Towards One-Minute 3D Multi-Parametric Quantitative Brain MRI using Partial Diffusion Models

MSc Thesis Biomedical Engineering

Tessel Huibregtsen

TU Delft

BIGR
Biomedical Imaging Group Rotterdam

Towards One-Minute 3D Multi-Parametric Quantitative Brain MRI using Partial Diffusion Models

by

Tessel Huibregtsen

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Wednesday July 16, 2025 at 10:45 AM.

Student number:	4876377	
Project duration:	November 18, 2024 – July 16, 2025	
Thesis committee:	Prof. J.A. Hernandez-Tamames	TU Delft, Erasmus MC
	Asst. Prof. D.H.J. Poot	Erasmus MC
	Asst. Prof. J.F. Veenland	Erasmus MC
	Asst. Prof. Q. Tao	TU Delft
	S. Wang, MSc	Erasmus MC

Cover: An image of the NGC 6888, also known as the Crescent or Brain Nebula.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

BIGR Biomedical Imaging Group Rotterdam

Preface

This thesis marks the final step towards my Master's degree in Biomedical Engineering at the TU Delft. The research was carried out at the Biomedical Imaging Group Rotterdam at the Erasmus Medical Center at the intersection of medicine, physics and deep learning.

The goal of this project was to enhance undersampled MR images acquired with an accelerated 3D multiparametric quantitative MRI (qMRI) sequence, using partial diffusion models. The motivation behind this work was to enable a 1.5-minute brain scan capable of providing radiologist with all the information that currently requires 20 to 30 minutes of scanning.

The past eight months have been both challenging and exciting, and I would like to express my gratitude to everyone who supported me during this time. To the BIGR group as a whole, thank you for welcoming me into your research group and inspiring me with all the amazing work you do. To the qMRI research line in particular, thank you for all your help into understanding the foundations of qMRI and the potential it holds for the future of MR. Then, to my daily supervisor, Shishuai, thank you for helping me navigate the project and for always encouraging me to further improve my work. Your support, both technically and personally, was invaluable. I also want to thank Dirk, whose technical insights and willingness to help out at any moment made tackling challenges much more manageable. And to Juan, thank you for your visionary enthusiasm about quantitative MRI and for reminding me of the broader impact of this work.

I am also very thankful to my family and friends, who, even though I wasn't always able to clearly explain what I was working on, kept asking, listening and supporting me. And to Max, thank you for your constant support, helping me think through problems, and for keeping me focused on the storyline of my thesis.

A special mention goes to Rob, whose beautiful image of the Crescent Nebula is on the cover of this thesis. This nebula, also known as the Brain Nebula, is invisible to the human eye but becomes visible by long-exposure astrophotography and advanced image-processing. Its resemblance to the human brain and the need for advanced post-processing to reveal it, make it a fitting metaphor for this thesis. In MRI, the raw signal is not directly visible or interpretable either. Only through advanced reconstruction techniques, such as partial diffusion models, we can recover meaningful, high-quality images from the undersampled, noisy data.

Working on this thesis has been a rewarding experience. I hope this work contributes, in some small way, to making quantitative MRI more accessible in clinical practice.

Tessel Huibregtsen
Delft, July 2025

Summary

Quantitative Magnetic Resonance Imaging (qMRI) aims to provide objective, reproducible measurements of tissue properties that could support clinical decision-making. However, its clinical adoption is hindered by the long scan times required for the acquisition of multiple contrast-weighted images associated with qMRI. The 3D silent Multi-Parametric Zero Echo Time (MuPa-ZTE) sequence already reduces scan time to approximately 5.5 minutes, but this remains too long to be added on top of routine 15–20 minutes clinical brain scans during the validation phase. Therefore, further acceleration to around 1.5 minutes is required. This is achieved by undersampling the 5.5-minute acquisition, which significantly reduces image quality.

In this thesis, we investigate the use of partial diffusion models, a class of generative models based on deep learning, to enhance these undersampled MR images acquired with an accelerated MuPa-ZTE sequence. The goal is to approach the image and quantitative map quality of the current 5.5-minute MuPa-ZTE acquisition using a 1.5-minute scan.

To this end, three variations of a partial denoising diffusion probabilistic model (DDPM) were trained on synthetic data simulating MuPa-ZTE acquisitions. Two types of synthetic data were generated for training: reference brain images depicting the ground truth contrasts without acquisition artifacts and undersampled brain images that simulate the 1.5-minute acquisition. The models aimed to improve the quality of the five contrast-weighted images associated with MuPa-ZTE, which were then used to generate proton density (PD), T_1, and T_2 maps through dictionary matching. The first model was trained only on the ground truth images, without access to undersampled images, the second model incorporated undersampled images indirectly through a learned latent representation and the third model had direct access to both image types during training, using a concatenation-based approach. All models were evaluated both on a synthetic test set and on in vivo data. Their outputs were qualitatively and quantitatively compared to synthetic 5.5-minute acquisition images to assess whether comparable image quality could be achieved with reduced acquisition time.

Results showed that all three partial DDPM variations improved the image quality of the 1.5-minute scan, with the concatenation-based model achieving the best overall performance. This model demonstrated effective noise suppression and improved preservation of structural detail in the contrast-weighted images, outperforming a conventional UNet baseline and approaching the quality of the 5.5-minute acquisition. The enhanced images directly translated into improved quantitative maps, with the quality of the T_1 and T_2 maps in some cases even exceeding those derived from the 5.5-minute acquisition.

These findings suggest that partial diffusion models can play an important role in accelerating qMRI with minimal decrease in image and map quality, bringing one-minute quantitative imaging closer to clinical feasibility.

Contents

1

Introduction

Magnetic Resonance Imaging (MRI) is a powerful, non-invasive imaging modality that provides rich structural or functional information. It is widely used for diagnostic purposes due to its ability to provide high-resolution, high-contrast images without ionizing radiation. Conventional MRI techniques are primarily used to produce contrast-weighted images for visual assessment. However, the pixel intensities in these images are typically qualitative and do not directly reflect underlying physical tissue properties [1].

In contrast, quantitative MRI (qMRI) aims to estimate intrinsic tissue parameters, such as relaxation times and proton density. These estimated tissue parameters are used to construct quantitative maps, which reflect the underlying tissue properties. This approach provides several unique advantages. The generated quantitative maps enable standardized comparison of data across different imaging conditions and patients [2], and the objective quantitative parameters can serve as biomarkers for specific diseases, potentially improving disease detection and monitoring [3]. Additionally, any type of contrast-weighted image can be synthesized from the generated quantitative maps, allowing radiologists to generate images optimized for specific diagnostic needs, without requiring extra scans.

Despite its advantages, qMRI typically requires the acquisition of multiple contrast-weighted images under systematically varied acquisition conditions, resulting in long scan times. This hinders clinical implementation. Reducing scan time while maintaining high-quality quantitative maps remains a significant challenge. A prominent trend in qMRI driven by this challenge is the development of fast 3D multi-parametric mapping techniques, which aim to efficiently estimate multiple intrinsic tissue parameters (such as T_1, T_2, and proton density) in a single scan [4]. Several innovative protocols have been proposed to achieve this, including Magnetic Resonance Fingerprinting (MRF) [5], 3D-QALAS [6], and Quantitative Transient-state Imaging (QTI) [7]. These techniques acquire a time series of contrast-weighted images under varying acquisition conditions, thereby encoding information about multiple tissue parameters in the signal evolution. Quantitative maps are then reconstructed from this signal evolution using predefined signal models.

Zero Echo Time (ZTE) imaging has been applied to 3D multi-parametric mapping in the 3D silent Multi-Parametric Zero TE (MuPa-ZTE) sequence, due to its unique advantages [8, 9, 10]. ZTE allows for imaging of tissues with very short T_2 relaxation times and its radial k-space acquisition scheme allows for further scan acceleration through undersampling. In addition to its speed, the MuPa-ZTE sequence offers the additional advantages of low acoustic noise due to limited gradient switching and motion robustness as it avoids phase accumulation by initiating readout immediately after excitation [8]. Currently, the MuPa-ZTE sequence requires approximately 5.5 minutes to acquire high-quality whole-brain quantitative maps. However, since MuPa-ZTE is a novel imaging protocol, it must first be validated alongside conventional brain MRI protocols, which typically require 15-20 minutes for equivalent multi-contrast information. During this validation phase, adding a 5.5-minute MuPa-ZTE acquisition on top of routine clinical scans is impractical. Therefore, to facilitate clinical feasibility, the MuPa-ZTE scan time needs to be reduced further, to approximately 1.5 minutes. Achieving this requires accelerating

the MuPa-ZTE acquisition by at least a factor of four. However, increasing the level of acceleration inherently leads to more significant undersampling, which introduces image artifacts, degraded contrast, and ultimately lower quality of the derived quantitative maps.

Recent developments in deep learning have demonstrated strong abilities in enhancing medical image quality. Convolutional neural networks, transformer-based architectures, and generative models have all been explored for various tasks such as image denoising, super-resolution, and artifact reduction in MRI [11, 12, 13]. Among these approaches, generative models, and especially diffusion models, have emerged as state-of-the-art for image synthesis and restoration [14]. Diffusion models work by modeling the data distribution through a learned denoising process where data is gradually corrupted with noise and then reconstructed through a series of iterative denoising steps. This enables the model to generate high-quality images from pure noise. A specific type of diffusion models, known as partial diffusion models [15], leverages the generative power of the diffusion model but only uses part of the forward and reverse diffusion process. By performing denoising over a limited number of steps, greater control over the generation process is maintained and computational complexity is reduced.

In this thesis, the use of partial diffusion models is investigated to enhance contrast-weighted images acquired with a four times accelerated MuPa-ZTE acquisition sequence. The goal is to improve the quality of this accelerated acquisition, such that the derived quantitative maps are of comparable quality to those obtained from the current state-of-the-art MuPa-ZTE acquisition, lasting approximately 5.5 minutes.

1.1. Research Questions

The primary objective of this thesis is to evaluate the use of partial diffusion models for improving the quality of contrast-weighted MR images acquired with an accelerated MuPa-ZTE sequence and to assess whether these enhanced images lead to more accurate quantitative maps. The research questions that will be answered specifically:

1. Can partial diffusion models improve the quality of contrast-weighted images acquired with a four times accelerated MuPa-ZTE sequence to reach the quality of images from a non-accelerated acquisition?
2. How do different architectural variations of the partial diffusion model affect the final image and map quality?
3. How does a partial diffusion-based approach compare to a conventional UNet baseline in this setting?
4. Can the proposed method lead to improved quality of quantitative maps without any further adjustments?

1.2. Thesis Structure

This thesis is organized into seven chapters following the standard academic reporting structure. Chapter 2 provides the theoretical background required for the thesis and outlines the current state-of-the-art. Chapter 3 explains the methodology, including the synthetic data generation, the model architecture, the training strategies, and the evaluation metrics. Chapter 4 outlines the experimental setup used to evaluate the different research questions. Chapter 5 presents the experimental results obtained, which are critically analyzed and discussed in Chapter 6. Chapter 7 concludes the thesis with a summary of the main contributions and answers to the research questions.

2

Theoretical Background

2.1. MRI principles

Magnetic Resonance Imaging (MRI) is based on the magnetic properties of atomic nuclei, most commonly those of hydrogen protons, due to their abundance in the human body. When placed in a strong external magnetic field (B_0), the magnetic moments of these hydrogen nuclei align either parallel or anti-parallel to the magnetic field, resulting in a net magnetization (M_0) along the longitudinal axis, defined along the direction of B_0. To generate a signal, a radio frequency (RF) pulse is applied at the same frequency at which the magnetic moments of the hydrogen protons precess around the external magnetic field, known as the Larmor frequency, ω_0. The Larmor frequency is determined by the gyromagnetic ratio of a proton, γ, and B_0:

$$\omega_0 = \gamma \times B_0 \tag{2.1}$$

This RF pulse flips the direction of M_0 away from alignment with the external magnetic field. Once the RF pulse is turned off, M_0 returns to equilibrium through two processes: longitudinal relaxation, where the longitudinal component of the magnetization vector, M_z, recovers to its equilibrium value, and transverse relaxation, where the transverse components, M_x and M_y, decay due to dephasing of the spins in the transverse plane. These dynamics are governed by the Bloch equations which describe the evolution of the magnetization vector $M(t)$. At a time t after an instantaneous $\pi/2$ RF pulse (that is, a 90° excitation pulse that tips magnetization into the transverse plane), the solution to the Bloch equations are:

$$M_x(t) = M_0 e^{-t/T_2^*} \sin(\omega t) \tag{2.2}$$

$$M_y(t) = M_0 e^{-t/T_2^*} \cos(\omega t) \tag{2.3}$$

$$M_z(t) = M_0(1 - e^{-t/T_1}) \tag{2.4}$$

Where M_0 is the equilibrium magnetization, ω the Larmor frequency, T_2^* the observed transverse relaxation time, and T_1 the longitudinal relaxation time. T_2^* reflects both T_2, the intrinsic spin-spin relaxation time of a tissue, and additional dephasing effects caused by magnetic field inhomogeneities. In an idealized situation, where the magnetic field is perfectly homogeneous, T_2^* is equal to T_2. T_1 and T_2 are tissue-specific and therefore determine the dynamics of the relaxation process. This process leads to the emission of electromagnetic signals, which are detected by the MRI receiver coils. To enable spatial localization of the electromagnetic signals, gradient magnetic fields are applied. These gradients introduce a linearly varying magnetic field strength, which causes the Larmor frequencies of the spins to vary linearly with position. This position-dependent variation of Larmor frequencies allows for the selective excitation of specific slices by applying RF pulses at corresponding frequencies, thereby encoding the signals spatially.

2.1.1. Contrast-Weighted Images

To generate images with specific contrast properties, MRI relies on pulse sequences: carefully timed combinations of RF pulses, gradient fields, and readouts. By varying the timing of these parameters, MRI can exploit the differences in tissue-specific relaxation times to enhance contrast between different types of tissues.

For conventional MRI, two key parameters that influence image contrast are the repetition time (TR) and the echo time (TE). TR is the time interval between two successive excitation pulses, while TE is the time between the excitation pulse and the signal readout. The choice of TR affects how much T_1-based recovery occurs between excitations: a long TR allows for full T_1 recovery and thus reduced T_1 contrast, whereas a short TR captures differences in how quickly tissues recover, enhancing T_1 contrast. In contrast, the choice of TE determines how much transverse relaxation occurs before signal readout: a long TE allows for more pronounced differences in transverse decay, enhancing T_2 contrast, while a short TE minimizes transverse decay by acquiring the signal before substantial decay occurs.

Three commonly used contrast types in clinical MRI are T_1-weighted, T_2-weighted, and proton density (PD)-weighted images. T_1-weighted images highlight differences in longitudinal relaxation time (T_1) and are thus acquired using short TR, to enhance T_1 contrast and short TE, to minimize T_2 contrast. Tissues with short T_1, such as fat, appear bright, while fluids with a long T_1, appear dark. These images are typically used for anatomical detail and post-contrast imaging [16]. T_2-weighted images focus on the transverse relaxation time, T_2, by using long TR and long TE. Tissues with a long T_2, such as fluids, appear bright which makes these images useful for detecting pathologies involving increased fluid content, such as inflammation [17]. Lastly, PD weighted images highlight the differences in proton density by using a long TR (to minimize T_1 effects) and a short TE (to minimize T_2 effects). These images are mainly used for anatomical detail. By carefully selecting the imaging parameters, MRI can be tailored to specific tissue characteristics, making it a powerful tool for a wide range of applications.

2.2. Quantitative MRI

As explained in the previous section, conventional MR pulse sequences are designed to encode different qualitative contrasts in the MR images by varying scanning settings such as the timing of the RF pulses. In contrast, quantitative MR pulse sequences are designed such that the recorded signals can be used to infer quantitative information about the physical properties of the underlying tissue, such as the longitudinal relaxation time (T_1), the transverse relaxation time (T_2), and the proton density (PD). To do so, qMRI protocols acquire multiple images under varying acquisition conditions, which are then fitted to a signal model that links the quantitative parameters and acquired signals.

The Bloch equations describe the evolution of the MR signal as a function of time and tissue properties and can therefore be used to predict the signal from a specific tissue type when subjected to a known pulse sequence. For example, in conventional quantitative T_1 mapping, several images are acquired with varying inversion recovery times, the time between an inversion pulse and the image acquisition, to capture different stages of longitudinal recovery. The resulting signal intensities are fitted voxel-wise to an exponential recovery curve based on the Bloch equations, to extract the T_1 value for each tissue location. Next to single parameter mapping, other sequences have also been developed that can map multiple quantitative parameters at once, known as multi-parametric sequences. By combining the variation of scanning parameters, multiple quantitative parameters can be estimated at once.

The advantage of qMRI lies in its ability to provide reproducible and comparable measurements across scanners, patients, and time points. However, one important challenge is that qMRI requires the acquisition of multiple images to be able to fit the signal evolution to the model, leading to prolonged scan times.

2.2.1. Accelerated qMRI

Multiple methods have been developed to overcome the long scan times typically associated with qMRI. A substantial class of methods to accelerate qMRI focuses on reconstruction-based approaches. As this thesis focuses on improving the weighted images associated with quantitative MRI, we will only elaborate on methods that use this intermediate step in the reconstruction pipeline. Methods that go directly from k-space to quantitative maps are out of scope.

Reconstruction-based methods aim to enhance image quality from undersampled data after acquisition. This type of method can be broadly categorized into two categories: model-based reconstruction methods and deep learning-based reconstruction methods. In model-based approaches, prior information about the signal is used to be able to reconstruct high-quality images from undersampled k-measurements. One widely used model-based approach is compressed sensing (CS), which exploits the inherent sparsity of MR images when transformed in different domains such as the wavelet or Fourier domain [18]. By transforming the data to these domains, the image can be undersampled without losing critical information. Similarly, low-rank and subspace reconstruction methods leverage redundant information in the spatiotemporal domains of the data to enable accelerated acquisitions [19, 20]. More recently, deep learning-based methods have been developed for accelerated qMRI reconstruction. Instead of modeling the reconstruction with physics, deep learning-based methods aim to train a deep neural network to do this reconstruction. Most current methods combine deep learning-based methods with model-based methods by for example integrating deep learning-based priors into traditional iterative reconstruction frameworks [21], or by incorporating MR signal models into the training process to ensure consistency with known physical dynamics [22].

2.3. Diffusion Model

Diffusion models, a class of generative models, have shown promising results in tasks such as image synthesis and image restoration [14]. Particularly, in this thesis, a diffusion model is used to improve the image quality of undersampled MR images acquired with an accelerated MuPa-ZTE sequence.

Diffusion models work by first gradually transforming an image into pure Gaussian noise through a controlled, step-by-step noising process. During training, the model learns to reverse this process by predicting the noise added at each timestep. Once trained, the model can iteratively denoise a Gaussian input, effectively generating new samples that resemble the original image distribution.

A widely used subclass of diffusion models is the Denoising Diffusion Probabilistic Model (DDPM), introduced by Ho et al. [23]. DDPMs model the corruption process as a Markov chain, where the next state (the next noisier image) only depends on the previous state (the previous image).

