# RESTORATION BASED GENERATIVE MODELS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Denoising diffusion models (DDMs) have recently attracted increasing attention by showing impressive synthesis quality. DDMs are built on a diffusion process that pushes data to the noise distribution and the models learn to denoise. In this paper, we establish the interpretation of DDMs in terms of image restoration (IR). Integrating IR literature allows us to use an alternative objective and diverse forward processes, not confining to the diffusion process. By imposing prior knowledge on the loss function grounded on MAP estimation, we eliminate the need for the expensive sampling of DDMs. Also, we propose a multi-scale training, which improves the performance compared to the diffusion process, by taking advantage of the flexibility of the forward process. Our model improves the quality and efficiency of both training and inference, Furthermore, we show the applicability of our model to inverse problems. We believe that our framework paves the way for designing a new type of flexible general generative model.

## 1 INTRODUCTION

Generative modeling is a prolific machine learning task that the models learn to describe how a dataset is distributed and generate new samples from the distribution. The most widely used generative models primarily differ in their choice of bridging the data distribution to a tractable latent distribution (Goodfellow et al., 2020; Kingma & Welling, 2014; Rezende et al., 2014; Rezende & Mohamed, 2015; Sohl-Dickstein et al., 2015; Chen et al., 2021a). In recent years, denoising diffusion models (DDMs) (Ho et al., 2020; Song & Ermon, 2019; Song et al., 2020b; Dockhorn et al., 2021) have drawn considerable attention by demonstrating remarkable results in terms of both high sample quality and likelihood. DDMs rely on a forward diffusion process that progressively transforms the data into Gaussian noise, and they learn to reverse the noising process. Albeit their enormous successes, their forward process is fixed as a diffusion process, which gives rise to a few limitations. To pull latent variables back to the data distribution, the denoising process requires thousands of network evaluations to sample a single instance. Many follow-up studies consider enhancing inference speed (Song et al., 2020a; Jolicoeur-Martineau et al., 2021; Tachibana et al., 2021) or grafting with other generative models (Xiao et al., 2021a; Vahdat et al., 2021; Zhang & Chen, 2021; Pandey et al., 2022).

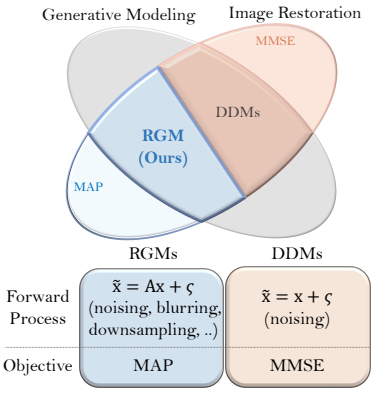


Figure 1: Comparison of DDMs and RGMs.

In this study, we focus on a different perspective. We interpret the DDMs through the lens of image restoration (IR), which is a family of inverse problems for recovering the original images from corrupted ones (Castleman, 1996; Gunturk & Li, 2018). The corruption arises in various forms, including noising (Buades et al., 2005; Rudin et al., 1992), blurring (Biemond et al., 1990), and downsampling (Farsiu et al., 2004). IR has been a long-standing problem because of its high practical value in various applications (Besag et al., 1991; Banham & Katsaggelos, 1997; Lehtinen et al., 2018; Ma et al., 2011). From an IR point of view, DDMs can be considered as IR models based on minimum mean square error (MMSE) estimation (Zervakis & Venetsanopoulos, 1991; Laumont et al., 2022), focusing only on the denoising task. Mathematically, IR is an ill-posed inverse problem in the sense that it does not admit a unique solution and hence, leads to instability in reconstruction (Hadamard, 1902). Owing to the ill-posedness of IR, MMSE which only measures data fidelity

produces impertinent results. DDMs alleviate this problem by leveraging costly Langevin sampling, and this inefficient inference scheme has been regarded as an indispensable tool in the literature of DDMs. By casting DRMs as IR models, however, the forward process need not be restricted to Gaussian noising, and ill-posedness can be detoured in ways other than Langevin dynamics.

Inspired by this observation, we propose a new flexible family of generative models, that we refer to as *restoration-based generative models (RGMs)*. First, we adopt an alternative objective; a maximum a posteriori (MAP) (Trussell, 1980; Hunt, 1977), which is predominantly used in IR. The MAP estimator compensates the ill-posedness by regularizing the data fidelity loss by a prior term. Many advent hand-crafted regularization schemes (Tikhonov, 1963; Donoho, 1995; Mallat, 1999; Baraniuk, 2007) encourage solutions to satisfy certain properties, such as smoothness and sparsity. However, for the purpose of density estimation, we implicitly parameterize a prior term as a variational regularization via GAN (Goodfellow et al., 2020) with a newly introduced random auxiliary variable. Our MAP approach retains the density estimating capability of DDMs at a much smaller computation cost. Secondly, unlike DDMs, which are buried in a Gaussian noising process, RGMs can be combined with other general degradation processes. As one instantiation, we design a multi-scale training that resolves the latent inefficiency of DDMs. Because the behavior of generative models is significantly affected by how the data distribution is transformed into a simple distribution, our approach opens the way for designing more flexible generative models. Our comprehensive empirical studies on image generation and inverse problems demonstrate that RGMs generate samples rivaling the quality of DDMs. Also, the inference of our model is several orders of magnitude faster than DDMs. In particular, our model achieve FID 2.47 on CIFAR10, with only seven number of network function evaluations.

## 2 Background

**Image Restoration** A common inverse problem arising in image processing, including denoising, deblurring, super-resolution, and inpainting, is the estimation of an image $\mathbf{x}$ given a corrupted image

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \xi, \tag{1}$$

where $\mathbf{A}$ is a matrix that models the degradation process, including blurring and downsampling kernels, and $\xi \sim \mathcal{N}(0, \mathbf{\Sigma})$ is an additive noise. A family of such problems are known as image restoration (IR). The inference of the image $\mathbf{x}$ from the noised one $\mathbf{y}$ is typically ill-posed, in the sense that the inverse problem (1) has multiple valid explanations (Hadamard, 1902). In other words, the noisy $\mathbf{y}$ does not have exactly one restoration $\mathbf{x}$. This is further exacerbated when the noise level is large. To produce consistent results, most methods use the maximum a posteriori (MAP) estimator:

$$\mathbf{x}^*_{\text{MAP}} = \text{argmax}_{\mathbf{x}} \ \log p(\mathbf{x} \mid \mathbf{y}) = \text{argmin}_{\mathbf{x}} \ f(\mathbf{x}, \mathbf{y}) + \lambda g(\mathbf{x}), \tag{2}$$

where $f(\mathbf{x}, \mathbf{y}) = -\log p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{2} \left\| \left(\mathbf{\Sigma}^\dagger\right)^{\frac{1}{2}} (\mathbf{A}\mathbf{x} - \mathbf{y}) \right\|_2^2$ is the data fidelity term with the pseudoinverse (Moore, 1920) $\mathbf{\Sigma}^\dagger$, $g$ is the prior term that encourages the reconstruction to satisfy some prior assumptions on $\mathbf{x}$, and a scalar $\lambda \geq 0$ controls the strength of the regularization. The regularization term $g$ is essential because it relieves the ill-posedness nature of the inverse problem by imposing the assumption about the desirable solution. Therefore, many researchers have been devoted to designing a proper $g$ (Rudin et al., 1992; Mallat, 1999; Lunz et al., 2018).

