

# RDDM: 残差去噪扩散模型

## 摘要

随着扩散模型在图像生成与恢复领域的广泛应用，传统基于去噪的扩散模型在图像恢复任务中的局限性逐渐显现，如缺乏可解释性等问题。在此背景下，残差去噪扩散模型 (RDDM) 应运而生，它提出了一种全新的双重扩散过程，将传统扩散过程解耦为残差扩散和噪声扩散，旨在解决单一去噪过程在图像恢复中的不可解释性问题，并统一图像生成与恢复任务。本文基于论文作者公开的源代码进行 RDDM 复现，通过对 CelebA、LOL 和 RainDrop 数据集进行实验比较不同采样策略的优劣，最后结果具体情况需要具体分析；此外，本文还验证了 RDDM 与 DDIM 在图像生成任务性能极小差距以及 RDDM 在处理退化图像问题上的有效性。

**关键词：**残差扩散模型；图像恢复；图像生成

## 1 引言

在扩散模型逐渐成为图像生成或者图像恢复任务首选主流模型的今天，一个常引发思考的问题是：基于去噪的扩散模型，如 DDPM [5]，DDIM [17] 等模型仅仅关注去噪方面是否存在局限性？当下，基于扩散原理的图像恢复方法 [6, 9, 11, 13, 14]，虽然将扩散模型拓展应用到了图像恢复任务里，采用退化图像作为条件输入，间接地引导反向生成过程，但是原始的去噪扩散过程依然保持不变。但实际上，由于退化图像是已知的，从纯粹的噪声状态开始的反向过程就显得有些多余。另外，因为扩散过程并未包含任何与退化图像相关的信息，所以其在图像恢复任务中缺乏明确的可解释性，正如图 1(a) 所展示的那样。

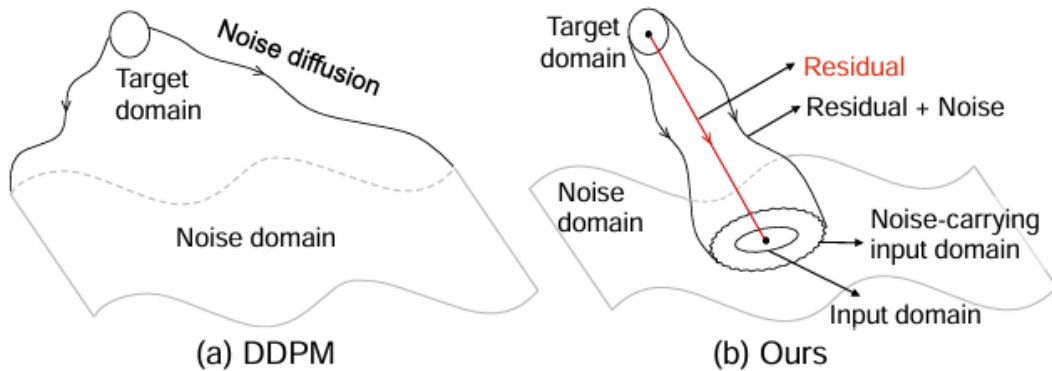


图 1. 前向过程的对比。(a) DDPM 的去噪扩散过程。(b) RDDM 的去噪扩散过程

针对这一问题，RDDM 这篇论文开创性地探索了一种全新的双重扩散过程，进而提出了残差去噪扩散模型（RDDM）。这一模型有效地解决了单一去噪过程在图像恢复中存在的可解释性问题。在 RDDM 架构中，以往的扩散过程被解耦为残差扩散和噪声扩散两个部分。残差扩散侧重确定性，其本质是从目标图像朝着条件输入图像进行的定向扩散；而噪声扩散则强调多样性，代表着扩散过程中所产生的随机扰动因素。因此，RDDM 能够统一对确定性或多样性有着不同程度需求的任务，像图像生成以及图像恢复等任务。相较于传统的基于去噪扩散的图像恢复方法，RDDM 的残差能够明确地指示前向扩散的方向，同时为图像恢复的逆向生成过程提供了指导，这一点在图 1(b) 中有着直观的体现。

## 2 相关工作

去噪扩散模型，诸如 DDPM [5]、SGM [18,19] 和 DDIM [17] 等，其最初的设计初衷是为了实现图像的生成任务。此后，基于 DDPM 和 DDIM 发展而来的图像恢复方法，像 DvSR [20]、SR3 [15] 和 WeatherDiffusion [11] 等，将退化后的图像作为条件输入到去噪网络中，然而，这些方法往往需要数量较大的采样步骤以及较大的批量规模，这在一定程度上增加了计算成本和时间开销。

