A Conditional Normalizing Flow for Accelerated Multi-Coil MR Imaging

Jeffrey Wen [1] **Rizwan Ahmad** [2] **Philip Schniter** [1]

Abstract

Accelerated magnetic resonance (MR) imaging attempts to reduce acquisition time by collecting data below the Nyquist rate. As an ill-posed inverse problem, many plausible solutions exist, yet the majority of deep learning approaches generate only a single solution. We instead focus on sampling from the posterior distribution, which provides more comprehensive information for downstream inference tasks. To do this, we design a novel conditional normalizing flow (CNF) that infers the signal component in the measurement operator's nullspace, which is later combined with measured data to form complete images. Using fastMRI brain and knee data, we demonstrate fast inference and accuracy that surpasses recent posterior sampling techniques for MRI. Code is available at `https://github.com/jwen307/mri_cnf`

1. Introduction

Magnetic resonance imaging (MRI) is a routine diagnostic imaging tool that has the potential to provide high-quality soft-tissue images without exposure to ionizing radiation. However, MRI exams are generally time-consuming, which reduces throughput, compromises patient comfort, and increases the likelihood of artifacts from patient motion. Scan time can be reduced by sampling below the Nyquist rate, but this makes the image reconstruction process more challenging. Hence, recovering high-accuracy images from highly subsampled MRI scans has become an active area of research (Knoll et al., 2020).

Many approaches have been proposed to recover MR images from subsampled measurements. Parallel imaging, which is available on all commercial scanners, takes advantage of the availability of multiple receiver coils. After estimating coil-sensitivity maps or interpolation kernels, methods like SENSE (Pruessmann et al., 1999) and GRAPPA (Griswold et al., 2002) can use subsampled data from multiple coils to remove aliasing artifacts in the final reconstruction. However, parallel imaging alone can typically allow only two- to three-fold acceleration of the acquisition process. For higher acceleration, methods based on compressed-sensing (CS) have been proposed (Lustig et al., 2007). The CS methods are framed as iteratively minimizing the sum of a data-fidelity term and a regularization term, where the regularization term incorporates prior knowledge about the images. The prior knowledge could be that the true images are sparse in some transform domain, as in traditional CS, or that the true images are preserved by some denoising function, as in "plug-and-play" recovery (Ahmad et al., 2020). Deep neural networks have also been proposed for MR image recovery, based on end-to-end approaches like (Zbontar et al., 2018; Eo et al., 2018; Sriram et al., 2020) or algorithmic unrolling (Hammernik et al., 2018). Yet another approach, known as compressed sensing with a generative model (CSGM) (Bora et al., 2017), trains a deep image generator and then optimizes its input to give the image that, after application of the forward model, best matches the measurements.

Although they achieve high reconstruction quality, the aforementioned methods provide only a point estimate. Yet, accelerated MRI is an ill-posed inverse problem, where there exist many possible reconstructions that are consistent with a given prior and set of subsampled measurements. Since small variations in image content can impact the final diagnosis, it is crucial for radiologists to know whether a visual structure is truly reflective of the patient anatomy or merely an imaging artifact. Problems of this form fall into the realm of uncertainty quantification (UQ) (Abdar et al., 2021).

One approach that facilitates UQ is Bayesian imaging, where the goal is not to compute a single "good" image estimate but rather to sample from the posterior distribution. The availability of a large batch of posterior samples enables many forms of UQ. For example, a simple approach is to generate the pixel-wise standard-deviation map, which quantifies which pixels are more trustworthy. A more involved approach is to construct a hypothesis test for the absence of a particular (multi-pixel) visual structure (Repetti et al., 2019).

[1]Dept. of ECE, The Ohio State University, Columbus, OH 43210, USA. [2]Dept. of BME, The Ohio State University, Columbus, OH 43210, USA. Correspondence to: Jeffrey Wen <wen.254@osu.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

In this paper, we focus on the task of sampling from the posterior, which facilitates future work that uses those samples for uncertainty quantification, adaptive sampling (Sanchez et al., 2020), counterfactual diagnosis (Chang et al., 2019), or other applications.

There exist several deep-learning based approaches to sample from the posterior, including those based on conditional generative adversarial networks (CGANs) (Isola et al., 2017; Adler & Öktem, 2018), conditional variational autoencoders (CVAEs) (Edupuganti et al., 2021; Tonolini et al., 2020), conditional normalizing flows (CNFs) (Ardizzone et al., 2019; Winkler et al., 2019), and score/Langevin/diffusion-based approaches (Kadkhodaie & Simoncelli, 2020; Laumont et al., 2022; Ho et al., 2020). In this paper, we focus on the CNF approach. Compared to the other methods, CNFs yield rapid inference and require only simple, likelihood-based training. In a recent super-resolution (SR) contest (Lugmayr et al., 2022), a CNF (by Song et al. (2022)) won, beating all CGAN, CVAE, and diffusion-based competitors.

Inspired by the success of CNFs in SR, we design the first CNF for accelerated multi-coil MRI. Previous applications of CNFs to MRI (Denker et al., 2021a) showed competitive results but were restricted to single-coil recovery of magnitude images. As the vast majority of modern MRI scanners capture multi-coil data, the extension to multi-coil, complex-valued data is crucial for real-world adoption. However, the order-of-magnitude increase in dimensionality makes this transition non-trivial. For this purpose, we propose a novel CNF that infers only the signal component in the nullspace of the measurement operator and combines its output with the measured data to generate complete images. Using fastMRI brain and knee data, we demonstrate that our approach outperforms existing posterior samplers based on CGANs (Adler & Öktem, 2018) and MRI-specific score/Langevin-based approaches (Jalal et al., 2021a; Chung & Ye, 2022a) in almost all accuracy metrics, while retaining fast inference and requiring minimal hyperparameter tuning.

2. Background

2.1. Measurement Model

In MRI, measurements of the D-pixel true image $\boldsymbol{i}_{\mathsf{true}} \in \mathbb{C}^D$ are collected in the spatial Fourier domain, known as the "k-space." In a multi-coil system with C coils, measurements from the cth coil can be written as

$$\boldsymbol{k}_c = \boldsymbol{P}\boldsymbol{F}\boldsymbol{S}_c\boldsymbol{i}_{\mathsf{true}} + \boldsymbol{\epsilon}_c \in \mathbb{C}^M, \tag{1}$$

where $\boldsymbol{P} \in \mathbb{R}^{M\times D}$ is a sampling matrix containing M rows of the $D \times D$ identity matrix $\boldsymbol{I}$, $\boldsymbol{F}$ is the $D \times D$ 2D unitary discrete Fourier transform (DFT) matrix, $\boldsymbol{S}_c \in \mathbb{C}^{D\times D}$ is the coil-sensitivity map of the cth coil, and $\boldsymbol{\epsilon}_c \in \mathbb{C}^M$ is measurement noise. We will assume that $\{\boldsymbol{S}_c\}_{c=1}^C$ have been obtained from ESPIRiT (Uecker et al., 2014), in which case $\sum_{c=1}^C \boldsymbol{S}_c^{\mathsf{H}}\boldsymbol{S}_c = \boldsymbol{I}$. In the case of single-coil MRI, $C = 1$ and $\boldsymbol{S}_1 = \boldsymbol{I}$.

We now rewrite the model in terms of the "coil images" $\boldsymbol{x}_c \triangleq \boldsymbol{S}_c\boldsymbol{i}$ and their corresponding "zero-filled" estimates $\boldsymbol{y}_c \triangleq \boldsymbol{F}^{\mathsf{H}}\boldsymbol{P}^\top\boldsymbol{k}_c$, and then stack all the coils together via $\boldsymbol{x}_{\mathsf{true}} \triangleq [\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_C^\top]^\top$ and $\boldsymbol{y} \triangleq [\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_C^\top]^\top$ to obtain

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_{\mathsf{true}} + \boldsymbol{\varepsilon}, \tag{2}$$

with $\boldsymbol{\varepsilon} = [(\boldsymbol{F}^{\mathsf{H}}\boldsymbol{P}^\top\boldsymbol{\epsilon}_1)^\top, \ldots, (\boldsymbol{F}^{\mathsf{H}}\boldsymbol{P}^\top\boldsymbol{\epsilon}_C)^\top]^\top$ and forward operator

$$\boldsymbol{A} = \operatorname{blkdiag}\left\{\boldsymbol{F}^{\mathsf{H}}\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{F}, \ldots, \boldsymbol{F}^{\mathsf{H}}\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{F}\right. . \tag{3}$$

To perform image recovery, one can first compute $\boldsymbol{y}$, then estimate $\widehat{\boldsymbol{x}}$ from $\boldsymbol{y}$, and finally either "coil-combine" to yield a complex-valued image estimate

$$\widehat{\boldsymbol{i}} = [\boldsymbol{S}_1^{\mathsf{H}}, \ldots, \boldsymbol{S}_C^{\mathsf{H}}]\widehat{\boldsymbol{x}} \tag{4}$$

or perform root-sum-of-squares (RSS) reconstruction to obtain a magnitude-only image estimate

$$|\widehat{\boldsymbol{i}}| = \sqrt{\textstyle\sum_{c=1}^C |\widehat{\boldsymbol{x}}_c|^2}. \tag{5}$$

In the "fully sampled" case, $M = D$ and so $\boldsymbol{y} = \boldsymbol{x}_{\mathsf{true}} + \boldsymbol{\varepsilon}$. But fully sampled acquisition is very slow, and so we are interested in accelerated MRI, where one collects $M<D$ measurements per coil to save time. This gives an "acceleration factor" of $R \triangleq D/M$, but it makes $\boldsymbol{A}$ rank deficient. In this latter case, accurate recovery of $\boldsymbol{x}_{\mathsf{true}}$ requires the use of prior information about $\boldsymbol{x}_{\mathsf{true}}$, such as the knowledge that $\boldsymbol{x}_{\mathsf{true}}$ is a vector of MRI coil images.

