Improving Diffusion-Based Image Restoration with Error Contraction and Error Correction

Qiqi Bao[1], Zheng Hui[2], Rui Zhu[3], Peiran Ren[2], Xuansong Xie[2], Wenming Yang[1]*

[1]Tsinghua University
[2]Institute for Intelligent computing, Alibaba Group
[3]City, University of London
bqq19@mails.tsinghua.edu.cn, zheng_hui@aliyun.com, rui.zhu@city.ac.uk,
peiran_r@sohu.com, xingtong.xxs@taobao.com, yang.wenming@sz.tsinghua.edu.cn

Abstract

Generative diffusion prior captured from the off-the-shelf denoising diffusion generative model has recently attained significant interest. However, several attempts have been made to adopt diffusion models to noisy inverse problems either fail to achieve satisfactory results or require a few thousand iterations to achieve high-quality reconstructions. In this work, we propose a diffusion-based image restoration with error contraction and error correction (DiffECC) method. Two strategies are introduced to contract the restoration error in the posterior sampling process. First, we combine existing CNN-based approaches with diffusion models to ensure data consistency from the beginning. Second, to amplify the error contraction effects of the noise, a restart sampling algorithm is designed. In the error correction strategy, the estimation-correction idea is proposed on both the data term and the prior term. Solving them iteratively within the diffusion sampling framework leads to superior image generation results. Experimental results for image restoration tasks such as super-resolution (SR), Gaussian deblurring, and motion deblurring demonstrate that our approach can reconstruct high-quality images compared with state-of-the-art sampling-based diffusion models.

Introduction

Low-level vision tasks in image restoration, such as image denoising, image super-resolution (SR), and image deblurring can be cast as inverse problems $y = A(x) + n$, where x stands for the original image, $A(\cdot)$ is the forward measurement operator and n represents the noise. The inverse problems aim to infer the underlying signal from measurements and yield a high-quality image.

Recently, diffusion models (Song et al. 2020; Song and Ermon 2020; Dhariwal and Nichol 2021; Ho, Jain, and Abbeel 2020; Nichol and Dhariwal 2021; Karras et al. 2022; Chung, Sim, and Ye 2022a; Mokady et al. 2022; Liu et al. 2023; Permenter and Yuan 2023) have shown state-of-the-art performance in image generation compared to Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). Diffusion models define a forward process by gradually adding *Gaussian* noise to the input data that maps data to noise. During the reverse sampling process, diffusion models start with a pure *Gaussian* noise image and progressively sample a less noisy image until reaching a clean one. Such diffusion models use the parameterized prior of the high-dimensional data distributions. In addition to the unconditional generative power, diffusion models have achieved remarkable success in solving inverse problems (Wang, Yu, and Zhang 2022; Abu-Hussein, Tirer, and Giryes 2022; Meng and Kabashima 2022; Kawar et al. 2022b; Song et al. 2023; Fabian, Tinaz, and Soltanolkotabi 2023; Murata et al. 2023; Song et al. 2021; Chung, Lee, and Ye 2023). Chung *et al*. (Chung et al. 2023) proposed a Diffusion Posterior Sampling (DPS) method for solving general noisy non-linear inverse problems. However, adapting denoising diffusion models from the pure *Gaussian* noise for image restoration behaves slowly in sampling. DPS requires a few hundreds iterations to achieve high-quality reconstructions. Denoising Diffusion Null-Space Model (DDNM) (Wang, Yu, and Zhang 2022) decomposes samples into the range-space and the null-space of the measurement. By refining the null-space during the reverse diffusion process, DDNM assures data consistency and incorporates priors from diffusion models. Though DDNM ensures data consistency from the beginning that helps reduce iterations, the ability to generate images with higher quality is constrained.

In this work, we depend on generative priors from pre-trained unconditional diffusion models. Data distribution is modeled regardless of the forward measurement operator

Figure 1: An example visualization of the intermediate results for the prediction of $\hat{x}_{0|t}$ in DPS (Chung et al. 2023).

*Corresponding author.

$A(\cdot)$ and can be generalized to different degradations. From Fig. 1, we can see that predicted images contain little information in the first line during the reverse diffusion process. The inaccuracy of the initial prediction in the early stage causes accumulated errors for image inverse tasks. In (Xu et al. 2023), Xu *et al*. find that SDE-based samplers consist of the discretization error along the trajectory. Both the initial prediction error and the discretization error lead to increased sampling steps to deliver higher sample quality. Therefore, we propose an error contraction strategy by introducing the accurate initial prediction and the restart sampling operation. First, to rectify estimation $\hat{x}_{0|t}$ to guarantee the data consistency in the initial phase, outputs from the existing neural network for instance the RealESRNet (Wang et al. 2021) are utilized. The reverse diffusion path is reduced to T' steps, where $T' < T$. The priors encapsulated from the pretrained neural network help generate a better initial point and contract errors in the initial phase. Second, rather than using SDE-based samplers in (Abu-Hussein, Tirer, and Giryes 2022; Chung et al. 2023; Wang, Yu, and Zhang 2022; Song et al. 2023; Chung, Lee, and Ye 2023), ODE-based samplers with restart sampling algorithm is involved. The deterministic backward processes reduce the discretization error, while the forward-backward restart sampling operation strengthens the contraction effect. Apart from the initialization error and the discretization error, the approximation error of the learned neural network and the natural image distribution prediction for the conditional image generation also affect the realness and consistency of the reconstructed images. We design an error correction strategy to solve the optimization problem. The error correction strategy is composed of two iterative steps: an efficient Adam optimization of the neural network's prediction and one step of gradient descent extended from the DPS framework.

Contributions. The main contributions are summarized as follows:

- We propose an error contraction strategy by integrating existing neural network priors and harnessing the restart sampling technique to achieve accurate reconstruction.
- We design an error correction strategy by imposing the prior term to correct the neural estimation and reconstructing y given the measurement model iteratively within the diffusion sampling framework.
- Compared with state-of-the-art methods, our model achieves superior performance on different image restoration tasks such as image SR, *Gaussian* deblurring, and *motion* deblurring.

