Robust Compressed Sensing MR Imaging with Deep Generative Priors

Anonymous Author(s)
Affiliation
Address
email

Abstract

The CSGM framework (Bora-Jalal-Price-Dimakis'17) has shown that deep generative priors can be powerful tools for solving inverse problems. However, this framework has been empirically successful only on certain datasets (for example, human faces and MNIST digits), and it is known to perform poorly on out-of-distribution samples. In this paper, we present the first successful application of the CSGM framework on clinical MRI data. We train a generative prior on brain scans from the fastMRI dataset, and show that posterior sampling via Langevin dynamics achieves high quality reconstructions. Furthermore, our experiments and theory show that posterior sampling is robust to changes in the ground-truth distribution and measurement process.

1 Introduction

Compressed sensing [24, 16] has enabled reductions to the number of measurements needed for successful reconstruction in a variety of imaging inverse problems. In particular, it has led to shorter scan times for magnetic resonance imaging (MRI) [59, 84], and most MRI vendors have released products leveraging this framework to accelerate clinical workflows. Despite their success, sparsity-based methods are limited by the achievable acceleration rates, as the sparsity assumptions are either hand-crafted or are limited to simple learned sparse codes [67, 68].

More recently, deep learning techniques have been used as powerful data-driven reconstruction methods for inverse problems [48, 63]. There are two broad families of deep-learning inversion techniques [63]: end-to-end supervised and distribution-learning approaches. End-to-end supervised inversion techniques use a training set of measured images and deploy convolutional neural networks (CNNs) and other architectures to learn the inverse mapping from measurements to image. Network architectures that include both CNN blocks and knowledge of the forward model have grown in popularity, as they combine deep learning with the compressed sensing optimization framework, see e.g. [33, 3, 60]. End-to-end methods are trained for specific imaging anatomy and measurement models and show excellent performance in these tasks. However, reconstruction quality is known to suffer when applied out of distribution, and recently has been shown to severely degrade [5, 20] under various types of natural measurement and anatomy perturbations.

In this paper we study deep-learning inversion techniques based on distribution learning. These models are trained without reference to measurements, and so easily adapt to changes in the measurement process. The most common family of such techniques, known also as Compressed Sensing with Generative Modeling (CSGM) [14] uses pre-trained generative models as priors. Generative models are extremely powerful at representing image statistics and CSGM has been successfully applied to numerous inverse problems [14, 34] including non-linear phase retrieval [35] and improved with invertible models [7], sparsity based deviations [22], image adaptivity [41], and intermediate layer optimization [19]. These methods have only recently been applied to MRI and have not yet been

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

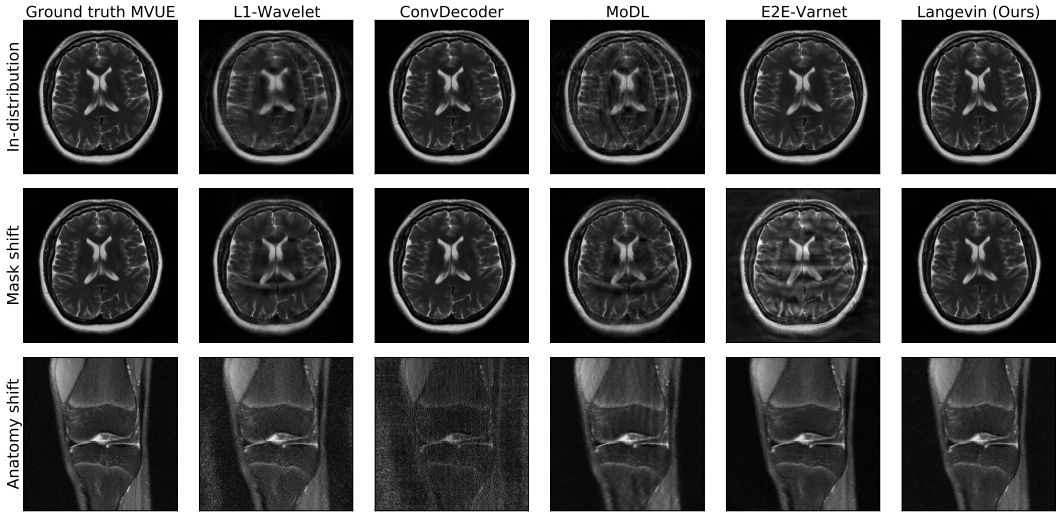

Figure 1: Comparison of reconstruction methods for in-distribution, sampling-shift, and anatomy-shift images. Our generative prior was trained *exclusively* on T2-weighted *brain* scans from the NYU fastMRI dataset, and the training of the generative prior was independent of the measurement model. Observers familiar with reading MR images [88] will notice that our reconstruction is similar in diagnostic value to E2E-Varnet, a supervised state of the art, when in-distribution and outperforms all competing methods out of distribution. In the latter case, competing methods often introduce artifacts that render them diagnostically unusable.

shown to be competitive with supervised end-to-end methods. The very recent work [52] trains a StyleGAN for magnitude-only DICOM images but requires the presence of side-information and studies Gaussian, real-valued measurements for reconstruction. The mismatch from the measurement model and the use of magnitude images are known to be problematic when evaluating performance [71]. Untrained and unamortized generators [37] have also been recently explored [20], showing promising results in some cases. Further, [18] studies the harder problem of learning a generative model for a class of images using only partial observations, as first proposed in AmbientGAN [15].

In this paper we train the first score-based generative model [74, 75] for MR images. We show that we can faithfully represent MR images without any assumptions on the measurement system. As a consequence, we are able to reconstruct retrospectively under-sampled MRI data under a variety of realistic sampling schemes. We show that our generative method is competitive with end-to-end supervised training when the test-data are matched to the training data and that it is robust to various out-of-distribution shifts, while end-to-end methods significantly degrade.

1.1 Contributions

- We successfully train a score-based deep generative model for complex-valued, T2-weighted brain MR images without any assumptions on the measurement scheme. When applied to multi-coil MRI reconstruction under the CSGM framework, we achieve competitive performance compared to end-to-end deep learning methods when the test-time data are sampled within distribution.
- We give evidence that posterior sampling should give high-quality reconstructions. First, we show that for any measurements (including the Fourier measurements in MRI) that posterior sampling with the correct prior is within constant factors of the optimal recovery method; second, even if the prior is wrong but gives α mass to the true distribution, we show that posterior sampling for Gaussian measurements is nearly optimal with just an additive $O(\log(1/\alpha))$ loss.
- We empirically show that our approach is robust to test-time distribution shifts including different sampling patterns and imaging anatomy. The former is unsurprising given that our model was trained without knowledge of the measurement scheme. As a consequence, our approach provides a degree of flexibility in choosing scan parameters – a common situation in routine clinical imaging. Perhaps surprisingly, the latter indicates that a specialized training set may offer sufficient regularization for a larger class of images. In contrast, we empirically show that end-to-end methods do not enjoy the same robustness guarantees, in some cases leading to severe degradation in reconstruction quality when applied to out-of-distribution MR images.

