



一种用于联合低光增强和人脸超分的深度学习网络

丛维仪, 郑卓然, 贾修一

引用本文:

丛维仪, 郑卓然, 贾修一. 一种用于联合低光增强和人脸超分的深度学习网络[J]. 智能系统学报, 2025, 20(1): 109–117.

CONG Weiyi, ZHENG Zhuoran, JIA Xiuyi. A deep learning network for joint low-light enhancement and face super-resolution[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(1): 109–117.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202406029>

您可能感兴趣的其他文章

基于双注意力模型和迁移学习的Apex帧微表情识别

Apex frame microexpression recognition based on dual attention model and transfer learning

智能系统学报. 2021, 16(6): 1015–1020 <https://dx.doi.org/10.11992/tis.202010031>

自步稀疏最优均值主成分分析

Sparse optimal mean principal component analysis based on self-paced learning

智能系统学报. 2021, 16(3): 416–424 <https://dx.doi.org/10.11992/tis.201911028>

神经网络多层特征信息融合的人脸识别方法

Face recognition method based on neural network multi-layer feature information fusion

智能系统学报. 2021, 16(2): 279–285 <https://dx.doi.org/10.11992/tis.201907053>

基于加权边缘弱化引导滤波的人脸光照补偿

Face illumination compensation based on weighted edge-weakening guided image filter

智能系统学报. 2018, 13(3): 373–379 <https://dx.doi.org/10.11992/tis.201612011>

一种新融合算法的维吾尔族人脸识别

A new fusion algorithm for uyghur face recognition

智能系统学报. 2018, 13(3): 431–436 <https://dx.doi.org/10.11992/tis.201710014>

行人重识别研究综述

Survey on pedestrian re-identification research

智能系统学报. 2017, 12(6): 770–780 <https://dx.doi.org/10.11992/tis.201706084>

DOI: 10.11992/tis.202406029

一种用于联合低光增强和人脸超分的深度学习网络

丛维仪, 郑卓然, 贾修一

(南京理工大学 计算机科学与工程学院, 江苏 南京 210094)

摘要: 在低光环境下, 人脸图像增强是许多任务的重要恢复方法。然而, 现有的低光环境下人脸超分辨率方法通常依赖于低光增强和超分算法的序列建模。遗憾的是, 由于优化目标之间的差异, 使用这种方法来增强人脸图像很容易导致伪影或噪声。为了应对这一挑战, 本文提出了一个端到端的低光人脸图像超分辨率网络 (low-light face super resolution network, LFSRNet)。该网络由浅层特征提取、深层特征提取和特征过滤上采样 3 个模块组成。首先浅层特征模块将输入的低光、低分辨率人脸图像映射到特征空间。随后, 深度特征提取模块对其进行亮度校正并细化结构。最后, 特征过滤上采样模块处理提取到的特征并重建人脸图像。此外, 为了更好地重建丢失的面部细节本文还设计了一个损失函数 faceMaskLoss。大量实验证明了所提模型的有效性。

关键词: 人脸超分辨率; 低光图像增强; 监督学习; 随机掩码; 损失函数; 深度学习; 局部特征提取; 全局特征提取
中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2025)01-0109-09

中文引用格式: 丛维仪, 郑卓然, 贾修一. 一种用于联合低光增强和人脸超分的深度学习网络 [J]. 智能系统学报, 2025, 20(1): 109-117.

英文引用格式: CONG Weiyi, ZHENG Zhuoran, JIA Xiuyi. A deep learning network for joint low-light enhancement and face super-resolution[J]. CAAI transactions on intelligent systems, 2025, 20(1): 109-117.

A deep learning network for joint low-light enhancement and face super-resolution

CONG Weiyi, ZHENG Zhuoran, JIA Xiuyi

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: In low-light environments, face image enhancement is used as a vital recovery method for many tasks. However, existing methods for face super-resolution in low-light environments usually relied on sequence modeling that combines low-light enhancement and super-resolution algorithms. Unfortunately, using this method to enhance a face image easily led to artifacts or noise because of the differences between the optimization objectives. To tackle this challenge, we proposed LFSRNet, an end-to-end low-light face image super-resolution network. Our network consisted of three modules: shallow feature extraction, deep feature extraction, and feature filtering upsampling. The shallow feature module initially mapped the input low-light, low-resolution face image into feature space. Subsequently, the deep feature extraction module performed luminance correction and refined the structure. Finally, the feature filtering upsampling module processed the extracted features and reconstructed the face image. Additionally, in order to better reconstruct the lost facial details, we also designed a loss function faceMaskLoss. Extensive experiments demonstrate the effectiveness of our proposed model.

Keywords: face super resolution; low-light image enhancement; supervised learning; random mask; loss function; deep learning; local feature extraction; global feature extraction

目前有许多基于人脸图像的视觉任务, 比如人脸对齐^[1-3]、人脸解析^[4-6]、人脸识别^[7-8] 和 3D 人

脸重构^[9-11]。这些任务的效果往往都会受到输入人脸图像质量的影响。然而, 由于光照和设备等因素的限制, 在低光环境下获取的人脸图像往往具有低分辨率, 低对比度的特点。

近些年来, 人脸图像超分 (face super resolution,

收稿日期: 2024-06-18.

基金项目: 国家自然科学基金项目 (62176123, 62476130).

通信作者: 贾修一. E-mail: jiaxy@njut.edu.cn.

