

REPORT ON THE PAPER

'SOLVING INVERSE PROBLEMS IN MEDICAL IMAGING WITH SCORE-BASED GENERATIVE MODELS'

Malek Ben Alaya
 Seminar on Machine Learning
 Winter Semester 2022-2023

INTRODUCTION

Image reconstruction from undersampled measurements is an important inverse problem in medical imaging. Supervised methods that try to solve this problem exist. These methods assume a fixed measurement process at training time and therefore fail to generalize to different measurement processes at test time. To mitigate this problem, the studied work proposes a new unsupervised method that solves linear inverse problems in context of medical imaging leveraging score-based generative models.

1 BACKGROUND

1.1 LINEAR INVERSE PROBLEMS

In a linear inverse problem, we are given an observation signal $y \in \mathbb{R}^m$ and the goal is to recover the original signal $x \in \mathbb{R}^n$, where the two signals satisfy the following underlying equation:

$$y = Ax + \epsilon, \quad (1)$$

where $A \in \mathbb{R}^{m \times n}$ is a measurement matrix and $\epsilon \in \mathbb{R}^m$ is a noise vector. We note that usually we have $n > m$. This makes the problem ill-posed and therefore makes the reconstruction challenging. In the context of medical imaging, the original signal to be reconstructed x is an image. As can be seen from Fig. 1, for MRI, the observation y is the subsampled k-space, where the fully sampled k-space is

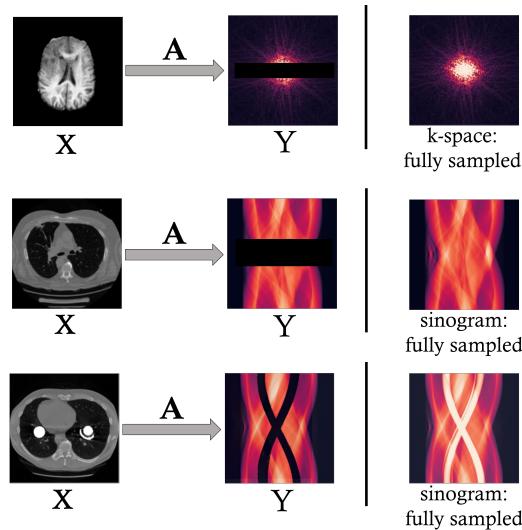


Figure 1: Examples of linear measurement processes in medical imaging. (Top) Medical Resonance Imaging (MRI). (Middle) Computed Tomography (CT). (Bottom) Metal Artifact Removal (MAR).

obtained through applying the Radon transformation to the original image x . As mentioned earlier, y being the subsampled k-space makes the reconstruction problem challenging. For the case where y is the fully sampled k-space, we can easily obtain a near perfect reconstruction of x by applying the inverse Radon transformation to y . The same thing applies for CT, with the only difference to MRI is that y in CT is the subsampled sinogram, where the fully sampled sinogram is obtained through applying the spatial Fourier

transformation to the original image x . Things are a bit different for MAR, where the idea is not to reconstruct the original image x which goes through the measurement process, but to remove the metallic objects from the original image, i.e reconstruct the original image without the metallic objects in it.

2 SOLVING INVERSE PROBLEMS WITH SCORE-BASED GENERATIVE MODELS

2.1 SCORE-BASED GENERATIVE MODELS: UNCONDITIONAL SAMPLING

The studied work proposes a new method that uses score-based generative models to solve linear inverse problems in medical imaging. To do that, it builds on existing methods that use score-based generative models to sample from a prior distribution $p(x)$. When performing unconditional sampling, i.e. sampling from $p(x)$, we start from a sample x_0 from the prior distribution $p(x)$ and iteratively add infinitesimal noise until we get to a sample x_1 , which should be an approximate sample from a pre-defined noise distribution $\pi(x)$. Subsequently, we reverse that process by iteratively reducing the noise in x_1 until we get to a sample x_0 , which should be an approximate sample from the prior distribution $p(x)$. The key point in the score based approach is to approximate the score function with a score model $s_\theta(x_t, t) \approx \nabla_{x_t} \log p_t(x_t)$. An illustration of the process is presented in Fig. 2.