- Our method can be used to obtain multiple samples from the posterior by running Langevin dynamics with different random initializations. This allows us to get multiple reconstructions which can be used to obtain confidence intervals for each reconstructed voxel and visualize our reconstruction uncertainty on a voxel-by-voxel resolution. Uncertainty quantification can be incorporated into end-to-end methods, e.g., using variational auto-encoders [25], but this requires changes to the architecture. Our method does not require any modification and multiple reconstruction samplers can be run in parallel.

Our main results are succinctly summarized in Figure 1: we achieve equivalent reconstruction performance using a reduced training set when evaluated in-distribution and *vastly outperform* competing methods when evaluated out-of-distribution.

1.2 Related Work

Generative priors have shown great utility to improving compressed sensing and other inverse problems, starting with [14], who generalized the theoretical framework of compressed sensing and restricted eigenvalue conditions [79, 24, 13, 16, 40, 12, 11, 26] for signals lying on the range of a deep generative model [30, 53]. Lower bounds in [50, 58, 47] established that the sample complexities in [14] are order optimal. The approach in [14] has been generalized to tackle different inverse problems [46, 35, 44, 8, 9, 66, 57, 34, 4]. Alternate algorithms for reconstruction include [22, 49, 64, 28, 27, 60, 37, 38]. The complexity of optimization algorithms using generative models have been analyzed in [29, 39, 55, 36].

Using compressed sensing for multi-coil MRI reconstruction has led to a rich body of work in the past two decades [59, 21, 81, 69]. See [23] and the recent special issue [43] for an overview of these methods. Classical approaches impose sparsity in a well-chosen basis, such as the wavelet domain [59], or apply shallow learning that leverages redundancy in the measurements [67, 68, 87]. Recent research has demonstrated the superior performance of deep neural networks for MRI image reconstruction [70, 33, 3, 76, 77]. A broad class of approaches is represented by end-to-end unrolled methods, which use deep networks as learned data priors in the image [3, 33, 76] or k-space domain [78]. Recent work has also investigated the performance of untrained methods [83, 38] for MR reconstruction and has shown competitive results. A much less explored line of research is MR image reconstruction with generative priors. The work in [62] proposes a CSGM-like algorithm that finetunes a pretrained generator, but no efficient stopping criterion is given.

2 System Model and Algorithm

2.1 Multi-coil Magnetic Resonance Imaging

MRI is a medical imaging modality that makes spatially sensitive measurements using an array of radio-frequency coils placed around the body. Each coil is spatially sensitive to a local region, and measurements are acquired directly in the spatial frequency, or *k-space*, domain. To decrease scan time, reduce operating costs, and improve patient comfort, a reduced number of k-space measurements are acquired in clinical use and reconstructed by incorporating explicit or implicit knowledge of the spatial sensitivity maps [72, 65, 31]. Formally, the vector of measurements $y_i \in \mathbb{C}^L$ acquired by the ith coil can be characterized by the forward model [65]:

$$y_i = PFS_i x^* + w_i, \quad i = 1, ..., N_c, \tag{1}$$

where $x^* \in \mathbb{C}^N$ is the image, S_i is the point-wise multiplication of the ith coil sensitivity map, F is the spatial Fourier transform operator, P represents the k-space sampling operator, and we assume $w_i \sim \mathcal{N}_c\left(0, \sigma^2 I\right)$ for simplicity. Importantly, note that the same under-sampling operator is applied to all N_c coils. Given multi-coil measurements y, sensitivity maps represented by S and the sampling operator P, the goal of MR image reconstruction is to estimate the underlying image variable x^*. This is formulated as solving the regularized optimization problem:

$$\arg\min_x \|y - Ax\|_2^2 + \lambda Q(x), \tag{2}$$

where we use the operator $A \in \mathbb{C}^{M \times N}$ to subsume the discrete approximation to all linear effects, and Q is a suitably chosen functional prior for the image variable x. For example, to enforce a

sparsity prior, one can penalize the ℓ_1 norm in the wavelet representation of x [59]. More recent approaches involve learned regularization terms parameterized by deep neural networks [70, 33, 3]. These models are typically trained *end-to-end* using a fixed training set and certain assumptions about the sampling operator. In the sequel, we present how score-based generative models can be combined with the posterior sampling [45] mechanism to reformulate (2) and achieve good quality reconstructions without any *a priori* assumptions about the sampling scheme.

When k-space is fully sampled at the Nyquist rate and no regularization is applied, the solution to (2) corresponds to the minimum-variance unbiased estimator (MVUE) of x^*, denoted by $\hat{x}_{\text{MVUE}}$ [65]. Given fully sampled k-space data, this estimate can act as a reference image for evaluating reconstruction error as well as for end-to-end training. Alternatively, a reference image called the root-sum-of-squares (RSS) estimate can be formed by taking the inverse Fourier transform of each coil and subsequently applying the ℓ_2 norm for each pixel across the coil dimension, i.e. $\hat{x}_{\text{RSS}} = \sqrt{\sum_{i=1}^{N_c} |(F^H y_i)|^2}$, where F^H is the Hermitian transpose of F (here the inverse DFT). Although the RSS estimate is a biased estimator, it is often used as it does not make any assumptions about the sensitivity maps, which are not explicitly measured by the MRI system. However, even if solving (2) results in perfect recovery of x^*, there will be a bias when comparing the result to $\hat{x}_{\text{RSS}}$ and thus the RSS and MVUE cannot be directly compared numerically.

2.2 Posterior Sampling

The algorithm we consider is *posterior sampling* [45]. That is, given an observation of the form $y = Ax^* + w$, where $y \in \mathbb{C}^M$, $A \in \mathbb{C}^{M \times N}$, $w \sim \mathcal{N}_c(0, \sigma^2 I)$, and $x^* \sim \mu$, the posterior sampling recovery algorithm outputs $\widehat{x}$ according to the posterior distribution $\mu(\cdot|y)$.

