

文章编号: 1671-7449(2024)03-0274-07

基于纹理和结构的双流人脸图像修复算法

王贺, 宋宏旭, 张震

(山西大学物理工程学院, 山西太原 030006)

摘要: 图像修复作为图像处理领域重要的研究方向而备受关注, 现有的图像修复模型只针对纹理或结构的某一方面, 忽略了二者在图像修复任务中是相辅相成的, 从而导致所修复出的图像不尽如意。鉴于此, 采用双支流的编解码器作为图像生成器的主干, 分别对应生成图像的纹理特征和结构特征, 达到结构约束纹理, 纹理引导结构的效果; 利用双向残差特征融合模块融合解码器生成的纹理特征和结构特征, 完成两种特征的信息交换; 并用多尺度上下文特征信息聚合模块丰富修复图像的细节特征。实验证明, 该方法在不同掩码率下 SSIM 及 PSNR 值均有提升。

关键词: 图像修复; 深度学习; 纹理和结构特征; 双向残差特征融合; 上下文特征信息聚合

中图分类号: TP391.41; TP18 **文献标识码:** A **doi:** 10.3969/j.issn.1671-7449.2024036

引用格式: 王贺, 宋宏旭, 张震. 基于纹理和结构的双流人脸图像修复算法[J]. 测试技术学报, 2024, 38(3): 274-280.

WANG He, SONG Hongxu, ZHANG Zhen. Two-stream face image restoration algorithm based on texture and structure[J]. Journal of Test and Measurement Technology, 2024, 38(3): 274-280.

Two-Stream Face Image Restoration Algorithm Based on Texture and Structure

WANG He, SONG Hongxu, ZHANG Zhen

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: As an important research direction in the field of image processing, image restoration has attracted much attention. The existing image restoration models only focus on one aspect of texture or structure, but ignore that the two complement each other in image restoration tasks, which leads to unsatisfactory images. In this paper, the codecs of double tributaries are used as the backbone of the image generator, corresponding to the texture features and structure features of the generated image respectively, so as to achieve the effect of structure constraint texture and texture guidance structure. Secondly, The bidirectional residual feature fusion module is used to fuse texture features and structure features generated by the decoder to complete the information exchange of the two features. Finally, The detailed features of the restored images are enriched with a multi-scale context feature information aggregation module. Experiments show that the proposed method can improve the SSIM and PSNR values under different mask rates.

Key words: image restoration; deep learning; texture and structural features; bidirectional residual feature fusion; contextual feature information aggregation

收稿日期: 2023-06-26

作者简介: 王贺(1983-), 女, 博士, 讲师, 主要从事机器学习和信息处理研究。E-mail: wanghe@sxu.edu.cn.

0 引言

图像修复是数字图像处理领域中非常重要的一个分支,当前,图像修复主要任务是通过使用计算机算法经过计算和学习,将损坏、模糊的图像数据恢复到图像最初的样子,甚至还可以更好^[1]。相比于图像处理领域的图像分类^[2]、目标检测^[3]、图像识别^[4]、图像去雾^[5]等方向,图像修复的任务不仅需要从未缺失的图像区域中提取图像特征,还需要对相应的缺失区域进行特征预测,以实现与未缺失区域相匹配的图像特征,从而形成一张完整的图像。进入 21 世纪后,随着智能手机、个人电脑等各种电子产品的普及,拍照已经成为一种时尚,当海量照片进行传输的过程中,图片损坏已不可避免,此时图像修复就显得尤为重要。

2016 年,Pathak 等^[6]首次提出将编码器解码器的深度学习方法应用于图像修复中,自此基于深度学习的图像修复成为主流。GAO 等^[7]提出在网络中加入反馈机制,在网络的训练阶段让网络进行自动优化。为了改善图像修复结果存在局部和全局不一致的问题,Iizuka 等^[8]设计了全局鉴别器和局部鉴别器,这样既能保证生成的图像符合全局语义,又可以提高局部图像和全局图像的一致性。以上网络模型只能处理缺失块位于图像中心附近的矩形区域,图像修复的质量有待提高。Liu 等^[9]设计出的部分卷积(Partial Convolution)可以很好地处理任意形状、大小、位置或距离图像边界任何距离的缺失块;Yu 等^[10]率先在网络中加入了注意力机制,改变了普通卷积处理图像特征

