Title

Bayesian Deep Learning Methods in Magnetic Resonance Imaging (MRI) and Simultaneous Positron Emission Tomography and Magnetic Resonance Imaging (PET/MRI)

Author

Leynes, Andrew Palmera

Publication Date

2022

Peer reviewed|Thesis/dissertation

Bayesian Deep Learning Methods in Magnetic Resonance Imaging (MRI) and Simultaneous Positron Emission Tomography and Magnetic Resonance Imaging (PET/MRI)

by

Andrew Palmera Leynes

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:
1644A2CD853841E...
Peder Larson
Chair

DocuSigned by:
Srikantan Nagarajan
DocuSigned by:
121B032C6073427...
Srikantan Nagarajan
Thomas Hope

Committee Members

Acknowledgements

No man is an Island, entire of it self;
every man is a piece of the Continent,
a part of the main.

John Donne
Devotions (1624)

This dissertation and my contributions to the research field would not be possible just by myself. At each step of the way, there were many people who provided their support and guidance. There are too many to mention so I apologize if I was not able to include your name here; you know that I am thankful that you were a part of my journey.

First, this dissertation would not be possible without the mentoring and support provided by Dr. Peder Larson. He introduced me to the world of MRI in my first months at UCSF and I ended up doing my Master's thesis with him. Now I'm completing my PhD dissertation with him as well. Peder always had his door open if you had any questions or wanted to talk through something, no matter how outlandish it was. He let me pursue these ridiculous ideas (at the time) that eventually formed my first papers and this dissertation. He gave me the support and independence to grow into my own as a scientist. For all these things, I am extremely thankful. It's difficult to find a better PI and advisor.

To Dr. Thomas Hope, whom I considered my mentor on the clinical aspects of my research, thank you for embedding in my mind: "don't let perfect get in the way of good". Without that single line, I might still be stuck on the first project by trying to make it

perfect.

To Dr. Sri Nagarajan, thank you for introducing me to the world of Bayesian inference and helping me navigate its twists and turns. It is a challenging subject to study and I thank you for being my resource whenever I get stuck on a concept. Even with your full schedule, you still managed to find the time to give us lectures.

To my qualifying exam committee: Drs. Miki Lustig, Michael Ohliger, Sharmila Majumdar, and (again) Sri Nagarajan. Thank you for challenging me and giving me guidance during my qualifying exam period. It has significantly shaped how I think about research projects.

To everyone in the Larson research group, past and present. Wenwen, Peng, Shuyu, Jeremy, thank you for welcoming me in the group as I was doing my Master's thesis and as I joined as a research specialist. Xucheng, Kirti, Abhe, Fei, Sule, Nikhil, thank you for all the scientific discussions and times outside of research. It's really unfortunate that Covid-19 put up a larger barrier between us all.

To the Vigneron group, who served as a second home in my time at UCSF: Brian (especially our beer-driven scientific and philosophical conversations on breaks from research), Hsin-Yu, Lucas, Dan, and those who have moved on from UCSF, Cornelius, Irene, Daniele, and Eugene M. Thank you for giving me an additional perspective on research and providing a warm and welcoming environment.

To my other scientific colleagues at UCSF: Janine, Melanie Morrison, Eugene O., Valentina, Claudia, Francesco, thank you for all the stimulating scientific discussions. I've learned a lot from all of you.

To Jed Chan, Peter Storey, and the rest of the UCSF Radiology Scientific Computing Services team, without whom all the research activities I've done would not be possible. Thank you for providing all the technical support to making sure our research goes as smoothly as possible.

To my MSBI cohort that were still around for the initial years: Liz P., Jay, and Jeff,

thanks for welcoming me to the US and keeping me company.

To my former housemate, Alejandro. It really was a great time living under the same roof. We never did end up doing that follow-up party. Thank you for all the rounds of food and beer and concerts.

To my parents, Argelie and Atoy Leynes, and my brothers, Ephraim and Jaime Leynes, thank you for all the love and support you've given me through the years. This is the highest educational degree possible in my career path, and I couldn't have reached this point without you.

Finally, to my wife, Ursu, and my daughter, Nathalie. This dissertation is dedicated to the two of you. Thank you, Ursu, for being so understanding and supportive of all the hours that I've put into this dissertation. Thank you for being my reviewer and test audience for my writing and presentations. The quality of the work in this dissertation would be difficult to achieve if I didn't have your feedback. Thank you for all the love and support all these years. And thank you, Nathalie, for constantly reminding me the purpose of all the work that I do. Thank you for reminding me to not to be too immersed in my work and remember that there is more to life than research, algorithms, and programming.

Andrew P. Leynes

March 2022

Abstract

Bayesian Deep Learning Methods in Magnetic Resonance Imaging (MRI) and Simultaneous Positron Emission Tomography and Magnetic Resonance Imaging (PET/MRI)

by

Andrew Palmera Leynes

Medical image reconstruction is the process of reproducing an image of an object from the measurements produced from a scanner through some physical process. Deep image reconstruction, which is utilizing deep learning in image reconstruction, has surpassed what is possible compared to classical approaches. However, one requirement of deep learning is a large dataset that covers as many cases as possible and covers different demographics, disease conditions, hardware configurations, and acquisition schemes. Thus, a major limiting factor in deep learning in medical image reconstruction is the training data available. In this dissertation, I address the challenge of *imperfect* and *small data* in deep learning. This is addressed through the use of Bayesian deep learning. Bayesian deep learning allows a practitioner to estimate uncertainty to quantify possible sources of error and reduce the tendency of models to overfit. This is used in two application domains: simultaneous positron emission tomography and magnetic resonance imaging (PET/MRI) attenuation correction and in sub-Nyquist accelerated MRI. The use of uncertainty in PET/MRI attenuation correction has allowed the correction of intrinsic dataset errors due to bowel air movement, lack of bone signal, misregistration, and metal implant artifacts. The use of uncertainty in sub-Nyquist accelerated MRI has reduced the overfitting behavior of reconstruction models and allowed for higher possible acceleration factors.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and outline

Medical image reconstruction is the process of reproducing an image of an object from the measurements produced from a scanner through some physical process [1]. Classical approaches to image reconstruction are based on the mathematical structure of the image acquisition process and mathematically-definable properties of natural images [2], [3]. In contrast, deep learning is a data-driven modeling technique driven by deep neural networks, large datasets of examples, and modern optimization techniques [4].

Deep image reconstruction, which is utilizing deep learning in image reconstruction, has surpassed what is possible compared to classical approaches [5].

For attenuation correction in simultaneous positron emission tomography and magnetic resonance imaging (PET/MRI), MRI cannot be directly used for attenuation correction since MRI measures nuclear spin density and it does not directly correspond to the PET photon attenuation coefficients [6]. Whereas this is possible with x-ray computed tomography (CT) as done in PET/CT since CT measures photon transmission [7]. Attenuation correction is important since it compensates for photon attenuation. The lack of attenuation correction leads to underestimation of radiotracer uptake concentration and image artifacts [7]. Thus, deep learning has been used to produce pseudo-CT images directly from MRI by training a deep learning model on a dataset of MRI and CT pairs [6], [8]. This has overcome all the

limitations of previous approaches and now allows for PET/MRI to approach the quantitative accuracy of PET/CT.

In accelerated magnetic resonance imaging (MRI), subsampling below the Nyquist limit is a common approach to reduce scan times since this typically does not require any hardware changes [9]–[11]. Utilizing deep learning has increased the possible acceleration (or subsampling) factors by $\sim$2 to 3$\times$ more than previous approaches [12]. This is done by training a deep learning model (or a deep learning augmented reconstruction algorithm) to produce the high-quality fully-sampled reference images from the subsampled MRI over a database of examples.

However, one requirement of deep learning is a large dataset that covers as many cases as possible and covers different demographics, disease conditions, hardware configurations, and acquisition schemes. Unlike computer vision and optical photography, data is much harder to come by in medical imaging. The current largest dataset, fastMRI [13], for medical image reconstruction consists of $\sim$170,000 raw 2D images compared to the $\sim$1.5 Million images in the ImageNet dataset [14].

Thus, a major limiting factor in deep learning in medical image reconstruction is the training data available.

In this dissertation, I address the challenge of *imperfect* and *small data* in deep learning. This is addressed through the use of Bayesian deep learning [15]–[18], a version of deep learning that incorporates Bayesian statistical inference methods into the deep learning models.

In chapter 2, I first discuss the preliminaries and background information needed for this dissertation. This chapter covers classical image reconstruction methods and deep learning methods. Most importantly, this chapter gives an overview of how Bayesian inference is used in medical imaging.

Subsequent chapters are then applications of Bayesian deep learning in specific problem domains.

In chapter 3 is the application of Bayesian deep learning for attenuation correction in

PET/MRI to address the problem of imperfect datasets. Acquiring a large dataset of perfectly matched MRI and CT images are practically impossible because the images are acquired at different times. The most common mismatches are due to movement of bowel air and registration errors due to different bed positions when imaging the pelvis. Furthermore, the presence of metal implants cause mismatch since the metal artifacts appear different in MRI versus CT. This chapter utilizes the measure of uncertainty in the produced synthetic CT images to assess sources of likely error, and uses the PET photon emission data to correct these regions.

In chapter 4 is the application of Bayesian deep learning in the limit of small data: $N = 1$. Deep image reconstruction methods require large datasets with fixed hardware configurations and acquisition schemes. However, as soon as there are changes in the hardware or the acquisition scheme, a new dataset is required and the models need to be trained to account for these changes. Furthermore, there are instances where it is impractical or impossible for have a large high-quality database of training examples. This chapter utilizes the statistical regularization capabilities of Bayesian inference to allow for deep image reconstruction at high acceleration factors without the need for a training dataset and in a scan-specific manner (much like classical reconstruction methods).

Finally, in chapter 5, I discuss the summary of the contributions of this dissertation and possible future lines of inquiry.

Chapter 2

Preliminaries

2.1 The medical image reconstruction problem

The first medical imaging modality was x-ray imaging or radiography [19]. Compared to magnetic resonance imaging (MRI) or positron emission tomography (PET), radiography was relatively simple. There are two components to an radiography system: an x-ray source and a photographic plate. The imaging process is then done as follows: x-rays would be emitted from a source and captured by a photographic plate, with an object or patient in between. The photographic plate then represents how many x-rays passed through a material. Image contrast and structure is then produced by the differing amounts of x-rays that pass through depending on what material was in the path of the x-rays. Bone would block more x-rays than soft tissues like muscle. Thus, bone appears dark on the photographic plate and muscle appears light. On display, this contrast is reversed by placing the exposed film on to a plate with a white backlight where bone when appears light and muscle appears dark. This is a straightforward imaging process that does not require mathematical image reconstruction because of the nature of the imaging process.

Modern medical imaging modalites like MRI, and PET are fundamentally different from x-ray due to the need for image reconstruction. Image reconstruction is the process of tak-

ing measured data and producing imaging from the acquired data. This process is required because, unlike radiography, the measurements you get from MRI and PET are not images themselves. In MRI, what's measured is the behavior of the nuclei under a certain pertubations of the electromagnetic field. In PET, what's measured is the number of photons that are captured by the detectors. Figure 2.1 shows a diagram of the imaging process for MRI and PET and visualizations of the data acquired.

It's clear from fig. 2.1 that the measured data are not visually-interpretable images, but there is some kind of structure in the data. Medical image reconstruction is the process of taking this data and what's known about the physics and image acquisition process to figuring out what the image is that produced these measurements. A detailed description of the physics and imaging of MRI can be found at [20], [21] and of the physics and imaging of PET can be found at [22], [23]

The medical image data measurement process can be formalized mathematically in the following way:

$$y = \mathcal{A}(x) \tag{2.1}$$

where y represents the measured data, $\mathcal{A}$ represents the physical process of obtaining measurements from the underlying image, x.

In MRI, the acquisition process consists of an object (the image with a certain contrast, x) being imaged with multiple receiver coils. Each receiver coil would see an image with a spatially-weighted signal intensity profile. Each image is then linearly encoded into waves of varying frequency (that depend on the underlying object) called k-space and the different wave frequencies or k-space locations are sampled. The final measured data are these wave responses for each receiver coil. The reconstruction problem is then to calculate the underlying image, x, from the wave responses.

In PET, the acquisition process consists of a radiotracer being injected into the body. This radiotracer gets localized in different regions of the body depending on the chemistry of the radiotracer. The acquisition process consists of the localized radiotracer emitting

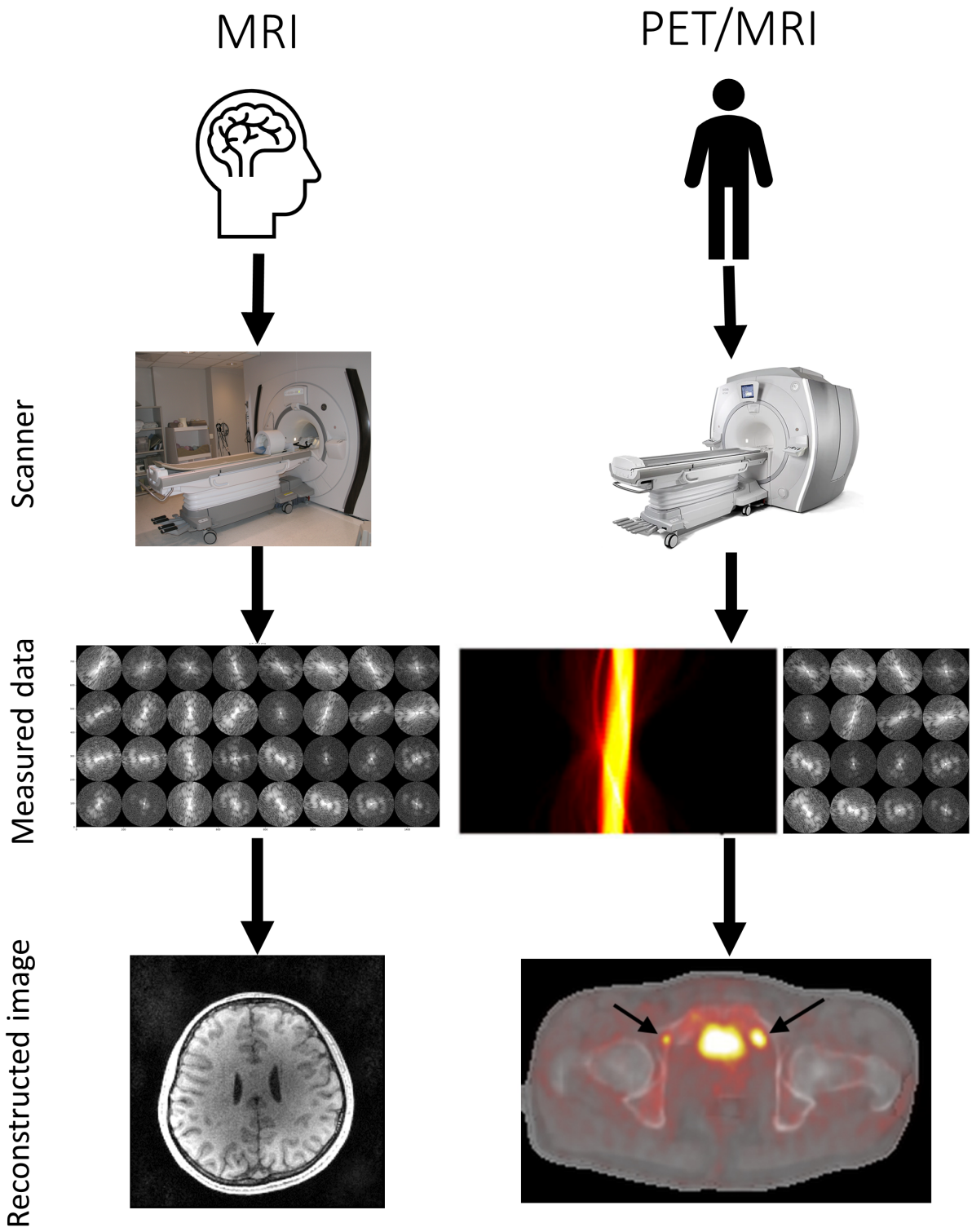

Figure 2.1: Visualizations of the measured data and reconstructed images from MRI and PET/MRI scanners.

photons that are then captured by a ring of detectors. The final measured data are the photon counts that occur at the detector ring along a line in the imaging field-of-view.

The image reconstruction problem would then be:

$$x = \mathcal{A}^{-1}(y) \tag{2.2}$$

where $\mathcal{A}^{-1}$ is the inverse process of going from measurements to the actual underlying image. Under certain cases and conditions such as non-accelerated MRI and high dose PET, $\mathcal{A}^{-1}$ is well-defined and can be directly used to produce the image from the measurements. But, using $\mathcal{A}^{-1}$ can be problematic if the problem is *underdetermined* or *ill-conditioned* even if it is possible to derive and calculate (such as in parallel imaging MRI [24], [25] and reduced dose filtered back-projection PET [26]) and there are cases where the inverse cannot be calculated at all (such as in compressed sensing MRI [27]). In the next section, classical methods to solve for x are discussed for PET and MRI.

2.2 Classical image reconstruction in MRI and PET

2.2.1 MR image reconstruction

MR image reconstruction is a relatively straightforward process when linear encoding gradient fields are used and the k-space sampling follows the Nyquist sampling rate [21], [28]. Because the linear encoding process corresponds to a multi-dimensional Fourier transform, an image from each coil may be produced using the inverse Discrete Fourier Transform (DFT):

$$x_i = \mathrm{DFT}^{-1}(y_i) \tag{2.3}$$

where i corresponds to each coil. The final image may then be produced using sum-of-squares coil combination [29] if there is more than 1 receiver coil:

$$x = \sqrt{\sum_{i=1}^{N} x_i^2} \tag{2.4}$$

An alternative approach to image reconstruction is by formulating the problem using linear algebra [24]:

$$y_i = \mathrm{DFT}(C_i x_i) \tag{2.5}$$

$$y = \begin{bmatrix} y_1 & y_2 & \dots & y_C \end{bmatrix}^T \tag{2.6}$$

where $y_i \in \mathbb{C}^M$ is a row vector of the measurements for each coil channel and y is a concatenation of y_i along the columns to form a vector with size $MC \times 1$, where $x \in \mathbb{C}^N$ is a column vector consisting of N voxels of the image, M is the number of k-space measurement locations and C is the number of coil channels. Equations (2.5) and (2.6) can be reformulated as a single linear equation:

$$y = Ax \tag{2.7}$$

where A is a matrix that represents the discretized forward model that results from eqs. (2.5) and (2.6). In this case, $A \in \mathbb{C}^{MC \times N}$ is the forward encoding operator with $MC \geq N$. Using linear least squares [30], the solution to the over-determined equation is:

$$x = (A^T A)^{-1} A^T y \tag{2.8}$$

However, note that A is typically not expressed explicitly for the whole image due its large size and memory limitations. For example, a typical modern 2D MRI acquisition may involve $M = 256 \times 256 = 65536$, $C = 8$, and $N = 256 \times 256 = 65536$, which gives A a size of $(65536)(8) \times 65536 = 524288 \times 65536$. If the matrix A is represented with 32-bit floating point numbers (4 bytes each number), then a machine would need 128 GigaBytes of

memory. Thus, practical implementations rely on splitting the problem into smaller pieces when possible (e.g., solving for a few pixels at a time in the original SENSE implementation [24]) or using iterative algorithms to reduce the computational complexity and memory requirements [31].

The solution approaches given above are for cases when the MRI is fully-sampled (i.e., sampling follows Nyquist sampling rate). Modern approaches to accelerated MRI have relied on sampling below the Nyquist sampling rate and require more sophisticated reconstruction approaches.

Accelerated MRI: sampling below the Nyquist limit

Parallel imaging and compressed sensing [25], [27] are approaches that allow for MRI data sampling that violate the Nyquist sampling limit and are called undersampled MRI. In this regime, the linear system of equations described in eq. (2.7) becomes *ill-conditioned*. A is *ill-conditioned* when the *condition number*, or the ratio of the first singular value and the last singular value, is large ($>> 1$). This is due to the k-space sampling pattern that introduces ambiguities to the solution space due to sampling below the Nyquist sampling limit. If A is ill-conditioned, then the reconstruction of x from the measurements y is sensitive to noise and model errors. Under this regime, more sophisticated reconstruction algorithms become necessary.

Modern image reconstruction in MRI is centered around the concept that the image reconstruction problem can be posed as an optimization problem. The beauty of posing the reconstruction as an optimization problem is that the reconstruction is no longer limited to analytical solutions that attempt to invert each stage of the image formation process. Rather, the optimization formulate is more flexible since any desired constraints or image properties can be imposed on the reconstructed image. For example, constraints on image smoothness or structure can be easily incorporated into an optimization formulation compared to analytic inversions.

Very generally, an optimization problem is one that attempts to find the solution to the following statement:

$$\hat{x} = \arg\min_{x} f(x) \tag{2.9}$$

or:

$$\hat{x} = \arg\max_{x} f(x) \tag{2.10}$$

where eq. (2.9) describes a minimization problem and eq. (2.10) describes a maximization problem. Equations (2.9) and (2.10) can be made equal as follows:

$$\hat{x} = \arg\min_{x} f(x) = \arg\max_{x} -f(x) \tag{2.11}$$

$$\hat{x} = \arg\max_{x} f(x) = \arg\min_{x} -f(x) \tag{2.12}$$

One of the most common approaches to solving $y = Ax$ is through *least-squares* optimization. We can re-arrange $y = Ax$ as follows:

$$y = Ax \tag{2.13}$$

$$y - Ax = 0 \tag{2.14}$$

We can see from eq. (2.14) that the solution to eq. (2.13) is $\hat{x}$ that gives Ax as close as possible to y. Then, we can define the following optimization function:

$$f(x) = \|y - Ax\|_2^2 \tag{2.15}$$

where:

$$\|x\|_2^2 = \sum_{n=1}^{N} |x_i|^2 \tag{2.16}$$

Thus, we have a minimization problem as follows:

$$\hat{x} = \arg\min_x f(x) = \arg\min_x \|y - Ax\|_2^2 \tag{2.17}$$

When this is solved analytically, the solution is eq. (2.8). However, the solution is very sensitive to noise. Thus, *regularization* is employed to stabilize the solution (i.e., reduce the sensitivity to noise and model errors) and to give a solution that has particular properties. For example, noise signal amplification in parallel imaging MRI causes the magnitude of signal intensities to be very large. In this case, *L_2-regularization*, also called *Tikhonov regularization*, is employed:

$$\hat{x} = \arg\min_x f(x) = \arg\min_x \|y - Ax\|_2^2 + \lambda \|x\|_2^2 \tag{2.18}$$

Adding the term $\|x\|_2^2$ suppresses the large signal intensities that may occur due to noise asignal amplification and λ controls the strength of signal suppression.

As mentioned above, desired structure like image smoothness may be incorporated. One such smoothness-enforcing expression is as follows:

$$\hat{x} = \arg\min_x f(x) = \arg\min_x \|y - Ax\|_2^2 + \lambda((\sum_{i=2}^{N} |x_i - x_{i-1}|^2) + (\sum_{j=2}^{N} |x_j - x_{j-1}|^2)) \tag{2.19}$$

where i indexes over horizontal pixel positions and j indexes over vertical pixel positions in 2D.

In compressed sensing MRI [27], the desired image property is one of *sparsity* where an image has the property that its transform produces only a few coefficients in some transform domain. A compressed sensing reconstruction is formulated as follows:

$$\hat{x} = \arg\min_x f(x) = \arg\min_x \|y - Ax\|_2^2 + \lambda \|Tx\|_1 \tag{2.20}$$

where T is a transform operator. Common choices for T are the discrete wavelet transform (DWT) [32] or total variation (TV) operator [27]. When the DWT is employed, the desired image structure is that the image can be described by a small subset of wavelets. When the TV operator is employed, the desired image structure is that the image gradients are sparse and causes the image to be piecewise-smooth.

The use of parallel imaging and compressed sensing to MRI has significantly decreased scan times without any hardware modifications.

2.2.2 PET image reconstruction

PET image reconstruction can be performed using filtered backprojection (FBP) or through iterative methods that are primarily based on Expectation Maximization. A detailed summary is provided in [2].

Filtered backprojection

Filtered backprojection is an analytical approach that obtains the images from the acquired data:

$$f(x,y) = \int_0^\pi d\phi \; p_{filtered}(x\cos(\phi) + y\sin(\phi), \phi) \tag{2.21}$$

$$p_{filtered}(s,\phi) = \int_{-FOV_r}^{FOV_r} ds' \; p(s',\phi) h(s-s') \tag{2.22}$$

where x and y are the spatial locations (in 2D), f is the signal intensity at the spatial location (x, y), $p_{filtered}(s, \phi)$ is the filtered projection data along the line with distance s from the center of rotation at an angle ϕ, $p(s, \phi)$ is the raw projection data, FOV_r is the radius of

the circular field-of-view, $h(s)$ is a filter kernel that filters the projection data to normalize the signal intensities and may additionally reduce noise and stabilize the reconstruction. However, FBP only captures that data are acquired in projections and does not capture the inherent statistical processes that occur in a PET imaging session. Incorporating the statistics in the reconstruction can reduce image artifacts and improve signal-to-noise ratio.

Iterative reconstruction with Expectation Maximization

Expectation maximization is an iterative method that allows for reconstruction of an image that considers the statistical processes that are inherent in the system. In the case of PET, the statistical process comes from the random radioactive decay of the radiotracer that induces photon emission.

Like in MRI, the PET acquisition process can be formulated as a discretized linear forward model as follows [2]:

$$y = P\lambda \tag{2.23}$$

where y is the mean number of events along a line-of-response (LOR), P is the detection probability matrix that encodes the probability of detecting a photon emission event along the LOR, and λ is the radiotracer concentration at each pixel. Equation (2.23) can be solved using Expectation Maximization (EM) and its variations [2]. Like undersampled accelerated MRI, P is typically *ill-conditioned* and requires *regularization* to stabilize the solution and is included in the EM algorithms through early stopping.

One common EM algorithm for PET reconstruction used in a clinical environment is the ordered subsets expectation maximization algorithm (OSEM) [2]. The cost function is:

$$Q(x,p) = \sum_{j=1}^{J} \left(-\sum_{i=1}^{N} a_{j,i}x_i + p_j \log(\sum_{i=1}^{N} a_{j,i}x_i) \right) \tag{2.24}$$

where J is the number of lines of response, N is the number of voxels in the reconstructed

image, x_i is the i-th voxel in the reconstructed image x, p_j are the photon counts along the j-th line of response, and $a_{j,i}$ is the coefficient of the system matrix that maps the i-th voxel to the j-th line of response. With the cost function defined, the EM updates are then:

$$x_i^{k+1} = x_i^k \frac{1}{\sum_{j' \in J_{k \bmod S}} a_{j',i}} \sum_{j' \in J_{k \bmod S}} a_{j,i} \frac{p_j}{\sum_{i'=1}^{N} a_{j,i'} f_{i'}^n}, \quad i = 1, ..., N \tag{2.25}$$

where n is the EM step number with a subset of the data, S is the number of disjoint subsets of the lines of response projection data, $J_{n \bmod S}$ is the k mod S subset of the projection data. Only a subset of the projection data is used in each iteration to accelerate reconstruction.

2.3 System calibration: when parts of the system are unknown

Section 2.2 discussed the reconstruction problem and solution methods when the imaging system under consideration is fully known (i.e., A in MRI and P in PET are fully determined). However, in most imaging scenarios, these missing portions of the system must be measured in a separate step, or estimated alongside image reconstruction.

2.3.1 MRI system calibration

In parallel imaging and compressed sensing MRI, the forward operator A can be decomposed as follows:

$$A = SFC \tag{2.26}$$

where S is the k-space sampling operator, F is the Fourier transform operator, and C is the coil sensitivity operator. S and F are known by design of the acquisition. This leaves C to be determined and is called the coil calibration problem. Knowledge of C is necessary in

order to solve for the unknown image x as described in section 2.2.1.

In SENSE-type formulations [24], C can be determined by an external coil sensitivity profile scan using a phantom. However, there are certain cases where imaging a phantom is not suitable for coil calibration such as for body imaging where the coil position may not be the same between the coil calibration scan and the actualy imaging scan. In these cases, a GRAPPA-type formulation [25], [33], [34] is utilized:

$$A = SGF \tag{2.27}$$

where G is the coil sensitivity kernel in k-space. This arises since a multiplication of the coil sensitivity profile and the image in image-space is equivalent to a convolution of the coil sensitivity kernel (the Fourier transform of the coil sensitivity profile) and the Fourier transform of the image in k-space:

$$G(\vec{k}) * X(\vec{k}) = G(\vec{k}) * \mathcal{F}\{x(\vec{r})\} = \mathcal{F}\{c(\vec{r})\} * \mathcal{F}\{x(\vec{r})\} = \mathcal{F}\{c(\vec{r})x(\vec{r})\} \tag{2.28}$$

Thus, rather than acquiring the coil sensitivity profiles, C, in a separate scan, G can be estimated from a fully-sampled low-frequency portion of k-space data in the imaging scan itself.

Several papers [25], [27], [31], [33] provide a more detailed description of the coil calibration procedures.

Joint estimation of image and coil sensitivity profiles

There are situations where it's not desirable or possible to take additional time for the coil calibration procedure. In these situations, *joint estimation* of coil sensitivity profiles and image content is utilized [35]–[37]. The joint estimation problem is formulated as an optimization problem:

$$\hat{x}, \hat{C} = \arg\min_{x,C} \|y - SFCx\| + R(x, C) \tag{2.29}$$

when formulating the problem in image-space [36], or:

$$\hat{k}, \hat{G} = \arg\min_{k,G} \|y - SGk\| + R(k, G) \tag{2.30}$$

when formulating the problem in k-space. R is some regularization operator.

Because x and C or k and G need to be solved jointly, the problem becomes a non-linear optimization problem and require more sophisticated procedures to solve.

2.3.2 PET system calibration

The PET imaging system operator can be decomposed as follows [38]:

$$P = P_{\text{det.sens.}} P_{\text{det.blur}} P_{\text{attn}} P_{\text{geom}} P_{\text{positronrange}} \tag{2.31}$$

where $P_{\text{det.sens.}}$ is the detection probability due to detector sensitivity, $P_{\text{det.blur}}$ is the detection probability due to a sinogram blurring factor, P_{attn} is the probability that a photon reaches the detector due to attenuation effects, P_{geom} is the probability that a photon reaches the detector based on the scanner geometry, and $P_{\text{positronrange}}$ is the probability that a photon is emitted in a certain position due to positron range effects. In this decomposition, $P_{\text{det.sens.}}$, $P_{\text{det.blur}}$, and P_{geom} are determined by the construction of the scanner and are fixed and known at the time of imaging; and, $P_{\text{positronrange}}$ is known based on the specific radiotracer used for imaging (e.g., $P_{\text{positronrange}} \approx I$ for ^{18}F-FDG); however, P_{attn} is dependent on the object being imaged and must be calibrated for at the time of imaging.

Calibration of P_{attn} can be done using positron transmission imaging [23] or using x-ray computed tomography (CT) [7].

Joint estimation of activity and attenuation

In cases where positron transmission imaging or CT is not possible or desirable, joint estimation of the radiotracer activity image (x) and the attenuation image (P_{attn}) can be done through optimization called maximum likelihood estimation of activity and attenuation (MLAA) [39], [40].

In the original MLAA paper [39], the likelihood function is:

$$L(\lambda, \mu) = \sum_i \left(-\exp\left(-\sum_k l_{ik}\mu_k \right) \left(\sum_k c_{ik}\lambda_k \right) + y_i \ln\left(\exp\left(-\sum_k l_{ik}\mu_k \right) \left(\sum_k c_{ik}\lambda_k \right) \right) \right) \tag{2.32}$$

where λ is the radiotracer activity concentration image, μ is the attenuation map, l_{ik} is the length of the intersection of the i-th line of response of the i-th detector with voxel k, μ_k is the attenuation at voxel k, c_{ik} is the sensitivity of the i-th detector to the radiotracer activity from voxel k without the effects of attenuation, and y_i is the measured number of photons from the i-th detector. An alternating optimization scheme is then used where either λ or μ is updated while keeping the other fixed at each iteration:

$$\lambda_k^{n+1} = \frac{\lambda_k^n}{\sum_i a_i c_{ik}} \sum_i \frac{c_{ik} y_i}{b_i^n} \tag{2.33}$$

$$\mu_k^{n+1} = \mu_k^n + \frac{\alpha}{D} \left(1 - \frac{\sum_i c_{ik} y_i}{\sum_i c_{ik} a_i b_i^n} \right) \tag{2.34}$$

with

$$a_i = \exp\left(-\sum_k l_{ik}\mu_k\right) \tag{2.35}$$

$$b_i^n = \sum_k c_{ik}\lambda_k^n \tag{2.36}$$

where n is the iteration number, b_i^n is the total activity along the i-th detector line for the n-th iteration using the current activity image estimate λ_i^n, D is the diameter of the reconstruction field-of-view, and α is a step-size parameter.

2.4 Deep learning in image reconstruction

Deep learning is a relatively new machine learning paradigm that relies on the use of deep neural networks. In image reconstruction, convolutional neural networks are most often used. A complete discussion of deep learning and convolutional neural networks (CNNs) can be found in [4]. However, deep learning is more than just using CNNs. There are many different ways that these CNNs can be utilized. There are three dominant approaches that CNNs are used in a larger architectural scheme:

1. Direct mapping
2. Unrolled networks
3. Generative modeling

Then, these different architectural paradigms are used in different learning paradigms that are based on what data are available:

1. Supervised learning
2. Unsupervised learning

3. Semi-supervised learning

4. Self-supervised learning

5. Deep image prior

2.4.1 Architectural paradigms

Direct mapping

Direct mapping methods are the simplest use of CNNs. These are when the solution to the inverse problem $x = f^{-1}(y)$ is directly learned through deep learning. Mathematically, the learning is stated as follows:

$$\hat{\theta} = \arg\min_{\theta} L(X, g_{CNN}(Y; \theta)) \tag{2.37}$$

where θ are the CNN parameters, L is some loss/objective function, and Y and X are input and output training examples, respectively. What gets learned is then an approximation to the inverse function:

$$f^{-1}(y) \approx g_{CNN}(y; \hat{\theta}) \tag{2.38}$$

For example, X can be the undersampled noisy image and Y would be the high-SNR fully-sampled image for image reconstruction. Another example is in the conversion of MRI to synthetic CT: X would be the MRI and Y would be the co-registered CT image.

On inference, the network is applied directly to any new observations y_{new}:

$$x_{new} = g_{CNN}(y_{new}; \hat{\theta}) \tag{2.39}$$

Unrolled networks

Unrolled networks are a different paradigm that is based on gradient descent optimization. For the minimization problem:

$$\hat{x} = \arg\min_{x} f(x) \tag{2.40}$$

The gradient descent update is as follows:

$$x_{k+1} = x_k - \rho \nabla_x f(x) \tag{2.41}$$

where ρ is a step-size factor. Under deep learning, we have training pairs Y_i, X_i that we can then assign as our start and end points in our gradient descent scheme. Suppose that we fix the number of gradient descent steps to be K, then we have:

$$x_0 = Y_i \tag{2.42}$$

$$x_K = X_i \tag{2.43}$$

Then, rather than explicitly calculating for $\nabla_x f(x)$, we can approximate it instead using a CNN:

$$\nabla_x f(x) \approx g_{CNN}(x; \theta) \tag{2.44}$$

and the updates are now:

$$x_{k+1} = x_k - \rho g_{CNN}(x; \theta) \tag{2.45}$$

Finally, we then have the following learning problem:

$$\hat{\theta} = \arg\min_{\theta} L(X, x_K) \tag{2.46}$$

$$x_{k+1} = x_k - \rho g_{CNN}(x; \theta) \tag{2.47}$$

$$x_0 = Y \tag{2.48}$$

This completes the mathematical description of the most basic unrolled network architecture. Alternative variants based on more sophisticated optimization schemes (e.g., regularized least-squares, alternating direction method of multipliers, etc.) are discussed in [41]. In a nutshell, the function of the CNN in this case is to provide the gradient steps from our starting point x_0 to a desired target point $x_K = Y_i$. The step size parameter ρ may be included in the parameters of the CNN to simplify the expressions:

$$\hat{\theta} = \arg\min_{\theta} L(X, x_K) \tag{2.49}$$

$$x_{k+1} = x_k - g_{CNN}(x; \theta) \tag{2.50}$$

$$x_0 = Y \tag{2.51}$$

and it is also possible to learn different parameters at each step rather than using the same parameters for all steps:

$$\hat{\Theta} = \arg\min_{\Theta} L(X, x_K) \tag{2.52}$$

$$\Theta = \{\theta_0, \theta_1, ..., \theta_k, ..., \theta_K\} \tag{2.53}$$

$$x_{k+1} = x_k - g_{CNN}(x; \theta_k) \tag{2.54}$$

$$x_0 = Y \tag{2.55}$$

On inference, it behaves similarly to the direct mapping method but with several applications of a CNN:

$$x_0 = y_{new} \tag{2.56}$$

$$x_{k+1} = x_k - g_{CNN}(x; \hat{\theta}) \tag{2.57}$$

$$x_{new} = x_K = x_{k-1} - g_{CNN}(y_{new}; \hat{\theta}) \tag{2.58}$$

Generative modeling

Generative modeling for inverse problems is one of the newest approaches to tackling solving inverse problems. Rather than forming a mapping from the observation y to the unknown x, a generative model of possible realizations of x are first learned. The major advantage of this method is that matched training pairs Y_i, X_i are not necessary; only X_i are needed:

$$\hat{\theta} = \arg\min_{\theta} L(X, g_{CNN}(Z; \theta)) \tag{2.59}$$

where Z is some latent variable (usually a 1D noise vector). However, inference is much more computationally intensive since another optimization problem needs to be solved:

$$\hat{z} = \arg\min_{z} f(y, g_{CNN}(z; \hat{\theta})) \tag{2.60}$$

where f is the objective function for fitting to the observations. For example in MRI image reconstruction, the objective function could be:

$$f(Y, g_{CNN}(z; \hat{\theta})) = \|Y - AX\|_2^2 \tag{2.61}$$

$$X = g_{CNN}(z; \hat{\theta}) \tag{2.62}$$

$$f(Y, g_{CNN}(z; \hat{\theta})) = \left\|Y - Ag_{CNN}(z; \hat{\theta})\right\|_2^2 \tag{2.63}$$

2.4.2 Learning paradigms

Supervised learning

Most of the architectural discussions above have been under the context of supervised learning (except for generative modeling). This is the most common learning paradigm in deep learning since typically, the two sides of the mapping are known and are easy to obtain, and only the function that maps between them ($f : X \to Y$ and/or $f : Y \to X$) is unknown.

