

中图法分类号: TP391.41; TP18 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-11

论文引用格式: Huang Jingle, Zhou Siwang. Dual-stage gradient-guided diffusion model for image super-resolution[J/OL]. Journal of Image and Graphics, XXXX:1-11. DOI: 10.11834/jig.260307. (黄京乐, 周四望. 双阶段梯度引导的扩散模型图像超分辨率方法[J/OL]. 中国图象图形学报, XXXX:1-11. DOI: 10.11834/jig.260307.) [DOI: 10.11834/jig.260307]

双阶段梯度引导的扩散模型图像超分辨率方法

黄京乐, 周四望

湖南大学计算机学院, 长沙 410082

摘要: 目的 现有基于卷积网络或Transformer的图像超分辨率方法受限于像素级损失函数的设计, 重建结果常出现细节模糊、高频纹理重建不足等问题。近年来扩散模型凭借强大的生成先验在感知质量上取得了显著提升, 但在纹理与边缘等细节区域容易生成与输入视觉内容不一致的幻觉纹理。针对上述问题, 提出一种双阶段梯度引导的扩散模型图像超分辨率方法, 从训练与推理两个阶段同时增强生成过程的视觉一致性。**方法** 训练阶段引入频域视觉一致性损失, 通过前向视觉算子刻画高分辨率图像的频谱结构, 并将其梯度注入噪声预测, 引导模型更准确地重建高频细节; 推理阶段基于扩散后验采样, 设计测量一致性与流形粘合的双重约束, 对每一步潜变量更新进行校正, 使采样轨迹既符合观测退化模型, 又保持在自然图像分布内。**结果** 在RealSR(real-world super-resolution)和DRealSR(diverse real-world super-resolution)两个数据集上与现有主流方法进行对比实验。客观指标方面, 与基线扩散超分辨率方法相比, 所提方法在RealSR数据集上的峰值信噪比(peak signal-to-noise ratio, PSNR)提升1.756 dB, 结构相似性(structural similarity index measure, SSIM)提升0.0115, 感知图像块相似度(learned perceptual image patch similarity, LPIPS)降低2.76%; 在DRealSR数据集上PSNR提升2.139 dB, SSIM提升0.0203, LPIPS降低10.88%。主观评价方面, 该方法在字符纹理等高频细节区域有效抑制了幻觉伪影, 结构准确性与纹理真实感优于对比方法。消融实验验证了各模块的有效性及其协同增益。**结论** 所提出的双阶段梯度引导的扩散模型图像超分辨率方法通过训练阶段的频域视觉引导与推理阶段的双重约束后验采样, 有效抑制了幻觉纹理, 提升了重建结果的保真度与结构一致性, 在感知质量与重建保真度的综合权衡上优于现有方法。

关键词: 图像超分辨率; 扩散模型; 梯度引导; 扩散后验采样; 重建保真度

Dual-stage gradient-guided diffusion model for image super-resolution

Huang Jingle, Zhou Siwang

Hunan University, Changsha 410082, China

Abstract: **Objective** Image super-resolution (ISR) aims to reconstruct high-resolution (HR) images from degraded low-resolution (LR) inputs, and is an important task in computer vision and image processing. Conventional methods based on convolutional neural networks (CNNs) or Transformer architectures have achieved considerable progress, but they are usually optimized with pixel-wise losses, such as mean squared error (MSE) or mean absolute error (MAE). These losses tend to produce over-smoothed results and insufficient high-frequency details, especially in regions with fine textures, characters, and sharp edges. In recent years, diffusion models have shown strong generative priors for image super-resolution and can produce perceptually realistic textures. However, their stochastic sampling process may generate hallucinated textures

收稿日期: 2026-05-28; 修回日期: 2026-07-24

基金项目: 国家自然科学基金项目(项目编号: 62572181); 湖南省自然科学基金项目(项目编号: 2026JJ80030)

Supported by: National Natural Science Foundation of China (No. 62572181); Natural Science Foundation of Hunan Province, China (No. 2026JJ80030)

that are inconsistent with the input LR observation, which weakens the structural reliability of reconstructed images. To address this problem, a dual-stage gradient-guided diffusion model for image super-resolution is proposed to enhance visual consistency from both the training and inference stages. **Method** The proposed method introduces explicit gradient guidance into two stages of a diffusion-based super-resolution pipeline. In the training stage, a frequency-domain visual consistency loss is used to supervise the noise prediction network. A forward visual operator extracts the spectral representation of the HR target image, and the gradient derived from this frequency-domain loss is injected into the noise prediction process. This strategy encourages the denoising network to learn latent representations that are more consistent with the frequency distribution of real HR images, thereby improving the reconstruction of high-frequency structures without modifying the base network architecture. In the inference stage, a dual-constraint diffusion posterior sampling strategy is designed to correct the latent variable update at each denoising step. The first constraint is measurement consistency, which enforces consistency between the generated HR result and the input LR observation under the degradation model. The second constraint is manifold gluing, which encourages the corrected latent variables to remain close to the natural image manifold and reduces structural drift during sampling. By combining training-stage frequency-domain visual guidance with inference-stage dual-constraint posterior sampling, the proposed framework suppresses hallucinated textures while preserving perceptual quality and structural consistency. **Result** Experiments are conducted on two super-resolution benchmarks, RealSR (real-world super-resolution) and DRealSR (diverse real-world super-resolution), under the $\times 4$ upscaling setting. The proposed method is compared with representative super-resolution methods through quantitative and qualitative evaluations. The evaluation metrics include peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) for fidelity assessment, learned perceptual image patch similarity (LPIPS) for perceptual similarity evaluation, CLIP image quality assessment (CLIP-IQA) and multi-scale image quality transformer (MUSIQ) for no-reference perceptual quality evaluation, and natural image quality evaluator (NIQE) for blind image quality assessment. Compared with the baseline diffusion-based super-resolution method, the proposed method improves PSNR by 1.756 dB, improves SSIM by 0.0115, and reduces LPIPS by 2.76% on the RealSR dataset. On the DRealSR dataset, the proposed method improves PSNR by 2.139 dB, improves SSIM by 0.0203, and reduces LPIPS by 10.88%. These results indicate that the proposed method improves reconstruction fidelity on both datasets while maintaining a reasonable balance between fidelity and perceptual quality. Qualitative comparisons further show that the proposed method effectively suppresses hallucination artifacts in high-frequency regions, such as character textures and complex patterns, and generates results with improved structural accuracy and texture fidelity. Ablation experiments verify that both the training-stage frequency-domain guidance and the inference-stage dual-constraint posterior sampling contribute to the final performance, and their combination provides complementary gains. **Conclusion** A dual-stage gradient-guided diffusion model for image super-resolution is proposed. By incorporating frequency-domain visual guidance during training and dual-constraint posterior sampling during inference, the proposed framework suppresses hallucinated textures and improves the fidelity and structural consistency of reconstructed images. Experimental results on RealSR and DRealSR demonstrate that the proposed method achieves a better balance between reconstruction fidelity and perceptual quality than the baseline diffusion-based method. Future work will focus on reducing the computational cost of gradient-guided sampling and extending the framework to blind super-resolution with unknown degradations.