时一视同仁的问题,加入注意力机制后能够对重要的特征进行重点处理,改善了修复图像模糊的问题;Liu 等^[11]提出基于纹理和结构的图像修复,浅层特征用于纹理,深层特征用于结构,改善了图像修复结果纹理和结构不一致的问题。

受阶段性修复图像方法^[12]的启发,本文采用一种双流网络进行图像修复,该网络模型采用两个编码器和解码器,以耦合的方式进行对接,分别利用纹理编码-解码器产生纹理特征,结构编码-解码器产生结构特征,达到结构特征约束纹理特征生成和纹理特征引导结构特征生成的效果。其次,在纹理特征和结构特征融合过程中加入双向残差特征融合模块和基于注意力的多尺度上下文特征信息聚合模块,进一步提升纹理特征和结构特征的融合效果和图像分辨率。

1 系统框架

为了使重建的人脸图像同时具有纹理和结构的语义,改善缺失区域与已知图像区域不一致的问题,本文采用双支流编码-解码器结构作为图像生成器的主干,分别构造生成纹理和生成结构的编码-解码器,形成并行且耦合的双支流图像生成器网络模型,这种组合方式既能够使纹理特征图和结构特征图单独建模,又能使二者相互补充,特征信息相互借鉴;该模型还运用了纹理和结构的双鉴别器来检验该模型的性能,以便于更好地进行模型优化。