2.3.1. Partial Diffusion Model

A major drawback of diffusion models is the computational cost associated with training and inference. These models require a large number of diffusion steps, typically in the range of 1000, to gradually transform data into pure Gaussian noise and then reconstruct it. Additionally, since inference typically starts from randomly sampled noise, there is limited control over the specific content of the output. However, for the aim of this thesis, improving the quality of accelerated weighted MR images, it is important to preserve and use information from the input image. Rather than generating an entirely new image from noise, the goal is to refine an existing image while maintaining its structural content. This motivates the use of partial diffusion models, as introduced by Zhao et al. [15].

Partial diffusion models rely on the same principles as standard DDPMs, but instead of diffusing the input all the way to pure noise, the forward process is only run for a partial number of steps. This means the input image is only partially noised, preserving much of its original structure. The reverse process begins from this intermediate state, which gives the model access to existing image features, while still benefiting from the flexibility and generative power of diffusion models.

Partial diffusion models have proven to be able to enhance the quality of qualitative contrast-weighted MR images [24]. This thesis will explore their use in multi-channel contrast-weighted images acquired by MuPa-ZTE and the influence on the quality of the quantitative maps generated from them.

3

Methods

This chapter describes the methodological steps taken to evaluate the use of partial diffusion models in enhancing undersampled MuPa-ZTE MRI acquisitions. The goal is to train and evaluate a model capable of producing high-quality contrast-weighted images from four times accelerated scans. First, due to the limited availability of clinical data for the MuPa-ZTE sequence, synthetic data were generated. Next, a partial diffusion model was developed to enhance the undersampled contrast-weighted MR images. Multiple architecture variations of this model were developed to explore their influence on the performance. The enhanced weighted images were subsequently used to construct quantitative maps using dictionary matching. Afterwards, the model performances were evaluated both qualitatively, by visual inspection, and quantitatively, by computing various image quality metrics. This was done for both the improved contrast-weighted images and the quantitative maps derived from them. Each component of this methodology will be described in more detail in this chapter.

3.1. Synthetic Data Generation

Obtaining sufficient clinical data to train deep neural networks remains a significant challenge in qMRI. This challenge stems from the complexity of adjusting MR sequences, the limited availability of volunteers, and the high cost and time associated with data acquisition. These constraints make it difficult to acquire a large, diverse training set, necessary for effectively training a deep neural network. To address this challenge, this work used a pipeline based on the work by Sabidussi et al. [25]. The pipeline was used to generate synthetic data that closely resembles that of MuPa-ZTE acquisitions. This approach allows for a virtually unlimited amount of training data to be generated, with precise control over acquisition settings and tissue parameters, thereby facilitating the creation of a large and diverse training set.

The synthetic data generation pipeline used consists of two main parts:

1. The generation of synthetic quantitative maps, representing the underlying tissue properties (PD, T_1 and T_2) of each voxel in a brain volume.
2. The simulation of MuPa-ZTE acquisitions using a forward model to generate the five contrast-weighted images.

3.1.1. Synthetic Quantitative Map Generation

The synthetic data generation pipeline is based on the assumption that the response of biological tissue to a given MRI pulse sequence can be modeled when the underlying tissue properties are well-defined. Therefore, the synthetic data generation starts by synthesizing realistic quantitative maps. To this end, segmented brain models from the BrainWeb database [26] are used, which provide voxel-wise tissue type labels. For each tissue class, the pipeline assigns mean PD, T_1, and T_2 values according to literature-based parameters (refer to Table 3.1) [25]. To simulate biological variability, the mean parameter value for each tissue type is sampled from a Gaussian distribution: $N(\mu_t, \sigma_p^2)$, where μ_t is the literature-based mean for tissue type t, and σ_p is the parameter-specific standard deviation. The

standard deviations are set as follows: 0.1 arbitrary units (a.u.) for PD, 0.3 seconds for T_1, and 0.01 seconds for T_2. Additionally, to model natural heterogeneity within tissue, voxel-wise noise is introduced, followed by Gaussian smoothing to achieve smooth transitions between neighboring voxels. The noise level to model natural heterogeneity within tissue is set to 0.05 a.u. for PD, 0.1 seconds for T_1, and 0.01 seconds for T_2.

Due to this introduced variety in the quantitative maps, multiple unique synthetic subjects can be generated from the same underlying brain model. For training purposes, we generated four synthetic subjects from each of 19 brain models. All brain models were resized to $181 \times 217 \times 181$ voxels. One additional brain model was used exclusively for testing, from which 10 synthetic subjects were generated to create a diverse evaluation set. In addition, we acquired two in vivo MRI datasets with the MuPa-ZTE sequence from two healthy volunteers for evaluation purposes. One in vivo dataset was only used for qualitative evaluation (see Appendix A), whereas the other in vivo dataset was used for both qualitative and quantitative evaluation.

Table 3.1: Mean Proton Densities and Relaxation Times of Various Tissues [25]

	PD (a.u.)	$\mathbf{T_1}$ (s)	$\mathbf{T_2}$ (s)
CSF	1.0	4.0	2.0
Gray Matter	0.90	1.4	0.11
White Matter	0.80	0.78	0.08
Fat	0.92	0.42	0.07
Muscle	0.85	1.3	0.05
Muscle/Skin	0.82	1.23	0.05
Skull	0.17	0.4	0.03
Vessels	0.95	1.98	0.275
Around Fat	0.40	0.9	0.08
Dura Matter	0.50	0.9	0.07
Bone Marrow	0.75	0.58	0.05

3.1.2. 3D Silent Multi-Parametric Zero TE MRI

For this thesis, we focused on accelerating the 3D silent Multi-Parametric Zero TE (MuPa-ZTE) sequence [8, 9, 10]. This sequence acquires five contrast-weighted images by using an interleaved acquisition scheme along a phyllotaxis-based k-space trajectory. Phyllotaxis is a radial sampling pattern that provides uniform coverage and allows flexible undersampling as spokes can easily be removed or added [27]. The sequence starts by acquiring a PD-weighted image, depicted in segment 1 in Figure 3.1. Next, in an interleaving manner, it acquires the remaining four contrast-weighted images with the use of magnetization preparation sequences that encode the desired contrast in the images. The full phyllotaxis-based k-space trajectory is split into cycles, and for each cycle, the same k-space segment is sampled for all four contrast-weighted images, also known as echoes. The sequence then repeats with a different k-space segment, until the entire k-space is covered for each echo. The first two cycles sample the center of k-space, with short-radius spokes, whereas the rest of the cycles sample the outer region of k-space, with long-radius spokes.

3.1.3. Forward Model and Synthetic Acquisition

To generate contrast-weighted images that realistically represent MuPa-ZTE acquisitions, a physics-based forward model was used that expresses how tissue responds to the MuPa-ZTE pulse sequence, based on its underlying tissue properties. The forward model used in this work takes the quantitative maps generated from the BrainWeb brain models as input and outputs the five contrast-weighted images, or echoes, generated in the MuPa-ZTE acquisition. For each voxel, the evolution of the magnetization over time is simulated through a sequence of preparation modules, ramp-up periods, and readouts, repeated over multiple iterations until steady state is reached.

The first echo, the PD-weighted contrast image, is simulated using a simplified acquisition model which

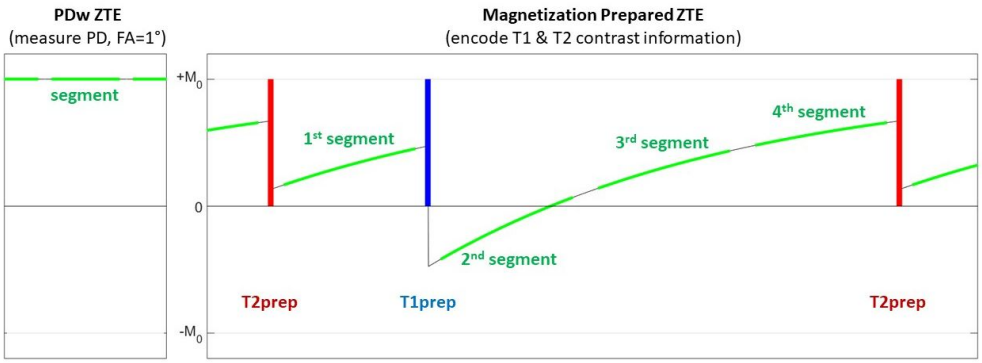

Figure 3.1: Magnetization evolution during the MuPa-ZTE sequence. Adapted from Wiesinger et al. [10]. First, a PD-weighted image is acquired by sampling the entire phyllotaxis-based k-space trajectory. Next, the remaining four contrast-weighted images, or echoes, are acquired using magnetization preparation modules (T2prep and T1prep) to encode T_2 and T_1 contrast, respectively. The k-space trajectory is split into cycles, and in each cycle, the same k-space segment is sampled for all four echoes. The sequence is repeated across cycles until the entire k-space is covered for each echo.

assumes immediate readout following excitation. This is appropriate for a Zero Echo Time (ZTE) acquisition like MuPa-ZTE, where readout occurs with minimal delay after excitation. The longitudinal magnetization, M_z, at equilibrium is computed according to the steady-state solution of the spoiled gradient (SPGR) signal equation, using the known PD and T_1 values from the generated quantitative maps [8]:

$$M_z = PD \cdot \frac{1 - e^{-\mathrm{TR}/T_1}}{1 - e^{-\mathrm{TR}/T_1} \cdot \cos(\mathrm{FA}_0)} \tag{3.1}$$

where TR is the repetition time, the time between two readouts, and FA_0 the flip angle, the angle at which the magnetization vector is tipped away from equilibrium by the RF pulse, used for Echo 0.

For the remaining four echoes (Echoes 2-5), the longitudinal magnetization of the i^{th} readout within a segment is modeled as [8]:

$$M_z^i = M_{\mathrm{prep}} \cdot \beta^i + M_{\mathrm{SPGR}} \cdot (1 - \beta^i) \tag{3.2}$$

where M_{prep} is the longitudinal magnetization immediately after the preparation module, and M_{SPGR} the steady-state SPGR signal given by:

$$M_{\mathrm{SPGR}} = \frac{1 - e^{-\mathrm{TR}/T_1}}{1 - e^{-\mathrm{TR}/T_1} \cdot \cos(\mathrm{FA})} \tag{3.3}$$

and β given by:

$$\beta = e^{-\mathrm{TR}/T_1} \cdot \cos(\mathrm{FA}) \tag{3.4}$$

The acquisition process is modeled by simulating the evolution of magnetization over multiple readouts and mapping this signal into k-space using a Non-Uniform Fast Fourier Transform (NUFFT). Specifically, for each readout, the longitudinal magnetization M_z^i is transformed into the transverse plane by applying the flip angle:

$$M_{xy}^i = M_z^i \cdot \sin(\mathrm{FA}) \tag{3.5}$$

This transverse magnetization is then transformed into k-space using the forward NUFFT. This process is repeated for each readout. The resulting k-space signals are accumulated across all readouts and subsequently reconstructed into an image using the adjoint NUFFT operator. In the provided pipeline, both the forward and adjoint NUFFT operations were implemented using the torchkbnufft library in Python [28]. Finally, the image is converted to a real-valued image by keeping the real part of the complex-valued image. In the reconstruction, k-space density compensation and simple sensitivity maps are also included to correct for non-uniform sampling.

The k-space density compensation is needed due to the radial sampling pattern in MuPa-ZTE, which leads to non-uniform sampling of k-space. With this pattern, the center of k-space is sampled more densely, and without correction, this leads to image artifacts like blurriness. The correction is done by introducing a density compensation factor, $\mathrm{dcomp}_{\mathrm{radial}}$, for each sample, which reflects the relative sampling density in the region of that sample. The density compensation factor for a sample at position $k(x, y, z)$ in a 3D radial sampling pattern is given by:

$$\mathrm{dcomp}_{\mathrm{radial}} = |k|^2 \tag{3.6}$$

However, in MuPa-ZTE, two types of spokes are used to cover different ranges in k-space: short-radius spokes and long-radius spokes. The first two cycles are used for the short-radius spokes, while the other 124 cycles are acquired with long-radius spokes, and therefore the outer region of k-space is more densely sampled than the center of k-space, when only using $\mathrm{dcomp}_{\mathrm{radial}}$. To account for this, an extra weighting factor is introduced based on the ratio between long-radius and short-radius spokes and their respective lengths:

$$w_s = \frac{N_l}{N_s} \cdot \frac{r_s}{r_l} \cdot \frac{n_l}{n_s} \tag{3.7}$$

with N_l and N_s, the total number of long-radius and short-radius spokes in the acquisition, r_l and r_s the radii of the long-radius and short-radius spokes, and n_l and n_s the number of samples on the long-radius and short-radius spokes, respectively. The total compensation factor for k-space samples on the short-radius spokes is then:

$$\mathrm{dcomp}_s = \mathrm{dcomp}_{\mathrm{radial}} \cdot w_s \tag{3.8}$$

Sensitivity maps, on the other hand, are essential in the reconstruction of multi-coil MR acquisitions. Due to the different locations of each receiver coil relative to the subject, coil sensitivity varies spatially, being more sensitive to nearby tissues and less sensitive to distant tissues. During reconstruction, each coil's density-compensated k-space data is first transformed to image space using the adjoint NUFFT. The resulting image from each coil is then multiplied by its corresponding sensitivity map and subsequently, these images are summed across all coils to produce the final reconstructed image. This ensures that regions far away from a coil are suppressed in the final image and regions closer to the coil are emphasized. In our implementation, a simplified model was used based on six coils. The created sensitivity maps are smooth Gaussian profiles centered around each coil, with the signal strength gradually decreasing with distance from the coil.

Together, the use of k-space density compensation and sensitivity maps ensures accurate image reconstruction despite the non-uniform sampling pattern and spatially varying coil sensitivities inherent to the acquisition process.

3.1.4. Long-Acquisition, Short-Acquisition, and Reference Images

To train and evaluate a model on its ability to enhance the quality of accelerated MR images, we require both the contrast-weighted images from the standard, long MuPa-ZTE acquisition, and their corresponding short-acquisition counterparts. The long-acquisition images are generated using the forward model as described in Subsection 3.1.3 with the complete MuPa-ZTE k-space trajectory consisting of 126 cycles. This synthetic acquisition simulates a real scan that would take 5 minutes and 37 seconds on the GE Healthcare Signa MR750 3.0T scanner. The additional weighting factor applied for density compensation for this acquisition, w_{s}, was computed to be 7.7483.

To simulate the accelerated imaging, we generate the short-acquisition images, by truncating the k-space trajectory as well as the k-space data used to reconstruct the weighted images. The MuPa-ZTE k-space trajectory starts with two cycles of short-radius spokes which densely sample the center of k-space. This is followed by a repeating sequence of four structurally distinct long-radius spoke cycles, each designed to sample different regions of the outer k-space. To preserve this structure, we truncated the trajectory after the first 34 ($= 2 + 4 \cdot 8$) cycles, which corresponds to approximately 27% of the full acquisition and would require around 1.5 minutes to acquire.

Since truncating the k-space trajectory changes the ratio between long-radius and short-radius spokes, the additional weighting factor for density compensation needs to be updated. The updated additional weighting factor, w_{s}, is computed as $7.7483 \cdot \frac{32}{124} = 1.9996$. Here, 32 and 124 represent the number

of long-radius spokes in the short and long acquisitions, respectively. Throughout this thesis, the 5.5-minute acquisition will be referred to as the long acquisition, and the 1.5-minute acquisition as the short acquisition.

The benefit of synthetic data generation is that we also have access to artifact-free images that can serve as idealized references. In this work, these reference contrast-weighted images are generated with the same forward model (see Subsection 3.1.3), but without the transformation to k-space and reconstruction. Instead, the magnetization is directly mapped to image space, without simulating acquisition and reconstruction effects, providing a clean reference. Figure 3.2 shows a representative slice from the reference 3D contrast-weighted volumes, alongside corresponding slices from the 3D volumes acquired using both the long-acquisition and short-acquisition methods.

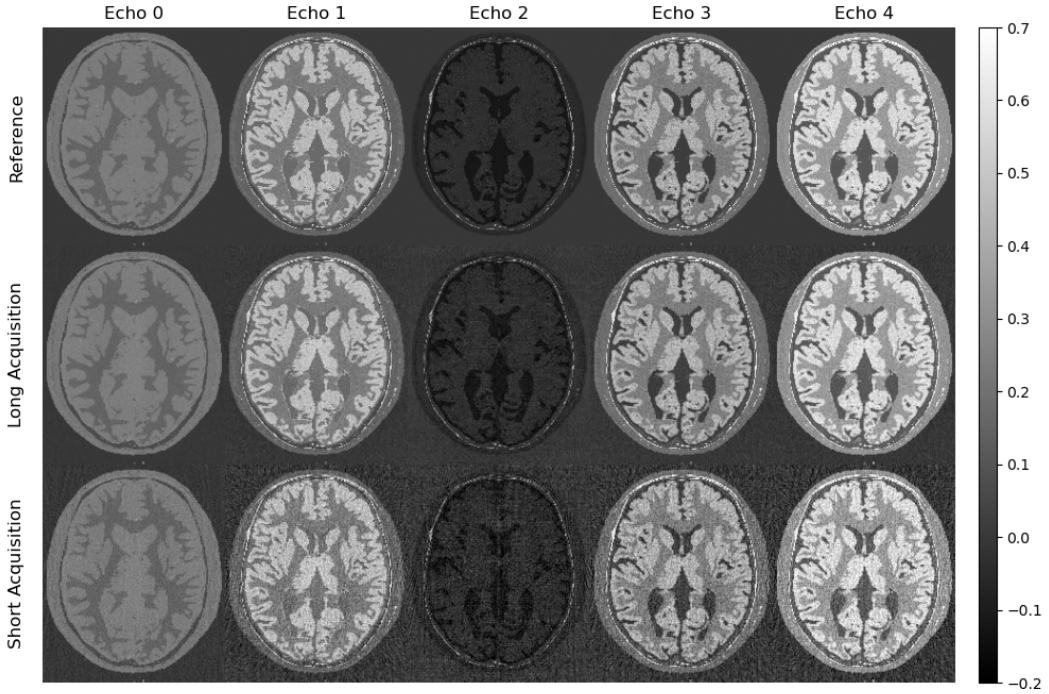

Figure 3.2: One slice of a 3D synthetically generated subject for the MuPa-ZTE acquisition. The top row shows the reference contrast-weighted images, the middle row the long-acquisition contrast-weighted images and the bottom row the short-acquisition contrast-weighted images. The columns show the different echoes.

3.2. Partial Diffusion Model Implementation

To enhance the quality of the short-acquisition images, a partial diffusion model was implemented. In this section, the architecture of the partial diffusion model will be outlined, including the variations researched in this thesis.

