

Measurement-Guided Consistency Model Sampling for Inverse Problems

Amirreza Tanevardi, Pooria Abbas Rad Moghadam, and Sajjad Amini

Abstract—Diffusion models have become powerful generative priors for solving inverse imaging problems, but their reliance on slow multi-step sampling limits practical deployment. Consistency models address this bottleneck by enabling high-quality generation in a single or only a few steps, yet their direct adaptation to inverse problems is underexplored. In this paper, we present a modified consistency sampling approach tailored for inverse problem reconstruction: the sampler’s stochasticity is guided by a measurement-consistency mechanism tied to the measurement operator, which enforces fidelity to the acquired measurements while retaining the efficiency of consistency-based generation. Experiments on Fashion-MNIST and LSUN Bedroom datasets demonstrate consistent improvements in perceptual and pixel-level metrics, including Fréchet Inception Distance, Kernel Inception Distance, peak signal-to-noise ratio, and structural similarity index measure, compared to baseline consistency sampling, yielding competitive or superior reconstructions with only a handful of steps.

Index Terms—consistency models, diffusion models, inverse problems, measurement-consistency sampling

I. INTRODUCTION

INVERSE problems play a central role in imaging science and engineering, with applications ranging from medical imaging [3] and audio signal processing [2] to remote sensing [8] and seismic imaging [1]. The task is to recover an unknown signal $\mathbf{x} \in \mathbb{R}^n$ from degraded measurements $\mathbf{y} \in \mathbb{R}^m$ obtained via a degradation operator A :

$$\mathbf{y} = A(\mathbf{x}) + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0, \sigma_y^2 \mathbf{I}). \quad (1)$$

Because A is often ill-conditioned, undersampled, or noisy, reconstruction is ill-posed [3] and hinges on strong priors. In a Bayesian view, the posterior can be calculated as:

$$p(\mathbf{x} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}) p(\mathbf{x}) \quad (2)$$

Thus, the maximum a posteriori (MAP) solution can be obtained by maximizing the log-posterior probability as:

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x}}{\operatorname{argmax}} \{ \log p(\mathbf{y} | \mathbf{x}) + \log p(\mathbf{x}) \} \quad (3)$$

where the log-likelihood term $\log p(\mathbf{y} | \mathbf{x})$ captures the *measurement fidelity*, ensuring that the estimated solution \mathbf{x} is consistent with the observed measurements \mathbf{y} . The log-prior term $\log p(\mathbf{x})$ incorporates prior knowledge about the random vector \mathbf{x} , such as smoothness, sparsity, or structural constraints

through the probability distribution function, and plays the role of *regularization* that stabilizes the solution when the problem is ill-posed or the measurements are noisy [8].

Diffusion models [12] are a class of generative models that learn to approximate the data distribution by reversing a gradual noising process. In the forward process, Gaussian noise is incrementally added to a clean data sample \mathbf{x}_0 through T steps, producing latent variables \mathbf{x}_t according to:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}), \quad (4)$$

where $\{\beta_t\}_{t=1}^T$ is a variance schedule. The generative process reverse this diffusion by learning a parameterized distribution as:

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mu_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}), \quad (5)$$

where the covariance $\sigma_t^2 \mathbf{I}$ is fixed according to the noise schedule, and the model only learns the mean $\mu_\theta(\mathbf{x}_t, t)$. Training is typically carried out by minimizing a simplified noise-prediction objective, where the model learns to estimate the noise added at each step.

Denoising Diffusion Restoration Models (DDRM) have recently emerged as powerful generative methods for posterior sampling in image inverse problems [4]. Building on the diffusion process, DDRM defines an iterative procedure to sample from the posterior of an inverse problem, using the diffusion model to capture the prior distribution. By modeling a forward noising process together with its reverse denoising dynamics, these priors enable posterior sampling strategies that produce high-quality reconstructions in a variety of applications, including super-resolution [7], inpainting [6], and deblurring [5]. More precisely, the posterior sampling in DDRM is done using the following iterations:

$$p_\theta^{(t)}(\bar{\mathbf{x}}_t^{(i)} | \mathbf{x}_{t+1}, \mathbf{y}) = \begin{cases} \mathcal{N}\left(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_{t+1}}, \eta^2 \sigma_t^2\right), & \text{if } s_i = 0, \\ \mathcal{N}\left(\bar{\mathbf{x}}_{\theta,t}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{\theta,t}^{(i)}}{\sigma_y / s_i}, \eta^2 \sigma_t^2\right), & \text{if } \sigma_t < \frac{\sigma_y}{s_i}, \\ \mathcal{N}\left((1 - \eta_b) \bar{\mathbf{x}}_{\theta,t}^{(i)} + \eta_b \bar{\mathbf{x}}^{(i)}, \sigma_t^2 - \frac{\sigma_y^2}{s_i^2} \eta_b^2\right), & \text{if } \sigma_t \geq \frac{\sigma_y}{s_i}, \end{cases} \quad (6)$$

where s_i are the singular values of the measurement operator A , $\bar{\mathbf{y}}^{(i)}$ denotes the i -th index of the measurements in the spectral domain. We also define $\bar{\mathbf{x}}_{\theta,t}^{(i)}$ as the i -th index of $\bar{\mathbf{x}}_{\theta,t} = \mathbf{V}^\top \mathbf{x}_{\theta,t}$ where $\bar{\mathbf{x}}_{\theta,t}$ is the model’s prediction of the clean component at step t . The parameters σ_t and σ_y control

the diffusion noise and measurement noise, respectively, while η and η_b adjust the stochasticity of sampling.

A major limitation, however, is that sampling from diffusion models typically requires hundreds of iterative steps, which restricts their applicability in time-sensitive or large-scale settings [21]

Consistency models (CMs) were proposed to address this computational bottleneck by learning a direct mapping $f_\theta(\mathbf{x}_t, t) \approx \mathbf{x}_0$ that projects a noisy latent \mathbf{x}_t to a clean sample in one or a few evaluations [10]. Training enforces self-consistency across noise levels and a boundary constraint at small noise, i.e.:

$$f_\theta(\mathbf{x}_t, t) = f_\theta(\mathbf{x}_{t'}, t') \quad \forall t, t' \in [\epsilon, T], \quad f_\theta(\mathbf{x}_\epsilon, \epsilon) = \mathbf{x}_\epsilon \quad (7)$$

which together ensure consistent predictions across noise scales and exact reconstruction at the lowest noise level. These constraints enable single-step or multi-step generation with dramatically reduced neural function evaluations (NFEs) [22]. These attractive speed–quality tradeoffs make CMs promising candidates as fast priors for inverse problems.

Directly applying CMs to inverse problems is nontrivial. Recent works (e.g., CoSIGN) mitigate this by introducing measurement conditioning via auxiliary encoders or by enforcing hard projections to satisfy $\|A(\mathbf{x}) - \mathbf{y}\|_2$ constraints, yielding few-step reconstructions competitive with slow diffusion solvers [11]. Nevertheless, designing simple, principled samplers that couple CM sampling dynamics with measurement fidelity while preserving single/few-step efficiency remains an open problem.

In this paper, we propose a novel deterministic sampling strategy that bridges this gap by coupling CM updates with an explicit measurement-driven guidance mechanism. Intuitively, the sampler adapts its stochasticity according to the measurement residual, injecting exploration when reconstructions deviate from the observations and producing stable deterministic updates as measurement fidelity improves. We demonstrate through experiments on Fashion-MNIST and LSUN Bedroom that this sampling method improves perceptual (FID, KID) and distortion (PSNR, SSIM) metrics over baseline CM sampling, achieving competitive reconstructions with only a handful of steps.

II. METHOD

Sampling from consistency models (CMs) is typically carried out through iterative denoising steps, where additional noise is injected to improve reconstruction quality and sample diversity. In the context of inverse problems, however, existing CM-based approaches generally underutilize the sampling process, relying instead on hard measurement constraints after sample generation to encourage reconstruction fidelity [11]

Denoising Diffusion Implicit Model (DDIM) sampling process [13] is a method used for diffusion models sampling, where a prediction of the clean sample is used to denoise the current intermediate sample to the next level. The same approach can be used with CMs, where instead of injecting noise into the sample each time, we use the DDIM step to achieve the next noisy latent for the few-step generative process [14].

