中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-18

论文引用格式:Meng Bo, Ding Zegang. A diffusion-inversion image super-resolution method with multi-scale feature enhancement and degradation-aware sampling[J/OL]. Journal of Image and Graphics, XXXX:1-18. DOI: 10.11834/jig.260487. (孟勃, 丁泽刚. 多尺度特征增强与退化感知采样的扩散反演图像超分辨率方法[J/OL]. 中国图象图形学报,XXXX:1-18. DOI: 10.11834/jig.260487.)[DOI:10.11834/jig.260487]

多尺度特征增强与退化感知采样的扩散反演图像超分辨率方法

孟勃[1,2,3],丁泽刚[1*]

1. 东北电力大学计算机学院,吉林　132012; 2. 电力机器人视觉技术研究与应用实验室,吉林　132012; 3. 东北电力大学电力机器人实验室,吉林　132012

摘　要:**目的**　扩散反演超分辨率(diffusion inversion super-resolution,InvSR)能够通过估计扩散中间状态所需噪声,以较少反向采样步数完成图像重建,但原始噪声预测网络在高分辨率特征压缩阶段缺少对局部边缘和细粒度纹理的显式保持,瓶颈层尚未协同建模跨区域结构上下文与局部纹理连续性,固定采样路径也难以适配不同退化输入。为提高扩散反演超分辨率模型的细节保真能力和采样路径适配能力,提出一种面向InvSR的细节保真噪声预测与采样路径自调度方法。**方法**　以InvSR模型为基线,从噪声预测网络和采样策略两方面进行改进。首先,设计高频保持下采样模块(high-frequency preserving downsampling module,HPDM),在高分辨率阶段引入多尺度深度卷积编码和邻域重排下采样,使特征压缩过程尽量保留边缘响应和邻域细节。其次,在瓶颈层引入结构纹理双分支融合模块(structure-texture dual-branch fusion module,STDFM),分别建模长程结构依赖和局部纹理连续性,并通过门控机制融合两类特征。最后,构建采样路径预测模块(sampling path prediction module,SPPM),利用离线恢复效用生成路径标签,训练轻量级卷积神经网络根据输入低质量图像预测时间步组合,实现输入相关的反向扩散推理。**结果**　在ImageNet-Test数据集上,与InvSR-1相比,本文方法的SSIM(structural similarity index measure)由0.6789提高至0.6808,NIQE(natural image quality evaluator)由4.3815降低至4.0004,PI(perceptual index)由3.0866降低至2.9565,CLIP-IQA(contrastive language-image pre-training based image quality assessment)由0.7093提高至0.7274,其中NIQE和PI在所比较方法中取得最优结果;模型参数量由33.84 M降低至28.47 M,减少约15.9%。与此同时,PSNR(peak signal-to-noise ratio)由24.14 dB降至23.65 dB,LPIPS(learned perceptual image patch similarity)由0.2517升至0.2634,表明不同质量指标之间仍存在一定权衡。在RealSR(real-world super-resolution)数据集上,本文方法的PSNR由24.50 dB提高至24.71 dB,NIQE和PI分别由4.2189、3.7779降低至3.9566、3.6148,CLIP-IQA由0.6918提高至0.6966;在RealSet80数据集上,NIQE和PI分别由4.0284、3.4666降低至3.8938、3.4479,但CLIP-IQA和MUSIQ(multi-scale image quality transformer)未优于InvSR-1。**结论**　所提方法在不改变InvSR主体扩散反演流程的基础上,通过高分辨率细节保真压缩、瓶颈层结构纹理协同建模和采样路径自调度,提高了噪声预测网络的细节保持能力和不同退化图像下的推理适配能力,为低步数扩散式图像超分辨率提供了一种有效改进思路。

关键词:单幅图像超分辨率;扩散模型;扩散反演;多尺度特征增强;噪声预测网络;采样路径预测;真实世界图像复原

A diffusion-inversion image super-resolution method with multi-scale feature enhancement and degradation-aware sampling

收稿日期:2026-08-05;**修回日期**:2026-10-29

*通信作者:丁泽刚　E-mail:2110493443@qq.com

Supported by:None

© 中国图象图形学报版权所有

Meng Bo[1,2,3], Ding Zegang[1*]

1. School of Computer Science, Northeast Dianli University, Jilin 132012, China; 2. Laboratory for Research and Application of Vision Technology for Power Robots, Northeast Dianli University, Jilin 132012, China; 3. Electric Power Robot Laboratory, Northeast Dianli University, Jilin 132012, China

Abstract: Objective Diffusion inversion super-resolution (InvSR) reconstructs images with only a few reverse sampling steps by estimating the noise required for intermediate diffusion states. However, the original noise prediction network lacks explicit preservation of local edges and fine-grained textures during high-resolution feature compression. In addition, the bottleneck layer does not jointly model cross-region structural context and local texture continuity, while the fixed sampling path is difficult to adapt to inputs with different degradation characteristics. To improve the detail preservation capability and sampling-path adaptability of diffusion inversion super-resolution, a detail-preserving noise prediction and adaptive sampling-path scheduling method for InvSR is proposed. **Method** The proposed framework jointly modifies the noise-prediction network and the sampling policy. First, a high-frequency preserving downsampling module (HPDM) replaces the attention-heavy first downsampling stage. A time-conditioned residual block retains the modulation of feature responses by the diffusion timestep. Three depthwise-convolution branches with 3×3, 5×5, and dilated 3×3 kernels encode local texture, slightly wider edge context, and expanded spatial context, respectively. Their outputs are concatenated, fused by a 1×1 convolution, and added to the input feature. PixelUnshuffle followed by a 1×1 projection then rearranges each 2×2 neighborhood into the channel dimension before compression, reducing irreversible loss of local ordering information. A linear spatial-attention block is retained only in the lower-resolution stage, where long-range relations can be modeled at lower cost. Second, a structure-texture dual-branch fusion module (STDFM) is inserted into the semantic bottleneck. Its structure branch uses linear spatial attention to capture global contours and cross-region dependencies, whereas its texture branch uses multi-scale depthwise convolutions to represent local gradients, orientation, and neighborhood continuity. A time-aware gate predicts the relative contribution of the two branches from the input content and timestep embedding, enabling the bottleneck to strengthen structural constraints when global evidence is weak and to emphasize local texture when fine detail is deficient. Third, a sampling path prediction module (SPPM) converts path selection into a sixteen-class degradation-aware classification problem. The candidate set contains sixteen timestep combinations covering one to five-step sampling. During offline label construction, every candidate is evaluated for each training sample using direction-aligned and normalized fidelity, structural, perceptual, and no-reference quality measures. Their weighted restoration utility determines the supervisory path label. A tiny depthwise convolutional network (TinyDWConvNet) with progressively expanded depthwise-separable convolutional stages, global average pooling, dropout, and a sixteen-way classifier learns to infer that label directly from a 128×128 low-quality input. The classifier contains 107,217 parameters and is evaluated only once at deployment; it does not require a ground-truth image, repeated reconstruction with all candidate paths, or online computation of image-quality measures. The three components are deliberately placed at different processing scales rather than stacked at a single resolution. HPDM protects high-frequency evidence before irreversible spatial contraction; STDFM operates after compression, when global relationships can be estimated efficiently; and SPPM acts outside the denoiser to allocate the reverse trajectory. This separation keeps each component aligned with a distinct source of error and avoids changing the pretrained diffusion prior. During reconstruction, the selected path determines only the sequence of reverse timesteps, while the enhanced noise predictor supplies the state-dependent noise estimate required at those timesteps. Consequently, the same pretrained diffusion backbone and interface remain available for one to five-step inference. The structural network and the path classifier use separate objectives, so the classifier can be extended without redesigning the main diffusion backbone. **Result** Training uses the LSDIR(A Large Scale Dataset for Image Restoration) dataset together with a 20,000-image subset of FFHQ (Flickr-Faces-HQ). Low-quality inputs are synthesized with the Real-ESRGAN (Real-world Enhanced Super-Resolution Generative Adversarial Network) degradation pipeline, and the scale factor is ×4 with 512×512 high-resolution crops and 128×128 low-resolution inputs. Evaluation is conducted on ImageNet-Test, the paired RealSR (Real-World Super-Resolution) dataset, and the unpaired RealSet80 dataset. Full-reference assessment uses peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch simi-

© 中国图象图形学报版权所有

larity (LPIPS); no-reference assessment uses the Natural Image Quality Evaluator (NIQE), Perceptual Index (PI), Multi-scale Image Quality Transformer (MUSIQ), and contrastive language-image pre-training based image quality assessment (CLIP-IQA). On ImageNet-Test, compared with one-step InvSR, the complete model decreases PSNR from 24.14 dB to 23.65 dB, while increasing SSIM from 0.6789 to 0.6808. It lowers NIQE from 4.3815 to 4.0004 and PI from 3.0866 to 2.9565, while increasing CLIP-IQA from 0.7093 to 0.7274. The parameter count decreases from 33.84 million to 28.47 million, a reduction of approximately 15.9%. The HPDM-plus-STDFM configuration provides the strongest fidelity-oriented ablation result, reaching 24.59 dB PSNR and 0.6902 SSIM with 28.37 million parameters, whereas adding SPPM shifts the balance toward input-dependent perceptual quality and produces the best CLIP-IQA score. On RealSR, the complete model decreases NIQE from 4.2189 to 3.9566 and PI from 3.7779 to 3.6148 relative to InvSR-1, although its full-reference and some semantic-perceptual scores do not all improve. On RealSet80, NIQE decreases from 4.0284 to 3.8938 and PI from 3.4666 to 3.4479, while CLIP-IQA and MUSIQ remain below the baseline. Visual comparisons show clearer local edges after HPDM and more continuous structural boundaries after STDFM. **Conclusion** The results indicate that high-resolution detail-preserving compression and bottleneck-level structure-texture coordination are complementary: the former retains evidence that would otherwise be lost before deep encoding, and the latter reorganizes that evidence under a broader structural context. Degradation-aware path prediction further changes reverse diffusion from a fixed schedule into an input-dependent decision without altering the pretrained diffusion backbone. The proposed method is therefore an effective joint structural and scheduling extension for low-step diffusion-inversion super-resolution. Its advantages are most consistent for ImageNet-Test and for no-reference naturalness on real degradations. The remaining variation among fidelity, semantic, and perceptual measures also exposes the current limitations: offline utility labels depend on a finite candidate set, one image-level path cannot accommodate different degradations within the same image, and path decisions are not optimized end to end with reconstruction. Future work will investigate larger and continuous timestep spaces, region-level or patch-level scheduling, joint optimization of the classifier and super-resolution model, and validation across additional degradation types, scale factors, and application domains.