从图像恢复任务的角度来看，这些方法中从纯粹噪声状态开始的反向过程，不仅显得多余，而且在效率方面表现欠佳。鉴于此，SDEdit [10]、ColdDiffusion [1]、InDI [3] 和 I2SB [7] 等方法提出了直接从降级图像或者带有噪声的降级图像出发来生成清晰图像的新思路。其中，InDI 和 I2SB 还构建了统一的图像生成与修复框架，这些框架与本文所提出的残差去噪扩散模型（RDDM）有着较为紧密的联系。具体而言，InDI、I2SB 以及本文提出的 RDDM 在正向扩散过程中，始终运用三个项（即输入图像  $I_{in}$ 、目标图像  $I_0$  和噪声  $\epsilon$ ）的混合模式，这相较于传统的基于去噪的扩散模型（其仅包含  $I_0$  和  $\epsilon$  两个项的混合）而言，是一种创新和突破。不过，InDI 和 I2SB 选择通过估计目标图像或其线性变换项来替代噪声估计，这实际上类似于 RDDM（SM-Res）中的一种特殊情况。

## 3 背景

去噪扩散模型 [5,16] 的最终目的是学习一个分布  $p(\theta) := \int p_\theta(I_{0:T}) dI_{1:T}$ ，从而能够近似目标的数据分布  $q(I_0)$ ，其中  $I_0$  是目标图像， $I_1 \cdots I_T$  是与  $I_0$  具有相同维度的添加了噪声的潜在空间图像。在前向过程中，使用固定的马尔科夫链将  $q(I_0)$  扩散成具有高斯分布的纯噪声图像：

$$q(I_{1:T}|I_0) := \prod_{i=1}^T q(I_i|I_{i-1}) \quad (1)$$

$$q(I_t|I_{t-1}) := \mathcal{N}(I_t; \sqrt{\alpha_t}I_{t-1}, (1 - \alpha_t)\mathbf{I}) \quad (2)$$

其中  $\alpha_{1:T} \in (0, 1]^T$ 。数据分布  $q(I_t|I_{t-1})$  也可以写成  $I_t = \sqrt{\alpha_t}I_{t-1} + \sqrt{1 - \alpha_t}\epsilon_{t-1}$ 。事实上，通过重参数化从  $I_0$  中采样  $I_t$  更简单：

$$I_t = \sqrt{\bar{\alpha}_t}I_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon \quad (3)$$

其中  $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ,  $\bar{\alpha} := \prod_{s=1}^t \alpha_s$ 。逆向生成过程是从  $p_\theta(I_T) \sim \mathcal{N}(I_T; \mathbf{0}, \mathbf{I})$  开始的马尔可夫链:

$$p_\theta(I_{0:T}) := p_\theta(I_T) \prod_{t=1}^T p_\theta(I_{t-1}|I_t) \quad (4)$$

$$p_\theta(I_{t-1}|I_t) := \mathcal{N}(I_{t-1}; \mu_\theta(I_t, t), \Sigma, \mathbf{I}) \quad (5)$$

其中,  $p_\theta(I_{t-1}|I_t)$  是可学习的转移概率 (方差时间表  $\Sigma_t$  是固定的)。与 DDPM 有关的简化的损失函数由  $p_\theta(I_0)$  的最大似然推导出, 即  $L(\theta) := \mathbb{E}_{I_0 \sim q(I_0), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} [\|\epsilon - \epsilon_\theta(I_t, t)\|^2]$  其中,  $\mathbb{E}$  表示数学期望。网络估计的噪声  $\epsilon_\theta$  可用于表示  $p_\theta(I_{t-1}|I_t)$  中的  $\mu_0$ 。还可以从概率分布  $p_\theta(I_{t-1}|I_t)$  逐步采样得到  $I_{t-1}$ 。