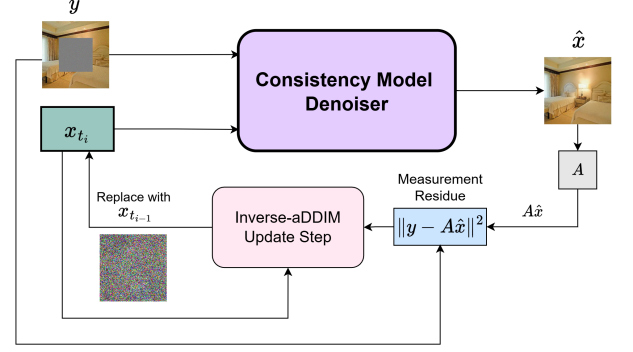


Fig. 1. Illustration of the proposed Inverse-aDDIM framework, where measurement consistency is enforced using the measurement residue

Adjusted DDIM (aDDIM) is an extension of DDIM that adapts the number and size of diffusion steps during sampling to accelerate generation while maintaining high-quality samples [14]

Motivated by the adjusted DDIM (aDDIM) sampler, we propose a measurement-aware sampler that adaptively adjusts its estimation of needed stochasticity according to the inverse problem residue. Our method reformulates the variance term present in the aDDIM formulation to incorporate direct measurement consistency, enabling application to arbitrary CMs trained for inverse tasks.

A. aDDIM Sampler

The variance-exploding update in aDDIM can be expressed as:

$$\hat{\mathbf{x}} = f_\theta(\mathbf{x}_t, t) \quad (8)$$

$$\hat{\epsilon} = \mathbf{x}_t - \hat{\mathbf{x}} \quad (9)$$

$$\mathbf{x}_s = \hat{\mathbf{x}} + \sqrt{\left(\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}\right) + \left(\frac{1 - \sqrt{\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}}}{\|\hat{\epsilon}\|}\right)^2 \mathbf{x}_{\text{var},t}} \hat{\epsilon} \quad (10)$$

where $s = t_{i-1}$ is the next noise level, $t = t_i$ is the current level, and $\mathbf{x}_{\text{var},t} = \eta \|\mathbf{x} - \hat{\mathbf{x}}\|^2$, with η a tunable hyperparameter and \mathbf{x} the teacher signal.

The effectiveness of this modification can be understood by comparing the squared norms of DDIM iterates with those of the hypothetical correct samples. Ideally, one would draw a clean sample $\mathbf{x}^* \sim p(\mathbf{x} | \mathbf{x}_t)$, and map it deterministically to the next latent:

$$\mathbf{x}_s^* = \left(1 - \sqrt{\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}}\right) \mathbf{x}^* + \sqrt{\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}} \mathbf{x}_t \quad (11)$$

Since \mathbf{x}^* carries the correct conditional variance, the resulting \mathbf{x}_s^* preserves the proper distribution.

DDIM, however, instead of sampling \mathbf{x}^* from the true conditional distribution, replaces \mathbf{x}^* with model prediction,

which is its conditional mean $\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x} \mid \mathbf{x}_t]$, yielding the deterministic update:

$$\mathbf{x}_s^{\text{DDIM}} = \left(1 - \sqrt{\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}}\right) \hat{\mathbf{x}} + \sqrt{\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}} \mathbf{x}_t \quad (12)$$

The conditional expectation of the difference of their squared norms thus becomes:

$$\mathbb{E}[\|\mathbf{x}_s^*\|^2 - \|\mathbf{x}_s^{\text{DDIM}}\|^2 \mid \mathbf{x}_t] = \mathbb{E}[\|\mathbf{x}_s^*\|^2 \mid \mathbf{x}_t] - \|\mathbf{x}_s^{\text{DDIM}}\|^2 \quad (13)$$

And since $\mathbf{x}_s^{\text{DDIM}} = \mathbb{E}[\mathbf{x}_s^* \mid \mathbf{x}_t]$:

$$\mathbb{E}[\|\mathbf{x}_s^*\|^2 \mid \mathbf{x}_t] - \|\mathbb{E}[\mathbf{x}_s^* \mid \mathbf{x}_t]\|^2 = \text{trace}(\text{Var}[\mathbf{x}_s \mid \mathbf{x}_t]) \quad (14)$$

