

基于3D先验特征的人脸超分辨率重建算法

姚汉群¹, 刘广文¹, 王超², 杨依宁³, 才华¹, 付强⁴

1. 长春理工大学 电子信息工程学院, 长春 130022;
2. 长春理工大学 空间光电技术国家与地方联合工程研究中心, 长春 130022;
3. 电磁空间安全全国重点实验室, 天津 300308;
4. 长春理工大学 空间光电技术研究所, 长春 130022)

摘要: 为有效解决复杂环境下人脸超分辨率特征恢复的问题, 提出一种全新的人脸超分辨率网络. 该网络通过融合3D渲染先验知识和双重注意力机制, 增强了对人脸空间位置和整体结构的理解, 同时提高了细节信息的恢复能力. 在数据集 CelebAMask-HQ 上的实验结果表明: 对放大4倍下采样的人脸, 该算法在峰值信噪比和结构相似性上达到 28.76 dB 和 0.827 5; 对放大8倍下采样的人脸, 峰值信噪比和结构相似性评价指标达到 26.29 dB 和 0.754 9. 与同类的 SAM3D 算法相比, 该算法在处理放大4倍下采样时的峰值信噪比和结构相似性上分别提升了 4.09, 1.93 个百分点, 在处理放大8倍下采样时上述两个指标分别提升了 2.02, 4.54 个百分点. 从而证明该算法的优越性, 也表明在实际应用中人脸的超分辨率恢复能获得更真实和清晰的视觉效果.

关键词: 机器视觉; 人脸超分辨率; 3D先验; 注意力机制

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2024)04-0895-10

Facial Super-resolution Reconstruction Algorithm Based on 3D Prior Features

YAO Hanqun¹, LIU Guangwen¹, WANG Chao², YANG Yining³, CAI Hua¹, FU Qiang⁴

1. School of Electronic Information Engineer, Changchun University of Science and Technology, Changchun 130022, China;
2. National and Local Joint Engineering Research Center for Space Optoelectronics Technology, Changchun University of Science and Technology, Changchun 130022, China;
3. National Key Laboratory of Electromagnetic Space Security, Tianjin 300308, China;
4. School of Opto-Electronic Engineer, Changchun University of Science and Technology, Changchun 130022, China)

Abstract: In order to effectively solve the problem of facial super-resolution feature recovery in complex environments, we proposed a novel facial super-resolution network. By integrating 3D rendering prior knowledge and a dual attention mechanism, the network enhanced the understanding of the facial spatial position and overall structure while improving the ability to recover detailed information. The experimental results on the CelebAMask-HQ dataset show that the proposed algorithm achieves peak signal-to-noise ratio and structural similarity of 28.76 dB and 0.827 5 for downsampled faces magnified by 4 times, and 26.29 dB and 0.754 9 for downsampled faces magnified

收稿日期: 2023-05-26.

第一作者简介: 姚汉群(1996—), 男, 汉族, 硕士研究生, 从事机器视觉与人脸识别的研究, E-mail: 531277576@qq.com. 通信

作者简介: 刘广文(1971—), 男, 汉族, 博士, 副教授, 从事信号与信息处理技术的研究, E-mail: lgwen_2003@126.com.

基金项目: 国家自然科学基金重大项目(批准号: 61890963)和吉林省科技发展计划项目(批准号: 20210204099YY).

by 8 times. Compared with the similar SAM3D algorithm, the proposed algorithm improves the peak signal-to-noise ratio and structural similarity by 4.09 and 1.93 percentage points when dealing with 4 times downsampling, and by 2.02 and 4.54 percentage points when dealing with 8 times downsampling, respectively. This proves the superiority of the proposed algorithm and also indicates that facial super-resolution recovery can achieve more realistic and clear visual effects in practical applications.

Keywords: machine vision; facial super-resolution; 3D prior; attention mechanism

0 引言

人脸超分辨率技术目前已被广泛关注和应用. 在生物识别技术, 如人脸识别^[1]中, 尤其是在处理低分辨率或质量较差的图像时, 超分辨率技术通过提高图像质量和增强细节显著提高了识别精度. 此外, 该技术在视频监控^[2]中也得到应用, 常用于从低质量的图像或视频中识别和提取面部信息. 人脸超分辨率技术在医疗诊断和治疗领域^[3], 例如皮肤病诊断和视网膜扫描等方面也发挥了重要作用. 但处理极度模糊或严重压缩的图像, 以及在复杂背景和不同光照条件下恢复面部细节等问题仍面临技术的挑战. 因此, 研究改进现有的超分辨率算法, 可进一步提高人脸图像的质量和识别精度.