## 4 本文方法

RDDM 旨在设计一种双向扩散过程, 以实现对图像生成与图像恢复的统一与阐释。RDDM 将 DDPM 中的  $I_T = \epsilon$  修改为  $I_T = I_{in} + \epsilon$ , 其中  $I_{in}$  代表图像恢复任务中的退化图像 (例如, 存在阴影, 低光或模糊图像)。对于图像生成,  $I_{in}$  被设为 0。这种修改使得 RDDM 与广泛应用的去噪扩散模型是兼容的, 例如,  $I_T = 0 + \epsilon$  即为图像生成中的纯噪声 ( $\epsilon$ )。对于图像恢复,  $I_T$  是待用噪声的恢复图像 ( $I_T + \epsilon$ ), 如图 2 第三列所示。修改后的正向过程是从  $I_0$  到  $I_T = I_{in} + \epsilon$  逐步将  $I_0$  降质为  $I_{in}$ , 并添加噪声  $\epsilon$ 。这就引出了双扩散过程: 从  $I_0$  到  $I_{in}$  的转化是残差扩散, 添加噪声的过程是噪声扩散。例如, 从无阴影图像  $I_0$  到带有噪声的阴影图像  $I_T$  的正向扩散过程涉及逐步添加阴影和噪声, 如图 2 中的第二行所示。

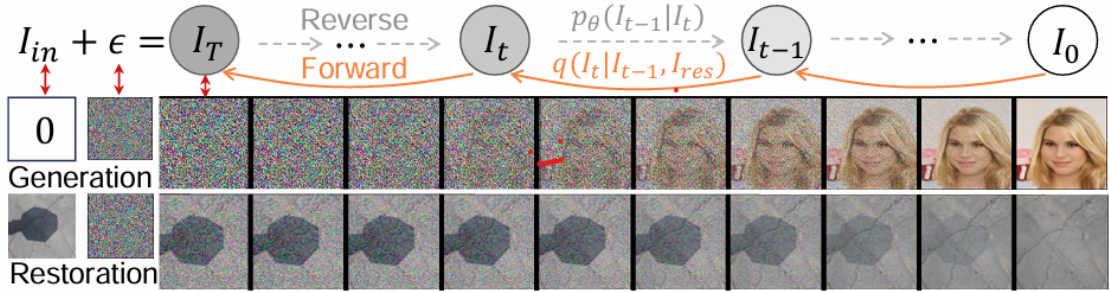


图 2. RDDM 中图像生成和恢复的统一框架 (以阴影去除任务为例)

在接下来的小节中, 我们将对 RDDM 背后的基础理论与方法进行详细阐述。受到残差学习 [4] 的启发, 我们在 4.1 小节重新定义了正向扩散过程的各个步骤。针对反向过程, 我们在 4.2 小节提出了一个训练目标, 用于预测正向过程中所注入的残差与噪声。在 4.3 小节, 我们提出了三种采样方法, 分别是残差预测 (SM-Res)、噪声预测 (SM-N) 以及“残差和噪声预测” (SM-Res-N)。

### 4.1 受扰动的定向残差扩散过程

为了模拟图像质量的逐渐退化和噪声的增加, 我们将 RDDM 中的单个前向处理步骤定义如下:

$$I_t = I_{t-1} + I_{res}^t, \quad I_{res}^t \sim \mathcal{N}(\alpha_t I_{res}, \beta_t^2 \mathbf{I}) \quad (6)$$

其中,  $I_{res}^t$  表示有从状态  $I_{t-1}$  到状态  $I_t$  的随机扰动和定向平均位移 (残差扩散),  $I_{res}^t$  中的残差  $I_{res}$  是  $I_{in}$  与  $I_0$  的差值 (即  $I_{res} = I_{in} - I_0$ ), 两个独立的系数表  $\alpha_t$  和  $\beta_t$  分别控制残差和噪声扩散。实际上, 类似于式子 (2), 从  $I_0$  采样  $I_t$  更简单,

$$\begin{aligned} I_t &= I_{t-1} + \alpha_t I_{res} + \beta_t \epsilon_{t-1} \\ &= I_{t-2} + (\alpha_t + \alpha_{t-1}) I_{res} + (\sqrt{\beta_{t-1}^2 + \beta_t^2}) \epsilon_{t-2} \\ &= \dots \\ &= I_0 + \bar{\alpha}_t I_{res} + \bar{\beta}_t \epsilon \end{aligned} \quad (7)$$

其中  $\epsilon_{t-1}, \dots, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ,  $\bar{\alpha}_t = \sum_{i=1}^t \alpha_i$  和  $\bar{\beta}_t = \sqrt{\sum_{i=1}^t \beta_i^2}$ 。如果  $t = T$ ,  $\bar{\alpha}_t = 1$  并且  $I_T = I_{in} + \bar{\beta}_T \epsilon$ 。  $\bar{\beta}_T$  用来控制图像恢复的噪声扰动强度, 比如  $\bar{\beta}_T^2 = 0.01$  用于图像恢复中的英文去噪。通过式子 (6) 可以定义前向过程中的联合概率分布:

$$q(I_{1:T} | I_0, I_{res}) := \prod_{t=1}^T q(I_t | I_{t-1}, I_{res}) \quad (8)$$

$$q(I_t | I_{t-1}, I_{res}) := \mathcal{N}(I_t; I_{t-1} + \alpha_t I_{res}, \beta_t^2 \mathbf{I}) \quad (9)$$

