hendrik Mattern, Marleen de Bruijne, Oliver Speck, and Andreas Nürnberger. Ds6, deformation-aware semi-supervised learning: Application to small vessel segmentation with noisy training data. *Journal of Imaging*, 8(10), 2022.
- [212] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [213] Kalpana Seshadrinathan, Thrasyvoulos N Pappas, Robert J Safranek, Junqing Chen, Zhou Wang, Hamid R Sheikh, and Alan C Bovik. Image quality assessment. In *The essential guide to image processing*, pages 553–595. Elsevier, 2009.
- [214] Chompunuch Sarasaen, Soumick Chatterjee, Andreas Nürnberger, and Oliver Speck. DDoS: Dynamic dual-channel u-net for improving deep learning based super-resolution of abdominal dynamic mri. In *38th Annual Scientific Meeting Congress of the European Society for Magnetic Resonance in Medicine and Biology, 34(Supplement 1): S6.O3., S44*. Springer, 10 2021.
- [215] Soumick Chatterjee, Chompunuch Sarasaen, Georg Rose, Andreas Nürnberger, and Oliver Speck. DDoS-UNet: Incorporating temporal information using dynamic dual-channel unet for enhancing super-resolution of dynamic mri. *arXiv preprint arXiv:2202.05355*, 2022.
- [216] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.
- [217] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE*

- conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.
- [218] C Lee Giles, Gary M Kuhn, and Ronald J Williams. Dynamic recurrent neural networks: Theory and applications. *IEEE Transactions on Neural Networks*, 5(2):153–156, 1994.
- [219] Barak A Pearlmutter. Learning state space trajectories in recurrent neural networks. *Neural Computation*, 1(2):263–269, 1989.
- [220] Tony Robinson, Mike Hochberg, and Steve Renals. The use of recurrent neural networks in continuous speech recognition. In *Automatic speech and speaker recognition*, pages 233–258. Springer, 1996.
- [221] Yan Huang, Wei Wang, and Liang Wang. Video super-resolution via bidirectional recurrent convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):1015–1028, 2017.
- [222] Wenhan Yang, Jiashi Feng, Guosen Xie, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Video super-resolution based on spatial-temporal recurrent residual networks. *Computer Vision and Image Understanding*, 168:79–92, 2018.
- [223] Benjamin Naoto Chiche, Arnaud Woiselle, Joana Frontera-Pons, and Jean-Luc Starck. Stable long-term recurrent video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 837–846, 2022.
- [224] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.
- [225] Soumick Chatterjee, Alessandro Sciarra, Max Dünnwald, Steffen Oeltze-Jafra, Andreas Nürnberger, and Oliver Speck. Retrospective motion correction of mr images using prior-assisted deep learning. In *Medical Imaging meets NeurIPS 2020*, 12 2020.
- [226] Jonah Casebeer, Zhepei Wang, and Paris Smaragdis. Multi-view networks for multi-channel audio classification. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 940–944. IEEE, 2019.
- [227] A Emre Kavur, N Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan, et al. Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. *Medical Image Analysis*, 69:101950, 2021.

- [228] Fernando Pérez-García, Rachel Sparks, and Sebastien Ourselin. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, page 106236, 2021.
- [229] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 1(10):e3, 2016.
- [230] D Randall Wilson and Tony R Martinez. The general inefficiency of batch training for gradient descent learning. *Neural networks*, 16(10):1429–1451, 2003.
- [231] Morium Akter Sara, Umme and Mohammad Shorif Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019.
- [232] Chompunuch Sarasaen, Soumick Chatterjee, Mario Bretkopf, Domenico Iuso, Georg Rose, and Oliver Speck. Breathing deformation model-application to multi-resolution abdominal mri. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2769–2772. IEEE, 2019.
- [233] Chompunuch Sarasaen, Soumick Chatterjee, Mario Bretkopf, Domenico Iuso, Georg Rose, and Oliver Speck. Generating breathing deformation model from low resolution 4d mri. In *2019 The 4th Image-Guided Interventions Conference (IGIC)*, 2019.
- [234] Chompunuch Sarasaen, Soumick Chatterjee, Georg Rose, Andreas Nürnberger, and Oliver Speck. Motion-robust dynamic abdominal mri using kt grasp and dynamic dual-channel training of super-resolution u-net (DDoS-UNet). In *Proc. The Joint Annual Meeting ISMRM-ESMRMB & ISMRT 31st Annual Meeting*, 5 2022.