**Denoising Generative Models** Denoising diffusion models (DDMs), such as DDPM (Ho et al., 2020), and score matching with Langevin dynamics (Song et al., 2020b), have recently emerged as the forefront of image synthesis research. Starting from the data distribution, they gradually corrupt the image $\mathbf{x}_0 \sim p_{\text{data}}$ into Gaussian noise over time through a forward Markovian diffusion process;

$$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0) = \prod_{t=0}^{T-1} q^{(t)}(\mathbf{x}_{t+1} \mid \mathbf{x}_t), \ \mathbf{x}_0 \sim p_{\text{data}}. \tag{3}$$

They pose the data generation as an iterative denoising procedure $p_\theta^{(t)}(\mathbf{x}_t \mid \mathbf{x}_{t+1})$, the reverse of the forward diffusion process:

$$p_\theta(\mathbf{x}_{0:T}) = p^{(T)}(\mathbf{x}_T) \prod_{t=0}^{T-1} p_\theta^{(t)}(\mathbf{x}_t \mid \mathbf{x}_{t+1}), \ \mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{4}$$

As they use linear diffusion process whose diffusion distributions $q^{(t)}\left(\mathbf{x}_{t+1} \mid \mathbf{x}_t\right)$ are modeled with conditional Gaussian distributions, we can deduces a tractable evidence lower bound (ELBO) (Sohl-Dickstein et al., 2015). The ELBO can be further simplified to the following objective (Ho et al., 2020; Song et al., 2020b):

$$\mathcal{L}\left(\theta\right) = \Sigma_{t=0}^{T} \mathbb{E}_{\mathbf{x}_0 \sim p_{\text{data}}, \mathbf{x}_t \sim q_{\sigma_t}\left(\mathbf{x}_t \mid \mathbf{x}_0\right)} \left[\lambda\left(t\right) \left\| G_\theta\left(\mathbf{x}_t, t\right) - \mathbf{x}_0 \right\|_2^2\right], \tag{5}$$

where $G_\theta$ is a neural network parametrized by $\theta$ that learns the noise by minimizing (5), and $\lambda\left(t\right) \geq 0$.

# 3 METHOD

## 3.1 DDMS ARE RESTORATION MODELS

We open this section by drawing the interpretation of DDMs in terms of restoration. DDMs use VPSDE or VESDE (Song et al., 2020b) as a forward process, and these two are known to be exchangeable with each other (Kim et al., 2022). Therefore, the rest of the paper focuses on VESDE. For a given noise level $\sigma_t$, the forward process of the VESDE is formulated as

$$\mathbf{x}_t = \mathbf{x}_0 + \xi,\ \xi \sim \mathcal{N}\left(\mathbf{0}, \sigma_t^2 \mathbf{I}\right),$$

which is the forward process (1) with identity degradation matrix $\mathbf{A} = \mathbf{I}$. As $\nabla_{\mathbf{x}_t} \log q_{\sigma_t}\left(\mathbf{x}_t \mid \mathbf{x}_0\right) = -\left(\mathbf{x}_t - \mathbf{x}_0\right)/\sigma_t^2$, the loss (5) for each forward step can be rewritten as the following minimum mean square error (MMSE) objective:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}, \mathbf{x}_t \sim \mathcal{N}\left(\mathbf{x}, \sigma_t^2 \mathbf{I}\right)} \left[\left\| G_\theta(\mathbf{x}_t, t) - \mathbf{x} \right\|_2^2\right]. \tag{6}$$

Therefore, DDMs are IR models which seek a denoiser $G_\theta$ for each $\sigma_t$ that minimizes the MMSE (6). MMSE loss is simple and straightforward to train, however, it confronts some apparent drawbacks. Since the MMSE only contains the recovery term, the solution is only optimized to ensure accordance with the degradation process. Therefore, it is affected by the ill-posedness. To be precise, when $\sigma_t$ is large, (1) becomes a highly ill-posed and possess many solutions for a given observation. In this case, MMSE solution averages all these candidate solutions, resulting in an atypical reconstruction. Formally, the solution to the MMSE for a corrupted data $\mathbf{x}_t$ is

$$\mathbf{x}_{\text{MMSE}}^* = \int \mathbf{x} p_{\text{data}}\left(\mathbf{x}\right) p_{\sigma_t}\left(\mathbf{x}_t \mid \mathbf{x}\right) d\mathbf{x}.$$

Recent works (Laumont et al., 2022; Kawar et al., 2021) have endeavored to resolve this problem by stochastic sampling, however, they suffer from notoriously low efficiency as they roll out thousands of trajectories. In a similar manner, DDMs utilize a sampling scheme that requires thousands of steps. In summary, there are two limitations of DDMs from the IR perspective:

1. The degradation process is restricted to Gaussian noising.
2. The inference efficiency is intrinsically low due to the MMSE estimator.

In the following sections, we show how to cope with these two limitations.

## 3.2 MAP ESTIMATION FOR GENERATION

As alluded in Section 3.1, DDMs can be regarded as MMSE grounded IR models, specialized in denoising. This observation brings us a new perspective on the design of a family of flexible generative models. As an alternative to MMSE, we propose a new generative model based on the MAP (2):

$$\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}, \mathbf{y} \sim \mathcal{N}\left(\mathbf{x}, \sigma^2 \mathbf{I}\right)} \left[\frac{1}{2\sigma^2} \left\| G_\theta(\mathbf{y}) - \mathbf{y} \right\|_2^2 + \lambda g(G_\theta(\mathbf{y}))\right], \tag{7}$$

where the second term delivers the prior knowledge of data distribution. MAP has been adopted as a standard approach for high-dimensional imaging problems and is known to be more relevant than MMSE in many applications (Saha et al., 2009; Bigdeli et al., 2019; Chen, 2016). By leveraging prior information on the solution, MAP approaches alleviate the ill-posedness of the inverse problem (1), without use of costly sampling of MMSE estimation. Therefore, carefully crafting the relevant prior term is crucial. We now show how one can execute an appropriate prior term for density estimation while alleviating the ill-posedness.