2.2. Posterior Sampling

In the case of MRI, the posterior distribution that we would ultimately like to sample from is $p_{i|k}(\cdot|\boldsymbol{k})$, where $\boldsymbol{k} \triangleq [\boldsymbol{k}_1^\top, \ldots, \boldsymbol{k}_C^\top]$. Equivalently, we could consider $p_{i|y}(\cdot|\boldsymbol{y})$ since $\boldsymbol{y}$ and $\boldsymbol{k}$ contain the same information. Another option is to sample from $p_{x|y}(\cdot|\boldsymbol{y})$ and then use (4) or (5) to combine coil images into a single image. We take the latter approach.

For CNFs and CGANs, posterior sampling is accomplished by designing a neural network that maps samples from an easy-to-generate latent distribution (e.g., white Gaussian) to the target distribution (i.e., the distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$, with density $p_{x|y}$). Once that network is trained, sample generation is extremely fast. For Langevin dynamics, an algorithm is run for hundreds or thousands of iterations to generate each sample, and each iteration involves calling a neural network. Consequently, the inference time is much longer than that of CNFs and CGANs.

2.3. Conditional Normalizing Flows

Normalizing flows (NF) (Dinh et al., 2015; 2017; Kingma & Dhariwal, 2018; Papamakarios et al., 2021) have emerged as powerful generative models capable of modeling complex data distributions. Normalizing flows learn an invertible mapping between a target data distribution and a simple latent distribution, generally a Gaussian. More concretely, for a latent sample $\boldsymbol{z}$ drawn from the latent distribution p_z, the normalizing flow defines an invertible transformation $f_{\boldsymbol{\theta}}(\cdot) : \mathbb{R}^Q \to \mathbb{R}^Q$. This transformation is parameterized by $\boldsymbol{\theta}$, and $\boldsymbol{x} = f_{\boldsymbol{\theta}}(\boldsymbol{z})$ defines a sample in the target data domain. This mapping of the latent distribution induces a probability in the target data domain with a probability density derived from the change-of-variable formula

$$\widehat{p}_x(\boldsymbol{x};\boldsymbol{\theta}) = p_z(f_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}))\ \det\left(\frac{\partial f_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right), \tag{6}$$

where $\det(\cdot)$ denotes the determinant. The goal of the normalizing flow is to approximate the underlying data distribution p_x with $\widehat{p}_x(\cdot;\boldsymbol{\theta})$. Given a set of data samples $\{\boldsymbol{x}^{(i)}\}_{i=1}^N$, the parameters $\boldsymbol{\theta}$ can be fit using a maximum likelihood loss

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln \widehat{p}_x(\boldsymbol{x}^{(i)};\boldsymbol{\theta}) \tag{7}$$

$$= \sum_{i=1}^{N} \ln p_z(f_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}^{(i)})) + \ln\ \det\left(\frac{\partial f_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}^{(i)})}{\partial \boldsymbol{x}^{(i)}}\right) \tag{8}$$

Once the training is complete, samples from the target distribution can be rapidly generated by drawing samples from the latent distribution and passing them through the normalizing flow $f_{\boldsymbol{\theta}}$.

It is worth noting that maximizing $L(\boldsymbol{\theta})$ is equivalent to minimizing the Kullback-Leibler (KL) divergence between $\widehat{p}_x(\cdot;\boldsymbol{\theta})$ and p_x (Papamakarios et al., 2021), which aligns with the goal of approximating p_x with $\widehat{p}_x(\cdot;\boldsymbol{\theta})$. The maximum-likelihood loss provides stable training with minimal hyperparameter tuning and has been shown to be robust to mode collapse.

Conditional normalizing flows (CNFs) (Ardizzone et al., 2021) generalize normalizing flows by adding a conditioning signal $\boldsymbol{y}$. With the CNF denoted as $h_{\boldsymbol{\theta}}(\cdot,\cdot) : \mathbb{R}^Q \times \mathbb{R}^Q \to \mathbb{R}^Q$, the forward process from the latent domain to the data domain is given by $\boldsymbol{x} = h_{\boldsymbol{\theta}}(\boldsymbol{z},\boldsymbol{y})$. For complex-valued, multi-coil MRI, we have $Q = 2CD$. The inclusion of $\boldsymbol{y}$ alters the objective of the CNF to approximating the unknown posterior distribution $p_{x|y}(\cdot|\boldsymbol{y})$ with $\widehat{p}_{x|y}(\cdot|\boldsymbol{y};\boldsymbol{\theta})$. As before, the change-of-variable formula implies the induced distribution

$$\widehat{p}_{x|y}(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}) = p_z(h_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x},\boldsymbol{y}))\ \det\left(\frac{\partial h_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{x}}\right), \tag{9}$$

where $h_{\boldsymbol{\theta}}^{-1}$ refers to the inverse mapping of $h_{\boldsymbol{\theta}}$ with respect to its first argument.

Given a dataset $\{(\boldsymbol{x}^{(i)},\boldsymbol{y}^{(i)})\}_{i=1}^N$, the maximum likelihood loss can be utilized to optimize the parameters $\boldsymbol{\theta}$

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln \widehat{p}_{x|y}(\boldsymbol{x}^{(i)}|\boldsymbol{y}^{(i)};\boldsymbol{\theta}) \tag{10}$$

$$= \sum_{i=1}^{N} \ln p_z(h_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}^{(i)},\boldsymbol{y}^{(i)})) + \ln\ \det\left(\frac{\partial h_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}^{(i)},\boldsymbol{y}^{(i)})}{\partial \boldsymbol{x}^{(i)}}\right) \tag{11}$$

CNFs have shown promising performance in solving inverse problems, such as super-resolution (Lugmayr et al., 2020; Kim & Son, 2021; Song et al., 2022), making it an exciting avenue of exploration for accelerated MRI. Denker et al. (2021a) developed a CNF for single-coil, magnitude-only knee images. This study showed promising initial results, but the limited scope did not demonstrate performance in the more realistic multi-coil, complex-valued domain. As this transition increases the dimensionality by an order of magnitude, non-trivial architectural changes are required. In this paper, we build on the latest advances in CNFs to create a method that is capable of generating high-quality posterior samples of multi-coil, complex-valued MRI images.

3. Method

Our CNF consists of two networks, a conditioning network $g_{\boldsymbol{\theta}}$ and a conditional flow model $h_{\boldsymbol{\theta}}$. The conditioning network takes the vector of zero-filled (ZF) coil-images $\boldsymbol{y}$ as input and produces features that are used as conditioning information by the flow model $h_{\boldsymbol{\theta}}$. Aided by the conditioning information, $h_{\boldsymbol{\theta}}$ learns an invertible mapping between samples in the latent space and those in the image space. Using the notation of Sec. 2.3, our overall CNF takes the form

$$\overline{h}_{\boldsymbol{\theta}}(\boldsymbol{z},\boldsymbol{y}) \triangleq h_{\boldsymbol{\theta}}(\boldsymbol{z}, g_{\boldsymbol{\theta}}(\boldsymbol{y})). \tag{12}$$

Recently, advancements of CNFs in the super-resolution literature have revealed useful insights for more general inverse problems. First, Lugmayr et al. (2020) suggested the use of a pretrained, state-of-the-art point-estimate network for the conditioning network $g_{\boldsymbol{\theta}}$. This network is then trained jointly with $h_{\boldsymbol{\theta}}$ using the loss in (11). This approach provides a functional initialization of $g_{\boldsymbol{\theta}}$ and allows $g_{\boldsymbol{\theta}}$ to learn to provide features that are useful for the maximum-likelihood training objective. We utilize a UNet from (Zbontar et al., 2018) for $g_{\boldsymbol{\theta}}$ since it has been shown to perform well in accelerated MRI. We first pre-train $g_{\boldsymbol{\theta}}$ for MRI recovery, and later we jointly train $g_{\boldsymbol{\theta}}$ and $h_{\boldsymbol{\theta}}$ together.

