Solving Inverse Problems with Ambient Diffusion

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

We provide the first framework to solve inverse problems with diffusion models learned from linearly corrupted data. Our method leverages a generative model trained on one type of corruption (e.g. highly inpainted images) to perform posterior sampling conditioned on measurements from a different forward process (e.g. blurred images). This fully unlocks the potential of ambient diffusion models that are essential in scientific applications where access to fully observed samples is impossible or undesirable. Our experimental evaluation shows that diffusion models trained on corrupted data can even outperform models trained on clean data for image restoration in both speed and performance.

1 Introduction

For certain scientific applications, it is expensive or impossible to get access to uncorrupted data [9, 13, 17] but effortless to acquire partially observed samples. It has also been shown that training generators on missing data reduces the memorization of the training set and hence corruption might be a design choice [11, 4, 25]. Prior works have shown how to train Generative Adversarial Networks (GANs) [3], flow models [20] and more recently diffusion models [11, 1, 19, 10, 24] on corrupted data. Yet, it has not been explored how to use models trained on a certain type of corruption (e.g. inpainted data) to solve inverse problems that arise from a different forward process (e.g. downsampling).

We propose the first framework to solve inverse problems with diffusion models learned from linearly corrupted data, as in Ambient Diffusion [11]. Ambient Diffusion models estimate the *ambient score*, i.e. how to best reconstruct given a *corrupted noisy input*. We show how to use these models for inverse problems outside of their training distribution. Our experiments show that Ambient Models outperform (in the high corruption regime) models trained on clean data. Further, they do so while being significantly faster. Our algorithm extends Diffusion Posterior Sampling [7] to Ambient Models and fully unlocks the potential of generative models trained on corrupted data for image restoration.

2 Method

Background and Notation. Diffusion models are typically trained (up to network reparametrizations) to reconstruct a clean image $\boldsymbol{x}_0 \sim p_0(\boldsymbol{x}_0)$ from a noisy observation $\boldsymbol{x}_t = \boldsymbol{x}_0 + \sigma_t \boldsymbol{\eta}, \quad \eta \sim \mathcal{N}(\mathbf{0}, I)$. Despite the simplicity of the training objective, diffusion models can approximately sample from $p(\boldsymbol{x})$ by running a discretized version of the Stochastic Differential Equation:

$$\mathrm{d}\boldsymbol{x} = -2\dot{\sigma}_t(\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] - \boldsymbol{x}_t)\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}, \tag{2.1}$$

where $\boldsymbol{w}$ is the standard Wiener process and $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$ is estimated by the trained neural network. Given a measurement $\boldsymbol{y}_{\mathrm{inf}} = A_{\mathrm{inf}}\boldsymbol{x}_0$, one can sample from the posterior distribution $p(\boldsymbol{x}_0|\boldsymbol{y}_{\mathrm{inf}})$ by

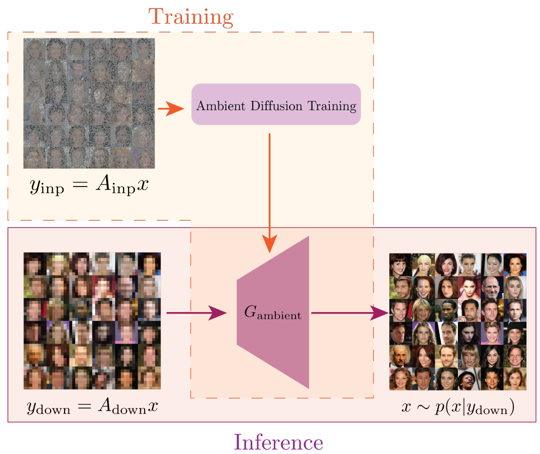


Figure 1: Illustration of the Ambient Diffusion Posterior Sampling (Ambient DPS) setting. During training, we only have access to linearly corrupted data from a forward operator A_{train}. We use this data and the Ambient Diffusion framework to learn a generative model, G_{ambient}, for the uncorrupted distribution, $p(\boldsymbol{x}_0)$. At inference time, we use the learned generator to sample from the posterior distribution $p(\boldsymbol{x}_0|\boldsymbol{y}_{\text{A}_{\text{inf}}})$, for measurements $\boldsymbol{y}_{\text{inf}}$ coming from a different forward operator, A_{inf}.

