

# 基于残差扩散模型的遥感超分辨率图像生成研究

左宪禹<sup>a,b</sup>, 田展硕<sup>a,b</sup>, 殷梦晗<sup>a</sup>, 党兰学<sup>a,b</sup>, 乔保军<sup>a,b</sup>, 刘扬<sup>a,b</sup>, 谢毅<sup>a,b</sup>

(河南大学 a.计算机与信息工程学院; b.河南省大数据分析与管理重点实验室, 河南 开封 475004)

**摘要:**传统基于扩散的图像超分辨率方法通常需要大量采样步骤,并且优化功能强大的扩散模型需要耗费大量运算时间。为了在有限的计算资源上实现训练,现有的加速采样技术往往会牺牲部分图像质量,导致超分辨率结果模糊。为了解决这一问题,提出了一种改进的、高效的残差超分辨扩散模型。通过构建马尔可夫链,在高分辨率图像和低分辨率图像之间移动残差来实现图像之间的转移,有效减少扩散步骤的数量。该方法保证了超分辨率结果的质量和灵活性,同时也提高了转移效率,消除了推理过程中需要的后加速及其相关图像细节特征的退化。实验证明,即使只执行15个采样步骤,所提的方法在合成数据集和真实数据集上也可以获得优于或至少可以与当前最先进方法相当的图像质量。

**关键词:**遥感;超分辨率;图像生成;残差移动;残差扩散模型

**中图分类号:** TP751

**文献标志码:** A

**文章编号:** 1000-2367(2025)03-0058-08

图像生成是近年来计算机视觉领域备受关注的的一个重要领域,也是计算需求最高的领域之一。在视觉处理中,图像超分辨率是一个关键问题,其旨在从低分辨率图像中恢复出高分辨率图像。特别是在遥感领域,超分辨率(super-resolution, SR)技术具有广阔的应用前景。随着对细粒度遥感应用需求的增长,高分辨率(high resolution, HR)遥感影像可以增强低分辨率(low resolution, LR)遥感图像<sup>[1]</sup>,有助于建筑物提取、小目标检测等后续任务的完成<sup>[2]</sup>。然而,受光学和传感器技术限制以及设备更新成本较高的影响,对地观测卫星的光谱和空间分辨率可能无法满足预期要求,导致遥感图像在大尺度或长时间序列应用中存在不足和质量问题<sup>[3]</sup>。超分辨率技术为遥感应用研究提供了一种有效且经济的替代方法,旨在从现有低分辨率观测数据中重建潜在的高分辨率图像的高频信息。

目前,基于深度学习的遥感超分辨率技术备受青睐,它通过学习低分辨率和高分辨率图像之间的空间映射,可以高效地预测低分辨率图像中缺失的高频信息<sup>[4]</sup>。与其他生成模型相比,扩散模型(diffusion model, DM)即使在无条件情况下也能轻松应用于修复、着色或基于笔画的合成任务<sup>[5-7]</sup>。作为基于可能性的模型,扩散模型避免了模式崩溃和训练不稳定性问题,通过充分利用参数共享,能够更好地建模自然图像的复杂分布,而无需像其他模型那样涉及数十亿个参数。

DM是一种概率模型类,它表现出模式覆盖行为,容易对数据中难以察觉的细节进行过度建模,导致计算资源过多。为了解决这个问题,研究人员提出了一种重新加权的变分目标方法<sup>[8]</sup>,该方法通过对初始去噪

**收稿日期:** 2024-03-12; **修回日期:** 2024-03-25.

**基金项目:** 国家自然科学基金(62176087);河南省高校科技创新团队支持计划(24IRTSTHN021);河南省科技攻关项目(242102240021).

**作者简介:** 左宪禹(1979—),男,河南开封人,河南大学教授,博士,研究方向为遥感大数据处理、高效能计算, E-mail: xianyu\_zuo@163.com.

**通信作者:** 谢毅(1983—),男,河南平顶山人,河南大学副教授,博士,研究方向为遥感超分辨率图像生成, E-mail: xieyi@henu.edu.cn.

**引用本文:** 左宪禹, 田展硕, 殷梦晗, 等. 基于残差扩散模型的遥感超分辨率图像生成研究[J]. 河南师范大学学报(自然科学版), 2025, 53(3): 58-65. (Zuo Xianyu, Tian Zhanshuo, Yin Menghan, et al. Remote sensing super-resolution image generation based on residual diffusion model[J]. Journal of Henan Normal University(Natural Science Edition), 2025, 53(3): 58-65. DOI:10.16366/j.cnki.1000-2367.2024.03.12.0002.)

