

基于扩散先验的脑部MRI超分辨率重建

熊承义^{ab}, 曹雨轩^{ab*}, 高志荣^c

(中南民族大学 a. 电子信息工程学院; b. 智能无线通信湖北省重点实验室; c. 计算机学院, 武汉 430074)

摘要 现有基于Transformer的MRI超分辨率方法虽具有良好的全局建模能力,但忽略了深度先验约束建模的重要性. 为此,提出了一种基于扩散先验的脑部MRI超分辨率方法,利用潜在扩散模型生成的先验来引导Transformer进行超分辨率重建,以提升MRI细节重建能力. 具体而言,采用两阶段协同训练策略:第一阶段通过真实图像潜编码构建内容先验;第二阶段引入扩散模型重构先验,并联合优化去噪与重建过程,实现无监督条件下的图像超分辨率. 此外,采用深度可分离卷积与置换自注意力机制,实现编码器的高效建模与感受野扩展. 在IXI多模态MRI数据集上的4倍超分辨率实验表明:所提出方法在提升重建图像主观质量与重建效率方面优于已有方法.

关键词 MRI超分辨率; 扩散先验; 置换自注意力; 深度可分离卷积

中图分类号 TP391.4 文献标志码 A 文章编号 1672-4321(2026)02-0202-10

doi:10.20056/j.cnki.ZNMDZK.20250840

Brain MRI super-resolution reconstruction based on diffusion priors

XIONG Chengyi^{ab}, CAO Yuxuan^{ab*}, GAO Zhirong^c

(South-Central Minzu University, a. College of Electronic and Information Engineering; b. Hubei Key Lab of Intelligent Wireless Communication; c. College of Computer Science, Wuhan 430074, China)

Abstract Transformer-based MRI super-resolution methods offer strong global modeling capabilities but often overlook the role of deep prior constraints. To address this problem, a brain MRI super-resolution method based on diffusion priors is proposed, where a latent diffusion model generates structural priors to guide the Transformer in restoring fine details. A two-stage training strategy is adopted: the first stage constructs a content prior from ground-truth latent encodings to pretrain the reconstruction network; the second stage introduces diffusion-based priors and jointly optimizes the denoising and reconstruction processes under unsupervised conditions. Additionally, depthwise separable convolutions and permuted self-attention are employed to enhance modeling efficiency and expand the receptive field. Experiments on the IXI multi-modal MRI dataset (4×SR) demonstrate superior reconstruction quality and efficiency of the method over existing methods.

Keywords MRI super-resolution; diffusion prior; permuted self-attention; depthwise separable convolution

磁共振成像(Magnetic Resonance Imaging, MRI)是一种广泛应用、无创且无辐射的成像技术,能够提供清晰的软组织结构和纹理信息,因此在临床诊断和医学研究中具有不可替代的重要性. 然而在实际应用中,获取高分辨率(High-Resolution, HR)MR图像通常需要较长的扫描时间,这不仅增加了患者的不适感,还容易产生运动伪影,导致图像质量下降. 分辨率与采集效率之间的这种权衡在临床实践

中构成了一项重大挑战. 为此,磁共振图像超分辨率(Super-Resolution, SR)技术^[1-2]应运而生,其目标是通过算法从低分辨率(Low-Resolution, LR)MR图像中恢复出高分辨率图像,从而在保持较短扫描时间的同时提升图像质量^[3].

早期的超分辨率方法主要基于插值算法(如双三次插值)或稀疏编码理论,但这些传统方法在细节恢复能力和抗噪性能方面存在明显局限,而深度

收稿日期 2025-04-08 * 通信作者 曹雨轩,研究方向:图像处理, E-mail: yuxuancao2023@163.com

作者简介 熊承义(1969-),男,教授,研究方向:图像处理与模式识别, E-mail: xiongcy@mail.scuec.edu.cn

基金项目 多谱信息处理技术国家重点实验室基金资助项目(6142113210303);中央高校基本科研业务费专项资金资助项目(CZZ25010)

学习方法已在该领域取得了显著进展.其中卷积神经网络(CNN)因其擅长捕捉局部空间特征,在MRI超分辨率重建任务的早期研究中被广泛应用.传统CNN模型如SRCNN^[4],通过堆叠卷积层直接建立低分辨率图像到高分辨率图像的映射关系.在此基础上,QIU等^[5]针对膝关节MRI设计了基于SRCNN的超分辨率网络.基于CNN的SR方法主要依赖局部卷积操作,难以建模图像中的长程依赖关系,导致其在复杂纹理重建和全局一致性方面存在局限.而Transformer凭借自注意力机制^[6]在自然语言处理领域的成功,在视觉任务中也展现出强大的全局建模能力.视觉Transformer(ViT)^[7]的提出首次将Transformer应用于图像处理任务,但由于其计算复杂度较高,难以直接应用于高分辨率图像处理.为此,Swin transformer^[8]引入了分层结构和移动窗口机制,在保证全局建模能力的同时大幅降低计算成本.近年来,Transformer逐渐被引入MRI超分辨率任务中^[9-10],例如,FORIGUA等^[11]提出的SuperFormer结合局部自注意力和3D位置编码,以融合多领域信息;TransMRSR^[12]则采用卷积与Transformer块结合的策略,分阶段提取局部特征并捕获全局依赖.尽管这些方法利用了Transformer强大的特征建模能力,但其计算复杂度随窗口尺寸的增大呈平方级增长.为降低计算成本,许多方法选择限制窗口大小,如SwinIR^[13]采用8×8窗口.然而,窗口过小会限制感受野,削弱全局信息建模能力.因此,如何在保持Transformer全局建模能力的同时提高计算效率并增强医学影像的真实细节恢复,仍是一个亟待解决的问题.