式子 (7) 定义了边缘概率分布  $q(I_t | I_0, I_{res}) = \mathcal{N}(I_t; I_0 + \bar{\alpha}_t I_{res}, \beta_t^2 \mathbf{I})$ 。本文的 RDDM 的前向过程是三个项 (即  $I_0$ ,  $I_{res}$  和  $\epsilon$ ) 的混合物, 即 RDDM 可以视作广泛使用的去噪扩散模型的扩展。

## 4.2 生成过程及训练目标

在正向过程 (式子 (7)) 中, 残差 ( $I_{res}$ ) 和噪声 ( $\epsilon$ ) 逐渐加入到  $I_0$  中, 然后合成  $I_t$ , 而从  $I_t$  到  $I_0$  的反向过程则是对正向过程中注入的残差和噪声进行估计。可以训练残差网络  $I_{res}^\theta(I_t, t, I_{in})$  来预测残差  $I_{res}$ , 训练噪声网络  $\epsilon_\theta(I_t, t, I_{in})$  去预测噪声  $\epsilon$ 。利用式子 (7), 可以得到估计的目标图像  $I_0^\theta = I_t - \bar{\alpha}_t I_{res}^\theta - \bar{\beta}_t \epsilon_\theta$ 。如果给出  $I_0^\theta$  和  $I_{res}^\theta$ , 则图像生成的过程可以定义为:

$$p_\theta(I_{t-1} | I_t) := q_\sigma(I_{t-1} | I_t, I_0^\theta, I_{res}^\theta) \quad (10)$$

其中, 从  $I_t$  到  $I_{t-1}$  的转移概率  $q_\sigma(I_{t-1} | I_t, I_0^\theta, I_{res}^\theta)^2$  为:

$$q_\sigma(I_{t-1} | I_t, I_0, I_{res}) = \mathcal{N}(I_{t-1}; I_0 + \bar{\alpha}_{t-1} I_{res} + \sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2} \frac{I_t - (I_0 + \bar{\alpha}_t I_{res})}{\bar{\beta}_t}, \sigma_t^2 \mathbf{I}) \quad (11)$$

其中  $\sigma_t^2 = \eta \bar{\beta}_t^2 \bar{\beta}_{t-1}^2 / \bar{\beta}_t^2$ ,  $\eta$  控制生成过程是随机的还是确定的。使用式子 (10) 和式子 (11), 可以通过以下方式从  $I_t$  中抽取  $I_{t-1}$ :

$$I_{t-1} = I_t - (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I_{res}^\theta - (\bar{\beta}_t - \sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}) \epsilon_\theta + \sigma_t \epsilon_t \quad (12)$$

其中,  $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 。当  $\eta = 1$  时, RDDM 具有和约束方差, 而 DDPM 具有保留方差。当  $\eta = 0$  (即  $\sigma_t = 0$  时), 采样过程时确定性的。

$$I_{t-1} = I_t - (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I_{res}^\theta - (\bar{\beta}_t - \bar{\beta}_{t-1}) \epsilon_\theta \quad (13)$$

另外，RDDM 用于训练的损失函数公式如下

$$L_{res}(\theta) := \mathbb{E}[\lambda_{res} \|I_{res} - I_{res}^\theta(I_t, t, I_{ub})\|^2] \quad (14)$$

$$L_\epsilon(\theta) := \mathbb{E}[\lambda_\epsilon \|\epsilon - \epsilon_\theta(I_t, t, I_{in})\|] \quad (15)$$

其中，超参数  $\lambda_{res}, \lambda_\epsilon \in \{0, 1\}$ ，并且训练的输入图像  $I_t$  通过式子 (7) 使用  $I_0, I_{res}$  和  $\lambda$  进行合成。也可以用  $I_{in}$  合成（用  $I_0 = I_{in} - I_{res}$  代替式子 (7) 中的  $I_0$ ），

$$I_t = I_{in} + (\bar{\alpha}_t - 1)I_{res} + \bar{\beta}_t\epsilon \quad (16)$$

