Advanced Medical Image Reconstruction: Deep Generative and Functional Approaches

by

Imraj Ravi Devia Singh

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Department of Computer Science
University College London
United Kingdom

June 26, 2025

Declaration: *I, Imraj Ravi Devia Singh, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.*

To my sister, Sukhmani.

Abstract

Medical imaging is pivotal in diagnosis and treatment planning, with modalities such as Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) providing critical insights into physiological and anatomical structures. The process of reconstructing images from raw scanner measurements, known as medical image reconstruction, is essential for producing clinically useful images. This thesis develops advanced methods for medical image reconstruction using deep generative and functional approaches, focusing primarily on PET and MRI.

First, we developed a Deep Image Prior (DIP) method for PET image reconstruction. DIP is an untrained deep generative model that leverages the inductive bias of network architectures to regularise the reconstruction process. Applied to fully three-dimensional problems, the DIP method demonstrated strong performance in moderate noise settings but encountered challenges at higher noise levels. To address this, we incorporated additional traditional regularisation techniques to stabilise the reconstruction and improve performance compared to traditional methods alone.

The second study investigates normalisation techniques for supervised PET image reconstruction. These supervised methods directly predict the reconstruction from raw measurements or an approximate reconstruction. We proposed and compared a range of normalisation techniques to handle the widely varying dynamic range of PET images. The normalisation strategies improved clinically relevant image assessment methods as compared with unnormalised models.

In the third study, we adapted Score-based Generative Models (SGMs), a trained deep generative model, for PET image reconstruction. This method

utilised a learned prior, sampled while ensuring consistency with PET measurements, which necessitated careful handling of PET data-consistency. We also developed normalisation strategies and extensions for fully 3D imaging. For paired PET and MRI images, we developed a guided reconstruction method that enforced consistency with both PET measurements and the corresponding MRI image. The results were compared with state-of-the-art methods, demonstrating improved performance.

Finally, we developed Adaptable Blobs (A-Blobs), a continuous functional image representation, for parallel MRI reconstruction. A-Blobs represents the image as a summation of Gaussians with parameterised locations, shapes, and intensities. Exploiting the properties of Gaussians, we derived an analytical forward model for parallel MRI, termed MR-Blob. This method exhibited favourable properties, including noise robustness and dense representational capacity.

Collectively, these studies expand the scope of medical image reconstruction and demonstrate the potential of deep generative and functional approaches to significantly advance the field.

Impact Statement

My research advances the field of medical image reconstruction by developing deep generative and functional approaches that integrate known physics into the reconstruction process. This work has significant implications both within academia and in clinical settings, benefiting clinicians, patients, and the broader scientific community.

A key impact of this research is the potential improvement in clinical outcomes. By leveraging advanced techniques for medical image reconstruction, we can produce higher-quality medical images with fewer artifacts and reduced noise. This enhancement allows for lower radiation doses or shorter scanning times, directly benefiting patients by increasing the safety and efficiency of medical imaging procedures.

Beyond direct patient benefits, the research contributes novel techniques to the field of medical imaging. For example, the application of DIP and SGMs for PET image reconstruction, as well as developing a continuous functional approach to image representation and reconstruction of parallel MRI image. For these studies I utilised open-source software such as: Synergistic Image Reconstruction Framework (SIRF) and Software for Tomographic Image Reconstruction (STIR).

I have contributed to open-source development and reproducibility that amplifies the impact on the scientific community. By making code, datasets, and methodologies publicly available, the research facilitates further innovation and bridges the gap between basic and clinical research. These open-source contributions enable other researchers to rigorously test advanced algorithms with realistic data, accelerating scientific progress —a trend that has already proved fruitful in the development of deep learning technologies.

The work has been well-received in the academic community, as evidenced by presentations at leading conferences such as the Medical Imaging Conference (MIC) 2022, Fully3D 2023, and the Medical Imaging with Deep Learning (MIDL) 2024, as well as invited talks at prestigious institutions like New York University, the University of Bremen, and Case Western Reserve University. These engagements have promoted the importance of open-source software and reproducible practices, fostering a culture of collaboration and transparency in research.

In summary, the impact of this research extends from improving patient outcomes and clinical practices to advancing the field of medical imaging and promoting scientific openness. The integration of deep learning with established reconstruction engines and the development of advanced methods have set new standards in the field. The wide-reaching benefits are expected to continue as these technologies are further refined and adopted.

Acknowledgements

This PhD experience is the culmination of over two decades of formal education, and the last 4 years have been the cherry on top. There is so much to acknowledge for that: First and foremost, I'd like to thank my supervisors - Simon, Kris and Bangti. I often think back to our first meetings at the start of COVID, and I picture myself: a nervous student with little to no knowledge of inverse problems, medical imaging, or deep learning, asking some very questionable questions. The kindness you showed me during that time, and throughout my period under your guidance, is something I deeply admire and strive to embody. Learning from you and working with you has been an enormous privilege. Thank you for welcoming me and giving me the space, time, and opportunities to grow into a researcher.

Three others who have taken the brunt of my questions are my close collaborators and friends - Zeljko, Alex and Riccardo. Our work we've done together and the time we've shared has been wonderful, and I've learnt an enormous amount from you all. I am so excited to see what you do in the future, either up close or from afar! Zeljko, thank you in-particular for your patience, rigour and gentle roasts - I owe a great deal of my development as a researcher to you.

The list of others who have joined, aided, and supported me on this journey is long - and probably rather weary of me! A non-exhaustive list includes the folk at the INM (Kjell, Robbie, Ander, Alex, Ashley, Rebecca, Francesca, Nicole, Sam, Statis, Toni and William), rest of the 169 crew (Gizem, Yolanne, Johannes and Cris), my first UCL friends (Snigda, Sophie, Yohn, David, Josie, Poppy), and those in the MIC circuit and SyneRBI ecosystem (Edo, Georg, Andrew). Thank you all for creating such a wholesome environment; it has been a pleasure and privilege to work alongside you.

Thank you to everyone at the i4Health CDT for shaping this PhD experience and giving me the opportunities to develop a new set of skills. I am also grateful to my funders, the EPSRC, for their support.

Outside the academic setting I'd like to thank my friends from B7, the few left from JGHS, and beyond. A special shout-out goes to Baz and Dan, undergrading and phd-ing really doesn't feel like enough time together. Rumours of a Navier-Blokes reunion tour? Closure problem, right?

To the Singh-Dhesis, Lakhans, Maans, Sunners, thank you for always being there for me and making so many great memories.

To my long-suffering better half Simmers, I strive to make your life as wonderful as you make mine, and I am so excited for our future. Who would've thought 7 years ago in the Felix office that I'd be acknowledging you in my thesis? Thank you for everything - and for getting me to write. Thank you also to your family who have been so kind and supportive.

To my family: despite all the adversity, we are getting there. Amar, Nadine, Oscar and Amira - I am so happy to see you all go from strength to strength. Thank you for your support, and wee Amira I am expecting a big hug. To my dear parents: you have always encouraged me to pursue what makes me happy. Thank you for trusting me, guiding me, and for supporting me through it all. Last, and certainly not least, thank you to my big sister, Suki. You have always been my biggest advocate - thank you for always being there for me.[1]

[1]Out of concern for being overly candid, I must also acknowledge the other vital aspects of my PhD experience. To this version of Hey by The Pixies, thank you for being a steadfast companion during countless study sessions. To the humble meal deal and the culinary misadventure of tortellini+ketchup, your sustenance powered me through an embarrassingly large portion of my time. And to the London slumlords and an economy battered by 14 years of austerity - when the going gets tough, the tough get going, albeit quite literally.

UCL Research Paper Declarations

Research Paper 1:

- **What is the title of the manuscript?** Deep Image Prior PET Reconstruction using a SIRF-Based Objective
- **Please include a link to or doi for the work:**
  https://doi.org/10.1109/NSS/MIC44845.2022.10399292
- **Where was the work published?** 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)
- **Who published the work?** IEEE
- **When was the work published?** 05 November 2022
- **List the manuscript's authors in the order they appear on the publication:** Imraj Singh (IS), Riccardo Barbano (RB), Robert Twyman (RT), Željko Kereta (ŽK), Bangti Jin (BJ), Simon Arridge (SA), Kris Thielemans (KT)
- **Was the work peer reviewed?** Yes.
- **Have you retained the copyright?** No.
- **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi** No.
  If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

  ☑ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

- **For multi-authored work, please give a statement of contribution covering all authors:** IS and RB share equal first author contribution. IS led the research project, designed the study, performed analysis, and contributed to the manuscript. RB provided critical insight, aided in study design and analysis, and contributed to the manuscript. RT provided the data used in the study. ŽK, BJ, SA and KT provided insight and reviewed the manuscript.
- **In which chapter(s) of your thesis can this material be found?** Chapter 3.

e-Signatures confirming that the information above is accurate:

Candidate: Imraj Singh

Date: 30 November 2024

Supervisor: Simon Arridge

Date: 02-12-2024

Research Paper 2:

- **What is the title of the manuscript?** 3D PET-DIP Reconstruction with Relative Difference Prior Using a SIRF-Based Objective
- **Please include a link to or doi for the work:** https://doi.org/10.48550/arXiv.2310.16846
- **Where was the work published?** 17th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine
- **Who published the work?** Fully3D
- **When was the work published?** 13 November 2023
- **List the manuscript's authors in the order they appear on the publication:** Imraj Singh (IS), Riccardo Barbano (RB), Željko Kereta (ŽK), Bangti Jin (BJ), Kris Thielemans (KT), Simon Arridge (SA)
- **Was the work peer reviewed?** Yes.
- **Have you retained the copyright?** Yes.
- **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi** No.
  If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

  ☑ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

- **For multi-authored work, please give a statement of contribution covering all authors:** IS led the research project, designed the study, performed analysis, and contributed to the manuscript. RB and ŽK provided critical insight, aided in study design and analysis, and contributed to the manuscript. BJ, KT and SA provided insight and reviewed the manuscript.
- **In which chapter(s) of your thesis can this material be found?** Chapter 3.

e-Signatures confirming that the information above is accurate:

Candidate: Imraj Singh

Date: 30 November 2024

Supervisor: Simon Arridge

Date: 02-12-2024

Research Paper 3:

- **What is the title of the manuscript?** Investigating Intensity Normalisation for PET Reconstruction with Supervised Deep Learning
- **Please include a link to or doi for the work:** https://doi.org/10.1109/NSSMICRTSD49126.2023.10337906
- **Where was the work published?** 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)
- **Who published the work?** IEEE
- **When was the work published?** 04 November 2023
- **List the manuscript's authors in the order they appear on the publication:** Imraj Singh (IS), Alexander Denker (AD), Željko Kereta (ŽK), Bangti Jin (BJ), Kris Thielemans (KT), Simon Arridge (SA)
- **Was the work peer reviewed?** Yes.
- **Have you retained the copyright?** No.
- **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi** No.
  If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

  ☑ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

- **For multi-authored work, please give a statement of contribution covering all authors:** IS led the research project, designed the study, performed analysis, and contributed to the manuscript. AD provided critical insight, aided in study design and analysis, and contributed to the manuscript. ŽK, BJ, KT and SA provided insight and reviewed the manuscript.
- **In which chapter(s) of your thesis can this material be found?** Chapter 4.

e-Signatures confirming that the information above is accurate:
Candidate: Imraj Singh
Date: 30 November 2024
Supervisor: Simon Arridge
Date: 02-12-2024

Research Paper 4:

- **What is the title of the manuscript?** Score-Based Generative Models for PET Image Reconstruction

- **Please include a link to or doi for the work:**
  https://doi.org/10.59275/j.melba.2024-5d51
- **Where was the work published?** Journal of Machine Learning for Biomedical Imaging
- **Who published the work?** MELBA
- **When was the work published?** 04 November 2022
- **List the manuscript's authors in the order they appear on the publication:** Imraj Singh (IS), Alexander Denker (AD), Riccardo Barbano (RB), Željko Kereta (ŽK), Bangti Jin (BJ), Kris Thielemans (KT), Peter Maass (PM), and Simon Arridge (SA)
- **Was the work peer reviewed?** Yes.
- **Have you retained the copyright?** Yes.
- **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi**
  Yes, https://doi.org/10.48550/arXiv.2308.14190.
  If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

  ☑ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

- **For multi-authored work, please give a statement of contribution covering all authors:** IS, AD, and RB share equal first author contribution. IS led the research project, designed the study, performed analysis, and contributed to the manuscript. AD and RB provided critical insight, aided in study design and analysis, and contributed to the manuscript. ŽK, BJ, KT, PM, and SA provided insight and reviewed the manuscript.
- **In which chapter(s) of your thesis can this material be found?** Chapter 5.

e-Signatures confirming that the information above is accurate:
Candidate: Imraj Singh
Date: 30 November 2024
Supervisor: Simon Arridge
Date: 02-12-2024

Research Paper 5:

- **What is the title of the manuscript?** MR-blob: Coordinate-Transformed Blobs for Parallel MRI Reconstruction
- **Please include a link to or doi for the work:**
  https://discovery.ucl.ac.uk/id/eprint/10181468/
- **Where was the work published?** 2023 Proceedings of the International Society for Magnetic Resonance in Medicine Conference

- **Who published the work?** ISMRM
- **When was the work published?** 04 May 2024
- **List the manuscript's authors in the order they appear on the publication:** Imraj Singh (IS), Željko Kereta (ŽK), Alexander Denker (AD), Riccardo Barbano (RB), Bangti Jin (BJ), Kris Thielemans (KT), and Simon Arridge (SA)
- **Was the work peer reviewed?** Yes.
- **Have you retained the copyright?** No.
- **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi** No.
  If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

  ☑ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

- **For multi-authored work, please give a statement of contribution covering all authors:** IS led the research project, designed the study, performed analysis, and contributed to the manuscript. ŽK, AD, RB, BJ, KT, PM, and SA provided insight and reviewed the manuscript.
- **In which chapter(s) of your thesis can this material be found?** Chapter 6.

e-Signatures confirming that the information above is accurate:
Candidate: Imraj Singh
Date: 30 November 2024
Supervisor: Simon Arridge
Date: 02-12-2024

Table of Contents

List of Acronyms

A-Blobs Adaptable Blobs.

AC Activity Concentration.

Adam Adaptive Moment Estimation.

ADMM Alternating Direction Method of Multipliers.

Amyloid ^{18}F-Florbetapir.

BR Background Roughness.

BSREM Block-Sequential Regularised Expectation Maximisation.

CFG Classifier-free Guidance.

CRC Contrast Recovery Coefficient.

CT Computed Tomography.

DDIM Denoising Diffusion Implicit Models.

DIP Deep Image Prior.

DPS Diffusion Posterior Sampling.

DSM Denoising Score Matching.

ELBO Evidence Lower Bound.

EM Expectation Maximisation.

EN Ensemble Noise.

ESPIRiT Eigenvector-based iterative Self-consistent Parallel Imaging Reconstruction.

FDG ^{18}F-FluoroDeoxyGlucose.

FFT Fast Fourier Transform.

FISTA Fast Iterative Shrinkage Thresholding Algorithm.

GANs Generative Adversarial Networks.

GRAPPA GeneRalized Autocalibrating Partial Parallel Acquisition.

HaarPSI Haar Wavelet-Based Perceptual Similarity Index.

iFFT inverse Fast Fourier Transform.

INRs Implicit Neural Representations.

iNUFFT inverse Non-Uniform Fast Fourier Transform.

IQA Image Quality Assessment.

KLDIV Kullbeck-Liebler Divergence.

LoR Line Of Response.

LPD Learned Primal-Dual.

MAP Maximum A Posteriori.

ML Maximum Likelihood.

MLEM Maximum Likelihood Expectation Maximisation.

MLP Multi-Layer Perceptron.

MR-Blob Magnetic Resonance A-Blobs.

MRI Magnetic Resonance Imaging.

MSE Mean Squared Error.

NLL Negative Log-Likelihood.

NUFFT Non-Uniform Fast Fourier Transform.

ODE Ordinary Differential Equation.

ODL Operator Discretization Library.

OSEM Ordered Subset Expectation Maximisation.

OSMAP-OSL Ordered Subset Maximum A Posteriori One Step Late.

PET Positron Emission Tomography.

PSNR Peak Signal-to-Noise Ratio.

QP Quadratic Prior.

RDP Relative Difference Prior.

ReLU Rectified Linear Unit.

RF Radio-Frequency.

ROI Region Of Interest.

SDE Stochastic Differential Equation.

SENSE SENSitivity Encoding.

SGD Stochastic Gradient Descent.

SGMs Score-based Generative Models.

SIRF Synergistic Image Reconstruction Framework.

SSIM Structural Similarity Index Measure.

STIR Software for Tomographic Image Reconstruction.

SVD Singular Value Decomposition.

ToF Time of Flight.

TV Total Variation.

VAEs Variational Auto-Encoders.

Notation

General Symbols and Conventions

Φ	An objective function.
$\mathbf{f}^\dagger$	Dagger indicates a quantity $\mathbf{f}$ that minimises (or maximises) the objective function, e.g. $\mathbf{f}^\dagger = \operatorname{argmin}_{\mathbf{f}} \Phi(\mathbf{f})$.
$\bar{\mathbf{f}}$	Bar indicates a mean quantity.
$\tilde{\mathbf{f}}$	Tilde indicates an approximate quantity.
f	Standard font usually indicates scalars or scalar valued function.
$\mathcal{H}$	Calligraphic font usually indicate operators, or more abstract spaces.
$\mathbb{R}$	Black board font indicates standard spaces.
p	Serif font indicates probability distributions.
i, j, k	Typically indices $i, j, k \in \mathbb{Z}$.
N	Used to denote the number of elements.
$\#S$	Cardinality of set S, i.e. the number of elements in S.
i	Imaginary number to distinguish it from indices.
$\mathcal{F}$	Fourier transform.
$\mathcal{P}_{\geq 0}[\cdot]$	Non-negativity projection operator
$\mathcal{N}$	Neighbourhood.
$\psi, \boldsymbol{\psi}, \boldsymbol{\Psi}$	Basis function, vector and transformation matrix respectively.
Ω, $f\|_\Omega$	An (ordered) set of sample points, and the corresponding evaluation of a function f on Ω producing a vector.

Inverse problems

x^{true}	True solution, image, object.
x	Solution, image, object.
$\mathbf{x}$	Discrete object vector with $N_{\mathbf{x}}$ elements.
y	Measurement(s).
$\mathbf{y}$	Discrete measurement vector with $N_{\mathbf{y}}$ elements.
$\bar{\mathbf{y}}$	Mean of the measurement vector.
$A^{\text{true}}_{\boldsymbol{\varepsilon}}(\cdot)$	True physical forward model.
$A(\cdot)$	Forward model or forward operator.
$\mathbf{A}$	Discrete linear forward operator with $N_{\mathbf{y}} \times N_{\mathbf{x}}$ elements.
$\mathcal{X}$	Reconstruction space or output space, i.e. images, volumes.
$\mathcal{Y}$	Measurement space, i.e. sinogram, k-space data.

Spaces

$\mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{C}, \mathbb{F}, \mathbb{Z}, \mathbb{N}$	Real numbers, non-negative real numbers, complex numbers, field $\in \{\mathbb{R}, \mathbb{C}\}$, integers, natural numbers.
$\mathbb{R}^{N_1 \times N_2}, \mathbb{C}^{N_1 \times N_2}$	Real and complex matrices of size $N_1 \times N_2$, note $N_2 = 1$ for a vector.
$[a, b]$	Closed interval $t \in [a, b] \iff a \leq t \leq b$.
(a, b)	Open interval $t \in (a, b) \iff a < t < b$.
$[a, b), (a, b]$	Half-open intervals $t \in [a, b) \iff a \leq t < b$ and $t \in (a, b] \iff a < t \leq b$.

Linear algebra

$\mathbf{f}$	Vectors.
f_i or $[\mathbf{f}]_i$	i-th element of the vector $\mathbf{f}$.
$\mathbf{H}$	Matrices, i.e. $\mathbf{H} \in \mathbb{R}^{N_{\mathbf{g}} \times N_{\mathbf{f}}}$.
h_{ij} or $[\mathbf{H}]_{ij}$	Element of i-th row and j-th column of matrix $\mathbf{H}$.
$\oslash$	Element-wise division.

$\|\cdot\|_p$	ℓ_p-norm.
$\mathbf{1}$	Vector of ones.
$\mathbf{I}$	Identity matrix.

Deep Learning

$\mathbf{x}^{(s)}$	Samples of dataset.
$\mathbf{y}_{\text{IN}}$	Input data or features.
$\mathbf{x}_{\text{GT}}$	Labels or target outputs.
$\{(\mathbf{y}_{\text{IN}}^{(s)}, \mathbf{x}_{\text{GT}}^{(s)})\}_{s=1}^{N_s}$	Supervised dataset comprised of N_s samples.
$\{\mathbf{x}_{\text{GT}}^{(s)}\}_{s=1}^{N_s}$	Unsupervised dataset comprised of N_s samples.
$\mathcal{G}_{\boldsymbol{\theta}}$	Neural network parameterised by $\boldsymbol{\theta}$.
r_{expected}	Expected risk.
$r_{\text{empirical}}$	Empirical risk.

Probability

$\mathsf{p}(\mathbf{x})$	Probability distribution function.
U	Uniform distribution.
N	Normal/Gaussian distribution.
B	Bernoulli distribution.
$\mathbb{E}[\mathbf{x}]$	Expectation of $\mathbf{x}$

List of Figures

List of Tables

Chapter 1

Introduction

The pipeline of medical imaging, from scanner to clinical outcome, has been developed by many disciplines, each contributing advancements both technological and scientific. The content of this thesis is concerned with image reconstruction. More concretely, given a set of - often indirect, noisy and undersampled - measurements how does one transform these measurements back into the underlying internal patient anatomy?

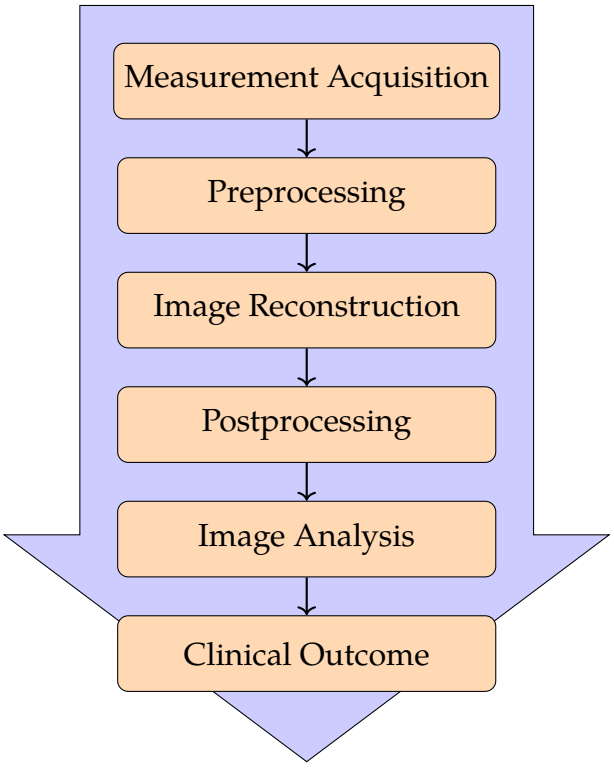


Figure 1.1: Flow diagram of the medical imaging pipeline.

Using indirect, or more precisely projection-based, measurements to re-

construct internal anatomy allows the imaging of internals without invasive surgery. It has been 120 years since the first non-invasive imaging - Wilhelm Röntgen's grainy X-ray of his own hand [193], and now non-invasive imaging is widely available in hospitals with Computed Tomography (CT), Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) among the modalities often available for clinicians.

The purpose of image reconstruction is to produce accurate, reliable, and ideally quick, reconstructions that serve to improve clinical outcomes whether for diagnosis or treatment planning. As well as the reconstruction procedure maintaining appropriate quality and speed, the acquisition procedure is developed to improve patient comfort, increase patient throughput, or reduce patient's absorbed radiation dose.[1] The image reconstruction is therefore constrained in the medical imaging pipeline seen above - Fig. 1.1. Upstream it is beneficial that the measurement acquisition be quicker, or with lower radiation doses to the patient; where the implications on the measurements is increased under-sampling and noise. Advanced image reconstruction aims to compensate for these degraded measurements whilst ensuring downstream analyses are minimally affected.

1.1 Inverse Problems

Medical image reconstruction is a classic example of an inverse problem, where the quantity of interest (derived from patient anatomy) is not measured directly. A general definition is that "inverse problems are concerned with determining causes from a desired or observed effect" [65]. In this context, the cause is the patient anatomy, and the observed effect are the measurements from a scanner. Examples of established clinical scanners include CT, PET, and MRI, where the measurements y and objects x are as follows:

> **CT**: The measurements y are the amplitudes of X-rays detected after irradiating the patient, while the object x is the attenuation of X-rays by the patient's tissue.

[1] Advanced acquisition methods can be developed to exploit known physical principles in novel ways, as described in Section 7.1.

PET: The measurements y are detected photons following the injection of a radiotracer, and the object x is the radiotracer distribution within the patient.

MRI: The measurements y are spatial encoded signals corresponding to specific tissue contrasts, such as T1, T2 or FLAIR, and the object x is the cumulative transverse magnetisation of hydrogen nucleui for this specific contrast.

These modalities are tomographic meaning that measurements are projections of the object (the patient). The measurements the outcome of the following true physical process:

$$y = \mathcal{A}_{\varepsilon}^{\text{true}}(x^{\text{true}}),$$

where $\mathcal{A}_{\varepsilon}^{\text{true}}$ is the true, often non-linear, physical model that incorporates the true noise process indicated by ε, y are the measurements, and x^{true} is the true sought-after object i.e. the solution. The true physical model is approximated with a forward model to respect the modality at hand. Solving the inverse problem amounts to attempting to invert the model, i.e. to go from the measurements back to the object that produced them.

The process of solving an inverse problem is referred to as the reconstruction process or simply reconstruction, herein these terms are used interchangeably. A key difficulty in inverse problems arises when they are ill-posed, where well-posedness is defined as a unique solution x existing that is stably dependent on the measurements y. An ill-posed inverse problem is one that violates the above definition.

For non-trivial inverse problems, acquiring fewer measurements and increasing noise leads to a more ill-posed problem. To stabilise the solution regularisation can be included to the reconstruction process, making the reconstruction better-posed. Broadly speaking, this is the established approach to image reconstruction: ensuring data-consistency between the measurements and the solution, as well as penalising the solution.

1.2 Deep Learning

Deep learning has advanced various fields by learning intricate patterns and representations from data with large complex neural network models [76]. The widespread adoption of deep learning in research is largely facilitated by the development of robust frameworks that offer highly optimised linear algebra routines and automatic differentiation.

In recent years, computer vision research has been dominated by deep learning with state-of-the-art performance on a range of tasks including image reconstruction. The approaches vary significantly in the amount of data, the type of data, and the computation needed. Supervised learning requires labelled paired data to train a model to map from the data to the label. The hope here is that the trained model would be able to generalise sufficiently to data outside the training distribution.

Unsupervised learning forgoes the label and trains only with data. For image reconstruction an unsupervised model requires test-time optimisation to reconstruct images consistent to measured data. This is an inherently more flexible framework, but at the cost of increased reconstruction time.

The inclusion of deep learning into medical image reconstruction requires careful attention. With the added complexity there are two new factors that need to be assessed: Interpretability and generalisability. An interpretable model would be able to give a clinician more faith in the reconstruction, whereas generalisability is particularly vital for medical image reconstruction where, for example, a rare pathology should not be obfuscated by a model that has not been trained to reconstruct it.

1.3 Motivations

A key motivation is to respect the known physics and intricacies of the imaging modalities while improving image quality by reducing noise and artifacts. Physics is respected by enforcing test-time data-consistency with the measurements, where the data-consistency term includes the physics via the forward model. By ensuring the methods produce data-consistent reconstructions, the

interpretability and generalisability of the methods can be improved. This is true for the studies on DIP, SGMs and A-Blobs, but not for supervised methods. Furthermore, the methods are unsupervised meaning that, at most, they only require a dataset of high-quality reference images, rather than labelled paired medical data that is difficult to obtain.

From a different perspective stated in Dimakis [55], the deep generative approaches (DIP and SGMs) to medical image reconstruction attempt to separate out the forward modelling from image modelling. Framing variational methods -discussed in Section 2.1.1- from this perspective, image modelling is done with hand-crafted regularisation. In a deep generative approach the network is leveraged to image model. At test-time both the forward model and image model are used iteratively to reconstruct an image that respects the model as well as the forward model. By separating out these models the generalisation of the approach is improved, for example using the same image model (i.e. a model restricting the solutions to PET images) could be used across forward models (i.e. different scanners, noise-levels) with minimal, or no, re-tuning.

The motivation of A-Blobs is to leverage a continuous functional image representation in order to improve reconstructed image quality. The A-Blobs are able to suppress noise due to their smoothness, and also it was of interest to develop them for an analytical formulation of the forward model in an effort to mitigate discretisation error of the forward model.

1.4 Outline

The background in Chapter 2 introduces literature on inverse problems, medical image reconstruction, and the application of deep learning for medical image reconstruction. The first section introduces the concept of inverse problems, framed through perspectives of regularisation and statistical approaches. These foundational principles are subsequently applied in the context of medical imaging reconstruction, focusing specifically on PET and MRI. This includes a discussion of the physical forward models that underpin these imaging modalities and a review of established algorithms used for image recon-

struction. The chapter concludes by exploring deep learning paradigms, including supervised and unsupervised learning, with a specific focus on how these paradigms have been adapted and applied to medical image reconstruction. This chapter provides the theoretical and conceptual foundation for the studies presented in the subsequent chapters.

In Chapter 3, the application of Deep Image Prior (DIP) methods to PET image reconstruction is explored in depth. This chapter builds on previous work [215, 218], focusing on extending DIP to address the challenges posed by low-count measurements in DIP imaging. The study demonstrates the adaptability and potential of DIP for improving image quality under these challenging conditions, setting the stage for further exploration of reconstruction methodologies.

Chapter 4 shifts the focus to supervised approaches for PET image reconstruction, with particular emphasis on normalisation techniques. The chapter investigates methods for improving image quality through post-processing and learned primal-dual approaches. These methodologies, described in detail in [216], highlight the role of normalisation in enhancing the robustness and accuracy of supervised learning-based reconstruction methods.

In Chapter 5, the use of Score-based Generative Models (SGMs) for PET image reconstruction is studied. This chapter introduces innovative adaptations of sampling techniques and proposes new normalisation procedures tailored for PET imaging. Additionally, it explores the integration of MR-guided reconstruction using SGMs and demonstrates the scalability of this approach to fully three-dimensional reconstructions. This work is published in [220].

The penultimate chapter, Chapter 6, introduces Adaptable Blobs (A-Blobs) as a continuous functional image representation and examines their application in parallel MRI reconstruction, coined Magnetic Resonance A-Blobs (MR-Blob). This study explores the advantages of using continuous representations for improving image reconstruction quality and computational efficiency in MRI, presenting a novel approach to parallel imaging.

Finally, Chapter 7 summarises the key findings of this thesis, and proposes directions for future research giving unpublished preliminary research.

1.4.1 Additional Contributions

In addition to the work presented in this thesis, my PhD studies include contributions to the following publications:

- "Development and evaluation of intraoperative ultrasound segmentation with negative image frames and multiple observer labels" [37].
- "Magnetic Resonance Fingerprinting with Total Nuclear Variation Regularisation" [217].
- "Data-driven approaches for electrical impedance tomography image segmentation from partial boundary data" [52].
- "Usability of PETSIRD, the PET Raw Data open format of the Emission Tomography Standardization Initiative (ETSI): results from ETSI's first hackathon" [114].

Furthermore, outside of publications I have contributed to the following:

- Contributed to the development of the open-source software projects Synergistic Image Reconstruction Framework (SIRF) and Software for Tomographic Image Reconstruction (STIR) [`https://www.ccpsynerbi.ac.uk/ccp-synerbi-prizes-2023-2024/`].
- Assisted in organising a training school on PET/MR reconstruction at the PSMR Conference 2022 [`https://hackmd.io/4bW24YXTSJ-zwJ_XdMn0wA`].
- Presented at a short course on image reconstruction and AI at the IEEE MIC Conference 2022 [`https://www.eventclass.org/contxt_ieee2022/scientific/online-program/session?s=SC-06`].
- Participated in a stochastic hackathon, contributing to the development of a framework for stochastic optimisation algorithms for PET reconstruction [`https://petpp.github.io/hackathon`].
- Recently participated in the PET Rapid Image Reconstruction (PETRIC) challenge, achieving joint second place [`https://petric.tomography.stfc.ac.uk/leaderboard/#timeseries`].

Chapter 2

Background

In this chapter we present a background on inverse problems, then focus on medical image reconstruction (specifically PET and MRI), and finally introduce deep learning for medical image reconstruction. This background aims to contextualise subsequent chapters by introducing pertinent literature, concepts, and topics.

2.1 Inverse Problems

As introduced in Section 1.1, consider again the true physical process of the form: $y = \mathcal{A}_\varepsilon^{\text{true}}(x^{\text{true}})$. The inverse problem is concerned with finding the solution x to this true physical process where the true physical model is $\mathcal{A}_\varepsilon^{\text{true}} : \mathcal{X} \to \mathcal{Y}$, known measurements are $y \in \mathcal{Y}$, and unknown object that is sought, i.e. the solution, $x \in \mathcal{X}$.

In practice, we typically do not have access to the true physical model $\mathcal{A}_\varepsilon^{\text{true}}$, as it is often too complex, highly non-linear, and operates over infinite-dimensional spaces. Instead this model is approximated and discretised as a forward model, which is also referred to as the forward operator, that approximates true physical model to sufficient accuracy. There exists a trade-off between forward modelling fidelity and computational cost, where including more physics into the forward model can lead to computationally prohibitive algorithms. Nonetheless the systematic error induced by approximation of the forward model is assumed to be within the noise level of acquisition [112]. The inverse prob-

lem can be now formulated as the seeking the solution to the more canonical forward problem:

$$y = \mathcal{A}(x),$$

where the forward model $\mathcal{A} \approx \mathcal{A}_{\varepsilon}^{\text{true}} : \mathcal{X} \to \mathcal{Y}$ is the approximated operator that maps the solution x to the measurements y.[1] The practical perspective of inverse problems is focused with determining the solution x, while the theoretical focus is on proving existence, uniqueness and stability of the solutions, as well as the convergence of numerical methods.[2] The inverse problem is said to be well-posed in the sense of Hadamard if the following properties are satisfied [106]:[3]

Existence: A solution x exists for any $y \in \mathcal{Y}$. (2.1)

Uniqueness: The solution x is unique for any $y \in \mathcal{Y}$. (2.2)

Stability: The solution x depends continuously on y. (2.3)

Violation of one, or more, of the above conditions renders the inverse problem ill-posed. Note that this is not a mathematical definition which would require specification of what a solution would be, what measurements are possible, the space $\mathcal{Y}$, and how continuity is measured [65].

In a practical setting the measurements are a discrete vector $\mathbf{y} \in \mathbb{F}^{N_{\mathbf{y}}}$ with $N_{\mathbf{y}}$ elements, and the solution is often parameterised onto a discrete grid of piece-wise constant basis functions called pixels in two-dimensions and voxels in three-dimensions. This discretises the solution as a vector $\mathbf{x} \in \mathbb{F}^{N_{\mathbf{x}}}$. Other representations of the solution are discussed in Section 2.2.1, and a new representation is described in Chapter 6. The discrete forward operator is a linear operator $\mathbf{A} \in \mathbb{F}^{N_{\mathbf{y}} \times N_{\mathbf{x}}}$. Thus the inverse problem seeks to find the solution to:

[1] This would mean $y \approx \mathcal{A}x$, but the approximation is omitted as is common in literature.

[2] Theoretical research often involves functional-analytic regularisation techniques, which is concerned predominantly with stabilising ill-posed inverse problems [65, 234].

[3] In fact the stability criteria was included, at least more clearly, after Hadamard [19, 110].

$$\mathbf{y} = \mathbf{A}\mathbf{x}.$$

This formulation is used in established medical image reconstruction approaches where $\mathbf{x}$ would be the image. A solution to the inverse problem can be obtained through inverting $\mathbf{A}$ such that $\mathbf{x}^\dagger = \mathbf{A}^{-1}\mathbf{y}$. Unfortunately this is only available for invertible forward operators $\mathbf{A}$. Another reconstruction method is least-squares. This method is applicable when $\mathbf{A}$ is not invertible, and a common approach utilises the Moore-Penrose pseudoinverse [172], denoted as $\mathbf{A}^+$, to obtain the least-squares solution.

The Moore-Penrose pseudoinverse is often computed via an Singular Value Decomposition (SVD) of $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ with $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^*$ [33], where the columns $[\mathbf{u}_1, \ldots, \mathbf{u}_{N_\mathbf{y}}] = \mathbf{U} \in \mathbb{F}^{N_\mathbf{y} \times N_\mathbf{y}}$ are the left singular vectors and columns $[\mathbf{v}_1, \ldots, \mathbf{v}_{N_\mathbf{x}}] = \mathbf{V} \in \mathbb{F}^{N_\mathbf{x} \times N_\mathbf{x}}$ are the right singular vectors and both matrices are unitary, i.e. $\mathbf{U}\mathbf{U}^* = \mathbf{I}$ and $\mathbf{V}\mathbf{V}^* = \mathbf{I}$ where * denotes the conjugate transpose and identities are of appropriate dimension. The diagonal elements of $\mathbf{\Sigma} \in \mathbb{R}_{\geq 0}^{N_\mathbf{y} \times N_\mathbf{x}}$ are the singular values and are ordered such that $[\mathbf{\Sigma}]_{11} \geq [\mathbf{\Sigma}]_{ii} \geq [\mathbf{\Sigma}]_{\text{rank}(\mathbf{A})}$. Here $\mathbf{\Sigma}^+ \in \mathbb{R}_{\geq 0}^{N_\mathbf{x} \times N_\mathbf{y}}$ is a diagonal matrix with:

$$[\mathbf{\Sigma}^+]_{ii} = \begin{cases} \frac{1}{[\mathbf{\Sigma}]_{ii}} & \text{for} \quad i < \text{rank}(\mathbf{A}) \\ 0 & \text{for} \quad \text{otherwise} \end{cases},$$

where it can be observed that the singular vectors that are related to smaller singular values are amplified with the pseudoinverse. This makes the decay of the singular values pertinent to the stability, and well-posedness, of the reconstruction method. As each singular value $[\mathbf{\Sigma}]_{ii}$ is associated with corresponding left and right singular vectors, larger singular values relate to low-frequency singular vectors, while smaller singular values correspond to high-frequency singular vectors.

To observe the instability of the reconstruction method, consider forward problem with additive noise $\boldsymbol{\varepsilon}$:

$$\mathbf{y}^{\boldsymbol{\varepsilon}} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}. \tag{2.4}$$

The least squares solution is given by:

$$\mathbf{x}^{\dagger} = \mathbf{A}^{+}\mathbf{y}^{\boldsymbol{\varepsilon}} = \mathbf{A}^{+}\mathbf{A}\mathbf{x} + \mathbf{A}^{+}\boldsymbol{\varepsilon} = \sum_{i=1}^{\operatorname{rank}(\mathbf{A})} \mathbf{v}_i \Bigg(\underbrace{\mathbf{v}_i^{*}\mathbf{x}}_{\substack{\text{solution} \\ \text{component}}} + \underbrace{\frac{\mathbf{u}_i^{*}\boldsymbol{\varepsilon}}{[\boldsymbol{\Sigma}]_{ii}}}_{\substack{\text{noise} \\ \text{component}}} \Bigg).$$

Here it can be clearly seen that the noise component is scaled by $\frac{1}{[\boldsymbol{\Sigma}]_{ii}}$, while the solution component is not. Focusing on the noise component:

$$\sum_{i=1}^{\operatorname{rank}(\mathbf{A})} \frac{1}{[\boldsymbol{\Sigma}]_{ii}} \mathbf{v}_i(\mathbf{u}_i^{*}\boldsymbol{\varepsilon}).$$

For small singular values $[\boldsymbol{\Sigma}]_{ii} \to 0$ the contribution of $\mathbf{u}_i^{*}\boldsymbol{\varepsilon}$ is amplified. Consequently, the components of the noise aligned with high-frequency singular vectors (the smaller singular values) are particularly amplified. This property makes the least-squares solution more sensitive to noise, and the reconstruction method is deemed more ill-posed.

In addition to noise amplification, the singular value spectrum of $\mathbf{A}$ can reveal information loss due to the null-space of the operator. In cases where the SVD of the forward operator $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{*}$ includes zero-valued singular values, directions in the solution space corresponding to those singular values lie within the null-space and are mapped to zero. This means that components aligned with the corresponding singular vectors are lost.

By considering the least squares solution of a noisy linear inverse problem, the ill-posed nature can be observed. The reconstruction method is limited by the fundamental constraints imposed by the singular value spectrum and the null-space of the forward operator.

To combat this sensitivity to noise, regularisation can be used to make the reconstruction method better-posed. While regularisation can stabilise the inversion, it cannot recover information lost to the null-space. In other words, regularisation can only partially address the ill-posed nature of the problem; it cannot restore details that $\mathbf{A}$ fails to capture in the first place.

2.1.1 Regularisation Perspective

Historically, regularisation has transformed the way ill-posed problems are perceived, challenging Hadamard's original assertion that only well-posed problems correspond to physical reality. Despite the initial scepticism, ill-posed problems have proven to be ubiquitous across scientific disciplines [6, 65].

Regularisation can make the reconstruction method better-posed, stabilising the inversion and mitigating noise amplification. However, it can also shift the solution away from the true solution, thus introducing bias. It is therefore important to consider the balance between bias and variance, which is referred to as the bias-variance trade-off in statistical learning [89].[4] This trade-off arises when efforts to improve the stability of the solution (by reducing variance) can lead to a loss in reconstruction accuracy (an increase in bias). Typically, this balance between bias and variance is controlled through a hyper-parameter, which allows adjustment between two goals: maintaining fidelity to the observed measurements (data-consistency) and enforcing a regularisation constraint that introduces a controlled bias. This tuning process helps ensure that the solution aligns well with the data while meeting desired stability criteria [202].

An example of the regularised reconstruction is the truncated pseudoinverse solution [112]:

$$\mathbf{x}^{\dagger} = \sum_{[\boldsymbol{\Sigma}]_{ii} \geq \tau} \mathbf{v}_i \left(\mathbf{v}_i^* \mathbf{x} + \frac{\mathbf{u}_i^* \boldsymbol{\varepsilon}}{[\boldsymbol{\Sigma}]_{ii}} \right),$$

where the hyper-parameter τ is a threshold set based on noise characteristics and the desired solution stability. By adjusting τ, one can control the bias-variance trade-off: increasing τ eliminates smaller singular values, removing high-frequency components of the solution. While this process dampens noise, an excessively high τ can result in information loss of the sought after solution. In imaging, high-frequency components are associated with both noise and large gradients that define structural details such as edges [202]. Thus, an

[4] It is important to note that density estimation methods in statistical learning often involve solving ill-posed problems [246]. As a result, regularisation is commonly used, making the bias-variance trade-off a central consideration.

overly large τ may adversely affect image quality. Conversely, if τ is set too low, noise may re-enter the solution, resembling the non-truncated case.

An alternative established reconstruction approach is to use variational methods [202]:

$$\mathbf{x}^{\dagger} = \underset{\mathbf{x}}{\operatorname{argmin}} \left[\Phi(\mathbf{x}) := \underbrace{\rho(\mathbf{A}\mathbf{x}, \mathbf{y}^{\boldsymbol{\varepsilon}})}_{\text{data-consistency}} + \underbrace{\lambda\eta(\mathbf{x})}_{\text{regularisation}} \right] \tag{2.5}$$

where ρ is the data-consistency between forward-projected solution $\mathbf{A}\mathbf{x}$ and noisy measurements $\mathbf{y}^{\boldsymbol{\varepsilon}}$, and η is the regularisation function that restricts $\mathbf{x}$ to a more feasible set of solutions. The hyper-parameter $\lambda \in \mathbb{R}_{\geq 0}$ is the regularisation parameter that balances the data-consistency and regularisation. The choice of the regularisation parameter is non-trivial and depends on the noise present on measurements and the assumptions on solution [202]. An example of strategy for setting the regularisation parameter is Morozov's Discrepancy Principle. This states that if there is a known discrepancy δ between the noisy measurements and measurements without noise, the regularisation parameter should be set such that $\|\mathbf{A}\mathbf{x}^{\dagger} - \mathbf{y}^{\boldsymbol{\varepsilon}}\| = \delta$ [81].

The optimisation problem outlined in Eqn. (2.5) can be solved using various optimisation algorithms. The choice of algorithm depends on the mathematical properties of the objective function Φ, as well as requirements for convergence speed and solution accuracy. Examples are described in Section 2.2.3.

Tikhonov regularisation was one of the first variational methods [106, 234], the objective function is given by:

$$\Phi_{\text{tikh}}(\mathbf{x}) := \|\mathbf{y}^{\boldsymbol{\varepsilon}} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\boldsymbol{\Gamma}\mathbf{x}\|_2^2,$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{N_{\mathbf{x}} \times N_{\mathbf{x}}}$ is the Tikhonov matrix. This solution can be obtained analytically from ${\mathbf{x}_{\text{tikh}}}^{\dagger} = (\mathbf{A}^*\mathbf{A} + \lambda\boldsymbol{\Gamma}^*\boldsymbol{\Gamma})^{-1}\mathbf{A}^*\mathbf{y}^{\boldsymbol{\varepsilon}}$, but this is often infeasible in-practice due to the expense of inverting a large matrix for high-dimensional problems and numerical instabilities arising from ill-conditioned matrices. For zeroth order Tikhonov, where $\boldsymbol{\Gamma} = \lambda\mathbf{I}$, the solution can alternatively be obtained through

an SVD, such that:

$$\mathbf{x}_{\text{tikh}}{}^{\dagger} = \sum_{i=1}^{\text{rank}(\mathbf{A})} \frac{\mathbf{v}_i([\boldsymbol{\Sigma}]_{ii}^2 \mathbf{v}_i^* \mathbf{x} + [\boldsymbol{\Sigma}]_{ii} \mathbf{u}_i^* \boldsymbol{\varepsilon})}{[\boldsymbol{\Sigma}]_{ii}^2 + \lambda^2}.$$

When comparing with the truncated pseudoinverse solution, Eqn. (2.1.1), we see that the high-frequency components of the solution can be damped through the regularisation parameter λ. Having the constant λ^2 in the denominator flattens the singular spectrum of $\mathbf{A}$ making the operator less ill-conditioned, and the reconstruction method better-posed. Note that as $\lambda \to 0$ then $\mathbf{x}_{\text{tikh}}{}^{\dagger}$ tends to the least squares solution.

It is important to note that regularisers within a variational method do not act alone. The zeroth order Tikhonov example penalises $\|\mathbf{x}\|_2^2$, which by itself has a trivial unique minima when $\mathbf{x} = \mathbf{0}$. Through the action of data-consistency and regularisation, the regularisation affects the solution locally, smoothing noise and regularising the solution.

Tikhonov regularisation and truncated pseudoinverse solutions are classical approaches. Recent advances in regularisation [19] use regularisation functions that extend beyond simple functional forms, leveraging advanced mathematical constructs such as sparsity-promoting norms, and total variation.

Sparsity-promoting norms encourage solutions with many elements that are exactly zero. This is typically achieved by applying an ℓ_1-norm regularisation to a transformed representation of the solution. The transformation is often chosen such that the solution becomes sparse in that domain — for example wavelet, Fourier, or other suitable domain where the solution exhibits sparsity. A general form for a sparsity-promoting regulariser is given as:

$$\eta_{\text{SP}}(\mathbf{x}) = \|\boldsymbol{\Psi}\mathbf{x}\|_1,$$

with the discrete sparsifying transform denoted $\boldsymbol{\Psi}$. In fact the truest measure of sparsity would be the ℓ_0 that is simply the number of non-zero components. However, solving for ℓ_0-norm solutions requires combinatorial algorithms, which are not feasible for high dimensions. The ℓ_1-norm is a convex relaxation of the ℓ_0 [63]. Further, regularisation that imposes a sparse structure is

related to the popular reconstruction approach of compressed sensing. Compressed sensing gives theoretical guarantees for the exact recovery of sparse solutions from fewer measurements than typically required [35, 58]. Unfortunately these guarantees only exist under assumptions that are violated in practical applications. Despite the lack of theoretical guarantees sparsity-promoting regularisers have been shown to work well in practice [147].

Total Variation (TV) regularisation is a sparsity-promoting regulariser on the gradient of the solution. This can be expressed as:

$$\eta_{\mathrm{TV}}(\mathbf{x}) = \|\nabla \mathbf{x}\|_1,$$

where $\boldsymbol{\Psi} = \nabla$ is the gradient operator. TV is effective at promoting piece-wise constant structures that are often sought after characteristics of the solution, because of this it is often referred to as a edge-preserving regulariser. That said, when over-regularised the solutions exhibit blocky artifacts that are unappealing [19]. Further development of TV include non-linear methods that smooth the solution based on local gradient magnitudes that can better reduce noise in the solution [173, 194].

All the reconstruction methods introduced thus far have been hand-crafted which involves the deliberate design of regularisation terms that encode specific assumptions or desired properties of the solution. While these methods have proven highly effective in stabilising reconstruction and preserving important features in the solutions, they rely on constructs that may not fully capture the underlying statistical properties of the sought-after solution.

2.1.2 Statistical Perspective

For the statistical perspective of inverse problems the measurements are modelled as samples drawn from a probability distribution characterising the noise corrupting the measurements. For the noisy forward problem described in Eqn. (2.4) it is common to have normally distributed noise $\boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I}\nu^2)$. The statistical inverse problem in that case can be expressed as:

$$\mathbf{y}^{\boldsymbol{\varepsilon}} \sim \mathsf{N}(\mathbf{A}\mathbf{x}, \mathbf{I}\nu^2), \tag{2.6}$$

where the mean of the measurements is $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x}$. The likelihood function $\mathsf{p}^{\mathrm{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x})$ quantifies the probability of observing measurements $\mathbf{y}^{\boldsymbol{\varepsilon}}$ given the solution $\mathbf{x}$. In statistical inference for inverse problems, our goal is to estimate properties of a distribution based on the observed data, the statistical properties of the noise, and -optionally- some prior knowledge of the solution.

There are two primary schools of thought for statistical inference applied to inverse problems: Frequentist and Bayesian.

2.1.2.1 Frequentist Inference

In the frequentist perspective, the solution $\mathbf{x}$ is treated as an unknown but fixed parameter, while the randomness arises solely from the measurements $\mathbf{y}$, which are considered random samples from their probability distribution. The objective is to estimate this fixed but unknown solution based on the observed data.

One of the foundational methods in this context is the Maximum Likelihood (ML), introduced by Fisher in 1922 [71]. The ML estimate seeks the solution that maximises the likelihood function:

$$\mathbf{x}^{\dagger}_{\mathrm{MLE}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, \mathsf{p}^{\mathrm{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x}).$$

For the case of normally distributed i.i.d noise, as Eqn. (2.6), the likelihood function is given by:

$$\mathsf{p}^{\mathrm{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\nu^n} \exp\left(-\frac{\|\mathbf{y}^{\boldsymbol{\varepsilon}} - \mathbf{A}\mathbf{x}\|_2^2}{2\nu^2}\right).$$

Maximising this likelihood is equivalent to minimising the negative log-likelihood, as the logarithm is a monotonically increasing function, leading to the least-squares method:

$$\mathbf{x}^{\dagger}_{\mathrm{MLE}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left[-\log \mathsf{p}^{\mathrm{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x}) \equiv \frac{1}{2\nu^2}\|\mathbf{y}^{\boldsymbol{\varepsilon}} - \mathbf{A}\mathbf{x}\|_2^2\right].$$

This is equivalent to the Moore-Penrose pseudoinverse solution described

in Section 2.1.1. As described previously, this solution is sensitive to noise in the measurements. Note that in a regularisation framework, discussed in Section 2.1.1, a statistical model for the measurements is often used to justify the data-consistency in a variational method [112]. Furthermore, the frequentist approach does admit regularised solutions, where regularisation improves the properties of the estimation procedure rather than including probabilistic beliefs of the prior.

2.1.2.2 Bayesian Inference

In contrast, the Bayesian perspective treats the solution $\mathbf{x}$ as a random variable with a prior distribution $\pi(\mathbf{x})$. This prior encapsulates any existing knowledge or assumptions about $\mathbf{x}$ before observing the data. When new measurements $\mathbf{y}^{\boldsymbol{\varepsilon}}$ are obtained, Bayes' theorem is used to update the prior distribution, resulting in the posterior distribution:

$$\mathsf{p}^{\text{post}}(\mathbf{x}|\mathbf{y}^{\boldsymbol{\varepsilon}}) = \frac{\mathsf{p}^{\text{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x})\pi(\mathbf{x})}{\mathsf{p}(\mathbf{y}^{\boldsymbol{\varepsilon}})}, \tag{2.7}$$

where $\mathsf{p}(\mathbf{y}^{\boldsymbol{\varepsilon}})$ is the marginal likelihood that normalises the posterior.

A common approach within Bayesian inference is computing the Maximum A Posteriori (MAP) estimate,[5] which seeks the solution $\mathbf{x}$ that maximises the posterior probability [112]:

$$\mathbf{x}^{\dagger}_{\text{MAP}} = \operatorname*{argmax}_{\mathbf{x}} \mathsf{p}^{\text{post}}(\mathbf{x}|\mathbf{y}^{\boldsymbol{\varepsilon}}).$$

Observe that the marginal likelihood is not a function of the solution meaning it does not affect the extrema location and it can be ignored for optimisation. By taking the negative log-likelihood and the posterior is maximised through minimising:

$$\mathbf{x}^{\dagger}_{\text{MAP}} = \operatorname*{argmin}_{\mathbf{x}} \left[-\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x}) - \log \pi(\mathbf{x}) + const \right].$$

This formulation reveals a direct connection between Bayesian inference and regularisation methods [106]. The term $-\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x})$ corresponds to

[5] Other estimators include posterior mean $\mathbb{E}[\mathbf{x}|\mathbf{y}] = \int \mathbf{x}\mathsf{p}^{\text{post}}(\mathbf{x}|\mathbf{y})d\mathbf{x}$.

the data-consistency term, ensuring the solution fits the observed data, while $-\log \pi(\mathbf{x})$ serves as a regularisation term that incorporates prior information about the solution.

The connection is further observed by considering a Gaussian prior $\pi(\mathbf{x}) = \mathsf{N}(\mathbf{0}, \chi^2\mathbf{I})$ with mean-zero and isotopic variance χ such that:

$$-\log \pi(\mathbf{x}) = \frac{1}{2\chi^2}\|\mathbf{x}\|_2^2 + const.$$

Supposing we have the same normally distributed likelihood as before, the MAP estimate is obtained from:

$$\begin{aligned}\mathbf{x}^{\dagger}_{\mathrm{MAP}} &= \underset{\mathbf{x}}{\operatorname{argmin}} \left[-\log \mathsf{p}^{\mathrm{lkhd}}(\mathbf{y}^{\boldsymbol{\varepsilon}}|\mathbf{x}) - \log \pi(\mathbf{x}) \right] \\ &= \underset{\mathbf{x}}{\operatorname{argmin}} \left[\frac{1}{2\nu^2}\|\mathbf{y}^{\boldsymbol{\varepsilon}} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{1}{2\chi^2}\|\mathbf{x}\|_2^2 \right].\end{aligned}$$

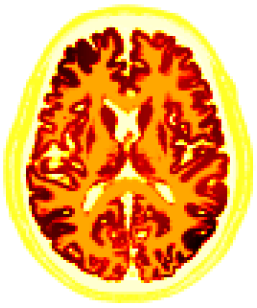
This is equivalent to zeroth order Tikhonov regularisation, where the regularisation term arises naturally from the prior [106]. The Bayesian framework thus provides a probabilistic justification for regularisation techniques, linking the choice of prior to the form of the regularisation.

