文章编号:1672-3961(2025)04-0018-11 DOI:10.6040/j.issn.1672-3961.0.2024.047

基于纹理和结构交互的人脸图像修复

周遵富[1,2],张乾[3*],石计亮[1,2],岳诗琴[4]
(1.贵州民族大学数据科学与信息工程学院, 贵州 贵阳 550025; 2.贵州省模式识别与智能系统重点实验室(贵州民族大学), 贵州 贵阳 550025; 3.贵州民族大学教务处, 贵州 贵阳 550025; 4.武汉理工大学汽车工程学院, 湖北 武汉 430070)

摘要:针对基于学习的人脸图像修复方法在提取深层特征时存在丢失上下文语义信息的问题,提出一种具有高效归一化注意力机制的生成器,有效提取人脸图像中的深层特征,并在多个尺度上更好地聚合低级和高级特征。为增强生成图像的一致性,提出一种具有残差主路径转换的双级门控特征融合模块,进一步融合解码后的纹理和结构信息,并设计一种增强的上下文特征聚合块,其中改进的提示生成块使提示参数在多尺度上进行特征间的交互,指导修复网络动态调整,生成现实、可信的人脸图像。试验结果表明,在 CelebA-HQ 数据集上,本研究方法的峰值信噪比 R_{PSN}、结构相似性 S_{SIM}、平均绝对误差 E_{MA}、Fréchet 初始距离 D_{FI} 分别为 37.74 dB、0.983 0、0.24%、1.489;在 LFW 数据集上,本研究方法的 R_{PSN}、S_{SIM}、E_{MA}、D_{FI} 分别为 39.19 dB、0.987 7、0.21%、3.555。与其他 5 种主流方法相比,本研究方法取得相当有竞争力的结果。定性和定量试验结果表明,本研究方法能有效恢复残损的人脸结构和纹理信息。

关键词:人脸图像修复;归一化注意力模块;提示生成模块;多尺度特征融合;双流判别器
中图分类号:TP391.41 **文献标志码**:A
引用格式:周遵富,张乾,石计亮,等. 基于纹理和结构交互的人脸图像修复[J]. 山东大学学报(工学版),2025,55(4):18-28.
ZHOU Zunfu, ZHANG Qian, SHI Jiliang, et al. Face image inpainting based on texture and structure interaction[J]. Journal of Shandong University (Engineering Science), 2025, 55(4):18-28.

Face image inpainting based on texture and structure interaction

ZHOU Zunfu[1,2], ZHANG Qian[3*], SHI Jiliang[1,2], YUE Shiqin[4]
(1. College of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, Guizhou, China; 2. Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University, Guiyang 550025, Guizhou, China; 3. Academic Affairs Office, Guizhou Minzu University, Guiyang 550025, Guizhou, China; 4. College of Automotive Engineering, Wuhan University of Technology, Wuhan 430070, Hubei, China)

Abstract: Aiming at the issue of losing contextual semantic information when extracting deep features by learning-based face image inpainting methods, a generator with an efficient normalized attention mechanism was proposed, which extracted deep features from face images more effectively and better aggregated low-level and high-level features at multiple scales. To enhance the consistency of the generated images, a bi-level gated feature fusion module with residual main path transformation was introduced. This module further fused decoded texture and structure information, and incorporated an enhanced contextual feature aggregation module, in which an improved prompt generation block enabled prompt parameters to interact between features at multiple scales, guiding the dynamic adjustment of the inpainting network to generate realistic and believable face images. Experimental results on the CelebA-HQ datasets showed that this research method achieved 37.74 dB, 0.983 0, 0.24%, and 1.489 in terms of peak signal-to-noise ratio (R_{PSN}), structural similarity (S_{SIM}), mean absolute error (E_{MA}), and Fréchet inception distance (D_{FI}). On the LFW dataset, the R_{PSN}, S_{SIM}, E_{MA}, and D_{FI} of this research method achieved 39.19 dB, 0.987 7, 0.21%, and 3.555. Compared with five other mainstream methods, this research method achieved quite competitive results. Qualitative and quantitative experiments demonstrated that this research method could effectively restore corrupted facial structure and texture information.

Keywords: face image inpainting; normalized attention module; prompt generation module; multi-scale feature fusion; two-stream discriminator

收稿日期:2024-03-04
基金项目:国家自然科学基金资助项目(62062024);贵州民族大学校级科研资助项目(GZMUZK[2021]YB23);贵州省教育厅自然科学研究资助项目(黔教技[2022]015 号)
第一作者简介:周遵富(1996—),男,贵州毕节人,硕士研究生,主要研究方向为计算机视觉。E-mail: zzf08100429@163.com
*通信作者简介:张乾(1984—),男,贵州贵定人,教授,硕士生导师,博士,主要研究方向为计算机视觉。E-mail: gzmuzq@gzmu.edu.cn

0　引言

图像修复是一种图像复原任务,其目的是在图像破损区域填充视觉上可信的语义信息,保持图像的全局一致性。图像修复广泛应用于生活中,例如图像编辑、目标移除和旧照片复原。

图像修复与大多数底层视觉任务一样,在近年取得了快速发展,主要归功于深度学习技术的广泛应用。在图像修复任务中,当图像缺损比例较大时,难以对图像的上下文语义信息推理。早期的图像修复方法通过在已知像素区域搜索最相似块,逐步填充缺失区域,在处理缺失区域较小的情况时表现良好[1-2]。基于学习的方法应用深度卷积神经网络构建能建模图像远距离关系的生成器,在不规则缺损条件下提取深层特征[3]。将基于学习的方法与传统基于补丁的方法结合用于图像修复[4-5],能生成语义关联度高的图像。文献[6]设计一个视觉结构重建(visual structure reconstruction, VSR)层,交互建模图像缺失区域的纹理和结构,通过在生成器中使用多个 VSR 层,在编码阶段逐步重建外部结构,在解码阶段重建内部结构,生成更高质量的结构信息;文献[7]在编码解码网络中引入有效性可迁移卷积(validness migratable convolution, VMC)和区域复合归一化(regional composite normalization, RCN)模块,其具备的动态选择机制使网络能避免遮挡区域无效信息的干扰,有效利用背景区域的有效信息;为了增强上下文语义推理,文献[8]设计一个聚合上下文转换(aggregated contextual-transformation, AOT)模块,捕捉信息丰富的远距离图像语义信息,增强判别器以合成清晰的纹理;文献[9]提出辅助权重适应(auxiliary weights adaptation, AWA)算法,有效确定可调控感知损失与可调控风格损失的权重,在训练过程中能较好地重构破损图像。但上述方法在面对较大残损区域时,生成的结果具有结构模糊性和纹理不一致性。

针对此问题,本研究提出一种新颖的双流人脸图像修复网络,将图像修复任务分为两个协同的重建阶段:结构约束的纹理合成和纹理引导的结构重建。通过这种协作方式,两个平行耦合的流单独建模并聚合,以实现互补。本研究应用双流判别器评估生成器的修复性能,以生成现实的像素和尖锐的边缘;设计一个新颖的残差主路径转换的双级门控特征融合模块(bi-level gated feature fusion module with residual main path transformation, BGFFM-RMPT),整合重建的结构和纹理特征图,增强整体一致性;提出增强的上下文特征聚合(enhanced contextual feature aggregation, ECFA)模块,其中的上下文注意力模块旨在捕捉来自图像中遥远空间位置信息,以增强生成图像的细节。受 Mish 激活函数的启发[10],本研究将原归一化注意力机制[11]中的 ReLU 激活函数替换为 Mish 激活函数,在生成器中添加改进的归一化注意力模块,加快模型的收敛速度。由于具备双重生成网络及专门设计的模块,本研究能重建更丰富的视觉结构和纹理信息。所提模型在 CelebA-HQ[12]和 LFW[13]数据集上进行广泛试验,在定性和定量方面取得优越的生成结果。