步骤进行欠采样来减少计算负担.然而,即使使用这种方法,DM 模型仍然需要大量计算,因为训练和评估这样的模型涉及 RGB 图像的高维空间中的重复函数评估和梯度计算.遥感图像具有覆盖范围广、分辨率高、多光谱信息、时间动态和数字处理能力等特点,图像具有复杂的高频特征,导致扩散模型采样期间计算资源呈指数级增加.为了增强这个强大的模型类的可访问性,同时显著减少其资源消耗,需要一种方法来减少训练和采样期间的计算需求.

本文提出了一种适用于遥感图像超分辨率的残差扩散模型(residual diffusion model, ResDM),包含更短马尔可夫链的有效扩散模型,用于在 HR 图像和相应的 LR 图像之间进行转换.马尔可夫链的初始状态收敛于 HR 图像的近似分布,而最终状态收敛于 LR 图像的近似分布.同时本文设计了一个过渡内核,通过扩散模型迭代步长递进式地学习 HR 图像在不同噪声强度下的噪声分布,并以 LR 图像为条件逐步转移 HR 与 LR 噪声图像之间的残差.该方法比现有的基于扩散的 SR 方法更有效,因为残差信息可以在几十步内快速传递.此外,本文的设计还允许对证据下界进行分析和简洁的表达,从而简化了对训练优化目标的归纳.在此基础上,本文进一步设计了一种高度灵活的噪声调度方案,该方案控制各步残差的移动速度和噪声强度.通过调整噪声调度超参数,能够有效促进恢复结果的逼真度与现实度的权衡.

综上所述,本文的工作做出了以下贡献:①本文的方法对高维数据的扩展更适用,使用 LR 图像近似分布引导的扩散,以提高扩散模型在高频细节上的集中度;②与标准的扩散模型相比,本文提出的残差扩散模型方法能够有效减少训练计算量,显著降低了推理成本;③有效减少了图像生成时间,模型仅需 15 步就能达到很好的生成质量效果.

## 1 相关工作

### 1.1 遥感超分辨率

遥感超分辨率技术旨在将观测到的低分辨率图像重建成对应的高分辨率图像,有助于提升高级遥感视觉任务的效果,如图像分割、目标识别和目标检索<sup>[9]</sup>.传统的遥感超分辨率方法通常采用插值技术,仅根据图像像素信息进行重建,但无法保证重建图像的质量.近年来,深度学习方法在遥感超分辨率技术中得到广泛应用,主要分为基于卷积网络、基于流、GAN 驱动和面向峰值信噪比(peak signal-to-noise ratio, PSNR)4 种类型<sup>[11]</sup>.

例如,超分辨率卷积神经网络(super-resolution convolutional neural network, SRCNN)<sup>[10]</sup>能够学习图像之间的线性映射,通过简单的卷积网络提取图像片段的特征映射,从而提高了重建质量并实现了快速的推理速度.然而,这种方法通过插值方法增加目标尺寸,无法实现端到端的低分辨率重建.另外,利用卷积层或残差网络等卷积结构直接从 LR 图像中提取特征,极大地提高了重建效率和质量.尽管基于卷积神经网络的高分辨率重建方法取得了一定成果,但由于遥感图像相较于自然图像更加复杂,要求模型对图像细节特征处理能力更高.因此,在遥感领域的超分辨率技术仍面临挑战,需要进一步研究和改进,以更好地适应遥感图像的特殊性.

LUGMAYR 等<sup>[11]</sup>提出的超分辨率流(super-resolution flow, SRFLOW)是一种基于流的方法,旨在学习高分辨率图像的条件分布,以实现具有自然和多样输出的高质量图像超分辨率.该方法使用可逆编码器将高分辨率图像映射到流空间潜在表示,并通过连接编码器、解码器和可逆流模块来避免训练不稳定性.然而,这种方法需要较高的训练成本,并提供较低的感知质量.

另外, XIONG 等<sup>[12]</sup>提出了超分辨率生成对抗网络(super-resolution generative adversarial network, SRGAN),采用感知损失函数生成高质量图像.类似地, WANG 等<sup>[13]</sup>介绍了增强型超分辨率生成对抗网络(enhanced super-resolution generative adversarial network, ESRGAN),结合了增强的超分辨率 GAN 和上级损失函数,以提高感知质量.基于 GAN 的方法将内容损失与对抗损失相结合,使其能够生成清晰的边缘和更丰富的纹理.

