CVPR
#8301

Posterior Sampling in Accelerated MRI using Denoising Diffusion Probabilistic Models

Anonymous CVPR submission

Paper ID 8301

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis ac risus vitae dui semper pulvinar. Nunc ex ligula, placerat at turpis sit amet, facilisis vulputate mauris. Donec convallis nunc dui, vel dignissim nunc porttitor ac. Donec consectetur neque ante. Vivamus ultricies nisi porttitor velit suscipit convallis. Aliquam in nisl elit. Donec pretium tellus ut nibh rutrum luctus luctus non libero. Maecenas ut libero porttitor, tempus nibh vel, tincidunt tellus. Curabitur semper enim neque, eu molestie est porttitor ac. Vivamus nec lorem felis.

1. Introduction

Accelerated Magnetic Resonance Imaging (MRI) aims to reduce acquisition time to improve patient comfort, lower costs and indirectly reduce motion artefacts. One way of achieving acceleration is undersampling the data in acquisition space, i.e., k-space. As k-space and image space are connected by a spatial Fourier transform, undersampling the acquisition violates the Nyquist sampling rate and creates an underdetermined inverse problem. Hence, a direct inverse Fourier transform on the undersampled k-space data leads to images with artefacts, rendering reconstructed images unsuited for further analysis.

Conventional methods such as parallel imaging [11, 25, 34] and compressed sensing [3, 18] have been applied successfully to the problem, though up to limited acceleration factors. Over the last decade, new advancements within deep learning have paved the way for further progress within accelerated MRI, and new deep learning-based models have been proposed [12, 22, 27, 30, 35]. These methods have improved the state-of-the-art drastically in terms of the reconstruction quality and possible undersampling rates. However, following the conventional methods, they focused on reconstructing a point estimate from the undersampled measurements, neglecting the unavoidable uncertainty in the reconstruction due to the missing data. In accelerated MRI, as the inverse problem is underdetermined,

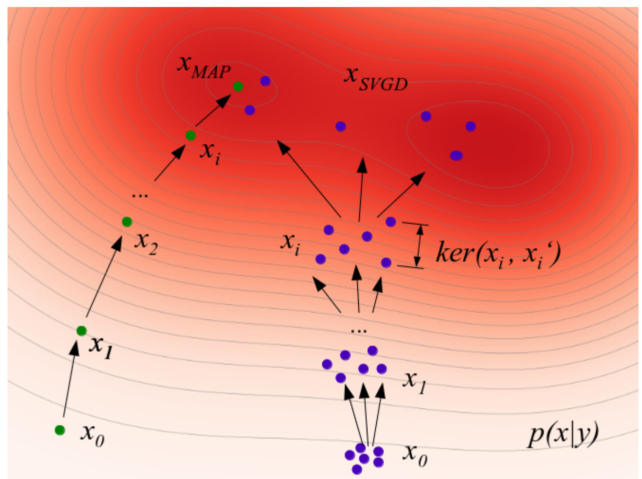

Figure 1. **PLACEHOLDER** - Similar figure here but fused with DDPM image

infinite numbers of different possible reconstructions exist for the given data. In a probabilistic sense, the distribution of these possible reconstructions can be written as a posterior distribution, i.e., the posterior of images given the k-space data. Then, reconstructing a single image boils down to a maximum-a-posteriori (MAP) estimation. This characterises the posterior only by a single image, ignoring the uncertainty and possible multi-modal posterior distribution.

One can normally approach the uncertainty estimation problem of MRI in two main manners. The first is the estimation of the model uncertainty, i.e., epistemic uncertainty, which is arising from the uncertainty in the model parameters, and in the practical case, often limited by the amount of data [10]. Methods with this approach capture the uncertainty in the trained model, but not the uncertainty due to the missing part of k-space. The second approach for uncertainty estimation focuses on the inversion uncertainty due to the missing data, where the aim is to capture the possibility of different reconstructions. The benefit of reconstructing multiple plausible samples potentially are two-fold. First, the different reconstructions can be fed into downstream tasks where, for example, uncertainty aware delineations of the anatomy can be found by using segmentation from multiple reconstructions. Similarly, diagnostic decisions can be made considering multiple images, reducing the chance of basing decisions on a faulty representation in a single image. Second, multiple reconstructions allow the uncertainty to be quantified on the image level, which can provide a measure of reconstruction quality, and potentially predict gross reconstruction errors, such as hallucinations [19].