By leveraging the Bayesian framework, we can derive regularisation techniques grounded in statistical principles, allowing for more flexible and data-adaptive approaches to defining a prior. Furthermore, other properties of the posterior distribution can be assessed which is useful to capture uncertainty in reconstruction [112].

In the next section, we focus on medical image reconstruction, with particular emphasis on PET and MRI. For the remainder of this thesis, we simplify the notation by omitting the $\boldsymbol{\varepsilon}$ superscript used for noisy measurements. Instead, $\mathbf{y}$ will denote the noisy measurements, as noise is inherently present in practical scenarios.

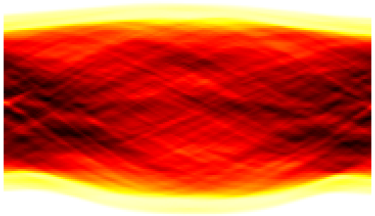
2.2 Medical Image Reconstruction

Medical image reconstruction is typically an ill-posed inverse problem. A range of medical imaging modalities exist but for this thesis we focus on PET and MRI. Figure 2.1 gives idealised measurements and images.

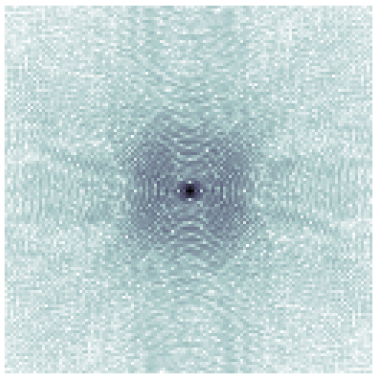


Idealised PET Image

Idealised MRI Image



PET Measurements (Sinogram)



MRI Measurements (k-space)

Figure 2.1: A comparison of idealised PET and MRI images from the BrainWeb dataset with their corresponding idealised raw measurements. The magnitude fully-sampled Cartesian k-space measurements are displayed on a logarithmic scale.

These modalities are distinct in what they measure and how they measure it, but both are non-invasive tomographic imaging techniques where measurements are projections. With undersampling and noise, the problems become increasingly ill-posed, primarily due to challenges associated with the null space of the forward operator and the amplification of noise, as outlined in Section 2.1.

As discussed in Section 1, medical image reconstruction is an integral part of the medical imaging pipeline. Upstream, the acquisition process often yields noisy and incomplete measurements, while downstream, it is crucial to ensure that the reconstructed images do not compromise subsequent analyses. In order to faithfully reconstruct medical images one must first ask the question:

2.2.1 What Are Medical Images?

Medical images in our context are visual representations of the interior of a body. These images are either two-dimensional $N_{\text{dim}} = 2$ images or three-dimensional $N_{\text{dim}} = 3$ volumes,[6] while dynamic imaging can include another temporal dimension. An image can be described as a scalar-valued function:

$$x(\mathbf{r}) : \mathbb{R}^{N_{\text{dim}}} \to \mathbb{F},$$

where $\mathbf{r} \in \mathbb{R}^{N_{\text{dim}}}$ is the spatial coordinate. Discretisation of these images is required for digital processing. The image is often represented as a linear combination of basis functions with associated coefficients of the form:

$$x(\mathbf{r}) = \sum_{b=1}^{N_b} c_{(b)} \psi_{(b)}(\mathbf{r}),$$

where N_b is the number of basis functions, $\psi_{(b)}$ is the b-th basis function, and $c_{(b)}$ is the b-th coefficient. The digital image is sampled at a discrete set of spatial locations, denoted $\Omega_{\mathbf{r}} = \{(r_{(j)})\}_{j=1}^{N_{\mathbf{r}}}$, where $N_{\mathbf{r}}$ is the number of spatial coordinate samples. The discretised image is denoted:

$$\mathbf{x} = x(\mathbf{r})|_{\Omega_{\mathbf{r}}} = \sum_{b=1}^{N_b} c_{(b)} \psi_{(b)}(\mathbf{r})|_{\Omega_{\mathbf{r}}}.$$

Here we use $|_{\Omega_{\mathbf{r}}}$ to represent the ordered evaluation at sample coordinates, producing a vector. A variety of basis functions have been proposed, where their properties lend themselves to different applications. For smoother basis-functions such as Gaussian and Kaiser-Bessel, often referred to blobs in literature [84, 208], their properties are favourable when noise suppression is re-

[6] Image is often used interchangeably with volume.

quired. The most prevalent basis function are piece-wise constant pixels or voxels.

The choice of discretisation significantly impacts both the systematic error in forward modelling and the ill-posedness of the inverse problem [106]. In Chapter 6, we propose a novel parameterisation that treats the image as a continuous Gaussian functions.

The representation of an image $\mathbf{x}$ with a fixed number of basis functions can be written as as a basis-transformation matrix. Let $\mathbf{\Psi} \in \mathbb{R}^{N_b \times N_\mathbf{r}}$ denote the basis-transformation matrix, where each entry $[\mathbf{\Psi}]_{ij} = \psi_{(i)}(\mathbf{r}_{(j)})$ represents the evaluation of the i-th basis function at the j-th coordinate.

$$\mathbf{x} = \mathbf{\Psi}\mathbf{c}.$$

For pixels and voxels, $N_b = N_\mathbf{r}$ and the basis-transformation matrix is effectively an identity matrix. Thus we omit the basis-function and coefficient notation, as is common in literature, when using piece-wise constant basis functions. Note that pixels and voxels resolution is defined by the support of each piece-wise constant basis function. Further discussion on alternatives to voxels is discussed in Chapter 6.

Medical images possess intrinsic properties independent of discretisation. More broadly, the characteristics of natural images have been extensively studied [70, 214] revealing key properties — such as spatial correlations, locality, compositionality, scale-invariance, and non-Gaussian statistics. In medical image reconstruction, incorporating these properties into priors or regularisers can enhance the fidelity of reconstruction:

- Spatial correlations refer to statistical dependencies that can exist over varying distances between pixels, capturing patterns or structures that repeat across an image. This concept encompasses broader relationships beyond immediate neighbours and is often associated with spatial stationarity, where statistical properties are consistent across regions.

- Locality refers specifically to the relationships between neighbouring pixels or patches, emphasising their strong correlation due to spatial prox-

imity. Regularisation methods, such as TV, leverage this property by promoting local consistency while preserving edges [194].

- Compositionality refers to the principle that images can be represented as a combination of simpler, semantically meaningful components, such as shapes, textures, or anatomical structures. This can be leveraged by representing images as combinations of known basis functions or predefined structures, enabling more interpretable and efficient representations [3].
- Scale invariance suggests that image structures are self-similar across scales. Multi-scale regularisation methods capture features at various levels, preserving both fine details and large-scale structures [149].
- Non-Gaussian statistics refer to pixel intensities and filter responses in natural and medical images deviate from Gaussian distributions, often exhibiting heavy tails that indicate a higher occurrence of extreme values. Modelling these non-Gaussian properties, such as with sparsity-promoting regularisers, enhances image reconstruction by better capturing the image's true structure [177].

Both PET and MRI images exhibit the discussed properties despite their fundamentally different imaging mechanisms. PET captures functional information by detecting photons that result from radiotracers that track metabolic processes [50], while MRI generates high-resolution structural images by detecting signals from the transverse magnetisation of hydrogen nuclei in response to magnetic fields [30]. These shared image properties arise from the intrinsic characteristics of natural images, but the specific way each modality generates and relates data to the underlying image depends on its forward model. The forward model encapsulates the physical and mathematical processes linking measurements to the image, and understanding it is crucial for accurate reconstruction, as detailed in the next section.

2.2.2 Modelling the True Physical Forward Process

Medical imaging measurements are the product of complex physics and practical hardware constraints. The true physical forward process is an operator that

converts the object into measurements acquired by the scanner. We introduce a brief review of the true physical forward process that produces the measured data for PET and MRI, as well as describing models for those processes.

2.2.2.1 PET Forward Model: From Radiotracer to Sinogram

PET is a non-invasive imaging modality that detects photons produced by the annihilation of a positron and an electron. The positron is emitted during the decay of a radiotracer, which is a biologically active molecule labelled with a positron-emitting radionuclide.

The PET imaging process begins with the administration of a radiotracer, which exposes patients to ionising radiation. Commonly used radiotracers include ^{18}F-FluoroDeoxyGlucose (FDG) for cancer detection, ^{18}F-Florbetapir (Amyloid) for Alzheimer's disease detection, and Oxygen-15 labelled water for blood flow studies [12]. These radionuclides are produced in cyclotrons. During β^+ decay, the radionuclide undergoes a transformation in which a proton is converted into a neutron, emitting a positron (e^+) and a neutrino (ν_e):

$$ {}^{a}_{z}\mathrm{X} \rightarrow {}^{a}_{z-1}\mathrm{Y} + e^+ + \nu_e, $$

where ${}^{a}_{z}\mathrm{X}$ is the parent radionuclide and ${}^{a}_{z-1}\mathrm{Y}$ is the daughter nuclide.

The kinetic energy of a positron is a characteristic of its specific radionuclide. Once emitted the positron travels and undergoes Coulomb elastic collisions with atomic nuclei and inelastic collisions with atomic electrons [140]. Typically, the positron must lose all its energy before annihilating with an electron [94]. This energy loss occurs gradually through inelastic collisions, each of which dissipates a small amount of energy, requiring many such interactions and the positron follows a tortuous path [11]. This expected distance that the positron travels is referred to as the positron range [140]. Since the positron's kinetic energy is radionuclide-dependent, the spatial resolution is inherently tied to the choice of radionuclide.

The positron-electron annihilation results in the emission of two 511 keV gamma photons travelling in approximately opposite directions (co-linear annihilation). However, due to the residual momentum of the positron-electron

pair, the annihilation photons are emitted at an angle deviating from 180 degrees. This non-collinearity introduces additional blurring in the reconstructed image, with the blurring effect increasing with the diameter of the scanner [213].

As the high-energy annihilation photons traverse the body, they interact with matter primarily through Compton scattering. Compton scattering is the dominant interaction in human tissue for photons between approximately the energies of 100 keV and 2 MeV [12]. The probability of scattering is governed by the Klein-Nishina equation [127], and this is the basis of single scatter simulations often used to account for scatter in the PET forward model [253].

Compton scattering as well as other interactions contribute to a macroscopic linear attenuation coefficient $\mu(\mathbf{r})$.[7] Attenuation is a measure of the probability that a photon will be attenuated by a unit length of the medium [12].

The PET scanner consists of an array of scintillation detectors arranged in a ring configuration. Each detector comprises a scintillator crystal coupled to a photodetector. When two detectors register photon events within a predefined time window, they are considered coincident events and the line connecting the two detectors defines a Line Of Response (LoR). The recorded coincidence events can be split into three categories: true coincidences where no scattering occurs and photons are from the same annihilation event; scattered coincidences that are from the same annihilation event although the photons are affected by scattering; and random coincidences where coincidences are not from the same annihilation event and could have undergone scattering [12]. Furthermore, the detectors are limited by dead time (periods when the detector cannot record new events), discretisation errors due to finite detector size, crystal penetration, inter-crystal scattering, and normalisation factors affecting detector sensitivity [32]. These hardware constraints are an important feature of the true physical forward process.

The coincident events for each LoR are accumulated and these are organised into a sinogram based on their spatial relationship. Considering a single ring, given a set of parallel LoRs forms a projection for a given angle α, where the one-

[7] Such as photoelectric absorption, pair production, Rayleigh scattering, triplet production and photo-nuclear reactions.

dimensional projection dimension s is perpendicular to the LoRs. Therefore, each element of the sinogram corresponds to a binned LoR with specific angle α and position s, i.e. the sinogram is parameterised by α and s.

The sinogram is the discrete representation of the Radon Transform of the two dimensional object $x(r_x, r_y)$. The Radon Transform $R(\alpha, s)$ is defined [106]:

$$Rx(\alpha, s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(r_x, r_y)\delta(r_x \cos\alpha + r_y \sin\alpha - s)dr_x dr_y,$$

δ is the Dirac delta function ensuring integration along the line $s = r_x \cos\alpha + r_y \sin\alpha$.[8] In three dimensions the transform is a ray transform, and the inter-ring coincidence events of multi-ring scanners are often stored in Michelograms [66]. This data-format is often high dimensional with the measurements $\mathbf{y}$ having $N_\mathbf{y}$ elements that are the product of the number of projections, projection angles, and inter-ring cross-planes. To lower the computational demand the discretisation can be made coarser via mashing and rebinning [245], this comes at the expense of error in forward modelling.

When developing the forward model $\mathbf{A}$ the aforementioned processes require correction and incorporation into the forward model. The resolution of the PET image $\mathbf{x} \in \mathbb{R}_{\geq 0}^{N_\mathbf{x}}$ is affected predominately by positron range and non-collinearity, this can be modelled as a blurring operator $\mathbf{A}^{\text{Res.}} \in \mathbb{R}^{N_\mathbf{x} \times N_\mathbf{x}}$.[9] However, resolution modelling is not always used in clinical practice due to the occasional increase in variability when quantifying small structures [182].

The attenuation correction is typically modelled as the diagonal matrix $\mathbf{A}^{\text{Attn.}} \in \mathbb{R}_{\geq 0}^{N_\mathbf{y} \times N_\mathbf{y}}$ [123], which gives the probability of a photon being detected for a given LoR. The detected signal of an LoR I_{LoR} is governed by the Beer-Lambert law:

$$I_{\text{LoR}} = I_0 \underbrace{e^{-\int_{\text{LoR}} \mu(l)dl}}_{\text{attenuation factor}},$$

where variable l lies on the LoR and I_0 is the intensity prior to attenuation. The linear attenuation coefficient can often be obtained from a CT scan. Al-

[8] Given a point source at $r_x = x_0$ and $r_y = y_0$, the Radon Transform simplifies to $R(\alpha, s) = \delta(s - x_0 \cos\alpha - y_0 \sin\alpha)$ which is a sinusoidal curve in the sinogram. This sinusoidal pattern is characteristic and gives the sinogram its name.

[9] PET modelling does not involve complex numbers, hence the restriction of image and forward modelling terms to real number.

ternatively there are methods that attempt to reconstruct as well as estimate attenuation [162].

To correct for scatter and random coincident events additive factors are included in the forward model. The aforementioned single scatter simulation can be used to correct for scattered photons and is denoted $\bar{\mathbf{s}} \in \mathbb{R}^{N_\mathbf{y}}$ [253]. The randoms can be separately corrected for and are denoted as $\bar{\epsilon} \in \mathbb{R}^{N_\mathbf{y}}$ [12].

Detector sensitivity refers to the ability of each detector to register incoming photons. The sensitivity of a detector can be modelled as a function of the detector's physical characteristics and position within the ring. In practice, normalisation factors are applied to compensate for sensitivity variations across the detectors and are denoted as a diagonal matrix $\mathbf{A}^{\text{Norm.}} \in \mathbb{R}^{N_\mathbf{y} \times N_\mathbf{y}}$ [12].

The final component of the PET forward model is the projection operator $\mathbf{A}^{\text{Proj.}} \in \mathbb{R}^{N_\mathbf{y} \times N_\mathbf{x}}$, this models the transformation from three-dimensional radiotracer distribution to number of events on an LoR. This is the most computational intensive operation, requiring projection operators that compute weighted line integrals through the volume. Sophisticated software exist for this computation [207, 233].

The PET forward model, or model of the mean of the measurements [187], is given by:

$$\bar{\mathbf{y}} = \mathbf{A}\mathbf{x} + \bar{\mathbf{b}} = \mathbf{A}^{\text{Norm.}}\mathbf{A}^{\text{Attn.}}\mathbf{A}^{\text{Proj.}}\mathbf{A}^{\text{Res.}}\mathbf{x} + \bar{\mathbf{s}} + \bar{\boldsymbol{\epsilon}}, \tag{2.8}$$

where both scatter and randoms contribute to the estimated background counts $\bar{\mathbf{b}} = \bar{\mathbf{s}} + \bar{\boldsymbol{\epsilon}}$ [180]. These background counts attempt to account for the true scattering and detection processes, but cannot account for the inherent noise on the measurements. This noise is due to the photon-counting process of detection, and given dimensionality of the measurements, concentration of radiotracer, and the duration of a scan, the measurements are a result of a low-counting photon process. This means that the measurements can be modelled as a sample from Poisson distribution with mean the expected measurements $\mathbf{y} \sim \mathsf{P}(\bar{\mathbf{y}})$ [21, 99].

2.2.2.2 MRI Forward Model: From Magnetic Dipoles to k-space

MRI is a non-invasive imaging technique that provides high-resolution structural information by exploiting the magnetic properties of, typically, hydrogen nuclei that are abundant in biological tissues (water and fat).[10] The protons within these hydrogen nuclei possess a quantum property called spin, which can be thought of as intrinsic angular momentum. In the absence of an external magnetic field, these spins are randomly oriented, resulting in no net magnetic dipole moment.

However, when subjected to a strong magnetic field B_0 of an MRI scanner, the spins tend to to align either parallel (lower energy state) or anti-parallel (higher energy state). This alignment creates a small, but detectable net magnetisation M_0 that point along B_0 [30]. This behaviour is governed by:

$$M_0 = N\frac{\gamma^2\hbar^2 B_0}{4k_B T},$$

where N is the number of spins per unit volume, γ is the gyromagnetic ratio, $\hbar$ is the reduced Planck constant, k_B is the Boltzmann constant, and T is the absolute temperature.[11] For context, the magnetic field strength of a typical clinical scanner ($B_0 \approx 3$ T) is more than 45,000 times stronger than the Earth's magnetic field.

In imaging applications, after alignment to the B_0 field -defined as $\mathbf{B} = [0, 0, B_0]^\top$ (i.e. along the z-direction)- the magnetic moment $\mathbf{M} = [M_x, M_y, M_z]^\top$, which is initially aligned as $\mathbf{M} = [0, 0, M_0]^\top$, is perturbed by a carefully designed Radio-Frequency (RF) pulses. These pulses are typically applied at the Larmor frequency ($\omega_0 = \gamma B_0$) that ensures the spins undergo nuclear magnetic resonance, allowing them to efficiently absorb energy from the RF pulse [181].

The RF pulses, typically applied perpendicular to B_0, rotate $\mathbf{M}$ into specified orientations. Each pulse can vary in amplitude, duration and phase, and is part of a pulse sequence designed to manipulate the spins for spatial encoding and contrast generation [30]. The time evolution of the magnetisation vector $\mathbf{M}$

[10] Originally, MRI was called Nuclear Magnetic Resonance Zeugmatographic Imaging [131, 132].

[11] It is important to note that M_0 is made significant due to the sheer number of spins N in a sample.

during and after the pulses is governed by the phenomenological macroscopic Bloch equations [24]:

$$\frac{d}{dt}\begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} = \begin{pmatrix} -\frac{1}{T_2} & \gamma B_z & -\gamma B_y \\ -\gamma B_z & -\frac{1}{T_2} & \gamma B_x \\ \gamma B_y & -\gamma B_x & -\frac{1}{T_1} \end{pmatrix}\begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \frac{1}{T_1} \end{pmatrix}, \tag{2.9}$$

where γ is the gyromagnetic ratio specific to the nucleus (for hydrogen, $\gamma/2\pi \approx 42.58$ MHz/T), T_1 is the longitudinal relaxation time, and T_2 is the transverse relaxation time. The Bloch equations omit additional factors such as diffusion and chemical exchange that can be can be introduced [151, 236].

The measurable component of the magnetisation is on the transverse plane $M_\perp = M_x + \mathrm{i}M_y$ and is referred to as the transverse magnetisation $M_\perp$. The transverse magnetisation is spatially varying, and to reconstruct an image spatial encoding is required. Spatial encoding in MRI is achieved using magnetic field gradients superimposed on the static B_0 field. These gradients, denoted $\frac{d\mathbf{B}}{dt} = [\frac{dB_x}{dt}, \frac{dB_y}{dt}, \frac{dB_z}{dt}]^\top$, create a spatial variation in the magnetic field, and thus in the Larmor frequency. These gradients are usually linear and they allow the encoding of spatial information into the phase and frequency of the transverse magnetisation [131]. The signal received by a receiver coil at time t can be expressed as:

$$s(t) = \int_{\Omega_\mathbf{r}} M_\perp(\mathbf{r}) e^{-i\phi(\mathbf{r},t)} d\mathbf{r}, \tag{2.10}$$

where $M_\perp(\mathbf{r})$ is transverse magnetisation at position $\mathbf{r}$ and $\phi(\mathbf{r}, t)$ is the accumulated phase at time t, given by:

$$\phi(\mathbf{r}, t) = \omega_0 t + \gamma \int_0^t \frac{d\mathbf{B}}{dt}\tau \cdot \mathbf{r} d\tau.$$

By substituting $\omega_0 = \gamma B_0$ and considering that the static field contributes a constant phase evolution common to all spins, we can focus on the phase accumulation due to the gradients:

$$\phi(\mathbf{r}, t) = \gamma \int_0^t \frac{d\mathbf{B}}{dt}\tau \cdot \mathbf{r} d\tau.$$

This remaining term encodes spatial information via the gradients. Defining the spatial frequency variables $\mathbf{k} = [k_x, k_y, k_z]^\top$ (also known as k-space trajectories):

$$\mathbf{k}(t) = \frac{1}{2\pi}\gamma \int_0^t \frac{d\mathbf{B}}{dt}\tau d\tau,$$

the accumulated phase becomes:

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

Substituting back into the signal equation, Eqn. (2.10), we obtain:

$$s(t) = \int_{\Omega_\mathbf{r}} M_\perp(\mathbf{r}) e^{-\mathrm{i}2\pi\mathbf{k}(t)\cdot\mathbf{r}} d\mathbf{r}.$$

This shows that the measured signal is the Fourier transform of the transverse magnetisation. In practice, the signal is sampled at discrete time points t_n, corresponding to discrete points in k-space $\mathbf{k}_n = \mathbf{k}(t_n)$. We denote the ordered set of measured k-space points as $\Omega_\mathbf{k}$, and the discretised signal measurements can be represented as:

$$s_n = \int_{\Omega_\mathbf{r}} M_\perp(\mathbf{r}) e^{-\mathrm{i}2\pi\mathbf{k}_n\cdot\mathbf{r}} d\mathbf{r}, \quad n = 1, \ldots, N_\mathbf{y},$$

where $N_\mathbf{y}$ is the number of measurements.

Further, multiple receiver coils can be used in a technique known as parallel imaging. Each coil is modelled as:

$$s_{(\kappa)}(t) = \int_{\Omega_\mathbf{r}} \upsilon_{(\kappa)}(\mathbf{r}) M_\perp(\mathbf{r}) e^{-\mathrm{i}2\pi\mathbf{k}(t)\cdot\mathbf{r}} d\mathbf{r},$$

where $\upsilon_{(\kappa)}$ represents the sensitivity of the κ-th coil out of N_κ coils. The coil sensitivity, $\upsilon_{(\kappa)}$, encodes spatial dependency, allowing for better spatial localisation of the transverse magnetisation. This is particularly valuable in accelerated imaging applications, where not all of k-space is sampled, resulting in undersampled measurements [53]. The discrete forward model for parallel imaging is often denoted:

$$\bar{\mathbf{y}} = \mathbf{A}^{\text{Fourier}}\mathbf{A}^{\text{Coils}}\mathbf{x} \tag{2.11}$$

Here the measurements are for all the coils $\mathbf{y} \in \mathbb{C}^{N_\kappa . N_\mathbf{y}}$, $\mathbf{A}^{\text{Fourier}} \in \mathbb{C}^{(N_\mathbf{y} N_\kappa)\times(N_\mathbf{x} N_\kappa)}$ is a discrete Fourier transform for the corresponding k-space coordinates $\Omega_\mathbf{k}$ that has a block-repeated structure, $\mathbf{A}^{\text{Coils}} \in \mathbb{C}^{(N_\mathbf{x} N_\kappa)\times N_\mathbf{x}}$ are the discrete coil sensitivities that have a stacked diagonal structure, $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ is the discrete image of transverse magnetisation. In practice there a few techniques to estimate the coil sensitivities, whether it explicit knowledge [179, 241], or calibration of the k-space measurements [80], this is discussed further in Section 2.2.3.2

In practical MRI systems, several factors affect the accuracy and quality of the forward model. These can be incorporated into the forward model as separate components. The finite relaxation times T_1 and T_2 cause the magnetisation to decay over time toward equilibrium $\mathbf{M} = [0, 0, M_0]^\top$. This decay can be modelled by including exponential attenuation in the signal [68]:

$$s_{(\kappa)}(t_n) = \int_{\Omega_\mathbf{r}} \upsilon_{(\kappa)}(\mathbf{r}) M_\perp(\mathbf{r}) e^{-\mathrm{i}2\pi\mathbf{k}_n\cdot\mathbf{r}} e^{-t_n/T_2^*} d\mathbf{r},$$

where T_2^* accounts for both T_2 relaxation and magnetic field inhomogeneities. Additionally, variations in $\mathbf{B}_0$ and susceptibility differences introduce phase errors [68]. These can be modelled by adding a spatially varying off-resonance frequency $\Delta\omega(\mathbf{r})$, adding this to the phase gives:

$$\phi(\mathbf{r}, t_n) = 2\pi\mathbf{k}_n \cdot \mathbf{r} + \Delta\omega(\mathbf{r}) t_n.$$

There are further imperfections such as gradient linearity, eddy currents, and hardware limitations that can be included too. Notwithstanding, in many cases the simplified forward model of receiver coils and Fourier transform suffices. This is denoted as:

$$\bar{\mathbf{y}} = \mathbf{A}\mathbf{x},$$

where $\mathbf{A}$ includes are linear approximations to MRI physics and hardware constraints. The noise affecting MRI k-space measurements additive, white com-

plex Gaussian noise [95].[12]

2.2.3 Model-based Medical Image Reconstruction

Model-based reconstruction approaches are among the most established and flexible methods in medical imaging. These models explicitly incorporates a mathematical model for the imaging process and follows the variational approaches discussed earlier in Section 2.1.1. In particularly the forward model represents the imaging process from image to measurements $\mathbf{A} : \mathcal{X} \to \mathcal{Y}$. In addition to leveraging the imaging process, prior information about the target image can be leveraged. Thus, in variational framework, the choice of regulariser is tailored for the properties of the imaging modality, see 2.2.1.

As described in Section 2.1.2, the noise model dictates the data-consistency term, which quantifies the discrepancy between the expected measurements $\bar{\mathbf{y}}$ and the actual observed measurements $\mathbf{y}$. This term must respect the noise model inherent to the measurements, as different imaging modalities are affected by different types of noise. For instance, in PET, the measurements are affected by Poisson noise due to the stochastic nature of radioactive decay and photon detection. Each element of the PET, sinogram data is modelled as a realisation from a Poisson random variable [12]. In MRI, the measurements are modelled as corrupted by white complex-valued Gaussian noise [30].

In the following sections, we will consider approaches for model-based reconstruction for both PET and MRI, highlighting important considerations, caveats, and algorithms used in practice.

2.2.3.1 PET Image Reconstruction

The PET scanner event detection is a photon-counting process. Photon-counting is intrinsically integer-valued, and is dictated by the half-life and spatial distribution of the radio-tracer. Given this process, PET measurements are most appropriately modelled as a realisation from a Poisson distribution:

$$\mathbf{y} \sim \mathsf{P}(\bar{\mathbf{y}}).$$

[12] The noise on magnitude images follows a Rician distribution, not Gaussian [36].

Given that the noise model is Poisson, the likelihood function, which ensures data consistency, is the Poisson likelihood. In practice, the Poisson Negative Log-Likelihood (NLL) is commonly used. Taking the logarithm, as discussed in Section 2.1.2, is valid because it is a monotonic transformation that preserves the location of the minimiser. Furthermore, the logarithm simplifies the product into a summation, and the negative sign is necessary when framing the problem as a minimisation task.[13] The Poisson NLL is expressed as:

$$-\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\bar{\mathbf{y}}) = -\log\left\{\prod_{i=1}^{N_{\mathbf{y}}} \mathsf{p}^{\text{lkhd}}(y_i|\bar{y}_i)\right\} \tag{2.12}$$

$$= -\log\left\{\prod_{i=1}^{N_{\mathbf{y}}} \exp(-\bar{y}_i)\frac{\bar{y}_i^{y_i}}{y_i!}\right\} \tag{2.13}$$

$$= \sum_{i=1}^{N_{\mathbf{y}}}\left(\bar{y}_i - y_i\log(\bar{y}_i) \quad \underbrace{+\log(y_i!)}_{\text{independent of } \bar{\mathbf{y}}},\right) \tag{2.14}$$

where the forward problem is the models of the mean measurements $\bar{\mathbf{y}}$:

$$\bar{\mathbf{y}} = \mathbf{A}\mathbf{x} + \bar{\mathbf{b}}.$$

the term independent of $\bar{\mathbf{y}}$ in Eqn. (2.14) would not affect a minimisation problem where $\mathbf{x}$ is the variable. Thus, omitting this term the NLL, or data-consistency is given by:

$$\rho_{\text{PET}}(\mathbf{y}|\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}) := \sum_{i}^{N_{\mathbf{y}}} [\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}]_i + y_i \log([\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}]_i).$$

The PET data-consistency ρ_{PET} is convex, however, the data-consistency is not defined $\log([\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}]_i) = \log(0)$. PET reconstruction engines, such as STIR, are able to account for this.

The gradient of the PET data-consistency is given by:

$$\frac{\partial \rho_{\text{PET}}(\mathbf{y}|\mathbf{A}\mathbf{x} + \bar{\mathbf{b}})}{\partial \mathbf{x}} := \mathbf{A}^\top \left[\mathbf{y} \oslash [\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}] - \mathbf{1}\right],$$

[13] In PET literature maximisation is more commonly used. Whereas minimisation is the standard approach in other fields such as machine learning and MRI image reconstruction.

where $\oslash$ denotes element-wise division and $\top$ denotes the adjoint. Challenges arise when any element of $[\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}]_i = 0$ while the corresponding $y_i > 0$, as this leads to a division by zero in the term $\mathbf{y} \oslash [\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}]$. However the background $\bar{\mathbf{b}}$ is usually non-zero and this is rarely observed in practice [180].

The PET measurements are typically affected by a large amount of noise, especially in the low-count regime, necessitating strong regularisation in the reconstruction algorithm to suppress noise amplification and stabilise the solution [12]. Moreover, the reconstructed image must satisfy the non-negativity constraint $\mathbf{x} \in \mathbb{R}^{N_\mathbf{x}}_{\geq 0}$ as the radiotracer cannot be negative. Additionally, due to patient-specific variations in metabolism and other factors, the dynamic range of $\mathbf{x}$ can vary significantly [161].

Maximum Likelihood Expectation Maximisation (MLEM) is an algorithm that updates the image estimate to obtain the ML estimate, $\mathbf{x}^{(k)}$ where k indicates the iterate, to maximise the likelihood as described in Section 2.1.2.1 to obtain the ML estimate [211]. For PET this update is given by:

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{x}^{(k)}}{\mathbf{A}^\top \mathbf{1}} \mathbf{A}^\top \left(\frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(k)} + \bar{\mathbf{b}}} \right).$$

The first iterate $\mathbf{x}^{(0)} \in \mathbb{R}^{N_\mathbf{x}}_{\geq 0}$ is strictly positive, and MLEM preserves non-negativity through the multiplicative updates. In practice a small positive value is included in the numerator of $\frac{\mathbf{x}^{(k)}}{\mathbf{A}^\top \mathbf{1}}$ to ensure the solution updates as a value of zero would cease to update. The term $\mathbf{A}^\top \mathbf{1}$ is called the sensitivity image that can be pre-computed, and $\frac{\mathbf{x}^{(k)}}{\mathbf{A}^\top \mathbf{1}}$ is often referred to as the Expectation Maximisation (EM) preconditioner.

In practice, MLEM is early-stopped to prevent the solution over-fitting to noise in the measurements. Further, the algorithm converges slowly, and the application of the forward operator $\mathbf{A}$ and adjoint $\mathbf{A}^\top$ is expensive due to the number of LoRs of the scanner. This motivates algorithms that converge quicker that include Ordered Subset Expectation Maximisation (OSEM) and Block-Sequential Regularised Expectation Maximisation (BSREM). These algorithms update the solution based on a subset (block) of the measurements, thus reducing forward operator $\mathbf{A}$ to only the rows corresponding to those measurements.

Ordered Subset Expectation Maximisation algorithm is an adaptation of MLEM that uses subsets [103]. A subset of the measurements is denoted $\{(\mathbf{y}_s)\}_{s=1}^{N_S}$ where the index s is a specific subset out of N_s disjoint subsets. These subsets often correspond to a set of equally spaced rows across all sinogram-measurements, which is suggested to balance to subsets to minimise variance between them [103, 180]. The accessing of these subsets can additionally be tuned to compensate for unbalanced subsets, which can aid convergence [96]. A subset of the background is denoted $\bar{\mathbf{b}}_s$, and corresponding rows of the forward operator is $\mathbf{A}_s$. This gives the OSEM update:

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{x}^{(k)}}{\mathbf{A}^\top \mathbf{1}} \mathbf{A}_{s(k)}^\top \left(\frac{\mathbf{y}_{s(k)}}{\mathbf{A}_{s(k)}\mathbf{x}^{(k)} + \bar{\mathbf{b}}_{s(k)}} \right). \tag{2.15}$$

The subset index $s(k) = (k \mod N_s) + 1$ cyclically iterates through the N_s subsets. This algorithm does not converge to the ML estimate [31, 180]. In clinical practice OSEM reconstructions are early-stopped and post-smoothed to supress noise [180]. The EM preconditioner $\frac{\mathbf{x}^{(k)}}{\mathbf{A}^\top \mathbf{1}}$ often includes small positive value similar to MLEM.

Maximum A Posteriori estimates do not rely on heuristics such as early-stopping and post-smoothing, and instead incorporate regularisation into the objective function, i.e. this can be viewed as a variational method, see Eqn. (2.5), of the form:[14]

$$\Phi(\mathbf{x}) := \rho_{\text{PET}}(\mathbf{y}|\mathbf{A}\mathbf{x} + \bar{\mathbf{b}}) + \lambda\eta(\mathbf{x}).$$

The inclusion of regularisation mitigates the need for early-stopping, with regularisation parameter λ weighing regularisation strength. This objective function can be split into subsets such that:

$$\Phi_s(\mathbf{x}) := \rho_{\text{PET}}(\mathbf{y}|\mathbf{A}_s\mathbf{x} + \bar{\mathbf{b}}_s) + \frac{\lambda}{N_s}\eta(\mathbf{x}), \tag{2.16}$$

[14] This muddles the functional analytic regularisation and Bayesian inferences approaches of Sections 2.1.1 and 2.1.2.2. The priors, Quadratic Prior and Relative Difference Prior, are only weakly motivated from a statistical perspective. Perhaps more correctly they can be referred to as regularisers or penalisations that penalise the solution to have desired properties. We refer to these "priors" as regularisers or penalisations henceforth.

with $\{(\Phi_s)\}_{s=1}^{N_s}$ sub-objective functions. For PET, regularisers that reduce noise and smooth the image are particularly effective. Two well-established regularisers in this context are the Quadratic Prior and the Relative Difference Prior [78, 79, 161].

Quadratic Prior encourages smoothness in the reconstructed image by penalising the squared differences between neighbouring voxels [78, 79]. It is defined as:

$$\eta_{\text{QP}}(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N_\mathbf{x}}\sum_{j\in\mathcal{N}_i} w_{ij}(x_i - x_j)^2,$$

where $\mathcal{N}_i$ denotes the set of neighbouring voxels of voxel i, and w_{ij} are weights that are often set to one or determined based on voxel distances. This prior is also equivalent to Tikhonov regularisation, see Section 2.1.1, provided the appropriate neighbourhood and weights.

The gradient of the QP with respect to $\mathbf{x}$ is:

$$\left[\frac{\partial \eta_{\text{QP}}(\mathbf{x})}{\partial \mathbf{x}}\right]_i = \sum_{j\in\mathcal{N}_i} w_{ij}(x_i - x_j).$$

Relative Difference Prior is designed to preserve edges while enforcing smoothness in homogeneous regions [161], it is defined as:

$$\eta_{\text{RDP}}(\mathbf{x}) := \sum_{i=1}^{N_\mathbf{x}}\sum_{j\in\mathcal{N}_i} w_{ij}\frac{(x_i - x_j)^2}{x_i + x_j + \gamma|x_i - x_j|},$$

where a small positive value is often included in the denominator to avoid division by zero, γ is the edge-preserving hyper-parameter, and w_{ij} is as before. The gradient of the RDP with respect to $\mathbf{x}$ is:

$$\left[\frac{\partial \eta_{\text{RDP}}(\mathbf{x})}{\partial \mathbf{x}}\right]_i = \sum_{j\in\mathcal{N}_i} w_{ij}\frac{(x_i - x_j)(\gamma|x_i - x_j| + x_i + 3x_j)}{(x_i + x_j + \gamma|x_i - x_j|)^2}.$$

RDP preserves anatomical edges, which is crucial in medical imaging [161]. Additionally, the gradient can be rewritten in terms of the ratio $r_{ij} = \frac{x_j}{x_i}$:

$$\left[\frac{\partial \eta_{\text{RDP}}(\mathbf{x})}{\partial \mathbf{x}}\right]_i = \sum_{j \in \mathcal{N}_j} w_{ij} \frac{(r_{ij} - 1)(\gamma |r_{ij} - 1| + r_{ij} + 3)}{(r_{ij} + 1 + \gamma |r_{ij} - 1|)^2},$$

this makes RDP more appropriate for PET as the gradient is stable to variations in dynamic range, meaning that the choice of λ is also easier to tune.

Block-Sequential Regularised Expectation Maximisation is an subset-based accelerated algorithm used in PET to solve the previously described objective function [4, 49]. BSREM is provably convergent under mild assumptions and employs an additive update given by [180]:

$$\mathbf{x}^{(k+1)} = \mathcal{P}_{\mathbf{x} \geq 0} \left[\mathbf{x}^{(k)} - \alpha^{(k)} D^{(k)}(\mathbf{x}^{(k)}) \nabla \Phi_{s(k)}(\mathbf{x}^{(k)}) \right]. \tag{2.17}$$

This is gradient descent with subset gradient $\Phi_{s(k)}$, where $s(k)$ again cyclically accesses the subsets. The k-th step-size is given by $\alpha^{(k)}$, preconditioner by $D^{(k)}(\cdot)$ which is often the EM preconditioner, and $\mathcal{P}_{\mathbf{x} \geq 0}[\cdot]$ is a non-negativity projection. The step-size is computed with $\alpha^{(k)} = \alpha_0/(\zeta n + 1)$, where α_0 is the initial step-size and ζ is a relaxation coefficient. This decaying step-size is required to ensure convergence.

2.2.3.2 MRI Image Reconstruction

MRI measurements are modelled as being corrupted by white complex Gaussian noise and the measurements are modelled as a realisation from a Gaussian distribution $\mathbf{y} \sim \mathsf{G}(\bar{\mathbf{y}}, \mathbf{I}\sigma^2)$ where σ is the standard deviation [51, 95]. The data-consistency for MRI measurements is derived from the Gaussian NLL:

$$
\begin{aligned}
-\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\bar{\mathbf{y}}) &= -\log\left\{\prod_i^{N_\mathbf{y}}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\bar{y}_i - y_i)^2}{2\sigma^2}\right)\right]\right\} \\
&= \sum_i^{N_\mathbf{y}}\left[\log(\sigma\sqrt{2\pi}) + \frac{1}{2\sigma^2}(\bar{y}_i - y_i)^2\right] \\
&= \underbrace{N_\mathbf{y}\log(\sigma\sqrt{2\pi})}_{\text{independent of }\bar{\mathbf{y}}} + \sum_i^{N_\mathbf{y}}\frac{1}{2\sigma^2}(\bar{y}_i - y_i)^2.
\end{aligned}
$$

Now consider the forward problem $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x}$. For simplicity and without loss of generality, the factor $\frac{1}{\sigma^2}$ can be absorbed into the regularisation parameter. Thus, the data-consistency term becomes:

$$
\rho_{\text{MRI}}(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2.
$$

The MRI data-consistency term is convex and has a Lipschitz continuous gradient. The gradient given by:

$$
\frac{\partial \rho_{\text{MRI}}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{A}^*\left(\mathbf{A}\mathbf{x} - \mathbf{y}\right).
$$

This is equivalent to least-squares function form described in Section 2.1.1

inverse Fast Fourier Transform (iFFT) can be directly applied to k-space measurements when sampled on a Cartesian grid with a low noise level. In this setting the iFFT performs well. The reconstruction is obtained directly from:

$$
\mathbf{x}^\dagger = \mathbf{A}^{-1}\mathbf{y},
$$

where $\mathbf{A}^{-1}$ denotes the iFFT operator [44]. The iFFT can be computed implicitly with a computational complexity of $\mathcal{O}(N \log N)$, with N being the number of voxels or k-space points. However, in clinical practice, acquiring fully sampled k-space data is time-consuming and may be impractical. To reduce scan times, undersampling strategies are employed, leading to incomplete k-

space data and an ill-posed inverse problem [68].

inverse Non-Uniform Fast Fourier Transform (iNUFFT) can be leveraged in scenarios where k-space data are acquired along non-Cartesian trajectories—such as radial, spiral, or other non-uniform sampling patterns that do not lie on a uniform grid. The Non-Uniform Fast Fourier Transform (NUFFT) can be used to efficiently compute the Fourier transform of non-uniformly sampled data [69, 166].

The NUFFT approximates the Fourier transform by interpolating non-uniform k-space samples onto a uniform grid, applying the FFT, and compensating for interpolation errors. The forward operator $\mathbf{A}$ in this context incorporates the NUFFT, and its Hermitian adjoint $\mathbf{A}^*$ is used in iterative reconstruction algorithms [69].

For undersampled k-space the forward operator has a significant null-space and requires regularisation, see Section 2.1. This is usually approached as a variational method, and a Total Variation regularisation is popular in MRI reconstruction.[15]

Total Variation (TV) regularisation is a widely used technique in MRI reconstruction that promotes piecewise smoothness while preserving sharp edges and fine details, this was briefly discussed in Section 2.1.1. A variety of discrete formulations of TV exist, and we focus on isotropic TV [43].

The discrete isotropic total variation of an image $\mathbf{x} \in \mathbb{R}^{N_\mathbf{x}}$ is defined as:

$$\eta_{\mathrm{TV}}(\mathbf{x}) = \sum_{i=1}^{N_\mathbf{x}} \sqrt{\sum_{j \in \mathcal{N}_i} w_{ij}(x_i - x_j)^2}, \tag{2.18}$$

in this case the neighbourhood $\mathcal{N}_i$ is defined such that $x_i - x_j$ is an (forward) finite difference approximation of a gradient for a two or three-dimensional image [43]. The weighting w_{ij} is often used in practice and depends on the resolution. This function is convex, but the gradient is non-smooth meaning that special consideration needs to be taken to ensure correct treatment.

[15] Note that TV has also been proposed for PET [61]. However it is more widely used in MRI.

Fast Iterative Shrinkage Thresholding Algorithm (FISTA) can be used to solve the regularised reconstruction problem efficiently, especially with non-smooth regularisation terms such as TV [17]. The FISTA iterative updates are:

$$\begin{aligned}
\mathbf{x}^{(k+1)} &= \text{prox}_{\gamma\lambda\eta}\left(\tilde{\mathbf{x}}^{(k)} - \gamma\nabla\rho_{\text{MRI}}(\tilde{\mathbf{x}}^{(k)})\right), \\
t^{(k+1)} &= \frac{1 + \sqrt{1 + 4(t^{(k)})^2}}{2}, \\
\tilde{\mathbf{x}}^{(k+1)} &= \mathbf{x}^{(k+1)} + \frac{t^{(k)} - 1}{t^{(k+1)}}(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}),
\end{aligned}$$

where $\tilde{\mathbf{x}}^{(k)}$ is an extrapolated point incorporating momentum to accelerate convergence, $\gamma > 0$ is the step size, chosen to satisfy certain convergence criteria - see [17] for details, $\nabla\rho_{\text{MRI}}(\tilde{\mathbf{x}}^k) = \mathbf{A}^*(\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{y})$ is the gradient of the data-consistency term, and $\text{prox}_\eta(\cdot)$ is the proximal operator associated with the regularisation term $\lambda\eta(\cdot)$ [27, 170]. The proximal operator for a function $\lambda\eta(\cdot)$ is defined as:

$$\text{prox}_{\gamma\lambda\eta}(\mathbf{v}) = \underset{\mathbf{x}}{\text{argmin}}\left(\frac{1}{2\gamma}\|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda\eta(\mathbf{x})\right).$$

For certain choices of $\eta(\mathbf{x})$, the underlying minimisation problem of the proximal operator has a closed-form solution. For example, if $\eta(\mathbf{x}) = \|\mathbf{x}\|_1$, the proximal operator corresponds to the soft-thresholding function [170]:

$$[\text{prox}_{\gamma\lambda\|\cdot\|_1}(\mathbf{v})]_i = \begin{cases} v_i - \gamma\lambda & \text{if} \quad v_i > \gamma\lambda, \\ 0 & \text{if} \quad |v_i| \leq \gamma\lambda, \\ v_i + \gamma\lambda & \text{if} \quad v_i < \gamma\lambda. \end{cases}$$

FISTA has a convergence rate of $\mathcal{O}(1/k^2)$[16]. This makes FISTA particularly suitable for large-scale MRI reconstruction problems where convergence speed is paramount.

Parallel MRI Reconstruction encompasses several established methods that include SENSitivity Encoding (SENSE), GeneRalized Autocalibrating Partial

[16]For functions such as MRI data-consistency with isotropic TV, see [17] for details.

Parallel Acquisition (GRAPPA), and Eigenvector-based iterative Self-consistent Parallel Imaging Reconstruction (ESPIRiT) [80, 179, 241].

SENSE poses parallel image reconstruction as a linear inverse problem by directly approximating the coil sensitivities $\mathbf{A}^{\text{Coils}}$ from Eqn. (2.11) [179]. GRAPPA treats the reconstruction problem as an interpolation problem in k-space [80]. In GRAPPA the missing lines of k-space are interpolated via a kernel determined from an auto-calibration region of k-space. The interpolated k-space is used to compute coil images that are combined to give a magnitude image. The third method ESPIRiT leverages ideas from both SENSE and GRAPPA, where an auto-calibration region is used to estimate coil sensitivities directly - see [241] for details. These methods leverage the redundancy of multi-coil measurements to mitigate the ill-posedness of accelerated parallel MRI reconstruction. As stated in [241], the methods can be described as subspace methods, where missing measurements are reconstructed by restricting a solution to a subspace. This concept is extended in Chapter 6, where sensitivities and image are defined a subspace of Gaussian functions.

2.2.4 Image Quality Assessment

Image Quality Assessment (IQA) is a critical aspect of medical imaging, particularly in the context of image reconstruction where the goal is to produce images that are diagnostically useful. IQA provides a means to quantify the performance of reconstruction algorithms. As described in [28], if a ground-truth image is available the assessment relies on a notion of distance between the reconstructed image and the ground-truth image. We typically distinguish quality metrics into global and local categories. Global assessment involves evaluating the full images, while local assessment focuses on specific regions of interest.

Important and widely used global IQA methods are Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). The MSE is a quality metric that computes:

$$\mathrm{MSE}(\mathbf{x}^\dagger, \mathbf{x}_{\mathrm{GT}}) = \frac{1}{N_{\mathbf{x}}} \sum_{i=1}^{N_{\mathbf{x}}} \left([\mathbf{x}^\dagger]_i - [\mathbf{x}_{\mathrm{GT}}]_i \right)^2, \tag{2.19}$$

which is simply the squared ℓ-2 norm $\|\mathbf{x}^\dagger - \mathbf{x}_{\mathrm{GT}}\|_2^2$ and a smaller MSE corresponds to closer agreement of the reconstruction with the ground-truth. The PSNR is defined as:

$$\mathrm{PSNR}(\mathbf{x}^\dagger, \mathbf{x}_{\mathrm{GT}}) = 10 \cdot \log_{10} \left(\frac{\max(\mathbf{x}_{\mathrm{GT}})^2}{\mathrm{MSE}(\mathbf{x}^\dagger, \mathbf{x}_{\mathrm{GT}})} \right), \tag{2.20}$$

where the higher the PSNR indicates higher reconstruction quality. However, an increase in PSNR, or decrease in MSE, may not correspond to an increase in perceptual quality [104]. Perceptual quality is distinct to the error that is assessed with MSE and PSNR, and is concerned with human interpretation of the image [252].

SSIM was developed to better correlate with perceived quality of an image [251], and is defined by:

$$\mathrm{SSIM}(\mathbf{x}^\dagger, \mathbf{x}_{\mathrm{GT}}) = \frac{1}{N_{\mathbf{x}}} \sum_{i=1}^{N_{\mathbf{x}}} l(x_i^\dagger, x_{\mathrm{GT},i}) c(x_i^\dagger, x_{\mathrm{GT},i}) s(x_i^\dagger, x_{\mathrm{GT},i}). \tag{2.21}$$

The functions l, c, and s compute measures of luminance, contrast and structure between the images, we refer the reader to [251] for further detail. Traditional metrics such as MSE and PSNR provide objective, quantitative measures of image quality but often fail to align with human visual perception. SSIM was introduced to address this limitation and attempts to improve correlation with perceived image quality. However, even SSIM does not fully capture the complexities of human visual perception, particularly in medical imaging contexts [29]. Observer studies, where human readers evaluate images for diagnostic tasks, remain the gold standard for IQA, but these studies are time-consuming and resource-intensive. Recent research [29] has shown that newer metrics like HaarPSI demonstrate stronger correlations with human observer performance, particularly for medical images, suggesting a promising alternative to traditional approaches [188].

2.2.4.1 IQA for Medical Image Reconstruction

In addition to the previously discussed IQA methods, specialised approaches exist for assessing the quality of medical image reconstructions. Given that measurements in medical imaging are often affected by noise, the evaluation of noise properties becomes a critical component in understanding overall reconstructed image quality.

For instance, in PET, significant research has been conducted to study methods for quantifying noise characteristics [235]. Moreover, the importance of localised IQA cannot be overstated, as medical images frequently contain region-specific information that is essential for clinical decision-making. Global IQA metrics may fail to detect subtle distortions in localised structures, such as small lesions, that are crucial for accurate diagnosis.

A frequently used local IQA method commonly used for quantification and detectability of lesions is Contrast Recovery Coefficient (CRC). Let $\mathbf{l} \in \mathbb{R}^{N_\mathbf{l}}$ represent the voxels values within the lesion Region Of Interest (ROI), and $\mathbf{d} \in \mathbb{R}^{N_\mathbf{d}}$ represent the voxel values with the background ROI, where $N_\mathbf{l}$ and $N_\mathbf{d}$ denotes the number of voxels in the respective ROIs. The mean value of each ROI is denoted $\bar{l} := \frac{1}{N_\mathbf{l}} \sum_{i \in N_\mathbf{l}} l_i$ and $\bar{d} := \frac{1}{N_\mathbf{d}} \sum_{i \in N_\mathbf{d}} d_i$. The CRC computes the ratio of relative intensities between the reconstructed ROIs and ground-truth image ROIs:

$$\mathrm{CRC}(\mathbf{l}, \mathbf{d}) := \left(\frac{\bar{l}}{\bar{d}} - 1 \right) / \left(\frac{\bar{l}_{\mathrm{GT}}}{\bar{d}_{\mathrm{GT}}} - 1 \right), \tag{2.22}$$

where the subscript GT denotes the ground-truth ROIs.

Noise can be studied in reconstructed images by analysing a background ROI with approximately uniform intensity, denoted as $\mathbf{d}$. The Background Roughness (BR) is then defined as:

$$\mathrm{BR}(\mathbf{d}) := \sqrt{\frac{\sum_{i \in N_\mathbf{d}} (d_i - \bar{d})^2}{N_\mathbf{d}}}. \tag{2.23}$$

A more accurate characterisation of the noise is Ensemble Noise (EN) [235]. This is computed from multiple reconstructions that correspond to multiple realisation of the noise on measurements, i.e. $\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(N_R)} \sim \mathsf{p}(\bar{\mathbf{y}})$ where N_R is the number of realisations. EN is computed on the background ROIs $\mathbf{d}$ as:

$$\mathrm{EN}(\mathbf{d}) := \frac{1}{N_{\mathbf{d}}} \sum_{k=1}^{N_{\mathbf{d}}} \sqrt{\frac{1}{N_R - 1} \sum_{j=1}^{N_R} \frac{(d_k^{(j)} - \bar{d}_k)^2}{\bar{d}_k}}, \tag{2.24}$$

where $d_k^{(j)}$ denotes the k-th element of the j-th realisation of the reconstructed background ROI. Further, the mean $\bar{d}_k$ taken is over realisations $\bar{d}_k = \frac{1}{N_R} \sum_{j=1}^{N_R} d_k^{(j)}$.

Although not strictly for IQA, evaluating the data-consistency of the reconstruction can help evaluate if the reconstruction method is respecting the measurements. For the PET data-consistency, Eqn. (2.14), the Poisson NLL is equivalent to the Kullbeck-Liebler Divergence (KLDIV) up to constant. The KLDIV is given by:

$$\mathrm{KLDIV}(\mathbf{x}) := \sum_{i=1}^{N_{\mathbf{y}}} \left[\bar{y}_i - y_i \log(\bar{y}_i) + y_i\right], \quad \bar{\mathbf{y}} = \mathbf{A}\mathbf{x} + \bar{\mathbf{b}} \tag{2.25}$$

This formulation highlights the discrepancy between the prediction of the measurement mean $\bar{y}$ and measurements, ensuring that the reconstruction respects the underlying measurement statistics. The additive constant y_i does not affect the assessment of consistency but is included for mathematical completeness.

Bias-variance graphs are often used in PET image reconstruction assessment often compares noise suppression properties of the reconstruction and the performance on IQA methods. As such IQA methods are often plotted against an estimate for the noise (BR or EN) in literature.
Further, reconstruction methods often have a tuning parameter that trades-off data-consistency for regularisation, or prior strength, e.g. the variational method see Section 2.1.1. Thus, multiple reconstructions can be obtained with varying degrees of data-consistency. As data consistency increases, the reconstruction begins fit the noise in the measurements, resulting in a higher estimate of noise. Simultaneously, the performance on IQA methods may decrease. This indicates that the reconstruction is under-regularised, as the variance in the measurements outweighs the bias introduced by the reconstruction method. This phenomenon exemplifies the bias-variance trade-off (described in Section 2.1.1). By plotting an estimate for the noise versus IQA methods one can observe the reconstructions methods performance with varying data-consistency, noting that low noise and high performance on IQA methods is sought.

