Parallel Diffusion Models of Operator and Image for Blind Inverse Problems

Hyungjin Chung[1*] Jeongsol Kim[1*] Sehui Kim[2] Jong Chul Ye[2]
[1]Dept. of Bio & Brain Engineering,
[2]Kim Jae Chul Graduate School of AI,
KAIST
{hj.chung, jeongsol, sehui.kim, jong.ye}@kaist.ac.kr

Figure 1. Representative results and overall concept of the proposed method. (a) Results of blind deblurring. Both the image and the kernel in the bottom right corner are jointly estimated with the proposed method. (b) Results of imaging through turbulence. (c) Evolution of joint reconstruction with the proposed method. 1^{st}, 2^{nd} row illustrate the change of $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ and $\hat{\boldsymbol{k}}_0(\boldsymbol{k}_t)$ through time as $t = 1 \rightarrow 0$, with the measurement and the kernel initialization given on the first column.

Abstract

Diffusion model-based inverse problem solvers have demonstrated state-of-the-art performance in cases where the forward operator is known (i.e. non-blind). However, the applicability of the method to blind inverse problems has yet to be explored. In this work, we show that we can indeed solve a family of blind inverse problems by constructing another diffusion prior for the forward operator. Specifically, parallel reverse diffusion guided by gradients from the intermediate stages enables joint optimization of both the forward operator parameters as well as the image, such that both are jointly estimated at the end of the parallel reverse diffusion procedure. We show the efficacy of our method on two representative tasks — blind deblurring, and imaging through turbulence — and show that our method yields state-of-the-art performance, while also being flexible to be applicable to general blind inverse problems when we know the functional forms. Code available: https://github.com/BlindDPS/blind-dps

This work was supported by the National Research Foundation of Korea under Grant NRF-2020R1A2B5B03001980, by the KAIST Key Research Institute (Interdisciplinary Research Group) Project, by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-02068, Artificial Intelligence Innovation Hub), and by the IITP grant funded by the Korea government(MSIT) (No. 2022-0-00984, Development of Artificial Intelligence Technology for Personalized Plug-and-Play Explanation and Verification of Explanation).

1. Introduction

Inverse problems subsume a wide set of important problems in science and engineering, where the objective is to recover the latent image from the corrupted measurement, generated by the forward operator. Considering the taxonomy, they can be split into two major categories — *non-blind* inverse problems, and *blind* inverse problems. The former considers the cases where the forward operator is known, and hence eases the problem. In contrast, the latter considers the cases where the operator is *unknown*, and thus the operator needs to be estimated together with the reconstruction of the latent image. The latter problem is considerably harder than the former problem, as joint minimization is typically much less stable.

In this work, we mainly focus on leveraging generative priors to solve inverse problems in imaging. Among many different generative model classes, diffusion models have established the new state-of-the-art. In diffusion models, we define the *forward* data noising process, which gradually corrupts the image into white Gaussian noise. The generative process is defined by the *reverse* of such process, where each step of reverse diffusion is governed by the score function [53]. With the recent surge of diffusion models, it has been demonstrated in literature that diffusion models are not only powerful generative models, but also excellent generative priors to solve inverse problems. Namely, one can either resort to iterative projections to the measurement subspace [13, 53], or estimate posterior sampling [11] to arrive at feasible solutions that meet the data consistency. For both linear [13, 27, 53] and some non-linear [11, 51] inverse problems, guiding unconditional diffusion models to solve down-stream inverse problems were shown to have stronger performance even when compared to the fully supervised counterparts.

Nevertheless, current solvers are strictly limited to cases where the forward operator is known and fixed. For example, [11, 27] consider non-blind deblurring with known kernels. The problem now boils down to optimizing only for the latent image, since the likelihood can be computed robustly. Unfortunately, in real world problems, knowing the kernel exactly is impractical. It is often the case where the kernel is also unknown, and we have to jointly estimate the image and the kernel. In such cases, not only do we need a prior model of the image, but we also need some proper prior model of the kernel [41, 55]. While conventional methods exploit, e.g. patch-based prior [55], sparsity prior [41], etc., they often fall short of accurate modeling of the distribution.

In this work, we aim to leverage the ability of diffusion models to act as strong generative priors and propose *BlindDPS* (Blind Diffusion Posterior Sampling) — constructing multiple diffusion processes for learning the prior of each component — which enable posterior sampling even when the operator is unknown. BlindDPS starts by initializing both the image and the operator parameter with Gaussian noise. Reverse diffusion progresses in parallel for both models, where the cross-talk between the paths are enforced from the approximate likelihood and the measurement, as can be seen in Fig. 2. With our method, both the image and the kernel starts with a coarse estimation, gradually getting closer to the ground truth as $t \to 0$ (see Fig. 1(c)).

In fact, our method can be thought of as a coarse-to-fine strategy naturally admitting a Gaussian scale-space representation [29, 36], which can be seen as a continuous generalization of the coarse-to-fine optimization strategy that most of the optimization-based methods take [41, 44]. Furthermore, our method is generally applicable to cases where we know the *structure* of the forward model a priori (e.g. convolution). To demonstrate the generality, we further show that our method can also be applied in imaging through turbulence. From our experiments, we show that the proposed method yields state-of-the-art performance while being generalizable to different inverse problems.

2. Background

Diffusion models Variance preserving (VP) diffusion models (i.e. DDPM [21]), in the score-based persepctive [53], define the forward noising process of the data $\boldsymbol{x}(t) \triangleq \boldsymbol{x}_t$, $t \in [0, 1]$ with a linear stochastic differential equation (SDE)

$$d\boldsymbol{x} = -\frac{\beta(t)}{2}\boldsymbol{x}dt + \sqrt{\beta(t)}d\boldsymbol{w}, \tag{1}$$

where $\beta(t)$ is the noise schedule, and $\boldsymbol{w}$ is the standard Brownian motion. One can define a proper noise schedule $\beta(t)$ such that the data distribution $\boldsymbol{x}(0) \sim p_0 = p_{\text{data}}$ is molded into the standard Gaussian distribution $\boldsymbol{x}(1) \sim p_1 \simeq \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Then, the corresponding reverse SDE is given by [2]

$$d\boldsymbol{x} = \left[-\frac{\beta(t)}{2}\boldsymbol{x} - \beta(t)\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)\right]dt + \sqrt{\beta(t)}d\bar{\boldsymbol{w}}, \tag{2}$$

where $\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)$ is the score function, typically approximated by denoising score matching (DSM) [56]

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{t,\boldsymbol{x}_t,\boldsymbol{x}_0}\left[\|\boldsymbol{s}_\theta(\boldsymbol{x}_t, t) - \nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{x}_t|\boldsymbol{x}_0)\|_2^2\right]. \tag{3}$$

Once trained, we can use the plug-in estimate $\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t) \simeq \boldsymbol{s}_\theta(\boldsymbol{x}_t, t)$ for the reverse diffusion in (2), and solve by discretization (e.g. ancestral sampling of [21]), effectively sampling from the prior distribution $p(\boldsymbol{x}_0)$.

