

# Accelerated Magnetic Resonance Imaging with Flow-Based Priors

MASTER'S THESIS

by  
Frederik Fraaz

supervised by  
Prof. Dr. Reinhard Heckel

## Abstract

Magnetic Resonance Imaging (MRI) is a powerful tool in clinical practice. But obtaining high quality images is a slow and costly process. Thus, an important direction of research is to reduce the acquisition time of MRI. There are two main ideas for reducing acquisition times: Undersampling the frequency domain representation of the measurement and acquiring multiple views in parallel (multi-coil MRI). Convolutional neural networks trained end-to-end achieve state-of-the-art image quality for accelerated multi-coil MRI. But these methods make assumptions about the undersampling operator during training and thus require retraining when the undersampling operator changes. This thesis uses a more flexible approach for accelerated multi-coil MRI that consists of first training a flow-based generator on image patches and then imposing it as a prior in the reconstruction. We evaluate the in-distribution performance of our method on the fastMRI multi-coil knee dataset and measure its robustness to distribution shifts on other MRI datasets. Our method achieves slightly better reconstruction quality than the best un-trained methods and slightly worse performance than neural networks trained end-to-end.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Normalizing Flow</b>	<b>5</b>
2.1	Change of variables . . . . .	5
2.2	Architecture . . . . .	6
2.2.1	Step of flow . . . . .	6
2.2.2	Multi-scale architecture . . . . .	7
2.3	Loss function . . . . .	8
<b>3</b>	<b>Accelerated Magnetic Resonance Imaging</b>	<b>8</b>
<b>4</b>	<b>Approach</b>	<b>9</b>
4.1	Preliminaries . . . . .	9
4.2	Prior-based image recovery . . . . .	10
<b>5</b>	<b>Experiments</b>	<b>11</b>
5.1	Training the normalizing flow . . . . .	11
5.1.1	Combining coil images . . . . .	12
5.1.2	Preprocessing . . . . .	12
5.1.3	Training details . . . . .	13
5.2	Reconstruction details . . . . .	14
5.3	Metrics . . . . .	15
5.4	In-distribution performance . . . . .	15
5.5	Robustness to distribution shifts . . . . .	17
5.5.1	Dataset shift . . . . .	17
5.5.2	Anatomy shift . . . . .	17
5.5.3	Adversarially-filtered shift . . . . .	18
5.6	Ablation study . . . . .	19
5.6.1	Architecture . . . . .	20
5.6.2	Regularization . . . . .	22
<b>6</b>	<b>Conclusion</b>	<b>23</b>
<b>A</b>	<b>Sample Reconstructions</b>	<b>30</b>

# 1 Introduction

Magnetic Resonance Imaging (MRI) is a powerful tool in clinical practice. But obtaining high quality images is a slow and costly process. Thus, an important direction of research is to reduce the acquisition time of MRI. There are two main ideas for reducing acquisition times: (1) Reduce the number of measurements, resulting in an undersampled frequency domain representation [Can06; LDP07] and (2) acquire multiple views of the body in parallel (multi-coil MRI) [Gri+02; Pru+99; SM97].

There exists a large amount of work focused on the task of obtaining high quality images from measurements where both of these ideas are combined, i.e., measurements consist of undersampled data from multiple coils. Traditionally, the problem of reconstructing an image from undersampled measurements has been solved by enforcing sparsity in some transform domain [Can06; LDP07].

More recently, convolutional neural networks (CNNs) trained end-to-end on a large dataset of MR images [Zbo+19] have shown superior performance [PW19; Sri+20]. But these methods make assumptions about the undersampling operator during training and thus require retraining when the sampling pattern or acceleration factor changes. A more flexible approach using un-trained CNNs has shown promising results as well, while not requiring training data [DH21; Hec19].

Bora et al. [Bor+17] introduced an algorithm for imposing a trained generative model as a prior in compressed sensing problems. Bora et al. [Bor+17] used generative adversarial networks (GANs) [Goo+14] and variational autoencoders (VAEs) [KW14] as image priors. GAN-based priors have shown promising results in the context of MRI [BZA20; KA21; Nar+19]. But GAN-based methods are compute intensive as they often require multiple independent restarts to achieve good reconstruction performance. Additionally, GAN-based methods usually rely on early stopping to avoid overfitting to the undersampled measurements [KBA21]. In a very recent work, Jalal et al. [Jal+21] found score-based generative models [SE20] to perform well as priors for MRI reconstruction.

In parallel, Dinh et al. [DKB15; DSB17] proposed flow-based generators. Similar to GANs and VAEs, flow-based generators model high-dimensional data from samples. Unlike GANs and VAEs, flow-based generators enable exact and efficient log-likelihood evaluation. They can thus be trained using maximum likelihood. Furthermore, flow-based generators are designed to be invertible and can be used for efficient image synthesis [DKB15].

Recently, flow-based generators have been used as priors for compressed

sensing problems [Asi+20; Hel+20]. Kelkar et al. [KBA21] were first to apply flow-based generators to accelerated MRI. Kelkar et al. [KBA21] obtain promising results for single-coil MRI, but they do not address the clinically more relevant problem of multi-coil MRI.

Our work applies flow-based priors to the problem of accelerated multi-coil MRI. The contributions of this thesis are as follows:

- We train a flow-based generator on image patches and then impose it as a prior for reconstructing an image from undersampled multi-coil MRI measurements. In contrast to end-to-end trained methods, we do not make any assumptions about the undersampling operator during training. Hence, a single trained prior can be used for different sampling patterns and undersampling ratios. Since we recover the image patch-wise, our approach can reconstruct images of arbitrary shapes. Anirudh et al. [ALT20] proposed patch-based priors in the context of GANs and Helminger et al. [Hel+20] applied the idea to flow-based generators.
- We evaluate the in-distribution performance of our approach on the fastMRI [Zbo+19] multi-coil knee dataset. Specifically, we train a flow-based generator on the training set and reconstruct undersampled measurements from the validation set. We compare the performance to trained methods (VarNet [Sri+20] and U-Net [RFB15]) as well as un-trained methods (ConvDecoder [DH21] and total variation (TV) regularization [BUF07]). Our method performs significantly better than TV and slightly better than ConvDecoder, but is outperformed by VarNet and U-Net.
- We measure the robustness of our prior-based method to distribution shifts. Following Darestani et al. [DCH21], we train our method on the fastMRI knee dataset and consider three shifts: (1) The dataset shift to the Stanford knee dataset [DCH21] (2) the anatomy shift to the fastMRI brain dataset and (3) the adversarially-filtered shift to the fastMRI-A dataset [DCH21]. Our method loses a similar amount of performance under these distribution shifts as trained and un-trained methods.

The remainder of this thesis is organized as follows: Section 2 introduces flow-based generators and Section 3 explains the basics of accelerated MRI. Section 4 covers our approach for prior-based image recovery. Section 5 evaluates the performance of our method and Section 6 concludes the thesis.

## 2 Normalizing Flow

Normalizing flows map data points to latent variables using a series of invertible nonlinear transformations. Their architecture enables efficient inference and image synthesis [DKB15].

More formally, normalizing flows are bijective mappings  $\mathbf{f}_\theta : X \rightarrow Z$  (with  $\mathbf{g}_\theta = \mathbf{f}_\theta^{-1}$ ) that map a random variable  $X \in \mathbb{R}^n$  to a latent variable  $Z \in \mathbb{R}^n$  [DSB17]. The parameters  $\theta$  of  $\mathbf{f}_\theta$  and  $\mathbf{g}_\theta$  are learned using maximum likelihood. We omit the subscript  $\theta$  in the remainder of this thesis. We use the terms normalizing flow, flow-based generator, generative flow and invertible neural network (INN) interchangeably.

### 2.1 Change of variables

The change of variables formula defines a relationship between the probability distributions  $p_X(\mathbf{x})$  and  $p_Z(\mathbf{z})$  [HKO01]:

$$p_X(\mathbf{x}) = p_Z(\mathbf{z}) |\det \mathbf{J}_f(\mathbf{z})|, \quad (1)$$

where  $\mathbf{J}_f$  is the Jacobian of  $f$ . We choose a tractable density for  $p_Z$ , a typical choice is the multivariate Gaussian distribution. Hence, the normalizing flow transforms the data distribution into a simpler latent distribution [DKB15].