In order to sample from the posterior, we use *Langevin Dynamics* [10]. Assuming we have access to $\nabla_x \log \mu(x|y)$, we can sample from $\mu(x|y)$ by running noisy gradient ascent:

$$x_{t+1} \leftarrow x_t + \frac{\eta_t}{2} \nabla_{x_t} \log \mu(x_t|y) + \sqrt{\eta_t}\zeta_t, \zeta_t \sim \mathcal{N}(0, 1). \tag{3}$$

Prior work [10] has shown that as $t \to \infty$ and $\eta_t \to 0$, Langevin dynamics will correctly sample from $\mu(x|y)$. In practice, vanilla Langevin Dynamics are slow to converge. Hence, the work in [73] proposes *annealed* Langevin Dynamics, where the marginal distribution of x at iteration t is modelled as $\mu_t = \mu * \mathcal{N}(0, \beta_t^2)$ and the generative model is trained to estimate the score function $f(x_t; \beta_t) := \nabla_{x_t} \log((\mu * \mathcal{N}(0, \beta_t^2)(x_t))$. Note that the number of annealing steps T, and the annealing noise strengths $\{\beta_t\}_{t=1}^T$ are hyperparameters that are used to train the score function, and are kept fixed during inference when we perform posterior sampling. Since the distribution of $y|x^*$ is Gaussian in Eqn (2), we obtain $\nabla_{x_t} \log \mu(y|x_t) = \frac{A^H(y - Ax_t)}{\sigma^2}$. We find that it is also helpful to anneal this term, and we set it to $\frac{A^H(y - Ax_t)}{\sigma^2 + \beta_t^2}$. An application of Bayes' rule gives: $\nabla_{x_t} \log \mu(x_t|y) = f(x_t; \beta_t) + \frac{A^H(y - Ax_t)}{\sigma^2 + \beta_t^2}$.

Putting everything together, our final algorithm is: for $x_0 \sim \mathcal{N}_c(0, I)$ and for all $t = 0, \cdots, T-1$,

$$x_{t+1} \leftarrow x_t + \frac{\alpha_t}{2} \left(f(x_t; \beta_t) + \frac{A^H(y - Ax_t)}{\beta_t + \sigma^2} \right) + \sqrt{\alpha_t}\zeta_t, \zeta_t \sim \mathcal{N}(0; I). \tag{4}$$

3 Theoretical Results

Background and Notation. We first introduce background and notation required for our theoretical results. $\|\cdot\|$ refers to the ℓ_2 norm. In this section alone, for simplicity of exposition, we will assume that all matrices and vectors are real valued.

For two probability distributions μ, ν on some normed space Ω, and for any $q \geq 1$, the Wasserstein-q [85, 6] and Wasserstein-∞ [17] distances are defined as:

$$\mathcal{W}_q(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \left(\mathop{\mathbb{E}}_{(u,v)\sim\gamma} [\|u - v\|^q] \right)^{1/q}, \quad \mathcal{W}_\infty(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \left(\gamma\text{-}\operatorname*{ess\,sup}_{(u,v)\in\Omega^2} \|u - v\| \right).$$

where $\Pi(\mu, \nu)$ denotes the set of joint distributions whose marginals are μ, ν. The above definition says that if $\mathcal{W}_\infty(\mu, \nu) \leq \varepsilon$, and $(u, v) \sim \gamma$, then $\|u - v\| \leq \varepsilon$ almost surely.

The $(\varepsilon, \delta)-$*approximate covering number* [45], is defined as the smallest number of ε-radius balls required to cover $1-\delta$ mass under a distribution.

Definition 3.1 ((ε, δ)-approximate covering number)**.** *Let μ be a distribution on $\mathbb{R}^N$. For some parameters $\varepsilon > 0, \delta \in [0,1]$, the* (ε, δ)-approximate covering number *of μ is defined as*

$$\mathrm{Cov}_{\varepsilon,\delta}(\mu) := \min\left\{k : \mu\left[\cup_{i=1}^{k} B(x_i, \varepsilon)\right] \geq 1-\delta, x_i \in \mathbb{R}^N\right\},$$

where $B(x, \varepsilon)$ is the ℓ_2 ball of radius ε centered at x.

Distributional robustness under Gaussian measurements. First, we consider mismatch between the ground-truth distribution, denoted by μ, and the generator distribution, denoted by ν. Prior work [45] has shown that if (i) $\mathcal{W}_q(\mu, \nu) \leq \varepsilon$ for some $q \geq 1$ and (ii) we are given $M \geq O(\log \mathrm{Cov}_{\varepsilon,\delta}(\mu))$ Gaussian measurements, then posterior sampling with respect to ν will recover $x^* \sim \mu$ up to an error of $\varepsilon/\delta^{1/q}$ with probability $1-\delta$. Closeness in Wasserstein distance is a reasonable assumption in certain examples, such as when μ is the distribution of celebrity faces and ν is the distribution of a generator trained on FlickrFaces [51]. However, this assumption is unsatisfactory when we consider distributions of abdominal and brain MR scans, for example, since images of these anatomies look entirely different.

We define the following weaker notion of distance between distributions. Informally, this new definition tells us that ν and μ are "close" if they can each be split into components which are close in $\mathcal{W}_\infty$ distance, such that the close components contain a sufficiently large fraction under ν and μ. Formally, this is defined as:

Definition 3.2 ((δ, α)-$\mathcal{W}_\infty$ distance)**.** *For two probability distributions ν and μ, and parameters $\delta, \alpha \in [0,1]$, the (δ, α)-$\mathcal{W}_\infty$ distance is defined as*

$$(\delta, \alpha)\text{-}\mathcal{W}_\infty(\mu, \nu) := \inf\{\varepsilon \geq 0 :$$
$$\exists \mu', \mu'', \nu', \nu'' \in \mathcal{M}(\mathbb{R}^N) \; s.t. \; \mu = (1-\delta)\mu' + \delta\mu'', \nu = (1-\alpha)\nu' + \alpha\nu'', \mathcal{W}_\infty(\mu', \nu') = \varepsilon.\}$$

Lemma A.1 highlights that this is a strict generalization of Wasserstein distances.

Since the (δ, α)-$\mathcal{W}_\infty$ distance is a generalization of Wasserstein distances, it is not clear that the main Theorem in [45] holds for distributions that are close in this new distance. The following result shows a rather surprising fact: if (δ, α)-$\mathcal{W}_\infty(\mu, \nu) \leq \varepsilon$ then posterior sampling with $m = O\left(\log\left(\frac{1}{1-\alpha}\right) + \log \mathrm{Cov}_{\varepsilon,\delta}(\mu)\right)$ measurements will still succeed with probability $\geq 1 - O(\delta)$.