Song et al. (2022) demonstrated the benefits of using "frequency-separation" when training a CNF for super-

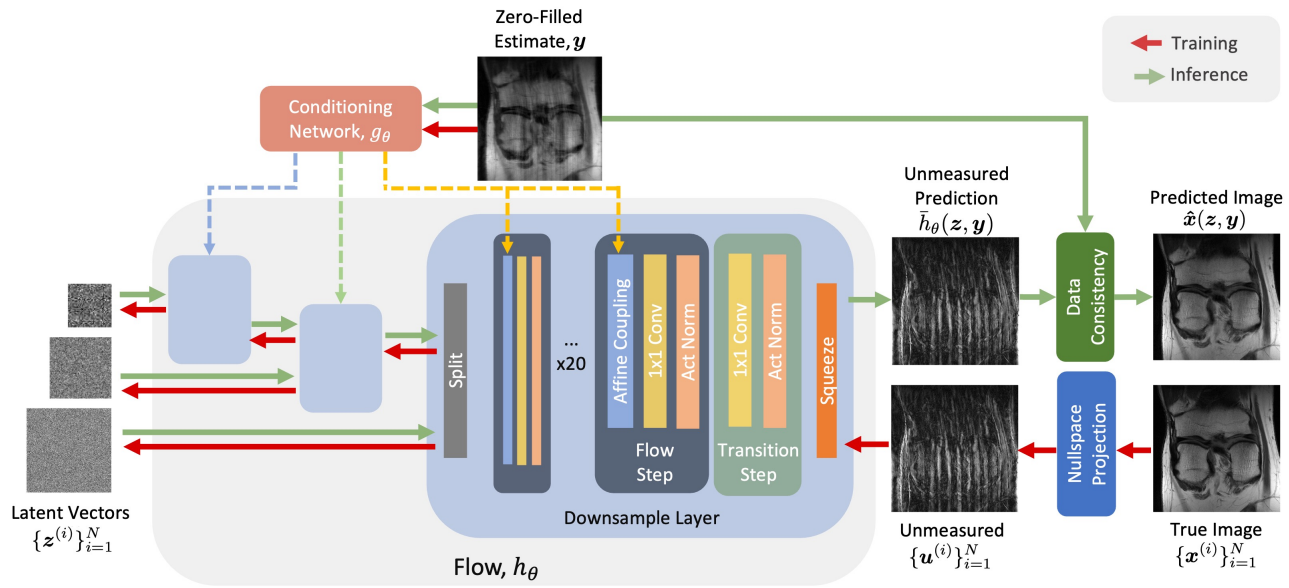

Figure 1. The architecture of our CNF. The conditioning network $g_{\boldsymbol{\theta}}$ takes in multi-coil zero-filled image estimates $\boldsymbol{y}$ and outputs features used by the flow model $h_{\boldsymbol{\theta}}$. The flow learns an invertible mapping between Gaussian random samples $\boldsymbol{z}^{(i)}$ and images $\boldsymbol{u}^{(i)}$ that are the projections of the training images $\boldsymbol{x}^{(i)}$ onto the non-measured subspace.

resolution. The authors argue that the low-resolution conditional image already contains sufficient information about the low-frequency components of the image, so the CNF can focus on recovering only the high-frequency information. The CNF output is then added to an upsampled version of the conditional image to yield an estimate of the full image.

We now generalize the frequency-separation idea to arbitrary linear models of the form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_{\text{true}} + \boldsymbol{\varepsilon}$ from (2) and apply the resulting procedure to MRI. Notice that (2) implies

$$\boldsymbol{A}^+\boldsymbol{y} = \boldsymbol{A}^+\boldsymbol{A}\boldsymbol{x}_{\text{true}} + \boldsymbol{A}^+\boldsymbol{\varepsilon} \tag{13}$$

where $(\cdot)^+$ denotes the pseudo-inverse. Here, $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{x}_{\text{true}}$ is recognized as the projection of $\boldsymbol{x}_{\text{true}}$ onto the row-space of $\boldsymbol{A}$, which we will refer to as the "measured space." Then

$$\boldsymbol{u}_{\text{true}} \triangleq (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x}_{\text{true}} \tag{14}$$

would be the projection of $\boldsymbol{x}_{\text{true}}$ onto its orthogonal complement, which we refer to as the "nullspace." Assuming that the nullspace has dimension > 0, we propose to construct an estimate $\widehat{\boldsymbol{x}}$ of $\boldsymbol{x}_{\text{true}}$ with the form

$$\widehat{\boldsymbol{x}}(\boldsymbol{z}, \boldsymbol{y}) = (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\overline{h}_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{y}) + \boldsymbol{A}^+\boldsymbol{y}, \tag{15}$$

where $\overline{h}_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{y})$ is our CNF-generated estimate of $\boldsymbol{u}_{\text{true}}$ and the $(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})$ in (15) strips off any part of $\overline{h}_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{y})$ that has leaked into the measured space. A similar approach was used in (Sønderby et al., 2017) for point estimation. Given training data $\{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)})\}_{i=1}^N$, the CNF $\overline{h}_{\boldsymbol{\theta}}(\cdot, \cdot)$ is trained to map code vectors $\boldsymbol{z}^{(i)} \sim p_z$ to the nullspace projections

$$\boldsymbol{u}^{(i)} \triangleq (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x}^{(i)} \tag{16}$$

using the measured data $\boldsymbol{y}^{(i)}$ as the conditional information. As a result of (15), the reconstructions $\widehat{\boldsymbol{x}}$ agree with the measurements $\boldsymbol{y}$ in that $\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{y}$. However, this also means that $\widehat{\boldsymbol{x}}$ inherits the noise $\boldsymbol{\varepsilon}$ corrupting $\boldsymbol{y}$, and so this data-consistency procedure is best used in the low-noise regime. In the presence of significant noise, the dual-decomposition approach (Chen & Davies, 2020) may be more appropriate.

In the accelerated MRI formulation (1)-(3), the matrix $\boldsymbol{A}$ is itself an orthogonal projection matrix, so that, in (15),

$$\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A} = \text{blkdiag}\{\boldsymbol{F}^{\mathsf{H}}\widetilde{\boldsymbol{P}}^\top\widetilde{\boldsymbol{P}}\boldsymbol{F}, \ldots, \boldsymbol{F}^{\mathsf{H}}\widetilde{\boldsymbol{P}}^\top\widetilde{\boldsymbol{P}}\boldsymbol{F}\ , \tag{17}$$

where $\widetilde{\boldsymbol{P}} \in \mathbb{R}^{(D-M)\times D}$ is the sampling matrix for the non-measured k-space. Also, $\boldsymbol{y}$ is in the row-space of $\boldsymbol{A}$, so

$$\boldsymbol{A}^+\boldsymbol{y} = \boldsymbol{y} \tag{18}$$

in (15). Figure 1 illustrates the overall procedure, using "data consistency" to describe (15) and "nullspace projection" to describe (16). In Sec. 4.2, we quantitatively demonstrate the improvements gained from designing our CNF to estimate only the nullspace component.

3.1. Architecture

The backbone of $g_{\boldsymbol{\theta}}$ is a UNet (Ronneberger et al., 2015) that mimics the design in (Zbontar et al., 2018), with 4 pooling layers and 128 output channels in the first convolution layer. The first layer was modified to accept complex-valued coil images. The inputs have $2C$ channels, where C is the number of coils, each with a real and imaginary component.

The outputs of the final feature layer of the UNet are processed by a feature-extraction network with L convolution layers. Together, the feature extraction network and the UNet make up our conditioning network $g_{\boldsymbol{\theta}}$. The output of each convolution layer is fed to conditional coupling blocks of the corresponding layer in $h_{\boldsymbol{\theta}}$.

For the flow model $h_{\boldsymbol{\theta}}$, we adopt the multi-scale RealNVP (Dinh et al., 2017) architecture. This construction utilizes L-layers and B-flow steps in each layer. A flow step consists of an activation normalization (Kingma & Dhariwal, 2018), a fixed 1×1 orthogonal convolution (Ardizzone et al., 2019), and a conditional coupling block (Ardizzone et al., 2021). Each layer begins with a checkerboard downsampling (squeeze layer) (Dinh et al., 2017) and a transition step made up of an activation normalization and 1×1 convolution. Layers end with a split operation that sends half of the channels directly to the output on the latent side. For all experiments, we use $L = 3$ and $B = 20$. The full architecture of $h_{\boldsymbol{\theta}}$ is specified in Fig. 1.