**Alleviation of ill-posedness** Unlike the general denoising task, it is necessary to bridge the image to the Gaussian noise to learn the data distribution. As the noise level increases, a single distorted observation has several solutions, which indicates that the ill-posedness deepens. This in turn degrades the expressiveness of the model for estimating the distribution. Therefore, it is difficult for the regularization term to remedy all ill-posedness on its own. We further offload the ill-posedness by imposing a priori knowledge by introducing a random auxiliary variable $\mathbf{z} \sim \mathcal{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$. Importantly, $\mathbf{z}$ enables us to obtain several solutions for a heavily degraded $\mathbf{x}_t$ through the guidance provided by $\mathbf{z}$. This eventually helps to pull the noised observations back to the data distribution, allowing for a rich representation for the density.

**Implicit Prior Knowledge** For density estimation, the knowledge about the data distribution should be properly encoded in the prior term $g$ of (7). The explicit density function with determined marginal distribution is intractable, but it can be learned with the aid of well-developed generative models. That is, we can design $g$ by integrating various existing generative models. In this paper, we learn an implicit representation of the data density by adopting generative adversarial network (GAN), which has shown promising results in many generative tasks, and use generator loss as a relevant prior term $g$. For each forward step, our MAP objective for our generator in conjunction with the GAN prior is given by:

$$\mathcal{L}(G_\theta) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}, \mathbf{y}\sim\mathcal{N}(\mathbf{x},\sigma^2\mathbf{I}), \mathbf{z}\sim\mathcal{N}(0,I)} \left[ \frac{1}{2\sigma^2} \|G_\theta(\mathbf{y},\mathbf{z}) - \mathbf{y}\|_2^2 + \lambda g_\phi(G_\theta(\mathbf{y},\mathbf{z})) \right], \tag{8}$$

where the first term is the data fidelity term, $g_\phi(\mathbf{x}) = \log(1 - D_\phi(\mathbf{x})) - \log D_\phi(\mathbf{x})$ is a learnable prior term that is trained coupled with a discriminator $D_\phi$, and $\lambda \geq 0$ is a hyperparameter. As conventional, we train the discriminator $D_\phi$ to minimize Jensen-Shannon divergence. Contrary to conventional MAP whose prior term is pre-defined, our approach tries to learn the prior term by coordinating with the denoiser $G_\theta$ through the discriminator. This end-to-end training allows our discriminative learning method to deliver promising performance. Note that when the discriminator is optimal, i.e. $D_\phi(x) = \frac{p(x)}{p(x)+q_\theta(x)}$, the loss (8) with $\lambda = 1$ agrees with

$$\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}, \mathbf{y}\sim\mathcal{N}(\mathbf{x},\sigma^2\mathbf{I})} \left[ D_{KL}(q_\theta(\mathbf{x}|\mathbf{y}) || p(\mathbf{x}|\mathbf{y})) \right] + \mathcal{H}(q_\theta), \tag{9}$$

where $\mathcal{H}$ denotes an entropy, which corresponds to training the model to learn the posterior distribution. The overall training procedure combined with all $\sigma \in \{\sigma_k\}_{k=1}^T$ is provided in Appendix B.2.

**Small Denoising Steps** A major downside of DDMs is their slow sampling procedure, which requires hundreds to thousands of denoising steps to obtain a single image. By adopting MAP approach and parametrizing the prior distribution through GAN, our model provides an avenue to offload the time-consuming sampling scheme and enables significantly small denoising steps. For small degradation we can obtain a restored image in one shot. But, as our restoration starts from the Gaussian noise, the data distribution is not completely estimated. Therefore, we perform the generation iteratively. In our experiments on CIFAR10, we generate a high-quality sample in four denoising steps, whereas most DDMs use hundreds to thousands steps.

### 3.3 Extension to General Restoration

In Section 3.2, we proposed a denoising generative model based on MAP objective. However, from the IR perspective, it is not necessary to restrict it to denoising ($\mathbf{A} = \mathbf{I}$) and it can be generalized to any family of degradation matrices $\mathbf{A}$ and noise factors $\mathbf{\Sigma}$ in (1). Utilizing the general forward process, we can learn the generative model by generalizing the loss function (8) as follows:

$$\mathcal{L}(G_\theta) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}, \mathbf{y}\sim\mathcal{N}(\mathbf{A}\mathbf{x},\mathbf{\Sigma}), \mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I})} \left[ \frac{1}{2} \left\| \left(\mathbf{\Sigma}^\dagger\right)^{\frac{1}{2}} \left(\mathbf{A} \cdot G_\theta(\mathbf{y},\mathbf{z}) - \mathbf{y}\right) \right\|_2^2 + \lambda g_\phi(G_\theta(\mathbf{y},\mathbf{z})) \right]. \tag{10}$$

Therefore, RGM has an flexible structure that can permeate any forward process, and aids in designing a new generative model. Here, we propose a new model established upon super-resolution (SR).

**Multi-scale RGM** Most DDMs maintain the image size during the diffusion process by adding noise to individual pixels. Consequently, they are very inefficient because they require a latent as much as dimension of pixel space that is much larger than the submanifold of the image space. Motivated by

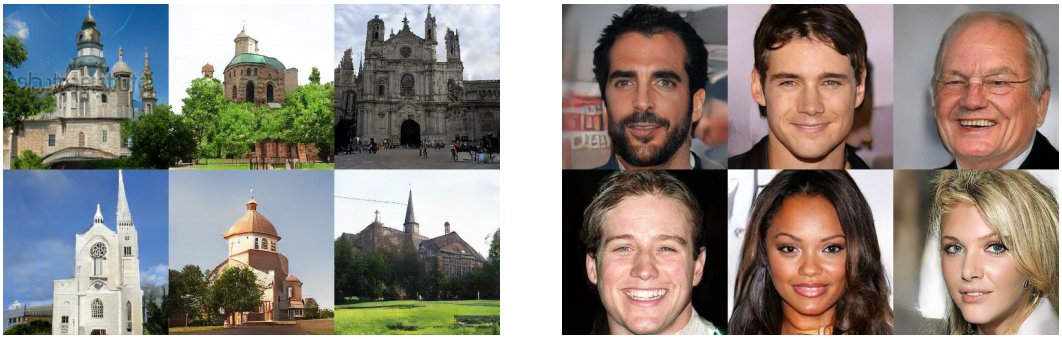

Figure 2: Generated samples on LSUN Church (left) and CelebA-HQ (right).

this, we take $\mathbf{A}$ as a block averaging filter that averages out $2 \times 2$ pixel values. Halving the image size at each coarsening step allows us a more expressive generative model with a lower-dimensional latent distribution. Moreover, multi-scale training has proven to be an effective strategy for synthesizing large scale images (Denton et al., 2015; Karras et al., 2017b; Reed et al., 2017). Therefore, our model produces strikingly realistic images by progressively extracting spatial information.