### 4.3 采样方法策略集

对于生成过程（从  $I_t$  到  $I_{t-1}$ ）， $I_t$  和  $I_{in}$  是已知的，因此  $I_{res}$  和  $\epsilon$  可以用式子 (16) 表示。此外，根据式子 (14)，(15)，(16)，RDDM 提出了如下三种采样方法：

- SM-RES：当  $\lambda_{res} = 1$  和  $\lambda_\epsilon = 0$  时，模型将通过网络预测残差  $I_{res}^\theta$ ，而用式子 (16) 对噪声  $\epsilon_\theta$  进行线性变换。
- SM-N：当  $\lambda_{res} = 0$  和  $\lambda_\epsilon = 1$  时，模型将通过网络预测噪声  $\epsilon_\theta$ ，而用式子 (16) 对残差  $I_{res}^\theta$  进行线性变换。
- SM-RES-N：当  $\lambda_{res} = 0$  和  $\lambda_\epsilon = 1$  时，模型将通过网络预测噪声和残差。

## 5 复现细节

### 5.1 代码说明

文章作者在今年陆续地将源代码公开，因此本文是基于作者已公开的 RDDM 功能模块的源代码进行复现的。RDDM 包含多个功能模块，每个功能模块在进行训练时都需要特定的操作。具体而言，在训练过程中，对于每一个功能模块，都需要提供一个训练数据集的文件列表文件（flist），这个文件列表文件详细记录了训练数据集中各个数据文件的路径或名称等相关信息。在实际操作中只需要通过 Python 内置的 os 模块和文件操作库即可完成。对于图像生成和图像恢复在不同数据集上的训练参数严格参照原文附录 Table 4 进行设置。

### 5.2 实验环境搭建

本文通过公开代码的环境文件 install.yaml 创建训练所需要的 conda 环境，即运行“conda env create -f install.yaml”。在搭建环境的过程遇到个别包安装失败，此时对这些安装失败的包进行单独地重装即可。此外，本文的训练工作都是在单张服务器显卡 NVIDIA GeForce RTX 4090 上完成，而图像恢复任务的评估工作是在 matlab 下完成，图像生成任务的评估则是在 python 文件完成。

表 1. 采样方法的分析

采样方法	图像生成 (CelebA)		低光照增强 (LOL)		去雨 (RainDrop)	
	FID(↓)	IS(↑)	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)
SM-Res	45.16	1.32	23.14	0.94	31.69	0.95
SM-N	36.05	1.49	15.56	0.64	20.06	0.71
SM-Res-N	40.73	1.35	23.84	0.93	32.17	0.96

## 6 实验结果分析

本文参照原文章所做的对比实验进行验证，因为实际情况比如硬件条件的不足，在图像生成任务时进行了适应性修改而图像恢复任务遵循原文附录中的参数设置，但是保证不同的方法除了方法本身的区别外，其它客观实验条件相同。

### 6.1 采样方法的分析

在实际应用中，不同的任务在不同环境中可能对应不同的最佳采样方法，本文仿照原文实验对人脸数据集 CelebA [8]、低/高光照数据集 LOL [2] 和雨点图片数据集 RainDrop [12] 进行了不同采样方法的实验分析。其中，CelebA 数据集上的采样步数为 10，LOL 数据上为 2，RainDrop 数据集上为 5。表 1 给出了三种采样方法的分析：

表 1 的数据显示，SM-Res 在图像恢复任务中的表现较为出色，但在图像生成方面，其 FID(Frechet Inception Distance) 值欠佳。与之相反，SM-N 在图像生成方面能产生更优的 FID 和 IS(Inception Score)。然而，在图像恢复任务中，SM-N 的表现不尽人意，例如在低光照增强任务中的 PSNR (Peak Signal-to-Noise Ratio) 仅为 15.56，在去雨任务中的 PSNR 为 20.06。这种差异可能是由于将  $I_{res}$  表示为  $\epsilon_\theta$  的线性变换对于图像恢复任务并不适配。这些不一致的结果应归因于残差侧重于确保确定性，而噪声侧重于体现多样性。在此实验结果的基础上，后续实验将使用 SM-N 进行图像生成，使用 SM-Res 进行 LOL 数据集上的低光照增强，使用 SM-Res-N 进行去雨图像恢复。



