Solving Poisson Inverse Problems in Phase Retrieval and Single Photon Emission Computerized Tomography

by

Zongyu Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan
2024

Doctoral Committee:

Professor Jeffrey A. Fessler, Chair
Professor Yuni K. Dewaraja
Professor Zhong He
Assistant Professor Qing Qu

Zongyu Li

zonyul@umich.edu

ORCID iD: 0000-0003-1813-1722

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to all those who have supported me throughout this journey of completing my PhD thesis. Their unwavering support, encouragement, guidance, and valuable insights have played a significant role in the successful completion of this work.

Firstly, I extend my deepest thanks to my advisors Dr. Jeff Fessler and Dr. Yuni Dewaraja for their exceptional mentorship and invaluable expertise. Their continuous support and constructive criticism pushed me beyond my limits and helped shape the direction of this research. Their insightful comments, suggestions, and critiques significantly contributed to enhancing the quality of my research.

Furthermore, I would like to acknowledge University of Michigan for providing necessary resources including access to literature databases as well as laboratory facilities that were essential for conducting experiments outlined in this thesis.

In addition, I sincerely appreciate the cooperation received from fellow researchers from University of Michigan and other universities. The intellectual exchanges we shared during seminars or informal discussions greatly enriched both my knowledge base and research outcomes.

Finally, heartfelt thanks go out to my parents and friends harbored immense faith in me during challenging times - your emotional support kept me motivated when doubts loomed over progress.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ACRONYMS

ADMM alternating direction method of multipliers

AP analytical projector

AWFS accelerated Wirtinger flow with score-based image prior

BM3D block-matching and 3D filtering

CCD charge-coupled device

CDI coherent diffractive imaging

CG conjugate gradient

CNN convolutional neural network

CRLB Cramér-Rao lower bound

CT computerized tomography

DDPM denoising diffusion probabilistic model

DFT discrete Fourier transform

DL deep learning

DNCNN Denoising convolutional neural network

DOLPH diffusion models for phase retrieval

DPM dose planning method

DRVH dose-rate volume histogram

DVK dose voxel kernel

EM expectation-maximization

ETM empirical transmission matrix

FFT fast (discrete) Fourier transform

FISTA fast iterative shrinkage-thresholding algorithm

FOV field of view

FWHM full width at half maximum

GAN generative adversarial network

GMM gaussian mixture models

GS Gerchberg Saxton

IID independent and identically distributed

LBFGS low-memory Broyden–Fletcher–Goldfarb–Shanno

MAE mean activity error

MAP maximum a posteriori

MBIR model-based image reconstruction

MC Monte Carlo

MLE maximum likelihood estimate

MLEM maximum likelihood expectation maximization

MM majorize minimize

MRI magnetic resonance imaging

NERF neural radiance field

NLM non-local means

NRMSE normalized root mean square error

ODWT orthogonal discrete wavelet transform

OSEM ordered-subset expectation maximization

PET positron emission tomography

PG poisson-gaussian

PGM proximal gradient method

PNP plug-and-play

POGM proximal optimized gradient method

PR phase retrieval

PSD positive semi-definite

PSF point spread function

PSNR peak signal-to-noise ratio

PV partial volume

RED regularization by denoising

RED-SD regularization by denoising with steepest descent

RELU Rectified Linear Unit

SAM segment anything model

SDP semi-definite programming

SFBP SPECT forward-backward projector

SOTA state-of-the-art

SPECT single photon emission computerized tomography

SSIM structural similarity index measure

STD standard deviation

TEW triple energy windows

TOF time-of-flight

TV total variation

UCRLB uniform Cramér-Rao lower bound

VAE variational autoencoder

VOI volumes of interest

VP virtual patient

WF Wirtinger flow

WFS Wirtinger flow with score-based image prior

XCAT extended cardiac-torso

^{68}Ga Gallium-68

^{90}Y Yttrium-90

^{177}Lu Lutetium-177

LIST OF SYMBOLS

$\boldsymbol{x}$ a signal or image

$\hat{\boldsymbol{x}}$ an image estimate

$\boldsymbol{y}$ measurements

$\boldsymbol{A}$ the measurement/system/forward model

$\|\cdot\|_p$ ℓ_p-norm

$\nabla\cdot$ the gradient operator

$\nabla^2\cdot$ the Hessian operator

$\mathsf{R}(\cdot)$ a regularization function

$\boldsymbol{B}'$ Hermitian transpose of $\boldsymbol{B}$

$\boldsymbol{B}^T$ transpose of $\boldsymbol{B}$

$\mathsf{diag}\{\cdot\}$ a diagonal matrix

ABSTRACT

We live in a world where many objects cannot be imaged directly and hence rely on reconstruction algorithms to solve the corresponding inverse imaging problems. However, lots of information is contaminated or even lost when samples are collected by imaging devices, so that the resulting inverse problem is ill-posed and challenging to solve. As the recorded photon arrivals by the sensor are often assumed to follow Poisson distributions, algorithms for solving Poisson inverse problems are crucial. This thesis tackles two applications where Poisson inverse problems arise: phase retrieval and single photon emission computerized tomography (SPECT).

For phase retrieval, we propose novel optimization algorithms working in low-count regimes, including a novel majorize-minimize (MM) algorithm, a modified Wirtinger flow algorithm using the observed Fisher information for step size and a generative image prior based on score matching. Our proposed algorithms lead to faster convergence rate and improved reconstruction quality evaluated both qualitatively and quantitatively.

For SPECT imaging, we focus on deep learning (DL) solutions including: 1) We propose end-to-end training of unrolled iterative convolutional neural network (CNN) using our memory efficient Julia toolbox for SPECT image reconstruction. 2) We propose a DL algorithm for joint dosimetry estimation and image deblurring for estimating patient's absorbed dose-rate distribution in radionuclide therapy. 3) We propose unsupervised coordinate-based learning for predicting missing SPECT projection views.

CHAPTER 1

Introduction

Imaging has a rich historical background and plays a vital role in various fields, including everyday applications for consumers, medical diagnostics [172], and astronomical research [154]. Continuous effort from researchers and engineers have been contributed to improving imaging techniques and obtain higher-quality images. One area of focus in the broad imaging field is *computational imaging*, which involves situations where direct camera imaging of objects (such as nano-structures or body organs) is not feasible. Instead, measurements are collected using imaging devices and the object needs to be reconstructed from these measurements. This process poses an *inverse problem* in imaging that can be challenging due to resource limitations of the systems involved.

One of challenges in *inverse problem* solving is the measurements are susceptible to noise corruption. This is because, the image sensors quantify the incident scene irradiance by tallying the discrete photon count within a specific period. Digital sensors use the photoelectric effect to convert photons into electrons, while film-based sensors rely on photosensitive chemical reactions. In both scenarios, due to individual photon arrivals being independent and random, there exists inherent uncertainty known as *Poisson process* which is signal-dependent and intrinsic to the underlying signal itself [85]. The signal-independent zero-mean additive Gaussian model is commonly used as an approximation for image noise. However, this model may not accurately represent the noise characteristics of imaging systems, particularly in low photon count scenarios. In such cases, the Poisson noise model, which takes into account photon noise statistics, may be a more suitable choice in such imaging applications.

Perhaps there are many applications that fall under the category of *inverse problems* considering Poisson noise statistics. This thesis primarily focuses on *phase retrieval* and single photon emission computerized tomography (SPECT). *Phase retrieval* is an essential problem in many optical imaging applications, as many imaging devices (*e.g.*, CCD cameras) can record only the magnitude (or square of magnitude) of signals and the phase information is lost. For example, in coherent diffractive imaging (CDI), a coherent beam source illuminates a sample of interest and a reference. When the beam hits the sample, it generates secondary electromagnetic waves that propagate until they reach a detector. By measuring the photon flux, the detector can capture and record a diffraction pattern. This pattern is roughly proportional to the square of Fourier transform magnitude of electric field associated with the illuminated objects [10, 11]. Recovering the structure of the sample from its diffraction pattern is a nonlinear *Poisson inverse problem* known as *phase retrieval*.

SPECT is a widely used nuclear imaging technique that provides visualizations of functional processes within the human body [153]. It plays a crucial role in assessing bodily functions, diagnosing diseases, and guiding treatment decisions. The SPECT procedure involves injecting a radioactive tracer into the bloodstream, which then gets taken up by specific tissues. Radioactive tracers consist of carrier molecules that are tightly bonded to radioactive atoms. The selection of these carrier molecules depends on the specific purpose of the scan, with some tracers utilizing molecules that interact with particular proteins or sugars in the body and even incorporating the patient's own cells. The tracer emits gamma rays that are detected by a gamma camera and converted into electrical signals. The gamma camera system is mounted on a rotating gantry that allows the detectors to move in an approximately circular motion around the patient lying on an examination bed. Some newer systems [1] are arranged in a ring geometry surrounding the patient. Thus the SPECT imaging system captures a 3-dimensional images depicting the distribution of the radioactive tracer throughout various bodily tissues and organs.

However, as mentioned earlier, obtaining high-quality measurements in imaging acquisition can be challenging, making it difficult to solve the corresponding *inverse problem*. For example, in SPECT, the measurement quality is known to be limited by the *scattering events* (*e.g.*, Compton scatter and coherent scatter), resolution of the collimator as well as patient motions during acquisition. All of these systems are further subject to (Poisson) noise and other nonidealities. These physical effects limit the quality of the measured samples and are almost unavoidable at a hardware level. Fortunately, there is much space to improve the image quality other than hardware level. *Computational imaging* techniques allow us to optimize an *inverse problem* by mathematically modeling the imaging system. Advances in computational imaging algorithms can potentially overcome challenges in the physical domain and enhance reconstruction quality. Therefore, using *computational imaging* algorithms to solve such *Poisson inverse problems* as discussed earlier is the central goal of this thesis.

[1] https://www.gehealthcare.com/products/molecular-imaging/starguide

A common way in *computational imaging* to solve *Poisson inverse problems* is to pose the reconstruction as an optimization problem consisting of a *data-fidelity* term associated with a Poisson likelihood, encouraging our reconstruction to be consistent with the measured samples, and a *regularization* term, enforcing our measurement-independent expectations or assumptions about the class of images we are reconstructing. Developing an effective regularization method is challenging, as it must be both flexible enough to represent all possible true images, yet be discriminating enough to reject noise and artifacts. Regularization reflects our prior beliefs or assumptions about the true images, for example, that an image having repeated textures can be represented with reduced dimensionality. As a result, a good regularizer can be designed to encourage lower-dimensional structure. As machine learning becomes popular in many fields, many image regularization methods are data-driven. These regularizers may not be explicit and are instead learned from a dataset of various images and are expected to be generalizable to unseen images. Such regularizers can range from simple, linear models (*e.g.*, total variation (TV)) to complex and nonlinear methods such as *deep learning* (DL) algorithms based on neural networks.

The emergence of large-scale data analysis and significant advancements in computing capabilities have resulted in *machine learning* methodologies being effective for solving *Poisson inverse problems*. Particularly, DL has emerged as a widely used approach that has proven successful across various fields. Unlike traditional machine learning approaches that rely heavily on feature engineering and manual extraction of relevant information from input data [187], DL methods are able to automatically learn intricate patterns and representations directly from raw data through a process called model training. With its remarkable capacity due to its millions or even trillions of learnable parameters, DL methods have led to significant progress in solving *Poisson inverse problems* in computational imaging, however, they also leave more open research questions and directions for future works.

1.1 Contributions

The contributions of works in this thesis can be summarized as follows.

1. Phase Retrieval.
    - We develop algorithms to address the challenges of low-count regimes in the Poisson phase retrieval problem. Our proposed approach involves a modified Wirtinger flow (WF) algorithm that uses a step size determined by the observed

Fisher information, as well as a novel curvature for majorize-minimize algorithms that incorporates a quadratic majorizer. Through simulated experimentation with different system matrices, we demonstrate the effectiveness and convergence properties of our methods. The simulation results reveal that our approaches not only enable successful recovery in extremely low-count scenarios but also outperform previous methods in terms of speed of convergence. This work is based on published papers [138, 139].

- In addition, we investigate situations where the measurements are influenced by a mixture of Poisson and Gaussian noise. We introduce a novel method called "AWFS" that employs accelerated WF with a score function as a generative prior. We provide theoretical analysis to demonstrate the convergence guarantee of the proposed algorithm at critical points. Results from simulations illustrate that our approach improves reconstruction quality in terms of both visual perception and numerical assessment. This work is based on [89, 137].

2. SPECT Imaging.

    - We develop an efficient Julia toolbox for modeling SPECT forward-backward projectors[2]. This toolbox uses multi-threading and in-place operations to enable parallel computing and reduce memory allocations. As a result, our proposed SPECT projector allows for efficient backpropagation during the training of deep learning regularized iterative algorithms in an end-to-end manner. This approach has shown potential for producing higher quality reconstructions compared to methods without end-to-end training. This work is based on published paper [134].
    - We propose a deep neural network (DblurDoseNet) for joint dosimetry estimation and image deblurring after SPECT reconstruction that produces accurate dose-rate distribution estimates as well as compensating for SPECT resolution effects. Evaluations both on phantoms and patients demonstrate that the proposed DblurDoseNet can outperform the current gold standard, *i.e.*, Monte Carlo (MC) based methods, and is also fast enough for real-time clinical use in radionuclide therapy dosimetry for treatment planning [136, 132].
    - We propose a neural network with unsupervised coordinate-based learning to predict missing SPECT projections before reconstruction. Our method aims to

[2] https://github.com/JuliaImageRecon/SPECTrecon.jl

decrease the acquisition time for SPECT by only obtaining a subset (*e.g.*, one fourth) of all projections. Our unsupervised approach achieves improved quality in image reconstruction when compared to linear interpolation methods used for the prediction of absent projection views. This work is based on published abstract [140].

1.2 Outline

Chapter 2 provides the necessary background on the mathematical frameworks we use for modelling Poisson inverse problems. It introduces the details of phase retrieval, SPECT image reconstruction and dosimetry. Chapter 3 introduced our proposed algorithms for the phase retrieval problem. Chapter 4 focuses on DL solutions in SPECT imaging, from acqusition, reconstruction, to post processing. Chapter 5 discusses current challenges faced and explores potential avenues for future research. One of the major challenges is addressing the limited amount of training data for large 3D images, prompting the exploration of techniques like transfer learning and active learning. Additionally, we delve into topics such as generative AI, multi-modality imaging, and optimization approaches from a computational perspective. Chapter 6 concludes this thesis.

The appendix provides more in depth algorithmic detail than is provided in the main chapters as well as other supplementary materials. In particular, Appendix A derives an improved curvature of the quadratic majorizer in the MM algorithm for Poisson phase retrieval in Chapter 3. Appendix B derives and analyzes the UCRLB for the WF algorithm in Chapter 3. Appendix C provided detailed analysis on the critical-point convergence guarantee of the "AWFS" algorithm in Chapter 3.

CHAPTER 2

Background

This chapter first introduces inverse problems mathematically and presents a generic framework for addressing such problems. We then present the background for phase retrieval and quantitative SPECT imaging in sufficient detail as two areas where inverse problems arise.

2.1 Inverse Problems

When indirectly collecting samples of some signals (or images) of interest, what we measure is often a function of that signal, be it blur, missing data, a projection, or something more complex. For the linear case, we can describe how the discrete measurements $\boldsymbol{y}$ relate to the underlying (possibly continuous) true signal $\boldsymbol{x}_{\text{true}}$, with the following equation:

$$\boldsymbol{y} \sim \text{Poisson}(\boldsymbol{A}\boldsymbol{x}_{\text{true}} + \bar{\boldsymbol{r}}). \tag{2.1}$$

Here, $\boldsymbol{A}$ represents a linear measurement model and $\bar{\boldsymbol{r}}$ denotes the mean background events such as scatters. Here we assume the noise vector elements have independent Poisson distributions, which is a realistic model for many applications. Estimating $\boldsymbol{x}_{\text{true}}$ from $\boldsymbol{y}$ and $\boldsymbol{A}$ is known as an *inverse problem*, and it is challenging because one usually does not have sufficient information to exactly recover $\boldsymbol{x}_{\text{true}}$.

To estimate $\boldsymbol{x}_{\text{true}}$ from $\boldsymbol{y}$ following (2.1), classically, filtering-based methods were developed, such as filtered back projection for CT [229] or Wiener filtering [35] for denoising problems. These algorithms are fast and efficient, but they do not model the physics of imaging systems so that the resulting reconstructed images often have insufficient quality for practical use.

Model-based image reconstruction (MBIR) methods, instead, first construct a mathematical model to represent the physics of imaging system, then solve an optimization problem built on top of that model. MBIR is a family of nonlinear reconstruction methods that has shown to be both flexible to a variety of inverse problems, and able to produce improved quality estimates. The following section discusses it in more detail.

2.1.1 Model-Based Image Reconstruction

From the Bayesian statistics perspective, the reconstruction problem can be described as maximizing the probability of our estimate given our measurements, *i.e.*, the posterior distribution $\mathsf{p}(\boldsymbol{x}|\boldsymbol{y})$. Using Bayes' rule, we can rewrite the posterior as

$$\mathsf{p}(\boldsymbol{x}|\boldsymbol{y}) = \frac{\mathsf{p}(\boldsymbol{y}|\boldsymbol{x})\,\mathsf{p}(\boldsymbol{x})}{\mathsf{p}(\boldsymbol{y})}. \tag{2.2}$$

Thus, to maximize the left-hand side, we equivalently maximize the right-hand side. When maximizing w.r.t. $\boldsymbol{x}$, we can drop $\mathsf{p}(\boldsymbol{y})$ as a constant scaling, and apply $-\log(\cdot)$ to write the problem as

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} -\log(\mathsf{p}(\boldsymbol{y}|\boldsymbol{x})) - \log(\mathsf{p}(\boldsymbol{x}))\,. \tag{2.3}$$

If the term $\log(\mathsf{p}(\boldsymbol{x}))$ is absent or $\boldsymbol{x}$ is assumed to follow a uniform distribution so that $\log(\mathsf{p}(\boldsymbol{x}))$ is a constant, (2.3) reduces to MLE, which is a special case of an extremum estimator from the frequentist perspective. In the context of MAP estimation, the first term is the negative log-likelihood and encodes the measurement dependence of our estimate. In the case of independent Poisson measurements as in (2.1), its negative log-likelihood is proportional to

$$\mathbf{1}'(\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}) - \boldsymbol{y}'\log(\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}})\,.$$

The second term, referred to as the negative log-prior, or just simply the prior, encodes our preconceived assumptions about which signals $\boldsymbol{x}$ are more probable irrespective of measurements.

MBIR methods relax the probabilistic interpretation of (2.3). To distinguish from the MAP interpretation, we refer to the first term as the *data-fidelity* term instead of the negative log-likelihood, and the second term as *regularization*, instead of a prior. These MAP terms are often used informally though, despite the lack of statistical interpretability of many regularization functions. Thus, in MBIR a common estimation problem for (2.1) may be posed as

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x};\beta), \quad f(\boldsymbol{x}) \triangleq \mathbf{1}'(\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}) - \boldsymbol{y}'\log(\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}), \tag{2.4}$$

where $\mathsf{R}(\cdot)$ is the regularization function penalizing deviation from our signal model, and β is a hyperparameter representing a trade-off between fit to data and $\mathsf{R}(\cdot)$.

As a concrete example, we could model a signal following Gaussian distribution with the ℓ_2 norm regularizer, $\mathsf{R}(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2$. Such an estimate often has a closed-form solution. If that solution is expensive to compute directly, it can be obtained via optimization algorithms such as gradient-based methods.

Developing effective signal and image models and corresponding regularization functions is challenging. Ideally, such models should be both broad enough to describe all plausible latent images, while discriminating enough to reject noise and artifacts.

2.1.2 Deep Learning for Inverse Problems

It can be difficult to come up with an explicit regularizer that can describe all plausible true images, so an option is to train a neural network to implicitly model the distribution of the latent images. Then we can balance the estimation between the likelihood and the output of neural network by various methods such as

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\beta}{2}\|\boldsymbol{x} - g_{\boldsymbol{\theta}}(\boldsymbol{x})\|_2^2, \tag{2.5}$$

where $f(\boldsymbol{x})$ is defined in (2.4), $g_{\boldsymbol{\theta}}$ denotes a class, or an *architecture*, of functions parameterized by *weights* $\boldsymbol{\theta}$. One can train $g_{\boldsymbol{\theta}}$ by minimizing some loss function that provides a metric of the quality of $\hat{\boldsymbol{x}}$ compared to the true signal $\boldsymbol{x}_{\text{true}}$. As (2.5) often does not have a closed-form solution, generally it is is minimized via some variant of stochastic (sub)gradient descent.

In artificial *neural networks*, the architecture $g_{\boldsymbol{\theta}}$ is designed as a composition of simpler functions, or *layers* that has learnable parameters, *e.g.*, fully connected layers, convolutional layers; and other untrained layers such as pooling layers and activation functions like RELU. Optimizing the parameters in such a layered structure requires efficient computation of parameter gradients, which can be achieved by applying the chain rule and backpropagating the results to earlier layers.

Combining different layers with different operations results in numerous possible neural-network architectures. For example, adding the output of previous layer to the current layer leads to the famous “ResNet” [86]; combining a series of downsampling and upsampling layers with concatenation on the channel dimension results in a U-shape network known as U-Net [184]. These deep neural networks are reported to provide SOTA performance on many imaging tasks. For example, U-Net achieved SOTA accuracy on the 2012 International Symposium on Biomedical Imaging (ISBI) challenge for segmentation of neuronal structures in electron microscopic stacks [184].

Despite these successes, modern deep learning methods have several disadvantages when it comes to solving inverse problems. First, it can be computationally expensive and memory hungry to backpropagate through the system matrix in unrolled iterative algorithms; and any change to the algorithm may require retraining the network. Another weakness is the concern of interpretability, *i.e.*, what kind of features that the neural network learns from the data. Such features can be hard to visualize for a deep network, *e.g.*, having millions or billions of parameters. Additionally, deep learning methods can generate unexpected and even unrealistic results, for example, Nataraj and Otazo [167] found that in some test cases of MRI brain scans with different pathologies, the trained networks removed some important organs such as very large tumors because they rarely showed up in training data.

In light of these strengths and weakness, one must keep in mind that the "best" algorithm may not exist, so that an algorithm that is adaptive to different scenarios can be more favorable, *e.g.*, having some tuning parameters that can be left to users to optimize. For example, in deep learning regularized MBIR, if one does not trust the neural network, one can set the regularization parameter in (2.5) to zero so that the algorithm reduces to traditional MLE.

2.2 Phase Retrieval [1]

Phase retrieval is an inverse problem with many applications in engineering and applied physics [96, 84], including radar [98], X-ray crystallography [161], astronomical imaging [48], Fourier ptychography [16, 253, 236, 212] and CDI [124]. In these applications, the sensing systems can only measure the magnitude (or the square of the magnitude) of the signal, which leads to

$$y_i \sim \mathsf{p}\big(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + b_i\big), \tag{2.6}$$

where $\mathsf{p}(\cdot)$ denotes a probability density function. Here, $\boldsymbol{a}_i' \in \mathbb{C}^N$ denotes the ith row of the system matrix $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ where $i = 1, \dots, M$, and $b_i \in \mathbb{R}_+$ denotes a known mean background signal for the ith measurement, e.g., as arising from dark current [200]. Fig. 2.1 illustrates the phase retrieval problem.

Usually, the sensing vectors $\{\boldsymbol{a}_i'\}$ are assumed to follow some structures, *e.g.*, IID random Gaussian, or the coefficients of DFT. For the random Gaussian case, Candes *et al.* [25] showed that $M \sim \mathcal{O}(N \log N)$ samples are sufficient to recover the signal; Bandeira *et al.* [7] posed a conjecture that $M = 4N - 4$ is necessary and sufficient to uniquely recover

[1] This section is largely taken from [139].

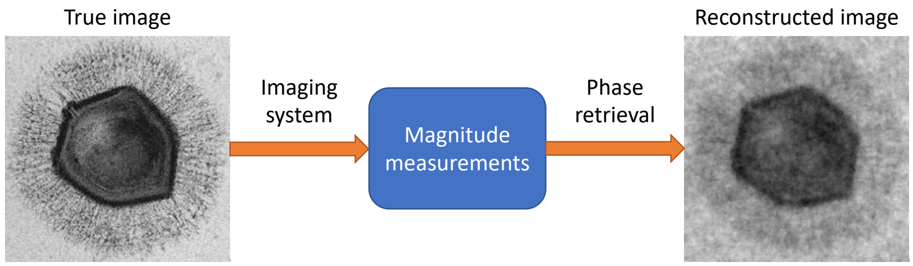


FIG 2.1 – Illustration of the phase retrieval problem.

the original signal from noiseless measurements. However, under very low-count regimes with noise, a much larger M is often needed to successfully reconstruct the signal. Additionally, when $\boldsymbol{A}$ corresponds to a Fourier transform, the measurements describe only the magnitudes of a signal's Fourier coefficients, and one usually does not have enough information to recover the signal; while the Fourier transform is injective, its point-wise absolute value is not [8]. So a common approach is to create redundancy in the measurement process by additional illuminations of the object using different masks [24]. Banderia *et al.* [8] showed that by using a set of $\mathcal{O}(\log M)$ random masks can increase the probability of recovering the signal.

Noise in the acquired measurements is another factor that has significant effects to the reconstruction quality. As the Gaussian noise and Poisson noise are the most common ones arise in imaging systems, we present an overview of phase retrieval with these noise models next.

2.2.1 Gaussian Phase Retrieval

In many previous works, the measurement vector $\boldsymbol{y} \in \mathbb{R}^M$ was assumed to have statistically independent elements following Gaussian distributions with variance σ^2:

$$y_i \sim \mathcal{N}(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + b_i, \sigma^2). \tag{2.7}$$

For this Gaussian noise model, the MLE of $\boldsymbol{x}$ corresponds to the following non-convex optimization problem

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{F}^N} g(\boldsymbol{x}), \quad g(\boldsymbol{x}) \triangleq \sum_i \left| y_i - b_i - \left|\boldsymbol{a}_i'\boldsymbol{x}\right|^2 \right|^2. \tag{2.8}$$

Here we overload the notation g (different from g in (2.5)). The field $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ depending on whether $\boldsymbol{x}$ is known to be real or complex. To solve (2.8), numerous algorithms have been proposed. One approach reformulates (2.8) by "matrix lifting" [25, 26, 194], where a rank-one matrix is introduced and if the rank constraint is relaxed, then the transformed problem is convex and can be solved by SDP. The SDP based algorithms can yield robust solutions but can be computationally expensive, especially on large-scale data. Another approach is WF [26] and its variants [102, 23, 203] that descend the cost function with a (projected/thresholded/truncated) Wirtinger gradient using an appropriate step size. In the classic WF algorithm [26], the gradient[2] for the Gaussian cost function (2.8) is

$$\nabla g(\boldsymbol{x}) = 4\boldsymbol{A}' \operatorname{diag}\{|\boldsymbol{A}\boldsymbol{x}|^2 - \boldsymbol{y} + \boldsymbol{b}\}\boldsymbol{A}\boldsymbol{x}. \tag{2.9}$$

To descend the cost function, Candes *et al.* [26] used a heuristic where the step size μ is rather small for the first few iterations and gradually increases as the iterations proceed. The intuition is that the gradient is noisy at the early iterations so a small step size is preferred. A drawback of this approach is that one needs to select hyper-parameters that control the growth of μ. An alternative approach is to perform backtracking for μ at each iteration [177], *i.e.*, by reducing μ until the cost function decreases sufficiently. This approach guarantees decreasing the cost function monotonically but can increase the compute time of the algorithm due to the variable number of inner iterations. Jiang *et al.* [102] derived the optimal step size for (2.8) and showed faster convergence rate than the heuristic step size when measurements are noiseless or follow IID Gaussian distribution. Cai *et al.* [23] proposed thresholded WF and showed it can achieve the minimax optimal rates of convergence, but that scheme requires an appropriate selection of tuning parameters. Soltanolkotabi [203] reformulated the phase retrieval problem as a nonconvex optimization problem and proved that projected Wirtinger gradient descent, when initialized in a neighborhood of the desired signal, has a linear convergence rate. However, it can be difficult to find an initial estimate satisfying the conditions mentioned in [203].

An alternative to cost function (2.8) (aka intensity model) is the magnitude model that works with the square root of $\boldsymbol{y}$. In particular, Gerchberg and Saxton [69] proposed an algorithm known as GS that introduced a new variable $\boldsymbol{\theta}$ to represent the phase, leading to the following joint optimization problem

$$\begin{aligned} \hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}} = &\underset{\boldsymbol{x}\in\mathbb{F}^N,\, \boldsymbol{\theta}\in\mathbb{C}^N}{\arg\min} \left\| \boldsymbol{A}\boldsymbol{x} - \operatorname{diag}\left\{\sqrt{\max\left(\boldsymbol{y}-\boldsymbol{b}, \boldsymbol{0}\right)}\right\}\boldsymbol{\theta}\right\|_2^2, \\ &\text{subject to} \quad |\theta_i| = 1,\ i = 1, \dots, N. \end{aligned} \tag{2.10}$$

[2] If $\boldsymbol{x} \in \mathbb{R}^N$, then all gradients w.r.t. $\boldsymbol{x}$ should be real and hence use only the real part of expressions like (2.9).

The square root in (2.10) is reminiscent of the Anscombe transform that converts a Poisson random variable into another random variable that approximately has a standard Gaussian distribution. However, that approximation is accurate when the Poisson mean is sufficiently large (*e.g.*, above 5), whereas we focuses on the lower-count regime in this work. The convergence and recovery guarantees of GS were studied in [168, 220]. In addition to matrix-lifting, WF, GS and their variants, several other algorithms have been proposed to solve phase retrieval problems under the assumption of the Gaussian measurement noise, including Gauss-Newton methods [68], LBFGS updates to approximate the Hessian in the Newton's method [99], MM methods [177], ADMM [141], and an iterative soft-thresholding with exact line search algorithm (STELA) [242].

2.2.2 Poisson Phase Retrieval

In many low-photon count applications [210, 80, 236, 9, 216, 126, 72], especially in [72], where 0.25 photon per pixel on average is considered, a Poisson noise model is more appropriate:

$$y_i \sim \text{Poisson}(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + b_i). \tag{2.11}$$

MLE of $\boldsymbol{x}$ for the model (2.11) corresponds to the following optimization problem

$$\begin{aligned} \hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{F}^N}{\arg\min}\, f(\boldsymbol{x}), \quad f(\boldsymbol{x}) &\triangleq \sum_i \psi(\boldsymbol{a}_i'\boldsymbol{x}; y_i, b_i), \\ \psi(v; y, b) &\triangleq (|v|^2 + b) - y\log\left(|v|^2 + b\right). \end{aligned} \tag{2.12}$$

Here, $f(\boldsymbol{x})$ denotes the negative log-likelihood corresponding to (2.11), ignoring irrelevant constants independent of $\boldsymbol{x}$, and the function $\psi(\cdot; y, b)$ denotes the marginal negative log-likelihood for a single measurement, where $v \in \mathbb{C}$. Because $|v|$ is real, it is helpful to re-write ψ in the form $\psi(v; y, b) = \phi(|v|; y, b)$, where

$$\phi(r; y, b) \triangleq (r^2 + b) - y\log\left(r^2 + b\right), \quad r \in \mathbb{R}_+. \tag{2.13}$$

One can verify that the function $\phi(r; y, b)$ is non-convex over $r \in \mathbb{R}_+$ when $0 < b < y$. That property, combined with the modulus within the logarithm in (2.12), makes (2.12) a challenging optimization problem. Similar problems for $b = 0$ have been considered previously [39, 16, 24, 37, 29], but many optical sensors also have Gaussian readout noise [253, 107] so that the mean background signal is unlikely to be zero. To accommodate the Gaussian readout noise, a more precise model would consider a sum of Gaussian and Poisson noise. However, the log likelihood for a Poisson plus Gaussian distribution is complicated, so a common approximation is to use a shifted Poisson model [199] that also leads to the cost function in (2.12). An alternative to the shifted Poisson model could be to work with an unbiased inverse transformation of a generalized Anscombe transform approximation [150, 212] or use a surrogate function that tightly upper bounds the challenging Poisson plus Gaussian maximum likelihood objective function and apply an MM algorithm.

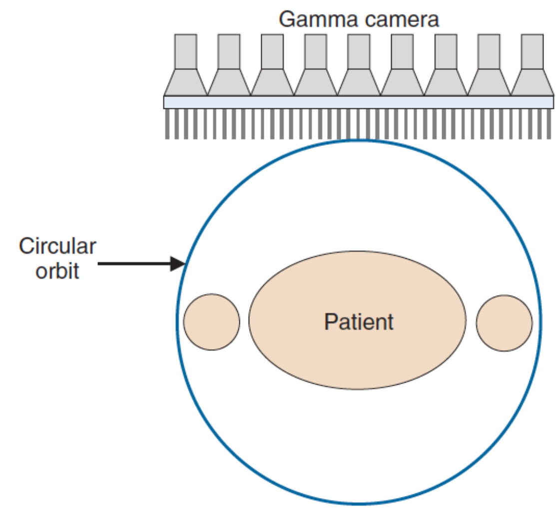
This work focuses on Poisson phase retrieval algorithms working in very low-count regimes that are discussed in more detail in Chapter 3.

2.3 Quantitative SPECT Imaging

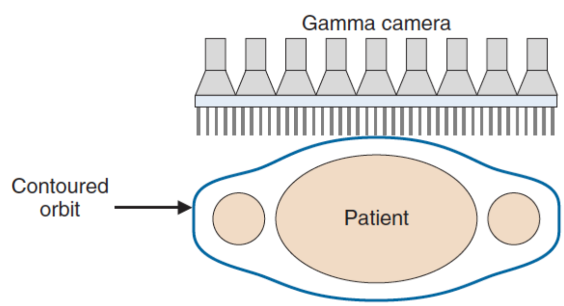
SPECT is a nuclear medicine technique that images spatial distributions of radioisotopes, by detecting gamma-rays that escape from the patient's body. A rotating gamma camera is used to acquire data for computed tomography imaging. SPECT plays a pivotal role in clinical diagnosis such as cardiac vascular diseases [33], tumor detection [174], and also to estimate radiation absorbed doses in nuclear medicine therapies [56]. For example, quantitative SPECT imaging with Lutetium-177 (^{177}Lu) in targeted radionuclide therapy (such as ^{177}Lu DOTATATE) is important in determining dose-response relationships in tumors and holds great potential for dosimetry-based individualized treatment [95]. Quantitative Yttrium-90 (^{90}Y) SPECT bremsstrahlung imaging is also valuable for estimating the activity distribution after radioembolization procedures for safety and absorbed dose verification [245].

2.3.1 SPECT Physics

Most SPECT systems are based on Gamma camera detectors that rotate around the patient body in circular or contoured orbit, as shown in Fig. 2.2. Each detector head is equipped with a parallel-hole collimator and can rotate independently to sample different projection angles, enabling a projection dataset to be acquired for tomographic reconstruction. Ideally, the signal recorded by one detector pixel at a certain rotation angle would be linearly proportional to the amount of activity contained along the ray through the patient corresponding to the location of that pixel. In practice, however, this ideal is not achieved due to the imaging system physics and statistics. For example, for a parallel-beam collimator shown in Fig. 2.3, its line spread response ideally would be an extended cylinder but actually it resembles a diverging cone [38, p. 285]. Furthermore, scattering events can be estimated only approximately, which can have a significant effect on the quality of the reconstructed image. Hence for almost all MBIR algorithms, a simplified SPECT imaging



(a)



(b)

FIG 2.2 – Illustration of circular (*top*) and contoured (*bottom*) orbits. Figure adapted from [38, p. 280].

model is used, as shown in Fig. 2.3, that models parallel-beam collimators, with depth-dependent attenuation and collimator point spread response, and any nonidealities are not modelled.

2.3.2 Scatter

Scattering is one way of interaction of radiation with matter. For example, Compton scattering is a "collision" between a photon and an electron of an atom; coherent scattering occurs between a photon and an atom as a whole. In that case, the atom often has a great mass so that the photon is deflected with essentially no loss of energy.

Scattered photons cannot be overlooked because they can be as large as 40% of non-scattered photons [38, p. 297], so that the presence of scattered events can result in reduced

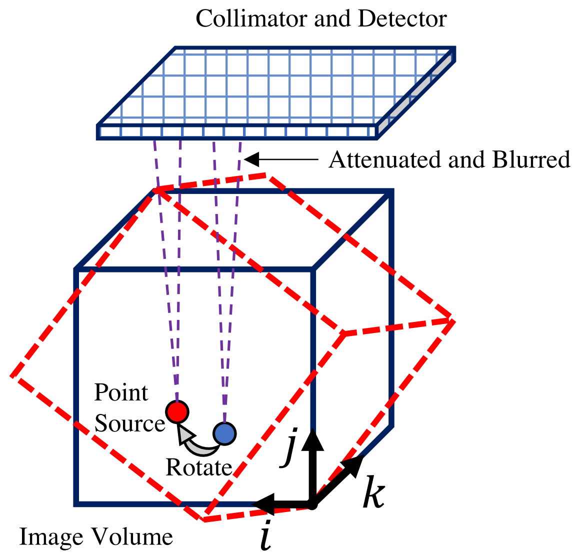


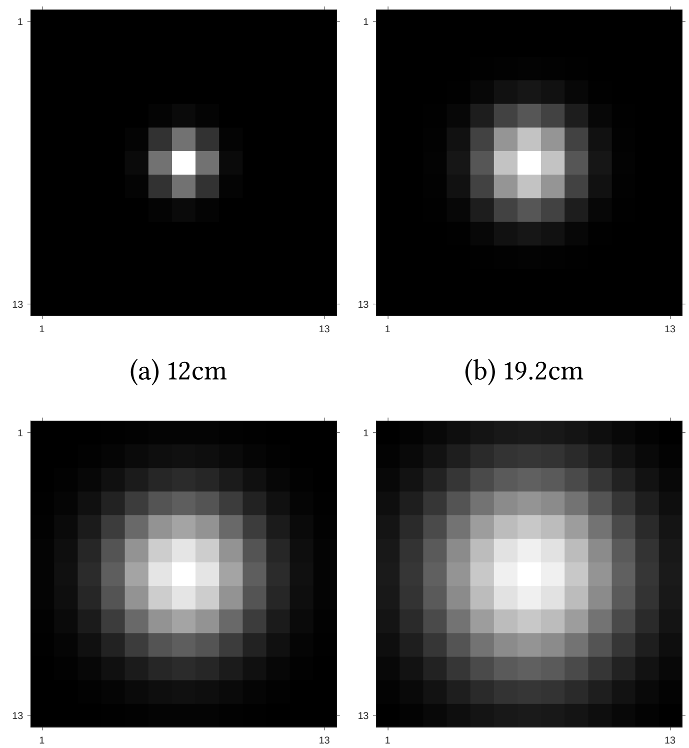
FIG 2.3 – SPECT imaging model for parallel-beam collimators, with depth-dependent attenuation and collimator point spread response.

image contrast and loss of important details. Therefore, methods for estimating scatters are essential in SPECT reconstruction.

One of the most commonly used methods for scatter correction is to simultaneously acquire counts with a photopeak window and two neighbouring scatter windows (known as the TEW approach). Then the acquired scatters are multiplied by a weighting factor that is determined experimentally. The TEW method is fast but can be inaccurate, especially for ^{90}Y bremsstrahlung photons where the energy spectrum is continuous. Another method is running MC simulations that fully model the physics of photon transport in the patient and camera and hence can provide more accurate scatter estimation; but that method requires running sufficiently large number of histories to generate results with low uncertainty, which can be computationally expensive. Several works proposed to use a deep neural network for scatter estimation, with the measured SPECT emission projection and attenuation map as inputs, and showed comparably accurate results as the MC method with faster compute time [231, 113, 101, 100].

2.3.3 Spatial Resolution

A SPECT camera is also known to suffer from limited spatial resolution due to the point (line) spread response of the collimator. Such blurring effects can significantly degrade the quality of reconstruction. For example, as shown in Fig. 2.4, the FWHM becomes very wide at larger distances, leading to a even more ill-posed reconstruction problem to solve.



(a) 12cm (b) 19.2cm

(c) 28.8cm (d) 39.2cm

FIG 2.4 – Point spread function at different depth locations (distance from collimator).

2.3.4 MLEM and OSEM

Numerous reconstruction algorithms have been proposed for SPECT reconstruction, of which the most popular ones are MBIR algorithms such as MLEM [197] and its variant OSEM [91]. These methods first construct a mathematical model for the SPECT imaging system, then maximize the (log-)likelihood using an appropriate statistical noise model (*e.g.*, Poisson noise). In particular, the cost function for MLEM has the form

$$f(\boldsymbol{x}) \triangleq \sum_{i=1}^{M} h([\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}]_i; y_i), \quad h(t; y) \triangleq t - y\log(t), \quad y \geq 0, t > 0, \tag{2.14}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ denotes the SPECT system matrix, $\boldsymbol{y} \in \mathbb{R}^M$ denotes total projections and $\bar{\boldsymbol{r}} \in \mathbb{R}^N$ denotes the mean of background events like scatters. To make the cost function separable, we follow the derivation in [52]:

$$\begin{aligned}
[\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}]_i &= \sum_{j=1}^{N} a_{ij}(x_j + \gamma_j) + \tilde{r}_i \\
&= \sum_{j=1}^{N} \left(\frac{a_{ij}\left(x_j^{(k)} + \gamma_j\right)}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} \right) \frac{x_j + \gamma_j}{x_j^{(k)} + \gamma_j} [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i + \frac{\tilde{r}_i}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i,
\end{aligned} \tag{2.15}$$

where k denotes the iteration number in MLEM or OSEM, j denotes the jth voxel; $\tilde{r}_i$ are non-negative constants

$$\tilde{r}_i \triangleq \bar{r}_i - \sum_{j=1}^{N} a_{ij}\gamma_j \geq 0. \tag{2.16}$$

As explained in [64], γ_j are any (user-selected) non-negative constants that satisfy the following constraints

$$\sum_{j=1}^{N} a_{ij}\gamma_j \leq \bar{r}_i, \quad i = 1, ..., M. \tag{2.17}$$

Because $h(\boldsymbol{x})$ is convex, using Jensen's inequality, we have

$$\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=1}^{M} h([\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}]_i; y_i) \\
&= \sum_{i=1}^{M} h\left(\sum_{j=1}^{N} \left(\frac{a_{ij}\left(x_j^{(k)} + \gamma_j\right)}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} \right) \frac{x_j + \gamma_j}{x_j^{(k)} + \gamma_j} [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i + \frac{\tilde{r}_i}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i \right) \\
&\leq \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\frac{a_{ij}\left(x_j^{(k)} + \gamma_j\right)}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} \right) h\left(\frac{x_j + \gamma_j}{x_j^{(k)} + \gamma_j} [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i \right) + \frac{\tilde{r}_i}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} h\left([\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i\right) \\
&= \sum_{j=1}^{N} \sum_{i=1}^{M} \left(\frac{a_{ij}\left(x_j^{(k)} + \gamma_j\right)}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} \right) h\left(\frac{x_j + \gamma_j}{x_j^{(k)} + \gamma_j} [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i \right) + \frac{\tilde{r}_i}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} h\left([\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i\right).
\end{aligned} \tag{2.18}$$

Thus, a majorizer can be constructed as

$$
\begin{aligned}
&H(\boldsymbol{x};\boldsymbol{x}_k) \triangleq \sum_{j=1}^{N} H_j(x_j;\boldsymbol{x}_k),\\
&H_j(x_j;\boldsymbol{x}_k) \triangleq \sum_{i=1}^{M} \left(\frac{a_{ij}\left(x_j^{(k)}+\gamma_j\right)}{[\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}]_i} \right) h\left(\frac{x_j+\gamma_j}{x_j^{(k)}+\gamma_j}[\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}]_i \right) + \frac{\tilde{r}_i}{[\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}]_i} h\left([\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}]_i\right).
\end{aligned}
\tag{2.19}
$$

Differentiating and setting the derivative of $H_j(x_j;\boldsymbol{x}_k)$ to zero leads to

$$
\begin{aligned}
0 = \frac{\partial H_j(x_j;\boldsymbol{x}_k)}{\partial x_j} &= \sum_{i=1}^{M} a_{ij}\dot{h}\left(\frac{x_j+\gamma_j}{x_j^{(k)}+\gamma_j}[\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}]_i \right)\\
&= \sum_{i=1}^{M} a_{ij} - \left(\frac{x_j^{(k)}+\gamma_j}{x_j+\gamma_j} \right) \sum_{i=1}^{M} a_{ij} \frac{y_i}{[\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}]_i},
\end{aligned}
\tag{2.20}
$$

after simplification yields

$$
\sum_{i=1}^{M} a_{ij}\left(x_j+\gamma_j\right) = \left(x_j^{(k)}+\gamma_j\right) \sum_{i=1}^{M} a_{ij} \frac{y_i}{[\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}]_i}. \tag{2.21}
$$

Setting $\boldsymbol{\gamma}$ to zero, as a common special case, leads to the following update for $\boldsymbol{x}_k$

$$
\boldsymbol{x}_{k+1} = \boldsymbol{x}_k \odot \left(\boldsymbol{A}'\left(\boldsymbol{y} \oslash \left(\boldsymbol{A}\boldsymbol{x}_k+\bar{\boldsymbol{r}}\right)\right)\right) \oslash \left(\boldsymbol{A}'\mathbf{1}\right). \tag{2.22}
$$

For OSEM, the system matrix $\boldsymbol{A}$ is split into ordered-subsets that usually correspond to a uniform subset of view angles (roughly uniform sampling of rows of $\boldsymbol{A}$), which leads to faster computation at each iteration.

2.3.5 Deep Learning Regularized EM

To minimize a cost function like (2.5) that has a deep learning regularizer, a common approach is to use alternating minimization. Specifically, with variable splitting $\boldsymbol{u} \triangleq g_{\boldsymbol{\theta}}(\boldsymbol{x})$, one can perform the following alternating update [142]:

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} \sum_{i=1}^{M} h([\boldsymbol{A}\boldsymbol{x}+\bar{\boldsymbol{r}}]_i; y_i) + \frac{\beta}{2}\|\boldsymbol{x}-\boldsymbol{u}_k\|_2^2,\\
\boldsymbol{u}_{k+1} &= g_{\boldsymbol{\theta}}(\boldsymbol{x}_{k+1}).
\end{aligned}
\tag{2.23}
$$

Then the derivative of the majorizer (2.20) becomes

$$\begin{aligned}
0 = \frac{\partial H_j(x_j; \boldsymbol{x}_k)}{\partial x_j} &= \sum_{i=1}^{M} a_{ij} \dot{h} \left(\frac{x_j + \gamma_j}{x_j^{(k)} + \gamma_j} [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i \right) + \beta \left(x_j - [\boldsymbol{u}_k]_j \right) \\
&= \sum_{i=1}^{M} a_{ij} \left(1 - \frac{y_i \left(x_j + \gamma_j \right)}{\left(x_j + \gamma_j \right) [\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} \right) + \beta \left(x_j - [\boldsymbol{u}_k]_j \right) \\
&= \sum_{i=1}^{M} a_{ij} - \left(\frac{x_j + \gamma_j}{x_j + \gamma_j} \right) \sum_{i=1}^{M} a_{ij} \frac{y_i}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i} + \beta \left(x_j - [\boldsymbol{u}_k]_j \right).
\end{aligned}$$

Again, let

$$e_j(\boldsymbol{x}_k) = \sum_{i=1}^{M} a_{ij} \frac{y_i}{[\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}}]_i}, \quad u_j(\beta) = \sum_{i=1}^{M} a_{ij} - \beta [\boldsymbol{u}_k]_j, \tag{2.24}$$

the derivative simplifies to

$$\beta x_j^2 + \left(\beta \gamma_j + u_j(\beta) \right) x_j + \gamma_j u_j(\beta) - \left(x_j + \gamma_j \right) e_j(\boldsymbol{x}_k) = 0. \tag{2.25}$$

The resulting vector update is

$$\hat{\boldsymbol{x}}_k = \frac{1}{2\beta} \left(-\boldsymbol{u}(\beta) + \sqrt{\boldsymbol{u}(\beta)^2 + 4\beta \boldsymbol{x}_k \odot \boldsymbol{e}\left(\boldsymbol{x}_k \right)} \right), \tag{2.26}$$

where

$$\boldsymbol{e}\left(\boldsymbol{x}_k \right) = \boldsymbol{A}' \left(\boldsymbol{y} \oslash \left(\boldsymbol{A}\boldsymbol{x}_k + \bar{\boldsymbol{r}} \right) \right), \quad \boldsymbol{u}(\beta) = \boldsymbol{A}' \boldsymbol{1} - \beta \boldsymbol{u}_k. \tag{2.27}$$

To compute $\boldsymbol{x}_{k+1}$, one must substitue $\hat{\boldsymbol{x}}_k$ back into $\boldsymbol{e}(\cdot)$ in (2.27), and repeat. Algorithm 1 summarizes the DL-regularized EM algorithm.

Algorithm 1: Deep learning regularized EM algorithm for SPECT image reconstruction.

Input: 3D projection measurements $\boldsymbol{y}$,
3D background measurements $\bar{\boldsymbol{r}}$,
system model $\boldsymbol{A}$, initial guess $\boldsymbol{x}_0$,
trained deep neural network $g_{\boldsymbol{\theta}}$,
outer iterations K

for $k = 0, ..., K - 1$ **do**
    $\boldsymbol{u}_{k+1} = g_{\boldsymbol{\theta}}(\boldsymbol{x}_k)$
    $\boldsymbol{x}_{k+1} \leftarrow$ repeat (2.26) until convergence tolerance or maximum # of inner iterations is reached
end

Output: $\boldsymbol{x}_K$

2.3.6 Dosimetry Estimation

Absorption of energy from ionizing radiation of radiotracers used for SPECT imaging or for dosimetry can cause damage to living tissues, so it is necessary to analyze the energy distribution in body tissues quantitatively to ensure an accurate therapeutic prescription to tumors or to assess potential risks to normal organs [38, p. 407]. Dosimetry estimation is also essential for clinical implementation of dosimetry-guided treatment planning in radionuclide therapy.

Radiation dose is defined as the quantity of radiation energy deposited in absorber per gram of absorber material. The basic unit is *gray* (Gy in short):

$$1\text{Gy} = 1 \text{ joule energy deposited per kg absorber.} \tag{2.28}$$

Chapter 4 discusses more details about dosimetry estimation methods and describe a new method for accurate and computationally efficient absorbed dosimetry estimation using a DL approach.

CHAPTER 3

Poisson Inverse Problems in Phase Retrieval[1]

3.1 Poisson Phase Retrieval in Low-count Regimes[2]

3.1.1 Motivation

Existing algorithms for Poisson phase retrieval are limited in the literature. Chen and Candes [37] proposed to solve the Poisson phase retrieval problem by minimizing a nonconvex functional as in the WF approach; Bian *et al.* [16] used Poisson MLE and truncated WF in Fourier ptychographic (FP) reconstruction. Zhang *et al.* [249] consider a scale square root of (2.11) for the common case with $b_i = 0$. Chang *et al.* [30] derived a TV-regularized ADMM algorithm for Poisson phase retrieval and established its convergence. Recently, Fatima *et al.* [63] proposed a double looped primal-dual majorize-minimize (PDMM) algorithm.

In this work, we propose novel algorithms for the Poisson phase retrieval problem and report empirical comparisons of the convergence speed and reconstruction quality of algorithms under a variety of experimental settings. In particular,

1. We propose a novel method for computing the step size for the WF algorithm that can lead to faster convergence compared to empirical step size [26], backtracking line search [177], optimal step size derived for the Gaussian noise model [102], and LBFGS updates to approximate the Hessian in Newton's method [99]. Moreover, our proposed method can be computed efficiently without any tuning parameter.