In this paper, we propose a method to draw multiple samples from the posterior distribution of the images given the k-space measurements. Though such methods have been proposed previously [1, 8, 31, 32], they suffer from either poor reconstruction performance or long run times. We propose to use denoising diffusion probabilistic models (DDPM) [13] allowing to reconstruct a set of samples images instead of a point estimate. As originally proposed, DDPM is an unconditional image generation method that learns to sample $p(x)$. However, in our method, we use a conditional extension, learning the posterior distribution $p(x|y)$. While there already are proposed methods to solve undersampled MRI reconstruction using DDPM or, akin, score based methods [5, 7, 23, 29, 36], our method consider combined multi-coil reconstruction by learning the sensitivity maps, without coil-compression or reconstructing data coil-by-coil, resulting in more faithful reconstructions.

The main contributions of the paper are as follows:
1) We propose to use DDPM to solve the inverse problem of undersampled multicoil MRI and show that our method can draw samples from the posterior.
2) In presence of undersampled data, we propose to learn the added noise conditioned on sampled data in the framework of DDPM.
3) We present extensive experiments using our method on the fastMRI dataset and compare with several other "state-of-the-art" methods.

2. Related work

2.1. Denoising Diffusion Probabilistic Models (DDPM)

DDPM [13] are unconditional generative models based on diffusion models proposed Sohl-Dickstein *et al.* [28] which propose to express a temporal diffusion process for image generation. The diffusion process can be seen as a Markov chain where each step consists of added Gaussian noise in the forward process and removal of such in the reversed process. In more detail, in the forward process, at each time step t, x_t is conditioned on x_{t-1} as

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \tag{1}$$

where a variance schedule $\beta_t \in [0, 1]$ is pre-defined. At large t the resulting x_t will be a Gaussian with isotropic variance. In order to sample a image x_0 from an isotropic Gaussian DDPM suggest to learn the reversed process, *i.e*.

$$q(x_{t-1}|x_t) = N(x_{t-1}; \mu_t, \hat{\beta}_t I). \tag{2}$$

In order to learn the reversed diffusion process DDPM use the variational bound on the negative log likelihood between the forward and reversed process. **More to write...**

2.2. Posterior Sampling and DDPM methods for Accelerated MRI

There is two primary forms of uncertainty in machine learning, epistemic and aleatoric uncertainty. To the first end, epistemic, or model uncertainty, for instance, can be modelled by Bayesian networks, performing multiple forward passes while sampling different model weights during inference [9, 10, 17] or alternatively, using sampling free methods [24]. These techniques have been applied to MR reconstruction [14, 20]; however, they do not address the ill-posed problem of the inversion due to missing data.

Second, aleatoric uncertainty, or data uncertainty, in reconstruction has also been addressed with numerous techniques. One of these is to use a heteroscedastic model, i.e. predicting pixel-wise variance alongside a mean prediction, effectively assuming an independent Gaussian posterior for each pixel [38]. However, the assumption that pixels are not inherent Gaussian distributed and ignoring the pixel-wise dependencies is a too unrealistic, rendering sampling from the posterior impossible.

Alternatively, MC-DDPM [36] suggest using a conditioned DDPM to solve single-coil MRI reconstruction only working in k-space. Likewise, Peng *et al.* [23], propose to use DDPM. Contrarily, they train an unconditional model to learn $p(x)$ and use measured data during reversed sampling process. Both methods don't propose how to solve multi-coil reconstruction and in their current form are very limited in clinical practice where multi-coil data is the norm. AdaDiff [7] too use diffusion priors to alternate between prior steps, learned using DDPM, and data consistency steps to reconstruct undersampled MRI. However, additionally AdaDiff use adversarial training to construct more truthfully reconstructions. Hence, AdaDiff only reconstruct a point estimate which does not estimate the posterior.

