

INDIGO+: A Unified INN-Guided Probabilistic Diffusion Algorithm for Blind and Non-Blind Image Restoration

Di You¹, Graduate Student Member, IEEE, and Pier Luigi Dragotti², Fellow, IEEE

Abstract—Generative diffusion models are becoming one of the most popular prior in image restoration (IR) tasks due to their remarkable ability to generate realistic natural images. Despite achieving satisfactory results, IR methods based on diffusion models present several limitations. First of all, most non-blind approaches require an analytical expression of the degradation model to guide the sampling process. Secondly, most existing blind approaches rely on families of pre-defined degradation models for training their deep networks. The above issues limit the flexibility of these approaches and so their ability to handle real-world degradation tasks. In this paper, we propose a novel INN-guided probabilistic diffusion algorithm for non-blind and blind image restoration, namely INDIGO and BlindINDIGO, which combines the merits of the perfect reconstruction property of invertible neural networks (INN) with the strong generative capabilities of pre-trained diffusion models. Specifically, we train the forward process of the INN to simulate an arbitrary degradation process and use the inverse to obtain an intermediate image that we use to guide the reverse diffusion sampling process through a gradient step. We also introduce an initialization strategy, to further improve the performance and inference speed of our algorithm. Experiments demonstrate that our algorithm obtains competitive results compared with recently leading methods both quantitatively and visually on synthetic and real-world low-quality images.

Index Terms—Blind image restoration, diffusion models, image restoration, invertible neural networks.

I. INTRODUCTION

In this paper, we explore a new way to employ diffusion models for image restoration. Image restoration (IR) is a typical inverse problem aiming to recover high-quality images from their noisy and degraded measurements. In a typical restoration problem, one observes $\mathbf{y} = \mathcal{H}(\mathbf{x}, \mathbf{n})$, where $\mathbf{y}$ is the degraded and noisy version of the original image $\mathbf{x}$ and $\mathbf{n}$ is some noise. The degradation process $\mathcal{H}$ can be linear or non-linear and is often unknown. In this paper, we classify these inverse problems as blind or non-blind IR problems based on whether we can

access labeled training pairs that are degraded in the same way as the measurement (fully supervised setting) or not. Specifically, the non-blind case includes two situations: either we know the expression of the forward operator, or we can simulate it without knowing its analytical expression and can therefore produce labeled image pairs. In the blind case, we do not have access to the labeled data and therefore do not have any way to simulate the actual degradation process.

This inverse problem is normally solved by addressing the classic trade-off between a data fidelity and a regularization term based on proper priors. While this approach goes back to the classic Tikhonov regularization, a wealth of new models and regularizers have emerged over the years often driven by the idea of sparsity. These have led to the development of many model-based reconstruction methods.

Recently, there has been a shift towards developing data-driven approaches, in particular based on deep learning architectures where the regularization is implicitly learned through the data and we refer to [5] for a recent review on the topic. The plug-and-play (PnP) framework [6] is a typical example where the prior can be learned through data. PnP is based on iterating between a step that enforces some forms of data consistency and a denoising step where the denoiser can be implemented with a deep neural network. In this context, the denoiser effectively acts as a data-driven regularizer. Interestingly, this heuristic has led to many remarkable results and we refer to [7] for a recent overview on the topic.

The generative prior of diffusion models [8], [9], [10], [11], [12] has now become one of the most popular priors in image restoration problems due to their remarkable ability to approximate the natural image manifold. A line of work [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] has focused on leveraging the rich image priors and strong generative capability of pretrained diffusion models to solve IR problems. Among them, earlier works [13], [14], [15], [16] have focused on linear degradation models and noiseless measurements. Two popular categories of approaches have then been proposed for investigating noisy and non-linear inverse problems. Decomposition-based approaches [17], [18], [19], [20] run singular value decomposition (SVD), range-null space decomposition or matrix decomposition on intermediate results during iterations to guide the sampling process. Gradient-based approaches [21], [22], [23], [24] propose to incorporate consistency-imposing gradient steps in between the reverse diffusion steps.

Received 9 January 2024; revised 7 June 2024; accepted 26 August 2024. Date of publication 9 September 2024; date of current version 15 January 2025. The work of Di You was supported by the President's Ph.D. Scholarship from Imperial College London. The guest editor coordinating the review of this article and approving it for publication was Dr. Zhihui Zhu. (Corresponding author: Di You.)

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: di.you22@imperial.ac.uk; p.dragotti@imperial.ac.uk).

Digital Object Identifier 10.1109/JSTSP.2024.3454957

Despite achieving satisfactory results, the aforementioned methods have inevitably limited generalization capabilities because their algorithms are designed under non-blind degradation settings. To alleviate this limitation, several diffusion model-based approaches [25], [26], [27] have been recently developed for unknown degradation operators in specific tasks such as image deblurring [25], [26] and low-light enhancement [27]. Towards a more diverse and complicated degradation process, DifFace [2] introduces a pre-trained IR network $g(\cdot)$ (e.g., based on CNN or Transformer) to obtain an initial distortion-invariant clean image as a starting point, x_N , for the subsequent diffusion sampling process. In the framework of DR2 [1], smooth results are first predicted by an iterative refinement similar to ILVR [13] during sampling and then further processed by a pre-trained IR network $g(\cdot)$ to achieve high-quality details. PGDiff [3] constrains the high-quality image space during the posterior sampling process with a constraint on the MSE between $g(y)$ and the denoised intermediate output obtained at step t of the diffusion reverse process. StableSR [4] uses the Stable Diffusion model [28] for image super-resolution and is further equipped with a time-aware encoder and a controllable feature wrapping module.

The above non-blind and blind IR approaches have demonstrated the effectiveness of the generative diffusion models for IR tasks. However, they are faced with the following limitations: (1) In the task of non-blind IR, most existing approaches require a closed-form expression of the degradation model to guide the sampling process. However, the image processing pipeline of many modern imaging systems is so complex that it is often impossible to describe it explicitly. (2) In the task of blind IR, most existing blind IR approaches rely on pre-defined degradation models for training the IR network $g(\cdot)$, which also limits their flexibility in real-world scenarios.

To address the above issues, we propose an INN-guided probabilistic diffusion algorithm for both non-blind¹ and blind image restoration. During the sampling process of diffusion model, we impose an additional data-consistency step by introducing an off-the-shelf light-weight invertible neural network (INN). Specifically, we pre-train the forward process of INN to simulate an arbitrary degradation process. At testing stage we alternate between an unconditional diffusion sampling step that gives us an intermediate image consistent with the diffusion model and a consistency step guided by the INN that forces the reconstruction to be consistent with the measurements. In particular, given at each step an estimated image, the forward part of the INN produces a coarse image which we then force to be consistent with the measurements and the details estimated by the diffusion process. We then use the inverse part of the INN as a reconstruction process to obtain an intermediate result that guides the next step of the reverse diffusion process. Therefore, our method guides the sampling towards satisfying the consistency constraint while maintaining rich details provided by the diffusion prior. In the task of non-blind IR, INN is pretrained with datasets on any specific degradation, so it is no longer limited by the requirement

of knowing the analytical expression of the degradation model. In the task of blind IR, we first initialize the parameters of INN by training it with synthetic dataset pairs that model different degradation processes. Then, by alternating between refining the INN parameters for the unknown degradation model and updating intermediate image results with the guidance of INN during sampling, our approach is more flexible and can handle unknown degradation settings in real-world scenarios.

We summarize our contributions as follows:

- We propose a novel INN-guided probabilistic diffusion algorithm for non-blind and blind image restoration, namely INDIGO and BlindINDIGO. In contrast to most existing approaches, our algorithm introduces prior degradation information to the diffusion reverse process by simulating it with INN, which help to boost IR performance and improve flexibility.
- To the best of our knowledge, this is the first attempt to combine the merits of the perfect reconstruction property of INN with strong generative prior of diffusion models for blind image restoration. With the help of INN, our algorithm effectively estimates the details lost in the degradation process and is able to handle arbitrary degradation processes.
- We further introduce an initialization strategy to accelerate our algorithm by reducing the number of timestep.
- Extensive experiments show that our approach for both non-blind and blind image restoration achieves state-of-the-art results compared with other methods on synthetically degraded and real low-quality images (see Fig. 1 for an example).

II. BACKGROUND

A. Review of Denoising Diffusion Probabilistic Models

Diffusion models, e.g. [8], [9], [10], [11], [12], [29], [30], sequentially corrupt training data with slowly increasing noise, and then learn to reverse this corruption in order to form a generative model of the data. Here we describe a classic diffusion model: denoising diffusion probabilistic model (DDPM) [8]. DDPM defines a T -step forward process transforming complex data distribution into simple Gaussian noise distribution and a T -step reverse process recovering data from noise. The forward process slowly adds random noise to data, where, in the typical setting, the added noise has a Gaussian distribution. Consequently, the forward process yields the present state x_t from the previous state x_{t-1} :

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t\mathbf{I}) \quad (1)$$

where x_t is the noisy image at time-step t , β_t is a predefined scale factor. As noted in [8], the above process allows us to sample an arbitrary state x_t directly from the input x_0 as follows:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon \quad (2)$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=0}^t \alpha_i$ and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$. For the reverse process, we can calculate the posterior distribution $q(x_{t-1}|x_t, x_0)$ using Bayes theorem and write the expression

¹The work on non-blind inverse problem was presented in part at IEEE MMSP conference 2023 [24].

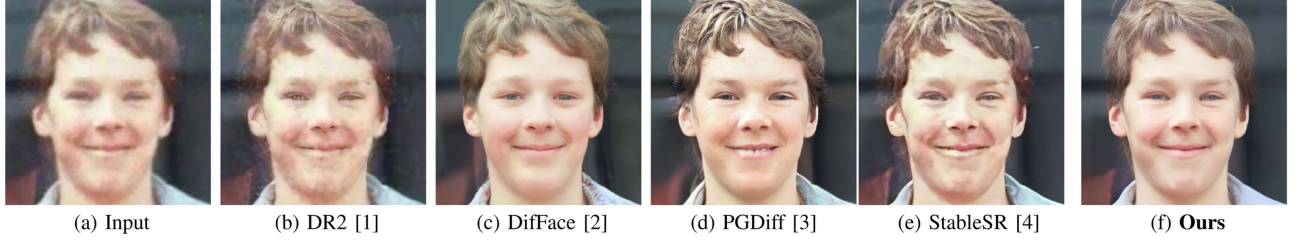


Fig. 1. Comparisons with state-of-the-art blind image restoration approaches [1], [2], [3], [4] on the real-world low-quality images. Our algorithm produces high-quality reconstruction results and preserves more details than the recent leading methods. (Zoom in for best view).

of $\mathbf{x}_{t-1}$ using (2) as follows:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon \right) + \sigma_t \mathbf{z}, \quad (3)$$

where $\sigma_t = \sqrt{\frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}} \beta_t$ and $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$. To predict the noise ϵ in the above equation, DDPM uses a neural network $\epsilon_\theta(\mathbf{x}_t, t)$ for each time-step t . To train $\epsilon_\theta(\mathbf{x}_t, t)$, DDPM uniformly samples a t from $\{1, \dots, T\}$ and updates the network parameters θ with the following gradient descent step:

$$\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\alpha_t} \mathbf{x}_0 + \sqrt{1 - \alpha_t} \epsilon, t)\|_2^2, \quad (4)$$

where $\mathbf{x}_0$ is a clean image from the dataset and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ is random noise. By replacing ϵ with the approximator $\epsilon_\theta(\mathbf{x}_t, t)$ in (3) and iterating it T times, DDPM can yield clean images $\mathbf{x}_0 \sim q(\mathbf{x})$ from initial random noises $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$, where $q(\mathbf{x})$ represents the image distribution in the training dataset.