The specific type of diffusion model used in this thesis, the DDPM, formalizes the diffusion process as a sequence of intermediate images, where each image is a slightly noisier version of the previous one. At each timestep t, Gaussian noise is added according to a predefined variance schedule: $\beta_1, ..., \beta_T$, with T the total number of diffusion steps. Instead of simulating this noise step-by-step, DDPMs allow for direct sampling of x_t from the original clean image x_0 at any timestep t:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon \tag{3.9}$$

Here, $\epsilon \sim \mathcal{N}(0, I)$ is standard Gaussian noise, and $\bar{\alpha}$ is the cumulative product of the noise scaling terms up to step t:

$$\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s \quad \text{where} \quad \alpha_t = 1 - \beta_t \tag{3.10}$$

This formula can be used to efficiently generate noisy inputs for training. During training, the loss function is defined as the squared Euclidean norm of the predicted noise, $\epsilon_\theta(x_t, t)$, across the entire 3D volume, as described in [23]:

$$\mathcal{L} = \mathbb{E}_{t,x_0,\epsilon}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2\right] \tag{3.11}$$

Therefore, the model learns to predict the noise added at each timestep.

During inference, the short-acquisition image is passed through the forward diffusion process for a predetermined number of timesteps, K, to generate the noisy input x_K. The model then performs denoising by iteratively estimating and removing the noise from x_t at each timestep t, with t ranging from K to 0. At each timestep t, the model predicts the added noise, $\epsilon_\theta(x_t, t)$. This predicted noise is removed from the current image to obtain a denoised version x_{t-1}, by reversing the diffusion process defined during the forward process:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right) + \sigma_t z \tag{3.12}$$

where $z \sim \mathcal{N}(0, I)$ is random Gaussian noise, and σ_t is derived from the noise variance schedule used during training:

$$\sigma_t = \sqrt{\beta_t} \tag{3.13}$$

This iterative procedure progressively denoises the image into a high-quality sample from the learned data distribution.

Our implementation is based on the open-source MONAI framework [29]. This framework contains various validated deep learning tools specifically tailored to applications in medical imaging. The main architecture used in this thesis is MONAI's implementation of a diffusion model [30], which is a 3D UNet-based network designed for denoising diffusion probabilistic models (DDPMs). The network was configured with three spatial dimensions, channel sizes of $32, 64, 64$, and one residual block per level. A linear beta schedule with 1000 timesteps was chosen, with the variance schedule linearly increasing from $\beta_{start} = 1 \times 10^{-4}$ to $\beta_{end} = 0.02$, as described in the original DDPM paper [23].

To condition the model on the current diffusion timestep, a timestep embedding is integrated into the network. The diffusion model architecture from MONAI uses a sinusoidal positional encoding to convert a scalar timestep into a higher-dimensional embedding. The resulting embedding is then passed through a small feedforward neural network, which projects the embedding into a representation that can be used by the UNet architecture. Finally, the timestep embedding is added to the network's residual block, which allows the model to adapt its denoising behavior based on the denoising timestep it is at.

3.2.1. Training strategy

The model was trained using the reference contrast-weighted images as input data and as output target. Due to memory constraints associated with entire 3D volumes, a dynamic patch sampling strategy was used. During each training iteration, a batch of 32 patches of size $32 \times 32 \times 32$ was sampled uniformly across the entire image volumes from the set of 76 synthetic subjects. To manage training duration, the total number of iterations over all epochs was fixed to 100,000. This ensured a feasible training time, while also exposing the model to a varied dataset. The batch size and patch size were selected based on empirical trade-offs between GPU memory limits and the need for spatial context.

For the training objective, the mean squared error (MSE) loss between the predicted noise and the true noise was used. The model was optimized using the Adam optimizer, with a fixed learning rate of 1×10^{-4}.

As in partial diffusion the full 1000 set of timesteps will not be used during the reverse diffusion process, there is also no need to train the model across all 1000 timesteps. Instead, during training, timesteps are randomly sampled only from the range [0, 100]. This approach ensures that the training time is used effectively to only learn the denoising steps in the range that will be used during the reverse diffusion process.

Training took approximately 7.5 hours on an NVIDIA A40 GPU with 48 GB of memory[1]. Early stopping was not implemented, as random patch sampling provided natural regularization across epochs.

3.2.2. Reverse Diffusion With Sliding Window Approach

Once the models are trained, model performance is evaluated on the held-out test set of 10 synthetic test subjects from the same brain model. Due to memory constraints, it is not possible to process an entire test subject at once. To address this, a sliding window approach, developed within the BIGR department, was used. First, the brain volume is passed through the forward diffusion process up to a predetermined number of diffusion steps. Next, a 3D sliding window, with predefined dimensions and a stride equal to half the window dimensions, is applied to draw patches from the input volume. To ensure complete coverage of the volume, the input is zero-padded such that an integer number of sliding windows fits in its dimensions. At each denoising step, the selected noised patch x_t^p, is passed to the model together with the current timestep, t, to predict the noise at this timestep. This predicted noise, $\epsilon_\theta(x_t^p, t)$, is then used to compute the denoised output x_{t-1}^p, according to the reverse diffusion process as described in Equation 3.12.

Important to note is that in the standard reverse diffusion process, the random noise, z in Equation 3.12, is sampled from a standard Gaussian distribution at each denoising step. However, to ensure consistency across overlapping regions from different sliding window patches, each voxel should use the same noise realization, regardless of the window in which it is evaluated. Therefore, at each timestep, a single noise value is sampled per voxel across the entire brain volume. Each sliding window then retrieves the noise values corresponding to the voxels it contains and uses those in the reverse diffusion process. After a denoising step at time t for each patch, all the output patches are merged. Overlapping voxels are combined using a Gaussian-weighted averaging, where voxels near the edge of the sliding window get a lower weight than voxels near the center. This helps reduce boundary artifacts. After merging, the next round of patches is drawn from the updated volume in the same sliding window manner and passed through the subsequent denoising step $(x_{t-1} \rightarrow x_{t-2})$. This process is repeated until the final denoised volume at x_0 is reconstructed. The approach is illustrated in Figure 3.3 for one slice of one channel of a 3D brain volume.

3.2.3. Model Variations

In this thesis, three variations of the partial diffusion model were evaluated: an unconditional model, a latent alignment model, and a concatenation-based conditional model. The unconditional model is trained as described in Subsection 3.2.1, with only the reference images. The training procedure of the unconditional DDPM is depicted in Figure 3.4.

This model serves as a baseline to assess the performance of the partial diffusion process by itself but has two notable limitations. First, since the model is not exposed to the short-acquisition images during training, it cannot learn features related to the specifics of this image distribution. Second, as explained in Subsection 2.3.1, the input image is noised during inference by applying forward diffusion steps. This inherently destroys some of the original structure information present in the short acquisition, making it challenging for an unconditioned model to retain and reconstruct fine details without any guidance.

To address these limitations, two conditional model variations were designed, both aiming to preserve and use the information present in the short acquisition. The latent alignment model follows the approach used in [24], where both the short-acquisition and the reference images are used during training. The goal is to train the model to map the distribution of short-acquisition images toward that of the reference images. The idea is that the latent spaces, x_t in this case, of both domains should be aligned throughout the diffusion process. This is achieved by linearly interpolating between the short-acquisition image x_t^{s} and the reference image x_t^{ref} at timestep t, according to a weight, w. Specifically, for a given diffusion timestep t, the input to the forward diffusion process is computed as:

$$x_{\mathrm{t}} = w \cdot x_t^{\mathrm{s}} + (1 - w) \cdot x_t^{\mathrm{ref}} \tag{3.14}$$

where $w = t/K_{train}$ and K_{train} is the maximum timestep used in the training process. This interpolation ensures that earlier diffusion steps are more influenced by the reference image, while later steps are

[1] Specifications available at: `https://www.nvidia.com/en-us/data-center/a40/`

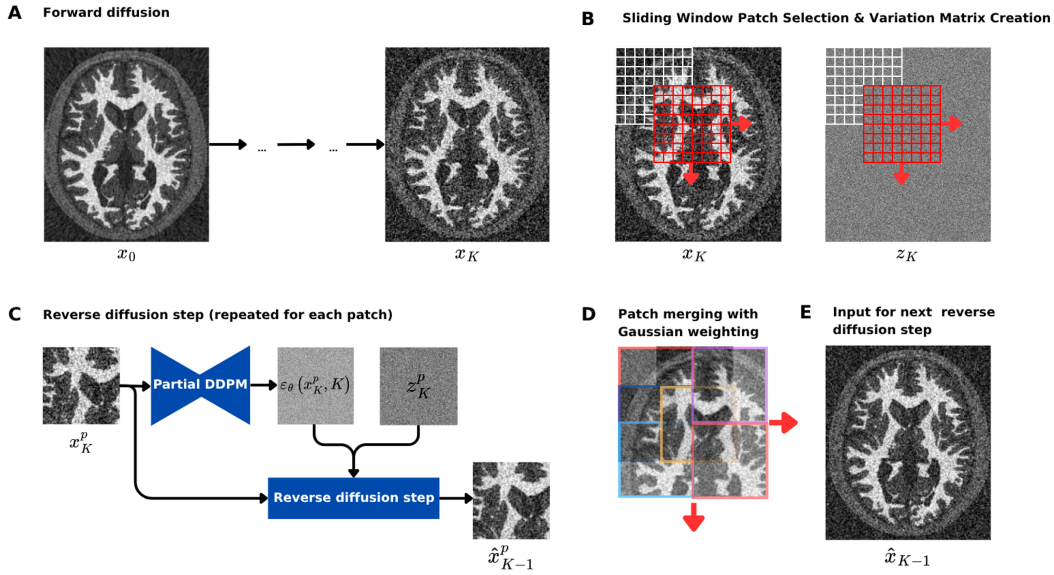

Figure 3.3: The sliding window approach used in inference. One slice of one channel is depicted for visualization purposes. (A) First, the input image is noised through the forward diffusion process up to the predetermined number of reverse diffusion steps, $t = K$. (B) The brain volume is divided into patches and a Gaussian volume of the same size is created for the realization of the noise. (C) Each patch is passed through the trained partial DDPM, which predicts the noise on the patch, $\epsilon_\theta(x_{\mathrm{K}}^p, \mathrm{K})$. Together with the realization of the noise of this patch, this predicted noise is used to perform a reverse diffusion step (Equation 3.12), which leads to $\hat{x}_{\mathrm{K}-1}^p$. (D) Once every patch is processed for the current timestep, all patches are merged using Gaussian weighting. (E) This generates $\hat{x}_{\mathrm{K}-1}$, which is again split into patches and processed in the same manner as described in (B), (C) and (D), until $t = 0$.

increasingly conditioned on the short-acquisition image. The interpolated image, x_{in}, is then treated as the input for the forward diffusion step at timestep t, after which Gaussian noise is added according to the DDPM noise schedule. The training process of the partial DDPM with latent alignment is depicted in Figure 3.5.

The concatenation-based conditional model, on the other hand, addresses the limitations of an unconditional model by explicitly maintaining access to the original short-acquisition image throughout the diffusion process. In this model, the short-acquisition input is concatenated channel-wise with the noised reference image during training. This allows the model to condition its denoising predictions not only on the progressively noised version of the reference, but also on the original short-acquisition input. During inference, the original short-acquisition input is concatenated with its own noised version. In this way, the model can preserve the structural information present in the original image which would otherwise be lost during forward diffusion, while still benefiting from the power of the diffusion model.

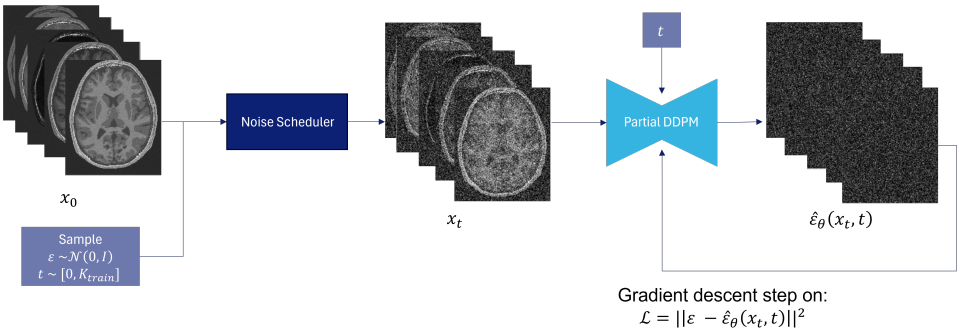

Figure 3.4: Training procedure of the unconditional partial DDPM. At each iteration, a timestep $t \sim [0, K_{\mathrm{train}}]$ and noise $\epsilon \sim \mathcal{N}(0, I)$ are sampled and applied to the clean reference image, x_0 to produce a noisy sample, x_t. The same timestep t is then passed to the partial DDPM along with x_t to predict the added noise, $\hat{\epsilon}_\theta(x_t, t)$, and the model is optimized using a mean squared error loss. Note that the figure is for illustrative purposes, in reality, the model was trained on patches due to memory constraints.

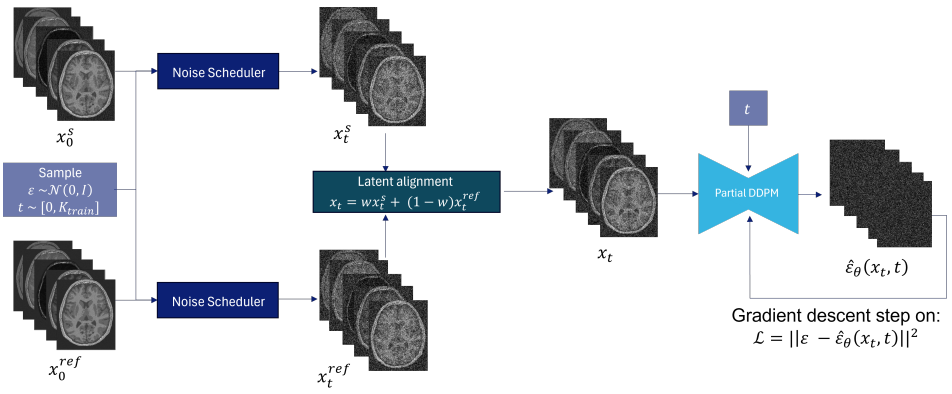

Figure 3.5: Training procedure of the partial DDPM with latent alignment. At each iteration, a timestep $t \sim [0, K_{\text{train}}]$ and noise $\epsilon \sim \mathcal{N}(0, I)$ are sampled and applied to the short-acquisition image, x_0^{s}, and the reference image, x_0^{ref}, to produce noisy samples x_t^{s} and x_t^{ref}. These two noised images are linearly interpolated according to Equation 3.14. The interpolated output, x_t, together with this timestep t are used as inputs to the network. The partial DDPM is trained to predict the added noise, $\hat{\epsilon}_\theta(x_t, t)$, and the model is optimized using a mean squared error loss. Note that the figure is for illustrative purposes, in reality, the model was trained on patches due to memory constraints.

The training process of the partial DDPM with concatenation is depicted in Figure 3.6.

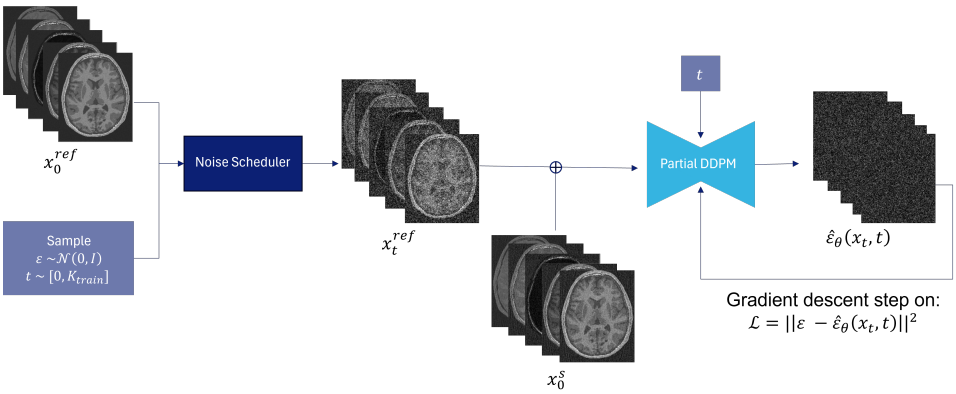

Figure 3.6: Training procedure of the partial DDPM with concatenation. At each iteration, a timestep $t \sim [0, K_{\text{train}}]$ and noise $\epsilon \sim \mathcal{N}(0, I)$ are sampled and applied to the reference image, x_0^{ref}, to produce a noisy sample, x_t^{ref}. The short-acquisition image x_0^{s} is concatenated with this noisy sample x_t^{ref} and used, together with the timestep t, as input to the network. The partial DDPM is trained to predict the added noise, $\hat{\epsilon}_\theta(x_t, t)$, and the model is optimized using a mean squared error loss. Note that the figure is for illustrative purposes, in reality, the model was trained on patches due to memory constraints.

3.2.4. Baseline UNet Comparison

To provide a performance baseline and evaluate the added value of diffusion-based models, a standard 3D UNet was also implemented and evaluated qualitatively on the same test subjects. The 3D UNet was implemented using the standard UNet framework by MONAI and adjusted such that it matched the UNet used in the partial diffusion models as accurately as possible [29]. Three spatial dimensions, channel sizes of $32, 64, 64$ and one residual block per level were set. The training settings were also identical, using the same patch drawing approach, a batch size of 32, learning rate of $1 \cdot 10^{-4}$ and a fixed number of 100,000 iterations. The model was trained to directly reconstruct a higher-quality version of the short-acquisition images by minimizing the mean squared error between its output and the corresponding long-acquisition reference images.

3.3. Dictionary Matching

As the ultimate goal of the MuPa-ZTE sequence is to generate high-quality quantitative maps with short acquisition times, it is important to evaluate how the enhanced contrast-weighted images, produced by the partial diffusion models, influence the quality of the resulting quantitative maps. To evaluate this, a PD, T_1, and T_2 map were generated from the quality-enhanced echo images. The PD map is

estimated by normalizing the PD-weighted image (Echo 0) using a high-signal average to correct for intensity inhomogeneity caused by external factors. For the T_1 and T_2 maps, dictionary matching was used to estimate the T_1 and T_2 values using Echo 1-4.

Dictionary matching is a model-based parameter estimation technique commonly used in quantitative MRI [31]. In dictionary matching, a large set of signal evolutions are simulated for a range of possible tissue properties using a signal model of the acquisition sequence. The dictionary matching implementation used in this work relies on the forward model described in Subsection 3.1.3 for this purpose. All signal evolutions are saved as dictionary entries, each representing the expected signal intensities across the last four echo times for a specific combination of tissue properties.

During the matching process, the observed signal evolution at each voxel is iteratively compared to the entries in the dictionary. The candidate T_1 and T_2 values are updated over a fixed number of iterations, converging to the dictionary entry that gives the closest match. This dictionary entry is selected and its corresponding tissue properties, T_1 and T_2 are assigned to the voxel. Finally, this allows for the mapping of PD, T_1, and T_2 values for the entire brain volume.