2.3 Deep learning for Medical Image Reconstruction

Deep learning is a subset of the broader field of machine learning, which is underpinned by statistical learning theory [89]. In a broad sense, the chief concept is learning from data, and statistical learning formalises this from a statistical viewpoint. There are several frameworks for learning from data, these include supervised and unsupervised learning [246]. In this section, we start by introducing supervised learning, followed by unsupervised learning specifically deep generative methods. Each of these concepts will be discussed in the context of medical image reconstruction.

In this section we use $\mathbf{y}_{\text{IN}}$ to denote an input to a network, and $\mathbf{x}_{\text{GT}}$ as a ground truth, reference or label.

2.3.1 Supervised Learning

In supervised learning we aim to find a function $\mathcal{G}_{\boldsymbol{\theta}} : \mathcal{Y}_{\text{IN}} \to \mathcal{X}_{\text{GT}}$ with parameters $\boldsymbol{\theta}$, where $\mathbf{y}_{\text{IN}} \in \mathcal{Y}_{\text{IN}}$ and $\mathbf{x}_{\text{GT}} \in \mathcal{X}_{\text{GT}}$ are the input and label respectively.[17] This function $\mathcal{G}_{\boldsymbol{\theta}}$ is an element of the hypothesis space $\mathcal{H}$ and is trained to minimise the expected risk between the prediction $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})$ and label $\mathbf{x}_{\text{GT}}$:

$$r_{\text{expected}}(\mathcal{G}_{\boldsymbol{\theta}}) = \mathbb{E}_{\mathbf{y}_{\text{IN}},\mathbf{x}_{\text{GT}}\sim \mathsf{p}(\mathbf{y}_{\text{IN}},\mathbf{x}_{\text{GT}})}\Phi(\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}), \mathbf{x}_{\text{GT}}),$$

where Φ is a distance often referred to as the loss function or objective function [89]. This is intractable as $\mathsf{p}(\mathbf{y}_{\text{IN}}, \mathbf{x}_{\text{GT}})$ is an unknown joint probability distribution. Instead, assuming access to a paired input-label dataset $\{(\mathbf{y}_{\text{IN}}^{(s)}, \mathbf{x}_{\text{GT}}^{(s)})\}_{s=1}^{N_s}$ where N_s is the number of samples, we use the tractable approximation - empirical risk:

$$r_{\text{empirical}}(\mathcal{G}_{\boldsymbol{\theta}}) = \frac{1}{N_s}\sum_{s=1}^{N_s}\Phi(\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}^{(s)}), \mathbf{x}_{\text{GT}}^{(s)}). \tag{2.26}$$

Empirical risk minimisation is the optimisation of the parameters $\boldsymbol{\theta}$ in order to minimise empirical risk of the model:

[17] The relationship $\mathcal{Y}_{\text{IN}} \to \mathcal{X}_{\text{GT}}$ is typically not a deterministic mapping between spaces, as is the case in inverse problems described in Section 1.1.

$$\boldsymbol{\theta}^{\dagger} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\, r_{\text{empirical}}(\mathcal{G}_{\boldsymbol{\theta}}).$$

The updating of the parameters is referred to as "learning" or "training" the model.

Deep learning defines the function $\mathcal{G}_{\boldsymbol{\theta}}$ as a deep neural network. The Multi-Layer Perceptron (MLP) was one of the first deep neural networks and is composed of multiple "layers" of weights $\mathbf{W}$, biases $\mathbf{b}$ and activation functions α [195]:

$$\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}) = \mathbf{W}^{(N_l)}\alpha(\ldots\alpha(\mathbf{W}^{(2)}\alpha(\mathbf{W}^{(1)}\mathbf{y}_{\text{IN}} + \mathbf{b}^{(1)}) + \mathbf{b}^{(2)})\ldots) + \mathbf{b}^{(N_l)},$$

where N_l is the number of layers, parameters are $\boldsymbol{\theta} = \{(\mathbf{W}^{(l)}, \mathbf{b}^{(l)})\}_{l=1}^{N_l}$, $\mathbf{W}^{(l)} : \mathbb{R}^{(l-1)} \to \mathbb{R}^{(l)}$, $\mathbf{b}^{(l)} \in \mathbb{R}^{(l)}$, and activation function is usually chosen as non-linear $\alpha : \mathbb{R}^{(l)} \to \mathbb{R}^{(l)}$. This structure is an example of a network architecture, where "deepness" arises from the stacking of multiple layers. The architecture can be tailored and fine-tuned to suit the specific requirements of the task at hand, as described in Section 2.3.3.

To minimise the empirical risk $r_{\text{empirical}}(\mathcal{G}_{\boldsymbol{\theta}})$, we need to optimise the parameters $\boldsymbol{\theta}$ of the model. However, due to the non-linear and non-convex nature of deep learning models, finding the global minimum is challenging, if not impossible [76]. Gradient-based optimisation algorithms are employed to iteratively update the parameters in a direction that reduces the loss function. The gradient is computed with the back-propagation algorithm that efficiently computes gradients [129].

The choice of optimisation algorithm has been key to the success of deep learning. Techniques like Stochastic Gradient Descent (SGD) introduce stochasticity by computing gradients on small, randomly selected subsets of the data (batches) [129]. This not only makes the computation more efficient, but also helps the model to generalise better by avoiding over-fitting to the training data. Specifically, the noise introduced by random batches allows SGD to escape sharp minima, which are often associated with poor generalisation, in favour of flatter minima that lead to better performance on unseen data [122]. Further-

more the inherent stochastic nature of SGD prevents the optimisation process from perfectly fitting the training set [85] and discourages memorisation [264].

From a different perspective, the optimisation process also influences the effective hypothesis space $\mathcal{H}$ that the model explores during training. While $\mathcal{H}$ theoretically includes all functions representable by the neural network architecture, practical modelling choices - such as architectural design and optimisation strategy- bias the search to a subspace of the hypothesis space [15, 159]. This phenomenon relates to the concept of inductive bias [18], which refers to the set of assumptions a learning algorithm relies on to make predictions for unseen inputs [16]. By constraining $\mathcal{H}$, inductive biases enable learning algorithms to generalise effectively beyond training data.

The architecture of the neural network is a key component that shapes the inductive bias. In the domain of computer vision, the U-Net architecture has proven to be particularly effective [255]. An overview of the U-Net is presented in Section 2.3.3, that bias the hypothesis toward the properties of medical images as described in Section 2.2.1.

An important optimisation algorithm in deep learning is Adaptive Moment Estimation (Adam) [124], which refines the training process by adaptively adjusting the step size for each parameter based on the first and second moments of the gradients. Adam achieves faster convergence and improved stability by combining the benefits of momentum and adaptive learning rates, where the learning rate acts as a scaling factor for the step size. These characteristics make Adam particularly effective in training deep neural networks, especially in scenarios where the magnitudes of gradients vary significantly between parameters [129].

This interplay between optimisation, model architecture, and data-fitting leads to the critical issue of bias-variance trade-off discussed in Section 2.1.1. Considering the expected prediction error at $\mathbf{y}_{\text{IN}}$:

$$\mathbb{E}[(\mathcal{G}_{\text{true}}(\mathbf{y}_{\text{IN}}) - \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}))^2] = \varsigma^2 + \text{Bias}^2 + \text{Variance},$$

where $\mathcal{G}_{\text{true}}$ is the true function, $\text{Bias} = \mathbb{E}[\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})] - \mathcal{G}_{\text{true}}(\mathbf{y}_{\text{IN}})$ measures the error introduced by model assumptions or restrictions, $\text{Variance} = \mathbb{E}[(\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}) -$

$\mathbb{E}[\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})])]^2$ quantifies model sensitivity to fluctuations in training data, and ς^2 is irreducible noise inherent to $\mathcal{G}_{\text{true}}(\mathbf{y}_{\text{IN}})$ [23, 89]. Under-fitting occurs when the model exhibits high bias, often due to strong restrictions on the hypothesis space, which limits the models ability to capture complex patterns in the data. Conversely, over-fitting arises when the model is overly complex, as excessive flexibility allows it to capture noise in the training data - increasing variance. The balance between bias and variance is influenced by the choice of model complexity, regularisation techniques, and optimisation choices, all of which play a critical role in achieving good generalisation [23].

Returning to medical image reconstruction, deep learning has significantly advanced the field in recent years [6, 92, 187]. Supervised learning in this context uses measurements (or approximate reconstructions) as model inputs $\mathbf{y}_{\text{IN}}$, and the model is trained to predict desired ground-truth images $\mathbf{x}_{\text{GT}}$. It is important to note that these ground-truth images (labels) are typically not available in medical imaging, and instead images of target quality are often used.

2.3.2 Supervised Approaches to Medical Image Reconstruction

Although supervised approaches are often state-of-the-art for medical image reconstruction, in this thesis these approaches are only of peripheral interest due the lack of data-consistency and lack of separation of image modelling from forward modelling [6], as discussed in Section 1.3. This will be explained further in the context of two established supervised methods: post-processing and Learned Primal-Dual (LPD) algorithms. Additionally, we develop these methods for PET in Chapter 4 where network architectures are given in Figs. 4.1 and 4.2 for post-processing and learned primal-dual respectively.

2.3.2.1 Post-Processing

One of the earliest and most straightforward deep learning approaches to medical image reconstruction is post-processing. In this method, a traditional reconstruction algorithm is often first applied to the measurements to produce an approximate image $\tilde{\mathbf{x}}$, this image is then refined using a neural network [109,

169]:[18]

$$\mathbf{x}^{\dagger} = \mathcal{G}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}),$$

where $\mathbf{x}^{\dagger}$ is the resulting reconstruction. The network parameters $\boldsymbol{\theta}$ are trained with empirical risk minimisation (Eqn. (2.26)), with a commonly used MSE loss function:

$$\boldsymbol{\theta}^{\dagger} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \frac{1}{N} \sum_{s=1}^{N_s} \|\mathcal{G}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}^{(s)}) - \mathbf{x}_{\mathrm{GT}}^{(s)}\|_2^2. \tag{2.27}$$

One popular example of a post-processing method is the FBPConvNet [109]. That paper proposed using a U-Net architecture motivating the choice by drawing parallels between the U-Net and traditional wavelet-based denoising methods [192]. The U-Net is ubiquitous in computer vision and it is described further in Section 2.3.3.

While post-processing methods like FBPConvNet are relatively simple and have shown promising results, they have limitations. The primary concern is the lack of data consistency; the network's output $\mathcal{G}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}})$ is not guaranteed to be consistent with the original measurements, $\|\mathbf{y} - \mathbf{A}\mathcal{G}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}})\| \gg \|\varepsilon\|$ where $\|\varepsilon\|$ captures known discrepancies such as noise in the measurements [6].

Furthermore, since the neural network operates solely in the image domain, it does not explicitly incorporate the forward model or the physics of the imaging system. This separation can lead to a dependence of the model on the forward model; where the approximate reconstruction $\tilde{\mathbf{x}}$ exhibits features specific to the forward model. This means that the model is less likely to generalise, for example across scanners or acquisition protocols [1].

The last critical issue is that the U-Net architecture used in post-processing typically has a large number of parameters, which increases the complexity of the model and the amount of training data required to prevent over-fitting. This high data requirement can be a significant limitation in medical imaging applications where paired data is scarce [203].

[18]There exists methods that predict directly from raw measurements [266]. However this requires partially learning known physics that is computationally expensive.

2.3.2.2 Learned Primal-Dual

To address some of the limitations of post-processing methods, the Learned Primal-Dual (LPD) algorithm was proposed [1]. This approach integrates the forward and adjoint operators $\mathbf{A}$ of the imaging system into the neural network architecture, which embeds known physics within the network.

LPD unrolls an iterative reconstruction algorithm into a fixed number of iterates [153], each consisting of learnable modules that process both in the primal (image) and dual (measurement) spaces. These updates were inspired by Primal-Dual Hybrid Gradient algorithm [38]. The method alternates between updating the primal variable $\mathbf{x}^{(k)}$ and dual variable $\mathbf{y}^{(k)}$:

$$\begin{aligned}\mathbf{y}^{(k+1)} &= \mathcal{D}_{\boldsymbol{\theta}}^{(k)}\left(\mathbf{y}^{(k)}, \mathbf{A}(\mathbf{x}^{(k)})\right),\\ \mathbf{x}^{(k+1)} &= \mathcal{P}_{\boldsymbol{\theta}}^{(k)}\left(\mathbf{x}^{(k)}, \mathbf{A}^{*}(\mathbf{y}^{(k+1)})\right),\end{aligned}$$

where the networks $\mathcal{P}_{\boldsymbol{\theta}}^{(k)}$ and $\mathcal{D}_{\boldsymbol{\theta}}^{(k)}$ are the k-th unrolled iterates for the primal and dual spaces respectively. The unrolled algorithm is denoted as:

$$\mathbf{x}^{\dagger} = \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}^{(0)}, \mathbf{x}^{(0)}), \quad \text{where} \quad \mathcal{G}_{\boldsymbol{\theta}} = \mathcal{D}_{\boldsymbol{\theta}}^{(k)} \circ \mathcal{P}_{\boldsymbol{\theta}}^{(k)} \ldots \mathcal{D}_{\boldsymbol{\theta}}^{(0)} \circ \mathcal{P}_{\boldsymbol{\theta}}^{(0)},$$

the initial dual variable is set as the measurements $\mathbf{y}^{(0)} = \mathbf{y}$, the initial value for primal $\mathbf{x}^{(0)}$ can be either zeros or an approximate reconstruction $\hat{\mathbf{x}}$, and the network is trained using the same supervised paradigm as in Eqn. (2.27).

An important advantage of the LPD algorithm is that it uses smaller neural networks for $\mathcal{P}_{\boldsymbol{\theta}}^{(k)}$ and $\mathcal{D}_{\boldsymbol{\theta}}^{(k)}$ compared to the large U-Net used in post-processing methods. These networks are typically shallow and have fewer parameters, reducing the complexity of the model. This simplicity means that the LPD algorithm generally requires less training data to achieve good performance, making it more practical in scenarios where paired data is limited. Furthermore, by explicitly incorporating the forward and adjoint operators, the LPD algorithm enhances the interpretability of the reconstruction process and ensures that the physics of the imaging system are included throughout the reconstruction.

However, the LPD algorithm still relies on supervised training with paired

data, requiring ground-truths from a diverse distribution. Such data is rarely available in practice for medical imaging [92, 187]. Furthermore, the absence of test-time data consistency further restricts its generalisability.

We now turn our attention to outlining the U-Net architecture that is the back-bone for many deep learning approaches in computer vision, including the post-processing approach discussed earlier.

2.3.3 Overview of the U-Net Architecture

The U-Net architecture is a deep convolutional neural network that has become the de facto standard for medical imaging tasks due to its remarkable ability to capture both global context and fine-grained local features [192]. Its name, "U-Net", originates from its distinct symmetric U-shaped design, which is composed of two main components: a contracting path (encoder) and an expanding path (decoder). The encoder progressively extracts hierarchical feature representations through downsampling, while the decoder reconstructs these features into a high-resolution output via upsampling.

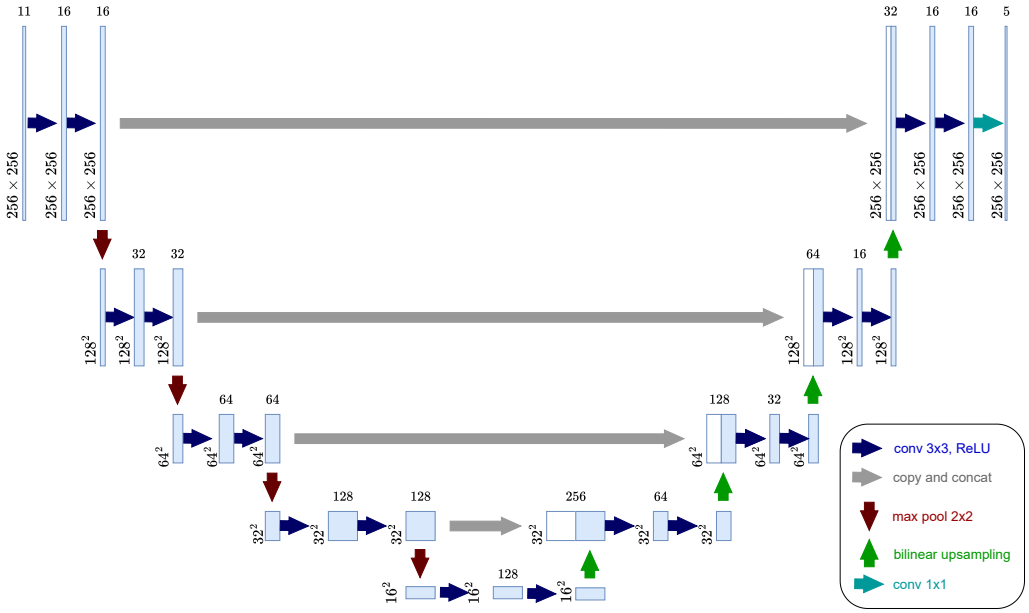


Figure 2.2: An example U-Net architecture illustrating the encoder, decoder, and skip connections.

Unlike traditional encoder-decoder architectures, the U-Net employs convolutional layers to extract hierarchical features across multiple scales. This multi-

scale feature extraction is crucial for medical images, where regions of interest may vary significantly in size and shape. There is a growing body of literature [109, 111, 144, 184, 255, 260] exploring the connections between U-Nets and wavelet decompositions, where wavelets are described in Section 6.2.2. This connection highlights the multi-scale nature of U-Net, and interpretation from a more classical signal-processing perspective.

2.3.3.1 Convolutional Layers

In the U-Net architecture convolutional layers are pivotal for feature extraction. The discrete convolution of a one-dimensional input data x and kernel w can be written as:

$$s(t) = \sum_{\tau} x(\tau)w(t - \tau),$$

where $s(t)$ represents the output features at position t. This formula represents the general discrete convolution operation, which sums the product of two sequences over their overlap. Here, w is a learnable kernel specifically designed for extracting features from x. In deep learning frameworks the operation is typically implemented without flipping the kernel [76]:

$$s(t) = \sum_{\tau} x(\tau)w(t + \tau),$$

which is more akin to cross-correlation. Regardless, this is referred to as a convolutional layer in the deep learning community, and it allows a network to extract useful data-driven kernels that can be used to decompose the input into sets of features [134].

An important property of convolutions is that they are shift-equivariant meaning that shifting the input signal results in a corresponding shift in the output signal. This property ensures that the convolutional operation is consistent across different positions in the input, this exploits the stationarity of natural images discussed in Section 2.2.1.

2.3.3.2 Activation functions

The activation functions introduce non-linearity that allows the network to learn complex patterns. Coupled with convolutional layers, the activation functions allow complex non-linear interactions between the features, which can make them richer and more informative [77]. A popular choice is the Rectified Linear Unit (ReLU) [158]:

$$\mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}. \tag{2.28}$$

This is equivalent to a non-negativity projection onto $\mathcal{P}_{\geq 0}(\cdot)$. This non-linearity is non-differentiable at zero, and creates sparse representations within the network [72]. This can be unfavourable, particularly when there is a bad initialisation or poor gradient step for network parameters that could impede optimisation. As such, there are a more advanced activations such as exponential linear units and scaled exponential linear units that are reported to improve convergence and performance [126].

2.3.3.3 Down-sampling

Down-sampling in the contracting path of the network reduces the spatial dimensions of feature maps, enabling the network to capture relationships across larger regions of the input, taking advantage of scale-invariance and spatial correlations within natural images, see Section 2.2.1. This is referred to a broadening the receptive field of the network and is often achieved using pooling layers or strided convolutions. Pooling layers decrease the spatial dimensions by computing the average or maximum over non-overlapping patches. Strided convolutions reduce spatial dimensions by performing convolution operations with stride k:

$$s(t) = \sum_{\tau} x(\tau) w(k \cdot t + \tau),$$

with a stride greater than one the dimension of input is decreased [263].

After down-sampling and feature extraction through convolutional layers,

the network produces a latent vector denoted as **z**. This latent vector encapsulates the essential information from the input in a compressed, lower-dimensional form [18], effectively distilling the input into the most salient features required for the task. This is often referred to as an information bottleneck and is related to the manifold hypothesis that is discussed in Section 2.3.4.1. Further, the elements of the latent vector can relate to different semantically meaningful components that make up natural images as discussed in Section 2.2.1.

With sufficient convolutions and down-sampling, this representation becomes invariant to small shifts in the input, meaning minor translations do not significantly impact the output. However, complete shift invariance is not guaranteed due to factors such as aliasing, and solutions to address these limitations have been proposed in the literature [265].

2.3.3.4 Up-sampling

Upsampling increases the spatial dimensions in the expanding path of the network. Transposed convolutions have been a popular option in the past, but can introduce chequerboard artifacts [164]. Modern upsampling typically uses nearest-neighbour or bilinear interpolation followed by standard convolutions.

2.3.3.5 Skip Connections

U-Nets have skip connections that link corresponding layers in the encoder and decoder paths. This preserves spatial information lost during down-sampling, which have been shown to be vital for image quality [255]. Skip connections also alleviate the vanishing gradient problem by providing direct pathways for gradient flow [90].

2.3.3.6 Attention Mechanism

Attention mechanisms, broadly, refer to techniques that allow neural networks to dynamically focus on the most relevant parts of their input. Unlike standard convolutions, which process local information within a fixed kernel size, attention mechanisms adaptively model relationships between different spatial or feature dimensions, regardless of distance [247].

In the context of a U-Net architecture, attention can be integrated by incorporating self-attention modules within the skip connections between encoder and decoder layers [165]. Self-attention computes pairwise relationships between all spatial positions in the input feature map to capture long-range dependencies.

Self-attention uses three learnable weight matrices: query $\mathbf{W_Q}$, key $\mathbf{W_K}$, and value $\mathbf{W_V}$, each of dimension $\mathbb{R}^{N_\mathbf{C} \times N_d}$. Given feature maps $\mathbf{X} \in \mathbb{R}^{N_\mathbf{H} \times N_\mathbf{W} \times N_\mathbf{C}}$, where $N_\mathbf{H} \times N_\mathbf{W}$ are the height and width of the input image or patch, and $N_\mathbf{C}$ is the number of channels, we first reshape $\mathbf{X}$ to $\mathbb{R}^{N_p \cdot N_\mathbf{C}}$. with $N_p = N_\mathbf{H} \times N_\mathbf{W}$. The query $\mathbf{Q}$, key $\mathbf{K}$ and value $\mathbf{V}$ are computed as:

$$\mathbf{Q} = \mathbf{X}\mathbf{W_Q}, \quad \mathbf{K} = \mathbf{X}\mathbf{W_K}, \quad \mathbf{V} = \mathbf{X}\mathbf{W_V},$$

resulting in matrices of dimension $\mathbb{R}^{N_p \times N_d}$. The attention score is computed as:

$$\mathbf{S} = \mathbf{Q}\mathbf{K}^\top \in \mathbb{R}^{N_p \times N_p},$$

representing similarities between different spatial positions. The attention map $\mathbf{A}$ is obtained by scaling and normalising the scores:

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{S}}{\sqrt{N_d}}\right),$$

where the softmax function ensures the attention weights sum to one across each row. The self-attention is then computed as:

$$\text{self-attention}(\mathbf{X}) = \mathbf{A}\mathbf{V} \in \mathbb{R}^{N_p \times N_d},$$

which is reshaped back into $\mathbb{R}^{N_\mathbf{H} \times N_\mathbf{W} \times N_d}$.

By integrating self-attention modules, the U-Net architecture effectively models global relationships within input features, enhancing performance by focusing on important regions across the entire image [165]. These modules are used on lower resolution scales of the U-Net-based score-model in Chapter 5 - see Fig. 5.2.

2.3.4 Unsupervised Learning - Deep Generative Models

Generative models attempt to estimate an unknown data distribution, denoted as $\pi(\mathbf{x})$,[19] using a dataset of samples $\left\{\mathbf{x}_{\mathrm{GT}}^{(s)}\right\}_{s=1}^{N_s}$ drawn from that distribution. Since this dataset is unpaired, generative models are typically categorised as unsupervised learning methods. The samples are used to define an empirical (discrete) distribution [23]. This empirical distribution is defined:

$$\mathsf{p}_{\mathrm{empirical}}(\mathbf{x}) = \frac{1}{N_s}\sum_{s=1}^{N_s}\delta(\mathbf{x}-\mathbf{x}_{\mathrm{GT}}^{(s)}), \tag{2.29}$$

where δ is a Dirac delta function that is one if $\mathbf{x}$ is a sample from the dataset $\left\{\mathbf{x}_{\mathrm{GT}}^{(s)}\right\}_{s=1}^{N_s}$ and zero otherwise. The generative model seeks to approximate empirical distribution with a continuous parameterised density $\mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x})$ such that:

$$\mathsf{p}_{\mathrm{empirical}}(\mathbf{x}) \approx \mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x}), \tag{2.30}$$

where $\boldsymbol{\theta}$ are the parameters of the generative model. The goal is that by approximating $\mathsf{p}_{\mathrm{empirical}}(\mathbf{x})$ the underlying true data distribution $\pi(\mathbf{x})$ is also approximated [23].

Traditional generative models, such as Gaussian Mixture Models and Hidden Markov Models, often rely on relatively simple statistical assumptions. These assumptions, while often well-motivated theoretically, can restrict the flexibility of such models, particularly when dealing with high-dimensional data. In these scenarios, complex relationships between features may exist, and traditional models often struggle to capture these intricacies due to the curse of dimensionality [89, 157].

Deep generative models address these limitations by leveraging the representational power of deep neural networks to model complex, non-linear dependencies within the data. Unlike traditional generative models, deep generative models are capable of learning highly intricate patterns, enabling them to capture more nuanced structures in high-dimensional data. This capacity allows deep generative models to generate realistic, high-quality samples, making them especially effective in applications such as image synthesis, natural

[19] This notation is used as this distribution is used as the prior in Chapter 5.

language generation, and other domains where traditional methods often fall short [76].

A central challenge for generative models is balancing accurate representation of the $\mathsf{p}_{\text{empirical}}(\mathbf{x})$ with generalisation to approximate $\pi(\mathbf{x})$. As stated in Eqn. (2.29), the empirical distribution represents the data points as Dirac delta functions, and fitting these exactly can lead to mode memorisation -an over-fitting regime where the model fails to generalise. This phenomenon is closely related to mode collapse, where the model captures only a subset of the data distribution, effectively assigning zero probability to some modes while memorising others [76].

Mode coverage is therefore a key aspect of generative modelling, as it reflects the model's generalisation to capture the full diversity of $\pi(\mathbf{x})$ encompassing all relevant modes. Achieving mode coverage is particularly challenging for high-dimensional data. A fundamental assumption that aids in addressing these challenges is the manifold hypothesis [76].

2.3.4.1 The Manifold Hypothesis

To define the manifold hypothesis, we must first understand the concept of a manifold. Consider the data $\mathbf{x} \in \mathbb{R}^{N_\mathbf{x}}$ where $\mathbb{R}^{N_\mathbf{x}}$ is often referred to as the ambient space. A subset $\mathcal{M} \subseteq \mathbb{R}^{N_\mathbf{x}}$ is a d-dimensional manifold if: For every point $p \in \mathcal{M}$, there is a open neighbourhood $\mathcal{N} \subset \mathbb{R}^{N_\mathbf{x}}$ containing p, and a function $\phi : \mathcal{N} \cap \mathcal{M} \to \mathbb{R}^d$, where ϕ is continuous, bijective, and invertible with continuous inverse (i.e. is a homeomorphism). This definition ensures that $\mathcal{M}$ is locally homeomorphic to the d-dimensional Euclidean space $\mathbb{R}^d$, even if its global structure may be complex or non-linear. Here, d is referred to as the intrinsic dimension of the manifold, while $N_\mathbf{x}$ is the ambient dimension [135].

Intuitively, this means that while $\mathcal{M}$ may be curved or have a complicated shape in the ambient space, any sufficiently small neighbourhood of a point on $\mathcal{M}$ resembles a flat d-dimensional space. The manifold hypothesis assumes that high-dimensional data lie approximately on such a d-dimensional manifold $\mathcal{M}$, where $d \lll N_\mathbf{x}$ [23].

This hypothesis is particularly significant in the context of deep genera-

tive modelling, such as Generative Adversarial Networks (GANs), Variational Auto-Encoders (VAEs) and Score-based Generative Models (SGMs). These models often aim to learn a latent space —a d-dimensional manifold— that captures the essential structure of high-dimensional data.[20] By assuming that data reside on or near a manifold, these models can map low-dimensional latent vector, denoted as $\mathbf{z} \in \mathcal{Z}$, to the high-dimensional data space. This mapping enables the generation of realistic samples while efficiently modelling the underlying data distribution.

2.3.4.2 Generative Adversarial Networks, Variational Auto-Encoders, and Score-based Generative Models

Among many existing methods, three established deep generative models for imaging are GANs, VAEs, and SGMs [55]. The models are denoted as $\mathcal{G}_{\boldsymbol{\theta}}$, where new samples are generated as $\mathbf{x} \sim \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})$, where the input $\mathbf{y}_{\text{IN}}$ is different for each model.

GANs learn the empirical probability distribution by employing a game-theoretic approach [77]. GANs consist of two networks: the generator $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{z})$ and discriminator $\mathcal{D}_{\boldsymbol{\theta}_D}(\mathbf{x})$, where $\mathbf{z}$ is the latent vector that is typically sampled from a tractable prior distribution such as Gaussian $\mathsf{N}(\mathbf{0}, \mathbf{I})$. The generator maps the latent space to the data space, where the latent space can be considered as the lower-dimensional manifold discussed earlier. The discriminator outputs a probability assessing whether the sample is real or generated.

The loss is formulated as a min-max game between generator and the discriminator:

$$\min_{\boldsymbol{\theta}} \max_{\boldsymbol{\theta}_D} \mathbb{E}_{\mathbf{x}_{\text{GT}} \sim \mathsf{p}_{\text{empirical}}}[\log \mathcal{D}_{\boldsymbol{\theta}_D}(\mathbf{x}_{\text{GT}})] + \mathbb{E}_{\mathbf{z} \sim \mathsf{p}_{\mathbf{z}}}[\log(1 - \mathcal{D}_{\boldsymbol{\theta}_D}(\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{z})))],$$

where $\mathsf{p}_{\mathbf{z}}$ is often Gaussian distribution. The generator attempts to minimise

[20] For SGMs, there is no explicit lower-dimensional latent space. However, the stochastic process discussed later provides a scale-space of simplified densities, which can be interpreted as an implicit latent space [97, 228]. That said, latent diffusion models explicitly incorporate latent vectors [191]. Note that, in practice, for all generative models, the latent spaces are rarely lower-dimensional than the ambient dimension.

discriminator's ability to distinguish between real and generated sampled, while the discriminator seeks to maximise this distinction.

GANs showed great empirical success in producing highly realistic images. Specifically the StyleGAN family of models [116–119] which are an advanced extension of GANs that introduce disentanglement in the latent space by conditioning the generation process on separate style and noise components, offering greater control over specific features in the generated samples. Despite this success they are prone to mode collapse, and as a result are notoriously hard to train [197].

GANs are characterised as implicit models as they do not explicitly model a likelihood function for the data, but instead learn the distribution through the adversarial interaction between generator and discriminator [77].

VAEs also use a latent vector $\mathbf{z}$ to generate samples from $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{z})$ [125], while the training maximises the likelihood of the data, $\mathsf{p}(\mathbf{x}) = \int \mathsf{p}(\mathbf{x}|\mathbf{z})\mathsf{p}(\mathbf{z})d\mathbf{z}$. Since direct optimisation of $\mathsf{p}(\mathbf{x})$ is intractable due to the integral over $\mathbf{z}$, the Evidence Lower Bound (ELBO) provides a tractable variational approximation instead [125]:

$$\log \mathsf{p}(\mathbf{x}) \geq \text{ELBO}_{\phi,\boldsymbol{\theta}}(\mathbf{x}) := \mathbb{E}_{\mathsf{q}_{\phi}(\mathbf{z}|\mathbf{x}_{\text{GT}})}[\log \mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})] - \rho_{\text{KLDIV}}(\mathsf{q}_{\phi}(\mathbf{z}|\mathbf{x})||\mathsf{p}(\mathbf{z})),$$

where $\mathsf{q}_{\phi}(\mathbf{z}|\mathbf{x}) \approx \mathsf{p}(\mathbf{z}|\mathbf{x})$, $\mathsf{p}(\mathbf{z})$ is tractable distribution - usually Gaussian, and ρ_{KLDIV} is the KLDIV. The first component of the ELBO $\mathbb{E}_{\mathsf{q}_{\phi}(\mathbf{z}|\mathbf{x})}[\log \mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})]$ ensures the generated images are of high fidelity, and $\rho_{\text{KLDIV}}(\mathsf{q}_{\phi}(\mathbf{z}|\mathbf{x})||\mathsf{p}(\mathbf{z}))$ ensures the latent space aligns with the prior that ensures mode coverage. Thus, VAEs can be considered capable of covering all the modes, generating diverse samples [125]. Unfortunately, the generated samples are often blurry and unrealistic [257]. This issue can be remedied by adapting the sampling process, using advanced architectures, or incorporating more expressive priors [243]. However, the inability to generate high-quality images remains the predominant challenge for VAEs.

SGMs model the data distribution through the score function - the gradient of the log-density $\nabla_{\mathbf{x}} \log \mathsf{p}(\mathbf{x})$ - rather than directly modelling $\mathsf{p}(\mathbf{x})$ itself [97, 222, 230]. This avoids the computation of a normalising constant that is often computationally intractable. To illustrate this consider the parameterised density:

$$\mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{e^{-\mathsf{f}_{\boldsymbol{\theta}}(\mathbf{x})}}{Z},$$

where $\mathsf{f}_{\boldsymbol{\theta}}$ is an energy function that parameterises the unnormalised log-density, Z is the normalisation ensuring $\int \mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x}) d\mathbf{x} = 1$. Maximising the parameterised density requires computing the normalisation constant, which involves integrating over $\mathbf{x}$. However, by taking the gradient of the log-density, the normalisation constant is no longer required:

$$\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}) = \nabla_{\mathbf{x}} \log \mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x}) = -\nabla_{\mathbf{x}} \log \mathsf{f}_{\boldsymbol{\theta}}(\mathbf{x}) - \underbrace{\nabla_{\mathbf{x}} \log Z}_{=0}.$$

Unfortunately, the score $\nabla_{\mathbf{x}} \log \mathsf{p}_{\boldsymbol{\theta}}(\mathbf{x})$ is not directly accessible. Instead, the gradient of the log-density of a tractable noise-corrupted distribution $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t)$ is used, where $\mathbf{x}_t$ represents a noisy version of the data.[21] Sampling from an SGMs is an iterative denoising process that progressively refines the noisy data $\mathbf{x}_t$ to sample from $\pi(\mathbf{x})$. This approach is computationally expensive, as it requires numerous evaluations of the score-model at different noise levels. We provide further details in Chapter 5.

These models were introduced to highlight the trilemma between three key properties necessary for real-world adoption of deep generative models [257], namely:

- **High-quality sampling**: Samples must be realistic and visually convincing.
- **Mode coverage and sample diversity**: The generated samples should re-

[21] **Notation disclaimer:** A probability density can be denoted p but for SGMs an additional subscript is used to emphasize a time dependence p_t that is associated with random vector $\mathbf{x}_t$. The gradient operator ∇ uses a subscript to denote the gradient with respect to that set of variables, i.e. $\nabla_{\mathbf{x}_t} f(\mathbf{x}_t) = \frac{\partial f}{\partial \mathbf{x}_t}$. The random vector $\mathbf{x}_t$ belongs to a stochastic process that is indexed in time t, and for notational simplicity the distinction between a random vector and realisation from the distribution of a random vector is omitted such that $\mathbf{x}_t \sim \mathsf{p}_t(\mathbf{x}_t)$ as is common in literature [230].

flect the entire data distribution, including regions of low density, to ensure diversity.

- **Computationally efficiency**: Sampling should be fast and computationally inexpensive.

This trade-off can be visualised diagrammatically in Fig. 2.3.

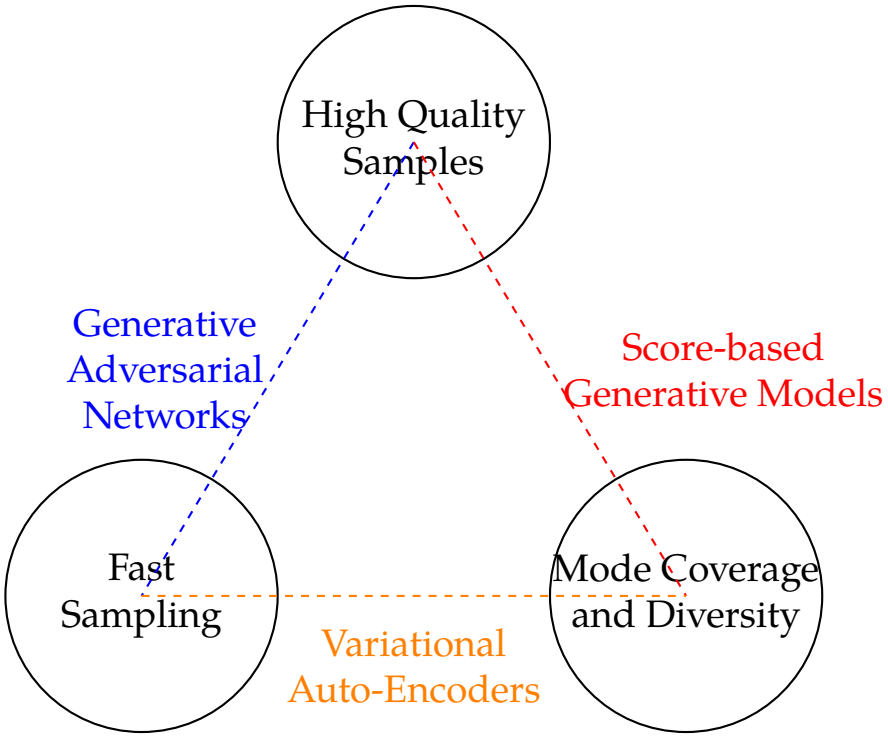


Figure 2.3: Trilemma diagram, recreated and mildly edited from [257].

Recent advancements in accelerated sampling techniques for SGMs have significantly improved their efficiency, making them the de facto standard for image generation [146]. These improvements address computational constraints while maintaining high-quality sampling and mode coverage, positioning SGMs as a versatile solution to the trilemma.

2.3.5 Unsupervised Approaches for Medical Image Reconstruction

In recent years, unsupervised learning approaches have gained significant attention in medical image reconstruction due to their ability to leverage inherent data structures without the need for large paired datasets [6]. The motivations

for this thesis, discussed in detail in Section 1.3, closely align with these advancements. Specifically, the approaches of interest integrate data consistency with known physics at test time and are agnostic to the forward model. This flexibility allows the same method to be applied across different forward models (i.e. different scanners), enhancing its generalisability.

Unsupervised approaches have shown promise in reconstructing high-quality images from undersampled or corrupted measurements. According to the taxonomy presented in [55], these deep generative models can be broadly categorised into trained and untrained models. Trained models require an unpaired dataset of ground-truth images $\{\mathbf{x}_{\text{GT}}^{(s)}\}_{s=1}^{N_s}$ for learning the underlying data distribution, whereas untrained models do not rely on any training data and instead only exploit inductive bias of the network.

Using the taxonomy of [55], this thesis focuses with two deep generative models; Deep Image Prior (DIP), an untrained deep generative model, and SGMs, a trained deep generative model. Another notable unsupervised approach involves using Implicit Neural Representations (INRs) for medical image reconstruction, where the image is represented as a continuous function. This approach relates to work in this thesis that also aims to reconstruct a continuous function.

As these approaches are central to this thesis, we will briefly describe them here, with additional literature discussed in the respective chapters.

2.3.5.1 Deep Image Prior

DIP parameterises an image as an output of a U-Net, denoted as $\mathcal{G}_{\boldsymbol{\theta}}$ [242]. DIP utilises the inductive biases of the U-Net architecture to implicitly regularise the reconstruction process. The parameters of the network are randomly initialised and input $\mathbf{y}_{\text{IN}}$ is typically static random noise meaning that no training data is required. The image is parameterised as $\mathbf{x} = \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})$, and reconstruction process

$$\boldsymbol{\theta}^{\dagger} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \rho(\mathbf{y}, \mathbf{A}\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})), \quad \text{where} \quad \mathbf{x}^{\dagger} = \mathcal{G}_{\boldsymbol{\theta}^{\dagger}}(\mathbf{y}_{\text{IN}}) \tag{2.31}$$

where $\mathbf{x}^\dagger$ is the reconstruction, ρ is the data-consistency term -see Section 2.2.3, $\mathbf{A}$ is the forward operator, and $\mathbf{y}$ are the measurements.

The U-Net is typically over-parameterised, meaning that DIP can fit most images, including noise. The key aspect of DIP is its inductive bias, which suppresses noise during optimisation, allowing natural features to be recovered first [242]. Therefore, it is crucial to stop the optimisation early, before the model begins to fit the noise -this is often referred to as early-stopping. For further literature and development of DIP for PET image reconstruction see Chapter 3.

2.3.5.2 Score-based Generative Models

SGMs can be used to solve inverse problems by integrating data consistency into the iterative denoising process [46, 107, 254]. More specifically, the trained score-model $s_{\boldsymbol{\theta}}$ approximates a data distribution π. From the Bayesian perspective of statistical inverse problems (see Section 2.1.2.2), this distribution serves as a prior. The relationship can be expressed as $\nabla_{\mathbf{x}} \log p_0(\mathbf{x}_0) = \nabla_{\mathbf{x}} \log \pi(\mathbf{x})$, where the subscript denotes that the equivalence with the prior holds only when the image is free of noise.

SGMs applied to inverse problems represent a burgeoning area of research [46, 254]. In Chapter 5, we explore the application of SGMs to PET image reconstruction and situate this work within the context of relevant literature.

2.3.5.3 Implicit Neural Representations

The final approach of note is the Implicit Neural Representations (INRs). INRs, also known as coordinate-based neural networks, represent signals such as images or volumes as continuous functions parameterised as a MLP, $\mathcal{G}_{\boldsymbol{\theta}}$ [221]. In the context of image reconstruction, INRs can model an image as a continuous function, $x(\mathbf{r})$, where $\mathbf{r}$ represents the spatial coordinate.

As discussed in Section 2.2.1, medical images are continuous functions with specific properties. INRs, like DIP, utilise inductive bias toward those properties to obtain high-quality reconstructions [210]. The reconstruction is obtained from:

$$\boldsymbol{\theta}^{\dagger} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \rho(\mathbf{y}, \mathbf{A}\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{r})|_{\Omega_{\mathbf{r}}}), \quad \mathbf{x}^{\dagger} = \mathcal{G}_{\boldsymbol{\theta}^{\dagger}}(\mathbf{r})|_{\Omega_{\mathbf{r}}},$$

where $\Omega_{\mathbf{r}}$ are the coordinates where the image is sampled, therefore $\mathcal{G}_{\boldsymbol{\theta}^{\dagger}}(\mathbf{r})|_{\Omega_{\mathbf{r}}}$ is an image.

In Chapter 6 this continuous "functional" approach to image representation is developed for image reconstruction with Adaptable Blobs (A-Blobs). Although not precisely an INR, the proposed representation bares resemblance, with many of the same challenges.

Chapter 3

Deep Image Prior

3.1 Introduction

In this chapter we develop DIP for PET image reconstruction and improve reconstruction stability and quality when measurements are low-count. The results presented are published in works [215, 218]. As described in Section 2.3.5.1, DIP is an unsupervised deep learning method that was originally proposed by Ulyanov et al, for image denoising and restoration [242]. The core concept is to parameterise the image as the output of a U-Net [192] architecture: $\mathbf{x} = \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})$, where parameters are updated according to Eqn. (2.31). DIP is untrained and the network is initialised randomly, where the input $\mathbf{y}_{\text{IN}}$ is often a static random image [242]. To regularise the inverse problem, DIP relies on the inductive bias of the U-Net, see Section 2.3.3 for more details.

DIP's ability to fit natural images before noise has been studied extensively and is partially attributed to spectral bias toward low spatial-frequencies [93, 212]; as natural images have a fast decaying power spectrum, i.e. more information is present in the lower frequencies [214], see Section 2.2.1. However, DIP is not merely a low-pass filter [57, 232], and completely removing high frequencies is undesirable because they encode important features such as image edges [173]. The inductive bias of the U-Net is more nuanced and can be tuned, which is described in Section 2.3.3.

Early-stopping is necessary to prevent fitting to noise with DIP. The early-stopping criterion is typically chosen through hyper-parameter selection on a

validation dataset. More recently a wide variety of other criterion have been proposed, for example tracking image-based variance metrics to indicate stopping [250].

The choice of optimisation algorithm also influences the inductive bias of the method, this referred to as implicit regularisation see Section 2.3. As analysed in [232], the Adam optimiser [124] causes larger changes in the hidden layers, and it was reported this enabled more informative features to be extracted from $\mathbf{y}$ earlier in the architecture, leading to Adam providing better reconstructions than SGD. In literature Adam is predominate choice for optimiser used in DIP [57, 232, 242].

3.2 DIP for Tomographic Inverse Problems

Tomographic inverse problems, where measurements are projections of the sought object, were not addressed in the original DIP work [242]. The first applications of DIP for tomographic problems are for CT and PET [9, 74]. In Baguer et al [9], additional regularisation is included within the objective function:

$$\boldsymbol{\theta}^{\dagger} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left[\Phi := \rho(\mathbf{y}, \mathbf{A}\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\mathrm{IN}})) + \lambda\eta(\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\mathrm{IN}}))\right], \quad \mathbf{x}^{\dagger} = \mathcal{G}_{\boldsymbol{\theta}^{\dagger}}(\mathbf{y}_{\mathrm{IN}}). \tag{3.1}$$

The input to the network $\mathbf{y}_{\mathrm{IN}}$ is static random noise, not the measurements $\mathbf{y}$. This objective function is in-line with classical variational approaches to inverse problems, see Section 2.1.1. The regularisation $\eta(\cdot)$ in Baguer et al, was discrete TV given by Eqn. (2.18). It was found that the regularisation was able to alleviate the need for robust stopping criteria, which is critical to prevent DIP over-fitting to noise. The regularisation effect was two-fold; regularisation via explicit TV regularisation, and implicit regularisation via early stopping of the optimisation with the inductive bias of the U-Net. For their optimisation they utilised Adam, which required access to the gradient $\frac{\partial \Phi}{\partial \boldsymbol{\theta}}$, this can be simply expanded to:

$$\frac{\partial \Phi}{\partial \boldsymbol{\theta}} = \frac{\partial \Phi}{\partial \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})} \cdot \frac{\partial \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})}{\partial \boldsymbol{\theta}}, \quad \text{where} \quad \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}) = \mathbf{x}. \tag{3.2}$$

The gradient $\frac{\partial \Phi}{\partial \mathbf{x}}$ is computed in established modern regularisation algorithms described in Section 2.2.3. Thus, modern reconstruction engines are able to compute these efficiently. While on the other-hand, the gradient computed from the output of the generator to network parameters $\frac{\partial \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})}{\partial \boldsymbol{\theta}}$ are obtained through automatic differentiation. Automatic differentiation is the back-bone of software packages for deep learning.

This gradient, Eqn. (3.2), therefore requires both the automatic differentiation and advanced model-based reconstruction techniques. In order to directly update the parameters of a network based on gradients computed from reconstruction engines one must integrate the reconstruction engine into a deep learning framework. In Bauger et al [9], this was already available in the software used. Namely, they used the Operator Discretization Library (ODL) reconstruction engine with CT models within the `PyTorch` deep learning framework [2, 171], this is elaborated upon in Section 3.2.1.

The application of DIP to PET Gong et al, was more involved due to the lack of integration of the PET forward model with a deep learning framework [74]. Additionally, they omitted the regularisation in Eqn. (3.1), and posed the problem as a constrained optimisation that is equivalent to Eqn. (2.31):

$$\boldsymbol{\theta}^{\dagger} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \rho_{\text{PET}}(\mathbf{y}, \mathbf{A}\mathbf{x} + \bar{\mathbf{b}}) \quad \text{s.t.} \quad \mathbf{x} = \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}),$$

with the PET forward model $\mathbf{A} \cdot + \bar{\mathbf{b}}$ as described in Section 2.2.2.1.

The constrained optimisation was written in the following augmented Lagrangian form:[1]

$$\rho_{\text{PET}}(\mathbf{y}, \mathbf{A}\mathbf{x} + \bar{\mathbf{b}}) + \frac{\beta}{2}\|\mathbf{x} - \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}) - \mathbf{u}\|^2 + \frac{\beta}{2}\|\mathbf{u}\|^2,$$

where $\mathbf{u}$ is the Lagrange multiplier and $\beta > 0$ is a hyper-parameter used to

[1] The standard augmented Lagrangian is usually written as $\rho_{\text{PET}}(\mathbf{y}, \mathbf{A}\mathbf{x} + \bar{\mathbf{b}}) + \mathbf{u}^{\top}(\mathbf{x} - \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})) + \frac{\beta}{2}\|\mathbf{x} - \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}})\|^2$. Gong et al opt to write the augmented Lagrangian in a non-standard format.

enforce the constraint and prevent $\mathbf{u}$ from becoming too large. The Alternating Direction Method of Multipliers (ADMM) was used to optimise it and the variable updates are given by:

$$\mathbf{x}^{(k+1)} = \underset{\mathbf{x}}{\operatorname{argmin}}\, \rho_{\text{PET}}(\mathbf{y}, \mathbf{A}\mathbf{x} + \bar{\mathbf{b}}) + \frac{\beta}{2}\|\mathbf{x} - \mathcal{G}_{\boldsymbol{\theta}^{(k)}}(\mathbf{y}_{\text{IN}}) + \mathbf{u}^{(k)}\|^2, \tag{3.3a}$$

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \|\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}) - (\mathbf{x}^{(k+1)} - \mathbf{u}^{(k)})\|^2 \tag{3.3b}$$

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{x}^{(k+1)} - \mathcal{G}_{\boldsymbol{\theta}^{(k+1)}}(\mathbf{y}_{\text{IN}}). \tag{3.3c}$$

ADMM splits the optimisation into a image reconstruction update and a deep learning parameter update, and alleviates the need to integrate the reconstruction engine within a deep learning framework. The first sub-problem Eqn. (3.3a) is a classical image reconstruction-type problem that is used in conventional model-based reconstruction methods. In fact it is a MAP estimate and it was solved using a modified EM update such as BSREM. The second sub-problem Eqn. (3.3b) is a problem that can be solved within deep learning frameworks. In Gong et al [74], a variety of algorithms were tested to solve Eqn. (3.3b) and it was determined that Limited-Broyden–Fletcher–Goldfarb–Shanno algorithm performed best. Furthermore, the tuning of β was vital to ensure stability of the method. It would therefore be beneficial to not split the optimisation and pose it as a single optimisation without this additional hyper-parameter β. To do this the forward operator needs to be integrated into a deep learning framework.

In Hashimoto et al [88], DIP was implemented as a single optimisation problem, thus reducing the number of hyper-parameters and the computational overhead; simplifying implementation. The forward model was stored as a sparse matrix that had an excessive GPU memory overhead. Furthermore, MSE was used as the data-consistency term, which is not appropriate for the noise affecting PET measurements, see Section 2.2.3.1. Their work was recently extended to 3D PET through slicing the forward operator and solving with a subset-based block iterative approach, as well as including additional regular-

isation, although this study did not compare with traditional MAP reconstructions [87].

3.2.1 Framing Medical Image Reconstruction in PyTorch

There exists a variety of medical image reconstruction engines that can be used for model-based reconstruction, including ODL which uses the ASTRA toolbox projectors for CT [2, 244], Gadgetron for MRI [83, 258], and Software for Tomographic Image Reconstruction (STIR) for emission tomography modalities such as PET [233].

Large-scale established reconstruction methods are often first-order optimisation methods. The computation of these first-order derivatives require access to the forward-projection, i.e. the application of $\mathbf{A}$ to $\mathbf{x}$, and the back-projection, which is the adjoint $\mathbf{A}^* \equiv \mathbf{A}^\top$ for PET.

First order optimisation methods are also the primary algorithms used in deep learning optimisation, where gradients are computed using automatic differentiation. Briefly, as an example, consider automatic differentiation of gradients in `PyTorch` via the chain-rule from objective function Φ:

$$\frac{\partial \Phi}{\partial \boldsymbol{\theta}} = \frac{\partial \rho(\mathbf{A}\mathbf{x} + \bar{\mathbf{b}})}{\partial \mathbf{x}} \cdot \frac{\partial \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\mathrm{IN}})}{\partial \boldsymbol{\theta}} + \lambda \frac{\partial \eta(\mathbf{x})}{\partial \mathbf{x}} \cdot \frac{\partial \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\mathrm{IN}})}{\partial \boldsymbol{\theta}}. \tag{3.4}$$

The derivative of the data-consistency term requires access to the adjoint of the forward operator: $\frac{\partial \rho(\mathbf{A}\mathbf{x}+\bar{\mathbf{b}})}{\partial \mathbf{x}} = \mathbf{A}^\top \frac{\partial \rho(\bar{\mathbf{y}})}{\partial \bar{\mathbf{y}}}$ where $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x} + \bar{\mathbf{b}}$. From this it can be observed that, given a linear operator $\mathbf{A}$, `PyTorch` requires access to the adjoint of the linear operator to back-propagate to the image $\mathbf{x}$.

To correctly integrate a linear operator within `PyTorch` requires allowing the auto-differentiation library access to the application of the linear operator and the application of the adjoint linear operator. This is done by extension of `torch.autograd` library with a custom `torch.autograd.Function`. This is what was done in ODL [2], to allow Baguer et al [9], to update DIP parameters based on CT measurements as a single optimisation.

In this work, rather than wrapping the forward operator, the full function and gradient can be wrapped; Φ and $\frac{\partial \Phi}{\partial \mathbf{x}}$ respectively. This is particularly im-

portant in the case of the PET objective function, where the Poisson NLL is convex but not Lipschitz continuous as noted in Section 2.2.3.1. It is therefore important to correctly handle the objective function to ensure stability of the optimisation. For this reason the SIRF objective function is directly wrapped, and used to back-propagate and update the network parameters $\boldsymbol{\theta}$.

3.3 Objectives

The application of deep learning to PET image reconstruction is challenging due to a lack of integration of the PET forward model into deep learning frameworks. This results in the need to formulate the problem via split-optimisation [74, 169], where the objective function is split into sub-problems such as EM with the forward model, and denoising with the network. Although these methods often utilise the same objective function, tuning of hyper-parameters is needed to ensure stability and reasonable convergence rates. An alternative approach integrates the forward model into the deep learning framework, and automatic differentiation can be applied from the full-objective function Φ through the forward model to the network parameters $\boldsymbol{\theta}$. The parameters are then updated via the gradient descent algorithm. This single optimisation allows for a simplified implementation, which reduces the number of hyper-parameters and the computational overhead.

The first objective of this chapter is implementing and testing the objective function wrapper, in a similar manner to Baguer et al. The wrapper integrated SIRF a multi-modality reconstruction framework into `PyTorch` an established deep learning framework. SIRF utilises the PET reconstruction engine STIR [233]. STIR benefits from a parallelised C++ backend, access to GPU-based projector [207], and clinically relevant PET regularisation penalties. Through using the SIRF objective, direct comparison between traditional reconstruction and DIP reconstruction was made.

The further objectives were scaling to fully 3D reconstruction and investigating the use of additional regularisation to alleviate the need for early-stopping.

