Removing Structured Noise using Diffusion Models

Tristan S.W. Stevens, *Student, IEEE*, Hans van Gorp, *Student, IEEE*, Faik C. Meral, Junseob Shin,
Jason Yu, Jean-Luc Robert and Ruud J.G. van Sloun, *Member, IEEE*

***Abstract*—Solving ill-posed inverse problems requires careful formulation of prior beliefs over the signals of interest and an accurate description of their manifestation into noisy measurements. Handcrafted signal priors based on e.g. sparsity are increasingly replaced by data-driven deep generative models, and several groups have recently shown that state-of-the-art score-based diffusion models yield particularly strong performance and flexibility. In this paper, we show that the powerful paradigm of posterior sampling with diffusion models can be extended to include rich, structured, noise models. To that end, we propose a joint conditional reverse diffusion process with learned scores for the noise and signal-generating distribution. We demonstrate strong performance gains across various inverse problems with structured noise, outperforming competitive baselines that use normalizing flows and adversarial networks. This opens up new opportunities and relevant practical applications of diffusion modeling for inverse problems in the context of non-Gaussian measurement models.**[1]

***Index Terms*—Diffusion models, posterior sampling, structured noise, inverse problems**

I. Introduction

MANY signal and image processing problems, such as denoising, compressed sensing, or phase retrieval, can be formulated as inverse problems that aim to recover unknown signals from (noisy) observations. These ill-posed problems are, by definition, subject to many solutions under the given measurement model. Therefore, prior knowledge is required for a meaningful and physically plausible recovery of the original signal. Bayesian inference and maximum a posteriori (MAP) solutions incorporate both signal priors and observation likelihood models. Choosing an appropriate statistical prior is not trivial and dependent on both the application as well as the recovery task.

Before deep learning, sparsity in some transformed domain has been the go-to prior in compressed sensing (CS) methods [1], such as iterative thresholding [2] or wavelet decomposition [3]. At present, deep generative modeling has established itself as a strong mechanism for learning such priors for inverse problem-solving. Both generative adversarial networks (GANs) [4] and normalizing flows (NFs) [5], [6] have been applied as natural signal priors for inverse problems in image recovery. These data-driven methods are more powerful compared to classical methods, as they can accurately learn the natural signal manifold and do not rely on assumptions such as signal sparsity or hand-crafted basis functions.

Recently, diffusion models have shown impressive results for both conditional and unconditional image generation and can be easily fitted to a target data distribution using score matching [7], [8]. These deep generative models learn the score of the data manifold and produce samples by reverting a diffusion process, guiding noise samples toward the target distribution. Diffusion models have achieved state-of-the-art performance in many downstream tasks and applications, ranging from state-of-the-art text-to-image models such as Stable Diffusion [9] to medical imaging [10]–[12]. Furthermore, understanding of diffusion models is rapidly improving and progress in the field is extremely fast-paced [13]–[17].

The iterative nature of the sampling procedure used by diffusion models renders inference slow compared to GANs and VAEs. However, many recent efforts have shown ways to significantly improve the sampling speed by accelerating the diffusion process. Inspired by momentum methods in sampling, [18] introduces a momentum sampler for diffusion models, which leads to increased sample quality with fewer function evaluations. [19] offers a new sampling strategy, namely Come-Closer-Diffuse-Faster (CCDF), which leverages the conditional quality of inverse problems. The reverse diffusion can be initialized from the observation instead of a sample from the base distribution, which leads to faster convergence for conditional sampling. [20] proposes a progressive distillation method that augments the training of the diffusion models with a student-teacher model setup. In doing this, they were able to drastically reduce the number of sampling steps. Lastly, many methods aim to execute the diffusion process in a reduced space to accelerate the diffusion process. While [21] restricts diffusion through projections onto subspaces, [22] and [9] run the diffusion in the latent space.

Despite this promise, current score-based diffusion methods for inverse problems are limited to measurement models with unstructured noise. In many image processing tasks, corruptions are however highly structured and spatially correlated. Relevant examples include interference, speckle, or haze. Nevertheless, current conditional diffusion models naively assume that the noise follows some basic tractable distribution (e.g. Gaussian or Poisson) [12], [23], [24]. Beyond the realm of diffusion models, [25] extended normalizing flow (NF)-based inference to structured noise applications. However, compared to diffusion models, NFs require specialized network architectures, which are computationally and memory expensive.

Given the promising outlook of diffusion models, we propose to learn score models for both the noise and the desired signal and perform joint inference of both quantities, coupled via

Manuscript submitted for review on July 31, 2023; revised August 16, 2023. This work was performed within the IMPULSE framework of the Eindhoven MedTech Innovation Center (e/MTIC, incorporating Eindhoven University of Technology and Philips Research), including a PPS supplement from the Dutch Ministry of Economic Affairs and Climate Policy.

Tristan S.W. Stevens (e-mail: t.s.w.stevens@tue.nl), Hans van Gorp and Ruud J.G. van Sloun are with the Electrical Engineering Department, Eindhoven University of Technology, the Netherlands. Faik C. Meral, Jun Seob Shin, Jason Yu, and Jean-Luc Robert are with Philips Research, Cambridge MA, USA.

[1] Code: https://github.com/tristan-deep/joint-diffusion

the observation model. The resulting sampling scheme enables solving a wide variety of inverse problems with structured noise.

The main contributions of this work are as follows:

- We propose a novel joint conditional approximate posterior sampling method to efficiently remove structured noise using diffusion models. Our formulation is compatible with many existing iterative sampling methods for score-based generative models.
- We show strong performance gains across various challenging inverse problems involving structured noise compared to competitive state-of-the-art methods based on NFs and GANs.
- We provide derivations for and comparison of three recent posterior sampling frameworks for diffusion models (ΠGDM, DPS, projection) as the backbone for our joint inference scheme.
- We demonstrate improved robustness on a range of out-of-distribution signals and noise compared to baselines.

II. Problem Statement

Many image reconstruction tasks can be formulated as an inverse problem with the basic form

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

where $\mathbf{y} \in \mathbb{R}^m$ is the noisy observation, $\mathbf{x} \in \mathbb{R}^d$ the desired signal or image, and $\mathbf{n} \in \mathbb{R}^m$ the additive noise. The linear forward operator $\mathbf{A} \in \mathbb{R}^{m \times d}$ captures the deterministic transformation of $\mathbf{x}$.

Maximum a posteriori (MAP) inference is typically used to find an optimal solution $\hat{\mathbf{x}}_{\text{MAP}}$ that maximizes posterior density $p_{X|Y}(\mathbf{x}|\mathbf{y})$:

$$\begin{aligned}\hat{\mathbf{x}}_{\text{MAP}} &= \arg\max_{\mathbf{x}} \log p_{X|Y}(\mathbf{x}|\mathbf{y}) \\ &= \arg\max_{\mathbf{x}} \left[\log p_{Y|X}(\mathbf{y}|\mathbf{x}) + \log p_X(\mathbf{x})\right], \end{aligned} \tag{2}$$

where $p_{Y|X}(\mathbf{y}|\mathbf{x})$ is the likelihood according to the measurement model and $\log p_X(\mathbf{x})$ the signal prior. Assumptions on the stochastic corruption process $\mathbf{n}$ are of key importance too, in particular for applications for which this process is highly structured. However, most methods assume i.i.d. Gaussian distributed noise, such that the forward model becomes $p_{Y|X}(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \sigma_N^2\mathbf{I})$. This naturally leads to the following simplified problem:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma_N^2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 - \log p_X(\mathbf{x}). \tag{3}$$

However, this naive assumption can be very restrictive as many noise processes are much more structured and complex. A myriad of problems can be addressed under the formulation of (1), given the freedom of choice for the noise source $\mathbf{n}$. Therefore, in this work, our aim is to solve a more broad class of inverse problems defined by any arbitrary noise distribution $\mathbf{n} \sim p_N(\mathbf{n}) \neq \mathcal{N}$ and signal prior $\mathbf{x} \sim p_X(\mathbf{x})$, resulting in the following, more general, MAP estimator:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max_{\mathbf{x}} \log p_N(\mathbf{y} - \mathbf{A}\mathbf{x}) - \log p_X(\mathbf{x}). \tag{4}$$

In this paper, we propose to solve this class of problems using flexible diffusion models. Moreover, diffusion models naturally enable posterior sampling, i.e. $\mathbf{x} \sim p_{X|Y}(\mathbf{x}|\mathbf{y})$, allowing us to take advantage of the benefits thereof [15], [26], [27] with respect to the MAP estimator which simply collapses the posterior distribution into a single point estimate.