与此同时,扩散模型(Diffusion Models, DMs)^[14]作为新一代生成式模型,与依赖对抗训练且易陷入模式崩溃的GAN不同,其采用一种概率生成框架,通过迭代去噪逐步将随机噪声转化为有意义的数据表示,从而实现数据分布的重构.这一过程不仅能够更全面地覆盖数据模式,还能提供更高的图像保真度,使得扩散模型成为高质量图像重建任务的理想选择.例如,CHUNG等^[15]基于反向扩散模型设计了一种MRI去噪和超分辨率联合方法,WU等^[16]则采用去噪扩散概率模型(DDPM)实现脑部MRI超分辨率重建.然而,传统扩散模型仍然面临着计算开销高、推理速度慢的问题.由于扩散模型依赖多次迭代去噪,每次推理通常需要数百步甚至上千步的计算,使其在实际应用中受限.在MRI超分辨率任务中,直接从噪声生成完整图像的方式存在一定

冗余,因为目标是增强已有的低分辨率图像,而非无条件从头生成新图像.因此,目前的研究主要聚焦在加速扩散推理,如通过改进采样策略减少推理步骤,或者采用潜在扩散模型(Latent Diffusion Model, LDM)^[17]在低维潜在空间中进行建模,以降低计算开销并提高推理效率.图像恢复模型DiffIR^[18]和图像去噪模型Hi-Diff^[19]将Transformer集成到潜在扩散模型的框架中,利用其作为解码器在扩散先验约束下重建出高质量图像.

针对上述问题和背景,本文基于扩散先验引导的Transformer架构,提出了一种用于脑部MRI超分辨率的高效扩散模型EDM-MSR(Efficient Diffusion Model for MRI Super-Resolution).通过轻量化的深度可分离卷积重新优化编码器潜在先验提取网络(Latent Prior Extraction Network, LPEN)和条件提取网络(Condition Extraction Network, CEN),用于提取HR和LR图像中的紧凑的潜在特征作为先验表示;受到SRFormer^[20]置换自注意力机制的启发,本文将其嵌入Transformer中设计先验引导动态Transformer(Prior-Guided Dynamic Transformer, PGDformer).PGDformer包括先验引导动态多头置换自注意力模块(Prior-Guided Dynamic Multi-Head Permuted Self-Attention, PDM-PSA)和先验引导动态前馈网络(Prior-Guided Dynamic Feed-Forward Network, PGD-FFN)两个部分,在保持计算效率的同时有效扩大感受野,并根据先验信息自适应地重组特征表示,从而增强模型对医学图像复杂结构的建模能力和重建质量.

1 提出的超分辨率重建方法EDM-MSR

本文提出的基于扩散先验的脑部MRI超分辨率重建方法,主要利用潜在扩散模型生成的先验来引导Transformer进行超分辨率重建.下文先给出模型的整体结构及算法流程,然后对相应训练阶段和推理阶段进行详细描述.

1.1 模型结构

本文提出的模型EDM-MSR的整体结构如图1所示,分为两个训练阶段.为了在无真实HR图像的情况下实现高质量超分辨率重建,本文采用两阶段协同训练策略,核心在于利用潜在扩散模型生成高质量先验表示以引导PGDformer.第一阶段通过有监督方式预训练PGDformer.首先将真实的HR图像和

LR 图像在通道维度拼接,并通过像素反混洗操作转换为空间分辨率较低、通道维度更高的结构化特征表示;随后利用 LPEN 提取融合了细节与全局信息的紧凑先验表示 Z ,引导 PGDformer 完成端到端重建.该阶段使 PGDformer 能够充分利用 Z 进行高质量重建,为后续无监督优化奠定基础.第二阶段针

对无真实 HR 图像的场景展开无监督建模,以 LPEN 生成的初始先验为起点,利用条件提取网络获得条件编码 C ,引导去噪网络迭代优化噪声估计;最终将逆扩散生成的精细化先验 \hat{Z} 输入 PGDformer,实现重建与去噪过程的联合优化,从而进一步提升图像重建质量并减少伪影.

