DOI:10.11918/202010013

风格感知和多尺度注意力的人脸图像修复

刘洪瑞[1,2],李硕士[1],朱新山[1,2],孙 浩[1],张 军[1]

(1.天津大学 电气自动化与信息工程学院,天津 300072;2.数字出版技术国家重点实验室,北京 100871)

摘 要: 人脸图像修复是计算机视觉领域中重建人脸图像的一项重要图像处理技术。现有人脸图像修复技术存在修复结果全局语义不合理的问题,这主要是由于现有技术的特征长程迁移能力不足,无法将破损图像中已知区域的信息合理地迁移到被遮蔽区域上。为此,本文在生成式对抗网络(generative adversarial network,GAN)框架下,构建了一种融合风格感知和多尺度注意力的编解码人脸图像修复模型。风格感知模块用于提取图像的全局语义信息,并利用提取的信息对编码逐级地进行渲染,以实现对修复过程的全局性调节;利用多尺度注意力模块对多尺度特征进行补丁块提取,并通过共享注意力得分和提取补丁块的矩阵乘法进行多尺度特征的长程迁移。在公开数据集 CelebA-HQ 上的实验结果表明:风格感知模块和多尺度注意力模块极大地增强了修复网络的特征长程迁移能力。相较于现有先进的人脸图像修复方案,本文所提出的模型在多种评价指标上均有显著的提升;修复结果的全局语义更加合理,并且在暗光条件下的修复效果更加自然。

关键词: 人脸图像修复;生成对抗网络;风格感知;多尺度注意力;长程迁移

中图分类号: TN911.73 **文献标志码:** A **文章编号:** 0367-6234(2022)05-0049-08

Style-aware and multi-scale attention for face image completion

LIU Hongrui[1,2], LI Shuoshi[1], ZHU Xinshan[1,2], SUN Hao[1], ZHANG Jun[1]

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;
2. State Key Laboratory of Digital Publishing Technology, Beijing 100871, China)

Abstract: Face image completion is an important image processing technique for reconstructing face images in the field of computer vision. The existing face image completion methods have the problem of unreasonable global semantics, which is mainly due to the lack of long-range transfer capability of the existing techniques that they are unable to reasonably transfer information from known regions in a broken image to occluded regions. To overcome the problem, a novel encoder-decoder face image completion network integrating style-aware and multi-scale attention was proposed under the framework of generative adversarial network (GAN). Specifically, the style-aware module was used to extract the global semantic information of an image, and the extracted information was employed to globally adjust the completion processing by rendering the encoding of the image level by level. The multi-scale attention module extracted patches of multi-scale features and performed a long-range transfer via matrix multiplication between a shared attention score and the extracted patches. Experimental results from the public dataset CelebA-HQ show that the style-aware module and the multi-scale attention module greatly enhanced the long-range transfer capability of the completion network. Compared with the existing state-of-the-art face image completion methods, the proposed model had significant improvement in various evaluation metrics. Meanwhile, the global semantics of the completion results were more reasonable and the completion effect was more natural under low lighting conditions.

Keywords: face image completion; generative adversarial network (GAN); style-aware; multi-scale attention; long-range transfer

人脸图像修复因其在电影工业、刑事侦破等方面的广泛应用,成为了计算机视觉领域的一个研究热点。传统的图像修复方法主要分为基于扩散和基于补丁块两类。第一类基于扩散的方法如文献[1-2]是将被遮挡区域周围的低维特征以迭代的方式传播进遮挡区域,第二类基于补丁块的方法如文献[3-4]则是在同一张或者多张图像中搜索相似的补丁块的方式来修复目标区域。这两类修复方法都只适用于填充纹理相似的背景图像,但人脸图像面部成分之间存在紧密的联系,其修复结果应具备全局语义的合理性,比如左右眼对称、肤色一致等。因此,基于扩散和基于补丁块的方法都不适用

收稿日期: 2020-10-08
基金项目: 国家自然科学基金(61972282,61971303);CCF 信息系统开放课题(CCFIS2018G02G04);北大方正集团有限公司数字出版技术国家重点实验室开放课题(Cndplab-2019-Z001)
作者简介: 刘洪瑞(1996—),男,硕士研究生
通信作者: 朱新山,xszhu@tju.edu.cn