Although the code that accompanies (Denker et al., 2021a) gives a built-in mechanism to scale their flow architecture to accommodate an increased number of input and output channels, we find that this mechanism does not work well (see Sec. 4.2). Thus, in addition to incorporating nullspace learning, we redesign several aspects of the flow architecture and training. First, to prevent the number of flow parameters from growing unreasonably large, our flow uses fewer downsampling layers (3 vs 6) but more flow steps per downsampling layer (20 vs 5), and we utilize one-sided (instead of two-sided) affine coupling layers. Second, to connect the conditioning network to the flow, Denker et al. (2021a) used a separate CNN for each flow layer and adjusted its depth to match the flow-layer dimension. We use a single, larger CNN and feed its intermediate features to the flow layers with matched dimensions, further preventing an explosion in the number of parameters. Third, our conditioning network uses a large, pretrained UNet, whereas Denker et al. (2021a) used a smaller untrained UNet. With our modifications, we grow the conditional network more than the flow network, which allows the CNF to better handle the high dimensionality of complex-valued, multi-coil data.

3.2. Data

We apply our network to two datasets: the fastMRI knee and fastMRI brain datasets (Zbontar et al., 2018). For the knee data, we use the non-fat-suppressed subset, giving 17286 training and 3592 validation images. We compress the measurements to $C = 8$ complex-valued virtual coils using (Zhang et al., 2013) and crop the images to 320×320 pixels. The sampling mask is generated using the golden ratio offset (GRO) (Joshi et al., 2022) Cartesian sampling scheme with an acceleration rate $R = 4$ and autocalibration signal (ACS) region of 13 pixels. We create the ZF coil-image vectors $\boldsymbol{y}$ by applying the mask and inverse Fourier transform to the fully sampled $\boldsymbol{k}_c$ given by the fastMRI dataset to obtain $\boldsymbol{y}_c = \boldsymbol{F}^{\mathsf{H}}\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{k}_c$ for all c, and then stack the coils to obtain $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_C^\top]^\top$. We create the ground-truth coil-image vectors $\boldsymbol{x}_{\mathsf{true}}$ using the same procedure but without the mask, i.e., $\boldsymbol{x}_c = \boldsymbol{F}^{\mathsf{H}}\boldsymbol{k}_c$ and $\boldsymbol{x}_{\mathsf{true}} = [\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_C^\top]^\top$.

With the brain data, we use the T2-weighted images and take the first 8 slices of all volumes with at least 8 coils. This provides 12224 training and 3352 validation images. The data is compressed to $C = 8$ virtual coils (Zhang et al., 2013) and cropped to 384×384 pixels. The GRO sampling scheme is again used with an acceleration rate $R = 4$ and a 32-wide ACS region. For both methods, the coil-sensitivity maps are estimated from the ACS region using ESPIRiT (Uecker et al., 2014). All inputs to the network are normalized by the 95th percentile of the ZF magnitude images.

3.3. Training

For both datasets, we first train the UNet in $g_{\boldsymbol{\theta}}$ with an additional 1×1 convolution layer to get the desired $2C$ channels. We train the UNet to minimize the mean-squared error (MSE) from the nullspace projected targets $\{\boldsymbol{u}^{(i)}\}_{i=1}^N$ for 50 epochs with batch size 8 and learning rate 0.003. Then, we remove the final 1×1 convolution and jointly train $g_{\boldsymbol{\theta}}$ and $h_{\boldsymbol{\theta}}$ for 100 epochs to minimize the negative log-likelihood (NLL) loss of the nullspace projected targets. For the brain data, we use batch size 8 and learning rate 0.0003. For the knee data, we use batch size 16 with learning rate 0.0005. All experiments use the Adam optimizer (Kingma & Ba, 2015) with default parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The full training takes about 4 days on 4 Nvidia V100 GPUs.

3.4. Comparison Methods

We compare against other methods that are capable of generating posterior samples for accelerated MRI. For the fastMRI brain data, we present results for the CGAN from (Adler & Öktem, 2018) and the Langevin method from (Jalal et al., 2021a). For the fastMRI knee data, we present results for the "Score" method from (Chung & Ye, 2022a) and the "sCNF" method from (Denker et al., 2021a).

For the CGAN, we utilize a UNet-based generator with 4 pooling layers and 128 output channels in the initial layer and a 5-layer CNN network for the discriminator. The generator takes $\boldsymbol{y}$ concatenated with a latent vector $\boldsymbol{z}$ as input. The model is trained with the default loss and hyperparameters from (Adler & Öktem, 2018) for 100 epochs with a learning rate of 0.001. For the Langevin method, we use the authors' implementation but with the GRO sampling mask described in Sec. 3.2.

The Score method is different than the other methods in that

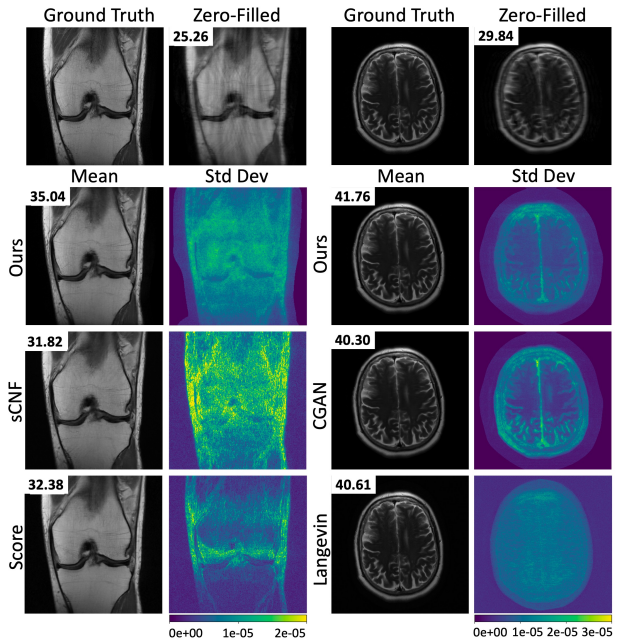

Figure 2. Mean images and pixel-wise standard-deviation maps computed from 8 and 32 posterior samples for the brain images and knee datasets, respectively. The standard-deviation maps show which pixels have the greatest reconstruction uncertainty. The corresponding PSNR is shown on each reconstruction.

it assumes that the k-space measurements $\boldsymbol{k}_c$ are constructed from true coil images $\boldsymbol{x}_{\text{true}}$ with magnitudes affinely normalized to the interval $[0, 1]$ and phases normalized to $[0, 1]$ radians. Although this normalization cannot be enforced on prospectively undersampled MRI data, Score fails without this normalization. So, to evaluate Score, we normalize each $\boldsymbol{k}_c$ using knowledge of the ground-truth $\boldsymbol{x}_c$, run Score, and un-normalized its output $\widehat{\boldsymbol{x}}_c$ for comparison with the other methods. Since the Score paper (Chung & Ye, 2022a) used RSS combining to compute $\widehat{\boldsymbol{i}}$, we do the same. For the Score method, we use $T = 200$ iterations and not the default value of $T = 2000$. This is because, when using posterior-sample averaging (see Sec. 3.5), the PSNR computed using 200 iterations is better than with 2000.

The sCNF method works only on single-coil magnitude data, and so we convert our multi-coil data to that domain in order to evaluate sCNF. To do this, we apply RSS (5) to ZF coil-images $\boldsymbol{y}$ and repeat the process for the true coil images $\boldsymbol{x}_{\text{true}}$. Using those magnitude images, we train sCNF for 300 epochs with learning rate 0.0005 and batch size 32.

3.5. Evaluation

We report results for several different metrics, including peak-signal-to-noise ratio (PSNR), structural-similarity index (SSIM) (Wang et al., 2004), Fréchet Inception Score (FID) (Heusel et al., 2017), and conditional FID (cFID) (Soloveitchik et al., 2021). PSNR and SSIM were computed on the average of P posterior samples $\{\boldsymbol{i}_p\}_{p=1}^{P}$, i.e.,

$$\widehat{\boldsymbol{i}}_{(P)} \triangleq \frac{1}{P}\sum_{p=1}^{P}\widehat{\boldsymbol{i}}_p \tag{19}$$

to approximate the posterior mean, while FID and cFID were evaluated on individual posterior samples $\widehat{\boldsymbol{i}}_p$. By default, we compute all metrics using magnitude reconstructions $|\widehat{\boldsymbol{i}}|$ rather than the complex-valued reconstructions $\widehat{\boldsymbol{i}}$, in part because competitors like sCNF generate only magnitude reconstructions, but also because this is typical in the MRI literature (e.g., the fastMRI competition (Zbontar et al., 2018)). So, for example, PSNR is computed as

$$\text{PSNR} \triangleq 10\log_{10}\left(\frac{D\max_d |[\boldsymbol{i}_{\text{true}}]_d|^2}{\big||\widehat{\boldsymbol{i}}_{(P)}| - |\boldsymbol{i}_{\text{true}}|\big|_2^2}\right), \tag{20}$$

where $[\cdot]_d$ extracts the dth pixel. For FID and cFID, we use the embeddings of VGG-16 (Simonyan & Zisserman, 2014) as (Kastryulin et al., 2022) found that this helped the metrics better correlate with the rankings of radiologists.