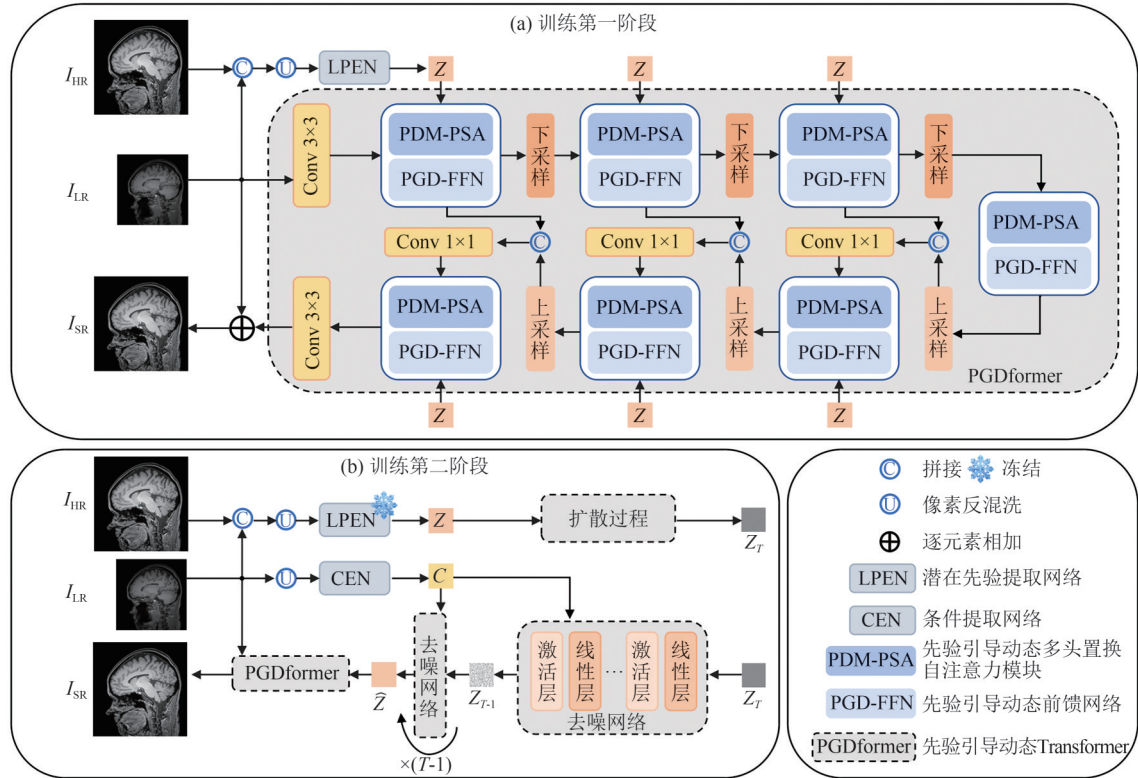


图1 EDM-MSR 整体结构图

Fig. 1 Overall structure of EDM-MSR

1.2 训练第一阶段

第一阶段(图 1(a))主要是联合训练潜在先验提取网络(LPEN)和先验引导动态 Transformer(PGDformer),从而得到更可靠的先验表示 Z .LPEN 的网络结构如图 2(a)所示,PGDformer 主要由两个模块组成:先验引导动态多头置换自注意力模块(PDM-PSA)和先验引导动态前馈网络(PGD-FFN).PDM-PSA 和 PGD-FFN 的网络结构分别如图 2(b)和图 2(c)所示.

1.2.1 潜在先验提取网络(LPEN)

潜在先验网络由 9 个残差块和 2 个线性层堆叠而成,其中每个残差块由“深度可分离卷积——激活——深度可分离卷积”组成,在保证潜在先验提取质量的同时有效降低计算成本.将输入的 I_{HR} 和 I_{LR} 在通道维度上进行拼接,经过像素反混洗操作后输出给 LPEN 来提取紧凑的潜在先验,从而得到先验表示 Z 如下:

$$Z = \text{LPEN}\left(\text{PixelUnshuffle}\left(\text{Concat}\left(I_{HR}, I_{LR}\right)\right)\right), \quad (1)$$

在后续超分辨率重建过程中, Z 作为动态调制参数来引导 PGDformer 进行重建.

1.2.2 先验引导动态 Transformer(PGDformer)

在基于 Transformer 架构的超分辨率模型中,大多数方法为控制计算成本将计算自注意力的窗口尺寸限制在 8×8 .增大多头自注意力(MSA)的窗口尺寸可有效提高超分辨率重建质量,但同时也会导致更大的计算负担.SRFormer 提出一种用于单图超分辨率任务的注意力机制——置换自注意力,可以在大窗口中高效计算自注意力,从而在不增加参数量和计算成本的前提下提升超分辨率性能.本文在 PGDformer 的设计过程中也应用到了这种思想.PGDformer 包括两个部分: PDM-PSA 和 PGD-FFN.

当获得 Z 后,它作为动态调节参数输入给 PDM-PSA 和 PGD-FFN 来引导重建过程,具体可表示为:

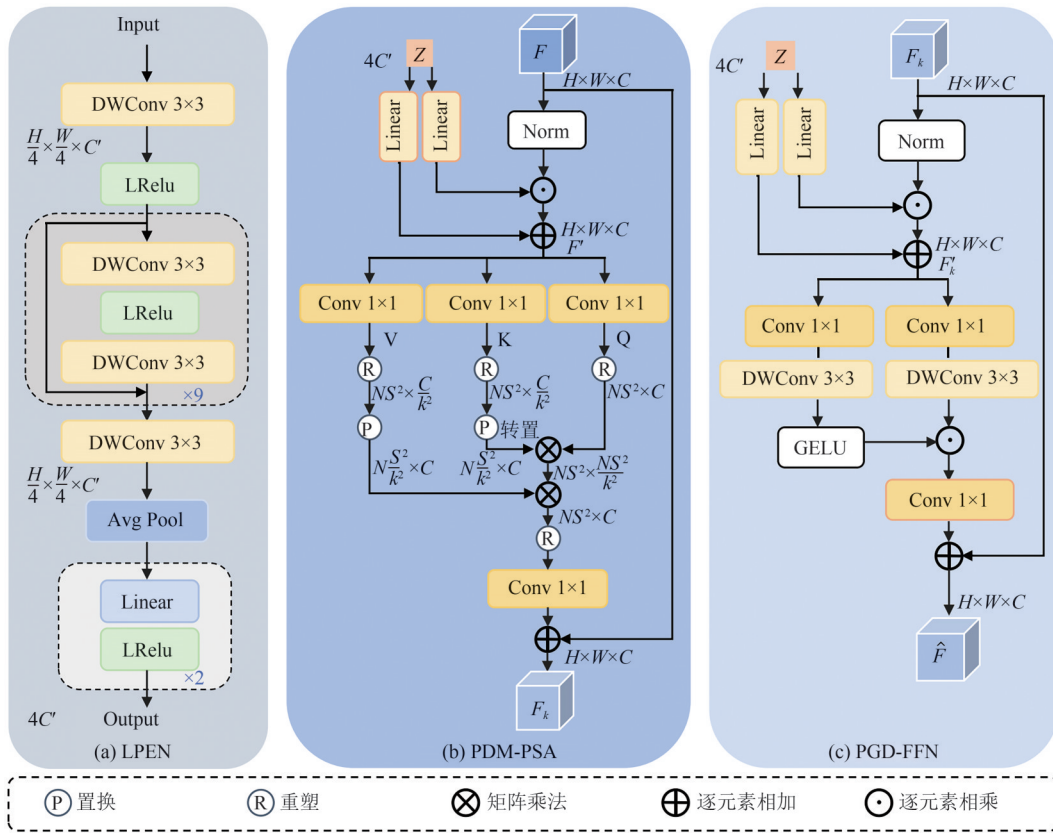


图2 LPEN、PDM-PSA 和 PGD-FFN 结构图

Fig. 2 Structure of LPEN, PDM-PSA and PGD-FFN

$$F' = \mathbb{L}(Z) \odot \text{Norm}(F) + \mathbb{L}(Z), \quad (2)$$

$$F'_k = \mathbb{L}(Z) \odot \text{Norm}(F_k) + \mathbb{L}(Z), \quad (3)$$

式中, F' 和 F'_k 均为中间输出特征图, $F' \in \mathbb{R}^{H \times W \times C}$, $F'_k \in \mathbb{R}^{H \times W \times C}$, C 表示特征图的通道维度, \mathbb{L} 表示线性层, \odot 表示逐元素相乘, $\text{Norm}(\cdot)$ 表示层归一化。