## 4 Experiments

This section evaluates the performance of the proposed RGMs on several benchmark datasets, including CIFAR10 (Krizhevsky et al., 2009) ($32 \times 32$ unconditional), CelebA-HQ (Liu et al., 2015) ($256 \times 256$), and LSUN Church (Yu et al., 2015) ($256 \times 256$). We also show the capability of RGMs for solving inverse problems. We parametrize our generator $G_\theta$ based on the UNet-like structure (Ronneberger et al., 2015) which was successfully used in NCSN++ architecture (Xiao et al., 2021a). The internal details of the implementation can be found in Appendix B.

**Setup** We implement two models: *RGM-D* is a model trained with the diffusion process, which is mainly used by DDMs. We also consider a multi-scale model whose degradation matrix is a $2 \times 2$ averaging filter. In this case, unlike the diffusion process, the image is corrupted by a downsampling filter together with additive noise. Therefore, the model (termed by *RGM-SR (naive)*) is demanded to conduct upsampling and denoising at the same time. To make it more effective, we explore another forward schedule that separates the downsampling and noising process and performs them alternatively. *RGM-SR* refers to the model to which this schedule is applied. (See Appendix B.1 for details).

### 4.1 2D Toy Example

We first employ a two-dimensional example to validate the effectiveness of prior knowledge of our MAP framework. We adopt a mixture of Gaussian with eight components (Grathwohl et al., 2018) as a target distribution. In Figure 3, we depict the benefits of our MAP approach over the MMSE approach. As illustrated in the top row, we diffuse the data distribution through four different noise levels. Each row from the second row to the bottom represents the learned distribution of our RGM, RGM without the auxiliary variable $\mathbf{z}$, and MMSE. First, the bottom row shows the failure of MMSE, where the modes of $p_{\text{data}}$ are connected and then missed. This tendency exacerbates as the noise level increases. Since the MMSE fails to reconstruct the data distribution even with a small rise in the noise level, the MMSE does not yield a satisfactory generative model with a small number of diffusion steps. Consequently, MMSE approaches, such as DDMs, require a large number of steps to stably recover the data distribution. On the other hand, by leveraging the prior term, our model generates samples

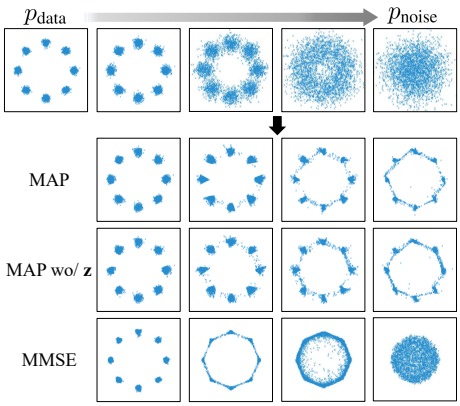


Figure 3: Comparison of recovering density by MMSE versus MAP.

Table 1: Results on unconditional generation of CIFAR10.

| Class | Model | FID (↓) | IS (↑) | NFE (↓) |
|---|---|---|---|---|
| **RGM** | RGM-D | 3.08 | 9.14 | 4 |
| | RGM-SR | 2.47 | 9.68 | 7 |
| **DDM** | DDPM (Ho et al., 2020) | 3.21 | 9.46 | 1000 |
| | NCSN (Song & Ermon, 2019) | 25.3 | 8.87 | 1000 |
| | Score SDE (VE) (Song et al., 2020b) | 2.20 | 9.89 | 2000 |
| | Score SDE (VP) (Song et al., 2020b) | 2.41 | 9.68 | 2000 |
| | Probability Flow (VP) (Song et al., 2020b) | 3.08 | 9.83 | 140 |
| | LSGM (Vahdat et al., 2021) | 2.10 | 9.87 | 147 |
| | DDIM (50 steps) (Song et al., 2020a) | 4.67 | 8.78 | 50 |
| | FastDDPM (T=50) (Kong & Ping, 2021) | 3.41 | 8.98 | 50 |
| | Recovery EBM (Gao et al., 2020) | 9.58 | 8.30 | 180 |
| | VDM (Kingma et al., 2021) | 4.00 | - | 1000 |
| | UDM (Kim et al., 2021) | 2.33 | 10.1 | 2000 |
| | Gotta Go Fast (Alexia Jolicoeur-Martineau, 2021) | 2.44 | - | 1000 |
| | Subspace Diffusion (Jing et al., 2022) | 2.17 | 9.94 | $\geq$ 1000 |
| | CLD (Dockhorn et al., 2021) | 2.25 | - | 2000 |
| | DEIS (Zhang & Chen, 2022) | 3.37 | 9.74 | 15 |
| | DDGAN (Xiao et al., 2021a) | 3.75 | 9.63 | 4 |
| | StyleGAN2+ES-DDPM (Lyu et al., 2022) | 5.52 | - | 101 |
| | Progressive Distillation (Salimans & Ho, 2022) | 3.00 | - | 4 |
| **GAN** | SNGAN+DGflow (Ansari et al., 2020) | 9.62 | 9.35 | 25 |
| | AutoGAN (Gong et al., 2019) | 12.4 | 8.60 | 1 |
| | TransGAN (Jiang et al., 2021) | 9.26 | 9.02 | 1 |
| | StyleGAN2 w/o ADA (Karras et al., 2020) | 8.32 | 9.18 | 1 |
| | StyleGAN2 w/ ADA (Karras et al., 2020) | 2.92 | 9.83 | 1 |
| **VAE** | NVAE (Vahdat & Kautz, 2020) | 23.5 | 7.18 | 1 |
| | Glow (Kingma & Dhariwal, 2018) | 48.9 | 3.92 | 1 |
| | PixelCNN (Van Oord et al., 2016) | 65.9 | 4.60 | 1024 |
| | VAEBM (Xiao et al., 2020) | 12.2 | 8.43 | 16 |

from the multimodal distribution significantly better. By imposing the prior knowledge, the MAP objective accurately estimates the density, which allows distribution recovery with a much smaller number of forward processes than the MMSE approach. Furthermore, we can observe the effect of the auxiliary variable $\mathbf{z}$ by comparing the second and third rows in Figure 3. MAP with $\mathbf{z}$ has higher sample quality. The effect of $\mathbf{z}$ is further amplified for more complex distributions, such as image data (See Figure 5 and Table 3). This synthetic experiment validates that our MAP approach in conjunction with the random auxiliary variable $\mathbf{z}$ enables accurate and efficient generative modeling.

### 4.2 Image Generation

We compare the performance of our method with several existing baselines. For quantitative comparison, we use Fréchet Inception Distance (FID) and Inception Score (IS) as the evaluation metrics. We also report the number of network function evaluations (NFE). For DDMs and RGMs, NFE value and real inference time are proportional. Following Song et al. (2020b); Dockhorn et al. (2021), we focus on the widely used CIFAR10 unconditional image generation benchmark and also validate the performance of RGMs on large-scale CelebA-HQ-256 images. Table 1 summarizes the quantitative evaluations on CIFAR10, and the results on CelebA-HQ is reported in Table 2. Qualitative performance is depicted in Figures 2 and 4.