In order to generate samples from the model, we sample latents  $\mathbf{z} \sim p_Z$  and pass them through the generator:

$$\mathbf{x} = \mathbf{g}(\mathbf{z}). \quad (2)$$

In practice,  $f$  is composed of a series of transformations:  $\mathbf{f} = \mathbf{f}_1 \circ \mathbf{f}_2 \cdots \circ \mathbf{f}_K$  [RM15]. We can thus express the relationship between an image  $\mathbf{x}$  and its latent representation  $\mathbf{z}$  as [KD18]:

$$\mathbf{x} \xleftarrow{\mathbf{f}_1} \mathbf{h}_1 \xleftarrow{\mathbf{f}_2} \mathbf{h}_2 \cdots \xleftarrow{\mathbf{f}_K} \mathbf{z}. \quad (3)$$

Using this relationship, the change of variables formula can be decomposed into individual transformations:

$$p_X(\mathbf{x}) = p_Z(\mathbf{z}) \prod_{k=1}^K |\det \mathbf{J}_{\mathbf{f}_k}(\mathbf{z})| \quad (4)$$

$$\log p_X(\mathbf{x}) = \log p_Z(\mathbf{z}) + \sum_{k=1}^K \log |\det \mathbf{J}_{\mathbf{f}_k}(\mathbf{z})|. \quad (5)$$

By choosing transformations  $\mathbf{f}_k$  whose Jacobians  $\mathbf{J}_{\mathbf{f}_k}$  are triangular matrices, the terms  $\log |\det \mathbf{J}_{\mathbf{f}_k}(\mathbf{z})|$  in (5) can be computed efficiently [DSB17].

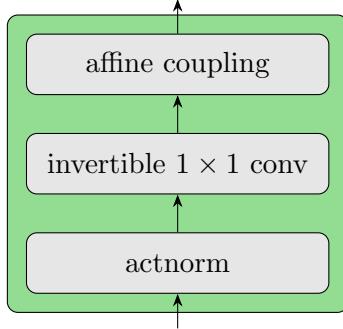


Figure 1: One step of flow (see [KD18]).

## 2.2 Architecture

We use Glow, an architecture for generative flows introduced by Kingma and Dhariwal [KD18], since it has a simple architecture and achieves good image quality on standard benchmarks for generative models.

### 2.2.1 Step of flow

The elementary building block of the Glow architecture, the step of flow (see Figure 1), consists of three components [KD18]:

**Actnorm** This affine transformation, introduced in [KD18], has an effect similar to batch normalization [IS15]. We refer the reader to [KD18] for details.

**Invertible  $1 \times 1$  convolution** This transformation performs a pixel-wise linear combination of channels [KD18]:

$$\mathbf{v} = \mathbf{x}\mathbf{U}, \quad (6)$$

where  $\mathbf{x}$  is the input tensor of shape  $H \times W \times C$ .  $H$  and  $W$  denote the height and width of the image and  $C$  is the number of channels.  $\mathbf{U}$  is initialized as a random  $C \times C$  rotation matrix and updated using gradient descent. The invertible  $1 \times 1$  convolution allows the affine coupling layers (explained next) in different flow steps to be applied to different channels [KD18].

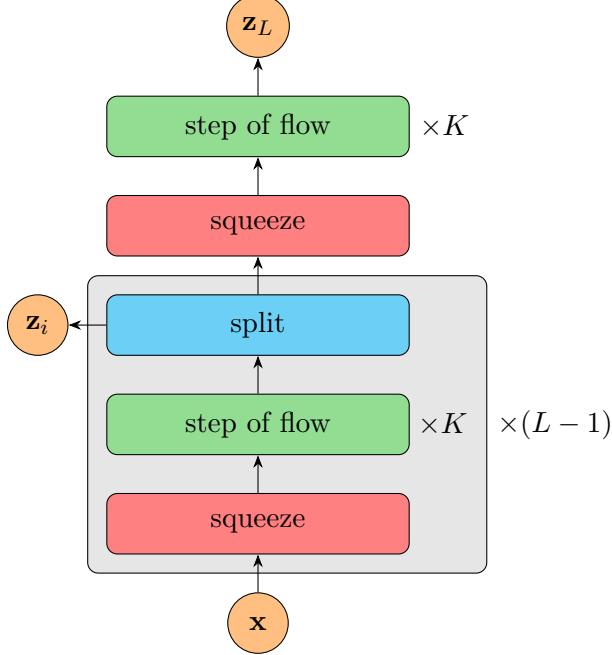


Figure 2: Multi-scale architecture (see [DSB17; KD18]).

**Affine coupling layer** This layer splits the input into two halves along the channel dimension and uses a learned transformation based on the first half to modify the second half [DSB17]:

$$\mathbf{v}_{1:c} = \mathbf{x}_{1:c} \quad (7)$$

$$\mathbf{v}_{c+1:C} = \mathbf{x}_{c+1:C} \odot \exp(\mathbf{s}(\mathbf{x}_{1:c})) + \mathbf{t}(\mathbf{x}_{1:c}), \quad (8)$$

where  $c < C$  and  $\odot$  denotes the Hadamard product. Since the affine coupling transforms only part of the input, it has a triangular Jacobian whose determinant does not depend on the Jacobian of  $\mathbf{s}$  or  $\mathbf{t}$ . Hence,  $\mathbf{s}$  and  $\mathbf{t}$  can be arbitrarily complex and are implemented using convolutional neural networks [DSB17].

### 2.2.2 Multi-scale architecture

Dinh et al. [DSB17] introduced the squeeze operation to reduce the spatial size while increasing the number of channels. This operation transforms a tensor of shape  $H \times W \times C$  into a  $\frac{H}{2} \times \frac{W}{2} \times 4C$  tensor [DSB17].

Figure 2 shows the multi-scale architecture. The model consists of  $L$  levels, each of which consists of a squeeze operation, followed by  $K$  steps of

flow and a split operation. The model factors out half of the channels at each level [DSB17]:

$$\mathbf{h}_0 = \mathbf{x} \quad (9)$$

$$(\mathbf{z}_{i+1}, \mathbf{h}_{i+1}) = \mathbf{f}_{i+1}(\mathbf{h}_i), \quad i = 0, \dots, L-2 \quad (10)$$

$$\mathbf{z}_L = \mathbf{f}_L(\mathbf{h}_{L-1}) \quad (11)$$

$$\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_L). \quad (12)$$

The  $\mathbf{z}_i$  from all levels are directly modeled as Gaussian random variables. Apart from reducing the computational and memory cost, the multi-scale architecture leads to intermediate levels of representation and distributes the loss throughout the network [DSB17].

### 2.3 Loss function

Given a dataset  $\mathcal{D}$  of  $N$  images, the generative flow is trained by minimizing the negative log-likelihood [DKB15]:

$$\mathcal{L}_{\text{train}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N -\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}). \quad (13)$$

As we saw in Section 2.1, we can compute the term  $\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})$  efficiently using the change of variables formula.

## 3 Accelerated Magnetic Resonance Imaging

In magnetic resonance imaging, measurements of an image  $\mathbf{x} \in \mathbb{C}^n$  are taken in the frequency domain:

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \text{noise}, \quad (14)$$

where  $\mathbf{F}$  is the 2D discrete Fourier transform matrix. This representation  $\mathbf{y} \in \mathbb{C}^n$  in the frequency domain is also known as k-space and directly measured by a so-called receiver coil. For a given k-space  $\mathbf{y}$ , we thus obtain an estimate of the spatial image  $\mathbf{x}$  by taking the inverse Fourier transform:

$$\hat{\mathbf{x}} = \mathbf{F}^{-1}\mathbf{y}. \quad (15)$$

In practice, multiple receiver coils are used. Each coil acquires a k-space measurement of the image:

$$\mathbf{y}_i = \mathbf{F}\mathbf{S}_i\mathbf{x} + \text{noise}, \quad i = 1, \dots, n_c, \quad (16)$$

where  $n_c$  is the number of receiver coils and  $\mathbf{S}_i$  is a complex-valued position-dependent coil sensitivity map. The multiplication of  $\mathbf{S}_i$  with  $\mathbf{x}$  is entry-wise. Denote by  $\mathbf{y}$  a vector that aggregates the measurements  $\mathbf{y}_i$  from all  $n_c$  coils.

Given the individual coil images  $\hat{\mathbf{x}}_i = \mathbf{F}^{-1}\mathbf{y}_i$ , root-sum-of-squares (RSS) [Roe+90] is a good estimate of the magnitude of the combined image:

$$\hat{\mathbf{x}} = \text{RSS}(\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{n_c}) = \sqrt{\sum_{i=1}^{n_c} |\hat{\mathbf{x}}_i|^2}. \quad (17)$$

In accelerated MRI, the k-space is undersampled, which leads to the forward operator:

$$\mathbf{y}_i = \mathbf{MFS}_i \mathbf{x} + \text{noise}, \quad i = 1, \dots, n_c, \quad (18)$$

where  $\mathbf{M}$  is a mask that sets some frequencies to zero. Hence, accelerated MRI poses the problem of recovering the image  $\mathbf{x}$  from  $n_c$  undersampled k-space measurements.

## 4 Approach

Our approach for accelerated MRI consists of two steps. First, we train a generative flow on patches of fully sampled MR images. Then, we use the trained model as a prior for recovering an image from undersampled k-space measurements.

### 4.1 Preliminaries

There are two main motivations for training the prior on patches rather than full images. First, training on patches reduces the computational and memory cost. Second, a patch-based prior allows us to recover images of arbitrary shapes [ALT20]. Next, we introduce some useful notation for working with patches.