Solvers of inverse problems that use diffusion models have shown remarkable performance and versatility, and can be divided into two groups. The first group of methods [31], [32], [33], [34], [35] has focused on designing and training conditional diffusion models suitable for image reconstruction tasks. The second group [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] has instead focused on keeping the training of unconditional diffusion models unaltered, and only modify the inference procedure to enable sampling from a conditional distribution. The approach proposed in this paper falls in the latter category and has the advantage of leveraging the pre-trained diffusion models to make them serve as a strong generative prior without the need of retraining diffusion models.

B. Wavelet Transform and Invertible Neural Networks

The wavelet transform is widely used in many imaging applications due to its ability to concentrate image features in a few large-magnitude wavelet coefficients, while small-value wavelet coefficients typically contain noise and can be shrunk or removed without affecting the image quality. The lifting scheme [36] is often used to construct a wavelet transform. As shown in Fig. 2(a), the forward wavelet transform converts the input signal into coarse and detail components and then the original signal is reconstructed by the inverse transform. Specifically, the lifting scheme first splits the signal $\mathbf{x} = (x_k)_{k \in \mathbb{Z}}$ into an even $\mathbf{x}_e = (x_{2k})_{k \in \mathbb{Z}}$ and an odd part $\mathbf{x}_o = (x_{2k+1})_{k \in \mathbb{Z}}$. A predictor is used to predict the odd part from the even part, and thus the difference between the odd part and its prediction

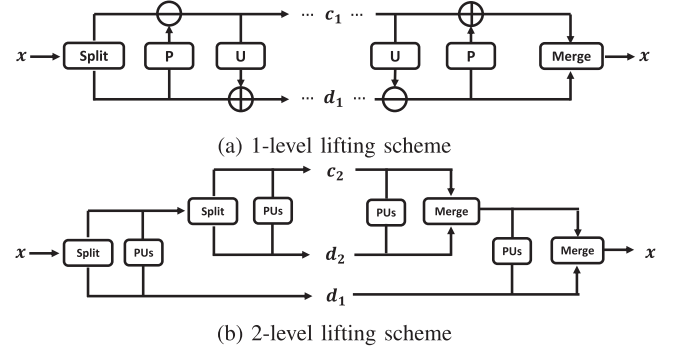


Fig. 2. The wavelet transform obtained using the lifting scheme.

reflects high-frequency details $\mathbf{d}$ of the signal. Based on this difference, the update step is used to adjust the even part to make it a smoother coarse version $\mathbf{c}$ of the original signal. The above lifting procedure implementing the forward wavelet transform can be described as:

$$\mathbf{d} = \mathbf{x}_o - P(\mathbf{x}_e), \quad \mathbf{c} = \mathbf{x}_e + U(\mathbf{d}). \quad (5)$$

The inverse transform can immediately be found by reversing the operations and flipping the signs. Therefore, the original signal can be recovered as follows:

$$\mathbf{x}_e = \mathbf{c} - U(\mathbf{d}), \quad \mathbf{x}_o = \mathbf{d} + P(\mathbf{x}_e). \quad (6)$$

The above equations illustrate that no matter how P and U are chosen, the scheme is always invertible and thus leads to critically sampled perfect reconstruction filter banks [36]. Furthermore, this scheme allows multiple levels and multiple pairs of predictors and updates (see Fig. 2(b)).

Inspired by the above idea, Huang et al. [37] propose a lifting-inspired invertible neural network (LINN) for image denoising. The forward transform of LINN non-linearly converts the input noisy image into coarse channel and detail channels. A denoising network performs the denoising operation on the detail part, and then the backward transform of the LINN reconstructs the denoised image using the original coarse channel and the denoised detail channels. In this architecture, INN consists of several invertible blocks where P and U in (5) and (6) become functions parameterized by neural networks. Specifically, the Predict and Update networks are applied alternatively to update the coarse and detail parts. The m -th pair of update and predict

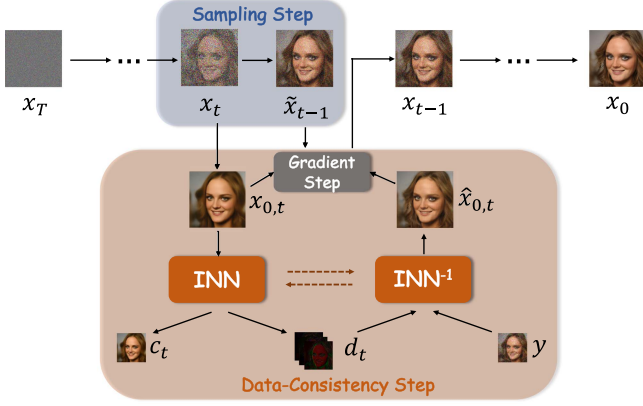


Fig. 3. Overview of our INDIGO for non-blind image restoration. Given a degraded image y during inference, the diffusion posterior sampling is guided by our data-consistency step with INN at each step t . We show the detailed algorithm in Algorithm 1.

operations of the k -th level INN can be expressed as:

$$d_m^k = d_{m-1}^k - P_m^k(c_{m-1}^k), \quad (7)$$

$$c_m^k = c_{m-1}^k + U_m^k(d_m^k), \quad (8)$$

where d_m^k and c_m^k denotes the updated detail part and coarse part using the m -th Predict network $P_m^k(\cdot)$ and Update network $U_m^k(\cdot)$, respectively. Similarly, the inverse transform of the k -th level INN can be expressed as:

$$c_{m-1}^k = c_m^k - U_m^k(d_m^k), \quad (9)$$

$$d_{m-1}^k = d_m^k + P_m^k(c_{m-1}^k). \quad (10)$$

There are also other choices for INN architectures, including coupling layer [38], affine coupling layer [39], reversible residual network [40] and i-RevNet architecture [41]. The invertible architecture that we design in this paper is based on the lifting-inspired invertible blocks in [37]. However, we use an alternative training strategy where we try to ensure that the coarse version produced by the network is as close as possible to the measured degraded image y .

III. INDIGO+ APPROACH

A. Overview

For a general image restoration problem $y = \mathcal{H}(x, n)$, we aim to obtain an image $\hat{x}$ that ensures data consistency while maintaining realistic textures. To simultaneously achieve these two goals, we leverage the merit of the perfect reconstruction property of INN and the strong generative prior of pretrained diffusion models. An overview of the proposed approach is shown in Figs. 3 and 4. We first train our INN so that its forward part $[c, d] = f_\phi(x)$ decomposes an image x in a coarse and detail part so that $c \approx \mathcal{H}(x, n)$. In other words, $f_\phi(\cdot)$ is trained to mimic the degradation process $\mathcal{H}$. Then during the diffusion posterior sampling process, we impose an additional data consistency step after each original unconditional sampling update. Specifically, we first utilize our pretrained INN to decompose

the intermediate result $x_{0,t}$ into the coarse part c_t that should approximate the degraded measurements and the detail part d_t that models the details lost during the degradation. We then replace c_t with the given observed measurements y . Next, the INN-optimized image $\hat{x}_{0,t}$ is constructed by inverse transform $f_\phi^{-1}(\cdot)$ of INN. Therefore, this INN-optimized result $\hat{x}_{0,t}$ guides the sampling towards satisfying the consistency constraint. Simultaneously, $\hat{x}_{0,t}$ maintains rich details obtained by diffusion posterior sampling without affecting data consistency. Then, the diffusion posterior sampling at the following step is guided by our data-consistent result, $\hat{x}_{0,t}$, through a gradient operation. Due to the fact that we train an INN to model the degradation process, our algorithm is more flexible than other methods and also more effective given that the invertibility property of the INN ensures that we compute implicitly the equivalent of an inverse at each iteration.

In the following subsections, we will explain in details how our approach can solve non-blind and blind inverse problems, respectively.

B. INDIGO for Non-Blind Image Restoration

In this subsection, we start with non-blind inverse problems and introduce the design of our INN and how it works in the diffusion process.

Modelling the degradation process with INN: By exploiting the invertibility of INN, we propose to treat its forward transform f_ϕ as a simulator of the degradation process and treat its inverse transform f_ϕ^{-1} as the reconstruction process. To realize this framework, we start with adopting the lifting-inspired invertible blocks in [37] (as in Section II-B), which can be expressed as follows:

$$[c, d] = f_\phi(x), \quad x = f_\phi^{-1}(c, d), \quad (11)$$

where the forward transform of INN generates the coarse and detail parts, c and d , while the inverse transform of INN can perfectly recover the input original image from c and d . To model the degradation process, we impose that c resembles y . Given a training set $\{x^i, y^i\}_{i=1}^N$, which contains N high-quality images and their low-quality counterparts, we optimize our INN with the following loss function:

$$L(\phi) = \frac{1}{N} \sum_{i=1}^N \|f_\phi^c(x^i) - y^i\|_2^2, \quad (12)$$

where ϕ denotes the set of learnable parameters of our INN and $f_\phi^c(x^i)$ and $f_\phi^d(x^i)$ denote the first and second part of the output of $f_\phi(x^i)$, respectively. Once we constrain one part of the output of $f_\phi(x^i)$ to be close to y , due to invertibility, the other part of the output will inevitably represent the detailed information lost during the degradation process.