如图 1 所示,整个网络模型有三部分组成,第一部分是特征生成器,第二部分是特征融合部分,第三部分是鉴别器部分。

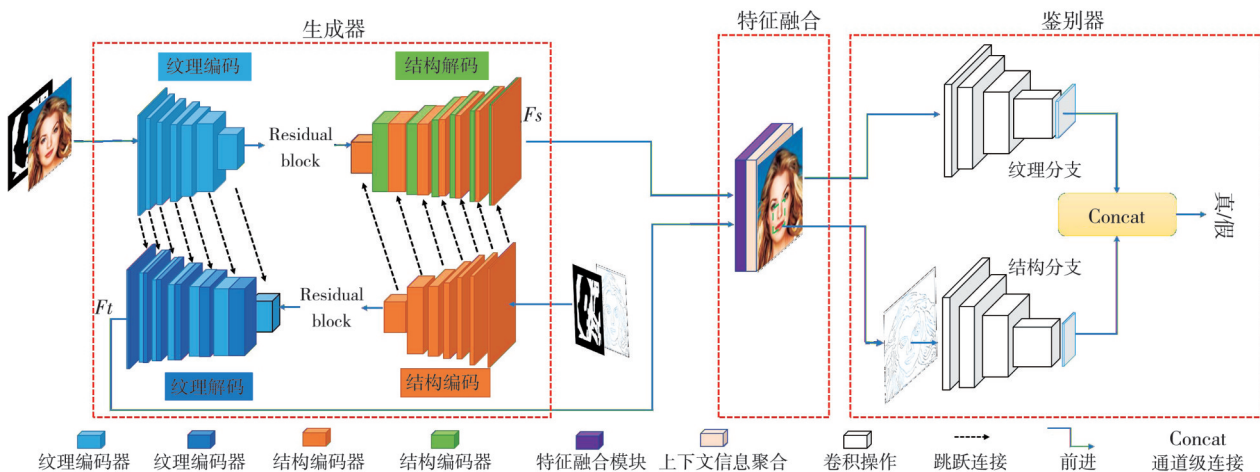


图 1 图像修复模型的总体框架

Fig. 1 General framework of the image restoration model

1) 生成器部分用来产生纹理特征和结构特征,所有编码器和解码器都含有7个卷积层,并采用部分卷积代替普通卷积,其中纹理编码器把真实图像和掩码图像像素相乘得到的受损图像作为输入,结构编码器把图像的边缘图和掩码图像像素相乘得到受损边缘图作为输入,同时结构编码器和纹理编码器每层生成的特征通过跳跃连接与解码器的对应层特征进行拼接,而后纹理编码器和结构解码器、结构编码器和纹理解码器构成新的编码-解码器结构,两条路径得到的纹理解码与结构解码分别产生了纹理特征 F_t 和结构特征 F_s ,在瓶颈处加入 Residual Block^[13]是为了在不影响网络性能的同时增加网络的深度,提取图像更多的深层特征信息; 2) 图像特征融合部分采用双向残差特征融合模块和基于注意力的多尺度上下文特征信息聚合模块,进一步提升纹理特征和结构特征的一致性,同时提升修复图像像素的能力; 3) 鉴别器针对其修复图像的纹理和结构进行鉴别,各包括5个卷积层,前3个卷积层的卷积核大小为4、步长为2,后2个卷积层的卷积核大小为 4×4 、步长为1,最后一层使用 Sigmoid 非线性激活函数,其他层使用斜率为0.2的 Leaky ReLU。结构分支与纹理分支的鉴别器结构相同,但其输入为修复图像的边缘图。最后,将两个分支的输出通

道维度上进行连接,计算出鉴别器联合损失,通过鉴别器输出的结果进行反向传播不断优化参数,直至达到最优的修复效果。

1.1 双向残差特征融合模块 (Bi-RFFA)

该模块用于融合解码器生成的纹理特征和结构特征,使得两种特征之间交换信息,利用控制门来控制特征信息交互的程度,通过这种融合使图像同时具有纹理和结构的语义。

图2所示为特征融合模块结构图。 F_t 为生成器生成的纹理图, F_s 为生成器生成的结构图; G_t 为控制纹理信息整合程度的控制门, G_s 为控制结构信息整合程度的控制门。该模块分2个支路:对于上支路而言, F_t 和 F_s 经过通道间连接操作(Concat)后连接在一起,随之经过C2F-ResSE模块将其输入到后边的线性映射 g ,为了得到控制特征信息交互时 F_t 的相关程度,将 g 的输出经过sigmoid激活函数转换成概率图,该概率图即为控制门 G_t ,并将其与 F_t 相乘,以确定 F_t 中各特征信息的保留程度,最后将保留下来的信息与 F_s 元素相加得到纹理感知结构特征 $F's$;作为对称,下支路也用类似的方式得到结构感知纹理特征 $F't$,最后通过通道级连接操作(Concat)将 $F's$ 和 $F't$ 连接得到Bi-RFFA模块的最终输出。