III. Background

Score-based diffusion models have been introduced independently as score-based models [28], [29] and denoising diffusion probabilistic modeling (DDPM) [30]. In this work, we will consider the formulation introduced by [8], which unifies both perspectives on diffusion models by expressing diffusion as a continuous-time process through stochastic differential equations (SDE). Diffusion models produce samples by reversing a corruption (noising) process. In essence, these models are trained to denoise their inputs for each timestep in the corruption process. Through iteration of this reverse process, samples can be drawn from a learned data distribution, starting from random noise.

The diffusion process of the data $\left\{\mathbf{x}_t \in \mathbb{R}^d\right\}_{t\in[0,1]}$ is characterized by a continuous sequence of Gaussian perturbations of increasing magnitude indexed by time $t \in [0, 1]$. Starting from the data distribution at $t = 0$, clean images are defined by $\mathbf{x}_0 \sim p(\mathbf{x}_0) \equiv p(\mathbf{x})$. Forward diffusion can be described using an SDE as follows:

$$\mathrm{d}\mathbf{x}_t = f(t)\mathbf{x}_t\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}, \tag{5}$$

where $\mathbf{w} \in \mathbb{R}^d$ is a standard Wiener process, $f(t) : [0, 1] \to \mathbb{R}$ and $g(t) : [0, 1] \to \mathbb{R}$ are the drift and diffusion coefficients, respectively. Moreover, these coefficients are chosen so that the resulting distribution $p(\mathbf{x}_1)$ at the end of the perturbation process approximates a predefined base distribution $p(\mathbf{x}_1) \approx \pi(\mathbf{x}_1)$. Furthermore, the transition kernel of the diffusion process can be defined in one step as $q(\mathbf{x}_t|\mathbf{x}_0) \sim \mathcal{N}(\mathbf{x}_t|\alpha_t\mathbf{x}_0, \beta_t^2\mathbf{I})$, where α_t and β_t can be analytically derived from the SDE.

Naturally, we are interested in reversing the diffusion process, so that we can sample from $\mathbf{x}_0 \sim p(\mathbf{x}_0)$. The reverse diffusion process is also a diffusion process given by the reverse-time SDE [8], [31]:

$$\mathrm{d}\mathbf{x}_t = \big[f(t)\mathbf{x}_t - g(t)^2 \underbrace{\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)}_{\text{score}}\big]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}_t \tag{6}$$

where $\bar{\mathbf{w}}_t$ is the standard Wiener process in the reverse direction. The gradient of the log-likelihood of the data with respect to itself, a.k.a. the *score function*, arises from the reverse-time SDE. The score function is a gradient field pointing back to the data manifold and can intuitively be used to guide a random sample from the base distribution $\pi(\mathbf{x})$ to the desired data distribution. Given a dataset $\mathcal{X} = \left\{\mathbf{x}_0^{(1)}, \mathbf{x}_0^{(2)}, \dots, \mathbf{x}_0^{(|\mathcal{X}|)}\right\} \sim p(\mathbf{x}_0)$, scores can be estimated by training a neural network $s_\theta(\mathbf{x}_t, t)$ parameterized by weights θ, with score matching techniques such as the denoising score matching (DSM) objective [7]:

$$\begin{aligned}\theta^* = \arg\min_{\theta} \mathbb{E}_{t\sim U[0,1]}\bigg\{&\mathbb{E}_{(\mathbf{x}_0,\mathbf{x}_t)\sim p(\mathbf{x}_0)q(\mathbf{x}_t|\mathbf{x}_0)} \\ &\Big[||s_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t|\mathbf{x}_0)||_2^2\Big]\bigg\}. \end{aligned} \tag{7}$$

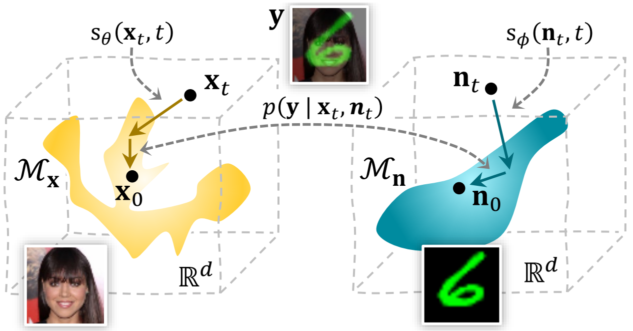


Fig. 1: Overview of the proposed joint posterior sampling method for removing structured noise using diffusion models. During the sampling process, the solutions for both signal and noise move toward their respective data manifold $\mathcal{M}$ through score models s_θ and s_ϕ. At the same time, the data consistency term ensures solutions that are in line with the (structured) noisy measurement $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$.

Given a sufficiently large dataset $\mathcal{X}$ and model capacity, DSM ensures that the score network converges to $s_\theta(\mathbf{x}_t, t) \simeq \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$. After training the time-dependent score model s_θ, it can be used to calculate the reverse-time diffusion process and solve the trajectory using numerical samplers such as the Euler-Maruyama algorithm. Alternatively, more sophisticated samplers, such as ALD [28], probability flow ODE [8], and Predictor-Corrector sampler [8], can be used to further improve sample quality.

These iterative sampling algorithms discretize the continuous time SDE into a sequence of time steps $\{0 = t_0, t_1, \dots, t_T = 1\}$, where a noisy sample $\hat{\mathbf{x}}_{t_i}$ is denoised to produce a sample for the next time step $\hat{\mathbf{x}}_{t_{i-1}}$. The resulting samples $\{\hat{\mathbf{x}}_{t_i}\}_{i=0}^{T}$ constitute an approximation of the actual diffusion process $\{\mathbf{x}_t\}_{t\in[0,1]}$.

IV. Method

A. Joint Posterior Sampling under Structured Noise

We are interested in posterior sampling under structured noise. We recast this as a joint optimization problem with respect to the signal $\mathbf{x}$ and noise $\mathbf{n}$ given by:

$$(\mathbf{x}, \mathbf{n}) \sim p_{X,N}(\mathbf{x}, \mathbf{n}|\mathbf{y}) \propto p_{Y|X,N}(\mathbf{y}|\mathbf{x}, \mathbf{n}) \cdot p_X(\mathbf{x}) \cdot p_N(\mathbf{n}). \tag{8}$$

Solving inverse problems using diffusion models requires conditioning of the diffusion process on the observation $\mathbf{y}$, such that we can sample from the posterior $p_{X|Y}(\mathbf{x}, \mathbf{n}|\mathbf{y})$. Therefore, we construct a *joint conditional* diffusion process $\{\mathbf{x}_t, \mathbf{n}_t|\mathbf{y}\}_{t\in[0,1]}$, in turn producing a *joint conditional* reverse-time SDE:

$$\begin{aligned} \mathrm{d}(\mathbf{x}_t, \mathbf{n}_t) = \big[f(t)(\mathbf{x}_t, \mathbf{n}_t) - \dots \\ g(t)^2 \nabla_{\mathbf{x}_t, \mathbf{n}_t} \log p(\mathbf{x}_t, \mathbf{n}_t|\mathbf{y})\big] \mathrm{d}t + g(t) \mathrm{d}\bar{\mathbf{w}}_t. \end{aligned} \tag{9}$$

	ΠGDM [34]	DPS [23]	Projection [8]
$\boldsymbol{\gamma}_t$	$\mathbf{y}$	$\mathbf{y}$	$\hat{\mathbf{y}}_t$
$\boldsymbol{\mu}_t$	$A\mathbf{x}_{0\vert t} + \mathbf{n}_{0\vert t}$	$A\mathbf{x}_{0\vert t} + \mathbf{n}_{0\vert t}$	$A\mathbf{x}_t + \mathbf{n}_t$
$\boldsymbol{\Sigma}_t$	$\left(r_t^2 \mathbf{A}\mathbf{A}^\mathsf{T} + q_t^2 I\right)$	$\rho^2 I$	$\rho^2 I$
λ	$\lambda' r_t^2 / g(t)^2$	$\lambda' \rho^2 / \left(g(t)^2 \vert\mathbf{y} - \mu\vert_2^1\right)$	$\lambda' \rho^2 / g(t)^2$
κ	$\kappa' q_t^2 / g(t)^2$	$\kappa' \rho^2 / \left(g(t)^2 \vert\mathbf{y} - \mu\vert_2^1\right)$	$\kappa' \rho^2 / g(t)^2$

TABLE I: Parameter choices for the Gaussian model of the noise-perturbed likelihood function in (11).