1　背景知识

1.1　传统的和基于学习的图像修复方法

截至目前,图像修复领域存在两类方法:传统的图像修复方法和基于学习的图像修复方法。传统的图像修复方法有基于扩散的方法[14-15]和基于补丁的方法[16-17]。前者利用偏微分方程和变分方法,从有效像素区域向缺失区域内部进行信息扩散;后者通过复制粘贴图像内容,从已知区域向未知区域传播。这两类方法在处理较小的缺失区域时,能取得较好的视觉效果,但当缺失区域较大时,由于无法学习图像中更多有效像素信息,不能合成大尺度的场景结构和目标。

基于学习的图像修复方法得到快速改进,将生成对抗网络(generative adversarial network, GAN)应用于图像补全任务中[18],学习人脸图像的深度特征,对图像的残缺部分更好地推理,为生成上下文语义连续、纹理结构协调的人脸图像提供基础。GAN 中含有一个生成器及一个能实时反馈信息的判别器,其中生成器接收含有噪声的向量并生成固定大小的人脸图像,判别器断定生成器输出的图像是否为真实样本。为了恢复图像的全局结构,两阶段的修复方法广泛应用于图像补全任务,第一阶段在缺失区域初步生成图像结构,在第二阶段的修复网络中利用初步修复的图像结构指导像素合成。文献[19-20]以边缘图为依据恢复图像的全局结构;文献[21]利用提取到的深层次特征预测图像中的轮廓和位置等信息,利用粗修复阶段生成的特征信息对图像整体颜色和亮度进行恢复。以上方法利用边缘图作为依据恢复图像的纹理及结构信息,取得较好的成果,但在残损图像中获得准确的边缘图存在非常大的挑战,在大面积缺损的情况下难以保

证生成逼真的图像。

1.2 注意力模块

注意力模块已经在计算机视觉领域得到广泛应用,将注意力模块应用于图像修复网络中,可以抑制无关区域特征,增强相关区域特征表达,确保网络更好地获取上下文语义信息,重建结构和纹理特征。文献[22]提出一种金字塔上下文编码器网络,在高级语义特征图和低级特征图之间用转移注意力补全缺失部位,确保视觉和语义的协调性;文献[23]提出一个含有粗修复网络和精修复网络的方案,在精修复网络中引入上下文注意力建模长期相关性,利用周围像素信息作参考以得到更好的预测;文献[24]在修复网络中引入长短期注意力层,利用解码器和编码器中的特征信息,改善外观一致性;文献[25]提出一种用于图像修复的多尺度注意力网络,全面分析从浅层细节到高层语义的特征,取得良好的图像修复结果。

1.3 激活函数

激活函数可视为卷积神经网络中的一个特殊层,即实现非线性映射的层。卷积神经网络在进行线性变换后,通常会接一个非线性激活函数,可以使数据分布重新映射,有助于增强卷积神经网络的非线性表达能力。文献[26]使用 ReLU 激活函数,其表达式简单易求导,加速了模型训练速度,正半轴导数恒为 1,弥合了 S 型激活函数存在的梯度爆炸这一鸿沟。但 ReLU 激活函数在负半轴的梯度为 0,在模型学习率设置较大时,会出现死亡神经元的情况。为缓解此问题,一种自正则化的激活函数 Mish 相比 ReLU 具有更强的模型表达能力和函数逼近能力,定义域包含负值区域,且负值区域具有梯度信息,有利于学习更丰富的特征表示,提升特征提取能力,解决训练过程中的梯度消失问题,使模型训练更加稳定。Mish 函数曲线的平滑性能减少模型训练过程中的抖动,让生成模型重构更逼真的人脸图像。

2 模型介绍

2.1 生成器

生成器是由 U-Net 变体构建的双流结构,如图 1 中生成器部分所示。在编码阶段,将破损人脸图像及其相应的边缘图分别投射到潜在空间,其中纹理分支注重于纹理特征,结构分支则针对结构特征。在解码阶段,纹理解码器通过从结构编码器中提取结构特征合成结构受限的纹理,结构解码器通过从纹理编码器中提取纹理特征恢复纹理引导的结构信息。采用上述双生成模式,结构和纹理能很好地相互补充,得到改善的结果。在基于编解码器的主干中,本研究加入 Mish 激活函数,更好地捕捉不规则边界信息。为了提高重建结构和纹理的一致性,本研究将两个分支输出的特征图通过专门设计的 BGFFM-RMPT 和 ECFA 模块进一步融合,结合跳跃连接优势,在多个尺度上融合低级和高级语义特征,最终对特征进行细化。

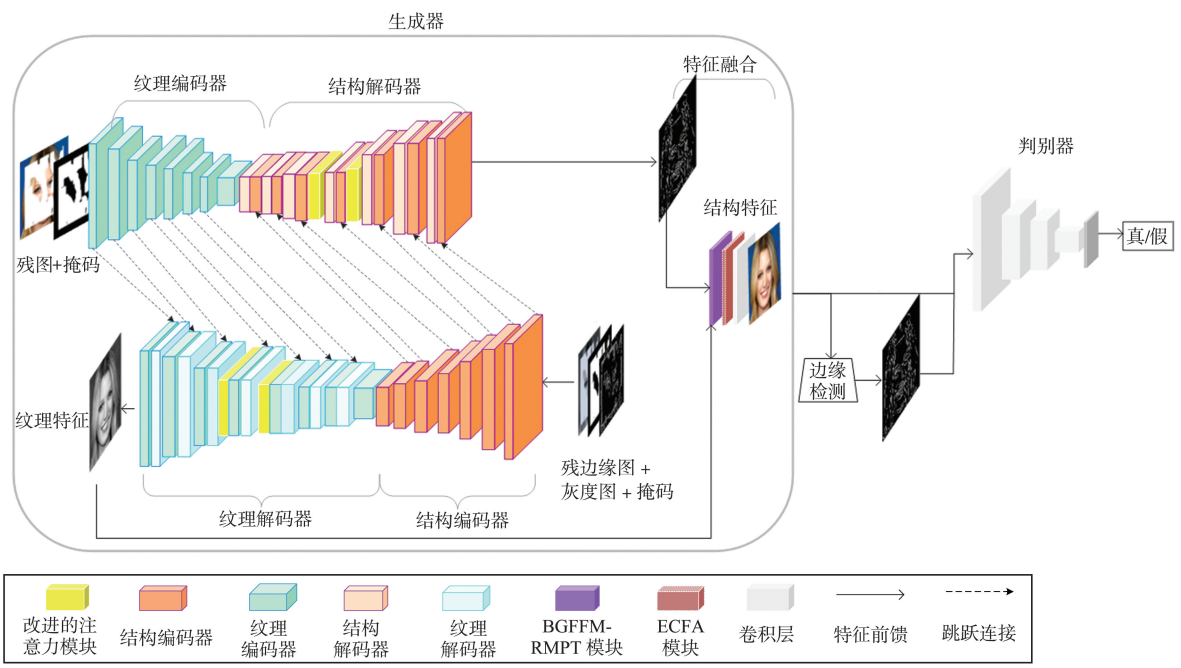

图 1 网络结构概览图

Fig.1 Overview of the network structure

2.1.1　BGFFM-RMPT 模块

受文献[27]启发,本研究在双级门控特征融合模块中加入残差主路径转换块,在空间和通道维度进一步整合解码后的纹理和结构特征,在两种信息之间进行交互。这种交互操作使纹理和结构特征被细化。BGFFM-RMPT 模块如图 2 所示。在模块中,解码器输出的纹理特征图为 $\boldsymbol{f}_{\mathrm{t}}$,结构特征图为 $\boldsymbol{f}_{\mathrm{s}}$。为了建立纹理感知的结构特征,利用软门控 $\boldsymbol{g}_{\mathrm{t}}$ 控制纹理信息的整合程度,$\boldsymbol{g}_{\mathrm{t}}$ 计算式为

$$\boldsymbol{g}_{\mathrm{t}}=\sigma(l_{\mathrm{m}}(\mathrm{Concat}(\boldsymbol{f}_{\mathrm{t}},\boldsymbol{f}_{\mathrm{s}}))),\tag{1}$$