人脸超分辨率方法主要分为三类: 基于插值的人脸超分辨率^[4-6]、基于学习的人脸超分辨率^[7-9]和深度学习的人脸超分辨率^[10-11]. 基于插值的方法利用人脸图像的空间冗余性进行超分辨率, 但难以复原高频细节; 基于学习的方法通过训练数据学习人脸的高频信息, 以提高重建质量; 深度学习的人脸超分辨率方法则通过卷积神经网络等复杂模型进一步提升人脸超分辨率的性能. 随着深度学习方法的应用, 人脸超分辨率技术已能通过神经网络, 利用局部的外观特性, 构建出低分辨率和高分辨率人脸图像之间的映射关系. 但大多数深度学习方法不能很好地利用面部结构和身份信息. 作为一种特定领域的超分辨率技术, 人脸超分辨率重建可用于恢复丢失的面部细节. 其成功的关键是如何有效利用人脸的先验知识, 从 1D 向量如身份和属性, 到 2D 图像如面部地标、面部热图和解析图, 再到 3D 模型. 因此, 发现新的人脸先验知识, 如何建模或表示这些先验知识, 以及如何将这些信息有机地集成到训练框架中, 都值得进一步研究.

为解决上述问题, 本文提出一种结合 3D 人脸先验知识的人脸超分辨率网络, 算法流程如图 1 所示. 该网络的核心是将 3D 人脸先验与超分辨率重建网络相融合, 并添加双重注意力模块, 以实现对面脸空间位置和整体结构更准确地把握, 突出细节恢复. 这样的设计不仅在复杂环境中提高了面部特征恢复的精度, 也在多尺度上优化了图像质量.

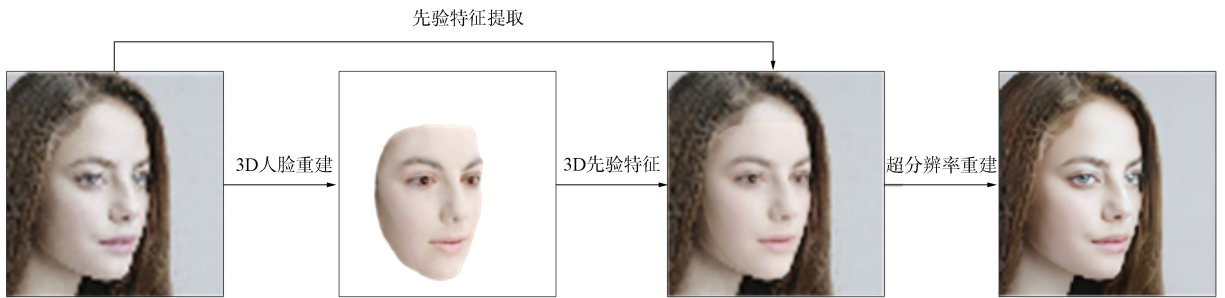


图 1 算法流程

Fig.1 Flow chart of algorithm

1 相关工作

人脸超分辨率技术旨在从低分辨率人脸图像中恢复高分辨率细节, 在现实生活中应用广泛, 如视频监控、人脸识别、虚拟现实等. 目前, 大部分研究工作集中在提高图像质量、增强面部特征细节、处

理面部遮挡和姿态变化等问题上. Dong 等^[12]提出了一种端到端的深度学习单图像超分辨率方法 SRCNN, 其具有轻量级结构、卓越的恢复质量, 并实现了快速在线应用, 但在处理遮挡、姿态变化等复杂场景时可能产生伪影和失真; Kim 等^[13]提出了一种高精度单图像超分辨率方法, 采用类 VGG-net 深度卷积网络, 通过增加网络深度显著提高准确性, 但可能导致过拟合和计算量过大; 为解决面部超分辨率问题, Chen 等^[14]提出了一种可端到端训练的 FSRNet 和 FSRGAN, 利用面部标记热图和解析图作为几何先验, 并引入了新评估指标, 但对于遮挡和姿态变化图像的处理仍有限; Zhang 等^[15-16]提出了残差密集网络 RDN 和深度残差通道注意力网络 RCAN, 分别针对深度卷积神经网络的层次特征和训练难题进行了优化, 并引入通道注意力机制, 通过考虑通道间的相互依赖自适应地调整通道特征. 但在处理复杂场景时上述方法仍面临局限性, 而且通道注意力可能无法充分捕获空间信息; Wang 等^[17]提出了一种基于空间特征变换(spatial feature transform, SFT)的方法, 通过语义分割概率图调整中间层特征, 以实现自然、逼真纹理的恢复; Hu 等^[18]提出了一种结合 3D 面部先验的人脸超分辨率方法 SAM3D, 首次将超分辨率与人脸 3D 先验信息融合, 并尝试利用空间注意力解决超分辨率问题, 推动了单图像超分辨率领域的发展, 实现了更高的准确性和更逼真的纹理恢复. 但现有人脸超分辨率方法在处理遮挡、姿态变化等复杂场景时仍存在一定的局限性, 例如可能产生模糊、细节丢失、结构扭曲、质感不自然等问题. 因此, 本文提出一种结合高维先验知识与双重注意力的人脸超分辨率算法.