We would like to factorize the posterior using our learned *unconditional* score model and tractable measurement model, given the joint formulation. Consequently, we construct two separate diffusion processes, defined by separate score models but entangled through the measurement model $p_{Y|X,N}(\mathbf{y}|\mathbf{x}, \mathbf{n})$. In addition to the original score model $s_\theta(\mathbf{x}, t)$, we introduce a second score model $s_\phi(\mathbf{n}_t, t) \simeq \nabla_{\mathbf{n}_t} \log p_N(\mathbf{n}_t)$, parameterized by weights ϕ, to model the expressive noise component $\mathbf{n}$. These two score networks can be trained independently on datasets for $\mathbf{x}$ and $\mathbf{n}$, respectively, using the objective in (7). The gradients of the posterior with respect to $\mathbf{x}$ and $\mathbf{n}$, used in (9), are now given by:

$$\begin{cases} \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t, \mathbf{n}_t|\mathbf{y}) \simeq s_\theta^*(\mathbf{x}_t, t) + \lambda \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t, \mathbf{n}_t) \\ \nabla_{\mathbf{n}_t} \log p(\mathbf{x}_t, \mathbf{n}_t|\mathbf{y}) \simeq s_\phi^*(\mathbf{n}_t, t) + \kappa \nabla_{\mathbf{n}_t} \log p(\mathbf{y}|\mathbf{x}_t, \mathbf{n}_t), \end{cases} \tag{10}$$

which simply factorizes the joint posterior into prior and likelihood terms using Bayes' rule from (8) for both diffusion processes. Following the literature on classifier-(free) diffusion guidance [32], [33] and diffusion for inverse problems [8], [23], [34], two Bayesian weighting terms, λ and κ, are also introduced. These terms are tunable hyper-parameters that weigh the importance of following the prior, $s_\theta^*(\mathbf{x}_t, t)$ and $s_\phi^*(\mathbf{n}_t, t)$, versus the measurement model, $\nabla_{\mathbf{n}_t} \log p(\mathbf{y}|\mathbf{x}_t, \mathbf{n}_t)$.

B. Data Consistency Rules

The resulting *true* noise-perturbed likelihood $p(\mathbf{y}|\mathbf{x}_t, \mathbf{n}_t)$ is generally intractable, unlike $p(\mathbf{y}|\mathbf{x}_0, \mathbf{n}_0)$. Different approximations have been proposed in recent works [8], [23], [24], [34]–[36]. Our method is agnostic to the type of data-consistency rule employed. To study its effect on the final output, we will implement three strong approaches proposed in literature, namely, Pseudoinverse-Guided Diffusion Models (ΠGDM) [34], Diffusion Posterior Sampling (DPS) [23], and projection [8].

In all methods, to ensure traceability of $p(\mathbf{y}|\mathbf{x}_t, \mathbf{n}_t)$, it is modeled as a Gaussian, namely:

$$p(\mathbf{y}|\mathbf{x}_t, \mathbf{n}_t) \approx \mathcal{N}(\boldsymbol{\gamma}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \tag{11}$$

where the three different methods employ different approximations for the parameters of the Normal distribution. In all three methods, the co-variance $\boldsymbol{\Sigma}_t$ is not a function of $\mathbf{x}_t$ or $\mathbf{n}_t$, and we can thus write the noise-perturbed likelihood score as:

$$\nabla_{\mathbf{x}_t, \mathbf{n}_t} \log p(\mathbf{y}|\mathbf{x}_t, \mathbf{n}_t) \approx \left[\nabla_{\mathbf{x}_t, \mathbf{n}_t} \boldsymbol{\mu}_t\right] \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{\gamma}_t - \boldsymbol{\mu}_t). \tag{12}$$

We will now derive the three different data-consistency rules for our joint-diffusion process. Additionally, Table I shows an overview of the choices made for each parameter in the different methods.

1) ΠGDM: The ΠGDM method starts with an approximation of $\mathbf{x}_t, \mathbf{n}_t$ toward $\mathbf{x}_0, \mathbf{n}_0$, which then allows the usage of the known relationship of $p(\mathbf{y}|\mathbf{x}_0, \mathbf{n}_0)$, as seen in (1). Since $\mathbf{y}$, $\mathbf{x}_t$, and $\mathbf{n}_t$ are conditionally independent given $\mathbf{x}_0$ and $\mathbf{n}_0$, we can write:

$$p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)=\int_{\mathbf{x}_0}\int_{\mathbf{n}_0}p(\mathbf{x}_0|\mathbf{x}_t)p(\mathbf{n}_0|\mathbf{n}_t)p(\mathbf{y}|\mathbf{x}_0,\mathbf{n}_0)\mathrm{d}\mathbf{n}_0\mathrm{d}\mathbf{x}_0, \tag{13}$$

which is a marginalization over $\mathbf{x}_0$ and $\mathbf{n}_0$. Now, we have substituted the intractability of computing $p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)$, for the intractability of computing (scores of) $p(\mathbf{x}_0|\mathbf{x}_t)$ and $p(\mathbf{n}_0|\mathbf{n}_t)$. ΠGDM then estimates $p(\mathbf{x}_0|\mathbf{x}_t)$ using variational inference (VI), where it models the reverse diffusion steps as Gaussians, which we extend here to the noise as well:

$$\begin{cases} p(\mathbf{x}_0|\mathbf{x}_t)\approx\mathcal{N}(\mathbf{x}_{0|t},r_t^2I) \\ p(\mathbf{n}_0|\mathbf{n}_t)\approx\mathcal{N}(\mathbf{n}_{0|t},q_t^2I), \end{cases} \tag{14}$$

where q_t^2 and r_t^2 represent the uncertainty or error made in the VI. The means of the Gaussian approximations $(\mathbf{x}_{0|t},\mathbf{n}_{0|t})$ are calculated using Tweedie's formula, which can be thought of as a one-step denoising process using our trained diffusion model to estimate the *true* $\mathbf{x}_0$ and $\mathbf{n}_0$:

$$\mathbf{x}_{0|t}=\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]=\frac{\mathbf{x}_t+\beta_t^2\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t)}{\alpha_t}\approx\frac{\mathbf{x}_t+\beta_t^2 s_\theta^*(\mathbf{x}_t,t)}{\alpha_t}, \tag{15}$$

with an analogous equation for $\mathbf{n}_{0|t}$. Here, α_t and β_t can be derived from the SDE formulation as mentioned in Section III. Substitution of the VI estimate (14) into equation (13), then results in an approximation of the noise-perturbed likelihood:

$$p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx\mathcal{N}(\boldsymbol{\gamma}_t;\boldsymbol{\mu}_t,\boldsymbol{\Sigma}_t)\begin{cases}\boldsymbol{\gamma}_t=\mathbf{y}\\ \boldsymbol{\mu}_t=\mathbf{A}\mathbf{x}_{0|t}+\mathbf{n}_{0|t}\\ \boldsymbol{\Sigma}_t=r_t^2\mathbf{A}\mathbf{A}^\intercal+q_t^2I.\end{cases} \tag{16}$$

Moreover, using ΠGDM, we derive the following estimated noise-perturbed likelihood scores:

$$\begin{cases}\nabla_{\mathbf{x}_t}\log p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx(\nabla_{\mathbf{x}_t}\mathbf{x}_{0|t})\ \mathbf{A}^\intercal\boldsymbol{\Sigma}_t^{-1}(\mathbf{y}-A\mathbf{x}_{0|t}-\mathbf{n}_{0|t})\\ \nabla_{\mathbf{n}_t}\log p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx(\nabla_{\mathbf{n}_t}\mathbf{n}_{0|t})\qquad\boldsymbol{\Sigma}_t^{-1}(\mathbf{y}-A\mathbf{x}_{0|t}-\mathbf{n}_{0|t}),\end{cases} \tag{17}$$

where $\nabla_{\mathbf{x}_t}\mathbf{x}_{0|t}$ and $\nabla_{\mathbf{n}_t}\mathbf{n}_{0|t}$ are the Jacobians of (15), which can be computed using automatic differentiation methods.

In ΠGDM, the Bayesian weighting terms λ and κ are not fixed scalars, rather these are chosen to be equal to the estimated VI variances, r_t^2 and q_t^2. Additionally, the diffusion coefficient $g(t)^2$ gets cancelled out in the weighting scheme. Lastly, in this work, we introduce the additional explicit scalars λ' and κ', to bring it in line with the other data consistency rules. Note that introducing these scalars is the same as scaling r_t^2 and q_t^2 by a fixed amount for all timesteps.