Background

Score-based Generative Formulation

Let a random variable x_0 with the data distribution $q_0(x_0) = p_{\text{data}}(x_0)$. Diffusion is the process of progressively adding *Gaussian* noise to the observation x_0 to transform $q_0(x_0)$ at time 0 to a normal distribution $q_T(x_T)$ at time T. Song *et al*. (Song et al. 2020) defined the forward SDE as

$$dx = f(t)\,xdt + g(t)\,dw_t, \tag{1}$$

where w_t is the standard Wiener process, and $f(t)$, $g(t)$ are the drift and diffusion coefficients, respectively. The forward process described in Eq. (1) has the corresponding reverse process from T to 0:

$$dx_t = \left[f(t)\,x_t - g^2(t)\,\nabla_x \log q_t(x_t)\right]dt + g(t)\,d\bar{w}_t. \tag{2}$$

where $\nabla_x \log q_t(x_t)$ is the *score function* of $q_t(x_t)$.

For the specific choice of $f(t) = -\frac{1}{2}\beta(t)$ and $g(t) = \sqrt{\beta(t)}$, VP SDE (Song et al. 2020) has the following form

$$dx = -\frac{\beta(t)}{2}xdt + \sqrt{\beta(t)}dw_t, \tag{3}$$

where $\beta(t) = \beta_{\min} + t(\beta_{\max} - \beta_{\min})$ is the noise schedule of the forward process. The corresponding reverse SDE of Eq. (3) is

$$dx = \left[-\frac{\beta(t)}{2}x - \beta(t)\,\nabla_{x_t}\log q_t(x_t)\right]dt + \sqrt{\beta(t)}d\bar{w}_t. \tag{4}$$

Song *et al*. (Song et al. 2020) proved that the ordinary differential equation (ODE) of Eq. (2), dubbed the probability flow ODE, is:

$$\frac{dx_t}{dt} = f(t)\,x_t - \frac{1}{2}g^2(t)\,\nabla_x \log q_t(x_t). \tag{5}$$

To estimate $\nabla_{x_t}\log q_t(x_t)$, Song *et al*. (Song et al. 2020) trained a time-dependent score-based model $s_\theta(x_t, t)$ via

$$\min_\theta \mathbb{E}_t\left\{\lambda(t)\,\mathbb{E}_{x_0,x_t}\left[\|s_\theta(x_t,t) - \nabla_{x_t}\log q_{0t}(x_t\,|x_0)\|_2^2\right]\right\}, \tag{6}$$

where $\lambda(t)$ is a positive weight coefficient, t is uniformly sampled from $[0, T]$, $x_t \sim q(x_t|x_0)$.

Forward and Reverse Diffusion Processes

We have the continuous version of the diffusion process for the denoising diffusion probabilistic model (DDPM) formulation (Ho, Jain, and Abbeel 2020) in Eq. (4). One forward step of (discrete) DDPM is

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon_{t-1}, \tag{7}$$

where $\epsilon_{t-1} \sim N(0, \mathbf{I})$. With the properties of *Gaussian*, we can sample x_t from x_0 as

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon, \tag{8}$$

where $\alpha_t = 1-\beta_t$ and $\overline{\alpha}_t = \prod_{i=1}^{t}\alpha_i$. One reverse sampling step is

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t,t)\right) + \sigma_t\epsilon_t, \tag{9}$$

where $s_\theta(x_t,t) = -\frac{\epsilon_\theta(x_t,t)}{\sqrt{1-\overline{\alpha}_t}}$.

Song *et al*. (Song, Meng, and Ermon 2020) proposed denoising diffusion implicit model formulation (DDIM) to enable a faster sampling process. One reverse sampling step in Eq. (9) is rewritten as

$$x_{t-1} = \sqrt{\overline{\alpha}_{t-1}}\hat{x}_{0|t}(x_t) + \sqrt{1-\overline{\alpha}_{t-1}-\sigma_{\eta_t}^2}\cdot\epsilon_\theta(x_t,t) + \sigma_{\eta_t}\epsilon_t, \tag{10}$$

where σ_{η_t} controls the stochasticity of the diffusion process. By setting $\sigma_{\eta_t} = 0$, the reverse process beyond the initial randomization becomes deterministic.

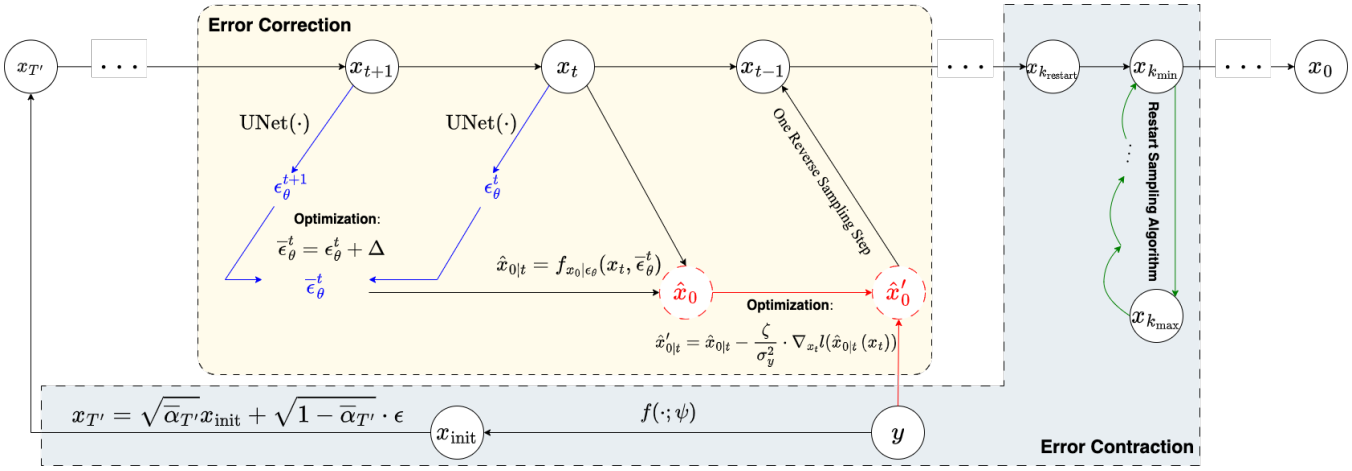

Figure 2: Illustration of our proposed DiffECC method. The gray boxes present the error contraction strategy. The yellow box presents the error correction strategy. The green arrows indicate the restart sampling operation. The blue arrows indicate the noise correction route. The red arrows indicate the corrected route of $\hat{x}_{0|t}$.