**Results** We can see that our model achieves the state-of-the-art FID score on CelebA-HQ-256 among restoraion-based models. On CIFAR10, our RGMs are comparable to the best existing DDMs and GAN models. Although the best denoising models obtain better results than ours on CIFAR10, they use a much larger number of denoising steps (e.g. ScoreSDE with VESDE requires 2000 steps). Progressive distillation achieves an FID score comparable to ours with quite reduced sampling cost, but they require considerable additional cost at training time due to their distillation process. Notably, our RGM-SR achieves FID 2.47 and IS 9.68 with only seven steps, which is state-of-the-art sampling FID performance when NFE is limited. The overall results confirm that our MAP estimation immediately eliminates the need for an expensive sampling scheme while still maintaining the density estimating capability of DDMs. Interestingly, RGM-SR outperforms RGM-D by a large margin even with far fewer latent variables than RGM-D. This improved performance may be attributed to the increase in NFE; however, the FID of RGM-D with $T = 8$ reported in Table 7 confirms that it is not.

Table 2: Results on generation of CelebA-HQ-256.

| Class | Model | FID (↓) | NFE (↓) |
|---|---|---|---|
| **RGM** | RGM-D | 7.15 | 4 |
| **DDM** | Score SDE (VP) (Song et al., 2020b) | 7.23 | 4000 |
| | Probability Flow (Song et al., 2020b) | 128.13 | 335 |
| | LSGM (Vahdat et al., 2021) | 7.22 | 23 |
| | UDM (Kim et al., 2021) | 7.16 | 2000 |
| | DDGAN (Xiao et al., 2021a) | 7.64 | 4 |
| **GAN** | PGGAN (Karras et al., 2017a) | 8.03 | 1 |
| | Adv. LAE (Pidhorskyi et al., 2020) | 19.2 | 1 |
| | VQ-GAN (Esser et al., 2021) | 10.2 | 1 |
| | DC-AE (Parmar et al., 2021) | 15.8 | 1 |
| | StyleSwin (Zhang et al., 2022) | 3.25 | 1 |
| **VAE** | NVAE (Vahdat & Kautz, 2020) | 29.7 | 1 |
| | VAEBM (Xiao et al., 2020) | 20.4 | 1 |
| | NCP-VAE (Aneja et al., 2021) | 24.8 | 1 |

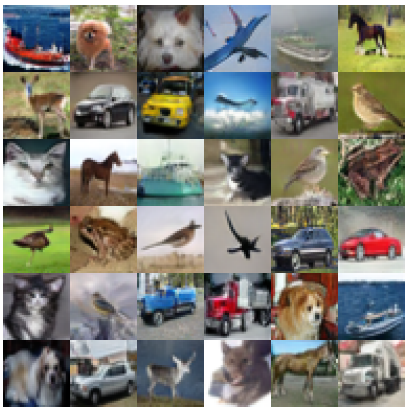

Figure 4: CIFAR10 generated samples.

The overall results indicate that the MAP estimation of RGMs is a promising way for generating high-quality samples in limited steps. More uncurated images can be founded in Appendix C.6.

### 4.3 Ablation Studies

This section is devoted to ablation analyses which show that all parts of our MAP objective, the data fidelity term, and the prior term together coupled with the auxiliary variable, and the regularization parameter, each play an important role in our performance of density estimation.

**On the role of $\mathbf{z}$** We include experimental results on LSUN, which demonstrate how the auxilliary variable $\mathbf{z}$ alleviates the ill-posedness of the inverse problem. By noising the upper-left image $\mathbf{x}_0$, we obtain the forward trajectory $\{\mathbf{x}_k\}_{k=1}^4$. The figures on the right are restored images of $\mathbf{x}_k$ by RGM-D together with four different $\mathbf{z}$. We can see that a reconstruction is almost unique when the noise level is small. But, as the noise level increases, a single $\mathbf{x}_k$ has various reconstructions. It is evident that assigning $\mathbf{z}$ helps generate different denoised images from a heavily degraded $\mathbf{x}_k$ through the guidance provided by $\mathbf{z}$. However, one might think that the ill-posedness is detoured by multi-step training using multiple $\sigma_k$ rather than through $\mathbf{z}$. This claim can be refuted using the result of RGM-D without $\mathbf{z}$ reported in Table 3. We observe the significant difference in FIDs of RGM-D with and without $\mathbf{z}$ under the same number of denoising steps, which indicates the effectiveness of $\mathbf{z}$.

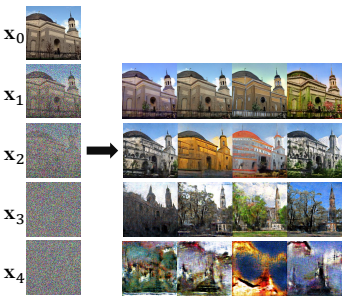


Figure 5: Study on effect of $\mathbf{z}$.

**On the effect of Varying $\lambda$** We investigate the sensitivity of the regularization parameter $\lambda$ in (8). Since it controls the relative importance between the fidelity term and the prior term, $\lambda$ is a trade-off hyperparameter that determines how much regularizes the joint distribution of $p_k$ and $p_{k+1}$. In Figure 6, we present FID scores measured on CIFAR10 with the same number of degradation steps ($T = 4$) and varying $\lambda$. We can see that our models are quite robust with respect to $\lambda$. An empirically observed sweet spot of $\lambda$ is $d/10 \leq \lambda \leq d$ for the image size $d$, in which FID is no longer improved outside this threshold. For small $\lambda$, the models put a lot of effort to recover the degradation, which hinders estimating data distribution. Choosing a large $\lambda$ also results in performance degeneration. There is another point that draws our attention. When $\lambda = \infty$, that is, when there is no fidelity term on the objective, FID scores completely deteriorate. In this case, the models are trained by the vanilla GAN loss. From the perspective of GAN, our MAP objective adds the fidelity term to the GAN loss function. We further observe that with the help of the fidelity term, our model enhances the mode-collapsing resiliency of GAN (See Figure 16). These results validate that the performance of RGMs owes to both fidelity and prior term, and the reliable regularization parameter is determined to balance these two terms.

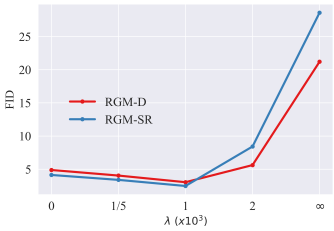


Figure 6: Study on effect of $\lambda$.

