

doi:10.3969/j.issn.1673-9833.2022.01.012

融合边缘条件的多个鉴别器生成对抗网络

朱文球, 汪晓毅, 黄史记

(湖南工业大学 计算机学院, 湖南 株洲 412007)

摘要: 针对现有图像处理方法仍然存在人脸补全后, 人脸五官等局部区域的视觉连通性较差的问题, 提出了一种包含全局鉴别网络、局部鉴别网络和人脸部位鉴别网络的边缘补全方法。其中全局鉴别网络鉴别全图的视觉连通性; 局部鉴别网络约束补全部分; 人脸部位鉴别网络约束补全图像效果。将残缺灰度图、残缺边缘图和掩膜图输入到边缘补全网络, 得到补全边缘图。然后将补全边缘图和残缺彩色图输入到图像补全网络, 得到补全图像。边缘补全网络和图像补全网络进行端对端连接, 形成一个完整的解决方案。在 CelebA 数据集上与对照组进行视觉连通性对比, 结果表明: 提出的算法能够更好地还原人脸部位的信息。

关键词: 生成对抗网络; 人脸补全; 深度学习; 多鉴别器; 自注意力机制

中图分类号: TP391.4

文献标志码: A

文章编号: 1673-9833(2022)01-0084-11

引文格式: 朱文球, 汪晓毅, 黄史记. 融合边缘条件的多个鉴别器生成对抗网络[J]. 湖南工业大学学报, 2022, 36(1): 84-94.

Multiple Discriminators Combining Edge Conditions to Generative Adversarial Network

ZHU Wenqiu, WANG Xiaoyi, HUANG Shiji

(College of Computer Science, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: In view of the flaw that the visual connectivity of such local areas as facial features is characterized with poor quality after face completion in existing image processing methods, an edge completion method, including global discriminator network, local discriminator network and face part discriminator network, has thus been proposed, among which the global discriminator network identifies the visual connectivity of the whole graph; the local discriminator network constraints the complement part; while the face part discriminator network constraints the complements the image effect. The completed edge image can be obtained by inputting the incomplete gray image, incomplete edge image and mask image into the edge completion network, to be followed by the input of the completed edge map and incomplete color map into the image completion network, thus obtaining the completed image, with edge completion network and image completion network connected end-to-end to form a complete solution. Visual connectivity was compared with the control group on CelebA dataset, the results show that the proposed algorithm is more effective in the restoration of the information of face parts.

Keywords: generative adversarial network; face inpainting; deep learning; multiple discriminators; self-attention mechanism

收稿日期: 2021-04-22

基金项目: 国家重点研发基金资助项目(2018AAA0100400); 湖南省自然科学基金资助项目(2021JJ50058); 湖南省教育厅开放平台创新基金资助项目(20K046); 湖南省战略性新兴产业科技攻关与重大科技成果转化基金资助项目(2019GK4009)

作者简介: 朱文球(1969-), 男, 湖南株洲人, 湖南工业大学教授, 硕士生导师, 主要研究方向为计算机视觉, 数字图像处理, 生物特征识别, E-mail: zwq@zhuzit.edu.cn

1 研究综述

在实际人脸识别场景中,很多时候会出现戴上眼镜、口罩等遮盖脸部的情况,此时由于获取的面部信息完整度不足,对人脸的识别造成了不便,从而降低了人脸识别的效率。此外,对于部分受损的人脸图像中人物的识别也变得更加困难。所以对于图像补全的研究也变得热门。

传统的图像补全和还原方法,主要依靠计算破损部分与周围像素的相似度,进行小范围的逐像素或者逐区块填充。A. Telea等^[1]提出了image inpainting算法,该算法从需要进行修补区域的边界开始,由边界到中心逐渐填充待补绘区域中的所有像素。Tang F.等^[2]提出了Criminisi算法,该算法通过缩小寻找匹配块的搜索范围并给出一种选择最优匹配块的准则以避免误差传播,但该算法在纹理与缺失部分并存时,容易发生纹理断裂和修补部位语义与全图不合的情况,导致训练效果不理想。J. Hays等^[3]提出引入外部数据库,利用外部数据库的内容检索相似部分,得到与缺失部分相似的内容进行填充,然而大量数据集中寻找匹配条件非常耗时,而且在缺失部分信息的复杂情况下,填充后的效果不一定尽如人意。基于这些问题,C. Barnes等^[4]提出了块匹配算法,逐区块搜索图像,每次搜索在指定的半径下随机选取一个值,然后收缩半径,直到搜索半径小于1个像素为止,最后对目标部分进行处理。由于块匹配算法对大面积处理范围效果不理想,He K. M.等^[5]提出了统计块概率修复方法,通过计算块概率填充空白区域实现较大面积的图像补全,取得了较好的效果。

虽然传统方法取得了一定的效果,但是面对图像纹理和内容复杂多变、大面积破损的情况下,补全效果并不理想。且在面对任意形状的修复任务中,传统的补全算法无法取得良好的效果。

随着生成对抗网络(generative adversarial network, GAN)^[6]的出现,研究者开始试图利用该方法,通过深度学习的方式训练修补缺失部分网络,对图像缺失部分进行基于深度学习的填充。生成对抗网络由Ian J. Goodfellow等人提出,算法基于“零和博弈”思想。GAN主要由两个部分组成,第一个部分是生成网络,输入数据并映射为图像;第二个部分是鉴别网络,对生成的数据进行真假判断。其中的生成网络希望最小化真实图像和生成图像的差距,而鉴别网络希望最大化真实图像和生成图像的差距。

