Deep Equilibrium Diffusion Restoration with Parallel Sampling

Jiezhang Cao[1], Yue Shi[1,2], Kai Zhang[3,*], Yulun Zhang[2], Radu Timofte[1,4], Luc Van Gool[1,5]
[1]ETH Zürich, [2]Shanghai Jiao Tong University, [3]Nanjing University, [4]University of Würzburg, [5]KU Leuven
https://github.com/caojiezhang/DeqIR

Abstract

Diffusion model-based image restoration (IR) aims to use diffusion models to recover high-quality (HQ) images from degraded images, achieving promising performance. Due to the inherent property of diffusion models, most existing methods need long serial sampling chains to restore HQ images step-by-step, resulting in expensive sampling time and high computation costs. Moreover, such long sampling chains hinder understanding the relationship between inputs and restoration results since it is hard to compute the gradients in the whole chains. In this work, we aim to rethink the diffusion model-based IR models through a different perspective, i.e., a deep equilibrium (DEQ) fixed point system, called ***DeqIR****. Specifically, we derive an analytical solution by modeling the entire sampling chain in these IR models as a joint multivariate fixed point system. Based on the analytical solution, we can conduct parallel sampling and restore HQ images without training. Furthermore, we compute fast gradients via DEQ inversion and found that initialization optimization can boost image quality and control the generation direction. Extensive experiments on benchmarks demonstrate the effectiveness of our method on typical IR tasks and real-world settings.*

1. Introduction

Image restoration (IR) aims at recovering a high-quality (HQ) image from a degraded input. Recently, diffusion models [37, 61] are attracting great attention because they can generate higher quality images than GANs [23] and likelihood-based models [44]. Based on diffusion models [37, 61], many IR methods [20, 41, 68] achieve compelling performance on different tasks. Directly using diffusion models in IR, however, suffers from some limitations.

First, diffusion model-based image restoration (DMIR) models rely on a long sampling chain to synthesize HQ images step-by-step, as shown in Figure 2 (a). As a result, it will lead to expensive sampling time during the inference. For example, DPS [20] based on DDPM [37] needs 1k sampling steps. To accelerate the sampling, some DMIR methods [41, 68, 87] use DDIM [61] to make a trade-off between computational cost and the restoration quality. Based on this, these methods can reduce sampling steps to 100 or even fewer. Unfortunately, it may degrade the sample quality when reducing the sampling steps [52]. It raises an interesting question: *is it possible to develop an alternative sampling method without sacrificing the sample quality?*

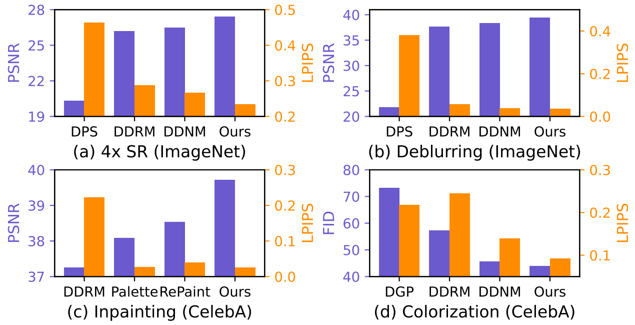
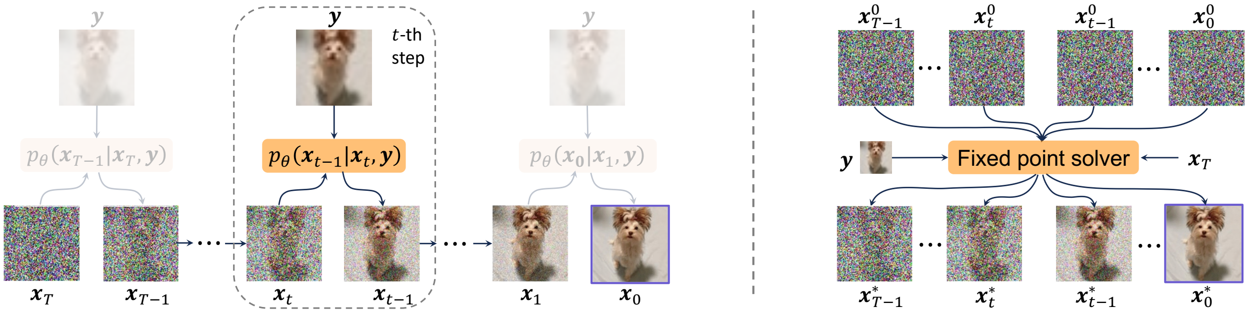


Figure 1. Comparisons of different zero-shot DMIR methods in various IR applications on different datasets.

Second, the long sampling chain makes understanding the relationship between the restoration and inputs difficult. In practice, sampling different Gaussian noises as inputs may have diverse results for some IR tasks (*e.g.*, inpainting and colorization). Such diversity is not necessary for some IR tasks, *e.g.*, super-resolution (SR) or deblurring. Nevertheless, different initializations may affect the quality of SR and deblurring. It raises the second question: *is it possible to optimize the initialization such that the generation can be improved or controlled?* However, it is difficult for existing methods to compute the gradient along the long sampling chain as they require storing the entire computational graph.

In this paper, we rethink the sampling process in IR from a deep equilibrium (DEQ) based on [57]. Specifically, we first derive a proposition to model the sampling chain as a fixed point system, achieving parallel sampling. Then, we use a DEQ solver to find the fixed point of the sampling chain. Last, we use modern automatic differentiation packages to compute the gradients with backpropagating and understand the relationship between input noise and restoration.

[*] Corresponding author.



(a) Most existing diffusion model-based IR (**sequential sampling**)

(b) Our zero-shot IR (**parallel sampling**)

Figure 2. Comparisons of sequential sampling and our parallel sampling.

We summarize our contributions as follows:

- We prove that the long sampling chain in DMIR can be formulated in a parallel way. Then we analytically formulate the generative process as a deep equilibrium fixed point system. Moreover, the generation has a convergence guarantee with few timesteps and iterations.
- Compared with most existing DMIR methods with sequential sampling, our method is able to achieve parallel sampling, as shown in Figure 2 (b). Moreover, our method can be run on multiple GPUs instead of a single GPU.
- Our model has more efficient gradients using DEQ inversion than existing DMIR methods which need a large computational graph for storing intermediate variables. The gradients can be computed through standard automatic differentiation packages. Moreover, we found that the initialization can be optimized with the gradients to improve the image quality and control the generation direction.
- Extensive experiments on benchmarks demonstrate the effectiveness of our zero-shot method on different IR tasks, as shown in Figure 1. Moreover, our method performs well in real-world applications that may contain unknown and non-linear degradations.

2. Related Work

Deep implicit learning (DIL). DIL attracts more and more attention and has emerging applications. Different from explicit learning, DIL is based on dynamical systems, *e.g.*, optimization [1, 24, 27, 32, 62], differential equation [17, 28, 34], or fixed-point system [3, 4, 35]. For the fixed-point system, DEQ [3] is a new type of implicit model and it models sequential data by directly finding the fixed point and optimizing this equilibrium. Recently, DEQ has been widely used in different tasks, *e.g.*, semantic segmentation [4], object detection [64, 65], robustness [46, 69, 75], optical flow estimation [5], and generative models like normalizing flow [51]. Notably, DEQ-DDIM [57] apply DEQs to diffusion models [37] by formulating this process as an equilibrium system. However, applying DEQs in diffusion model-based IR methods is non-trivial because the generative process is complex, and formulating such a process is very challenging.