running the process:

$$\mathrm{d}\boldsymbol{x} = -2\dot{\sigma}_t\sigma_t \left(\frac{\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] - \boldsymbol{x}_t}{\sigma_t} + \underbrace{\nabla \log p(\boldsymbol{y}_{\text{inf}}|\boldsymbol{x}_t)}_{\text{likelihood term}} \right) \mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}. \tag{2.2}$$

For most forward operators it is intractable to write the likelihood in closed-form. Hence, several approximations have been proposed to use diffusion models for inverse problems [7, 18, 17, 26, 8, 12, 14]. One of the simplest and most effective approximations is Diffusion Posterior Sampling (DPS) [7]. DPS estimates $\boldsymbol{x}_0$ using $\boldsymbol{x}_t$ and uses the conditional likelihood $p(\boldsymbol{y}_{\text{inf}}|\hat{\boldsymbol{x}}_0)$ instead of the intractable term, i.e. DPS approximates $p(\boldsymbol{y}_{\text{inf}}|\boldsymbol{x}_t)$ with $p(\boldsymbol{y}_{\text{inf}}|\boldsymbol{x}_0 = \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t])$. The update rule becomes:

$$\mathrm{d}\boldsymbol{x} = -2\dot{\sigma}_t\sigma_t \left(\frac{\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] - \boldsymbol{x}_t}{\sigma_t} + \gamma_t \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}_{\text{inf}}|\boldsymbol{x}_0 = \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]) \right) \mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}, \tag{2.3}$$

where γ_t is a tunable guidance parameter.

Ambient Diffusion Posterior Sampling. As mentioned, in some settings we do not have uncorrupted training data but we have access to lossy measurements that we want to leverage to train a diffusion model for the clean distribution.

The authors of [11] consider the setting of having access to linearly corrupted data $\{\boldsymbol{y}_0 = A_{\text{train}} x_0, A_{\text{train}}\}$, where the distribution of A_{train}, denoted as $p(A_{\text{train}})$, is assumed to be known. For this corruption setting, they provide a framework to learn the best restoration model for $\boldsymbol{x}_0$ given any noisy and linearly corrupted observation $\boldsymbol{y}_{t,\text{train}} = A_{\text{train}}(\boldsymbol{x}_0 + \sigma_t \boldsymbol{\eta})$, for $\boldsymbol{\eta} \sim \mathcal{N}(0, I_n)$. In other words, Ambient Diffusion learns $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{y}_{t,\text{train}}, A_{\text{train}}]$ for all noise levels t, as long as some technical conditions on the corruption process are satisfied.

DPS requires access to $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$ to approximately sample from $p(\boldsymbol{x}_0|\boldsymbol{y}_{\text{inf}})$. Since Ambient Diffusion models can only work with corrupted inputs, we propose the following update rule instead:

$$\mathrm{d}\boldsymbol{x} = -2\dot{\sigma}_t\sigma_t \left(\underbrace{\frac{\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{y}_{t,\text{train}}, A_{\text{train}}] - \boldsymbol{x}_t}{\sigma_t}}_{\text{Ambient Score}} + \gamma_t \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}_{\text{inf}}|\boldsymbol{x}_0 = \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{y}_{t,\text{train}}, A_{\text{train}}]) \right) \mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}, \tag{2.4}$$

for a fixed $A_{\text{train}} \sim p(A_{\text{train}})$. Comparing this to the DPS update rule (E.q. 2.3), all the $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$ terms have been replaced with their ambient counterparts, i.e. with $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{y}_{t,\text{train}}, A_{\text{train}}]$. We remark that, similar to DPS, the proposed algorithm is an approximation to sampling from the true posterior distribution $\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{y}_{\text{inf}}]$. We term our approximate sampling algorithm for solving inverse problems with diffusion models learned from corrupted data **Ambient DPS**.

3 Experiments

Setup. In this section, we evaluate the performane of Ambient DPS, that uses diffusion models trained on corrupted data, and we compare it to DPS, that uses diffusion models trained on clean data. For our experiments, we use the models from the Ambient Diffusion [11] that are trained on randomly inpainted data with different erasure probabilities. Specifically, for AFHQ we use the Ambient Models with erasure probability $p \in \{0.2, 0.4, 0.6, 0.8\}$ and for Celeb-A we use the pretrained models with $p \in \{0.6, 0.8, 0.9\}$.