Diffusion Models for Inverse Problems

The general form of inverse problems can be formulated as

$$y = A(x_0) + n, x_0 \in \mathbb{R}^D, y \in \mathbb{R}^d, n \in \mathbb{R}^d, \tag{11}$$

where $A(\cdot) : \mathbb{R}^D \to \mathbb{R}^d$ is the known forward measurement operator and $n \sim N(\mathbf{0}, \sigma_y^2 \mathbf{I})$ is the white *Gaussian* noise. The likelihood function can be written as

$$p(y|x_0) = \frac{1}{\sqrt{(2\pi)^n \sigma_y^{2n}}} \exp\left[-\frac{\|y - A(x_0)\|_2^2}{2\sigma_y^2}\right], \tag{12}$$

with mean $A(x_0)$ and variance σ_y^2. The goal for the inverse problems is to recover $\hat{x}_0 \in \mathbb{R}^D$ from a degraded image y.

As in (Chung, Sim, and Ye 2022b; Chung et al. 2023; Wang, Yu, and Zhang 2022; Zhu et al. 2023), we can use diffusion models to solve inverse problems by replacing the score function in Eq. (4) with the conditional score function $\nabla_{x_t} \log p(x_t|y)$. By Bayes rule, we can derive the following equation:

$$\nabla_{x_t} \log p(x_t|y) = \underbrace{\nabla_{x_t} \log p(x_t)}_{unconditional\ score} + \underbrace{\nabla_{x_t} \log p(y|x_t)}_{adversarial\ gradient} \cancel{-\nabla_{x_t} \log p(y)}. \tag{13}$$

To adjust the control intensity, Classifier Guidance scales the adversarial gradient by a γ parameter:

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(x_t) + \gamma \nabla_{x_t} \log p(y|x_t), \tag{14}$$

where the first term can be approximated with the pre-trained score function $s_\theta(x_t, t)$, and the second term is the guidance term with the conditional score of $p_t(y|x_t)$.

Method

Error Contraction Strategy

Accurate initial prediction. To restore the data distribution of the high-quality image from the degraded counterpart y, the marginal distribution can be written as

$$p(x_0|y) = \int p_\theta(x_T) \prod_{t=1}^{T} p_\theta^{(t)}(x_{t-1}|x_t, y) dx_{1:T}. \tag{15}$$

In Fig. 1, the predicted $\hat{x}_{0|t}$ is seriously destroyed when t is close to $t = T$. Since x_{t-1} is yield by sampling from $p(x_{t-1}|x_t, \hat{x}_{0|t})$, the corrupted prediction for $\hat{x}_{0|t}$ makes the reverse sampling process converges slowly. The observation indicates that decreasing the estimation errors from the initialization would improve image reconstruction with data consistency. To satisfy the data consistency from the initial phase, we simply modify Eq. (15) to reconstruct the diffusion posterior distribution $p(x_0|y)$ by

$$p(x_0|y) = \int p_\theta(x_{T'}|y) \prod_{t=1}^{T'} p_\theta^{(t)}(x_{t-1}|x_t, y) dx_{1:T'}, \tag{16}$$

where $T' < T$ represents the starting timestep. Parameters in the main diffusion process are defined as the noise $Schedule_1$. The reverse transition distribution $p_\theta(x_{t-1}|x_t, y)$ is described in the next part. Now the goal turns to design the transition distribution of $p_\theta(x_{T'}|y)$.

Inspired by (Chung, Sim, and Ye 2022a; Wang et al. 2023; Yue and Loy 2022), rather than applying the initial randomization, we start with more accurate initial prediction. The transition distribution $p_\theta(x_{T'}|y)$ is formulated as a *Gaussian* distribution:

$$p_\theta(x_{T'}|y) = N(x_{T'}; \sqrt{\overline{\alpha}_{T'}} x_{\text{init}}, (1 - \overline{\alpha}_{T'})\mathbf{I}). \tag{17}$$

Via the reparameterization trick, we have the forward diffusion process represented as

$$\begin{aligned} x_{\text{init}} &= f(y; \psi) \\ x_{T'} &= \sqrt{\overline{\alpha}_{T'}} x_{\text{init}} + \sqrt{1 - \overline{\alpha}_{T'}} \cdot \epsilon, \epsilon \sim N(\mathbf{0}, \mathbf{I}), \end{aligned} \tag{18}$$

where $f(\cdot; \psi)$ is a pre-trained image restoration network (like Real-ESRNet (Wang et al. 2021), MPRNet (Zamir et al.

Algorithm 1: DiffECC

Require: $T, T', A(\cdot), y, f(\cdot;\psi), \mathrm{UNet}(\cdot), t_{\mathrm{cond}}, \zeta, K, k_{\mathrm{restart}}, k_{\mathrm{skip}}$