具体来说,在 **PDM-PSA** 中,输入特征图 F' 将被分割成 N 个不重叠的平方窗口 X , 每个窗口的维度 $X \in \mathbb{R}^{NS^2 \times C}$, 其中 S 为每个窗口的边长, 表示窗口的尺寸, N 是窗口的数量, C 是通道数. 接下来, 使用线性层 \mathbb{L}_Q , \mathbb{L}_K 和 \mathbb{L}_V , 通过线性变换将每个窗口 X 嵌入得到 $Q \in \mathbb{R}^{NS^2 \times C}$, $K \in \mathbb{R}^{NS^2 \times C/k^2}$ 以及 $V \in \mathbb{R}^{NS^2 \times C/k^2}$. 其中, Q 保持和窗口 X 一样的通道维度, 而 K 和 V 的通道维度被压缩至 C/k^2 (k 为 token 的缩减因子, 本文中设置为 2). 为了让更多的 token 参与到自注意力的计算并且避免增加计算开销, 本文将 K 和 V 的空间信息转换到通道维度, 得到置换后的 $K_p \in \mathbb{R}^{NS^2/k^2 \times C}$, $V_p \in \mathbb{R}^{NS^2/k^2 \times C}$, 在这种方式下, K 和 V 的窗口尺寸将会减少至 $S/k \times S/k$, 但是通道维度没有改变, 从而可以保证每个注意力头生成的注意力图的表现力. 最后, 置换自注意力的计算可表示为:

$$\text{PSA}(Q, K_p, V_p) = \text{Softmax}\left(\frac{QK_p^\top}{\sqrt{d_k}} + B\right)V_p, \quad (4)$$

式中, B 为对齐的相对位置嵌入矩阵, K_p^\top 为 K_p 的转置矩阵, $\sqrt{d_k}$ 为标准化因子. **PDM-PSA** 的整个过程可以定义为:

$$F_k = \Phi_c \text{PSA}(Q, K_p, V_p) + F, \quad (5)$$

在 **PGD-FFN** 中, 首先使用 1×1 点卷积来聚合来自不同通道的信息, 然后使用 3×3 深度可分离卷积来聚合来自空间相邻像素的信息, 同时使用门控机制来增强信息编码. **PGD-FFN** 的整个过程可以定义为:

$$\hat{F} = \Phi_c \left(\text{GELU}(\Phi_c \Phi_d(F'_k)) \odot (\Phi_c \Phi_d(F'_k)) \right) + F_k, \quad (6)$$

式(5)和式(6)中, Φ_c 为 1×1 点卷积, Φ_d 为 3×3 深度可分离卷积.

1.2.3 损失函数

在第一阶段的训练中, 为了获得更可靠的先验表示, 本文对 LPEN 和 PGDformer 同时进行训练优化. 由于在超分辨率任务中, 基于 L_1 范数的损失函数有助于保留图像的锐度并且降低平滑的现象, 故在本文中, 重建图像像素级的损失使用基于 L_1 范数

的损失函数:

$$\mathcal{L}_{\text{rec}} = \|I_{\text{SR}} - I_{\text{HR}}\|_1, \quad (7)$$

式中, $\|\cdot\|_1$ 为 L_1 范数, I_{HR} 为真实的高分辨率图像, I_{SR} 为重建后的超分辨率图像.

1.3 训练第二阶段

第二阶段(图 1(b))主要是联合训练扩散模型(DM)、条件提取网络(CEN)以及 PGDformer, 旨在通过训练具备强大数据映射能力的扩散模型来学习真实 MR 图像的先验知识分布, 进而来引导并进一步增强 PGDformer 的重建过程. 扩散模型包括前向扩散和反向去噪两个过程.

1.3.1 前向扩散过程

首先利用第一阶段训练好的 LPEN 来获得先验表示 Z , 然后利用其在前向扩散过程中采样 $Z_T \in \mathbb{R}^{4C'}$, 具体表示为:

$$q(Z_T|Z) = \mathcal{N}\left(Z_T; \sqrt{\bar{\alpha}_T} Z, (1 - \bar{\alpha}_T)I\right), \quad (8)$$

式中, T 为总迭代次数, \mathcal{N} 表示高斯分布, $\alpha = 1 - \beta$, $\bar{\alpha}_t = \prod_{i=0}^t \alpha_i$, 其中 $t = (1, \dots, T)$, $\beta_{1:T} \in (0, 1)$ 为控制噪声方差的超参数.

1.3.2 反向去噪过程

反向去噪过程可以定义为从 Z_T 到 \hat{Z} 向后运行的马尔可夫链. 由于 Z 是一种紧凑的先验表示, 故在去噪过程中只需很少的迭代次数就可获得良好的估计. 从 Z_t 到 Z_{t-1} 的概率分布可定义为:

$$q(Z_{t-1}|Z_t, Z_0) = \mathcal{N}\left(Z_{t-1}; \mu_t(Z_t, Z_0), \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t I\right), \quad (9)$$

$$\mu_t(Z_t, Z_0) = \frac{1}{\sqrt{\alpha_t}} \left(Z_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon \right) \quad (10)$$

式中, ϵ 表示 Z_t 中的噪声, 在训练中利用去噪网络 e_θ 来预测每一步的噪声 ϵ . 受潜在扩散模型(LDM)的影响, 本文利用条件提取网络(CEN)从 LR 图像中获得紧凑的条件特征, 用于引导去噪网络进行去噪, 具体表示为:

$$C = \text{CEN}(\text{PixelUnshuffle}(I_{\text{LR}})), \quad (11)$$

式中, CEN 是与 CPEN 具有相同结构的网络. 因此, 整个去噪网络 e_θ 将会利用 Z_t 和 C 来预测噪声, 具体表示为:

$$Z_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(y_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} e_\theta(Z_t, C, t) \right) + \sqrt{1 - \alpha_t} \epsilon_t, \quad (12)$$

式中, $\epsilon_t \sim \mathcal{N}(0, I)$. 在采样 T 次后, DM 可以生成与原始 HR 图像分布一致的先验表示 \hat{Z} , 进而利用 \hat{Z} 来指

导 PGDformer 进行超分辨率重建, 进一步优化其重建性能.