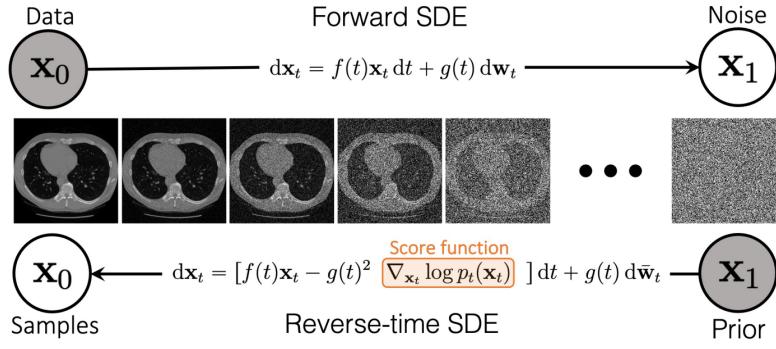


Figure 2: We can gradually add noise to images by following the trajectory of an SDE. By using a score model to estimate the score function, we can approximate the inverse SDE and then solve it to produce new images from noise.

2.2 SCORE-BASED GENERATIVE MODELS: CONDITIONAL SAMPLING

When solving inverse problems, the goal is not to sample from the prior distribution $p(x)$, but rather to sample from the conditional distribution $p(x | y)$. A possible procedure to do that is to first condition the original stochastic process $\{x_t\}_{t \in [0,1]}$ on y , which yields a new conditional stochastic process $\{x_t | y\}_{t \in [0,1]}$. In a subsequent step, we reverse $\{x_t | y\}_{t \in [0,1]}$ by solving the following reverse-time SDE (Song et al., 2020):

$$dx_t = [f(t)x_t - g(t)^2 \nabla_{x_t} \log p_t(x_t | y)] dt + g(t)d\bar{w}_t, \quad t \in [0, 1], \quad (2)$$

where $f : [0, 1] \rightarrow \mathbb{R}$, $g : [0, 1] \rightarrow \mathbb{R}$ and $\{w_t \in \mathbb{R}^n\}_{t \in [0,1]}$ denotes a standard Wiener process. Finally, for $t = 0$, the sample $(x_0 | y)$ that we get should be an approximate sample from $p(x | y)$ and should therefore solve the linear inverse problem. The problem when solving Eq.(2) resides in the computation of the score function $\nabla_{x_t} \log p_t(x_t | y)$. Some solutions exist for that problem. One is to estimate $\nabla_{x_t} \log p_t(x_t | y)$ with a new score model $s_\theta(x_t, y, t)$ that depends explicitly on y . This method requires paired data $\{(x_i, y_i)\}_{i=1}^N$, which makes it a supervised method and is therefore not considered in this work. An unsupervised alternative is to solve Eq.(2) using the unconditional score model $s_\theta(x_t, t)$. This approach usually requires computing an SVD of A (Kadkhodaie & Simoncelli, 2020; Kawar et al., 2021), which is generally problematic for medical imaging problems as it can be very expensive. Alternatively, this work proposes a new solution to solve Eq.(2) using the unconditional score model $s_\theta(x_t, t)$ and an efficient decomposition of A instead of the SVD.

2.3 SOLVE REVERSE CONDITIONAL SDE USING $s_\theta(x_t, t)$ AND DECOMPOSITION OF A

2.3.1 DECOMPOSITION OF A

Without loss of generality, we assume that $\text{rank}(A) = m$. In that case, according to Proposition 1 in Appendix A, the measurement matrix A can be decomposed in the following way $A = \mathcal{P}(\Lambda)T$. We illustrate this decomposition for CT/MRI in Fig. 3. The matrix T is independent of the number of measurements and is the Radon/Fourier Transformation in CT/MRI. $\text{Diag}(\Lambda)$ is a diagonal matrix, which can be interpreted as a subsampling mask on the sinogram/k-space and $\mathcal{P}(\Lambda)$ is an operator that subsamples the sinogram/k-space according to this subsampling mask into an observation y with a smaller size than original image x .

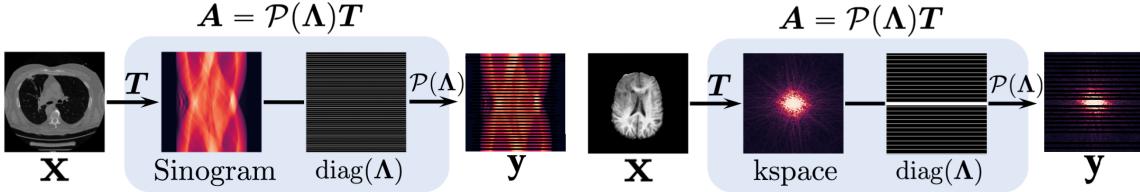


Figure 3: Linear measurement processes for sparse-view CT (left) and undersampled MRI (right).

