A Regularized Conditional GAN for Posterior Sampling in Image Recovery Problems

Matthew Bendel, Rizwan Ahmad, and Philip Schniter

THE OHIO STATE UNIVERSITY

Supported by NIH grant R01EB029957

37th Conference on Neural Information Processing Systems

December 2023

Image Inverse Problems

Goal: Recover image $\boldsymbol{x}$ from measurements $\boldsymbol{y} = \mathcal{M}(\boldsymbol{x})$:

- $\mathcal{M}(\cdot)$ masks, distorts, and/or corrupts $\boldsymbol{x}$ with noise.
- Solution typically posed as finding single best recovery $\widehat{\boldsymbol{x}}$, known as "point-estimation" [1]

Challenges with point-estimation:

- Inability to navigate the perception-distortion tradeoff [2]
- Inability to quantify reconstruction uncertainty
- Problems with fairness in $\widehat{\boldsymbol{x}}$

Solution: Sample from posterior distribution $p_{\mathsf{x}|\mathsf{y}}(\boldsymbol{x}|\boldsymbol{y}) = \dfrac{p_{\mathsf{y}|\mathsf{x}}(\boldsymbol{y}|\boldsymbol{x})p_{\mathsf{x}}(\boldsymbol{x})}{\int p_{\mathsf{y}|\mathsf{x}}(\boldsymbol{y}|\boldsymbol{x})p_{\mathsf{x}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}$

Existing approaches:

- Conditional VAEs [3, 4], conditional NFs [5, 6], conditional GANs [7, 8]
- Langevin [9] / Diffusion [10, 11] methods

Our contribution

Our approach: A novel regularized conditional Wasserstein GAN

- Generator outputs $\widehat{\boldsymbol{x}}_i = G_{\boldsymbol{\theta}}(\boldsymbol{z}_i, \boldsymbol{y})$ for code realization $\boldsymbol{z}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
- Discriminator $D_{\boldsymbol{\phi}}$ aims to distinguish true pair $(\boldsymbol{x}, \boldsymbol{y})$ from fake pair $(\widehat{\boldsymbol{x}}_i, \boldsymbol{y})$
- We jointly train networks $G_{\boldsymbol{\theta}}$ and $D_{\boldsymbol{\phi}}$ via

$$\min_{\boldsymbol{\theta}} \max_{\boldsymbol{\phi}} \left\{ \mathbb{E}_{\mathsf{x},\mathsf{z},\mathsf{y}}\{D_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) - D_{\boldsymbol{\phi}}(G_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y})\} + \mathcal{R}(\boldsymbol{\theta}) - \mathcal{L}_{\mathsf{gp}}(\boldsymbol{\phi}) \right\}$$

Proposed regularization:

- Regularization based on a supervised-L1 penalty & standard-deviation reward:

$$\mathcal{R}(\boldsymbol{\theta}) \triangleq \underbrace{\mathbb{E}_{\mathsf{x},\mathsf{z}_1,\dots,\mathsf{z}_P,\mathsf{y}}\left\{\|\boldsymbol{x} - \widehat{\boldsymbol{x}}_{(P)}\|_1\right\}}_{\triangleq\, \mathcal{L}_{1,P}(\boldsymbol{\theta})} - \beta_{\mathsf{std}} \underbrace{\textstyle\sum_{i=1}^{P} \mathbb{E}_{\mathsf{z}_1,\dots,\mathsf{z}_P,\mathsf{y}}\left\{\|\widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_{(P)}\|_1\right\}}_{\triangleq\, \mathcal{L}_{\mathsf{std},P}(\boldsymbol{\theta})}$$

  where $\widehat{\boldsymbol{x}}_{(P)} = \frac{1}{P}\sum_{i=1}^{P} \widehat{\boldsymbol{x}}_i$ is the average of P posterior samples
- In the simple Gaussian case where $\widehat{x}_i = \mu + \sigma z_i$ with $z_i \sim \mathcal{N}(0,1)$, $\boldsymbol{\theta} = [\mu, \sigma]$, and true $p_{\mathsf{x}|\mathsf{y}}(\cdot|\boldsymbol{y}) = \mathcal{N}(\mu_0, \sigma_0)$, we prove recovery of the correct posterior, e.g.,