©《智能系统学报》编辑部版权所有

FSR)被提出用来将低分辨率人脸映射回高分辨率人脸。得益于深度学习的发展,已有许多优秀的工作被提出来解决FSR任务。借助于注意力机制,Chen等^[12]通过关注人脸图像中不同尺度下特征来加强细节的恢复。与常规图像相比,人脸图像往往拥有固定的结构,可以通过某些先验(人脸特征图、人脸热力图等)来恢复图像,因此,Chen等^[13]提出一种融合先验估计和人脸超分的卷积神经网络(convolutional neural network, CNN),通过约束人脸结构信息实现高分辨率人脸的修复。考虑到恢复人脸需要综合局部和全局信息,Gao等^[14]提出一种CNN-Transformer结构的网络,借助设计的面部结构注意力单元高效地提取特征信息。然而,包含上述方法在内的大多数FSR工作在设计时都仅考虑正常光照下的图像,忽视了低光这一常见的情况。

如何对低光人脸图像进行超分是一个问题。一种方法是将其分解为两种相对常见的任务,即低光增强和人脸图像超分。因此,串联执行这两种任务是一种解决问题的思路。根据执行顺序不同,可以分为两种情况:低光增强+图像超分,图像超分+低光增强。然而因为两种任务的优化目标并不完全一致,所以会出现次优化的问题。另一种思路是为低光人脸超分方法设计专门的算法,也有部分学者做出了尝试。Guo等^[15]提出一种基于生成式对抗网络(generative adversarial networks, GAN)的低光人脸超分方法,通过一个额外网络粗略恢复人脸,然后再进行超分操作,然而,这种方式与前面提及的二阶段方法几乎一致。Ding等^[16]则从另一种思路入手,借助面部特征先验补充缺失的信息,但是为了约束信息,作者在文中使用了多达8个损失函数,极大增加了训练的不稳定性。因此,提出一种新的端到端的低光人脸超分方法十分重要。

本文提出一种新的低光人脸超分网络(low-light face super resolution network, LFSRNet)来解决上述问题。与经典的超分模型类似,它由浅层特征提取、深层特征提取和特征过滤上采样3个模块组成。浅层特征提取模块使用卷积提取浅层信息并将提取到的低频信息传递到后续模块。鉴于低光照会使人脸图像更加难以提取,所以本文设计了一个基于CNN的亮度矫正人脸结构信息提取模块。此外,为了综合全局和局部信息,模型将其与Transformer相结合组成模型中的深度特征提取模块。最后,特征过滤上采样模块会对浅层和深层特征做进一步的融合得到重构后的正常光照、高分辨率的人脸图像。此外,为了提高恢

复人脸的质量,本文还设计了一种新的人脸掩码损失来约束重建过程。

1 相关工作

1.1 低光增强

低光增强旨在将欠曝图像恢复成同等大小正常光照下的图像。现有的低光增强工作可以被分为传统方法^[17-20]和基于深度学习的方法^[21-24]。Lore等^[25]通过堆叠稀疏去噪自动编码器提出了LL-Net(light-weight lane line detection network),能够同时实现低光增强和图像去噪。结合Retinex理论,Wei等^[26]提出了RetinexNet,通过分解、调整和重构3个阶段完成对图像中照度分量的矫正。Wang等^[27]则将低光增强看作一个残差学习问题并提出了DLN(deep lightening network)模型,通过迭代地执行亮化和暗化过程,学习低照度图像和正常光照图像之间的残差,然后借助于学习到的残差修复图像。这些方法只能修复图像中的光照信息而不可以提高图像的分辨率。

1.2 人脸图像超分

人脸图像超分的目标是将低分辨率的人脸图像重构成为高分辨率的图像。近些年来,已有大量的方法被提出来解决这个任务^[28-32]。根据是否使用人脸图像的特殊信息,现有方法可以被分为一般的方法和基于先验的方法。Lu等^[33]提出了一种基于分散注意力的方法SISN(split-attention in split-attention network),通过对特征通道的拆分,使得网络能够关注于信息丰富的区域并加强对这些区域的修复。Gao等^[14]则认为现有的串行结构无法高效地利用low-level的信息,同时修复面部信息需要综合局部和全局的信息,因此提出了一种CNN-Transformer混合的U-Net的人脸超分网络。与一般方法不同,基于先验的方法希望通过引入人脸先验来增强对恢复图像细节的约束。Chen等^[13]首先将人脸特征热图和人脸解析图作为先验引入到FSR任务中,通过约束生成图像和GT(ground truth)之间的先验距离来引导图像重构。Yang等^[34]则考虑了另一种引入先验的方式,将GAN模型先在HQ(high quality)图像上预训练,然后将其作为解码器嵌入到网络中来提供足够的信息完成重构。然而现有的人脸图像超分方法只关注正常光照下的人脸图像,很少有方法考虑到低光这一常见的退化因素。

1.3 低光人脸超分

相较于正常光照下的人脸超分,低光人脸超分目前很少有人研究。Wang等^[35]考虑到串联低

光增强和人脸超分方法的不足, 设计了一个双工特征融合模块, 该方法双向融合并嵌入低光增强和人脸超分之间的特征, 从而提升了二阶段方法的效果。Guo 等^[15]提出了一种两阶段的方法 IEF-SR (illumination-enhanced face super-resolution) 来恢复低光人脸图像, IEF-SR 首先使用一个粗恢复网络调节输入图像的亮度, 然后使用一个多尺度的 GAN 模型渐进式地生成高分辨率人脸。Ding 等^[16]则考虑利用先验信息, 借助 U-Dense 和 Hour-Glass 设计了一种端到端的低光人脸超分网络, 该方法在重构图像的同时引入了先验估计和低光增强, 从而缓解了低光对超分任务的影响。与上述方法不同的是, 本文所提模型在考虑图像的局部

