



## 基于多尺度特征融合的人脸图像修复方法

白宗文<sup>1</sup>, 弋婷婷<sup>1</sup>, 周美丽<sup>1</sup>, 魏 崑<sup>2</sup>

(1. 延安大学 物理与电子信息学院, 陕西 延安 716000; 2. 西安理工大学 计算机学院, 西安 710048)

**摘要:**传统图像修复方法在修复受损区域较大的图像时会出现修复结果过于平滑或模糊的现象,并且较难重建合理的人脸图像结构。在传统生成对抗网络的鉴别器中引入多尺度特征融合方法,将不同深度的特征图经过上采样后直接相加,使浅层信息和深层信息有效结合。通过借助高层特征把握图像的整体规律,同时利用低层特征填充人脸图像的细节纹理,进而使一张图像的分辨率及其语义特征相互融合,实现有效的人脸图像修复。在CelebA数据集上的实验结果表明,该方法的峰值信噪比、相似性结构、L1损失指标均优于区域归一化方法,取得了较好的视觉效果。

**关键词:**图像修复;生成对抗网络;多尺度特征融合;鉴别器;高层特征;低层特征

开放科学(资源服务)标志码(OSID):



**中文引用格式:**白宗文, 弋婷婷, 周美丽, 等. 基于多尺度特征融合的人脸图像修复方法[J]. 计算机工程, 2021, 47(5): 213-220, 228.

**英文引用格式:**BAI Zongwen, YI Tingting, ZHOU Meili, et al. Face image inpainting method based on multi-scale feature fusion[J]. Computer Engineering, 2021, 47(5): 213-220, 228.

## Face Image Inpainting Method Based on Multi-Scale Feature Fusion

BAI Zongwen<sup>1</sup>, YI Tingting<sup>1</sup>, ZHOU Meili<sup>1</sup>, WEI Wei<sup>2</sup>

(1. School of Physics and Electronic Information, Yan'an University, Yan'an, Shaanxi 716000, China;

2. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China)

**[Abstract]** When dealing with severely corrupted images, the traditional image inpainting methods often produce blurry or too smooth regions in the restored images, and have difficulty in reconstructing a reasonable face image structure. To address the problem, this paper introduces a multi-scale feature fusion method into the discriminator of the traditional Generative Adversarial Network (GAN), which directly adds the upsampled feature maps of different depth to achieve effective fusion of shallow and deep information. Then this method grasps the overall pattern of the image with the help of high-level features, and fills the detail texture of the face image with low-level features, so that the resolution of the image can be fused with its semantic features for effective face image restoration. The experimental results on the dataset of CelebA show that the proposed method outperforms the regional normalization method in terms of Peak Signal to Noise Ratio (PSNR), structural similarity, L1 loss indicators, achieving ideal visual effects.

**[Key words]** image inpainting; Generative Adversarial Network (GAN); multi-scale feature fusion; discriminator; high-level feature; low-level feature

DOI: 10.19678/j.issn.1000-3428.0060053

### 0 概述

图像修复是指利用受损图像周围合理可信的信息来重建图像的损坏区域,此技术被广泛应用于目标移除<sup>[1]</sup>、图像翻译<sup>[2]</sup>和图像编辑<sup>[3]</sup>等工程应用领域。现有研究表明,图像修复质量取决于受损面积的大小、图像自身结构特征和图像语义特征等因素。

传统修复方法对于受损面积较小、图像自身结构简单的受损图像具有良好的效果,但其依赖于当前单幅图像未受损区域的信息,通过扩散或者统计计算对修复区域进行信息填充,没有利用复杂的结构特征信息或者高层语义特征,修复结果难以达到语义级的一致性。近年来,随着深度学习技术的兴起,基

**基金项目:**国家自然科学基金(61761042, 6194112);陕西省自然科学基金(2020JM-556);延安大学科研计划(CXY201909, YDZ2019-05);陕西省能源大数据智能处理重点实验室开放基金(IPBED10, IPBED7)。

**作者简介:**白宗文(1979—),男,副教授、博士,主研方向为深度学习;弋婷婷,硕士研究生;周美丽,副教授;魏 崑,副教授、博士。

**收稿日期:**2020-11-18 **修回日期:**2021-01-10 **E-mail:** ydbzw@yau.edu.cn

于深度网络模型的图像修复取得了良好的效果,尤其是人脸图像修复,这一技术的修复质量、视觉效果以及在不同领域中的应用成为关注的焦点。目前,如何使受损图像保持像素层和语义层的一致性,并且避免损坏区域与未损坏区域的不完全融合成为人脸图像修复的关键问题。采用深度网络模型进行图像修复,除了利用当前受损图像的自身信息以外,还可以结合图像结构特征和语义特征获得更优的修复质量和视觉效果,使修复的合理性得到进一步提高。

通常传统的图像修复网络都会引入特征归一化方法进行辅助训练,但对于人脸图像修复网络而言,采用这种跨空间的归一化方法容易导致在提取受损区域以及未受损区域的特征时出现内部协方差偏移的问题<sup>[4-5]</sup>,极大地影响了最终的修复质量。这一问题在区域归一化<sup>[6]</sup>方法中能够得到有效解决,但现有的图像修复方法大多是从单一尺度进行图像特征提取,较少将语义信息与不同尺度的特征进行融合。

为改善针对人脸细节的修复效果,并且有效保留高层特征的结构化信息,使整体的修复效果更符合人类视觉感知,本文通过结合多尺度特征融合方法和生成对抗网络(Generative Adversarial Network, GAN),构建一种新的网络结构——多尺度特征融合对抗网络。