The *patch* operator ( $\mathcal{P}$ ) takes an image  $\mathbf{x}$  as input and returns  $n_p$  non-overlapping patches [ALT20]:

$$(\mathbf{x}_1, \dots, \mathbf{x}_{n_p}) = \mathcal{P}(\mathbf{x}). \quad (19)$$

Its inverse, the *compose* operator ( $\mathcal{C} = \mathcal{P}^{-1}$ ), combines  $n_p$  patches into one image:

$$\mathbf{x} = \mathcal{C}(\mathbf{x}_1, \dots, \mathbf{x}_{n_p}). \quad (20)$$

Assuming a trained flow-based generator  $\mathbf{g}$ , the *generative compose* operator ( $\mathcal{T}$ ) takes  $n_p$  latent vectors  $(\mathbf{z}_1, \dots, \mathbf{z}_{n_p})$  as input and generates an image [ALT20]:

$$\mathbf{x} = \mathcal{T}(\mathbf{z}_1, \dots, \mathbf{z}_{n_p}) = \mathcal{C}(\mathbf{g}(\mathbf{z}_1), \dots, \mathbf{g}(\mathbf{z}_{n_p})), \quad (21)$$

i.e.,  $\mathcal{T}$  passes the latent vectors through  $\mathbf{g}$  to obtain image patches and then applies the compose operator  $\mathcal{C}$  to combine the patches into a single image. Note that only the latent vectors vary from patch to patch, while the generator  $\mathbf{g}$  has the same parameters across all patches. We can thus implement  $\mathcal{T}$  efficiently by passing the  $n_p$  latent vectors through  $\mathbf{g}$  as a single batch  $\tilde{\mathbf{z}} = [\mathbf{z}_1, \dots, \mathbf{z}_{n_p}]^T$ .

## 4.2 Prior-based image recovery

We recover an image from the undersampled k-space  $\mathbf{y}$  by minimizing:

$$\mathcal{L}(\tilde{\mathbf{z}}) = \sum_{i=1}^{n_c} \left\| \mathbf{y}_i - \mathbf{M} \mathbf{F} \hat{\mathbf{S}}_i \mathcal{T}(\tilde{\mathbf{z}}) \right\|^2 + \gamma \|\tilde{\mathbf{z}}\|^2, \quad (22)$$

where  $\|\cdot\|$  is the Euclidean norm and  $\gamma$  is a hyperparameter. We estimate  $\hat{\mathbf{S}}_i$  from  $\mathbf{y}$  using ESPIRiT [Uec+14]. Our formulation of the loss (22) is based on [Asi+20; Bor+17; DH21; KBA21]. Note that our prior generates only a single image  $\hat{\mathbf{x}}_{\mathbf{g}} = \mathcal{T}(\tilde{\mathbf{z}})$ , which is then multiplied with  $\hat{\mathbf{S}}_i$  to obtain the individual coil images.

The first term in (22) enforces data consistency with the known frequencies in  $\mathbf{y}$ . We regularize the optimization problem implicitly by initializing  $\tilde{\mathbf{z}}$  to zero, i.e., we start with the most likely image under our prior and only make steps toward less likely images that improve data consistency with the measured k-space  $\mathbf{y}$  [Asi+20].

Additionally, we explicitly penalize latent representations with a large Euclidean norm through the second term in (22). The intuition behind this regularization strategy is that, assuming a Gaussian latent distribution, latents with smaller Euclidean norm correspond to more likely images [Asi+20]. We minimize (22) using an iterative first order method.

**Enforcing data consistency** After taking  $n_{\text{iter}}$  steps with respect to the loss (22), our estimate of the image is  $\hat{\mathbf{x}}_{\mathbf{g}} = \mathcal{T}(\tilde{\mathbf{z}})$ . We can improve this estimate by replacing its frequencies with those of the undersampled k-space  $\mathbf{y}$

where the mask  $\mathbf{M}$  is nonzero [DH21]. Formally, we can express this operation as [ZFZ19]:

$$\hat{\mathbf{x}}_{i,\text{dc}} = \mathbf{F}^{-1} \left( \mathbf{y}_i + (\mathbf{1} - \mathbf{M}) \mathbf{F} \hat{\mathbf{S}}_i \hat{\mathbf{x}}_{\mathbf{g}} \right), \quad i = 1, \dots, n_c, \quad (23)$$

where  $\mathbf{1}$  is an all-ones matrix with same shape as  $\mathbf{M}$ .

We obtain the final result by combining the individual coil images into a single real image using RSS:

$$\hat{\mathbf{x}}_{\text{dc}} = \sqrt{\sum_{i=1}^{n_c} |\hat{\mathbf{x}}_{i,\text{dc}}|^2}. \quad (24)$$

## 5 Experiments

In this section, we evaluate our method for prior-based image recovery from undersampled multi-coil MRI data. Section 5.1 and 5.2 cover training and reconstruction details, respectively. Section 5.3 introduces the metrics used to evaluate reconstruction performance. Section 5.4 evaluates the in-distribution performance of our method on knee images from the fastMRI dataset [Zbo+19]. Section 5.5 measures the robustness to distribution shifts. In Section 5.6 we study the effect of different hyperparameter choices on reconstruction performance.

Throughout this section, we compare our method to four other methods for MRI reconstruction:

- U-Net [RFB15], a baseline trained method.
- VarNet [Sri+20], the state-of-the-art trained method.
- TV regularization [BUF07], a classical un-trained method.
- ConvDecoder [DH21], the state-of-the-art un-trained method.

### 5.1 Training the normalizing flow

We train the flow-based prior on complex-valued patches with spatial dimensions  $P \times P$ , where the patch size  $P$  is a hyperparameter.

### 5.1.1 Combining coil images

The generative model  $\mathbf{g}$  has to fulfill two constraints that originate from its use in the optimization problem (22). First, the prior should generate a complex-valued image. Second, the prior should generate a single image rather than  $n_c$  individual coil images. To obtain suitable training data, we thus need a method that combines a complex-valued multi-coil image into a single complex-valued image.

Root-sum-of-squares gives a good estimate of the magnitude of the combined image, but loses the phase. An alternative approach for combining multiple coil images into a single image, called the *reduce* operator ( $\mathcal{R}$ ), uses estimates of the sensitivity maps [Sri+20]:

$$\mathcal{R}(\mathbf{x}_1, \dots, \mathbf{x}_{n_c}) = \sum_{i=1}^{n_c} \hat{\mathbf{S}}_i^* \mathbf{x}_i. \quad (25)$$

The problem with (25) is that estimating the sensitivity maps leads to inaccuracies. In order to increase the accuracy of the reduce operator, we replace the magnitude with the RSS estimate (17). Formally, we can express this operation as:

$$\mathcal{R}'(\mathbf{x}_1, \dots, \mathbf{x}_{n_c}) = \mathbf{r}(\cos \varphi + j \sin \varphi), \quad (26)$$

with the magnitude  $\mathbf{r} = \text{RSS}(\mathbf{x}_1, \dots, \mathbf{x}_{n_c})$ , the phase  $\varphi = \arg(\mathcal{R}(\mathbf{x}_1, \dots, \mathbf{x}_{n_c}))$  and the imaginary unit  $j$ .

With this modification, the inaccuracy of the estimated sensitivity maps is only contained in the phase of the image, but not in its magnitude. Using  $\mathcal{R}'$  in place of  $\mathcal{R}$  to obtain training images can improve reconstruction performance of the trained prior.

### 5.1.2 Preprocessing

The fastMRI dataset [Zbo+19] contains volumes of MRI measurements, each of which is a collection of slices (one slice corresponds to one image). In order to obtain patches for training the generative flow, we start with a fully sampled k-space volume. We then apply the following steps:

1. Select the middle slice from the volume (also known as the mid-slice).
2. Apply the inverse Fourier transform to obtain a (complex-valued) image of shape  $n_c \times H \times W$ .
3. Center crop to obtain an image of shape  $n_c \times 320 \times 320$ .

4. Combine the individual coil images into a single image of shape  $320 \times 320$  using  $\mathcal{R}$  or  $\mathcal{R}'$  (see Section 5.1.3 for when we use  $\mathcal{R}'$  in place of  $\mathcal{R}$ ).
5. Normalize the image by dividing each pixel by the maximum of the magnitude of the image.
6. Random crop to obtain a patch of shape  $P \times P$ .

We train the flow-based prior on these normalized image patches. Note that we do not make any assumptions about the undersampling mask  $\mathbf{M}$ . Hence, a single trained prior can be used for different sampling patterns and undersampling ratios.