3.4. Model Evaluation

All three model types were evaluated qualitatively as well as quantitatively, both on different synthetic test subjects as well as on an in vivo subject. The qualitative assessment was done by visual inspection of the output contrast-weighted images and the corresponding quantitative maps. This evaluation focused on identifying visual improvements, such as sharper details, reduced noise, and preservation of structural features.

The quantitative assessment was done by calculating five commonly used image quality metrics for the contrast-weighted images, as well as the derived quantitative maps: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and High Frequency Error Norm (HFEN). Each metric focuses on different aspects of image quality. A brief description of these metrics is provided below. To ensure the evaluation focused solely on the relevant part of the image, all metrics were computed only within the brain region, using a brain mask to set all background voxels to zero prior to evaluation.

MAE measures the average absolute difference between the predicted and reference images across all pixels. It is a simple and interpretable metric, but it does not account for spatial structure or perceptual quality [32].

RMSE, on the other hand, calculates the root of the average squared differences between the predicted and reference image. It places more emphasis on large errors than MAE, due to the squaring operation, making it more sensitive to outliers [32]. Like MAE, RMSE remains a pixel-wise measure and it does not account for spatial structure or perceptual quality.

PSNR is a logarithmic measure (in decibels (dB)) that compares the maximal signal value with the power of the noise. It is a commonly used metric in image quality quantification, but it still lacks sensitivity to perceptual quality and spatial structure.

SSIM, on the other hand, is designed specifically to match human visual perception of image quality [33]. Therefore, it takes into account luminance, contrast, and spatial structure rather than pixel-wise error metrics. This makes it useful to assess perceptual image quality.

Lastly, HFEN measures the difference in high-frequency content between the predicted and reference images by applying a high-pass filter before computing the error [34]. This metric is sensitive to structural detail loss and blurring and thus very relevant in the context of this thesis.

MAE, RMSE, and HFEN are error-based metrics, where lower values indicate better performance, with zero being the ideal. PSNR and SSIM, on the other hand, are similarity-based metrics, where higher values are preferred. PSNR has a theoretical upper bound of infinity as it divides by the mean squared error, but values typically range between 20-50 dB, while SSIM ranges from 0 to 1, with 1 indicating perfect similarity.

4

Experimental Setup

In this chapter we describe the experimental setup used to evaluate the proposed partial diffusion model. The aim was to assess three different architecture variants of the partial DDPM, as introduced in Chapter 3, under consistent and controlled conditions.

The goal of the experiments was to compare the ability of each partial DDPM to enhance undersampled contrast-weighted images acquired by 4x accelerated MuPa-ZTE. A secondary objective was to evaluate whether these enhanced contrast-weighted images naturally led to improved quantitative maps. In addition, we investigated how the number of denoising steps in the reverse denoising process influences the performance of each model.

4.1. Training

Three model architectures were compared, the unconditional partial DDPM, which was trained using only the reference images, the partial DDPM with latent alignment, which was trained using an interpolation between the latent representations of the undersampled and reference image as a function of the timestep during forward diffusion, and the partial DDPM with concatenation, which was trained using the reference image, but with the undersampled image concatenated channel-wise.

Each model was trained using the same 76 synthetic subjects, identical hyperparameters (learning rate, batch size, number of epochs), and the same partial diffusion noise scheduler. The specifics of these settings are described in Chapter 3.

4.2. Evaluation

The evaluation consisted of two parts: the synthetic evaluation and the in vivo evaluation. For the synthetic evaluation, the held-out test brain model was used and for the in vivo evaluation the MuPa-ZTE acquisition of a volunteer was used. Both evaluations consisted of a qualitative and a quantitative component. The quantitative evaluation metrics included the mean absolute error (MAE), the root mean squared error (RMSE), the high frequency error norm (HFEN), the peak signal-to-noise ratio (PSNR), and the structural similarity index measure (SSIM). These are defined and discussed in Section 3.4.

4.2.1. Synthetic Evaluation

Ten synthetic test subjects were generated from the held-out brain model that was not used during training. The undersampled images of each synthetic subject were enhanced using all three model architectures. To assess the effect of the number of reverse diffusion steps, inference was performed at 11 different numbers of diffusion steps, denoted as K:

$$K \in \{5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\} \tag{4.1}$$

At each K, enhanced contrast-weighted images were generated from the undersampled images for all 10 subjects. During inference, the 3D sliding window approach was used, as discussed in 3.2.2, with a

sliding window size of 32x32x32 and a stride of 16x16x16. These outputs were assessed qualitatively based on visual inspection of the accuracy of the anatomical structures, the noise suppression, and the sharpness of boundaries. Quantitative image quality metrics were computed voxel-wise within the brain mask, comparing the enhanced contrast-weighted images with the reference synthetic images. For comparison, these metrics were also computed for the original undersampled image and the fully sampled images.

For each model, the optimal number of reverse diffusion steps K was selected as the smallest value at which the model's performance remained within 1% of its performance maximum, across all evaluation metrics, in order to minimize inference time while maintaining high image quality. For this purpose, image quality metrics were first averaged across all echoes for each subject. Afterward, these per-subject averages were combined by computing the mean across subjects, resulting in a single mean value per metric per K. The optimal K was selected based on these values.

Using the optimal K, enhanced contrast-weighted images of all 10 subjects were generated. These were used to generate quantitative PD, T_1, and T_2 maps for each model, using the dictionary matching approach as described in Section 3.3. The resulting maps were compared qualitatively and quantitatively to the synthetic reference maps, in the same manner as for the enhanced contrast-weighted images.

4.2.2. In Vivo Evaluation

To assess generalizability to real-world data, the three models were evaluated on two in vivo subjects. This subject underwent a long MuPa-ZTE acquisition. However, as the original scan parameters were not recorded in full detail, the acquisition protocol did not exactly match the one used in the forward model for generating the training data. Specifically, the total number of cycles in the k-space trajectory was 102 (instead of the 126 used in training). This data was truncated in a similar manner as for the synthetic data, where the first 27 cycles ($\sim 26\%$) were kept to generate the five contrast-weighted echoes acquired with a 4x accelerated, short MuPa-ZTE acquisition.

These short-acquisition images were used as input to the models, where each model used its previously determined optimal K. The resulting enhanced short-acquisition images were assessed qualitatively in the same manner as the enhanced synthetic contrast-weighted images. However, unlike the synthetic data, no ground truth contrast-weighted images are available for the in vivo subject. To be able to do a meaningful quantitative evaluation, an additional long MuPa-ZTE acquisition was acquired of the same subject during the same scanning session. This second long acquisition was used as the reference for all quantitative evaluations. The original long acquisition, from which the short-acquisition input was derived, could not serve as a valid reference, as it inherently shares noise characteristics and potential artifacts with the short-acquisition input. Using an independently acquired full MuPa-ZTE scan as the reference minimizes bias and allows for a more objective assessment of the model's ability to enhance short-acquisition contrast-weighted images. The image quality metrics were also computed for the short-acquisition image with respect to this reference long acquisition and for the first long acquisition with respect to the reference long acquisition. No image registration was applied, as the scans were acquired consecutively in the same session, and no significant subject motion was expected between acquisitions.

Next, the five enhanced short-acquisition contrast-weighted images of the in vivo subject were used to generate PD, T_1, and T_2 maps. These maps were again evaluated both qualitatively and quantitatively in the same manner as the quantitative maps generated from the synthetic contrast-weighted images but with respect to the reference long MuPa-ZTE acquisition.

In addition to the primary in vivo evaluation, a second in vivo dataset was available from a subject who underwent a single long MuPa-ZTE acquisition with the same acquisition parameters as those used for training the models. As no additional reference scan was available for this subject, only qualitative evaluation was possible. This dataset was therefore not included in the main analysis but was used to further visually assess the model's performance on unseen real-world data. The resulting enhanced contrast-weighted images and corresponding quantitative maps are included in Appendix A for completeness.

5

Results

In this chapter we present the results of the experimental evaluation of three variations of the partial DDPM. The results are organized into two main sections: the results on the synthetic dataset and the results on the in vivo dataset. Each section contains both qualitative and quantitative evaluations. Two types of comparisons were carried out between the short-acquisition data and their enhanced counterparts. The first comparison is done in the contrast-weighted image domain, while the other comparison is done in the quantitative map domain.

5.1. Results on Synthetic Dataset

5.1.1. Influence of partial diffusion steps in inference

Figure 5.1 shows that all three models are capable of improving all five evaluated metrics compared to the short-acquisition input images for specific values of K. The unconditional partial DDPM and latent alignment variant produce similar results across the evaluated K values, with optimal performance at $K = 10$. Beyond this point, the performance deteriorates. In contrast, the concatenation-based partial DDPM exhibits consistent improvement in metrics as K increases, although the rate of improvement levels off after $K = 50$. The optimal K was identified as the smallest number of reverse diffusion steps that achieved performance within 1% of the maximum performance, in order to minimize inference time while maintaining high image quality. Based on this criterion, the optimal value for K was determined to be 10 for both the unconditional partial DDPM and the latent alignment variant, and 50 for the concatenation-based model.

5.1.2. Contrast-Weighted Images Comparison

In the following experiments, we use $K = 10$ for the unconditional partial DDPM and the latent alignment variant, and $K = 50$ for the concatenation-based model. First, the enhanced contrast-weighted images were compared visually across the models. A representative slice of the 3D brain volume of one of the test subjects, enhanced with each of the three models, is visualized in Figure 5.2 for visual assessment.

To further facilitate visual comparison, Figure 5.3 presents pixel-wise difference images computed between the reference and all other images for the same subject and slice as presented in Figure 5.2. Figure 5.2 shows that all three partial DDPM variants reduce noise across all echoes, compared to the short-acquisition input. The outputs of the unconditional partial DDPM and the latent alignment variant still demonstrate some residual noise compared to the long-acquisition echoes, whereas the concatenation-based model produces outputs with minimal visible noise compared to the long-acquisition and reference images. The UNet output appears notably more blurred compared to the partial DDPM outputs. Figure 5.3 further highlights that the concatenation model achieves the lowest residual error among the three DDPM variants. However, its output still contains slightly more deviation from the reference image than the long acquisition. Interestingly, all models demonstrate the greatest improvement for Echo 2, which shows the least amount of residual error in the outputs.

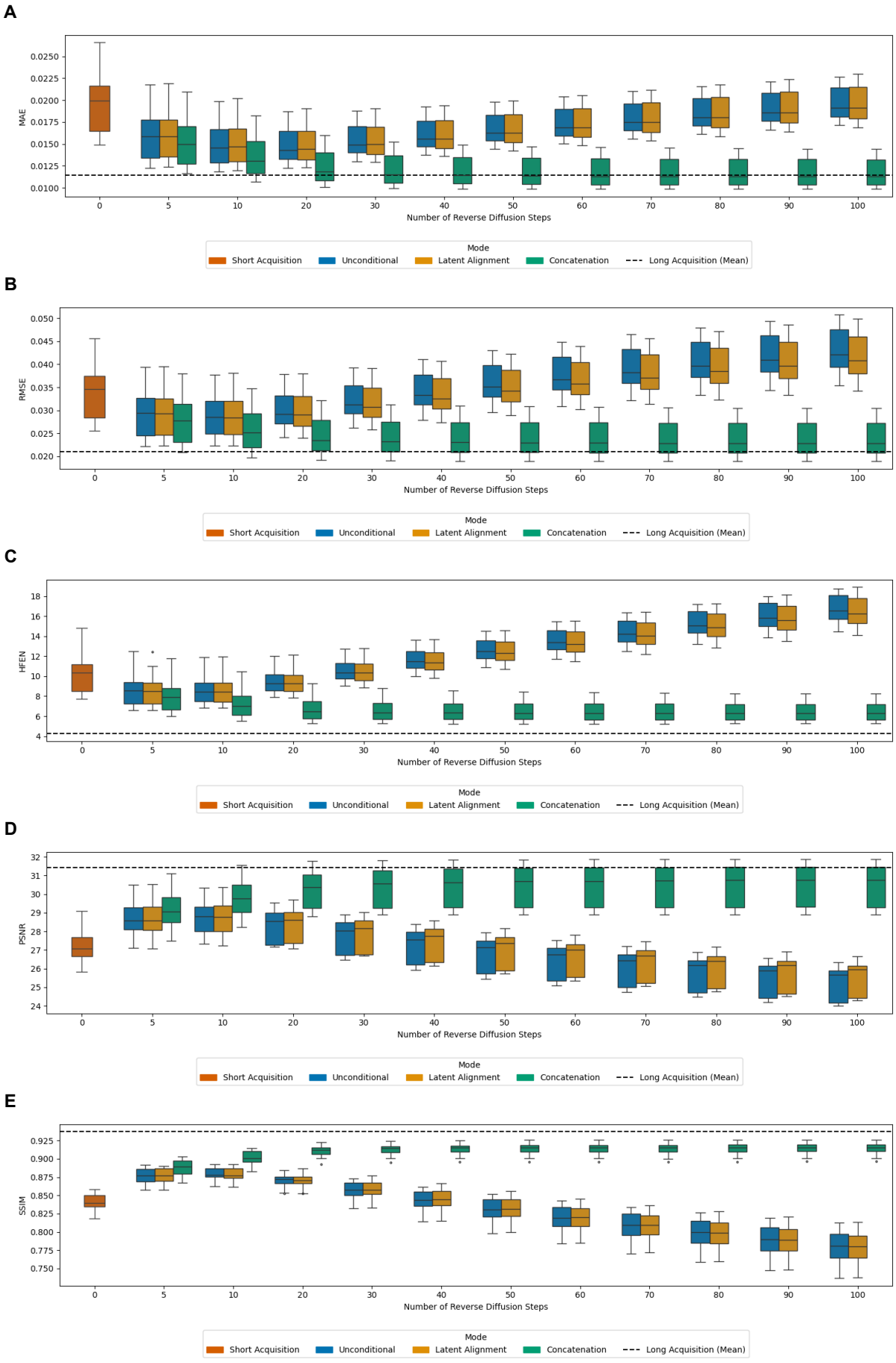

Figure 5.1: Boxplots of five image quality metrics (A-E) computed across 11 different reverse diffusion step counts for three different models, evaluated on 10 synthetic test brain volumes. All metrics are calculated with respect to the reference image and are computed using a brain mask. Timestep zero corresponds to the short-acquisition input images. The dashed line indicates the mean metric value for the long-acquisition images, also with respect to the reference image. Subplots show: (A) MAE, (B) RMSE, (C) HFEN, (D) PSNR, and (E) SSIM.

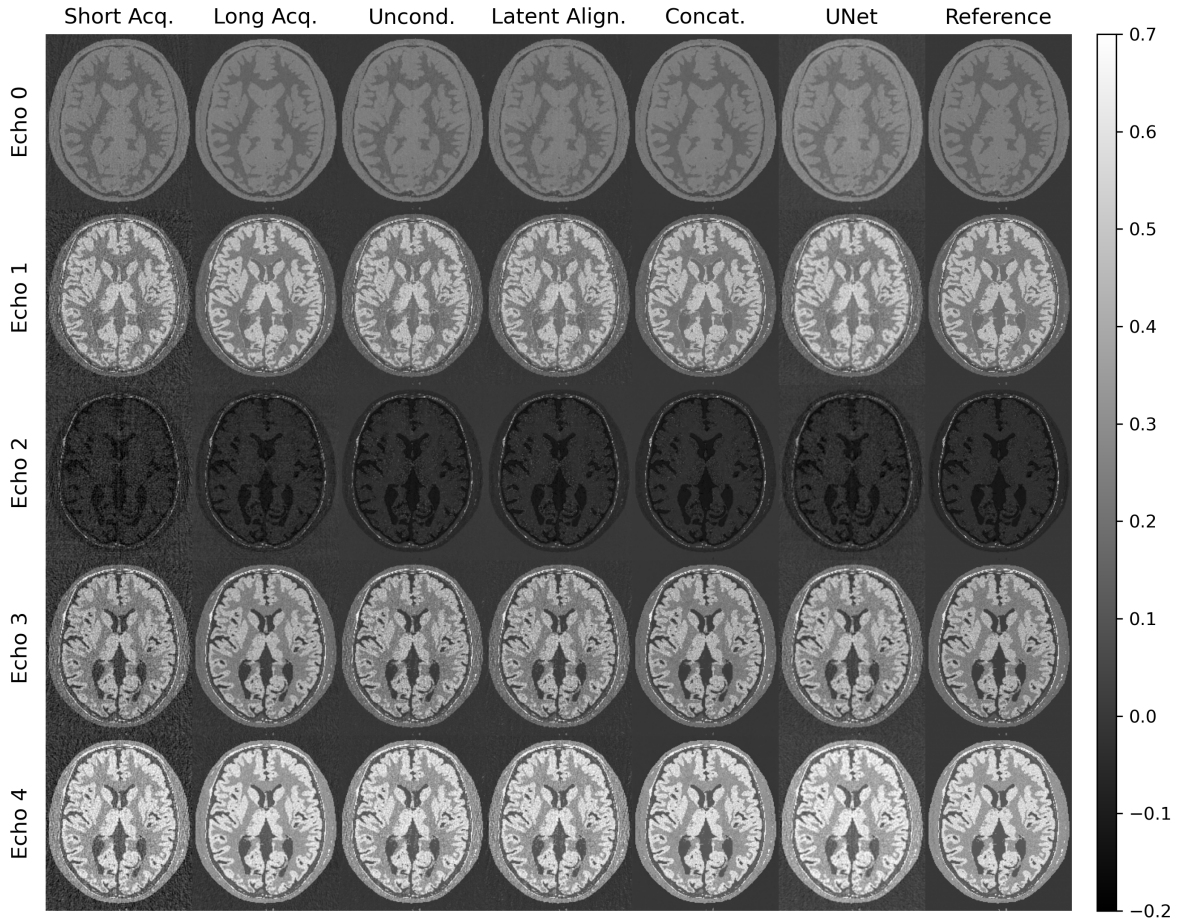

Figure 5.2: Contrast-weighted images for a synthetic test subject (single slice shown). The first column displays the five short-acquisition input echoes provided to all models. The second column displays the corresponding long-acquisition echoes, while the last column shows the reference echoes. The third, fourth, and fifth columns show the output of the unconditional partial DDPM, the latent alignment variant, and the concatenation-based model, respectively. The sixth column displays the output of a basic UNet. The rows display the five echoes associated with the MuPa-ZTE acquisition.

