Spring 5-15-2023

Deep Learning for Tomographic Image Reconstruction Guided by Generative Models and Image Science

Sayantan Bhadra
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering
Department of Computer Science & Engineering

Dissertation Examination Committee:
Ulugbek S. Kamilov, Chair
Mark A. Anastasio, Co-Chair
Abhinav K. Jha
Brendan Juba
Fan Lam
Alvitta Ottley

Deep Learning for Tomographic Image Reconstruction Guided by Generative Models and Image Science

by
Sayantan Bhadra

A dissertation presented to
the McKelvey School of Engineering
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2023
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CSGM	Compressed Sensing using Generative Models
CT	Computed Tomography
DGM	Deep Generative Model
DIP	Deep Image Prior
DPS	Diffusion Posterior Sampling
FID	Fréchet Inception Distance
FOM	Figure-of-merit
GAN	Generative Adversarial Network
IAGAN	Image-adaptive Generative Adversarial Network
IND	In-distribution
INN	Invertible Neural Network
MCMC	Markov Chain Monte Carlo
MRI	Magnetic Resonance Imaging
OOD	Out-of-distribution
PLS-TV	Penalized Least Squares with Total Variation penalty
ProGAN	Progressively-growing Generative Adversarial Network
PULSE	Photo Upsampling via Latent Space Exploration
SSIM	Structural Similarity Index Measure
StyleGAN	Style-based Generative Adversarial Network
TV	Total Variation

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Mark Anastasio, for his excellent mentorship and kind support over these years. Through his guidance, he has instilled a deep sense of curiosity, passion, and attention to detail that have contributed immensely to my personal and scientific growth.

I thank my other committee members, Dr. Ulugbek Kamilov, Dr. Abhinav Jha, Dr. Brendan Juba, Dr. Fan Lam, and Dr. Alvitta Ottley, for offering valuable inputs at different stages of my research. I also appreciate the enormous help and advice I received from Dr. Frank Brooks and Dr. Umberto Villa during our collaboration on various research projects.

It has been a joy to be a part of the Computational Imaging Science Laboratory and spend time doing research with colleagues who are motivated, friendly, and always eager to help each other. Dr. Yang Lou, Dr. Yujia Chen, and Dr. Seonyeong Park guided me patiently during my initial years. I also value the exceptional contributions made by Dr. Weimin Zhou and Mr. Varun Kelkar during our collaboration on multiple projects.

During these years, I have enjoyed the company of friends like Garv, Prashant, Hai, and Daniel, who have always kept me in good spirits through various ups and downs. I have shared many memorable moments with them and others who have enriched this journey.

I would not have undertaken this path without the encouragement of my teachers during my high school and undergraduate years. They laid a strong foundation and inspired me to pursue scientific research.

Finally, I thank my parents and my entire family for their unconditional love and unwavering support. Their dedication to my well-being and growth has made this achievement possible.

Sayantan Bhadra

Washington University in St. Louis
May 2023

Dedicated to my parents and my brother.

ABSTRACT OF THE DISSERTATION

Deep Learning for Tomographic Image Reconstruction Guided by Generative Models and Image Science

by

Sayantan Bhadra

Doctor of Philosophy in Computer Science

Washington University in St. Louis, 2023

Professor Ulugbek S. Kamilov, Chair

Professor Mark A. Anastasio, Co-Chair

Tomographic image reconstruction is generally an ill-posed inverse problem. Such inverse problems are typically regularized using prior knowledge of the sought-after object property. Recently, deep neural networks have been actively investigated for regularizing image reconstruction problems by learning a prior for the object properties from training images. Deep generative models such as generative adversarial networks (GANs) have demonstrated the ability to learn object distributions comprehensively and synthesize high-quality images. This dissertation explores novel generative model-constrained reconstruction methods that employ state-of-the-art GANs in the context of ill-posed tomographic imaging problems. The symbiotic relationship between image science and deep learning to enable responsible artificial intelligence (AI) applications in medical imaging is also demonstrated. In the first part of the dissertation, an image reconstruction method is proposed (IAGAN-TV), which extends the IAGAN method introduced in ill-posed image restoration problems with an improved regularization strategy and employs a progressively growing GAN architecture in an image-adaptive framework. We demonstrate the ability of the IAGAN-TV method to recover fine structures in ill-posed image reconstruction problems, which cannot be achieved using sparsity-promoting penalties alone. The stability and generalization properties of the

proposed method are established. In the second part, a formal definition of "hallucinations" is introduced in the context of image reconstruction using fundamental image science principles derived from linear operator theory. We demonstrate how the ability to define image hallucinations allows the quantification of false structures in reconstructed images, enabling preliminary assessments of deep learning-based reconstruction methods via virtual imaging trials. In the final part of the dissertation, a method to produce multiple data-consistent solutions to image reconstruction problems is proposed (PULSE++) that employs a style-based GAN architecture (StyleGAN). The PULSE++ method extends the PULSE method introduced in single-image super-resolution tasks to general ill-posed inverse problems. The proposed method improves the performance of PULSE by stabilizing the core optimization method and utilizing more accurate statistical knowledge of the StyleGAN latent space. The scalability of the PULSE++ method and its effectiveness with different practical measurement noise distributions is demonstrated. We illustrate how the ability to produce multiple data-consistent solutions using the PULSE++ method enabled new assessments of imaging systems, such as uncertainty quantification in image reconstruction.

Chapter 1

Introduction

Inverse problems are ubiquitous in scientific computing, with wide-ranging applications in computer science, medicine, astronomy, geophysics, and many other fields. In such problems, algorithms are designed to utilize a set of observations or measurements in order to estimate the parameters characterizing the source which produced them. Because of this common underlying principle, in recent years advancements in algorithms for inverse problems in computer science, such as various applications in computer vision and machine learning, have been translated into other practical areas of scientific computing, and vice-versa. This dissertation focuses on one such interplay, where we explore how state-of-the-art machine learning methods in computer vision, in particular deep neural networks and generative models, can advance tomographic image reconstruction, while themselves being evolved guided by imaging science principles.

1.1 Overview and Motivation

Modern imaging systems are computed in nature and require an appropriate image reconstruction method for estimating an object from a collection of tomographic measurements [70]. In practice, the acquired measurement data are noisy, and at times incomplete, in which case the associated inverse problem will be ill-posed. For example, to accelerate the data-acquisition in magnetic resonance imaging (MRI), undersampled measurement data can be purposely acquired [146]. In such cases, image reconstruction methods that seek to estimate approximate but potentially useful estimates of the object property require regularization. Regularization strategies incorporate appropriate prior knowledge of the object, known as object priors in the Bayesian parlance, into the reconstruction process. For example, sparsity-promoting regularization strategies have found great success in recent years [33, 121, 113]. More recently, a variety of data-driven methods have been proposed

whereby a more accurate object prior is learned from large databases of existing imaging data. Many data-driven methods employ deep neural networks, otherwise known as deep learning (DL) [84, 100, 113].

A majority of the early development of DL-based algorithms for image reconstruction had been focused on supervised learning methods [47, 113]. While they have shown great promise, supervised learning methods need to be retrained or fine-tuned when the measurement process changes due to a change in imaging system parameters, a common occurrence in practical clinical settings. This is often impractical due to the long training time of such methods and the wide variety of imaging system parameters that are adjusted to optimize imaging system performance [19]. In order to circumvent these issues, more recently unsupervised learning methods have been promoted that employ deep generative models (DGMs) [47] as priors for regularizing the image reconstruction problem [28]. DGMs such as generative adversarial networks (GANs) [48], normalizing flows [43] and diffusion models [123, 60] have recently shown great promise in modeling distributions of high-dimensional images, that includes modern medical imaging applications [122, 79]. Inspired by image-adaptive GANs (IAGANs) [5], in Chapter 3, a GAN-based reconstruction method is proposed which incorporated the prior learned by a state-of-the-art model (ProGAN) [72] within an optimization framework that also employed sparsity-promoting regularization. As a result of imposing a more accurate prior, fine structures could be recovered from highly undersampled and noisy measurements that are relevant for medical diagnosis but may be oversmoothed in images reconstructed with traditional sparsity-promoting penalties alone. This was the first application of generative model-constrained optimization [28] in ill-posed tomographic imaging applications.

While both supervised and unsupervised DL-based methods have continued to evolve and demonstrated encouraging results in ill-posed inverse problems [108], an analysis of the prior information learned by these deep networks and their ability to generalize to data that may lie outside the training distribution is still being explored. For example, an inaccurate prior might lead to false structures being hallucinated in the reconstructed image and that is a cause for serious concern in medical imaging [64, 49, 11, 90]. However, a mathematical definition for such false structures did not exist previously, which prevented the detection and quantification of hallucinations produced by DL-based methods [107]. In Chapter 4, a formal definition of hallucinations is proposed for general ill-posed inverse problems using linear operator theory for image formation [19, 17]. The proposed formalism can be employed

to separate sources of systemic error, like measurement error or noise, from errors that arise due to inaccurate priors imposed during image reconstruction.

So far, our discussion on regularization of ill-posed tomographic imaging problems has been restricted to recovering single image estimates, which is the norm. However, by definition, ill-posed inverse problems admit an infinite number of solutions, and there may be multiple objects that are all consistent with the same measurement data. The ability to generate such alternate solutions is important because it may enable new assessments of imaging systems. In principle, this can be achieved by means of posterior sampling methods [133, 104]. In recent years, deep neural networks have been employed for posterior sampling using approximate Markov Chain Monte Carlo (MCMC) techniques with promising results [126, 120, 106, 68]. However, such methods are not yet for use with practical tomographic imaging applications. On the other hand, empirical sampling methods [126, 9, 16] may be computationally feasible for large-scale imaging systems and enable uncertainty quantification for practical applications. Empirical sampling involves solving a regularized inverse problem within a stochastic optimization framework to obtain alternate data-consistent solutions [16]. In Chapter 5, a new empirical sampling method is proposed that computes multiple solutions of a tomographic inverse problem that are application-relevant and consistent with the same acquired measurement data. The method operates by repeatedly solving an optimization problem in the latent space of a style-based generative adversarial network (StyleGAN), and was inspired by the Photo Upsampling via Latent Space Exploration (PULSE) method [102] that was developed for super-resolution tasks. The proposed method, termed as PULSE++, extends the PULSE method to general ill-posed inverse problems such as tomographic image reconstruction, and improves the quality of data-consistent solutions by employing more accurate statistical knowledge of the StyleGAN latent space and stabilizing the stochastic optimization method involved.

1.2 Outline and Contributions

The outline of the dissertation and our contributions in each topic are described as follows.

In Chapter 2, we formally state the image reconstruction problem. Salient aspects of linear operator theory and regularization in image reconstruction are reviewed.

In Chapter 3, we introduce the IAGAN-TV algorithm as the first application of a generative model-constrained iterative optimization method for highly ill-posed tomographic imaging problems. This work was proposed and described previously [26, 81, 82]. The contributions of our work in computer science and medical imaging fields respectively are summarized below.

Contributions to the computer science field: The IAGAN method had been introduced for ill-posed image restoration problems in computer vision with promising results. We demonstrate that additional regularization in the form of a sparsity-promoting penalty, such as the total variation (TV) penalty [35], can improve reconstruction quality. The proposed method was termed as IAGAN-TV. The stability and generalization properties of the IAGAN-TV method were established using bias-variance analysis and performing image reconstruction from out-of-distribution measurements. A preliminary study was performed comparing the performance of IAGAN-TV, which employs an explicit generative model (GAN) with representation error [28], against generative model-constrained reconstruction using an explicit generative model (normalizing flows) without representation error.

Contributions to the medical imaging field: Our work generalized the IAGAN framework to tomographic imaging applications and demonstrated the effectiveness of generative model-constrained reconstruction with imaging operators relevant for modern medical imaging modalities, such as multi-coil MRI. Due to imposing a stronger prior, the IAGAN-TV algorithm enabled image reconstruction that could recover fine structures in highly ill-posed settings, such as accelerated MRI, which may be infeasible using traditional sparsity-promoting penalties alone. Furthermore, due to its stability and generalization properties, the IAGAN-TV algorithm can be reliably applied in clinical applications while being a data-driven method. Since the IAGAN-TV algorithm does not require knowledge of the measurement process during training of the GAN, there is no need for retraining when imaging system parameters are varied, which provides a practical advantage when compared with emerging supervised learning-based methods.

In Chapter 4, we mathematically formalize the concept of hallucinations for general linear inverse problems such as image reconstruction. This work was proposed and described previously [22, 23]. The contributions of our work in computer science and medical imaging fields respectively are summarized below.

Contributions to the computer science field: Image hallucinations have previously been studied in the context of image super-resolution [14, 93, 139, 45]. However, a formal definition of hallucinations did not exist, due to which false structures that may be incorrectly introduced by the imposed prior in image restoration tasks could not be quantitatively analyzed. We proposed the concept of hallucination maps using linear operator theory, which enabled detection and assessment of image hallucinations, that cannot be achieved using traditional error maps. Additionally, we presented preliminary comparison studies on generalization properties of image restoration methods using the same neural network architecture (U-Net) [116] but different algorithms – a data-driven and image domain-learning method [69] versus a non-data-driven and iterative optimization method (DIP) [134, 92]. It was observed that the data-driven method in [69] may produce more hallucinations that can negatively impact the downstream task as compared to the non-data-driven method DIP with out-of-distribution data at test time. This highlighted the importance of using hallucination maps in simulation studies to assess generalization properties of emerging image restoration algorithms that employ deep learning.

Contributions to the medical imaging field: While DL-based methods have recently shown promise in tomographic image reconstruction, evidence suggests that they may be unstable and produce false structures [11, 49]. The resulting artifacts can possibly lead to an incorrect medical diagnosis, thus limiting the application of DL-based methods in clinical settings. Our work addressed the urgent need to quantitatively analyze false structures that arise due to the imposed prior, i.e. hallucinations, and enabled new assessments of instability and generalization properties of DL-based reconstruction methods via virtual imaging trials. The effectiveness of hallucination maps in isolating errors due to data-driven and non-data-driven priors was demonstrated using a stylized MRI system and datasets consisting of adult and pediatric brain MRI images. Additionally, the proposed formalism of hallucinations is general and can be applied across imaging modalities.

In Chapter 5, we propose the PULSE++ method to obtain multiple data-consistent and application-relevant solutions of ill-posed tomographic imaging problems. This work was proposed and described previously [24]. The contributions of our work in computer science and medical imaging fields respectively are summarized below.

Contributions to the computer science field: We generalized and extended the PULSE algorithm [102], originally introduced in single image super-resolution tasks to obtain multiple high-resolution images consistent with the same low-resolution image, to general inverse problems such as image reconstruction. The PULSE method operates by repeatedly solving an optimization problem in the latent space of a style-based generative adversarial network (StyleGAN) [76], and constraining the solution to lie near the manifold of the StyleGAN generator. However, deviation from the manifold of the StyleGAN generator may lead to artifacts and produce images that are not relevant to the application. Furthermore, the PULSE method employs inaccurate statistics of the StyleGAN latent space and performs optimization over a limited search space to prevent instability, which leads to lack of data consistency. Our proposed method, PULSE++, ensures that each solution lies strictly on the manifold of the StyleGAN generator, and the optimization method is redesigned for increased stability such that the entire StyleGAN latent space is included in the domain of the objective function. Additionally, PULSE++ employs accurate statistical knowledge of the StyleGAN latent space. These modifications allowed PULSE++ to achieve the required data consistency while producing solutions that are more relevant to the application since they are constrained to lie on the manifold of StyleGAN generator. The proposed method also outperformed a recent diffusion model-based posterior sampling method (DPS) [124, 68] in terms of maintaining data consistency of alternate solutions. The ability of PULSE++ to produce multiple data-consistent solutions in the presence of exact Poisson noise was demonstrated, which is infeasible with current DPS methods [38].

Contributions to the medical imaging field: The PULSE++ method was employed to produce multiple data-consistent and application-relevant solutions at scale using two different stylized tomographic imaging modalities – MRI and X-ray computed tomography (CT). The ability of PULSE++ to produce diverse data-consistent solutions of ill-posed tomographic imaging problems will enable new types of assessment and refinement of imaging systems. These include computation of uncertainty maps [128, 119] to reveal reconstruction risk, estimating figures-of-merit (FOMs) that describe the likelihood of hallucinated false structures and analyzing the impact of the null space of a linear imaging operator [19] in new, problem-specific ways. Additionally, ensembles of data-consistent solutions obtained using PULSE++ will enable the design of numerical experiments to reveal image reconstruction instabilities [49] and advance task-informed adaptive imaging procedures [39, 18]. The proposed method is general and can be employed for any imaging modality and measurement noise distribution.

Chapter 2

Background on Image Reconstruction

2.1 The Image Reconstruction Problem

A digital imaging system can be described as a continuous-to-discrete (CD) mapping [19, 10]:

$$\mathbf{g} = \boldsymbol{\mathcal{H}} f(\mathbf{r}) + \mathbf{n}, \tag{2.1}$$

where $f(\mathbf{r}) \in \mathbb{L}_2(\mathbb{R}^d)$ is a function of continuous variables that represents the object being imaged, the vector $\mathbf{g} \in \mathbb{E}^M$ denotes the measured data samples and $\mathbf{n} \in \mathbb{E}^M$ is the measurement noise. The CD operator $\boldsymbol{\mathcal{H}} : \mathbb{L}_2(\mathbb{R}^d) \to \mathbb{E}^M$ describes the action of the imaging system. In practice, discrete-to-discrete (DD) imaging models are often employed as approximations to the true CD imaging model. In a DD model, an N-dimensional approximation of $f(\mathbf{r})$ is utilized [19, 10]:

$$f_a(\mathbf{r}) = \sum_{n=1}^{N} [f]_n \psi_n(\mathbf{r}), \tag{2.2}$$

where the subscript a stands for approximate, $[f]_n$ is the n-th element of the coefficient vector $\mathbf{f} \in \mathbb{E}^N$ and $\psi_n(\mathbf{r})$ is the n-th expansion function. When $\boldsymbol{\mathcal{H}}$ is linear, the DD imaging system can be expressed as

$$\mathbf{g} \approx \boldsymbol{\mathcal{H}} f_a(\mathbf{r}) + \mathbf{n} = \sum_{n=1}^{N} [f]_n \boldsymbol{\mathcal{H}} \psi_n(\mathbf{r}) + \mathbf{n} \equiv \mathbf{H}\mathbf{f} + \mathbf{n}, \tag{2.3}$$

where $\mathbf{H} : \mathbb{E}^N \to \mathbb{E}^M$ is the system matrix constructed using $\boldsymbol{\mathcal{H}}$ and $\{\psi_n(\mathbf{r})\}_{n=1}^N$. Image reconstruction methods based on Eq. (2.3) seek to estimate $\mathbf{f}$ from $\mathbf{g}$, after which the approximate object function $f_a(\mathbf{r})$ can be determined by use of Eq. (2.2). Since the coefficient vector $\mathbf{f}$ is the sought-after representation of the object function $f(\mathbf{r})$, $\mathbf{f}$ is also referred to as the object vector. A well-known expansion function is the pixel expansion function. For

two-dimensional objects $f(\mathbf{r})$ with $\mathbf{r} = (x, y)$, the pixel expansion function can be expressed as [10]:

$$\psi_n(\mathbf{r}) = \begin{cases} 1, & \text{if } |x - x_n| \text{ and } |y - y_n| \leq \frac{\beta}{2} \\ 0, & \text{otherwise} \end{cases} \tag{2.4}$$

where $\mathbf{r}_n = (x_n, y_n)$ represents the coordinates of the n-th grid point of a uniform Cartesian lattice and β denotes the spacing between the lattice points. When a pixel expansion function is employed, the corresponding coefficient vector $\mathbf{f}$ directly provides a digital image representation of the continuous object function $f_a(\mathbf{r})$.

2.2 Description of Measurement and Null Subspaces

For the DD imaging model described by Eq. (2.3), the properties of $\mathbf{H}$ affect the ability to estimate $\mathbf{f}$ uniquely and stably. In the absence of measurement noise, $\mathbf{f}$ can be determined uniquely from measurements $\mathbf{Hf}$ when $\mathbf{H}$ is injective or if $\mathbf{f}$ is known to lie in a subset S of $\mathbb{E}^N$ and the restriction $\mathbf{H}|_S$ is injective. The ability to stably reconstruct an estimate of $\mathbf{f}$ is also of fundamental importance. Stability is a way of quantifying how "close" two estimates $\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2$ of $\mathbf{f}$ will be, if they are estimated from two "close" measurement vectors $\mathbf{g}_1$ and $\mathbf{g}_2$ respectively. For instance, $\mathbf{g}_1$ and $\mathbf{g}_2$ may correspond to the same object but differ due to them having two different measurement noise realizations. A popular notion of stability is based on how the ℓ_2-distance between $\hat{\mathbf{f}}_1$ and $\hat{\mathbf{f}}_2$ relates to that between $\mathbf{g}_1$ and $\mathbf{g_2}$ [15]:

$$\|\hat{\mathbf{f}}_1 - \hat{\mathbf{f}}_2\|_2 \leq \alpha \|\mathbf{g}_1 - \mathbf{g}_2\|_2, \tag{2.5}$$

where α is a constant that is additionally required to be smaller than a tolerance value ϵ.

The ability to estimate $\mathbf{f}$ stably can be analyzed through the lens of the singular value decomposition (SVD) of $\mathbf{H}$ [19]:

$$\mathbf{H} = \sum_{n=1}^{R} \sqrt{\mu_n} \mathbf{v}_n \mathbf{u}_n^{\dagger}. \tag{2.6}$$

Here, $\mathbf{u}_n$ and $\mathbf{v}_n$ are the singular vectors of $\mathbf{H}$ and $\sqrt{\mu_n}$ are the singular values. The vector $\mathbf{u}_n^\dagger$ is the adjoint of $\mathbf{u}_n$ and the integer $R > 0$ denotes the rank of $\mathbf{H}$, where $\mathbf{H}$ is not necessarily full-rank. The singular values $\sqrt{\mu_n}$ are ordered such that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_R > 0$.

A pseudoinverse-based estimate of $\mathbf{f}$ can be computed as $\hat{\mathbf{f}}_{pinv} \equiv \mathbf{H}^+\mathbf{g}$, where the linear operator $\mathbf{H}^+$ denotes the Moore-Penrose pseudoinverse of $\mathbf{H}$ that can be expressed as

$$\mathbf{H}^+ = \sum_{n=1}^{R} \frac{1}{\sqrt{\mu_n}} \mathbf{u}_n \mathbf{v}_n^\dagger. \tag{2.7}$$

From Eq. (2.3), due to the linearity of $\mathbf{H}^+$, $\hat{\mathbf{f}}_{pinv}$ can be represented as

$$\hat{\mathbf{f}}_{pinv} = \mathbf{H}^+\mathbf{g} \approx \mathbf{H}^+(\mathbf{H}\mathbf{f} + \mathbf{n}) = \mathbf{H}^+\mathbf{H}\mathbf{f} + \mathbf{H}^+\mathbf{n}. \tag{2.8}$$

Due to the presence of the term $\mathbf{H}^+\mathbf{n}$ in Eq. (2.8), when the trailing singular values of $\mathbf{H}$ are small, α in Eq. (2.5) is large, leading to unstable estimates of $\mathbf{f}$. In this scenario, a truncated pseudoinverse can be defined as

$$\mathbf{H}_P^+ = \sum_{n=1}^{P} \frac{1}{\sqrt{\mu_n}} \mathbf{u}_n \mathbf{v}_n^\dagger, \tag{2.9}$$

where the integer $P \leq R$ is chosen such that, for a given tolerance ϵ, $\mathbf{H}_P^+\mathbf{g}$ is a stable, linear estimate of $\mathbf{f}$ according to Eq. (2.5) with $\mu_P > 1/\epsilon^2 \geq \mu_{P+1}$. The truncated pseudoinverse can be used to form projection operators that project $\mathbf{f} \in \mathbb{E}^N$ onto orthogonal subspaces – the 'generalized' null and measurement spaces [40]. The generalized null space of $\mathbf{H}$, denoted by $\mathcal{N}_P(\mathbf{H})$, is spanned by the singular vectors $\{\mathbf{u}_n\}_{n=P+1}^N$ that correspond to singular values satisfying $\sqrt{\mu_n} \leq 1/\epsilon$. The orthogonal complement of the generalized null space is the generalized measurement space $\mathcal{N}_P^\perp(\mathbf{H})$.

Definition 2.2.1 (*Generalized measurement and null components*)**.** Let $\mathbf{H}$ and $\mathbf{H}_P^+$ denote the forward and truncated pseudoinverse operators, described in Equations (2.3) and (2.9) respectively. Let $\mathbf{H}_P$ denote the truncated forward operator, defined as

$$\mathbf{H}_P = \sum_{n=1}^{P} \sqrt{\mu_n} \mathbf{v}_n \mathbf{u}_n^\dagger. \tag{2.10}$$

Note that $\mathbf{H}_P^+ = (\mathbf{H}_P)^+$. The coefficient vector $\mathbf{f}$ can be uniquely decomposed as $\mathbf{f} = \mathbf{f}_{meas} + \mathbf{f}_{null}$, where $\mathbf{f}_{meas} \in \mathcal{N}_P^\perp(\mathbf{H})$ and $\mathbf{f}_{null} \in \mathcal{N}_P(\mathbf{H})$ are specified as

$$\mathbf{f}_{meas} = \mathcal{P}_{meas}\mathbf{f} = \mathbf{H}_P^+\mathbf{H}\mathbf{f} = \mathbf{H}_P^+\mathbf{H}_P\mathbf{f}, \tag{2.11}$$

and

$$\mathbf{f}_{null} = \mathcal{P}_{null}\mathbf{f} = [\mathbf{I}_N - \mathbf{H}_P^+\mathbf{H}]\mathbf{f} = [\mathbf{I}_N - \mathbf{H}_P^+\mathbf{H}_P]\mathbf{f}. \tag{2.12}$$

Here, the projection operators $\mathcal{P}_{meas}$ and $\mathcal{P}_{null}$ project $\mathbf{f}$ to $\mathcal{N}_P^\perp(\mathbf{H})$ and $\mathcal{N}_P(\mathbf{H})$ [19], and $\mathbf{I}_N$ is the identity operator in $\mathbb{E}^N$.

In special cases where the singular values $\sqrt{\mu_n}$ and the tolerance ϵ are such that $P = R$, the generalized null space is spanned by the singular vectors $\{\mathbf{u}_n\}_{n=R+1}^N$ with singular values $\sqrt{\mu_n} = 0$. In such cases, the generalized null space reduces to the true null space

$$\mathcal{N}_P(\mathbf{H}) = \mathcal{N}(\mathbf{H}) \equiv \{\mathbf{f} \in \mathbb{E}^N | \mathbf{H}\mathbf{f} = \mathbf{0}\}, \tag{2.13}$$

where $\mathbf{0}$ is the zero vector in $\mathbb{E}^M$. Correspondingly, the true measurement space is the orthogonal complement of the true null space. By definition, the true null space contains those object vectors that are mapped to the zero measurement data vector and hence are 'invisible' to the imaging system.

Having obtained the generalized measurement and null components of $\mathbf{f}$, the approximate object function $f_a(\mathbf{r})$ can also be decomposed into generalized measurement and null components:

$$\begin{aligned} f_a(\mathbf{r}) &= \sum_{n=1}^{N} [f]_n \psi_n(\mathbf{r}) \\ &= \sum_{n=1}^{N} [f_{meas}]_n \psi_n(\mathbf{r}) + \sum_{n=1}^{N} [f_{null}]_n \psi_n(\mathbf{r}) \\ &= f_{a,meas}(\mathbf{r}) + f_{a,null}(\mathbf{r}). \end{aligned} \tag{2.14}$$

Note that for all $\mathbf{g}_1, \mathbf{g}_2 \in \mathbb{E}^M$, $\|\mathbf{H}_P^+\mathbf{g}_1 - \mathbf{H}_P^+\mathbf{g}_2\| \leq (1/\sqrt{\mu_P})\|\mathbf{g}_1 - \mathbf{g_2}\|$, whereas for all $\eta \in \mathcal{N}_P(\mathbf{H})$, $\|\eta\| \geq \|\mathbf{H}\eta\|/\sqrt{\mu_{P+1}}$. Hence, for a given $\mathbf{f} \in \mathbb{E}^N$, $\mathbf{f}_{meas}$ is the component of $\mathbf{f}$ that can be stably estimated via the truncated pseudoinverse from the measurement data. Contrarily,

$\mathbf{f}_{null}$ cannot be stably estimated from the measurement data alone; additional information provided through priors and regularization is needed to estimate this component.

2.3 Regularization in Tomographic Image Reconstruction

As discussed above, in order to obtain a stable estimate of $\mathbf{f}$ from incomplete and/or noisy measurements, imposition of prior knowledge about the object is generally needed. A flexible method of incorporating priors in the estimation of $\mathbf{f}$ is through the Bayesian formalism, where $\mathbf{f}$, $\mathbf{g}$ and $\mathbf{n}$ are treated as instances of random vectors with distributions p_{f}, p_{g} and p_{n} respectively [15]. It is assumed that p_{f}, i.e. the distribution over all objects is known, and is called the *prior*. By Bayes' theorem, the posterior distribution $p_{\mathbf{f}|\mathbf{g}}$, given by

$$p_{\mathsf{f}|\mathsf{g}}(\mathbf{f}|\mathbf{g}) = \frac{p_{\mathsf{g}|\mathsf{f}}(\mathbf{g}|\mathbf{f})p_{\mathsf{f}}(\mathbf{f})}{p_{\mathsf{g}}(\mathbf{g})}, \tag{2.15}$$

characterizes the probability density over all possible values of the object given the prior and the noise model. Estimates such as the *maximum a posteriori* (MAP) estimate $\operatorname{argmax}_{\mathbf{f}} p_{\mathsf{f}|\mathsf{g}}(\mathbf{f}|\mathbf{g})$ can then be obtained from the posterior.

Regularization via penalization is an alternative formalism to incorporate prior knowledge. Here, the image reconstruction task can be formulated as an optimization problem such as [10]

$$\hat{\mathbf{f}} = \operatorname*{argmin}_{\mathbf{f}} \mathcal{J}(\mathbf{g}, \mathbf{H}\mathbf{f}) + \lambda \mathcal{R}(\mathbf{f}), \tag{2.16}$$

where the data fidelity term $\mathcal{J}(\mathbf{g}, \mathbf{H}\mathbf{f})$ enforces the estimate $\hat{\mathbf{f}}$ when acted upon by $\mathbf{H}$ to agree with the observed measurement data $\mathbf{g}$ and the penalty term $\mathcal{R}(\mathbf{f})$ encourages the solution to be consistent with the assumed prior. The hyper-parameter λ controls the trade-off between data fidelity and regularization. Often, the penalty term $\mathcal{R}(\mathbf{f})$ is hand-crafted to encode priors such as the smoothness of natural images or sparsity of natural images in some transform domain [113]. The solution obtained through this formalism can be interpreted as the MAP estimate obtained from the Bayesian formalism in Eq. (2.15), with $p_{\mathbf{f}}(\mathbf{f}) = \exp(-\lambda\mathcal{R}(\mathbf{f}))$ and $p_{\mathsf{g}|\mathbf{f}}(\mathbf{g}|\mathbf{f}) = \exp(-\mathcal{J}(\mathbf{g}, \mathbf{H}\mathbf{f}))$. A formulation of Eq. (2.16) which

is commonly solved is the penalized least-squares optimization problem with the sparsity-promoting total variation (TV) semi-norm penalty (PLS-TV) [35]:

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{f}\|_{TV}. \tag{2.17}$$

Proximal-gradient methods are employed to solve the PLS-TV optimization problem [35, 21]. While sparsity-promoting regularization is effective for many ill-posed tomographic imaging problems [33, 121, 113], such hand-crafted penalties do not comprehensively describe complex properties of the sought-after object and may not produce useful images when the measurements are extremely sparse, such as highly accelerated MRI applications [146].

Regularization can also be interpreted as restricting the possible solutions to a subset $S_\mu \subset \mathbb{E}^N$, with S_μ being a member of a family of subsets parameterized by μ. The reconstruction procedure can then be represented by a possibly nonlinear mapping $\mathcal{B}_\mu : \mathbb{E}^M \to S_\mu$, with the image estimate given by $\hat{\mathbf{f}} = \mathcal{B}_\mu(\mathbf{g})$. Ideally, it is desirable that $\mathcal{B}_\mu$ satisfies the stability criterion described in Eq. (2.5). Recently, methods that implicitly or explicitly learn a regularizer from existing data have been proposed. Methods based on dictionary learning and learning sparsifying transforms were some of the earliest applications of such data-driven regularization [111, 130, 112, 117]. However, the most actively investigated data-driven regularization methods involve learning from training data by use of deep neural networks, popularly known as deep learning [47, 137]. Deep learning has been employed in different ways to explicitly or implicitly impose priors in image reconstruction problems. For example, within the context of an end-to-end learned reconstruction mapping, a prior is imposed that is implicitly specified by the distribution of training data and network topology. Such methods, referred to as *supervised learning* [47], were the earliest proposed deep learning-based methods for highly ill-posed image reconstruction problems with promising results [69, 53]. However, since they are learned end-to-end, such supervised learning methods involve knowledge of the measurement process during training, and need to be re-trained or fine-tuned when image acquisition parameters are altered. Additionally, some supervised learning methods may be unstable and/or may not generalize well to out-of-distribution measurements [11, 49]. Recently, *unsupervised learning* [47] methods have been proposed for ill-posed inverse problems, which may circumvent these issues and operate by learning the manifold of the object distribution using deep generative models (DGMs) [47, 28]. A comprehensive survey of the current state of deep learning-based methods in tomographic image reconstruction can be found in recent reviews [113, 101, 54].