2 算法设计

本文提出一种新的人脸超分辨率重建算法, 该算法的核心是引入人脸的 3D 先验信息, 系统网络结构如图 2 所示. 首先, 根据 UNET 解码器对多尺度面部特征进行提取, 利用预训练的网络权重预测低分辨率人脸图像的 Basel 人脸模型(Basel face model, BFM)^[19]系数, 进而实现 3D Morphable Model 算法^[20](3DMM)的三维重建. 本文将该三维重建的渲染模型与多尺度图像一同输入到空间特征变换模块, 从而获得各对应尺度的 3D 面部先验. 其次, 将这些 3D 先验特征与先验网络相结合, 在不同的尺度上并入 3D 先验信息. 最后, 在网络的末端采用空间注意力和通道注意力机制, 以有效整合网络的信息, 最终得到超分辨率人脸图像.

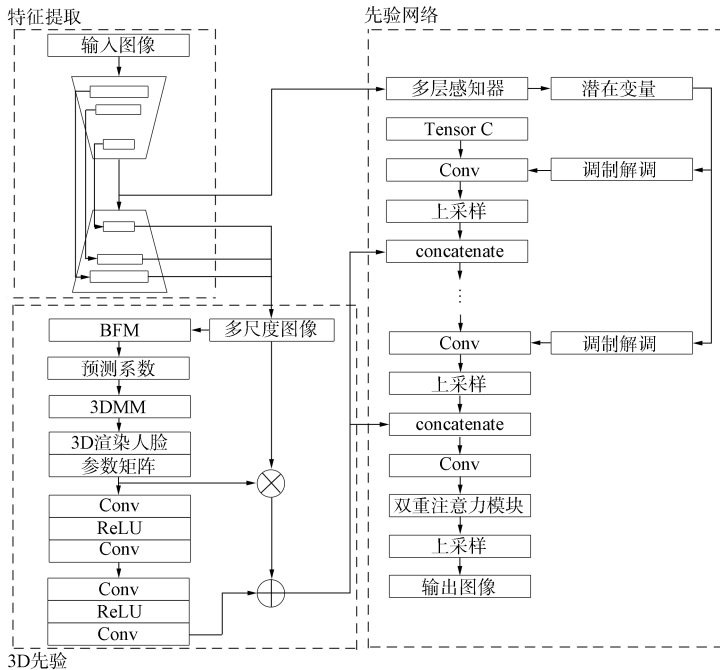


图 2 系统网络结构

Fig. 2 Architecture of system network

2.1 3D 人脸先验

2.1.1 3D 人脸重建算法

人脸的三维模型可通过从二维图像中提取关键信息进行重建, 3DMM 能根据特定的面部属性, 如面部特征和身份纹理等参数获得 3D 面部先验. 这种方法充分利用了人脸的独特属性, 从而使重建过程更准确、有效. 在 3D 面部重建过程中, 3DMM 结合了从 2D 图像中提取的关键信息, 以生成具有细致形状和纹理特征的 3D 面部模型. 重建得到的三维面部模型可保留原始图像中的面部特征, 呈现出清晰锐利的面部成分, 使其在视觉上更真实、生动.

为获得精确的 3D 渲染人脸, 本文对 3DMM 模型进行改进, 通过训练 ResNet-50 网络预测 BFM 中的形状参数、纹理参数和光照参数, 然后利用这些参数回归 3DMM 模型, 获得 3D 系数向量, 构建出精确的面部结构定位的 3D 面部图像. 3D 系数向量表示为 $\mathbf{C} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \mathbf{t}) \in \mathbb{R}^{239}$, 其中 $\boldsymbol{\alpha} \in \mathbb{R}^{80}$, $\boldsymbol{\beta} \in \mathbb{R}^{64}$, $\boldsymbol{\delta} \in \mathbb{R}^{80}$, $\boldsymbol{\gamma} \in \mathbb{R}^9$, $\mathbf{t} \in \mathbb{R}^6$ 分别表示身份、面部表情、纹理、照度和面部姿态参数. 利用这些参数可生成具有丰富细节和准确定位的 3D 面部图像. $\boldsymbol{\alpha}$ 、 $\boldsymbol{\beta}$ 和 $\boldsymbol{\delta}$ 是用于生成 3D 人脸的相应系数向量, 将人脸系数向量转换为人脸图像的 3D 形状 S 和纹理 T :

$$S = S(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \bar{S} + B_{\text{id}}\boldsymbol{\alpha} + B_{\text{exp}}\boldsymbol{\beta}, \quad (1)$$

$$T = T(\boldsymbol{\delta}) = \bar{T} + B_t, \quad (2)$$

其中: \bar{S} 表示平均人脸形状; \bar{T} 表示平均人脸纹理; B_{id} 表示身份的主成分分析基, B_{exp} 表示表情的主成分分析基, B_t 表示纹理的主成分分析基, 且 $B_{\text{id}}, B_{\text{exp}}, B_t$ 均对基矩阵用标准差进行了缩放.