Key words: single image super-resolution; diffusion model; diffusion inversion; multi-scale feature enhancement; noise prediction network; sampling path prediction; real-world image restoration

论文引用格式：DOI:10.11834/jig.260487

0 引　言

单幅图像超分辨率(single image super-resolution,SISR)旨在根据低分辨率或退化图像恢复高分辨率图像,在老照片修复、监控视频增强、医学图像辅助分析、遥感图像解译和数字文化遗产保护等场景中具有重要应用价值。传统插值方法虽然实现简单、推理开销较低,但通常只能恢复平滑的低频结构,难以重建真实纹理和高频边缘。深度学习方法通过端到端特征学习显著提升了超分辨率性能,但真实输入往往同时受到模糊、噪声、压缩、下采样及成像设备差异等多种因素影响,其退化过程难以由单一双三次模型准确刻画。因此,如何在复杂退化条件下兼顾重建保真度、视觉真实感与推理效率,仍是图像超分辨率研究的重要问题(孙巍等,2025)。

深度超分辨率方法经历了从卷积神经网络、生成对抗网络到Transformer的持续发展。SRCNN(super-resolution convolutional neural network)、VDSR(very deep super-resolution)、EDSR(enhanced deep super-resolution network)、RDN(residual dense network)和RCAN(residual channel attention network)方法通过加深网络、残差或密集连接以及注意力机制不断提高像素级重建精度(Dong等,2016;Kim等,2016;Lim等,2017;Zhang Y等,2018a;Zhang Y等,2018b);针对深层特征冗余及利用不足问题,相关研究通过自适应多分支特征融合与循环机制,增强不同层次特征的利用和细节重建能力(刘紫阳等,2026);SRGAN(super-resolution generative adversarial network)、ESRGAN(enhanced super-resolution generative adversarial network)、BSRGAN(blind super-resolution generative adversarial network)和Real-ESRGAN(real-world enhanced super-resolution generative adversarial network)进一步利用对抗学习及复杂退化建模改善感知质量和真实场景泛化能力

© 中国图象图形学报版权所有

(Ledig等,2017;Wang X等,2018;Zhang K等,2021;Wang X等,2021)。随后,SwinIR(Swin Transformer for Image Restoration)、Restormer和HAT(Hybrid Attention Transformer)借助自注意力增强长距离依赖建模能力(Liang等,2021;Zamir等,2022;Chen等,2023)。然而,上述判别式或对抗式方法对缺失纹理的生成能力仍受到训练数据和退化模型限制,在严重未知退化条件下容易出现过度平滑或伪纹理。

扩散模型凭借强大的生成先验为真实世界图像超分辨率提供了新的技术路线。SR3(super-resolution via repeated refinement)将条件扩散用于图像逐步细化,StableSR、DiffBIR和SeeSR等进一步利用预训练文本到图像扩散模型中的语义与纹理先验,提高复杂真实退化下的恢复能力(Saharia等,2023;Wang J等,2024;Lin等,2024;Wu等,2024a;Rombach等,2022)。但传统扩散模型通常依赖数十甚至数百次迭代去噪,推理开销较高。为此,ResShift通过残差迁移缩短扩散链,SinSR(Diffusion-Based Image Super-Resolution in a Single Step)和OSEDiff(One-Step Effective Diffusion Network for Real-World Image Super-Resolution)进一步将扩散式超分辨率压缩到单步推理(Yue等,2023;Wang Y等,2024;Wu等,2024b),推动扩散超分辨率由高质量生成向低步数高效恢复发展。

近年来,高效扩散超分辨率进一步向一步蒸馏、流模型以及灵活时间步控制方向发展。PiSA-SR(pixel-level and semantic-level adjustable super-resolution)通过像素级与语义级双LoRA(low-rank adaptation)分别调节失真去除和语义细节生成,实现可控的一步真实世界超分辨率(Sun等,2025);TSD-SR(target score distillation for super-resolution)引入目标分数蒸馏和分布感知采样,在单步推理条件下强化真实细节恢复(Dong等,2025)。FluxSR基于流匹配模型设计流轨迹蒸馏,将多步生成过程压缩为一步恢复(Li等,2025);FlowSR则将低分辨率到高分辨率的恢复过程建模为一致性校正流,并利用快—慢时间调度提高一步推理的效率与稳定性(Xu等,2025)。另一方面,InvSR通过部分噪声预测构造扩散中间状态,使模型能够以1~5个反向步骤进行灵活恢复(Yue等,2025),表明采样时间步本身也是影响低步数扩散超分辨率性能的重要因素。

最近的研究进一步关注时间步利用、生成质量优化及一步模型的可控性。TADSR(time-aware one-step diffusion network for super-resolution)针对固定时间步难以充分利用不同扩散阶段生成先验的问题,引入时间感知编码与时间感知变分分数蒸馏,使单步模型能够依据时间步条件调节保真度与真实感(Zhang T等,2026);GDPO-SR(group direct preference optimization for one-step generative image super-resolution)将组直接偏好优化引入一步生成式超分辨率,通过在线样本相对优势和区域属性感知奖励进一步改善纹理与平滑区域的恢复质量(Yi等,2026)。这些研究表明,近期扩散超分辨率已经从单纯减少采样次数逐渐转向对扩散时间、采样过程和恢复偏好的精细控制。

尽管现有方法显著提高了扩散超分辨率的推理效率,但仍存在进一步优化空间。一方面,低步数推理对中间状态预测和局部细节表征更加敏感,高分辨率特征在连续压缩过程中容易损失边缘和细粒度纹理;另一方面,现有一步或少步方法的推理时间步、采样步数或时间步组合大多采用预先设定或人工指定的全局策略。不同低质量图像具有明显不同的模糊程度、噪声水平、高频衰减和纹理复杂度,统一采样路径难以同时兼顾轻度与重度退化输入。针对输入退化特征自动判断所需反向扩散路径,使简单样本避免冗余计算、复杂样本获得充分恢复,仍有进一步研究空间。

针对上述问题,本文以InvSR为基线,从特征保持、结构纹理建模和采样策略3个方面进行改进,提出多尺度特征增强与退化感知采样的扩散反演图像超分辨率方法。主要贡献如下:

1)构建高频保持下采样模块(high-frequency preserving downsampling module,HPDM),通过多尺度深度卷积与像素反重排增强高分辨率特征压缩过程中的局部边缘和细粒度纹理保持能力;

2)设计结构纹理双分支融合模块(structure-texture dual-branch fusion module,STDFM),在瓶颈层分别建模长程结构依赖和局部纹理连续性,并通过门控机制进行自适应融合;

3)提出采样路径预测模块(sampling path prediction module,SPPM),依据低质量输入的退化特征预测候选时间步组合,将固定反向扩散过程转化为输入相关的采样路径选择。

© 中国图象图形学报版权所有

1 算法描述

1.1 基线模型与整体结构

InvSR的核心思想是把低分辨率输入映射到预训练扩散模型的中间状态，使后续采样无须从纯高斯噪声开始。设低质量输入图像为$\boldsymbol{y}_0$，高质量目标图像为$\boldsymbol{x}_0$，噪声预测网络f_w接收退化图像及扩散时间步t，输出构造中间状态所需的噪声估计。相较从随机噪声逐步去噪的方法，InvSR将采样起点前移到更接近目标图像的位置，因而可用较少反向步数完成重建。

原始噪声预测网络由输入卷积、下采样块、中间块和输出卷积构成。对于128×128到512×512的×4超分辨率，浅层高分辨率特征主要承担边缘、纹理和局部结构提取；在该阶段堆叠空间注意力会增加开销，普通下采样又可能损失高频信息。所提网络以HPDM重构第1个下采样阶段，在较低分辨率阶段保留线性注意力，并在语义瓶颈处以STDFM融合结构与纹理证据。

总体流程如图1所示。输入经初始卷积得到特征$\boldsymbol{F}_0$，第1阶段经HPDM得到$\boldsymbol{F}_1$，第2阶段以线性注意力和卷积前馈块得到$\boldsymbol{F}_2$，瓶颈层经STDFM产生$\boldsymbol{F}_m$，最后由输出卷积预测噪声图。SPPM在扩散采样前根据低质量图像的退化特征选择候选时间步组合。该设计使高分辨率阶段侧重边缘与纹理保持，瓶颈层侧重结构与纹理融合，推理阶段侧重输入相关的采样路径适配。