## 6.2 图像生成

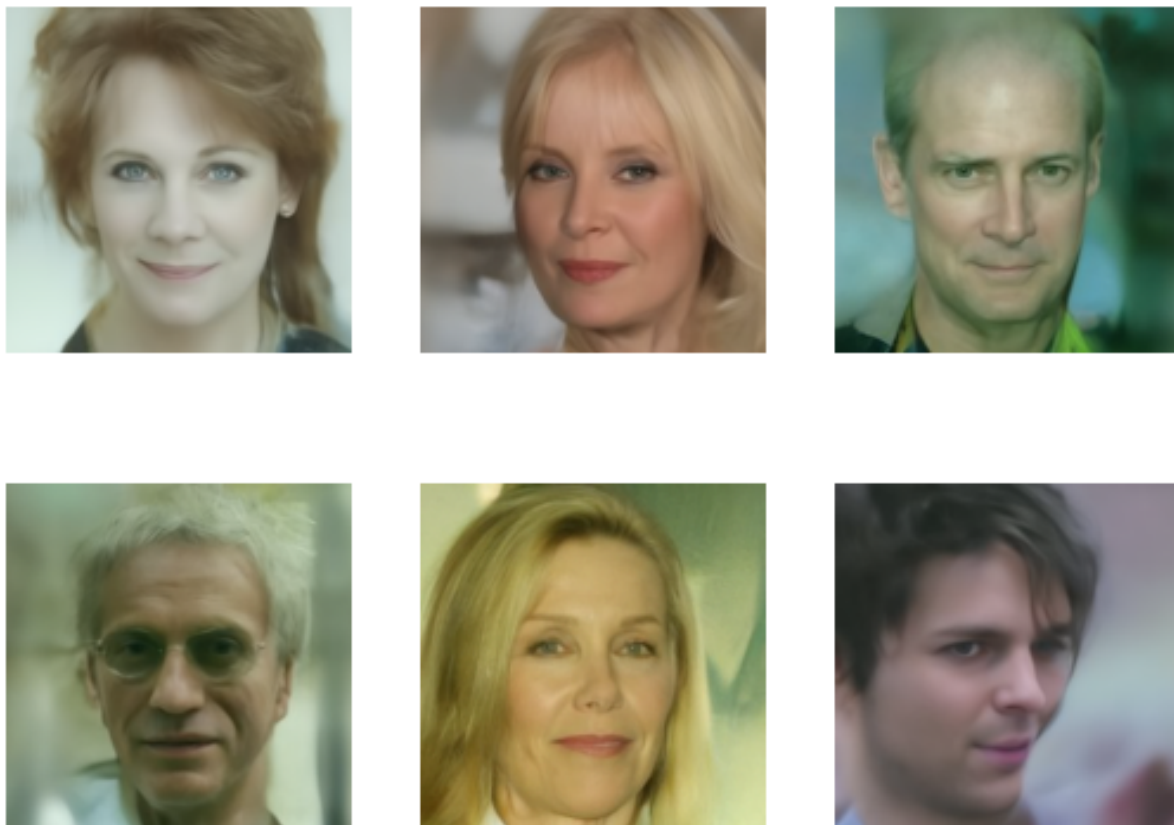


图 3. CelebA  $256 \times 256$  数据集上的图像生成

图 3 显示了 10 个采样步数下生成的人脸图像。原文不仅通过公式推导证明了 RDDM 经过系数变换可以生成和 DDIM 一样的图像，也做了相应的实验证明。本小节将通过改变两个模型的采样步数对 CelebA 数据集进行定量分析，其结果如图 3 所示。图 3 验证了系数变换前后生成图像的质量几乎相同，此外很明显可以观察到如果采用较大的采样步数（如  $\text{steps}=100$ ），性能上相较其他采样步数会有较大（大约相差 1.2），这可能时由系数变换过程中引入的数值误差引起的。这些误差可能会在采样步骤逐渐增大时不断累积和放大。