利用面部形状 S 计算面部的法向量. 对每个顶点 p , 计算与其相邻的每个面 f 的法向量 \mathbf{VN}_p :

$$\mathbf{VN}_p = \frac{\sum_{f \in P_p} ((s_{f_0} - s_{f_1}) \times (s_{f_1} - s_{f_2}))}{\left\| \sum_{f \in P_p} ((s_{f_0} - s_{f_1}) \times (s_{f_1} - s_{f_2})) \right\|}, \quad (3)$$

其中, $s_{f_i} (i=0, 1, 2)$ 表示面部形状 S 的第 f_i 个顶点, $f = \{f_0, f_1, f_2\}$ 表示三角形面的顶点索引, P_p 表示与顶点 p 相邻的面的集合.

根据角度 $\boldsymbol{\theta} = (\theta_x, \theta_y, \theta_z)$ 计算旋转矩阵 \mathbf{R} , 对面部形状 S 应用旋转矩阵 \mathbf{R} 并加上姿态参数 \mathbf{t} , 得到旋转后的面部形状 S_r , 并重新计算旋转后的面部法向量 \mathbf{VN}_r :

$$\mathbf{R} = \mathbf{R}_z(\theta_z)\mathbf{R}_y(\theta_y)\mathbf{R}_x(\theta_x), \quad (4)$$

$$S_r = S \cdot \mathbf{R} + \mathbf{t}, \quad \mathbf{VN}_r = \mathbf{VN}_p \cdot \mathbf{R}, \quad (5)$$

其中 $\mathbf{R}_x(\theta_x)$, $\mathbf{R}_y(\theta_y)$ 和 $\mathbf{R}_z(\theta_z)$ 分别是关于 x, y, z 轴的旋转矩阵.

使用球谐函数和照度系数 $\boldsymbol{\gamma}$ 对面部纹理 T 进行照明, 得到面部颜色 \mathbf{FC} 和照明强度 L :

$$\mathbf{FC} = \begin{pmatrix} L_r T_r \\ L_g T_g \\ L_b T_b \end{pmatrix}, \quad (6)$$

$$L = \sum_{i=0}^8 \gamma_i Y_i(\mathbf{VN}_r), \quad (7)$$

其中 L_r, L_g, L_b 分别为红、绿、蓝通道的照明强度, T_r, T_g, T_b 分别为红、绿、蓝通道的面部纹理, Y_i 为球谐函数.

最后, 返回旋转后的面部形状 S_r 、旋转后的面部法向量 \mathbf{VN}_r 、面部颜色 \mathbf{FC} 和面部三角形顶点索引 tri .

本文将 U-Net 转换得到的不同尺度的低分辨率图像输入到预训练好的 3D 渲染分支中, 然后通过该分支预测 BFM 模型中的身份、纹理等参数, 再将这些参数输入到 3DMM 模型中, 生成渲染图像. 在此过程中, 将照度参数与面部姿态参数进行级联, 以便与身份、纹理等参数进行整合. 为使参数矩阵与低分辨率图像大小相同, 本文采用零填充操作. 最后将处理后的参数矩阵与渲染图像进行级联.

通过上述流程, 可实现将 U-Net 生成的低分辨率图像与 3D 渲染分支预测的面部参数相结合, 从而生成具有更丰富细节和更精确定位的 3D 渲染人脸. 这种方法不仅能提高面部重建的质量, 同时也

有助于在处理复杂场景和不同尺度的图像时保持稳定性和准确性. 综上, 3D 人脸重建算法如下.

算法 1 三维人脸重建.

- 步骤 1) 程序: 三维人脸重建(C , 人脸模型);
- 步骤 2) 在 BFM 模型中训练 N 预测 $\alpha, \beta, \delta, \gamma, t$;
- 步骤 3) 获得三维系数向量 $x = (\alpha, \beta, \delta, \gamma, t) \in \mathbb{R}^{239}$, 通过 3DMM 回归;
- 步骤 4) $S = S(\alpha, \beta) = \bar{S} + B_{id}\alpha + B_{exp}\beta$;
- 步骤 5) $T = T(\delta) = \bar{T} + B_t$;
- 步骤 6) 使用 S 和式(3)计算 VN ;
- 步骤 7) 使用 θ 和式(4)计算 R ;
- 步骤 8) $S_r = S \cdot R + t$;
- 步骤 9) $VN_r = VN_p \cdot R$;
- 步骤 10) 使用 T, V, γ 和式(6), (7)计算 FC 和 L ;
- 步骤 11) $tri \leftarrow$ 来自 BFM 的面部缓冲区;
- 步骤 12) 返回 S_r, VN_r, FC, tri ;
- 步骤 13) 结束程序;
- 步骤 14) 程序: 创建模型(v, f, c);
- 步骤 15) 创建一个空的三维模型 obj;
- 步骤 16) 循环: 在 VN 和 FC 中的每个顶点 v_i 及对应的颜色 c_i ;
- 步骤 17) 将顶点 v_i 和颜色 c_i 添加到三维模型 obj;
- 步骤 18) 结束循环;
- 步骤 19) 循环: 在 f 中的每个三角形 f_i ;
- 步骤 20) 将三角形 f_i 添加到 obj;
- 步骤 21) 结束循环;
- 步骤 22) 结束程序.