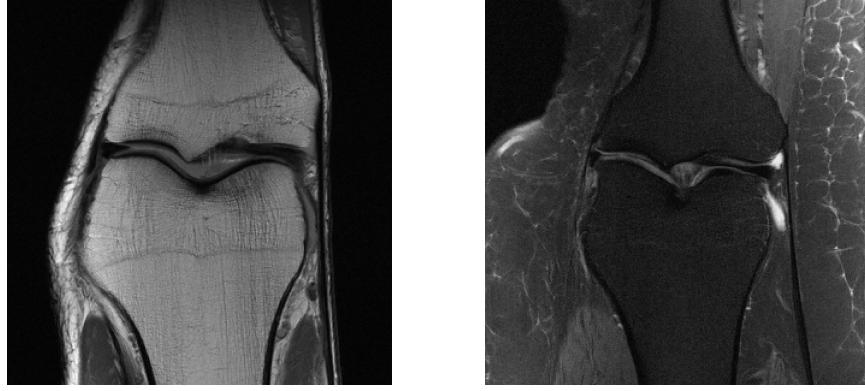
### 5.1.3 Training details

As explained in Section 2.2, we use the Glow architecture proposed by Kingma and Dhariwal [KD18]. We replace the exponential function in the affine coupling layer with the sigmoid function as in the official implementation of [KD18]. We rescale the output of the sigmoid to  $(0.5, 1)$  as proposed by Behrmann et al. [Beh+21] to improve stability in the inverse direction.

Our model has two input channels, one for the real and imaginary part of the patch, respectively. We train the model for 15000 iterations with a batch size of 16 using the Adam optimizer [KB15]. We use a learning rate of  $10^{-4}$  with linear warmup over the first 300 iterations and default values for  $\beta_1$  and  $\beta_2$ .

We train the model on the fastMRI knee training set, which contains images of two acquisition types: With and without fat suppression. Figure 3 shows typical knee images. We found that training separate priors for each acquisition type results in better reconstruction performance than training a single prior for images of both acquisition types (see Section 5.6 for a comparison). Furthermore, the architecture that works best for fat-suppressed images differs from the one for non-fat-suppressed images.

**Architecture for images with fat suppression** We use  $\mathcal{R}$  to combine coil images during preprocessing. We train on image patches with a spatial size of  $64 \times 64$ . We use  $L = 3$  levels with  $K = 32$  steps each. We use a stack of three convolutional layers with 256 channels to implement the transformations in the affine coupling layers.



(a) Image without fat suppression. (b) Image with fat suppression.

Figure 3: Typical mid-slice images from the fastMRI knee dataset.

**Architecture for images without fat suppression** We use  $\mathcal{R}'$  to combine coil images during preprocessing. We train on image patches with a spatial size of  $128 \times 128$ . We use  $L = 3$  levels with  $K = 10$  steps each. We use the U-Net [RFB15] to implement the transformations in the affine coupling layers as proposed by Kothari et al. [Kot+21]. Each U-Net consists of two downsampling and upsampling blocks. We use 32 and 64 output channels for the convolutional layers in the first and second downsampling block, respectively.

## 5.2 Reconstruction details

We use the Adam optimizer to solve the optimization problem (22). We use a learning rate of 0.08 for non-fat-suppressed images and a learning rate of 0.02 for fat-suppressed images. We run the optimization for 1000 iterations.

Others [Asi+20; KBA21] obtain the best results by setting the regularization parameter  $\gamma$  to 0 when using flow-based priors to solve compressed sensing problems. But we find that for multi-coil MRI reconstruction,  $\gamma = 10^{-3}$  and  $\gamma = 10^{-2}$  work best for non-fat-suppressed images and fat-suppressed images, respectively. Details on this choice can be found in Section 5.6.

We experimented with an overlap between adjacent patches as described by Helminger et al. [Hel+20], but did not notice an improvement in reconstruction performance compared to non-overlapping patches. Thus, we reconstruct without overlap between patches in all of our experiments. For images whose height  $H$  or width  $W$  is not divisible by the patch size  $P$ ,

we reconstruct an image of shape  $(\lceil \frac{H}{P} \rceil \cdot P, \lceil \frac{W}{P} \rceil \cdot P)$  and crop to  $(H, W)$  before passing the image through the forward model.

We normalize the undersampled k-space by dividing each pixel by the maximum of its zero-filled reconstruction. We obtain the zero-filled reconstruction by first applying the inverse Fourier transform to the undersampled k-space and then applying the RSS operator. This form of normalization allows the flow-based model to generate patches in a dynamic range similar to the one it was trained on, which is critical for achieving good performance.

### 5.3 Metrics

We measure image quality in terms of peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [Wan+04] and visual information fidelity (VIF) [SB06]. We use the RSS image obtained from fully sampled data as the ground truth for our comparisons.

To make our results directly comparable to those obtained by Darestani and Heckel [DH21] with other methods, we scale the ground truth image such that it has the same mean and standard deviation as the reconstructed image. Furthermore, we make image-based comparisons, i.e., we use the dynamic range of a single slice rather than that of the entire volume for computing scores [DH21]. We refer the reader to [DH21] for a discussion of these choices.

### 5.4 In-distribution performance

We use the same setup as Darestani and Heckel [DH21]: We evaluate on multi-coil knee images from the fastMRI dataset. Specifically, we use 4x accelerated mid-slice images of all 199 volumes in the validation set.

Table 1 shows the results obtained with our method compared to trained and un-trained methods. For VarNet, U-Net, ConvDecoder and TV we use the results from [DH21] and refer the reader to that paper for details of the parameters used for each of these methods.

Our method achieves the best VIF scores of all evaluated methods, but is outperformed by end-to-end trained methods (VarNet and U-Net) according to SSIM. Our method obtains slightly higher scores than ConvDecoder. In contrast to our trained prior, ConvDecoder is an un-trained method that solely relies on its convolutional architecture as an image prior. Furthermore, our method significantly outperforms TV.

It should be noted that the scores obtained with our method are sensitive to how the ground truth image and the reconstructed image are scaled.

Table 1: Performance on the fastMRI knee validation set (4x accelerated). Rows are sorted by SSIM. Marginal errors denote 95% confidence intervals. Results for VarNet, U-Net, ConvDecoder and TV are taken from [DH21].

	VIF	SSIM	PSNR
VarNet	$0.6456 \pm 0.0122$	<b><math>0.8342 \pm 0.0084</math></b>	<b><math>34.20 \pm 0.18</math></b>
U-Net	$0.5955 \pm 0.0147$	$0.7883 \pm 0.0097$	$32.04 \pm 0.21$
INN Prior (Ours)	<b><math>0.7524 \pm 0.0191</math></b>	$0.7834 \pm 0.0117$	$32.07 \pm 0.32$
ConvDecoder	$0.6823 \pm 0.0217$	$0.7753 \pm 0.0145$	$31.67 \pm 0.39$
TV	$0.4412 \pm 0.0272$	$0.6977 \pm 0.0141$	$30.20 \pm 0.41$

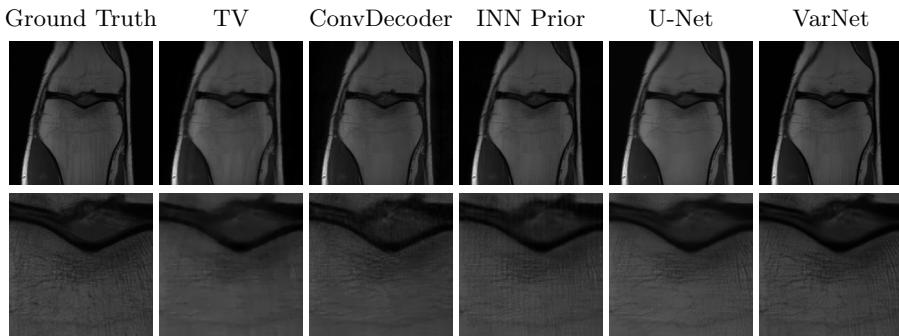


Figure 4: Sample reconstructions for a 4x accelerated measurement from the fastMRI knee validation set. The bottom row shows zoomed in center regions.

If we scale the mean and standard deviation of the reconstructed image to the ground truth image (instead of vice versa as in Table 1), we obtain a significantly lower VIF of 0.5743, a slightly higher SSIM of 0.7924 and a slightly higher PSNR of 32.34 for our method. In contrast to that, the scores for VarNet change only marginally depending on the choice of scaling.

Figure 4 shows sample reconstructions for a 4x accelerated measurement from the fastMRI knee validation set. More reconstructions can be found in Appendix A.

## 5.5 Robustness to distribution shifts

So far we have studied reconstruction performance on images that come from the same distribution as the training set. In this section, we evaluate the robustness of our method to distribution shifts. These shifts are relevant in practice, as it is common to have different patient populations or different machines for acquisition at test time compared to the training dataset [DCH21].

Darestani et al. [DCH21] introduced several experiments in order to evaluate the robustness of image reconstruction methods. They studied the robustness of trained methods (U-Net and VarNet) as well as un-trained methods (both traditional and neural network based), but they did not study methods that use a trained prior. When evaluating the robustness of our method, we use the same experimental setup as Darestani et al. [DCH21] to make our results directly comparable to their results using other methods.

