融合关键点约束和边缘信息的人脸超分辨率算法

刘锦博[1,2*] 肖 露[1,2] 邓 锴[1,2] 高天宇[1,2]
LIU Jinbo XIAO Lu DENG Kai GAO Tianyu

摘 要 人脸超分辨率方法存在边缘信息利用不足的状况，这致使轮廓模糊且细节缺失。基于此，文章提出一种融合关键点约束与边缘信息的重建算法。设计了双网络协同迭代反馈机制，借助交互式反馈逐步校正关键点位置，进而提高重建质量。建立结构纹理融合约束模块，把语义级关键点先验和像素级边缘信息进行多层次融合。运用自适应多尺度注意力机制，动态调节局部与全局特征的贡献比例，增强多尺度面部特征的表达能力。实验显示，所提方法在公开数据集上的综合性能比主流对比方法更佳，提高了重建人脸图像的视觉质量。

关键词 人脸超分辨率；边缘信息；关键点约束；自适应多尺度注意力；双网络协同迭代反馈

doi：10.3969/j.issn.1672-9528.2026.07.003

0 引言

人脸超分辨率（face super-resolution, FSR）技术是图像重建领域的重要分支，能够从低分辨率（low-resolution, LR）人脸图像中重建高分辨率（high-resolution, HR）细节。FSR 在视频监控、人脸识别和医学影像等领域具有广泛应用[1]。近年来，随着深度学习技术的发展，通过引入人脸先验信息，重建性能获得了极大的提升，例如使用面部关键点[2]、语义分割图[3]、三维形状模型[4]等去约束重建过程。然而，当前的方法还存在两方面的限制，一方面低分辨率输入会使得关键点估计不够准确，进而影响重建结构的一致性；另一方面过度依赖语义层先验却忽略了像素级边缘信息，造成高频细节恢复不充分。

为解决上述问题，研究者提出多种改进策略。2024 年，Hajian 等[5]提出的 MSRFSR（multi-stage refinement for face super-resolution）方法通过多阶段特征精化增强迭代协作，重点优化关键点估计精度。2025 年，Liu 等[6]利用面部语义特征显式引导注意力机制，缓解边缘模糊问题；Li 等[7]则探索基于图像内部自相似性先验的注意力聚合，提出了 SANet（self-similarity prior and attention network）模型，同时提升关键点定位和边缘恢复性能。尽管这些方法在特定场景下取得进展，但在极端低分辨率条件下（如 16×16 输入），仍面临关键点定位不准确和边缘细节丢失的双重挑战。

针对 FSR 中关键点估计不准确和边缘高频细节恢复不足的核心问题，本文从双网络协同优化、多源信息融合和注意力机制改进三个层面展开研究。首先，针对 LR 输入导致先验信息失真的问题，设计双网络协同迭代反馈机制，通过人脸重建网络（face reconstruction network, FRN）与关键点估计网络（keypoint estimation network, KEN）的交互式反馈，逐步修正关键点位置并提升重建质量；其次，针对单一先验信息约束有限的局限，构建结构纹理融合约束模块（structure-texture fusion constraint module, STFCM），将语义级关键点先验与像素级边缘信息进行多层次融合，实现从整体轮廓到局部细节的全方位约束；最后，针对传统注意力机制难以兼顾局部与全局特征的问题，采用自适应多尺度注意力机制（adaptive multi-scale attention, AMSA），通过可学习的权重参数动态调整局部卷积特征与全局非局部特征的贡献比例，增强网络对多尺度面部特征的表达能力。上述三个模块相互配合，协同作用，共同提升 FSR 重建的准确性和真实感。

1 方法

1.1 整体架构

本文提出的重建方法主要由 FRN 和 KEN 两部分组成，通过双网络协同迭代反馈机制实现交互式优化。所提出方法的整体架构如图 1 所示，FRN 负责从 LR 输入逐步恢复 HR 人脸图像，包含初始上采样模块（initial upsampling module, IUM）、STFCM 和最终上采样模块（final upsampling module, FUM），其中 STFCM 由自适应多尺度注意力块（adaptive

