

基于多特征融合的图片修复算法^{*}蒋行国^{1,2} 黎明¹

(1. 四川轻化工大学自动化与信息工程学院 宜宾 644000; 2. 人工智能四川省重点实验室 宜宾 644000)

摘要: 针对现有图片修复算法修复结果存在结构一致性差和纹理细节不足等问题,在生成对抗网络(GAN)的框架下,提出一种基于多特征融合的图片修复算法。首先,采用双编码-解码结构提取纹理和结构特征信息,并引入快速傅里叶卷积残差块,有效捕获全局上下文特征。然后,通过注意力特征融合(AFF)模块完成结构与纹理特征之间的信息交换,提高图像的全局一致性。并利用密集连接特征聚合(DCFA)模块在多个尺度上提取丰富的语义特征,进一步提升修复图像的一致性和准确性,以呈现更精细的内容。实验结果表明,在破损区域占比为40%~50%时,相较于最优对比算法,所提算法在 CelebA-HQ 数据集上 PSNR 和 SSIM 分别提高 1.18% 和 0.70%,FID 降低 3.99%。在 Paris StreetView 数据集上 PSNR 和 SSIM 分别提高 1.17% 和 0.50%,FID 降低 2.29%。实验证明所提算法能有效修复大面积破损图像,修复结果具有更合理的结构和丰富的纹理细节。

关键词: 图像修复;生成对抗网络;多特征融合;快速傅里叶卷积残差块

中图分类号: TP391.4;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Image inpainting algorithm based on multi-feature fusion

Jiang Xingguo^{1,2} Li Ming¹

(1. School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;

2. Artificial Intelligence Key Laboratory of Sichuan Province, Yibin 644000, China)

Abstract: Aiming at the problems of poor structural consistency and insufficient texture details in the inpainting results of existing image inpainting algorithms, an image inpainting algorithm based on multi-feature fusion was proposed under the framework of generative adversarial network (GAN). Firstly, the dual encoder-decoder structure is used to extract the texture and structure feature information, and the fast Fourier convolution residual block is introduced to effectively capture the global context features. Then, the information exchange between structure and texture features was completed through the attention feature fusion (AFF) module to improve the global consistency of the image. The dense connected feature aggregation (DCFA) module was used to extract rich semantic features at multiple scales to further improve the consistency and accuracy of the inpainted image, so as to present more detailed content. Experimental results show that, compared with the optimal comparison method, the proposed algorithm improves PSNR and SSIM by 1.18% and 0.70% respectively, and reduces FID by 3.99% on the CelebA-HQ dataset when the proportion of damaged regions is 40%~50%. On the Paris Street View dataset, PSNR and SSIM are increased by 1.17% and 0.50%, respectively, and FID is reduced by 2.29%. Experimentally, it is proved that the suggested algorithm can effectively repair large broken images, and the repaired images have a more sensible structure and richer texture details.

Keywords: image inpainting; generative adversarial network; fast Fourier convolution residual block; multi-feature fusion

0 引言

图像修复是指对缺损或损坏的图像区域填充像素信

息,使得恢复出来的图像具有语义合理且视觉上真实的效果^[1],是计算机视觉领域中的重要研究方向。在实际生活中,被广泛应用于文物复原^[2]、目标物移除^[3]、老照片修

复^[4]等诸多领域,具有广阔的应用价值和商业价值。

传统的图像修复方法利用局部特征作为主要的修复依据,缺乏全局上下文信息,无法识别和重建全局结构,导致边界不自然,这种方法对于较大区域缺损时无法捕捉图像的全局语义结构,修复结果与原始图像存在明显差异,往往会比较模糊。近年来,深度学习被广泛用于解决图像修复问题,并取得令人满意的结果,逐渐成为图像修复领域的主流方法。与传统方法相比,基于深度学习的图像修复方法能理解更深层次的语义特征,以获得具有语义一致的修复结果。Liu等^[5]提出部分卷积代替普通卷积,使用自动更新的二进制掩码识别出缺失区域的位置,可以很好的处理多种形状的缺失,但随着缺损区域的增大,会出现结构扭曲,纹理细节丢失的现象。Zheng等^[6]提出了一种双路径组成的多元图像补全方法,并通过长短注意力机制来学习不同样式的全局和局部特征,但当图像存在多处不连续缺失区域时会出现结构一致性差等问题。Zeng等^[7]提出一种金字塔上下文编码器网络的图像修复算法,可以从多个尺度学习图像的语义信息,但修复图像仍存在结构扭曲和纹理出现伪影等缺陷。

现有的图像修复算法已经取得了不错的修复效果,能够获取到一些图像深层语义信息,但在修复缺损区域面积大,结构纹理复杂的图像时,由于缺乏结构与纹理先验知识的引导,难以提取更多的细节特征,从而导致生成图像纹理模糊、结构扭曲。针对上述问题,本文提出了一种基于多特征融合的图像修复算法,该算法采用双生成器结构实现纹理信息和结构信息的生成,并引入快速傅里叶卷积残差块捕获更多的全局上下文特征,增强网络的细节修复能力。在特征融合阶段设计了注意力特征融合(attention feature fusion, AFF)模块,增强纹理和结构的融合,提高全局图像的一致性,从而使得修复图像同时具有纹理和结构的语义。并设计了密集连接特征聚合(densely connected feature aggregation, DCFA)模块,以充分挖掘图像高级语义特征间上下文关系,增强图像局部之间的相关性,生成具有特征连续、语义一致的修复图像。

1 相关工作

1.1 传统的图像修复方法

传统的图像修复方法可以分为基于扩散的方法^[8-9]和基于补丁的方法^[10-12]。基于扩散的方法是利用图像缺失周边的已知像素信息,根据原始图像的特征向内扩散到缺损区域进行修复。Bertalmio等^[8]提出基于偏微分方程的图像修复算法,该算法沿着等照度线将上下文信息从边界连续传播到图像缺失区域,修复的区域会自动填充,但该算法在处理较大缺失区域时无法修复其内部。Ballester等^[9]提出了一个基于图像灰度和梯度方向的联合插值扩散,以自动方式将等值线平滑延伸到缺失数据的孔洞中,是一个变分问题。这种基于扩散的图像修复方法仅适用于小面积区