### 1.2 扩散模型

近年来,对生成扩散模型的研究引起了广泛的关注.扩散模型是由非平衡态热力学驱动的生成模型,可以分为正向过程和逆向过程.在 2020 年,谷歌<sup>[5]</sup>开发了去噪扩散概率模型,采用 U-Net 结构的自编码器来预

测噪声,并通过独立的分支网络学习高斯分布.OpenAI也提出了一种类别导向的扩散模型—Guided diffusion<sup>[14]</sup>,以加强去噪扩散概率模型的网络结构.该方法通过计算目标类别与分类分数之间的交叉熵损失来确定梯度.并通过大量实验证明扩散模型可以实现比当前最先进的生成模型更高质量的图像样本.

最近,一些研究已经探讨了扩散模型在超分辨率(SR)中的应用.一种流行的方法是在每个步骤中将低分辨率(LR)图像与噪声结合,并重新训练扩散模型.LI等<sup>[15]</sup>提出了扩散概率模型的单图像超分辨率,这是第一个基于扩散的单图像超分辨率(single image super-resolution, SISR)模型,证明了利用扩散模型进行 SISR 任务是可行且有前景的.

现阶段比较流行的方法是利用无条件的预训练扩散模型作为先验,并结合额外的约束来指导反向过程.ROMBACH等<sup>[7]</sup>提出了潜在扩散模型(latent diffusion model, LDM).他们将扩散模型转化为强大而灵活的生成器,可用于一般的条件输入,如文本或边界框,并在卷积方式速率合成中实现高分辨率.然而,由于在训练过程中需要多次通过整个网络进行前向和反向传递,仍然需要高昂的计算成本.

在本文中提出的 ResDM 设计了一个离散的马尔可夫链来描述高分辨率图像和低分辨率图像之间的过渡,不仅提高了迭代训练速度,加快了收敛速度,还针对数据量进行了压缩,节省了大量的计算空间.

## 2 残差偏移的扩散模型

本文提出了一种适宜于遥感超分辨率的扩散模型 ResDM.为了降低训练扩散模型以合成高分辨率图像所需的计算成本,采用了 LR 图像与 HR 图像之间的残差训练方法.该模型允许通过对相应损失项进行欠采样来忽略在感知上不相关的细节.在扩散模型内部,采用了 U-resNet 网络进行参数训练,并引入了双重注意力机制,以连接编码器和解码器中的特征信息,从而保留更多的图像高频细节信息.

为了便于描述,LR 图像和 HR 图像分别用  $y$  和  $x$  表示.此外,模型训练时预设  $y$  和  $x$  具有相同的空间分辨率,通过在必要时使用最近邻插值对 LR 图像  $y$  进行预上采样,以便实现相同空间分辨率的预处理操作,预处理后的  $y$  和  $x$  将分别表示为  $y_0$  和  $x_0$ .图 1 展示了本文提出的模型架构图.

扩散模型的迭代生成范式已被证明在捕获复杂分布方面非常有效,这启发了本文以迭代方式处理超分辨率(SR)问题.本文提出了一个包含更短马尔可夫链的有效扩散模型,用于在高分辨率(HR)图像和相应的低分辨率(LR)图像之间进行转换.马尔可夫链的初始状态收敛于 HR 图像的近似分布,而最终状态收敛于 LR 图像的近似分布.接下来将详细介绍为实现超分辨率而构建这样一个马尔可夫链的过程.

前向过程:本文的残差偏移借鉴了 ResShift<sup>[16]</sup>的想法,将 LR 和 HR 图像之间的残差表示为  $e_0 = y_0 - x_0$ .通过一个长度为  $T$  的马尔可夫链逐渐移动它们的残差  $e_0$  来从  $x_0$  过渡到  $y_0$ .首先引入了一个移动序列  $\{\eta_t\}_{t=1}^T$ ,它随时间步长  $T$  单调增加,满足  $\eta_1 \rightarrow 0, \eta_T \rightarrow 1$ .其残差概率分布表示为:

$$q(x_t | x_{t-1}, y_0) = \mathcal{N}(x_t; x_{t-1} + \alpha_t e_0, \kappa^2 \alpha_t I), t = 1, 2, \dots, T, \quad (1)$$

其中,  $\alpha_t = \eta_t - \eta_{t-1}, t > 1, \alpha_1 = \eta_1$ . ResShift<sup>[16]</sup>证明了在任何时间步长  $t$  的边缘分布都是解析可积的,即

$$q(x_t | x_0, y_0) = \mathcal{N}(x_t; x_0 + \eta_t e_0, \kappa^2 \eta_t I), t = 1, 2, \dots, T, \quad (2)$$