2. We derive an MM algorithm with quadratic majorizer using a novel curvature. We show theoretically that our proposed curvature is sharper than the curvature derived from the upper bound of the second derivative of the Poisson maximum likelihood cost function.

[1]This chapter is based on [138, 139, 137].
[2]This section is based on [139].

3. We present numerical simulation results under random Gaussian, canonical DFT, masked DFT and empirical transmission system matrix settings for low-count data, *e.g.*, 0.25 photon per pixel. We show that under such experimental settings, algorithms derived from the Poisson noise model produce consistently higher reconstruction quality than algorithms derived from Gaussian noise model, as expected. Furthermore, the reconstruction quality is further improved by incorporating regularizers that exploit assumed properties of the signal.

4. We compare the convergence speed (in terms of cost function and PSNR vs. time) of WF with Fisher information with other methods for step size (backtracking line search, optimal Gaussian) and LBFGS quasi-Newton method. We also compare the convergence speed of regularized WF with MM and ADMM, using smooth regularizers such as corner-rounded anisotropic TV. For both cases, our proposed WF Fisher algorithm converges the fastest under all system matrix settings.

Overloaded Notations: $\boldsymbol{A}, \boldsymbol{x}, \boldsymbol{b}, \boldsymbol{y}$ used in this chapter are defined in (2.6); $f(\cdot)$ and $\psi(\cdot)$ are defined in (2.12); $\phi(\cdot)$ is defined in (2.13).

3.1.2 Methods

3.1.2.1 Wirtinger flow (WF)

This section describes the modified WF algorithm with proposed step-size approach based on Fisher information. To generalize the WF algorithm to the Poisson cost function (2.12), the most direct approach simply replaces the gradient (2.9) by (3.1) in the WF framework [249] and perform backtracking the find a good step size μ. We propose a faster alternative next. We treat $0 \log 0$ as 0 in (2.12) because a Poisson random variable with zero mean can only take the value 0. With this assumption, one can verify that ψ has the following well-defined ascent direction (negative of descent direction [250]) and a second derivative:

$$\begin{aligned}
\dot{\psi}(v; y, b) &= 2v\left(1 - \frac{y}{|v|^2 + b}\right), \quad v \in \mathbb{C}. \\
\ddot{\psi}(v; y, b) &= \mathrm{sign}(v)\left(2 + 2y\frac{|v|^2 - b}{(|v|^2 + b)^2}\right), \\
|\ddot{\psi}(v; y, b)| &\leq 2 + \frac{y}{4b}.
\end{aligned} \tag{3.1}$$

Fisher information for Poisson model. We first make a quadratic approximation along the gradient direction of the cost function at each iteration, and then apply one step of Newton's method to minimize that 1D quadratic. Because computing the Hessian can be computationally expensive in large-scale problems, we follow the statistics literature by replacing the Hessian by the observed Fisher information when applying Newton's method [213, 123]. Our Fisher approach is based on the fact that the observed Fisher information is the negative Hessian matrix of the incomplete data log-likelihood functions evaluated at the observed data, and hence can provide a good approximation to the Hessian with enough data [158]. Moreover, the Fisher information matrix is always PSD, and avoids calculation of second derivatives. Using Fisher information in gradient-based algorithms has a long history in statistics and is central to Fisher's method of scoring [213, 170, 92, 123].

Specifically, we first approximate the 1D line search problem associated with (2.12) by the following Taylor series

$$\mu_k = \arg\min_{\mu \in \mathbb{R}} f_k(\mu), \tag{3.2}$$
$$f_k(\mu) \triangleq f(\boldsymbol{x}_k - \mu \nabla f(\boldsymbol{x}_k)) \approx f(\boldsymbol{x}_k) - \|\nabla f(\boldsymbol{x}_k)\|_2^2 \, \mu + \frac{1}{2} \nabla f(\boldsymbol{x}_k)' \nabla^2 f(\boldsymbol{x}_k) \nabla f(\boldsymbol{x}_k) \mu^2,$$

where one can verify that the minimizer is

$$\mu_k = \frac{\|\nabla f(\boldsymbol{x}_k)\|_2^2}{\operatorname{real}\{\nabla f(\boldsymbol{x}_k)' \nabla^2 f(\boldsymbol{x}_k) \nabla f(\boldsymbol{x}_k)\}}. \tag{3.3}$$

We next approximate the Hessian matrix $\nabla^2 f(\boldsymbol{x})$ using the observed Fisher information matrix:

$$\begin{aligned} \nabla^2 f(\boldsymbol{x}) &\approx \boldsymbol{I}(\boldsymbol{x}, \boldsymbol{b}) \\ &\triangleq \mathbb{E}_{\boldsymbol{y}} \left[\nabla^2 f(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{b}) \middle| \boldsymbol{x}, \boldsymbol{b} \right] \\ &= \mathbb{E}_{\boldsymbol{y}} \left[(\nabla f(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{b})) \, (\nabla f(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{b}))' \middle| \boldsymbol{x}, \boldsymbol{b} \right] \\ &= \boldsymbol{A}' \mathbb{E}_{\boldsymbol{y}} \left[\left(\dot{\psi}.(\boldsymbol{v}; \boldsymbol{y}, \boldsymbol{b}) \right) \left(\dot{\psi}.(\boldsymbol{v}; \boldsymbol{y}, \boldsymbol{b}) \right)' \middle| \boldsymbol{v}, \boldsymbol{b} \right] \boldsymbol{A}. \end{aligned} \tag{3.4}$$

Here the dot subscript notation $\dot{\psi}.(\boldsymbol{v}; \boldsymbol{y}, \boldsymbol{b})$ denotes element-wise application of the function $\dot{\psi}$ to its arguments (as in the Julia language), so the gradient $\dot{\psi}.(\boldsymbol{v}; \boldsymbol{y}, \boldsymbol{b})$ is a vector in $\mathbb{C}^M$. One can verify that the marginal Fisher information for a single term $\psi(v; y, b)$ is

$$\bar{I}(v, b) = \mathbb{E}_y \left[\left| \dot{\psi}(v; y, b) \right|^2 \middle| v, b \right] = \frac{4|v|^2}{|v|^2 + b}, \quad v \in \mathbb{C}, \; b > 0. \tag{3.5}$$

Substituting (3.5) into (3.4) using the statistical independence of the elements of the gradient vector, and then substituting (3.4) into (3.3) yields the simplified step-size expression

$$\mu_k \triangleq \frac{\|\nabla f(\boldsymbol{x}_k)\|_2^2}{\boldsymbol{d}_k' \, \boldsymbol{D}_1 \, \boldsymbol{d}_k} \in \mathbb{R}_+, \tag{3.6}$$

where $\boldsymbol{d}_k \triangleq \boldsymbol{A}\nabla f(\boldsymbol{x}_k)$ and $\boldsymbol{D}_1 \triangleq \mathsf{diag}\{\bar{I}.(\boldsymbol{A}\boldsymbol{x}_k, \boldsymbol{b})\}$. (Careful implementation avoids redundant matrix-vector products.)

This approach removes all tuning parameters other than number of iterations. In addition, using the observed Fisher information leads to a larger step size than using the best Lipschitz constant of (2.12), *i.e.*, $\max_i\{2 + y_i/(4b_i)\}$ when $b_i > 0$, hence accelerating convergence. To facilitate fair comparisons in subsequent sections, we also derive a Fisher information step size for the Gaussian noise model here. The marginal Fisher information for the scalar case of the Gaussian cost function (2.8) is

$$\bar{I}(v, b) = \mathbb{E}_y\left[\left|4|v|(|v|^2 - b - y)\right|^2 \middle| v, b\right] = 16|v|^2(|v|^2 + b). \tag{3.7}$$

one can also derive a convenient step size μ_k for the WF algorithm for the Gaussian model (2.8) using its observed Fisher information to approximate the exact Hessian.

WF with regularization. To potentially improve the reconstruction quality, one often adds a regularizer or penalty to the Poisson log-likelihood cost function, leading to a cost function of the form

$$\Psi(\boldsymbol{x}) = f(\boldsymbol{x}) + \beta \, \mathsf{R}(\boldsymbol{x}) \,. \tag{3.8}$$

The general methods in this work are adaptable to many regularizers, but for simplicity we focus on regularizers that are based on the assumption that $\boldsymbol{T}\boldsymbol{x}$ is approximately sparse, for a $K \times N$ matrix $\boldsymbol{T}$. In particular, we used the corner-rounded anisotropic TV matrix for regularization. Because the WF algorithm requires a well-defined gradient, we replaced the ℓ_1 norm term with a Huber function regularizer of the form

$$\begin{aligned} R(\boldsymbol{x}) =& \boldsymbol{1}' h.(\boldsymbol{T}\boldsymbol{x}; \alpha) = \min_{\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{T}\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \alpha\|\boldsymbol{z}\|_1, \\ h(t; \alpha) \triangleq& \begin{cases} \frac{1}{2}|t|^2, & |t| < \alpha, \\ \alpha|t| - \frac{1}{2}\alpha^2, & \text{otherwise}, \end{cases} \end{aligned} \tag{3.9}$$

which involves solving for $\boldsymbol{z}$ analytically in terms of $\boldsymbol{x}$. This smooth regularizer is suitable for gradient-based methods like WF and for quasi-newton methods like LBFGS, as well as for versions of MM and ADMM. We refer to (3.9) as "TV regularization" even though it is technically (anisotropic) "corner rounded" TV.

For the smooth regularizer (3.9), we majorize the Huber function $h(t)$ using quadratic polynomials with the optimal curvature using the ratio $\dot{h}(z)/z$ [90, p. 184], so that the step size μ_k becomes

$$\begin{aligned}
\mu_k &\triangleq \frac{\|\nabla \tilde{f}(\boldsymbol{x}_k)\|_2^2}{\nabla \tilde{f}(\boldsymbol{x}_k)' \left(\boldsymbol{A}'\boldsymbol{D}_1\boldsymbol{A} + \beta \boldsymbol{T}'\boldsymbol{D}_2\boldsymbol{T}\right) \nabla \tilde{f}(\boldsymbol{x}_k)}, \\
\nabla \tilde{f}(\boldsymbol{x}_k) &\triangleq \nabla f(\boldsymbol{x}_k) + \beta \boldsymbol{T}' \dot{h}.(\boldsymbol{T}\boldsymbol{x}; \alpha), \\
\boldsymbol{D}_2 &\triangleq \mathsf{diag}\{\min.(\alpha \oslash |\boldsymbol{T}\boldsymbol{x}_k|, 1)\},
\end{aligned} \tag{3.10}$$

where $\oslash$ denotes element-wise division.

Algorithm 2 summarizes the (regularized) WF algorithm for the Poisson model that uses the observed Fisher information for the step size and the optional gradient truncation for noise reduction.

Algorithm 2: Regularized WF algorithm for the Poisson model

Input: $\boldsymbol{A}, \boldsymbol{y}, \boldsymbol{b}, \boldsymbol{x}_0, \boldsymbol{T}, \beta$ and n (number of iterations)
for $k = 0, ..., n-1$ **do**
  $\nabla \tilde{f}(\boldsymbol{x}_k) = \boldsymbol{A}' \dot{\boldsymbol{\psi}}.(\boldsymbol{A}\boldsymbol{x}_k; \boldsymbol{y}, \boldsymbol{b}) + \beta \boldsymbol{T}' \dot{h}.(\boldsymbol{T}\boldsymbol{x}_k)$
  if cost function is regularized **then**
    $\mu_k \leftarrow$ Computed by (3.10)
  else
    $\mu_k \leftarrow$ Computed by (3.3)
  end
  $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \mu_k \nabla \tilde{f}(\boldsymbol{x}_k)$
end
Output: $\boldsymbol{x}_n$

3.1.2.2 Majorize minimize (MM)

This section introduces our proposed MM algorithm with a quadratic majorizer using a novel curvature formula for the Poisson phase retrieval problem.

An MM algorithm [93] is a generalization of the expectation-maximization (EM) algorithm that solves an optimization problem by iteratively constructing and solving simpler surrogate optimization problems. Quadratic majorizers are very common in MM algorithms because they have closed-form solutions and are well-suited to conjugate gradient methods.

The bounded curvature property derived in (3.1) enables us to derive an MM algorithm [17] with a quadratic majorizer for (2.12), as illustrated in Fig. 3.1 for the real case.

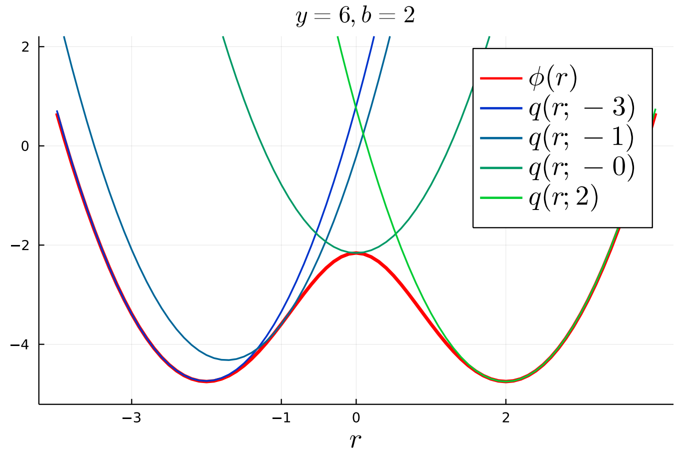


FIG 3.1 – Quadratic majorizers for the non-convex Poisson log-likelihood function $\phi(r; y, b)$ when $y = 6$ and $b = 2$.

With a bit more work to generalize to $\mathbb{C}^N$, a quadratic majorizer for (2.12) has the form

$$q(\boldsymbol{x}; \boldsymbol{x}_k) \triangleq f(\boldsymbol{x}_k) + \text{real}\Big\{(\boldsymbol{x} - \boldsymbol{x}_k)'\boldsymbol{A}'\dot{\boldsymbol{\psi}}.(\boldsymbol{A}\boldsymbol{x}_k; \boldsymbol{y}, \boldsymbol{b})\Big\} + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)'\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_k), \tag{3.11}$$

where $\boldsymbol{W}$ denotes a diagonal curvature matrix. From (3.1), one choice of $\boldsymbol{W}$ uses the maximum of $\ddot{\psi}$:

$$\boldsymbol{W}_{\max} \triangleq \mathsf{diag}\{2 + \boldsymbol{y}/(4\boldsymbol{b})\} \in \mathbb{R}^{M\times M}. \tag{3.12}$$

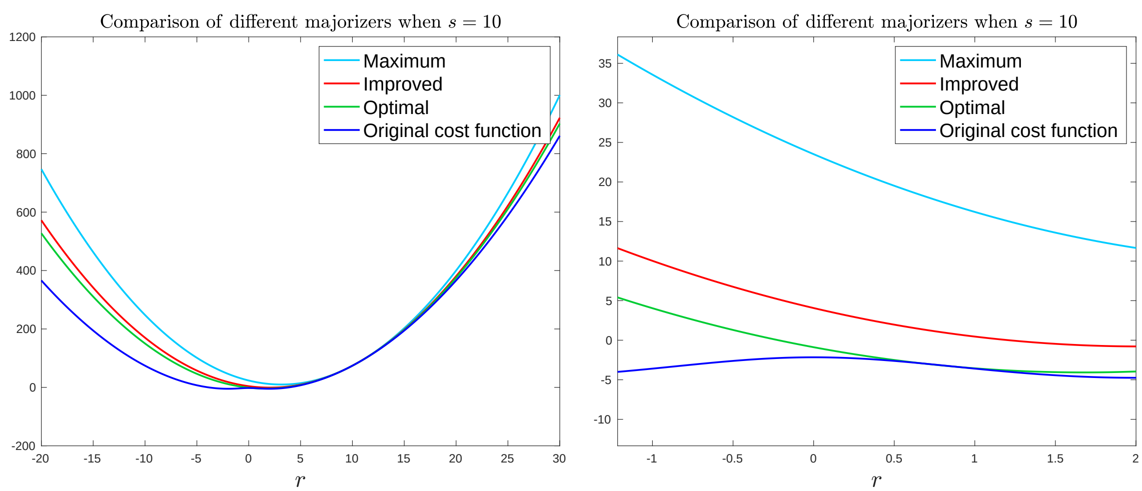
However, $\boldsymbol{W}_{\max}$ is suboptimal because the curvature of a quadratic majorizer of $\psi(v; \cdot)$ varies with $v = [\boldsymbol{A}\boldsymbol{x}_k]_i$. For example, when $|v| \to \infty$, then (2.12) is dominated by the quadratic term having curvature = 2; so if y is large and b is small, then $\boldsymbol{W}_{\max}$ can be much greater than the optimal curvature 2. Thus, instead of using $\boldsymbol{W}_{\max}$ to build majorizers, we propose to use the following improved curvature:

$$\begin{aligned} \boldsymbol{W}_{\text{imp}} &\triangleq \mathsf{diag}\{\boldsymbol{c}.(\boldsymbol{A}\boldsymbol{x}_k; \boldsymbol{y}, \boldsymbol{b})\} \in \mathbb{R}^{M\times M}, \\ c(s; y, b) &\triangleq \begin{cases} \ddot{\psi}\left(\frac{b+\sqrt{b^2+b|s|^2}}{|s|}; y, b\right), & s \neq 0, \\ 2, & s = 0. \end{cases} \end{aligned} \tag{3.13}$$

One can verify $\lim_{s\to 0} c(s; y, b) = 2$ so (3.13) is continuous over $s \in \mathbb{C}$. The Appendix A proves that (3.13) provides a majorizer in (3.11) and is an improved curvature compared to $\boldsymbol{W}_{\max}$, though it is not necessarily the sharpest possible [51]; the sharpest (optimal) curvature $c_{\text{opt}}(s)$ in real case can be expressed as

$$c_{\text{opt}}(s) = \sup_{r \neq s} \frac{2\left(\phi(r) - \phi(s) - \dot{\phi}(s)(r-s)\right)}{(r-s)^2}, \tag{3.14}$$

where $\phi(\cdot)$ is the marginal Poisson cost function defined in (2.13). However, (3.14) usually does not have a closed-form solution due to its transcendental derivative; while our $\boldsymbol{W}_{\text{imp}}$ has a simpler form and is more efficient to compute. Fig. 3.2 visualizes the quadratic majorizer with different curvatures and the original Poisson cost function (2.12). We find the optimal curvature numerically by first discretizing r and then finding the supremum over all discrete segments.



(a) (b)

FIG 3.2 – Comparison of quadratic majorizers with maximum, improved and the optimal curvatures, for $y = 6$ and $b = 2$, visualized around $r = 0.5$. All three curves touch at the point $r = s = 10$ by construction.

If any constraint or regularizer is absent, the quadratic majorizer (3.11) associated with (3.12) or (3.13) leads to the following MM update:

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x} \in \mathbb{F}^N} q(\boldsymbol{x}; \boldsymbol{x}_k) = \boldsymbol{x}_k - (\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}'\dot{\boldsymbol{\psi}}.(\boldsymbol{A}\boldsymbol{x}_k; \boldsymbol{y}, \boldsymbol{b}). \tag{3.15}$$

When N is large, the matrix inverse operation in (3.15) is impractical, so we run a few inner iterations of CG to descend the quadratic majorizer and hence descend the original cost function.

3.1.2.3 Regularized MM

For the regularized cost function (3.8), one can use the quadratic majorizer (3.11) as a starting point. If the regularizer is prox-friendly, then the minimization step of an MM algorithm for the regularized optimization problem is

$$\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x} \in \mathbb{F}^N}{\arg\min}\, q(\boldsymbol{x}; \boldsymbol{x}_k) + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1. \tag{3.16}$$

can be solved by proximal gradient methods [49, 15, 112]. To solve (3.16), we can use the POGM with adaptive restart [112] that provides faster worst-case convergence bound than the FISTA [15].

For non-proximal friendly regularizers, we can "smooth" it using the Huber function (3.9), leading to the optimization problem of the form

$$\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x} \in \mathbb{F}^N}{\arg\min}\, q(\boldsymbol{x}; \boldsymbol{x}_k) + \beta 1' h.(\boldsymbol{T}\boldsymbol{x}; \alpha), \tag{3.17}$$

and we use nonlinear CG for this minimization, with step sizes based on Huber's quadratic majorizer.

Algorithm 3 summarizes our MM algorithm with quadratic majorizer using the improved curvature (3.13).

Algorithm 3: MM algorithm for the Poisson model

```
Input: A, y, b, x_0 and n (number of iterations)
for k = 0, ..., n − 1 do
    Build q(x; x_k) (3.11) using W_imp (3.13)
    if cost function is regularized then
        if T is prox-friendly then
            Update x_k by (3.16) using POGM
        else
            Update x_k by (3.17) using CG
        end
    else
        Update x_k by (3.15) or CG
    end
end
Output: x_n
```

3.1.3 Implementation Details

This section introduces the implementation details of algorithms discussed in the previous section and our experimental setup for the numerical simulation (Section 3.1.4). We ran all algorithms on a server with Ubuntu 16.04 LTS operating system having Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20GHz and 187 GB memory. All elements in the measurement vector $\boldsymbol{y}$ were simulated to follow independent Poisson distributions per (2.11). All algorithms were implemented in Julia v1.7.3. All the timing results presented in Section 3.1.4 were averaged across 10 independent test runs.

3.1.3.1 Initialization

Luo *et al.* [148] proposed the optimal initialization strategy under random Gaussian system matrix setting with Poisson noise. Since this work focuses on low-count regimes, the scale factor κ in [148] is a very small number so that $y - \kappa \approx y$. Therefore, we used $\tilde{\boldsymbol{x}}_0$, the leading eigenvector of $\boldsymbol{A}' \,\mathsf{diag}\{\boldsymbol{y} \oslash (\boldsymbol{y} + 1)\}\boldsymbol{A}$ (instead of $\boldsymbol{A}' \,\mathsf{diag}\{(\boldsymbol{y} - \kappa) \oslash (\boldsymbol{y} + 1)\}\boldsymbol{A}$) as an initial estimate of $\boldsymbol{x}$.

To accommodate signals of arbitrary scale, we scaled that leading eigenvector using a nonlinear least-square fit:

$$\hat{\alpha} = \operatorname*{arg\,min}_{\alpha \in \mathbb{R}} \| \boldsymbol{y} - \boldsymbol{b} - |\alpha \boldsymbol{A} \tilde{\boldsymbol{x}}_0|^2 \|_2^2 = \frac{\sqrt{(\boldsymbol{y} - \boldsymbol{b})' |\boldsymbol{A} \tilde{\boldsymbol{x}}_0|^2}}{\|\boldsymbol{A} \tilde{\boldsymbol{x}}_0\|_4^2}. \tag{3.18}$$

Finally, our initial estimate is the element-wise absolute value of $\hat{\alpha}\boldsymbol{x}_0$ if $\boldsymbol{x}$ is known to be real and nonnegative; and is $\hat{\alpha}\boldsymbol{x}_0$ otherwise.

3.1.3.2 Ambiguities

To handle the global phase ambiguity, *i.e.*, all the algorithms can recover the signal only to within a constant phase shift due to the loss of global phase information, before quantitative comparisons, we corrected the phase of $\hat{\boldsymbol{x}}$ by

$$\hat{\boldsymbol{x}}_{\text{corrected}} \triangleq \operatorname{sign}\left(\langle \hat{\boldsymbol{x}}, \boldsymbol{x} \rangle\right) \hat{\boldsymbol{x}}. \tag{3.19}$$

3.1.3.3 System matrix and True signals

System matrix. We investigated 4 different choices for the system matrix $\boldsymbol{A}$: complex random Gaussian matrix (having 80000 rows), canonical DFT (with reference image), masked

DFT matrix (with 20 masks) and a transmission matrix (ETM) that is acquired empirically through physical experiments [28, 159].

For the canonical DFT, we used a reference image as used in holographic CDI [9], specifically, the measurements follow

$$\boldsymbol{y} \sim \text{Poisson}(|\mathcal{F}\{[\boldsymbol{x}, \boldsymbol{0}, \mathcal{R}]\}|^2 + \boldsymbol{b}), \tag{3.20}$$

where $\mathcal{F}$ denotes DFT and $\mathcal{R}$ denotes a known reference image. This work uses the reference image shown in Fig. 3.3, taken by screen shot from [9].

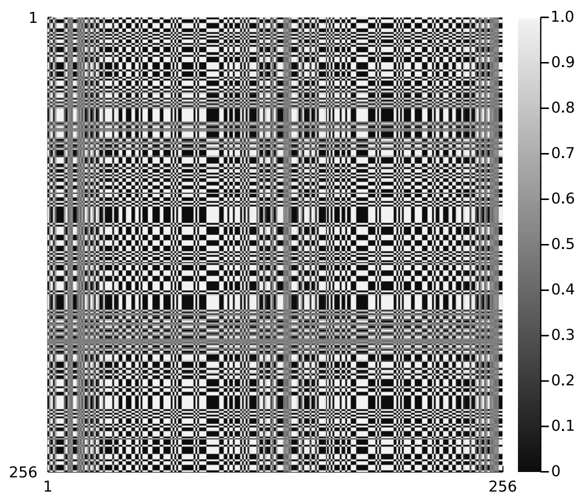


FIG 3.3 – Reference image from [9] used in holographic CDI and our canonical DFT experiments.

For the masked DFT case, the measurement vector $\boldsymbol{y}$ in the Fourier phase retrieval problem has elements with means given by

$$\mathbb{E}[y[\tilde{n}]] = \left| \sum_{n=0}^{N-1} x[n] e^{-\imath 2\pi n \tilde{n} / \tilde{N}} \right|^2 + b[\tilde{n}], \tag{3.21}$$

where $\tilde{N} = 2N - 1$ (here we consider the over-sampled case), and $\tilde{n} = 0, ..., \tilde{N} - 1$. After introducing redundant masks, the measurement model becomes

$$\mathbb{E}[y_l[\tilde{n}]] = \left| \sum_{n=0}^{N-1} x[n] D_l[n] e^{-\imath 2\pi n \tilde{n} / \tilde{N}} \right|^2 + b_l[\tilde{n}], \tag{3.22}$$

where $\mathbb{E}[\boldsymbol{y}_l] \in \mathbb{R}^{\tilde{N}}$ for $i = 1, \ldots, L$ and D_l denotes the lth of L masks. Our experiment used $L = 21$ masks to define the overall system matrix $\boldsymbol{A} \in \mathbb{C}^{L\tilde{N} \times N}$, where the first mask has full sampling and the remaining 20 have sampling rate 0.5 with random sampling patterns.

We scaled each system matrix by a constant factor such that the average count of measurement vector $\boldsymbol{y}$ is 0.25, and the background counts $\boldsymbol{b}$ are set to be all 0.1.

True images. We considered 4 images as the true images in our experiments Fig. 3.4 shows such images; (b) is from [152], (c) is from [9], (d)-(f) are from [159]. We used subfigure (a) for experiments with random Gaussian system matrix, (b) for masked DFT matrix, (c) for canonical DFT matrix and (d) for empirical transmission matrix, respectively.

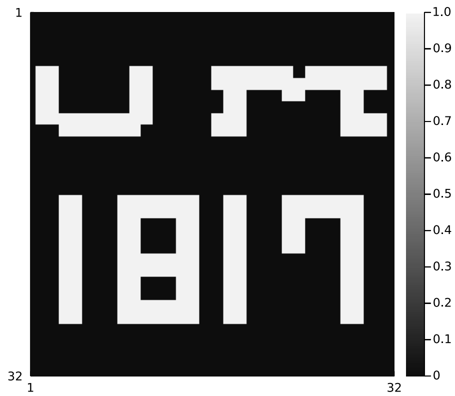
3.1.4 Numerical Simulation Results

3.1.4.1 Convergence speed of WF with Fisher information

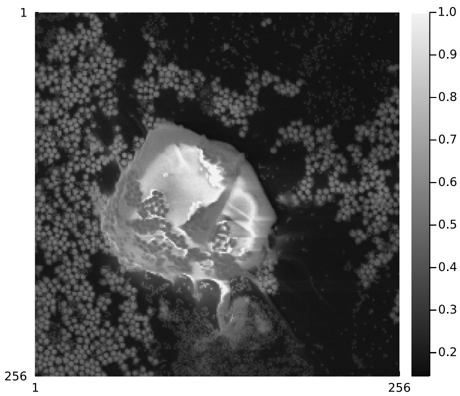
This section compares convergence speeds, in terms of cost function vs. time and PSNR vs. time, between WF using our proposed Fisher information for step size, and empirical step size [24], backtracking line search [177], the optimal step size for the Gaussian noise model [102], and LBFGS quasi-Newton to approximate the Hessian in Newton's method [99]. The LBFGS algorithm was from the "Optim.jl" Julia package [163].

Fig. 3.5 shows that, for all system matrix choices, WF with Fisher information converged faster (in terms of decreasing the cost function) than all other methods; the LBFGS algorithm had comparable convergence speed as WF with backtracking line search. We found that WF with the empirical step size did not converge using hyper-parameters in [24] so we excluded those results in Fig. 3.5. The backtracking approach, although slower than Fisher approach per wall-time, is faster per-iteration. However, the step size found by backtracking line search could be sensitive to hyper-parameter choices. For the WF algorithm with optimal step size (derived based on Gaussian noise model [102]), we conjectured that it reached a non-stationary point that has larger cost function value than those of other methods, as expected.

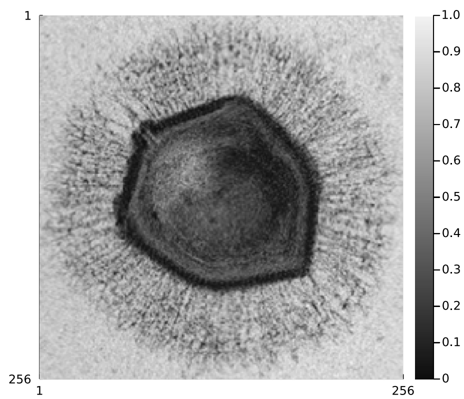
In terms of PSNR, we found that in random Gaussian, masked DFT and empirical transmission cases, WF with Fisher information increased the PSNR faster than all other methods; WF with optimal Gaussian step size led to lower PSNR, perhaps again due to reaching a sub-optimal minimizer. However, for the canonical Fourier case, we found that all methods started decreasing PSNR after several iterations. The algorithms may be more sensitive to noise in the canonical Fourier matrix setting, especially in the low-count regime considered here. Apparently WF with optimal Gaussian step size overfits the noise more slowly due to its sub-optimal step size under Poisson noise.



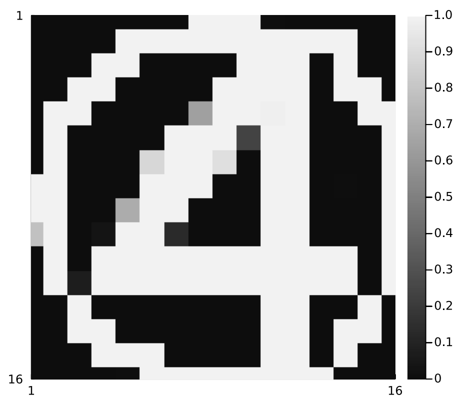
(a)



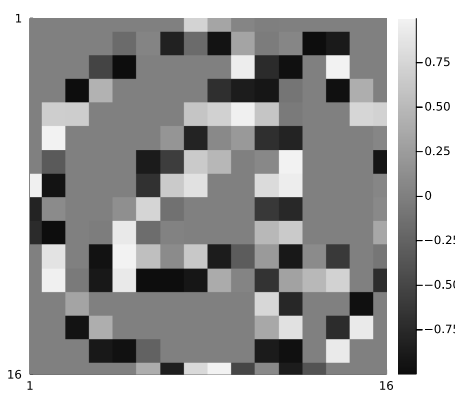
(b) [152]



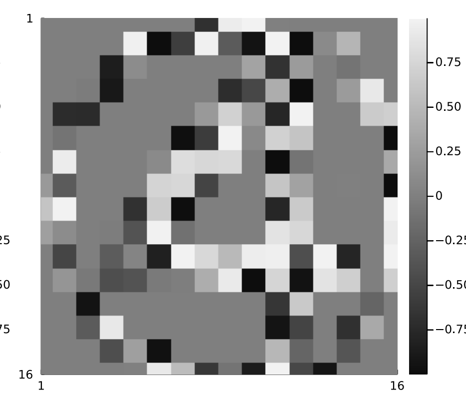
(c) [9]



(d) [159]

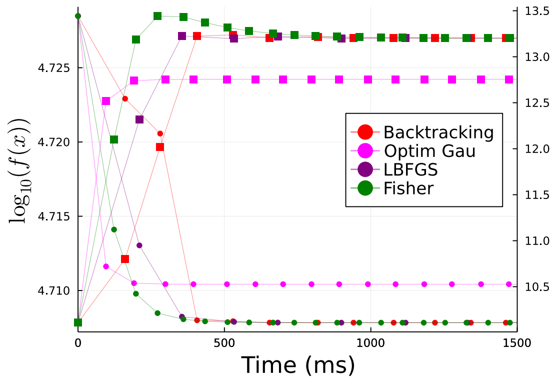


(e) Real part of (d)

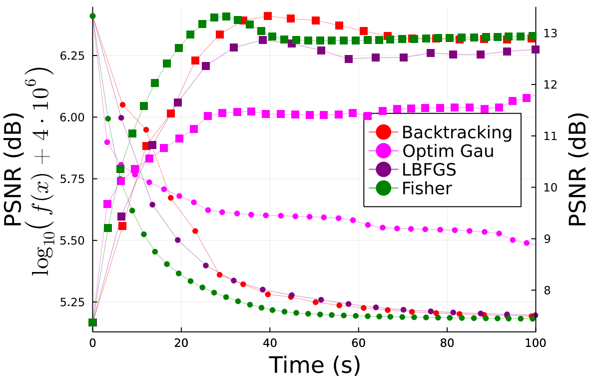


(f) Imaginary part of (d)

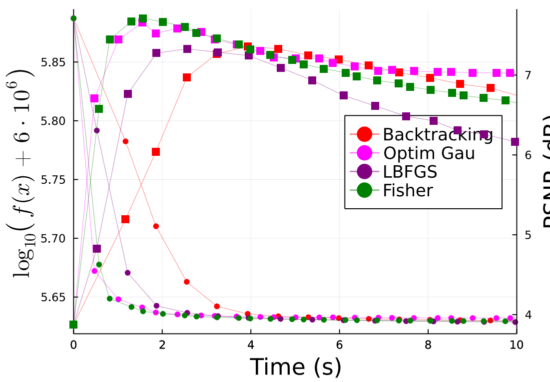
FIG 3.4 – True images used in the simulations. Subfigure (d) shows the magnitude of a complex image.



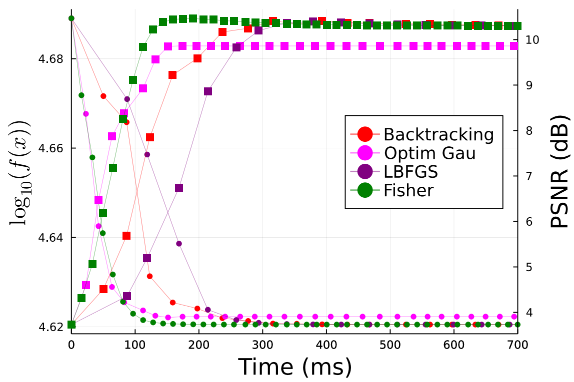
(a) Random Gaussian system matrix.



(b) Masked DFT system matrix.



(c) Canonical DFT system matrix.



(d) Empirical transmission matrix.

FIG 3.5 – Comparison of convergence speed for various WF methods and LBFGS under different system matrix settings. The "Optim Gau" curve is WF using the curvature from [102] that is optimal for Gaussian noise. The circle marker corresponds to the cost function and the square marker corresponds to PSNR.

3.1.4.2 Comparison of Poisson and Gaussian algorithms

This section compares the reconstruction quality, *i.e.*, the NRMSE to the true signal, between WF derived from the Gaussian noise (2.8), and WF derived from the Poisson noise model (2.12) as well as regularized WF under different system matrix settings. We used corner-rounded TV regularizer with $\beta = 32$ and $\alpha = 0.1$ in the regularized WF algorithm.

Fig. 3.6 shows that algorithms derived from the Poisson model yielded consistently better reconstruction quality (lower NRMSE) than algorithms derived from the Gaussian model, as expected. Furthermore, by incorporating regularizer that exploits the assumed property of the true signal, the NRMSE was further decreased. Spectral initialization worked well in random Gaussian matrix setting, but not for other system matrices, as expected from its theory. The WF Gaussian approach failed to reconstruct in masked and canonical

DFT system matrix setting. Since incorporating appropriate regularizers helps algorithms yield higher quality reconstructions, a question is naturally raised about which regularized algorithm converges the fastest. The next subsection presents such comparisons.

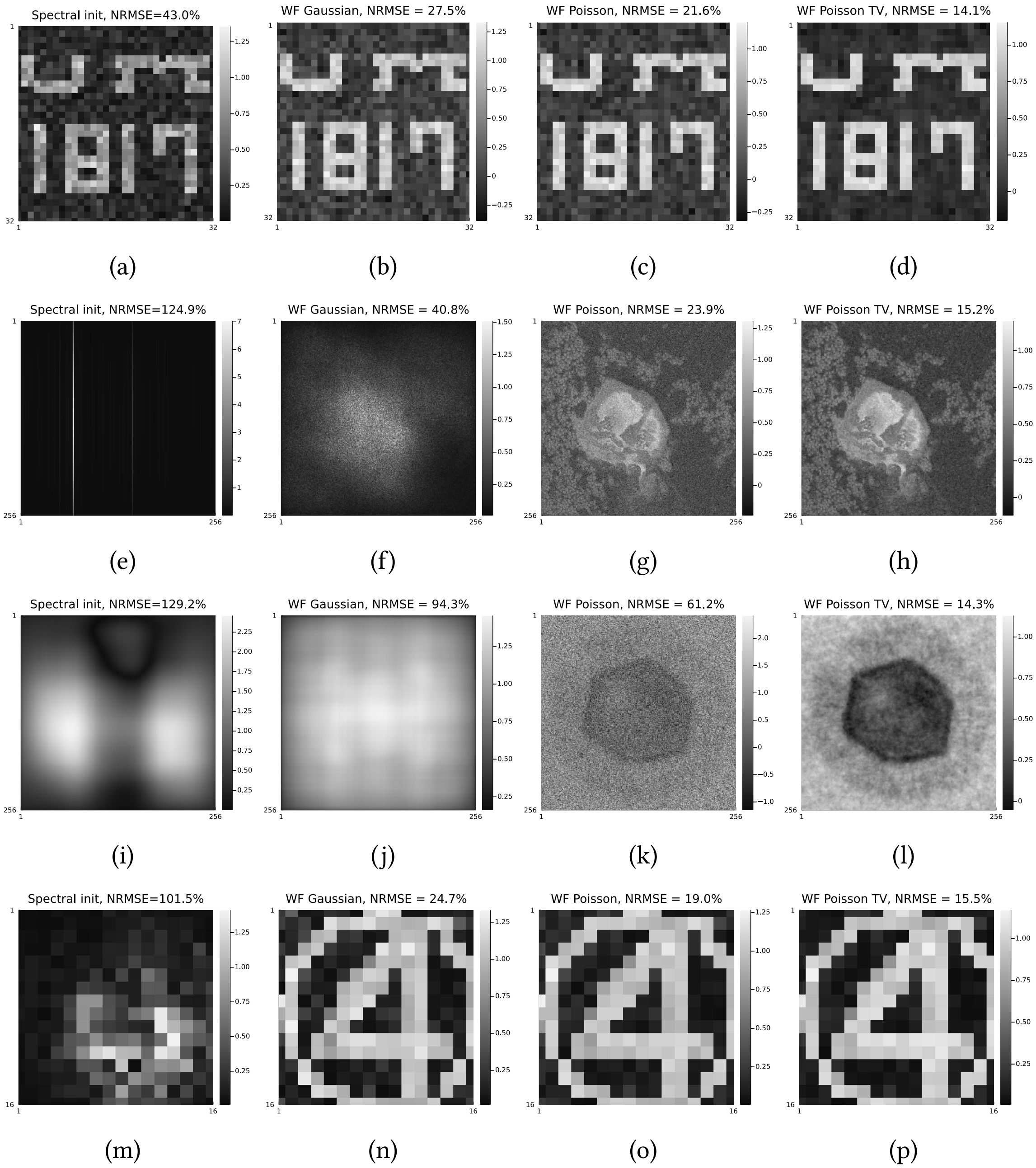


FIG 3.6 – Reconstruction quality comparison between four methods (left to right): the optimal Poisson spectral initialization [148], the WF Gaussian method, the WF Poisson method, and WF Poisson with TV regularization. System matrices: (a)-(d) random Gaussian; (e)-(h) masked DFT; (i)-(l) canonical DFT with reference image; (m)-(p) ETM. Magnitude of complex images shown. All WF algorithms used the proposed Fisher information for step size.

3.1.4.3 Convergence speed of regularized Poisson algorithms

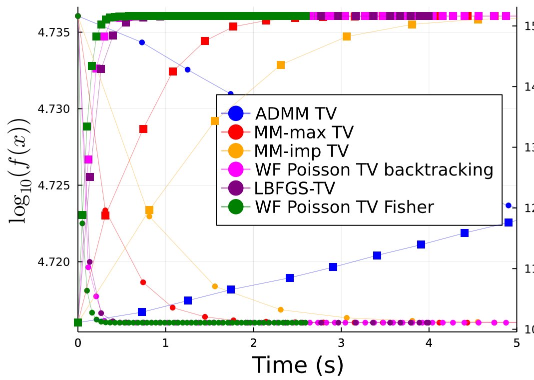
As discussed in Section 3.1.2, many algorithms can be modified to accommodate regularizers. We compared the convergence speeds of regularized Poisson algorithms (WF Fisher, WF backtracking, LBFGS, MM and ADMM), with a smooth regularizer (corner-rounded TV), under different system matrix settings. Based on Fig. 3.5, we did not run simulations of regularized WF with empirical step size and with Gaussian optimal step size, due to their non-converging trend and sub-optimal solution, respectively. For all other algorithms, we set the regularization parameters to be $\beta = 32$ and $\alpha = 0.1$ (defined in (3.8) and (3.9)).

Fig. 3.7 shows that the regularized WF with our proposed Fisher information for step size converged the fastest compared to other methods under all different system matrices. The LBFGS again had a comparable convergence speed as WF using backtracking line search. The MM algorithm with improved curvature, was slower in wall-time due to extra computation per iteration, but was faster per iteration due to its sharper curvature. In masked and canonical Fourier case, however, MM with improved curvature was faster than the maximum curvature in wall-time comparison, which can be attributed to large magnitude low frequency components in the coefficients of the Fourier transform.

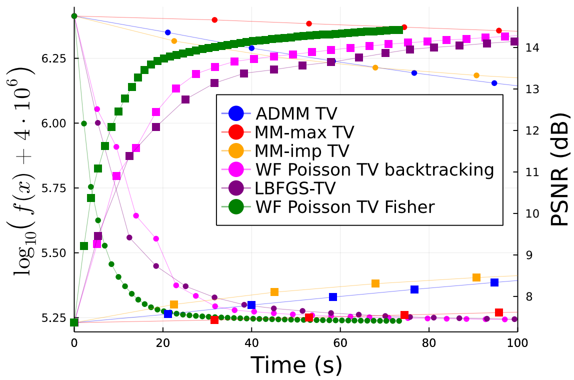
3.1.5 Discussion

Current methods for phase retrieval mostly focus on MLE for Gaussian noise; fewer algorithms were derived for Poisson noise [37, 16, 30]. Here we proposed a novel WF algorithm and an MM algorithm and then did an empirical study on the convergence speed as well as reconstruction quality of several Poisson phase retrieval algorithms. In our proposed WF algorithm, we first replaced the gradient term in Gaussian WF (2.9) with its Poisson counterpart (3.1). Then we did a quadratic approximation of the cost function and applied one iteration of Newton's method to define an "optimal" step size. We then proposed to use the observed Fisher information to approximate the Hessian when computing the step size, which is a common method in computational statistics. Moreover, the Fisher information matrix is guaranteed to be positive semi-definite and is more computationally efficient compared to the Hessian. To further illustrate our proposed method of using Fisher information to approximate the Hessian, Fig. 3.8 visualizes these two matrices (in marginal forms).

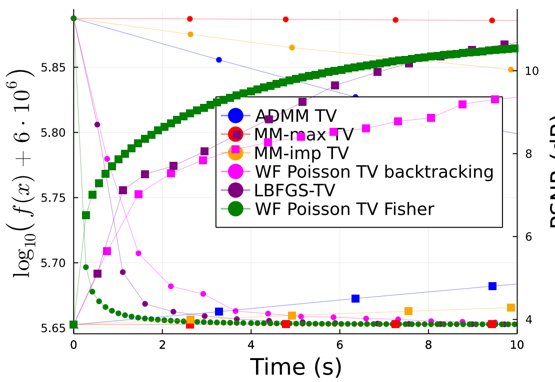
As shown in Fig. 3.8, the Hessian is noisy and can have some negative elements. Such undesirable features can lead to unstable step size calculations. In contrast, the elements in Fisher information matrix are non-negative and less noisy. We ran some experiments and found that when the background counts b_i are large, using the noisy Hessian to calculate



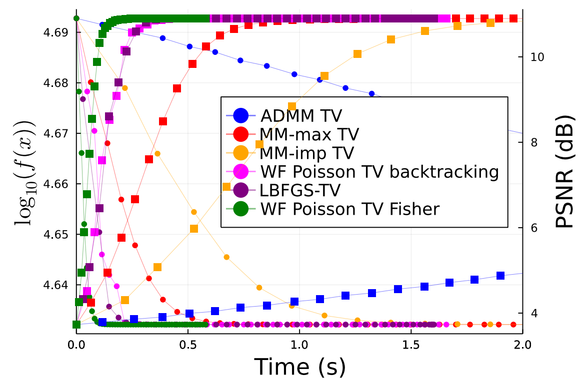
(a) Random Gaussian system matrix.



(b) Masked DFT system matrix.



(c) Canonical DFT system matrix.



(d) Empirical transmission matrix.

FIG 3.7 – Comparison of convergence speed of variant algorithms with corner-rounded TV regularizer. The circle marker corresponds to cost function and the square marker corresponds to PSNR.

the step size can lead to divergence of the cost function, due to the negative values in the marginal second derivative. Setting such negative values in the second derivative to zero is a possible solution, but we found that approach led to slower convergence than using the Fisher information. One potential alternative to our approach is to use the empirical Fisher information, but that may be suboptimal since the empirical Fisher information does not generally capture second-order information [120].

To accommodate our WF algorithm with non-smooth regularizers, *e.g.*, $\|\boldsymbol{T}\boldsymbol{x}\|_1$, we used a Huber function to approximate the ℓ_1 norm with a quadratic function around zero, so that the Wirtinger gradient is well-defined everywhere. A limitation of this work is that we did not consider other regularizers in our experiments, though our algorithms can be generalized to handle other smooth regularizers with minor modifications. One drawback

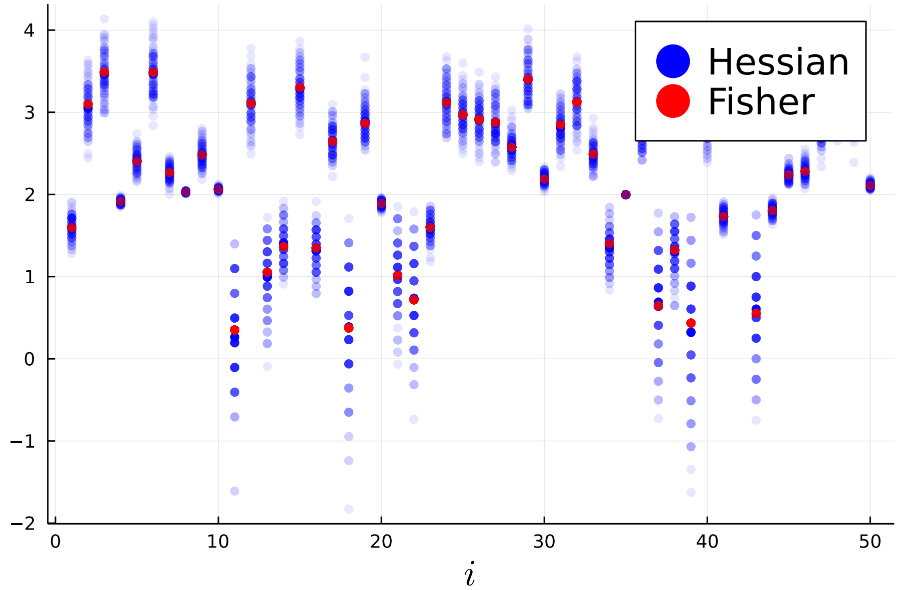


FIG 3.8 – Visualization of the marginal Hessian (3.1) and the marginal observed Fisher information (3.5). The horizontal axis denotes the ith element in the marginal Hessian/Fisher. Data were simulated with a random Gaussian matrix and 100 independent realizations.

of TV regularization is that it assumes piece-wise uniform latent images so it lacks generalizability to other kinds of images, One way to address this is to train deep neural networks [181, 248] with a variety of images, potentially leading to better generalizability.

3.1.6 Conclusion

This work proposed and compared algorithms based on MLE and regularized MLE for phase retrieval from Poisson measurements, in low-photon count regimes. We proposed a novel method that used the Fisher information to compute the step size in the WF algorithm; this approach eliminates all parameter tuning except the number of iterations. We also proposed a novel MM algorithm with improved curvature compared to the one derived from the upper bound of the second derivative of the cost function.

Simulation results experimented on random Gaussian matrix, masked DFT matrix, canonical DFT matrix and an empirical transmission matrix showed that: 1) For unregularized algorithms, the WF algorithm using our proposed Fisher information for step size converged faster than using empirical step size, backtracking line search, optimal step size for Gaussian noise model and LBFGS. Moreover, our proposed Fisher step size can be computed efficiently without any tuning parameter. 2) As expected, algorithms derived from the Poisson noise model produce consistently better reconstruction quality

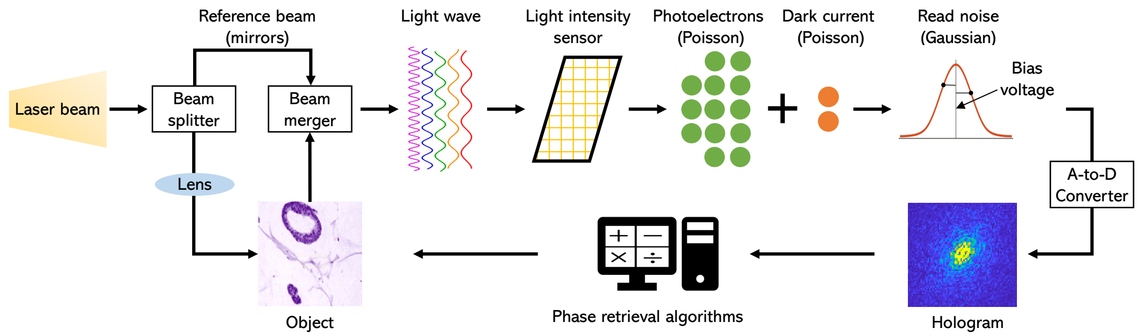


FIG 3.9 – Illustration of Poisson and Gaussian noise statistics in holographic phase retrieval.

than algorithms derived from the Gaussian noise model for low-count data. Furthermore, by incorporating regularizers that exploit the assumed properties of the true signal, the reconstruction quality can be further improved. 3) For regularized algorithms with smooth corner-rounded TV regularizer, WF with Fisher information converges faster than WF with backtracking line search, LBFGS, MM and ADMM.