Song *et al.* provide recommendations for choosing the variance of the VI [34], namely $r_t^2=\frac{\beta^2}{\beta^2-1}$, when the noise model is a known tractable distribution, which we adopt. Additionally, since we here introduce the notion of modeling $\mathbf{n}$ using a different diffusion model, we also set the variance of the VI estimate of $p(\mathbf{n}_0|\mathbf{n}_t)$ to $q_t^2=r_t^2$, as it is subjected to a similar SDE trajectory.

2) DPS: Diffusion Posterior Sampling (DPS) [23] also leverages Tweedie's formula in order to estimate $\mathbf{x}_{0|t}$ and $\mathbf{n}_{0|t}$. However, unlike ΠGDM, DPS does not leverage VI with Gaussian posteriors. Instead, a Gaussian error with diagonal covariance and variance ρ^2 is assumed, which again we can adapt to our problem as such:

$$p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx\mathcal{N}(\boldsymbol{\gamma}_t;\boldsymbol{\mu}_t,\boldsymbol{\Sigma}_t)\begin{cases}\boldsymbol{\gamma}_t=\mathbf{y}\\ \boldsymbol{\mu}_t=\mathbf{A}\mathbf{x}_{0|t}+\mathbf{n}_{0|t}\\ \boldsymbol{\Sigma}_t=\rho^2I,\end{cases} \tag{18}$$

resulting in the following scores:

$$\begin{cases}\nabla_{\mathbf{x}_t}\log p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx\frac{1}{\rho^2}(\nabla_{\mathbf{x}_t}\mathbf{x}_{0|t})\ \mathbf{A}^\intercal(\mathbf{y}-A\mathbf{x}_{0|t}-\mathbf{n}_{0|t})\\ \nabla_{\mathbf{n}_t}\log p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx\frac{1}{\rho^2}(\nabla_{\mathbf{n}_t}\mathbf{n}_{0|t})\qquad(\mathbf{y}-A\mathbf{x}_{0|t}-\mathbf{n}_{0|t}).\end{cases} \tag{19}$$

Note the difference between equations (17) and (19). The former employs a non-diagnonal covariance matrix, while the latter uses a simple diagonal approximation. In other words, DPS does not take into account how the variance of the estimation of $\mathbf{x}_{0|t}$ gets mapped to $\mathbf{y}$, in the case of a non-diagonal measurement matrix $\mathbf{A}$. The authors of DPS [23] then propose to rescale the noise, or step size, of the noise-perturbed likelihood score by a fixed scalar divided by the norm of the noise-perturbed likelihood. Additionally, the diffusion coefficient $g(t)^2$ gets cancelled out in the weighting scheme. Again, we achieve that here by choosing λ and κ appropriately.

3) Projection: The projection method [8] takes another approach altogether in comparison with ΠGDM and DPS. Instead of relating $\mathbf{x}_t,\mathbf{n}_t$ toward $\mathbf{x}_0,\mathbf{n}_0$, it relates $\mathbf{y}$ to $\mathbf{y}_t$ and then uses the following approximation:

$$p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx p(\hat{\mathbf{y}}_t|\mathbf{x}_t,\mathbf{n}_t), \tag{20}$$

where $\hat{\mathbf{y}}_t$ is a sample from $p(\mathbf{y}_t|\mathbf{y})$, and $\{\mathbf{y}_t\}_{t\in[0,1]}$ is an additional stochastic process that essentially corrupts the observation along the SDE trajectory together with $\mathbf{x}_t$. Note that in the case of a linear measurement $p(\mathbf{y}_t|\mathbf{y})$ is tractable, and we can easily compute $\hat{\mathbf{y}}_t=\alpha_t\mathbf{y}+\beta_t\mathbf{A}\mathbf{z}$, using the reparameterization trick with $\mathbf{z}\in\mathbb{R}^d\sim\mathcal{N}(\mathbf{0},\mathbf{I})$, see [10].

We then use the measurement model of equation (1), which is normally only defined for time $t=0$, and apply it to the current timestep t. In this approximation, we assume that we make a Gaussian error with diagonal covariance and standard deviation ρ^2 as

$$p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx\mathcal{N}(\boldsymbol{\gamma}_t;\boldsymbol{\mu}_t,\boldsymbol{\Sigma}_t)\begin{cases}\boldsymbol{\gamma}_t=\hat{\mathbf{y}}_t\\ \boldsymbol{\mu}_t=\mathbf{A}\mathbf{x}_t+\mathbf{n}_t\\ \boldsymbol{\Sigma}_t=\rho^2I.\end{cases} \tag{21}$$

Calculating the score of (21) with respect to both $\mathbf{x}_t$ and $\mathbf{n}_t$ then results in:

$$\begin{cases}\nabla_{\mathbf{x}_t}\log p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx\frac{1}{\rho^2}\mathbf{A}^\intercal(\hat{\mathbf{y}}_t-\mathbf{A}\mathbf{x}_t-\mathbf{n}_t)\\ \nabla_{\mathbf{n}_t}\log p(\mathbf{y}|\mathbf{x}_t,\mathbf{n}_t)\approx\frac{1}{\rho^2}\qquad(\hat{\mathbf{y}}_t-\mathbf{A}\mathbf{x}_t-\mathbf{n}_t).\end{cases} \tag{22}$$

Simillar to DPS, we also reweigh the scores in order to cancel out both $g(t)^2$ and $1/\rho^2$, using λ and κ, see Table I.

V. Related Work

In this section, we will cover two other works that tackle inverse problems with structured noise with the use of competitive deep generative models, namely normalizing flows (NF) and generative adversarial networks (GAN). These methods will

Algorithm 1: Joint posterior sampling with ΠGDM for score-based diffusion models

```
Require: $T, s_\theta, s_\phi, \lambda, \kappa, r_t^2, q_t^2, \mathbf{y}$
$\mathbf{x}_T \sim \pi(\mathbf{x}), \mathbf{n}_1 \sim \pi(\mathbf{n}), \Delta t \leftarrow \frac{1}{T}$

for $i = T-1$ to $0$ do
    $t \leftarrow \frac{i+1}{T}$
    // Data consistency steps
    $\mathbf{x}_{0|t} \leftarrow (\mathbf{x}_t + \beta_t^2 s_\theta^*((\mathbf{x}_t, t))/\alpha_t$
    $\mathbf{n}_{0|t} \leftarrow (\mathbf{n}_t + \beta_t^2 s_\phi^*((\mathbf{n}_t, t))/\alpha_t$
    $\boldsymbol{\mu}_t \leftarrow \mathbf{A}\mathbf{x}_{0|t} + \mathbf{n}_{0|t}$
    $\boldsymbol{\Sigma}_t \leftarrow r_t^2 \mathbf{A}\mathbf{A}^\intercal + q_t^2 I$
    $\mathbf{x}_t \leftarrow \mathbf{x}_t - \lambda r_t^2 (\nabla_{\mathbf{x}_t} \mathbf{x}_{0|t}) \mathbf{A}^\intercal \boldsymbol{\Sigma}_t^{-1} (\mathbf{y} - \boldsymbol{\mu}_t)$
    $\mathbf{n}_t \leftarrow \mathbf{n}_t - \kappa q_t^2 (\nabla_{\mathbf{n}_t} \mathbf{n}_{0|t}) \quad \boldsymbol{\Sigma}_t^{-1} (\mathbf{y} - \boldsymbol{\mu}_t)$
...
...
    // Unconditional diffusion steps
    $\mathbf{x}_{t-\Delta t} \leftarrow \mathbf{x}_t - f(t)\mathbf{x}_t \Delta t$
    $\mathbf{x}_{t-\Delta t} \leftarrow \mathbf{x}_{t-\Delta t} + g(t)^2 s_\theta^*(\mathbf{x}_t, t)\Delta t$
    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
    $\mathbf{x}_{t-\Delta t} \leftarrow \mathbf{x}_{t-\Delta t} + g(t)\sqrt{\Delta t}\mathbf{z}$

    $\mathbf{n}_{t-\Delta t} \leftarrow \mathbf{n}_t - f(t)\mathbf{n}_t \Delta t$
    $\mathbf{n}_{t-\Delta t} \leftarrow \mathbf{n}_{t-\Delta t} + g(t)^2 s_\phi^*(\mathbf{n}_t, t)\Delta t$
    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
    $\mathbf{n}_{t-\Delta t} \leftarrow \mathbf{n}_{t-\Delta t} + g(t)\sqrt{\Delta t}\mathbf{z}$
end
return: $\mathbf{x}_0$
```

serve as baselines in all of our experiments in which we evaluate the presented diffusion-based denoiser. There are no separate diffusion methods amongst the baselines, as our approach is the first to address structured noise in inverse problem settings. That being said, we do show the compatibility of our method with current state-of-the-art guided diffusion samplers. Finally, the NF- and GAN-based methods discussed in the following section rely on MAP estimation, see Section II, whereas we perform posterior sampling.

