文章编号:1674-2869(2024)05-0564-10

# 实际场景人脸超分辨率算法综述

朱柏霖,卢 涛*,王依伊,饶茜雅,赵康辉,张彦铎
武汉工程大学计算机科学与工程学院,湖北 武汉 430205

**摘 要**:人脸超分辨率能够有效提高低分辨率人脸图像的分辨率和质量,因而在视频监控、刑事侦查、娱乐等领域得到了广泛应用。然而实际场景中成像系统、记录设备、传输介质和处理方法不完善,导致噪声、模糊等降低图像质量的多种降质过程以不规则的方式组合,仅假设明确的降质过程来训练网络模型无法满足实际应用需求。针对这些实际场景中人脸超分辨率存在的多样化降质过程,分别介绍了非盲降质人脸超分辨率技术和盲降质超分辨率技术原理、人脸超分辨率领域常用的数据集和评价指标以及代表性工作的主客观重建结果。未来相关研究应聚焦多模态信息决策融合和张量融合,提升重建图像特征维度和时域相似性;通过大规模预训练和对抗学习等,提升模型泛化能力;研究身份一致性算法以及迁移学习等技术对处理复杂成像条件的影响。

**关键词**:人脸超分辨率;实际场景;深度学习;降质过程
**中图分类号**:TP391.4 **文献标识码**:A **DOI**:10.19843/j.cnki.CN42-1779/TQ.202211016

## Review of real-world face super-resolution algorithms

*ZHU Bolin*, *LU Tao**, *WANG Yiyi*, *RAO Xiya*, *ZHAO Kanghui*, *ZHANG Yanduo*
School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

**Abstract**: Face super-resolution enhances resolution and quality of low-resolution facial images, finding wide applications in fields of video surveillance, criminal investigation, and entertainment. However, in real-world scenarios, imperfect imaging systems, recording equipment, transmission media, and processing methods resulted in irregular combinations of degradation processes that reduce image quality, like noise and blurring. Training network models based on explicit degradation processes alone cannot meet practical needs. This paper reviews the principles of non-blind and blind face super-resolution techniques, the commonly used datasets and evaluation metrics and the subjective and objective reconstruction results of representative works in the field of face super-resolution. Future related researches should focus on multi-modal information decision fusion and tensor fusion to improve the feature dimension and temporal similarity of the reconstructed images; enhance the generalization ability of the model through large-scale pre-training and adversarial learning; investigate the impact of identity consistency algorithms and technologies such as transfer learning on complex imaging conditions.

**Keywords**: face super-resolution; real-world; deep learning; degradation process

图像或视频是人类感知客观世界最直观的方式之一,在通信、遥感、医学、娱乐等领域中得到了广泛的应用[1],其分辨率高低直接影响采样信息的重建保真度。然而传输系统的带宽约束、图像传感器的物理限制、超声影像的质量不足和网络视频的码率限制等因素,降低了实际场景图像分

**收稿日期**:2022-11-19
**基金项目**:国家自然科学基金(62072350,62171328)
**作者简介**:朱柏霖,硕士研究生。Email:bberllin@foxmail.com
***通信作者**:卢 涛,博士,教授。Email:lut@wit.edu.cn
**引文格式**:朱柏霖,卢涛,王依伊,等.实际场景人脸超分辨率算法综述[J].武汉工程大学学报,2024,46(5):564-573.

辨率和质量，无法满足高质量需求[2]。超分辨率（super-resolution，SR）技术能够提高数字图像/视频的空间/时间分辨率、细节和清晰度，在多个实际应用中展现了重要价值[3]。

在视频监控、人脸识别、图像复原等实际应用场景中，人脸图像扮演着至关重要的角色[4]。摄像头与待捕获人脸距离不同，影响捕获人脸尺度大小，且捕获人脸存在不同的面部姿势和表情，影响面部信息保真度，导致人脸图像失真、模糊、被遮挡，影响人脸识别、表情分析等应用的准确性和可靠性，增加了人脸超分辨率（face super-resolution，FSR）任务的复杂性，引起世界范围内学者们的广泛关注[5]。

FSR技术作为图像SR技术的特定领域，旨在提高低分辨率（low-resolution，LR）人脸图像的清晰度[6]。由于LR图像中信息的缺失，使得求解高分辨率（high-resolution，HR）图像成为一个具有多个可能解的问题，其可视为欠定方程的求解问题：引入人脸先验等约束条件获得稳定的方程解。现有大多数FSR技术仅对明确的降质过程进行建模求解，难以处理实际场景中降质过程复杂不明确的情况。

鉴于此，根据实际场景降质过程是否明确，对FSR技术进行调研综述。按照不同发展阶段总结FSR的发展历程，详细阐述基于非盲降质和盲降质的FSR技术。最后根据可视化对比结果，剖析影响SR算法性能的因素，探讨未来发展方向。

## 1　FSR技术与发展历程

### 1.1　FSR定义

FSR技术旨在提高LR人脸图像的分辨率，求解欠定方程恢复图像细节，即使用不同的重建函数获得不同的解。通过一般退化过程得到重建后的人脸图像，如式（1）所示。

$$I_{\mathrm{SR}} = \mathit{Ø}^{-1}\left(I_{\mathrm{LR}}, \delta\right) = F\left(I_{\mathrm{LR}}, \delta\right) \tag{1}$$

其中，$\mathit{Ø}^{-1}(\cdot)$表示人脸图像降质过程的逆运算，$I_{\mathrm{SR}}$表示重建后的人脸图像，$\delta$表示重建函数的相关参数（如学习率、梯度参数等），$I_{\mathrm{LR}}$表示LR人脸图像，$F(\cdot)$表示SR模型。