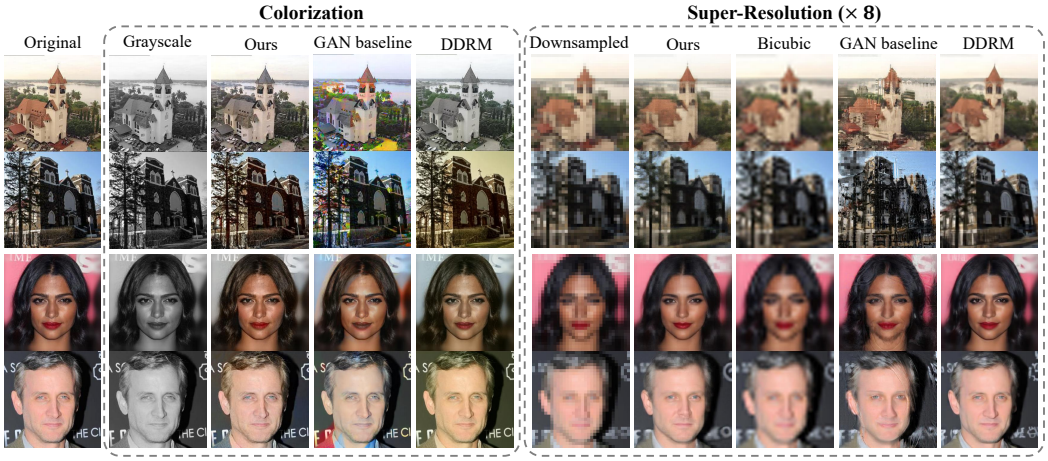


Figure 7: Colorization (left) and super-resolution (right) results on LSUN and CelebA-HQ datasets.

**On the forward process schedule** Since the forward process determines the way of connecting the data and latent distributions, it significantly affects the performance of models. The first important factor is the number of forward steps $T$, which is directly related to NFE. In Table 3, we ablate the effect of $T$. When $T = 1$, it may be difficult for the model to directly approximate the data distribution from the Gaussian noise. This is reflected in the poor FID score. We also study the forward process schedule of the SR model. We can observe that the separation of the same forward process into two steps makes the model easier to learn, and this brings the performance enhancement of RGM-SR compared to RGM-SR (naive).

Table 3: Ablation Studies.

| Model | FID (↓) |
|---|---|
| RGM-D ($\lambda = \infty$) | 32.5 |
| RGM-D ($T = 1$) | 14.6 |
| RGM-D ($T = 4$) | 3.04 |
| RGM-D (wo / $\mathbf{z}$) | 3.87 |
| RGM-SR (naive) | 3.17 |
| RGM-SR | 2.47 |

### 4.4 Inverse Problems

While our model was originally devised to generate images, we further show the applicability of RGMs to inverse problems. Recently, a promising approach in imaging inverse problems is to leverage a learned denoiser as an alternative to the proximal operator of splitting algorithms (Romano et al., 2017; Hurault et al., 2021). Such methodology is referred to as Plug-and-Play (PnP) algorithms (Venkatakrishnan et al., 2013). In a similar spirit, we utilize the trained RGMs as a modular part of the PnP algorithms to solve various inverse problems. In this section, we testify our RGM-D for two inverse problems; super-resolution and colorization, by plugging our model into Douglas-Rachford Splitting algorithm (Lions & Mercier, 1979). Details can be found in Appendix B.3.

**Results** We compare the performance of our model with current-leading models: We compare our model with DDRM (Kawar et al., 2022), which solves inverse problems with a pre-trained DDPM by a posterior sampling scheme. As a GAN baseline, we adopt StyleSwin (Zhang et al., 2022) and reconstruct the image by optimizing over the latent vector (Pan et al., 2021). We also consider bicubic interpolation as a baseline for super-resolution. We observe that our model is capable of reconstructing faithful and realistic images, as evident in Figure 7. Compared with baselines, our model produces high-quality reconstructions across all the datasets. In particular, our model shows promising performance for colorization. These results show the applicability of RGMs to PnP prior, and this will bring a range of potential applications, including image segmentation, conditional generation, and other imaging inverse problems. Additional quantitative and qualitative results are provided in Appendix C.4.

**Comparison of RGM-D and RGM-SR** We investigate the effect of the degradation process used during training on the performance of solving inverse problems. We compare the reconstruction performance of RGM-D and RGM-SR that are trained on different degradation processes by applying both models to denoising and super-resolution (SR) tasks on CIFAR10. Quantitative results are presented in Table 4. We can see that the RGM-SR that is trained based on SR actually performs the SR task better. Also, we can observe a similar tendency for denoising. The results confirm that the degradation process used in training actually helps in solving the corresponding inverse problem.

Table 4: Quantitative comparison of RGM-D and RGM-SR on image reconstruction.

| Model | Super-Resolution | | | | Denoising | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ($\times 2$) | | ($\times 4$) | | ($\sigma = 10/255$) | | ($\sigma = 20/255$) | | ($\sigma = 40/255$) | |
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| RGM-D | 26.63 | 0.88 | 20.84 | 0.58 | **30.11** | **0.93** | **26.57** | **0.86** | **24.23** | **0.80** |
| RGM-SR | **27.42** | **0.90** | **21.14** | **0.59** | 29.41 | 0.92 | 25.87 | 0.84 | 23.53 | 0.77 |

## 5 Related Work

In recent years, denoising-based generative models (Ho et al., 2020; Song & Ermon, 2019; Song et al., 2020b) have emerged as a class of density estimation models, first sparked by (Sohl-Dickstein et al., 2015). They define a sampling process as the reverse of a forward diffusion process that maps data to Gaussian noise by consecutively adding a small portion of the noise to the input data. As they faithfully estimate the data distribution and generate high-fidelity samples, they have rapidly been applied to various domains such as conditional generation (Lee et al., 2022; Ho et al., 2022a), audio synthesis(Kong et al., 2021; Popov et al., 2021), medical imaging (Song et al., 2021; Chung & Ye, 2022), video generation (Ho et al., 2022b; Yang et al., 2022), and 3D point cloud generation (Lyu et al., 2021). Their major drawback is slow and expensive inference. Many studies have been dedicated to circumventing this downside by developing a fast numerical solver (Jolicoeur-Martineau et al., 2021; Zhang & Chen, 2022; Tachibana et al., 2021; Liu et al., 2022) or using an alternative noising process such as non-Markovian (Song et al., 2020a), a second-order Langevin dynamics (Dockhorn et al., 2021), and non-linear diffusion processes (De Bortoli et al., 2021; Chen et al., 2021b). Another line of work improves sampling efficiency by incorporating it into other generative models, including GAN (Xiao et al., 2021a; Lyu et al., 2022), and VAE (Vahdat & Kautz, 2020). Xiao et al. (2021a) which enjoys small sampling steps by using GAN is one of our related work. However, they did not introduce GAN from a MAP perspective and our model requires less training iteration to achieve the same performance. (See Appendix A for details.) Moreover, there have been distillation approaches (Salimans & Ho, 2022; Meng et al., 2022). On a side note, all the aforementioned models use the Gaussian noising process as the forward process.