式中,Concat(·)为将 $\boldsymbol{f}_{\mathrm{t}}$ 和 $\boldsymbol{f}_{\mathrm{s}}$ 在通道维度进行拼接,$l_{\mathrm{m}}(\cdot)$ 为由 3×3 卷积核实现的映射函数,$\sigma(\cdot)$ 为 Sigmoid 激活函数。通过 $\boldsymbol{g}_{\mathrm{t}}$ 可自适应地将 $\boldsymbol{f}_{\mathrm{t}}$ 合并到 $\boldsymbol{f}_{\mathrm{s}}$,得到纹理感知的结构特征

$$\boldsymbol{f}'_{\mathrm{s}}=\alpha(\boldsymbol{g}_{\mathrm{t}}\odot\boldsymbol{f}_{\mathrm{t}})\oplus\boldsymbol{f}_{\mathrm{s}},\tag{2}$$

式中,α 为初始化为 0 的训练参数,⊙和⊕分别为元素间相乘和元素间相加。

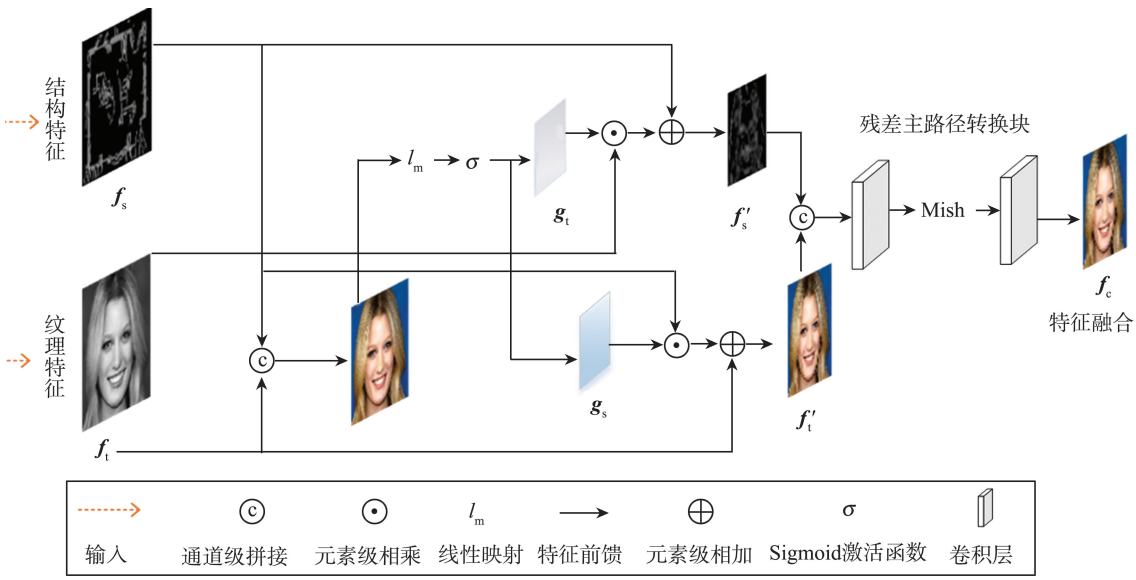

图 2　BGFFM-RMPT 模块

Fig.2　BGFFM-RMPT module

同理,结构感知的纹理特征

$$\boldsymbol{f}'_{\mathrm{t}}=\beta(\boldsymbol{g}_{\mathrm{s}}\odot\boldsymbol{f}_{\mathrm{s}})\oplus\boldsymbol{f}_{\mathrm{t}},\tag{3}$$

式中:β 为初始化为 0 的训练参数;$\boldsymbol{g}_{\mathrm{s}}$ 为用于控制结构信息整合程度的软门控,$\boldsymbol{g}_{\mathrm{s}}=\boldsymbol{g}_{\mathrm{t}}$。将 $\boldsymbol{f}'_{\mathrm{s}}$和 $\boldsymbol{f}'_{\mathrm{t}}$融合,以通道拼接的方式获得聚合的特征图,将其传输至残差主路径转换模块中细化特征,其中残差主路径转换模块由两个 3×3 卷积层和一个 Mish 激活函数层组成,以提高模型的表达能力。上述过程的表达式为

$$\boldsymbol{f}_{\mathrm{c}}=\mathrm{RMPT}(\mathrm{Concat}(\boldsymbol{f}'_{\mathrm{s}},\boldsymbol{f}'_{\mathrm{t}})),\tag{4}$$

式中,RMPT 为残差主路径转换模块函数,$\boldsymbol{f}_{\mathrm{c}}$ 为聚合的特征图。

2.1.2　ECFA 模块

为精准捕捉有助于填补缺失区域的有效信息,本研究设计一个增强的上下文特征聚合块,如图 3 所示,增强输入图像松散特征间的关联性,使预测图像具有全局和局部一致性。注意力模块具有建模远程依赖性的超强能力,避免信息流失问题。在本研究中引入上下文注意力模块,该模块具有很强的自适应和全局感知能力,使模型能够根据输入自适应调整,更好地学习和利用图像内在的语义和结构信息,显著提高修复结果的一致性和纹理质量。受文献[28]的启发,本研究提出改进的提示生成模块,使提示参数适应不同的输入特征信息,提高提示质量和表达能力。改进的提示生成模块还可以使提示参数在多个尺度上与修复网络的特征进行交互,丰富退化类型信息。在本研究中采用多尺度特征聚合编码多个尺度中丰富的语义特征,更好地平衡精度和复杂度,以处理更复杂的缺损情况,生成高度可信的人脸图像。

通过计算查询特征和所有其他位置特征之间的相关性,上下文注意力模块可以从整个图像中捕捉与当前位置相关的全局上下文信息,对于修复大范围缺失区域时保持语义和结构的一致性非常关键。输入一个聚合的特征图 $\boldsymbol{f}_{\mathrm{c}}\in\mathbf{R}^{C\times HW}$,其中 C 为通道的维度,H 为图像的高度,W 为图像的宽度,从背景区域提取 3×3 的块,将其重塑为卷积核。对于特征图的第 i 个背景块 $\boldsymbol{f}_{\mathrm{c}_i}$与第 j 个前景块 $\boldsymbol{f}_{\mathrm{c}_j}$的匹配度 $\boldsymbol{s}_{i,j}\in\mathbf{R}^{HW\times HW}$,用余弦相似度 $\hat{\boldsymbol{s}}_{i,j}$度量两个区域的相似性,再用缩放的 Softmax 衡量两个部分间的匹配度。上述过程的计算式为

$$\hat{\boldsymbol{s}}_{i,j}=\left\langle \frac{\boldsymbol{f}_i}{\|\boldsymbol{f}_i\|},\frac{\boldsymbol{f}_j}{\|\boldsymbol{f}_j\|}\right\rangle, \tag{5}$$

$$\boldsymbol{s}_{i,j}=\frac{\exp(\hat{\boldsymbol{s}}_{i,j})}{\sum_k \exp(\hat{\boldsymbol{s}}_{i,k})}, \tag{6}$$