1. 郑州西亚斯学院工学部 河南郑州 451100
2. 河南省智能制造数字孪生工程研究中心 河南郑州 451150
[基金项目] 郑州西亚斯学院 2024 年度科研资助项目（2024XKE117）；2025 年河南省科技发展计划（科技攻关）项目（252102411005）；2025 年度河南省高等学校重点科研项目（25B520016）

multi-scale attention block, AMSAB）、区域特征调制块（region feature modulation block, RFMB）和轮廓增强重建块（contour enhancement reconstruction block, CERB）组成，RFMB 依赖关键点热图输入。KEN 由预处理模块（preprocessing module, PrePM）、增强沙漏模块（refinement hourglass module, RHM）和后处理模块（post-processing module, PostPM）组成，从重建的人脸图像中提取关键点热图，为人脸重建提供几何约束。此外，FRN 和 KEN 分别构建了一条反馈连接，将 STFCM 和 RHM 本次迭代的输出用于下一次模块的输入，实现信息迭代传递和优化。

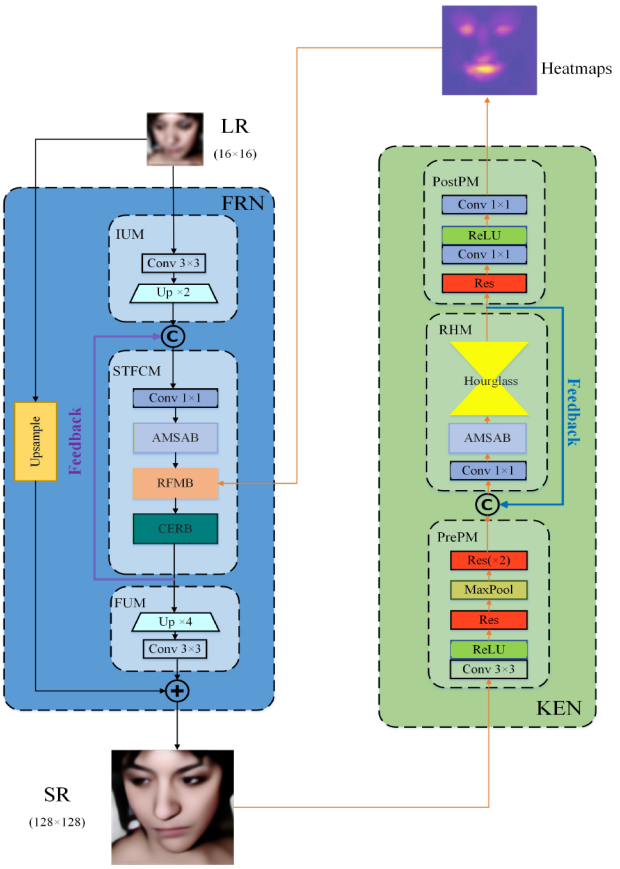

图 1 所提方法整体架构图

整体流程采用 8 倍上采样策略：FRN 首先通过 IUM 将 16×16 的 LR 输入放大至 32×32，然后通过 STFCM 进行特征增强和高频细节恢复，最后由 FUM 生成 128×128 的 HR 输出。之后，HR 输出被进一步送入 KEN 中，以提取关键点热图。在迭代过程中，KEN 从当前迭代的重建结果中提取关键点热图，并将其反馈至 FRN，指导下一轮迭代的人脸重建。

1.2 双网络协同迭代反馈机制

LR 输入致使先验估计不准确，遂设计双网络协同迭代反馈机制。在具体实现进程里，每次迭代有三个步骤。其一，FRN 借助上一次迭代的关键点热图生成初步重建结果；其二，KEN 剖析重建结果，提取更精准的关键点信息；其三，将更新后的关键点热图送入 FRN，以优化下一次重建。经由这种迭代过程，能逐步提高关键点估计的准确性、人脸重建的质量。

对于 n 次迭代，迭代反馈过程公式为：

$$I_{SR}^{n} = \mathrm{FUM}\left(\mathrm{STFCM}\left(\mathrm{IUM}\left(I_{LR}\right), F_{STFCM}^{n-1}, H^{n-1}\right)\right) + \mathrm{UP}\left(I_{LR}\right) \quad (1)$$

$$H^{n} = \mathrm{PostPM}\left(\mathrm{RHM}\left(\mathrm{PrePM}\left(I_{SR}^{n}\right), R_{RHM}^{n-1}\right)\right) \quad (2)$$