$\kappa$  是一个控制噪声强度的超参数.关于标准差  $\kappa \sqrt{\alpha_t}$ ,目的是  $x_t$  和  $x_{t-1}$  之间过渡平滑,这是因为假设图像数据落在  $[0, 1]$  的范围内,  $x_t$  和  $x_{t-1}$  之间的预期距离可以由  $\sqrt{\alpha_t}$  来限制.

$$\max[(x_0 + \eta_t e_0) - (x_0 + \eta_{t-1} e_0)] = \max[\alpha_t e_0] < \alpha_t < \sqrt{\alpha_t},$$

其中,  $\max[\cdot]$  表示像素最大化操作.关于均值参数,即  $x_0 + \alpha_t e_0$ ,推导出方程(2)中的边缘分布,并且  $x_1$  和

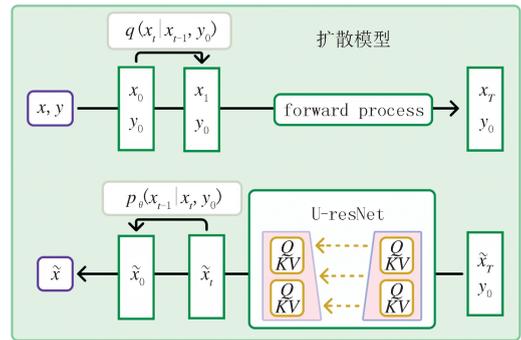


图1 ResDM的模型框架

Fig. 1 ResDM model architecture

$x_T$  边缘分布分别收敛为  $\mu_{x_0}(\cdot)$  和  $\mathcal{N}(\cdot; y_0, \kappa^2 I)$ , 分别作为 HR 图像和 LR 图像的两个近似分布. 通过这样一种方式构造马尔可夫链, 可以通过对给定的 LR 图像  $y_0$  进行反采样来处理 SR 任务.

反向过程: 反向过程旨在估计后验分布  $p(x_0 | y_0)$ , 通过以下公式表示,

$$p_\theta(x_0 | y_0) = \int p(x_T | y_0) \prod_{t=1}^T p_\theta(x_{t-1} | x_t, y_0) dx_{1:T}, \quad (3)$$

其中,  $p(x_T | y_0) \approx \mathcal{N}(x_T | y_0, \kappa^2 I)$ ,  $p_\theta(x_{t-1} | x_t, y_0)$  是具有可学习参数  $\theta$  从  $x_t$  到  $x_{t-1}$  逆跃迁核. 本文假设  $p_\theta(x_{t-1} | x_t, y_0) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, y_0, t), \sum_\theta(x_t, y_0, t))$ .  $\theta$  的优化是通过最小化负证据下界来实现的, 即

$$\min_\theta \sum_t D_{\text{KL}}[q(x_{t-1} | x_t, x_0, y_0) \| p_\theta(x_{t-1} | x_t, y_0)], \quad (4)$$

其中,  $D_{\text{KL}}[* \| *]$  表示 Kullback-Leibler (KL) 散度. 其中更多的数学细节可以在文献[5, 17]中找到.

方程(4)结合方程(1)和方程(2), 根据贝叶斯公式获得目标分布  $q(x_{t-1} | x_t, x_0, y_0)$  的后验分布, 显式表示如下:

$$q(x_{t-1} | x_t, x_0, y_0) = \mathcal{N}(x_{t-1} | \frac{\eta_{t-1}}{\eta} x_t + \frac{\alpha_t}{\eta_t} x_0, \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t I), \quad (5)$$

该推导的详细过程参考文献[16]. 考虑到方差参数与  $x_t$  和  $y_0$  无关, 因此设置  $\sum_\theta(x_t, y_0, t) = \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t I$ . 对

于均值参数  $\mu_\theta(x_t, y_0, t) = \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t I$ , 重参数化表示如下:

$$\mu_\theta(x_t, y_0, t) = \frac{\eta_{t-1}}{\eta_t} x_t + \frac{\alpha_t}{\eta_t} f_\theta(x_t, y_0, t), \quad (6)$$

其中,  $f_\theta$  是一个具有参数  $\theta$  的深度神经网络, 旨在预测  $x_0$ . 本文尝试了不同的  $\mu_\theta$  参数化形式, 并发现方程(6)表现出更强的稳定性和性能. 基于方程(6), 将方程(5)中的目标函数简化如下:

$$\min_\theta \sum_t \| f_\theta(x_t, y_0, t) - x_0 \|_2^2. \quad (7)$$

### 3 实验结果

本节首先介绍了使用的数据集、实验细节以及评估指标. 随后展示了扩散模型 ResDM 在生成高分辨率图像方面的效果. 此外, 本文进行了大量的消融实验, 对 ResDM 模型的性能进行了系统性比较. 图 2 为模型训练使用的部分数据集以及模型训练完成后所生成的图像. 最后, 对视觉实验和评价指标进行了综合比较.