Unsupervised learning

Unsupervised learning is different from supervised learning since there are no training pairs available. Only data from a single domain X is available. Under this regime, the task is to find a set of patterns that fit the data and satisfy certain mathematically-defined properties (e.g., sparsity). The most common example of unsupervised learning is clustering, where data is attempted to be grouped together. Under the context of image reconstruction, the *classical* image reconstruction tasks can be considered as unsupervised learning. Since the reconstructed image is a pattern that fits the data in (usually) a different domain and may satisfy certain properties (e.g., sparsity of Wavelet coefficients). The generative modeling framework discussed in section 2.4.1 is also an example of unsupervised learning. This domain is still an active research area in deep image reconstruction.

Semi-supervised learning

Semi-supervised learning is a hybrid of supervised and unsupervised learning techniques. It is called *semi-supervised* because although you may have data from two domains X and Y, they may not be exactly matched or paired. An example of semi-supervised learning is domain translation using generative adversarial networks (GANs) [42]. In MRI reconstruction, one example that utilized GANs in a semi-supervised learning scheme with a database of unrelated undersampled k-space data [43]. In their work, the supervision signals come

from comparing the synthetic undersampled data of their reconstructed images with real undersampled k-space data (with a different sampling pattern) from a different scan, and their approach is unsupervised because there are no matching pairs of k-space data with the same undersampling pattern nor are there fully-sampled images to learn from.

Self-supervised learning

Self-supervised learning is a variant of unsupervised and semi-supervised learning where only the network parameters are shared across samples and ground-truth reference data is not available. For example in an unsupervised image reconstruction regime, a CNN may be trained to produce a reconstructed image from a latent code z that fits the measured data $\hat{y}$ through the model $y = Ag_{CNN}(z;\theta)$ in a GAN framework. In this scenario, the generated images $\hat{x} = g_{CNN}(z;\theta)$ would be compared against a dataset $X = \{x_1, x_2, ..., x_M\}$ using a discrimator CNN $D(\hat{x}, X)$. In contrast in a self-supervised scenario, there would be no discriminator to compare against the reconstructed images against the database of images. The generative CNN g_{CNN} would only be trained with measurements $Y = \{y_1, y_2, ..., y_N\}$.

Deep image prior

The deep image prior framework [44] is (simply) self-supervised learning taken to smallest possible dataset size ($N = 1$). It is very similar to classical image reconstruction regimes with the exception of using a CNN as a generative model for producing the reconstructed image from a random noise vector z:

$$\hat{\theta} = \arg\min_{\theta} L(y, f(g_{CNN}(z;\theta))) \tag{2.64}$$

$$y = f(x) \tag{2.65}$$

$$x = g_{CNN}(z;\theta) \tag{2.66}$$

where L is some loss function appropriate to the reconstruction problem. In MRI reconstruction, L would correspond to regularized least squares.

The unique aspect about deep image prior is that it is the first work to demonstrate the effectiveness of convolutional neural networks as implicit priors for image reconstruction.

2.5 Bayesian deep learning in medical image reconstruction

2.5.1 Probabilistic modeling and Bayesian inference

A probabilistic model consists of the following ingredients:

1. Probability distributions for random variables
2. Relationships between random variables

As an example, let us revisit the model in 2.1:

$$y = \mathcal{A}(x) \tag{2.67}$$

Suppose that $\mathcal{A}$ defines a linear projection model (a) that has its measurements corrupted by additive noise:

$$y = ax + n \tag{2.68}$$

where n is a random variable. In many different situations (including in MRI), the additive noise n is a zero-mean Gaussian-distributed random variable with variance σ^2:

$$n \sim Gaussian(0, \sigma^2) \tag{2.69}$$

where $\sim$ means that the random variable is drawn from a certain distribution. More explicitly, we can write the probability density function:

$$p(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{n^2}{2\sigma^2}) \tag{2.70}$$

where $p(n)$ is the continuous probability density function of n. Given eqs. (2.68) and (2.70), we can derive the following conditional distribution or *likelihood function*:

$$L(x|y) = p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y-ax)^2}{2\sigma^2}) \tag{2.71}$$

Note the subtle notational difference between the conditional probability, $p(y|x)$, and the likelihood function, $L(x|y)$. The (supposed) rationale for the L notation is because L is ultimately a function of x because y is the data that is observed and is thus constant. However, it flips around the dependencies and commonly causes confusion (including for me). Because of this, I will avoid using the notation for the likelihood function $L(x|y)$ and will only use the conditional probability form, $p(y|x)$ moving forward.

Now, given the likelihood, we can then apply the principle of *maximum likelihood* to estimate the unknown variable x:

$$\hat{x} = \arg\max_x p(y|x) = \arg\max_x \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y-ax)^2}{2\sigma^2}) \tag{2.72}$$

$$= \arg\max_x -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y-ax)^2 - \frac{1}{2\sigma^2}(y-ax)^2 \tag{2.73}$$

$$= \arg\max_x -\frac{1}{2\sigma^2}(y-ax)^2 \tag{2.74}$$

$$\hat{x} = \arg\min_x (y-ax)^2 \tag{2.75}$$

assuming σ is known, which gives us the least squares optimization problem as before. We may also add regularization terms if so desired.

In most discussions of statistical estimation in signal processing and image reconstruc-

tion (the so-called frequentist perspective), we would stop the modeling and inference around here. However, there is still one more thing that is undefined that separates the Bayesian perspective from the frequentist perspective of modeling: defining $p(x)$. In frequentist terminology, only n is considered as a random variable and x is considered as a deterministic object (with only a single value!) but is currently unknown. However, in the Bayesian perspective, everything that is not directly known or measured in the problem is treated as a random variable. So, x is also a random variable that has a probability distribution.

This is a fundamental distinction in how probabilities and uncertainties are viewed. Because a quantity is fundamentally not known (even though there is only one unique value for it), the quantity is viewed to have a probability distribution. In this case, the probability distribution captures the uncertainty of knowledge about the exact value of the variable. In contrast, the frequentist approach enforces right from the very beginning that there is only one true unique value. A great video lecture around this topic can be found at [45].

Historically, the Bayesian perspective seemed to have been one of the dominant perspectives before the rise of frequentist statistics. In Laplace's *A Philosophical Essay on Probabilities* [46], he states (translated into English):

> The curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance.
>
> Probability is relative, in part to this ignorance, in part to our knowledge. We know that of three or a greater number of events a single one ought to occur; but nothing induces us to believe that one of them will occur rather than the others. In this state of indecision it is impossible for us to announce their occurrence with certainty. It is, however, probable that one of these events, chosen at will, will not occur because we see several cases equally possible which exclude its occurrence, while only a single one favors it.

In other words: all of reality was considered to be deterministic and uncertainty was

only due to the lack of knowledge. Of course, this was before the age of quantum mechanics where (so far) there appears to be fundamentally probabilistic quantities. Furthermore, it's not that there are many solutions to a problem, it is that there is one solution but you do not know which of the many possibilities is the truth due to the lack of knowledge.

So, following this *Bayesian* perspective, let us write a more complete description of the probabilities. First, we start with a description of the joint distribution of all the variables in the problem:

$$p(y, x, a, \sigma) = p(y|x, a, \sigma)p(x, a, \sigma) \tag{2.76}$$

Given that we know that the measurement problem is $y = ax + n$ (and a is known to have the value $\bar{a}$), we can assume in this instance that x, a, and n are independent and thus $p(x, a, \sigma) = p(x)p(a)p(\sigma)$. We can further assume that σ is known and was measured. We will then have:

$$p(y, x, a, \sigma) = p(y|x, a, \sigma)p(x)p(a)p(\sigma) = p(y|x, \bar{a}, \sigma_{\text{measured}})p(x) \tag{2.77}$$

with $p(\sigma) = \delta(\sigma - \sigma_{\text{measured}})$ and $p(a) = \delta(a - \bar{a})$. With the random variable y being observed (having the value $\bar{y}$), a is known, and σ is measured, the only unknown is x. We can then utilize *Bayes theorem of inverse probability* or *Bayes theorem* for short. Generally, Bayes theorem allows to you define p(everything unknown|everything known). To solve for $p(x|\bar{y}, \bar{a}, \sigma)$:

$$p(x|\bar{y}, \bar{a}, \sigma) = \frac{p(\bar{y}, x, \bar{a}, \sigma)}{p(\bar{y})} = \frac{p(\bar{y}, x, \bar{a}, \sigma)}{\iiint p(\bar{y}, x, a, \sigma)dxdad\sigma} = \frac{p(\bar{y}|x, \bar{a}, \sigma_{\text{measured}})p(x)}{\int p(\bar{y}|x, \bar{a}, \sigma_{\text{measured}})p(x)dx} \tag{2.78}$$

$$= \frac{1}{Z}p(\bar{y}|x, \bar{a}, \sigma_{\text{measured}})p(x) \tag{2.79}$$

with $Z = p(\bar{y}) = \int p(\bar{y}|x, \bar{a}, \sigma_{\text{measured}})p(x)dx$. Since we have a probability distribution for

the parameter of interest, we define an estimator for x that extracts a single value from the conditional probability distribution $p(x|\bar{y}, \bar{a}, \sigma)$. A common choice is the conditional mean estimator:

$$\hat{x} = E_{p(x|\bar{y},\bar{a},\sigma)}[x] = \int x p(x|\bar{y}, \bar{a}, \sigma) dx \tag{2.80}$$

where $E_{p(x|\bar{y},\bar{a},\sigma)}$ denotes taking the expectation with the distribution $p(x|\bar{y}, \bar{a}, \sigma)$. Note that I indicated the distribution that the expectation is taking over. This is because you may take the expectation over different distributions, and it helps to be clear under which distribution the expectation is being calculated. This becomes especially important when *variational Bayesian inference* techniques are used.

Alternatively, we may also use the *maximum a posteriori* (MAP) estimator that takes the maximum value of $p(x|\bar{y}, \bar{a}, \sigma)$, which may be solved using optimization:

$$\begin{aligned}
\hat{x} &= \arg\max_x p(x|\bar{y}, \bar{a}, \sigma) && (2.81)\\
&= \arg\max_x \frac{1}{Z} p(\bar{y}|x, \bar{a}, \sigma_{\text{measured}}) p(x) && (2.82)\\
&= \arg\max_x \log(p(\bar{y}|x, \bar{a}, \sigma_{\text{measured}})) + \log p(x) - \log Z && (2.83)\\
&= \arg\max_x \log(p(\bar{y}|x, \bar{a}, \sigma_{\text{measured}})) + \log p(x) && (2.84)\\
&= \arg\max_x -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - ax)^2 + \log p(x) && (2.85)\\
&= \arg\min_x \frac{1}{2\sigma^2}(y - ax)^2 + R(x) && (2.86)
\end{aligned}$$

which is equivalent to penalized least-squares with $R(x) = -\log p(x)$.

All the derivations above are only for a univariate model. When multivariate models are considered (e.g., in imaging), the derivation must consider the multivariate nature of the problem. A common mistake is to consider a single image as a univariate variable. An image

is composed of many different pixels and is thus multivariate. Each pixel must be treated appropriately. As an example, let us extend the model above to a multivariate imaging setting such as multi-coil MRI. Our data generating model is now as follows:

$$Y = AX + U \tag{2.87}$$

where $X \in \mathbb{R}^{N\times M}$, $A \in \mathbb{R}^{K\times N}$, $U \in \mathbb{R}^{K\times M}$, N is the number of voxels, M is the number of coils, and K is the number of k-space sampling locations. U is additive noise that is zero-mean and independent across the k-space sampling locations, but may be correlated across the coils:

$$p(u_{11}, u_{12}, ..., u_{km}, ..., u_{KM}) = \prod_{k=1}^{K} \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}U_k^T \Sigma^{-1} U_k\right) \tag{2.88}$$

$$U_k = \begin{bmatrix} u_{k1} & u_{k2} & ... & u_{kM} \end{bmatrix}^T \tag{2.89}$$

with positive-definite covariance matrix $\Sigma \in \mathbb{R}^{M\times M}$. Suppose then that the coil covariance matrix, Σ, is measured and the forward model operator, A, is fully known. The probability distributions for these can then be defined as:

$$p(\sigma_{11}, \sigma_{12}, ..., \sigma_{ij}, ..., \sigma_{MM}) = \prod_{i=1}^{M}\prod_{j=1}^{M} \delta(\sigma_{ij} - \bar{\sigma}_{ij}) \tag{2.90}$$

$$p(a_{11}, a_{12}, ..., a_{kn}, ..., a_{KN}) = \prod_{k=1}^{K}\prod_{n=1}^{N} \delta(a_{ij} - \bar{a}_{ij}) \tag{2.91}$$

where σ_{ij} and $\bar{a}_{ij}$ are the known values, and δ is the Dirac delta function. With the known noise distribution and data generating model, we can derive the likelihood:

$$p(Y|X,\bar{A},\bar{\Sigma}) = \prod_{k=1}^{K} p(Y_k|X,\bar{A}_k,\bar{\Sigma}) = \prod_{k=1}^{K} \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(Y_k - \bar{A}_k X)^T \Sigma^{-1}(Y_k - \bar{A}_k X)\right) \tag{2.92}$$

$$Y_k = \begin{bmatrix} y_{k1} & y_{k2} & ... & y_{kM} \end{bmatrix}^T \tag{2.93}$$

$$A_k = \begin{bmatrix} a_{k1} & a_{k2} & ... & a_{kN} \end{bmatrix}^T \tag{2.94}$$

and by Bayes theorem:

$$p(X|\bar{Y},\bar{A},\bar{\Sigma}) = \frac{p(\bar{Y},X,\bar{A},\bar{\Sigma})}{p(\bar{Y})} \tag{2.95}$$

$$= \frac{p(\bar{Y},X,\bar{A},\bar{\Sigma})}{\int p(\bar{Y},X,A,\sigma) dx_{11}...dx_{NM} da_{11}...da_{KN} d\sigma_{11}...d\sigma_{MM}} \tag{2.96}$$

$$= \frac{p(\bar{Y}|X,\bar{A},\bar{\Sigma})p(X)}{\int p(\bar{Y}|X,\bar{A},\bar{\Sigma})p(X) dx_{11}...dx_{NM}} \tag{2.97}$$

$$p(X|\bar{Y},\bar{A},\bar{\Sigma}) = \frac{1}{Z} p(\bar{Y}|X,\bar{A},\bar{\Sigma})p(X) \tag{2.98}$$

where $\bar{Y}$ is the measured values of Y, and $Z = \int p(\bar{Y}|X,\bar{A},\bar{\Sigma})p(X)dx_{11}...dx_{NM}$. The conditional mean estimator is then:

$$\hat{X} = E_{p(X|\bar{Y},\bar{A},\bar{\Sigma})}[X] = \int X p(X|\bar{Y},\bar{A},\bar{\Sigma}) dx_{11}...dx_{NM} \tag{2.99}$$

and the MAP estimator is:

$$\hat{X} = \arg\min_X \sum_{k=1}^{K} \frac{1}{2}(Y_k - A_k X)^T \Sigma^{-1}(Y_k - A_k X) - \log p(X) \tag{2.100}$$

The MAP estimator is equivalent to solving a penalized least-squares problem and there are multitudes of optimization techniques that can be leveraged to solve the problem for

specific forms of $p(X)$. For example, when $p(X)$ is based on a Laplace distribution:

$$p(X) = \prod_{n=1}^{N} \prod_{m=1}^{M} p(x_{nm}) = \prod_{n=1}^{M} \prod_{m=1}^{M} \frac{\lambda}{2} \exp\left(-\lambda |x_{nm}|\right) \tag{2.101}$$

this is equivalent to L1-regularized least squares or LASSO [47]:

$$\hat{X} = \arg\min_{X} \sum_{k=1}^{K} \frac{1}{2} (Y_k - A_k X)^T \Sigma^{-1} (Y_k - A_k X) + \lambda \left\| X \right\|_1 \tag{2.102}$$

If we further suppose that $\Sigma = \sigma^2 I$, then we get the much more familiar form of L1-regularized least squares regression:

$$\hat{X} = \arg\min_{X} \sum_{k=1}^{K} \frac{1}{2\sigma^2} (Y_k - A_k X)^T (Y_k - A_k X) + \lambda \left\| X \right\|_1 \tag{2.103}$$

$$= \arg\min_{X} \frac{1}{2\sigma^2} \left\| Y - AX \right\|_2^2 + \lambda \left\| X \right\|_1 \tag{2.104}$$

In contrast, solving the inference problem defined in eq. (2.98) is an active area of research and highly depends on the form of the prior $p(X)$. I refer the reader to the following references for discussions on solving the inference problem using classical techniques: [48]–[51]. To briefly summarize, solving the inverse probability inference problem is analytically tractable in only very specific cases: one of which is that the distributions must belong to the *exponential family* [48]. For all other models, numerical or approximate techniques are necessary. When the dimensionality is low (i.e., not so many variables, 10's-100's of variables, depending on compute capacity), then numerical Monte Carlo approaches can be used to solve the inference problem. However, in imaging problems and especially in deep learning, the models are very high-dimensional (e.g., a single 1024×1024 image has 1 million parameters; each pixel is one unknown parameter) and approximate Bayesian inference techniques become necessary [48], [52], [53].

2.5.2 Approximate Bayesian Inference

In a nutshell, approximate Bayesian inference techniques either modify the model to make it analytically tractable, or modify the model that the dimensionality is greatly reduced. This is especially important in deep learning and medical imaging reconstruction as the number of parameters that we deal with is on the order of millions. Generally, the goal of approximate Bayesian inference is to find an approximate form of the true posterior that is tractable to solve:

$$p(X|\bar{Y},\bar{A},\bar{\Sigma}) \approx q(X|\bar{Y},\bar{A},\bar{\Sigma}) \tag{2.105}$$

where q is some function that approximates p using some metric of similarity (e.g., Kullback-Leibler divergence)

There are several texts that go over the different approximations methods [48], [52]–[56] so I will only discuss the few that are relevant to this thesis: variational Bayesian inference [52]–[54], and Monte-Carlo Dropout [57].

Variational Bayesian Inference

Variational Bayesian inference uses the Kullback-Leibler divergence (KL-divergence) as the backbone for forming the approximation. The KL-divergence is defined as follows:

$$D_{KL}(q(X|\bar{Y},\bar{A},\bar{\Sigma})||p(X|\bar{Y},\bar{A},\bar{\Sigma})) = -\int q(X|\bar{Y},\bar{A},\bar{\Sigma}) \log \frac{p(X|\bar{Y},\bar{A},\bar{\Sigma})}{q(X|\bar{Y},\bar{A},\bar{\Sigma})} dX \tag{2.106}$$

$$= E_{q(X|\bar{Y},\bar{A},\bar{\Sigma})}[\log \frac{p(X|\bar{Y},\bar{A},\bar{\Sigma})}{q(X|\bar{Y},\bar{A},\bar{\Sigma})}] \tag{2.107}$$

under the same context of image reconstruction as above. We can derive what's called the *Evidence Lower Bound* (ELBO) that allows us to solve for q. We start by expanding the marginal likelihood or *evidence*, $p(\bar{Y})$:

$$\log p(\bar{Y}) = \log \int p(X, \bar{Y}, \bar{A}, \bar{\Sigma}) dX \tag{2.108}$$

$$= \log \int \frac{p(X, \bar{Y}, \bar{A}, \bar{\Sigma}) q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})} dX \tag{2.109}$$

$$= \log E_{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}[\frac{p(X, \bar{Y}, \bar{A}, \bar{\Sigma})}{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}] \tag{2.110}$$

Applying Jensen's inequality, we then get:

$$\log E_{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}[\frac{p(X, \bar{Y}, \bar{A}, \bar{\Sigma})}{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}] \geq E_{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}[\log \frac{p(X, \bar{Y}, \bar{A}, \bar{\Sigma})}{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}] = L(q(X|\bar{Y}, \bar{A}, \bar{\Sigma})) \tag{2.111}$$

Thus,

$$\log p(\bar{Y}) \geq E_{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}[\log \frac{p(X, \bar{Y}, \bar{A}, \bar{\Sigma})}{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}] \tag{2.112}$$

The principle of *evidence maximization* [48] is to make the data that we observed ($\bar{Y}$) the most likely across all possibilities of all the other variables (X in this example, but may also include A, Σ, and any other unknown variables) and (if we're using variational inference) all the approximating functions (q). The evidence maximization framework is typically used in model selection when there are no approximations involved. However, it becomes extremely useful for selecting the best possible approximation given the constraints.

A possible approximate Bayesian inference reconstruction can defined as follows:

$$q(x_i) = Normal(\mu_{x_i}, \sigma_{x_i}) \tag{2.113}$$

$$q(X|\bar{Y}, \bar{A}, \bar{\Sigma}) = \prod_{i=1}^{N} q(x_i) \tag{2.114}$$

and the optimization problem is then finding the values of the variational parameters

$\{\mu_{x_i}, \sigma_{x_i}\} \forall i$:

$$\mu_{x_1}, ..., \mu_{x_N}, \sigma_1, ..., \sigma_{x_N} = \underset{\mu_{x_1}, ..., \mu_{x_N}, \sigma_1, ..., \sigma_{x_N}}{\arg\max} E_{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}[\log \frac{p(X, \bar{Y}, \bar{A}, \bar{\Sigma})}{q(X|\bar{Y}, \bar{A}, \bar{\Sigma})}] \tag{2.115}$$

One such approach is through *Variational Expectation Maximization* [53].

Bayesian Deep Learning and Monte-Carlo Dropout

In deep learning, the goal optimization is to learn network parameters that minimize the given loss function. In this case, the objective functions given above for image reconstruction is modified to solve for network parameters. For example, in a generative modeling paradigm we may have:

$$Y = Ag_{CNN}(Z; \theta) + N \tag{2.116}$$

$$L(q(\theta|\bar{Y}, \bar{A}, \bar{\Sigma}, Z)) = E_{q(\theta|\bar{Y}, \bar{A}, \bar{\Sigma}, Z)}[\log \frac{p(\theta, \bar{Y}, \bar{A}, \bar{\Sigma}, Z)}{q(\theta|\bar{Y}, \bar{A}, \bar{\Sigma}, Z)}] \tag{2.117}$$

where θ are the network parameters and Z is a random noise vector. This is more involved to solve due to the requirement of evaluating the expectation E_q. One approach to solve this is to use a Monte Carlo approximation of the expectation function and using stochastic gradient descent [58], [59]. However, this requires implementation changes that are not trivial and highly depend on the particular network architecture under consideration.

Monte-Carlo Dropout (MC-Dropout) MC-Dropout [57] is an alternative method that requires little to no changes to the standard deep learning training algorithm and can be deployed for practically any deep convolutional neural network. The theory is fundamentally based on variational Bayesian inference and the major benefit of MC-Dropout is in the ease how it can be implemented. The use of MC-Dropout is summarized in two steps:

1. Train the network with dropout layers [60] (if not already), while keeping everything else the same (e.g., loss function, regularization, etc.)

2. On inference, continue using dropout, and generate T inference samples using different dropout masks.

Once the inference samples are produced, final results can then be derived such as taking pixel-wise mean and standard deviation for the final reconstructed image and uncertainty estimates, respectively.

$$X = g_{CNN}(Z;\theta) \tag{2.118}$$

$$\bar{X} = \int g_{CNN}(Z;\theta)q(\theta)d\theta \approx \frac{1}{T}\sum_{i=1}^{T} g_{CNN}(Z;\theta_i) \tag{2.119}$$

$$\sigma_X^2 = \int (g_{CNN}(Z;\theta) - \bar{X})^2 q(\theta)d\theta \approx \frac{1}{T}\sum_{i=1}^{T} (g_{CNN}(Z;\theta) - \bar{X})^2 \tag{2.120}$$

where $\theta_i \sim q(\theta)$.

2.5.3 Summary

Taking a Bayesian approach to medical image reconstruction allows you to not only capture a single solution, but to estimate a whole family of possible solutions. This leads to estimates of uncertainty that can capture limitations in the training dataset during inference. This feature is utilized in the work in chapter 3. Furthermore, by having a family of solutions that takes into consideration the statistics of the problem, this can give more accurate reconstructions in the face of noise and undersampling in MRI reconstruction, as shown in the work in chapter 4.

Chapter 3

Attenuation Coefficient Estimation for PET/MRI With Bayesian Deep Learning pseudo-CT and Maximum Likelihood Estimation of Activity and Attenuation (UpCT-MLAA)

A major remaining challenge for magnetic resonance-based attenuation correction methods (MRAC) is their susceptibility to sources of MRI artifacts (e.g. implants, motion) and uncertainties due to the limitations of MRI contrast (e.g. accurate bone delineation and density, and separation of air/bone). I propose using a Bayesian deep convolutional neural network that, in addition to generating an initial pseudo-CT from MR data, also produces uncertainty estimates of the pseudo-CT to quantify the limitations of the MR data. These outputs are combined with MLAA reconstruction that uses the PET emission data to improve the attenuation maps. With the proposed approach (UpCT-MLAA), I demonstrate accurate estimation of PET uptake in pelvic lesions and show recovery of metal implants.

In patients without implants, UpCT-MLAA had acceptable but slightly higher RMSE than Zero-echo-time and Dixon Deep pseudo-CT when compared to CTAC. In patients with metal implants, MLAA recovered the metal implant; however, anatomy outside the implant region was obscured by noise and crosstalk artifacts. Attenuation coefficients from the pseudo-CT from Dixon MRI were accurate in normal anatomy; however, the metal implant region was estimated to have attenuation coefficients of air. UpCT-MLAA estimated attenuation coefficients of metal implants alongside accurate anatomic depiction outside of implant regions.

This chapter was previously published at [61].

3.1 Prior work

My prior work surrounding this topic can be found at [8], [62]–[64]. Initially, a segmentation and model-based approached was used to produce pseudo-CT images from a combination of zero echo time (ZTE), Dixon fractional fat, and Dixon fractional water MRI [62] to address the problem of missing bone in the pseudo-CT. This was then expanded to use deep learning to avoid the manual segmentation and produce a synthetic CT directly from MRI [63]. I later found that ZTE may not be completely necessary since bone can be reliably estimated using only Dixon MRI [8]. Other groups have corroborated this finding [65]. And on the sequence optimization side, I leveraged optimization techniques to identify what kind of features the input MRI needs in order to produce the most realistic CT images possible [64]. I found that the bone delineation needs to be much sharper to obtain the same sharpness of bone definition that CT provides. However on the MRI sequence side, it is still an open question how to achieve that kind of image sharpness.

3.2 Introduction

The quantitative accuracy of simultaneous positron emission tomography and magnetic resonance imaging (PET/MRI) depends on accurate attenuation correction. Simultaneous imag-

ing with positron emission tomography and computed tomography (PET/CT) is the current clinical gold standard for PET attenuation correction since the CT images can be used for attenuation correction of 511keV photons with piecewise-linear models [7]. Magnetic resonance imaging (MRI) measures spin density rather than electron density and thus cannot directly be used for PET attenuation correction.

A comprehensive review of attenuation correction methods for PET/MRI can be found at [66] . Briefly, current methods for attenuation correction in PET/MRI can be grouped into the following categories: atlas-based, segmentation-based, and machine learning-based. Atlas-based methods utilize a CT atlas that is generated and registered to the acquired MRI [67]–[70]. Segmentation-based methods use special sequences such as ultrashort echo-time (UTE) [71]–[75] or zero echo-time (ZTE) [76]–[80] to estimate bone density and Dixon sequences [81]–[83] to estimate soft tissue densities. Machine learning-based methods, including deep learning methods, use sophisticated machine learning models to learn mappings from MRI to pseudo-CT images [63], [65], [84]–[88] or PET transmission images [89]. There have also been methods that estimate attenuation coefficient maps from the PET emission data [90], [91] or directly correct PET emission data [92]–[94] using deep learning.

For PET alone, an alternative method for attenuation correction is "joint estimation", also known as maximum likelihood estimation of activity and attenuation (MLAA) [39], [40]. Rather than relying on an attenuation map that was measured or estimated with another scan or modality, the PET activity image (λ-map) and PET attenuation coefficient map (μ-map) are estimated jointly from the PET emission data only. However, MLAA suffers from numerous artifacts and high noise [95].

In PET/MRI, recent methods developed to overcome the limitations of MLAA include using MR-based priors [96], [97] constraining the region of joint estimation [98], or using deep learning to denoise the resulting λ-map and/or μ-map from MLAA [99]–[102]. Mehranian and Zaidi's [97] approach of using priors improved MLAA results however this was not demonstrated on metal implants. Ahn et al and Fuin et al's methods [96], [98] that also use

priors were able to recover metal implants in the PET image reconstruction, but the μ-maps were missing bones and other anatomical features. Furthermore, their methods require a manual or semi-automated segmentation step to delineate the regions where to apply the correct priors (such as the metal implant region). The approaches by Hwang et al [100]–[102] and Choi et al [99] that utilize supervised deep learning resulted in anatomically correct and accurate μ-maps; however, the method was not demonstrated in the presence of metal implants.

Utilizing supervised deep learning is considered a very promising method for accurate and precise PET/MRI attenuation correction. However, the main limitation of a supervised deep learning method is the finite data set that needs to have a diverse set of well-matched inputs and outputs.

In PET/MRI, the presence of metal implants complicates training because there are resulting metal artifacts in both CT and MRI. Furthermore, the artifacts appears differently: a metal implant produces a star-like streaking pattern with high Hounsfield unit values in the CT image [103] and a signal void in the MRI image [96]. This makes registration between MRI and CT images difficult and the artifacts lead to intrinsic errors in the training dataset.

In addition, there will arguably always be edge cases and rare features that cannot be captured with enough representation in a training data set. Images of humans can have rare features not easily obtained (e.g., missing organs due to surgery, a new or uncommon implant). Under these conditions, a standard supervised deep learning approach may produce incorrect predictions and the user (or any downstream algorithm) will be unaware of the errors.

A recent study by Ladefoged et al [104] demonstrated the importance of a high-quality data set in deep learning-based brain PET/MRI attenuation correction. A large, diverse set of at least 50 training examples were required to achieve robustness and they highlighted that the remaining errors and limitations in deep learning-based MR attenuation correction were due to "abnormal bone structures, surgical deformation, and metal implants."

In this work, I propose the use of supervised Bayesian deep learning to estimate predictive uncertainty to detect rare or previously unseen image structures and estimate intrinsic errors that traditional supervised deep learning approaches cannot.

Bayesian deep learning provides tools to address the limitations of a finite training dataset: the estimation of epistemic and predictive uncertainty [105]. A general introduction to uncertainties in machine learning can be found at [106].

Epistemic uncertainty is the uncertainty on learned model parameters that arises due to incomplete knowledge or, in the case of supervised machine learning, the lack of training data. Epistemic uncertainty is manifested as a diverse set of different model parameters that fit the training data.

The epistemic uncertainty of the model can then be used to produce predictive uncertainty that capture if there are any features or structures that deviate from the training dataset on a test image. This allows for the detection of rare or previously unseen image structures without explicitly training to identify these structures.

Typical supervised deep learning approaches do not capture the epistemic nor predictive uncertainty because only one set of model parameters are learned and only a single prediction is produced (e.g., a single pseudo-CT image).

In this work for PET/MRI attenuation correction, the predictive uncertainty is used to automatically weight the balance between the deep learning μ-map prediction from MRI and the μ-map estimates from the PET emission data from MLAA. When the model is expected to have good performance on a region in a test image, then MLAA has minimal contribution. However, when the model is expected to have poor performance on regions in a test image, then MLAA has a stronger contribution to the attenuation coefficient estimates of those regions.

Specifically, I extend the framework of Ahn et al's MLAA regularized with MR-based priors [96] and generate MR-based priors with a Bayesian convolutional neural network (BCNN) [57] that additionally provides a predictive uncertainty map to automatically modulate the

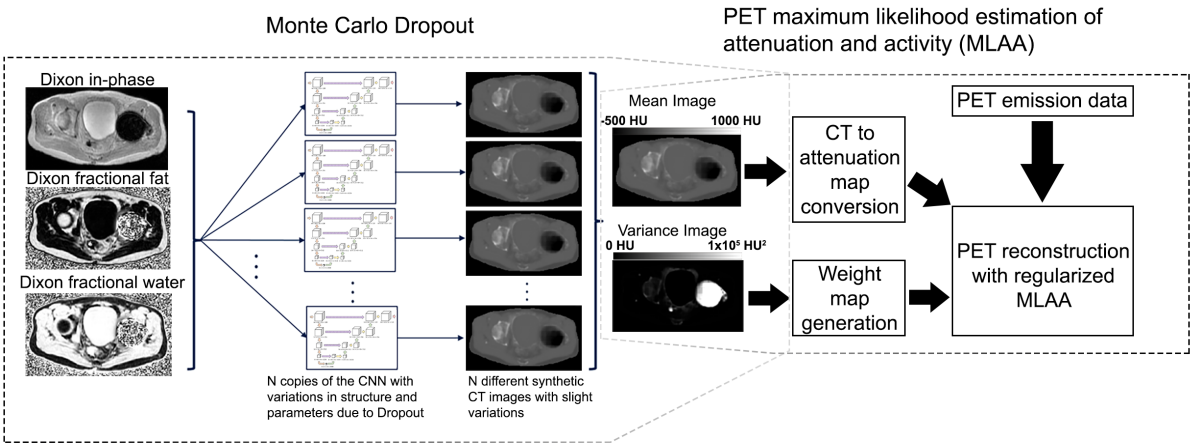

Figure 3.1: Schematic flow of UpCT-MLAA. Monte Carlo Dropout is first performed with the BCNN, then the outputs are provided as inputs to PET reconstruction with regularized MLAA.

strength of the MLAA priors. I demonstrate a proof-of-concept methodology that produces anatomically correct, accurate, and precise μ-maps with high SNR that can recover metal implants for PET/MRI attenuation correction in the pelvis.

3.3 Materials and Methods

UpCT-MLAA is composed of two major elements: initial pseudo-CT characterization with Bayesian deep learning through Monte Carlo Dropout [57] and PET reconstruction with regularized MLAA [96]. The algorithm is depicted in fig. 3.1 and each component is described in detail below.

3.3.1 Bayesian Deep Learning

The architecture of the BCNN is shown in fig. 3.2. It was based on the U-net-like network in [63] with the following modifications: (1) Dropout [57], [60] was included after every convolution, (2) the patch size was increased to 64×64×32 voxels, and (3) each layer's number of channels was increased by 4 times to compensate for the reduction of information capacity due to the Dropout. The PyTorch software package [107] (v0.4.1, http//pytorch.org) was

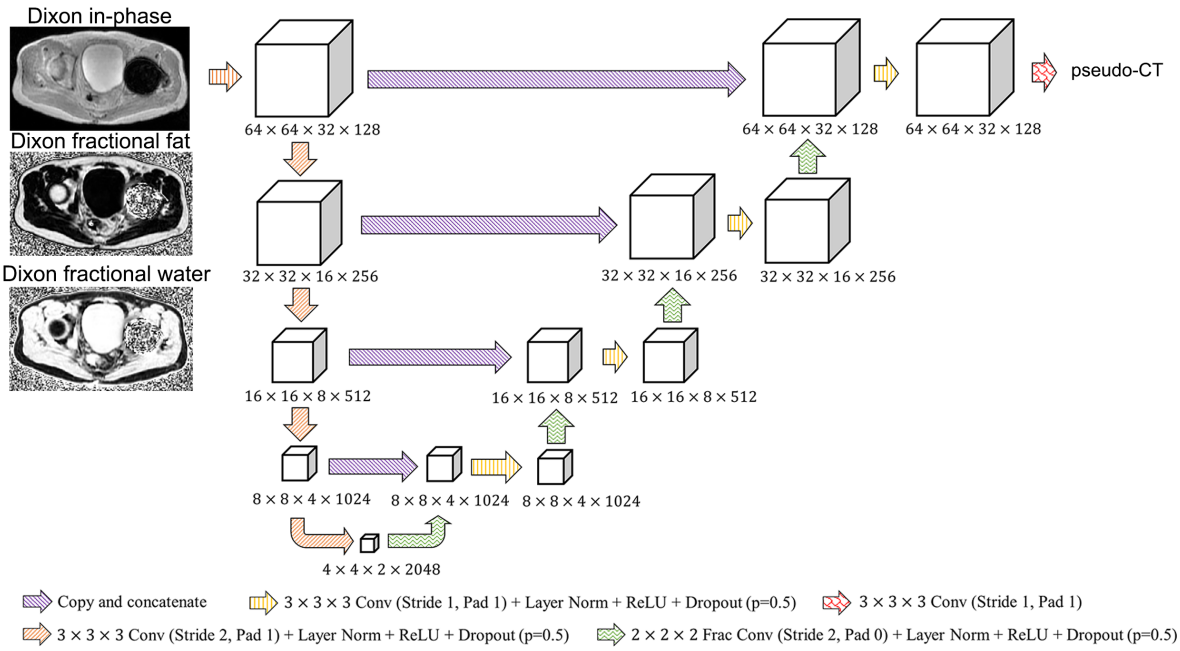

Figure 3.2: Deep convolutional neural network architecture used in this work.

used. Inputs to the model were volume patches of the following dimensions and size: 64 pixels × 64 pixels × 32 pixels × 3 channels. Each channel was a volume patch of the bias-corrected and fat-tissue normalized Dixon in-phase image, Dixon fractional fat image, and Dixon fractional water image, respectively, at the same spatial locations [8]. The output was a corresponding pseudo-CT image with size 64 pixels × 64 pixels × 32 pixels × 1 channel. ZTE MRI was not used as inputs to this model since it has been demonstrated that accurate HU estimates can be achieved with only the Dixon MR pulse sequence [8], [65].

Model Training

Model training was performed similarly to my previous work [8]. The loss function was a combination of an L1-loss, gradient difference loss (GDL), and Laplacian difference loss (LDL):

$$Loss = |\mathbf{y} - \hat{\mathbf{y}}| + \lambda_{GDL}\left(|\nabla_x \mathbf{y} - \nabla_x \hat{\mathbf{y}}|^2 + |\nabla_y \mathbf{y} - \nabla_y \hat{\mathbf{y}}|^2 + |\nabla_z \mathbf{y} - \nabla_z \hat{\mathbf{y}}|^2\right) + \lambda_{LDL}\left(|\Delta \mathbf{y} - \Delta \hat{\mathbf{y}}|^2\right) \tag{3.1}$$

where ∇ is the gradient operator, Δ is the Laplacian operator, $\mathbf{y}$ is the ground-truth CT image patch, and $\hat{\mathbf{y}}$ is the output pseudo-CT image patch with $\lambda_{GDL} = 0.01$ and $\lambda_{LDL} = 0.01$. The Adam optimizer [108] (learning rate $= 1 \times 10^{-5}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$) was used to train the neural network. An L2 regularization ($\lambda = 1 \times 10^{-5}$) on the weights of the network was used. He initialization [109] was used and a mini-batch of 4 volumetric patches was used for training on two NVIDIA GTX Titan X Pascal (NVIDIA Corporation, Santa Clara, CA, USA) graphics processing units. The models were trained for approximately 68 hours to achieve 100,000 iterations.