1.3.3 损失函数

在第二阶段的训练中, 通过总损失函数 \mathcal{L} 来联合训练 DM、CEN 和 PGDformer, 具体表示为:

$$\mathcal{L} = \mathcal{L}_{\text{diff}} + \mathcal{L}_{\text{rec}}, \quad (13)$$

$$\mathcal{L}_{\text{diff}} = \frac{1}{4C'} \sum_{i=1}^{4C'} |\hat{Z}(i) - Z(i)|, \quad (14)$$

式中, $\mathcal{L}_{\text{diff}}$ 为扩散模型的损失, \mathcal{L}_{rec} 为重建图像像素级的损失.

1.4 推理阶段

在推理阶段, 本文仅使用反向去噪过程, 具体流程如图 3 所示. 首先利用 CEN 从 I_{LR} 中获得潜在的条件特征 C , 然后随机采样高斯噪声 Z_T , 去噪网络利用 Z_T 和 C , 在 T 次迭代后估计出 \hat{Z} , 进而引导 PGDformer 重建出最终的 SR 图像 I_{SR} .

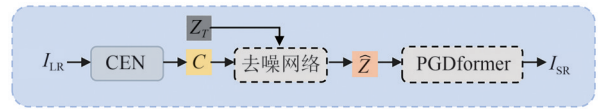


图 3 推理阶段过程

Fig. 3 The process of inference

2 实验

2.1 实验设置

2.1.1 数据集

本实验采用公共脑部 MRI 数据集 IXI (详见 <http://brain-development.org/ixi-dataset/>), 该数据集包括 581 个 3D T1 加权体积、578 个 3D T2 加权体积以及 578 个 3D 质子密度(PD)加权体积图像. 本实验分别从 T1、T2 和 PD 三种模态的数据集中各选取前 500 个 3D 体积, 并按照 7:1:2 的比例划分为训练集、验证集和测试集; 随后, 从每个 3D 图像中提取 50 张分辨率为 256×256 的高质量 2D 切片; 最终, 每种模态的数据集均包含 17500 张训练切片、2500 张验证切片和 5000 张测试切片. 在实际磁共振成像过程中, 随着采样点的减少, MR 图像的分辨率会显著降低, 并受到光学特性等因素的影响, 从而出现一定程度的模糊. 为了更贴近真实应用场景, 本实验采用双三次插值与高斯模糊相结合的双重降质方式来模拟生成低分辨率图像.

2.1.2 训练细节

对于 PGDformer, 本文采用 4 层编码-解码结构, 用于多尺度特征提取以及动态调整参数. 从第 1 层

到第4层,在PDM-PSA中注意力头设置为 $[1, 2, 4, 8]$,通道数为 $[32, 64, 128, 256]$,窗口大小设置为 16×16 .对于扩散模型,总的时间步长设置为4,扩散系数 β_t 从 $\beta_1 = 0.1$ 线性增长到 $\beta_T = 0.99$,去噪网络的线性层数为5.采用Adam优化器来训练整个模型,初始学习率设置为0.0002,使用学习率衰减策略,分别在10万次、20万次、30万次和40万次迭代时衰减至原学习率的一半,总迭代次数为50万次.所有实验均在Pytorch深度学习框架下进行,GPU为两张NVIDIA GTX 1080Ti, Python版本为3.9.

2.2 对比实验

2.2.1 客观指标分析

为了验证本文方法的有效性,与双三次插值法(Bicubic)、SwinIR^[13]、DiffIR^[18]、EDSR^[21]、ESRGAN^[22]、SR3^[23]和SinSR^[24]等7种有代表性的超分辨率方法在不同放大因子(2×和4×)下进行对比,评价指标采用PSNR和SSIM.

对比结果如表1所示.从表1可以看出,当放大因子为2×时,EDM-MSR在PD和T2模态上均取得了最优的PSNR和SSIM结果.在PD模态下,相比于次优方法DiffIR提升了0.09 dB的PSNR,同时SSIM也达到了最高值0.9805;与基于CNN的EDSR和基于Transformer的SwinIR相比,EDM-MSR的PSNR分别提高了1.06 dB和0.62 dB,表明扩散模型和Transformer的融合设计能够更有效地捕捉图像的高频细节;在4×放大因子下,EDM-MSR在PD、T1和T2三种模态下均取得了最优结果,PSNR相比于次优方法分别提高了0.35 dB、0.29 dB和0.24 dB,同时SSIM也取得最优值.总体而言,EDM-MSR通过轻量化的潜在空间建模和多头置换自注意力机制实现了更优的重建精度.

为了进一步评估不同扩散模型在MRI超分辨率任务中的计算效率,本文对SR3、SinSR、DiffIR以及本文提出的EDM-MSR进行参数量、步长和推理时间的对比分析.实验基于4×放大因子,在T2模态数据上进行推理时间统计,结果如表2所示.可以看出,SR3由于其参数量高达155 M,并且使用1000步的步长,因此推理时间达到92.56 s,计算效率较低.而SinSR采用1步去噪的方式,推理时间仅为1.51 s,但由于步长过短,对最终重建的图像质量有所影响.DiffIR和EDM-MSR均采用4步去噪,但EDM-MSR的参数量仅为20 M,相比于DiffIR(26 M)更轻量化,并且推理时间进一步缩短至3.75 s,比DiffIR(4.26 s)提升了约12%的推理效率,同时仍能

表1 不同方法在2×和4×放大因子下的平均PSNR/SSIM对比
Tab. 1 Average PSNR/SSIM comparison of different algorithms at 2× and 4× amplification factor