在生成对抗网络思想的影响下,D. Pathak等^[7]提出了采用编码-解码(encoder-decoder)的管道模

型,通过基于上下文像素预测驱动的无监督的视觉特征学习算法,利用周围图像信息推断缺失部位。S. Iizuka等^[8]在上下文编码器的基础上,提出了全局鉴别器和局部鉴别器,强化对缺失部分的训练,同时使用膨胀卷积增大感受野,使网络能够得到距离更远的信息。而Yu J. H.等^[9]提出的上下文注意力网络之后,Zhang H.等^[10]提出的自注意力生成对抗网络中,解决了与周围区域不一致的扭曲结构或者模糊纹理的问题,既能合成新的图像结构,又能利用好周边的图像特征作为参考,让预测的结果更好。此后Yu J. H.等^[11]还结合上下文注意力机制,提出了门控卷积,判别器不再是局部和全局判别器的组合,而是使用了SN-Batch GAN的判别器,其生成图像的语义合理性和视觉连通性在大面积图像补全中是出色的。曹琨等^[12]提出了基于条件生成对抗网络的人脸补全算法,提取了人脸五官脸部特征,并将其作为约束信息输入生成器和鉴别器中,生成器负责重构,鉴别器结合损失函数生成完整人脸。该方法基于条件能作用于局部信息,但仍然只是较为简单的基于Encoder-Decoder的信息提取,而没能使用多个卷积块进行足够深入的特征提取。Xie C. H.等^[13]提出了基于可学习双向注意图的图像修复,作者认为基于卷积神经网络(convolution neural network, CNN)的算法并不能处理好非规则空洞修复,而且会产生颜色差异和模糊问题。提出前后向注意力图合并为可学习的双向注意力图提高图像的视觉质量,尽管效果上比较好,但由于没有边缘信息的约束,在简单纹理上的补全效果不错,但是针对人脸补全的任务表现一般。而在EdgeConnect中,K. Nazeri等^[14]提出了“先线条,后颜色”的理念,使用了基于边缘条件的图像复原。通过边缘修复网络得到修复好的边缘信息,然后利用所得到的边缘修复图像作为条件,拼接残缺图像数据,输入到图像修复网络,从而利用边缘信息通过图像修复网络还原图像。其本质就是将图像修复的步骤分为高频信息和低频信息的补全,但是它使用网络的鉴别器只有全局鉴别,修复细节比拥有局部网络和拥有人脸五官鉴别网络相比,在图像生成的纹理上比带有多个鉴别器的方法稍微逊色一些。

为了解决边缘补全方法中只关注图像整体性,而忽视补全部分和人脸五官的视觉连通性的问题,本文以EdgeConnect为基础,提出以下思路:

1) 提出人脸部位(眼睛、鼻子和嘴巴)鉴别器的方法,当补全边缘图像和破损图像经过图像补全网络的处理之后,对生成的补全人脸图的眼睛、鼻子和嘴巴进行鉴别,从而让人脸图像的五官部分更加符合

人们正常的认知。

2) 提出基于边缘条件增强掩膜和正常区域特征差异的门控卷积块, 在筛选有效区域和无效区域的同时, 通过对门控卷积提取特征和输入数据相加进行信息增强, 从而更加精准地区分正常区域和需要补全的区域。

3) 提出局部鉴别器, 当补全图像经过图像补全网络的处理之后, 让补全后的图像整体拥有更好的视觉连通性。

2 结合边缘条件补全与多个鉴别器组合的网络模型

本文提出的结合边缘条件补全与多个鉴别器组合的网络模型如图 1 所示。网络模型由两个部分组成, 分别是边缘补全网络和图像补全网络。其中边缘补全网络负责修补缺损人脸图像的边缘, 图像补全网络以补全边缘为条件, 对人脸图像进行补全。

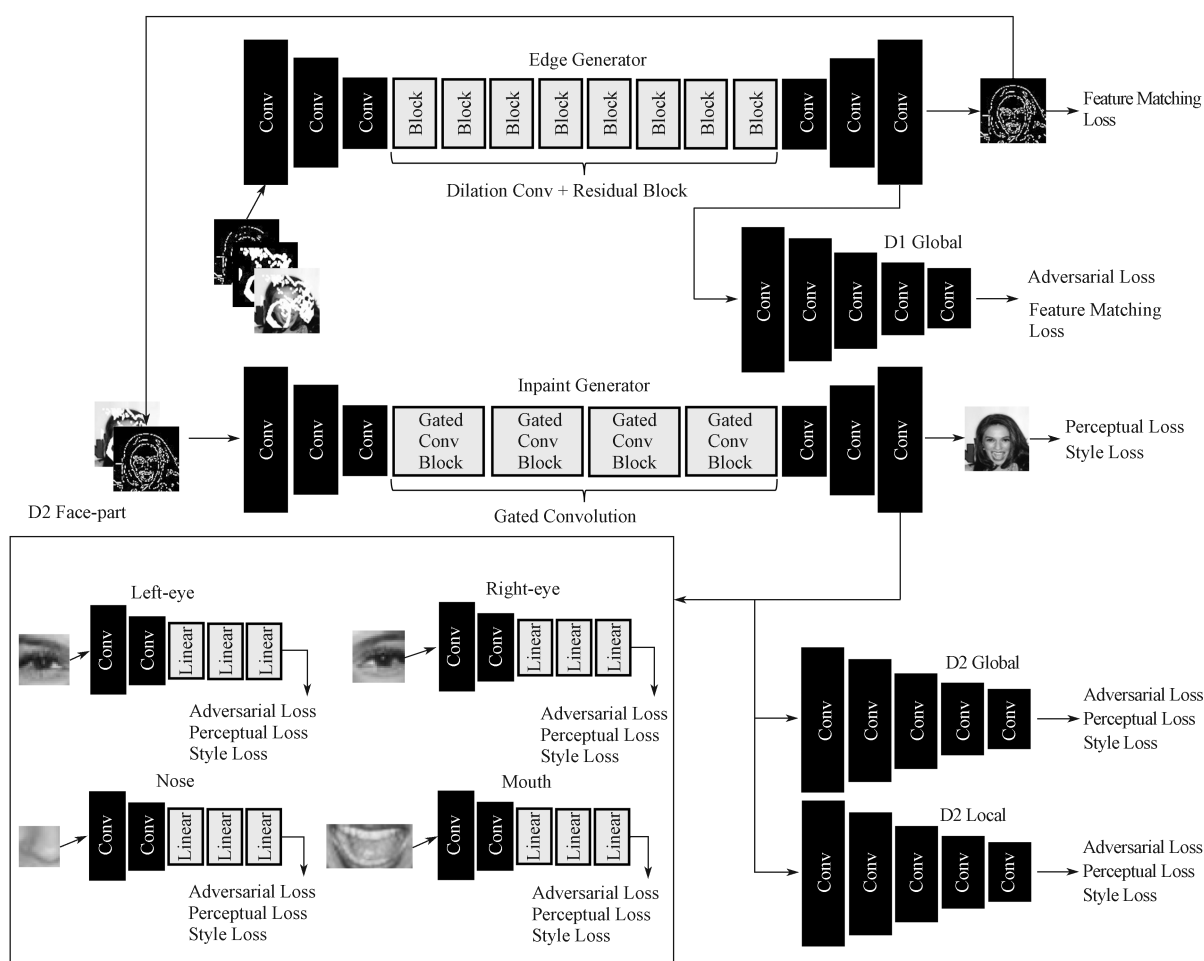


图 1 结合边缘条件补全与多个鉴别器组合的网络模型

Fig. 1 A network model with edge condition completion and multiple discriminators combined

模型对数据的处理分为 3 个步骤: 第一, 以没有掩膜的边缘图作为目标, 输入掩膜、带掩膜的 Canny 边缘图和带掩膜的灰度图, 通过训练边缘补全网络模型, 使得 Edge Generator 能够产生预测的补全边缘图; 第二, 使用边缘补全网络输出的预测补全边缘图, 结合残缺的彩色人脸图输入至图像补全网络, 输出预测的修复人脸图; 第三, 将边缘补全网络和图像补全网络进行级联, 以掩膜、带掩膜的边缘图和带掩膜的灰度图作为输入, 以真实图像作为目标进行联合训练,