3.3.2 Pseudo-CT prior and weight map

Generation of the pseudo-CT estimate and variance image was performed through Monte Carlo Dropout [57] with the BCNN described above. The Monte Carlo Dropout inference is outlined in fig. 3.1. A total of 243 Monte Carlo samples were performed to generate a pseudo-CT estimate and a variance map:

$$\mathbf{pCT} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{f}_i(\mathbf{x}) \tag{3.2}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} \left(\mathbf{f}_i(\mathbf{x}) - \mathbf{pCT}\right)^2 \tag{3.3}$$

where $\mathbf{f}_i$ is a sample of the BCNN with Dropout, $\mathbf{x}$ is the input Dixon MRI, and N is the number of Monte Carlo samples. Inference took approximately 40 minutes per patient on 8 NVIDIA K80 graphics processing units.

The pseudo-CT estimate was converted to a μ-map with a bilinear model [7] and the

variance map was converted to a weight map with a range of 0.0 to 1.0 with the following empirical transformation:

$$\mathbf{w}\left(\vec{r}\right) = \frac{1}{1 + \exp\left(0.1\left(\left(\frac{\sigma^2(\vec{r})}{1000}\right) - 25\right)\right)} \tag{3.4}$$

where $\sigma^2\left(\vec{r}\right)$ is the variance at voxel position $\vec{r}$. The sigmoidal transformation was calibrated by inspecting the resulting variance maps. It was designed such that the transition band of the sigmoid covers the range of variances in the body and finally saturates at uncertainty values of bowel air and metal artifact regions. With the constants chosen, the transition band of the sigmoid corresponds to variances of 0 to ∼100,000 HU^2 (standard deviations of 0 to ∼300 HU). The weight map was then linearly scaled to have a range of 1×10^3 to 5×10^6, called β_{MR}. The low β_{MR} values correspond to regions with high uncertainty and thus the estimation for these regions would be dominated by the emission data.

The empirical transformation was a heuristic developed to be compatible with the MLAA algorithm of Ahn et al [96]. In the work of Ahn et al, the weight prior, β_{MR}, was defined only by two discrete values: one for the implant region and one for outside the implant region. My work maintained the use of upper and lower bounds and I used a smooth curve that allows for smoothness as the input values goes towards the saturation values. The predicted variance values have a range of $[0,\ \infty)$. I set an upper-bound threshold for variance values based on the observed ranges and visual inspection of which anatomic structures the high-variance regions corresponded to. The complete transformation function then has the following requirement: $f : [0,\ \sigma^2_{max}] \rightarrow [\beta_{MR_{min}}, \beta_{MR_{max}}]$ and I desired a smooth transition towards the saturation values. Thus, I chose a sigmoid function for initial transformation the transformation of the variance values: $f : [0,\ \sigma^2_{max}\ ] \rightarrow [0,\ 1]$, and the linear transformation re-maps it to the range needed for $\beta_{MR} : g : [0,\ 1] \rightarrow [\beta_{MR_{min}}, \beta_{MR_{max}}]$.

The weight map was additionally processed to set weights outside the body (e.g. air voxels) to 0.0 so that these were not included in MLAA reconstruction. A body mask was

generated by thresholding (> -400 HU) the pseudo-CT estimate. The initial body mask was morphologically eroded by a 1-voxel radius sphere. Holes in the body were then filled in with the imfill function (Image Processing Toolbox, MATLAB 2014b) at each axial slice. The body masks were then further refined by removing arms as in my previous work [62].

3.3.3 Sources of uncertainty and variations

There are three different predictive uncertainties that are utilized in my work: total voxel uncertainty–and its components–patch uncertainty, and voxel-wise uncertainty.

Total voxel uncertainty is the combination of patch uncertainty (uncertainty due to changes in input patch) and uncertainty of each voxel for the same input patch (uncertainty due to changes in the model). These can be decoupled and independently estimated.

Patch uncertainty comes from variations of the response of the CNN due to changes in the input data. Whereas voxel uncertainty (for the same input patch) come from variations of the network parameters with respect to the same input. Mathematically, the predictive likelihood for a single voxel can be written completely as follows:

$$p\left(y_i^* |\ \mathbf{X}^*,\ \mathbf{X},\ \mathbf{Y}\right) = \frac{1}{N} \sum_{x^* \in \mathbf{X}^*} \int p\left(y^* | x^*, \theta\right) p\left(\theta | \mathbf{X}, \mathbf{Y}\right) d\theta \tag{3.5}$$

where y_i^* is the predicted value at the i-th voxel, $\mathbf{X}^*$ is the set of neighboring and overlapping input patches, N is the number of patches to predict the value of a single voxel, θ are the network parameters, $\mathbf{X}, \mathbf{Y}$ are the training input/output pairs, and $p\left(\theta | \mathbf{X}, \mathbf{Y}\right)$ is the posterior distribution of the network parameters given the training pairs that is learned during model training.

The final predicted value is obtained by taking the expectation of the model predictions over the predictive likelihood:

$$p\left(y_i^* |\ \mathbf{X}^*,\ \mathbf{X},\ \mathbf{Y}\right) = \frac{1}{N} \sum_{x^* \in \mathbf{X}^*} p\left(y^* \middle| x^*, \hat{\theta}\right) \tag{3.6}$$

$$\approx \frac{1}{NM} \sum_{x^* \in \mathbf{X}^*} \sum_{m=1}^{M} y_{i,m}^* \qquad \text{(By Monte-Carlo approximation of the integrals)} \tag{3.7}$$

$$= \frac{1}{NM} \sum_{x^* \in \mathbf{X}^*} \sum_{m=1}^{M} f_{CNN}\left(x^*, \theta_m\right) \tag{3.8}$$

And the variance is:

$$\sigma_{y_i^*}^2 = E\left[\left(y_i^* - \widehat{y_i^*}\right)^2\right] = \int \left(y_i^* - \widehat{y_i^*}\right)^2 \left(\frac{1}{N} \sum_{x^* \in \mathbf{X}^*} \int p\left(y^* | x^*, \theta\right) p\left(\theta | \mathbf{X}, \mathbf{Y}\right) d\theta\right) dy_i^* \tag{3.9}$$

$$\approx \frac{1}{NM} \sum_{x^* \in \mathbf{X}^*} \sum_{m=1}^{M} \left(y_{i,m}^* - \widehat{y_i^*}\right)^2 \qquad \text{(By Monte-Carlo approximation of the integrals)} \tag{3.10}$$

$$= \frac{1}{NM} \sum_{x^* \in \mathbf{X}^*} \sum_{m=1}^{M} \left(f_{CNN}\left(x^*, \theta_m\right) - \widehat{y_i^*}\right)^2 \tag{3.11}$$

where M is the number of Monte-Carlo samples used in inference.

Voxel uncertainty corresponds to the following term in the predictive likelihood:

$$\int p\left(y^* | x^*, \theta\right) p\left(\theta | \mathbf{X}, \mathbf{Y}\right) d\theta \tag{3.12}$$

and corresponds to the following summation in the prediction and variance:

$$\sum_{m=1}^{M} f_{CNN}\left(x^*, \theta_m\right) \tag{3.13}$$

$$\sum_{m=1}^{M} \left(f_{CNN}\left(x^*, \theta_m\right) - \widehat{y_i^*}\right)^2 \tag{3.14}$$

The patch uncertainty comes from averaging the predictions of different input patches for each single voxel and corresponds to the summation in the prediction and variance:

$$\sum_{x^* \in \mathbf{X}^*} f_{CNN}\left(x^*, \theta_m\right) \tag{3.15}$$

$$\sum_{x^* \in \mathbf{X}^*} \left(f_{CNN}\left(x^*, \theta_m\right) - \widehat{y_i^*}\right)^2 \tag{3.16}$$

Suppose that there is no model uncertainty, and the network parameters are fixed to be $\hat{\theta}$ (only one set of network parameters used in all inferences). The predictive likelihood will then be:

$$p\left(y_i^* |\ \mathbf{X}^*,\ \mathbf{X},\ \mathbf{Y}\right) = \frac{1}{N} \sum_{x^* \in \mathbf{X}^*} p\left(y^* \middle| x^*, \hat{\theta}\right) \tag{3.17}$$

And the final predicted value will be:

$$\widehat{y_i^*} = \frac{1}{N} \sum_{x^* \in \mathbf{X}^*} f_{CNN}\left(x^*, \hat{\theta}\right) \tag{3.18}$$

And the variance will be

$$\sigma_{y_i^*}^2 = \frac{1}{N} \sum_{x^* \in \mathbf{X}^*} \left(f_{CNN}\left(x^*, \hat{\theta}\right) - \widehat{y_i^*}\right)^2 \tag{3.19}$$

Suppose that we do not process overlapping patches and only extract voxel uncertainty, the predictive likelihood will be:

$$p\left(y_i^* |\ x^*,\ \mathbf{X},\ \mathbf{Y}\right) = \int p\left(y^* | x^*, \theta\right) p\left(\theta | \mathbf{X}, \mathbf{Y}\right) d\theta \tag{3.20}$$

And the final predicted value will be:

$$\widehat{y_i^*} = \frac{1}{M} \sum_{m=1}^{M} f_{CNN}\left(x^*, \theta_m\right) \tag{3.21}$$

And the variance will be:

$$\sigma^2_{y_i^*} = \frac{1}{M}\sum_{m=1}^{M}\left(f_{CNN}\left(x^*, \theta_m\right) - \widehat{y_i^*}\right)^2 \tag{3.22}$$

Thus, patch uncertainty and (patch-specific) voxel uncertainty can be independently obtained but are tightly coupled together when calculating total voxel uncertainty. In the final prediction for this work, we utilize total voxel uncertainty that incorporates both patch uncertainty and (patch-specific) voxel uncertainty.

3.3.4 Uncertainty estimation and pseudo-CT prior for robust Maximum Likelihood estimation of Activity and Attenuation (UpCT-MLAA)

UpCT-MLAA is a combination of the outputs of the BCNN and regularized MLAA. The process is depicted in fig. 3.1. MRI and CT images of patients without metal implants were used to train the BCNN.

I explicitly trained the network only on patients without metal implants to force the BCNN to extrapolate on the voxel regions containing metal implant (i.e. "out-of-distribution" features) to maximize the uncertainty in these regions.

Thus, a high variance ($\geq\sim 1\times10^5$ HU^2) emerged in implant regions compared to a low variance in normal anatomy (0 to $\sim 2.5\times10^4$ HU^2) with the uncertainty estimation as can be seen in fig. 3.1. The μ-map estimate and the weight map were then provided to the regularized MLAA [96] to perform PET reconstruction (5 iterations with 28 subsets, each iteration consists of 1 TOF-OSEM iteration and 5 ordered subsets transmission (OSTR) iterations, β_{MR} as described above, $\beta_{smooth} = 2\times10^4$). Specifically, the MR-based regularization term in MLAA is:

$$R_{MR}\left(\mu\right) = \sum_i \frac{\beta_{MR_i}}{2}\left(\mu_i - \mu_i^{MR}\right)^2 \tag{3.23}$$

where i indexes over each voxel in the volume. μ^{MR} is determined from the mean pseudo-CT image and β_{MR} is determined from the variance image through the weight map transformation. The formulation in eq. (3.23) is slightly different than in Section 2.3.2 of [96] but has the same effect.

3.4 Patient Studies

The study was approved by the local Institutional Review Board (IRB). Patients who were imaged with PSMA-11 signed a written informed consent form while the IRB waived the requirement for informed consent for FDG and DOTATATE studies.

Patients with pelvic lesions were scanned using an integrated 3 Tesla time-of-flight PET/MRI system [110] (SIGNA PET/MR, GE Healthcare, Chicago, IL, USA). The patient population consisted of 29 patients (Age = 58.7 ± 13.9 years old, 16 males, 13 females): 10 patients without implants were used for model training, 16 patients without implants were used for evaluation with a CT reference, and three patients with implants were used for evaluation in the presence of metal artifacts.

3.4.1 PET/MRI Acquistion

The PET acquisition on the evaluation set was performed with different radiotracers: ^{18}F-FDG (11 patients), ^{68}Ga-PSMA-11 (7 patients), ^{68}Ga-DOTATATE (1 patient). The PET scan had 600 mm transaxial field-of-view (FOV) and 25 cm axial FOV, with time-of-flight timing resolution of approximately 400 ps. The imaging protocol included a six bed-position whole-body PET/MRI and a dedicated pelvic PET/MRI acquisition. The PET data were acquired for 15-20 min during the dedicated pelvis acquisition, during which clinical MRI sequences and the following MRAC sequences were acquired: Dixon (FOV = $500 \times 500 \times 312$ mm, resolution = 1.95×1.95 mm, slice thickness = 5.2 mm, slice spacing = 2.6 mm, scan time = 18 s) and ZTE MR (cubical FOV = $340 \times 340 \times 340$ mm, isotropic resolution =

$2 \times 2 \times 2$ mm, 1.36 ms readout duration, FA = 0.6°, 4 μs hard RF pulse, scan time = 123 s).

3.4.2 CT Imaging

Helical CT images of the patients were acquired separately on different machines (GE Discovery STE, GE Discovery ST, Siemens Biograph 16, Siemens Biograph 6, Philips Gemini TF ToF 16, Philips Gemini TF ToF 64, Siemens SOMATOM Definition AS) and were co-registered to the MR images using the method outlined below. Multiple CT protocols were used with varying parameter settings (110-130 kVp, 30-494 mA, rotation time = 0.5 s, pitch = 0.6-1.375, 11.5-55 mm/rotation, axial FOV = 500-700 mm, slice thickness = 3-5 mm, matrix size = 512×512). Pre-processing consisted of filling in bowel air with soft-tissue HU values and copying arms from the Dixon-derived pseudo-CT due to the differences in bowel air distribution and the CT scan being acquired with arms up, respectively [62]. MRI and CT image pairs were co-registered using the ANTS [111] registration package and the SyN diffeomorphic deformation model with combined mutual information and cross-correlation metrics [8], [62], [63].

3.4.3 PET reconstructions

In addition to UpCT-MLAA, additional PET reconstructions were performed for comparison.

For each patient without metal implants: (1) UpCT-MLAA was performed and time-of-flight ordered subsets expectation maximization with a point spread function model (TOF-OSEM) [112] (transaxial field of view (FOV) = 600 mm, 2 iterations, 28 subsets, matrix size = 192×192, 89 slices of 2.78 mm thickness) with two μ-maps: (2) ZeDD-CTAC, (3) initial AC estimate of the BCNN (BpCT-AC) and (4) CTAC, for comparison. BpCT-AC is a surrogate for ZeDD-CTAC but without the use of a specialized MR sequence.

For each patient with metal implants, UpCT-MLAA was performed along with (1) naive MLAA, (2 to 4) regularized MLAA with increasing regularization parameters ($\beta_{MR} = [1 \times 10^3,\ 7 \times 10^5,\ 5 \times 10^6]$, constant over the volume), (5) TOF-OSEM with BpCT-AC,

and (6) TOF-OSEM with CTAC for comparison.

3.4.4 Data Analysis

Image error analysis and lesion-based analysis were performed for patients without metal implants: the average (μ) and standard deviation (σ) of the error, mean-absolute-error (MAE), and root-mean-squared-error (RMSE) were computed over voxels that met a minimum signal amplitude and/or signal-to-noise criteria [63]. Global HU and PET SUV comparisons were only performed in voxels with amplitudes > -950 HU in the ground-truth CT to exclude air, and a similar threshold of > 0.01 cm^{-1} attenuation in the CTAC was used for comparison of AC maps. Bone and soft-tissue lesions were identified by a board-certified radiologist. Bone lesions are defined as lesions inside bone or with lesion boundaries within 10 mm of bone [113]. A Wilcoxon signed-rank test was used to compare the SUVmax biases compared to CTAC of individual lesions.

In the cases where a metal implant was present, I qualitatively examined the resulting AC maps of the different reconstructions and quantitatively compared SUVmax with reference CTAC PET. High uptake lesions and lesion-like objects were identified on the PET images reconstructed with UpCT-MLAA and separated into two categories: (1) in-plane with the metal implant, and (2) out-plane of the metal implant. A Wilcoxon signed-rank test was used to compare the SUV and SUVmax values between the different reconstruction methods and CTAC PET.

3.5 Results

3.5.1 Monte Carlo Dropout

Representative images of the output of the BCNN with Monte Carlo Dropout is shown in fig. 3.3. The same mask used for the weight maps was used to remove voxels outside the body. The pseudo-CT images visually resemble the ground-truth CT images for patients

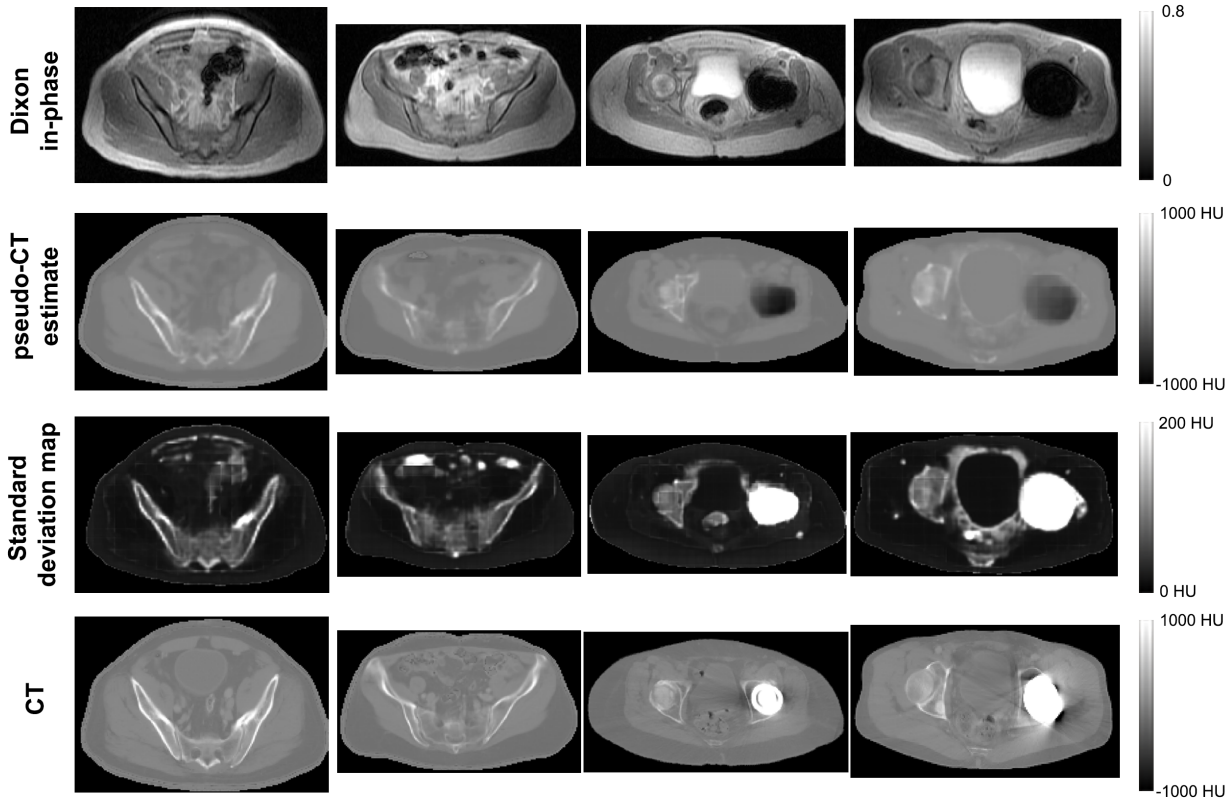

Figure 3.3: Representative intermediate image outputs of the BCNN with Monte Carlo Dropout compared to the reference CT images for patients without metal implants (columns 1 and 2) and patients with metal implants (columns 3 and 4). The voxel-wise standard deviation map is shown instead of variance for better visual depiction. Regions with high standard deviation correspond to bone, bowel air, skin boundary, implants, blood vessels, and regions with likely modeling error (e.g. around the bladder in the standard deviation map in the rightmost column.)

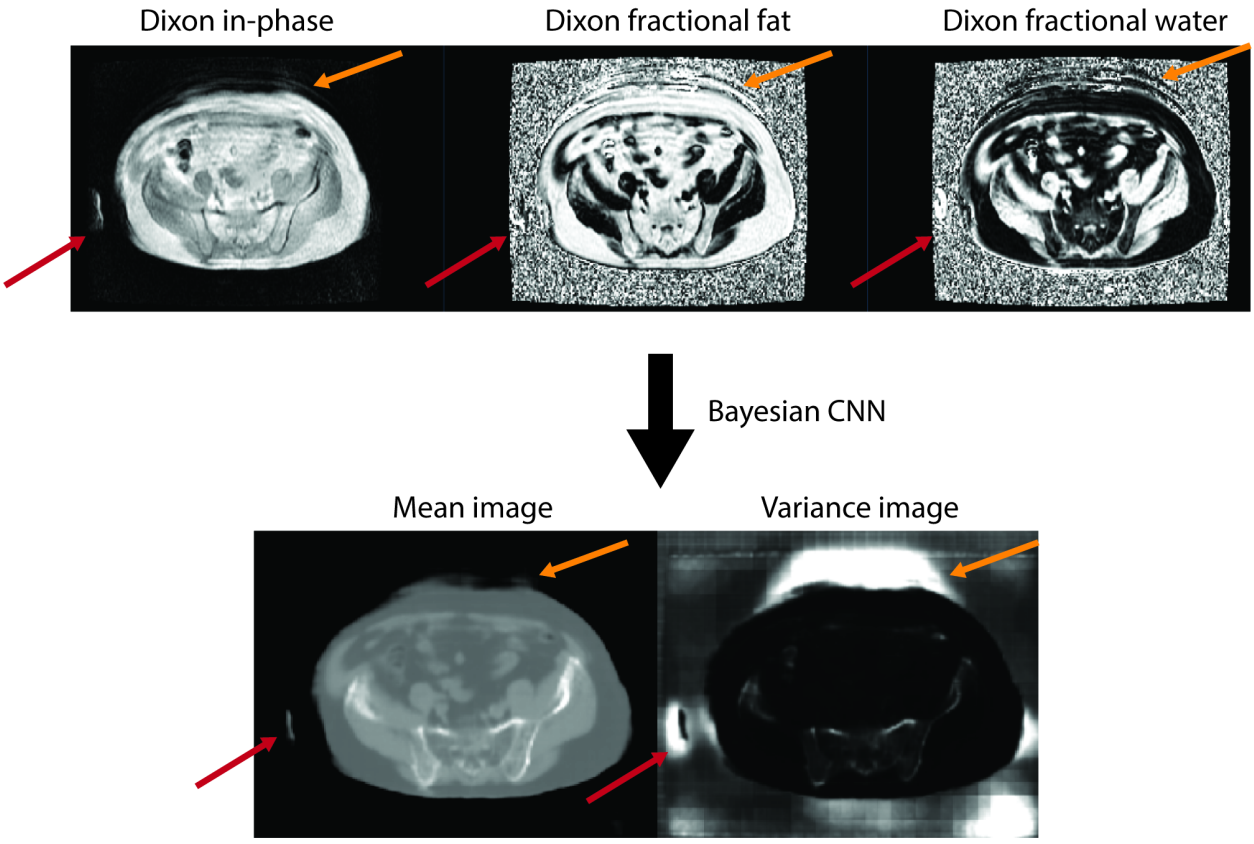

Figure 3.4: Result of uncertainty estimation on without body masks on an MRI with motion artifacts due to breathing and arm truncation due to edge of field-of-view inhomogeneity. The orange arrow points to a region where motion artifacts were present and the dark red arrow points to a region with arm truncation. Both artifact regions were highlighted in the variance image without the network being explicitly trained to highlight these regions.

without implants. While in patients with implants, the metal artifact region in the MRI was assigned air HU values. Nonetheless, the associated standard deviation maps highlighted image structures that the network had high predictive uncertainty. The most important of which are air pockets and the metal implant. The BCNN highlighted these regions and structures in the standard deviation image without being explicitly trained to do so.

An additional example of the uncertainty estimation is provided in fig. 3.4. The input MRI had motion artifacts due to breathing and arm truncation due to inhomogeneity at the edge of the FOV. Like the metal implants, the BCNN highlighted the motion artifact region and arm truncation in the variance image without being explicitly trained to do so.

3.5.2 Patients without implants

The PET reconstruction results for the patients without implants are summarized in fig. 3.5. The RMSE is reported along with the average (μ) and standard deviation (σ) of the error as RMSE ($\mu \pm \sigma$). Additional results for the pseudo-CT, AC maps, and PET data are provided in figs. 3.6 to 3.9.

Pseudo-CT results

The total RMSE for the pseudo-CT compared to gold-standard CT across all volumes were 98 HU (-13 ± 97 HU) for ZeDD-CT and 95 HU (-6.5 ± 94 HU) for BpCT. The BpCT is the same pseudo-CT image used in UpCT-MLAA.

Attenuation coefficient map results

The total RMSE for the AC maps compared to gold-standard CTAC across all volumes were 3.1×10^{-3} cm^{-1} ($-5.0 \times 10^{-4} \pm 3.1 \times 10^{-3}$ cm^{-1}) for ZeDD-CTAC, 3.2×10^{-3} cm^{-1} ($-3.8 \times 10^{-5} \pm 3.2 \times 10^{-3}$ cm^{-1}) for BpCT-AC, and 3.5×10^{-3} cm^{-1} ($-2.6 \times 10^{-5} \pm 3.5 \times 10^{-3}$ cm^{-1}) for UpCT-MLAA-AC.

PET images

The total RMSE for PET images compared to gold-standard CTAC PET across all volumes were 0.023 SUV (-0.005 ± 0.023 SUV) for ZeDD PET, 0.022 SUV ($-8.1 \times 10^{-5} \pm 0.022$ SUV) for BpCT-AC PET and 0.027 SUV ($1.5 \times 10^{-4} \pm 0.027$ SUV) for UpCT-MLAA PET.

Lesion uptake and SUVmax

The results for lesion analysis for patients without implants are shown in fig. 3.5. There were 30 bone lesions and 60 soft tissue lesions across the 16 patient datasets. The RMSE w.r.t. CTAC PET SUV and SUVmax are summarized in table 3.1. For SUVmax of bone lesions, no significant difference was found for ZeDD PET and BpCT-AC PET ($p = 0.116$)

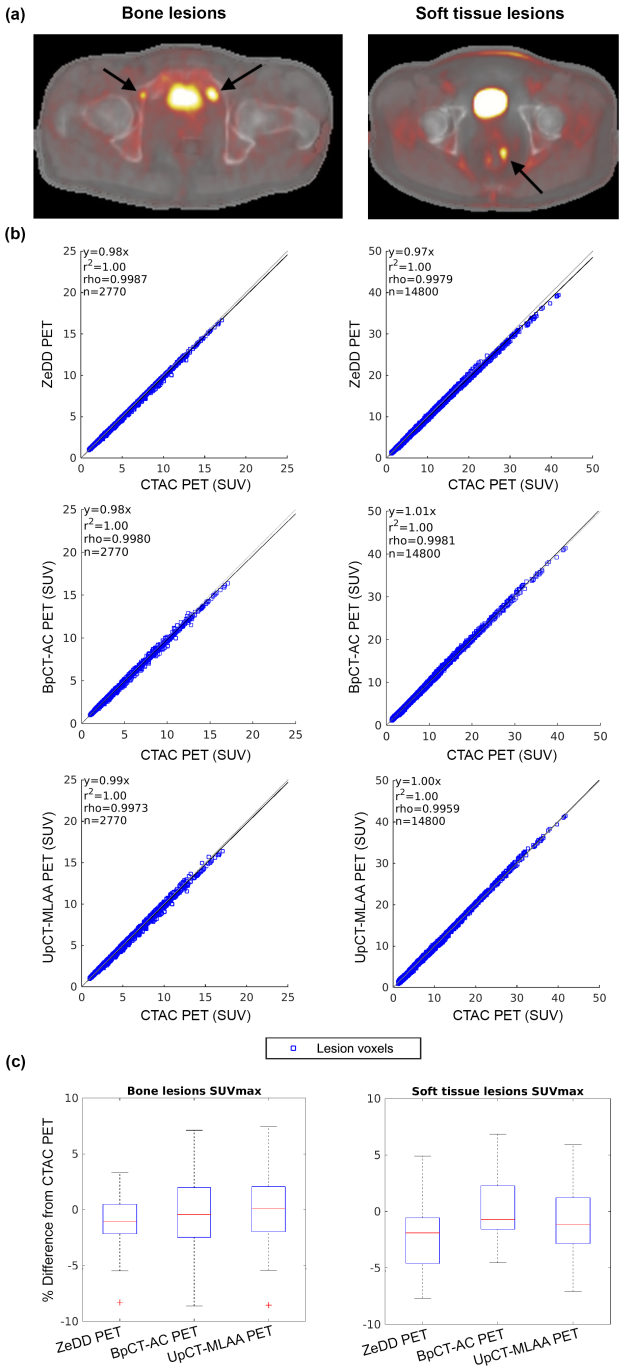

Figure 3.5: Representative images of bone and soft tissue lesions for patients without implants (A, reproduced from (20)), scatter plots of SUV in every lesion voxel (B), and box plots of the SUVmax in each lesion (C). This shows that BpCT-AC and UpCT-MLAA-AC is near equivalent to ZeDD-CTAC in patients without implants when comparing to CTAC.

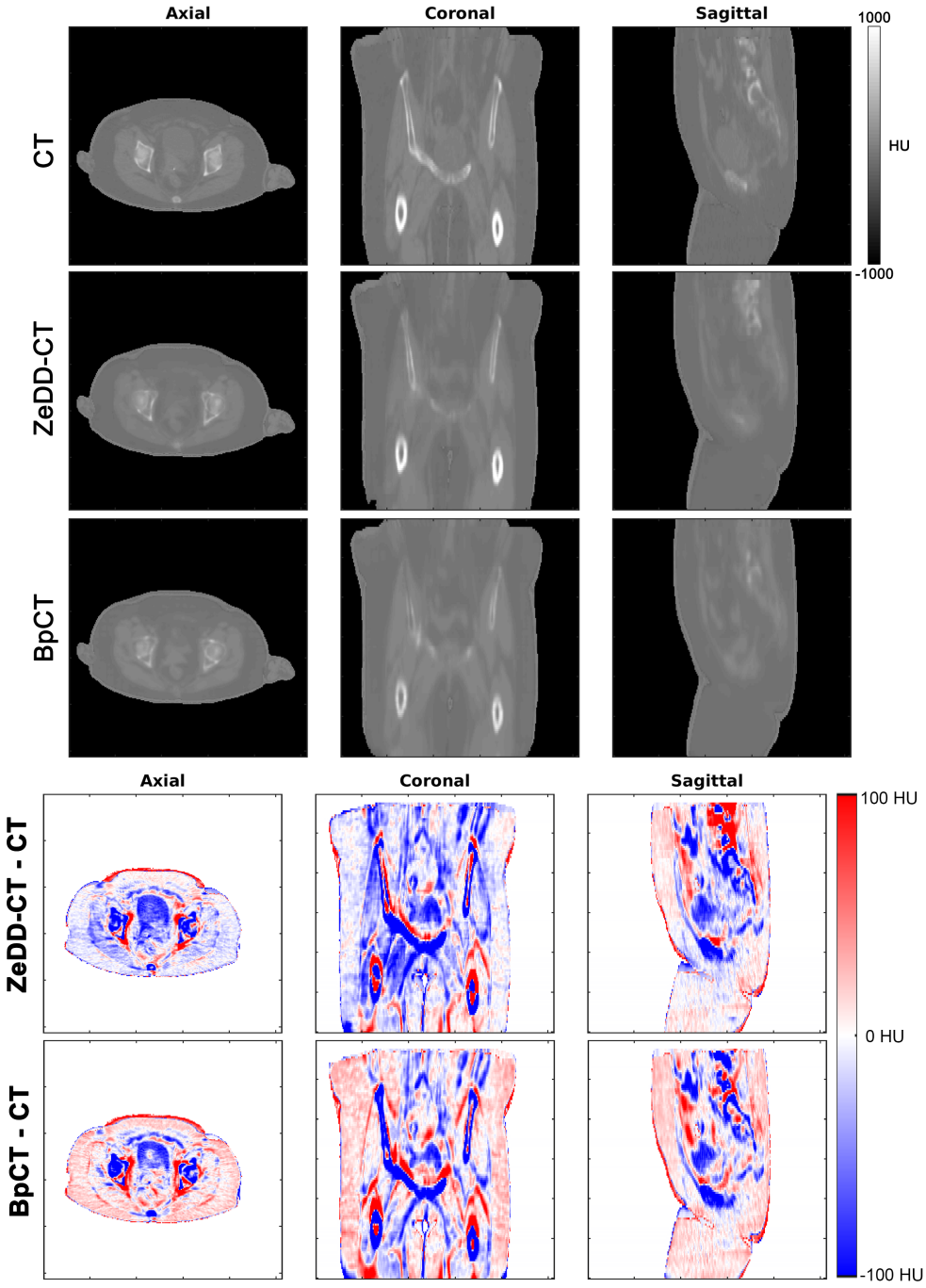

Figure 3.6: CT and difference images of pseudo-CT images for one representative case without implants. Both ZeDD-CT and BpCT have the least RMSE for this patient (RMSE=98.5 HU, and RMSE=98.9 HU, respectively).

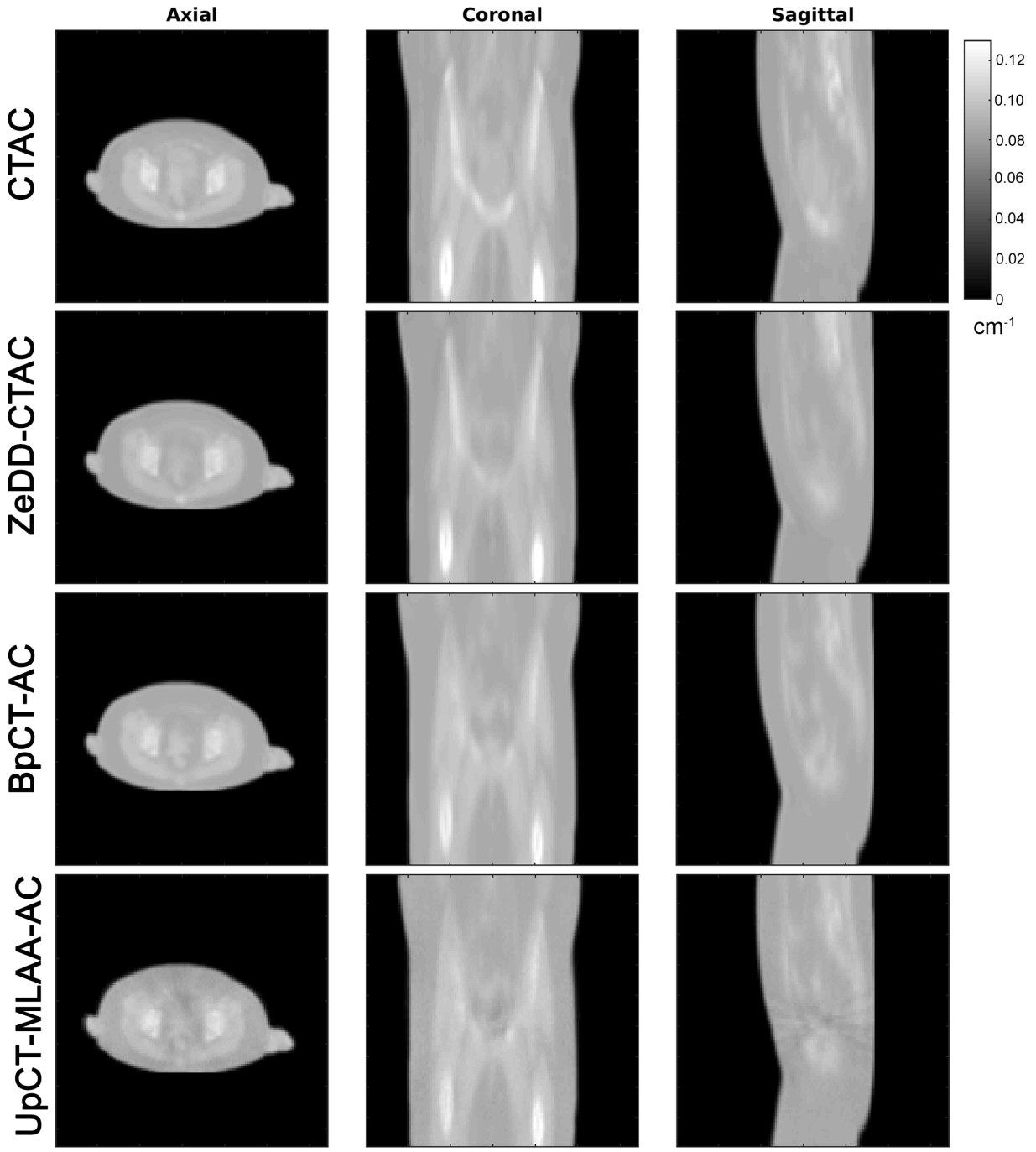

Figure 3.7: CTAC and the different AC maps produced from the different methods for one representative case without implants.

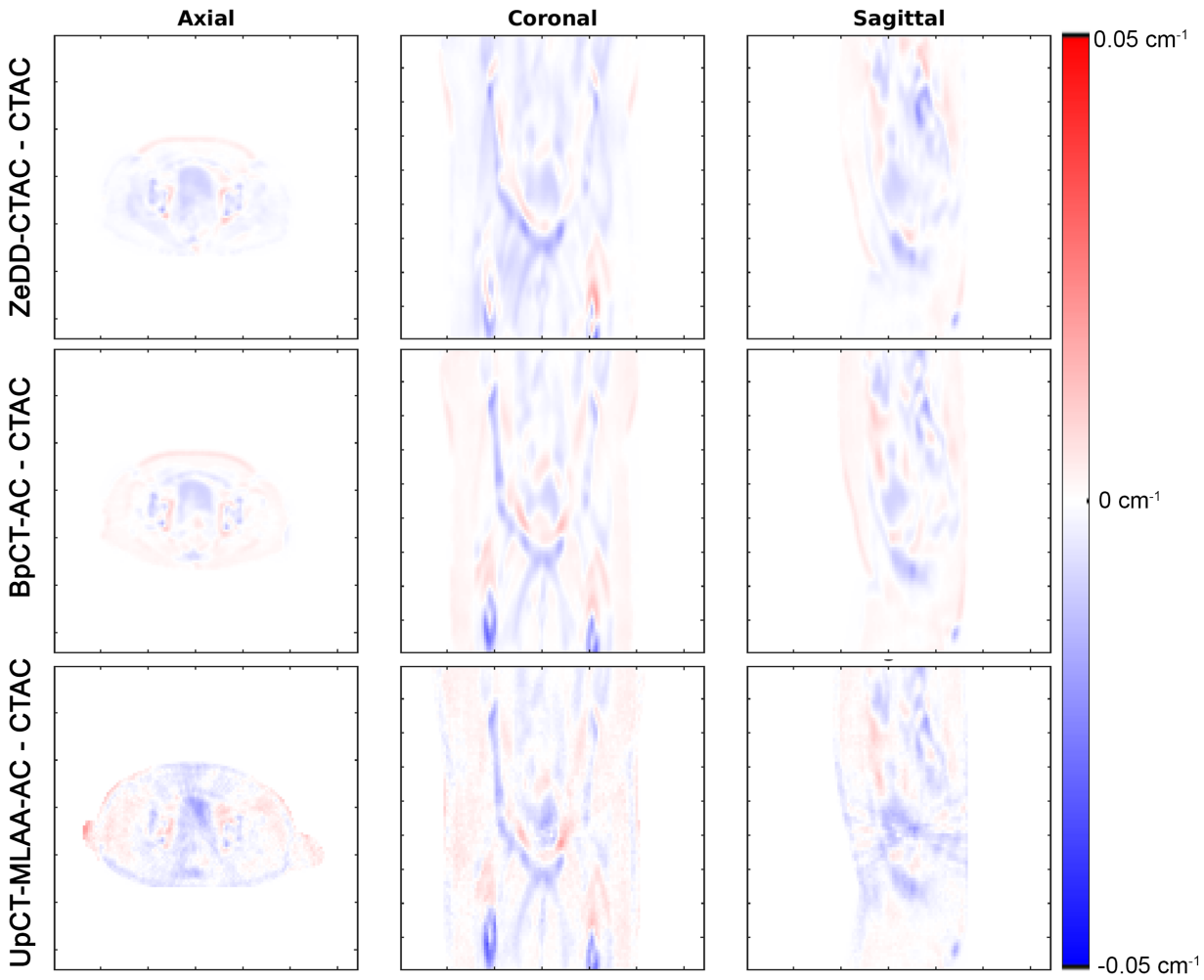

Figure 3.8: Difference images of the AC methods compared to ground-truth CTAC for one representative case without implants.