Key words: image super-resolution; diffusion model; gradient guidance; diffusion posterior sampling; reconstruction fidelity

论文引用格式: DOI: 10.11834/jig.260307

0 引言

图像超分辨(Image super-resolution, ISR)旨在从退化的低分辨率(low resolution, LR)图像中重建具

有丰富细节的高分辨率(high resolution, HR)图像。现有基于卷积网络或Transformer的图像超分辨率方法以像素级损失为主要训练目标(Lim等, 2017; Liang等, 2021), 重建结果在感知层面存在细节模糊、高频纹理重建不足等问题(刘紫阳等, 2026)。为改善感知质量, 生成对抗网络(generative adver-

serial network, GAN)方法通过对抗训练建模高质量图像分布,在纹理清晰度上取得显著进展(Ledig 等, 2017),但其固有的训练不稳定性与模式崩溃问题限制了模型的实用性(谭明奎 等, 2021),且对复杂退化的泛化能力有限。近年来,扩散模型凭借对复杂图像分布的强建模能力在图像生成领域取得了突破性进展,催生了一系列将扩散模型引入图像超分辨率任务的研究,在生成逼真纹理细节方面展现出强大的先验建模能力(Rombach 等, 2022; 韦炎炎 等, 2025)。

现有基于扩散模型的图像超分辨率方法主要围绕如何将 LR 信息融入扩散生成网络展开研究,其核心在于网络结构与训练策略的设计。早期工作采用通道拼接策略:Saharia 等(2023)将插值放大后的 LR 图像与带噪 HR 图像在通道维度拼接后送入去噪网络;Ho 等(2022)在此基础上进一步将该策略扩展至级联框架。但这种方式需要针对特定退化模型从头训练,灵活性有限。后续工作转向在预训练扩散模型的基础上,通过更精细的特征编码与对齐机制融合 LR 信息。StableSR 利用编码器提取特征注入生成模型(Wang J Y 等, 2024);FaithDiff 通过联合微调与对齐模块,使退化输入特征与扩散过程中的噪声潜变量实现更好的对齐(Chen 等, 2025);ControlSR 利用潜在 LR 嵌入对 ControlNet 的控制信号施加更强约束(Wan 等, 2024);StructSR 引入结构感知筛选机制,在推理早期筛选与 LR 结构相似度最高的候选图像作为结构参考,减轻结构偏差(Li Y C 等, 2025)。这些方法在感知质量与结构保真度上取得了不同程度的改善,但其共同局限在于:LR 信号仅在特征层面进行隐式调节,难以对扩散先验施加牢固的观测约束,导致模型在生成高频细节区域时容易产生与真实视觉内容不一致的幻觉纹理。如图 1 所示,以 StableSR 为例,其生成结果虽然视觉清晰,却将 HR 图像中的“mm”字符错误生成“mos”和“mes”,表现出明显的幻觉纹理问题。

与上述通过网络结构融合 LR 信息的思路不同,扩散后验采样(diffusion posterior sampling, DPS)从逆问题求解的视角出发,提出了一种在扩散模型采样过程中施加观测约束的通用框架(Chung 等, 2023)。该方法将 ISR 视为后验采样问题,在每一步去噪迭代中利用数据一致性损失对潜变量进行梯度校正,从而增强生成结果与观测数据的一致性。后

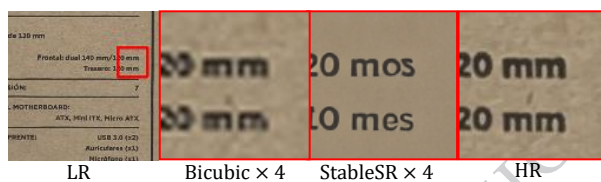


图 1 扩散超分辨率方法的幻觉纹理问题

Fig. 1 Hallucination artifacts in diffusion-based super-resolution

续工作对这一框架进行了不同维度的扩展:Rout 等(2023)将后验采样扩展至潜在扩散模型,Wang X D 等(2025)设计了即插即用的引导模块以适配不同的逆问题设定。在超分辨率领域,Li X Y 等(2025)提出 DiffISR,将梯度引导思想引入训练阶段,通过在训练过程中注入视觉一致性梯度增强模型对高频细节的感知能力,在红外图像超分辨率任务中取得了显著效果。尽管 DiffISR 专注于红外图像超分辨率,但其将梯度引导引入训练阶段的思路具有普适性,为自然图像超分辨率任务提供了思路。这一工作表明,梯度引导机制不仅可用于推理校正,还可在训练阶段优化模型的先验建模能力。

受 DPS 梯度引导思想与 DiffISR 训练阶段引导实践的启发,本文提出一种面向条件扩散超分辨率模型的双阶段梯度引导框架,在训练阶段和推理阶段同时引入显式约束。训练阶段通过频域视觉引导增强模型对高频结构的感知能力;推理阶段通过测量一致性约束与流形粘约束联合校正采样轨迹,以增强结果与输入观测之间的一致性。双阶段机制分别作用于先验建模和后验校正两个环节,前者改善模型对高频细节的学习能力,为推理阶段提供更好的先验起点;后者进一步补偿先验模型的残余偏差,减轻推理过程中可能出现的结构偏差与幻觉纹理。该框架可在不修改基础网络结构的前提下集成到现有条件扩散超分辨率模型中。

本文的主要贡献如下:1)在训练阶段,引入基于前向视觉算子的频域视觉引导策略,将高频结构约束的梯度信息直接注入噪声预测过程,增强模型对高频细节的建模能力;2)在推理阶段,设计兼顾测量一致性与流形粘合的扩散后验采样机制,以双重梯度校正每一步的生成轨迹,在保持感知质量的同时抑制幻觉纹理;3)在 RealSR、DRealSR 数据集上的实验表明,所提方法能够提升重建结果的保真度,并在感知质量与结构一致性之间取得较好的平衡。消融

研究确认了双阶段机制的独立贡献与协同增益。

1 方法

本文提出一种面向条件扩散超分辨率模型的双阶段梯度引导方法。该方法在训练和推理两个阶段引入显式梯度引导:训练阶段通过频域视觉一致性损失增强模型对高频细节的建模能力;推理阶段通过测量一致性约束与流形粘约束校正采样轨迹。二者分别作用于先验学习和后验采样过程,从而增强重建结果与输入观测之间的一致性。