### 5.5.1 Dataset shift

First, we evaluate how well a prior trained on the fastMRI knee dataset works on the Stanford dataset, a knee dataset created by Darestani et al. [DCH21]. Since the properties of images in the Stanford dataset differ from those in the fastMRI dataset, we use the Stanford dataset to measure out-of-distribution performance. We refer the reader to [DCH21] for details of the Stanford dataset. Since the Stanford dataset contains only fat-suppressed images, we use the 99 fat-suppressed images from the fastMRI knee validation set as a measure of in-distribution performance [DCH21].

Figure 5 shows the results. We observe that our prior-based method loses a similar amount of performance as other methods when going from fastMRI to Stanford. This result is in line with the findings of Darestani et al. [DCH21], who observed a linear relationship between in-distribution and out-of-distribution performance for trained and un-trained methods, similar to what has been observed for classification problems in [Rec+19; Tao+20].

### 5.5.2 Anatomy shift

In our next experiment, we study the robustness of our method to anatomy shifts. We reconstruct 100 randomly selected images from the fastMRI brain validation set with a prior that was trained only on knee images. As a measure of in-distribution performance, we use 100 randomly selected images from the fastMRI knee validation set.

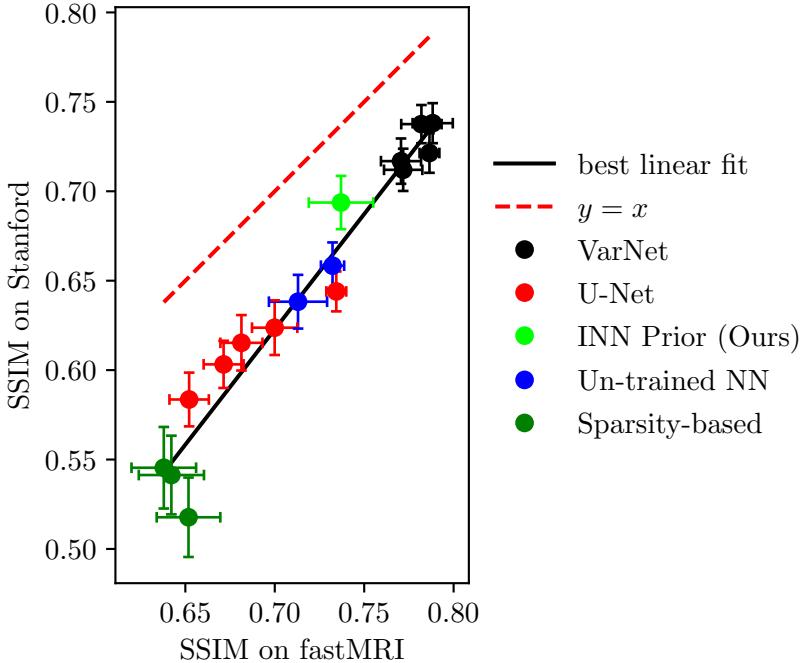


Figure 5: Our method loses a similar amount of performance as other methods when reconstructing images from the Stanford dataset. Note that the SSIM scores on the x-axis include only the fat-suppressed images from the fastMRI dataset. Hence, these scores are lower than in Table 1. Error bars denote 95% confidence intervals. Multiple points for a given method correspond to different variants. The figure is taken from [DCH21]. We added our method (in light green) to the methods evaluated in that paper.

Figure 6 shows the results. Similar to the results obtained by Darestani et al. [DCH21] for other methods, our method achieves higher SSIM scores on the brain dataset than on the knee dataset, even when trained only on knee images. As Darestani et al. [DCH21] noted, this effect comes from the fact that brain images are naturally easier to reconstruct.

### 5.5.3 Adversarially-filtered shift

Darestani et al. [DCH21] also studied whether there are knee images that are naturally difficult to reconstruct or whether one method does particularly well on images that other methods struggle with. To answer this

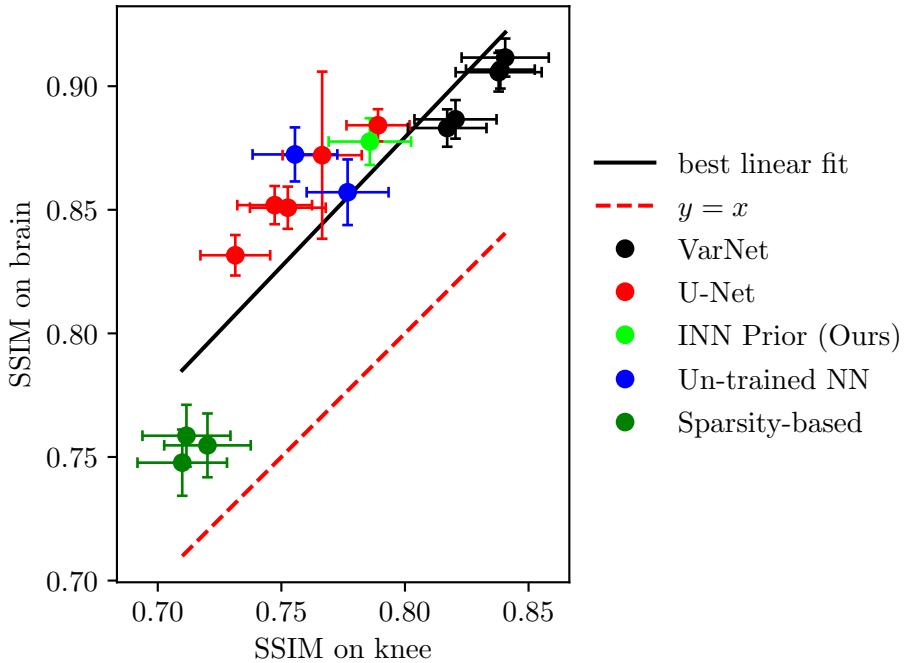


Figure 6: Similar to other methods, our method achieves higher SSIM scores on the fastMRI brain dataset, even when trained only on knee images. Error bars denote 95% confidence intervals. Multiple points for a given method correspond to different variants. The figure is taken from [DCH21]. We added our method (in light green) to the methods evaluated in that paper.

question, they created the fastMRI-A dataset, which includes the 10% of mid-slices from the fastMRI knee validation set on which the i-RIM architecture [PW19] achieves the lowest SSIM scores. As a measure of in-distribution performance, we use the same subset of the fastMRI knee validation set as in Section 5.5.2.

Figure 7 shows the results on the fastMRI-A dataset. We observe that our prior-based method loses a similar amount of performance as other methods when reconstructing challenging images.

## 5.6 Ablation study

In this section, we study how the choice of architecture and regularization strategy affect reconstruction performance. Unless otherwise noted, we use

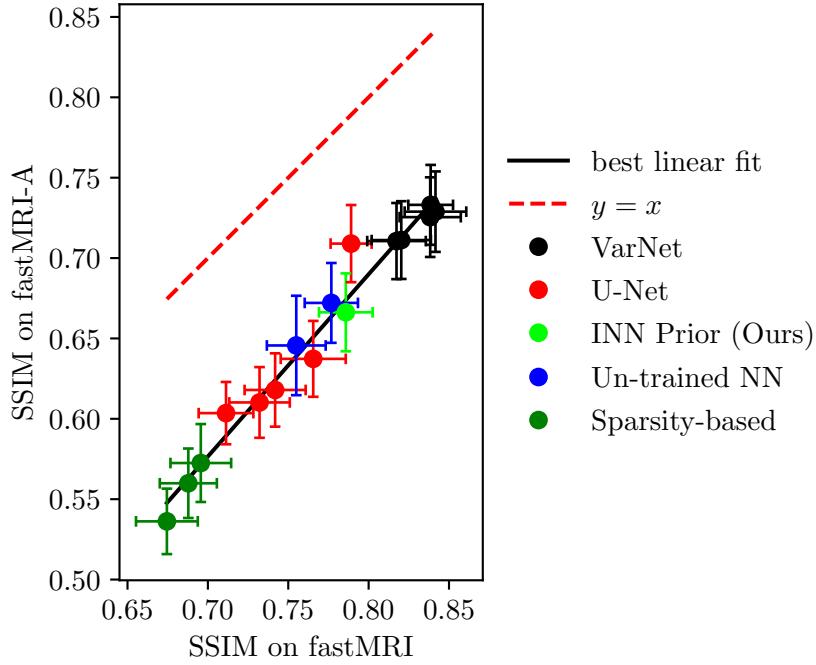


Figure 7: Our method loses a similar amount of performance as other methods when reconstructing images from the fastMRI-A dataset. Error bars denote 95% confidence intervals. Multiple points for a given method correspond to different variants. The figure is taken from [DCH21]. We added our method (in light green) to the methods evaluated in that paper.

the same setup as in Section 5.4, i.e., we use all mid-slice images from the fastMRI multi-coil knee validation set (4x accelerated).

### 5.6.1 Architecture

In order to achieve the final performance described in the previous sections, we experimented with different choices for the architecture of the prior. Here, we focus on three of them: Using separate priors for each acquisition type, the patch size of the prior and the choice of affine coupling.