Chapter 3

Tomographic Image Reconstruction with Image-adaptive Priors Learned by use of Generative Adversarial Networks

3.1 Overview

Deep generative models such as generative adversarial networks (GANs) [48] have shown great promise in estimating the prior distributions for images. Bora *et al.* [28] developed an unsupervised learning method called Compressed Sensing using Generative Models (CSGM) for incorporating such learned priors in ill-posed inverse problems, where instead of sparsity-promoting penalties as in the PLS-TV formulation in Eq. (2.17), the regularization is performed by enforcing the image to lie in the range of a pre-trained generative model. However, in practice, it is difficult for a GAN to span all possible images that may arise from the true distribution. Hence, by constraining the reconstructed image to lie in the range of the generator, a potential lack of fidelity may be introduced between the reconstructed image and the observed measurements in the measurement space of the imaging operator $\mathbf{H}$ [19]. In order to mitigate the problem of limited representation capabilities of a GAN, Hussein *et al.* [5] proposed an image-adaptive GAN-based (IAGAN) framework, where the trained generative model parameters are also tuned to be consistent with the observed measurement data. This results in a higher fidelity with the observed measurements while still maintaining the learned prior over the imaging object obtained by pre-training the GAN. However, the IAGAN method had been applied only to image restoration problems in computer vision. It was necessary for the IAGAN method to be validated using tomographic imaging operators

in order to be applicable for medical imaging. Additionally, the IAGAN method relied on early stopping in a stochastic gradient descent (SGD) framework to prevent overfitting to noise. Alternative regularization techniques that could mitigate the need for early stopping needed to be explored.

In this study, we extend the IAGAN method to general ill-posed inverse problems such as image reconstruction, and demonstrate the performance of the method using both simulated and experimental measurements. A state-of-the-art GAN called Progressive Growing of GANs (ProGAN) [72] was trained on the publicly available NYU fastMRI dataset [146] containing knee MRI images and raw measurements. It was observed that the trained generator could produce knee images similar to ground truth images in the training dataset. The learned generative model was employed in the IAGAN framework to reconstruct images from highly subsampled k-space data belonging to a previously unseen validation dataset. It is demonstrated that by using an image-adaptive GAN-based reconstruction method on incomplete measurement data, we can obtain high-fidelity images and recover fine structures relevant for medical diagnosis that may be oversmoothed by sparsity-promoting penalties. The stability and generalization properties of the IAGAN method are analyzed.

3.2 Generative Adversarial Networks (GANs)

Generative adversarial networks (GANs) [48] are deep generative models that have recently shown encouraging results in learning distributions of images and demonstrated high-quality image synthesis performance, including medical applications [122]. In GANs, a generator network and a discriminator network are trained through an adversarial process [48]. Here, we consider a true object image $\mathbf{f} \in \mathbb{R}^N$ sampled from a data distribution $p_{\mathbf{f}}$. The generator maps a random vector $\mathbf{z} \in \mathbb{R}^k$ to a synthetic object image $\hat{\mathbf{f}} = G(\mathbf{z}; \theta_G)$, where $G : \mathbb{R}^k \to \mathbb{R}^N$ is the mapping represented by a neural network with parameters θ_G. The discriminator is an inference network, parameterized by θ_D, that represents a mapping $D : \mathbb{R}^N \to \mathbb{R}$ of the input image ($\mathbf{f}$ or $\hat{\mathbf{f}}$) to a real-valued scalar. In the adversarial process, D is trained to maximally differentiate the synthetic image $\mathbf{f}$ from the true image $\hat{\mathbf{f}}$, and G is trained to maximally fool D such that the generated synthetic image $\hat{\mathbf{f}}$ is wrongly classified as a true image. This adversarial process can be represented by a two-player minimax game with value function

$V(D, G)$:

$$\min_{\theta_G} \max_{\theta_D} V(D, G) = \mathbb{E}_{\mathbf{f} \sim p_{\mathbf{f}}}[l\left(D(\mathbf{f}; \theta_D)\right)] + \mathbb{E}_{\mathbf{z} \sim p_z}[l(1 - D\left(G(\mathbf{z}; \theta_G)\right))], \tag{3.1}$$

where $l(.)$ represents a suitable objective function. Let θ_G^* and θ_D^* denote the optimal parameters for G and D respectively after stable convergence has been reached in the above minimax game. When the global optimum of this minimax game is achieved as G and D are given sufficient capacity, i.e., in the non-parametric limit, the synthetic images generated by the generator G can not be differentiated from the true images by using any observer, and the synthetic image distribution $p_{\hat{\mathbf{f}}}$ equals the true image distribution $p_{\mathbf{f}}$: $p_{\hat{\mathbf{f}}} = p_{\mathbf{f}}$.

3.3 Progressive Growing of GANs (ProGANs)

In practice, however, stabilization of GAN training has been known to be difficult due to the adversarial learning process [47], which has served as a bottleneck in using GANs to reliably generate high-resolution images. Recently, Karras *et al.* [72] proposed a training strategy for GANs that has mitigated the stabilization problem of GAN training to great effect and resulted in GANs being able to generate realistic natural images at resolutions as high as 1024×1024 pixels. In this novel learning strategy called Progressive Growing of GANs (ProGAN), the training starts from low-resolution images and layers are added progressively to both the generator and the discriminator networks to increase the resolution (Fig. 3.1). Such a progressive training strategy, along with a novel value function $V(G, D)$ based on the Wasserstein metric [88, 12, 51] resulted in higher stability in the training process and significantly improved the synthesis quality of GANs for high-resolution images.

Figure 3.1: ProGAN: Training starts with generator G and discriminator D corresponding to low spatial resolution of 4x4 pixels. As training progresses, layers are added to G and D to gradually increase the spatial resolution of the generated images towards the final resolution, which for our study is 256×256.

3.4 Image-Adaptive GAN-Based Reconstruction (IA-GAN) for Tomographic Imaging

Once a GAN has been stably trained and can generate images similar to samples from the true data distribution, the learned generator network can be used as a prior for solving linear inverse problems such as Eq. (2.1). In the context of tomographic imaging, the learned prior can be employed for reconstructing images from incomplete and/or corrupted measurement data. Bora *et al.* [28] proposed the reconstructed image to be the one that is constrained to lie in the range of the pre-trained generator and agrees with the measurement data in a least squares sense through the imaging operator $\mathbf{H}$ (CSGM):

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}G(\mathbf{z};\theta_G^*)\|_2^2, \tag{3.2}$$

and the reconstructed image $\hat{\mathbf{f}} := G(\hat{\mathbf{z}}, \theta_G^*)$. However, achieving a perfect generator is practically difficult, and thus it is not feasible for the range of a GAN's generator to span all possible images that may arise from the distribution. As a result, the reconstructed image obtained using the formulation in Eq. (3.2) inherently contains *representation error*, defined as [28]

$$\rho_G(\mathbf{f}) := \min_{\mathbf{z}} \|G(\mathbf{z}) - \mathbf{f}\|_2^2. \tag{3.3}$$

The presence of representation error may lead to a lack of fidelity with respect to the observed measurements, and produce realistic but false structures in the reconstructed image. In order to mitigate this problem, Hussein *et al.* [5] proposed an image-adaptive GAN-based (IAGAN) reconstruction framework where the image is constrained to lie in the range of G while at the same time the trained generator's parameter weights are further tuned to enforce consistency with the observed measurement data. With the parameters of G now denoted by θ and initialized as θ_G^*, $\mathbf{z}$ and θ are jointly minimized:

$$\hat{\mathbf{z}}, \hat{\theta} = \underset{\mathbf{z},\theta}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}G(\mathbf{z}; \theta)\|_2^2, \tag{3.4}$$

with the reconstructed image as $\hat{\mathbf{f}} := G(\hat{\mathbf{z}}; \hat{\theta})$. The authors also proposed to initialize the latent vector $\mathbf{z}$ with the optimal solution obtained from approximately solving Eq. (3.2) to improve the stability of the optimization problem. Since the generator network is differentiable, any suitable stochastic gradient-based method may be applied to solve Eq. (3.4). Additionally, regularization on the generative model in the form of a sparsity-promoting penalty may be added to the IAGAN framework to mitigate artifacts resulting from severe data incompleteness and/or when the measurements contain a high level of noise. The optimization problem in Eq. (3.4) may be modified as follows:

$$\hat{\mathbf{z}}, \hat{\theta} = \underset{\mathbf{z},\theta}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}G(\mathbf{z}; \theta)\|_2^2 + \lambda \mathcal{R}(G(\mathbf{z}; \theta)), \tag{3.5}$$

with θ initialized as θ_G^*, where $\mathcal{R}(.)$ is a suitable sparsity-promoting penalty function and λ is a hyperparameter that controls the strength of regularization. The reconstructed image is $\hat{\mathbf{f}} := G(\hat{\mathbf{z}}; \hat{\theta})$. Inspired by [92], we propose $\mathcal{R}(.)$ to be the TV penalty, and we define the method represented by Eq. (3.5) as IAGAN-TV.

Recently, deep generative models have been designed which theoretically possess zero representation error, such as invertible neural networks (INNs) using normalizing flows [43, 87].

While the synthesis quality of INNs is inferior to state-of-the-art GANs such as ProGANs and training of INNs does not computationally scale well, they may be employed in a CSGM framework without adaptation of parameters due to the absence of representation error. Kelkar *et al.* [81] proposed an INN-based CSGM reconstruction method using latent space projection and TV regularization (INN Proj. TV), with the optimization problem stated as

$$\begin{aligned} \hat{\mathbf{z}} &= \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}G(\mathbf{z};\theta)\|_2^2 + \lambda \|G(\mathbf{z};\theta)\|_{TV} \\ s.t.\ \mathbf{z}_{1:N-k} &= \mathbf{0}, \end{aligned} \tag{3.6}$$

where the latent vector $\mathbf{z} \in \mathbb{R}^N$ and k is a hyperparameter which determines the sparsity of $\mathbf{z}$. The reconstructed image is $\hat{\mathbf{f}} := G(\hat{\mathbf{z}};\theta)$.

In [81], we compared the performance of the IAGAN-TV and INN Proj. TV methods in ill-posed image reconstruction problems which highlighted the trade-offs offered by each method.

3.5 Numerical Studies

Numerical studies were conducted to assess the effectiveness of the proposed method, especially in terms of recovering fine object features. The reduction in the appearance of realistic but false features and oversmoothing artifacts was studied. The impact of early stopping and the addition of a TV penalty term in the IAGAN method were compared in terms of their regularization abilities. Our studies were divided into two parts - (1) reconstruction from stylized, simulated undersampled single-coil MRI measurements (henceforth referred to as the simulation study), and (2) reconstruction from emulated experimental undersampled single-coil MRI measurements (henceforth referred to as the emulated experimental study). For the simulation study, in-distribution images, i.e. the images that come from the same distribution as the training dataset, as well as out-of-distribution images were considered. The proposed method was compared to traditional sparsity-based, as well as the INN-based reconstruction method. For the comparisons, traditional image quality metrics such as the root mean squared error (RMSE) defined as the discrete error norm $\|\hat{\mathbf{f}} - \mathbf{f}\|_2$, as well as structural similarity (SSIM) index [140] were utilized. Where applicable, bias-variance tradeoff calculations were carried out to assess the robustness of the algorithms.

3.5.1 Datasets and Imaging System

Simulation Study

The generative models were trained on single channel 2D MRI images of size 256 $\times$ 256. The following two datasets were employed for training, as well as evaluation in the case of in-distribution images:

- *fastMRI knee dataset*: 15000 non-fat suppressed, proton density (PD) weighted coronal knee images from the NYU fastMRI dataset.
- *fastMRI brain dataset*: 12000 T1-weighted axial adult brain images from the NYU fastMRI dataset.

Randomly-selected samples of images from the training dataset as well as images generated by the ProGAN and INN are shown in Fig. 3.2. The Fréchet Inception Distance (FID) scores [57, 94] for the INN and the ProGAN were computed for an initial assessment of the image synthesis performance of each generative model. A lower FID score generally indicates superior image synthesis quality. The official Python implementation [59] was employed to compute the FID scores shown in Table 3.1. It can be observed that the FID score for the ProGAN is significantly lower compared to the INN for both knee and brain datasets, which validates the superior image synthesis performance of ProGAN compared to the invertible generative model. For evaluation of reconstruction performance on out-of-distribution images, images from a pediatric epilepsy resection MRI dataset containing anomalies [97, 96] were used, along with generative models trained on the fastMRI brain dataset. Evaluating the robustness of a reconstruction method on out-of-distribution images is relevant because (i) in practice, test images may not exactly correspond to the training data distribution and a practitioner might be oblivious to these small differences, and (ii) it is of interest to examine the scenario of *transfer compressed sensing*, where learned priors from one dataset are employed to recover images from a closely related but different test distribution, due to the unavailability of sufficient data to learn the priors (for example, data including rare anomalies.)

Simulated undersampled single-coil MR measurements were employed as a proxy for experimental MRI k-space measurements. Variable density Poisson disc sampling patterns shown

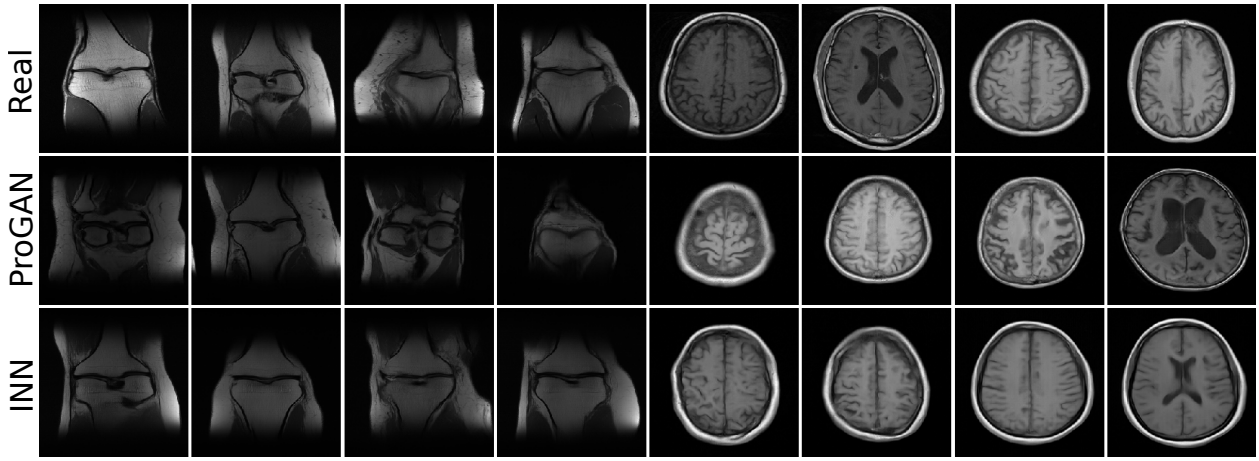


Figure 3.2: Samples from the training dataset and images generated by the ProGAN and INN. All images are displayed in the grayscale range of $[0, 1]$.

in Fig. 3.3 corresponding to $R = 8$ and $R = 20$ undersampling ratios were utilized, which retain low frequencies and randomly sample higher frequencies with a variable density [127, 95].

Emulated Experimental Study

Data for training the generative models were prepared in the following way. The fastMRI initiative database provides *emulated single-coil k-space measurements*, each of which is a complex-valued linear combination of responses from multiple coils of raw multi-coil k-space data [131]. These fully sampled k-space measurements were used to generate complex-valued images via the inverse fast Fourier transform (IFFT). They were divided into a training dataset for training the generative models, and a test dataset. The complex-valued images were converted to two-channel real images for training, and generative models with two-channel output were trained. Image reconstruction was performed directly from retrospectively undersampled emulated single-coil measurements, and the image estimates were compared with the reconstructions from the corresponding fully sampled k-space measurements in the test dataset. A Cartesian random undersampling mask with $R = 4$ was used for the retrospective undersampling.

For evaluating the reconstruction performance on the above-described image types, a validation image and a test dataset was used for each of the image types. These images were kept unseen during training. The regularization parameters for all the reconstruction methods

were tuned on the respective validation image for each image type, and the parameter setting showing the best RMSE performance was chosen. The tuned parameters were used in the reconstruction of images from the unseen test datasets. Performance metrics and their statistical significance were reported on these test datasets for all the image types.

3.5.2 Network Architecture and Training

The progressive GANs (ProGANs) were trained using the original implementation provided by Karras *et al.* [73]. The default settings for the training parameters were employed in this study, and the default latent space dimensionality of 512 was maintained. The training was performed on a system with an Intel Xeon E5-2620v4 CPU @ 2.1 GHz and 4 NVIDIA TITAN X GPUs. The algorithms are implemented in Python 3.6/Tensorflow 1.14. The employed INN architecture was adapted from Kingma and Dhariwal [87] and utilized the same settings for the training parameters as in the official implementation by Kingma and Dhariwal [1]. The INN was trained on a system with a 2x 20-core IBM POWER9 Central Processing Unit (CPU) @ 2.4GHz, and 4 16 GB NVIDIA V100 Graphical Processing units (GPUs) for a period of about 2.5 days [85].

Dataset	ProGAN	INN
Knee	22.72	75.06
Brain	10.67	82.41

Table 3.1: FID scores of the generative models. A lower FID is correlated with improved visual quality of generated images [57].

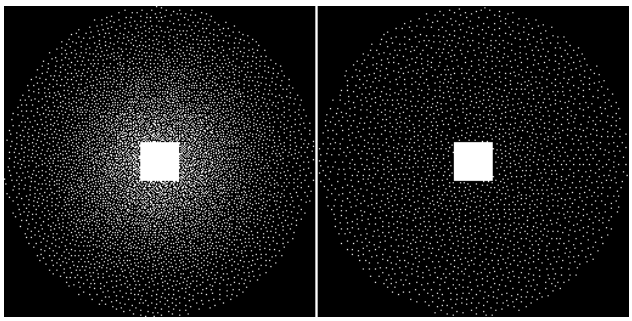


Figure 3.3: 8-fold (left) and 20-fold (right) undersampling masks

3.5.3 Image Reconstruction

Next, the performances of the following reconstruction methods were qualitatively and quantitatively compared - (i) penalized least squares with TV regularization (PLS-TV) solved with the fast iterative shrinkage and thresholding algorithm (FISTA) [20], (ii) the method proposed by Bora *et al.* [29], i.e. the problem stated in Eq. (3.2), with a ProGAN [71] trained

as described in Sec. 3.4 as the generative model (henceforth referred to as CSGM-GAN), (iii) Image-adaptive GAN-based reconstruction with TV regularization described in Eq. (3.5) (IAGAN-TV), and (iv) INN-based reconstruction using latent space projection and TV regularization, described in Eq. (3.6) (INN Proj. TV). For the simulation study, coronal knee and axial brain images were reconstructed from simulated measurements corresponding to the $R = 8$ and $R = 20$ undersampling ratios for this comparison. This was done with noiseless measurements, as well as measurements with i.i.d Gaussian noise with 20 dB per-pixel SNR.

Next, the approaches described above were employed to reconstruct anomalous pediatric brain images from 8-fold simulated undersampled measurements with 20 dB SNR [96]. The generative models used to reconstruct the pediatric brain image were trained on axial adult brain images from the previously described NYU fastMRI dataset.

Finally, for the emulated experimental study, image reconstruction was performed from four-fold retrospectively undersampled emulated single-coil measurements. The image estimates were compared to IFFT-based reconstructions from fully sampled measurements.

3.6 Results

3.6.1 Simulation Study: Reconstruction from Undersampled Measurements with 20 dB Measurement SNR

Figures 3.4 and 3.5 display reconstructed images of a coronal knee test image from 8-fold and 20-fold noisy undersampled measurements respectively. One key observation is that for 8-fold subsampling, all algorithms except for CSGM-GAN performed well, in terms of RMSE and SSIM. This was because the 8-fold variable density Poisson disc undersampling mask is designed in order to keep the low frequency information intact, and randomly sample only the high frequency information with a variable density. It should be noted that due to the representation error, the CSGM-GAN reconstruction retained highly realistic features, some of which, were false. Further, it should be noted that the IAGAN-TV and the INN-based method seem to have performed the best in terms of recovering the finer features of the

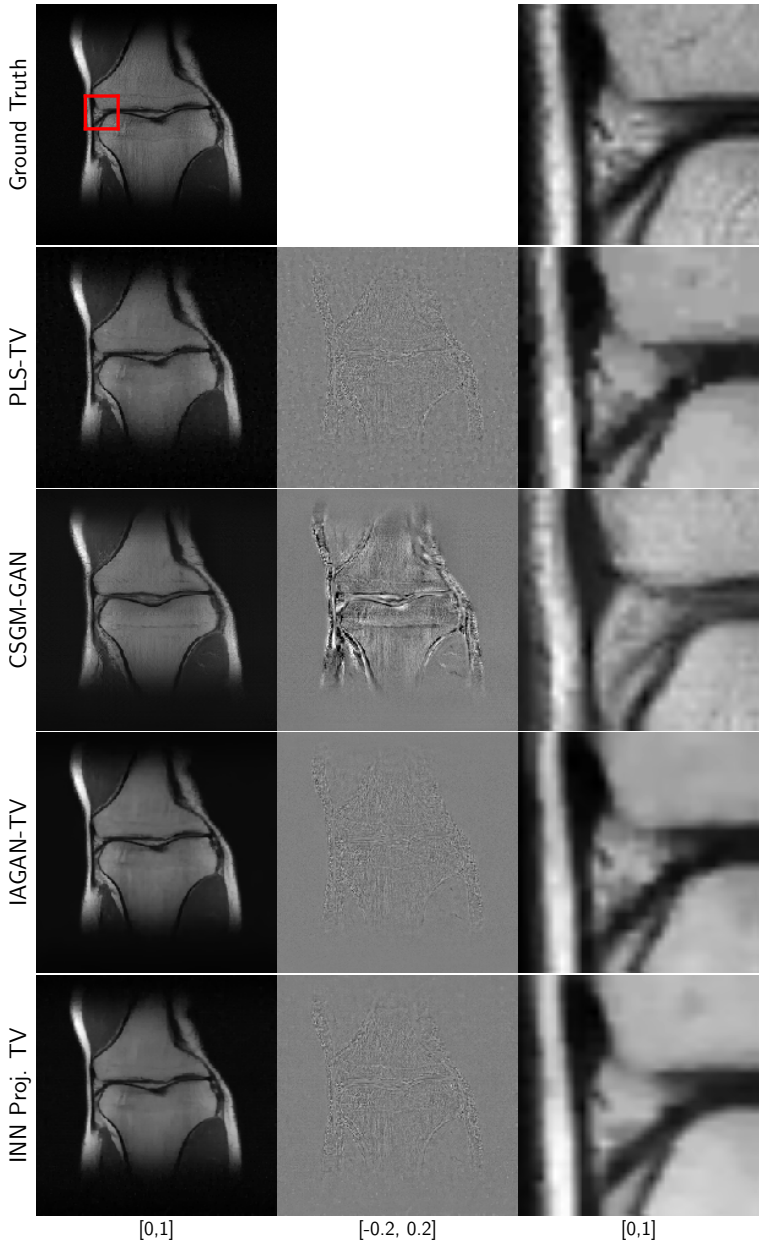


Figure 3.4: Ground truth, difference plots and reconstruction results for a coronal PD weighted knee image without fat suppression, with 8-fold undersampling and 20 dB measurement SNR. The RMSE and SSIM values are displayed in Table 3.2.

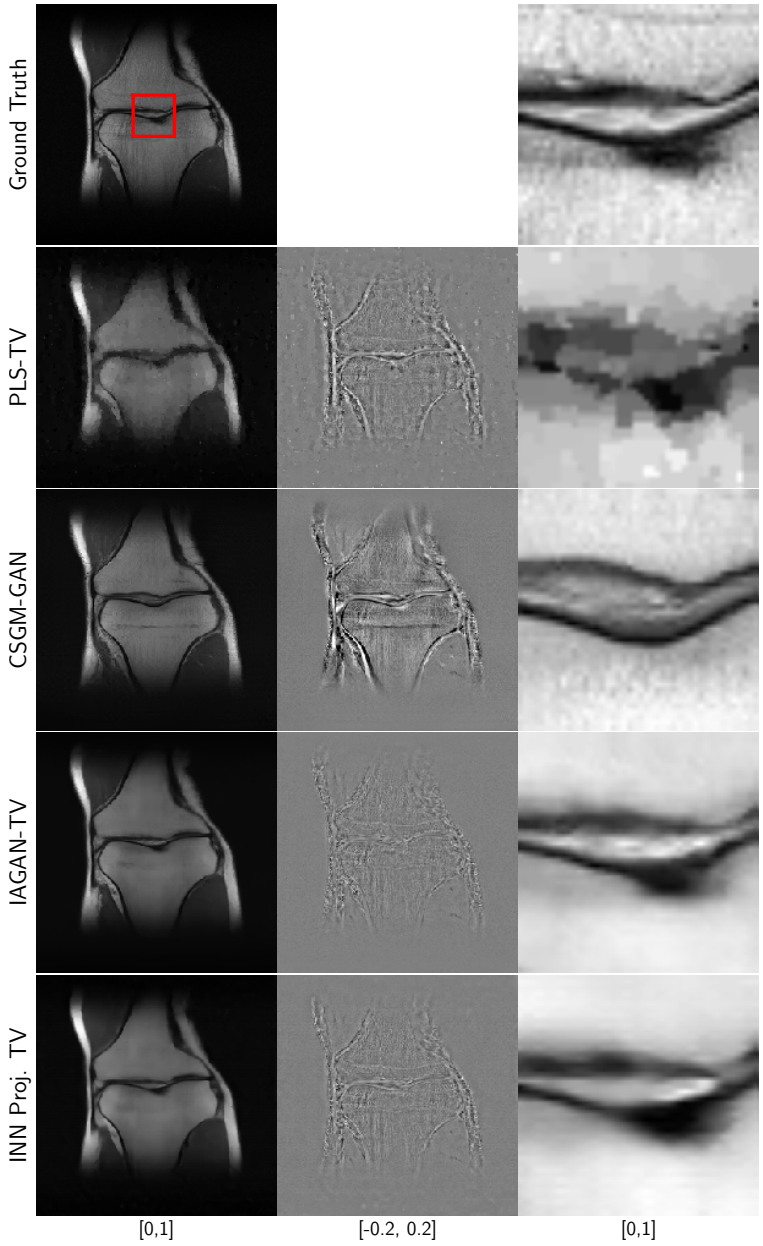


Figure 3.5: Ground truth, difference plots and reconstruction results for a coronal PD weighted knee image without fat suppression, with 20-fold undersampling and 20 dB measurement SNR. The RMSE and SSIM values are displayed in Table 3.2.

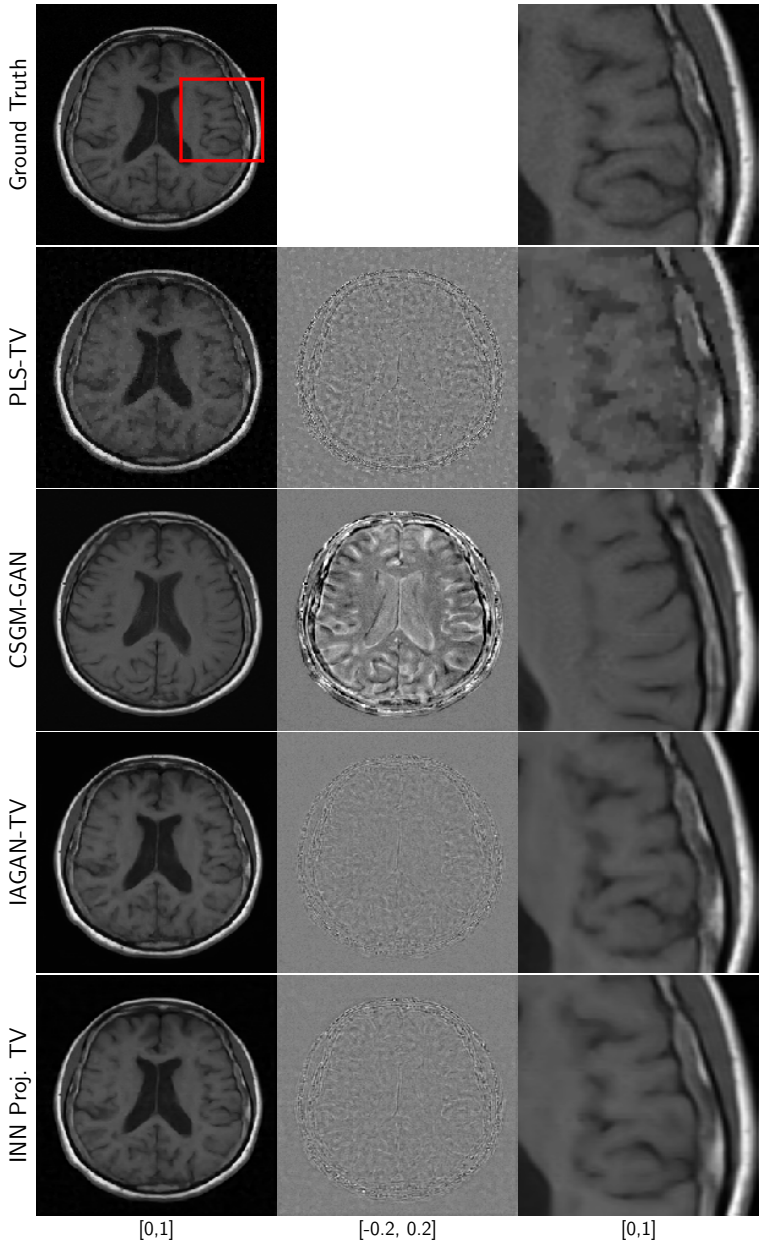


Figure 3.6: Ground truth, difference plots and reconstruction results for an axial T1 weighted brain image, with 8-fold undersampling and 20 dB measurement SNR. The RMSE and SSIM values are displayed in Table 3.2.

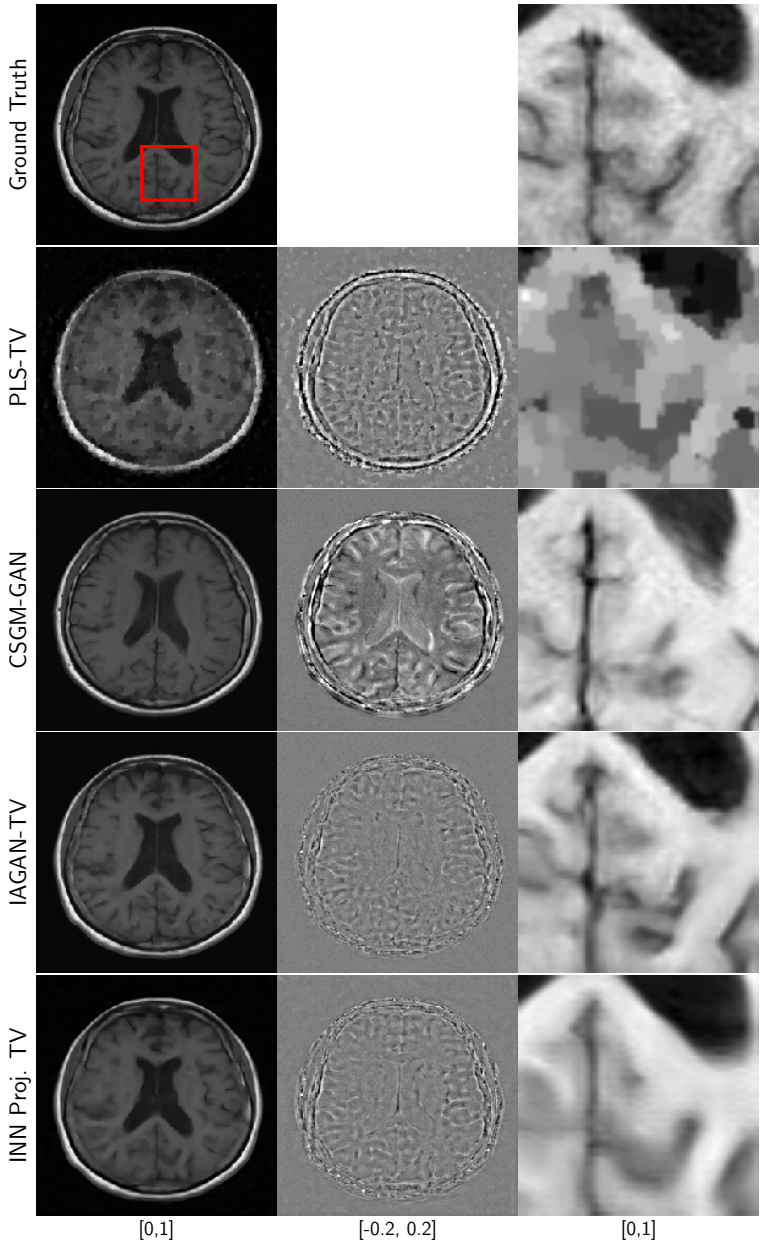


Figure 3.7: Ground truth, difference plots and reconstruction results for an axial T1 weighted brain image, with 20-fold undersampling and 20 dB measurement SNR. The RMSE and SSIM values are displayed in Table 3.2.

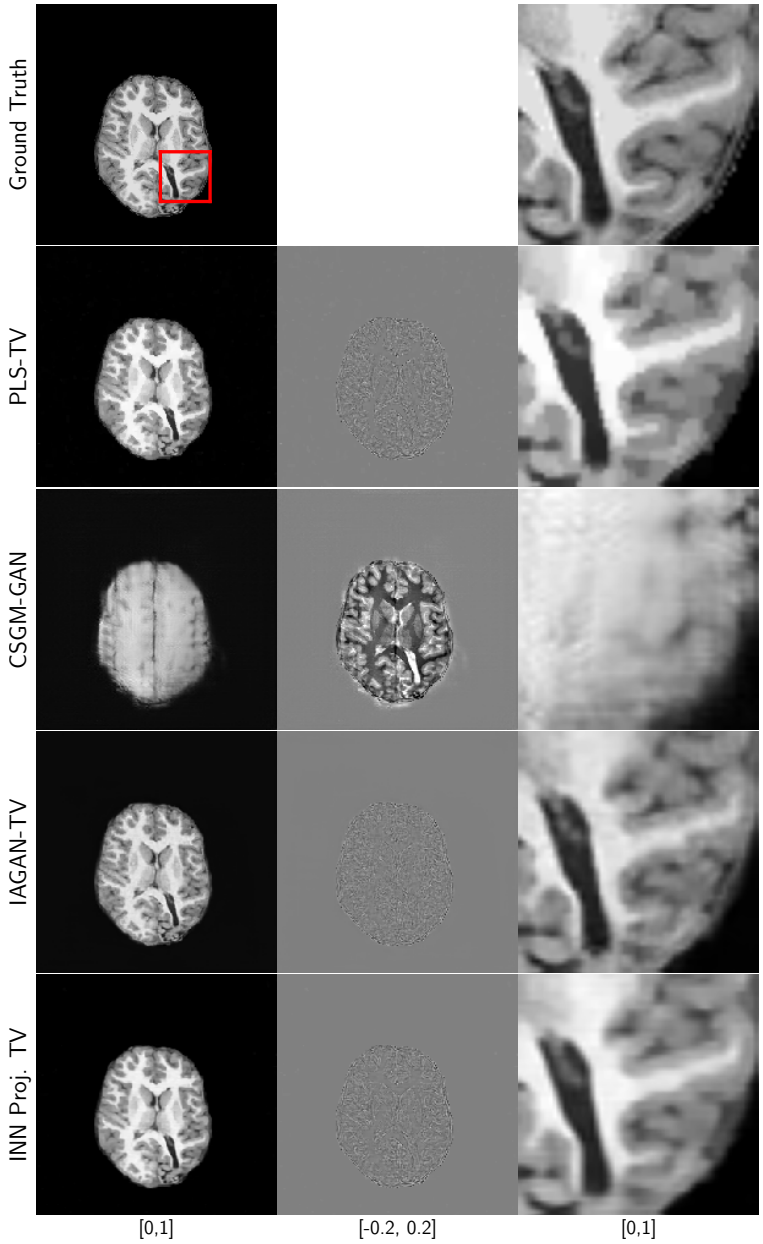


Figure 3.8: Ground truth, difference plots and reconstruction results for an axial T1 weighted pediatric brain image with anomaly, with 8-fold undersampling and 20 dB measurement SNR. The CSGM-GAN method is unable to adapt to the domain shift and produces large errors, while the IAGAN-TV method demonstrates good generalization performance and produces much lower errors.

image. As shown in Fig. 3.5, for 20-fold undersampling, the PLS-TV reconstruction has characteristic smoothing artifacts due to its reliance on only TV regularization.