1.1 问题定义

ISR任务可建模为线性逆问题 $\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{n}$, 其中, $\mathbf{x}_0$ 表示 HR 图像, $\mathbf{y}$ 表示 LR 观测, $\mathcal{A}(\cdot)$ 表示退化或观测算子, $\mathbf{n}$ 为测量噪声。该任务旨在根据 $\mathbf{y}$ 恢复符合自然图像分布的高分辨率结果, 即从后验分布 $p(\mathbf{x}_0 | \mathbf{y})$ 中进行采样。

在潜在扩散超分辨率框架中, 编码器 $\mathcal{E}$ 将 $\mathbf{x}_0$ 和 $\mathbf{y}$ 分别映射到潜在空间, 其过程表示为

$$\mathbf{z}_0 = \mathcal{E}(\mathbf{x}_0), \quad \mathbf{z}_y = \mathcal{E}(\mathbf{y}) \quad (1)$$

式中, $\mathbf{z}_0, \mathbf{z}_y$ 分别为 $\mathbf{x}_0, \mathbf{y}$ 在潜在空间中的表示。在扩散过程中, $\mathbf{z}_0$ 经逐步加噪得到 $\mathbf{z}_t$ 。去噪网络 ϵ_θ 以 $(\mathbf{z}_t, \mathbf{z}_y, t)$ 为输入预测噪声, 其标准训练目标为

$$\mathcal{L}_{\text{SR}} = \mathbb{E}_{\mathbf{z}_0, \mathbf{z}_y, t} [\|\epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t)\|_2^2] \quad (2)$$

式中, ϵ 为标准高斯噪声, $\epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t)$ 为去噪网络的噪声预测, $\mathbb{E}[\cdot]$ 表示数学期望, t 为时间步。

为统一描述训练阶段与推理阶段的显式引导机制, 从后验得分的角度刻画潜变量的更新方向。标准条件扩散模型学习以 $\mathbf{z}_y$ 为条件的先验分布 $p(\mathbf{z}_0 | \mathbf{z}_y)$, 在此基础上引入额外的显式约束信号, 定义为 $\tilde{\mathbf{y}}$, 根据贝叶斯恒等式, 后验得分可写为

$$\begin{aligned} \nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t | \mathbf{z}_y, \tilde{\mathbf{y}}) = & \nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t | \mathbf{z}_y) \\ & + \nabla_{\mathbf{z}_t} \log p(\tilde{\mathbf{y}} | \mathbf{z}_t, \mathbf{z}_y) \end{aligned} \quad (3)$$

式中, 第一项表示以 LR 编码为条件的扩散先验得分, 由 ϵ_θ 直接近似; 第二项表示在 $\mathbf{z}_y$ 之外额外施加的观测约束。由于第二项难以解析求解, 本文采用 Tweedie 公式 (Efron, 2011; Chung 等, 2023) 进行近似, 任意时刻 t 的洁净潜变量估计值 $\hat{\mathbf{z}}_0$ 可写为

$$\hat{\mathbf{z}}_0(\mathbf{z}_t) = \frac{\mathbf{z}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t)}{\sqrt{\bar{\alpha}_t}} \quad (4)$$

式中, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ 为累积信号保留系数, $\alpha_s = 1 - \beta_s$, $\beta_s \in (0, 1)$ 为第 s 个时间步的噪声调度系数。进一步, 将潜变量解码到像素空间, 对应的图像估计表示为

$$\hat{\mathbf{x}}_0(\mathbf{z}_t) = \mathcal{D}(\hat{\mathbf{z}}_0(\mathbf{z}_t)) \quad (5)$$

式中, $\mathcal{D}$ 为解码器。

参考 DPS 推导框架, 假设 $\tilde{\mathbf{y}}$ 与观测映射结果 $\mathcal{A}(\mathbf{x}_0)$ 之间满足加性高斯观测误差模型, 其中, $\mathcal{A}(\cdot)$ 为阶段相关的观测映射算子, 并利用 $\hat{\mathbf{x}}_0(\mathbf{z}_t)$ 作为 $\mathbf{x}_0$ 的点估计。在该假设下, $\log p(\tilde{\mathbf{y}} | \mathbf{z}_t, \mathbf{z}_y)$ 可近似为负的二范数平方误差形式。因此, 额外观测约束项可近似为重建误差对潜变量的梯度, 可表示为

$$\nabla_{\mathbf{z}_t} \log p(\tilde{\mathbf{y}} | \mathbf{z}_t, \mathbf{z}_y) \approx -\nabla_{\mathbf{z}_t} \|\tilde{\mathbf{y}} - \mathcal{A}(\hat{\mathbf{x}}_0(\mathbf{z}_t))\|_2^2 \quad (6)$$

训练阶段由于可直接获得 $\mathbf{x}_0$, 可取 $\mathcal{A}(\cdot)$ 为频域视觉变换, 此时 $\tilde{\mathbf{y}} = \mathcal{A}(\mathbf{x}_0)$ 表示由 $\mathbf{x}_0$ 导出的目标约束; 推理阶段取 $\mathcal{A}(\cdot)$ 为下采样算子, 此时 $\tilde{\mathbf{y}} = \mathbf{y}$ 为 LR 观测。因此, 两个阶段可写为统一形式的显式观测约束建模, 只是 $\mathcal{A}(\cdot)$ 与 $\tilde{\mathbf{y}}$ 的具体定义不同。

结合式 (3) 和式 (6), 并将额外观测约束项的权重记为 λ , 后验得分的完整近似形式可写为

$$\begin{aligned} \nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t | \mathbf{z}_y, \tilde{\mathbf{y}}) \approx & \frac{\epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t)}{\sqrt{1 - \bar{\alpha}_t}} \\ & - \lambda \nabla_{\mathbf{z}_t} \|\tilde{\mathbf{y}} - \mathcal{A}(\hat{\mathbf{x}}_0(\mathbf{z}_t))\|_2^2 \end{aligned} \quad (7)$$

式中, 第一项由扩散去噪网络给出, 体现以 $\mathbf{z}_y$ 为条件的生成先验; 第二项为显式观测约束项, 驱动采样轨迹向与约束信号一致的方向修正。基于这一框架, 本文在训练阶段令 $\mathcal{A}(\cdot)$ 为频域变换, 利用 $\tilde{\mathbf{y}} = \mathcal{A}(\mathbf{x}_0)$ 构造频域显式观测约束, 增强模型对高频结构的建模能力; 在推理阶段令 $\mathcal{A}(\cdot)$ 为下采样算子, 利用 $\tilde{\mathbf{y}} = \mathbf{y}$ 构造测量一致性约束与流形粘约束, 校正采样轨迹并抑制幻觉纹理。

1.2 训练阶段

1.2.1 频域视觉引导的训练算法

训练阶段通过频域视觉一致性损失度量当前预测结果与 HR 图像之间的高频差异, 并将该损失对潜变量 $\mathbf{z}_t$ 的梯度注入噪声预测过程, 使模型在参数学习阶段增强对高频细节的感知能力。该机制不依赖于特定的视觉算子, 此处以频域幅度表征作为默认实现, 但同样适用于其他可微的特征映射。完整的训练流程如算法 1 所示。

