计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(4):215−225 [doi: 10.15888/j.cnki.csa.009456]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于边缘图与多尺度特征融合的图像修复①

黄　健, 王　虎, 赵小飞

(西安科技大学 通信与信息工程学院, 西安 710600)
通信作者: 王　虎, E-mail: 1323279271@qq.com

摘　要: 针对现有的图像修复方法在面对大规模图像缺损和不规则破损区域修复时, 修复结果出现生成结构与原图像语义不符以及纹理细节模糊等问题, 本文提出一种利用生成边缘图的多尺度特征融合图像修复算法——MSFGAN (multi-scale feature network model based on edge condition). 模型采用两阶段网络设计, 使用边缘图作为修复条件对修复结果进行结构约束. 首先, 使用 Canny 算子提取待修复图像的边缘图进行完整边缘图生成; 然后利用完整的边缘图结合待修复图像进行图像修复. 为了弥补图像修复算法中经常出现的问题, 提出一种融入了注意力机制的多尺度特征融合模块 (attention mechanism multi-fusion convolution block, AM block), 实现受损图像的特征提取和特征融合. 在图像修复网络解码器部分引入跳跃链接, 将高级语义提取和底层特征进行融合实现高质量细节纹理修复. 在 CelebA 和 Places2 数据集上的测试结果显示, MSFGAN 修复质量上比当前修复方法有一定提升, 其中在 20%–30% 掩码比例中, SSIM 平均提升 0.0291, PSNR 提升 1.535 dB, 使用消融实验验证了当前优化和创新点在图像修复任务中的有效性.

关键词: 深度学习; 生成对抗网络; 边缘生成; 图像修复; 注意力机制

引用格式: 黄健,王虎,赵小飞.基于边缘图与多尺度特征融合的图像修复.计算机系统应用,2024,33(4):215–225. http://www.c-s-a.org.cn/1003-3254/9456.html

Image Restoration Based on Edge Map and Multi-scale Feature Fusion

HUANG Jian, WANG Hu, ZHAO Xiao-Fei

(College of Communication and Information Technology, Xi’an University of Science and Technology, Xi’an 710600, China)

Abstract: In the face of large-scale image defects and irregular damage areas, existing image restoration methods often produce results with structural inconsistencies and blurry texture details. This study proposes an image restoration algorithm using the generated edge map and multi-scale feature fusion—MSFGAN (multi-scale feature network model based on edge condition). The model adopts a two-stage network design, using the edge map as a restoration condition to constrain the structural aspects of the restoration results. Firstly, the Canny operator is used to extract the edge map of the image to be restored, generating a complete edge map. Then, the complete edge map is combined with the image to be restored for image restoration. To address common issues in image restoration algorithms, an Attention Mechanism Multi-Fusion convolution block (AM block) is proposed, integrating an attention mechanism for feature extraction and fusion of damaged images. Skip connections are introduced in the decoder part of the image restoration network to fuse high-level semantics and low-level features, achieving high-quality detail and texture restoration. Test results on the CelebA and Places2 datasets show that MSFGAN has improved restoration quality compared to current methods. In the 20%–30% mask ratio, the average improvement of SSIM is 0.0291, and PSNR improvement is 1.535 dB. Ablation experiments

① 基金项目: 陕西省重点研发计划 (2023-YBGY-255); 陕西省科技厅工业攻关 (2022GY-115) ; 陕西省教育厅服务地方企业 (22JC050)
收稿时间: 2023-09-13; 修改时间: 2023-11-09; 采用时间: 2023-11-24; csa 在线出版时间: 2024-03-01
CNKI 网络首发时间: 2024-03-07

© 中国科学院软件研究所　http://www.c-s-a.org.cn

validate the effectiveness of the proposed optimization and innovations in image restoration tasks.

Key words: deep learning; generative adversarial network (GAN); edge generation; image restoration; attention mechanism

图像修复是图像处理领域的重要分支, 根本目的在于对图像中的破损部分和缺失部分进行复原. 图像修复的难点在于利用破损区域周围的像素信息合成结构连贯、语义合理的结构. 图像修复技术在现实生活中的应用极为广泛, 例如: 老旧照片的修复, 图像水印去除、对图像中的指定对象进行去除等[1].

图像修复工作由早期的基于扩散的图像修复技术和基于补丁的图像修复技术发展为基于深度学习的图像修复技术. 基于深度学习的图像修复技术克服了传统技术中的修复结果单一、结构不合理以及修复过程搜索算法时间长的问题, 即可以结合待修复图像的上下文语义信息生成原始图像或者数据库中不包含的像素信息. 近年来, 随着生成对抗网络 (generative adversarial network, GAN)[2]在计算机视觉领域的应用, 图像修复任务逐渐发展为条件生成任务. 使用生成对抗网络的图像修复技术实现了高级语义信息和低级像素信息的链接, 在网络对图像进行压缩提取语义信息后, 结合提取到的语义信息重新生成新的图片并和原图像完整部分拼接得到修复结果. 但是修复结果往往会出现生成结果模糊、过于平滑以及修复边缘结构模糊的现象.

基于生成对抗网络的图像修复技术最早由上下文编码器 (context encoder, CE)[3]提出, 该网络由编解码器结构将大小为 256×256 的破损形状为中心矩形的图像进行修复, 修复结果存在着纹理模糊和生成结构不合理的现象. 后续的研究往往使用引入注意力机制和空洞卷积 (dilated convolution)[4]的方式扩大卷积感受野来提升修复质量. 但是这种方式往往忽略了破损部分相邻像素包含的特征信息, 导致修复结果在修复边缘处表现不佳.

针对上述问题, 本文提出了一种基于边缘图条件的多尺度特征融合网络 (multi-scale feature network model based on edge condition). 该模型使用 StackGAN[5]的两阶段网络理念, 使用两个阶段分别对修复结果进行边缘预测和图像修复, 其中使用一阶段生成的边缘图对二阶段图像修复网络进行结构生成引导, 该模型的主要设计内容分为以下几个方面.

(1) 本文结合门控卷积和混合空洞卷积的思想提出 AM block (attention mechanism multi-fusion convolution block) 作为两阶段网络生成器的中间层. AM block 中使用不同扩张率的空洞卷积核和通道注意力机制可以确保网络在进行特征提取的时候可以充分提取不同感受野的特征信息并进行充分融合, 在模块中融入的门控机制可以使网络在边缘处区分破损区域和完整区域, 充分利用有效信息.