**Separate priors** First, we compare the performance of a common prior for both acquisition types (with and without fat suppression) to the performance of a separate prior for each acquisition type. In the case of separate priors we

Table 2: Reconstruction performance (SSIM) on the fastMRI knee validation set (4x accelerated) of a separate prior for each acquisition type compared to a common prior. Marginal errors denote 95% confidence intervals.

	non-fat-suppressed	fat-suppressed
separate priors	$0.8292 \pm 0.0078$	$0.7370 \pm 0.0180$
common prior	$0.7862 \pm 0.0112$	$0.7134 \pm 0.0174$

use the parameters described in Sections 5.1 and 5.2. When reconstructing images of both acquisition types with a single prior, the architecture optimized for fat-suppressed images generalizes better to non-fat-suppressed images than vice versa. We train on images of both acquisition types in the case of a single prior. Table 2 compares the performance of two separate priors to that of a single prior. As intuitively expected, the priors optimized for images of one specific acquisition type perform better than the common prior.

**Patch size** Next, we study the effect of the patch size on reconstruction performance. Table 3 shows the architectures used for different patch sizes. For  $P = 32$  we use an architecture for non-fat-suppressed images that differs slightly from the one described in Section 5.1. With the default architecture, the spatial dimensions of the intermediate representations at deeper levels would become too small, since the U-Net in the affine coupling as well as the squeeze operation reduce the spatial size.

Table 4 shows the performance achieved with different patch sizes. We observe that a patch size of 128 works best for non-fat-suppressed images and a patch size of 64 works best for fat-suppressed images. Inference throughput is roughly constant as a function of patch size: Independent of patch size, it takes around 14 minutes to reconstruct an image on a single GPU.

**Affine coupling** Here, we look at the choice of affine coupling. As noted earlier, a standard stack of convolutions as in [KD18] works best for fat-suppressed images and a U-Net coupling as in [Kot+21] works best for non-fat-suppressed images. Surprisingly, reconstruction performance of a prior is not necessarily related to the quality of samples generated from randomly drawn latent vectors.

Figure 8 shows an example of this effect. Figures 8a and 8b show samples from a generative flow with standard affine coupling and U-Net affine

Table 3: Architectures used for different patch sizes. As affine coupling, we use a stack of three convolutional layers for fat-suppressed images and the U-Net for non-fat-suppressed images. All models have roughly the same number of parameters.

	non-fat-suppressed	fat-suppressed
32	$L = 2, K = 15$	$L = 3, K = 32$
64	$L = 3, K = 10$	$L = 3, K = 32$
128	$L = 3, K = 10$	$L = 3, K = 32$

Table 4: Reconstruction performance (SSIM) on the fastMRI knee validation set (4x accelerated) for priors with different patch sizes. Marginal errors denote 95% confidence intervals.

	non-fat-suppressed	fat-suppressed
32	$0.8151 \pm 0.0093$	$0.7347 \pm 0.0181$
64	$0.8214 \pm 0.0074$	$0.7370 \pm 0.0180$
128	$0.8292 \pm 0.0078$	$0.7293 \pm 0.0177$

coupling, respectively. The samples generated by the model with standard affine coupling look sharper and more similar to the distribution it was trained on ( $64 \times 64$  patches of non-fat-suppressed mid-slice images) than those of the model with U-Net affine coupling. Furthermore, the model with standard affine coupling achieves a better negative log-likelihood of 4.00 on the validation set compared to a negative log-likelihood of 4.12 for the model with U-Net affine coupling.

But the model with U-Net affine coupling works better when imposed as a prior for reconstructing undersampled measurements. When reconstructing all 100 non-fat-suppressed mid-slice images from the fastMRI knee validation set (4x accelerated), we obtain a SSIM of 0.8060 with the standard affine coupling and a SSIM of 0.8214 with the U-Net affine coupling.

### 5.6.2 Regularization

Finally, we study the importance of explicit regularization of the latent vector. Recall our optimization problem from Section 4.2:

$$\mathcal{L}(\tilde{\mathbf{z}}) = \sum_{i=1}^{n_c} \left\| \mathbf{y}_i - \mathbf{MF}\hat{\mathbf{S}}_i \mathcal{T}(\tilde{\mathbf{z}}) \right\|^2 + \gamma \|\tilde{\mathbf{z}}\|^2. \quad (27)$$

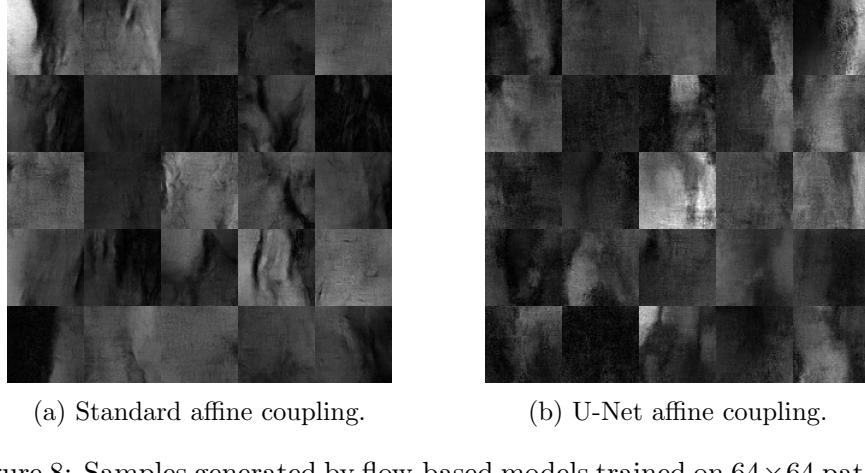


Figure 8: Samples generated by flow-based models trained on  $64 \times 64$  patches of non-fat-suppressed mid-slice images.

As noted earlier,  $\gamma$  is a regularization parameter that penalizes latent vectors with low likelihoods. Inspired by an experiment by Asim et al. [Asi+20], we measure reconstruction performance as a function of  $\gamma$ . This experiment is performed on 20 randomly selected mid-slice images from the fastMRI knee validation set.

Figures 9 and 10 show the results for non-fat-suppressed and fat-suppressed images, respectively. As explained in Section 4.2, we perform a data consistency step at the end of the optimization. Perhaps surprisingly, the  $\gamma$  that leads to the best performance before applying data consistency differs from the one that leads to the best final performance after the data consistency step. This effect is especially pronounced in the case of fat-suppressed images, where  $\gamma = 10^{-6}$  works best before enforcing data consistency, but  $\gamma = 10^{-2}$  leads to significantly higher performance when taking the data consistency step into account.

## 6 Conclusion

In this thesis, we applied flow-based priors to the problem of accelerated multi-coil MRI. Our method is patch-based, which allows us to reconstruct images of arbitrary shapes. Since we do not make any assumptions about the undersampling operator during training, a trained prior can be used for different sampling patterns and undersampling ratios.

Our method outperforms the best un-trained methods that enable the

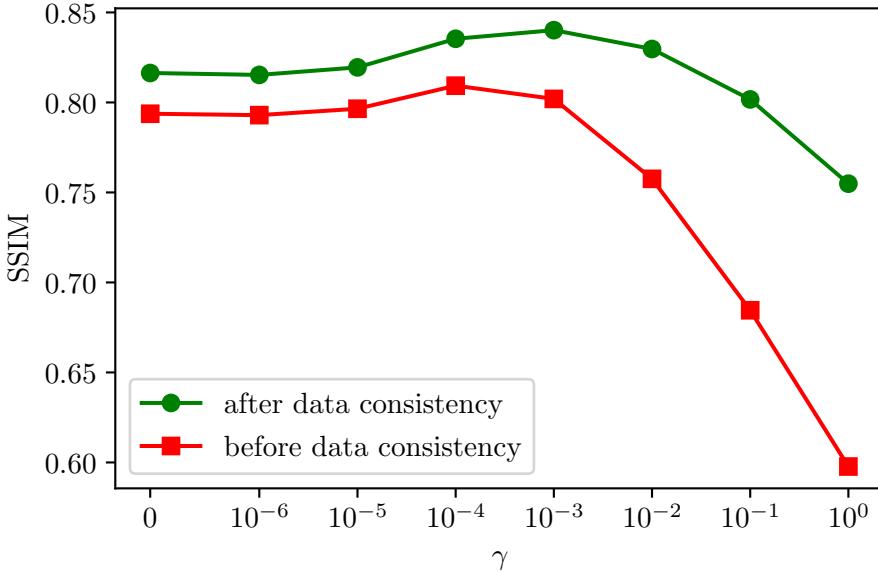


Figure 9: Reconstruction performance as a function of  $\gamma$  for a set of 12 non-fat-suppressed images from the fastMRI knee validation set (4x accelerated).

same flexibility. However, it is outperformed by end-to-end trained methods like VarNet which require retraining when the undersampling operator changes. Furthermore, we measured the robustness of our method to distribution shifts and found that flow-based priors lose a similar amount of performance under these shifts as trained and un-trained methods.

As our method has considerably lower inference throughput than end-to-end trained methods, an interesting direction of research would be to reduce the computation time of prior-based reconstruction methods.