算法 1: 频域视觉引导的训练算法。

输入: 高分辨率图像 $\mathbf{x}_0$, 低分辨率图像 $\mathbf{y}$, 引导比例系数 ρ , 编码器 $\mathcal{E}$, 解码器 $\mathcal{D}$, 前向视觉算子 $\mathcal{V}$, 时间步总数 T 。

输出: 优化后的模型参数 θ 。

- 1) repeat
- 2) $\mathbf{x}_0 \sim q(\mathbf{x}_0), \tilde{\mathbf{y}} \leftarrow \mathcal{V}(\mathbf{x}_0), \mathbf{z}_y \leftarrow \mathcal{E}(\mathbf{y});$
- 3) $\mathbf{z}_0 \leftarrow \mathcal{E}(\mathbf{x}_0);$
- 4) $t \sim \text{Uniform}(\{1, \dots, T\});$
- 5) $\epsilon \sim \mathcal{N}(0, \mathbf{I});$
- 6) $\mathbf{z}_t \leftarrow \sqrt{\bar{\alpha}_t} \mathbf{z}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon;$
- 7) $\hat{\mathbf{z}}_0 \leftarrow (\mathbf{z}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t)) / \sqrt{\bar{\alpha}_t};$
- 8) $\hat{\mathbf{x}}_0 \leftarrow \mathcal{D}(\hat{\mathbf{z}}_0);$
- 9) $\mathcal{L}_{\text{visual}} \leftarrow \|\tilde{\mathbf{y}} - \mathcal{V}(\hat{\mathbf{x}}_0)\|_2^2;$
- 10) $\hat{\epsilon}_\theta \leftarrow \epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t) + \rho \sqrt{1 - \bar{\alpha}_t} \nabla_{\mathbf{z}_t} \mathcal{L}_{\text{visual}};$
- 11) Take gradient descent step on: $\|\hat{\epsilon}_\theta\|_2^2;$

1.2.2 前向视觉算子定义

训练阶段将观测映射算子 $\mathcal{A}(\cdot)$ 实例化为前向视觉算子 $\mathcal{V}(\cdot)$ 。 $\mathcal{V}(\cdot)$ 设计为可插拔形式, 在具体实现中采用频域幅度表征, 若后续需要替换为其他特征映射, 该形式也保持成立。具体地, 对输入图像 $\mathbf{x}$ 进行二维快速傅里叶变换, 表示为

$$\mathcal{F}(u, v) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \mathbf{x}(h, w) \exp\left(-2\pi i \left(\frac{uh}{H} + \frac{vw}{W}\right)\right) \quad (8)$$

式中, H 和 W 分别表示 $\mathbf{x}$ 的高度和宽度, h 和 w 为空间位置索引, u 和 v 为频域坐标索引。随后算子 $\mathcal{V}(\cdot)$ 依次执行频谱中心化、对数幅度压缩及零均值单位方差归一化, 完整的映射过程可写为

$$\mathcal{V}(\mathbf{x}) = \text{Norm}\left(\log\left(1 + |\text{Shift}(\mathcal{F}(\mathbf{x}))|\right)\right) \quad (9)$$

式中, $\text{Shift}(\cdot)$ 表示频谱中心化, $\text{Norm}(\cdot)$ 表示统计归一化操作。 $\tilde{\mathbf{y}} = \mathcal{V}(\mathbf{x}_0)$ 为训练阶段的频域视觉引导量。

1.2.3 视觉一致性损失与梯度注入

利用 Tweedie 近似得到 $\hat{\mathbf{z}}_0$, 并解码得到像素空间的预测 $\hat{\mathbf{x}}_0 = \mathcal{D}(\hat{\mathbf{z}}_0)$ 。频域视觉一致性损失定义为

$$\mathcal{L}_{\text{visual}} = \|\tilde{\mathbf{y}} - \mathcal{V}(\hat{\mathbf{x}}_0)\|_2^2 \quad (10)$$

该损失直接约束频域表征, 有助于抑制纹理偏移和局部区域的幻觉纹理。与将 $\mathcal{L}_{\text{visual}}$ 作为辅助损失叠加不同, 本文将其对 $\mathbf{z}_t$ 的梯度注入噪声预测, 可

写为

$$\hat{\epsilon}_\theta(\mathbf{z}_t, \mathbf{z}_y, t) = \epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t) + \rho \sqrt{1 - \bar{\alpha}_t} \nabla_{\mathbf{z}_t} \mathcal{L}_{\text{visual}} \quad (11)$$

式中, ρ 为训练阶段频域引导权重。 $\sqrt{1 - \bar{\alpha}_t}$ 用于使梯度注入强度与不同时间步的噪声水平相匹配。最终训练目标可写为

$$\mathcal{L} = \mathbb{E}_{\mathbf{z}_t, \mathbf{z}_y, t} [\|\hat{\epsilon}_\theta\|_2^2] \quad (12)$$

训练过程中固定编码器和解码器的参数, 仅更新去噪网络参数 θ 。

1.3 推理阶段

1.3.1 基于测量一致性与流形粘合的双重约束扩散后验采样算法

训练阶段增强了模型先验对高频细节的建模能力, 但推理过程仍可能因随机采样而偏离输入结构。由于推理阶段无法获取 HR 图像, 因此不再使用训练阶段的频域视觉引导, 而是利用 LR 观测构造测量一致性约束, 并结合流形粘约束校正采样轨迹。完整推理流程如算法 2 所示。

1.3.2 标准扩散推进

在每个时间步 t , 去噪网络以 LR 条件嵌入 $\mathbf{z}_y$ 为

算法 2: 双重约束扩散后验采样算法。

输入: 低分辨率图像 $\mathbf{y}$, 时间步总数 T , 测量引导权重 η , 流形粘合权重 γ , 噪声调度 $\{\bar{\alpha}_t\}_{t=1}^T$, 编码器 $\mathcal{E}$, 解码器 $\mathcal{D}$, 退化算子 $\mathcal{A}$ 。