2.3.2 INCORPORATE OBSERVATION \mathbf{Y} INTO UNCONDITIONAL SAMPLING PROCESS

In order to solve Eq.(2) using the unconditional score model $s_\theta(x_t, t)$, the idea is to make minimal changes to the unconditional sampling process to incorporate the observation y into it. As mentioned earlier, sampling directly from the conditional stochastic process $\{\mathbf{x}_t \mid \mathbf{y}\}_{t \in [0,1]}$ is complicated. So we define a new stochastic process $\{\mathbf{y}_t\}_{t \in [0,1]}$ that relates to $\{\mathbf{x}_t\}_{t \in [0,1]}$ through the following equation $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \alpha(t)\epsilon$, where $\alpha : [0, 1] \rightarrow \mathbb{R}$. The key point for $\{\mathbf{y}_t\}_{t \in [0,1]}$ is that \mathbf{y}_t is a noisy version of the observation \mathbf{y} and that for $t \rightarrow 0$, $\{\mathbf{y}_t\}_{t \in [0,1]} \rightarrow \mathbf{y}$. With some reformulations, we can prove that $\mathbf{y}_t = \alpha(t)\mathbf{y} + \beta(t)\mathbf{A}\mathbf{z}$, where $\beta : [0, 1] \rightarrow \mathbb{R}$ and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. This stochastic process is easy to sample from. So the idea is to make samples of the original stochastic process $\{\mathbf{x}_t\}_{t \in [0,1]}$ consistent with the samples of $\{\mathbf{y}_t \mid \mathbf{y}\}_{t \in [0,1]}$, which means ensuring that the following linear consistency condition $\mathbf{A}\mathbf{x}_t \approx \mathbf{y}_t$ holds at each sampling step. To that end, minimal changes are applied to the iterative unconditional sampling process to transform it into a conditional sampling process. In general, an iterative unconditional sampler iterates according to the following equation

$$\hat{\mathbf{x}}_{t_{i-1}} = \mathbf{h}(\hat{\mathbf{x}}_{t_i}, \mathbf{z}_i, s_\theta(\hat{\mathbf{x}}_{t_i}, t_i)), \quad i = N, N-1, \dots, 1, \quad (3)$$

where $\hat{\mathbf{x}}_{t_N} \sim \pi(\mathbf{x})$ and $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. At each iteration i , a noisy sample $\hat{\mathbf{x}}_{t_i}$ is fed to the function h , which reduces the noise therein using the unconditional score model $s_\theta(x_t, t)$ to produce the next sample at iteration $i - 1$. To enforce the linear consistency condition between the samples of $\{\mathbf{x}_t\}_{t \in [0,1]}$ and samples of $\{\mathbf{y}_t \mid \mathbf{y}\}_{t \in [0,1]}$, we add an additional step to the iteration rule in Eq.(3), which leads to

$$\text{intermediate step } \hat{\mathbf{x}}'_{t_i} = \mathbf{k}(\hat{\mathbf{x}}_{t_i}, \hat{\mathbf{y}}_{t_i}, \lambda) \text{ ensure consistency: } \mathbf{A}\hat{\mathbf{x}}'_{t_i} \approx \hat{\mathbf{y}}_{t_i} \quad (4a)$$

$$\hat{\mathbf{x}}_{t_{i-1}} = \mathbf{h}(\hat{\mathbf{x}}'_{t_i}, \mathbf{z}_i, s_\theta(\hat{\mathbf{x}}_{t_i}, t_i)), \quad i = N, N-1, \dots, 1, \quad (4b)$$