A. Normalizing Flows

Whang *et al.* propose to use normalizing flows to model both the data and the noise distributions [25]. Normalizing flows are a special class of likelihood-based generative models that make use of an invertible mapping $G : \mathbb{R}^d \to \mathbb{R}^d$ to transform samples from a base distribution $p_Z(\mathbf{z})$ into a more complex multimodal distribution $\mathbf{x} = G(\mathbf{z}) \sim p_X(\mathbf{x})$. The invertible nature of the mapping G allows for exact density evaluation through the change of variables formula:

$$\log p_X(\mathbf{x}) = \log p_Z(\mathbf{z}) + \log |\det J_{G^{-1}}(\mathbf{x})|, \tag{23}$$

where J is the Jacobian that accounts for the change in volume between densities. Since exact likelihood computation is possible through the flow direction G^{-1}, the parameters of the generator network can be optimized to maximize likelihood of the training data. Subsequently, the inverse task is solved using the MAP estimation in (4):

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \left\{\log p_{G_N}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \log p_{G_X}(\mathbf{x})\right\}, \tag{24}$$

where G_N and G_X are generative flow models for the noise and data respectively. Analog to that, the solution can be solved in the latent space rather than the image space as follows:

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} \left\{\log p_{G_N}(\mathbf{y} - \mathbf{A}(G_X(\mathbf{z}))) + \lambda \log p_{G_X}(G_X(\mathbf{z}))\right\}. \tag{25}$$

Note that in (25) a smoothing parameter λ is added to weigh the prior and likelihood terms, as was also done in [25]. The optimal $\hat{\mathbf{x}}$ or $\hat{\mathbf{z}}$ can then be found by applying gradient ascent on equations (24) or (25), respectively.

B. Generative Adversarial Networks

Generative adversarial networks are implicit generative models that can learn the data manifold in an adversarial manner [37]. The generative model is trained with an auxiliary discriminator network that evaluates the generator's performance in a minimax game. The generator $G(\mathbf{z}) : \mathbb{R}^l \to \mathbb{R}^d$ maps latent vectors $\mathbf{z} \in \mathbb{R}^l \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ to the data distribution of interest. The structure of the generative model can also be used in inverse problem solving [4]. The objective can be derived from (2) and is given by:

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left\{||\mathbf{y} - AG_X(\mathbf{z})|| + \lambda ||z||_2^2\right\}, \tag{26}$$

where λ weights the importance of the prior with the measurement error. Similar to NF, the optimal $\hat{\mathbf{z}}$ can be found using gradient ascent. The ℓ_2 regularization term on the latent variable is proportional to negative log-likelihood under the prior defined by G_X, where the subscript denotes the density that the generator is approximating. While this method does not explicitly model the noise, it remains an interesting comparison, as the generator cannot reproduce the noise found in the measurement and can only recover signals that are in the range of the generator. Therefore, due to the limited support of the learned distribution, GANs can inherently remove structured noise. However, the representation error (i.e. observation lies far from the range of the generator [4]) imposed by the structured noise comes at the cost of recovery quality.

VI. Implementation Details

All data and noise models are trained on the CelebA dataset [38] and the MNIST dataset with 10000 and 27000 training samples, respectively. Images are resized to 64 × 64 pixels. We test on a randomly selected subset of 100 images and use both the peak signal-to noise ratio (PSNR) and structural similarity index (SSIM) to evaluate our results. Automatic hyperparameter tuning for optimal inference was performed for all baseline methods on a small validation set of only 5 images. All parameters used for training and inference can be found in the provided code repository linked in the paper.

A. Proposed Method

For both the score models, we use the NCSNv2 architecture as introduced in [29]. Given the two separate datasets, one for the data and one for the structured noise, two separate score models can be trained independently. This allows for easy adaptation of our method, since many existing trained score models can be reused. Only during inference, the two priors are combined through the proposed sampling procedure as described in Algorithm 1, using the adapted Euler-Maruyama sampler. We use the following SDE: $f(t) = 0$, $g(t) = \sigma^t$ with $\sigma = 25$ to define the diffusion trajectory. During each experiment, we run the sampler for $T = 600$ iterations.

B. Baseline Methods

The closest to our work is the flow-based noise model proposed by [25], discussed in Section V-A, which will serve as our main baseline. To boost the performance of this baseline and to make it more competitive we moreover replace the originally-used RealNVP [39] with the Glow architecture [40]. Glow is a widely used flow model highly inspired by RealNVP, with the addition of 1×1 convolutions before each coupling layer. We use the exact implementation found in [5], with a flow depth of $K = 18$, and number of levels $L = 4$, which has been optimized for the same CelebA dataset used in this work and thus should provide a fair comparison with the proposed method.

Secondly, GANs as discussed in Section V-B are used as a comparison. We train a DCGAN [41], with a generator latent input dimension of $l = 100$. The generator architecture consists of 4 strided 2D transposed convolutional layers, having 4×4 kernels yielding feature maps of 512, 256, 128 and 64. Each convolutional layer is followed by a batch normalization layer and ReLU activation.

Lastly, depending on the reconstruction task, classical non-data-driven methods are used as a comparison. For denoising experiments, we use the block-matching and 3D filtering algorithm (BM3D) [42], and in compressed sensing experiments, LASSO with wavelet basis [43]. Except for the flow-based method of [25], none of these methods explicitly model the noise distribution. Still, they are a valuable baseline, as they demonstrate the effectiveness of incorporating a learned structured noise prior rather than relying on simple noise priors.

VII. Experiments

We subject our method to a variety of inverse problems such as denoising and compressed sensing, all with an element of additive structured noise. To test the method's robustness, we repeat the experiments on both out-of-distribution (OoD) data and OoD noise. To show the capabilities of our method beyond the domain of natural images, we test the joint-conditional diffusion method on radio-frequency ultrasound data. Lastly, we compare the methods' computational performance. The proposed method outperforms the baselines both qualitatively and quantitatively in all experiments.

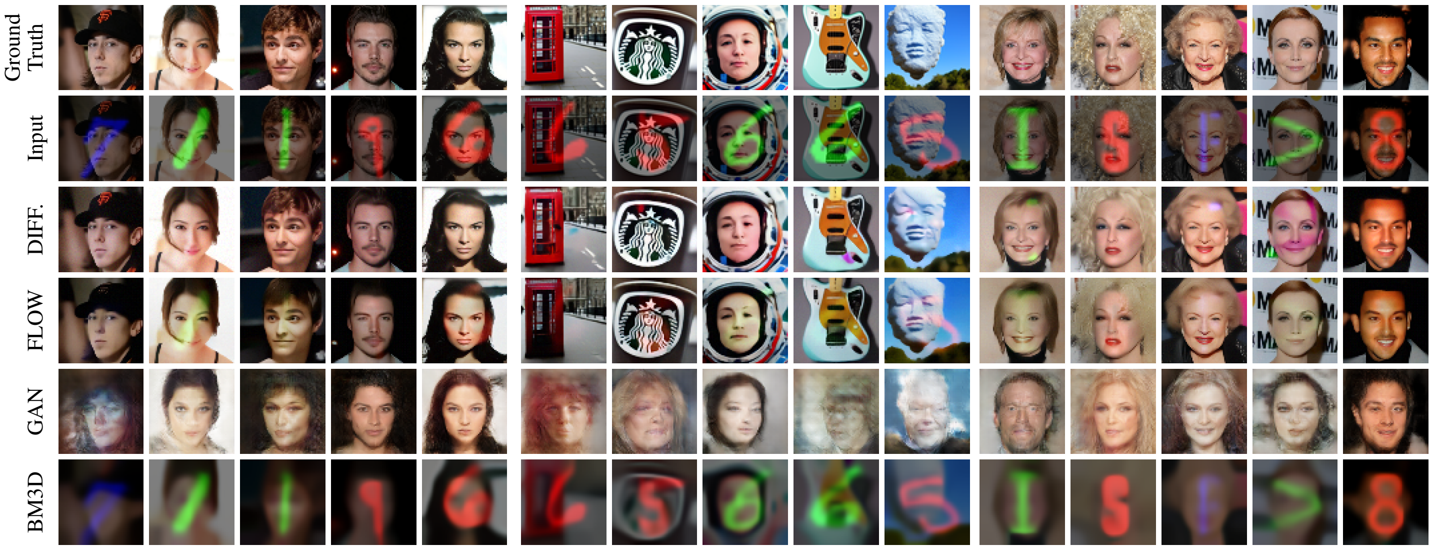
Removing MNIST digits: For comparison with [25], we recreate an experiment introduced in their work, where MNIST digits are added to CelebA faces. The corruption process is defined by $\mathbf{y} = 0.5 \cdot \mathbf{x}_{\text{CelebA}} + 0.5 \cdot \mathbf{n}_{\text{MNIST}}$. In this experiment, the noise score network s_ϕ is trained on the MNIST dataset. Fig. 4a shows a quantitative comparison of our method with all baselines. Furthermore, a random selection of test samples is shown in Fig. 2a for qualitative analysis. Diffusion and flow-based methods are able to recover the underlying signal, with the diffusion method preserving more details. Furthermore, we observe that for the flow-based method, initialization from the measurement is necessary to reproduce the results in [25] since random initialization does not converge. The GAN method is also able to remove the digits, but cannot accurately reconstruct the faces as it is unable to project the observation onto the range of the generator. Similarly, the BM3D denoiser fails to recover the underlying signal, confirming the importance of prior knowledge of the noise in this experiment. The metrics in Fig. 4a support these observations.