和全局信息的同时, 设计了一个人脸掩码损失函数来指导模型的学习。

2 本文方法

本节中, 首先会对所提的 LFSRNet 的整体结构进行叙述, 然后详细解释模型的组成部分, 最后介绍所提的人脸掩码损失函数以及其他用来约束方法的损失函数。

2.1 模型整体架构

在介绍模型时, 本文使用 $I_{LR} \in \mathbb{R}^{3 \times h \times w}$, $I_{SR} \in \mathbb{R}^{3 \times H \times W}$ 和 $I_{HR} \in \mathbb{R}^{3 \times H \times W}$ 分别表示输入图像、模型生成的输出图像和相对应的真实图像, 模型的整体结构如图 1 所示。

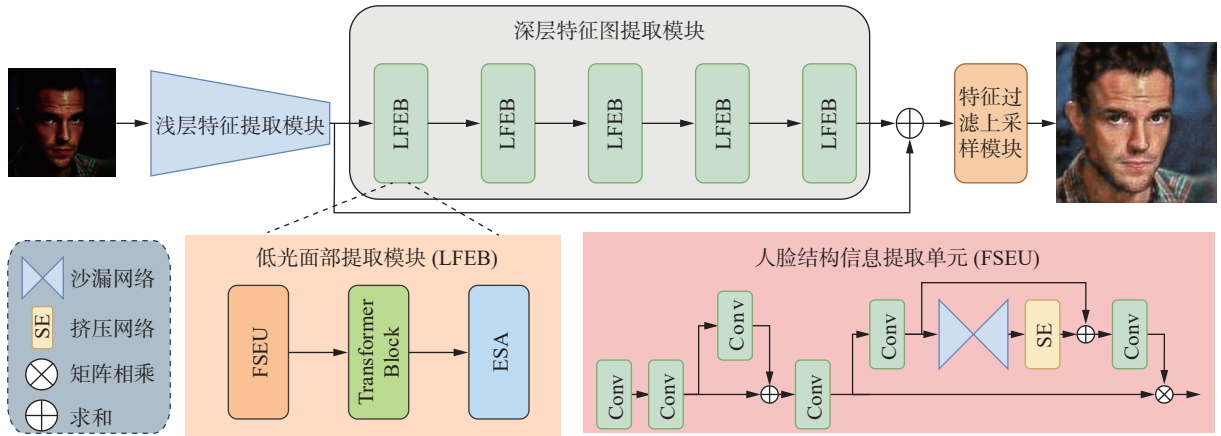


图 1 LFSRNet 模型的框架

Fig. 1 Structure of the LFSRNet model

主要包含 3 部分: 浅层特征提取模块、深层特征提取模块和特征过滤上采样模块。对于一张输入的低光低分辨率图像 I_{LR} , 模型使用一个 3×3 的卷积来提取得到浅层特征 $F_0 \in \mathbb{R}^{C \times h \times w}$:

$$F_0 = \text{Conv}_{3 \times 3}(I_{LR})$$

式中 C 是特征图的通道维度。

然后, 模型通过深度特征提取模块提取深度特征 $F_{DF} \in \mathbb{R}^{C \times h \times w}$:

$$F_{DF} = H_{DF}(F_0)$$

式中 H_{DF} 是由 N 个堆叠的低光人脸特征图提取块 (low-light face extraction blocks, LFEB) 组成, 有关于 LFEB 的细节会在后续小节进行阐述。

之后, 模型将浅层特征 F_0 与深度特征 F_{DF} 相融合得到的 $F_S \in \mathbb{R}^{C \times h \times w}$:

$$F_S = F_0 + F_{DF}$$

最后, 通过一个特征过滤上采样模块重构得到 I_{SR} 。

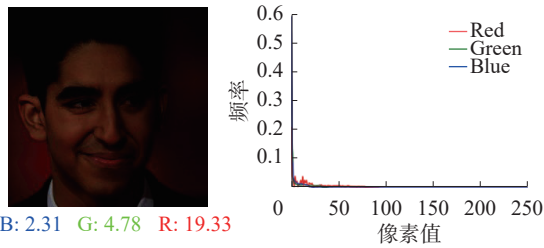
$$I_{SR} = H_{UP}(F_S)$$

2.2 深层特征提取

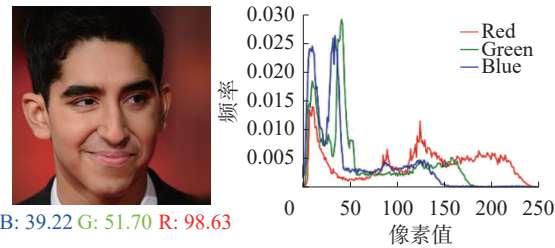
人脸细节的修复通常需要融合局部和全局信

息。目前基于纯卷积的方法在建模局部信息上表现得很出色, 但是难以捕获全局信息; 同时 Transformer^[36] 因为其优秀的全局建模能力而被广泛用于计算机视觉任务。因此, 本文设计了一个 CNN-Transformer 混合架构的低光人脸信息提取块 (low-light face extraction block, LFEB)。它的结构在图 1 中给出, 主要包含一个人脸结构信息提取单元 (facial structure extraction unit, FSEU), 一个 Transformer Block 和一个 ESA (enhanced spatial attention) 块。在本文中, 模型使用 5 个 LFEB 组成方法中的深层特征提取模块。

图 2 给出了一张正常光照下人脸图像及其对应低光版本的 RGB 色彩直方图。可以发现, 低光图像三通道的像素值主要集中在图像的左边, 图 2 中人脸图像下对应的颜色均值也能表明这点。这表明亮度矫正的过程可以被理解为像素值增加的过程。因此, 如图 1 所示, 本文在 FSEU 的前半部分使用了一种投影相加的方法对输入特征图进行亮度矫正。



(a) 低光环境下人脸图像及其通道像素平均值



(b) 正常光照下人脸图像及其通道像素平均值

图 2 不同光照条件下的色彩直方图和通道像素平均值

Fig. 2 Color histograms and channel pixel averages under different lighting conditions