方法	倍数	PD		T1		T2	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
		↑	↑	↑	↑	↑	↑
Bicubic		31.04	0.9007	30.11	0.8972	30.80	0.9001
EDSR		37.98	0.9763	36.63	0.9715	37.59	0.9786
ESRGAN		37.41	0.9514	36.57	0.9712	37.07	0.9772
SwinIR	2×	38.42	0.9777	36.65	0.9715	37.62	0.9773
SR3		37.52	0.9677	36.35	0.9712	37.49	0.9768
DiffIR		<u>38.95</u>	<u>0.9796</u>	37.85	<u>0.9725</u>	<u>38.51</u>	<u>0.9795</u>
SinSR		38.52	0.9795	36.72	0.9722	37.69	0.9774
EDM-MSR(本文)		39.04	0.9805	<u>37.67</u>	0.9739	38.75	0.9801
Bicubic		26.47	0.8174	25.53	0.7928	25.61	0.8098
EDSR		30.54	0.9056	29.92	0.8878	30.02	0.8972
ESRGAN		30.62	0.9058	29.74	0.8819	29.89	0.8857
SwinIR	4×	30.97	0.9068	30.01	<u>0.8910</u>	30.60	0.9015
SR3		29.67	0.8885	28.65	0.8463	29.31	0.8849
DiffIR		<u>31.67</u>	<u>0.9115</u>	<u>30.09</u>	0.8907	<u>30.89</u>	<u>0.9076</u>
SinSR		31.52	0.9102	30.05	0.8887	30.86	0.9064
EDM-MSR(本文)		32.02	0.9197	30.38	0.8993	31.13	0.9135

注:蓝色和下划线字体分别表示各列最优和次优结果.

保证高质量的重建结果.

表2 放大因子为4×时在T2模态上的模型效率分析

Tab. 2 Model efficiency analysis on T2 modality at 4× amplification factor			
方法	参数量/M	步长	推理时间/s
SR3	155	1000	92.56
DiffIR	26	4	4.26
SinSR	115	1	1.51
EDM-MSR(本文)	20	4	3.75

2.2.2 视觉质量分析

为评估不同超分辨率方法在MRI图像重建中的视觉表现,本文在放大因子为4×的条件下,分别在PD、T1和T2三种模态数据集上进行了定性对比分析.图4-图6展示了各方法在不同模态下的重建结果:第1行为整体重建图像,第2行为感兴趣区域的局部放大图,第3行为对应的误差图.误差图颜色越接近蓝色表示重建误差越小,即重建效果越好;颜色越亮则说明误差较大,图像质量较差.

从图中可以看出,传统插值方法在细节恢复方面存在明显不足,生成的图像边缘模糊,纹理信息

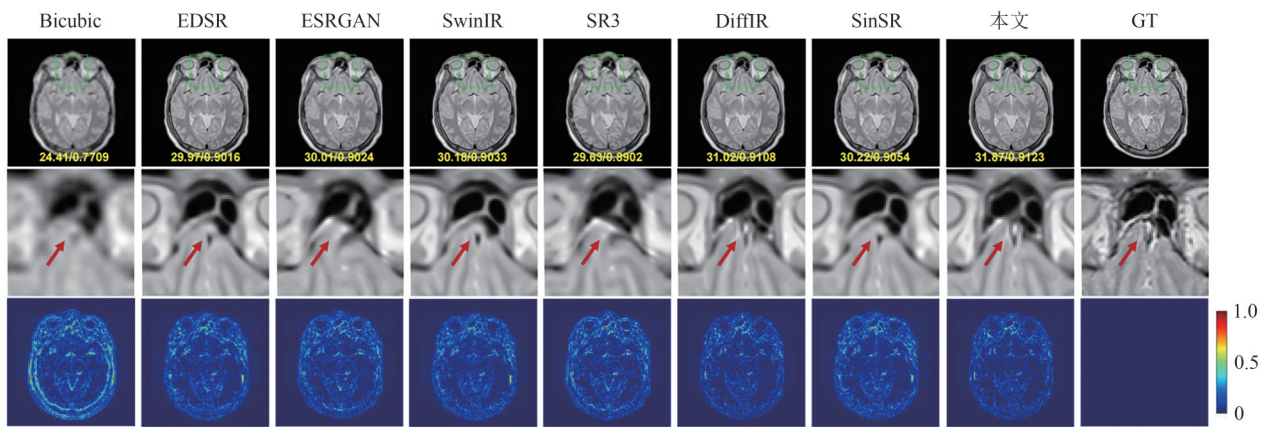


图4 放大因子为4×时不同SR方法在PD模态下的定性结果对比(包括重建图像、局部放大图和误差图)

Fig. 4 Qualitative comparison of different SR methods on the PD modality at 4× amplification factor (including reconstructed images, zoomed-in regions, and error maps)

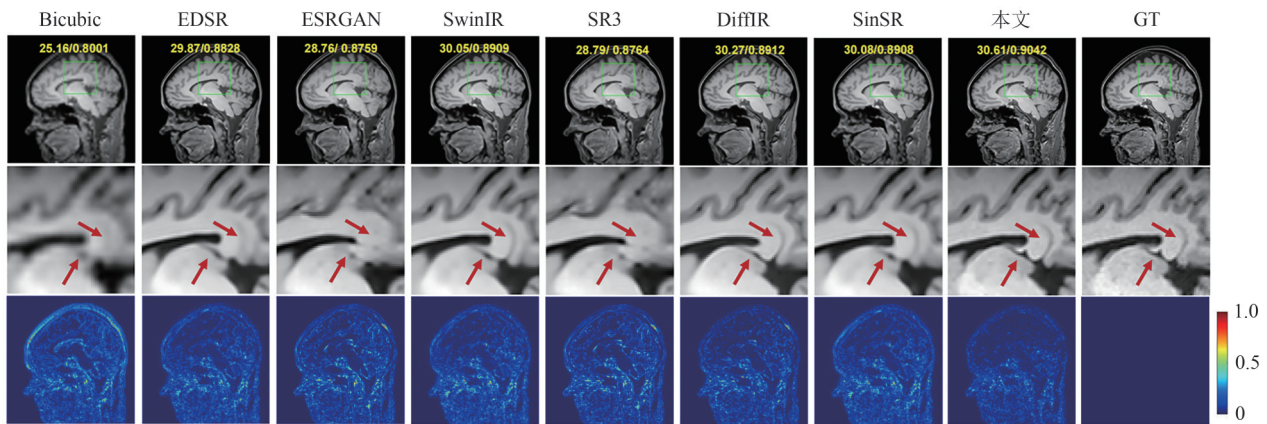


图5 放大因子为4×时不同SR方法在T1模态下的定性结果对比(包括重建图像、局部放大图和误差图)