Recently, the literature has begun to replace the additive Gaussian noising process with other transforms. Breaking away from the diffusion process, (Rissanen et al., 2022) proposed a forward blurring process inspired by heat dissipation. They suggest a new generation process, but they specialize in the proposed blurring process and cannot be incompatible with other degradation processes. Possibly the closest study to our work is Cold Diffusion (Bansal et al., 2022) which generalizes the diffusion process to arbitrary image transformations. It seems to use a general transform similar to our models, but Cold Diffusion only uses deterministic degradation processes by entirely removing additive Gaussian noise, which hinders its density estimation performance. Also, they use the MMSE objective, still requiring an array of several forward steps. We include a comparison with these related works in Appendix C.3.

## 6 Conclusion and Future Work

In this study, we presented a general framework for modeling efficient generative models through the lens of IR. Compared to DDMs whose both forward and reverse processes are fixed to thousands of Gaussian steps, our approach provides more flexible models that eliminate expensive sampling and can enjoy versatile forward processes. We eliminated the usage of slow sampling by taking on MAP estimation and incorporating implicit prior information through GAN. In addition,we propose a multi-scale method as an example of the usability of various forward processes. The experimental results showed that the image quality obtained was on par with the leading DDMs, and we achieved state-or-the-art performance using a limited number of forward steps. We hope that this work provides a broad view of modeling useful generative models.

Our model has two degrees of freedom: One is how to parametrize the prior knowledge, and the other is the choice of the forward process. This opens up interesting directions for future research. In addition to the GAN we used, the prior term can be constructed in different ways. It could also be interesting to explore other degradation transformations. Moreover, analyzing the effect of the random auxiliary variable would be an worthwhile direction for future research. Future work could include the comprehensive design of a convergence guaranteed PnP algorithm for application to various inverse problems. We leave these further extensions to future work.

## REFERENCES


Rémi Piché-Taillefer Tal Kachman Ioannis Mitliagkas Alexia Jolicoeur-Martineau, Ke Li. Gotta go fast when generating data with score-based models. *arXiv preprint arXiv:2105.14080*, 2021.

Jyoti Aneja, Alex Schwing, Jan Kautz, and Arash Vahdat. A contrastive learning approach for training variational autoencoder priors. *Advances in neural information processing systems*, 34:480–493, 2021.

Abdul Fatir Ansari, Ming Liang Ang, and Harold Soh. Refining deep generative models via discriminator gradient flow. *arXiv preprint arXiv:2012.00780*, 2020.

Mark R Banham and Aggelos K Katsaggelos. Digital image restoration. *IEEE signal processing magazine*, 14(2):24–41, 1997.

Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie S Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. *arXiv preprint arXiv:2208.09392*, 2022.

Richard G Baraniuk. Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4): 118–121, 2007.

Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1):1–20, 1991.

Jan Biemond, Reginald L Lagendijk, and Russell M Mersereau. Iterative methods for image deblurring. *Proceedings of the IEEE*, 78(5):856–883, 1990.

Siavash Bigdeli, David Honzátko, Sabine Süsstrunk, and L Andrea Dunbar. Image restoration using plug-and-play cnn map denoisers. *arXiv preprint arXiv:1912.09299*, 2019.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL `https://openreview.net/forum?id=B1xsqj09Fm`.

Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pp. 60–65. Ieee, 2005.

Kenneth R Castleman. *Digital image processing*. Prentice Hall Press, 1996.

Tianrong Chen, Guan-Horng Liu, and Evangelos A Theodorou. Likelihood training of schr\" odinger bridge using forward-backward sdes theory. *arXiv preprint arXiv:2110.11291*, 2021a.

Tianrong Chen, Guan-Horng Liu, and Evangelos A Theodorou. Likelihood training of schrodinger bridge using forward-backward sdes theory. *arXiv preprint arXiv:2110.11291*, 2021b.

Yunjin Chen. Higher-order mrfs based image super resolution: why not map? *IET Image Processing*, 10(4):297–303, 2016.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical Image Analysis*, pp. 102479, 2022.

Giannis Daras, Mauricio Delbracio, Hossein Talebi, Alexandros G Dimakis, and Peyman Milanfar. Soft diffusion: Score matching for general corruptions. *arXiv preprint arXiv:2209.05442*, 2022.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021.

Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. *Advances in neural information processing systems*, 28, 2015.

Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped langevin diffusion. *arXiv preprint arXiv:2112.07068*, 2021.

David L Donoho. De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3): 613–627, 1995.

Gintare Karolina Dziugaite, Daniel M Roy, and Zoubin Ghahramani. Training generative neural networks via maximum mean discrepancy optimization. *arXiv preprint arXiv:1505.03906*, 2015.

Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12873–12883, 2021.

Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multiframe super resolution. *IEEE transactions on image processing*, 13(10):1327–1344, 2004.

Ruiqi Gao, Yang Song, Ben Poole, Ying Nian Wu, and Diederik P Kingma. Learning energy-based models by diffusion recovery likelihood. *arXiv preprint arXiv:2012.08125*, 2020.

Donald Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE transactions on Image Processing*, 4(7):932–946, 1995.

Xinyu Gong, Shiyu Chang, Yifan Jiang, and Zhangyang Wang. Autogan: Neural architecture search for generative adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3224–3234, 2019.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

Bahadir Gunturk and Xin Li. *Image restoration*. CRC Press, 2018.

Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pp. 49–52, 1902.

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

Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *J. Mach. Learn. Res.*, 23:47–1, 2022a.

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *arXiv preprint arXiv:2204.03458*, 2022b.

Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. *arXiv preprint arXiv:2209.05557*, 2022.

Bobby R. Hunt. Bayesian methods in nonlinear digital image restoration. *IEEE Transactions on Computers*, 26(03):219–229, 1977.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. *arXiv preprint arXiv:2110.03220*, 2021.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. *arXiv preprint arXiv:2201.13256*, 2022.

Yifan Jiang, Shiyu Chang, and Zhangyang Wang. Transgan: Two transformers can make one strong gan. *arXiv preprint arXiv:2102.07074*, 1(3), 2021.

Bowen Jing, Gabriele Corso, Renato Berlinghieri, and Tommi Jaakkola. Subspace diffusion generative models. *arXiv preprint arXiv:2205.01490*, 2022.

Alexia Jolicoeur-Martineau, Ke Li, Rémi Piché-Taillefer, Tal Kachman, and Ioannis Mitliagkas. Gotta go fast when generating data with score-based models. *arXiv preprint arXiv:2105.14080*, 2021.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017a.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017b.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems*, 33:12104–12114, 2020.

Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022.

Dongjun Kim, Seungjae Shin, Kyungwoo Song, Wanmo Kang, and Il-Chul Moon. Score matching model for unbounded data score. *arXiv preprint arXiv:2106.05527*, 2021.