Sampling with the guidance of pretrained INN: In the unconditionally trained DDPM [8], the reverse diffusion process iteratively samples x_{t-1} from $p(x_{t-1}|x_t)$ to yield clean images $x_0 \sim q(x)$ from initial random noise $x_T \sim \mathcal{N}(0, I)$. Here, we rewrite (3) with the pre-trained approximator $\epsilon_\theta(x_t, t)$ and split

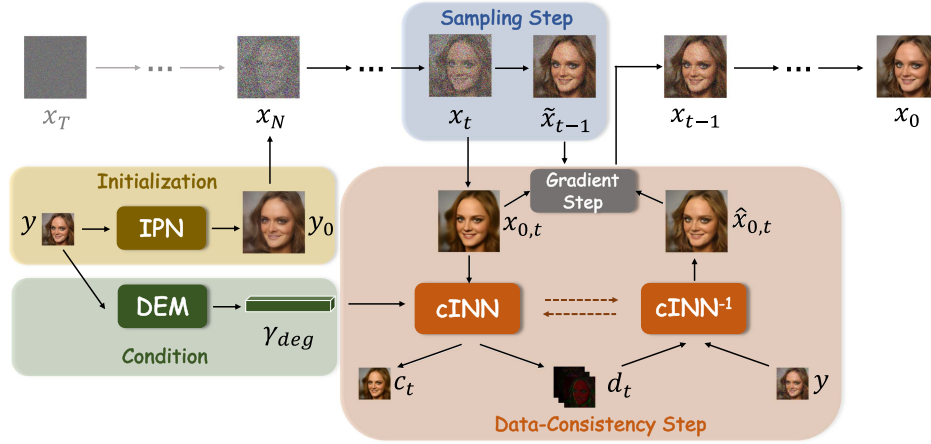


Fig. 4. Overview of our BlindINDIGO for blind image restoration. Given a degraded image y during inference, our approach first predicts a clean version y_0 with the initialization prediction network (IPN) and extract an implicit degradation embedding γ_{deg} with the degradation estimation module (DEM). Next, starting from a diffused y_0 , the diffusion posterior sampling is guided by our data-consistency step with INN at each step t . We show the detailed algorithm in Algorithm 2.

it into the following two equations:

$$x_{0,t} = \frac{1}{\sqrt{\alpha_t}}(x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_\theta(x_t, t)) \quad (13)$$

and

$$x_{t-1} = \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}\beta_t}}{1 - \bar{\alpha}_t}x_{0,t} + \sigma_t z. \quad (14)$$

As illustrated in (13), $x_{0,t}$ is the predicted clean image from the noisy image x_t . To solve inverse problems, we need to refine each unconditional transition using y to ensure data consistency. In our proposed algorithm, we impose our data-consistency step by modifying the clean image $x_{0,t}$ instead of the noisy image x_t .

As shown in Algorithm 1, we impose an additional data consistency step (in blue) with our off-the-shelf INN after each original unconditional sampling update. In this additional step, we apply the forward transform $f_\phi(\cdot)$ to the intermediate result $x_{0,t}$ leading to the decomposition of $x_{0,t}$ into coarse and detail part c_t, d_t respectively. We then replace the coarse part c_t with the measurements y . The INN-optimized $\hat{x}_{0,t}$ is then generated by applying the inverse transform $f_\phi^{-1}(\cdot)$ to $\{y, d_t\}$. Thus, the INN-optimized $\hat{x}_{0,t}$ is composed of the coarse information y and the details generated by the diffusion process. To incorporate the INN-optimized $\hat{x}_{0,t}$ into the DDPM algorithm, we update x_t with the guidance of the gradient of $\|\hat{x}_{0,t} - x_{0,t}\|_2^2$. With the help of INN, our algorithm effectively estimates the details lost in the degradation process and is no longer limited by the requirement of knowing the exact expression of the degradation model, since the degradation is learned through data.

C. BlindINDIGO for Blind Image Restoration

In the previous section, our INN is learned using a fully supervised approach given that we assume to have access to a training set $\{x^i, y^i\}_{i=1}^N$ where the degradation in y^i is fully consistent with the degradation of the actual measurements. In practical scenarios, many images undergo complex and unknown degradation processes. Some works [25], [26] solve this by assuming a closed-form expression of the degradation process

Algorithm 1: Non-Blind INDIGO.

Input: Corrupted image y , gradient scale ζ , INN f_ϕ .
Output: Output image x_0 conditioned on y

$x_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for t from T to 1 **do**
 $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $z = \mathbf{0}$
 $x_{0,t} = \frac{1}{\sqrt{\alpha_t}}(x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_\theta(x_t, t))$
 $\tilde{x}_{t-1} = \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}\beta_t}}{1 - \bar{\alpha}_t}x_{0,t} + \sigma_t z$
 $c_t, d_t = f_\phi(x_{0,t})$
 $\hat{x}_{0,t} = f_\phi^{-1}(y, d_t)$
 $x_{t-1} = \tilde{x}_{t-1} - \zeta \nabla_{x_t} \|\hat{x}_{0,t} - x_{0,t}\|_2^2$
end
return x_0

and then they predict the parameters in this expression. In this work, by simulating several degradation processes and through finetuning, our approach can deal with unknown, linear and non-linear degradation processes. The algorithm is described in Fig. 4 and Algorithm 2.

Conditional INN for blind image restoration: Since one set of parameters ϕ in $f_\phi(x)$ can simulate one type of degradation, we take different degradation labels as an additional input to guide the forward and inverse transform of INN, i.e., $f_\phi(x, \gamma_{deg})$, to simulate multiple different degradation processes. To extract the degradation information γ_{deg} , we utilize a pre-trained Degradation Estimation Module (DEM) to model the degradation implicitly in the latent feature space, since real-world degradations are usually too complex to be modeled with an explicit combination of multiple degradation types. As depicted in Fig. 5, we keep the basic invertible blocks of [37] and modulate them with the degradation vector γ_{deg} . Specifically, the image is split into two parts by a splitting operator. Then the Prediction Network (PNNet) conditioned on the coarse part aims to predict the detail part, while the Update Network (UNet) conditioned on the detail part is used to adjust the coarse part to make it smoother. The Prediction and Update networks are applied alternatively to generate the coarse and detail parts c and d . The details of how the degradation vector γ_{deg} controls the features are

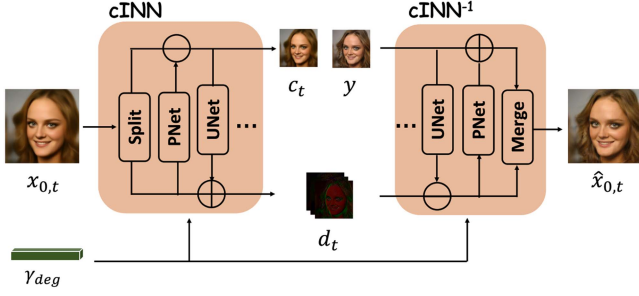


Fig. 5. The forward and inverse transform of our INN during inference.

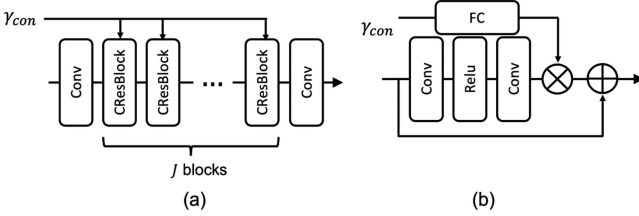


Fig. 6. (a) The architecture of the PNet/UNet. (b) The architecture of the CResBlock [42].

Algorithm 2: BlindINDIGO.

Input: Corrupted image $\mathbf{y}$, gradient scale ζ , pretrained INN $f_\phi(\cdot)$, pretrained IPN $g_\omega(\cdot)$, implicit degradation embedding γ_{deg} extracted by pre-trained DEM from $\mathbf{y}$, learning rate l for optimizing INN.

Output: Output image $\mathbf{x}_0$ conditioned on $\mathbf{y}$

$\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

$\mathbf{x}_N = \sqrt{\bar{\alpha}_N} g_\omega(\mathbf{y}) + \sqrt{1 - \bar{\alpha}_N} \boldsymbol{\eta}$

for t from N to 1 **do**

$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$

$\mathbf{x}_{0,t} = \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{x}_t, t))$

$\tilde{\mathbf{x}}_{t-1} = \frac{\sqrt{\bar{\alpha}_t(1-\bar{\alpha}_{t-1})}}{1-\bar{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}\beta_t}}{1-\bar{\alpha}_t} \mathbf{x}_{0,t} + \sigma_t \mathbf{z}$

$\mathbf{c}_t, \mathbf{d}_t = f_\phi(\mathbf{x}_{0,t}, \gamma_{deg})$

$\hat{\mathbf{x}}_{0,t} = f_\phi^{-1}(\mathbf{y}, \mathbf{d}_t, \gamma_{deg})$

$L_{total} = \lambda_F L_F(\mathbf{c}_t, \mathbf{y}) + \lambda_I L_I(\hat{\mathbf{x}}_{0,t}, \mathbf{x}_{0,t})$

$\mathbf{x}_{t-1} = \tilde{\mathbf{x}}_{t-1} - \zeta \nabla_{\mathbf{x}_t} L_{total}$

$\phi \leftarrow \phi - l \nabla_\phi \|\mathbf{c}_t - \mathbf{y}\|_2^2$

end

return $\mathbf{x}_0$

shown in Fig. 6. The degradation vector γ_{deg} is passed through a fully-connected layer and a channel-wise multiplication is added to the original Residual block [43] to control the summation weight. To train this conditional INN, the loss function becomes:

$$L(\phi) = \frac{1}{N} \sum_{i=1}^N \|f_\phi^c(\mathbf{x}^i, \gamma_{deg}^i) - \mathbf{y}^i\|_2^2, \quad (15)$$

where $f_\phi^c(\mathbf{x}^i, \gamma_{deg}^i)$ denotes the first part of the output of $f_\phi(\mathbf{x}^i, \gamma_{deg}^i)$ and the implicit degradation vector γ_{deg}^i is generated by a pretrained DEM $h_\kappa(\cdot)$. Here, the training set $\{\mathbf{x}^i, \mathbf{y}^i, \gamma_{deg}^i\}_{i=1}^N$ contains N high-quality images, their low-quality counterparts, and the implicit degradation vector $\gamma_{deg}^i = h_\kappa(\mathbf{y}^i)$ generated by a pretrained DEM.

Guiding posterior sampling with INN: Similar to non-blind INDIGO, we still apply the forward transform $f_\phi(\cdot)$ to the intermediate result $\mathbf{x}_{0,t}$ and then replace its coarse part $\mathbf{c}_t$ with the measurements $\mathbf{y}$. The INN-optimized $\hat{\mathbf{x}}_{0,t}$ is then generated by applying the inverse transform $f_\phi^{-1}(\cdot)$. The invertibility of the INN allows us to compute the gradient step in either the measurement or the image domain. In the non-blind case, we only operate in the image domain. For the blind case, to further improve the reconstruction performance, we operate in both domains. Specifically, we take the gradient of the following loss:

$$L_{total} = \lambda_F L_F(\mathbf{c}_t, \mathbf{y}) + \lambda_I L_I(\hat{\mathbf{x}}_{0,t}, \mathbf{x}_{0,t}), \quad (16)$$

where we constrain the measurement space with $L_F(\mathbf{c}_t, \mathbf{y})$ and the high-quality image space with $L_I(\hat{\mathbf{x}}_{0,t}, \mathbf{x}_{0,t})$. Here, λ_F and λ_I denote the loss weights of L_F and L_I , respectively. In our implementation, we set $\lambda_F = 2.5$, $\lambda_I = 1$, $L_F(\mathbf{c}_t, \mathbf{y}) = \|\mathbf{c}_t - \mathbf{y}\|_2^2$ and $L_I(\hat{\mathbf{x}}_{0,t}, \mathbf{x}_{0,t}) = \|V(\hat{\mathbf{x}}_{0,t}) - V(\mathbf{x}_{0,t})\|_2^2$, where $V(\cdot)$ denotes the feature embedding space of VGG16 [45] network. We discuss the detailed implementation and ablation study in Section IV-C5.