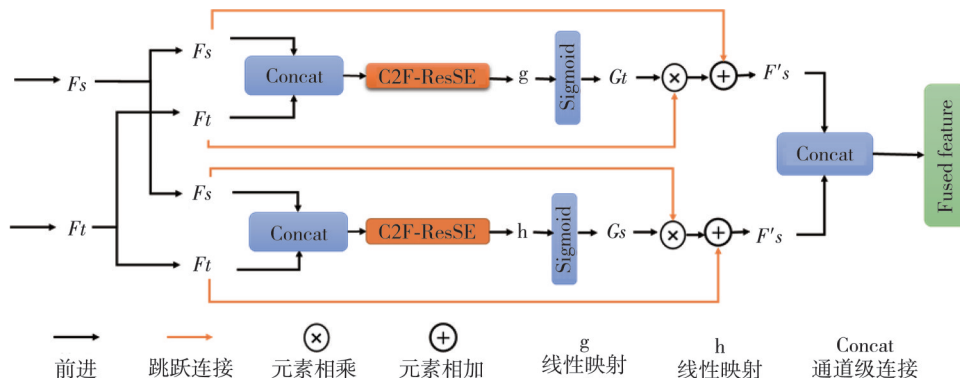


图2 特征融合模块
Fig. 2 Feature fusion block

图3所示为C2F-ResSE结构图,该模块充分结合了C2F和ResSE Block的优点。

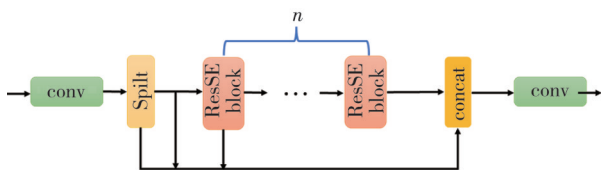


图3 C2F-ResSE模块
Fig. 3 C2F-ResSE block

C2F^[14]是发掘于目标检测YOLOV8的一种算法,它能够使模型在运行过程中保证轻量化的同时获得更加丰富的梯度流信息,在其中加入ResSE模块^[15],使其形成C2F-ResSE模块。主要过程为特征图先经过一个卷积核大小为 1×1 、步长为1的卷积之后,再对其进行分片操作(Split),这个操作主要是针对通道进行分片,而后再经过 n 个ResSE模块,同时把每个ResSE模块的输出

特征与分片特征经过通道间连接操作(Concat), 最后再经过大小为 1×1 、步长为1的卷积处理。

1.2 基于注意力的多尺度上下文特征信息聚合模块(MSCFA)

此模块采用多尺度特征信息聚合的方式^[16]丰富语义特征, 能够很好地平衡特征的准确性和复杂性。为了使图像修复网络不再对图像的通道特征和像素特征一视同仁, 加入FA 模块(Feature Attention)组成了基于多尺度的上下文特征信息聚合模块(MSCFA), 从而提升特征信息的聚合能力。

如图4所示, 经过特征融合模块生成的图像先

经过卷积操作生成特征映射 F , 再经过提取补丁(patch)将特征映射以 3×3 为单位划分成不重叠的补丁, 并计算出这些补丁之间的相似程度, 利用softmax函数得出每个补丁的相关性得分图(F' contextual)。将所有的补丁在得分图的关注下重组后获得重构的特征图(F_{rec}), 然后采用不同膨胀率的膨胀卷积进行卷积操作, 膨胀率分别为1、2、4、8, 得到不同尺度的特征图, 这样能够充分利用图像特征信息, 通过像素相加的方式把不同尺度特征图整合到一起得到精细化后的特征图 F_c , 再把 F_c 和 F 经过通道级连接操作(Concat)后输入FA 模块进行处理, 最后通过反卷积得到修复图像。