#### 3.1 数据集

本研究的实验采用了 NWPU-RESISC45 数据集, 这是一个被广泛应用于遥感图像分类的数据集, 由西北工业大学收集和维持. 该数据集包含来自 45 个不同场景类别 (如城市、农田、森林、沙漠等) 的大量高分辨率遥感图像样本. 每个类别包含 700 张图像, 每张图像的大小为  $256 \times 256$  像素, 具有图像规模大、信息丰富等特点. 在本研究中, 用于训练的分辨率为  $256 \times 256$  的高分辨率图像是从 NWPU 数据集上的沙滩分类中提取的. 具体而言, 训练集包含 600 张图像, 测试集包含 70 张图像, 验证集包含 30 张图像.



(a) 原始高分辨率图像 (HR); (b) 具有相同空间分辨率的低分辨率图像 (LR); (c) 残差低分辨率图像; (d) 模型生成的超分辨率图像 (SR).

图2 实验中部分数据集以及模型生成图像

Fig.2 Partial datasets and model generated images in the experiment

### 3.2 实验细节

本研究的所有实验均在一台 NVIDIA Tesla V100 图形处理器上进行.首先,将大型图像统一裁剪为  $256 \times 256$  分辨率,作为神经网络的输入,并利用数据增强方法,如翻转和镜像,以提高模型的鲁棒性.LR 图像是通过核估计和噪声注入实现的真实世界超分辨率的降解管道合成的.采用 Adam 算法对 ResDM 进行训练,默认设置为 PyTorch-1.13.1,每轮训练使用 8 张图像.在训练过程中,采用参数化方法对扩散模型进行线性初始化,步长为 15.学习速率固定为  $5e-5$ ,并在  $5 \times 10^5$  次迭代中更新权值参数.在网络架构方面,骨干网络采用 attUNet,包括 16 层残差网络、8 层带有注意机制的残差网络、5 层下采样网络和 5 层上采样网络.为了确保实验评价的稳定性,对最终的模型参数进行了 3 次实验,并取平均值以获得公平的性能评估.

### 3.3 评价指标

在本实验中,采用了 4 个指标来评价模型的性能.

Inception 分数<sup>[18]</sup>(inception score, IS  $\uparrow$ ): IS 指标通过评估生成图像的清晰度和多样性来衡量图像质量.它利用生成图像在所有类别中边缘分布的熵来描述样本的多样性,并通过计算它们的 KL 散度来评估图像质量. KL 散度越大,图像质量越好.

Fréchet Inception 距离<sup>[19]</sup>(fréchet inception distance, FID  $\downarrow$ ): FID 是生成模型的常用评估指标之一,用于评估生成样本的质量.它是一种衡量两组图像数据集相似性的方法,通过计算真实图像和生成图像特征向量之间的距离来评估.

峰值信噪比(peak signal-to-noise ratio, PSNR  $\uparrow$ ): PSNR 是衡量图像或视频质量的指标,通常用于比较原始图像和经过压缩或其他处理后的图像之间的差异. PSNR 数值越高,表示图像质量越好.

结构相似性指数(structural similarity index, SSIM  $\uparrow$ ): SSIM 是一种用于衡量图像相似性的指标,考虑了亮度、对比度和结构等方面. SSIM 的取值范围在  $-1$  到  $1$  之间,数值越接近  $1$  表示图像越相似.

### 3.4 比较方法

为了验证该方法在 RS 超分辨率生成任务上的有效性,本文将 ResDM 与其他领先模型进行了定性和定量比较,比较模型如下.

迭代细化的扩散模型(image super-resolution via iterative refinement, SR3):这是一种通过重复细化实现图像超分辨率的方法,将去噪扩散概率模型应用于条件图像生成,并通过随机迭代去噪过程执行超分辨率.在不同放大倍数的超分辨率任务中,SR3 展现出强大的性能.

细节补充的扩散模型(diffusion model with detail complement for super-resolution of remote sensing, DMDC)<sup>[20]</sup>:不同于传统优化模型对图像理解不足的情况,将扩散模型作为生成模型首次引入到遥感超分辨率任务中,并将低分辨率图像作为条件信息来指导图像生成.考虑到生成模型可能无法准确地恢复特定的小对象和复杂的场景,提出了细节补充任务,以提高恢复能力,同时提出了联合像素约束损失和去噪损失来优化逆扩散方向.该方法证明在小而密集目标的遥感超分辨率中具有优越性.