实现端到端的图像补全。

在测试阶段, 通过加载已经训练完成的模型进行残缺人脸补全测试, 需要输入残缺人脸图和掩膜图, 经过级联后的人脸补全网络产生补全的人脸图像。

2.1 边缘补全网络

边缘补全网络由边缘补全生成网络 (edge generator, 简称 G1) 和边缘补全鉴别网络 (discriminator, 简称 D1) 组成。其中 G1 网络是生成补全边缘的图像, 输入掩膜灰度图 (被遮挡的灰度

人脸图)、掩膜边缘图(被遮掩的边缘图)和掩模图(残缺部分为白,完整部分为黑的二值图),以真实边缘为标签(训练目标),级联输入数据,并由 G1 运算之后,生成补全后的边缘图,然后通过 D1 计算对抗损失^[6],并与计算 D1 的每个卷积层通过 LeakyReLU 激活后得出的值之间的 L1 损失加起来,得出特征匹

配损失^[14],并用这个损失数值对关联它的网络进行反向传播。G1 的目的是要最小化生成边缘图和真实边缘图之间的差距,以提高生成器生成图像的质量,而 D1 则是要尽可能扩大这个差距,从而让 G1 的进步变得更加困难,当然得到的结果也会更好。

边缘补全网络结构如图 2 所示。

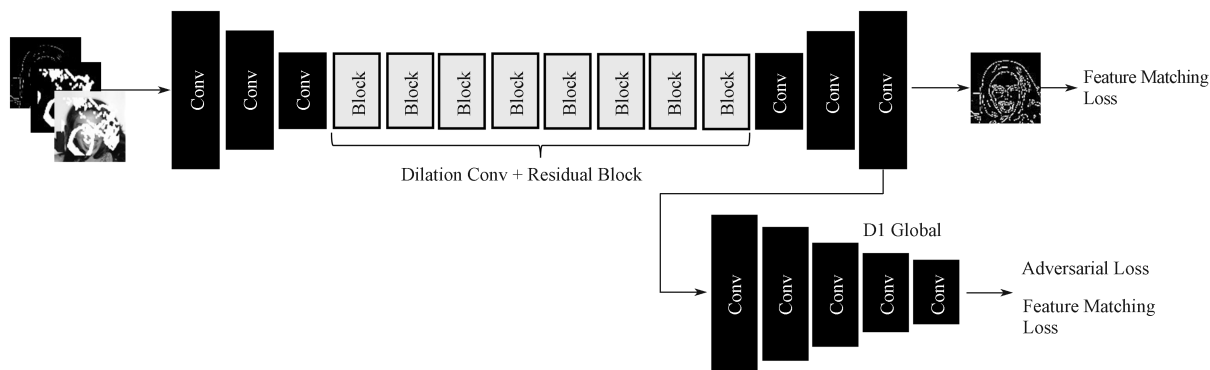


图 2 边缘补全网络结构图

Fig. 2 Edge-inpainting network structure diagram

如图 2 所示,边缘补全网络模型(其生成部分用 G1 代称)中, G1 由 3 个卷积层、8 个残差块、3 个反卷积层级联所组成。其中卷积层的卷积核大小分别为 7×7 、 4×4 和 4×4 ,反卷积层由大小为 4×4 、 4×4 和 7×7 的卷积核所组成,卷积层和反卷积层都在卷积操作之后拥有谱归一化(spectral normalization)处理和设置 ReLU 激活函数,卷积层的第一个卷积和反卷积层的最后一个卷积分别做了反射填充处理。

通过多层的卷积、残差网络^[15]和反卷积的标准化处理,对不同信息源进行融合,以提取深层次的边缘纹理结构特征。针对网络深度太大而导致的梯度消弭问题,使用残差块以避免。

而 D1 以真实边缘人脸图为目标(标签),与 G1 进行对抗,在 G1 能力越来越强的时候, D1 在参数设置合理的情况下,也能相应提升其对 G1 生成的补全边缘图的鉴别能力,利用 Canny 边缘检测器提取图像的边缘特征,作为鉴别器学习的符合预期的正样本,通过对网络 G1 生成的不符合实验预期的负样本和 Canny 边缘特征进行融合,以此提升 D1 的鉴别能力,监督 G1 生成更加符合原图的边缘信息图像。

2.2 基于自注意力机制的图像补全

2.2.1 图像补全网络的组成

图像补全网络由图像补全生成网络(inpaint generator,简称 G2)和图像补全鉴别网络(discriminator,简称 D2)组成。其中 G2 网络的作

用是生成补全图像,输入补全边缘图和残缺的人脸图像,以真实图像为标签,级联输入数据,并由 G2 运算之后,生成补全后的图像,然后通过 D2 的各个鉴别网络计算对抗损失^[6]、感知损失^[16]和风格损失^[16]。图像补全网络的损失是由 D2 Face-Part 网络对抗损失的总和,并用这个损失数值对关联它的网络进行反向传播。G2 的目的是要 G2 通过训练,对参数动态学习实现对有效区域和掩膜区域的有效区分,减少掩膜对图像补全的不利影响,从而让图像补全的颜色以及结构更加合理,最小化生成补全图和真实图之间的差距,以提高生成器生成图像的质量。而 D2 则是要尽可能扩大这个差距,从而让 G2 与 D2 鉴别网络组合对抗,并促进 G2 进步。图像补全网络结构见图 3。

如图 3 所示, G2 由 3 个卷积层、4 个提出的门控卷积、3 个反卷积层级联组成。其中卷积层的卷积核大小分别为 7×7 、 4×4 和 4×4 ,反卷积层分别由大小为 4×4 、 4×4 和 7×7 的卷积核所组成,卷积层和反卷积层都在卷积操作之后拥有谱归一化处理和设置 LeakyReLU 激活函数,卷积层的第一个卷积和反卷积层的最后一个卷积分别做了反射填充处理。

D2 以真实人脸图为目标,与 G2 进行对抗,在 G2 能力越来越强的时候, D2 在参数设置合理的情况下,能相应提升其对 G2 生成的补全图像鉴别能力,利用 G2 生成的掩膜边缘特征对图像进行补全。作为鉴别器学习的符合预期的正样本,通过对网络 G2 生成的不符合实验预期的负样本和掩膜边缘特征进行