The term $\text{trace}(\text{Var}[\mathbf{x}_s \mid \mathbf{x}_t])$ can be calculated as:

$$\text{trace}(\text{Var}[\mathbf{x}_s \mid \mathbf{x}_t]) = \left(1 - \sqrt{\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}}\right)^2 \text{trace}(\text{Var}[\mathbf{x} \mid \mathbf{x}_t])$$

and $\text{trace}(\text{Var}[\mathbf{x} \mid \mathbf{x}_t])$ is approximated by $\eta \|\mathbf{x} - \hat{\mathbf{x}}\|^2$ in aDDIM. $\text{Var}[\mathbf{x}_s \mid \mathbf{x}_t]$ is the variance that has been “dropped” by DDIM. Thus, DDIM produces iterates of systematically smaller norm, often manifesting as over-smoothed samples. The adjustment of the DDIM compensates for this by increasing the noise estimate that must be added to $\hat{\mathbf{x}}$. This increase is proportional to the current conditional variance of \mathbf{x}_s .

Reference [14] uses this sampling algorithm within their training paradigm, where it requires a teacher signal \mathbf{x} , which can be either a training-set data point (in case of consistency training) or the output of a teacher model (in case of consistency distillation). This requirement makes converting aDDIM to a form suitable for any CM inverse problem solver non-trivial.

B. Adapting to Inverse Problems

To extend the aDDIM framework to inverse problems, we propose replacing the variance term $\mathbf{x}_{\text{var},t}$ with the *inverse residue* defined as $\|\mathbf{y} - A\hat{\mathbf{x}}\|^2$, which directly reflects fidelity to the measurement \mathbf{y} . The above replacement yields a measurement-aware sampler that preserves the variance-compensating effect of aDDIM while removing the need for a teacher signal. The full sampling step can be formalized as:

$$\hat{\mathbf{x}} = f_\theta(\mathbf{x}_t, \mathbf{y}, t), \quad \hat{\epsilon} = \mathbf{x}_t - \hat{\mathbf{x}}$$

$$\mathbf{x}_s = \hat{\mathbf{x}} + \sqrt{\left(\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}\right) + \left(\frac{1 - \sqrt{\frac{s^2 - t_{\min}^2}{t^2 - t_{\min}^2}}}{\|\hat{\epsilon}\|}\right)^2 \gamma \|\mathbf{y} - A\hat{\mathbf{x}}\|^2} \hat{\epsilon}$$

Note that the final sample will be the direct output of the CM $f_\theta(\mathbf{x}_t, \mathbf{y}, t)$ after a few iterations of the update.

To justify this substitution, consider a linear forward operator A . Expanding the expectation of the residue gives:

$$\mathbb{E}[\|\mathbf{y} - A\hat{\mathbf{x}}\|^2] = \mathbb{E}[\|A(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{n}\|^2] \quad (15)$$

$$= \|A(\mathbf{x} - \hat{\mathbf{x}})\|^2 + \text{trace}(\Sigma_n), \quad (16)$$

where \mathbf{n} denotes measurement noise with covariance Σ_n . Assuming Gaussian noise, we obtain:

$$\mathbb{E}[\|\mathbf{y} - A\hat{\mathbf{x}}\|^2] = \|A(\mathbf{x} - \hat{\mathbf{x}})\|^2 + m\sigma_y^2. \quad (17)$$

where m is the measurement dimension. Bounding the first term with eigenvalues of A yields:

$$\mathbb{E}[\|\mathbf{y} - A\hat{\mathbf{x}}\|^2] \leq \lambda_{\max}(A^T A) \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + m\sigma_y^2 \quad (18)$$

$$= \|A\|_2^2 \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + m\sigma_y^2 \quad (19)$$

where $\|A\|_2^2$ is the spectral norm of the measurement model. The constant offset $m\sigma_y^2$ can be ignored in practice, while $\|A\|_2^2$ can be absorbed into the hyperparameter. Giving:

$$\mathbb{E}[\|\mathbf{y} - A\hat{\mathbf{x}}\|^2] \leq \eta' \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \quad (20)$$

Thus, the expected inverse residue remains less than $\|\mathbf{x} - \hat{\mathbf{x}}\|^2$, ensuring that our proposed substitution preserves the variance-compensating property of aDDIM while not overshooting the correct variance trace estimate. This modification effectively adapts the algorithm to the measurement-consistency setting of inverse problems.

The general CM sampling process using the Inverse-aDDIM algorithm is illustrated in Fig. 1, where the new noisy latent obtained through our proposed algorithm is denoised via the Inverse CM solver in a few-step iterative manner.