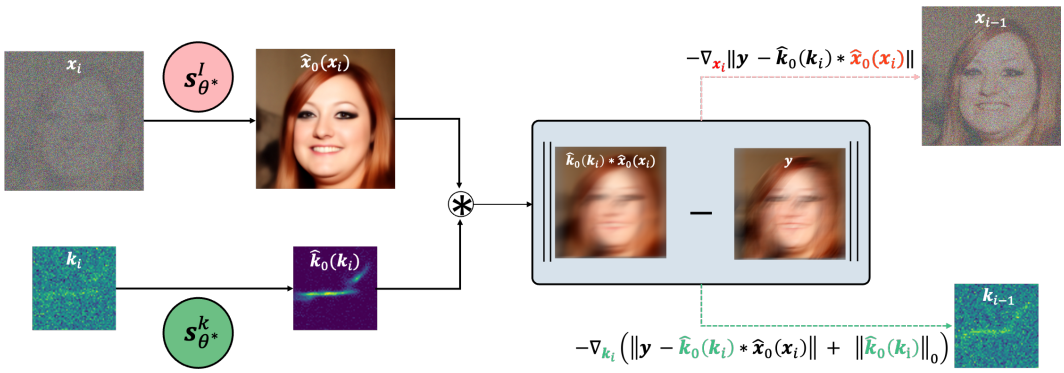

Figure 2. Description of BlindDPS. From the intermediate (noisy) estimate $\boldsymbol{x}_i, \boldsymbol{k}_i$, we achieve the denoised representation $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_i), \hat{\boldsymbol{k}}_0(\boldsymbol{k}_i)$ through Tweedie's formula with the score functions $\boldsymbol{s}^i_{\theta^*}, \boldsymbol{s}^k_{\theta^*}$. The residual $\|\boldsymbol{y} - \hat{\boldsymbol{k}}_0 * \hat{\boldsymbol{x}}_0\|$ is computed with the denoised estimates, and the residual-minimizing gradients are applied parallel to both diffusion processes.

Diffusion posterior sampling (DPS) Consider the following Gaussian measurement model

$$p(\boldsymbol{y}|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{y}|\mathcal{H}(\boldsymbol{x}_0), \sigma^2\boldsymbol{I}),\ \boldsymbol{y} \in \mathbb{R}^m,\ \boldsymbol{x}_0 \in \mathbb{R}^n, \quad (4)$$

where $\boldsymbol{y}$ is the corrupted measurement, $\boldsymbol{x}_0$ is the latent image that we wish to estimate, and $\mathcal{H}$ is the forward operator. As the problem is often ill-posed, it is desirable to be able to sample from the posterior distribution $p(\boldsymbol{x}_0|\boldsymbol{y})$. By Bayes' rule, we have for a general timestep t,

$$\begin{aligned}\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y}) &= \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t) \quad (5)\\ &\simeq \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t) + \boldsymbol{s}_{\theta^*}(\boldsymbol{x}_t, t), \quad (6)\end{aligned}$$

where we can plug (6) into the reverse diffusion (2) to sample from $p(\boldsymbol{x}_0|\boldsymbol{y})$, i.e.

$$\begin{aligned}d\boldsymbol{x} = (-\frac{\beta(t)}{2}\boldsymbol{x} - \beta(t)[\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t)\\ + \boldsymbol{s}_{\theta^*}(\boldsymbol{x}_t, t)])dt + \sqrt{\beta(t)}d\bar{\boldsymbol{w}}. \quad (7)\end{aligned}$$

Note that the time-conditional log-likelihood $\log p(\boldsymbol{y}|\boldsymbol{x}_t)$ is intractable in general. However, it was shown in the work of DPS [11] that we can use an approximation to arrive at

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{y}|\boldsymbol{x}_t) \simeq \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)),$$

where

$$\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) := \frac{1}{\sqrt{\bar{\alpha}(t)}}(\boldsymbol{x}_t + (1 - \bar{\alpha}(t))\boldsymbol{s}_{\theta^*}(\boldsymbol{x}_t, t)) \quad (8)$$

is the denoised estimate of $\boldsymbol{x}_t$ in the VP-SDE context given by the Tweedie's formula [17]. Hence, one can use the following tractable reverse SDE to sample from the posterior distribution

$$\begin{aligned}d\boldsymbol{x} = (-\frac{\beta(t)}{2}\boldsymbol{x} - \beta(t)[\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t))\\ + \boldsymbol{s}_{\theta^*}(\boldsymbol{x}_t, t)])dt + \sqrt{\beta(t)}d\bar{\boldsymbol{w}}, \quad (9)\end{aligned}$$

where we observe that $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t))$ can be efficiently computed using analytical likelihood, and backpropagation through the score function, i.e.

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{y}) \simeq \boldsymbol{s}_{\theta^*}(\boldsymbol{x}_t) - \frac{1}{\sigma^2}\nabla_{\boldsymbol{x}_t}\|\boldsymbol{y} - \mathcal{H}(\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t))\|_2^2.$$

However, one should note that the method in (9) is only applicable when the forward model $\mathcal{H}$ is fixed, and hence cannot be directly used for solving *blind* inverse problems.

Blind inverse problem Blind inverse problems consider the case where the forward model $\mathcal{H}$ is unknown. Among them, we focus on the case where the forward operator is parameterized with $\boldsymbol{\varphi}$, and we need to estimate the *parameter* $\boldsymbol{\varphi}$. Specifically, consider the following forward model

$$\boldsymbol{y} = \mathcal{H}_{\boldsymbol{\varphi}}(\boldsymbol{x}) + \boldsymbol{n}, \quad (10)$$

where $\boldsymbol{\varphi}$ is the parameter of the forward model, $\boldsymbol{x}$ is the ground truth image, and $\boldsymbol{n}$ is some noise. Here, both $\boldsymbol{\varphi}, \boldsymbol{x}$ are unknown, and should be estimated. A classical way to solve (10) is to optimize for the following

$$\min_{\boldsymbol{x},\boldsymbol{\varphi}} \quad \frac{1}{2}\|\mathcal{H}_{\boldsymbol{\varphi}}(\boldsymbol{x}) - \boldsymbol{y}\|^2 + R_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}) + R_{\boldsymbol{x}}(\boldsymbol{x}), \quad (11)$$

where $R_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}), R_{\boldsymbol{x}}(\boldsymbol{x})$ are regularization functions for $\boldsymbol{\varphi}, \boldsymbol{x}$, respectively, which can also be thought of as the negative log prior for each distribution, e.g. $R(\cdot) = -\log p(\cdot)$.