$x_{\mathrm{init}} = f(y;\psi)$ ▷ Obtain a good initial prediction via the pretrained CNN-based network
Respace the intervals of the diffusion process in T, and obtain the parameters in this schedule as noise $Schedule_1$.
$\epsilon \sim N(\mathbf{0}, \mathbf{I})$ ▷ Sample noise
$x_{T'} = \sqrt{\overline{\alpha}_{T'}} x_{\mathrm{init}} + \sqrt{1-\overline{\alpha}_{T'}} \cdot \epsilon$ ▷ Starting state at time T'
for $t = T'$ **to** t_{cond} **do**
  $\epsilon_\theta^t = \mathrm{UNet}(x_t, t)$ ▷ Noise estimation
  $\overline{\epsilon}_\theta^t = \epsilon_\theta^t + \arg\min_{\Delta} \left(d(\epsilon_\theta^{t+1}, (\epsilon_\theta^t + \Delta))\right)$ ▷ Error correction for the noise estimation ϵ_θ^t at time t
  $\hat{x}_{0|t} = \frac{1}{\sqrt{\overline{\alpha}_t}}(x_t - \sqrt{1-\overline{\alpha}_t} \cdot \overline{\epsilon}_\theta^t)$ ▷ Calculate $\hat{x}_{0|t}$ from the estimated noise
  $\hat{x}'_{0|t} = \hat{x}_{0|t} - \frac{\zeta}{\sigma_y^2} \cdot \nabla_{x_t} \left\| y - A\left(\hat{x}_{0|t}\left(x_t\right)\right)\right\|_2^2$ ▷ Error correction for the prediction of $\hat{x}_{0|t}$ at time t
  $\hat{\epsilon}_\theta^t = \frac{1}{\sqrt{1-\overline{\alpha}_t}}(x_t - \sqrt{\overline{\alpha}_t} \cdot \hat{x}'_{0|t})$ ▷ Calculate $\hat{\epsilon}_\theta^t$ from the predicted reconstruction result
  $x_{t-1} = \sqrt{\overline{\alpha}_{t-1}} \hat{x}'_{0|t} + \sqrt{1-\overline{\alpha}_{t-1}} \cdot \hat{\epsilon}_\theta^t$ ▷ One deterministic step of reverse diffusion sampling
  if $t = k_{\mathrm{restart}}$ **then**
    Respace the intervals of the diffusion process in K, and obtain the parameters in this schedule as noise $Schedule_2$.
    $\hat{x}_0^r = \hat{x}'_0$
    $k_{\max} = K - k_{\mathrm{skip}}$
    $x_{k_{\max}} = \sqrt{\overline{\alpha}_{k_{\max}}} \cdot \hat{x}_0^r + \sqrt{1-\overline{\alpha}_{k_{\max}}} \cdot \epsilon$ ▷ Starting state for restart operation
    for $k = k_{\mathrm{skip}-1}, ..., 0$ **do**
      $\epsilon_\theta^k = \mathrm{UNet}(x_k, k)$
      $\hat{x}_{0|k} = \frac{1}{\sqrt{\overline{\alpha}_k}}(x_k - \sqrt{1-\overline{\alpha}_k} \cdot \epsilon_\theta^k)$
      $x_{k-1} = \sqrt{\overline{\alpha}_{k-1}} \hat{x}_{0|k} + \sqrt{1-\overline{\alpha}_{k-1}} \cdot \epsilon_\theta^k$ ▷ One deterministic sampling step in the restart sampling process
    end for
    $x_{t-1} = \sqrt{\overline{\alpha}_{t-1}} \cdot \hat{x}_0 + \sqrt{1-\overline{\alpha}_{t-1}} \cdot \epsilon$
  end if
end for
for $t = t_{\mathrm{cond}} - 1$ **to** 0 **do**
  $\epsilon_\theta^t = \mathrm{UNet}(x_t, t)$ ▷ Noise estimation
  $\hat{x}_{0|t} = \frac{1}{\sqrt{\overline{\alpha}_t}}(x_t - \sqrt{1-\overline{\alpha}_t} \cdot \epsilon_\theta^t)$ ▷ Predicted $\hat{x}_{0|t}$
  $x_{t-1} = \sqrt{\overline{\alpha}_{t-1}} \hat{x}_{0|t} + \sqrt{1-\overline{\alpha}_{t-1}} \cdot \epsilon_\theta^t$ ▷ One deterministic step of reverse diffusion sampling
end for

2021), MIRNet (Zamir et al. 2020, 2022)) with parameter ψ. Instead of employing Real-ESRGAN (GAN-based), we resort to Real-ESRNet (CNN-based trained with MAE loss). The principal intention is to rely on diffusion reverse sampling to synthesize image detail information. In addition, the CNN-based solution is more common and simpler.

Restart sampling algorithm. In the DDIM fashion, we can get the final one-step sampling expression as

$$x_{t-1} = \sqrt{\overline{\alpha}_{t-1}} \hat{x}_{0|t}(y) + \sqrt{1-\overline{\alpha}_{t-1}-\sigma_{\eta_t}^2} \cdot \epsilon_\theta(x_t, t) + \sigma_{\eta_t} \epsilon_t. \tag{19}$$

In our case, since the noise term σ_{η_t} may not be strong enough and can cause the discretization error, we set $\sigma_{\eta_t} = 0$. Instead, we extend the idea in (Xu et al. 2023) and propose the restart sampling operation to amplify the error contraction effects of the noise in a deterministic backward processes to reduce the discretization error simultaneously.

In the restart sampling algorithm, the back-and-forth step is performed in a new time interval. We respace the intervals of the diffusion process in K. Parameters in the restart process are defined as the noise $Schedule_2$. The amount of added noise in the restart forward process is larger than the small single-step noise in Eq. (19), thus amplifying the error contraction effect. We set the predicted $\hat{x}'_0$ at time k_{restart} as $\hat{x}_0^r$, being the input to the restart sampling algorithm. In the restart forward process, a substantial amount of noise is added to transit the $\hat{x}_0^r$ from $k = 0$ to $k = k_{\mathrm{skip}}$,

$$\begin{aligned} \hat{x}_0^r &= \hat{x}'_0 \\ k_{\max} &= K - k_{\mathrm{skip}} \\ x_{k_{\max}} &= \sqrt{\overline{\alpha}_{k_{\max}}} \cdot \hat{x}_0^r + \sqrt{1-\overline{\alpha}_{k_{\max}}} \cdot \epsilon, \end{aligned} \tag{20}$$

where k_{skip} represents the number steps to skip during the restart diffusion process. A restart backward process runs the backward ODE.