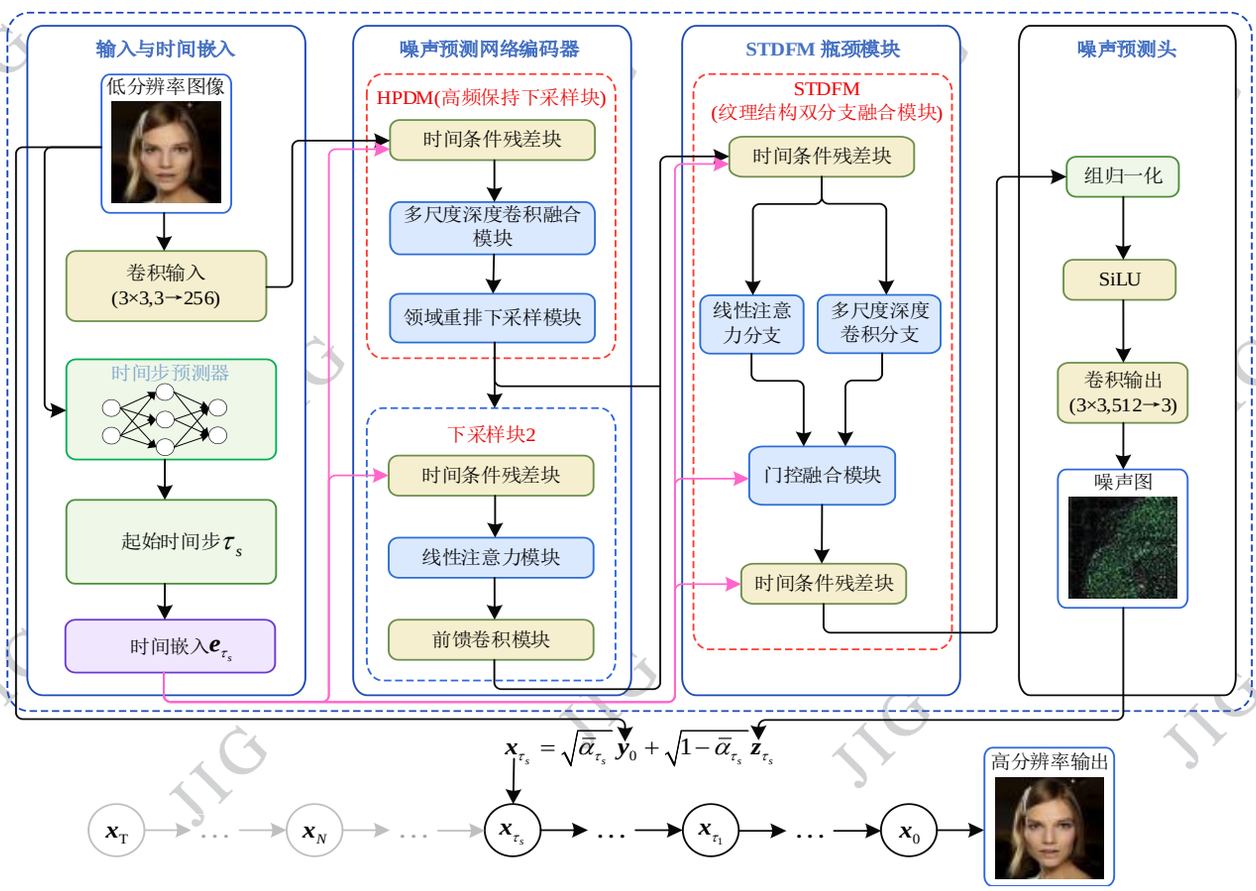


图1 改进InvSR超分辨率方法总体框架

Fig. 1 Overall framework of the improved InvSR super-resolution method

1.2 高频保持下采样模块

原始高分辨率下采样块采用“残差块—注意力—下采样”的堆叠形式。为在降低空间分辨率时显式保存局部边缘和细粒度纹理，HPDM由时间条件残差块、多尺度深度卷积融合及像素反重排下采样组成，其结构如图2所示。

时间条件残差块保持扩散时间步对特征分布的调控能力。该块以残差学习为基础(He等，2016)，

© 中国图象图形学报版权所有

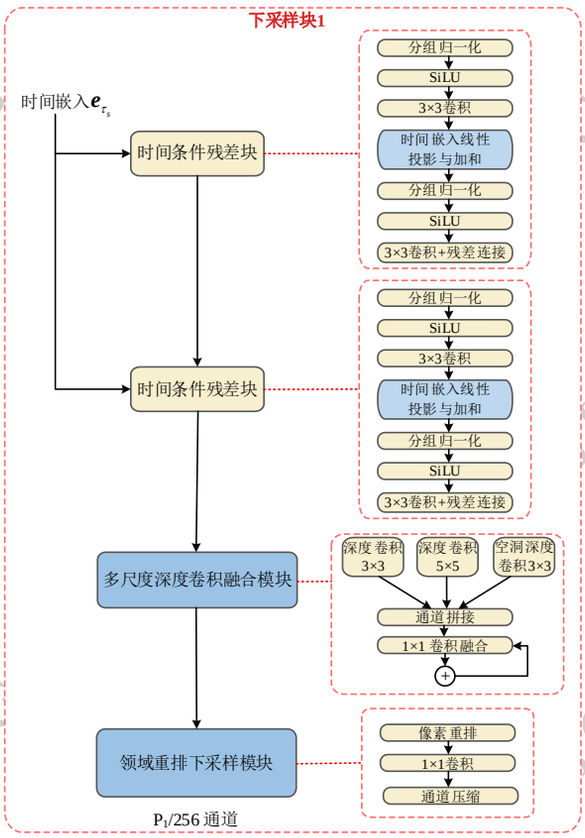


图2　高频保持下采样模块结构

Fig. 2　Architecture of the high-frequency preserving downsampling module

并采用条件调制把时间嵌入注入特征变换(Perez等,2018)。设输入特征为$\boldsymbol{F}$,时间嵌入为$\boldsymbol{e}_{\tau_s}$,并令$\boldsymbol{A}=\mathrm{Conv}_1\left(\sigma\left(\mathrm{GN}(\boldsymbol{F})\right)\right)$,则残差变换表示为

$$\boldsymbol{F}'=\boldsymbol{F}+\mathrm{Conv}_2\left(\sigma\left(\mathrm{GN}\left(\boldsymbol{A}+\varphi\left(\boldsymbol{e}_{\tau_s}\right)\right)\right)\right) \tag{1}$$

式中,$\boldsymbol{A}$为第一层卷积输出特征;GN表示group normalization;σ表示SiLU激活函数(sigmoid linear unit);其中,$\varphi\left(\boldsymbol{e}_{\tau_s}\right)$表示由时间嵌入映射得到的调制项;$\mathrm{Conv}_1$和$\mathrm{Conv}_2$表示残差支路中的卷积映射。该形式使时间步信息直接参与中间特征响应调整,从而保持噪声估计与扩散状态的一致性。

多尺度深度卷积融合用于提取不同感受野下的局部结构。深度可分离卷积可显著降低参数量和计算量(Howard等,2017;Sandler等,2018)。3×3深度卷积分支提取基础纹理,5×5分支覆盖稍大范围边缘,膨胀率为2的3×3分支补充扩展上下文。令3路分支输出分别为$\boldsymbol{F}_3$、$\boldsymbol{F}_5$和$\boldsymbol{F}_d$,随后沿通道维拼接,经1×1卷积融合并与输入残差相加:

$$\begin{aligned}&\boldsymbol{F}_3=\mathrm{DW}_{3\times3}(\boldsymbol{F}),\quad \boldsymbol{F}_5=\mathrm{DW}_{5\times5}(\boldsymbol{F}),\\&\boldsymbol{F}_d=\mathrm{DW}_{3\times3,d=2}(\boldsymbol{F}),\\&M(\boldsymbol{F})=\mathrm{Conv}_{1\times1}\left(\left[\boldsymbol{F}_3,\boldsymbol{F}_5,\boldsymbol{F}_d\right]\right),\boldsymbol{F}_{\mathrm{out}}=\boldsymbol{F}+M(\boldsymbol{F}).\end{aligned} \tag{2}$$

像素反重排下采样以PixelUnshuffle和1×1卷积替代步长卷积。对$\boldsymbol{F}\in\mathrm{R}^{C\times H\times W}$,2×2邻域先重排为$\boldsymbol{F}_u\in\mathrm{R}^{4C\times H/2\times W/2}$,再映射到目标通道数。空间邻域的排列关系被显式保存在通道维,可减少直接降采样造成的高频信息丢失。第2个下采样阶段的分辨率已经降低,因此保留一个线性复杂度空间注意力模块,以补充长距离依赖。

1.3　结构纹理双分支融合模块

中间块位于编码器最深层,空间分辨率较低而语义抽象程度较高,适合整合整体结构与局部纹理。STDFM在该位置并行设置结构分支和纹理分支,并通过时间感知门控进行融合,结构如图3所示。结构分支采用线性空间注意力建模整体轮廓、长程依赖和跨区域上下文;纹理分支采用多尺度深度卷积提取局部梯度、纹理方向和邻域连续性。两分支均保持输入特征的空间尺寸和通道数不变。线性空间注意力分支的输入、输出均为$B\times C\times H\times W$。

设结构分支输出为$\boldsymbol{F}_{\mathrm{g}}$,纹理分支输出为$\boldsymbol{F}_{\mathrm{l}}$,$\boldsymbol{F}_{\mathrm{g}},\boldsymbol{F}_{\mathrm{l}}\in\mathrm{R}^{B\times C\times H\times W}$,其中$B$为批量大小,$C$为通道数,$H$和$W$为空间尺寸。当前网络配置下中间层通道数$C=512$。两分支沿通道维拼接得到$[\boldsymbol{F}_{\mathrm{g}},\boldsymbol{F}_{\mathrm{l}}]\in\mathrm{R}^{B\times2C\times H\times W}$,并经1×1卷积映射$C$通道。扩散时间步$t$首先经正弦时间投影得到256维表示,再通过时间嵌入模块映射$\boldsymbol{e}_{\tau_s}\in\mathrm{R}^{B\times512}$,在门控模块中$\tau_s$进一步经线性映射$W_t\in\mathrm{R}^{C\times512}$投影至$C$维,并重塑为$B\times C\times1\times1$,随后沿空间维广播。时间投影维度为256,时间嵌入维度为512。

因此门控权重表示为

$$\boldsymbol{G}=\mathrm{sigmoid}\left(\mathrm{Conv}_{1\times1}\left(\left[\boldsymbol{F}_{\mathrm{g}},\boldsymbol{F}_{\mathrm{l}}\right]\right)+\mathrm{reshape}\left(\boldsymbol{e}_{\tau_s}\boldsymbol{W}_t^{\mathrm{T}}\right)\right) \tag{3}$$

其中$\boldsymbol{G}\in\mathrm{R}^{B\times C\times H\times W}$。因此$\boldsymbol{G}$不是单一标量或仅含通道信息的权重,而是随通道和空间位置变化的门控权重图。门控值越接近1,结构分支贡献越大;越接近0,纹理分支贡献越大。融合特征表示为

$$\boldsymbol{F}_{\mathrm{f}}=\mathrm{Conv}_{1\times1}\left(\boldsymbol{G}\odot\boldsymbol{F}_{\mathrm{g}}+(1-\boldsymbol{G})\odot\boldsymbol{F}_{\mathrm{l}}\right) \tag{4}$$

式中,$\odot$表示逐元素乘法;$\boldsymbol{F}_{\mathrm{f}}\in\mathrm{R}^{B\times C\times H\times W}$为融合输出。随后将$\boldsymbol{F}_{\mathrm{f}}$与进入双分支前的中间特征进行残差相加,再经时间条件残差块进一步更新,从而在保持原有特征信息的同时完成结构与纹理的自适应融

© 中国图象图形学报版权所有

合。门控根据输入内容和扩散时间步调节两类特征的贡献，使瓶颈层在结构缺失明显时增强全局约束，在局部纹理不足时提高纹理分支权重。HPDM 先在特征压缩前保存高频证据，STDFM 再在瓶颈层整合结构与纹理，二者共同构成多尺度特征增强的噪声预测网络。