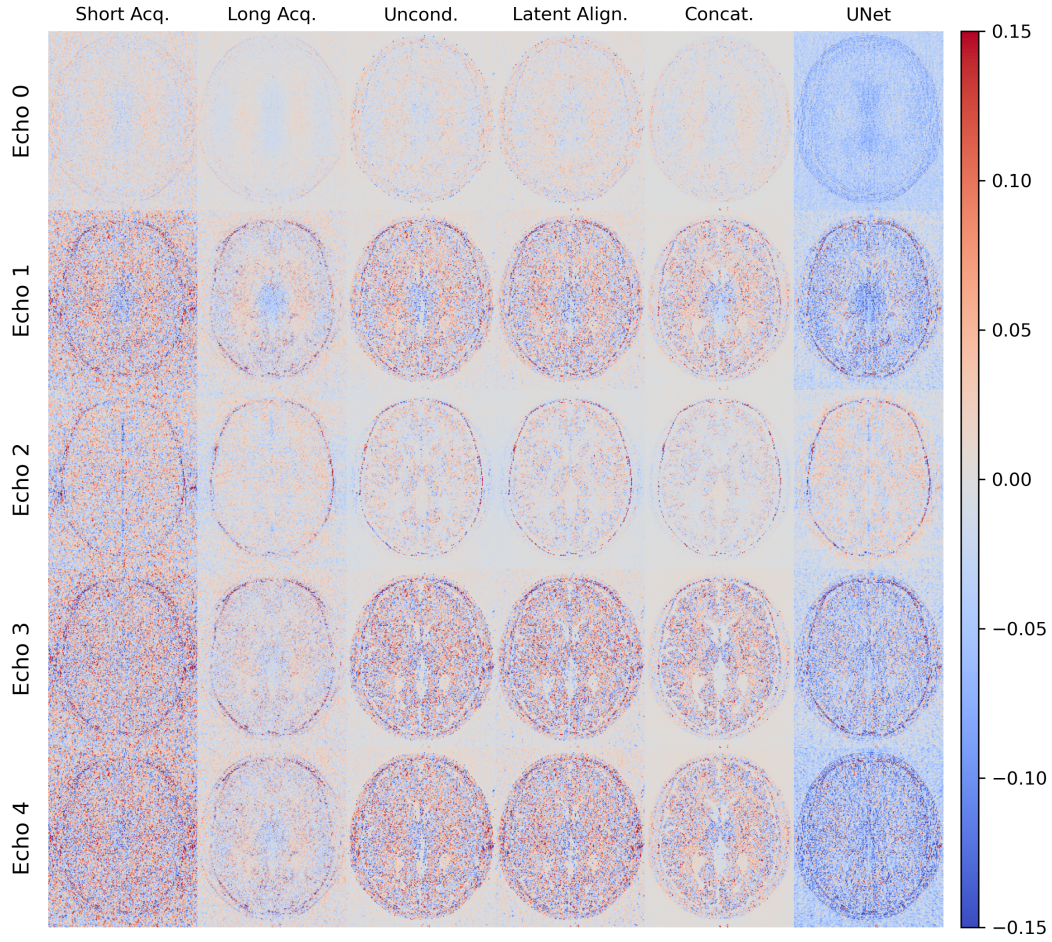

Figure 5.3: Difference images for the same synthetic test subject as presented in Figure 5.2 (single slice shown). Each column presents the pixel-wise difference between the reference images and the images from the corresponding acquisition or model indicated in the column title.

Figure 5.4 and 5.5 show two zoomed-in patches from the same synthetic brain model as depicted in Figure 5.2, to highlight the differences in detail preservation and noise reduction across the models. In Figure 5.4, red arrows indicate a thin light grey structure with clear boundaries in the reference image. These boundaries are blurred in the short-acquisition image due to noise but are restored in the result of the concatenation-based partial DDPM. The unconditional partial DDPM and the latent alignment variant also recover the structure, but it remains noisy. The UNet output is more blurred compared to the outputs of the partial DDPMs and the reference image. Light blue arrows in Figure 5.4 point to white matter regions that are noticeably denoised in the concatenation-based partial DDPM when compared to the short-acquisition input image. While the results of the unconditional partial DDPM and the latent alignment variant also show some denoising, this is less pronounced compared to the concatenation-based model. The UNet output also shows less noise but has a mean intensity shift in this white matter region.

In Figure 5.5, red arrows indicate a boundary region between two tissue types. While this boundary appears relatively straight in the reference image, it is noticeably degraded in the short-acquisition image. The UNet model reduces the noise but produces a very blurred boundary. In contrast, both the unconditional partial DDPM and the latent alignment variant show improved sharpness, yet the boundary remains irregular. The concatenation-based partial DDPM, on the other hand, successfully recovers a sharp, straight boundary that closely resembles that of the reference image.

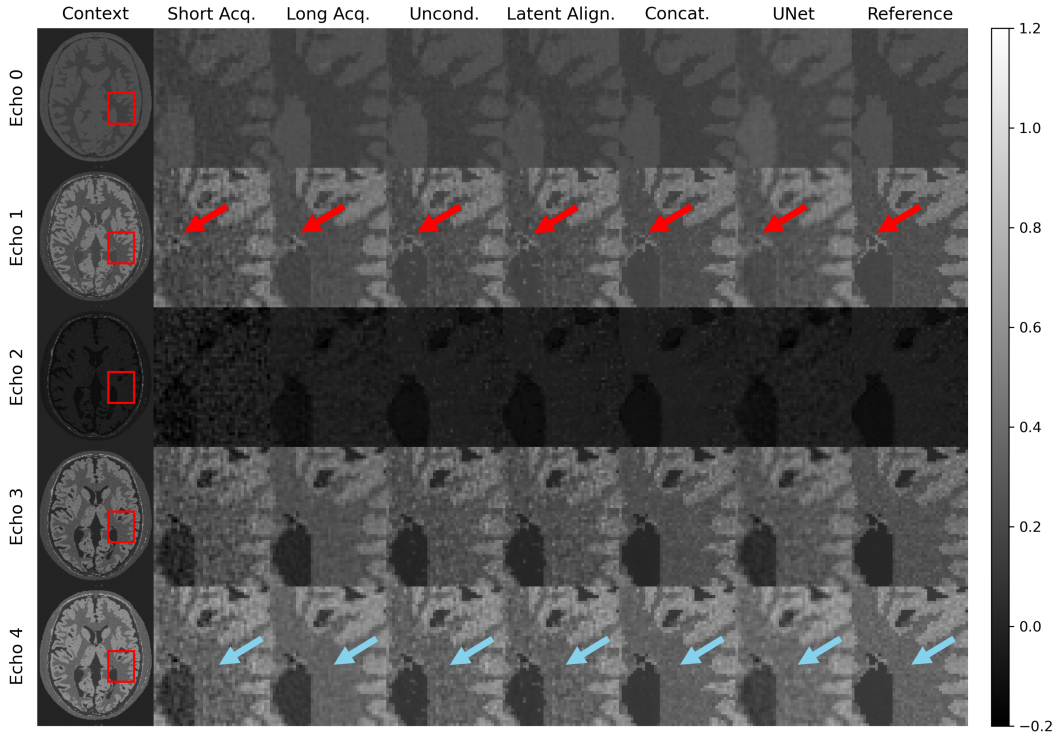

Figure 5.4: Visual comparison of the three models on a 40x48 patch from one of the synthetic test subjects. The first column shows the context of the patch. The rows and other columns are arranged in the same order as in Figure 5.2. Red arrows indicate a thin, light grey structure whose boundary sharpness varies across the models. Blue arrows indicate a white matter region where noise levels vary across the models.

In addition to the qualitative assessment, we analyzed the mean image quality metrics for each model using their respective selected K values. All three model variations were able to achieve lower MAE, RMSE and HFEN and higher PSNR and SSIM compared to the short-acquisition input images (MAE: 0.02, RMSE: 0.034, HFEN: 10.4, PSNR: 27.2, SSIM: 0.84), see Figure 5.1. Out of the three model variations, the concatenation-based partial DDPM achieved lower mean values for MAE (0.012), RMSE (0.024) and HFEN (6.5), and higher mean values for PSNR (30.4) and SSIM (0.91) compared to the unconditional and latent alignment variants which achieved MAE of 0.015, RMSE of 0.029, HFEN of 8.8, PSNR of 28.7, and SSIM of 0.88. The concatenation-based model achieved MAE and RMSE mean values similar to the long acquisition (MAE: 0.012, RMSE: 0.021).

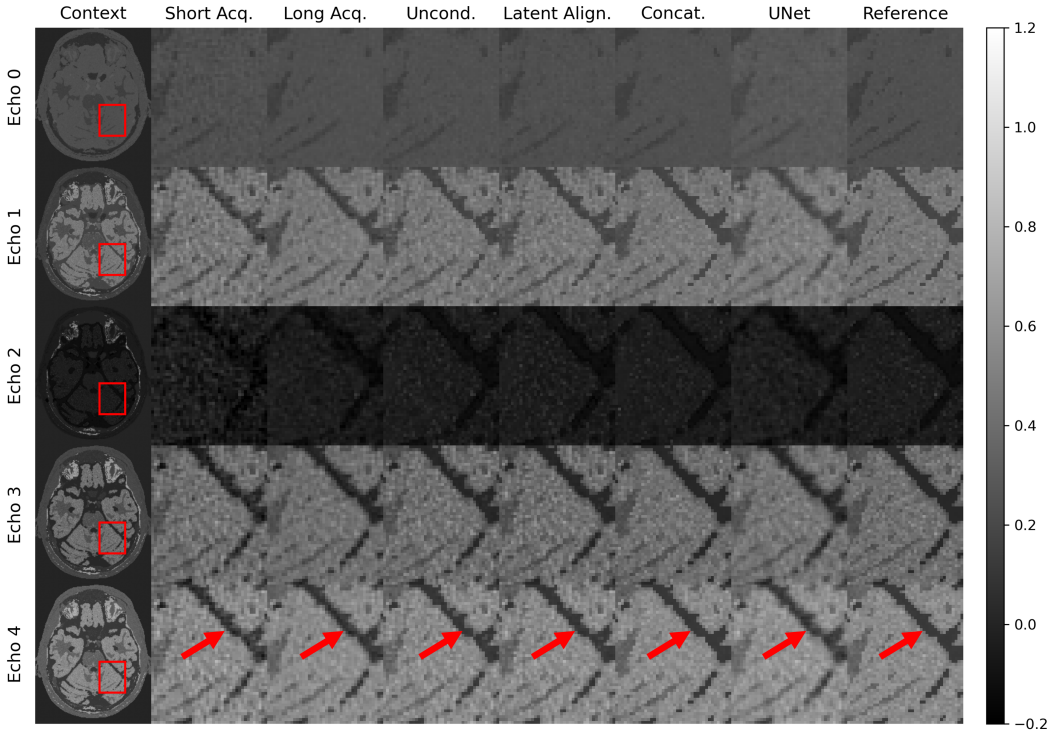

Figure 5.5: Visual comparison of the three models on a 40x48 patch from one of the synthetic test subjects. The first column shows the context of the patch. The rows and other columns are arranged in the same order as in Figure 5.2. Red arrows indicate a straight tissue boundary, which is differently recovered across the models.

5.1.3. Quantitative Maps Comparison

Next to the contrast-weighted image comparison, the quantitative maps reconstructed from the enhanced contrast-weighted images were also compared. Figure 5.6 presents a single slice from the 3D quantitative maps generated using the outputs of each of the three models, for one test subject. Figure 5.7 shows the corresponding difference images between the reference maps and the maps indicated by the column titles. The resulting T_1 and T_2 maps from all three models show reduced noise compared to the short-acquisition maps. This improvement is especially notable in the T_2 maps, where tissue contrast is also visibly enhanced. The unconditional partial DDPM and the latent alignment variant lead to PD and T_1 maps that resemble their long-acquisition counterparts. Their T_2 maps even appear slightly less noisy than the long-acquisition T_2 map, which is supported by the reduced residual error visible in Figure 5.7. The concatenation-based partial DDPM very closely resembles the reference maps, especially for T_1 and T_2. This is supported by the lowest residual error visible for both the T_1 and T_2 maps in Figure 5.7. The structures in the PD map are visually similar, however a slight global intensity shift is observed. This shift is also visible in the long-acquisition PD map. Figure 5.7 shows this global intensity shift as an overall negative residual error without much structure visible.

The quantitative maps were also analyzed quantitatively across the different models. The image quality metrics for the T_1, T_2, and PD maps are shown in Figures 5.8-5.10, respectively. Figure 5.8 shows that all three models improve the image quality metrics for the T_1 maps compared to the short-acquisition T_1 maps in the synthetic test subjects. Notably, the concatenation-based partial DDPM outperforms the long-acquisition images across all metrics. Similarly, Figure 5.9 demonstrates that all three models improve the metrics for the T_2 maps compared to the short-acquisition T_2 maps. In this case, all three models achieve metrics exceeding that of the long-acquisition images, except for the SSIM of the unconditional partial DDPM and the latent alignment variant. In contrast, Figure 5.10 shows a different trend for the PD maps. While the concatenation-based partial DDPM improves all metrics relative to the short-acquisition maps, the other two models only improve the MAE, RMSE, and PSNR but not the HFEN and SSIM. Notably, none of the models improve the metrics for the PD maps beyond that of the long-acquisition image (MAE: 0.03, RMSE: 0.052, HFEN: 9.0, PSNR: 27.7, SSIM: 0.94).

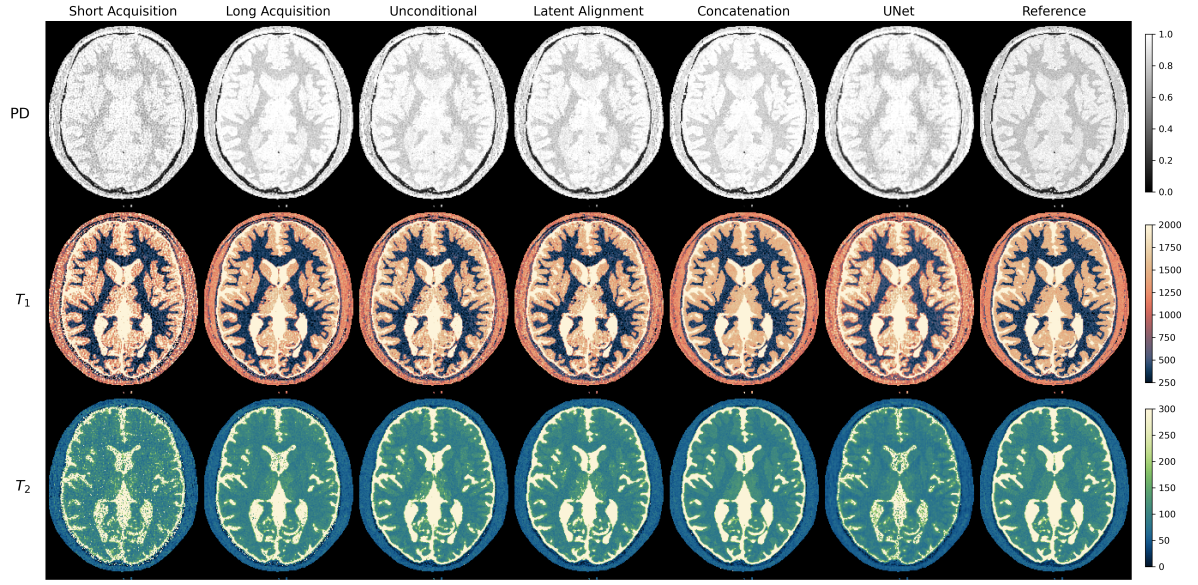

Figure 5.6: Single slice of the PD, T_1, and T_2 quantitative maps of a synthetic test subject reconstructed from the short-acquisition contrast-weighted images, long-acquisition contrast-weighted images, the result of the unconditional partial DDPM, the result of the partial DDPM with latent alignment, the result of the concatenation-based partial DDPM, the result of the UNet, and the reference contrast-weighted images. The background was manually set to zero for all quantitative maps. T_1 and T_2 maps are in milliseconds, and PD is in arbitrary units.

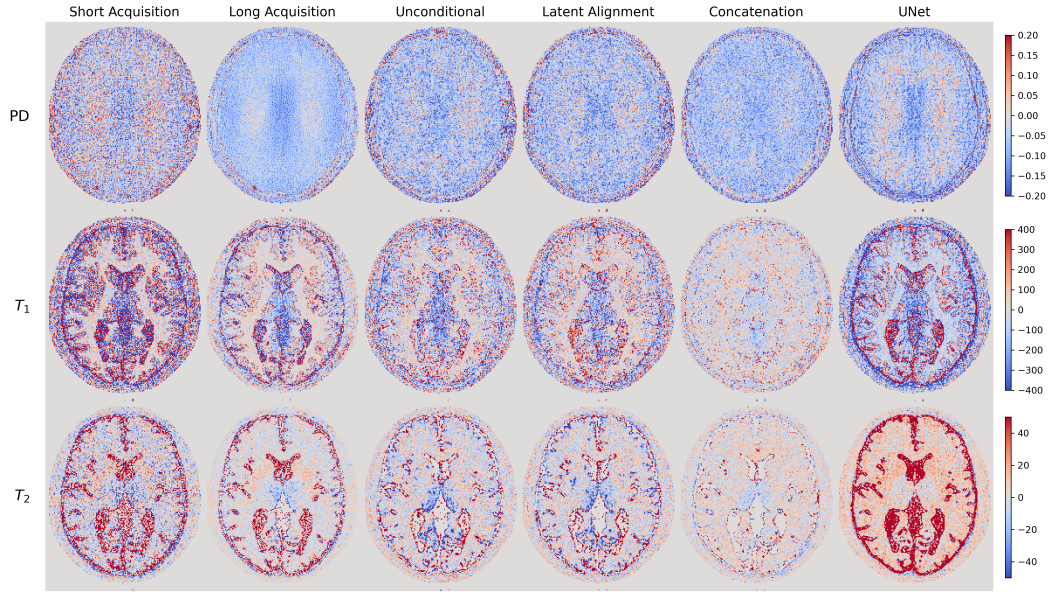

Figure 5.7: Difference images of the same slice and synthetic test subject as in Figure 5.6, showing the PD, T_1, and T_2 quantitative maps. Each column presents the pixel-wise difference between the reference map and the map generated using the images from the corresponding acquisition or model indicated in the column title. T_1 and T_2 maps are in milliseconds, and PD is in arbitrary units.