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{\mathcal{L}_{1,P}(\boldsymbol{\theta}) - \beta_{\mathsf{std}}\mathcal{L}_{\mathsf{std},P}(\boldsymbol{\theta})\right\}$$

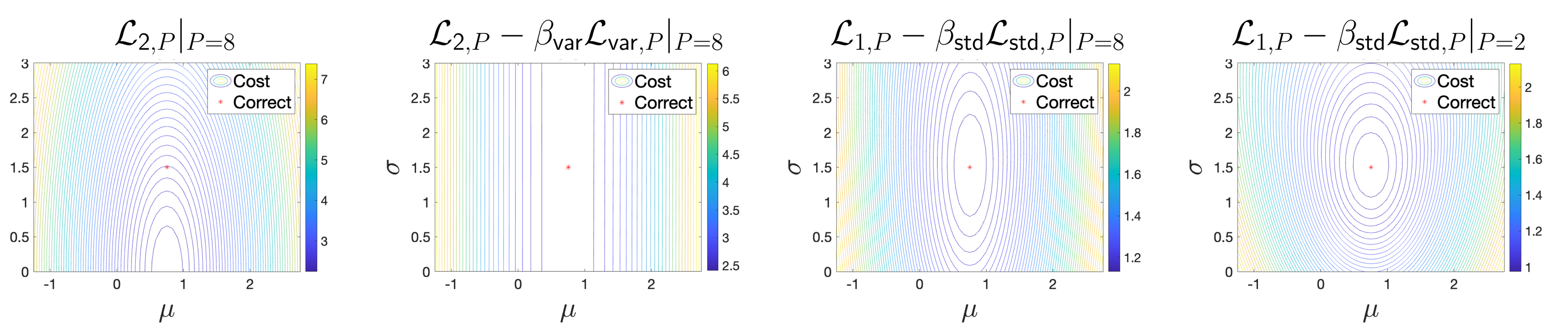
  yields

$$\widehat{\boldsymbol{\theta}} = [\mu_0, \sigma_0]$$

  for any $P \geq 2$ when $\beta_{\mathsf{std}} = \frac{1}{P}\sqrt{\frac{1}{(P-1)(P+1)}} \triangleq \beta_{\mathsf{std}}^{\mathcal{N}}$
- We also prove the failure of L2/variance-based regularizations, e.g.,

$$\widetilde{\mathcal{R}}(\boldsymbol{\theta}) \triangleq \underbrace{\mathbb{E}_{\mathsf{x},\mathsf{z}_1,\dots,\mathsf{z}_P,\mathsf{y}}\left\{\|\boldsymbol{x} - \widehat{\boldsymbol{x}}_{(P)}\|_2^2\right\}}_{\triangleq\, \mathcal{L}_{2,P}(\boldsymbol{\theta})} - \beta_{\mathsf{var}} \underbrace{\textstyle\sum_{i=1}^{P} \mathbb{E}_{\mathsf{z}_1,\dots,\mathsf{z}_P,\mathsf{y}}\left\{\|\widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_{(P)}\|_2^2\right\}}_{\triangleq\, \mathcal{L}_{\mathsf{var},P}(\boldsymbol{\theta})}$$

  L2 alone induces mode collapse & L2/variance can't recover posterior variance:

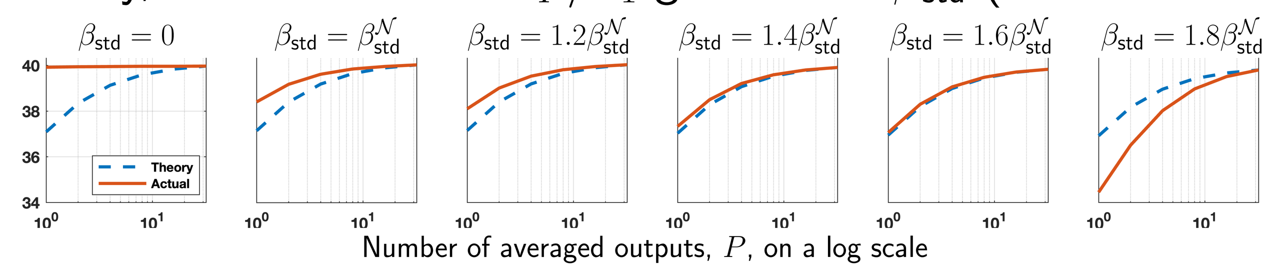


Auto-tuning β_{std}

- To tune β_{std}, we first establish that, for *true* posterior samples $\{\widehat{\boldsymbol{x}}_i\}$, we have