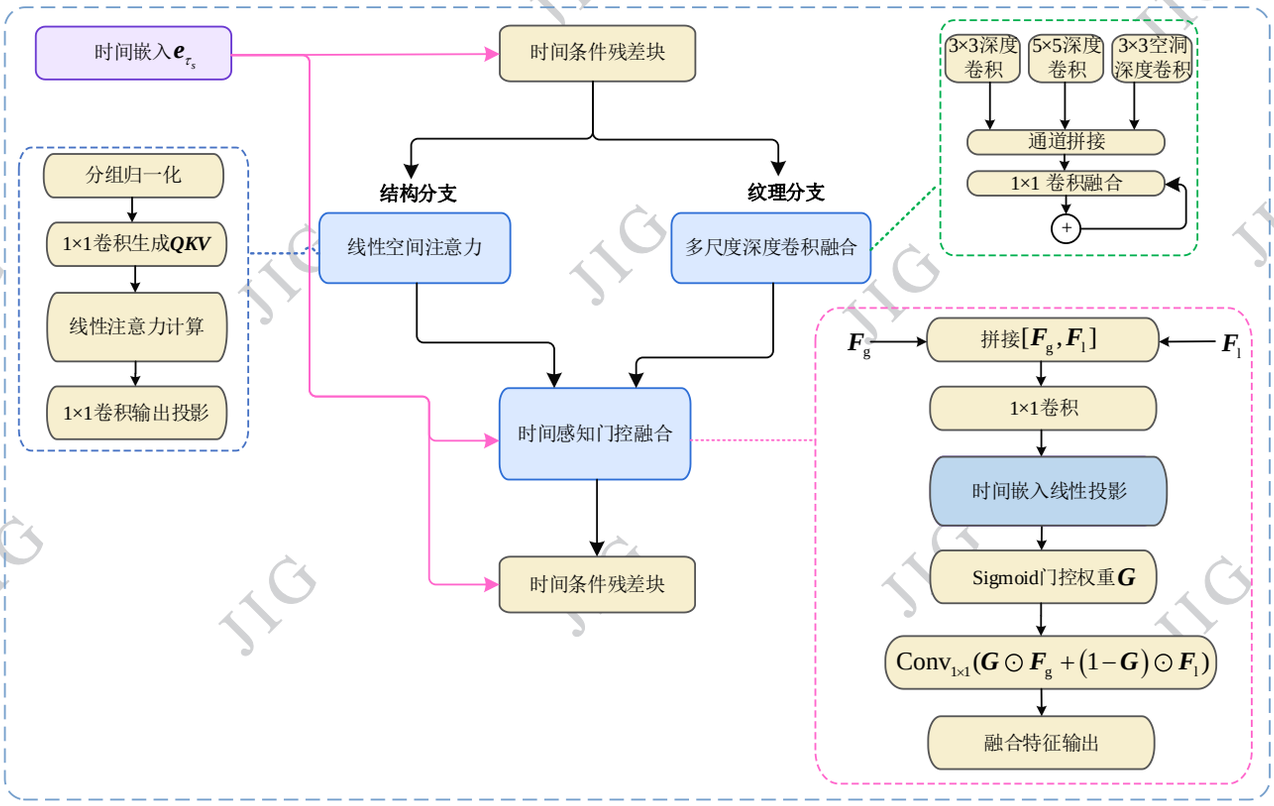


图3　结构纹理双分支融合模块结构

Fig. 3　Architecture of the structure-texture dual-branch fusion module

1.4　采样路径预测模块

InvSR 支持不同时间步组合下的扩散反演采样，但固定采样路径难以适配不同退化输入。退化较轻且结构完整的图像采用较长路径可能产生冗余计算，模糊、压缩伪影或高频缺失严重的图像采用过短路径又可能恢复不足。SPPM 把采样路径选择建模为输入退化特征驱动的分类问题，结构和路径预测流程如图4所示。

图像退化表征通常与边缘清晰度、梯度能量、高频衰减和纹理复杂度相关（Pech-Pacheco 等，2000；Pertuz 等，2013）。为覆盖不同退化复杂度，对不同采样步数及时间步组合进行预实验。候选路径覆盖一步至五步，采样候选路径集合定义为

$$\begin{aligned}S=\{&[100],[150],[200],[250],\\&[150,100],[200,100],[200,150],\\&[150,100,50],[200,100,50],[200,150,100],\\&[250,100,50],[250,150,50],[250,200,150],\\&[200,150,100,50],[250,200,150,100],\\&[250,200,150,100,50]\}\end{aligned}\tag{5}$$

分类网络 C_θ 以 128×128 低质量图像 y_i 为输入，输出16类路径概率

$$p_i=C_\theta(y_i),\quad p_i\in\mathbf{R}^{16}\tag{6}$$

网络采用 TinyDWConvNet（tiny depthwise convolutional network）：输入卷积把通道数由3扩展至24，随后以深度可分离卷积依次扩展为48、96、160和224通道，经自适应平均池化、丢弃率为0.2的Dropout及16分类全连接层输出结果，参数量为107 217。预测类别为

$$\hat{k}_i=\underset{k}{\mathrm{argmax}}\,p_{i,k}\tag{7}$$

© 中国图象图形学报版权所有

式中，$\hat{k}_i$为样本的预测路径类别，$p_{i,k}$为第k类候选路径的预测概率。

训练阶段离线构造路径监督标签。采用PSNR（peak signal-to-noise ratio）、SSIM（structural similarity index measure）、LPIPS（learned perceptual image patch similarity）、NIQE（natural image quality evaluator）、PI（perceptual index）、CLIP-IQA（contrastive language-image pre-training based image quality assessment）和MUSIQ（multi-scale image quality transformer）7项指标构建综合恢复效用，其中PSNR、SSIM、CLIP-IQA和MUSIQ为正向指标，LPIPS、NIQE和PI为负向指标。对第m项指标q的方向进行统一后，采用极差归一化：

$$r_{i,k}^{(m)}=\frac{q_{i,k}^{(m)}-\min\limits_{j}q_{i,j}^{(m)}}{\max\limits_{j}q_{i,j}^{(m)}-\min\limits_{j}q_{i,j}^{(m)}+\varepsilon} \tag{8}$$

式中，$q_{i,k}^{(m)}$为方向统一后的指标值，$r_{i,k}^{(m)}$为第i个样本采用第k条路径时第m项指标的归一化效用，ε为防止分母为0的微小常数。该归一化用于消除不同指标量纲及变化范围差异，使综合效用反映各路径在不同质量维度上的相对优劣。

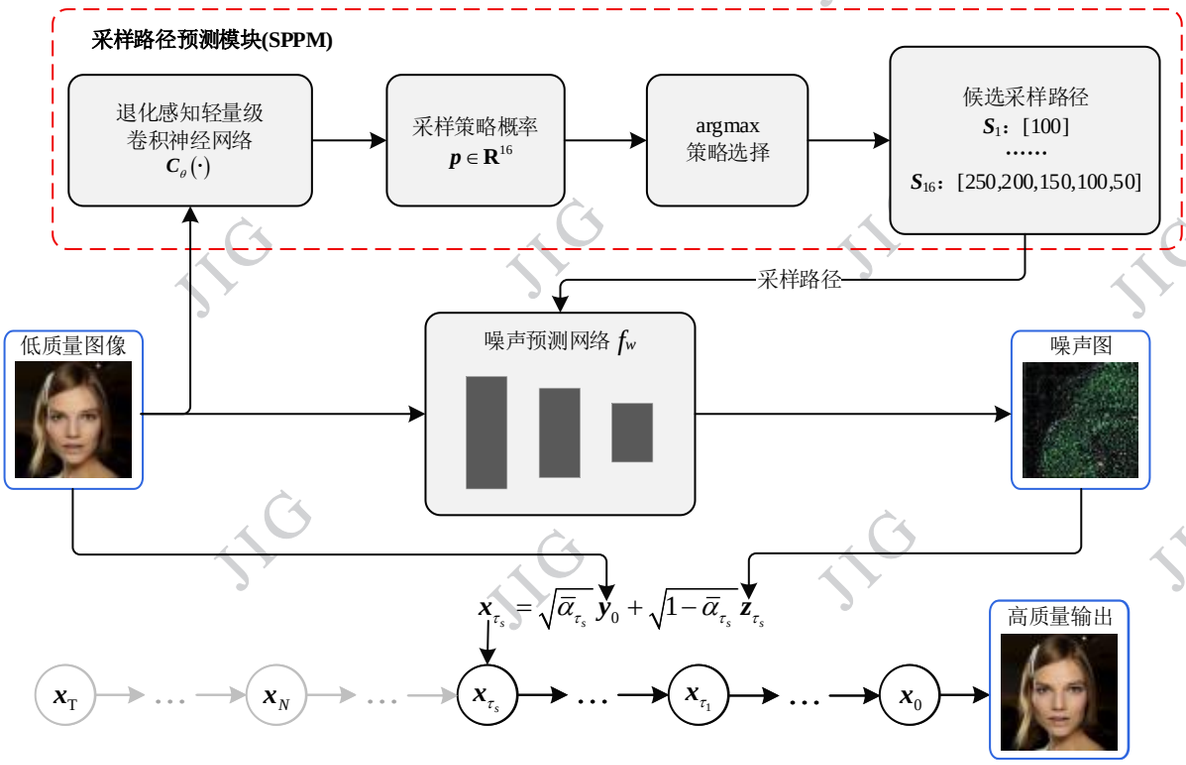


图4　采样路径预测与扩散反演推理流程

Fig. 4　Sampling path prediction and diffusion-inversion inference pipeline

指标权重由验证集确定。首先采用HPDM与STDFM联合生成的7项指标等权综合评价全部候选固定路径，并确定平均综合效用最高的[100]为参考固定路径。随后以原始InvSR为性能基准，计算[100]在各指标上的方向统一相对提升率：

$$\Delta_m=\begin{cases}\dfrac{Q_m^{[100]}-Q_m^{\mathrm{InvSR}}}{Q_m^{\mathrm{InvSR}}}, m\in\boldsymbol{M}^+\\ \dfrac{Q_m^{\mathrm{InvSR}}-Q_m^{[100]}}{Q_m^{\mathrm{InvSR}}}, m\in\boldsymbol{M}^-\end{cases} \tag{9}$$