域缺损,当出现较大的区域缺损或已知区域与缺失区域的相似信息较少时,无法准确地恢复缺损的细节信息。

基于补丁的方法是从图像上搜索与缺失区域相似的补丁块来填充缺失区域的像素值。Criminisi等^[10]提出了一种基于块的纹理合成算法,其通过从完好区域搜寻与破损区域相似度最高的图像块,将纹理和结构信息传播到图像的缺失区域来实现修复。Barnes等^[11]提出一种随机最近邻补丁匹配算法,通过采样查找图像块之间估计的最近邻匹配,然后利用图像块之间的连贯性将匹配点传播到待修复区域进行修复。Huang等^[12]提出了一种自动指导增强补丁匹配的补全算法,通过将补丁偏移和变化的先验概率转换为补全算法的拓展补丁距离,以及采样方案中整合平移规律约束完成自动补全。这种基于补丁的图像修复方法由于缺乏对图像的高级理解,对于修复结构纹理复杂的图像产生纹理模糊。因此,传统的图像修复方法难以获取到图像更深层的语义特征,具有较大的局限性。

1.2 基于深度学习的图像修复方法

近年来,随着深度学习在计算机视觉领域的发展,深度学习技术在图像修复领域取得了一系列卓越的成果。例如,Pathak等^[13]提出的上下文编码器(context encoders, CE)图像修复方法,使用了编码器-解码器的架构,利用图像缺损区域附近的像素信息来对未缺损区域的信息进行补全,完成图像的精细修复。Iizuka等^[14]使用全局和局部判别器的修复方法,分别关注修复图像的整体结构信息与纹理细节信息,保证了整体语义连贯性和生成细节的完整性。Yu等^[15]提出粗略到细化的两阶段生成式修复网络结构,在细修复网络中加入上下文注意力模块,增强图像局部和全局结构的一致性。Liu等^[5]提出部分卷积,对不规则掩码修复有一定效果。Yu等^[16]针对部分卷积在更新不规则掩码过程中的缺陷进行改进,提出了门控卷积,进一步提高网络对细节的修复能力。Liu等^[17]提出了一种结合结构和纹理的相互编码-解码器(mutual encoder-decoder, MED)图像修复网络,使用特征均衡模块对结构和纹理特征进行融合,获得更好的修复结果。Li等^[18]提出递归特征推理模块,可以反复推断缺损边界的特征映射,将其作为推理依据从缺损区域由外向内补全特征信息。Guo等^[19]提出了一种纹理与结构双生成的图像修复方法,实现结构与纹理特征信息的整合,从而提高生成图像的质量。Liu等^[20]提出空间概率多样性归一化(spatially probabilistic diversity normalization, SPDNorm),可以在缺失区域生成逼真的视觉效果。孙琪等^[21]提出部分卷积和多尺度融合的图像修复方法,融合深层和浅层的图像特征对人脸图像进行修复。王山豹等^[22]提出一种多模态注意力机制生成网络的图像修复算法,通过多尺度特征融合得到精细的修复结果。Chen等^[23]提出了一种残差特征注意网络(residual feature attention network, RFA-Net),通过残差注意力模块自适应学习关键特征,有效解决了修复结果中纹理模糊和边缘不

清晰等问题。

2 基于多特征融合图像修复算法

本文提出的基于多特征融合的图像修复算法总体框架如图 1 所示,采用生成对抗网络^[24]为基础模型。首先,使用生成器网络中的双编码-解码结构获取图像的结构和纹

理特征信息,并引入快速傅里叶卷积残差块,提高对全局关键信息的提取能力。然后,在通过结构解码器和纹理解码器重构图像的结构和纹理特征,将输出的结构特征与纹理特征经过注意力特征融合模块和密集连接特征聚合模块组成的特征融合网络进行更深层次的融合,保证了生成图像的一致性。