Diffusion model-based image restoration. Previous image restoration (IR) methods [25, 26] use convolutional neural networks (CNN) to achieve impressive performance on IR. Up to now, many researchers propose to design the network architecture using residual blocks [14, 42, 82], GANs [7–9, 36, 54, 56, 66, 67, 83], attention [10–13, 15, 16, 21, 47–49, 71, 74, 76–78, 80, 85], and others [29, 30, 38, 43, 73, 81] to improve the IR performance.

Recently, denoising diffusion probabilistic models (DDPM) [37] developed a powerful class of generative models that can synthesize high-quality images [23] from noise step-by-step. Based on the diffusion models, existing IR methods [20, 41, 68] can be divided into supervised methods and zero-shot methods. The supervised methods aim to train a conditional diffusion model in the image space [59, 60, 70, 79] or the latent space [50, 58, 63, 72]. However, these methods need training diffusion models for the specific degradations and have limited generalization performance to other degradations in different IR tasks.

For zero-shot IR methods, they use a pre-trained diffusion model (*e.g.*, DDPM [37] and DDIM [61]) to restore images without training [20, 41, 68]. For example, based on a given reference image, ILVR [19] guides the generative process in DDPM and generates high-quality images. Based on DDPM, DPS [20] solves the inverse problems via approximation of the posterior sampling using 1000 steps of the manifold-constrained gradient. Similar to DPS, DiffPIR [87] integrates the traditional plug-and-play method into the diffusion models. Repaint [53] also employs a pre-trained DDPM as the generative prior for the image inpainting task. To accelerate the sampling, there are some IR methods using DDIM. For example, DDRM [41] applies a pre-trained denoising diffusion generative model to solve a linear inverse problem with 20 sampling steps. This method uses SVD on the degradation operator, which is similar to SNIPS [40]. Based on SVD, DDNM [68] applies range-null space decomposition in linear image inverse problem and refines the null-space iteratively. Here, DDNM uses DDIM as the base sampling strategy with 100 sampling steps. However, all of these methods use the serial sampling chain, resulting in a long sampling time and expensive computational cost.

3. Preliminaries

Image restoration. Image restoration aims at synthesizing high-quality image $\hat{\boldsymbol{x}}$ from a degraded observation $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}) + \boldsymbol{n}_\sigma$, where $\boldsymbol{A}$ is some degradation (*e.g.*, bicubic), $\boldsymbol{x}$ is the original image, and $\boldsymbol{n}_\sigma$ is a non-linear noise (*e.g.*, white Gaussian noise) with the level σ. The solution can be obtained by optimizing the following problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} 1/2\sigma^2 \left\|\boldsymbol{A}(\boldsymbol{x}) - \boldsymbol{y}\right\|_2^2 + \lambda\mathcal{R}(\boldsymbol{x}), \tag{1}$$

where $\mathcal{R}(\boldsymbol{x})$ is a reguralization term with a trade-off parameter λ, *e.g.*, sparsity and Tikhonov regularization.

Diffusion models. DDPM [37] is a generative model that can synthesize high-quality images with a forward process (*i.e.*, diffusion process) and a reverse process. The forward process gradually introduces noise from Gaussian distribution $\mathcal{N}(\cdot)$ with specific noise levels to the data, *i.e.*,

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\boldsymbol{I}\right), \tag{2}$$

where $\bar{\alpha}_t := \Pi_{s=1}^t \alpha_s$, $\alpha_t := 1 - \beta_t$ and β_t is a variance. For the reverse process, the previous state $\boldsymbol{x}_{t-1}$ can be predicted with $\tilde{\boldsymbol{\mu}}_t$ and $\tilde{\sigma}_t$, which is formulated as:

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0), \tilde{\sigma}_t^2\boldsymbol{I}\right), \tag{3}$$

where $\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t, \boldsymbol{x}_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon})$ and $\tilde{\sigma}_t^2 := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. Here, the noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$ can be estimated by $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)$ in each time-step. To apply $\tilde{\boldsymbol{\mu}}_t$ to the image inverse problem, one can replace $\boldsymbol{x}_0$ with $\hat{\boldsymbol{x}}_{0|t}$ conditioned on the degraded image $\boldsymbol{y}$, *i.e.*,

$$\boldsymbol{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\boldsymbol{x}}_{0|t} + \frac{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t + \tilde{\sigma}_t\boldsymbol{\epsilon}, \tag{4}$$

where $\hat{\boldsymbol{x}}_{0|t}$ can be estimated by using a degradation $\boldsymbol{A}$ to map the denoised image $\boldsymbol{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t))$ in the degradation space [68], *i.e.*,

$$\hat{\boldsymbol{x}}_{0|t} = \boldsymbol{A}^\dagger\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A})\boldsymbol{x}_{0|t}, \tag{5}$$

where $\boldsymbol{A}^\dagger$ is the pseudo-inverse of $\boldsymbol{A}$.

Deep equilibrium models. Deep equilibrium models (DEQs) [3] are infinite depth feed-forward networks that can find fixed points in the forward pass. Given an input injection $\boldsymbol{x}$, an hidden state $\boldsymbol{\nu}^{k+1}$ can be predicted by using an equilibrium layer f_θ parametrized by θ, *i.e.*,

$$\boldsymbol{\nu}^{k+1} = f_\theta\left(\boldsymbol{\nu}^k; \boldsymbol{x}\right), k = 0, \ldots, L{-}1. \tag{6}$$

When increasing the depth towards infinity, the model tends to converge to a fixed point (equilibrium) $\boldsymbol{\nu}^*$, *i.e.*,

$$\lim_{k\to\infty} f_\theta\left(\boldsymbol{\nu}^k; \boldsymbol{x}\right) = f_\theta\left(\boldsymbol{\nu}^*; \boldsymbol{x}\right) = \boldsymbol{\nu}^*. \tag{7}$$

To solve the equilibrium state $\boldsymbol{\nu}^*$, one can use some fixed point solvers, like Broyden's method [6], or Anderson acceleration [2], and it can be accelerated by the neural solver [17] in the inference.

4. Methodology

4.1. Deep Equilibrium Diffusion Restoration

Most existing zero-shot IR methods [20, 41, 68] restore high-quality images step-by-step with long serial sampling chains. Such an inherent property comes from the diffusion models, and it will lead to expensive sampling time and high computation costs. This issue may be intractable if we need a gradient by backpropagating through the long sampling chains which often result in out-of-memory in the experiments. To address this issue, we present a main modeling contribution in this paper.

Fixed point modeling. Motivated by [57], our goal is to formulate diffusion model-based IR as a deep equilibrium fixed point system. Specifically, given a degraded image $\boldsymbol{y}$ and Gaussian noise $\boldsymbol{x}_T$, the sampling chain $\boldsymbol{x}_{0:T-1}$ can be treated as multivariable of the DEQ fixed point system, we first formulate $\boldsymbol{x}_{0:T-1}$ as follows:

$$\boldsymbol{x}_{0:T-1} = F(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y})), \tag{8}$$

where $\boldsymbol{x}_T \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{y}$ are the input injections, and $F(\cdot)$ is a function that performs sequential data across all the sample steps simultaneously. To formulate the function F in Eqn. (8), we first provide the following proposition for the parallel sampling.