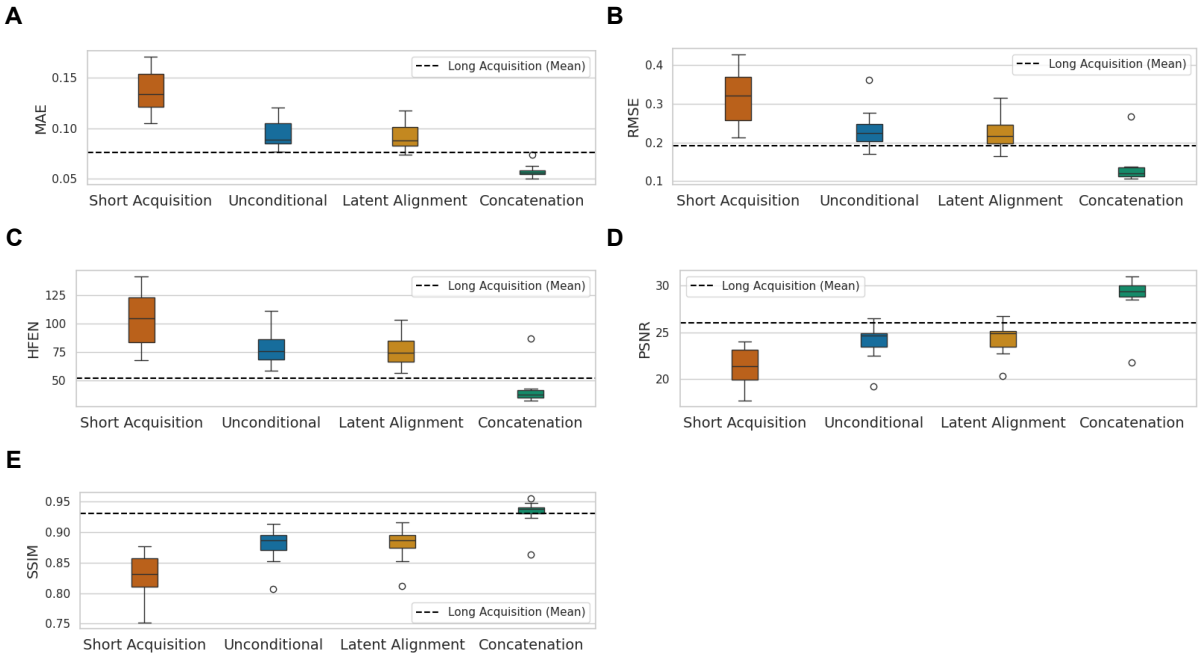

Figure 5.8: Boxplots of five image quality metrics (A–E) for T_1 maps of ten test subjects reconstructed from the short-acquisition contrast-weighted images, the result of the unconditional partial DDPM, the result of the partial DDPM with latent alignment, and the result of the concatenation-based partial DDPM. A dashed line marks the mean for long-acquisition maps. Metrics are computed relative to the reference image while using a brain mask. Subplots show: (A) MAE, (B) RMSE, (C) HFEN, (D) PSNR, and (E) SSIM.

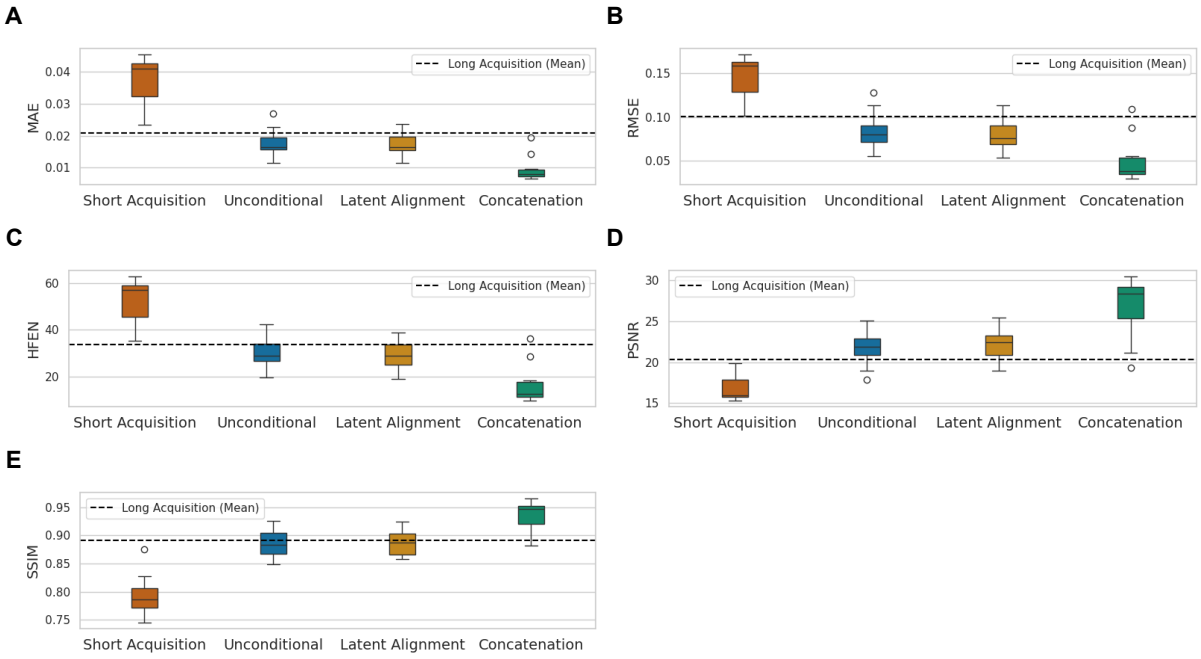

Figure 5.9: Boxplots of five image quality metrics (A–E) for T_2 maps of ten test subjects reconstructed from the short-acquisition contrast-weighted images, the result of the unconditional partial DDPM, the result of the partial DDPM with latent alignment, and the result of the concatenation-based partial DDPM. The mean metric value for long-acquisition maps is depicted as a dashed line. Metrics are computed relative to the reference image while using a brain mask. Subplots show: (A) MAE, (B) RMSE, (C) HFEN, (D) PSNR, and (E) SSIM.

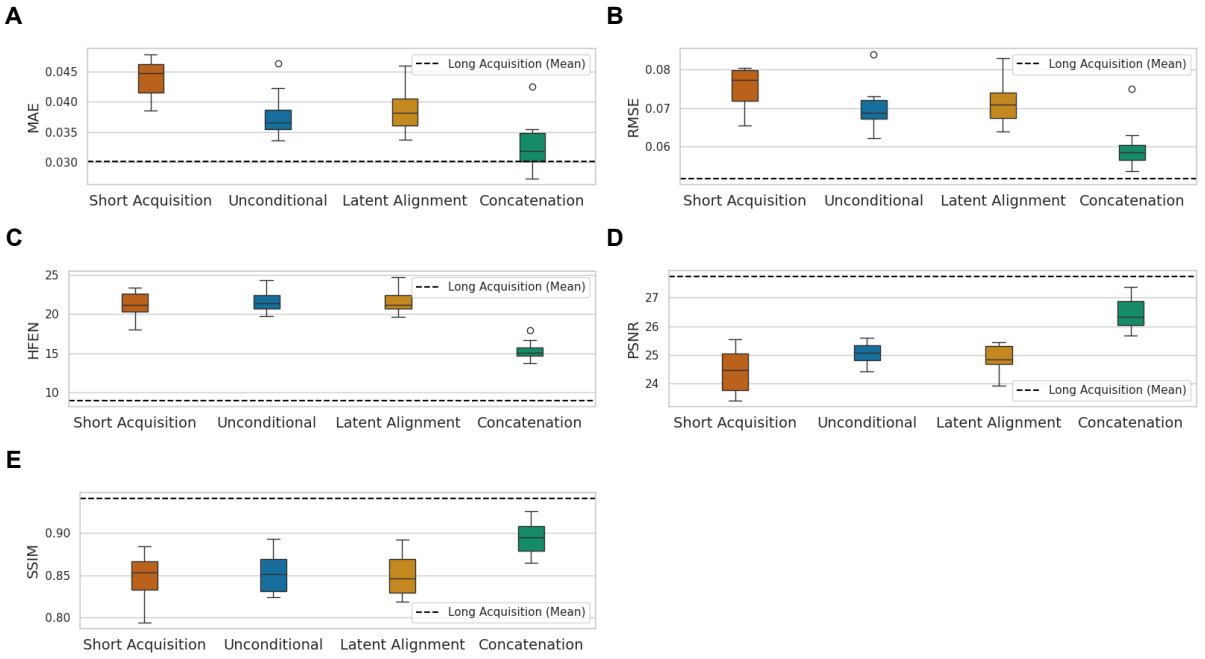

Figure 5.10: Boxplots of five image quality metrics (A–E) for PD maps of ten test subjects reconstructed from the short-acquisition contrast-weighted images, the result of the unconditional partial DDPM, the result of the partial DDPM with latent alignment, and the result of the concatenation-based partial DDPM. The mean metric value for long-acquisition maps is depicted as a dashed line. Metrics are computed relative to the reference image while using a brain mask. Subplots show: (A) MAE, (B) RMSE, (C) HFEN, (D) PSNR, and (E) SSIM.

5.2. Results on In Vivo Dataset

5.2.1. Contrast-Weighted Images Comparison

The performance of the models was also evaluated on an in vivo subject. In Figure 5.11, a representative slice for the short-acquisition in vivo contrast-weighted images, the long-acquisition contrast-weighted image, from which the short-acquisition images were created, the outputs of the three models, and the UNet output is displayed. As a reference image, an additionally acquired long-acquisition of the same volunteer is displayed.

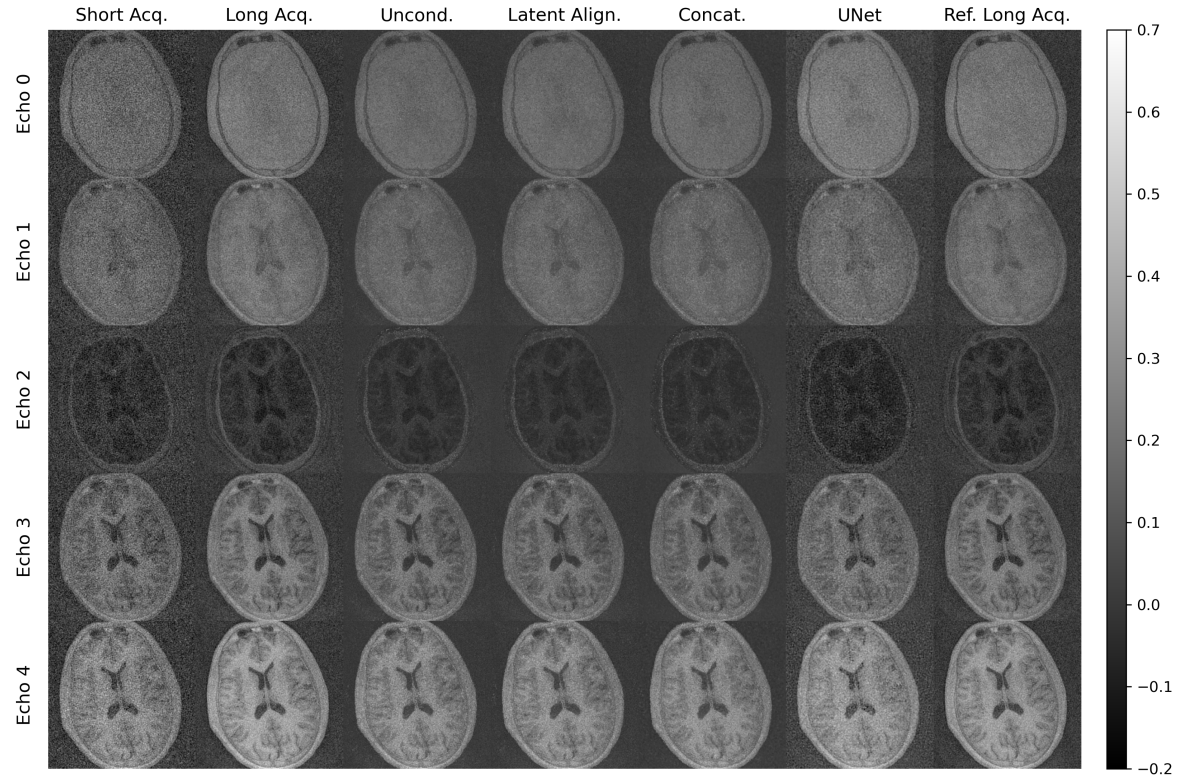

Figure 5.11: Contrast-weighted images for the in vivo subject (single slice shown). The first column displays the five short-acquisition input echoes provided to all models. The second column displays the corresponding long-acquisition echoes, while the last column shows the reference long-acquisition echoes. The third, fourth, and fifth columns show the output of the unconditional partial DDPM, the latent alignment variant, and the concatenation-based model, respectively. The sixth column displays the output of a basic UNet.

Figure 5.11 shows less noise in the resulting brain models of all three partial DDPM variants compared to the short-acquisition input echoes. The background in the partial DDPM outputs is almost completely denoised, even compared to both the long-acquisition and reference long-acquisition reference echoes. However, the brain structures in Echoes 4 and 5 are less sharp in the partial DDPM outputs than in the long-acquisition and the reference long-acquisition images. The UNet model output has less noise compared to the short-acquisition input, but significantly more noise than the partial DDPM outputs. Figure 5.12 further supports these observations by demonstrating the reduced residual error in the outputs of the partial DDPM models compared to the short-acquisition input. Similar to the results on the synthetic dataset, the model outputs show the least residual error in Echo 2 compared to the other echoes.

Image quality metrics were computed for these outputs with respect to the reference long-acquisition images. As before, calculations were performed within a brain mask. Table 5.1 displays the mean values across the five echoes for one in vivo subject. For comparison, the same metrics were also computed for the short- and long-acquisition in vivo images relative to the same reference. Table 5.1 shows that all image quality metrics improve for all three partial DDPMs compared to the short-acquisition input data. Additionally, all partial DDPM metrics are similar to those of the long-acquisition image. Interestingly, all three partial DDPMs achieve similar metrics.

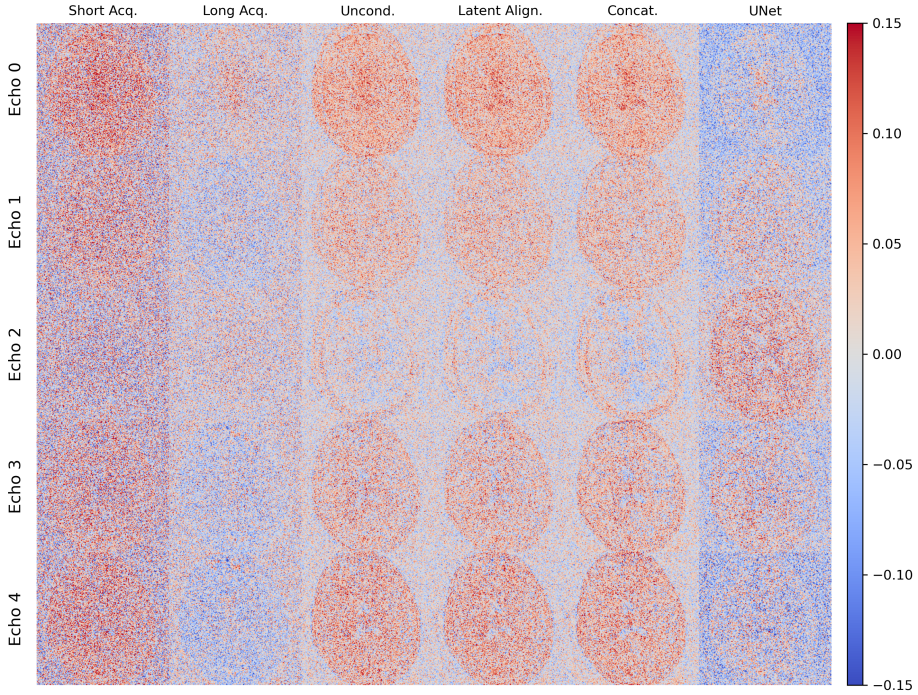

Figure 5.12: Difference images for the in vivo subject in Figure 5.11 (single slice shown). Each column presents the pixel-wise difference between the reference long-acquisition images and the images from the corresponding acquisition or model indicated in the column title.

Table 5.1: Mean Image Quality Metrics across Echoes for Contrast-Weighted Images of an In Vivo Subject, Computed with Respect to the Reference Long Acquisition

	Short Acq.	Unconditional	Latent Alignment	Concatenation	Long Acq.
MAE	0.039	0.028	0.028	0.029	0.028
RMSE	0.065	0.048	0.047	0.048	0.047
HFEN	21.6	16.2	16.1	16.8	16.7
PSNR	27.07	29.76	29.85	29.63	29.80
SSIM	0.62	0.71	0.71	0.70	0.72

5.2.2. Quantitative Maps Comparison

Similar to the synthetic evaluation, quantitative maps were reconstructed from the enhanced contrast-weighted images of the in vivo subject. Figure 5.13 demonstrates that all three partial DDPMs produce quantitative maps with more clearly defined edges than the short-acquisition maps. Figure 5.14 shows the corresponding difference images between the reference maps and the corresponding maps indicated by the column titles. Especially for the T_2 maps, the brain structures are barely visible in the short-acquisition T_2 map but are successfully recovered in all three models. The enhanced T_2 maps show substantially improved tissue contrast and structural definition, approaching the quality of the reference long-acquisition. However, the enhanced T_2 maps seem to not properly recover the CSF layer between the brain and the skull for all three models. Figure 5.14 shows that there is a large positive residual error in the CSF layer for all three models.

For the T_1 maps, all partial DDPM variants show reduced noise compared to the short-acquisition input and remain in the same intensity scaling as the long-acquisition map. However, they show a consistent negative bias relative to the reference long-acquisition map, and the model outputs also appear noisier than the long-acquisition T_1 map. Interestingly, out of the three models, the concatenation-based model shows the highest noise levels within the brain tissue, with some edge blurring visible.

The PD maps are more difficult to evaluate visually, as the long-acquisition and reference long-acquisition PD maps do not show the same contrast. The difference images in Figure 5.14 show that the residual error decreases for all three models, but more residual error is left compared to the long-acquisition maps.

In contrast to the partial DDPM outputs, the UNet results show significantly higher noise levels across all parameter maps and also contain a pronounced mean intensity shift in the T_1 map compared to the reference long-acquisition map.

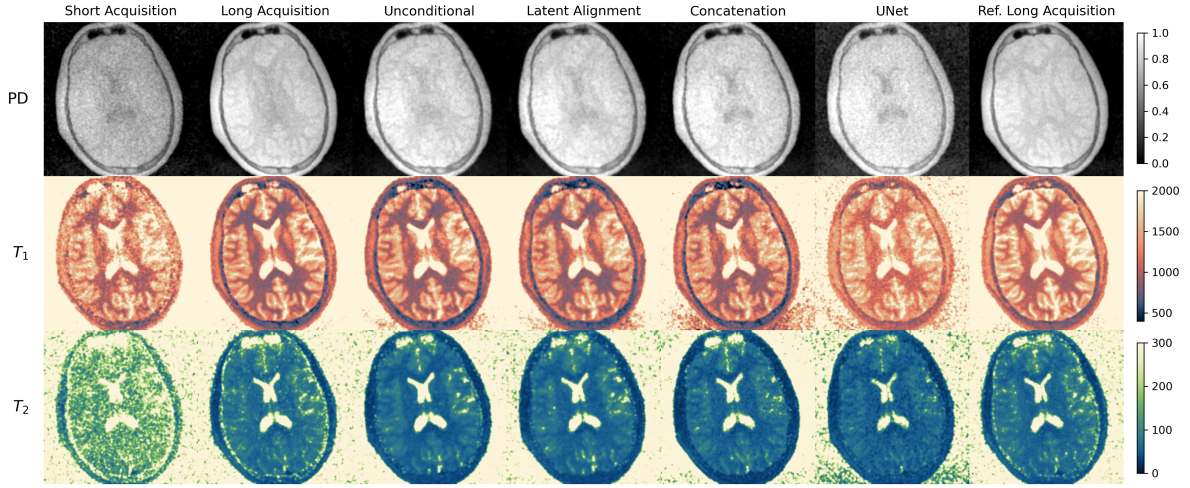

Figure 5.13: Single slice of the PD, T_1, and T_2 maps for an in vivo subject reconstructed from the short-acquisition contrast-weighted images, long-acquisition contrast-weighted images, the result of the unconditional partial DDPM, the result of the partial DDPM with latent alignment, the result of the concatenation-based partial DDPM, the result of the UNet, and a reference long acquisition. Gaussian smoothing was applied to all images using $\sigma = 0.9$. T_1 and T_2 maps are in milliseconds, and PD is in arbitrary units.