Theorem 3.3. *Let $\delta, \alpha \in [0,1]$, and $\varepsilon > 0$ be parameters. Let μ, ν be arbitrary distributions over $\mathbb{R}^N$ satisfying (δ, α)-$\mathcal{W}_\infty(\mu, \nu) \leq \varepsilon$. Let $x^* \sim \mu$ and suppose $y = Ax^* + \xi$, where $A \in \mathbb{R}^{M \times N}$ and $\xi \in \mathbb{R}^M$ are i.i.d. Gaussian normalized such that $A_{ij} \sim \mathcal{N}(0, 1/M)$ and $\xi_i \sim \mathcal{N}(0, \sigma^2/M)$, with $\sigma \gtrsim \varepsilon$. Given y and the fixed matrix A, let $\widehat{x}$ be the output of posterior sampling with respect to ν.*

Then for $M \geq O\left(\log\left(\frac{1}{1-\alpha}\right) + \min(\log \mathrm{Cov}_{\sigma,\delta}(\mu), \log \mathrm{Cov}_{\sigma,\delta}(\nu))\right)$, there exists a universal constant $c > 0$ such that with probability at least $1 - e^{-\Omega(M)}$ over A, ξ,

$$\Pr_{x^* \sim \mu, \widehat{x} \sim \nu(\cdot|y)}\left[\|x^* - \widehat{x}\| \geq c(\varepsilon + \sigma)\right] \leq 2\delta + 2e^{-\Omega(M)}.$$

For our running example of ν being a generator trained on brain scans, and μ the distribution of abdominal scans, we can set ν' to be the distribution of our generator restricted to abdominal scans, and we can let μ' be the distribution restricted to "inliers" in μ. This shows that even if our generator places an *exponentially small* probability mass(i.e., $1 - \alpha \ll 1$) on the set of abdominal scans, we can still recover abdominal scans with a *polynomial additive* increase in the number of measurements (i.e., $\log(1/(1-\alpha))$).

Near-optimality under arbitrary measurement processes. The previous result required Gaussian matrices to handle the distribution shift. Our next result shows that for an *arbitrary* measurement process, and assuming that there is no distribution shift between the generator and the ground truth distribution, posterior sampling is almost the best algorithm for this *fixed* measurement process.

Theorem 3.4. *Let $x^* \sim \mu$ and let $y = \mathcal{A}(x^*)$ be measurements generated from x^* for some arbitrary forward operator $\mathcal{A} : \mathbb{R}^N \to \mathbb{R}^M$. Then if there exists an algorithm that uses y as inputs and outputs x' such that $\|x^* - x'\| \leq \varepsilon$ with probability $1-\delta$, then posterior sampling $\widehat{x} \sim \mu(\cdot|y)$ will satisfy $\|x^* - \widehat{x}\| \leq 2\varepsilon$ with probability $\geq 1 - 2\delta$.*

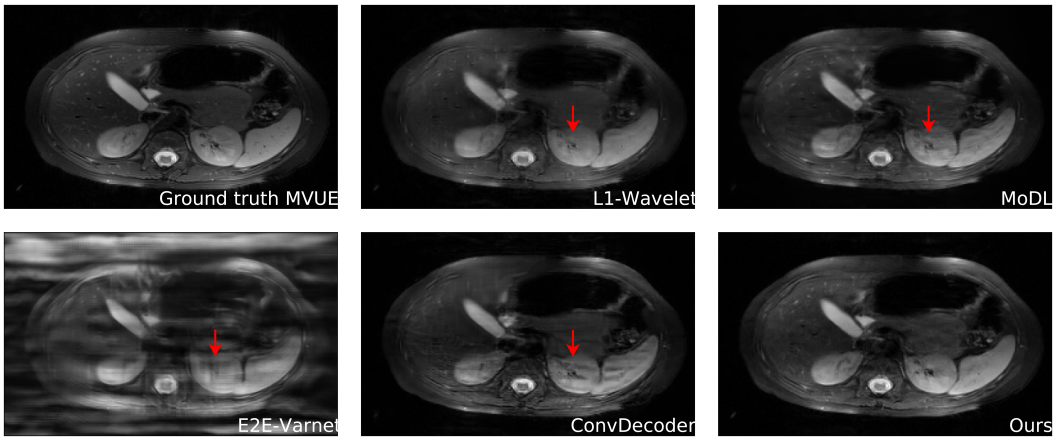

Figure 2: Comparative reconstructions of a 2D abdominal scan with uniform random under-sampling in the horizontal direction at $R = 4$. None of the methods were trained to reconstruct abdomen MRI. Our method uses a score-based generative model trained on brain images (as explained), and obtains excellent qualitative abdomen reconstructions. The red arrows point to artifacts that occur for prior methods and indicate missing details (for ℓ_1-Wavelet and MoDL) or severe artifacts (for E2E-VarNet and the ConvDecoder) in the kidney structure, which makes these scans unacceptable for clinical use [88].

Remark on combining these results. Our theoretical results above show that posterior sampling is (1) highly robust to distribution shift under Gaussian measurements, and (2) accurate with arbitrary measurements without distribution shift. A natural hope would be to combine these two results and show that it is robust to distribution shift under Fourier measurements. Unfortunately, this is *not* true for general distributions: for example, if μ and ν are both random distributions over Fourier-sparse signals, then Fourier measurements will usually give zero information about the signal, so cannot convince the sampler to sample near μ rather than ν.

4 Experimental Results

We perform retrospective under-sampling in all experiments, i.e., given fully-sampled k-space measurements from the NYU fastMRI [54, 89] and Stanford MRI [1] datasets, we apply sampling masks and evaluate the performance of all considered algorithms on the reconstructed data. Depending on scan parameters (e.g., 3D scans for the Stanford knee data in Appendix E), we carefully slice the data in the proper dimension to not commit any inverse crime [32, 71].

We first highlight that an advantage of the proposed approach is the invariance to the sampling scheme during training. In contrast, this is a design choice that must be made for supervised end-to-end methods, which are trained on equispaced, vertical sampling masks, following the fastMRI guidelines [89]. As our results show, this affords us a significant degree of robustness across a wide distribution of sampling masks during inference.

We train our score-based model, NCSNv2 [74], on a small subset of scans from the NYU fastMRI brain dataset. Specifically, we train using T2-weighted images at a field strength of 3T for a total of 14,539 2D training slices. We calculate the MVUE from the fully sampled data and use the ESPIRiT algorithm [81, 42] applied to the fully-sampled central portion of k-space to estimate the sensitivity maps. The backbone network for our model is a RefineNet [56]. Since the generator's output is expected to be complex-valued, we treat the real and imaginary parts as separate image channels. Details about the architectures are given in Appendix F.