Future work includes precomputing and tabluting the optimal curvature for the quadratic majorizer, establishing sufficient conditions for global convergence, investigating algorithms with other kind of regularizers (*e.g.*, deep learning methods [181, 248]), investigating sketching methods for large problem sizes [149], and testing Poisson phase retrieval algorithms under a wider variety of experimental settings.

3.2 Poisson-Gaussian Phase Retrieval with Score-based Image Prior[3]

3.2.1 Motivation

In practical scenarios, the measurements $\boldsymbol{y}$ are often contaminated by both PG noise. The Poisson distribution is due to the photon counting and dark current [221]. The Gaussian statistics stem from the readout structures (*e.g.*, analog-to-digital converter (ADC)) of common cameras. Fig. 3.9 illustrates the PG mixed noise statistics in the holographic PR. Because the PG likelihood is complicated, most previous works [25, 24, 194, 102, 23, 203, 177, 23, 70, 168, 220, 68, 99, 141, 242, 207, 26, 230, 210, 80, 236, 9, 216, 126, 135, 72, 139, 63, 37, 16, 249, 30] approximate the Poisson noise statistics by the central limit theorem and

[3]This section is based on [89, 137].

work with a substitute Gaussian log-likelihood estimate problem or use the Poisson maximum likelihood model but simply disregard Gaussian readout noise. Other more complicated approximation methods have also been proposed, such as the shifted Poisson model [199], the unbiased inverse transformation of a generalized Anscombe transform [151, 212], and the majorize-minimize algorithm [62]. However, these approximate methods can lead to a suboptimal solution after optimization that results in a lower-quality reconstruction. Apart from the likelihood modeling, the regularizer $\mathsf{R}(\boldsymbol{x})$ provides prior information about underlying object characteristics that may aid in resolving ill-posed inverse problems. Beyond simple choices of $\mathsf{R}(\boldsymbol{x})$ such as TV or the L1-norm of coefficients of wavelet transform [50], deep learning (DL)-integrated algorithms for solving inverse problems in computational imaging have been reported to be the state-of-the-art [169]. The trained networks can be used as an object prior for regularizing the reconstructed image to remain on a learned manifold [19]. Incorporating a trained denoising network as a regularizer $\mathsf{R}(\cdot)$ led to methods such as plug-and-play (PNP) [27, 251, 106] and regularization by denoising (RED) [182]. In contrast to training a denoiser using clean images, there is growing popularity of self-supervised image denoising approaches that do not require clean data as the training target [130, 13, 223]. In addition to training a denoiser as regularizer, generative model-based priors have also been proposed [5, 225]. Recently, diffusion models have gained significant traction for image generation [204, 88, 57, 206]. These probabilistic image generation models start with a clean image and gradually increase the level of noise added to the image, resulting in white Gaussian noise. Then in the reverse process, a neural network is trained to learn the noise in each step to generate or sample a clean image as in the original data distribution. The score-based diffusion models estimate the gradients of data distribution and can be used as plug-and-play priors for inverse problems [82] such as image deblurring and MRI and CT reconstruction [129, 97, 45, 47, 145, 205]. However, the realm of using score-based models to perform phase retrieval is relatively unexplored; previous relevant works [198, 82] applied DDPM to PR but with less realistic system models and under solely Gaussian or Poisson noise statistics.

In summary, our contribution is three-fold:

- We present a new algorithm known as accelerated WF with a score-based image prior (*i.e.*, $\nabla \mathsf{R}(\boldsymbol{x})$) to address the challenge of holographic PR problem in the presence of PG noise statistics.

- Theoretically, we derive a Lipschitz constant for the holographic PR's PG log-likelihood and subsequently demonstrate the critical points convergence guarantee of our proposed algorithm.

- Simulation experiments demonstrate that: 1) Algorithms with the PG likelihood model yield superior reconstructions in comparison to those relying solely on either the Poisson or Gaussian likelihood models. 2) With the proposed score-based prior as regularization, the proposed approach generates higher quality reconstructions and is more robust to variation of noise levels without any parameter tuning compared to alternative state-of-the-art methods.

3.2.2 Methods

3.2.2.1 Score Function and Diffusion Models

Let $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ denote a model for the prior distribution of the latent image $\boldsymbol{x}$; the score function is then defined as[4] $s_{\boldsymbol{\theta}}(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log p_{\boldsymbol{\theta}}(\boldsymbol{x})$. Consider a sequence of positive noise scales (for white Gaussian $\mathcal{N}(0, \sigma_k^2)$): $\sigma_1 > \sigma_2 > \cdots > \sigma_K$, with σ_K being small enough so that noise of this level does not visibly affect the image, and σ_1 depending on the application. Score matching can be used to train a noise conditional score network [219, 204] as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathbb{E}_{\boldsymbol{x},\tilde{\boldsymbol{x}}} \left[\left(s_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma_k) - \frac{\boldsymbol{x} - \tilde{\boldsymbol{x}}}{\sigma_k^2} \right)^2 \right],$$
$$\text{where} \quad \boldsymbol{x} \sim p(\boldsymbol{x}), \quad \tilde{\boldsymbol{x}} \sim \boldsymbol{x} + \mathcal{N}(0, \sigma_k^2 \boldsymbol{I}). \tag{3.23}$$

With enough data, the neural network $s_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma)$ is expected to learn the distribution $p_{\sigma(\boldsymbol{x})} = \int p(\boldsymbol{x}) p_\sigma(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{x}$ where $p_\sigma(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}, \sigma^2 \boldsymbol{I})$. To sample from the prior, the method of Langevin dynamics is frequently used [204]. To leverage diffusion models for solving inverse problems, previous methods generally recast the reconstruction problem as a conditional generation or sampling problem [198, 82, 206, 44, 205, 43]. This involves relying on the capacity of diffusion models to produce high-quality images while complying with data-fidelity constraints. However, in applications where data collection is costly, *i.e.*, with a limited amount of training data, it is often challenging to train a diffusion model that can generate high-quality images even in an unconditional way. Under these conditions, we found that the score function learned during training diffusion models can serve as an effective image prior, which can capture certain data characteristics when trained for the denoising prediction in the reverse process of the diffusion model. Similar to previous works [82] that uses the score function as a PNP prior, here we also incorporate the score function as a regularization in the optimization objective for solving the PR problem. We

[4] This definition differs from the score function in statistics where the gradient is taken w.r.t. $\boldsymbol{\theta}$ of $\log p_{\boldsymbol{\theta}}(\boldsymbol{x})$.

believe this is a more efficient scheme for incorporating diffusion priors especially for applications with a limited amount of training data, a very common situation in the optical imaging sector.

3.2.2.2 Likelihood Modeling and WF

Based on the physical model as demonstrated in Fig. 3.9, we model the system matrix $\boldsymbol{A}$ by the (oversampled and scaled) discrete Fourier transform applied to a concatenation of the sample $\boldsymbol{x}$, a blank image (representing the holographic separation condition [126]) and a known reference image $\mathcal{R}$, similar to (3.20), the measurements $\boldsymbol{y}$ follow the Poisson plus Gaussian distribution:

$$\boldsymbol{y} \sim \mathcal{N}(\text{Poisson}\left(|\boldsymbol{A}(\boldsymbol{x})|^2 + \bar{\boldsymbol{b}}\right), \sigma^2 \boldsymbol{I}), \quad \boldsymbol{A}(\boldsymbol{x}) \triangleq \alpha \mathcal{F}\{[\boldsymbol{x}, \boldsymbol{0}, \mathcal{R}]\}. \tag{3.24}$$

Here σ^2 denotes the variance of Gaussian noise, and α denotes a scaling factor (quantum efficiency, conversion gain, etc.) after applying the Fourier transform. So that the negative log-likelihood of (3.24) is

$$g_{\text{PG}}(\boldsymbol{x}) = \sum_{i=1}^{M} g_i(\boldsymbol{x}), \quad g_i(\boldsymbol{x}) \triangleq -\log\left(\sum_{n=0}^{\infty} \frac{e^{-\left(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + \bar{b}_i\right)} \cdot \left(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + \bar{b}_i\right)^n}{n!} \cdot \frac{e^{-\left(\frac{(y_i - n)^2}{\sqrt{2}\sigma}\right)}}{\sqrt{2\pi\sigma^2}}\right). \tag{3.25}$$

Here M denotes the length of $\boldsymbol{y}$, $\boldsymbol{a}_i'$ denotes the ith row of $\boldsymbol{A}$ (since $\boldsymbol{A}$ is linear). WF can be used for estimating $\boldsymbol{x}$:

$$\begin{aligned} &\nabla g_{\text{PG}}(\boldsymbol{x}) = 2\boldsymbol{A}' \,\mathsf{diag}\{\phi_i(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + b_i; y_i)\}\boldsymbol{A}\boldsymbol{x}, \\ &\phi(u; v) \triangleq 1 - \frac{s(u, v-1)}{s(u, v)}, \quad s(a, b) \triangleq \sum_{n=0}^{\infty} \frac{a^n}{n!} e^{-\left(\frac{b-n}{\sqrt{2}\sigma}\right)^2}. \end{aligned} \tag{3.26}$$

Lemma. The function $\phi(u)$ is Lipschitz differentiable and the Lipschitz constant for $\dot{\phi}(u)$ is:

$$\max\{|\ddot{\phi}(u)|\} \triangleq \mu = \left(1 - e^{-\frac{1}{\sigma^2}}\right) e^{\frac{2y_{\max}-1}{\sigma^2}}, \text{where} \quad y_{\max} = \max_{i \in \{1,\ldots,M\}} \{y_i\}. \tag{3.27}$$

The proof is given in [40].

Theorem 1 *Assume $|x_j|$ is bounded above by C for each j, a Lipschitz constant of $\nabla g_{\mathrm{PG}}(\boldsymbol{x})$ is*

$$\mathcal{L}(\nabla g_{\mathrm{PG}}) \triangleq 2\|\boldsymbol{A}\|_2^2 \left(2C^2\|\boldsymbol{A}\|_\infty^2\, \tilde{y}_{\max} + \left|1 - C^2\, \|\boldsymbol{A}\|_\infty^2\, \tilde{y}_{\max}\right|\right),$$
$$\tilde{y}_{\max} \triangleq \left(1 - e^{-\frac{1}{\sigma^2}}\right) e^{\frac{2y_{\max}-1}{\sigma^2}}. \tag{3.28}$$

where $y_{\max}$ is $\max_i\{\|y_i|\}, i = 1, \ldots, M$.

Proof: Let $g_{\mathrm{PG}}(\boldsymbol{x})$ denote a function that maps a vector $\boldsymbol{x} \in \mathbb{R}^N$ to a scalar; it is the sum of each $g_i(\boldsymbol{x}) \triangleq \phi_i(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + b_i; y_i)$ over $i = 1, \ldots, M$. Let $\boldsymbol{g}(\boldsymbol{x})$ denote a function that maps a vector $\boldsymbol{x} \in \mathbb{R}^N$ to the measurement space $\boldsymbol{y} \in \mathbb{R}^M$; it is the concatenation of each $g_i(\boldsymbol{x})$. So $\nabla g_{\mathrm{PG}}(\boldsymbol{x}) \in \mathbb{R}^N$, $\nabla^2 g_{\mathrm{PG}}(\boldsymbol{x}) \in \mathbb{R}^{N\times N}$, and $\nabla \boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}^{M\times N}$.

By the chain rule, the Hessian of g_{PG} is

$$\nabla^2 g_{\mathrm{PG}}(\boldsymbol{x}) = 2\boldsymbol{A}' \left(\mathsf{diag}\{\boldsymbol{A}\boldsymbol{x}\}\nabla \boldsymbol{g}(\boldsymbol{x}) + \mathsf{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\boldsymbol{A}\right). \tag{3.29}$$

Assume $|x_j|$ is bounded above by C for each j. Then it follows that $\|\,\mathsf{diag}\{\boldsymbol{A}\boldsymbol{x}\}\|_2 \leq C\|\boldsymbol{A}\|_\infty$ by the construction of matrix-vector multiplication, leading to a Lipschitz constant for $\nabla g_{\mathrm{PG}}(\boldsymbol{x})$:

$$\mathcal{L}(\nabla g_{\mathrm{PG}}) = 2C\|\boldsymbol{A}\|_2\,\|\boldsymbol{A}\|_\infty\,\|\nabla \boldsymbol{g}(\boldsymbol{x})\|_2 + 2\|\boldsymbol{A}\|_2^2\,\|\,\mathsf{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\|_2. \tag{3.30}$$

Here $\mathcal{L}(\nabla g_{\mathrm{PG}})$ denotes a Lipschitz constant for ∇g_{PG}, not necessarily the best one. To compute $\|\nabla \boldsymbol{g}(\boldsymbol{x})\|_2$, we substitute the Lipschitz constant of $\dot{\phi}(u)$ into (3.26) and apply Lemma 3.2.2.2, leading to

$$\|\nabla \boldsymbol{g}(\boldsymbol{x})\|_2 \leq 2C\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_\infty \left(1 - e^{-\frac{1}{\sigma^2}}\right) e^{\frac{2y_{\max}-1}{\sigma^2}}. \tag{3.31}$$

To compute $\|\,\mathsf{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\|_2$, let

$$t \in [b, \max_i\{|\boldsymbol{a}_i'\boldsymbol{x}|^2\} + b] \subseteq \mathcal{T} \triangleq [b, C^2\|\boldsymbol{A}\|_\infty^2 + b]. \tag{3.32}$$

From the fact that $\dot{\phi}(t) \leq 1$ by its construction, one can derive that

$$\|\,\mathsf{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\|_2 = \|\boldsymbol{g}(\boldsymbol{x})\|_\infty \leq \max_{t\in\mathcal{T}}\{|\dot{\phi}(t)|\} \leq \left|1 - C^2\|\boldsymbol{A}\|_\infty^2 \max\{|\ddot{\phi}(t)|\}\right|. \tag{3.33}$$

Combining (3.30), (3.31) and (3.33) completes the proof of Theorem 1.

However, due to the infinite sum in Poisson-Gaussian log-likelihood (3.25), we approximate $s(a, b)$ with a finite sum following [40]:

$$s(a,b) \approx \sum_{n=0}^{n^+} \frac{a^n}{n!} e^{-\left(\frac{b-n}{\sqrt{2}\sigma}\right)^2}, \quad n^+ = \lceil n^* + \delta\sigma \rceil, \tag{3.34}$$

with n^* given by

$$\begin{aligned} n^* &= \sigma \mathcal{W}\left(\frac{a}{\sigma^2} e^{b/\sigma^2}\right) \\ &\approx \sigma \left(\frac{b}{\sigma^2}\log\left(\frac{a}{\sigma^2}\right) - \log\left(\frac{b}{\sigma^2}\log\left(\frac{a}{\sigma^2}\right)\right)\right) \\ &= \frac{b}{\sigma}\log\left(\frac{a}{\sigma^2}\right) - \sigma\log\left(\frac{b}{\sigma^2}\log\left(\frac{a}{\sigma^2}\right)\right), \end{aligned} \tag{3.35}$$

where $\mathcal{W}(\cdot)$ denotes the Lambert function. The accuracy of this approximation is controlled by δ. Reference [40] provides a comprehensive analysis on the maximum error value of the truncated sum (3.34) and found the bound was very precise.

3.2.2.3 Accelerated WF with Score-based Image Prior

Algorithm 4: Our proposed accelerated WF with score-based image prior.

Input: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, momentum factor $\eta_0 = 1$, step size factor β, weighting factor γ, truncation operator $\mathcal{P}_C(\cdot) \rightarrow [0, C]$; initial image $\boldsymbol{x}_0$, initial auxiliary variables $\boldsymbol{z}_0 = w_0 = \boldsymbol{v}_0 = \boldsymbol{x}_0$, initialize $\sigma_1 > \sigma_2 > \cdots > \sigma_K$.

for $k = 1 : K$ **do**
- **for** $t = 1 : T$ **do**
  - Set step size $\mu = \beta\sigma_k^2$.
  - Set $\Delta z_{t,k} = \frac{\eta_{t-1,k}}{\eta_{t,k}}(\boldsymbol{z}_{t,k} - \boldsymbol{x}_{t,k})$.
  - Set $\Delta x_{t,k} = \frac{\eta_{t-1,k}-1}{\eta_{t,k}}(\boldsymbol{x}_{t,k} - \boldsymbol{x}_{t-1,k})$.
  - Set $\boldsymbol{w}_{t,k} = \mathcal{P}_C\left(\boldsymbol{x}_{t,k} + \Delta z_{t,k} + \Delta x_{t,k}\right)$.
  - Compute $s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k)$ and $s_{\boldsymbol{\theta}}(\boldsymbol{w}_{t,k}, \sigma_k)$.
  - Set $\boldsymbol{z}_{t+1,k} = \boldsymbol{w}_{t,k} - \mu\left(\nabla g_{\mathrm{PG}}(\boldsymbol{w}_{t,k}) + s_{\boldsymbol{\theta}}(\boldsymbol{w}_{t,k}, \sigma_k)\right)$.
  - Set $\boldsymbol{v}_{t+1,k} = \boldsymbol{x}_{t,k} - \mu\left(\nabla g_{\mathrm{PG}}(\boldsymbol{x}_{t,k}) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k)\right)$.
  - Set $\eta_{t+1,k} = \frac{1}{2}\left(1 + \sqrt{1 + 4\eta_{t,k}^2}\right)$.
  - Set $\boldsymbol{x}_{t+1,k} = \mathcal{P}_C\left(\gamma_{t,k}\boldsymbol{z}_{t+1,k} + (1-\gamma_{t,k})\boldsymbol{v}_{t+1,k}\right)$.
- **end**

end

Output: Return $\boldsymbol{x}_{T,K}$.

For accelerating the WF algorithm, we followed the implementation of [131] as its convergence guarantee was proved. Assuming that the true score function can be learned properly, when we have a trained score function $s_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma)$ by applying (3.23), the gradient descent algorithm for MAP estimation has the form: $\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \mu(\nabla g(\boldsymbol{x}_t) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \sigma_k))$. Algorithm 4 summarizes our proposed AWFS algorithm. In a similar fashion as Langevin dynamics, we choose σ_k to be a descending scale of noise levels. In practice, we generally use each noise level a fixed number of times, with geometrically spaced noise levels between some lower and upper bound. The step size factor β in Algorithm 4 can be selected empirically, but we show that the Lipschitz constant of the gradient $\nabla g_{\mathrm{PG}}(\boldsymbol{x}_t) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \sigma_k)$ exists as demonstrated in Theorem 2 (the proof is given in the Appendix C).

We assume that the data allows the neural network to learn the score function well, *i.e.*, $s_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma) \approx \nabla \log(p_\sigma(\boldsymbol{x}))$, and $p_\sigma(\boldsymbol{x}) = p(\boldsymbol{x}) \circledast \mathcal{N}(0, \sigma^2)$, where $\circledast$ denotes (circular) convolution. One can show that $\nabla \log(p_\sigma(\boldsymbol{x}))$ is Lipschitz continuous on $[-C, C]^N$. The proof is given in Appendix C. Using $p_\sigma(\boldsymbol{x})$, we define the smoothed posterior as

$$p_\sigma(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r}) \propto p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}, \bar{\boldsymbol{b}}, \boldsymbol{r}) p_\sigma(\boldsymbol{x}). \tag{3.36}$$

Theorem 2 *For a smooth density function $p_{\sigma_k}(\boldsymbol{x})$ that has finite expectation with $\sigma_k > 0$, the Lipschitz constant of $\nabla g_{\mathrm{PG}}(\boldsymbol{x}_{t,k}) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k)$ exists when each element in $\boldsymbol{x}_{t,k}$ satisfies $0 < |x_j| < C$ for each j. Furthermore, if the weighting factor $\gamma \in \{0, 1\}$ is chosen appropriately following [131], i.e., according to the higher posterior probability between $p_{\sigma_k}(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{A}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ and $p_{\sigma_k}(\boldsymbol{v}|\boldsymbol{y}, \boldsymbol{A}, \bar{\boldsymbol{b}}, \boldsymbol{r})$; then with sufficiently small β, the inner iteration sequence $\{\boldsymbol{x}_{t,k}\}$ generated by Algorithm 4 is bounded, and any accumulation point of $\{\boldsymbol{x}_{t,k}\}$ is a critical point of the posterior distribution $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ in* (3.36).

Proof: By Lipschitz continuity of $\log(p_\sigma(\boldsymbol{x}))$, and from the design of Algorithm 4, $\boldsymbol{x}_{t,k}$ and $\boldsymbol{w}_{t,k}$ are both bounded between $[0, C]$ for all t, k, so the Lipschitz constant $\mathcal{L}^*$ of $\nabla g_{\mathrm{PG}}(\cdot) + s_{\boldsymbol{\theta}}(\cdot)$ exists. With the step size μ satisfying $0 < \mu < \frac{1}{\mathcal{L}^*}$, and the weighting factor $\gamma \in \{0, 1\}$ being chosen according to the higher posterior probability between $p_{\sigma_k}(\boldsymbol{z}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ and $p_{\sigma_k}(\boldsymbol{v}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ (see [131]), we satisfy all conditions in Theorem 1 of [131], which establishes the critical-point convergence of the sequence $\boldsymbol{x}_{t,k}$ generated by Algorithm 4 for any $\sigma_k, k = 1, \ldots, K$. Hence the sequence $\boldsymbol{x}_{t,k}$ generated by Algorithm 4 converges as $t \to \infty$ to a critical-point of the posterior $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ for any σ_k.

3.2.3 Experiment

3.2.3.1 Experiment Settings

Dataset. We tested all algorithms on three datasets: 162 histopathology images related to breast cancer [3] (train/val/test is 122/20/20); 920 images from CelebA dataset [143] (train/val/test is 800/100/20); and 720 images from a homemade CT-density dataset (train/val/test is 600/100/20). The CT-density dataset was generated from SPECT/CT images for Yttrium-90 radionuclide therapy after applying the CT-to-density calibration curve [132]. Although the size of training datasets are relatively small compared to typical datasets such as ImageNet or LSUN [88, 206] that have millions of images, we do not require the score functions to learn image priors strong enough to generate realistic images from white Gaussian noise; rather, it is sufficient for the priors to be able to denoise moderately noisy images.

System Model. Similar to (3.20), we define the system matrix to be discrete Fourier transform of the concatenation of the true image $\boldsymbol{x}$, a blank image $\mathbf{0}$ and a reference image $\mathcal{R}$ with scaling and oversampling. We set the scaling factor α to be in the range $[0.02, 0.035]$ so that the average counts per pixel range from 6 to 25; the oversampled ratio is set to 2. We set $\mathcal{R}$ to be a binary random image similar to what was used in [126]. The standard deviation of the Gaussian read noise added to the measurements $\boldsymbol{y}$ was set as $\sigma \in [0.5, 1.5]$.

Implemented Algorithms. For unregularized algorithms, we implemented Gaussian WF, Poisson WF and Poisson-Gaussian WF. For regularized algorithms, we implemented smoothed total variation (TV) based on the Huber function [90, p. 184] and PNP/RED methods with the DL denoiser [252]: PNP-ADMM [218], PNP-PGM [105], and RED-SD [182]. We also implemented the RED-SD algorithm with "Noise2Self" zero-shot image denoising network [13] (RED-SD-SELF). For diffusion models, we implemented DOLPH [198] and our proposed AWFS. The implementation details of each algorithm can be found in the appendix of [137]. We used spectral initialization [148] for the Gaussian PR and Poisson PR methods; we then used the output results from Poisson PR to initialize other algorithms. We ran all algorithms until convergence in normalized root mean squared error (NRMSE) or reached the maximum number of iterations (*e.g.*, 50).

To evaluate the robustness and limitation of these algorithms, we first tuned the parameters for each algorithm at the noise level when $\alpha = 0.030$ and $\sigma = 1$, and then held them fixed throughout all experiments (Table 3.1, Table 3.2, Fig. 3.13 and Fig. 3.14). In practice the ground truths are unknown, so oracle tuning of test datasets is infeasible (though

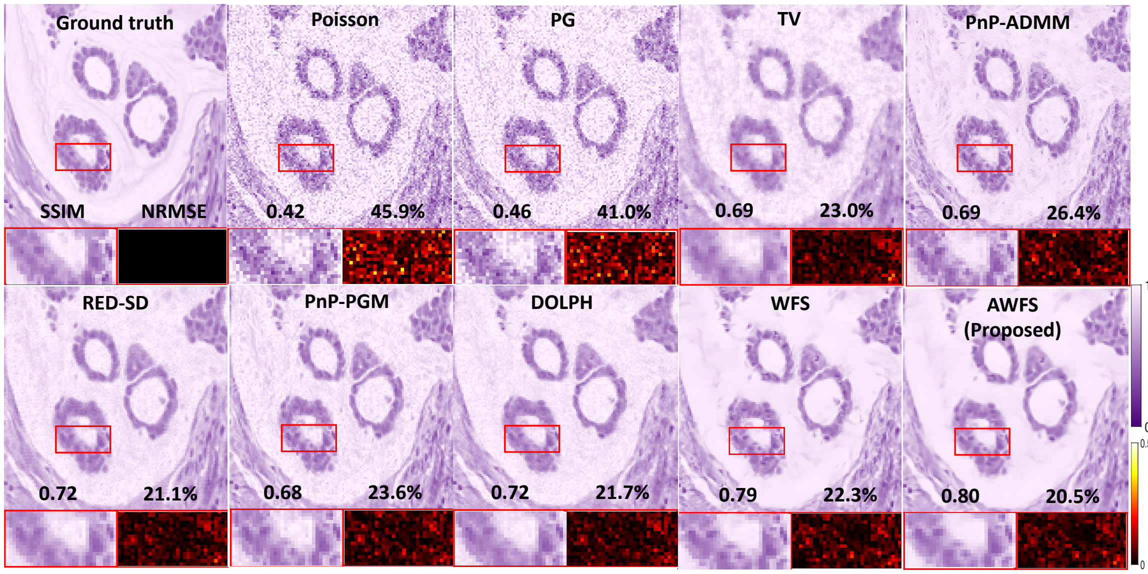


FIG 3.10 – Reconstructed images on dataset [3]. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. α and σ were set to 0.02 and 1, respectively.

some form of cross validation may be possible). Though the numbers reported could fluctuate after careful refinement, *e.g.*, by performing grid search on tuning parameters, such techniques would potentially impede the algorithm's practical use.

Network Training. For PNP denoising networks, we trained all denoisers on different noise levels $\sigma \in \{9, 11, 13, 15\}$ and found that $\sigma = 15$ worked the best on our data. We also used the denoiser scaling technique from [238] to dynamically adjust the performance of all PNP methods. To perform score matching, we applied 20 geometrically spaced noise levels between 0.005 and 0.1 on each of the training images. All networks were implemented in PyTorch and trained on an NVIDIA Quadro RTX 5000 GPU using the ADAM optimizer [115] for 1000 epochs with the best one being selected based off the validation error, *i.e.*, the mean squared error (MSE) loss.

3.2.3.2 Results

We compared all implemented algorithms both qualitatively, by visualizing the reconstructed images and residual errors, and quantitatively, by computing the NRMSE and structural similarity index measure (SSIM). Due to the global phase ambiguity, *i.e.*, all the algorithms can recover the signal only to within a constant phase shift due to the loss of global phase information, we corrected the phase of $\hat{\boldsymbol{x}}$ by $\hat{\boldsymbol{x}}_{\text{corrected}} \triangleq \operatorname{sign}(\langle \hat{\boldsymbol{x}}, \boldsymbol{x}_{\text{true}} \rangle)\, \hat{\boldsymbol{x}}$.

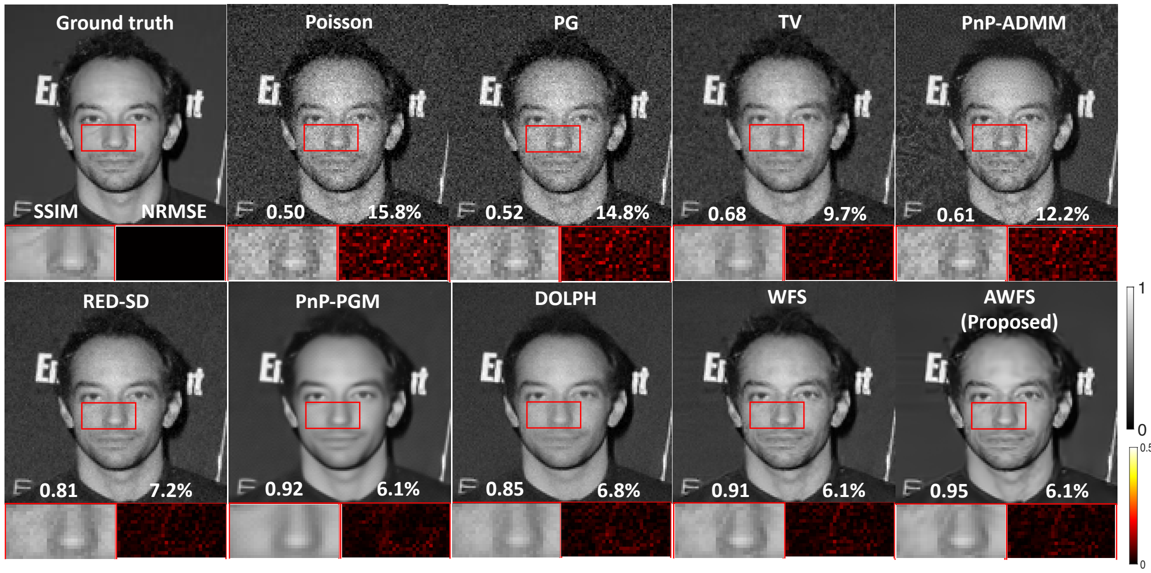


FIG 3.11 – Reconstructed images on celebA dataset [143]. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. α and σ were set to 0.035 and 1, respectively.

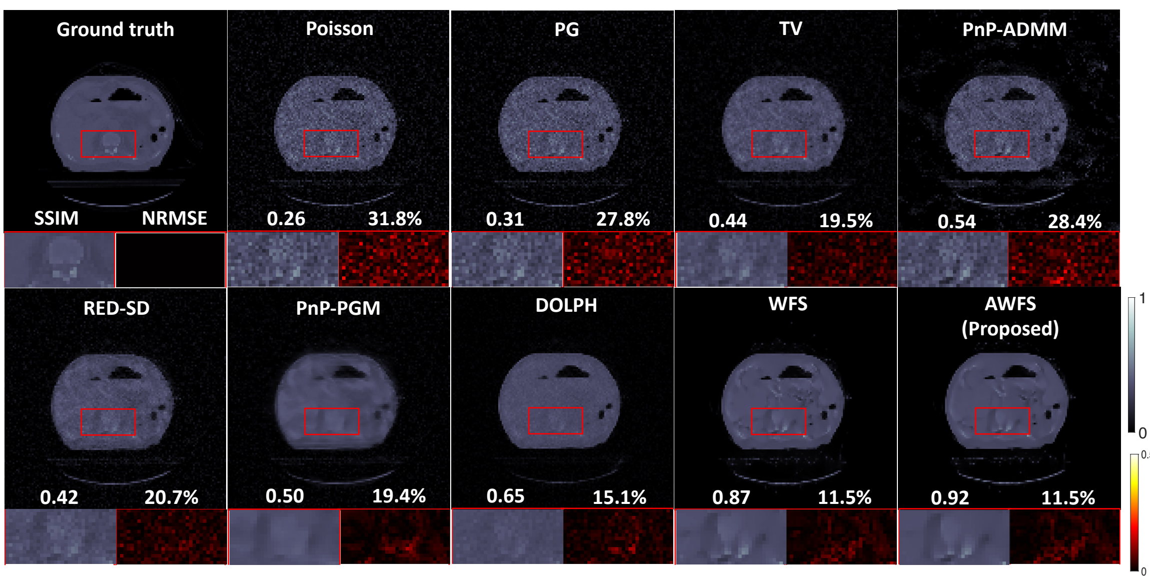
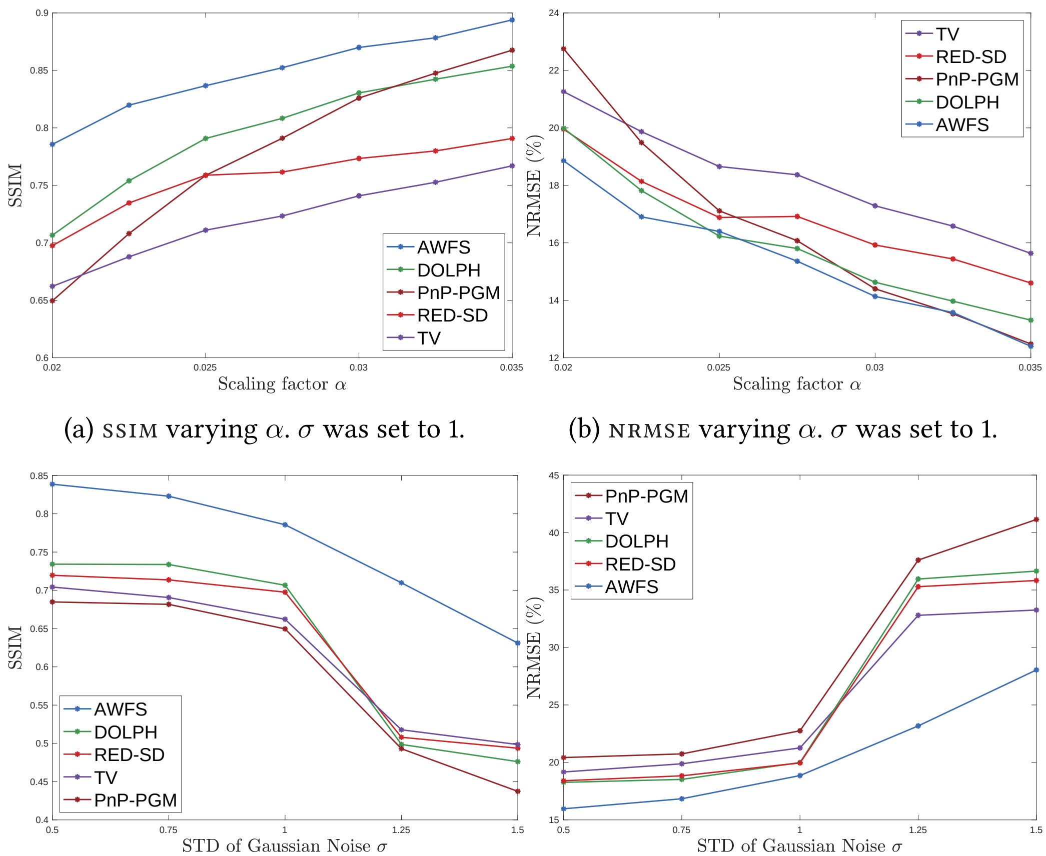


FIG 3.12 – Reconstructed images on CT-density dataset. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. α and σ were set to 0.035 and 1, respectively.

Fig. 3.10, Fig. 3.11 and Fig. 3.12 visualize reconstructed images generated by algorithms mentioned in the previous section. The WF with PG likelihood outperforms WF with Poisson likelihood with a consistently higher SSIM and lower NRMSE. Moreover, we found unregularized Gaussian WF failed to reconstruct images similar to what was reported in

TBL 3.1 – SSIM and NRMSE for Poisson and PG likelihoods. Results were averaged across 7 different noise levels by varying $\alpha \in 0.02 : 0.005 : 0.035$ in (3.24).

Likelihood	Unregularized (SSIM/NRMSE)		DOLPH (SSIM/NRMSE)		AWFS (SSIM/NRMSE)	
DataSet: Histopathology [3]						
Poisson	0.54 ± 0.18	31.7 ± 10.2	0.72 ± 0.13	19.5 ± 6.1	0.83 ± 0.06	16.2 ± 3.7
Poisson-Gaussian	0.57 ± 0.18	28.9 ± 9.0	0.80 ± 0.06	16.0 ± 2.9	$\mathbf{0.85 \pm 0.05}$	$\mathbf{15.4 \pm 3.7}$
DataSet: CelebA [143]						
Poisson	0.39 ± 0.10	24.5 ± 11.4	0.61 ± 0.12	15.6 ± 10.6	0.72 ± 0.16	15.2 ± 11.8
Poisson-Gaussian	0.42 ± 0.10	21.8 ± 9.1	0.71 ± 0.11	$\mathbf{13.7 \pm 11.1}$	$\mathbf{0.74 \pm 0.15}$	14.8 ± 11.9
DataSet: CT-Density						
Poisson	0.19 ± 0.06	48.9 ± 13.1	0.38 ± 0.11	25.6 ± 7.5	0.84 ± 0.08	17.8 ± 4.3
Poisson-Gaussian	0.24 ± 0.06	40.8 ± 9.5	0.55 ± 0.08	20.0 ± 3.3	$\mathbf{0.88 \pm 0.05}$	$\mathbf{16.4 \pm 3.7}$



(a) SSIM varying α. σ was set to 1.

(b) NRMSE varying α. σ was set to 1.

(c) SSIM varying σ; α was set as 0.02.

(d) NRMSE varying σ; α was set as 0.02.

FIG 3.13 – Comparison of SSIM and NRMSE varying scaling factor $\alpha \in [0.02, 0.035]$ and STD of Gaussian noise $\sigma \in [0.25, 1.5]$ defined in (3.24).

[177]. Of the regularized algorithms with PG likelihood, our proposed AWFS had less visual

TBL 3.2 – SSIM and NRMSE using Poisson Gaussian likelihood with different regularization/image prior approaches. Results were averaged across 7 different noise levels by varying $\alpha \in 0.02 : 0.005 : 0.035$ in (3.24). WFS* runs the same number of iterations as AWFS whereas WFS† runs more iterations until convergence.

Dataset	Histopathology [3]		CelebA [143]		CT-Density	
Methods	SSIM	NRMSE (%)	SSIM	NRMSE (%)	SSIM	NRMSE (%)
Unregularized	0.57 ± 0.18	28.9 ± 9.0	0.42 ± 0.10	21.8 ± 9.1	0.24 ± 0.06	40.8 ± 9.5
RED-SD-SELF [13]	0.66 ± 0.13	21.9 ± 4.5	0.60 ± 0.09	15.9 ± 10.6	0.34 ± 0.04	28.1 ± 4.1
PNP-ADMM [218]	0.71 ± 0.11	20.7 ± 4.2	0.56 ± 0.08	16.7 ± 8.1	0.55 ± 0.03	31.2 ± 2.7
TV regularizer	0.72 ± 0.11	18.2 ± 3.9	0.64 ± 0.07	14.4 ± 8.6	0.41 ± 0.03	23.7 ± 2.8
RED-SD [182]	0.76 ± 0.09	16.8 ± 3.6	0.69 ± 0.11	13.9 ± 10.9	0.38 ± 0.04	25.9 ± 4.0
PNP-PGM [105]	0.78 ± 0.11	16.5 ± 4.5	$\mathbf{0.74 \pm 0.14}$	$\mathbf{13.5 \pm 11.3}$	0.42 ± 0.07	24.6 ± 4.4
DOLPH [198]	0.80 ± 0.06	16.0 ± 2.9	0.71 ± 0.11	13.7 ± 11.1	0.55 ± 0.08	20.0 ± 3.3
WFS*	0.76 ± 0.12	18.2 ± 5.5	0.63 ± 0.16	16.9 ± 11.8	0.53 ± 0.17	21.3 ± 7.6
WFS†	0.83 ± 0.06	16.2 ± 4.0	0.70 ± 0.16	15.7 ± 11.8	0.74 ± 0.13	17.3 ± 4.8
AWFS (Proposed)	$\mathbf{0.85 \pm 0.05}$	$\mathbf{15.4 \pm 3.7}$	$\mathbf{0.74 \pm 0.15}$	14.8 ± 11.9	$\mathbf{0.88 \pm 0.05}$	$\mathbf{16.4 \pm 3.7}$

noise and achieved greater detail recovery compared to other methods, as evidenced by the zoomed-in area in these figures.

For quantitative evaluations, Table 3.1 exemplifies the effect of using our proposed PG likelihood as compared to the simpler Poisson likelihoods. We did not run the Gaussian likelihood with DOLPH or AWFS due to the abysmal performance with this likelihood. In all cases, usage of the PG likelihood results in improved image quality in terms of both metrics. Table 3.2 consists of experiments using the PG likelihood and shows the efficacy of the proposed AWFS method over other methods. In particular, our AWFS had superior quantitative performance over all other compared methods on the histopathology and CT-density datasets; in contrast, the PNP-PGM showed the lowest NRMSE on celebA dataset. This is likely due to higher randomness in celebrity faces because the effectiveness of generative models can vary depending on the dataset used. Thus, when provided with a small amount of training data with high randomness, image denoising models (DNCNN) may be more effective than generative models.

We also tested the robustness of the leading algorithms in Table 3.2, by varying both scaling factor α and STD of Gaussian noise σ. Fig. 3.13 and Fig. 3.14 illustate results, where our AWFS algorithm had the highest SSIM and lowest NRMSE. In Fig. 3.14, AWFS demonstrated minimal variations in SSIM and NRMSE metrics than DOLPH as evidenced by the smaller discrepancies in SSIM (0.17 vs. 0.23) and NRMSE (12.6% vs. 18.2%) when σ varies

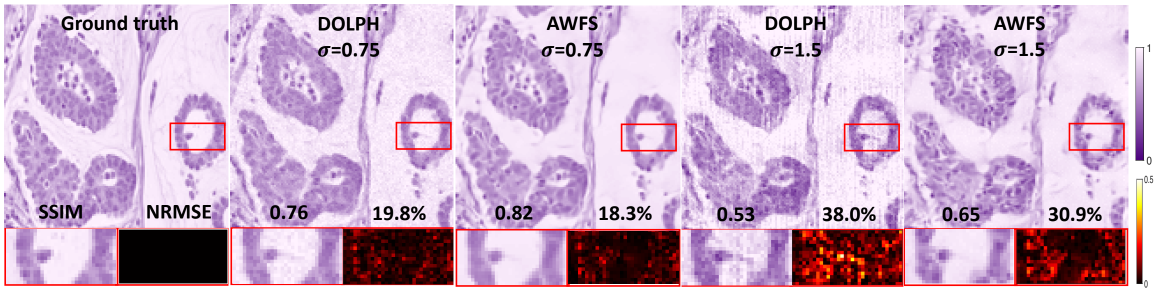


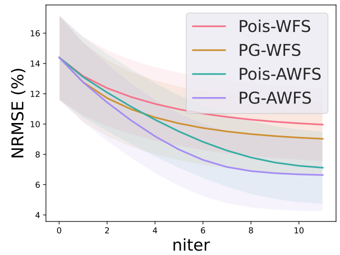
FIG 3.14 – Reconstructed images by DOLPH [198] and our proposed AWFS method under different σ values. Scaling factor α was set to 0.02 (defined in (3.24)).

from 0.75 to 1.5. Fig. 3.15 compares the convergence rate of AWFS vs. WFS for the Poisson and PG likelihood, respectively. Under a variety of noise levels, AWFS consistently converged faster than WFS in terms of number of iterations.

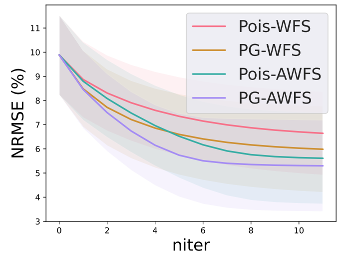
3.2.4 Discussion

PR has a long-standing history in the field of signal processing and imaging. Pioneering works such as the error reduction and hybrid input-output algorithms by Gerchberg Saxton [70] and Fienup [67] have been proposed to address this problem. These iterative algorithms involve constraints imposed on evaluations between the image domain and frequency domain. However, these methods have limitations in terms of the quality of reconstructed images and their convergence remains uncertain [244]. Another approach to solving PR problems is through compressed sensing and optimization techniques like WF [26], matrix lifting [25, 24, 194], MM [177] and ADMM [141]. This work focuses on the WF algorithm due to being straightforward to incorporate with the DL regularizer for the image prior. The likelihood modelling of the noise statistics existing in the measurement is also critical. Previous studies have primarily focused on modelling either Gaussian or Poisson likelihood only, but in practical scenarios, both types of noise are often encountered. Therefore, this work contributes to a more practical perspective of addressing the holographic PR problem by using a PG likelihood and incorporating state-of-the-art deep learning image priors.

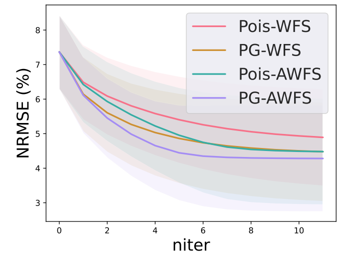
In our evaluation of three datasets, we consistently observed that the use of PG likelihood yielded superior performance compared to using either Poisson or Gaussian likelihood alone, as expected. Additionally, the results obtained from the CT-density dataset were generally of lower quality than those from the other two datasets. This can be attributed to lower average counts per pixel (many zero pixels near the image borders).



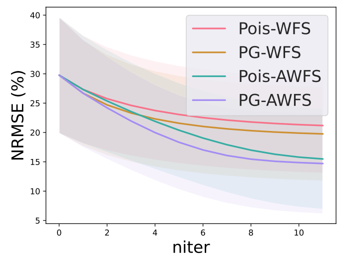
(a) Histopathology dataset: $\alpha = 0.02$.



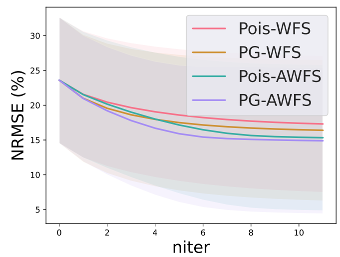
(b) Histopathology dataset: $\alpha = 0.0275$.



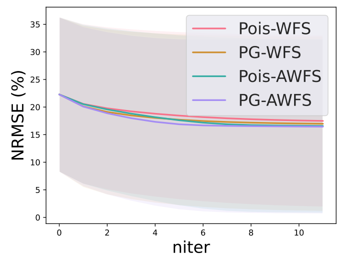
(c) Histopathology dataset: $\alpha = 0.035$.



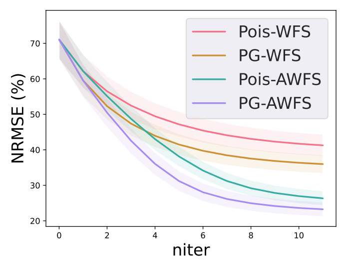
(d) CelebA dataset: $\alpha = 0.02$.



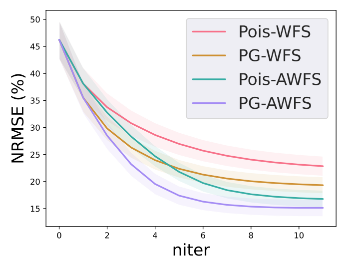
(e) CelebA dataset: $\alpha = 0.0275$.



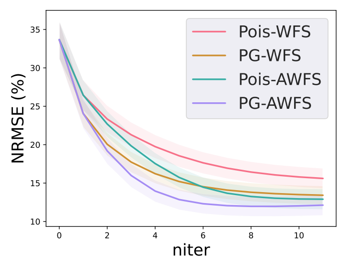
(f) CelebA dataset: $\alpha = 0.035$.



(g) CT-density dataset: $\alpha = 0.02$.



(h) CT-density dataset: $\alpha = 0.0275$.



(i) CT-density dataset: $\alpha = 0.035$.

FIG 3.15 – Comparing AWFS vs. WFS with NRMSE vs. number of iterations under different noise levels. The curves and shadows represent the mean and standard deviation, respectively.

Using a DL image prior can be considered from two perspectives: training a denoiser or training to learn the density distribution of images. In our work, we applied both approaches and observed that the effectiveness of these methods differed depending on the dataset tested. Specifically, in the Histopathology dataset [3] and the CT-density dataset, where the images share similar structures, the generative models performs better even when trained with limited data. In the case of the CelebA dataset [143], which includes a wide variety of celebrity faces, generative models did not exhibit as strong performance as denoiser methods when trained on limited data. This is likely due to the fact that generating high-quality images is generally more challenging than removing noise from existing images and may necessitate a larger training dataset. We have also noticed that the PNP-ADMM method provided unsatisfying reconstruction quality, possibly attributable to the non-zero duality gap and slow convergence for non-convex problems [222]. We plan to investigate it further in the future.

The effectiveness of accelerated WF compared to vanilla WF is due to the non-convexity of the PR problem. Although recent advances in geometric landscape analysis of PR can guarantee that all local minimizers are global even with random initialization [22], in practice the measurements are contaminated by noise so that many more measurements are required for the cost function to have a benign geometric landscape.

Despite the promising results achieved with our proposed AWFS approach, there are several limitations of our work. First, the approximate calculation of the infinite sum in (3.25) is accurate but computationally expensive. Future work should seek ways to accelerate this calculation while maintaining accuracy. Second, we did not implement and test the accelerated WF applied on the diffusion posterior sampling method [43], for which the network is fine-tuned from a pretrained state-of-the-art diffusion model. This approach has the potential to advance current methods in PR problem and we will investigate it in the future.

3.2.5 Conclusion

We proposed a novel algorithm based on **A**ccelerated **W**irtinger **F**low and **S**core-based image prior (AWFS) for Poisson-Gaussian holographic phase retrieval. With evaluation on simulated experiments, we demonstrated that our proposed AWFS algorithm had the best reconstruction quality both qualitatively and quantitatively and was more robust to various noise levels, compared to other state-of-the-art methods. Furthermore, we proved that our proposed algorithm has a critical-point convergence guarantee. Therefore, our approach has much promise for translation in real-world applications encountering phase retrieval problems.

CHAPTER 4

Poisson Inverse Problems in SPECT Imaging[1]

4.1 Training End-to-End Unrolled Iterative Neural Networks for SPECT Image Reconstruction[2]

4.1.1 Movitation

Although MLEM and OSEM (introduced in Chapter 2) have achieved great success in clinical uses, they are known to suffer from a trade-off between recovery and noise as well as limited compensation for SPECT resolution effects. To address that trade-off, regularization-based reconstruction methods have been proposed [173, 65, 122]. For example, Panin *et al.* [173] proposed TV regularization for SPECT reconstruction. However, TV regularization may lead to "blocky" images and over-smoothing the edges. One way to overcome blurring edges is to incorporate anatomical boundary side information from CT images [54], but that method requires accurate organ segmentation in advance. Chun *et al.* [41] proposed to use NLM filters that exploit the self-similarity of patches in images for regularization, yet that method is computationally expensive and hence not quite practical. In general, choosing an appropriate regularizer can be challenging and moreover, these traditional regularized algorithms lack generalizability to images that do not follow assumptions made by the prior.

With the recent success of deep learning and especially CNN, DL methods have been reported to outperform conventional algorithms in many medical imaging applications such as in MRI [246, 241, 178], CT [162, 31] and PET reconstruction [180, 114, 155]. However, fewer DL approaches to SPECT reconstruction appear in the literature. Shao *et al.* [192] proposed "SPECTnet" with a two-step training strategy that learns the transformation from projection space to image space as an alternative to the traditional OSEM algorithm. Shao

[1]This chapter is based on [134, 132, 140].
[2]This section is based on [134].

et al. [193] also proposed a DL method that can directly reconstruct the activity image from the SPECT projection data, even with reduced view angles. Mostafapour *et al.* [164] trained a neural network that maps non-attenuation-corrected SPECT images to those corrected by CT images as a post-processing procedure to enhance the reconstructed image quality.