其中，$\boldsymbol{M}^+$和$\boldsymbol{M}^-$分别表示正向和负向指标集合。为提高对固定路径下改善不足的关注，对7项提升率进行归一化并采用平滑反向赋权：

$$\hat{\Delta}_m=\frac{\Delta_m-\min\limits_{j}\Delta_j}{\max\limits_{j}\Delta_j-\min\limits_{j}\Delta_j+\varepsilon},w_m=\frac{2-\hat{\Delta}_m+\varepsilon}{\sum\limits_{j=1}^{7}(2-\hat{\Delta}_j+\varepsilon)} \tag{10}$$

由验证集得到最终权重分别为：PSNR 0.1689、SSIM 0.1716、LPIPS 0.1862、NIQE 0.1049、PI 0.0931、CLIP-IQA 0.1473和MUSIQ 0.1281，各权重之和为1。最终，第i个样本在第k条路径上的综合恢复效用及监督类别定义为

© 中国图象图形学报版权所有

$$U(\boldsymbol{y}_i,\boldsymbol{S}_k)=\sum_{m=1}^{7}w_m r_{i,k}^{(m)},\ \sum_{m=1}^{7}w_m=1$$
$$k_i^*=\underset{k\in\{1,\ldots,16\}}{\operatorname{argmax}}\,U(\boldsymbol{y}_i,\boldsymbol{S}_k). \tag{11}$$

式中，k_i^*对应的独热标签训练分类网络，推理阶段仅需对低质量输入执行一次分类，根据预测类别选择对应采样路径并传入InvSR，无须GT(ground truth)、遍历候选路径或在线计算图像质量指标，从而实现退化相关的自适应采样。由此，SPPM以较小额外开销实现不同退化输入的差异化调度。通过上述设计，采样时间步由固定经验参数转化为输入相关的动态决策。退化较轻的图像可采用较短路径以减少冗余计算，结构模糊、纹理缺失或伪影较强的图像则可选择更充分的反向扩散过程。

2 实验与分析

2.1 数据集与实验环境

训练数据采用LSDIR(A Large Scale Dataset for Image Restoration)大规模图像复原数据集(Li等，2023)和FFHQ(Flickr-Faces-HQ)20 000幅人脸图像子集(Karras等，2019)。测试包括合成退化与真实退化数据：ImageNet-Test由ImageNet验证集构造(Deng等，2009)；RealSR(real-world super-resolution)提供成对的低质量图像和高质量真值(Cai等，2019)；RealSet80无真值，仅用于无参考评价。低质量输入通过Real-ESRGAN(Real-world Enhanced Super-Resolution Generative Adversarial Network)退化管线(Wang X等，2021)合成，包含模糊、下采样、噪声、JPEG(Joint Photographic Experts Group)压缩及高阶组合。验证集从ImageNet training split随机抽取1000张统一裁剪/缩放到512×512。

表1 数据集与实验环境设置

Table 1 Datasets and experimental settings

项目	设置
训练数据	LSDIR + FFHQ 20 000幅图像子集
退化方式	Real-ESRGAN degradation pipeline
超分倍率	×4
裁剪尺寸	HR 512×512；LR 128×128
运行环境	Python 3.10；PyTorch 2.4.0；CUDA 12.1
硬件平台	4块NVIDIA A100 GPU

2.2 评价指标

全参考指标包括峰值信噪比(PSNR)(Huynh-Thu和Ghanbari，2008)、结构相似性(SSIM)(Wang Z等，2004)和学习感知图像块相似性(LPIPS)(Zhang R等，2018)；无参考指标包括自然图像质量评价(NIQE)(Mittal等，2013)、感知指数(PI)(Blau等，2018)、多尺度图像质量评价(MUSIQ)(Ke等，2021)及基于对比语言—图像预训练模型的图像质量评价(CLIP-IQA)(Wang J等，2023)。PSNR和SSIM在YCbCr (luma, blue-difference chroma and red-difference chroma)空间的Y通道计算，其余指标在RGB(red-green-blue)/sRGB(standard RGB)空间计算。PSNR、SSIM、CLIP-IQA和MUSIQ越大越好，LPIPS、NIQE和PI越小越好。

2.3 训练与推理流程

训练阶段，从高分辨率图像(high-resolution image，HR)中随机裁剪512×512图像块，并利用Real-ESRGAN退化模型依次施加模糊、缩放、噪声和压缩等操作，生成128×128低分辨率图像(low-resolution image，LR)，以模拟真实场景中复杂且混合的退化过程。为适配潜空间扩散模型输入，低质量图像先经双三次插值上采样至目标尺寸，再送入VAE(variational autoencoder)编码器得到干净潜变量；随后依据随机采样的扩散时间步加入高斯噪声，构造含噪潜变量，并输入SD U-Net(Stable Diffusion U-Net)获得预测潜变量。同时，高分辨率图像经同一VAE编码器得到真实潜变量，作为重建监督。低质量图像用于提供退化条件，低质量图像+扩散时间步二者共同输入包含HPDM和STDFM的噪声预测网络，以估计扩散反演所需噪声。

预测结果与真实潜变量之间计算扩散损失和LPIPS感知损失，利用判别器比较真实结果与预测结果并形成判别损失。各项损失联合反向传播，用于更新噪声预测网络及判别模块参数。训练采用AdamW(adaptive moment estimation with decoupled weight decay)优化器(Loshchilov和Hutter，2019)，前2000次迭代进行学习率预热，随后学习率保持为5×10^{-5}，共训练10万次迭代。损失函数由重建损失、LPIPS感知损失和GAN(generative adversarial network)对抗损失组成，重建损失权重为1.0，LPIPS损失权重为2.0，GAN损失权重为0.1。预训练SD U-Net、VAE及LPIPS网络冻结，仅更新噪声预测网

© 中国图象图形学报版权所有

络和判别器。训练过程中从{250, 200, 150, 100}中随机选择扩散起始时间步，使模型能够适应不同反演位置，同时保证结构改造前后的数据处理、退化方式和优化协议保持一致，整体流程如图5所示。

推理阶段，首先将低质量图像输入采样路径预测网络，由SPPM根据输入退化特征输出对应的路径类别；随后将输入图像送入改进噪声预测网络，得到相应扩散中间状态的噪声估计；最后在预训练扩散模型中按照预测得到的时间步组合执行反向采样，并经VAE解码生成高分辨率结果。与固定采样路径相比，该流程能够针对不同输入退化程度选择不同反向扩散轨迹。由此，模型在结构上兼顾高频细节保持、局部纹理连续性和全局语义表达，在推理阶段则依据输入特征自适应分配采样计算量，在保持InvSR低步数推理优势的同时提高采样策略的灵活性与适配性。

2.4 消融实验

为验证3个改进模块的独立贡献及组合效果，在ImageNet-Test上进行消融实验。Baseline表示原始InvSR-1；“HPDM”“STDFM”和“SPPM”列中的“√”表示启用对应模块。表2列出全部8种组合。

单独加入HPDM后，参数量由33.84M降至26.24M，PSNR由24.14dB升至24.43dB，NIQE由4.3815降至4.3348，说明邻域重排和多尺度局部编码能够在减少参数的同时改善像素重建精度与自然统计特性。

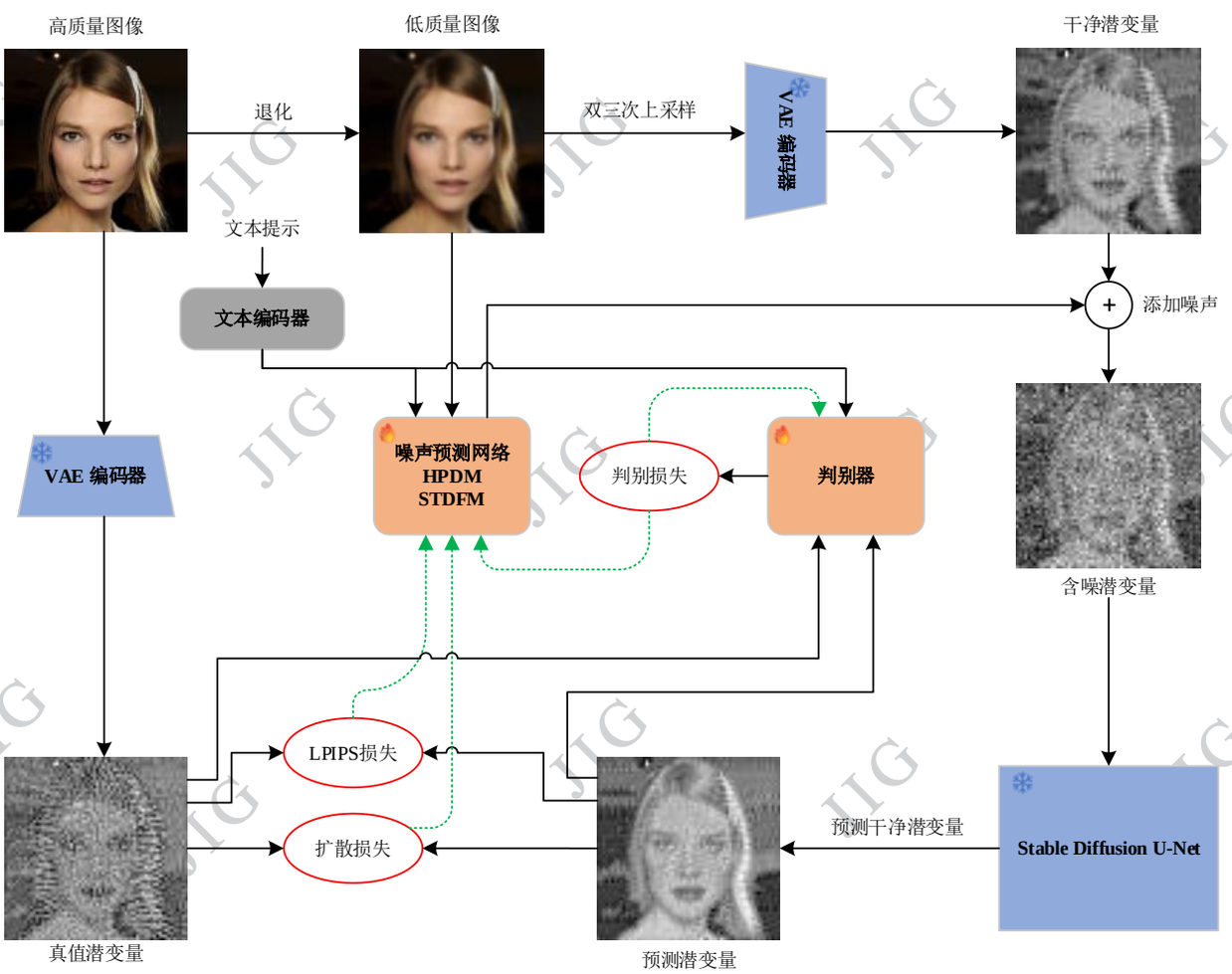


图5 多尺度特征增强网络训练流程

Fig. 5 Training pipeline of the multi-scale feature-enhanced network

单独加入STDFM仅使PSNR小幅提高，其余指标下降，表明瓶颈融合依赖前端保留足够的高频证据。HPDM与STDFM联合时获得最强的保真度结果：PSNR为24.59dB，SSIM为0.6902，NIQE为4.1441，PI为2.9613，参数量为28.37M。完整模型加入SPPM后，CLIP-IQA达到0.7274，SSIM、NIQE和