参数$\delta$的训练目标函数如式（2）所示：

$$\hat{\delta} = \arg\min_{\delta} L\left(I_{\mathrm{SR}}, I_{\mathrm{HR}}\right) \tag{2}$$

其中，$L(\cdot)$表示损失函数。

### 1.2　图像降质过程定义

由于图像在获取、传输或处理过程中可能遭受质量损失，通常涉及模糊和噪声等多种退化因素的复合作用，给定LR人脸图像$I_{\mathrm{LR}}$和HR人脸图像$I_{\mathrm{HR}}$，图像的一般降质过程如式（3）所示。

$$I_{\mathrm{LR}} = \mathit{Ø}\left(I_{\mathrm{HR}}, \theta\right) \tag{3}$$

其中，$\mathit{Ø}(\cdot)$表示图像降质过程的函数，$\theta$表示模型参数，包括模糊核、下采样操作和噪声等。

在可控的非盲降质条件下，给定式（4）以描述图像退化函数。

$$I_{\mathrm{LR}} = \left(I_{\mathrm{HR}}\right)\downarrow_{s} \tag{4}$$

其中，$\downarrow$表示下采样操作，$s$表示比例因子，这类下采样操作多为单一的双线性插值或者双三次插值，然而这种退化模型过于单一，并未考虑到实际场景下的噪声、模糊等因素。为此，一个更贴近实际场景下的退化过程被提出，如式（5）所示。

$$I_{\mathrm{LR}} = J\left(\left(I_{\mathrm{HR}} \circledcirc k\right)\downarrow_{s} + n\right) \tag{5}$$

其中$k$、$\circledcirc$、$n$和$J(\cdot)$分别代表模糊核、卷积运算、噪声和图像压缩函数。

### 1.3　FSR技术发展历程

2000年，Baker等[7]首次提出FSR概念，通过数学理论模型还原LR人脸图像。随后Liu等[8]使用全局-局部双分支模型和主成分分析（principal component analysis，PCA）法重建人脸边缘。在这些开创性工作的启发下，越来越多的工作使用数字图像处理技术对公式（1-3）进行求解。

Yang等[9]通过高、低分辨率图像块计算稀疏编码的系数。Kim等[10]采用梯度下降的核脊回归（kernel ridge regression，KRR）算法拟合映射函数。Wang等[11]结合高斯、拉普拉斯范数推导贝叶斯框架最优解。Chang等[12]使用流行学习和邻域插值少量数据拟合重建过程。随着FSR数字处理技术的发展，大多数工作采用稀疏编码、流形学习等数字建模恢复人脸信息。

在深度学习有了突破性进展后，Zhou等[13]将深度学习用于FSR技术，设计双通道卷积网络提出人脸表征。Huang等[14]使用像素级卷积神经网络（convolutional neural networks，CNN）与小波变换预测相应的小波系数。但基于像素损失训练往往导致输出过于平滑，为解决这一问题，Goodfellow等[15]首次提出生成对抗网络（generative adversarial networks，GAN），恢复更多细节的人脸图像并减少重建图像过于平滑的问题。随后Cai等[16]使用GAN的生成器恢复遮挡图像，融合多种损失函数指导鉴别器筛选无遮挡HR人脸图像，在跨数据库测试中表现出更好的模型泛化性能。

采用深度学习的FSR技术取得了显著的进展，但难以处理式(5)中噪声、压缩等降质过程，导致真实图像退化模型与假设的建模模型不一致。为此，2018年，Bulat等[17]首次提出对真实世界SR重建任务进行尝试。

## 2 实际场景FSR技术

2021年，Jiang等[4]较全面地调研了基于深度学习的FSR技术研究。考虑到实际场景图像降质过程的多样性，根据降质过程是否明确已知将FSR技术分为基于非盲降质和基于盲降质的FSR技术。

基于非盲降质的FSR技术利用人脸先验处理明确的降质过程，在部分实际场景数据集中取得不错的效果，但泛化性能较差；基于盲降质的FSR技术学习单一模型应对1个或多个退化过程的实例，不需要额外的先验信息或参考图像就能提高模型泛化能力和重建结果的真实程度。

### 2.1 基于非盲降质的FSR技术

基于非盲降质的FSR技术通常假设图像存在特定降质过程，针对不同降质过程单独建模，可分为空间下采样、噪声、模糊及多种降质过程组合的FSR技术。

2.1.1 空间下采样FSR技术 经典的SR任务将HR图像经过双线性插值或双三次插值等下采样操作得到LR图像。程超月[18]和邵奔[19]利用经过空间下采样处理得到LR图像的身份先验，通过空间注意力残差网络(spatial attention residual network，SPARNet)[20]重建清晰的人脸关键结构。上述方法取得了较高的峰值信噪比(peak signal-to-noise ratio，PSNR[21])，但PSNR仅关注像素差异，忽视了图像感知质量，且重建效果过于平滑，缺乏细节纹理信息。

为此，Zhang等[22]提出一种基于身份先验的有监督像素级GAN，重建出真实且纹理复杂的视觉效果。林旺庆[23]基于数据扩充和人脸属性、关键点等面部先验知识增强高频细节的捕捉能力。Xin等[24]提出一种结合像素级和语义级信息的密集连接网络，提高了面部纹理细节的恢复能力。

考虑到重建图像中包含的原始图像面部信息，He等[25]提出堆叠GAN的端到端网络，恢复结果保留输入面部的身份细节。为了改善恢复图像细节的不良伪影，Wang等[26]将密集卷积网络和谱归一化GAN相结合。虽然上述方法取得了更好的人脸轮廓细节和面部质感，但其训练的不稳定性导致模型难以收敛，重建结果有时会缺少必要的高频细节。