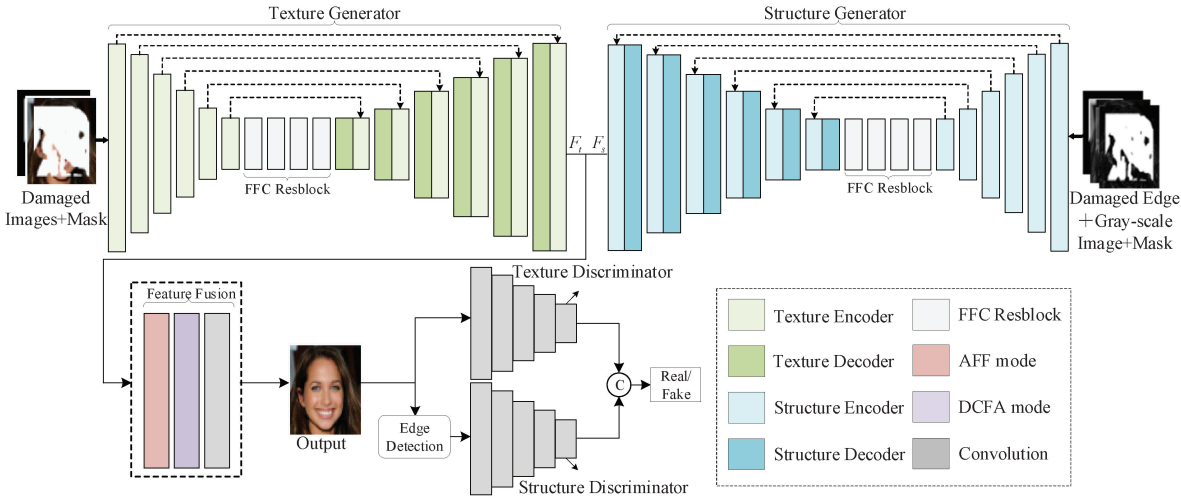


图 1 整体网络框架

Fig. 1 Overall network framework

2.1 生成器

生成器对图像的结构和纹理特征信息分别进行编码操作和解码操作。在编码阶段,纹理编码器的输入通道为 2,通过输入破损图像和掩码图像来进行纹理信息的编码。结构编码器输入通道为 3,则通过输入破损边缘图像、破损灰度图像以及掩码图像来进行结构信息的编码。在解码阶段,双解码器对结构和纹理特征进行解码,并利用跳跃连接技术指导解码操作,将编码过程中的信息连接到解码过程中,可以在多个尺度上更好地将低级和高级特征组合在一起,能够更好地还原出缺失区域。在生成器网络中,为了更有效地从边界捕捉信息,使用部分卷积替代原有的普通卷积。与普通卷积相比,部分卷积操作只针对已知区域的像素,更有利于不规则破损图像修复。

2.2 快速傅里叶卷积残差块

在生成器网络中使用了快速傅里叶卷积残差块,使得网络在早期层捕获尽可能多的全局上下文信息,同时残差结构可以避免深度网络中的梯度消失和爆炸问题。快速傅里叶卷积(fast fourier convolution, FFC)^[25]基于通道对图像特征进行快速傅里叶变换。如图 2 所示,FFC 将输入特征按通道分为局部分支和全局分支,局部分支使用核大小为 3×3 的常规卷积进行特征图的局部更新。全局分支不仅会用 3×3 的卷积核对信息进行卷积处理,还通过快速傅里叶变换获取频域信息,关联全局上下文信息。最终由局部分支和全局分支的输出进行交叉融合得到最终输出结果。

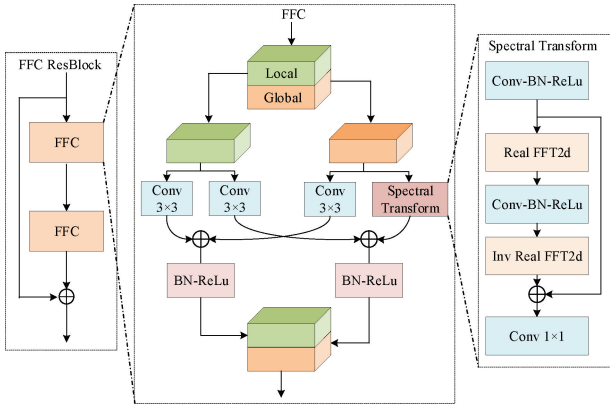


图 2 快速傅里叶卷积残差块结构

Fig. 2 Fast fourier convolution residual block structure

2.3 注意力特征融合模块

为了将双解码器生成的纹理和结构特征进行信息交换与融合,设计了注意力特征融合模块。结构如图 3 所示,在其中嵌入了 ECA 注意力^[26],通过考虑每个通道内的全局信息关系,为每个通道分配不同的权重,有效地捕捉不同通道之间的关系,从而进行特征通道信息的最优选择。ECA 注意力首先对输入特征进行全局平均池化操作获取每个通道的全局信息,然后使用卷积核大小为 k 且权重共享的一维卷积进行卷积操作。最后,在经过 Sigmoid 函数生成的权重与输入特征的对元素相乘得到输出特征。该模块还采用了软门控控制集成程度,使得输出的结

果同时具有结构与纹理的语义。模块由两条路径组成,上面一条路径构建结构感知的纹理特征,将纹理解码器与结构解码器生成的纹理特征与结构特征分别表示为 F_t 与 F_s , 控制纹理信息整合的软门控 G_s 表示为:

$$G_s = \sigma(\text{ECA}(h(\text{Concat}(F_s, F_t)))) \quad (1)$$

式中: $\sigma(\cdot)$ 表示 sigmoid 激活函数, $h(\cdot)$ 表示核大小为 3×3 卷积实现的映射函数, $\text{ECA}(\cdot)$ 表示经过 ECA 注意力捕捉重要通道信息的操作, $\text{Concat}(\cdot)$ 表示通道的连接操作。

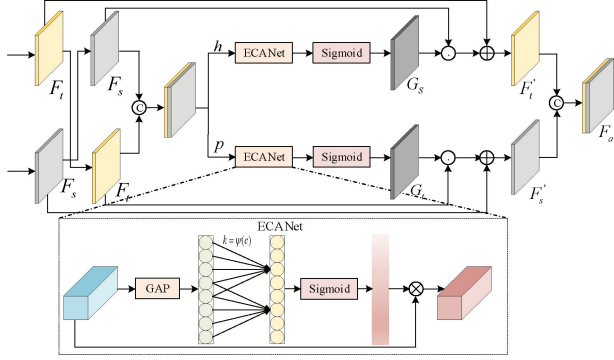


图 3 注意力特征融合模块结构

Fig. 3 Attention feature fusion module structure

通过 G_s 自适应的将 F_s 动态合并到 F_t 中, 可以表示为:

$$F_t' = \alpha(G_s \odot F_s) \oplus F_t \quad (2)$$

式中: α 表示被初始化为 0 的训练参数, \odot 表示元素乘法操作, \oplus 表示元素加法操作。

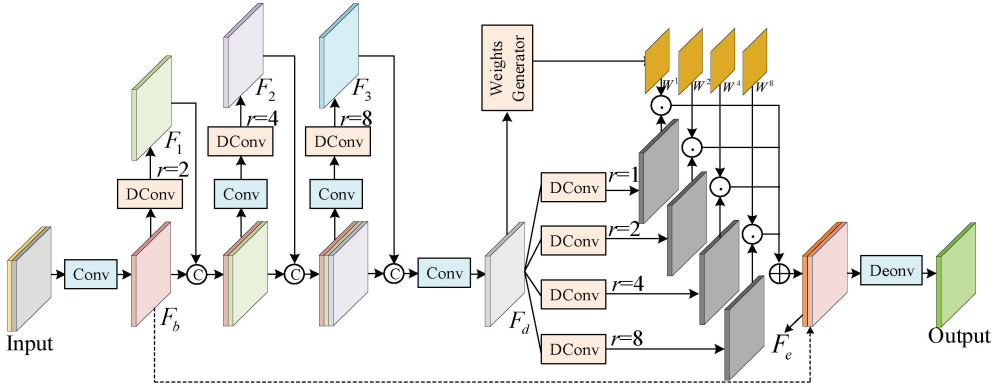


图 4 密集连接特征聚合模块结构

Fig. 4 Densely connected feature aggregation module structure

然后, 为了进一步捕捉多尺度的语义特征, 采用了 4 组扩展率不同的扩展卷积重构特征图, 计算公式如下:

$$F_d^k = \text{Conv}_k(F_d) \quad (10)$$

式中: Conv_k 表示扩展率为 k , $k \in \{1, 2, 4, 8\}$ 的扩展卷积层。

然后将捕捉到的语义特征送入像素级权重生成器 P_w

同理, 下面一条路径构建纹理感知的结构特征 F_s' , 表示为:

$$G_t = \sigma(\text{ECA}(p(\text{Concat}(F_s, F_t)))) \quad (3)$$

$$F_s' = \beta(G_t \odot F_t) \oplus F_s \quad (4)$$

式中: p 表示核大小为 3×3 卷积实现的映射函数, β 表示被初始化为零的训练参数。

最后, 将两条路径输出的结构感知的纹理特征 F_t' 和纹理感知的结构特征 F_s' 通过通道连接进行特征融合得到最后的特征图 F_o , 可以表示为:

$$F_o = \text{Concat}(F_t', F_s') \quad (5)$$

2.4 密集连接特征聚合模块

为了增强图像不同层次特征之间的相关性, 确保图像的整体一致性, 设计了密集连接特征聚合模块。模块结构如图 4 所示, 该模块先经过一个密集连接扩展卷积操作, 捕捉远程依赖关系, 从而更好的有效的捕获到全局与局部特征。并采用多尺度特征聚合在多个尺度上对丰富的语义特征进行编码, 使得不同尺度的特征相互补充, 从而应对更具挑战的情况。首先, 输入特征分别经过 3 组扩张率不同的扩展卷积得到不同的尺度特征图, 并依次进行通道连接操作整合到一起。具体计算公式如下:

$$F_1 = \text{Conv}_{k_1}(F_b) \quad (6)$$

$$F_2 = \text{Conv}_{k_2}(\text{Conv}(\text{Concat}(F_1, F_b))) \quad (7)$$

$$F_3 = \text{Conv}_{k_3}(\text{Conv}(\text{Concat}(F_1, F_2, F_b))) \quad (8)$$

$$F_d = \text{Conv}(\text{Concat}(F_1, F_2, F_3, F_b)) \quad (9)$$

式中: $F_i (i = 1, 2, 3)$ 是中间变量, $\text{Conv}_{k_1}, \text{Conv}_{k_2}, \text{Conv}_{k_3}$ 分别表示扩展率为 2, 4, 8 的扩展卷积层, $\text{Conv}(\cdot)$ 表示普通的 1×1 卷积, $\text{Concat}(\cdot)$ 表示通道的连接操作。

中, P_w 可以预测像素级的权重图, 能够更好地聚合多尺度语义特征。 P_w 由一系列卷积操作与激活函数组成。具体来说, P_w 有两个普通卷积层, 卷积核大小分别为 3 和 1, 并且在这两个卷积操作后都采用非线性 Relu 激活函数。像素级权重图计算公式如下:

$$W^1, W^2, W^4, W^8 = \text{Slice}(\text{Softmax}(P_w(F_d))) \quad (11)$$

式中: $\text{Softmax}()$ 和 $\text{Slice}()$ 分别表示在通道维度上的 softmax 操作和 Slice 操作。

最后将多尺度的语义特征 F_d^k 与像素权重图 W^k 进行逐元素相乘,然后将不同尺度上的特征图进行加权求和,生成更精细特征图 F_e :

$$F_e = (F_d^1 \odot W^1) \oplus (F_d^2 \odot W^2) \oplus (F_d^4 \odot W^4) \oplus (F_d^8 \odot W^8) \quad (12)$$

2.5 判别器

判别器是对生成器生成的修复图像与真实图像进行判别,并通过不断学习优化模型参数来提升对修复图像的判别能力,通过生成器和判别器相互对抗提升两则的表征能力,解决修复图像存在伪影模糊的问题,确保破损区域与非破损区域内纹理结构的一致性。判别器包括纹理判别器与结构判别器,两个判别器的结构相同,都由一系列卷积操作和激活函数组成。具体来说,包括了 3 个步长为 2 与两个步长为 1 的普通卷积层组成,卷积核大小均为 4,在其中最后一层卷积层使用 Sigmoid 激活函数将输出映射到 0~1,在其余四层卷积层都使用了具有线性修正单元 LeakyReLU 激活函数,斜率设置为 0.2。不同于纹理判别器直接将生成图像与真实图像作为输入,结构判别器输入真实图像的边缘图通过残差块进行检测,并且将灰度图像与边缘图一起输入到结构判别器中,对结构判别器的对抗损失进行优化,这样可以在修复过程中确保全局结构信息的一致性。为防止修复网络在训练过程中判别器过早拟合导致生成器无法继续学习的问题,在判别器网络中采用谱归一化技术^[27],使判别器网络满足 Lipschitz 连续性,改善网络训练不稳定的问题。

2.6 损失函数

为了保证生成图像的语义一致性和视觉保真度,采用多种损失函数相结合进行训练,优化图像修复之后的结果。总的损失为:

$$L_{\text{joint}} = \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{perc}} L_{\text{perc}} + \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{style}} L_{\text{style}} + \lambda_{\text{inter}} L_{\text{inter}} \quad (13)$$