Though promising results were reported with these methods, most of them worked in 2D whereas 3D is used in practice [192, 193]. Furthermore, there has yet to be an investigation of end-to-end training of DL regularizers that are embedded in unrolled SPECT iterative statistical algorithms such as regularized EM. End-to-end training is popular in machine learning and other medical imaging fields such as MRI image reconstruction [196], and is reported to meet data-driven regularization for inverse problems [165]. But for SPECT image reconstruction, such training is nontrivial to implement due to its complicated system matrix. Alternative training methods have been proposed, such as sequential training [142, 186, 171, 46] and gradient truncation [155]; these methods were shown to be effective, though they could yield sub-optimal reconstruction results due to approximations to the training loss gradient. Another approach is to construct a neural network that also models the SPECT system matrix, like in "SPECTnet" [192], but this approach lacks interpretability compared to algorithms like unrolled DL-regularized EM, *e.g.*, if one sets the regularization parameter to zero, then the latter becomes identical to the traditional EM.

As an end-to-end training approach has not yet been investigated for SPECT image reconstruction, this section first describes a SPECT forward-backward projector written in the open-source and high-performance Julia language that enables efficient auto-differentiation. Then we compare the end-to-end training approach with other alternative methods. Our contribution is summarized as follows:

- We provide a Julia implementation of forward-backward projector for SPECT image reconstruction, where the backprojector is the exact adjoint of the forward projector. Our Julia projector supports multi-threading on CPU for accelerating computation. We also provide an efficient Julia GPU implementation (by eliminating explicit scalar indexing) and PyTorch implementation for completeness. Our code is open source and is available at this link.

- Our Julia projector has comparable speed and accuracy compared to a public available Matlab-based projector[3], while using much less memory ($\sim$5%). Compared to MC methods for primary component, our Julia projector achieved a good approximation while being much faster.

[3]Available at `http://web.eecs.umich.edu/~fessler/irt/irt`.

- Simulation results based on ^{177}Lu XCAT phantoms and VP phantoms with ^{90}Y show that the unrolled DL regularized EM algorithm, when trained end-to-end with our proposed Julia projector, achieved the best reconstruction quality evaluated by line profiles and quantitative metrics like MAE and NRMSE, compared to other training methods such as sequential training and gradient truncation (ignoring gradients w.r.t. the forward-backward projector).

- Simulation results based on ^{177}Lu VP phantoms show that all learning-based methods achieved comparable reconstruction quality with the traditional OSEM method.

Overloaded Notations: $f(\cdot)$ and $g_{\boldsymbol{\theta}}$ defined in (2.5); $\boldsymbol{x}$, $\boldsymbol{A}$, $\boldsymbol{y}$, $\bar{\boldsymbol{r}}$, $h(\cdot)$ defined in (2.14); $\boldsymbol{u}_k$ defined in (2.23).

4.1.2 Methods

4.1.2.1 Implementation of Julia SPECT projector

Our Julia implementation of SPECT projector is based on [247], modeling parallel-beam collimator geometries. Our projector also accounts for attenuation and depth-dependent collimator response. We did not model the scattering events like Compton scatter and coherent scatter of high energy gamma rays within the object.

For the forward projector, at each rotation angle, we first rotate the 3D image matrix $\boldsymbol{x} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{y} \times n_\mathrm{z}}$ according to the third dimension by its projection angle θ_l (typically $2\pi(l-1)/n_\mathrm{view}$); l denotes the view index, which ranges from 1 to n_view and n_view denotes the total number of projection views. We implemented and compared both bilinear interpolation and 3-pass 1D linear interpolation [58] with zero padding boundary condition for image rotation. For attenuation correction, we first rotated the three-dimensional attenuation map $\boldsymbol{\mu} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{y} \times n_\mathrm{z}}$ (obtained by transmission tomography) also by θ_l. Assuming n_y is the index corresponding to the closest plane of $\boldsymbol{x}$ to the detector, then we model the accumulated attenuation factor $\bar{\boldsymbol{\mu}}$ for each view angle as

$$\bar{\mu}(i,j,k;l) = \mathrm{e}^{-\Delta_\mathrm{y}\left(\frac{1}{2}\mu(i,j,k;l) + \sum_{s=j+1}^{n_\mathrm{y}} \mu(i,s,k;l)\right)}, \tag{4.1}$$

where i, j, k denotes the 3D voxel coordinate and Δ_y denotes the voxel size for the (first and) second coordinate. Next, for each y slice (an (x,z) plane for a given j index) of the rotated and attenuated image, we convolved with the appropriate slice of the depth-dependent point spread function $\boldsymbol{p} \in \mathbb{R}^{p_\mathrm{x} \times p_\mathrm{z} \times n_\mathrm{y} \times n_\mathrm{view}}$ using a 2D FFT. Here we used replicate padding for both the i and k coordinates. The view-dependent PSF accommodates non-circular orbits. Finally, the forward projection operation simply sums the rotated, blurred and attenuated activity image $\boldsymbol{x}$ along the second coordinate j. Algorithm 5 summarizes the forward projector, where $\circledast$ denotes a 2D convolution operation.

Algorithm 5: SPECT forward projector

Input: 3D image $\boldsymbol{x} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{y} \times n_\mathrm{z}}$,
3D attenuation map $\boldsymbol{\mu} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{y} \times n_\mathrm{z}}$,
4D point spread function $\boldsymbol{p} \in \mathbb{R}^{p_\mathrm{x} \times p_\mathrm{z} \times n_\mathrm{y} \times n_\mathrm{view}}$,
voxel size Δ_y.
Initialize: $\boldsymbol{v} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{z} \times n_\mathrm{view}}$ as all zeros.
for $l = 1, ..., n_\mathrm{view}$ **do**
  $\tilde{\boldsymbol{x}} \leftarrow$ rotate $\boldsymbol{x}$ by θ_l
  $\tilde{\boldsymbol{\mu}} \leftarrow$ rotate $\boldsymbol{\mu}$ by θ_l
  for $j = 1, ..., n_\mathrm{y}$ **do**
    $\bar{\boldsymbol{\mu}} \leftarrow$ calculate by (4.1) using $\tilde{\boldsymbol{\mu}}$
    $\tilde{x}(i, j, k) \mathrel{*}= \bar{\mu}(i, j, k)$
    $v(i, k, l) \mathrel{+}= \tilde{x}(i, j, k; l) \circledast p(i, k; j, l)$
  end
end
Output: $\boldsymbol{v} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{z} \times n_\mathrm{view}}$

All of these steps are linear, so hereafter, we use $\boldsymbol{A}$ to denote the forward projector, though it is not stored explicitly as a matrix. Because each each step is linear, each step has an adjoint operation. Overall, the backward projector is the adjoint of $\boldsymbol{A}$ that satisfies

$$\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}'\boldsymbol{y} \rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y}. \tag{4.2}$$

The exact adjoint of (discrete) image rotation is not simply a discrete rotatation of the image by $-\theta_l$. Instead, one should also consider the adjoint of linear interpolation. For the adjoint of convolution, we assume the point spread function is symmetric along coordinates i and k so that the adjoint convolution operator is just the forward convoluation operator along with the adjoint of replicate padding. Algorithm 6 summarizes the SPECT backward projector.

To accelerate the for-loop process, we used multi-threading to enable projecting or backprojecting multiple angles at the same time. To reduce memory, we pre-allocated necessary arrays and used fully in-place operations inside the for-loop. To further accelerate auto-differentiation, we customized the chain rule to use the linear operator $\boldsymbol{A}$ or $\boldsymbol{A}'$ as the Jacobian when calling $\boldsymbol{A}\boldsymbol{x}$ or $\boldsymbol{A}'\boldsymbol{y}$ during backpropagation. We implemented and tested our projector in Julia v1.6; we also implemented a GPU version in Julia (using CUDA.jl) that runs efficiently on a GPU by eliminating explicit scalar indexing. For

Algorithm 6: SPECT backward projector

Input: 3D view $\boldsymbol{v} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{z} \times n_\mathrm{view}}$,
3D attenuation map $\boldsymbol{\mu} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{y} \times n_\mathrm{z}}$,
4D point spread function $\boldsymbol{p} \in \mathbb{R}^{p_\mathrm{x} \times p_\mathrm{z} \times n_\mathrm{y} \times n_\mathrm{view}}$,
voxel size Δ_y.
Initialize: $\boldsymbol{x} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{y} \times n_\mathrm{z}}$ as all zeros.
for $l = 1, ..., n_\mathrm{view}$ **do**
  $\tilde{\boldsymbol{\mu}} \leftarrow$ rotate $\boldsymbol{\mu}$ by θ_l
  for $j = 1, ..., n_\mathrm{y}$ **do**
    $\bar{\boldsymbol{\mu}} \leftarrow$ calculate by (4.1) using $\tilde{\boldsymbol{\mu}}$
    $\tilde{v}(i, k, l) \leftarrow$ adjoint of $v(i, k, l) \circledast p(i, k, j, l)$
    $\tilde{x}(i, j, k; l) \leftarrow \tilde{v}(i, k, l) \cdot \bar{\mu}(i, j, k; l)$
  end
  $\boldsymbol{x}$ += adjoint rotate $\tilde{\boldsymbol{x}}$ by θ_l
end
Output: $\boldsymbol{x} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{y} \times n_\mathrm{z}}$

completeness, we also provide a PyTorch version but without multi-threading support, in-place operations nor the exact adjoint of image rotation.

4.1.2.2 Unrolled DL regularized EM algorithm

The DL regularized EM algorithm is summarized in Algorithm 1. To train $g_{\boldsymbol{\theta}}$, the most direct way is to unroll the EM algorithm and train end-to-end with an appropriate target; this supervised approach requires backpropagating through the SPECT system model, which is not trivial to implement with previous SPECT projection tools. There are several non-end-to-end training methods such as sequential training [142] that first train $\boldsymbol{u}_k$ by the target and then plug into (2.26) at each iteration. This method must use non-shared weights for the neural network per each iteration. Another method is gradient truncation [155] that ignores the gradient w.r.t. the system matrix $\boldsymbol{A}$ and its adjoint $\boldsymbol{A}'$ during backpropagation. Both of these training methods, though reported to be effective, may be sub-optimal because they approximate the overall training loss gradients.

4.1.2.3 Phantom Dataset and Simulation Setup

We used simulated XCAT phantoms [189] and virtual patient phantoms for experiment results presented in the next section. Each XCAT phantom was simulated to approximately follow the activity distributions observed when imaging patients after ^{177}Lu DOTATATE therapy. We set the image size to $128 \times 128 \times 80$ with voxel size $4.8 \times 4.8 \times 4.8\mathrm{mm}^3$.

Tumors of various shapes and sizes (5-100mL) were located in the liver as is typical for patients undergoing this therapy.

For virtual patient phantoms, we consider two radionuclides: ^{177}Lu and ^{90}Y. For ^{177}Lu phantoms, the true images were from PET/CT scans of patients who underwent diagnostic ^{68}Ga DOTATATE PET/CT imaging (Siemens Biograph mCT) to determine eligibility for ^{177}Lu DOTATATE therapy. The ^{68}Ga DOTATATE distribution in patients is expected to be similar to ^{177}Lu and hence can provide a reasonable approximation to the activity distribution of ^{177}Lu in patients for DL training purposes but at higher resolution. The PET images had size $200 \times 200 \times 577$ and voxel size $4.073 \times 4.073 \times 2$ mm^3 and were obtained from our Siemens mCT (resolution is 5–6 mm FWHM [201]) and reconstructed using the standard clinic protocol: 3D OSEM with three iterations, 21 subsets, including resolution recovery, time-of-flight, and a 5mm (FWHM) Gaussian post-reconstruction filter. The density maps were also generated using the experimentally derived CT-to-density calibration curve.

For ^{90}Y phantoms, the true activity images were reconstructed (using a previously implemented 3D OSEM reconstruction with CNN based scatter estimation [231]) from ^{90}Y SPECT/CT scans of patients who underwent ^{90}Y microsphere radioembolization in our clinic.

In total, we simulated 4 XCAT phantoms, 8 ^{177}Lu and 8 ^{90}Y virtual patient phantoms. All image data have University of Michigan Institutional Review Board (IRB) approval for retrospective analysis. For all simulated phantoms, we selected the center slices covering the lung, liver and kidney corresponding to SPECT axial FOV (39cm).

Then we ran SIMIND MC program [144] to generate the radial position of SPECT camera for 128 view angles. The SIMIND model parameters for ^{177}Lu were based on ^{177}Lu DOTATATE patient imaging in our clinic (Siemens Intevo with medium energy collimators, a 5/8" crystal, a 20% photopeak window at 208 keV, and two adjacent 10% scatter windows) [55]. For ^{90}Y, a high-energy collimator, 5/8" crystal, and a 105 to 195 keV acquisition energy window was modeled as in our clinical protocol for ^{90}Y bremsstrahlung imaging. Next we approximated the point spread function for ^{177}Lu and ^{90}Y by simulating point source at 6 different distances (20, 50, 100, 150, 200, 250mm) and then fitting a 2D Gaussian distribution at each distance. The camera orbit was assumed to be non-circular (auto-contouring mode in clinical systems) with the minimum distance between the phantom surface and detector set at 1 cm.

4.1.3 Experiment Results

4.1.3.1 Comparison of Projectors

We used an XCAT phantom to evaluate the accuracy and memory-efficiency of our Julia projector.

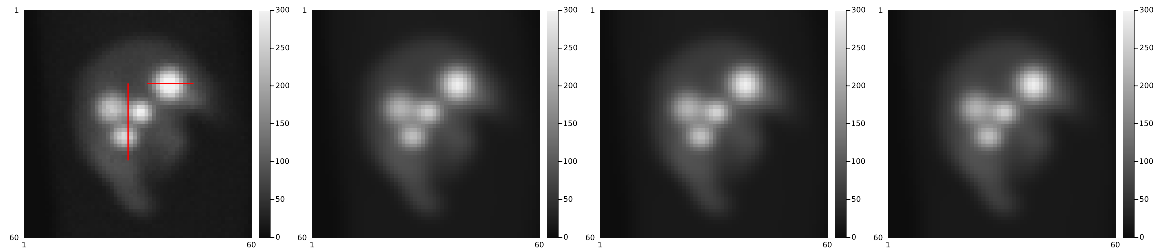
Accuracy. We first compared primary projection images and profiles generated by our Julia projector with those from MC simulation and the Matlab projector. For results of MC, we ran two SIMIND simulations for 1 billion histories using ^{177}Lu and ^{90}Y as radionuclide source, respectively. Each simulation took about 10 hours using a 3.2 GHz 16-Core Intel Xeon W CPU on MacOS. The Matlab projector was originally implemented and compiled in C99 and then wrapped by a Matlab MEX file as a part of the Michigan Image Reconstruction Toolbox (MIRT) [66]. The physics modelling of the Matlab projector was the same as our Julia projector except that it only implemented 3-pass 1D linear interpolation for image rotation. Unlike the memory-efficient Julia version, the Matlab version pre-rotates the patient attenuation map for all projection views. This strategy saves time during EM iterations for a single patient, but uses considerable memory and scales poorly for DL training approaches involving multiple patient datasets.

Fig. 4.1 compared the primary projections generated by different methods without adding Poisson noise. Visualizations of image slices and line profiles illustrate that our Julia projector (with rotation based on 3-pass 1D interpolation) is almost identical to the Matlab projector, while both give a reasonably good approximation to the MC.

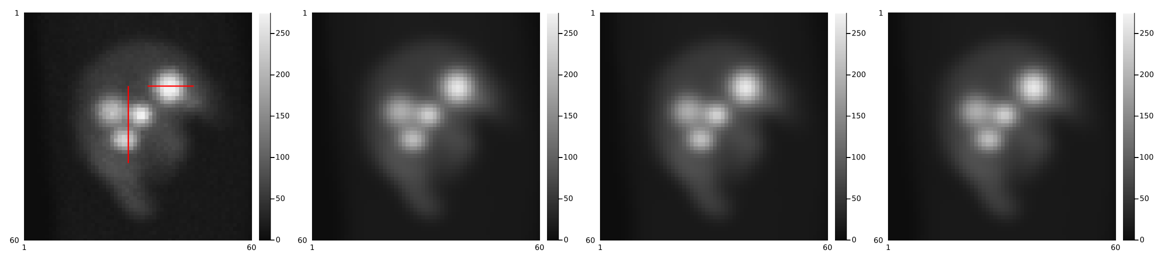
Speed and Memory Use. We then compared the memory use and compute times between our Julia projector and the Matlab projector using different number of threads when projecting a $128 \times 128 \times 80$ image. Fig. 4.2 shows that our Julia projector has comparable computing time for a single projection with 128 view angles using different number of CPU threads, while only uses a very small fraction of memory (∼5%) and pre-allocation time (∼1%) compared to the Matlab projector.

4.1.3.2 Comparison of DL-regularized EM Using Different Training Methods

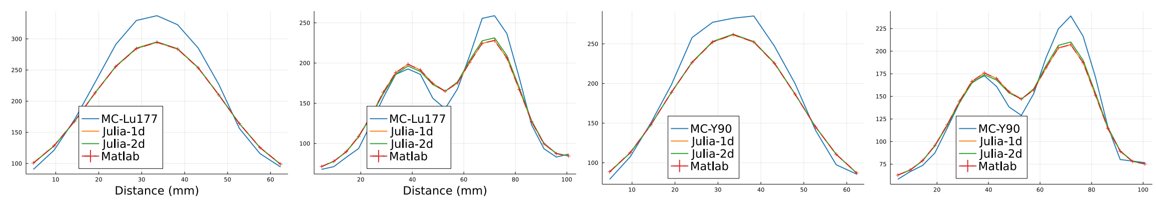
This section compares end-to-end training with other training methods that have been used previously for SPECT, namely the gradient truncation and sequential training. We implemented an unrolled DL-regularized EM algorithm with 3 outer iterations, each of which had one inner iteration. Here we used only 3 outer iterations (compared to previous works such as [155]) because we used the 16-iteration 4-subset OSEM reconstructed



(a) MC (^{177}Lu) (b) Julia 1D (^{177}Lu) (c) Julia 2D (^{177}Lu) (d) Matlab (^{177}Lu)



(e) MC (^{90}Y) (f) Julia 1D (^{90}Y) (g) Julia 2D (^{90}Y) (h) Matlab (^{90}Y)



(i) ^{177}Lu hori. profile (j) ^{177}Lu vert. profile (k) ^{90}Y hori. profile (l) ^{90}Y vert. profile

FIG 4.1 – Projections generated by MC simulation, Matlab projector and our Julia projector with 3-pass 1D linear interpolation and 2D bilinear interpolation for image rotation, using ^{177}Lu and ^{90}Y radionuclides. Subfigure (i)-(l) show line profiles across tumors as shown in subfigure (a) and (e), respectively. MC projections were scaled to have the same total activities as the Matlab projector per FOV.

image as a warm start for all reconstruction algorithms. We set the regularization parameter (defined in (2.5)) as $\beta = 1$. The DL regularizer was a 3-layer 3D CNN, where each layer had a $3 \times 3 \times 3$ convolutional filter followed by RELU activation (except the last layer). We added the input image $\boldsymbol{x}_k$ to the output of CNN following the common residual learning strategy [86]. End-to-end training and gradient truncation could also work with a shared weights CNN approach, but were not included here for fair comparison purpose, since the sequential training only works with non-shared weights CNN. All the neural networks were initialized by with the same parameters (drawn from a Gaussian distribution) and trained on an Nvidia RTX 3090 GPU by minimizing mean square error (loss) using AdamW optimizer [146] with a constant learning rate 0.002.

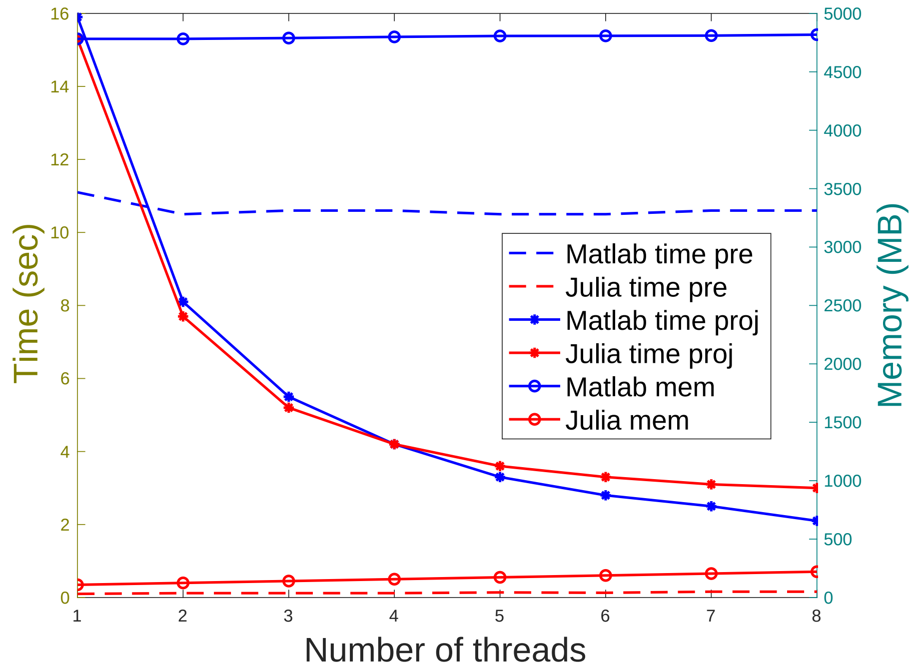


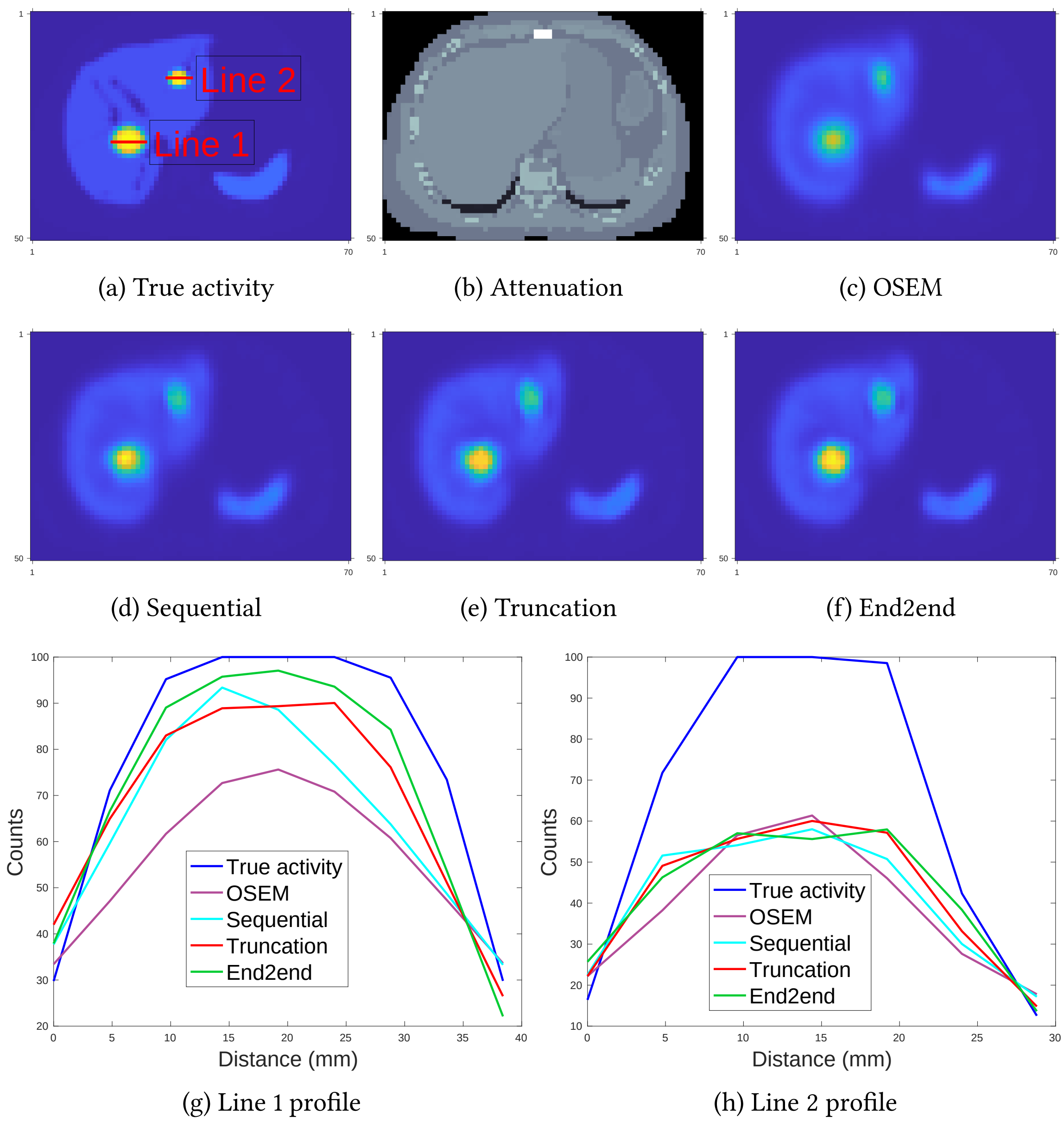
FIG 4.2 – Time and memory comparison between Matlab projector and our Julia projector for projecting 128 view angles of a $128 \times 128 \times 80$ image. “time pre” denotes the time cost for pre-allocating necessary arrays before projection; “time proj” denotes the time cost for a single projection; “mem” denotes the memory usage. All methods were tested on MacOS with a 3.8 GHz 8-Core Intel Core i7 CPU.

Besides line profiles for qualitative comparison, we also used MAE and NRMSE as quantitative evaluation metrics. All activity images were scaled by a factor that normalized the whole activity to 1 MBq per FOV before comparison.

Results on ^{177}Lu XCAT Phantoms. We evaluated these reconstruction algorithms using 4 ^{177}Lu XCAT phantoms we simulated. We generated the primary projections by calling forward operation of our Julia projector and then added uniform scatters with 10% of the primary counts before adding Poisson noise. Of the 4 phantoms, we used 2 for training, 1 for validation and 1 for testing.

Fig. 4.3 shows that the end-to-end training yielded incrementally better reconstruction of the tumor in the liver center over OSEM, sequential training and gradient truncation. Fig. 4.3 (g) also illustrates this improvement by the line profile across the tumor. For the tumor at the top-right corner of the liver, all methods had comparable performance; this can be attributed to the small tumor size (5mL) for which PV effects associated with SPECT resolution are higher; and hence its recovery is even more challenging.

Table 4.1 demonstrates that all DL methods (sequential training, gradient truncation and end-to-end training) consistently had lower reconstruction error than the traditional



(a) True activity (b) Attenuation (c) OSEM

(d) Sequential (e) Truncation (f) End2end

(g) Line 1 profile (h) Line 2 profile

FIG 4.3 – Qualitative comparison of different training methods and OSEM tested on ^{177}Lu XCAT phantoms. Subfigure (a)-(c): true activity map, attenuation map and OSEM reconstruction (16 iterations and 4 subsets); (d)-(f): regularized EM using sequential training, gradient truncation, end2end training, respectively; (g) and (h): line profiles in (a).

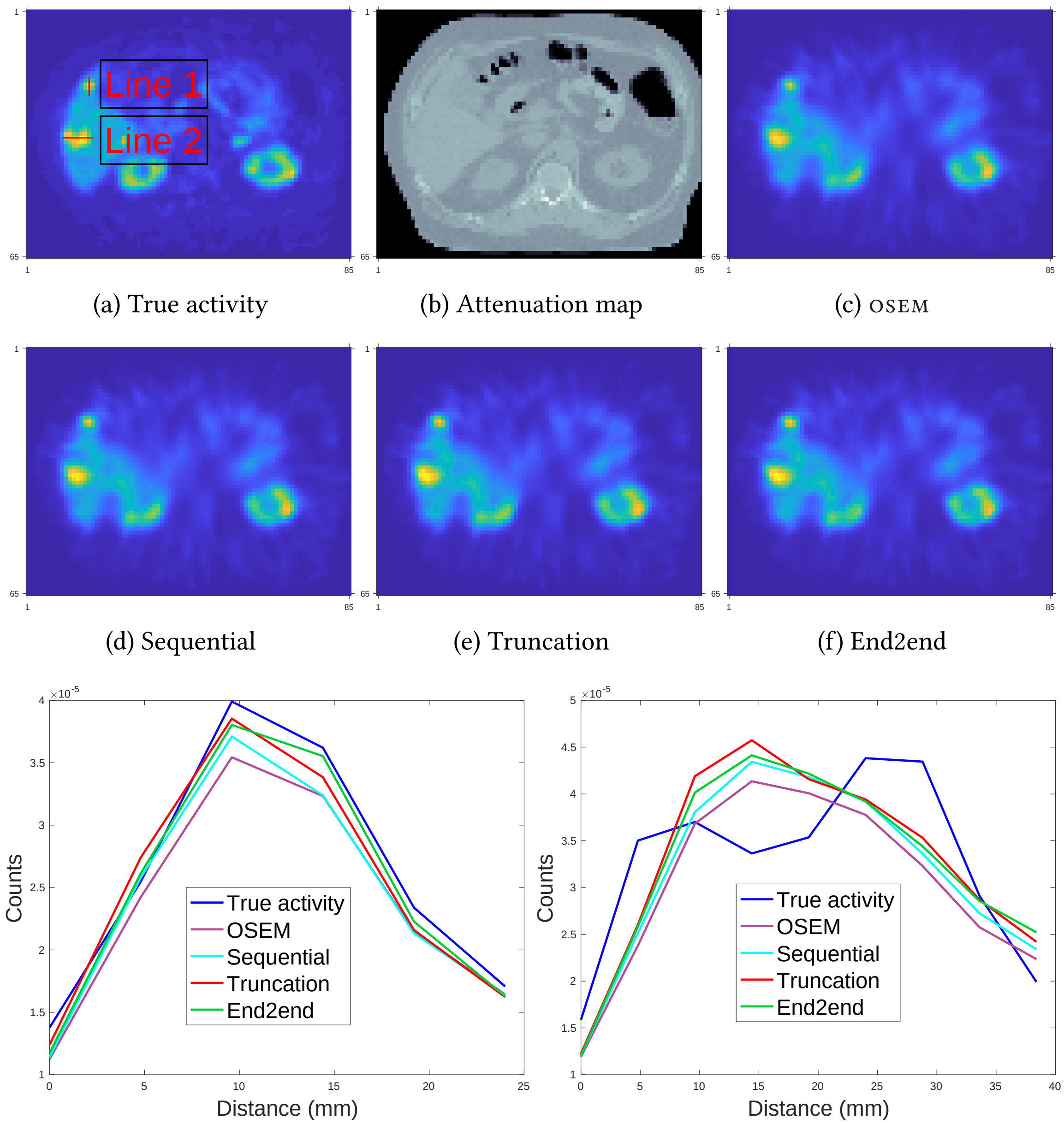
OSEM method. Among all DL methods, the proposed end-to-end training had lower MAE over nearly all lesions and organs than other training methods. The relative reduction in MAE by the end-to-end training was up to 36% (for lesion 3) compared to sequential training. End-to-end training also had lower NRMSE for most lesions and organs, and was otherwise comparable to other training methods. The relative improvement compared to sequential training was up to 29% (for lesion 3).

TBL 4.1 – MAE(%) and NRMSE (%) for ^{177}Lu XCAT phantoms.

MAE				
Lesion/Organ	OSEM	Sequential	Truncation	End2end
Lesion 1 (67mL)	11.8	4.7	3.8	**3.4**
Lesion 2 (10mL)	19.2	8.8	9.8	**8.5**
Lesion 3 (9mL)	25.1	19.2	14.5	**12.3**
Lesion 4 (5mL)	42.5	39.7	**37.9**	**37.9**
Liver	5.6	4.5	3.4	**2.3**
Lung	3.0	2.4	1.3	**1.2**
Spleen	13.2	10.3	8.1	**7.6**
Kidney	14.8	13.7	12.8	**11.8**
NRMSE				
Lesion/Organ	OSEM	Sequential	Truncation	End2end
Lesion 1 (67mL)	27.3	20.2	19.5	**19.0**
Lesion 2 (10mL)	26.1	17.6	16.2	**15.4**
Lesion 3 (9mL)	28.0	22.2	17.6	**15.8**
Lesion 4 (5mL)	43.0	40.9	**39.2**	39.5
Liver	28.5	**24.3**	24.6	24.8
Lung	32.2	30.7	**29.7**	30.7
Spleen	25.3	21.6	20.0	**19.3**
Kidney	40.7	39.4	39.5	**39.0**

Results on ^{177}Lu VP Phantoms. Next we present test results on 8 ^{177}Lu virtual patient phantoms. Out of 8 ^{177}Lu phantoms, we used 4 for training, 1 for validation and 3 for testing.

Fig. 4.4 shows that the improvement of all learning-based methods was limited compared to OSEM, which is also evident from line profiles. For example, in Fig. 4.4 (g), where the line profile is drawn on a small tumor. We found that OSEM already gave a fairly accurate estimate, so we did not observe as much improvement as we had seen on ^{177}Lu XCAT phantoms for end-to-end training or even learning-based methods. Table 4.2 also demonstrates this behavior. The OSEM method had substantially lower MAE and NRMSE compared to the errors shown for ^{177}Lu XCAT data (cf Table 4.1). Moreover, all learning-based methods had comparable accuracy. For example, sequential training performed the best on lesions; gradient truncation was the best on liver and lung in terms of MAE; end-to-end training had the lowest NRMSE on kidney and spleen. Perhaps this could be due to different local minimizers were reached when training neural networks; a more comprehensive study would be needed to verify this conjecture.



(a) True activity (b) Attenuation map (c) OSEM

(d) Sequential (e) Truncation (f) End2end

(g) Line 1 profile (h) Line 2 profile

FIG 4.4 – Qualitative comparison of different training methods and OSEM tested on ^{177}Lu VP phantoms. Subfigure (g) and (h) correspond to line profiles marked in (a).

TBL 4.2 – MAE(%) and NRMSE(%) for ^{177}Lu VP phantoms.

MAE				
Lesion/Organ	OSEM	Sequential	Truncation	End2end
Lesion (6-152mL)	6.1	**5.0**	5.6	5.3
Liver	4.2	3.6	**1.5**	2.2
Healthy liver	3.8	4.1	**2.0**	2.2
Lung	4.5	3.8	**1.2**	2.6
Kidney	5.7	4.9	**2.3**	**2.3**
Spleen	1.6	**1.4**	2.4	1.8
NRMSE				
Lesion/Organ	OSEM	Sequential	Truncation	End2end
Lesion (6-152mL)	17.5	**16.7**	17.7	16.8
Liver	19.5	**18.9**	19.0	19.0
Healthy liver	21.6	**21.3**	21.8	21.8
Lung	23.2	**23.0**	25.5	24.9
Kidney	18.6	17.9	17.3	**16.9**
Spleen	15.1	13.7	13.1	**12.5**

Results on ^{90}Y VP Phantoms. For completeness, we also tested with 8 ^{90}Y virtual patient phantoms. Of the 8 phantoms, we used 4 for training, 1 for validation and 3 for testing.

Fig. 4.5 compares the reconstruction quality among OSEM, sequential training, gradient truncation and end-to-end training on two test slices (subfigure (a)-(f) and (g)-(l)). Visually, the end-to-end training reconstruction gives the closest estimate to the true activity, this is also evident through the line profiles (subfigure (m) and (n)) across the tumor and the liver.

Table 4.3 reports the MAE and NRMSE for lesions and organs across all testing phantoms. Similar to the qualitative assessment (Fig. 4.5), the end-to-end training also produced lower errors consistently across all testing lesions and organs. For instance, compared to sequential training/gradient truncation, the end-to-end training reduced MAE on average by 20.7%/13.2%, 35.0%/22.9% and 55.6%/42.7% for lesion, healthy liver and lung, respectively. The NRMSE was also reduced by 14.5%/7.3%, 16.0%/9.2% and 15.7%/6.7% for lesion, healthy liver and lung, respectively. All learning-based methods consistently had lower errors than the OSEM method.

TBL 4.3 – MAE(%) and NRMSE(%) for ^{90}Y VP phantoms.

MAE				
Lesion/Organ	OSEM	Sequential	Truncation	End2end
Lesion (3-356mL)	33.9	26.6	24.3	**21.1**
Liver	25.1	18.5	16.3	**11.6**
Healthy liver	25.3	24.3	20.5	**15.8**
Lung	90.2	65.7	51.0	**29.2**
NRMSE				
Lesion/Organ	OSEM	Sequential	Truncation	End2end
Lesion (3-356mL)	37.0	31.1	28.7	**26.6**
Liver	30.2	22.7	21.3	**19.6**
Healthy liver	32.0	28.2	26.1	**23.7**
Lung	63.4	59.8	54.0	**50.4**

4.1.3.3 Results at intermediate iterations

One potential problem associated with end-to-end training (and gradient truncation) is that the results at intermediate iterations could be unfavorable, because they are not directly trained by the targets [118]. Here, we examined the images at intermediate iterations and did not observe any problems as illustrated in Fig. 4.6, where images at each iteration gave a fairly accurate estimate to the true activity. Perhaps under the shallow-network setting (*e.g.*, 3 layers used here, with only 3 outer iterations), the network for each iteration was less likely to overfit the training data. Another reason could be due to the non-shared weights setting so that the network could learn suitable weights for each iteration.

4.1.4 Discussion

In previous DL-based iterative algorithms for SPECT image reconstruction, the neural networks were trained either sequentially [142] or with gradient truncation [156], due to memory inefficient forward-backward projectors so that backpropagating through the projectors is not easy. This work investigated a new SPECT projector using Julia that is an open-source, high performance and cross-platform language. With comparisons between MC and a public available Matlab-based projector, we verified the accuracy, speed and memory-efficiency of our Julia projector. These favorable properties support efficient backpropagation when training end-to-end unrolled iterative reconstruction algorithms. Most modern DL algorithms process multiple data batches in parallel, so memory efficiency is of great importance for efficient training and testing neural networks. To that extent, our Julia projector is much more suitable than our previously developed Matlab-based projector.

We used a simple regularized EM algorithm as an example to test end-to-end training and other training methods on different datasets including ^{177}Lu XCAT phantoms, ^{177}Lu and ^{90}Y virtual patient phantoms. Simulation results demonstrated the effectiveness of end-to-end training on these datasets. For example, end-to-end training improved the MAE of lesion/liver in ^{90}Y phantoms by 21%/37% and 13%/29% compared to sequential training and gradient truncation. This improvement could be attributed to a better DL regularizer when trained in end-to-end fashion. Although the end-to-end training yielded the lowest reconstruction error on both ^{177}Lu XCAT phantoms and ^{90}Y VP phantoms, the reconstruction errors on ^{177}Lu VP phantoms were comparable for all methods. This could be due to CNN architecture choice and hyperparameters (such as number of iterations) in the EM algorithm, which we will explore in the future. Another reason could be with the nonuniform activity in VP phantoms where the recovery is generally higher than activity for the XCAT phantom (MAE reported in Table 4.1 and Table 4.2) because the assigned "true" activities at the boundaries of organs did not drop sharply, and instead, were blurred out. And therefore the OSEM algorithm was fairly competitive as reported in Table 4.2; in ^{90}Y VP results, the OSEM performs worse than learning-based methods, which could be attributed to the high downscatter associated with ^{90}Y SPECT due to the continuous bremsstrahlung energy spectrum.

We found all learning methods did not work very well for small tumors (*e.g.*, 5mL), potentially due to the worse PV effect. Reducing PV effects in SPECT images has been studied extensively [175, 83]. Xie *et al.* [233] trained a deep neural network to learn the mapping between PV-corrected and non-corrected images. Incorporating their network into our reconstruction model using transfer learning is an interesting future direction.

Although the previous sections showed promising results, this work has several limitations. First, we did not test numerous hyperparameters and CNN architectures, nor with a wide variety of phantoms and patients for different radionuclides therapies. Another limitation is that we did not investigate more advanced parallel computing methods such as distributed computing using multiple computers to further accelerate our Julia implementation of SPECT forward-backward projector. Such acceleration is feasible using existing Julia packages if needed. The compute times reported in Fig. 4.2 show that the method needs a few seconds per 128 projection views using 8 threads, which is already feasible for scientific investigation. Finally, we did not explore the effect of model mismatch between our Julia projector and MC simulation when plugged into the end-to-end training framework, but we expect the neural network can learn to compensate for the mismatch after training.

4.1.5 Conclusion

This section presents a Julia implementation of a backpropagatable SPECT forward-backward projector that is accurate, fast and memory-efficient compared to MC and a previously developed analytical Matlab-based projector. Simulation results based on ^{177}Lu XCAT phantoms, ^{90}Y and ^{177}Lu VP phantoms demonstrate that: 1) End-to-end training yielded the lowest MAE and NRMSE when tested on XCAT phantoms and ^{90}Y VP phantoms, compared to other training methods (such as sequential training and gradient truncation) and OSEM. 2) For ^{177}Lu VP phantoms, we observed all training methods yielded comparable reconstruction results as the OSEM method. These results indicate that end-to-end training, which is feasible with our developed Julia projector, is worth investigating for SPECT reconstruction.

4.2 DblurDoseNet: A Deep Neural Network for SPECT Dosimetry Estimation and Resolution Recovery[4]

4.2.1 Motivation

Accurate and computationally efficient methods for patient-specific absorbed dose estimation are also essential for clinical implementation of dosimetry-guided treatment planning in radionuclide therapy. For example, current ^{177}Lu DOTATATE therapy for neuroendocrine tumors uses a fixed activity basis (4 cycles of 7.4 GBq), whereas SPECT/CT imaging-based dosimetry after one cycle can be used to individualize the next administration to potentially enhance tumor response while keeping toxicity to critical organs like kidney at an acceptable level [60]. Traditionally, the mean absorbed doses in volumes of interest (VOI) are the reported quantity. However, voxel-level calculation enables consideration of multiple alternative dose metrics, such as statistics from dose-rate volume histogram (DRVH) analyses that are potentially more relevant to treatment planning. Explicit MC radiation transport using the patient's emission (PET or SPECT) and anatomical images (CT) as input is broadly accepted as the gold standard for voxel-level patient-specific dosimetry; however, it is computationally expensive to generate estimates with low statistical uncertainty. In contrast, faster and simpler DVK convolution methods [18] can be inaccurate in

[4]This section is based on [132].

the presence of heterogeneous tissues, *e.g.*, at the liver-lung or bone-marrow interfaces. Moreover, even though MC is theoretically accurate, the dose accuracies of both MC and DVK methods are degraded by reconstruction artifacts and the limited spatial resolution of SPECT and PET images.

There is an increased interest in studies that apply deep neural networks in nuclear medicine applications [215, 255, 190, 217, 4]. However, deep learning applications in radionuclide therapy dosimetry are limited [2, 128, 79, 78]. Akhavanallaf *et al.* [2] employed a modified ResNet [86] that represented voxel S-values kernels [18] to predict the distribution of the deposited energy in whole-body organ-level dosimetry and demonstrated comparable performance to the direct MC approach. Lee *et al.* [128] implemented a 3D UNet [184] that used PET and CT-based density image patches to predict 3D voxel-level dose-rate maps. Götz *et al.* [79] proposed a hybrid method based on a combination of a modified UNet and an empirical mode decomposition of density maps to enhance the accuracy/reliability of radiation dose estimation. Götz *et al.* [78] also trained a neural network to predict DVK for dosimetry in ^{177}Lu targeted radionuclide therapies. Despite promising results, a limitation of the training approaches in these prior studies [2, 128, 79, 78] is that they used MC-generated dose-rate maps derived from each patient's measured SPECT or PET images as the training label, which are degraded by the camera spatial resolution and reconstruction artifacts. Moreover, the concept of residual learning can be adopted in a CNN dosimetry model by exploiting a fast DVK convolution dose-rate map as an initial estimate. Residual learning for image denoising was first proposed to improve the effectiveness and efficiency of a denoising CNN [252] and was further applied to low-dose PET and CT reconstruction [235, 32].

The aim of this study was to develop a deep learning-based absorbed dose-rate estimation method that can overcome the accuracy-efficiency trade-off associated with current voxel dosimetry methods and attempt to learn to reduce the degrading effects of spatial resolution and reconstruction artifacts. Specifically, we used dose-rate estimates directly corresponding to phantom (virtual patient) activity maps as the training label, instead of the patient SPECT-derived dose-rate images (Fig. 4.7). Furthermore, unlike prior studies where a CNN was trained to directly estimate the dose-rate map or S-value kernels, we first used the approximate physics-based fast Fourier transform (FFT) DVK convolution method (with density scaling) to produce initial estimates, and then trained the CNN to learn the subtle residual differences between the initial estimate and the true dose-rate maps. We trained and tested the proposed CNN for SPECT/CT imaging-based dosimetry following ^{177}Lu DOTATATE therapy of neuroendocrine tumors (NETs).

4.2.2 Methods

4.2.2.1 Virtual Patient Phantom Generation for Training and Testing

Fig. 4.8 is an overview of our data generation and training process. To define the true activity maps of virtual patient phantoms, we chose to use PET instead of SPECT-based activity maps because PET offers substantially higher spatial resolution than SPECT as evident in the top branch of Fig. 4.8. These images were readily available because, prior to ^{177}Lu DOTATATE, patients underwent diagnostic ^{68}Ga DOTATATE PET/CT imaging (Siemens Biograph mCT) to determine eligibility for therapy.

The ^{68}Ga DOTATATE distribution in patients is expected to be similar to the ^{177}Lu DOTATATE distribution and hence our virtual patient phantoms can provide a reasonable approximation to the activity distribution of ^{177}Lu patients. The PET images (of size $200 \times 200 \times 577$, voxel size is $4.073 \times 4.073 \times 2\text{mm}^3$) were obtained from our Siemens mCT (resolution is 5-6mm FWHM [202]) and reconstructed using the standard clinic protocol: 3D OSEM with 3 iterations, 21 subsets that included resolution recovery, time-of-flight (TOF), and a 5mm (FWHM) Gaussian post-reconstruction filter. We selected 14 such PET images from our clinic database to generate anthropomorphic phantoms for training and testing, with University of Michigan Institutional Review Board (IRB) approval for retrospective analysis. The selected cases covered a diverse range with regards to sex (9 males and 5 females), age (35 to 88 years), weight (49 kg to 100 kg), and lesions of different sizes and location (within and outside the liver). The PET/CT images were first extracted into 195 slices with 0.2 cm slice width that covered the SPECT field-of-view (39 cm) with the liver and kidney centered, which is the typical region imaged following ^{177}Lu DOTATATE. Meanwhile, the corresponding density maps were generated using an experimentally derived CT-to-density calibration curve.

Next, ^{177}Lu SPECT projections corresponding to each phantom's activity/density maps were generated using the SIMIND MC code [144] (Fig. 4.8 top branch) simulating approximately 2 billion histories per projection. The SIMIND model parameters were based on ^{177}Lu patient imaging in our clinic (Siemens Intevo with medium energy collimators, a 5/8" crystal, a 20% photopeak window at 208 keV and two adjacent 10% scatter windows). Poisson noise was simulated after the 128 projection views were scaled to a count-level in the range of 3 to 20 million total counts, corresponding to the range in post-therapy imaging. SPECT reconstruction used an in-house 3D OSEM algorithm with CT-based attenuation correction, triple energy window scatter correction and collimator-detector response modeling (4 subsets and 16 iterations, $128 \times 128 \times 81$ matrix with voxel size

$4.8 \times 4.8 \times 4.8\text{mm}^3$, no Gaussian smoothing). All images were finally registered into CT image space ($512 \times 512 \times 130$ with voxel size $0.98 \times 0.98 \times 3\text{mm}^3$).

Out of 14 virtual patient phantoms, we randomly selected 9 for training and 5 for testing. Out of the training dataset, to assess under/over-fitting, we randomly selected 20% of the total slices to serve as a validation dataset.

4.2.2.2 Patient Data

In addition to the above virtual patients, our testing data included a total of 42 scans from 12 patients imaged at up to 4 time points during the first week following cycle 1 of standard ^{177}Lu DOTATATE (7.4 GBq). The images were acquired as part of an ongoing University of Michigan IRB approved research study, where all subjects signed an informed consent form. SPECT acquisition time was 25 minutes and all other SPECT imaging reconstruction parameters were as described above for the phantom simulation. The CT was performed in low-dose mode (120 kVp; 15 – 80 mAs) with free breathing.

4.2.2.3 Monte Carlo Dosimetry and Dose Voxel Kernel Convolution

Monte Carlo. The Monte Carlo code that we used, called Dose Planning Method (DPM), was originally developed and validated for fast dose-rate estimation in external beam radiotherapy [191]. Previously, we adapted and benchmarked DPM for internal radionuclide therapy applications [228]. Because DPM was optimized specifically for voxel-level electron/photon dose computations with full radiation transport, it is faster than using general-purpose MC codes for voxel-level dose estimation. We used DPM to generate the ground truth training labels (Fig. 4.8) by simulating approximately 1 billion histories to generate dose-rate maps with reasonably low statistical uncertainty. For example, with 1 billion histories for the phantom results shown in Fig. 4.10 and Fig. 4.11, the average statistical uncertainty across the kidney and lesions was less than 0.1% for both the ground truth MC run and the SPECT +MC run. (Obtained from the uncertainty images available from DPM).

DVK Convolution with Density Scaling. To provide DVK dose-rate maps for residual learning, ^{177}Lu soft tissue (1.04g/cm^3) voxel kernels were generated using DPM. The beta particle kernel size was $9 \times 9 \times 9$ and the photon kernel size was $99 \times 99 \times 99$ (both with voxel size $0.98 \times 0.98 \times 3\text{mm}^3$). We convolved the SPECT image with the DVK using fast Fourier transform (FFT)-convolution. Since using homogeneous soft tissue kernels neglects tissue inhomogeneities, we applied density scaling that has been shown to be a reasonable correction in past reports [59]. Here, after convolution, each voxel was scaled by 1.04 ($\mathrm{g/cm^3}$) and divided by the local voxel density value ($\mathrm{g/cm^3}$) derived from the CT scan. Because our goal was to generate a reasonably accurate and quick initial estimate for the residual learning process, we did not pursue other more sophisticated approaches [77, 127] that account for tissue heterogeneities. To address the very high dose-rate estimate in extra low-density regions, *e.g.*, air gaps, we set the dose-rate in regions where the density is less than 0.1 $\mathrm{g/cm^3}$ to zero.

4.2.2.4 Network: DblurDoseNet

Our network design considers the decay properties of ^{177}Lu and the physics of beta/photon interaction in tissue. The mean energy of the emitted electrons in the beta decay of ^{177}Lu is 134 keV and the maximum energy is 497 keV, and the corresponding continuous slowing down approximation (CSDA) ranges (in water) are 0.3 mm and 1.8 mm, respectively [20]. The gamma-rays associated with ^{177}Lu are low in intensity (113 keV (6.2%) and 208 keV (10.4%)), and hence, the absorbed dose is dominated by the beta component.

The input to the DVK method was an entire 3D SPECT image volume and its output was a 3D dose-rate map. In principle, a CNN could be designed similarly. However, for ^{177}Lu considering the short beta particle range in tissue and the low photon contribution, we designed a more memory efficient CNN that used a pack of 11 adjacent slices of the SPECT and density images at a time to produce one output dose-rate map corresponding to the middle slice of that pack. The CNN was applied with an 11-slice sliding window to all axial slices using padding that replicated the first and last slices at the top and bottom boundaries, respectively. Thus, the input to CNN was two arrays of size $512 \times 512 \times 11$ (with voxel size $0.98 \times 0.98 \times 3\mathrm{mm}^3$) and the output was an array of size 512×512 that corresponded to the dosimetry of the middle slice in the input arrays. During training and testing, these $512 \times 512 \times 11$ packs could be processed sequentially, but GPU devices could accelerate the processing by parallel computation.