空间下采样方法的训练样本与实际场景图像存在域差，影响测试结果。上述方法仅考虑人脸图像本质上的放缩问题、关注LR图像和HR图像间的映射关系，并未关注实际场景多种因素的影响，泛化能力较差。

2.1.2 噪声FSR技术 针对噪声干扰的LR人脸图像，需要同时解决空间下采样的域差问题和噪声抑制问题。为此，Chen等[27]提出一种基于时序门控集成网络的多尺度噪声人脸复原方法。Nagar等[28]能够在高密度脉冲噪声下保证鲁棒性。Aakerberg等[29]首次对实际场景下LR人脸进行任意尺度的重建。上述方法可以高效重建噪声人脸图像，但处理复杂面部细节和复杂噪声时往往会增加边缘伪影、模糊等。

为此，Pang等[30]提出一种渐进式采样与样本调度相结合的策略。Miao等[31]通过小波变换分解频域分量，指导模型恢复复杂面部细节。为了解决混合高斯噪声和脉冲噪声的FSR问题，Tang等[32]提出一种自适应为像素分配权重以自动识别混合噪声的方法，在实际场景数据集上表现良好。

不同去噪技术有效抑制了LR图像的噪声，且可以应用于各种类型的人脸图像，具有高适用性。但其计算复杂度较高，过度去噪时会丢失图像细节，影响重建图像的真实性。

2.1.3 模糊FSR技术 模糊的LR人脸图像会导致细节丢失，针对模糊人脸图像首先进行模糊去除步骤，然后再应用SR算法。例如，辛经纬[33]将图像的语义表征、概率分布与人脸属性相结合，增加强噪声和模糊的人脸数据集的重建效果。未考虑个人身份信息的人脸验证方法的重建图像可能继承他人身份。为了解决这个问题，Xu等[34]提出一种基于身份保持的高度模糊人脸图像重建方法，提取不同层次身份信息和语义特征。

去模糊技术恢复了LR图像细节，适用于由运动模糊、镜头模糊等引起的图像模糊。然而，现有技术难以解决运动估计、模糊核估计等复杂问题，且过度去模糊同样会放大图像噪声，影响重建质量。

(4)多种降质过程组合的FSR技术

当同时存在模糊、噪声或空间下采样的不同组合的人脸图像时，首先需要分析其包含的降质过程，针对不同降质过程采用不同方法。如Tu等[35]提出多尺度退化人脸恢复模型，使用姿态学习重建正面化的人脸图像。

多种降质过程组合的FSR技术考虑到实际场

景下的多种退化因素，得以模拟更符合实际场景下人脸图像的降质过程。

目前，非盲降质的FSR技术主要用于解决带有噪声、模糊等因素的FSR问题。通过构建实际场景的LR到HR人脸图像数据对，可以在实际场景下获得更好的重建效果。

### 2.2　基于盲降质的FSR技术

基于盲降质的FSR技术使用单一退化模型应对多种退化组合，如Li等[36]使用实际退化观测值和同身份的高质量图像修复未知退化。Yang等[37]隐式计算错误映射，优先分解/融合不同子空间，动态处理不同退化特征。然而上述方法生成模型会导致重建结果过度平滑，为此，Guo等[38]将预训练GAN生成的准确几何先验融入盲降质模型；Teng等[39]提出集成生成先验和人脸几何先验的协作网络，均提高了模型泛化能力。考虑到面部上下文信息与人脸先验的相互作用，Wang等[40]引入空间注意机制对上下文信息和先验信息进行建模，提高还原图像的真实感和保真度。Ge等[41]设计小波鉴别器端到端重建更多身份细节保留的图像，提高严重退化下身份保留细节提取能力。

尽管上述方法在处理训练数据的复杂退化问题上取得了显著成果，但仍难以模拟复杂退化过程，且需要频繁调整保真度、感知损失等超参数。为此，Yue等[42]基于扩散模型，通过建立低质量图像到高质量图像的后验分布，提高面部形状和细节的恢复能力。Gao等[43]基于扩散模型设计多尺度深度反向投影网络，提高不同尺度恢复图像的质量。Chen等[44]使用预训练稳定扩散的生成先验指导自注意力网络训练，在真实世界公共测试数据集上达到了最先进的性能。为了提高重建结果的真实程度和扩散模型的泛化能力，Yang等[45]单独建模退化属性，指导反向扩散过程。Gao等[46]提出扩散-信息-扩散框架，高度泛化不同退化场景和异构域下的人脸特征。

为了快速学习新任务并克服训练和测试人脸图像场景不一致的问题，使用元学习方法训练，根据输入图像特性构建一个测试时调整参数的自适应模型，让模型获得一种“学习”的能力。Xia等[47]使用元学习泛化大量外部数据，快速适应特定的复杂退化并提取隐式退化信息。Hu等[48]设计了一种元恢复模块处理各种退化因素，有良好的实际应用前景。

基于扩散模型的方法分步添加如式(5)所示的模糊核、噪声等模拟降质过程，有效应对多种降质过程并生成更稳定、清晰的图像。但生成速度慢，处理非图像或结构化数据时泛化能力弱。而元学习方法可以快速适应新场景、新任务，具有高泛化性，但极其依赖任务分布和数据质量。

由于实际场景捕获的低质量(low quality，LQ)人脸图像，其退化过程往往过于复杂，退化函数无法用单一退化模型精确表达。而基于盲降质的FSR技术旨在重建未知退化的LQ人脸图像，引入了更全面的退化形式，随机采样各个降质过程控制退化的严重程度，有效处理现实任务。