输出: 重建的高分辨率图像 $\mathcal{D}(\mathbf{z}_0)$ 。

- 1) $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I}), \mathbf{z}_y \leftarrow \mathcal{E}(\mathbf{y});$ // 初始化
- 2) for $t = T$ to 1 do:
- 3) $\hat{\epsilon}_\theta \leftarrow \epsilon_\theta(\mathbf{z}_t, \mathbf{z}_y, t);$ // 噪声预测
- 4) $\hat{\mathbf{z}}_0 \leftarrow (\mathbf{z}_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon}_\theta) / \sqrt{\bar{\alpha}_t};$ // 洁净潜在量估计
- 5) $\epsilon \sim \mathcal{N}(0, \mathbf{I});$ // 随机扰动采样
- 6) $\mathbf{z}_{t-1}' \leftarrow [\sqrt{\bar{\alpha}_t} (1 - \bar{\alpha}_{t-1}) / (1 - \bar{\alpha}_t)] \mathbf{z}_t$
 $+ [\sqrt{\bar{\alpha}_{t-1}} \beta_t / (1 - \bar{\alpha}_t)] \hat{\mathbf{z}}_0 + \bar{\sigma}_t \epsilon$;
// 标准扩散更新
- 7) $\mathbf{z}_{t-1}'' \leftarrow \mathbf{z}_{t-1}' - \eta \nabla_{\mathbf{z}_{t-1}'} \|\mathbf{y} - \mathcal{A}(\mathcal{D}(\hat{\mathbf{z}}_0))\|_2^2;$ // 测量一致性校正
- 8) $\mathbf{z}_{t-1} \leftarrow \mathbf{z}_{t-1}'' - \gamma \nabla_{\mathbf{z}_{t-1}''} \|\hat{\mathbf{z}}_0 - \mathcal{E}(\mathcal{A}^T \mathbf{y} + (\mathbf{I} - \mathcal{A}^T \mathcal{A}) \mathcal{D}(\hat{\mathbf{z}}_0))\|_2^2;$ // 流形粘合校正
- 9) return: $\mathcal{D}(\mathbf{z}_0)$ 。

引导,预测噪声 $\epsilon_\theta(z_t, z_t, t)$ 。根据 Tweedie 公式得到 $\hat{z}_0$,并通过标准扩散更新得到中间状态 z'_{t-1} 。该过程仅受扩散先验驱动,尚未施加任何观测约束。

1.3.3 测量一致性约束与流形粘合约束

推理阶段将 $\mathcal{A}(\cdot)$ 实例化为下采样算子,并令 $\tilde{y} = y$,测量一致性损失定义为

$$\mathcal{L}_{\text{meas}}(z_t) = \frac{1}{2} \|y - \mathcal{A}\mathcal{D}(\hat{z}_0)\|_2^2 \quad (13)$$

引入测量引导权重 $\eta > 0$,将该损失对 z_t 的梯度作用于 z'_{t-1} ,其计算过程可写为

$$z''_{t-1} = z'_{t-1} - \eta \nabla_{z_t} \mathcal{L}_{\text{meas}}(z_t) \quad (14)$$

该步骤增强了生成结果与 LR 观测间的测量一致性。

仅依靠测量一致性约束难以保证采样轨迹始终接近数据流形。为同时满足流形一致性与观测对齐,引入 Rout 等(2023)提出的流形粘合目标。其测量修正操作定义为

$$\tilde{x}_0 = \mathcal{A}^T y + (I - \mathcal{A}^T \mathcal{A}) \mathcal{D}(\hat{z}_0) \quad (15)$$

式中, $\mathcal{A}^T$ 表示 $\mathcal{A}$ 的伴随算子,在图像超分辨率任务中, $\mathcal{A}$ 为下采样算子, $\mathcal{A}^T$ 为对应的上采样算子。第一项将观测信息投影回像素空间,第二项保留未被 LR 观测直接约束的高频残差信息。 $\tilde{x}_0$ 可理解为同时包含观测约束信息与模型先验细节的测量修正结果,作为后续流形粘合约束的参考目标。流形粘合损失定义为

$$\mathcal{L}_{\text{glue}}(z_t) = \frac{1}{2} \|\hat{z}_0 - \mathcal{E}(\tilde{x}_0)\|_2^2 \quad (16)$$

式(16)在潜在空间中约束 $\hat{z}_0$ 与 $\mathcal{E}(\tilde{x}_0)$ 保持一致,使测量修正结果与当前潜在预测结果相匹配。引入流形粘合权重 $\gamma > 0$,最终更新结果可写为

$$z_{t-1} = z''_{t-1} - \gamma \nabla_{z_t} \mathcal{L}_{\text{glue}}(z_t) \quad (17)$$

流形粘合约束有助于协调先验去噪、测量一致性校正与潜在流形约束之间的关系,缓解不同约束项之间的潜在冲突。

2 实验

本节通过对比实验、消融实验和超参数分析验证所提方法的有效性,分别评估整体重建性能、双阶段引导模块的贡献及关键参数对重建保真度与感知质量权衡的影响。

2.1 实验配置

实验选用 DIV2K、Flickr2K、DIV8K 和 Outdoor-SceneTraining 构建 HR 训练集,验证集采用 DIV2K-Val。训练阶段采用 Real-ESRGAN (Wang X T 等, 2021)的二阶退化流程在线生成 LR 图像;测试阶段选用 RealSR (Cai 等, 2019)和 DRealSR (Wei 等, 2020)数据集,将 HR 图像中心裁剪为 512×512 的图像块,并用放大倍率为 $\times 4$ 的抗混叠双三次下采样生成 128×128 的 LR 图像。

所有实验均在配备 Intel Core i9-14900KF CPU 与单张 NVIDIA GeForce RTX 4090D GPU (24GB 显存)的硬件平台上完成。操作系统为 Windows 11,软件环境配置为 Python 3.10.18 和 CUDA 11.8,深度学习框架基于 PyTorch 2.1.0 和 PyTorch Lightning 2.5.6 实现。模型以 StableSR 预训练权重作为初始化,并在上述训练集上进行微调。训练采用 AdamW 优化器,学习率为 5×10^{-5} ,批量大小设为 2,通过 4 次梯度累积实现等效批量大小为 8 的参数更新,共训练 15000 次迭代。扩散模型采用 DDPM 调度方案,时间步 T 设为 1000。实验采用方法部分所述的双阶段梯度引导框架,其中 ρ 、 η 、 γ 分别设为 1.0、0.1 和 0.03。推理阶段采用重采样策略将扩散步从 1000 步缩减至 200 步,下采样算子 $\mathcal{A}$ 采用可微的抗混叠双三次下采样实现,固定随机种子为 42。

对比实验选取 RealESRGAN+ (Wang X T 等, 2021)、StableSR (Wang J Y 等, 2024)、ControlSR (Wan 等, 2024)和 StructSR (Li Y C 等, 2025)作为比较方法。所有方法均使用其开源代码和官方预训练权重,在相同的测试集上进行评估。

评估指标分为有参考指标和无参考指标。有参考指标包括峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity index measure, SSIM) 和感知图像块相似度 (learned perceptual image patch similarity, LPIPS)。其中, PSNR 和 SSIM 在 YCbCr 颜色空间的 Y 通道上计算,用于评价亮度保真度与结构一致性; LPIPS 基于预训练深度网络特征衡量重建图像与参考图像之间的感知差异。无参考指标包括 CLIP-IQA (CLIP image quality assessment)、MUSIQ (multi-scale image quality transformer) 和 NIQE (natural image quality evaluator),分别从视觉-语言语义先验、多尺度质量预测和自然场景统计角度评价图像感知质量。