2.1.2 空间特征变换

SFT 模块^[17]在超分辨率和人脸超分辨率领域具有自适应性、高表达能力、性能良好和鲁棒性的优势, 它根据输入的先验信息调整特征图, 使网络适应各种场景. 利用 3D 先验信息和人脸相关知识, SFT 模块更好地保留了面部细节和特征, 提高了模型的性能、质量和鲁棒性, 实现了高质量的人脸重建. 这里引入 SFT 层, 根据输入的 3D 人脸先验信息自适应地调整特征图, 从而使网络能更适应各种场景和不同的输入数据, 其结构如图 3 所示.

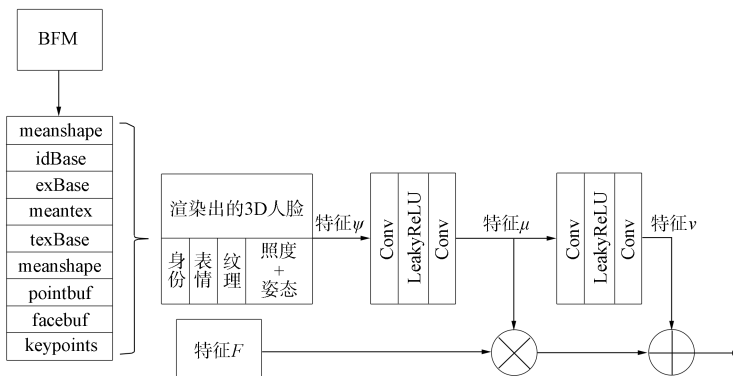


图 3 空间特征变换模块结构

Fig. 3 Structure of spatial feature transformation module

SFT 模块的主要功能是学习映射函数 Θ , 根据 3D 先验信息调制参数对 (μ, ν) . 参数对用于在空间上调节每个中间特征图的仿射变换, 使 SFT 层的输出可根据调制参数对进行自适应控制. 首先利用 3D 脸部先验信息提取出深层特征 ψ :

$$(\mu, \nu) = \Theta(\phi), \tag{8}$$

其中 μ 和 ν 分别表示卷积得到的不同程度的深度特征. 然后, 利用这些调制参数对对中间特征图 F 进行空间方向的转换(如缩放和平移等操作):

$$\text{SFT}(F|\mu, \nu) = \mu \otimes F + \nu, \tag{9}$$

其中 F 表示特征图, \otimes 表示逐元素相乘. 在执行空间方向的转换后, 获得融合了 3D 先验信息的高级特征.

通过上述流程可将重建的渲染模型与多尺度图像一起传入 SFT 模块, 得到对应尺度的 3D 面部先验特征, 并为后续 3D 先验特征与先验网络相结合, 在不同尺度上拼接先验信息做准备.

2.2 面部先验算法

在获取到各尺度的 3D 先验信息后, 在先验网络中将对应尺度的 3D 先验与生成先验进行融合, 以得到一个更具丰富信息的人脸先验.

2.2.1 先验网络

由一个预训练的面部生成对抗网络(generative adversarial networks, GAN)通过其卷积层的权重捕获了面部的分布, 即为生成先验. 本文利用改进的 StyleGAN2^[21]为任务提供多样和丰富的面部细节. StyleGAN2 在训练过程中学习了人脸的潜在分布即潜在变量, 并将这种变量嵌入到其卷积层的权重中. 因此, 这些权重构成了一种面部的生成先验, 即在生成新的人脸图像时, 网络会根据这个先验分布生成符合人脸特征分布的新样本.

为生成丰富多样的面部细节、便于与多尺度 3D 先验融合, 并减少迭代优化所需的时间, 本文采用以下策略: 先利用调制过程将潜在变量映射到生成器的不同层次, 而非直接生成最终图像; 再通过卷积上采样, 保留多尺度深层特征; 最后将这些特征与带有 3D 先验信息的高级特征拼接, 并继续执行上采样操作.

在上述过程中, 潜在变量通过嵌入到每个卷积层, 为每个分辨率尺度生成人脸特征. 这种方法能有效融合多尺度深层特征, 生成高质量的超分辨率图像, 同时保留丰富的面部细节, 并减少迭代优化所需时间.

2.2.2 双重注意力

DANetHead(dual attention network head)是一种自注意力机制的应用, 用于捕获图像中的长距离依赖关系, 其结构如图 4 所示. 将 DANetHead 添加至先验网络后, 可帮助模型更好地捕获全局上下文信息, 从而提升超分辨率图像的质量. DANetHead 引入了两个并行的自注意力模块, 即位置注意力模块(position attention module, PAM)和通道注意力模块(channel attention module, CAM), 它们共同捕获长距离上下文信息. 将 PAM 和 CAM 并行应用于输入特征图, 并将它们的输出加在一起, 得到最终的上下文特征.