where $\hat{\mathbf{x}}_{t_N} \sim \pi(\mathbf{x})$, $\hat{\mathbf{y}}_{t_i} \sim p_{t_i}(\mathbf{y}_{t_i} \mid \mathbf{y})$, and $0 \leq \lambda \leq 1$ is a hyper-parameter. Conversely to the unconditional sampler, here we do not directly feed the noisy sample $\hat{\mathbf{x}}_{t_i}$ to the function h to produce the next sample, but we first make it consistent with the sample $\hat{\mathbf{y}}_{t_i}$ at the intermediate step and then feed it to h . The iteration function $\mathbf{k}(\cdot, \hat{\mathbf{y}}_{t_i}, \lambda) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ promotes data consistency by solving a proximal optimization step (Nesterov, 2003; Hammerl et al., 2021; Boyd et al., 2004). The hyperparameter λ in $\mathbf{k}(\hat{\mathbf{x}}_{t_i}, \hat{\mathbf{y}}_{t_i}, \lambda)$ balances between $\hat{\mathbf{x}}'_{t_i} \approx \hat{\mathbf{x}}_{t_i}$ and $\mathbf{A}\hat{\mathbf{x}}'_{t_i} \approx \hat{\mathbf{y}}_{t_i}$. For $\lambda = 0$, $\hat{\mathbf{x}}'_{t_i} = \mathbf{k}(\hat{\mathbf{x}}_{t_i}, \hat{\mathbf{y}}_{t_i}, 0) = \hat{\mathbf{x}}_{t_i}$, which means the sampler in Eq.(4b) performs unconditional generation as the constraint $\mathbf{A}\hat{\mathbf{x}}'_{t_i} = \hat{\mathbf{y}}_{t_i}$ is completely ignored. For $\lambda = 1$, $\hat{\mathbf{x}}'_{t_i} = \mathbf{k}(\hat{\mathbf{x}}_{t_i}, \hat{\mathbf{y}}_{t_i}, 1)$ satisfies $\mathbf{A}\hat{\mathbf{x}}'_{t_i} = \hat{\mathbf{y}}_{t_i}$ exactly. Therefore, $\lambda = 1$ should only be used for the case where the measurement process is not noisy. When the measurement is noisy, we choose $0 < \lambda < 1$ to allow slackness in the constraint $\mathbf{A}\hat{\mathbf{x}}'_{t_i} = \hat{\mathbf{y}}_{t_i}$. In practice, λ is tuned automatically on a validation dataset using Bayesian optimization. We note that $\mathbf{k}(\hat{\mathbf{x}}_{t_i}, \hat{\mathbf{y}}_{t_i}, \lambda)$ is computed using the following expression

$$\mathbf{k}(\hat{\mathbf{x}}_{t_i}, \hat{\mathbf{y}}_{t_i}, \lambda) = \mathbf{T}^{-1} [\lambda \mathbf{\Lambda} \mathcal{P}^{-1}(\mathbf{\Lambda}) \hat{\mathbf{y}}_{t_i} + (1 - \lambda) \mathbf{\Lambda} \mathbf{T} \hat{\mathbf{x}}_{t_i} + (\mathbf{I} - \mathbf{\Lambda}) \mathbf{T} \hat{\mathbf{x}}_{t_i}], \quad (5)$$

where $\mathcal{P}^{-1}(\mathbf{\Lambda}) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ denotes any right inverse of $\mathcal{P}(\mathbf{\Lambda})$. We note that this expression makes use of the decomposition of A mentioned earlier. In Fig. 4, we provide an example of an unconditional sampler (Euler-Maruyama sampler) in form of Algorithm 1, which is transformed into a conditional sampler using the proposed approach in form of Algorithm 2. We can observe how this is achieved by adding/modifying just three lines of pseudo-code.

3 EXPERIMENTS AND RESULTS

In this section, we benchmark the proposed method against existing supervised, unsupervised and learning-free techniques to perform sparse-view CT reconstruction, undersampled MRI reconstruction and MAR making use of three different datasets.

Datasets For sparse-view CT reconstruction and MAR, we use the LIDC (Armato III et al., 2011) and LDCT datasets. The LIDC dataset is comprised of images of human lungs, with a (130K/ 1K images) training/ testing split. The LDCT dataset is comprised of images of heads, chests and abdomens of humans, with a (47K/ 1K images) training/ testing split. For undersampled MRI reconstruction, the BraTS (Menze et al., 2014) dataset is used, which comprises images of human brains, with a (297K/ 1K images) training/ testing split.

Evaluation For the evaluation of the considered methods, this work makes use of the PSNR and SSIM similarity metrics. These metrics compare the similarity between the original ground truth image and the corresponding reconstructed image.