## 3　图像质量评估指标与FSR技术数据集

### 3.1　图像质量评估指标

图像质量评估指标一般分为主观评价指标和客观评价指标[49]。主观评价指标主要依靠人的感知，易受主观想法和环境因素的干扰，难以形成标准的评价结果。而客观评价指标从数学角度出发计算具体数值，可以直观表明评价结果。客观评价指标包括PSNR、结构相似度(structural similarity，SSIM)[50]、自然图像质量评估(natural image quality evaluator，NIQE)[51]、图像感知相似度(learned perceptual image similarity，LPIPS)[52]、平均主观得分(mean opinion score，MOS)[53]和弗里歇特初始距离(Fréchet inception distance，FID)[54]。

两张图像的像素差异值越小，PSNR越高，仅关注图像中像素对齐的距离，可能导致PSNR更高但视觉感知更差。

SSIM用于衡量图像间的亮度、对比度和结构的相似性。若两张图像的SSIM越高，即图像失真程度越小，更符合视觉感知。

NIQE利用图像和预先建立的多元高斯之间的距离去衡量图像质量。NIQE值越小，图像质量越好，视觉效果更佳。

LPIPS也称感知损失，用于度量图片间的差别，相比PSNR和SSIM，LPIPS更符合人的感知情况。图像感知相似度$F_{\mathrm{LPIPS}}$的计算公式如下：

$$F_{\mathrm{LPIPS}}=\sum_{l}\frac{1}{H_lW_l}\sum_{h,w}\left\|\boldsymbol{\omega}_l\times\left(\hat{y}^l_{\mathrm{hw}}-\hat{y}^l_{0\mathrm{hw}}\right)\right\|_2^2 \tag{10}$$

其中$H_l$和$W_l$为第$l$层网络层对应的高和宽，$h$和$w$代表$H_l$和$W_l$的索引，$\hat{y}^l_{\mathrm{hw}}$和$\hat{y}^l_{0\mathrm{hw}}$分别为单位归一化第$l$层网络层后的预测特征值和参考特征值(通常是真值)。LPIPS首先使用向量$\boldsymbol{\omega}_l$缩放激活通道。

在计算$l$层特征堆栈后，对每层输出进行激活后在通道维度中进行单位归一化。最后利用$\boldsymbol{\omega}_l$在通道上计算最终的LPIPS值。LPIPS值越低表示两张图像越相似，反之差异越大。

MOS通过人为的测试图像评定分数，最后根据平均分得出结论。MOS值越大，表示成像的结果越好，但如果测试人员过少，MOS往往不够准确。

FID更加关注生成图像和真实图像的分布建模之间的差异，经常用于评估人脸图像的视觉质量。视觉质量越好，FID值越小。

### 3.2 FSR算法数据集

表1列举了常见的人脸图像数据集。其中特征点是指根据人脸生理结构预定义的面部关键位置，通常包括眼角、眉毛、鼻尖、嘴角等，特征点的数量和精确度直接影响后续处理任务的性能。

CelebA为人脸属性数据集，包括10 177个身份的202 599张人脸图片。其中每张图片均有特征标记，包括人脸标注框（bounding box，BBox）、5个人脸特征点坐标和40个属性标记（包括微笑、戴眼镜等）。

AFLW是一个大规模面部对齐数据集，包含从Flickr获取的各个姿势和表情，除了特征点，还提供矩形框和椭圆框的脸部位置标注。

AFW是早期人脸关键点检测常用数据集，包含205张图像，其中有473张标记的人脸，每个人脸提供6个关键点和3个姿势角度的标注。

300W数据集的每个图像上不止包含1张人脸，但每张图像上只标注1张人脸。

WiderFace共32 203张图像，包括393 703张人脸，面部尺寸、姿态、表情等方面变化较大，广泛应用于实际场景FSR工作。

WFLW数据集提供了10 000张图像，除了标注98个关键点之外还包括遮挡、姿态、妆容、光照等信息标注。可以对数据集上的姿态、遮挡等鲁棒性进行简单的评估。

**表1 人脸图像数据集**

Tab. 1 Face image datasets

| 数据集 | 发布时间/年 | 数量/个 | 特征点/个 |
|---|---|---|---|
| CelebA[55] | 2015 | 202 599 | 5 |
| LFPW[56] | 2011 | 1 432 | 29 |
| AFLW[57] | 2011 | 25 993 | 21 |
| Helen[58] | 2012 | 2 330 | 68 |
| AFW[59] | 2012 | 205 | 6 |
| 300W[60] | 2013 | 3 827 | 68 |
| WiderFace[61] | 2015 | 32 203 | × |
| WFLW[62] | 2018 | 10 000 | 98 |

## 4 现有FSR技术实验分析

对比非盲降质和盲降质的FSR经典方法的重建效果时，使用的数据集均为CelebA。表2展示了4、8和16倍放大因子下小波超分辨率网络（wavelet-based super-resolution network，Wavelet-SRNet）[14]、SPARNet、小波域生成对抗网络（wavelet-domain generative adversarial network，WaveletSRGAN）[60]、人脸对齐超分辨率网络（super-resolution with face alignment network，Super-FAN）[63]和属性增强卷积神经网络（attribute augmented convolutional neural network，AACNN）[64]的主客观指标，图1展示了不同非盲降质FSR技术对LR图像的4、8和16倍重建效果，其中HR经过对应倍数的双三次插值下采样得到LR，Bicubic表示对LR进行双三次插值上采样。结合图1和表2可以看出，Wavelet-SRNet的重建效果并不理想，牙齿等面部细节特征丢失严重，其原因是采用基于均方误差损失的残差网络减少了高频特征分量，且网络深度不足限制了复杂图像的细节纹理。类似的还有AACNN和SPARNet采用大量的池化层和卷积层，对输入的LR图像提取语义特征，导致信息丢失和语义混淆。虽然上述文献的PSNR等指标表现良好，但重建图像过于平滑。