相较于自然图像,人脸图像往往具有固定的结构并且模型可以通过提取这些固定结构来恢复人脸细节。具体来说,模型使用 HourGlass^[37] 和 SE(squeeze and excitation) 组成了一个 人脸结构注意力分支,使得模型能够更加关注包含人脸特征的通道。光图像往往被暗光掩盖了许多关键信息,亮度矫正和人脸结构信息提取的顺序是经过考虑的。因此,本文先对输入图像做亮度矫正然后进行结构信息的提取。

为了进一步提取全局信息,模型引入了多轴视觉 Transformer^[38](multi-axis vision Transformer, MaxViT) 模块并将其结构在图 3 中给出,图 3 中 FFN(feed forward network)代表线性层。

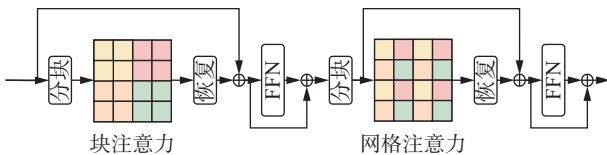


图 3 多轴视觉 Transformer 结构

Fig. 3 Structure of multi-axis vision Transformer

MaxViT 利用换轴自注意力(multi-axis self-attention, Max-SA)来建模像素之间的关系。对于一个给定的输入图像,其特征图为 $X \in \mathbb{R}^{C \times h \times w}$, 它首先被调整为 $\begin{bmatrix} C & \frac{h \times w}{P \times P} & P \times P \end{bmatrix}$, 也就是被划分为 $\frac{h \times w}{P \times P}$ 个 $P \times P$ 大小的图像块,然后对每个块进行块自注意力操作。显然,这种方式虽然避免了执行全局注意力的高额花费,但是因为只交互图像块内部的信息,所以表示能力会有所下降。为了

减少局部注意力带来的负面效果,MaxViT 转化坐标轴维度并将特征图转为 $\begin{bmatrix} C & G \times G & \frac{h \times w}{G \times G} \end{bmatrix}$, 之后在大小 $G \times G$ 的网格中执行网格注意力操作。值得注意的是,虽然网格注意力操作也是只对局部执行自注意力,但是从图 3 不难看出,网格中混合了其他图像块中的像素点,因此网格注意力也可以被认为是一种全局注意力。借助于 MaxViT 块的帮助,模型能够以较小的计算成本完成对全局信息的建模。最后,模型使用一个 ESA^[39-40] 块对特征做进一步的提升。

2.3 特征过滤上采样模块

图 4 是特征过滤上采样模块的结构。在最后的融合阶段,模型使用逐元素相加的方式来混合所提取到的特征。然而,由于浅层特征提取模块并没有对亮度进行矫正,因此混合后的特征会混有低光噪声。为了避免低光噪声对最后生成图像的影响,模型引入一个通道注意力(channel attention, CA)模块^[41]对融合后特征进行过滤。

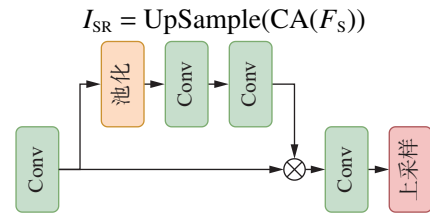


图 4 特征过滤上采样模块结构

Fig. 4 Structure of feature filtering upsample module

2.4 损失函数

为了提升模型在人脸图像上的恢复效果,本文设计了一个人脸掩码损失函数(faceMaskLoss, L_{mask})。具体来说,模型首先使用 MTCNN(multi-task cascaded convolutional networks)^[42] 检测图像中的人脸区域,然后随机遮掩人脸区域,约束模型更多地关注被遮掩的部分,从而增强对人脸细节的还原。 L_{mask} 的计算方式为

$$I_{\text{face}} = \text{MTCNN}(I_{\text{HR}})$$

$$I_{\text{mask}} = \text{Gen}(I_{\text{face}}, r)$$

$$L_{\text{mask}} = \|I_{\text{SR}} \times I_{\text{mask}} - I_{\text{HR}} \times I_{\text{mask}}\|_1$$

式中: I_{face} 表示检测到的人脸区域,可以被表示为 (x_1, x_2, y_1, y_2) 这种四元组的形式; Gen 表示生成掩码的方式; r 表示掩码比率,取值在 0~1, 1 表示人脸中所有的像素值都被计算, 0 则相反; I_{mask} 表示生成的随机掩码。

关于表示生成掩码的方式(生成函数)Gen, 本文提出了基于像素和基于分块的 2 种方法。基于像素的方式使用参数为 r 的伯努利分布得到与人脸区域同样大小的 0、1 值作为掩码。基于分块

的方法则根据掩码比率 r 计算分块数量并生成随机分块, 需要注意的是在基于分块的方法并不需要禁止分块之间存在重叠。本文将在实验部分对上述 2 种方法进行对比。

除了上述的 L_{mask} , 模型还使用 L_1 损失和快速傅里叶变换 (FFT) 损失, 它们的计算方式为

$$L_1 = \|I_{\text{SR}} - I_{\text{HR}}\|_1$$

$$L_{\text{FFT}} = \|\text{FFT}(I_{\text{SR}}) - \text{FFT}(I_{\text{HR}})\|_1$$

最后, 训练时使用的所有损失函数计算公式可以表示为

$$L = w_1 \times L_1 + w_2 \times L_{\text{FFT}} + w_3 \times L_{\text{mask}}$$

式中 w_1 、 w_2 和 w_3 都是平衡系数。

3 实验仿真及结果分析

3.1 数据集

为了展示所提模型的效果, 本文进行了 4 倍超分实验。由于没有通用的低光人脸超分数据集, 本文将正常光照下的人脸数据集 CelebA-HQ 和 Helen 数据集转换为低光条件下的合成数据集 CelebA-LLF 和 Helen-LLF, 使用的公式为

$$I_{\text{LR}} = \text{Dark}(I_{\text{HR}}) \downarrow_s$$

式中: s 为下采样因子; Dark 是暗化函数, 可以表示为

$$\text{Dark}(I) = \beta \times (\alpha \times I)^\gamma + 0.002 \times \mathcal{N}(0, 1)$$