## 1 相关工作

图像修复的目的是重建图像中丢失或损坏的部分,从而恢复原图像的清晰像素并保持其与周围场景的一致性。目前,图像修复技术已在计算机视觉领域得到广泛应用,成为图像处理一个关键且极具挑战的研究课题。本节从非基于学习的传统方法和基于学习的方法两个方面出发,对图像修复方法做简要的归纳分析,并在现有研究的基础上,针对目前方法存在的问题提出解决方案。

在非基于学习的传统方法中,本文主要介绍基于扩散(diffusion-based)和基于补丁(patch-based)两类方法。传统方法主要依赖受损图像的未破坏区域信息来填充受损区域,可以视为像素级的图像修复,因而修复的结果很难达到语义级的一致性。在基于学习的方法中,本文主要介绍通过学习图像的语义推断缺失内容的图像修复方法。

### 1.1 非基于学习的传统方法

#### 1.1.1 基于扩散的方法

MARCELO B等人提出通过基于扩散的方法<sup>[7-8]</sup>

来填充损坏的区域,实际就是将可用的图像信息传播到需要修复的区域,即从孔的边界传播到最小变化方向的信息,通常将扩散过程用等照度线或热扩散等物理模型建模。在使用此类方法时,修复区域过大会导致运算量急剧增加,且由外边界向区域中心的信息扩散过程会因距离问题速度变慢。此外,扩散过程会使修复区域过于平滑,并且在含有丰富纹理的情况下,能够利用的图像块的相似性会减弱,导致匹配到合适纹理块的难度加大,从而使修复区域变得模糊,不能得到理想的修复结果。

#### 1.1.2 基于补丁的方法

对于图像和视频的合成与分析,基于补丁的采样方法是一种被普遍使用的主流方法。同时,不同应用中使用的搜索技术以及交互性程度也各有差异。文献[9]提出一种应用于区域填充和目标移除的基于范例(exemplar-based)的图像修复算法。该算法基于样本的纹理合成效率和定性性能,同时兼顾周围线性结构对图像的约束,引入优先级函数来决定目标函数的填充顺序,并通过基于补丁的采样过程提高计算效率。在应用方面,该算法在去除大的遮挡物以及修复细划痕方面效果显著。文献[10]提出的结构传播图像处理方法解决了显著性结构断裂的问题,其以全局优化问题描述结构传播,同时利用动态规划方法进行求解。当用户指定多条曲线时,采用置信度传播算法寻求补丁。但如果用复杂曲线表示缺失部分时,会大幅增加图像块间的匹配难度。文献[11]提出一种快速寻找图像块之间近似临近匹配的随机算法,核心思想是先通过采样找到合适的匹配点,再利用自然图像中的连贯性将匹配点快速传播到周围区域。文献[12]提出一种利用中层结构线索自动引导基于补丁的图像完成算法,该算法在拓展补丁距离以及采样方案中整合平移规律约束,并将其作为补丁偏移和转换的先验概率来达到最终目的。同时该算法还可以自动恢复到基线平行平面(fronto-parallel)图像,解决了缺乏可靠平面检测的问题。区别于传统的纹理合成算法,该算法能够平衡分析和合成过程,提高修复性能。

### 1.2 基于学习的方法

传统的图像修复方法大多假定在平滑性、局部相似性等约束条件下通过偏微分方程、能量函数、热扩散方程等物理模型进行建模来实现图像修复,只适用于缺少少量信息的图像,是典型的局部化操作,缺乏对整体图像的统一认知。近年来,随着深度学

习在计算机视觉领域的应用范围不断扩大, 基于深度学习的图像修复技术成为研究热点。与传统方法不同, 基于深度学习的修复方法能够模仿人类进行图像修复的过程, 实现从像素级到语义级的多层次融合修复, 并取得具有高保真度的修复结果。其中, 基于GAN的图像生成方法在修复工作中被广泛应用, 其对随机噪声进行映射学习, 能够生成丰富多样的能够以假乱真的新图像, 实现“无中生有”效果。最早将GAN引入到图像修复领域的Context Encode网络<sup>[13]</sup>既可以理解整个图像的内容, 又可以为缺失的部分生成一个合理的假设。但该方法不能处理任意图像的修补掩膜, 且仅限于修复低分辨率图像, 缺乏对局部一致性的考虑。文献<sup>[14]</sup>提出的图像修复方法同时考虑局部一致性和全局一致性, 并且利用完全卷积神经网络在任何形状缺失区域内填充修复任意分辨率的图像。该方法可适应各种场景, 如在文献<sup>[15]</sup>提到的人脸修复就是利用全局-局部判别器的思想。

基于学习的深度框架通过自动捕捉图像的内在层次来理解图像内容, 并生成高层次的语义特征来修复缺失内容, 一般可以取得比传统方法更优的结果。基于GAN的图像修复可以在特定类型的图像修复中取得较好的效果, 但是对于高分辨率的图像修复, 会因图像中内容不相似而导致难以收敛, 从而加大GAN的训练难度。而多任务学习框架<sup>[16]</sup>结合结构嵌入和注意机制, 通过训练共享生成器, 利用图像结构知识的整合来辅助图像的修复。在有关注意力机制的研究中, 对孔洞区域与上下文之间的相关性建模被用于图像修复任务, 例如文献<sup>[17]</sup>提出利用多尺度上下文注意模块(MUSICAL)进行不同尺度下的特征图交换, 通过该方法可以使某些注意层获得预期的特征图, 从而避免误用具有丰富纹理背景的信息。此外, 文献<sup>[18]</sup>方法通过引入可学习的双向注意力机制(LBAM)使可学习的注意力图能够适应任意不规则孔洞, 并在卷积过程中有效传播。文献<sup>[19]</sup>方法通过引入连贯性语义注意层(CSA)增强了受损区域的语义相关性和特征连续性。该方法包含粗、精两个分支, 将其加入CSA加入到精细网络层中, 不仅可以保持上下文结构一致, 而且可以预测缺失部分。