式中：I_{LR} 为低分辨率输入，I_{SR}^{n} 为第 n 次迭代的人脸重建结果，H^{n} 为第 n 次迭代的关键点热图，F_{STFCM}^{n-1} 为 STFCM 的上次输出，用于下一次迭代的输入，R_{RHM}^{n-1} 为 RHM 的上次输出，用于下一次迭代的输入，UP 为上采样操作，用于复用 LR 特征。

在初始状态下（n=0），由于缺乏关键点热图引导，FRN 移除 RFMB，仅保留 IUM、CERB 和 FUM，生成初始重建结果。随后，KEN 提取初始面部关键点热图，启动迭代过程。

1.3 结构纹理融合约束模块 STFCM

为了发挥出边缘信息、关键点信息之间的互补作用，从而设计出了 STFCM。此模块涵盖三个核心组件，分别是 AMSAB、RFMB、CERB。

1.3.1 自适应多尺度注意力块 AMSAB

人脸特征既包含局部纹理细节，又存在全局结构依赖。传统卷积网络在捕获局部特征方面表现比较出色，然而在对长距离依赖关系进行建模操作的时候却存在一定局限[8]。基于此，本文使用 AMSAB，借助可学习参数来动态调整局部与非局部特征的权重，进而让网络能够同时兼顾局部细节、全局结构，AMSAB 的架构如图 2 所示。

首先，利用 3×3 卷积提取局部特征：

$$F_{local} = \mathrm{Conv}_{3\times3}(F_{in}) \quad (3)$$

随后，通过三个独立的 1×1 卷积将输入映射为查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$、值 $\boldsymbol{V}$，并基于缩放点积注意力计算非局部特征：

$$\boldsymbol{Q} = \mathrm{Conv}_{1}^{Q}(F_{in}), \quad \boldsymbol{K} = \mathrm{Conv}_{1}^{K}(F_{in}), \quad \boldsymbol{V} = \mathrm{Conv}_{1}^{V}(F_{in}) \quad (4)$$

$$\boldsymbol{A} = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right), \quad F_{nonlocal} = \boldsymbol{A}\boldsymbol{V} \quad (5)$$

式中：d_k 为键特征维度，$\boldsymbol{A}$ 为注意力权重矩阵。进而，通过可学习参数 α 自适应融合局部与非局部特征：

$$F_{attn} = \alpha F_{local} + (1-\alpha)F_{nonlocal} \quad (6)$$

式中：$\alpha=\sigma(\alpha_0)$，α_0 为可学习标量，其作用是对两类特征的重要性予以动态平衡。最终，借助门控机制、残差连接从而获取到输出的结果：