式中〈 〉为内积。

通过矩阵乘法得到 $\boldsymbol{f}_c$ 的加权平均形式

$$\tilde{\boldsymbol{f}}=\boldsymbol{f}_c\cdot\boldsymbol{s}_{i,j}。 \tag{7}$$

通过拼接 $\boldsymbol{f}_c$ 和 $\tilde{\boldsymbol{f}}$ 得到重建特征图 $\boldsymbol{f}_{rec}$，使用一个 1×1 的卷积层保留 $\boldsymbol{f}_c$ 的原始通道数。整个过程从背景块中借鉴相似的特征信息对缺失区域建模。将 $\boldsymbol{f}_{rec}$ 输入改进的提示生成模块中，以准确捕捉特征信息。

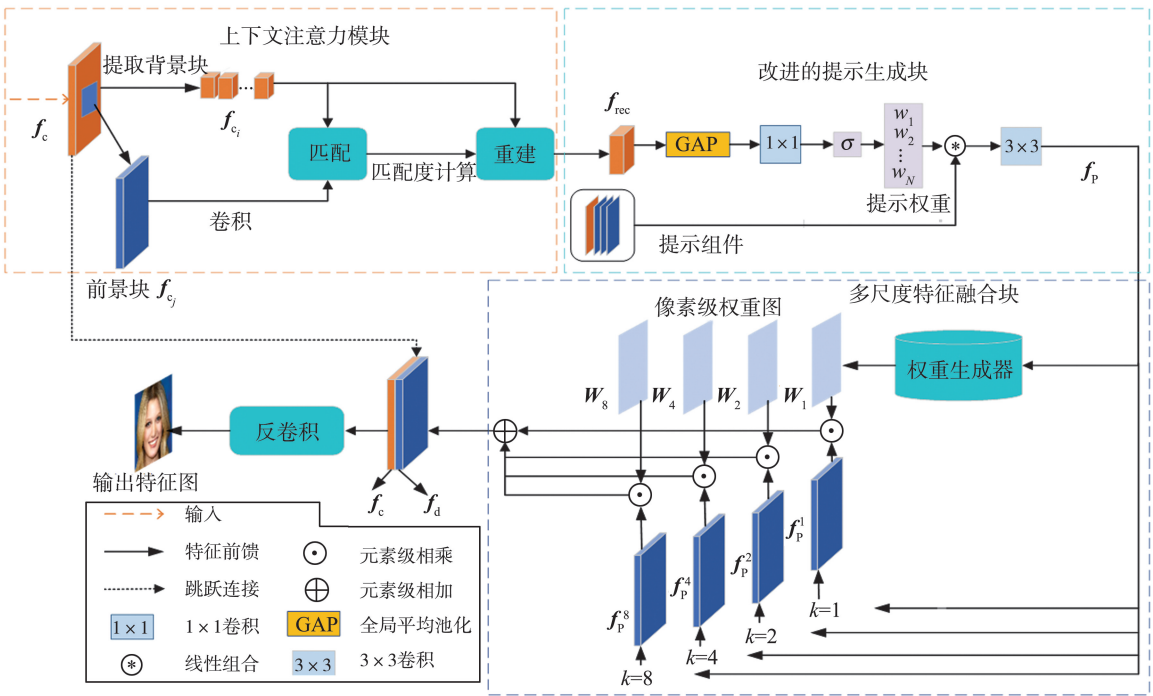

图 3 增强的上下文特征聚合块

Fig.3 Enhanced contextual feature aggregation block

在推理时，修复网络必须能处理不同分辨率的图像。因此，本研究采用双线性上采样操作，将提示组件提升到与输入特征相同的大小。改进的提示生成模块（improved prompt generation module，IPGM）中，提示组件 $\boldsymbol{P}_c$ 构成一组可学习的参数，与输入的特征相互作用，嵌入退化信息，从输入特征中动态预测基于注意力的权重。这些权重反映不同提示组件对输入特征的重要性，将其应用到提示组件中，得到以输入为条件的提示 $\boldsymbol{f}_P$。这些权重参数包含退化类型的上下文信息，指导修复网络动态调整。IPGM 还创建了一个共享空间，以促进提示组件之间的相关信息共享。为了从输入特征图 $\boldsymbol{f}_{rec}$ 中生成提示权重，IPGM 在空间维度上应用全局平均池化生成特征向量 $\boldsymbol{v}\in\mathbf{R}^{\hat{C}}$，其中 $\hat{C}$ 为空间维度。$\boldsymbol{v}$ 经过通道降维卷积层得到一个紧凑的特征向量，进行归一化操作得到提示权重；利用这些权重对提示组件进行调整，输入一个带有批量归一化和 Mish 激活函数的 3×3 卷积层，得到 $\boldsymbol{f}_P$。IPGM 的实现过程可概括为

$$\boldsymbol{f}_P=\mathrm{Conv}_{3\times3}\left(\sum_{c=1}^{N}\mathrm{Softmax}(\mathrm{Conv}_{1\times1}(\mathrm{GAP}(\boldsymbol{f}_{rec})))\boldsymbol{P}_c\right), \tag{8}$$

式中，$\mathrm{Conv}_{3\times3}(\cdot)$ 为 3×3 的卷积层，$\mathrm{Softmax}(\cdot)$ 为归一化操作，$\mathrm{Conv}_{1\times1}(\cdot)$ 为 1×1 的卷积层，$\mathrm{GAP}(\cdot)$ 为全局平均池化层。

在重建特征图时，采用 4 组不同扩张率的卷积层提取人脸图像的多尺度语义信息，得到不同扩张率的提示

$$\boldsymbol{f}_P^k=\mathrm{Conv}_k(\boldsymbol{f}_P), \tag{9}$$

式中：$\mathrm{Conv}_k(\cdot)$ 为膨胀卷积层；k 为扩张率，$k\in\{1,2,4,8\}$。

为更好地整合多尺度语义特征，引入一个像素级权重图生成器 G_w，预测像素级的权重图。在试验中，权重图生成器由两个卷积层组成，核大小分别为 3×3 和 1×1，每个核都经过 Mish 非线性激活，权重图生成器的输出通道设置为 4。像素级权重图

$$\boldsymbol{W}=\mathrm{Softmax}(G_w(\boldsymbol{f}_P))。 \tag{10}$$

以通道级方式 $\mathrm{Slice}(\cdot)$ 将 $\boldsymbol{W}$ 切分成 4 个像素

级特征图

$$\boldsymbol{W}_1,\ \boldsymbol{W}_2,\ \boldsymbol{W}_4,\ \boldsymbol{W}_8=\mathrm{Slice}(\boldsymbol{W})。\tag{11}$$