Related, Chung *et al.* [5] and Song *et al.* [29], propose to use score based methods to learn the prior step in a defined reversed denoising process. In difference to DDPM based methods score based methods learn the gradient towards the posterior point estimate instead of learning the added noise like DDPM. Regarding multi-coil reconstruction, Chung *et al.* don't show improvements and Song *et al.* only propose a solution for singe-coil data.

Another method for posterior sampling is to use a variational autoencoder (VAE) [16] to encode the zero-filled image into the latent space, and take local latent samples to decode into image space, while also improving the image quality [8]. Similarly, Tezcan *et al*. [32], use a VAE, but take samples from a global latent distribution adhering to the measured data and decode images from these by explicitly incorporating the measured data into the images. Though this brings better sample quality, it comes at a high computational cost. Both papers do not compare to the state-of-the-art in the public challenge dataset fastMRI [37].

3. Method

In this section, we will first briefly formulate the accelerated MRI reconstruction problem. Then, we will present our proposed method using DDPM for posterior sampling.

3.1. Accelerated MRI

In accelerated MR, the forward imaging process is modelled as

$$y = \mathbf{E}x + \omega, \tag{3}$$

with $x \in \mathbb{C}^N$ as the fully sampled image, $y \in \mathbb{C}^{\kappa M}$ as the measured undersampled k-space data (M<N), $E : \mathbb{C}^N \rightarrow \mathbb{C}^{\kappa M}$ as the linear undersampled encoding operation and $\omega \sim N(0, I\sigma_{ksp})$ as the noise in k-space. We assume isotropic noise without loss of generality, as data can be whitened otherwise and for training and sampling we neglect ω. The encoding matrix is composed of the coil sensitivities S of κ coils, the Fourier transform F and finally the undersampling operation U. As in the general case E is not full rank, the problem has infinitely many solutions.

From a Bayesian perspective [2], one can write the log posterior of the images given the k-space data as $\log p(x|y) = \log p(y|x) + \log p(x)$, where the first term corresponds to the the data likelihood given as $\log p(y|x) = -\frac{1}{2\sigma^2}||y - Ex||$ due to the Gaussian noise in Eq. 3. The second term, $\log p(x)$ is the prior term, which represents our a-priori belief about the data and can be be explicitly chosen [18] or learned [33].

3.2. DDPM for Accelerated MRI

First assume a denoising diffusion process as a Markov chain in k-space from $x = \hat{x}_0 + Ux_0$ to $x_T = Ux_0 + \epsilon$, where $\hat{x}_0 = U^T x_0$ and $\epsilon \sim N(\mu_{Ux_0}, \sigma^2_{Ux_0})$. Differently to normal DDPM, we use an adaptive Gaussian distribution based on mean and standard deviation of the measured data. Each step in the diffusion process can be divided into two parts. First a step in the unsampled followed by a data consistency step ensuring conditional reconstruction. For each step the the first part in the forward process is defined for the unsampled data as

$$q(\hat{x}_t|x_{t-1}) = N(x_t; \sqrt{(1-\beta_t)}x_{t-1}, \beta_t I) \tag{4}$$

where a predefined variance schedule $\beta_t \in [0, 1]$ is considered. Since each step is purely a Gaussian we can condition on x_0 and with $\hat{\alpha}_t = \prod_{s=1}^{t}(1-\beta_t)$, equation 4 can be written conditioned on x_0 as

$$q(\hat{x}_t|x_0) = N(x_t; \sqrt{(\hat{\alpha}_t)}x_{t-1}, (1-\hat{\alpha}_t)I). \tag{5}$$

To sample an arbitrary $\hat{x}_t$ we can then simply use $\hat{x}_{t-1} = \sqrt{\hat{\alpha}_t}\hat{x}_0 + \frac{\beta_t}{\sqrt{1-\alpha_t}}\epsilon$. This helps during training as seen in algorithm .1 where in each forward and backward step a time point is first sampled and then loss is calculated for that step. A reversed process step, $x_{t-1} \rightarrow x_t$, for the unsampled data can then be written as

$$q(\hat{x}_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \beta_t I), \tag{6}$$