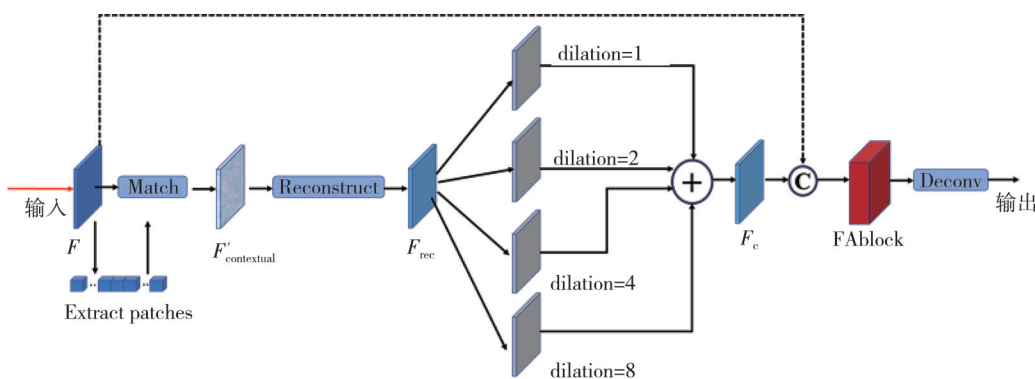


图4 MSCFA的结构图

Fig. 4 Structure of MSCFA

如图5所示, FA 模块由通道注意力^[17]和像素注意力^[18]通过串联的方式连接后形成, 图像特征

先经过通道注意力得到对于通道维度的关注后, 再经过像素注意力后能使得到的图像像素更高。

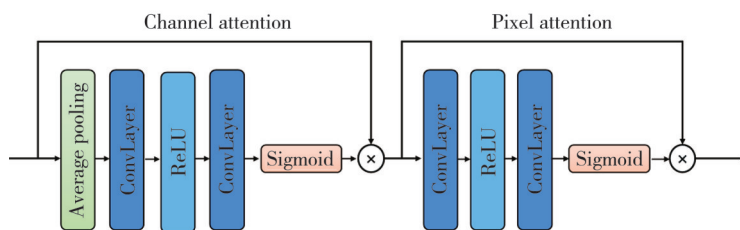


图5 FA 模块

Fig. 5 FA block

2 实验结果与分析

2.1 实验准备

实验配置为: intel i7-11700CPU, 64 位 Windows 10 操作系统, Nvidia GeForce RTX 3060Ti, 内存大小为 16 G, 实验框架为 Pytorch 深度学习框架, PyTorch1.10.2, CUDA11.3。

本文使用的数据集为公开的人脸图像数据集 CelebA-HQ, 该数据集中包含了 3 万张高清人脸

图像, 涵盖了不同的肤色、性别、年龄等多种样式的人脸图像。本文把 CelebA-HQ 数据集按 14:1 的比例划分训练集和测试集, 其中训练集有 28 000 张, 测试集有 2 000 张; 随机掩码数据集取自 NVIDIA 不规则掩码数据集, 总共 12 000 张不同程度缺失的随机形状掩码, 按照(0, 10%], (10%, 20%], (20%, 30%], (30%, 40%], (40%, 50%], (50%, 60%]的缺失比率划分为 6 个等级, 每个等级的掩码数量包含 2 000 张; 使用 Adam 算法优化模型, 训练学习率恒定设置为 0.000 2,

batchsize 设置为 2, 训练迭代次数为 30 万次, 图片输入尺寸为 256×256 。

2.2 评估标准

本文使用峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 和结构相似性 (Structural Similarity, SSIM) 对人脸图像修复结果进行评判。SSIM 衡量的是两幅图像在光线、对比度和结构信息上的相似性; PSNR 是反映峰值功率与噪声功率比值的失真度的指标。PSNR 和 SSIM 越大, 表明修复模型产生的修复图像失真越小, 与原图像相似度更高, 即图像修复的效果更好。两个指标的表达式分别为式(1)和式(3)。

$$PSNR(x, y) = 10 \lg \left(\frac{MAX_x^2}{MSE} \right), \quad (1)$$

式中: MAX_x 为修复图像中最大的灰度值, 此方法中取值为 256, MSE 为修复图像和真实图像的均方误差, 表达式为式(2)。

$$MSE(x, y) = \frac{1}{H \times W} \sum_{i=0}^H \sum_{j=0}^W (x(i, j) - y(i, j))^2, \quad (2)$$

式中: x 为修复结果图; y 为真实图; H , W 分别为图像的高度和宽度。

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (3)$$

式中: x 为修复结果图; y 为真实图; μ_x 和 μ_y 分别为修复结果图和真实图的均值; σ_x^2 和 σ_y^2 分别为二者的方差; σ_{xy} 为协方差; C_1 , C_2 为常数。

2.3 主观评价

图 6 所示为不同算法修复模型在掩码率为 (50%, 60%) 的修复结果对比, 其中 DeepFill2^[19] 修复模型在此掩码率下的修复结果失真较大, 在脸颊和嘴部特征方面出现了不同程度的失真, 同时修复图像较模糊; Pconv^[11] 作为首个能够处理随机形状掩码的修复模型, 在处理随机形状掩码时相较之前的修复模型有较大提升, 但仍有修复局限性, 突出表现在修复区域与未缺失图像区域不一致方面; Shift-Net^[20] 对细节特征如眼部和嘴部特征的修复较明显, 但在图像的整体色彩和下巴部分的修复欠佳; CSTDG^[21] 修复人脸时在细节和内容方面的处理相对比较细腻, 但是仍有部分区域的细节特征有待改善, 尤其是在眼部特征的处理方面; 本文改进模型在纹理特征和结构特征的