$$\mathcal{E}_P \triangleq \mathbb{E}\{\|\widehat{\boldsymbol{x}}_{(P)} - \boldsymbol{x}\|_2^2|\boldsymbol{y}\} = \frac{P+1}{P}\mathcal{E}_{\mathsf{mmse}} \quad \text{and so} \quad \frac{\mathcal{E}_P}{\mathcal{E}_1} = \frac{P+1}{2P}$$

- Experimentally, we observe that $\mathcal{E}_P/\mathcal{E}_1$ grows with β_{std} (see red curves below):



- We adapt β_{std} so that validation $\widehat{\mathcal{E}}_P/\widehat{\mathcal{E}}_1$ matches theoretical $\mathcal{E}_P/\mathcal{E}_1$ at $P = 8$:

$$\beta_{\mathsf{std}} \leftarrow \beta_{\mathsf{std}} + \mu_{\mathsf{std}}\left(\left[\frac{P+1}{2P}\right]_{\mathsf{dB}} - \left[\frac{\widehat{\mathcal{E}}_P}{\widehat{\mathcal{E}}_1}\right]_{\mathsf{dB}}\right) \quad \text{for some } \mu_{\mathsf{std}} > 0$$

Quantifying performance using CFID

We quantify posterior-approximation accuracy using Conditional FID [12]:

1. Here the goal is to compute the conditional Wasserstein-2 distance

$$\mathrm{CWD} \triangleq \mathbb{E}_{\mathsf{y}}\left\{W_2(p_{\mathsf{x}|\mathsf{y}}(\cdot, \boldsymbol{y}), p_{\widehat{\mathsf{x}}|\mathsf{y}}(\cdot, \boldsymbol{y}))\right\}$$

2. $(\boldsymbol{x}, \widehat{\boldsymbol{x}})$ are replaced by Inception-v3 or VGG-16 embeddings $(\underline{\boldsymbol{x}}, \underline{\widehat{\boldsymbol{x}}})$, like in FID
3. $p_{\underline{\mathsf{x}}|\mathsf{y}}$ & $p_{\underline{\widehat{\mathsf{x}}}|\mathsf{y}}$ approximated by $\mathcal{N}(\boldsymbol{\mu}_{\underline{\mathsf{x}}|\mathsf{y}}, \boldsymbol{\Sigma}_{\underline{\mathsf{xx}}|\mathsf{y}})$ & $\mathcal{N}(\boldsymbol{\mu}_{\underline{\widehat{\mathsf{x}}}|\mathsf{y}}, \boldsymbol{\Sigma}_{\underline{\widehat{\mathsf{x}}\widehat{\mathsf{x}}}|\mathsf{y}})$, like in FID, giving