于人脸图像修复。

随着生成对抗网络[5]的发展,人脸图像修复技术取得了显著的进步。现有基于生成对抗网络的人脸图像修复方案主要分为 4 类:基于全连接层的方案、基于大尺度空洞卷积的方案、基于文献[6]提出的 U-net 结构的方案、基于注意力模块的方案。文献[7]提出一种上下文编码器,其使用全连接层来完成面部特征的长程迁移,但是由于这类方案无法有效地利用局部信息,修复效果会出现模糊及局部语义不合理的现象。针对这一问题,文献[8]考虑到大尺度空洞卷积具有较强的信息扩散能力,将其用于特征长程迁移以完成人脸修复任务。在此基础上,文献[9]和文献[10]对空洞卷积进行了改进,分别提出了局部卷积以及门控卷积,为修复网络所有层中每个空间位置的每个通道提供了一个可学习的动态特征选择机制,提高了网络对掩模形状的适应性。然而,虽然空洞卷积理论上的感受野很大,但是其更加关注局部信息,这往往会导致修复结果出现左右眼不对称等全局语义不合理的问题。文献[11]考虑到小尺度特征具有更大感受野的优点,借鉴 U-net 的多尺度网络架构,通过对小尺度特征进行多次卷积实现特征的长程迁移,然后在编码器和解码器间使用了多次跳跃连接恢复图像细节。但是,小尺度卷积本质上还是一种局部操作,修复结果同样会出现全局语义不合理的问题。此外,文献[12]和文献[13]分别提出了内容注意力模块或长短注意力模块,这两个模块具有优良的特征长程迁移能力。但是,这两个模块在大尺度特征上使用时会导致显存占用过多,因此无法在多尺度特征上使用,修复结果仍然会出现全局语义不合理的问题。

综上所述,现有方案都存在修复结果全局语义不合理的问题,其本质原因在于它们对特征的长程迁移能力不足。为了解决这一问题,提出了基于风格感知和多尺度注意力的人脸图像修复网络。

1 基于风格感知和多尺度注意力的人脸图像修复方法

提出一种基于风格感知和多尺度注意力的人脸修复方法(style-aware and multi-scale attention for face image completion, SA-MA-FIC),由人脸拓扑结构预测器、人脸图像修复生成器和判别器 3 部分组成(见图 1)。针对修复图像全局语义不合理的问题,该网络在生成器中设计了风格感知模块对修复过程进行全局性地调节并设计了多尺度注意力模块用于多尺度特征的长程迁移。本节分为 4 个部分,分别详述模型整体设计、风格感知模块、多尺度注意力模块以及损失函数。

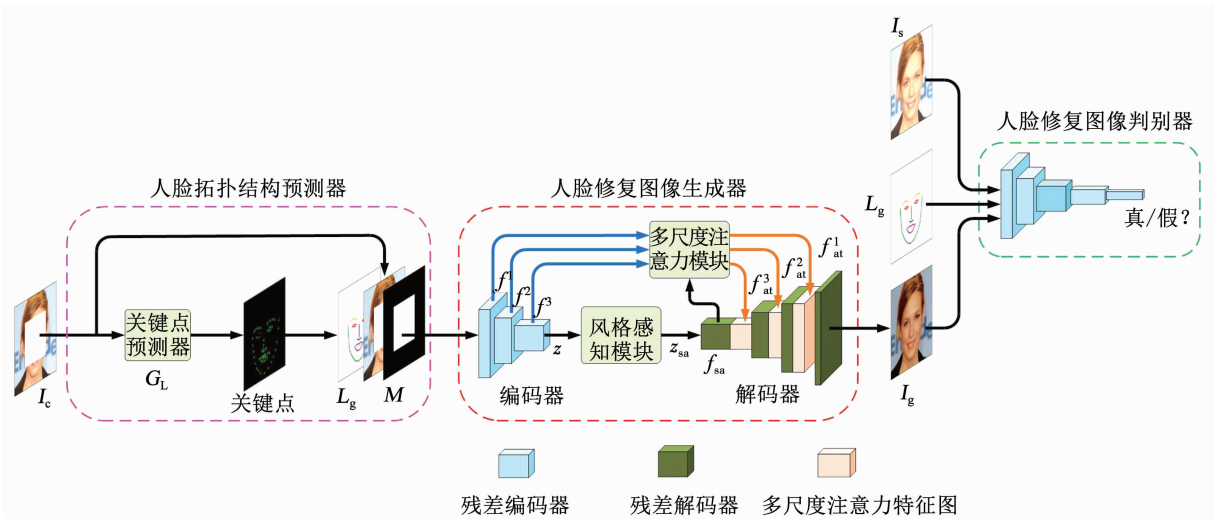

图 1 基于风格感知和多尺度注意力的人脸图像修复

Fig. 1 Style-aware and multi-scale attention for face image completion

1.1 模型整体设计

为了在人脸图像修复任务中取得良好的修复效果,本文在模型结构设计上,借鉴了文献[5]提出的生成对抗网络。该网络引入了文献[14 – 15]提出的纳什平衡建立问题模型,其结构见图 2。博弈双方分别为一个生成器 G 和一个判别器 D,生成器的目标是尽量去学习真实数据的分布,判别器的目标是尽量准确判别输入数据是真实数据 X 还是生成器生成的数据 $G(Z)$;为了达到各自的目标,生成器和判别器通过不断对抗训练来提升各自的生成能力和判别能力,这个训练过程就是寻找二者之间的纳什平衡。