式中: $\alpha \in (0.9, 1)$, $\beta \in (0.5, 1)$, $\gamma \in (1, 5)$, $\mathcal{N}(0, 1)$ 则表示一个标准的正态分布。

3.2 训练细节和指标

本文基于 PyTorch 和 BasicSR 将所提模型并在一张 NVIDIA RTX 3090 上完成了实验。在训练时, 使用的优化器是 Adam。初始的学习率设置为 2×10^{-4} , 并随着迭代次数的增加不断衰减, 衰减权重设置为 0.5。对于损失函数, 在训练时将 L_{mask} 的掩码比例 r 设置为 0.5, 权重系数 w_1 、 w_2 和 w_3 分别设置为 1、0.05 和 0.01。此外, 本文在训练时还使用随机旋转和水平翻转作为数据增强的方法。本文从 CelebA-LLF 数据集中选取 25 000 张图片作为训练集, 并将 CelebA-LLF 数据集中剩余的 5 000 张图片和 Helen-LLF 中的 50 张图片作为测试集。

本文用 3 个指标——峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity, SSIM) 和可学习的感知图像块相似性 (learned perceptual image patch similarity, LPIPS) 评估超分辨率结果的质量。PSNR 和 SSIM 是在转换后的 YCbCr 空间的 Y 通道上计算的, 它们的值越大, 表示生成图像质量越好。而 LPIPS 则是衡

量图像感知质量的常用指标, LPIPS 越小, 表明图像的感知相似度越高。

3.3 消融实验

为了验证所提模块的有效性以及不同掩码生成方法的差异, 本文对 FSEU 和 L_{mask} 进行了消融实验, 实验结果见表 1。

表 1 消融实验
Table 1 Ablation experiment

不同策略	PSNR	SSIM	LPIPS
w/o FSEU	24.238 3	0.748 8	0.326 2
w/o L_{mask}	24.243 4	0.750 3	0.334 3
w $L_{\text{pixelmask}}$	24.225 5	0.750 8	0.331 2
LF SRNet	24.495 4	0.758 7	0.319 5

表 1 中第 1 行和第 4 行表明所提的 FSEU 可以有效提高模型恢复低光人脸图像的能力。而对比第 2 行和第 4 行的结果可以发现, 引入 L_{mask} 对于提升模型的修复效果也是有所帮助的。最后, 通过比较第 3 行和第 4 行, 可以发现基于像素的掩码方法要差于基于分块的掩码方法。为了更加形象地说明两种方法之间的差异, 本文从数据集中随机抽取了一张图像并在它上面可视化了两种不同的掩码方式, 如图 5 所示。从图 5 中不难看出, 基于分块的掩码方法能够更好地反映图像中像素之间的联系, 因此效果更好。



图 5 不同掩码策略的可视化

Fig. 5 Visualization of different mask strategies

3.4 与其他方法对比

目前, 关于低光人脸超分辨率的研究非常有限, 只有少数工作^[15-16,37]对此进行了研究。不幸的是, 这些工作并没有公布它们的训练代码, 因此无法与它们进行对比。为了进行比较, 本文使用引言中提到的两种方法对现有的超分辨率方法进行了修改。实验中使用的超分变率方法有自然图像超分方法 EDSR (enhanced deep super-resolution network)^[43]、SwinIR (swin Transformer image restoration)^[44] 和 Omni-SR (omni super-resolution)^[45], 面部超分方法 CTCNet (CNN Transformer cooperation network)^[14] 和 SPARNet (spatial attention residual network)^[12] 和一个低光自然图像超分方法 LSR (lightening super-resolution)^[46]。低光增强的

方法则使用了 DLN(deep lightning network)^[27]。实验结果如表 2 所示,其中单一方法表示直接进行低光人脸超分($I_{LL} \rightarrow I_{NS}$)。在使用+连接的方法中,DLN 在前表示先进行低光增强($I_{LL} \rightarrow I_{NL}$),再

进行超分辨率($I_{NL} \rightarrow I_{NS}$);DLN 在后则表示先进行超分辨率($I_{LL} \rightarrow I_{LS}$),再进行低光增强($I_{LS} \rightarrow I_{NS}$)。值得注意的是,表 2 中给出的所有实验结果都是在 3.1 节所提数据集上重新训练后的结果。

表 2 CelebA-LLF 和 Helen-LLF 上的指标结果
Table 2 Metrics results on CelebA-LLF and Helen-LLF