For example, consider blind deconvolution from camera motion blur as illustrated in Fig. 3(a). The forward model reads

$$\boldsymbol{y} = \boldsymbol{k} * \boldsymbol{x} + \boldsymbol{n}, \quad (12)$$

where $\boldsymbol{k}$ is the blur kernel, corresponding to the parameter $\boldsymbol{\varphi}$. On the other hand, although the "real" forward model

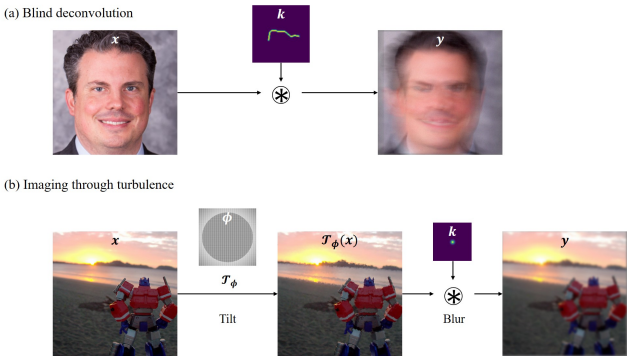

Figure 3. Illustration of the imaging forward model. (a) Blind deconvolution, (b) Imaging through turbulence

for atmospheric turbulence is rarely directly used in practice due to the highly complicated nature of the wave propagation theory, the tilt-blur model is often used [5, 6, 49], as the model is simple but fairly accurate. Specifically, the visualization of such imaging process is shown in Fig. 3(b), which can be mathematically described by

$$\boldsymbol{y} = \boldsymbol{k} * \mathcal{T}_{\boldsymbol{\phi}}(\boldsymbol{x}) + \boldsymbol{n}, \tag{13}$$

where $\mathcal{T}$ is the tilt operator parameterized by the tilt vector field $\boldsymbol{\phi}$. To remove the scale ambiguity between the kernel and image, the magnitude and the polarity constraints of kernels are often used:

$$\mathbf{1}^T\boldsymbol{k} = 1, \boldsymbol{k} \succeq 0. \tag{14}$$

Then, the success of the optimization algorithm (11) with the forward models (12) or (13) under the constraint (14) depends on two factors: 1) How closely the prior-imposing functions $R_{\{\boldsymbol{x},\boldsymbol{k}\}}$ estimate the true prior, and 2) how well the optimization procedure finds the minimum value. Conventional methods are sub-optimal in both aspects. First, the prior (e.g. sparsity [41], dark channel [44], implicit from deep networks [47]) functions do not fully represent the *true* prior. Second, the optimization process is unstable and hard to tune. For instance, [41, 44] requires different weighting parameters *per image*, and often fails during the abrupt changes in the stage transition during coarse-to-fine optimization strategy. In section 3, we show that our method can solve both of these problems.

3. BlindDPS

In DPS [11], the authors used the diffusion prior for $R_{\boldsymbol{x}}$ by training a score function that models $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$. As for blind inverse problems, a prior model for the parameter $p(\boldsymbol{\varphi})$ should also be specified. In this regard, our proposal is to use the diffusion prior also for the forward model parameter by estimating $\nabla_{\boldsymbol{\varphi}} \log p(\boldsymbol{\varphi})$. With such choice, one can model a much more accurate prior for the parameters compared to the conventional choices. In the following, we detail on how to build our method *BlindDPS*, focusing on blind deconvolution. The method for imaging through turbulence can be derived in a completely analogous fashion, where the details can be found in Supplementary section B.1.

Key idea. In blind deblurring (deconvolution), the probabilistic forward model is specified as follows

$$p(\boldsymbol{y}|\boldsymbol{x}_0, \boldsymbol{k}_0) := \mathcal{N}(\boldsymbol{y}|\boldsymbol{k}_0 * \boldsymbol{x}_0, \sigma^2\boldsymbol{I}), \tag{15}$$

where $\boldsymbol{k}_0$ is the random variable of the convolution kernel. As $\boldsymbol{x}_0$ and $\boldsymbol{k}_0$ are independent, the posterior probability is given as

$$p(\boldsymbol{x}_0, \boldsymbol{k}_0|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x}_0, \boldsymbol{k}_0)p(\boldsymbol{x}_0)p(\boldsymbol{k}_0). \tag{16}$$

Note that our aim is to use implicit diffusion priors for both $p(\boldsymbol{x}_0)$ and $p(\boldsymbol{k}_0)$ through their score functions. One can easily take pre-trained score functions for the image. Similarly, the score function for the kernel can also be estimated from standard DSM (3) to get $\boldsymbol{s}^k_{\theta^*}(\boldsymbol{k}, t) \simeq \nabla_{\boldsymbol{k}_t} \log p_t(\boldsymbol{k}_t)$. Note that performing DSM to achieve $\boldsymbol{s}^k_{\theta^*}$ costs much less than training the image score function $\boldsymbol{s}^i_{\theta^*}$, as the distribution is much simpler, and the dimensionality of the vector $\boldsymbol{k}$ is also sufficiently smaller than $\boldsymbol{x}$.

On the other hand, again from the independence of $\boldsymbol{x}_0$ and $\boldsymbol{k}_0$, we are able to construct two separate reverse diffusion processes of identical form:

$$d\boldsymbol{x} = \left[-\frac{\beta(t)}{2}\boldsymbol{x} - \beta(t)\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)\right] dt + \sqrt{\beta(t)}d\bar{\boldsymbol{w}},$$
$$d\boldsymbol{k} = \left[-\frac{\beta(t)}{2}\boldsymbol{k} - \beta(t)\nabla_{\boldsymbol{k}_t} \log p(\boldsymbol{k}_t)\right] dt + \sqrt{\beta(t)}d\bar{\boldsymbol{w}}.$$

Note that the two reverse SDEs are only able to sample from the marginals — $p(\boldsymbol{x}_0)$, $p(\boldsymbol{k}_0)$. However, one can define the dependency between $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{k}$ from the posterior probability. Using Bayes' rule in (16) for general t, we have

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t, \boldsymbol{k}_t|\boldsymbol{y}) = \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t, \boldsymbol{k}_t) + \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t),$$
$$\nabla_{\boldsymbol{k}_t} \log p(\boldsymbol{x}_t, \boldsymbol{k}_t|\boldsymbol{y}) = \nabla_{\boldsymbol{k}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t, \boldsymbol{k}_t) + \nabla_{\boldsymbol{k}_t} \log p(\boldsymbol{k}_t).$$