(2) 为了改进图像修复中容易出现的修复边缘结构不合理以及修复结果缺乏高质量纹理的现象, 本文在第 2 阶段图像修复网络的编解码器中引入跳跃连接, 使得网络在解码过程中兼顾底层特征的有效信息.

(3) 网络使用谱归一化马尔可夫判别器配合改进后的生成器网络组成生成对抗网络结构进行训练, 以生成结构合理, 纹理清晰的图像修复结果.

1 相关工作

早期的图像修复是提取图像自身有效信息通过算法来对破损区域进行填充, 重点在如何实现有效信息的提取和破损区域周围信息的有效融合[6].

基于像素扩散的修复方法中 Drori 等人[7]通过计算破损图像的上下文信息计算样本的相似程度然后使用算法来对图像进行修复, 其中计算样本的相似程度需要占用大量的计算资源. Criminisi 等人[8]提出结合纹理复制的修复算法, 通过计算补丁优先级计算优先修复的补丁, 对图像破损区域边缘逐渐修复. Hays 等人[9]建立外部数据库, 使用待修复图像相关信息搜索相似的图像, 提取相似匹配度高的部分进行填充, 但是这依赖巨大的搜索数据库, 而且修复效果不佳.

基于样本的修复方法中 Levin 等人[10]利用对图像信息的统计, 通过建立局部特征直方图在图像上构建指数族分布寻找和当前图像特征最为接近的样本, 然后使用样本对缺损部分进行填充. Barnes 等人[11]使用快速最近邻算法搜索和当前图像相似度最高的样本对图像进行修复, 快速最邻近算法降低了搜索样本时的内存占用和计算成本.

由此可见传统修复算法以计算样本相似度和相关搜索算法为基础在当前图像信息或者其他样本信息的

© 中国科学院软件研究所 http://www.c-s-a.org.cn

集合中搜索最为合适的修补样本, 在小范围的破损图像或者指定位置破损的图像中可以取得较为理想效果. 然而该类修复方式没有对图像的语义理解, 仅依靠自身信息的扩散对缺损部分填充, 无法在被修复区域产生原图像中没有的结构和纹理, 所以在面对大面积破损区域以及恢复没有对应样本的补丁时效果不好, 并且在修复的过程中需要大量计算成本.

随着深度神经网络的发展, Pathak 等人[3]在 2016 年提出上下文编码器 (CE), 将编码器解码器应用到图像修复上, 编码器来提取缺损图像的语义和纹理的深层语义表征, 解码器用来还原图像, 融合对抗损失后能进一步恢复图像的细节纹理. 不同于传统方法, 上下文编码器可以学习缺损图像的语义信息, 从而生成图像中原本没有的部分. 由此基于深度学习的图像方法得到了证实, 更多深度学习的算法开始被研究. 在此基础上 Wang 等人[12]使用多种卷积提取特征, 在解码过程进行融合, 多列中间层特征融合使上下文编码器更好地获取不同尺度的特征信息, 使修复结果的结构更合理.

Yan 等人[13]使用 U-Net[14]结构设计生成网络, 将低层次的语义信息和高层次的语义信息进行特征融合, 网络中提出的移位链接层结合了 Criminisi 等人[8]提出的逐层修复的方法, 将深度学习方法和传统方法结合, 使用图像有效部分的纹理来对破损区域进行填充. 神经网络确保对图像的语义信息和特征信息进行充分提取, 移位链接层确保修复纹理来自于图像本身纹理信息, 在小规模破损的图像修复任务中可以取得高质量的修复结果. 之后, Liu 等人[15]提出了一种部分卷积, 通过对像素有效性进行二值划分, 同时使用掩码更新的策略优化了修复伪影的问题, 但自动更新掩码机制过于不稳定, 且粗略的二值划分不能充分地利用有效信息.

Nazeri 等人[16]提出了一种先修复边缘再修复纹理的两阶段图像修复方法 EdgeConnect. 利用算法提取图像的边缘信息进行边缘预测, 或者使用外部导入提前设置好的边缘图像进行指导, 修复结果在结构合理程度上有着较好表现. Yu 等人[17]在 StackGAN[5]架构体系下设计了一个由粗到细的两阶段图像修复网络, 使用上下文注意力来计算缺损部分和未缺损部分的关联程度, 但是修复过程中往往会出现结构失真等情况. 次年 Yu 等人[18]继续提出引入用户指导草图来配合网络对指定部分进行修复. Sagong 等人[19]提出了一个使用语义注意力模块 (contextual attention module, CAM) 和共享编码解码结构的图像修复方法使用共享参数来对特征进行提取以及特征解码. 之后, Shin 等人[20]继续在 PESPI[19]的基础上提出了使用速率自适应扩张卷积层优化后的网络, 根据逐渐提升的扩张率提取特征信息.

2019 年 Liu 等人[21]在提出了一种带有关联语义注意力机制的方法, 简称 CSA 模型. CSA 模型使用其特有的注意力机制计算图像损坏区域与图像未破损区域之间的相似权重, 通过相似度进行匹配, 还处理了图像损坏区域深层特征之间的关联性. Li 等人[22]设计了一个循环特征推理模块来进行阶段式的修复工作利用相邻像素之间的关联加强生成区域的约束, 以生成更高质量的修复图. Zeng 等人[23]在 Nazeri 等人[16]提出的网络结构基础上设计了特征融合中间层, 使用单阶段网络缩小了网络体量并加强了特征信息融合, 修复结果有着较高细粒度和细节纹理, 但往往出现边缘模糊和断层的现象.

由当前阶段的研究可以看出, 基于深度学习的图像修复方法, 由于存在着语义不充分、网络缺乏结构约束、高级语义无法融合底层特征信息等原因, 导致修复结果出现结构不合理、生成纹理模糊、生成图像和原图像语义不一致等问题. 基于此, 本文从增加约束条件和优化网络特征提取、融合来进行研究. 在约束条件方面, 使用单独的网络来生成边缘图对图像修复进行结构约束, 在特征融合方面, 提出使用 AM block 为中间层和跳跃连接编码器解码器使网络加强完整部分的特征提取和融合, 解码过程将高级语义信息和底层特征信息拼接融合以生成高质量纹理结构.

2 模型设计

现有的基于深度学习的图像修复算法往往难以保证生成结果的结构合理性和高质量纹理. 如何确保生成部分和未破损区域在边缘交界处结构连贯和生成高质量纹理修复结果是当前研究的主要问题. 为此本文提出一种利用生成边缘图为条件的图像修复方法.