which also can be written as the simplified step wise version $\hat{\mu}_t = \frac{1}{\sqrt{\alpha_t}}x_t + \frac{\beta_t}{\sqrt{1-\hat{\alpha}_t}}\epsilon_\theta(x_t, t)$ and $\hat{x}_t = \mu_t + \sqrt{beta_t}z$, where $z \sim N(\mu_{Ux_0}, \sigma^2_{Ux_0})$. Since $x_T = Ux_0 + \epsilon$ the reversed process is conditioned on the measured data Ux_0, this is used later during both training and inference. After one step in the diffusion process we propose to use a data consistency step as

$$x_{t-1} = U^T\hat{x}_{t-1} + Ux_0. \tag{7}$$

$\mu_\theta(x_t, t)$ predicts the mean for the posterior distribution and similar to the DDPM paper we suggest to learn the added noise ϵ, instead of the true mean $\sqrt{\hat{\alpha}_t}x_0 + \sqrt{(1-\hat{\alpha}_t)}\epsilon$, and then calculate the mean with $\sqrt{\hat{\alpha}_t}x_0 + \sqrt{(1-\hat{\alpha}_t)}\omega_\theta(x_t, t)$. Where $\epsilon_\theta(x_t, t)$ is the output of the model. During training, time-steps t are randomly sampled and the simplified loss function to minimise, described further in the DDPM paper, is defined as

$$L = \mathbb{E}_{t, x_0, \epsilon}\Big[||\epsilon - \epsilon_\theta(\sqrt{\hat{\alpha}_t}x_0 + \sqrt{1-\hat{\alpha}_t}\epsilon, t)||\Big] \tag{8}$$

Notice that, differently from the original DDPM loss function, we use L1 norm instead of L2 and this is due to the k-space values tend to be small for our dataset.

So far, our method is not so different than original DDPM except that we use data constancy steps, include measured data in the diffusion process and diffusion process is defined in k-space. One other difference, not yet mentioned, is that we learn to predict the noise in coil-combined image space. Experimentally, we have seen that denoising in image space is easier than denoising in k-space and combining the coil to image space let the model get information across coils and without computational expensive coil-wise prediction. We propose to use same architecture as DDPM in the image domain and denoising each step in image space. This could be written as

$$\epsilon_\theta(x_t, t) = \mathcal{F}\mathcal{E}_{\theta_2}M_{\theta_1}(\mathcal{R}_{\theta_2}\mathcal{F}^{-1}x_t, t). \tag{9}$$

Here, M_θ is the same autoencoder network architecture proposed in DDPM, $\mathcal{F}$ and $\mathcal{F}^{-1}$ is the normal and inverse discrete Fourier operator and $\mathcal{E}$ respectively $\mathcal{R}$ is the coil expand and reduce operator used for notation simplification. The expand operation $\mathcal{E}_{\theta_2} = (\hat{S}_1 x, \hat{S}_2 x, ..., \hat{S}_N x)$ use learned coil sensitivity maps to predict the S coil operator. In the same way, the reduce operator $\mathcal{R}_{\theta_2} = \sum_{j=1}^{N} \hat{S}_j^* x_j$ is a simplification of the S^{-1} coil operator. Here N denotes number of coils.

Similar to Sriram *et al.* [30] we propose to learn the sensitivity maps with a shallow UNet [26]. We predict each coil sensitivity map, $\hat{S}_n$, coil-wise and is trained jointly with the denoising network to minimise the loss in Equation 8. We also normalise the sensitivity maps to ensure $\sum_{j=1}^{N} \hat{S}S^* = I$. Maybe add that since we learn the noise it's not necessarily the sensitivity maps rather arbitrary mapping from several coils to one and back.

The pseudo-code for our method can be followed as

Algorithm .1 Training and Sampling

procedure TRAIN(X,U)
  while ! validation loss convergence **do**
    $t \sim [0, 1000]$
    $x_0 \sim X$
    $\epsilon \sim N(\mu_{Ux_0}, \sigma^2_{Ux_0})$
    $\hat{x}_{t-1} = \sqrt{\hat{\alpha}_t}\hat{x}_0 + \frac{\beta_t}{\sqrt{1-\alpha_t}}\epsilon$
    $x_{t-1} = U^T \hat{x}_{t-1} + Ux_0$
    $\hat{\epsilon} = \mathcal{F}\mathcal{E}_{\theta_2} M_{\theta_1}(\mathcal{R}_{\theta_2}\mathcal{F}^{-1}x_t, t)$ ▷ Forward Pass
    $L = ||U^T\epsilon - U^T\hat{\epsilon}||$ ▷ Calculate Loss
    $\theta = \theta - \lambda\nabla_\theta L$ ▷ Backward Pass and Optimisation step
  end while
end procedure