3.4 Experimental Methods

In this study we consider the objective function of the form given in Eqn. (3.1). The PET data-consistency term ρ is the Poisson NLL while both QP and RDP regularisation were tested as regularisation,these are described in Section 2.2.3.1.[2] The RDP edge-preservation parameter was set to $\gamma = 2$ for all tests, this is in-line with clinical settings.

3.4.1 Dataset

The dataset comprised of a single voxelised XCAT torso phantom [209], with activity concentrations representative of an FDG study. Measurements were obtained from numerical GATE photon emission simulation of back-to-back 511 keV photon emissions [108]. The simulated scanner was a General Electric Discovery 690 [22] with uniform crystal efficiencies, and the simulation was performed via the STIR-GATE-Connection [239] where normalisation, attenuation, randoms and scatter were estimated from the Monte-Carlo data and incorporated within the forward model, as described in Section 2.2.2.1. Cylindrical hot lesions were inserted into the abdominal wall, liver, lung and spine. Note that cardiac motion, respiratory motion, and radioactive decay were not modelled.

In 2D, the data were binned into sinograms with 288 projection angles extracted for a 1 ring, and for 3D all 24 rings were used with all ring differences. The 3D data was also split into datasets containing 250 ("lower") and 1200 ("higher") million coincidence events, and the reconstruction volume had dimensions of $47 \times 128 \times 128$ with voxel-size $3.27 \times 4.0 \times 4.0$mm.

3.4.2 Established Reconstruction Methods

The established reconstruction algorithms were different in 2D and 3D cases. In 2D, the ML estimate was computed with OSEM, Eqn. (2.15), while Green's Ordered Subset Maximum A Posteriori One Step Late (OSMAP-OSL) [78, 79] was used to compute MAP estimates with QP and RDP, denoted OSMAP-OSL-

[2] For this study STIR's RDP was exposed within SIRF.

QP and OSMAP-OSL-RDP respectively. The OSMAP-OSL is a simple adaption of OSEM given by:

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{x}^{(k)}}{\mathbf{A}^\top \mathbf{1} + \lambda \nabla_{\mathbf{x}} \eta(\mathbf{x})} \mathbf{A}_{s(k)}^\top \left(\frac{\mathbf{y}}{\mathbf{A}_{s(k)} \mathbf{x}^{(k)} + \bar{\mathbf{b}}_{s(k)}} \right),$$

where $\nabla_{\mathbf{x}} \eta(\mathbf{x})$ is the gradient of the regulariser with respect to the image, and other terms are as before, see Eqn. (2.15). This algorithm is non-convergent for more than one subset, as such it was only used in the 2D case, where the operator and measurements were split into four subsets. Results were given after 200 epochs of updates.

In 3D, we utilise the BSREM algorithm, Eqn. (2.17), to compute the MAP estimate with RDP and 34 subsets. This is denoted BSREM-RDP, and the relaxation parameter η was chosen to ensure convergence within 1000 epochs, where an image based convergence metric was used.

3.4.3 Image Quality Assessment

As ground-truth images were known we used standard PET methods for IQA. Specifically we used CRC and BR, see Eqns. (2.22) and (2.23) respectively. On the CRC-BR graphs the further to the top-left corresponds to better reconstructions that correspond to better noise suppression whilst recovering the contrast of the lesion. In 2D multiple regularisation parameters were varied λ for both DIP approaches and comparison methods. Evaluating the DIP approach after the approximate convergence enables us to observe the agreement of established methods and DIP; validating the wrapping of the objective function.

In 3D, the DIP approaches were evaluated throughout the optimisation trajectory and compared to converged MAP estimates. From this we can observe whether DIP is able to out-perform established methods independent of the choice of stopping criterion.

3.4.4 Proposed DIP for PET Methods

Three DIP approaches were tried in 2D: DIP without additional regularisation PET-DIP, with QP regularisation PET-DIP+QP, and with RDP regularisation

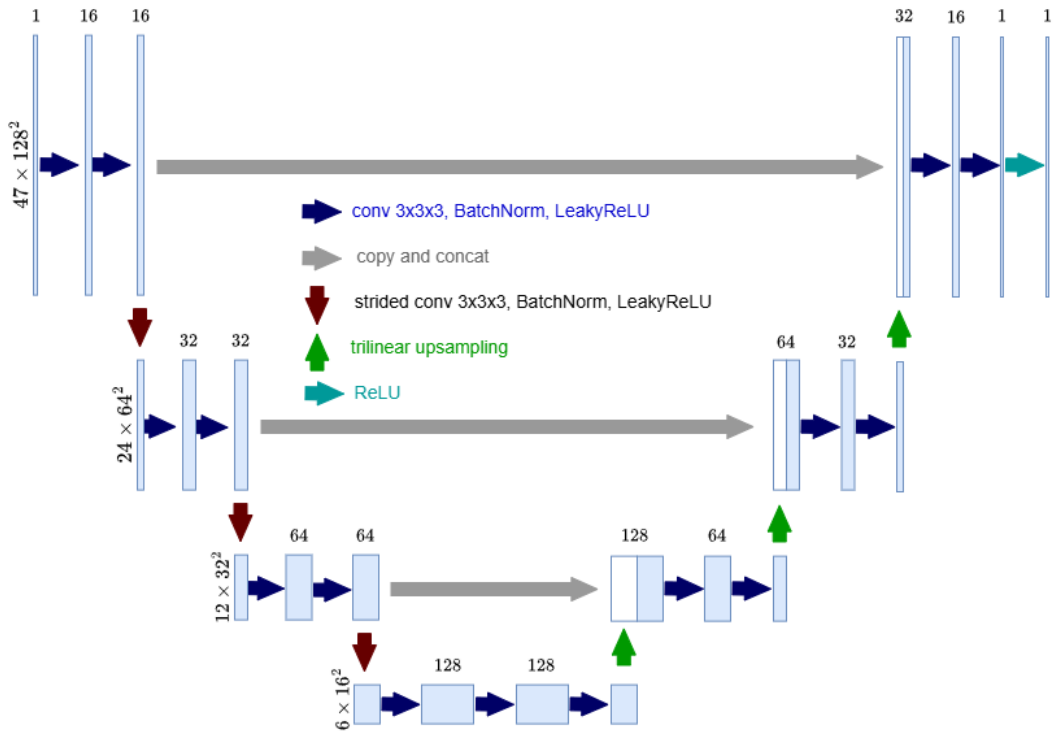


Figure 3.1: U-Net used for the 3D PET-DIP.

PET-DIP+RDP. In 3D only PET-DIP and PET-DIP+RDP were tested.

For all studies a U-Net [192] architecture was used, see Section 2.3.3. The network is over-parameterised utilising an architecture that used linear-upsampling and strided convolutions to change scale, with the number of features compensating for the increase/decrease of spatial dimensionality. Batch normalisation and Leaky ReLU were included after each convolution. Skip connections were also present between encoding and decoding paths of the network. A ReLU was used on the network output as a non-negativity constraint. The network in 3D was unaltered from previous work [74, 87] as shown in Fig. 3.1, while two network choices - 2 and 6 scale U-Nets - were evaluated in 2D. The termination criterion was fixed at 50,000 epochs in 2D, and 20,000 epochs in 3D. Optimisation was performed with the first-order Adam optimiser. In 3D an initial learning rate of 1.0 was cosine annealed to 0.0 over the 20,000 epochs.

The following Table 3.1 summarises the methods in this study.

Approach Name	Optimisation Algorithm Used	Explicit Regulariser	2D or 3D
MLE	OSEM	None	2D
PET-DIP	Adam	None	Both
PET-DIP+QP	Adam	QP	2D
OSMAP-OSL-QP	OSMAP-OSL	QP	2D
PET-DIP+RDP	Adam	RDP	Both
OSMAP-OSL-RDP	OSMAP-OSL	RDP	2D
BSREM-RDP	BSREM	RDP	Both

Table 3.1: Algorithms used in 2D and 3D

3.5 Results and Discussion

In 2D, a feasibility study was conducted and the obtain results compared DIP with classical MAP estimates computed with OSMAP-OSL. In 3D, the network architecture is brought in-line with published work on applying DIP for PET image reconstruction [74, 87], and results are presented for two noise levels. Instead of using an early-stopping criteria in 2D, the network was run for a fixed number of iterations. In 3D, the quality metrics were monitored along the optimisation trajectory.

3.5.1 2D Results

The wrapper was tested with PET-DIP, PET-DIP+QP and PET-DIP+RDP with two architectures of 2-scale and 6-scale. In Fig. 3.2 we see solutions of the different regularisation strengths, the λ values used are $\lambda \in \{20, 10, 7.5, 5, 2.5, 1\}\times 10^{-4}$ for OSMAP-OSL-QP and PET-DIP+QP, this is similarly presented for results with RDP with $\lambda \in \{20, 10, 7.5, 5, 2.5, 1\} \times 10^{-2}$. Given the sweeps of λ, it can be seen that RDP is better at suppressing noise (lower BR) while retaining a high lesion detectability (high CRC). This is unsurprising as RDP was hand-crafted to have edge-preserving properties [161]. Qualitatively, the effect of DIP can be observed in Fig. 3.3. The first row compares approximate ML estimates parameterised as an image and as the output of the U-Net. From the inset, DIP can be observed to have smoother solutions, indicating the noise impedance properties of DIP.

DIP was run for 50,000 epochs and it was observed that the objective func-

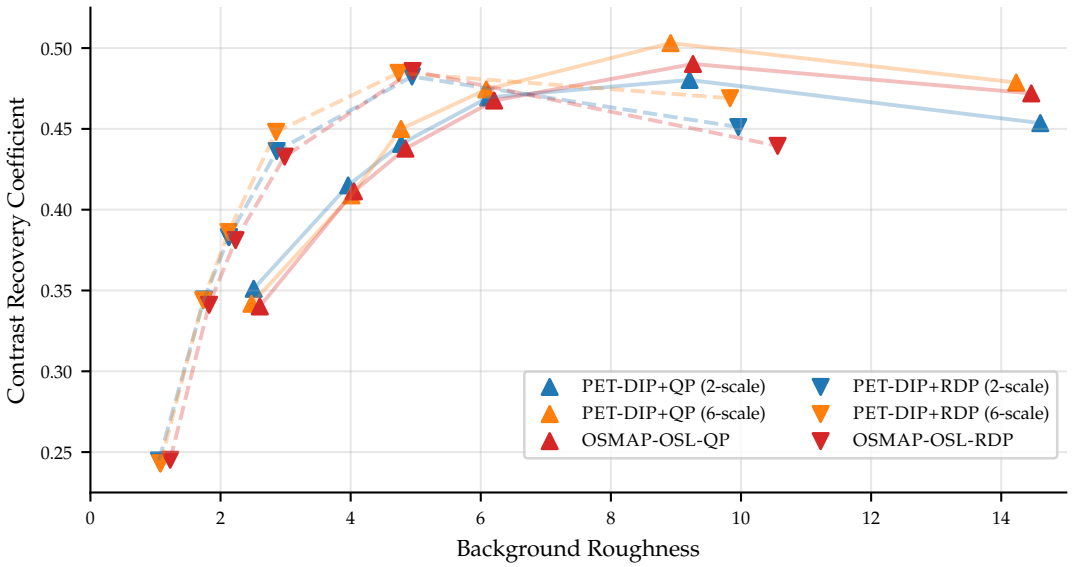


Figure 3.2: For the baseline, we use 200 epochs of the OSMAP-OSL modification of OSEM [103] with 4 ordered subsets. For DIP both 6-scale and 2-scale U-Net architectures are tested and 50,000 epochs are used. OSMAP and DIP results on CRC of lung/lung lesion ROIs and STD of lung ROI. Both the QP and RDP are tested with various λ values: QP $\lambda \in \{20, 10, 7.5, 5, 2.5, 1\} \times 10^{-4}$ and RDP $\lambda \in \{20, 10, 7.5, 5, 2.5, 1\} \times 10^{-2}$.

tion did not continue to decay; meaning that the reconstruction was approximately converged. The quantitative and qualitative discrepancy between DIP and comparison methods can be attributed to the non-convex optimisation of U-Net parameters. This means that the optima could be a local, rather than a global minima. Further, the OSMAP-OSL is a non-convergent algorithm, this latter points was addressed in 3D work where a convergent algorithm is used.

Nonetheless, these 2D results demonstrate consistent results between the baseline and DIP with wrapped SIRF objective. The consistency in results indicate that the value and gradient of the objective function, from SIRF, was correctly passed to `PyTorch`.

3.5.2 3D Results

In 3D, DIP was evaluated on two count levels. For the BSREM-RDP reconstructions, a set of eleven regularisation values λ were used for each count level. For the higher count (lower noise) $\lambda \in [3.125, 31.25] \times 10^{-3}$; for lower count (higher noise) $\lambda \in [1.5, 15] \times 10^{-2}$. The largest and smallest values in the range represent over-regularised and under-regularised solutions respectively.

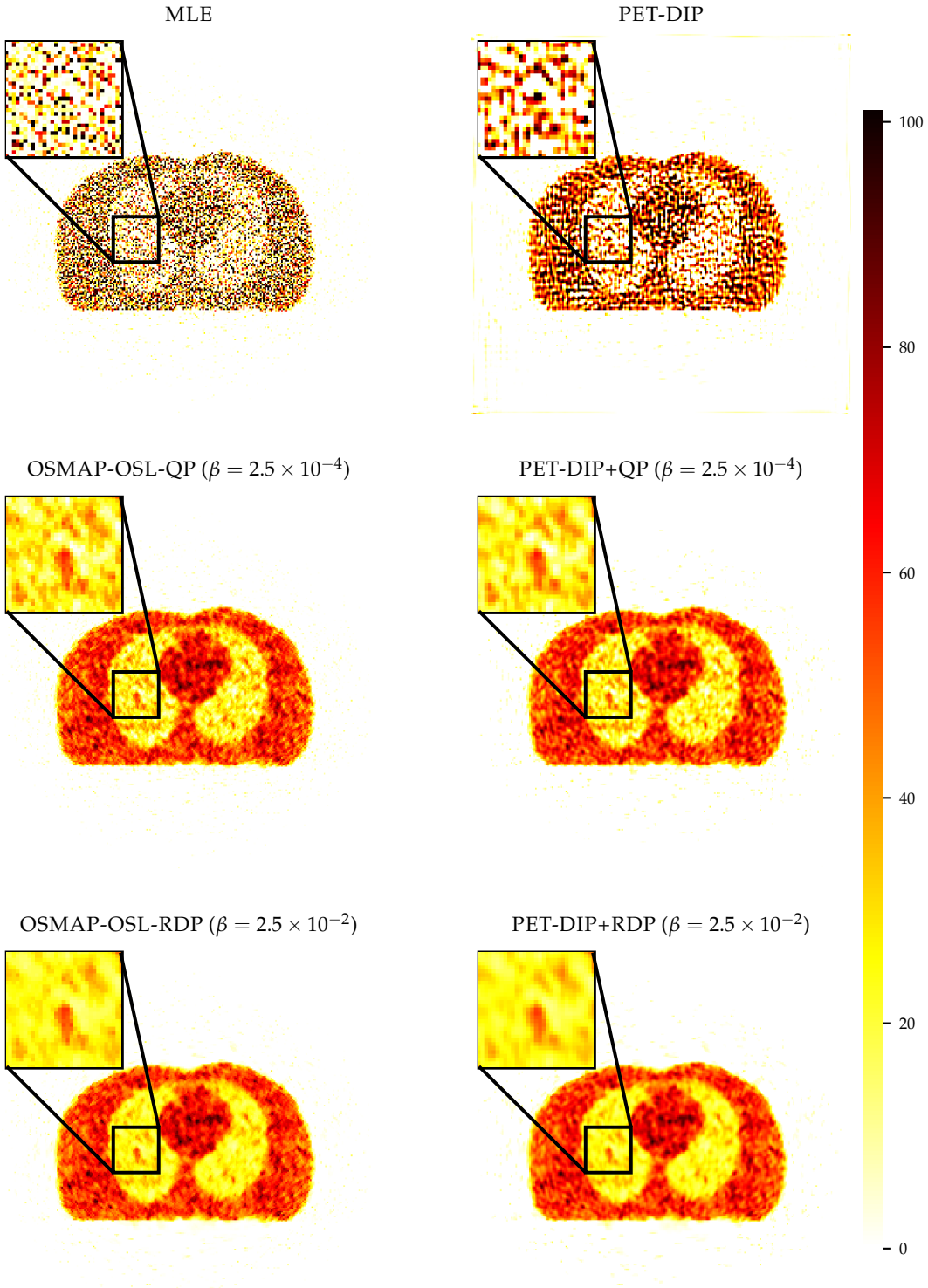


Figure 3.3: OSEM, OSMAP and DIP (6-scale) reconstruction images. DIP - Maximum Likelihood Estimate (MLE) was obtained after 50,000 epochs, and OSEM after 200 epochs with 4 subsets. The λ values for regularised reconstructions were determined from greatest CRC values for RDP and QP from Fig. 3.2. The inset shows the reconstruction of the lung lesion ROI.

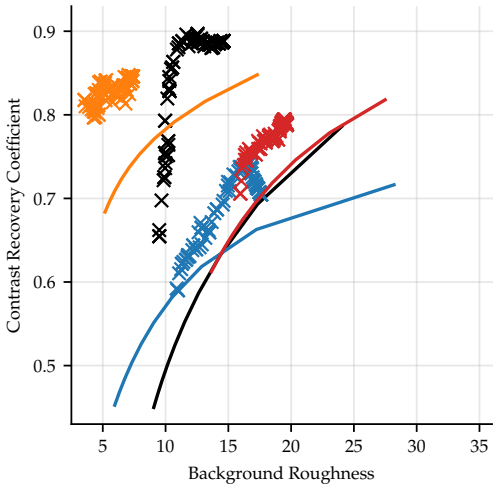
DIP results are shown for the minimum-loss reconstruction every 100 epochs between 10,000 and 20,000 epochs. For PET-DIP+RDP, four regularisation values λ were used for each count level: higher count $\lambda \in \{3.125, 12.5, 21.875, 31.25\} \times 10^{-3}$; lower count $\lambda \in \{1.5, 6.0, 10.5, 15.0\} \times 10^{-2}$. For the higher count Fig. 3.4a, unregularised DIP considerably out-performs BSREM-RDP across all lesions. For the lower count Fig. 3.4b, this is not the case. Including RDP in the objective function, improves the trajectory of PET-DIP+RDP significantly; improved image quality metrics are observed in both lower and higher count see Figs. 3.4c and 3.4d. However, the improvement is not consistent across all lesions. Namely, as observed in Fig. 3.4d, the spine lesion CRC with PET-DIP+RDP is notably worse than that of BSREM-RDP reconstruction. Furthermore, with RDP there is a large drop in CRC for the abdominal lesion in higher count data. This could be due to the abdominal lesions' location at the edge of the axial field-of-view, where noise is higher as sensitivity is lower. These issues of lesion dependence on local sensitivity, contrast and surrounding activity have been observed and investigated with non-DIP reconstruction [238].

We observed that PET-DIP+RDP worked best when λ was lowest but non-zero. Results for DIP reconstruction of lower count data for individual lesions with different regularisation strengths are shown in Figs. 3.5a, 3.5b, 3.5c and 3.5d.

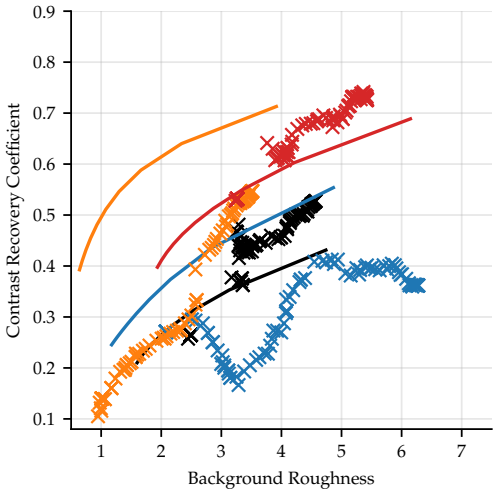
The results for best regularisation strength in higher and lower count data are shown in Figs. 3.4c and 3.4d respectively, and DIP alone are given in Figs. 3.4a and 3.4b. Qualitative visual comparisons of the lower count reconstructions are given in Fig. 3.6.

Table 3.2: GPU memory requirements on tested data for explicit vs. our implicit projector; estimated from sinogram/image sizes as well as 8-byte sparse element-size, and observing GPU memory usage respectively. Memory requirements for the 3D U-Net used in forward and backward modes, and maximum image volume allowable on a 24 GB GPU.

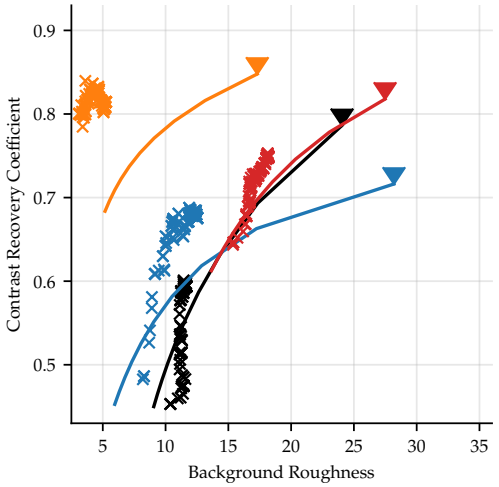
Projector		3D U-Net		
Explicit matrix	Implicit (ours)	Forward	Backward	Maximum Volume
> 100 GB	< 1 GB	0.65 GB	0.88 GB	300^3



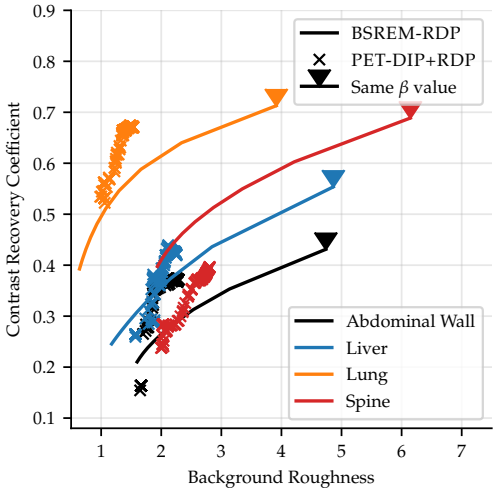
(**a**) PET-DIP comparison on higher count measurements



(**b**) PET-DIP comparison on lower count measurements



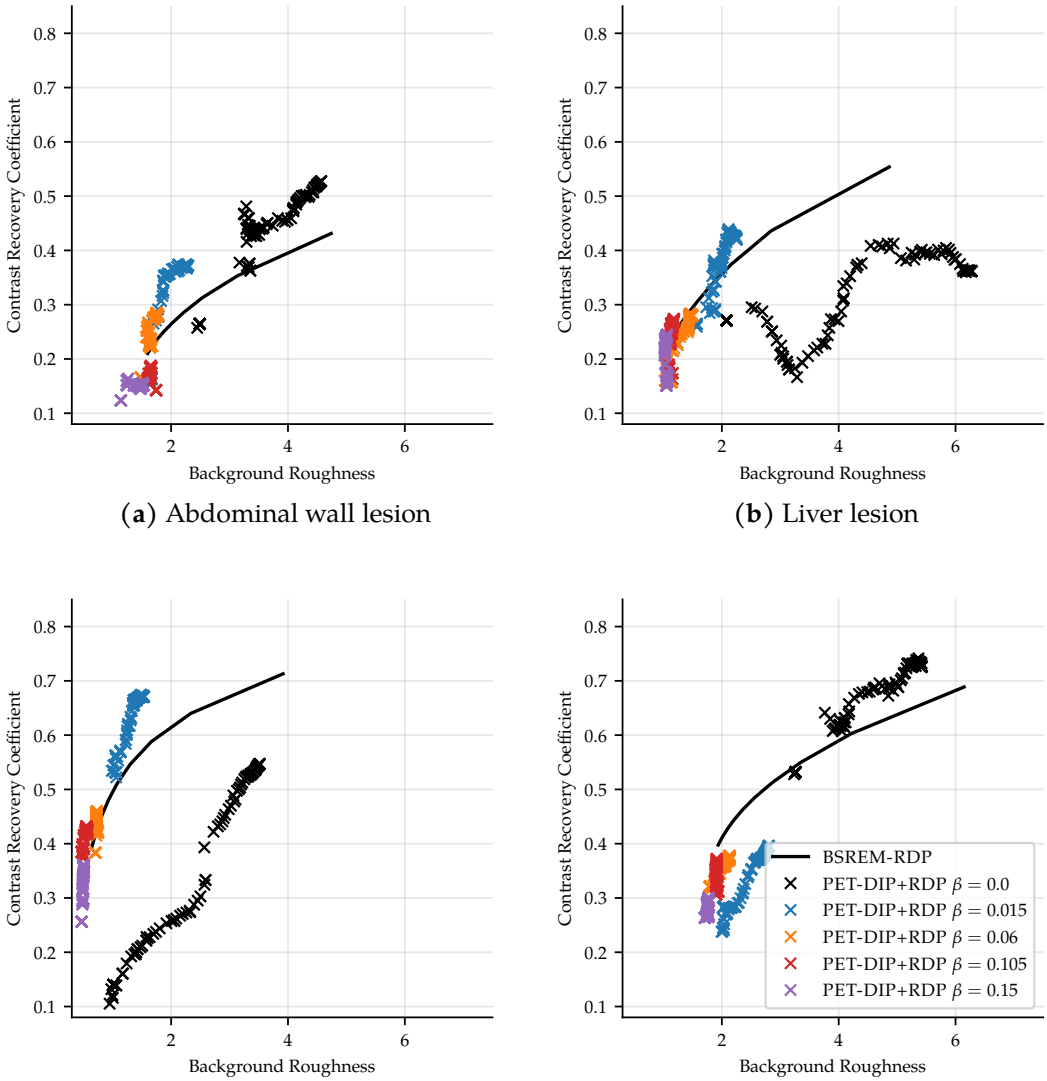
(**c**) PET-DIP+RDP ($\lambda = 3.125 \times 10^{-3}$ comparison on higher count measurements, marker indicates the same λ BSREM-RDP reconstruction



(**d**) PET-DIP+RDP ($\lambda = 1.5 \times 10^{-2}$) comparison on higher count measurements, marker indicates the same λ BSREM-RDP reconstruction.

Figure 3.4: Contrast Recovery Coefficient vs. Background Roughness for abdominal wall, liver, lung and spine lesions. Solid lines correspond to the BSREM-RDP reconstruction with different regularisation strengths. Cross markers DIP reconstructions every 100 epochs.

A single NVIDIA RTX 3090 with 24GB of dedicated memory (VRAM) was used in this study. PARALLELPROJ [207] was used for the projection operator, both the forward and adjoint are implemented in CUDA (GPU-specific language). One full gradient 3D PET DIP iteration took $\approx$ 2.4s, therefore 13.3 hours for the 20,000 iterations. This included costly copying to-and-from the GPU which is currently necessary for integration with SIRF. The wrapper could be



(**a**) Abdominal wall lesion

(**b**) Liver lesion

(**c**) Lung lesion

(**d**) Spine lesion

Figure 3.5: Contrast Recovery Coefficient vs. Background Roughness for lower count measurements with 5 different regularisation strengths. Solid lines correspond to the BSREM-RDP reconstruction with different regularisation strengths. Cross markers DIP reconstructions every 100 epochs.

developed further by interfacing directly with the projector through a CUDA-based `PyTorch` wrapper which would keep operations on the GPU and arrays saved in VRAM, speeding up computation. Run-time could also be reduced by the use of subsets in PET-DIP+RDP. This will be pursued in the future as it would be an important step in developing efficient deep learning techniques for PET reconstruction.

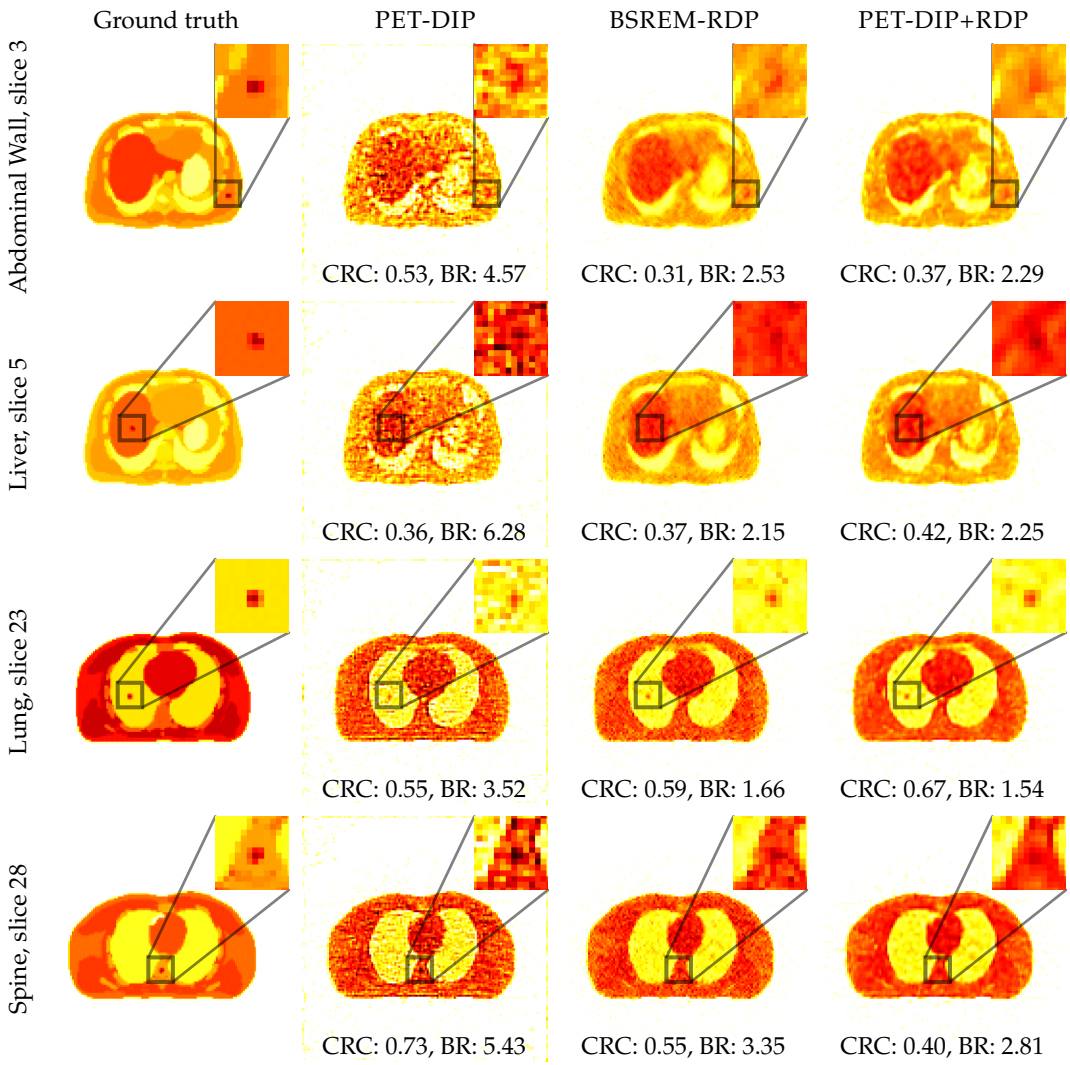


Figure 3.6: Axial slices taken through the centre of lesions. Slices of ground truth emission and lower count data reconstructions with PET-DIP, BSREM-RDP ($\lambda = 4.5 \times 10^{-2}$), and PET-DIP+RDP ($\lambda = 1.5 \times 10^{-2}$). CRC and BR values quotes are for the lesion shown in the slice. DIP reconstructions are the minimum-loss solutions over 20,000 epochs. Colour-scales between reconstructed image slices are kept constant.

3.6 Conclusions

Although DIP is attractive due to requiring no ground-truth training data, it was found to not constitute a stable prior for PET. With the push toward ever-more ill-posed PET reconstruction (reducing dose and acquisition time) [143], DIP does not seem a viable avenue alone for reconstruction. Furthermore, supervision via hyper-parameter tuning is needed to ensure performance gains from DIP. This bias in reconstruction can limit the generalisability and could be a factor in lower count data reconstructions failing with DIP alone. If this is

the case, this would require DIP architectures and early-stopping criteria to be tuned for given count levels, hampering chances of actual adoption in clinical practice.

The introduction RDP to PET-DIP+RDP is a double-edged sword. On one hand, the DIP reconstruction is stabilised having less dependency on the stopping criteria, which is observed in literature [9]. On the other hand, we observed that this introduces issues with lower sensitivity slices requiring different penalty strengths to higher sensitivity slices. Additionally, finding the appropriate penalty strength requires further tuning.

By leveraging established reconstruction engines for PET, the sophisticated forward model and can integrated into deep learning framework by wrapping. This opens the door for further research, specifically for methods that separate forward modelling from the image modelling, which is used in Chapter 5.

3.6.1 Limitations

The focus of this work was first evaluating the effectiveness of the the `PyTorch` wrapper, and then investigating additional regularisation to alleviate the need for early stopping. For this work only a single dataset of the *in-silico* XCAT torso phantom was used. Further studies would require multiple datasets that would ideally be *in-vivo*, which are hard to obtain in practice. Additionally, the effects of cardiac, respiratory motion and radioactive decay were not included in the simulation, limiting the realism of the simulations.

This work did not investigate the choice of stopping criterion, but this is a critical component of DIP. Without a stopping criteria there is no decision on which reconstruction to use along the optimisation trajectory. As can be seen in Fig. 3.4d, the reconstructions can perform worse than the established method BSREM-RDP.

For 20,000 iterations of 3D DIP required 13.3 hours to obtain a reconstruction. This is prohibitively time consuming for clinical practice.

3.6.2 Avenues of Further Research

To accelerate the reconstruction one can educate DIP by pre-training it on a task such as post-processing an approximate reconstruction, see Section 2.3.2.1. This would mean the static random input would be replaced with the approximate reconstruction. At test-time the network can be further tuned in a DIP fashion to fit the measurements. This approach was proposed for micro CT reconstruction [14]. Further, the idea of fine-tuning a supervised approach based on test-time measurements has been described in [47] and was shown to combat the distribution shift between training and test-time.

In literature it is well-known that while DIP is typically over-parameterised, the number of parameters that significantly impact the output are lower. As it was noted by Tachella et al [232]:

"The effective degrees of freedom are significantly smaller than the actual number of weights in the network, being fully characterized by the architecture and initialization of the network."

This is a motivation for the deep decoder architecture [91], which replaces the over-parameterised U-Net with an under-parameterised network. Further, recent work by colleagues Riccardo et al [13], provide an interesting avenue of research. In that work a low-rank approximation of DIP is sought using a data-driven approach. More specifically, a singular value decomposition of network weights is used to define a sub-space of DIP that is then optimised over for subsequent reconstructions. Constraining DIP to a sub-space of a prior image could provide a natural way of introducing guidance information into DIP reconstruction.

Rather than pre-defining a subspace, one could pre-define that the output must have a low-rank structure. Using lightweight networks DIP could learn decomposition into this low-rank structure using methods such as [198].

Guided reconstruction of PET-DIP with an MRI image was proposed by [74], here the image reconstruction was guided by substituting the input static random image with an aligned MRI image. In that case it was unclear how and if the MRI image is used by the algorithm. Designing a U-Net, as described in [255], to ensure guidance is enforced would be a powerful avenue for the fu-

ture. Thorough understanding the inductive bias of the U-Net is the paramount to the success of DIP. This is still poorly understood, but an active area of research.

Chapter 4

Post-Processing and Learned Primal Dual

4.1 Introduction

Supervised approaches for medical image reconstruction often provide state-of-the-art performance when the test data is sufficiently in-distribution with the training data, and they are widespread in the literature [6] -see Section 2.3.2 for more details. For this reason, they are often treated as the gold standard to compare against for advanced image reconstruction.

In this chapter, we focus on the application of supervised approaches for PET image reconstruction [86, 187]. PET images of Activity Concentration (AC) are inherently non-negative and exhibit significant variations in dynamic range between patients. These variations arise from factors such as patient size, radio-tracer dose, and metabolism differences, making PET image modelling challenging. The non-negativity is physically realistic since AC cannot be negative. To address these challenges, normalisation techniques have been proposed in related imaging fields, such as MRI [163], and more recently in PET imaging [155]. However, there are no studies specifically focused on developing and applying such normalisation techniques for deep learning-based PET image reconstruction.

In this work, we consider two popular supervised learning techniques -post-processing and LPD- which are described in Sections 2.3.2.1 and 2.3.2.2 respec-

tively. For post-processing the input $\mathbf{y}_{\text{IN}} = \tilde{\mathbf{x}}$ is an approximate reconstruction, and for LPD the input is the measurement $\mathbf{y}_{\text{IN}} = \mathbf{y}$. These methods have been previously applied to PET image reconstruction without accounting for the wide dynamic range [82, 113]. In this investigation, we propose the use of a simple linear instance normalisation to stabilise the image distribution, and the results presented are published in [216].

The objective of this study is to investigate the effect of linear instance normalisation on supervised approaches for PET image reconstruction.

4.2 Normalisation Approaches

Given a supervised dataset $\{(\mathbf{y}_{\text{IN}}^{(s)}, \mathbf{x}_{\text{GT}}^{(s)})\}_{s=1}^{N_s}$ with N_s samples of target reconstructions $\mathbf{x}_{\text{GT}}^{(s)}$ and corresponding measurements $\mathbf{y}$, the training loss with linear normalisation factor c is given by:

$$\min_{\theta} \sum_{s=1}^{N_s} \| \frac{1}{c^{(s)}} \mathcal{G}_{\boldsymbol{\theta}}(c^{(s)} \mathbf{y}_{\text{IN}}^{(s)}) - \mathbf{x}_{\text{GT}}^{(s)} \|_2^2,$$

where $c^{(s)}$ the sample-dependent intensity normalisation. Normalisation is used to stabilise the variation of the inputs to the network. The chosen normalisation techniques depended solely on the measurements, or the approximate OSEM reconstruction. In this work four normalisations were tested and are summarised in Table. 4.1. Additionally, no normalisation denoted No Norm was a baseline, where $c^{(s)} = 1$.

Table 4.1: Normalisation techniques investigated for PET reconstruction with supervised learning.

	MeanY	MeanCY	MeanOSEM	MaxOSEM
$c^{(s)}$	$\frac{1}{\text{Mean}(\mathbf{y}^{(s)})}$	$\frac{1}{\text{Mean}(\mathbf{y}^{(s)} - \bar{\mathbf{b}}^{(s)})}$	$\frac{1}{\text{Mean}(\tilde{\mathbf{x}}^{(s)})}$	$\frac{1}{\max(\tilde{\mathbf{x}}^{(s)})}$

4.3 Experimental Methods

4.3.1 Dataset

The BrainWeb dataset comprises of 20 anatomical brain phantoms [8] that are segmented into regions of: background, cerebrospinal fluid, gray matter, white matter, fat, muscle, muscle/skin, skull, blood vessels, connective tissue (region around fat), dura matter, and bone marrow. These regions are assigned intensities based on the simulated radio-tracer. In this study FDG radio-tracer was simulated with the volumes further perturbed by three realisations of random distortions, giving 60 ground truth volumes in total [205]. The random distortions perturbed the contrast, added a low-bias field to grey matter, and included blurring and noise. This degraded these target images as well as expanded the training set. The high resolution ($2\,\text{mm}^3$) phantoms defined ground truth PET and CT volumes, where CT volumes were used to compute attenuation factors used for measurement simulation. The test set (subject 04) included ground-truths with simulated hot elliptical lesions. The true measurements $\mathbf{y}_{\text{GT}}$ were simulated with a single crystal ring GE Discovery MI acquisition geometry with resolution, attenuation factors and sensitivity models included, this was done in `pyParallelProj` [207]. The noise level was set by re-scaling the true counts to ensure a prescribed true count per emission voxel. For supervised training data the prescribed true counts per emission voxel were $5, 10, 50$, and for test data it was 2.5. After scaling, a constant background based on the mean true counts was added before applying Poisson noise. The approximate forward model for reconstruction was modelled with low resolution images ($8\,\text{mm}^3$) to avoid an inverse crime [112]. Note, axial slices were used to simulate measurements and the investigation is strictly in 2D.

4.3.2 Network Details

For both methods, the network architectures used were changed minimally from DivαL implementations [138]. The PET-LPD was adapted by using the OSEM reconstruction as initialisation of primal channels, and by including an affine forward operator with sample-specific multiplicative $\mathbf{A}^{(s)}$ and additive

factors $\bar{\mathbf{b}}^{(s)}$, see Section. 2.2.2.1. To reduce computational time, a total of 3 unrolled iterations were used for PET-LPD network. The PET-U-Net input was the same OSEM reconstruction after 1 epoch with 34 subsets. PET-U-Net used a U-Net architecture with 1,783,249 trainable parameters, and the PET-LPD used sets of convolutional filters at each iteration with 132,300 trainable parameters. The intensity normalisation was passed as an input parameter to the networks. For PET-U-Net the OSEM image was normalised and output of the U-Net was unnormalised. For PET-LPD the input to each network iteration was normalised and output unnormalised.

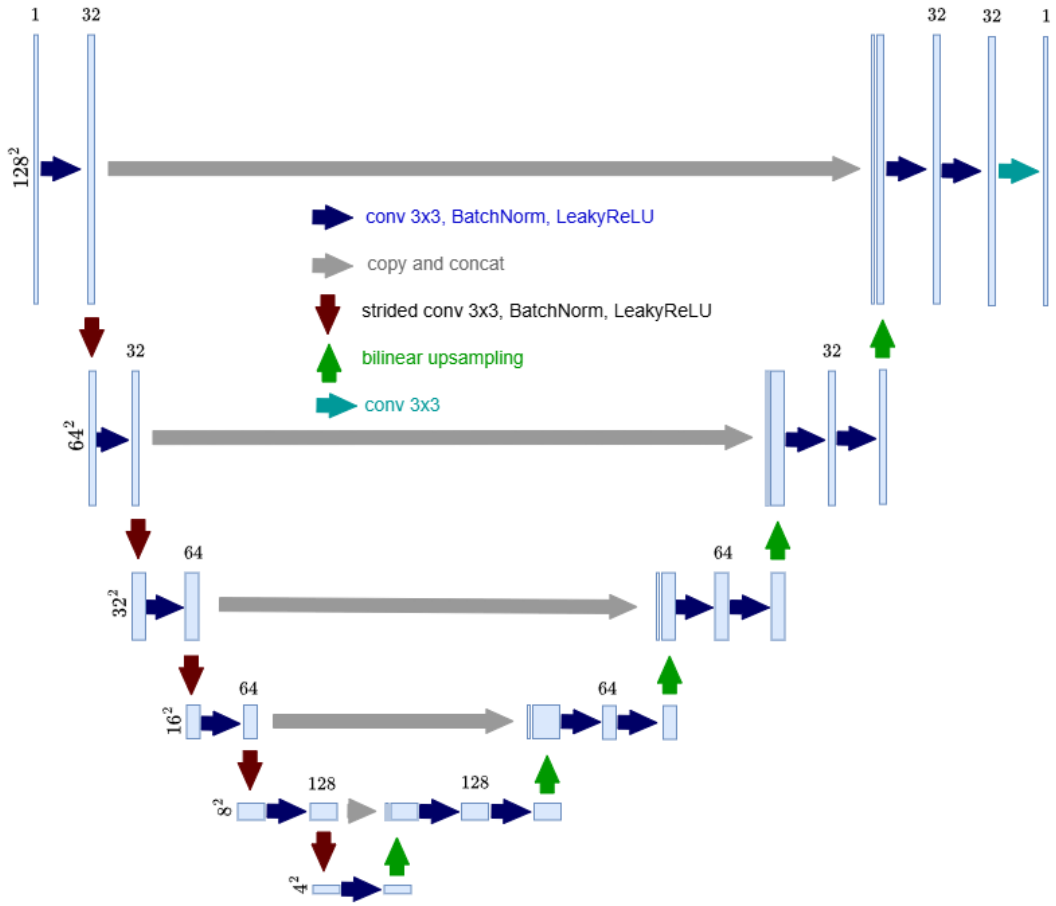


Figure 4.1: Network architecture for the post-processing U-Net.

The PET-U-Net was trained for 250 epochs using the Adam optimiser with a learning rate of 3×10^{-4}. The PET-LPD model was trained for 100 epochs using the Adam optimiser with a learning rate of 1×10^{-3}. For both models, the model with the lowest validation error (based on a held-out validation set) was used for evaluation.

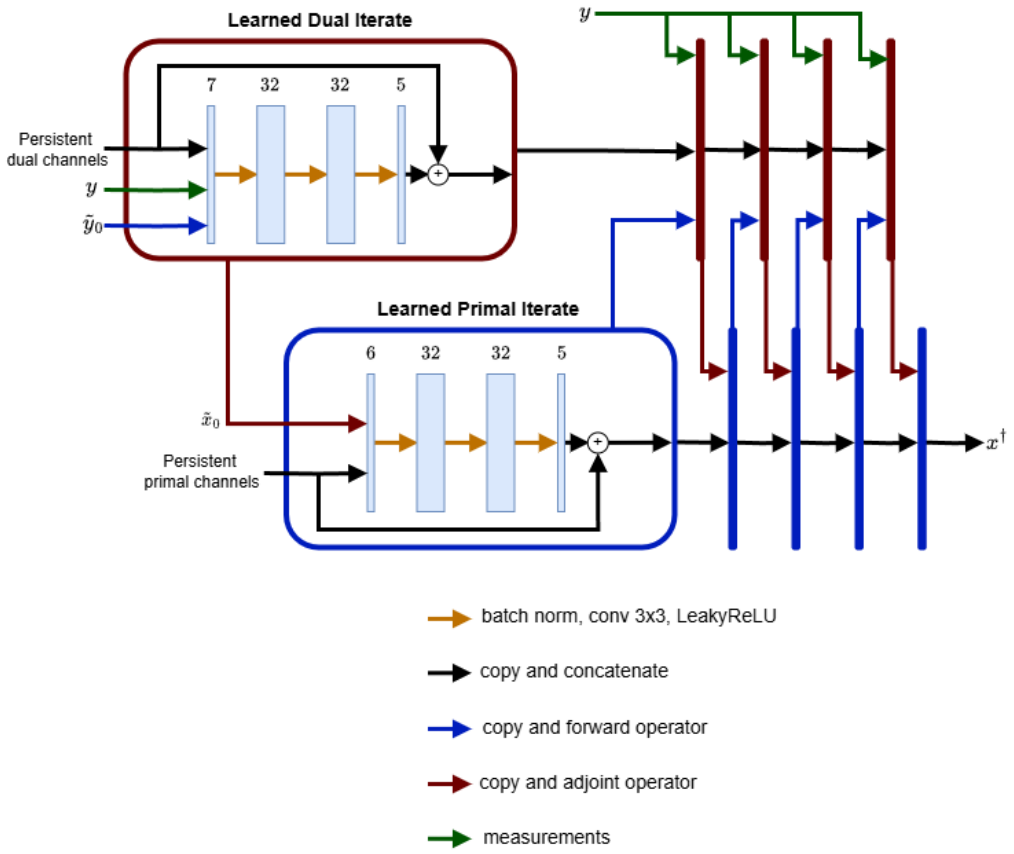


Figure 4.2: Network architecture for the learned primal-dual network. The first iteration is expanded to see internal layers. The initial persistent channels and projected channels ($\hat{\mathbf{x}}_0$ and $\hat{\mathbf{y}}_0$) are initialised with raw measurements or the OSEM image for dual and primal iterates respectively.

4.3.3 Image Quality Assessment

For this investigation we compute CRC, EN and PSNR as described in Section 2.2.4. Additionally, KLDIV between clean measurements $\mathbf{y}$ and estimate of the mean of the measurements $\bar{\mathbf{y}} = \mathbf{A}\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{y}_{\text{IN}}) + \bar{\mathbf{b}}$.

4.4 Results and Discussion

Results are shown for out-of-distribution test data (2.5 true counts per volume with lesions) in Tables 4.2 and 4.3 for PET-U-Net and PET-LPD respectively.

In addition, Figure 4.3 gives a qualitative comparison of sample with No Norm and MeanCY. We find that models with normalisation benefit from higher CRC and lower KLDIV. The EN is higher with normalisation, meaning reconstructions are more sensitive to perturbations in noise, and is a result of being

Table 4.2: OSEMConvNet: Mean (standard error) of the quality metrics on 80 sample test set. Bold indicates the best performance. Confidence is based on a paired t-test.

	No Norm	MeanY	MeanCY	MeanOSEM	MaxOSEM
CRC	0.749	**0.805*****	0.788***	0.803***	0.758***
	(0.014)	(0.015)	(0.014)	(0.015)	(0.017)
EN	**0.216**	0.248	0.244	0.247	0.264
	(0.005)	(0.007)	(0.007)	(0.008)	(0.008)
PSNR	27.55	27.93***	**27.94*****	**27.94*****	27.08*
	(0.38)	(0.40)	(0.40)	(0.39)	(0.43)
KLDIV	68128	**63068*****	63115***	63069***	63667***
	(127)	(246)	(245)	(246)	(268)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.3: Learned Primal-Dual: Mean (standard error) of the quality metrics on 80 sample test set. Bold indicates the best performance.

	No Norm	MeanY	MeanCY	MeanOSEM	MaxOSEM
CRC	0.749	0.876***	0.862***	**0.893*****	0.874***
	(0.016)	(0.013)	(0.013)	(0.011)	(0.011)
EN	**0.163**	0.230	0.227	0.278	0.250
	(0.005)	(0.008)	(0.008)	(0.006)	(0.008)
PSNR	27.94	28.38***	**28.50*****	27.89*	28.00**
	(0.44)	(0.44)	(0.44)	(0.41)	(0.43)
KLDIV	63925	63153***	63027***	**62958*****	63089***
	(239)	(225)	(243)	(238)	(249)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

more data-consistent. Normalisation based on OSEM images shows instability with regards to PSNR and can under-perform as compared with No Norm. More work is needed to conclusively establish the best normalisation practice, although it has been shown that normalisation practices have significant impact on generalisability. This can be seen by the statistically significant $p < 0.01$ paired T-test against No Norm.

4.5 Conclusion

In this study four linear intensity normalisation approaches were investigated for supervised PET image reconstruction. The normalisation approaches provided a statistically significant impact on generalisability, allowing the model

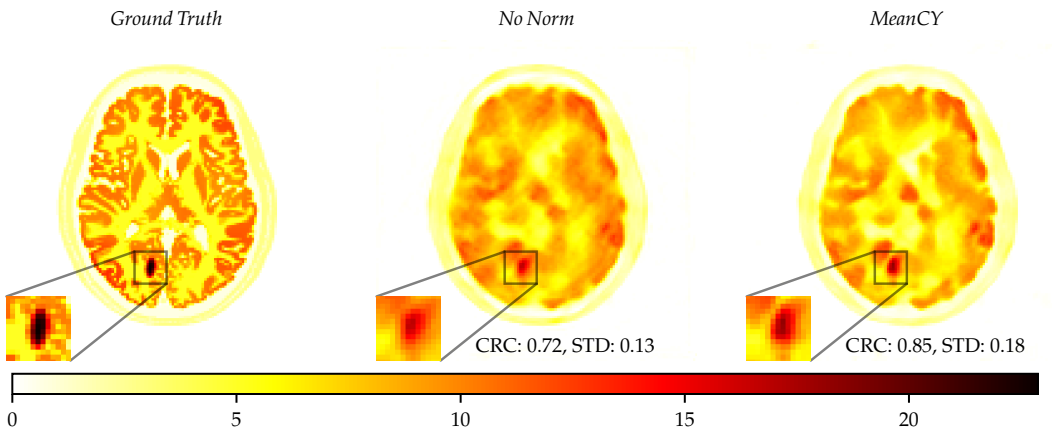


Figure 4.3: Reconstruction for Learned Primal-Dual with a noise level equivalent to an average of 2.5 true counts per emission voxel with no normalisation and normalisation by MeanCY.

to fit the measurements more closely and improve CRC. Specifically, MeanCY normalisation was deemed best performing and it was used in supervised comparison methods against SGM reconstructions in Chapter 5.

4.5.1 Limitations

Due to computational overhead, and lack of ground-truth data, 2D *in-silico* evaluation was performed limiting applicability to clinical practice.

The linear choice of normalisation could be overly restrictive, and non-linear techniques can further improve stability of the intensity distribution. Further, an estimate for the whole body concentration of the injected dose of the radiotracer would allow for normalisation in a similar fashion to standardised uptake values. Nonetheless these has been shown to require standardisation due to variations in patient-related factors and technical factors [155].

Chapter 5

Score-based Generative Models

5.1 Introduction

This chapter introduces and develops SGMs for PET image reconstruction, with the presented results published in [220]. This study addresses the unique challenges of PET imaging, such as its widely varying dynamic range and inherent non-negativity -see Section 2.2.2.1 for details. Additionally, the method is extended to include MR-guided SGM reconstruction and fully 3D reconstruction, expanding its applicability.

The methods discussed in this chapter integrate known physics with data consistency, following a similar approach to DIP in Chapter 3. However, unlike DIP, which operates in an untrained paradigm requiring only raw measurements and the forward model, SGMs rely on an unpaired dataset of high-quality target images. These target images define an empirical distribution, $p_{\text{empirical}}$, which approximates the true sought-after distribution π -see Section 2.3.4. As described briefly in Section 2.3.5.2, the score model is used to approximate this prior π when applying trained generative models to inverse problems, including PET image reconstruction.

The subsequent sections expand on these concepts, presenting an extensive review of background literature before detailing the application of SGMs for PET image reconstruction.

5.2 What is a Diffusion Model?

Recently, the term "diffusion model" has come into widespread use within the computer vision community. It broadly refers to a class of deep generative models that leverages the (Stein) score-matching objectives to train model $\mathcal{G}_{\boldsymbol{\theta}}$ to approximate the score $\nabla_{\mathbf{x}} \log \pi(\mathbf{x})$ of the empirical distribution [105]:

$$L_{\text{SM}}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}\sim\pi(\mathbf{x})}\left[\|\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}) - \nabla_{\mathbf{x}} \log \pi(\mathbf{x})\|_2^2\right].$$

This objective requires the true score of the distribution, $\nabla_{\mathbf{x}} \log \pi(\mathbf{x})$, which is not directly accessible in practice. Instead Vincent [249] gave the connection between score-matching and denoising auto encoders, showing that Denoising Score Matching (DSM) can be used as theoretically-sound and tractable alternative objective:

$$\underset{\boldsymbol{\theta}}{\text{argmin}}\left\{L_{\text{DSM}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim U[0,T]}\mathbb{E}_{\mathbf{x}_0\sim\pi}\mathbb{E}_{\mathbf{x}_t\sim \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_0)}\left[\|\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_0)\|_2^2\right]\right\}. \tag{5.1}$$

Here, the subscript t denotes a time-dependency that defines the noisy image $\mathbf{x}_t$. This noisy image $\mathbf{x}_t$ is a sampled from a distribution p_t, which is defined as a convolution of a tractable distribution with the complex distribution π. The tractable distribution is typically Gaussian $\mathsf{N}(\mathbf{0}, \mathbf{I})$. As time approaches the terminal time $t \to T$, the terminal distribution approaches the tractable distribution (i.e. $\mathsf{p}_T := \mathsf{N}(\mathbf{0}, \mathbf{I})$), while the initial distribution is defined as $\mathsf{p}_0 := \pi$. These properties ensure that, over time, the distribution is progressively smoothed. For a single dimension this is illustrated in Fig. 5.1, where the bimodal 1D π is successively smoothed to become a unimodal Gaussian at $t = T$.