Here, in order to estimate the time-conditional log-likelihood $\log p(\boldsymbol{y}|\boldsymbol{x}_t, \boldsymbol{k}_t)$ which is intractable in general, we need the following result:

Theorem 1. *Under the same conditions in [11], we have*

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{y}|\boldsymbol{x}_t, \boldsymbol{k}_t) \simeq \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t), \hat{\boldsymbol{k}}_0(\boldsymbol{k}_t))$$
$$\nabla_{\boldsymbol{k}_t} \log p_t(\boldsymbol{y}|\boldsymbol{x}_t, \boldsymbol{k}_t) \simeq \nabla_{\boldsymbol{k}_t} \log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t), \hat{\boldsymbol{k}}_0(\boldsymbol{k}_t)).$$

Remark 1. *Our theorem holds as long as $\boldsymbol{x}_t$, $\boldsymbol{k}_t$ are independent. Note that the theorem can be further generalized to handle more random variables whenever the independence between the variables is established. In other*

words, we can construct arbitrary many diffusion procedures for each component of the forward model, which can be solved analogous to the approximation proposed in Theorem 1. This result will be useful when we solve the problem of imaging through turbulence in in Supplementary section B.1.

Using Theorem 1, we finally arrive at the following reverse SDEs

$$d\boldsymbol{x} = (-\frac{\beta(t)}{2}\boldsymbol{x} - \beta(t)[\nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t), \hat{\boldsymbol{k}}_0(\boldsymbol{k}_t)) + \boldsymbol{s}^i_{\theta^*}(\boldsymbol{x}_t, t)])dt + \sqrt{\beta(t)}d\bar{\boldsymbol{w}}, \tag{17}$$

$$d\boldsymbol{k} = (-\frac{\beta(t)}{2}\boldsymbol{k} - \beta(t)[\nabla_{\boldsymbol{k}_t}\log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t), \hat{\boldsymbol{k}}_0(\boldsymbol{k}_t)) + \boldsymbol{s}^k_{\theta^*}(\boldsymbol{k}_t, t)])dt + \sqrt{\beta(t)}d\bar{\boldsymbol{w}}. \tag{18}$$

The system of equations (17),(18) are now numerically solvable as the gradient of the log likelihood is analytically tractable. Specifically, for the Gaussian measurement, we have

$$\nabla_{\boldsymbol{x}_t}\log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{k}}_0) = -\frac{1}{\sigma^2}\nabla_{\boldsymbol{x}_t}\|\boldsymbol{y} - \hat{\boldsymbol{k}}_0 * \hat{\boldsymbol{x}}_0\|_2^2. \tag{19}$$

Combined with the ancestral sampling steps [21], our algorithm for posterior sampling of blind deblurring is formally given in Algorithm 1. Here, note that we choose to take static step size times the gradient of the norm instead of taking time-dependent step sizes times the gradient of the squared norm, as it was shown to be effective despite its simplicity [11]. Furthermore, in order to impose the usual condition (14), we define a set $C := \{\boldsymbol{k}|\mathbf{1}^T\boldsymbol{k} = 1, \boldsymbol{k} \succeq 0\}$, and project onto the set through $\mathcal{P}_C(\hat{\boldsymbol{k}}_0)$ in Algorithm 1, after the estimation of $\hat{\boldsymbol{k}}_0$ at each intermediate step. For visual illustration of the proposed method, see Fig. 2.

Augmenting diffusion prior with sparsity. Implementing (17),(18) directly induces fairly stable results with the correct choice of α. Here, we go a step further and adopt a lesson from the classic literature. As we often wish to estimate blur kernels that are sparse, we promote sparsity *only* to the kernel that we are estimating by augmenting the diffusion prior with ℓ_0/ℓ_1 regularization. The minimization strategy for the kernel then becomes

$$\boldsymbol{k}_{i-1} = \boldsymbol{k}'_{i-1} - \alpha\left(\|\boldsymbol{y} - \hat{\boldsymbol{k}}_0 * \hat{\boldsymbol{x}}_0\|_2 + \lambda R_{\boldsymbol{k}}(\hat{\boldsymbol{k}}_0)\right), \tag{20}$$

where λ is the regularization strength, and the choice of $R_{\boldsymbol{k}}(\cdot) := \ell_0/\ell_1$ regularization depends on the type of the dataset. With such augmentation, reconstruction can be further stabilized.

Interpretation in Gaussian scale-space. (Gaussian) Scale-space theory [36] states that one can represent signals in multiple scales by gradually convolving with Gaussian filters. As adding Gaussian noise to random vectors in the forward pass of the diffusion has a dual relation in the density domain (i.e. convolution with Gaussian kernels), one can think of the diffusion process as a realization of one such process. Thus, the reverse diffusion process can be interpreted as a coarse-to-fine synthesis evolving through the Gaussian scale-space, which is most visible by visualizing $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t), \hat{\boldsymbol{k}}_0(\boldsymbol{k}_t)$ when evolving through $t = 1 \to 0$ (see Fig. 1(c)).

For blind deconvolution problems, in order to achieve optimal quality, it is a standard practice to start the optimization process at a coarse scale by down-sampling, and sequentially upsample with a pre-determined schedule to refine the estimates [41, 42]. However, the discretized schedule is typically abrupt (e.g. [41, 42] uses 8 discretization) and ad-hoc. On the other hand, by using the reverse diffusion process, we are granted with a natural, smooth schedule of evolution, which can be thought of as a continuous generalization of the coarse-to-fine reconstruction strategy. This could be another reason why the proposed method is able to dramatically outperform the conventional methods.