Structured noise with compressed sensing: In this experiment, the corruption process is defined by $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}_{\text{sine}}$ with a random Gaussian measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ and a noise with sinusoidal variance $\sigma_k \propto \exp\bigl(\sin\bigl(\frac{2\pi k}{16}\bigr)\bigr)$ for each pixel k, which we use to train s_ϕ on. The subsampling factor is defined by the size of the measurement matrix d/m. In Fig. 5a the results of the compressed sensing experiment and the comparison with the baselines are shown for an average standard deviation of $\sigma_N = 0.2$ and subsampling of factor $d/m = 2$. Similar to the results of the previous experiment, the diffusion method is more robust to the shift in distribution and is able to deliver high-quality recovery under the structured noise setting. In contrast, the flow-based method under-performs when subjected to the OoD data.

Removing sinusoidal noise: The corruption process is defined by $\mathbf{y} = \mathbf{x} + \mathbf{n}_{\text{sine}}$ where the noise variance $\sigma_k \propto \exp\bigl(\sin\bigl(\frac{2\pi k}{16}\bigr)\bigr)$ follows a sinusoidal pattern along each row of the image k. In this experiment, the score network s_ϕ is trained on a dataset generated with 1D sinusoidal noise samples $\mathbf{n}_{\text{sine}}$. See Fig. 3 for a comparison of our method to the flow-based method for varying noise variances. Both methods perform quite well, with the diffusion method having a slight edge. A visual comparison in Fig. 6, however, reveals that the diffusion method preserves more detail in general.

Out-of-distribution data: Additionally, to test the robustness of the methods, we recreate all experiments with out-of-distribution (OoD) data generated using a stable-diffusion text-to-image model [9] as well as random data from ImageNet. We use the exact same hyperparameters and trained models as in previous experiments. Qualitative and quantitative results can be found in Fig.2b and Fig.4b, respectively. Similarly to the findings of [5] and [25], the flow-based method is robust to OoD data, unlike the GAN. We empirically show that the diffusion method is also resistant to OoD data in inverse tasks with complex noise structures and even outperforms the flow-based method.

Out-of-distribution noise: In real-world applications, it can be challenging to accurately obtain noise samples for training the noise diffusion model. Even though not trivial, in many practical cases, noise can be measured, isolated from signals, or simulated. Still, these will be subjected to some distribution



(a) CelebA with MNIST noise (b) Out-of-distribution data (c) Out-of-distribution noise (TMNIST)

Fig. 2: Qualitative results on the removing MNIST digits (noise) from CelebA (signal) experiment, comparing our diffusion-based method (joint posterior sampling) to the baselines[2]: †FLOW, ‡GAN, and §BM3D.

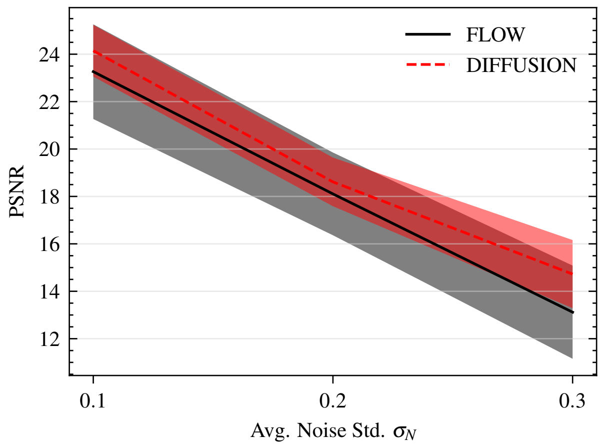


Fig. 3: Comparison of PSNR values for varying sinusoidal noise variances in the removing sinusoidal noise experiment. Shaded areas represent the standard deviation on the metric.

shift with respect to the true noise signals. Therefore, we extend the existing removing MNIST digits experiment with two OoD noise variants: (1) samples drawn from the TMNIST-Alphabet dataset containing different characters, and (2) random translations applied to noise (digits). As can be seen in Table II, our method is able to consistently outperform the most competitive flow baseline in the OoD noise experiments. Again, no retuning of hyperparameters or retraining of models was performed. In general, the random translation seemed to be a more challenging task compared to the TMNIST-alpha characters.

Performance: To highlight the difference in inference time between our method and the baselines, benchmarks are performed on a single 12GBytes NVIDIA GeForce RTX 3080 Ti, see Table III. A quick comparison of inference times reveals a $4\times$ (ΠGDM) or $10\times$ (Projection) difference in speed between ours and the flow-based method. All the deep generative models need approximately an equal amount of iterations $(T \approx 600)$ to converge. However, given the same modeling capacity, the flow model requires substantially more trainable parameters compared to the diffusion method. This is mainly due to the restrictive requirements imposed on the architecture to ensure tractable likelihood computation. It should be noted that in this work no improvements are applied to speed up the diffusion process, such as CCDF [19], for the diffusion method, leaving room for even more improvement in future work.

TABLE II: Results for the experiments with OoD noise.

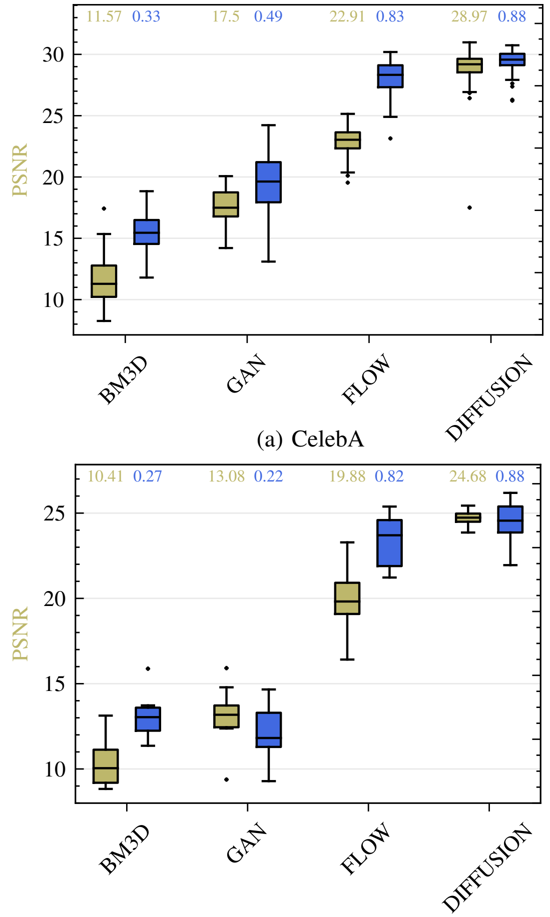
	Dataset	TMNIST		translation	
		PSNR	SSIM	PSNR	SSIM
*DIFF.	CelebA	25.94 ± 2.4	0.851 ± 0.04	23.63 ± 4.1	0.893 ± 0.04
†FLOW	CelebA	22.61 ± 1.1	0.826 ± 0.05	22.96 ± 1.1	0.837 ± 0.05
*DIFF.	OoD	22.59 ± 2.4	0.858 ± 0.06	21.55 ± 3.0	0.895 ± 0.05
†FLOW	OoD	20.06 ± 1.8	0.831 ± 0.08	20.54 ± 2.1	0.839 ± 0.08

TABLE III: Inference performance benchmark for all methods.