Super-FAN额外进行了面部对齐，虽然更好地

**表2 非盲降质FSR技术定量比较[4]**

Tab. 2 Quantitative comparison for non-blind FSR methods

| 方法 | PSNR | | | SSIM | | | LPIPS | | |
|---|---|---|---|---|---|---|---|---|---|
| | 4倍 | 8倍 | 16倍 | 4倍 | 8倍 | 16倍 | 4倍 | 8倍 | 16倍 |
| Wavelet-SRNet | 30.92 | 26.83 | 23.13 | 0.908 | 0.787 | 0.634 | 0.051 | 0.089 | 0.160 |
| SPARNet | 31.71 | 27.44 | 23.68 | 0.913 | 0.804 | 0.674 | 0.048 | 0.089 | 0.139 |
| WaveletSRGAN | 30.72 | 25.55 | 22.78 | 0.907 | 0.765 | 0.625 | 0.045 | 0.092 | 0.148 |
| Super-FAN | 31.17 | 27.08 | 23.42 | 0.905 | 0.788 | 0.652 | 0.040 | 0.058 | 0.125 |
| AACNN | 31.30 | 26.68 | 22.98 | 0.907 | 0.773 | 0.626 | 0.052 | 0.100 | 0.171 |

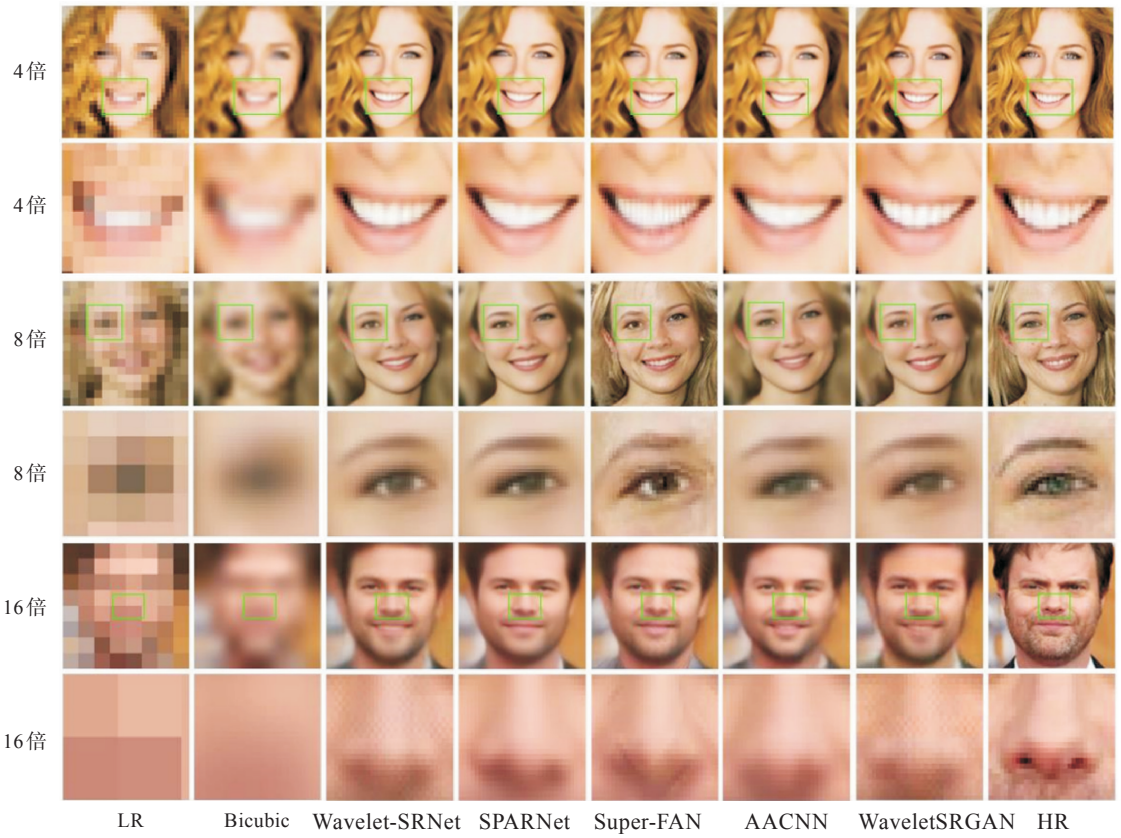


**图1　非盲降质FSR技术对LR图像的4倍、8倍和16倍重建效果**[4]

Fig. 1　The 4x, 8x, and 16x reconstruction effects of the non-blind degraded FSR technique on LR images

保留了面部结构，但重建结果中过度强调细节成分，生成了不必要的信息。这是因为热图损失的权重过大，过于强调面部对齐的精度。而WaveletSRGAN和Wavelet-SRNet虽然客观指标不突出，但人脸轮廓和细节质量更强，因为其引入了对抗损失，视觉效果上更加优秀。

基于盲降质的FSR技术的定量比较结果如表3所示。图2展示了盲降质FSR技术对LQ图像的16倍放大因子的可视化重建结果，测试集均采用CelebAHQ，Bilinear表示对LQ进行双线性插值上采样。结合图2和表3可以看出，使用双三次插值重建人脸图像获得了较高的PSNR，但难以恢复有意义的面部细节。深度人脸词典网络（deep face dictionary network，DFDNet）[67]过度依赖高质量参考图像，且采用的K均值生成字典对于人脸多样性存在局限，仅能捕捉有限的人脸变化和细节，导致重建结果出现色块，难以重建细节特征。渐进语义感知风格转换生成对抗网络（progressive semantic-aware style transformation for blind face restoration with GAN，PSFR-GAN）[66]重建出较多不存在的信息，这是因为其过度依赖解析图指导恢复过程。尽管采用多尺度特征调节细节信息的丰富度，但受限于人脸样本的多样性和质量，应用到实际场景仍需要提高模型泛化能力。生成性面部先验的生成对抗网络（generative facial prior with GAN，GPF-GAN）[68]的重建效果不理想，存在过多失真和虚假信息，这是因为预训练生成的面部先验不够准确，导致模型泛化能力不足。