2.2 对比实验

2.2.1 定量对比

图2以归一化的PSNR和CLIP-IQA为双轴,展示了保真度与感知质量的权衡关系。图中(1,1)表示两项指标同时最优的理想点,虚线弧是过所提方法的等距轮廓线,弧线外侧的方法距理想点更远。所提方法在两个数据集上都更接近理想点,表明其在提升重建保真度、抑制幻觉纹理的同时,有效均衡了感知质量,保真度与感知质量的权衡优于对比方法。表1的定量结果支撑了这一结论:相较于基线StableSR,所提方法的PSNR、SSIM、LPIPS效果提升,表明该方法能够增强重建结果的保真度与结构一致性;部分无参考指标虽未达到最优,但这源于扩散超分辨率中保真度和感知质量的固有权衡,即通过约束生成轨迹向输入内容靠拢,有效抑制了幻觉纹理,代价是适度牺牲了生成的感知多样性。与RealESRGAN+相比,所提方法在PSNR上略低,但在CLIP-IQA等感知质量指标上更具优势;与ControlSR和StructSR相比,所提方法虽未在所有感知指标上取得最优,但在有参考指标上表现更好。因此,所提方法的主要目标并非追求单一感知分数最优,而是在抑制幻觉纹理的前提下尽可能保持合理的感知质量与结构一致性,有参考指标的整体提升更能直接反映双阶段引导机制的实际技术贡献。

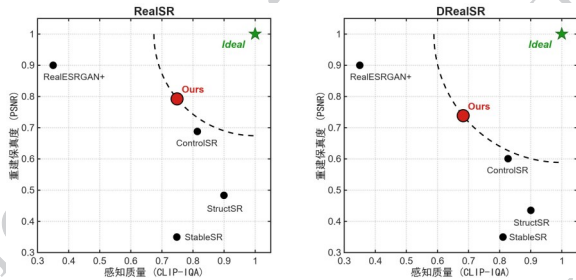


图2 保真度-感知质量权衡对比

Fig. 2 Fidelity - perception trade-off

为评估推理阶段的资源消耗,我们在相同硬件平台下统计各方法的峰值显存占用。RealESRGAN+、StableSR、ControlSR、StructSR和本文方法的峰值显存分别为0.17、7.07、11.73、7.12和24.93GB。结果表明,本文方法显存占用较高,主要由于推理阶段引入双重约束,需要构建并保留用于梯度求导的计算图。结合表1和图2可知,该方法以较高的推理阶段资源消耗为代价,提升了重建保真

度与结构一致性,并在整体保真度-感知质量权衡上表现更优。

2.2.2 定性对比

图3展示了所提方法与各对比方法在RealSR和DRealSR数据集上的视觉定性对比结果,图中额外纳入了双三次插值(Bicubic)上采样与预训练LDM(Latent Diffusion Model)(Rombach等,2022)直接推理的结果作为辅助参照。以包含字符纹理的样本Nikon_010为例,Bicubic和LDM整体较平滑,高频细节不足;RealESRGAN+、StructSR和ControlSR未能正确重建文本信息,产生了明显的字符伪影;StableSR感知质量较高,但仍存在字符形变;所提方法则较好复原了字符细节,在结构准确性与纹理真实感上表现最优。在其他包含数字或复杂纹理的样本中,所提方法生成的图像在结构清晰度和纹理真实感上与HR参考图最为接近,这一定性观察与表1中本文方法在PSNR/SSIM上的提升趋势一致,进一步验证了双阶段引导机制在抑制幻觉纹理、提升结构一致性方面的有效性。

2.3 消融实验

2.3.1 消融配置

表2展示了各模块消融实验的定量结果,其中训练阶段的频域视觉引导记为F-Guide(frequency-domain visual guidance),推理阶段的双重约束后验采样记为D-DPS(dual-constraint diffusion posterior sampling)。观察表明,完整方法在两数据集上的PSNR和SSIM均取得最优结果,说明F-Guide与D-DPS联合使用能够有效提升重建保真度。与基线方法相比,仅启用F-Guide时,PSNR、LPIPS和NIQE均有改善,表明F-Guide有助于提升重建结果与参考图像之间的感知相似性;仅启用D-DPS时,PSNR提升更为明显,说明D-DPS对增强观测一致性和提高像素级保真度具有更直接的作用。两种模块联合后进一步取得最优保真度结果,表明二者具有一定互补作用。需要注意的是,完整方法在部分无参考指标上并非最优,说明更强的保真度约束可能会影响无参考质量评价偏好的纹理表现,重建保真度与无参考感知质量之间存在权衡。总体而言,两阶段机制更有利于获得与参考图像一致的高保真重建结果,并在一定程度上兼顾保真度与感知质量。

2.3.2 超参数分析

所提方法涉及三个关键超参数:训练阶段的频

表 1 不同方法在 RealSR、DRealSR 数据集上的定量比较(4倍图像超分辨率)

Table 1 Quantitative comparison of different methods on RealSR and DRealSR datasets for ×4 image super-resolution							
数据集	方法	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-IQA ↑	MUSIQ ↑	NIQE ↓
RealSR	RealESRGAN+	27.196	0.8024	0.2093	0.5376	66.947	5.7595
	StableSR	25.014	0.7417	0.2646	0.6665	<u>70.039</u>	5.8895
	ControlSR	26.356	0.7255	0.2918	<u>0.6877</u>	71.033	5.2505
	StructSR	25.543	0.7490	0.2990	0.7156	69.755	6.7630
	本文	<u>26.770</u>	<u>0.7532</u>	<u>0.2573</u>	0.6667	67.499	<u>5.6681</u>
DRealSR	RealESRGAN+	31.906	0.8768	0.2002	0.5340	61.458	6.7209
	StableSR	28.876	0.7987	0.2830	0.6886	<u>65.803</u>	<u>6.7183</u>
	ControlSR	30.256	0.8046	0.2927	<u>0.6941</u>	67.285	6.4165
	StructSR	29.348	<u>0.8225</u>	0.3023	0.7187	65.464	7.9381
	本文	<u>31.015</u>	0.8190	<u>0.2522</u>	0.6457	61.113	6.8182