式中: L_{rec} 为重建损失, L_{perc} 为感知损失, L_{adv} 为对抗损失, L_{style} 为风格损失, L_{inter} 为中间损失, $\lambda_{\text{rec}}, \lambda_{\text{perc}}, \lambda_{\text{style}}, \lambda_{\text{adv}}, \lambda_{\text{inter}}$ 分别为各损失的权重参数。

1) 重建损失

为了保证图像重建前后的一致性,引入了重建损失 L_{rec} ,用图像的修复结果 I_{out} 和图像的真实值 I_{gt} 之间的 L_1 距离来衡量两幅图像之间的像素差异,其计算过程如公式所示:

$$L_{\text{rec}} = E[\|I_{\text{out}} - I_{\text{gt}}\|_1] \quad (14)$$

2) 感知损失

由于仅仅从像素层面进行优化约束难以保证高层语义信息的完整性,引入了感知损失 L_{perc} 约束深层结构的一致性。感知损失借助预训练的 VGG 网络进行计算,公式如下:

$$L_{\text{perc}} = E\left[\sum_i \|\varphi_i(I_{\text{out}}) - \varphi_i(I_{\text{gt}})\|_1\right] \quad (15)$$

式中: $\varphi_i(\cdot)$ 为预训练 VGG 网络第 i 层输出的特征图。

3) 风格损失

引入了风格损失 L_{style} 可以使修复区域与周围环境在风格上融合,提高修复结果的自然度。 L_{style} 定义如下:

$$L_{\text{style}} = E\left[\sum_i \|\psi_i(I_{\text{out}}) - \psi_i(I_{\text{gt}})\|_1\right] \quad (16)$$

式中: $\psi_i(\cdot) = \varphi_i(\cdot)^T \varphi_i(\cdot)$ 表示由特征图 φ_i 构造的 Gram 矩阵。

4) 对抗损失

对抗损失 L_{adv} 对生成器的训练过程进行约束,可以确保生成图像的视觉真实性及纹理结构的一致性,定义为:

$$L_{\text{adv}} = \min_{G_D} \max_{E_{I_{\text{gt}}}, E_{E_{\text{gt}}}} [\log D(I_{\text{gt}}, E_{\text{gt}})] + E_{I_{\text{out}}}, E_{\text{out}} [\log(1 - D(I_{\text{out}}, E_{\text{out}}))] \quad (17)$$

式中: E_{gt} 为完整的边缘图, E_{out} 为生成的边缘图。

5) 中间损失

为了提高生成器中的纹理解码器和结构解码器捕捉纹理和结构特征的能力,在 F_s 和 F_t 上引入了中间监督损失 L_{inter} , 定义为:

$$L_{\text{inter}} = L_{\text{structure}} + L_{\text{texture}} = BCE(E_{\text{gt}}, p_s(F_s)) + L_1(I_{\text{gt}}, P_t(F_t)) \quad (18)$$

式中: P_t 与 P_s 为残差块和卷积得到的投影函数。

3 实验结果与分析

3.1 数据集与实验设置

在公共数据集 CelebA-HQ^[28] 与 Paris StreetView^[29] 验证所提算法的有效性。CelebA-HQ 数据集包含了 30 000 幅图像,实验选取 28 000 幅图像作为训练集,剩下的 2 000 幅图像作为测试集。Paris StreetView 数据集包含了 14 900 幅训练图像和 100 幅测试图像。不规则掩膜图像来自 Liu 等^[6]提出的随机掩码数据集,按照掩码大小分为 4 个区间(10%~20%, 20%~30%, 30%~40%, 40%~50%)。在训练过程中,所有数据集图像的大小调整为 256×256。

实验运行环境为 Windows 10(64 bit)操作系统,硬件配置为 Intel(R) Core(TM) i5-12400F,内存为 16 GB,显卡为 NVIDIA GeForce RTX 3060,显存为 12 GB。深度学习框架采用 pytorch1.7.0。在训练中,批次大小设置为 4,采用 Adam 优化器,网络学习率在初始训练阶段设置为 2×10^{-4} ,当趋于收敛时使用 5×10^{-5} 的学习率为进行微调,判别器的学习率设置为 2×10^{-5} 。损失函数各部分权重参数分别设置为 $\lambda_{\text{rec}}=10$, $\lambda_{\text{perc}}=0.1$, $\lambda_{\text{adv}}=0.1$, $\lambda_{\text{style}}=250$ 和 $\lambda_{\text{inter}}=1$ 。

3.2 定性分析

为了验证所提算法的优越性,选取了图像修复领域中几个主流算法进行定性对比分析实验,对比算法分别是 PRVS^[30], RFR^[18], ICT^[31], CTSDG^[19] 与 AOT-GAN^[32]。图 5 与图 6 展示了不同方法在 CelebA-HQ 与 Paris

StreetView 数据集上的定性比较结果,图中最后两行为放大的局部细节展示。从图 5 中可看出,PRVS 算法生成的人脸具有不连续现象,修复结果存在结构扭曲。RFR 算法在小区缺损表现良好,但在大区域缺损时修复结果会产生模糊现象。ICT 算法生成的人脸图像存在边界伪影和局部色差的问题。CTSDG 算法与 AOT-GAN 算法修复的面部轮廓更加合理,具有一定的结构性,但在眼睛、嘴巴等细节特征不明显。与不同方法相比,所提方法生成的人脸图像具有合理的语义和纹理结构信息,可以完整地恢复出面

部特征,在视觉效果上保证了结构的连贯性。从图 6 可以看出,PRVS 算法生成的图像具有明显的伪影且生成了多余信息。RFR 算法无法完成合理的修复,生成的图像存在语义不合理、视觉上不真实。ICT 算法生成的图像存在局部纹理模糊和伪影。CTSDG 算法生成的图像在修复区域修复的纹理细节不够精细。AOT-GAN 算法修复结果出现像素模糊和修复区域存在不连贯的现象。与不同方法相比,所提方法生成的图像具有合理的内容和更清晰的纹理,在细节方面与原图有较高的一致性。