Proposition 1 **(Parallel sampling)** *Given a degradation matrix $\boldsymbol{A}$, a degraded image $\boldsymbol{y}$ and a Gaussian noise image $\boldsymbol{x}_T \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, for $k \in [1, \ldots, T]$, the state $\boldsymbol{x}_{T-k}$ can be predicted by previous states $\{\boldsymbol{x}_{T-k+1}, \ldots, \boldsymbol{x}_T\}$, i.e.,*

$$\begin{aligned}\boldsymbol{x}_{T-k} =& \frac{\sqrt{\bar{\alpha}_{T-k}}}{\sqrt{\bar{\alpha}_T}}\left(\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A}\right)\boldsymbol{x}_T + \boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{z}_{T-k+1} \\ &+ \sum_{s=T-k}^{T-1}\frac{\sqrt{\bar{\alpha}_{T-k}}}{\sqrt{\bar{\alpha}_s}}\left(\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A}\right)\boldsymbol{z}_{s+1},\end{aligned} \tag{9}$$

where $\boldsymbol{z}_s = c_s^0\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_s, s) + \sqrt{\bar{\alpha}_{s-1}}\boldsymbol{A}^\dagger\boldsymbol{y} + c_s^1\boldsymbol{\epsilon}_s$, the coefficients are defined as $c_s^0 := c_s^2 - \sqrt{(1-\bar{\alpha}_s)/\alpha_s}(\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A})$, $c_s^1 := \sqrt{1-\bar{\alpha}_s}\eta$ and $c_s^2 := \sqrt{1-\bar{\alpha}_s}\sqrt{1-\eta^2}, 0 \leq \eta < 1$.

Proof Please refer to the proofs in Supplementary. □

From the proposition, $\boldsymbol{x}_{T-k}$ is related to subsequent states $\boldsymbol{x}_{T-k+1:T}$ and the degraded image $\boldsymbol{y}$. It means that our method is different from most existing diffusion model-based IR methods which update $\boldsymbol{x}_t$ based only on $\boldsymbol{x}_{t+1}$. Based on our proposition, the timestep T can be small using DDIM [61]. In addition, the proposition can be extended to start from the intermediate state. Motivated by [79], we can predict the intermediate state using a restoration model (*e.g.*, [18, 48, 84, 86]) to provide prior information from the restoration model during the sampling processing when the degradation matrix $\boldsymbol{A}$ is unknown or inaccurate.

Algorithm 1 Implementation of $\mathrm{RootSolve}(\cdot)$

```
Require: A degraded image $\boldsymbol{y}$, a pre-trained diffusion model, timesteps $T$, iterations $K$, an integer parameter $m \geq 1$
Initialize $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{x}_i^{(0)} = \boldsymbol{x}_T, i = 0, \dots, T-1$
Calculate $\mathbf{x}_{0:T-1}^{(1)} = F\left(\mathbf{x}_{0:T-1}^{(0)}; (\boldsymbol{x}_T, \boldsymbol{y})\right)$
for $k$ from 1 to $K$ do
    $m_k = \min\{m, k\}$
    $\boldsymbol{G}_k = [g_{k-m_k}, \dots, g_k]$
    Solve least-squares problem for $\boldsymbol{\alpha}=[\alpha_0, \dots, \alpha_{m_k}]$
        $\boldsymbol{\alpha}_k = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{G}_k \boldsymbol{\alpha}\|_2, s.t., \boldsymbol{\alpha}^\top \mathbf{1} = 1$
    Update the sequence
    $\boldsymbol{x}_{0:T-1}^{(k+1)} = \sum_{i=0}^{m_k} (\boldsymbol{\alpha}_k)_i F\left(\boldsymbol{x}_{0:T-1}^{(k-m_k+i)}; (\boldsymbol{x}_T, \boldsymbol{y})\right)$
end for
return $\boldsymbol{x}_0^* := \boldsymbol{x}_0^{K+1}$
```

Based on our proposed proposition, we can formulate the right side of Eqn. (9) as $\boldsymbol{x}_{T-k} = f(\boldsymbol{x}_{T-k+1:T}; \boldsymbol{y})$. Then, we can write all sampling steps as a "fully-lower-triangular" inference process, *i.e.*,

$$\begin{bmatrix} \boldsymbol{x}_{T-1} \\ \boldsymbol{x}_{T-2} \\ \vdots \\ \boldsymbol{x}_0 \end{bmatrix} = \begin{bmatrix} f(\boldsymbol{x}_T; \boldsymbol{y}) \\ f(\boldsymbol{x}_{T-1:T}; \boldsymbol{y}) \\ \vdots \\ f(\boldsymbol{x}_{1:T}; \boldsymbol{y}) \end{bmatrix}, \tag{10}$$

where the function f can be implemented in all sequential states in parallel, corresponding to Eqn. (8), *i.e.*, $\boldsymbol{x}_{0:T-1} = F(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y}))$. To find the solution to the fixed point of Eqn. (10), we apply commonly used fixed point solvers like Anderson acceleration [2] which can accelerate the convergence of the fixed-point sequence. To this end, we first define the residual $g(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y})) = F(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y})) - \boldsymbol{x}_{0:T-1}$. Then, we can directly input the residual to the Anderson acceleration solver and obtain the final converged fixed point, *i.e.*,

$$\boldsymbol{x}^*_{0:T-1} = \mathrm{RootSolve}\left(g(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y}))\right), \tag{11}$$

where $\boldsymbol{x}_0^*$ is our desired result at the end of sampling, and $\mathrm{RootSolve}(\cdot)$ is a fixed point solver using Anderson acceleration, which is implemented in Algorithm 1. For convenience, we define $g_k := g(\boldsymbol{x}_{0:T-1}^{(k)}; (\boldsymbol{x}_T, \boldsymbol{y}))$. Note that $\mathrm{RootSolve}(\cdot)$ can be implemented in the PyTorch package, and we use the same hyper-parameters as [57]. Moreover, Algorithm 1 is guaranteed to converge to a fixed point, which is verified in the experiment sections. Note that we do not train all functions and diffusion models.

Compared with most existing diffusion model-based IR methods [20, 68], our method operating all states in parallel results in more accurate estimations of the intermediate latent states $\boldsymbol{x}_t$, requiring fewer sampling steps. It implies that we are able to obtain the better final sample $\boldsymbol{x}_0^*$ based on these accurately estimated intermediate latent states $\boldsymbol{x}_t$.

Algorithm 2 Initialization Optimization via DEQ inversion

```
Require: A degraded image $\boldsymbol{y}$, a pre-trained diffusion model, update rate $\lambda$, total steps $S$.
Initialize $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \boldsymbol{x}_i = \boldsymbol{x}_T, i = 0, \dots, T-1$
for steps from 1 to $S$ do
    Disable gradient computation, and compute $\boldsymbol{x}^*_{0:T-1}$ according to Algorithm 1
        $\boldsymbol{x}^*_{0:T-1} = \mathrm{RootSolve}\left(g(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y}))\right)$
    Enable gradient computation, and compute loss and use the 1-step grad $\partial\mathcal{L}/\partial\boldsymbol{x}_T$
    Update $\boldsymbol{x}_T$ with a gradient descent:
        $\boldsymbol{x}_T \leftarrow \boldsymbol{x}_T + \lambda\partial\mathcal{L}/\partial\boldsymbol{x}_T$
end for
return $\boldsymbol{x}_T$
```

4.2. Initialization Optimization via DEQ Inversion

Different initializations have diverse generations in some IR tasks, *e.g.*, colorization and inpainting. However, such diversity of generation is hard to control, and it is harmful to SR or deblurring which requires guaranteeing the identity. To address this, we provide an interesting perspective to explore the initialization of our diffusion model.