As shown in Fig. 4.9, we first concatenated the input activity/density maps along the channel dimension, and then applied three 3D convolutional layers (with kernel size $7 \times 7 \times 5$, $7 \times 7 \times 3$, $7 \times 7 \times 3$, respectively) to extract depth features. Next, we implemented a 2D U-Net [184] that had 4 down-sample and up-sample layers, where the first convolutional layer in the 2D U-Net had 16 filters. After each down-sample layer, the number of filters at the next convolutional layer was increased by a factor of 2 until it reached 128. We added the DVK dose-rate map to the 2D U-Net output, as in the common residual learning approach. Finally, we obtained the CNN dose-rate map estimate after setting the dose-rate value in very low-density voxels ($\rho < 0.1\mathrm{g/cm^3}$) to zero.

The CNN was trained by minimizing the mean square error between the ground-truth and CNN dose-rate maps using a batch size of 32. We used the Adam optimizer [115] with a dynamic learning rate (an initial value 0.001 with ReduceOnPlateau management strategy) and trained our CNN for 200 epochs on two Nvidia Tesla V100 GPUs. The training/validation loss converged visually to 288/410 after 4 hours of training. To cover different input count levels, we normalized each SPECT activity map so that all its voxels summed to one, and then inversely scaled the dose-rate map estimate accordingly. To potentially improve convergence during training, we also scaled the normalized SPECT and dose-rate maps with a constant value so that they have a similar range as the density maps.

4.2.2.5 Evaluation in Test Phantoms

In test phantoms, we used MC with the phantom activity and density maps to calculate the ground truth dose-rate maps for performance evaluation. The estimated dose-rate maps generated from SPECT/CT with the DVK (with density scaling), MC (with 1 billion histories) and CNN methods were evaluated qualitatively by visual comparison of images, line profiles and dose-rate-volume histograms (DRVH) with those corresponding to the ground truth. For quantitative evaluations, we used the following metrics:

1. *Dose-rate error.* For each volume of interest (VOI), the dose-rate error is the absolute error across the whole VOI calculated relative to the ground truth. This error was calculated for the mean absorbed dose and DRVH statistics (DR10, DR30, DR70, DR90), corresponding to the minimum dose-rate to 10%, 30%, 70%, 90% of the VOI, respectively.

2. *Normalized root mean square error (NRMSE).*

3. *Ensemble noise.* The ensemble noise in spherical VOIs defined in non-tumoral liver or spleen was calculated across 3 (M=3) Poisson noise realizations as:

$$\text{Noise} = \frac{\sqrt{\frac{1}{n_p}\sum_{j\in\text{VOI}}\left(\frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{\boldsymbol{x}}_m[j]-\mu_j\right)^2\right)}}{\frac{1}{n_p}\sum_{j\in\text{VOI}}\mu_j}\times 100\%, \tag{4.3}$$

where

$$\mu_j \triangleq \frac{1}{M}\sum_{m=1}^{M}\hat{\boldsymbol{x}}_m[j], \tag{4.4}$$

where n_p is the total number of voxels in the VOI, $\hat{\boldsymbol{x}}_m[j]$ denotes the jth voxel in the estimated dose-rate image of the mth Poisson noise realization.

The lesion VOIs for these quantitative evaluations were defined manually on CT of SPECT/CT-guided by baseline diagnostic CT or MRI by a radiologist with abdomen imaging expertise. Organ contours were defined using semi-automatic CT segmentation tools. The healthy liver was defined as liver minus lesions in the liver.

4.2.3 Results

4.2.3.1 Virtual Patient Phantom Test Results

Qualitative Assessment. Generally, there was better visual agreement between CNN dose-rate maps and the ground-truth than between DVK/MC dose-rate maps and the ground-truth. The example images and line profiles in Fig. 4.10 and Fig. 4.11 and the DRVHs in Fig. 4.12 provide qualitative evidence of the superior performance of the CNN across multiple regions (kidney, abdominal lesion, lung lesion).

Quantitative Assessment. Fig. 4.13 compares the mean dose-rate error and NRMSE in lesions and organs across all test phantoms. Similar to the results of the qualitative assessment (Fig. 4.10, Fig. 4.11 and Fig. 4.12), the CNN also consistently showed superior results compared to DVK and MC in quantitative evaluations (Fig. 4.13). For instance, compared to DVK and MC, the CNN estimates showed an average improvement (in mean dose-rate error) of 52%/20%, 55%/53%, 66%/50%, 66%/62%, 48%/49% and 58%/39% in healthy liver, lesion, left kidney, right kidney, spleen and lumbar vertebra, respectively. The NRMSE was also substantially lower for the CNN than for DVK and MC across all VOIs (Fig. 4.13). The average improvement (in NRMSE) demonstrated by the CNN compared to DVK/MC was 10%/9%, 18%/17%, 11%/12%, 9%/10%, 26%/27% and 18%/10% in healthy liver, lesion, left kidney, right kidney, spleen and lumbar vertebra, respectively. In addition to the improvement in the average values, the maximum errors (denoted by the error bars in Fig. 4.13) were also consistently lower with CNN compared to DVK and MC. In Fig. 4.13, all three methods showed the highest errors for lesion and lumbar vertebra regions. This was attributed to the smaller size of these VOIs compared to other organs and the corresponding increase in partial volume effects. In the case of lumbar vertebra, relevant to bone marrow dosimetry, the very low uptake in these regions also contributed to higher dose-rate errors. For lesions and lumbar vertebra that had a relatively large sample size (15 and 18), a paired t-test demonstrated that the differences of mean dose-rate error and NRMSE between CNN and MC (DVK), as shown in Fig. 4.13, were statistically significant. Moreover, DRVHs statistics

(DR10, DR30, DR70, DR90) as demonstrated in Fig. 4.13 also shows the superiority of the CNN compared to DVK.

Noise Evaluation. Table 4.4 shows a consistent reduction of ensemble noise in background VOIs with an average of 21% and 27% improvement demonstrated by the CNN compared to DVK and MC (running 1 billion histories), where MC had the highest level of noise due to its statistical nature.

TBL 4.4 – Ensemble noise from 3 realizations for DVK, MC and CNN across all test phantoms. Number of voxels ranges from 2527 to 23411.

Ensemble Noise	Background Region	DVK	MC	CNN
Phantom #1	Liver & Spleen	4.6%	6.1%	**3.4%**
Phantom #2	Liver	12.6%	13.3%	**9.2%**
Phantom #3	Liver	14.0%	14.6%	**12.9%**
Phantom #4	Liver	20.3%	19.6%	**14.8%**
Phantom #5	Spleen	7.1%	7.6%	**5.8%**

4.2.3.2 Patient Results

In patients, where there was no known ground-truth, results were instead compared visually. Fig. 4.14 and Fig. 4.15 show examples of dose-rate maps corresponding to high count (day 1 post-therapy) and low-count (day 7 post-therapy) imaging conditions post-^{177}Lu DOTATATE. Although concrete conclusions could not be drawn, as there was no known ground truth; visual inspections implied potential reduction of SPECT spatial resolution effects on dose-rate accuracy by our DblurDoseNet. For instance, with the CNN, the enlarged kidney map and line profiles of Fig. 4.14 show a larger decrease in dose-rate in the medulla and renal pelvis areas, which could be due to the expected lower physiological ^{177}Lu uptake in this part of the kidney compared to the cortex region. In addition, in Fig. 4.15, the lesion with a necrotic center demonstrated a larger drop in dose-rate at the center with the CNN compared to DVK or MC, which could be due to the expected lower uptake associated with necrosis. Moreover, to demonstrate the generalizability of our CNN on patient data, we tested our CNN using 42 SPECT/CT scans of 12 patients and then compared with DVK and MC dose-rate maps in terms of the mean dose-rate and DRVH statistics (DR10, DR30, DR70 and DR90) across lesions and kidneys. As demonstrated in Table 4.5, there was a strong agreement between CNN and MC for mean dose-rate in kidneys; and for mean dose-rate in lesions, CNN showed higher values than MC, which could be partially due to the compensation of SPECT resolution effects. For DRVH statistics shown in

Fig. 4.12, the CNN and MC results also agreed well in kidney; but for lesions, the CNN measurement showed a lower dose-rate value in DR70 (DR90) and a higher dose-rate value in DR30 (DR10), compared to MC and DVK. The DRVHs in the lesions might be improved because the blurring effects caused by the limited SPECT camera resolution would lead to a higher DR70 (DR90) and a lower DR30 (DR10), but concrete conclusions could not be made due to the absence of ground truth.

TBL 4.5 – Dose-rate values (mean dose-rate and DRVH statistics) for DVK, MC and CNN methods averaged across all 42 scans from 12 patients. Minimum and maximum values are shown in parenthesis. The medians (ranges) for the VOI volumes are: lesion: 15 mL (2.3 mL – 582 mL); left kidney: 192 mL (105 mL – 275 mL); right kidney: 180 mL (122 mL – 259 mL). *: Reported dose-rates are normalized to 1 MBq in field-of-view.

		Dose-rate*(nGy/MBq-sec)		
		DVK	MC	CNN
Lesion	Mean dose-rate	13.7 (0.2 – 87.9)	13.9 (0.3 – 88.9)	14.4 (0.3 – 88.8)
	DR10	25.1 (0.4 - 177)	25.4 (0.4 - 179)	27.8 (0.5 - 188)
	DR30	15.8 (0.3 – 122)	16.1 (0.3 – 123)	16.9 (0.3 – 127)
	DR70	8.2 (0.2 – 48.1)	8.3 (0.3 – 48.5)	8.1 (0.3 – 48.8)
	DR90	5.4 (0.1 – 17.1)	5.6 (0.1 – 19.3)	5.0 (0.1 – 22.9)
Left kidney	Mean dose-rate	3.7 (0.7 – 8.6)	3.8 (0.7 – 8.7)	3.8 (0.7 – 8.3)
	DR10	6.0 (1.4 - 12.7)	6.1 (1.4 - 12.8)	5.8 (1.4 - 12.3)
	DR30	4.6 (1.0 – 10.7)	4.7 (1.0 – 11.0)	4.6 (1.0 – 10.7)
	DR70	2.7 (0.2 – 7.0)	2.7 (0.2 – 7.0)	2.8 (0.2 – 7.0)
	DR90	1.6 (0.1 – 3.8)	1.7 (0.1 – 3.8)	1.6 (0.1 – 3.6)
Right kidney	Mean dose-rate	4.2 (0.7 – 9.1)	4.3 (0.8 – 9.2)	4.2 (0.8 – 9.1)
	DR10	7.1 (1.5 - 17.2)	7.2 (1.6 - 17.2)	7.0 (1.8 - 15.2)
	DR30	5.2 (1.0 – 11.8)	5.3 (1.1 – 11.8)	5.3 (1.1 – 12.2)
	DR70	2.8 (0.2 – 7.3)	2.9 (0.2 – 7.3)	2.8 (0.2 – 7.7)
	DR90	1.6 (0.1 – 4.4)	1.7 (0.1 – 4.5)	1.5 (0.1 – 4.0)

4.2.3.3 Comparing Performance with a Non-residual Network and a 2D Network

To demonstrate the effectiveness of residual learning and the 3D convolutional feature extractor that we implemented, we also compared our proposed CNN with a CNN that had the same architecture but without residual learning (not adding the DVK dose-rate map to the output of 2D U-Net); and to a CNN without 3D feature extractor (a purely 2D U-Net where we treated the depth dimension of input as channels). The non-residual CNN and the 2D CNN were trained using the same hyper-parameters and the same training data as for the proposed CNN. All the testing used the same test phantoms demonstrated in the previous section. As shown in Table 4.6, quantitative comparisons across all test phantoms showed superior results of our proposed CNN (DblurDoseNet) for almost all

VOIs except for some cases where all the networks show comparable results. Based on these promising results, we believed the idea of residual learning was effective and it was beneficial to include a few 3D convolutional layers to extract 3D information rather than using only 2D convolutions.

TBL 4.6 – Mean (maximum) dose-rate error and NRMSE comparison between CNN with and without residual learning and with 2D and 3D networks evaluated across VOIs in all test phantoms.

	Mean Dose-rate Error			NRMSE		
	3D w/ res (DblurDoseNet)	3D w/o res	2D w/ res	3D w/ res (DblurDoseNet)	3D w/o res	2D w/ res
Healthy liver	1.4% (2.3%)	5.5% (7.0%)	**1.2% (3.1%)**	**19.6% (33.2%)**	21.6% (35.1%)	23.2% (33.3%)
Lesion	**5.3% (13.0%)**	6.9% (12.5%)	6.0% (13.9%)	**21.2% (32.5%)**	21.4% (31.5%)	21.8% (38.0%)
Liver	**1.9% (3.5%)**	5.7% (7.6%)	1.9% (4.8%)	**20.6% (26.3%)**	21.6% (27.6%)	22.8% (26.6%)
Left kidney	**0.9% (2.1%)**	5.2% (6.5%)	1.8% (3.8%)	19.2% (22.9%)	20.1% (22.0%)	**19.0% (20.8%)**
Right kidney	**1.8% (5.1%)**	5.8% (12.6%)	2.6% (7.5%)	**19.6% (21.5%)**	20.5% (24.3%)	20.0% (23.6%)
Spleen	2.5% (6.2%)	6.3% (9.5%)	**2.2% (6.4%)**	**13.1% (17.7%)**	14.4% (19.9%)	13.2% (18.2%)
Lumbar Vertebra	11.1% (27.4%)	**10.5% (27.2%)**	12.1% (30.6%)	33.0% (51.4%)	32.9% (49.1%)	**32.7% (50.2%)**

4.2.3.4 Time Cost

We compared the computation times of the different methods for generating a dose-rate map corresponding to the typical $512 \times 512 \times 130$ patient SPECT/CT image size on CPU (Intel Core i9 @2.3 GHz) or GPU (Tesla V100). DVK with density scaling took 20 seconds on the CPU and 10 seconds on the GPU. DPM MC code took 60 minutes simulating 1 billion histories (for both ground truth and test phantoms/patients) on the CPU while running DPM on a GPU is not an option at this time (we are unaware of any MC code for internal therapy running on a GPU). The CNN took about 20 minutes on the CPU and about 20 seconds using the GPU. After considering the DVK pre-computation time for the residual learning network, the total GPU time cost for the CNN with residual learning is about 20+10 seconds.

4.2.4 Discussion

Reliable voxel-level dosimetry requires reliable dose-rate images at multiple timepoints as well as dependable co-registration and fitting of the dose-rate vs. time data estimated at the voxel-level. Performing reliable voxel-level co-registration and fitting to generate dose maps can be challenging, but the feasibility has been demonstrated [94, 188]. In this work, we focused on generating reliable dose-rate maps. With evaluation both on virtual patient phantoms that covered clinically relevant conditions and patients who underwent ^{177}Lu DOTATATE therapy in our clinic, we demonstrated that our CNN using residual learning framework could provide fast and accurate dose-rate estimation. Despite using only moderate amount of training data, DburDoseNet provided consistently superior performance over conventional voxel dosimetry in terms of resolution, accuracy and noise across multiple regions including kidneys, lumbar vertebra and lesions in soft-tissue and lung. Importantly, for clinical implementation, the CNN voxel dose-rate map for a $512 \times 512 \times 130$ patient image could be generated in $\sim$30 seconds, which was a fraction of the time associated with running MC, the current gold standard. Although generating the ground-truth labels for training by MC was computationally expensive, this effort was needed only once at training time, for a given SPECT imaging system.

The main limitation to accurate voxel-level patient specific dose-rate estimation with non-learning-based methods is the poor spatial resolution associated with the input SPECT (or PET) images. This issue was evident in our results where the theoretically accurate MC-based calculation only slightly outperformed DVK with density scaling. In contrast, by using the true activity map-based dose-rate estimates for training, our CNN has the ability to "learn" the physics of dose deposition and to compensate for the SPECT resolution effects that both lead to blurring of the conventional (non-learning-based) dose-rate maps, as demonstrated in the phantom results (Fig. 4.10, Fig. 4.11, Fig. 4.12, Fig. 4.13 and Table 4.4). In patient studies, potential mitigation of SPECT resolution effects was demonstrated empirically. In Fig. 4.14, the CNN-based estimates show sharper line profiles and larger drops in dose-rate over the medulla area of the kidney, analogous to the illustration of Fig. 4.7. In Fig. 4.15, the larger drop of dose-rate in the necrotic center of a tumor, may reflect what is expected based on physiology. Although test results were promising over 42 scans originating from 12 patients, further testing is planned as more patient images become available. We did not investigate training with more virtual patients, because simulating ^{177}Lu SPECT projections by full MC simulation was computationally expensive. Furthermore, we found that our CNN, trained by 9 virtual patient phantoms, was able to generate promising dose-rate estimates across a diverse range of test cases. We expect that applying self/weakly-supervised training may address the computational inefficiency of simulating ^{177}Lu SPECT projections in the future. In addition, due to the lack of ground truth, we were unable to make concrete conclusions about the performance of our CNN on test patient data. But the uncertainty of our CNN can be quantified by generating confidence maps [125, 14, 243] using Bayesian networks [109], an ensemble of multiple networks [121], or an extension of the probabilistic U-Net [119], which can be one direction to investigate in the future.

The mean dose-rate errors shown in Fig. 4.13, especially for lesions, were generally lower than one would expect based on reported activity recovery in quantitative ^{177}Lu

SPECT phantom studies. For example, for 72 OSEM updates, activity recovery of only 80% was reported for a 26.5 mL volume "hot" sphere in a "warm" background region [214]. The results of the current study showed lower errors because, unlike in a physical phantom, the assigned "true" activity values at the boundary of the structures in our PET-based virtual patients did not drop off sharply, and instead, were blurred out. Moreover, in Fig. 4.13, all 3 methods showed the largest mean dose-rate error for lesions and lumbar vertebra, as expected due to the relatively smaller sizes of these structures compared to other organs, and hence partial volume effects associated with SPECT resolution were higher. The large error for the lumbar region with DVK ($\sim$25%) was likely to be due to the heterogenous tissue within this region, which includes cortical bone, trabecula bone, yellow and red marrow. Regarding DVK, the simple density scaling that was performed in our study was potentially inadequate for this region. Furthermore, the ^{177}Lu uptake in a lumbar region was very low, so the cross-dose contribution to dose-rate there, including the photon cross-dose, could be significant. Our $99 \times 99 \times 99$ photon kernel may have been insufficient to capture the full photon cross dose contribution to the lumbar vertebra. Our study did not include standard partial volume correction using volume-dependent recovery coefficients (RCs) because such methods provide only a mean dose, not a voxel-level correction. Furthermore, the limitations of standard RC methods due to dependence on object shape, activity distribution and target-to-background ratios are well known. Voxel-level partial volume correction is much more challenging [211] and their applications in SPECT are not well established. Our results demonstrated that training using true dose-rate maps could reduce the need for such corrections to compensate for resolution effects.

To define our virtual patient activity maps, we chose to use ^{68}Ga DOTATATE PET/CT to exploit the availability of these images that had finer resolution than SPECT and showed similar uptake patterns as ^{177}Lu DOTATATE. Despite the standard practice of using ^{68}Ga PET or ^{177}Lu SPECT as a theranostic pair, some differences between the two distributions were to be expected, but we did not expect this to impact our CNN performance because the PET images were used only to define the virtual patient phantoms and not in the training process itself, as proposed in another study [239]. Ideally, however, images of higher resolution than clinical PET should be considered as the true representation of the activity map of patients when generating the virtual patient training set, but usually they are not readily available. To circumvent this issue, we also investigated using phantoms with piece-wise uniform uptake in CT-defined organs/lesions for training (such as XCAT [189] in Fig. 4.7), but we found that such training led to unnaturally uniform dose-rate maps when tested on patient images. We also fed our CNN with an all-zero activity map, as a sanity check to our proposed framework as well as implementation. The output dose-rate map, as expected, was all zeros. This illustrates that if there is no apparent signal in the reconstructed SPECT, then there will not be any unexpected nonzero values in the dose-rate map. A possible alternative to our PET-based virtual patient activity maps is to assign distributions based on high-resolution animal models, for instance, ex-vivo autoradiography showing uptake distribution of DOTATATE in kidney [157].

Our results also demonstrate the advantage of residual learning framework exploiting the fast DVK approach as an initial estimate, which was not utilized in the prior studies [2, 128, 79, 78]. We also conjectured that incorporating residual learning could not only improve performance on the test data, but also accelerate the training process. Other than using a fast DVK approach for residual learning, an alternative was to generate a quick MC (low number of histories) estimate, which was not explored here. Another advantage of our network is that we first implemented a couple of depth feature extractor layers that shrink the 3D input into 2D at the beginning of our network. Compared to fully 3D approaches, this approach leads to a network having fewer parameters (because 2D kernels have fewer parameters than 3D kernels), so it is less likely to overfit the training data, avoiding a common problem in deep learning applications for medical imaging, where only moderate amount of training data is available. Another option that we did not investigate is to use 2.5D CNN architectures [254]. A potential drawback of our proposed CNN is possible discontinuity of pixel values in coronal slices; however, we did not observe such discontinuity, presumably due to the 11-slice sliding window.

We expect that training a single CNN, as we did in the current study, is simpler than training 2 separate CNNs to learn the dosimetry and SPECT resolution effects. Typically, there will be 3 stages needed to train 2 separate CNNs; stage 1: training CNN-A for SPECT resolution; stage 2: training CNN-B for dosimetry; stage 3: jointly fine-tuning CNN-A and CNN-B. Compared to our proposed end-to-end network (DblurDoseNet), which only involves one training stage, such 3-stage of training will be more complex and potentially inefficient. However, only through comprehensive comparisons can one draw definite conclusions between these two approaches, which we expect to undertake in the future. Although our study only investigated ^{177}Lu dosimetry, we expect that by changing the training dataset and making minor modifications to the architecture, our CNN approach can be extended to other radionuclides including ^{90}Y that is a pure-beta emitter and I-131 that has significant beta and gamma contributions to the dose-rate.

4.2.5 Conclusion

We constructed and tested a residual CNN that was trained on virtual patient phantom images to learn the mapping from SPECT/CT images to the corresponding dose-rate maps. We took the novel approach of using a single CNN to learn not only the dose-rate estimation but also to compensate for blurring of the dose-rate map due to poor SPECT resolution. Across multiple regions such as kidney, lumbar vertebra and lesions in both soft tissue and lung, the proposed residual DburDoseNet was able to outperform conventional voxel-level dosimetry methods, including the current "gold standard" MC, in terms of accuracy, noise and speed. Patient specific voxel-level dose rate maps can be generated in $\sim$30 secs on GPU; hence the CNN approach has much promise for real-time clinical use in radionuclide therapy dosimetry for treatment planning.

4.3 Efficient Super Resolution Network (ESR-Net) for SPECT Image Reconstruction

4.3.1 Motivation

One limitation of previous deep learning methods for SPECT image reconstruction is that they assume the training images have the same voxel size as the reconstructed images, which can be suboptimal when training images having finer voxel sizes (or higher resolution) are available. In reality, objects (e.g., patients) being scanned are continuous-space functions, whereas the reconstructed images are always digital and lie in a finite-dimensional vector space. This type of model mismatch has rarely been considered in supervised learning methods for image reconstruction problems, so such methods generally involve some type of "inverse crime" in their formulation and evaluation [104]. Directly applying conventional algorithms [42, 142, 156, 73, 74] to training and testing images with finer voxel sizes is conceptually straightforward, but can be very slow due to the heavy computations in forward and backward projections. We propose a novel method that can enhance the resolution of the reconstruction by training the regularizer using images having finer voxel sizes, while maintaining the computational efficiency by performing forward and backward projections using coarser voxels that are more suitable for the collimator resolution. We call the proposed **E**fficient **S**uper-**R**esolution network "ESR-Net". [133].

4.3.2 Methods

The key idea of ESR-Net is to let forward and backward projections work with coarse voxel sizes whereas regularization works with finer voxel sizes. We formulate image reconstruction as the following optimization problem:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \geq 0}{\arg\min}\, f(\boldsymbol{T}\boldsymbol{x}) + \beta R(\boldsymbol{x}), \quad f(\boldsymbol{v}) \triangleq \mathbf{1}'\left(\boldsymbol{A}\boldsymbol{v} + \bar{\boldsymbol{r}}\right) - \boldsymbol{y}' \log(\boldsymbol{A}\boldsymbol{v} + \bar{\boldsymbol{r}}), \tag{4.5}$$

where $\boldsymbol{A}$ denotes the SPECT system model, $\boldsymbol{y}$ denotes noisy measurements that are assumed to follow independent Poisson distributions. $\bar{\boldsymbol{r}}$ denotes the mean background events. $\boldsymbol{T}$ denotes a 3D downsampling matrix, and $\boldsymbol{x}$ denotes a finely sampled image. This work focuses on CNN-based regularizer $R(\boldsymbol{x})$. Starting from (4.5), we use an unfolded set of updates inspired by a block coordinate descent (BCD) algorithm [142], leading to the iteration update of the form

$$\begin{aligned} \boldsymbol{u}_{k+1} &= r_\theta(\boldsymbol{x}_k), \qquad (4.6) \\ \boldsymbol{x}_{k+1} &= \underset{\boldsymbol{x} \geq 0}{\arg\min}\, f(\boldsymbol{T}\boldsymbol{x}) + \frac{\beta}{2}\|\boldsymbol{x} - \boldsymbol{u}_{k+1}\|_2^2 \\ &\approx \frac{1}{2\beta}\left(\sqrt{h^2(\boldsymbol{u}_{k+1}) + 4\beta \boldsymbol{x}_k \odot e(\boldsymbol{x}_k)} - h(\boldsymbol{u}_{k+1})\right), \end{aligned}$$

where r_θ denotes a neural network with parameter θ, $\odot$ denote element-wise multiplication. (This is an alternating update, but it is not actually a descent, and we use a small number of outer iterations and do not focus on convergence in the limit as the number of iterations increase.) Subscript k denotes the iteration number. $h(\cdot)$ and $e(\cdot)$ are

$$\begin{aligned} h(\boldsymbol{u}_{k+1}) &\triangleq \boldsymbol{T}'\boldsymbol{A}'\mathbf{1} - \beta \boldsymbol{u}_{k+1}, \\ e(\boldsymbol{x}_k) &\triangleq \boldsymbol{T}'\boldsymbol{A}'\left(\boldsymbol{y} \oslash \left(\boldsymbol{A}\boldsymbol{T}\boldsymbol{x}_k + \bar{\boldsymbol{r}}\right)\right). \qquad (4.7) \end{aligned}$$

In (4.6), we ran one iteration of regularized EM algorithm to approximate the minimizer w.r.t. to $\boldsymbol{x}$, when $\boldsymbol{u}_{k+1}$ is held fixed, per each outer iteration, hence the "$\approx$" symbol. Fig. 4.16 shows the proposed ESR-Net architecture.

4.3.3 Results

We compared our proposed ESR-Net with the OSEM algorithm and BCD-Net [142]. Both OSEM and BCD-Net work with 4.8mm^3 voxel sizes ($128 \times 128 \times 80$ image size), whereas ESR-Net works with 1.6mm^3 voxels (and hence $384 \times 384 \times 240$ image size), so we resized the reconstructed images of OSEM and BCD-Net into 1.6mm^3 voxel size using trilinear interpolation before comparison. We trained BCD-Net and ESR-Net with the same regularization parameter $\beta = 0.1$ and the same CNN architecture; the only difference is that the BCD-Net was trained using activity maps of size $128 \times 128 \times 80$ that were downsampled from the original true activity maps. The regularized EM algorithm for BCD-Net and ESR-Net was implemented in Julia using the "SPECTrecon" package[5]. Fig. 4.17 shows that the proposed ESR-Net visually improves the reconstruction of a necrotic tumor as well as resolution of spleen over the OSEM and the BCD-Net significantly. For quantitative comparison, we used MAE and NRMSE as evaluation metrics. Table 4.7 and Table 4.8 show that the proposed ESR-Net consistently had the lowest NRMSE over all organs for both XCAT and virtual patient phantoms, compared to OSEM and BCD-Net. Results shown in "Lesion*" were averaged across all lesions.

TBL 4.7 – MAE and NRMSE of organs tested on XCAT phantoms.

Organ/MAE/NRMSE(%)	OSEM	BCD-Net	ESR-Net
Lesion*	14.8/37.4	17.4/36.1	**9.6/31.4**
Kidney	11.5/53.8	11.8/53.2	**11.3/47.3**
Liver	1.5/47.7	**1.1**/46.9	1.2/**41.7**
Spleen	3.4/42.2	3.7/41.3	**3.4/31.8**
Lung	9.3/53.3	9.2/53.4	**4.6/44.6**

TBL 4.8 – MAE and NRMSE of organs for virtual patient phantoms.

Organ/MAE/NRMSE(%)	OSEM	BCD-Net	ESR-Net
Lesion*	6.6/24.8	5.9/22.9	**5.0/22.8**
Kidney	6.4/24.6	4.9/22.6	**2.8/21.1**
Liver	6.8/25.6	6.6/24.8	**3.6/23.2**
Spleen	10.6/22.8	8.7/19.6	**4.6/16.6**

4.3.4 Conclusion and Future Work

We plan to extend the current phantom dataset to cover more realistic virtual patients, but one challenge is that these virtual patients images suffer from SPECT resolution effects

[5]Available at `https://github.com/JuliaImageRecon/SPECTrecon.jl`.

and hence are blurry. As the ESR-Net uses high-resolution images for training, we plan to deblur the OSEM reconstructed images using a modern SOTA blind image deblurring algorithm [226]. The original algorithm is implemented only for 2D images, and we plan to implement the 3D version and then compare with the original 2D version (deblur slice by slice). Then we plan to generalize the "ESR-Net" to other radionuclides. We also will think about how to adapt this work to reduce the "inverse crime" aspect of supervised learning methods for image reconstruction.

4.4 Sparse View SPECT Image Reconstruction with Neural Radiance Field Representation (Sperf)

4.4.1 Motivation

SPECT acquisition under low-count conditions such as encountered in Lu-177 imaging can take 20-30 mins per field-of-view with standard systems, which is particularly challenging when multiple bed positions are needed. The goal of this work is to reduce SPECT acquisition time by only acquiring a subset of total projections (e.g., 30 of 120) and then use a neural network with single-scan self-supervised coordinate-based learning to predict missing projections before reconstruction.

Neural radiance field (NERF) [160] is one of the SOTA methods for 2D view synthesis of 3D objects. It takes a low-dimensional coordinate vector as input to the neural network (multilayer perceptron) and predicts volume density and view-dependent emitted radiance at that spatial location. Then it uses classic volume rendering techniques to project the neural network's outputs into an image. NERF was successfully applied to predict missing projection views in CT image reconstruction [208].

Current deep learning method for SPECT projection view synthesis is to train a 3D CNN to map from sampled projections to missing projections using supervised learning [185]. That method was reported to be effective but can be computationally expensive at both training and testing phase, and it requires a very large amount of training data to have generalizability to unseen data.

4.4.2 Methods

To train our network, we input the coordinates of 30 projections views (retrospectively under-sampling of the standard 120 views) from ^{177}Lu SPECT/CT. The coordinates are 5 dimensional: pixel position (i, j), sine and cosine of the view angle and radial position. The training was performed by self-supervised learning from a single patient scan; the training target was the corresponding measured total projection counts and triple energy window scatter estimates in those views, and hence does not need separate training data. During testing, we input the coordinates of the missing 90 projection views and the output was the corresponding predicted counts (Fig. 4.18). The intuition of this training framework was to mimic traditional interpolation methods but with a data-driven interpolation strategy. Our network was a fully connected network with 12 layers and 256 neurons for each layer. It was trained for 300 epochs by minimizing mean square error loss with Adam optimizer. The training/testing dataset is SPECT/CT projection data corresponding to 8 patients who were imaged (single or multi-bed positions with 18-25 mins/bed) following ^{177}Lu DOTATATE and ^{177}Lu PSMA radiopharmaceutical therapies. Reconstructions were done by running 50 iterations of MLEM algorithm including depth dependent resolution recovery. Lesions and normal organs segmented on CT were applied to co-registered SPECT reconstructions. Performance was evaluated by calculating the mean difference and normalized root mean square error (NRMSE) in counts relative to the reconstruction with all 120 views (reference). We compared our method with traditional (non-learning) based linear interpolation for generating the missing views.

4.4.3 Results

4.4.3.1 Synthesized Projections

We evaluated the accuracy of synthesized projections using different methods, including linear interpolation, supervised learning ([185]), and unsupervised learning (our proposed method). Our evaluation was conducted on 9 scans from 3 patients who were undertaking ^{177}Lu 177 DOTATATE radiation therapy. The SPECT scans were taken from the first day of therapy to a week into the treatment.

Fig. 4.19 and Table 4.9 showed that our proposed method produced the most accurate synthesized projections, closely matching the ground truth data (measured projections). The performance of supervised learning is hindered by the limited amount of training data available, leading to insufficient training for the U-Net model [185]. In contrast, our proposed unsupervised method is zero-shot learning and hence does not require any training data.

4.4.3.2 Reconstruction

We used a ^{177}Lu VP phantom of size $128 \times 128 \times 80$ for training and testing. We ran SIMIND MC program to simulate 128 projection views. The network was trained by taking

TBL 4.9 – Comparing synthesized projections using linear interpolation, supervised learning [185] and our proposed method. Results were based on 9 patient scans.

Method	Linear Interpolation	U-Net [185]	Ours
NRMSE (%)	14.0±3.0	17.3±1.9	**11.1±2.3**

the 3D coordinates of 64 sampled projections as input and the target was the projection values. After training, the network predicts missing projection view values by inputting the coordinates of the missing projections. Hence, the training procedure is target-specific, *i.e.*, one needs to retrain the network if the target is changed. However, NERF is very shallow (just a few linear layers) so that retraining can be done at acceptable run-time cost.

We compared OSEM with 1) 128 projections; 2) 64 projections (sampled uniformly from 128 projections); 3) 64 sampled projections plus 64 synthesized projections from NERF. Fig. 4.20 shows that the proposed method ("osem-NERF") visually looks less noisy than OSEM with 64 views and also has lower NRMSE at most slices (subfigure (a,c,d)). However, we found that one problem with our method is the suppression of the maximum values, *e.g.*, in Fig. 4.20 (b), the maximum activity value at lesion is noticeably lower than other methods, which leads to worse NRMSE.

4.4.3.3 Reducing number of views vs. reducing counts per view

An interesting experiment is to compare reducing number of projections views vs. reducing number of counts per view. We did an experiment on an XCAT phantom. First we obtained 128 projection views by running a SIMIND MC simulation with a billion history. We then scaled the total projection counts to reflect realistic count levels in SPECT scans, specifically 4 million and 40 million counts. Next, the OSEM reconstruction was performed with 16 iterations and 4 subsets. The resulting image from this process was used as the reference for subsequent comparisons.

To simulate low-count reconstruction, the number of total projection counts was scaled down to one fourth of the original count (4 million to 1 million, 40 million to 10 million) using various methods: 128 views with one fourth of original counts per view, 64 views with half of original counts per view, 32 views with original counts per view, 16 views with two times original counts per view, 8 views with four times original counts per view. We ran 16 iterations, 4 subsets OSEM algorithm for each low-count reconstruction. The experiment was repeated 5 times with different random seeds for simulating Poisson noise.

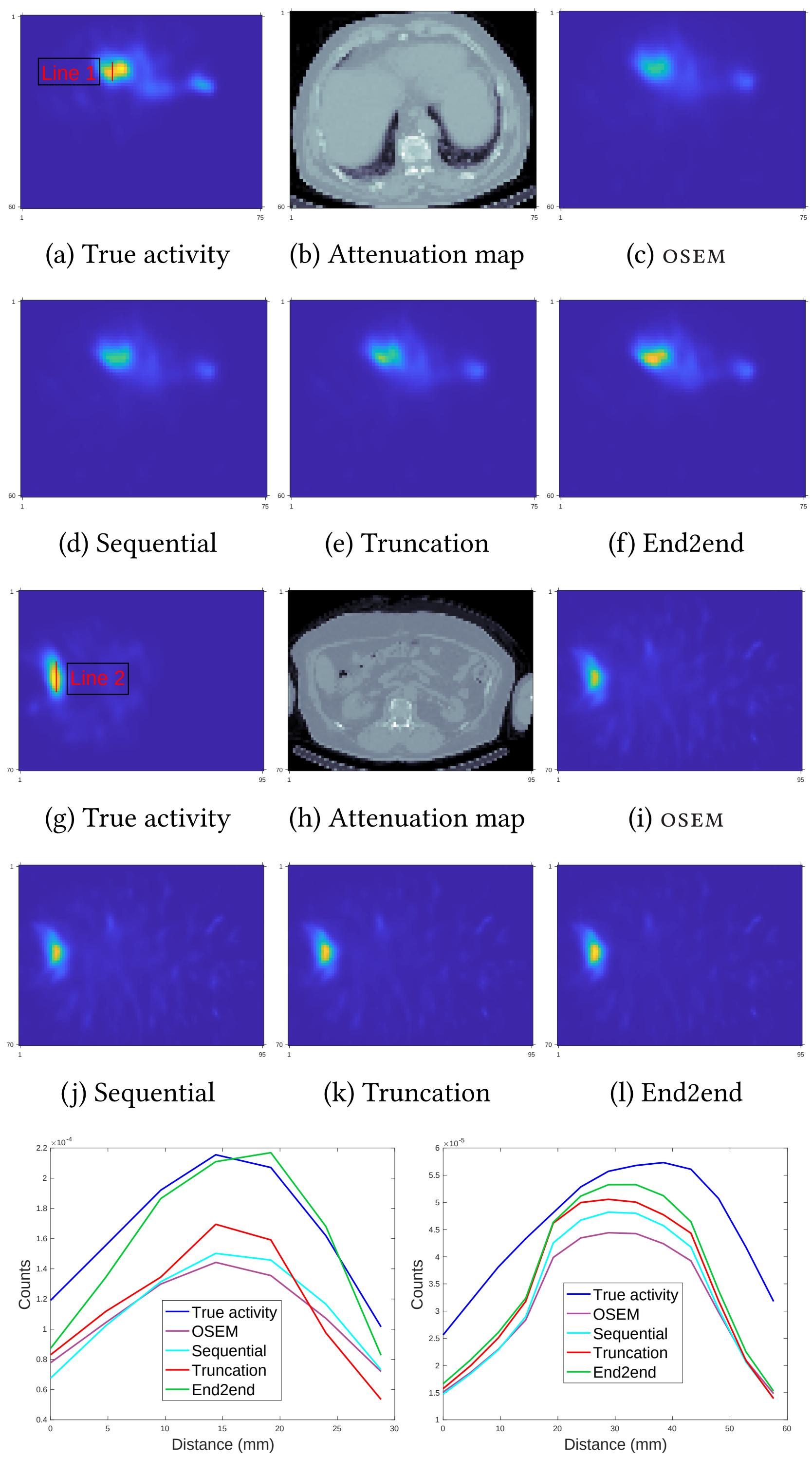
Fig. 4.21 and Table 4.10 compare the OSEM reconstruction quality using different methods to reduce counts. For 16, 32, 64, or 128 views, similar reconstruction quality was achieved both qualitatively and quantitatively. However, when using only 8 views, artifacts were observed along with significantly higher NRMSE values. This demonstrates that both reducing the number of views and decreasing the counts per view have similar effects on the quality of the reconstructed images, except in extreme situations (such as just having 8 views).

TBL 4.10 – Comparing reducing number of projection views vs. reducing number of counts per view. Results (in NRMSE%) were averaged on five random seeds.

Count level/NRMSE(%)	8 views	16 views	32 views	64 views	128 views
1 million	23.5±0.4	19.9±0.5	18.9±0.5	18.4±0.2	18.8±0.1
10 million	16.5±0.2	7.4±0.1	6.3±0.1	5.9±0.1	6.0±0.1

4.4.4 Conclusion and Future Work

This study demonstrated the potential effectiveness of coordinate-based learning for predicting SPECT projections. Our proposed method led to improved reconstructions both qualitatively and quantitatively from only a subset of measured projections, potentially enabling reduction in whole-body ^{177}Lu SPECT acquisitions time. Future directions include exploring newer versions of NERF that embed the input positions with different methods, *e.g.*, Fourier embedding [166, 12, 209, 110]. We also expect to further evaluate our method on an extended dataset, and compare with other machine learning methods for predicting SPECT projections.



(a) True activity (b) Attenuation map (c) OSEM

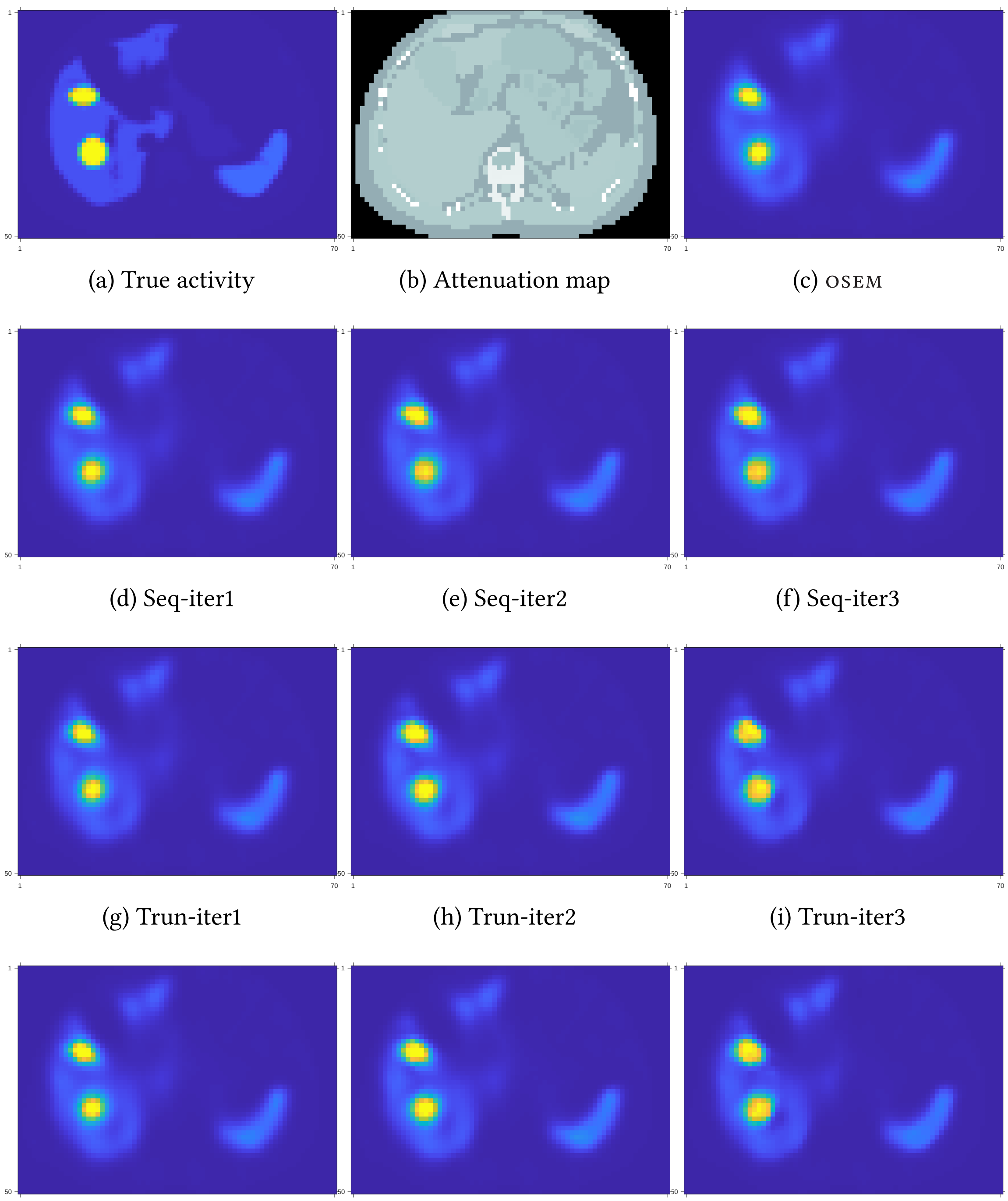
(d) Sequential (e) Truncation (f) End2end

(g) True activity (h) Attenuation map (i) OSEM

(j) Sequential (k) Truncation (l) End2end

(m) Line 1 profile (n) Line 2 profile

FIG 4.5 – Qualitative comparison of different training methods and OSEM tested on ^{90}Y VP phantoms. Subfigure (a)-(f) and (g)-(l) show two slices from two testing phantoms. Subfigure (m) and (n) correspond to line profiles in (a) and (g), respectively.



(a) True activity (b) Attenuation map (c) OSEM

(d) Seq-iter1 (e) Seq-iter2 (f) Seq-iter3

(g) Trun-iter1 (h) Trun-iter2 (i) Trun-iter3

(j) End-iter1 (k) End-iter2 (l) End-iter3

FIG 4.6 – Visualization of intermediate iteration results of different training methods. Subfigure (d)-(f): sequential training; (g)-(i): gradient truncation; (j)-(l): end-to-end training.

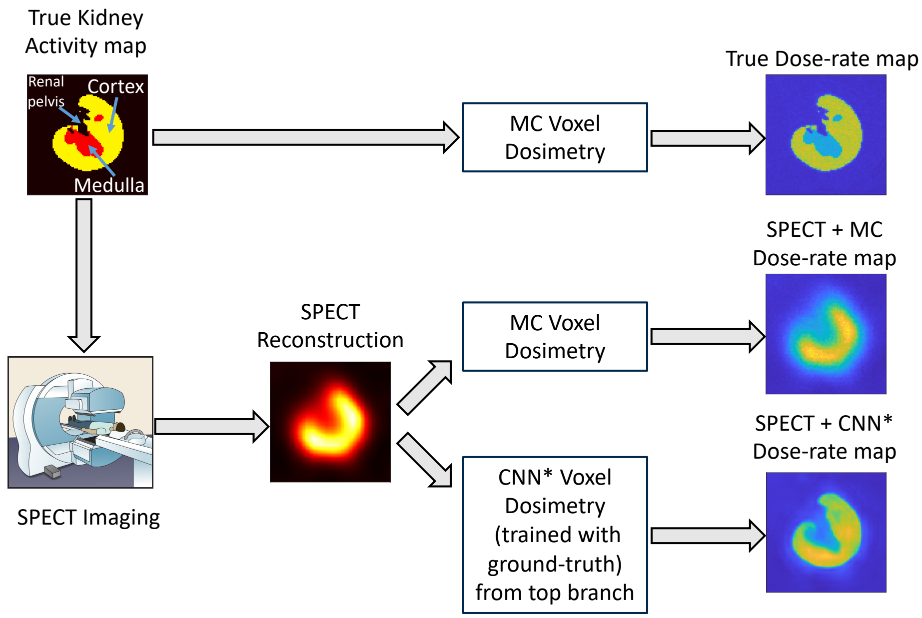


FIG 4.7 – Illustration of blurring of dose-rate maps due to the limited resolution of the SPECT-based input activity map and the potential for a learning-based method to outperform MC, the current gold-standard. The CNN* used in this illustration was trained and tested on different XCAT phantoms [189].

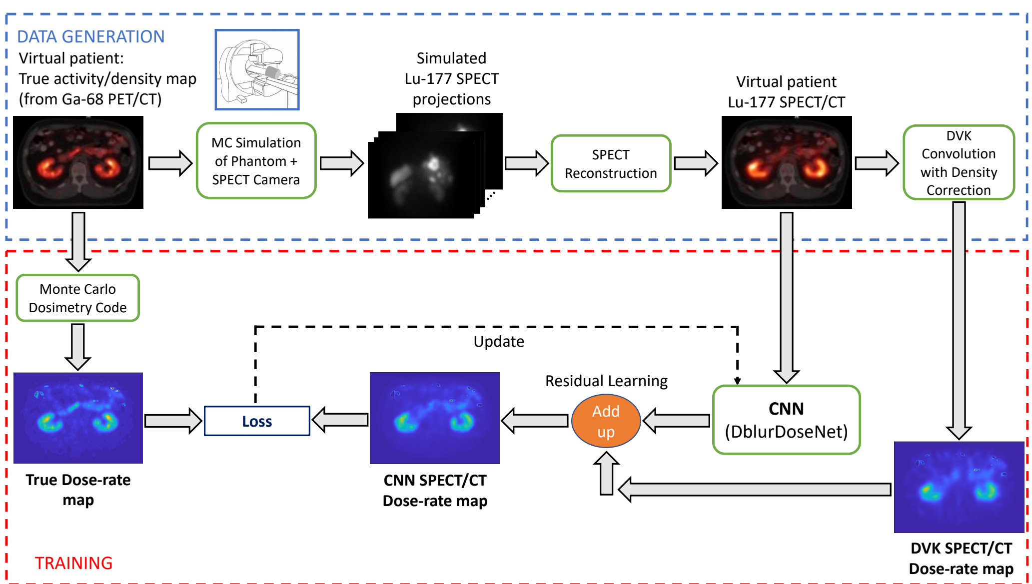


FIG 4.8 – Overview of phantom data generation for training/testing and the network training process.

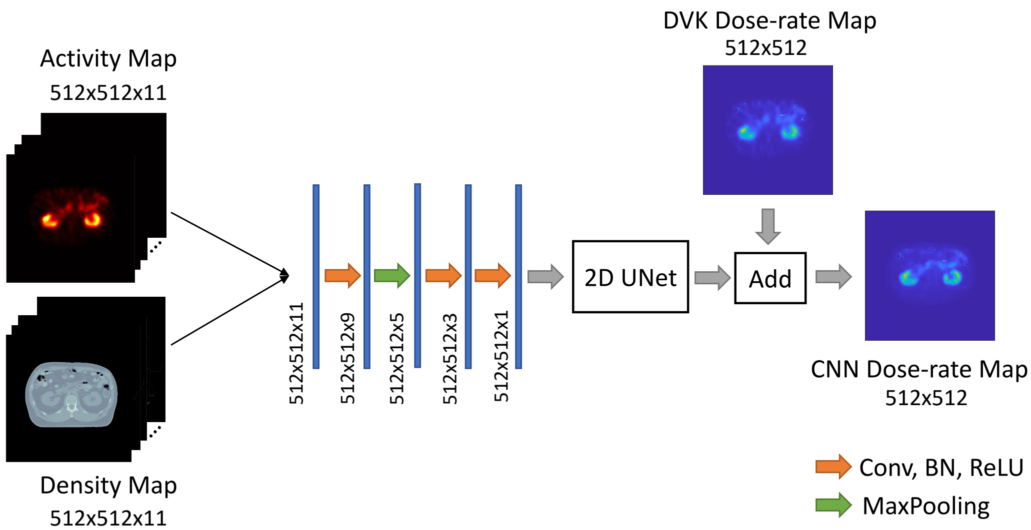


FIG 4.9 – The architecture of our DblurDoseNet.

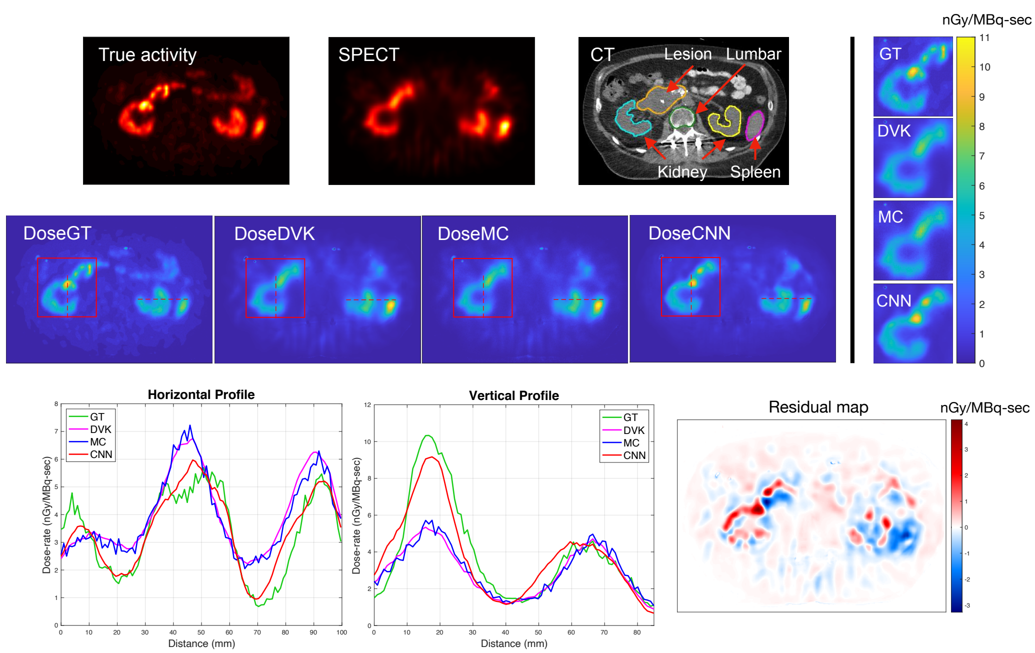


FIG 4.10 – One slice of the test virtual patient phantom #2. The top two branches show the true activity map defined based on ^{68}GA PET, SPECT and CT images, the ground truth dose-rate map and the dose-rate images from the different methods (DVK, MC, CNN). The bottom branch shows line profiles across the kidney and the residual map (the difference between CNN and DVK dose-rate map). The dose-rate units were normalized to 1 MBq in the field-of-view in all figures.