For each quantitative map, image quality metrics were computed individually with respect to the relevant reference long-acquisition map. The metrics for the PD, T_1, and T_2 maps are presented in Table 5.2. The metrics were calculated using a brain mask. All three models demonstrate improved image quality metrics across all three quantitative maps compared to the short-acquisition maps. The T_1 and PD map metrics achieved by the models are comparable to those of the long-acquisition maps, while the T_2 map metrics are slightly improved from those of the long-acquisition map. All three models achieve very similar performance across the different quantitative maps. It is also worth noting that all image quality metrics indicate a substantially larger mismatch between the maps and the reference long-acquisition maps compared to the results obtained for the synthetic dataset, even for the short-acquisition and first long-acquisition in vivo maps.

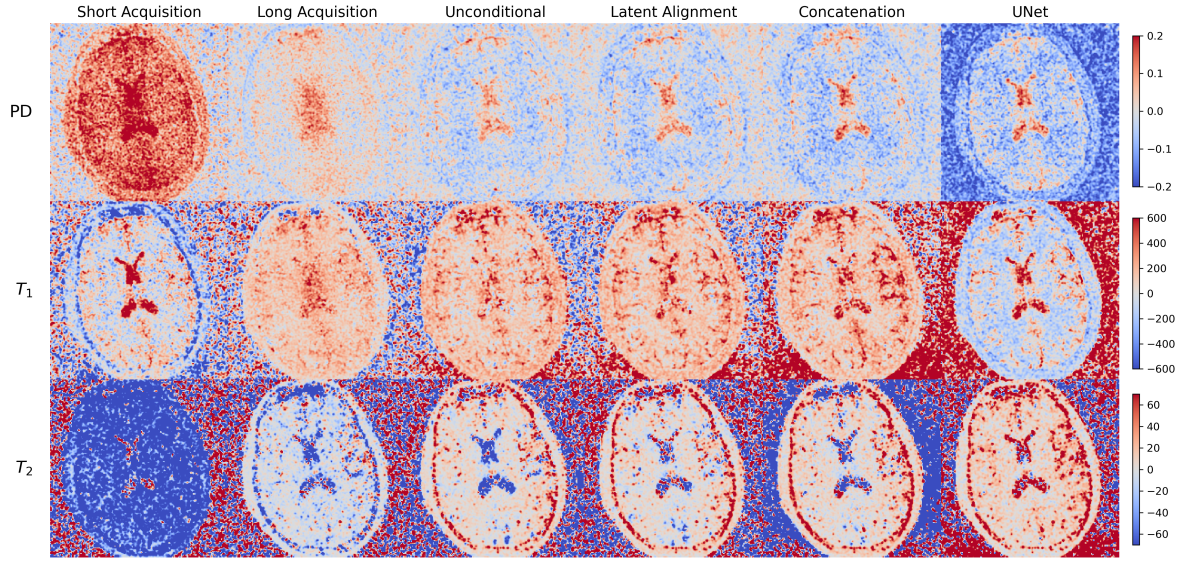

Figure 5.14: Difference images of the same slice of the in vivo subject as in Figure 5.6, showing the PD, T_1, and T_2 quantitative maps. Each column presents the pixel-wise difference between the reference long-acquisition map and the map generated using the images from the corresponding acquisition or model indicated in the column title. T_1 and T_2 maps are in milliseconds, and PD is in arbitrary units.

Table 5.2: Mean Image Quality Metrics for PD, T_1 and T_2 Maps of an In Vivo Subject, Computed with Respect to the Reference Long Acquisition

		Short Acq.	Unconditional	Latent Align.	Concatenation	Long Acq.
PD	**MAE**	0.160	0.103	0.102	0.102	0.104
	RMSE	0.264	0.171	0.169	0.169	0.173
	HFEN	84.1	59.9	59.0	61.9	58.1
	PSNR	23.91	27.67	27.79	27.76	27.60
	SSIM	0.54	0.64	0.65	0.65	0.66
T_1	**MAE**	0.499	0.360	0.357	0.374	0.362
	RMSE	0.941	0.745	0.746	0.765	0.752
	HFEN	305.8	243.2	242.8	264.5	241.6
	PSNR	12.57	14.60	14.59	14.36	14.52
	SSIM	0.48	0.56	0.56	0.54	0.57
T_2	**MAE**	0.116	0.065	0.064	0.068	0.073
	RMSE	0.297	0.209	0.208	0.218	0.225
	HFEN	91.6	65.6	65.4	75.5	71.0
	PSNR	10.55	13.61	13.64	13.24	12.95
	SSIM	0.37	0.46	0.45	0.46	0.44

6

Discussion

In this chapter, we interpret the findings presented in Chapter 5 and evaluate their implications for accelerated quantitative MRI. This thesis investigated three partial DDPM variants for enhancing contrast-weighted images acquired with a four times accelerated MuPa-ZTE sequence, to enable 1.5-minute quantitative MRI acquisitions while maintaining image quality. Additionally, this thesis investigated whether these enhanced contrast-weighted images directly translate to improved quantitative maps. The concatenation-based partial DDPM consistently achieved the highest performance across the synthetic dataset. While the unconditional and latent alignment variants also improved the short-acquisition images, they underperformed relative to the concatenation-based model. Importantly, all three models successfully transferred to in vivo data, although the inter-model comparison was more challenging due to the lack of a ground truth. The enhanced contrast-weighted images directly translated to improved quantitative maps, with T_1 and T_2 maps showing even more relative enhancement than the contrast-weighted images they were generated from. PD maps are also enhanced from their short-acquisition counterparts, but to a lesser extent than T_1 and T_2 maps. This thesis demonstrates that partial DDPMs could enable significant acceleration of the MuPa-ZTE sequence, potentially reducing acquisition times to one minute, while preserving image quality. In this discussion, we explore these research findings within the broader context of accelerated MRI, address the limitations, and identify promising directions for further research.

6.1. Interpretation of Results

6.1.1. Effect of the number of reverse diffusion steps

The ablation study on the number of reverse diffusion steps, K, showed that both the unconditional and latent alignment variants achieve their best performance at a relatively low number of diffusion steps ($K = 10$). This outcome can be explained by looking at the nature of the diffusion process. With a large number of steps, the short-acquisition input image undergoes extensive forward diffusion, progressively degrading the available prior information. Consequently, the model starts with a heavily degraded input, which makes it challenging to recover the correct high-quality output. However, to effectively leverage the power of a diffusion model, a sufficient number of steps is required. This introduces a trade-off between preserving prior information while also exploiting maximum denoising capacity. Based on this balance, we selected $K = 10$ as the optimal value for these two models. The concatenation-based partial DDPM does not suffer from this limitation, as it maintains direct access to the original, undiffused input image through its architecture. By concatenating the untouched input with the diffused images, it benefits from both the available prior information of the original data as well as the denoising capacity of the diffusion process. This explains why its performance continued to improve with increasing K. However, the rate of performance improvement plateaued beyond a certain point, likely because additional denoising steps offer limited additional performance once most of the recoverable information has already been reconstructed. As an increased number of reverse diffusion steps leads to increased computational complexity, a trade-off has to be made between maximum performance and minimized inference time. To address this, the value for K that resulted in performance within 1% of the maximum

performance for all evaluated metrics was selected. For the concatenation-based model this occurred at $K = 50$.

6.1.2. Model Performance on Synthetic Data

The experiments on synthetic contrast-weighted images demonstrated that the concatenation-based partial DDPM outperformed both the unconditional and latent alignment variant across all evaluated metrics. This model's direct access to the original multi-echo input through concatenation enhances detail preservation and reduces high-frequency noise. In contrast, the unconditional and latent alignment variants improved upon the short-acquisition input but consistently underperformed relative to the concatenation-based model. However, the unconditional and latent alignment models did perform similarly to each other, with differences <1% across all metrics. This similarity in performance suggests that the linear latent space alignment approach does not provide meaningful benefits compared to the unconditional partial DDPM. The underlying issue might be the fundamental mismatch between the training objective and the desired transformation. The loss function used in training the models optimizes the noise prediction for both the short-acquisition and reference images, rather than explicitly learning the mapping between their respective data distributions. This could restrict the model's ability to properly benefit from including the short-acquisition input data in the training process. While previous work showed modest improvements with latent alignment, these improvements were limited to a small value for K and were not observed in our experiments even at comparable K values, potentially due to the different undersampling patterns employed [24]. Further research could explore whether modifying the loss function to explicitly penalize a distribution mismatch, allows for learning the relationship between the short-acquisition and reference data better.

6.1.3. Model Performance on In Vivo Data

The experiments on in vivo contrast-weighted images demonstrated that all three partial DDPM variants improved the quality of the short-acquisition MuPa-ZTE images, achieving image quality metrics that closely approached those of a long acquisition. These results provide initial evidence that the model performance generalizes to in vivo data, although further validation across a diverse set of in vivo subjects is required. A critical limitation of the in vivo evaluation is the lack of a ground truth. The reference long acquisition inherently contains noise and reconstruction artifacts that prevent absolute performance measurements. Consequently, the calculated metrics for the in vivo data primarily indicate relative improvement over the short-acquisition input, rather than absolute reconstruction accuracy.

The challenge of establishing a reliable ground truth is illustrated by the relatively low inter-acquisition similarity between the two long in vivo acquisitions (SSIM: 0.72), compared to the similarity observed in synthetic data (SSIM: 0.93 between the long-acquisition and reference images). This discrepancy suggests either a significant noise component or some subtle motion between the acquisitions and further emphasizes the limitations of using one imperfect acquisition as a reference for another. Nevertheless, the observed improvements are not limited to quantitative metrics: they are also evident in the visual inspection of the contrast-weighted images of both the primary in vivo dataset (Figure 5.11) and the secondary in vivo dataset (Appendix A), giving qualitative support for the relative enhancement achieved by the partial DDPMs. Notably, the models demonstrated robustness to acquisition parameter variations, as the primary in vivo dataset used fewer k-space cycles (27 vs 34 in training) for the short-acquisition images, yet still achieved image quality similar to the long-acquisition images. This suggests that the partial DDPMs have learned generalizable features of the MuPa-ZTE sequence. For the unconditional partial DDPM, this demonstrates its robustness to increased undersampling factors, while for the latent alignment and concatenation-based variants, it shows successful generalization beyond the specific undersampling factor they were trained on.

Another important insight from the in vivo results is that models trained exclusively on synthetic data successfully transferred to real acquisitions. This demonstrates the effectiveness of synthetic MuPa-ZTE data in capturing the underlying signal properties of necessary for generalization to real acquisitions. The use of synthetic datasets provides significant practical advantages for model development. Synthetic datasets can consist of virtually unlimited and controllable training samples. This can be used to systematically improve model performance and robustness without the need for extensive in vivo data collection. For instance, if the MuPa-ZTE acquisition parameters are modified, generating new synthetic data that reflect these changes and retraining the model accordingly, is a straightforward, fast,

procedure that can be completed within days. In contrast, acquiring a sufficient number of in vivo scans with a new protocol for retraining would require months of patient recruitment and scanning.

6.1.4. Quantitative Maps Enhancement

To test whether the models are useful clinically, the next objective was to determine whether the contrast-weighted image enhancements translated into enhanced quantitative parameter maps directly, the primary output of the MuPa-ZTE sequence. For synthetic test data, the concatenation-based model recovers T_1 and T_2 maps that closely resemble the reference maps. Remarkably, even the underperforming models in contrast-weighted image enhancement, the unconditional partial DDPM, and latent alignment variant, achieve image quality metrics comparable to those of the long-acquisition maps for the T_1 maps, and even exceed the long-acquisition metrics for the T_2 maps.

The increased performance in quantitative mapping, exceeding the improvement observed in contrast-weighted images, likely stems from the fact that the model enhances all five echoes simultaneously. By processing echoes jointly, the model can exploit temporal correlations in the signal evolution, which forms the basis for T_1 and T_2 parameter estimation. Specifically, the model might improve the temporal consistency of the signal evolution across the echoes more than it improves the image quality of the individual echoes, resulting in more accurate parameter maps even though individual echoes show less enhancement.

The PD maps, however, showed consistently lower image quality across the three models compared to the long-acquisition reference. PD maps reflect the underlying proton density scaling factor that determines voxel-wise signal magnitude, and therefore depend significantly on preserving absolute signal intensities. While the models successfully reconstruct the temporal signal evolution, necessary for T_1 and T_2 estimation, the absolute intensity scaling required for the PD maps seems to be altered. This limitation suggests that for accurate PD mapping, additional constraints or normalization might be required to ensure accurate intensity scaling across voxels.

The in vivo quantitative maps demonstrate that contrast-weighted image enhancement also translates to improved parameter estimation in real clinical data. However, the absence of a ground truth severely limits a thorough quantitative evaluation of the performances. Unlike synthetic data where the ground truth is known, in vivo validation can only show relative improvements over short-acquisition inputs. The observed tissue contrast differences in PD maps between model variants emphasize the need for further validation across additional in vivo subjects.

6.1.5. Comparison against UNet baseline

In comparison to the partial DDPMs, the basic UNet baseline produces visibly more blurred output contrast-weighted images for both the synthetic images as the in vivo subject, a known limitation of convolutional neural networks (CNN) trained with a mean squared error (MSE) loss. Single-pass CNNs tend to oversmooth fine structures due to local averaging effects and the MSE loss function does not penalize this blurring. The partial DDPM does not suffer from this limitation, as the iterative denoising process allows for progressive refinement of image details. Additionally, the loss function in the partial DDPM is focused on predicting the noise, rather than the output image directly. This allows the model to recover high-frequency information over multiple steps, preserving sharper edges and small details. These results demonstrate the improved capability of partial DDPMs compared to standard UNet architectures for enhancing undersampled MuPa-ZTE images.

6.2. Limitations

While the results presented in this thesis demonstrate the potential of partial DDPMs for enhancing MuPa-ZTE contrast-weighted images and their corresponding quantitative maps, several limitations should be acknowledged to contextualize these findings and guide future research.

One of the main limitations of this thesis is the limited number of test data available. The synthetic test set, while varied in T_1, T_2, and PD values, was limited to ten subjects with identical brain anatomy derived from a single brain model. This lack of variation in test subject anatomy prevents the thorough evaluation of the model across different brain morphologies that could be encountered in clinic. Therefore, it is unclear how the models would perform on scans from patients with significantly different brain

structures, such as lesions or large brain tumors.

One in vivo subject was used to test the generalizability of the model to data outside of the training data distribution. Although this provided promising results, a more thorough evaluation on multiple in vivo subjects is necessary. One subject does not allow for any statistical assessment of performance variability and thus limits the confidence in the reported overall performance.

Another limitation of this thesis is the simplifications incorporated in the synthetic data generation. Most importantly, the current pipeline assumes each voxel contains a single tissue type. In reality, however, voxels often contain a mixture of tissue types, particularly at tissue boundaries. The current approach ignores these partial volume effects, even though they can significantly influence the signal evolution captured. Therefore, the partial DDPMs are never exposed to such effects in training and are likely to underperform in regions where these effects are very prominent. Additionally, the synthetic data does not fully capture the complexities of real MRI artifacts including B_0 inhomogeneities, RF field variation or patient motion. This could limit model robustness when these effects become very pronounced during acquisition.

Lastly, in training, a patch-based approach was applied to avoid memory constraints. However, this approach limits the model's access to global anatomical context, which might affect performance in larger structural variations as seen in abnormal brain structures. The current patch size restricts the field of view available to the network and therefore limits the amount of information the model can use to improve the quality of the image. This could be problematic for patients with atypical anatomy.

6.3. Clinical Impact

This thesis has shown that using partial DDPMs allows for significantly enhancing the quality of undersampled contrast-weighted images and more importantly the quality of the quantitative maps that follow from them. In the broader context, this implies that partial DDPMs could lead to multiparametric quantitative MR acquisitions being performed in 1.5 minutes or less, while still producing high-quality quantitative maps. Such a reduction in scan time would have a significant impact in both clinical and research settings. Faster scans would increase patient comfort and allow for shorter waiting lists for MRI appointments. Additionally, these high-quality quantitative maps acquired in 1.5 minutes could be used to detect biomarkers for specific diseases and could allow for more precise disease monitoring. As accurate quantitative maps are only dependent on tissue properties, they facilitate standardized comparisons across scanners and patients, improving the reproducibility and reliability of longitudinal studies. Lastly, once the quantitative maps are acquired, any type of conventional contrast-weighted image can be synthetically generated using the appropriate forward models. This could allow radiologists to tailor the image contrasts to specific diagnostic needs without requiring additional scan time, allowing for more flexible and personal workflows.

6.4. Future Work

To achieve the full potential of partial DDPMs, the current models should be validated on more in vivo data, including pathologies, to properly evaluate performance in a clinical setting. Additionally, the models should be validated on a phantom with known T_1 and T_2 values, to assess the accuracy of the quantitative maps reconstructed from the enhanced contrast-weighted images.

To further improve the current models, we propose two directions of further research: increasing patch size in training and exploring alternative ways to incorporate data consistency. First, increasing the patch size could provide the model with a larger spatial context, potentially improving reconstruction quality. Although a larger patch size would require a smaller batch size due to memory constraints, this trade-off might be worthwhile if it leads to improved performance. Secondly, the current best performing model, the concatenation-based partial DDPM, incorporates data consistency in a relatively limited way. Future work could explore alternative approaches that enforce the data consistency more strongly to ensure the enhanced images adhere to the measured data, while still benefiting from the generative power of the partial DDPM.

Another potential advantage of diffusion models, not explored in this thesis, is their inherent stochasticity, which allows for generating multiple outputs from the same input. This property could be used to

estimate model uncertainty, which could inform radiologist on the reliability of the output and therefore support them in clinical decision-making. Additionally, generating multiple outputs and averaging them could improve performance by reducing stochastic variance in the final result.

Next to these validations and improvements, the current performance of the model suggests that even higher undersampling factors might be feasible. Future work should train models on progressively undersampled synthetic images to establish the limits of partial DDPM reconstruction. Specifically, it should determine the maximum undersampling factor beyond which partial DDPMs can no longer reliably recover long-acquisition contrast-weighted image quality.

7

Conclusion

This thesis investigated the potential of partial diffusion models to enhance contrast-weighted images acquired with a four times undersampled MuPa-ZTE sequence and addressed four key research questions.

First, we investigated whether partial diffusion models can improve the image quality of the contrast-weighted images acquired by the 1.5-minute MuPa-ZTE acquisition to reach that of the non-accelerated acquisition images. The results demonstrated that all three partial DDPM variations are able to improve the quality of the short-acquisition images and that it is possible to reach the quality of the non-accelerated acquisition images using a partial DDPM.