网络为生成对抗网络, 生成器和判别器迭代训练以生成高质量修复结果. 网络使用两阶段设计, 每个网络都可以进行独立训练. 一阶段网络为边缘图修复网络, 目的在对提取出的破损边缘图进行修复, 以生成合

© 中国科学院软件研究所 http://www.c-s-a.org.cn

理的边缘结构图. 第 2 阶段为图像修复网络, 使用第 1 阶段生成的边缘图作为修复条件对破损图像进行修复, 确保生成结构合理. 判别器使用谱归一化马尔可夫判别器来判断生成图像是否接近原图像, 通过迭代博弈训练来优化网络参数以得到合理的修复效果. 整体网络结构图如图 1 所示.

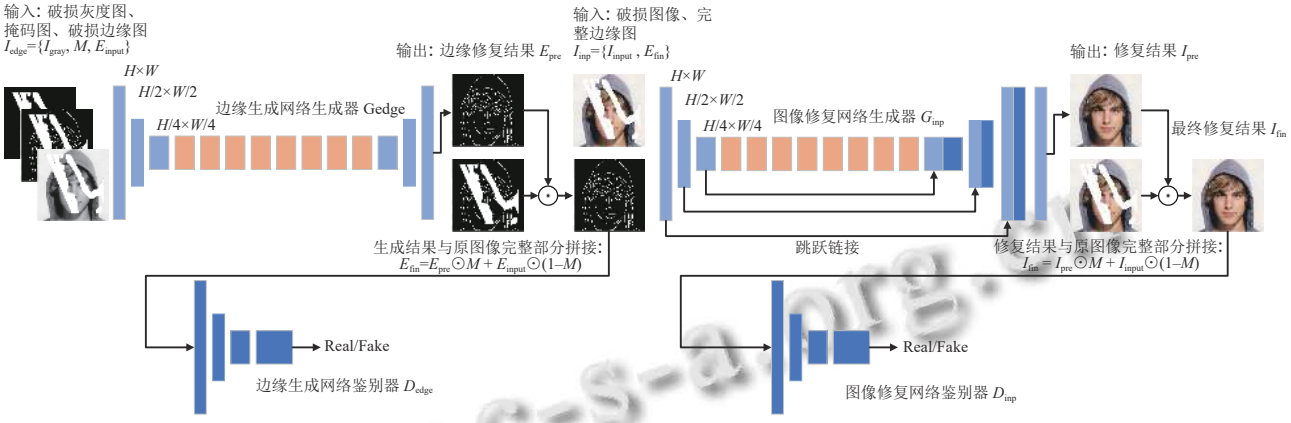


图 1 整体网络结构

2.1 生成网络结构

在图像修复的第 1 阶段, 即图像边缘图的生成阶段, 将修复图像I_{input}进行灰度化处理, 得到单通道的灰度图I_{gray}, 接着使用 Canny 算子对灰度图进行边缘提取得到图像的受损边缘图E_{input}; 最后将受损图像灰度图I_{gray}、受损部分掩码图M和E_{input}通道维度合并为联合输入I_{edge}送入边缘生成网络. 如式 (1) 所示.

$$I_{edge}=\{I_{gray}, M, E_{input}\} \tag{1}$$

经过边缘生成器G_{edge}和判别器D_{edge}联合训练后, 边缘生成器G_{edge}可以对破损图像的完整边缘图进行预测, 得到完整的修复边缘结果E_{pre}. 如式 (2) 所示.

$$E_{pre}=G_{edge}(I_{edge}) \tag{2}$$

使用二值掩码图M拼接原破损图像边缘图E_{input}和生成结果边缘图E_{pre}, 保留原破损图像的有效信息, 得到最终修复边缘图E_{fin}, 如式 (3) 所示. 将原破损图像I_{input}和修复边缘图E_{fin}作为第 2 阶段修复网络的输入I_{inp}, 如式 (4) 所示.

$$E_{fin}=E_{pre}\odot M+E_{input}\odot(1-M) \tag{3}$$

$$I_{inp}=\{I_{input}, E_{fin}\} \tag{4}$$

经过图像修复网络生成器G_{inp}和判别器D_{inp}联合训练后, 生成器G_{inp}可以根据输入I_{inp}生成修复结果I_{pre}, 如式 (5) 所示. 使用二值掩码M对原破损图的有效区域进行保留得到最终结果I_{fin}, 如式 (6) 所示.

$$I_{pre}=G_{inp}(I_{inp}) \tag{5}$$

$$I_{fin}=I_{pre}\odot M+I_{input}\odot(1-M) \tag{6}$$

边缘生成网络和图像修复网络的生成器部分采用相似结构, 编码器使用两次下采样将输入进行特征提取, 使用融入注意力机制的多尺度特征融合模块代替普通的残差模块作为网络中间层, 在提升网络感受野的同时保证获取信息的连续性, 融入的门控机制确保了对原图像有效部分的利用. 最终使用与编码器结构对称的解码器对修复信息进行解码, 实现图像修复, 其中图像修复网络将编码过程中每次得到的特征图使用跳跃连接和解码过程中相同尺寸的特征图进行拼接使得解码过程可以充分利用编码过程得到的底层特征信息.

2.2 判别网络和损失函数

生成对抗网络的判别器使用谱归一化马尔可夫判别器扩大判别器感受野, 满足 Lipschitz 连续提高训练稳定性. 在两阶段网络中分别在每个阶段设置相同的判别器, 两个网络均可以独立训练. 在端对端的两阶段图像修复网络训练中, 去除第 1 阶段边缘修复网络的判别器D_{edge}, 仅保留修复网络中的判别器进行训练.