汇总多尺度语义特征信息,合成精细的特征图

$$\boldsymbol{f}_{\mathrm{d}}=(\boldsymbol{f}_{\mathrm{P}}^{1}\odot\boldsymbol{W}_1)\oplus(\boldsymbol{f}_{\mathrm{P}}^{2}\odot\boldsymbol{W}_2)\oplus(\boldsymbol{f}_{\mathrm{P}}^{4}\odot\boldsymbol{W}_4)\oplus(\boldsymbol{f}_{\mathrm{P}}^{8}\odot\boldsymbol{W}_8)。\tag{12}$$

此外,在 ECFA 模块中采用跳跃连接[29],防止由于块移动操作造成语义损害,将一对卷积层和反卷积层无缝嵌入网络架构中,以提高计算效率,得到最终修复的人脸图像。

2.2　判别器

在文献[30-31]的启发下,本研究引入一个双流判别器[27],通过评估纹理和结构特征统计区分原图像和生成图像。判别器如图 1 所示。纹理分支包括:3 个卷积层,核大小为 4×4,步距为 2;2 个卷积层,核大小为 4×4,步距为 1;最后一层使用 Sigmoid 非线性激活函数。结构分支类似于纹理分支,输入的边缘图由一个残差块检测,接着是核大小为 1×1 的卷积层。2 个分支的输出在通道维度上拼接,计算出对抗性损失,更好地指导对残损人脸图像的修复。

2.3　损失函数

2.3.1　符号及定义

定义$\{(\boldsymbol{X}_i,\boldsymbol{I}_{\mathrm{gt}})\}_i$为训练数据集,其中$\boldsymbol{X}_i$和$\boldsymbol{I}_{\mathrm{gt}}$分别为残损图像和原始图像。将$\boldsymbol{X}_i$输入含有参数化的生成器 G 中,预测出缺损位置的像素,得到输出图像

$$\boldsymbol{I}_{\mathrm{out}}=G(\boldsymbol{X}_i,\theta),\tag{13}$$

式中 θ 为模型优化参数。

本研究提出的模型用联合损失训练,包括平均误差损失、总方差损失、对抗性损失、感知损失和风格损失,以呈现视觉上真实、语义上合理的人脸图像。联合损失

$$L_{\mathrm{union}}(\theta)=\boldsymbol{\lambda}_{\mathrm{p}}\cdot(L_{\mathrm{mae}}(\theta)\ L_{\mathrm{adv}}(\theta)\ L_{\mathrm{tv}}(\theta))^{\mathrm{T}}+\boldsymbol{\lambda}_{\mathrm{f}}\cdot(L_{\mathrm{per}}(\theta)\ L_{\mathrm{sty}}(\theta))^{\mathrm{T}},\tag{14}$$

式中:$L_{\mathrm{mae}}(\cdot)$为平均绝对误差损失;$L_{\mathrm{adv}}(\cdot)$为对抗性损失;$L_{\mathrm{tv}}(\cdot)$为总方差损失;$\boldsymbol{\lambda}_{\mathrm{p}}$为权衡参数向量;$\boldsymbol{\lambda}_{\mathrm{f}}$为感知损失和风格损失的权重向量;$L_{\mathrm{per}}(\theta)$为感知损失,$L_{\mathrm{per}}(\theta)=E\left[\sum_i\|\boldsymbol{\phi}_i(\boldsymbol{I}_{\mathrm{out}})-\boldsymbol{\phi}_i(\boldsymbol{I}_{\mathrm{gt}})\|_1\right]$,其中 E 为特征空间中的期望,$\boldsymbol{\phi}_i$ 为输出视觉几何组(visual geometry group, VGG)网络第 i 层的特征图;$L_{\mathrm{sty}}(\theta)$为风格损失,$L_{\mathrm{sty}}(\theta)=E\left[\sum_i\|\boldsymbol{G}_i(\boldsymbol{I}_{\mathrm{out}})-\boldsymbol{G}_i(\boldsymbol{I}_{\mathrm{gt}})\|_1\right]$,其中 $\boldsymbol{G}_i(\cdot)$为格拉姆矩阵,$\boldsymbol{G}_i(\cdot)=\boldsymbol{\phi}_i(\cdot)\boldsymbol{\phi}_i(\cdot)^{\mathrm{T}}$。感知损失和风格损失在特征空间计算,用于更好地恢复结构和纹理信息。

2.3.2　可调控的感知和风格损失

为更好地挖掘辅助损失潜力,在训练过程中需要可调控的权重衡量不同类别损失项的贡献程度。本研究引入文献[9]中的可调控感知损失(tunable perceptual loss, TPL)和可调控风格损失(tunable style loss, TSL),进一步应用辅助权重适应算法,使用一组连续值的权重独立对每个损失项进行加权,自适应地调整为最佳权重,减少对网格的搜索,极大缩短训练时间。在训练模型 K 次后,用学习感知图像块相似度指标衡量修复性能。可调控感知损失的深层特征从预训练的 VGG-16 网络前 3 个最大池化层中提取,可调控感知损失 $L'_{\mathrm{per}}(\theta)$ 是感知损失的超集,即

$$L'_{\mathrm{per}}(\theta)=(\lambda_{\mathrm{p,max}}\odot\sigma(\eta_{\mathrm{p}}))\cdot(L^1_{\mathrm{per}}(\theta)\ L^2_{\mathrm{per}}(\theta)\ \cdots\ L^N_{\mathrm{per}}(\theta))^{\mathrm{T}},\tag{15}$$

式中,$\lambda_{\mathrm{p,max}}$为 TPL 的最大权重,η_{p} 为连续型参数,$L^N_{\mathrm{per}}(\theta)$为第 N 个感知损失。

可调控风格损失在标准风格损失的基础上增加 η_{s},η_{s} 为可调控风格损失的连续型参数。η_{p} 和 η_{s} 由 Sigmoid 激活函数激活通过 $\lambda_{\mathrm{p,max}}$ 和 TSL 的最大权重 $\lambda_{\mathrm{s,max}}$ 重定,最终的权值范围分别为 $0\sim\lambda_{\mathrm{p,max}}$ 和 $0\sim\lambda_{\mathrm{s,max}}$。可调控风格损失为

$$L'_{\mathrm{sty}}(\theta)=(\lambda_{\mathrm{s,max}}\odot\sigma(\eta_{\mathrm{s}}))\cdot(L^1_{\mathrm{sty}}(\theta)\ L^2_{\mathrm{sty}}(\theta)\ \cdots\ L^N_{\mathrm{sty}}(\theta))^{\mathrm{T}},\tag{16}$$

式中 $L^N_{\mathrm{sty}}(\theta)$为第 N 个风格损失。新的总损失

$$L_{\mathrm{total}}(\theta)=\boldsymbol{\lambda}_{\mathrm{p}}\cdot(L_{\mathrm{mae}}(\theta)\ L_{\mathrm{adv}}(\theta)\ L_{\mathrm{tv}}(\theta))^{\mathrm{T}}+L'_{\mathrm{per}}(\theta)+L'_{\mathrm{sty}}(\theta)。\tag{17}$$

3　试验结果与分析

本研究提出的方法基于 Pytorch 2.0.1 版本进行开发,使用 Linux 系统、NVIDIA GPU A100 图形处理器进行试验。模型的训练以端到端的方式进行,优化器为 AdamW,优化器参数分别设置为 $\beta_1=0.5$ 和 $\beta_2=0.999$。学习率为 0.001。参考文献[9],TPL 和 TSL 中的 $\lambda_{\mathrm{p,max}}$ 和 $\lambda_{\mathrm{s,max}}$ 分别为 2 和 750。