潜在扩散模型(high-resolution image synthesis with latent diffusion models, LDM):通过将图像输入到功能强大的预训练自动编码器的潜在空间中,然后将生成过程分解为一系列去噪自动编码器的顺序应用.这种方法允许引导机制来控制图像生成过程,而无需重新训练模型.与基于像素的方法相比,这种方法显著降低了计算需求.

残差移位的扩散模型(efficient diffusion model for image super-resolution by residual shifting, Res-Shift):通过在高分辨率图像和低分辨率图像之间传输残差,模型显著提高了传输效率.这种方法能够明显减少扩散步骤的数量,从而消除了推理过程中的后处理加速及其相关的性能下降.

### 3.5 扩散步长和超参数

在实验部分,使用方程 1 中提出的转移分布来减少马尔可夫链中的扩散步骤  $T$ .通过调整超参数  $p$  来灵活控制残差移位的速度.本文对相同的超参数进行控制,并比较了  $10 \sim 50$  步长的图像生成质量.同时,本文还比较了相同步长下不同噪声强度的超分辨率图像生成质量.具体结果如表 1 所示.

### 3.6 综合数据评价

根据 NWPU-RESISC45 数据集上的实验结果对本文提出的方法与最先进方法进行比较分析,如表 2 所

示,可以得出以下观点:本文提出的方法 ResDM 在 PSNR 指标上表现最佳,表现出较优的图像重建质量.在保持结构相似性方面,SSIM 仅略低于 ResShift 方法,表现出色.这一优势主要源于本文设计精良的扩散模型,该模型从 LR 图像的微小扰动开始,而非传统假设的白色高斯噪声.相比其他方法,FID 分数最低,表明生成的图像更接近真实图像;IS 得分最高,显示出生成的图像多样化且分类自信.从实验结果来看,本文的 ResDM 在大多数指标上表现出色,PSNR、FID 和 IS 3 个指标表现更好,表明其提供了高质量的图像重建,具有良好的多样性和分类置信度.

表 1 噪声强度和步长对图像质量的影响

Tab. 1 The effect of noise intensity and step size on image quality

步长	噪声强度	PSNR ↑	SSIM ↑	步长	噪声强度	PSNR ↑	SSIM ↑
15	0.3	26.31	0.679 1	10	0.3	26.29	0.679 7
15	1.0	26.23	0.677 3	15	0.3	26.31	0.679 1
15	1.5	26.17	0.676 2	30	0.3	26.05	0.672 5
15	2.0	26.14	0.673 5	50	0.3	25.98	0.668 9

表 2 不同方法在 NWPU 验证数据集上的定量结果

Tab. 2 Quantitative results of different methods on NWPU validation datasets

方法	步长	PSNR ↑	SSIM ↑	FID ↓	IS ↑	方法	步长	PSNR ↑	SSIM ↑	FID ↓	IS ↑
SR3	2 000	23.04	0.630	60.12	79.80	ResShift	15	24.75	0.680	57.56	88.00
DMDC	2 000	24.15	0.625	69.28	86.30	本文(ResDM)	15	26.31	0.679	57.21	88.80
LDM	2 000	25.01	0.610	57.32	88.50	本文(ResDM)	1 000	25.16	0.623	59.39	85.32

### 3.7 生成效率

表 3 对比了本文方法与其他扩散模型在不同生成步长下的运行时间性能.结果显示,在测试的所有方法相同生成步长下,本文方法均表现出比其他扩散模型更快的运行速度.随着步数的增加,ResDM 相对于其他 DM 的效率改善变得更加显著.这一改善主要归功于高分辨率图像和低分辨率图像之间传输残差缩短了模型的迭代次数,减少了大量的计算规模.需要注意的是,由于传统扩散模型受到像素级别噪声推理的限制,本文方法在 15 步长就能达到传统扩散模型 100 步长甚至更多步长的效果.因此,可以推断出 ResDM 在生成效率方面具有显著优势,特别是在图像生成过程更复杂且步数更多时.因此,在对速率要求高且资源有限的应用中,本文的 ResDM 是更为理想的选择.