## 2 融合多尺度特征的人脸图像修复

本文将多尺度特征融合方法与生成对抗网络相

结合, 构建一种多尺度特征融合对抗网络用于人脸图像修复。

### 2.1 生成对抗网络

生成对抗网络(GAN)<sup>[20]</sup>的设计主要受博弈论中“零和博弈”的启发。将GAN引入深度学习神经网络中, 即利用生成器 $G$ 和判别器 $D$ 相互对抗博弈的促进过程, 使 $G$ 学习到的数据分布能够生成更逼真的图像。在训练过程中, 判别器 $D$ 的目标即鉴别出 $G$ 生成的虚假照片 $G(z)$ , 使 $D(G(z))$ 的概率尽可能小, 而生成器 $G$ 的目标是使生成的图像 $G(z)$ 尽可能地欺骗鉴别器, 使 $D(G(z))$ 的概率尽可能大。由此就产生了 $G$ 和 $D$ 不断动态博弈的过程。这个不断循环往复的对抗过程能够使最终的生成图像达到“以假乱真”的效果。GAN原理框图如图1所示。

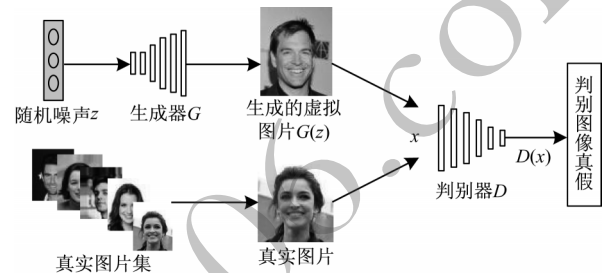


图1 GAN原理框图

Fig.1 Principle block diagram of GAN

### 2.2 特征金字塔网络

特征金字塔网络(Feature Pyramid Network, FPN)最早用于解决物体检测中的多尺度问题, 特别是对小物体的检测效果显著<sup>[21]</sup>, 其利用常规CNN模型高效提取图片中各维度特征, 本质是一种加强主干网络的特征表达方法。

为更直观地理解FPN的思想原理, 基于ResNet101给出FPN网络模型的结构原理框图, 如图2所示(彩色效果见《计算机工程》官网HTML版), 其中, 红色长方体表示卷积及下采样过程, 橙色长方体代表普通卷积过程, 蓝色长方体代表最终生成的特征图。由图2可知, 一个深层FPN在“自上而下”(图中“自左向右”)提取特征时使用每个阶段最后的输出特征进行横向连接。此外,  $1 \times 1$ 卷积的使用也可以灵活改变通道数, 使各层通道数最终保持一致。而利用“自上而下”的最邻近插值法进行上采样的过程, 可使最终的特征图既有合适的分辨率又有表达能力较强的语义信息, 而 $3 \times 3$ 卷积的使用则有效地避免了混叠效应。