To achieve this, we first define a general loss function that can provide additional information. Specifically, given a degraded image $\boldsymbol{y}$ and the output of $\mathrm{RootSolve}$, *i.e.*, $\boldsymbol{x}_0^*$, then the loss can be defined as

$$\mathcal{L} = \ell\left(\phi(\boldsymbol{x}_0^*), \varphi(\boldsymbol{y})\right), \tag{12}$$

where ℓ can be L_2 loss or perceptual loss. For example, ϕ can be $\boldsymbol{A}$ and φ is an identity function; or ϕ is an identity function and φ is a pre-trained IR model [18, 48]. Based on the loss, we apply the implicit function theorem to compute the gradients of the loss $\mathcal{L}$ *w.r.t.* $\boldsymbol{x}_T$, *i.e.*,

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{x}_T} = -\frac{\partial\mathcal{L}}{\partial\boldsymbol{x}^*_{0:T}}\left(J_g^{-1}\big|_{\boldsymbol{x}^*_{0:T}}\right)\frac{\partial F(\boldsymbol{x}^*_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y}))}{\partial\boldsymbol{x}_T}, \tag{13}$$

where $J_g^{-1}\big|_{\boldsymbol{x}^*_{0:T}}$ is inverse Jacobian of $g(\boldsymbol{x}_{0:T-1}; (\boldsymbol{x}_T, \boldsymbol{y}))$ evaluated at $\boldsymbol{x}^*_{0:T}$. In practical, we use an approximation version, *i.e.*, $\boldsymbol{M} \approx J_g^{-1}|_{\boldsymbol{x}^*_{0:T}}$, *e.g.*, 1-step gradient (*i.e.*, $\boldsymbol{M} = \boldsymbol{I}$) [31–33]. Note that the pre-trained diffusion model is frozen. The gradients can be computed by using standard autograd packages in PyTorch. Then, $\boldsymbol{x}_T$ can be updated along the gradient, as shown in Algorithm 2.

Different from existing diffusion model-based IR methods which have a large computational graph to store the gradients in the whole process, our method is more efficient due to the DEQ inversion. In addition, with the help of the inversion method, our zero-shot IR methods can be extended to supervised learning by replacing the loss (12) with $\mathcal{L} = \|\boldsymbol{x}_0^* - \boldsymbol{x}_0\|_F^2$ which we leave it in the future work.

Datasets	Methods	2×SR			4×SR			Deblur (Gaussian)			Deblur (anisotropic)			NFEs
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	/Iters
ImageNet	Baseline	29.63	0.875	0.165	25.15	0.699	0.351	18.22	0.529	0.433	20.86	0.544	0.480	-
	DGP [55]	22.32	0.583	0.426	18.35	0.398	0.529	21.81	0.522	0.472	20.77	0.459	0.504	1500
	DPS [20]	22.40	0.597	0.405	20.34	0.488	0.464	22.04	0.569	0.394	21.82	0.561	0.381	1000
	ILVR [19]	23.36	0.613	0.334	22.76	0.583	0.383	-	-	-	-	-	-	100
	DiffPIR [87]	27.16	0.790	0.214	24.31	0.649	0.350	25.32	0.673	0.296	23.37	0.535	0.439	100
	DDRM [41]	31.43	0.906	0.117	26.21	0.745	0.288	40.70	0.978	0.040	37.69	0.964	0.057	20
	DDNM [68]	31.81	0.908	0.097	26.49	0.753	0.266	**43.83**	**0.989**	**0.018**	38.40	0.970	0.038	100
	DeqIR (Ours)	**32.35**	**0.913**	**0.082**	**27.47**	**0.781**	**0.230**	43.42	0.987	0.021	**39.47**	**0.973**	**0.036**	15
CelebA-HQ	Baseline	35.87	0.953	0.099	30.12	0.857	0.240	18.94	0.704	0.337	23.16	0.727	0.354	-
	DGP [55]	28.61	0.809	0.279	25.25	0.690	0.405	27.02	0.738	0.372	25.73	0.663	0.426	1500
	DPS [20]	28.71	0.818	0.219	25.01	0.710	0.282	27.56	0.775	0.229	26.91	0.754	0.234	1000
	ILVR [19]	27.31	0.783	0.234	27.09	0.775	0.245	-	-	-	-	-	-	100
	DiffPIR [87]	32.51	0.882	0.156	28.60	0.795	0.228	30.63	0.835	0.197	29.32	0.802	0.232	100
	DDRM [41]	**36.76**	0.953	0.074	31.91	0.880	0.149	43.06	0.983	0.036	41.27	0.976	0.053	20
	DDNM [68]	36.37	0.950	0.065	31.86	0.876	**0.136**	46.99	0.991	0.021	43.43	0.983	0.037	100
	DeqIR (Ours)	36.63	**0.954**	**0.062**	**32.22**	**0.889**	0.155	**47.18**	**0.992**	**0.019**	**43.57**	**0.984**	**0.036**	15

Table 1. Quantitative results of zero-shot IR methods (including **super-resolution** and **deblurring**) on ImageNet and CelebA-HQ. Best results are highlighted as **first**, second and third.

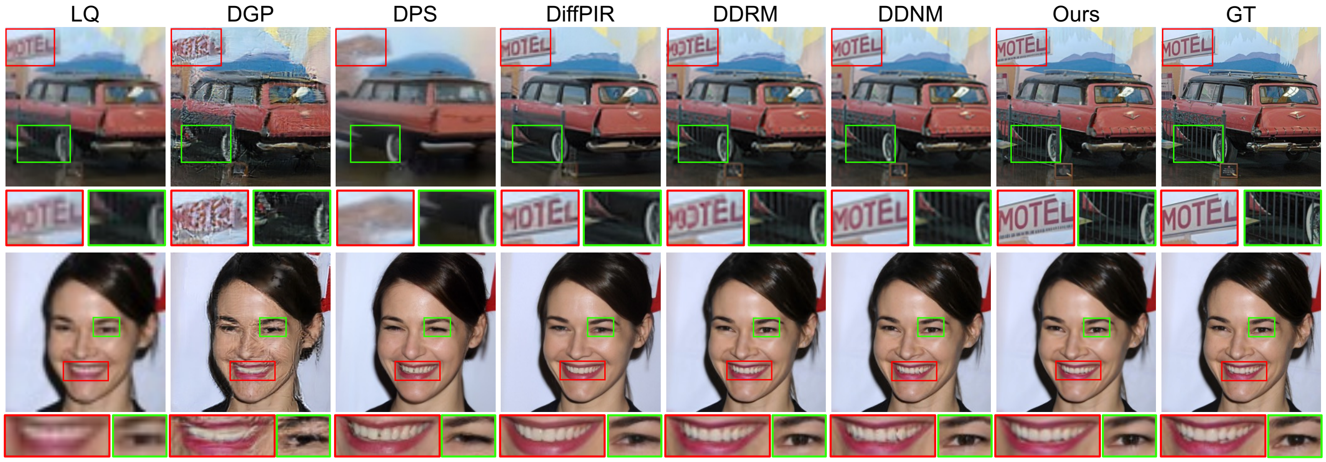


Figure 3. Qualitative results of zero-shot 4× super-resolution methods on ImageNet and CalabA-HQ.