本文选择在先验网络的最后利用空间注意力与通道注意力联系整个网络的信息, 提升模型在处理复杂场景和多样化表达时的性能.

2.3 损失函数

训练模型的学习目标包括: 限制输出以接近真实的重建损失 L_{rec} ^[22]、恢复真实纹理的对抗性损失 L_{adv} ^[21, 23]、与重建相关的皮肤光照强度损失 L_{per} ^[24]、进一步增强面部细节的面部成分损失 L_{comp} ^[21, 25] 以及身份保持损失 L_{id} ^[22]. 通过这些学习目标的组合, 模型可在恢复图像时有效提升质量并保持身份一致性.

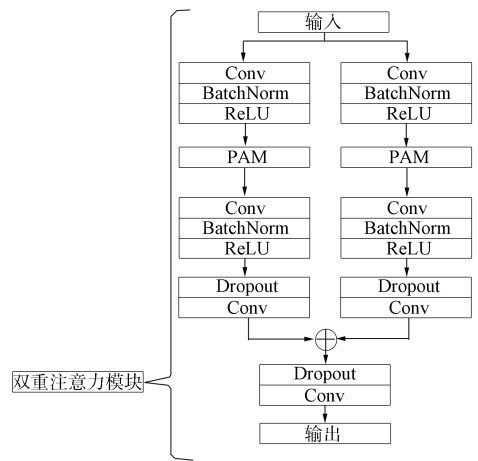


图 4 双重注意力模块结构

Fig. 4 Structure of dual attention module

3 实 验

实验选择数据集 CelebAMask-HQ 验证算法性能. 数据集 CelebAMask-HQ 是一个大规模的人脸图像数据集, 是从数据集 CelebA(celebFaces attributes dataset)中选择的 30 000 张高分辨率人脸图像. 每个图像都有对应于 CelebA 的面部属性的分割蒙版, 尺寸为 512×512. 在数据集 CelebA 中, 将数据集分为训练集、验证集和测试集. 通过对数据集进行划分, 可确保在训练、验证和测试过程中具有良好的数据分布和不同的人脸样本, 从而有效地评估模型在各种场景下的性能. 本文按与官方相同的比例随机划分并处理数据集.

3.1 实验数据及对比方法

训练阶段使用 24 000 张来自数据集 CelebAMask-HQ^[26]中未遮挡、带有帽子和饰品等小范围遮挡及非正面姿态等可识别出人脸关键点的图像, 测试阶段则使用同数据集中的 3 000 张图像. 遵循现有超分辨率方法的协议, 通过随机双三次下采样方法生成低分辨率图像 LQ 作为输入. 低分辨率人脸图像通过下采样至 128×128 像素(×4 尺度)和 64×64 像素(×8 尺度)生成.

为评估本文人脸超分辨率网络的性能, 将本文算法与 6 种目前最先进的超分辨率和人脸超分辨率重建算法进行定性和定量比较, 包括: 非常深的卷积网络 VDSR; 超分辨率的卷积神经网络 SRCNN; 残差密集网络 RDN; 非常深的残差注意力通道网络 RCAN; 深度端到端的可训练的人脸超分辨网络 FSRNet 以及使用人脸 3D 先验的 SAM3D 算法.

3.2 数据分析

3.2.1 定量结果

利用峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性(structural similarity index measure, SSIM)两个客观评价指标分析对比各算法的性能. 表 1 列出了不同算法在人脸图像放大 4 倍时的评价指标对比结果. 表 2 列出了不同算法在人脸图像放大 8 倍时的评价指标对比结果.

表 1 不同算法在人脸图像放大 4 倍时的评价指标对比结果

Table 1 Comparison results of evaluation metrics for different algorithms whent enlarging facial images by 4 times

算法	PSNR/dB	SSIM	算法	PSNR/dB	SSIM
FSRNET	25.37	0.754 0	RCAN	26.25	0.798 7
VDSR	26.36	0.751 6	BICUBIC	25.34	0.598 8
SRCNN	25.84	0.780 4	SAM3D	27.63	0.811 8
RDN	25.88	0.619 2	本文	28.76	0.827 5

表 2 不同算法在人脸图像放大 8 倍时的评价指标对比结果

Table 2 Comparison results of evaluation metrics for different algorithms when enlarging facial images by 8 times

算法	PSNR/dB	SSIM	算法	PSNR/dB	SSIM
FSRNET	22.27	0.584 3	RCAN	23.61	0.656 2
VDSR	22.26	0.582 0	BICUBIC	20.71	0.535 6
SRCNN	22.79	0.613 9	SAM3D	25.77	0.722 1
RDN	22.31	0.587 5	本文	26.29	0.754 9