Finetuning our INN during sampling: By replacing the INN in Algorithm 1 with the above pretrained conditional INN, our approach can deal with multiple inverse problems. However, both conditional INN and the DEM are trained with synthetic degradation data pairs that may not model exactly the actual degradation. In real-world scenarios with more complex degradations, the parameters of our INN need to be refined to simulate the degradation processes more accurately. We achieve this by finetuning the parameters in INN at testing stage. This is done at the end of each iteration as shown in Algorithm 2. In this step given the current estimated image $\mathbf{x}_{0,t}$, the parameters of our INN, f_ϕ , are updated to reduce $L_F(\mathbf{c}_t, \mathbf{y})$ in (16).

Accelerating our algorithm with initialization: To accelerate our algorithm by reducing the number of timesteps, we introduce an initialization strategy. As observed in [2], [14], starting from a single forward diffusion with better initialization instead of Gaussian noise significantly reduces the number of sampling steps in the diffusion posterior sampling process. Following [2], and as shown in Algorithm 2, we first produce an initial restored image, $g_\omega(\mathbf{y})$, using an image restoration method. By construction, this image would already look consistent with the measurements but lacks details. Then it is forward-diffused (noise-added) to generate the starting point of sampling, $\mathbf{x}_N = \sqrt{\bar{\alpha}_N} g_\omega(\mathbf{y}) + \sqrt{1 - \bar{\alpha}_N} \boldsymbol{\eta}$. In our setting, $g_\omega(\cdot)$ is the SwinIR [46] method trained with L2 Loss.

D. Algorithm 1 vs. Algorithm 2

In summary, Algorithm 1 can be used in fully supervised settings and for the case of the single, fixed degradation. It requires training data pairs which are degraded in the same way as the testing dataset to train the INN. To evaluate Algorithm 1, we need to train different INNs for different degradations. Algorithm 2 can be used for multiple degradations including unseen cases. For example, to reconstruct measurement $\mathbf{y}$ from degradation model $\mathcal{H}$ during inference, in the non-blind case (Algorithm 1) we have access to labeled dataset $\{\mathbf{x}^i, \mathbf{y}^i = \mathcal{H}(\mathbf{x}^i)\}_{i=1}^N$ to train

TABLE I
QUANTITATIVE (PSNR↑/FID↓/LPIPS↓/NIQE↓) COMPARISON ON 4× SR WITH DIFFERENT LEVELS OF GAUSSIAN NOISE

Methods	Noise $\sigma=0$				Noise $\sigma=0.05$				Noise $\sigma=0.10$			
	PSNR↑	FID↓	LPIPS↓	NIQE↓	PSNR↑	FID↓	LPIPS↓	NIQE↓	PSNR↑	FID↓	LPIPS↓	NIQE↓
ILVR [13]	27.43	44.04	0.2123	5.4689	26.42	60.27	0.3045	4.6527	24.60	88.88	0.4833	4.4888
DDRM [18]	28.08	65.80	0.1722	4.4694	27.06	45.90	0.2028	4.8238	26.16	45.49	0.2273	4.9644
DPS [22]	26.67	32.44	0.1370	4.4890	25.92	31.71	0.1475	4.3743	24.73	31.66	0.1698	4.2388
Ours	28.15	22.33	0.0889	4.1564	27.16	26.64	0.1215	4.1004	26.25	28.89	0.1399	3.9659
Ours-DDIM	28.06	24.51	0.0966	4.2524	27.19	28.07	0.1271	4.3619	26.22	30.85	0.1488	4.2189

Bold texts represent the best performance.

our INN, while in the blind case (Algorithm 2) we need to generate a dataset with several distortion models: $\{x^i, y^i = \mathcal{H}^j(x^i)\}_{i=1}^N$, $\mathcal{H}^j \in \{\mathcal{H}^j\}_{j=1}^J$ to train our INN and the correct $\mathcal{H}$ may not belong to $\{\mathcal{H}^j\}_{j=1}^J$.

IV. EXPERIMENTAL RESULTS

A. Results on Non-Blind Image Restoration

1) *Implementation Details:* Empirically, we set $\zeta = 0.5$ and $T = 1000$ in Algorithm 1. To implement our INN, we follow the structure of the invertible blocks in [37]. Specifically, our INN consists of 2 levels of the lifting-inspired invertible blocks and each block is constructed using the same set of 4 pairs of PNet and Unet as in [37]. Each PNet/Unet network consists of an input convolutional layer, 2 residual blocks with depth-wise separable convolution layers, and an output convolutional layer. The number of feature channels in PUNet is set to 32 and the spatial filter size in depth-wise separable convolutional layers is set to 5. The total number of learnable parameters of our INN is 0.71M.

We test our method on the FFHQ 256×256 1k validation dataset [47], CelebA HQ 256×256 1k validation dataset [48], and a real-world SR dataset DRealSR [44]. For the face photo reconstruction, we utilize a pre-trained unconditional diffusion model trained on the FFHQ training dataset by [22] and select 10k images from the FFHQ training dataset to train our INN. For the natural image reconstruction on [44], we utilize the pre-trained unconditional diffusion model trained on ImageNet [49] by [10] and use DRealSR [44] training dataset to train our INN. We apply our proposed method to three settings for inverse problems: bicubic downsampling with/without noise, non-linear degradation model based on combining downsampling with jpeg compression, real-world degradation model. In this section, we assume the degradation model is known, so for the first two settings, we synthesize training and testing data using the same degradation. For the third setting, we train our INN with DRealSR training dataset [44], which is a large-scale diverse SR benchmark obtained by zooming digital single-lens reflex (DSLR) cameras to collect real low-resolution (LR) and high-resolution (HR) images. After simulating its degradation with our INN, we test our algorithm with DRealSR test dataset [44]. The reconstruction results are evaluated with PSNR, FID [50], LPIPS [51] and NIQE [52].

2) *Results With Bicubic Downsampling Degradation Model:* We compare our method with 3 state-of-the-art methods based on diffusion models: ILVR [13], DDRM [18], and DPS [22]. As shown in Table I, we evaluate all methods on the problem

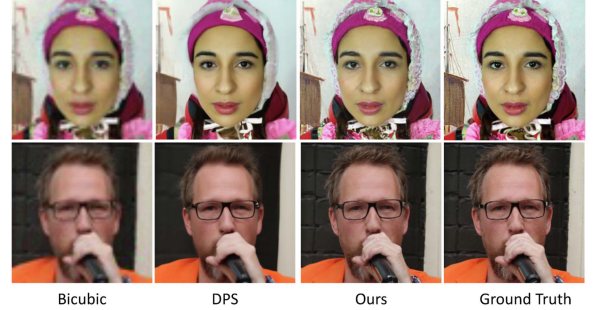


Fig. 7. Comparisons with state-of-the-art image restoration approach [22] on solving the **non-blind** super-resolution problem (x4) on FFHQ validation dataset.

of bicubic downsampling (4×) with different levels of Gaussian noise on the FFHQ dataset. Please note that in the first setting which we named ‘ours’, we apply our consensus strategy (see Appendix A). To further accelerate the sampling strategy, we use DDIM [9] as the sampling strategy with time step 250 and present the results as ‘ours-DDIM’. Further discussion on sampling with DDIM can be seen in Section IV-C6. One can observe that our method outperforms all baseline methods in all metrics. Moreover, as it can be seen in Fig. 7, our algorithm produces high-quality reconstruction results while preserving realistic details.

3) *Results With Non-Linear Degradation Model:* As described in the introduction, our flexible INN can handle a variety of degradation processes. In this setting, a bicubic downsampling operator is first performed on a high-quality image and then a JPEG compression degradation is applied to save the low-resolution image into JPEG format, where the jpeg factor and downsampling scale are 20 and 4, respectively. As shown in Fig. 8, our algorithm can still produce high-quality images from heavily degraded input measurements. It can be seen that our results produce realistic details while ensuring data consistency. Note that the methods we compared against in Section IV-A2 are not able to support this type of degradation, so for this case, we only show our results.

4) *Results With Real Degradation Model:* In Fig. 9, we show the results of our algorithm on reconstructing images from real-world degradation processes using DRealSR [44] with scale factor 4. As for Section IV-A3, since we simulate the degradation with INN, our solution is no longer limited by the requirement of knowing the exact expression of the degradation process. We can observe that our algorithm can produce, also in this case, high-quality results for real-world degradation.



Fig. 8. Results of our algorithm on solving the **non-blind** inverse problem with Jpeg compression on CelebA HQ validation dataset.

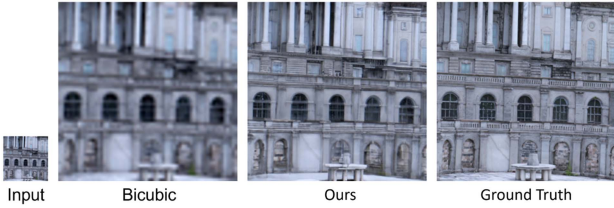


Fig. 9. Result of our algorithm on reconstructing real images from DRealSR [44] with resolution enhancement by a factor 4 per direction (**non-blind**).

B. Results on Blind Image Restoration

1) *Implementation Details*: Empirically, we set $T = 1000$, $N = 400$, $\zeta = 1.5$, and $l = 1e-5$ in Algorithm 2. Based on the framework for the non-blind setting, we adopt CResblocks [42] for our PNet/UNet, where the number of feature channels is set to 32 and the dimension of γ_{deg} is 128. The total number of learnable parameters of our INN is 0.91M. We directly utilize the pre-trained implicit degradation estimator in [53] as DEM and we follow DifFace [2] to implement our IPN by using a pre-trained SwinIR [46] Network.

We test our method on blind face restoration and blind natural image restoration. For the face photo reconstruction, we utilize a pre-trained unconditional diffusion model trained on the FFHQ training dataset by [2] and select 10k images from the FFHQ training dataset to train our INN. For the natural image reconstruction on ImageNet validation dataset [49], we utilize the pre-trained unconditional diffusion model trained on ImageNet [49] by [10] and use DIV2K [54] training dataset to train our INN. Following previous works [1], [55], [56], [57], [58], [59], we adopt a commonly used degradation model as follows to synthesize training data:

$$\mathbf{y} = [(\mathbf{x} \circledast \mathbf{k}_\sigma)_{\downarrow r} + \mathbf{n}_\delta]_{\text{JPEG}_q}, \quad (17)$$

where the high quality image $\mathbf{x}$ is first convolved with a Gaussian blurring kernel $\mathbf{k}_\sigma$ followed by a downsampling operation with a scale factor r . After that, additive white Gaussian noise $\mathbf{n}_\delta$ is added to the image, and finally the noisy image is compressed by JPEG with quality factor q . To train our INN, we set $r = 4$,

and randomly sample values of σ , δ and q from the intervals [3, 9], [5, 50], and [30, 80] respectively. The reconstruction results are evaluated with PSNR, LPIPS [51], BRISQUE [60], Identity Similarity (IDS) using the ArcFace similarity [61].