判别网络由 5 次卷积核大小为 4×4 的下采样组成, 每次得到的特征图通道分别为 64、128、256、512、1. 使用的谱归一化判别器通过将网络的 Lipschitz 常数限制为 1, 优化对抗训练. 在第 1 阶段边缘修复网络的训

© 中国科学院软件研究所 http://www.c-s-a.org.cn

练中使用对抗损失、特征匹配损失和 L1 损失. 第 1 阶段联合损失函数如示 (7) 所示.

$$\mathcal{L}_{\text{edge}} = \lambda_{\text{adv1}}\mathcal{L}_{\text{adv1}} + \lambda_{\text{FM}}\mathcal{L}_{\text{FM}} + \lambda_{\text{L1}}\mathcal{L}_{\text{L1}} \tag{7}$$

其中, λ_{adv1} 、 λ_{L1} 和 λ_{FM} 为超参数. $\mathcal{L}_{\text{adv1}}$ 为第 1 阶段对抗网络的对抗损失, $\mathcal{L}_{\text{FM}}$ 为特征匹配损失 (feature matching loss), 取判别器中间层和与训练好的 VGG 网络中间层计算损失, 促进生成器生成与真实图像相近的结果来稳定训练过程. 特征匹配损失如式 (8) 所示.

$$\mathcal{L}_{\text{FM}}(G, D_k) = \left[\sum_{i=1}^{T}\frac{1}{N}\left\|D_1^k(E_{\text{gt}}) - D_1^k(E_{\text{fin}})\right\|_1\right] \tag{8}$$

其中, E_{gt} 代表真实图像的边缘图, N_k 代表判别器 D_{edge} 的第 k 个激活层, D_k 代表判别器中间层的输出, T 代表判别器中间层的数量. 在第 2 阶段图像修复网络使用对抗损失、L1 损失、感知损失和风格损失作为联合损失训练. 第 2 阶段网络联合损失函数如式 (9) 所示:

$$\begin{aligned}\mathcal{L}_{\text{impating}} =& \lambda_{\text{L1}}\mathcal{L}_{\text{L1}} + \lambda_{\text{adv2}}\mathcal{L}_{\text{adv2}} \\ &+ \lambda_{\text{prec}}\mathcal{L}_{\text{prec}} + \lambda_{\text{style}}\mathcal{L}_{\text{style}}\end{aligned} \tag{9}$$

其中, λ_{L1} 、 λ_{adv2} 、 λ_{prec} 和 λ_{style} 为超参数, $\mathcal{L}_{\text{L1}}$ 为 L1 损失, $\mathcal{L}_{\text{adv2}}$ 为第 2 阶段网络对抗损失, $\mathcal{L}_{\text{prec}}$ 为感知损失, $\mathcal{L}_{\text{style}}$ 为风格损失; 感知损失和风格损失在深度学习的图像修复任务中的有效性在以往的研究中得到了验证. 感知损失降低修复结果和真实图像之前的 L1 距离, 感知损失函数如式 (10) 所示.

$$\mathcal{L}_{\text{prec}} = \mathbb{E}\left[\sum_i \frac{1}{N_i}\left\|\phi_i(I_{\text{gt}}) - \phi_i(I_{\text{pred}})\right\|_1\right] \tag{10}$$

其中, ϕ_i 为预训练网络第 i 层激活图. 风格损失为原始图像与修复图像对应的格拉姆矩阵的 L1 距离. 风格损失通过卷积来提取图像的纹理特征, 利用协方差矩阵可以得到纹理信息. 风格损失函数如式 (11) 所示.

$$\mathcal{L}_{\text{style}} = \mathbb{E}_j\left[\left\|G_j^{\phi}(\widetilde{I}_{\text{pred}}) - G_j^{\phi}(\widetilde{I}_{\text{gt}})\right\|_1\right] \tag{11}$$

其中, G_j^{ϕ} 是 ϕ_i 的激活映射构造出的 $C_j \times C_j$ 的格拉姆矩阵. ϕ_i 为当前卷积网络第 i 个卷积层, $\widetilde{I}_{\text{pred}}$ 为网络生成结果, $\widetilde{I}_{\text{gt}}$ 为真实图像.

2.3 融入注意力机制的多尺度特征提取模块

在生成器编码部分对图像进行特征提取后, 如何充分获得特征图中语义信息进行特征融合以生成结构合理的修复结果一直是图像修复任务中的难题. 之前的算法通常使用空洞卷积进行堆叠以获得更大范围的感受野. 混合空洞卷积 (hybrid dilated convolution, HDC)[24]的理论说明使用相同扩张率的空洞卷积堆叠的时候虽然会扩大感受野, 但是丧失了获取信息的连贯性, 在图像修复任务中待修复部分边缘会存在大量有效信息, 连续的空洞卷积对图像修复这类需要连续周边信息的任务不适用, 往往会产生棋盘格效应以及待修复边缘结构不连贯等问题.

因此本文提出一种多尺度特征融合模块 AM block, 使用不同扩张率的空洞卷积对特征图进行卷积计算, 并将得到的特征图在通道维度上拼接. 使用门控的机制和通道注意力机制对不同的空洞卷积进行权重分配. 使生成器在获取更大的感受野的同时保证对周边信息的利用, 同时门控机制和通道注意力机制使得网络在提取特征的时候利用软门控区分待修复图像中的有效像素信息, 并在特征融合时重新对不同卷积核提取的特征信息进行权重重新分配. AM block 结构图如图 2 所示.

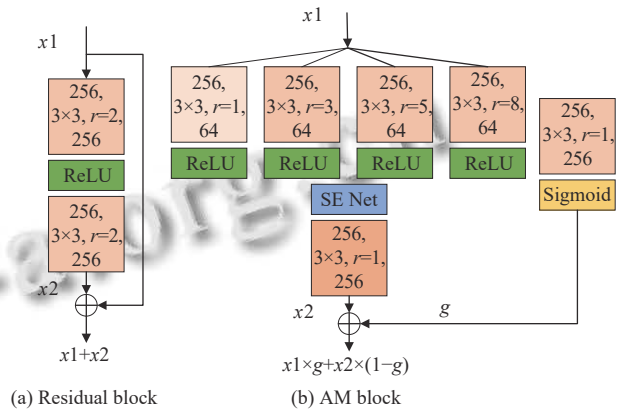


图 2 AM block 结构图

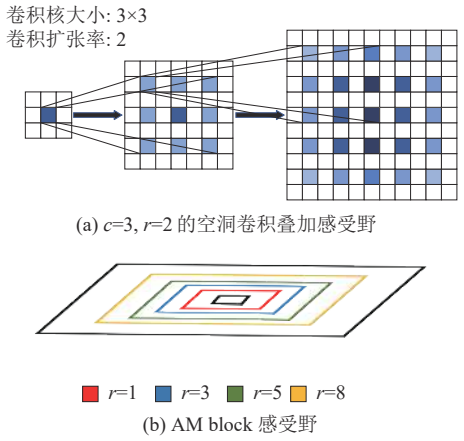
模块内使用 4 种不同扩张率且最大公倍数为 1 的空洞卷积, 扩张率分别为 1、3、5、8 在确保获得更广泛感受野的同时, 使用扩张率为 1 的普通卷积获得像素周边的信息. 多个空洞卷积叠加感受野和本文方法感受野如图 3 所示.