$$F_{out} = F_{attn} \odot F_{in} + F_{in} \quad (7)$$

式中：$\odot$ 表示元素级乘法，F_{out} 表示 AMSAB 的最终输出。残差连接能够对深层网络里的梯度消失问题起到缓解作用，并保持原始特征信息。

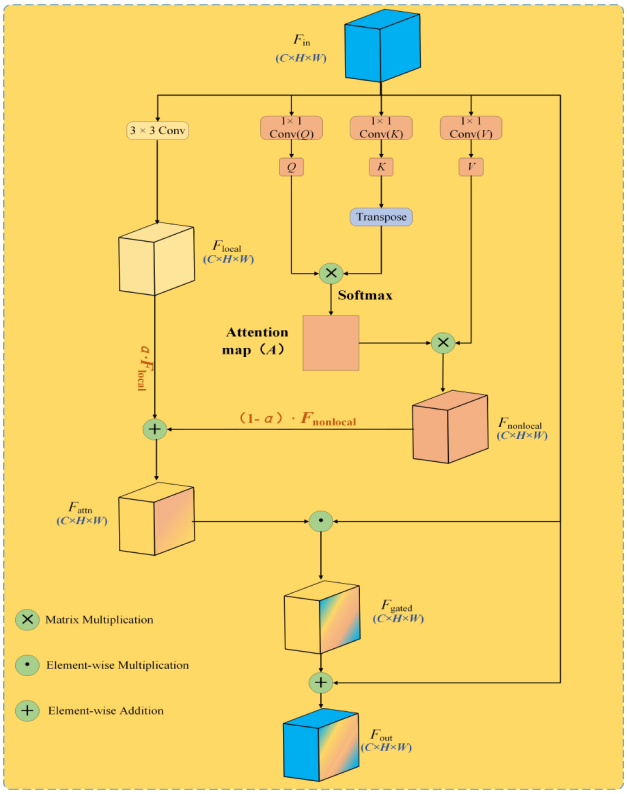

图 2 AMSAB 架构图

1.3.2 区域特征调制块 RFMB

传统方法是直接把关键点热图和特征图进行拼接，这样很难达成精确的组件级调制。而 RFMB 会把 68 通道的关键点热图按照面部组件来分组，利用 Softmax 函数去计算组件注意力权重，针对不同的面部区域实施差异化调制，以此增强眼睛、鼻子、嘴巴等关键区域的重建效果。

面部热图的划分公式为：

$$H_i = \sum_{j\in G_i} h_j \tag{8}$$

式中：H_i 表示第 i 个面部组件的聚合热图，h_j 表示第 j 个关键点的热图，G_i 表示第 i 个面部组件包含的关键点索引集合。

区域特征调制过程的计算公式为：

$$M_i = \mathrm{Softmax}(H_i) \tag{9}$$

$$F_i = \mathrm{GConv}_i(F_{\mathrm{AMSAB}}) \tag{10}$$

$$F_{\mathrm{RFMB}} = \sum_{i=1}^{K} M_i \odot F_i \tag{11}$$

式中：M_i 表示第 i 个组件的注意力热图，GConv_i 表示第 i 组的分组卷积，F_{AMSAB} 表示 RFMB 的输入特征，K 表示组件数量，F_i 表示调制后的第 i 组特征，F_{RFMB} 表示区域特征调制块的输出。

1.3.3 轮廓增强重建块 CERB

边缘信息对保持轮廓锐度至关重要，但现有方法通常依赖独立边缘提取网络，增加计算负担。本文采用轻量化边缘模块，通过平均池化操作提取边缘特征，为重建过程提供像素级结构约束。边缘块架构如图 3 所示，首先通过平均池化层对输入特征进行平滑，然后将原始特征减去平滑特征获得边缘特征，最后将边缘特征与输入拼接作为输出。采用核大小分别为 5 和 9 的平均池化层，进一步设计低尺度边缘块（low-scale edge block, LEB）和高尺度边缘块（high-scale edge block, HEB），分别用于提取低分辨率特征和高分辨率特征中的边缘信息。

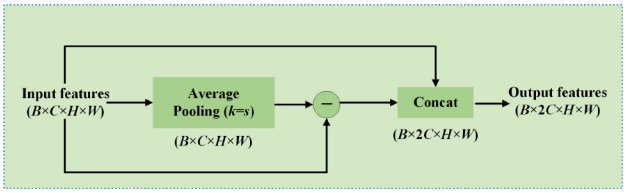

图 3 边缘块架构图

为整合低分辨率和高分辨率特征的互补信息，本文采用多投影组设计以构建 CERB。通过在不同尺度投影组中插入边缘块提取多尺度边缘信息，增强轮廓恢复效果。

定义 F_{LR}^{K}、F_{HR}^{k}、E_{LR}^{K} 和 E_{HR}^{k} 分别为第 K 个投影组的低分辨率特征、高分辨率特征、低分辨率边缘特征和高分辨率边缘特征。投影组数量 K=3 时的 CERB 架构如图 4 所示。

计算过程为：

$$F_{\mathrm{HR}}^{k} = \mathrm{DeConv}_k(\mathrm{LConv}_{1\times1}^{k}(\mathrm{Concat}(F_{\mathrm{LR}}^{k-1}, E_{\mathrm{LR}}^{k-1}))) \tag{12}$$