在公开数据集 CelebA-HQ、LFW 和文献[32]提出的不规则掩码数据集上,将图像大小设置为 256 像素×256 像素。在 CelebA-HQ 数据集中抽取 27 000张图像作训练,3 000 张图像作测试;在 LFW 数据集中抽取 11 910 张图像作训练,1 323 张图像作测试。本研究所有试验在两个公共数据集上训练 30 万次,通过定性和定量评价方式与当前优秀的图像修复方法进行比较。通过消融试验分析本研究提出的各模块对修复性能的影响。

3.1 定性评价

为客观比较不同图像修复方法的效果,使用本研究方法与渐进式重建视觉结构(progressive reconstruction of visual structure, PRVS)[6]、动态选择网络(dynamic selection network, DSNet)[7]、AOT[8]、辅助损失重加权(auxiliary loss reweighting, ALR)[9]和沙漏注意网络(hourglass attention network, HAN)[33]方法进行比较。各方法在CelebA-HQ数据集中不规则损坏图像上的定性比较结果如图4所示,各方法在LFW数据集上的定性比较结果如图5所示。由图4(a)~(h)可以看出,本研究方法与其他方法相比,在眼睛和耳朵等部位的修复结果具有少量的失真现象,更接近真实图像。由图4(i)~(p)可以看出,本研究方法在帽子边缘处具有少量的伪影,能生成相邻像素过渡自然的人脸图像,而其他方法生成的人脸图像具有伪影和相邻像素过渡不自然的现象。由图4(q)~(x)可以看出,本研究方法成功修复被遮挡的眼镜,生成具有真实性和一致性的人脸图像,而其他方法生成的结果具有冗余信息。在图5(a)~(h)中,与其他5种方法相比,本研究提出的方法在帽子和线条等部位的修复结果具有更好的清晰性和连贯性。在图5(i)~(p)中,与其他5种方法相比,本研究方法在耳朵及左肩部位的修复效果更自然,其中HAN方法虽然很好地恢复了遮挡的鼻子,但在白色背景处生成较多的冗余信息,对衣襟的修复效果较差。在图5(q)~(x)中,与其他5种方法相比,本研究提出的方法对损坏的眼睛和帽子等处的修复结果具有一致的结构信息和清晰的纹理细节。综上,由可视化比较结果可知,在CelebA-HQ和LFW数据集上使用不规则掩码数据集时,本研究方法的修复结果具有全局一致性和清晰的纹理细节。

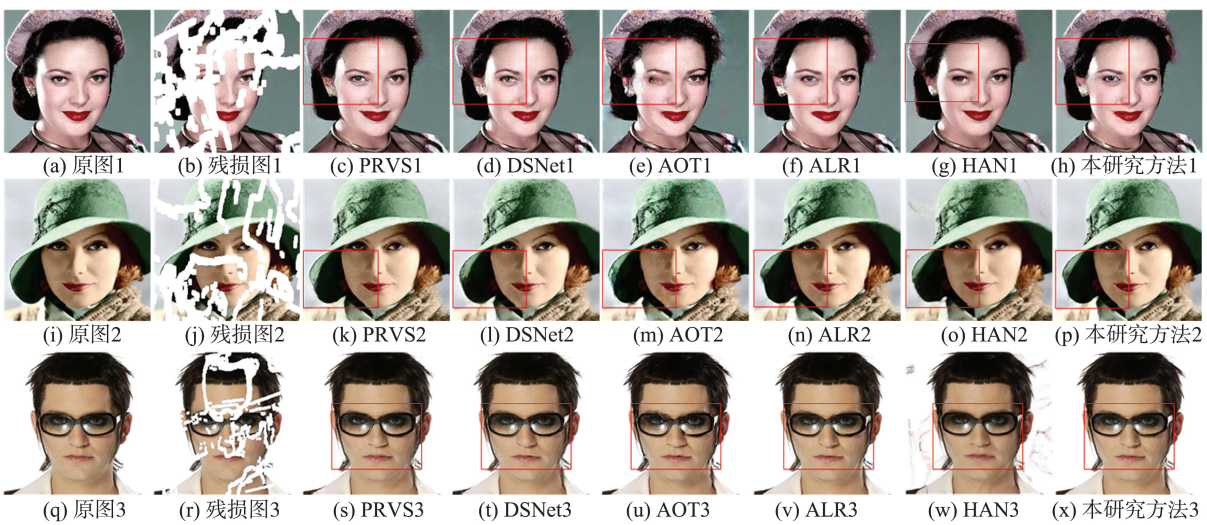

图4 各方法在CelebA-HQ数据集上的定性比较结果

Fig.4 Qualitative comparison of methods on the CelebA-HQ datasets

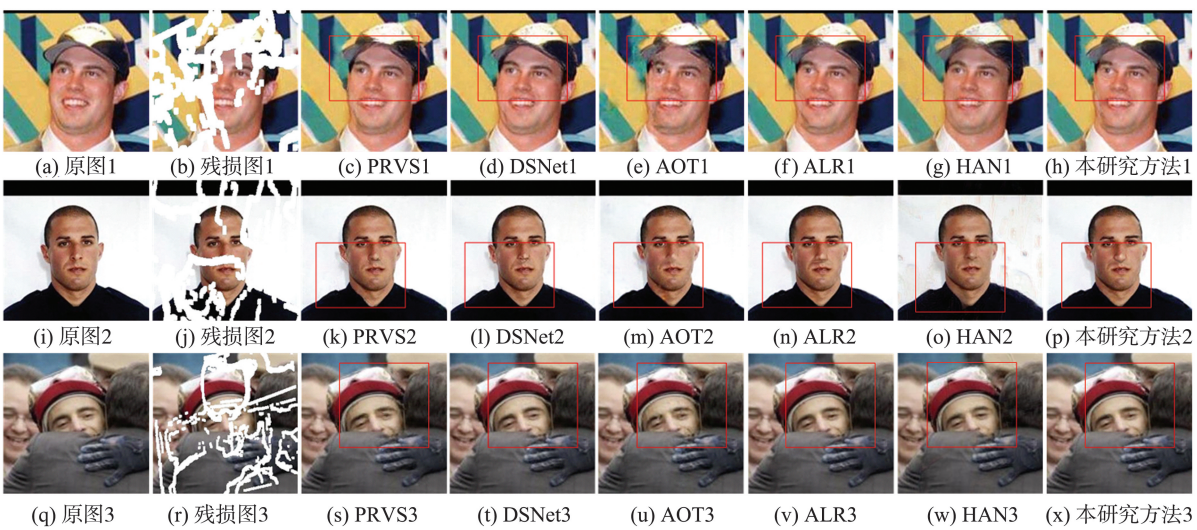

图5 各方法在LFW数据集上的定性比较结果

Fig.5 Qualitative comparison of methods on the LFW datasets

3.2　定量评价

为客观评价本研究的图像修复效果,使用峰值信噪比 R_{PSN}、平均绝对误差 E_{MA}、结构相似性 S_{SIM} 和 Fréchet 初始距离 D_{FI} 作为评价指标,其中 R_{PSN} 衡量图像噪声水平;E_{MA} 衡量生成图像与真实图像在像素级别的误差;S_{SIM} 衡量原图与修复图的相似度;由于网络中使用感知损失和风格损失,本研究采用 D_{FI} 衡量生成图像与真实图像间的特征分布。R_{PSN} 和 S_{SIM} 越高,图像修复效果越好;E_{MA} 越低,生成的图像质量越高;D_{FI} 越低,图像在高层特征上越相似,感知质量越好。这些评价指标的组合能在像素级别和特征层面衡量模型的修复效果。在 CelebA-HQ 和 LFW 数据集上,将本研究方法与 PRVS、DSNet、AOT、ALR 和 HAN 方法进行比较,对比结果如表 1、2 所示,其中最优结果加粗表示。