5. Experiments

Experiment settings. We conduct typical IR tasks, including SR, deblurring, colorization, and inpainting. Specifically, we consider 2× and 4× bicubic downsampling for SR, Gaussian and anisotropic for deblurring, use an average grayscale operator in colorization, and use text and stripe masks in inpainting. For convenience, we choose ImageNet [22] and CelebA-HQ [39] with 100 classes [87] and the image size of 256×256 for validation, which have the same trend on 1k classes. For fair comparisons, we use the same pre-trained diffusion models [23] and [53] for ImageNet and CelebA-HQ, respectively. More details are put in Supplementary.

Evaluation metrics. We use PSNR, SSIM and LPIPS as the evaluation metrics for most IR tasks. For the task of colorization, we use the Consistency metric [68] and FID because PSNR and SSIM cannot reflect the performance [68]. In general, higher PSNR and SSIM, and lower LPIPS and FID mean better performance. In addition, we report the number of NFEs (timesteps) or iterations for each method.

5.1. Evaluation on Image Super-Resolution

We compare our method with a GAN-based IR method (*e.g.*, DGP) and SOTA zero-shot diffusion model-based IR methods (*e.g.*, DPS [20], DiffPIR [87], DDRM [41] and DDNM [68]) on ImageNet and CelebA-HQ datasets. In addition, we use the bicubic upscaling as a baseline for SR.

In Table 1, our method outperforms most methods under different metrics on both ImageNet and CelebA-HQ. In particular, compared with the competitive IR method DDNM, our method on ImageNet surpasses it by an LPIPS margin of up to 0.036, and by a PSNR margin of up to 0.98dB. Moreover, our method only needs 15 iteration steps, compared with DDNM (100 steps). We provide more details and quantitative results of other scales in Supplementary.

For the qualitative results, our method achieves the best visual quality containing more realistic textures, as shown in Figure 3. These visual comparisons align with the quantitative results, demonstrating the effectiveness of our method. More visual results are put in Supplementary Materials.

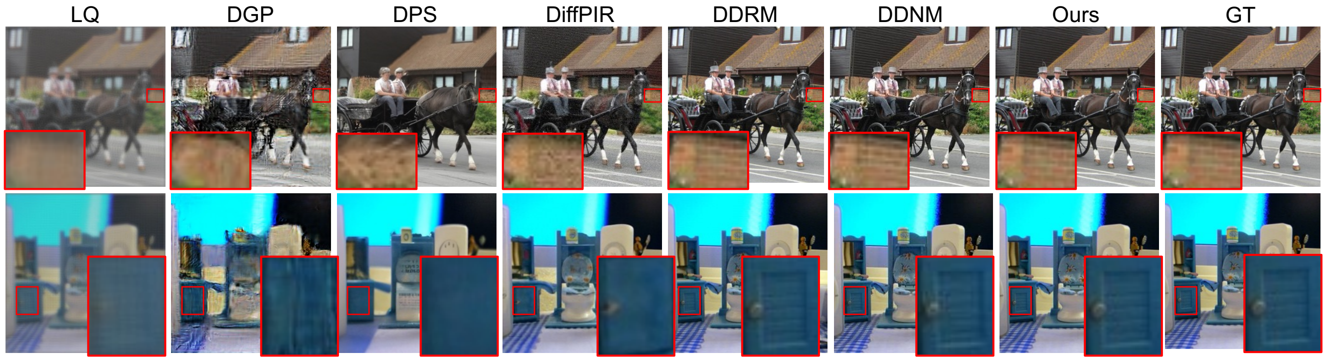


Figure 4. Qualitative results of zero-shot image deblurring (Gaussian) methods.

Methods	Text mask			Stripe mask		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Baseline	14.55	0.642	0.515	9.02	0.131	0.730
Palette [59]	38.09	0.978	0.027	25.91	0.733	0.343
DDRM [41]	37.25	0.969	0.223	34.34	0.933	0.223
RePaint [53]	38.54	0.974	0.039	36.25	0.951	0.086
DDNM [68]	39.45	0.980	**0.023**	36.75	**0.957**	**0.076**
DeqIR (Ours)	**39.72**	**0.981**	0.026	**36.99**	0.948	0.091

Table 2. Comparisons of zero-shot **inpainting** methods on CelebA.

Methods	ImageNet			CelebA-HQ		
	Cons↓	LPIPS↓	FID↓	Cons↓	LPIPS↓	FID↓
Baseline	0	0.196	90.93	0	0.210	70.69
DGP [55]	-	0.256	99.86	-	0.218	73.24
DDRM [41]	265.08	0.223	79.42	472.25	0.245	57.29
DDNM [68]	45.07	0.186	77.21	51.43	0.139	45.73
DeqIR (Ours)	**43.15**	**0.171**	**70.94**	**50.16**	**0.092**	**43.98**

Table 3. Quantitative results of zero-shot **colorization** methods.

5.2. Evaluation on Image Deblurring

We compare the same zero-shot IR methods used in the SR task. In addition, we use $\boldsymbol{A}^{\dagger}\boldsymbol{y}$ as a baseline. In this experiment, we mainly consider Gaussian and anisotropic kernels to evaluate the performance of all models.

In Table 1, the quantitative results show that our method achieves the best performance on all datasets, except for Gaussian deblurring on ImageNet. Compared with DDNM [68], the PSNR improvement of our method can be up to 1.07dB for anisotropic deblurring. In Figure 4, our generated images have the best visual quality with more realistic details which are close to GT images. We provide more quantitative and qualitative results (including more kernels) in Supplementary Materials.

5.3. Evaluation on Image Inpainting

For the image inpainting task, we compare our method with SOTA inpainting methods, including Palette [59], RePaint [53], DDRM [41] and DDNM [68]. We also use $\boldsymbol{A}^{\dagger}\boldsymbol{y}$ as a baseline. In addition, we consider the text mask and stripe mask as examples and show the results on CelebA-HQ in Table 2. The results of more masks and results on ImageNet are put in Supplementary Materials.

In Table 2, our method outperforms Palette [59] and DDRM [41] significantly, and has comparable performance with RePaint [53] and DDNM [68]. In Figure 6, taking the "mouth" in the generated face images as an example, our method generates clear structures and details that are not only more realistic but also more reasonable compared to other inpainting methods. In contrast, other methods may introduce blur artifacts.

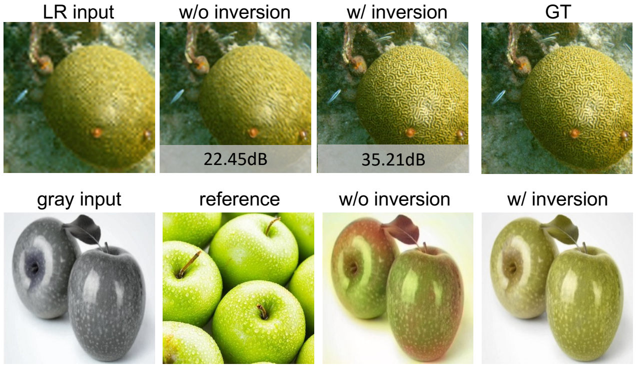


Figure 5. Interesting applications of DEQ inversion.

5.4. Evaluation on Image Colorization

We compare our method with SOTA methods (*i.e.*, DGP [55], DDRM [41] and DDNM [68]). We also use $\boldsymbol{A}^{\dagger}\boldsymbol{y}$ as a baseline. In addition to LPIPS, we additionally use the Consistency metric and FID to evaluate the image quality.