For the brain data, we compute all metrics on 72 random test images in order to limit the Langevin image generation time to 4 days. We generate complex-valued images using the coil-combining method in (4) and use $P = 32$ posterior samples to calculate cFID[1], FID[1], PSNR, and SSIM. (For the reference statistics of FID, we use the entire training dataset.) Because FID and cFID are biased by small sample sizes, we also compute FID[2] and cFID[2] with 2484 test samples and $P = 8$ for our method and the CGAN.

With the knee data, we follow a similar evaluation procedure except that, to comply with the evaluation steps of Score, we generate magnitude-only signals using the root-sum-of-square (RSS) combining from (5). Also, we computed metrics on 72 randomly selected slices in order to bound the image generation time of Score to 6 days with $P = 8$. We use $P = 8$ for all metrics, but for FID[2] and cFID[2], we use 2188 test samples.

When computing inference time for all methods, we use a single Nvidia V100 with 32GB of memory and evaluate the time required to generate one posterior sample.

4. Results

Table 1 reports the quantitative metrics for the knee dataset. It shows that our method outperforms sCNF by a significant margin in all metrics except inference time. By using information from multiple coils and a more advanced architecture, our method shows the true competitive potential of CNFs in realistic accelerated MR imaging.

Table 1 also shows that our method surpasses Score in all metrics except FID[1], even though Score benefited from im-

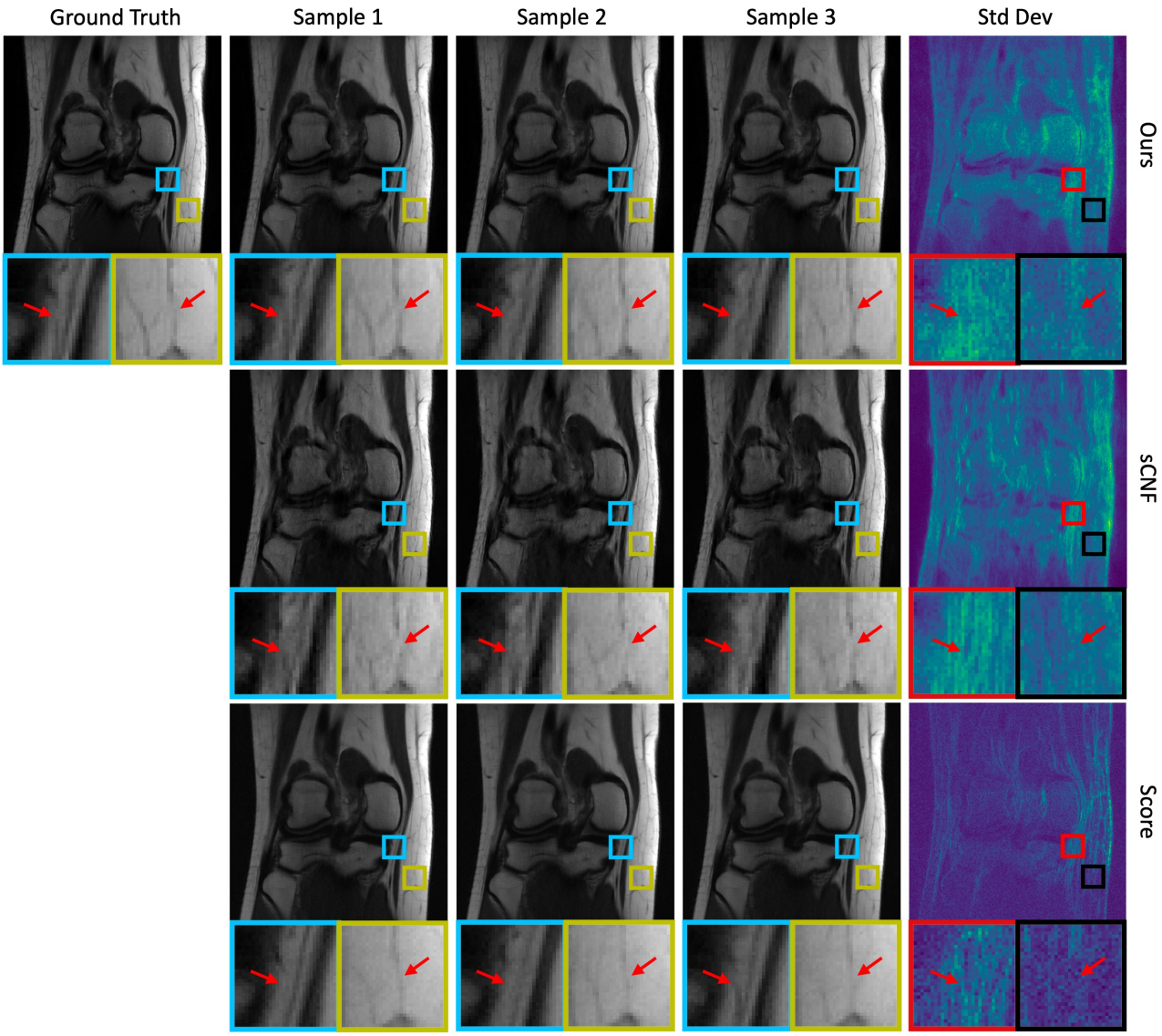

Figure 3. Examples of posterior samples and standard-deviation maps for the knee data. The samples show important structural variations. This demonstrates the advantages of generating multiple reconstructions and computing a pixel-wise standard-deviation map.

practical ground-truth normalization. Compared to Score, our method generated posterior samples $8000\times$ faster. Furthermore, our method (and sCNF) will see a speedup when multiple samples are generated because the conditioning network $g_{\boldsymbol{\theta}}$ needs to be evaluated only once per P generated samples for a given $\boldsymbol{y}$. For example, with the knee data, we are able to generate $P = 32$ samples in 1.41 seconds, corresponding to 44 milliseconds per sample, which is a $2.5\times$ speedup over the value reported in Table 1.

Table 2 reports the quantitative results for the brain dataset. The table shows that we outperform the Langevin and CGAN methods in all benchmarks except inference time. While our method is a bit slower than the CGAN, it is orders of magnitude faster than the Langevin approach.

We show the mean images and standard-deviation maps for the fastMRI knee and brain experiments in Fig. 2. For the knee data, our method captures texture more accurately than the sCNF method and provides a sharper representation than the Score method. All of the brain methods provide a visually accurate representation to the ground truth, but the Langevin method provides a more diffuse variance map, with energy spread throughout the image.

In Fig. 3, we plot multiple posterior samples, along with zoomed-in regions, to illustrate the changes across independently drawn samples for each method. The standard-deviation maps are generated using $P = 8$ posterior samples, three of which are shown. From the zoomed-in regions, it can be seen that several samples are consistent with the

Model	PSNR (dB)↑	SSIM↑	FID[1]↓	FID[2]↓	cFID[1]↓	cFID[2]↓	Time
Score	34.15 ± 0.19	0.8764 ± 0.0036	**4.49**	—	4.49	—	15 min
sCNF	32.93 ± 0.17	0.8494 ± 0.0047	7.32	5.78	8.49	6.51	**66 ms**
Ours	**35.23 ± 0.22**	**0.8888 ± 0.0046**	4.68	**2.55**	**3.96**	**2.44**	108 ms

Table 1. Average performance on non-fat-suppressed fastMRI knee data, with standard error reported after the ±. PSNR, SSIM, FID[1], and cFID[1] are computed for 72 test images and $P = 8$ posterior samples. FID[2], and cFID[2] are computed for 2188 test samples and $P = 8$ posterior samples. Time to the generation of one posterior sample.

Model	PSNR (dB)↑	SSIM↑	FID[1]↓	FID[2]↓	cFID[1]↓	cFID[2]↓	Time
Langevin	37.88 ± 0.41	0.9042 ± 0.0062	6.12	—	5.29	—	14 min
CGAN	37.28 ± 0.19	0.9413 ± 0.0031	5.38	4.06	6.41	4.28	**112 ms**
Ours	**38.85 ± 0.23**	**0.9495 ± 0.0012**	**4.13**	**2.37**	**4.15**	**2.44**	177 ms

Table 2. Average performance on non-fat-suppressed fastMRI brain data, with standard error reported after the ±. PSNR, SSIM, FID[1], and cFID[1] are computed for 72 test images and $P = 32$ posterior samples. FID[2] and cFID[2] are computed using 2484 test samples and $P = 8$. Time to the generation of one posterior sample.

ground truth while others are not (although they may be consistent with the measured data). Regions of high posterior variation can be flagged from visual inspection of the standard-deviation map and further investigated through viewing multiple posterior samples for improved clinical diagnoses.

Our method presents observable, realistic variations of small anatomical features in the zoomed-in regions. The variations are also registered in the standard-deviation map. Both the posterior samples and the standard-deviation map could be used by clinicians to assess their findings. Comparatively, our method demonstrates variation that is spread across the entire image, while in the Score method, the variation is mostly localized to small regions. Since it is difficult to say which standard-deviation map is more useful or correct, the interpretation of these maps could be an interesting direction for future work. The sCNF also demonstrates variation, but it is mostly driven by residual aliasing artifacts.