2) *Results on Synthetic Degradation*: We test our method on CelebA HQ 512×512 1k validation dataset [48] and ImageNet validation dataset [49] on synthetic degradation. We compare our method with 4 state-of-the-art methods based on diffusion models: PGDiff [3], DifFace [2], DR2 [1], and StableSR [4]. To evaluate these methods on different levels of degradation, we test them on mild ($\sigma = 4$, $\delta = 15$, $q = 70$), medium ($\sigma = 6$, $\delta = 25$, $q = 50$) and severe ($\sigma = 8$, $\delta = 35$, $q = 30$) degradations respectively. We provide the quantitative comparison on different levels of degradations in Table II and we see that our approach achieves significant gains over existing works, in particular, for medium and severe degradations. To further accelerate the sampling strategy, we use DDIM [9] as the sampling strategy with $T = 250$, $N = 100$ and present the results as ‘ours-DDIM’. One can observe that DDIM can speed up the reconstruction process with a slight reduction in perceptual quality. Further discussion on sampling with DDIM can be seen in Section IV-C6. Qualitative comparisons in Figs. 10–12 demonstrate the superiority of our BlindINDIGO in comparison to existing methods on different levels of degradations. Fig. 13 demonstrates the robustness of our approach, enabling its application to a variety of categories, including cats, dogs, and lions. Here, we measure the runtime of all the approaches on an Nvidia RTX 3080 GPU and show complexity comparison in Table III. The parameter numbers here do not include the parameters of the pre-trained diffusion model.

3) *Results With Real Degradation Model*: We also apply our BlindINDIGO on a real-world dataset CelebChild [57] for evaluating the generalization of the proposed method. CelebChild-Test contains 180 child faces of celebrities collected from the Internet. They are low-quality and many of them are black-and-white old photos. Since no ground-truth images are available for this setting, we compare the image quality using BRISQUE [60] and compare the image fidelity with Identity Similarity (IDS) using the ArcFace similarity [61] in Table IV. We observe that our approach achieves best or second best scores, demonstrating its superiority in the generation of high-quality images and effectiveness in preserving identity. The qualitative comparison on CelebChild [57] is shown in Fig. 14.

C. Analysis and Discussion

1) *Comparison to Conditional Diffusion Model and Standard Supervised Learning Approach*: Different from training a conditional diffusion model for a specific inverse problem from scratch, we keep the same pre-trained diffusion model and only modify the inference procedure with the guidance of INN to enable sampling from a conditional distribution. This strategy can efficiently leverage the trained diffusion model trained on huge amount of data to make it serve as a strong generative prior in different inverse problems. In addition, it saves the cost of training, since we only need to train the INN (0.71M). Furthermore, our blind INDIGO does not necessarily need to have access to

TABLE II
QUANTITATIVE (PSNR $\uparrow$ /LPIPS $\downarrow$ /IDS $\uparrow$) COMPARISON ON 4 $\times$ SR WITH DIFFERENT LEVELS OF DEGRADATIONS

Methods	Mild			Medium			Severe		
	PSNR $\uparrow$	LPIPS $\downarrow$	IDS $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	IDS $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	IDS $\uparrow$
PGDiff [3]	23.38	0.2668	0.9509	22.56	0.2883	0.9442	21.96	0.3062	0.9371
DifFace [2]	24.32	0.2782	0.9441	23.94	0.2896	0.9362	23.45	0.3014	0.9270
DR2 [1]	25.86	0.2775	0.9312	24.23	0.2916	0.9105	23.34	0.3030	0.9169
StableSR [4]	25.70	0.2155	0.9545	24.46	0.2371	0.9552	23.62	0.2559	0.9538
Ours	25.41	0.2142	0.9662	24.85	0.2360	0.9585	24.16	0.2534	0.9558
Ours-DDIM	25.72	0.2297	0.9634	25.09	0.2479	0.9570	24.31	0.2550	0.9390

Bold texts represent the best performance.

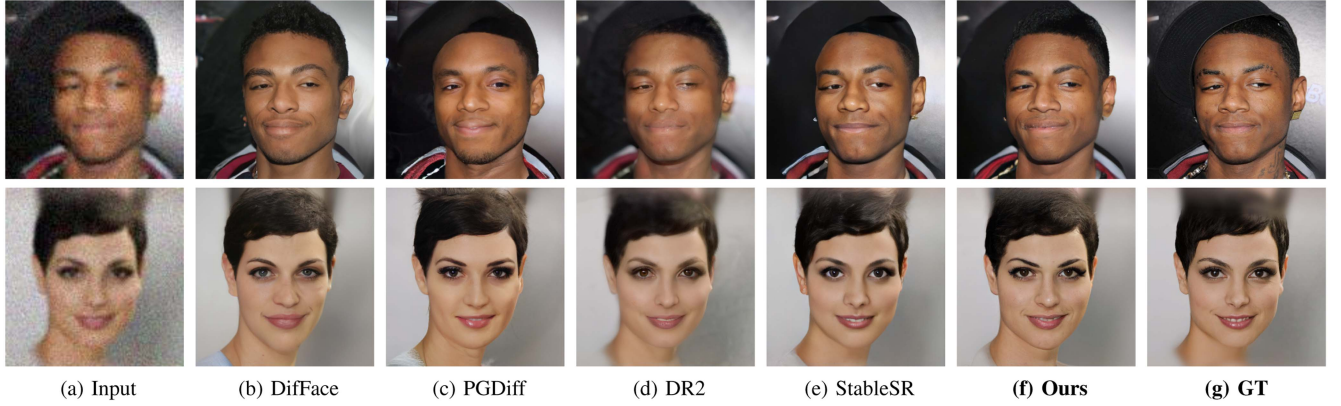


Fig. 10. Comparisons on 4x **blind** SR with mild degradation on CelebA-HQ.

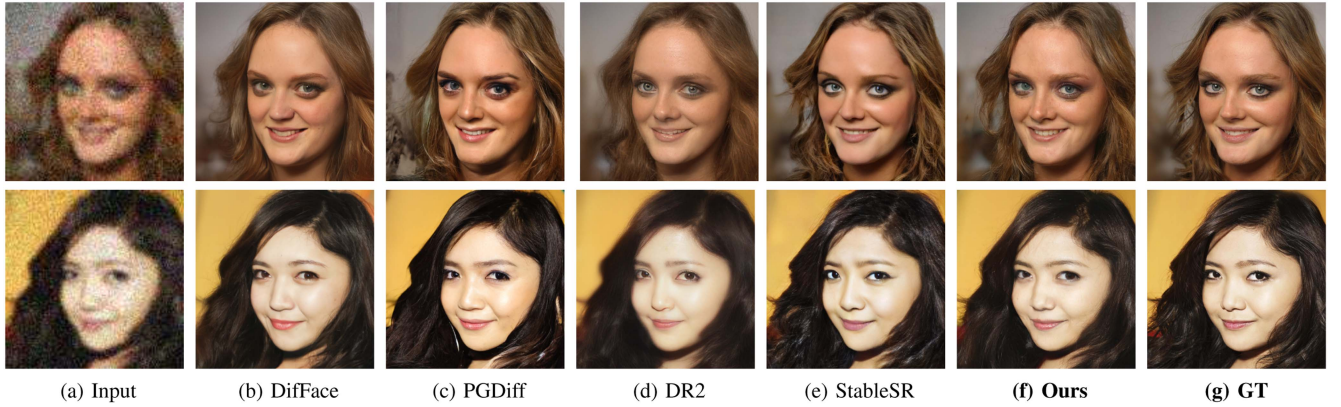


Fig. 11. Comparisons on 4x **blind** SR with medium degradation on CelebA-HQ.

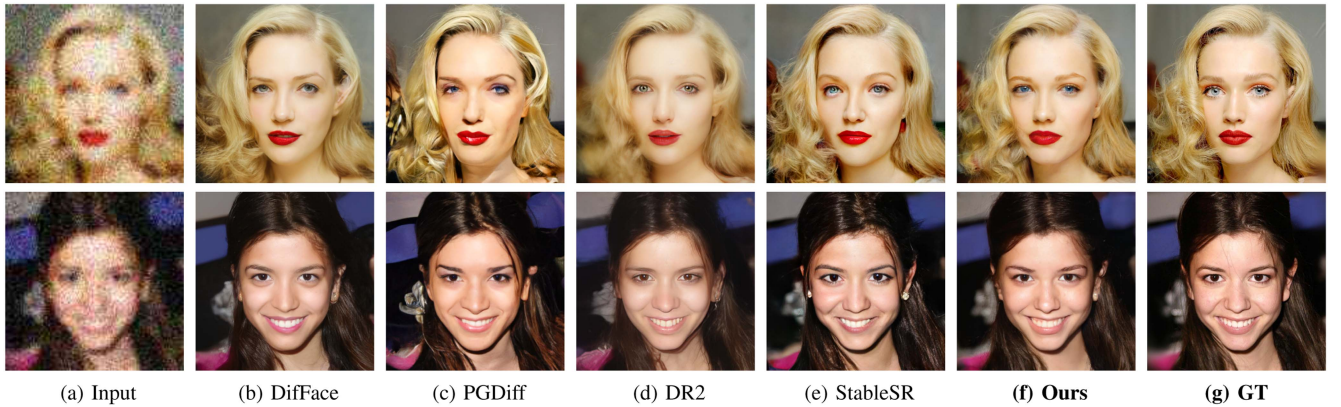


Fig. 12. Comparisons on 4x **blind** SR with severe degradation on CelebA-HQ.

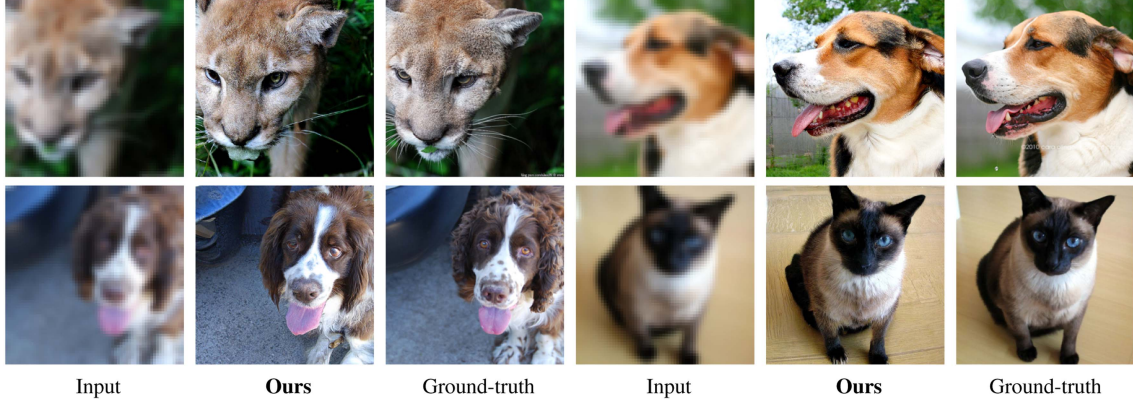


Fig. 13. Result of our BlindINDIGO on 4x super-resolution on the ImageNet dataset.