方法	CelebA-LLF			Helen-LLF		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
EDSR ^[43]	23.673 8	0.739 8	0.362 0	19.625 3	0.788 7	0.379 6
DLN ^[27] + EDSR ^[43]	22.628 6	0.735 5	0.377 3	17.878 9	0.703 0	0.462 9
EDSR ^[43] + DLN ^[27]	22.479 7	0.743 6	0.336 9	17.338 3	0.717 2	0.428 3
Omni-SR ^[45]	23.869 3	0.736 7	0.389 1	22.346 9	0.822 7	0.327 9
DLN ^[27] + Omni-SR ^[45]	22.643 2	0.736 9	0.380 5	21.616 2	0.752 0	0.330 5
Omni-SR ^[45] + DLN ^[27]	22.536 9	0.747 5	0.346 4	20.197 5	0.766 8	0.273 4
SwinIR ^[44]	24.042 0	0.747 1	0.351 7	22.166 0	0.843 3	0.242 7
DLN ^[27] + SwinIR ^[44]	22.666 6	0.739 7	0.358 3	20.640 8	0.795 1	0.333 4
SwinIR ^[44] + DLN ^[27]	22.596 0	0.751 8	0.320 7	20.381 5	0.814 2	0.275 3
CTCNet ^[14]	24.365 9	0.754 4	0.328 7	21.454 1	0.858 5	0.233 8
DLN ^[27] + CTCNet ^[14]	22.671 5	0.740 1	0.357 5	22.557 2	0.851 0	0.277 3
CTCNet ^[14] + DLN ^[27]	18.662 7	0.722 4	0.336 7	22.463 1	0.853 0	0.200 7
SPARNet ^[12]	24.284 0	0.754 8	0.325 3	22.678 5	0.858 9	0.199 8
DLN ^[27] + SPARNet ^[12]	22.660 4	0.751 4	0.357 3	22.257 4	0.849 8	0.240 0
SPARNet ^[12] + DLN ^[27]	22.581 1	0.751 3	0.331 3	21.825 9	0.850 7	0.228 5
LSR ^[46]	23.912 2	0.741 3	0.356 8	21.362 1	0.756 7	0.354 7
LFSRNet	24.495 4	0.758 7	0.319 5	22.944 5	0.859 3	0.199 2

注:加粗表示最好的结果。

从表 2 可以看出,本文方法在 PSNR、SSIM 和 LPIPS 这 3 个指标上都达到了最优,这表明与对比方法相比,本文方法能更好地将低光、低分辨率的人脸图像重建为正常光照下的高分辨率人脸图像。此外,通过观察表 2 中的结果,可以得出以下结论:1)对比同一方法的 3 种不同实现策略,端到端训练的 PSNR 值往往优于两阶段方法,这表明使用直接训练方法可以有效提高生成图像的质量,显示了端到端方法的优越性。2)比较 2 种两阶段方法,先低光增强后人脸超分的方法在 PSNR 方面的指标往往更好,而在 SSIM 和 LPIPS 方面的性能则弱于先人脸超分后低光增强

的方法。本文推测这种现象是因为生成图像的质量受后执行任务的影响更大,人脸超分方法往往侧重于生成高质量的图像,而低光增强方法则侧重于恢复图像的结构信息,提高图像的感知质量。同时,这也表明二者的优化目标并不一致,再次说明了使用端到端方法来实现低光人脸超分任务是更好的选择。

图 6 和图 7 给出了对 CelebA-LLF 和 Helen-LLF 的定性分析结果。在精心设计的 FSEU 和 L_{mask} 的帮助下,本文方法可以有效地捕捉低光下的人脸图像信息。因此,重建结果对比方法更接近真实图像。



图 6 CelebA-LLF 数据集上的测试结果

Fig. 6 Results on the CelebA-LLF dataset

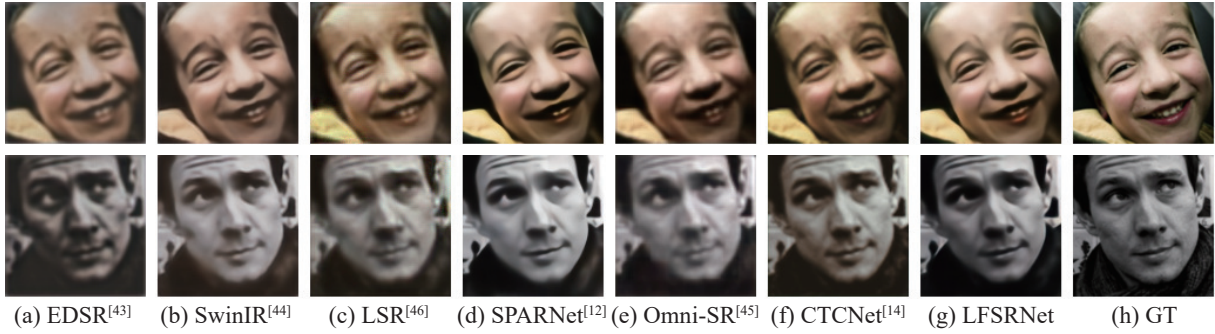


图 7 Helen-LLF 数据集上的测试结果

Fig. 7 Results on the Helen-LLF dataset

3.5 人脸对齐

为了进一步验证生成图像质量, 本文在 CelebA-LLF 上进行了人脸对齐任务。人脸对齐算法根据输入的人脸图像计算人脸的特征标点的位置, 如眼睛和鼻子。本文使用 MTCNN^[42] 来估计不同超分辨率方法生成的图像, 并使用均方误差 (mean-square error, MSE) 来评估人脸配准任务的效果。MSE 用 R_{MSE} 来表示, 计算方法为

$$R_{\text{MSE}} = \frac{\sum_{i=1}^n \sum_{j=1}^m \|S_i^j - H_i^j\|_2}{n \times m}$$

式中: n 是图片数量, m 是面部标记点数量, S_i^j 表示第 i 张图像中的第 j 个特征点, H_i^j 与之相似。

最终结果如表 3 所示, MSE 值越小表示人脸对齐效果越好。表 3 结果表明本文提出的 LFSR-Net 比其他方法有更好的人脸对齐性能, 侧面印证了所提方法生成图像更接近真实图像。

表 3 人脸对齐的指标结果

Table 3 Metrics result of face alignment

算法	R_{MSE}
EDSR ^[43]	12.098 3
SwinIR ^[44]	12.365 4
Omni-SR ^[45]	14.494 4
SPARNet ^[12]	10.954 6
CTCNet ^[14]	10.692 3
LSR ^[46]	10.245 1
LFSRNet	9.460 8