表 3 ResDM 与 DDIM、LDM 在数据集上的效率和性能比较

Tab. 3 Comparison of efficiency and performance of ResDM with DDIM and LDM on datasets

步长	方法	运行时间/s	方法	运行时间/s	方法	运行时间/s
15	ResDM	6.23	LDM	7.64	DDIM	8.53
30	ResDM	10.97	LDM	12.39	DDIM	14.32
100	ResDM	27.75	LDM	31.72	DDIM	36.49
200	ResDM	56.65	LDM	63.19	DDIM	72.10
500	ResDM	139.55	LDM	157.39	DDIM	173.63
1 000	ResDM	407.35	LDM	452.27	DDIM	511.97

### 3.8 训练效率

在本文的研究中,对比了不同扩散模型方法在每轮训练的平均用时,如表 4 所示.每个方法均采用 3 次训练迭代后的平均用时,其中步长采用原论文中推荐步长.在迭代训练过程中,由于内存限制,ResDM 模型在迭代训练时,每轮训练能够有效减少训练时间和计算资源的消耗.由于受到内存限制,每轮的 8 张图像无法完全加载,因此每轮分为 4 个小批次,每次加载 2 张图像,结果可以看到 ResDM 模型能够显著减少训练时间.这一结果验证了本文提出的方法在扩散模型训练中的有效性.

表 4 ResDM 在训练时每轮训练用时情况与其他方法对比

Tab. 4 Comparison of ResDM training time per round with other methods during training

方法	DDPM	SR3	DDIM	LDM	ResDM
步长	2 000	2 000	1 000	2 000	15
时间/s	19 809.56	15 014.54	10 549.71	5 692.65	1 379.24

### 3.9 实验总结

由于扩散模型图像生成受到噪声概率影响,无法确保生成的图像达到统一的图像质量,因此在图像生成效率和模型训练效率测试中生成的图像质量略有差异.同样,无论是图像生成质量还是生成效率,也无法完全统一模型训练效率.这是因为本文提出的包含更短马尔可夫链的残差扩散模型,马尔可夫链的初始状态收敛于 HR 图像的近似分布,而最终状态收敛于 LR 图像的近似分布,这导致模型随着扩散步长增大会出现模型训练过拟合现象,模型会过多地记住训练数据的细节,而无法良好地泛化到测试数据,出现在训练数据上表现很好,但在测试数据集上表现不佳的现象.表 1 中数据证明了 ResDM 模型随着步长的增加生成图像整体质量反而有所下降,表 2 中 ResDM 模型步长为 1 000 时的图像质量数据横向对比也如预期一样,不如 ResDM 模型步长为 15 的图像质量,因此也可以得出结论,在相似生成效率情况下 ResDM 与其他扩散模型图像质量各有优劣,但是在 ResDM 方法理想条件下生成的图像具有良好的多样性和分类置信度.

总的来说,基于以上实验结果,本文提出的 ResDM 方法有以下优势:①有效改善了扩散模型在遥感图像生成领域难以适配图片、训练周期长、生成图像慢的不足.本文提出的方法能有效减少训练计算量.②在优化大量训练时间和内存消耗后,依然能够提供高质量的遥感超分辨图像、提高图像生成速度的同时还具有良好的多样性和分类置信度.

## 4 结 论

本文提出了一种高效的 SR 扩散模型 ResDM.与现有基于扩散的 SR 方法需要大量迭代才能获得满意结果不同,本文的方法通过在高分辨率图像和低分辨率图像之间移动残差来实现图像之间的转移,能够有效减少训练时间和计算资源的消耗.在不降低生成图像质量的前提下,ResDM 显著提高了模型的训练和采样效率.本文方法的核心思想是将扩散模型的高分辨率图像破坏为低分辨率图像,而不是高斯白噪声,从而有效地缩短扩散模型的长度.

实验证明了本文提出的方法在合成数据集和真实数据集上的优越性.然而,由于遥感图像类型多样,研究发现在进行超分辨图像生成时,针对单一类遥感图像能够取得较好的结果,但训练多类型图像效果还有待改进,目前来说,虽然扩散模型与其他图像生成模型相比在生成多类型图像方面表现更好,并且已有研究人员提出改进的训练技术,可以用扩散模型生成多类型的自然图像,但扩散模型在针对遥感多类型图像生成领域还没有成熟的相关解决方法,在未来也是一个很好的研究方向.因此,后续研究工作将进一步探索对生成遥感图像场景的控制,以在更加复杂的情况下更好地渲染更大的图像.本文的研究工作为发展更有效的扩散模型以解决遥感 SR 问题提供了有价值的理论支撑.