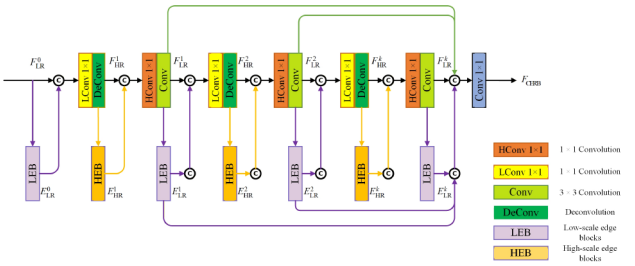

图 4 CERB 架构图

$$F_{\mathrm{LR}}^{k} = \mathrm{Conv}_{3\times3}^{k}(\mathrm{HConv}_{1\times1}^{k}(\mathrm{Concat}(F_{\mathrm{HR}}^{k}, E_{\mathrm{HR}}^{k}))) \tag{13}$$

最后，融合各投影组的特征得到输出：

$$F_{\mathrm{CERB}} = \mathrm{Conv}_{1\times1}(\mathrm{Concat}([F_{\mathrm{LR}}^{1}, F_{\mathrm{LR}}^{2}, ..., F_{\mathrm{LR}}^{K}], [E_{\mathrm{LR}}^{1}, E_{\mathrm{LR}}^{2}..., E_{\mathrm{LR}}^{K}])) \tag{14}$$

CERB 的输出 F_{CERB} 即为 STFCM 的最终输出。

1.4 损失函数设计

为平衡重建精度和感知质量，本文采用联合损失函数优化模型训练，包括重建损失、热图损失、感知损失和对抗损失。

重建损失 L_{rec} 采用 L_1 范数度量重建图像与真实图像的像素差异：

$$L_{\mathrm{rec}} = \frac{1}{N}\sum_{i=1}^{N}\left\| I_{\mathrm{SR}}^{i} - I_{\mathrm{GT}}^{i} \right\|_1 \tag{15}$$

热图损失 L_{heat} 采用 L_2 范数约束预测关键点热图与真实热图的一致性：

$$L_{\mathrm{heat}} = \frac{1}{N}\sum_{i=1}^{N}\left\| H_{\mathrm{SR}}^{i} - H_{\mathrm{GT}}^{i} \right\|_2 \tag{16}$$

感知损失 L_{per} 利用预训练 LightCNN 提取高层特征，优化感知质量：

$$L_{per}=\frac{1}{N}\sum_{i=1}^{N}\left\|\phi(I_{SR}^{i})-\phi(I_{GT}^{i})\right\|_2 \tag{17}$$

式中：$\phi(\cdot)$ 表示预训练 LightCNN 提取的特征，对抗损失引入判别器网络增强重建结果的真实感，判别器损失 L_D 和生成器损失 L_G 分别定义为：

$$L_D=-\mathbb{E}[\log D(I_{GT})]-\mathbb{E}[\log(1-D(G(I_{LR})))] \tag{18}$$

$$L_G=-\mathbb{E}[\log D(G(I_{LR}))] \tag{19}$$

式中：$D(\cdot)$ 和 $G(\cdot)$ 分别表示判别器和生成器。总体损失函数为：

$$L_{total}=L_{rec}+\lambda_1 L_{heat}+\lambda_2 L_{per}+\lambda_3 L_{adv} \tag{20}$$

2 实验

2.1 实验和超参数设置

本文在 CelebA 和 Helen 两个公开数据集上进行训练和测试，所有图像裁剪至 128×128 分辨率，通过双三次下采样生成 16×16 的低分辨率输入，实现 8 倍放大重建。使用 OpenFace 工具提取人脸关键点[9]。评估指标包括峰值信噪比（peak signal-to-noise ratio, PSNR）、结构相似性（structural similarity index, SSIM）、学习感知图像块相似度（learned perceptual image patch similarity, LPIPS）。实验环境为 NVIDIA TITAN X GPU，采用 PyTorch 框架实现。超参数设置如表 1 所示。

表 1 超参数设置

参数名称	设置值
迭代次数 N	3
关键点数量	68
投影组数量 K	6
中间特征通道	48
边缘块池化核大小 s	LEB:5, HEB:9
优化器	Adam (β_1=0.9，β_2=0.99，ε=10^{-8})
初始学习率	1×10^{-4}
批处理大小	16
总迭代次数	150 000
学习率衰减	50 000/100 000/130 000 减半
热图损失权重 λ_1	1×10^{-2}
感知损失权重 λ_2	1×10^{-1}
对抗损失权重 λ_3	5×10^{-3}