4 结束语

本文提出了一种针对低光环境的端到端人脸超分网络 LFSRNet。该方法通过对人脸的局部和全局建模, 有效地重建了清晰的人脸图像。同时, 为了进一步提高恢复效果, 本文提出了一种人脸掩码损失函数, 以引导模型更多地关注修复

人脸区域, 这种引导损失的效果通过分块掩码的方式得到了验证。大量实验表明, 本文方法优于目前最先进的方法。在未来的工作中, 可以尝试构建一个更轻量级的网络来重建低光环境下的低分辨率人脸图像。

参考文献:

- [1] KOWALSKI M, NARUNIEC J, TRZCINSKI T. Deep alignment network: a convolutional neural network for robust face alignment[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017: 2034–2043.
- [2] 贺怀清, 陈琴, 惠康华. 基于深度可分离沙漏网络的快速人脸对齐[J]. 计算机工程与设计, 2021, 42(8): 2316–2323.
HE Huaiqing, CHEN Qin, HUI Kanghua. Fast face alignment based on deep separable hourglass network[J]. Computer engineering and design, 2021, 42(8): 2316–2323.
- [3] LIU Yaojie, JOURABLOO A, REN W, et al. Dense face alignment[C]//2017 IEEE International Conference on Computer Vision Workshops. Venice: IEEE, 2017: 1619–1628.
- [4] LIN Jinpeng, YANG Hao, CHEN Dong, et al. Face parsing with RoI tanh-warping[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 5647–5656.
- [5] LIU Yinglu, SHI Hailin, SHEN Hao, et al. A new dataset and boundary-attention semantic segmentation for face parsing[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2020, 34(7): 11637–11644.
- [6] ZHENG Qingping, DENG Jiankang, ZHU Zheng, et al. Decoupled multi-task learning with cyclical self-regulation for face parsing[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 4146–4155.
- [7] 屈东东, 贺利乐, 何林. 改进的轻量化人脸识别算法[J].

- 智能系统学报, 2023, 18(3): 544–551.
- QU Dongdong, HE Lile, HE Lin. Improved lightweight face recognition algorithm[J]. *CAAI transactions on intelligent systems*, 2023, 18(3): 544–551.
- [8] 狄岚, 矫慧文, 梁久祯. 稀疏综合字典学习的小样本人脸识别[J]. *智能系统学报*, 2021, 16(2): 218–227.
- DI Lan, JIAO Huiwen, LIANG Jiuzhen. Sparse comprehensive dictionary learning for small-sample face recognition[J]. *CAAI transactions on intelligent systems*, 2021, 16(2): 218–227.
- [9] 章毅, 吕嘉仪, 兰星, 等. 结合面部动作单元感知的三维人脸重建算法[J]. *软件学报*, 2024, 35(5): 2176–2191.
- ZHANG Yi, LYU Jiayi, LAN Xing, et al. AU-aware algorithm for 3D facial reconstruction[J]. *Journal of software*, 2024, 35(5): 2176–2191.
- [10] 黄丰智, 范鹏, 范智瀚, 等. 基于散斑结构光编码的三维人脸重建方法研究[J]. *计算机应用与软件*, 2023, 40(9): 218–221, 228.
- HUANG Fengzhi, FAN Peng, FAN Zhihan, et al. 3d face reconstruction method based on speckle structured light coding[J]. *Computer applications and software*, 2023, 40(9): 218–221, 228.
- [11] ZHANG Tianke, CHU Xuangeng, LIU Yunfei, et al. Accurate 3D face reconstruction with facial component tokens[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 8999–9008.
- [12] CHEN Chaofeng, GONG Dihong, WANG Hao, et al. Learning spatial attention for face super-resolution[J]. *IEEE transactions on image processing*, 2021, 30: 1219–1231.
- [13] 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. Salt Lake City: IEEE, 2018: 2492–2501.
- [14] GAO Guangwei, XU Zixiang, LI Juncheng, et al. CTCNet: a CNN-Transformer cooperation network for face image super-resolution[J]. *IEEE transactions on image processing*, 2023, 32: 1978–1991.
- [15] GUO Kehua, HU Min, REN Sheng, et al. Deep illumination-enhanced face super-resolution network for low-light images[J]. *ACM transactions on multimedia computing, communications, and applications*, 2022, 18(3): 1–19.
- [16] DING Xin, HU Ruimin, WANG Zhongyuan. Face enhancement and hallucination in the wild[J]. *Neural computing and applications*, 2023, 35(3): 2399–2412.
- [17] ABDULLAH-AL-WADUD M, KABIR M H, ALI AKBER DEWAN M, et al. A dynamic histogram equalization for image contrast enhancement[C]//2007 Digest of Technical Papers International Conference on Consumer Electronics. Las Vegas: IEEE, 53(2): 593–600.
- [18] GUO Xiaojie, LI Yu, LING Haibin. LIME: low-light image enhancement via illumination map estimation[J]. *IEEE transactions on image processing*, 2017, 26(2): 982–993.
- [19] IBRAHIM H, PIK KONG N S. Brightness preserving dynamic histogram equalization for image contrast enhancement[J]. *IEEE transactions on consumer electronics*, 2007, 53(4): 1752–1758.
- [20] LI Mading, LIU Jiaying, YANG Wenhan, et al. Structure-revealing low-light image enhancement via robust retinex model[J]. *IEEE transactions on image processing*, 2018, 27(6): 2828–2841.
- [21] 赵明华, 汶怡春, 都双丽, 等. 基于照度与场景纹理注意力图的低光图像增强[J]. *中国图象图形学报*, 2024, 29(4): 862–874.
- ZHAO Minghua, WEN Yichun, DU Shuangli, et al. Low-light image enhancement algorithm based on illumination and scene texture attention map[J]. *Journal of image and graphics*, 2024, 29(4): 862–874.
- [22] 吴巨峰, 赵训刚, 周强, 等. 基于对比学习的低光照图像增强[J]. *计算机科学*, 2023, 50(S1): 525–530.
- WU Jufeng, ZHAO Xungang, ZHOU Qiang, et al. Contrastive learning for low-light image enhancement[J]. *Computer science*, 2023, 50(S1): 525–530.
- [23] ZHANG Yonghua, GUO Xiaojie, MA Jiayi, et al. Beyond brightening low-light images[J]. *International journal of computer vision*, 2021, 129(4): 1013–1037.
- [24] ZHU Anqi, ZHANG Lin, SHEN Ying, et al. Zero-shot restoration of underexposed images via robust retinex decomposition[C]//2020 IEEE International Conference on Multimedia and Expo. London: IEEE, 2020: 1–6.
- [25] LORE K G, AKINTAYO A, SARKAR S. LLNet: a deep autoencoder approach to natural low-light image enhancement[J]. *Pattern recognition*, 2017, 61: 650–662.
- [26] WEI Chen, WANG Wenjing, YANG Wenhan, et al. Deep retinex decomposition for low-light enhancement[C]//Proceedings of 29th British Machine Vision Conference. London: BMVA, 2018: 155–155.
- [27] WANG Liwen, LIU Zhisong, SIU Wanchi, et al. Lightening network for low-light image enhancement[J]. *IEEE transactions on image processing*, 2020, 29: 7984–7996.
- [28] HE Jingwen, SHI Wu, CHEN Kai, et al. GCFSR: a generative and controllable face super resolution method without facial and GAN priors[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 1879–1888.
- [29] 姚肇亮, 刘宇男, 张姗姗, 等. 基于多尺度特征融合的属性已知人脸图像超分辨率[J]. *南京理工大学学报*, 2022, 46(5): 571–578.
- YAO Zhaoliang, LIU Yunan, ZHANG Shanshan, et al. Multi-scale feature fusion based attribute-aware face im-