## 参 考 文 献

- [1] FERNANDEZ-BELTRAN R, LATORRE-CARMONA P, PLA F. Single-frame super-resolution in remote sensing: a practical overview [J]. International Journal of Remote Sensing, 2017, 38(1): 314-354.
- [2] 毛盼娣,徐道连. 高效单图像超分辨率重建: 深监督对称蒸馏网络[J]. 河南师范大学学报(自然科学版), 2023, 51(6): 57-65.  
MAO P D, XU D L. Efficient single-image super-resolution: deeply-supervised symmetric distillation network [J]. Journal of Henan Normal University (Natural Science Edition), 2023, 51(6): 57-65.
- [3] WANG Y, BASHIR S M A, KHAN M, et al. Remote sensing image super-resolution and object detection: benchmark and state of the art [J]. Expert Systems with Applications, 2022, 197: 116793.
- [4] SAHARIA C, HO J, CHAN W, et al. Image super-resolution via iterative refinement [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4713-4726.
- [5] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [J]. Advances in Neural Information Processing Systems, 2020, 33:

6840-6851.

- [6] DHARIWAL P, NICHOL A. Diffusion models beat GANs on image synthesis[EB/OL]. [2024-02-15]. <https://arxiv.org/abs/2105.05233v4>.
- [7] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans; IEEE Press, 2022.
- [8] GABBUR P. Improved DDIM sampling with moment matching Gaussian mixtures[EB/OL]. [2024-02-13]. <https://arxiv.org/abs/2311.04938v2>.
- [9] YUAN Z Q, ZHANG W K, TIAN C Y, et al. Remote sensing cross-modal text-image retrieval based on global and local information[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5620616.
- [10] DONG C, LOY C C, HE K M, et al. Learning a deep convolutional network for image super-resolution[M]. Cham: Springer International Publishing, 2014: 184-199.
- [11] LUGMAYR A, DANELLJAN M, VAN GOOL L, et al. SRFflow: learning the super-resolution space with normalizing flow[M]. Cham: Springer International Publishing, 2020: 715-732.
- [12] XIONG Y F, GUO S X, CHEN J S, et al. Improved SRGAN for remote sensing image super-resolution across locations and sensors[J]. Remote Sensing, 2020, 12(8): 1263.
- [13] WANG X T, YU K, WU S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[M]. Cham: Springer International Publishing, 2019: 63-79.
- [14] NICHOL A Q, DHARIWAL P. Improved denoising diffusion probabilistic models [C]//International conference on machine learning. [S.l.]: PMLR, 2021.
- [15] LI H Y, YANG Y F, CHANG M, et al. SRDiff: Single image super-resolution with diffusion probabilistic models[J]. Neurocomputing, 2022, 479: 47-59.
- [16] YUE Z S, WANG J Y, LOY C C. Resshift: Efficient diffusion model for image super-resolution by residual shifting[EB/OL]. [2024-02-16]. <https://arxiv.org/pdf/2307.12348.pdf>.
- [17] SOHL-DICKSTEIN J, WEISS E, MAHESWARANATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics [C]//International conference on machine learning. [S.l.]: PMLR, 2015.
- [18] BARRATT S, SHARMA R. A note on the inception score[EB/OL]. [2024-02-16]. <https://arxiv.org/pdf/1801.01973>.
- [19] YU Y, ZHANG W, DENG Y. Frechet inception distance (fid) for evaluating gans, 2021[EB/OL]. [2024-02-16]. <https://www.researchgate.net/publication/354269184>.
- [20] LIU J, YUAN Z Q, PAN Z Y, et al. Diffusion model with detail complement for super-resolution of remote sensing[J]. Remote Sensing, 2022, 14(19): 4834.

## Remote sensing super-resolution image generation based on residual diffusion model

Zuo Xianyu<sup>a,b</sup>, Tian Zhanshuo<sup>a,b</sup>, Yin Menghan<sup>a</sup>, Dang Lanxue<sup>a,b</sup>, Qiao Baojun<sup>a,b</sup>, Liu Yang<sup>a,b</sup>, Xie Yi<sup>a,b</sup>

(a. School of Computer and Information Engineering; b. Henan Provincial Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng 475004, China)

**Abstract:** Traditional diffusion-based image super-resolution methods usually require a large number of sampling steps and a lot of computational time to optimize powerful diffusion models. In order to realize training on limited computational resources, existing accelerated sampling techniques often sacrifice image quality and lead to ambiguous super-resolution results. To address this problem, an improved and efficient residual super-resolution diffusion model is proposed in this paper. The transfer between images is achieved by constructing a Markov chain that shifts the residuals between high-resolution images and low-resolution images, significantly reducing the number of diffusion steps. The method ensures the quality and flexibility of the super-resolution results, while improving the transfer efficiency and eliminating the post-acceleration and its associated image detail feature degradation required in the inference process. It is experimentally demonstrated that even with only 15 sampling steps, the method in this paper achieves image quality better than or at least comparable to current state-of-the-art methods on both synthetic and real datasets.

**Keywords:** remote sensing; super-resolution; image generation; residual shifting; residual diffusion model