In Table 3, our method achieves the best performance on both ImageNet and CelebA-HQ under different metrics. As shown in Figure 7, our method restores images with reasonable color. In contrast, other methods may restore part of the color (as observed in the "tree") or unreasonable color (*e.g.*, evident in the "building" in DGP [55]).

5.5. Evaluation on DEQ Inversion

We extend our method using DEQ inversion to interesting applications, *e.g.*, SR with optimized initialization (top) and reference-based colorization (bottom), as shown in Figure 5. We found that optimizing the initialization is able to improve PSNR and control the generation in the desired direction. More details and results are put in Supplementary.

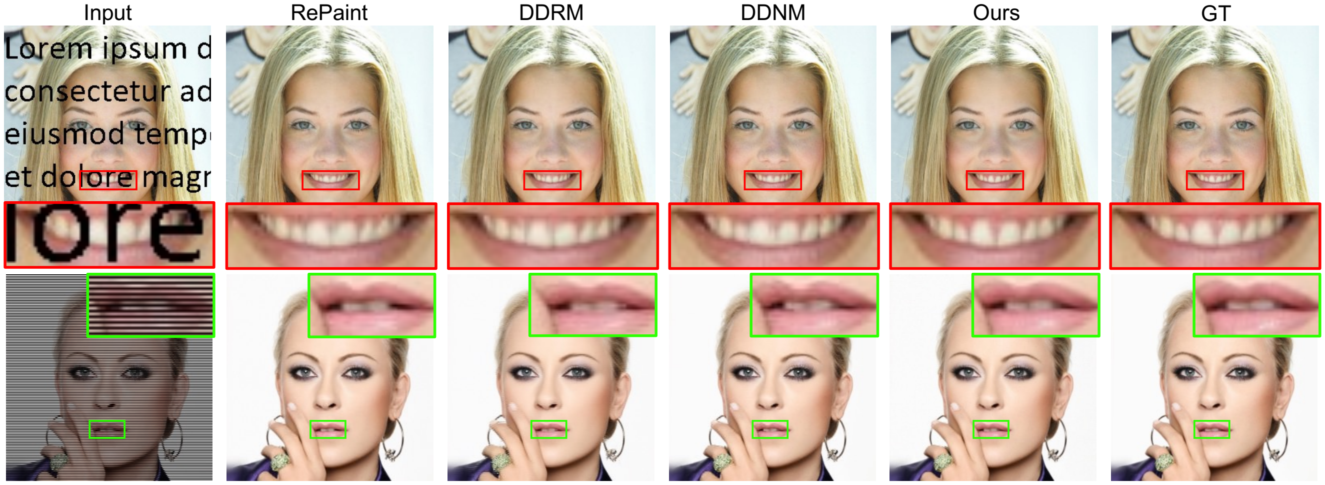


Figure 6. Qualitative results of image inpainting methods on CelebA-HQ.

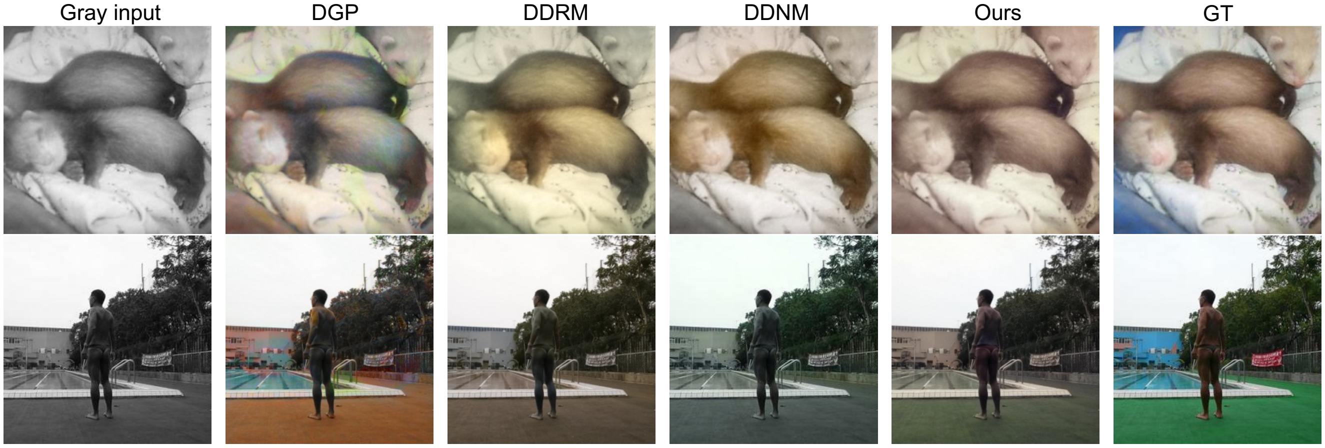


Figure 7. Qualitative results of image colorization methods on ImageNet.

5.6. Ablation Study

Effect of timesteps. We study the impact of timesteps in our diffusion models. Specifically, we change the number of timesteps from 2 to 35. In Figure 8 (left), the image quality improves with additional timesteps until it stabilizes. However, more timesteps lead to a larger memory and slower convergence. To trade off between performance and efficiency, we set the timesteps to 20 in this experiment.

Effect of iterations. We investigate the impact of varying the number of iterations in the Anderson acceleration in Figure 8 (middle). Increasing the number of iterations results in improved performance. As we can see, 15 iterations are sufficient to converge to satisfactory results.

Effect of hyper-parameter η. We further investigate the influence of the hyper-parameter η in our proposed analytic formulation, *i.e.*, Eqn. (9). In Figure 8 (right), different values of the hyper-parameter have different effects on the performance. Larger values introduce more noise in the generated image, while smaller values may limit the restoration performance. Therefore, we set the hyper-parameter η to 0.15 in this task.

5.7. Diversity of Generation

To investigate the ability of our method, we show diverse results for different tasks in Figure 9. With different seeds, our method is able to generate diverse images with realistic details on inpainting and colorization. For 32× SR, the input face image is severely degraded, and the generated faces are realistic but they are difficult to retain the identity.

5.8. Real-World Applications

Our method can be applied in real-world settings which may have unknown, non-linear and complex degradations.

Old photo restoration. The degradations in old photo restoration suffer from non-linear and unknown artifacts. Such artifacts are often covered by a hand-drawn mask (denoted by $\boldsymbol{A}_{\text{mask}}$). The degradation can be a composite of $\boldsymbol{A}_{\text{mask}}$ and a colorization degradation (denoted by $\boldsymbol{A}_{\text{color}}$), and its pseudo-inverse can also be constructed by hand. In Figure 10 (top), our method achieves a remarkable enhancement with facial details, effectively reducing the visible artifacts while preserving finer details. The inpainting and colorization results serve as a compelling illustration of the effectiveness of our old photo restoration technique.