© 中国图象图形学报版权所有

表2　ImageNet-Test数据集上的消融实验结果

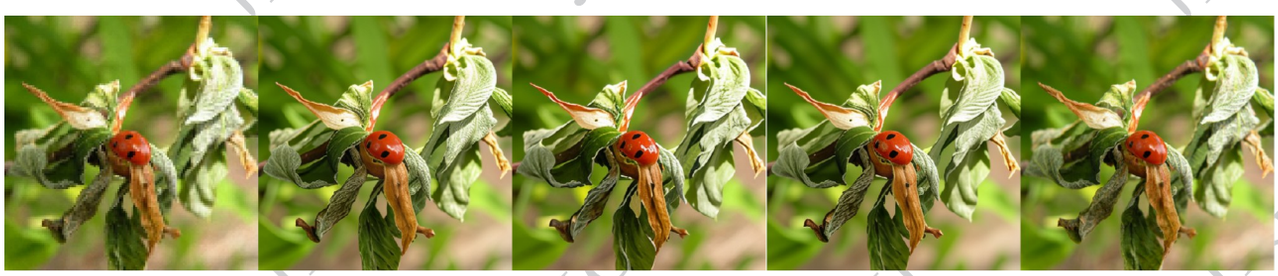
Table 2　Ablation study on the ImageNet-Test dataset

编号	HPDM	STDFM	SPPM	PSNR/dB ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PI ↓	CLIP-IQA ↑	MUSIQ ↑	参数量/M
1				24.14	0.6789	0.2517	4.3815	3.0866	0.7093	**72.2900**	33.84
2	√			24.43	0.6742	0.2869	4.3348	3.1358	0.6883	69.1270	26.24
3		√		24.21	0.6717	0.2844	4.4296	3.2762	0.6727	68.4051	28.35
4			√	<u>24.46</u>	<u>0.6854</u>	**0.2493**	4.2920	2.9746	<u>0.7138</u>	71.0065	33.95
5	√	√		**24.59**	**0.6902**	<u>0.2512</u>	<u>4.1441</u>	<u>2.9613</u>	0.7079	70.2591	28.37
6	√		√	24.20	0.6685	0.2876	4.4161	3.1473	0.6999	69.8088	26.35
7		√	√	23.98	0.6666	0.2864	4.5416	3.3050	0.6856	69.0989	28.46
8	√	√	√	23.65	0.6808	0.2634	**4.0004**	**2.9565**	**0.7274**	<u>71.0955</u>	28.47

注：粗体表示该指标的最优结果，下划线表示次优结果。

PI优于Baseline，但PSNR由24.14dB降至23.65dB，LPIPS由0.2517升至0.2634，说明SPPM更侧重输入相关的感知质量与自然度适配，而不同质量指标之间存在一定权衡。

图6从左至右给出LR、InvSR、InvSR+HPDM、InvSR+HPDM+STDFM和GT。HPDM使局部边缘与纹理响应更加清晰，继续加入STDFM后结构边界连续性和局部纹理协调性进一步改善。



(a)LR (b)InvSR (c)InvSR+HPDM (d)InvSR+HPDM+STDFM (e)GT

图6　HPDM与STDFM协同作用的可视化比较

Fig. 6　Visual comparison of the complementary effects of HPDM and STDFM

2.5　采样路径对比实验

为验证SPPM路径预测有效性及固定采样基线公平性，在验证集比较16条候选固定路径，覆盖1～5步及不同时间组合。采用7项指标评价，统一方向并归一化后等权计算综合效用。SPPM验证集分类准确率为87.3%。ImageNet-Test最优路径中，[100]占比46.47%；[150]、[200]和[250]分别占比9.00%、8.47%和5.13%；[250,200,150,10050]占比16.37%，其余路径占比较低。随后在ImageNet-Test比较最佳固定、随机、SPPM预测和Oracle四种策略。[100]综合效用最高，固定用于测试；随机路径逐图随机选取并取平均；SPPM仅依据低质量输入预测，不使用GT；Oracle利用GT遍历候选路径并选取最高综合效用结果，作为候选路径集合下的理论上限。该对比在统一候选集合、评价指标与测试条件下开展，可避免采样步数和路径设置差异造成的基线偏差，衡量路径预测对不同退化样本的适应能力。通过固定、随机、预测和理论最优策略对照，可全面分析所提模块在路径选择准确性、恢复效用及推理策略合理性方面的实际作用。

由表4可见，随机路径整体性能相对较低，说明无差别的路径选择难以稳定适配不同退化输入。最佳固定路径相比随机策略取得更好的整体结果，但其对所有图像采用相同采样轨迹，适应性仍然有限。SPPM预测路径在多数指标上优于最佳固定路径，并进一步接近Oracle路径的性能，表明逐图自适应采样能够有效改善不同退化输入下的恢复性能，其性能提升并非来源于选取较弱的固定基线，而是来源

© 中国图象图形学报版权所有

表3　不同固定采样路径的平均综合效用比较

Table 3　Comparison of average comprehensive utility for different fixed sampling paths

固定采样路径	采样步数	平均综合效用
[100]	1	0.7204
[150]	1	0.6887
[200]	1	0.6213
[250]	1	0.5201
[150,100]	2	0.6082
[200,100]	2	0.4973
[200,150]	2	0.5411
[150,100,50]	3	0.5348
[200,100,50]	3	0.4317
[200,150,100]	3	0.4617
[250,100,50]	3	0.2954
[250,150,50]	3	0.2670
[250,200,150]	3	0.3446
[200,150,100,50]	4	0.3902
[250,200,150,100]	4	0.2603
[250,200,150,100,50]	5	0.3731

于输入相关的路径预测能力。

2.6　定量对比

对比方法涵盖真实世界生成对抗超分辨率、扩散先验及高效扩散超分辨率模型，包括Real-ESRGAN（Wang X等，2021）、DiffBIR（Lin等，2024）、SeeSR（Wu等，2024a）、ResShift（Yue等，2023）、SinSR（Wang Y等，2024）、OSEDiff（Wu等，2024b）、InvSR（Yue等，2025），以及近年来提出的PiSA-SR（Sun等，2025）、FlowSR（Xu等，2025）和GDPO-SR（Yi等，2026）。其中，PiSA-SR和FlowSR代表一步高效超分辨率方法，GDPO-SR通过偏好优化进一步提升一步生成式超分辨率的恢复质量。ImageNet-Test和RealSR具有参考图像，采用全参考与无参考指标进行评价；RealSet80无参考图像，仅采用无参考指标。

由表5可见，相较InvSR-1，本文方法的SSIM由0.6789提升至0.6808，NIQE由4.3815降至4.0004，PI由3.0866降至2.9565，CLIP-IQA由0.7093提升至0.7274，同时参数量由33.84M降至28.47M，降幅约15.9%。PSNR、LPIPS和MUSIQ虽未超过基线，但所提方法在显著降低参数量的同时改善了结构一致性、图像自然度及感知质量。

由表6可见，在RealSR上，本文方法相较InvSR-1将PSNR与CLIP-IQA分别由24.50dB、0.6918提升至24.71dB、0.6966，NIQE由4.2189降至3.9566，PI由3.7779降至3.6148，说明输入相关的路径选择有助于改善真实退化图像的重建精度、自然统计特性和感知质量。但SSIM、LPIPS及MUSIQ未超过InvSR-1，表明图像级统一路径仍难以同时适配同一图像内部的不同退化区域。

由表7可见，在RealSet80上，本文方法的NIQE由4.0284降至3.8938，PI由3.4666降至3.4479，但CLIP-IQA和MUSIQ分别低于InvSR-1。该结果说明SPPM可在一定程度上改善未知真实退化下的自然统计特性，但语义一致性和多尺度感知质量仍存在波动。

2.7　可视化对比

图7选取树枝、建筑、人脸和蔬菜4类场景，对多种方法进行比较。树枝样本考察细长结构连续性和枝杈恢复，本文方法保持枝干走向，部分细长结构更连续，明显断裂和异常着色减少。建筑样本用于检验边缘、墙体纹理和人物轮廓，重建主体结构清楚，伪影较少。人脸样本重点观察眼部、鼻部轮廓及皮肤纹理，本文方法在增强清晰度的同时，较稳定保持整体轮廓和主要五官结构，未见明显过度锐化。蔬菜样本包含密集纹理和复杂颜色分布，所得结果边

表4　不同采样路径选择策略的定量比较

Table 4　Quantitative comparison of different sampling path selection strategies

路径选择策略	PSNR/dB ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PI ↓	CLIP-IQA ↑	MUSIQ ↑
最佳固定路径	24.46	0.6988	0.2543	4.2213	3.1148	0.6559	67.0717
随机路径	23.18	0.6732	0.2837	4.5855	3.4209	0.6742	67.5049
SPPM预测路径	23.65	0.6808	0.2634	4.0004	2.9565	0.7274	71.0955
Oracle路径	24.65	0.6930	0.2468	4.1985	3.1051	0.6765	68.0821