Similar observations can be made for the results of reconstruction of an axial brain image from 8-fold and 20-fold undersampled measurements, as shown in Fig. 3.6 and Fig. 3.7, respectively. In addition, it should be noted that for the 8-fold undersampling case, some of the finer features, such as the folds in the brain, are difficult to recover using PLS-TV, but were successfully recovered with both IAGAN-TV and the INN-based reconstruction. For the 20-fold undersampling case, all the methods face challenges in recovering finer features such as the folds of the brain, and may produce oversmoothened features or even realistic hallucinations. Finally, the results for the reconstruction of the pediatric brain image are shown in Fig. 3.8. Here, the out-of-distribution image was accurately recovered by INN Proj. TV and the proposed method, but not in the case of CSGM-GAN. Here, the poor performance of CSGM-GAN could also be due to domain shifts unrelated to anatomical features.

3.6.2 Impact of Additional TV Regularization in IAGAN

In the IAGAN method, early stopping was employed as a means to prevent overfitting to measurement noise. However, determining the early stopping iteration is often a difficult task due to semi-convergence in overparameterized optimization problems [6]. We demonstrate this phenomenon using the simulation study with 8-fold undersampled measurements from the knee object in Fig. 3.5 with a noise level of SNR = 20 dB. With the TV regularization parameter $\lambda = 0$ in Eq. (3.5), the reconstructed images at different iterations are visualized in Fig. 3.9. It can be observed that the images progressively become noisy, and the RMSE and SSIM values decrease beyond iteration 5000. With TV penalty parameter $\lambda = 300$, however, the images do not become progressively noisy and the RMSE and SSIM metrics remain the same beyond iteration 5000, as shown in Fig. 3.10. The convergence properties of the IAGAN method were investigated with and without TV penalty. The cost function and the RMSE and SSIM metrics across iterations were plotted with different values of the TV regularization parameter λ. With $\lambda = 0$, the cost function decreases with iterations, but the RMSE and SSIM metrics reach their optimal values around iteration 2000 and then start to decrease, demonstrating semi-convergence behavior. As λ is increased to 10, the semi-convergence behavior still exists but is less pronounced. For higher values of λ such as $\lambda = 300$ and $\lambda = 5000$, the convergence behavior improves significantly, demonstrating

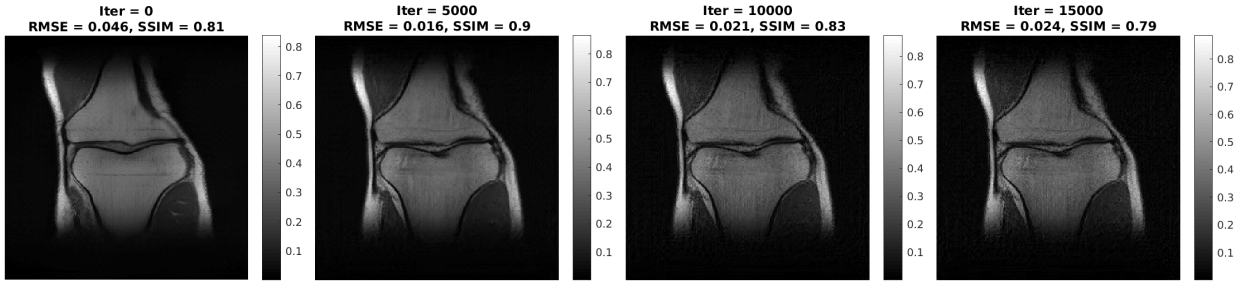


Figure 3.9: Reconstructed images at different iterations with TV penalty parameter $\lambda = 0$ in the IAGAN-TV method. The images progressively become noisy as the number of iterations increases. The RMSE and SSIM values decrease beyond iteration 5000.

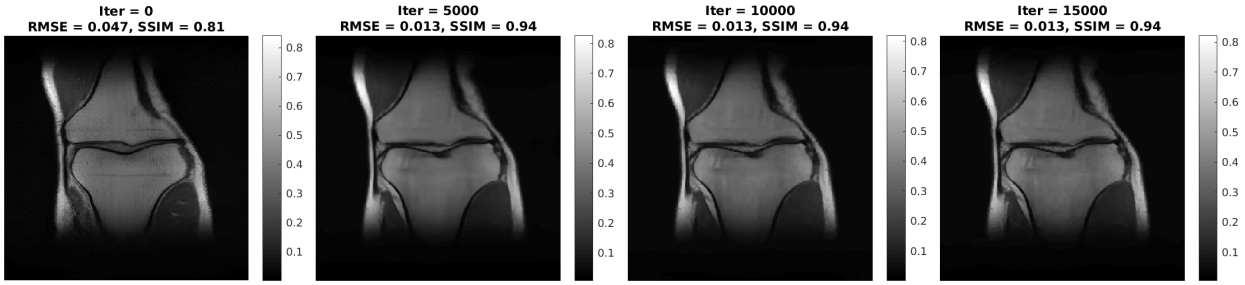
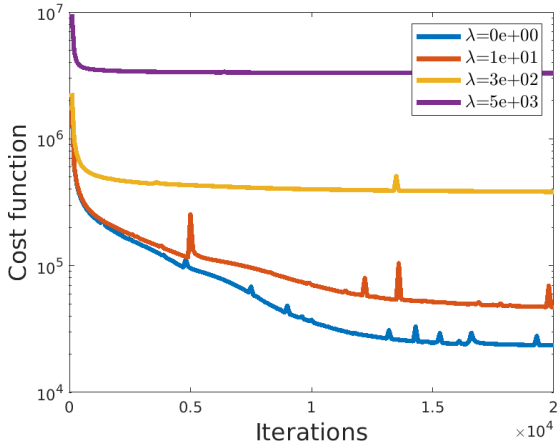


Figure 3.10: Reconstructed images at different iterations with TV penalty parameter $\lambda = 300$ in the IAGAN-TV method. The images do not progressively become noisy and semi-convergence behavior is not observed. The RMSE and SSIM values remain unchanged beyond iteration 5000.

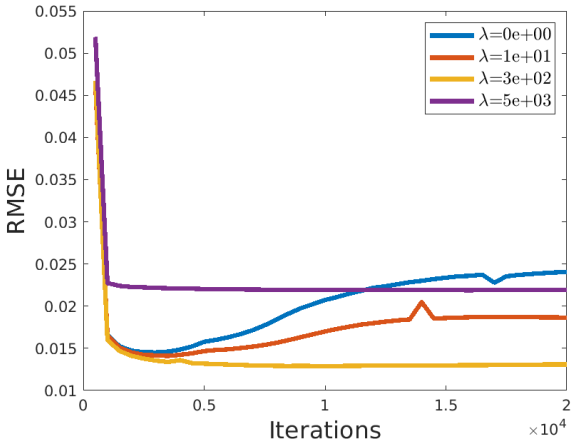
that inclusion of the TV penalty leads to more stable behavior of the IAGAN method. Additionally, improved image quality can be achieved in terms of RMSE and SSIM metrics using TV regularization without early stopping as compared to early stopping with $\lambda = 0$. The optimal value of λ is chosen in our studies by performing grid search over a held-out validation dataset.

3.6.3 Emulated Experimental Study

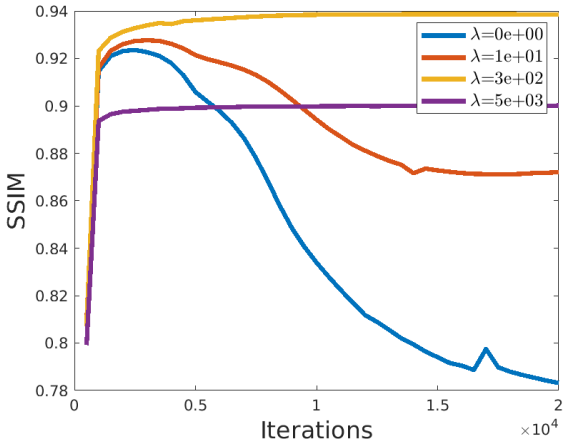
The absolute value of the image reconstructed by use of the IAGAN-TV method from four-fold retrospectively undersampled emulated single-coil measurements is shown in Fig. 3.12, along with the IFFT-based reconstruction from fully sampled data and difference plots. Here, we see that the proposed method demonstrates superior performance compared to PLS-TV, which oversmooths the image, and CSGM-GAN, which introduces distinct false structures.



(a) Loss curve



(b) RMSE curve



(c) SSIM curve

Figure 3.11: Plots of (a) the cost function, (b) RMSE and (c) SSIM vs. iterations with TV regularization parameter $\lambda = 0, 10, 300, 5000$. The optimal values of RMSE and SSIM are achieved with $\lambda = 300$. Semi-convergence behavior is observed with $\lambda = 0$ and $\lambda = 300$, and the convergence behavior improves with increase in λ.

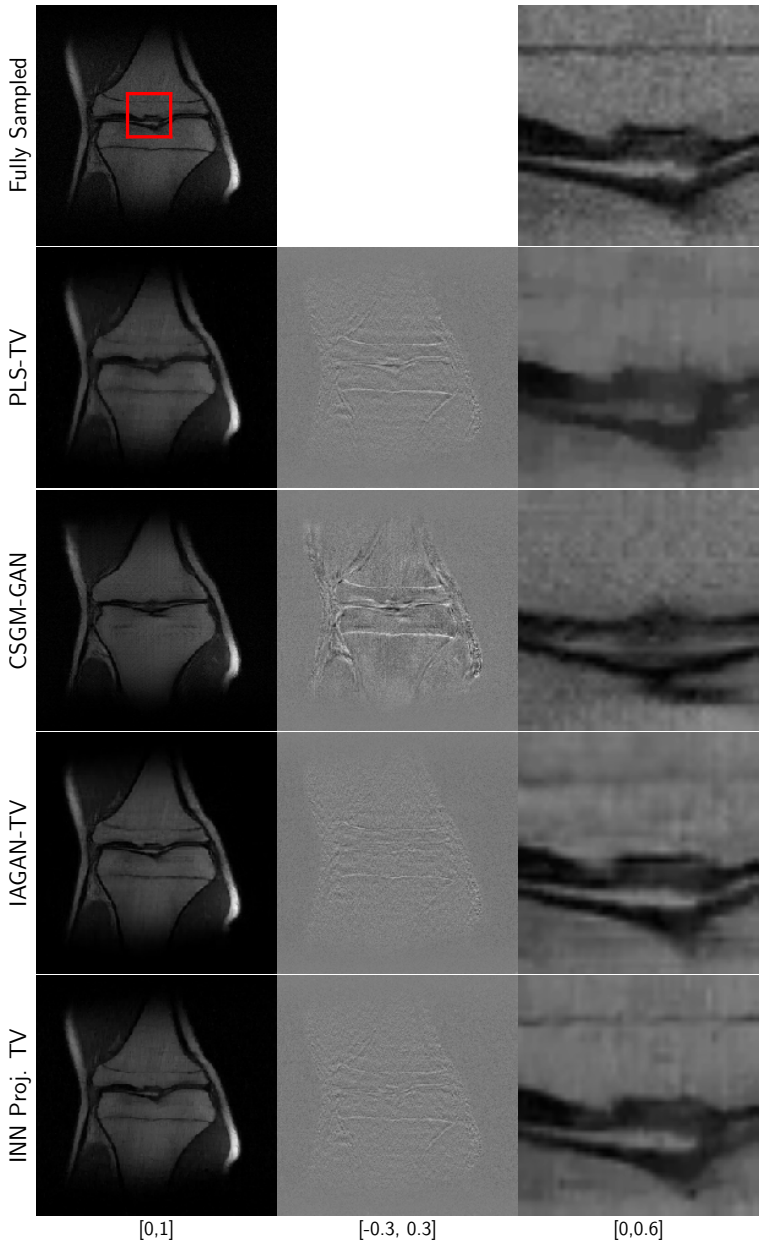


Figure 3.12: The absolute value of coronal PD weighted knee images reconstructed from emulated single-coil measurements with Cartesian four-fold retrospective undersampling.

3.6.4 Root Mean Square Error and Structural Similarity

Root mean-squared error (RMSE) and structural similarity (SSIM) index values over an ensemble of 50 test images from the in-distribution study were calculated. Ensemble mean and standard deviation of these values are displayed in Tables 3.2. It can be noted that,

Table 3.2: Comparison of RMSE and SSIM for different algorithms for undersampled data with 20 dB measurement SNR, for the simulation study, computed on an ensemble of 50 images. The values outside the parentheses denote the ensemble mean values of the metric, where as the values inside the parentheses denote the standard deviation (SD) of the metric.

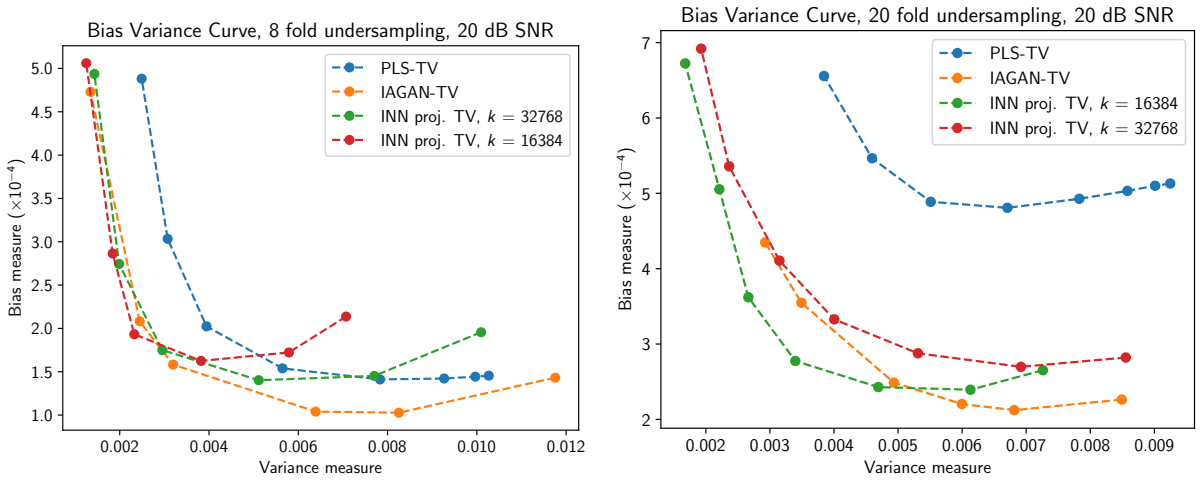
Algorithm	Knee (in dist.) 8x		Knee (in dist.) 20x		Brain (in dist.) 8x		Brain (in dist.) 20x	
	RMSE mean (RMSE SD)	SSIM mean (SSIM SD)	RMSE mean (RMSE SD)	SSIM mean (SSIM SD)	RMSE mean (RMSE SD)	SSIM mean (SSIM SD)	RMSE mean (RMSE SD)	SSIM mean (SSIM SD)
PLS-TV	0.0122 (0.0033)	0.9736 (0.0108)	0.0178 (0.0050)	0.9556 (0.0172)	0.0228 (0.0033)	0.9609 (0.0093)	0.0473 (0.0098)	0.8798 (0.0337)
CSGM-GAN	0.0381 (0.0157)	0.8808 (0.0485)	0.0389 (0.0154)	0.8753 (0.0470)	0.0721 (0.0318)	0.8174 (0.0633)	0.0725 (0.0249)	0.8153 (0.0598)
IAGAN-TV	**0.0099** (0.0026)	**0.9844** (0.0064)	**0.0140** (0.0041)	**0.9705** (0.0135)	**0.0148** (0.0024)	**0.9794** (0.0061)	**0.0246** (0.0043)	**0.9483** (0.0146)
INN Proj. TV	0.0102 (0.0027)	0.9829 (0.0070)	0.0147 (0.0042)	0.9678 (0.0137)	0.0163 (0.0028)	0.9723 (0.0086)	0.0262 (0.0049)	0.9414 (0.0177)

across several categories, the performance of the INN Proj. TV and the proposed method was comparable and the best among all the methods compared, although the proposed method outperformed INN Proj. TV for some of the image categories by a small margin.

The statistical significance of the differences between the reconstruction methods was tested using the one way repeated measures ANOVA test, followed by post-hoc paired samples t-tests between pairs of algorithms, with the Bonferroni correction. Since the metrics obtained from CSGM-GAN violated some of the assumptions of the ANOVA test, it was left out of the statistical significance study. It was observed that INN Proj. TV and the proposed approach are both statistically significantly better than PLS-TV (with p-value $< 10^{-17}$ for the in-distribution images, p-value $< 10^{-12}$ for the out-of-distribution images, and p-value $< 10^{-6}$ for the emulated experimental study). For the in-distribution images in the simulation study, there is a small but statistically significant difference between the performance of IAGAN-TV and INN Proj. TV, with the proposed approach performing better (with p-value $< 10^{-5}$).

3.6.5 Bias-Variance Tradeoff

Although the evaluation of perceptual quality and quantitative evaluation in terms of RMSE and SSIM indicate the superiority of the IAGAN-based reconstruction method as compared to more traditional approaches, a task-based assessment of reconstruction algorithms is necessary to determine the superiority of one reconstruction algorithm to the other [19]. However, such a detailed task-based assessment of generative model-based reconstruction algorithms is a substantial task in itself, and remains a topic for future study. Here, an analysis of the



(a) (b)

Figure 3.13: Bias-variance tradeoff analysis for (a) 8-fold and (b) 20-fold undersampling comparing PLS-TV, IAGAN-TV and INN Proj. TV, while sweeping the TV regularization parameter for each method.

bias-variance trade-off is provided for the PLS-TV, IAGAN-TV and INN Proj. TV methods by performing a sweep over the TV regularization parameter for each case.

Bias-variance analysis was performed on images reconstructed from simulated measurements corresponding to both 8-fold and 20-fold undersampling patterns, with 20 dB measurement SNR. The ground truth used for this study was an image from the fastMRI knee dataset. Stylized, simulated undersampled single-coil MRI measurements were used. A dataset of reconstructed images $\{\hat{\mathbf{f}}^{(i)}\}_{i=1}^{q}$ from measurements with $q = 100$ independent noise realizations was considered for every regularization setting. The bias $\mathbf{b}$ and the variance σ_i of a pixel i were calculated as:

$$\mathbf{b} = \frac{1}{q}\sum_{i=1}^{q}\hat{\mathbf{f}}^{(i)} - \mathbf{f} \tag{3.7}$$

$$\sigma_j^2 = \frac{1}{q-1}\left(\hat{\mathbf{f}}_j^{(i)} - \frac{1}{q}\sum_{i=1}^{q}\hat{\mathbf{f}}_j^{(i)}\right)^2, \tag{3.8}$$

where $\mathbf{f}$ is the ground truth image. As a summary measure, the average squared bias $\frac{1}{N}\|\mathbf{b}\|_2^2$ versus the average variance $\frac{1}{N}\sum_{j=1}^{N}\sigma_j^2$ was plotted. Figures 3.13a and 3.13b show the bias-variance curves for 8 and 20-fold undersampling, respectively.

As can be seen, the bias and variance curves for the IAGAN-TV method lie below the curves for PLS-TV, which is indicative of superior performance over a range of regularization values. This also indicates that, while the transition from an over-smoothed image to a noisy image is such that intermediate images could be both noisy and oversmoothed, this trade-off is better for the proposed reconstruction approach. It should be noted that for the 8-fold undersampling case in Fig. 3.13a, the bias-variance curve for the proposed method consistently remains below PLS-TV, while the curve for INN-based method may lie above PLS-TV in certain ranges of the corresponding TV regularization parameter.

3.7 Summary

This study demonstrated the use of an image-adaptive GAN-based algorithm (IAGAN-TV) that employed a state-of-the-art GAN to regularize ill-posed image reconstruction problems. Reconstructed images from undersampled Fourier measurements illustrate that using learned GAN-based priors in an image-adaptive framework can mitigate the impact of representation error in GANs and recover fine features in the image which may be oversmoothed by traditional sparsity-based reconstruction methods. It was observed that the image-adaptive framework can successfully recover fine image structures with experimental data. It was demonstrated how image quality could be enhanced and the need for early stopping could be mitigated by adding the sparsity-promoting TV penalty in the IAGAN framework. The stability and generalization properties of the IAGAN-TV method were established using bias-variance analysis and validation studies with out-of-distribution data. Additionally, the performance of the IAGAN-TV method was compared with a generative model-constrained method that employed an invertible neural network (INN) with no representation error. It was observed that the IAGAN-TV method and the INN-based reconstruction method had similar performance, with the former having the additional advantage that state-of-the-art GANs such as ProGANs possess superior image synthesis quality and are scalable.

There remain important topics for future investigation. In our studies, the ProGAN was employed to learn the object distribution by training on reconstructed images from fully-sampled measurements. However, if measurement noise is high, the reconstructed images are also noisy, leading to poor image synthesis performance by ProGANs [150]. In such cases, if the imaging operator and measurement noise distribution are known approximately,

progressively-growing AmbientGANs [30, 149] can be utilized to learn object distributions directly from the measurements with high image synthesis performance.

It will be important to investigate the performance of the IAGAN-TV method with evolving GAN architectures such as BigGAN [32] and StyleGAN [76, 78] that promote lower representation error. The implementation of the IAGAN-TV method can also be extended to three-dimensional imaging systems by using GANs trained to synthesize three-dimensional images [61]. Furthermore, since the image-adaptive framework is general, it can be extended to ill-posed inverse problems in other tomographic imaging modalities, such as X-ray CT and photoacoustic tomography (PACT) [138]. Finally, for a comprehensive evaluation of the IAGAN-TV method, task-based assessments need to be performed that take into account all physical and statistical factors.

Chapter 4

On Hallucinations in Tomographic Image Reconstruction

4.1 Overview

As described in Chapter 2 and demonstrated using the IAGAN-TV algorithm in Chapter 3, deep learning-based methods have inspired a new wave of reconstruction methods that implicitly or explicitly learn the prior distribution from a set of training images in order to regularize the reconstruction problem. However, certain supervised learning-based methods have also raised concerns regarding their robustness [64, 49, 11, 90] and their ability to generalize to measurements that may lie outside the distribution of the training data [11, 13, 81]. This is particularly relevant in the field of medical imaging where novel abnormalities can be present in the observed measurement data that may not be encountered even with a large training dataset. Moreover, simulation studies have shown that deep learning-based reconstruction methods are inherently unstable, i.e. small perturbations in the measurement may produce large differences in the reconstructed image [49, 11].

The potential lack of generalization of deep learning-based reconstruction methods as well as their innate unstable nature may cause false structures to appear in the reconstructed image that are absent in the object being imaged. These false structures may arise due to the reconstruction method incorrectly estimating parts of the object that either did not contribute to the observed measurement data or cannot be recovered in a stable manner, a phenomenon that can be termed as *hallucination*. The presence of such false structures in reconstructed images can possibly lead to an incorrect medical diagnosis. Hence, there is an urgent need to investigate the nature and impact of false structures arising out of hallucinations from deep learning-based reconstruction methods for tomographic imaging.

The topic of image hallucinations has previously been studied within the context of image super-resolution [14, 93, 139, 45]. In image super-resolution, the term hallucination generally refers to high-frequency features that are introduced into the high-resolution image but do not exist in the measured low-resolution image. Hallucinations can also be realized in more general inverse problems such as image reconstruction. In such cases, the structure of the imaging operator null space is generally more complicated and the hallucinations may not be confined to high-frequency structures [19]. However, a formal definition of hallucinations within the context of such inverse problems has not been reported.

This study proposes a way to mathematically formalize the concept of hallucinations for general linear imaging systems that is consistent with both the mathematical notion of a hallucination in image super-resolution and the intuitive notion of hallucinations as "artifacts or incorrect features that occur due to the prior that cannot be produced from the measurements". In addition, the notion of a *task-informed* or *specific* hallucination map is introduced. Through preliminary numerical studies, the behavior of different reconstruction methods under the proposed formalism is illustrated. It is shown that, in certain cases, traditional error maps are insufficient for visualizing and detecting specific hallucinations.

4.2 Definition of Hallucination Maps

When comparing or evaluating image reconstruction methods, it may be useful to visualize and quantify false structures that cannot be stably reconstructed from the measurements. Such structures have been colloquially referred to as being 'hallucinated' and are attributable to use of an imperfect reconstruction prior. Error maps that display the difference between the reconstructed image estimate and the true object are commonly employed to assess reconstruction errors. Artifacts revealed by error maps encompass a broad range of deviations that can appear in a reconstructed image with respect to its depiction of the object function being imaged. For example, incorrect modeling of the system matrix $\mathbf{H}$ or measurement noise can lead to artifacts. Consequently, as demonstrated in Fig. 4.1, it may not be possible to isolate and label the artifacts attributable to the reconstruction prior from the error map alone. A possible way to circumvent this is to compute separate error maps for the null and measurement components of the reconstructed image estimate. However, precise definitions for hallucinations in these sub-spaces have been lacking. Instead, such artifacts

that are attributable to the prior are better seen and can be labeled as such once the true object and its reconstructed estimate have been decomposed into the measurement and null components. This can be better understood when the error map $\hat{\mathbf{f}}^{EM}$ is expressed as

$$\begin{aligned}\hat{\mathbf{f}}^{EM} &= \hat{\mathbf{f}} - \mathbf{f} \\ &= (\hat{\mathbf{f}}_{meas} - \mathbf{f}_{meas}) + (\hat{\mathbf{f}}_{null} - \mathbf{f}_{null}) \\ &= (\hat{\mathbf{f}}_{meas} - \hat{\mathbf{f}}_{tp}) + (\hat{\mathbf{f}}_{null} - \mathbf{f}_{null}) + (\hat{\mathbf{f}}_{tp} - \mathbf{f}_{meas}).\end{aligned} \tag{4.1}$$

The first term in Eq. (4.1) indicates the difference introduced due to the imposed prior in the reconstruction method between $\hat{\mathbf{f}}_{meas}$ and $\hat{\mathbf{f}}_{tp}$, which is the stable estimate of $\mathbf{f}$ that lies in $\mathcal{N}_P^{\perp}(\mathbf{H})$. The second term in Eq. (4.1) refers to the error that arises due to an inaccurate estimation of the generalized null component of $\mathbf{f}$ due to the imposed prior. On the other hand, the term $(\hat{\mathbf{f}}_{tp} - \mathbf{f}_{meas})$ in Eq. (4.1) does not involve any priors and describes the error that arises due to model mismatch and measurement noise. Thus, it can be observed that the error map contains different types of errors in the generalized measurement and null space of $\mathbf{H}$, and that may or may not be affected by the prior imposed in a reconstruction method. Consequently, in order to comprehensively describe false structures that arise only due to an incorrect prior, it is essential to decompose $\hat{\mathbf{f}}$ and $\mathbf{f}$ into their generalized measurement and null components and define error measures accordingly. There are no established methods for defining such false structures from the perspective of the generalized measurement space and null space of the imaging operator.

In order to visualize and quantify hallucinations in tomographic images, measurement and null space hallucination maps are formally defined below. The proposed definitions are general and can be applied to analyze hallucinations produced by any reconstruction method that seeks to invert a linear imaging model. The defined hallucination maps will permit isolation of image artifacts that cannot be stably reconstructed from the measurement data and are attributable to the implicit or explicit reconstruction prior.

4.2.1 Hallucination Map in the Generalized Measurement Space

Let $\hat{\mathbf{f}}$ denote the estimate of the coefficient vector $\mathbf{f}$ obtained from $\mathbf{g}$ by use of an image reconstruction method. It is desirable that the projection of $\hat{\mathbf{f}}$ onto the generalized measurement space $\mathcal{N}_P^{\perp}(\mathbf{H})$, i.e. $\hat{\mathbf{f}}_{meas}$, should be near the truncated pseudoinverse solution $\hat{\mathbf{f}}_{tp} \equiv \mathbf{H}_P^{+}\mathbf{g}$.

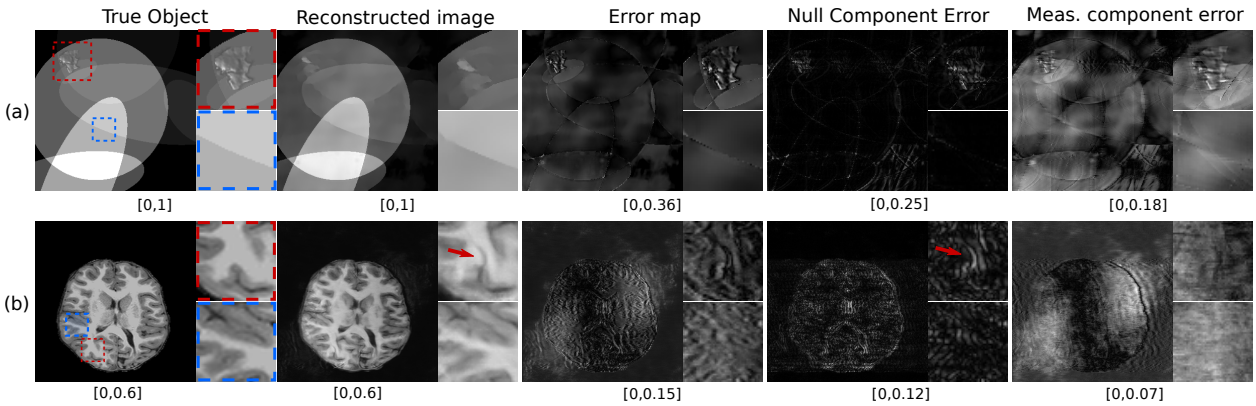


Figure 4.1: From left-to-right are examples of a true object, a reconstructed estimate of the object produced by use of a U-Net from tomographic measurements, the total error map, the error in the null component of the reconstructed object, and the error in the measurement component of the reconstructed object. The two rows correspond to different objects. In each case, the true object is outside the respective training data distribution of the U-Net and phase noise was added to the measurements prior to image reconstruction.

This would ensure that $\hat{\mathbf{f}}_{meas}$ is consistent with the estimate of $\mathbf{f}$ that can be stably recovered from $\mathbf{g}$. However, due to the imposed regularization in a reconstruction method, there may be discrepancies in $\hat{\mathbf{f}}_{meas}$ with respect to the stable estimate $\hat{\mathbf{f}}_{tp}$ in the generalized measurement space $\mathcal{N}_P^{\perp}(\mathbf{H})$. In order to quantify such differences, a hallucination map in the generalized measurement space is defined as follows.

Definition 4.2.1 (*Generalized measurement space hallucination map*)**.** As previously defined, let $\hat{\mathbf{f}}$ be an image estimate obtained by use of a reconstruction method and let $\hat{\mathbf{f}}_{tp}$ be the truncated pseudoinverse solution. The hallucination map in the measurement space is defined as,

$$\hat{\mathbf{f}}_{meas}^{HM} \equiv \hat{\mathbf{f}}_{meas} - \hat{\mathbf{f}}_{tp}. \tag{4.2}$$

It should be noted that the computation of the hallucination map in the generalized measurement space requires no knowledge of the true object and simply reveals errors in the measurement component of $\hat{\mathbf{f}}$ with respect to the stably computed estimate $\hat{\mathbf{f}}_{tp}$.

For use in cases where pixel expansion functions are not employed, it is useful to translate the definition of hallucination maps to the subspace of the object space $\mathbb{L}_2(\mathbb{R}^d)$ spanned by

a generic basis $\{\psi_n(\mathbf{r})\}_{i=1}^N$. By use of Eq. (2.2), the estimate of $f_a(\mathbf{r})$ can be represented as

$$\hat{f}_a(\mathbf{r}) = \sum_{n=1}^{N} [\hat{f}]_n \psi_n(\mathbf{r}). \tag{4.3}$$

The hallucination map $\hat{f}_{a,meas}^{HM}(\mathbf{r})$ can be defined in the space $\mathbb{L}_2(\mathbb{R}^d)$ as

$$\hat{f}_{a,meas}^{HM}(\mathbf{r}) \equiv \sum_{n=1}^{N} [\hat{f}_{meas}^{HM}]_n \psi_n(\mathbf{r}). \tag{4.4}$$

4.2.2 Hallucination Map in the Generalized Null Space

As reviewed in Sec. 2.2, to estimate the generalized null vector $\mathbf{f}_{null}$ from $\mathbf{g}$, reconstruction methods that impose appropriate priors are required. Hence, to accurately capture the effect of the prior on the reconstructed image, a definition of hallucinations must satisfy the following two desiderata:

- The definition must involve the assessment of how accurate the estimate $\hat{\mathbf{f}}_{null} = \mathcal{P}_{null}\hat{\mathbf{f}}$ is as compared to the true generalized null vector $\mathbf{f}_{null}$.
- Since no prior is used in obtaining $\hat{\mathbf{f}}_{tp}$, the definition must ensure that $\hat{\mathbf{f}}_{tp}$ does not have any null space hallucinations.