List of Publications

The following publications (as the first author, the co-first author, and the co-author) were published during the course of the study.

Journal Publications

1. **Chompunuch Sarasaen**^a, Soumick Chatterjee^b, Mario Breilkopf, Georg Rose, Andreas Nürnbergger, and Oliver Speck. "Fine-tuning deep learning model parameters for improved super-resolution of dynamic mri with prior-knowledge." *Artificial Intelligence in Medicine* 121, 2021: 102196 ¹.
2. Soumick Chatterjee^b, **Chompunuch Sarasaen**^a, Georg Rose, Andreas Nürnbergger, and Oliver Speck. "DDoS-UNet: Incorporating temporal information using dynamic dual-channel unet for enhancing super-resolution of dynamic mri." *arXiv preprint arXiv:2202.05355*, 2022. (has been submitted to the journal of *Artificial Intelligence in Medicine* and being reviewed).
3. Soumick Chatterjee, Arnab Das, Chirag Mandal, Budhaditya Mukhopadhyay, Manish Vipinraj, Aniruddh Shukla, Rajatha Nagaraja Rao, **Chompunuch Sarasaen**, Oliver Speck, and Andreas Nürnbergger. "TorchEsegeta: Framework for Interpretability and Explainability of Image-Based Deep Learning Models." *Applied Sciences* 12, no. 4, 2022: 1834.
4. Soumick Chatterjee, Mario Breilkopf, **Chompunuch Sarasaen**, Hadya Yassin, Georg Rose, Andreas Nürnbergger, and Oliver Speck. "ReconResNet: Regularised residual learning for MR image reconstruction of Undersampled Cartesian and Radial data." *Computers in Biology and Medicine* 143, 2022: 105321.
5. Soumick Chatterjee, Kartik Prabhu, Mahantesh Pattadkal, Gerda Bortsova, **Chompunuch Sarasaen**, Florian Dubost, Hendrik Mattern, Marleen de Bruijne, Oliver Speck, and Andreas Nürnbergger. "Ds6, deformation-aware semi-supervised learning: Application to small vessel segmentation with noisy training data." *Journal of Imaging*. 2022; 8(10): 259.

¹*a* and *b* have equal contribution

6. Soumick Chatterjee, Fatima Saad, **Chompunuch Sarasaen**, Suhita Ghosh, Rupali Khatun, Petia Radeva, Georg Rose, Sebastian Stober, Oliver Speck, and Andreas Nürnberger. "Exploration of interpretability techniques for deep covid-19 classification using chest x-ray images." arXiv preprint arXiv:2006.02570, 2020.