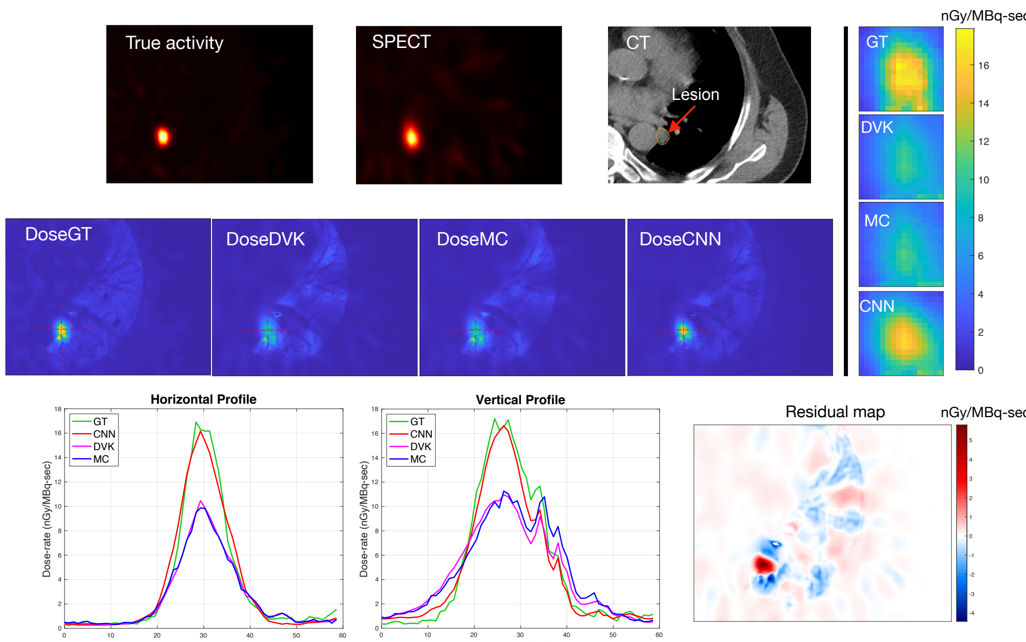


FIG 4.11 – One slice of the test virtual patient phantom #5. The top two branches show the true activity map defined based on ^{68}Ga PET, SPECT and CT images, the ground truth dose-rate map and the dose-rate images from the different methods (DVK, MC, CNN). The bottom branch shows line profiles across the kidney and the residual map (the difference between the CNN and DVK dose-rate maps).

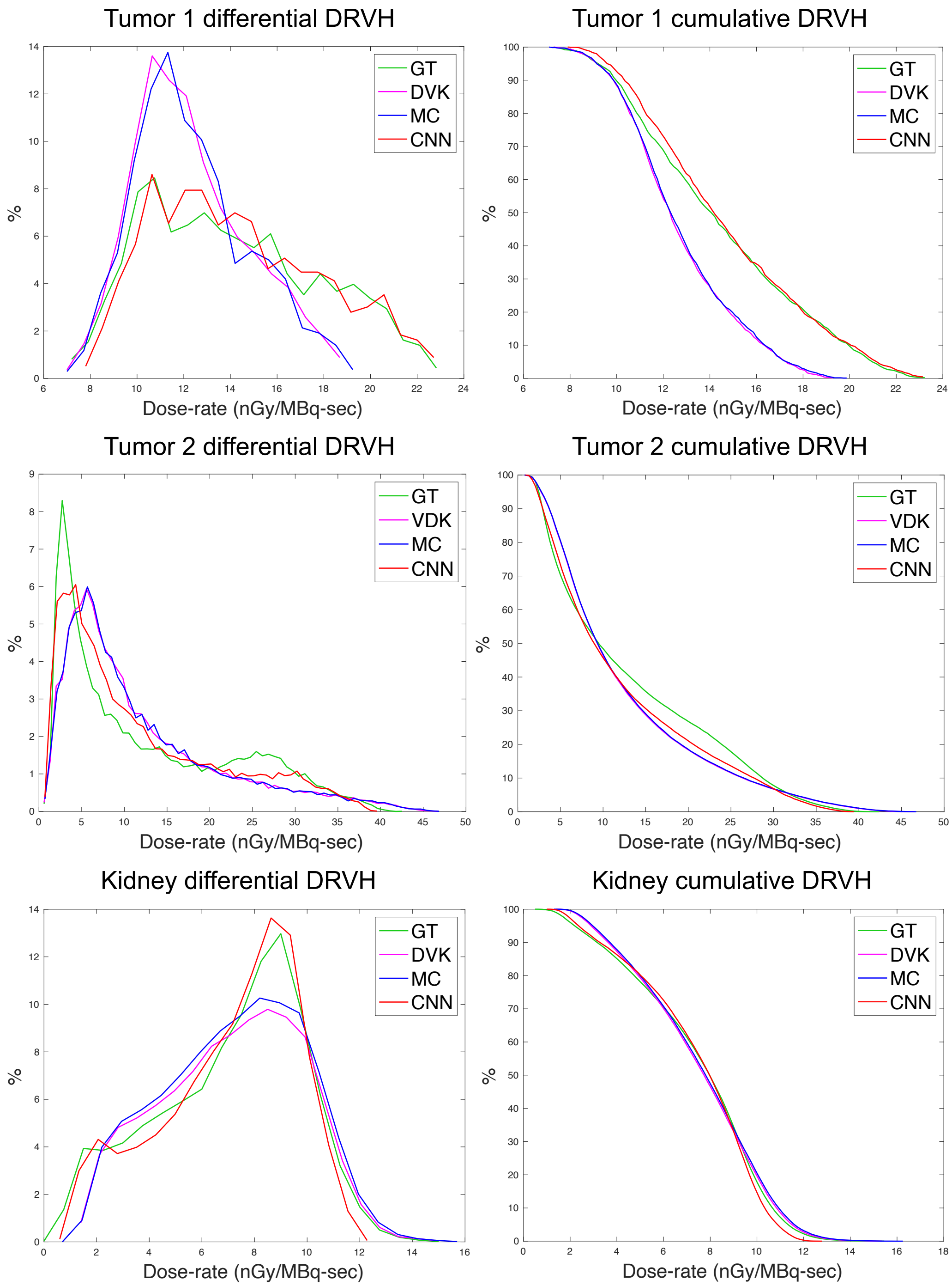


FIG 4.12 – Tumor & kidney differential and cumulative dose-rate volume histograms corresponding to DVK, MC, CNN and the ground-truth dose-rate maps of virtual patient phantoms. The sizes of tumor 1 and tumor 2 are 4mL and 65 mL, respectively.

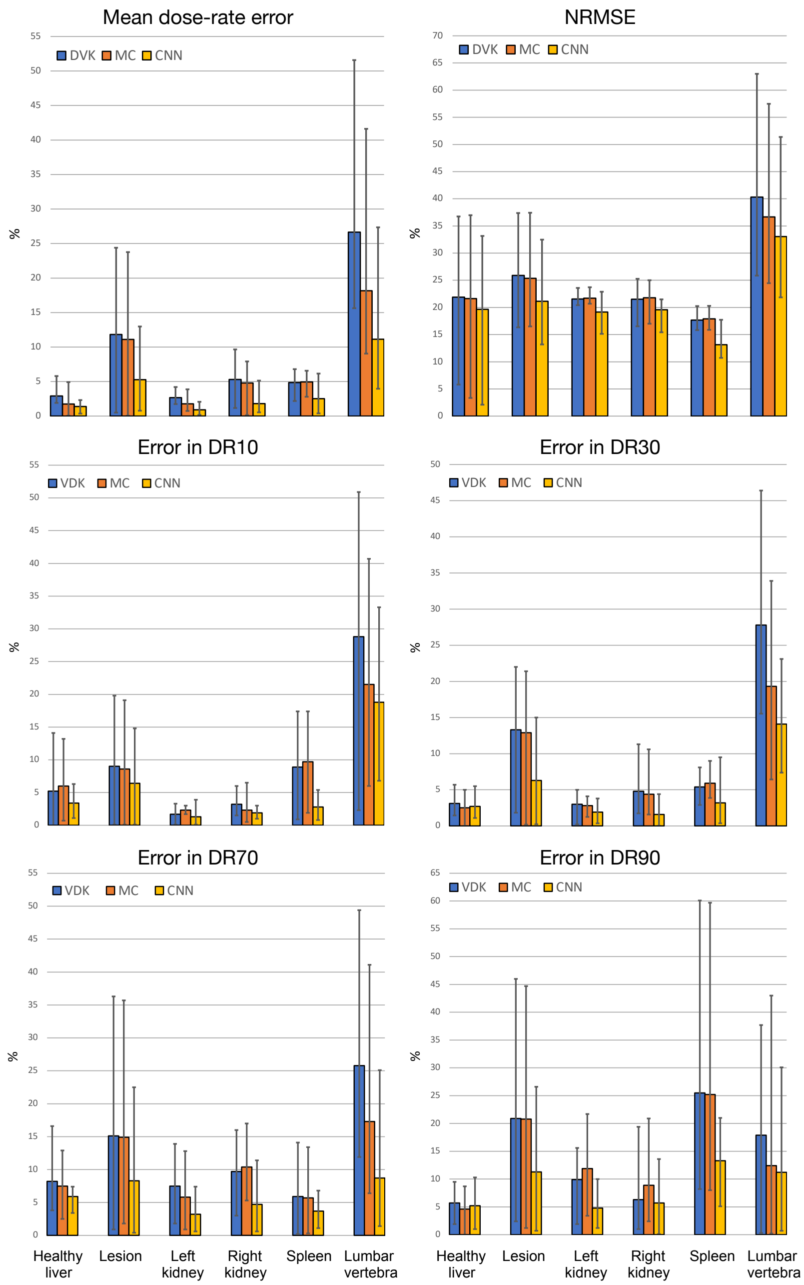


FIG 4.13 – Mean dose-rate error, NRMSE and error in DRVH statistics (DR10, DR30, DR70, DR90) comparison of DVK, MC and CNN relative to ground-truth dose-rate map across all test phantoms. Median (range) VOI volumes are: healthy liver (liver minus lesions): 1607mL (1164mL – 2262mL); lesion: 16mL (4mL – 181mL); Left kidney: 177mL (98mL – 211mL); Right kidney: 156mL (76mL – 249mL); Spleen: 191mL (131mL – 467mL); Lumbar vertebra L2 to L5: 54mL (34mL – 68mL).

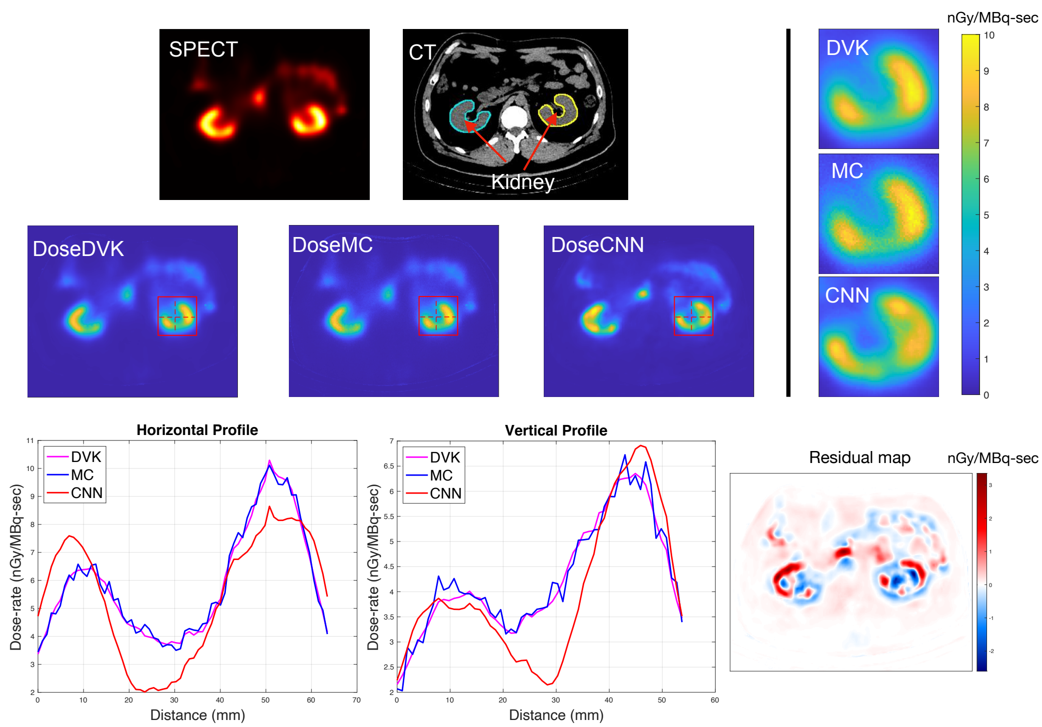


FIG 4.14 – One slice across kidney of the input images (SPECT, CT) and output DVK, CT, CNN dose-rate maps and line profiles for a patient imaged after ^{177}Lu DOTATATE (at day 1 post-therapy). The residual map is the difference between CNN and DVK dose-rate map.

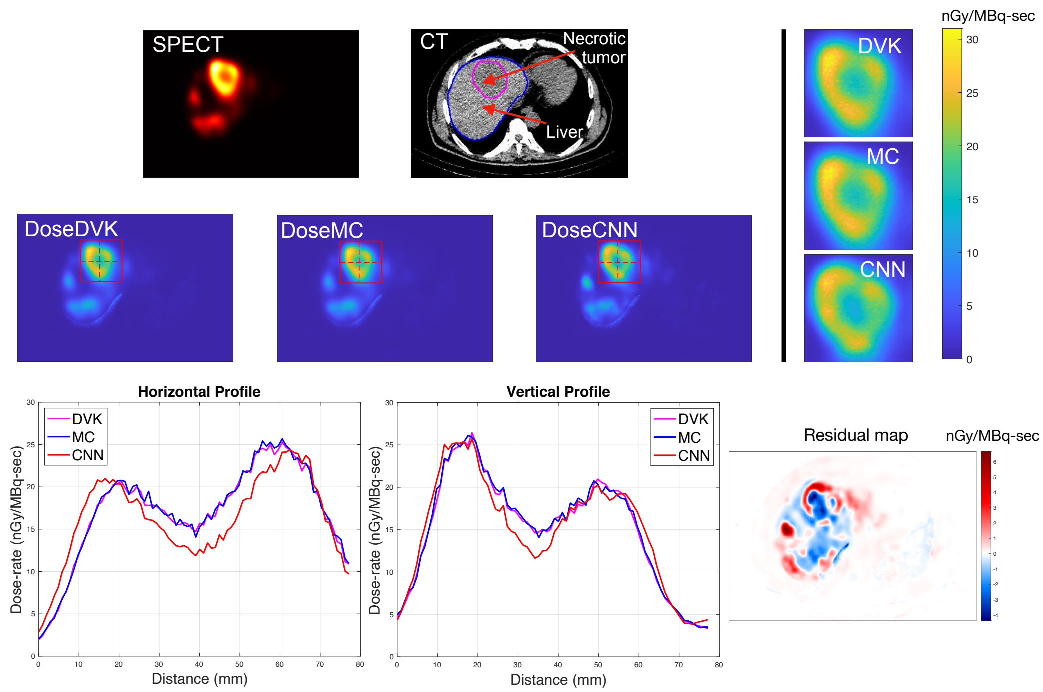


FIG 4.15 – One slice across lesion of the input images (SPECT, CT) and output DVK, MC, CNN dose-rate maps and line profiles for a patient imaged after ^{177}Lu DOTATATE (at day 7 post-therapy). The residual map is the difference between CNN and DVK dose-rate map.

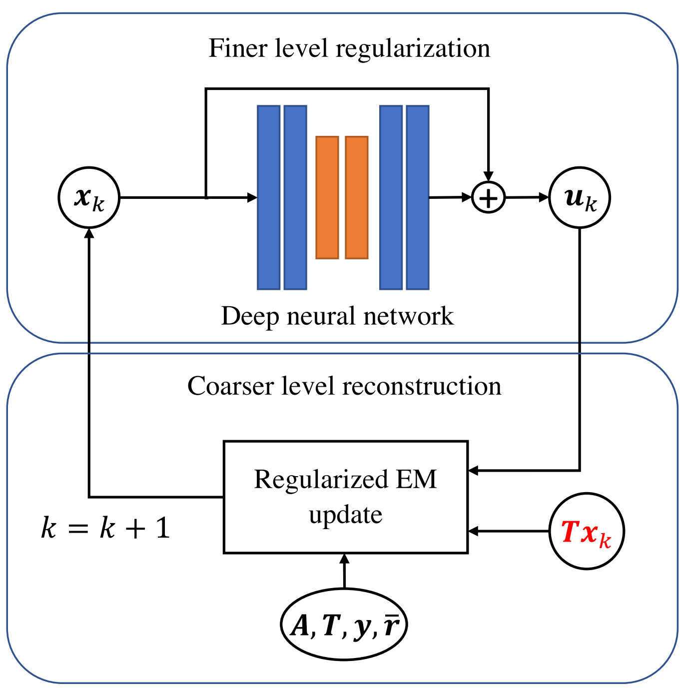
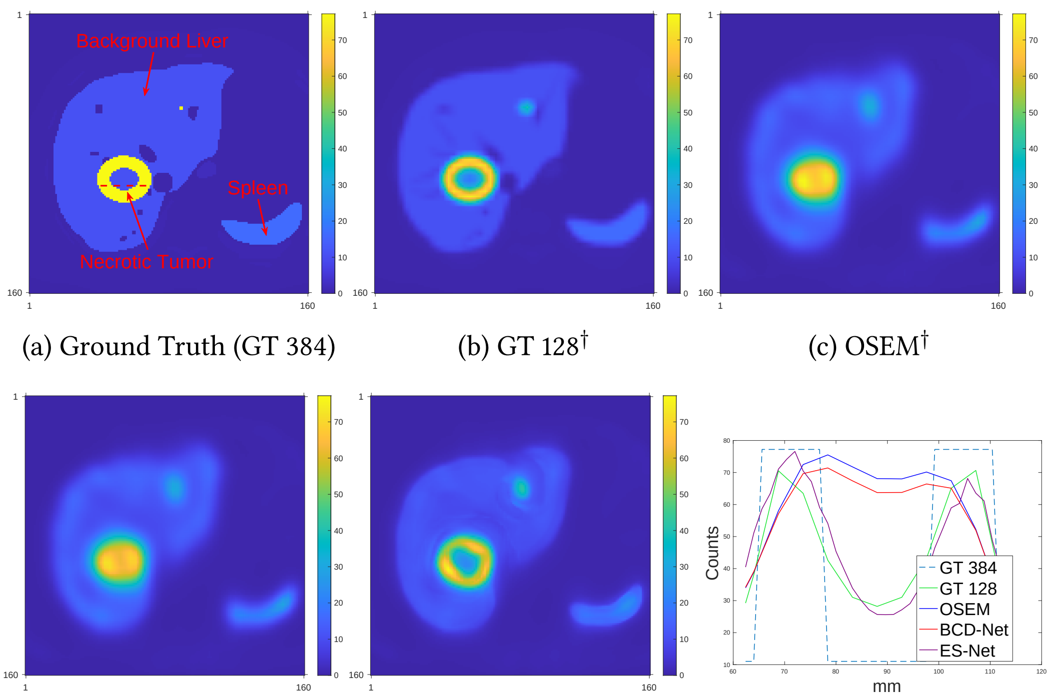


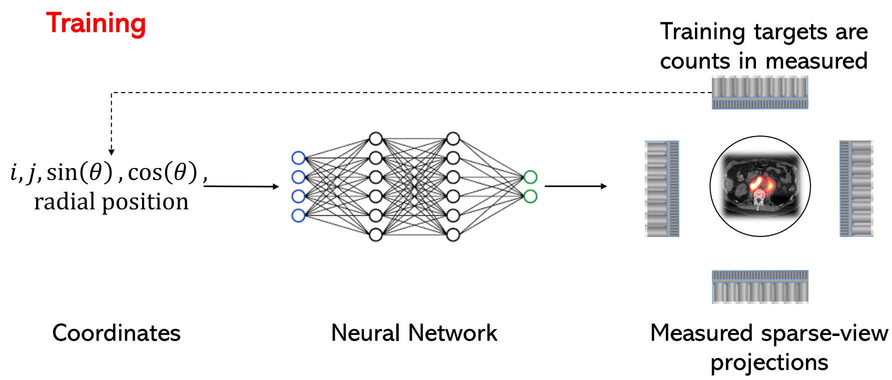
FIG 4.16 – Architecture of proposed ESR-Net.



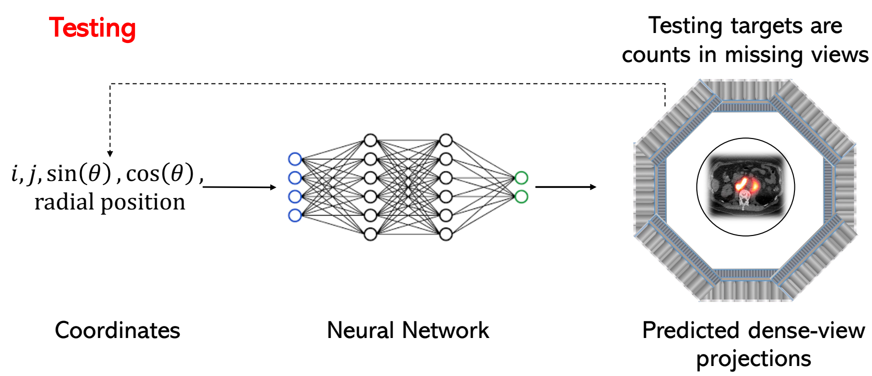
(a) Ground Truth (GT 384) (b) GT 128† (c) OSEM†

(d) BCD-Net† [142] (e) ESR-Net (proposed) (f) Line profile

FIG 4.17 – Qualitative comparison of different methods on test XCAT phantom. † denotes after interpolation (image size 128×128×80→384×384×240 with voxel size $4.8\text{mm}^3 \rightarrow 1.6\text{mm}^3$). Subfigure (f) shows the line profile over a necrotic tumor.



(a) Training framework.



(b) Testing framework.

FIG 4.18 – Training and testing frameworks of our proposed unsupervised coordinate learning method.

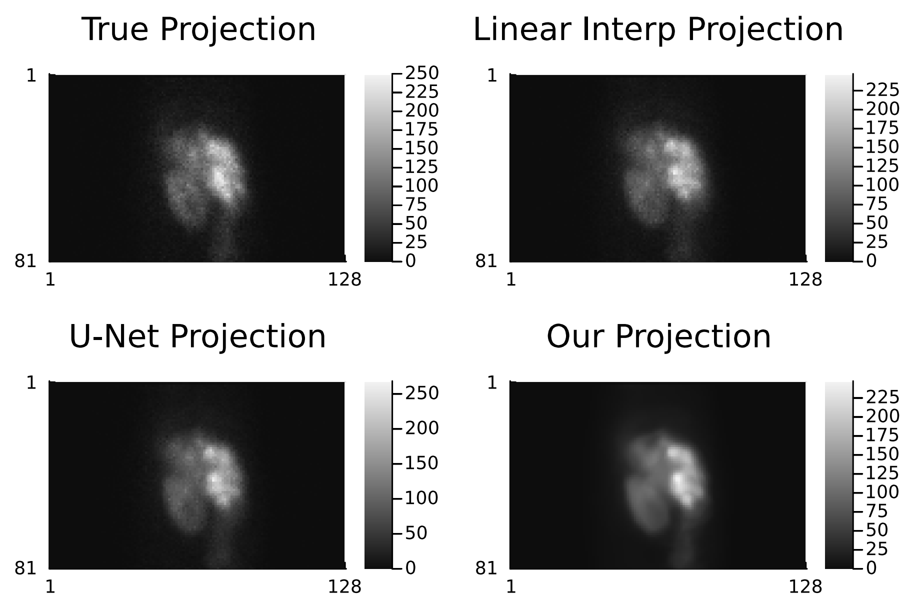


FIG 4.19 – Comparison of synthsized projections by linear interpolation (NRMSE=18.3%), U-Net [185] (NRMSE=18.6%), and our method (NRMSE=13.9%).

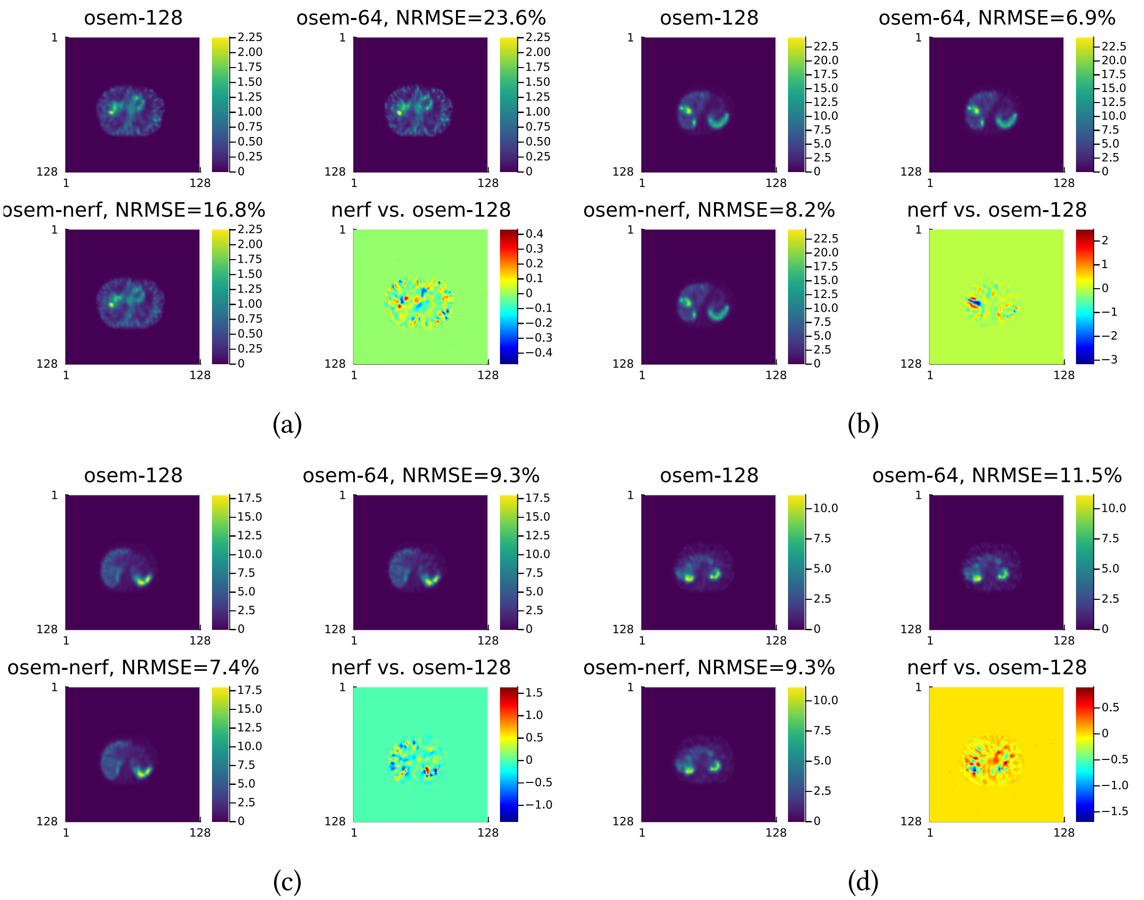
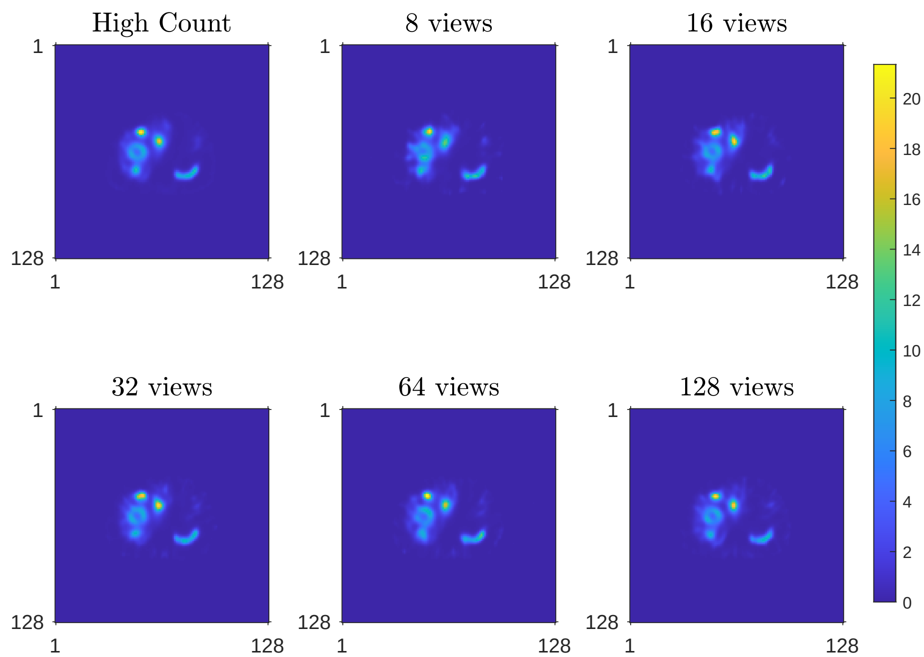
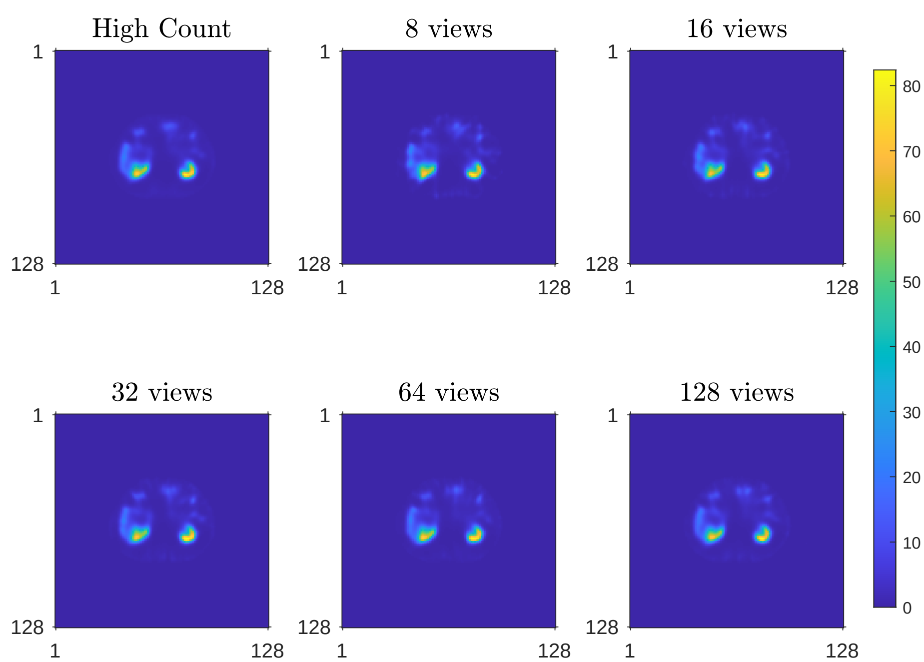


FIG 4.20 – Comparison of OSEM reconstruction results. “osem-128” denotes OSEM with 128 views; “osem-64” denotes OSEM with 64 views; “osem-NERF” denotes OSEM with 64 sampled views and 64 synthesized views from NERF.



(a) Count level is 1 million.



(b) Count level is 10 million.

FIG 4.21 – Comparing reducing number of projection views vs. reducing number of counts per view. The high count reconstructions have 4 times higher total number of counts and 128 projections.

CHAPTER 5

Discussion and Future Works

Poisson inverse problems are an intriguing and flexible category of mathematical and computational difficulties that have a wide range of applications in science and engineering. These problems revolve around the task of reconstructing unknown parameters or functions based on measured data that adhere to Poisson MLE. The allure of Poisson inverse problems lies in their capacity to reveal concealed information from noisy or incomplete observations, thus making them immensely valuable in various fields including medical imaging, environmental science, materials characterization, and astrophysics. This thesis focuses on two specific applications of Poisson inverse problems, namely phase retrieval and SPECT imaging. Thus far, we have presented several effective algorithms for resolving these types of Poisson inverse problems. This chapter discusses the challenges associated with these applications and explores potential future directions that can be investigated based on the research findings presented in this PhD thesis thus far.

5.1 Learning on "SmArge" data

A significant portion of my research focuses on the challenges associated with working with large 3D data when limited training data is available. Thus, the name "SmArge" originated from a combination of the words "small" and "large", representing the concept of working with a limited amount of data in extensive size. This section puts forth several ideas for training machine learning algorithms on such "SmArge" data.

5.1.1 Transfer Learning

Transfer learning is a technique that enables the application of knowledge gained from one problem domain to another related domain [227]. It involves leveraging pre-trained models or learned representations and adapting them to new tasks, which can significantly reduce the need for large amounts of labeled data in the target task. By using transfer learning as illustrated in Fig. 5.1, we can make more efficient use of pretrained large foundation models and achieve high performance in various domains.

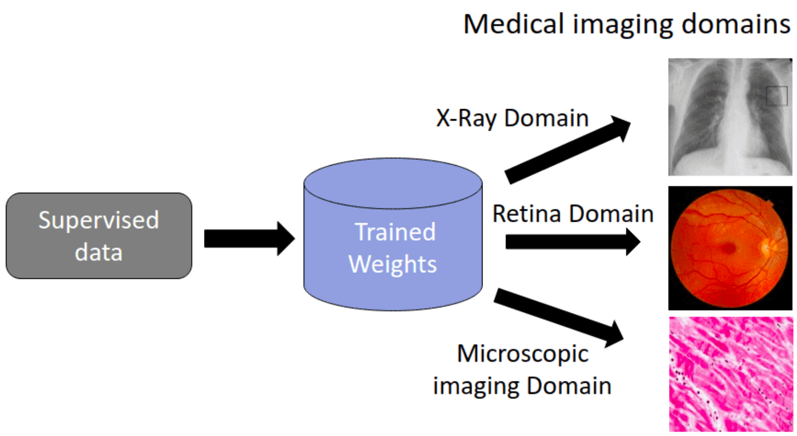


FIG 5.1 – Illustration of transfer learning on medical images. Figure adopted from `https://theaisummer.com/medical-imaging-transfer-learning/`.

For example, one can use finetuned Segment Anything Model (SAM) [117] for SPECT tumor segmentation. The SAM is a promptable method that allows user to either draw a bounding box or a point inside the object to be segmented and hence allow transfer learning to new image segmentation tasks. SAM was pretrained on over 1 billion masks of 11 million images, and was reported to have often competitive with or even superior performance compared to prior methods. Fig. 5.2 demonstrates the architecture of foundation model for segmentation (SAM). It has three inter-connected components: a promptable segmentation task, a segmentation model (SAM), and a data engine for collecting dataset of over 1 billion masks.

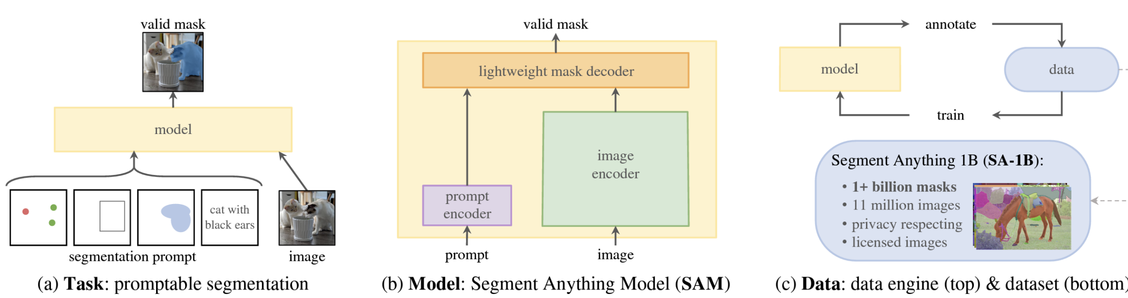


FIG 5.2 – The architecture of SAM. Figure is adopted from [117].

We did a preliminary test of finetuned SAM for tumor segmentation on SPECT images, and it achieved promising results. Fig. 5.3 shows that the finetuned SAM outperformed the supervised U-Net approach, as well as the thresholding approach (the bottom row of Fig. 5.3), which is an empirical method that selects pixels that is larger than the threshold in a user-specified region.

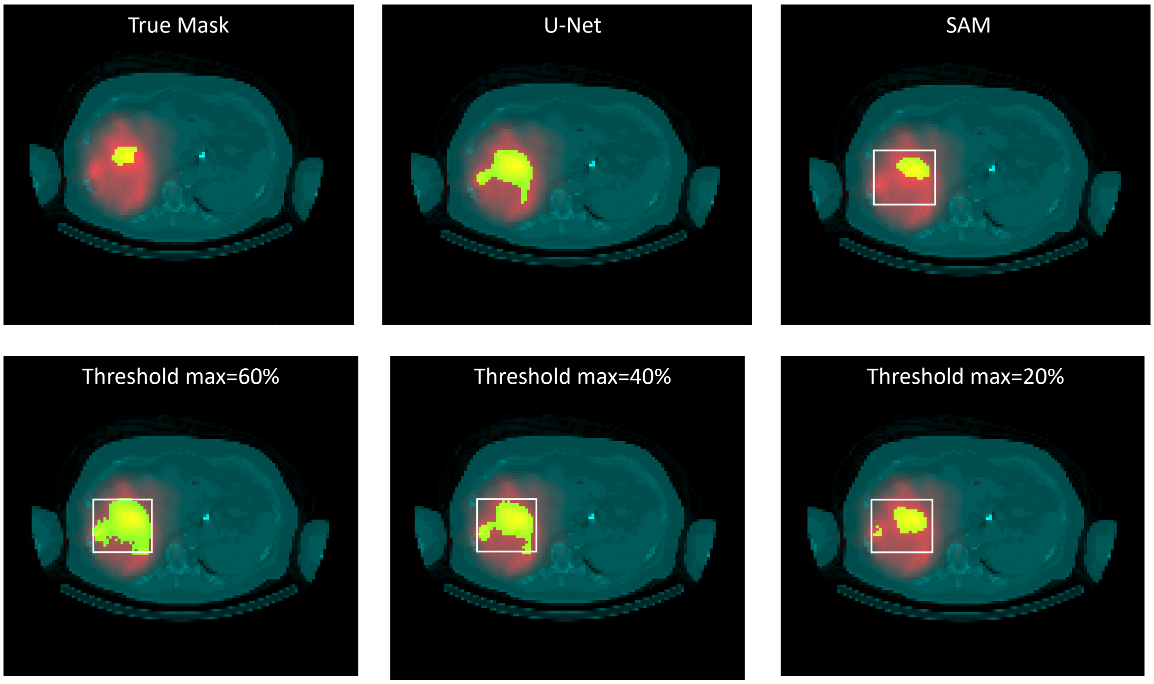


FIG 5.3 – Comparison of different tumor segmentation methods on SPECT images.

However, a challenge in transfer learning is that the majority of pre-trained foundational models are designed to work with 2D data, while medical images predominantly consist of 3D data [36]. Therefore, one potential future work is to investigate and develop transfer learning techniques specifically designed for 3D medical imaging data. If such models are not readily available, one can adapt existing 2D transfer learning techniques to work with 3D medical data in a slice by slice fashion.

5.1.2 Unsupervised Learning

Unsupervised learning is another option for dealing with limited 3D training data [179]. It involves training models on unlabeled data, allowing them to uncover patterns and structures without the need for explicit supervision signals. This approach has gained attention due to its potential usefulness in scenarios where labeled data is scarce or expensive to obtain. Several techniques have been developed in unsupervised learning, such as clustering algorithms and autoencoders, which aim to discover hidden relationships within the data.

These methods have yielded promising results across various domains including natural language processing, computer vision, and anomaly detection.

For SPECT imaging, one can apply unsupervised learning to scatter correction (illustrated in Fig. 5.4) by running short time MC simulation for a rough estimation for the scatters, and then apply unsupervised denoising approaches such as noise2noise [130], noise2self [13], etc. One can also use unsupervised learning in SPECT image reconstruc-

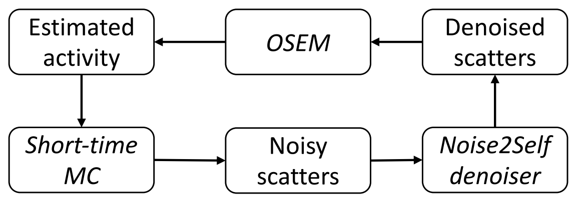


FIG 5.4 – Illustration of using unsupervised denoising methods for scatter correction.

tion algorithms as shown in Fig. 5.5. This involves dividing projection views into separate subsets and employing self-consistency loss between each subset during the training of a neural network model [240]. Unsupervised learning can also be used in SPECT image

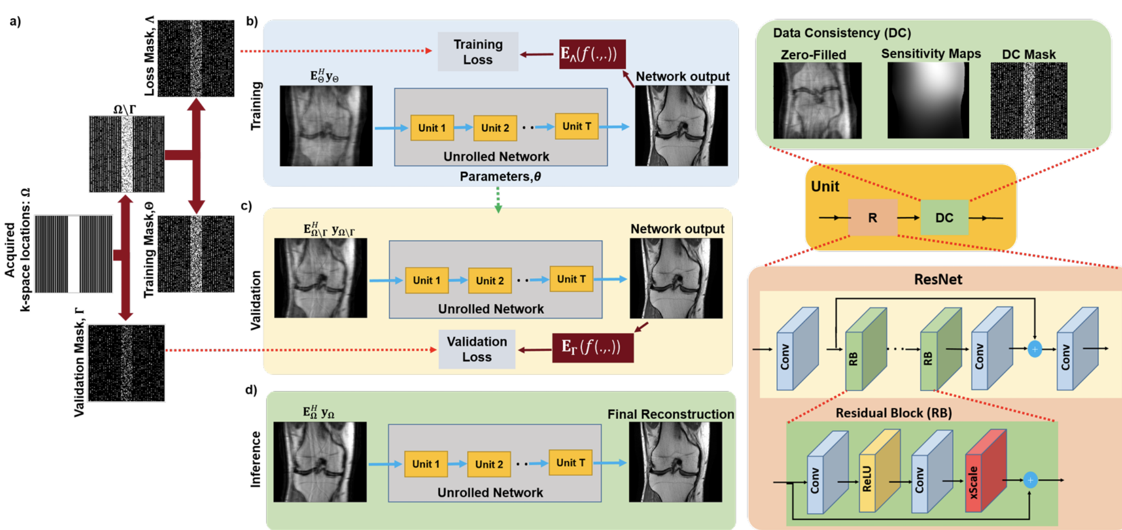


FIG 5.5 – Illustration of using zero-shot unsupervised learning for MRI image reconstruction. For SPECT reconstruction, one can replace MRI k-space acquisitions by SPECT acquisitions. Figure adopted from [240].

segmentation by employing unsupervised domain adaptation as shown in Fig. 5.6. This approach involves initially training a neural network with labeled data from a source domain, and then fine-tuning the pretrained network on unlabeled data from the target domain through self-learning. Self-learning can be achieved by minimizing the consistency loss between different realizations of probabilistic models or incorporating a segmentation head for early features and using the final predictions of the network as pseudo-labels to further refine the model [195, 234].

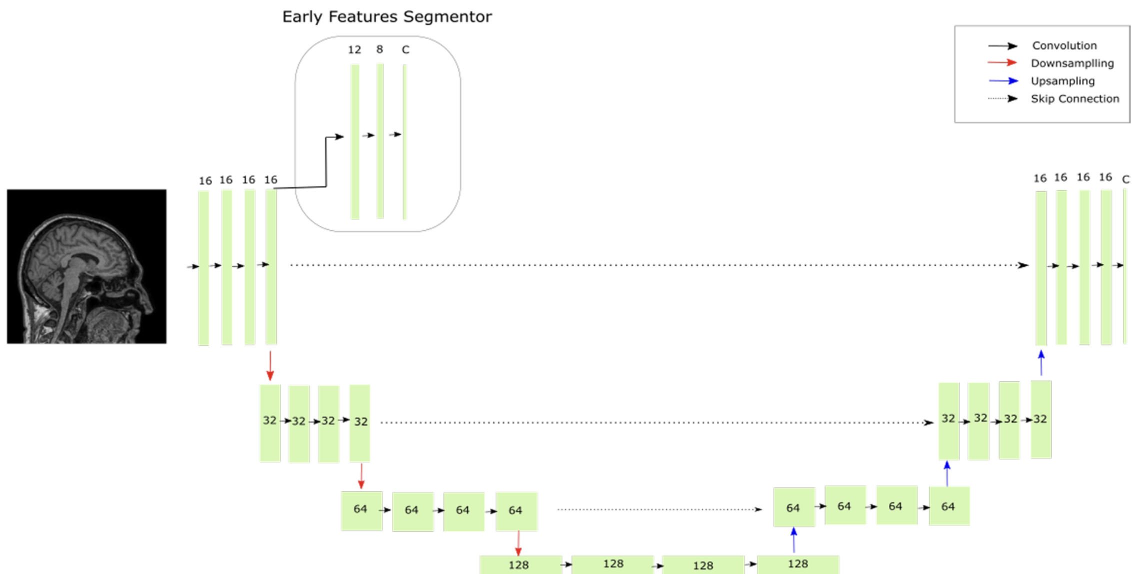


FIG 5.6 – Unsupervised method for image segmentation. Self-learning can be done by having self-consistency loss between the early segmenter and the final segmenter. Figure adopted from [195].

5.1.3 Patch-based Models

Patch-based machine learning models have gained significant attention in recent years [21, 224]. This approach is particularly useful in medical imaging where high-quality labeled datasets are often limited. The patch-based models use small image patches as the input for training and inference processes, enabling them to capture local spatial information efficiently. By extracting features from these localized patches, patch-based models can effectively classify or segment medical images, even with a small amount of annotated data available. One drawback of the patch-based method is its inability to capture global information from the image. Therefore, a potentially more effective approach would be to use a patch pyramid (as illustrated in Fig. 5.7), where patches are extracted at each level of the pyramid. The concept of patch pyramid can be applied to score-matching diffusion models, where the score function of the entire image can be expressed as a composite of the score functions of image patches at various scales. By learning and incorporating the score functions from these patch pyramids, the model can capture both local and global information, resulting in more accurate score-matching and potentially leading to improved performance on various medical imaging tasks.

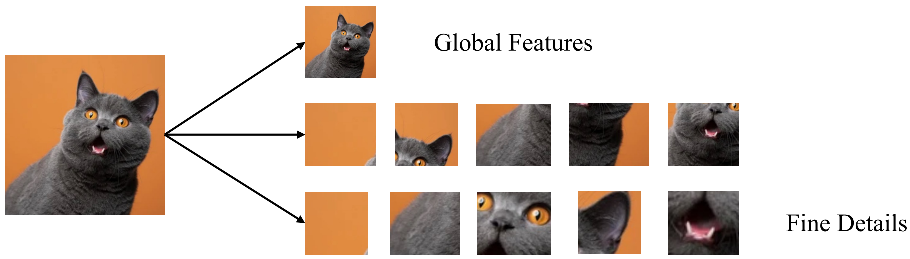


FIG 5.7 – Patch Pyramid for extracting image features at different scales. Adaptive/selective sampling can be applied to choose patches with richer information for training.

5.1.4 Learning 3D representation with 2D CNN

Another idea to train machine learning models on 3D data is to utilize 2D CNNs for representing 3D features. This is motivated by the fact that 2D CNNs have fewer parameters, reducing the risk of overfitting when training with limited data. An example that demonstrates this concept is the "Swap-Net" model, which alternates between swapping the spatial dimension and the channel dimension at different layers, allowing for 2D convolutions to be applied at every two-dimensional coordinate (*e.g.*, xy, xz, yz) within the three-dimensional data as illustrated in Fig. 5.8.

5.2 Generative AI

Generative AI has emerged as a popular technology in the field of medical imaging, revolutionizing the way we diagnose and treat various healthcare conditions. DL models such as GAN and VAE have been proposed to create highly realistic and high-resolution medical images [76, 116]. These generative models can generate synthetic images of organs, tissues, and anomalies, filling gaps in the limited datasets available for training and aiding in tasks like disease detection, tumor segmentation, and treatment planning. Generative AI for medical imaging holds great promise in improving the accuracy and efficiency of diagnostic procedures, enabling early disease detection, and ultimately enhancing patient care while reducing the burden on healthcare professionals.

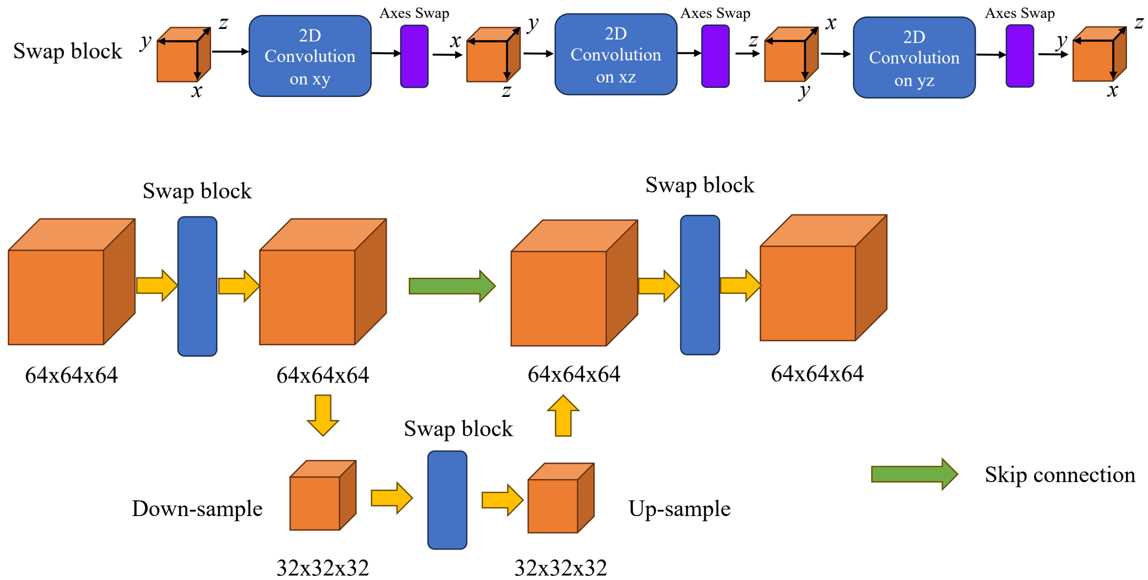


FIG 5.8 – Architecture of Swap-Unet. The idea of swapping axes was from [237].