融合, 以此提升 D2 的鉴别能力, 监督 G2 生成更符合原图的图像。

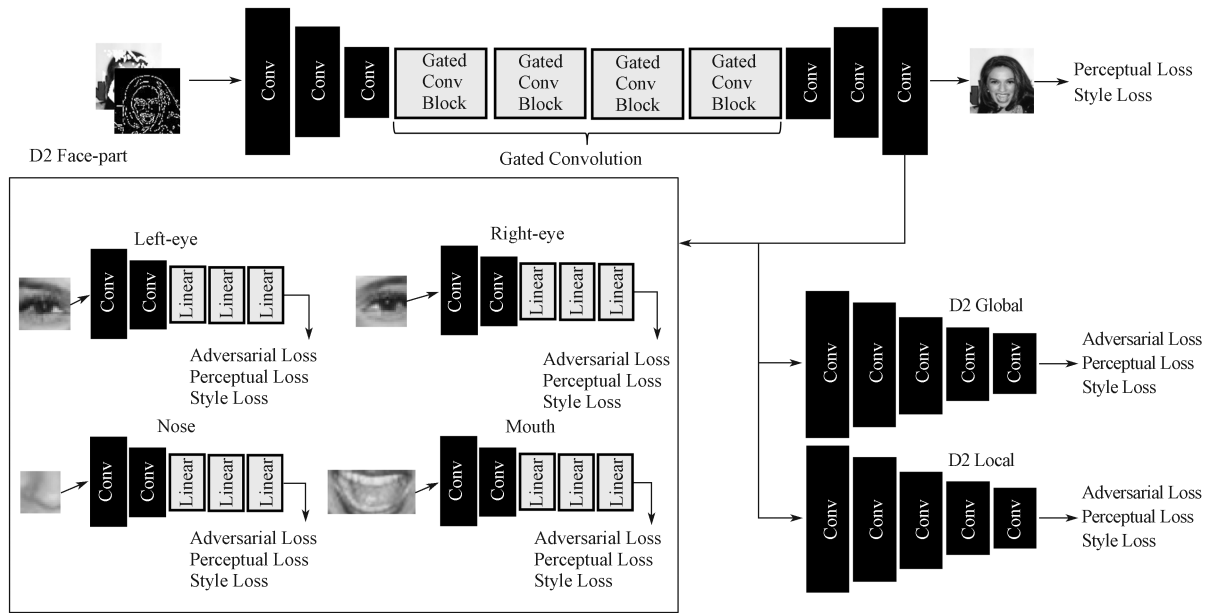


图 3 图像补全网络结构图

Fig. 3 Image-inpainting network structure diagram

2.2.2 自注意力机制的设计与实现

普通卷积在任意掩膜下补全图像有局限性, 它以滑动方式提取局部特征, 滑动窗口下的像素被默认为都是有效的。但对于图像补全而言, 窗口下包含掩膜的边界时, 其无效像素和有效像素均会被卷积窗口处理, 从而导致信息模糊, 以及出现补全部分语义不符合实际结果的现象。

门控卷积 (gated convolution) 通过对应每个通道内的可学习的特征选择机制, 将所有层响应空间位置的像素融合到一起, 达到对部分卷积进行泛化的目的, 从而能增进正常区域和掩膜区域的过渡, 解决将所有像素视为有效像素的问题。文献 [17] 引用了基于门控卷积的部分卷积算法, 虽然有效地区分了正常部分和需要补全的部分, 但是其所使用的部分卷积公式^[11]是不可导的, 会导致反向传播的梯度数据出现过或者过小的极端情况, 从而导致训练过程具有不稳定性, 增加了训练过程的不稳定性。所以对此提出可导的部分卷积操作替代 EdgeConnect^[14] 的普通卷积, 式 (1) 为部分卷积操作公式。

$$\begin{cases} \mathbf{M}_c = \mathbf{M} \otimes k_{\frac{1}{9}}, \\ \mathbf{F}_{\text{conv}} = \mathbf{W}^T \mathbf{F}^i, \\ \mathbf{F}_{\text{out}} = \mathbf{F}_{\text{conv}} \odot g_A(\mathbf{M}_c), \\ \mathbf{M}' = g_M(\mathbf{M}_c). \end{cases} \quad (1)$$

式中: 后两个公式为部分卷积公式; $\mathbf{F}_{\text{conv}}$ 为当前滑动窗口的特征值; $\mathbf{F}_{\text{out}}$ 为 $\mathbf{F}_{\text{conv}}$ 与重整化函数的部分卷

积同或运算后的特征值; $\mathbf{F}^i$ 为输入当前窗口滑动的特征值; $\mathbf{W}^T$ 为卷积层的转置权重; $\mathbf{M}$ 为当前的掩膜图; $\mathbf{M}_c$ 为用卷积核当前的 $1/9$ 得到的掩膜图, 然后判定是否大于 0 决定是否更新为 1; $\mathbf{M}'$ 为更新之后的掩膜; g_A 和 g_M 分别为:

$$\begin{cases} g_A(x) = \begin{cases} a \exp(-\gamma_l(x-\mu)^2), & \text{if } x < \mu, \\ 1 + (a-1) \exp(-\gamma_r(x-\mu)^2), & \text{else;} \end{cases} \\ g_M(x) = (\text{LeakyRelu}(x))^{\alpha}. \end{cases} \quad (2)$$

式中: g_M 为更新掩膜图操作; g_A 为重整化 (Renorm) 操作, 其下标 A 和 M 分别是归一化权重和掩膜图的矩阵; $x = \mathbf{M}_c$ 、 $\alpha = 1.1$ 、 $\mu = 2.0$ 、 $\gamma_l = 1.0$ 、 $\gamma_r = 1.0$ 、 a 默认取 1.0, 它们皆为用户自定义常量。

这些 soft 将会随着其结果被导入公式 (1)。每次通过部分卷积的操作后, 用式 (2) 更新掩膜图。公式 (2) 是式 (1) 的附属公式, 代入式 (2) 产生结果至式 (1), 得出更平滑的部分卷积进行反向传播, 增强训练过程的稳定性。

门控卷积的操作公式如下:

$$\begin{cases} \mathbf{G}_{y,x} = \sum \sum \mathbf{W}_g \cdot \mathbf{I}, \\ \mathbf{F}_{y,x} = \sum \sum \mathbf{W}_f \cdot \mathbf{I}, \\ \mathbf{O}_{y,x} = \phi(\mathbf{F}_{y,x}) \odot \phi(\mathbf{G}_{y,x}). \end{cases} \quad (3)$$

式中: $\mathbf{F}_{y,x}$ 和 $\mathbf{G}_{y,x}$ 分别为 Feature 卷积和 Gating 卷积; y, x 为当前滑动区域的中心点坐标; $\mathbf{W}_g$ 为选取掩膜和非掩膜空间的卷积滤波器; $\mathbf{W}_f$ 为选取掩膜和非掩

膜空间人脸图像过渡区的卷积滤波器; I/O 分别为输入卷积滤波和输出卷积滤波。

Gating 卷积和 Sigmoid 激活函数实现动态特征选择, 即选取掩膜覆盖空间和正常空间; Feature 卷积和 LeakyReLU 激活函数实现特征提取, 即选取部