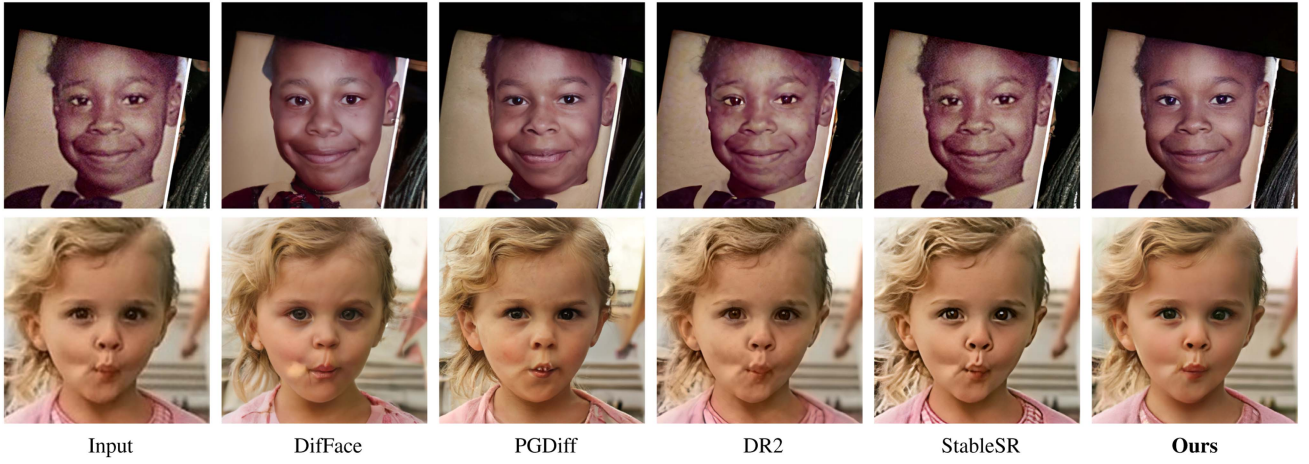


Fig. 14. Comparisons with blind state-of-the-art image restoration approaches [1], [2], [3], [4] on real-world dataset celaba-child.

TABLE III
COMPLEXITY COMPARISON WITH STATE-OF-THE-ART BLIND IMAGE RESTORATION APPROACHES (RUNTIME/PARAMS/ITERATIONS)

	DR2	DifFace	PGDiff	StableSR	Ours
Runtime	0.96s	3.36s	59.37s	126.07s	43.85s
Params	86M	16M	17M	94M+105M	16M+1M
Iterations	100	100	1000	1000	400

TABLE IV
QUANTITATIVE (BRISQUE/IDENTITY SIMILARITY (IDS)) COMPARISON ON REAL-WORLD FACE IMAGE RESTORATION

Methods	Brisque ↓	IDS ↑
PGDiff [3]	<u>21.82</u>	0.8013
DifFace [2]	25.96	0.7697
DR2 [1]	25.01	0.9503
StableSR [1]	25.30	0.9302
Ours	20.35	<u>0.9315</u>

Bold and underlined texts represent the best and second best performance.

labeled datasets, while conditional diffusion models need them. In this section, we compare our approach with two representative methods from conditional diffusion models: I2SB [62] and

standard supervised learning approaches SwinIR [46] on 4x SR with different levels of Gaussian noise on Imagenet validation dataset. Table V shows that our approach achieves competitive reconstruction results in the noiseless case and outperforms significantly the other methods in noisy settings.

2) *Analysis on the Sampling Process of Our Approach:* We show the visual results of the sampling process in Fig. 15. To clearly illustrate the comparison, we present the noisy results during the sampling process in Fig. 15(a), and the clean results in Fig. 15(b).

Firstly, as shown in Algorithm 2, given a degraded image y during inference, our approach first predicts a clean version y_0 with the Initialization Prediction Network (IPN). As shown in the first row of Fig. 15(b), IPN effectively predict a smooth result y_0 which still lacks details compared to our final result x_0 .

Secondly, since we set $T = 1000$ and $N = 400$ in Algorithm 2, we start from x_{400} instead of x_{1000} . As shown in Fig. 15(a), the noise gradually decreases with iterations from $t = 400$ to 0. And the second row of Fig. 15(b) shows the effectiveness of our data-consistency step guided by the INN.

3) *Effect of Our INN:* Our INN is designed to simulate the degradation process, as in $[c, d] = f_\phi(x, \gamma_{deg})$. To demonstrate

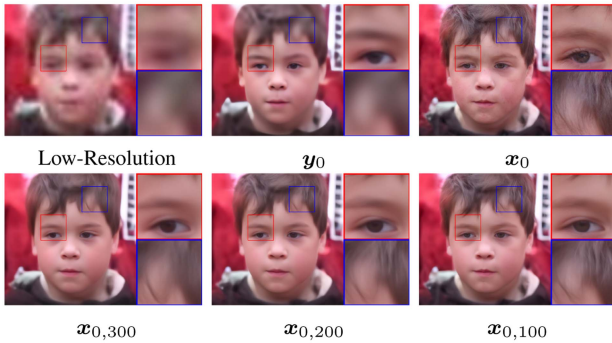
TABLE V
QUANTITATIVE (PSNR $\uparrow$ /FID $\downarrow$) COMPARISON ON 4 $\times$ SR WITH DIFFERENT LEVELS OF GAUSSIAN NOISE

Methods	$\sigma=0$		$\sigma=0.05$		$\sigma=0.10$		Time
	PSNR $\uparrow$	FID $\downarrow$	PSNR $\uparrow$	FID $\downarrow$	PSNR $\uparrow$	FID $\downarrow$	
SwinIR	24.60	118.55	24.29	125.41	23.51	152.07	0.11s
I2SB	25.46	55.27	21.90	138.53	17.62	248.15	36.20s
Ours	26.28	79.98	25.70	81.85	24.58	99.08	54.27s

Bold and underlined texts represent the best and second best performance.



(a) Noisy intermediate results.



(b) Clean intermediate results.

Fig. 15. Sampling process of our approach.

the ability of our INN to simulate different levels of degradation, we present the results of c and d in the second and third rows of Fig. 18. As a reference, the first row shows the degraded measurements y . From left to right, the level of degradation gradually increases, and the coarse part c in the second row also follow this trend.

4) *Effect of Step Size*: The value of step size ζ is essentially the weight that is given to the data consistency of the inverse problem. Fig. 16 shows results with different step sizes ζ in our non-blind INDIGO in the case of noise level $\sigma = 0.1$. One can observe that with low values of the step size, the results we obtain have lower consistency with the given measurements. On the other hand, setting the step size value too high leads to artefacts that tend to amplify the noise. Therefore, we set the step size $\zeta = 0.5$ and $\zeta = 1.5$ by default in our non-blind INDIGO and BlindINDIGO, respectively.

5) *Effect of Loss Function*: We explore different loss function designs for our BlindINDIGO in Table VIII, where $L_{pix,y}$, $L_{pix,img}$, $L_{fea,img}$ denote MSE loss in measurement space between c_t and y , MSE loss in image space between $\hat{x}_{0,t}$, $x_{0,t}$ and perceptual loss in image space between $\hat{x}_{0,t}$, $x_{0,t}$, respectively. One can observe that the result with loss function $L_{pix,y}$ achieves best PSNR, while the result with loss function $L_{pix,y}$ and $L_{fea,img}$ achieves best perceptual quality. We use the loss function $L_{pix,y}$ and $L_{fea,img}$ in all our experiments for the blind case.

TABLE VI
PERFORMANCE OF OUR APPROACH WITH DIFFERENT NFE USING DDPM AND DDIM IN TERMS OF PSNR $\uparrow$ /LPIPS $\downarrow$ ON 4 $\times$ SR

Sampling	NFE	PSNR	LPIPS
DDPM	1000 (default)	28.12	0.0806
DDIM	500	28.21	0.0847
DDIM	250	28.31	0.0914
DDIM	200	28.20	0.0986
DDIM	100	26.01	0.1709

6) *Sampling With DDIM*: To accelerate the sampling process, we use DDIM [9] as the sampling strategy, which skips steps in the reverse process to speed up the DDPM generating process. We show the performance in terms of PSNR and LPIPS with respect to the change in NFEs (number of neural function evaluations) on the first 100 images of the testing dataset in Table VI. One can observe that our algorithm achieves good performance when $NFE \geq 250$, whereas when $NFE = 100$, performance deteriorates significantly. Moreover, as it can be seen in Fig. 19, both versions of our algorithm, that is, the default version based on DDPM with $T = 1000$ and the fast version based on DDIM with $T = 250$, produce high-quality reconstruction results while preserving realistic details. In addition, we apply DDIM as the sampling strategy in our BlindINDIGO and present the results in Table VII. As shown in Table VII, we investigate effects of different components in our BlindINDIGO. We apply our BlindINDIGO on 3 settings: synthetic seen degradation, synthetic unseen degradation and real unseen degradation. For synthetic seen degradation, we test with medium degradation ($\sigma = 6$, $\delta = 25$, $q = 50$). For synthetic unseen degradation, we set $\sigma = 40$, $\delta = 0$, and $q = 100$. Both of them are evaluated on a subset (first 100 images) of CelebA HQ validation dataset. For the real unseen degradation setting, we evaluate our approach on a subset (first 100 images) of WIDER-Test dataset. One can observe the effect of DDIM sampling from cases 1 and 2, from cases 5 and 6, or from cases 8 and 9 in Table VII. Overall, this analysis confirms that our approach is compatible with DDIM and that DDIM can be used to accelerate the reconstruction process with only a small compromise to the perceptual quality.

7) *Effect of Initialization*: As shown in Table VII, we investigate the effect of the initialization in our BlindINDIGO. When comparing cases 3 and 4, or when comparing cases 9 and 11, we can see that with our initialization strategy, our approach performs better in terms of both reconstruction accuracy and image quality.