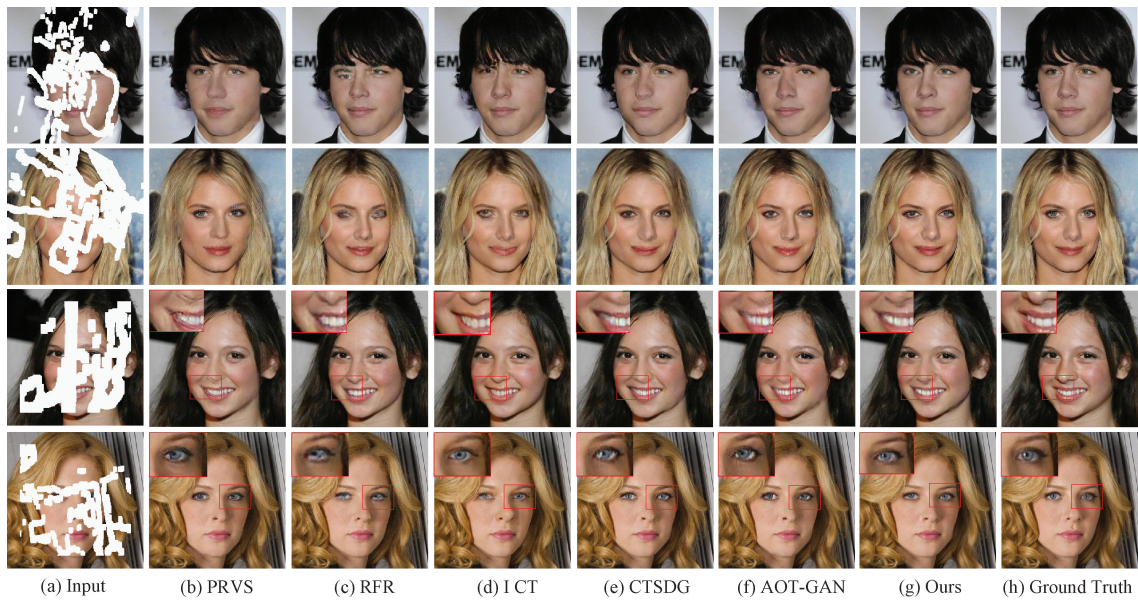


图 5 不同方法在 CelebA-HQ 数据集上的定性效果对比

Fig. 5 Qualitative effect comparison of different methods on CelebA-HQ dataset

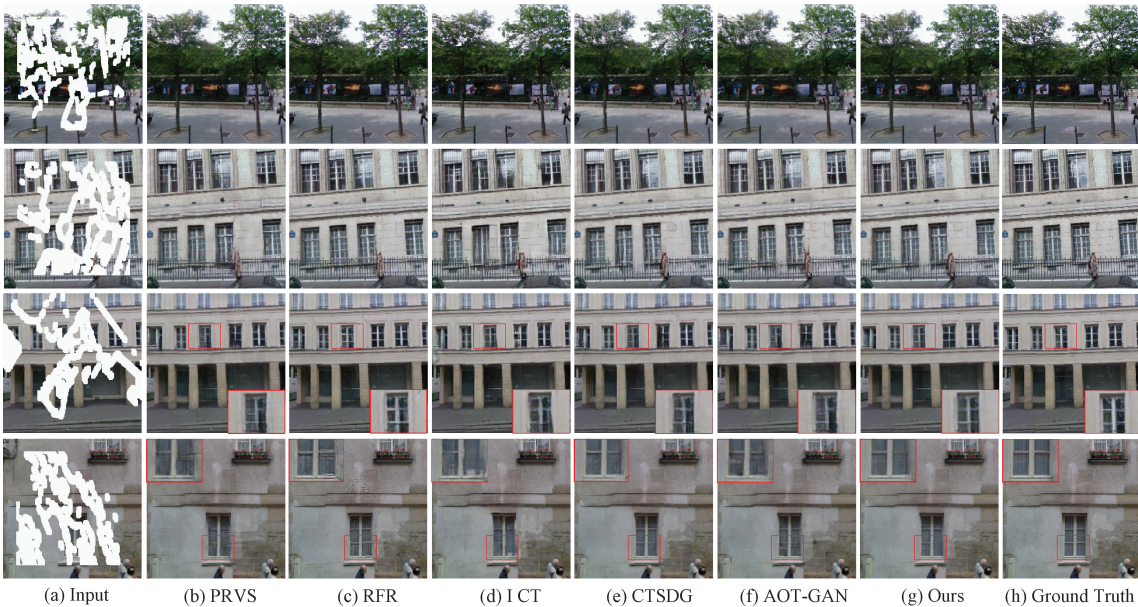


图 6 不同方法在 Paris StreetView 数据集上的定性效果对比

Fig. 6 Qualitative effect comparison of different methods on the Paris StreetView dataset

3.3 定量分析

为了客观评价所提算法的有效性,采用峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity, SSIM) 以及距离得分 (frechet inception distance, FID) 进行定量分析。其中 PSNR 定义重建图像和原始图像的失真程度,SSIM 通过估计图像的亮度、对比度和结构特征,衡量图像之间的相似度,PSNR 和 SSIM 值越高表示图像修复的效果越好。FID 是计算真实图像和修复图像之间距离的性能指标,考虑了图像的多样性、结构和语义方面的差异,FID 值越低,表示图像修复效果越好。