With these in mind, a hallucination map $\hat{\mathbf{f}}_{null}^{HM}$ in the generalized null space $\mathcal{N}_P(\mathbf{H})$ is defined as follows.

Definition 4.2.2 (*Generalized null space hallucination map*)**.** Consider a pixel-wise indicator function $\mathbb{1} : \mathbb{R}^N \to \mathbb{R}^N$ such that for any $\vartheta \in \mathbb{R}^N$,

$$[\mathbb{1}(\vartheta)]_n = \begin{cases} 1, & \text{if } [\vartheta]_n \neq 0 \\ 0, & \text{if } [\vartheta]_n = 0. \end{cases} \tag{4.5}$$

Then, the hallucination map $\mathbf{f}_{null}^{HM} \in \mathbb{E}^N$ can be defined as

$$\hat{\mathbf{f}}_{null}^{HM} \equiv \mathbb{1}(\hat{\mathbf{f}}_{null}) \odot (\hat{\mathbf{f}}_{null} - \mathbf{f}_{null}), \tag{4.6}$$

where $\odot$ denotes the Hadamard product or element-wise multiplication. Note that the indicator function in the definition ensures that $\hat{\mathbf{f}}_{tp}$ does not possess any null space hallucinations, since no prior was imposed.

It is important to highlight that, for the computation of the hallucination map in the generalized null space, one must have full knowledge of the generalized null component of the true object. This is in contrast to the hallucinations in the generalized measurement space, where the knowledge of the generalized measurement component of the true object is not required. This simply reflects that, according to the provided definitions, the generalized null space hallucination maps depict errors in the reconstructed null component of the object, while the generalized measurement space hallucination maps depict errors in the component of the object that can be stably reconstructed via a truncated pseudoinverse operator from the observed measurement data.

This difference in the two definitions is associated with the fact that $\hat{\mathbf{f}}_{tp}$ is close to $\mathbf{H}_P^+\mathbf{H}\mathbf{f}$ if the measurement noise is small in the sense of Eq. (2.5), and/or the model error is negligible. Hence, the proposed definition of $\hat{\mathbf{f}}_{meas}^{HM}$ is able to reveal the effect of the prior on the reconstructed generalized measurement space component, without requiring the true object. In this sense, there is no analog of a stably reconstructed component like $\hat{\mathbf{f}}_{tp}$ in the null space; hence invoking the true null component is necessary for defining $\hat{\mathbf{f}}_{null}^{HM}$. Note that due to our definition, $\hat{\mathbf{f}}_{meas}^{HM}$ may also be influenced by the different noise propagation characteristics of the methods employed to form $\hat{\mathbf{f}}_{tp}$ and $\hat{\mathbf{f}}$ and therefore may not solely quantify errors associated with the prior.

It should also be noted that the errors introduced by the prior in the measurement space can be remedied by adopting a reconstruction method that penalizes measurement space hallucinations without any prior knowledge of the object, e.g., via a data consistency constraint [11] or null space shuttle procedure [40]. Accordingly, for such constrained image reconstruction methods, analyzing hallucinations in the null space is critical towards understanding the effect of the prior on the image estimate.

Similar to the hallucination map in the generalized measurement space, the hallucination map $\hat{f}_{a,null}^{HM}(\mathbf{r})$ can be defined as

$$\hat{f}_{a,null}^{HM}(\mathbf{r}) \equiv \sum_{n=1}^{N} [\hat{f}_{null}^{HM}]_n \psi_n(\mathbf{r}). \tag{4.7}$$

According to the proposed definitions, the truncated pseudoinverse solution $\hat{\mathbf{f}}_{tp}$ has zero hallucination in both the generalized measurement space and null space. However, that does not necessarily imply that $\hat{\mathbf{f}}_{tp}$ is without artifacts, since $\hat{\mathbf{f}}_{tp}$ ignores $\mathbf{f}_{null}$ completely. The computation of $\hat{\mathbf{f}}_{tp}$ leads to the recovery of only $\mathbf{f}_{meas}$ that can be estimated stably. When other regularized reconstruction methods attempt to reduce artifacts by imposing priors to estimate $\mathbf{f}_{null}$, a trade-off is made between the estimation of $\mathbf{f}_{meas}$ and $\mathbf{f}_{null}$ that can potentially lead to hallucinations in the generalized measurement space and null space.

4.2.3 Specific Hallucination Maps

The use of objective, or task-based, measures of image quality for evaluating imaging systems has been widely advocated [19]. However, the hallucination maps as defined in Section 4.2 do not incorporate any task-specific information. In particular, $\hat{\mathbf{f}}_{null}^{HM}$ may contain an abundance of structures or textures, some of which may not confound an observer on a specified diagnostic task. Hence, it may be useful to identify those structures or textures in the hallucination maps that are task-relevant. One possible way to accomplish this is to process the hallucination map via an image processing transformation T, such that potentially task-relevant features or textures are localized while others are suppressed [34, 37]. Formally, this can be described as:

$$\hat{\mathbf{f}}_{null}^{SHM} = T\hat{\mathbf{f}}_{null}^{HM}, \tag{4.8}$$

where the processed pixel map $\hat{\mathbf{f}}_{null}^{SHM}$ that preserves task-specific information is referred to as a *specific hallucination map*. Note that the design of the transformation T is application-dependent, as it should localize those structures or textures from the hallucination map that are relevant to a specified task. Moreover, the specification of the observer (which could be a human or computational procedure) who will perform the task should also influence the design of T, as the extent to which hallucinations impact observer performance will vary. While requiring significant effort to formulate, specific hallucination maps open up the possibility of comparing reconstruction methods based on their propensities for creating hallucinations that influence task performance.

The complete procedure for computing measurement and null space hallucination maps, as well as the specific hallucination map, is presented in Algorithm 1.

Algorithm 1: Procedure for computation of measurement and null space hallucination maps from measurement data $\mathbf{g}$, system matrix $\mathbf{H}$, true object $\mathbf{f}$ and reconstructed image $\hat{\mathbf{f}}$

1 Compute the truncated pseudoinverse solution:

$$\hat{\mathbf{f}}_{tp} = \mathbf{H}_P^+ \mathbf{g}$$

2 Compute the generalized measurement component of $\hat{\mathbf{f}}$:

$$\hat{\mathbf{f}}_{meas} = \mathcal{P}_{meas}\hat{\mathbf{f}} = \mathbf{H}_P^+ \mathbf{H}\hat{\mathbf{f}}$$

3 Compute the generalized null components of $\mathbf{f}$ and $\hat{\mathbf{f}}$:

$$\mathbf{f}_{null} = \mathcal{P}_{null}\mathbf{f} = [\mathbf{I}_N - \mathbf{H}_P^+ \mathbf{H}]\mathbf{f},$$
$$\hat{\mathbf{f}}_{null} = \mathcal{P}_{null}\hat{\mathbf{f}} = [\mathbf{I}_N - \mathbf{H}_P^+ \mathbf{H}]\hat{\mathbf{f}}.$$

4 **Measurement space hallucination map**:

$$\hat{\mathbf{f}}_{meas}^{HM} = \hat{\mathbf{f}}_{meas} - \hat{\mathbf{f}}_{tp}$$

5 **Null space hallucination map**:

$$\hat{\mathbf{f}}_{null}^{HM} = \mathbb{1}(\hat{\mathbf{f}}_{null}) \odot (\hat{\mathbf{f}}_{null} - \mathbf{f}_{null})$$

6 Apply image processing transformation T on $\mathbf{f}_{null}^{HM}$ to obtain the specific halucination map:

$$\hat{\mathbf{f}}_{null}^{SHM} = T\hat{\mathbf{f}}_{null}^{HM}$$

4.3 Numerical Studies

Numerical studies were conducted to demonstrate the utility of the proposed hallucination maps. Although the focus of these preliminary studies is on null space hallucination maps, the presented analyses could readily be repeated by use of measurement space hallucination maps. Hallucination maps were employed to compare the behavior of data-driven and model-based image reconstruction methods under different conditions.

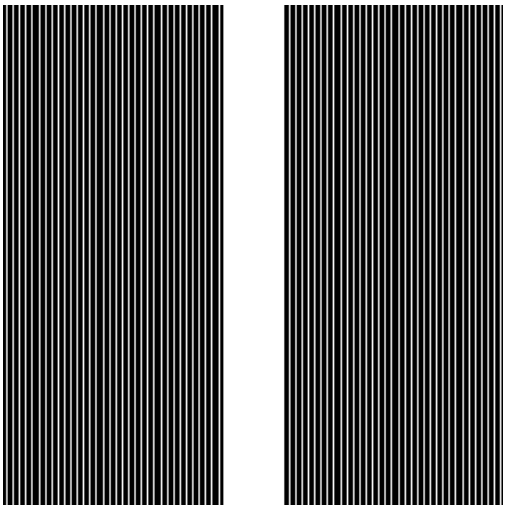


Figure 4.2: Sampling mask

4.3.1 Stylized Imaging System

A stylized two-dimensional (2D) single-coil magnetic resonance (MR) imaging system was considered. It should be noted that the assumed imaging operator was not intended to accurately model a real-world MR imager. Instead, the purpose of the presented simulation studies is only to demonstrate the potential utility of hallucination maps. Hence physical factors such as coil sensitivity and bias field inhomogeneity were not considered. Fully-sampled k-space data were emulated by applying the 2D Fast Fourier Transform (FFT) on the digital objects described below. Independent and identically distributed (i.i.d.) Gaussian noise was added to the real and imaginary components of the complex-valued k-space data [8] in the training dataset for the U-Net as well as in the test data during evaluation with different reconstruction methods. Additionally, in the test dataset, zero-mean random uniform phase noise [144] was introduced into the k-space measurements to emulate modeling errors [19]. A uniform Cartesian undersampling mask with an undersampling factor of 3 was applied on the fully-sampled k-space data to obtain undersampled measurements, as shown in Fig. 4.2. The k-space lines that were not sampled were subsequently zero-filled. The Moore-Penrose pseudoinverse $\mathbf{H}^{+}$ was applied by performing the inverse 2D Fast Fourier Transform (IFFT) on the zero-filled k-space data. Since the true pseudoinverse was considered without any truncation of singular values, the hallucination map in the generalized null space in our studies corresponds to the hallucination map in the true null space.

4.3.2 Reconstruction Methods

Both data-driven and non-data-driven image reconstruction methods were investigated as described below. The data-driven method considered was a U-Net based method [69, 55, 65], while the non-data-driven methods were PLS-TV and Deep Image Prior with TV penalty (DIP-TV). These reconstruction methods and their implementations are discussed below.

U-Net Reconstruction

The U-Net based reconstruction method employs image-domain learning, where a mapping is learned from an initial image estimate that contains artifacts due to undersampling to an accurate estimate of the true object. In our studies, the initial image estimate that was input to the U-Net was obtained by applying the pseudoinverse on the k-space data. Such initial estimates obtained from multiple measurement data were then employed as inputs to a convolutional neural network (CNN), which was trained in order to produce artifact-free images, similar to images from the ground truth distribution. As is common practice, the CNN architecture used in this study is the U-Net [116]. A U-Net consists of two CNNs that represent a downsampling path followed by an upsampling path respectively, and skip connections [44] between similar levels in the downsampling and upsampling paths. Let the initial estimate from the measurement data be denoted as $\mathbf{f}'$ and the function computed by the U-Net be represented as $\mathcal{B}(\mathbf{f}';\theta)$ where $\mathcal{B} : \mathbb{E}^N \to \mathbb{E}^N$ and θ denotes the weight parameters of the U-Net. Given a training data set of initial estimate-ground truth pairs $\left\{\mathbf{f}'_i, \mathbf{f}_i\right\}_{i=1}^{D}$ where D is the size of the training data set, the optimal weight parameters θ^* are learned by approximately solving the following optimization problem:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{D} \mathcal{L}\Big(\mathcal{B}(\mathbf{f}'_i, \theta), \mathbf{f}_i\Big), \tag{4.9}$$

where $\mathcal{L}(\cdot,\cdot)$ is a suitable loss function. In this work, mean absolute error was used as the loss function [148]. The model for the U-Net was based on the single-coil baseline U-Net architecture provided in [146]. A stochastic gradient-based method known as RMSProp [129] was employed to solve the optimization problem in Eq. (4.9). After this iterative scheme for training the U-Net reached convergence, the trained U-Net was used to reconstruct images from a previously unseen test measurement dataset, where an initial estimate $\mathbf{f}'_{test}$

computed from a test measurement data was employed to obtain the reconstructed image $\hat{\mathbf{f}}_{test} = \mathcal{B}(\mathbf{f}'_{test}, \theta^*)$. The training and testing of the U-Net based reconstruction was performed using PyTorch [110] code available at https://github.com/facebookresearch/fastMRI.

PLS-TV

As described previously, in the PLS-TV method Eq. (2.17) is solved using proximal-gradient methods. In this study, PLS-TV reconstruction was performed for a dataset of measurements using the Berkeley Advanced Reconstruction Toolbox (BART) [132]. BART performs PLS-TV reconstruction using the augmented Lagrangian based optimization method proposed in [7]. The regularization parameter λ in Eq. (2.16) was chosen by first performing image reconstruction on a subset of the dataset, with different values of λ. The value of λ which provided the lowest mean of the root mean squared error (RMSE) metric over the subset was chosen, and used for image reconstruction of all the images in the dataset.

DIP

Recently, Ulyanov *et al.* [134] showed that a CNN $G : \mathbb{R}^k \to \mathbb{E}^N$ with randomly initialized weights θ and random input $\mathbf{z} \in \mathbb{R}^k$ can be an effective regularizer for image restoration problems such as denoising, super-resolution and inpainting. This method of regularization, known as deep image prior (DIP), utilizes the observation that the structure of deep convolutional networks captures several low-level image statistics and is biased towards smooth, natural images. Van Veen *et al.* [135] extended the DIP framework to applications in tomographic imaging from incomplete measurements with encouraging results. Essentially, image reconstruction using the DIP method can be formulated in terms of the following optimization problem:

$$\begin{aligned} \theta^* &= \underset{\theta}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}G(\mathbf{z};\theta)\|_2^2, \\ \hat{\mathbf{f}} &:= G(\mathbf{z};\theta^*) \end{aligned} \tag{4.10}$$

where $\mathbf{z}$ and θ are randomly initialized.

It has been shown in [134, 135] that the DIP method overfits the measurement noise upon convergence. Hence, further regularization may be required, either in the form of early stopping or with the addition of penalties in the optimization problem in Eq. (4.10). Inspired by [92], in our experiments, image reconstruction using the DIP method with TV regularization (DIP-TV) was performed by approximately solving the following optimization problem:

$$\begin{aligned} \theta^* &= \underset{\theta}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}G(\mathbf{z}, \theta)\|_2^2 + \lambda \|G(\mathbf{z}, \theta)\|_{TV}, \\ \hat{\mathbf{f}} &:= G(\mathbf{z}; \theta^*) \end{aligned} \tag{4.11}$$

where $\mathbf{z}$ and θ were randomly initialized, and λ is the regularization parameter. The same U-Net architecture employed for the U-Net based reconstruction was employed for DIP-TV, and was implemented in TensorFlow [2]. Similar to the implementation of the PLS-TV method as outlined in Sec. 4.3.2, the regularization parameter λ for the TV penalty in Eq. (4.11) was chosen by first performing image reconstruction on a subset of the dataset, with different values of λ. Subsequently, the value which provided the lowest mean RMSE over the subset was chosen to perform image reconstruction from all the measurements. The optimization problem in Eq. (4.11) was approximately solved using a stochastic gradient algorithm called Adam [86].

4.3.3 Training, Validation and Test Data

For the U-Net based reconstruction method, training was performed on 2D axial adult brain MRI images from the NYU fastMRI Initiative database [146]. These will be referred to as the in-distribution (IND) images. The training and validation datasets contained 2500 and 500 images, respectively. For testing, both IND and out-of-distribution (OOD) images were considered. The OOD images were obtained from a pediatric epilepsy resection MRI dataset [98]. Both the IND and OOD testing datasets contained 69 images. It should be noted that the OOD images differed from the IND images in several aspects, such as the nature of the objects (adult for IND and pediatric for OOD) as well as the use of different MR systems used to obtain the true object images in each case. All images were of dimension 320×320.

After creating the training, validation and test datasets, neural network training was performed with the IND training and validation datasets for the U-Net method. At test time,

images were reconstructed from both IND and OOD test datasets using the U-Net, PLS-TV and DIP methods.

4.3.4 Computation of Hallucination Maps

After images were reconstructed from the testing data, null space hallucination maps $\hat{\mathbf{f}}_{null}^{HM}$ were computed. The quantities $\hat{\mathbf{f}}_{null}$ and $\mathbf{f}_{null}$, as required by Eq. (4.6), were computed according to Eq. (2.12). Subsequently, specific null space hallucination maps $\hat{\mathbf{f}}_{null}^{SHM}$ were also computed. In this preliminary study, these maps were designed for the purpose of localizing regions where coherent structures, as opposed to random errors, were present in $\hat{\mathbf{f}}_{null}^{HM}$. Such structured hallucinations could be relevant to certain signal detection tasks. To accomplish this, the transformation T in Eq. (4.8) was implemented as follows.

First, the region of support of each object was identified using Otsu's method [66] and binary support masks were formed for each object. The support masks were applied on the $\hat{\mathbf{f}}_{null}^{HM}$ such that errors in the reconstructed image that lie outside the region of support could be ignored. Subsequently, histogram equalization was performed. A 2D Gaussian filter with kernel width of 7 was applied on the histogram-equalized map in order to obtain a smooth distribution of intensities across the hallucination map. The width of the Gaussian filter was chosen heuristically in this study. Finally, a binary threshold was applied where the cut-off value was set to the 95-th percentile of intensity values in the processed map, such that intensities below the threshold were set to zero and intensities above the threshold were set to 1. From the thresholded maps, connected components that had a size of less than 100 pixels ($\approx 0.1\%$ of total number of pixels in each image) were eliminated to remove localized regions with negligible dimensions, resulting in the specific hallucination maps $\hat{\mathbf{f}}_{null}^{SHM}$ for our studies. This procedure for computing the action of T was repeated for all $\hat{\mathbf{f}}_{null}^{HM}$ computed from both the IND and OOD test datasets for each reconstruction method. It should be noted that this procedure serves only as a simplistic example of the computation of a specific hallucination map, and there is no suggestion that it is optimal in any sense.

Finally, conventional error maps were computed as the difference between the reconstructed estimate $\hat{\mathbf{f}}$ and the true object $\mathbf{f}$. In order to demonstrate the potential utility of the specific hallucination maps over processed versions of conventional error maps, *specific error maps*

were formed by acting T on the error maps. The codes employed in our numerical studies are available at https://github.com/comp-imaging-sci/hallucinations-tomo-recon.

4.4 Results

The numerical results are organized as follows. First, an illustration of null space hallucination maps is provided for different reconstruction methods, in order to demonstrate their utility in highlighting false structures that may be introduced due to the imposed prior. Differences in the null space hallucination maps corresponding to the data-driven U-Net method when applied to IND and OOD data are examined. This is followed by a demonstration of the difference in the quantitative performance of the U-Net method on IND and OOD data. The performance of the U-Net is compared with the non-data-driven methods in our studies – PLS-TV and DIP – in terms of metrics derived by use of null space hallucination maps. While the results below focus on null space hallucination maps, similar analysis can also be performed on measurement space hallucination maps. Examples of measurement space hallucination maps are provided in Appendix A.

4.4.1 Differences Between Error and Hallucination Maps

Reconstructed images and corresponding error maps and null space hallucination maps from an IND measurement are shown in Fig. 4.3. It can be observed that, for all the reconstruction methods, the error map and the null space hallucination map have different characteristics in some regions of the image. This is because the error map contains false structures due to hallucinations as well as all other factors, whereas the null space hallucination map only contains errors due to the imposed prior. These differences can also be observed from the computed specific error maps and specific null space hallucination maps. As expected, the U-Net method performs well, leading to mostly low-intensity regions in the null space hallucination map. In one of the regions that is featured in the specific hallucination map for all the reconstruction methods, it can be seen that the U-Net has lower hallucinations since it is able to faithfully recover fine structures in the region. Such fine structures were oversmoothed in the reconstructed images that were obtained by use of the PLS-TV and DIP methods, leading to higher hallucinations. On the other hand, all the reconstructed

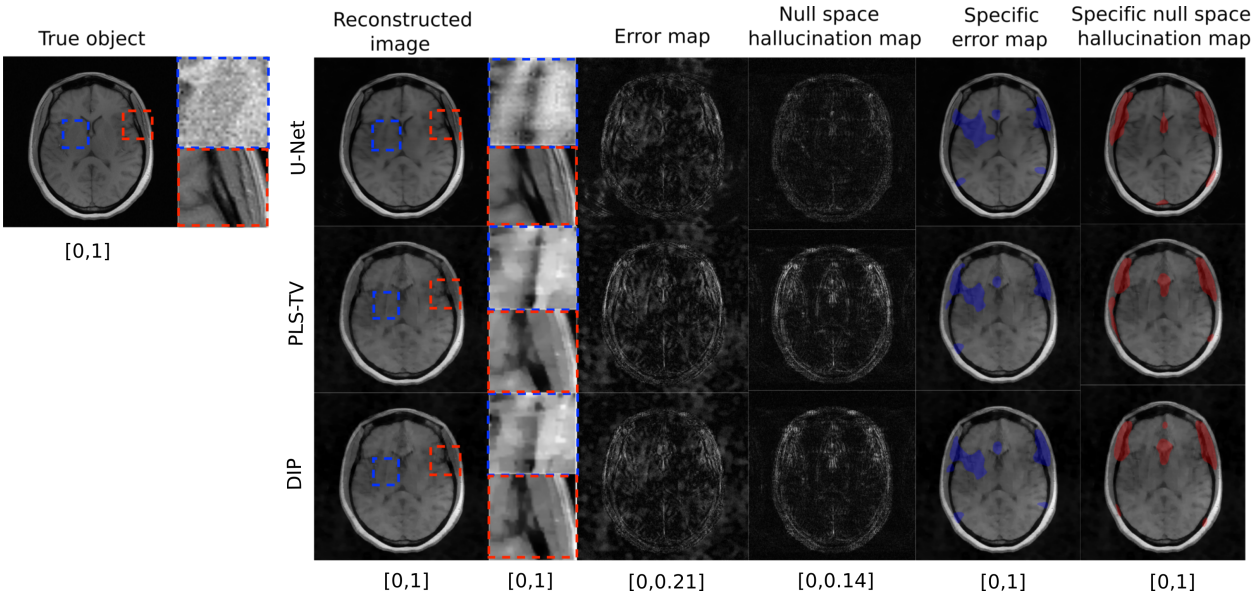


Figure 4.3: Example of a true object and reconstructed images along with error maps and hallucination maps (null space) for IND data with different reconstruction methods – U-Net (top), PLS-TV (middle) and DIP (bottom). Expanded regions are shown to the right of the reconstructed images. The specific error map (blue) and specific null space hallucinations map (red) are overlaid on the reconstructed images for each method. The image estimated by the U-Net method has visibly lower hallucinations in the null space compared to PLS-TV and DIP. The region within the red bounding box is one of the locations that contains hallucinations for all the reconstruction methods. In this region, the U-Net method shows mild hallucinations compared to PLS-TV and DIP. Fine structures in this region appear to be oversmoothed in the image estimates obtained by use of PLS-TV and DIP. A false structure is also shown (within the blue bounding box region) that appears for all the reconstruction methods due to the phase noise and not due to the imposed prior, and hence cannot be classified as a hallucination.

images also contain a distinct false structure that is revealed in the specific error map but not the specific hallucination map. This is an example of a false structure that can exist in reconstructed images, but may not necessarily be classified as a hallucination.

To further demonstrate the different characteristics of error maps and null space hallucination maps for this IND study, scatter plots of the centroids of the detected regions in each type of map corresponding to the ensemble of IND reconstructed images from all three reconstruction methods are shown in Fig. 4.4 (top row). From these scatter plots, it can be observed that there is a high amount of variance in the locations of the detected regions in the specific error maps as compared to the detected regions in the specific hallucination maps. The

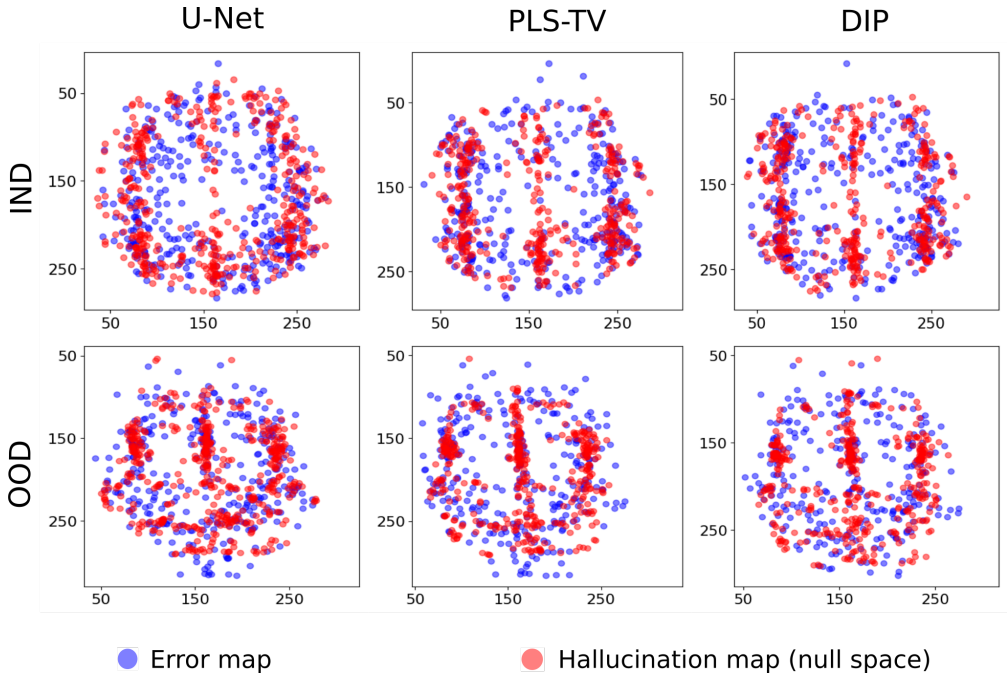


Figure 4.4: Scatter plots for centroids of localized regions in specific error maps and specific null space hallucination maps with different reconstruction methods for IND (top) and OOD (bottom) data. Note that for each type of data distribution and for all the reconstruction methods, the centroids of the regions detected from the error map have a higher variance compared to the hallucination map as well as some degree of non-overlap.

latter typically appear in similar regions across the ensemble of reconstructed images for all the methods. Furthermore, the concentrations of centroids for the detected regions in both types of maps have some degree of non-overlap. These observations reflect the fact that, due to additional sources of error such as measurement noise and model error that are also typically random in nature, the regions in the reconstructed images that are revealed by the error map can sometimes be different from those revealed by the null space hallucination map that considers error only due to an inaccurate prior.

As the distribution shifts to OOD, as shown in Fig. 4.5, the null space hallucination map for the U-Net method appears comparable to the hallucination maps obtained by use of PLS-TV and DIP. False structures that can be identified as hallucinations appear in the image reconstructed by the U-Net method. The higher error for the U-Net method is a result of the change of distribution and the method's inability to generalize well to data that are significantly out of distribution with respect to the training data. The change of distribution results in significant inaccuracies in the null component of the reconstructed

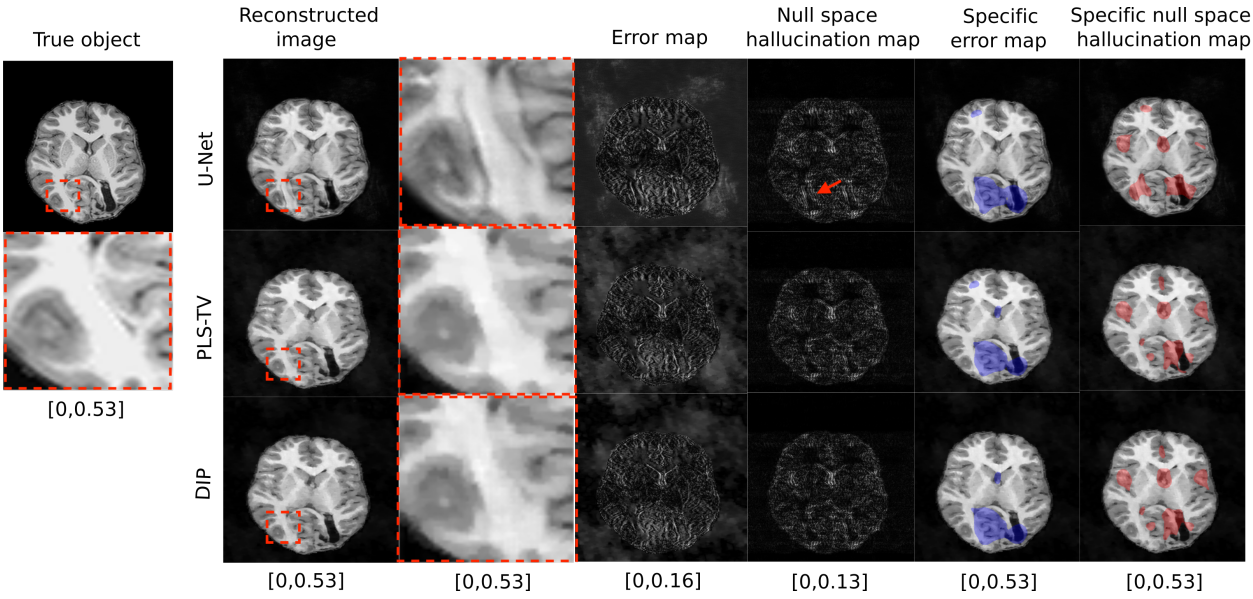


Figure 4.5: Example of true object and reconstructed images along with error map and hallucination maps (null space) for OOD data with different reconstruction methods – U-Net (top), PLS-TV (middle) and DIP (bottom). Expanded regions are shown to the right of the reconstructed images. The specific error map (blue) and specific null space hallucinations map (red) are overlaid on the reconstructed images for each method. The image estimated by the U-Net method has some distinct false structures (region within red bounding box) that do not exist in the reconstructed images obtained by using PLS-TV and DIP. This region is also highlighted in the specific null space hallucination map for the U-Net method which suggests that the false structure is a hallucination.

estimate produced by the U-Net. Under such circumstances, it can be useful to identify and localize hallucinations due to inaccuracies in the imposed data-driven regularization through the null space hallucinations.

As shown in Fig. 4.5 and consistent with the IND results discussed above, the localized regions detected in the specific error map and specific hallucination map for the OOD cases are generally different. Scatter plots of the centroids of the detected regions in the specific error maps and specific hallucination maps confirm this and are displayed in Fig. 4.4 (bottom row). For all the reconstruction methods, the error map centroids again have a higher variance and are located away from clusters of hallucination map centroids in some regions. In other words, under such circumstances, one cannot rely on only the error maps without considering the corresponding hallucination maps in order to estimate where hallucinations due to the imposed prior are likely to be localized in a reconstructed image.

Although hallucination maps can reveal false structures, the impact of the false structures on specific applications requires further analysis. For example, a false structure may be classified as a *false positive structure* or a *false negative structure* [63, 36]. A false positive structure is one which is absent in the true object but present in the reconstructed image, whereas a false negative structure denotes the opposite. While an important topic, the classification of hallucinations is beyond the scope of this dissertation.

4.4.2 Investigation of Structured Hallucinations

Additional studies were conducted to validate that the specific hallucination maps actually revealed regions in the image that contain significant errors. To accomplish this, two empirical probability distribution functions (PDFs) were estimated that describe the average SSIM values computed over two non-overlapping regions in the reconstructed images for the OOD case. One region corresponded to the support of the specific hallucination maps described above and the second region was spanned by all other pixels in the image. The two empirical PDFs are shown in Fig. 4.6a and reveal that the mode of the distribution corresponding to the SSIM averaged over the structured hallucination regions is demonstrably lower than that describing the average SSIM values over the background regions.

The empirical PDFs that described the SSIM value averaged over the structured hallucination regions were also compared for each of the three reconstruction methods. As shown in Fig. 4.6b, for the IND case, the images reconstructed by use of the U-Net had significantly higher SSIM values, on average, in the structured hallucination regions as compared to both the PLS-TV and DIP methods. This can be attributed to network training with a sufficiently large amount of IND data. However, for the OOD case in Fig. 4.6c, because null space hallucinations increased for the U-Net method, the corresponding reconstructed images had lower SSIM values on average as compared with DIP in the support of the null space hallucination maps. The medians of ensemble SSIM values in these support regions for all the reconstruction methods with IND and OOD data are shown in Table 4.1. It should be noted that, for both the IND and OOD cases, the DIP method was implemented with the same network architecture as the U-Net-based method. Thus, when there is a shift in the testing data distribution, some data-driven methods such as the U-Net method may not provide any significant improvement in the estimate of the null component compared

to model-based methods that do not employ training data. However, the data-driven methods involve the additional risk of hallucinating false structures. These observations gained through hallucination maps provide insight into the impact of the data-driven nature of the prior imposed by pre-trained neural networks.

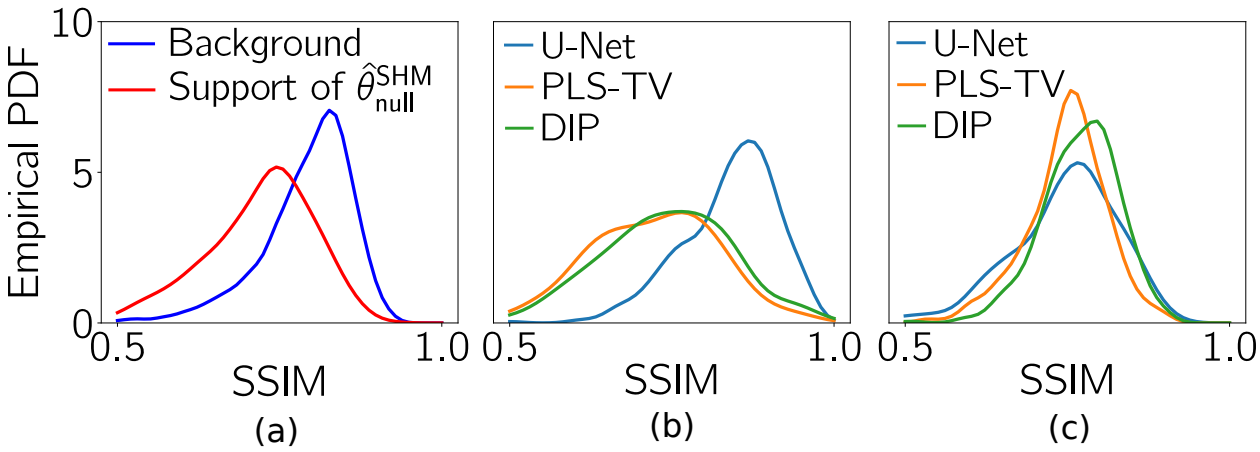


Figure 4.6: (a) Empirical PDF of SSIM values in the structured hallucination regions (support of $\hat{\mathbf{f}}^{SHM}_{null}$) and the regions spanned by the remaining pixels in the support of the image (background), respectively, for the U-Net method with OOD data. Empirical PDFs of SSIM values in the structured hallucination regions for all three reconstruction methods with (b) IND and (c) OOD data respectively.