处理方面比较突出, 基本还原了真实图像的全部特征, 相比其他几种修复模型不论在细节特征还是轮廓特征方面都处理较好。



图 6 不同算法修复模型结果对比

Fig. 6 Results comparison of different algorithms for repairing the model

图 7 为本文模型在 6 种不同掩码率下且对不同姿势和不同肤色的人脸图像的修复结果。

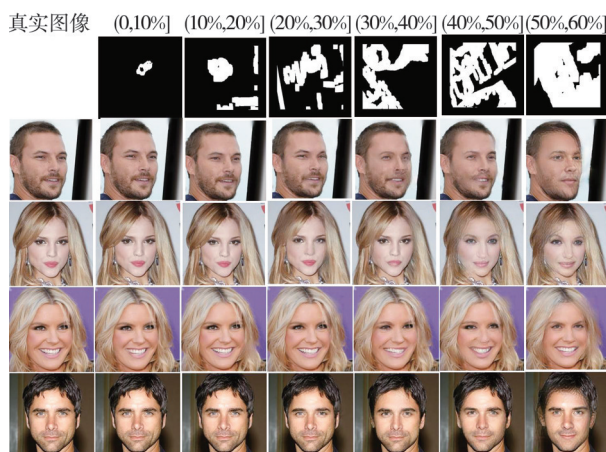


图 7 本文改进模型在不同掩码率下人脸修复结果

Fig. 7 Face restoration results of the improved model in this paper under different mask rates

从图中可以看出, 本文模型对正面人脸图像的修复效果较好, 尤其是在掩码率为 (50%, 60%) 的情况下, 相比侧面人脸的修复效果好, 不但很好地还原了真实图像, 而且修复区域和已知区域之间的衔接比较自然, 视觉上能够达到很好的要求。随着掩码率的增大, 侧面人脸的修复图像会出现不同程度的模糊和失真, 但依然可以清楚地看出人脸的轮廓和一些细节信息。

2.4 客观评价

为了验证实验的准确性, 各对比模型均使用人脸数据集 CelebA-HQ 和 NVIDIA 掩码数据集,

同时采用相同的训练集、测试集以及相同的实验设备,对比实验分别就 DeepFill2, PConv, Shift-

Net, CTSDG 4 种模型在不同掩码率下的 PSNR 和 SSIM 进行计算,结果如表 1 所示。

表 1 各个模型评价指标对比

Tab. 1 Comparison of evaluation indicators of each model

掩码率/%	PSNR/dB					SSIM/%				
	DeepFill2	Shift-Net	PConv	CTSDG	本文改进	DeepFill2	Shift-Net	PConv	CTSDG	本文改进
(0,10]	33.43	36.21	38.42	39.27	40.78	95.97	96.34	97.26	98.5	98.86
(10,20]	25.08	30.75	32.54	33.05	34.26	90.19	91.96	93.47	95.75	96.61
(20,30]	22.23	27.46	28.99	29.57	30.60	83.74	84.25	86.24	92.04	93.46
(30,40]	20.24	24.50	26.64	27.03	27.90	77.68	77.29	81.63	87.75	89.64
(40,50]	18.67	21.78	24.7	25.04	25.79	71.66	70.75	75.45	82.92	83.24
(50,60]	17.14	18.02	21.85	21.97	22.47	63.32	64.65	70.47	74.86	76.48

从表 1 中看出,在掩码率为(0,10%]时,本文的修复模型 SSIM 指标上达到 98.86%,PSNR 达到了 40.78 dB,很大程度上还原了真实图像的样子,相比 4 种模型中最优的 CTSDG 在 SSIM 上有 0.36% 的提升,在 PSNR 上有 1.51 dB 的提升;在掩码率为(50%,60%]时,本文所提修复模型在 SSIM 上达到了 75.48%,PSNR 达到 22.47 dB,同样在 SSIM 比 CTSDG 提升了 1.62%,PSNR 提升了 0.5 dB。总体上,该模型在各个掩码率下提高了还原真实图像的力度,同时相比其他几种算法效果都有较大的提升。