Algorithm 1 BlindDPS — Blind Deblurring

Require: $N, \boldsymbol{y}, \alpha, \{\tilde{\sigma}_i\}_{i=1}^N, \lambda, R_{\boldsymbol{k}}(\cdot)$
$\boldsymbol{x}_N, \boldsymbol{k}_N \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
for $i = N-1$ **to** 0 **do**
  $\hat{\boldsymbol{s}}^i \leftarrow \boldsymbol{s}^i_{\theta^*}(\boldsymbol{x}_i, i)$
  $\hat{\boldsymbol{s}}^k \leftarrow \boldsymbol{s}^k_{\theta^*}(\boldsymbol{k}_i, i)$
  $\hat{\boldsymbol{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\boldsymbol{x}_i + \sqrt{1-\bar{\alpha}_i}\hat{\boldsymbol{s}}^i)$
  $\hat{\boldsymbol{k}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\boldsymbol{k}_i + \sqrt{1-\bar{\alpha}_i}\hat{\boldsymbol{s}}^k)$
  $\hat{\boldsymbol{k}}_0 \leftarrow \mathcal{P}_C(\hat{\boldsymbol{k}}_0)$
  $\boldsymbol{z}_i, \boldsymbol{z}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
  $\boldsymbol{x}'_{i-1} \leftarrow \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\boldsymbol{x}_i + \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\hat{\boldsymbol{x}}_0 + \tilde{\sigma}_i\boldsymbol{z}_i$
  $\boldsymbol{k}'_{i-1} \leftarrow \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\boldsymbol{k}_i + \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\hat{\boldsymbol{k}}_0 + \tilde{\sigma}_i\boldsymbol{z}_k$
  $\boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}'_{i-1} - \alpha\nabla_{\boldsymbol{x}_i}\|\boldsymbol{y} - \hat{\boldsymbol{k}}_0 * \hat{\boldsymbol{x}}_0\|_2$
  $\mathcal{L}_{\boldsymbol{k}} \leftarrow \|\boldsymbol{y} - \hat{\boldsymbol{k}}_0 * \hat{\boldsymbol{x}}_0\|_2 + \lambda R_{\boldsymbol{k}}(\hat{\boldsymbol{k}}_0)$
  $\boldsymbol{k}_{i-1} \leftarrow \boldsymbol{k}'_{i-1} - \alpha\nabla_{\boldsymbol{k}_i}\mathcal{L}_{\boldsymbol{k}}$
end for
return $\boldsymbol{x}_0, \boldsymbol{k}_0$

4. Experiments

4.1. Experimental setup

Dataset. For blind deblurring, we use FFHQ 256×256 [26], and AFHQ-dog 256×256 [10]. We choose 1k validation set for FFHQ, and the full 500 test images for AFHQ-dog. We leverage pre-trained (image) score functions, as in the experimental setting of [13]. For imaging through turbulence, we use FFHQ 256×256 and ImageNet 256×256 [15]. For both blind inverse problems, we add Gaussian measurement noise with $\sigma = 0.02$. Full details on

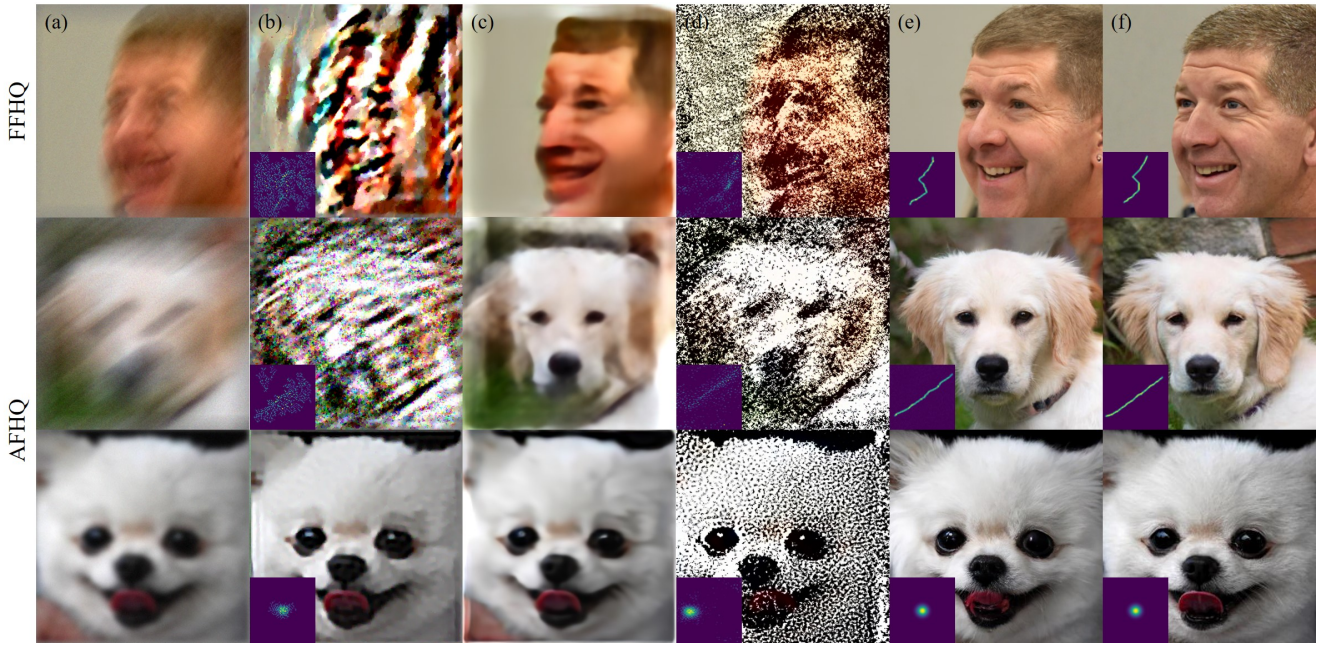

Figure 4. Blind deblurring results. (row 1): FFHQ 256×256 motion deblurring, (row 2): AFHQ 256×256 motion deblurring. (row 3): AFHQ 256×256 Gaussian deblurring. (a) Measurement, (b) Pan-DCP [44], (c) MPRNet [61], (d) SelfDeblur [47], (e) BlindDPS (**ours**), (f) Ground truth. For (c), kernel not shown as the method only estimate images.

	FFHQ (256 × 256)						AFHQ (256 × 256)					
	Motion			Gaussian			Motion			Gaussian		
Method	FID ↓	LPIPS ↓	PSNR ↑	FID ↓	LPIPS ↓	PSNR ↑	FID ↓	LPIPS ↓	PSNR ↑	FID ↓	LPIPS ↓	PSNR ↑
BlindDPS (ours)	**29.49**	**0.281**	**22.24**	**27.36**	**0.233**	**24.77**	**23.89**	**0.338**	**20.92**	**20.54**	**0.287**	**23.63**
SelfDeblur [47]	270.0	0.717	10.83	235.4	0.686	11.36	300.5	0.768	9.081	172.2	0.662	11.53
MPRNet [61]	111.6	0.434	17.40	95.12	0.337	20.75	131.8	0.521	14.85	46.19	0.366	20.51
DeblurGANv2 [32]	220.7	0.571	17.75	185.5	0.529	19.69	186.2	0.597	17.64	86.87	0.523	20.29
Pan-DCP [44]	214.9	0.520	15.41	92.70	0.393	20.50	214.0	0.704	11.87	57.14	0.392	20.97
Pan-ℓ_0 [41]	242.6	0.542	15.53	109.1	0.415	19.94	235.0	0.627	15.34	62.76	0.395	21.41
Perrone *et al.* [45]	156.8	0.492	16.08	85.3	0.363	20.66	197.7	0.588	16.68	59.32	0.375	21.38

Table 1. Quantitative evaluation (FID, LPIPS, PSNR) of blind deblurring task on FFHQ and AFHQ. **Bold**: Best, under: second best.