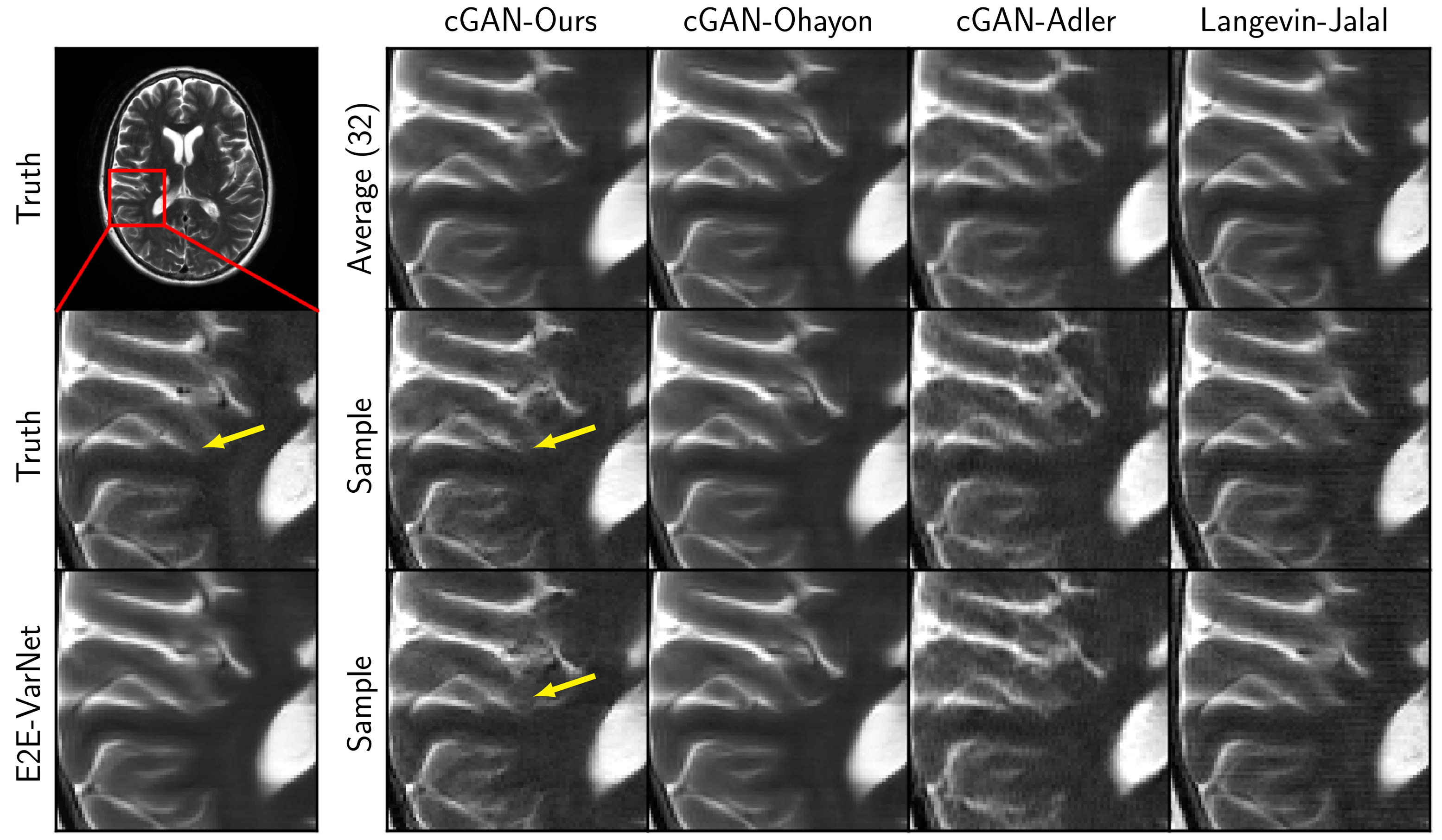
$$\mathrm{CFID} \triangleq \mathbb{E}_{\mathsf{y}}\left\{\|\boldsymbol{\mu}_{\underline{\mathsf{x}}|\mathsf{y}} - \boldsymbol{\mu}_{\underline{\widehat{\mathsf{x}}}|\mathsf{y}}\|_2^2 + \mathrm{tr}\left[\boldsymbol{\Sigma}_{\underline{\mathsf{xx}}|\mathsf{y}} + \boldsymbol{\Sigma}_{\underline{\widehat{\mathsf{x}}\widehat{\mathsf{x}}}|\mathsf{y}} - 2\left(\boldsymbol{\Sigma}_{\underline{\mathsf{xx}}|\mathsf{y}}^{1/2}\boldsymbol{\Sigma}_{\underline{\widehat{\mathsf{x}}\widehat{\mathsf{x}}}|\mathsf{y}}\boldsymbol{\Sigma}_{\underline{\mathsf{xx}}|\mathsf{y}}^{1/2}\right)^{1/2}\right]\right\}$$