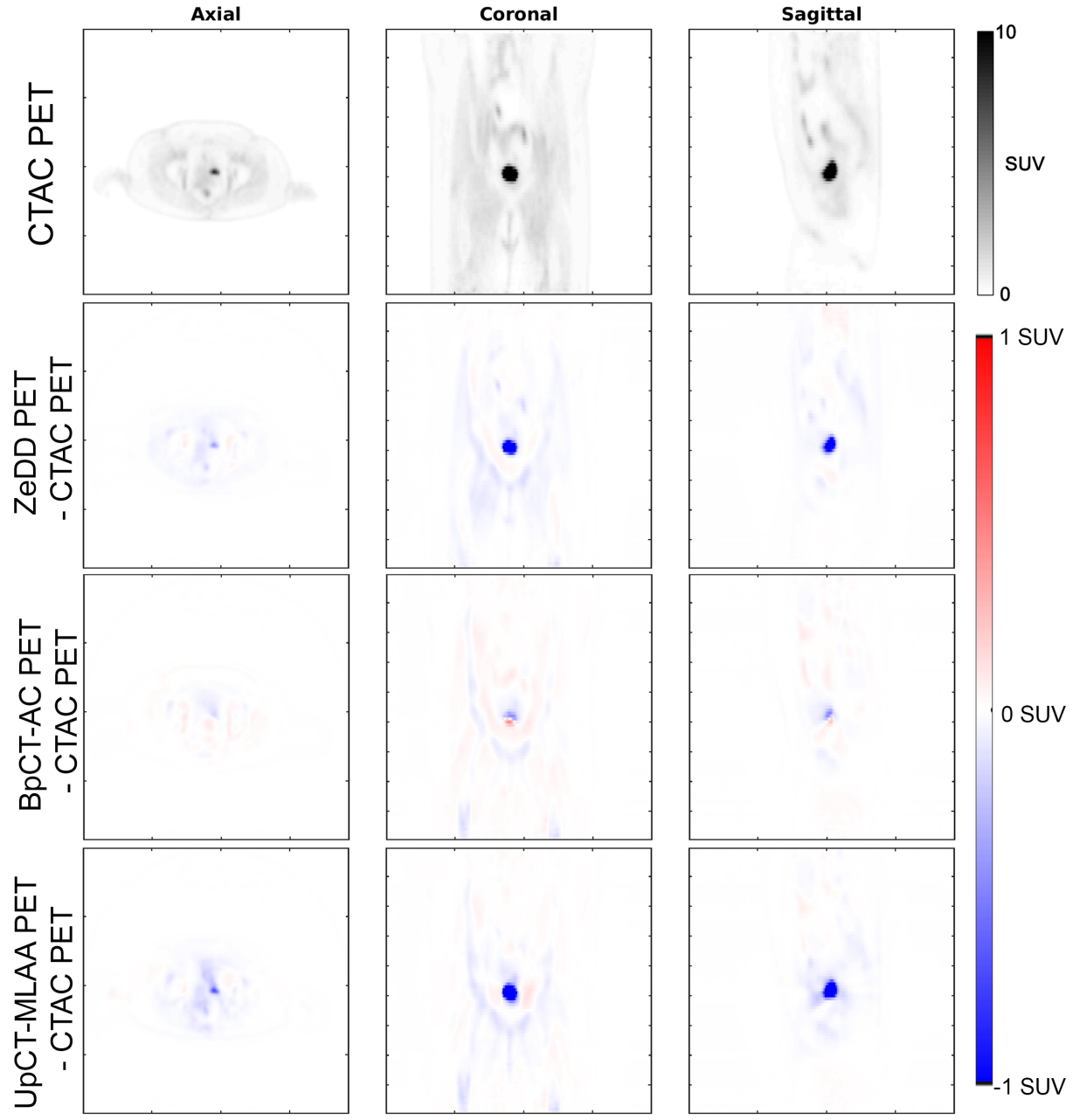

Figure 3.9: PET images of CTAC PET and difference images of the AC methods compared to CTAC PET for one representative case without implants.

Table 3.1: Lesion SUV errors over the volume compared to CTAC in patients without implants

Bone Lesions	
Method	SUV RMSE ($\mu \pm \sigma$)
ZeDD-CTAC	2.6 % (-1.3 ± 2.3 %)
BpCT-AC	3.2 % (-0.9 ± 3.1%)
UpCT-MLAA	3.6 % (-0.3 ± 3.6 %)
Method	SUV_{max} RMSE ($\mu \pm \sigma$)
ZeDD-CTAC	2.6 % (-1.3 ± 2.3 %)
BpCT-AC	3.1 % (-0.3 ± 3.1%)
UpCT-MLAA	3.4 % (0.03 ± 3.4 %)
Soft tissue Lesions	
Method	SUV RMSE ($\mu \pm \sigma$)
ZeDD-CTAC	4.4 % (-2.9 ± 3.3 %)
BpCT-AC	3.5 % (0.01 ± 3.5%)
UpCT-MLAA	4.6 % (-1.1 ± 4.5 %)
Method	SUV_{max} RMSE ($\mu \pm \sigma$)
ZeDD-CTAC	4.1 % (-2.3 ± 3.4 %)
BpCT-AC	3.4 % (-0.1 ± 3.4%)
UpCT-MLAA	4.8 % (-1.6 ± 4.5 %)

while PET ZeDD PET and UpCT-MLAA PET were significantly different ($p = 0.037$). For SUVmax of soft tissue lesions, ZeDD PET and BpCT-AC PET were significantly different ($p < 0.001$) while no significant difference was found between ZeDD PET and UpCT-MLAA PET ($p = 0.16$).

3.5.3 Patients with metal implants

Figures 3.10 and 3.11 show the different AC maps generated with the different reconstruction processes and associated PET image reconstructions on two different radiotracers (^{18}F-FDG and ^{68}Ga-PSMA) and fig. 3.12 shows the summary of the SUVmax results. Additional results for pseudo-CT, AC maps, and PET images are provided in figs. 3.13 to 3.18.

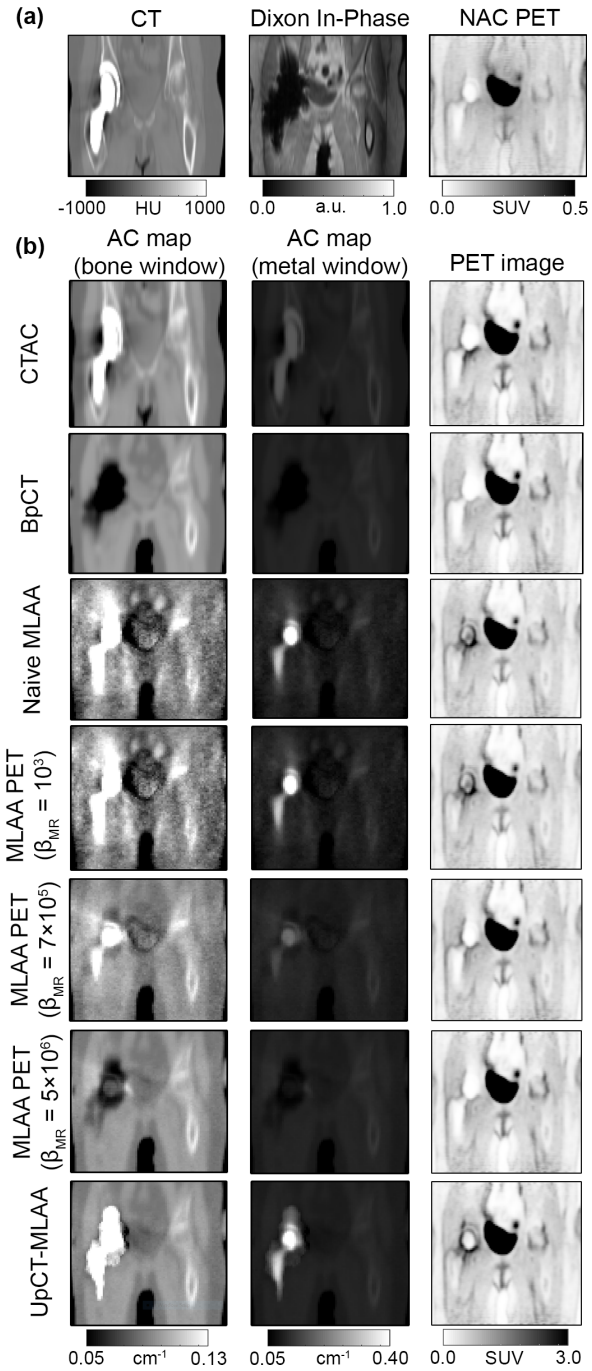

Figure 3.10: Representative images from metal implant patient #3 imaged with ^{18}F-FDG. Shown are the CT, Dixon in-phase, and NAC PET images (a), AC maps (b, first and second column), and associated PET reconstructions (b, third column). The AC maps are shown in two different window levels to highlight bone and soft tissue (b, first column) and the metal implant (b, second column).

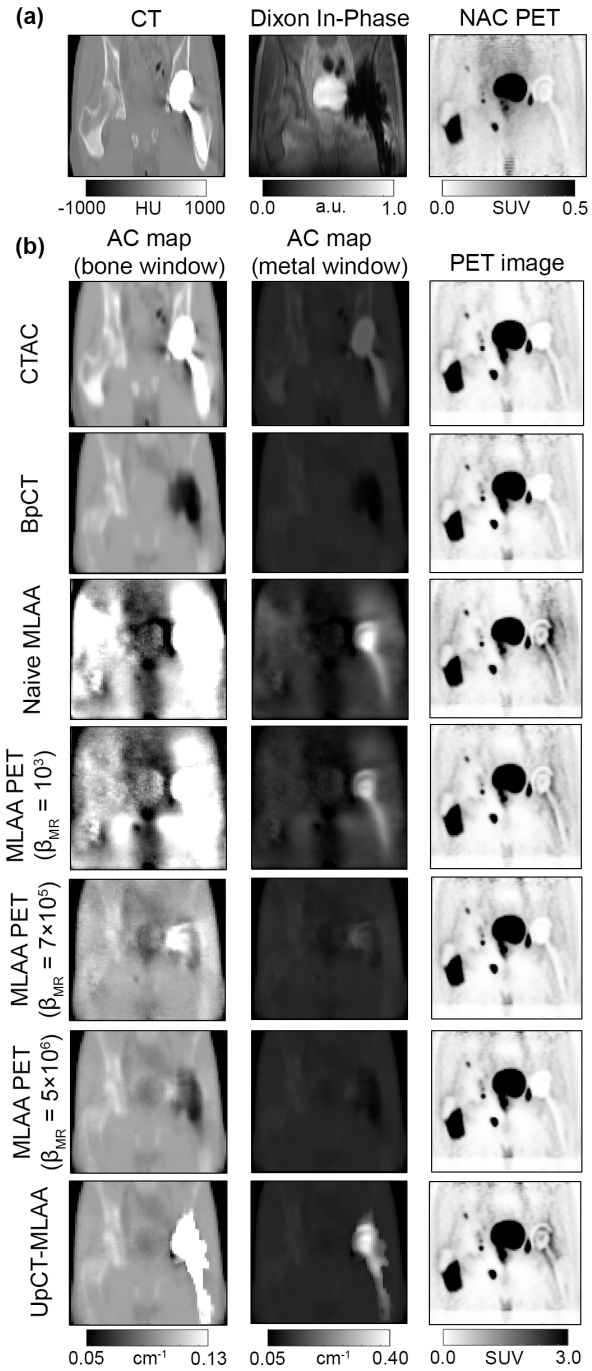

Figure 3.11: Representative images from metal implant patient #1 imaged with ^{68}Ga-PSMA. Shown are the CT, Dixon in-phase, and NAC PET images (a), AC maps (b, first and second column), and associated PET reconstructions (b, third column). The AC maps are shown in two different window levels to highlight bone and soft tissue (b, first column) and the metal implant (b, second column).

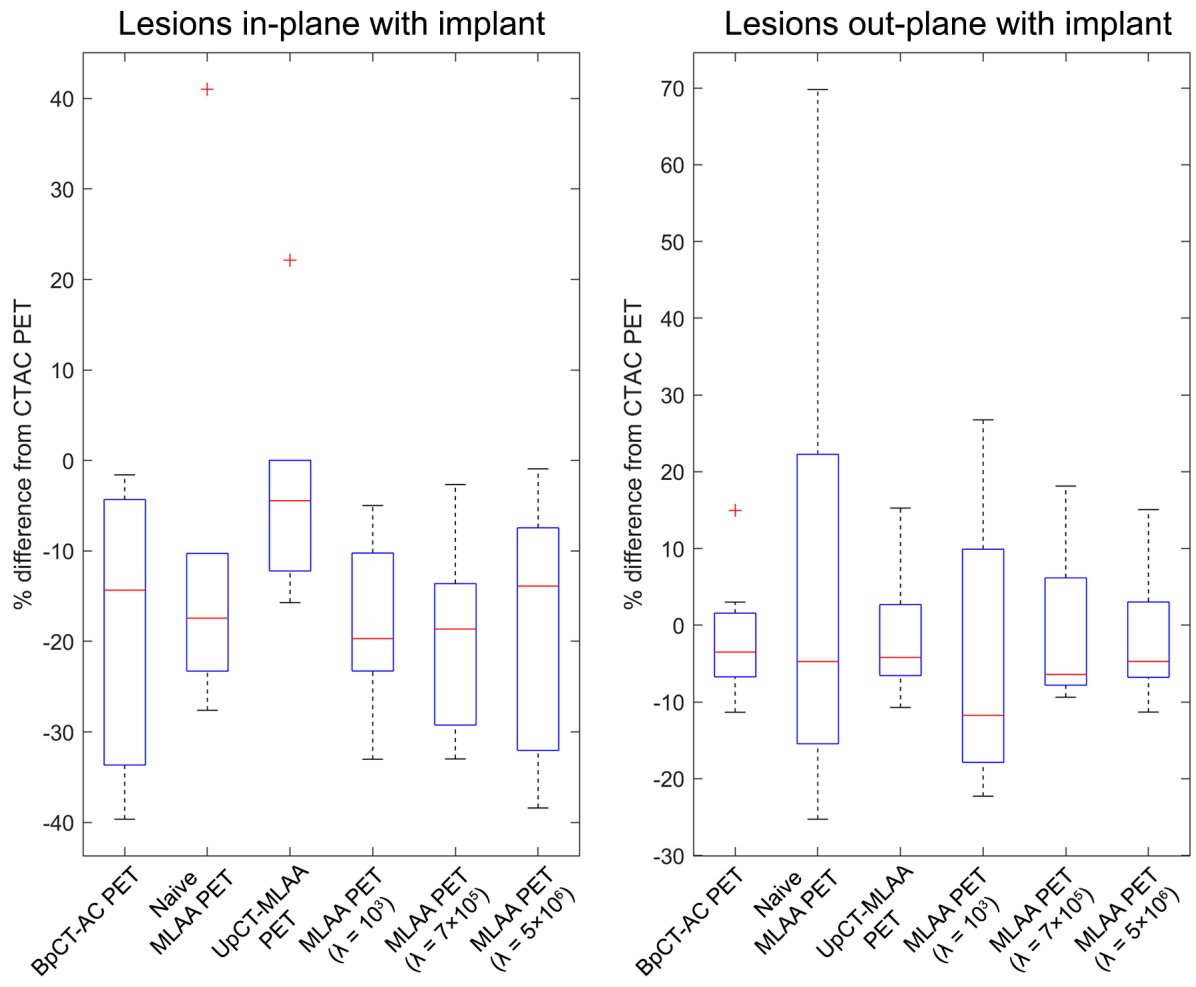

Figure 3.12: Box plot summarizing the results comparing to CTAC PET for patients with implants. The red crosses denote outliers.

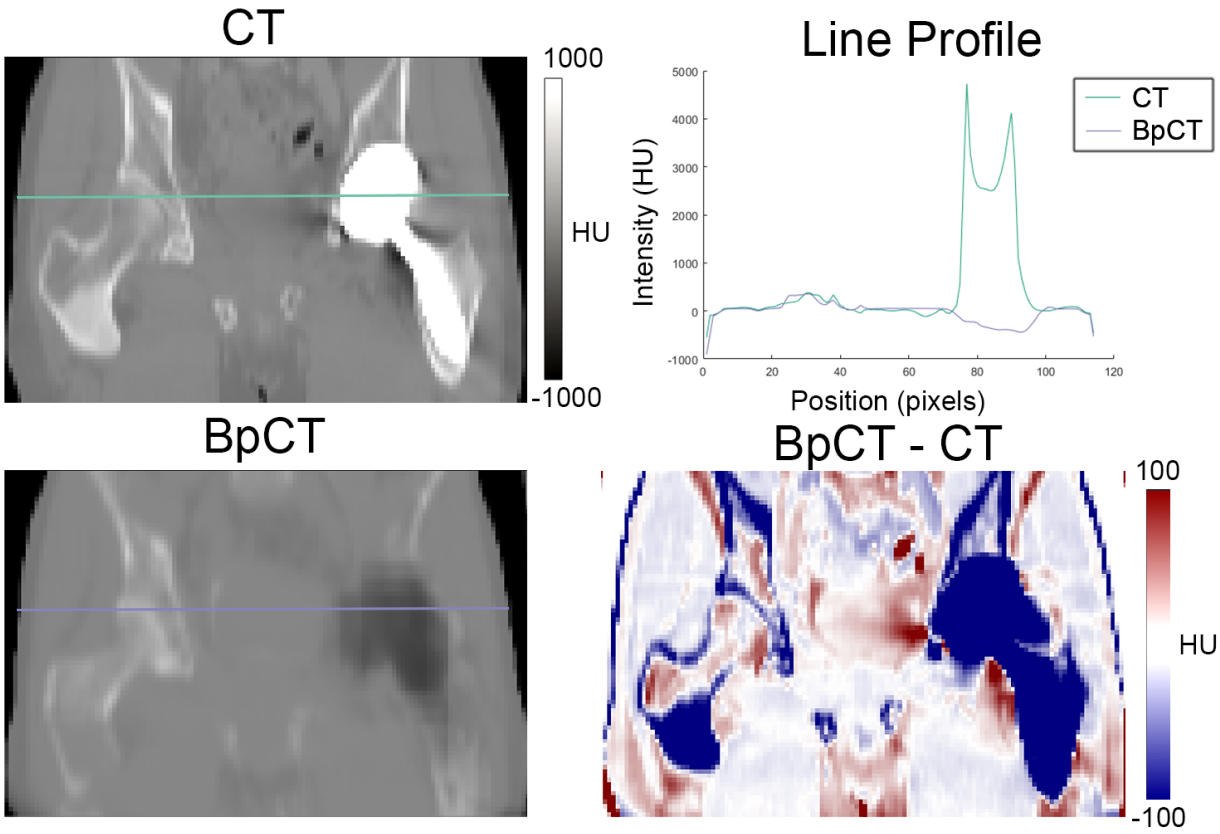

Figure 3.13: CT, pseudo-CT, line profiles, and difference images for a patient with a metal implant imaged with PSMA.

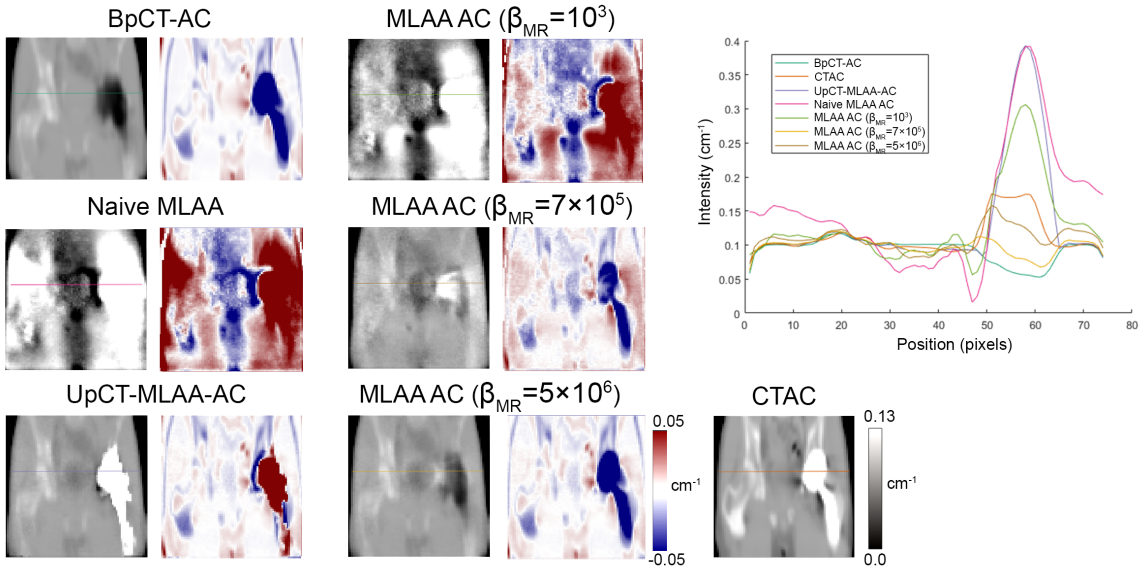

Figure 3.14: AC maps, line profile, and difference images for a patient with a metal implant imaged with PSMA.

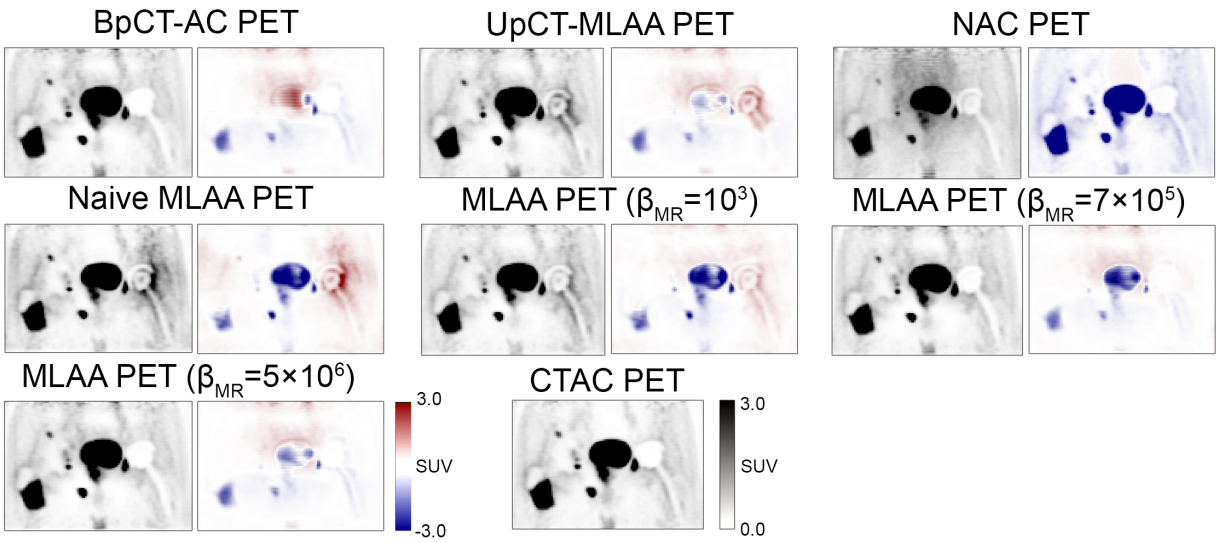

Figure 3.15: PET images and difference images for a patient with a metal implant imaged with PSMA.

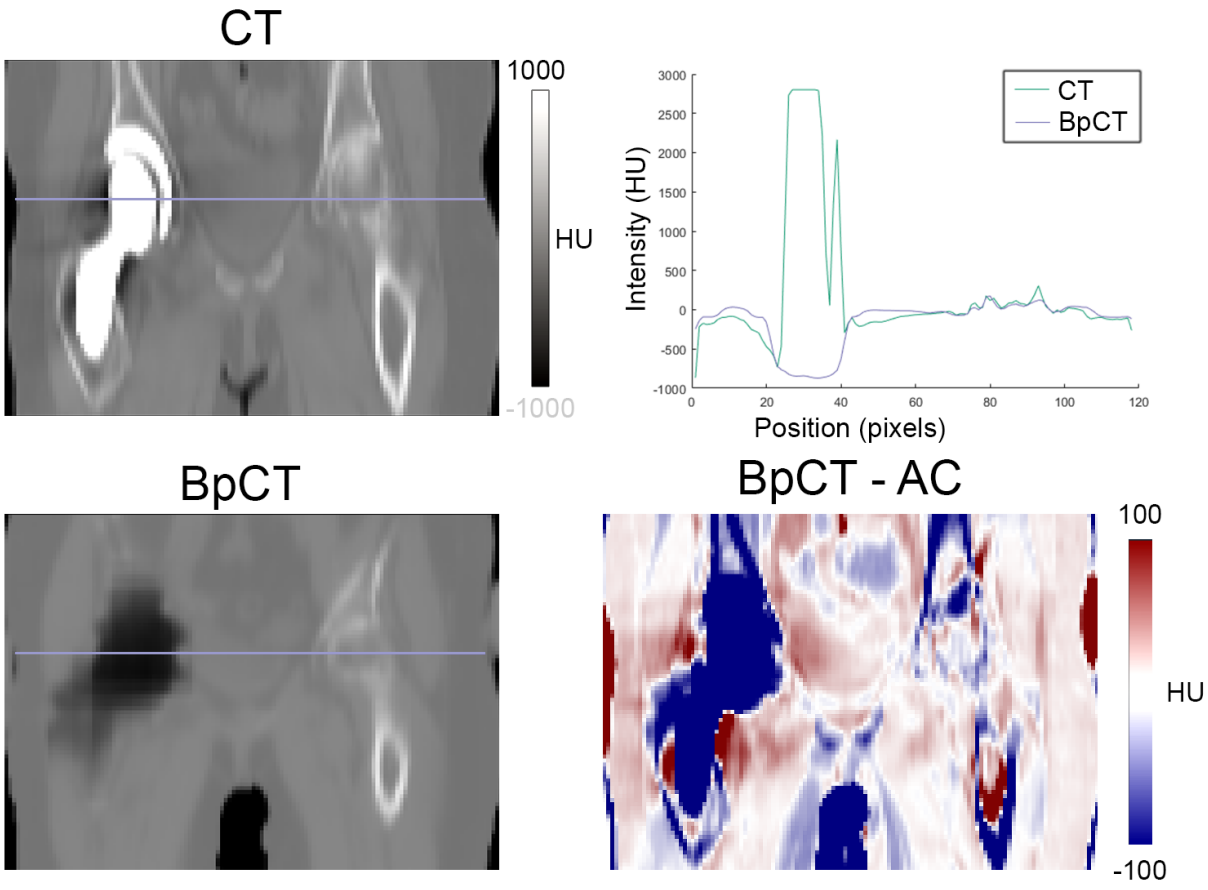

Figure 3.16: CT, pseudo-CT, line profiles, and difference images for a patient with a metal implant imaged with FDG.

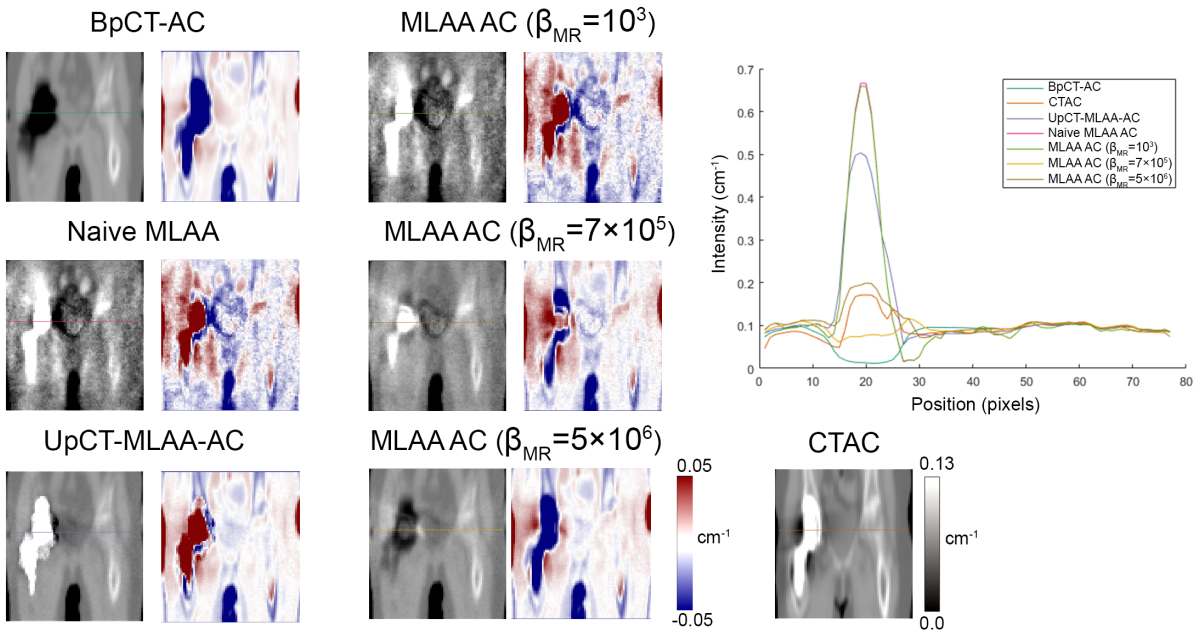

Figure 3.17: AC maps, line profile, and difference images for a patient with a metal implant imaged with FDG.

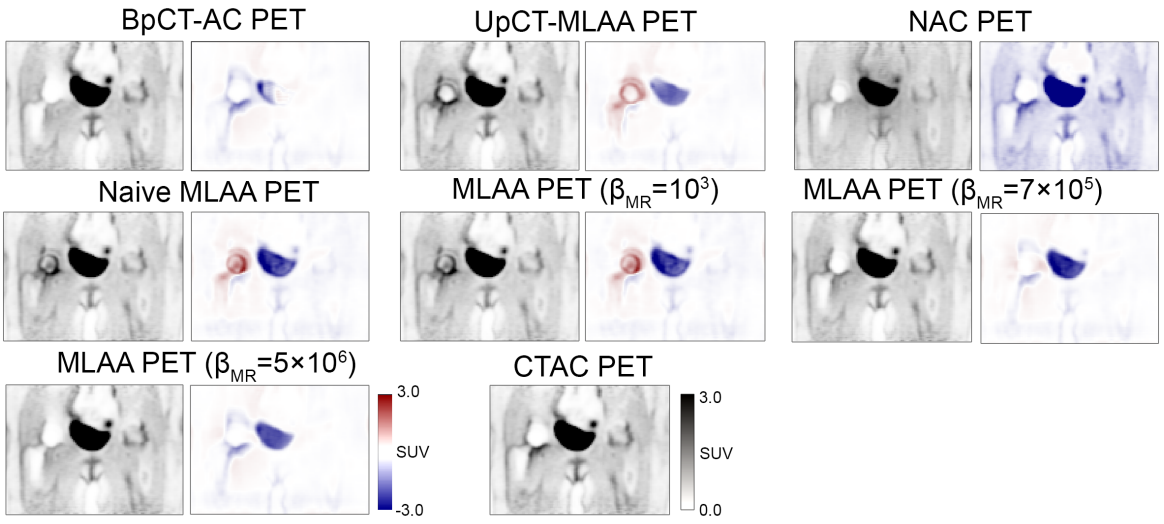

Figure 3.18: PET images and difference images for a patient with a metal implant imaged with FDG.

Metal implant recovery

Figure 3.10b (1st and 2nd column) and fig. 3.11b (1st and 2nd column) show the AC map estimation results.

BpCT-AC filled in the metal implant region with air since the metal artifact in MRI appears as a signal void. Although reconstructing using naive MLAA recovers the metal implant, the AC map was noisy and anatomical structures were difficult to depict. The addition of regularization (increasing β_{MR}) reduces the noise, however over-regularization eliminates the presence of the metal implant. The use of a different radiotracer also influenced reconstruction performance: the MLAA-based methods performed worse when the tracer was ^{68}Ga-PSMA compared to ^{18}F-FDG with low regularization. In contrast, UpCT-MLAA-AC recovered the metal implant while maintaining high SNR depiction of anatomical structures outside the implant region for both radiotracers. The high attenuation coefficients were constrained in the regions where high variance was measured (or where the metal artifact was present on the BpCT AC maps).

PET image reconstruction

Figure 3.10b (3rd column) and fig. 3.11b (3rd column) show the PET image reconstructions results.

Qualitatively, the MLAA-based methods (UpCT-MLAA and Standard MLAA) show uptake around the implant, whereas BpCT-AC PET and CTAC PET show the implant region without any uptake. When compared to the NAC PET, the MLAA-based methods better match what is depicted within the implant region. Quantitatively, table 3.1 summarizes the SUV results for voxels in-plane of the metal implant and out-plane of the metal implant.

SUVmax quantification

Figure 3.12 shows the comparisons of SUVmax of lesions in-plane and out-plane of the metal implant and tables 3.2 and 3.3 list the RMSE values for SUV and SUVmax. There were

Table 3.2: SUV errors over the volume compared to CTAC

Method	SUV RMSE ($\mu,\ \sigma$, p-value)
In-plane with metal implant	
BpCT-AC	0.10 SUV ($\mu = -0.014, \sigma = 0.10,\ p < 0.001$)
Naive MLAA	0.19 SUV ($\mu = 0.06, \sigma = 0.18,\ p < 0.001$)
MLAA PET ($\beta_{MR} = 10^3$)	0.16 SUV ($\mu = 0.047, \sigma = 0.15,\ p < 0.001$)
MLAA PET ($\beta_{MR} = 7 \times 10^5$)	0.09 SUV ($\mu = 0.001, \sigma = 0.09,\ \ p < 0.001$)
MLAA PET ($\beta_{MR} = 5 \times 10^6$)	0.09 SUV ($\mu = -0.008, \sigma = 0.09,\ \ p < 0.001$)
UpCT-MLAA	0.12 SUV ($\mu = 0.018\ SUV, \sigma = 0.12\ SUV,\ p < 0.001$)
Out-plane with metal implant	
BpCT-AC	0.086 SUV ($\mu = 0.009\ SUV, \sigma = 0.085\ SUV,\ p < 0.001$)
Naive MLAA	0.14 SUV ($\mu = 0.018\ SUV, \sigma = 0.14\ SUV,\ p < 0.001$)
MLAA PET ($\beta_{MR} = 10^3$)	0.13 SUV ($\mu = 0.016, \sigma = 0.13,\ \ p < 0.001$)
MLAA PET ($\beta_{MR} = 7 \times 10^5$)	0.09 SUV ($\mu = 0.012, \sigma = 0.09,\ \ p < 0.001$)
MLAA PET ($\beta_{MR} = 5 \times 10^6$)	0.09 SUV ($\mu = 0.010, \sigma = 0.09,\ \ p < 0.001$)
UpCT-MLAA	0.086 SUV ($\mu = 0.012\ SUV, \sigma = 0.085\ SUV,\ p < 0.001$)

Table 3.3: Lesion SUV_{max} percent errors

Method	SUV_{max} % RMSE ($\mu,\ \sigma, p - value$)
In-plane with metal implant ($n = 6$)	
BpCT-AC	24.2% ($\mu = -18.0\%, \sigma = 16.2\%,\ p = 0.03$)
Naive MLAA	26.9% ($\mu = -9.2\%,\ \sigma = 25.3\%,\ p = 0.31$)
MLAA PET ($\beta_{MR} = 10^3$)	21.0% ($\mu = -18.5\%, \sigma = 9.9\%,\ p = 0.03$)
MLAA PET ($\beta_{MR} = 7 \times 10^5$)	22.2% ($\mu = -19.3\%, \sigma = 10.9\%,\ p = 0.03$)
MLAA PET ($\beta_{MR} = 5 \times 10^6$)	23.1% ($\mu = -17.8\%, \sigma = 14.8\%,\ p = 0.03$)
UpCT-MLAA	13.6% ($\mu = -2.4\%,\ \sigma = 13.4\%,\ p = 0.44$)
Out-plane with metal implant ($n = 15$)	
BpCT-AC	6.9% ($\mu = -2.5\%,\ \sigma = -6.5\%,\ p = 0.07$)
Naive MLAA	27.9% ($\mu = 4.7\%,\ \sigma = 27.5\%,\ p = 0.72$)
MLAA PET ($\beta_{MR} = 10^3$)	18.7% ($\mu = -5.2\%, \sigma = 18.0\%,\ p = 0.28$)
MLAA PET ($\beta_{MR} = 7 \times 10^5$)	9.6% ($\mu = -1.3\%, \sigma = 9.5\%,\ p = 0.33$)
MLAA PET ($\beta_{MR} = 5 \times 10^6$)	7.4% ($\mu = -2.1\%, \sigma = 7.1\%,\ p = 0.21$)
UpCT-MLAA	7.1% ($\mu = -1.9\%,\ \sigma = 6.8\%,\ p = 0.19$)

6 lesions in-plane and 15 lesions out-plane with the metal implants across the 3 patients with implants. Only UpCT-MLAA provided relatively low SUVmax quantification errors on lesions both in-plane and out-plane of the metal implant.

For lesions in-plane of the metal implant, BpCT-AC PET had large underestimation of SUVmax, naive MLAA PET had better mean estimation of SUVmax but had a large standard deviation. The addition of light regularization to MLAA improves the RMSE by decreasing the standard deviation at the cost of increased mean error. Increasing regularization increases RMSE but reduces the bias error with increased standard deviation. UpCT-MLAA PET had the best agreement with CTAC PET. Only Naive MLAA and UpCT-MLAA had results where a significant difference could not be found when compared to CTAC ($p > 0.05$).