经过不同扩张率的卷积核提取特征后, 使用 Leaky-ReLU 函数激活, 进行通道维度上的拼接, 引入 SE-Net (squeeze-and-excitation networks)[25]对通道进行新的权重分配, SE-Net 使得网络学习对不同卷积核获得的结

© 中国科学院软件研究所 http://www.c-s-a.org.cn

果的利用率. SE-Net 如图 4 所示.

空洞卷积在图像修复任务中对于所有的空间像素都视为有效像素并进行计算, 这导致网络在面对不规则修复的任务时无法区分破损部分和完整部分. 在之前的研究中部分卷积使用更新掩码图的方式对缺损部分和完整部分进行区分, 每次从边缘修复一部分图像, 并更新掩码使修复后的部分变为有效像素, 空洞卷积如图 5 所示.



(a) c=3, r=2 的空洞卷积叠加感受野

(b) AM block 感受野

图 3 单一空洞卷积叠加感受野和本文方法感受野

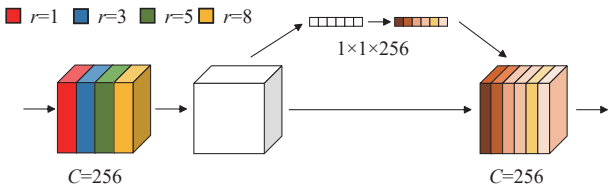


图 4 挤压激励网络 SE-Net

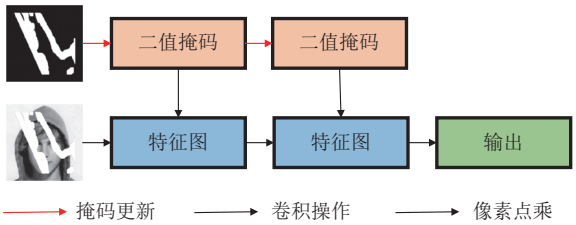


图 5 部分卷积

虽然部分卷积可以有效使网络区分有效像素和无效像素, 但是仅是简单地将有效部分像素置为 1, 无效部分置为 0, 不能对具体的边缘部分进行重要程度划分.

和部分卷积相比, 门控卷积使用空间注意力机制的思想, 给每个通道上的空间提供动态的特征选择. 门控卷积使网络学习每个阶段的掩码状态, 确定相应的软掩码值, 不只使网络区别有效部分和无效部分, 而是动态地进行划分. 门控卷积结构图如图 6 所示.

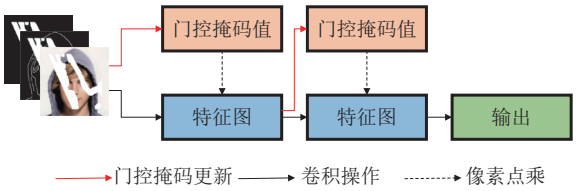


图 6 门控卷积

3 实验结果与分析

本实验在 NVIDIA RTX3080 上运行, 深度学习框架使用 PyTorch 1.12.1, 开发环境为 Python 3.6, 使用 Adam 优化器优化训练, 参数设置为 0.01 和 0.9. 实验分别在 Places2 环境数据集和 CelebA 人脸数据集上对 MSFGAN 网络模型的有效性进行验证.

CelebA 人脸数据集是由香港中文大学提供的开放数据集, 共 200 000 张不同环境下的人脸图片. 实验随机选择数据集中的 100 000 张片人脸图像进行裁剪作为网络训练集, 从剩余图片中随机选取 1 000 张图像用于检验网络修复效果. Places2 场景数据集包含各种场景和人物的 1 800 000 张图像, 选择地点和建筑相关的 150 000 张图像作为训练集进行训练, 随机选取未训练的 1 000 张图像作为测试集. 掩码数据集使用 Liu 等人[21]提供的不规则掩码数据集. 不规则掩码数据集包含不同掩码比例的掩码, 掩码区域占比从 10%–60%, 其中每种掩码比例有不相同的掩码图 2 000 张, 掩码区域分为中心集中掩码和全图掩码两种, 使用所有掩码图进行训练. 实验中使用到的 3 种数据集样本如图 7 所示.

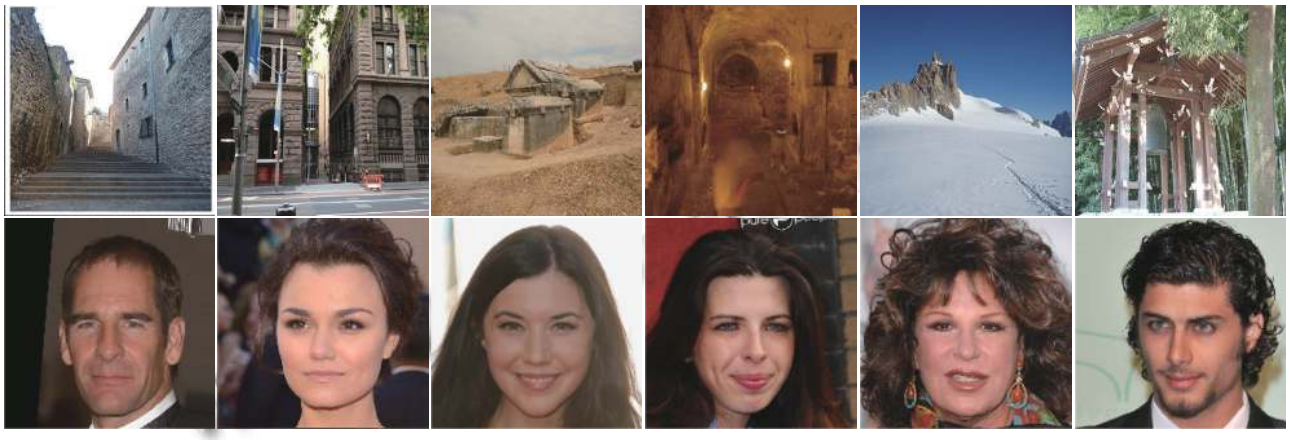
实验将本文方法和当前主流方法 Contextual Attention (CA)[17]、RFR-net (RFR)[22]和 AOT-GAN (AOT)[23]进行对比测试.

3.1 定量分析

实验在 Places2 数据集上对 4 种修复方法进行定量测试. 使用相同的图像数据集和不同程度破损的掩码进行测试. 实验在 Places2 数据集中选取每种破产程度的图像 500 张, 共计 2 500 张图像. 定量分析采用图像修复经典指标为峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似度 (structural similarity, SSIM). 峰值信噪比定义为重建图像和原始图像的失真程度,

峰值信噪比数值越高, 表示修复结果和原图像的相似度越高. SSIM 通过估计亮度, 感知结构相似程度和对比度, 衡量人主观观测到的图像风格相似程度. 定量分析结果如表 1 所示.