	FFHQ (256 × 256)				AFHQ (256 × 256)			
	Motion		Gaussian		Motion		Gaussian	
Method	MSE ↓	MNC ↑	MSE ↓	MNC ↑	MSE ↓	MNC ↑	MSE ↓	MNC ↑
BlindDPS (ours)	**0.003**	**0.955**	**0.000**	**0.995**	**0.003**	**0.930**	**0.001**	**0.991**
SelfDeblur [47]	0.021	0.323	0.020	0.266	0.021	0.268	0.020	0.272
Pan-DCP [44]	0.020	0.425	0.016	0.478	0.020	0.365	0.016	0.481
Pan-ℓ_0 [41]	0.020	0.454	0.016	0.518	0.020	0.398	0.015	0.517

Table 2. Quantitative evaluation (MSE, MNC [23]) of kernel estimation on FFHQ and AFHQ. **Bold**: Best, under: second best.

experimental setup can be found in supplementary section F

Evaluation. We use three metrics—Frechet inception distance (FID), learned Perceptual Image Patch Similarity (LPIPS), and peak signal-to-noise-ratio (PSNR)—for quantitatively measuring the performance of the image reconstruction. For kernel estimation, we use mean-squared-error (MSE), and maximum of normalized convolution (MNC) [23] , which is computed by

$$\text{MNC} := \max\left(\frac{\tilde{\boldsymbol{k}} * \boldsymbol{k}^*}{\|\tilde{\boldsymbol{k}}\|_2 \|\boldsymbol{k}^*\|_2}\right), \tag{21}$$

where $\tilde{\boldsymbol{k}}, \boldsymbol{k}^*$ are the estimated, and the ground truth kernels, respectively.

4.2. Results

Blind deblurring. Motion deblurring results are presented in Fig. 1(a) and Fig. 4. As our setting for motion deblurring imposes a rather aggressive degradation with a large blur kernel, most of the prior arts fail catastrophically, not being able to generate a feasible solution. In contrast, our method accurately captures both the kernel and the image with sharpness. Similar trend can be seen for Gaussian deblurring presented in the third row of Fig. 4. Other meth-

Method	FFHQ (256 × 256)			ImageNet (256 × 256)		
	FID ↓	LPIPS ↓	PSNR ↑	FID ↓	LPIPS ↓	PSNR ↑
BlindDPS (ours)	**27.35**	**0.247**	24.49	**51.25**	**0.341**	19.59
TSR-WGAN [25]	58.30	0.258	**26.29**	69.80	0.369	17.67
ILVR [9]	65.50	0.370	21.48	85.21	0.494	18.09
MPRNet [61]	116.2	0.411	19.68	78.24	0.421	20.34
DeblurGANv2 [32]	225.9	0.561	18.40	60.31	0.393	**21.56**

Table 3. Quantitative evaluation (FID, LPIPS, PSNR) of imaging through turbulence task on FFHQ and ImageNet. **Bold**: Best, under: second best.

ods fall far short of BlindDPS in the sense that they either produce reconstructions that are blurry with inaccurate blur kernel estimation, or fails dramatically (e.g. SelfDeblur). Furthermore, the proposed method establishes the state-of-the-art in all quantitative metrics, which can be seen in Table 1 and Table 2.

Figure 5. Blind deblurring results on diverse data and kernel type/size.(a)ImageNet, (b)USC-SIPI, (c) kernels from Levin *et al.* [34]

Diverse images and kernels. To demonstrate the generalizability of BlindDPS, we apply our method to more diverse images and kernels in Fig. 5. We see that BlindDPS indeed yields high-quality reconstructions in all the presented cases. Notably, when smaller sized kernel is used, we can simply center-crop the inferred kernel after the inference without any additional treatment.

Imaging through turbulence. We show the reconstruction results in Fig. 1(b) and Fig. 6, with quantitative metrics in Table 3. Consistent with the results from blind deblurring, BlindDPS outperforms the comparison methods in most cases, effectively removing both the blur and the tilt from the measurement. Notably, our method outperforms *all* other methods by large margins on perceptual metrics (i.e. FID, LPIPS). For PSNR, the proposed method often slightly underperforms against supervised learning approaches, which is to be expected, as for reconstructions from heavy degradations, retrieving the high-frequency details often penalizes such distortion metrics [3].

4.3. Ablation studies

We perform two ablation studies to verify our design choices: 1) using the diffusion prior for the forward model parameters, and 2) augmenting the diffusion prior with the sparsity prior. Details on the experimental setup along with further analysis can be found in Supplementary section C.

Diffusion prior for the forward model. One may question why the score function for the kernel is necessary in the first place, since one could also estimate the kernel solely through gradient descent using the gradient of the likelihood. In fact, this corresponds to using the uniform prior for the kernel distribution, which we compare against the proposed diffusion prior (BlindDPS) in Fig. 7. We clearly see that using the uniform prior yields heavily distorted result, with poorly estimated kernel. From this experiment, we observe that using another diffusion process specifically for the forward model is crucial for the performance.

Effect of sparsity regularization. One design choice made in BlindDPS is the additional sparsity regularization applied to kernels. Here, we analyze the effect of such regularization. In Table C.1, we report on quantitative metrics for the kernel, depending on the regularization weight λ. Clearly, setting $\lambda = 0.0$ induces inferior performance especially for motion deblurring. When setting $\lambda \geq 0.1$ however, we can see that one can achieve good performance regardless of the chosen weight value. As diffusion priors have been shown to have surprisingly high generalization capacity [14, 24], we choose a mild weight value of $\lambda = 1.0$, which gives visually appealing results without down-weighting the influence of diffusion priors too much.

5. Discussion and Related Works

This work follows the line of endeavors to develop methods that can solve inverse problems through diffusion models. Methods that are based on iterative projection onto convex sets (POCS) were the first to be developed, iterating between the denoising step, and the projection step [9, 13, 14, 52, 53]. Methods that attempt to approximate posterior sampling via annealed Langevin dynamics (ALD) [24], and singular value decomposition (SVD) [27] were proposed, with the latter showing particular robustness to noisy measurements.