We underline that all the Ambient Models have worse performance for unconditional generation compared to the models trained with clean data (i.e. the models trained with $p = 0.0$). The goal of this work is to explore the conditional generation performance of Ambient Models, where the conditioning is in the measurements $\boldsymbol{y}_{\mathrm{inf}}$, and compare it with models trained on uncorrupted data. To ensure that Ambient Models do not have an unfair advantage, we test only on restoration tasks that are different from the ones encountered in their training. Specifically, we use models trained on random inpainting and we evaluate on Gaussian Compressed Sensing [2] and Super Resolution.

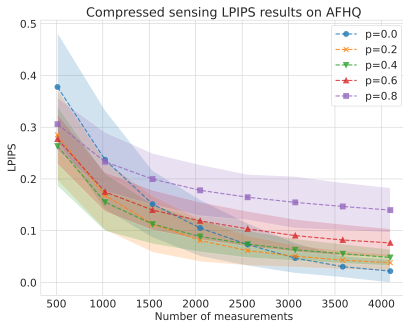
Hyperparameters. The only tunable parameters for DPS (Eq. 2.3) and Ambient DPS (Eq. 2.4) are in the scheduling of the magnitude of the measurements likelihood term. In all the experiments in the DPS paper, this term is kept constant throughout the diffusion sampling trajectory and the authors recommend selecting a value in the range between $[0.1, 10]$. We follow this recommendation and we keep this term constant. The value of the step size for each model is selected with a hyperparameter search in the recommended range. For all our experiments, we follow exactly the DPS implementation provided in the official code repository of the paper. The other parameter that impacts performance is the number of steps we are going to run each algorithm for, i.e. the discretization level of the SDEs of Equations 2.3, 2.4. Typically, the higher the number of steps the better the performance since the discretization error decreases [5, 6]. For the performance results, we run each method for a number of steps $\in \{50, 100, 150, 200, 250, 300\}$, and we report the best result among them.

Results. Figure 2 presents Gaussian Compressed Sensing reconstruction results (i.e. reconstructing a signal from Gaussian random projections). We show MSE and LPIPS performance metrics for the AFHQ dataset as we vary the number of measurements. The results are given for models that are *trained with inpainted* images at different levels of corruption, indicated by the erasure probability p. As shown in the Figure, the model trained with clean data outperforms the models trained with corrupted data when the number of measurements is high. However, as we reduce the number of measurements, Ambient Models outperform the models trained with clean data in the very low measurements corruption regime. To the best of our knowledge, there is no known theoretical argument that explains this performance cross-over and understanding this further is an interesting research direction. Similar results are presented in Figure 3 for the task of super-resolution at AFHQ. The model trained on clean data ($p = 0$) slightly outperforms the Ambient Models in both LPIPS and MSE for reconstructing a $2\times$ downsampled image, as expected. Yet, as the resolution decreases, there is again a cross-over in performance and models trained on corrupted data start to outperform the models trained on uncorrupted data. We include results for LPIPS and MSE for Compressed Sensing and Downsampling in FFHQ and Celeb-A in the Appendix (Figs 5, 6, 7, 8).

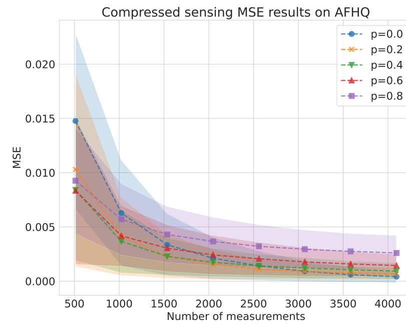
Finally, we ablate how the number of sampling steps affects the performance. The MSE results for Compressed Sensing with 4000 measurements on AFHQ are shown in Figure 4. As shown, the higher the erasure probability p during training, the better the Compressed Sensing performance of the model for low Number of Function Evaluations (NFEs). Models trained with higher corruption are faster since they require fewer steps for the same performance. For increased NFEs, the models that are trained on clean(er) data finally outperform. This result is consistent across different datasets (AFHQ, FFHQ, CelebA), reconstruction tasks (Compressed Sensing, Downsampling) and metrics (MSE, LPIPS) (Figures 9, 10, 11, 12, 13 in the Appendix).