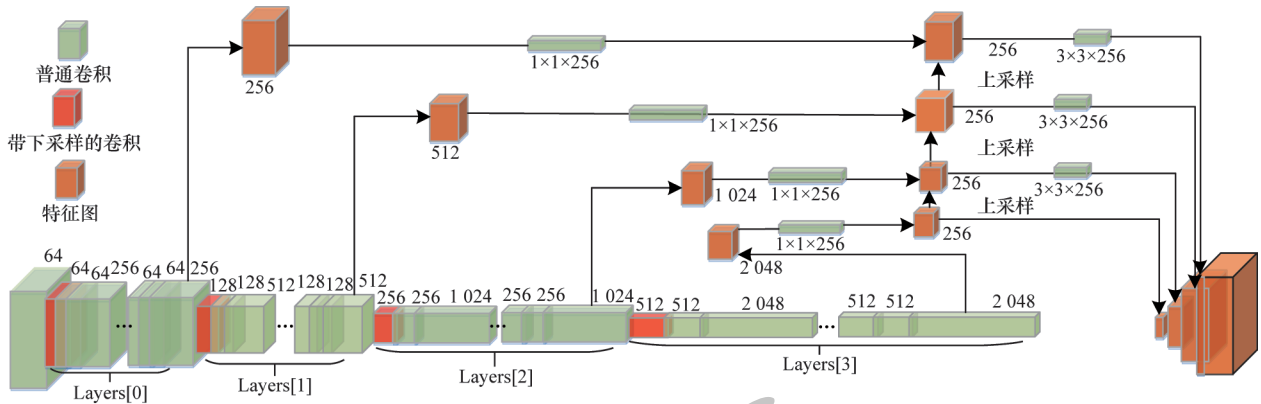


图2 基于 ResNet101的 FPN原理框图

Fig.2 Principle block diagram of FPN based on ResNet101

2.3 多尺度特征融合的网络框架

文献[5]设计的网络采用区域归一化(Area Normalization, RN)方法,实验结果表明其优于实例归一化(Instance Normalization, IN)方法。因此,本文以该网络作为主要架构,继续在网络的生成器中沿用区域归一化方法,而对于生成器的考虑,本文则利用多尺度特征融合的方法,使生成器部分提取到比传统网络语义更强、有效信息更精确的特征。本文设计的基于多尺度特征融合对抗框图如图3所示(彩色效果见《计算机工程》官网HTML版)。与PatchGAN<sup>[22-23]</sup>中判别器只是简单地使用卷积网络提取单一图像特征的方法不同,本文方法从多尺度特征角度考虑对判别器的特征提取方式进行改变。在图3中的判别器部分,最左边从底层到顶层(浅紫色块)首先利用传统的卷积方法对图片进行特征提取,然后通过1×1卷积(左边蓝色箭头)扩大通道数,这样可以增加有效信息的采集,而中间所得到的特征图(深紫色块)利用上采样及加和(棕色向下箭头)的方式有效地融合了高层特征的结构语义信息和不同层分辨率的细节纹理,再利用3×3卷积(右边浅橙色箭头)调节后得到的特征图各自经过一个自适应池化(淡橙色块),最后将得到的特征有效的连接(浅黄色块)在一起作为输出。经过这样一个不断融合的过程,可在提升判别器判别力的同时促进生成器生成更逼真的图像,由此增强对图像的修复能力。

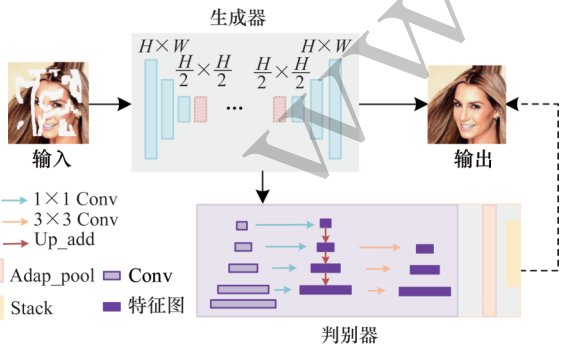


图3 基于多尺度特征融合的对抗框架

Fig.3 Adversarial framework based on multi-scale feature fusion

由于高层特征的语义信息可以对单幅图像的高频结构特征进行有效展现,因此能够缓解修复图像

时缺少丰富的轮廓信息而造成的平滑现象。而对于细节纹理的获得,将低频信息的高分辨率填充到损坏的底层部分,使图像在视觉上更清晰自然。基于以上两点分析,本文通过图像高层特征把握图像整体规律,利用低层特征填充高层特征中被遮挡的孔洞,进而将合适的分辨率以及语义特征相融合,结合损坏的单幅图像的纹理信息与到边界明确的结构信息,提高图像修复的质量。

此外,对于不同分辨率的图像,本文采用的1×1小卷积灵活调节所需的通道数。例如在提取人脸图像信息时将原始的通道数扩展到256,虽然参数增加了近2倍,但是更有效地保存了小目标(如人眼)的细节及结构信息。同时本文对于更大卷积的使用也做了一定考虑,例如将1×1卷积换成3×3卷积,理论分析这样可以保留更多的有效信息,但是会造成参数急剧增加,导致训练速度变慢,若换用更大的卷积对设备的配置将提出更高的要求,极大地增加了成本输出。

2.4 损失函数

生成对抗网络的损失函数是在生成器和鉴别器之间交替进行的,同时采用了不同的损失函数进行计算。本文模型损失函数的使用与基线方法大致相同,用 $I_{gt}$ 表示输入的原图,用 $M$ 表示二进制掩膜,两者经过Khatri-Rao积(简称KR积,符号表示为 $\odot$ )运算,得到带掩膜的损坏图像,计算公式如下:

$$\tilde{I}_{gt} = I_{gt} \odot (1 - M) \tag{1}$$

其中, $\tilde{I}_{gt}$ 表示输入的待修复图像。经过本文模型的修复融合结果用公式表达如下:

$$I_{pred} = \tilde{I}_{gt} \times M + I_{gt} \times (1 - M) \tag{2}$$

其中, $I_{pred}$ 即为模型预测的修复结果。因此,该模型的对抗损失为:

$$L_{adv} = E_{I_{gt}} [\lg D(I_{gt})] + E_{(\tilde{I}_{gt}, I_{pred})} [\lg (1 - D(I_{pred}))] \tag{3}$$

此外,在对损失函数进行计算时,本文利用标签平滑方式<sup>[24]</sup>将原有真实标签和虚假标签分别改为0.9及0.1,这样可以提高模型的泛化能力,使模型的训练结果具有更强的鲁棒性。

本文引入风格损失函数<sup>[25]</sup>进行相关判断,如式(4)所示:

$$L_{\text{style}} = E_i \left[ \left\| G_i^\phi(I_{\text{pred}}) - G_i^\phi(I_{\text{gt}}) \right\|_1 \right] \quad (4)$$

其中,  $G_i^\phi$  是  $C_i \times C_i$  Gram 矩阵由激活图  $\phi_i$  构建而成。由于要表达图像的纹理特征, 因此此处引入了 Gram 矩阵, 这种求两个特征图内积的结果和结构没有太大关系, 因此, 可以用来度量纹理特征和结构位置并无直接联系的图像的细节纹理。

为增强提取特征的细节, 本文引入了感知损失<sup>[26]</sup>, 表达式如下:

$$L_{\text{prec}} = E \left[ \sum_j \frac{1}{N_j} \left\| \phi_j(I_{\text{gt}}) - \phi_j(I_{\text{pred}}) \right\|_1 \right] \quad (5)$$

其中,  $\phi_j$  表示预训练 VGG<sup>[27]</sup> 网络得到的激活图。通过预训练出的网络来捕捉图像的细节特征可以得到更好的修复结果。

将以上损失与 L1 损失结合起来构成本文的整体目标损失, 如式(6)所示:

$$L_{\text{total}} = \lambda_l L_l + \lambda_{\text{adv}} L_{\text{adv}} + \lambda_s \lambda_{\text{style}} + \lambda_p \lambda_{\text{prec}} + \varepsilon \quad (6)$$