For lesions out-plane of the metal implant, the trend is reverse for BpCT-AC PET and the MLAA methods. BpCT-AC PET had the best agreement with CTAC PET and the MLAA methods showed decreasing RMSE with increasing regularization. UpCT-MLAA had the second-best agreement with CTAC PET. No significant difference could be found for all methods when compared to CTAC ($p > 0.05$)

3.6 Discussion

This chapter presents the use of a Bayesian deep convolutional neural network to enhance MLAA by providing an accurate pseudo-CT prior alongside predictive uncertainty estimates that automatically modulate the strength of the priors (UpCT-MLAA). The method was evaluated in patients without and with implants with pelvic lesions. The performance for metal implant recovery and uptake estimation in pelvic lesions in patients with metal implants was characterized. This is the first work that demonstrated an MLAA algorithm for PET/MRI that was able to recover metal implants while also accurately depicting detailed anatomic structures in the pelvis. This is also the first work to synergistically combine supervised Bayesian deep learning and MLAA in a coherent framework for simultaneous

PET/MRI reconstruction in the pelvis. The UpCT-MLAA method demonstrated similar quantitative uptake estimation of pelvic lesions to a state-of-the-art attenuation correction method (ZeDD-CT) while additionally providing the capability to perform reasonable PET reconstruction in the presence of metal implants and removing the need of a specialized MR pulse sequence.

One of the major advantages of using MLAA is that it uses the PET emission data to estimate the attenuation coefficients alongside the emission activity. This gives MLAA the capability to truly capture the underlying imaging conditions that the PET photons undergo. This is especially important in simultaneous PET/MRI where true ground-truth attenuation maps cannot be derived. Currently, the most successful methods for obtaining attenuation maps are through deep learning-based methods [63], [65], [84]–[90]. However, these methods are inherently supervised model-based techniques and have limited capacity to capture imaging conditions that were not present in the training set nor conditions that cannot be reliably modeled, such as the movement and mismatch of bowel air and the presence of metal artifacts. Since MLAA derives the attenuation maps from the PET emission data, MLAA can derive actual imaging conditions that supervised model-based techniques are unable to capture. Furthermore, this eliminates the need for specialized MR pulse sequence (such as ZTE for bone) since the bone AC would be estimated by MLAA instead. This would allow for more accurate and precise uptake quantification in simultaneous PET/MRI.

To the best of my knowledge, only a few other methods combines MLAA with deep learning [99]–[102]. Their methods apply deep learning to denoise an MLAA reconstruction by training a deep convolutional neural network to produce an equivalent CTAC from MLAA estimates of activity and attenuation maps. This method inherently requires ground-truth CTAC maps to train the deep convolutional neural network and thus is affected by the same limitations that supervised deep learning and model-based methods have. Unlike their method, my method (UpCT-MLAA) preserves the underlying MLAA reconstruction while still providing the same reduction of crosstalk artifacts and noise.

My approach is different from all other approaches because I leverage supervised Bayesian deep learning uncertainty estimation to detect rare and previously unseen structures in pseudo-CT estimation. There are only a few previous works that estimate uncertainty on pseudo-CT generation [114], [115]. Klages et al [115] utilized a standard deep learning approach and extracted patch uncertainty but did not assess their method on cases with artifacts or implants. Hemsley et al [114] utilized a Bayesian deep learning approach to estimate total predictive uncertainty and similarly demonstrated high uncertainty on metal artifacts. Both approaches were intended for radiotherapy planning and this work is the first to apply uncertainty estimation towards PET/MRI attenuation correction. I demonstrated how likely μ-map errors can be detected and resolved with the use of PET emission data through MLAA.

High uncertainty was present in many different regions. Metal artifact regions had high uncertainty because they were explicitly excluded in the training process—i.e., an out-of-distribution structure. Air pockets had high uncertainty likely because of the inconsistent correspondence of air between MRI and CT—i.e., intrinsic dataset errors. Other image artifacts (such as motion due to breathing) have high uncertainty likely due to the rare occurrence of these features in the training dataset and its inconsistency with the corresponding CT images. Bone had high uncertainty since there is practically no bone signal in the Dixon MRI. Thus, the CNN likely learned to derive bone value based on the surrounding structure and the variance image shows the intrinsic uncertainty and limitations of estimating bone HU values from Dixon MRI. Again, these regions were highlighted by being assigned high uncertainty without the network being explicitly trained to identify these regions.

On evaluation with patients without implants, I demonstrated that BpCT was a sufficient surrogate of ZeDD-CT for attenuation correction across all lesion types: BpCT provided comparable SUV estimation on bone lesions and improved SUV estimation on soft tissue lesions. However, the BpCT images lacked accurate estimation of bone HU values that resulted in average underestimation of bone lesion SUV values (-0.9%). The average

underestimation was reduced with UpCT-MLAA (−0.3%). Although the mean underestimation values improved, the RMSE of UpCT-MLAA was higher than BpCT-AC (3.6% vs. 3.2%, respectively) due to the increase in standard deviation (3.6% vs. 3.1%, respectively). This trend was more apparent for soft tissue lesions. The RMSE, mean error, and standard deviation were worse for UpCT-MLAA vs. BpCT. Since the PET/MRI and CT were acquired in separate sessions, possibly months apart, there may be significant changes in tissue distribution. This could explain the increase in errors of BpCT-AC under UpCT-MLAA.

On the patients with metal implants, UpCT-MLAA was the most comparable to CTAC across all lesion types. Notably, there was an opposing trend in the PET SUVmax results for lesions in/out-plane of the metal implant between BpCT-AC and the MLAA methods. These were likely due to the sources of data for reconstruction. BpCT-AC has attenuation coefficients estimated only from the MRI whereas Naïve MLAA has attenuation coefficients estimated only from the PET emission data. The input MRI were affected by large metal artifacts due to the metal implants that makes the regions appear to be large pockets of air. Thus, in BpCT-AC, the attenuation coefficients of air were assigned to the metal artifact region. For lesions in-plane of the implant, this led to a large bias due to the bulk error in attenuation coefficients and a large variance due to the large range of attenuation coefficients with BpCT-AC, while this is resolved with MLAA. For lesions out-plane of the implant, the opposite trend arises. For MLAA the variance is large due to the noise in the attenuation coefficient estimates. This is resolved in BpCT-AC since the attenuation coefficients are learned for normal anatomical structures that are unaffected by metal artifacts. The combination of BpCT with MLAA through UpCT-MLAA resolved these disparities.

A major challenge to evaluate PET reconstructions in the presence of metal implants is that typical CT protocols for CTAC produce metal implant artifacts that may cause overestimation of uptake and thus does not serve as a true reference. Since my method relies on time-of-flight MLAA, I believe that my method would produce a more accurate AC map, and therefore more accurate SUV map. This is demonstrated by the lower SUV¬max

estimates of UpCT-MLAA compared to CTAC PET. However, to have precise evaluation, a potential approach to evaluate UpCT-MLAA is to use metal artifact reduction techniques on the CT acquisition [103] or by acquiring transmission PET images [116].

Accurate co-registration of CT and MRI with metal implant artifacts was a limitation since the artifacts present themselves differently. Furthermore, the CT and MRI images were acquired in separate sessions. These can be mitigated by acquiring images sequentially in a tri-modality system [117].

Another limitation of this study was the small study population. Having a larger population would allow evaluation with a larger variety of implant configurations and radiotracers and validation of the robustness of the attenuation correction strategy.

Finally, the performance of the algorithm can be further improved. In this study, I only sought to demonstrate the utility of uncertainty estimation with a Bayesian deep learning regime for the attenuation correction in the presence of metal implants: that the structure of the anatomy is preserved and implants can be recovered while still providing similar PET uptake estimation performance in pelvic lesions. My proposed UpCT-MLAA was based on MLAA regularized with MR-based priors [89], which can be viewed as uni-modal Gaussian priors. I speculate that this could be further improved by using Gaussian mixture priors for MLAA as in [97]. The major task to combine these methods would be to learn the Gaussian mixture model parameters from patients with implants. With additional tuning of the algorithm and optimization of the BCNN, UpCT-MLAA can potentially produce the most accurate and precise attenuation coefficients in all tissues and in any imaging conditions.

3.7 Conclusion

I have developed and evaluated an algorithm that utilizes a Bayesian deep convolutional neural network that provides accurate pseudo-CT priors with uncertainty estimation to enhance MLAA PET reconstruction. The uncertainty estimation allows for the detection of "out-of-

distribution" pseudo-CT estimates that MLAA can subsequently correct. I demonstrated quantitative accuracy in pelvic lesions and recovery of metal implants in pelvis PET/MRI.

Chapter 4

Scan-specific Self-supervised Bayesian Deep Non-linear Inversion for Undersampled MRI Reconstruction (DNLINV)

Magnetic resonance imaging is subject to slow acquisition times due to the inherent limitations in data sampling. Recently, supervised deep learning has emerged as a promising technique for reconstructing sub-sampled MRI. However, supervised deep learning requires a large dataset of fully-sampled data. Although unsupervised or self-supervised deep learning methods have emerged to address the limitations of supervised deep learning approaches, they still require a database of images. In contrast, scan-specific deep learning methods learn and reconstruct using only the sub-sampled data from a single scan. Current scan-specific approaches require a fully-sampled auto calibration scan region in k-space that cost additional scan time. Here, we introduce Scan-Specific Self-Supervised Bayesian Deep Non-Linear Inversion (DNLINV) that does not require an auto calibration scan region. DNLINV utilizes a deep image prior-type generative modeling approach and relies on approximate

Bayesian inference to regularize the deep convolutional neural network. We demonstrate our approach on several anatomies, contrasts, and sampling patterns and show improved performance over existing approaches in scan-specific calibrationless parallel imaging and compressed sensing.

4.1 Introduction

Magnetic resonance imaging is subject to slow acquisition times due to the inherent limitations in data sampling. The acquisition speed may be increased by either having gradients with higher slew rates, or by sub-sampling data below the Nyquist limit as in Parallel Imaging [9], [10], and Compressed Sensing [11]. Sub-sampling of data has been the focus of most research efforts as this allows for acceleration of scans without costly hardware changes.

Parallel imaging relies on the redundancy offered by multiple coils when evenly sub-sampling k-space. By having multiple coils with its own local sensitivity profile, each coil image would result in different aliasing patterns. SENSE and SENSE-like approaches model the problem as a solution to a linear system of equations that unravels the aliasing artifacts to solve for the unaliased image with known coil sensitivity profiles [9], [118]. Alternatively, GRAPPA and GRAPPA-like approaches attempt to fill in the missing k-space regions using an auto calibration scan (ACS) [10]. At high acceleration rates, residual aliasing and severe noise amplification are encountered in these methods [25]. In 2D MRI at 3 Tesla, $\sim 2\times$ to $3\times$ acceleration factors are typically possible with these methods.

Compressed sensing relies on a different mechanism. Rather than relying on the redundancy of multiple coils with even sub-sampling, compressed sensing requires that the sub-sampling pattern would result in the aliasing having a noise-like appearance in a sparsifying domain [119]. Then, non-linear algorithms are used to remove the noise-like aliasing patterns [119]. Compressed sensing can achieve higher acceleration factors than SENSE and GRAPPA, up to $\sim 12.5\times$ in certain applications such as MR Angiography [120]. Neverthe-

less, at high acceleration rates, various artifacts may be encountered that obscure anatomy and are difficult to identify [120].

Supervised deep learning has emerged as a promising technique for reconstructing sub-sampled MRI [12]. Supervised deep learning approaches learn to reconstruct images from sub-sampled data using a database of images consisting of matched pairs of sub-sampled data and fully-sampled data. These methods have shown possible acceleration factors that exceed what is possible in parallel imaging and compressed sensing, up to an additional $\sim 2\times$ to $3\times$ more acceleration [12]. However, supervised deep learning requires a large dataset of fully-sampled data.

There are scenarios where acquiring a large high-quality database of fully-sampled data may be impractical or even impossible. For example, simultaneously having 1 mm spatial resolution at 100 ms temporal resolution in 3D dynamic imaging is not possible with current hardware configurations and has required sophisticated reconstruction algorithms [121]. Large high-quality databases are also challenging for SNR limited techniques such as non-proton imaging (e.g. 2H, 23Na, and 31P) and chemical exchange saturation transfer (CEST) MRI. Finally, obtaining large datasets of rare disease conditions are impractical. In one recent study for a differential diagnosis machine learning system, there were only two cases of Susac syndrome that were imaged with MRI over a 10-year period [122].

Recently, unsupervised or self-supervised deep learning have emerged to address the limitations of supervised deep learning approaches [43], [123]–[126]. Rather than estimating the fully-sampled image from the sub-sampled images, these methods utilize only sub-sampled data without the need for fully-sampled data. Unsupervised approaches train a deep neural network using a set of sub-sampled data from different scans whereas self-supervised approaches train a deep neural network using only sub-sampled data of each individual scan with network parameters shared across examples (thus still utilizing a database). Different approaches to accomplish this are currently being actively explored: some have utilized generative adversarial networks (GANs) [43], while others have directly used model-based

approaches [123]–[126]. These methods have shown competitive performance compared to fully supervised methods and have been demonstrated to be superior to classical reconstruction approaches when fully-sampled data is unavailable. However, these unsupervised and self-supervised methods fundamentally still require a database of images.

Unlike the deep learning approaches mentioned above, scan-specific deep learning methods learn and reconstruct using only the sub-sampled data from a single scan (N=1). The major benefit of a scan-specific approach is that it removes the need for a training database and removes the dependency of performance on the size and quality of the training dataset. Scan-specific approaches operate at the lower bound of data sampling and data set size. Thus, scan-specific approaches are generalizable to any hardware configuration, contrast mechanism, pulse sequence, or data sampling scheme, much like classical reconstruction methods.

Currently, there are only a few scan-specific deep learning approaches that exist [127]–[130] compared to approaches that use a training database. These scan-specific methods train a deep convolutional neural network to learn how to interpolate the sub-sampled regions of k-space from the fully-sampled auto-calibration scan region (like GRAPPA) [127], [128], [130],, or rely on ESPIRiT to estimate the coil sensitivity maps and use deep learning to only reconstruct the image [129]. Scan-specific methods are found to have superior performance compared to classical reconstruction approaches and competitive performance compared to supervised deep learning approaches. However, the fully-sampled auto calibration scan region in k-space costs additional scan time.

In this context, ideas from calibrationless parallel imaging and compressed sensing acquisitions that are interesting jointly estimate image content and coil sensitivity profiles without the need for the auto calibration region [36], [37]. However, there are currently no existing scan-specific deep learning methods for calibrationless parallel imaging and compressed sensing.

We introduce Scan-Specific Self-Supervised Bayesian Deep Non-Linear Inversion (DNLINV)

for under-sampled MRI reconstruction. Our contribution is two-fold. First, we utilize a network architecture that allows for the joint estimation of image content and coil sensitivity profiles. Second, we use Bayesian deep learning to statistically regularize the network and reduce overfitting to noise. We demonstrate our approach on several anatomies, sampling patterns, and show superiority over existing approaches in scan-specific calibrationless parallel imaging and compressed sensing.

4.2 Theory

4.2.1 Generative modeling

We define x as the underlying true image. Then, using a SENSE signal model [9], the acquired coil-by-coil k-space data can be written as a linear equation:

$$y_i = SFC_i x + n \quad | \quad 1 \leq i \leq N \tag{4.1}$$

where N is the number of coils, y_i is the k-space data for the i-th coil, C_i is the coil sensitivity matrix for the i-th coil, F is the Fourier transform operator, S is the k-space sampling operator, and x is the underlying MR image. The coil sensitivities are acquired either through a calibration scan [9] or by estimating the sensitivities from an autocalibrating scan region [118]. The optimization problem then reduces to an inverse problem of obtaining x from the k-space measurements y_i.

Using compressed sensing for reconstruction leads to the following optimization problem:

$$\hat{x} = \arg\min \|y - Ax\|_2^2 + \lambda_1 \|\Psi x\|_1 \tag{4.2}$$

$$y = [y_1 \quad y_2 \quad \dots \quad y_N]^T \tag{4.3}$$

$$A = [SFC_1 \quad SFC_2 \quad \dots \quad SFC_N]^T \tag{4.4}$$

Where y is the complete measured k-space data from all coils, A is the forward operator, x is the image to be estimated, Ψ is a sparsifying transformation, and λ_1 is a regularization parameter. This enforces that the image x would be sparse in the transformed domain defined by Ψ. Common choices for Ψ are the Wavelet transform or the finite difference operator [11].

We can recast the compressed sensing model as a generative model in the following manner:

$$\hat{z} = \arg\min_z \frac{1}{2} \left\| y - A\Psi^{-1} z \right\|_2^2 + \lambda_1 \left\| z \right\|_1 \tag{4.5}$$

Instead of solving directly for x, we solve for the sparse coefficients z and generate x from the known transformation Ψ^{-1}:

$$\hat{x} = \Psi^{-1} \hat{z} \tag{4.6}$$

In general, we can write this generative modeling problem as follows:

$$\hat{z} = \arg\min_z \frac{1}{2} \left\| y - Ag(z) \right\|_2^2 + R(z) \tag{4.7}$$

And produce the final image estimate as follows:

$$\hat{x} = g(\hat{z}) \tag{4.8}$$

Where g is a function that generates an image x from a set of coefficients z, and R is a regularization function that specifies the desired properties of z.

We may also want to estimate the generative function g in addition to estimating z.

$$\hat{z}, \hat{g} = \arg\min_{z,g} \frac{1}{2} \|y - Ag(z)\|_2^2 + R(z, g) \tag{4.9}$$

This general generative modeling perspective can then be used to describe the Deep Image Prior [44] method. Deep Image Prior is a generative modeling technique that generates an image from a random noise vector using a learned convolutional neural network. Mathematically, we can write it as:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{2} \|y - Ag_{CNN}(z; \theta)\| \, R(\theta) \tag{4.10}$$

Where g_{CNN} is a deep convolutional neural network, and θ are the parameters of the neural network. The parameters θ can be found using stochastic gradient descent using backpropagation [131] with an optimizer such as Adam [108]. The final image can be obtained similarly:

$$\hat{x} = g_{CNN}(z; \hat{\theta}) \tag{4.11}$$

4.2.2 Joint estimation of image and coil sensitivities

In calibrationless parallel imaging and compressed sensing MRI, the lack of an autocalibration scan region means that the image and coil sensitivities in eq. (1) must be solved jointly. Thus, the problem becomes non-linear. The ENLIVE [36] approach formulates the problem as a low-rank matrix factorization problem:

$$\hat{x}, \hat{C} = \arg\min_{x,C} \frac{1}{2} \|y - SFCx\|_2^2 + \lambda_2(\|x\|_2^2 + \|C\|_2^2) \tag{4.12}$$

$$C = [C_1 \; C_2 \; ... \; C_N]^T \tag{4.13}$$

where λ_2 is a regularization parameter. We can reformulate the deep image prior ap-

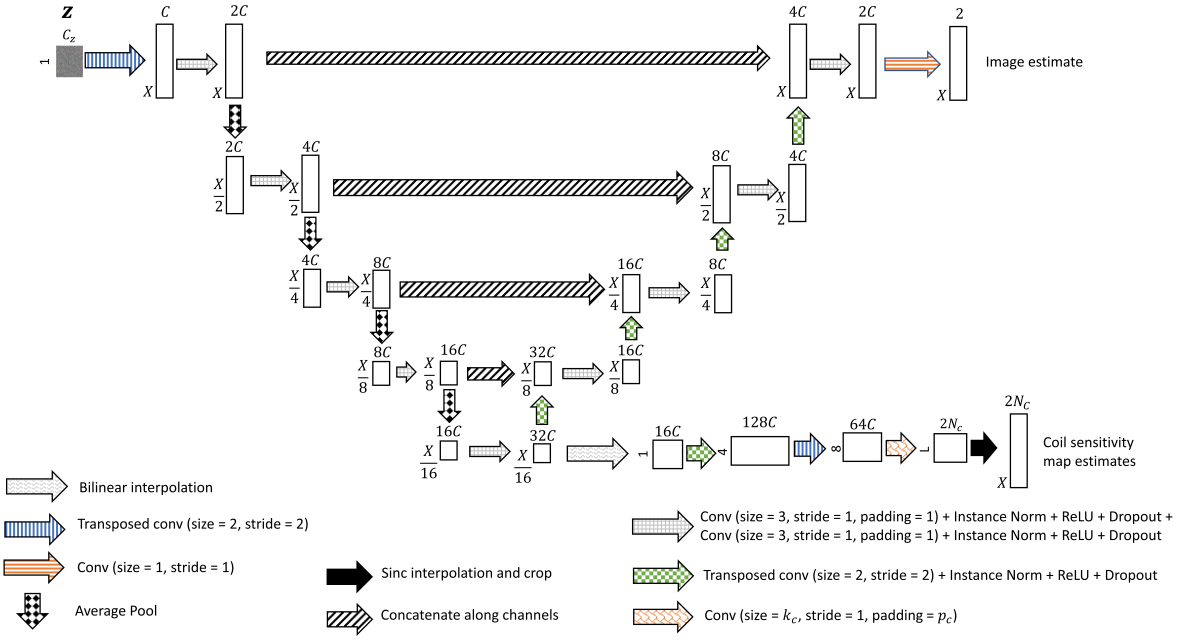

Figure 4.1: Deep convolutional neural network architecture used for generative modeling. The spatial size is denoted at the bottom left of each block (e.g., X) and the number of channels is denoted at the top of each block (e.g., C). The final image estimate and each coil image have 2 channels for real and imaginary dimensions, respectively.

proach by considering a new deep convolutional neural network (DCNN) architecture that simultaneously produces the image and coil sensitivities. The DCNN architecture is shown in fig. 4.1. Mathematically, the optimization problem is then:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{2} \left\| y - SF g_{C_{CNN}}(z;\theta) g_{x_{CNN}}(z;\theta) \right\|_2^2 + R(\theta) \tag{4.14}$$

$$x, C = g_{CNN}(z;\theta) \tag{4.15}$$

This can also be solved using learning and inference procedures typical for deep neural networks.

4.2.3 Bayesian inference

The Deep Image Prior (DIP) formulation (eq. (4.10)) is known to overfit to noise [44], [132]. In fact, stochastic gradient descent is simply gradient descent and deep neural networks achieve zero loss under a scan-specific self-supervised regime (training N=1) where cross-validation cannot be performed [133]. Thus, we rely on Bayesian inference combat the overfitting problem.

In contrast to regularized least-squares optimization (which is equivalent to maximum a posteriori (MAP) estimation) performing Bayesian inference provides uncertainty estimates that has been demonstrated to reduce overfitting of the Deep Image Prior model [44], [132]. In particular, we utilize variational inference [52] and Monte-Carlo Dropout [57].

Reformulating the DIP problem (eq. (4.14))) under Bayesian probabilistic modeling, we obtain the following:

$$y = A g_{CNN}(z;\theta) + n \tag{4.16}$$

$$n \sim \text{Normal}(0, \Sigma) \tag{4.17}$$

$$p\left(\theta|y,z\right)=\frac{p\left(y|\theta,z\right)p\left(\theta,z\right)}{p\left(y\right)} \tag{4.18}$$

However, unlike deep image prior that only solves for the network parameters θ, we also perform inference for the input coefficients to the network z to solve for all unknowns in DNLINV:

$$p\left(\theta,z|y\right)=\frac{p\left(y|\theta,z\right)p\left(\theta,z\right)}{p\left(y\right)} \tag{4.19}$$

We then define the following priors:

$$p\left(z_j\right)=\text{Normal}\left(0,1\right) \tag{4.20}$$

$$p\left(\theta_s\right)=\text{Normal}\left(0,\lambda\right) \tag{4.21}$$

$$p\left(z\right)=\prod_{j=1}^{J}p\left(z_j\right) \tag{4.22}$$

$$p\left(\theta\right)=\prod_{s=1}^{S}p\left(\theta_s\right) \tag{4.23}$$

where λ is the precision for the parameter θ_s, J is the number of latent variables, and S is the total number of network parameters.

Since directly solving for the true posterior is intractable, we utilize a mean-field approximation to the posterior as follows:

$$p\left(\theta,z|y\right)\approx q\left(\theta|y\right)q(z|y) \tag{4.24}$$

Then, each of the approximate posteriors can be solved for independently. The full derivation is provided in 4.A.

To solve for the approximate posterior of the network parameters $q(\theta|y)$, Monte-Carlo Dropout is a method that converts a deep convolutional neural network into a Bayesian deep convolutional network. This is an implicit method that solves for $q(\theta|y)$ through numerical sampling of dropout masks.

To solve for $q\left(z|y\right)$, we derive the evidence lower bound (ELBO) [52], [134], and utilize the reparameterization trick [52], [134]. The ELBO cost function has the following general form:

$$\mathcal{L} = E_{q(z|y)q(\theta|y)}\left[\log \frac{p\left(y|z,\theta\right)p\left(z,\theta\right)}{q\left(z|y\right)q\left(\theta|y\right)}\right] \tag{4.25}$$

and we restrict q(z) to be:

$$q\left(z_j\right) = \text{Normal}\left(\mu_{z_j}, \sigma_{z_j}\right),\ q\left(z\right) = \prod_{j=1}^{J} q\left(z_j\right) \tag{4.26}$$

This leads to the following cost function:

$$\begin{aligned}\mathcal{L} = &-\frac{N}{2}(\log 2\pi + \text{logdet}(\Sigma_c)) - \frac{1}{2M}\sum_{m=1}^{M}\sum_{k=1}^{K}\left(y_k - S_k F g_C\left(z^{(m)};\theta^{(m)}\right)g_x\left(z^{(m)};\theta^{(m)}\right)\right)^T \\ &\Sigma_c^{-1}\left(y_k - S_k F g_C\left(z^{(m)};\theta^{(m)}\right)g_x\left(z^{(m)};\theta^{(m)}\ \right)\right) \\ &- 0.5\sum_{j=1}^{J}\sum_{m=1}^{M}\left(z_j^{(m)}\right)^2 + 0.5\sum_{j=1}^{J}\log\left(\sigma_{z_j}^2\right) + 0.5\end{aligned} \tag{4.27}$$

where y is the observed k-space data, $z^{(m)}$ is m-th sample of the input latent code to the network, $\theta^{(m)}$ is the m-th sample of the network parameters, M is the number of Monte Carlo samples, Σ_c is the coil covariance matrix, g_C is the coil sensitivity map outputs of the network, g_x is the image output of the network, σ_{z_j} is the standard deviation of the approximate posterior $q(z_j)$, J is the number of latent variables, K is the number of k-space sampling points, and N is the total number of k-space samples across all coils.

Finally, the optimization problem is:

$$\underset{\mu_{z_j},\sigma_{z_j}\theta,\Sigma_C}{\arg\min}\ \mathcal{L} \tag{4.28}$$

4.2.4 Implementation details

The optimization problem in eq. (4.28) is solved using backpropagation stochastic gradient descent (SGD) using the Adam optimizer [108].

The nature of MC-Dropout is implicit: there is no explicit realization of $q(\theta|y)$ nor explicit cost function for it. Furthermore, we utilize decoupled weight decay [135] that places the gradient descent update corresponding to the Gaussian prior on the network parameters to the optimization update and thus does not appear in the cost function. Therefore, in the implementation of the optimization, we proceed using only $\mathcal{L}$ as the cost function but each step of SGD is performed with a different dropout mask (as in MC-Dropout). The network parameters are updated at each SGD step including the decoupled weight decay updates.

The coil noise covariance matrix Σ_C is decomposed to a lower triangular Cholesky factorization:

$$\Sigma_C = LL^T \tag{4.29}$$

and L is estimated instead to guarantee that Σ_C is positive semi-definite. The computational complexity is like that of deep learning training. The optimization is based stochastic gradient descent (SGD) that has less than linear convergence. Using reparameterization gradient variational inference [52] reduces the variance of the gradient estimates. Having a larger number of Monte-Carlo samples per SGD step reduces the gradient estimate variance and accelerates convergence at the cost of requiring more memory per SGD step.

The final image and coil sensitivity estimates are then constructed through Monte-Carlo sampling of the expectation:

$$\hat{x} = E_{q(z)q(\theta)}\left[g_x\left(z;\theta\right)\right] \approx \frac{1}{M}\sum_{m=1}^{M} g_x\left(z^{(m)};\theta^{(m)}\right) \tag{4.30}$$

$$\hat{C} = E_{q(z)q(\theta)}\left[g_C\left(z;\theta\right)\right] \approx \frac{1}{M}\sum_{m=1}^{M} g_C\left(z^{(m)};\theta^{(m)}\right) \tag{4.31}$$

4.3 Numerical Experiments

In the spirit of reproducible research, we provide our code and scripts to reproduce all experiments in this paper at `https://www.github.com/LarsonLab/dnlinv`. All reconstruction parameters of the experiments can be found in the code and the results shown in this paper are immediately provided in the repository.

4.3.1 Evaluation metrics

The evaluation metrics used were peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). PSNR was calculated as:

$$MSE = \sum_{r=1}^{R}\left(x_r - x_{ref,r}\ \ \right)^2 \tag{4.32}$$

$$PSNR = -20\log I_{max} + 10\log MSE \tag{4.33}$$

where x_r is the r-th voxel of the reconstructed sum-of-squares image, $x_{ref,r}$ is the r-th voxel of the reference fully-sampled sum-of-squares image, and R is the total number of voxels in an image. SSIM was calculated in the same manner as in the fastMRI challenge [13].

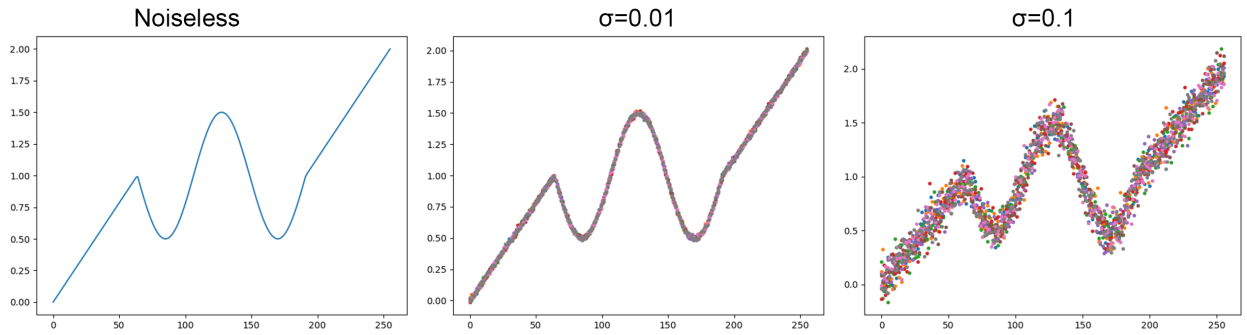

Figure 4.2: (left) 1d noiseless signal, (middle) 1d signal with noise $\sigma = 0.01$, (right) 1d signal with noise $\sigma = 0.1$. There are 8 samples per location for the noisy signals.

Table 4.1: List of inference methods for 1d denoising experiment. MSE = mean-square error, DIP = deep image prior, MC-z = monte carlo on z, MC-Dropout = monte carlo dropout, MAP = maximum a posteriori, DNLINV = deep non-linear inversion

Inference method	Loss function	Inference on z	Inference on θ
Deep image prior (*dip*)	MSE	Fixed z	MAP θ
DIP MC-z (*dip_mc_inference*)	MSE	Draw from $p(z)$	MAP θ
DIP MC-Dropout (*dip_mc_dropout*)	MSE	Fixed z	Implicit $q(\theta)$
DIP MC-z MC-Dropout (*dip_mc_z_mc_dropout*)	MSE	Draw from $p(z)$	Implicit $q(\theta)$
DNLINV	ELBO	Estimate $q(z)$	Implicit $q(\theta)$

4.3.2 1D Denoising

The data was simulated with a noise standard deviation of 0.01 and 0.1 with 8 random samples to emulate the multiple sampling provided by multiple coils. The noiseless and noisy signals are shown in fig. 4.2. The data was then denoised using deep image prior and DNLINV. Note again that the only difference between deep image prior and DNLINV is the inference method.

We include an ablation study of the effect of the inference method on the 1d denoising experiment. The list of inference methods 1d denoising experiments is provided in table 4.1.

Deep image prior was recommended to be used in conjunction with early stopping due to its overfitting behavior [44], [132]. Thus, we also keep track of performance vs. number

of optimization steps.

In addition, we include the denoising performance the maximum likelihood estimator for Gaussian noise (mean of samples).

The performance metric was peak signal-to-noise ratio (PSNR) for the 1d denoising experiments.

4.3.3 SENSE numerical phantom image reconstruction

A SENSE numerical phantom experiment was performed where coil sensitivity profiles are known. An evenly spaced sampling scheme was used to maximize the effect of noise amplification. A birdcage coil configuration with 8 channels was simulated using the SigPy python package (v0.1.16) [136]. A Shepp-Logan phantom was simulated with a maximum signal intensity of 1.0. Additive white Gaussian noise ($\sigma = 0.01$) was added to the image. Deep image prior and DNLINV reconstructions were performed and compared.

4.3.4 In vivo retrospective undersampling

Data from ENLIVE [36] and fastMRI [13] were used for retrospective undersampling experiments.

ENLIVE data

For ENLIVE data, we use both head and knee data under high acceleration factor parallel imaging and under calibrationless parallel imaging and compressed sensing sampling schemes.

Head data was retrospectively undersampled with a CAIPIRINHA sampling pattern [137] at 4x, 9x, 16, and 25x acceleration factors and were reconstructed using L2-ESPIRiT [118], ENLIVE [36], Deep Image Prior [44], and DNLINV, and retrospectively undersampled with a calibrationless variable-density poisson disk sampling pattern [11] at 4x, 7x, and 8.5x acceleration factors and were reconstructed using simultaneous autocalibrating and k-space estimation (SAKE) [37], ENLIVE, Deep Image Prior, and DNLINV.

Knee data was retrospectively undersampled with a calibrationless variable-density poisson disk sampling pattern [11] at 4x, 7x, and 8.5x acceleration factors and were reconstructed using SAKE, ENLIVE, Deep Image Prior, and DNLINV. Additional knee data used in ENLIVE were from the Stanford 3D fast-spin-echo dataset provided in mridata.org and were used to evaluate performance over a dataset.

fastMRI data

We performed similar experiments under autocalibrated and calibrationless settings using fastMRI knee data [13]. Images were retrospectively undersampled with calibrationless parallel imaging and compressed sensing sampling patterns (acceleration factors = [2, 3, 4]) were reconstructed with ENLIVE, DNLINV, and the pre-trained U-net fastMRI. Additionally, retrospectively undersampled with autocalibrated parallel imaging and compressed sensing sampling patterns (acceleration factors = [2, 3, 4]) were reconstructed using L1-ESPIRiT, ENLIVE, DNLINV, and a pre-trained U-net from the fastMRI project [13]. A uniform distribution for selecting k-space lines was used for both sampling patterns.

4.3.5 Ablation study

Finally, we performed an ablation study using the knee ENLIVE data at 5x acceleration. Noise estimation is performed in cases where acquiring noise samples cannot be done and we investigate whether it is needed for accurate reconstructions. Then, the use of non-linear activations (e.g., ReLU) in Deep image prior and DNLINV is what differentiates them from multi-layer convolutional sparse coding [138]. We gradually removed portions of DNLINV and assessed the performance impact. DNLINV was compared against DNLINV without noise estimation (keeping Σ_C constant), deep image prior, DIP MC-z, and DNLINV with linear activations (without ReLU).

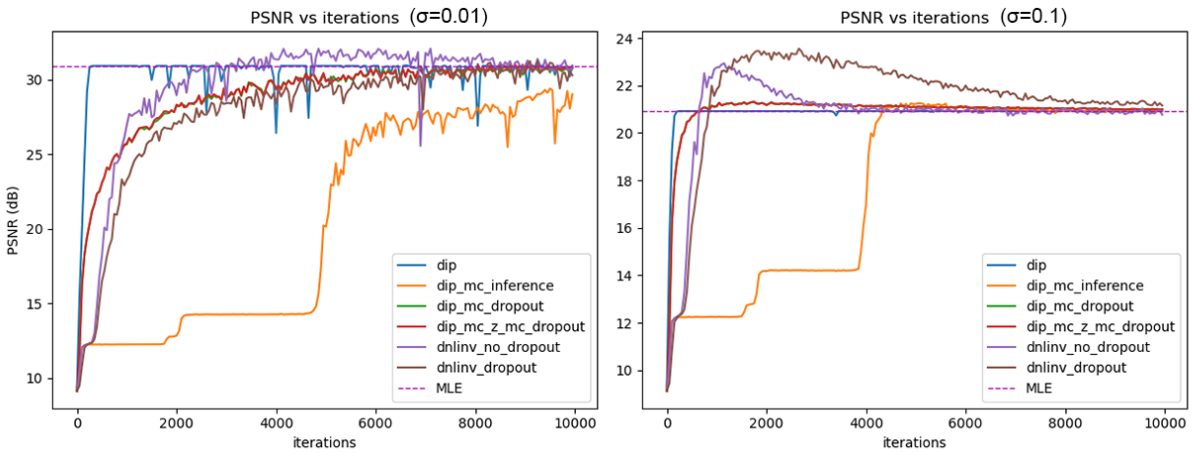

Figure 4.3: 1d denoising experiment results demonstrating the effect of inference method in both (left) high SNR and (right) low SNR regimes. Peak signal-to-noise ratio (PSNR) is plotted over the number of optimization iterations for the inference methods described in table 4.1. The optimal maximum likelihood estimator (MLE) PSNR is shown for comparison.

4.4 Results

4.4.1 1D Denoising Experiments

The results for the 1d denoising experiments are shown on fig. 4.3. Deep image prior by itself ("dip") has the same performance as the maximum likelihood estimator for signals with Gaussian noise. The addition of drawing the latent variables from a prior distribution ("dip_mc_inference") deteriorated performance. Adding MC-Dropout to DIP ("dip_mc_dropout" and "dip_mc_z_mc_dropout") improved denoising performance, even when drawing latent variables from a prior distribution. Finally, DNLINV with dropout ("dnlinv_dropout") achieved the highest possible peak signal-to-noise ratio (PSNR) for the low SNR scenario.