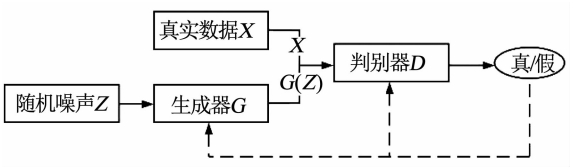

图 2　生成对抗网络结构

Fig. 2　Structure of the generative adversarial network

本文通过建立一个基于生成对抗网络的模型进行人脸图像修复，模型由人脸修复图像生成器 G 和判别器 D 两部分组成，见图 1。其中，生成器学习由破损图像 I_c 到生成图像 I_g 的映射，令 I_g 符合真实图像的数据分布；判别器则学习准确区分真、假样本的能力。

生成器的修复过程分为 3 步：第一步，通过编码器对输入信息进行编码，得到多尺度特征 $\{f^l \mid l=1,2,3\}$ 及编码 z；第二步，使用风格感知模块对编码 z 进行全局性调节得到风格编码 z_{sa}，在小尺度特征上完成全局语义的修复；第三步，依据 z_{sa} 的语义信息使用多尺度注意力模块，对多尺度特征进行长程迁移，合理恢复面部细节信息，最终将其与 z_{sa} 输入解码器进行解码，得到人脸修复图像 I_g。以上修复过程定义为

$$I_g = G(I_c, L_g) \tag{1}$$

判别器的真样本和假样本采用了 (L_g, I_g) 和 (L_g, I_s) 所组成的图像对，其中 L_g 是 I_g 的人脸拓扑结构图。该人脸结构拓扑图是由文献[16]提出的人脸关键点预测器 G_L 预测破损图像中 68 个人脸关键点，并对它们分别设置了颜色及连线得到的。不同关键点设置不同颜色可以帮助修复网络区分五官，使用连线取代单独的关键点可使用户轻松实现对人脸图像的编辑。通过这种真假样本对的设计，既可以有效地提高 I_g 的图像质量，又可以保证 I_g 符合 L_g 的拓扑结构，这会极大地提高训练的稳定性。

1.2　风格感知模块

针对人脸图像修复结果全局语义不合理的问题，本文设计了风格感知模块予以解决，该模块主要由风格渲染和风格提取两个并行的通道组成，见图 3。该模块利用风格提取通道提取输入信息的整体风格，并利用该风格在风格渲染通道中以文献[17]提出的自适应实例归一化(adaption instance normalization，AdaIN)的方式对输入编码 z 逐级地进行渲染，从而实现对修复过程的全局性调节。

在风格提取通道中，首先使用自注意力模和残差块对 256 × 32 × 32 的编码 z 进行重要信息的提取并压缩，得到维度为 512 × 4 × 4 的特征；然后将该特征通过全连接层映射为一个 512 × 1 × 1 的风格向量，这是一种全局性的操作，风格向量的每一个值均为 512 × 4 × 4 个特征值的加权求和。因此，通过上述操作获得的风格向量能够有效地反映全局语义信息。

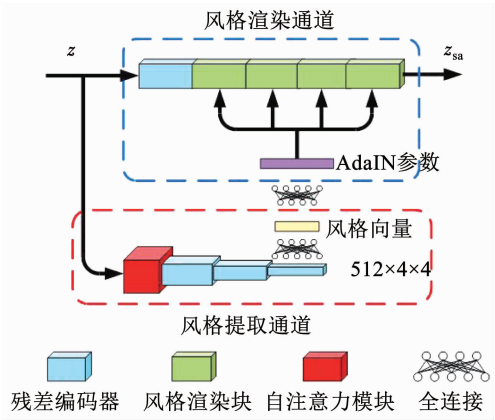

图 3　风格感知模块

Fig. 3　Style-aware module

在风格渲染通道中，为了实现对高维语义的调节，首先使用一个残差块将 256 × 32 × 32 的编码 z 进行下采样，得到维度为 512 × 16 × 16 的小尺度特征；然后利用多级风格渲染块对其进行风格渲染。风格渲染块的具体结构见图 4，其以 AdaIN 的方式进行两次渲染，渲染的过程定义为

$$f^*_{i,j} = \frac{f_{i,j} - \mu(f_{i,j})}{\sigma(f_{i,j})} \times \beta_{i,j} + \alpha_{i,j} \tag{2}$$

式中：$f_{i,j}$ 为输入特征，$\mu(f_{i,j})$ 和 $\sigma(f_{i,j})$ 分别为 $f_{i,j}$ 的均值和方差，$(\alpha_{i,j}, \beta_{i,j})$ 为由风格向量映射而来的 AdaIN 仿射参数，$f^*_{i,j}$ 为经过渲染后的特征，i 为第 i