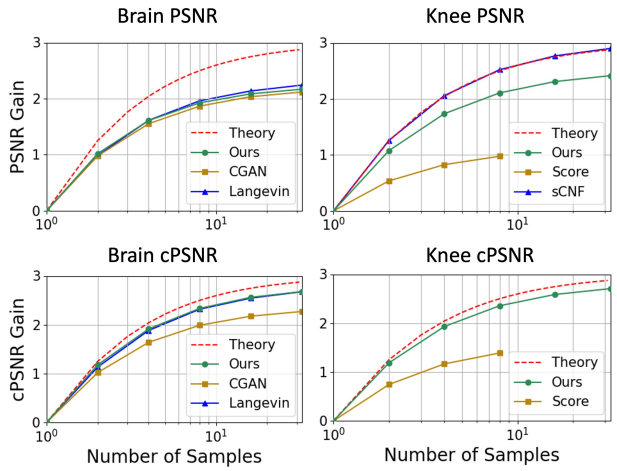

Figure 4. The gain in (magnitude) PSNR and complex PSNR of the P-sample mean estimate $\widehat{\boldsymbol{i}}_{(P)}$ versus P, for both brain and knee data. Note the ≈ 3 dB increase as P grows from 1 to infinity.

4.1. PSNR Gain versus P

It is well known that the minimum MSE (MMSE) estimate of $\boldsymbol{i}$ from $\boldsymbol{y}$ equals the conditional mean $\mathrm{E}\{\boldsymbol{i}|\boldsymbol{y}\}$, i.e., the mean of the posterior distribution $p_{i|y}(\cdot|\boldsymbol{y})$. Thus, one way to approximate the MMSE estimate is to generate many samples from the posterior distribution and average them, as in (19). Bendel et al. (2022) showed that the MSE

$$\mathcal{E}_P \triangleq \mathrm{E}\left[\|\widehat{\boldsymbol{i}}_{(P)} - \boldsymbol{i}_{\mathsf{true}}\|_2^2 \; \boldsymbol{y}\right] \tag{21}$$

of the P-posterior-sample average $\widehat{\boldsymbol{i}}_{(P)}$ obeys $\mathcal{E}_1/\mathcal{E}_P = 2P/(P+1)$. So, for example, the SNR increases by a factor of two as P grows from 1 to ∞. The same thing should happen for PSNR, as long as the PSNR definition is consistent with (21). For positive signals (i.e., magnitude images) the PSNR definition from (20) is consistent with (21), but for complex signals we must use "complex PSNR"

$$\mathrm{cPSNR} \triangleq 10\log_{10}\left(\frac{D\max_d |[\boldsymbol{i}_{\mathsf{true}}]_d|^2}{\|\widehat{\boldsymbol{i}}_{(P)} - \boldsymbol{i}_{\mathsf{true}}\|_2^2}\right). \tag{22}$$

As RSS combining provides only a magnitude estimate, we compute the coil-combined estimate for our method and Score to evaluate cPSNR behavior for the knee dataset.

One may then wonder whether a given approximate posterior sampler has a PSNR gain versus P that matches the theory. In Fig. 4, we answer this question by plotting the PSNR gain and the cPSNR gain versus $P \in \{1, 2, 4, 8, 16, 32\}$ for the various methods under test (averaged over all 72 test samples). There we see that our method's cPSNR curve matches the theoretical curve well for both brain and knee data. As expected, our (magnitude) PSNR curve does not match the theoretical curve. The cPSNR curves of the Score and CGAN methods fall short of the theoretical curve by

Model	PSNR (dB)↑	SSIM↑	FID^2↓	cFID^2↓
(Denker et al., 2021a)	17.61 ± 0.20	0.6665 ± 0.0072	16.02	16.68
+ Data Consistency	27.27 ± 0.21	0.7447 ± 0.0061	16.92	18.56
+ Architectural Changes	33.87 ± 0.23	0.8715 ± 0.0049	4.48	4.50
+ Nullspace Learning	**35.23 ± 0.22**	**0.8888 ± 0.0046**	**2.55**	**2.44**

Table 3. Ablation Study: Performance on non-fat-suppressed fastMRI knee data, with standard error reported after the ±. Each line adds a new contribution to the model of the previous line. Metrics are computed as described in Sec. 3.5

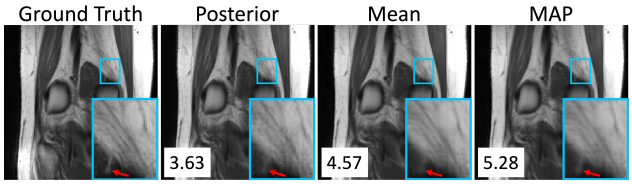

Figure 5. Examples of a ground-truth image, one posterior sample, an average of $P = 8$ posterior samples, and a MAP estimate. The log posterior density in units of bits-per-dimension is shown in the bottom right corner of each image.

a large margin, but interestingly, the Langevin method's cPSNR curve matches ours almost perfectly. sCNF's PSNR gain curve matches the theoretical one almost perfectly, which provides further empirical evidence that CNF methods accurately sample from the posterior distribution.

4.2. Ablation Study

To evaluate the impact of our contributions to CNF architecture and training design, we perform an ablation study using the fastMRI knee dataset. We start with the baseline model in (Denker et al., 2021a), modified to take in 16 channels instead of 1, and scale it up using the built-in mechanism in the author's code. We train this model for 300 epochs with batch size 32 and learning rate 0.0001 to minimize the NLL of the multicoil targets $\{\boldsymbol{x}^{(i)}\}$, since higher learning rates were numerically unstable. Table 3 shows what happens when we add each of our contributions. First, we add data consistency (15) to the evaluation of the baseline. We then add the architectural changes described in Sec. 3.1, and finally we add nullspace learning to arrive at our proposed method. From Table 3, it can be seen that each of our design contributions yielded a significant boost in performance, and that nullspace learning was a critical ingredient in our outperforming the Score method in Table 1. For this ablation study, all models were trained following the procedure outlined in Sec. 3.3 (except for the learning rate of the baseline).

4.3. Maximum a Posteriori (MAP) Estimation

Because CNFs can evaluate the posterior density of a signal hypothesis (recall (9)), they can be used for posteriori (MAP) estimation, unlike CGANs.

Due to our data-consistency step (15), we find the MAP estimate of $\boldsymbol{x}$ using

$$\widehat{\boldsymbol{x}}_{\mathsf{MAP}} = \widehat{\boldsymbol{u}}_{\mathsf{MAP}} + \boldsymbol{A}^{+}\boldsymbol{y} \tag{23}$$

$$\widehat{\boldsymbol{u}}_{\mathsf{MAP}} = \arg\max_{\boldsymbol{u}\in\mathrm{null}(\boldsymbol{A})} \ln \widehat{p}_{u|y}(\boldsymbol{u}|\boldsymbol{y}). \tag{24}$$

Note the CNF output $\boldsymbol{u}$ is constrained to the nullspace of $\boldsymbol{A}$. From (17), this nullspace is spanned by the columns of

$$\boldsymbol{W} \triangleq \mathrm{blkdiag}\left\{\boldsymbol{F}^{\mathsf{H}}\widetilde{\boldsymbol{P}}^{\top}, \ldots, \boldsymbol{F}^{\mathsf{H}}\widetilde{\boldsymbol{P}}^{\top}\right. , \tag{25}$$

which are orthonormal, and so $\widehat{\boldsymbol{u}}_{\mathsf{MAP}} = \boldsymbol{W}\widetilde{\boldsymbol{k}}_{\mathsf{MAP}}$ with

$$\widetilde{\boldsymbol{k}}_{\mathsf{MAP}} = \arg\max_{\widetilde{\boldsymbol{k}}} \ln \widehat{p}_{u|y}(\boldsymbol{W}\widetilde{\boldsymbol{k}}|\boldsymbol{y};\boldsymbol{\theta}) \tag{26}$$

$$= \arg\max_{\widetilde{\boldsymbol{k}}} \left[\ln p_z(\overline{h}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{W}\widetilde{\boldsymbol{k}}, \boldsymbol{y})) + \ln \det\left(\frac{\partial \overline{h}_{\boldsymbol{\theta}}^{-1}(\widetilde{\boldsymbol{u}}, \boldsymbol{y})}{\partial \widetilde{\boldsymbol{u}}} \Bigg|_{\widetilde{\boldsymbol{u}}=\boldsymbol{W}\widetilde{\boldsymbol{k}}} \right) \right]. \tag{27}$$

For this maximization, we use the Adam optimizer with 5000 iterations and a learning rate of 1×10^{-8}. Above, $\widetilde{\boldsymbol{k}}$ can be recognized as the unmeasured k-space samples.