2.2 对比实验

在 CelebA 和 Helen 数据集上与当前主流方法进行对比，包括面向 PSNR 的方法（MSFSR[10]、DIC[11]、UFSRNet[12]）和面向感知质量的方法（MRRNet[13]、DICGAN[11]）。PSNR 和 SSIM 对比结果如表 2 所示。LPIPS 对比结果如表 3 所示。

表 2 PSNR/SSIM 对比结果

方法	CelebA		Helen	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
MSFSR	26.26	0.753 1	25.53	0.720 2
MRRNet	26.62	0.739 8	26.48	0.778 0
DIC	26.87	0.790 4	26.94	0.799 3
DICGAN	25,96	0.741 7	25.87	0.753 2
UFSRNet	27.20	0.794 8	27.02	0.802 1
本文方法	27.52	0.799 8	27.21	0.808 9
本文方法（GAN）	26.23	0.747 8	26.15	0.763 2

从表 2 可以看出，本文方法在两个数据集上均取得最优的 PSNR 和 SSIM。在 CelebA 上 PSNR 达 27.52 dB，较次优方法 UFSRNet 提升 0.32 dB；在 Helen 上 PSNR 达 27.21 dB，提升 0.19 dB。表明 CERB 通过融合关键点约束与边缘信息，有效提升了像素级重建精度。本文方法（GAN）的 PSNR 下降属于 GAN 引入感知损失后的典型现象。

表 3 LPIPS 对比结果

方法	CelebA	Helen
	LPIPS↓	LPIPS↓
MRRNet	0.096 0	0.093 1
DICGAN	0.095 9	0.093 6
本文方法（GAN）	0.085 1	0.090 2

从表 3 可以看出，本文方法（GAN）在两个数据集上均取得最低的 LPIPS，优于 MRRNet 和 DICGAN，表明关键点约束与边缘信息的引入有效提升了面部细节的感知质量。综合来看，本文方法在像素级精度和感知质量两个维度均表现最优，验证了融合关键点约束与边缘信息的有效性。

图 5 展示了不同方法的 8 倍人脸超分辨率重建结果。从局部放大区域可以看出，MSFSR 和 MRRNet 重建结果较为模糊，面部轮廓和细节丢失严重；DIC 和 UFSRNet 虽能恢复大致轮廓，但纹理细节仍不够清晰；DICGAN 存在明显伪影和纹理失真。本文方法在面部轮廓和细节恢复上更接近真实高分辨率图像（ground truth, GT），本文方法（GAN）在感知质量上进一步提升，面部细节更为自然清晰。

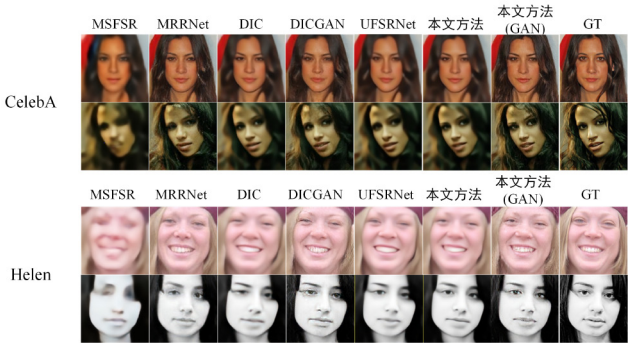

图 5 8 倍人脸超分辨率重建结果对比

2.3 消融实验

为验证各模块的有效性，设计三个变体在Helen数据集上进行消融研究：Baseline（移除CERB、RFMB和AMSAB）、Model1（在Baseline基础上添加RFMB）、Model2（在Model1基础上添加AMSAB）。消融实验结果如表4所示。

表4 消融实验

方法	PSNR↑	SSIM↑
Baseline	26.81	0.789 7
Model1	26.98	0.799 3
Model2	27.10	0.800 7
本文方法	27.21	0.808 9

从表4可以看出，各模块的引入均带来了性能提升。Model1较Baseline提升0.17 dB，表明RFMB的多尺度特征融合有效增强了特征表达；Model2较Model1提升0.12 dB，表明AMSAB的自适应注意力机制有助于关键区域的特征聚合；本文方法较Model2进一步提升0.11 dB，表明CERB融合关键点约束与边缘信息有效提升了重建精度。各模块逐步叠加均带来增益，验证了其各自的有效性。