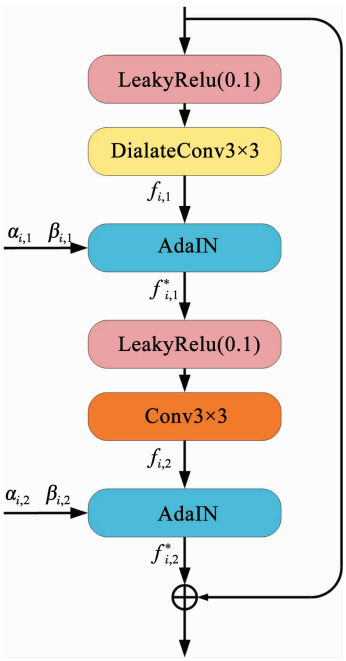

图 4　风格渲染块

Fig. 4　Style rendering block

个风格渲染块，j 为每个风格渲染块的第 j 次 AdaIN 操作。

为了提升渲染效果，在风格渲染块中使用空洞卷积来增大感受野。连续使用风格渲染块即可实现在不同感受野下的风格渲染。风格渲染块所需的 AdaIN 仿射参数是由风格提取通道所提取的风格向量，通过一个全连接层映射而来的，因此经其调节后的编码z_{sa}具备全局语义的合理性。

1.3 多尺度注意力模块

如上节所述，输入信息经过编码和风格感知模块的全局性调节后，得到了具备全局语义合理性的编码z_{sa}，但在这个过程中丢失了很多细节信息，直接用z_{sa}解码无法获取理想的修复图像。为此，可以通过将多尺度特征$\{f^l \mid l=1,2,3\}$输入解码器来恢复丢失的细节信息。但是，多尺度特征是由破损图像编码得到的，其对应破损区域位置的信息会有所缺失，直接使用多尺度特征又会再次导致修复结果全局语义的不合理。为此，参照文献[18]提出的自注意力模块对单尺度特征长程迁移的方案，本文设计了一种多尺度注意力模块，通过对多尺度特征进行长程迁移，获得合理的多尺度特征$\{f_{at}^l \mid l=1,2,3\}$，再利用其逐级恢复缺失区域的细节信息。

多尺度注意力模块由补丁提取和补丁加权求和两步组成。第一步，参照文献[12]对多尺度特征$\{f^l \mid l=1,2,3\}$进行补丁提取，见图 5(a)：首先分别以 4、2、1 的步长提取对应大小分别为 $4\times4\times c_1$、$2\times2\times c_2$、$1\times1\times c_3$的补丁块$\{p_i^l \mid l=1,2,3;i=1,2,\cdots,1\,024\}$；然后将提取到的补丁块逐块拼接起来，记作$\{\text{patches}^l \mid l=1,2,3\}$，维度分别为$(1\,024\times c_1)\times4\times4$、$(1\,024\times c_2)\times2\times2$、$(1\,024\times c_3)\times1\times1$，其中$c_l$为多尺度特征$f^l$的通道数。第二步，对由多尺度特征所提取补丁块进行加权求和得到$\{f_{at}^l \mid l=1,2,3\}$，见图 5(b)：首先对$z_{sa}$使用残差块上采样得到风格感知特征图$f_{sa}$；再进行$1\times1$卷积和 softmax$(Q^{\mathrm{T}}Q)$操作得到维度为 $1\,024\times32\times32$ 的自注意力得分 λ；然后分别使用patches^1、patches^2、patches^3和 λ 进行矩阵相乘，来实现对补丁块的加权求和，从而完成了对多尺度特征的长程迁移，其计算过程为

$$p_j^l = \sum_{i=1}^{N} \lambda_{j,i}\, p_i^l \tag{3}$$

式中：p_i^l是对编码器特征f^l所提取的第 i 个补丁，$\lambda_{j,i}$是不同补丁块的权重得分，p_j^l是f_{at}^l特征图中的第 j 个补丁。

值得注意的是，在多尺度特征图上直接使用自注意力模块对多尺度特征值加权求和，同样可以实现多尺度特征的长程迁移。然而，这种实现方式需要计算 3 个尺度逐渐膨胀的自注意力得分，维度分别为 $1\,024\times32\times32$、$4\,096\times64\times64$、$16\,384\times128\times128$。这将会导致庞大的显存占用和计算量，因此无法用于实际场景中。而本文所设计的多尺度注意力模块，仅需计算一个共用且维度仅为 $1\,024\times32\times32$ 的自注意力得分，就实现了对多尺度补丁块的加权求和。该模块不仅可以取得与在多个尺度分别使用自注意力模块同样的特征长程迁移效果，而且可以有效地减少显存的占用及计算量。