4 Conclusions

We presented a simple framework based on DPS for solving inverse problems with Ambient Diffusion models. We showed that diffusion models trained on missing data are state-of-the-art inverse problem solvers for high corruption levels even if the forward process at inference time is different from the

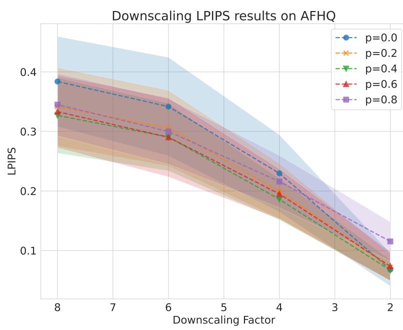


(a) LPIPS per Number of Measurements.

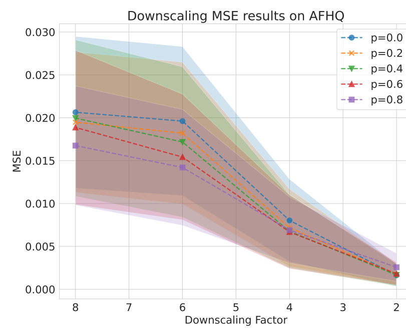


(b) MSE per Number of Measurements.

Figure 2: Compressed Sensing results, AFHQ: performance metric and standard deviation. As shown, the model trained with clean data ($p = 0.0$) only outperforms the models trained with corrupted data for more than 1000 measurements, in both LPIPS and MSE.

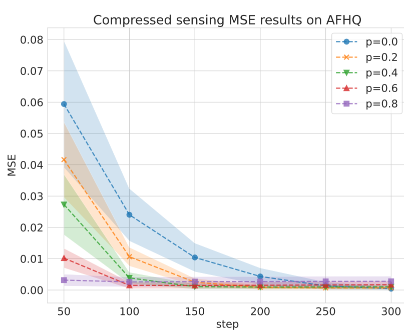


(a) LPIPS per downsampling factor.

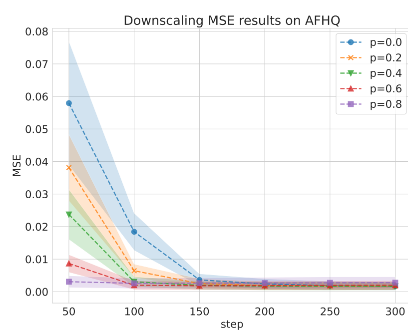


(b) MSE per downsampling factor.

Figure 3: Super-resolution results, AFHQ: Performance metric and standard deviation. The model trained with clean data ($p = 0.0$) performs worse, except at downscaling factor 2.



(a) Compressed Sensing with 4000 measurements per Number of Function Evaluations (NFEs).



(b) $2\times$ Super-Resolution per Number of Function Evaluations (NFEs).

Figure 4: Speed performance plots for AFHQ.

one used during training. Our framework fully unlocks the potential of Ambient Diffusion models that are critical in applications where access to full data is impossible or undesirable.

References

[1] Asad Aali, Marius Arvinte, Sidharth Kumar, and Jonathan I Tamir. "Solving Inverse Problems with Score-Based Generative Priors learned from Noisy Data". In: *arXiv preprint arXiv:2305.01166* (2023) (page 1).

[2] Richard G Baraniuk. "Compressive sensing [lecture notes]". In: *IEEE signal processing magazine* 24.4 (2007), pp. 118–121 (page 3).

[3] Ashish Bora, Eric Price, and Alexandros G Dimakis. "AmbientGAN: Generative models from lossy measurements". In: *International conference on learning representations*. 2018 (page 1).

[4] Nicolas Carlini, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramer, Borja Balle, Daphne Ippolito, and Eric Wallace. "Extracting training data from diffusion models". In: *32nd USENIX Security Symposium (USENIX Security 23)*. 2023, pp. 5253–5270 (page 1).

[5] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R Zhang. "Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions". In: *arXiv preprint arXiv:2209.11215* (2022) (page 3).

[6] Sitan Chen, Giannis Daras, and Alex Dimakis. "Restoration-degradation beyond linear diffusions: A non-asymptotic analysis for ddim-type samplers". In: *International Conference on Machine Learning*. PMLR. 2023, pp. 4462–4484 (page 3).

[7] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. "Diffusion Posterior Sampling for General Noisy Inverse Problems". In: *The Eleventh International Conference on Learning Representations*. 2023. URL: `https://openreview.net/forum?id=OnD9zGAGT0k` (pages 1, 2).

[8] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. "Improving diffusion models for inverse problems using manifold constraints". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 25683–25696 (page 2).