Data distribution	U-Net	PLS-TV	DIP
IND	**0.84**	0.71	0.73
OOD	0.75	0.73	**0.76**

Table 4.1: Median of SSIM values from ensembles of images reconstructed by use of the U-Net, PLS-TV and DIP methods that were computed in the support region of specific null space hallucination maps. In these regions, the U-Net method has the highest median SSIM for IND data, while for OOD data the DIP method has the highest median SSIM.

4.4.3 Bias Maps and Hallucinations

A *bias map*, defined as

$$\mathbf{b} := \mathbb{E}\hat{\mathbf{f}} - \mathbf{f}, \tag{4.12}$$

determines the expected deviation of an image estimate from the true object, and as such, may include contributions from an incorrect prior, as well as those from incorrect measurement and noise models. Hence, the bias map may be correlated with the hallucination maps, but may display significant differences from it based on the average behavior of the inaccuracies in the measurement and noise models. For example, Fig. 4.7 shows the bias map computed using a dataset of images estimated from simulated undersampled MRI measurements with fixed phase noise and iid Gaussian additive noise, along with the error map and the null space hallucination map for an IND and an OOD image. The corresponding true objects are shown in Figures 4.3 and 4.1b respectively. Figure 4.7 shows that the bias map retains clusters of artifacts from the error map that are due to the phase noise. Hence, although the bias maps are correlated with both the hallucination maps and the error map, each provides a different kind of information.

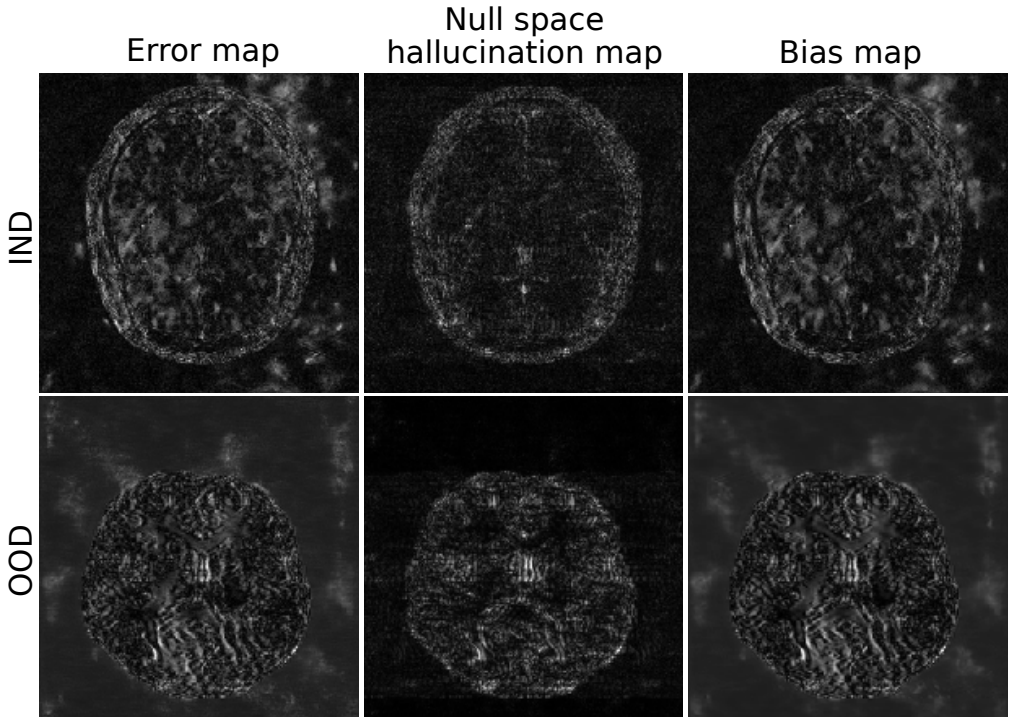


Figure 4.7: An error map, a null space hallucination map and a bias map for IND and OOD images estimated by use of the U-Net method. The corresponding true objects are shown in Figures 4.3 and 4.1b respectively. The bias map was computed over a dataset of 100 images estimated from a single set of simulated measurements with fixed phase noise and different realizations of the iid Gaussian noise. The bias map contains contributions from both the model error, as well as inaccuracies in the prior.

4.5 Summary

While regularization via sparsity-promoting penalties in an optimization-based reconstruction framework is commonly employed, emerging learning-based methods that employ deep neural networks have shown the potential to improve reconstructed image quality further by learning priors from existing data. However, an analysis of the prior information learned by deep networks and their ability to generalize to data that may lie outside the training distribution is still being explored. Additionally, there are open questions and concerns about the stability of such networks when applied for image reconstruction. While it has been understood that use of an inaccurate prior might lead to false structures, or hallucinations, being introduced in the reconstructed image, formal definitions for hallucinations within the context of tomographic image reconstruction have not been reported.

In this work, by use of concepts from linear operator theory, formal definitions for hallucination maps in linear tomographic imaging problems are introduced. These provide the opportunity to isolate and visualize image hallucinations that are contained within the measurement or null spaces of a linear imaging operator. The measurement space hallucination map permits the analysis of errors in the measurement space component of a reconstructed object estimate with respect to the component of the object that can be stably computed from a given set of measurement data. Alternatively, the null space hallucination map permits analysis of errors in the null space component of a reconstructed object estimate with respect to the true object null space component. These errors are caused solely by the reconstruction prior. Both maps can be employed to systematically investigate the impact of different priors utilized in image reconstruction methods. Finally, the notion of a specific hallucination map was also introduced, which can be formulated to reveal hallucinations that are relevant to a specified image-based inference.

Numerical studies were performed with simulated undersampled measurements from a stylized single-coil MRI system. Both data-driven and non-data-driven methods were investigated to demonstrate the utility of the proposed hallucination maps. It was observed that null space hallucination maps can be particularly useful as compared to traditional error maps when assessing the effect of data-driven regularization strategies with out-of-distribution data. Furthermore, it was shown that structured hallucinations with data-driven methods that are caused due to a shift in the data distribution may ultimately lead to significant artifacts in the reconstructed image.

The computation of the projection operations as described in Eq. (2.11) and Eq. (2.12) via the SVD may be infeasible for large-scale problems. Wilson and Barrett [141] proposed an iterative method to compute $\mathbf{f}_{meas}$ and $\mathbf{f}_{null}$ without explicit computation of the SVD of $\mathbf{H}$. Alternatively, randomized SVD [52] is a relatively computationally efficient algorithm that can be employed to estimate these quantities. Kuo *et al.* [89] recently proposed a method to learn null space projection operations that can significantly reduce the computational burden. It may also be expected that the importance of analyzing hallucinations in image reconstruction can further stimulate the development of efficient methods for implementing projection operators. The development of such computationally efficient methods for large-scale problems remains an active area of research.

It should be noted that the proposed definition of hallucination maps is general and can be applied to any linear imaging system and reconstruction method, provided that the computation of the projection operators $\mathcal{P}_{meas}$ and $\mathcal{P}_{null}$ is feasible. Depending on the sampling pattern involved in the data acquisition process, different system matrices $\mathbf{H}$ will have different null space characteristics. This, in turn, may lead to different properties in the corresponding hallucination maps that would allow a comparison of reconstruction methods under a variety of data acquisition strategies.

The proposed framework is most useful in situations where the generalized null component of the true object is significant and hence strong priors need to be incorporated in the reconstruction method via regularization. If the generalized null component is relatively small compared to the generalized measurement component, the need for strong regularization during reconstruction is diminished. This, in turn, would imply that hallucinations are likely to be minimal or non-existent due to the imposed weak regularization and hence computing hallucination maps may not be necessary. In such situations, computing only the error map may be sufficient to assess the reconstruction method.

There remain important topics for future investigation. Beyond the framework presented, it will be important to derive objective figures-of-merit (FOMs) from ensembles of hallucination maps. Furthermore, the probability of occurrence of hallucinations can be potentially quantified from ensembles of hallucination maps. While understanding the interplay between hallucinations and image reconstruction priors is important in preliminary studies, ultimately, image reconstruction methods should be objectively evaluated with consideration of all physical and statistical factors.

Chapter 5

Mining the Manifolds of StyleGANs for Multiple Data-Consistent Solutions of Ill-Posed Tomographic Imaging Problems

5.1 Overview

Most medical image reconstruction methods available today are designed to produce a single estimate of the object, which is known as the *maximum a posteriori* (MAP) point estimate when interpreted in a statistical framework. However, in the presence of data noise or incompleteness, multiple objects can exist that are consistent with a given set of measurement data. Moreover, there is generally no guarantee that the produced object estimate will be the most accurate or useful (with respect to a specific clinical task) among the multiple possible objects that are consistent with the measured data. This is especially true for many deep learning-based image reconstruction methods, which are often based on heuristic designs and can have an enhanced propensity for producing hallucinated structures [23]. These hallucinated structures are of particular concern for medical imaging applications because such structures may not always be readily identifiable as artifacts and therefore the images can appear plausible but are, in fact, incorrect.

The ability to identify multiple objects that are consistent with a given set of measurement data is of significant importance to the assessment and refinement of data-acquisition designs and image reconstruction procedures. For example, from a collection of distinct data-consistent objects, uncertainty maps [128, 119] can be computed. Such maps can be

employed to reveal the reliability of a reconstructed image corresponding to a given data-acquisition design. The ability to identify multiple data-consistent objects could also permit analysis of the impact of the null space of a linear imaging operator in new, problem-specific, ways and enable the design of numerical experiments to reveal image reconstruction instabilities [49]. Moreover, a new capability to generate ensembles of data-consistent objects is needed to advance task-informed adaptive imaging procedures [39, 18].

The generation of multiple solutions to an inverse problem is consistent with the goal of Bayesian inversion methods [133]. In imaging applications, this can be conceptually achieved by sampling from the posterior distribution that describes the sought-after object conditioned on a set of measurement data [104]. This is a holy grail of image reconstruction, but it remains generally impractical in medical imaging applications due to their large scale [50]. In recent years, computational procedures for accomplishing *approximate* posterior sampling in limited-scale problems have been proposed that employ deep neural networks combined with Markov chain Monte Carlo (MCMC) sampling methods or Langevin dynamics [126, 120, 106, 68]. While promising, the efficacy of such methods for use with large-scale medical image reconstruction problems remains a topic of investigation. Furthermore, such posterior sampling methods are limited to imaging systems with specific noise distributions, typically Gaussian, and may not work directly when the noise distribution is different, e.g. Poisson [38]. To circumvent the computational challenges of posterior sampling methods and their lack of generalizability to different noise distributions, *empirical sampling* [126, 16, 9] can be performed to obtain multiple distinct objects that are consistent with a given set of measurement data. In an empirical sampling method, multiple data-consistent solutions are obtained by solving a regularized inverse problem within a stochastic optimization framework [125]. Empirical sampling methods, while not guaranteeing true posterior sampling, can be computationally feasible for large-scale imaging systems. Moreover, in contrast to MCMC-based posterior sampling methods in which the samples are generated sequentially, alternate solutions obtained via empirical sampling are independent of one another and can be obtained in parallel, thus providing reductions in computation times.

It may be possible that when multiple solutions are sought from a single acquisition of measurement data, the data-consistent objects may contain unrecognizable structures that are irrelevant to the medical imaging application at hand. Such situations may arise when there is no constraint imposed on the objects to ensure that they are relevant to a specified imaging application. One way to achieve empirical sampling that produces application-relevant and

data-consistent objects is to constrain the process by use of a deep generative model [47] that characterizes the distribution of to-be-imaged objects. For the single image super-resolution problem (SISR), an optimization-based technique called Photo Upsampling via Latent Space Exploration (PULSE) was proposed in [102] to generate diverse photorealistic high-resolution images from a single low-resolution image. A state-of-the-art deep generative model known as StyleGAN [76] was employed to characterize the distribution of high resolution images. However, due to the complex nature of the latent space of the StyleGAN, the optimization method in PULSE violates the constraint that the generated image resides in the range of the StyleGAN. Consequently, the high-resolution images may contain artifacts that are irrelevant to the imaging application. Furthermore, there remains an important need to extend this method for general tomographic inverse problems and quantitatively investigate the data-consistency of the generated samples.

In this work, the following problem is addressed: *Assume a StyleGAN describing a distribution of to-be-imaged objects, a tomographic measurement model, and a single acquisition of incomplete and noisy measurement data are provided. Find a collection of distinct objects that are consistent with the same acquired measurement data (in a to-be-prescribed sense) and reside in the range of the StyleGAN.* A key motivation for formulating this problem is to establish an application-relevant empirical sampling method that can be employed in preliminary assessments and refinements of data-acquisition designs and tomographic imaging technologies via virtual imaging trials. A method for solving this problem is proposed that is referred to as the PULSE++ method. The PULSE++ method represents the first extension of the PULSE methodology for use with tomographic imaging problems. The optimization method used in PULSE is redesigned in the PULSE++ method to facilitate a more effective search in the complex latent space of a StyleGAN and ensures the constraint that the generated image resides in the range of the StyleGAN. Additionally, by utilizing improved assumptions about the statistics of object embeddings in the latent space of the StyleGAN, the ability of the PULSE++ method to produce diverse, application-relevant, and data-consistent objects is enhanced as compared to the original PULSE method. It should be noted that the PULSE++ method may be considered as an image reconstruction method if a single estimate is sought from the multiple available solutions using statistical methods, and more objectively based on the task involved. In this study, PULSE++ is not utilized as an image reconstruction method, but rather as an efficient empirical sampling technique that can be employed to assess and refine a given data acquisition design in new,

problem-specific ways, e.g. by enabling computation of reliable uncertainty maps. Two different stylized tomographic imaging modalities are systematically studied that also involve different measurement noise distributions. Uncertainty maps that quantify the degree of variability of data-consistent alternate solutions are computed.

5.2 Salient Features of the StyleGAN Latent Space

State-of-the-art deep generative models such as the StyleGAN [76] hold the potential for characterizing the distribution of finite-dimensional approximations of to-be-imaged objects [150, 61]. Let $G : \mathbb{R}^k \rightarrow \mathbb{R}^N$ denote a parameterized deep generative model with L layers, where $k \ll N$. The generator network $G(\mathbf{z})$ is trained such that it maps a k-dimensional latent vector $\mathbf{z} \in \mathcal{Z}$ sampled from a known distribution, such as a standard Gaussian distribution, to an image that is representative of the distribution formed by the training images. The generator network G in a StyleGAN is composed of two networks: a mapping network G_m and a synthesis network G_s [76]. G_m is a fully connected neural network (FCNN) [47] that maps the latent vector $\mathbf{z}$ to an intermediate style latent vector $\mathbf{w} \in \mathbb{R}^k$. Subsequently, the style latent vector $\mathbf{w}$ is replicated L times, and each duplicate style latent vector $\mathbf{w}$ is passed through a learned affine transformation [62] that encodes semantic information and input to one of the L layers in the synthesis network G_s. Each such vector that is input to G_s controls a specific style or semantic attribute in the generated image. The collection of L copies of the vector $\mathbf{w} \in \mathbb{R}^k$ is represented as the latent matrix $\mathbf{W} \in \mathbb{R}^{k \times L}$ and the corresponding intermediate latent space is denoted as $\mathcal{W}$. Additionally, the StyleGAN contains a set of L noise latent vectors $\mathbf{\Phi} \equiv \{\boldsymbol{\phi}_i\}_{i=1}^{L}$ such that $\boldsymbol{\phi}_i \in \mathbb{R}^{p_i}$, where $p_i = 4^{(1+\lceil \frac{i}{2} \rceil)}$ [76]. Each noise latent vector $\boldsymbol{\phi}_l$ serves as an input to layer l in G_s. These noise latent vectors are sampled from standard Gaussian distributions and multiplied by learned scaling factors [76] that enable additional stochastic variability in the fine details of the generated images.

In addition to a superior performance in image synthesis, the style-specific control that is gained with a StyleGAN generator can be leveraged to perform meaningful semantic transformations of objects in tomographic imaging applications [80, 46, 118]. To perform such semantic transformations, an embedding for the given object must be obtained first in the latent space of the StyleGAN. Abdal *et al.* [3] proposed an efficient embedding algorithm

that involved solving an optimization problem in an *extended* latent space $\mathcal{W}^+ \equiv \mathbb{R}^{k \times L} \supset \mathcal{W}$. Penalty terms such as *GEOCROSS* [102] have been proposed that promote the embedding in the extended latent space $\mathcal{W}^+$ to be close to the latent space $\mathcal{W}$, which in turn encourages the embedded object to be near the range of the generator network G_s with the latent space $\mathcal{W}$ [143].

Recently, a number of studies posited that the problem of embedding via optimization may be better conditioned by utilizing a modified StyleGAN latent space that possessed a more well-defined structure. It was empirically observed that the application of a certain computationally cheap and invertible transformation $\mathcal{T} : \mathbb{R}^{k \times L} \to \mathbb{R}^{k \times L}$ produced a matrix $\mathbf{V} = \mathcal{T}(\mathbf{W})$, where the columns of $\mathbf{V}$ denoted as $\{\mathbf{v}_i\}_{i=1}^{L}$ approximately followed the standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_k)$ [102, 143, 151]. In particular, $\mathcal{T}$ is the composition of a leaky rectified linear unit (ReLU) [47] with an affine whitening transformation [151]. Additional details regarding the transformation operator $\mathcal{T}$ are described in Appendix B. In what follows the spaces $\mathcal{V} \subset \mathbb{R}^{k \times L}$ and $\mathcal{V}^+ \equiv \mathbb{R}^{k \times L}$ denote the images through $\mathcal{T}$ of the spaces $\mathcal{W}$ and $\mathcal{W}^+$, respectively. Using this transformation, the synthesis network G_s can be equivalently represented in terms of a generator network $\tilde{G}(\mathbf{V}, \mathbf{\Phi}) := G_s(\mathcal{T}^{-1}(\mathbf{V}), \mathbf{\Phi})$. It was reported in [102, 143, 151] that performing optimization in the latent space $\mathcal{V}^+$ with the generator network $\tilde{G}$ allowed for the use of simple regularizers and led to more accurate embeddings.

It has been empirically observed that the embedding quality can be further improved by optimizing over the noise latent vectors $\mathbf{\Phi}$ in addition to $\mathbf{V}$ [4, 78]. While being effective, the optimization of noise latent vectors $\mathbf{\Phi}$, which have a significantly higher number of parameters compared to the object dimensions, may be unstable and lead to significant artifacts in the embedded object [4]. Accordingly, it has been suggested that the optimization over $\mathbf{\Phi}$ should be performed by imposing appropriate regularization [78].

5.3 Empirical Sampling with PULSE

Menon *et al.* [102] employed such preferable embedding properties of the StyleGAN latent space $\mathcal{V}^+$ in a SISR task, with the goal of mining the manifold of the generator network $\tilde{G}(\mathbf{V}, \mathbf{\Phi})$ to discover multiple photo-realistic high-resolution images that are consistent with the same low-resolution image. The proposed framework was termed as Photo Upsampling

via Latent Space Exploration (PULSE). The PULSE method belongs to a broader class of generative model-constrained methods for ill-posed inverse problems in imaging, known as compressed sensing using generative models (CSGM) [28, 26, 81]. The embedding problem within the CSGM framework is highly non-convex. However, gradient-based methods have been observed to find *good* local optima in a computationally feasible manner, thus providing solutions to the inverse problem that are compatible with the measurement data [28].

In SISR, the measured data $\mathbf{g} \in \mathbb{E}^M$ is a low-resolution version of the sought-after image $\mathbf{f} \in \mathbb{E}^N$, where $M \leq N$. The imaging operator $\mathcal{H} : \mathbb{E}^N \to \mathbb{E}^M$ in SISR is a degradation operator that removes the higher spatial frequencies from $\mathbf{f}$. The CSGM optimization problem in PULSE was formulated to recover a high-resolution estimate $\hat{\mathbf{f}} \equiv \tilde{G}(\hat{\mathbf{V}}, \hat{\mathbf{\Phi}})$ from the low-resolution image $\mathbf{g}$, stated as

$$\begin{aligned} \hat{\mathbf{V}}, \hat{\mathbf{\Phi}} = \underset{\mathbf{V},\mathbf{\Phi}}{\operatorname{argmin}} \Big\{ \mathcal{J}(\mathbf{g}, \tilde{G}(\mathbf{V}, \mathbf{\Phi})) + \lambda_g GEOCROSS(\mathbf{V}) \Big\}, \\ s.t.\ \mathbf{V} \in \mathcal{V}^+, \mathbf{v}_i \in S^{k-1}(\sqrt{k})\ \forall i \in \{1 \dots L\}, \\ \boldsymbol{\phi}_i \in S^{p_i-1}(\sqrt{p_i})\ \forall i \in \{1 \dots L_{\boldsymbol{\phi}}\}, \\ \boldsymbol{\phi}_i = \overline{\boldsymbol{\phi}}_i\ \forall i \in \{L_{\boldsymbol{\phi}} + 1 \dots L\}. \end{aligned} \tag{5.1}$$

Above $\lambda_g > 0$ is a regularization hyperparameter, $S^{d-1}(r) \equiv \{\mathbf{a} \in \mathbb{R}^d \,|\, \|\mathbf{a}\|_2 = r\}$ denotes the spherical surface of radius r in d-dimensions, $L_{\boldsymbol{\phi}}$ denotes the number of levels for which the noise latent vectors are optimized, and $\overline{\boldsymbol{\phi}}_i$ $(i = L_{\boldsymbol{\phi}} + 1, \dots, L)$ denote the higher-resolution noise latent vectors that are randomly chosen and kept fix during the optimization. The motivation for imposing the norm constraints on the style latent $\{\mathbf{v}_i\}$ and noise $\{\boldsymbol{\phi}_i\}$ vectors was based on the "soap bubble effect" [102] observed for standard multivariate Gaussian vectors in high-dimensional spaces, as discussed in Sec. 5.4. To mitigate instabilities due to optimization over $\mathbf{\Phi}$, additional regularization was imposed by optimizing over only the initial $L_{\boldsymbol{\phi}} = 5$ low-resolution noise latent vectors in $\mathbf{\Phi}$, while keeping the high-resolution noise latent vectors fixed. The data fidelity term $\mathcal{J}(\mathbf{g}, \tilde{G}(\mathbf{V}, \mathbf{\Phi}))$ was chosen to be a suitable ℓ_p-norm. The penalty term $GEOCROSS(\mathbf{V})$ is defined as the sum of pairwise geodesic distances among the L style latent vectors in $\mathbf{V}$ on $S^{k-1}(\sqrt{k})$ [102]. Equation (5.1) was approximately solved by jointly optimizing over $(\mathbf{V}, \mathbf{\Phi})$ and using projected gradient descent with the Adam optimizer [86]. Multiple runs of the optimization problem were performed, and for each run, $\mathbf{V}$ and $\mathbf{\Phi}$ were randomly initialized by sampling from standard Gaussian distributions. Due to the high degree of non-convexity in Eq. (5.1), varying the initialization

on each run produced solutions corresponding to different local minima of the optimization problem. Thus, after the completion of all the CSGM runs, a diverse set of photo-realistic high-resolution images could be obtained that were significantly different from each other while being qualitatively consistent with the same observed low-resolution image.

Despite results that were promising qualitatively, a quantitative validation of the method was omitted. First, the choice of the data fidelity function $\mathcal{J}(\mathbf{g}, \tilde{G}(\mathbf{V}, \mathbf{\Phi}))$ as an ℓ_p-norm was not justified in terms of the statistics of the noise distribution in the low-resolution image. Second, the tolerance level for data consistency to accept a high-resolution image estimate was chosen in an arbitrary fashion irrespective of the measurement noise distribution. It is therefore difficult to determine the degree to which data consistency is being preserved. Third, by performing optimization in the latent space $\mathcal{V}^+$ instead of $\mathcal{V}$, the PULSE method allows the generated image to lie outside the range of the pre-trained StyleGAN, and utilizes the *GEOCROSS* term to limit the extent to which the range of the StyleGAN is extended. This is conflicting with our objective of finding data-consistent images that reside in the range of the StyleGAN, and may result in images not being application-relevant and containing artifacts [142]. Forth, while the PULSE method, as well as previous studies such as [143, 151, 80], employed the prior assumption that the style latent vectors $\{\mathbf{v}_i\}$ follow a standard multivariate Gaussian distribution, no rigorous quantitative evaluation was performed to justify the accuracy of this ansatz. While the lack of such quantitative validation may still be acceptable for a computer vision task where the objective is to obtain diverse photo-realistic face images from a given low-resolution image, a quantitative assessment of the method is critical for a proper assessment of tomographic imaging systems.

5.4 Statistical Validation of the Gaussianized Latent Space in StyleGAN

In this section, a statistical study is described that demonstrates that the underlying assumption of a Gaussian structure in the StyleGAN latent space $\mathcal{V}^+$ in the PULSE method is inaccurate. The validation study was performed using two StyleGAN models trained on medical image datasets (MRI-StyleGAN and CT-StyleGAN), as well as the open-sourced StyleGAN model trained on human face images of size 1024×1024 [76] (Face-StyleGAN) that was employed in the PULSE method for SISR [102]. The training of StyleGANs was

performed by adapting an open-sourced TensorFlow-based code [75]. MRI-StyleGAN was trained using 60,000 axial knee images of size 256 × 256 pixels extracted from the NYU fastMRI dataset [146]. CT-StyleGAN was trained using 60,000 X-ray CT chest images of size 512×512 pixels extracted from the NIH DeepLesion dataset [145]. For both the MRI and CT StyleGAN models, each image was normalized to lie in the range [0,1] prior to training. The default training hyperparameters of the open-sourced StyleGAN model were employed. The MRI-StyleGAN model was trained using 2 NVIDIA TITAN X GPUs, while training of the CT-StyleGAN model was performed using 4 NVIDIA V100 GPUs. Both the models were trained for ∼1 day. The MRI-StyleGAN and CT-StyleGAN models along with their pre-trained weights are provided in the code repository that accompanies this paper [25]. For a quantitative assessment of the trained StyleGAN models, the Fréchet Inception Distance (FID) [58] was computed. A lower value of the FID score for a GAN indicates better generative performance. The FID score in each case was computed using 20,000 training images and 20,000 StyleGAN-generated images. The FID scores for the MRI-StyleGAN and the CT-StyleGAN were 9.71 and 38.65 respectively. These values fall within the range of FID scores observed for GANs trained on standard MRI and CT datasets [122]. The FID score for the Face-StyleGAN as reported in [76] was 4.40. It should be noted, however, that a definitive metric for the quantitative evaluation of GANs remains an active area of research [41, 31, 122, 83].

After training the StyleGAN models, the validity of the Gaussian prior assumption on the latent space $\mathcal{V}$—and thus of the norm constraint in the PULSE CSGM in Eq. (5.1)—was investigated. The validation study was based on the following well-known theorem on standard Gaussian distributions [136]:

Theorem 5.4.1. Let $\mathbf{a} \in \mathbb{R}^d$ be a standard Gaussian vector, i.e. $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Then $\|\mathbf{a}\|_2^2 \sim \chi^2(d)$, where d is the degree of freedom of a χ^2-distribution.

As a corollary to Theorem 5.4.1, a necessary condition for a style latent vector $\mathbf{v}$ to follow a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_k)$ is that $\|\mathbf{v}\|_2^2 \sim \chi^2(k)$. With higher values of the degree of freedom k, the $\chi^2(k)$ distribution concentrates around the mode given by $\max(0, k-2)$. To verify whether this condition holds true for the pre-trained MRI, CT and Face StyleGAN models, 10^7 random realizations of a style latent vector $\mathbf{v} = \mathcal{T}G_m(\mathbf{z})$ were generated by sampling $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$. The probability density function (PDF) of $\|\mathbf{v}\|_2^2$, denoted as $\pi(\|\mathbf{v}\|_2^2)$, was then estimated from those realizations. Figure 5.1 shows a comparison between $\pi(\|\mathbf{v}\|_2^2)$ and the PDF of the $\chi^2(k)$ distribution for the MRI, CT and Face

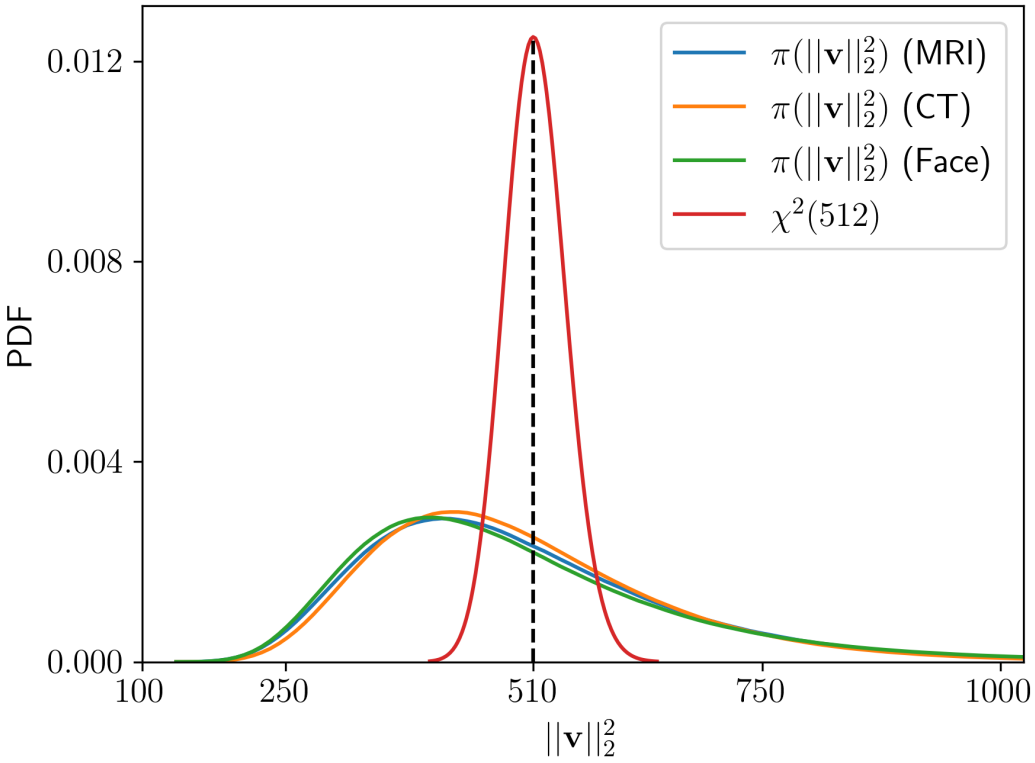


Figure 5.1: Comparison of $\pi(\|\mathbf{v}\|_2^2)$ with the PDF of $\chi^2(k)$ for the MRI, CT and Face StyleGAN models ($k = 512$). The estimated PDF $\pi(\|\mathbf{v}\|_2^2)$ has heavier tails and differs significantly from the PDF of $\chi^2(k)$ for all three models, and thus invalidates the soap bubble effect argument exploited in PULSE.

StyleGAN models, for $k = 512$. As expected, the $\chi^2(k)$ distribution is highly concentrated around the mode $k - 2$, i.e. 510. On the contrary, the estimated PDF $\pi(\|\mathbf{v}\|_2^2)$ has much heavier tails and strongly differs from the PDF of $\chi^2(k)$ for all three StyleGAN models and hence it is evident that the soap bubble effect does not manifest in the $\mathcal{V}$ latent space.

In summary, this study invalidated the previously-held assumption that $\mathbf{v} \in \mathbb{R}^k$ is a standard Gaussian random vector, and presents a strong argument against constraining each $\mathbf{v}$ to lie on the spherical surface $S^{k-1}(\sqrt{k})$. Based on this finding and the limitations of the PULSE method described in Sec. 5.3, an enhanced version of the PULSE method is proposed for empirical sampling, hereafter referred to as PULSE++.

5.5 Generating Multiple Data-Consistent Solutions using PULSE++

The proposed PULSE++ method is designed to explore the manifold of the StyleGAN more efficiently as compared to PULSE and produce diverse solutions that are application-relevant while promoting higher data consistency. The applied modifications and the re-formulation of the CSGM optimization problem in PULSE++ are detailed below.