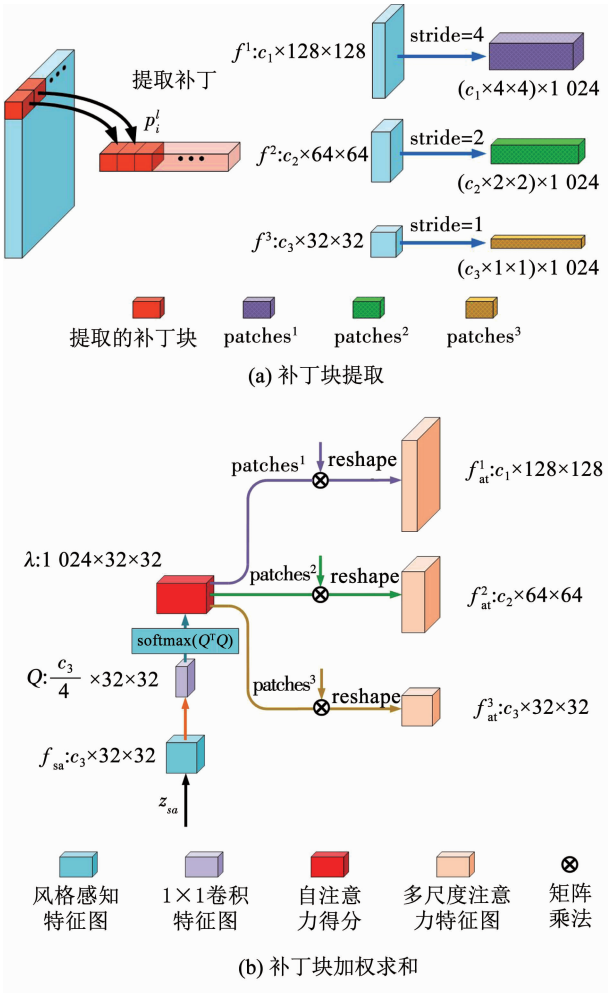

图 5 多尺度注意力模块

Fig. 5 Multi-scale attention module

1.4 损失函数

修复网络的目标是使修复图像I_g和原始图像I_s尽可能的相似。为此，本文引入重构损失，目的在于缩小I_g与I_s之间的像素级差异，表示为

$$L_{rec} = \| I_g - I_s \|_1 \tag{4}$$

式中 $\|\cdot\|_1$为L_1范数。

为提高修复结果的视觉质量，同时保证I_g与其所对应的L_g拓扑结构一致，引入了判别损失。该损失能够有效缩小修复图像I_g和真实图像I_s之间的数

据分布差异。判别损失参照文献[19]提出的LSGAN进行设计，对抗损失函数和判别模型的对抗损失函数分别见式(5)、(6)：

$$L_{\text{adv}}^{\text{G}} = \| D(I_{\text{g}}, L_{\text{g}}) - 1 \|_2 \tag{5}$$

$$L_{\text{adv}}^{\text{D}} = \| D(I_{\text{s}}, L_{\text{g}}) - 1 \|_2 + \| D(I_{\text{g}}, L_{\text{g}}) \|_2 \tag{6}$$

式中：$\| \cdot \|_2$是L_2范数，$(I_{\text{s}}, L_{\text{g}})$为真样本，$(I_{\text{g}}, L_{\text{g}})$为假样本，$D(I_{\text{s}}, L_{\text{g}})$、$D(I_{\text{g}}, L_{\text{g}})$分别为真样本、假样本的判别器输出。

此外，为了提高修复图像I_{g}与原始图像I_{s}之间的知觉相似性，本文引入了文献[20]提出的感知损失L_{pc}对在预训练网络不同层输出的特征距离加以惩罚。同时，为了去除修复图像中由于编码器上采样所导致的人工痕迹，引入了文献[21]提出的风格损失L_{style}对修复图像I_{g}和原始图像I_{s}在预训练网络不同层输出的协方差距离加以限制。

总的损失函数为

$$L_{\text{G}} = \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{adv}} L G_{\text{adv}} + \lambda_{\text{pc}} L_{\text{pc}} + \lambda_{\text{style}} L_{\text{style}} \tag{7}$$

式中λ_{rec}、λ_{adv}、λ_{pc}、λ_{style}为权重因子，具体取值见实验细节。

2　实验结果与分析

为验证 SA-MA-FIC 的优越性，设计了对比实验，将其与目前几种先进算法进行了定性和定量比较；另外，设计了消融实验，以验证 SA-MA-FIC 中风格感知模块和多尺度注意力模块的有效性。

2.1　数据集

本文在公开的人脸数据集 CelebA-HQ 上进行了训练和测试。该数据集共有 30 000 张图片，本文选取了 29 000 张进行训练，1 000 张进行测试。所有图片的尺寸均缩放到 256 × 256 × 3 的大小。训练和测试的掩码采用了随机矩形掩模以及破损比例为 10%~50% 的任意形状掩模数据集[9]。