生成对抗先验嵌入网络（GAN prior embedded network，GPEN）[65]将GAN嵌入深度神经网络，使用LQ人脸图像微调视觉效果。结构和生成先验融合网络（shape and generative prior integrated network，SGPN）[69]三维重建面部几何形状以自适应融合特征。GPEN和SGPN在人脸形状重建和人脸细节生成上达到良好平衡，重建出更逼真的人脸图像。虽然两者的重建结果良好，但SGPN在PSNR值和SSIM值相当的情况下，在LPIPS和FID

**表3　盲降质FSR技术定量比较**[68]

Tab. 3　Quantitative comparison for blind FSR methods

| 方法 | PSNR | SSIM | LPIPS | FID |
|---|---|---|---|---|
| Bilinear | 24.20 | 0.669 0 | 0.605 1 | 129.22 |
| GPEN | 22.93 | 0.604 8 | 0.292 9 | 13.34 |
| PSFR-GAN | 23.07 | 0.603 5 | 0.323 2 | 24.29 |
| DFDNet | 23.07 | 0.577 5 | 0.355 6 | 25.08 |
| GFP-GAN | 22.06 | 0.589 4 | 0.311 9 | 20.21 |
| SGPN | 23.10 | 0.614 6 | 0.269 8 | 7.21 |




**图2 盲降质FSR技术对LQ图像的16倍重建效果**[68]

Fig. 2 The 16x reconstruction effect of blind FSR techniques on LQ images

评估指标上结果更好,表明重建图像更接近于原始图像,能够得到更清晰的局部细节。这是因为SGPN的集成网络能够更有效地利用预训练生成的先验信息,相对于GPEN限制求解空间,SGPN能够更好地捕捉并模拟真实人脸图像的分布特征。

在FSR任务中,既要追求高客观指标,也要保持良好的主观效果。引入对抗损失的GAN方法在PSNR和SSIM相当的情况下,可以重建更好的人脸轮廓和细节质量。同时,能否更有效地利用预训练的先验信息,采用更好的重建面部几何形状、高生成能力的网络等,对处理降质过程具有决定性作用。

## 5 总结与展望

在非盲降质问题中,基于传统损失函数(如MSE)的方法的重建结果并不理想,其原因是网络结构设计不足以捕捉复杂图像细节特征,导致重建图像过于平滑或出现信息丢失等问题。引入对抗损失等方法后,人脸轮廓、细节质量都有所改善。一些盲降质FSR技术在PSNR指标上更突出,但重建结果缺乏面部细节,感知质量较差。其中高效利用先验的方法如SGPN在主观评估指标上表现更好。现有大多数LR图像通过空间下采样得到,无法应对部分实际场景需求,且实际场景降质过程难以拟合,限制模型泛化能力。

使用大模型可以更好适应特定场景的图像特点,例如改善夜间监控或低光照环境对高精度人脸图像的需求。这是因为大模型具有强大的特征提取能力和表示能力,能够拟合更广泛的数据分布提高泛化性能。设计大模型需要考虑计算效率和资源消耗,可以引入大规模预训练和对抗学习等技术,进一步提升模型的泛化能力。借鉴一般图像SR任务中良好的网络结构,将其扩展为更适合FSR任务的高效深度网络。

实际场景的人脸数据集可以更好地反应真实世界的多样性和复杂性,涵盖不同种族、年龄、性别、面部特征等。通过模拟更真实的图像降质过程生成的LR人脸图像,使训练的网络更适用于实际应用场景。

在实际应用的不同成像条件下,如何保持人脸图像的个体身份不变是一个重点。为了同时提高视觉质量,可以研究维持身份一致性的算法,增强如皮肤纹理、瞳孔等感知细节,使网络更适用于真实世界人脸图像。

深度、纹理和光照等多模态信息提供额外的上下文联系。融合这些特征张量和时间信息可以帮助模型更好地理解人脸的三维结构和表面特性。然而现有大多数方法主要应用计算机视觉技术,忽视了文本、声音和图像等多模态算法,可以通过深度信息指导纹理恢复、光照信息改善面部阴影和高光效果等对现有方法进行改进。

不同光谱、光源位置、光照强度等都会对人脸图像产生影响。在背光环境下,人脸可能会偏暗看不清细节,且人脸与摄像机镜头的相对位置决定人脸姿态的多样性。针对这个问题,可以引用人脸皮肤色彩等先验信息减少光照影响,使用迁移学习或增量学习技术快速适应不同光照、表情等变化。

**参考文献**

[1] 刘长新,吴宁,胡俐蕊,等.基于递归门控卷积的遥感图像超分辨率研究[J].计算机科学,2024,51(2):205-216.

[2] 李嫣,任文琦,张长青,等.基于真实退化估计与高频引导的内窥镜图像超分辨率重建[J].自动化学报,2024,50(2):334-347.

[3] 江俊君,程豪,李震宇,等.深度学习视频超分辨率技术综述[J].中国图象图形学报,2023,28(7):1927-1964.