分的掩膜和非掩膜区的过渡图, 然后通过 Gating 和 Feature 的点乘, 更有效地选取到有用的信息。而提出的门控卷积具有更强的像素选择性, 使得卷积能够在更大范围的像素缺失下仍能精确描述局部特征。

普通卷积和门控卷积的结构如图 4 所示。

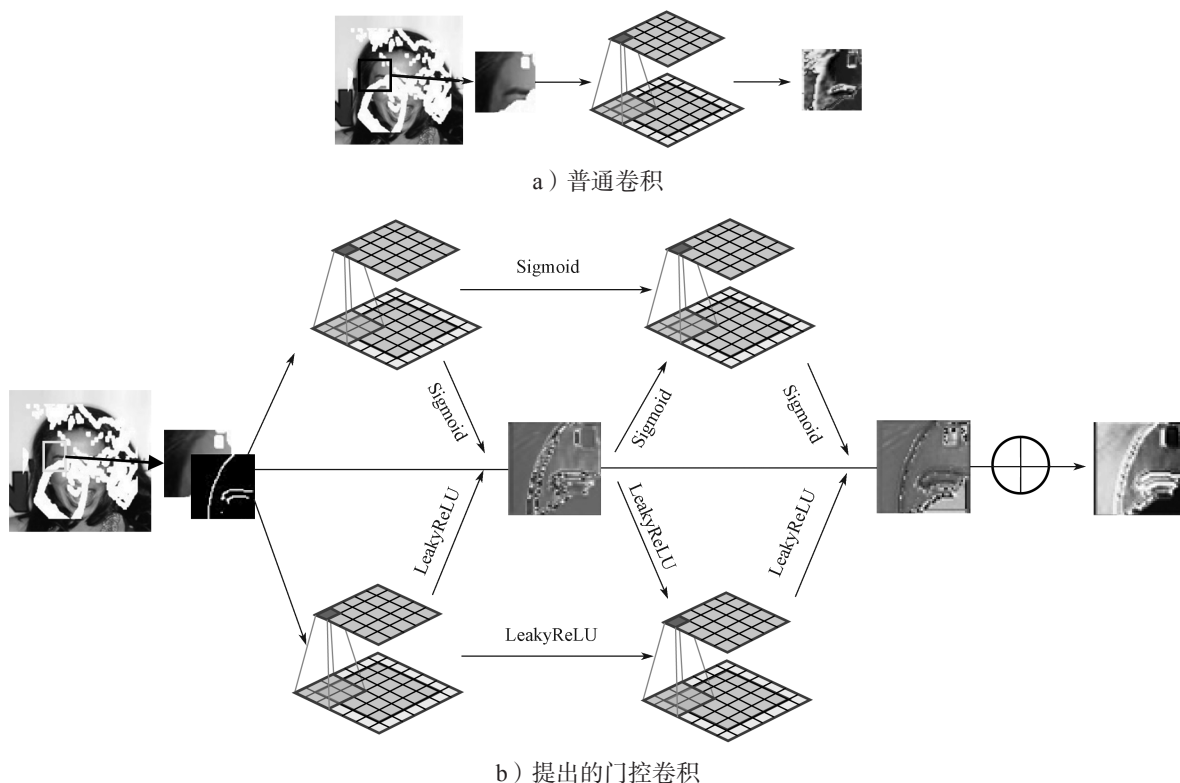


图 4 普通卷积和门控卷积的结构示意图

Fig. 4 Structures of normal convolution and gated convolution

针对门控卷积对于掩膜边界两侧仍有少量平滑的过渡区, 导致了门控卷积的卷积窗口也出现了对部分无效像素视为有效像素的问题, 提出了使用自注意力机制强化人脸遮挡部分和完整部分的区别。

提出的自注意力机制能够通过对其中的两个子门控卷积产生的不同特征与输入相加, 得出掩膜边界两侧更加明确的特征差异, 处理掩膜区和非掩膜区的过渡。门控卷积不但拥有残差块所拥有的长距离下保地留特征的能力, 同时具备门控卷积本身所自带的增强边缘两侧区分特征的能力, 从而能更加精确补全掩膜部分, 缓解补全区域和正常区域的边界像素影响网络的判断。

3 面向部分补全的鉴别器网络

3.1 结合结构与纹理的局部鉴别网络

针对图像补全网络只有全局鉴别网络而导致的补全部分还原不理想问题, 提出了局部鉴别网络。

提出的局部鉴别网络 (见图 5) 是图像补全网络

中鉴别网络模块的一部分, 它的网络结构和全局鉴别网络一致, 由 5 层卷积块组成, 每层卷积块包含有一层谱归一化的卷积层和 LeakyReLU 激活层 (最后一个卷积块没有 LeakyReLU 激活层)。

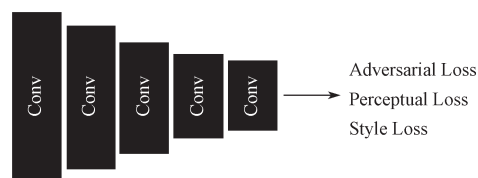


图 5 局部鉴别网络结构图

Fig. 5 Structure of local discriminator in inpaint network

由补全部分和对应位置的真实部分输入全局鉴别网络, 生成两个 15×15 的矩阵, 然后将这些矩阵输入对抗网络。过程由局部对抗损失式 (4) 进行计算, 通过公式分别计算各自的损失函数:

$$L_{adv, l} = L_{adv}(I_{gt, l}, I_{pred, l}) = E_{gt, l} [\ln D_2(I_{gt, l})] + E_{pred, l} \{ \ln [1 - D_2(I_{pred, l})] \} \quad (4)$$

式中: $L_{adv, l}$ 为局部对抗损失; L_{adv} 为对抗损失函

数; $I_{gt,l}$ 为真实局部图像; $I_{pred,l}$ 为补全局部图像。

$E_{gt,l}[\ln D_2(I_{gt,l})]$ 和 $E_{pred,l}\{\ln[1-D_2(I_{pred,l})]\}$ 分别为真实局部图像和补全局部图像的分布函数, 其中, $E_{gt,l}$ 为局部真实部分的损失, $E_{pred,l}$ 为局部预测部分的损失, D_2 为图像补全网络中的鉴别网络损失。

文中约定, 在该类对抗损失函数中, 其意义仅有适用范围的区别, 如下标“pred”代表这个数据是网络输出的预测结果; 下标“gt”即为对应 pred 类型的真实数据; 下标标注了“l”即为满足 D2-Local 的输入; 下标“fp”为符合 D2-Facepart 的输入格式。