从数据中可以看出, MSFGAN 在复原结构准确度和主观视觉符合度上取得了最优的效果. 可以看出在受损面积小的情况下, 各算法都可以得到较高的修复效果, 但是当破损面积增大到 20% 以上时, MSFGAN 的多尺度融合和边缘引导的效果开始显现. 在具体数值上, MSFGAN 在 5 种不同掩码下的平均 SSIM 值达到了 0.904 0, PSNR 值达到了 27.22 dB, 取得了较好的结果. 对比实验结果如图 8 所示.



(a) Places2 数据集和 CelebA 人脸数据集

0%–10% 10%–20% 20%–30% 30%–40% 40%–50% 50%–60%

(b) 不规则掩码数据集

图 7 本文使用数据集示例

表 1 Places2 数据集上不同算法在不同掩码比例下的 PSNR (dB)/SSIM 值

掩码比例	CA	RFR	AOT	MSFGAN
1%–10%	31.185/0.969 5	32.256/0.970 5	33.204/0.975 6	33.954/0.988 4
10%–20%	26.304/0.880 1	28.146/0.909 6	28.756/0.921 1	29.231/0.951 4
20%–30%	24.260/0.849 7	25.146/0.866 3	25.328/0.875 2	26.863/0.904 3
30%–40%	22.135/0.802 0	22.632/0.827 2	22.965/0.863 1	23.956/0.897 0
40%–50%	18.119/0.743 1	21.126/0.749 5	21.635/0.742 2	22.105/0.779 1

3.2 消融实验

为了验证 AM block 和边缘图引导在图像修复工作中的有效性, 对本文提出方法在 Places2 数据集上进行消融实验. 分别对不使用边缘图引导且残差网络的中间层网络 (Res block)、使用边缘图引导的以残差网络为中间层的网络 (edge + Res block)、不使用边缘图引导的 AM block 中间层网络 (AM block) 和本文提出方法使用边缘引导和 AM block 网络 (MSFGAN) 进行对照实验. 实验结果如图 9 所示.

可以看出, 含有边缘引导的网络可以使修复结果的结构更加平滑, 生成结构更加合理, AM block 使修复结果的纹理效果更精细, 细粒度更高, 这得益于不同感受野的卷积和门控机制, 使修复阶段可以融合更多的像素信息并弹性区分有效部分. 使用 PSNR 和 SSIM 对 4 种网络进行指标检测, 可以看出两种模块对提升图像结果的主观感受和与真实图片的相似度均有提升. 对比结果如表 2 所示.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