5.5.1 Imposing Accurate Priors on Style and Noise Latent Vectors in StyleGAN

In the PULSE++ method, the style latent vectors in Eq. (5.1) are constrained to lie in $\mathcal{V}$ instead of the extended style latent space $\mathcal{V}^+$. This ensures that each alternate solution resides in the range of the StyleGAN. Consequently, the *GEOCROSS* penalty term is no longer required since the constraint $\mathbf{v} = \mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_L$ is already imposed. Additionally, regularization of the style latent and noise vectors in PULSE++ is imposed using accurate statistical knowledge instead of the approximations employed in the PULSE method. To account for the heavy tails of the estimated PDF $\pi(\|\mathbf{v}\|_2^2)$ in the CSGM optimization problem, PULSE++ replaces the spherical constraint $\mathbf{v} \in S^{k-1}(\sqrt{k})$ with a constraint in an annular region $\mathcal{A}$ defined as

$$\mathcal{A} := \{\mathbf{v} \in \mathbb{R}^k | \delta_{min} \leq \|\mathbf{v}\|_2 \leq \delta_{max}\}. \tag{5.2}$$

The inner and outer radii $\delta_{min} < \sqrt{k}$ and $\delta_{max} > \sqrt{k}$ are chosen such that the probability of $\mathbf{v}_i$ lying inside $\mathcal{A}$ is equal to a pre-determined parameter $\gamma \in (0, 1)$. Specifically, δ_{min} and δ_{max} are chosen such that

$$\Pi_{\text{emp}}(\delta_{min}) = 1 - \Pi_{\text{emp}}(\delta_{max}) = \frac{\gamma}{2}, \tag{5.3}$$

where Π_{emp} is the empirical cumulative distribution function (ECDF) of $\|\mathbf{v}\|_2$. The hyper-parameter γ can be chosen based on the desired trade-off between data consistency and the degree to which the alternate solutions are representative of the training distribution. A projection operator $\mathcal{P}_{\mathcal{A}} : \mathbb{R}^k \mapsto \mathcal{A}$ is defined to constrain $\mathbf{v}$ in the annular region $\mathcal{A}$ described

in Eq. (5.2), stated as

$$\mathcal{P}_{\mathcal{A}}(\mathbf{v}) = \begin{cases} \delta_{min}\dfrac{\mathbf{v}}{\|\mathbf{v}\|_2}, & \text{if } \|\mathbf{v}\|_2 < \delta_{min}, \\ \mathbf{v}, & \text{if } \delta_{min} \leq \|\mathbf{v}\|_2 \leq \delta_{max}, \\ \delta_{max}\dfrac{\mathbf{v}}{\|\mathbf{v}\|_2}, & \text{if } \|\mathbf{v}\|_2 > \delta_{max}. \end{cases} \tag{5.4}$$

In a conventional CSGM formulation [28], regularization of the latent vectors for which the statistical distribution is known beforehand is performed by adding a penalty term that represents the negative log-probability of the density function. In PULSE, however, regularization on the Gaussian noise latent vectors in $\mathbf{\Phi}$ was performed by employing the soap bubble effect and imposing a strict norm constraint, resulting in an approximate Gaussian prior. In order to use the full knowledge of the prior distribution of $\mathbf{\Phi}$, the strict norm constraint on $\mathbf{\Phi}$ in Eq. (5.1) was relaxed, and instead a penalty term was added in $\mathcal{R}(\mathbf{V}, \mathbf{\Phi})$ in the form of a negative log-probability density function of the Gaussian noise latent vectors in $\mathbf{\Phi}$.

Incorporating the modifications to Eq. (5.1) described above, the CSGM optimization problem in PULSE++ is stated as:

$$\begin{aligned} \hat{\mathbf{V}}, \hat{\mathbf{\Phi}} = \underset{\mathbf{V}, \mathbf{\Phi}}{\operatorname{argmin}} \mathcal{L}(\mathbf{V}, \mathbf{\Phi}) := \mathcal{J}(\mathbf{g}, \tilde{G}(\mathbf{V}, \mathbf{\Phi})) + \frac{1}{2}\sum_{i=1}^{L} \|\boldsymbol{\phi}_i\|_2^2 \\ s.t.\ \mathbf{V} \in \mathcal{V} \text{ and } \mathbf{v}_i \in \mathcal{A}\ \forall i \in \{1, \ldots, L\}, \end{aligned} \tag{5.5}$$

where $\mathcal{A}$ is defined in Eq. (5.2).

5.5.2 Two-stage Optimization Approach to Improve Stability and Data Consistency

Menon *et al.* [102] observed that jointly optimizing over the style latent vectors in $\mathbf{V}$ and all noise latent vectors in $\mathbf{\Phi}$ using stochastic gradient descent was unstable and produced visually unrealistic images. To prevent such instabilities, the PULSE method optimizes only over the noise latent vectors in $\mathbf{\Phi}$ corresponding to the 5 lowest resolution levels and keeps

the other noise latent vectors fixed. While this trick improves the synthesis quality, it significantly limits the degree of data consistency that can be achieved. We hypothesize that the unstable nature of the non-convex CSGM optimization problem when including high-resolution noise latent vectors in the search space is due to overparameterization, which allows potentially infinite global minima [109]. Under such circumstances, random initialization may produce global minima solutions that contain artifacts. We propose that the risk of encountering such artifacts can be mitigated with a better initialization in the style latent space $\mathcal{V}$ before including $\mathbf{\Phi}$ as optimization variables. There remained a scope for investigating and mitigating instabilities when optimizing over all noise latent vectors in $\mathbf{\Phi}$ in the CSGM optimization problem. When all L vectors in $\{\boldsymbol{\phi}_i\}_{i=1}^{L}$ are included in the search space, the CSGM optimization problem in Eq. (5.1) is overparameterized. For example, the MRI-StyleGAN described in Sec. 5.4 generates images with size $N = 256^2 = 65536$, while $\mathbf{\Phi}$ has a total of 109220 variables. Hence, if all the L noise latent vectors in $\mathbf{\Phi}$ are included in the search space of Eq. (5.1), the non-convex CSGM optimization problem is overparameterized and may permit infinitely many *global* minima. Oymak and Soltanolkotabi [109] demonstrated that in such cases, the iterates obtained using first-order stochastic gradient methods may converge fast to global minima that are nearest to the initial point. In the case of ill-posed inverse problems such as image-superresolution or image reconstruction, these global minima may represent different data-consistent solutions. However, due to the highly expressive and corrugated nature of the manifold of the StyleGAN [76], many of these global minima solutions may contain significant artifacts [3]. Since a global minimum solution would lie close to the initial point, random initialization would increase the likelihood of producing such undesirable global minima solutions. Instead, if the optimization problem is initialized at a point lying close to the projection of the measured object on the style latent space before optimization of $\mathbf{\Phi}$, the global minimum will be close to the projection and reduce the risk of unknown artifacts while ensuring data consistency. Towards that end, a two-stage optimization approach is proposed in PULSE++. In the first stage, optimization was performed only on $\mathbf{V}$ to find a good initial vector in the style latent space $\mathcal{V}$ which agrees well with the measurements, while the noise latent vectors were randomly sampled and kept fixed. At the end of this stage, the intermediate estimate was denoted as $\hat{\mathbf{f}}_1 := \tilde{G}(\hat{\mathbf{V}}_1, \overline{\mathbf{\Phi}})$, where $\overline{\mathbf{\Phi}}$ were the randomly initialized noise latent vectors. If the data fidelity of $\hat{\mathbf{f}}_1$ did not satisfy the pre-set tolerance level ϵ_1 (c.f. Section 5.5.4), the solution was rejected, otherwise optimization progressed to the final stage where $\mathbf{V}$ and *all* noise latent vectors in $\mathbf{\Phi}$ were jointly optimized. Finally, the approximate solution $\hat{\mathbf{f}}_2 := \tilde{G}(\hat{\mathbf{V}}_2, \hat{\mathbf{\Phi}}_2)$ was obtained when the

maximum number of iterations was reached. The approximate solution $\hat{\mathbf{f}}_2$ was considered to be data-consistent if the data fidelity term $\mathcal{J}(\mathbf{g}, \hat{\mathbf{f}}_2)$ was less than a more strict tolerance level ϵ_2 (c.f. Section 5.5.4). The complete procedure for performing empirical sampling with PULSE++ is detailed in Algorithm 2. The function *Optimize()* in Algorithm 2 which approximately solved the CSGM optimization problem in each stage is described in the next section.

Algorithm 2: Empirical sampling with PULSE++

Input: Measurement data $\mathbf{g}$, forward operator $\mathcal{H}(\cdot)$, objective function $\mathcal{L}(\mathbf{V}, \mathbf{\Phi})$ from Eq. (5.5); projection operator $\mathcal{P}_{\mathcal{A}}(\mathbf{V})$ from Eq. (5.4); annulus parameter γ, learning rate lr of Adam optimizer, number of steps in first ($n^{(1)}_{steps}$) and second ($n^{(2)}_{steps}$) optimization stages; number of alternate solutions T; acceptance tolerances ϵ_1 and ϵ_2

Output: Set $\hat{\mathbf{F}}$ of data consistent object estimates

$\hat{\mathbf{F}} \leftarrow \{\}$
for $t \in \{1, \ldots, T\}$ **do**
  Initialize $\mathbf{V}^{[0]}$ and $\mathbf{\Phi}^{[0]}$ with *i.i.d.* entries $\sim \mathcal{N}(0, 1)$
  `// Optimization over {V} only`
  $\hat{\mathbf{V}}_1 \leftarrow$ *Optimize*$(\mathcal{L}(\cdot, \mathbf{\Phi}^{[0]}), \{\mathbf{V}\}, n^{(1)}_{steps})$
  $\hat{\mathbf{f}}_{t1} \leftarrow \tilde{G}(\hat{\mathbf{V}}_1, \mathbf{\Phi}^{[0]})$
  if $\mathcal{J}(\mathbf{g}, \hat{\mathbf{f}}_{t1}) > \epsilon_1$ **then**
    break `/* Reject solution */`
  end
  `// Joint optimization over {V, Φ}`
  $\{\hat{\mathbf{V}}_2, \hat{\mathbf{\Phi}}_2\} \leftarrow$ *Optimize*$(\mathcal{L}(\cdot, \cdot), \{\hat{\mathbf{V}}, \mathbf{\Phi}^{[0]}\}, n^{(2)}_{\text{steps}})$
  $\hat{\mathbf{f}}_{t2} \leftarrow \tilde{G}(\hat{\mathbf{V}}_2, \hat{\mathbf{\Phi}}_2)$
  if $\mathcal{J}(\mathbf{g}, \hat{\mathbf{f}}_{t2}) \leq \epsilon_2$ **then**
    $\hat{\mathbf{F}} \leftarrow \hat{\mathbf{F}} \cup \{\hat{\mathbf{f}}_{t2}\}$
  end
end
return $\hat{\mathbf{F}}$

5.5.3 Solution of the CSGM Optimization Problem

The CSGM optimization problem in Eq. (5.5) was approximately solved in a sequential manner in accordance with the two-stage approach outlined in Sec. 5.5.2. Projected gradient descent was employed using the Adam method to approximately solve the optimization

problem. Initially, the optimization variables $\boldsymbol{\Theta}$ included only the style latent matrix $\mathbf{V}$, and the noise latent vectors in $\boldsymbol{\Phi}$ were subsequently added to $\boldsymbol{\Theta}$ at the end of the first stage of optimization. The values of the optimization variables $\boldsymbol{\Theta}^{[j]}$ at iteration j were first updated using one step of the Adam algorithm. Denoting the gradient update of $\mathbf{V}^{[j]}$ as $\mathbf{V}^{[j+\frac{1}{2}]}$, the projection operator in Eq. (5.4) was then applied independently on each of the L latent vectors that represented the columns of $\mathbf{V}^{[j+\frac{1}{2}]}$. The projection operations were performed in order to enforce the necessary constraint on the ℓ_2-norm of each column of the style latent matrix $\mathbf{V}^{[j+\frac{1}{2}]}$ that was defined by the annular region $\mathcal{A}$. For notational convenience, these L independent projection operations will be simply denoted as $\mathbf{V}^{[j+1]} \leftarrow \mathcal{P}_{\mathcal{A}}(\mathbf{V}^{[j+\frac{1}{2}]})$. Similarly, with a slight abuse of notation, $\{\mathbf{V}^{[j+1]}, \boldsymbol{\Phi}^{[j+1]}\} \leftarrow \mathcal{P}_{\mathcal{A}}(\{\mathbf{V}^{[j+\frac{1}{2}]}, \boldsymbol{\Phi}^{[j+\frac{1}{2}]}\})$ will indicate the projection of the latent style vectors $\mathbf{V}^{[j+\frac{1}{2}]}$ leaving the latent noise vectors $\boldsymbol{\Phi}$ unalterated. The Adam update followed by the projection operation $\mathcal{P}_{\mathcal{A}}(\{\mathbf{V}^{[j+\frac{1}{2}]}, \boldsymbol{\Phi}^{[j+\frac{1}{2}]}\})$ constituted a single projected-gradient step of the optimization procedure that produced the next iterates $\mathbf{V}^{[j+1]}$ and $\boldsymbol{\Phi}^{[j+1]}$. It should be noted that the objective function in Eq. (5.5) may not decrease monotonically with each such projected-gradient step[115]. Thus, the current best estimates $\hat{\mathbf{V}}_i$ and $\hat{\boldsymbol{\Phi}}_i$ of Eq. (5.5) were updated with the iterates $\mathbf{V}^{[j+1]}$ and $\boldsymbol{\Phi}^{[j+1]}$ only if $\mathcal{L}(\mathbf{V}^{[j+1]}, \boldsymbol{\Phi}^{[j+1]}) < \mathcal{L}(\hat{\mathbf{V}}_i, \hat{\boldsymbol{\Phi}}_i)$, where the subscript i is 1 and 2 for the first stage and second stage of optimization respectively. The function *Optimize()* for performing the projected gradient descent steps as described above is outlined in Algorithm 3.

Algorithm 3: Projected gradient descent optimizer for PULSE++

Function *Optimize*($\mathcal{L}(\cdot)$, $\boldsymbol{\Theta}$, n_{steps})**:**
  Initialize $\hat{\boldsymbol{\Theta}} \leftarrow \boldsymbol{\Theta}$, $\boldsymbol{\Theta}^{[0]} \leftarrow \boldsymbol{\Theta}$
  for $j \in \{0, \ldots, n_{\text{steps}}\}$ **do**
    $\boldsymbol{\Theta}^{[j+\frac{1}{2}]} \leftarrow \text{Adam}(\mathcal{L}(\boldsymbol{\Theta}^{[j]}))$
    $\boldsymbol{\Theta}^{[j+1]} \leftarrow \mathcal{P}_{\mathcal{A}}(\boldsymbol{\Theta}^{[j+\frac{1}{2}]})$
    if $\mathcal{L}(\boldsymbol{\Theta}^{[j+1]}) < \mathcal{L}(\hat{\boldsymbol{\Theta}})$ **then**
      $\hat{\boldsymbol{\Theta}} \leftarrow \boldsymbol{\Theta}^{[j+1]}$
    end
  end
  return $\hat{\boldsymbol{\Theta}}$

5.5.4 Establishing Rules for Accepting Data-Consistent Alternate Solutions

The tolerances ϵ_1 and ϵ_2 used in Algorithm 2 are defined via a generalization of the Morozov discrepancy principle [105] to account for the finite capacity of the generative model $\tilde{G}$ [28]. To this aim, given an object $\mathbf{f} \in \mathbb{R}^N$, sampled from the object distribution $\mathcal{D}$, two embeddings $\mathbf{f}_1, \mathbf{f}_2 \in \mathbb{R}^N$ are constructed as follows. The image $\mathbf{f}_1 = \tilde{G}(\hat{\mathbf{V}}_1, \overline{\mathbf{\Phi}})$ is the direct embedding of $\mathbf{f}$ on the range of $\tilde{G}$ when the representation error $\|\mathbf{f} - \tilde{G}(\mathbf{V}, \overline{\mathbf{\Phi}})\|_2^2$ is minimized only with respect to the latent style vector $\mathbf{V}$ for a randomly chosen (but fixed) $\overline{\mathbf{\Phi}}$. Similarly, the image $\mathbf{f}_2 = \tilde{G}(\hat{\mathbf{V}}_2, \hat{\mathbf{\Phi}}_2)$ is the direct embedding of $\mathbf{f}$ on the range of $\tilde{G}$ when the representation error is jointly minimized with respect to both $\mathbf{V}$ and $\mathbf{\Phi}$. Then the tolerances ϵ_1 and ϵ_2 are defined as

$$\epsilon_i := \mathbb{E}_{\mathbf{f},\mathbf{g}}\big[\mathcal{J}(\mathbf{g}, \mathbf{f})\big] + \mathbb{E}_{\mathbf{f}}\big[\mathcal{J}(\mathbf{H}\mathbf{f}, \mathbf{f}_i)\big], \quad i = 1, 2. \tag{5.6}$$

Above, the first term represents the expected value of the data fidelity $\mathcal{J}$ over the joint distribution of objects $\mathbf{f}$ and corresponding noisy measurement data $\mathbf{g}$. The second term stems for the representation error of the GAN and it is given by the expected value over the object distribution of the mismatch between the output of the imaging operator $\mathbf{H}$ applied to the object $\mathbf{f}$ and its embedding $\mathbf{f}_i$. For certain measurement noise distributions, the term $\mathbb{E}_{\mathbf{f},\mathbf{g}}\big[\mathcal{J}(\mathbf{g}, \mathbf{H}\mathbf{f})\big]$ is analytically defined. For example, in the case of Gaussian additive noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_N)$, $\mathbb{E}_{\mathbf{f},\mathbf{g}}\big[\mathcal{J}(\mathbf{g}, \mathbf{f})\big] = \frac{M}{2}$ where M is the number of measurements. For other measurement noise distribution and for the term stemming from representation error, Eq. (5.6) is numerically evaluated by replacing expectations over the object and data distribution with a Monte Carlo estimate from samples of a held-out test dataset not included during training of the StyleGAN.

5.6 Numerical Studies

Numerical studies were conducted to demonstrate the ability of the proposed PULSE++ method to produce multiple data-consistent solutions from the same tomographic measurements. Two stylized tomographic imaging systems were considered: one acquiring incomplete Fourier space measurements (Sec. 5.6.1) and the other acquiring X-ray fan-beam CT projection data (Sec. 5.6.2). The advantage of PULSE++ over PULSE with respect to

preserving data consistency is established from these numerical studies. The PULSE and PULSE++ methods were established by adapting an open-source implementation of PULSE for SISR in PyTorch [103]. Since the training of a StyleGAN does not require any knowledge of the measurement process, the same pre-trained StyleGAN was employed for different sampling conditions in the PULSE and PULSE++ methods. The network architecture and pre-trained weights of MRI-StyleGAN and CT-StyleGAN were transferred from TensorFlow to PyTorch [27].

For purposes of comparison, alternate solutions were also computed from incomplete Fourier space measurements by implementing a recently proposed approximate posterior sampling method [68] that employs a score-based diffusion model and annealed Langevin dynamics [123]. This approximate posterior sampling method will be referred to as diffusion posterior sampling (DPS) in the studies below. The DPS method requires training a state-of-the-art score-based generative model known as NCSNv2 [124], which was performed by adapting an open-sourced implementation. The same training dataset of axial knee MRI images that was used to train MRI-StyleGAN was also employed to train the NCSNv2 model. The DPS method was performed by adapting a previous implementation [67]. Similar to StyleGAN, the NCSNv2 model is also trained independent of the measurement process, and hence DPS was implemented with the same pre-trained NCSNv2 for different sampling conditions. The code repository for the numerical studies in our paper has been published [25].

5.6.1 Stylized Imager that Acquires Incomplete Fourier Space Measurements

A stylized imaging system that acquires incomplete 2D Fourier space (or k-space) measurements was considered. It should be noted that the goal of these preliminary studies was only to demonstrate and compare the performance of the proposed PULSE++ method against the original PULSE method and score-based posterior sampling, with respect to producing multiple data-consistent solutions from the same measurement data acquired under identical conditions. Hence, there is no attempt to model the real-world complexities of data-acquisition in MRI. Two different axial knee images of size 256×256 that belong to the NYU fastMRI dataset were considered to serve as objects $\mathbf{f}$, as shown in Fig. 5.2. It should be noted that the images representing these objects, denoted as Knee 1 and Knee 2, were not included in the training dataset for the MRI-StyleGAN and the NCSNv2 models.

Incomplete and noisy k-space data were simulated from these two objects for demonstrating the numerical studies. The acceleration factor was defined as $R = N/M > 1$, where M denotes the number of k-space samples measured and N is the dimension of the object. The imaging operator was modeled as $\mathcal{H} = \mathbf{M}\mathcal{F}$, where $\mathcal{F}$ denotes the 2D Fast Fourier Transform (FFT) and the binary matrix $\mathbf{M} \in \{0,1\}^{M\times N}$ represents a random Cartesian k-space sampling mask. The measurement noise $\mathbf{n} \in \mathbb{E}^M$ was sampled from an i.i.d complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_M)$ [8], i.e. the real and imaginary components of $\mathbf{n}$ are Gaussian with zero mean and $\frac{\sigma^2}{2}$ variance. In the numerical experiments, the size of the image was set to $N = 256^2$. Two different acceleration factors $R = 3, 4$ and two levels of noise $\sigma = 0.03, 0.05$ were investigated. The random Cartesian sampling masks were generated using open-sourced codes [147]. The MRI-StyleGAN model described in Sec. 5.4 was employed for performing empirical sampling with PULSE and PULSE++ using the same k-space data for each combination of σ and R. Since the measurement noise was i.i.d. Gaussian, the data fidelity term was specified as $\mathcal{J}(\mathbf{g}, \hat{\mathbf{f}}) = \frac{1}{\sigma^2}\|\mathbf{g} - \mathcal{H}\hat{\mathbf{f}}\|_2^2$, where $\mathbf{g} \in \mathbb{R}^M$ is the k-space data and $\hat{\mathbf{f}} \in \mathbb{R}^N$ denotes the estimated object. The learning rate lr of the Adam optimizer in the CSGM formulation for the PULSE and PULSE++ methods was set as 0.4, similar to the original implementation of PULSE in [102]. The total number of iterations performed to obtain each alternate solution was 5000, out of which the first $n^{(1)}_{\text{steps}} = 4500$ steps involved optimization over only $\mathbf{V}$, followed by joint optimization over $\mathbf{V}$ and $\mathbf{\Phi}$ for the remaining $n^{(2)}_{\text{steps}} = 500$ iterations. The initial step size η for annealed Langevin dynamics in DPS [123] was 5×10^{-8}. An alternate solution obtained using either PULSE or PULSE++ was completed in $\sim$5 minutes, while each alternate solution from the score-based posterior sampling method using the NCSNv2 model was computed in $\sim$7 minutes, with all the experiments performed using an NVIDIA 1080 Ti GPU.

5.6.2 CT Imaging System with Limited Angular Range

Numerical studies were also performed with incomplete and noisy X-ray CT measurements from a stylized CT imaging system. The objective of these studies was to demonstrate the ability of the PULSE++ method to perform empirical sampling in higher-dimensional spaces (resolution 512×512 pixels) and imaging systems with non-Gaussian measurement noise distribution, in a computationally feasible manner. The CT-StyleGAN model introduced in Sec. 5.4 was employed for these studies to find alternate solutions from the same X-ray

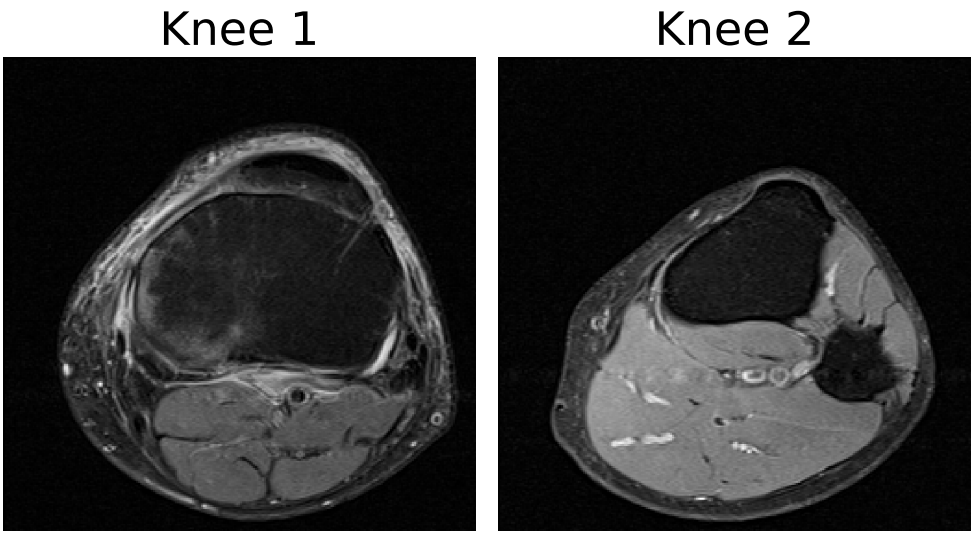


Figure 5.2: The objects Knee 1 and Knee 2 with size 256×256 from which noisy and incomplete k-space measurements were generated. Both the objects are displayed in the grayscale range of $[0, 1]$.

photon projection data. Two separate CT lung images of size 512×512 pixels were selected from the NIH DeepLesion dataset to represent objects $\mathbf{f}$ from which measurement data were simulated. It should be noted that these images were not included during training of the StyleGAN. The objects, denoted as Lung 1 and Lung 2, are shown in Fig. 5.3. The maximum linear attenuation coefficient values in Lung 1 and Lung 2 were 0.046 mm^{-1} and 0.063 mm^{-1}, respectively. The physical unit of each pixel (px) was 0.82 mm [145]. A fan-beam geometry with a linear detector array and a monoenergetic source was assumed. Projection data were simulated for 120 views spanning the limited angular range $[0°, 119°]$. The noiseless X-ray measurements $\bar{\mathbf{g}} \in \mathbb{R}^M$ from an object $\mathbf{f} \in \mathbb{R}^N$ were modeled as [42]

$$\bar{\mathbf{g}} = I_0 \exp(-\mathbf{H}\mathbf{f}), \tag{5.7}$$

where the system matrix $\mathbf{H} \in \mathbb{E}^{M \times N}$ is the fan-beam projector and I_0 is the intensity of an unattenuated beam. The fan-beam projector $\mathbf{H}$ was implemented using the Air Tools II library [56]. The noisy intensity measurements $\mathbf{g} \in \mathbb{R}^M$ were Poisson-distributed with a mean of $\bar{\mathbf{g}}$ [91]. Higher values of I_0 result in a higher signal-to-noise ratio (SNR) in the intensity measurements. Since the aim of this simulation study was primarily to assess the ability of the PULSE++ method to perform empirical sampling with high-dimensional objects, additional physical factors required to accurately model a real-world CT imaging system

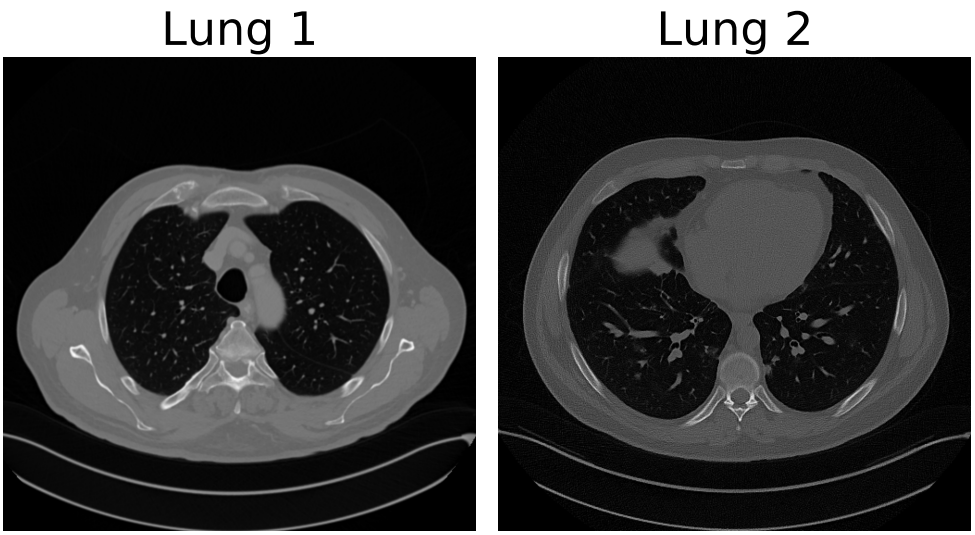


Figure 5.3: The objects Lung 1 and Lung 2 with size 512×512 from which noisy and incomplete X-ray projection data were generated. Both the objects are displayed in the grayscale range of $[0, 1]$.

such as beam spectrum, photon scattering and dark current effects were not considered. Numerical studies were conducted using simulated projection data with values of $I_0 = 10^3$ and $I_0 = 10^5$, which correspond to measurement data that have different levels of photon noise. The data fidelity term $\mathcal{J}(\mathbf{g}, \hat{\mathbf{f}})$ in Eq. (5.5) was defined as the Kullback-Leibler (KL) divergence between the noisy measurement data $\mathbf{g}$ and the forward projection data $\hat{\mathbf{g}} = I_0 \exp(-\mathbf{H}\hat{\mathbf{f}})$ [70]. The number of gradient descent iterations was $n_{steps}^{(1)} + n_{steps}^{(2)} = 1000$, of which the initial $n_{steps}^{(1)} = 900$ iterations were performed in the style latent space $\mathcal{V}$ only. Each CSGM run of the PULSE++ method with the CT measurement data completed in ~ 9 minutes on an NVIDIA 1080 Ti GPU.

5.6.3 Ablation Study

An ablation study was performed to critically assess the impact of each of the enhancements introduced to develop the PULSE++ method from the baseline PULSE method. Intermediate methods that represent each of these enhancements were implemented as described below:

PULSE$_1$

The inaccurate strict norm constraint $\mathbf{v}_i \in S^{k-1}(\sqrt{k})$ in Eq. (5.1) was replaced with the constraint $\mathbf{v}_i \in \mathcal{A}$, where $\mathcal{A}$ is an annulus as defined in Eq. (5.2). Additionally, the constraint $\boldsymbol{\phi}_i \in S^{p_i-1}(\sqrt{p_i})$ denoting an approximate Gaussian prior in PULSE was substituted with the statistically consistent log-probability density function penalty:

$$\begin{aligned}\hat{\mathbf{V}}, \hat{\boldsymbol{\Phi}} = \underset{\mathbf{V},\boldsymbol{\Phi}}{\operatorname{argmin}} \mathcal{L}(\mathbf{V}, \boldsymbol{\Phi}) := \Big\{\mathcal{J}(\mathbf{g}, \tilde{G}(\mathbf{V}, \boldsymbol{\Phi})) + \mathcal{R}(\mathbf{V}, \boldsymbol{\Phi})\Big\}, \\ s.t. \mathbf{V} \in \mathcal{V}^+ \; \mathbf{v}_i \in \mathcal{A} \; \forall i \in \{1, \ldots, L\},\end{aligned} \tag{5.8}$$

where the regularization term was given by

$$\mathcal{R}(\mathbf{V}) = \lambda_c CROSS(\mathbf{V}) + \frac{1}{2}\sum_{i=1}^{L} \|\boldsymbol{\phi}_i\|_2^2. \tag{5.9}$$

Above, the pairwise Euclidean distance $CROSS(\mathbf{V}) \equiv \sum_{i=1}^{L-1}\sum_{j=i+1}^{L} \|\mathbf{v}_i - \mathbf{v}_j\|_2^2$ [102] was used in place of the $GEOCROSS$ term in the original PULSE formulation since latent vectors in $\mathcal{A}$ may have different norms, and λ_c is the regularization hyperparameter.

PULSE$_2$

The PULSE optimization problem in Eq. (5.1) is performed in the style latent space $\mathcal{V}$ instead of $\mathcal{V}^+$ to ensure the solution lies exactly on the StyleGAN generator's manifold and the $GEOCROSS$ penalty term was removed.

PULSE$_3$

The PULSE$_1$ optimization problem in Eq. (5.8) was performed in the style latent space $\mathcal{V}$ instead of $\mathcal{V}^+$ to ensure the solution lies exactly on the StyleGAN generator's manifold and the $CROSS$ penalty term was removed.

PULSE$_4$

The PULSE optimization problem in Eq. (5.8) was performed in the style latent space $\mathcal{V}$ and all the noise latent vectors in $\mathbf{\Phi}$ were included as optimization variables. The optimization problem was approximately solved in two stages with $\mathbf{\Phi}$ included as optimization variables only in the second stage.

The combination of all the enhancements as described above represents the PULSE++ method.

5.6.4 Uncertainty Quantification

The *uncertainty map* $\hat{\mathbf{f}}^{UM}$ was computed as the pixel-wise standard deviation of the alternate solutions $\{\hat{\mathbf{f}}_t\}_{t=1}^{T}$, where T is the number of alternate solutions. Additionally, uncertainty maps were computed separately for the measurable and null space components [19] of the alternate solutions. The measurable and null space component of an object $\mathbf{f} \in \mathbb{R}^N$ in the domain of system matrix $\mathbf{H} \in \mathbb{E}^{M \times N}$ are defined as $\mathbf{f}_{meas} = \mathbf{H}^{+}\mathbf{H}\mathbf{f}$ and $\mathbf{f}_{null} = [\mathbf{I}_N - \mathbf{H}^{+}\mathbf{H}]\mathbf{f}$ respectively, where $\mathbf{H}^{+}$ is the Moore-Penrose pseudoinverse of $\mathbf{H}$. The null space component $\mathbf{f}_{null}$ is "invisible" to $\mathbf{H}$, and only $\mathbf{f}_{meas}$ contributes to the forward data $\mathbf{H}\mathbf{f}$. The uncertainty maps of the measurable and null space components of alternate solutions were denoted as $\hat{\mathbf{f}}_{meas}^{UM}$ and $\hat{\mathbf{f}}_{null}^{UM}$ respectively. If multiple solutions are consistent with the same measurement data, it is expected that the variability expressed by $\hat{\mathbf{f}}_{null}^{UM}$ will be higher than that expressed by $\hat{\mathbf{f}}_{meas}^{UM}$. Additionally, three figures-of-merit (FOMs) were computed from the uncertainty maps to characterize the degree of variability associated with the alternate solutions obtained via each method. The *total uncertainty* FOM $\|\hat{\mathbf{f}}^{UM}\|_2^2$ is the total estimated variance from all the pixels in an alternate solution. Similarly, the uncertainty FOMs associated with the measurable and null space components of alternate solutions are $\|\hat{\mathbf{f}}_{meas}^{UM}\|_2^2$ and $\|\hat{\mathbf{f}}_{null}^{UM}\|_2^2$ respectively. It should be noted that, in theory, $\|\hat{\mathbf{f}}^{UM}\|_2^2 = \|\hat{\mathbf{f}}_{meas}^{UM}\|_2^2 + \|\hat{\mathbf{f}}_{null}^{UM}\|_2^2$ since $\hat{\mathbf{f}}_{meas}$ and $\hat{\mathbf{f}}_{null}$ are orthogonal to each other. However, there may be small discrepancies between those quantities due to floating point arithmetic and numerical approximations in the iterative computation of $\hat{\mathbf{f}}_{meas}$ and $\hat{\mathbf{f}}_{null}$ [19].