该公式即为局部对抗网络的损失函数。它负责鉴别补全部分的真实程度, 从而使得补全结果不至于偏离真实性。

一般的局部鉴别网络基于鉴别真实性, 但因为只判断补全部分本身的真实程度, 对生成部分的风格制约力较弱, 导致生成部分影响全图的视觉连通性。而且在高层重建时, 颜色、纹理和确切的形状等因素并没有被考虑到。所以提出了使用风格损失和感知损失进行额外地约束。局部鉴别网络的风格损失计算公式如下:

$$L_{style,l} = L_{style}(\hat{y}, y) = \sum_{i=1}^4 \|G_i^\phi(\hat{y}) - G_i^\phi(y)\|_1, \quad (5)$$

式中: $L_{style,l}$ 为局部风格损失; L_{style} 为风格损失函数; $\hat{y}$ 和 y 分别为补全局部图像和对应补全部分的真实局部图像; G_i^ϕ 是对应 i 值的网络层, 其中的 i 取 1~4, 分别对应于预先训练的 VGG-19^[18] 网络的 relu2_2、relu3_4、relu4_4、relu5_2 层的激活图, ϕ 为随 i 取的网络层; $\|\cdot\|_1$ 为 1 范数, 即直线距离。

激活映射还用于计算风格损失, 用于测量激活映射协方差之间的差异。通过各个图像特征的欧几里得距离来衡量在感知上与真实部分不相似的程度。

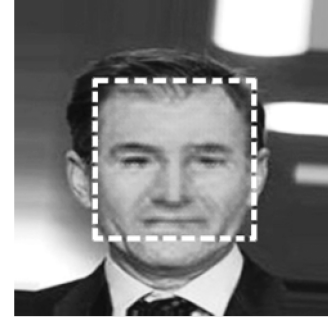
风格损失虽然一定程度上纠正了纹理和像素的补全误差, 但是对于图像补全部分的形状和结构没有很好保留。为了解决风格损失只保留纹理和颜色信息, 而没有对形状和结构信息进行有效保留的问题, 提出感知损失, 以对生成结果进行结构和形状上的限制。局部鉴别网络的感知损失如式 (6):

$$L_{perc,l} = L_{perc}(I_{gt,l}, I_{pred,l}) = \sum_{i=1}^4 \|\Phi_i(I_{gt,l}) - \Phi_i(I_{pred,l})\|_1. \quad (6)$$

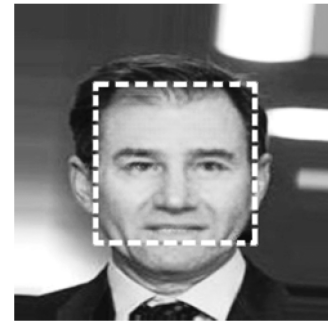
式中: $L_{perc,l}$ 为局部感知损失; L_{perc} 为感知损失函数; Φ_i 为网络层, Φ_i 中的 i 对应预训练网络第 i 层的激活图, 即应在 ImageNet 数据集上预先训练的 VGG-19^[18] 网络的 relu1_1、relu2_1、relu3_1、relu4_1 和

relu5_1 层的激活层。

从各个激活映射提取结构和形状特征进行计算, 以测量激活之间的欧几里得距离。从而利于高层信息的重构。不同人像补全方法对图像局部的影响如图 6 所示。



a) 对抗损失



b) 结合损失



c) 原图

图 6 不同人像补全方法对图像局部的影响

Fig. 6 Influence of different portrait completion methods on the local image

如图 6 所示, 卷积滤波会结合非掩膜区域的信息, 通过对风格、纹理和结构进行预测, 为信息重构提供依据, 根据这些信息进行卷积, 从而使得补全部分的风格与全图更加贴近, 纹理上更接近真实的人脸图, 结构上更加贴近人脸部位的真实轮廓。

将式 (4)~(6) 相加, 得到如下局部鉴别网络的总损失函数:

$$L_{local} = L_{adv,l} + L_{style,l} + L_{perc,l}. \quad (7)$$

提出局部鉴别网络的主要作用, 是对图像补全后

的部分进行对抗计算, 得出补全部位的对抗损失, 从而起到对补全部位语义合理性的保障作用。

3.2 基于人脸局部部位约束的鉴别网络

由于一般的局部鉴别网络本质上和全局鉴别网络一样, 只关注整体性, 而且局部鉴别网络在人脸补全时, 很难对人脸图像的纹理进行细节描述。因而提出一个特殊的局部鉴别网络用于专门处理人脸部位

的生成细节, 让补全部分与全图符合视觉连通性。

人脸部位鉴别网络 (D2 Face-Part) (图7) 是特殊的局部鉴别网络, 人脸部位鉴别网络一共由4个子网络组成, 这4个子网络分别是左眼鉴别网络、右眼鉴别网络、鼻子鉴别网络和嘴巴鉴别网络。它针对了人脸的关键部分, 即左眼、右眼、嘴巴和鼻子, 也就是我们所说的“五官”。

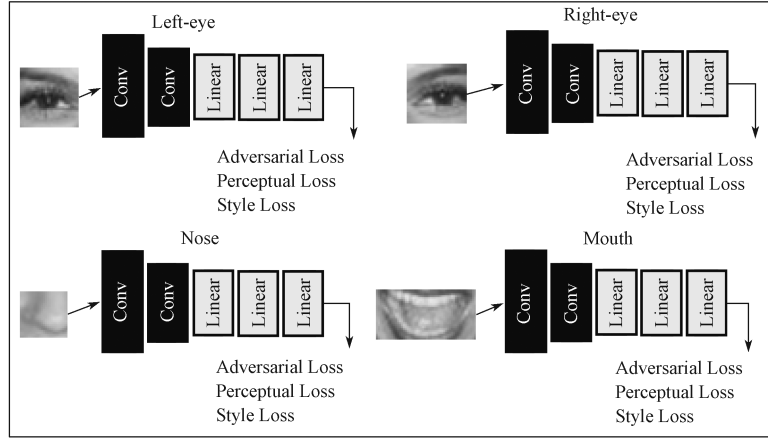


图7 人脸部位鉴别网络结构图

Fig. 7 Structures of face-part discriminator in inpainting network

人脸部位鉴别网络会将补全人脸图像和真实人脸图像的左眼、右眼、鼻子和嘴巴提取, 并送入人脸部位鉴别网络, 该网络会分别送进对应的4个网络, 最后生成4个得分, 然后分别计算各自的对抗损失, 对抗损失以1为真, 以0为假。这些得分经过对抗损失计算后, 将这4个对抗损失数值相加并取平均, 从而得到公式(8):

$$L_{adv,fp} = L_{adv}(I_{gt}, I_{pred}) = \frac{1}{4} \sum_{i=1}^4 \{ E_{gt,i} [\ln D_2(I_{gt,i})] + E_{pred,i} [1 - \ln D_2(I_{pred,i})] \} \quad (8)$$