Dongjun Kim, Seungjae Shin, Kyungwoo Song, Wanmo Kang, and Il-Chul Moon. Soft truncation: A universal training technique of score-based diffusion model for high precision score estimation. In *International Conference on Machine Learning*, pp. 11201–11228. PMLR, 2022.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

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

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *International conference on machine learning*, 2014.

Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018.

Zhifeng Kong and Wei Ping. On fast sampling of diffusion probabilistic models. *arXiv preprint arXiv:2106.00132*, 2021.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *International Conference on Learning Representations*, 2021.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using plug & play priors: when langevin meets tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737, 2022.

Sang-gil Lee, Heeseung Kim, Chaehun Shin, Xu Tan, Chang Liu, Qi Meng, Tao Qin, Wei Chen, Sungroh Yoon, and Tie-Yan Liu. Priorgrad: Improving conditional denoising diffusion models with data-driven adaptive prior. *International Conference on Learning Representations*, 2022.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

Chun-Liang Li, Wei-Cheng Chang, Yu Cheng, Yiming Yang, and Barnabás Póczos. Mmd gan: Towards deeper understanding of moment matching network. *Advances in neural information processing systems*, 30, 2017.

Yujia Li, Kevin Swersky, and Rich Zemel. Generative moment matching networks. In *International conference on machine learning*, pp. 1718–1727. PMLR, 2015.

Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. *arXiv preprint arXiv:2202.09778*, 2022.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pp. 3730–3738, 2015.

Sebastian Lunz, Ozan Öktem, and Carola-Bibiane Schönlieb. Adversarial regularizers in inverse problems. *Advances in neural information processing systems*, 31, 2018.

Zhaoyang Lyu, Zhifeng Kong, Xudong Xu, Liang Pan, and Dahua Lin. A conditional point diffusion-refinement paradigm for 3d point cloud completion. *arXiv preprint arXiv:2112.03530*, 2021.

Zhaoyang Lyu, Xudong Xu, Ceyuan Yang, Dahua Lin, and Bo Dai. Accelerating diffusion models via early stop of the diffusion process. *arXiv preprint arXiv:2205.12524*, 2022.

Jianhua Ma, Jing Huang, Qianjin Feng, Hua Zhang, Hongbing Lu, Zhengrong Liang, and Wufan Chen. Low-dose computed tomography image restoration using previous normal-dose scan. *Medical physics*, 38(10):5713–5731, 2011.

Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

Chenlin Meng, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

Chenlin Meng, Ruiqi Gao, Diederik P Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. *arXiv preprint arXiv:2210.03142*, 2022.

Eliakim H Moore. On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, 26: 394–395, 1920.

Radford M Neal. *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto Toronto, ON, Canada, 1993.

Khai Nguyen, Nhat Ho, Tung Pham, and Hung Bui. Distributional sliced-wasserstein and applications to generative modeling. *arXiv preprint arXiv:2002.07367*, 2020.

Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Kushagra Pandey, Avideep Mukherjee, Piyush Rai, and Abhishek Kumar. Diffusevae: Efficient, controllable and high-fidelity generation from low-dimensional latents. *arXiv preprint arXiv:2201.00308*, 2022.

Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and trends® in Optimization*, 1 (3):127–239, 2014.

Gaurav Parmar, Dacheng Li, Kwonjoon Lee, and Zhuowen Tu. Dual contradistinctive generative autoencoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 823–832, 2021.

Stanislav Pidhorskyi, Donald A Adjeroh, and Gianfranco Doretto. Adversarial latent autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14104–14113, 2020.

Vadim Popov, Ivan Vovk, Vladimir Gogoryan, Tasnima Sadekova, and Mikhail Kudinov. Grad-tts: A diffusion probabilistic model for text-to-speech. In *International Conference on Machine Learning*, pp. 8599–8608. PMLR, 2021.

Scott Reed, Aäron Oord, Nal Kalchbrenner, Sergio Gómez Colmenarejo, Ziyu Wang, Yutian Chen, Dan Belov, and Nando Freitas. Parallel multiscale autoregressive density estimation. In *International conference on machine learning*, pp. 2912–2921. PMLR, 2017.

Edward T Reehorst and Philip Schniter. Regularization by denoising: Clarifications and new interpretations. *IEEE transactions on computational imaging*, 5(1):52–67, 2018.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pp. 1278–1286. PMLR, 2014.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. *arXiv preprint arXiv:2206.13397*, 2022.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

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*, pp. 234–241. Springer, 2015.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Saikat Saha, Yvo Boers, Hans Driessen, Pranab Kumar Mandal, and Arunabha Bagchi. Particle based map state estimation: A comparison. In *2009 12th International Conference on Information Fusion*, pp. 278–283. IEEE, 2009.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

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*, 2020b.

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.

Hideyuki Tachibana, Mocho Go, Muneyoshi Inahara, Yotaro Katayama, and Yotaro Watanabe. It\^{o}-taylor sampling scheme for denoising diffusion probabilistic models using ideal derivatives. *arXiv preprint arXiv:2112.13339*, 2021.

Andrei Nikolaevich Tikhonov. On the regularization of ill-posed problems. In *Doklady Akademii Nauk*, volume 153, pp. 49–52. Russian Academy of Sciences, 1963.

H Trussell. The relationship between image restoration by the maximum a posteriori method and a maximum entropy method. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28 (1):114–117, 1980.

Arash Vahdat and Jan Kautz. Nvae: A deep hierarchical variational autoencoder. *Advances in Neural Information Processing Systems*, 33:19667–19679, 2020.

Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. *Advances in Neural Information Processing Systems*, 34:11287–11302, 2021.

Aaron Van Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *International conference on machine learning*, pp. 1747–1756. PMLR, 2016.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948. IEEE, 2013.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Zhisheng Xiao, Karsten Kreis, Jan Kautz, and Arash Vahdat. Vaebm: A symbiosis between variational autoencoders and energy-based models. *arXiv preprint arXiv:2010.00654*, 2020.

Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. *arXiv preprint arXiv:2112.07804*, 2021a.

Zhisheng Xiao, Qing Yan, and Yali Amit. Ebms trained with maximum likelihood are generator models trained with a self-adverserial loss. *arXiv preprint arXiv:2102.11757*, 2021b.

Ruihan Yang, Prakhar Srivastava, and Stephan Mandt. Diffusion probabilistic modeling for video generation. *arXiv preprint arXiv:2203.09481*, 2022.

Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.

ME Zervakis and AN Venetsanopoulos. A class of noniterative estimators for nonlinear image restoration. *IEEE transactions on circuits and systems*, 38(7):731–744, 1991.

Bowen Zhang, Shuyang Gu, Bo Zhang, Jianmin Bao, Dong Chen, Fang Wen, Yong Wang, and Baining Guo. Styleswin: Transformer-based gan for high-resolution image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11304–11314, 2022.

Qinsheng Zhang and Yongxin Chen. Diffusion normalizing flow. *Advances in Neural Information Processing Systems*, 34:16280–16291, 2021.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*, 2022.