Accelerated MRI reconstruction results

We reconstructed fastMRI [13] multicoil T2 brain data at acceleration $R = 8$

Model	CFID↓	FID↓	PSNR*↑	SSIM*↑	LPIPS*↓	DISTS*↓	Time (4)↓
E2E-VarNet [14]	7.82	8.40	**36.49**	0.9220	0.0575	0.1253	316ms
Langevin-Jalal [15]	7.34	14.32	33.90	0.9137	0.0579	0.1086	14 min
cGAN-Adler [8]	10.10	10.77	33.51	0.9111	0.0614	0.1252	**217 ms**
cGAN-Ohayon [16]	6.04	11.05	34.92	0.9222	0.0532	0.1128	**217 ms**
cGAN-Ours	**4.87**	**7.72**	35.42	**0.9257**	**0.0379**	**0.0877**	**217 ms**

- Metrics with * are reported for the optimal averaging constant P



- Posterior samples from our cGAN show meaningful variations (see arrows)
- Posterior samples from cGAN-Ohayon (i.e., L2 alone) show no variation
- Posterior samples from cGAN-Adler and Langevin-Jalal show unwanted visual artifacts

Navigating the perception-distortion tradeoff

There's a fundamental tradeoff [2] between the *perceptual quality* (PQ) and *distortion* on $\widehat{\boldsymbol{x}}$

- PQ: Any distance between $\widehat{\boldsymbol{x}}$ and manifold of "clean" images
- Distortion: Any distance between $\widehat{\boldsymbol{x}}$ and the true $\boldsymbol{x}$

Posterior samplers can navigate this tradeoff by averaging P samples!