Algorithm 1 Unconditional sampling	Algorithm 2 Inverse problem solving
Require: N	Require: N, \mathbf{y}, λ
1: $\hat{\mathbf{x}}_1 \sim \pi(\mathbf{x}), \Delta t \leftarrow \frac{1}{N}$	1: $\hat{\mathbf{x}}_1 \sim \pi(\mathbf{x}), \Delta t \leftarrow \frac{1}{N}$
2: for $i = N - 1$ to 0 do	2: for $i = N - 1$ to 0 do
3: $t \leftarrow \frac{i+1}{N}$	3: $t \leftarrow \frac{i+1}{N}$
4: $\hat{\mathbf{x}}_{t-\Delta t} \leftarrow \hat{\mathbf{x}}_t - f(t)\hat{\mathbf{x}}_t\Delta t$	4: $\hat{\mathbf{y}}_t \sim p_{0t}(\mathbf{y}_t \mathbf{y})$
5: $\hat{\mathbf{x}}_{t-\Delta t} \leftarrow \hat{\mathbf{x}}_{t-\Delta t} + g(t)^2 s_{\theta*}(\hat{\mathbf{x}}_t, t)\Delta t$	5: $\hat{\mathbf{x}}_t \leftarrow \mathbf{T}^{-1}[\lambda \Lambda \mathcal{P}^{-1}(\Lambda) \hat{\mathbf{y}}_t + (1 - \lambda) \Lambda \mathbf{T} \hat{\mathbf{x}}_t + (\mathbf{I} - \Lambda) \mathbf{T} \hat{\mathbf{x}}_t]$
6: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$	6: $\hat{\mathbf{x}}_{t-\Delta t} \leftarrow \hat{\mathbf{x}}_t - f(t)\hat{\mathbf{x}}_t\Delta t$
7: $\hat{\mathbf{x}}_{t-\Delta t} \leftarrow \hat{\mathbf{x}}_{t-\Delta t} + g(t)\sqrt{\Delta t} \mathbf{z}$	7: $\hat{\mathbf{x}}_{t-\Delta t} \leftarrow \hat{\mathbf{x}}_{t-\Delta t} + g(t)^2 s_{\theta*}(\hat{\mathbf{x}}_t, t)\Delta t$
8: return $\hat{\mathbf{x}}_0$	8: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
	9: $\hat{\mathbf{x}}_{t-\Delta t} \leftarrow \hat{\mathbf{x}}_{t-\Delta t} + g(t)\sqrt{\Delta t} \mathbf{z}$
	10: return $\hat{\mathbf{x}}_0$

Figure 4: Converting an unconditional sampler to an inverse problem solver using the proposed approach.

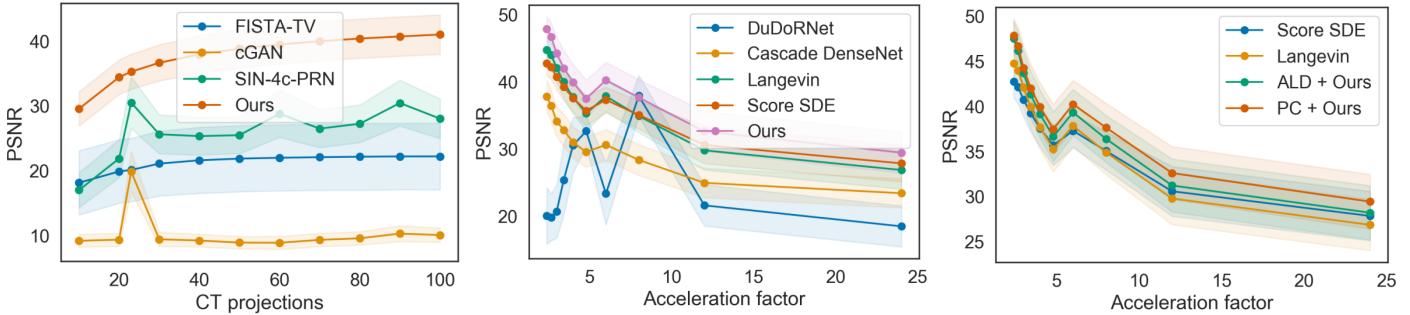


Figure 5: Performance vs. numbers of measurements. Shaded areas represent standard deviation. (Left) CT on LIDC. (Center) MRI on BraTS (Right) Comparing score-based generative models for undersampled MRI reconstruction on BraTS.