注:加粗、下划线字体表示各列最优、次优结果。↑表示数值越大越好,↓表示数值越小越好。

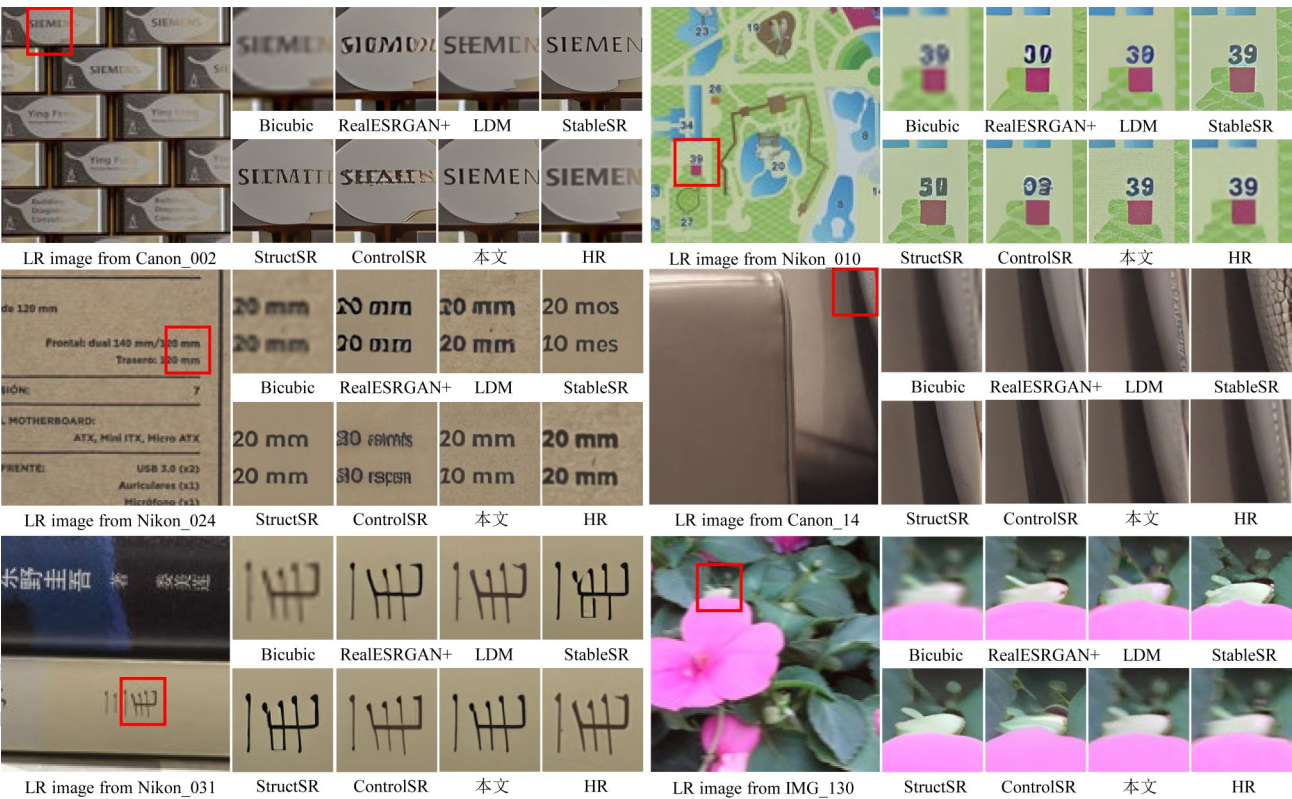


图 3 不同方法在 RealSR 和 DRealSR 数据集上的视觉效果比较

Fig. 3 Visual comparison of different methods on RealSR and DRealSR datasets

域引导权重 ρ 、推理阶段的测量引导权重 η 和流形粘合权重 γ 。由于调整 ρ 涉及模型重训,计算开销较大,实验中固定 $\rho = 1.0$ 。针对 η 与 γ ,在 DRealSR 数据集上进行单变量控制实验,超参数采用非等距选取值,以覆盖弱约束、中等约束和较强约束等不同

引导强度区间。结果如表 3 所示。

当固定 $\eta = 0.1$ 并调整 γ 时,随着 γ 增大,PSNR 持续提升,SSIM 总体呈上升趋势,而 CLIP-IQA 与 MUSIQ 呈下降趋势,表明增强流形粘合约束有助于提高重建保真度,但过强约束会压制模型的感知多

表 2 各模块消融实验结果
Table 2 Ablation study results of different modules

数据集	F-Guide	D-DPS	PSNR	SSIM	LPIPS	CLIP-IQA	MUSIQ	NIQE
RealSR	—	—	25.014	0.7417	0.2646	0.6665	70.039	5.8895
	√	—	25.589	<u>0.7466</u>	0.2555	0.6563	68.992	5.4251
	—	√	<u>26.045</u>	0.7394	0.2754	0.6833	<u>69.272</u>	5.8019
	√	√	26.770	0.7532	<u>0.2573</u>	<u>0.6667</u>	67.499	<u>5.6681</u>
DRealSR	—	—	28.876	0.7987	0.2830	0.6886	65.803	6.7183
	√	—	29.179	<u>0.7973</u>	<u>0.2701</u>	0.6663	63.938	6.5415
	—	√	<u>30.185</u>	<u>0.8003</u>	0.2855	<u>0.6864</u>	<u>64.349</u>	<u>6.6225</u>
	√	√	31.015	0.8190	0.2522	0.6457	61.113	6.8182

注:加粗、下划线字体表示各列最优、次优结果。√表示启用相应模块,—表示未启用相应模块。

样性。LPIPS 整体下降,并在 $\gamma = 0.05$ 时达到最低;但相比 $\gamma = 0.03, \gamma = 0.05$ 和 $\gamma = 0.1$ 下的 CLIP-IQA 与 MUSIQ 已出现进一步的明显下滑。因此选取 $\gamma = 0.03$ 作为保真度与感知质量之间的平衡配置。

当固定 $\gamma = 0.03$ 并调整 η 时,保真度指标随 η 增大持续提升,说明测量一致性约束有助于增强重建结果与 LR 观测之间的一致性。但随着 η 增大,CLIP-IQA 与 MUSIQ 明显下降,NIQE 整体上升,表明过大的测量引导权重会降低感知质量。综合权衡保真度指标与感知质量指标,选取 $\eta = 0.1$ 作为最终配置。

表 3 超参数 η 与 γ 分析
Table 3 Analysis of hyperparameters η and γ

η	γ	PSNR	SSIM	LPIPS	CLIP-IQA	MUSIQ	NIQE
0.1	0.005	30.561	0.8086	0.2633	0.6540	61.779	6.7387
	0.01	30.688	0.8119	0.2596	0.6518	61.620	6.7472
	0.03	31.015	0.8190	0.2522	0.6457	61.113	6.8182
	0.05	31.478	0.8367	0.2284	0.6330	60.592	6.9391
	0.1	31.829	0.8358	0.2349	0.6279	59.602	6.8973
0.01		29.870	0.8084	0.2611	0.6707	63.082	6.7837
0.05		30.745	0.8259	0.2403	0.6545	61.701	6.7861
0.1	0.03	31.015	0.8190	0.2522	0.6457	61.113	6.8182
0.2		32.335	0.8533	0.2071	0.6183	59.249	6.9078
0.5		33.713	0.8732	0.1790	0.5598	55.582	7.0209