Model		# trainable parameters	Inference time [ms]
*DIFF.	(Proj.)	8.9M	5605
	(DPS)		16818
	(ΠGDM)		16094
†FLOW		25.8M	61853
‡GAN		3.9M	59
§BM3D		–	29

Comparison Data Consistency Methods: Although all three diffusion-based data consistency methods, as discussed in this section, outperform the baselines of Section V, ΠDGM provides the most consistent results with lower variance between samples, as shown in Fig. 7a. Empirically, this trend continues to be seen in the out-of-distribution datasets; see Fig. 7b. This is not surprising as ΠDGM has a more

[2] *Ours, † [25], ‡ [4], § [42], ¶ [43]



(a) CelebA

(b) Out-of-distribution data

Fig. 4: Quantitative results using PSNR (green) and SSIM (blue) for the removing MNIST digits experiment of the (a) CelebA and (b) out-of-distribution datasets.

sophisticated approximation for the noise-perturbed likelihood score compared to DPS and the projection method. A visual comparison is shown in Fig. 7c.

VIII. Discussions

Inverse problems are powerful tools for inferring unknown signals from observed measurements and have been at the center of many signal and image processing algorithms. Strong priors, often those learned through deep generative models, have played a crucial role in guiding these inferences, especially in the context of high-dimensional data. While complex priors on the signal are commonly employed, noise sources are often assumed to be simply distributed, drastically reducing the effectiveness of inverse problems in structured noise settings.

In this work, we address this limitation by introducing a novel joint posterior sampling technique. We not only leverage deep generative models to learn strong priors for the signal, but we also extend our approach to incorporate priors on the noise distribution. To achieve this, we employ an additional diffusion model that has been trained specifically to capture the characteristics of structured noise. Furthermore, we show the compatibility of our method with three existing posterior sampling techniques (projection, DPS, ΠGDM). We demonstrate our method on natural and out-of-distribution data and noise and achieve increased performance over the state-of-the-art and established conventional methods for complex inverse tasks. Additionally, the diffusion-based method is substantially easier to train using the score matching objective compared to other deep generative methods that rely on constrained neural architectures or adversarial training.

While our method is considerably faster and better in removing structured noise compared to the flow-based method [25], it is not ready (yet) for real-time inference and is still slow compared to GANs [4] and classical methods. Fortunately, research into accelerating the diffusion process is on its way. In addition, although a simple sampling algorithm was adopted in this work, many more sampling algorithms for score-based diffusion models exist. For example, the predictor-corrector (PC) sampler has been shown to improve sample quality [8]. Future work should explore this wide increase in design space to understand the limitations and possibilities of more sophisticated sampling schemes in combination with the proposed joint diffusion method. Furthermore, the range of problems to which we can apply the proposed method can be expanded into non-linear likelihood models as well as data from other domains, such as medical imaging. Lastly, the connection between diffusion models and continuous normalizing flows through the neural ODE formulation [44] is not investigated but is of great interest given the comparison with the flow-based method in this work.

IX. Conclusions

In this work, we presented a framework for removing structured noise using diffusion models. The proposed joint posterior sampling technique for diffusion models has been shown to effectively remove highly structured noise and outperform baselines in both image quality and computational performance. Additionally, it exhibits enhanced robustness in out-of-distribution scenarios. Our work provides an efficient addition to existing score-based conditional sampling methods by incorporating knowledge of the noise distribution, whilst supporting a variety of guided diffusion samplers. Future work should focus on speeding up the relatively slow inference process of diffusion models and furthermore investigate the applicability of the proposed method outside the realm of natural images.

References

[1] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.

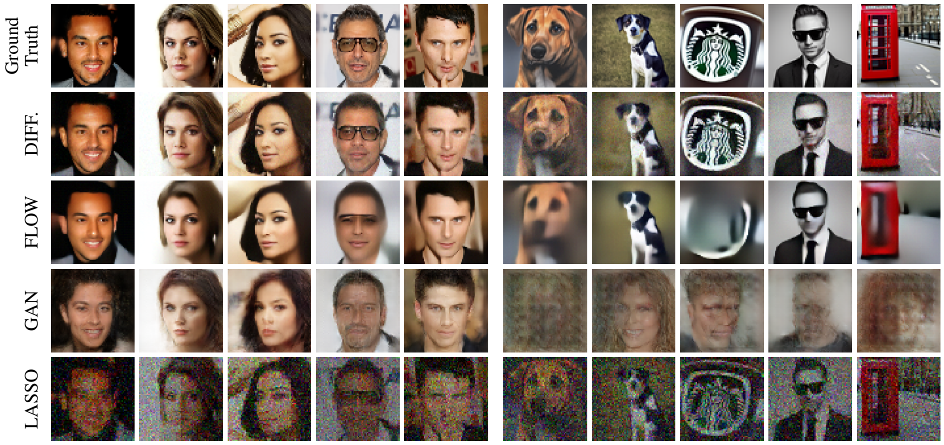
[2] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[3] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.

[4] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, “Compressed sensing using generative models,” in *International Conference on Machine Learning*. PMLR, 2017, pp. 537–546.

[5] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand, “Invertible generative models for inverse problems: mitigating representation error and dataset bias,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 399–409.

[6] X. Wei, H. van Gorp, L. Gonzalez-Carabarin, D. Freedman, Y. C. Eldar, and R. J. van Sloun, “Deep unfolding with normalizing flow priors for inverse problems,” *IEEE Transactions on Signal Processing*, vol. 70, pp. 2962–2971, 2022.



(a) CelebA with structured noise

(b) Out-of-distribution data

Fig. 5: Results on the compressed sensing with structured noise experiment, comparing our diffusion-based method to the baselines.

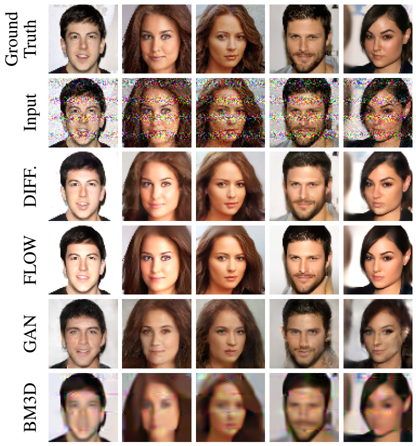
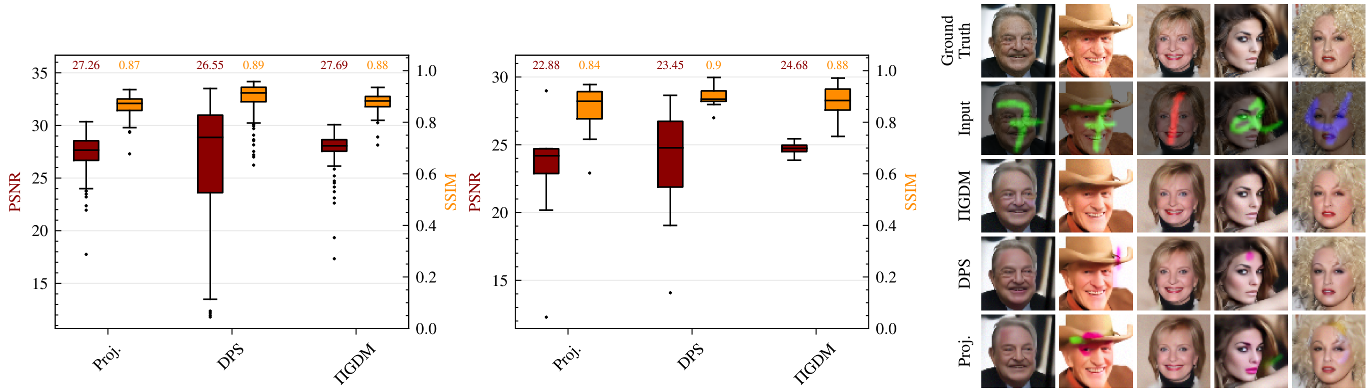


Fig. 6: Results on the removing sinusoidal noise experiment.



(a) CelebA + MNIST

(b) Out-of-distribution data

(c) CelebA + MNIST

Fig. 7: Comparison of the projection, DPS an ΠGDM data-consistency rules used in the joint posterior sampling method. Qualitative (c) and quantitative results are shown using PSNR (red) and SSIM (orange) for the removing MNIST digits experiment on images of the (a) CelebA and (b) out-of-distribution datasets.

[7] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[8] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2020.