[9] The Event Horizon Telescope Collaboration et al. "First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole". In: *The Astrophysical Journal Letters* 875.1 (Apr. 2019), p. L4. DOI: `10.3847/2041-8213/ab0e85`. URL: `https://dx.doi.org/10.3847/2041-8213/ab0e85` (page 1).

[10] Zhuo-Xu Cui, Chentao Cao, Shaonan Liu, Qingyong Zhu, Jing Cheng, Haifeng Wang, Yanjie Zhu, and Dong Liang. "Self-score: Self-supervised learning on score-based models for mri reconstruction". In: *arXiv preprint arXiv:2209.00835* (2022) (page 1).

[11] Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alexandros G Dimakis, and Adam Klivans. "Ambient Diffusion: Learning Clean Distributions from Corrupted Data". In: *arXiv preprint arXiv:2305.19256* (2023) (pages 1–3).

[12] Berthy T Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L Bouman, and William T Freeman. "Score-Based diffusion models as principled priors for inverse imaging". In: *arXiv preprint arXiv:2304.11751* (2023) (page 2).

[13] Angela F Gao, Oscar Leong, He Sun, and Katherine L Bouman. "Image Reconstruction without Explicit Priors". In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–5 (page 1).

[14] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. "Diffusion models as plug-and-play priors". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 14715–14728 (page 2).

[15] Reinhard Heckel and Paul Hand. *Deep Decoder: Concise Image Representations from Untrained Non-convolutional Networks*. 2019. arXiv: `1810.03982 [cs.CV]`.

[16] Yuyang Hu, Mauricio Delbracio, Peyman Milanfar, and Ulugbek S. Kamilov. "A Restoration Network as an Implicit Prior". In: (2023). arXiv:2310.01391.

[17] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. "Robust compressed sensing mri with deep generative priors". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 14938–14954 (pages 1, 2).

[18] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. "Denoising Diffusion Restoration Models". In: *Advances in Neural Information Processing Systems* (page 2).

[19] Bahjat Kawar, Noam Elata, Tomer Michaeli, and Michael Elad. "GSURE-Based Diffusion Model Training with Corrupted Data". In: *arXiv preprint arXiv:2305.13128* (2023) (page 1).

[20] Varun A Kelkar, Rucha Deshpande, Arindam Banerjee, and Mark A Anastasio. “AmbientFlow: Invertible generative models from incomplete, noisy measurements”. In: *arXiv preprint arXiv:2309.04856* (2023) (page 1).

[21] Kwanyoung Kim and Jong Chul Ye. “Noise2score: tweedie’s approach to self-supervised image denoising without clean images”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 864–874.

[22] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. “Noise2Noise: Learning image restoration without clean data”. In: *arXiv preprint arXiv:1803.04189* (2018).

[23] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk. “From denoising to compressed sensing”. In: *IEEE Transactions on Information Theory* 62.9 (2016), pp. 5117–5144.

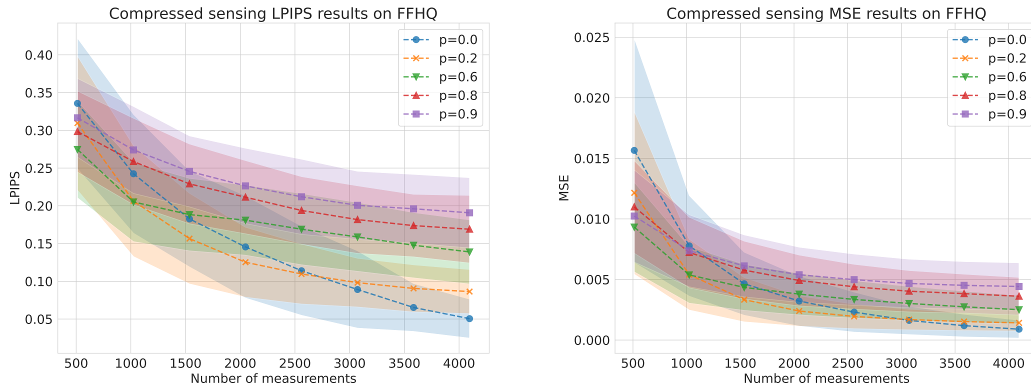
[24] Charles Millard and Mark Chiew. “A Theoretical Framework for Self-Supervised MR Image Reconstruction Using Sub-Sampling via Variable Density Noisier2Noise”. In: *IEEE Transactions on Computational Imaging* 9 (2023), pp. 707–720. DOI: 10.1109/TCI.2023.3299212 (page 1).