Returning back to Eqn. (5.1), the time t is sampled uniformly from the interval $[0, T]$, the clean image is sampled from the image distribution π (i.e. it is a sample from a dataset), and $\mathbf{x}_t$ is a sampled from the conditional distribution $\mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_0)$. This conditional distribution is constructed to be tractable, in fact it is simply scaled summation of two random vectors $\gamma_t\mathbf{x}_0 + \nu_t\boldsymbol{\varepsilon} \sim \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_0) \equiv \mathsf{N}(\mathbf{x}_t; \gamma_t\mathbf{x}_0, \nu_t^2\mathbf{I})$ where $\boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{x}_0 \sim \pi$, where γ_t and ν_t are scalar-valued

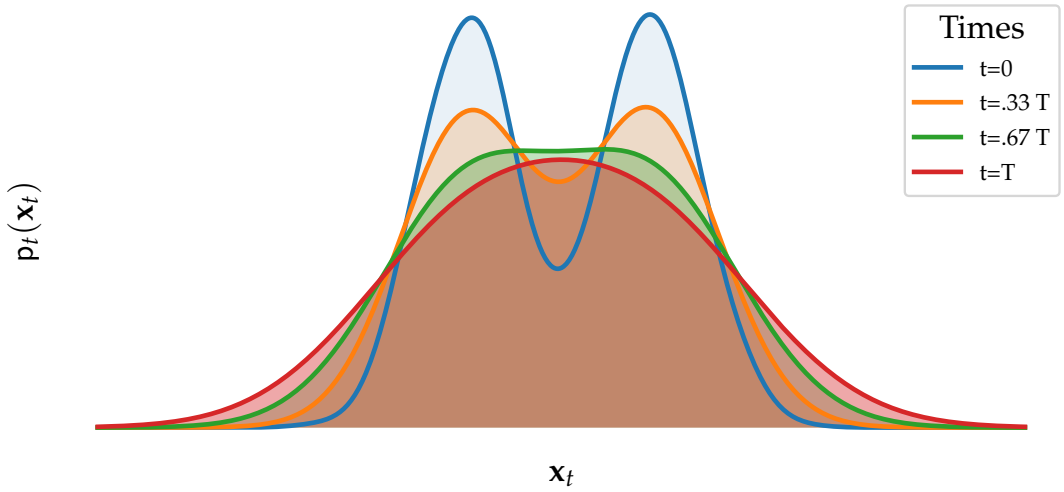


Figure 5.1: One-dimensional probability density scale space. Convolution of the bimodal π distribution at $t = 0$ with a Gaussian untill $t = T$.

functions of t that are defined later. The conditional score is therefore defined as:

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0) = \nabla_{\mathbf{x}_t} \left[-\frac{1}{2\nu_t^2} \|\mathbf{x}_t - \gamma_t \mathbf{x}_0\|_2^2 + const. \right] = -\frac{1}{\nu_t^2}(\mathbf{x}_t - \gamma_t \mathbf{x}_0) \tag{5.2}$$

$$= -\frac{\boldsymbol{\varepsilon}}{\nu_t} \quad \text{where} \quad \boldsymbol{\varepsilon} = \frac{\mathbf{x}_t - \gamma_t \mathbf{x}_0}{\nu_t}. \tag{5.3}$$

The score-model is trained to approximate this score which is the noise added to the image $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \approx -\frac{\boldsymbol{\varepsilon}}{\nu_t}$. Thus the model $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$ can be used to denoise $\mathbf{x}_t$. Other parameterisations exist that can train the model as image denoising or noise prediction [190], for this thesis only the score-based parameterisation is only considered.

There have been a various motivations and formulations of diffusion models, these include denoising diffusion probabilistic modelling [97, 222], and score-matching Langevin dynamics [227]. In this thesis the formulation chosen unifies these approaches through the lens of a Stochastic Differential Equation (SDE). This formulation is coined SGMs [230].

Compared with denoising diffusion probabilistic modelling [97, 222], SGMs generalise t to be continuous rather than discrete. The continuous forward diffusion process is defined by an Itô SDE

$$d\mathbf{x}_t = \mathbf{h}(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}_t, \quad \mathbf{x}_0 \sim \pi, \tag{5.4}$$

where $\{\mathbf{x}_t\}_{t\in[0,T]}$ is a stochastic process indexed by time t. Each random vector $\mathbf{x}_t$ is associated with a family of time-dependent density $\mathsf{p}_t(\mathbf{x}_t)$. The multivariate Wiener process $\{\mathbf{w}_t\}_{t\geq 0}$ is the standard Brownian motion. Starting at the image distribution π, the vector-valued drift function is a $\mathbf{h}(\cdot, t) : \mathbb{R}^{N_\mathbf{x}} \to \mathbb{R}^{N_\mathbf{x}}$ and the diffusion function $g(t) : \mathbb{R} \to \mathbb{R}$ are chosen such that the terminal distribution at $t = T$ approximates the standard Gaussian, $\mathsf{p}_T \approx \mathsf{N}(\mathbf{0}, \mathbf{I})$. Thus, the forward diffusion process maps the image distribution π to a tractable, analytic distribution. In this thesis the variance preserving SDE is used [97] and defines the drift and diffusion as $\mathbf{h}(\mathbf{x}_t, t) = -\frac{\beta(t)}{2}\mathbf{x}_t$ and $g(t) = \sqrt{\beta(t)}$.[1] The resulting SDE of the form:

$$d\mathbf{x}_t = -\frac{\beta(t)}{2}\mathbf{x}_t dt + \sqrt{\beta(t)}d\mathbf{w}_t, \tag{5.5}$$

where $\beta(t) = \beta_{\text{min}} + t(\beta_{\text{max}} - \beta_{\text{min}})$ is a linear schedule with $\beta_{\text{min}} = 0.1$ and $\beta_{\text{max}} = 10$ with a terminal time $T = 1$. As described previously, it can be observed that this chosen formulation Eqn. (5.5) is the summation of two scaled independent random vectors at any time point. With the tractable distribution dominating toward $t = 1$. The transition kernel for the variance preserving SDE is a Gaussian and has closed-from mean and variance [200], i.e. $\mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_0) = \mathsf{N}(\mathbf{x}_t; \gamma_t\mathbf{x}_0, \nu_t^2\mathbf{I})$, with coefficients

$$\gamma_t = \exp\left(-\frac{1}{2}\int_0^t \beta(s)ds\right), \quad \nu_t^2 = 1 - \exp\left(-\int_0^t \beta(s)ds\right). \tag{5.6}$$

Using the terms above and Eqns. (5.1) and (5.2), the denoising score matching loss can be rewritten as:

$$L_{\text{DSM}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim U[0,1]}\mathbb{E}_{\mathbf{x}_0\sim\pi}\mathbb{E}_{\boldsymbol{\varepsilon}\sim\mathsf{N}(\mathbf{0},\mathbf{I})}\left[\omega_t \left\| \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) + \frac{\boldsymbol{\varepsilon}}{\nu_t}\right\|_2^2\right], \tag{5.7}$$

[1] This is equivalent to denoising diffusion probabilistic models [97] in the limit when the time discretisation approaches continuous time.

where the choice of $\omega_t = \nu_t^2$ is used to approximate maximum likelihood training [226]. Note that from Eqn. 5.2, $\mathbf{x}_t = \nu_t \boldsymbol{\varepsilon} + \gamma_t \mathbf{x}_0$ the score-model $\mathcal{G}_{\boldsymbol{\theta}}$ attempts to predict the noise $-\boldsymbol{\varepsilon}/\nu_t$.

Having trained a model to approximate the score, it can be used to reverse the forward diffusion process; to start with pure-noise, and apply the score-model iteratively to attain a sample from π. This is available as the forward stochastic process Eqn. (5.4) has an existing reverse stochastic process [5] that is defined

$$d\mathbf{x}_t = [\mathbf{h}(\mathbf{x}_t, t) - g(t)^2 \nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t)]dt + g(t)d\bar{\mathbf{w}}_t, \tag{5.8}$$

that runs backwards in time. The Wiener process $\{\bar{\mathbf{w}}_t\}_{t\geq 0}$ is time-reversed Brownian motion, and the score $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t)$ approximated with the trained score-model $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t) \approx \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$ so that the approximate reverse SDE is

$$d\mathbf{x}_t = \left[-\frac{\beta(t)}{2}\mathbf{x}_t - \beta(t)\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\right]dt + \sqrt{\beta(t)}d\bar{\mathbf{w}}_t. \tag{5.9}$$

Using the score-model the reverse diffusion process can be sampled by first drawing a sample from the terminal distribution $\mathbf{x}_1 \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$, and then simulating the reverse SDE Eqn. (5.9) backwards in time until $t = 0$. The latter can be achieved by Euler-Maruyama schemes [230].[2]

The Euler-Maruyama scheme is classical method for approximate the solutions of an SDE. This scheme uses an equidistant time discretisation of the form $0 = t_0 \leq \ldots t_k \cdots \leq t_N = 1$ for $N \in \mathbb{Z}$, with a time step $\Delta t = -1/N$. A single iteration of Euler-Maruyama for Eqn. (5.9) SDE is given by:

$$\mathbf{x}_{t_{k-1}} = \mathbf{x}_{t_k} + \left[-\frac{\beta(t_k)}{2}\mathbf{x}_{t_k} - \beta(t_k)\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_{t_k}, t_k)\right]\Delta t + \sqrt{|\Delta t|\beta(t_k)}\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I}), \tag{5.10}$$

This formulation requires a fine discretisation Δt to correctly approximate as solution of the reverse SDE, leading to slow generation of samples. This is because larger step sizes can cause non-convergence in high dimensional spaces

[2] With a finite time this formulation can be shown to be equivalent to denoising diffusion probabilistic models [97, 230].

due to the randomness of the Wiener process [128, Chapter 11]. Additionally the method is Markovian, meaning the next step is only dependent on the current step and not the history of previous iterates.

As derived in [230, Appendix D.1], for the above reverse SDE there exists a deterministic Ordinary Differential Equation (ODE), referred to as the probability flow ODE, with the same marginal density as the reverse SDE:

$$\frac{d\mathbf{x}_t}{dt} = \underbrace{-\frac{\beta(t)}{2}\mathbf{x}_t}_{\text{linear}} + \underbrace{\frac{\beta(t)}{2}\nabla_{\mathbf{x}_t}\log \mathsf{p}_t(\mathbf{x}_t)}_{\text{non-linear}}. \tag{5.11}$$

Unlike the reverse SDE, this ODE is deterministic, producing a smooth trajectory that continuously evolves towards the target distributions. These properties have led to the development of faster non-Markovian samplers [120, 146]. The next section introduces a different perspective on SGMs, which motivates quicker sampling methods that leverage the probability flow ODE.

5.2.1 The Denoising Perspective and Denoising Diffusion Implicit Models

The key insight to move beyond the slow, stochastic reverse time SDE is the connection to denoising. The link between SGMs is made explicit through the relationship between the score and the conditional expectation $\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$ - the mean estimate of $\mathbf{x}_0$ given the noisy $\mathbf{x}_t$. The connection is formally derived using the first order Miyasawa relationship [152]:

$$\nabla_{\mathbf{x}_t}\log \mathsf{p}_t(\mathbf{x}_t) = \mathbb{E}\left[\nabla_{\mathbf{x}_t}\log \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_0)|\mathbf{x}_t\right],$$

given that the condition score is $\nabla_{\mathbf{x}_t}\log \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_0) = -\frac{1}{\nu_t^2}(\mathbf{x}_t - \gamma_t\mathbf{x}_0)$, we can substitute this into the expression and rearrange to solve for the conditional mean:

$$\nabla_{\mathbf{x}_t}\log \mathsf{p}_t(\mathbf{x}_t) = \mathbb{E}\left[-\frac{1}{\nu_t^2}(\mathbf{x}_t - \gamma_t\mathbf{x}_0)|\mathbf{x}_t\right]$$
$$\mathbb{E}\left[\mathbf{x}_0|\mathbf{x}_t\right] = \frac{\mathbf{x}_t + \nu_t^2\nabla_{\mathbf{x}_t}\log \mathsf{p}_t(\mathbf{x}_t)}{\gamma_t}.$$

By replacing the true score $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t)$ with our trained score-model $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$, we obtain a remarkable result known as Tweedie's Formula [40, 59]. This formula provides a one-step denoising estimate of the original data:

$$\hat{\mathbf{x}}_0(\mathbf{x}_t) := \frac{\mathbf{x}_t + \nu_t^2 \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}{\gamma_t} \approx \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$$

This equation shows that a trained score-model can be directly used as a denoiser to predict the clean image from any noisy input in a single step. While powerful, this alone does not accelerate the standard generative process.

Denoising Diffusion Implicit Models (DDIM) leverages the one-step denoiser $\hat{\mathbf{x}}_0(\mathbf{x}_t)$ to formulate a much faster sampling procedure. The core idea is to define a new, non-Markovian forward process that, while being a purely mathematical construct, is guaranteed to have the same marginals $p_t(\mathbf{x}_t|\mathbf{x}_0)$ as the original forward SDE. This ensures that the same pre-trained score-model $\mathcal{G}_{\boldsymbol{\theta}}$ can be used.

The benefit of this alternative process is that it allows for a corresponding reverse process that is not strictly tied to the small, stochastic steps of the SDE. Instead, it can be understood as a numerical discretisation of the deterministic probability flow ODE [146, 225]. This perspective frames the DDIM sampler as an iterative refinement process: at each step t_k, it uses the Tweedie estimate $\hat{\mathbf{x}}_0(\mathbf{x}_{t_k})$ to predict the final clean image and then takes a large, direct step toward it.

The general DDIM update rule to get from $\mathbf{x}_{t_k}$ to $\mathbf{x}_{t_{k-1}}$ is formulated as:

$$\mathbf{x}_{t_{k-1}} = \gamma_{t_{k-1}} \hat{\mathbf{x}}_0(\mathbf{x}_{t_k}) - \nu_{t_k} \sqrt{\nu_{t_{k-1}}^2 - \zeta_{t_k}^2} \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_{t_k}, t_k) + \zeta_{t_k} \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I}). \tag{5.12}$$

The parameter ζ_{t_k} controls the stochasticity of the update.[3] Different choices of ζ_t result in different sampling schemes. In this work, $\zeta_{t_k} = \zeta \psi_{t_k}$ with a hyperparameter $\zeta \in [0, 1]$, controlling the amount of stochasticity in the sampling, and $\psi_{t_k} = \nu_{t_{k-1}} / \nu_{t_k} \sqrt{1 - \gamma_{t_k} / \gamma_{t_{k-1}}}$ [225].

[3] In fact it is only truly DDIM when $\zeta_{t_k} = 0$

5.3 Inverse Problems and Score-based Generative Models

From a Bayesian perspective of inverse problems, see Section 2.1.2.2, we are concerned with sampling from the posterior defined:

$$\mathsf{p}^{\text{post}}(\mathbf{x}|\mathbf{y}) \propto \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\mathbf{x})\pi(\mathbf{x}), \tag{5.13}$$

where the likelihood p^{lkhd} is known, given the forward model and noise model of the measurements $\mathbf{y}$. The goal in the application of SGMs for inverse problems is to approximate the prior π in order to sample from the posterior. This aligns with the motivations of the thesis, where the image modelling is separated from the forward modelling.

While SGMs can be applied to inverse problems through trained conditional mechanisms [98], we focus on the application of unconditional SGMs for inverse problems. In this case data-consistency with measurements is required during the sampling of the score-model to sample the posterior. There exist two main directions for incorporating data-consistency with an unconditional score-model:

1. **Approximation of $\mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t)$**

2. **Projection-type data-consistency**

The term $\mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t)$ is often referred to a the noisy or time-dependent likelihood. It is derived by considering a posterior $\mathsf{p}_t^{\text{post}}(\mathbf{x}_t|\mathbf{y})$ that has the form:

$$\mathsf{p}_t^{\text{post}}(\mathbf{x}_t|\mathbf{y}) \propto \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t)\mathsf{p}_t(\mathbf{x}_t),$$

where the posterior $\mathsf{p}_0^{\text{post}}(\mathbf{x}_0|\mathbf{y})$ at $t = 0$ is precisely the posterior we would like the to sample from in Eqn. (5.13). Taking the gradient-log of the posterior gives:

$$\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{post}}(\mathbf{x}_t|\mathbf{y}) = \nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) + \underbrace{\nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t)}_{\text{score}},$$
$$\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{post}}(\mathbf{x}_t|\mathbf{y}) \approx \nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) + \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t),$$

the gradient-log of the time-dependent prior is the score that is approximated by the score-model. To sample from $\mathsf{p}_0^{\text{post}}(\mathbf{x}_0|\mathbf{y})$ requires conditioning the reverse SDE the measurements. This can be obtained by substituting in the gradient-log of the posterior into the reverse SDE (Eqn. (5.9)), which results in:

$$d\mathbf{x}_t = \left[-\frac{\beta(t)}{2}\mathbf{x}_t - \beta(t)(\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) + \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)) \right] dt + \sqrt{\beta(t)} d\bar{\mathbf{w}}_t. \tag{5.14}$$

Sampling this reverse SDE to $t = 0$ results in a sample of the posterior $\mathsf{p}^{\text{post}}(\mathbf{x}|\mathbf{y})$, and repeated sampling allows for the estimation of higher posterior moments. In spite of this, the common practice only computes a single sample for the reconstruction, due to computational overhead of repeated sampling [42, 107, 229].

The term $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t)$ is only available in analytic form at $t = 0$, i.e. when it is the likelihood $\nabla_{\mathbf{x}} \log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\mathbf{x})$. The approximation of the noisy likelihood is the basis for several methods for formulating SGMs for inverse problems. One approximation is the the annealed likelihood [107, 185]:

$$\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) \approx \lambda_t \nabla_{\mathbf{x}_t} \log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t). \tag{5.15}$$

This approximation is referred as the ANNLD approximation, where the penalty strength λ_t typically increases as $t \to 0$. This enforces the likelihood more closer to $t = 0$, as $\mathsf{p}_t^{\text{lkhd}} \approx \mathsf{p}^{\text{lkhd}}$ is more valid closer to $t = 0$.

A recent popular choice for the approximation of the noisy likelihood is Diffusion Posterior Sampling (DPS) [40]. Rather than computing the likelihood based-off the noisy image $\mathbf{x}_t$, the Tweedie's estimate $\hat{\mathbf{x}}_0$ is used to compute the conditional mean, from which the likelihood gradient is computed.

$$\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) \approx \lambda_t \nabla_{\mathbf{x}_t} \log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\hat{\mathbf{x}}_0(\mathbf{x}_t)), \tag{5.16}$$

this approximation is denoted DPS. Given that the Tweedie's estimate Eqn. (**??**) applies the score-model to denoise $\mathbf{x}_t$, the gradient therefore requires back-propagation through the score-model, adding extra computational overhead in comparison to ANNLD approximation, and the time-dependent penalty λ_t is still required. Nonetheless, this approximation has been shown perform favourably on several image reconstruction tasks [40].

The projection-type data consistency modifies the DDIM sampling rule in Eqn. (5.12) for data consistent conditional generation [41, 224, 267]. These methods generally consist of three steps: Estimate the denoised image $\mathbf{x}_0$ using Tweedie's estimate $\hat{\mathbf{x}}_0(\mathbf{x}_{t_k})$; update $\hat{\mathbf{x}}_0(\mathbf{x}_{t_k})$ with a data consistency step; and add the noise back, according to the DDIM update rule, in order to get a sample for the next time step t_{k-1}. Importantly, with this approach there is no need to estimate the gradient of the time-dependent likelihood $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t)$ as data consistency is only enforced on Tweedie's estimate at $t = 0$.

The conditional DDIM samplers differ predominately in the implementation of the data consistency update. Decomposed Diffusion Sampling (DDS) [41] proposes to align Tweedie's estimate with the measurements by running p steps of a Conjugate Gradient (CG) scheme for minimising the negative log-likelihood at each sampling step. Let $\text{CG}^{(p)}(\hat{\mathbf{x}}_0)$ denote the p-th CG update initialised with $\hat{\mathbf{x}}_0(\mathbf{x}_{t_k})$. This can be seen as an approximation to the conditional expectation, i.e. $\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t, \mathbf{y}] \approx \text{CG}^{(p)}(\hat{\mathbf{x}}_0)$ [186]. Using this approximation, the update step for DDS can be written as:

$$\mathbf{x}_{t_{k-1}} = \gamma_{t_{k-1}} \text{CG}^{(p)}(\hat{\mathbf{x}}_0) - \nu_{t_k} \sqrt{\nu_{t_{k-1}}^2 - \zeta_{t_k}^2} \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_{t_k}, t_k) + \zeta_{t_k} \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim \mathsf{G}(\mathbf{0}, \mathbf{I}), \tag{5.17}$$

This projection method is denoted DDS, and by leveraging DDIM the sampling is greatly accelerated, as the fine time discretisation for sampling the reverse SDE is not needed. Above, the amount of data consistency is controlled, coarsely, by the number of steps p to prevent over-fitting to noise in the mea-

surements.

5.4 Image Modelling - the PET Intricacies

The use of an unconditional score-model for inverse problems aids in generalisation as the model is trained to only sample from the prior distribution π. Through data consistency one can attempt to sample the posterior distribution, i.e. condition the score-model sampling on the measurements. The training and sampling of the score-model must be adapted given intricacies of the modality at hand. More specifically, the score-model must faithfully and robustly model images obtained from the modality.

PET images of Activity Concentration (AC) are inherently non-negative and have greatly varying dynamic ranges between patients [161]. The non-negativity is physically realistic as AC cannot be negative, and the dynamic range variations are due the patient size, radio-tracer dose and metabolism differences between patients. These factors make the image modelling of PET images challenging.

5.4.1 PET Normalisation for Score-based Generative Models

As it was noted in Chapter 4, normalisation is an important technique that can improve image reconstruction when networks are trained. This can be partially attributed to overcoming bias toward intensity levels that appear more frequently in the training set. Consequently, the network might struggle to handle new images with unseen intensity levels, leading to instability in the learning and evaluation process [237]. Although unsupervised, a score-model may exhibit these instability too when trained on PET images of a widely varying dynamic range. It is importance to stabilise this widely varying dynamic range to aid generalisability. Furthermore, for SGMs it is noted in literature that the noisy image $\mathbf{x}_t$ must converge to noise similarly for all target images defining the empirical dataset [145]. This is because the forward diffusion process is predefined with the rate at which it converges to Gaussian is dependent on the dynamic range of the target images. A larger intensity scale of images

will converge to pure noise slower. This can make training and sampling more unstable [145].

Input normalisation is a standard deep learning methodology to deal with intensity shifts and normalise the inputs to the network. We propose input normalisation to ensure that the score-model $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$ is able to estimate the score of images with arbitrary intensity values. In the training stage, each training image $\mathbf{x}_0$ is normalised to ensure voxel intensities are within a certain range. To do this, a training normalisation factor c_{train} is introduced that when applied ensures the average emission per emission voxels (a voxel with non-zero intensity value) is 1. This is computed as:

$$c_{\text{train}} = c(\mathbf{x}_0) := \frac{\sum_{j=1}^{N_{\mathbf{x}}} [\mathbf{x}_0]_j}{\#\{j : [\mathbf{x}_0]_j > 0\}}, \tag{5.18}$$

where the numerator $\sum_{j=1}^{N_{\mathbf{x}}} [\mathbf{x}_0]_j$ computes the total emission in the image and the denominator $\#\{j : [\mathbf{x}_0]_j > 0\}$ is the number of emission voxels. The normalisation factor is incorporated into the DSM training objective function by rescaling the initial image, yielding the objective

$$\mathbb{E}_{t\sim U[0,1]}\mathbb{E}_{\mathbf{x}_0\sim\pi}\mathbb{E}_{\boldsymbol{\varepsilon}\sim\mathsf{G}(\mathbf{0},\mathbf{I})}\mathbb{E}_{c\sim\mathsf{U}[\frac{c_{train}}{2},\frac{3c_{train}}{2}]}\left[\nu_t^2\left\|\mathcal{G}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}_t, t) + \frac{\boldsymbol{\varepsilon}}{\nu_t}\right\|_2^2\right], \tag{5.19}$$

with $\tilde{\mathbf{x}}_t = \gamma_t \mathbf{x}_0/c + \nu_t \boldsymbol{\varepsilon}$. Compared with Eqn. 5.7, the scale-factor in range $c \sim \mathsf{U}[\frac{c_{train}}{2}, \frac{3c_{train}}{2}]$ is used to encourage the score-model to be more robust with respect to misestimations of the normalisation constant during sampling.

The incorrect estimation of the normalisation constant is due to the true number of emission voxels $\#\{j : [\mathbf{x}_0]_j > 0\}$ is not known during sampling. A surrogate for the number of emission voxels can be obtained by estimating it from an approximate OSEM reconstruction. The OSEM is fast to compute often using a single epoch of OSEM from non-negative initialisation. The resulting sampling normalisation factor is given by:

$$c_{\text{OSEM}} = \frac{\sum_{j=1}^{m} [\mathbf{x}_{\text{OSEM}}]_j}{\#\{j : [\mathbf{x}_{OSEM}]_j > Q_{0.01}\}}, \tag{5.20}$$

where $Q_{0.01}$ defines the 1% percentile of $\mathbf{x}_{\text{OSEM}}$ values. This threshold is heuristically chosen to ensure that noise and reconstruction artefacts do not cause an over-estimation of the number of emission voxels.

In proposed sampling methods the normalisation constant c_{OSEM} is applied as a factor scaling the time-dependent likelihood such that $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|c_{\text{OSEM}}\mathbf{x}_t)$. At final time step $t = 0$, the output $\mathbf{x}_0$ is rescaled by c_{OSEM} to recover the correct intensity level. In projection-type data consistency the Tweedie's estimate is rescaled.

5.4.2 Modifications of Sampling Methods

The sampling schemes and approximations in Section 5.3 were originally proposed for inverse problems with Gaussian NLL. To work with a Possion NLL the sampling technique must be modified to take into account the inherent non-negativity. The work on DPS [40] also considers inverse problems with Poisson noise model, but utilises a Gaussian approximation, which is known to be unsuitable for PET reconstruction in the event of the low photon count [21, 99].

For methods that approximate the gradient-log of the noisy likelihood, i.e. ANNLD and DPS, the non-negativity required for the Possion NLL can be ensured with a non-negativity projection. This type of projection can be seen in the context of guided diffusion where the iterates $\mathbf{x}_{t_k}$ are projected to a specified domain after each sampling step [142, 196]. However, this creates a mismatch between the forward and reverse SDEs. In literature it is noted that this mismatch results in artefacts in the reconstructions and may even lead to divergence of the sampling [145]. We observed that thresholding all negative values of $\mathbf{x}_{t_k}$ leads to a divergence of the sampling process. Therefore only threshold the input to the Possion NLL, i.e. with $\nabla_{\mathbf{x}_t} \log \mathsf{p}^{\text{lkhd}}$ being the Possion NLL, see Eqn. (5.15), for the PET-ANNLD approximation we use:

$$\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) \approx \lambda_t^{\text{ANNLD}} \nabla_{\mathbf{x}_t} \log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|c_{\text{OSEM}}\mathcal{P}_{\geq 0}[\mathbf{x}_t]), \tag{5.21}$$

and likewise for PET-DPS, i.e. DPS in Eqn. (5.16):

$$\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) \approx \lambda_t^{\text{DPS}} \nabla_{\mathbf{x}_t} \log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|c_{\text{OSEM}} \mathcal{P}_{\geq 0}[\hat{\mathbf{x}}_0(\mathbf{x}_t)]), \tag{5.22}$$

where $\mathcal{P}_{\geq 0}$ is the non-negativity projection. This leads to a perturbed likelihood gradient that is not computed on the true iterate $\mathbf{x}_t$, but only on the projection. In order to reconstruct the PET image the reverse SDE is solved using the specific approximation (PET-Annld or PET-DPS) as the likelihood term. The algorithms for PET-Annld and PET-DPS are given in Algos. 1 and 2 respectively.

In addition to gradient based approximations, the projection-type data consistency is modified. This reduces reconstruction time by modifying the DDIM-based sampling rule DDS [41, 267], and the new rule is coined PET-DDS. This rule enforces data consistency by projecting the Tweedie's estimate $\hat{\mathbf{x}}_0(\mathbf{x}_{t_k})$ toward the measurement manifold by running gradient descent steps of a MAP objective, see Section 2.2.3.1. The update is given by:

$$\mathbf{x}_{t_k}^{(0)} = \hat{\mathbf{x}}_0(\mathbf{x}_{t_k}) \tag{5.23}$$

$$\mathbf{x}_{t_k}^{(i+1)} = \mathcal{P}_{\mathbf{x}\geq 0}\left[\mathbf{x}_{t_k}^{(i)} + \mathbf{D}(\mathbf{x}_{t_k}^{(i)}) \nabla_{\mathbf{x}} \Phi(\mathbf{x}_{t_k}^{(i)})\right] \tag{5.24}$$

$$i = 0, \ldots, p-1$$

$$\mathbf{x}_{t_{k-1}} = \gamma_{t_{k-1}} \mathbf{x}_{t_k}^{(p)} - \nu_{t_k} \sqrt{\nu_{t_{k-1}}^2 - \zeta_{t_k}^2} \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_{t_k}, t_k) + \zeta_{t_k} \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I}), \tag{5.25}$$

where $\mathbf{D}(\mathbf{x}) = \text{diag}\left\{\max(\mathbf{x}, 10^{-4})/\mathbf{A}^\top \mathbf{1}\right\}$ is the preconditioner. The objective Φ is:

$$\Phi(\mathbf{x}^{(i)}) = -\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|c_{\text{OSEM}} \mathbf{x}^{(i)}) + \lambda^{\text{DDS}} \|\mathbf{x}^{(i)} - \hat{\mathbf{x}}_0\|_2^2. \tag{5.26}$$

The term $\|\mathbf{x}^{(i)} - \hat{\mathbf{x}}_0\|_2^2$ can be considered an anchoring term that ensures the update does not deviate too far from the Tweedie's estimates. Given this formulation the objective Eqn. (5.26) is convex, and gradient descent could converge to a unique minimum. To alleviate the computational burden of running to convergence, a fixed number of iterations p are chosen to ensure sufficient data consistency. This approach implicitly enforces regularisation due to early stopping from the Tweedie's estimate initialisation. The amount of regularisation is

therefore defined by the number of iteration p as well as penalty strength λ^{DDS}.

The different proposed sampling schemes for PET are summarised in Table. 5.1.

Table 5.1: Summary of different sampling schemes proposed for PET.

Method	**Data consistency with** $-\log \mathsf{p}^{\mathbf{lkhd}}$
PET-ANNLD	$-\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}\|c_{\text{OSEM}}\mathcal{P}_{\geq 0}[\mathbf{x}])$ 5.21
PET-DPS	$-\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}\|c_{\text{OSEM}}\mathcal{P}_{\geq 0}[\hat{\mathbf{x}}_0(\mathbf{x})])$ 5.22
PET-DDS	p steps of MAP objective 5.24 and **??**

Algorithm 1 PET-ANNLD

Require: Measurements $\mathbf{y}$
Require: Number of steps N
Require: Time discretisation $0 = t_0 \leq \cdots \leq t_N = 1$
  $\mathbf{x}_{t_N} \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$ ▷ Sample initial noise
  for $k = N-1, \ldots, 1$ **do**
    $\mathcal{G}_{\boldsymbol{\theta}} \leftarrow \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_{t_{k+1}}, t_{k+1})$
    $\boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$
    $\Delta t \leftarrow t_k - t_{k+1}$
    $\tilde{\mathbf{x}}_{t_k} \leftarrow \mathbf{x}_{t_{k+1}} + [g(\mathbf{x}_{t_{k+1}}, t_{k+1}) - g(t_{k+1})^2 \mathcal{G}_{\boldsymbol{\theta}}]\Delta t + g(t_{k+1})\sqrt{|\Delta t|}\boldsymbol{\varepsilon}$ ▷ Unconditional score update
    $\mathbf{x}_{t_k} \leftarrow \tilde{\mathbf{x}}_{t_k} - g(t_{k+1})^2 \lambda^{\text{ANNLD}}_{t_{k+1}} \nabla_{\mathbf{x}_{t_{k+1}}} \log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|c_{\text{OSEM}}\mathcal{P}_{\geq 0}[\mathbf{x}_{t_{k+1}}])\Delta t$ ▷ Data consistency step
  end for
  $\mathbf{x}^\dagger \leftarrow c_{\text{OSEM}}\mathbf{x}_{t_1}$

5.5 Further adaptions for PET image reconstruction

In addition to the intricacies of the PET image modelling, there are other adaptations that can be considered given the available side-information acquired with PET measurements, as well as considerations for fully 3D PET image reconstruction.

In more recent years the emergence of PET/MR scanners allows for the simultaneous acquisition of both PET and MRI measurements. As described in 2.2.1, these modalities provide complementary information. Additionally, clinical PET images are fully 3D voxelised volumes. Training a score-model to generate fully 3D volumes is extremely computationally expensive as well

Algorithm 2 PET-DPS

Require: Measurements $\mathbf{y}$
Require: Number of steps N
Require: Time discretisation $0 = t_0 \leq \cdots \leq t_N = 1$
Require: Transition density $p(\mathbf{x}_t|\mathbf{x}_0) = \mathsf{N}(\mathbf{x}_t; \gamma_t \mathbf{x}_0, \nu_t^2 \mathbf{I})$
$\mathbf{x}_{t_N} \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$ ▷ Sample initial noise
for $k = N-1, \ldots, 1$ **do**
$\mathcal{G}_{\boldsymbol{\theta}} \leftarrow \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_{t_{k+1}}, t_{k+1})$
$\hat{\mathbf{x}}_0(\mathbf{x}_{t_{k+1}}) \leftarrow \gamma_{t_{k+1}}^{-1}(\mathbf{x}_{t_{k+1}} + \nu_{t_{k+1}}^2 \mathcal{G}_{\boldsymbol{\theta}})$ ▷ Compute Tweedie estimate
$\boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$
$\Delta t \leftarrow t_k - t_{k+1}$
$\tilde{\mathbf{x}}_{t_k} \leftarrow \mathbf{x}_{t_{k+1}} + [\mathbf{h}(\mathbf{x}_{t_{k+1}}, t_{k+1}) - g(t_{k+1})^2 \mathcal{G}_{\boldsymbol{\theta}}]\Delta t + g(t_{k+1})\sqrt{|\Delta t|}\boldsymbol{\varepsilon}$ ▷ Unconditional score update
$\ell \leftarrow -\log \mathsf{p}^{\mathrm{lkhd}}(\mathbf{y}|c_{\mathrm{OSEM}}\mathcal{P}_{\geq 0}[\hat{\mathbf{x}}_0(\mathbf{x}_t)])$
$\mathbf{x}_{t_k} \leftarrow \tilde{\mathbf{x}}_{t_k} + \lambda_{t_{k+1}}^{\mathrm{DPS}}/\ell \, \nabla_{\mathbf{x}_{t_{k+1}}} \log \mathsf{p}^{\mathrm{lkhd}}(\mathbf{y}|c_{\mathrm{OSEM}}\mathcal{P}_{\geq 0}[\hat{\mathbf{x}}_0(\mathbf{x}_{t_{k+1}})])\Delta t$ ▷ Data consistency step
end for
$\mathbf{x}^\dagger \leftarrow c_{\mathrm{OSEM}}\mathbf{x}_{t_1}$

as requiring access to many 3D volumes of target image quality. These factors necessitate the use of a 2D score-model for fully 3D reconstruction. Further, the computation of the fully 3D forward model is expensive and reducing this computational burden is important in allowing reconstruction times within clinically acceptable time-frames.

5.5.1 Guided Reconstruction with an MR image

MRI is able to provide high resolution images that highlight the structure of soft-tissue. This is complementary to PET images that give lower resolution functional information. This complementary nature of the modalities can be leveraged to regularise the inverse problems making it better-posed [7]. More specifically, additional MR images can be used to regularise the PET image reconstruction by encoding the common anatomical features through edges or other hand-crafted regularisers [10, 26, 62, 223]. Using a high quality reconstructed MR image to guide PET image reconstruction is referred to as guided reconstruction [60]. In a statistical sense, this corresponds to formulating a conditional prior $\pi(\mathbf{x}|\mathbf{x}_{\mathrm{MR}})$, where $\mathbf{x}_{\mathrm{MR}}$ is the guidance MR image. The gradient-log posterior is then given by:

$$\nabla_{\mathbf{x}} \log \mathsf{p}^{\text{post}}(\mathbf{x}|\mathbf{y}, \mathbf{x}_{\text{MR}}) = \nabla_{\mathbf{x}} \log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\mathbf{x}) + \nabla_{\mathbf{x}} \log \pi(\mathbf{x}|\mathbf{x}_{\text{MR}}),$$

here it is assumed that both $\mathbf{y}$ and $\mathbf{x}_{\text{MR}}$ are conditionally independent given $\mathbf{x}$. The likelihood $\mathsf{p}^{\text{lkhd}}(\mathbf{y}|\mathbf{x})$ is given by the Poisson noise model and $\pi(\mathbf{x}|\mathbf{x}_{\text{MR}})$ is a prior conditioned on the MR image $\mathbf{x}_{\text{MR}}$. The conditional prior requires the training of a conditional score-model $\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t, \mathbf{x}_{\text{MR}}) \approx \nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_{\text{MR}})$. Given this conditional score, the reverse SDE can be written as:

$$d\mathbf{x}_t = [\mathbf{h}(\mathbf{x}_t, t) - g(t)^2 \left(\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_{\text{MR}})\right)]dt + g(t)d\bar{\mathbf{w}}_t. \quad (5.27)$$

As before both PET-Annld or PET-DPS can be used to approximate the score of the time dependent likelihood $\nabla_{\mathbf{x}_t} \log \mathsf{p}_t^{\text{lkhd}}(\mathbf{y}|\mathbf{x}_t)$. Within SGM literature, the score-model with an additional conditioning input is known as a conditional score-model. Conditional score-models are popular and are the basis of established text-to-image approaches such as Stable Diffusion [191]. Additionally this conditional approach has been applied for PET image denoising [75], where the conditional input is a noisy PET image and is trained with paired target PET image. To train a conditional score-model, a paired dataset of target images and conditioning input (i.e. the MR image) is needed. This paired dataset is difficult to obtain, and Classifier-free Guidance (CFG) alleviates the burden by requiring only a partially paired dataset. This consists of paired images as well as only target PET images without MR image input. CFG trains an unconditional and conditional score-model by passing an image of zeroes as the MR image during training. This means that the same model can be used for unconditional and conditional sampling; by passing zeroes or the MR image respectively. This is denoted as:

$$\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t, \mathbf{x}_{\text{MR}}) \approx \nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t|\mathbf{x}_{\text{MR}}) \quad \text{and} \quad \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t, t, \mathbf{x}_{\text{MR}} = \mathbf{0}) \approx \nabla_{\mathbf{x}_t} \log \mathsf{p}_t(\mathbf{x}_t),$$

The corresponding training objective for CFG requires an adaption of DSM;

$$\mathbb{E}_{t\sim U[0,T]}\mathbb{E}_{\mathbf{x}_0,\mathbf{x}_{\mathrm{MR}}\sim\pi}\mathbb{E}_{\mathbf{x}_t\sim p_t(\mathbf{x}_t|\mathbf{x}_0)}\mathbb{E}_{q\sim\mathsf{B}(q)}\Big[\omega_t\|\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t,t|\mathbf{x}_{\mathrm{MR}})-\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t|\mathbf{x}_0)\|_2^2\Big]\}. \tag{5.28}$$

Note that q is an additional scalar that drawn from Bernoulli distribution that nullifies the conditioning MR image setting it to zeroes. After training, CFG can use both conditional and unconditional score-models in a linear combination to control the amount of guidance from the MR image:

$$\bar{\mathcal{G}}_{\boldsymbol{\theta}}(\mathbf{x}_t;t,\mathbf{x}_{\mathrm{MR}})=(1+w)\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t,t,\mathbf{x}_{\mathrm{MR}})-w\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_t,t,\mathbf{0}),$$

the weighting w is the guidance strength. It is this combined score-model $\tilde{\mathcal{G}}_{\boldsymbol{\theta}}$ that can be inserted directly into the sampling to guide reconstruction with additional MR image.

5.5.2 Fully 3D reconstruction

PET images in clinic are typically full 3D volumes, with photons detected across multiple detector rings. This forward model in this case is inherently 3D and computationally expensive. These two aspects need to be addressed to develop clinically orientated algorithms for PET image reconstruction. The PET-DDS algorithm is further adapted for 3D reconstruction as it is the most scalable.

The score-model $\mathcal{G}_{\boldsymbol{\theta}}$ is trained in 2D slices due to the computation and data demand of training a fully 3D score-model. More specifically, the dataset only contained axial slices meaning that the score-model provided a prior on the axial plane, leaving the orthogonal direction without regularisation. Additional regularisation in the orthogonal direction was used to ensure consistency across axial slices. The RDP was used in this axial direction by formulating a 1D neighbourhood penalising variation along the axial z-direction that is denoted η_z.

To further accelerate the computation of the data consistency update, block-sequential subset objective gradients were computed. This was inspired by established PET image reconstruction algorithms BSREM and OSEM; that cyclically iterate through a set of N_s objective subsets to accelerate convergence see

Section 2.2.3.1. There is however a trade-off, by utilising subsets the individual updates may not converge, and the likelihood is only based off a subset of all available measurements; making it noisier and less accurate. However, the acceleration would allow for clinically feasible reconstruction times. The subset objectives were ordered in a staggered fashion meaning that equally spaced projections are taken across the angles of all the projections. Further, when these subsets are accessed they are accessed in the Herman-Meyer order [96] in an attempt to maximise orthogonality between subsets.

Formalising this the PET-DDS algorithm is given in Algo. 3. Note that the hyper-parameters λ^{DDS} and p control the amount of data-consistency.

Algorithm 3 PET-DDS

Require: Measurements $\mathbf{y}$
Require: Number of steps N
Require: Time discretisation $0 = t_0 \leq \cdots \leq t_N = 1$
Require: Transition density $p(\mathbf{x}_t|\mathbf{x}_0) = \mathsf{N}(\mathbf{x}_t; \gamma_t \mathbf{x}_0, \nu_t^2 \mathbf{I})$
Require: Number of inner optimisation steps p, number of subsets N_s
Require: Stochasticity $\{\zeta_t\}_{t\geq 0}$
  $\mathbf{x}_{t_N} \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$ ▷ Sample initial noise
  for $k = N-1, \ldots, 1$ **do**
    $\mathcal{G}_{\boldsymbol{\theta}} \leftarrow \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{x}_{t_{k+1}}, t_{k+1})$
    $\text{Noise}(\mathbf{x}_{t_{k+1}}, \mathcal{G}_{\boldsymbol{\theta}}) \leftarrow -\nu_{t_{k+1}} \sqrt{\nu_{t_k}^2 - \eta_{t_{k+1}}^2}\, \mathcal{G}_{\boldsymbol{\theta}}$
    $\mathbf{x}_{t_{k+1}}^{(0)} = \hat{\mathbf{x}}_0(\mathbf{x}_{t_{k+1}}) \leftarrow \gamma_{t_{k+1}}^{-1}(\mathbf{x}_{t_{k+1}} + \nu_{t_{k+1}}^2 \mathcal{G}_{\boldsymbol{\theta}})$ ▷ Compute Tweedie estimate
    for $i = 0, \ldots, p-1$ **do** ▷ Inner optimisation for data consistency
      $s \leftarrow (p(N-k) + i \mod N_s) + 1$
      $\Phi_s(\mathbf{x}_{t_{k+1}}^{(i)}) \leftarrow -\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\mathbf{A}_s c_{\text{OSEM}} \mathbf{x}_{t_{k+1}}^{(i)} + \bar{\mathbf{b}}_s) + (\lambda^{\text{RDP}} \eta_z(\mathbf{x}_{t_{k+1}}^{(i)}) + \lambda^{\text{DDS}} \|\mathbf{x}_{t_{k+1}}^{(i)} - \hat{\mathbf{x}}_0(\mathbf{x}_{t_{k+1}})\|_2^2)/N_s$
      $\mathbf{x}_{t_{k+1}}^{(i+1)} \leftarrow \mathcal{P}_{\geq 0}\left[\mathbf{x}_{t_{k+1}}^{(i)} + \mathbf{D}(\mathbf{x}_{t_{k+1}}^{(i)}) \nabla_{\mathbf{x}} \Phi_s(\mathbf{x}_{t_{k+1}}^{(i)})\right]$
    end for
    $\boldsymbol{\varepsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$
    $\mathbf{x}_{t_k} \leftarrow \gamma_{t_k} \mathbf{x}_{t_{k+1}}^{(p)} + \text{Noise}(\mathbf{x}_{t_{k+1}}, \mathcal{G}_{\boldsymbol{\theta}}) + \zeta_{t_{k+1}} \boldsymbol{\varepsilon}$
  end for
  $\hat{\mathbf{x}} \leftarrow c_{\text{OSEM}} \mathbf{x}_{t_1}$

5.6 Experimental Methods

We aim to evaluate the effectiveness of SGMs for PET image reconstruction, focusing on the applicability of SGMs as a prior; independent of the optimal

regularisation parameter.

5.6.1 Network Architecture

In this work, we adopt the architecture proposed by [54].[4] This model is shown in Fig. 5.2 and is derived from the widely used U-Net framework [192], featuring a decoder designed as a sequence of residual blocks (ResBlock) combined with downsampling layers, and an encoder comprising residual blocks with upsampling operations. At the smaller spatial resolution (32×32, 16×16 and 8×8) global attention layers are introduced to enhance representation, see Section 2.3.3. To integrate the timestep into each residual block, the architecture leverages adaptive group normalisation (AdaGN) layers, formulated as $\text{AdaGN}(h, e) = e_s\text{GroupNorm}(h) + e_b$, where h are intermediate features and $e = [e_s, e_b]$ embeds the time step. This embedding utilises a sinusoidal and MLP time embeddings. For the MRI-guided model, the clean MRI image is incorporated as an additional input channel to the network.

5.6.2 Dataset

Due to data-availability it was chosen to have *in-silico* experiments based on the BrainWeb dataset. This consisted of 20 patient-realistic volumes [8]. There were two sets of simulation conducted; one for the target PET images and one for the noisy measurements (i.e. the raw scanner data). The experimental methods differed between 2D and 3D testing, but the same score-model was used for both.

The score-model learns a target empirical distribution, and therefore only target PET images are required for training. The score-model was trained on axial slices of 19 PET simulated volumes, with three realisations of random distortions as per [205], see Section 4.3 for more details. A FDG tracer was simulated and zero intensity slices were removed resulting in 4569 training slices in total. The training slices were healthy, with no lesions, and with distortions that were deemed realistic. For the MR guided score-model, MR images that correspond to the PET image were simulated and used during training.

[4] Available at `https://github.com/openai/guided-diffusion`.

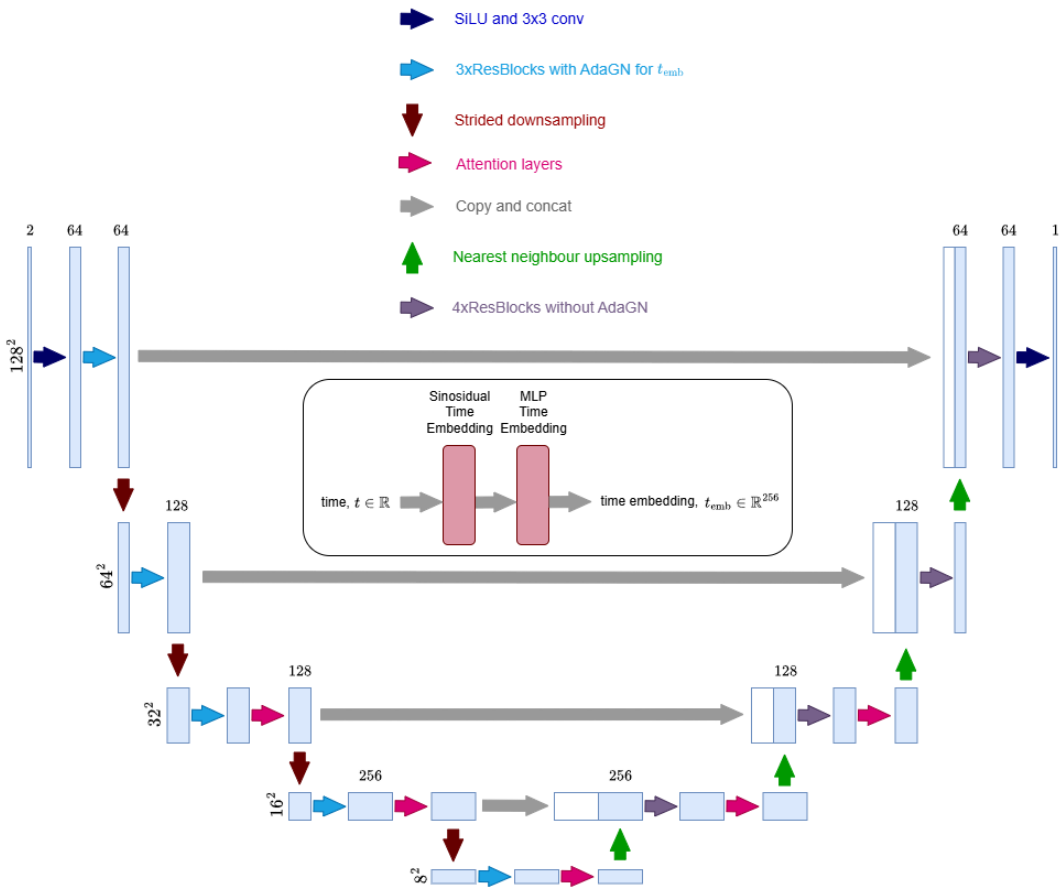


Figure 5.2: Enter Caption

In 2D evaluation, the image simulation was conducted the same as the training data, but on 20 equidistant axial slices of the held-out subject 04 evaluation volume. These slices were further diversified by the addition of simulated lesions as local regions of elliptical hyper-intensity within soft-tissue. The images were then scaled to ensure that the total counts divided by the emission volume was on average 2.5 or 10, corresponding to low and normal dose PET activity concentration respectively. Measurements were then simulated using `pyParallelProj` [207] and included resolution, attenuation (from simulated CT image), sensitivity and background contamination giving clean measurements. These measurements were then corrupted with Poisson noise with 10 different noise realisations. The forward model for these measurements included approximations of the aforementioned resolution, attenuation, detector sensitivities and background, as well as the scanner geometry, see Section 2.2.2.1 for more details. The total true counts averaged over the evaluation set

were 122, 808 and 491, 232 for the low and normal dose PET measurements respectively.

For 3D evaluation the image simulation used the full subject 04 evaluation volume, with the addition of ellipsoidal lesions of hyper-intensity within soft-tissue. In this setting two tracers were simulated, a FDG tracer as well as an Amyloid tracer to provide further out-of-distribution evaluation. The measurements were simulated with a Siemens Biograph mMR scanner geometry [115], and detector sensitivities and attenuation were simulated and included in the forward model using SIRF and STIR [167, 233]. Poisson noise was added to the measurements and 5 realisations were obtained. The noise level was equivalent to 40 million counts without background, corresponding to low dose PET scan. The projector and measurements were ordered into 28 staggered subsets unless otherwise specified.

5.6.3 Comparison Reconstruction Methods

The comparison methods for 2D included the post-processing and unrolled iterative methods described in Section 4; the PET-U-Net and PET-LPD respectively. These supervised methods were trained on measurements of noise levels 5, 10 or 50 without lesions. Additionally, the PET image denoising approach developed in [75] was compared. This method is denoted as Annld (OSEM) as it is equivalent to PET-Annld with the MSE ensure data-consistency with the OSEM image, i.e. $\|\mathbf{x}_{\text{OSEM}} - \mathbf{x}_t\|_2^2$. This surrogates the likelihood on measurements, and the same score-model for both sampling methods.

In 3D, reconstructions were compared against state-of-the-art DIP unsupervised method described in Chapter 3, as well as converged MAP estimates with the RDP regulariser computed with BSREM -see Section 2.2.3.1.

5.6.4 Image Quality Assessment

Two global quality methods of PSNR and SSIM were used to quantify overall fidelity of reconstructions against the simulated ground truths. Two further metrics were computed locally to assess lesion detectability and noise level, these were CRC and EN respectively. These are described Section 2.2.4. In

2D the KLDIV between measurements $\mathbf{y}$ and estimated measurements $\bar{\mathbf{y}}$ was computed to quantify data consistency. All metrics are averaged over the respective evaluation set and across realisations. With algorithms that required tuning of a penalty strength, a range of penalty strengths were swept to produce bias-variance graphs. This aided in the assessment of the SGM as a prior. The graphs have EN along the x-axis as a surrogate for penalty strength, as it quantifies the noise in the reconstructions across realisations of the measurements, where higher EN corresponds to greater data consistency. The converse, lower EN corresponding to more influence of the prior, is not necessarily true in the case of SGM prior, where multiple reconstructions with high likelihood under the model increase EN.

5.7 Results and Discussion

In 2D, we compare the different SGM sampling methods - ANNLD (OSEM), PET-ANNLD, PET-DPS and PET-DDS - against one another and against state-of-the-art supervised methods - PET-U-NET and PET-LPD. Two noise levels are evaluated as well as with and without lesions. MR-guided SGM reconstructions are compared with non-guided SGM reconstruction.

In 3D the results are given with lesions and two tracers for PET-DDS. These are compared with classical MAP estimates and state-of-the-art DIP reconstructions. All results are computed on a single NVIDIA GeForce RTX 3090.

5.7.1 2D Results without Lesions

The penalty strengths λ_t^{ANNLD} and λ_t^{DPS} are both time dependent, while λ^{DDS} is not. The parameters are defined $\lambda_t^{\text{ANNLD}} = \lambda^{\text{ANNLD}}(1-t)$ as proposed in [107], and $\lambda_t^{\text{DPS}} = \frac{\lambda^{\text{DPS}}}{\log \mathsf{p}^{\text{lkhd}}(\mathbf{y}|\hat{\mathbf{x}}_0)}$ as proposed in [40]. For PET-DDS the number of iterations was first tuned to ensure that reconstructions over-fit to noise in the measurements, then the penalty strength λ^{DDS} was increased to increase the influence of the score-model. In 2D the number of projection steps for PET-DDS were set to 4 for noise level 2.5 and 15 for noise level 10.

Figs. 5.3 and 5.4 show that results for the four SGM methods vary greatly.

Notably, the EN for PET-ANNLD is significantly higher at both noise levels 10 and 2.5. When we apply the non-negativity projection $\mathcal{P}_{\mathbf{x}\geq 0}[\mathbf{x}_t]$ to the noisy image, we set all negative values to zero. This operation causes a mismatch between the training distribution —which includes negative values in the noisy images— and the conditional sampling used in PET-ANNLD. This mismatch leads to a divergence in the sampling process, resulting in a considerable increase in EN compared to the other methods.