We use an ℓ_1-Wavelet regularized reconstruction algorithm [59] as a parallel imaging and compressed sensing baseline. This aims to solve the optimization problem given in (2) with $Q(x) = ||Wx||_1$, where W is a Wavelet transform. We use the publicly available implementation from the BART toolbox [82, 80] and optimize the regularization hyper-parameter using the same subset of samples from the brain dataset that was used to train our method. We find that $\lambda = 0.01$ performs the best on the training data and use this value for all experiments.

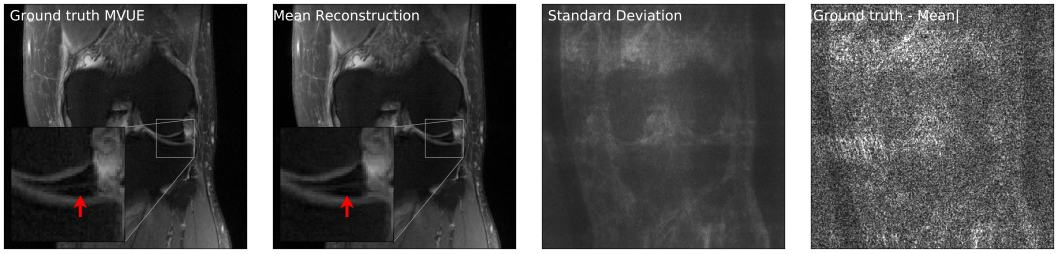

Figure 3: Our method successfully recovers fine details and can provide an estimate of the reconstruction error. The left column shows a knee from the fastMRI dataset, along with an annotated meniscus tear (indicated by red arrow in zoomed inset). Given measurements at an acceleration factor of $R = 4$, we obtain 48 independent reconstructions via posterior sampling. The second column shows the pixel-wise average of reconstructions, the third column shows the pixel-wise standard deviation, and the fourth column shows the magnitude of the error between the ground truth and the mean reconstruction. Note that our generative prior has never seen such a pathology, as it was trained on T2-weighted brain scans.

We consider three different deep learning baselines: MoDL [3], E2E-VarNet [76], and the ConvDecoder architecture [20]. The first two methods, MoDL and E2E-VarNet, use supervised learning and belong to the general class of unrolled optimization methods, while the ConvDecoder is a self-supervised method that optimizes an individual model per scan.

We train the MoDL baseline using a residual network backbone on the same training dataset as our method, at an acceleration factor $R = 4$ and equispaced under-sampling, with a supervised SSIM loss on the MVUE image. For the E2E-VarNet baseline, we use the pretrained model from the official fastMRI repository [2]. This model is trained at a very large scale, using a total number of 92,590 2D scans at acceleration factors $R = \{4, 8\}$, much larger and varied than our training set. Since the model is originally trained on RSS images, we fine tune it for a number of 5 epochs with a supervised SSIM loss on the MVUE images. For the ConvDecoder baseline, we use the architecture for brain data in [20] and optimize the number of fitting iterations on a subset of samples from the training data. We find that 10000 iterations are sufficient to reach a stable average performance at $R = 4$.

We evaluate reconstruction performance using the MVUE of the fully sampled data as a reference image and measure the structural similarity index (SSIM) [86] between the absolute values of the reconstruction and ground-truth MVUE images.

4.1 In-Distribution Performance

In this experiment, we test all models using the same forward model that matches the training conditions for the baselines: vertical, equispaced sampling patterns. Examples of various sampling patterns are shown in Appendix B.

Figure 1 (top row) and Figure 4a show qualitative and quantitative results, respectively, for the case when there is no distributional mismatch between the training and inference sampling patterns. From Figure 4a, we notice that our method surpasses all baselines in the SSIM metric at a broad range of acceleration factors. Despite the performance gap, randomly-chosen samples in Figure 1 (top row) show that our qualitative reconstructions are comparable with those of E2E-VarNet and outperform all other baselines, which still have residual aliasing artifacts. The quantitative drop suffered by E2E-VarNet can be explained by the difference between the RSS and MVUE estimates, and the fact that fine tuning may not cover this gap. This also highlights the importance of qualitative evaluations in medical image reconstruction and the limitations of existing image quality metrics [61].

We find that ℓ_1-Wavelet suffers both qualitatively and quantitatively at high acceleration factors, while the ConvDecoder is also a competitive architecture, but incurs a large computational cost. When benchmarked on an NVIDIA RTX 2080Ti GPU, our method takes 16 minutes and 0.95 GB of memory to reconstruct a high-resolution brain scan, whereas the ConvDecoder takes longer than 80 minutes and 6.6 GB of memory. While our method is limited by the inference time and is not in the range of end-to-end models (where reconstruction takes at most on the order of seconds and 3.5 GB of memory), multiple scans can be reconstructed in parallel due to the reduced memory footprint.