- age super-resolution[J]. Journal of Nanjing University of Science and Technology, 2022, 46(5): 571–578.
- [30] 陈泓佑, 陈帆, 和红杰, 等. 基于多任务对抗和抗噪对抗学习的人脸超分辨率算法[J]. 模式识别与人工智能, 2022, 35(10): 863–880.
- CHEN Hongyou, CHEN Fan, HE Hongjie, et al. Face super-resolution algorithm based on multi-task adversarial and antinoise adversarial learning[J]. Pattern recognition and artificial intelligence, 2022, 35(10): 863–880.
- [31] MENON S, DAMIAN A, HU Shijia, et al. PULSE: self-supervised photo upsampling via latent space exploration of generative models[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 2434–2442.
- [32] WANG Chenyang, JIANG Junjun, ZHONG Zhiwei, et al. Spatial-frequency mutual learning for face super-resolution[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 22356–22366.
- [33] LU Tao, WANG Yuanzhi, ZHANG Yanduo, et al. Face hallucination via split-attention in split-attention network [C]//Proceedings of the 29th ACM International Conference on Multimedia. Virtual Event: ACM, 2021: 5501–5509.
- [34] YANG Tao, REN Peiran, XIE Xuansong, et al. GAN prior embedded network for blind face restoration in the wild[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 672–681.
- [35] WANG Yuanzhi, LU Tao, YAO Yuan, et al. Learning to hallucinate face in the dark[J]. IEEE transactions on multimedia, 2023, 26: 2314–2326.
- [36] HAN Kai, WANG Yunhe, CHEN Hanting, et al. A survey on vision transformer[J]. IEEE transactions on pattern analysis and machine intelligence, 2023, 45(1): 87–110.
- [37] NEWELL A, YANG Kaiyu, DENG Jia. Stacked hour-glass networks for human pose estimation[C]//Proceedings of the 14th European Conference on Computer Vision. London: Springer, 2016: 483–499.
- [38] TU Zhengzhong, TALEBI H, ZHANG Han, et al. MaxViT: multi-axis vision Transformer[C]//Proceedings of the 17th European Conference on Computer Vision. London: Springer, 2022: 459–479.
- [39] KONG Fangyuan, LI Mingxi, LIU Songwei, et al. Residual local feature network for efficient super-resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022: 765–775.
- [40] LIU Jie, ZHANG Wenjie, TANG Yuting, et al. Residual feature aggregation network for image super-resolution[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 2356–2365.
- [41] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11531–11539.
- [42] ZHANG Kaipeng, ZHANG Zhanpeng, LI Zhifeng, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE signal processing letters, 2016, 23(10): 1499–1503.
- [43] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017: 1132–1140.
- [44] LIANG Jingyun, CAO Jiezhong, SUN Guolei, et al. SwinIR: image restoration using swin transformer[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal: IEEE, 2021: 1833–1844.
- [45] WANG Hang, CHEN Xuanhong, NI Bingbing, et al. Omni aggregation networks for lightweight image super-resolution[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 22378–22387.
- [46] RASHEED M T, SHI Daming. LSR: Lightening super-resolution deep network for low-light image enhancement[J]. Neurocomputing, 2022, 505: 263–275.

作者简介:



丛维仪, 硕士研究生, 主要研究方向为图像增强。E-mail: congweiyi@njust.edu.cn。



郑卓然, 博士研究生, 主要研究方向为深度学习和图像增强。E-mail: zhengzr@njust.edu.cn。



贾修一, 教授, 博士生导师, 中国计算机学会杰出会员, 主要研究方向为机器学习、粒计算和计算机视觉。主持国家自然科学基金项目 4 项。发表学术论文 100 余篇。E-mail: jiaxy@njust.edu.cn。