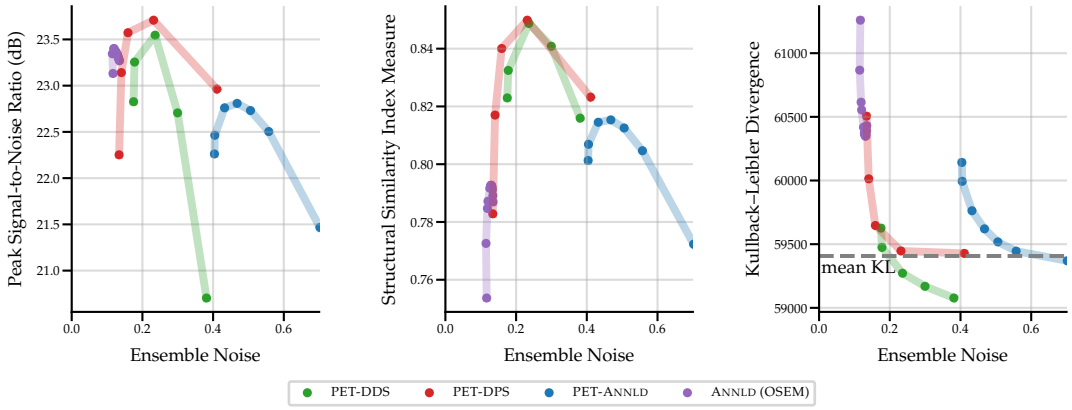


Figure 5.3: Results for BrainWeb **without lesions** with noise level **10** for different penalty strengths. The Standard Deviation is computed over reconstructions of noise realisations **y**. The points represent different values of the parameter λ.

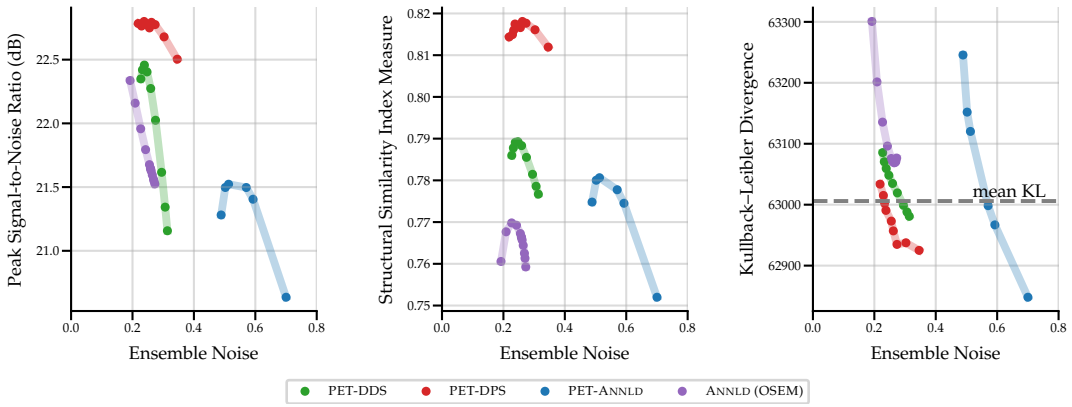


Figure 5.4: Results for BrainWeb **without lesions** with noise level **2.5** for different penalty strengths. Standard deviation is across reconstructions from different realisations of measurements.

PET-DPS is the best performing method on global quality metrics, with the highest PSNR and SSIM, with low EN. However, it is also computationally the most expensive, requiring 1000 steps with back-propagation through the score

Table 5.2: The computing time of a single reconstruction averaged over 5 reconstructions.

Method	**PET-ANNLD**	**PET-DPS**	**PET-DDS**
Time (s)	41.52	60.64	3.90

model. PET-DDS performs competitively with a much lower computational overhead of 100 steps without score-model back-propagation. In Table 5.2, the computational time for one single reconstruction is compared. PET-ANNLD and PET-DPS are largely comparable in terms of inference efficiency, and are about ten times slower than PET-DDS. The difference in computing times can be attributed to the fact that PET-DDS requires fewer time steps through the use of the accelerated DDIM sampling method.

At noise level 10, Fig. 5.3, the KLDIV is lower for PET-DDS than other methods meaning that it is able to ensure data consistency whilst still being competitive in-terms of other metrics. Note that the "mean KL" is the divergence between the measurements with and without noise, and is a threshold indicating that the reconstruction is sufficiently data consistent.

Table 5.3: The mean quality score and standard error using the best hyperparameters for each method for BrainWeb **without** lesions for noise level **2.5** (out-of-distribution) and **10** (in-distribution). The penalty strength used for each SGM method is denoted by λ. The best SGM is highlighted in grey, and overall best metric is underlined. Supervised methods are trained on data with noise levels 5, 10 and 50.

Noise Level	**Method**	**PSNR,** λ	**SSIM,** λ
2.5	**ANNLD (OSEM)**	22.38±0.82, 0.527	0.770±0.02, 3.08
	PET-ANNLD	21.52±0.84, 12.0	0.781±0.01, 12.0
	PET-DPS	22.80±0.81, 650.	<u>0.818</u>±0.01, 750.
	PET-DDS	22.46±0.82, 0.25	0.789±0.02, 0.2
	PET-LPD	<u>23.06</u>±0.85, N/A	0.806±0.01, N/A
	PET-U-NET	22.80±0.82, N/A	0.798±0.01, N/A
10	**ANNLD (OSEM)**	23.40±0.84, 0.2	0.793±0.02, 0.9
	PET-ANNLD	22.81±0.87, 10.	0.815±0.01, 10.
	PET-DPS	23.70±0.83, 400.	0.850±0.01, 400.
	PET-DDS	23.55±0.77, 0.025	0.849±0.01, 0.025
	PET-LPD	<u>24.74</u>±0.91, N/A	<u>0.861</u>±0.01, N/A
	PET-U-NET	24.52±0.85, N/A	0.868±0.01, N/A

In Table 5.3 results are given with penalty strengths that give the best PSNR

or SSIM values. Results are also given for PET-U-Net and PET-LPD that are trained on measurement noise levels of 5, 10 and 50. This means that noise level 2.5 is more out-of-distribution than noise level 10 for supervised methods. At noise level 10 the supervised approaches are the best, whilst at noise level 2.5 SGM approaches are more competitive. Between the noise levels PET-LPD has a decrease fo 6.7% and 6.6% in PSNR and SSIM, while PET-DPS has as moderate drop of 3.4% and 3.8% respectively. In the table it is of note that there is a pronounced change in the selected λ values between noise levels 2.5 and 10 for PET-DDS, especially when compared with other methods. This is due to the p number of BSREM-like data consistency updates run, where at the noise level 2.5 less steps are needed to fit to noise than at noise level at 10 necessitating much stronger regularisation.

5.7.2 2D Results with Lesions

Both the score-model and supervised methods were trained on images without lesions. Evaluation on images with lesions gives an insight into the generalisation to out-of-distribution data, as well as allowing the computation CRC to evaluate lesion detectability - a clinically pertinent metric. The bias-variance graphs in Figs. 5.5 and 5.6 show similar trends to the without lesion evaluation with regard to PSNR, SSIM and KLDIV. However, PET-DDS shows significant improvements to CRC as compared with other SGM methods.

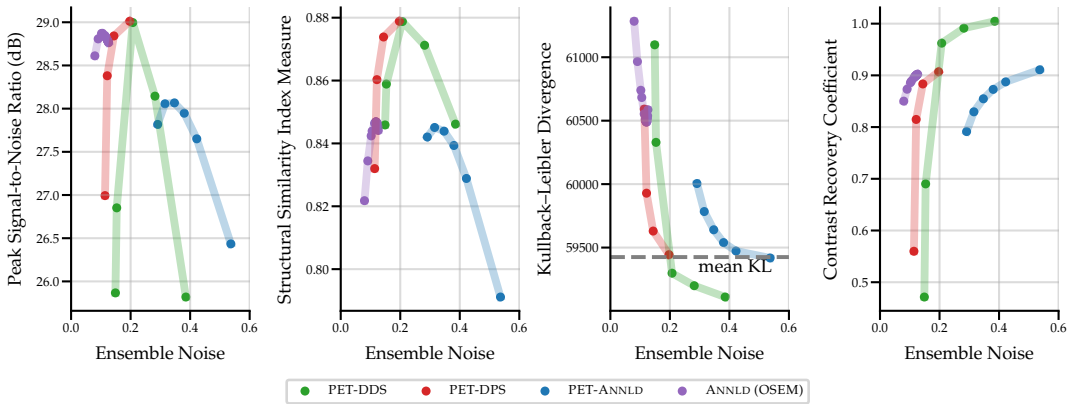


Figure 5.5: Results for BrainWeb **with lesions** with noise level **10** for different penalty strengths. The Standard Deviation is computed over reconstructions of different noise realisations **y**. The points represent different values of the parameter λ.

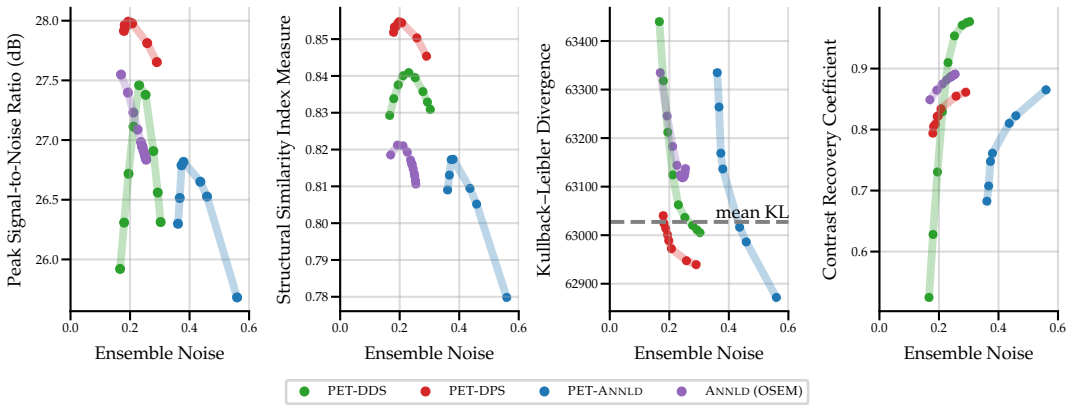


Figure 5.6: Results for BrainWeb **with lesions** with noise level **2.5** for different penalty strengths. Standard deviation is across reconstructions from different realisations of measurements.The points represent different values of the parameter λ.

It can be seen in Table. 5.4 that the CRC is much improved for SGM methods compared to supervised methods across both noise levels. This performance increase can again be attributed to SGMs being unsupervised and more generalisable as compared to supervised methods. Further, PET-DDS performs the best with regards CRC across all methods. Given this performance on the most clinically relevant metric and the low computational overhead, it was chosen that this was the most appropriate method to evaluate in a MR guided and 3D settings.

The choice of quality metric has implicit bias in the features that are assessed, see Section 2.2.4. While both global and local measures are desirable for PET reconstructions, they are not always consistent. For example smoothing may suppress the noise and increase PSNR and SSIM, but it can reduce the contrast, leading to a worse CRC. Likewise, if the method preserves details, it may increase the contrast at the expense of having a higher noise in the image and decreasing PSNR and SSIM. Thus, the results in Table 5.4 indicate that PET-DDS is more effective than PET-DDS in enhancing the contrast of the ROIs, but less effective in suppressing the noise.

[5]Regularised due to denoised score estimate initialisation.

Table 5.4: Results using the best hyperparameters for each method for BrainWeb **with lesions** for noise level **2.5** and **10**. The penalty strength used for each SGM method is denoted by λ. The best score-based method is highlighted in grey. The overall best score per noise level is underlined.

Noise Level	Method	PSNR, λ	SSIM, λ	CRC, λ
2.5	ANNLD (OSEM)	27.60±0.87, 0.527	0.821±0.02, 1.71	0.891±0.02, 50.
	PET-ANNLD	26.82±0.90, 12.	0.817±0.02, 12.	0.908±0.03, 50.
	PET-DPS	27.99±0.85, 625.	<u>0.855</u>±0.01, 650.	0.886±0.02, 1500.
	PET-DDS	27.46±0.83, 0.15	0.841±0.01, 0.15	<u>0.977</u>±0.01, 0.01
	PET-LPD	<u>28.40</u>±0.92, N/A	0.853±0.01, N/A	0.865±0.03, N/A
	PET-U-NET	27.74±0.83, N/A	0.836±0.01, N/A	0.805±0.03, N/A
10	ANNLD (OSEM)	28.87±0.93, 0.25	0.847±0.01, 0.9	0.902±0.02, 4.
	PET-ANNLD	28.07±0.94, 10.	0.845±0.01, 7.5	0.911±0.02, 20.
	PET-DPS	29.01±0.87, 400.	0.878±0.01, 400.	0.920±0.02, 550.
	PET-DDS	28.99±0.88, 0.025	0.879±0.01, 0.025	<u>1.00</u>±0.01, $0.^5$
	PET-LPD	<u>30.07</u>±0.96, N/A	<u>0.894</u>±0.01, N/A	0.904±0.02, N/A
	PET-U-NET	29.41±0.82, N/A	0.889±0.01, N/A	0.865±0.03, N/A

5.7.3 MR Guided 2D Results

MR guided reconstruction with CFG provides the flexibility adjust the level of guidance from the MR image. Three guidance strengths were evaluated $w = 0.25, 0.5, 1.0$, where smaller values of $w \in [0, 1]$ correspond to stronger guidance. The evaluation was conducted with a noise level of 2.5; with and without lesions.

The qualitative results without lesions are shown in Fig. 5.7. These results demonstrate that guidance suppresses hallucinations in grey matter tracts while improving PSNR and SSIM. From the bias-variance graphs in Fig. 5.8, we observe that guided reconstruction achieves high PSNR and SSIM, while lowering EN and increasing KLDIV. This suggests that the guided reconstruction is effectively performing as an image translation process, transforming MR images into PET images. This transformation is sufficient to achieve high performance on global quality metrics, even with elevated KLDIV. Since the images are well-aligned and not significantly out-of-distribution, image translation is expected to perform well. Increasing EN, by increasing data-consistency, reduces performance on global quality metrics. Compared with unguided recon-

struction, both data consistency and global quality metrics show improvement with guidance.

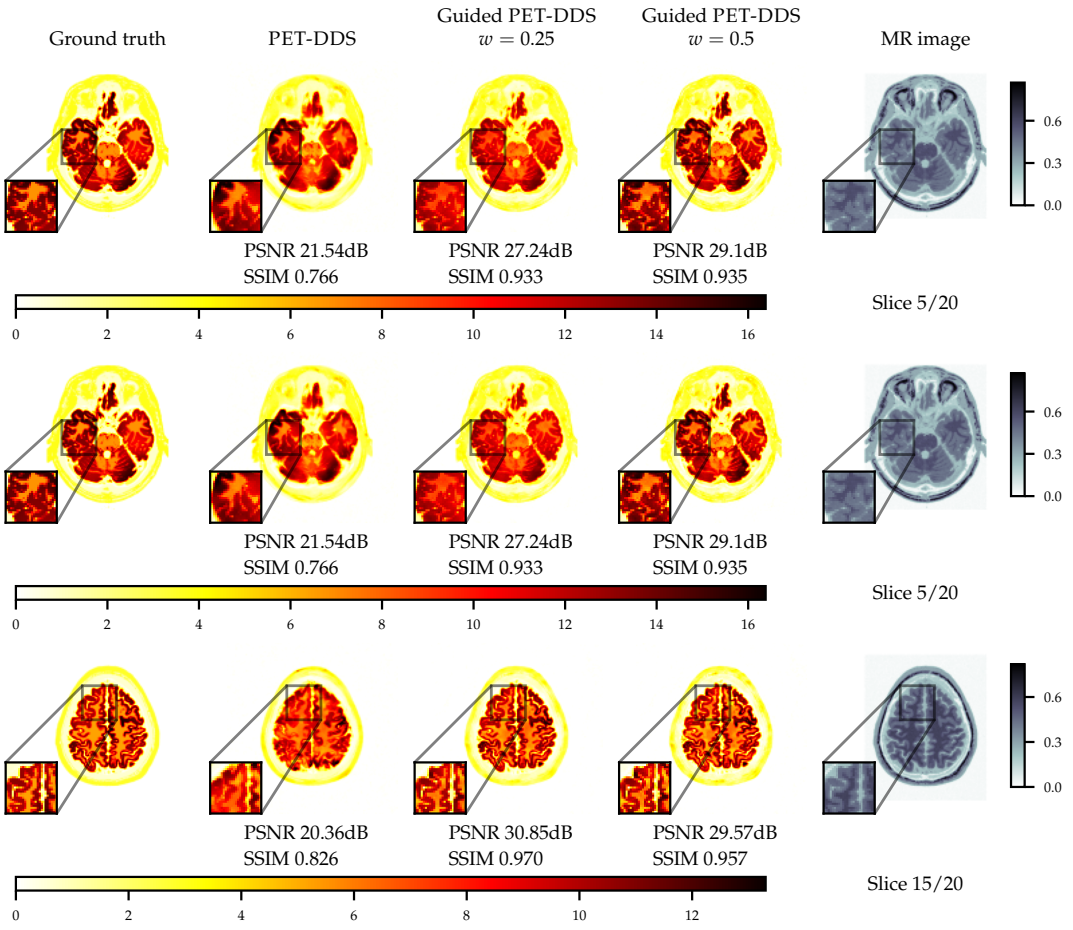


Figure 5.7: Comparisons of the PET-DDS MR guided vs. unguided at noise level **2.5** **without lesions**.

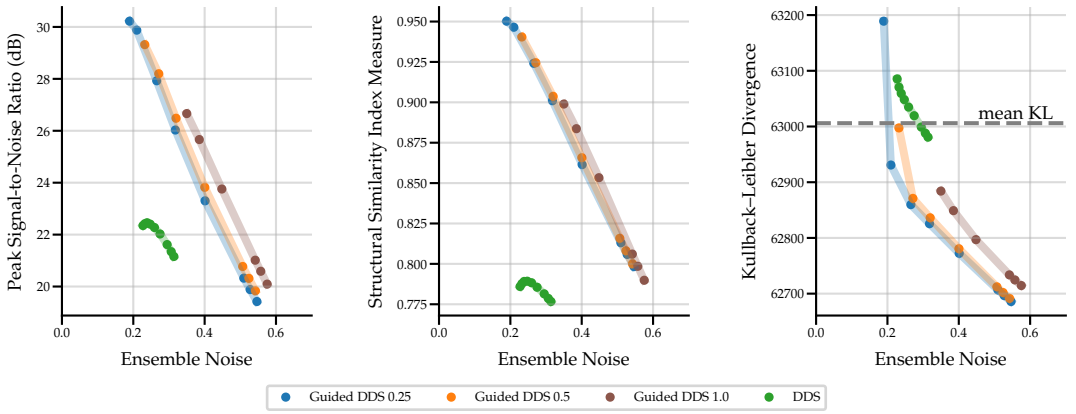


Figure 5.8: Results for 2D reconstruction guided vs unguided **without lesions** for noise level **2.5**. The points represent different values of the parameter λ.

For evaluation with lesions, the MR image does not provide any indica-

tion of abnormal pathology and represents a worst-case scenario for guidance, where the clinically relevant pathology is absent in the guidance modality. As shown in Fig. 5.9, guidance suppresses the lesion in the reconstruction, resulting in a reduction in CRC, while still improving global metrics PSNR and SSIM.

The bias-variance graphs in Fig. 5.10 further emphasizes this observation, highlighting the biases inherent in different quality metrics. Locally, the lesion is not well recovered, but globally, the guidance successfully reconstructs common anatomical features, boosting global quality metrics. This demonstrates both the potential risks of guidance and the critical need to evaluate local and global quality metrics independently.

The quantitative results in Table 5.10 reinforce this finding: while hallucination suppression globally improves metrics due to the recovery of common anatomy, local quality metrics suffer when the guidance is mismatched.

Table 5.5: Results using the best hyperparameters for SGM methods for noise level **2.5** with **MR image guidance**. The penalty strength used for each SGM method is denoted by λ. The best method by performance metric is highlighted in grey for with/without lesion. The penalty strength is tuned for each method individually.

	without lesions		**with lesions**		
	PSNR, λ	**SSIM,** λ	**PSNR,** λ	**SSIM,** λ	**CRC,** λ
DDS (w/o MR)	22.46, 0.25	0.789, 0.2	27.46, 0.15	0.841, 0.15	0.910, 0.01
DDS $w = 0.25$	30.22, 0.35	0.950, 0.35	31.21, 0.15	0.954, 0.25	0.726, 0.0
DDS $w = 0.5$	29.32, 0.25	0.940, 0.25	31.12, 0.15	0.946, 0.25	0.778, 0.0
DDS $w = 1.0$	26.66, 0.15	0.899, 0.15	29.31, 0.1	0.906, 0.15	0.939, 0.0

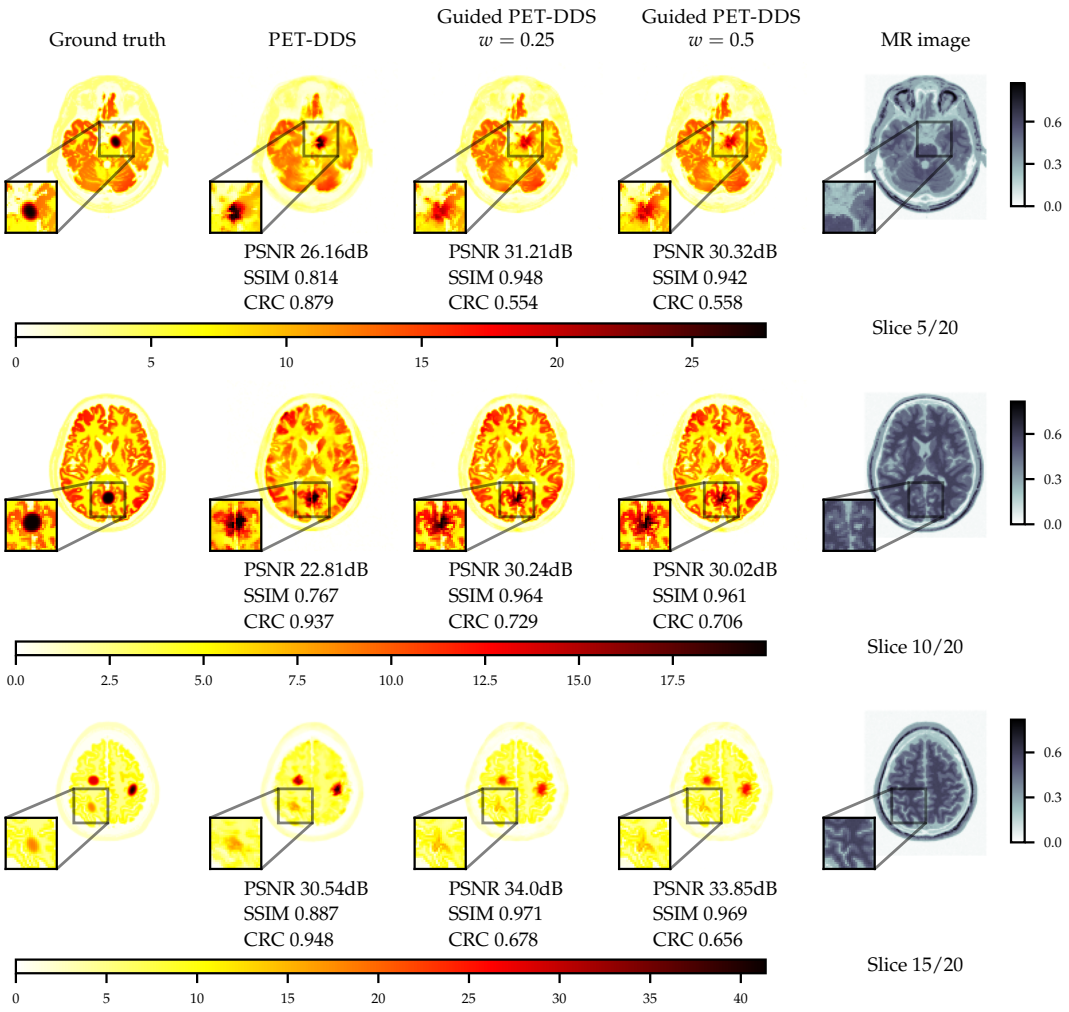


Figure 5.9: Comparisons of single slice reconstructions with the PET-DDS MR guided vs. unguided at noise level **2.5 without lesion** (top) and **with lesion** (bottom).

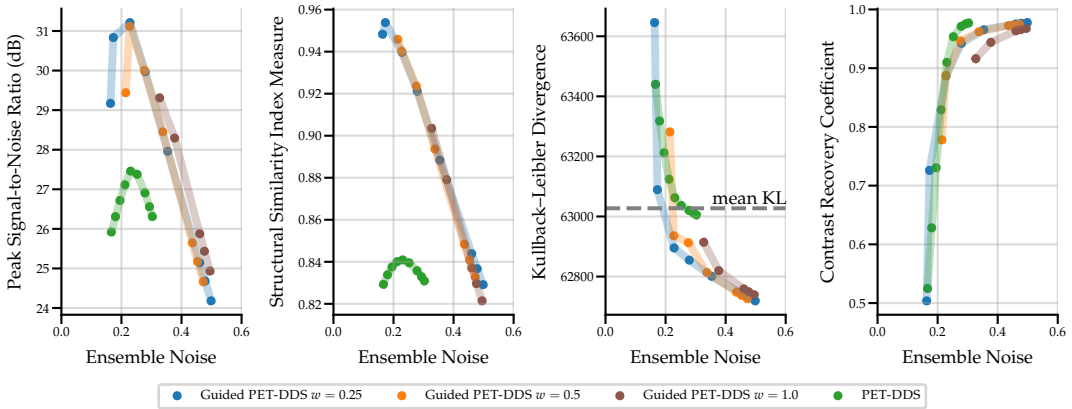


Figure 5.10: Results for 2D reconstruction guided vs unguided **with lesion** for noise level **2.5**. The points represent different values of the parameter λ.

5.7.4 3D Results

The 3D PET-DDS was analysed for fully 3D PET reconstruction with two tracers and simulated lesions. The performance was evaluated with and without RDP regularisation in the z-direction perpendicular to axial slices (termed RDPz), and subset-based data consistency updates as in Eqn. (**??**). The use of subsets accelerate reconstruction by evaluating the forward model block sequentially. The trade off between quality metrics and number of subsets can be seen in Table. 5.6. The higher the number of the subset, the coarser the approximation of the gradient update with regards to the full update. The number of subsets used for subsequent experiments used in clinical practice was 28, and this was the same used in BSREM computed MAP estimates.

Table 5.6: 3D PET-DDS+RDPz computing time with different numbers of subsets. Quality metrics computed on the first realisation of FDG tracer measurements using 3D PET-DDS+RDPz with $\lambda = 158.0$ and $\beta = 21.9$. Best values are highlighted in grey.

Number of subsets	**1**	**4**	**7**	**14**	**28**	**42**
Reconstruction time (min)	47.8	13.6	8.6	5.1	3.4	2.8
PSNR	25.91	25.90	25.89	25.84	25.72	25.57
SSIM	0.927	0.927	0.927	0.925	0.922	0.919
CRC	0.990	0.990	0.990	0.985	0.987	0.988

From the bias-variance graphs Figs. 5.11 and 5.12 for FDG and Amyloid tracers respectively, the improvements to PSNR and SSIM are apparent with PET-DDS+RDPz as opposed to PET-DDS. Both PET-DDS+RDPz and PET-DDS outperform the comparison methods in terms of CRC, while RDP provides the highest PSNR values.

From the reconstructions given in Figs. 5.13 and 5.14, it can be observed that the SGM-based methods provide sharper, less blurry, images. This gives an insight into the success of RDP reconstruction for PSNR as this metric is biased toward blurry images, meaning blurry images still give good performance for that metric. Comparing PET-DDS and PET-DDS+RDPz it can be seen that the axial slices are coherent for both methods, but for coronal and sagittal there are discontinuities between slices due to the 2D score-model only being applied axially. This highlights the need for additional regularisation to ensure

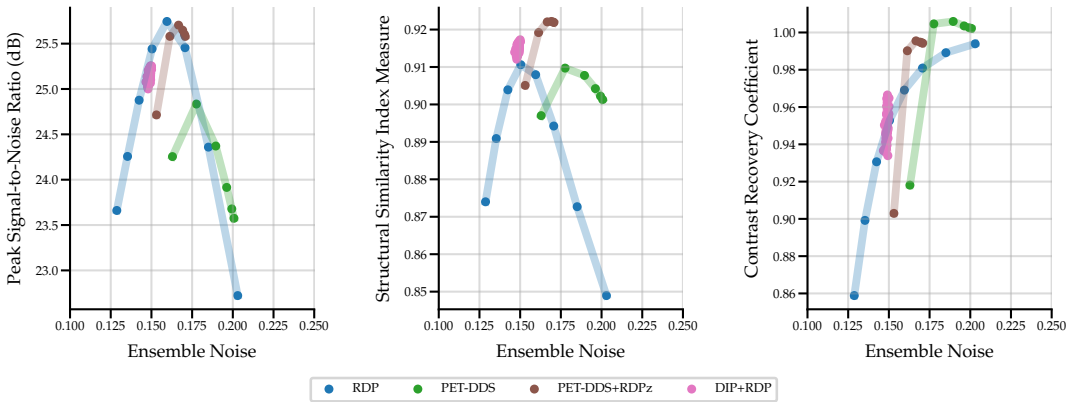


Figure 5.11: Results for 3D reconstruction using the FDG tracer for different penalty values. PET-DDS-RDP$_z$ β = 21.9, and DIP+RDP β = 0.1. Standard deviation is across reconstructions from different realisations of measurements. For DIP, the points corresponds to various number of optimisation steps. For the other methods, the points represent different values of the parameter λ.

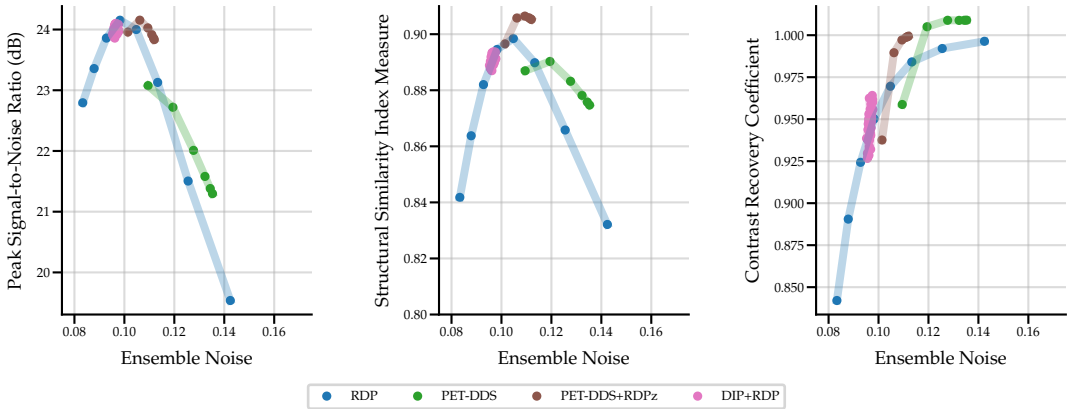


Figure 5.12: Results for 3D reconstruction using the Amyloid tracer for different penalty values. PET-DDS-RDP$_z$ β = 21.9, and DIP+RDP β = 0.1. Standard deviation is across reconstructions from different realisations of measurements. For DIP, the points corresponds to various number of optimisation steps. For the other methods, the points represent different values of the parameter λ.

coherence in the direction orthogonal to that which the score-model is applied. From Table. 5.7, the ability for the SGM methods to work well for both FDG and Amyloid can be observed and highlights the generalisation capability of SGMs.

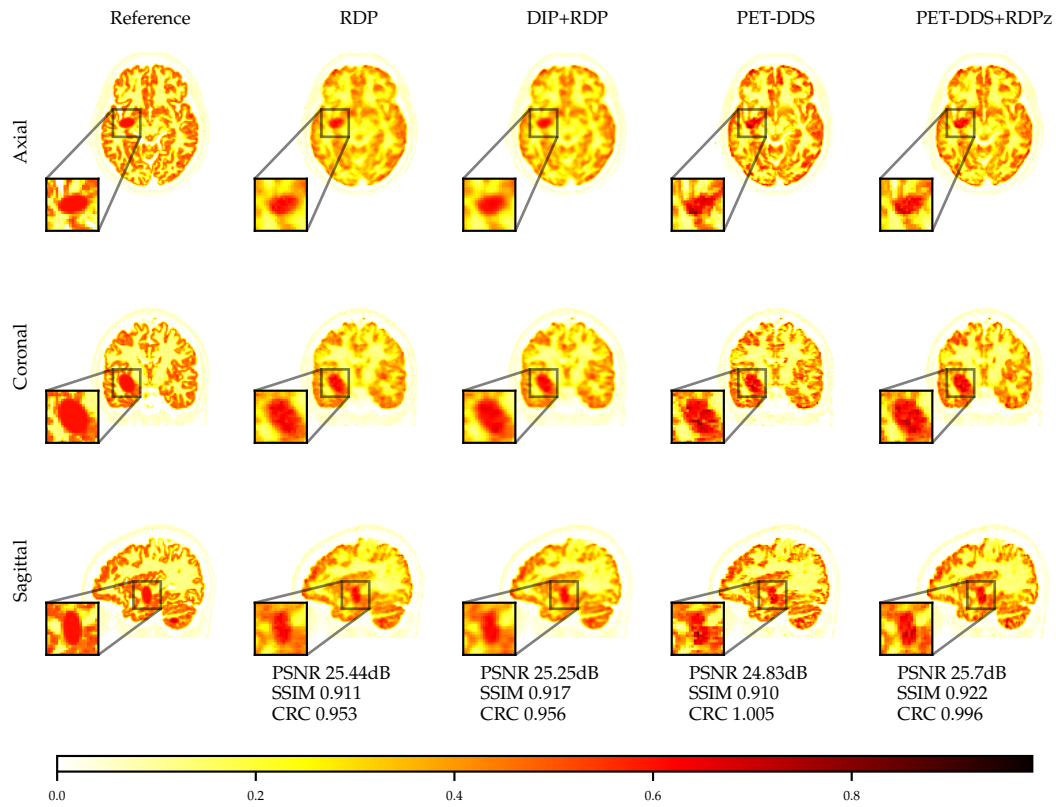


Figure 5.13: 3D reconstruction for the different methods with FDG tracer, and metrics computed on the inset lesion.

Table 5.7: Results using the best hyperparameters for each method for 3D BrainWeb data with FDG and Amyloid tracers. The penalty strength used for each SGM method is denoted by λ. The best performing method is highlighted in grey.

Tracer	Method	PSNR, λ	SSIM, λ	CRC, λ
FDG	**RDP**	25.74, 1.81	0.911, 2.77	0.994, 0.5
	DIP+RDP	25.26, 9,800	0.917, 10,800	0.966, 9,500
	PET-DDS	24.83, 398	0.910, 398	1.01, 158
	PET-DDS+RDPz	25.70, 158	0.922, 63.1	0.996, 158
Amyloid	**RDP**	24.15, 2.77	0.898, 1.81	0.996, 0.5
	DIP+RDP	24.10, 10,200	0.894, 10,800	0.964, 9,500
	PET-DDS	23.08, 1000	0.890, 398	1.009, 10
	PET-DDS+RDPz	24.15, 398	0.906, 158	0.999, 10

5.8 Conclusion

In this work, we adapted SGMs for PET image reconstruction by incorporating PET-specific constraints, such as Poisson noise and non-negativity, into popular sampling techniques. Additionally, we introduced a measurement-based

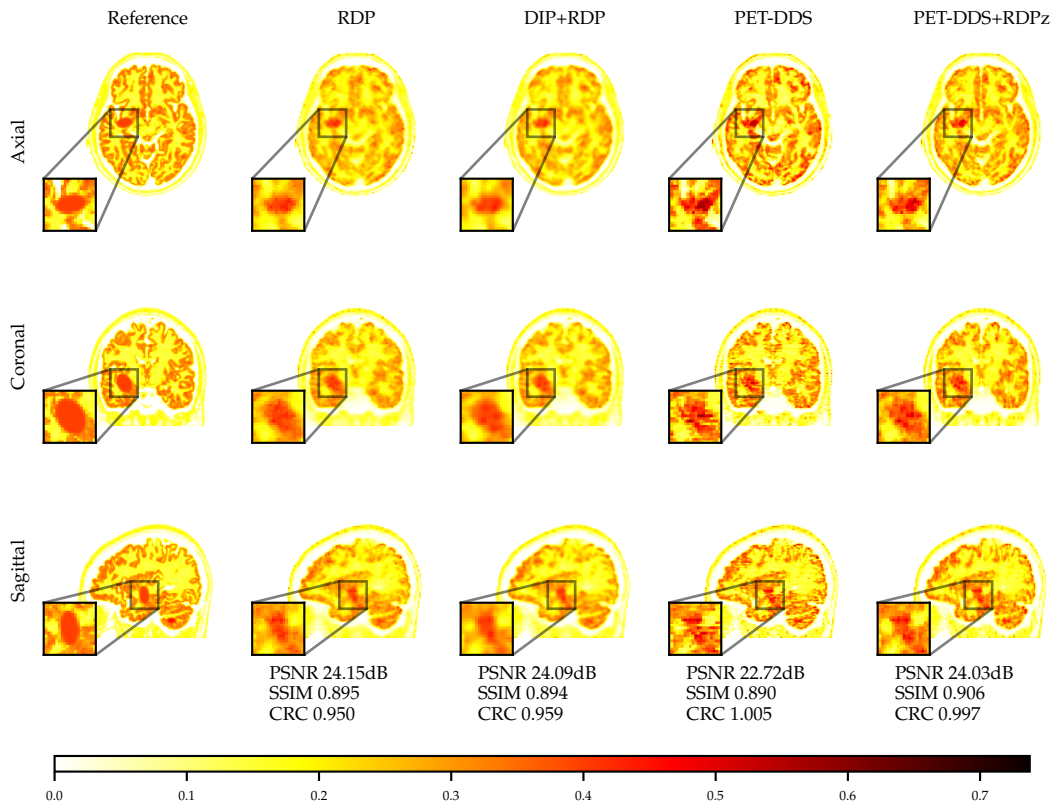


Figure 5.14: 3D reconstruction for the different method with Amyloid tracer, and metrics computed on inset lesion.

normalisation technique to stabilise the dynamic range. Experimental results demonstrate that integrating these PET-specific constraints within SGMs leads to improved image quality, effectively leveraging the strengths of SGMs for image modelling while separating the forward modelling, where known physics are integrated into the sampling. The obtained samples approximately represent the posterior p^{post}.

Further work incorporating reflected SGMs could introduce non-negativity into the sampling procedure in a more principled manner [145], further enhancing the theoretical robustness of the reconstruction. Additional further work is described in Section 7.2.

While the U-Net architecture has proven highly effective in modelling complex image structures within the SGMs framework, its theoretical foundations remain an area for ongoing study, as explored in [111, 255]. Moreover, there is a need for deeper theoretical validation of whether SGMs accurately sample from the true posterior $p(\mathbf{x}|\mathbf{y})$, especially in the context of inverse problems such

as PET imaging. Recent work [67] provides a foundation for establishing the mathematical guarantees and convergence properties of these methods.

Finally, our approach highlights the potential of guided SGM reconstruction with multimodal data, such as MR-guided PET imaging. Preliminary results suggest that incorporating an additional MR guidance image via CFG can enhance reconstruction quality. An interesting extension could be the synergistic reconstruction for PET/MR with joint prior [7, 45, 139].

5.8.1 Limitations

Despite the promising outcomes, several limitations warrant discussion. First and foremost, all experiments were conducted *in-silico* using simulated data. While simulation studies allow for controlled experimentation and initial validation, they do not fully capture the complexities and variabilities present in clinical PET imaging. The ground truth images used to define the prior distribution π may possess higher quality and less variability than real clinical data, potentially leading to reconstructions that are unrealistically good compared to what is achievable in practice.

Furthermore, our investigation into the generalisation capabilities of the score model was limited. The model was trained on patient-realistic slices devoid of lesions and tested on slices containing lesions. Although this provides an initial assessment of the model's ability to generalise to unseen pathological features, it does not encompass the full spectrum of anatomical and pathological variations found in diverse patient populations. A comprehensive evaluation on more extensive and varied datasets, including real *in-vivo* data, is essential to thoroughly understand the biases and limitations of SGMs in a clinical context.

Chapter 6

Adaptable-Blobs

6.1 Introduction

In this chapter, we introduce the Gaussian radial basis function "blobs" with adaptable covariance (shape) and centroid (location), for which we coin the term A-Blobs. This A-Blobs representation is a continuous functional representation that is developed to represent images and formulate inverse problems -with relevant literature briefly outlined in Sections 2.2.1 and 2.3.5.3.

It is important to note that methods such as DIP and SGMs applied data-consistency principles to ensure faithfulness to measurements. Despite their success and apparent generalisability, the underlying mechanisms of these methods remain poorly understood. The neural networks used in previous chapters, primarily U-Nets, are highly over-parameterised, obscuring the specific factors contributing to their effectiveness. Additionally, these networks are constrained to fixed image sizes, such as 128×128, which limits their flexibility and generalisability. The A-Blobs allow for a continuous function representation that is more interpretable as it is a linear sum of basis functions.

This chapter begins with an overview of relevant literature and discussion on alternatives to voxel-based representations. We then develop the A-Blobs framework for parallel MRI and present the results published in [219].

6.2 Alternatives to Voxels

As described in Section 2.2.1, a 3D medical image is defined as an object in a continuous three-dimensional space given by coordinate $\mathbf{r} = [r_x, r_y, r_z]^\top$. The object is a scalar-valued function $x(\mathbf{r}) : \mathbb{R}^3 \to \mathbb{F}$ where the value is the quantity of interest, i.e. AC for PET or transverse magnetisation for MRI etc. This function is discretised as is required for digital computation. The most common choice for discretisation is voxels. This choice is informed by their simplicity, linearity, regularity, uniformity, and efficiency in representation. However, if the object to be represented takes up only a small portion of the overall volume domain this can lead to an inefficient representation. Additionally, the piece-wise constant nature of voxels can easily represent high frequency details corresponding to noise and artifacts. There exists a trade-off when considering the representation for medical image reconstruction [84]. The representations considered are a discrete linear combinations of basis-functions:

$$x(\mathbf{r}) = \sum_{b=1}^{N_b} c_{(b)} \psi_{(b)}(\mathbf{r}) \tag{6.1}$$

where $c_{(b)}$ and $\psi_{(b)}$ are the b-th coefficient and basis respectively. The summation is the linear combination of all N_b basis functions. Note that this gives the functional definition of an image, and a digital image is sampled at a discrete set of spatial locations. The discrete set of spatial locations is denoted $\Omega_{\mathbf{r}} = \{(\mathbf{r}_{(i)})\}_{i=1}^{N_{\mathbf{r}}} \subset \mathbb{R}^{N_{\text{dim}}}$, where $N_{\mathbf{r}}$ are the number of sample points for the discrete image. This defines the discrete image as:[1]

$$\mathbf{x} = x(\mathbf{r})|_{\Omega_{\mathbf{r}}}.$$

Given a fixed number of basis functions N_b and $N_{\mathbf{r}}$, the basis-transformation matrix is given as:

$$\mathbf{\Psi} \in \mathbb{F}^{N_b \times N_{\mathbf{r}}} \quad \text{where} \quad [\mathbf{\Psi}]_{ij} = \psi_{(i)}(\mathbf{r}_{(j)}).$$

From this basis-transformation matrix the discrete image is:

[1] There is an implicit assumption that the image is well-approximated or exactly represented within this subspace.

$$\mathbf{x} = \mathbf{\Psi c},$$

where $\mathbf{c} \in \mathbb{F}^{N_b}$ is the coefficient vector. In this context a basis vector is an element of the basis-transformation matrix such that $\mathbf{\Phi} = [\boldsymbol{\psi}_{(1)}, \boldsymbol{\psi}_{(2)}, \ldots, \boldsymbol{\psi}_{(N_b)}]$ where the basis vector is $\boldsymbol{\psi}_{(b)} = \psi_{(b)}|_{\Omega_{\mathbf{r}}}$.

6.2.1 Local Basis Functions

Voxels are a regular grid of non-overlapping piece-wise constant local basis functions, where $N_b = N_{\mathbf{r}}$ and $\mathbf{\Psi} = \mathbf{I}$. They have a fixed size that define the resolution of the discretised image. Other local basis function have been proposed, for example triangular, B-Splines, Gaussian, and Kaiser-Bessel [141].

Globally-supported functions such as Gaussians decay quickly in space but are globally defined; as such they are often truncated in this context to be locally defined [84, 208]. The choice of local basis function can have favourable properties for the resulting medical image reconstruction. For example, as shown in [150], 3D PET image reconstruction can benefit greatly from using alternative basis functions; improving contrast recovery and reducing noise. Smoother basis functions such as Gaussians and Kaiser-Bessel are often referred to as "blobs" in this literature, and we adopt this convention.

6.2.2 Non-local and Data-driven Adaptive Basis Functions

Non-local basis functions can capture long-range dependencies within an image and represent global structures that allow for more compressed representations. A canonical example of non-local basis functions is the Fourier basis, which represents images in terms of global frequency components without spatial localisation. Wavelets [48] incorporate both frequency and spatial localisation, offering a hybrid between local and non-local basis functions. This makes them useful in image processing, as they provide a joint space-frequency representation of an image. Wavelet transforms can be continuous or discrete, and they may form either linearly independent basis functions or frames.

In a discrete setting the basis-transformation matrix $\mathbf{\Psi}$ has linearly indepen-

dent basis vectors $\boldsymbol{\psi}_{(b)}$ if the only solution to $\boldsymbol{\Psi}\mathbf{c} = \mathbf{0}$ is $\mathbf{c} = \mathbf{0}$. When $\boldsymbol{\Psi}$ spans the entire space $\mathbf{x} \in \mathcal{X}$ the discrete image can be exactly represented by $\mathbf{x} = \boldsymbol{\Psi}\mathbf{c}$. In other words, if $\boldsymbol{\Psi}$ is linearly independent and spans $\mathcal{X}$, it forms a basis, in the strict sense, for $\mathcal{X}$.

The Haar wavelet is one of the simplest wavelets, providing a compactly supported, linearly independent basis. Each Haar wavelet vector is piecewise constant and localised in space. The Haar basis vectors are shift and scaled versions of the Haar wavelet function, that in 1D is:

$$\psi(r) = \begin{cases} 1 & 0 \leq r < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq r < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The resulting basis-transformation matrix is orthonormal - i.e. $\boldsymbol{\Psi}^\top\boldsymbol{\Psi} = \mathbf{I}$. Many other families of other wavelets exist [48, 148], one family is used for image compression with the JPEG image file format - the most widely used image compression standard in the world [102].[2] Wavelets have also been extensively used in image reconstruction as natural images have been observed to be sparser when represented with wavelet coefficients, allowing for denoising of MRI reconstructions via shrinkage [147]. A reconstruction method enforcing sparsity of wavelets coefficients is formulated as:

$$\mathbf{x}^\dagger = \underset{\mathbf{x}}{\operatorname{argmin}}\, \rho(\mathbf{y}, \mathbf{A}\mathbf{x}) + \lambda \|\boldsymbol{\Psi}^{-1}\mathbf{x}\|_1,$$

where $\boldsymbol{\Psi}^{-1}$ is a discrete sparsifying transform - in-fact the matrix $\boldsymbol{\Psi}$ is usually orthonormal meaning $\boldsymbol{\Psi}^{-1} = \boldsymbol{\Psi}^\top$. As described in Section 2.1.1, the ℓ_1-norm encourages sparsity.

A frame for $\mathcal{X}$ is the matrix $\boldsymbol{\Psi}$ that satisfies the frame inequality:

$$a\|\mathbf{x}\|_2^2 \leq \|\boldsymbol{\Psi}^\top\mathbf{x}\|_2^2 \leq b\|\mathbf{x}\|_2^2,$$

for all $\mathbf{x} \in \mathcal{X}$, with $0 < a \leq b < \infty$ as frame bounds. Frames allow for over-

[2] Originally JPEG used cosine transform but the more modern standard (JPEG-2000) uses a wavelet transform.

completeness, meaning $\mathbf{\Psi}$ may have more columns than the dimension of $\mathcal{X}$, enabling robust representations even with linearly dependent columns.

The Gabor wavelet forms an over-complete frame with linearly dependent set of basis vectors. A Gabor wavelet in one dimension is defined as:

$$\phi(r) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{r^2}{2\sigma^2}} e^{\mathrm{i}2\pi u r},$$

where u is the central frequency and σ controls the scale of the Gaussian envelope. The redundancy of this over-complete frame provides a robust representation even in the presence of noise or incomplete data, making it ideal for applications where resilience to perturbations is beneficial [136, 148].

In contrast to being hand-crafted, basis vectors can be defined from a dataset as is the case with dictionary learning [3]. This approach is among the earliest example of machine learning for medical image reconstruction. For Electrical Impedance Tomography, Principle Component Analysis was used to define a dictionary to constrain the solution with a basis vectors defined from an anatomical dataset [248]. In a similar vein to wavelets the discrete image can be represented by:

$$\mathbf{x} = \mathbf{D}\mathbf{c},$$

where the dictionary $\mathbf{D}$ is a collection of N_a atoms $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2 ..., \mathbf{d}_{N_a}]$ that, when linearly combined, can approximate the image [64]. While dictionary learning is a powerful technique, dictionaries do not need to be data-driven. As described in [39], dictionaries can include over-complete basis vectors such as the Gabor wavelets introduced earlier. As an image can be non-uniquely represented, there is a possibility for an adaptive representation, i.e. choosing out of many representations one that is most appropriate for the task at hand. For inverse problems this can be formulated as:

$$\mathbf{c}^\dagger = \underset{\mathbf{c}}{\operatorname{argmin}} \rho(\mathbf{y}, \mathbf{A}\mathbf{D}\mathbf{c}) + \lambda \|\mathbf{c}\|_1, \quad \text{where} \quad \mathbf{x} = \mathbf{D}\mathbf{c}.$$

Similar to wavelet reconstruction, sparsity is encouraged through the ℓ_1-norm. The goals of adaptive representation, as outlined in [39], are: sparse-

ness of the representation within the dictionary, higher resolution than non-adaptive methods, and speed of computation of the representation.

6.2.3 Implicit Neural Representations

INRs are a new paradigm that represents continuous images using neural networks, see Section 2.3.5.3 for a brief overview. With INRs MLPs are used to map spatial coordinates directly to image values [221]. INRs leverage the universal function approximation property of neural networks to approximate continuous image functions.

Considering the MLP as formulated in Eqn. (2.3.1), INRs are fit to an image $\mathbf{x}_{\text{GT}}$ through the following optimisation:[3]

$$\theta^{\dagger} = \underset{\theta}{\operatorname{argmin}} \|\mathbf{x}_{\text{GT}} - \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{r})\big|_{\Omega_{\mathbf{r}}}\|_2^2.$$

The choice of activation function of the MLP plays a crucial role in the representational capacity of the network. For example, sinusoidal activation functions [221], Gaussian activations [183], and complex Gabor wavelet activations [199] have been proposed to improve the network's ability to represent complex signals. These different activations have different properties that can make the optimisation more prone to poor local-minima, and if initialised incorrectly the network may not be able to successfully represent the object [183].

This highlights the two open problems with INRs: Sensitivity to initialisation, where non-convex objective function can lead the optimisation to poorly performing local-minima. Given the high resolution and dimensionality of medical images the evaluation of $\mathcal{G}_{\boldsymbol{\theta}}$ at every sample point $N_{\mathbf{r}}$ becomes quickly infeasible.

That said, the ability to approximate more complex images increases significantly with depth of the network layers. This representation can be considered as a structured dictionary [261], where the network's parameters implicitly define a set of low-rank basis functions as they have shared parameters. Further, INRs have been applied to inverse problems, where the reconstruction is cast

[3] As the output of layer is often linear, we can consider an MLP as a linear sum of a non-linear low-rank set of basis functions.

as an optimisation problem over the network parameters, minimising the discrepancy between the measurements and the forward model applied to the network's output [210]. Note, that in that work the network parameters were initialised with a prior image.

6.2.4 Gaussian Splatting

For 3D graphics, specifically radiance fields, Gaussian Splatting methods are developing into the de facto choice for volumetric rendering [121]. In Gaussian Splatting the volume is represented with Gaussian functions that are fitted, optimising the covariance (shape), centroid (location) and other parameters, to the volume of interest. This fitting and subsequent rendering utilises the fast splatting procedure to approximate the projection of the 3D Gaussians to a 2D plane. This models the fact that multiple images from multiple views are used to generate a 3D volume. Splatting allows for these images to be quickly approximated, allowing for the fitting of Gaussians as well as rendering of new views [268].

The formulation of Gaussians to represent a volumetric object linear summation of Gaussians -Eqn. (6.1). In this case the basis ψ is dependent on the centroid and covariance of the Gaussian, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively, such that:

$$\psi(\mathbf{r}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left(-\frac{(\mathbf{r}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{r}-\boldsymbol{\mu})}{2}\right).$$

In the original Gaussian splatting paper the covariance was parameterised as three components scaling, rotation, and skew.

6.2.5 Alternatives to Voxels for Medical Image Reconstruction

Analysing the previously introduced alternatives to voxels requires a set of criteria specific to medical image reconstruction. In literature, work has already analysed local basis functions for CT image reconstruction [84]. In that paper a set of favourable basis function properties is defined as:

1. The basis set should exhibit linear independence, meaning no basis function can be expressed as a linear combination of the others. This ensures

uniqueness of the coefficients $c_{(b)}$, avoiding redundancy.

2. Power of Approximation: As the number of basis functions N_b increases, the approximation error $\|x(\mathbf{r}) - \sum_{b=1}^{N_b} c_{(b)}\psi_{(b)}(\mathbf{r})\|$ should decrease at a desirable rate. Ideally, the error decreases rapidly with increasing N_b, indicating an efficient representation.

3. Insensitivity to Shift of Basis Function Set: A shift in the placement of the basis functions within the domain should not cause a significant drop in the approximation power. This property ensures robustness to the spatial configuration of the basis functions.

4. Efficient Computation of Projections and Back-Projections: The basis functions should allow for efficient computation of the forward model (projections) and its adjoint (back-projections), which are essential operations in medical image reconstruction.

5. Efficient Implementation of Reconstruction Constraints: Constraints such as non-negativity or spatial support should be easily incorporated into the reconstruction algorithm, possibly by choosing basis functions that inherently satisfy certain constraints.

6. Fidelity of Visual Appearance: The reconstructed images should be visually accurate, preserving important features and details necessary for clinical interpretation.

The insensitivity to a shift of the basis functions is not relevant for adaptive or data-driven basis functions. As such, this criteria can be ignored. In-fact, the sensitivity to shifts indicates the basis functions has been informed by the structure of represented object, which can be useful.

For adaptive basis sets like overcomplete dictionaries and frames, the strong linear independence criterion can be relaxed. While overcomplete dictionaries introduce redundancy, sparsity-promoting techniques can be used to select a subset of basis functions, leading to unique and efficient representations [63]. Frames, being overcomplete, provide stable and robust representations, but the

non-uniqueness requires additional constraints like sparsity for practical use [148].

Additionally, for strong linear independence the space in-which the image lies needs to be of fixed regularity or resolution. If an image can be represented with a linear independent basis at a fixed resolution, if the resolution is increased (number of sample points is increased) the previously linear independent basis is under-complete. The contrary is true too, if the resolution is decreased (fewer sample points) the basis would then be over-complete. Linear independence can only be sought for images of fixed resolution. For reconstruction of images that are resolution-agnostic the strong linear independence is unattainable.