The trend recently shifted towards leveraging the denoised estimate via Tweedie's formula at the inter-

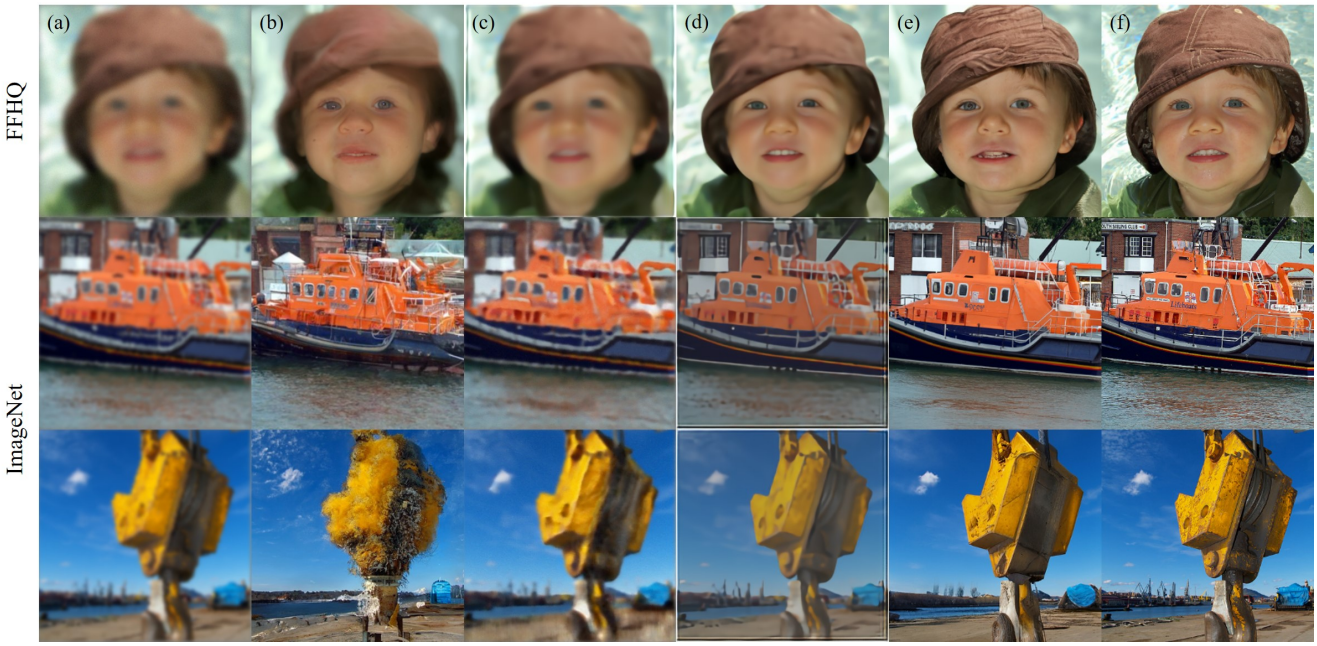

Figure 6. Reconstruction of imaging through turbulence. (row 1): FFHQ 256×256, (row 2-3): ImageNet 256×256. (a) Measurement, (b) ILVR [9], (c) MPRNet [61], (d) TSR-WGAN [25], (e) BlindDPS (**ours**), (f) Ground truth.

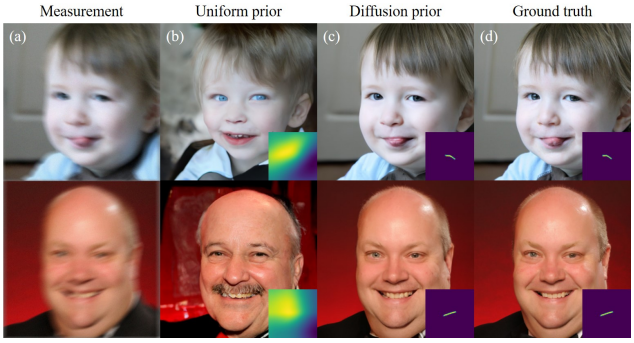

Figure 7. Ablation study: uniform prior vs. diffusion prior. (a) Measurement, (b) uniform prior, (c) diffusion prior, (d) ground truth.

mediate steps under various names — manifold constrained gradient (MCG) [12], gradient guidance [22], and reconstruction-based method [28]. Diffusion posterior sampling (DPS) [11] is the method that is the most similar to ours, showing that such method is an approximation of the posterior sampling process. However, none of the methods so far considered blind inverse problem, and to the best of our knowledge, we are the first to show that posterior sampling with diffusion scales to blind settings.

Limitations and future directions. As BlindDPS performs joint minimization on multiple factors (e.g. kernel, tilt-map, image), it is typically less robust than the non-blind reconstruction scheme. At times, the solution diverges when the parameters are incorrectly tuned. Such instability when we experiment with datasets with high diversity e.g. ImageNet. As we find ImageNet diffusion prior to be volatile, it is tricky to scale our method to use for out-in-the-wild blurred images, and limit the scope of this work to simulated ones. For imaging through turblence, it is often the case where the tilt map is incorrectly estimated whereas the kernel and the ground truth image are accurately estimated. Furthermore, as we train and use specified diffusion score functions for each of the component, inference speed is delayed, due to the additional forward/backward passes through the newly involved score functions. When the forward functional involves estimating additional parameters, the number of score functions required will scale linearly, not being efficient with complex functional forms. Finally, we note that our method is yet to solve the *truly blind* case, where we do not know the functional form of the forward mapping. Solving the truly blind case would be an interesting direction of future studies.

6. Conclusion

In this work, we proposed BlindDPS, a framework for solving blind inverse problems by jointly estimating the parameters of the forward measurement operator and the image to be reconstructed. We theoretically show how we can construct a system of reverse SDEs to approximate posterior sampling for blind inverse problems, by using multiple score functions designed to estimate each part of the component. With extensive experiments, we show that BlindDPS establishes state-of-the-art performance on both blind deblurring and imaging through turbulence, even when the degradation and the measurement noise are heavy.

References

[1] Nantheera Anantrasirichai, Alin Achim, Nick G Kingsbury, and David R Bull. Atmospheric turbulence mitigation using complex wavelet-based fusion. *IEEE Transactions on Image Processing*, 22(6):2398–2408, 2013. 18

[2] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. 2

[3] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6228–6237, 2018. 7

[4] 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. 17

[5] Wai Ho Chak, Chun Pong Lau, and Lok Ming Lui. Subsampled turbulence removal network. *arXiv preprint arXiv:1807.04418*, 2018. 4, 15, 18

[6] Stanley H Chan. Tilt-then-blur or blur-then-tilt? clarifying the atmospheric turbulence model. *IEEE Signal Processing Letters*, 29:1833–1837, 2022. 4, 15, 18

[7] Tony F Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE transactions on Image Processing*, 7(3):370–375, 1998. 17

[8] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021. 17

[9] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning method for denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 7, 8, 19, 20, 25, 26

[10] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8188–8197, 2020. 5, 18

[11] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3, 4, 5, 8, 12, 13, 18, 19