In Figure 5, we show an example of a MAP estimate along with the ground truth image, one sample from the posterior, a $P = 8$ posterior-sample average, and their corresponding log-posterior-density values. As expected, the MAP estimate has a higher log-posterior-density that the other estimates. Visually, the MAP estimate is slightly sharper than the sample average but contains less texture details than the single posterior sample.

5. Conclusion

In this work, we present the first conditional normalizing flow for posterior sample generation in multi-coil accelerated MRI. To do this, we designed a novel conditional normalizing flow (CNF) that infers the signal component in the measurement operator's nullspace, whose outputs are later combined with information from the measured space. In experiments with fastMRI brain and knee data, we demonstrate improvements over existing posterior samplers for MRI. Compared to score/Langevin-based approaches, our inference time is four orders-of-magnitude faster. We also illustrate how the posterior samples can be used to quantify uncertainty in MR imaging. This provides radiologists with additional tools to enhance the robustness of clinical diagnoses. We hope this work motivates additional exploration of posterior sampling for accelerated MRI.

Acknowledgements

This work was supported in part by the National Institutes of Health under Grant R01-EB029957.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion, 76:243–297, 2021.

Adler, J. and Öktem, O. Deep Bayesian inversion. arXiv:1811.05910, 2018.

Ahmad, R., Bouman, C. A., Buzzard, G. T., Chan, S., Liu, S., Reehorst, E. T., and Schniter, P. Plug and play methods for magnetic resonance imaging. IEEE Signal Process. Mag., 37(1):105–116, March 2020.

Ardizzone, L., Bungert, T., Draxler, F., Köthe, U., Kruse, J., Schmier, R., and Sorrenson, P. Framework for Easily Invertible Architectures (FrEIA). https://github.com/vislearn/FrEIA, 2018. Accessed: 2022-11-05.

Ardizzone, L., Lüth, C., Kruse, J., Rother, C., and Köthe, U. Guided image generation with conditional invertible neural networks. arXiv:1907.02392, 2019.

Ardizzone, L., Kruse, J., Lüth, C., Bracher, N., Rother, C., and Köthe, U. Conditional invertible neural networks for diverse image-to-image translation. arXiv:2105.02104, 2021.

Bendel, M., Ahmad, R., and Schniter, P. A regularized conditional GAN for posterior sampling in inverse problems. arXiv:2210.13389, 2022.

Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. In Proc. Int. Conf. Mach. Learn., pp. 537–546, 2017.

Chang, C.-H., Creager, E., Goldenberg, A., and Duvenaud, D. Explaining image classifiers by counterfactual generation. In Proc. Int. Conf. on Learn. Rep., 2019.

Chen, D. and Davies, M. E. Deep decomposition learning for inverse imaging problems. In Proc. European Conf. Comp. Vision, pp. 510–526, 2020.

Chung, H. and Ye, J. C. Score-based diffusion models for accelerated MRI. Med. Image Analysis, 80:102479, 2022a.

Chung, H. and Ye, J. C. score-MRI. https://github.com/HJ-harry/score-MRI, 2022b. Accessed: 2022-12-15.

Denker, A., Schmidt, M., Leuschner, J., and Maass, P. Conditional invertible neural networks for medical imaging. J. Imaging, 7(11):243, 2021a.

Denker, A., Schmidt, M., Leuschner, J., and Maass, P. cinn-for-imaging. https://github.com/jleuschn/cinn_for_imaging, 2021b. Accessed: 2022-10-08.

Dinh, L., Krueger, D., and Bengio, Y. NICE: Non-linear independent components estimation. In Proc. Int. Conf. on Learn. Rep. Workshops, 2015.

Dinh, L., Sohl-Dickstein, J., and Bengio, S. Density estimation using Real NVP. In Proc. Int. Conf. on Learn. Rep., 2017.

Edupuganti, V., Mardani, M., Vasanawala, S., and Pauly, J. Uncertainty quantification in deep MRI reconstruction. IEEE Trans. Med. Imag., 40(1):239–250, January 2021.

Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.-J., and Hwang, D. KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn. Reson. Med., 80(5):2188–2201, 2018.

Falcon, W. et al. Pytorch lightning, 2019. URL https://github.com/PyTorchLightning/pytorch-lightning.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., and Haase, A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med., 47(6):1202–1210, 2002.

Hammernik, K., Klatzer, T., Kobler, E., Recht, M. P., Sodickson, D. K., Pock, T., and Knoll, F. Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med., 79(6):3055–3071, 2018.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In Proc. Neural Inf. Process. Syst. Conf., volume 30, 2017.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In Proc. Neural Inf. Process. Syst. Conf., volume 33, pp. 6840–6851, 2020.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. In Proc. IEEE Conf. Comp. Vision Pattern Recog., pp. 1125–1134, 2017.

Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A., and Tamir, J. Robust compressed sensing MRI with deep generative priors. In Proc. Neural Inf. Process. Syst. Conf., 2021a.

Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A., and Tamir, J. csgm-mri-langevin. https://github.com/utcsilab/csgm-mri-langevin, 2021b. Accessed: 2021-12-05.

Joshi, M., Pruitt, A., Chen, C., Liu, Y., and Ahmad, R. Technical report (v1.0)–pseudo-random cartesian sampling for dynamic MRI. arXiv:2206.03630, 2022.

Kadkhodaie, Z. and Simoncelli, E. P. Solving linear inverse problems using the prior implicit in a denoiser. arXiv:2007.13640, 2020.

Kastryulin, S., Zakirov, J., Pezzotti, N., and Dylov, D. V. Image quality assessment for magnetic resonance imaging. arXiv:2203.07809, 2022.

Kim, Y. and Son, D. Noise conditional flow model for learning the super-resolution space. arXiv:1606.02838, 2021.

Kingma, D. and Dhariwal, P. Glow: Generaetive flow with invertible 1x1 convolutions. In Proc. Neural Inf. Process. Syst. Conf., pp. 10236–10245, 2018.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In Proc. Int. Conf. on Learn. Rep., 2015.

Knoll, F., Hammernik, K., Zhang, C., Moeller, S., Pock, T., Sodickson, D. K., and Akcakaya, M. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. IEEE Signal Process. Mag., 37(1):128–140, January 2020.

Laumont, R., Bortoli, V. D., Almansa, A., Delon, J., Durmus, A., and Pereyra, M. Bayesian imaging using plug & play priors: When Langevin meets Tweedie. SIAM J. Imag. Sci., 15(2):701–737, 2022.

Lugmayr, A., Danelljan, M., Van Gool, L., and Timofte, R. SRFlow: Learning the super-resolution space with normalizing flow. In Proc. European Conf. Comp. Vision, 2020.

Lugmayr, A., Danelljan, M., Timofte, R., Kim, K.-w., Kim, Y., Lee, J.-y., Li, Z., Pan, J., Shim, D., Song, K.-U., et al. NTIRE 2022 challenge on learning the super-resolution space. In Proc. IEEE Conf. Comp. Vision Pattern Recog., pp. 786–797, 2022.

Lustig, M., Donoho, D., and Pauly, J. M. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn. Reson. Med., 58(6):1182–1195, 2007.

Ong, F. and Lustig, M. SigPy: A python package for high performance iterative reconstruction. In Proc. Annu. Meeting ISMRM, volume 4819, 2019.

Papamakarios, G., Nalisnick, E. T., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. J. Mach. Learn. Res., 22(57):1–64, 2021.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. PyTorch: An imperative style, high-performance deep learning library. In Proc. Neural Inf. Process. Syst. Conf., pp. 8024–8035, 2019.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. SENSE: Sensitivity encoding for fast MRI. Magn. Reson. Med., 42(5):952–962, 1999.

Repetti, A., Pereyra, M., and Wiaux, Y. Scalable bayesian uncertainty quantification in imaging inverse problems via convex optimization. SIAM J. Imag. Sci., 12(1):87–118, 2019.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proc. Intl. Conf. Med. Image Comput. Comput. Assist. Intervent., pp. 234–241, 2015.

Sanchez, T., Krawczuk, I., Sun, Z., and Cevher, V. Uncertainty-driven adaptive sampling via GANs. In Proc. Neural Inf. Process. Syst. Workshop, 2020.

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.

Soloveitchik, M., Diskin, T., Morin, E., and Wiesel, A. Conditional Frechet inception distance. arXiv:2103.11521, 2021.

Sønderby, C. K., Caballero, J., Theis, L., Shi, W., and Huszár, F. Amortised MAP inference for image super-resolution. In Proc. Int. Conf. on Learn. Rep., 2017.

Song, K.-U., Shim, D., Kim, K.-w., Lee, J.-y., and Kim, Y. FS-NCSR: Increasing diversity of the super-resolution space via frequency separation and noise-conditioned normalizing flow. In Proc. IEEE Conf. Comp. Vision Pattern Recog. Workshop, pp. 968–977, June 2022.

Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C. L., Yakubova, N., Knoll, F., and Johnson, P. End-to-end variational networks for accelerated MRI reconstruction. In Proc. Intl. Conf. Med. Image Comput. Comput. Assist. Intervent., pp. 64–73, 2020.