表 1　各方法在 CelebA-HQ 数据集上的定量比较结果
Table 1　Quantitative comparison of methods on the CelebA-HQ dataset

方法	掩码比例/%	R_{PSN}/dB	S_{SIM}	E_{MA}	D_{FI}
PRVS	[1,10)	35.46	0.969 9	0.005 1	1.977
	[10,20)	32.41	0.933 6	0.009 3	4.548
	[20,30)	30.56	0.904 3	0.012 3	6.564
	[30,40)	28.48	0.873 6	0.016 0	9.795
	[40,50)	26.54	0.827 7	0.021 0	9.564
DSNet	[1,10)	35.26	0.971 3	0.004 4	1.874
	[10,20)	34.27	0.957 2	0.006 4	2.511
	[20,30)	32.03	0.932 9	0.009 2	**3.572**
	[30,40)	29.29	0.901 3	0.013 4	**6.518**
	[40,50)	27.10	0.858 5	0.018 6	**6.908**
AOT	[1,10)	34.86	0.967 6	0.011 4	2.686
	[10,20)	32.59	0.941 6	0.015 0	4.567
	[20,30)	30.35	0.912 6	0.018 8	6.658
	[30,40)	27.60	0.866 8	0.025 8	13.798
	[40,50)	25.00	0.805 8	0.036 0	20.383
ALR	[1,10)	37.46	0.982 7	0.002 5	1.491
	[10,20)	35.07	0.966 2	0.004 5	2.429
	[20,30)	32.62	0.943 5	0.006 9	3.663
	[30,40)	29.70	0.912 7	0.011 0	7.875
	[40,50)	27.21	0.869 1	0.016 8	9.707
HAN	[1,10)	36.45	0.975 8	0.004 6	**1.471**
	[10,20)	32.88	0.940 2	0.008 9	3.480
	[20,30)	30.94	0.911 7	0.012 0	5.338
	[30,40)	28.87	0.887 0	0.015 7	7.819
	[40,50)	27.13	0.854 5	0.020 1	6.496
本研究方法	[1,10)	**37.74**	**0.983 0**	**0.002 4**	1.489
	[10,20)	**35.39**	**0.967 7**	**0.004 3**	**2.234**
	[20,30)	**32.88**	**0.945 3**	**0.006 7**	3.574
	[30,40)	**29.88**	**0.914 0**	**0.010 7**	7.405
	[40,50)	**27.28**	**0.869 3**	**0.016 5**	10.754

表 2　各方法在 LFW 数据集上的定量比较结果
Table 2　Quantitative comparison of methods on the LFW dataset

方法	掩码比例/%	R_{PSN}/dB	S_{SIM}	E_{MA}	D_{FI}
PRVS	[1,10)	35.07	0.967 5	0.006 9	5.544
	[10,20)	34.06	0.956 8	0.008 5	7.086
	[20,30)	32.17	0.940 8	0.010 6	9.791
	[30,40)	28.91	0.914 2	0.014 7	17.523
	[40,50)	26.68	0.877 2	0.019 9	17.339
DSNet	[1,10)	35.85	0.976 5	0.004 0	4.255
	[10,20)	35.71	0.971 1	0.005 2	4.942
	[20,30)	33.49	0.956 5	0.007 4	6.817
	[30,40)	29.76	0.928 6	0.011 7	**13.067**
	[40,50)	**27.31**	0.890 9	0.017 0	**13.221**
AOT	[1,10)	36.44	0.974 8	0.009 5	5.433
	[10,20)	34.73	0.961 7	0.012 0	7.120
	[20,30)	32.28	0.942 9	0.015 3	10.823
	[30,40)	28.30	0.902 1	0.023 1	22.006
	[40,50)	25.74	0.854 6	0.032 7	23.033
ALR	[1,10)	38.74	0.986 8	0.002 2	3.688
	[10,20)	36.75	0.978 6	0.003 6	4.916
	[20,30)	34.12	0.964 6	0.005 5	7.067
	[30,40)	29.92	0.937 2	0.009 9	14.947
	[40,50)	27.11	0.897 7	0.015 8	16.625
HAN	[1,10)	36.87	0.978 7	0.004 4	4.353
	[10,20)	33.27	0.944 2	0.008 5	9.129
	[20,30)	30.87	0.908 3	0.012 3	15.958
	[30,40)	27.87	0.877 3	0.017 2	23.381
	[40,50)	25.85	0.846 8	0.023 0	21.610
本研究方法	[1,10)	**39.19**	**0.987 7**	**0.002 1**	**3.555**
	[10,20)	**37.07**	**0.980 0**	**0.003 4**	**4.709**
	[20,30)	**34.33**	**0.966 2**	**0.005 3**	**6.803**
	[30,40)	**30.03**	**0.938 4**	**0.009 6**	15.267
	[40,50)	27.21	**0.898 3**	**0.015 6**	17.556

由表 1、2 可知:本研究提出的方法在 E_{MA} 和 S_{SIM} 指标上取得了最优结果,表明本研究提出的基于纹理和结构的相互指导修复方法能生成高质量、高保真的人脸图像;在 R_{PSN}、D_{FI} 指标上,虽然本研究方法中个别指标不是最好的,但也具有相当的竞争力;DSNet 具备的动态选择机制对无约束自然场景人脸图像的大面积缺失区域复原效果略好。综上,优越的定量比较结果表明,本研究方法对人脸图像修复任务是有效的。

3.3　消融试验

为了研究各模块对图像修复效果的具体贡献,本研究在 LFW 数据集上进行消融试验,具体设置如下。

试验 1:以 ALR 为基准[9],生成器中未加入改

进的注意力模块,在解码器中应用 Mish 激活函数,引入上下文注意力模块、改进的提示生成模块、多尺度特征融合模块、含有残差块主路径转换的双级门控特征融合模块。

试验 2:在试验 1 的基础上,将解码器中的激活函数替换为 Leaky ReLU 激活函数。

试验 3:在试验 2 的基础上,去掉残差主路径转换模块。

试验 4:在试验 1 的基础上,在生成器中添加改进的归一化注意力模块。

不同模块在图像修复任务上的可视化结果如图 6 所示。对比试验 1 和 2,由图 6(a)~(d)可以看出,试验 1 的生成结果连续性比试验 2 好,边缘更清晰;在图 6(g)~(j)中,试验 1 的生成结果在帽子和黑色衣领处的连续性优于试验 2,证明 Mish 激活函数提升了模型的表征能力。对比试验 2 和 3,在图 6(d)、(e)中,试验 3 生成了冗余信息(如花纹和头发上方白色线条部分),试验 2 生成的线条更清晰;在图 6(j)、(k)中,试验 2 在帽子的白线、拇指间和黑色衣领处保留了更丰富的细节信息,而试验 3 丢失了应有的图像信息,验证了本研究提出的残差主路径转换模块的重要性。对比试验 1 和 4,试验 4 的模型整合了各个模块的优点,在人脸的各个遮挡部位能得到综合改善。