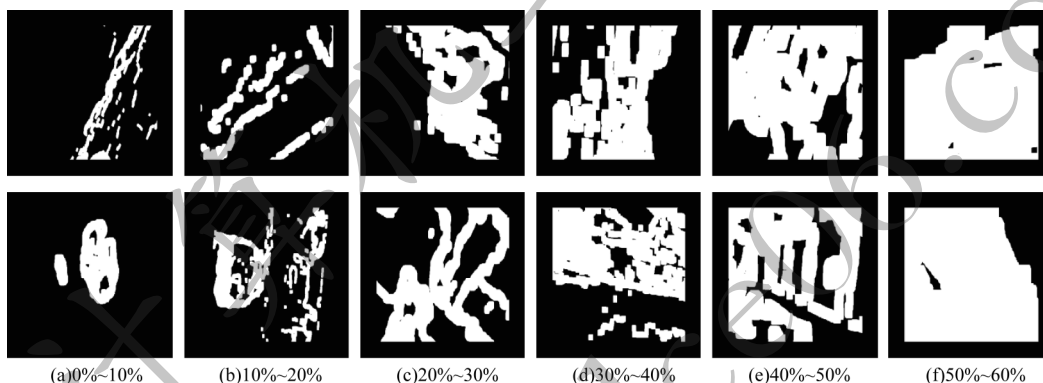


图4 不规则掩膜各等级区域展示图

Fig.4 Regional display images of various grades of irregular mask

### 3.2 实验步骤与评价指标

通过 PyTorch 实现本文方法模型, 在 Ubuntu14.0 环境下利用 3 块 1080Ti GPU 进行训练, 批量设为 14, 同时使用 Adam<sup>[30]</sup> 进行优化。

采用 3 种评价指标对修复完成的图像进行最终的质量评估, 分别是 L1 损失、峰值信噪比 (Peak Signal to Noise Ratio, PSNR)<sup>[31]</sup>、结构相似性 (Structural Similarity, SSIM)<sup>[32]</sup>。

损失函数 L1 基于逐像素比较差异, 然后取绝对值对图像进行相应的重建能力评估, 计算公式如下:

$$L_{L_1} = \frac{1}{N} \sum_{p \in P} |x(p) - y(p)| \quad (7)$$

其中,  $x(p)$  表示真实图像的像素值,  $y(p)$  表示修复图像的像素值。

PSNR 是衡量图像质量的另一种指标, 计算 PSNR 要先对图像的均方误差 (Mean Square Error, MSE) 进行计算, 计算公式如下:

$$M_{\text{MSE}} = \frac{1}{HW} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} [I(i,j) - K(i,j)]^2 \quad (8)$$

其中,  $\varepsilon$  表示无限接近 0 的最小正数, 能够对整体的目标损失加以约束, 使产生的结果得到校正, 提高修复的整体性能。此外, 本文所设置的引入损失函数的权重分配为  $\lambda_l = 1, \lambda_{\text{adv}} = 0.1, \lambda_s = \lambda_p = 0.0001$ 。

## 3 实验与结果分析

### 3.1 数据集

在 CelebA 数据集<sup>[28]</sup> 上使用本文方法进行实验, 其中训练集有 162 770 张人脸图像, 验证集有 19 867 张图像, 测试集有 19 962 张图像。实验还使用不规则掩膜 (irregular mask) 数据集<sup>[29]</sup>, 其中包含 12 000 张不规则掩膜图像, 掩膜区域占整张图像 0%~60%。该数据集掩膜区域分为 6 个等级, 不规则掩膜各等级区域展示如图 4 所示, 其中包括 0%~10%、10%~20%、20%~30%、30%~40%、40%~50%、50%~60% 这 6 种遮挡率的掩膜形式, 且各部分均分别包含 2 000 张掩膜图像。

其中,  $H, W$  是单色图像  $I, K$  的尺寸大小,  $I(i, j)$  是真实图像的像素表示,  $K(i, j)$  是修复图像的像素表示。由此, PSNR 的计算公式表示如下:

$$P_{\text{PSNR}} = 10 \times \lg \left( \frac{\text{MAX}_I^2}{M_{\text{MSE}}} \right) \quad (9)$$

其中,  $\text{MAX}_I$  表示图像中某点颜色的最大值, 例如本文每个采样点用 8 位表示, 则其颜色最大值为 255。

SSIM 主要通过亮度 (luminance)、对比度 (Contrast) 和结构 (Structure) 三方面对两张图像进行最终的结构相似性评估, 计算公式如下:

$$S_{\text{SSIM}}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (10)$$

其中,  $x$  和  $y$  表示真实值和修复值,  $\mu_x$  和  $\mu_y$  表示真实图像和修复图像的均值,  $\sigma_x$  和  $\sigma_y$  表示真实图像和修复图像的方差,  $\sigma_{xy}$  表示协方差,  $c_1$  和  $c_2$  为常数, 取值分别为  $0.01^2$  和  $0.03^2$ 。