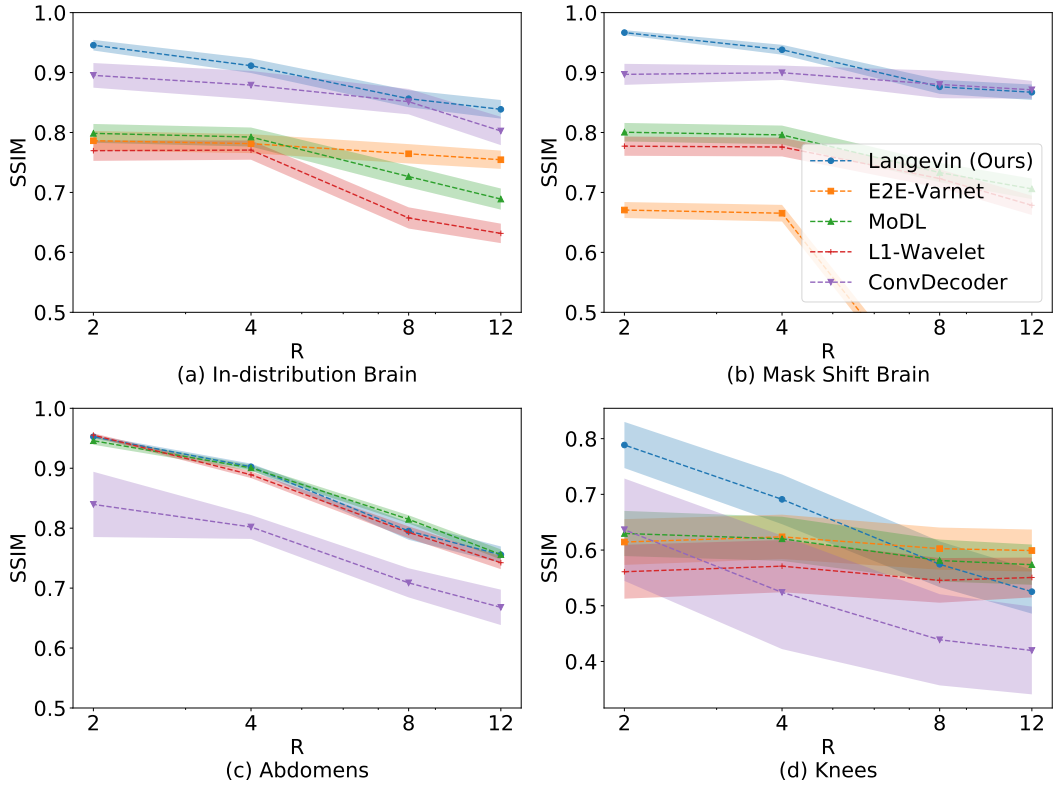

Figure 4: Average test SSIM in various scenarios, across a range of accelerations factors R. Higher R indicates a smaller number of acquired measurements. Our approach shows the best performance and lowest reconstruction variance both in- and out-of-distribution at test-time. Shaded regions indicate 95% confidence intervals.

4.2 Out-of-Distribution Performance

Test-time sampling pattern shifts. Here we consider shifts in the forward sampling operator at test-time, while still evaluating on the same anatomy as the training conditions. We measure robustness by evaluating the average incurred performance loss when the sampling pattern changes. Recall that our proposed approach does not use any explicit information about the sampling pattern P during training, hence we anticipate the highest degree of robustness.

Figure 1 (middle row) shows qualitative reconstructions when the measurements are obtained from an equispaced, horizontal sampling mask, with an acceleration factor $R = 8$. It can be observed that the reconstructions output by both end-to-end methods and the ℓ_1-Wavelet show significant aliasing artifacts that make the scans unacceptable for clinical usage. From the statistical results in in Figure 4b, it can be seen that our method retains peak average performance, as well as a significantly lower variance compared to all other baselines.

Furthermore, this experiment reveals that MoDL is more robust to this type of mask shift when compared to E2E-VarNet, even though it uses a smaller network and does not leverage the same pre-training algorithm. This is explained by the fact that E2E-VarNet does not use external sensitivity map estimates. Instead, this model trains a deep neural network for end-to-end map estimation. While this improves performance on in-distribution samples, the performance drop in Figure 1 (middle row) is strong evidence that accurate sensitivity map estimation is vital for robust generalization, and both our proposed approach and MoDL benefit from the external ESPIRiT algorithm, which is compatible with different sampling patterns.

Test-time anatomy shifts. We now consider the much more difficult problem of reconstructing different anatomies than the ones seen during training. This was previously investigated in [20], which concluded that all methods suffer a drastic shift due to the various changes in scan parameters between body parts. In contrast to prior work, our main finding is that the proposed score-based

model retains a significant degree of robustness under these shifts, and outputs excellent qualitative reconstructions. In some cases, end-to-end methods retain robustness as well .

Figure 4c shows SSIM scores obtained on reconstructed abdominal scans obtained from [1] at an acceleration factor of $R = 4$. This represents both an anatomy and sampling pattern shift, and it can be seen that our method, MoDL and the ℓ_1-Wavelet algorithm retain their competitive advantage, while the ConvDecoder and E2E-VarNet suffer severe performance losses. Figure 2 further shows a qualitative comparison of a reconstructed abdominal scan, with highlighted artifacts.

Finally, Figure 4d shows SSIM scores obtained on fastMRI knee reconstructions, while Figure 1 (bottom row) shows the accompanying qualitative plots. This anatomy is challenging especially because of the poor signal-to-noise ratio conditions, which can be seen even in the ground-truth image. It can be noticed that this is the most severe shift for all methods, but our approach still shows the best performance at $R = 2, 4$ and a significantly lower variance.

4.3 Uncertainty Estimation

Our method can also provide uncertainty estimates for each reconstructed pixel by running multiple reconstruction samplers. For a given observation y, we can obtain independent samples $\widehat{x}_1, \cdots, \widehat{x}_K \sim P(\cdot|y)$, for K sufficiently large. Now, using the conditional mean estimate $\bar{x} = \sum_{i=1}^{K} \widehat{x}_i / K$, we can compute the pixel-wise standard deviation $\sqrt{\sum_{i=1}^{K} |\widehat{x}_i - \bar{x}|^2 / K}$, and this gives an estimate of the error in each pixel. As shown in Fig 3, the pixel-wise standard deviation is a good estimate of the ground truth error $|x^* - \bar{x}|$. Additionally, notice that the reconstructions are able to recover fine details such as the annotated meniscus tear[1] in Fig 3 and predict low uncertainty for these features. We again emphasize that our model was not trained on knee scans. For more examples, see Appendix C.

5 Conclusions

This paper reports the first successful application of the CSGM framework for robust multi-coil MR image reconstruction under realistic sampling conditions, and provides theoretical evidence for the robustness of posterior sampling. Our score-based model was trained on a small subset of brain MRI scans without any explicit information about the sampling scheme. This shows state-of-the-art performance under severe distributional shifts, making our model applicable in a wide range of clinical settings.

Our method shows a considerable degree of generalization to out-of-distribution samples such as abdomen and knee MRI, even when trained exclusively on brain MRI. Notably, these scans were acquired using different MRI vendors with different pulse sequence parameters and at different institutions. Nonetheless, we observed that further hyperparameter tuning leads to improved results, in agreement with recent work [20]. We postulate that adding a small set of diverse training samples to our generative model could further improve robustness, and we hypothesize that these samples may not necessarily be restricted to MR images.

The results presented in this work represent an important step to applying deep learning models in the clinic, as there is a natural variation in sampling, image orientation, receive coils, scanner hardware, and anatomy in clinical practice. Though promising, our initial results were still limited to fast spin-echo imaging only and all data were retrospectively under-sampled. Further study is required to demonstrate prospective performance in a larger body of heterogeneous MRI data. Our method also currently requires a high compute cost at inference time, as well as the need for a pre-trained generative model. Clinical use requires fast reconstruction in addition to fast scanning. Future work should investigate whether score-based models can be trained without a fully-sampled training set as well as investigate approaches to reducing computation time.

[1] `https://discuss.fastmri.org/t/219`

References

[1] http://mridata.org/.

[2] https://github.com/facebookresearch/fastMRI.

[3] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, 38(2):394–405, 2018.

[4] Rushil Anirudh, Jayaraman J Thiagarajan, Bhavya Kailkhura, and Peer-Timo Bremer. Mimicgan: Robust projection onto image manifolds with corruption mimicking. *International Journal of Computer Vision*, pages 1–19, 2020.

[5] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.