4.4.2 SENSE image reconstruction experiments

The result for the SENSE image reconstruction experiment is shown on fig. 4.4. Deep image prior was severely affected by the noise amplification, which is expected based on the DIP method as well as where the parallel imaging problem is under-determined in this high undersampling regime. DNLINV, on the other hand, was largely unaffected by the noise

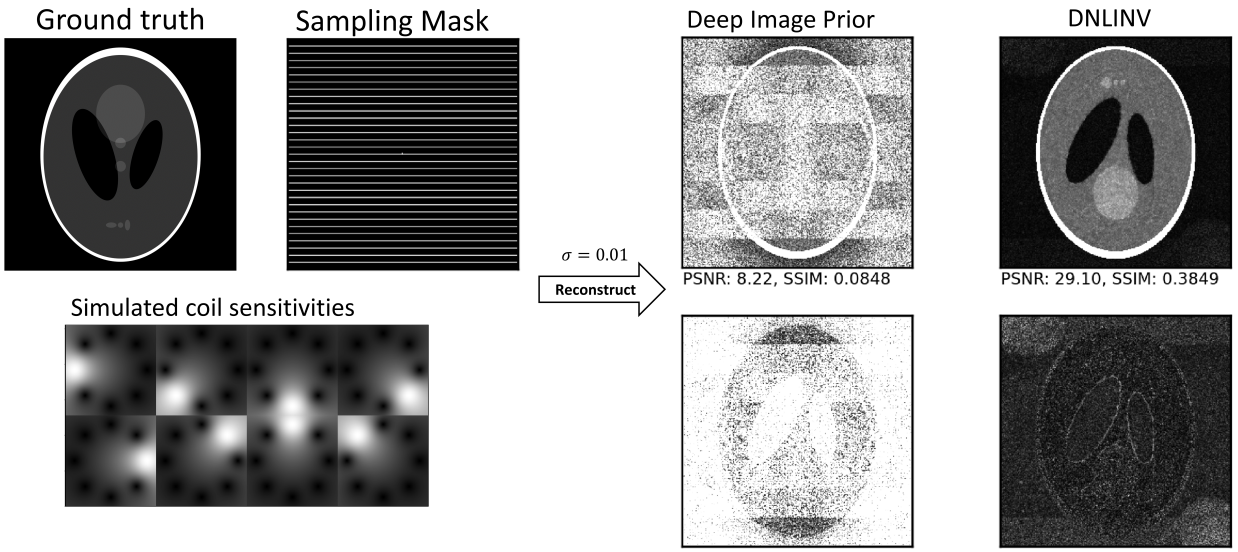

Figure 4.4: SENSE phantom reconstruction experiment for a simulated 8-channel birdcage coil with 8x undersampling. The reconstructed images are on the top row of the right half of the figure and absolute difference images compared to the noiseless ground truth are shown below each reconstruction. The absolute difference image intensities were amplified by 3x to improve visibility. Deep Image Prior shows noise overfitting due to the uncertainty in the reconstruction as characterized by the parallel imaging g-factor. DNLINV is robust to the amplified noise. Both reconstructed images were at 7500 iterations.

amplification as was desired.

Reconstructions at different number of optimization steps are shown in fig. 4.5. Deep image prior immediately overfits to the noise at just a few hundred iterations whereas DNLINV slowly generated image content. DNLINV took around 2500 iterations to reach the peak PSNR and SSIM (although residual aliasing artifacts remained) and took about 7500 iterations to largely eliminate the aliasing artifacts. Like in the 1d denoising experiments, DNLINV beings to overfit to noise at 10,000 iterations and eventually overfits to the noise at 20,000 iterations.

4.4.3 In vivo retrospective undersampling experiments

Figure 4.6 show the main results for the calibrationless parallel imaging and compressed sensing sampling patterns. Aggregate PSNR and SSIM results over different datasets are provided in figs. 4.14 and 4.15. Additional results for calibrationless as well as autocali-

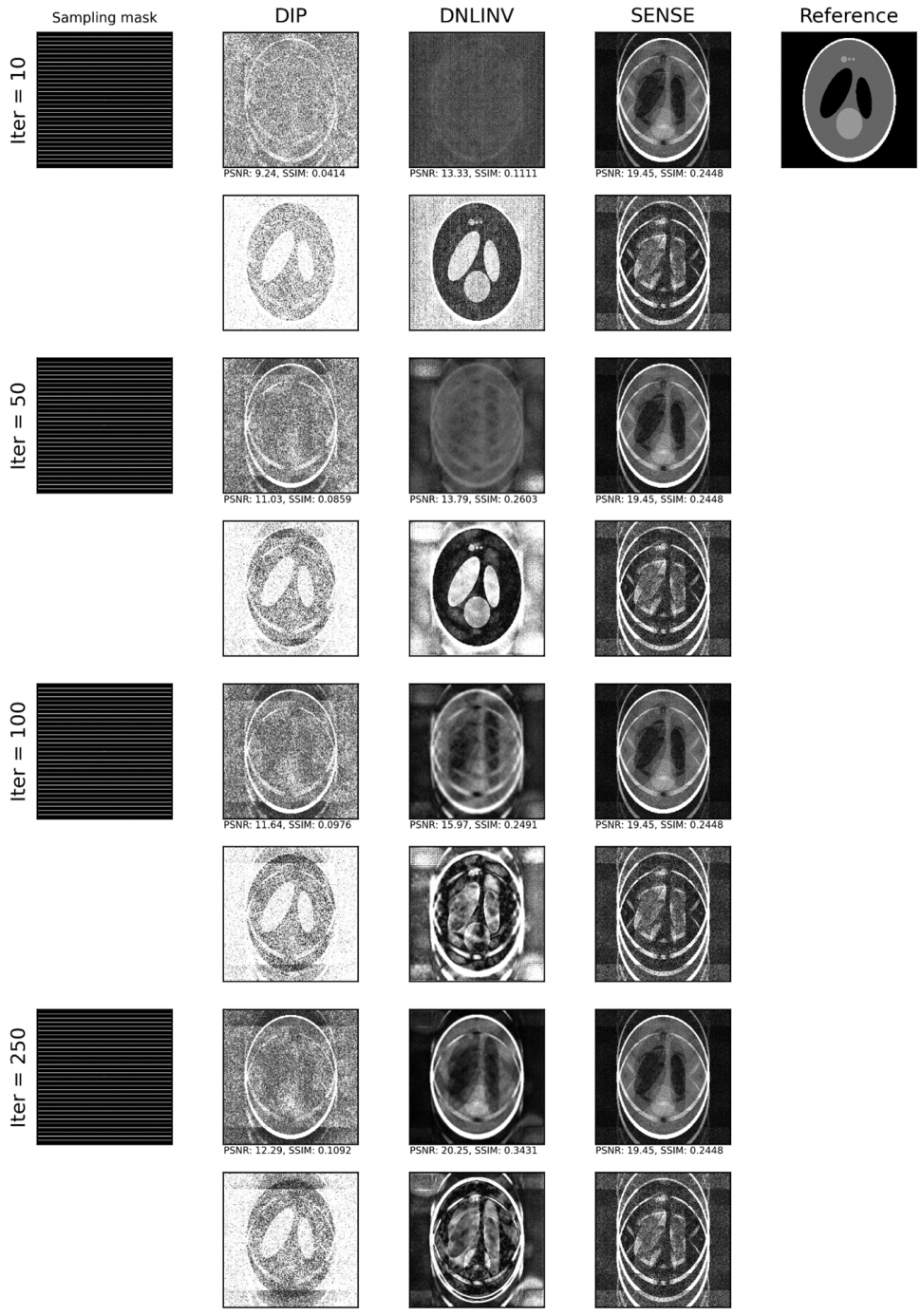

Figure 4.5: Phantom image reconstructions at different iterations for Deep Image Prior (2nd column), Deep Non-linear Inversion (DNLINV) (3rd column), and SENSE ($\lambda_{L2} = 0.005$) (4th column). Under each image is the absolute difference image compared to the noiseless Shepp-Logan phantom with the intensities amplified by 3x.

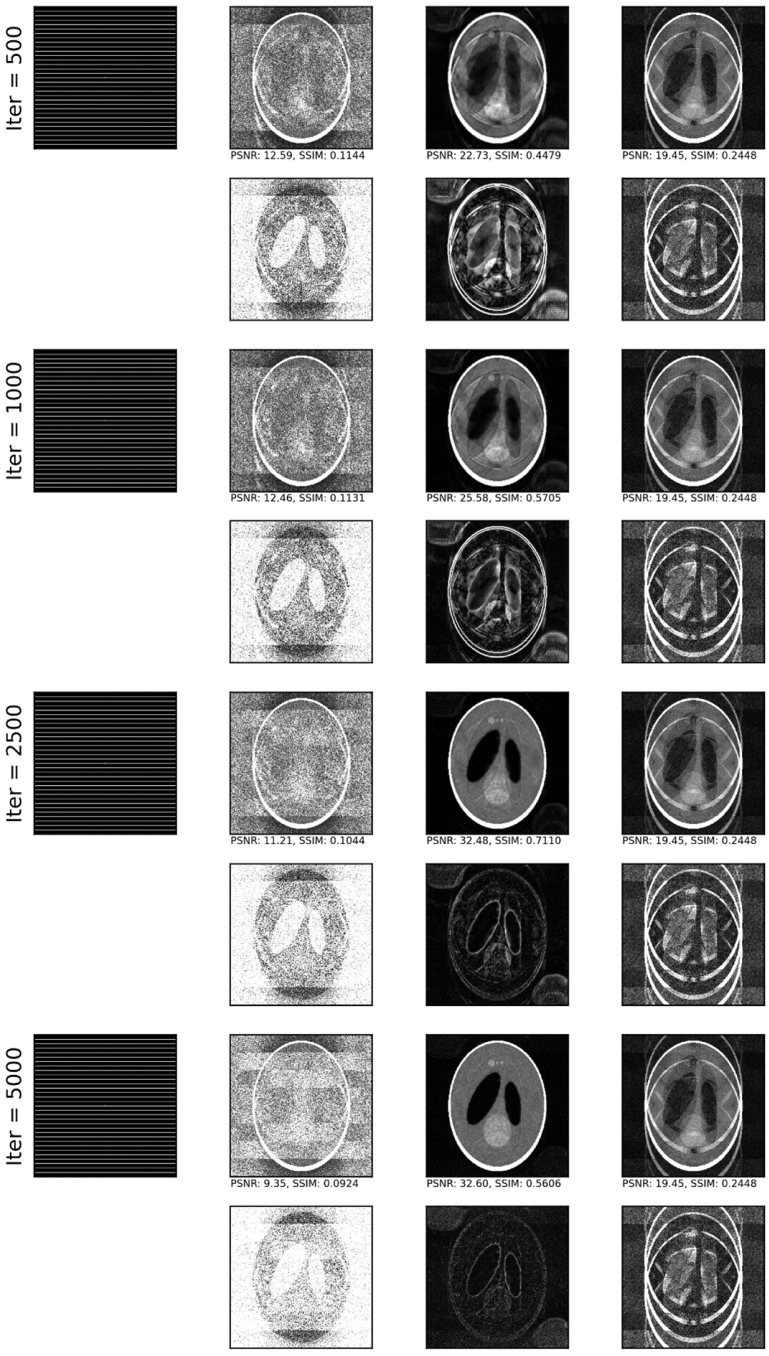

Figure 4.5: (cont'd) Phantom image reconstructions at different iterations for Deep Image Prior (2nd column), Deep Non-linear Inversion (DNLINV) (3rd column), and SENSE ($\lambda_{L2} = 0.005$) (4th column). Under each image is the absolute difference image compared to the noiseless Shepp-Logan phantom with the intensities amplified by 3x.

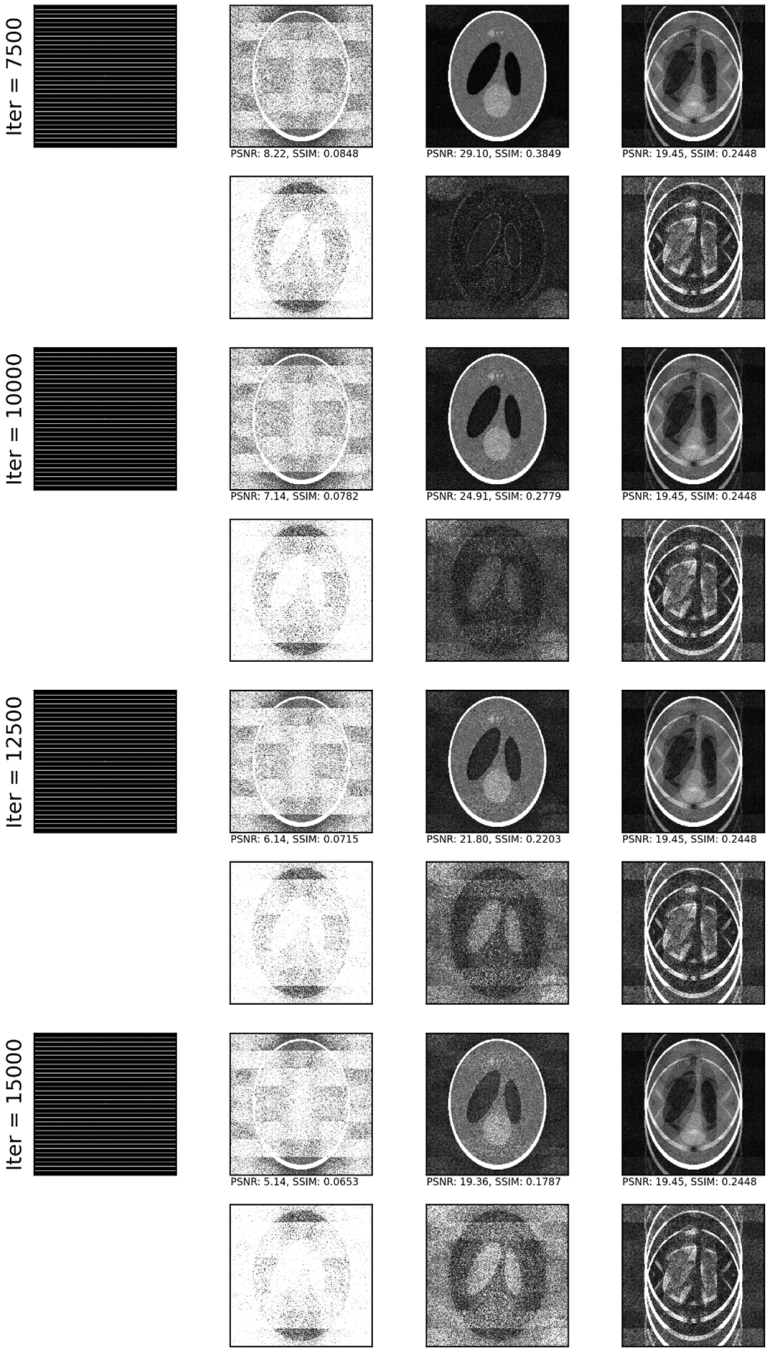

Figure 4.5: (cont'd) Phantom image reconstructions at different iterations for Deep Image Prior (2nd column), Deep Non-linear Inversion (DNLINV) (3rd column), and SENSE ($\lambda_{L2} = 0.005$) (4th column). Under each image is the absolute difference image compared to the noiseless Shepp-Logan phantom with the intensities amplified by 3x.

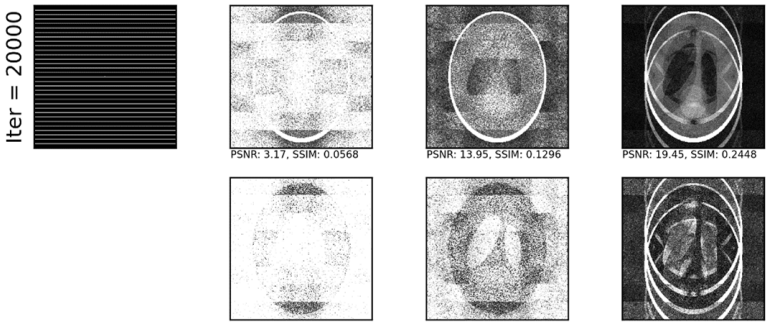

Figure 4.5: (cont'd) Phantom image reconstructions at different iterations for Deep Image Prior (2nd column), Deep Non-linear Inversion (DNLINV) (3rd column), and SENSE ($\lambda_{L2} = 0.005$) (4th column). Under each image is the absolute difference image compared to the noiseless Shepp-Logan phantom with the intensities amplified by 3x.

brated sampling patterns are shown in figs. 4.8 to 4.12. For the calibrationless and autocalibrated 2D undersampling patterns (fig. 4.6 and figs. 4.8 and 4.9), DNLINV provided similar performance to prior reconstruction algorithms at low acceleration factors. At moderate acceleration factors, classical reconstructions performed well but had increasing noise whereas DNLINV was able to reconstruct with similar quality to low acceleration factors. At high acceleration factors, classical reconstruction algorithms failed to properly reconstruct the images while DNLINV was still able to provide a good reconstruction. The aggregate results for calibrationless sampling patterns shown in fig. 4.14 demonstrate the improved performance in a dataset compared to ENLIVE and DIP.

Similar results are seen for the calibrationless 1D undersampling results in fig. 4.7 and fig. 4.10, although in this case DNLINV shows residual aliasing artifacts at moderate and high undersampling factors. For the aggregate dataset results in fig. 4.15, DNLINV had the highest PSNR but was 2nd only to the pre-trained fastMRI U-net in terms of SSIM. In the autocalibrated 1D undersampling (figs. 4.11 and 4.12), DNLINV typically out-performed ESPIRiT and ENLIVE, but the pre-trained fastMRI U-net network was the top algorithm.

For the head ENLIVE image in fig. 4.6, the highest acceleration factor tested at 8.5x was notable for DNLINV because of the seemingly large error. However, on closer inspection in fig. 4.13, the error comes from errors in coil sensitivity estimates while the image content

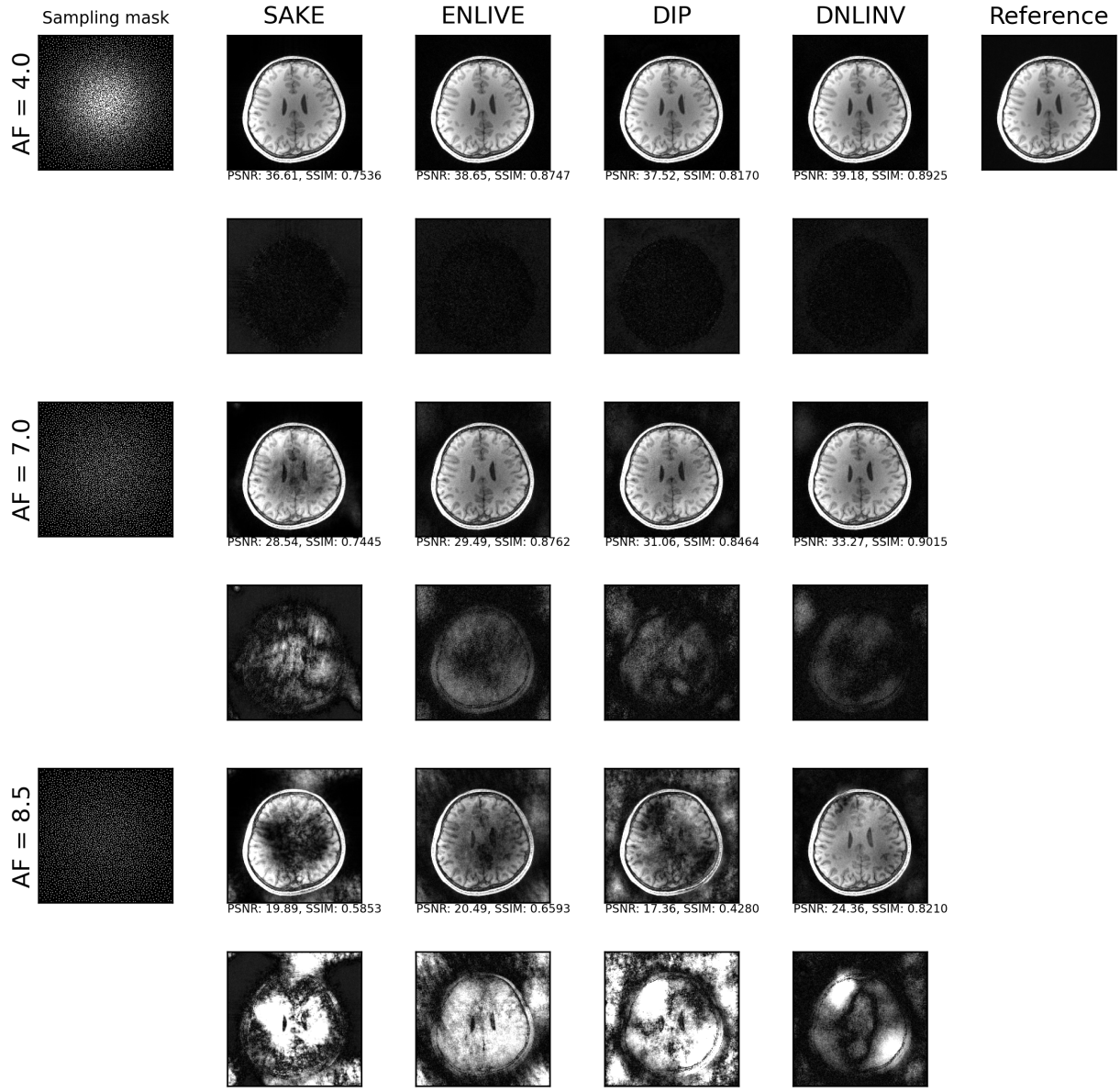

Figure 4.6: Head calibrationless poisson disk retrospective undersampling results comparing SAKE, ENLIVE, DIP, and DNLINV. The absolute difference image are shown below each reconstruction and the error intensities were amplified by 3x to improve visibility. While all methods perform qualitatively similar for the lowest AFacceleration factor (AF) = 4.0, DNLIV showed fewer artifacts and higher PSNR/SSIM at moderate to high AF.

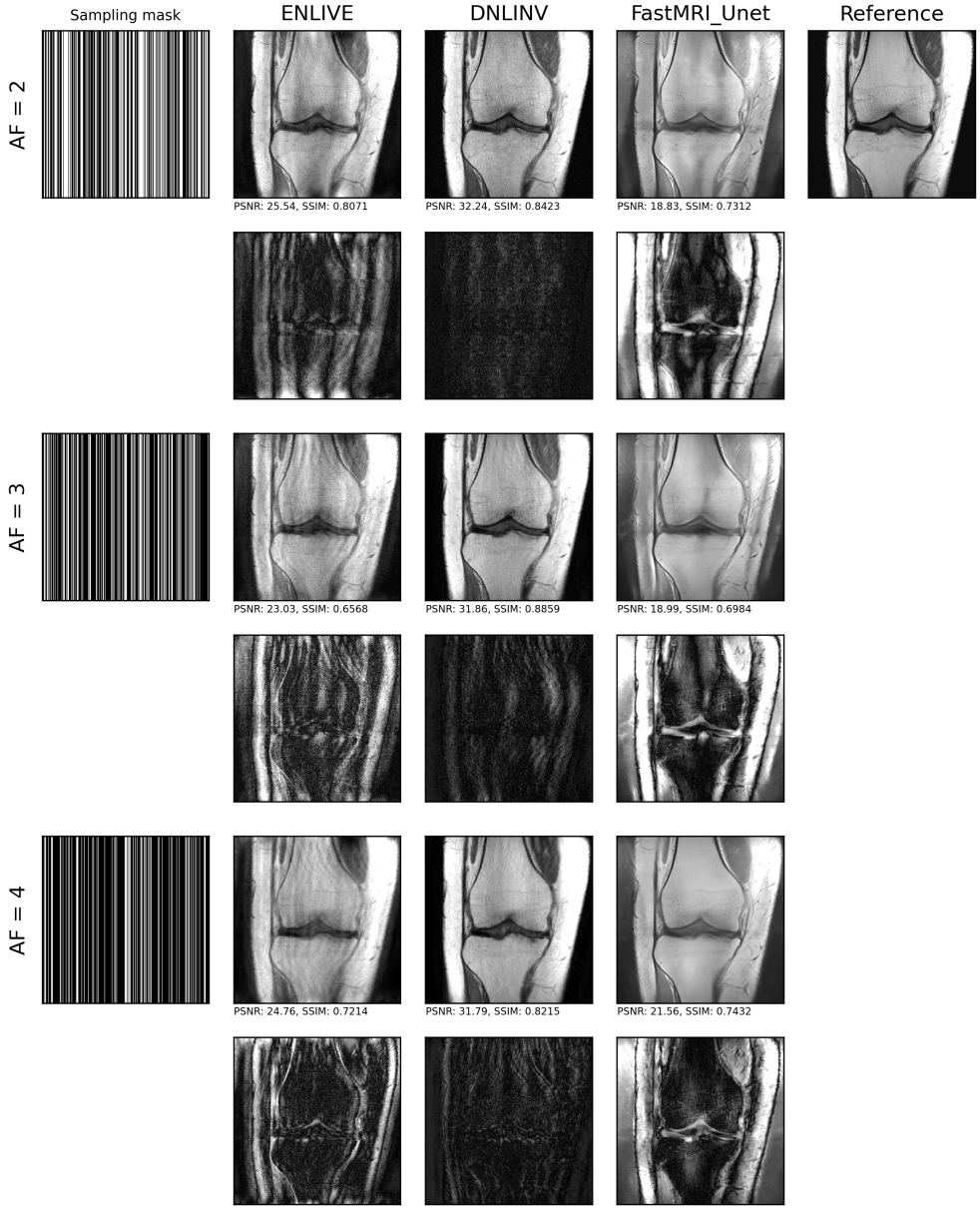

Figure 4.7: Knee 1d calibrationless retrospective undersampling comparing ENLIVE, DNLINV, and the fastMRI pre-trained Unet reconstructions. The absolute difference image are shown below each reconstruction and the error intensities were amplified by 3x to improve visibility. Even at the lowest AF = 2.0, only DNLINV is the only method largely free of parallel imaging artifacts. At higher acceleration factors. What some aliasing artifacts appear at AF = 3.0 and 4.0 in DNLINV, the error is still much smaller compared to the other methods.

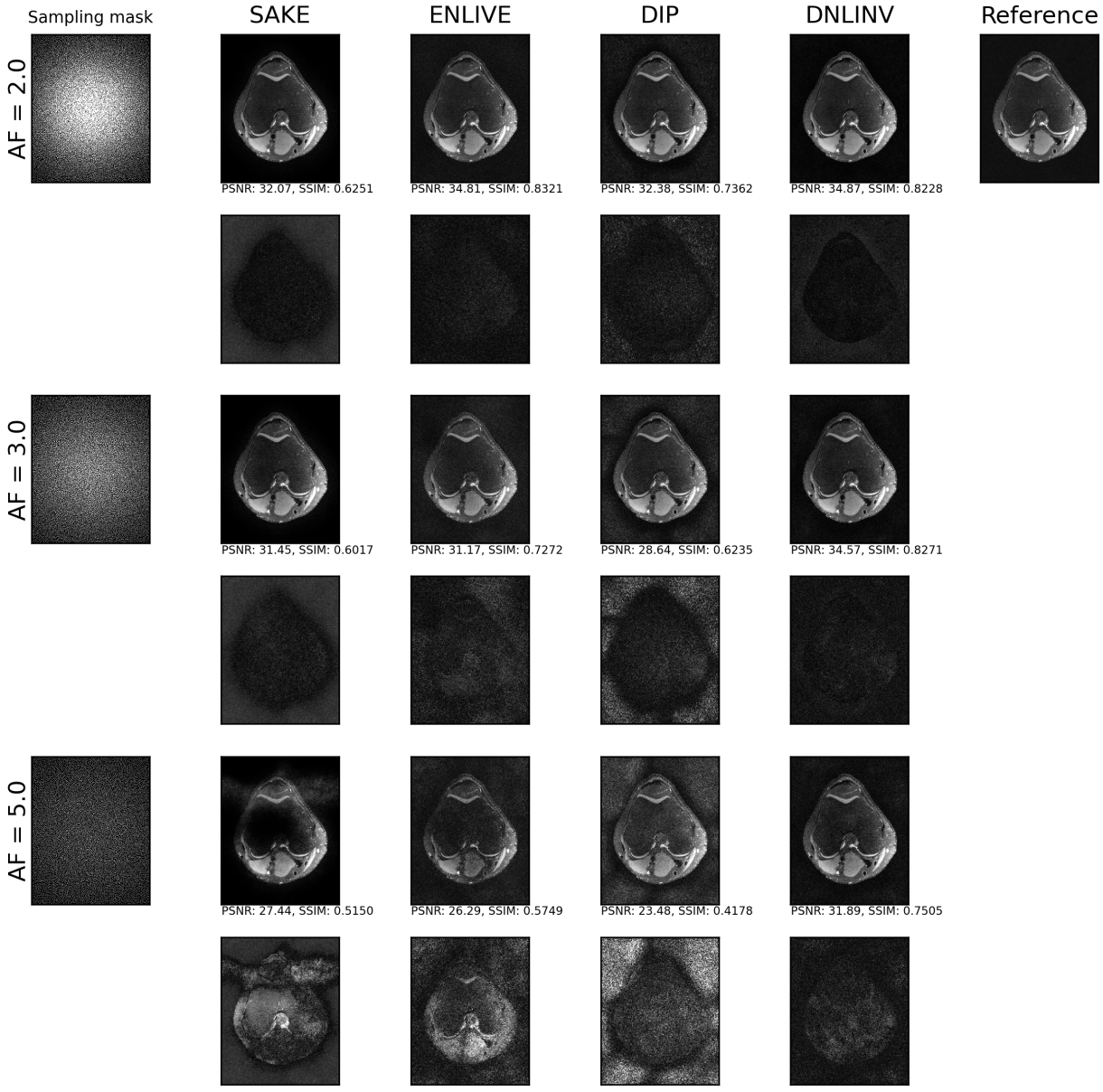

Figure 4.8: Knee calibrationless poisson disk retrospective undersampling results. The absolute difference images w.r.t. the reference image is shown under each reconstruction and the magnitudes are multiplied by 3 to better visualize the error regions.

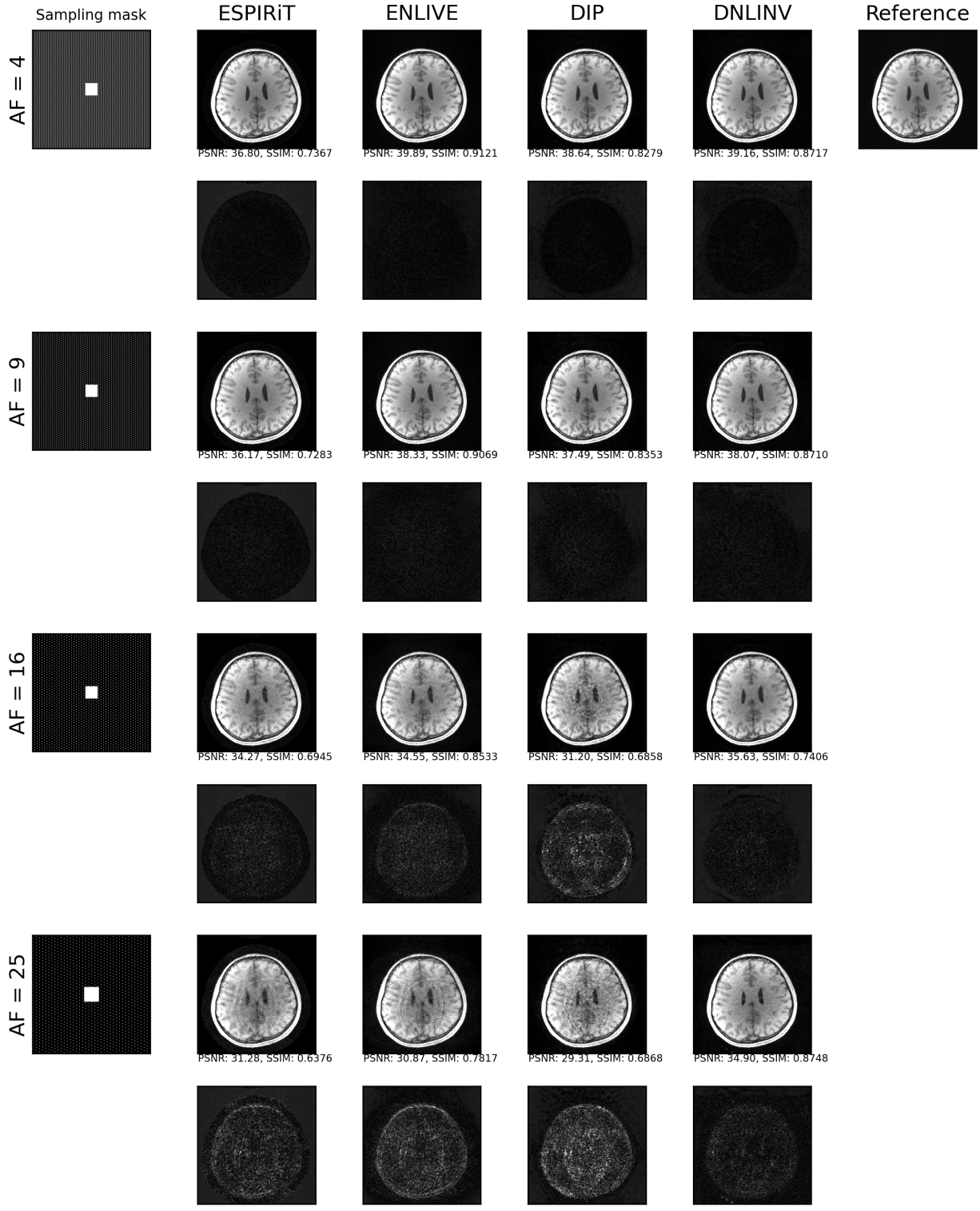

Figure 4.9: Head CAIPIRINHA retrospective undersampling results. The absolute difference images w.r.t. the reference image is shown under each reconstruction and the magnitudes are multiplied by 3 to better visualize the error regions.

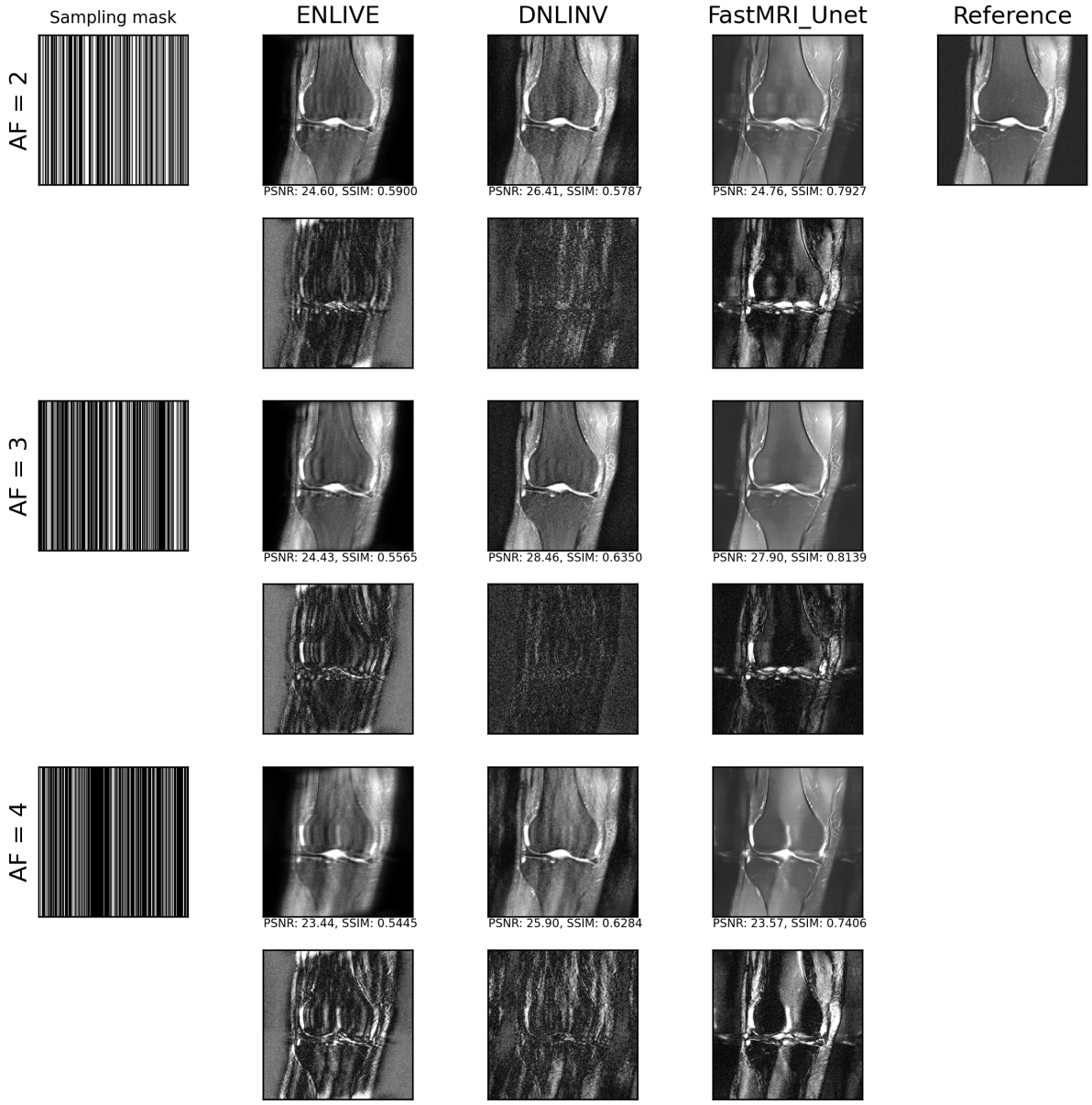

Figure 4.10: Proton-density fat-suppressed knee 1d calibrationless retrospective undersampling. The absolute difference images w.r.t. the reference image is shown under each reconstruction and the magnitudes are multiplied by 3 to better visualize the error regions.