式中: $L_{adv,fp}$ 为脸部部位对抗损失; $E_{gt,i} [\ln D_2(I_{gt,i})]$ 和 $E_{pred,i} [1 - \ln D_2(I_{pred,i})]$ 分别代表真实人脸部位图像的分布函数和补全人脸部位图像的分布函数, 其中 $I_{gt,i}$ 为第 i 个真实人脸部位; $I_{pred,i}$ 为第 i 个补全图像的人脸部位。

式(8)即为 D2 Face-Part 与 G2 对抗的损失函数。这个函数的作用是希望 G2 (图像补全生成网络) 在训练的时候更加关注人脸“五官”的生成可靠性, 从而在人脸部位的微观层面上更加重视其效果。

与改进的局部网络相同, 为了克服只有对抗损失所带来的仅关注图像本身的真假, 而不关注其纹理、结构的弊端。提出对人脸每个部位的鉴别器使用风格损失函数和感知损失函数, 以此使得生成的部分令全

图拥有更良好的视觉连通性。式(9)和式(10)分别为人脸部位的风格损失函数和感知损失函数:

$$L_{style,fp} = L_{style}^{\phi,j}(\hat{y}, y) = \frac{1}{4} \sum_{j=1}^4 \sum_{i=1}^4 \|G_j^{\phi}(\hat{y}_i) - G_j^{\phi}(y_i)\|_1, \quad (9)$$

$$L_{perc,fp} = L_{perc}^{\phi,j}(I_{gt,fp}, I_{pred,fp}) = \frac{1}{4} \sum_{j=1}^5 \sum_{i=1}^4 \|\Phi_j(I_{gt,i}) - \Phi_j(I_{pred,i})\|_1. \quad (10)$$

式中: $L_{style,fp}$ 为脸部部位风格损失; $L_{style}^{\phi,j}$ 为脸部部位风格损失函数; $\hat{y}_i$ 和 y_i 为预测人脸部位与真实人脸部位; $L_{perc,fp}$ 为脸部部位感知损失; $L_{perc}^{\phi,j}$ 为脸部部位感知损失函数; $I_{gt,fp}$ 为真实人脸部位; $I_{pred,fp}$ 为预测人脸部位; $I_{gt,i}$ 为真实人脸部位; $I_{pred,i}$ 为预测人脸部位。

式(9)通过激活映射计算风格损失, 以测量激活映射协方差之间的差异。通过各个图像特征的欧几里得距离来衡量在感知上与真实部分不相似的程度。计算中所提取的特征被作为风格进行欧几里得距离计算差异程度, 以此对补全的人脸部位进行纹理上的约束, 从而保证了纹理和像素上均符合全图的主题。式(10)对补全后人脸图像整体的“五官”进行约束, 提升补全后“五官”的视觉连通性。

测量激活映射协方差之间的差异。通过各个图像特征的欧几里得距离来衡量在感知上与真实部分的相似程度。计算中所提取的特征被作为人脸图像的

风格特征,进行欧几里得距离计算真实人脸部位和补全人脸部位之间风格上的差异程度,以此对补全的人脸部位进行纹理上的约束,从而保证了纹理和像素上符合全图的主题。对补全后人脸图像整体的“五官”进行约束,提升补全后“五官”的视觉连通性。人脸部位鉴别网络还通过网络内已经被归一化处理的卷积滤波,利用人脸部位未遮掩部分的风格进行风格重构,从而让人脸五官在清晰度和纹理上更加与人脸图像的主题贴近。

D2 Face-Part 还通过网络内已经被归一化处理的卷积滤波,利用人脸部位未遮掩部分的风格进行风格重构,从而让人脸五官在清晰度和纹理上与人脸图像的主题更加贴近(图8)。



a) 对抗损失



b) 结合损失



c) 不同补全方法对人脸五官的影响

图8 使用不同损失方法的补全效果对比

Fig. 8 Comparison of inpainting effect using different loss methods

将式(8)~(10)相加,得到G2与D2 Face-Part生成对抗的总损失:

$$L_{fp} = L_{adv,fp} + L_{style,fp} + L_{perc,fp} \quad (11)$$

总体来看,面向图像补全的鉴别网络由3个鉴别网络组成,它们分别从人脸图像补全后的整体性、人脸图像补全部分的合理性、以及人脸图像“五官”生成,更加符合逻辑。利用D2 Face-Part对人脸的结构进行分析,从而让眼睛、鼻子和嘴巴在结构上更加地接近真实形态。通过外部损失的约束,实现对G2在整体、局部以及纹理、结构和像素上形成多方面的监督效果。

4 试验

4.1 试验设计

使用CelebA^[19]中的202 599张数据集进行训练、测试和评估,在训练之前,对每张人脸图重新标度为256像素×56像素×3像素。实验环境以Windows10为平台,配有Pytorch1.7,在Python3.8上实现,处理器为i5-9400F 2.9 GHz,内存为32 GB,显卡为RTX2070 SUPER 8 GB。学习率默认取0.000 1,采用Adam^[20]梯度下降法反向传播更新梯度,其Beta1与Beta2分别取0.0和0.9。

4.2 试验结果和分析

本文采用了近年来比较经典且表现效果良好的算法进行对比研究,分别是文献[7]的Context-Encoder方法、文献[8]的Globally-Locally方法和文献[14]的EdgeConnect方法。为了直观地表现本文提出补全算法的优越性,使用峰值信噪比(peak signal-to-noise ratio, PSNR)衡量补全后的图像与原图像之间的差距,PSNR的数值越大,则补全图像的表现效果越好。同时,为了体现本文提出的算法在补全图像的结构真实程度,引用结构相似性(structural similarity, SSIM)衡量补全图像与原图结构的差距,SSIM以0为最低分,分数越高,表明补全图像的结构在逻辑上就越符合原图结构的标准,最高分是1。图9为块遮挡、随机遮挡和随机遮挡下五官的补全效果图。

图9a和图9b分别是块遮挡补全和非规则遮挡补全的结果,共有8行7列,行代表不同的测试图像,列代表同一类型或者同一个方法在不同图像的表现。其中第1列代表被特定掩膜遮挡的原图,第2列是特定类型遮挡原图的掩膜图,第3列到第6列分别是文献[7]、文献[8]、文献[14]和本文提出方法在不同掩膜遮挡下的补全结果,第7列是原图。

图9c共5行4列,每一行代表不同的方法实现的结果,分别是文献[7]的Context-Encoder方法、文献[8]的Globally-Locally方法、文献[14]的

EdgeConnect 方法、本文提出的方法和原图。每一列则分别代表左眼、右眼、鼻子和嘴巴。



a) 块遮挡补全结果



b) 随机遮挡补全结果



c) 随机遮挡下的五官还原结果

图9 块遮挡、随机遮挡和随机遮挡下五官的补全效果图

Fig. 9 Restoration results of facial features under block occlusion, random occlusion and random occlusion