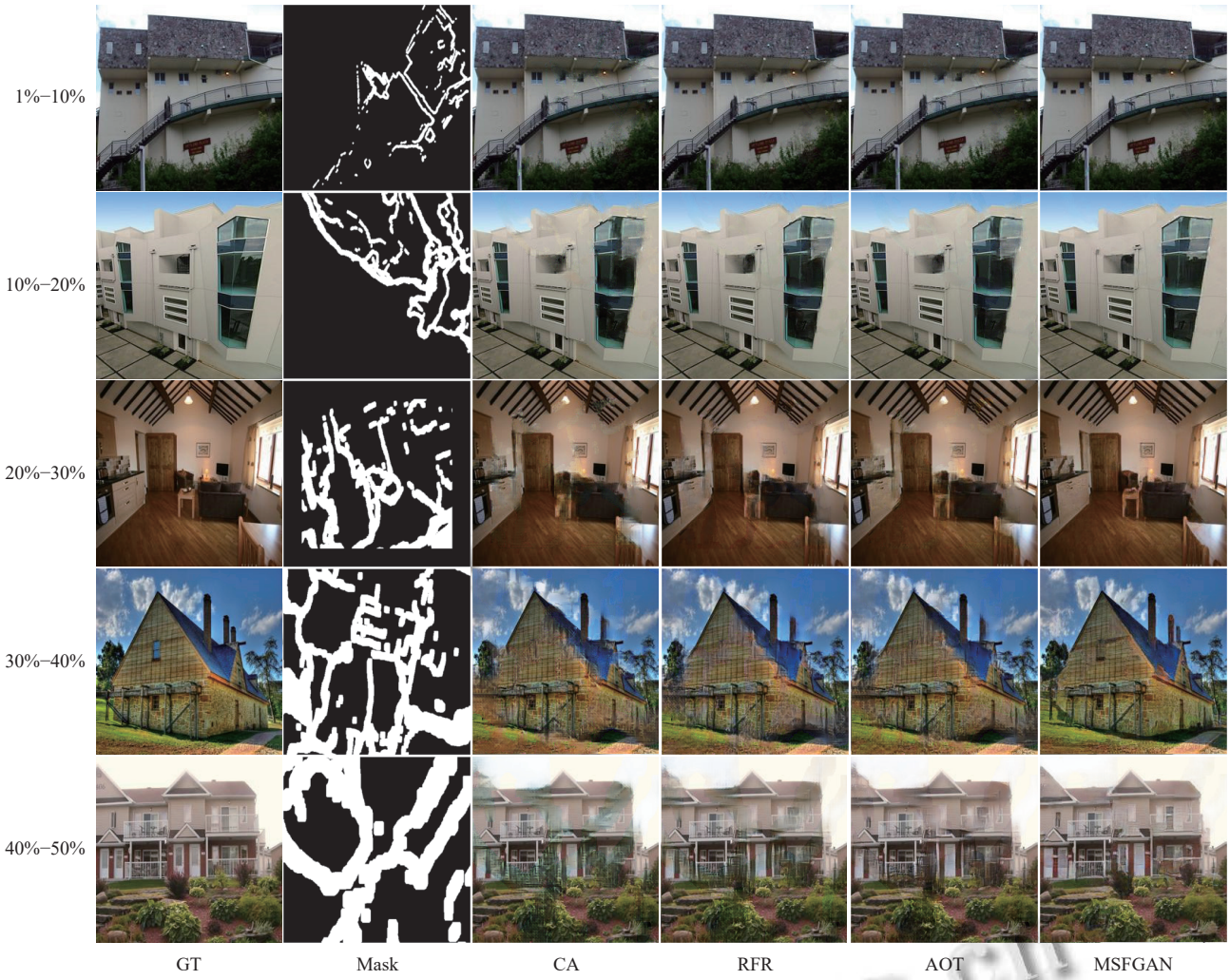


图 8　相同掩码下和主流算法的对照实验

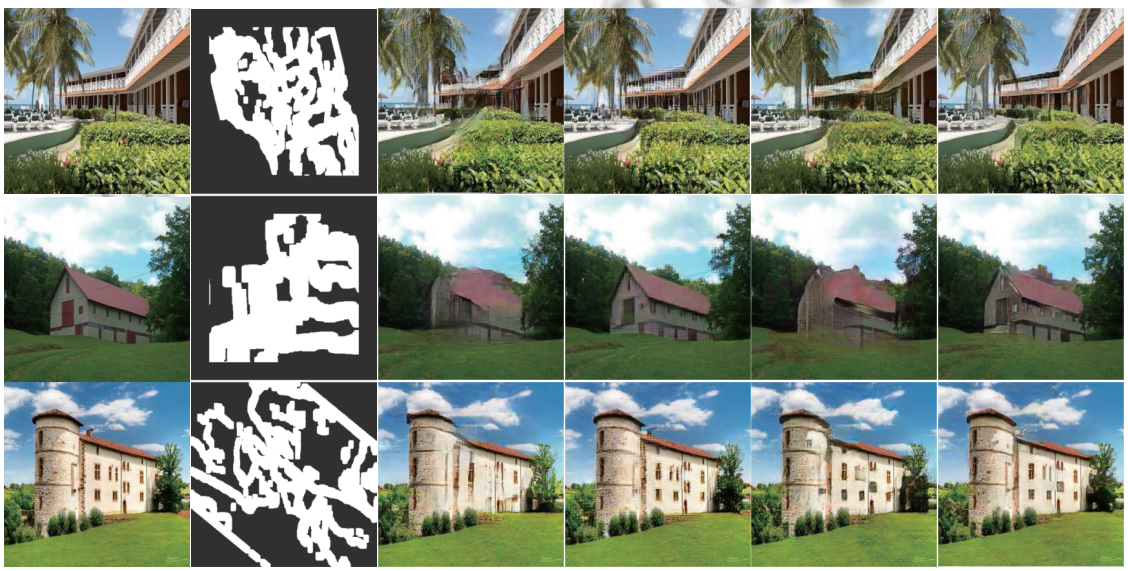


图 9　消融实验

© 中国科学院软件研究所 http://www.c-s-a.org.cn

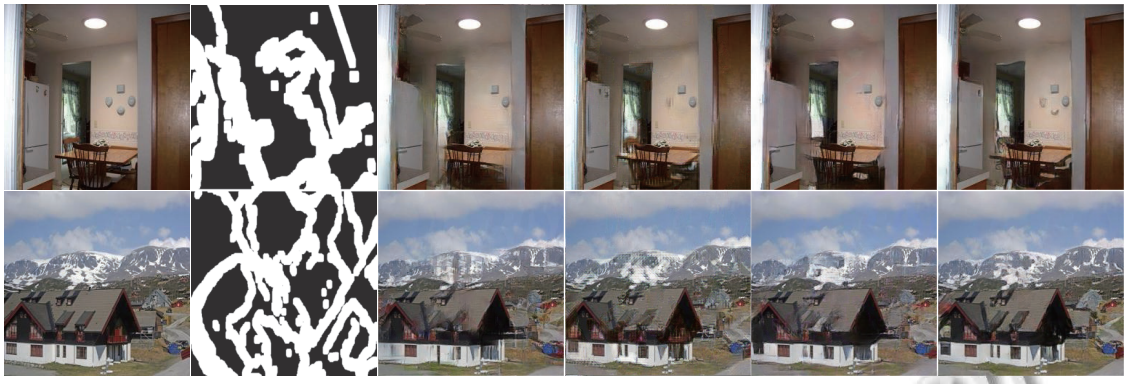


图 9 消融实验 (续)

表 2 在 Palce2 数据集上进行消融实验各网络 PSNR (dB)/SSIM 值

掩码比例	Res block	edge + Res block	AM block	MSFGAN
1%–10%	29.148/0.921 3	31.248/0.965 2	30.960/0.974 9	33.954/0.988 4
10%–20%	27.336/0.896 6	28.112/0.942 2	27.223/0.936 8	29.231/0.951 4
20%–30%	24.226/0.863 3	25.632/0.875 5	25.328/0.876 3	26.863/0.904 3
30%–40%	20.746/0.820 2	22.228/0.853 3	22.117/0.856 4	23.956/0.897 0
40%–50%	20.224/0.725 6	21.475/0.752 2	21.366/0.750 0	22.105/0.779 1

3.3 边缘生成实验

上述定量实验和消融实验的结果展示了本文方法在边缘图引导和中间层改进在图像修复任务中的有效性. 由于本文网络两阶段修复的特性, 边缘生成网络能否生成清晰合理的边缘图至关重要, 因此为了测试边缘生成网络在较大规模破损的图像修复任务以及人脸生成任务的表现, 本文设计了边缘生成实验, 对边缘生成网络在 CelebA 人脸数据集上进行训练. 随机选取了 2 万张人脸图片以及 1 000 张掩码比例为 40%–50% 的掩码图并进行了训练集和测试集的随机划分, 边缘生成网络使用 Canny 算子对破损边缘图行预处理来生成破损的边缘图, 判别器将生成器生成的边缘图和真实图像经过相同参数的 Canny 算子提取的完整边缘图进行判别并进行迭代训练.

实验表明, 得益于边缘生成网络编码器解码器的设计结构以及本文提出模块 AM block 的多尺度特征融合中间层, 网络在编码阶段充分提取了受损图像的特征并在中间层部分进行了充分的特征融合, 网络学习到了人脸修复任务中的关键点, 并生成了语义合理的边缘图. 可以看出, 在面对大面积修复任务以及人脸修复任务等对语义要求严格的相关修复任务时, MSFGAN 仍能生成主观合理、纹理清晰的边缘图来引导图像修复任务. 边缘生成实验如图 10 所示.