应注意的是,上述3种评价指标均依据灰度图像进行评估。

### 3.3 结果分析

通过以上实验参数的设置,本文方法最终的修复结果如图5所示。从人类视觉感知角度可以明显看出,本文方法所取得的修复图像与真实图像相比几乎没有差别。

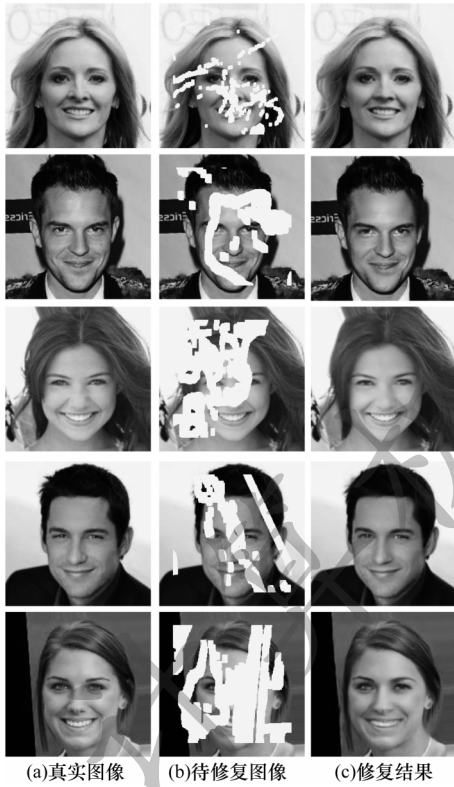


图5 本文方法修复结果  
Fig.5 Repair results of the proposed method

#### 3.3.1 定量分析

将本文方法与区域归一化(RN)方法进行对比,两种方法都是在测试数据集数量为19 870张CelebA人脸图像上进行测试,输入的图像最终尺寸都会调整为256像素×256像素。分别设置5组实验进行图像质量指标测试,结果如图6~图8所示。由图6和图7可以看出,本文方法所得结果的折线图均位于RN方法之上,表明本文方法修复结果效果更佳。由图8可以看出,本文方法的L1损失折线位于下方,同样反映了本文方法的修复效果更优。在10%~20%的掩膜区域,将1 000张图像与输出的19 870张人脸图像相比,本文方法得到的测试结果都优于RN方法,其中,PSNR相比于RN方法提高了2.54 dB,SSIM测试结果提高了0.02,L1损失的值下降了0.20%。在其他的掩膜区域(20%~30%、30%~40%、

40%~50%)中依照同样的方法进行测试,而最后的0%~60%是对各区域掩膜等比随机选取(各区域分别约为333张),最终随机共取了2 000张不规则掩膜与19 870张人脸图像进行测试。由测试数据可知,对于0%~60%区域的掩膜,本文方法相比于RN方法的PSNR值提高了3.14 dB,SSIM提高了0.05,而L1损失下降了0.67%。