[25] Gowthami Somepalli, Vasu Singla, Micah Goldblum, Jonas Geiping, and Tom Goldstein. “Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models”. In: *arXiv preprint arXiv:2212.03860* (2022) (page 1).

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

[27] Hongkai Zheng, Weili Nie, Arash Vahdat, and Anima Anandkumar. “Fast Training of Diffusion Models with Masked Transformers”. In: *arXiv preprint arXiv:2306.09305* (2023).

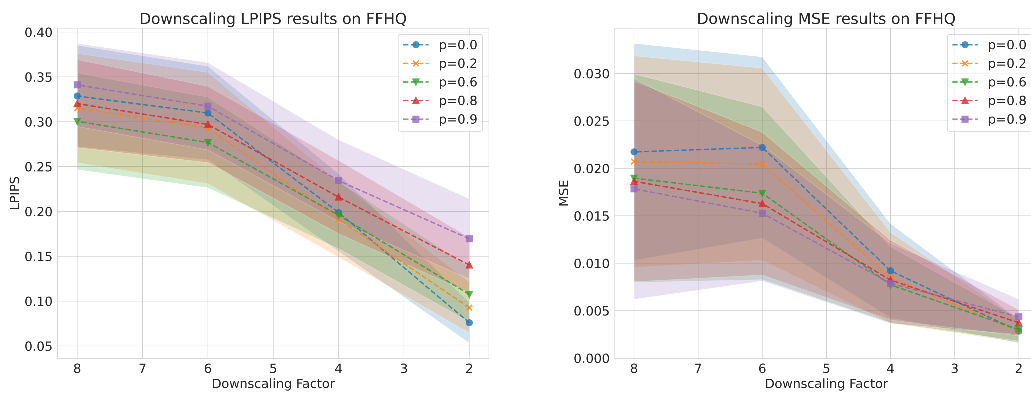
A Additional Performance Results



(a) LPIPS per Number of Measurements.

(b) MSE per Number of Measurements.

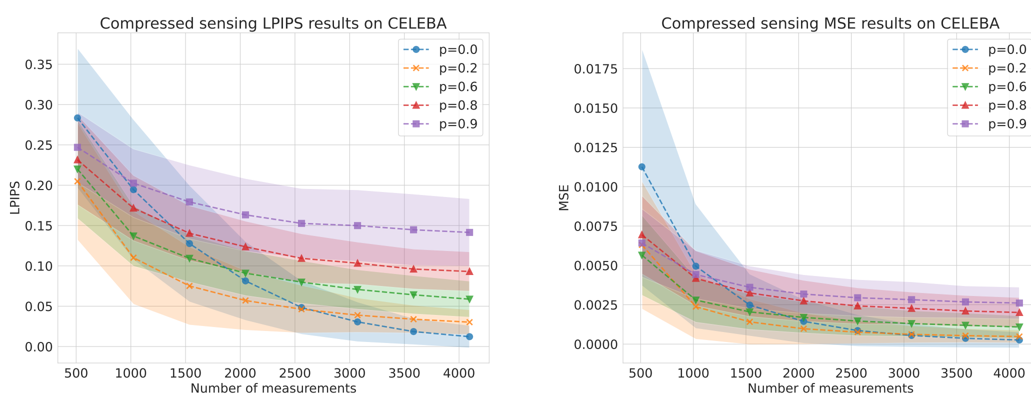
Figure 5: Compressed Sensing Results for FFHQ.



(a) LPIPS per Downscaling Factor.

(b) MSE per Downscaling Factor.

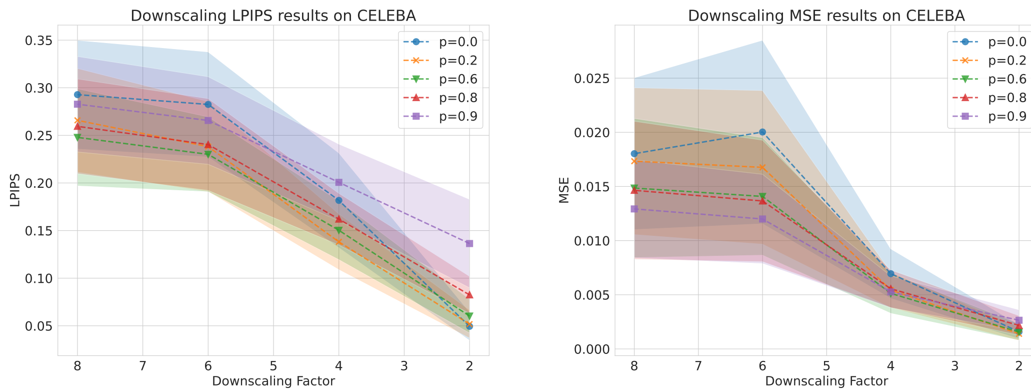
Figure 6: Downscaling Results for FFHQ.