[6] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[7] Muhammad Asim, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: mitigating representation error and dataset bias. *arXiv preprint arXiv:1905.11672*, 2019.

[8] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Blind image deconvolution using deep generative priors. *arXiv preprint arXiv:1802.04073*, 2018.

[9] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Solving bilinear inverse problems using deep generative priors. *CoRR, abs/1802.04073*, 3(4):8, 2018.

[10] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Seminaire de probabilités XIX 1983/84*, pages 177–206. Springer, 1985.

[11] Richard G Baraniuk, Volkan Cevher, Marco F Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010.

[12] Richard G Baraniuk and Michael B Wakin. Random projections of smooth manifolds. *Foundations of computational mathematics*, 9(1):51–77, 2009.

[13] Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

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

[15] Ashish Bora, Eric Price, and Alexandros G Dimakis. Ambientgan: Generative models from lossy measurements. *ICLR*, 2:5, 2018.

[16] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592, 2008.

[17] Thierry Champion, Luigi De Pascale, and Petri Juutinen. The ∞-Wasserstein distance: Local solutions and existence of optimal transport maps. *SIAM Journal on Mathematical Analysis*, 40(1):1–20, 2008.

[18] EK Cole, JM Pauly, SS Vasanawala, and F Ong. Unsupervised mri reconstruction with generative adversarial networks. arxiv 2020. *arXiv preprint arXiv:2008.13065*.

[19] Giannis Daras, Joseph Dean, Ajil Jalal, and Alexandros G Dimakis. Intermediate layer optimization for inverse problems using deep generative models. *arXiv preprint arXiv:2102.07364*, 2021.

[20] Mohammad Zalbagi Darestani, Akshay Chaudhari, and Reinhard Heckel. Measuring robustness in deep learning based compressive sensing. *arXiv preprint arXiv:2102.06103*, 2021.

[21] Anagha Deshmane, Vikas Gulani, Mark A Griswold, and Nicole Seiberlich. Parallel mr imaging. *Journal of Magnetic Resonance Imaging*, 36(1):55–72, 2012.

[22] Manik Dhar, Aditya Grover, and Stefano Ermon. Modeling sparse deviations for compressed sensing using generative models. *arXiv preprint arXiv:1807.01442*, 2018.

[23] Mariya Doneva. Mathematical models for magnetic resonance imaging reconstruction: An overview of the approaches, problems, and future research areas. *IEEE Signal Processing Magazine*, 37(1):24–32, 2020.

[24] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[25] Vineet Edupuganti, Morteza Mardani, Shreyas Vasanawala, and John Pauly. Uncertainty quantification in deep mri reconstruction. *IEEE Transactions on Medical Imaging*, 40(1):239–250, 2021.

[26] Yonina C Eldar and Moshe Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Transactions on Information Theory*, 55(11):5302–5316, 2009.

[27] Alyson K Fletcher, Parthe Pandit, Sundeep Rangan, Subrata Sarkar, and Philip Schniter. Plug-in estimation in high-dimensional linear inverse problems: A rigorous analysis. In *Advances in Neural Information Processing Systems*, pages 7440–7449, 2018.

[28] Alyson K Fletcher, Sundeep Rangan, and Philip Schniter. Inference in deep networks in high dimensions. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1884–1888. IEEE, 2018.

[29] Fabian Latorre Gómez, Armin Eftekhari, and Volkan Cevher. Fast and provable admm for learning with generative priors. *arXiv preprint arXiv:1907.03343*, 2019.

[30] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[31] Mark A. Griswold, Peter M. Jakob, Robin M. Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

[32] Matthieu Guerquin-Kern, Laurent Lejeune, Klaas Paul Pruessmann, and Michael Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 31(3):626–636, 2011.

[33] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, 79(6):3055–3071, 2018.

[34] Paul Hand and Babhru Joshi. Global guarantees for blind demodulation with generative priors. In *Advances in Neural Information Processing Systems*, pages 11531–11541, 2019.

[35] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, pages 9136–9146, 2018.

[36] Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. *arXiv preprint arXiv:1705.07576*, 2017.

[37] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. *arXiv preprint arXiv:1810.03982*, 2018.

[38] Reinhard Heckel and Mahdi Soltanolkotabi. Compressive sensing with un-trained neural networks: Gradient descent finds the smoothest approximation. *arXiv preprint arXiv:2005.03991*, 2020.

[39] Chinmay Hegde. Algorithmic aspects of inverse problems using generative models. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 166–172. IEEE, 2018.

[40] Chinmay Hegde, Michael Wakin, and Richard G Baraniuk. Random projections for manifold learning. In *Advances in neural information processing systems*, pages 641–648, 2008.

[41] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Image-adaptive gan based reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3121–3129, 2020.

[42] Siddharth Iyer, Frank Ong, Kawin Setsompop, Mariya Doneva, and Michael Lustig. Sure-based automatic parameter selection for espirit calibration. *Magnetic Resonance in Medicine*, 84(6):3423–3437, 2020.

[43] Mathews Jacob, Jong Chul Ye, Leslie Ying, and Mariya Doneva. Computational mri: Compressive sensing and beyond [from the guest editors]. *IEEE Signal Processing Magazine*, 37(1):21–23, 2020.

[44] Gauri Jagatap and Chinmay Hegde. Phase retrieval using untrained neural network priors. 2019.

[45] Ajil Jalal, Sushrut Karmalkar, Alex Dimakis, and Eric Price. Compressed sensing with approximate priors via conditional resampling.

[46] Ajil Jalal, Liu Liu, Alexandros G Dimakis, and Constantine Caramanis. Robust compressed sensing using generative models. *Advances in Neural Information Processing Systems*, 33, 2020.

[47] Shirin Jalali and Xin Yuan. Solving linear inverse problems using generative models. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 512–516. IEEE, 2019.

[48] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[49] Maya Kabkab, Pouya Samangouei, and Rama Chellappa. Task-aware compressed sensing with generative adversarial networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[50] Akshay Kamath, Sushrut Karmalkar, and Eric Price. Lower bounds for compressed sensing with generative models. *arXiv preprint arXiv:1912.02938*, 2019.

[51] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

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

[53] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[54] Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, Hersh Chandarana, et al. fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, 2020.

[55] Qi Lei, Ajil Jalal, Inderjit S Dhillon, and Alexandros G Dimakis. Inverting deep generative models, one layer at a time. In *Advances in Neural Information Processing Systems*, pages 13910–13919, 2019.

[56] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1925–1934, 2017.

[57] Zhaoqiang Liu, Selwyn Gomes, Avtansh Tiwari, and Jonathan Scarlett. Sample complexity bounds for 1-bit compressive sensing and binary stable embeddings with generative priors. *arXiv preprint arXiv:2002.01697*, 2020.