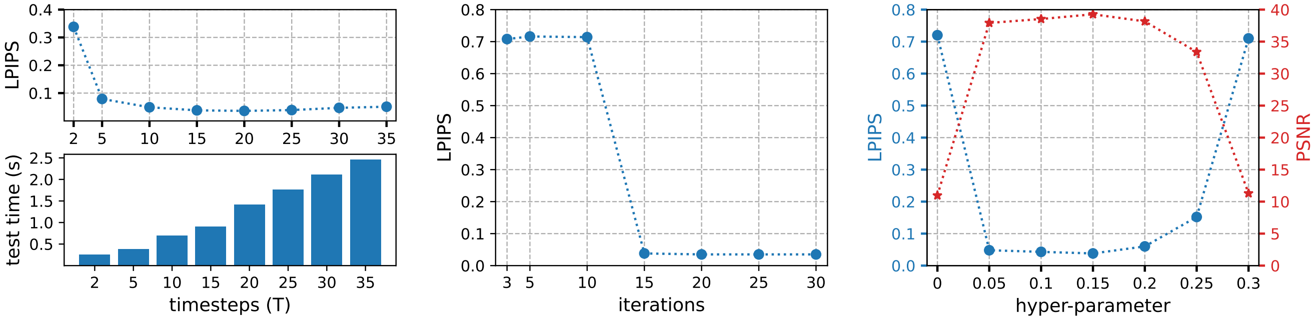


Figure 8. Ablation study of timesteps (left), iteration (middle) and hyper-parameters (right) for anisotropic deblurring on ImageNet.

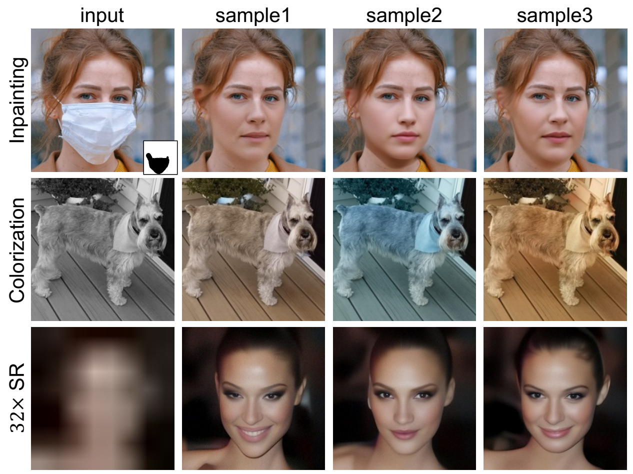


Figure 9. Diversity of generation of our method.

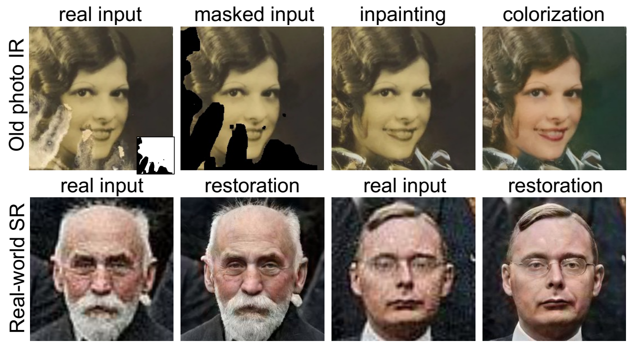


Figure 10. Real-world applications of our method.

Methods	DPS [20]	DDRM [41]	DDNM [68]	Ours-10	Ours-15	Ours-20
Time (s)	468.85	**5.26**	12.67	12.11	16.53	21.19
PSNR↑	21.82	37.69	38.40	38.58	39.21	**39.47**

Table 4. Running time of different methods. ∗-T: T timesteps.

Methods	ImageNet			CelebaA-HQ		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SRGAN [45]	24.83	0.696	0.245	31.16	0.868	0.164
BSRGAN [83]	23.65	0.651	0.331	27.80	0.808	0.216
LDM [58]	22.34	0.606	0.318	27.18	0.783	0.208
DiffIR [72]	**29.25**	**0.814**	**0.235**	**34.96**	**0.924**	**0.121**
DeqIR (Ours)	27.44	0.782	**0.235**	32.19	0.887	0.154

Table 5. Comparisons of supervised learning methods and our zero-shot method on ImageNet for 4× SR.

Real-world SR. Real-world degradations may have non-Gaussian noise, unknown compression noise and downscaling. We use a restoration model [80] to provide the prior information to the input noise. As shown in Figure 10 (bottom), our method achieves good robustness to the real noise. Notably, our method successfully preserves the facial identity and produces realistic results with rich details.

Arbitrary size. Our method can also be used in images with arbitrary sizes. Similarly to [48, 68], we crop a large-size image as multiple overlapped patches and then test each patch. Last we concatenate the generation as the final results. We put the results in Supplementary due to the limited space.

5.9. Further Experiments

Running time. We compare the running time of different methods for anisotropic deblurring on ImageNet. For fair comparisons, we evaluate all methods on 256×256 input images on NVIDIA TITAN RTX using their publicly available code. In Table 4, our method with 10 steps has a comparable running time to DDNM [68]. DDRM [41] with 20 steps is faster than our method, but it is worse than our method.

Comparisons with supervised learning. We compare our zero-shot method with supervised learning methods in Table 5. Our method outperforms GAN-based methods and LDM [58], but it is worse than DiffIR [72]. However, these methods have limited generalization on other tasks.

6. Conclusion

In this paper, we have proposed a novel zero-shot diffusion model-based IR method, called DeqIR. Specifically, we model diffusion model-based IR generation as a deep equilibrium (DEQ) fixed point system. Our IR method can conduct parallel sampling, instead of long sequential sampling in traditional diffusion models. Based on the DEQ inversion, we are able to explore the relationship between the restoration and initialization. With the initialization optimization, the restoration performance can be improved and the generation direction can be guided with additional information. Extensive experiments demonstrate that our proposed DeqIR achieves better performance on different IR tasks. Moreover, our DeqIR can be generalized to real-world applications.

Acknowledgments. This work was partly supported by The Alexander von Humboldt Foundation.

References

[1] Brandon Amos and J Zico Kolter. Optnet: Differentiable optimization as a layer in neural networks. In *ICML*, 2017. 2

[2] Donald G Anderson. Iterative procedures for nonlinear integral equations. *Journal of the ACM*, 1965. 3, 4

[3] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *NeurIPS*, 2019. 2, 3

[4] Shaojie Bai, Vladlen Koltun, and J Zico Kolter. Multiscale deep equilibrium models. In *NeurIPS*, 2020. 2

[5] Shaojie Bai, Zhengyang Geng, Yash Savani, and J Zico Kolter. Deep equilibrium optical flow estimation. In *CVPR*, 2022. 2

[6] Charles G Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of computation*, 1965. 3

[7] Jiezhang Cao, Yong Guo, Qingyao Wu, Chunhua Shen, Junzhou Huang, and Mingkui Tan. Adversarial learning with local coordinate coding. In *ICML*, 2018. 2

[8] Jiezhang Cao, Langyuan Mo, Yifan Zhang, Kui Jia, Chunhua Shen, and Mingkui Tan. Multi-marginal wasserstein gan. In *NeurIPS*, 2019.

[9] Jiezhang Cao, Yong Guo, Qingyao Wu, Chunhua Shen, Junzhou Huang, and Mingkui Tan. Improving generative adversarial networks with local coordinate coding. *TPAMI*, 2020. 2

[10] Jiezhang Cao, Yawei Li, Kai Zhang, and Luc Van Gool. Video super-resolution transformer. *arXiv preprint arXiv:2106.06847*, 2021. 2

[11] Jiezhang Cao, Jingyun Liang, Kai Zhang, Yawei Li, Yulun Zhang, Wenguan Wang, and Luc Van Gool. Reference-based image super-resolution with deformable attention transformer. In *ECCV*, 2022.

[12] Jiezhang Cao, Jingyun Liang, Kai Zhang, Wenguan Wang, Qin Wang, Yulun Zhang, Hao Tang, and Luc Van Gool. Towards interpretable video super-resolution via alternating optimization. In *ECCV*, 2022.