## Code

Code is available at [https://github.com/ffraaz/flow\\_based\\_priors](https://github.com/ffraaz/flow_based_priors).

## Acknowledgments

We thank Mohammad Zalbagi Darestani for providing us with a VarNet checkpoint and the scores for VarNet, U-Net, ConvDecoder and TV used in Figures 5 to 7.

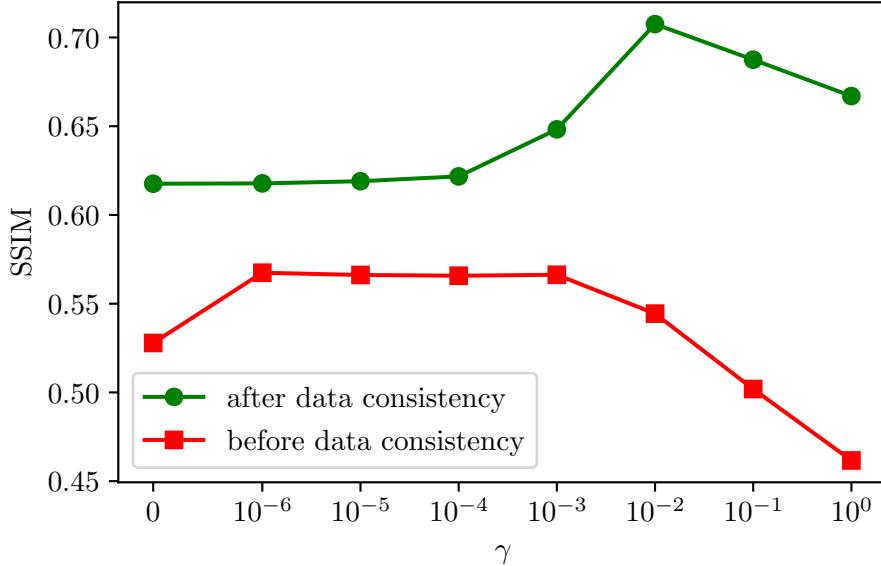


Figure 10: Reconstruction performance as a function of  $\gamma$  for a set of 8 fat-suppressed images from the fastMRI knee validation set (4x accelerated).

## References

- [ALT20] R. Anirudh, S. Lohit, and P. Turaga. “Generative patch priors for practical compressive image recovery”. In: *arXiv:2006.10873* (2020).
- [Asi+20] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand. “Invertible generative models for inverse problems: Mitigating representation error and dataset bias”. In: *International Conference on Machine Learning*. 2020, pp. 399–409.
- [Beh+21] J. Behrmann, P. Vicol, K.-C. Wang, R. B. Grosse, and J. Jacobsen. “Understanding and mitigating exploding inverses in invertible neural networks”. In: *AISTATS*. 2021.
- [Bor+17] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. “Compressed sensing using generative models”. In: *International Conference on Machine Learning*. 2017, pp. 537–546.
- [BUF07] K. Block, M. Uecker, and J. Frahm. “Undersampled radial MRI with multiple coils. Iterative image reconstruction using a to-

- tal variation constraint”. In: *Magnetic Resonance in Medicine* (2007), pp. 1086–1098.
- [BZA20] S. Bhadra, W. Zhou, and M. A. Anastasio. “Medical image reconstruction with image-adaptive priors learned by use of generative adversarial networks”. In: *arXiv:2001.10830* (2020).
- [Can06] E. J. Candès. “Compressive sampling”. In: *International Congress of Mathematicians*. 2006, pp. 1433–1452.
- [DCH21] M. Z. Darestani, A. S. Chaudhari, and R. Heckel. “Measuring robustness in deep learning based compressive sensing”. In: *International Conference on Machine Learning*. 2021, pp. 2433–2444.
- [DH21] M. Z. Darestani and R. Heckel. “Accelerated MRI with untrained neural networks”. In: *arXiv:2007.02471* (2021).
- [DKB15] L. Dinh, D. Krueger, and Y. Bengio. “NICE: Non-linear independent components estimation”. In: *International Conference on Learning Representations*. 2015.
- [DSB17] L. Dinh, J. Sohl-Dickstein, and S. Bengio. “Density estimation using real NVP”. In: *International Conference on Learning Representations*. 2017.
- [Goo+14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. “Generative adversarial nets”. In: *Advances in Neural Information Processing Systems*. 2014.
- [Gri+02] M. Griswold, P. Jakob, R. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. “Generalized autocalibrating partially parallel acquisitions (GRAPPA)”. In: *Magnetic Resonance in Medicine* (2002), pp. 1202–1210.
- [Hec19] R. Heckel. “Regularizing linear inverse problems with convolutional neural networks”. In: *arXiv:1907.03100* (2019).
- [Hel+20] L. Helminger, M. Bernasconi, A. Djelouah, M. Gross, and C. Schroers. “Blind image restoration with flow based priors”. In: *arXiv:2009.04583* (2020).
- [HKO01] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. 2001.

- [IS15] S. Ioffe and C. Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International Conference on Machine Learning*. 2015, pp. 448–456.
- [Jal+21] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir. “Robust compressed sensing MRI with deep generative priors”. In: *Advances in Neural Information Processing Systems*. 2021.
- [KA21] V. A. Kelkar and M. A. Anastasio. “Prior image-constrained reconstruction using style-based generative models”. In: *arXiv:2102.12525* (2021).
- [KB15] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *International Conference on Learning Representations*. 2015.
- [KBA21] V. A. Kelkar, S. Bhadra, and M. A. Anastasio. “Compressible latent-space invertible networks for generative model-constrained image reconstruction”. In: *IEEE Transactions on Computational Imaging* (2021), pp. 209–223.
- [KD18] D. P. Kingma and P. Dhariwal. “Glow: Generative flow with invertible 1x1 convolutions”. In: *Advances in Neural Information Processing Systems*. 2018.
- [Kot+21] K. Kothari, A. Khorashadizadeh, M. de Hoop, and I. Dokmanić. “Trumpets: Injective flows for inference and inverse problems”. In: *arXiv:2102.10461* (2021).
- [KW14] D. P. Kingma and M. Welling. “Auto-encoding variational Bayes”. In: *International Conference on Learning Representations*. 2014.
- [LDP07] M. Lustig, D. Donoho, and J. M. Pauly. “Sparse MRI: The application of compressed sensing for rapid MR imaging”. In: *Magnetic Resonance in Medicine* (2007), pp. 1182–1195.
- [Nar+19] D. Narnhofer, K. Hammernik, F. Knoll, and T. Pock. “Inverse GANs for accelerated MRI reconstruction”. In: *Wavelets and Sparsity XVIII*. 2019, pp. 381–392.
- [Pru+99] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger. “SENSE: Sensitivity encoding for fast MRI”. In: *Magnetic Resonance in Medicine* (1999).

- [PW19] P. Putzky and M. Welling. “Invert to learn to invert”. In: *Advances in Neural Information Processing Systems*. 2019.
- [Rec+19] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar. “Do ImageNet classifiers generalize to ImageNet?” In: *International Conference on Machine Learning*. 2019, pp. 5389–5400.
- [RFB15] O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional networks for biomedical image segmentation”. In: *arXiv:1505.04597* (2015).
- [RM15] D. Rezende and S. Mohamed. “Variational inference with normalizing flows”. In: *International Conference on Machine Learning*. 2015, pp. 1530–1538.
- [Roe+90] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller. “The NMR phased array”. In: *Magnetic Resonance in Medicine* (1990), pp. 192–225.
- [SB06] H. Sheikh and A. Bovik. “Image information and visual quality”. In: *IEEE Transactions on Image Processing* (2006), pp. 430–444.
- [SE20] Y. Song and S. Ermon. “Improved techniques for training score-based generative models”. In: *Advances in Neural Information Processing Systems*. 2020.
- [SM97] D. K. Sodickson and W. J. Manning. “Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays”. In: *Magnetic Resonance in Medicine* (1997), pp. 591–603.
- [Sri+20] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson. “End-to-end variational networks for accelerated MRI reconstruction”. In: *arXiv:2004.06688* (2020).
- [Tao+20] R. Taori, A. Dave, V. Shankar, N. Carlini, B. Recht, and L. Schmidt. “Measuring robustness to natural distribution shifts in image classification”. In: *Advances in Neural Information Processing Systems*. 2020, pp. 18583–18599.
- [Uec+14] 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”. In: *Magnetic Resonance in Medicine* (2014), pp. 990–1001.

- [Wan+04] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. “Image quality assessment: From error visibility to structural similarity”. In: *IEEE Transactions on Image Processing* (2004), pp. 600–612.
- [Zbo+19] J. Zbontar et al. “fastMRI: An open dataset and benchmarks for accelerated MRI”. In: *arXiv:1811.08839* (2019).
- [ZFZ19] H. Zheng, F. Fang, and G. Zhang. “Cascaded dilated dense network with two-step data consistency for MRI reconstruction”. In: *Advances in Neural Information Processing Systems*. 2019.