由表 1 和表 2 可见: 在放大 4 倍时, BICUBIC 算法重构的人脸图像效果最差, 两个指标均为最低, 视觉效果模糊; VDSR, SRCNN 和 RDN 算法在重构人脸图像质量上略有提升, 但效果不明显, 仍存在视觉模糊和人脸轮廓边缘不清晰等问题; 相比之下, RCAN 和 FSRNET 算法的性能稍好, 但仍无法达到令人满意的效果; 而本文算法和 SAM3D 算法在重构质量上有显著提升, 特别是本文算法, 不仅在放大 4 倍时的 PSNR 和 SSIM 指标超过了其他算法, 且在放大 8 倍时, 尽管所有算法的评价指标都有所下降, 但本文算法仍保持了相对较高的性能, 证明了其在处理高倍率放大时的优越性. 尽管 SAM3D 算法在放大 4 倍和 8 倍时的性能也相对较好, 但在放大 8 倍时的 PSNR 和 SSIM 指标仍低于本文算法. 表明虽然 SAM3D 算法能在一定程度上提升重构质量, 但在处理放大更高倍率时, 其性能增幅有限. 本文算法在处理不同放大倍率的人脸图像重构任务时, 都表现了优异性能, 无论是在客观的

PSNR 和 SSIM 评价指标上,还是在视觉效果上,都超越了其他对比算法.从而证明了本文算法在处理复杂的人脸图像重构任务时的有效性和鲁棒性.

3.2.2 定性评价

本文算法在放大 8 倍下的定性结果如图 5 所示.由图 5 可见,本文算法成功地恢复了清晰且富有细节的面部特征,例如鼻子、眼睛和嘴巴.



图 5 人脸图像放大 8 倍时不同算法的效果对比

Fig. 5 Comparison of effects of different algorithms when enlarging facial images by 8 times

对比其他算法,本文算法在高频特征的恢复上也展现了显著优势,而这些区域在其他算法中常表现模糊.相反,其他算法在恢复过程中会在面部特征区域(如眼睛和鼻子)产生一些伪影,特别是当面部图像被部分遮挡或显示不完全时.而本文引入 3D 人脸先验后的方法展示出更清晰、更锐利且无明显重影伪像的人脸结构,这表明引入的 3D 先验在辅助网络理解面部的空间位置和整体结构上具有至关重要的作用,极大减少了常见的人脸特征伪像和人脸属性的显著失真.由于融合了高维先验知识和双重注意力机制,这不仅使本文算法能更有效地捕捉空间和特征关系,还使其在恢复过程中能专注于关键细节.这种策略有效地解决了早期研究中遇到的遮挡、姿态变化、过拟合和计算负荷等挑战.尤其双重注意力机制使本文算法能更有效地捕捉通道和空间信息,从而克服了仅依赖通道注意力机制的限制.因此,无论是在图像客观评分还是综合主观评价上,本文算法都能重构出具有更佳重建效果的人脸图像.

3.3 消融实验

下面进行消融研究以证明每个模块的有效性.根据 PSNR 和 SSIM 在 8 倍尺度测试数据上比较使用和不使用渲染的 3D 人脸先验和双重注意力模块的拟议网络.表 3 列出了添加 3D 先验和双重注意力模块前后的评价指标对比结果.由表 3 可见,不使用 3D 渲染人脸和注意力的方法通常无法捕获清晰的人脸结构,而添加 3D 渲染先验后显示了更清晰和更锐利的面部结构.通过同时使用 3D 先验和注意

力模块,视觉质量得到进一步改善.表1和表2中的定量比较也说明了所添加的渲染先验和注意力模块的有效性.

表3 添加3D先验和双重注意力模块前后的评价指标对比

Table 3 Comparison of evaluation metrics before and after adding 3D prior and dual attention modules

模型	PSNR/dB	SSIM
3D先验	25.14	0.745 3
3D先验+双重注意力	26.29	0.754 9

综上所述,为解决现有方法在处理复杂场景时人脸图像产生的伪影、失真、模糊、细节丢失、结构扭曲、质感不自然等问题,本文提出了一种基于3D渲染先验和双重注意力模块的人脸超分辨率网络.引入3D渲染先验使网络能更好地理解人脸的空间位置和整体结构,有助于在复杂场景下恢复更精细的面部特征.双重注意力模块允许本文模型关注重要的局部细节和全局上下文信息,从而提高面部特征恢复的准确性和图像质量.在数据集 CelebAMask-HQ 上进行实验的结果表明,本文算法在 PSNR 和 SSIM 指标上均优于现有的先进算法,证明了该算法在处理带有帽子、饰品等小范围遮挡和非正面姿态场景时的优越性.