8) *Effect of Finetuning*: As discussed in Section III-C, in real-world scenarios with more complex degradations, the

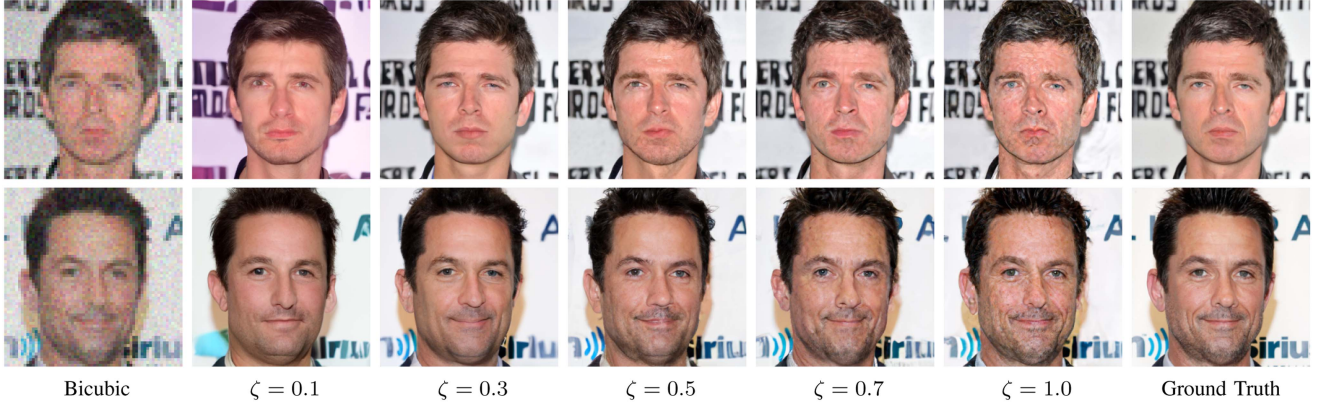


Fig. 16. Ablation study on the choice of step size schedule for our INDIGO.

TABLE VII
ABLATION STUDY ON BLINDINDIGO

	Case	Sampling	T	N	Initialization	Finetuning	PSNR $\uparrow$	LPIPS $\downarrow$	Time
Synthetic Seen	1	DDPM	1000	400	✓	✓	25.08	0.2334	43.85s
	2	DDIM	250	100	✓	✓	25.28	0.2443	11.03s
	3	DDIM	250	100	✓	✗	25.31	0.2442	9.25s
	4	DDIM	250	100	✗	✗	24.08	0.3455	9.19s
Synthetic Unseen	5	DDPM	1000	400	✓	✓	24.67	0.2196	43.85s
	6	DDIM	250	100	✓	✓	25.18	0.2280	11.03s
	7	DDIM	250	100	✓	✗	23.19	0.2546	9.25s
Real Unseen	8	DDPM	1000	400	✓	✓	120.42	4.2694	43.85s
	9	DDIM	250	100	✓	✓	127.36	4.2297	11.03s
	10	DDIM	250	100	✓	✗	136.37	4.6045	9.25s
	11	DDIM	250	100	✗	✓	155.14	5.7365	10.94s

TABLE VIII
ABLATION STUDY ON THE LOSS FUNCTION IN TERMS OF PSNR $\uparrow$ /LPIPS $\downarrow$ IN THE CASE OF MEDIUM DEGRADATION ON A SUBSET OF CELEBAHQ-TEST DATASET

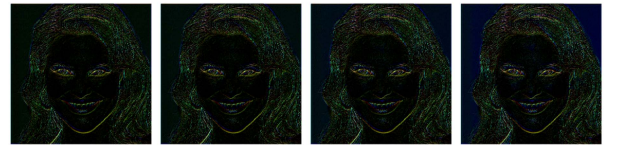
Strategies	PSNR	LPIPS
No guidance during sampling	24.00	0.2685
$L_{pix,img}$	24.52	0.2535
$L_{pix,y}$	25.24	0.2452
$L_{pix,y} + L_{pix,img}$	25.21	0.2441
$L_{pix,y} + L_{fea,img}$ (default)	25.00	0.2323
$L_{pix,y} + L_{fea,img} + L_{pix,img}$	24.95	0.2330

Fig. 17. Some examples where our algorithm with the *pre-trained* INN does not perform well under severe and complex real-world degradation conditions (second column). This issue is fixed with our *fine-tuning* strategy (third column).

(a) Low-resolution inputs with different degradation levels.



(b) Coarse parts generated by forward INN with different conditions.



(c) Detail parts generated by forward INN with different conditions.

Fig. 18. Effect of our INN.

parameters of our INN need to be refined to simulate the degradation process more accurately. We achieve this by finetuning the parameters of our INN at testing stage. We investigate the effect of finetuning in our BlindINDIGO in Table VII. One can observe that our finetuning strategy contributes to performance

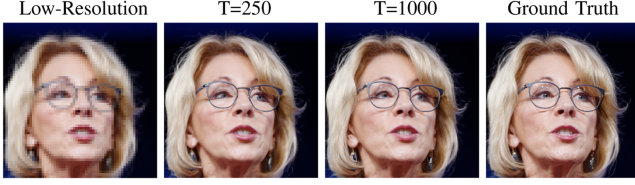


Fig. 19. Results of our approach with $T = 1000$ and $T = 250$ (with DDIM) on 4x super-resolution task.

improvement in both synthetic unseen degradation (cases 6 and 7) and real-world unseen degradation (cases 9 and 10). As shown in Fig. 17, without finetuning, the output image becomes blurry due to the inaccurate simulation of the degradation process. Finetuning fixes this issue.

V. CONCLUSION

In this paper, we have introduced a novel approach that fully leverages the power of pre-trained generative diffusion models for inverse problems. We achieve this by introducing an INN that enforces that the generative process of the diffusion model be consistent with the measurements. This leads to a simple way to effectively sample from the posterior rather than the prior as in unconditional diffusion.

Besides being very effective, the approach is extremely flexible since the degradation process can be learned from data and refined at testing stage if necessary. In the non-blind case, since we pre-train the forward process of INN to simulate an arbitrary degradation process, we are no longer limited by the requirement of knowing the analytical expression of the degradation model and we can handle highly non-linear degradation processes. In the blind case, we can handle unknown degradations due to our approach that at testing stage alternately refine the INN to better simulate the unknown degradation and update intermediate results with the guidance of INN during reverse diffusion sampling.

Experiments demonstrate that our algorithm obtains competitive results both quantitatively and visually on synthetic and real-world low-quality images.

APPENDIX A CONSENSUS STRATEGY

We propose a novel consensus strategy for our INDIGO. Our insight is that the Langevin iteration in (14) has a random term $\mathbf{z}$. We can therefore create several parallel versions of that iteration by using different realization of $\mathbf{z}$. In this way our method estimate several enhanced versions of the corrupted image that can then be combined. However, instead of directly averaging the outputs of our algorithm, we adopt an averaging operation during the guidance of the gradient after each sampling step as shown in Algorithm 3. It is noteworthy that our strategy supports averaging multiple results. However, we show only the case of averaging two results in the algorithm for simplicity.

In Table IX, we show our INDIGO with up to four parallel versions with our consensus strategy in the non-blind case. It

Algorithm 3: INDIGO with Consensus Strategy.

Input: Corrupted image $\mathbf{y}$, gradient scale ζ , pretrained INN $f_\phi(\cdot)$.

Output: Output image $\mathbf{x}_0$ conditioned on $\mathbf{y}$

$\mathbf{x}_T^1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

$\mathbf{x}_T^2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for t from T to 1 **do**

$\mathbf{z}^1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z}^1 = \mathbf{0}$

$\mathbf{z}^2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z}^2 = \mathbf{0}$

$\mathbf{x}_{0,t}^1 = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t^1 - \sqrt{1 - \alpha_t}\epsilon_\theta(\mathbf{x}_t^1, t))$

$\mathbf{x}_{0,t}^2 = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t^2 - \sqrt{1 - \alpha_t}\epsilon_\theta(\mathbf{x}_t^2, t))$

$\hat{\mathbf{x}}_{t-1}^1 = \frac{\sqrt{\alpha_t}(1 - \alpha_{t-1})}{1 - \alpha_t}\mathbf{x}_t^1 + \frac{\sqrt{\alpha_{t-1}\beta_t}}{1 - \alpha_t}\mathbf{x}_{0,t}^1 + \sigma_t\mathbf{z}^1$

$\hat{\mathbf{x}}_{t-1}^2 = \frac{\sqrt{\alpha_t}(1 - \alpha_{t-1})}{1 - \alpha_t}\mathbf{x}_t^2 + \frac{\sqrt{\alpha_{t-1}\beta_t}}{1 - \alpha_t}\mathbf{x}_{0,t}^2 + \sigma_t\mathbf{z}^2$

$\mathbf{c}_t^1, \mathbf{d}_t^1 = f_\phi(\mathbf{x}_{0,t}^1)$

$\mathbf{c}_t^2, \mathbf{d}_t^2 = f_\phi(\mathbf{x}_{0,t}^2)$

$\hat{\mathbf{x}}_{0,t}^1 = f_\phi^{-1}(\mathbf{y}, \mathbf{d}_t^1)$

$\hat{\mathbf{x}}_{0,t}^2 = f_\phi^{-1}(\mathbf{y}, \mathbf{d}_t^2)$

$\mathbf{x}_{t-1}^1 = \hat{\mathbf{x}}_{t-1}^1 - \zeta \nabla_{\mathbf{x}_t^1} (\|\hat{\mathbf{x}}_{0,t}^1 - \mathbf{x}_{0,t}^1\|_2^2 + \|\hat{\mathbf{x}}_{0,t}^2 - \mathbf{x}_{0,t}^1\|_2^2)$

$\mathbf{x}_{t-1}^2 = \hat{\mathbf{x}}_{t-1}^2 - \zeta \nabla_{\mathbf{x}_t^2} (\|\hat{\mathbf{x}}_{0,t}^2 - \mathbf{x}_{0,t}^2\|_2^2 + \|\hat{\mathbf{x}}_{0,t}^1 - \mathbf{x}_{0,t}^2\|_2^2)$

end

return $\mathbf{x}_0$

TABLE IX
ABLATION STUDY ON OUR CONSENSUS STRATEGY ON NON-BLIND 4X SUPER-RESOLUTION

Noise level	Strategies	PSNR	PSNR gain
30	Baseline	23.97	–
	Averaging 2 results	24.12	0.15
	Averaging 3 results	24.51	0.54
	Averaging 4 results	24.34	0.37
50	Baseline	22.35	–
	Averaging 2 results	22.56	0.21
	Averaging 3 results	23.03	0.68
	Averaging 4 results	22.91	0.56
80	Baseline	20.51	–
	Averaging 2 results	20.91	0.40
	Averaging 3 results	21.50	0.99
	Averaging 4 results	21.48	0.97

TABLE X
ABLATION STUDY ON THE CONSENSUS STRATEGY OF OUR BLINDINDIGO ON 4X BLIND SR WITH MEDIUM DEGRADATION

	PSNR $\uparrow$	LPIPS $\downarrow$	Time
w/o Consensus	25.31	0.2442	9.25s
w/ Consensus	25.26	0.2413	19.98s

can be seen that our consensus strategy with 3 parallel versions achieves the best performance. Also, we can observe that as the noise level increases, the PSNR gain brought by our strategy is more significant. Therefore, in the non-blind case, we apply our consensus strategy with 3 parallel results. In the blind case, as shown in Table X, we observe that although our consensus strategy brings gains in terms of LPIPS, it also has an impact on the runtime. Taking into account the tradeoff between improved image quality and the additional computational time required, we decide not to employ our consensus strategy in the blind case.