3 结论

本文针对人脸超分辨率中关键点估计不准确和边缘高频细节恢复不足的问题，提出一种融合关键点约束和边缘信息的重建算法。该方法通过双网络协同迭代反馈机制实现关键点估计与人脸重建的相互增强，利用结构纹理融合约束模块将语义级关键点先验与像素级边缘信息进行多层次融合，并采用自适应多尺度注意力机制增强多尺度特征表达能力。CelebA和Helen数据集上的实验结果显示，所提方法在PSNR、SSIM和LPIPS上均领先于对比方法，重建人脸的边缘更清晰、结构更完整。未来工作将从模型轻量化和实际监控场景部署两个方向开展进一步研究。

参考文献：

[1] Liu Jinbo, Liu Zhonghua, Ou Weihua, et al. CFGPFSR: a generative method combining facial and GAN priors for face super-resolution[J/OL]. Neural Processing Letters, 2024[2026-06-17]. https://link.springer.com/article/10.1007/s11063-024-11562-8.

[2] Chen Yu, Tai Ying, Liu Xiaoming, et al. FSRNet: end-to-end learning face super-resolution with facial priors[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2018: 2492-2501.

[3] Wang Chenyang, Jiang Junjun, Zhong Zhiwei, et al. Super-resolving face image by facial parsing information[J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2023, 5(4): 435-448.

[4] Qu Chengchao, Monari E, Schuchert T, et al. Patch-based facial texture super-resolution by fitting 3D face models[J]. Machine Vision and Applications, 2019, 30: 557-586.

[5] Hajian A, Aramvith S. MSRFSR: multi-stage refining face super-resolution with iterative collaboration between face recovery and landmark estimation[J]. IEEE Access, 2024, 12: 56951-56972.

[6] Liu Hao, Yang Yang, Liu Yunxia. W-Net: a facial feature-guided face super-resolution network[J]. Image and Vision Computing, 2025, 159: 105549.

[7] Li Ling, Zhang Yan, Yuan Lin, et al. SANet: face super-resolution based on self-similarity prior and attention integration[J]. Pattern Recognition, 2025, 157: 110854.

[8] Liu Yuhao, Chu Zhenzhong. A dynamic fusion of local and non-local features-based feedback network on super-resolution[J]. Symmetry. 2023, 15(4): 885.

[9] Baltrusaitis T, Zadeh A, Lim Y C, et al. OpenFace 2.0: facial behavior analysis toolkit[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018).NewYork:ACM,2018: 59-66.

[10] Zhang Yunchen, Wu Yi, Chen Liang. MSFSR: a multi-stage face super-resolution with accurate facial representation via enhanced facial boundaries[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.Piscataway:IEEE,2020: 504-505.

[11] Ma Cheng, Jiang Zhenyu, Rao Yongming, et al. Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation[PP/OL].(2020-03-29)[2026-06-17].https://doi.org/10.48550/arXiv.2003.13063.

[12] Wang Tongguan, Xiao Yang, Cai Yuxi, et al. UFSRNet: u-shaped face super-resolution reconstruction network based on wavelet transform[J]. Multimedia Tools and Applications, 2024, 83: 67231-67249.

[13] Huang Weikang, Lan Shiyong, Wang Wenwu, et al. Face super-resolution with spatial attention guided by multiscale receptive-field features[C]//Artificial Neural Networks and Machine Learning – ICANN 2022.Berlin:Springer,2022: 145-157.

【作者简介】

刘锦博（1999—），通信作者（email:18538710341@163.com），男，河南郑州人，硕士研究生，助教，研究方向：计算机视觉、深度学习。

肖露（1997—），女，河南南阳人，硕士研究生，助教，研究方向：计算机视觉、人工智能。

邓锴（1997—），男，河南南阳人，硕士研究生，助教，研究方向：算力网络、未来网络。

高天宇（2007—），男，河南濮阳人，本科，研究方向：计算机视觉、人工智能。

（收稿日期：2026-05-13 修回日期：2026-06-30）