表 1、2 为不同算法在 CelebA-HQ 数据集与 Paris StreetView 数据集上的定量实验对比。从测试结果来看,

与最优的对比方法相比,当使用小掩膜(10%~20%掩膜率)遮挡时,在 CelebA-HQ 数据集上 PSNR 和 SSIM 分别提高 0.43% 和 0.42%,FID 降低 5.25%。在 Paris StreetView 数据集上 PSNR 降低 0.37%,SSIM 提高 0.32%,FID 降低 4.34%。当使用大掩膜(40%~50%掩膜率)遮挡时,在 CelebA-HQ 数据集上 PSNR 和 SSIM 分别提高 1.18% 和 0.70%,FID 降低 3.99%。在 Paris Street View 数据集上 PSNR 和 SSIM 分别提高 1.17% 和 0.50%,FID 降低 2.29%。所提算法在 CelebA-HQ 数据集上的各项指标相较于对比方法均取得最优的结果,在 Paris StreetView 数据集上虽然没有在每项指标中取得最优的结果,但从整体数据来看存在明显的优势。证明所提算法修复结果更接近于真实值,修复效果最优。

表 1 不同方法在 CelebA-HQ 数据集上的定量比较结果							
Table 1 Quantitative comparison results of different methods on CelebA-HQ dataset							
评价指标	掩膜率/%	PRVS	RFR	ICT	CTSDG	AOT-GAN	Ours
PSNR ↑	10~20	30.325	31.234	32.642	33.142	31.925	33.284
	20~30	27.569	28.735	28.935	29.213	28.856	29.624
	30~40	25.158	26.167	27.286	27.453	26.763	27.837
	40~50	22.318	23.057	25.034	25.135	23.864	25.432
SSIM ↑	10~20	0.926	0.958	0.957	0.961	0.956	0.965
	20~30	0.895	0.912	0.919	0.928	0.917	0.935
	30~40	0.817	0.842	0.881	0.885	0.868	0.896
	40~50	0.763	0.808	0.849	0.852	0.832	0.858
FID ↓	10~20	15.62	8.43	3.84	3.43	4.89	3.25
	20~30	19.48	12.58	7.69	4.73	9.55	4.42
	30~40	26.32	16.40	10.28	7.58	10.72	6.75
	40~50	37.21	19.83	14.73	9.77	15.89	9.38

表 2 不同方法在 Paris StreetView 数据集上的定量比较结果							
Table 2 Quantitative comparison results of different methods on the Paris StreetView dataset							
评价指标	掩膜率/%	PRVS	RFR	ICT	CTSDG	AOT-GAN	Ours
PSNR ↑	10~20	29.259	30.532	30.372	30.879	29.826	30.765
	20~30	26.419	27.935	27.658	28.124	27.748	28.562
	30~40	24.325	24.453	24.274	26.531	25.153	26.841
	40~50	22.142	22.684	22.372	23.251	22.738	23.523
SSIM ↑	10~20	0.873	0.886	0.884	0.933	0.882	0.936
	20~30	0.855	0.865	0.861	0.896	0.858	0.905
	30~40	0.794	0.817	0.819	0.857	0.823	0.854
	40~50	0.738	0.788	0.791	0.803	0.776	0.807
FID ↓	10~20	16.41	8.64	7.53	5.53	6.26	5.29
	20~30	18.29	10.53	10.62	7.23	11.28	6.58
	30~40	28.52	18.51	16.77	9.87	16.43	9.52
	40~50	41.42	22.24	19.49	12.64	23.59	12.35

3.4 消融实验

实验在 CelebA-HQ 数据集进行消融实验,在破损率为 40%~50%的随机掩码图像上进行测试,来分析所提算法各个模块的有效性,进行了 3 组实验。w/o FFC 表示去掉网络中的快速傅里叶卷积残差块,w/o AFF 表示去掉网络中的注意力特征融合模块,w/o DCFA 表示去掉网络中的密集连接特征聚合模块。图 7 展示了消融实验的定性结果对比。从图中可以看出,缺少任何一个模块的修复结果视觉质量都会出现一定程度的下降,通过比较所提算法

修复的图像更加真实,有效缓解了修复区域出现纹理细节缺失,全局一致性更加连贯。表 3 为消融实验定量对比。从表中结果可以看出,相较于所提算法,无 FFC 残差块网络的 PSNR 降低 0.96%,SSIM 降低 0.47%,FID 提高 3.84%;无 AFF 模块网络的 PSNR 降低 2.30%,SSIM 降低 1.05%,FID 提高 9.49%;无 DCAF 模块网络的 PSNR 降低 2.95%,SSIM 降低 1.40%,FID 提高 13.54%。缺少任何一个模块后,PSNR 与 SSIM 都有所降低,FID 值提高,通过比较验证了模型各个模块的有效性和必要性。

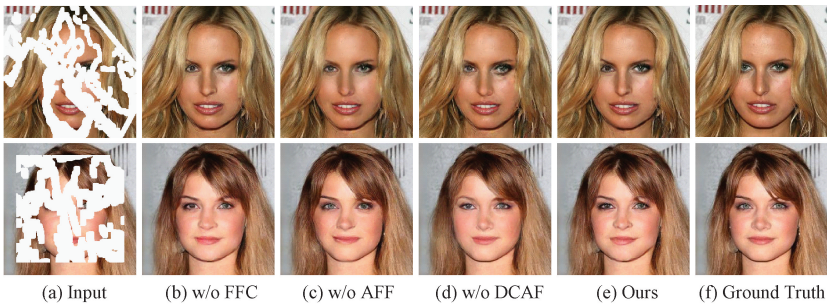


图 7 消融实验定性比较结果

Fig.7 Qualitative comparison results of ablation experiments

表 3 消融实验定量比较结果

Table 3 Quantitative comparison results of ablation experiments

方法	PSNR ↑	SSIM ↑	FID ↓
w/o FFC	25.187	0.854	9.74
w/o AFF	24.846	0.849	10.27
w/o DCFA	24.683	0.846	10.65
本文	25.432	0.858	9.38