Error Correction Strategy

Using the ODE-based sampler, we can derive a general update formula for the conditional diffusion as

$$x_{t-1} = \sqrt{\overline{\alpha}_{t-1}} \hat{x}'_{0|t}(y) + \sqrt{1-\overline{\alpha}_{t-1}} \cdot \hat{\epsilon}_\theta^t, \tag{21}$$

where the following forms are iteratively used for $\hat{x}'_{0|t}(y)$ and $\hat{\epsilon}^t_\theta$:

$$\hat{x}'_{0|t} = \begin{cases} \hat{x}_0\left(x_t\right) - \dfrac{\zeta}{\sigma_y^2} \cdot \nabla_{x_t} \left\| y - A\left(\hat{x}_0\left(x_t\right)\right)\right\|_2^2 \\ f_{x_0|\epsilon_\theta}(x_t, \bar{\epsilon}^t_\theta), \text{where } \bar{\epsilon}^t_\theta = \epsilon^t_\theta + \ \Delta, \end{cases} \tag{22}$$

$$\hat{\epsilon}_\theta = \begin{cases} \epsilon^t_\theta + \ \arg\min_{\Delta} \left(d(\epsilon^{t+1}_\theta, (\epsilon^t_\theta + \Delta))\right) \\ f_{\epsilon_\theta|x_0}(x_t, \hat{x}'_{0|t}), \text{where } \hat{x}'_{0|t} = \hat{x}_{0|t} - \dfrac{\zeta}{\sigma_y^2} \cdot \nabla_{x_t} \ell(\hat{x}_{0|t}), \end{cases} \tag{23}$$

where $f_{x_0|\epsilon_\theta}(\cdot)$ represents the function of predicting x_0 from ϵ_θ and $f_{\epsilon_\theta|x_0}(\cdot)$ represents the function of predicting ϵ_θ from x_0.

Extended the idea in DPS (Chung et al. 2023) and ΠGDM (Song et al. 2023), data consistency in our method is imposed as

$$\hat{x}'_0 \leftarrow \hat{x}_0 - \gamma \cdot \nabla_{x_t} \ell(\hat{x}_0\left(x_t\right)), \tag{24}$$

where $\nabla_{x_t} \ell(\hat{x}_0\left(x_t\right))$ represents the computation of the gradient. Specifically, we use the Jensen approximation from DPS (Chung et al. 2023) as

$$p(y|x_t) \simeq p(y|\hat{x}_0). \tag{25}$$

Using the likelihood function in Eq. (12), we get the correcting step under the *Gaussian* measurement model as

$$\nabla_{x_t} \log p(y|x_t) \simeq -\frac{1}{\sigma_n^2} \nabla_{x_t} \left\| y - A\left(\hat{x}_0\left(x_t\right)\right)\right\|_2^2. \tag{26}$$

The update of $\hat{x}'_0$ is calculated as

$$\hat{x}'_0 = \hat{x}_0 - \frac{\zeta}{\sigma_y^2} \cdot \nabla_{x_t} \left\| y - A\left(\hat{x}_0\left(x_t\right)\right)\right\|_2^2. \tag{27}$$

An efficient Adam optimizer (Kingma and Ba 2014) is implied to correct the neural estimation. As for a clean image, the backward diffusion is expected to reach the fixed point at each time step with $\epsilon_\theta(x_t, t) = \epsilon$. For image restoration tasks, the initial inputs are contaminated with unknown degradations. In this way, we design the neural estimation correction by combining the current denoiser output with the previous denoiser output as the regularization term. We form the optimization as

$$\bar{\epsilon}^t_\theta = \epsilon^t_\theta + \arg\min_{\Delta} \left(d(\epsilon^{t+1}_\theta, (\epsilon^t_\theta + \Delta))\right), \tag{28}$$

where the l_1 loss is used for the distance metric $d(\cdot, \cdot)$.

Based on the above discussion, we summarize the detailed algorithm of our proposed method namely DiffECC in Algorithm 1. The overall framework of our sampling method is demonstrated in Fig. 2.

Experiments

We test our proposed method on image SR, *Gaussian* deblurring and *motion* deblurring. In particular, the forward measurement operator for image SR is performed with bicubic down-sampling. For *Gaussian* deblurring, the kernel has the size of 61× 61 with a standard deviation of 3.0. The *motion* deblurring is with the kernel size of 61× 61 and the intensity value being 0.5. All tasks can be formulated by convolving the kernels with ground truth images.

Experimental Setup

Dataset. For vision tasks using face images, we test our experiment on the Flickr Faces High Quality (FFHQ) dataset (Karras, Laine, and Aila 2019). We sample 1k images for evaluation, which are of size 256×256 pixels. For vision tasks using natural images, we evaluate quantitative results on the ImageNet test dataset (Deng et al. 2009) as (Kawar et al. 2022a), with 1k validation images of size 256×256 pixels. All images are normalized to the range $[0, 1]$. Problem-specific pre-trained diffusion models for face images and natural images are taken from (Choi et al. 2021) and (Dhariwal and Nichol 2021) respectively.

Quantitative metrics. For quantitative comparison, we evaluate different methods with the standard distortion metrics Peak Signal Noise Ratio (PSNR) (dB) and Structural Similarity Index (SSIM) (Wang et al. 2004) (higher is better), as well as widely-used perceptual metrics Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018) and Frechet Inception Distance (FID) (Heusel et al. 2017) (lower is better). PSNR and SSIM measure the faithfulness of reconstructed images, which is not important but necessary for image restoration tasks. LPIPS measures the perceptual similarity between the generated image and the original high-quality image. FID evaluates the quality and diversity between generated distribution and data distribution. The sampling time is measured by the number of function evaluations (NFE).

Experimental Results

We perform comparisons with four state-of-the-art methods, including DPS (Chung et al. 2023), denoising diffusion restoration models (DDRM) (Kawar et al. 2022a), DDNM (Wang, Yu, and Zhang 2022) and denoising diffusion models for plug-and-play IR (DiffPIR) (Zhu et al. 2023). The same pre-trained diffusion models, degradation kernels, and validation datasets are employed for all methods in comparisons for fairness.

Quantitative evaluations on FFHQ and ImageNet 256×256-1k validation datasets are provided in Table 1 and Table 2 respectively. The qualitative comparisons for image 4× SR with $\sigma_n = 0.05$ are shown in Fig. 3. The results demonstrate that DiffECC achieves superior performance compared to other methods. We provide extended quantitative and qualitative results with different scaling factors and noise values in the supplementary. Comparison results for real-world image restoration where the forward measurement operator $A(\cdot)$ is unknown are given in the supplementary.