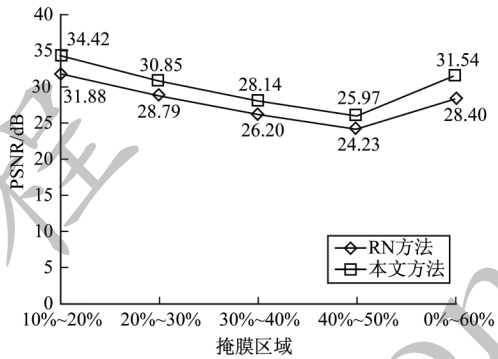


图6 PSNR指标对比  
Fig.6 PSNR index comparison

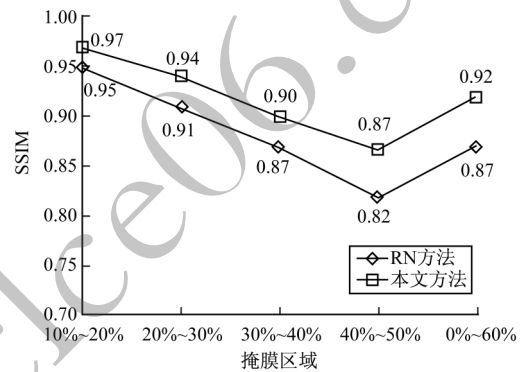


图7 SSIM指标对比  
Fig.7 SSIM index comparison

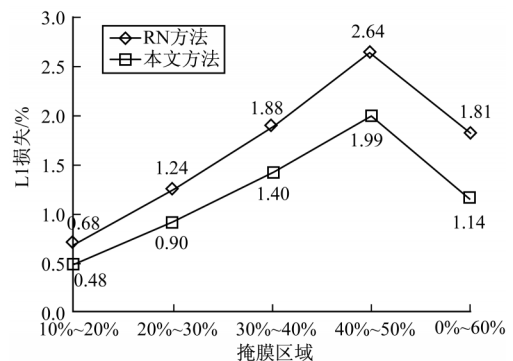


图8 L1损失对比  
Fig.8 L1 loss comparison

#### 3.3.2 定性分析

本文方法与RN方法的修复效果对比如图9所示,可以看出,本文方法修复的图像清晰自然,在视觉上显得更为合理流畅。

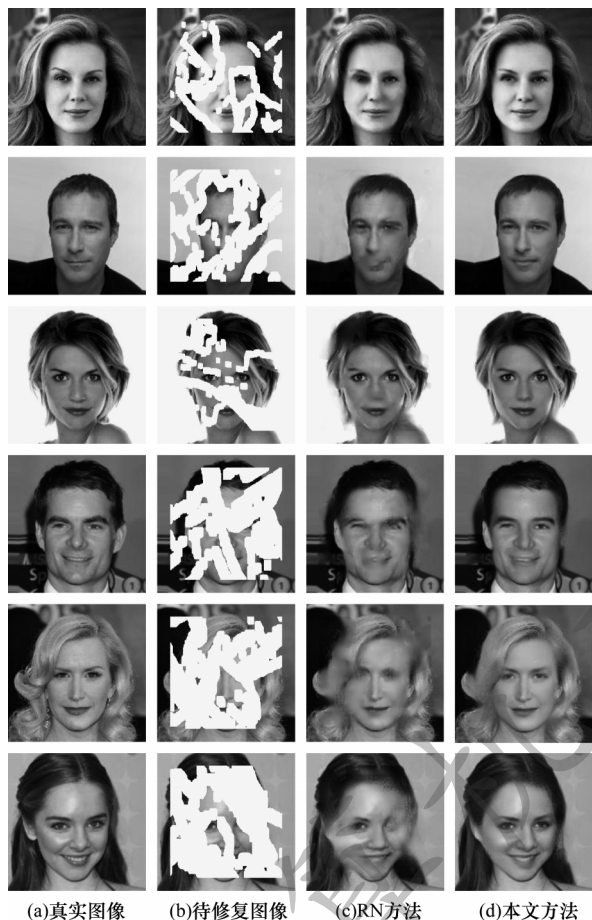


图 9 不同方法修复效果对比

Fig.9 Comparison of repair effects of different methods

2 种方法在细节方面的修复对比如图 10 所示, 通过对对比放大框里各个人眼的修复结果可以看出, 本文方法在细节纹理及结构方面的修复效果更佳, 本文方法修复的人眼没有模糊的重影, 结构纹理清晰明了。

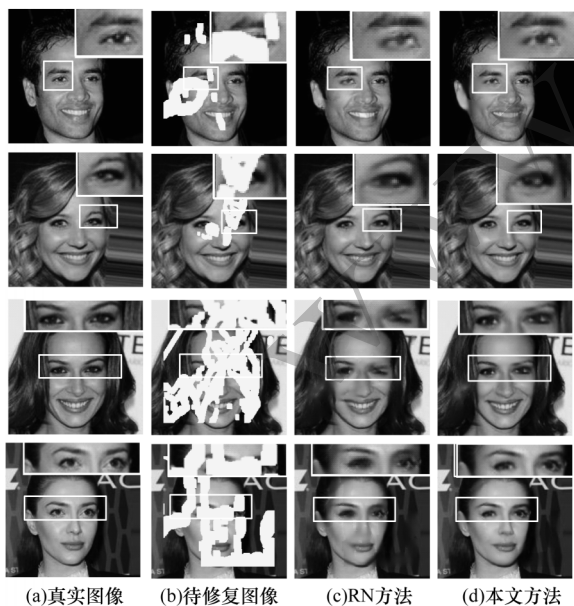


图 10 修复结果细节对比

Fig.10 Comparison of repair results in detail

在不规则大掩膜对整个面部几乎完全遮挡的情况下, 2 种方法的结果对比如图 11 所示, 可以看出, 依然是本文方法取得的修复效果更优, 例如在第 1 行中面部除了人眼未遮挡的情况下, 本文方法最终修复的人脸显示了原图的微笑表情, 且修复结果在视觉感知下结构分明。

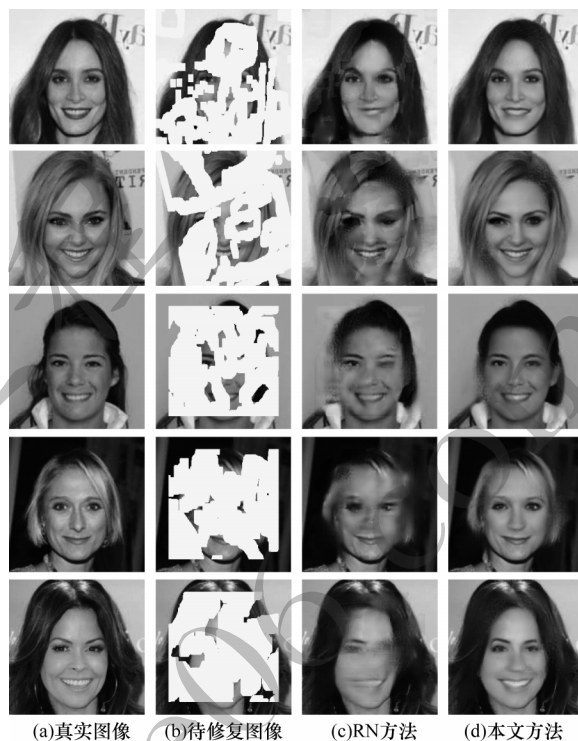


图 11 面部大面积遮挡图像修复效果对比

Fig.11 Comparison of repair effects for images with facial large-area occlusion

图 9 和图 11 的结果表明本文方法可以产生视觉上更合理的修复效果, 而 RN 方法所得到的结果存在扭曲的结构和模糊的纹理, 例如图 9 第 2 行和第 3 行采用 RN 方法修复的人脸存在伪影与模糊的结构, 图 9 最后 3 行利用 RN 修复的人脸面部大范围存在模糊现象。相比之下, 本文方法可以产生视觉上真实及合理的图像。

#### 4 结束语

本文提出一种多尺度特征融合的图像修复方法。通过 GAN 将语义信息与纹理信息有效融合, 保证生成内容的边缘和纹理与原图的内容基本一致, 从而使最终的修复效果满足视觉上的可视性和语义上的合理性。本文在对抗网络的鉴别器中引入 FPN 方法, 同时增加原有的网络通道数, 从而得到合理的图像结构和清晰的真实纹理。实验结果表明, 本文方法的 PSNR、SSIM、L1 损失等指标均优于 RN 方法, 其 PSNR 指标平均提升 2.28 dB, SSIM 指标平均提升 0.04, L1 损失平均下降 0.47%。通过观察定性分析环节展示的人脸图像也可明显看出, 本文方法在

人脸图像修复方面较RN方法能够取得更好的视觉效果。后续将采用两个判别器提升网络的鉴别能力,充分利用图像的有效特征进行图像修复,同时在对抗网络的生成器中引入多尺度特征融合的方法,进一步优化网络生成修复图像时的性能表现。