2.2　实验细节

实验平台选用 NVIDIA RTX2080Ti 显卡，内存为 32 GB，操作系统为 Ubuntu18.04；使用 Pytorch v1.2.0 架构来搭建修复网络；网络参数初始化方式选用正交初始化；优化器使用 Adam 优化器，设置动量衰减指数$\beta_1 = 0$、$\beta_2 = 0.9$，学习率$l_{\text{r}} = 0.000\ 1$；损失函数的权重λ_{rec}、λ_{adv}、λ_{pc}、λ_{style}分别设置为 1、0.1、0.1、250。

本文将 SA-MA-FIC 与 CA[12]、GC[10]、LaFIn[13]、PIC[16]算法在数据集 CelebA-HQ 上进行对比，后 4 者均采用原作者提供的代码及其网络权重进行实验。

2.3　定性分析

为了直观地展示本文提出的人脸图像修复方法的优越性，将该方法与 CA、GC、LaFIn、PIC 4 种方案的修复结果进行了定性评估，见图 6。

尽管 CA、GC 的内容注意力模块以及 PIC 的长短注意力模块均具有一定的特征长程迁移能力，但由于它们均没有合理的人脸先验知识，因此修复结果的拓扑结构往往不合理，人脸五官呈现出明显的扭曲(图 6 第 1 ~3 行)；LaFIn 虽然引入了关键点提供人脸拓扑结构的先验知识，但是其修复结果仍然存在全局语义不合理的问题，尤其表现为肤色不一致(图 6(e)第 2、3 行)以及左右眼亮度、瞳色不对称(图 6(e)第 4、5 行)。SA-MA-FIC 使用风格感知模块对修复过程进行了全局性的调节；并使用多尺度注意力模块对多尺度特征进行长程迁移，有效地恢复了细节信息。因此，相比于上述 4 种方案，SA-MA-FIC 能够生成拓扑结构和全局语义均合理的修复图像，并且对暗光图像的修复效果更加自然真实(图 6(f)第 2、3 行)。

2.4　定量分析

为了更加客观地展示本文提出的人脸图像修复方法的优越性，将该方法与 CA、GC、LaFIn、PIC 4 种方案的修复结果进行了定量评估。评估指标包括文献[22]提出的峰值信噪比(peak signal to noise ratio，PSNR)、结构相似性(structure similarity index，SSIM)、文献[23]提出的图像补丁感知相似性(learned perceptual image patch similarity，LPIPS)，这些指标分别用来衡量修复图像与原始图像的像素级差异、整体相似度以及感知相似度。其中，PSNR 和 SSIM 指标越高说明修复效果越好，而 LPIPS 则相反。

定量评价结果见表 1。从测试结果来看，SA-MA-FIC 的 PSNR 和 SSIM 指标均明显高于其他几种方案，LPIPS 指标则低于它们，这说明与现有先进方案相比，本文提出的 SA-MA-FIC 具有更强的人脸图像修复能力。

2.5　消融研究

为了验证本文所提出的风格感知模块和多尺度注意模块的有效性，设计了 SA-MA-FIC 的两种变体 MA-FIC 和 SA-FIC。MA-FIC 仅使用多尺度注意模块，去除了风格渲染块中的 AdaIN 操作；SA-FIC 仅使用风格感知模块，直接将未经长程特征迁移的多尺度特征输入解码器。实验将 SA-MA-FIC 与二者进行了详细地对比，结果见图 7。

图 6　在 CelebA-HQ 测试集上的定性评价

Fig. 6　Qualitative evaluation on CelebA-HQ dataset

表 1　在 CelebA-HQ 测试集上的定量评估结果

Tab. 1　Quantitative evaluation results on CelebA-HQ test set

掩模类型	CA[12]			GC[10]			LaFIn[13]			PIC[16]			SA-MA-FIC		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
10%~20%	27.73	0.937	0.069	31.17	0.971	0.034	31.88	0.974	0.028	33.17	0.979	0.030	35.09	0.986	0.018
20%~30%	24.67	0.884	0.111	27.95	0.943	0.059	28.75	0.950	0.046	29.61	0.958	0.048	31.15	0.969	0.034
30%~40%	22.32	0.817	0.158	25.59	0.906	0.088	26.36	0.917	0.069	26.67	0.920	0.075	28.23	0.943	0.055
40%~50%	20.41	0.736	0.213	23.69	0.862	0.121	24.38	0.876	0.096	24.25	0.870	0.107	26.01	0.910	0.080
random	24.82	0.897	0.067	26.65	0.927	0.053	27.54	0.936	0.043	25.91	0.911	0.053	28.07	0.942	0.041

2.5.1　风格感知模块的作用

实验结果表明由于缺乏风格感知模块的全局性调节，MA-FIC 的修复结果会出现左右眼不对称的现象(图 7(b)第 1 ~3 行)以及在暗光条件下的明显失真(图 7(b)第 4 行)。此外，MA-FIC 的修复结果有时会出现大面积伪影，SA-MA-FIC 则有效避免了上述问题(图 7(d)第 5 行)。