[4] JIANG J J, WANG C Y, LIU X M, et al. Deep learning-based face super-resolution: a survey [J]. ACM Computing Surveys, 2023, 55(1):13.
[5] 卢涛,章瑾,陈白帆,等.多尺度自适应配准的视频超分辨率算法[J].武汉工程大学学报,2016,38(2):178-184,194.
[6] 郭婷,张天序,郭诗嘉.一种红外图像和宽光谱融合的人脸识别算法[J].武汉工程大学学报,2022,44(3):320-324.
[7] BAKER S, KANADE T. Hallucinating faces[C]// Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition. Piscataway, NJ: IEEE, 2000: 83-88.
[8] LIU C, SHUM H Y, ZHANG C S. A two-step approach to hallucinating faces: global parametric model and local nonparametric model [C]// Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2001.
[9] YANG J C, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
[10] KIM K I, KWON Y. Single-image super-resolution using sparse regression and natural image prior [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(6): 1127-1133.
[11] WANG Z Y, HAN Z, HU R M, et al. Noise robust face hallucination employing Gaussian-Laplacian mixture model [J]. Neurocomputing, 2014, 133: 153-160.
[12] CHANG H, YEUNG D Y, XIONG Y M, et al. Super-resolution through neighbor embedding [C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2004.
[13] ZHOU E, FAN H Q, CAO Z M, et al. Learning face hallucination in the wild [C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Palo Alto, CA: Association for the Advancement of Artificial Intelligence, 2015: 3871-3877.
[14] HUANG H B, HE R, SUN Z N, et al. Wavelet-SRNet: a wavelet-based CNN for multi-scale face super resolution [C]//2017 IEEE International Conference on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2017: 1699-1706.
[15] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial net [C]// Proceedings of the 27th International conference on Neural Information Processing Systems. Cambridge, MA: MIT press, 2014: 2672-2680.
[16] CAI J C, HAN H, SHAN S G, et al. FCSR-GAN: joint face completion and super-resolution via multi-task learning [J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2020, 2(2): 109-121.
[17] BULAT A, YANG J, TZIMIROPOULOS G. To learn image super-resolution, use a GAN to learn how to do image degradation first [C]//Computer Vision-ECCV 2018 Workshops. Switzerland: Springer, 2018: 187-202.
[18] 程超月.基于先验信息的人脸超分辨率重建技术研究[D].北京:北京邮电大学,2021.
[19] 邵奔.身份约束深度低分辨率人脸识别[D].武汉:中南民族大学,2021.
[20] CHEN C F, GONG D H, WANG H, et al. Learning spatial attention for face super-resolution [J]. IEEE Transactions on Image Processing, 2020, 30: 1219-1231.
[21] HUYNH-THU Q, GHANBARI M. Scope of validity of PSNR in image/video quality assessment [J]. Electronics Letters, 2008, 44(13): 800-801.
[22] ZHANG M L, LING Q. Supervised pixel-wise GAN for face super-resolution [J]. IEEE Transactions on Multimedia, 2020, 23: 1938-1950.
[23] 林旺庆.基于先验知识的人脸超分辨与基于样本扩充的低分辨率人脸识别方法研究[D].厦门:厦门大学,2020.
[24] XIN J W, WANG N N, GAO X B, et al. Residual attribute attention network for face image super-resolution [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 9054-9061.
[25] HE J J, ZHENG J J, SHEN Y, et al. Facial image synthesis and super-resolution with stacked generative adversarial network [J]. Neurocomputing, 2020, 402: 359-365.
[26] WANG M X, CHEN Z X, WU Q M J, et al. Improved face super-resolution generative adversarial networks [J]. Machine Vision and Applications, 2020, 31: 22.
[27] CHEN Z B, LIN J X, ZHOU T K, et al. Sequential gating ensemble network for noise robust multiscale face restoration [J]. IEEE Transactions on Cybernetics, 2021, 51(1): 451-461.
[28] NAGAR S, JAIN A, SINGH P K, et al. Pixel-wise dictionary learning based locality-constrained representation for noise robust face hallucination [J]. Digital Signal Processing, 2020, 99: 102667.
[29] AAKERBERG A, NASROLLAHI K, MOESLUND

T B. Real-world super-resolution of face-images from surveillance cameras [J]. IET Image Processing, 2022, 16(2): 442-452.

[30] PANG Y, MAO J W, HE L B, et al. An improved face image restoration method based on denoising diffusion probabilistic models [J]. IEEE Access, 2024,12:3581-3596.

[31] MIAO Y Q, DENG J K, HAN J G. WaveFace: authentic face restoration with efficient frequency recovery [Z/OL]. (2024-05-19) [2024-05-23]. https://doi.org/10.48550/arXiv.2403.12760

[32] TANG S Z, SHU Z Q. Mixed noise face hallucination via adaptive weighted residual and nuclear-norm regularization [J]. Applied Intelligence, 2023, 53: 11979-11996.

[33] 辛经纬. 面向视频监控的人脸图像超分辨率重建算法研究[D]. 西安:西安电子科技大学,2022.

[34] XU Y, ZOU H Y, HUANG Y, et al. Super-resolving blurry face images with identity preservation [J]. Pattern Recognition Letters, 2021, 146: 158-164.

[35] TU X G, ZHAO J, LIU Q K, et al. Joint face image restoration and frontalization for recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1285-1298.

[36] LI X M, LIU M, YE Y T, et al. Learning warped guidance for blind face restoration [C]//Computer Vision-ECCV 2018. Switzerland: Springer, 2018: 278-296.

[37] YANG L B, WANG P, GAO Z N, et al. Implicit subspace prior learning for dual-blind face restoration [Z/OL]. (2020-10-12)[2024-05-04]. https://doi.org/10.48550/arXiv.2010.05508.