注:固定 $\eta = 0.1$ 分析 γ 的影响,固定 $\gamma = 0.03$ 分析 η 的影响。

3 结 论

针对扩散模型图像超分辨率中容易产生幻觉纹理、结构一致性问题,提出了一种双阶段梯度引导的扩散模型图像超分辨率方法。该方法分别从训练阶段和推理阶段增强生成过程的约束能力:训练阶段利用频域视觉引导提升模型对高频结构的建模能力,推理阶段通过测量一致性与流形粘约束校正采样轨迹,从而减少生成结果与输入观测之间的结构偏差。实验结果表明,所提方法在不引入额外网络结构的前提下,能够提升重建结果的保真度,并在保真度与感知质量之间取得较好的平衡。

由于推理阶段需要在采样过程中多次计算梯度,计算开销有所增加;同时,测量一致性约束依赖已知下采样算子,因此在未知退化或复杂退化场景中的适用性仍需进一步提高。后续工作将探索更高效的梯度校正策略,进一步开展训练阶段频域视觉引导权重 ρ 的敏感性分析,并研究面向盲图像超分辨率的自适应退化建模方法。

参考文献 (References)

Cai J R, Zeng H, Yong H W, Cao Z S and Zhang L. 2019. Toward real-world single image super-resolution: a new benchmark and a new model//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3086-3095 [DOI: 10.1109/ICCV.2019.00318]

Chen J Y, Pan J S and Dong J X. 2025. FaithDiff: unleashing diffusion

- priors for faithful image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 28188-28197 [DOI: 10.1109/CVPR52734.2025.02625]
- Chung H, Kim J, McCann M T, Klasky M L and Ye J C. 2023. Diffusion posterior sampling for general noisy inverse problems//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview
- Efron B. 2011. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106 (496): 1602-1614 [DOI: 10.1198/jasa.2011.tm11181]
- Ho J, Saharia C, Chan W, Fleet D J, Norouzi M and Salimans T. 2022. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 23(47): 1-33
- Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE Computer Society: 105-114 [DOI: 10.1109/CVPR.2017.19]
- Li X Y, Wang Z R, Zou Y, Chen Z X, Ma J, Jiang Z Y, et al. 2025. DiffISR: a diffusion model with gradient guidance for infrared image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 7534-7544 [DOI: 10.1109/CVPR52734.2025.00706]
- Li Y C, Liang D, Ding T Y and Huang S J. 2025. StructSR: refuse spurious details in real-world image super-resolution//Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia, PA, USA: AAAI Press: 5022-5030 [DOI: 10.1609/aaai.v39i5.32532]
- Liang J Y, Cao J Z, Sun G L, Zhang K, Van Gool L and Timofte R. 2021. SwinIR: image restoration using Swin Transformer//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Montreal, QC, Canada: IEEE: 1833-1844 [DOI: 10.1109/ICCVW54120.2021.00210]
- Lim B, Son S H, Kim H W, Nah S J and Lee K M. 2017. Enhanced deep residual networks for single image super-resolution//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA: IEEE Computer Society: 1132-1140 [DOI: 10.1109/CVPRW.2017.151]
- Liu Z Y, Yang Y, Huang S Y and Wang S Z. 2026. Recursive super-resolution network with adaptive feature fusion. *Journal of Image and Graphics*, 31(4): 1044-1060 (刘紫阳, 杨勇, 黄淑英, 王书昭. 2026. 自适应特征融合的超分辨率重建循环网络. *中国图象图形学报*, 31(4): 1044-1060) [DOI: 10.11834/jig.250332]
- Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE: 10684-10695 [DOI: 10.1109/CVPR52688.2022.01042]
- Rout L, Raouf N, Daras G, Caramanis C, Dimakis A and Shakkottai S. 2023. Solving linear inverse problems provably via posterior sampling with latent diffusion models//Proceedings of the Advances in Neural Information Processing Systems 36. New Orleans, LA, USA: Curran Associates, Inc.: 49960-49990 [DOI: 10.52202/075280-2174]
- Saharia C, Ho J, Chan W, Salimans T, Fleet D J and Norouzi M. 2023. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4713-4726 [DOI: 10.1109/TPAMI.2022.3204461]
- Tan M K, Xu S K, Zhang S H and Chen Q. 2021. A review on deep adversarial visual generation. *Journal of Image and Graphics*, 26(12): 2751-2766 (谭明奎, 许守恺, 张书海, 陈奇. 2021. 深度对抗视觉生成综述. *中国图象图形学报*, 26(12): 2751-2766) [DOI: 10.11834/jig.210252]
- Wan Y H, Jiang P T, Hou Q B, Zhang H, Chen J W, Cheng M M, et al. 2024. ControlSR: taming diffusion models for consistent real-world image super resolution[EB/OL]. [2026-05-13]. <https://arxiv.org/abs/2410.14279v2>
- Wang J Y, Yue Z S, Zhou S C, Chan K C K and Loy C C. 2024. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132(12): 5929-5949 [DOI: 10.1007/s11263-024-02168-7]
- Wang X D, Wang P, Li Z Y and Yuan X. 2025. Plug-and-play diffusion models for inverse problems with data consistency projection//Proceedings of the IEEE International Conference on Image Processing Workshops. Anchorage, AK, USA: IEEE: 350-355 [DOI: 10.1109/ICIPW68931.2025.11386346]
- Wang X T, Xie L B, Dong C and Shan Y. 2021. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Montreal, QC, Canada: IEEE: 1905-1914 [DOI: 10.1109/ICCVW54120.2021.00217]
- Wei P X, Xie Z W, Lu H N, Zhan Z Y, Ye Q X, Zuo W M, et al. 2020. Component divide-and-conquer for real-world image super-resolution//Proceedings of the European Conference on Computer Vision. Glasgow, UK: Springer: 101-117 [DOI: 10.1007/978-3-030-58598-3_7]
- Wei Y Y, Mao T Y, Li B A, Wang F, Li F, Zhang Z, et al. 2025. Visual and large multimodal models promote image restoration and enhancement: research progress. *Journal of Image and Graphics*, 30(5): 1197-1219 (韦炎炎, 毛天一, 李柏昂, 王飞, 李锋, 张召, 等. 2025. 视觉模型及多模态大模型推进图像复原增强研究进展. *中国图象图形学报*, 30(5): 1197-1219) [DOI: 10.11834/jig.240436]

作者简介

黄京乐,女,硕士研究生,主要研究方向为图像超分辨率。E-
© 中国图象图形学报版权所有

mail:s2410w0951@hnu.edu.cn

周四望,通信作者,男,教授,博士生导师,主要研究方向为物

联网数据处理、图像处理、深度学习。 E-mail:swzhou@hnu.
edu.cn