另外，由表 2 可知，与 MA-FIC 相比，SA-MA-FIC 的 PSNR、SSIM、LPIPS 指标均呈现显著提升，这说明风格感知模块能够有效地提升修复图像的全局语义合理性。

2.5.2　多尺度注意模块的作用

实验结果表明，当左右眼都被遮蔽时，SA-FIC 能够保证修复结果的全局语义合理性(图 7(c)第 3、4 行)；但是，当只有一只眼睛被遮蔽时，由于 SA-FIC 特征长程迁移能力的不足，就无法再保证修复结果全局语义的合理(图 7(c)第 1、2 行)。

此外，由表 2 可知，虽然 SA-FIC 的 PSNR、SSIM 指标均基本与 SA-MA-FIC 持平，但是 SA-FIC 的 LPIPS 性能则明显不及 SA-MA-FIC，这说明多尺度注意力模块能够有效提升修复结果与原始图像的知觉相似度。但是，多尺度注意力模块的设计上存在

一定缺陷，如图 7 第 6 行所示，SA-MA-FIC、MA-FIC 修复结果的右眼部分存在明显的失真，而 SA-FIC 反而不存在这一问题。这是由于多尺度注意力模块是基于多尺度补丁块进行特征长程迁移的，然而对于大角度的侧脸图像而言，左眼和右眼的轮廓并不相同，使用这种迁移方式并不合理。

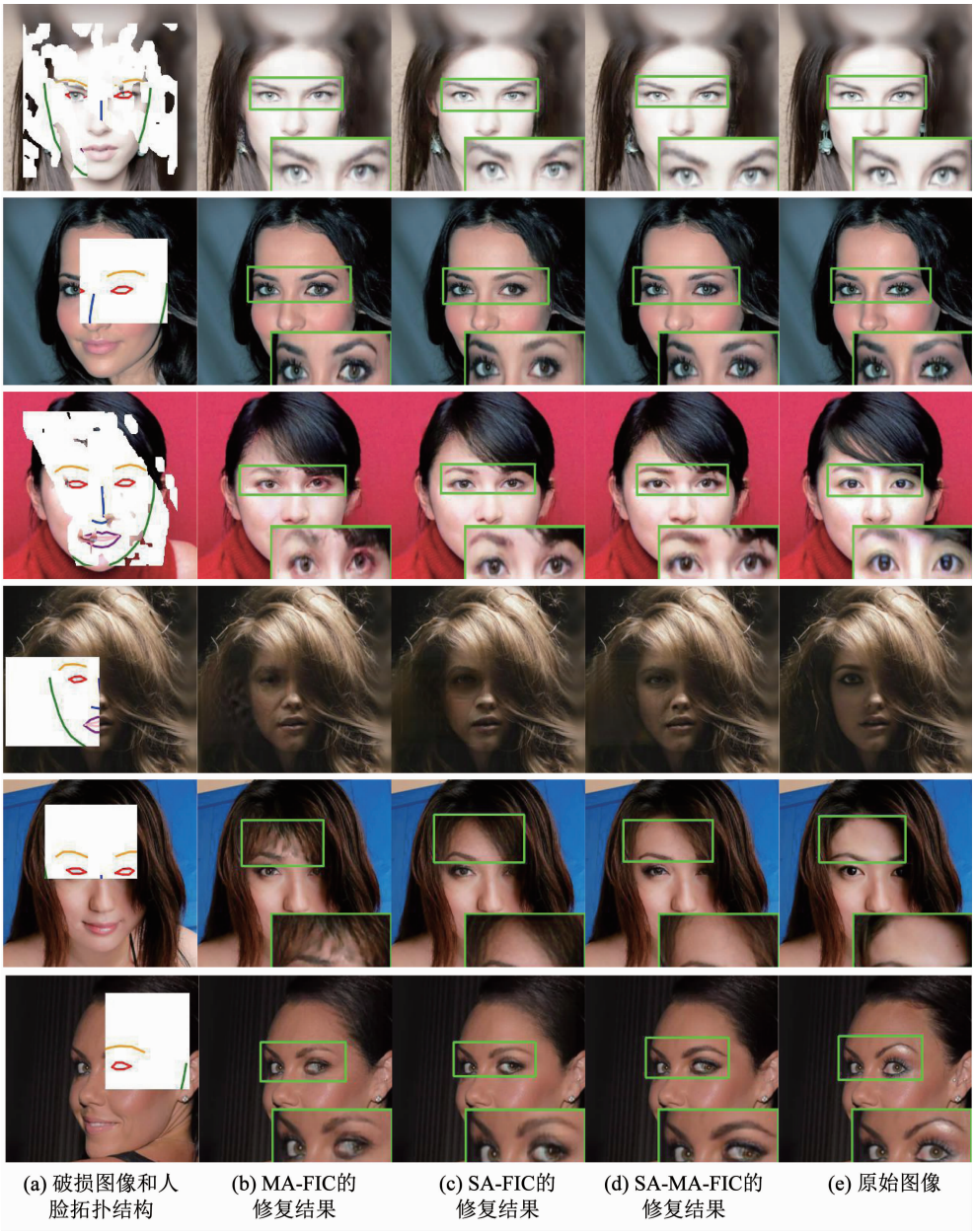

图 7　本文模型的不同变体在 CelebA-HQ 测试集上的定性评价

Fig. 7　Qualitative evaluations on CelebA-HQ dataset across different variations of the proposed model