Conference Papers and Abstracts

1. **Chompunuch Sarasaen**, Soumick Chatterjee, Georg Rose, Andreas Nürnberger, and Oliver Speck. "Motion-robust dynamic abdominal mri using kt grasp and dynamic dual-channel training of super-resolution u-net (DDoS-UNet)." ISMRM 2022.
2. Soumick Chatterjee^b, **Chompunuch Sarasaen**^a, Georg Rose, Andreas Nürnberger, and Oliver Speck. "DDoS-UNet: Incorporating temporal information using dynamic dual-channel unet for enhancing super-resolution of dynamic mri." In Medical Imaging with Deep Learning (MIDL), Zürich, Switzerland, 7 2022.
3. **Chompunuch Sarasaen**, Soumick Chatterjee, Fatima Saad, Mario Breitkopf, Andreas Nürnberger, and Oliver Speck. "Fine-tuning deep learning model parameters for improved super-resolution of dynamic mri with prior-knowledge." In Proc. The Annual Meeting of The International Society for Magnetic Resonance in Medicine (ISMRM), 5 2021.
4. Soumick Chatterjee, **Chompunuch Sarasaen**, Alessandro Sciarra, Mario Breitkopf, Steffen Oeltze-Jafra, Andreas Nürnberger, and Oliver Speck. "Going beyond the image space: undersampled mri reconstruction directly in the k-space using a complex valued residual neural network." In Proc. The Annual Meeting of The International Society for Magnetic Resonance in Medicine (ISMRM), 5 2021.
5. **Chompunuch Sarasaen**, Soumick Chatterjee, Andreas Nürnberger, and Oliver Speck. "DDoS: Dynamic dual-channel u-net for improving deep learning based super-resolution of abdominal dynamic mri." In 38th Annual Scientific Meeting Congress of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), 34(Supplement 1): S6.O3., S44. Springer, 10 2021.
6. Soumick Chatterjee, Mario Breitkopf, **Chompunuch Sarasaen**, Hadya Yassin, Georg Rose, Andreas Nürnberger, and Oliver Speck. "ReconResNet: Regularised residual learning for MR image reconstruction of Undersampled Cartesian and Radial data." In Medical Imaging with Deep Learning (MIDL), Lübeck, Germany, 7 2021.
7. **Chompunuch Sarasaen**, Soumick Chatterjee, Andreas Nürnberger, and Oliver Speck. "Super resolution of dynamic mri using deep learning, enhanced by prior-

- knowledge." In 37th Annual Scientific Meeting Congress of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), 33(Supplement 1): S03.04, S28-S29. Springer, 10 2020.
8. **Chompunuch Sarasaen**, Soumick Chatterjee, Mario Breitkopf, Domenico Iuso, Georg Rose, and Oliver Speck. "Generating breathing deformation model from low resolution 4d mri." In 2019 The 4th Image-Guided Interventions Conference (IGIC), 2019.
 9. Soumick Chatterjee, Mario Breitkopf, **Chompunuch Sarasaen**, Georg Rose, Andreas Nürnberger, and Oliver Speck. "Comparison between the usage of same and different variable density undersampling patterns for deep learning based mri reconstruction." In 2019 The 4th Image-Guided Interventions Conference (IGIC), 2019.
 10. Soumick Chatterjee, Mario Breitkopf, **Chompunuch Sarasaen**, Georg Rose, Andreas Nürnberger, and Oliver Speck. "A deep learning approach for reconstruction of undersampled cartesian and radial data." In 36th Annual Scientific Meeting Congress of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), 10 2019.
 11. **Chompunuch Sarasaen**, Soumick Chatterjee, Mario Breitkopf, Domenico Iuso, Georg Rose, and Oliver Speck. "Breathing deformation model-application to multi-resolution abdominal mri." In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 2769–2772. IEEE, 2019.

Declaration of Honor

I hereby declare that I produced this thesis without prohibited external assistance and that none other than the listed references and tools have been used. I did not make use of any commercial consultant concerning graduation. A third party did not receive any nonmonetary perquisites neither directly nor indirectly for activities which are connected with the contents of the presented thesis.

All sources of information are clearly marked, including my own publications.

In particular I have not consciously:

- Fabricated data or rejected undesired results
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data
- Plagiarized data or publications
- Presented the results of other researchers in a distorted way

I do know that violations of copyright may lead to injunction and damage claims of the author and also to prosecution by the law enforcement authorities. I hereby agree that the thesis may need to be reviewed with an electronic data processing for plagiarism.

This work has not yet been submitted as a doctoral thesis in the same or a similar form in Germany or in any other country. It has not yet been published as a whole.

Magdeburg, Januar 5, 2024

Chompunuch Sarasaen