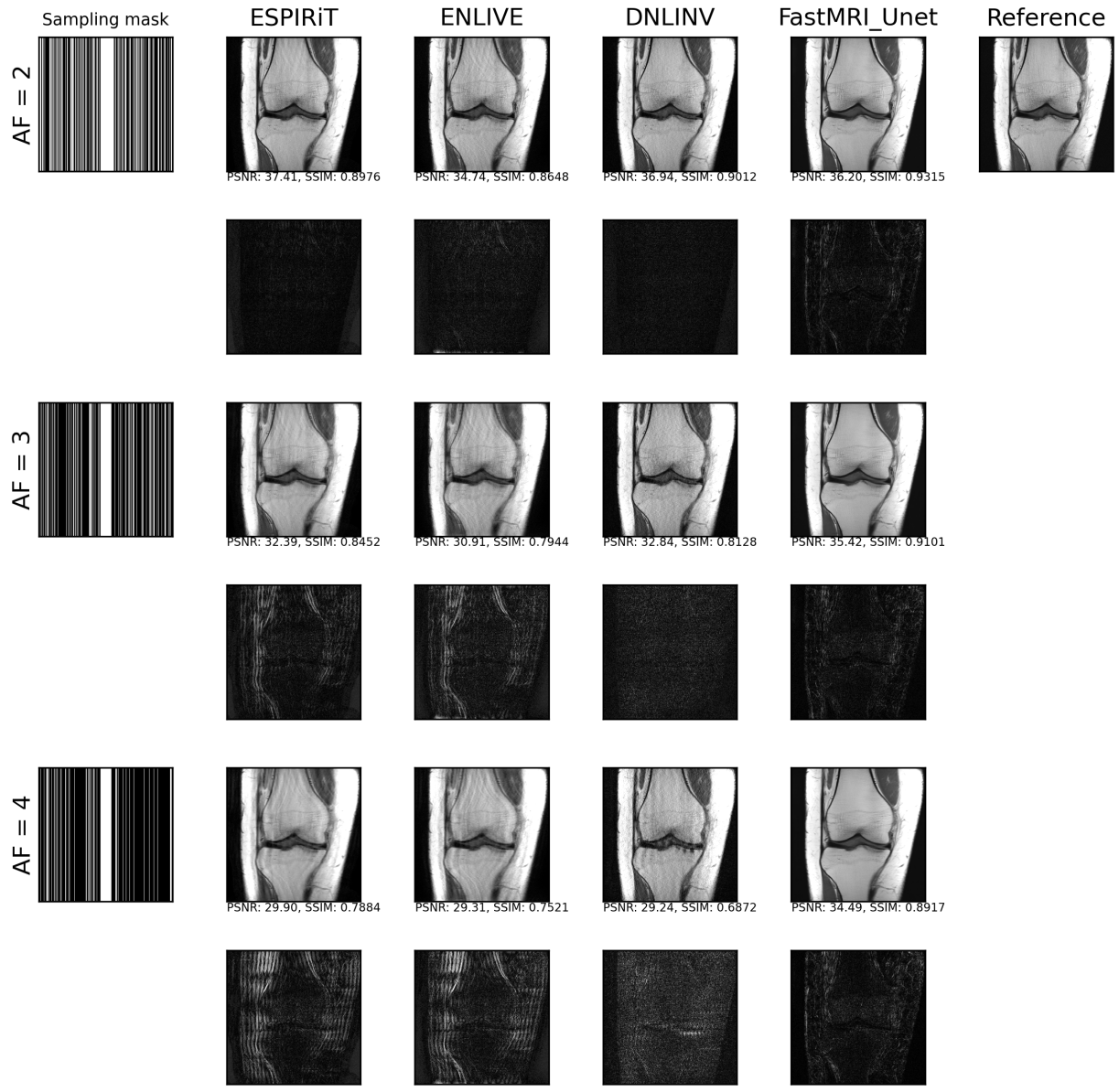

Figure 4.11: Proton-density knee 1d autocalibrated retrospective undersampling. The absolute difference images w.r.t. the reference image is shown under each reconstruction and the magnitudes are multiplied by 3 to better visualize the error regions.

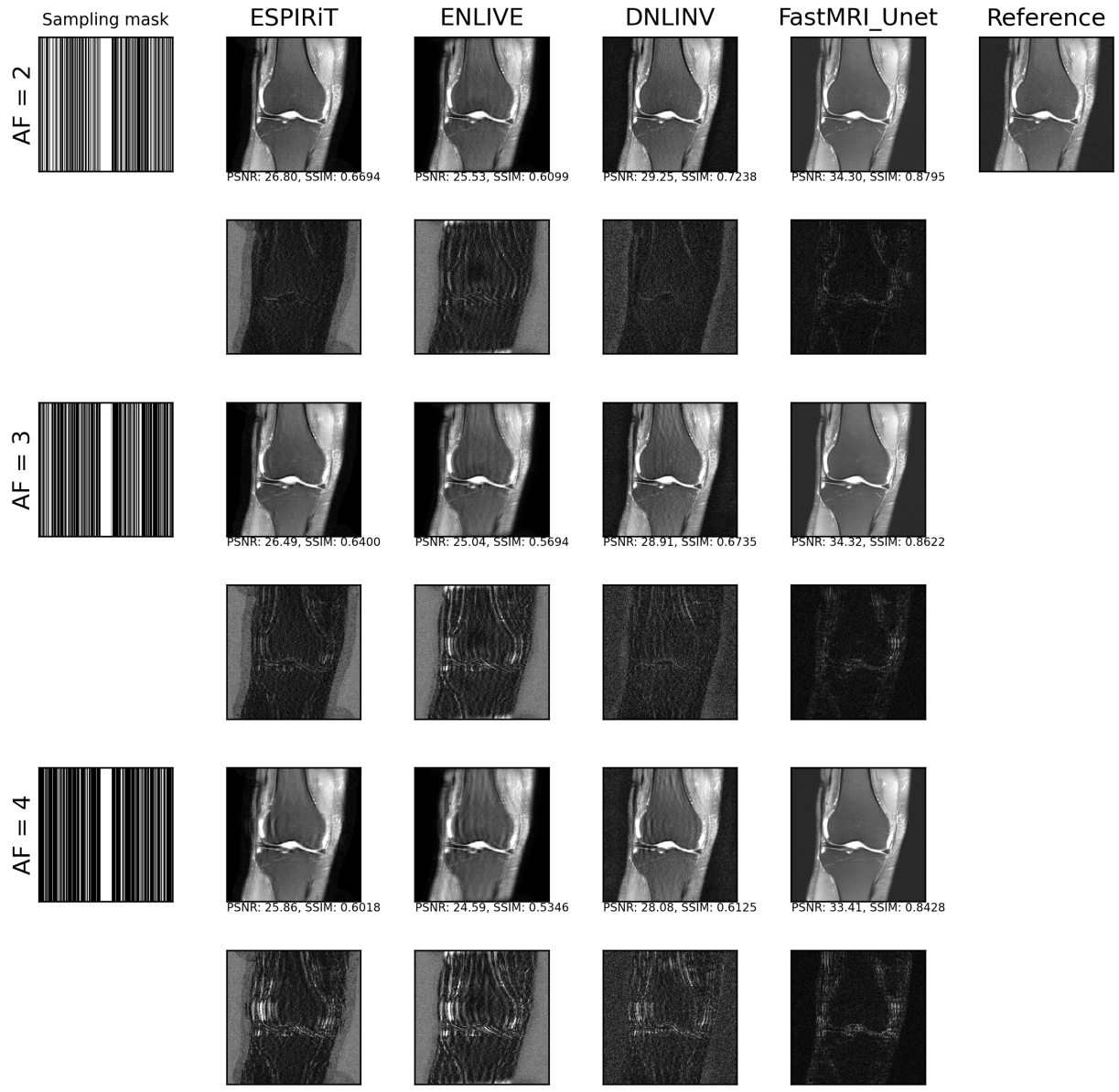

Figure 4.12: Proton-density fat-suppressed knee 1d autocalibrated retrospective undersampling. The absolute difference images w.r.t. the reference image is shown under each reconstruction and the magnitudes are multiplied by 3 to better visualize the error regions.

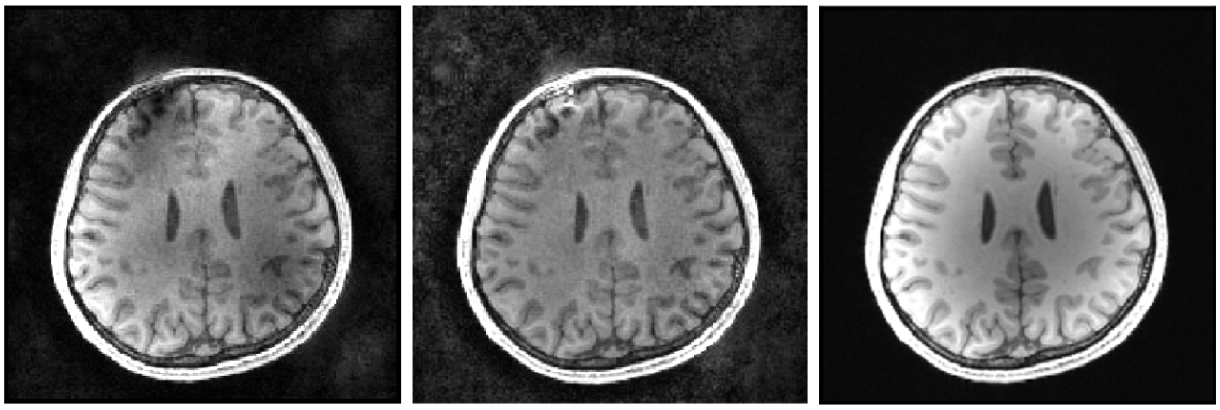

Figure 4.13: DNLINV sum-of-squares image reconstruction (left), DNLINV SENSE image reconstruction (middle), and fully-sampled sum-of-squares image (right)

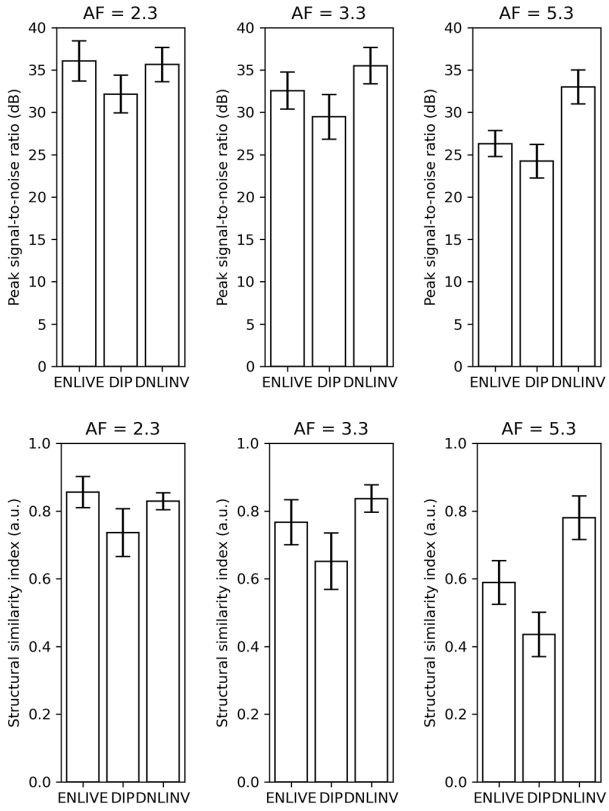

Figure 4.14: Aggregate reconstruction results for a calibrationless sampling pattern using the Stanford 3d knee FSE dataset from mridata.org (n=20). Error bars correspond to 1 standard deviation of the metrics.

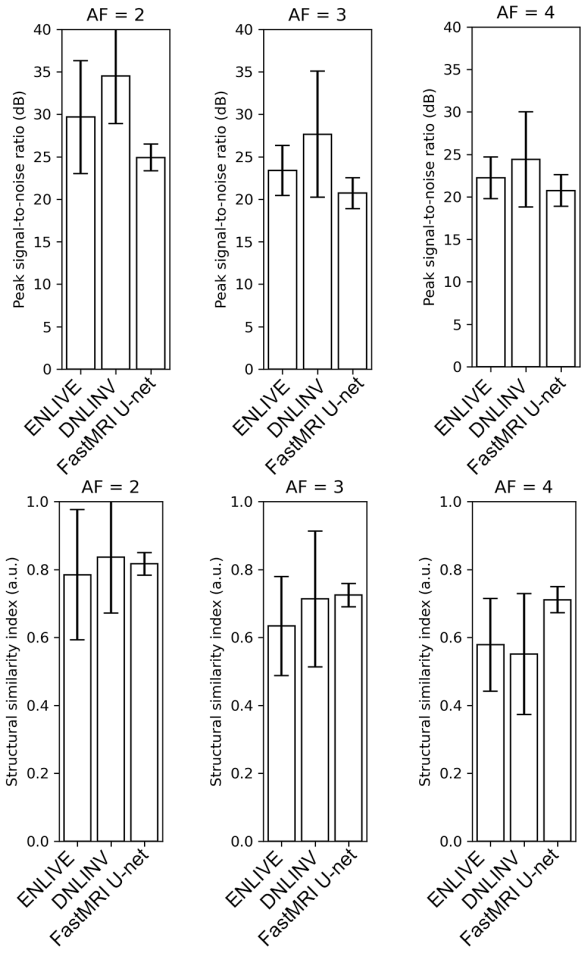

Figure 4.15: Aggregate reconstruction results for a calibrationless sampling pattern from the validation set of the FastMRI 3T proton-density knee dataset (n=10). Error bars correspond to 1 standard deviation of the metrics.

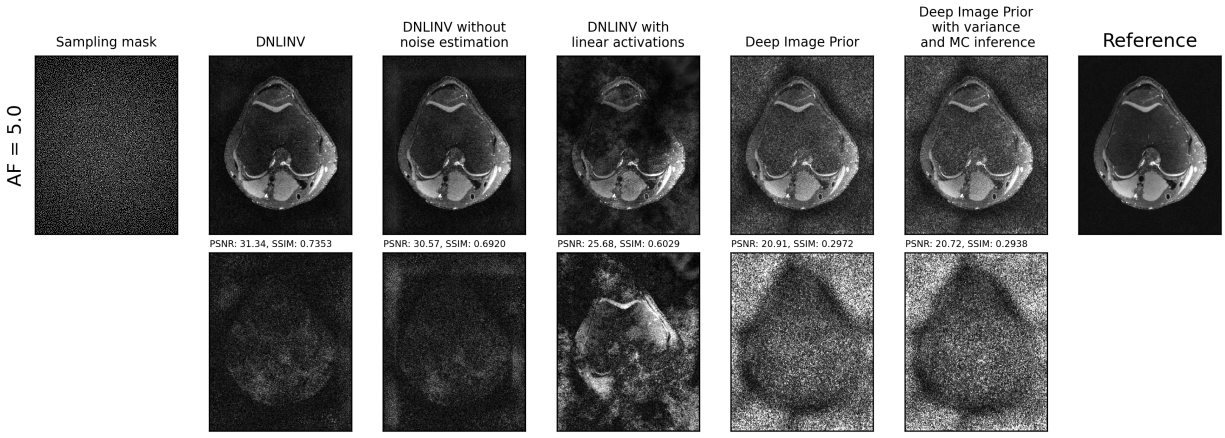

Figure 4.16: Ablation experiment in a knee. The absolute difference image compared to the reference are shown below each reconstruction and the error intensities were amplified by 3x to improve visibility.

was in fact sufficiently recovered with minimal artifacts.

fig. 4.16 shows the ablation experiment, illustrating the value of each component of the DNLINV method. While DIP shows no obvious artifacts, it suffers from clear noise amplification. As in the 1D denoising example, adding MC inference slightly degrades performance and is no better than standard DIP. DNLINV suffers from artifacts or noise amplification without ReLU non-linear activations or noise estimation, while adding these methods achieves the best performance.

4.5 Discussion

In this paper, we introduced deep non-linear inversion (DNLINV). DNLINV is more robust to noise than existing methods and thus can reconstruct data at higher acceleration factors than previously possible, especially for calibrationless parallel imaging and compressed sensing. Calibrationless parallel imaging and compressed sensing is especially difficult at higher acceleration factors due to not only undersampling of the image content, but also due to the undersampling of the low frequency k-space locations that are crucial for accurate coil sensitivity estimation. We demonstrated that noise deteriorates the quality of coil sensitivity

estimates and thus leads to failure in image reconstruction. With the inclusion of probabilistic Bayesian inference in the reconstruction, we introduce noise robustness that allow for successful coil sensitivity estimation and image reconstruction at higher acceleration factors than were previously possible.

To the best of our knowledge, there is no other scan-specific nor self-supervised deep learning method that incorporates Bayesian deep learning in reconstruction. Thus, DNLINV is the first of its kind to utilize Bayesian inference to minimize overfitting. Other methods have utilized splitting the dataset into disjoint sets to separately learn the regularizer and the data fitting [124], use the autocalibation scan region for learning k-space interpolation [127], [128], [130], or rely on an architecture search to find an optimal network architecture [129].

The major component that makes DNLINV successful is the use of Bayesian evidence maximization [48], [52], using the ELBO cost function. The evidence maximization framework is what allows the network complexity to be better controlled. In deep image prior, only L_p regularization is available to control the network complexity and since the network is overparameterized, the network still overfits as has been observed by the original authors and other works as well [44], [132]. To the best of our knowledge, there is no other work that exists for deep self-supervised MR image reconstruction that utilizes Bayesian evidence maximization. We demonstrated the impact that this can on image reconstruction.

The next major component is the use of ReLU non-linear activations in the network. With the use of only linear activations, the network behaves like a convolutional coding model [138]. We demonstrated that using only linear layers lead to failure in image reconstruction (fig. 4.16), showing the necessity of the non-linearities. The use of ReLU activations make the network like a multi-layer semi-non-negative convolutional coding model [139].

The scan-specific and self-supervised nature of our approach allows for it to perform much like classical reconstruction algorithms and can be used for any hardware configuration, contrast, pulse sequence, or sampling pattern without any re-training. Supervised

reconstruction algorithms have difficult handling test cases that are far from the training distribution. In the test example using the fastMRI data under the calibrationless context (fig. 4.7 and fig. 4.10), the pre-trained U-net produced significant artifacts due to the lack of the autocalibration region that was used for training data.

By eliminating the need of the autocalibration region, calibrationless parallel imaging and compressed sensing acquisitions achieve the highest possible sub-sampling factors. Thus far, only deep learning methods trained on a database have been demonstrated to reconstruct calibrationless parallel imaging and compressed sensing data [140], [141]. The advantage of our scan-specific approach is that it can be used even in the most experimental of hardware configurations and pulse sequences.

DNLINV is able to compete with supervised approaches but only up to a moderate acceleration factors. Supervised approaches have the benefit of learning from a database in specific circumstances whereas DNLINV does not.

4.5.1 Limitations

DNLINV still tends to overfit to noise over time. This is due to the variational approximations used. This mode-seeking and overfitting behavior is a well-known characteristic of variational Bayesian inference [52]. Despite this, iterative methods such as ESPIRiT and ENLIVE have similar behavior and the number of iterations is chosen as a hyperparameter in reconstruction. Using more sophisticated and computationally intensive inference methods such as through Langevin dynamics [132], expectation propagation [142], or ensemble methods [143] may be used to ultimately conquer the overfitting aspects.

There are several other limitations to DNLINV. The Bayesian inference used has heavy compute requirements due to the Monte Carlo approximation of the expectations and integrals and the high noise of SGD. The reconstructions performed take 15 minutes to 2 hours on GPU depending on number of iterations required for convergence. Possible ways to accelerate this process is by pre-training on a database to initialize the network to a good

configuration and thus require less iterations on reconstruction (i.e., transfer learning) [129]. Alternatively, meta-learning approaches may be used to learn an optimizer that also reduces the number of gradient steps needed (i.e., learning-to-learn [144]). Finally, the use of deep convolutional neural networks has a large memory requirement. Possible ways to save on memory is with the use of more compact neural network architectures (e.g., DCGAN [145]) or by reducing the overall size of the network using an architecture search [129].

4.A Derivation of cost function

Using Bayes theorem, we can write the inversion problem generally as follows:

$$p(z, \theta|y) = \frac{p(y|z, \theta)p(z, \theta)}{p(y)} \tag{4.34}$$

With the generative model defined as

$$y = Ag_C(z; \theta)g_x(z; \theta) + n \tag{4.35}$$

And the additive noise n is defined as

$$n \sim \frac{1}{\sqrt{(2\pi)^{N_c} \det |\Sigma_c|}} \exp\left(-\frac{1}{2} n^T \Sigma_c^{-1} n\right) \tag{4.36}$$

where y is the observed k-space data, N_c is the number of coil channels, z is the input latent code to the network, θ are the network parameters, Σ_c is the coil covariance matrix, A is the forward operator, g_C is the coil sensitivity map outputs of the network, and g_x is the image output of the network. In the case of linearly-encoded MRI:

$$A = SF \tag{4.37}$$

where F is the Fourier transform operator and S is the k-space sampling operator.

The Bayesian evidence maximization framework in Chapter 3.4 of [48] poses the following

optimization problem:

$$\max_{M} p(y|M) \tag{4.38}$$

where M is a model for the data. This optimization problem searches the space of all possible models and attempts to find the model that explains the data the best. In the context of DNLINV, the evidence maximization function becomes:

$$\max_{p(z,\theta|y)} p(y) = \max_{p(z,\theta|y)} E_{p(z,\theta|y)}[\log(\frac{p(y|z,\theta)p(z,\theta)}{p(z,\theta|y)})] \tag{4.39}$$

since the model M is fixed to be a deep convolutional neural network. In effect, this cost function solves for the posterior distribution of the network parameters and latent code that makes the observation the most likely.

Derivation:

$$\max_{z,\theta} p(y) = \max_{z,\theta} \log p(y) \tag{4.40}$$

$$\log p(y) = \log p(y) \iint p(z,\theta|y)dzd\theta \tag{4.41}$$

(Multiplication by 1 since the integral results to 1)

$$= \iint \log p(y) p(z,\theta|y)dzd\theta \tag{4.42}$$

($\log p(y)$ can be placed inside the integral since y is not the integration variable)

$$= \iint \log(\frac{p(y|z,\theta)p(z,\theta)}{p(z,\theta|y)}) p(z,\theta|y)dzd\theta \tag{4.43}$$

(by expansion of $p(y) = \frac{p(y|z,\theta)p(z,\theta)}{p(z,\theta|y)}$)

$$= E_{p(z,\theta|y)}[\log(\frac{p(y|z,\theta)p(z,\theta)}{p(z,\theta|y)})] \tag{4.44}$$

(using the definition of the expectation)

The problem now shifts to solving for the posterior distribution $p(z,\theta|y)$. Given the non-linear nature of the generative model, solving for the posterior is intractable and thus, a variational approximation must be used (Chapter 10.1 in [48]). We define a factorial variational approximation (Chapter 10.1 in [48]):

$$p(z,\theta|y) \approx \ q(z,\theta|y) \approx q(z|y)q(\theta|y) \tag{4.45}$$

This considers that the approximate posterior for z and θ are independent. These two approximations cause variational inference to fit to local modes of the posterior space. Now, we utilize the Kullback-Leibler (KL) divergence (Chapter 10.1 in [48]). to formulate the cost function to fit the approximate posteriors q to the true posteriors p:

$$\min_{q(z|y),q(\theta|y)} D_{KL}(q(z|y)q(\theta|y)||p(z,\theta|y)) \tag{4.46}$$

To solve the above minimization problem, we derive the Evidence Lower Bound [52] under the model that uses approximate posteriors:

$$D_{KL}(q(z|y)q(\theta|y)||p(z,\theta|y)) = -E_{q(z|y)q(\theta|y)}[\log \frac{p(z,\theta|y)}{q(z|y)q(\theta|y)}] \tag{4.47}$$

$$= -E_{q(z|y)q(\theta|y)}[\log \frac{\frac{p(y|z,\theta)p(z,\theta)}{p(y)}}{q(z|y)q(\theta|y)}] \tag{4.48}$$

$$= -E_{q(z|y)q(\theta|y)}[\log \frac{p(y|z,\theta)p(z,\theta)}{p(y)q(z|y)q(\theta|y)}] \tag{4.49}$$

$$= E_{q(z|y)q(\theta|y)}[\log p(y) - \log \frac{p(y|z,\theta)p(z,\theta)}{q(z|y)q(\theta|y)}] \tag{4.50}$$

$$D_{KL}(q(z|y)q(\theta|y)||p(z,\theta|y)) = \log p(y) - E_{q(z|y)q(\theta|y)}[\log \frac{p(y|z,\theta)p(z,\theta)}{q(z|y)q(\theta|y)}] \tag{4.51}$$

$$D_{KL}(q(z|y)q(\theta|y)||p(z,\theta|y)) = \log p(y) + \mathcal{L} \tag{4.52}$$

Since $\log p(y)$ is a constant (because the data has already been observed, even though the form is unknown), then minimizing the KL-divergence is equivalent to maximizing $\mathcal{L}$, which is known as the "Evidence Lower Bound" or "Variational Free Energy". This name is derived by rearranging the equation above and noting that because the KL-divergence is a distance metric it must be ≥ 0, then $\log p(y) \geq \mathcal{L}$. Equation (4.52) may also be derived using Jensen's inequality (Section 2.2 and A.1 in [52]).

Our cost function is then:

$$\mathcal{L} = E_{q(z|y)q(\theta|y)}[\log \frac{p(y|z,\theta)p(z,\theta)}{q(z|y)q(\theta|y)}] \tag{4.53}$$

Expanding this:

$$
\begin{aligned}
E_{q(z|y)q(\theta|y)}[\log \frac{p(y|z,\theta)p(z,\theta)}{q(z|y)q(\theta|y)}] =& E_{q(z|y)q(\theta|y)}[\log p(y|z,\theta) + \log p(z,\theta) \\
& - \log q(z|y)q(\theta|y)] \qquad (4.54)
\end{aligned}
$$

$$
\begin{aligned}
=& E_{q(z|y)q(\theta|y)}[\log p(y|z,\theta) + \log p(z,\theta)] \\
& - E_{q(z|y)q(\theta|y)}[\log q(z|y)] - E_{q(z|y)q(\theta|y)}[\log q(\theta|y)] \qquad (4.55)
\end{aligned}
$$

$$
\begin{aligned}
=& E_{q(z|y)q(\theta|y)}[\log p(y|z,\theta) + \log p(z,\theta)] \\
& - \iint q(z|y)q(\theta|y) \log q(z|y) dz d\theta \\
& - \iint q(z|y)q(\theta|y) \log q(\theta|y) dz d\theta \qquad (4.56)
\end{aligned}
$$

$$
\begin{aligned}
=& E_{q(z|y)q(\theta|y)}[\log p(y|z,\theta) + \log p(z,\theta)] \\
& - \int \left(\int q(\theta|y) d\theta \right) q(z|y) \log q(z|y) dz \\
& - \int \left(\int q(z|y) dz \right) q(\theta|y) \log q(\theta|y) d\theta \qquad (4.57)
\end{aligned}
$$

(by Fubini's theorem)

$$
\begin{aligned}
=& E_{q(z|y)q(\theta|y)}[\log p(y|z,\theta) + \log p(z,\theta)] \\
& - \int q(z|y) \log q(z|y) dz - \int q(\theta|y) \log q(\theta|y) d\theta \qquad (4.58)
\end{aligned}
$$

$$
\begin{aligned}
=& E_{q(z|y)q(\theta|y)}[\log p(y|z,\theta) + \log p(z,\theta)] \\
& + H[q(z|y)] + H[q(\theta|y)] \qquad (4.59)
\end{aligned}
$$

Assuming that $p(z,\theta) = p(z)p(\theta)$, we then arrive at the general form of the cost function for DNLINV:

$$
\mathcal{L} = E_{q(z|y)q(\theta|y)} \left[\log p\left(y|z,\theta\right) + \log p\left(z\right) + \log p(\theta) \right] + H\left[q\left(z|y\right)\right] + H\left[q\left(\theta|y\right)\right] \qquad (4.60)
$$

4.B Numerically solving the inference problem

Because the expectation $E_{q(z|y)q(\theta|y)}$ is intractable, we solve this using stochastic optimization [52]. Stochastic optimization approximates the expectation with Monte-Carlo samples by drawing from the approximate posteriors q. The cost function then becomes:

$$\mathcal{L} = \frac{1}{M}\sum_{m=1}^{M} \log p\left(y\middle|z^{(m)}, \theta^{(m)}\right) + \log p\left(z^{(m)}\right) + \log p\left(\theta^{(m)}\right) + H\left[q\left(z|y\right)\right] + H\left[q\left(\theta|y\right)\right] \quad (4.61)$$

where M is the number of Monte Carlo samples, $z^{(m)} \sim q(z)$ and $\theta^{(m)} \sim q(\theta)$ are sampled for each Monte Carlo sample.

We define the following priors:

$$p\left(z_j\right) = \text{Normal}\left(0, 1\right) \quad (4.62)$$

$$p\left(\theta_s\right) = \text{Normal}\left(0, \lambda\right) \quad (4.63)$$

$$p\left(z\right) = \prod_{j=1}^{J} p\left(z_j\right) \quad (4.64)$$

$$p\left(\theta\right) = \prod_{s=1}^{S} p\left(\theta_s\right) \quad (4.65)$$

where J is the number of latent variables, λ is the precision for the parameter θ_s, and S is the total number of network parameters.

Monte-Carlo dropout (MC-Dropout) is an approach that converts a deep convolutional neural network into a Bayesian deep convolutional neural network [57]. Using MC-Dropout in the deep convolutional neural network gives us $q(\theta|y)$ by training the network with dropout and sampling the different dropout masks during inference.

With $q(\theta)$ taken care of using MC-Dropout, we now solve for $q(z)$. Vanilla stochastic variational inference has high variance estimates of the ELBO, thus we utilize the "reparameterization trick" [52], [134], to reduce the variance of the gradient estimates. The reparameterization trick is performed as follows:

First, we define a factorized approximation for $q(z)$:

$$q\left(z_j\right) = \text{Normal}\left(\mu_{z_j}, \sigma_{z_j}\right) \tag{4.66}$$

$$q\left(z\right) = \prod_{j=1}^{J} q(z_j) \tag{4.67}$$

where J is the number of latent variables. By defining $q(z_j)$ as a Gaussian distribution, we may reparametrize z_j as follows:

$$z_j = \mu_{z_j} + \sigma_{z_j}\epsilon_j \tag{4.68}$$

where

$$\epsilon_j^{(m)} \sim \text{Normal}\left(0,\ 1\right) \tag{4.69}$$

Thus, ϵ_j is sampled and a sample of z_j is constructed using eq. (4.68):

$$z_j^{(m)} = \mu_{z_j} + \sigma_{z_j}\epsilon_j^{(m)} \tag{4.70}$$

The reparametrization trick removes dependencies on an unknown probability distribution and exposes the parameters of the approximate distribution, μ_j and σ_{z_j}, to the back-propagation computational graph, which can then be solved for as a typical parameter in deep learning.

Plugging in the specifics that are known about the problem (eqs. (4.35), (4.36), (4.66) and (4.67)) to the general form of the cost function (eq. (4.60)), the cost function becomes:

$$\mathcal{L} = -\frac{N}{2}(\log 2\pi + \text{logdet}(\Sigma_c)) - \frac{1}{2M}\sum_{m=1}^{M}\sum_{k=1}^{K}\left(y_k - S_k F g_C\left(z^{(m)};\theta^{(m)}\right) g_x\left(z^{(m)};\theta^{(m)}\right)\right)^T$$
$$\Sigma_c^{-1}\left(y_k - S_k F g_C\left(z^{(m)};\theta^{(m)}\right) g_x\left(z^{(m)};\theta^{(m)}\ \right)\right)$$
$$- 0.5\sum_{j=1}^{J}\sum_{m=1}^{M}\left(z_j^{(m)}\right)^2 + 0.5\sum_{j=1}^{J}\log\left(\sigma_{z_j}^2\right) + 0.5 \tag{4.71}$$

where y is the observed k-space data, $z^{(m)}$ is m-th sample of the input latent code to the network, $\theta^{(m)}$ is the m-th sample of the network parameters, M is the number of Monte Carlo samples, Σ_c is the coil covariance matrix, g_C is the coil sensitivity map outputs of the network, g_x is the image output of the network, σ_{z_j} is the standard deviation of the approximate posterior $q(z_j)$, J is the number of latent variables, K is the number of k-space sampling points, and N is the total number of k-space samples across all coils.

Since both approaches for solving are based on sampling and stochastic gradient descent, we may simply proceed with the optimization procedure the same as standard deep learning, with the addition of sampling $z^{(m)} \sim q(z|y)$ and dropout masks at each iteration.

To improve estimates of the ELBO, multiple Monte Carlo samples may be drawn at each optimization step. Additionally, we utilize decoupled weight decay [135] to induce L_2-regularization on the network parameters and does not appear in the cost function.

Chapter 5

Summary and Future Directions

This dissertation was focused on the applications of Bayesian deep learning to problems in medical image reconstruction. A rule of thumb in deep learning is you need hundreds of thousands of images that cover as many cases as possible to properly deploy it in practice. This is typically not the case in medical image reconstruction due to the many possible variations in hardware, acquisition schemes, and disease conditions. Furthermore, training data may be imperfect to due many factors such as registration errors or limits in possible signal-to-noise ratio (SNR). Thus, I utilize Bayesian deep learning to address the challenges in *imperfect data* and *small data* regimes.

In the case of imperfect data, the application domain was on attenuation correction for simultaneous PET/MRI. Pairs of MRI and CT images were used to train a deep convolutional neural network so that pseudo-CT images may be generated from the MRI. However, there were two major challenges in this application. First was the challenge of co-registration and anatomical correspondence between the MRI and CT. Since the images were acquired at different times, the patient body position are different and deformable registration techniques are employed. However, there are still errors that remain due to natural changes inside the body: such as those due to the movement of bowel air, or changes in tissue composition over time. Additionally, the presence of metal implants pose a challenge since metal implants

produce different imaging artifacts on MRI than CT. These are intrinsic dataset errors that cannot be eliminated. Bayesian deep learning was used to quantify the uncertainty to due these intrinsic dataset errors. Once the uncertainty was localized and quantified, additional information from the PET emission data was then used to correct for these regions of error and allow for more accurate attenuation correction in PET/MRI.

In the case of small data, the application domain was on accelerated MRI reconstruction that relied on below Nyquist sub-sampling to achieve the acceleration. Deep image reconstruction methods require large datasets with fixed hardware configurations and acquisition schemes. However, as soon as there are changes in the hardware or the acquisition scheme, a new dataset is required and the models need to be trained to account for these changes. Furthermore, there are instances where it is impractical or impossible for have a large high-quality database of training examples. Thus I developed DNLINV, a *scan-specific* method that is applicable to any acquisition scheme or hardware configuration that uses only the measurement data of a single scan itself, much like classical image reconstruction methods. The use of Bayesian deep learning allows us to incorporate the full statistical model into the reconstruction and quantify the uncertainty surrounding the solution space. This allows DNLINV to overcome the overfitting behavior that is present in highly accelerated MRI and has pushed the possible acceleration factors another step further.

These are two specific contributions that Bayesian deep learning can provide however there are still limitations to this work and the general approach. Addressing these would likely be the immediate steps for future research.

First, this was a first exploration of the feasibility of using Bayesian deep learning in these specific applications. Thus, there were only a (relatively) small number of cases explored. There needs to be more validation of the uncertainty metrics on a larger variety of cases to establish baselines and benchmarks [146]. In particular, characterizing false positive and false negative behavior of the uncertainty maps is important to better understand the behavior of the uncertainty quantification.

Second, Bayesian deep learning is computationally intensive. This is due to the requirement of recognizing the space of possible solutions and taking these into account when solving the problems. This addition requires more compute than standard deep learning methods. Utilizing new developments in theory or more efficient algorithms [52], [147]–[151] would be important to make Bayesian deep learning more accessible in practice.

Third, utilizing the mean-field variational approximation to achieve Bayesian deep learning still leads to overfitting in certain situations. Modern advances that allow for approximate Bayesian inference beyond the mean-field variational approximation would largely resolve any remaining issues surrounding the overfitting behavior [52], [152]–[154].

These are the most pressing limitations to develop the methodology further. Other possible lines of inquiry are then more related to how Bayesian deep learning and uncertainty quantification is used in a larger workflow.

The current uncertainty estimates may not necessarily be semantically meaningful. Other MR image reconstruction research groups have started exploring uncertainty quantification in MRI reconstruction [155] however there have been no examples of the semantic behavior of uncertainty quantification on disease findings. Incorporation of a method such as Counterfactual Latent Uncertainty Explanations (CLUE) [156] would be valuable for interpretability.

Finally, the uncertainty measures itself may be utilized in a larger clinical workflow. I've only demonstrated two possible use-cases of the uncertainty metrics. Once the uncertainty has been provided, how exactly can it be incorporated into the clinical workflow aside just its algorithmic applications to improve performance [157]? Can it be a human-readable heat map? How can it be used to guide image interpretation and decision making? These are broader questions that medical imaging scientists would be better posed to answer than computer scientists who are more focused on the algorithmic details [158].

In summary, Bayesian deep learning allows for the quantification of uncertainty that recognizes the space of possible solutions and is specifically designed to make a user aware of the limitations of the training data, unlike standard deep learning approaches. This

feature has allowed for advances in PET/MRI attenuation correction and accelerated MRI reconstruction.

References

[1] J. A. Fessler, "Medical image reconstruction: A brief overview of past milestones and future directions," *arXiv:1707.05927 [physics]*, Jul. 2017. arXiv: 1707.05927 [physics].

[2] M. Defrise, P. E. Kinahan, and C. J. Michel, "Image Reconstruction Algorithms in PET," in *Positron Emission Tomography*, D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, Eds., London: Springer-Verlag, 2005, pp. 63–91, ISBN: 978-1-85233-798-8. DOI: 10.1007/1-84628-007-9_4.

[3] M. S. Hansen and P. Kellman, "Image reconstruction: An overview for clinicians," *Journal of Magnetic Resonance Imaging*, vol. 41, no. 3, pp. 573–585, 2015, ISSN: 1522-2586. DOI: 10.1002/jmri.24687.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[5] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image Reconstruction is a New Frontier of Machine Learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1289–1296, Jun. 2018, ISSN: 1558-254X. DOI: 10.1109/TMI.2018.2833635.

[6] C. Catana, "Attenuation correction for human PET/MRI studies," *Physics in Medicine & Biology*, vol. 65, no. 23, 23TR02, Nov. 2020, ISSN: 0031-9155. DOI: 10.1088/1361-6560/abb0f8.

[7] P. E. Kinahan, B. H. Hasegawa, and T. Beyer, "X-ray-based attenuation correction for positron emission tomography/computed tomography scanners," *Seminars in Nuclear*

Medicine, vol. 33, no. 3, pp. 166–179, Jul. 2003, ISSN: 0001-2998. DOI: 10.1053/snuc.2003.127307.

[8] A. P. Leynes and P. E. Z. Larson, "Synthetic CT Generation Using MRI with Deep Learning: How Does the Selection of Input Images Affect the Resulting Synthetic CT?" In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 6692–6696. DOI: 10.1109/ICASSP.2018.8462419.

[9] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999, ISSN: 1522-2594. DOI: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S.

[10] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002, ISSN: 1522-2594. DOI: 10.1002/mrm.10171.

[11] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007, ISSN: 1522-2594. DOI: 10.1002/mrm.21391.

[12] F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akcakaya, "Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 128–140, Jan. 2020, ISSN: 1558-0792. DOI: 10.1109/MSP.2019.2950640.

[13] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui, "fastMRI: An Open Dataset

and Benchmarks for Accelerated MRI," *arXiv:1811.08839 [physics, stat]*, Nov. 2018. arXiv: 1811.08839 [physics, stat].

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *CVPR09*, 2009.

[15] R. M. Neal, *Bayesian Learning for Neural Networks*, ser. Lecture Notes in Statistics 118. New York: Springer, 1996, ISBN: 978-0-387-94724-2.

[16] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[17] A. G. Wilson, "The Case for Bayesian Deep Learning," *arXiv:2001.10995 [cs, stat]*, Jan. 2020. arXiv: 2001.10995 [cs, stat].

[18] L. V. Jospin, W. Buntine, F. Boussaid, H. Laga, and M. Bennamoun, "Hands-on Bayesian Neural Networks – a Tutorial for Deep Learning Users," *arXiv:2007.06823 [cs, stat]*, Jan. 2022. arXiv: 2007.06823 [cs, stat].

[19] W. G. Bradley, "History of Medical Imaging," *Proceedings of the American Philosophical Society*, vol. 152, no. 3, pp. 349–361, 2008, ISSN: 0003-049X.