© 中国图象图形学报版权所有

表5　不同方法在ImageNet-Test数据集上的×4超分辨率定量比较

Table 5　Quantitative comparison of ×4 super-resolution on the ImageNet-Test dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PI ↓	CLIP-IQA ↑	MUSIQ ↑	参数量/M
Real-ESRGAN	26.62	<u>0.7523</u>	0.2303	4.4909	3.7234	0.5090	64.8186	16.70
DiffBIR-50	25.72	0.6695	0.2795	4.5875	3.2260	0.6900	69.7089	385.43
SeeSR-50	26.69	0.7422	0.2187	4.3825	3.4742	0.5868	71.2412	751.50
ResShift-4	**27.33**	**0.7530**	**0.1998**	5.8700	4.3643	0.6147	65.5860	118.59
SinSR-1	<u>26.98</u>	0.7304	<u>0.2209</u>	5.2623	3.8189	0.6618	67.7593	118.59
OSEDiff-1	23.95	0.6756	0.2624	4.7157	3.3775	0.6818	70.3928	8.50
InvSR-1	24.14	0.6789	0.2517	4.3815	3.0866	0.7093	72.2900	33.84
PiSA-SR-1	25.22	0.6971	0.2305	<u>4.0811</u>	<u>2.9892</u>	<u>0.7476</u>	**73.1519**	1300
FlowSR-1	25.57	0.7064	0.2278	4.1588	3.0804	**0.7569**	72.5601	982
GDPO-SR-1	25.02	0.6909	0.2347	4.3065	3.1580	0.7468	<u>72.7257</u>	1770
本文方法	23.65	0.6808	0.2634	**4.0004**	**2.9565**	0.7274	71.0955	28.47

注：粗体表示该指标的最优结果，下划线表示次优结果。

表6　不同方法在RealSR数据集上的×4超分辨率定量比较

Table 6　Quantitative comparison of ×4 super-resolution on the RealSR dataset

方法	PSNR/dB ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PI ↓	CLIP-IQA ↑	MUSIQ ↑	参数量/M
Real-ESRGAN	25.85	<u>0.7734</u>	0.2728	4.6766	4.4881	0.4898	59.6803	16.70
DiffBIR-50	24.83	0.6642	0.3864	**3.7366**	**3.3661**	0.6857	65.3934	385.43
SeeSR-50	**26.20**	0.7555	0.2806	4.5358	4.1464	0.6824	66.3757	751.50
ResShift-4	25.77	0.7453	0.3395	6.9113	5.4013	0.5994	57.5536	118.59
SinSR-1	26.02	0.7097	0.3993	6.2547	4.7183	0.6634	59.2981	118.59
OSEDiff-1	23.89	0.7030	0.3288	5.3310	4.3584	0.7008	65.4806	8.50
InvSR-1	24.50	0.7262	0.2872	4.2189	3.7779	0.6918	67.4586	33.84
PiSA-SR-1	<u>26.11</u>	**0.8494**	**0.2094**	4.3991	4.0241	0.7032	67.8927	1300
FlowSR-1	25.66	0.7401	0.2811	4.2121	3.8773	**0.7149**	<u>67.9030</u>	982
GDPO-SR-1	25.75	0.7307	<u>0.2664</u>	4.5981	4.0772	<u>0.7054</u>	**68.5741**	1770
本文方法	24.71	0.7128	0.3045	<u>3.9566</u>	<u>3.6148</u>	0.6966	67.1822	28.47

注：粗体表示该指标的最优结果，下划线表示次优结果。

界、表面纹理及色彩自然性较协调。相较InvSR-1，本文方法在样本中呈现更连续的轮廓和更清晰的局部纹理；与多步扩散方法相比，本文方法在自适应采样路径下获得具有竞争力的视觉效果。可视化结果用于补充定量指标，不替代全参考和无参考指标的综合分析。

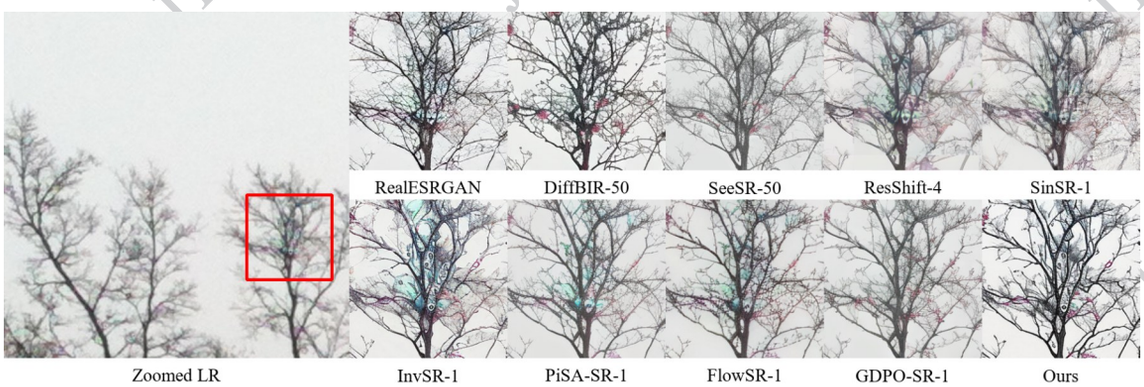
© 中国图象图形学报版权所有

表7　不同方法在RealSet80数据集上的无参考指标比较

Table 7　No-reference image quality comparison on the RealSet80 dataset

方法	NIQE ↓	PI ↓	CLIP-IQA ↑	MUSIQ ↑
Real-ESRGAN	4.1568	3.8852	0.6189	64.4957
DiffBIR-50	**3.8630**	**3.2117**	<u>0.7404</u>	67.9806
SeeSR-50	4.3678	3.7429	0.7114	69.7658
ResShift-4	5.9866	4.8318	0.6515	61.7967
SinSR-1	5.6243	4.2830	0.7228	64.0573
OSEDiff-1	4.3457	3.8219	0.7093	68.8202
InvSR-1	4.0284	3.4666	0.7291	69.8055
PiSA-SR-1	4.0033	3.5805	0.7325	<u>70.8440</u>
FlowSR-1	3.9075	3.4785	**0.7504**	70.7141
GDPO-SR-1	4.3315	3.7084	0.7349	**71.8842**
本文方法	<u>3.8938</u>	<u>3.4479</u>	0.7014	68.8004

注:粗体表示该指标的最优结果,下划线表示次优结果。

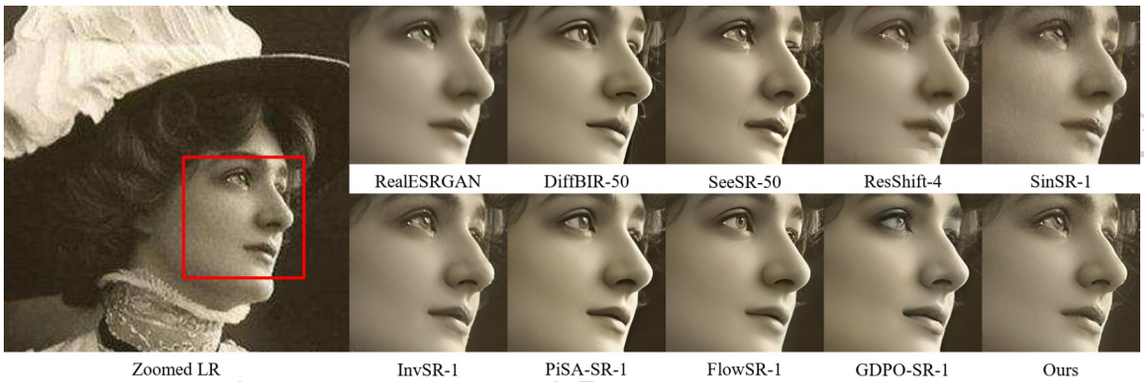


（a）　树枝

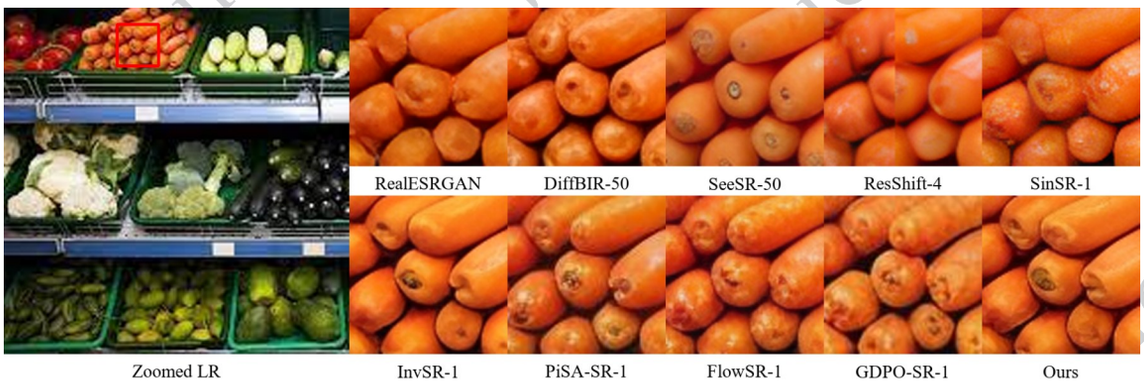


（b）　建筑

© 中国图象图形学报版权所有



（c） 人脸



（d） 蔬菜

图7 超分辨率重建结果的可视化比较

Fig. 7 Visual comparison of super-resolution reconstruction results

2.8 误差分析与局限性

从退化类型、图像内容和采样路径3方面分析：对退化较轻或整体结构完整的图像，SPPM通常选择较短路径，可减少冗余采样，但局部高频纹理密集区域仍可能恢复不足；对严重模糊、压缩伪影明显或纹理缺失较多的图像，较长路径提供更充分的反向扩散过程，但受原始信息缺失限制，部分区域仍会产生纹理重建误差。

当前SPPM采用图像级预测，同一图像内部的空间非均匀退化只能由一条路径处理，局部适配能力有限；离线效用标签依赖有限的候选时间步组合，且不同评价指标之间存在保真度—感知质量权衡。因而，完整模型在ImageNet-Test上主要表现为自然度与感知质量指标改善，而全参考保真度指标存在一定波动。

3 结 论

本文提出多尺度特征增强与退化感知采样的扩散反演图像超分辨率方法。HPDM通过多尺度深度卷积与像素反重排保留局部边缘和细粒度信息；STDFM分别建模长程结构关系与局部纹理连续性，并以时间感知门控完成融合；SPPM根据输入退化特征选择时间步组合，使推理由固定路径转化为输入相关的路径决策。