[38] GUO X M, YI L, ZOU H, et al. Generative facial prior for large-factor blind face super-resolution [J]. Journal of Physics: Conference Series, 2021, 2078: 012045.

[39] TENG Z, YU X S, WU C D. Blind face restoration via multi-prior collaboration and adaptive feature fusion [J]. Frontiers in Neurorobotics, 2022, 16: 797231.

[40] WANG Z X, ZHANG J W, CHEN T S, et al. Restoreformer++: towards real-world blind face restoration from undegraded key-value pairs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023,45(12):15462-15476.

[41] GE G J, SONG Q, ZHU G B, et al. BFRFormer: transformer-based generator for real-world blind face restoration [Z/OL]. (2024-02-29) [2024-05-04]. https://doi.org/10.48550/arXiv.2402.18811

[42] YUE Z S, LOY C C. Difface: blind face restoration with diffused error contraction[Z/OL]. (2022-12-13) [2023-12-11]. https://doi. org/10. 48550/arXiv. 2212. 06512

[43] GAO J H, TANG N, ZHANG D X. A multi-scale deep back-projection backbone for face super-resolution with diffusion models [J]. Applied Sciences, 2023, 13(14): 8110.

[44] CHEN X X, TAN J F, WANG T, et al. Towards real-world blind face restoration with generative diffusion prior [Z/OL]. (2024-03-18) [2024-05-04]. http://doi. org/10.485501arxiv.2312.15736.

[45] YANG P Q, ZHOU S C, TAO Q Y, et al. PGDiff: guiding diffusion models for versatile face restoration via partial guidance [C]//Advances in Neural Information Processing Systems. 2023:32194-32214.

[46] GAO N, LI J, HUANG H B, et al. DiffMAC: diffusion manifold hallucination correction for high generalization blind face restoration[Z/OL]. (2024-05-15) [2024-05-23]. https: //doi. org/ 10. 48550/arXiv.2403.10098

[47] XIA B, TIAN Y P, ZHANG Y L, et al. Meta-learning based degradation representation for blind super-resolution [J]. IEEE Transactions on Image Processing, 2023, 32: 3383-3396.

[48] HU X C, ZHANG Z, SHAN C F, et al. Meta-USR: a unified super-resolution network for multiple degradation parameters [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32 (9): 4151-4165.

[49] 张凯兵,朱丹妮,王珍,等.超分辨图像质量评价综述[J].计算机工程与应用,2019,55(4):31-40,47.

[50] WANG Z, BOVIK A C. A universal image quality index [J]. IEEE Signal Processing Letters, 2002, 9 (3): 81-84.

[51] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a "completely blind" image quality analyzer [J]. IEEE Signal Processing Letters, 2013, 20(3): 209-212.

[52] GHAZANFARI S, GARG S, KRISHNAMURTHY P, et al. R-LPIPS: an adversarially robust perceptual similarity metric[Z/OL]. (2023-07-27)[2023-07-31]. https://doi.org/10.48550/arXiv.2307.15157

[53] STREIJL R C, WINKLER S, HANDS D S. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives [J]. Multimedia Systems, 2016, 22(2): 213-227.

[54] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium [C]//Advances in Neural Information Processing Systems 30 (NIPS

2017). Red Hook, NY: Curran Associates Inc., 2017: 6626-6640.

[55] LIU Z W, LUO P, WANG X G, et al. Deep learning face attributes in the wild [C]//2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 3730-3738.

[56] BELHUMEUR P N, JACOBS D W, KRIEGMAN D J, et al. Localizing parts of faces using a consensus of exemplars[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2930-2940.

[57] KÖSTINGER M, WOHLHART P, ROTH P M, et al. Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization[C]//2011 IEEE International Conference on Computer Vision Workshops. Piscataway, NJ: IEEE Computer Society, 2011: 2144-2151.

[58] LE V, BRANDT J, LIN Z, et al. Interactive facial feature localization [C]//Computer Vision-ECCV 2012. Berlin: Springer, 2012: 679-692.

[59] ZHU X X, RAMANAN D. Face detection, pose estimation, and landmark localization in the wild [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 2879-2886.

[60] HUANG H B, HE R, SUN Z N, et al. Wavelet domain generative adversarial network for multi-scale face hallucination [J]. International Journal of Computer Vision, 2019, 127(6/7): 763-784.

[61] YANG S, LUO P, LOY C C, et al. Wider face: a face detection benchmark[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 5525-5533.

[62] WU W Y, QIAN C, YANG S, et al. Look at boundary: a boundary-aware face alignment algorithm [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, CA: IEEE, 2018:2029-2138.

[63] BULAT A, TZIMIROPOULOS G. Super-FAN: integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, CA: IEEE, 2018:109-117.

[64] LEE C H, ZHANG K P, LEE H C, et al. Attribute augmented convolutional neural network for face hallucination [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, CA: IEEE, 2018:834-842.

[65] YANG T, REN P R, XIE X S, et al. GAN prior embedded network for blind face restoration in the wild[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, CA: IEEE, 2021:672-681.

[66] CHEN C F, LI X M, YANG L B, et al. Progressive semantic-aware style transformation for blind face restoration [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, CA: IEEE,2021:11891-11900.

[67] LI X M, CHEN C F, ZHOU S C, et al. Blind face restoration via deep multi-scale component dictionaries [C]// Computer Vision-ECCV 2020. Berlin: Springer, 2020:399-415.

[68] WANG X T, LI Y, ZHANG H L, et al. Towards real-world blind face restoration with generative facial prior[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, CA: IEEE,2021:9164-9174.

[69] ZHU F D, ZHU J W, CHU W Q, et al. Blind face restoration via integrating face shape and generative priors[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, CA: IEEE,2022:7662-7671.

本文编辑:霍淑利