3.1 CT RECONSTRUCTION

Method	Projections <small>(test time)</small>	LIDC 320 × 320		LDCT 512 × 512	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑
FBP	23	10.18±1.38	0.230±0.072	10.11±1.19	0.302±0.078
FISTA-TV	23	20.08±4.89	0.799±0.061	21.88±4.42	0.850±0.067
cGAN	23	19.83±3.07	0.479±0.103	19.90±2.52	0.545±0.065
Neumann	23	17.18±3.79	0.454±0.128	18.83±3.29	0.525±0.073
SIN-4c-PRN	23	30.48±3.99	0.895±0.047	34.82±3.55	0.877±0.116
Ours	10	29.52±2.63	0.823±0.061	28.96±4.41	0.849±0.086
	20	34.40±2.66	0.895±0.048	36.80±4.50	0.936±0.058
	23	35.24±2.71	0.905±0.046	37.41±4.62	0.941±0.057

Figure 6: Results for sparse-view CT reconstruction on LIDC and LDCT. FISTA-TV is a standard iterative reconstruction method that does not need training. cGAN, Neumann, and SIN-4c-PRN are supervised learning techniques trained with 23 projection angles.

Fig. 6 shows the results for sparse-view CT reconstruction on LIDC and LDCT, where the supervised learning methods were trained with 23 projection angles. We observe that the proposed method outperforms the other methods. It still also outperforms the other methods even when given less measurement at test time (20) than the other methods (23). Fig. 5 (Left) investigates the effects of using different number of measurements at test time on the performance of the considered methods for CT reconstruction, where it shows the PSNR on the test set vs. the number of CT projections at test time for LIDC. As you can see, for the two supervised learning methods cGAN and SIN-4c-PRN, if the number of projections angles at test time is not equal to the number of projections at training time (23), their performance drops significantly. On the other side, the performance of the proposed method increases as the number of CT projections increases. The proposed method generalizes therefore better than the supervised learning methods to different number of measurements at test time. Fig. 7 shows a visual comparison of the sparse-view CT reconstruction quality for the considered methods. It is clear to see that the proposed method succeeds in reconstructing the ground truth image very well and that it has a better reconstruction quality than other methods, e.g., Neumann or FISTA-TV.

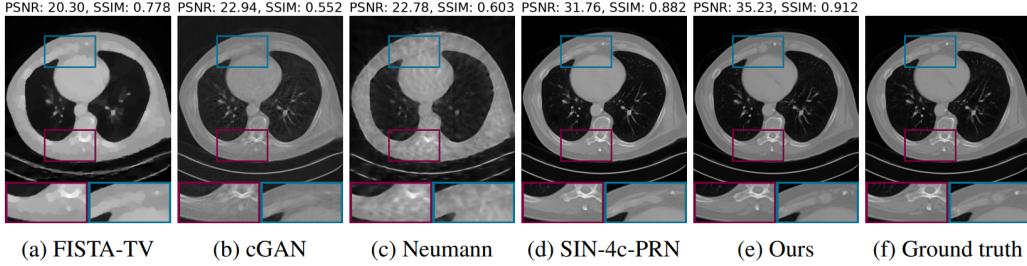


Figure 7: Examples of sparse-view CT reconstruction results on LDCT, all with 23 projections.

3.2 METALLIC ARTIFACT REMOVAL

Fig. 8 shows the results for the MAR task on the LIDC dataset. We observe that the proposed method outperforms the learning-free and supervised methods, where the supervised methods were designed and trained specifically for MAR. On the other side, the proposed method was not trained specifically for MAR, but uses the same score model that solves the CT reconstruction task to solve the MAR task. This means the proposed method has the ability to generalize to different measurement processes as it can solve both CT reconstruction and MAR tasks with the same score model.

3.3 MRI RECONSTRUCTION

Method	PSNR↑	SSIM↑
LI	26.30 ± 2.62	0.910 ± 0.028
cGANMAR	27.27 ± 1.96	0.927 ± 0.060
SNMAR	27.28 ± 1.43	0.937 ± 0.048
Ours	32.16 ± 2.32	0.939 ± 0.022

Figure 8: MAR results on LIDC. First method is a learning-free technique. The others are supervised techniques.