6.3 Adaptable Blobs

Blobs, as described in the preceding sections, are a family of smooth local-basis functions that can be used to represent an image. In the rest of this chapter the term A-Blobs is used to describe Gaussian blobs with parameterised covariance (shape) and centroid (location). This is related to Gaussian splatting, where A-Blobs are Gaussian splats without splatting procedure [268].

Another perspective from Poggio and Girosi [176], is that A-Blobs is an approximation scheme, that can be mapped into a neural network. This has links with Regularization Networks and Hyper Basis Functions, as an approximation scheme an important question, neglected in this thesis, is: "how many samples are needed to achieve a given degree of accuracy?" This accuracy is dependent on the number of coordinate $N_{\mathbf{r}}$ and the smoothness of the function to be approximated [231].

In contrast with the original 3D Gaussian splatting paper [121], more recent work [56] parameterises the inverse covariance through the Cholesky decomposition with a lower triangular matrix $\mathbf{L}$ such that $\mathbf{\Sigma}^{-1} = \mathbf{L}^{\top}\mathbf{L}$. This ensures the covariance is semi-positive definite for any real values of the lower diagonal, i.e. there are no non-negativity constraints as required by the earlier parameterisation. Additionally, the eigenvalues that dictate the scale of each A-Blobs are

simply the squares of the leading diagonal terms of $\mathbf{L}$. This allows for more controllability of the blobs, where minimum or maximum scale can be enforced, and the skew of the blobs is simply the ratio between the largest and smallest eigenvalue, i.e. the condition number. For these reasons the lower triangular matrix parameterisation is used in this chapter where the basis function can be written:

$$\psi(\mathbf{r};\boldsymbol{\mu},\mathbf{L}) = \exp\left(-\frac{\|\mathbf{L}(\mathbf{r}-\boldsymbol{\mu})\|_2^2}{2}\right).$$

Further we define the upper triangular matrix $\mathbf{U}$ as $\mathbf{U} = \mathbf{L}^{-\top}$ which in-turn defines $\mathbf{L} = \mathbf{U}^{-\top}$, this is important for notational simplicity in the subsequent sections such that $\Sigma = \mathbf{U}^\top\mathbf{U}$.

Given the unique properties of Gaussians, the forward model for inverse problems can begin to be formulated directly. This work is in 2D where we first demonstrate image representation and Fourier image representation, and then define a forward model for Parallel MRI. The latter is developed with additional regularisation functions to aid the reconstruction of the MRI images from noisy and undersampled parallel MRI measurements.

6.3.1 Image Representation and Fourier Image Representation

A single adaptable blob is given by:

$$b(\mathbf{r};c,\boldsymbol{\mu},\mathbf{L}) = c\psi(\mathbf{r};\boldsymbol{\mu},\mathbf{L}) = c\exp\left(-\frac{\|\mathbf{L}(\mathbf{r}-\boldsymbol{\mu})\|_2^2}{2}\right),$$

where c, $\boldsymbol{\mu}$ and $\mathbf{L}$ are parameters to be optimised. Only the non-zero elements of $\mathbf{L}$ are used in the parameterisation, ensuring that the lower triangular form is enforced. The image is represented as a linear combination of these A-Blobs:

$$x(\mathbf{r};\mathbf{c},\boldsymbol{\Gamma}) = \sum_{b=1}^{N_b} b(\mathbf{r};c_{(b)},\boldsymbol{\mu}_{(b)},\mathbf{L}_{(b)}) = \sum_{b=1}^{N_b} \mathbf{c}_{(b)}\exp\left(-\frac{\|\mathbf{L}_{(b)}(\mathbf{r}-\boldsymbol{\mu}_{(b)})\|^2}{2}\right),$$

with N_b the number of A-Blobs discretising the object, $\mathbf{c} = \{(c_{(b)})\}_{b=1}^{N_b}$ and $\boldsymbol{\Gamma} = \{(\mathbf{L}_{(b)},\boldsymbol{\mu}_{(b)})\}_{b=1}^{N_b}$.

To assess the ability of A-Blobs to represent a medical image in 2D, a varying numbers of A-Blobs are used to represent a single axial T2-weighted BrainWeb MR image. As before, the notation $x(\mathbf{r})|_{\Omega_\mathbf{r}}$ is used to denote the A-Blobs representation evaluated at all the pixel locations, with $N_\mathbf{r} = 256 \times 256$. Further, the shorthand $\|\mathbf{x}_{gt} - x(\mathbf{r})|_{\Omega_\mathbf{r}}\|_2^2$ is the MSE between ground-truth pixel represented BrainWeb and A-Blobs represented BrainWeb.

In addition to the fully parameterised A-Blobs representation, we compare with A-Blobs representations that are partially parameterised. More specifically, increasing levels of parameterisation are evaluated. Starting with only the coefficients $\mathbf{c}$ parameterised on fixed grid with fixed size and isotropic shape, similar to previous work [84, 208]. Note, isotropic blobs have one scaling factor controlling the size that can be described as $\mathbf{L} = s\mathbf{I}$, where s is the scale. For the next level of parameterisation the coefficient and the scale of the blobs is parameterised whilst being on a fixed grid. The next level additionally parameterises the centroid of the blobs, and finally the full A-Blobs parameterises coefficient, centroid and full covariance through the lower triangular matrix. Table. 6.1 summarises the levels of parameterisation:

Parameterisation	coeff	coeff & iso cov	coeff, iso cov & cent	A-Blobs
$\boldsymbol{\theta} :=$	$c_{(b)}$	$c_{(b)}, s_{(b)}\mathbf{I}$	$c_{(b)}, \boldsymbol{\mu}_{(b)}, s_{(b)}\mathbf{I}$	$c_{(b)}, \boldsymbol{\mu}_{(b)}, \mathbf{L}_{(b)}$

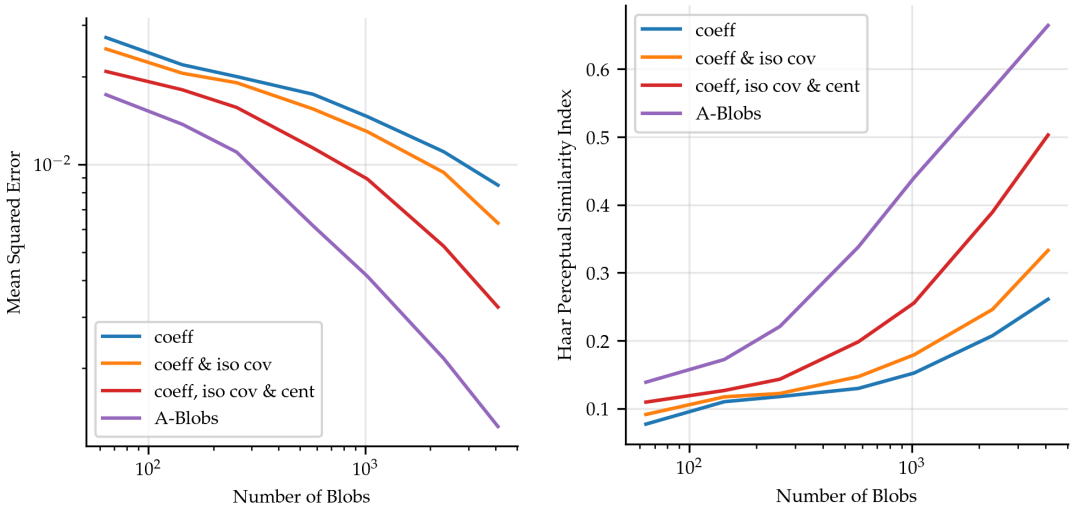
Table 6.1: Different parameterisations evaluated for image representation with A-Blobs

To represent the image $\mathbf{x}_{\text{GT}}$ with A-Blobs, the following optimisation problem is solved:

$$\boldsymbol{\theta}^* \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\mathbf{x}_{gt} - x(\mathbf{r}; \boldsymbol{\theta})|_{\Omega_\mathbf{r}}\|_2^2. \tag{6.2}$$

The above optimisation problem is solved using Adam optimiser [124]. The initialisation of the A-Blobs is as a grid of isotopic covariance scaled such that the overlap between A-Blobs is able to fully cover the domain as described in [176, 208], i.e. $s \propto \sqrt{\Delta x}$ with Δx the spacing between blob centroids on the grid. The size of the grid is set as $[8 \times 8,\ 12 \times 12,\ 16 \times 16,\ 24 \times 24,\ 32 \times 32,\ 48 \times 48,\ 64 \times 64]$.

Given the properties of Gaussian functions the Fourier-transformed A-Blobs



(**a**) Image quality assessed with Mean Squared Error vs number of A-Blobs.

(**b**) Image quality assessed with Haar Wavelet-Based Perceptual Similarity Index vs number of A-Blobs.

Figure 6.1: Plots assessing image representation of A-Blobs with varying degree of parameterisation and number of A-Blobs.

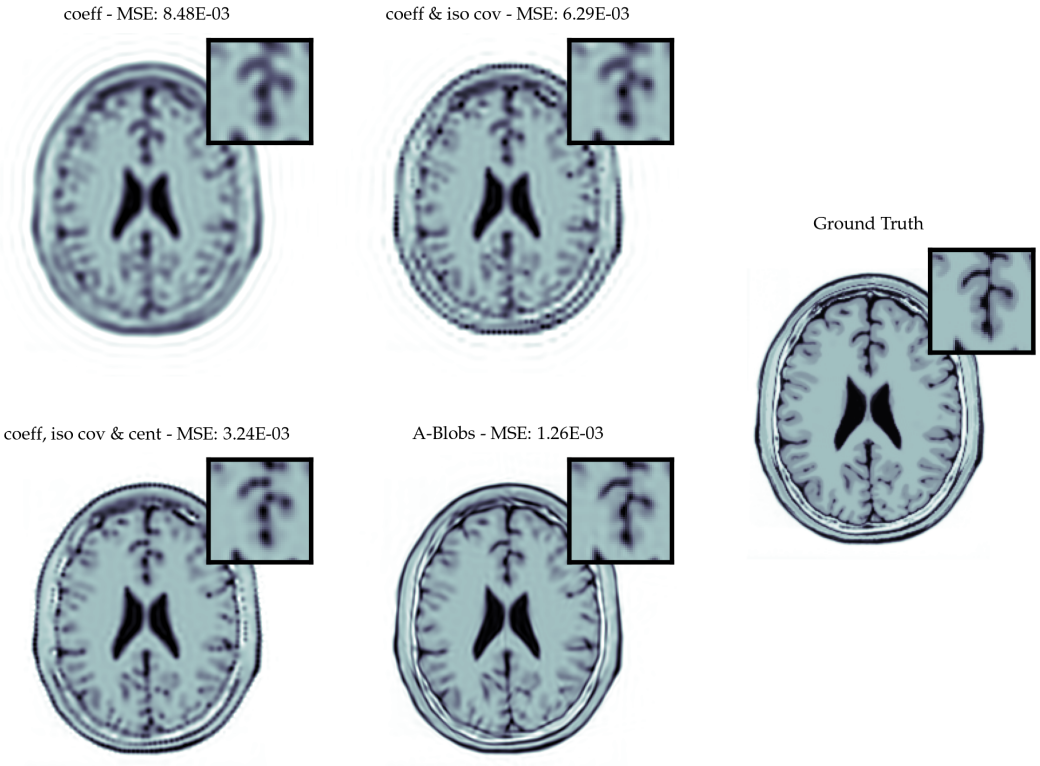


Figure 6.2: Images of changing the parameterisation, while keeping number of A-Blobs the same $N_b = 4096$. Mean Squared Error (MSE) is given for each parameterisation.

can be formulated as:

$$\mathcal{F}[b](\mathbf{k}; \mathbf{c}, \boldsymbol{\mu}, \mathbf{U}) = c(2\pi)^{\frac{N_{dim}}{2}} |\det \mathbf{U}| \exp\left(-\mathbf{k}^\top \boldsymbol{\mu} \mathrm{i} - \frac{1}{2}\|\mathbf{U}\mathbf{k}\|_2^2\right). \tag{6.3}$$

The Fourier A-Blobs $\mathcal{F}[x]$ is the summation of N_b analytical Fourier blobs each with their own parameters:

$$\mathcal{F}[x](\mathbf{k}; \mathbf{c}, \boldsymbol{\Gamma}) = \sum_{b=1}^{N_b} \mathcal{F}[b](\mathbf{k}; c_{(b)}, \boldsymbol{\mu}_{(b)}, \mathbf{U}_{(b)}) \tag{6.4}$$

$$= \sum_{b=1}^{N_b} c_{(b)}(2\pi)^{\frac{N_{dim}}{2}} |\det \mathbf{U}_{(b)}| \exp\left(-\mathbf{k}^\top \boldsymbol{\mu}_{(b)} \mathrm{i} - \frac{1}{2}\|\mathbf{U}_{(b)}\mathbf{k}\|_2^2\right). \tag{6.5}$$

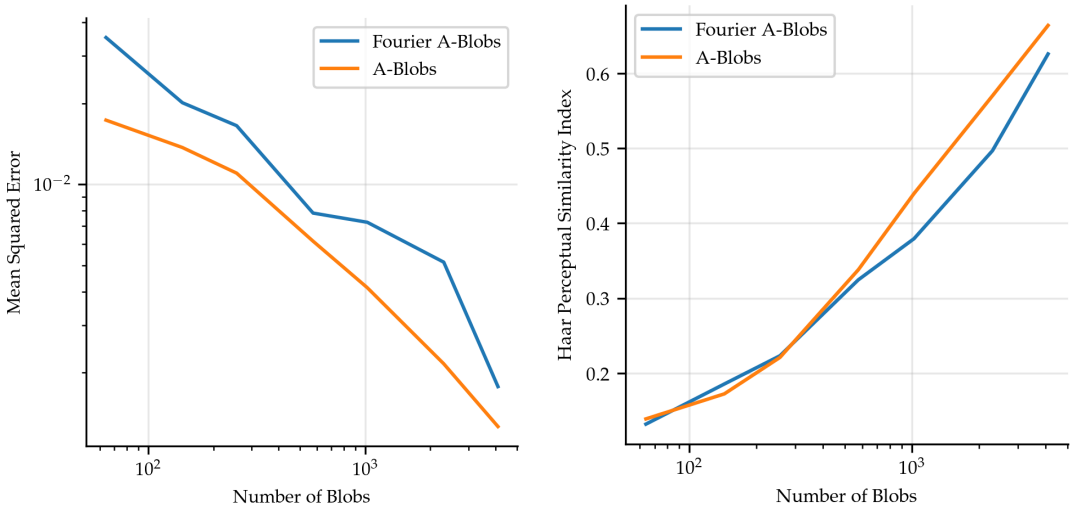
Given this analytical representation of Fourier A-Blobs, the Fourier-transformed BrainWeb $\mathcal{F}[\mathbf{x}_{\mathrm{GT}}]$ can be used to fit parameters $\mathbf{c}$ and $\boldsymbol{\Gamma}$. The set of k-space, or Fourier, coordinates are denoted $\Omega_{\mathbf{k}} = \{(\mathbf{k}_{(s)})\}_{s=1}^{N_{\mathbf{k}}}$ and correspond to full grid of 256×256 k-space coordinates. Initialised with the previously described isotropic grid of A-Blobs, the Fourier A-Blobs are optimised with the objective function:

$$\mathbf{c}^\dagger, \boldsymbol{\Gamma}^\dagger = \underset{\mathbf{c},\boldsymbol{\Gamma}}{\operatorname{argmin}} \|\mathcal{F}[\mathbf{x}_{gt}] - \mathcal{F}[x](\mathbf{k}; \mathbf{c}, \boldsymbol{\Gamma})|_{\Omega_{\mathbf{k}}}\|_2^2. \tag{6.6}$$

We compare the reconstruction accuracy of parameters obtained from Eqn. (6.6) by evaluating A-Blobs on $\Omega_{\mathbf{r}}$. Thus the quality assessment methods MSE and HaarPSI are compute between $x(\mathbf{r}; \mathbf{c}, \boldsymbol{\Gamma})|_{\Omega_{\mathbf{r}}}$ and $\mathbf{x}_{\mathrm{GT}}$.

Discussion and Important Observations: Quantitatively from Fig. 6.1 both MSE and HaarPSI improve with increased number of A-Blobs and increased amount of parameterisation. The ability of the increasingly parameterised A-Blobs can be qualitatively observed in Fig. 6.2, where isotopic covariances lead to distinctly blob-like artifacts. The full A-Blobs parameterisation gives an image where A-Blobs are skewed with important features of the underlying anatomy.

It is observed that there is good agreement between Fourier A-Blobs and A-Blobs representation as is seen in Fig. 6.3. Although, for Fourier A-Blobs the optimisation Eqn. (6.6) minimises discrepancy in k-space, this results in in-



(**a**) Image quality assessed with Mean Squared Error.

(**b**) Image quality assessed with Haar Wavelet-Based Perceptual Similarity Index.

Figure 6.3: Plots assessing image representation of A-Blobs with varying numbers of A-Blobs and parameters obtained through optimisation with spatial or frequency measurements, A-Blobs or Fourier A-Blobs respectively.

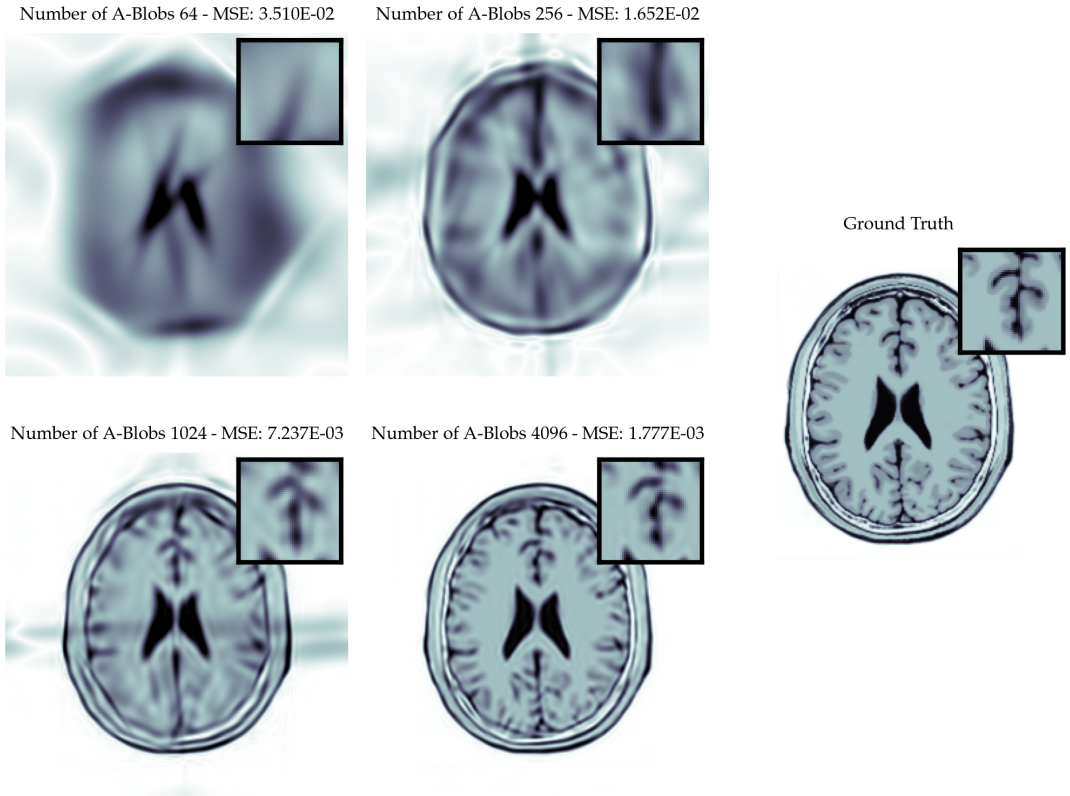


Figure 6.4: Images of Fourier A-Blobs with varying numbers of A-Blobs. Mean Squared Error (MSE) is given for each number of A-Blobs.

creased error when computing MSE in the spatial domain. The images given in Fig. 6.4, show the Fourier A-Blobs images for varying numbers of A-Blobs. Ar-

tifacts are seen with 64, 256 and 1024 A-Blobs with a small number large overly skewed A-Blobs visible. This could be due to two factors; $\mathcal{F}[\mathbf{x}_{\text{GT}}]$ is computed with an FFT that assumes a periodic boundary condition this is not assumed for analytical Fourier A-Blobs, and the optimisation procedure was not tuned further from spatial representation leading to poor local minima. Further, by virtue of fitting to Fourier data with a finite number of A-Blobs leads to an approximation error that may be small in frequency domain but large in the spatial domain. In other-words, regularisation, see Section 2.1.1, may be needed to encourage approximations that adhere to specific spatial assumptions, potentially removing the large overly skewed A-Blobs.

The optimisation and initialisation of A-Blobs requires careful attention. It was observed that the non-linear optimisation is prone to stalling in poorly performing minima, and incorrect initialisation (i.e. placing Gaussian's centroids away from image structure) leads to poor image representation. Unlike INRs, a more principled initialisation is attainable by initialising as a uniform grid. This is what is done in [208], and further described as the "Practical Algorithm" in [176].

For this experiment and all subsequent experiments the Adam optimisation algorithm is used [124]. We observed that the adaptive nature of Adam, that rescales parameter-wise step-sizes based on running averages, allowed for stable optimisation. For these reasons Adam is also what is commonly used for INRs. The parameterisation with "coeff" is a linear and convex optimisation, which is a linear least-squares problem, and the full A-Blobs parameterisation is a non-linear least-squares problem. The formulation of A-Blobs is more straightforward than INRs meaning that more sophisticated non-linear least-squares optimisation algorithms could be used. An example would be Levenberg-Marquardt [154], in fact this was recently applied for Gaussian splatting [100]. Further to this, the objective functions considered are separable non-linear least-squares problems as they are comprised of a linear combination of non-linear functions. For this class of problem more specialised algorithms such as VARPRO could be leveraged [73]. However, the application of these advanced optimisation algorithms is left for further work.

In the next section we introduce MR-Blob, the formulation of A-Blobs for parallel MRI reconstruction.

6.4 MR-Blob Reconstruction

In parallel MRI the measurements are acquired simultaneously across multiple receiver coils. As described in Section 2.2.2.2, these coils have a sensitivity that further spatially encode the measurements. This enables accelerated acquisition with fewer measurements wile maintaining high quality reconstructions, as the multiple coil sensitivities can be leveraged in advanced reconstruction algorithms. The coil sensitivities are not known *a-priori* and the estimation of sensitivities and reconstruction of the image can be cast as a non-linear inverse problem [240]. Other established methods linearise the problem by pre-computing estimates of the coils or by calibrating for the affect of the coils, GRAPPA, SENSE and ESPIRiT [80, 179, 241].

In this work we explicitly model the coil sensitivities as single isotropic Gaussian functions, and we leverage the specific properties of Gaussians to develop an analytic forward model with A-Blobs approximating the image. A coil sensitivity is modelled as:

$$b_{(\kappa)}(\mathbf{r}; \boldsymbol{\mu}_{(\kappa)}, s_{(\kappa)}) = \exp\left(-\frac{s_{(\kappa)}^2 \mathbf{I} \|(\mathbf{r} - \boldsymbol{\mu}_{(\kappa)})\|_2^2}{2}\right).$$

with κ denoting the κ-th coil with isotropic scale $s_{(\kappa)}$ and centroid $\boldsymbol{\mu}_{(\kappa)}$. The coil sensitivity acts such that signal closer to the centre is stronger and decays in space exponentially. A coil image, the image obtained from a coil, is modelled as the product of Gaussians [174]:

$$\sum_{b=1}^{N_b} \underbrace{b_{(\kappa)}(\mathbf{r}; \boldsymbol{\mu}_{(\kappa)}, s_{(\kappa)})}_{\text{Coil sensitivity}} \cdot \underbrace{x(\mathbf{r}; \mathbf{c}_{(b)}, \boldsymbol{\mu}_{(b)} \mathbf{L}_{(b)})}_{\text{A-Blobs image}} = \sum_{b=1}^{N_b} a_{(\kappa,b)} \exp\left(-\frac{(\mathbf{r} - \boldsymbol{\mu}_{(\kappa,b)})^\top \boldsymbol{\Sigma}_{(\kappa,b)}^{-1} (\mathbf{r} - \boldsymbol{\mu}_{(\kappa,b)})}{2}\right),$$

where

$$
\begin{aligned}
\boldsymbol{\Sigma}_{(\kappa,b)}^{-1} &= (\mathbf{L}_{(b)}^{\top}\mathbf{L}_{(b)} + s_{(\kappa)}^{2}\mathbf{I}),\\
\boldsymbol{\mu}_{(\kappa,b)} &= (\mathbf{L}_{(b)}^{\top}\mathbf{L}_{(b)} + s_{(\kappa)}^{2}\mathbf{I})^{-1}(\mathbf{L}_{(b)}^{\top}\mathbf{L}_{(b)}\boldsymbol{\mu}_{(b)} + s_{(\kappa)}^{2}\mathbf{I}\boldsymbol{\mu}_{(\kappa)})\\
a_{(\kappa,b)} &= c_{(b)}\exp\left(-\frac{(\boldsymbol{\mu}_{(b)} - \boldsymbol{\mu}_{(\kappa)})^{\top}(\mathbf{U}_{(b)}^{\top}\mathbf{U}_{(b)} + s_{(\kappa)}^{-2}\mathbf{I})^{-1}(\boldsymbol{\mu}_{(b)} - \boldsymbol{\mu}_{(\kappa)})}{2}\right).
\end{aligned}
$$

MRI measurements acquired by a coil are in the frequency (k-space) domain. These can be modelled in k-space coordinates $\mathbf{k} = [k_x,\ k_y]^{\top}$ by Fourier transforming $b_{(\kappa)}(\mathbf{r}) \cdot x(\mathbf{r})$:

$$
\begin{aligned}
&\mathcal{F}\{b_{\kappa}\cdot x\}(\mathbf{k}; s_{\kappa}, \boldsymbol{\mu}_{\kappa}, \mathbf{c}, \boldsymbol{\Gamma}) =\\
&\qquad \sum_{b=1}^{N_b}(2\pi)^{\frac{N_{dim}}{2}}\sqrt{\det(\boldsymbol{\Sigma}_{(\kappa,b)})}a_{(\kappa,b)}\exp\left(-i\mathbf{k}^{\top}\boldsymbol{\mu}_{(\kappa,b)} - \frac{1}{2}\mathbf{k}^{\top}\boldsymbol{\Sigma}_{(\kappa,b)}\mathbf{k}\right),
\end{aligned}
$$

where $\mathbf{c} = \{(c_{(b)})\}_{b=1}^{N_b}$, $\boldsymbol{\Gamma} = \{(\boldsymbol{\mu}_{(b)}, \mathbf{L}_{(b)})\}_{b=1}^{N_b}$ and we define $\boldsymbol{\Gamma}_{\kappa} = \{(\boldsymbol{\mu}_{(\kappa)}, s_{(\kappa)})\}_{\kappa=1}^{N_{\kappa}}$. The MR-blob is defined:

$$
\text{MR-blob}(\mathbf{k}; \mathbf{c}, \boldsymbol{\Gamma}, \boldsymbol{\Gamma}_{\kappa}) := \begin{pmatrix} \mathcal{F}(b_{(1)} \cdot x)(\mathbf{k}; s_{(1)}, \boldsymbol{\mu}_{(1)}, \mathbf{c}, \boldsymbol{\Gamma}) \\ \vdots \\ \mathcal{F}\{b_{(N_{\kappa})} \cdot x\}(\mathbf{k}; s_{(N_{\kappa})}, \boldsymbol{\mu}_{(N_{\kappa})}, \mathbf{c}, \boldsymbol{\Gamma}) \end{pmatrix}.
$$

The same parameters $\mathbf{c}$ and $\boldsymbol{\Gamma}$ are used across all coils. The inverse problem is seeking the solution to:

$$
\text{MR-blob}(\mathbf{k}; \mathbf{c}, \boldsymbol{\Gamma}, \boldsymbol{\Gamma}_{\kappa})|_{\Omega_{\mathbf{k}}} = \mathbf{y},
$$

where parallel MRI measurements are $\mathbf{y}$ with $N_{\kappa} \in \mathbb{N}$ coils, the measurement are defined on $\Omega_k = \{(\mathbf{k}_{(i)})\}_{i=1}^{N_{\mathbf{k}}}$ with $N_{\mathbf{k}}$ sampled coordinates.

In the variational framework of inverse problems[202], see Section 2.1.1, we recover the parameters of the reconstruction through optimising:

$$
\mathbf{c}^{\dagger}, \boldsymbol{\Gamma}^{\dagger}, {\boldsymbol{\Gamma}_{\kappa}}^{\dagger} \in \min_{\mathbf{c},\boldsymbol{\Gamma},\boldsymbol{\Gamma}_{\kappa}} \left\{\frac{1}{N_{\mathbf{k}}}\|\text{MR-blob}(\mathbf{k}; \mathbf{c}, \boldsymbol{\Gamma}, \boldsymbol{\Gamma}_{\kappa})|_{\Omega_{\mathbf{k}}} - \mathbf{y}\|_2^2 + \lambda\eta(x(\mathbf{r}; \mathbf{c}, \boldsymbol{\Gamma}))\right\}, \qquad (6.7)
$$

where the first term promotes data-consistency, and the second is a regulariser with strength λ. We directly optimise over blob parameters $\mathbf{c}, \mathbf{\Gamma}$, as well as over the coil sensitivities $\mathbf{\Gamma}_\kappa$. Thus, we jointly reconstruct the image and optimise the coil sensitivities. The reconstruction is given by:

$$x(\mathbf{r}; \mathbf{c}^\dagger, \mathbf{\Gamma}^\dagger)|_{\Omega_\mathbf{r}},$$

where the spatial sample coordinates $\Omega_\mathbf{r}$ are determined from the image as a regular grid corresponding to pixel centroids.

We consider two regularisers. The first penalises the skewness of the A-Blobs:

$$\eta_{\text{cond}}(\mathbf{\Gamma}) := \frac{1}{N_b} \sum_{b=1}^{N_b} \text{cond}(\mathbf{L}_{(b)}^\top \mathbf{L}_{(b)}) - 1, \tag{6.8}$$

by penalising the skewness, we penalise overly elongated A-Blobs and encourage a shape closer to isotropic. The condition number for a blob is computed in 2D as: $\text{cond}(\mathbf{L}_{(b)}^\top \mathbf{L}_{(b)}) = \frac{\max([\mathbf{L}_{(b)}^2]_{11}, [\mathbf{L}_{(b)}^2]_{22})}{\min([\mathbf{L}_{(b)}^2]_{11}, [\mathbf{L}_{(b)}^2]_{22})}$.

The other regulariser considered is more akin to traditional regularisation, and is continuous TV:

$$\eta_{\text{TV}}(\mathbf{r}; \mathbf{c}, \mathbf{\Gamma}) := \frac{1}{N_r} = \|\nabla_\mathbf{r} x(\mathbf{r}; \mathbf{c}, \mathbf{\Gamma})|_{\Omega_\mathbf{r}}\|_1,$$

where the spatial sampling is the same that is used for the image.

6.5 Experimental Methods

Three methods were compared to illustrate the features of reconstruction with pixels (M1), A-Blobs (M2) and MR-blobs (M3). The NUFFT with coil sensitivities forward model as $\mathbf{A} : \mathbb{C}^{N_\mathbf{r}} \to \mathbb{C}^{N_\mathbf{k} \times N_\kappa}$. The corresponding objective functions are given by

Algorithm 4 MR-blob reconstruction: the ADAM optimisation algorithm is used to update the parameters [124]. All operations are element-wise.

Require: Measurements $\mathbf{y}$
Require: MR-blob parameters $\mathbf{c}, \mathbf{\Gamma}, \mathbf{\Gamma}_\kappa$
Require: Number of iterations N_i
Require: Stepsize α
Require: Exponential decay rates for the moment estimates β_1, β_2
Require: Objective function $\Phi(\Omega; \mathbf{c}, \mathbf{\Gamma}, \mathbf{\Gamma}_\kappa)$
Require: Initial parameters $\mathbf{c}^{(0)}, \mathbf{\Gamma}^{(0)}, \mathbf{\Gamma}_\kappa^{(0)}$
Require: Initialise 1^{st} and 2^{nd} moment vectors $\mathbf{m}^{(0)}, \mathbf{v}^{(0)}$ respectively.
for $i = 0, \ldots, N_i - 1$ **do**
$\mathbf{g}^{(i)} \leftarrow \nabla_{\mathbf{c}^{(i)}, \mathbf{\Gamma}^{(i)}, \mathbf{\Gamma}_\kappa^{(i)}} \Phi(\Omega; \mathbf{c}^{(i)}, \mathbf{\Gamma}^{(i)}, \mathbf{\Gamma}_\kappa^{(i)})$ ▷ Get gradient using AutoDiff
$\mathbf{m}^{(i+1)} \leftarrow \beta_1 \cdot \mathbf{m}^{(i)} + (1 - \beta_1) \cdot \mathbf{g}^{(i)}$ ▷ Update biased first moment estimate
$\mathbf{v}^{(i+1)} \leftarrow \beta_2 \cdot \mathbf{v}^{(i)} + (1 - \beta_2) \cdot \left(\mathbf{g}^{(i)}\right)^2$ ▷ Update biased second moment estimate
$\hat{\mathbf{m}}^{(i+1)} \leftarrow \mathbf{m}^{(i+1)} / (1 - \beta_1^i)$ ▷ Bias-corrected first moment estimate
$\hat{\mathbf{v}}^{(i+1)} \leftarrow \mathbf{v}^{(i+1)} / (1 - \beta_2^i)$ ▷ Bias-corrected second moment estimate
$\{\mathbf{c}^{(i+1)}, \mathbf{\Gamma}^{(i+1)}, \mathbf{\Gamma}_\kappa^{(i+1)}\} \leftarrow \{\mathbf{c}^{(i)}, \mathbf{\Gamma}^{(i)}, \mathbf{\Gamma}_\kappa^{(i)}\} - \alpha \cdot \hat{\mathbf{m}}^{(i+1)} / (\sqrt{\hat{\mathbf{v}}^{(i+1)}} + \epsilon)$
end for
$\mathbf{c}^{(N)}, \mathbf{\Gamma}^{(N)}, \mathbf{\Gamma}_\kappa^{(N)}$

$$\text{M1 (Pixels):} \min_{\mathbf{f}} \left\{ \frac{1}{N_\mathbf{k}} \|\mathbf{A}\mathbf{f} - \mathbf{y}\|_2^2 + \frac{\lambda}{N_\mathbf{r}} \|\nabla \mathbf{f}\|_1 \right\}$$

$$\text{M2 (Blobs):} \min_{\mathbf{c}, \mathbf{\Gamma}} \left\{ \frac{1}{N_\mathbf{k}} \|\mathbf{A}x(\mathbf{r}; \mathbf{c}, \mathbf{\Gamma})|_{\Omega_\mathbf{r}} - \mathbf{y}\|_2^2 + 10 \cdot \eta_{\text{cond}}(\mathbf{\Gamma}) + \lambda \eta_{\text{TV}}(\mathbf{r}; \mathbf{c}, \mathbf{\Gamma}) \right\}$$

$$\text{M3 (MR-blob):} \min_{\mathbf{\Gamma}_\kappa, \mathbf{c}, \mathbf{\Gamma}} \left\{ \frac{1}{N_\mathbf{k}} \|\text{MR-blob}(\mathbf{k}; \mathbf{c}, \mathbf{\Gamma}, \mathbf{\Gamma}_\kappa)|_{\Omega_\mathbf{k}} - \mathbf{y}\|^2 + 10 \cdot \eta_{\text{cond}}(\mathbf{\Gamma}) + \lambda \eta_{\text{TV}}(\mathbf{r}; \mathbf{c}, \mathbf{\Gamma}) \right\},$$

TV regularisation with the same λ was tested for all three methods, see Section 2.2.3.2 for details on TV. For M2 and M3 we include a constant skew-penalisation and evaluate TV at points corresponding to pixel centres.

For evaluation a 128×128 Shepp-Logan phantom was used. Measurements were simulated via NUFFT with 25 radial spokes and a golden angle spoke offset; this resulted in an affective 9.7× acceleration. The measurements were degraded with complex white Gaussian noise to an SNR of 20dB. The coil sen-

sitivities were included in the simulation and modelled as a Gaussian function as implemented in `torchkbnufft` [156].

For IQA, PSNR, SSIM, and CRC were computed. CRC was evaluated on the small lower three ellipsoids of the Shepp-Logan. Additionally, data-consistency was computed for the reconstructions, giving an indication of how closely the forward projected reconstruction agrees with measurements.

The gradients of 6.7 are obtain using automatic (reverse-mode) differentiation and the Adam optimiser is used [124]. The initialisation was a 30×30 equispaced grid of isotropic blobs.

For *in-vivo* MR-Blob we evaluated on a single 16-coil *in-vivo* sample from the FastMRI dataset [262]. As noted in [240], the non-linear estimation of sensitivities and reconstruction of the image is an undetermined problem with a free parameter. This free parameter g acts by giving the infinite set of solutions $\{\boldsymbol{\kappa} \cdot g, \mathbf{x}/g\}$ with $\boldsymbol{\kappa}$ the coil sensitivity. In our formulation this is restricted by using unnormalised isotropic Gaussians such that $\max(|b_{(\kappa)}|) = 1$. To evaluate the performance of the coil sensitivity estimation we compare against ESPIRiT estimated coil sensitivities [241], as is shown in Fig. 6.7. We use fully sampled k-space and the resulting reconstruction compared to root-sum-of-squares image. "The root-sum-of-squares is an absolute sum of squares of the coil images, which are then square-rooted."

6.6 Results and Discussion

From Fig. 6.5, there is good quantitative agreement between all three methods, this is particularly true at higher regularisation values. At lower regularisation values the deviation is most prominent in Data-Consistency. Both A-Blobs-based methods (M2 & M3) show a larger deviation. This could be due to the high noise regime of the measurements and the difficulty of representing non-smooth noise with smooth functions; the reconstruction is restricted to a subspace that does not approximate noise well.

The number of parameters for A-Blobs methods were 2.6× less parameters than pixels. Additionally, the piece-wise-constant structure of the Shepp-Logan

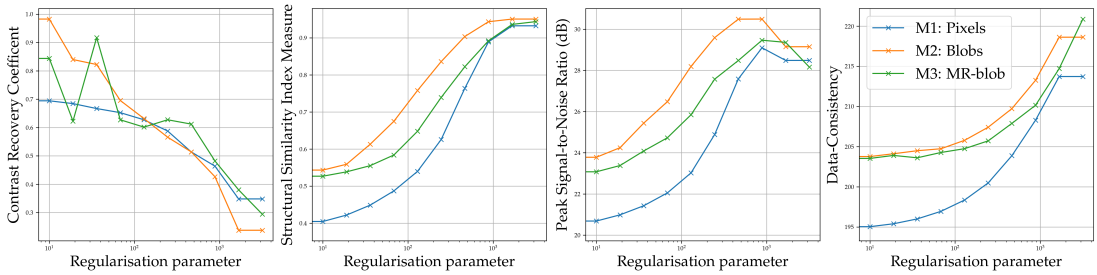


Figure 6.5: CRC, SSIM, PSNR and data-consistency for all three methods swept over the same range of regularisation parameters ($\log_{10}(\lambda) \in \{1.00, 1.28, 1.56, 1.83, 2.11, 2.39, 2.67, 2.94, 3.22, 3.50\}$).

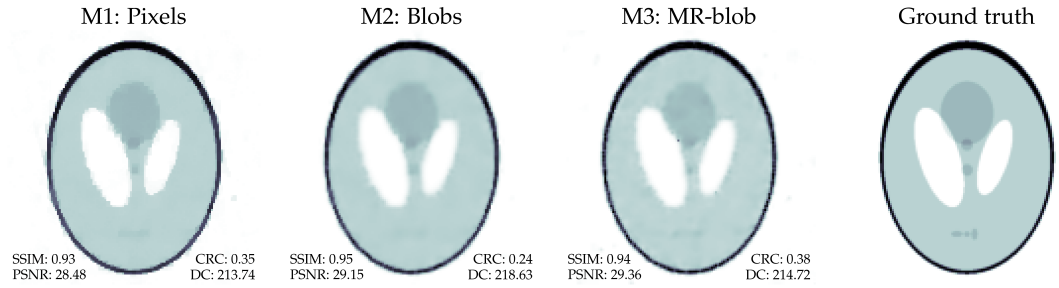


Figure 6.6: Well-regularised reconstructions ($\lambda = 1,668$) from noisy, under-sampled measurements. Note the instability exhibited in MR-blob reconstruction.

phantom is not well approximated by smooth basis functions such as blobs. In spite of this in Fig. 6.6 the reconstruction quality is visually and quantitatively competitive.

The optimisation of MR-blob (M3) includes the parameters of the coil-sensitivities that are initially centred in the image. The optimisation allows the coils to translate and re-scale to fit the measurements. This resulted in the coil sensitivities and coil images shown in Fig. 6.7. Here it can be seen that the MR-Blob coils are able to approximate the ground truth coils.

The M2 and M3 reconstructions exhibit basis functions with prominent structure, see Fig. 6.8. It is noted that M3 gave fewer zero contrasts, less skewed blobs and a larger range of scales compared to M2.

The same forward model used for simulation is used for reconstruction for M1 and M2, this means that for these methods an inverse crime was committed [112]. This means that the same forward model used for simulation is used to solve the inverse problem, potentially giving over-performing results. This makes the competitive performance of M3 more impressive.

With respect to *in-vivo* MR-Blob, in Fig. 6.9 we observe that the reconstruc-

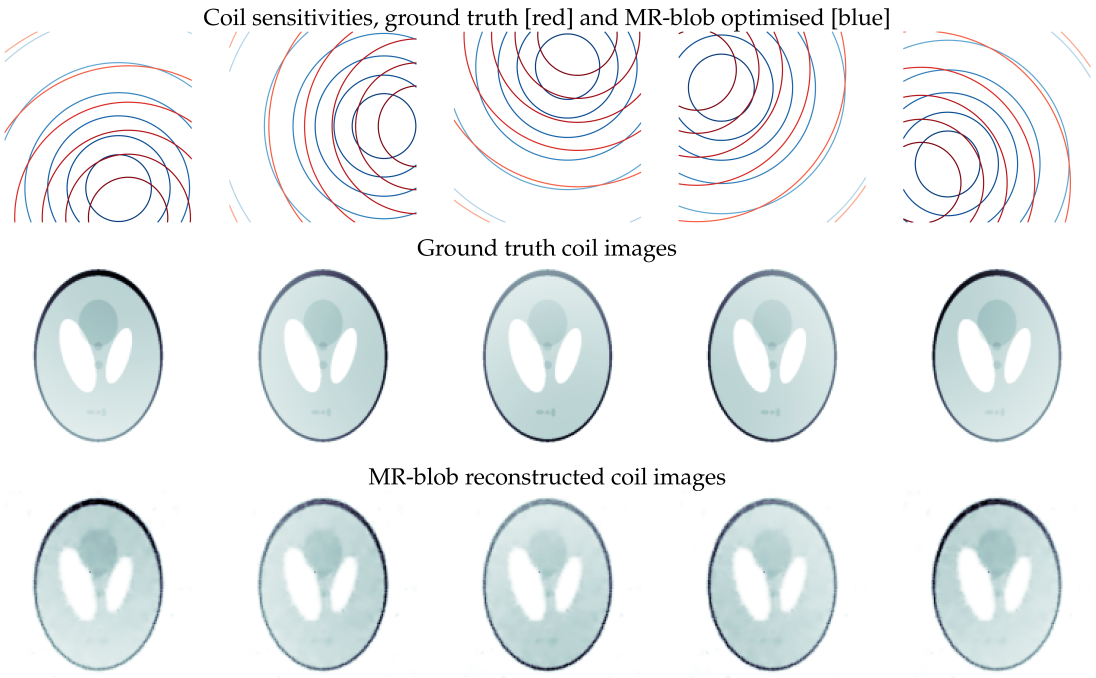


Figure 6.7: Top row: Ground truth (red) and MR-blob optimised (blue) coil sensitivities, MR-blob coils were initialised at the centre at set scale. Middle row: Ground truth coil images. Bottom row: Well-regularised ($\lambda = 1,668$) MR-blob reconstructed coil images. Columns correspond to individual coils.

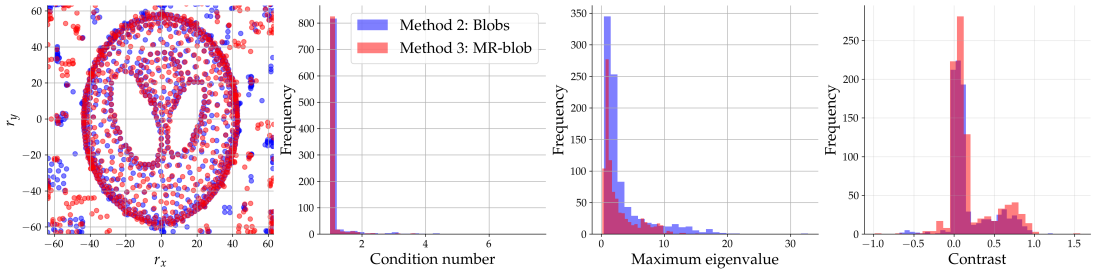


Figure 6.8: Coordinate transformed blob parameters of well-regularised ($\lambda = 1,668$) M2 and M3 reconstructions. Left: centroid of A-Blobs. Centre left: Condition number of covariance (skewness). Centre right: Maximum eigenvalue (scale) of covariance. Right: Contrast of A-Blobs.

tion is visually similar but there is a discrepancy in the scale of the image. This can be attributed to the free-parameter discussed earlier. Further work is needed to ensure robust scaling.

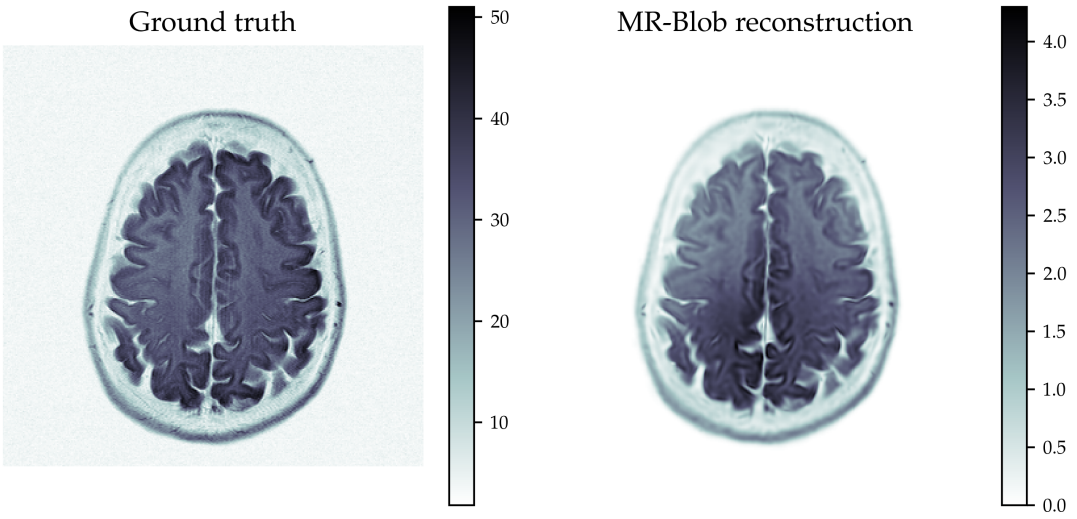


Figure 6.9: MR-Blob reconstruction qualitative comparison against root-sum-of-squares reconstruction.

From Fig. 6.10 we observe that isotropic Gaussians give a coarse approximation to the coil sensitivities obtained from ESPIRiT. The coil sensitivities were initialised at the centre of the image and the optimised sensitivities are shown. For coils 1 and 16 the approximation is closer to that observed with ESPIRiT and for coil 9 there is a larger discrepancy.

As there were 16 coils and fully sampled 320×640 k-space data, the computational cost for updates was prohibitively large. As such the parameters were updated based on subsets of the measurements in a block-sequential fashion with an Adam optimiser. In spite of this the results are promising.

6.7 Conclusion

Over the last few years has been a renewed interest in alternatives to voxels [137, 178, 199, 221]. In this chapter we proposed A-Blobs as an alternate, and forward model directly. It was shown that by increasing the parameterisation of A-Blobs we enable the blobs to adapt more flexibly to the structure of the underlying function; reducing the approximation error of image and Fourier

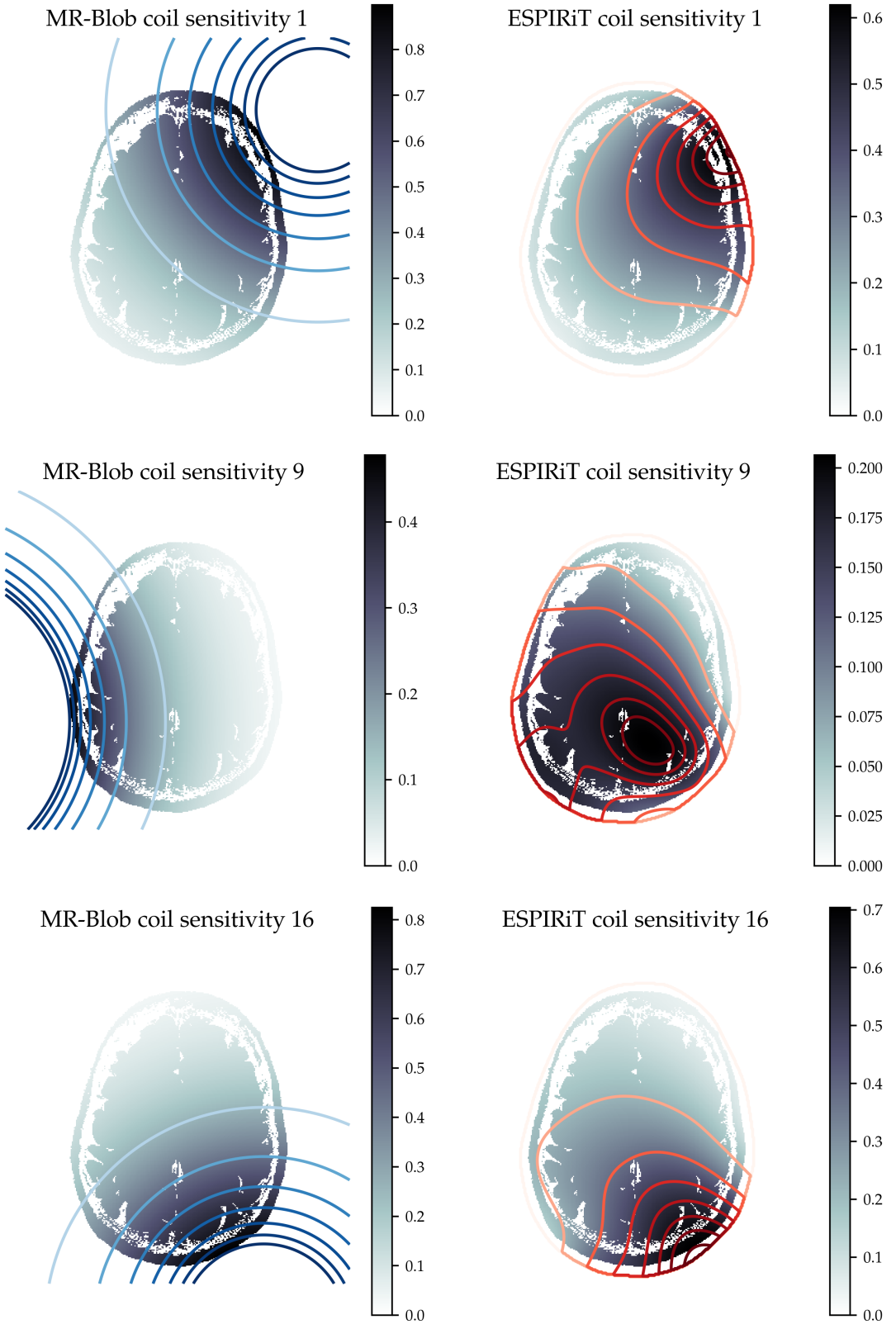


Figure 6.10: Coil sensitivities from coils 1, 9 and 16 obtained from MR-Blob and ESPIRiT. An image of a masked-magnitude coil sensitivity is overlayed with iso-contours of the coil sensitivity. Separate colour bars are given, and MR-Blob coils were initialised at the centre with equal scale.

image representation significantly.

Further, it is known that Gaussian mixtures, that A-Blobs can be interpreted as, are able to approximate densities to arbitrary precision given N_b [175]. With this insight the decrease of approximation error with N_b is expected and observed in experiments. This indicates that arbitrary precision can be attained in the case of A-Blobs, although rigorous analysis is required to prove this.

For MR-Blob, the modelling of coil sensitivities as a single isotropic Gaussian could be overly restrictive and potentially increase model error see Section 2.1. We give results for both *in-silico* and *in-vivo* reconstruction that show competitive performance with voxelised representations, but more rigorous analysis is required to validate the use of isotropic coils. It is worth noting that modelling as isotropic Gaussians we effectively constrain the coil sensitivities to a subspace, and more established approaches can be observed to constrain to a subspace too. albeit less restrictive [80, 179, 241].

The optimisation of A-Blobs utilised Adam. With the tractable formulation of A-Blobs it would be of interest to utilise advanced algorithms for non-linear objective functions [100]. The reconstruction could benefit from initialisation based on an approximate reconstruction, and development of continuous regularisation terms [178]. Additional further work is described in Section 7.2.

Chapter 7

Conclusions

This thesis investigates advanced methodologies in medical image reconstruction, focusing on PET and MRI imaging. The overarching motivation was to incorporate known physics through data-consistency at test time, enabling more robust and generalisable image reconstruction methods. The studies presented span a range of innovative approaches, each addressing specific challenges and contributing significantly to the field.

The first study developed DIP for PET reconstruction, a method that requires no training data, and has been shown to be effective for PET reconstruction in prior studies [74]. This study validated these results as well as developing it further by stabilising reconstruction in a lower-count measurements regime. By introducing established regularisation into the objective function, the need for robust stopping criteria was alleviated, and DIP+RDP outperformed established methods on ability to recover lesions. An important aspect of this study was the integration of a PET forward model into a deep learning framework, for this SIRF, STIR and `PyTorch` software were leveraged.

The next study considered supervised methods for PET reconstruction. There we accounted for the intricacies of PET, specifically the widely varying dynamic range. We developed normalisation approaches for two supervised approaches, post-processing and LPD, and demonstrated improvements in CRC and data-consistency.

The previously developed insights informed our next work on the application of SGMs for PET image reconstruction. Specifically, the integrating PET

forward models into deep learning frameworks, and accounting for the widely varying dynamic range of PET with normalisation. In the study we developed sampling methods to deal with the non-negativity inherent to PET. We also proposed MR guidance and scaling to 3D. The evaluation showed that the proposed methods were competitive with supervised methods, even outperforming them on data further outside of the training distribution. This further demonstrates the generalisability enabled through data-consistency. With guided reconstruction we observed that hallucinations were suppressed, although lesions were also affected. The scaling to 3D improved reconstruction quality as compared to comparison methods, and was accelerated only require 3.4 minutes per volume.