procedure SAMPLE(Ux_0)
  $\epsilon \sim N(\mu_{Ux_0}, \sigma^2_{Ux_0})$
  for t = T,...,1 **do**
    $z \sim N(\mu_{Ux_0}, \sigma^2_{Ux_0})$
    $\epsilon_\theta = \mathcal{F}\mathcal{E}_{\theta_2} M_{\theta_1}(\mathcal{R}_{\theta_2}\mathcal{F}^{-1}x_t, t)$ ▷ Forward Pass
    $\hat{x}_{t-1} = \frac{1}{\alpha_t}x_t - \frac{\beta_t}{\sqrt{1-\hat{\alpha}_t}}\epsilon_\theta$ ▷ Reversed Step
    $x_{t-1} = U^T \hat{x}_{t-1} + Ux_0.$ ▷ Ensure Data Consistency
  end for
end procedure

Notice that the measurement values $x \in \mathcal{C}$, however, M_θ, $\mathcal{E}_theta$ and $\mathcal{R}_theta$ act $\in \mathcal{R}$. We solve this by concatenating the real and imaginary parts of the input in channel dimension. We recognise that this is a simplification but is used due to memory constraints. For further work, complex convolutions as in Cole *et al.* [6] will be explored.

3.3. Implementation Details

We use the same architecture as in Nichol and Dhhariwal [21] for the DDPM model, *e.g.* an UNet network with multi-head attention layers and where time-step t is injected to the hidden layers with GroupNorm. We set input channels to 64 to accommodate for the sensitivity predictor network and kept residual blocks to 3. The sensitivity predictor network was design following E2E VarNet [30], using a shallow UNet with 16 input channels and 4 down sampling layers. The network is trained with the proposed weighted loss in Eq. (8) with adam optimiser [15] and a learning rate of 10^{-4}. We implemented our method using PyTorch on a NVIDIA A100 GPU. We trained our method for xx epochs which took xx time.

We decided to stick to the original DDPM model instead of the improved DDPM [21] as a proof of concept. An extension to improved DDPM is straight forward and should not introduce specific problems related to MRI.

4. Results

4.1. Experiments, Dataset and Baselines

We evaluate our proposed method using T2 weighted brain images, a subset of the publicly available fastMRI dataset [37]. The dataset contains a large number of fully sampled raw k-space data and has lately become commonly used for evaluating methods for accelerated MRI. The subset used is all T2 weighted images from the fastMRI official train and validation dataset and hence reported numbers are not directly comparable to competition results. The complete set of fastMRI was not used due to computational limits. We use an equispaced cartesian undersampling mask with an undersampling factor of eight, as proposed in the fastMRI work. We choose to evaluate an undersampling factor of eight over four considering the imposed higher uncertainty and a high tendency for hallucinations when point estimates are made [4].

We compare our method with the ...

5. Conclusion

In this paper we proposed a SVGD based method that uses approximate gradients from the VarNet reconstruction and a latent space for its kernel to sample from the posterior of the MR inverse problem. The method yielded good reconstructions and uncertainty estimations, showing promise for MRI as well as other inverse problems.

Acknowledgement

-

References

[1] Jonas Adler and Ozan Öktem. Deep bayesian inversion. *arXiv preprint arXiv:1811.05910*, 2018. 2

[2] Daniela Calvetti and Erkki Somersalo. Inverse problems: From regularization to bayesian inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(3):e1427, 2018. 3

[3] Rick Chartrand and Wotao Yin. Iteratively reweighted algorithms for compressive sensing. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3869–3872, 2008. 1

[4] Akshay S Chaudhari, Christopher M Sandino, Elizabeth K Cole, David B Larson, Garry E Gold, Shreyas S Vasanawala, Matthew P Lungren, Brian A Hargreaves, and Curtis P Langlotz. Prospective deployment of deep learning in mri: a framework for important considerations, challenges, and recommendations for best practices. *Journal of Magnetic Resonance Imaging*, 54(2):357–371, 2021. 4