### 参考文献

- [ 1 ] PATEL H, PETHANI H, PIPALIYA K, et al. Object remove using structure oriented in image inpainting[J]. National Journal of System and Information Technology, 2015, 8(2):88-96.
- [ 2 ] LIU M Y, BREUEL T, KAUTZ J. Unsupervised image-to-image translation networks[C]//Proceedings of Neural Information Processing Systems Conference. Long Beach, USA; [s. n. ], 2017:700-708.
- [ 3 ] JO Y, PARK J. SC-FEGAN: face editing generative adversarial network with user's sketch and color[C]//Proceedings of International Conference on Computer Vision. Washington D. C. , USA ; IEEE Press, 2019:1745-1753.
- [ 4 ] WIESLER S, RICHARD A, RALF S, et al. Mean-normalized stochastic gradient for large-scale deep learning [C]//Proceedings of 2014 IEEE International Conference on Acoustics. Washington D. C. , USA ; IEEE Press, 2014:180-184.
- [ 5 ] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02)[2020-10-10]. <http://arxiv.org/abs/1502.03167>.
- [ 6 ] YU Tao, GUO Zongyu, JIN Xin, et al. Region normalization for image inpainting [C]//Proceedings of International Conference on Artificial Intelligence. [S. l. ]: AAAI, 2020:12733-12740.
- [ 7 ] BERTALMIÓ M, SAPIRO G, CASELLES V, et al. Image inpainting[C]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA; ACM Press, 2000:417-424.
- [ 8 ] BERTALMIÓ M, VESE L A, SAPIRO G, et al. Simultaneous structure and texture image inpainting[J]. IEEE Transactions on Image Processing, 2003, 12(8):882-889.
- [ 9 ] CRIMINISI A, PEREZ P, TOYAMA K, et al. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9):1200-1212.
- [ 10 ] SUN Jian, YUAN Li, JIA Jiaya, et al. Image completion with structure propagation[C]//Proceedings of International Conference on Computer Graphics and Interactive Techniques. New York, USA; ACM Press, 2005:861-868.
- [ 11 ] 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.
- [ 12 ] HUANG J, KANG S B, AHUJA N, et al. Image completion using planar structure guidance [C]//Proceedings of International Conference on Computer Graphics and Interactive Techniques. New York, USA; ACM Press, 2014:1-10.
- [ 13 ] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA ; IEEE Press, 2016:2536-2544.
- [ 14 ] IIZUKA S, SIMOSERRA E, ISHIKAWA H, et al. Globally and locally consistent image completion [J]. ACM Transactions on Graphics, 2017, 36(4):1-14.
- [ 15 ] LI Yijun, LIU Sifei, YANG Jimei, et al. Generative face completion [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA ; IEEE Press, 2017:3911-3919.
- [ 16 ] YANG Jie, QI Zhiqian, SHI Yong, et al. Learning to incorporate structure knowledge for image inpainting[C]//Proceedings of International Conference on Artificial Intelligence. [S. l. ]: AAAI, 2020:12605-12612.
- [ 17 ] WANG Ning, LI Jingyuan, ZHANG Lefei, et al. MUSICAL: multi-scale image contextual attention learning for inpainting[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China; [s. n. ], 2019:3748-3754.
- [ 18 ] XIE Chaozhao, LIU Shaohui, LI Chao, et al. Image inpainting with learnable bidirectional attention maps[C]//Proceedings of International Conference on Computer Vision. Washington D. C. , USA ; IEEE Press, 2019:8858-8867.
- [ 19 ] LIU Hongyu, JIANG Bin, XIAO Yi, et al. Coherent semantic attention for image inpainting[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA ; IEEE Press, 2019:4170-4179.
- [ 20 ] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Proceedings of Conference and Workshop on Neural Information Processing Systems. New York, USA; ACM Press, 2014:2672-2680.
- [ 21 ] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA ; IEEE Press, 2017:2117-2125.
- [ 22 ] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA ; IEEE Press, 2017:1125-1134.
- [ 23 ] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA ; IEEE Press, 2017:2223-2232.
- [ 24 ] HE Tong, ZHANG Zhi, ZHANG Hang, et al. Bag of tricks for image classification with convolutional neural networks [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA ; IEEE Press, 2019:558-567.

(上接第 220 页)

- [25] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]// Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA : IEEE Press, 2016:2414-2423.
- [26] JOHNSON J, ALAHI A, FEI F L. Perceptual losses for real-time style transfer and super-resolution [C]// Proceedings of 2016 European Conference on Computer Vision. Berlin, Germany : Springer, 2016:694-711.
- [27] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2020-10-10]. <https://arxiv.org/abs/1409.1556>.
- [28] LIU Ziwei, LUO Ping, WANG Xiaogang, et al. Deep learning face attributes in the wild [C]// Proceedings of International Conference on Computer Vision. New York, USA : ACM Press, 2015:3730-3738.
- [29] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions [C]// Proceedings of 2018 European Conference on Computer Vision. Berlin, Germany : Springer, 2018:85-100.
- [30] ADAM K D, BA J. Adam: a method for stochastic optimization[EB/OL]. (2017-01-30)[2020-10-10]. <https://arxiv.org/abs/1412.6980>.
- [31] ALAIN H, ZIOU D. Image quality metrics: PSNR vs. SSIM [C]// Proceeding of the 20th International Conference on Pattern Recognition. Washington D. C. , USA : IEEE Press, 2010:2366-2369.
- [32] WANG Z, 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):600-612.

编辑 金胡考