[12] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems*, 2022. 8

[13] Hyungjin Chung, Byeongsu Sim, and Jong Chul 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. 2, 5, 7, 18

[14] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated MRI. *Medical Image Analysis*, page 102479, 2022. 7

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5, 18

[16] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. 18

[17] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. 3

[18] David L Fried. Probability of getting a lucky short-exposure image through turbulence. *JOSA*, 68(12):1651–1658, 1978. 18

[19] Xiang Gao, Meera Sitharam, and Adrian E Roitberg. Bounds on the jensen gap, and implications for mean-concentrated distributions. *arXiv preprint arXiv:1712.05267*, 2017. 12

[20] Yiming Guo, Xiaoqing Wu, Chun Qing, Changdong Su, Qike Yang, and Zhiyuan Wang. Blind restoration of images distorted by atmospheric turbulence based on deep transfer learning. In *Photonics*, volume 9, page 582. MDPI, 2022. 18

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020. 2, 5

[22] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *arXiv preprint arXiv:2204.03458*, 2022. 8

[23] Zhe Hu and Ming-Hsuan Yang. Good regions to deblur. In *European conference on computer vision*, pages 59–72. Springer, 2012. 6

[24] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. In *Advances in Neural Information Processing Systems*, volume 34, pages 14938–14954, 2021. 7

[25] Darui Jin, Ying Chen, Yi Lu, Junzhang Chen, Peng Wang, Zichao Liu, Sheng Guo, and Xiangzhi Bai. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. *Nature Machine Intelligence*, 3(10):876–884, 2021. 7, 8, 18, 19, 20, 25, 26

[26] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 5, 18

[27] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *ICLR Workshop on Deep Generative Models for Highly Structured Data*, 2022. 2, 7

[28] Bahjat Kawar, Jiaming Song, Stefano Ermon, and Michael Elad. Jpeg artifact correction using denoising diffusion restoration models. *arXiv preprint arXiv:2209.11888*, 2022. 8

[29] Jan J Koenderink. The structure of images. *Biological cybernetics*, 50(5):363–370, 1984. 2

[30] Dilip Krishnan, Terence Tay, and Rob Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR 2011*, pages 233–240. IEEE, 2011. 17

[31] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 17

[32] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019. 6, 7, 17, 19

[33] Brett Levac, Ajil Jalal, and Jonathan I Tamir. Accelerated motion correction for mri using score-based generative models. *arXiv preprint arXiv:2211.00199*, 2022. 16

[34] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971. IEEE, 2009. 7, 17

[35] Lerenhan Li, Jinshan Pan, Wei-Sheng Lai, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Learning a discriminative prior for blind image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6616–6625, 2018. 17

[36] Tony Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270, 1994. 2, 5

[37] Ce Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009. 18

[38] Junmin Liu, Peipei Wang, Xiaoke Zhang, Yanliang He, Xinxing Zhou, Huapeng Ye, Ying Li, Shixiang Xu, Shuqing Chen, and Dianyuan Fan. Deep learning based atmospheric turbulence compensation for orbital angular momentum beam distortion and communication. *Optics express*, 27(12):16671–16688, 2019. 18

[39] Zhiyuan Mao, Nicholas Chimitt, and Stanley H Chan. Image reconstruction of static and dynamic scenes through anisoplanatic turbulence. *IEEE Transactions on Computational Imaging*, 6:1415–1428, 2020. 18

[40] Zhiyuan Mao, Ajay Jaiswal, Zhangyang Wang, and Stanley H Chan. Single frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model. *arXiv preprint arXiv:2207.10040*, 2022. 18

[41] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. l_0-regularized intensity and gradient prior for deblurring text images and beyond. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):342–355, 2016. 2, 4, 5, 6, 17, 19

[42] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1628–1636, 2016. 5, 17

[43] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 17

[44] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Deblurring images via dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 40(10):2315–2328, 2017. 2, 4, 6, 17, 19, 21, 22, 23, 24

[45] Daniele Perrone and Paolo Favaro. Total variation blind deconvolution: The devil is in the details. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2909–2916, 2014. 6, 17, 19

[46] Shyam Nandan Rai and CV Jawahar. Removing atmospheric turbulence via deep adversarial learning. *IEEE Transactions on Image Processing*, 31:2633–2646, 2022. 18

[47] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3341–3350, 2020. 4, 6, 17, 19, 21, 22, 23, 24

[48] Christian J Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Learning to deblur. *IEEE transactions on pattern analysis and machine intelligence*, 38(7):1439–1451, 2015. 17

[49] Masao Shimizu, Shin Yoshimura, Masayuki Tanaka, and Masatoshi Okutomi. Super-resolution from image sequence under influence of hot-air optical turbulence. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 4, 15

[50] Slavko Simic. On a global upper bound for jensen's inequality. *Journal of mathematical analysis and applications*, 343(1):414–419, 2008. 12

[51] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023. 2

[52] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, 2021. 7

[53] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *9th International Conference on Learning Representations, ICLR*, 2021. 2, 7

[54] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 769–777, 2015. 17

[55] Libin Sun, Sunghyun Cho, Jue Wang, and James Hays. Edge-based blur kernel estimation using patch priors. In *IEEE international conference on computational photography (ICCP)*, pages 1–8. IEEE, 2013. 2

[56] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. 2

[57] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022. 17

[58] Yuan Xie, Wensheng Zhang, Dacheng Tao, Wenrui Hu, Yanyun Qu, and Hanzi Wang. Removing turbulence effect via hybrid total variation and deformation-guided kernel regression. *IEEE Transactions on Image Processing*, 25(10):4943–4958, 2016. 18

[59] Xiangyu Xu, Jinshan Pan, Yu-Jin Zhang, and Ming-Hsuan Yang. Motion blur kernel estimation via deep learning. *IEEE Transactions on Image Processing*, 27(1):194–205, 2017. 17

[60] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 17

[61] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 6, 7, 8, 17, 19, 21, 22, 23, 24, 25, 26

[62] Xiang Zhu and Peyman Milanfar. Removing atmospheric turbulence via space-invariant deconvolution. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):157–170, 2012. 18

[63] Wangmeng Zuo, Deyu Meng, Lei Zhang, Xiangchu Feng, and David Zhang. A generalized iterated shrinkage algorithm for non-convex sparse coding. In *Proceedings of the IEEE international conference on computer vision*, pages 217–224, 2013. 17

[64] Wangmeng Zuo, Dongwei Ren, David Zhang, Shuhang Gu, and Lei Zhang. Learning iteration-wise generalized shrinkage–thresholding operators for blind deconvolution. *IEEE Transactions on Image Processing*, 25(4):1751–1764, 2016. 17