Tonolini, F., Radford, J., Turpin, A., Faccio, D., and Murray-Smith, R. Variational inference for computational imaging inverse problems. J. Mach. Learn. Res., 21(179): 1–46, 2020.

Uecker, M., Lai, P., Murphy, M. J., Virtue, P., Elad, M., Pauly, J. M., Vasanawala, S. S., and Lustig, M. ESPIRiT–an eigenvalue approach to autocalibrating parallel MRI:

Where SENSE meets GRAPPA. Magn. Reson. Med., 71 (3):990–1001, 2014.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process., 13(4): 600–612, Apr. 2004.

Winkler, C., Worrall, D., Hoogeboom, E., and Welling, M. Learning likelihoods with conditional normalizing flows. arXiv preprint arXiv:1912.00042, 2019.

Zbontar, J., Knoll, F., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., Parente, M., Geras, K. J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M., Vincent, P., Pinkerton, J., Wang, D., Yakubova, N., Owens, E., Zitnick, C. L., Recht, M. P., Sodickson, D. K., and Lui, Y. W. fastMRI: An open dataset and benchmarks for accelerated MRI. arXiv:1811.08839, 2018.

Zhang, T., Pauly, J. M., Vasanawala, S. S., and Lustig, M. Coil compression for accelerated imaging with Cartesian sampling. Magn. Reson. Med., 69(2):571–582, 2013.

A. Accelerated MRI Simulation Procedure

We outline the procedure for simulating the accelerated MRI problem. The fastMRI datasets provide the fully sampled multi-coil k-space, i.e., $\{\boldsymbol{k}_c\}_{c=1}^C$ with $M = D$. To obtain the ground truth coil-images $\{\boldsymbol{x}_c\}_{c=1}^C$, we take the inverse Fourier transform of the fully sampled k-space measurement, i.e., $\boldsymbol{x}_c = \boldsymbol{F}^{\mathsf{H}}\boldsymbol{k}_c$, wherein we assume that the noise $\boldsymbol{\epsilon}_c$ in (1) is negligible. To obtain the zero-filled images $\boldsymbol{y}_c$, we take the inverse Fourier transform after masking the fully-sampled k-space measurement $\boldsymbol{k}_c$, i.e., $\boldsymbol{y}_c = \boldsymbol{F}^{\mathsf{H}}\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{k}_c$. This procedure is illustrated in Fig. 6. In real-world accelerated MRI, the data acquisition process would collect masked k-space $\boldsymbol{P}\boldsymbol{k}_c$ directly.

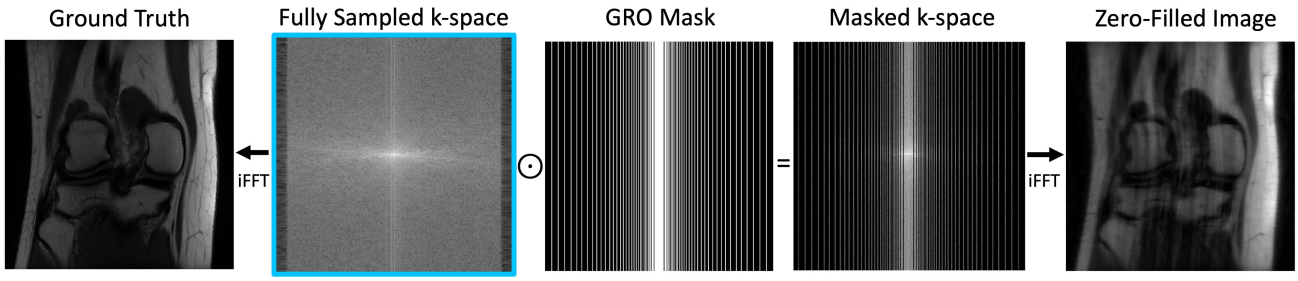

Figure 6. A visual illustration of simulating accelerated MRI. Given the fully sampled k-space $\boldsymbol{k}_c$ highlighted in blue, we obtain the ground truth $\boldsymbol{x}_c$ by applying the inverse Fourier transform $\boldsymbol{F}^{\mathsf{H}}$. The zero-filled image $\boldsymbol{y}_c$ is acquired by applying the sampling mask $\boldsymbol{P}^\top\boldsymbol{P}$ to fully sampled $\boldsymbol{k}_c$ and then taking the inverse Fourier Transform $\boldsymbol{F}^{\mathsf{H}}$.

B. Implementation Details

For our machine learning framework, we use PyTorch (Paszke et al., 2019) and PyTorch lightning (Falcon et al., 2019). To implement the components of the CNF, we use the Framework for Easily Invertible Architectures (FrEIA) (Ardizzone et al., 2018). For the Score, sCNF, and Langevin methods, we utilize the authors' implementations at (Chung & Ye, 2022b), (Denker et al., 2021b), and (Jalal et al., 2021b), respectively. ESPIRiT coil-estimation and coil-combining are implemented using the SigPy package (Ong & Lustig, 2019).

C. Brain Posterior Samples

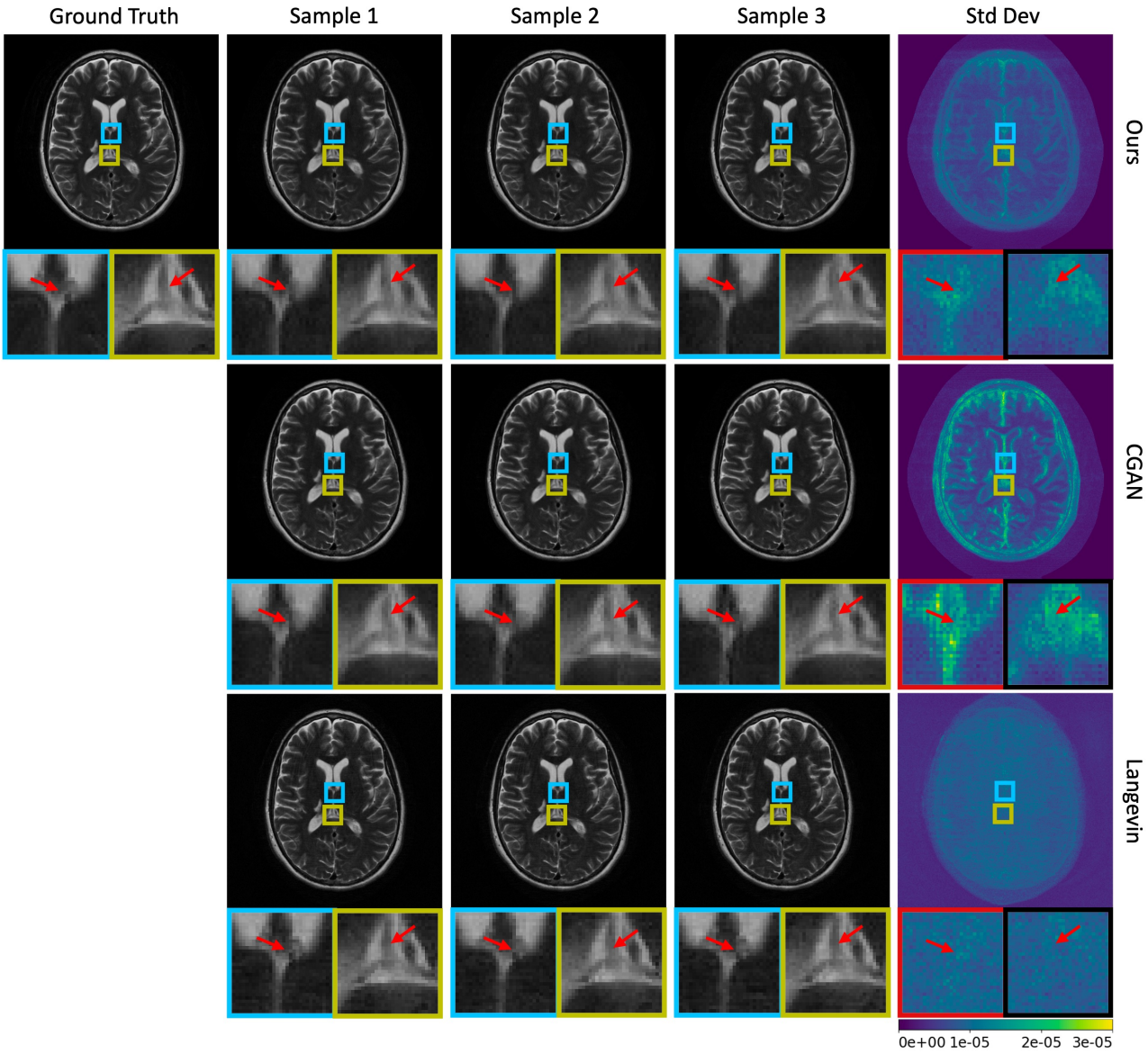

Figure 7. Examples of posterior samples and standard-deviation maps for the brain images, both with zoomed regions.