Fig. 5 Qualitative comparison of different SR methods on the T1 modality at 4× amplification factor (including reconstructed images, zoomed-in regions, and error maps)

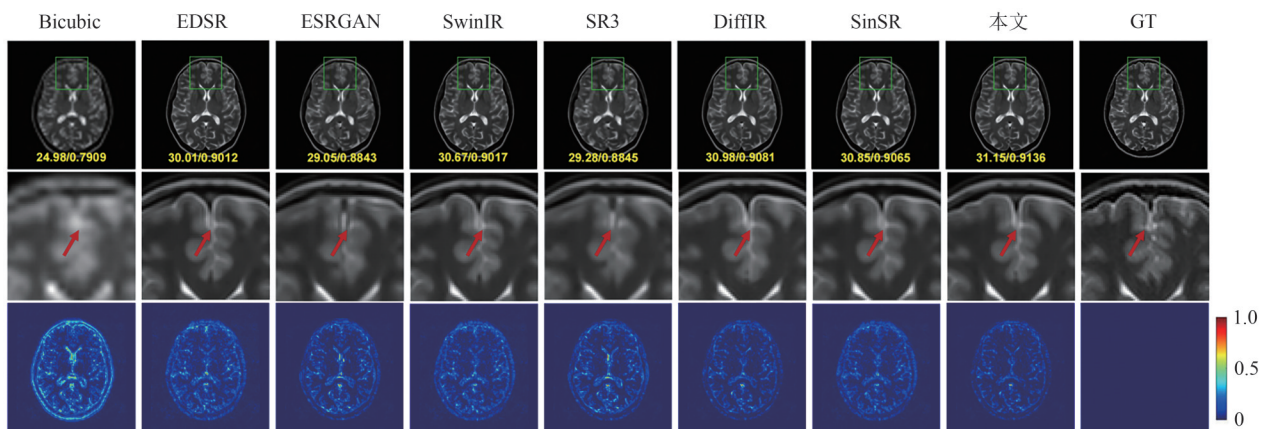


图6 放大因子为4×时不同SR方法在T2模态下的定性结果对比(包括重建图像、局部放大图和误差图)

Fig. 6 Qualitative comparison of different SR methods on the T2 modality at 4× amplification factor (including reconstructed images, zoomed-in regions, and error maps)

丢失严重.基于扩散模型的方法(如DiffIR、SinSR以及EDM-MSR)在恢复结构信息方面表现较优,能够更准确地重建组织细节.值得注意的是,本文提出的方法EDM-MSR在多个模态下均展现出更优的视觉质量.在局部放大区域中,相比于基于CNN和基

于Transformer的SR方法,EDM-MSR生成的图像具有更清晰的边缘和更丰富的纹理信息,并且误差较小,表明该方法在医学图像的高频信息方面具有更好的表现.

2.3 消融实验

为了验证本文提出的 EDM-MSR 模型中各个关键组件对 MRI 超分辨率重建任务的有效性,设计了一系列消融实验,主要针对扩散先验表示、条件信息、联合训练策略、置换自注意力以及扩散模型的反向去噪过程中的迭代次数进行分析.不同的变体均采用相同的训练策略,完成训练后在放大因子为 4× 的 T2 数据集上进行测试.

2.3.1 扩散先验的有效性

在本研究的完整模型中,利用紧凑的扩散先验表示来指导 PGDformer 进行超分辨率重建,以增强

重建性能.为了验证扩散先验的作用,在对照组中仅保留第一阶段的训练,去除第二阶段的条件信息提取网络和 PGDformer 重建模块,将此变体记为 Baseline.实验结果如表 3 中第 1 行所示,可见有明显下降,PSNR 从 31.13 dB 下降至 30.06 dB,SSIM 从 0.9135 下降至 0.8997.由于没有了包含高频细节的先验表示的引导,故对最终的纹理细节重建产生了较大的影响.这表明扩散模型可以有效捕捉脑部 MR 图像的先验分布,并为超分辨率重建提供关键的高频信息引导.

表 3 放大因子为 4× 时,在 T2 数据集上不同组件对模型重建性能的有效性比较

Tab. 3 Comparison of the effects of different variants on model performance on T2 at 4× amplification factor

变体	组件				评价指标	
	扩散先验	条件信息	联合训练	置换自注意力	PSNR	SSIM
Baseline	×	×	×	×	30.06	0.8997
w/o CEN	√	×	√	√	30.89	0.9115
w/o Joint	√	√	×	√	30.56	0.9032
w/o PSA	√	√	√	×	30.98	0.9126
EDM-MSR	√	√	√	√	31.13	0.9135

2.3.2 条件信息的有效性

在完整模型中,第二阶段的训练中加入了条件信息提取模块,以充分利用低分辨率图像的特征信息.为验证该部分的贡献,在对照组实验中去除了条件提取网络,并保持模型其他组件不变,将此变体记为 w/o CEN.实验结果如表 3 中第 2 行所示,PSNR 和 SSIM 均有下降,这表明条件信息提取网络可以为扩散模型提供更准确的引导信息,从而提升最终的重建质量.

2.3.3 联合训练策略的有效性

传统的两阶段训练方法首先独立训练自动编码器,进而在固定编码器参数的基础上训练解码器.本文的联合训练策略则通过在第二阶段中同时优化自动编码器(CEN)和解码器(PGDformer),以实现更紧密的特征融合和信息利用.为了验证联合训练策略的有效性,在第二阶段实验中只训练 CEN 和扩散模型,并且仅利用扩散损失来优化网络,完成训练后将其和第一阶段的 PGDformer 直接结合起来重建图像,将此变体记为 w/o Joint.实验结果如表 3 中第 3 行所示,PSNR 由 31.13 dB 下降至 30.56 dB,SSIM 从 0.9135 下降至 0.9032.这表明,在分开训练的情况下,扩散模型和 PGDformer 无法充分协同,导致模型在重建过程中对细节的恢复能力下降.因此,联合

训练策略可以实现端到端优化,从而有效提升了模型的整体性能.