Ablation Studies

Effects of x_{init}. For inverse problems, we perform ablation studies analyzing the effectiveness of starting from different initial predictions in the reverse process. First, we take Real-ESRNet and MPRNet on *motion* deblurring and obtain PSNR scores with 29.27 and 29.64 respectively. Though the predicted error using the CNN model for pre-processing different degradations is contracted less than 1 after transformed to $x_{T'}$ after multiplying a factor of $\sqrt{\bar{\alpha}_{T'}}$, the accuracy of the predicted regressed image affects the result

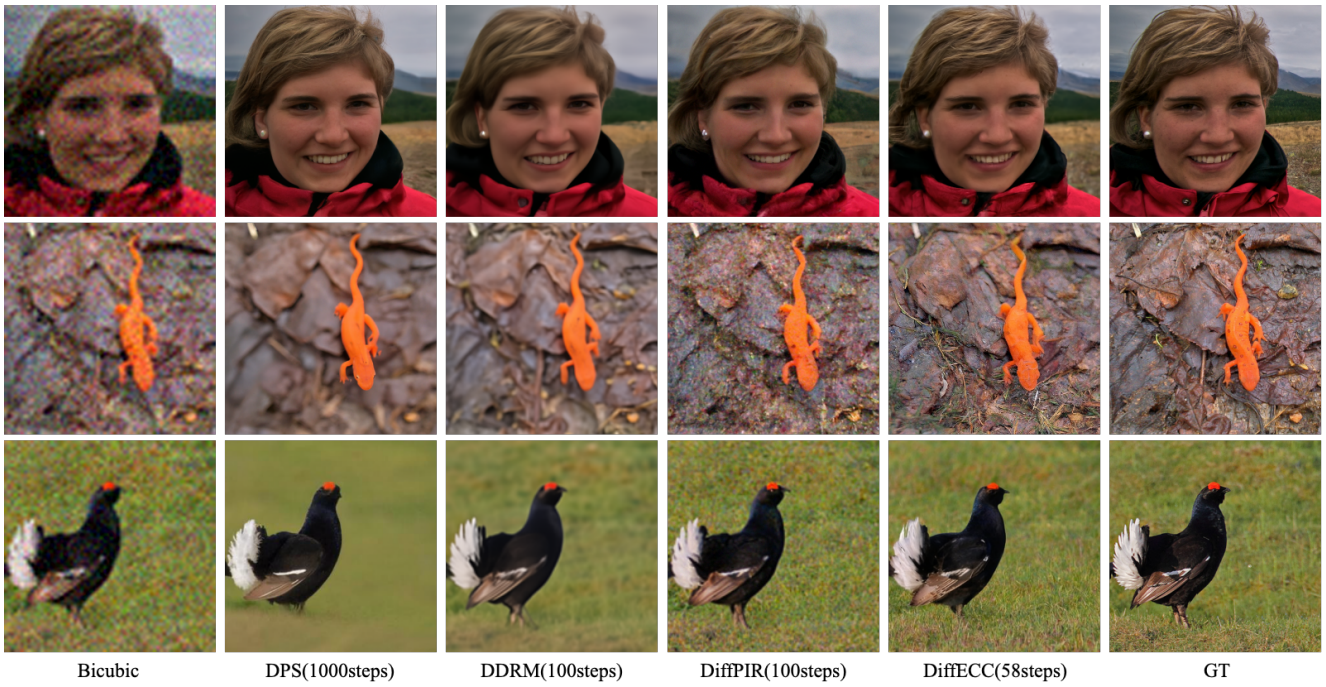

Figure 3: Visual comparisons of $\times 4$ image SR ($\sigma_n = 0.05$) on FFHQ and ImageNet 256×256-1k validation datasets.

FFHQ		Super-resolution (×4)				Deblur (*Gaussian*)				Deblur (*motion*)			
Method	NFEs ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
DiffECC	58	**28.47**	**0.9140**	**0.1843**	**24.08**	**26.23**	**0.8789**	**0.2455**	**27.17**	**27.06**	**0.8922**	**0.2465**	**23.67**
DiffPIR	100	26.73	0.8812	0.2571	25.36	24.85	0.8670	0.2838	28.27	26.98	0.8887	0.2477	24.98
DDRM	100	27.52	0.8758	0.2455	45.84	25.50	0.8427	0.2813	52.10	-	-	-	-
DPS	1000	24.02	0.8333	0.3034	34.56	25.34	0.8424	0.2581	28.37	21.61	0.7961	0.3266	30.83

Table 1: Quantitative results (PSNR, SSIM, LPIPS, and FID) of solving inverse problems: super-resolution, *Gaussian* deblur and *Motion* deblur with $\sigma_n = 0.05$ on FFHQ 256×256-1k validation dataset. Black colors in bold indicate the best scores.

ImageNet		Super-resolution (×4)				Deblur (gaussian)				Deblur (motion)			
Method	NFEs ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
DiffECC	58	**23.88**	**0.7815**	**0.3470**	**43.04**	**22.35**	**0.7667**	**0.3961**	**55.14**	**24.04**	**0.8045**	**0.3470**	**48.32**
DiffPIR	100	22.99	0.7045	0.4157	55.45	21.71	0.7246	0.4286	60.7	13.55	0.2877	0.6899	162.06
DDRM	100	22.36	0.7221	0.3869	56.54	22.84	0.7092	0.4290	75.37	-	-	-	-
DPS	1000	21.07	0.7213	0.4612	67.46	19.76	0.5990	0.4342	65.62	19.18	0.6772	0.4647	66.88

Table 2: Quantitative results (PSNR, SSIM, LPIPS, and FID) of solving inverse problems: super-resolution, *Gaussian* deblur and *Motion* deblur with $\sigma_n = 0.05$ on ImageNet 256×256-1k validation dataset. Black colors in bold indicate the best scores.

of the diffusion model. Since Real-ESRNet obtains coarse blind restoration results, in general, the output obtained by Real-ESRNet can be used as the initial value. We further perform $4\times$ noisy SR ($\sigma_n = 0.03$) experiment on images. Quantitative comparisons are listed in Table 3.