[9] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 684–10 695.

[10] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *International Conference on Learning Representations*, 2021.

[11] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, "Robust compressed sensing mri with deep generative priors," *Advances in Neural Information Processing Systems*, vol. 34, pp. 14 938–14 954, 2021.

[12] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated mri," *Medical Image Analysis*, p. 102479, 2022.

[13] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving diffusion models for inverse problems using manifold constraints," *arXiv preprint arXiv:2206.00941*, 2022.

[14] A. Bansal, E. Borgnia, H.-M. Chu, J. S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein, "Cold diffusion: Inverting arbitrary image transforms without noise," *arXiv preprint arXiv:2208.09392*, 2022.

[15] G. Daras, Y. Dagan, A. Dimakis, and C. Daskalakis, "Score-guided intermediate level optimization: Fast langevin mixing for inverse problems," in *International Conference on Machine Learning*. PMLR, 2022, pp. 4722–4753.

[16] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," *arXiv preprint arXiv:2206.00364*, 2022.

[17] C. Luo, "Understanding diffusion models: A unified perspective," *arXiv preprint arXiv:2208.11970*, 2022.

[18] G. Daras, M. Delbracio, H. Talebi, A. G. Dimakis, and P. Milanfar, "Soft diffusion: Score matching for general corruptions," 2022.

[19] H. Chung, B. Sim, and J. C. Ye, "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 413–12 422.

[20] T. Salimans and J. Ho, "Progressive distillation for fast sampling of diffusion models," in *International Conference on Learning Representations*, 2021.

[21] B. Jing, G. Corso, R. Berlinghieri, and T. Jaakkola, "Subspace diffusion generative models," *arXiv preprint arXiv:2205.01490*, 2022.

[22] A. Vahdat, K. Kreis, and J. Kautz, "Score-based generative modeling in latent space," *Advances in Neural Information Processing Systems*, vol. 34, pp. 11 287–11 302, 2021.

[23] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," *arXiv preprint arXiv:2209.14687*, 2022.

[24] X. Meng and Y. Kabashima, "Diffusion model based posterior sampling for noisy linear inverse problems," *arXiv preprint arXiv:2211.12343*, 2022.

[25] J. Whang, Q. Lei, and A. Dimakis, "Solving inverse problems with

a flow-based noise model," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 146–11 157.

[26] A. Jalal, S. Karmalkar, A. Dimakis, and E. Price, "Instance-optimal compressed sensing via posterior sampling," in *International Conference on Machine Learning*. PMLR, 2021, pp. 4709–4720.

[27] B. Kawar, G. Vaksman, and M. Elad, "Snips: Solving noisy inverse problems stochastically," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21 757–21 769, 2021.

[28] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[29] ——, "Improved techniques for training score-based generative models," *Advances in neural information processing systems*, vol. 33, pp. 12 438–12 448, 2020.

[30] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[31] B. D. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982.

[32] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[33] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.

[34] J. Song, A. Vahdat, M. Mardani, and J. Kautz, "Pseudoinverse-guided diffusion models for inverse problems," in *International Conference on Learning Representations*, 2023.

[35] B. T. Feng, J. Smith, M. Rubinstein, H. Chang, K. L. Bouman, and W. T. Freeman, "Score-based diffusion models as principled priors for inverse imaging," *arXiv preprint arXiv:2304.11751*, 2023.

[36] M. A. Finzi, A. Boral, A. G. Wilson, F. Sha, and L. Zepeda-Núñez, "User-defined event sampling and uncertainty quantification in diffusion models for physical dynamical systems," in *International Conference on Machine Learning*. PMLR, 2023, pp. 10 136–10 152.

[37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[38] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[39] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real nvp," *arXiv preprint arXiv:1605.08803*, 2016.

[40] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," *Advances in neural information processing systems*, vol. 31, 2018.

[41] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[42] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image processing: algorithms and systems, neural networks, and machine learning*, vol. 6064. SPIE, 2006, pp. 354–365.

[43] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[44] Y. Song, C. Durkan, I. Murray, and S. Ermon, "Maximum likelihood training of score-based diffusion models," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 1415–1428.

Tristan Stevens holds a B.Sc. and M.Sc. degree (cum laude) in electrical engineering from the Eindhoven University of Technology, the Netherlands, earned in 2019 and 2021 respectively. Currently, he is working towards a Ph.D. degree with the Biomedical Diagnostics Laboratory at the Eindhoven University of Technology, Eindhoven, The Netherlands. His research is focused on the intersection of signal processing and probabilistic deep learning for medical ultrasound imaging. Additionally, he had the opportunity to work at Philips Research in Cambridge, USA, gaining experience in industry research centered around medical imaging.

Hans van Gorp received the B.Sc. and M.Sc. degrees in electrical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2018 and 2020, respectively. He is currently working towards a Ph.D. degree at both the Eindhoven University of Technology and Philips Sleep and Respiratory Care. His primary research interests include deep generative modeling, signal processing, and AI for medical applications, especially in the field of automatic sleep stage analysis.

F. Can Meral holds a B.Sc., M.Sc. degree in mechanical engineering from the Middle East Technical University, Ankara Turkey and Koç University, Istanbul Turkey, earned in 2003 and 2005 respectively. He received his Ph.D. from the University of Illinois at Chicago in 2010. He was a postdoctoral research fellow at Brigham and Women's Hospital and Harvard Medical School in Boston from 2010 to 2015. He joined Philips Research North America in 2015. Currently, he is a Senior Systems Engineer with Philips Ultrasound.

Junseob Shin was born in Seoul, Korea in 1986. He received his B.S. (Magna Cum Laude) and M.S. degrees in bioengineering from University of California, San Diego, CA, USA in 2008 and 2009, respectively, and his Ph.D. degree in biomedical engineering from the University of Southern California in Los Angeles, CA, USA, in 2014. He served as a postdoctoral research associate with the Earth and Environmental Sciences Division, Los Alamos National Laboratory, NM, USA from 2014 to 2015 before working as a senior scientist at Philips Research North America, Cambridge, MA, USA from 2015 to 2023. Currently, he is a principal ultrasound engineer at Accupulse Medical Inc. His current research interests include ultrasound beamforming, intracardiac echocardiography, 3-D imaging, and deep learning for ultrasound image formation and artifact reduction.

Jason Yu received his Ph.D. degree in electrical and computer engineering from Duke University in 2013. Originally from the Boston area of Massachusetts, he also received a B.Sc. degree in electrical engineering from Tufts University in 2007 and an M.Sc. degree in electrical and computer engineering from Duke University in 2009. While at Duke, his research focused on adaptive beamforming and MIMO array processing techniques for MTI radar, including development of a low-power MIMO array testbed. He then worked at MIT Lincoln Laboratory from 2013 to 2019 in the Airborne Radar Systems and Techniques group, where his work focused on radar signal processing, radar beamforming, radar systems analysis, and radar prototype development. In 2019, he started working at Philips, where his work focused on medical ultrasound image formation and beamforming, medical ultrasound image quality, and medical ultrasound color Doppler imaging.

Jean-Luc Robert Jean-luc Robert is a Senior Scientist, and Image Formation lead at Philips research North America. He received his M.Sc. as well as his Ph.D. degree in physics, in 2003 and 2007, respectively from Institut Langevin, under Mathias Fink's supervision. His research covers wave theory, array processing (beamforming) and signal processing to improve image quality of ultrasound systems, as well as provide quantification of tissue. This has led him to be one of the first person to discover ultra-fast imaging, and to be a pioneer of compressive sensing and deep learning in ultrasound. Jean-luc has contributed to 45 scientific publications and 80 patent applications and received numerous awards at Philips. Since 2020, Jean-luc has taken additional responsibilities in leading the global image formation portfolio at Philips research and shape the roadmap.

Ruud van Sloun is an Associate Professor at the Department of Electrical Engineering at the Eindhoven University of Technology in the Netherlands. He received the M.Sc. and Ph.D. degree (both cum laude) in Electrical Engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2014, and 2018, respectively. From 2019-2020 he was a Visiting Professor with the Department of Mathematics and Computer Science at the Weizmann Institute of Science, Rehovot, Israel, and from 2020-2023 he was a kickstart AI fellow at Philips Research. He received an ERC starting grant, an NWO VIDI grant, an NWO Rubicon grant, and a Google Faculty Research Award. His current research interests include closed-loop image formation, deep learning for signal processing and imaging, active signal acquisition, model-based deep learning, compressed sensing, ultrasound imaging, and probabilistic signal and image reconstruction.