2.5 消融实验

在消融实验中进行了 4 组对比实验:1)不使用 C2F-ResSE 和 MSCFA 模块;2)单独使用 C2F-ResSE 模块;3)单独使用 MSCFA 模块;4)C2F-ResSE 和 MSCFA 模块均使用,在(0,60%]的平均掩码率下得到表 2 所示结果。

表 2 消融实验各模块对比表

Tab. 2 Comparison of the modules of the ablation experiment

C2F-ResSE	MSCFA	PSNR/dB	SSIM/%
×	×	28.32	87.63
×	√	29.65	89.06
√	×	29.37	88.51
√	√	30.04	90.12

从表 2 中可以看出,缺少任何一个模块后 PSNR 和 SSIM 都会有所下降,单独使用 MSCFA 模块比两模块均未使用时,PSNR 平均提升 1.33 dB,SSIM 平均提升 1.61%;单独使用 C2F-ResSE 模块时,PSNR 平均提升 1.05 dB,SSIM 提升 1.35%;同时使用这 2 个模块时,PSNR 平均提升 1.72 dB 左右,SSIM 平均提升 2.49%。

3 结 论

本文从图像的纹理和结构出发,利用双编码器解码器结构进行图像修复,充分利用了图像的纹理和结构特征信息,通过加入特征融合模块和上下文特征信息聚合模块,对人脸图像数据集 CelebA-HQ 在(0,60%]掩码率下进行修复,改善了以往修复图像出现的模糊现象,使得修复出的人脸图像不论从主观还是客观上都有一定程度的改善,得到的修复图像像素有所提高,视觉效果较好,在峰值信噪比(PSNR)和结构相似性(SSIM)上相比其他几种人脸修复算法也都有所提升。

参考文献:

[1] 朱新山, 钱永军, 孙彪, 等. 基于深度神经网络的图像修复取证算法[J]. 光学学报, 2018, 38(11): 105-113.
 ZHU Xinshan, QIAN Yongjun, SUN Biao, et al. Image inpainting forensics algorithm based on deep neural network[J]. Acta Optica Sinica, 2018, 38(11): 105-113. (in Chinese)

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

[3] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.

[4] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[5] DONG H, PAN J, XIANG L, et al. Multi-scale

- boosted dehazing network with dense feature fusion [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2154-2164.
- [6] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2536-2544.
- [7] GAO R, GRAUMAN K. On-demand learning for deep image restoration [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 1095-1104.
- [8] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion [J]. ACM Transactions on Graphics, 2017, 36(4): 1-14.
- [9] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 89-105.
- [10] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 5505-5514.
- [11] LIU H, JIANG B, SONG Y, et al. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations [C]//16th European Conference on Computer Vision-ECCV, 2020: 725-741.
- [12] 芮志超, 郭艳艳. 基于上下文特征提取的边缘生成三阶段图像修复算法 [J]. 测试技术学报, 2024, 38(1): 34-40.
RUI Zhichao, GUO Yanyan. Three-stage image inpainting algorithm by edge generation based on contextual feature extraction [J]. Journal of Test and Measurement Technology, 2024, 38(1): 34-40. (in Chinese)
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] SUN Y, CHEN G, ZHOU T, et al. Context-aware cross-level fusion network for camouflaged object detection [C]//IJCAI International Joint Conference on Artificial Intelligence, 2021: 1025-1031.
- [15] ZHANG W, WANG Y, NI B, et al. Fully context-aware image inpainting with a learned semantic pyramid [J]. Pattern Recognition, 2023, 143: 109741.
- [16] YI Z L, TANG, Q, AZIZI S, et al. Contextual residual aggregation for ultra high-resolution image inpainting [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7505-7514.
- [17] XU K, BA J, KIROUS R, et al. Show, attend and tell: neural image caption generation with visual attention [C]//International Conference on Machine Learning, PMLR, 2015: 2048-2057.
- [18] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 294-310.
- [19] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 4470-4479.
- [20] YAN Z, LI X, LI M, et al. Shift-net: image inpainting via deep feature rearrangement [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [21] GUO X, YANG H, HUANG D. Image inpainting via conditional texture and structure dual generation [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 14114-14123.