2.3.4 置换自注意力的有效性

在 PGDformer 中引入了置换自注意力机制,在保持计算复杂度可控的同时,有效扩大了感受野,从而增强了全局信息的建模能力.为验证置换自注意力机制的有效性,在对照组实验中将置换自注意力替换成普通的 3×3 深度卷积,将此变体记为 w/o PSA.实验结果如表 3 中第 4 行所示,PSNR 和 SSIM 相比于完整模型均有下降,表明在大窗口范围内执行注意力计算能够更充分地捕捉全局上下文信息,从而提升重建质量.

2.3.5 扩散模型的迭代次数

扩散模型的反向去噪过程通常需要多次迭代,以逐步恢复高质量图像.为探讨不同迭代次数对重建性能的影响,本文设置了 6 种迭代次数 $T = \{1, 2, 4, 8, 16, 32\}$,对 PSNR 变化趋势进行分析,实验结果如图 7 所示.当 $T=1$ 时,反向过程生成的先验表示不足,导致 PSNR 显著偏低;随着 T 的增加,重建性能明显提升,当 $T=4$ 时已接近饱和,此后继续增加迭代次数对性能提升的作用趋于减弱(如从 $T=4$ 增加至 $T=32$,PSNR 提升不足 0.01 dB).这说明在高度压缩的潜在空间中,仅需较少的反向步骤即可获得稳定有

效的先验表示.综合考虑性能提升与计算开销,本文最终选取 $T=4$ 作为去噪步数,以在精度与效率之间取得平衡.

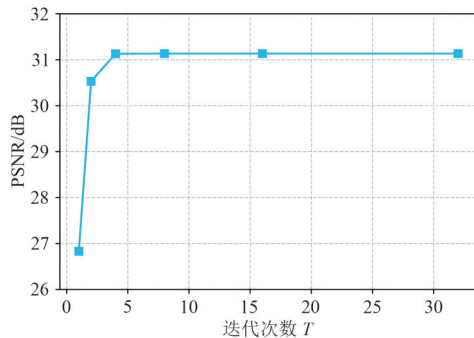


图 7 扩散模型中迭代次数 T 的消融研究

Fig. 7 Ablation study of the number of iteration T in the diffusion model

3 结论

本文提出了一种面向脑部 MRI 的超分辨率重建方法,利用潜在扩散模型生成的先验来引导动态 Transformer (PGDformer) 进行重建.通过深度可分离卷积实现先验提取的轻量化设计,引入置换自注意力机制实现在大窗口中高效计算自注意力,从而提升全局特征建模能力,有效提升了重建图像的结构表达与高频纹理细节还原能力. IXI 多模态 MRI 数据集上的实验结果表明:本文方法在重建图像主客观质量和重建效率等方面均展现出较好的结果.

参 考 文 献

- [1] LI Y, SIXOU B, PEYRIN F. A review of the deep learning methods for medical images super resolution problems[J]. IRBM, 2021, 42(2): 120-133.
- [2] WANG Z, CHEN J, HOI S C H. Deep learning for image super-resolution: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3365-3387.
- [3] JI Z, ZOU B, KUI X, et al. Deep learning-based magnetic resonance image super-resolution: A survey [J]. Neural Computing and Applications, 2024, 36(21): 12725-12752.
- [4] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution [C]// Computer Vision-ECCV 2014. Cham: Springer, 2014: 184-199.
- [5] QIU D, ZHANG S, LIU Y, et al. Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning[J]. Computer Methods and Programs in Biomedicine, 2020, 187: 105059.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [J]. arXiv: 2020.2010.11929.
- [8] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2022: 9992-10002.
- [9] YAN C, SHI G, WU Z. SMIR: A Transformer-Based Model for MRI super-resolution reconstruction [C]// 2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE). Hefei: IEEE, 2022: 1-6.
- [10] FENG C M, YAN Y, FU H, et al. Task transformer network for joint MRI reconstruction and super-resolution [C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer, 2021: 307-317.
- [11] FORIGUA C, ESCOBAR M, ARBELAEZ P. SuperFormer: Volumetric transformer architectures for MRI super-resolution [C]// Simulation and Synthesis in Medical Imaging. Cham: Springer, 2022: 132-141.
- [12] HUANG S, LIU X, TAN T, et al. TransMRSR: Transformer-based self-distilled generative prior for brain MRI super-resolution [J]. The Visual Computer, 2023, 39(8): 3647-3659.
- [13] LIANG J, CAO J, SUN G, et al. SwinIR: Image restoration using swin transformer [C]// 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal: IEEE, 2021: 1833-1844.
- [14] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [C]// Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: ACM, 2020: 6840-6851.
- [15] CHUNG H, LEE E S, YE J C. MR image denoising and super-resolution using regularized reverse diffusion [J]. IEEE Transactions on Medical Imaging, 2023, 42(4): 922-934.
- [16] WU Z, CHEN X, XIE S, et al. Super-resolution of brain MRI images based on denoising diffusion probabilistic model [J]. Biomedical Signal Processing and Control, 2023, 85: 104901.
- [17] 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, 2022: 10674-10685.
- [18] XIA B, ZHANG Y, WANG S, et al. DiffIR: Efficient diffusion model for image restoration [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris:IEEE, 2024: 13049-13059.
- [19] CHEN Z, ZHANG Y, LIU D, et al. Hierarchical integration diffusion model for realistic image deblurring[J]// arXiv: 2023,2305.12966.
- [20] ZHOU Y, LI Z, GUO C L, et al. SRFormer: Permuted self-attention for single image super-resolution [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris:IEEE, 2024: 12734-12745.
- [21] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu:IEEE, 2017: 1132-1140.
- [22] WANG X, YU K, WU S, et al. ESRGAN: Enhanced super-resolution generative adversarial networks [C]// Computer Vision-ECCV 2018 Workshops. Cham: Springer, 2019: 63-79.
- [23] 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.
- [24] WANG Y, YANG W, CHEN X, et al. SinSR: Diffusion-based image super-resolution in a single step [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle:IEEE, 2024: 25796-25805.

(责编 曹东,校对 雷建云)