[20] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*. Elsevier, 2004.

[21] D. G. Nishimura, *Principles of Magnetic Resonance Imaging*. Standford Univ., 2010.

[22] M. N. Wernick and J. N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT*. Elsevier, 2004.

[23] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, *Physics in Nuclear Medicine*. Elsevier Health Sciences, 2012.

[24] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999, ISSN: 1522-2594. DOI: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S.

[25] A. Deshmane, V. Gulani, M. A. Griswold, and N. Seiberlich, "Parallel MR imaging," *Journal of Magnetic Resonance Imaging*, vol. 36, no. 1, pp. 55–72, 2012, ISSN: 1522-2586. DOI: 10.1002/jmri.23639.

[26] J. Ollinger and J. Fessler, "Positron-emission tomography," *IEEE Signal Processing Magazine*, vol. 14, no. 1, pp. 43–55, 1997. DOI: 10.1109/79.560323.

[27] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007, ISSN: 1522-2594. DOI: 10.1002/mrm.21391.

[28] T. A. Gallagher, A. J. Nemeth, and L. Hacein-Bey, "An Introduction to the Fourier Transform: Relationship to MRI," *American Journal of Roentgenology*, vol. 190, no. 5, pp. 1396–1405, May 2008, ISSN: 0361-803X, 1546-3141. DOI: 10.2214/AJR.07.2874.

[29] E. G. Larsson, D. Erdogmus, R. Yan, J. C. Principe, and J. R. Fitzsimmons, "SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging," *Journal of Magnetic Resonance*, vol. 163, no. 1, pp. 121–123, Jul. 2003, ISSN: 10907807. DOI: 10.1016/S1090-7807(03)00132-0.

[30] S. Boyd, S. P. Boyd, and L. Vandenberghe, "Least-squares and linear programming," in *Convex Optimization*, Cambridge university press, 2004.

[31] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magnetic Resonance in Medicine*, vol. 46, no. 4, pp. 638–651, 2001, ISSN: 1522-2594. DOI: 10.1002/mrm.1241.

[32] 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.

[33] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002, ISSN: 1522-2594. DOI: 10.1002/mrm.10171.

[34] M. Lustig and J. M. Pauly, "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magnetic Resonance in Medicine*, n/a–n/a, 2010, ISSN: 07403194, 15222594. DOI: 10.1002/mrm.22428.

[35] J. P. Haldar and J. Zhuo, "P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data," *Magnetic Resonance in Medicine*, vol. 75, no. 4, pp. 1499–1514, 2016, ISSN: 1522-2594. DOI: 10.1002/mrm.25717.

[36] H. C. M. Holme, S. Rosenzweig, F. Ong, R. N. Wilke, M. Lustig, and M. Uecker, "ENLIVE: An Efficient Nonlinear Method for Calibrationless and Robust Parallel Imaging," *Scientific Reports*, vol. 9, no. 1, p. 3034, Feb. 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-39888-7.

[37] P. J. Shin, P. E. Z. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion: Calibrationless Parallel Imaging," *Magnetic Resonance in Medicine*, vol. 72, no. 4, pp. 959–970, Oct. 2014, ISSN: 07403194. DOI: 10.1002/mrm.24997.

[38] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar, "High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner," vol. 43, no. 4, pp. 1001–1013, Apr. 1998, ISSN: 0031-9155. DOI: 10.1088/0031-9155/43/4/027.

[39] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens, "Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms," *IEEE Transactions on Medical Imaging*, vol. 18, no. 5, pp. 393–403, May 1999, ISSN: 0278-0062. DOI: 10.1109/42.774167.

[40] A. Rezaei, M. Defrise, G. Bal, C. Michel, M. Conti, C. Watson, and J. Nuyts, "Simultaneous Reconstruction of Activity and Attenuation in Time-of-Flight PET," *IEEE Transactions on Medical Imaging*, vol. 31, no. 12, pp. 2224–2233, Dec. 2012, ISSN: 1558-254X. DOI: 10.1109/TMI.2012.2212719.

[41] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 141–151, Jan. 2020, ISSN: 1558-0792. DOI: 10.1109/MSP.2019.2950557.

[42] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *arXiv:1703.10593 [cs]*, Nov. 2018. arXiv: 1703.10593 [cs].

[43] E. K. Cole, F. Ong, S. S. Vasanawala, and J. M. Pauly, "Fast Unsupervised MRI Reconstruction Without Fully-Sampled Ground Truth Data Using Generative Adversarial Networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3988–3997.

[44] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep Image Prior," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1867–1888, Jul. 2020, ISSN: 1573-1405. DOI: 10.1007/s11263-020-01303-4.

[45] Richard McElreath, *Bayesian Statistics without Frequentist Language*, Bayes@Lund2017, Apr. 2017.

[46] P. S. Laplace, *A Philosophical Essay on Probabilities*, trans. by F. L. Emory and F. W. Truscott, https://www.gutenberg.org/ebooks/58881, Feb. 2019.

[47] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996, ISSN: 0035-9246.

[48] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. New York: Springer, 2006, ISBN: 978-0-387-31073-2.

[49] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, R. Christensen, W. Johnson, A. Branscum, T. E. Hanson, B. P. Carlin, T. A. Louis, R. G. Miller, D. Lunn, C. Jackson, N. Best, A. Thomas, D. Spiegelhalter, T. Leonard, T. S. Ferguson, P. Sprent, J. Q. Smith, J. Lawson, H. C. Tuckwell, B. J. T. Morgan, M. Woodward, D. A. G. Rees, L. L. Kupper, B. H. Neelon, and S. M. O'Brien, *Bayesian Data Analysis, Third Edition*. 2014.

[50] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Second. New York: Chapman and Hall/CRC, Mar. 2020, ISBN: 978-0-429-02960-8. DOI: 10.1201/9780429029608.

[51] M. A. Tanner, *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, ser. Springer Series in Statistics. New York, NY: Springer New York, 1996, ISBN: 978-1-4612-8471-0 978-1-4612-4024-2. DOI: 10.1007/978-1-4612-4024-2.

[52] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, "Advances in Variational Inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 2008–2026, Aug. 2019, ISSN: 1939-3539. DOI: 10.1109/TPAMI.2018.2889774.

[53] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, Nov. 2008, ISSN: 1558-0792. DOI: 10.1109/MSP.2008.929620.

[54] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An Introduction to Variational Methods for Graphical Models," *Machine Learning*, vol. 37, no. 2, pp. 183–233, Nov. 1999, ISSN: 1573-0565. DOI: 10.1023/A:1007665907178.

[55] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, ser. UAI'01, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Aug. 2001, pp. 362–369, ISBN: 978-1-55860-800-9.

[56] M. W. Seeger and D. P. Wipf, "Variational Bayesian Inference Techniques," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 81–91, Nov. 2010, ISSN: 1053-5888. DOI: 10.1109/MSP.2010.938082.

[57] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of The 33rd International Conference on Machine Learning*, PMLR, Jun. 2016, pp. 1050–1059.

[58] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight Uncertainty in Neural Network," in *Proceedings of the 32nd International Conference on Machine Learning*, PMLR, Jun. 2015, pp. 1613–1622.

[59] K. Shridhar, F. Laumann, and M. Liwicki, "Uncertainty Estimations by Softplus normalization in Bayesian Convolutional Neural Networks with Variational Inference," *arXiv:1806.05978 [cs, stat]*, May 2019. arXiv: 1806.05978 [cs, stat].

[60] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014, ISSN: 1533-7928.

[61] A. P. Leynes, S. Ahn, K. A. Wangerin, S. S. Kaushik, F. Wiesinger, T. A. Hope, and P. E. Z. Larson, "Attenuation Coefficient Estimation for PET/MRI With Bayesian Deep Learning pseudo-CT and Maximum Likelihood Estimation of Activity and At-

tenuation," *IEEE Transactions on Radiation and Plasma Medical Sciences*, pp. 1–1, 2021, ISSN: 2469-7303. DOI: 10.1109/TRPMS.2021.3118325.

[62] A. P. Leynes, J. Yang, D. D. Shanbhag, S. S. Kaushik, Y. Seo, T. A. Hope, F. Wiesinger, and P. E. Z. Larson, "Hybrid ZTE/Dixon MR-based attenuation correction for quantitative uptake estimation of pelvic lesions in PET/MRI," *Medical Physics*, vol. 44, no. 3, pp. 902–913, 2017, ISSN: 2473-4209. DOI: 10.1002/mp.12122.

[63] A. P. Leynes, J. Yang, F. Wiesinger, S. S. Kaushik, D. D. Shanbhag, Y. Seo, T. A. Hope, and P. E. Z. Larson, "Zero-Echo-Time and Dixon Deep Pseudo-CT (ZeDD CT): Direct Generation of Pseudo-CT Images for Pelvic PET/MRI Attenuation Correction Using Deep Convolutional Neural Networks with Multiparametric MRI," *Journal of Nuclear Medicine*, vol. 59, no. 5, pp. 852–858, May 2018, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.117.198051.

[64] A. P. Leynes and P. E. Z. Larson, "Visualizing the 'ideal' input MRI for synthetic CT generation with a trained deep convolutional neural network: Can we improve the inputs for deep learning models?" In *Proceedings of the 27th Annual Meeting of the International Society for Magnetic Resonance in Medicine*, Montreal, QC, Canada, May 2019.

[65] A. Torrado-Carvajal, J. Vera-Olmos, D. Izquierdo-Garcia, O. A. Catalano, M. A. Morales, J. Margolin, A. Soricelli, M. Salvatore, N. Malpica, and C. Catana, "Dixon-VIBE Deep Learning (DIVIDE) Pseudo-CT Synthesis for Pelvis PET/MR Attenuation Correction," *Journal of Nuclear Medicine*, vol. 60, no. 3, pp. 429–435, Mar. 2019, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.118.209288.

[66] J. S. Lee, "A Review of Deep-Learning-Based Approaches for Attenuation Correction in Positron Emission Tomography," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 2, pp. 160–184, Mar. 2021, ISSN: 2469-7303. DOI: 10.1109/TRPMS.2020.3009269.

[67] H. Arabi and H. Zaidi, "One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI," *European Journal of Nuclear Medicine and Molecular Imaging*, pp. 1–15, Jun. 2016, ISSN: 1619-7070, 1619-7089. DOI: 10.1007/s00259-016-3422-5.

[68] M. Hofmann, I. Bezrukov, F. Mantlik, P. Aschoff, F. Steinke, T. Beyer, B. J. Pichler, and B. Schölkopf, "MRI-Based Attenuation Correction for Whole-Body PET/MRI: Quantitative Evaluation of Segmentation- and Atlas-Based Methods," *Journal of Nuclear Medicine*, vol. 52, no. 9, pp. 1392–1399, Sep. 2011, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.110.078949.

[69] T. Sekine, A. Buck, G. Delso, E. E. G. W. ter Voert, M. Huellner, P. Veit-Haibach, and G. Warnock, "Evaluation of Atlas-Based Attenuation Correction for Integrated PET/MR in Human Brain: Application of a Head Atlas and Comparison to True CT-Based Attenuation Correction," *Journal of Nuclear Medicine*, vol. 57, no. 2, pp. 215–220, Feb. 2016, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.115.159228.

[70] S. Wollenweber, S. Ambwani, G. Delso, A. Lonn, R. Mullick, F. Wiesinger, Z. Piti, A. Tari, G. Novak, and M. Fidrich, "Evaluation of an Atlas-Based PET Head Attenuation Correction Using PET/CT amp; MR Patient Data," *IEEE Transactions on Nuclear Science*, vol. 60, no. 5, pp. 3383–3390, Oct. 2013, ISSN: 0018-9499. DOI: 10.1109/TNS.2013.2273417.

[71] J. Cabello, M. Lukas, S. Förster, T. Pyka, S. G. Nekolla, and S. I. Ziegler, "MR-Based Attenuation Correction Using Ultrashort-Echo-Time Pulse Sequences in Dementia Patients," *Journal of Nuclear Medicine*, vol. 56, no. 3, pp. 423–429, Mar. 2015, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.114.146308.

[72] M. R. Juttukonda, B. G. Mersereau, Y. Chen, Y. Su, B. G. Rubin, T. L. S. Benzinger, D. S. Lalush, and H. An, "MR-based attenuation correction for PET/MRI neurological studies with continuous-valued attenuation coefficients for bone through

a conversion from R2* to CT-Hounsfield units," *NeuroImage*, vol. 112, pp. 160–168, May 2015, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2015.03.009.

[73] V. Keereman, Y. Fierens, T. Broux, Y. D. Deene, M. Lonneux, and S. Vandenberghe, "MRI-Based Attenuation Correction for PET/MRI Using Ultrashort Echo Time Sequences," *Journal of Nuclear Medicine*, vol. 51, no. 5, pp. 812–818, May 2010, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.109.065425.

[74] C. N. Ladefoged, D. Benoit, I. Law, S. Holm, A. Kjær, L. Højgaard, A. E. Hansen, and F. L. Andersen, "Region specific optimization of continuous linear attenuation coefficients based on UTE (RESOLUTE): Application to PET/MR brain imaging," *Physics in Medicine and Biology*, vol. 60, no. 20, p. 8047, 2015, ISSN: 0031-9155. DOI: 10.1088/0031-9155/60/20/8047.

[75] B. K. Navalpakkam, H. Braun, T. Kuwert, and H. H. Quick, "Magnetic Resonance–Based Attenuation Correction for PET/MR Hybrid Imaging Using Continuous Valued Attenuation Maps:" *Investigative Radiology*, vol. 48, no. 5, pp. 323–332, May 2013, ISSN: 0020-9996. DOI: 10.1097/RLI.0b013e318283292f.

[76] G. Delso, F. Wiesinger, M. Carl, G. McKinnon, M. Khalighi, E. ter Voert, F. Barbosa, M. Huellner, and P. Veit-Haibach, "ZTE-based clinical bone imaging for PET/MR," *Journal of Nuclear Medicine*, vol. 56, no. supplement 3, pp. 1806–1806, May 2015, ISSN: 0161-5505, 2159-662X.

[77] A. P. Leynes, J. Yang, D. D. Shanbhag, S. S. Kaushik, Y. Seo, T. A. Hope, F. Wiesinger, and P. E. Z. Larson, "Hybrid ZTE/Dixon MR-based attenuation correction for quantitative uptake estimation of pelvic lesions in PET/MRI," *Medical Physics*, vol. 44, no. 3, pp. 902–913, Mar. 2017, ISSN: 2473-4209. DOI: 10.1002/mp.12122.

[78] T. Sekine, E. E. G. W. ter Voert, G. Warnock, A. Buck, M. Huellner, P. Veit-Haibach, and G. Delso, "Clinical Evaluation of Zero-Echo-Time Attenuation Correction for Brain 18F-FDG PET/MRI: Comparison with Atlas Attenuation Correction," *Journal*

of Nuclear Medicine, vol. 57, no. 12, pp. 1927–1932, Dec. 2016, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.116.175398.

[79] F. Wiesinger, M. Bylund, J. Yang, S. Kaushik, D. Shanbhag, S. Ahn, J. H. Jonsson, J. A. Lundman, T. Hope, T. Nyholm, P. Larson, and C. Cozzini, "Zero TE-based pseudo-CT image conversion in the head and its application in PET/MR attenuation correction and MR-guided radiation therapy planning," *Magnetic Resonance in Medicine*, vol. 80, no. 4, pp. 1440–1451, 2018, ISSN: 1522-2594. DOI: 10.1002/mrm.27134.

[80] J. Yang, F. Wiesinger, S. Kaushik, D. Shanbhag, T. A. Hope, P. E. Z. Larson, and Y. Seo, "Evaluation of Sinus/Edge-Corrected Zero-Echo-Time–Based Attenuation Correction in Brain PET/MRI," *Journal of Nuclear Medicine*, vol. 58, no. 11, pp. 1873–1879, Nov. 2017, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.116.188268.

[81] T. Koesters, K. P. Friedman, M. Fenchel, Y. Zhan, G. Hermosillo, J. Babb, I. O. Jelescu, D. Faul, F. E. Boada, and T. M. Shepherd, "Dixon Sequence with Superimposed Model-Based Bone Compartment Provides Highly Accurate PET/MR Attenuation Correction of the Brain," *Journal of Nuclear Medicine*, vol. 57, no. 6, pp. 918–924, Jun. 2016, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.115.166967.

[82] D. H. Paulus, H. H. Quick, C. Geppert, M. Fenchel, Y. Zhan, G. Hermosillo, D. Faul, F. Boada, K. P. Friedman, and T. Koesters, "Whole-Body PET/MR Imaging: Quantitative Evaluation of a Novel Model-Based MR Attenuation Correction Method Including Bone," *Journal of Nuclear Medicine*, vol. 56, no. 7, pp. 1061–1066, Jul. 2015, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.115.156000.

[83] S. Wollenweber, S. Ambwani, A. Lonn, D. Shanbhag, S. Thiruvenkadam, S. Kaushik, R. Mullick, H. Qian, G. Delso, and F. Wiesinger, "Comparison of 4-Class and Continuous Fat/Water Methods for Whole-Body, MR-Based PET Attenuation Correction,"

IEEE Transactions on Nuclear Science, vol. 60, no. 5, pp. 3391–3398, Oct. 2013, ISSN: 0018-9499. DOI: 10.1109/TNS.2013.2278759.

[84] Y. Chen, C. Ying, M. M. Binkley, M. R. Juttukonda, S. Flores, R. Laforest, T. L. S. Benzinger, and H. An, "Deep learning-based T1-enhanced selection of linear attenuation coefficients (DL-TESLA) for PET/MR attenuation correction in dementia neuroimaging," *Magnetic Resonance in Medicine*, vol. 86, no. 1, pp. 499–513, 2021, ISSN: 1522-2594. DOI: 10.1002/mrm.28689.

[85] K. Gong, J. Yang, P. E. Z. Larson, S. C. Behr, T. A. Hope, Y. Seo, and Q. Li, "MR-Based Attenuation Correction for Brain PET Using 3-D Cycle-Consistent Adversarial Network," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 2, pp. 185–192, Mar. 2021, ISSN: 2469-7303. DOI: 10.1109/TRPMS.2020.3006844.

[86] P. K. Han, D. E. Horng, K. Gong, Y. Petibon, K. Kim, Q. Li, K. A. Johnson, G. E. Fakhri, J. Ouyang, and C. Ma, "MR-based PET attenuation correction using a combined ultrashort echo time/multi-echo Dixon acquisition," *Medical Physics*, vol. 47, no. 7, pp. 3064–3077, 2020, ISSN: 2473-4209. DOI: 10.1002/mp.14180.

[87] F. Liu, H. Jang, R. Kijowski, T. Bradshaw, and A. B. McMillan, "Deep Learning MR Imaging–based Attenuation Correction for PET/MR Imaging," *Radiology*, p. 170 700, Sep. 2017, ISSN: 0033-8419. DOI: 10.1148/radiol.2017170700.

[88] L. Tao, J. Fisher, E. Anaya, X. Li, and C. S. Levin, "Pseudo CT Image Synthesis and Bone Segmentation From MR Images Using Adversarial Networks With Residual Blocks for MR-Based Attenuation Correction of Brain PET Data," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 2, pp. 193–201, Mar. 2021, ISSN: 2469-7303. DOI: 10.1109/TRPMS.2020.2989073.

[89] K. D. Spuhler, J. Gardus, Y. Gao, C. DeLorenzo, R. Parsey, and C. Huang, "Synthesis of Patient-Specific Transmission Data for PET Attenuation Correction for PET/MRI Neuroimaging Using a Convolutional Neural Network," *Journal of Nuclear Medicine*,

vol. 60, no. 4, pp. 555–560, Apr. 2019, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.118.214320.

[90] H. Arabi and H. Zaidi, "Deep learning-guided estimation of attenuation correction factors from time-of-flight PET emission data," *Medical Image Analysis*, vol. 64, p. 101 718, Aug. 2020, ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101718.

[91] F. Hashimoto, M. Ito, K. Ote, T. Isobe, H. Okada, and Y. Ouchi, "Deep learning-based attenuation correction for brain PET with various radiotracers," *Annals of Nuclear Medicine*, vol. 35, no. 6, pp. 691–701, Jun. 2021, ISSN: 1864-6433. DOI: 10.1007/s12149-021-01611-w.

[92] X. Dong, Y. Lei, T. Wang, K. Higgins, T. Liu, W. J. Curran, H. Mao, J. A. Nye, and X. Yang, "Deep learning-based attenuation correction in the absence of structural information for whole-body positron emission tomography imaging," *Physics in Medicine & Biology*, vol. 65, no. 5, p. 055 011, Mar. 2020, ISSN: 0031-9155. DOI: 10.1088/1361-6560/ab652c.

[93] I. Shiri, H. Arabi, P. Geramifar, G. Hajianfar, P. Ghafarian, A. Rahmim, M. R. Ay, and H. Zaidi, "Deep-JASC: Joint attenuation and scatter correction in whole-body 18F-FDG PET using a deep residual network," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, no. 11, pp. 2533–2548, Oct. 2020, ISSN: 1619-7089. DOI: 10.1007/s00259-020-04852-5.

[94] J. Yang, D. Park, G. T. Gullberg, and Y. Seo, "Joint correction of attenuation and scatter in image space using deep convolutional neural networks for dedicated brain 18F-FDG PET," *Physics in Medicine & Biology*, vol. 64, no. 7, p. 075 019, Apr. 2019, ISSN: 0031-9155. DOI: 10.1088/1361-6560/ab0606.

[95] Y. Berker and Y. Li, "Attenuation correction in emission tomography using the emission data-A review: AC in PET and SPECT using emission data," *Medical Physics*, vol. 43, no. 2, pp. 807–832, Jan. 2016, ISSN: 00942405. DOI: 10.1118/1.4938264.

[96] S. Ahn, L. Cheng, D. D. Shanbhag, H. Qian, S. S. Kaushik, F. P. Jansen, and Florian Wiesinger, "Joint estimation of activity and attenuation for PET using pragmatic MR-based prior: Application to clinical TOF PET/MR whole-body data for FDG and non-FDG tracers," *Physics in Medicine & Biology*, vol. 63, no. 4, p. 045 006, 2018, ISSN: 0031-9155. DOI: 10.1088/1361-6560/aaa8a6.

[97] A. Mehranian and H. Zaidi, "Joint Estimation of Activity and Attenuation in Whole-Body TOF PET/MRI Using Constrained Gaussian Mixture Models," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1808–1821, Sep. 2015, ISSN: 0278-0062. DOI: 10.1109/TMI.2015.2409157.

[98] N. Fuin, S. Pedemonte, O. A. Catalano, D. Izquierdo-Garcia, A. Soricelli, M. Salvatore, K. Heberlein, J. M. Hooker, K. V. Leemput, and C. Catana, "PET/MRI in the Presence of Metal Implants: Completion of the Attenuation Map from PET Emission Data," *Journal of Nuclear Medicine*, vol. 58, no. 5, pp. 840–845, May 2017, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.116.183343.

[99] B.-H. Choi, D. Hwang, S.-K. Kang, K.-Y. Kim, H. Choi, S. Seo, and J.-S. Lee, "Accurate Transmission-Less Attenuation Correction Method for Amyloid-β Brain PET Using Deep Neural Network," *Electronics*, vol. 10, no. 15, p. 1836, Jan. 2021. DOI: 10.3390/electronics10151836.

[100] D. Hwang, K. Y. Kim, S. K. Kang, S. Seo, J. C. Paeng, D. S. Lee, and J. S. Lee, "Improving the Accuracy of Simultaneously Reconstructed Activity and Attenuation Maps Using Deep Learning," *Journal of Nuclear Medicine*, vol. 59, no. 10, pp. 1624–1629, Oct. 2018, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.117.202317.

[101] D. Hwang, S. K. Kang, K. Y. Kim, S. Seo, J. C. Paeng, D. S. Lee, and J. S. Lee, "Generation of PET Attenuation Map for Whole-Body Time-of-Flight 18F-FDG PET/MRI Using a Deep Neural Network Trained with Simultaneously Reconstructed Activity and Attenuation Maps," *Journal of Nuclear Medicine*, vol. 60,

no. 8, pp. 1183–1189, Aug. 2019, ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.118.219493.

[102] D. Hwang, S. K. Kang, K. Y. Kim, H. Choi, S. Seo, and J. S. Lee, "Data-driven respiratory phase-matched PET attenuation correction without CT," *Physics in Medicine & Biology*, vol. 66, no. 11, p. 115 009, May 2021, ISSN: 0031-9155. DOI: 10.1088/1361-6560/abfc8f.

[103] J. Y. Huang, J. R. Kerns, J. L. Nute, X. Liu, P. A. Balter, F. C. Stingo, D. S. Followill, D. Mirkovic, R. M. Howell, and S. F. Kry, "An evaluation of three commercially available metal artifact reduction methods for CT imaging," *Physics in Medicine and Biology*, vol. 60, no. 3, pp. 1047–1067, Jan. 2015, ISSN: 0031-9155. DOI: 10.1088/0031-9155/60/3/1047.

[104] C. N. Ladefoged, A. E. Hansen, O. M. Henriksen, F. J. Bruun, L. Eikenes, S. K. Øen, A. Karlberg, L. Højgaard, I. Law, and F. L. Andersen, "AI-driven attenuation correction for brain PET/MRI: Clinical evaluation of a dementia cohort and importance of the training group size," *NeuroImage*, vol. 222, p. 117 221, Nov. 2020, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2020.117221.

[105] A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017, pp. 5574–5584.

[106] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Machine Learning*, vol. 110, no. 3, pp. 457–506, Mar. 2021, ISSN: 1573-0565. DOI: 10.1007/s10994-021-05946-3.

[107] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *NeurIPS Autodiff Workshop 2017*, Oct. 2017.

[108] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Dec. 2014. arXiv: 1412.6980 [cs].

[109] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1026–1034. DOI: 10.1109/ICCV.2015.123.

[110] C. S. Levin, S. H. Maramraju, M. M. Khalighi, T. W. Deller, G. Delso, and F. Jansen, "Design Features and Mutual Compatibility Studies of the Time-of-Flight PET Capable GE SIGNA PET/MR System," *IEEE Transactions on Medical Imaging*, vol. 35, no. 8, pp. 1907–1914, Aug. 2016, ISSN: 0278-0062. DOI: 10.1109/TMI.2016.2537811.

[111] B. Avants, N. Tustison, M. Stauffer, G. Song, B. Wu, and J. Gee, "The Insight ToolKit image registration framework," *Frontiers in Neuroinformatics*, vol. 8, no. 44, Apr. 2014. DOI: 10.3389/fninf.2014.00044.

[112] A. M. Alessio, C. W. Stearns, S. Tong, S. G. Ross, S. Kohlmyer, A. Ganin, and P. E. Kinahan, "Application and Evaluation of a Measured Spatially Variant System Model for PET Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 938–949, Mar. 2010, ISSN: 0278-0062. DOI: 10.1109/TMI.2010.2040188.

[113] A. Samarin, C. Burger, S. D. Wollenweber, D. W. Crook, I. A. Burger, D. T. Schmid, G. K. von Schulthess, and F. P. Kuhn, "PET/MR imaging of bone lesions – implications for PET quantification from imperfect attenuation correction," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 39, no. 7, pp. 1154–1160, Apr. 2012, ISSN: 1619-7070, 1619-7089. DOI: 10.1007/s00259-012-2113-0.

[114] M. Hemsley, B. Chugh, M. Ruschin, Y. Lee, C.-L. Tseng, G. Stanisz, and A. Lau, "Deep Generative Model for Synthetic-CT Generation with Uncertainty Predictions," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou,

D. Racoceanu, and L. Joskowicz, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 834–844, ISBN: 978-3-030-59710-8. DOI: 10.1007/978-3-030-59710-8_81.

[115] P. Klages, I. Benslimane, S. Riyahi, J. Jiang, M. Hunt, J. O. Deasy, H. Veeraraghavan, and N. Tyagi, "Patch-based generative adversarial neural network models for head and neck MR-only planning," *Medical Physics*, vol. 47, no. 2, pp. 626–642, 2020, ISSN: 2473-4209. DOI: 10.1002/mp.13927.

[116] S. L. Bowen, N. Fuin, M. A. Levine, and C. Catana, "Transmission imaging for integrated PET-MR systems," *Physics in Medicine and Biology*, vol. 61, no. 15, pp. 5547–5568, Jul. 2016, ISSN: 0031-9155. DOI: 10.1088/0031-9155/61/15/5547.

[117] P. Veit-Haibach, F. P. Kuhn, F. Wiesinger, G. Delso, and G. von Schulthess, "PET–MR imaging using a tri-modality PET/CT–MR system with a dedicated shuttle in clinical routine," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 26, no. 1, Oct. 2012, ISSN: 0968-5243, 1352-8661. DOI: 10.1007/s10334-012-0344-5.

[118] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014, ISSN: 1522-2594. DOI: 10.1002/mrm.24751.

[119] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed Sensing MRI," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, Mar. 2008, ISSN: 1053-5888. DOI: 10.1109/MSP.2007.914728.

[120] O. N. Jaspan, R. Fleysher, and M. L. Lipton, "Compressed sensing MRI: A review of the clinical literature," *The British Journal of Radiology*, vol. 88, no. 1056, p. 20 150 487, Sep. 2015, ISSN: 0007-1285. DOI: 10.1259/bjr.20150487.

[121] F. Ong, X. Zhu, J. Y. Cheng, K. M. Johnson, P. E. Z. Larson, S. S. Vasanawala, and M. Lustig, "Extreme MRI: Large-scale volumetric dynamic imaging from continuous

non-gated acquisitions," *Magnetic Resonance in Medicine*, vol. 84, no. 4, pp. 1763–1780, 2020, ISSN: 1522-2594. DOI: 10.1002/mrm.28235.

[122] A. M. Rauschecker, J. D. Rudie, L. Xie, J. Wang, M. T. Duong, E. J. Botzolakis, A. M. Kovalovich, J. Egan, T. C. Cook, R. N. Bryan, I. M. Nasrallah, S. Mohan, and J. C. Gee, "Artificial Intelligence System Approaching Neuroradiologist-level Differential Diagnosis Accuracy at Brain MRI," *Radiology*, vol. 295, no. 3, pp. 626–637, Jun. 2020, ISSN: 0033-8419, 1527-1315. DOI: 10.1148/radiol.2020190283.

[123] O. Senouf, S. Vedula, T. Weiss, A. Bronstein, O. Michailovich, and M. Zibulevsky, "Self-supervised Learning of Inverse Problem Solvers in Medical Imaging," in *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, Q. Wang, F. Milletari, H. V. Nguyen, S. Albarqouni, M. J. Cardoso, N. Rieke, Z. Xu, K. Kamnitsas, V. Patel, B. Roysam, S. Jiang, K. Zhou, K. Luu, and N. Le, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 111–119, ISBN: 978-3-030-33391-1. DOI: 10.1007/978-3-030-33391-1_13.

[124] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3172–3191, 2020, ISSN: 1522-2594. DOI: 10.1002/mrm.28378.

[125] J. I. Tamir, S. X. Yu, and M. Lustig, "Unsupervised Deep Basis Pursuit: Learning inverse problems without ground-truth data," Sep. 2019.

[126] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "RARE: Image Reconstruction Using Deep Priors Learned Without Groundtruth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1088–1099, Oct. 2020, ISSN: 1941-0484. DOI: 10.1109/JSTSP.2020.2998402.

[127] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil, "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging," *Magnetic Resonance in Medicine*, vol. 81, no. 1, pp. 439–453, 2019, ISSN: 1522-2594. DOI: 10.1002/mrm.27420.

[128] S. A. H. Hosseini, C. Zhang, S. Weingärtner, S. Moeller, M. Stuber, K. Ugurbil, and M. Akçakaya, "Accelerated coronary MRI with sRAKI: A database-free self-consistent neural network k-space reconstruction for arbitrary undersampling," *PLOS ONE*, vol. 15, no. 2, e0229418, Feb. 2020, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0229418.

[129] M. Z. Darestani and R. Heckel, "Accelerated MRI With Un-Trained Neural Networks," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 724–733, 2021, ISSN: 2333-9403. DOI: 10.1109/TCI.2021.3097596.

[130] Y. Arefeen, O. Beker, J. Cho, H. Yu, E. Adalsteinsson, and B. Bilgic, "Scan-specific artifact reduction in k-space (SPARK) neural networks synergize with physics-based reconstruction to accelerate MRI," *Magnetic Resonance in Medicine*, vol. n/a, no. n/a, ISSN: 1522-2594. DOI: 10.1002/mrm.29036.

[131] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, ISSN: 0028-0836. DOI: 10.1038/323533a0.

[132] Z. Cheng, M. Gadelha, S. Maji, and D. Sheldon, "A Bayesian Perspective on the Deep Image Prior," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA: IEEE, Jun. 2019, pp. 5438–5446, ISBN: 978-1-72813-293-8. DOI: 10.1109/CVPR.2019.00559.

[133] S. Du, J. Lee, H. Li, L. Wang, and X. Zhai, "Gradient Descent Finds Global Minima of Deep Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, PMLR, May 2019, pp. 1675–1685.

[134] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv:1312.6114 [cs, stat]*, May 2014. arXiv: 1312.6114 [cs, stat].

[135] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *International Conference on Learning Representations*, Sep. 2018.

[136] F. Ong and M. Lustig, "SigPy: A python package for high performance iterative reconstruction," *Proceedings of the International Society of Magnetic Resonance in Medicine, Montreal, QC*, vol. 4819, May 2019.

[137] K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, and L. L. Wald, "Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty," *Magnetic Resonance in Medicine*, vol. 67, no. 5, pp. 1210–1224, 2012, ISSN: 1522-2594. DOI: 10.1002/mrm.23097.

[138] T. M. Quan and W.-K. Jeong, "Compressed sensing reconstruction of dynamic contrast enhanced MRI using GPU-accelerated convolutional sparse coding," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2016, pp. 518–521. DOI: 10.1109/ISBI.2016.7493321.

[139] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A Deep Semi-NMF Model for Learning Hidden Representations," p. 9,

[140] A. Pramanik, H. K. Aggarwal, and M. Jacob, "Deep Generalization of Structured Low-Rank Algorithms (Deep-SLR)," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4186–4197, Dec. 2020, ISSN: 1558-254X. DOI: 10.1109/TMI.2020.3014581.

[141] A. Pramanik, H. Aggarwal, and M. Jacob, "Calibrationless Parallel MRI Using Model Based Deep Learning (C-MODL)," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2020, pp. 1428–1431. DOI: 10.1109/ISBI45749.2020.9098490.

[142] J. Zhao, X. Liu, S. He, and S. Sun, "Probabilistic inference of Bayesian neural networks with generalized expectation propagation," *Neurocomputing*, vol. 412, pp. 392–398, Oct. 2020, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2020.06.060.

[143] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017, pp. 6402–6413.

[144] M. Andrychowicz, M. Denil, S. Gómez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. de Freitas, "Learning to learn by gradient descent by gradient descent," in *Advances in Neural Information Processing Systems*, vol. 29, Curran Associates, Inc., 2016.

[145] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *arXiv:1511.06434 [cs]*, Nov. 2015. arXiv: 1511.06434 [cs].

[146] Z. Nado, N. Band, M. Collier, J. Djolonga, M. W. Dusenberry, S. Farquhar, Q. Feng, A. Filos, M. Havasi, R. Jenatton, G. Jerfel, J. Liu, Z. Mariet, J. Nixon, S. Padhy, J. Ren, T. G. J. Rudner, F. Sbahi, Y. Wen, F. Wenzel, K. Murphy, D. Sculley, B. Lakshminarayanan, J. Snoek, Y. Gal, and D. Tran, "Uncertainty Baselines: Benchmarks for Uncertainty & Robustness in Deep Learning," *arXiv:2106.04015 [cs]*, Jan. 2022. arXiv: 2106.04015 [cs].

[147] A. Kamath, D. Gnaneshwar, and M. Valdenegro-Toro, "Know Where To Drop Your Weights: Towards Faster Uncertainty Estimation," *arXiv:2010.14019 [cs, stat]*, Oct. 2020. arXiv: 2010.14019 [cs, stat].

[148] S. T. Radev, U. K. Mertens, A. Voss, L. Ardizzone, and U. Köthe, "BayesFlow: Learning complex stochastic models with invertible neural networks," *arXiv:2003.06281 [cs, stat]*, Dec. 2020. arXiv: 2003.06281 [cs, stat].

[149] J. Rock, T. Azevedo, R. de Jong, D. Ruiz-Muñoz, and P. Maji, "On Efficient Uncertainty Estimation for Resource-Constrained Mobile Applications," *arXiv:2111.09838 [cs]*, Dec. 2021. arXiv: 2111.09838 [cs].

[150] A. Tripp, E. Daxberger, and J. M. Hernández-Lobato, "Sample-Efficient Optimization in the Latent Space of Deep Generative Models via Weighted Retraining," in *NeurIPS Workshop on Bayesian Deep Learning*, p. 14.

[151] L. H. Nguyen and J.-A. Goulet, "Analytically Tractable Inference in Neural Networks – An Alternative to Backpropagation," in *NeurIPS Workshop on Bayesian Deep Learning*, p. 5.

[152] V. Fortuin, "Priors in Bayesian Deep Learning: A Review," *arXiv:2105.06868 [cs, stat]*, May 2021. arXiv: 2105.06868 [cs, stat].

[153] A. Alexos, A. Boyd, and S. Mandt, "Structured Stochastic Gradient MCMC: A hybrid VI and MCMC approach," in *NeurIPS Workshop on Bayesian Deep Learning*, p. 27.

[154] M. Zhang, P. Hayes, and D. Barber, "Generalization Gap in Amortized Inference," in *NeurIPS Workshop on Bayesian Deep Learning*, p. 6.

[155] D. Narnhofer, A. Effland, E. Kobler, K. Hammernik, F. Knoll, and T. Pock, "Bayesian Uncertainty Estimation of Learned Variational MRI Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 41, no. 2, pp. 279–291, Feb. 2022, ISSN: 1558-254X. DOI: 10.1109/TMI.2021.3112040.

[156] J. Antorán, A. Weller, U. Bhatt, T. Adel, and J. M. Hernández-Lobato, "Getting A Clue: A Method For Explaining Uncertainty Estimates," in *NeurIPS Workshop on Bayesian Deep Learning*, 2021, p. 34.

[157] R. Mehta, T. Christinck, T. Nair, A. Bussy, S. Premasiri, M. Costantino, M. M. Chakravarthy, D. L. Arnold, Y. Gal, and T. Arbel, "Propagating Uncertainty Across Cascaded Medical Imaging Tasks for Improved Deep Learning Inference," *IEEE Trans-*

actions on Medical Imaging, vol. 41, no. 2, pp. 360–373, Feb. 2022, ISSN: 1558-254X. DOI: 10.1109/TMI.2021.3114097.

[158] *Bayesian Deep Learning Workshop — NeurIPS*, http://bayesiandeeplearning.org.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
EFAFE692F3A4477...
Author Signature

2/24/2022
Date