model-1 denotes that x'_T is calculated as

$$T' = T = 50, x'_T = \sqrt{\overline{\alpha}_{T'}} x_0 + \sqrt{1 - \overline{\alpha}_{T'}} \cdot \epsilon. \quad (29)$$

model-2 denotes that x'_T is calculated as

$$T = 100, T' = 50, x'_T = \sqrt{\overline{\alpha}_{T'}} x_0 + \sqrt{1 - \overline{\alpha}_{T'}} \cdot \epsilon. \quad (30)$$

model-3 denotes that the initialization is constructed by the DDIM sampling inversion technique. DDIM inversion procedure is the inverted scheduler of DDIM (Song, Meng, and Ermon 2020) scheduler. The reversed ODE process in the limit of small steps is calculated as

$$x_{t+1} = \sqrt{\frac{\alpha_{t+1}}{\alpha_t}} x_t + \left(\sqrt{\frac{1}{\alpha_{t+1}} - 1} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \cdot \epsilon_\theta(x_t, t). \quad (31)$$

We formulate the trajectory from x_0 to $x_{T'}$, where $T =$

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
model-1	22.38	0.7244	0.3838	49.52
model-2	24.39	0.7937	0.3394	43.48
model-3	24.55	0.8002	0.3983	96.20
model-4	**24.68**	**0.8004**	**0.3079**	**35.52**

Table 3: Quantitative evaluation of the image with different initialization strategy from ImageNet validation.

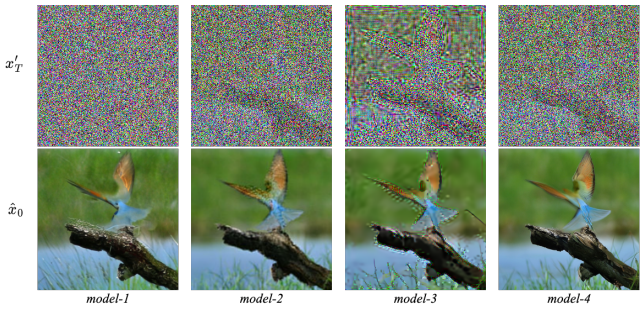

Figure 4: Visualization of restoration results with different initialization process.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
wo restart	23.97	0.7764	0.3187	37.59
w restart	**24.68**	**0.8004**	**0.3079**	**35.52**

Table 4: Quantitative evaluation of the restart sampling strategy.

$100, T' = 50$.

model-4 indicates that the output of the simple Real-ESRNet model is used as the initial prediction. The formulation is calculated as:

$$T = 100, T' = 50, x_{T'} = \sqrt{\overline{\alpha}_{T'}} f(y;\psi) + \sqrt{1-\overline{\alpha}_{T'}}\epsilon. \tag{32}$$

From Table 3 we can see that combining the generally pretrained neural network Real-ESRNet with the diffusion model attributes to an error contraction. Visualization of different starting points are shown in Fig. 4.

Effects of restart sampling algorithm. To investigate the effect of restart sampling algorithm, we perform two experiments: without restart sampling algorithm (T'=58, NFEs=58) and with restart sampling algorithm (T'=50, NFEs=58). It is evident from Table 4 that the restart strategy harnesses and enhances the reconstruction ability by providing error contraction effects. To illustrate the effects of the hyperparameters K and k_{skip} in restart strategy, we show the reconstructed images of SR samples in Fig. 5. Hyperparameters are fixed as $K = 40, k_{\text{skip}} = 32$ since the generated images tend to be more stable.

Effects of components in error correction strategy. To analyze the impact of the iteratively corrected $\hat{x}'_{0|t}$ and $\hat{\epsilon}_\theta$ in error correction strategy, we show in Table 5 how quantitative results change with different operations.

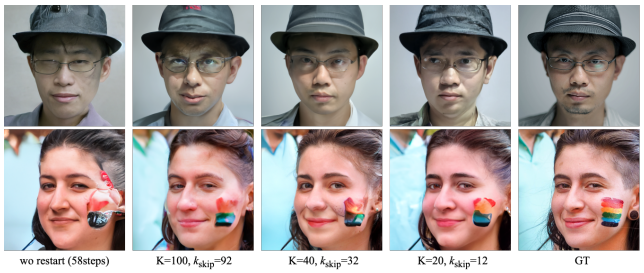

Figure 5: Effect of hyperparameters K and k_{skip}.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Model-1	23.43	0.7179	0.3615	54.61
Model-2	23.91	0.7763	0.3192	38.24
Model-3	**24.68**	**0.8004**	**0.3079**	**35.52**

Table 5: Quantitative evaluation of the error correction strategy.

In *Model-1*, the reverse diffusion process is calculated as

$$\begin{aligned} \hat{x}'_{0|t} &= \hat{x}_0(x_t) - \frac{\zeta}{\sigma_y^2}\cdot\nabla_{x_t}\|y - A(\hat{x}_0(x_t))\|_2^2 \\ x_{t-1} &= \sqrt{\overline{\alpha}_{t-1}}\hat{x}'_{0|t} + \sqrt{1-\overline{\alpha}_{t-1}}\cdot\epsilon_\theta^t. \end{aligned} \tag{33}$$

In *Model-2*, the reverse diffusion process is calculated as

$$\begin{aligned} \hat{x}'_{0|t} &= \hat{x}_0(x_t) - \frac{\zeta}{\sigma_y^2}\cdot\nabla_{x_t}\|y - A(\hat{x}_0(x_t))\|_2^2 \\ \hat{\epsilon}_\theta^t &= f_{\epsilon_\theta|x_0}(x_t, \hat{x}'_{0|t}) \\ x_{t-1} &= \sqrt{\overline{\alpha}_{t-1}}\hat{x}'_{0|t} + \sqrt{1-\overline{\alpha}_{t-1}}\cdot\hat{\epsilon}_\theta^t. \end{aligned} \tag{34}$$