REFERENCES

- [1] Z. Wang et al., “DR2: Diffusion-based robust degradation remover for blind face restoration,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 1704–1713.
- [2] Z. Yue and C. C. Loy, “Difface: Blind face restoration with diffused error contraction,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–15, 2024.
- [3] P. Yang, S. Zhou, Q. Tao, and C. C. Loy, “PGDiff: Guiding diffusion models for versatile face restoration via partial guidance,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2023, pp. 32194–32214.
- [4] J. Wang, Z. Yue, S. Zhou, K. C. Chan, and C. C. Loy, “Exploiting diffusion prior for real-world image super-resolution,” *Int. J. Comput. Vis.*, pp. 1–21, 2024.
- [5] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, “Deep learning techniques for inverse problems in imaging,” *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 39–56, May 2020.
- [6] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 945–948.
- [7] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, “Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications,” *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 85–97, Jan. 2023.
- [8] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 6840–6851.
- [9] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–22.
- [10] P. Dhariwal and A. Nichol, “Diffusion models beat gans on image synthesis,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 8780–8794.
- [11] Y. Song and S. Ermon, “Generative modeling by estimating gradients of the data distribution,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2019, pp. 11895–11907.
- [12] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” in *Proc. Int. Conf. Learn. Representations*, 2022, pp. 1–36.
- [13] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, “ILVR: Conditioning method for denoising diffusion probabilistic models,” in *Proc. Int. Conf. Comput. Vis.*, 2021, pp. 14347–14356.
- [14] H. Chung, B. Sim, and J. C. Ye, “Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12403–12412.
- [15] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. V. Gool, “RePaint: Inpainting using denoising diffusion probabilistic models,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11451–11461.
- [16] Z. Kadkhodaie and E. P. Simoncelli, “Solving linear inverse problems using the prior implicit in a denoiser,” in *Proc. NeurIPS Workshop Deep Learn. Inverse Problems*, 2020, pp. 1–19.
- [17] B. Kawar, G. Vaksman, and M. Elad, “SNIPS: Solving noisy inverse problems stochastically,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 21757–21769.
- [18] B. Kawar, M. Elad, S. Ermon, and J. Song, “Denoising diffusion restoration models,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2022, pp. 23593–23606.
- [19] Y. Wang, J. Yu, and J. Zhang, “Zero-shot image restoration using denoising diffusion null-space model,” in *Proc. Int. Conf. Learn. Representations*, 2023, pp. 1–31.
- [20] Y. Song, L. Shen, L. Xing, and S. Ermon, “Solving inverse problems in medical imaging with score-based generative models,” in *Proc. Int. Conf. Learn. Representations*, 2022, pp. 1–18.
- [21] J. Song, A. Vahdat, M. Mardani, and J. Kautz, “Pseudoinverse-guided diffusion models for inverse problems,” in *Proc. Int. Conf. Learn. Representations*, 2023, pp. 1–30.
- [22] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, “Diffusion posterior sampling for general noisy inverse problems,” in *Proc. Int. Conf. Learn. Representations*, 2023, pp. 1–30.
- [23] J. Liu et al., “DOLCE: A model-based probabilistic diffusion framework for limited-angle ct reconstruction,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 10498–10508.
- [24] D. You, A. Floros, and P. L. Dragotti, “INDigo: An INN-guided probabilistic diffusion algorithm for inverse problems,” in *2023 IEEE 25th Int. Workshop Multimedia Signal Process.*, 2023, pp. 1–6.
- [25] H. Chung, J. Kim, S. Kim, and J. C. Ye, “Parallel diffusion models of operator and image for blind inverse problems,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 6059–6069.
- [26] N. Murata et al., “GibbsDDRM: A partially collapsed gibbs sampler for solving blind inverse problems with denoising diffusion restoration,” in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 25501–25522.
- [27] B. Fei et al., “Generative diffusion prior for unified image restoration and enhancement,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 9935–9946.
- [28] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10674–10685.
- [29] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, “Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 5775–5787.
- [30] J. Ho and T. Salimans, “Classifier-free diffusion guidance,” in *Proc. NeurIPS Workshop Deep Generative Models Downstream Appl.*, 2021, pp. 1–8.
- [31] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, “Image super-resolution via iterative refinement,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2023.
- [32] C. Saharia et al., “Palette: Image-to-image diffusion models,” in *Proc. ACM SIGGRAPH 2022 Conf. Proc.*, 2022, pp. 1–10.
- [33] H. Sahak, D. Watson, C. Saharia, and D. Fleet, “Denoising diffusion probabilistic models for robust image super-resolution in the wild,” 2023, *arXiv:2302.07864*.
- [34] H. Li et al., “SRDiff: Single image super-resolution with diffusion probabilistic models,” *Neurocomputing*, vol. 479, pp. 47–59, 2022.
- [35] X. Qiu, C. Han, Z. Zhang, B. Li, T. Guo, and X. Nie, “DiffBFR: Bootstrapping diffusion model for blind face restoration,” in *Proc. 31st ACM Int. Conf. Multimedia*, New York, NY, USA, 2023, pp. 7785–7795.
- [36] I. Daubechies and W. Sweldens, “Factoring wavelet transforms into lifting steps,” *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.
- [37] J.-J. Huang and P. L. Dragotti, “WINNet: Wavelet-inspired invertible network for image denoising,” *IEEE Trans. Image Process.*, vol. 31, pp. 4377–4392, 2022.
- [38] L. Dinh, D. Krueger, and Y. Bengio, “Nice: Non-linear independent components estimation,” in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–13.
- [39] L. Dinh, J. Sohl-Dickstein, and S. Bengio, “Density estimation using real NVP,” 2016, *arXiv:1605.08803*.
- [40] A. N. Gomez, M. Ren, R. Urtasun, and R. B. Grosse, “The reversible residual network: Backpropagation without storing activations,” *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 2214–2224, 2017.
- [41] J.-H. Jacobsen, A. W. Smeulders, and E. Oyallon, “i-RevNet: Deep invertible networks,” in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–11.
- [42] J. He, C. Dong, and Y. Qiao, “Interactive multi-dimension modulation with dynamic controllable residual learning for image restoration,” in *Proc. Comput. Vis.—ECCV 2020, 16th Eur. Conf. Proc., Part XX 16*, Aug. 23–28, 2020, pp. 53–68.
- [43] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [44] P. Wei et al., “Component divide-and-conquer for real-world image super-resolution,” in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 101–117.
- [45] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.
- [46] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, “Swinir: Image restoration using swin transformer,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1833–1844.
- [47] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4396–4405.
- [48] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of GANs for improved quality, stability, and variation,” in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–26.
- [49] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [50] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6626–6637.

- [51] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.
- [52] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [53] B. Xia et al., "Knowledge distillation based degradation estimation for blind super-resolution," in *Proc. Int. Conf. Learn. Representations*, 2023, pp. 1–18.
- [54] R. Timofte, E. Agustsson, L. V. Gool, M.-H. Yang, and L. Zhang, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1110–1121.
- [55] C. Chen, D. Gong, H. Wang, Z. Li, and K.-Y. K. Wong, "Learning spatial attention for face super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 1219–1231, 2021.
- [56] T. Yang, P. Ren, X. Xie, and L. Zhang, "GAN prior embedded network for blind face restoration in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 672–681.
- [57] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9164–9174.
- [58] S. Zhou, K. C. Chan, C. Li, and C. C. Loy, "Towards robust blind face restoration with codebook lookup transformer," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2022, pp. 30599–30611.
- [59] Y. Gu et al., "VQFR: Blind face restoration with vector-quantized dictionary and parallel decoder," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 126–143.
- [60] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [61] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4685–4694.
- [62] G.-H. Liu, A. Vahdat, D.-A. Huang, E. A. Theodorou, W. Nie, and A. Anandkumar, "I²sb: Image-to-image schrödinger bridge," in *Proc. 40th Int. Conf. Mach. Learn.*, 2023, pp. 22 042–22 062.



Pier Luigi Dragotti (Fellow, IEEE) received the Laurea (*summa cum laude*) degree in electronic engineering from the University of Naples Federico II, Naples, Italy, in 1997, and the master's degree in communications systems and the Ph.D. degree from the Swiss Federal Institute of Technology of Lausanne (EPFL), Lausanne, Switzerland, in 1998 and 2002, respectively. He has held several visiting positions, in particular, he was a visiting student with Stanford University, Stanford, CA, USA, in 1996, Summer

Researcher with the Mathematics of Communications Department, Bell Labs, Murray Hill, NJ, USA, in 2000, Visiting Scientist with the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2011, and Visiting Scholar with Trinity College, Cambridge, U.K., in 2020. Before joining Imperial College London, London, U.K., in 2002, he was a Senior Researcher with EPFL working on distributed signal processing for the Swiss National Competence Center in Research on Mobile Information and Communication Systems. He is currently a Professor of signal processing with the Department of Electrical and Electronic Engineering, Imperial College London. His research interests include sampling theory and its applications, computational imaging, and model-based deep learning. Dr. Dragotti was an Elected Member of the IEEE Image, Video and Multidimensional Signal Processing Technical Committee, Elected Member of the IEEE Signal Processing Theory and Methods Technical Committee, and IEEE Computational Imaging Technical Committee. In 2011, he was the recipient of the Prestigious ERC Starting Investigator Award (consolidator stream). He was also IEEE SPS Distinguished Lecturer (2021–2022), Editor-in-Chief of IEEE TRANSACTIONS ON SIGNAL PROCESSING (2018–2020), Technical Co-Chair of the European Signal Processing Conference in 2012 and Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING from 2006 to 2009.



Di You (Graduate Student Member, IEEE) received the bachelor's degree in electronic and information engineering from the Dalian University of Technology, Dalian, China, in 2019, and the master's degree in computer applications technology from Peking University, Shenzhen, China, in 2022. She is currently working toward the Doctoral degree with Electrical and Electronic Engineering Department, Imperial College London, London, U.K., under the supervision of Prof. Pier Luigi Dragotti. Her research interests include the areas of computer vision, signal processing, and deep learning, specifically for inverse problems. She was awarded the President's Ph.D. Scholarship by Imperial College London.

ing, and deep learning, specifically for inverse problems. She was awarded the President's Ph.D. Scholarship by Imperial College London.