参 考 文 献

- [1] 王绪冬. 基于超分辨率重建的低分辨率人脸识别 [D]. 南京: 南京大学, 2021. (WANG X D. Low-Resolution Face Recognition Based on Face Hallucination Method [D]. Nanjing: Nanjing University, 2021.)
- [2] 陈贵强, 何军, 罗顺尧. 基于改进 CycleGAN 的视频监控人脸超分辨率恢复算法 [J]. 计算机应用研究, 2021, 38(10): 3172-3176. (CHEN G Q, HE J, LUO S C. Improved Video Surveillance Face Super-resolution Recovery Algorithm Based on CycleGAN [J]. Application Research of Computers, 2021, 38(10): 3172-3176.)
- [3] CHOU E, TAN M, ZOU C, et al. Privacy-Preserving Action Recognition for Smart Hospitals Using Low-Resolution Depth Images [EB/OL]. (2018-11-25)[2023-01-10]. <https://arxiv.org/abs/1811.09950>.
- [4] BÄTZ M, EICHENSEER A, SEILER J, et al. Hybrid Super-resolution Combining Example-Based Single-Image and Interpolation-Based Multi-image Reconstruction Approaches [C]//2015 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ: IEEE, 2015: 58-62.
- [5] ZHANG L Z, ZHANG W, LU G M, et al. Feature-Level Interpolation-Based GAN for Image Super-resolution [J]. Personal and Ubiquitous Computing, 2021, 26(4): 955-1010.
- [6] SANCHEZ-BEATO A, PAJARES G. Noniterative Interpolation-Based Super-resolution Minimizing Aliasing in the Reconstructed Image [J]. IEEE Transactions on Image Processing, 2008, 17(10): 1817-1826.
- [7] WU W, LIU Z, HE X H, et al. Learning-Based Super-resolution Using Kernel Partial Least Squares [J]. Image and Vision Computing, 2011, 29(6): 394-407.
- [8] JIA K, GONG S G. Generalized Face Super-resolution [J]. IEEE Transactions on Image Processing, 2008, 17(6): 873-886.
- [9] GAJJAR P P, JOSHI M V. New Learning Based Super-resolution: Use of DWT and IGMRF Prior [J]. IEEE Transactions on Image Processing, 2010, 19(5): 1201-1213.
- [10] YANG Z Y, SHI P, PAN D, et al. A Survey of Super-resolution Based on Deep Learning [C]//2020 International Conference on Culture-Oriented Science & Technology (ICCST). Piscataway, NJ: IEEE, 2020: 514-518.
- [11] JIANG J J, WANG C Y, LIU X M, et al. Deep Learning-Based Face Super-resolution: A Survey [J]. ACM Computing Surveys (CSUR), 2021, 55(1): 13-1-13-36.
- [12] DONG C, LOY C C, HE K M, et al. Image Super-resolution Using Deep Convolutional Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.
- [13] KIM J, LEE J K, LEE K M, et al. Accurate Image Super-resolution Using Very Deep Convolutional Networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1646-1654.
- [14] CHEN Y, TAI Y, LIU X M, et al. Fsrnet: End-to-End Learning Face Super-resolution with Facial Priors [C]//

- Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 2492-2501.
- [15] ZHANG Y L, LI K P, LI K, et al. Image Super-resolution Using Very Deep Residual Channel Attention Networks [C]//Proceedings of the European Conference on Computer Vision (ECCV). New York: ACM, 2018: 286-301.
- [16] ZHANG Y L, TIAN Y P, KONG Y, et al. Residual Dense Network for Image Super-resolution [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 2472-2481.
- [17] WANG X T, YU K, DONG C, et al. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 606-615.
- [18] HU X B, REN W Q, LAMASTER J, et al. Face Super-resolution Guided by 3D Facial Priors [C]//16th European Conference on Computer Vision. Berlin: Springer, 2020: 763-780.
- [19] PAYSAN P, KNOTHE R, AMBERG B, et al. A 3D Face Model for Pose and Illumination Invariant Face Recognition [C]//2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway, NJ: IEEE, 2009: 296-301.
- [20] BLANZ V, VETTER T. A Morphable Model for the Synthesis of 3D Faces [C]//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 1999: 187-194.
- [21] VIAZOVETSKYI Y, IVASHKIN V, KASHIN E, et al. Stylegan2 Distillation for Feed-Forward Image Manipulation [C]//16th European Conference on Computer Vision. Berlin: Springer, 2020: 170-186.
- [22] WANG X T, LI Y, ZHANG H L, et al. Towards Real-World Blind Face Restoration with Generative Facial Prior [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 9168-9178.
- [23] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Networks [J]. Communications of the ACM, 2020, 63(11): 139-144.
- [24] DENG Y, YANG J L, XU S C, et al. Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2019: 285-295.
- [25] WANG T C, LIU M Y, ZHU J Y, et al. High-Resolution Image Synthesis and Semantic Manipulation with Conditional Gans [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 8798-8807.
- [26] KARRAS T, AILA T, LAINE S, et al. Progressive Growing of Gans for Improved Quality, Stability, and Variation [EB/OL]. (2017-10-27)[2023-01-10]. <https://arxiv.org/abs/1710.10196>.

(责任编辑:韩 啸)