(a) LPIPS per Number of Measurements.

(b) MSE per Number of Measurements.

Figure 7: Compressed Sensing Results for Celeb-A.

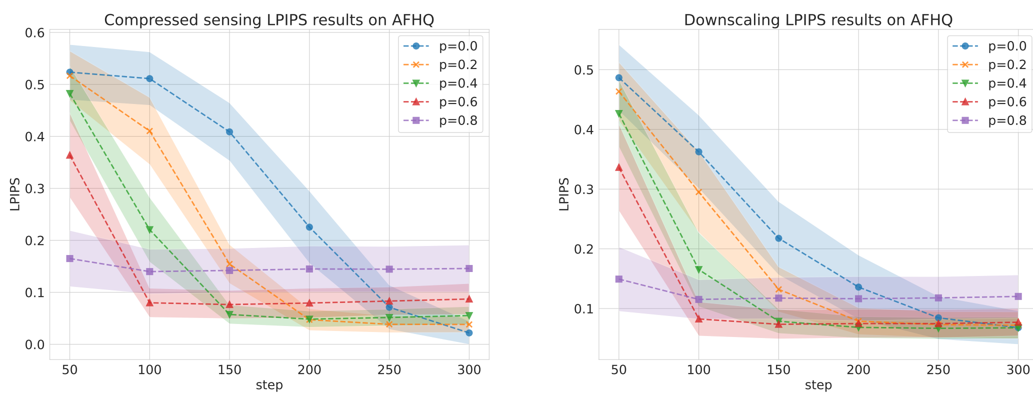


(a) LPIPS per Downscaling Factor.

(b) MSE per Downscaling Factor.

Figure 8: Downscaling Results for Celeb-A.

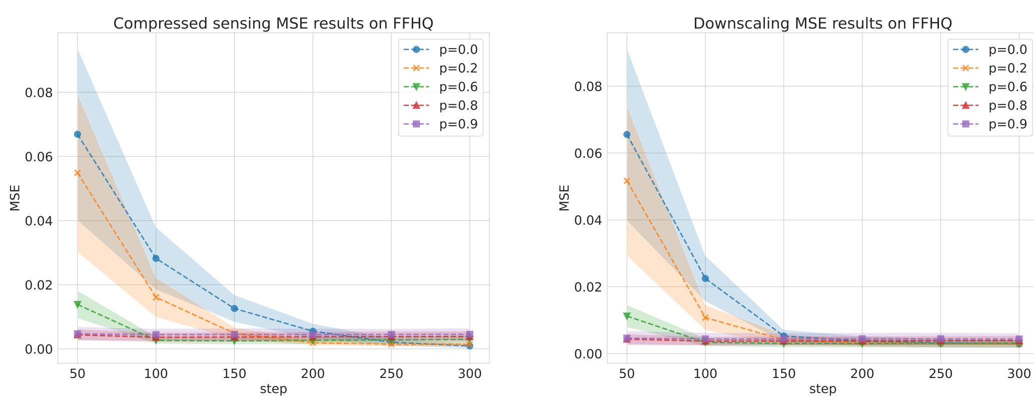
B Additional Speed Results



(a) Compressed Sensing with 4000 measurements per Number of Function Evaluations (NFEs).

(b) $2\times$ Super-Resolution per Number of Function Evaluations (NFEs).

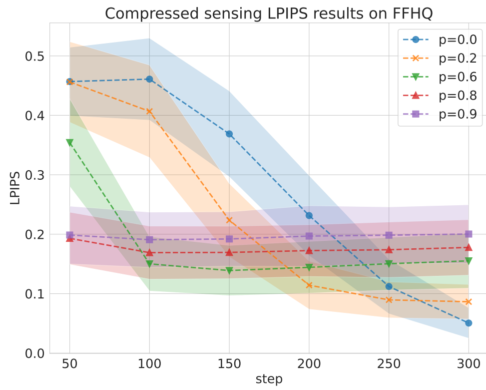
Figure 9: Speed LPIPS performance plots for AFHQ.