Truth	$P = 1$	$P = 2$	$P = 4$	$P = 32$

PSNR:	32.32	33.67	34.53	**35.42**
LPIPS:	0.0418	**0.0379**	0.0421	0.0539

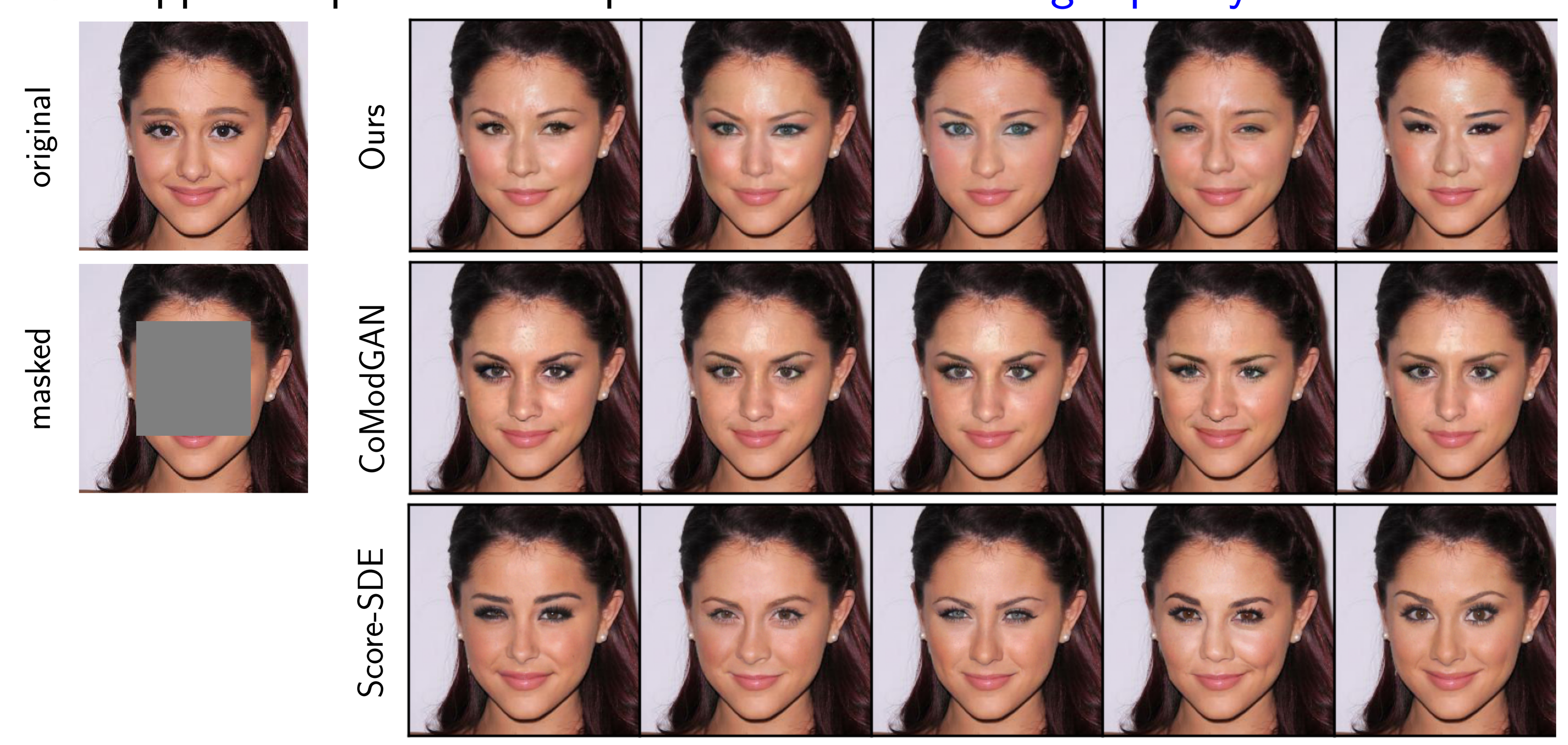
better PQ ⟵ ⟶ better dist

Large-scale image completion results

We inpainted a centered 128x128 square on 256x256 CelebA-HQ faces

Model	CFID↓	FID↓	Time (128)↓
Score-SDE [10]	5.11	7.92	48 min
CoModGAN [17]	5.29	8.50	**217 ms**
cGAN (ours)	**4.69**	**7.45**	**217 ms**

- The cGANs generated samples 13 000x faster than Score-SDE!
- Our approach produces samples which are both high quality and diverse



References

[1] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. Springer Science & Business Media, 2006.
[2] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," CVPR, 2018.
[3] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," NeurIPS 2015.
[4] F. Tonolini, J. Radford, A. Turpin, D. Faccio, and R. Murray-Smith, "Variational inference for computational imaging inverse problems," JMLR 2020.
[5] C. Winkler, D. Worrall, E. Hoogeboom, and M. Welling, "Learning likelihoods with conditional normalizing flows," *arXiv preprint arXiv:1912.00042*, 2019.
[6] L. Ardizzone, C. Lüth, J. Kruse, C. Rother, and U. Köthe, "Guided image generation with conditional invertible neural networks," *arXiv:1907.02392*, 2019.
[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros "Image-to-image translation with conditional adversarial networks," CVPR 2017.
[8] J. Adler and O. Öktem, "Deep Bayesian inversion," *arXiv:1811.05910*, 2018.
[9] Y. Song and S. Ermon, "Improved techniques for training score-based generative models," NeurIPS 2020.
[10] Y.Song, J.Sohl-Dickstein, D.P.Kingma, A.Kumar, S.Ermon, B.Poole, "Score-based generative modeling through stochastic differential equations," ICLR 2021.
[11] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J.-C. Ye, "Diffusion posterior sampling for general noisy inverse problems," ICLR 2023.
[12] M. Soloveitchik, T. Diskin, E. Morin, and A. Wiesel, "Conditional Frechet inception distance," *arXiv:2103.11521*, 2021.
[13] J. Zbontar *et al.*, "fastMRI: An open dataset and benchmarks for accelerated MRI," *arXiv:1811.08839*, 2018.
[14] A. Sriram *et al.*, "End-to-end variational networks for accelerated MRI reconstruction," MICCAI 2020.
[15] A. Jalal, M. Arvinte, G. Daras, E. Price, A. Dimakis, and J. Tamir, "Robust compressed sensing MRI with deep generative priors," NeurIPS 2021.
[16] G. Ohayon, T. Adrai, G. Vaksman, M. Elad, and P. Milanfar, "High perceptual quality image denoising with a posterior sampling CGAN," ICCVW 2021.
[17] S.Zhao, J.Cui, Y.Sheng, Y.Dong, X.Liang, E.I.Chang, Y.Xu, "Large scale image completion via co-modulated generative adversarial networks," ICLR 2021.