表 2　模型的不同变体在 CelebA-HQ 测试集上的定量评估结果

Tab. 2　Quantitative evaluation results on CelebA-HQ test set across different variations of the proposed model

掩模类型	MA-FIC			SA-FIC			SA-MA-FIC		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
10%~20%	34.64	0.984	0.019	34.66	0.985	0.020	35.09	0.986	0.018
20%~30%	30.80	0.966	0.037	30.88	0.966	0.037	31.15	0.969	0.034
30%~40%	27.97	0.939	0.060	28.13	0.941	0.060	28.23	0.943	0.055
40%~50%	25.78	0.904	0.086	25.88	0.910	0.085	26.01	0.910	0.080
random	27.79	0.937	0.044	27.87	0.942	0.044	28.07	0.942	0.041

3 结 论

针对人脸图像修复结果全局语义不合理的问题,提出了一种基于风格感知和多尺度注意力的人脸修复方法。首先,设计了一种风格感知模块,实现了对人脸图像修复过程的全局调节;另外,为了提高修复网络的特征长程迁移能力,设计了一种多尺度注意力模块,有效地恢复了面部细节信息。与多种先进方法的对比实验表明,本文所提方法的修复结果在主观视觉上的效果更加自然逼真,在客观的像素级相似度、整体相似度和感知相似度指标上也得到了显著提升,有效地解决了人脸修复图像全局语义不合理的问题。但是,对于大角度的侧脸图像而言,本文方案的修复结果仍然会出现左眼和右眼不对称的现象,未来可通过开发对侧脸鲁棒的多尺度特征长程迁移模块来解决这一问题。

参考文献

[1] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting[C]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 2000: 417. DOI: 10.1145/344779.344972

[2] ELAD M, STARCK J L, QUERRE P, et al. Simultaneous cartoon and texture image inpainting using morphological component analysis [J]. Applied and Computational Harmonic Analysis, 2005, 19(3): 340. DOI: 10.1016/j.acha.2005.03.005

[3] 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): 24. DOI: 10.1145/1576246.1531330

[4] CRIMINISI A, PÉREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200. DOI: 10.1109/TIP.2004.833105

[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems. Montreal: NIPS, 2014: 2672

[6] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234. DOI: 10.1007/978-3-319-24574-4_28

[7] 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. Seattle: IEEE, 2016: 2536. DOI: 10.1109/CVPR.2016.278

[8] LIZUKA S, SIMO-SERRA E, LSHIKAWA H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics, 2017, 36(4): 1. DOI: 10.1145/3072959.3073659

[9] LIU Guilin, 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. Munich: Springer, 2018: 85. DOI: 10.1007/978-3-030-01252-6_6

[10] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution [C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE, 2019: 4471. DOI: 10.1109/ICCV.2019.00457

[11] JO Y, PARK J. SC-FEGAN: Face editing generative adversarial network with user's sketch and color[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE, 2019: 1745. DOI: 10.1109/ICCV.2019.00183

[12] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5505. DOI: 10.1109/CVPR.2018.00577

[13] ZHENG C, CHAM T J, CAI J. Pluralistic image completion[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 1438. DOI: 10.1109/CVPR.2019.00153

[14] NASH J F. Non-cooperative games[J]. Annals of Mathematics, 1951, 54(2): 286. DOI: 10.2307/1969529

[15] NASH J F. Equilibrium points in n-person games[J]. National Academy of Sciences, 1950, 36(1): 48. DOI:10.1073/pnas.36.1.48

[16] YANG Y, GUO X, MA J, et al. Generative landmark guided face inpainting[C]//Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision. Cham: Springer, 2020: 14. DOI: 10.1007/978-3-030-60633-6_2

[17] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 1501. DOI: 10.1109/ICCV.2017.167

[18] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [C]//Proceedings of the International Conference on Machine Learning. Long Beach: International Machine Learning Society, 2019: 7354

[19] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2794. DOI: 10.1109/ICCV.2017.304

[20] JUSTIN J, ALEXANDRE A, LI Feifei. Perceptual losses for real-time style transfer and super-resolution [C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 694. DOI: 10.1007/978-3-319-46475-6_43

[21] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2016: 2414. DOI: 10.1109/CVPR.2016.265

[22] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600. DOI: 10.1109/TIP.2003.819861

[23] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 586. DOI:10.1109/CVPR.2018.00068

(编辑 苗秀芝)