不同模型在多个评价指标上的表现如表 3 所示,其中最优结果加粗表示。对比试验 1 和 2 可以看出,在解码器中使用 Mish 激活函数能取得较好的结果,原因是该激活函数能提高模型的泛化能力;对比试验 2 和 3 可以看出,未添加本研究提出的残差主路径转换模块时取得的结果较差;通过试验 4 的结果可知,当加入注意力模块后,虽然个别定量结果并非最优,但模型的修复效果总体得到提升,如在大面积缺损条件下能得到更好的 R_{PSN},最优结果为 26.07 dB。综上,从定性和定量两个方面全面验证了本研究提出的各模块设计的有效性。

(a) 原图1 (b) 残损图1 (c) 试验1-1 (d) 试验2-1 (e) 试验3-1 (f) 试验4-1

(g) 原图2 (h) 残损图2 (i) 试验1-2 (j) 试验2-2 (k) 试验3-2 (l) 试验4-2

图 6 消融研究的视觉示例

Fig.6 Visual examples of ablation study

表 3 框架组件的消融研究

Table 3 Ablation study of the framework components

方案	掩码比例/%	R_{PSN}/dB	S_{SIM}	方案	掩码比例/%	R_{PSN}/dB	S_{SIM}
试验 1	[1,10)	**39.24**	**0.987 9**	试验 2	[1,10)	37.78	0.977 4
	[10,20)	**37.12**	**0.980 6**		[10,20)	36.16	0.970 8
	[20,30)	**34.40**	**0.967 1**		[20,30)	33.87	0.958 0
	[30,40)	29.97	**0.939 1**		[30,40)	29.79	0.931 0
	[40,50)	27.13	**0.899 0**		[40,50)	26.99	0.891 9
	[50,60)	26.00	**0.874 2**		[50,60)	25.90	0.868 2

表 3(续)

方案	掩码比例/%	R_{PSN}/dB	S_{SIM}	方案	掩码比例/%	R_{PSN}/dB	S_{SIM}
试验 3	[1,10)	38.35	0.986 2	试验 4	[1,10)	39.19	0.987 7
	[10,20)	36.45	0.978 0		[10,20)	37.07	0.980 0
	[20,30)	33.84	0.963 5		[20,30)	34.33	0.966 2
	[30,40)	29.72	0.935 3		[30,40)	**30.03**	0.938 4
	[40,50)	26.99	0.894 4		[40,50)	**27.21**	0.898 3
	[50,60)	25.86	0.868 7		[50,60)	**26.07**	0.873 6

3.4　模型实测应用

为验证各模块设计的有效性,本研究对真实场景下的人脸图像进行实测,如图 7 所示。由图 7 可以看出,本研究所提模型能灵活处理不规则缺失的人脸图像,实现良好的修复性能。

(a) 原图1　(b) 残损图1　(c) 修复图1

(d) 原图2　(e) 残损图2　(f) 修复图2

图 7　实测样例

Fig.7　Actual measurement sample

4　结论

为了增强生成图像的全局一致性,本研究提出一种新颖的用于人脸图像修复的双流网络结构,通过在生成器中引入改进的归一化注意力模块,有效提取人脸图像的深层纹理及结构特征;设计一种基于残差主路径转换的双级门控特征融合模块,进一步细化深层特征,优化生成图像的一致性。此外,本研究还提出一种基于生成提示的上下文特征聚合模块,生成更精细的纹理结构信息。试验结果表明,与其他主流方法比较,本研究方法在 CelebA-HQ 和 LFW 数据集上能有效修复残损人脸面部细节信息。然而,当面对具有大面积的不规则残损图像时,本研究修复后的人脸图像与原始图像仍存在一定的差异,后续研究将改进鉴别器和损失函数,以更好地监督生成器在大范围缺损情况下能生成更逼真的人脸图像。

参考文献:

[1] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing[J]. ACM Transactions on Graphics, 2009, 28(3): 1-11.

[2] EFROS A A, FREEMAN W T. Image quilting for texture synthesis and transfer[C]//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA:ACM, 2001: 341-346.

[3] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 4470-4479.

[4] YAN Z, LI X, LI M, et al. Shift-Net: image inpainting via deep feature rearrangement[C]//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018: 3-19.

[5] LIU H, JIANG B, XIAO Y, et al. Coherent semantic attention for image inpainting[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 4170-4179.

[6] LI J, HE F, ZHANG L, et al. Progressive reconstruction of visual structure for image inpainting[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 5961-5970.

[7] WANG N, ZHANG Y, ZHANG L. Dynamic selection network for image inpainting[J]. IEEE Transactions on Image Processing, 2021, 30: 1784-1798.

[8] ZENG Y, FU J, CHAO H, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(7): 3266-3280.

[9] HUI S, ZHOU S, DENG Y, et al. Auxiliary loss reweighting for image inpainting[EB/OL]. (2022-04-22) [2024-04-20]. https://arxiv.org/abs/2111.07279

[10] MISRA D. Mish: a self regularized non-monotonic neural activation function[EB/OL]. (2019-08-26) [2024-04-20]. https://arxiv.org/abs/1908.08681

[11] LIU Y, SHAO Z, TENG Y, et al. NAM: normalization-based attention module[EB/OL]. (2021-11-24) [2024-04-20]. https://arxiv.org/abs/2111.12419

[12] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation [EB/OL]. (2018-02-26) [2024-04-20]. https://arxiv.org/abs/1710.10196

[13] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments[EB/OL]. (2008-09-16)[2024-04-20]. https://inria.hal.science/inria-00321923v1/document

[14] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting[C]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New Orleans, USA: ACM, 2000: 417-424.

[15] TSCHUMPERLÉ D, DERICHE R. Vector-valued image regularization with PDEs: a common framework for different applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(4): 506-517.

[16] DARABI S, SHECHTMAN E, BARNES C, et al. Image melding: combining inconsistent images using patch-based synthesis[J]. ACM Transactions on Graphics, 2012, 31(4): 1-10.

[17] BUYSSENS P, DAISY M, TSCHUMPERLÉ D, et al. Exemplar-based inpainting: technical review and new heuristics for better geometric reconstructions[J]. IEEE Transactions on Image Processing, 2015, 24(6): 1809-1824.

[18] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: Curran Associates, 2014: 2672-2680.

[19] NAZERI K, NG E, JOSEPH T, et al. EdgeConnect: generative image inpainting with adversarial edge learning [EB/OL]. (2019-01-11) [2024-04-20]. https://arxiv.org/abs/1901.00212

[20] XIONG W, YU J, LIN Z, et al. Foreground-aware image inpainting[C]//Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 5833-5841.

[21] REN Y, YU X, ZHANG R, et al. StructureFlow: image inpainting via structure-aware appearance flow[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 181-190.

[22] ZENG Y H, FU J L, CHAO H Y, et al. Learning pyramid-context encoder network for high-quality image inpainting[C]//Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 1486-1494.

[23] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 5505-5514.

[24] ZHENG C X, CHAM T J, CAI J F. Pluralistic image completion[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 1438-1447.

[25] QIN J, BAI H, ZHAO Y. Multi-scale attention network for image inpainting[J]. Computer Vision and Image Understanding, 2021, 204: 103155.

[26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[27] GUO X, YANG H, HUANG D. Image inpainting via conditional texture and structure dual generation[C]//Proceedings of the 18th IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 14114-14123.

[28] POTLAPALLI V, ZAMIR S W, KHAN S, et al. PromptIR: prompting for all-in-one blind image restoration[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates, 2023: 71275-71293.

[29] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015: 234-241.

（下转第39页）