III. EXPERIMENTS

A. Experimental Setup

To validate the effectiveness of the proposed algorithm, we conducted experiments using the CoSIGN base model [11] on the LSUN Bedroom dataset at a resolution of 256×256 [15]. In this framework, a trainable ControlNet is employed to impose measurement constraints and guide the generation process, while keeping the CM backbone frozen [11]. In our study, we did not retrain any of the available checkpoints; instead, we trained ControlNet for the linear deblurring task, which was not previously included.

We further evaluated our approach on the Fashion-MNIST dataset at a resolution of 28×28 [16]. For this dataset, we adopted the base model from [17], where the auxiliary network does not directly enforce the measurement by guiding the output generation but instead produces a representation of it, which is then provided to the CM.

Our evaluation covers three distinct inverse problems. The measurements were generated as follows: (i) for super-resolution, LSUN bedroom images were downsampled to one-quarter of their original resolution in both dimensions, while Fashion MNIST images were downsampled to half; (ii) for inpainting, a central square region with half the side length of the image was masked; and (iii) for the deblurring task, a Gaussian blur kernel with a standard deviation of 5 for the LSUN bedroom dataset and 3 for the Fashion MNIST dataset was applied. Following [11], Gaussian noise with standard deviation $\sigma_y = 0.05$ was added to all measurements.

We also evaluated our method on a non-linear deblurring task for the LSUN Bedroom dataset, where images were blurred using a learned kernel-based nonlinear model, despite the absence of a theoretical foundation.

TABLE I
PERFORMANCE COMPARISON OF BASELINE AND INVERSE-ADDIM
ACROSS FOUR INVERSE TASKS ON THE LSUN BEDROOM DATASET.

Task	Method	PSNR \uparrow	SSIM \uparrow	KID \downarrow	FID \downarrow
4 \times Super-resolution	Baseline	26.03	0.771	5.67	40.90
	Inverse-aDDIM	25.70	0.765	5.50	40.44
Linear Deblurring	Baseline	24.41	0.700	8.61	41.57
	Inverse-aDDIM	24.02	0.690	8.32	40.90
Inpainting	Baseline	22.51	0.830	4.69	39.82
	Inverse-aDDIM	22.11	0.830	4.54	39.44
Nonlinear Deblurring	Baseline	25.85	0.815	3.10	40.70
	Inverse-aDDIM	25.81	0.811	3.07	39.90

TABLE II
PERFORMANCE COMPARISON OF BASELINE AND INVERSE-ADDIM
ACROSS THREE INVERSE TASKS ON THE FASHION-MNIST DATASET.

Task	Method	PSNR \uparrow	SSIM \uparrow	KID \downarrow	FID \downarrow
2 \times Super-resolution	Baseline	18.76	0.518	14.2	22.34
	Inverse-aDDIM	18.97	0.557	12.1	19.85
Linear Deblurring	Baseline	18.22	0.509	15.6	23.09
	Inverse-aDDIM	18.55	0.541	12.9	19.58
Inpainting	Baseline	17.75	0.523	11.5	19.84
	Inverse-aDDIM	17.99	0.569	9.6	17.52

B. Evaluation

The baseline used for comparison is the standard few-step consistency model sampling introduced in [10]. In the case of the CoSIGN model, this corresponds to sampling with the soft measurement constraint alone. Across all experiments, we employed 2-step generation, which consistently yielded the best results. We evaluated performance using both pixel-level and perceptual metrics. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [18] were used to assess reconstruction fidelity. To measure perceptual quality and alignment with human visual perception, we employed Kernel Inception Distance (KID) [19], multiplied by 10^3 , and Fréchet Inception Distance (FID) [20].

C. Results

Quantitative comparisons are reported in Tables I and II, while qualitative results are presented in Figs. 2 and 3. On Fashion-MNIST, our method consistently outperforms the baseline across all metrics. On LSUN Bedroom, Inverse-aDDIM achieves superior KID and FID scores, while maintaining competitive PSNR and SSIM performance. These results highlight the adaptability of our approach to different inverse problem settings and CM solvers. The slight reduction in pixel-level metrics observed on the LSUN Bedroom dataset can be attributed to their sensitivity to hallucinated details, which are often necessary to achieve perceptually realistic reconstructions.



Fig. 2. Visual comparison on four inverse problems on the LSUN Bedroom dataset: super-resolution (top row), deblurring (second row), inpainting (third row), and nonlinear deblurring (bottom row).