[58] Zhaoqiang Liu and Jonathan Scarlett. Information-theoretic lower bounds for compressive sensing with generative models. *arXiv preprint arXiv:1908.10744*, 2019.

[59] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[60] Morteza Mardani, Enhao Gong, Joseph Y Cheng, Shreyas S Vasanawala, Greg Zaharchuk, Lei Xing, and John M Pauly. Deep generative adversarial neural networks for compressive sensing mri. *IEEE transactions on medical imaging*, 38(1):167–179, 2018.

[61] Allister Mason, James Rioux, Sharon E Clarke, Andreu Costa, Matthias Schmidt, Valerie Keough, Thien Huynh, and Steven Beyea. Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of mr images. *IEEE transactions on medical imaging*, 39(4):1064–1072, 2019.

[62] Dominik Narnhofer, Kerstin Hammernik, Florian Knoll, and Thomas Pock. Inverse gans for accelerated mri reconstruction. In *Wavelets and Sparsity XVIII*, volume 11138, page 111381A. International Society for Optics and Photonics, 2019.

[63] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *arXiv preprint arXiv:2005.06001*, 2020.

[64] Parthe Pandit, Mojtaba Sahraee-Ardakan, Sundeep Rangan, Philip Schniter, and Alyson K Fletcher. Inference with deep generative priors in high dimensions. *arXiv preprint arXiv:1911.03409*, 2019.

[65] Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[66] Shuang Qiu, Xiaohan Wei, and Zhuoran Yang. Robust one-bit recovery via relu generative networks: Improved statistical rates and global landscape analysis. *arXiv preprint arXiv:1908.05368*, 2019.

[67] Saiprasad Ravishankar and Yoram Bresler. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041, 2011.

[68] Saiprasad Ravishankar and Jeffrey A. Fessler. Data-driven models and approaches for imaging. In *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, page MW2C.4. Optical Society of America, 2017.

[69] Sebastian Rosenzweig, Hans Christian Martin Holme, Robin N Wilke, Dirk Voit, Jens Frahm, and Martin Uecker. Simultaneous multi-slice mri using cartesian and radial flash and regularized nonlinear inversion: Sms-nlinv. *Magnetic resonance in medicine*, 79(4):2057–2066, 2018.

[70] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503, 2017.

[71] Efrat Shimron, Jonathan Tamir, Ke Wang, and Michael Lustig. Subtle inverse crimes: Naively using publicly available images could make reconstruction results seem misleadingly better! *Proceedings of The ISMRM*, 2021.

[72] Daniel K Sodickson and Warren J Manning. Simultaneous acquisition of spatial harmonics (smash): fast imaging with radiofrequency coil arrays. *Magnetic resonance in medicine*, 38(4):591–603, 1997.

[73] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, pages 11918–11930, 2019.

[74] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *arXiv preprint arXiv:2006.09011*, 2020.

[75] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[76] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational networks for accelerated mri reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 64–73. Springer, 2020.

[77] Anuroop Sriram, Jure Zbontar, Tullie Murrell, C. Lawrence Zitnick, Aaron Defazio, and Daniel K. Sodickson. Grappanet: Combining parallel imaging with deep learning for multi-coil mri reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[78] Anuroop Sriram, Jure Zbontar, Tullie Murrell, C Lawrence Zitnick, Aaron Defazio, and Daniel K Sodickson. Grappanet: Combining parallel imaging with deep learning for multi-coil mri reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14315–14322, 2020.

[79] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[80] Martin Uecker, Christian Holme, Moritz Blumenthal, Xiaoqing Wang, Zhengguo Tan, Nick Scholand, Siddharth Iyer, Jon Tamir, and Michael Lustig. mrirecon/bart: version 0.7.00, March 2021.

[81] Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala, and Michael Lustig. Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. *Magnetic resonance in medicine*, 71(3):990–1001, 2014.

[82] Martin Uecker, Frank Ong, Jonathan I Tamir, Dara Bahri, Patrick Virtue, Joseph Y Cheng, Tao Zhang, and Michael Lustig. Berkeley advanced reconstruction toolbox. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 23, 2015.

[83] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.

[84] Shreyas S. Vasanawala, Marcus T. Alley, Brian A. Hargreaves, Richard A. Barth, John M. Pauly, and Michael Lustig. Improved pediatric mr imaging with compressed sensing. *Radiology*, 256(2):607–616, 2010. PMID: 20529991.

[85] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[86] Zhou Wang and Alan C Bovik. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 2(1):1–156, 2006.

[87] Bihan Wen, Saiprasad Ravishankar, Luke Pfister, and Yoram Bresler. Transform learning for magnetic resonance image reconstruction: From model-based learning to building neural networks. *IEEE Signal Processing Magazine*, 37(1):41–53, 2020.

[88] Catherine Westbrook. *Handbook of MRI technique*. John Wiley & Sons, 2014.

[89] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An open dataset and benchmarks for accelerated MRI. 2018.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Section 3, 4, 4.3
   (b) Did you describe the limitations of your work? [Yes] See Section 2 regarding the differences between RSS and MVUE images. As we need to model sensitivity maps of the MR machine, we are unable to get RSS images, although there is no clear answer as to whether there is a best choice among MVUE and RSS images. Additionally, Section 4 details the additional memory requirements and running time of our method in comparison to end-to-end methods.
   (c) Did you discuss any potential negative societal impacts of your work? [No] Our work is very specific to MR imaging, and we are unaware of ethical concerns in this space as we used publicly available and anonymized MRI data.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes] Yes we have read the ethics review guidelines.
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] In the statements of Theorem 3.3, Theorem 3.4.
   (b) Did you include complete proofs of all theoretical results? [Yes] In Appendix A
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Our GitHub link in Appendix F contains all relevant information

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] In Appendix F

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] All quantitative plots have 95% confidence intervals.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Yes, in section 4 and Appendix F

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [Yes] Our algorithm uses the NCSNv2 generative model, NYU fastMRI dataset, and Stanford MRI datasets, all of which have been cited.

(b) Did you mention the license of the assets? [Yes] In Appendix F

(c) Did you include any new assets either in the supplemental material or as a URL? [Yes] To the best of our knowledge, we are the first to train generative models for complex valued MR scans, and we include this in the GitHub link in Appendix F

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] We used the NYU fastMRI and Stanford MRI datasets, both of which have this covered in their terms of agreement.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] We used the NYU fastMRI and Stanford MRI datasets, both of which have been anonymized.

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] We didn't do any human subject experiments

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] We didn't do any human subject experiments

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] We didn't do any human subject experiments