5.7 Results

5.7.1 Empirical Sampling from Fourier Space Measurements

Visual Assessment

Samples of alternate solutions generated by PULSE++ ($\gamma = 0.1$), PULSE ($\lambda_g = 10^4$) and DPS ($\eta = 5 \times 10^{-8}$) from the same k-space data produced by Knee 1, corresponding to $R = 3$ and $\sigma = 0.03$, are shown in Fig. 5.4. For each method, the alternate solutions exhibit considerable diversity while being produced by use of the same measurement data. However, among the three methods, only the alternate solutions produced by the PULSE++ method satisfied the stipulated data consistency criterion based on Morozov's discrepancy principle. This establishes an advantage of the PULSE++ method over PULSE and DPS in preserving the data consistency of alternate solutions.

Figure 5.5 shows samples of data-consistent alternate solutions obtained with PULSE++ $\{\gamma = 0.1\}$ from k-space produced by Knee 1 when the sampling pattern or the noise level is varied, e.g. with sampling conditions $\{R = 3, \sigma = 0.05\}$ and $\{R = 4, \sigma = 0.03\}$. Additionally, using the same pre-trained StyleGAN model, data-consistent alternate solutions were obtained with PULSE++ $\{\gamma = 0.9\}$ from k-space data with $R = 3$ and $\sigma = 0.03$ corresponding to Knee 2 (Fig. 5.2), as shown in Fig. 5.6.

Ablation Study

An ablation study was performed to comprehensively assess the improvement in data consistency yielded by the PULSE++ method, as outlined in Sec. 5.6.3, using alternate solutions from the k-space data corresponding to Knee 1 and parameters $R = 3$ and $\sigma = 0.03$. For evaluation, 100 alternate solutions were computed using each of the methods PULSE ($\lambda_g = 2 \times 10^5$), PULSE$_1$($\gamma = 0.1, \lambda_c = 10^3$), PULSE$_2$, PULSE$_3$($\gamma = 0.1$), PULSE$_4$($\lambda_g = 2 \times 10^5$) and PULSE++ ($\gamma = 0.1$). A box plot of the data fidelity values obtained with each method is shown in Fig. 5.7. It was observed that the improved statistical assumptions about the style latent vectors $\mathbf{V}$ and noise latent vectors $\mathbf{\Phi}$ produced significantly lower data fidelity values with less variance, resulting in higher data consistency and stability. The combination

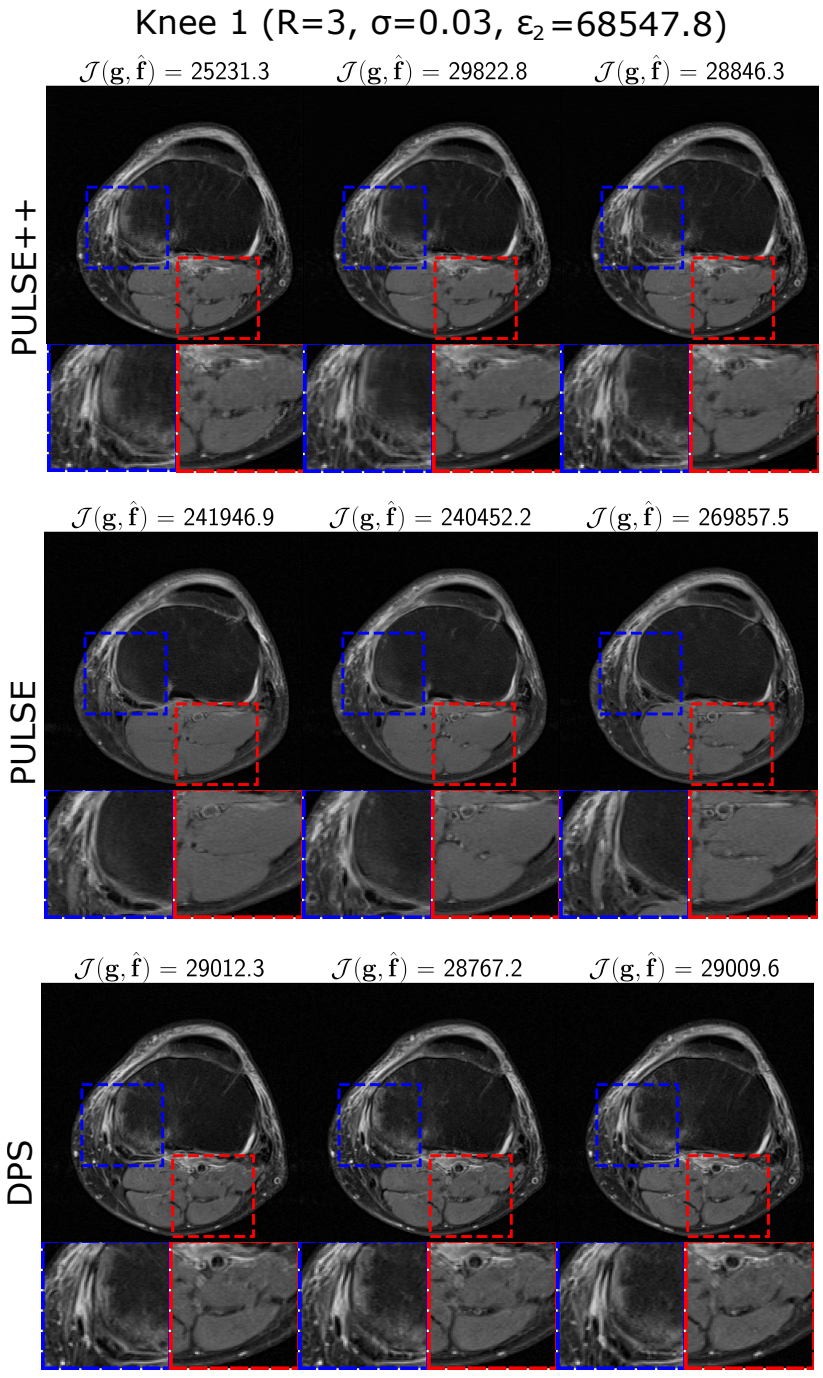


Figure 5.4: Samples of alternate solutions obtained from the same k-space data produced from Knee 1 for $R = 3$ and $\sigma = 0.03$. Zoomed-in images are shown below each alternate solution that demonstrate distinct structures. However, the alternate solutions produced by PULSE have significantly higher data fidelity compared to PULSE++ and DPS methods, which produced data-consistent solutions. All the alternate solutions are displayed in the grayscale range $[0, 1]$.

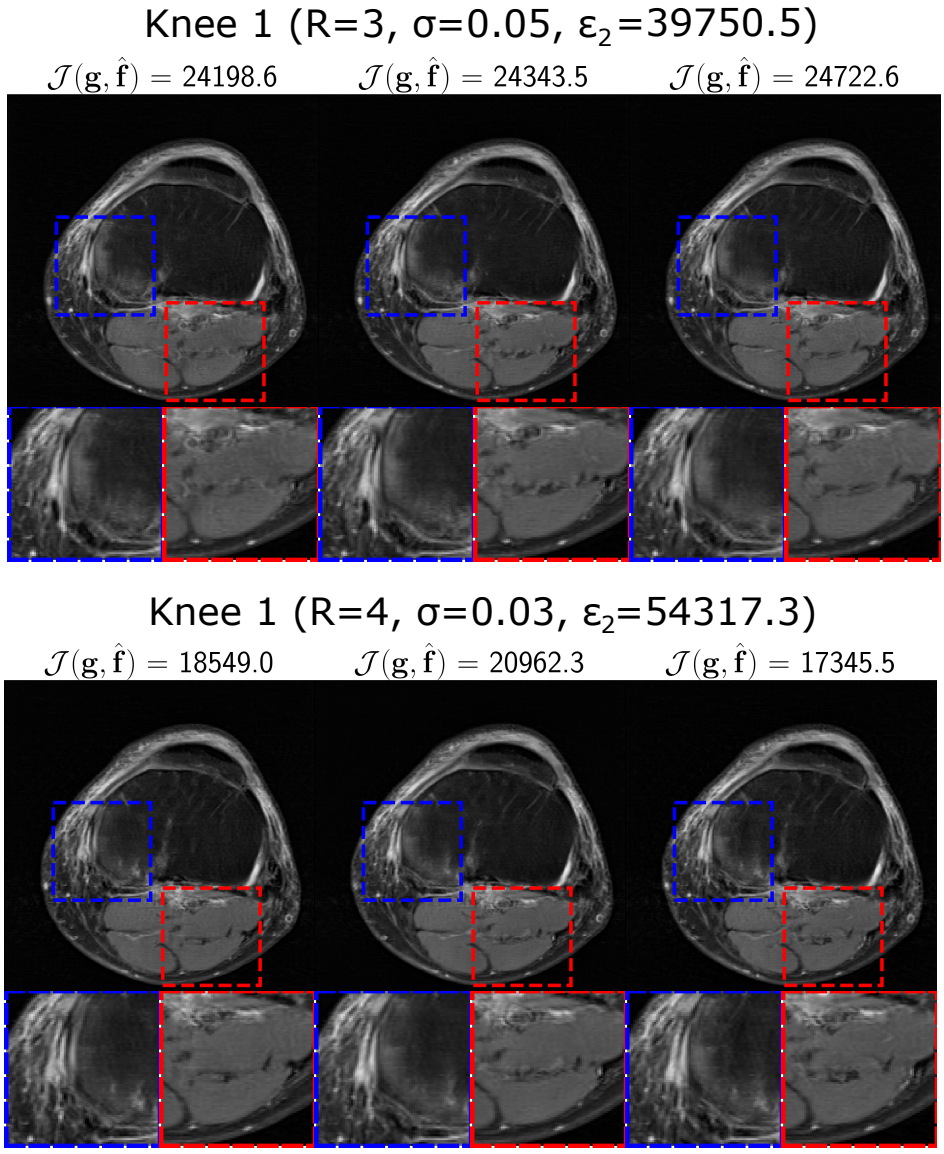


Figure 5.5: Alternate data-consistent solutions obtained using PULSE++ from k-space data produced by Knee 1 for different sampling conditions $\{R = 3, \sigma = 0.05\}$ (top) and $\{R = 4, \sigma = 0.03\}$ (bottom) using the same MRI-StyleGAN model as in Fig. 5.4. Zoomed-in images are shown below each alternate solution that demonstrate distinct structures. The alternate solutions are displayed in the grayscale range $[0, 1]$.

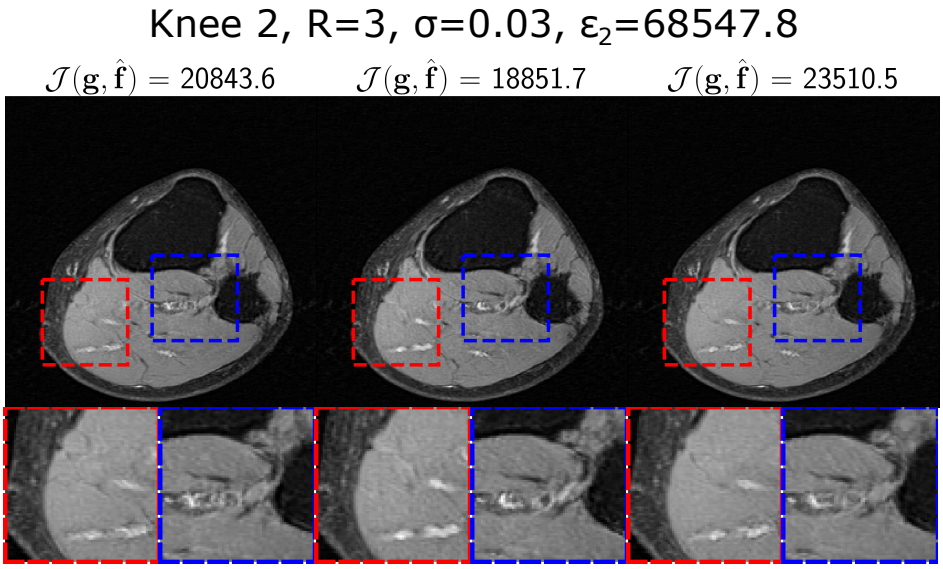


Figure 5.6: PULSE++ can produce alternate data-consistent solutions for k-space data from different objects within the same distribution on which the StyleGAN is trained, as shown here for Knee 2. Zoomed-in images are shown below each alternate solution that demonstrate distinct structures. All the alternate solutions are displayed in the grayscale range $[0, 1]$.

of all the enhancements introduced in PULSE++ leads to more stable solutions that lie in the range of the StyleGAN generator while being data-consistent.

Comparison of Data Consistency

The data fidelity from 100 samples obtained using the PULSE++, PULSE and DPS methods corresponding to k-space from Knee 1 with parameters $R = 3$ and $\sigma = 0.03$ is shown in the box plot in Fig. 5.8. It was observed that the PULSE++ method produced data-consistent solutions, while the PULSE method produced significantly higher data fidelity values and could not achieve the desired data consistency. This demonstrated the impact of imposing more accurate statistical assumptions on the style and noise latent spaces of the StyleGAN and introduction of the stable two-stage optimization procedure in PULSE++. Along with PULSE++, the DPS method also achieved data-consistent solutions which are visually plausible, as demonstrated in Fig. 5.4.

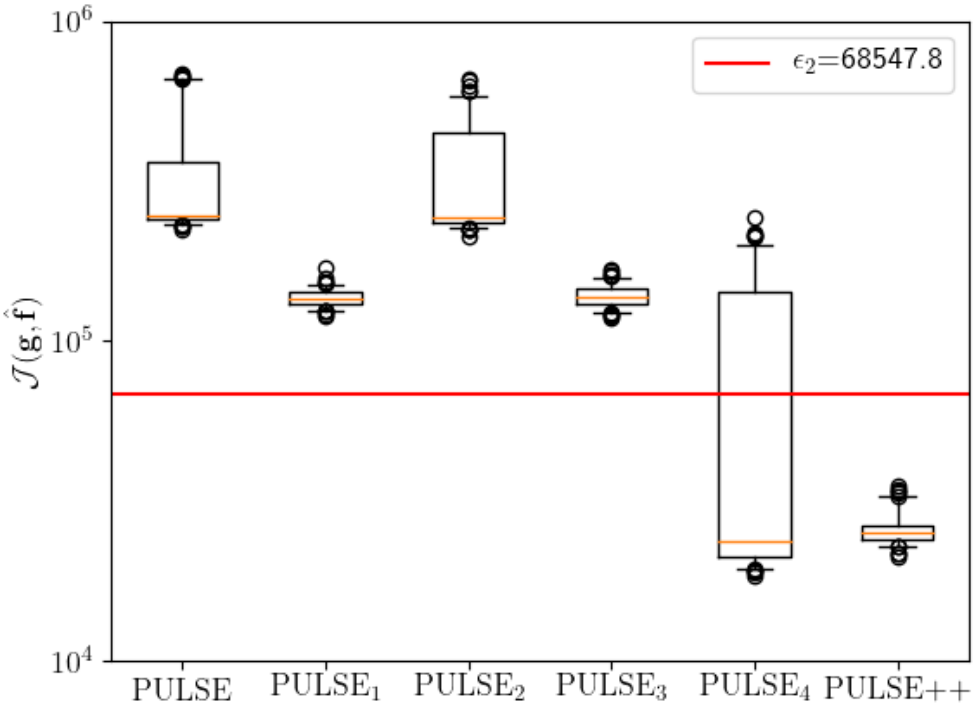


Figure 5.7: A box plot of data fidelity values of alternate solutions obtained with different methods from the k-space corresponding to Knee 1 with $R = 3$ and $\sigma = 0.03$. For each method, data fidelity values of 100 alternate solutions were plotted. The methods include PULSE, the different variants of PULSE as described in Sec. 5.6.3 that represent each enhancement and PULSE++. The plot demonstrates the improvement in data consistency and stability achieved with the modifications introduced in regularization and optimization space in PULSE to produce the PULSE++ method.

Uncertainty Quantification

From each set of k-space data, uncertainty maps were computed from $T = 100$ alternate solutions obtained with each of the three methods – PULSE++, PULSE and DPS. The uncertainty maps corresponding to Knee 1 and system parameters $R = 3$ and $\sigma = 0.03$ are shown in Fig. 5.9. For the PULSE++ and DPS methods, it was observed that uncertainty was primarily due to variations in the null space component, while uncertainty estimated using PULSE is mainly due to variations in the measurable component. This demonstrates that the canonical PULSE method is unsuitable for uncertainty estimation and confirms the need for PULSE++ for more accurate uncertainty estimation.

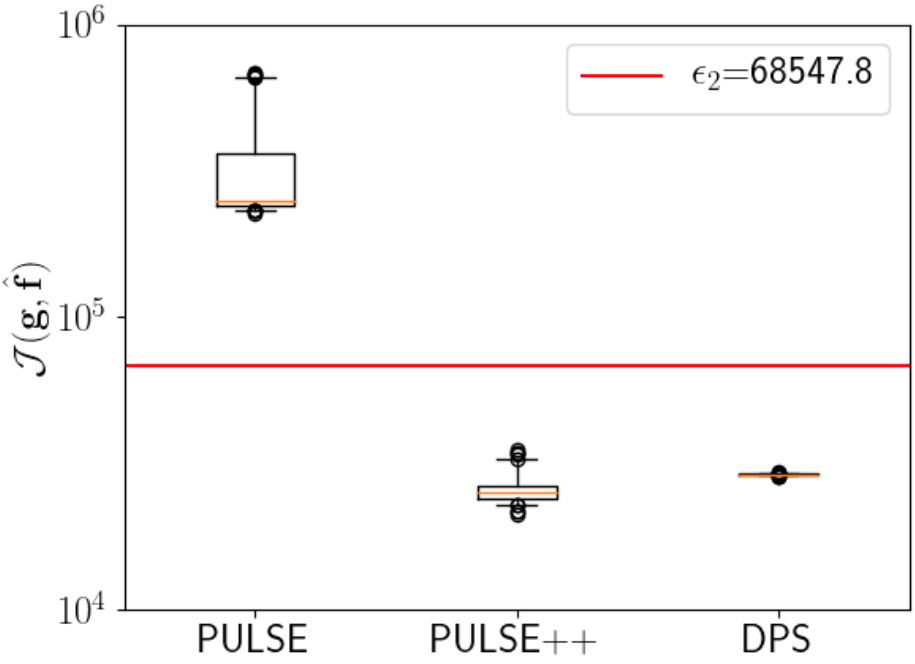


Figure 5.8: A box plot of data fidelity values of alternate solutions obtained using PULSE, PULSE++ and DPS methods from the k-space corresponding to Knee 1 with $R = 3$ and $\sigma = 0.03$. The plot demonstrates the ability of PULSE++ to achieve data-consistent solutions as opposed to PULSE, and validates the modifications introduced in regularization and optimization space in PULSE++. Alternate solutions obtained using the DPS method are also data-consistent.

Table 5.1 summarizes the uncertainty FOMs of the alternate solutions and their measurable and null space components, corresponding to PULSE++, PULSE and DPS methods for objects Knee 1 and Knee 2 with different system parameter settings. In both PULSE++ and DPS, the uncertainty FOM in the null space component was significantly higher than compared to that in the measurable component, while it was the opposite for the PULSE method. This is consistent with the uncertainty maps in Fig. 5.9. As expected, when the acceleration factor R was increased, there was a decrease in the uncertainty FOM in the measurable component while the uncertainty FOM in the null space component increased. Predictably, in general, an increase in the noise level σ for the same value of R also increased the uncertainty FOM in the measurable component of the alternate solutions. Furthermore, for the same acceleration factor R and noise level σ, both PULSE and DPS methods possessed a consistently higher uncertainty FOM in the measurable component as compared to PULSE++. This corroborates that the PULSE++ method significantly reduces the risk of data inconsistency for generating alternate solutions.

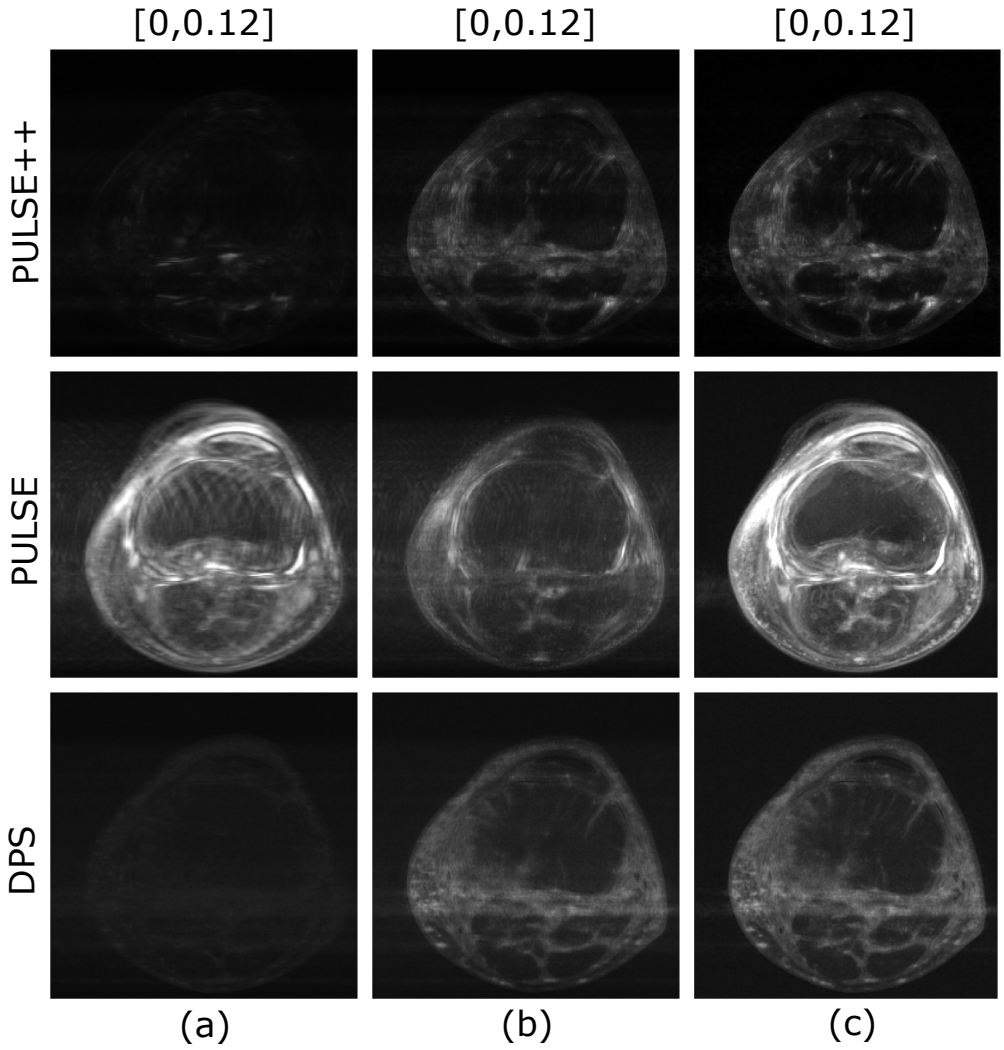


Figure 5.9: Uncertainty maps (a) $\hat{\mathbf{f}}^{UM}_{meas}$, (b) $\hat{\mathbf{f}}^{UM}_{null}$ and (c) $\hat{\mathbf{f}}^{UM}$ from the same k-space data produced by Knee 1 with $R = 3$ and $\sigma = 0.03$. The grayscale range of each type of uncertainty map is [0,0.12]. The PULSE++ method has significantly lower uncertainty in the measurable component compared to the PULSE method, indicating enhanced data consistency in alternate solutions produced by PULSE++.

5.7.2 Empirical Sampling from Limited-Angle CT Measurements

Visual Assessment

Samples of data-consistent alternate solutions obtained with the PULSE++ method using the same CT-StyleGAN model are shown in Fig. 5.10, corresponding to limited-angle projection data from Lung 1 ($I_0 = 10^5$, $I_0 = 10^3$) and Lung 2 ($I_0 = 10^5$). The alternate solutions in

Table 5.1: Summary of uncertainty FOMs of alternate solutions from the same k-space data for different values of R and σ

Object	Method	R	σ	$\|\hat{\mathbf{f}}^{UM}_{meas}\|_2^2$	$\|\hat{\mathbf{f}}^{UM}_{null}\|_2^2$	$\|\hat{\mathbf{f}}^{UM}\|_2^2$
Knee 1	PULSE++	3	0.03	2.31	11.89	14.20
	PULSE	3	0.03	103.42	41.75	145.17
	DPS	3	0.03	7.15	28.26	35.41
	PULSE++	3	0.05	3.56	11.53	15.09
	PULSE	3	0.05	86.35	41.62	127.98
	DPS	3	0.05	11.68	30.31	41.99
	PULSE++	4	0.03	1.41	19.60	21.01
	PULSE	4	0.03	84.63	74.09	158.72
	DPS	4	0.03	8.32	62.45	70.77
Knee 2	PULSE++	3	0.03	1.54	11.94	13.48
	PULSE	3	0.03	48.12	31.70	79.82
	DPS	3	0.03	6.90	18.94	25.84
	PULSE++	3	0.05	2.73	10.48	13.21
	PULSE	3	0.05	48.69	35.19	83.88
	DPS	3	0.05	10.17	20.92	31.09
	PULSE++	4	0.03	1.21	16.33	17.54
	PULSE	4	0.03	34.83	48.34	83.17
	DPS	4	0.03	5.33	36.82	42.15

each case displayed considerable variability in fine-scale structures. This illustrates the ability of the proposed PULSE++ method to produce diverse data-consistent solutions from the same measurement data for high-dimensional objects and containing Poisson noise, which is computationally infeasible with currently available posterior sampling methods [38].

Uncertainty Quantification

Uncertainty maps were computed for $T = 100$ alternate solutions corresponding to projection data from Lung 1 and Lung 2 with $I_0 = 10^5$ and $I_0 = 10^3$ obtained using PULSE++, along with their measurable and null space components in the domain of the fan-beam projector $\mathbf{H}$. The uncertainty maps for Lung 1 are shown in Fig. 5.11. The total uncertainty for the alternate solutions and their measurable and null space components for both Lung 1 and Lung 2 are shown in Table 5.2. Again, it was observed that the uncertainty in the alternate

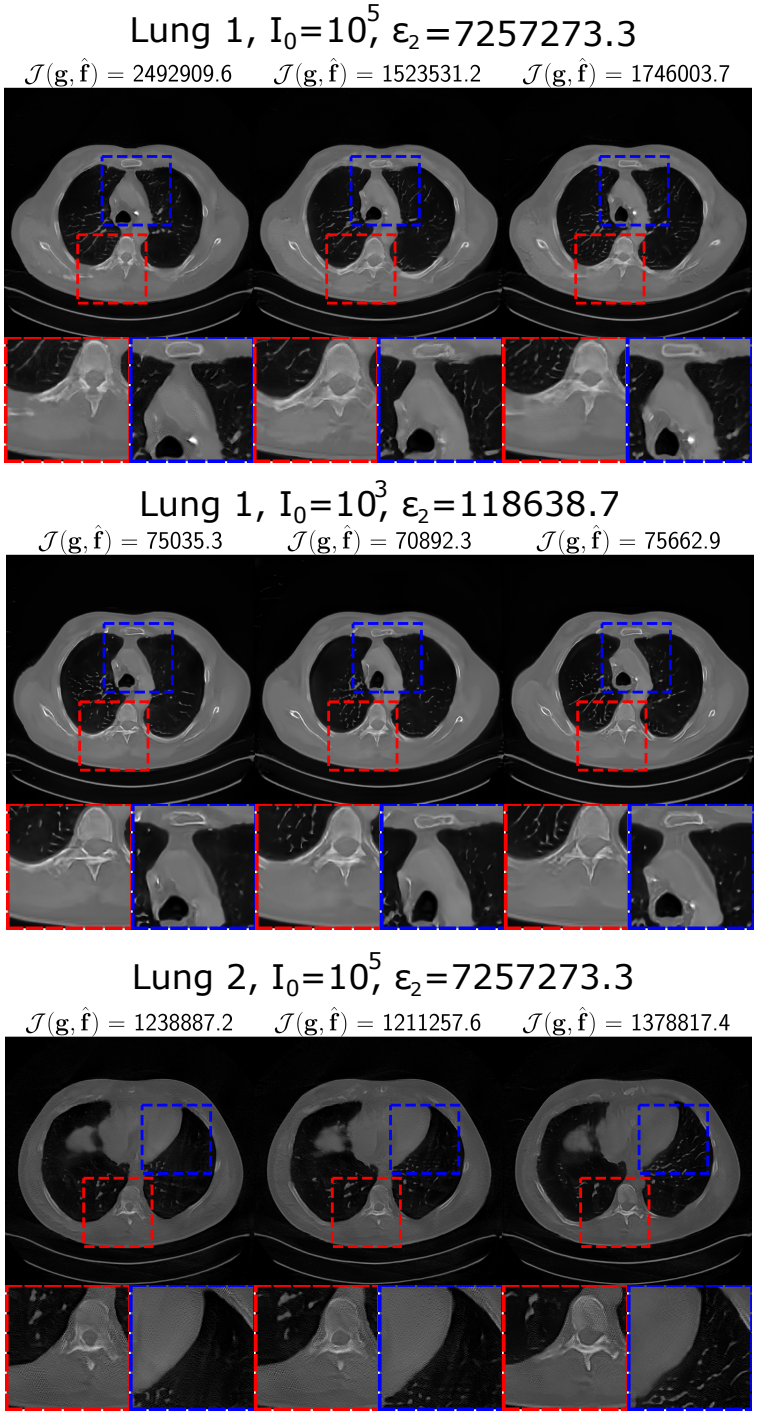


Figure 5.10: Alternate data-consistent solutions obtained using the PULSE++ method using the CT-StyleGAN model under different settings, for projection data from Lung 1 ($I_0 = 10^5$, $I_0 = 10^3$) and Lung 2 ($I_0 = 10^5$). Zoomed-in images from the alternate solutions demonstrate diversity in a number of fine-scale structures. The grayscale range of the alternate solutions is [0,1].

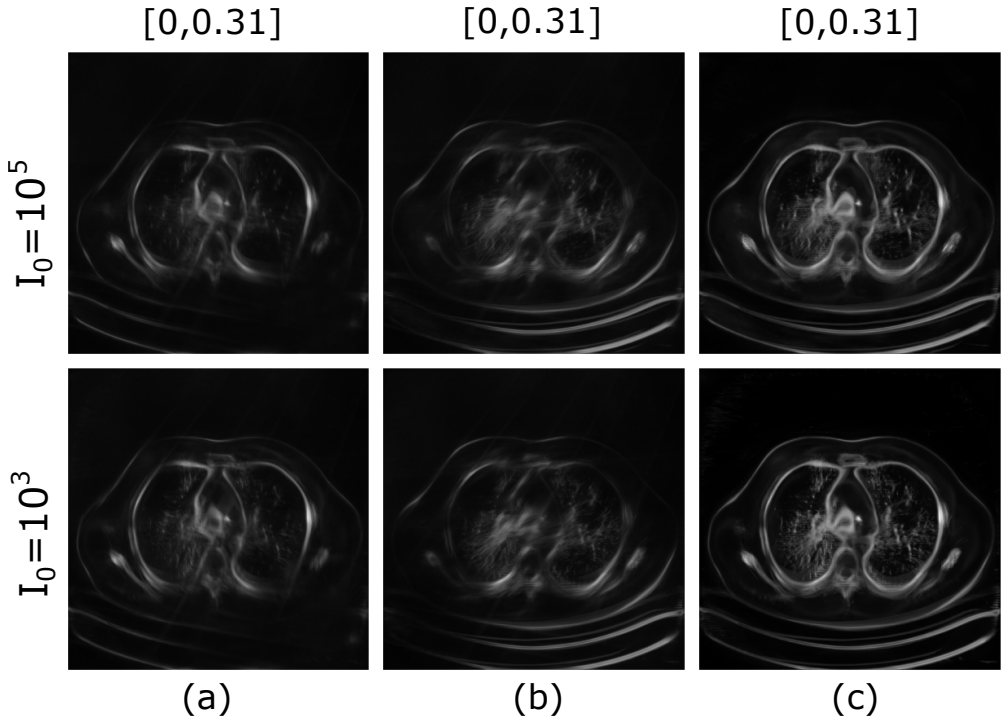


Figure 5.11: Uncertainty maps (a) $\hat{\mathbf{f}}^{UM}_{meas}$, (b) $\hat{\mathbf{f}}^{UM}_{null}$ and (c) $\hat{\mathbf{f}}^{UM}$ obtained with PULSE++ from CT measurements corresponding to Lung 1 for $I_0 = 10^5$ and $I_0 = 10^3$. The grayscale range of each type of uncertainty map is [0,0.31]. It is evident that in both cases, the uncertainty is primarily in the null space component. The variability in the measurable component is higher for $I_0 = 10^3$ due to a lower SNR in the projection data.

solutions was primarily due to variations in their null space component. Since the projection data for $I_0 = 10^5$ had higher SNR, the uncertainty FOM in the measurable component was expectedly lower as compared to that for $I_0 = 10^3$.

Table 5.2: Summary of uncertainty FOMs of alternate solutions obtained with PULSE++ from the same projection data using different values of I_0

Object	I_0	$\|\hat{\mathbf{f}}^{UM}_{meas}\|_2^2$	$\|\hat{\mathbf{f}}^{UM}_{null}\|_2^2$	$\|\hat{\mathbf{f}}^{UM}\|_2^2$
Lung 1	10^5	205.30	297.95	504.08
	10^3	239.74	281.09	521.51
Lung 2	10^5	124.00	404.64	530.15
	10^3	147.61	307.55	455.98

5.8 Summary

In this work, an empirical sampling method, called PULSE++, was proposed that employed generative model-constrained reconstruction with a StyleGAN to obtain multiple objects that are consistent with the same acquired tomographic measurement data. The proposed method represents an extension of the PULSE method that was originally developed for single-image super-resolution applications, but employs improved statistical assumptions regarding the StyleGAN latent space and methods to improve the stability of the optimization problem. It was demonstrated that the PULSE++ method was able to find data-consistent objects, whereas the PULSE method could not. Uncertainty maps were computed, and it was observed that the PULSE++ method consistently estimated lower uncertainty in the measurable space component compared to PULSE and a state-of-the-art diffusion model-based posterior sampling method (DPS), which suggests that PULSE++ is more suitable to perform reliable uncertainty quantification. Additionally, it was illustrated that the PULSE++ method is scalable and performs well with different measurement noise distributions.