图 10 人脸修复任务边缘生成效果及修复结果

© 中国科学院软件研究所 http://www.c-s-a.org.cn

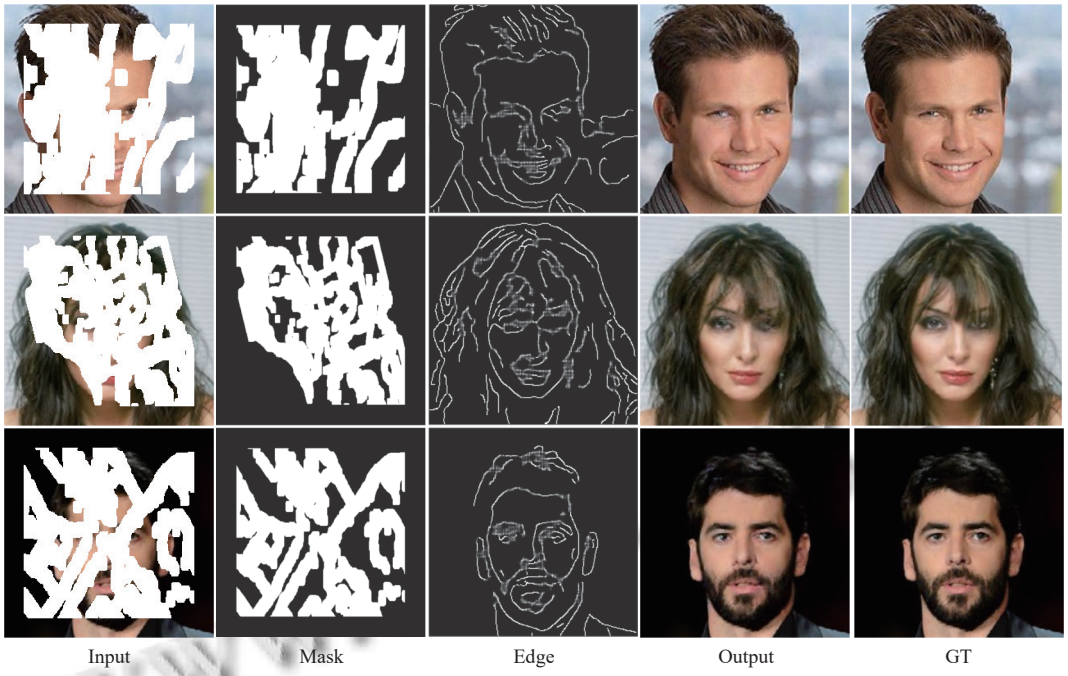


图 10 人脸修复任务边缘生成效果及修复结果 (续)

4 结论与展望

本文针对当前基于深度学习图像修复方法中出现的问题提出了一种基于边缘图和多尺度特征融合的图像修复算法. 首先使用边缘网络对破损边缘图进行修复补全, 生成结构合理符合语义的结构边缘; 接着利用生成的边缘图对第 2 阶段图像修复网络进行结构约束, 以生成结构合理的高质量修复结果.

在 CelebA 数据集和 Places2 数据集上和当前主流的修复方法进行了定向实验. 结果表明本文算法在生成合理结构和修复边缘细节上优于其他算法. 在相同掩码程度的图像修复中进行消融实验, 得到当前提出的方法对网络特征提取和图像生成有一定价值. 在实验过程中发现, 在处理高分辨率的图像修复任务时出现了局部纹理过于平滑, 修复区域模糊的情况, 此外由于网络依赖完整边缘图, 所以如何生成边缘图和处理高像素图像修复任务将在后续工作中进行优化.

参考文献

1 李月龙, 高云, 闫家良, 等. 基于深度神经网络的图像缺损修复方法综述. 计算机学报, 2021, 44(11): 2295–2316.

2 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2672–2680.

3 Pathak D, Krähenbühl P, Donahue J, *et al*. Context encoders: Feature learning by inpainting. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 2536–2544.

4 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. Proceedings of the 4th International Conference on Learning Representations. San Juan, 2016.

5 Zhang H, Xu T, Li HS, *et al*. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 5908–5916.

6 罗海银, 郑钰辉. 图像修复方法研究综述. 计算机科学与探索, 2022, 16(10): 2193–2218.

7 Drori I, Cohen-Or D, Yeshurun H. Fragment-based image completion. ACM Transactions on Graphics, 2003, 22(3): 303–312. [doi: 10.1145/882262.882267]

8 Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 2004, 13(9): 1200–1212. [doi: 10.1109/TIP.2004.833105]

9 Hays J, Efros AA. Scene completion using millions of photographs. ACM Transactions on Graphics, 2007, 26(3): 4–es. [doi: 10.1145/1276377.1276382]

10 Levin A, Zomet A, Weiss Y. Learning how to inpaint from global image statistics. Proceedings of the 9th IEEE

© 中国科学院软件研究所 http://www.c-s-a.org.cn

International Conference on Computer Vision. Nice: IEEE, 2003. 305–312.

11 Barnes C, Shechtman E, Finkelstein A, *et al*. PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics, 2009, 28(3): 24.

12 Wang Y, Tao X, Qi XJ, *et al*. Image inpainting via generative multi-column convolutional neural networks. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 329–338.

13 Yan ZY, Li XM, Li M, *et al*. Shift-Net: Image inpainting via deep feature rearrangement. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.

14 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.

15 Liu GL, Reda FA, Shih KJ, *et al*. Image inpainting for irregular holes using partial convolutions. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, 2018. 89–105.

16 Nazeri K, Ng E, Joseph T, *et al*. EdgeConnect: Generative image inpainting with adversarial edge learning. arXiv: 1901.00212, 2019.

17 Yu JH, Lin Z, Yang JM, *et al*. Generative image inpainting with contextual attention. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 5505–5514.

18 Yu JH, Lin Z, Yang JM, *et al*. Free-form image inpainting with gated convolution. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 4470–4479.

19 Sagong MC, Shin YG, Kim SW, *et al*. PEPSI: Fast image inpainting with parallel decoding network. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 11352–11360.

20 Shin YG, Sagong MC, Yeo YJ, *et al*. Pepsi++: Fast and lightweight network for image inpainting. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 252–265. [doi: 10.1109/TNNLS.2020.2978501]

21 Liu HY, Jiang B, Xiao Y, *et al*. Coherent semantic attention for image inpainting. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 4169–4178.

22 Li JY, Wang N, Zhang LF, *et al*. Recurrent feature reasoning for image inpainting. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 7757–7765.

23 Zeng YH, Fu JL, Chao HY, *et al*. Aggregated contextual transformations for high-resolution image inpainting. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(7): 3266–3280. [doi: 10.1109/TVCG.2022.3156949]

24 Wang PQ, Chen PF, Yuan Y, *et al*. Understanding convolution for semantic segmentation. Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe: IEEE, 2018. 1451–1460.

25 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.

(校对责编: 牛欣悦)

© 中国科学院软件研究所 http://www.c-s-a.org.cn