[13] Jiezhang Cao, Qin Wang, Yongqin Xian, Yawei Li, Bingbing Ni, Zhiming Pi, Kai Zhang, Yulun Zhang, Radu Timofte, and Luc Van Gool. Ciaosr: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution. In *CVPR*, 2023. 2

[14] Lukas Cavigelli, Pascal Hager, and Luca Benini. Cas-cnn: A deep convolutional neural network for image compression artifact suppression. In *IJCNN*, 2017. 2

[15] 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 *CVPR*, 2021. 2

[16] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 2

[17] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *NeurIPS*, 2018. 2, 3

[18] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *CVPR*, 2023. 3, 4

[19] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. In *ICCV*, 2021. 2, 5

[20] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023. 1, 2, 3, 4, 5, 8

[21] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, 2019. 2

[22] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 5

[23] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, 2021. 1, 2, 5

[24] Josip Djolonga and Andreas Krause. Differentiable learning of submodular models. In *NeurIPS*, 2017. 2

[25] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang. Compression artifacts reduction by a deep convolutional network. In *ICCV*, 2015. 2

[26] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *TPAMI*, 2015. 2

[27] Priya L Donti, David Rolnick, and J Zico Kolter. Dc3: A learning method for optimization with hard constraints. In *ICLR*, 2021. 2

[28] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. In *NeurIPS*, 2019. 2

[29] Xueyang Fu, Zheng-Jun Zha, Feng Wu, Xinghao Ding, and John Paisley. Jpeg artifacts reduction via deep convolutional sparse coding. In *ICCV*, 2019. 2

[30] Xueyang Fu, Menglu Wang, Xiangyong Cao, Xinghao Ding, and Zheng-Jun Zha. A model-driven deep unfolding method for jpeg artifacts removal. *TNNLS*, 2021. 2

[31] Samy Wu Fung, Howard Heaton, Qiuwei Li, Daniel McKenzie, Stanley Osher, and Wotao Yin. Fixed point networks: Implicit depth models with jacobian-free backprop. *arXiv preprint arXiv:2103.12803*, 2021. 4

[32] Zhengyang Geng, Meng-Hao Guo, Hongxu Chen, Xia Li, Ke Wei, and Zhouchen Lin. Is attention better than matrix decomposition? In *ICLR*, 2020. 2

[33] Zhengyang Geng, Xin-Yu Zhang, Shaojie Bai, Yisen Wang, and Zhouchen Lin. On training implicit models. In *NeurIPS*, 2021. 4

[34] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *ICLR*, 2022. 2

[35] Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph neural networks. In *NeurIPS*, 2020. 2

[36] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. In *NeurIPS*, 2017. 2

[37] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020. 1, 2, 3

[38] Xixi Jia, Sanyang Liu, Xiangchu Feng, and Lei Zhang. Focnet: A fractional optimal control network for image denoising. In *CVPR*, 2019. 2

[39] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *ICLR*, 2018. 5

[40] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *NeurIPS*, 2021. 2

[41] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS*, 2022. 1, 2, 3, 5, 6, 8

[42] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016. 2

[43] Yoonsik Kim, Jae Woong Soh, Jaewoo Park, Byeongyong Ahn, Hyun-Seung Lee, Young-Su Moon, and Nam Ik Cho. A pseudo-blind convolutional neural network for the reduction of compression artifacts. *TCSVT*, 2019. 2

[44] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *NeurIPS*, 2021. 1

[45] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 8

[46] Mingjie Li, Yisen Wang, and Zhouchen Lin. Cerdeq: Certifiable deep equilibrium model. In *ICML*, 2022. 2

[47] Wenbo Li, Zhe Lin, Kun Zhou, Lu Qi, Yi Wang, and Jiaya Jia. Mat: Mask-aware transformer for large hole image inpainting. In *CVPR*, 2022. 2

[48] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCVW*, 2021. 3, 4, 8

[49] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *TIP*, 2024. 2

[50] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Ben Fei, Bo Dai, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023. 2

[51] Cheng Lu, Jianfei Chen, Chongxuan Li, Qiuhao Wang, and Jun Zhu. Implicit normalizing flows. In *ICLR*, 2021. 2

[52] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *NeurIPS*, 2022. 1

[53] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022. 2, 5, 6

[54] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *CVPR*, 2020. 2

[55] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *TPAMI*, 2021. 5, 6

[56] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *CVPR*, 2016. 2

[57] Ashwini Pokle, Zhengyang Geng, and J Zico Kolter. Deep equilibrium approaches to diffusion models. In *NeurIPS*, 2022. 1, 2, 3, 4

[58] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 2, 8

[59] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH*, 2022. 2, 6

[60] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *TPAMI*, 2022. 2

[61] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 1, 2, 3

[62] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. 2

[63] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. In *arXiv preprint arXiv:2305.07015*, 2023. 2

[64] Shuai Wang, Yao Teng, and Limin Wang. Deep equilibrium object detection. In *ICCV*, 2023. 2

[65] Tiancai Wang, Xiangyu Zhang, and Jian Sun. Implicit feature pyramid network for object detection. *arXiv preprint arXiv:2012.13563*, 2020. 2

[66] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *ECCVW*, 2018. 2

[67] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCVW*, 2021. 2

[68] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *ICLR*, 2023. 1, 2, 3, 4, 5, 6, 8

[69] Colin Wei and J Zico Kolter. Certified robustness for deep equilibrium models via interval bound propagation. In *ICLR*, 2021. 2

[70] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *CVPR*, 2022. 2

[71] Bin Xia, Yucheng Hang, Yapeng Tian, Wenming Yang, Qingmin Liao, and Jie Zhou. Efficient non-local contrastive attention for image super-resolution. In *AAAI*, 2022. 2

[72] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *ICCV*, 2023. 2, 8

[73] Bin Xia, Yulun Zhang, Yitong Wang, Yapeng Tian, Wenming Yang, Radu Timofte, and Luc Van Gool. Knowledge distillation based degradation estimation for blind super-resolution. In *ICLR*, 2023. 2

[74] Chaohao Xie, Shaohui Liu, Chao Li, Ming-Ming Cheng, Wangmeng Zuo, Xiao Liu, Shilei Wen, and Errui Ding. Image inpainting with learnable bidirectional attention maps. In *ICCV*, 2019. 2

[75] Zonghan Yang, Tianyu Pang, and Yang Liu. A closer look at the adversarial robustness of deep equilibrium models. In *NeurIPS*, 2022. 2

[76] Zili Yi, Qiang Tang, Shekoofeh Azizi, Daesik Jang, and Zhan Xu. Contextual residual aggregation for ultra high-resolution image inpainting. In *CVPR*, 2020. 2

[77] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *CVPR*, 2018.

[78] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *ICCV*, 2019. 2

[79] Zongsheng Yue and Chen Change Loy. Difface: Blind face restoration with diffused error contraction. *arXiv preprint arXiv:2212.06512*, 2022. 2, 3

[80] 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 *CVPR*, 2022. 2, 8

[81] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Aggregated contextual transformations for high-resolution image inpainting. *TVCG*, 2022. 2

[82] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *TPAMI*, 2021. 2

[83] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *ICCV*, 2021. 2, 8

[84] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *ECCV*, 2016. 3

[85] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 2

[86] Shangchen Zhou, Kelvin C.K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *NeurIPS*, 2022. 3

[87] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *CVPRW*, 2023. 1, 2, 5