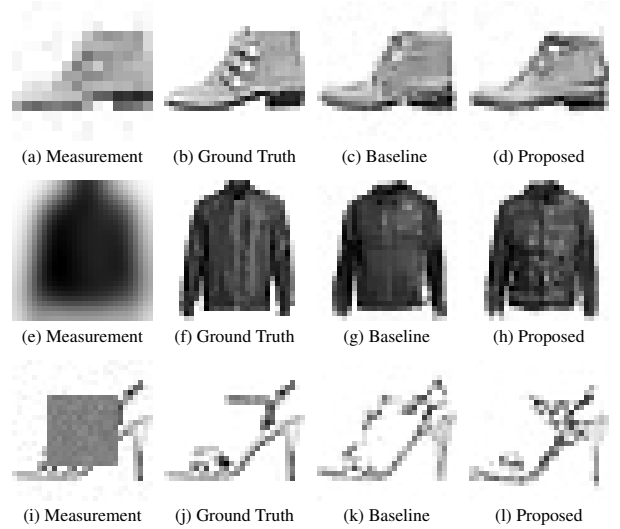


Fig. 3. Visual comparison on three inverse problems: super-resolution (top), deblurring (middle), and inpainting (bottom) on the Fashion-MNIST dataset.

IV. CONCLUSION

We proposed Inverse-aDDIM, a measurement-aware sampling method for consistency models that promotes measurement consistency to inverse problems solvers through variance compensation via inverse residues. It preserves efficiency while improving reconstruction quality across linear and nonlinear tasks. Experiments on Fashion-MNIST and LSUN Bedroom confirm its effectiveness, with future work aimed at more complex operators and task-specific constraints.

REFERENCES

- [1] J. Virieux and S. Operto, “An overview of full-waveform inversion in exploration geophysics,” *Geophysics*, vol. 74, no. 6, pp. WCC1–WCC26, 2009.
- [2] E. Moliner, J. Lehtinen, and V. Välimäki, “Solving audio inverse problems with a diffusion model,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2023, pp. 1–5.
- [3] Y. Song, L. Shen, L. Xing, and S. Ermon, “Solving inverse problems in medical imaging with score-based generative models,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
- [4] B. Kavar, M. Elad, S. Ermon, and J. Song, “Denoising diffusion restoration models,” in *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 23 593–23 606, 2022.
- [5] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, “DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 8878–8887.
- [6] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, “Semantic image inpainting with deep generative models,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5485–5493.
- [7] M. Haris, G. Shakhnarovich, and N. Ukita, “Deep back-projection networks for super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 1664–1673.
- [8] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, “Compressed sensing using generative models,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 537–546.
- [9] H. Chung, J. Kim, and J. C. Ye, “Direct diffusion bridge using data consistency for inverse problems,” in *Adv. Neural Inf. Process. Syst.*, 2023.
- [10] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, “Consistency models,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2023.
- [11] J. Zhao, B. Song, and L. Shen, “CoSIGN: Few-step guidance of consistency model to solve general inverse problems,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Lecture Notes in Computer Science, vol. 13627, Springer, 2024, pp. 108–126.
- [12] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 6840–6851, 2020.
- [13] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2020.
- [14] J. Heek, E. Hoogeboom, and T. Salimans, “Multistep consistency models,” arXiv:2403.06807v3 [cs.LG], Nov. 2024.
- [15] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, “LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop,” arXiv:1506.03365 [cs.CV], 2015.
- [16] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms,” arXiv:1708.07747 [cs.LG], 2017.
- [17] M. Schmitt, V. Pratz, U. Köthe, P.-C. Bürkner, and S. T. Radev, “Consistency models for scalable and fast simulation-based inference,” in *Adv. Neural Inf. Process. Syst.*, 2024.
- [18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [19] M. Binkowski, D. J. Sutherland, M. Arbel, and A. Gretton, “Demystifying MMD GANs,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018.
- [20] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local Nash equilibrium,” in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [21] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, “ILVR: Conditioning method for denoising diffusion probabilistic models,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 14367–14376.
- [22] S. Luo, Y. Tan, L. Huang, J. Li, and H. Zhao, “Latent consistency models: Synthesizing high-resolution images with few-step inference,” arXiv:2310.04378v1 [cs.CV], Oct. 2023.