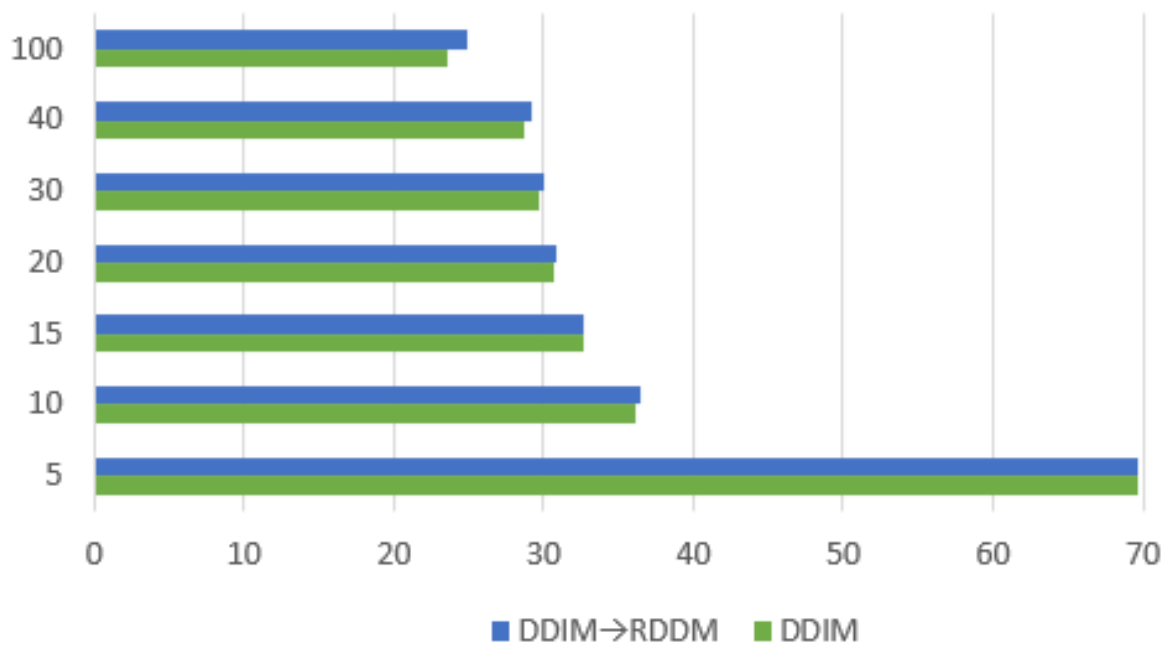


图 4. CelebA  $256 \times 256$  数据集上不同采样步数的图像生成的 FID 分数图

### 6.3 图像恢复

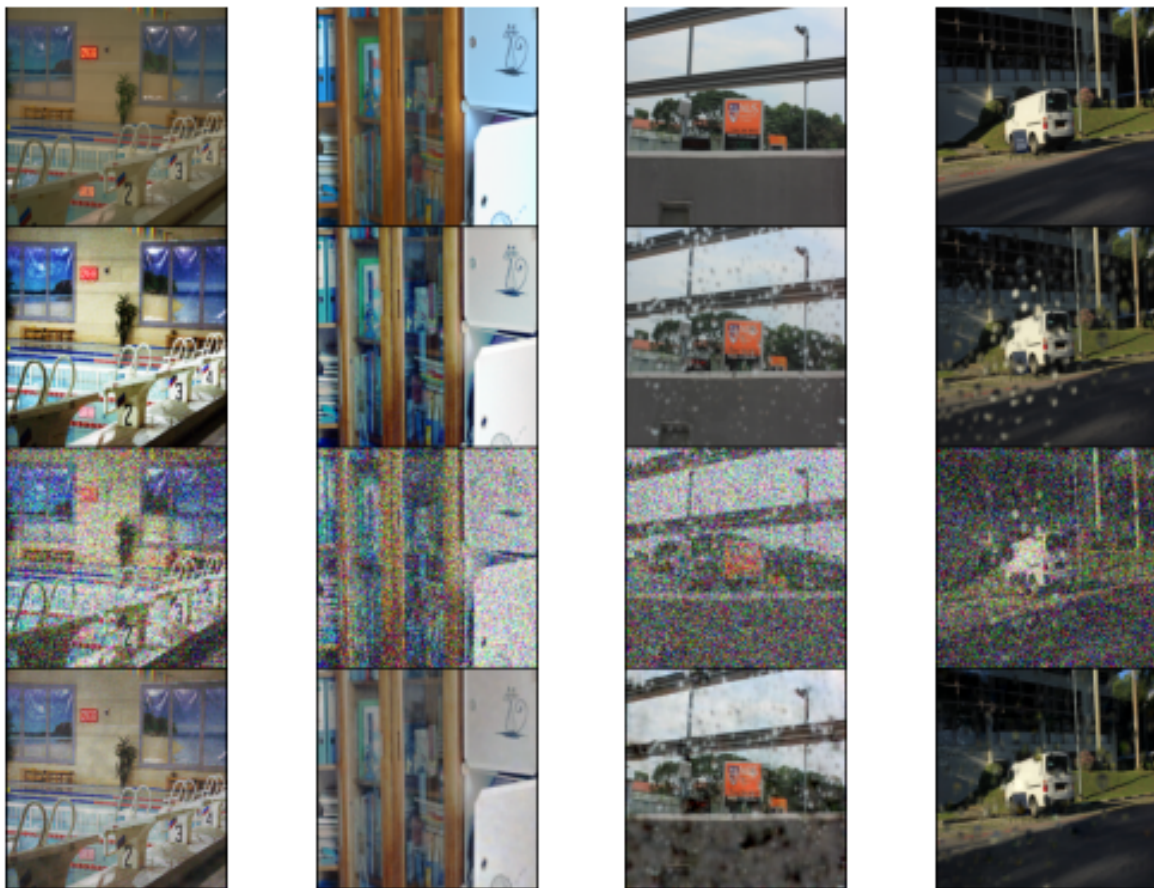


图 5. LOL 数据集和 RainDrop 数据集训练样例图



图 5 展示了 LOL 数据集和 RainDrop 数据集在 RDDM 图像恢复中的样图，其并不是从纯噪声图像开始恢复，而是从退化图像开始。图 6 呈现了这两个数据集的恢复效果。对于 LOL 数据集，经过处理后，室内场景的亮度和清晰度有了显著提升，书架上的书籍和室内设施的纹理等细节都清晰地展现出来，这表明所采用的图像恢复算法在处理 LOL 数据集的退化问题上表现出色。而对于 RainDrop 数据集，处理后的图像中雨滴效果明显减少，车辆、道路设施、背景中的建筑和道路标志等细节都得到了很好的恢复，图像整体质量有了很大的提高，这体现出 RDDM 在去除雨滴和恢复因雨滴导致的图像退化方面有很强的能力。总体来看，这些数据集的采样图和恢复效果图展示了图像从退化到恢复的过程，并且所使用的图像恢复算法在处理 LOL 和 RainDrop 数据集的图像恢复任务时都取得了较好的效果。



图 6. LOL 数据集和 RainDrop 数据集恢复效果（第二行为恢复图）

## 7 总结与展望

本文介绍了残差去噪扩散模型（RDDM），这是一种创新的双扩散模型，通过将传统扩散过程解耦为残差扩散和噪声扩散，解决了单一去噪过程在图像恢复中的不可解释性问题。RDDM 重新定义了正向扩散过程，采用两个独立系数调度控制残差和噪声扩散速度，并提出部分路径无关的生成过程。同时，本文对 RDDM 方法进行了复现，本文首先验证了不同的采样策略在不同的退化环境下的效果，在此基础上得出不同环境下对应的最佳采样策略；此外还验证了 RDDM 和主流扩散模型 DDIM 的等效性，虽然受限于硬件资源，但是得出的结果基本和原论文一致，即 RDDM 通过系数变换可以和主流的扩散模型在图像生成的图片上保持相近的质量，而在图像恢复上能还原出清晰的图像。

通过复现 RDDM，深入理解了扩散模型在图像领域的应用和创新方向。学到了如何通过巧妙地设计模型架构来解决问题，比如利用残差来增强模型在图像恢复中的可解释性。在

未来工作中，一方面除了继续复现 RDDM 的其他功能模块，还可以基于 RDDM 进行进一步优化，例如探索更有效的系数调度策略、改进网络结构以提升性能。另一方面，可以将 RDDM 的思想拓展到其他领域，如视频处理或 3D 数据处理等，同时尝试与其他先进技术结合，开发出功能更强大、应用更广泛的模型，推动相关领域的发展。

## 参考文献

- [1] Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. *Advances in Neural Information Processing Systems*, 36, 2024.
- [2] Wenhan Yang Jiaying Liu Chen Wei, Wenjing Wang. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018.
- [3] Mauricio Delbracio and Peyman Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *arXiv preprint arXiv:2303.11435*, 2023.
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- [6] Yeying Jin, Wenhan Yang, Wei Ye, Yuan Yuan, and Robby T Tan. Shadowdiffusion: Diffusion-based shadow removal using classifier-driven attention and structure preservation. *arXiv preprint arXiv:2211.08089*, 2, 2022.
- [7] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar.  $I^2$  sb: Image-to-image schrödinger bridge. *arXiv preprint arXiv:2302.05872*, 2023.
- [8] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.
- [9] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022.
- [10] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

- [11] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10346–10357, 2023.
- [12] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [13] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- [14] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- [15] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4713–4726, 2022.
- [16] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.
- [17] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.
- [18] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.
- [19] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.
- [20] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022.