[5] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical Image Analysis*, page 102479, 2022. 2

[6] Elizabeth Cole, Joseph Cheng, John Pauly, and Shreyas Vasanawala. Analysis of deep complex-valued convolutional neural networks for mri reconstruction and phase-focused applications. *Magnetic resonance in medicine*, 86(2):1093–1109, 2021. 4

[7] Salman UH Dar, Şaban Öztürk, Yilmaz Korkmaz, Gokberk Elmas, Muzaffer Özbey, Alper Güngör, and Tolga Çukur. Adaptive diffusion priors for accelerated mri reconstruction. *arXiv preprint arXiv:2207.05876*, 2022. 2

[8] Vineet Edupuganti, Morteza Mardani, Shreyas Vasanawala, and John Pauly. Uncertainty quantification in deep mri reconstruction. *IEEE Transactions on Medical Imaging*, 40(1):239–250, 2020. 2, 3

[9] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 2

[10] Ethan Goan and Clinton Fookes. Bayesian neural networks: An introduction and survey. In *Case Studies in Applied Bayesian Data Science*, pages 45–87. Springer, 2020. 1, 2

[11] 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: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002. 1

[12] 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. 1

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 2

[14] Shi Hu, Nicola Pezzotti, and Max Welling. Learning to predict error for mri reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 604–613. Springer, 2021. 2

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[16] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2

[17] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017. 2

[18] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing mri. *IEEE signal processing magazine*, 25(2):72–82, 2008. 1, 3

[19] Matthew J Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, et al. Results of the 2020 fastmri challenge for machine learning mr image reconstruction. *IEEE transactions on medical imaging*, 40(9):2306–2317, 2021. 2

[20] Dominik Narnhofer, Alexander Effland, Erich Kobler, Kerstin Hammernik, Florian Knoll, and Thomas Pock. Bayesian uncertainty estimation of learned variational mri reconstruction. *arXiv preprint arXiv:2102.06665*, 2021. 2

[21] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021. 4

[22] Arghya Pal and Yogesh Rathi. A review of deep learning methods for mri reconstruction. *arXiv preprint arXiv:2109.08618*, 2021. 1

[23] Cheng Peng, Pengfei Guo, S Kevin Zhou, Vishal Patel, and Rama Chellappa. Towards performant and reliable undersampled mr reconstruction via diffusion model sampling. *arXiv preprint arXiv:2203.04292*, 2022. 2

[24] Janis Postels, Francesco Ferroni, Huseyin Coskun, Nassir Navab, and Federico Tombari. Sampling-free epistemic uncertainty estimation using approximated variance propagation. In *ICCV*, 2019. 2

[25] 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. 1

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4

[27] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for mr image reconstruction. In *International Conference on Information Processing in Medical Imaging*, pages 647–658. Springer, 2017. 1

[28] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using

nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 2

[29] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*, 2021. 2

[30] 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. 1, 4

[31] He Sun and Katherine L Bouman. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. *arXiv preprint arXiv:2010.14462*, 9, 2020. 2

[32] Kerem C Tezcan, Christian F Baumgartner, and Ender Konukoglu. Sampling possible reconstructions of undersampled acquisitions in mr imaging. *arXiv preprint arXiv:2010.00042*, 2020. 2, 3

[33] Kerem C. Tezcan, Christian F. Baumgartner, Roger Luechinger, Klaas P. Pruessmann, and Ender Konukoglu. Mr image reconstruction using deep density priors. *IEEE Transactions on Medical Imaging*, 38(7):1633–1642, 2019. 3

[34] 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. 1

[35] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 514–517. IEEE, 2016. 1

[36] Yutong Xie and Quanzheng Li. Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. *arXiv preprint arXiv:2203.03623*, 2022. 2

[37] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018. 3, 4

[38] Zizhao Zhang, Adriana Romero, Matthew J Muckley, Pascal Vincent, Lin Yang, and Michal Drozdzal. Reducing uncertainty in undersampled mri reconstruction with active acquisition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2049–2058, 2019. 2