5.2.1 Diffusion Models

Diffusion models, a category of deep generative models, have recently become a very hot topic in the field of machine learning research [88, 206]. These models are designed to capture and simulate complex probabilistic processes, making them particularly well-suited for tasks that involve understanding and generating sequential data. Diffusion models have gained popularity in various fields, such as natural language processing, image synthesis, and data generation. They are rooted in stochastic differential equations and Bayesian inference. The key idea behind diffusion models is to iteratively update the data by applying a series of transformations, which gradually make the data more similar to the target distribution. These transformations are designed to be reversible, meaning they can be undone. The diffusion models are reported to have the remarkable capability to model uncertainty and generate data that is coherent and controlled. As a result, they play a transformative role in the development of realistic content with high quality and contextual relevance.

5.2.2 PET-guided Diffusion for SPECT Image Reconstruction

To enhance the effectiveness of medical image reconstruction using diffusion models, a possible approach is to use guided diffusion as shown in Fig. 5.9. Guided diffusion is an extension of the standard diffusion model that incorporates supplementary guidance or conditioning information during the data generation process [57]. A future direction of this thesis can be to use PET images as the conditioning information when training a diffusion model for SPECT image reconstruction. This is motivated by the fact that PET images have higher resolution and exhibit similar activity distribution as SPECT images.

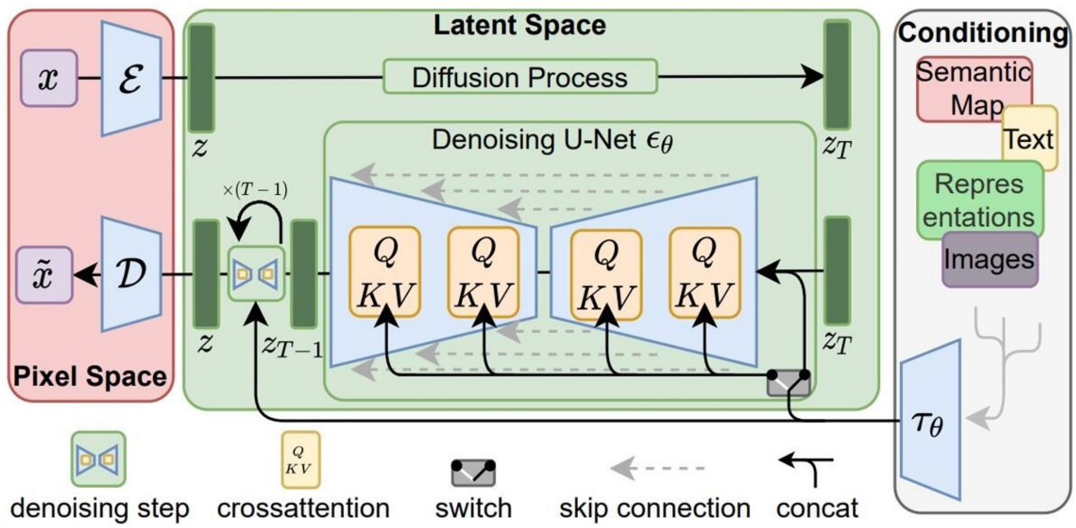


FIG 5.9 – Architecture of stable diffusion model [183]. For SPECT image reconstruction, one first train a SPECT diffusion model with PET images as conditions. After training, one apply the diffusion model to posterior sampling methods.

5.3 Optimization Methods

Optimization methods play a crucial role in solving inverse problems they allow one to find solutions or parameters that minimize a custom objective function. Therefore, it is important to derive robust optimization algorithms that result in faster convergence and enhance the quality of reconstruction.

5.3.1 Stochastic Expectation-Maximization with Variance Reduction (SVREM)

Expectation-Maximization (EM) is a popular tool but the whole dataset is needed in the E-step. Stochastic EM reduces the cost of E-step by stochastic approximation, but has a slower asymptotic convergence rate. Chen *et al.* [34] proposed SVREM and showed that it had the same exponential asymptotic convergence rate as EM. Kereta *et al.* [111] applied SVREM to PET image reconstruction and showed faster convergence rate compared to other accelerated gradient descent algorithms [108, 53, 1, 103, 232]. Applying SVREM to SPECT image reconstruction could be an interesting future direction.

CHAPTER 6

Conclusion

This thesis demonstrates algorithms that aim to solve Poisson inverse problems in phase retrieval and SPECT imaging. For phase retrieval, our contributions are novel algorithms [138, 139, 137] that have faster convergence speed and lead to improved image reconstruction quality. For example, we propose modifications to the WF algorithm. Our method determines the step size based on observed Fisher information and incorporates a quadratic majorizer into our majorize-minimize approaches. We demonstrate that our methods are effective and exhibit favorable convergence properties [139]. Furthermore, we explore cases involving measurements affected by a combination of Poisson and Gaussian noise. We propose the use of an innovative technique called "AWFS" which uses accelerated WF with a score function as a generative prior. Theoretical analysis is conducted to showcase the critical point convergence guarantee of our algorithm. Simulation results demonstrate that our approach enhances reconstruction quality in terms of both visual perception and numerical assessment.

For SPECT imaging, we focus on DL solutions [134, 132, 140]. We develop a Julia toolbox [134] enables efficient modeling of SPECT forward-backward projectors with parallel computing and minimized memory allocations. This facilitates effective backpropagation during deep learning regularized iterative algorithm training, resulting in higher quality reconstructions compared to non-end-to-end methods. Moreover, we propose DblurDoseNet [132], a deep neural network for joint dosimetry estimation and image deblurring after SPECT reconstruction. It accurately estimates dose-rate distribution and compensates for SPECT resolution effects. Evaluations on phantoms and patients show that DblurDoseNet outperforms conventional dosimetry methods while being fast enough for real-time clinical use in radionuclide therapy dosimetry. Additionally, we propose a neural network with unsupervised learning to predict missing SPECT projections. Our method aims to decrease acquisition time by obtaining only a subset of all projections. Our method outperforms linear interpolation techniques used to predict missing projection views in terms of the achieved image reconstruction quality [140].

As mentioned in Chapter 5, there are several potential avenues for further research. These include investigating transfer learning techniques like finetuned SAM for tumor segmentation in SPECT images, exploring unsupervised methods for scatter correction in SPECT imaging, and incorporating PET-guided diffusion into the reconstruction of SPECT images. Similar methods can be employed to address 3D phase retrieval problems as well. We will be excited to see explorations on these research directions and believe they have the potential to improve the accuracy and efficiency of algorithms for solving Poisson inverse problems.

APPENDIX A

Proof of the Proposed Improved Curvature Formula

This appendix proves that the improved curvature formula defined in (3.13) provides a majorizer for the negative log-likelihood of the Poisson model in Chapter 3. For simplicity, we drop the subscript i and irrelevant constants and focus on the negative log-likelihood for real case for simplicity as in (2.13). One can generalize the majorizer derived here for (2.13) to the complex case by taking the magnitude and some other minor modifications.

First, we consider some simple cases:

- If $y = 0$, then (2.13) is a quadratic function, so no quadratic majorizer is needed.
- If $b = 0$ and $y > 0$ then (2.13) has unbounded 2nd derivative so no quadratic majorizer exists.
- If $b = 0$ and $r = 0$, then y must be zero because a Poisson random variable with zero mean can only take the value 0. Thus again quadratic majorizer is not needed.

So hereafter we assume that $y > 0,\ b > 0$. Under these assumptions, the derivatives of (2.13) are:

$$\dot{\phi}(r) = 2r\left(1 - \frac{y}{r^2 + b}\right), \tag{A.1}$$

$$\ddot{\phi}(r) = 2 + 2y\frac{r^2 - b}{(r^2 + b)^2}, \tag{A.2}$$

$$\phi^{(3)}(r) = \frac{2yr(3b - r^2)}{(r^2 + b)^3}, \tag{A.3}$$

where $\phi^{(3)}(r)$ denotes the third derivative. Clearly, $\dot{\phi}(r)$ is convex on $(-\infty, -\sqrt{3b}]$ and $[0, \sqrt{3b}]$, and concave on $[-\sqrt{3b}, 0]$ and $[\sqrt{3b}, +\infty)$, based on the sign of $\phi^{(3)}(r)$.

A quadratic majorizer of $\phi(\cdot)$ at point s has the form:

$$\Phi(r;s) = \phi(s) + \dot{\phi}(s)(r-s) + \frac{1}{2}c(s)(r-s)^2, \tag{A.4}$$

its derivative (w.r.t. r) is:

$$\dot{\Phi}(r;s) = c(s)(r-s) + \dot{\phi}(s). \tag{A.5}$$

By design, this kind of quadratic majorizer satisfies $\Phi(s;s) = \phi(s)$ and $\dot{\Phi}(s;s) = \dot{\phi}(s)$. From (A.3), we note that $r^2 = 3b$ is a maximizer of $\ddot{\phi}$ so the maximum curvature is:

$$\ddot{\phi}(r) \leq 2y\frac{2b}{(4b)^2} + 2 = 2 + \frac{y}{4b}. \tag{A.6}$$

Proposition: $\Phi(r;s)$ defined in (A.4) is a majorizer of $\phi(r)$ when $c(s) = c_{\text{imp}}(s)$, where:

$$c_{\text{imp}}(s) \triangleq \begin{cases} \ddot{\phi}(u(s))\,, & s \neq 0, \\ \lim_{s\to 0} \ddot{\phi}(u(s))\,, & s = 0, \end{cases} \tag{A.7}$$

where

$$u(s) \triangleq \frac{b + \sqrt{b^2 + bs^2}}{s}. \tag{A.8}$$

By construction, the proposed curvature $c(s)$ is at most the max curvature given in (A.6).

Proof: Because of the symmetry of $\ddot{\phi}(r)$, it suffices to prove the proposition for $s \geq 0$ without loss of generality. First we consider some trivial cases:

1. If $s = 0$, one can verify $\lim_{s\to 0} \ddot{\phi}(u(s)) = 2$. In this case, $\Phi(r;s)$ is simply

$$\Phi(r;0) = \phi(0) + \frac{1}{2}c(0)r^2 = r^2 + b - y\log(b) \geq r^2 + b - y\log\left(r^2 + b\right) = \phi(r). \tag{A.9}$$

2. If $s = \sqrt{3b}$, one can verify

$$\ddot{\phi}(g(\sqrt{3b})) = 2 + \frac{y}{4b}, \tag{A.10}$$

which equals the maximum curvature.

Hereafter, we consider only $s > 0$ and $s \neq \sqrt{3b}$. To proceed, it suffices to prove

$$\forall r \in (-\infty, s],\ \dot{\phi}(r) \geq \dot{\Phi}(r;s), \quad \forall r \in [s, +\infty),\ \dot{\phi}(r) \leq \dot{\Phi}(r;s), \tag{A.11}$$

because if (A.11) holds, then $\forall \tilde{r} < s$:

$$\Phi(s;s) - \Phi(\tilde{r};s) = \int_{\tilde{r}}^{s} \dot{\Phi}(r;s)dr \leq \int_{\tilde{r}}^{s} \dot{\phi}(r)dr = \phi(s) - \phi(\tilde{r}), \tag{A.12}$$

and $\forall \tilde{r} > s$:

$$\Phi(\tilde{r};s) - \Phi(s;s) = \int_{s}^{\tilde{r}} \dot{\Phi}(r;s)dr \geq \int_{s}^{\tilde{r}} \dot{\phi}(r)dr = \phi(\tilde{r}) - \phi(s). \tag{A.13}$$

Together with $\Phi(s;s) = \phi(s)$, we have shown that (A.11) implies $\Phi(r;s) \geq \phi(r),\ \forall r \in \mathbb{R}$.

Substituting $\dot{\Phi}(r;s) = c(s)(r-s) + \dot{\phi}(s)$ into (A.11), one can verify that showing (A.11) becomes showing

$$c_{\text{imp}}(s) \geq \frac{\dot{\phi}(r) - \dot{\phi}(s)}{r - s}, \quad \forall r \in \mathbb{R},\ r \neq s. \tag{A.14}$$

Furthermore, when $s > 0$, the parabola $\Phi(\cdot;s)$ is symmetric about its minimizer:

$$\delta = \delta(s) \triangleq \arg\min_{r} \Phi(r;s) = s - \frac{\dot{\phi}(s)}{c_{\text{imp}}(s)} = \frac{s\,\ddot{\phi}(u(s)) - \dot{\phi}(s)}{\ddot{\phi}(u(s))} \geq 0. \tag{A.15}$$

This minimizer is nonnegative because $\dot{\phi}(s) \leq 2s$ and

$$c_{\text{imp}}(s) = \ddot{\phi}(u(s)) = 2 + \frac{ys^2(b + \sqrt{b^2 + bs^2})}{b(b + s^2 + \sqrt{b^2 + bs^2})^2} \geq 2. \tag{A.16}$$

Thus, if $\phi(r) \leq \Phi(r;s)$ when $r \geq 0$, we have $\phi(-r) = \phi(r) \leq \Phi(r;s) \leq \Phi(-r;s) = \Phi(r+2\delta;s)$, so it suffices to prove (A.14) only for $r \geq 0$, which simplifies (A.14) to showing

$$c_{\text{imp}}(s) \geq \frac{\dot{\phi}(r) - \dot{\phi}(s)}{r - s}, \quad \forall r \geq 0,\ r \neq s. \tag{A.17}$$

In short, if (A.17) holds, then $\Phi(r;s) \geq \phi(r),\ \forall r \in \mathbb{R}$.

To prove (A.17), we exploit a useful property of $c_{\text{imp}}(s)$. Under geometric view, $c_{\text{imp}}(s)$ defines (the ratio of) an affine function connecting points $(u(s), \dot{\phi}(u(s)))$ and $(s, \dot{\phi}(s))$ is tangent to $\dot{\phi}(r)$ at point $r = u(s)$, so that one can verify

$$\ddot{\phi}(u(s)) = c_{\text{imp}}(s) = \frac{\dot{\phi}(u(s)) - \dot{\phi}(s)}{u(s) - s}, \quad u(s) \neq s. \tag{A.18}$$

The reason why $u(s) \neq s$ is that one can verify $u(s) = s$ implies $s = \sqrt{3b}$ for $s > 0$ that has already been proved above.

Let $\xi(r) = (\dot{\phi}(r) - \dot{\phi}(s))/(r - s)$, where $r \geq 0$ and $r \neq s$, plugging in $\dot{\phi}(r)$ and $\dot{\phi}(s)$ yields:

$$\xi(r) = 2 + \frac{2y(sr - b)}{(s^2 + b)(r^2 + b)}. \tag{A.19}$$

Differentiating w.r.t. r leads to:

$$\dot{\xi}(r) = \frac{2y}{s^2 + b} \cdot \frac{-sr^2 + 2br + bs}{(r^2 + b)^2}, \tag{A.20}$$

where one can verify the positive root of $-sr^2 + 2br + bs = 0$ is $u(s)$ that is given by (A.8).

Together with $\dot{\xi}(r) > 0$ when $r \in (0, u(s))$ and $\dot{\xi}(r) < 0$ when $r \in (u(s), \infty)$, we have (A.17) holds because $\xi(r)$ achieves its maximum at $\xi(u(s))$:

$$\xi(r) \leq \xi(u(s)) = c_{\text{imp}}(s). \tag{A.21}$$

APPENDIX B

Uniform Cramér–Rao Lower Bound Analysis for Phase Retrieval Algorithms

This appendix[1] derives and analyzes the UCRLB for the phase retrieval problem, and then compares the bias-variance trade-off between phase retrieval algorithms (*e.g.*, Wirtinger flow, Gerchberg-Saxton, phaselift, MM and ADMM) that were derived from MLE where the measurements follow i.i.d. Gaussian distribution. We also consider regularizers that exploit the assumed properties of the latent signal, *e.g.*, ℓ_2 norm and ℓ_1 norm (approximated by the Huber function) that corresponds to the sparsity of finite differences (anisotropic TV) or of the detailed coefficients of a discrete wavelet transform. Simulation results show that many phase retrieval algorithms can be biased so that the classical CRLB fails to bound their variance. Regularized algorithms that better approximate the properties of the true signal have better bias-variance trade-offs (when compared to UCRLB) and lower reconstruction error.

B.1 Motivation

It is well known that the variance of any unbiased estimator is bounded by the CRLB, however, many estimators, *e.g.*, derived from regularized maximum likelihood (maximum a posteriori in Bayesian setting) are typically biased. Hence their variance cannot be bounded by the classical CRLB. Hero *et al.* [87] proposed the UCRLB, which is a bound on the smallest attainable variance that can be achieved using any estimator with bias gradient of which norm is bounded by a constant. We analyzed the UCRLB for the challenging phase retrieval problem.

There have been several studies involving the classical CRLB for phase retrieval in the literature [176, 141, 6]. Balan and Bekkerman [6] derived and analyzed the CRLB for

[1]This work is based on [135].

two different types of phase retrieval problems. Qian *et al.* [176] proposed a novel method known as feasible point pursuit (FPP) that is based on quadratically constrained quadratic programming (QCQP) and is measured against the classical CRLB. However, we argue that the classical CRLB might not be very useful because it is unknown whether an estimator determined by an iterative phase retrieval algorithm is unbiased or not, especially for those with tuning parameters or regularizers. Instead, UCRLB should be used to evaluate these algorithms, as will be presented next.

B.2 Methods

As discussed in the previous section, a limitation of the classical CRLB analysis is that unbiased estimation is often not practical. Instead, the uniform CRLB is often used for biased estimation [87]. For simplicity, we only consider the scalar UCRLB, which is the smallest attainable variance of a single element of the true signal. Following Theorem 1 in [87], the scalar UCRLB can be written as

$$\begin{aligned} &\mathrm{Var}(\hat{\boldsymbol{x}}_j) \geq B_\gamma \triangleq (\boldsymbol{e}_j + \boldsymbol{v}_\gamma)' \boldsymbol{F}^{+} (\boldsymbol{e}_j + \boldsymbol{v}_\gamma), \\ &\boldsymbol{v}_\gamma \triangleq -(\gamma \boldsymbol{C} + \boldsymbol{F}^{+})^{-1} \boldsymbol{F}^{+} \boldsymbol{e}_j, \\ &\boldsymbol{F} \triangleq \frac{4}{\sigma^2} \boldsymbol{G}\boldsymbol{G}', \quad \boldsymbol{G} \triangleq \mathrm{real}\{\boldsymbol{A}' \,\mathsf{diag}\{|\boldsymbol{A}\boldsymbol{x}|\}\}, \end{aligned} \tag{B.1}$$

where $\boldsymbol{F}^{+}$ denotes the Moore–Penrose inverse of $\boldsymbol{F}$, $\boldsymbol{C}$ is a positive definite matrix, and $\boldsymbol{e}_j$ is a unit vector whose the jth element is 1. The scalar UCRLB can be approximated by a set of points $\left(\|\boldsymbol{v}_\gamma\|_{\boldsymbol{C}}, \sqrt{B_\gamma}\right)$ with varying γ and an appropriate choice of $\boldsymbol{C}$. We used $\boldsymbol{C} = \boldsymbol{I}$ for simplicity (so $\|\boldsymbol{v}_\gamma\|_{\boldsymbol{C}} = \|\boldsymbol{v}_\gamma\|_2$).

With the UCRLB, the limiting variance of an estimator becomes a function of the norm of bias gradient, where empirically we approximate the bias gradient by [87]:

$$\nabla \boldsymbol{b}(\hat{\boldsymbol{x}}) \triangleq \nabla_{\boldsymbol{x}} \left(\mathbb{E}[\hat{\boldsymbol{x}}] - \boldsymbol{x}\right) \approx \frac{1}{L-1} \sum_{l=1}^{L} \left(\hat{\boldsymbol{x}}(\boldsymbol{y}_l) - \overline{\hat{\boldsymbol{x}}}\right) \left(-\nabla f(\boldsymbol{x}; \boldsymbol{y}_l)\right)' - \boldsymbol{I}, \tag{B.2}$$

where L denotes the number of realizations of $\boldsymbol{y}$, $\boldsymbol{I}$ denotes an identity matrix, and $\hat{\boldsymbol{x}}(\boldsymbol{y}_l)$ denotes the estimator based on $\boldsymbol{y}_l$. $\overline{\hat{\boldsymbol{x}}}$ is the sample mean of $\hat{\boldsymbol{x}}(\boldsymbol{y}_l)$ shown as follows

$$\overline{\hat{\boldsymbol{x}}} \triangleq \frac{1}{L} \sum_{l=1}^{L} \hat{\boldsymbol{x}}(\boldsymbol{y}_l). \tag{B.3}$$

Next, we define the norm of bias gradient for the jth element in $\nabla \boldsymbol{b}(\hat{\boldsymbol{x}})$ by $\|\delta(\hat{\boldsymbol{x}}_j)\|_{\boldsymbol{C}}$ with

$$\delta(\hat{\boldsymbol{x}}_j) \triangleq \nabla \boldsymbol{b}(\hat{\boldsymbol{x}})' \boldsymbol{e}_j. \tag{B.4}$$

Then, we estimate the variance of $\hat{\boldsymbol{x}}_j$ (the jth element in $\hat{\boldsymbol{x}}$) by the sample variance

$$\operatorname{Var}(\hat{\boldsymbol{x}}_j) \approx \boldsymbol{e}_j' \left(\frac{1}{L-1} \sum_{l=1}^{L} \left(\hat{\boldsymbol{x}}(\boldsymbol{y}_l) - \overline{\hat{\boldsymbol{x}}}\right) \left(\hat{\boldsymbol{x}}(\boldsymbol{y}_l) - \overline{\hat{\boldsymbol{x}}}\right)' \right) \boldsymbol{e}_j. \tag{B.5}$$

Finally we compare the point $\left(\|\delta(\hat{\boldsymbol{x}}_j)\|_{\boldsymbol{C}}, \sqrt{\operatorname{Var}(\hat{\boldsymbol{x}}_j)}\right)$ with the UCRLB, *i.e.*, set of points $\left(\|\boldsymbol{v}_\gamma\|_{\boldsymbol{C}}, \sqrt{B_\gamma}\right)$, to illustrate the bias-variance trade-off of the corresponding estimator.

B.3 Experiments

B.3.1 Compared Algorithms

We compared the following algorithms with the UCRLB:

- Unregularized WF [26].
- Regularized WF with ℓ_2 norm regularizer and ℓ_1 norm approximated by the Huber function $h.(\boldsymbol{T}\boldsymbol{x}; \alpha)$. We set $\boldsymbol{T}$ to be the TV matrix or the detailed coefficients of ODWT.
- GS [69].
- PhaseLift with variant β for the nuclear norm [24].
- PRIME [177].
- ADMM [141].

B.3.2 Results

Fig. B.1 (b) shows that the WF and PRIME algorithm are biased with its variance below the classical CRLB but above the UCRLB, which illustrates the limitation of the classical CRLB; WF with ℓ_2 norm regularization (WF-ridge) achieves the lowest variance for some β [61], but at the cost of the largest bias; WF-TV is closer to the UCRLB compared to WF-ODWT, presumably due to the true signal is piece-wise uniform, which better matches the assumption of TV regularization. Fig. B.1 (c) and Fig. B.1 (d) compare the NRMSE of

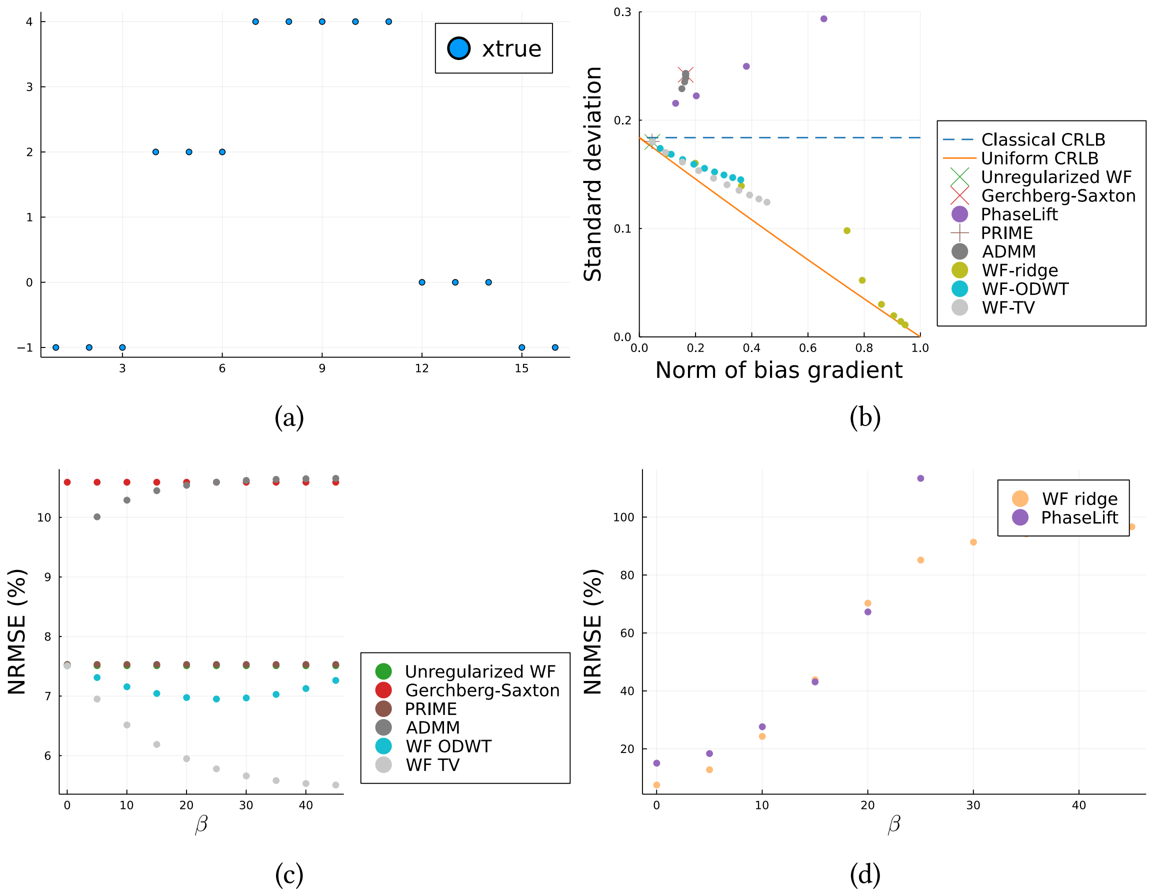


FIG B.1 – UCRLB and NRMSE comparison of phase retrieval algorithms. Subfigure (a) shows the true signal.

phase retrieval algorithms. Algorithms based on the "magnitude model", *e.g.*, PhaseLift and ADMM worked worse than unregularized WF and PRIME; for regularized algorithms, we found that WF with TV regularization yielded better reconstruction than with ODWT, perhaps due to TV regularization better fits the assumed property of the true signal.

APPENDIX C

Proof of the Critical Point Convergence for the "AWFS" Algorithm

This appendix proves the "AWFS" algorithm in Chapter 3 has a critical point convergence guarantee.

It was already shown that ∇g_{PG} is Lipschitz continuous, so the remaining problem is to find a Lipschitz constant for $s(\boldsymbol{x}, \sigma)$. We assume that the data allows the neural network to learn the score function well, *i.e.*, $p_\sigma(\boldsymbol{x}) = p(\boldsymbol{x}) \circledast \mathcal{N}(0, \sigma^2)$, where $\circledast$ denotes (circular) convolution. We start with some well-known lemmas where the proofs can be readily found in [81, 147]. We include our proofs just for completeness.

Lemma C.1: The Fourier transform (and inverse transform) of an absolutely integrable function is continuous.

Proof of lemma C.1: Let f be absolutely integrable and let $\tilde{f}$ be its Fourier transform. We have

$$|\tilde{f}(w+h) - \tilde{f}(w)| = \left| \int f(x)(e^{-2\pi j x(w+h)} - e^{-2\pi j w x}) dx \right| \leq \int |f(x)||e^{2\pi j x h} - 1| dx \\ \leq \max(|e^{2\pi j x h} - 1|) \int |f(x)| dx \leq 2 \int |f(x)| dx. \quad \text{(C.1)}$$

Using absolute integrability of f, we see $|\tilde{f}(w+h) - \tilde{f}(w)|$ tends to 0 as h tends to 0, so $\tilde{f}$ is uniformly continuous, which also implies it is continuous.

The proof of the inverse transform follows similarly.

Lemma C.2: Suppose a sequence of functions $f_i : \mathbb{R} \to \mathbb{R}$ converges in the L^1 to some function f, and that each f_i is absolutely integrable. Then f is also absolutely integrable.

Proof: Because

$$\lim_{i \to \infty} \int_{-\infty}^{\infty} |f_i(x) - f(x)| dx = 0 \quad \text{(C.2)}$$

and that $\int_{-\infty}^{\infty} |f_i(x)|dx < \infty$. It follows that

$$\int_{-\infty}^{\infty} |f(x)|dx < \int_{-\infty}^{\infty} |f(x) - f_i(x)|dx + \int_{-\infty}^{\infty} |f_i(x)|dx, \tag{C.3}$$

for any i. The second integral is always finite, and for sufficiently large i, the first integral must be finite as it converges to 0. Hence it is possible to find i such that both integrals converge, so f is absolutely integrable.

Proposition C.1: The derivative of $\log(p_\sigma(\boldsymbol{x}))$ is bounded on the interval $[-C, C]$.
Proof: We start by dropping constant factors and using derivative of a convolution, we have

$$\frac{d}{dx}(\log(p(x) \circledast \mathcal{N}(0, \sigma^2))) \sim \frac{\mathcal{F}^{-1}(\imath x \mathcal{F}(p(x)) \cdot \mathcal{F}(\mathcal{N}(0, \sigma^2)))}{p(x) \circledast \mathcal{N}(0, \sigma^2)} \sim \frac{\mathcal{F}^{-1}(xe^{-x^2} \cdot \mathcal{F}(p(x)))}{p(x) \circledast \mathcal{N}(0, \sigma^2)} \tag{C.4}$$

where $\mathcal{F}$ denotes Fourier transform. The denominator is continuous and since x lies in a closed interval by assumption, has a lower bound $M > 0$ by the extreme value theorem. We next consider the numerator.
By [75, pp. 65], a sequence of Gaussian mixture models (GMM) can be used to approximate any smooth probability distribution in L^2 convergence. Furthermore, L^2 convergence implies L^1 convergence. Hence, consider a sequence of GMM f_i that converge in L^1 to $p(x)$. By linearity of Fourier transform, $\mathcal{F}(f_i(x))$ must be a linear combination of terms of the form $e^{-(x-\mu_i)^2/c_i}$ for some c_i. Thus, the numerator $xe^{-x^2} \cdot \mathcal{F}(f_i(x))$ is a finite linear combination of terms of the form $xe^{-(x-\mu_i)^2/c_i}$, each of which are absolutely integrable. Therefore, we have a sequence of functions, each of which are absolutely integrable, that converge in L^1 to $xe^{-x^2} \cdot \mathcal{F}(p(x))$, so by Lemma C.2, this is also absolutely integrable. By Lemma C.1, the inverse Fourier transform of this is continuous. Finally, again by the extreme value theorem and using the boundedness of x, the numerator is bounded above by some $M' > 0$. Hence, the entire expression (C.4) is bounded above by M'/M.

Lemma C.3: Suppose we have an everywhere twice differentiable function of two variables $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$. Then $\frac{\partial^2}{\partial x \partial y} \log f(x, y)$ is bounded if the following three conditions are met:

1. $\frac{\partial^2}{\partial x \partial y} f(x, y)$ is bounded.
2. f itself is bounded below by a positive number and also bounded above.
3. ∇f is bounded.

Proof: Suppose we have $f(x, y)$ satisfying those three conditions. We compute the second partial derivative of its log:

$$\frac{\partial}{\partial x} \log f(x, y) = \frac{\frac{\partial}{\partial x} f(x, y)}{f(x, y)}. \tag{C.5}$$

and

$$\frac{\partial^2}{\partial x \partial y} \log f(x, y) = \frac{(\frac{\partial^2}{\partial x \partial y} f(x, y)) f(x, y) - (\frac{\partial}{\partial x} f(x, y))(\frac{\partial}{\partial y} f(x, y))}{f(x, y)^2}. \tag{C.6}$$

From the second condition, the denominator is bounded below by a positive number, so it suffices to consider the boundedness of the numerator. The first term of the numerator is a product of two quantities, the first of which is bounded by the first condition and the second of which is bounded by the second condition. The second term of the numerator is also a product of two quantities, both of which are bounded by the third condition. Thus, this shows $\frac{\partial^2}{\partial x \partial y} \log f(x, y)$ is bounded.

Proposition C.2: The gradient of $\log(p_\sigma(\boldsymbol{x}))$ is Lipschitz continuous on $[-C, C]^N$.

Proof: By renaming the variables, and redefining $f(x, y) = p(x, y, \cdots)$, we may consider the boundedness of $\frac{\partial^2}{\partial x \partial y} (\log f(x, y) \circledast \mathcal{N}(0, \sigma^2 \boldsymbol{I}))$ on $[-C, C]^2$. To apply Lemma C.3 to remove the log, we need to verify the three conditions. Define $g(x, y) = f(x, y) \circledast \mathcal{N}(0, \sigma^2 \boldsymbol{I})$. The second condition is readily verified to be true: By assumption, x and y take values on a closed interval, thus by the extreme value theorem, so does $g(x, y)$. Further, g is a convolution of positive numbers and so the output is always positive, hence, the lower bound of this closed interval is a positive number, verifying this condition.

For the third condition, we need to consider boundedness of $\frac{\partial}{\partial x} f(x, y) \circledast \mathcal{N}(0, \sigma^2 \boldsymbol{I})$. This is nearly identical to Proposition C.2, with the only difference being we have some general function in terms of only x $f(x, y)$ instead of a probability distribution $p(x)$. The proof of that lemma is readily adapted for this case with the only condition needing verification being the absolute integrability over x of $f(x, y)$. In fact, this is clear because f is always positive; hence, integrating over f with respect to x must yield a finite number as integrating a second time over y yields 1.

It thus suffices to consider boundedness of $h(x, y) = \frac{\partial^2}{\partial x \partial y} (f(x, y) \circledast \mathcal{N}(0, \sigma^2 \boldsymbol{I}))$. It is assumed that f is smooth and the convolution of smooth functions is smooth, which implies $f(x, y) \circledast \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ is smooth. Hence h is differentiable, so it is continuous. By the extreme value theorem, as x and y take values on a closed interval, h must be bounded. By Lemma C.3, the entries of the Hessian of the score function are bounded. Therefore, a Lipschitz constant of $s_{\boldsymbol{\theta}}(\boldsymbol{x})$ exists.

Proof of Theorem 2: By Proposition C.2, and from the design of Algorithm 4, $\boldsymbol{x}_{t,k}$ and $\boldsymbol{w}_{t,k}$ are both bounded between $[0, C]$ for all t, k, so the Lipschitz constant $\mathcal{L}^*$ of $\nabla g_{\mathrm{PG}}(\cdot) + s_{\boldsymbol{\theta}}(\cdot)$ exists. With the step size μ satisfying $0 < \mu < \frac{1}{\mathcal{L}^*}$, and the weighting factor $\boldsymbol{\gamma} \in \{0, 1\}$ being chosen according to whichever higher posterior probability between $p_{\sigma_k}(\boldsymbol{z}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ and $p_{\sigma_k}(\boldsymbol{v}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ (where $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}, \bar{\boldsymbol{b}}, \boldsymbol{r}) \propto p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}, \bar{\boldsymbol{b}}, \boldsymbol{r}) p_{\sigma_k}(\boldsymbol{x})$), then we satisfy all conditions in Theorem 1 of [131], which establishes the critical-point convergence of the inner iteration sequence $\boldsymbol{x}_{t,k}$ in Algorithm 4 for the posterior distribution $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$. Similar convergence analysis can be found in [71].

BIBLIOGRAPHY

[1] S. Ahn and J. Fessler. "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms". In: *IEEE Transactions on Medical Imaging* 22.5 (2003), pp. 613–626. DOI: 10.1109/TMI.2003.812251 (cit. on p. 110).

[2] A. Akhavanallaf, I. Shiri, H. Arabi, and H. Zaidi. "Whole-body voxel-based internal dosimetry using deep learning". In: *Eur. J. Nucl. Med. Mol. Imaging* 48.3 (2021). DOI: 10.1007/s00259-020-05013-4 (cit. on pp. 69, 80).

[3] A. Aksac, D. Demetrick, T. Ozyer, *et al.*. "BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis". In: *MC Res Notes* 12.82 (2019). DOI: https://doi.org/10.1186/s13104-019-4121-7 (cit. on pp. 45, 46, 48, 49, 51).

[4] H. Arabi and H. Zaidi. "Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy". In: *Eur. J. Hybrid Imaging* 23 (2020). DOI: 10.1186/s41824-020-00086-8 (cit. on p. 69).

[5] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand. "Invertible generative models for inverse problems: mitigating representation error and dataset bias". In: *Proceedings of the 37th International Conference on Machine Learning* 119 (2020) (cit. on p. 39).

[6] R. Balan and D. Bekkerman. "The Cramer-Rao Lower Bound in the Phase Retrieval Problem". In: *SampTA*. 2019, pp. 1–5. DOI: 10.1109/SampTA45681.2019.9030920 (cit. on p. 117).

[7] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson. "Saving phase: Injectivity and stability for phase retrieval". In: *Appl. Comp. Harm. Anal.* 37.1 (2014), pp. 106–125. DOI: 10.1016/j.acha.2013.10.002 (cit. on p. 9).

[8] A. S. Bandeira, Y. Chen, and D. G. Mixon. "Phase retrieval from power spectra of masked signals". In: *Info. Infer.* 3.2 (2014), pp. 83–102. DOI: 10.1093/imaiai/iau002 (cit. on p. 10).

[9] D. A. Barmherzig and J. Sun. "Low-photon holographic phase retrieval". In: *OSA: Computational Optical Sensing and Imaging*. 2020 (cit. on pp. 12, 30–32, 38).

[10] D. A. Barmherzig, J. Sun, E. J. Candes, T. Lane, and P.-N. Li. "Dual-Reference Design for Holographic Phase Retrieval". In: *2019 13th International conference on Sampling Theory and Applications (SampTA)*. 2019, pp. 1–4. DOI: 10.1109/SampTA45681.2019.9030848 (cit. on p. 2).

[11] D. A. Barmherzig, J. Sun, P.-N. Li, T. Lane, and E. J. Candès. "Holographic phase retrieval and reference design". In: *Inverse problems* 35.9 (), pp. 94001–. DOI: 10.1088/1361-6420/ab23d1 (cit. on p. 2).

[12] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman. "Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields". In: *ICCV* (2023) (cit. on p. 87).

[13] J. Batson and L. Royer. “Noise2Self: Blind denoising by self-supervision”. In: *Proceedings of the International Conference on International Conference on Machine Learning*. 2019 (cit. on pp. 39, 45, 49, 105).

[14] C. F. Baumgartner, K. C. Tezcan, K. Chaitanya, A. M. Hötker, U. J. Muehlematter, K. Schawkat, A. S. Becker, O. Donati, and E. Konukoglu. “PHiSeg: Capturing Uncertainty in Medical Image Segmentation”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II*. 2019, pp. 119–127. DOI: 10.1007/978-3-030-32245-8_14 (cit. on p. 78).

[15] A. Beck and M. Teboulle. “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”. In: *SIAM J. Imaging Sci.* 2.1 (2009), 183–202. DOI: 10.1137/080716542 (cit. on p. 28).

[16] L. Bian, J. Suo, J. Chung, X. Ou, C. Yang, F. Chen, and Q. Dai. “Fourier ptychographic reconstruction using Poisson maximum likelihood and truncated Wirtinger gradient”. In: *Nature Sci. Rep.* 6.1 (2016). DOI: 10.1038/srep27384 (cit. on pp. 9, 12, 21, 35, 38).

[17] D. Böhning and B. G. Lindsay. “Monotonicity of quadratic approximation algorithms”. In: *Ann. Inst. Stat. Math.* 40.4 (Dec. 1988), 641–63. DOI: 10.1007/BF00049423 (cit. on p. 26).

[18] W. Bolch, L. Bouchet, R. JS., *et al.*. “MIRD pamphlet No. 17: the dosimetry of nonuniform activity distributions–radionuclide S values at the voxel level”. In: *J. Nucl. Med.* 40.1 (1999) (cit. on pp. 68, 69).

[19] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. “Compressed Sensing Using Generative Models”. In: *Proceedings of the International Conference on Machine Learning (ICML)*. Ed. by D. Precup and Y. W. Teh. Vol. 70. PMLR, Aug. 2017, pp. 537–546. DOI: 10.5555/3305381.3305437 (cit. on p. 39).

[20] D. Brice. “Stopping powers for electrons and positrons, ICRU report 37”. In: *Nuc. Inst. Meth. Phys. Res.* 12.1 (1985). DOI: 10.1016/0168-583X(85)90718-9 (cit. on p. 72).

[21] A. Brutzkus, A. Globerson, E. Malach, A. R. Netser, and S. Shalev-Schwartz. “Efficient Learning of CNNs using Patch Based Features”. In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato. Vol. 162. Proceedings of Machine Learning Research. PMLR, 2022, pp. 2336–2356 (cit. on p. 106).

[22] J.-F. Cai, M. Huang, D. Li, and Y. Wang. “Nearly optimal bounds for the global geometric landscape of phase retrieval”. In: *Inverse Problems* 39.7 (June 2023), p. 075011. DOI: 10.1088/1361-6420/acdab7 (cit. on p. 52).

[23] T. T. Cai, X. Li, and Z. Ma. “Optimal Rates of Convergence for Noisy Sparse Phase Retrieval via Thresholded Wirtinger Flow”. In: *Annals Stat.* 44.5 (2016), pp. 2221–2251. DOI: 10.1214/16-AOS1443 (cit. on pp. 11, 38).

[24] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski. “Phase retrieval via matrix completion”. In: *SIAM J. Imaging Sci.* 6.1 (2013), 199–225. DOI: 10.1137/110848074 (cit. on pp. 10, 12, 31, 38, 50, 119).

[25] E. J. Candes, T. Strohmer, and V. Voroninski. “PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming”. In: *Comm. Pure Appl. Math.* 66.8 (Aug. 2013), 1241–74. DOI: 10.1002/cpa.21432 (cit. on pp. 9, 11, 38, 50).

[26] E. Candes, X. Li, and M. Soltanolkotabi. “Phase Retrieval via Wirtinger Flow: Theory and Algorithms”. In: *IEEE Trans. Info. Theory* 61.4 (Apr. 2015), pp. 1985–2007. DOI: 10.1109/TIT.2015.2399924 (cit. on pp. 11, 21, 38, 50, 119).

[27] S. H. Chan, X. Wang, and O. A. Elgendy. “Plug-and-Play ADMM for Image Restoration: Fixed-Point Convergence and Applications”. In: *IEEE Transactions on Computational Imaging* 3.1 (2017), pp. 84–98. DOI: 10.1109/TCI.2016.2629286 (cit. on p. 39).

[28] R. Chandra, Z. Zhong, J. Hontz, V. McCulloch, C. Studer, and T. Goldstein. “PhasePack: A Phase Retrieval Library”. In: *Asil. Conf. Sig. Sys. Comp.* (2017), pp. 1617–1621 (cit. on p. 30).

[29] H. Chang and S. Marchesini. “Denoising Poisson phaseless measurements via orthogonal dictionary learning”. In: *Optics Express* 26.16 (Aug. 2018), 19773–96. DOI: 10.1364/OE.26.019773 (cit. on p. 12).

[30] H. Chang, Y. Lou, Y. Duan, and S. Marchesini. “Total variation–based phase retrieval for Poisson noise removal”. In: *SIAM J. Imag. Sci.* 11.1 (2018), pp. 24–55. DOI: 10.1137/16M1103270 (cit. on pp. 21, 35, 38).

[31] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang. “LEARN: Learned Experts’ Assessment-Based Reconstruction Network for Sparse-Data CT”. In: *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1333–1347. DOI: 10.1109/TMI.2018.2805692 (cit. on p. 53).

[32] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang. “Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network”. In: *IEEE Transactions on Medical Imaging* 36.12 (2017), pp. 2524–2535. DOI: 10.1109/TMI.2017.2715284 (cit. on p. 69).

[33] J. Chen, E. Garcia, R. Folks, *et al.*. “Onset of left ventricular mechanical contraction as determined by phase analysis of ECG-gated myocardial perfusion SPECT imaging: development of a diagnostic tool for assessment of cardiac mechanical dyssynchrony”. In: *J. Nucl. Cardiol.* 12.6 (2005), pp. 687–695. DOI: 10.1016/j.nuclcard.2005.06.088 (cit. on p. 13).

[34] J. Chen, J. Zhu, Y. W. Teh, and T. Zhang. “Stochastic Expectation Maximization with Variance Reduction”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. Curran Associates, Inc., 2018 (cit. on p. 109).

[35] J. Chen, J. Benesty, Y. Huang, and S. Doclo. “New insights into the noise reduction Wiener filter”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 14.4 (2006), pp. 1218–1234. DOI: 10.1109/TSA.2005.860851 (cit. on p. 6).

[36] S. Chen, K. Ma, and Y. Zheng. *Med3D: Transfer Learning for 3D Medical Image Analysis*. 2019. arXiv: 1904.00625 [cs.CV] (cit. on p. 104).

[37] Y. Chen and E. J. Candes. “Solving random quadratic systems of equations is nearly as easy as solving linear systems”. In: *Comm. Pure Appl. Math.* 70.5 (May 2017), 822–83. DOI: 10.1002/cpa.21638 (cit. on pp. 12, 21, 35, 38).

[38] S. R. Cherry, J. A. Sorenson, and M. E. Phelps. *Physics in nuclear medicine*. 2012 (cit. on pp. 13, 14, 20).

[39] K. Choi and A. D. Lanterman. “Phase retrieval from noisy data based on minimization of penalized I-divergence”. In: *J. Opt. Soc. Am. A* 24.1 (Jan. 2007), 34–49. DOI: 10.1364/JOSAA.24.000034 (cit. on p. 12).

[40] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot. “A Convex Approach for Image Restoration with Exact Poisson–Gaussian Likelihood”. In: *SIAM Journal on Imaging Sciences* 8.4 (2015), pp. 2662–2682. DOI: 10.1137/15M1014395 (cit. on pp. 41, 43).

[41] S. Y. Chun, J. A. Fessler, and Y. K. Dewaraja. “Non-local means methods using CT side information for I-131 SPECT image reconstruction”. In: *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. 2012, pp. 3362–3366. DOI: 10.1109/NSSMIC.2012.6551766 (cit. on p. 53).

[42] Y. Chun and J. A. Fessler. “Deep BCD-Net Using Identical Encoding-Decoding CNN Structures for Iterative Image Recovery”. In: *2018 IEEE IVMSP*. 2018, pp. 1–5. DOI: 10.1109/IVMSPW.2018.8448694 (cit. on p. 81).

[43] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye. “Diffusion Posterior Sampling for General Noisy Inverse Problems”. In: *The Eleventh International Conference on Learning Representations*. 2023 (cit. on pp. 40, 52).

[44] H. Chung, B. Sim, D. Ryu, and J. C. Ye. “Improving Diffusion Models for Inverse Problems using Manifold Constraints”. In: *Advances in Neural Information Processing Systems*. Vol. 35. 2022, pp. 25683–25696 (cit. on p. 40).

[45] H. Chung and J. C. Ye. “Score-based diffusion models for accelerated MRI”. In: *Medical Image Analysis* 80 (2022), p. 102479. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2022.102479 (cit. on p. 39).

[46] G. Corda-D’Incan, J. A. Schnabel, and A. J. Reader. “Memory-Efficient Training for Fully Unrolled Deep Learned PET Image Reconstruction With Iteration-Dependent Targets”. In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 6.5 (2022), pp. 552–563. DOI: 10.1109/TRPMS.2021.3101947 (cit. on p. 54).

[47] Z.-X. Cui, C. Cao, S. Liu, Q. Zhu, J. Cheng, H. Wang, Y. Zhu, and D. Liang. *Self-Score: Self-Supervised Learning on Score-Based Models for MRI Reconstruction*. 2022. arXiv: 2209.00835 [eess.IV] (cit. on p. 39).

[48] J. Dainty and J. Fienup. “Phase retrieval and image reconstruction for astronomy”. In: *Imag. Recov. Theory Appl.* 13 (Jan. 1987), pp. 231–275 (cit. on p. 9).

[49] I. Daubechies, M. Defrise, and C. De Mol. “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”. In: *Comm. Pure Appl. Math.* 57.11 (Nov. 2004), 1413–57. DOI: 10.1002/cpa.20042 (cit. on p. 28).

[50] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992. DOI: 10.1137/1.9781611970104. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611970104 (cit. on p. 39).

[51] J. de Leeuw and K. Lange. “Sharp quadratic majorization in one dimension”. In: *Comp. Stat. Data Anal.* 53.7 (May 2009), 2471–84. DOI: 10.1016/j.csda.2009.01.002 (cit. on p. 26).

[52] A. R. De Pierro. “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography”. In: *IEEE Trans. Med. Imag.* 14.1 (Mar. 1995), 132–7. DOI: 10.1109/42.370409 (cit. on p. 17).

[53] A. Defazio, F. Bach, and S. Lacoste-Julien. “SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives”. In: *Advances in Neural Information Processing Systems.* Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger. Vol. 27. Curran Associates, Inc., 2014 (cit. on p. 110).

[54] Y. Dewaraja, K. Koral, and J. Fessler. “Regularized reconstruction in quantitative SPECT using CT side information from hybrid imaging”. In: *Phys. Med. Biol.* 55.9 (2010), pp. 2523–2539. DOI: 10.1088/0031-9155/55/9/007 (cit. on p. 53).

[55] Y. Dewaraja, D. Mirando, A. Peterson, *et al.*. “A pipeline for automated voxel dosimetry: application in patients with multi-SPECT/CT imaging following 177Lu peptide receptor radionuclide therapy”. In: *J. Nucl. Med.* (2022). DOI: 10.2967/jnumed.121.263738 (cit. on p. 58).

[56] Y. Dewaraja, S. Wilderman, K. Koral, M. Kaminski, and A. Avram. “Use of integrated SPECT/CT imaging for tumor dosimetry in I-131 radioimmunotherapy: a pilot patient study”. In: *Cancer Biother Radiopharm* 24.4 (2009), pp. 417–426. DOI: 10.1089/cbr.2008.0568 (cit. on p. 13).

[57] P. Dhariwal and A. Nichol. “Diffusion Models Beat GANs on Image Synthesis”. In: *Advances in Neural Information Processing Systems.* Vol. 34. 2021, pp. 8780–8794 (cit. on pp. 39, 108).

[58] E. V. R. Di Bella, A. B. Barclay, R. L. Eisner, and R. W. Schafer. “A comparison of rotation-based methods for iterative reconstruction algorithms”. In: *IEEE Trans. Nuc. Sci.* 43.6 (Dec. 1996), 3370–6. DOI: 10.1109/23.552756 (cit. on p. 55).

[59] A. Dieudonné, R. Hobbs, R. Lebtahi, F. Maurel, *et al.*. In: *J. Nuc. Med.* 54.2 (2013). DOI: 10.2967/jnumed.112.105825 (cit. on p. 71).

[60] U. Eberlein, M. Cremonesi, and M. Lassmann. “Individualized Dosimetry for Theranostics: Necessary, Nice to Have, or Counterproductive?” In: *J. Nucl. Med.* 58 (2017). DOI: 10.2967/jnumed.116.186841 (cit. on p. 68).