The proposed PULSE++ method is general and, in principle, can be applied to any tomographic imaging system. While the studies in this paper concerned with two-dimensional imaging systems, the PULSE++ method can be extended to three-dimensional imaging systems by use of three-dimensional StyleGAN architectures [61]. Furthermore, the proposed framework may be readily adapted for use with future style-based deep generative models [78, 74].

The use of a StyleGAN in the PULSE++ method presents certain challenges. The StyleGAN must be sufficiently well-trained and should accurately represent the to-be-imaged object distribution. This can be challenging in diagnostic imaging applications, where the objects can contain varying pathologies that may not be fully represented in the StyleGAN training data [122, 41]. As such, the representation error of the StyleGAN should be acknowledged when employing PULSE++. However, because the PULSE++ method is intended to facilitate early-stage assessments of new imaging technologies, this representation error may be more tolerable than it would be if it was intended as an approximate Bayesian reconstruction method for clinical use.

There remain additional topics for future studies. In the presented studies, the true imaging operator was assumed to be known. When the goal is to facilitate the assessment of imaging

technologies in virtual imaging studies, this may not be a limitation. However, the impact of modeling errors on the performance of the PULSE++ method when applied to experimental measurements remains an important topic for investigation [113]. Additionally, it will be important to explore the application of the PULSE++ method for analyzing image reconstruction instabilities [49, 23] and enabling adaptive imaging procedures [18, 39].

Chapter 6

Conclusion

In this dissertation, we developed and investigated machine learning and deep learning methods for advancing tomographic image reconstruction, guided by modern generative models and fundamental principles of image science. We explored the symbiotic relationship between deep learning and image science which enables the development of more efficient neural network-based methods for computer vision tasks and new assessments of medical imaging systems.

In Chapter 3, we implemented generative model-constrained image reconstruction for ill-posed tomographic imaging problems by use of generative adversarial networks (GANs). We generalized the image-adaptive GAN-based reconstruction framework (IAGAN) for image restoration tasks in computer vision to an image reconstruction task involving tomographic imaging operators. The IAGAN method enabled imposition of stronger priors learned by state-of-the-art GANs such as ProGANs as compared to employing sparsity-promoting penalties only, while removing the possibility of plausible false structures that may be introduced due to representation error in GANs. Furthermore, we proposed a novel regularization technique in which a total variation (TV) penalty is added to the IAGAN objective function instead of performing early stopping to circumvent issues arising due to semi-convergence. The proposed method, termed as IAGAN-TV, improved upon the canonical IAGAN method with higher performance in terms of traditional image quality metrics such as root mean square error (RMSE) and structural similarity (SSIM), and mitigated the need for early stopping with improved convergence properties. Using numerical studies for accelerated MRI, the performance of the IAGAN-TV method was compared with a traditional sparsity-based image reconstruction method (PLS-TV) and a generative model-constrained reconstruction method based on invertible neural networks (INNs) which theoretically possess zero representation error but with lower image synthesis quality and scalability as compared with GANs. The IAGAN-TV method demonstrated its ability to preserve fine structures in the image that

are smoothed by PLS-TV, and maintained competitive performance in terms of traditional image-quality metrics as compared to the INN-based method. The generalization of the IAGAN-TV method was illustrated with experiments where a clear domain shift existed (adult brain vs pediatric brain), and the stability of the IAGAN-TV method was established using bias-variance analysis.

In Chapter 4, we presented a formal definition of hallucinations in ill-posed inverse problems using linear operator theory to quantify false structures that arise due to inaccurate priors. This definition was based on the mathematics of decomposition of objects into measurable and null components, which is a fundamental principle of image science. Using numerical studies with Fourier space measurements, it was demonstrated that traditional error maps may be insufficient to isolate false structures that originate only due to the imposed prior, while the proposed hallucination maps enabled identification and quantification of such false structures. Additionally, the concept of specific hallucination maps to identify task-specific hallucinations was demonstrated with a proof-of-concept study. Both data-driven and non-data-driven reconstruction methods were investigated to analyze hallucinations produced by them. It was observed that when the measurement data is out-of-distribution, end-to-end reconstruction networks such as the U-Net may be prone to producing false structures as quantified by hallucination maps. This illustrated how hallucination maps can be employed in virtual imaging trials for assessing the stability and generalization performance of deep learning-based reconstruction methods.

In Chapter 5, we proposed PULSE++, a method to produce multiple data-consistent solutions of ill-posed tomographic image reconstruction problems by mining the manifold of a StyleGAN generator. We extended the PULSE method (Photo Upsampling using Latent Space Exploration) introduced in single-image super-resolution (SISR) tasks by improving the accuracy of the imposed priors on the StyleGAN latent space and stabilizing the optimization method, which significantly improved data consistency of the alternate solutions. We illustrated how the PULSE++ method can produce multiple data-consistent solutions by performing numerical studies with different stylized imaging systems at scale (MRI and X-ray CT) and with different measurement noise distributions, such as Gaussian and Poisson distributions. The ability of PULSE++ to produce multiple data-consistent solutions enabled new assessments of imaging systems, such as uncertainty quantification and computing corresponding figures-of-merit (FOMs).

Many topics remain for future investigation. It is critical to evaluate GANs and modern deep generative models (DGMs) with respect to diagnostic tasks [83, 41] for objectively assessing the accuracy of the prior imposed in generative model-constrained reconstruction methods. Since many practical imaging systems are three-dimensional, the viability of generative model-constrained methods, such as by use of 3D-StyleGANs [61], needs to be studied. Moreover, it will be crucial to investigate generative model-constrained reconstruction using evolved DGMs, such as advanced StyleGANs [76, 77], vector-quantized variational autoencoders (VQ-VAEs) [114], denoising diffusion probabilistic models (DDPMs) [60, 38], among others. The computation of more informative FOMs from hallucination maps, such as probability of hallucination relevant to a particular task with a given imaging system and reconstruction method, remains an important challenge. Finally, it will be important to investigate the utility of multiple data-consistent solutions available using PULSE++ for analyzing image hallucinations and enabling adaptive imaging [18].

References

[1] Calculation of scale term #62, Github issues, 2018. https://github.com/openai/glow/issues/62.

[2] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. TensorFlow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.

[3] R. Abdal, Y. Qin, and P. Wonka. Image2Stylegan: How to embed images into the StyleGAN latent space? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4432–4441, 2019.

[4] R. Abdal, Y. Qin, and P. Wonka. Image2stylegan++: How to edit the embedded images? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8296–8305, 2020.

[5] S. Abu Hussein, T. Tirer, and R. Giryes. Image-Adaptive GAN based Reconstruction. *AAAI Conference on Artificial Intelligence*, 2020.

[6] B. M. Afkham, J. Chung, and M. Chung. Learning regularization parameters of inverse problems via deep neural networks. *Inverse Problems*, 37(10):105017, 2021.

[7] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2010.

[8] S. Aja-Fernández and G. Vegas-Sánchez-Ferrero. Statistical analysis of noise in MRI. *Switzerland: Springer International Publishing*, 2016.

[9] K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett, D. Bintley, et al. First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *The Astrophysical Journal Letters*, 875(1):L4, 2019.

[10] M. A. Anastasio and R. W. Schoonover. Basic principles of inverse problems for optical scientists. *digital Encyclopedia of Applied Physics*, pages 1–24, 2003.

[11] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences*, 2020.

[12] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv e-prints*, page arXiv:1701.07875, Jan 2017.

[13] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *International Conference on Machine Learning*, pages 399–409. PMLR, 2020.

[14] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002.

[15] G. Bal. Introduction to inverse problems. *Lecture Notes-Department of Applied Physics and Applied Mathematics, Columbia University, New York*, 2012.

[16] J. M. Bardsley, A. Solonen, H. Haario, and M. Laine. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1895–A1910, 2014.

[17] H. H. Barrett. Is there a role for image science in the brave new world of artificial intelligence? *Journal of Medical Imaging*, 7(1), 2020.

[18] H. H. Barrett, L. R. Furenlid, M. Freed, J. Y. Hesterman, M. A. Kupinski, E. Clarkson, and M. K. Whitaker. Adaptive spect. *IEEE transactions on medical imaging*, 27(6):775–788, 2008.

[19] H. H. Barrett and K. J. Myers. *Foundations of Image Science*. John Wiley & Sons, 2013.

[20] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[21] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[22] S. Bhadra, V. A. Kelkar, F. J. Brooks, and M. A. Anastasio. Assessing regularization in tomographic imaging via hallucinations in the null space. In *Medical Imaging 2021: Image Perception, Observer Performance, and Technology Assessment*, volume 11599, pages 132–138. SPIE, 2021.

[23] S. Bhadra, V. A. Kelkar, F. J. Brooks, and M. A. Anastasio. On hallucinations in tomographic image reconstruction. *IEEE transactions on medical imaging*, 40(11):3249–3260, 2021.

[24] S. Bhadra, U. Villa, and M. A. Anastasio. Mining the manifolds of deep generative models for multiple data-consistent solutions of ill-posed tomographic imaging problems. *arXiv preprint arXiv:2202.05311*, 2022.

[25] S. Bhadra, U. Villa, and M. A. Anastasio. Mining the manifolds of deep generative models for multiple data-consistent solutions of tomographic imaging problems - pyTorch implementation. 2022.

[26] S. Bhadra, W. Zhou, and M. A. Anastasio. Medical image reconstruction with image-adaptive priors learned by use of generative adversarial networks. In *Medical Imaging 2020: Physics of Medical Imaging*, volume 11312, page 113120V. International Society for Optics and Photonics, 2020.

[27] P. Bialecki and T. Viehmann. PyTorch implementation of the StyleGAN generator. 2019.

[28] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 537–546. JMLR. org, 2017.

[29] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 537–546. JMLR. org, 2017.

[30] A. Bora, E. Price, and A. G. Dimakis. AmbientGAN: Generative models from lossy measurements. In *International Conference on Learning Representations (ICLR)*, 2018.

[31] A. Borji. Pros and cons of GAN evaluation measures: New developments. *arXiv preprint arXiv:2103.09396*, 2021.

[32] A. Brock, J. Donahue, and K. Simonyan. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[33] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[34] G. Castellano, L. Bonilha, L. Li, and F. Cendes. Texture analysis of medical images. *Clinical radiology*, 59(12):1061–1069, 2004.

[35] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[36] K. Cheng, F. Calivá, R. Shah, M. Han, S. Majumdar, and V. Pedoia. Addressing the false negative problem of deep learning MRI reconstruction models by adversarial attacks and robust training. In *Medical Imaging with Deep Learning*, pages 121–135. PMLR, 2020.

[37] C. L. Chowdhary and D. Acharjya. Segmentation and feature extraction in medical imaging: a systematic review. *Procedia Computer Science*, 167:26–36, 2020.

[38] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

[39] E. Clarkson, M. A. Kupinski, H. H. Barrett, and L. Furenlid. A task-based approach to adaptive and multimodality imaging. *Proceedings of the IEEE*, 96(3):500–511, 2008.

[40] M. M. Deal and G. Nolet. Nullspace shuttles. *Geophysical Journal International*, 124(2):372–380, 1996.

[41] R. Deshpande, M. A. Anastasio, and F. J. Brooks. A method for evaluating the capacity of generative adversarial networks to reproduce high-order spatial context. *arXiv preprint arXiv:2111.12577*, 2021.

[42] Q. Ding, Y. Long, X. Zhang, and J. A. Fessler. Modeling mixed Poisson-Gaussian noise in statistical image reconstruction for X-ray CT. *Arbor*, 1001:48109, 2016.

[43] L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[44] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*, pages 179–187. Springer, 2016.

[45] A. Fawzi, H. Samulowitz, D. Turaga, and P. Frossard. Image inpainting through neural networks hallucinations. In *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, pages 1–5. Ieee, 2016.

[46] L. Fetty, M. Bylund, P. Kuess, G. Heilemann, T. Nyholm, D. Georg, and T. Löfstedt. Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Zeitschrift für Medizinische Physik*, 30(4):305–314, 2020.

[47] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[48] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[49] N. M. Gottschling, V. Antun, B. Adcock, and A. C. Hansen. The troublesome kernel: why deep learning for inverse problems is typically unstable. *arXiv preprint arXiv:2001.01258*, 2020.

[50] P. J. Green, K. Łatuszyński, M. Pereyra, and C. P. Robert. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25(4):835–862, 2015.

[51] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein GANs. *CoRR*, abs/1704.00028, 2017.

[52] N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[53] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic resonance in medicine*, 79(6):3055–3071, 2018.

[54] K. Hammernik and F. Knoll. Machine learning for image reconstruction. In *Handbook of Medical Image Computing and Computer Assisted Intervention*, pages 25–64. Elsevier, 2020.

[55] Y. Han and J. C. Ye. Framing U-Net via deep convolutional framelets: Application to sparse-view CT. *IEEE transactions on medical imaging*, 37(6):1418–1429, 2018.

[56] P. C. Hansen and J. S. Jørgensen. Air tools ii: algebraic iterative reconstruction methods, improved implementation. *Numerical Algorithms*, 79(1):107–137, 2018.

[57] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in neural information processing systems*, pages 6626–6637, 2017.

[58] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017.

[59] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Two time-scale update rule for training gans: Fréchet inception distance (fid). *Github*, 2017.

[60] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[61] S. Hong, R. Marinescu, A. V. Dalca, A. K. Bonkhoff, M. Bretzner, N. S. Rost, and P. Golland. 3d-StyleGAN: A style-based generative adversarial network for generative modeling of three-dimensional medical images. In *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections*, pages 24–34. Springer, 2021.

[62] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.

[63] Y. Huang, A. Preuhs, M. Manhart, G. Lauritsch, and A. Maier. Data consistent CT reconstruction from insufficient data with learned prior images. *arXiv preprint arXiv:2005.10034*, 2020.

[64] Y. Huang, T. Würfl, K. Breininger, L. Liu, G. Lauritsch, and A. Maier. Some investigations on robustness of deep learning in limited angle tomography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 145–153. Springer, 2018.

[65] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo. Deep learning for undersampled MRI reconstruction. *Physics in Medicine & Biology*, 63(13):135007, 2018.

[66] A. K. Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[67] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir. CSGM-MRI-Langevin. 2021.

[68] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir. Robust compressed sensing MRI with deep generative priors. *arXiv preprint arXiv:2108.01368*, 2021.

[69] 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*, 26(9):4509–4522, 2017.

[70] A. C. Kak, M. Slaney, and G. Wang. Principles of computerized tomographic imaging. *Medical Physics*, 29(1):107–107, 2002.

[71] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[72] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

[73] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation – Official TensorFlow implementation of the ICLR 2018 paper. In *Github*. 2018. https://github.com/tkarras/progressive_growing_of_gans.

[74] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila. Alias-free generative adversarial networks. In *Proc. NeurIPS*, 2021.

[75] T. Karras, S. Laine, and T. Aila. StyleGAN - official TensorFlow implementation. 2018.

[76] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

[77] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of StyleGAN. *arXiv preprint arXiv:1912.04958*, 2019.

[78] 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*, pages 8110–8119, 2020.

[79] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof. Diffusion models for medical image analysis: A comprehensive survey. *arXiv preprint arXiv:2211.07804*, 2022.

[80] V. A. Kelkar and M. A. Anastasio. Prior image-constrained reconstruction using style-based generative models. *arXiv preprint arXiv:2102.12525*, 2021.

[81] V. A. Kelkar, S. Bhadra, and M. A. Anastasio. Compressible latent-space invertible networks for generative model-constrained image reconstruction. *IEEE Transactions on Computational Imaging*, 2021.

[82] V. A. Kelkar, S. Bhadra, and M. A. Anastasio. Medical image reconstruction using compressible latent space invertible networks. In *Medical Imaging 2021: Physics of Medical Imaging*, volume 11595, pages 465–473. SPIE, 2021.

[83] V. A. Kelkar, D. S. Gotsis, F. J. Brooks, P. KC, K. J. Myers, R. Zeng, and M. A. Anastasio. Assessing the ability of generative adversarial networks to learn canonical medical image statistics. *arXiv preprint arXiv:2204.12007*, 2022.

[84] B. Kelly, T. P. Matthews, and M. A. Anastasio. Deep learning-guided image reconstruction from incomplete data. *arXiv preprint arXiv:1709.00584*, 2017.

[85] V. Kindratenko, D. Mu, Y. Zhan, J. Maloney, S. H. Hashemi, B. Rabe, K. Xu, R. Campbell, J. Peng, and W. Gropp. HAL: Computer System for Scalable Deep Learning. In *Practice and Experience in Advanced Research Computing*, pages 41–48. 2020.

[86] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[87] D. P. Kingma and P. Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018.

[88] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE signal processing magazine*, 34(4):43–59, 2017.

[89] J. Kuo, J. Granstedt, U. Villa, and M. A. Anastasio. Learning a projection operator onto the null space of a linear imaging operator. In *Medical Imaging 2021: Physics of Medical Imaging*, volume 11595, page 115953X. International Society for Optics and Photonics, 2021.

[90] M.-H. Laves, M. Tölle, and T. Ortmaier. Uncertainty estimation in medical image denoising with Bayesian deep image prior. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis*, pages 81–96. Springer, 2020.

[91] S. Leng, M. Bruesewitz, S. Tao, K. Rajendran, A. F. Halaweish, N. G. Campeau, J. G. Fletcher, and C. H. McCollough. Photon-counting detector CT: system design and clinical applications of an emerging technology. *Radiographics*, 39(3):729–743, 2019.

[92] J. Liu, Y. Sun, X. Xu, and U. S. Kamilov. Image restoration using total variation regularized deep image prior. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7715–7719. IEEE, 2019.

[93] W. Liu, D. Lin, and X. Tang. Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 478–484. IEEE, 2005.

[94] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet. Are gans created equal? a large-scale study. In *Advances in neural information processing systems*, pages 700–709, 2018.

[95] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE signal processing magazine*, 25(2):72–82, 2008.

[96] A. Maallo, T. Liu, E. Freud, C. Patterson, and M. Behrmann. Pediatric epilepsy resection MRI dataset, 2019. https://doi.org/10.1184/R1/9856205.

[97] A. M. S. Maallo, E. Freud, T. T. Liu, C. Patterson, and M. Behrmann. Effects of unilateral cortical resection of the visual cortex on bilateral human white matter. *NeuroImage*, 207:116345 1–12, 2020.

[98] A. M. S. Maallo, E. Freud, T. T. Liu, C. Patterson, and M. Behrmann. Effects of unilateral cortical resection of the visual cortex on bilateral human white matter. *NeuroImage*, 207:116345, 2020.

[99] A. L. Maas, A. Y. Hannun, A. Y. Ng, et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Citeseer, 2013.

[100] M. T. McCann, K. H. Jin, and M. Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017.

[101] M. T. McCann and M. Unser. Biomedical image reconstruction: From the foundations to deep neural networks. *arXiv preprint arXiv:1901.03565*, 2019.

[102] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 2437–2445, 2020.

[103] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. 2020.

[104] A. Mohebi and P. Fieguth. Posterior sampling of scientific images. In *International Conference Image Analysis and Recognition*, pages 339–350. Springer, 2006.

[105] V. A. Morozov. On the solution of functional equations by the method of regularization. In *Doklady Akademii Nauk*, volume 167, pages 510–512. Russian Academy of Sciences, 1966.

[106] L. Mosser, O. Dubrule, and M. J. Blunt. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *arXiv preprint arXiv:1806.03720*, 2018.

[107] M. J. Muckley, B. Riemenschneider, A. Radmanesh, S. Kim, G. Jeong, J. Ko, Y. Jun, H. Shin, D. Hwang, M. Mostapha, et al. Results of the 2020 fastmri challenge for machine learning mr image reconstruction. *IEEE transactions on medical imaging*, 40(9):2306–2317, 2021.

[108] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.

[109] S. Oymak and M. Soltanolkotabi. Overparameterized nonlinear learning: Gradient descent takes the shortest path? In *International Conference on Machine Learning*, pages 4951–4960. PMLR, 2019.

[110] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019.

[111] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE transactions on medical imaging*, 30(5):1028–1041, 2010.

[112] S. Ravishankar and Y. Bresler. Learning sparsifying transforms. *IEEE Transactions on Signal Processing*, 61(5):1072–1086, 2012.

[113] S. Ravishankar, J. C. Ye, and J. A. Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1):86–109, 2019.

[114] A. Razavi, A. van den Oord, and O. Vinyals. Generating diverse high-fidelity images with vq-vae-2. In *Advances in neural information processing systems*, pages 14866–14876, 2019.

[115] S. J. Reddi, S. Kale, and S. Kumar. On the convergence of Adam and beyond. *arXiv preprint arXiv:1904.09237*, 2019.

[116] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[117] S. Ravishankar and Y. Bresler. Data-driven learning of a union of sparsifying transforms model for blind compressed sensing. *IEEE Transactions on Computational Imaging*, 2(3):294–309, 2016.

[118] K. Schutte, O. Moindrot, P. Hérent, J.-B. Schiratti, and S. Jégou. Using StyleGAN for visual interpretability of deep learning models on medical images. *arXiv preprint arXiv:2101.07563*, 2021.

[119] R. Shaw, C. H. Sudre, S. Ourselin, and M. J. Cardoso. Estimating MRI image quality via image reconstruction uncertainty. *arXiv preprint arXiv:2106.10992*, 2021.

[120] A. Siahkoohi, G. Rizzuti, M. Louboutin, P. A. Witte, and F. J. Herrmann. Deep Bayesian inference for task-based seismic imaging. In *KAUST*, 03 2021. Talk at KAUST.

[121] E. Y. Sidky and X. Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777, 2008.

[122] Y. Skandarani, P.-M. Jodoin, and A. Lalande. GANs for medical image synthesis: An empirical study. *arXiv preprint arXiv:2105.05318*, 2021.

[123] Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

[124] Y. Song and S. Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020.

[125] J. C. Spall. Stochastic optimization. In *Handbook of computational statistics*, pages 173–201. Springer, 2012.

[126] H. Sun and K. L. Bouman. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. *arXiv preprint arXiv:2010.14462*, 9, 2020.

[127] M. U. *et al.* Bart toolbox for computational magnetic resonance imaging, doi: 10.5281/zenodo.592960. 2015.

[128] J. Tick, A. Pulkkinen, and T. Tarvainen. Image reconstruction with uncertainty quantification in photoacoustic tomography. *The Journal of the Acoustical Society of America*, 139(4):1951–1961, 2016.

[129] T. Tieleman and G. Hinton. Lecture 6.5—RMSProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.

[130] S. Tiwari, K. Kaur, and K. Arya. A study on dictionary learning based image reconstruction techniques for big medical data. In *Handbook of Multimedia Information Security: Techniques and Applications*, pages 377–393. Springer, 2019.

[131] M. Tygert and J. Zbontar. Simulating single-coil mri from the responses of multiple coils. *arXiv preprint arXiv:1811.08026*, 2018.

[132] M. Uecker, F. Ong, J. I. Tamir, D. Bahri, P. Virtue, J. Y. Cheng, T. Zhang, and M. Lustig. Berkeley advanced reconstruction toolbox. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 23, 2015.

[133] T. J. Ulrych, M. D. Sacchi, and A. Woodbury. A bayes tour of inversion: A tutorial. *Geophysics*, 66(1):55–69, 2001.

[134] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[135] D. Van Veen, A. Jalal, M. Soltanolkotabi, E. Price, S. Vishwanath, and A. G. Dimakis. Compressed sensing with deep image prior and learned regularization. *arXiv preprint arXiv:1806.06438*, 2018.

[136] R. Vershynin. Random vectors in high dimensions. *Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press*, 3:38–69, 2018.

[137] G. Wang. A perspective on deep imaging. *IEEE access*, 4:8914–8924, 2016.

[138] K. Wang and M. A. Anastasio. Photoacoustic and thermoacoustic tomography: image formation principles. In *Handbook of Mathematical Methods in Imaging*. 2015.

[139] N. Wang, D. Tao, X. Gao, X. Li, and J. Li. A comprehensive survey to face hallucination. *International journal of computer vision*, 106(1):9–30, 2014.

[140] 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*, 13(4):600–612, 2004.

[141] D. W. Wilson and H. H. Barrett. Decomposition of images and objects into measurement and null components. *Optics Express*, 2(6):254–260, 1998.

[142] Z. Wu, D. Lischinski, and E. Shechtman. Stylespace analysis: Disentangled controls for stylegan image generation, 2020. *URL: https://arxiv. org/abs/2011.12799. doi*, 10, 2020.

[143] J. Wulff and A. Torralba. Improving inversion and generation diversity in StyleGAN using a Gaussianized latent space. *arXiv preprint arXiv:2009.06529*, 2020.

[144] F. Xiaoyu, L. Qiusheng, and S. Baoshun. Compressed sensing MRI with phase noise disturbance based on adaptive tight frame and total variation. *IEEE Access*, 5:19311–19321, 2017.

[145] K. Yan, X. Wang, L. Lu, and R. M. Summers. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of Medical Imaging*, 5(3):036501, 2018.

[146] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, et al. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018.

[147] J. e. a. Zbontar. fastMRI. 2018.

[148] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016.

[149] W. Zhou, S. Bhadra, F. J. Brooks, J. L. Granstedt, H. Li, and M. A. Anastasio. Advancing the AmbientGAN for learning stochastic object models. In *Medical Imaging 2021: Image Perception, Observer Performance, and Technology Assessment*, volume 11599, page 115990A. International Society for Optics and Photonics, 2021.

[150] W. Zhou, S. Bhadra, F. J. Brooks, H. Li, and M. A. Anastasio. Learning stochastic object models from medical imaging measurements by use of advanced AmbientGANs. *arXiv preprint arXiv:2106.14324*, 2021.

[151] P. Zhu, R. Abdal, Y. Qin, J. Femiani, and P. Wonka. Improved styleGAN embedding: Where are the good latents? *arXiv preprint arXiv:2012.09036*, 2020.

Appendix A

Examples of measurement space hallucination maps

The measurement space hallucination map is denoted as

$$\hat{\mathbf{f}}_{meas}^{HM} \equiv \hat{\mathbf{f}}_{meas} - \hat{\mathbf{f}}_{tp}, \tag{A.1}$$

which describes the consistency between the measurement component of the reconstructed image $\hat{\mathbf{f}}_{meas}$ with respect to the truncated pseudoinverse solution $\hat{\mathbf{f}}_{tp}$ that can be stably obtained from the measurement data $\mathbf{g}$. Furthermore, unlike the null space hallucination map, the computation of $\hat{\mathbf{f}}_{meas}^{HM}$ does not require the knowledge of the true object $\mathbf{f}$. On the other hand, the error map between $\hat{\mathbf{f}}_{meas}$ and $\mathbf{f}_{meas}$ is a similar but different error quantity that lies in the measurement space $\mathcal{N}_P^{\perp}(\mathbf{H})$ and requires the knowledge of $\mathbf{f}$:

$$\hat{\mathbf{f}}_{meas}^{EM} \equiv \hat{\mathbf{f}}_{meas} - \mathbf{f}_{meas}. \tag{A.2}$$

In some cases, $\hat{\mathbf{f}}_{meas}^{HM}$ and $\hat{\mathbf{f}}_{meas}^{EM}$ may not convey the same information due to the differences that can exist between $\mathbf{f}_{meas}$ and $\hat{\mathbf{f}}_{tp}$. These differences arise when there is significant measurement noise in the imaging system or due to modeling error in $\mathbf{H}$, or both. With the true imaging operator denoted as $\tilde{\mathbf{H}}$ and the assumed imaging operator as $\mathbf{H}$, Eq. (2.3) can be re-written as

$$\mathbf{g} = \tilde{\mathbf{H}}\mathbf{f} + \mathbf{n}. \tag{A.3}$$

Accordingly, $\hat{\mathbf{f}}_{tp}$ can be expressed as

$$\hat{\mathbf{f}}_{tp} = \mathbf{H}_P^{+}\mathbf{g} \approx \mathbf{H}_P^{+}(\tilde{\mathbf{H}}\mathbf{f} + \mathbf{n}) = \mathbf{H}_P^{+}\tilde{\mathbf{H}}\mathbf{f} + \mathbf{H}_P^{+}\mathbf{n}. \tag{A.4}$$

On the other hand, $\mathbf{f}_{meas}$ is represented as

$$\mathbf{f}_{meas} \equiv \mathbf{H}_P^{+}\mathbf{H}\mathbf{f}. \tag{A.5}$$

It can be observed from Eq. (A.4) and Eq. (A.5) that, when either or both of the quantities $||\mathbf{n}||_2^2$ and $||\mathbf{H} - \tilde{\mathbf{H}}||_2^2$ is non-trivial, $||\hat{\mathbf{f}}_{tp} - \mathbf{f}_{meas}||_2^2$ is likely to be significant. In such cases, $\hat{\mathbf{f}}_{meas}^{HM}$ and $\hat{\mathbf{f}}_{meas}^{EM}$ may represent different information.

Figure S.A.1 shows examples of measurement space hallucination maps for images reconstructed by the U-Net method corresponding to an in-distribution (IND) object and an out-of-distribution (OOD) object. It can be observed that the measurement component error map has appreciable differences compared to the measurement space hallucination map. These differences can be attributed to the presence of non-trivial measurement noise as well as additional phase noise disturbance during simulation of the k-space data resulting in model error, since the phase noise is unknown and assumed to be zero during reconstruction. The measurement space hallucination maps as shown here may provide an insight into how well a given reconstruction method maintains data consistency and be compared with other reconstruction methods.

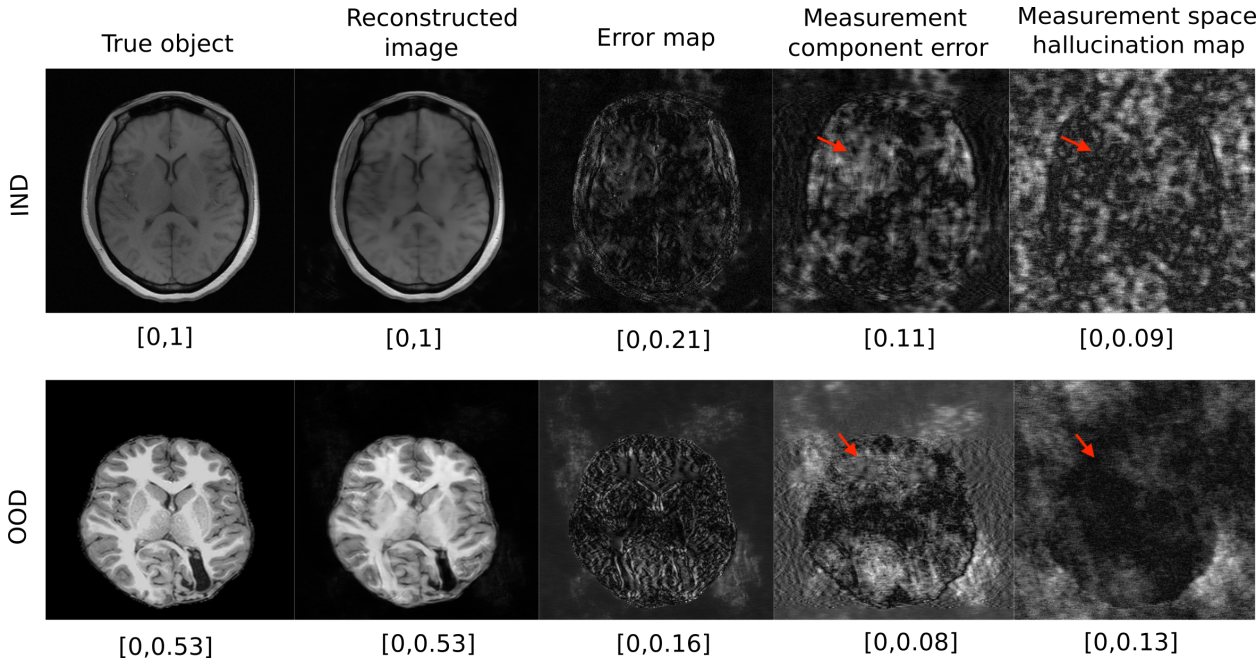


Figure A.1: Measurement space hallucination maps for reconstructed images using the U-Net method corresponding to an IND (above) and an OOD (below) object. Note that the measurement component error map and the measurement space hallucination map are appreciably different. The red arrows point towards a region in each type of object where such differences can be clearly seen.

Appendix B

Invertible transformation between style latent spaces in StyleGAN

The vector $\mathbf{w} \in \mathbb{R}^k$ in the latent space $\mathcal{W}$ is produced by passing the initial latent vector $\mathbf{z} \in \mathbb{R}^k$ through the fully-connected mapping network G_m [76]. The final layer of the network G_m consists of the point-wise non-linear LeakyReLU activation function [99] with parameter 0.2, i.e. $\text{LeakyReLU}_{0.2}(\cdot)$. The inverse of the function $\text{LeakyReLU}_{0.2}(\cdot)$ is the function $\text{LeakyReLU}_{5.0}(\cdot)$. Consequently, the output $\mathbf{w}' \in \mathbb{R}^k$ of the mapping network G_m that just precedes the final non-linear activation layer $\text{LeakyReLU}_{0.2}(\cdot)$, can be obtained as $\mathbf{w}' \equiv \text{LeakyReLU}_{5.0}(\mathbf{w})$. The mean vector $\boldsymbol{\mu}_{\mathbf{w}'}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}'}$ were estimated by computing the sample mean and sample covariance respectively from 10^7 random samples of $\mathbf{w}'$. These samples were generated by randomly sampling from $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_k)$, and then passing them through the mapping network G_m followed by inverting the final LeakyReLU activation function. Subsequently, a whitening transformation is applied on the vector $\mathbf{w}'$ to produce the resultant vector $\mathbf{v} \in \mathbb{R}^k$ in latent space $\mathcal{V}$ as

$$\mathbf{v} \equiv {\boldsymbol{\Sigma}_{\mathbf{w}'}}^{-\frac{1}{2}}(\mathbf{w}' - \boldsymbol{\mu}_{\mathbf{w}'}). \tag{B.1}$$

Thus, the invertible transform function $\mathcal{T} : \mathbb{R}^k \to \mathbb{R}^k$ from the latent space $\mathcal{W}$ to the latent space $\mathcal{V}$ is expressed as

$$\begin{aligned} \mathbf{v} &= {\boldsymbol{\Sigma}_{\mathbf{w}'}}^{-\frac{1}{2}}(\text{LeakyReLU}_{5.0}(\mathbf{w}) - \boldsymbol{\mu}_{\mathbf{w}'}) \\ &= \mathcal{T}(\mathbf{w}). \end{aligned} \tag{B.2}$$