4.2.1 遮挡还原补全结果分析

不同方法对块遮挡、非规则遮挡状况的试验结果, 如表 1~2 所示。由表 1 和 2 的结果可看出, 本文提出的方法优于其它的对比组。图 9a 和图 9b 中第 3 列图像补全后仍然存在部分噪声, 对细节的还原水平总体较差, 并且生成部分的色调和光暗度明显与全图不同。第 4 列的人脸特征过于扁平, 人脸部位的特

征还原度较低, 并没有很好的体现人脸的五官。第 5 列在随机遮挡的条件下还原效果较好, 但在块遮挡下没能很好地还原人脸的轮廓, 而且在补全色调上没有很好地契合全图的基本色调。第 6 列在对比组的基础上, 除了基本避免了噪声的出现外, 还在补全部分的色调和人脸轮廓上更加符合原图, 且在人脸的五官的细节纹理上表现得更好。

表 1 不同方法对块遮挡状况补全结果

Table 1 Block occlusion occlusion results under different methods

使用方法	PSNR	SSIM
Context-Encoder	21.03	0.946 7
Globally-Locally	26.88	0.921 8
EdgeConnect	23.96	0.865 3
Ours	27.89	0.950 8
平均	24.94	0.921 2

表 2 不同方法对非规则遮挡状况补全结果

Table 2 Results of irregular occlusion under different methods

使用方法	PSNR	SSIM
Context-Encoder	14.89	0.782 4
Globally-Locally	18.08	0.829 7
EdgeConnect	21.00	0.773 1
Ours	22.76	0.887 1
平均	19.18	0.818 0

4.2.2 人脸部位补全结果分析

表 3 为在非规则遮挡下人脸特征补全结果。

表 3 非规则遮挡下人脸特征部位补全结果

Table 3 Results of facial feature location (irregular occlusion)

使用方法	人脸位置 (PSNR/SSIM)			
	左眼	右眼	鼻子	嘴
Context-Encoder	18.99/0.679 8	18.99/0.679 8	18.99/0.679 8	18.99/0.679 8
Globally-Locally	18.66/0.667 5	18.66/0.667 5	18.66/0.667 5	18.66/0.667 5
Edge-Connect	16.16/0.664 0	16.16/0.664 0	16.16/0.664 0	16.16/0.664 0
Ours	19.37/0.705 1	19.37/0.705 1	19.37/0.705 1	19.37/0.705 1
平均	18.30/0.679 1	18.30/0.679 1	18.30/0.679 1	18.30/0.679 1

结合图 9 对表 3 显示的结果进行分析, 图 9c 中的第 1 行补全后的人脸部位细节模糊, 而且有明显的噪声, 结构上明显不清晰。第 2 行补全结构不明显, 而且补全纹理偏于扁平, 人脸部位体现不明显。第 3 行色调表达过于偏离全图, 而且对人脸部位的补全细节不理想。第 4 行在有遮挡物的情况下还原效果稍差, 但总体上对人脸部位补全效果相比对照组更佳, 色调、光暗度上均符合视觉连通性, 纹理结构相比对照组更加清晰。

5 结语

对图像修复网络增加了局部鉴别和人脸部位鉴

别的网络,以 EdgeConnect 构成的网络为基础,将图像修复网络的残差块修改为门控卷积块,同时提出局部鉴别网络和人脸部位鉴别网络,并将其整合进图像补全网络。在大面积破损、结构复杂以及随机噪声的干扰下,仍然能完成对图像的补全或者还原,补全之后与原图的对比结果较为良好。

但对于人脸面对的方向不同的情况时,其还原效果不尽如人意。对于出现的问题,课题组将会进一步研究如何在训练时充分考虑人脸方向的问题,这也是我们以后将要改进的方向。

参考文献:

- [1] TELEA A. An Image Inpainting Technique Based on the Fast Marching Method[J]. Journal of Graphics Tools, 2004, 9(1): 23-34.
- [2] TANG F, YING Y T, WANG J, et al. A Novel Texture Synthesis Based Algorithm for Object Removal in Photographs[C]// Advances in Computer Science - ASIAN 2004. Higher-Level Decision Making. Heidelberg: Springer, 2004: 248-258.
- [3] HAYS J, EFROS A A. Scene Completion Using Millions of Photographs[J]. ACM Transactions on Graphics, 2007, 26(3): 4.
- [4] BARNES C, GOLDMAN D B, SHECHTMAN E, et al. The PatchMatch Randomized Matching Algorithm for Image Manipulation[J]. Communications of the ACM, 2011, 54(11): 103-110.
- [5] HE K M, SUN J. Computing Nearest-Neighbor Fields Via Propagation-Assisted KD-Trees[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012: 111-118.
- [6] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.
- [7] PATHAK D, KRÄHENBÜHL P, DONAHUE J, et al. Context Encoders: Feature Learning by Inpainting[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos: IEEE Computer Society Press, 2016: 2536-2544.
- [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] YU J H, LIN Z, YANG J M, et al. Generative Image Inpainting with Contextual Attention[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5505-5514.
- [10] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-Attention Generative Adversarial Networks[EB/OL]. [2021-03-02]. <https://arxiv.org/abs/1805.08318v2>.
- [11] YU J H, LIN Z, YANG J M, et al. Free-Form Image Inpainting with Gated Convolution[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019, DOI: 10.1109/ICCV.2019.00457.
- [12] 曹 琨, 吴 飞, 骆立志, 等. 基于条件生成对抗网络的人脸补全算法 [J]. 传感器与微系统, 2019, 38(6): 129-132.
CAO Kun, WU Fei, LUO Lizhi, et al. Face Completion Algorithm Based on Condition Generation Adversarial Network[J]. Transducer and Microsystem Technologies, 2019, 38(6): 129-132.
- [13] XIE C H, LIU S H, LI C, et al. Image Inpainting with Learnable Bidirectional Attention Maps[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 8857-8866.
- [14] NAZERI K, NG E, JOSEPH T, et al. EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning[EB/OL]. [2020-03-11]: ArXiv: 1901.00212[Cs.CV]. <https://arxiv.org/abs/1901.00212>.
- [15] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [16] JOHNSON J, ALAHI A, LI F F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[C]// Computer Vision – ECCV 2016. Amsterdam: Springer International Publishing, 2016: 694-711.
- [17] 王富平, 李文楼, 刘 颖, 等. 结合边缘信息和门卷积的人脸修复算法 [J]. 计算机科学与探索, 2021, 15(1): 150-162.
WANG Fuping, LI Wenlou, LIU Ying, et al. Face Inpainting Algorithm Combining Edge Information with Gated Convolution[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(1): 150-162.
- [18] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2015, 4(10): 1409-1421.
- [19] LIU Z W, LUO P, WANG X G, et al. Deep Learning Face Attributes in the Wild[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 3730-3738.
- [20] KINGMA D, BA J. Adam: a Method for Stochastic Optimization[J]. Computer Research Repository, 2014, 12: 1-6.

(责任编辑: 申 剑)