实验表明，本文方法在ImageNet-Test上将InvSR-1的SSIM由0.6789提升至0.6808，NIQE由4.3815降至4.0004，PI由3.0866降至2.9565，CLIP-IQA由0.7093提升至0.7274，同时参数量由33.84M降至28.47M。在RealSR上，PSNR由24.50dB提升至24.71dB，NIQE和PI分别降至3.9566和3.6148；在RealSet80上，NIQE和PI分别降至3.8938和3.4479。消融实验表明，HPDM与STDFM在像素保真和结构纹理建模方面具有互补作用，SPPM进一步改善图像自然度和感知质量，但部分保真度及语义感知指标仍存在一定波动。

该方法仍受离线路径标签、有限候选时间步以及图像级路径决策的限制，在不同数据集上尚难同

© 中国图象图形学报版权所有

时兼顾像素保真度、感知质量与语义一致性。后续将扩展连续或更大范围的时间步搜索,研究区域级或块级路径调度以及分类器与超分辨率模型的联合优化,并在更多真实退化类型、放大倍率和跨域数据集上进一步验证其泛化能力与计算效率。

参考文献(References)

Blau Y, Mechrez R, Timofte R, Michaeli T and Zelnik-Manor L. 2018. The 2018 PIRM challenge on perceptual image super-resolution// Proceedings of the European Conference on Computer Vision Workshops. Munich, Germany: Springer: 334-355 [DOI: 10.1007/978-3-030-11021-5_21]

Cai J, Zeng H, Yong H, Cao Z 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, Republic of Korea: IEEE: 3086-3095 [DOI: 10.1109/ICCV.2019.00318]

Chen X, Wang X, Zhou J, Qiao Y and Dong C. 2023. Activating more pixels in image super-resolution Transformer//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 22367-22377 [DOI: 10.1109/CVPR52729.2023.02142]

Deng J, Dong W, Socher R, Li L J, Li K and Fei-Fei L. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 248-255 [DOI: 10.1109/CVPR. 2009. 5206848]

Dong C, Loy C C, He K and Tang X. 2016. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2): 295-307 [DOI: 10.1109/TPAMI.2015.2439281]

Dong L, Fan Q, Guo Y, Wang Z, Zhang Q, Chen J, et al. 2025. TSD-SR: one-step diffusion with target score distillation for real-world image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 23174-23184 [DOI: 10.1109/CVPR52734.2025.02158]

He K, Zhang X, Ren S and Sun J. 2016. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. [2026-08-01]. https://arxiv. org/abs/1704.04861 [DOI: 10.48550/arXiv. 1704. 04861]

Huynh-Thu Q and Ghanbari M. 2008. Scope of validity of PSNR in image/ video quality assessment. Electronics Letters, 44(13): 800-801 [DOI: 10.1049/el:20080522]

Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4401-4410 [DOI: 10.1109/CVPR.2019.00453]

Ke J, Wang Q, Wang Y, Milanfar P and Yang F. 2021. MUSIQ: multi-scale image quality Transformer//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5148-5157 [DOI: 10.1109/ICCV48922.2021.00510]

Kim J, Lee J K and Lee K M. 2016. Accurate image super-resolution using very deep convolutional networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1646-1654 [DOI: 10.1109/CVPR.2016.182]

Ledig C, Theis L, Huszár 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, USA: IEEE: 4681-4690 [DOI: 10.1109/CVPR.2017.19]

Li J, Cao J, Guo Y, Li W and Zhang Y. 2025. One diffusion step to real-world super-resolution via flow trajectory distillation//Proceedings of the 42nd International Conference on Machine Learning. Vancouver, Canada: PMLR, 267: 34044-34053

Li Y, Zhang K, Liang J, Cao J, Liu C, Gong R, et al. 2023. LSDIR: a large scale dataset for image restoration//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver, Canada: IEEE: 1775-1787 [DOI: 10.1109/ CVPRW59228.2023.00178]

Liang J, Cao J, Sun G, 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, Canada: IEEE: 1833-1844 [DOI: 10.1109/ICCVW54120.2021.00210]

Lim B, Son S, Kim H, Nah S 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, USA: IEEE: 136-144 [DOI: 10.1109/CVPRW.2017.151]

Lin X, He J, Chen Z, Lyu Z, Dai B, Yu F, et al. 2024. DiffBIR: toward blind image restoration with generative diffusion prior//Proceedings of the European Conference on Computer Vision. Milan, Italy: Springer: 430-448 [DOI: 10.1007/978-3-031-73202-7_25]

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]

Loshchilov I and Hutter F. 2019. Decoupled weight decay regularization// Proceedings of the International Conference on Learning Represen-

© 中国图象图形学报版权所有

tations［DOI：10.48550/arXiv.1711.05101］

Mittal A，Soundararajan R and Bovik A C. 2013. Making a completely blind image quality analyzer. IEEE Signal Processing Letters，20（3）：209-212［DOI：10.1109/LSP.2012.2227726］

Pech-Pacheco J L，Cristóbal G，Chamorro-Martínez J and Fernúndez-Valdivia J. 2000. Diatom autofocusing in brightfield microscopy：a comparative study//Proceedings of the International Conference on Pattern Recognition. Barcelona，Spain：IEEE：314-317［DOI：10.1109/ICPR.2000.903548］

Perez E，Strub F，de Vries H，Dumoulin V and Courville A. 2018. FiLM：visual reasoning with a general conditioning layer//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans，USA：AAAI Press，32（1）：3942-3951［DOI：10.1609/aaai.v32i1.11671］

Pertuz S，Puig D and Garcia M A. 2013. Analysis of focus measure operators for shape-from-focus. Pattern Recognition，46（5）：1415-1432［DOI：10.1016/j.patcog.2012.11.011］

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，USA：IEEE：10684-10695［DOI：10.1109/CVPR52688.2022.01042］

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］

Sandler M，Howard A，Zhu M，Zhmoginov A and Chen L C. 2018. MobileNetV2：inverted residuals and linear bottlenecks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Salt Lake City，USA：IEEE：4510-4520［DOI：10.1109/CVPR.2018.00474］

Sun L，Wu R，Ma Z，Liu S，Yi Q and Zhang L. 2025. Pixel-level and semantic-level adjustable super-resolution：a dual-LoRA approach//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，USA：IEEE：2333-2343［DOI：10.1109/CVPR52734.2025.00223］

Sun W，Wang Q Z，Chen X L，Ding C，Yan Q S，Zhang L，et al. 2025. Development of real-world image super-resolution. Journal of Image and Graphics，30（6）：1576-1592（孙巍，王乾宙，陈雪凌，丁晨，闫庆森，张磊，等. 2025. 真实场景下图像超分辨技术现状与趋势. 中国图象图形学报，30（6）：1576-1592）［DOI：10.11834/jig.240775］

Wang J，Chan K C K and Loy C C. 2023. Exploring CLIP for assessing the look and feel of images//Proceedings of the AAAI Conference on Artificial Intelligence，37（2）：2555-2563［DOI：10.1609/aaai.v37i2.25353］

Wang J，Yue Z，Zhou S，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，Yu K，Wu S，Gu J，Liu Y，Dong C，et al. 2018. ESRGAN：enhanced super-resolution generative adversarial networks//Proceedings of the European Conference on Computer Vision Workshops. Munich，Germany：Springer：63-79［DOI：10.1007/978-3-030-11021-5_5］

Wang X，Xie L，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，Canada：IEEE：1905-1914［DOI：10.1109/ICCVW54120.2021.00217］

Wang Y，Yang W，Chen X，Wang Y，Guo L，Chau L P，et al. 2024. SinSR：diffusion-based image super-resolution in a single step//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：25796-25805［DOI：10.1109/CVPR52733.2024.02437］

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［DOI：10.1109/TIP.2003.819861］

Wu R，Yang T，Sun L，Zhang Z，Li S and Zhang L. 2024a. SeeSR：towards semantics-aware real-world image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：25456-25467［DOI：10.1109/CVPR52733.2024.02405］

Wu R，Sun L，Ma Z and Zhang L. 2024b. One-step effective diffusion network for real-world image super-resolution//Advances in Neural Information Processing Systems 37. Vancouver，Canada：Neural Information Processing Systems Foundation，Inc.：92529-92553［DOI：10.52202/079017-2938］

Xu J，Li W，Sun H，Li F，Wang Z，Peng L，et al. 2025. Fast image super-resolution via consistency rectified flow//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu，USA：IEEE：11755-11765［DOI：10.1109/ICCV51701.2025.01093］

Yi Q，Li S，Wu R，Sun L，Zhang Z and Zhang L. 2026. GDPO-SR：group direct preference optimization for one-step generative image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver，USA：IEEE：2177-2187

Yue Z，Wang J and Loy C C. 2023. ResShift：efficient diffusion model for image super-resolution by residual shifting//Advances in Neural Information Processing Systems 36. New Orleans，USA：Neural Information Processing Systems Foundation，Inc.：13294-13307［DOI：10.52202/075280-0583］

Yue Z，Liao K and Loy C C. 2025. Arbitrary-steps image super-resolution via diffusion inversion//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，USA：IEEE：23153-23163［DOI：10.1109/CVPR52734.

© 中国图象图形学报版权所有

2025.02156]

Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: efficient Transformer for high-resolution image restoration//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5728-5739 [DOI: 10.1109/CVPR52688.2022.00564]

Zhang K, Liang J, Van Gool L and Timofte R. 2021. Designing a practical degradation model for deep blind image super-resolution//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4791-4800 [DOI: 10.1109/ICCV48922.2021.00475]

Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]

Zhang T, Duan Z P, Guo C L, Jiang P T, Li B, Cheng M M, et al. 2026. Time-aware one step diffusion network for real-world image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver, USA: IEEE: 30543-30552

Zhang Y, Tian Y, Kong Y, Zhong B and Fu Y. 2018a. Residual dense network for image super-resolution//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2472-2481 [DOI: 10.1109/CVPR.2018.00262]

Zhang Y, Li K, Li K, Wang L, Zhong B and Fu Y. 2018b. Image super-resolution using very deep residual channel attention networks//Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer: 286-301 [DOI: 10.1007/978-3-030-01234-2_18]

作者简介

孟勃，女，教授，博士生导师，主要研究方向为数字图像处理、计算机视觉、深度学习算法、电力机器人视觉、目标检测与跟踪。E-mail：2323871167@qq.com

丁泽刚，通信作者，男，硕士研究生，主要研究方向为计算机视觉和扩散模型。E-mail：2110493443@qq.com

© 中国图象图形学报版权所有