Method	24× Acceleration		8× Acceleration		4× Acceleration	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Cascade DenseNet	23.39 ± 2.17	0.765 ± 0.042	28.35 ± 2.30	0.845 ± 0.038	30.97 ± 2.33	0.902 ± 0.028
DuDoRNet	18.46 ± 3.05	0.662 ± 0.093	37.88 ± 3.03	0.985 ± 0.007	30.53 ± 4.13	0.891 ± 0.071
Score SDE	27.83 ± 2.73	0.849 ± 0.038	35.04 ± 2.11	0.943 ± 0.016	37.55 ± 2.08	0.960 ± 0.013
Langevin	28.80 ± 3.21	0.873 ± 0.039	36.44 ± 2.28	0.952 ± 0.016	38.76 ± 2.32	0.966 ± 0.012
Ours	29.42 ± 3.03	0.880 ± 0.035	37.63 ± 2.70	0.958 ± 0.015	39.91 ± 2.67	0.965 ± 0.013

Figure 9: Results for undersampled MRI reconstruction on BraTS. First two methods are supervised learning techniques trained with 8x acceleration. The others are unsupervised techniques.

Fig. 9 shows the results of undersampled MRI reconstruction on the BraTS dataset. We observe that the proposed method outperforms the supervised and unsupervised methods on the 24x and 4x acceleration tasks, where the supervised methods were trained with 8x acceleration. For the 8x acceleration task, the proposed method ranks second, with comparable performance to the top supervised method DuDoRNet, which ranks first. Fig. 5 (Center) shows the PSNR vs. the acceleration factor at test time for MRI reconstruction on BraTS. We observe that for the supervised method DuDoRNet, if the acceleration factor at test time is not equal to the acceleration factor at training time (8), its performance drops significantly. On the other side, the unsupervised learning methods Langevin and Score SDE can generalize better than DuDoRNet to different acceleration factors at test time. The proposed method can generalize even better as its PSNR curve is almost always above all other curves. From Fig. 5 (Right), We observe that whether the proposed method uses annealed Langevin dynamics (ALD) or Predictor-Corrector (PC) as a sampler, it outperforms existing score-based methods. Additionally, we observe that the proposed method performs better when using PC than when using the ALD sampler, showing the effectiveness of using more sophisticated sampling techniques in the proposed method.

4 CONCLUSION

To conclude, the studied paper proposes a new method that uses score-based generative models to solve linear inverse problems in medical imaging. The proposed method makes minimal changes to any iterative score-based unconditional sampler to incorporate the observation y into its sampling process and convert it to an iterative conditional sampling process, which can solve linear inverse problems. Throughout the paper, the proposed method was benchmarked against other existing supervised, unsupervised and learning-free techniques for various linear inverse problems in medical imaging using different datasets. In the conducted experiments, the proposed method performed better than existing methods in CT reconstruction and MAR, showed better generalization capabilities to different number of measurements at test time in CT and MRI reconstruction compared to supervised learning methods and showed that it can generalize to different measurement processes by solving both CT reconstruction and MAR tasks using the same score model. Additionally, the proposed method did not always rank first regarding MRI reconstruction.

REFERENCES

- Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931, 2011.
- Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- Kerstin Hammernik, Jo Schlemper, Chen Qin, Jinming Duan, Ronald M Summers, and Daniel Rueckert. Systematic evaluation of iterative deep neural networks for fast parallel mri reconstruction with sensitivity-weighted coil combination. *Magnetic Resonance in Medicine*, 86(4):1859–1872, 2021.
- Zahra Kadkhodaie and Eero P Simoncelli. Solving linear inverse problems using the prior implicit in a denoiser. *arXiv preprint arXiv:2007.13640*, 2020.
- Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.
- Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.
- Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.
- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

A APPENDIX

Proposition 1. If $\text{rank}(\mathbf{A}) = m$, then there exist an invertible matrix $\mathbf{T} \in \mathbb{R}^{n \times n}$, and a diagonal matrix $\Lambda \in \{0, 1\}^{n \times n}$ with $\text{tr}(\Lambda) = m$, such that $\mathbf{A} = \mathcal{P}(\Lambda)\mathbf{T}$. Here $\mathcal{P}(\Lambda) \in \{0, 1\}^{m \times n}$ is an operator that, when multiplied with any vector $\mathbf{a} \in \mathbb{R}^n$, reduces its dimensionality to m by removing each i -th element of \mathbf{a} for $i = 1, 2, \dots, n$ if $\Lambda_{ii} = 0$.