Secondly, the influence of the different partial DDPM variations was assessed. The concatenation-based partial DDPM outperformed unconditional and latent alignment variations consistently across quality measures for the synthetic test set on the contrast-weighted images. This showed that direct access to undiffused input data is beneficial for model performance. Inter-model comparisons for the in vivo data were not possible due to the absence of a ground truth. Here, all three models showed similar performance.

Thirdly, we compared the models to a baseline UNet model. Our qualitative evaluations showed that the partial DDPMs outperformed the UNet model. The models were particularly superior in preserving fine details and sharp edges. The iterative denoising process in the partial DDPMs thus enables superior high-frequency information recovery.

Lastly, we evaluated the quality of the quantitative PD, T_1, and T_2 maps generated from the enhanced contrast-weighted images. The results showed that enhanced contrast-weighted images directly translated into improved T_1 and T_2 maps, with the improvement in the maps exceeding that observed in the individual input images. This suggests that joint enhancement across contrasts better restores the signal evolution, amplifying the performance in the quantitative mapping domain beyond what is seen in the image domain alone. PD maps showed more limited, but still measurable improvement.

This thesis demonstrates that partial diffusion models can significantly enhance the quality of contrast-weighted images acquired with a four times undersampled MuPa-ZTE sequence while preserving image quality. The concatenation-based partial DDPM showed the highest performance across the image quality metrics. These findings suggest that partial DDPMs could be a promising solution for achieving one-minute quantitative MRI acquisitions, which could have a substantial impact on patient comfort and clinical efficiency.

By demonstrating that high-quality 3D multiparametric quantitative brain MRI could be achieved in 1.5 minutes, this work brings the vision of clinical quantitative MRI one step closer.

References

1. Bloem JL, Reijnierse M, Huizinga TWJ, and Helm-van Mil AHM van der. MR signal intensity: staying on the bright side in MR image interpretation. RMD Open. 2018 Jun 20; 4:e000728. DOI: 10.1136/rmdopen-2018-000728
2. Gracien RM, Maiworm M, Brüche N, Shrestha M, Nöth U, Hattingen E, Wagner M, and Deichmann R. How stable is quantitative MRI? – Assessment of intra- and inter-scanner-model reproducibility using identical acquisition sequences and data analysis programs. NeuroImage. 2020 Feb 15; 207:116364. DOI: 10.1016/j.neuroimage.2019.116364
3. Hubbard Cristinacce PL, Keaveney S, Aboagye EO, Hall MG, Little RA, O'Connor JPB, Parker GJM, Waterton JC, Winfield JM, and Jauregui-Osoro M. Clinical translation of quantitative magnetic resonance imaging biomarkers – An overview and gap analysis of current practice. Physica Medica. 2022 Sep 1; 101:165–82. DOI: 10.1016/j.ejmp.2022.08.015
4. Seiler A, Nöth U, Hok P, Reiländer A, Maiworm M, Baudrexel S, Meuth S, Rosenow F, Steinmetz H, Wagner M, Hattingen E, Deichmann R, and Gracien RM. Multiparametric Quantitative MRI in Neurological Diseases. Frontiers in Neurology. 2021 Mar 8; 12. DOI: 10.3389/fneur.2021.640239
5. Ma D, Gulani V, Seiberlich N, Liu K, Sunshine JL, Duerk JL, and Griswold MA. Magnetic resonance fingerprinting. Nature. 2013 Mar; 495:187–92. DOI: 10.1038/nature11971
6. Fujita S, Hagiwara A, Hori M, Warntjes M, Kamagata K, Fukunaga I, Andica C, Maekawa T, Irie R, Takemura MY, Kumamaru KK, Wada A, Suzuki M, Ozaki Y, Abe O, and Aoki S. Three-dimensional high-resolution simultaneous quantitative mapping of the whole brain with 3D-QALAS: An accuracy and repeatability study. Magnetic Resonance Imaging. 2019 Nov; 63:235–43. DOI: 10.1016/j.mri.2019.08.031
7. Gómez PA, Cencini M, Golbabaee M, Schulte RF, Pirkl C, Horvath I, Fallo G, Peretti L, Tosetti M, Menze BH, and Buonincontri G. Rapid three-dimensional multiparametric MRI with quantitative transient-state imaging. Scientific Reports. 2020 Aug 13; 10:13769. DOI: 10.1038/s41598-020-70789-2
8. Ljungberg E, Damestani N, Wood T, Lythgoe DJ, Zelaya F, Williams SCR, Solana AB, Barker GJ, and Wiesinger F. Silent zero TE MR neuroimaging: Current state-of-the-art and future directions. Progress in Nuclear Magnetic Resonance Spectroscopy. 2021 Apr 1; 123:73–93. DOI: 10.1016/j.pnmrs.2021.03.002
9. Ljungberg E, Wood T, Solana AB, Barker G, and Wiesinger F. Silent Motion-Corrected ZTE for Quantitative T1 and T2 mapping. Joint Annual Meeting ISMRM-ESMRMB ISMRT 31st Annual Meeting. London, England, UK, 2022 :0351. DOI: 10.58530/2022/0351
10. Wiesinger F, McKinnon G, Kaushik S, Solana AB, Ljungberg E, Vogel M, Takei N, Schulte R, Pirkl C, Cozzini C, Nuñez-Gonzalez L, Hernandez Tamames JA, and Engström M. 3D Silent Parameter Mapping: Further refinements & quantitative assessment. Vol. 29. Proceedings of the International Society Magnetic Resonance in Medicine, 2021 :1828
11. Sarvamangala DR and Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evolutionary Intelligence. 2022 Mar 1; 15:1–22. DOI: 10.1007/s12065-020-00540-3
12. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, and Fu H. Transformers in medical imaging: A survey. Medical Image Analysis. 2023 Aug 1; 88:102802. DOI: 10.1016/j.media.2023.102802

13. Oulmalme C, Nakouri H, and Jaafar F. A systematic review of generative AI approaches for medical image enhancement: Comparing GANs, transformers, and diffusion models. International Journal of Medical Informatics. 2025 Jul 1; 199:105903. DOI: 10.1016/j.ijmedinf.2025.105903

14. Kazerouni A, Aghdam EK, Heidari M, Azad R, Fayyaz M, Hacihaliloglu I, and Merhof D. Diffusion models in medical imaging: A comprehensive survey. Medical Image Analysis. 2023 Aug 1; 88:102846. DOI: 10.1016/j.media.2023.102846

15. Zhao K, Hung ALY, Pang K, Zheng H, and Sung K. PartDiff: Image Super-resolution with Partial Diffusion Models. 2023 Jul 21. DOI: 10.48550/arXiv.2307.11926. arXiv: 2307.11926[eess]

16. Chen Y, Almarzouqi SJ, Morgan ML, and Lee AG. T1-Weighted Image. *Encyclopedia of Ophthalmology*. Ed. by Schmidt-Erfurth U and Kohnen T. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018 :1747–50. DOI: 10.1007/978-3-540-69000-9_1228

17. Chen Y, Almarzouqi SJ, Morgan ML, and Lee AG. T2-Weighted Image. *Encyclopedia of Ophthalmology*. Ed. by Schmidt-Erfurth U and Kohnen T. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018 :1750–2. DOI: 10.1007/978-3-540-69000-9_1229

18. Geethanath S, Reddy R, Konar AS, Imam S, Sundaresan R, D. R. RB, and Venkatesan R. Compressed Sensing MRI: A Review. Critical Reviews in Biomedical Engineering. 2013; 41:183–204. DOI: 10.1615/CritRevBiomedEng.2014008058

19. Zimmermann M, Abbas Z, Dzieciol K, and Shah N. Accelerated Parameter Mapping of Multiple-Echo Gradient-Echo Data Using Model-Based Iterative Reconstruction. IEEE Trans. Med. Imaging 2018; 37:626–37. DOI: 10.1109/TMI.2017.2771504

20. Muslu Y, Tamada D, Roberts N, Cashen T, Mandava S, Kecskemeti S, Hernando D, and Reeder S. Free-breathing, fat-corrected T_1 mapping of the liver with stack-of-stars MRI, and joint estimation of T_1, PDFF, R_2^*, and B_1^+. Magn. Reson. Med. 2024; 92:1913–32. DOI: 10.1002/mrm.30182

21. Singhal V and Majumdar A. Reconstructing multi-echo magnetic resonance images via structured deep dictionary learning. Neurocomputing. 2020 Sep; 408:135–43. DOI: 10.1016/j.neucom.2019.11.107

22. Heydari A, Ahmadi A, Kim T, and Bilgic B. Joint MAPLE: Accelerated joint T_1 and T_2^* mapping with scan-specific self-supervised networks. Magn. Reson. Med. 2024; 91:2294–309. DOI: 10.1002/mrm.29989

23. Ho J, Jain A, and Abbeel P. Denoising Diffusion Probabilistic Models. 2020. arXiv: 2006.11239 [cs.LG]

24. Zhao K, Pang K, Hung ALY, Zheng H, Yan R, and Sung K. MRI Super-Resolution With Partial Diffusion Models. IEEE Transactions on Medical Imaging. 2025 Mar; 44:1194–205. DOI: 10.1109/TMI.2024.3483109

25. Sabidussi ER, Klein S, Caan MWA, Bazrafkan S, Den Dekker AJ, Sijbers J, Niessen WJ, and Poot DHJ. Recurrent inference machines as inverse problem solvers for MR relaxometry. Medical Image Analysis. 2021 Dec; 74:102220. DOI: 10.1016/j.media.2021.102220

26. Aubert-Broche B, Griffin M, Pike GB, Evans AC, and Collins DL. Twenty New Digital Brain Phantoms for Creation of Validation Image Data Bases. IEEE Transactions on Medical Imaging. 2006 Nov; 25:1410–6. DOI: 10.1109/TMI.2006.883453

27. Piccini D, Littmann A, Nielles-Vallespin S, and Zenge MO. Spiral phyllotaxis: The natural way to construct a 3D radial trajectory in MRI. Magnetic Resonance in Medicine. 2011 Apr 5; 66:1049–56. DOI: 10.1002/mrm.22898

28. Muckley MJ, Stern R, T M, and Knoll F. TorchKbNufft: A High-Level, Hardware-Agnostic Non-Uniform Fast Fourier Transform. *ISMRM Workshop on Data Sampling & Image Reconstruction*. Source code available at https://github.com/mmuckley/torchkbnufft. 2020

29. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, Murrey B, Myronenko A, Zhao C, Yang D, Nath V, He Y, Xu Z, Hatamizadeh A, Myronenko A, Zhu W, Liu Y, Zheng M, Tang Y, Yang I, Zephyr M, Hashemian B, Alle S, Darestani MZ, Budd C, Modat M, Vercauteren T, Wang G, Li Y, Hu Y, Fu Y, Gorman B, Johnson H, Genereaux B, Erdal BS, Gupta V, Diaz-Pinto A, Dourson A, Maier-Hein L, Jaeger PF, Baumgartner M, Kalpathy-Cramer J, Flores M, Kirby J, Cooper LAD, Roth HR, Xu D, Bericat D, Floca R, Zhou SK, Shuaib H, Farahani K, Maier-Hein KH, Aylward S, Dogra P, Ourselin S, and Feng A. MONAI: An open-source framework for deep learning in healthcare. 2022. arXiv: 2211.02701 [cs.LG]

30. Pinaya WHL, Graham MS, Kerfoot E, Tudosiu PD, Dafflon J, Fernandez V, Sanchez P, Wolleb J, Costa PFd, Patel A, Chung H, Zhao C, Peng W, Liu Z, Mei X, Lucena O, Ye JC, Tsaftaris SA, Dogra P, Feng A, Modat M, Nachev P, Ourselin S, and Cardoso MJ. Generative AI for Medical Imaging: extending the MONAI Framework. 2023 Jul 27. DOI: 10.48550/arXiv.2307.15208. arXiv: 2307.15208[eess]

31. Shafieizargar B, Byanju R, Sijbers J, Klein S, Dekker AJ den, and Poot DHJ. Systematic review of reconstruction techniques for accelerated quantitative MRI. Magnetic Resonance in Medicine. 2023; 90:1172–208. DOI: 10.1002/mrm.29721

32. Willmott CJ and Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research. 2005 Dec 19; 30:79–82. DOI: 10.3354/cr030079

33. Wang Z, Bovik A, Sheikh H, and Simoncelli E. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing. 2004 Apr; 13:600–12. DOI: 10.1109/TIP.2003.819861

34. Ravishankar S and Bresler Y. MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning. IEEE Transactions on Medical Imaging. 2011 May; 30:1028–41. DOI: 10.1109/TMI.2010.2090538

A

Extra Qualitative In Vivo Evaluation

An extra in vivo acquisition was available for qualitative evaluation. This acquisition used the exact same parameters as used in the forward model for generating the synthetic data. Figure A.1 shows the contrast-weighted images enhanced with the three different models for this extra in vivo dataset. Here, it is also clear that all three models are able to improve the image quality of the short-acquisition images and that the outputs resemble the long-acquisition images. The baseline UNet shows a less noisy, but blurred version of the short-acquisition images.

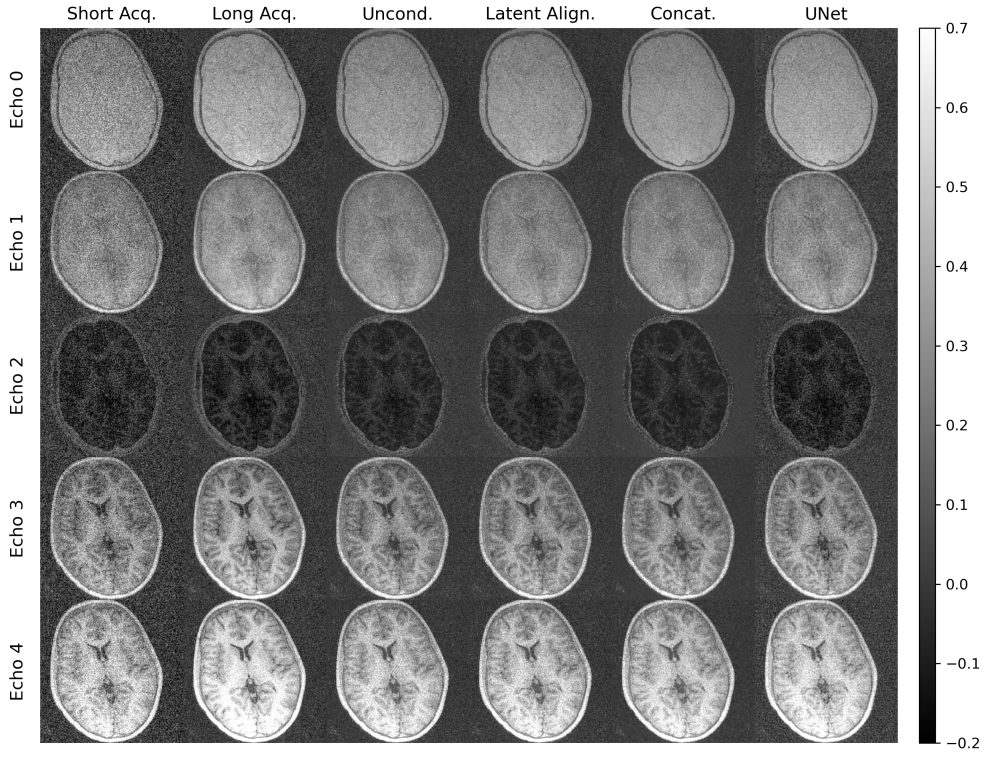

Figure A.1: Contrast-weighted images for an extra in vivo subject (single slice shown). The first column displays the five short-acquisition input echoes provided to all models and the second column displays the corresponding long-acquisition echoes. The third, fourth and fifth column show the output of the unconditional partial DDPM, the latent alignment variant and the concatenation-based model, respectively. The sixth column displays the output of a basic UNet. The rows display the five echoes associated with the MuPa-ZTE acquisition.

In Figure A.2 the quantitative maps generated from the contrast-weighted images in Figure A.1 are displayed. All three models decrease the amount of noise compared to the short-acquisition maps and also improve the contrast between the different tissues. The three models show similar results and the outputs resemble the long-acquisition maps. The outputs generated from the images enhanced with the baseline UNet also show improved quality compared to the short-acquisition maps, but both the T_1 and T_2 maps show a bias compared to the long-acquisition maps.

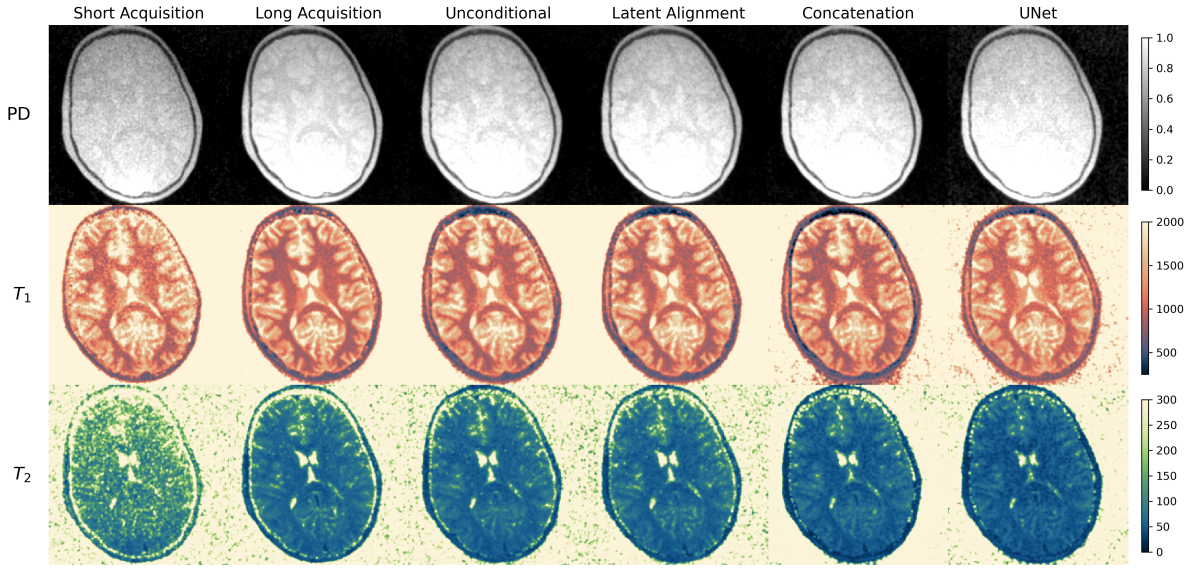

Figure A.2: Single slice of the PD, T_1 and T_2 maps for an extra in vivo subject reconstructed from (1) short-acquisition contrast-weighted images, (2) long-acquisition contrast-weighted images, (3), the result of the unconditional partial DDPM, (4) the result of the partial DDPM with latent alignment, (5) the result of the concatenation-based partial DDPM and (6) the result of the UNet. Gaussian smoothing was applied to all images using $\sigma = 0.9$. T_1 and T_2 maps are in milliseconds, PD is in arbitrary units.