4 结 论

本文提出了一种基于多特征融合的图像修复算法。该算法通过双编码-解码结构获取图像的纹理与结构特征信息,并引入了快速傅里叶卷积残差块有效捕捉全局上下文特征。构建了注意力特征融合模块与密集连接特征聚合模块增强图像结构和纹理的一致性。在 CelebA-HQ 数据集和 Paris StreetView 数据集上进行了广泛的实验。定性分析表明,所提算法修复的图像具有更好的视觉效果。定量分析表明,所提算法对比算法具有更好的性能。

参考文献

[1] 童俊毅,张银胜,张培琰,等. 基于双阶段多尺度生成对抗网络的图像复原方法[J]. 国外电子测量技术,2024, 43(6):50-58.
TONG J Y, ZHANG Y SH, ZHANG P Y, et al. Image restoration method based on two-stage multi-scale generative adversarial network [J]. Foreign Electronic Measurement Technology, 2024, 43(6): 50-58.

[2] HOU M L, YANG S, HU Y G, et al. Novel method for virtual restoration of cultural relics with complex geometric structure based on multiscale spatial geometry [J]. ISPRS International Journal of Geo-Information, 2018, 7(9): 353.
[3] 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: 7508-7517.
[4] 高伟,吴顺. 基于多尺度注意力半监督学习的老照片划痕修复[J]. 计算机工程,2022,48(10):245-251,261.
GAO W, WU SH. Scratch repairing of old photos based on multi-scale attention semi-supervised learning [J]. Computer Engineering, 2022, 48(10): 245-251,261.
[5] LIU G L, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C]. European Conference on Computer Vision (ECCV), 2018: 85-100.
[6] ZHENG CH X, CHAM T J, CAI J F. Pluralistic image completion [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019: 1438-1447.
[7] ZENG Y H, FU J L, CHAO H Y, et al. Learning pyramid-context encoder network for high-quality image inpainting [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019: 1486-1494.
[8] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting [C]. 27th Annual Conference on Computer Graphics and Interactive Techniques,2000: 417-424.

- [9] BALLESTER C, BERTALMIO M, CASELLES V, et al. Filling-in by joint interpolation of vector fields and gray levels[J]. IEEE Transactions on Image Processing, 2001, 10(8): 1200-1211.
- [10] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting [J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.
- [11] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing [J]. ACM Transactions on Graphics (TOG), 2009, 28(3): 24.
- [12] HUANG J B, KANG S B, AHUJA N, et al. Image completion using planar structure guidance[J]. ACM Transactions on Graphics(TOG), 2014, 33(4): 1-10.
- [13] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2536-2544.
- [14] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics(ToG), 2017, 36(4): 1-14.
- [15] YU J H, LIN ZH, YANG J M, et al. Generative image inpainting with contextual attention[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5505-5514.
- [16] YU J H, LIN ZH, YANG J M, et al. Free-form image inpainting with gated convolution[C]. IEEE/CVF International Conference on Computer Vision, 2019: 4471-4480.
- [17] LIU H Y, JIANG B, SONG Y B, et al. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations [C]. Computer Vision-ECCV 2020: 16th European Conference, Springer International Publishing, 2020: 725-741.
- [18] LI J Y, WANG N, ZHANG L F, et al. Recurrent feature reasoning for image inpainting[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7760-7768.
- [19] GUO X F, YANG H Y, HUANG D. Image inpainting via conditional texture and structure dual generation[C]. IEEE/CVF International Conference on Computer Vision, 2021: 14134-14143.
- [20] LIU H Y, WAN Z Y, HUANG W, et al. PD-GAN: Probabilistic diverse gan for image inpainting [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 9371-9381.
- [21] 孙琪, 翟锐, 左方, 等. 基于部分卷积和多尺度特征融合的人脸图像修复模型[J]. 计算机工程与科学, 2023, 45(2): 304-312.
SUN Q, ZHAI R, ZUO F, et al. Facial image inpainting based on partial convolution and multi-scale feature integration [J]. Computer Engineering and Science, 2023, 45(2): 304-312.
- [22] 王山豹, 梁栋, 沈玲. 利用多模态注意力机制生成网络的图像修复[J]. 计算机辅助设计与图形学学报, 2023, 35(7): 1109-1121.
WANG SH B, LIANG D, SHEN L. Image inpainting with multi-modal attention mechanism generative network[J]. Journal of Computer-Aided Design and Computer Graphics, 2023, 35(7): 1109-1121.
- [23] CHEN M, ZANG SH R, AI ZH H, et al. RFA-Net: Residual feature attention network for fine-grained image inpainting [J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105814.
- [24] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63(11): 139-144.
- [25] CHI L, JIANG B R, MU Y D. Fast Fourier convolution [J]. Advances in Neural Information Processing Systems, 2020, 33: 4479-4488.
- [26] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [27] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [J]. ArXiv preprint arXiv: 1802.05957, 2018.
- [28] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation [J]. ArXiv preprint arXiv: 1710.10196, 2017.
- [29] DOERSCH C, SINGH S, GUPTA A, et al. What makes paris look like paris[J]. Communications of the ACM, 2015, 58(12): 103-110.
- [30] LI J Y, HE F X, ZHANG L F, et al. Progressive reconstruction of visual structure for image inpainting[C]. IEEE/CVF International Conference on Computer Vision, 2019: 5962-5971.
- [31] WAN Z Y, ZHANG J B, CHEN D D, et al. High-fidelity pluralistic image completion with transformers[C]. IEEE/CVF International Conference on Computer Vision, 2021: 4692-4701.
- [32] ZENG Y H, FU J L, CHAO H Y, et al. Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 29(7): 3266-3280.

作者简介

蒋行国(通信作者), 博士, 副教授, 主要研究方向为智能信息处理。

E-mail: tonny_jiang@suse.edu.cn

黎明, 硕士研究生, 主要研究方向为计算机视觉和图像处理。

E-mail: 322085404122@stu.suse.edu.cn