[61] Y. Eldar. “Minimum variance in biased estimation: bounds and asymptotically optimal estimators”. In: *IEEE Trans. Sig. Proc.* 52.7 (2004), pp. 1915–1930. DOI: 10.1109/TSP.2004.828929 (cit. on p. 119).

[62] G. Fatima and P. Babu. “PGPAL: A monotonic iterative algorithm for Phase-Retrieval under the presence of Poisson-Gaussian noise”. In: *IEEE Sig. Proc. Letters* (2022), pp. 1–1. DOI: 10.1109/LSP.2022.3143469 (cit. on p. 39).

[63] G. Fatima, Z. Li, A. Arora, and P. Babu. “PDMM: A Novel Primal-Dual Majorization-Minimization Algorithm for Poisson Phase-Retrieval Problem”. In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 1241–1255. DOI: 10.1109/TSP.2022.3156014 (cit. on pp. 21, 38).

[64] J. A. Fessler and A. O. Hero. “Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms”. In: *IEEE Trans. Im. Proc.* 4.10 (Oct. 1995), 1417–29. DOI: 10.1109/83.465106 (cit. on p. 17).

[65] J. Fessler. “Penalized weighted least-squares image reconstruction for positron emission tomography”. In: *IEEE Trans. Med. Imaging* 13.2 (1994), pp. 290–300. DOI: 10.1109/42.293921 (cit. on p. 53).

[66] J. Fessler. *Michigan Image Reconstruction Toolbox* (cit. on p. 59).

[67] J. Fienup. “Phase Retrieval Algorithms: A Comparison”. In: *Appl. Opt.* 21 (1982) (cit. on p. 50).

[68] B. Gao and Z. Xu. “Phaseless recovery using the Gauss-Newton method”. In: *IEEE Trans. Sig. Proc.* 65.22 (Nov. 2017), 5885–96. DOI: 10.1109/tsp.2017.2742981 (cit. on pp. 12, 38).

[69] R. W. Gerchberg and W. O. Saxton. “A practical algorithm for the determination of phase from image and diffraction plane pictures”. In: *Optik* 35.2 (Apr. 1972), 237–46 (cit. on pp. 11, 119).

[70] R. W. Gerchberg and W. O. Saxton. “Practical Algorithm for Determination of Phase from Image and Diffraction Plane Pictures”. In: *OPTIK* 35.2 (1972), 237–246. ISSN: 0030-4026 (cit. on pp. 38, 50).

[71] S. Ghadimi and G. Lan. “Accelerated gradient methods for nonconvex nonlinear and stochastic programming”. In: *Math. Program.* 156 (2016). DOI: https://doi.org/10.1007/s10107-015-0871-8 (cit. on p. 124).

[72] A. Gnanasambandam and S. H. Chan. “Image Classification in the Dark Using Quanta Image Sensors”. In: *ECCV 2020*. Cham: Springer International Publishing, 2020, pp. 484–501 (cit. on pp. 12, 38).

[73] K. Gong, D. Wu, K. Kim, J. Yang, G. E. Fakhri, Y. Seo, and Q. Li. “EMnet: an unrolled deep neural network for PET image reconstruction”. In: *Med. Imag. 2019*. Vol. 10948. SPIE, 2019, pp. 1203–1208. DOI: 10.1117/12.2513096 (cit. on p. 81).

[74] K. Gong, D. Wu, K. Kim, J. Yang, T. Sun, G. E. Fakhri, Y. Seo, and Q. Li. “MAPEM-Net: an unrolled neural network for Fully 3D PET image reconstruction”. In: *Fully3D*. Vol. 11072. SPIE, 2019, pp. 109–113. DOI: 10.1117/12.2534904 (cit. on p. 81).

[75] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016 (cit. on p. 122).

[76] 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*. Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger. Vol. 27. Curran Associates, Inc., 2014 (cit. on p. 107).

[77] T. Götz, C. Schmidkonz, E. Lang, A. Maier, T. Kuwert, and P. Ritt. “A comparison of methods for adapting 177Lu dose-voxel-kernels to tissue inhomogeneities”. In: *Phys. Med. Bio.* 64.24 (2019). DOI: 10.1088/1361-6560/ab5b81 (cit. on p. 72).

[78] T. Götz, E. Lang, C. Schmidkonz, T. Kuwert, and B. Ludwig. “Dose voxel kernel prediction with neural networks for radiation dose estimation”. In: *Z. Med. Phys.* 31.1 (2021). DOI: 10.1016/j.zemedi.2020.09.005 (cit. on pp. 69, 80).

[79] T. Götz, C. Schmidkonz, S. Chen, S. Al-Baddai, T. Kuwert, and E. Lang. “A deep learning approach to radiation dose estimation”. In: *Phys. Med. Biol.* 65.3 (2020). DOI: 10.1088/1361-6560/ab65dc (cit. on pp. 69, 80).

[80] A. Goy, K. Arthur, S. Li, and G. Barbastathis. "Low photon count phase retrieval using deep learning". In: *Phys. Rev. Lett.* 121.24 (Dec. 2018), p. 243902. DOI: 10.1103/PhysRevLett.121.243902 (cit. on pp. 12, 38).

[81] L. Grafakos. *Classical and Modern Fourier Analysis.* Pearson/Prentice Hall, 2004 (cit. on p. 121).

[82] A. Graikos, N. Malkin, N. Jojic, and D. Samaras. "Diffusion Models as Plug-and-Play Priors". In: *Advances in Neural Information Processing Systems.* 2022 (cit. on pp. 39, 40).

[83] A. Grings, C. Jobic, T. Kuwert, *et al.*. In: *EJNMMI Phys.* (2022). DOI: 10.1186/s40658-022-00446-2 (cit. on p. 67).

[84] P. Grohs, S. Koppensteiner, and M. Rathmair. "Phase retrieval: uniqueness and stability". In: *SIAM Review* 62.2 (2020), 301–50. DOI: 10.1137/19M1256865 (cit. on p. 9).

[85] S. W. Hasinoff. "Photon, Poisson Noise". In: *Computer Vision: A Reference Guide.* Ed. by K. Ikeuchi. Boston, MA: Springer US, 2014, pp. 608–610. ISBN: 978-0-387-31439-6. DOI: 10.1007/978-0-387-31439-6_482 (cit. on p. 1).

[86] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* June 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90 (cit. on pp. 8, 60, 69).

[87] A. Hero, J. Fessler, and M. Usman. "Exploring estimator bias-variance tradeoffs using the uniform CR bound". In: *IEEE Trans. Sig. Proc.* 44.8 (1996), pp. 2026–2041. DOI: 10.1109/78.533723 (cit. on pp. 117, 118).

[88] J. Ho, A. Jain, and P. Abbeel. "Denoising Diffusion Probabilistic Models". In: 33 (2020). Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, pp. 6840–6851 (cit. on pp. 39, 45, 108).

[89] J. Hu, Z. Li, X. Xu, L. Shen, and J. A. Fessler. "Poisson-Gaussian Holographic Phase Retrieval with Score-based Image Prior". In: *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems.* 2023 (cit. on pp. 4, 38).

[90] P. J. Huber. *Robust statistics.* New York: Wiley, 1981 (cit. on pp. 25, 45).

[91] H. M. Hudson and R. S. Larkin. "Accelerated image reconstruction using ordered subsets of projection data". In: *IEEE Trans. Med. Imag.* 13.4 (Dec. 1994), 601–9. DOI: 10.1109/42.363108 (cit. on p. 16).

[92] H. Hudson and J. Ma. "Fisher's method of scoring in statistical image reconstruction: comparison of Jacobi and Gauss-Seidel iterative schemes". In: *Statistical Methods in Medical Research* 3.1 (1994), pp. 41–61. DOI: 10.1177/096228029400300104 (cit. on p. 23).

[93] D. R. Hunter and K. Lange. "A tutorial on MM algorithms". In: *American Statistician* 58.1 (Feb. 2004), 30–7. DOI: 10.1198/0003130042836 (cit. on p. 25).

[94] P. Jackson, J.-M. Beauregard, M. Hofman, T. Kron, A. Hogg, and R. Hicks. "An automated voxelized dosimetry tool for radionuclide therapy based on serial quantitative SPECT/CT imaging". In: *Med. Phys.* (2013). DOI: 10.1118/1.4824318 (cit. on p. 77).

[95] P. A. Jackson, M. S. Hofman, R. J. Hicks, M. Scalzo, and J. Violet. "Radiation Dosimetry in 177Lu-PSMA-617 Therapy Using a Single Posttreatment SPECT/CT Scan: A Novel Methodology to Generate Time- and Tissue-Specific Dose Factors". In: *Journal of Nuclear Medicine* 61.7 (2020), pp. 1030–1036. ISSN: 0161-5505. DOI: 10.2967/jnumed.119.233411 (cit. on p. 13).

[96] K. Jaganathan, Y. C. Eldar, and B. Hassibi. *Phase retrieval: an overview of recent developments*. 2015 (cit. on p. 9).

[97] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir. "Robust Compressed Sensing MRI with Deep Generative Priors". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 14938–14954 (cit. on p. 39).

[98] P. Jaming. "Phase retrieval techniques for radar ambiguity problems." In: *J. Four. Anal. Appl.* 5.4 (1999), pp. 309–329 (cit. on p. 9).

[99] L. Ji and Z. Tie. "On gradient descent algorithm for generalized phase retrieval problem". In: *2016 IEEE 13th International Conference on Signal Processing (ICSP)* (2016), pp. 320–325. DOI: 10.1109/ICSP.2016.7877848 (cit. on pp. 12, 21, 31, 38).

[100] Y. Jia, Z. Li, A. Akhavanallaf, J. A. Fessler, and Y. K. Dewaraja. "90Y SPECT scatter estimation and voxel dosimetry in radioembolization using a unified deep learning framework". In: *EJNMMI Phys.* (2023). DOI: https://doi.org/10.1186/s40658-023-00598-9 (cit. on p. 15).

[101] Y. Jia, Z. Li, J. Fessler, and Y. Dewaraja. "Y-90 Bremsstrahlung SPECT Scatter Estimation and Voxel Dosimetry Using a Deep Learning Pipeline". In: *Journal of Nuclear Medicine* 64.supplement 1 (2023), P1211–P1211. ISSN: 0161-5505. eprint: https://jnm.snmjournals.org/content (cit. on p. 15).

[102] X. Jiang, S. Rajan, and X. Liu. "Wirtinger flow method with optimal stepsize for phase retrieval". In: *IEEE Signal Proc. Letters* 23.11 (Nov. 2016), 1627–31. DOI: 10.1109/LSP.2016.2611940 (cit. on pp. 11, 21, 31, 33, 38).

[103] R. Johnson and T. Zhang. "Accelerating Stochastic Gradient Descent using Predictive Variance Reduction". In: *Advances in Neural Information Processing Systems*. Vol. 26. 2013 (cit. on p. 110).

[104] J. Kaipioa and E. Somersalo. "Statistical inverse problems: Discretization, model reduction and inverse crimes". In: *J. Comp. Appl. Math.* 198.2 (Jan. 2007), 493–504. DOI: 10.1016/j.cam.2005.09.027 (cit. on p. 81).

[105] U. S. Kamilov, H. Mansour, and B. Wohlberg. "A Plug-and-Play Priors Approach for Solving Nonlinear Imaging Inverse Problems". In: *IEEE Signal. Proc. Let.* 24.12 (Dec. 2017), pp. 1872–1876 (cit. on pp. 45, 49).

[106] U. S. Kamilov, C. B. Bouman, G. T. Buzzard, and B. Wohlberg. "Plug-and-Play Methods for Integrating Physical and Learned Models in Computational Imaging". In: *IEEE Signal Process. Mag.* 40.1 (Jan. 2023), pp. 85–97 (cit. on p. 39).

[107] I. Kang, F. Zhang, and G. Barbastathis. "Phase extraction neural network (PhENN) with coherent modulation imaging (CMI) for phase retrieval at low photon counts". In: *Optics Express* 28.15 (July 2020), 21578–600. DOI: 10.1364/OE.397430 (cit. on p. 12).

[108] L. Kaufman. "Implementing and Accelerating the EM Algorithm for Positron Emission Tomography". In: *IEEE Transactions on Medical Imaging* 6.1 (1987), pp. 37–51. DOI: 10.1109/TMI.1987.4307796 (cit. on p. 110).

[109] A. Kendall, V. Badrinarayanan, and R. Cipolla. "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding". In: *Proceedings of the British Machine Vision Conference (BMVC)*. 2017, pp. 57.1–57.12. DOI: 10.5244/C.31.57 (cit. on p. 78).

[110] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". In: *ACM Transactions on Graphics* 42.4 (July 2023) (cit. on p. 87).

[111] Ž. Kereta, R. Twyman, S. Arridge, K. Thielemans, and B. Jin. "Stochastic EM methods with variance reduction for penalised PET reconstructions". In: *Inverse Problems* 37.11 (2021), p. 115006. DOI: 10.1088/1361-6420/ac2d74 (cit. on p. 109).

[112] D. Kim and J. A. Fessler. "Adaptive Restart of the Optimized Gradient Method for Convex Optimization". In: *Journal of Optimization Theory and Applications* 178.1 (July 2018), pp. 240–63. DOI: 10.1007/s10957-018-1287-4 (cit. on p. 28).

[113] H. Kim, Z. Li, J. Son, J. A. Fessler, Y. K. Dewaraja, and S. Y. Chun. "Physics-Guided Deep Scatter Estimation by Weak Supervision for Quantitative SPECT". In: *IEEE Transactions on Medical Imaging* 42.10 (2023), pp. 2961–2973. DOI: 10.1109/TMI.2023.3270868 (cit. on p. 15).

[114] K. Kim, D. Wu, K. Gong, J. Dutta, J. H. Kim, Y. D. Son, H. K. Kim, G. El Fakhri, and Q. Li. "Penalized PET Reconstruction Using Deep Learning Prior and Local Linear Fitting". In: *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1478–1487. DOI: 10.1109/TMI.2018.2832613 (cit. on p. 53).

[115] D. Kingma and J. Ba. "Adam: A method for stochastic optimization". In: *IEEE International Conference on Learning Representations (ICLR)*. 2015 (cit. on pp. 46, 73).

[116] D. P. Kingma and M. Welling. "Auto-Encoding Variational Bayes". In: *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. 2014 (cit. on p. 107).

[117] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. *Segment Anything*. 2023. arXiv: 2304.02643 [cs.CV] (cit. on p. 103).

[118] F. Knoll. "Rise of the machines (in MR image reconstruction)". In: *IMA workshop on Computational Imaging*. 2019 (cit. on p. 66).

[119] S. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. Maier-Hein, S. M. A. Eslami, D. Jimenez Rezende, and O. Ronneberger. "A Probabilistic U-Net for Segmentation of Ambiguous Images". In: *Advances in Neural Information Processing Systems*. Vol. 31. 2018 (cit. on p. 78).

[120] F. Kunstner, P. Hennig, and L. Balles. "Limitations of the empirical Fisher approximation for natural gradient descent". In: *Adv. Neural Info. Proc. Sys.* Vol. 32. 2019 (cit. on p. 36).

[121] B. Lakshminarayanan, A. Pritzel, and C. Blundell. "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles". In: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, pp. 6405–6416 (cit. on p. 78).

[122] D. Lalush and B. Tsui. "A generalized Gibbs prior for maximum a posteriori reconstruction in SPECT". In: *Phys. Med. Biol.* 38.6 (1993), pp. 729–741. DOI: 10.1088/0031-9155/38/6/007 (cit. on p. 53).

[123] K. Lange. "A gradient algorithm locally equivalent to the EM Algorithm". In: *J. Royal Stat. Soc. Ser. B* 57.2 (1995), 425–37 (cit. on p. 23).

[124] T. Latychevskaia. "Iterative phase retrieval in coherent diffractive imaging: practical issues". In: *Appl. Opt.* 57.25 (2018), pp. 7187–7197. DOI: 10.1364/AO.57.007187 (cit. on p. 9).

[125] 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: Second International Workshop, UNSURE 2020, and Third International Workshop, GRAIL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings.* 2020, pp. 81–96. DOI: 10.1007/978-3-030-60365-6_9 (cit. on p. 78).

[126] H. Lawrence, D. A. Barmherzig, H. Li, M. Eickenberg, and M. Gabrie. *Phase retrieval with holography and untrained priors: Tackling the challenges of low-photon nanoscale imaging.* To appear in JMLR. 2020 (cit. on pp. 12, 38, 41, 45).

[127] M. Lee, J. Kim, J. Paeng, K. Kang, J. Jeong, D. Lee, and J. Lee. "Whole-body voxel-based personalized dosimetry: The multiple voxel s-value approach for heterogeneous media with nonuniform activity distributions". In: *J. Nuc. Med.* 59.7 (2018). DOI: 10.2967/jnumed.117.201095 (cit. on p. 72).

[128] M. Lee, D. Hwang, J. Kim, and J. Lee. "Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry". In: *Sci. Rep.* 9.1 (2019). DOI: 10.1038/s41598-019-46620-y (cit. on pp. 69, 80).

[129] S. Lee, H. Chung, J. Kim, and J. C. Ye. *Progressive Deblurring of Diffusion Models for Coarse-to-Fine Image Synthesis.* 2022. arXiv: 2207.11192 [cs.CV] (cit. on p. 39).

[130] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. "Noise2Noise: Learning Image Restoration without Clean Data". In: *Proceedings of the 35th International Conference on Machine Learning.* Vol. 80. Proceedings of Machine Learning Research. 2018, pp. 2965–2974 (cit. on pp. 39, 105).

[131] H. Li and Z. Lin. "Accelerated Proximal Gradient Methods for Nonconvex Programming". In: *Advances in Neural Information Processing Systems.* Vol. 28. 2015 (cit. on pp. 44, 124).

[132] Z. Li, J. A. Fessler, J. K. Mikell, S. J. Wilderman, and Y. K. Dewaraja. "DblurDoseNet: A deep residual learning network for voxel radionuclide dosimetry compensating for SPECT imaging resolution". In: *Med. Phys.* 49.2 (Feb. 2022), 1216–30. DOI: 10.1002/mp.15397 (cit. on pp. 4, 45, 53, 68, 111).

[133] Z. Li, Y. K. Dewaraja, and J. A. Fessler. "ESR-Net: An Efficient Image Super-resolution Network for SPECT Reconstruction". In: *2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).* 2022 (cit. on p. 81).

[134] Z. Li, Y. K. Dewaraja, and J. A. Fessler. "Training End-to-End Unrolled Iterative Neural Networks for SPECT Image Reconstruction". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 7.4 (2023), pp. 410–420. DOI: 10.1109/TRPMS.2023.3240934 (cit. on pp. 4, 53, 111).

[135] Z. Li and J. A. Fessler. "Uniform Cramér–Rao Lower Bound Analysis for Phase Retrieval Algorithms". In: *2022 Asilomar Conference on Signals, Systems, and Computers.* 2022 (cit. on pp. 38, 117).

[136] Z. Li, J. A. Fessler, J. K. Mikell, S. J. Wilderman, and Y. K. Dewaraja. "A Deep Residual Learning Network for Practical Voxel Dosimetry in Radionuclide Therapy". In: *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).* 2020, pp. 1–4. DOI: 10.1109/NSS/MIC42677.2020.9507764 (cit. on p. 4).

[137] Z. Li, J. Hu, X. Xu, L. Shen, and J. A. Fessler. *Poisson-Gaussian Holographic Phase Retrieval with Score-based Image Prior.* 2023. arXiv: 2305.07712 (cit. on pp. 4, 21, 38, 45, 111).

[138] Z. Li, K. Lange, and J. A. Fessler. "Poisson Phase Retrieval With Wirtinger Flow". In: *2021 IEEE International Conference on Image Processing (ICIP).* 2021, pp. 2828–2832. DOI: 10.1109/ICIP42928.2021.9506139 (cit. on pp. 4, 21, 111).

[139] Z. Li, K. Lange, and J. A. Fessler. "Poisson Phase Retrieval in Very Low-Count Regimes". In: *IEEE Transactions on Computational Imaging* 8 (2022), pp. 838–850. DOI: 10.1109/TCI.2022.3209936 (cit. on pp. 4, 9, 21, 38, 111).

[140] Z. Li, X. Xu, J. Hu, J. A. Fessler, and Y. K. Dewaraja. "Reducing SPECT acquisition time by predicting missing projections with single-scan self-supervised coordinate-based learning". In: *Journal of Nuclear Medicine* 64 (supplement 1) (cit. on pp. 5, 53, 111).

[141] J. Liang, P. Stoica, Y. Jing, and J. Li. "Phase Retrieval via the Alternating Direction Method of Multipliers". In: *IEEE Sig. Proc. Letters* 25.1 (2018), pp. 5–9. DOI: 10.1109/LSP.2017.2767826 (cit. on pp. 12, 38, 50, 117, 119).

[142] H. Lim, I. Y. Chun, Y. K. Dewaraja, and J. A. Fessler. "Improved Low-Count Quantitative PET Reconstruction With an Iterative Neural Network". In: *IEEE Trans. Med. Imaging* 39.11 (2020), pp. 3512–3522. DOI: 10.1109/TMI.2020.2998480 (cit. on pp. 18, 54, 57, 66, 81, 82, 97).

[143] Z. Liu, P. Luo, X. Wang, and X. Tang. "Deep Learning Face Attributes in the Wild". In: *Proceedings of International Conference on Computer Vision (ICCV).* Dec. 2015 (cit. on pp. 45, 47–49, 51).

[144] M. Ljungberg. "The SIMIND Monte Carlo Program". In: (2012). DOI: 10.1201/b13073-8 (cit. on pp. 58, 70).

[145] Y. Long, J. A. Fessler, and J. M. Balter. "A 3D forward and back-projection method for X-ray CT using separable footprint". In: *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med.* Winner of poster award. 2009, 146–9 (cit. on p. 39).

[146] I. Loshchilov and F. Hutter. "Decoupled Weight Decay Regularization". In: *ICLR.* 2019 (cit. on p. 60).

[147] D. G. Luenberger. *Optimization by Vector Space Methods.* 1st. USA, 1997. ISBN: 047155359X (cit. on p. 121).

[148] W. Luo, W. Alghamdi, and Y. M. Lu. "Optimal Spectral Initialization for Signal Recovery With Applications to Phase Retrieval". In: *IEEE Trans. Sig. Proc.* 67.9 (2019), pp. 2347–2356. DOI: 10.1109/TSP.2019.2904918 (cit. on pp. 29, 34, 45).

[149] Y. Luo, W. Huang, X. Li, and A. R. Zhang. *Recursive importance sketching for rank constrained least squares: algorithms and high-order convergence.* 2022 (cit. on p. 38).

[150] M. Makitalo and A. Foi. "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising". In: *IEEE Trans. Im. Proc.* 20.1 (Jan. 2011), 99–109. DOI: 10.1109/TIP.2010.2056693 (cit. on p. 13).

[151] M. Makitalo and A. Foi. "Optimal Inversion of the Generalized Anscombe Transformation for Poisson-Gaussian Noise". In: *IEEE Trans. Imag. Proc.* 22.1 (2013), pp. 91–103. DOI: 10.1109/TIP.2012.2202675 (cit. on p. 39).

[152] S. Marchesini. *Ptychography Gold Ball Example Dataset.* July 2017. DOI: 10.11577/1454414 (cit. on pp. 31, 32).

[153] G. Mariani, L. Bruselli, T. Kuwert, *et al.*. "A review on the clinical uses of SPECT/CT". In: *Eur. J. Nucl. Med. Mol. Imaging* 37.10 (2010). DOI: doi:10.1007/s00259-010-1390-8 (cit. on p. 2).

[154] M. Masias, J. Freixenet, X. Lladó, and M. Peracaula. "A review of source detection approaches in astronomical images". In: *Monthly Notices of the Royal Astronomical Society* 422.2 (Apr. 2012), pp. 1674–1689. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.20742.x. eprint: https://academic.oup.com/mnras/article-pdf/422/2/1674/3511408/mnras0422-1674.pdf (cit. on p. 1).

[155] A. Mehranian and A. J. Reader. "Model-Based Deep Learning PET Image Reconstruction Using Forward-Backward Splitting Expectation Maximisation". In: *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).* 2019, pp. 1–4. DOI: 10.1109/NSS/MIC42101.2019.9059998 (cit. on pp. 53, 54, 57, 59).

[156] A. Mehranian and A. J. Reader. "Model-Based Deep Learning PET Image Reconstruction Using Forward–Backward Splitting Expectation–Maximization". In: *IEEE Trans. Rad. Plas. Med. Sci.* 5.1 (2021), pp. 54–64. DOI: 10.1109/TRPMS.2020.3004408 (cit. on pp. 66, 81).

[157] M. Melis, E. Krenning, B. Bernard, *et al.*. "Localisation and mechanism of renal retention of radiolabelled somatostatin analogues". In: *Eur. J. Nucl. Med. Mol. Imaging* 32.10 (2005). DOI: 10.1007/s00259-005-1793-0 (cit. on p. 80).

[158] L. Meng. *Method for Computation of the Fisher Information Matrix in the Expectation-Maximization Algorithm.* 2016. DOI: 10.48550/ARXIV.1608.01734 (cit. on p. 23).

[159] C. A. Metzler, M. K. Sharma, S. Nagesh, R. G. Baraniuk, O. Cossairt, and A. Veeraraghavan. "Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset". In: *Proc. Intl. Conf. Comp. Photography.* 2017, 1–16. DOI: 10.1109/ICCPHOT.2017.7951483 (cit. on pp. 30–32).

[160] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *Commun. ACM* 65.1 (2021), pp. 99–106. ISSN: 0001-0782. DOI: 10.1145/3503250 (cit. on p. 84).

[161] R. P. Millane. "Phase retrieval in crystallography and optics". In: *J. Opt. Soc. Am. A* 7.3 (Mar. 1990), 394–411. DOI: 10.1364/JOSAA.7.000394 (cit. on p. 9).

[162] J. Minnema, A. Ernst, E. M. van, *et al.*. "A review on the application of deep learning for CT reconstruction, bone segmentation and surgical planning in oral and maxillofacial surgery". In: *Dentomaxillofac Radiol* (2022). DOI: doi:10.1259/dmfr.20210437 (cit. on p. 53).

[163] P. K. Mogensen and A. N. Riseth. "Optim: A mathematical optimization package for Julia". In: *J. of Open Source Software* 3.24 (2018), p. 615. DOI: 10.21105/joss.00615 (cit. on p. 31).

[164] S. Mostafapour, F. Gholamiankhah, S. Maroufpour, M. Momennezhad, M. Asadinezhad, S. R. Zakavi, H. Arabi, and H. Zaidi. "Deep learning-guided attenuation correction in the image domain for myocardial perfusion SPECT imaging". In: *J. Comp. Design and Engineering* 9.2 (Mar. 2022), pp. 434–447. DOI: 10.1093/jcde/qwac008 (cit. on p. 54).

[165] S. Mukherjee, M. Carioni, O. Öktem, and C.-B. Schönlieb. "End-to-end reconstruction meets data-driven regularization for inverse problems". In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan. Vol. 34. Curran Associates, Inc., 2021, pp. 21413–21425 (cit. on p. 54).

[166] T. Müller, A. Evans, C. Schied, and A. Keller. "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding". In: *ACM Trans. Graph.* 41.4 (July 2022), 102:1–102:15. DOI: 10.1145/3528223.3530127 (cit. on p. 87).

[167] G. Nataraj and R. Otazo. "Investigating Robustness to Unseen Pathologies in Model-Free Deep Multicoil Reconstruction". In: *Proceedings of the 2020 ISMRM Workshop on Sampling and Reconstruction.* 2020 (cit. on p. 9).

[168] P. Netrapalli, P. Jain, and S. Sanghavi. "Phase Retrieval Using Alternating Minimization". In: *IEEE Trans. Sig. Proc.* 63.18 (2015), pp. 4814–4826. DOI: 10.1109/TSP.2015.2448516 (cit. on pp. 12, 38).

[169] G. Ongie, A. Jalal, C. A. Metzler, R. Baraniuk, A. G. Dimakis, and R. M. Willett. "Deep Learning Techniques for Inverse Problems in Imaging". In: *IEEE Journal on Selected Areas in Information Theory* 1 (2020), pp. 39–56 (cit. on p. 39).

[170] M. R. Osborne. "Fisher's Method of Scoring". In: *International Statistical Review* 60.1 (1992), pp. 99–117. ISSN: 03067734, 17515823 (cit. on p. 23).

[171] B. Ozturkler, A. Sahiner, M. Pilanci, S. Vasanawala, J. Pauly, and M. Mardani. "Scalable and interpretable neural MRI reconstruction via layer-wise training". In: *International Society for Magnetic Resonance in Medicine.* 2021 (cit. on p. 54).

[172] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsaftaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, K. Huang, K. S. Nikita, B. P. Veasey, M. Zervakis, J. H. Saltz, and C. S. Pattichis. "AI in Medical Imaging Informatics: Current Challenges and Future Directions". In: *IEEE Journal of Biomedical and Health Informatics* 24.7 (2020), pp. 1837–1857. DOI: 10.1109/JBHI.2020.2991043 (cit. on p. 1).

[173] V. Panin, G. Zeng, and G. Gullberg. "Total variation regulated EM algorithm [SPECT reconstruction]". In: *IEEE Trans. Nuc. Sci.* 46.6 (1999), pp. 2202–2210. DOI: 10.1109/23.819305 (cit. on p. 53).

[174] M. Perri, P. Erba, and D. Volterrani. "Octreo-SPECT/CT imaging for accurate detection and localization of suspected neuroendocrine tumors". In: *Q. J. Nucl. Med. Mol. Imaging* 52.4 (2008), pp. 323–333 (cit. on p. 13).

[175] P. Pretorius and M. King. "Diminishing the impact of the partial volume effect in cardiac SPECT perfusion imaging". In: *Med. Phys.* 36.1 (2009). DOI: 10.1118/1.3031110 (cit. on p. 67).

[176] C. Qian, N. D. Sidiropoulos, K. Huang, L. Huang, and H. C. So. "Phase Retrieval Using Feasible Point Pursuit: Algorithms and Cramér–Rao Bound". In: *IEEE Trans. Sig. Proc.* 64.20 (2016), pp. 5282–5296. DOI: 10.1109/TSP.2016.2593688 (cit. on pp. 117, 118).

[177] T. Qiu, P. Babu, and D. P. Palomar. "PRIME: phase retrieval via majorization-minimization". In: *IEEE Trans. Sig. Proc.* 64.19 (Oct. 2016), 5174–86. DOI: 10.1109/TSP.2016.2585084 (cit. on pp. 11, 12, 21, 31, 38, 48, 50, 119).

[178] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong. "Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss". In: *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1488–1497. DOI: 10.1109/TMI.2018.2820120 (cit. on p. 53).

[179] K. Raza and N. K. Singh. "A Tour of Unsupervised Deep Learning for Medical Image Analysis". In: *Curr. Med. Imaging* 17.9 (2021). DOI: 10.2174/1573405617666210127154257 (cit. on p. 104).

[180] A. J. Reader, G. Corda, A. Mehranian, C. d. Costa-Luis, S. Ellis, and J. A. Schnabel. "Deep Learning for PET Image Reconstruction". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.1 (2021), pp. 1–25. DOI: 10.1109/TRPMS.2020.3014786 (cit. on p. 53).

[181] T. Remez, O. Litany, R. Giryes, and A. M. Bronstein. "Class-Aware Fully Convolutional Gaussian and Poisson Denoising". In: *IEEE Trans. Image Proc.* 27.11 (2018), pp. 5707–5722. DOI: 10.1109/TIP.2018.2859044 (cit. on pp. 37, 38).

[182] Y. Romano, M. Elad, and P. Milanfar. "The Little Engine That Could: Regularization by Denoising (RED)". In: *SIAM Journal on Imaging Sciences* 10.4 (2017), pp. 1804–44. DOI: 10.1137/16M1102884 (cit. on pp. 39, 45, 49).

[183] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. "High-resolution image synthesis with latent diffusion models". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 10684–10695 (cit. on p. 109).

[184] O. Ronneberger, P. Fischer, and T. Brox. "U-net: convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention*. 2015, 234–41. DOI: 10.1007/978-3-319-24574-4_28 (cit. on pp. 8, 69, 72).

[185] T. Rydén, E. M. Van, I. Marin, J. Svensson, and P. Bernhardt. "Deep-Learning Generation of Synthetic Intermediate Projections Improves 177Lu SPECT Images Reconstructed with Sparsely Acquired Projections". In: *J. Nucl. Med.* 62.4 (2021). DOI: 10.2967/jnumed.120.245548 (cit. on pp. 84–86, 99).

[186] A. Sahiner, M. Mardani, B. Ozturkler, M. Pilanci, and J. M. Pauly. "Convex Regularization behind Neural Reconstruction". In: *International Conference on Learning Representations*. 2021 (cit. on p. 54).

[187] B. G. Sakelaris, Z. Li, J. Sun, S. Banerjee, V. Booth, and E. Gourgou. "Modelling learning in Caenorhabditis elegans chemosensory and locomotive circuitry for T-maze navigation". In: *European Journal of Neuroscience* 55.2 (2022), pp. 354–376. DOI: https://doi.org/10.1111/ejn.15560. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.15560 (cit. on p. 3).

[188] D. Sarrut, A. Halty, J. Badel, L. Ferrer, and M. Bardiès. "Voxel-based multimodel fitting method for modeling time activity curves in SPECT images". In: *Med. Phys.* 44.12 (2017). DOI: 10.1002/mp.12586 (cit. on p. 77).

[189] W. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. Tsui. "4D XCAT phantom for multimodality imaging research". In: *Med. Phys.* 37.9 (2010), pp. 4902–4915. DOI: 10.1118/1.3480985 (cit. on pp. 57, 79, 90).

[190] R. Seifert, M. Weber, E. Kocakavuk, C. Rischpler, and D. Kersting. "Artificial Intelligence and Machine Learning in Nuclear Medicine: Future Perspectives". In: *Semin. Nucl. Med.* 51.2 (2021). DOI: 10.1053/j.semnuclmed.2020.08.003 (cit. on p. 69).

[191] J. Sempau, S. Wilderman, and A. Bielajew. "DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations". In: *Phys. Med. Bio.* 45.8 (2000). DOI: 10.1088/0031-9155/45/8/315 (cit. on p. 71).

[192] W. Shao, S. Rowe, and Y. Du. "SPECTnet: a deep learning neural network for SPECT image reconstruction". In: *Ann Transl Med* 9.9 (2021). DOI: 10.21037/atm-20-3345 (cit. on pp. 53, 54).

[193] W. Shao, M. G. Pomper, and Y. Du. "A Learned Reconstruction Network for SPECT Imaging". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.1 (2021), pp. 26–34. DOI: 10.1109/TRPMS.2020.2994041 (cit. on pp. 53, 54).

[194] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. "Phase Retrieval with Application to Optical Imaging: A contemporary overview". In: *IEEE Sig. Proc. Mag.* 32.3 (2015), pp. 87–109. DOI: 10.1109/MSP.2014.2352673 (cit. on pp. 11, 38, 50).

[195] R. Sheikh and T. Schultz. "Unsupervised Domain Adaptation for Medical Image Segmentation via Self-Training of Early Features". In: *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*. Ed. by E. Konukoglu, B. Menze, A. Venkataraman, C. Baumgartner, Q. Dou, and S. Albarqouni. Vol. 172. Proceedings of Machine Learning Research. PMLR, July 2022, pp. 1096–1107 (cit. on p. 106).

[196] G. Shen, K. Dwivedi, K. Majima, T. Horikawa, and Y. Kamitani. "End-to-End Deep Image Reconstruction From Human Brain Activity". In: *Front. Comput. Neurosci.* (2019). DOI: 10.3389/fncom.2019.00021 (cit. on p. 54).

[197] L. A. Shepp and Y. Vardi. "Maximum likelihood reconstruction for emission tomography". In: *IEEE Trans. Med. Imag.* 1.2 (Oct. 1982), 113–22. DOI: 10.1109/TMI.1982.4307558 (cit. on p. 16).

[198] S. Shoushtari, J. Liu, and U. S. Kamilov. *DOLPH: Diffusion Models for Phase Retrieval.* 2022. arXiv: 2211.00529 [eess.IV] (cit. on pp. 39, 40, 45, 49, 50).

[199] D. L. Snyder, A. M. Hammoud, and R. L. White. "Image recovery from data acquired with a charge-coupled-device camera". In: *J. Opt. Soc. Am. A* 10.5 (May 1993), 1014–23. DOI: 10.1364/JOSAA.10.001014 (cit. on pp. 13, 39).

[200] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White. "Compensation for readout noise in CCD images". In: *J. Opt. Soc. Am. A* 12.2 (Feb. 1995), 272–83. DOI: 10.1364/JOSAA.12.000272 (cit. on p. 9).

[201] A. Soderlund, J. Chaal, G. Tjio, *et al.*. "Beyond 18F-FDG: Characterization of PET/CT and PET/MR Scanners for a Comprehensive Set of Positron Emitters of Growing Application–18F, 11C, 89Zr, 124I, 68Ga, and 90Y". In: *J. Nucl. Med.* 56.8 (2015), pp. 1285–1291. DOI: 10.2967/jnumed.115.156711 (cit. on p. 58).

[202] A. Soderlund, J. Chaal, G. Tjio, J. Totman, M. Conti, and D. Townsend. “Beyond 18F-FDG: Characterization of PET/CT and PET/MR scanners for a comprehensive set of positron emitters of growing application—18F, 11C, 89Zr, 124I, 68Ga, and 90Y”. In: *J. Nuc. Med.* 56.8 (2015). DOI: 10.2967/jnumed.115.156711 (cit. on p. 70).

[203] M. Soltanolkotabi. “Structured Signal Recovery From Quadratic Measurements: Breaking Sample Complexity Barriers via Nonconvex Optimization”. In: *IEEE Trans. Info. Theory* 65.4 (2019), pp. 2374–2400. DOI: 10.1109/TIT.2019.2891653 (cit. on pp. 11, 38).

[204] Y. Song and S. Ermon. “Generative Modeling by Estimating Gradients of the Data Distribution”. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019 (cit. on pp. 39, 40).

[205] Y. Song, L. Shen, L. Xing, and S. Ermon. “Solving Inverse Problems in Medical Imaging with Score-Based Generative Models”. In: *International Conference on Learning Representations*. 2022 (cit. on pp. 39, 40).

[206] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. 2021 (cit. on pp. 39, 40, 45, 108).

[207] J. Sun, Q. Qu, and J. Wright. “A geometric analysis of phase retrieval”. In: *2016 IEEE International Symposium on Information Theory (ISIT)*. 2016, pp. 2379–2383. DOI: 10.1109/ISIT.2016.7541725 (cit. on p. 38).

[208] Y. Sun, J. Liu, M. Xie, B. Wohlberg, and U. S. Kamilov. “CoIL: Coordinate-Based Internal Learning for Tomographic Imaging”. In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 1400–1412. DOI: 10.1109/TCI.2021.3125564 (cit. on p. 84).

[209] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng. “Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains”. In: *NeurIPS* (2020) (cit. on p. 87).

[210] P. Thibault and M. Guizar-Sicairos. “Maximum-likelihood refinement for coherent diffractive imaging”. In: *New J. of Phys.* 14.6 (June 2012), p. 063004. DOI: 10.1088/1367-2630/14/6/063004 (cit. on pp. 12, 38).

[211] B. Thomas, V. Cuplov, A. Bousse, *et al.*. “PETPVC: a toolbox for performing partial volume correction techniques in positron emission tomography”. In: *Phys. Med. Biol.* 61.22 (2016). DOI: 10.1088/0031-9155/61/22/7975 (cit. on p. 79).

[212] X. Tian. “Fourier ptychographic reconstruction using mixed Gaussian-Poisson likelihood with total variation regularisation”. In: *Electronics Letters* 55.19 (Sept. 2019), 1041–3. DOI: 10.1049/el.2019.1141 (cit. on pp. 9, 13, 39).

[213] D. M. Titterington. “Recursive parameter estimation using incomplete data”. In: *J. Royal Stat. Soc. Ser. B* 46.2 (1984), 257–67 (cit. on p. 23).

[214] J. Tran-Gia and M. Lassmann. “Characterization of Noise and Resolution for Quantitative 177Lu SPECT/CT with xSPECT Quant”. In: *J. Nucl. Med.* 60.1 (2019). DOI: 10.2967/jnumed.118.211094 (cit. on p. 79).

[215] C. Uribe, S. Mathotaarachchi, V. Gaudet, *et al.*. “Machine Learning in Nuclear Medicine: Part 1-Introduction”. In: 60.4 (2019). DOI: 10.2967/jnumed.118.223495 (cit. on p. 69).

[216] I. Vazquez, I. E. Harmon, J. C. R. Luna, and M. Das. "Quantitative phase retrieval with low photon counts using an energy resolving quantum detector". In: *J. Opt. Soc. Am. A* 38.1 (Jan. 2021), 71–9. DOI: 10.1364/JOSAA.396717 (cit. on pp. 12, 38).

[217] P. Veit-Haibach, I. Buvat, and K. Herrmann. "EJNMMI supplement: bringing AI and radiomics to nuclear medicine". In: *Eur. J. Nucl. Med. Mol. Imaging* 46.13 (2019). DOI: 10.1007/s00259-019-04395-4 (cit. on p. 69).

[218] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg. "Plug-and-Play Priors for Model Based Reconstruction". In: *Proc. IEEE Global Conf. Signal Process. and Inf. Process.* Austin, TX, USA, Dec. 2013, pp. 945–948 (cit. on pp. 45, 49).

[219] P. Vincent. "A Connection Between Score Matching and Denoising Autoencoders". In: *Neural Computation* 23.7 (2011), pp. 1661–1674. DOI: 10.1162/NECO_a_00142 (cit. on p. 40).

[220] I. Waldspurger. "Phase Retrieval With Random Gaussian Sensing Vectors by Alternating Projections". In: *IEEE Trans. Info. Theory* 64.5 (2018), pp. 3301–3312. DOI: 10.1109/TIT.2018.2800663 (cit. on pp. 12, 38).

[221] L. Wang. "The Poisson Channel With Varying Dark Current Known to the Transmitter". In: *IEEE Transactions on Information Theory* 65.8 (2019), pp. 4966–4978. DOI: 10.1109/TIT.2019.2911474 (cit. on p. 38).

[222] Y. Wang, W. Yin, and J. Zeng. "Global Convergence of ADMM in Nonconvex Nonsmooth Optimization". In: *J. Sci. Comput.* 78.1 (Jan. 2019), pp. 29–63. ISSN: 0885-7474. DOI: 10.1007/s10915-018-0757-z (cit. on p. 52).

[223] Z. Wang, J. Liu, G. Li, and H. Han. "Blind2Unblind: Self-Supervised Image Denoising with Visible Blind Spots". In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022, pp. 2017–2026. DOI: 10.1109/CVPR52688.2022.00207 (cit. on p. 39).

[224] Z. Wang, Y. Jiang, H. Zheng, P. Wang, P. He, Z. Wang, W. Chen, and M. Zhou. "Patch Diffusion: Faster and More Data-Efficient Training of Diffusion Models". In: *arXiv preprint arXiv:2304.12526* (2023) (cit. on p. 106).

[225] X. Wei, H. van Gorp, L. Gonzalez-Carabarin, D. Freedman, Y. C. Eldar, and R. J. G. van Sloun. "Deep Unfolding With Normalizing Flow Priors for Inverse Problems". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 2962–2971. DOI: 10.1109/TSP.2022.3179807 (cit. on p. 39).

[226] F. Wen, R. Ying, Y. Liu, P. Liu, and T.-K. Truong. "A Simple Local Minimal Intensity Prior and an Improved Algorithm for Blind Image Deblurring". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.8 (2021), pp. 2923–2937. DOI: 10.1109/TCSVT.2020.3034137 (cit. on p. 84).

[227] K. Wiess, T. M. Khoshgoftaar, and D. Wang. "A survey of transfer learning". In: *J. Big Data* 3.9 (2016). DOI: https://doi.org/10.1186/s40537-016-0043-6 (cit. on p. 102).

[228] S. Wilderman and Y. Dewaraja. "Method for fast CT/SPECT-based 3D Monte Carlo absorbed dose computations in internal emitter therapy". In: *IEEE Trans. Nuc. Sci.* 54.1 (2007). DOI: 10.1109/TNS.2006.889164 (cit. on p. 71).

[229] M. Willemink and P. Noël. “The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence”. In: *Eur. Radiol.* 29.5 (2019). DOI: doi:10.1007/s00330-018-5810-7 (cit. on p. 6).

[230] Z. Wu, Y. Sun, J. Liu, and U. Kamilov. “Online Regularization by Denoising with Applications to Phase Retrieval”. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. 2019, pp. 3887–3895. DOI: 10.1109/ICCVW.2019.00482 (cit. on p. 38).

[231] H. Xiang, H. Lim, J. A. Fessler, and Y. Dewaraja. “A deep neural network for fast and accurate scatter estimation in quantitative SPECT/CT under challenging scatter conditions”. In: *Eur. J. Nuc. Med. Mol. Im.* 47 (May 2020), 2956–67. DOI: 10.1007/s00259-020-04840-9 (cit. on pp. 15, 58).

[232] L. Xiao and T. Zhang. “A Proximal Stochastic Gradient Method with Progressive Variance Reduction”. In: *SIAM Journal on Optimization* 24.4 (2014), pp. 2057–2075. DOI: 10.1137/140961791 (cit. on p. 110).

[233] H. Xie, Z. Liu, L. Shi, *et al.*. “Segmentation-free Partial Volume Correction for Cardiac SPECT using Deep Learning”. In: *Journal of Nuclear Medicine (supplement 2)* 63 (2022) (cit. on p. 67).

[234] Q. Xie, Y. Li, N. He, M. Ning, K. Ma, G. Wang, Y. Lian, and Y. Zheng. “Unsupervised Domain Adaptation for Medical Image Segmentation by Disentanglement Learning and Self-Training”. In: *IEEE Transactions on Medical Imaging* (2022), pp. 1–1. DOI: 10.1109/TMI.2022.3192303 (cit. on p. 106).

[235] J. Xu, E. Gong, J. Pauly, and G. Zaharchuk. *200x Low-dose PET Reconstruction using Deep Learning*. 2017. DOI: 10.48550/ARXIV.1712.04119 (cit. on p. 69).

[236] R. Xu, M. Soltanolkotabi, J. P. Haldar, W. Unglaub, J. Zusman, A. F. J. Levi, and R. M. Leahy. *Accelerated Wirtinger flow: A fast algorithm for ptychography*. 2018 (cit. on pp. 9, 12, 38).

[237] X. Xu, M. Klasky, M. McCann, and J. A. Fessler. “Swap-Net: A Memory-Efficient 2.5D Cascading Network for 3D Image Reconstruction”. In: (2023) (cit. on p. 108).

[238] X. Xu, J. Liu, Y. Sun, B. Wohlberg, and U. S. Kamilov. “Boosting the Performance of Plug-and-Play Priors via Denoiser Scaling”. In: *54th Asilomar Conf. on Signals, Systems, and Computers*. 2020, pp. 1305–1312. DOI: 10.1109/IEEECONF51394.2020.9443410 (cit. on p. 46).

[239] S. Xue, A. Gafita, A. Afshar-Oromieh, M. Eiber, A. Rominger, and K. Shi. “Voxel-wise Prediction of Post-therapy Dosimetry for 177Lu-PSMA I&T Therapy using Deep Learning”. In: *Journal of Nuclear Medicine* 61.supplement 1 (2020), pp. 1424–1424 (cit. on p. 79).

[240] B. Yaman, S. A. H. Hosseini, and M. Akcakaya. “Zero-Shot Self-Supervised Learning for MRI Reconstruction”. In: *International Conference on Learning Representations*. 2022 (cit. on p. 105).

[241] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin. “DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction”. In: *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1310–1321. DOI: 10.1109/TMI.2017.2785879 (cit. on p. 53).

[242] Y. Yang, M. Pesavento, Y. C. Eldar, and B. Ottersten. “Parallel Coordinate Descent Algorithms for Sparse Phase Retrieval”. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019, pp. 7670–7674. DOI: 10.1109/ICASSP.2019.8683363 (cit. on pp. 12, 38).

[243] R. Yasarla and V. M. Patel. “Uncertainty Guided Multi-Scale Residual Learning-Using a Cycle Spinning CNN for Single Image De-Raining”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 8397–8406. DOI: 10.1109/CVPR.2019.00860 (cit. on p. 78).

[244] Z. Yuan and H. Wang. “Phase retrieval via reweighted Wirtinger flow”. In: *Appl. Opt.* 56.9 (Mar. 2017), pp. 2418–2427. DOI: 10.1364/AO.56.002418 (cit. on p. 50).

[245] J. Yue, T. Mauxion, D. Reyes, *et al.*. “Comparison of quantitative Y-90 SPECT and non-time-of-flight PET imaging in post-therapy radioembolization of liver cancer”. In: *Med. Phys.* 43.10 (2016). DOI: 10.1118/1.4962472 (cit. on p. 13).

[246] G. Zeng, Y. Guo, J. Zhan, *et al.*. “A review on deep learning MRI reconstruction without fully sampled k-space”. In: *BMC Med. Imaging* 21.1 (2021). DOI: 10.1186/s12880-021-00727-9 (cit. on p. 53).

[247] G. Zeng and G. Gullberg. “Frequency domain implementation of the three-dimensional geometric point response correction in SPECT imaging”. In: *Conference Record of the 1991 IEEE Nuclear Science Symposium and Medical Imaging Conference*. 1991, 1943–1947 vol.3. DOI: 10.1109/NSSMIC.1991.259256 (cit. on p. 55).

[248] Z. Zha, B. Wen, X. Yuan, J. Zhou, and C. Zhu. “Simultaneous Nonlocal Low-Rank And Deep Priors For Poisson Denoising”. In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022, pp. 2320–2324. DOI: 10.1109/ICASSP43922.2022.9746870 (cit. on pp. 37, 38).

[249] H. Zhang, Y. Liang, and Y. Chi. “A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms”. In: *J. Mach. Learning Res.* 18.141 (2017), 1–35 (cit. on pp. 21, 22, 38).

[250] H. Zhang and D. P. Mandic. “Is a complex-valued stepsize advantageous in complex-valued gradient learning algorithms?” In: *IEEE Trans. Neural Net. Learn. Sys.* 27.12 (Dec. 2016), 2730–5. DOI: 10.1109/TNNLS.2015.2494361 (cit. on p. 22).

[251] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte. “Plug-and-Play Image Restoration With Deep Denoiser Prior”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2022), pp. 6360–6376. DOI: 10.1109/TPAMI.2021.3088914 (cit. on p. 39).

[252] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising”. In: *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3142–3155. DOI: 10.1109/TIP.2017.2662206 (cit. on pp. 45, 69).

[253] Y. Zhang, P. Song, and Q. Dai. “Fourier ptychographic microscopy using a generalized Anscombe transform approximation of the mixed Poisson-Gaussian likelihood”. In: *Optics Express* 25.1 (Jan. 2017), 168–79. DOI: 10.1364/OE.25.000168 (cit. on pp. 9, 12).

[254] A. Ziabari, D. H. Ye, S. Srivastava, K. D. Sauer, J.-B. Thibault, and C. A. Bouman. “2.5D Deep Learning For CT Image Reconstruction Using A Multi-GPU Implementation”. In: *2018 52nd Asilomar Conference on Signals, Systems, and Computers* (2018), pp. 2044–2049. DOI: 10.1109/ACSSC.2018.8645364 (cit. on p. 80).

[255] K. Zukotynski, V. Gaudet, C. Uribe, *et al.*. “Machine Learning in Nuclear Medicine: Part 2-Neural Networks and Clinical Aspects”. In: *J. Nucl. Med.* 62.1 (2021). DOI: 10.2967/jnumed.119.231837 (cit. on p. 69).