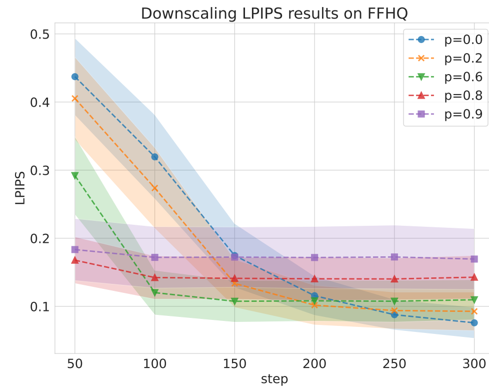
(a) Compressed Sensing with 4000 measurements per Number of Function Evaluations (NFEs).

(b) $2\times$ Super-Resolution per Number of Function Evaluations (NFEs).

Figure 10: Speed MSE performance plots for FFHQ.

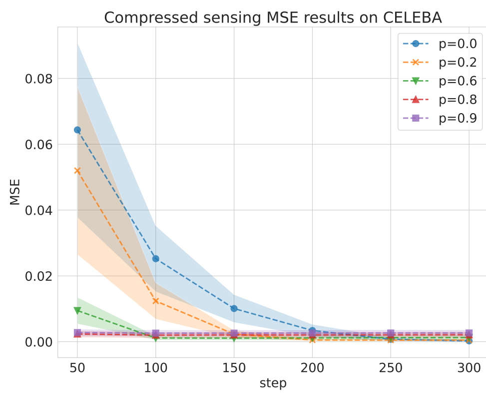


(a) Compressed Sensing with 4000 measurements per Number of Function Evaluations (NFEs).

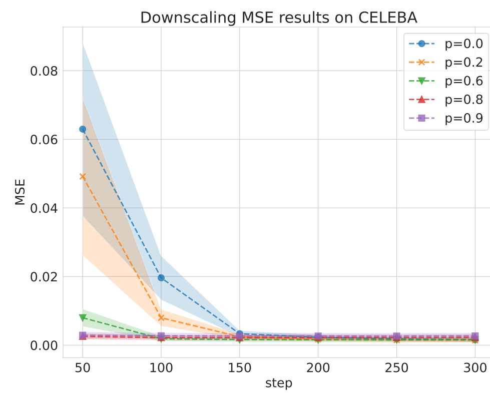


(b) 2× Super-Resolution per Number of Function Evaluations (NFEs).

Figure 11: Speed LPIPS performance plots for FFHQ.

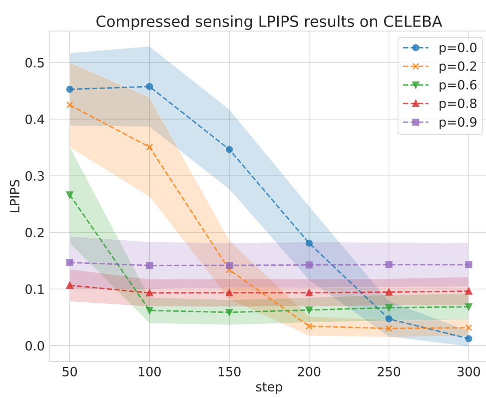


(a) Compressed Sensing with 4000 measurements per Number of Function Evaluations (NFEs).

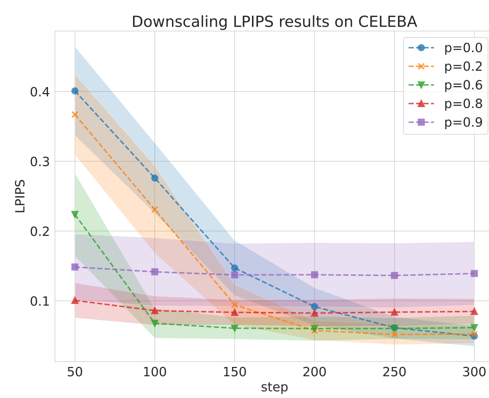


(b) 2× Super-Resolution per Number of Function Evaluations (NFEs).

Figure 12: Speed MSE performance plots for Celeb-A.



(a) Compressed Sensing with 4000 measurements per Number of Function Evaluations (NFEs).



(b) 2× Super-Resolution per Number of Function Evaluations (NFEs).

Figure 13: Speed LPIPS performance plots for Celeb-A.