The final study focused on a functional approach using A-Blobs. By parameterising images as continuous functions, this method offers a promising alternative to voxel-based representations. A forward model for parallel MRI reconstruction was developed and allowed for direct formulation of A-Blobs in k-space. Initial results indicated competitive reconstruction quality, but further work is needed to refine the approach, address modelling limitations, and validate its applicability in clinical settings.

With these concluding remarks it would be remiss to not include recent trends in acquisition for PET and MRI. This is because advanced medical image reconstruction is constrained by the acquisition process.

7.1 Trends in Advanced Acquisition

For PET a key development neglected in this thesis has been the development of increase Time of Flight (ToF) resolution, where ToF allows for localisation of the annihilation event along an LoR [133]. With a fine enough resolution it has been noted that one could obtain "reconstruction-free" Positron Emission Imaging [206].[1] Furthermore, detectors are growing smaller and the length of scanners (i.e. Total-Body PET) is growing longer meaning that most LoRs detect one or zero photons, resulting in extremely sparse sinograms [160, 259].

[1] With fine enough ToF resolution the modality ceases to be tomographic and measurements can be viewed as a point cloud.

This motivates the use of list-mode data [12], and an important challenge is how to deal with the highly heterogeneous sensitivities of the scanner [25].

From an MRI perspective, there has been an interest in low field MRI, which uses $B_0 \approx 1 - 199$ mT rather than ≥ 1.5 T used in clinical practice [201].[2] The problems encountered in this regime are low signal-to-noise ratio of measurements (driven by the inherent supra-linear relationship with B_0 [101]), and strong B_0 field inhomogeneities, requiring sophisticated treatment for image reconstruction [204]. On the other hand, ultra high field MRI with $B_0 > 7$ T are being developed for clinical use [256]. The challenges here are the increased RF field inhomogeneity and power deposition in tissue [130], which can be somewhat mitigated by parallel RF transmission [168]. However, there still exists many technical challenges [256].

The advanced acquisition described above provide interesting challenges with respect to medical image reconstruction. In the next section we will present future work that is in-line with the studies presented in this thesis.

7.2 Future Work

This section highlights preliminary, unpublished results as part of future work. Further research directions for each developed method are detailed in their respective chapters.

SGM-based Unsupervised Anomaly Localisation Training the score-model exclusively on healthy patient data could enable the detection of pathological abnormalities [20], such as lesions, by observing increased posterior uncertainty in affected regions. This approach could serve as a valuable tool for anomaly localisation. An illustrative example of this concept is provided in Appendix A, demonstrating its potential in guided reconstruction scenarios.

3D A-Blobs and Volumetric Regularisation Given the simple analytical and parallelisable formulation of A-Blobs, we implement CUDA-kernels for 3D A-Blobs by analytically computing the value and gradient, as is done with Gaus-

[2] Ultra low field MRI exist that have even weaker magnetic strength.

sian Splatting [121]. Furthermore, we investigate the use of volumetric regularisation based on sparsity to remove A-Blobs that do not contribute sufficiently to the representation. This is shown in Appendix B.

Bibliography

[1] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[2] J. Adler *et al.*, *Odlgroup/odl: Odl 0.7.0*, version v0.7.0, Oct. 2018.

[3] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[4] S. Ahn and J. A. Fessler, "A globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 623–626, 2003.

[5] B. D. O. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982.

[6] S. R. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[7] S. R. Arridge, M. J. Ehrhardt, and K. Thielemans, "(An overview of) synergistic reconstruction for multimodality/multichannel imaging methods," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2200, p. 20 200 205, 2021.

[8] B. Aubert-Broche, M. Griffin, G. B. Pike, A. C. Evans, and D. L. Collins, "Twenty new digital brain phantoms for creation of validation image data bases," *IEEE Transactions on Medical Imaging*, vol. 25, no. 11, pp. 1410–1416, 2006.

[9] D. O. Baguer, J. Leuschner, and M. Schmidt, "Computed tomography reconstruction using deep image prior and learned reconstruction methods," *Inverse Problems*, vol. 36, no. 094004, 24pp, 2020.

[10] B. Bai, Q. Li, and R. M. Leahy, "Magnetic resonance-guided positron emission tomography image reconstruction," in *Seminars in Nuclear Medicine*, Elsevier, vol. 43, 2013, pp. 30–44.

[11] B. Bai, A. Ruangma, R. Laforest, Y.-C. Tai, and R. M. Leahy, "Positron range modeling for statistical pet image reconstruction," in *Proceedings of the IEEE Conference NSS/MIC/RTSD*, IEEE, vol. 4, 2003, pp. 2501–2505.

[12] D. L. Bailey, M. N. Maisey, D. W. Townsend, and P. E. Valk, *Positron emission tomography*. Springer, 2005, vol. 2.

[13] R. Barbano, J. Antoran, J. Leuschner, J. M. Hernández-Lobato, B. Jin, and. Kereta, "Image reconstruction via deep image prior subspaces," *Transactions on Machine Learning Research*, 2024.

[14] R. Barbano *et al.*, "An educated warm start for deep image prior-based micro ct reconstruction," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 1210–1222, 2022.

[15] P. L. Bartlett, A. Montanari, and A. Rakhlin, "Deep learning: A statistical viewpoint," *Acta Numerica*, vol. 30, pp. 87–201, 2021.

[16] P. W. Battaglia *et al.*, "Relational inductive biases, deep learning, and graph networks," *arXiv preprint arXiv:1806.01261*, 2018.

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

[18] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[19] M. Benning and M. Burger, "Modern regularization methods for inverse problems," *Acta Numerica*, vol. 27, pp. 1–111, 2018.

[20] C. I. Bercea, P. C. Cattin, J. A. Schnabel, and J. Wolleb, "Denoising diffusion models for anomaly localization in medical images," *arXiv preprint arXiv:2410.23834*, 2024.

[21] M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini, "Image deblurring with Poisson data: From cells to galaxies," *Inverse Problems*, vol. 25, no. 12, pp. 123006, 26, 2009, ISSN: 0266-5611,1361-6420.

[22] V. Bettinardi, L. Presotto, E. Rapisarda, M. Picchio, L. Gianolli, and M. C. Gilardi, "Physical performance of the new hybrid PET/CT Discovery-690," *Medical Physics*, vol. 38, no. 10, pp. 5394–5411, 2011.

[23] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4.

[24] F. Bloch, "Nuclear induction," *Physical Review*, vol. 70, no. 7–8, pp. 460–474, Oct. 1946, ISSN: 0031-899X.

[25] A. Boquet-Pujadas, P. Del Aguila Pla, and M. Unser, "PET rebinning with regularized density splines," in *Proceedings of IEEE International Symposium on Biomedical Imaging*, 2023, pp. 1–4.

[26] J. E. Bowsher *et al.*, "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors," in *Proceedings of the IEEE Conference NSS/MIC/RTSD*, IEEE, vol. 4, 2004, pp. 2488–2492.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[28] A. Breger *et al.*, "A study of why we need to reassess full reference image quality assessment with medical images," *arXiv preprint arXiv:2405.19097*, 2024.

[29] A. Breger *et al.*, "A study on the adequacy of common iqa measures for medical images," *arXiv preprint arXiv:2405.19224*, 2024.

[30] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.

[31] J. Browne and A. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 687–699, 1996.

[32] T. F. Budinger, "PET instrumentation: What are the limits?" In *Seminars in Nuclear Medicine*, Elsevier, vol. 28, 1998, pp. 247–267.

[33] S. L. Campbell and C. D. Meyer, *Generalized inverses of linear transformations*. SIAM, 2009.

[34] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[35] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[36] A. Cárdenas-Blanco, C. Tejos, P. Irarrazaval, and I. Cameron, "Noise in magnitude magnetic resonance images," *Concepts in Magnetic Resonance Part A: An Educational Journal*, vol. 32, no. 6, pp. 409–416, 2008.

[37] L. F. Chalcroft *et al.*, "Development and evaluation of intraoperative ultrasound segmentation with negative image frames and multiple observer labels," in *Simplifying Medical Ultrasound: Second International Workshop, ASMUS*, Springer, 2021, pp. 25–34.

[38] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120–145, 2011.

[39] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[40] H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *Proceedings of the International Conference on Learning Representations*, 2023.

[41] H. Chung, S. Lee, and J. C. Ye, "Decomposed diffusion sampler for accelerating large-scale inverse problems," 2024.

[42] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," *Medical Image Analysis*, vol. 80, p. 102 479, 2022.

[43] L. Condat, "Discrete total variation: New definition and minimization," *SIAM Journal on Imaging Sciences*, vol. 10, no. 3, pp. 1258–1290, Jan. 2017, ISSN: 1936-4954.

[44] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965, ISSN: 1088-6842.

[45] G. Corda-D'Incan, J. Schnabel, and A. J. Reader, "Syn-net for synergistic deep-learned PET-MR reconstruction," in *Proceedings of the IEEE Conference NSS/MIC/RTSD*, IEEE, 2021.

[46] G. Daras *et al.*, "A survey on diffusion models for inverse problems," *arXiv preprint arXiv:2410.00083*, 2024.

[47] M. Z. Darestani, J. Liu, and R. Heckel, "Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing," in *Proceedings of the International Conference on Machine Learning*, PMLR, 2022, pp. 4754–4776.

[48] I. Daubechies, "Ten lectures on wavelets," *SIAM*, 1992.

[49] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum" a posteriori" estimates in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 20, no. 4, pp. 280–288, 2001.

[50] M. Defrise, P. E. Kinahan, and C. J. Michel, "Image reconstruction algorithms in PET," D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, Eds., pp. 63–91, 2005.

[51] A. J. Den Dekker and J. Sijbers, "Data distributions in magnetic resonance images: A review," *Physica Medica*, vol. 30, no. 7, pp. 725–741, 2014.

[52] A. Denker *et al.*, "Data-driven approaches for electrical impedance tomography image segmentation from partial boundary data," *Applied Mathematics for Modern Challenges*, 2024.

[53] A. Deshmane, V. Gulani, M. A. Griswold, and N. Seiberlich, "Parallel MR imaging," *Journal of Magnetic Resonance Imaging*, vol. 36, no. 1, pp. 55–72, 2012.

[54] P. Dhariwal and A. Q. Nichol, "Diffusion models beat GANs on image synthesis," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2021, pp. 8780–8794.

[55] A. G. Dimakis, "Deep generative models and inverse problems," in *Mathematical Aspects of Deep Learning*, Cambridge, UK: Cambridge University Press, 2022, pp. 400–421.

[56] S. Diolatzis, T. Zirr, A. Kuznetsov, G. Kopanas, and A. Kaplanyan, "N-dimensional gaussians for fitting of high dimensional functions," in *Proceedings of the ACM SIGGRAPH Conference*, 2024, pp. 1–11.

[57] S. Dittmer, T. Kluth, P. Maass, and D. Otero Baguer, "Regularization by architecture: A deep prior approach for inverse problems," *Journal of Mathematical Imaging and Vision*, vol. 62, no. 3, pp. 456–470, Oct. 2019.

[58] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[59] B. Efron, "Tweedie's formula and selection bias," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1602–1614, 2011.

[60] M. J. Ehrhardt, "Multi-modality imaging with structure-promoting regularizers," in *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, K. Chen, C.-B. Schönlieb, X.-C. Tai, and L. Younces, Eds., Cham, Switzerland: Springer International Publishing, 2021, pp. 1–38, ISBN: 978-3-030-03009-4.

[61] M. J. Ehrhardt, P. Markiewicz, and C.-B. Schönlieb, "Faster PET reconstruction with non-smooth priors by randomization and preconditioning," *Physics in Medicine & Biology*, vol. 64, no. 22, p. 225 019, 2019.

[62] M. J. Ehrhardt *et al.*, "PET reconstruction with an anatomical MRI prior using parallel level sets," *IEEE Transactions on Medical Imaging*, vol. 35, no. 9, pp. 2189–2199, 2016.

[63] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Science & Business Media, 2010.

[64] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[65] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems* (Mathematics and its Applications). Kluwer Academic Publishers Group, Dordrecht, 1996.

[66] F. H. Fahey, "Data acquisition in PET imaging," *Journal of Nuclear Medicine Technology*, vol. 30, no. 2, pp. 39–49, 2002.

[67] P. Fernsel, Ž. Kereta, and A. Denker, "Convergence properties of score-based models using graduated optimisation for linear inverse problems," *arXiv preprint arXiv:2404.18699*, 2024.

[68] J. A. Fessler, "Model-based image reconstruction for MRI," *IEEE Signal Processing magazine*, vol. 27, no. 4, pp. 81–89, 2010.

[69] J. Fessler and B. Sutton, "Nonuniform fast fourier transforms using min-max interpolation," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 560–574, 2003.

[70] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, p. 2379, Dec. 1987.

[71] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, vol. 222, no. 594-604, pp. 309–368, 1922.

[72] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323.

[73] G. Golub and V. Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, R1, 2003.

[74] K. Gong, C. Catana, J. Qi, and Q. Li, "PET image reconstruction using deep image prior," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1655–1665, 2018.

[75] K. Gong, K. Johnson, G. El Fakhri, Q. Li, and T. Pan, "PET image denoising based on denoising diffusion probabilistic model," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 51, no. 2, pp. 358–368, 2024.

[76] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[77] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 27, 2014.

[78] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified em algorithm," *IEEE Transactions on Medical Imaging*, vol. 9, no. 1, pp. 84–93, 1990.

[79] P. J. Green, "On use of the EM algorithm for penalized likelihood estimation," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 52, no. 3, pp. 443–452, 1990.

[80] M. A. Griswold *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.

[81] C. W. Groetsch, "Comments on Morozov's Discrepancy Principle," in *Improperly Posed Problems and Their Numerical Treatment: Conference Held at the Mathematisches Forschungsinstitut, Oberwolfach, September 26–October 2, 1982*, G. Hämmerlin and K.-H. Hoffmann, Eds. Basel: Birkhäuser Basel, 1983, pp. 97–104, ISBN: 978-3-0348-5460-3.

[82] A. Guazzo and M. Colarieti-Tosti, "Learned primal dual reconstruction for PET," *Journal of Imaging*, vol. 7, no. 12, 2021.

[83] M. S. Hansen and T. S. Sørensen, "Gadgetron: An open source framework for medical image reconstruction," *Magnetic Resonance in Medicine*, vol. 69, no. 6, pp. 1768–1776, Jul. 2012, ISSN: 1522-2594.

[84] K. M. Hanson and G. W. Wecksung, "Local basis-function approach to computed tomography," *Applied Optics*, vol. 24, no. 23, pp. 4028–4039, 1985.

[85] M. Hardt, B. Recht, and Y. Singer, "Train faster, generalize better: Stability of stochastic gradient descent," in *Proceedings of the International Conference on Machine Learning*, PMLR, 2016, pp. 1225–1234.

[86] F. Hashimoto, Y. Onishi, K. Ote, H. Tashima, A. J. Reader, and T. Yamaya, "Deep learning-based pet image denoising and reconstruction: A review," *Radiological physics and technology*, vol. 17, no. 1, pp. 24–46, 2024.

[87] F. Hashimoto, Y. Onishi, K. Ote, H. Tashima, and T. Yamaya, *Fully 3d implementation of the end-to-end deep image prior-based PET image reconstruction using block iterative algorithm*, Jul. 2023.

[88] F. Hashimoto, K. Ote, and Y. Onishi, "PET image reconstruction incorporating deep image prior and a forward projection model," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 6, no. 8, pp. 841–846, 2022.

[89] T. Hastie, *The elements of statistical learning: Data mining, inference, and prediction*, 2009.

[90] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[91] R. Heckel *et al.*, "Deep decoder: Concise image representations from untrained non-convolutional networks," in *Proceedings of the International Conference on Learning Representations*, 2019.

[92] R. Heckel, M. Jacob, A. Chaudhari, O. Perlman, and E. Shimron, "Deep learning for accelerated and robust mri reconstruction," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 37, no. 3, pp. 335–368, 2024.

[93] R. Heckel and M. Soltanolkotabi, "Denoising and regularization via exploiting the structural bias of convolutional generators. r. heckel and m. soltanolkotabi," in *Proceedings of the Eighth International Conference on Learning Representations*, 2020.

[94] W. Heitler, *The quantum theory of radiation*. Courier Corporation, 1984.

[95] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Medical physics*, vol. 12, no. 2, pp. 232–233, 1985.

[96] G. T. Herman and L. B. Meyer, "Algebraic reconstruction techniques can be made computationally efficient (positron emission tomography application)," *IEEE Transactions on Medical Imaging*, vol. 12, no. 3, pp. 600–609, 1993.

[97] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2020.

[98] J. Ho and T. Salimans, "Classifier-free diffusion guidance," 2021.

[99] T. Hohage and F. Werner, "Inverse problems with Poisson data: Statistical regularization theory, applications and algorithms," *Inverse Problems*, vol. 32, no. 9, pp. 093001, 56, 2016, ISSN: 0266-5611,1361-6420.

[100] L. Höllein, A. Božič, M. Zollhöfer, and M. Nießner, "3DGS-LM: faster gaussian-splatting optimization with levenberg-marquardt," *arXiv preprint arXiv:2409.12892*, 2024.

[101] D. I. Hoult and R. Richards, "The signal-to-noise ratio of the nuclear magnetic resonance experiment," *Journal of Magnetic Resonance (1969)*, vol. 24, no. 1, pp. 71–85, 1976.

[102] G. Hudson, A. Léger, B. Niss, I. Sebestyén, and J. Vaaben, "JPEG-1 standard 25 years: Past, present, and future reasons for a success," *Journal of Electronic Imaging*, vol. 27, no. 4, pp. 040 901–040 901, 2018.

[103] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 601–609, 1994.

[104] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[105] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, vol. 6, no. 24, pp. 695–709, 2005.

[106] K. Ito and B. Jin, *Inverse problems: Tikhonov theory and algorithms*. World Scientific, 2014, vol. 22.

[107] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir, "Robust compressed sensing MRI with deep generative priors," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2021, pp. 14 938–14 954.

[108] S. Jan *et al.*, "Gate v6: A major enhancement of the gate simulation platform enabling modelling of ct and radiotherapy," *Physics in Medicine & Biology*, vol. 56, no. 4, p. 881, 2011.

[109] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[110] F. John, "Continuous dependence on data for solutions of partial differential equations with a prescribed bound," *Communications on pure and applied mathematics*, vol. 13, no. 4, pp. 551–585, 1960.

[111] Z. Kadkhodaie, F. Guth, E. P. Simoncelli, and S. Mallat, "Generalization in diffusion models arises from geometry-adaptive harmonic representations," in *Proceedings of the International Conference on Learning Representations*, 2024.

[112] J. Kaipio and E. Somersalo, "Statistical inverse problems: Discretization, model reduction and inverse crimes," *Journal of Computational and Applied Mathematics*, vol. 198, no. 2, pp. 493–504, 2007.

[113] S. Kaplan and Y. Zhu, "Full-dose PET image estimation from low-dose PET image using deep learning: A pilot study," *Journal of Digital Imaging*, vol. 32, no. 5, pp. 773–778, 2019.

[114] N. Karakatsanis *et al.*, *Usability of PETSIRD, the PET raw data open format of the emission tomography standardization initiative (ETSI): Results from ETSI's first hackathon*, 2024.

[115] A. M. Karlberg, O. Sæther, L. Eikenes, and P. E. Goa, "Quantitative comparison of PET performance - siemens biograph mCT and mMR," *European Journal of Nuclear Medicine and Molecular Imaging: Physics*, vol. 3, no. 1, 2016.

[116] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 33, pp. 12 104–12 114, 2020.

[117] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

[118] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.

[119] T. Karras *et al.*, "Alias-free generative adversarial networks," *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 34, pp. 852–863, 2021.

[120] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 35, NeurIPS, 2022, pp. 26 565–26 577.

[121] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d Gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics (ToG)*, vol. 42, no. 4, pp. 1–14, 2023.

[122] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," in *Proceedings of the International Conference on Learning Representations*, 2016.

[123] P. E. Kinahan, D. Townsend, T. Beyer, and D. Sashin, "Attenuation correction for a combined 3D PET/CT scanner," *Medical physics*, vol. 25, no. 10, pp. 2046–2053, 1998.

[124] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, 2015.

[125] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of the International Conference on Learning Representations*, ICLR, 2014.

[126] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 30, 2017.

[127] O. Klein and Y. Nishina, "Über die streuung von strahlen durch frei elektronen nach der neue relativistische quantum dynamics von dirac," *Journal for Physics*, vol. 52, no. 11, pp. 853–868, 1929.

[128] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer Berlin Heidelberg, 1992, ISBN: 9783662126165.

[129] M. Kochenderfer, *Algorithms for Optimization*. The MIT Press Cambridge, 2019.

[130] O. Kraff and H. H. Quick, "7T: Physics, safety, and potential clinical applications," *Journal of Magnetic Resonance Imaging*, vol. 46, no. 6, pp. 1573–1589, 2017.

[131] C.-M. Lai and P. Lauterbur, "True three-dimensional image reconstruction by nuclear magnetic resonance zeugmatography," *Physics in Medicine & Biology*, vol. 26, no. 5, p. 851, 1981.

[132] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature*, vol. 242, no. 5394, pp. 190–191, Mar. 1973, ISSN: 1476-4687.

[133] P. Lecoq *et al.*, "Roadmap toward the 10 ps time-of-flight PET challenge," *Physics in Medicine & Biology*, vol. 65, no. 21, 21RM01, 2020.

[134] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[135] J. Lee, *Introduction to topological manifolds*. Springer Science & Business Media, 2010, vol. 202.

[136] T. S. Lee, "Image representation using 2D gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, 1996.

[137] P. Lesonen, V. Wettenhovi, V. Kolehmainen, A. Pulkkinen, and M. Vauhkonen, "Anatomy-guided multi-resolution image reconstruction in PET," *Physics in Medicine & Biology*, vol. 69, no. 10, p. 105 023, 2024.

[138] J. Leuschner, M. Schmidt, D. Otero Baguer, D. Erzmann, and M. Baltazar, *DIVαl library*, version v0.5.8, 2021.

[139] B. Levac, A. Jalal, K. Ramchandran, and J. I. Tamir, "MRI reconstruction with side information using diffusion models," pp. 1436–1442, 2023.

[140] C. S. Levin and E. J. Hoffman, "Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution," *Physics in Medicine & Biology*, vol. 44, no. 3, p. 781, 1999.

[141] R. M. Lewitt, "Alternatives to voxels for image representation in iterative reconstruction algorithms," *Physics in Medicine & Biology*, vol. 37, no. 3, p. 705, 1992.

[142] X. Li, J. Thickstun, I. Gulrajani, P. Liang, and T. B. Hashimoto, "Diffusion-LM improves controllable text generation," in *Proceedings to the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2022, pp. 4328–4343.

[143] M. E. Lindemann, V. Stebner, A. Tschischka, J. Kirchner, L. Umutlu, and H. H. Quick, "Towards fast whole-body PET/MR: Investigation of PET image quality versus reduced PET acquisition times," *PLOS ONE*, vol. 13, no. 10, pp. 1-1–5, Oct. 2018.

[144] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-cnn for image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 773–782.

[145] A. Lou and S. Ermon, "Reflected diffusion models," in *Proceedings of the International Conference on Machine Learning*, vol. 202, PMLR, 2023, pp. 22 675–22 701.

[146] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 35, NeurIPS, 2022, pp. 5775–5787.

[147] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[148] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1999.

[149] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[150] S. Matej and R. M. Lewitt, "Practical considerations for 3-d image reconstruction using spherically symmetric volume elements," *IEEE Transactions on Medical Imaging*, vol. 15, no. 1, pp. 68–78, 1996.

[151] H. M. McConnell, "Reaction rates by nuclear magnetic resonance," *The Journal of Chemical Physics*, vol. 28, no. 3, pp. 430–431, 1958.

[152] K. Miyasawa, "An empirical bayes estimator of the mean of a normal population," *Bulletin of the International Statistical Institute*, vol. 38, no. 181-188, pp. 1–2, 1961.

[153] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[154] J. J. Moré, "The levenberg-marquardt algorithm: Implementation and theory," in *Proceedings of the Biennial Numerical analysis Conference*, Springer, 2006, pp. 105–116.

[155] A. Mortazi, J. K. Udupa, D. Odhner, Y. Tong, and D. A. Torigian, "Post-acquisition standardization of positron emission tomography images," *Frontiers in Nuclear Medicine*, vol. 3, p. 1 210 931, 2023.

[156] M. J. Muckley, R. Stern, T. Murrell, and F. Knoll, "TorchKbNufft: A high-level, hardware-agnostic non-uniform fast Fourier transform," in *ISMRM Workshop on Data Sampling & Image Reconstruction*, 2020.

[157] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[158] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the International Conference on Machine Learning*, 2010, pp. 807–814.

[159] B. Neyshabur, R. Tomioka, and N. Srebro, "In search of the real inductive bias: On the role of implicit regularization in deep learning," in *Proceedings of the International Conference on Learning Representations*, 2015.

[160] Q. K.-T. Ng, E. K. A. Triumbari, N. Omidvari, S. R. Cherry, R. D. Badawi, and L. Nardo, "Total-body PET/CT -first clinical experiences and future perspectives," in *Seminars in Nuclear Medicine*, Elsevier, vol. 52, 2022, pp. 330–339.

[161] J. Nuyts, D. Bequé, P. Dupont, and L. Mortelmans, "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," *IEEE Transactions on Nuclear Science*, vol. 49, no. 1, pp. 56–60, 2002.

[162] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens, "Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms," *IEEE Transactions on Medical Imaging*, vol. 18, no. 5, pp. 393–403, 1999.

[163] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143–150, 2000.

[164] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, e3, 2016.

[165] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," in *Proceedings of the Conference on Medical Imaging with Deep Learning*, 2018.

[166] A. Oppenheim, D. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the fast fourier transform," *Proceedings of the IEEE*, vol. 59, no. 2, pp. 299–301, 1971.

[167] E. Ovtchinnikov *et al.*, "SIRF: Synergistic image reconstruction framework," *Computer Physics Communications*, vol. 249, p. 107 087, 2020.

[168] F. Padormo, A. Beqiri, J. V. Hajnal, and S. J. Malik, "Parallel transmission for ultrahigh-field imaging," *NMR in Biomedicine*, vol. 29, no. 9, pp. 1145–1161, 2016.

[169] C. D. Pain, G. F. Egan, and Z. Chen, "Deep learning-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, no. 9, pp. 3098–3118, 2022.

[170] N. Parikh, S. Boyd, *et al.*, "Proximal algorithms," *Foundations and trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[171] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 32, NeurIPS, 2019.

[172] R. Penrose, "A generalized inverse for matrices," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 51, no. 3, pp. 406–413, 1955.

[173] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[174] K. B. Petersen, M. S. Pedersen, *et al.*, "The matrix cookbook," *Technical University of Denmark*, vol. 7, no. 15, p. 510, 2008.

[175] K. N. Plataniotis and D. Hatzinakos, "Gaussian mixtures and their applications to signal processing," *Advanced Signal Processing Handbook*, pp. 89–124, 2000.

[176] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1481–1497, 1990.

[177] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[178] M. Pourya, A. Boquet-Pujadas, and M. Unser, "A box-spline framework for inverse problems with continuous-domain sparsity constraints," *IEEE Transactions on Computational Imaging*, 2024.

[179] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.

[180] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Physics in Medicine & Biology*, vol. 51, no. 15, R541, 2006.

[181] I. I. Rabi, J. R. Zacharias, S. Millman, and P. Kusch, "A new method of measuring nuclear magnetic moment," *Physical Review*, vol. 53, no. 4, pp. 318–318, Feb. 1938, ISSN: 0031-899X.

[182] A. Rahmim, J. Qi, and V. Sossi, "Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls," *Medical Physics*, vol. 40, no. 6Part1, p. 064 301, 2013.

[183] S. Ramasinghe and S. Lucey, "Beyond periodicity: Towards a unifying framework for activations in coordinate-MLPs," in *Proceedings of the European Conference on Computer Vision*, Springer, 2022, pp. 142–158.

[184] Z. Ramzi, K. Michalewicz, J.-L. Starck, T. Moreau, and P. Ciuciu, "Wavelets in the deep learning era," *Journal of Mathematical Imaging and Vision*, vol. 65, no. 1, pp. 240–251, 2023.

[185] Z. Ramzi, B. Rémy, F. Lanusse, J.-L. Starck, and P. Ciuciu, "Denoising score-matching for uncertainty quantification in inverse problems," 2020.

[186] S. Ravula, B. Levac, A. Jalal, J. Tamir, and A. Dimakis, "Optimizing sampling patterns for compressed sensing MRI with diffusion generative models," in *Proceedings of the Conference on Advances of Neural Information Processing Systems/Workshop on Deep Learning and Inverse Problems*.

[187] A. J. Reader, G. Corda-D'Incan, A. Mehranian, C. da Costa-Luis, S. Ellis, and J. A. Schnabel, "Deep learning for PET image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 1–25, 2020.

[188] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A haar wavelet-based perceptual similarity index for image quality assessment," *Signal Processing: Image Communication*, vol. 61, pp. 33–43, 2018.

[189] R. Remmert, *Theory of complex functions*. Springer Science & Business Media, 1991, vol. 122.

[190] F. D. S. Ribeiro and B. Glocker, *Demystifying variational diffusion models*, 2024.

[191] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 684–10 695.

[192] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.

[193] W. C. Röntgen, "On a new kind of rays," *Science*, vol. 3, no. 59, pp. 227–231, 1896.

[194] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[195] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[196] C. Saharia *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2022, pp. 36 479–36 494.

[197] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 29, 2016.

[198] V. Saragadam, R. Balestriero, A. Veeraraghavan, and R. G. Baraniuk, *Deeptensor: Low-rank tensor decomposition with deep network priors*, 2022.

[199] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, "Wire: Wavelet implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 507–18 516.

[200] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations* (Institute of Mathematical Statistics Textbooks). Cambridge University Press, Cambridge, 2019, ISBN: 978-131-6649-46-6.

[201] M. Sarracanie and N. Salameh, "Low-field mri: How low can we go? a fresh view on an old debate," *Frontiers in Physics*, vol. 8, p. 172, 2020.

[202] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational methods in imaging*. Springer, 2009, vol. 167.

[203] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for MR image reconstruction," in *Proceedings of the International Conference on Information Processing in Medical Imaging*, Springer, 2017, pp. 647–658.

[204] D. Schote, L. Winter, C. Kolbitsch, G. Rose, O. Speck, and A. Kofler, "Joint B_0 and image reconstruction in low-field mri by physics-informed deep-learning," *IEEE Transactions on Biomedical Engineering*, 2024.

[205] G. Schramm, *Simulated brainweb PET/MR data sets for denoising and deblurring*, 2021.

[206] G. Schramm, "Reconstruction-free positron emission imaging: Fact or fiction?" *Frontiers in Nuclear Medicine*, vol. 2, p. 936 091, 2022.

[207] G. Schramm and K. Thielemans, "PARALLELPROJ—an open-source framework for fast calculation of projections in tomography," *Frontiers in Nuclear Medicine*, vol. 3, Jan. 2024, ISSN: 2673-8880.

[208] M. Schweiger and S. R. Arridge, "Image reconstruction in optical tomography using local basis functions," *Journal of Electronic Imaging*, vol. 12, no. 4, pp. 583–593, 2003.

[209] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4d xcat phantom for multimodality imaging research," *Medical physics*, vol. 37, no. 9, pp. 4902–4915, 2010.

[210] L. Shen, J. Pauly, and L. Xing, "NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 770–782, 2022.

[211] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

[212] Z. Shi, P. Mettes, S. Maji, and C. G. Snoek, "On measuring and controlling the spectral bias of the deep image prior," *International Journal of Computer Vision*, vol. 130, no. 4, 2022, ISSN: 15731405.

[213] K. Shibuya *et al.*, "Limit of spatial resolution in FDG-PET due to annihilation photon non-collinearity," in *Proceedings of the World Congress on Medical Physics and Biomedical Engineering*, Springer, 2007, pp. 1667–1671.

[214] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, no. 1, pp. 1193–1216, Mar. 2001.

[215] I. R. Singh, R. Barbano, Ž. Kereta, B. Jin, K. Thielemans, and S. Arridge, "3D PET-DIP reconstruction with relative difference prior using a SIRF-based objective," in *Proceedings to the Seventeenth International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Fully3D, 2023.

[216] I. R. Singh, A. Denker, B. Jin, K. Thielemans, and S. Arridge, "Investigating intensity normalisation for PET reconstruction with supervised deep learning," in *Proceedings of the IEEE Conference NSS/MIC/RTSD*, 2023.

[217] I. R. Singh, O. Jaubert, B. Jin, K. Thielemans, and S. Arridge, "Magnetic resonance fingerprinting with total nuclear variation regularisation," in *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine*, London, UK, May 2022, p. 2515.

[218] I. R. Singh *et al.*, "Deep image prior PET reconstruction using a sirf-based objective," in *Proceedings of the IEEE Conference NSS/MIC/RTSD*), IEEE, 2022, pp. 1–2.

[219] I. R. Singh *et al.*, "MR-blob: Coordinate-transformed blobs for parallel MRI reconstruction," in *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine*, Singapore, SG, May 2024, p. 1911.

[220] I. R. Singh *et al.*, "Score-based generative models for PET image reconstruction," *Machine Learning for Biomedical Imaging*, vol. 2, no. Special Issue for Generative Models, pp. 547–585, 2024.

[221] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 33, pp. 7462–7473, 2020.

[222] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the International Conference on Machine Learning*, ICML, 2015, pp. 2256–2265.

[223] S. Somayajula, C. Panagiotou, A. Rangarajan, Q. Li, S. R. Arridge, and R. M. Leahy, "PET image reconstruction using information theoretic anatomical priors," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 537–549, 2011.

[224] B. Song, S. M. Kwon, Z. Zhang, X. Hu, Q. Qu, and L. Shen, "Solving inverse problems with latent diffusion models via hard data consistency," in *Proceedings of the International Conference on Learning Representations*, 2024.

[225] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proceedings of the International Conference on Learning Representations*, ICLR, 2021.

[226] Y. Song, C. Durkan, I. Murray, and S. Ermon, "Maximum likelihood training of score-based diffusion models," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2021, pp. 1415–1428.

[227] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2019, pp. 11 895–11 907.

[228] Y. Song and S. Ermon, "Improved techniques for training score-based generative models," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, NeurIPS, 2020, pp. 12 438–12 448.

[229] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *Proceedings of the International Conference on Learning Representations*, ICLR, 2022.

[230] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *Proceedings of the International Conference on Learning Representations*, ICLR, 2021.

[231] C. J. Stone, "Optimal global rates of convergence for nonparametric regression," *The Annals of Statistics*, pp. 1040–1053, 1982.

[232] J. Tachella, J. Tang, and M. Davies, "The neural tangent link between cnn denoisers and non-local filters," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8618–8627.

[233] K. Thielemans *et al.*, "STIR: Software for tomographic image reconstruction release 2," *Physics in Medicine & Biology*, vol. 57, no. 4, p. 867, 2012.

[234] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," in *Doklady akademii nauk*, Russian Academy of Sciences, vol. 151, 1963, pp. 501–504.

[235] S. Tong, A. M. Alessio, and P. E. Kinahan, "Noise and signal properties in PSF-based fully 3D PET image reconstruction: An experimental evaluation," *Physics in Medicine & Biology*, vol. 55, no. 5, p. 1453, 2010.

[236] H. C. Torrey, "Bloch equations with diffusion terms," *Physical Review*, vol. 104, no. 3, pp. 563–565, Nov. 1956, ISSN: 0031-899X.

[237] D. Tran, J. Snoek, and B. Lakshminarayanan, "Practical uncertainty estimation and out-of-distribution robustness in deep learning," NeurIPS Tutorial, Google Brain, 2020.

[238] Y.-J. Tsai *et al.*, "Benefits of using a spatially-variant penalty strength with anatomical priors in PET reconstruction," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 11–22, 2019.

[239] R. Twyman *et al.*, "A demonstration of stir-gate-connection," in *Proceedings of the IEEE Conference NSS/MIC/RTSD*, IEEE, 2021, pp. 1–3.

[240] M. Uecker, T. Hohage, K. T. Block, and J. Frahm, "Image reconstruction by regularized nonlinear inversion—joint estimation of coil sensitivities and image content," *Magnetic Resonance in Medicine*, vol. 60, no. 3, pp. 674–682, 2008.

[241] M. Uecker *et al.*, "ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magnetic resonance in medicine*, vol. 71, no. 3, pp. 990–1001, 2014.

[242] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 9446–9454.

[243] A. Vahdat and J. Kautz, "NVAE: A deep hierarchical variational autoencoder," *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 33, pp. 19 667–19 679, 2020.

[244] W. van Aarle *et al.*, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015, ISSN: 0304-3991.

[245] S. Vandenberghe, M. E. Daube-Witherspoon, R. M. Lewitt, and J. S. Karp, "Fast reconstruction of 3D time-of-flight PET data by axial rebinning and transverse mashing," *Physics in Medicine & Biology*, vol. 51, no. 6, p. 1603, 2006.

[246] V. Vapnik, "Statistical learning theory," *John Wiley & Sons*, vol. 2, pp. 831–842, 1998.

[247] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[248] M. Vauhkonen, J. Kaipio, E. Somersalo, and P. Karjalainen, "Electrical impedance tomography with basis constraints," *Inverse Problems*, vol. 13, no. 2, p. 523, 1997.

[249] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[250] H. Wang, T. Li, Z. Zhuang, T. Chen, H. Liang, and J. Sun, *Early stopping for deep image prior*, 2023.

[251] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[252] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. Springer International Publishing, 2006, ISBN: 9783031022388.

[253] C. C. Watson, D. Newport, and M. E. Casey, "A single scatter simulation technique for scatter correction in 3D PET," in *Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Springer, 1996, pp. 255–268.

[254] G. Webber and A. J. Reader, "Diffusion models for medical image reconstruction," *BJR| Artificial Intelligence*, ubae013, 2024.

[255] C. Williams, F. Falck, G. Deligiannidis, C. C. Holmes, A. Doucet, and S. Syed, "A unified framework for U-Net design and analysis," *Advances in Neural Information Processing Systems*, vol. 36, pp. 27 745–27 782, 2023.

[256] S. N. Williams, P. McElhinney, and S. Gunamony, "Ultra-high field MRI: Parallel-transmit arrays and RF pulse design," *Physics in Medicine & Biology*, vol. 68, no. 2, 02TR02, 2023.

[257] Z. Xiao, K. Kreis, and A. Vahdat, "Tackling the generative learning trilemma with denoising diffusion GANs," in *Proceedings of the International Conference on Learning Representations*, ICLR, 2022.

[258] H. Xue, S. Inati, T. S. Sørensen, P. Kellman, and M. S. Hansen, "Distributed MRI reconstruction using gadgetron-based cloud computing," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1015–1025, Mar. 2014, ISSN: 0740-3194.

[259] S. Yamamoto *et al.*, "Development of ultrahigh resolution Si-PM-based PET system using 0.32 mm pixel scintillators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 836, pp. 7–12, 2016.

[260] J. C. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 991–1048, 2018.

[261] G. Yüce, G. Ortiz-Jiménez, B. Besbinar, and P. Frossard, "A structured dictionary perspective on implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19 228–19 238.

[262] J. Zbontar *et al.*, *fastMRI: An open dataset and benchmarks for accelerated MRI*, 2018. arXiv: 1811.08839.

[263] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the European Conference on Computer Vision*, Springer Verlag, 2014, pp. 818–833.

[264] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021.

[265] R. Zhang, "Making convolutional networks shift-invariant again," in *Proceedings of the International Conference on Machine Learning*, PMLR, 2019, pp. 7324–7334.

[266] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[267] Y. Zhu *et al.*, "Denoising diffusion models for plug-and-play image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*NTIRE*), 2023.

[268] M. Zwicker, H. Pfister, J. Van Baar, and M. Gross, "Surface splatting," in *Proceedings of the Conference on Computer Graphics and Interactive Techniques*, 2001, pp. 371–378.

Appendix A

SGM-based Unsupervised Anomaly Localisation

The guided reconstruction approach presented in Section 5.5.1 facilitates the translation from healthy MR images to healthy PET images in the absence of data consistency constraints. However, when data consistency is enforced using measurements from anatomies with lesions, a distributional mismatch arises. This mismatch occurs because the score-model, trained on healthy patients, is applied to measurements from patients with pathologies. By sampling multiple reconstructions using the proposed guided reconstruction method (Section 5.5.1), we can assess the variance across these reconstructions to localise abnormalities. Specifically, ten reconstructions were computed from a single set of measurements. The resulting mean, variance, and variance-to-mean ratio are displayed in Fig. A.1, which highlight areas indicative of anomalies.

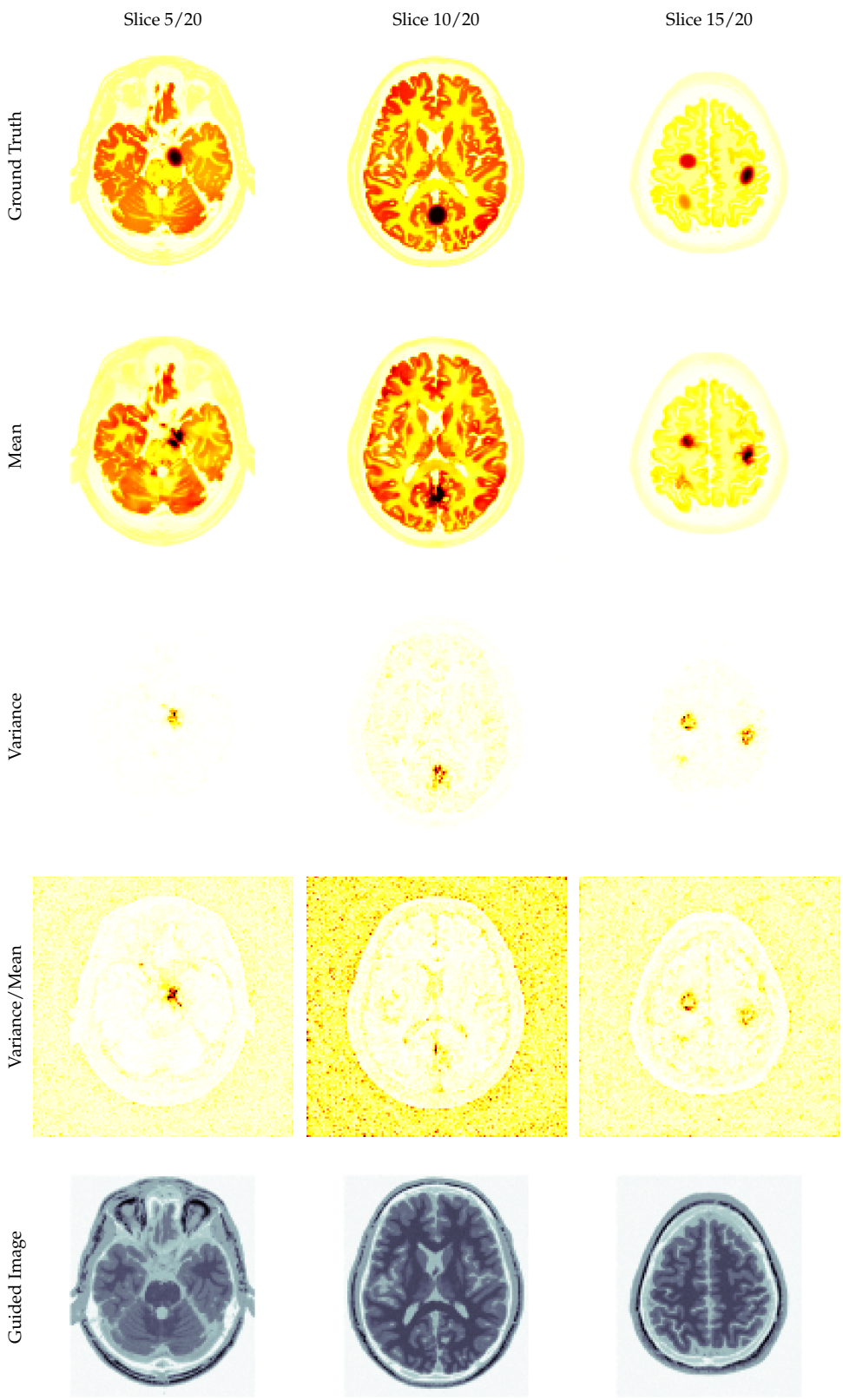


Figure A.1: Guided reconstruction of score-model trained on healthy patients and tested with measurements of an unhealthy patient. Ten reconstructions are sampled and the mean and variance are shown.

Appendix B

3D A-Blobs and Volumetric Regularisation

In the previous implementation `AutoGrad` was used to automatically compute the gradients using reverse-mode automatic differentiation. Given the simplicity of A-Blobs the gradients with respect to the parameters $c \in \mathbb{C}$, $\boldsymbol{\mu} \in \mathbb{R}^3$ and $\mathbf{L} \in \mathbb{R}^{3\times3}$ can be derived. Starting with the definition of a single adaptable blob:

$$b(\mathbf{r}; c, \boldsymbol{\mu}, \mathbf{L}) = c \exp\left(-\frac{\|\mathbf{L}(\mathbf{r}-\boldsymbol{\mu})\|_2^2}{2}\right),$$

The gradients are:

$$\begin{aligned}
\frac{\partial b(\mathbf{r}; c, \boldsymbol{\mu}, \mathbf{L})}{\partial c} &= \exp\left(-\frac{1}{2}\|\mathbf{L}(\mathbf{r}-\boldsymbol{\mu})\|_2^2\right), \\
\frac{\partial b(\mathbf{r}; c, \boldsymbol{\mu}, \mathbf{L})}{\partial \boldsymbol{\mu}} &= \mathbf{L}^\top \mathbf{L}(\mathbf{r}-\boldsymbol{\mu}) b(\mathbf{r}), \\
\frac{\partial b(\mathbf{r}; c, \boldsymbol{\mu}, \mathbf{L})}{\partial \mathbf{L}} &= \left[\mathbf{L}(\mathbf{r}-\boldsymbol{\mu})(\mathbf{r}-\boldsymbol{\mu})^\top b(\mathbf{r})\right]_{\text{lower}}.
\end{aligned}$$

The derivatives of the Fourier-transformed A-Blobs are:

$$\frac{\partial \mathcal{F}[b](\mathbf{k})}{\partial c} = (2\pi)^{\frac{N_{dim}}{2}} |\det(\mathbf{U})| \exp\left(-\mathbf{i}\mathbf{k}^\top \boldsymbol{\mu} - \frac{1}{2}\|\mathbf{U}\mathbf{k}\|_2^2\right) \tag{B.2a}$$

$$\frac{\partial \mathcal{F}[b](\mathbf{k})}{\partial \boldsymbol{\mu}} = -\mathbf{i}\mathbf{k}\mathcal{F}[b](\mathbf{k}) \tag{B.2b}$$

$$\frac{\partial \mathcal{F}[b](\mathbf{k})}{\partial u_{ij}} = \begin{cases} (-\mathbf{k}^\top \mathbf{U}\mathbf{k})\,[\mathcal{F}[b](\mathbf{k})]_{\text{upper}} & \text{if } i \neq j \\ (\frac{1}{u_{ij}} - \mathbf{k}^\top \mathbf{U}\mathbf{k})\,[\mathcal{F}[b](\mathbf{k})]_{\text{upper}} & \text{if } i = j. \end{cases} \tag{B.2c}$$

Note that special attention need to by taken to ensure $\boldsymbol{\mu} \in \mathbb{R}^3$ and $\mathbf{L} \in \mathbb{R}^{3\times 3}$, which requires taking the conjugate Wirtinger derivative [189].

In addition to the explicit gradients above, we can also derive a sparsity promoting volumetric regularisation, where the volume of an A-Blobs is $v = c(u_{11} \cdot u_{22} \cdot u_{33})^2$ and $\det(\mathbf{U}) = u_{11} \cdot u_{22} \cdot u_{33}$.[1]

$$\eta(\mathbf{c}, \boldsymbol{\Gamma}) = \lambda \sum_{b=1}^{N_b} |v_{(b)}|,$$

the proximal operator corresponds to the soft-thresholding function [170]:

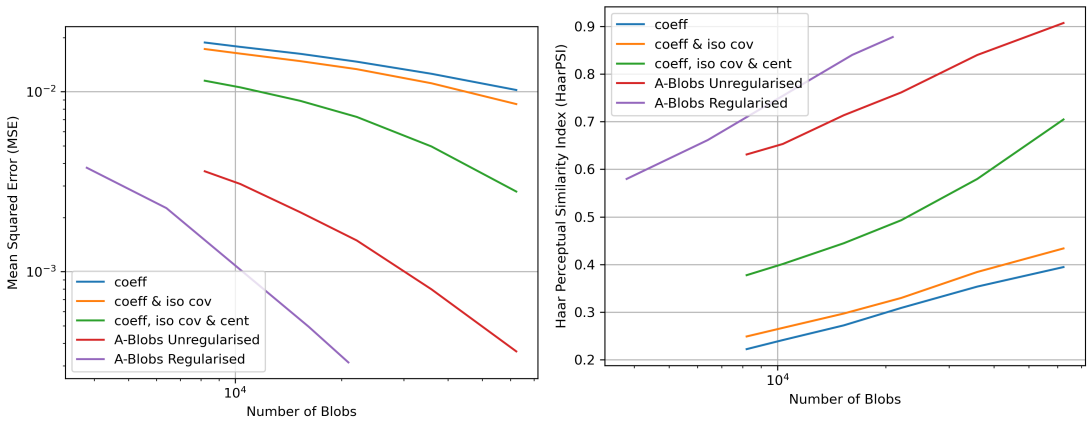
$$[\text{prox}_{\gamma\lambda\|\cdot\|_1}(\mathbf{v})]_{(b)} = \begin{cases} v_{(b)} - \gamma\lambda & \text{if} \quad v_{(b)} > \gamma\lambda, \\ 0 & \text{if} \quad |v_{(b)}| \leq \gamma\lambda, \\ v_{(b)} + \gamma\lambda & \text{if} \quad v_{(b)} < \gamma\lambda, \end{cases}$$

where γ described in Section 2.2.3.2. This is equivalent to:

$$[\text{prox}_{\gamma\lambda\|\cdot\|_1}(\mathbf{v})]_{(b)} = \begin{cases} c_{(b)} - \frac{\gamma\lambda}{([\mathbf{U}_{(b)}]_{11}\cdot[\mathbf{U}_{(b)}]_{22}\cdot[\mathbf{U}_{(b)}]_{33})^2} & \text{if} \quad v_{(b)} > \gamma\lambda, \\ 0 & \text{if} \quad |v_{(b)}| \leq \gamma\lambda, \\ c_{(b)} + \frac{\gamma\lambda}{([\mathbf{U}_{(b)}]_{11}\cdot[\mathbf{U}_{(b)}]_{22}\cdot[\mathbf{U}_{(b)}]_{33})^2} & \text{if} \quad v_{(b)} < \gamma\lambda. \end{cases}$$

In Fig. B.1 we show results for image representation in 3D for various parameterisation and with the proposed regularisation. In Fig. B.2 we give qualitative results for the various parameterisations in 3D.

[1] This is related to reweighted ℓ_1 minimisation [34], where we reweigh the coefficient based on the determinant or the covariance.



(**a**) Image quality assessed with Mean Squared Error vs number of A-Blobs.

(**b**) Image quality assessed with Haar Wavelet-Based Perceptual Similarity Index vs number of A-Blobs.

Figure B.1: Plots assessing image representation of A-Blobs with varying degree of parameterisation and number of A-Blobs.

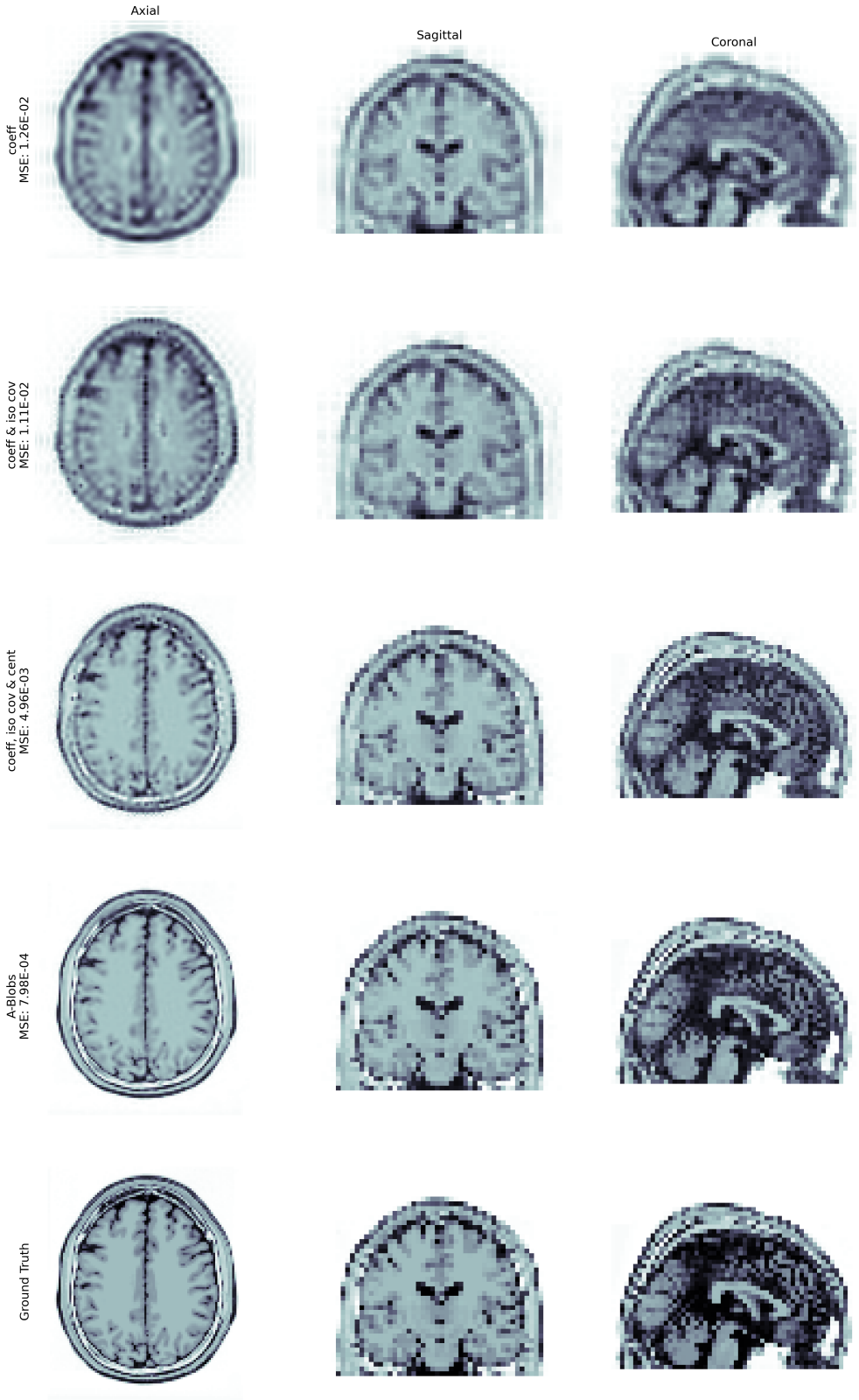


Figure B.2: Images of changing the parameterisation, while keeping number of A-Blobs the same. Mean Squared Error (MSE) is given for each parameterisation.