## A Sample Reconstructions

Figures 11 to 14 show reconstructions of measurements from the fastMRI multi-coil knee validation set (4x accelerated).



Figure 11: Sample reconstructions of non-fat-suppressed measurements from the fastMRI multi-coil knee validation set (4x accelerated).

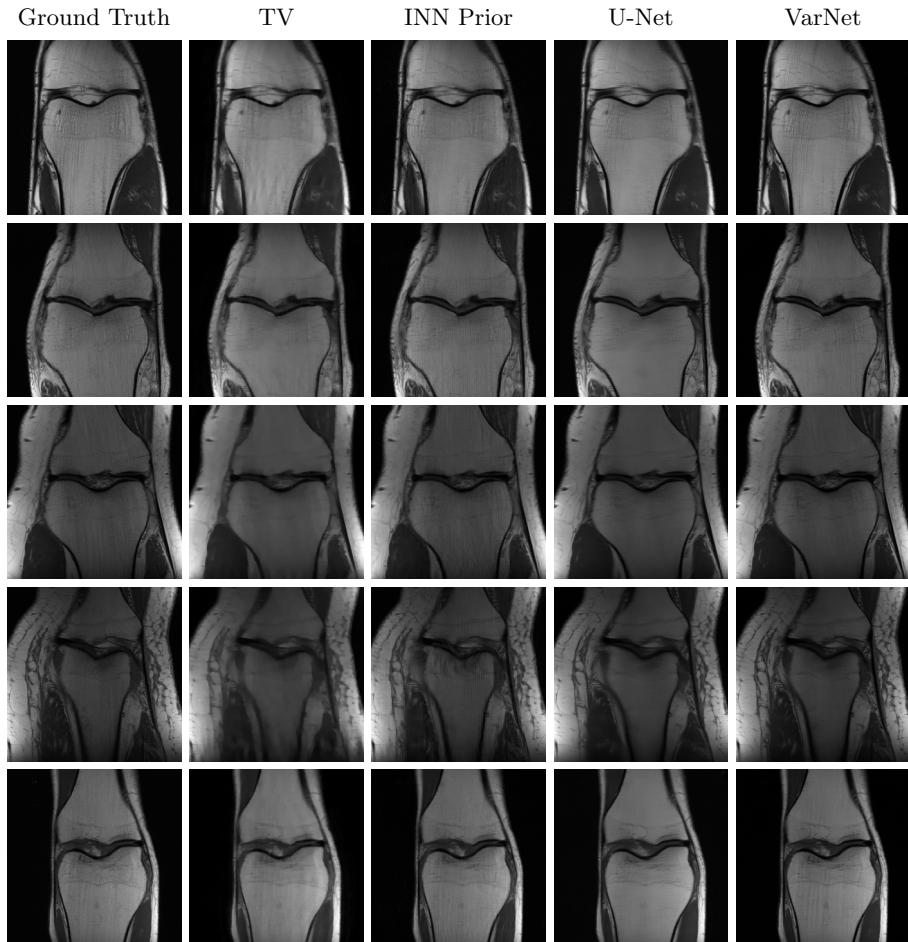


Figure 12: Sample reconstructions of non-fat-suppressed measurements from the fastMRI multi-coil knee validation set (4x accelerated).

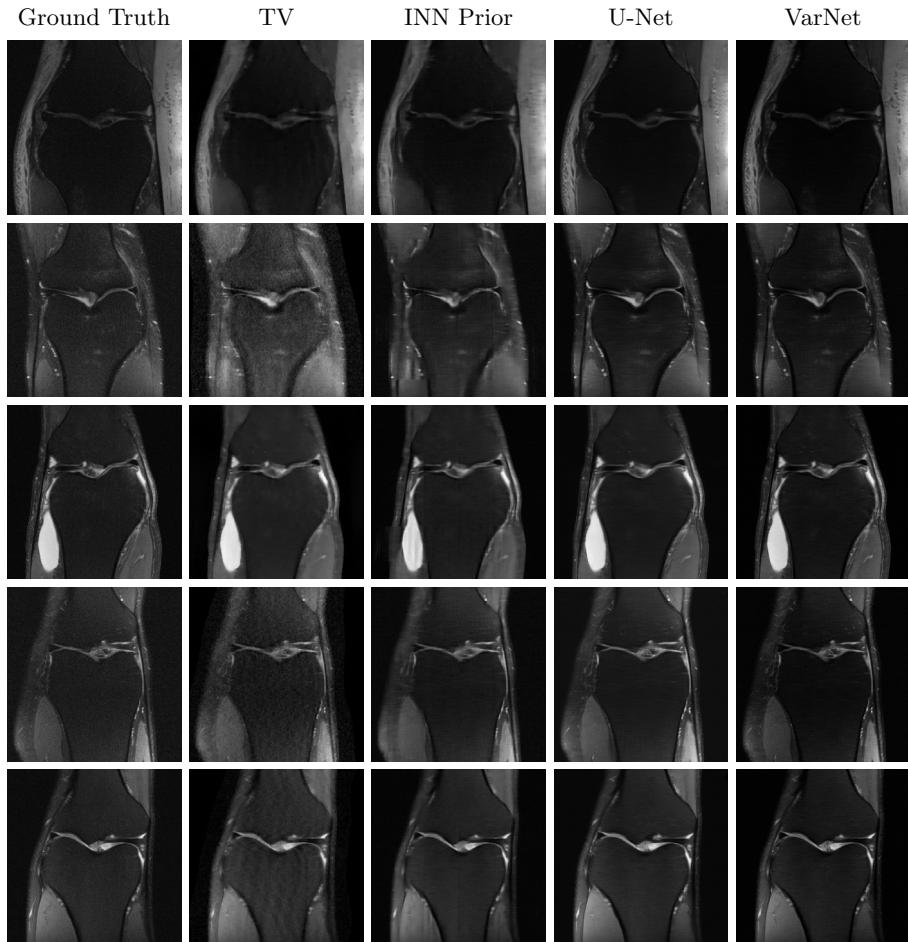


Figure 13: Sample reconstructions of fat-suppressed measurements from the fastMRI multi-coil knee validation set (4x accelerated).

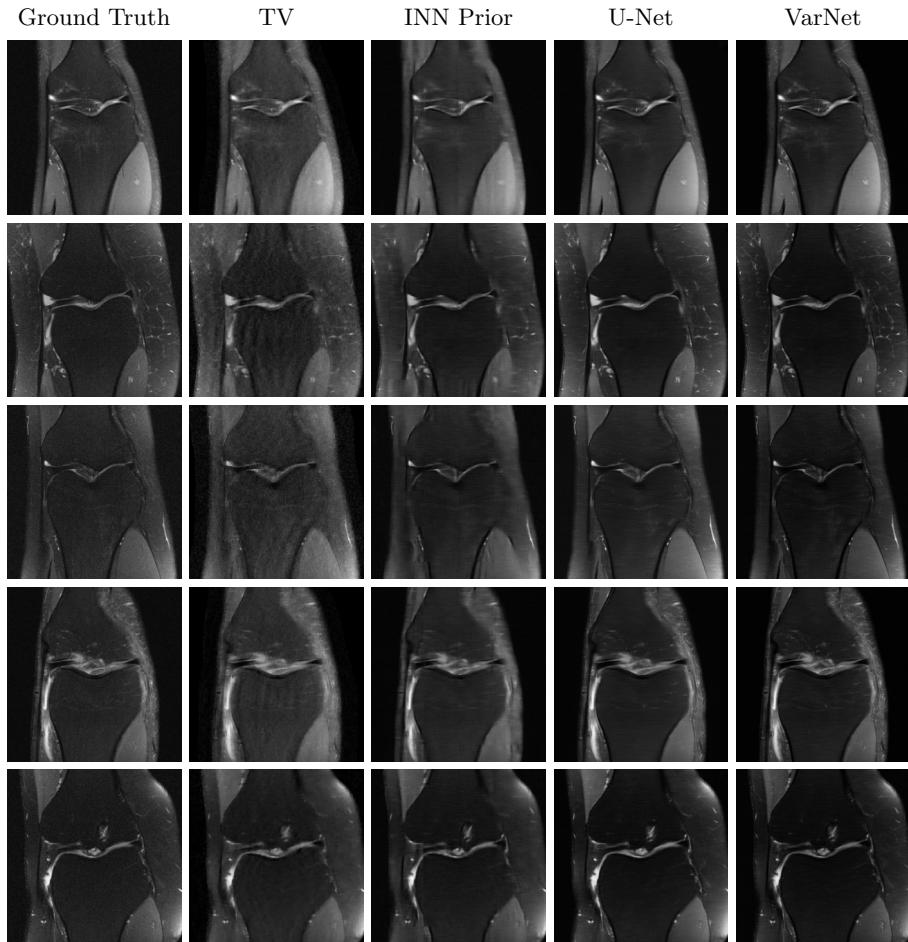


Figure 14: Sample reconstructions of fat-suppressed measurements from the fastMRI multi-coil knee validation set (4x accelerated).