In *Model-3*, the reverse diffusion process is calculated as

$$\begin{aligned} \overline{\epsilon}_\theta^t &= \epsilon_\theta^t + \arg\min_{\Delta}\left(d(\epsilon_\theta^{t+1}, (\epsilon_\theta^t + \Delta))\right) \\ \hat{x}_{0|t} &= f_{x_0|\epsilon_\theta}(x_t, \overline{\epsilon}_\theta^t) \\ \hat{x}'_{0|t} &= \hat{x}_{0|t} - \frac{\zeta}{\sigma_y^2}\cdot\nabla_{x_t}\left\|y - A\left(\hat{x}_{0|t}(x_t)\right)\right\|_2^2 \\ \hat{\epsilon}_\theta^t &= f_{\epsilon_\theta|x_0}(x_t, \hat{x}'_{0|t}) \\ x_{t-1} &= \sqrt{\overline{\alpha}_{t-1}}\hat{x}'_{0|t} + \sqrt{1-\overline{\alpha}_{t-1}}\cdot\hat{\epsilon}_\theta^t. \end{aligned} \tag{35}$$

Conclusion

In this paper, we introduce a diffusion model-based sampling technique with error contraction and error correction strategies for image restoration, referred to as DiffECC. Specifically, by integrating existing neural network techniques and interweaving a restart diffusion sampling process, the error contraction method improves the visual quality for inverse problems. In the error correction method, we incorporate the denoiser into optimization algorithms with iterative correction in the backward sampling process. Extensive experimental results highlight the superior performance of DiffECC in comparison to other methods.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (Nos. 62171251 & 62311530100) and the Special Foundations for the Development of Strategic Emerging Industries of Shenzhen (Nos.JSGG 20211108092812020 & CJGJZD 20210408092804011).

References

Abu-Hussein, S.; Tirer, T.; and Giryes, R. 2022. ADIR: Adaptive Diffusion for Image Reconstruction. *arXiv preprint arXiv:2212.03221*.

Choi, J.; Kim, S.; Jeong, Y.; Gwon, Y.; and Yoon, S. 2021. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*.

Chung, H.; Kim, J.; Mccann, M. T.; Klasky, M. L.; and Ye, J. C. 2023. Diffusion Posterior Sampling for General Noisy Inverse Problems. In *The Eleventh International Conference on Learning Representations*.

Chung, H.; Lee, S.; and Ye, J. C. 2023. Fast Diffusion Sampler for Inverse Problems by Geometric Decomposition. *arXiv preprint arXiv:2303.05754*.

Chung, H.; Sim, B.; and Ye, J. C. 2022a. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12413–12422.

Chung, H.; Sim, B.; and Ye, J. C. 2022b. Improving Diffusion Models for Inverse Problems using Manifold Constraints.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34: 8780–8794.

Fabian, Z.; Tinaz, B.; and Soltanolkotabi, M. 2023. DiracDiffusion: Denoising and Incremental Reconstruction with Assured Data-Consistency. *arXiv preprint arXiv:2303.14353*.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33: 6840–6851.

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

Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401–4410.

Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022a. Denoising Diffusion Restoration Models. In *Advances in Neural Information Processing Systems*.

Kawar, B.; Song, J.; Ermon, S.; and Elad, M. 2022b. Jpeg artifact correction using denoising diffusion restoration models. *arXiv preprint arXiv:2209.11888*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Liu, X.; Park, D. H.; Azadi, S.; Zhang, G.; Chopikyan, A.; Hu, Y.; Shi, H.; Rohrbach, A.; and Darrell, T. 2023. More control for free! image synthesis with semantic diffusion guidance. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 289–299.

Meng, X.; and Kabashima, Y. 2022. Diffusion Model Based Posterior Samplng for Noisy Linear Inverse Problems. *arXiv preprint arXiv:2211.12343*.

Mokady, R.; Hertz, A.; Aberman, K.; Pritch, Y.; and Cohen-Or, D. 2022. Null-text Inversion for Editing Real Images using Guided Diffusion Models. *arXiv preprint arXiv:2211.09794*.

Murata, N.; Saito, K.; Lai, C.-H.; Takida, Y.; Uesaka, T.; Mitsufuji, Y.; and Ermon, S. 2023. GibbsDDRM: A Partially Collapsed Gibbs Sampler for Solving Blind Inverse Problems with Denoising Diffusion Restoration. *arXiv preprint arXiv:2301.12686*.

Nichol, A. Q.; and Dhariwal, P. 2021. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, 8162–8171. PMLR.

Permenter, F.; and Yuan, C. 2023. Interpreting and Improving Diffusion Models Using the Euclidean Distance Function. *arXiv preprint arXiv:2306.04848*.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Song, J.; Vahdat, A.; Mardani, M.; and Kautz, J. 2023. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*.

Song, Y.; and Ermon, S. 2020. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33: 12438–12448.

Song, Y.; Shen, L.; Xing, L.; and Ermon, S. 2021. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1905–1914.

Wang, Y.; Yu, J.; and Zhang, J. 2022. Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model. *arXiv preprint arXiv:2212.00490*.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Wang, Z.; Zhang, Z.; Zhang, X.; Zheng, H.; Zhou, M.; Zhang, Y.; and Wang, Y. 2023. DR2: Diffusion-based Robust Degradation Remover for Blind Face Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1704–1713.

Xu, Y.; Deng, M.; Cheng, X.; Tian, Y.; Liu, Z.; and Jaakkola, T. 2023. Restart Sampling for Improving Generative Processes. *arXiv preprint arXiv:2306.14878*.

Yue, Z.; and Loy, C. C. 2022. DifFace: Blind Face Restoration with Diffused Error Contraction. *arXiv preprint arXiv:2212.06512*.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2020. Learning Enriched Features for Real Image Restoration and Enhancement. In *ECCV*.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2021. Multi-Stage Progressive Image Restoration. In *CVPR*.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2022. Learning enriched features for fast image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 45(2): 1934–1948.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhu, Y.; Zhang, K.; Liang, J.; Cao, J.; Wen, B.; Timofte, R.; and Gool, L. V. 2023. Denoising Diffusion Models for Plug-and-Play Image Restoration. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (NTIRE)*.