

# 一种基于局部属性生成对抗网络的人脸修复算法

蒋 斌 刘虹雨 杨 超 涂文轩 赵子龙  
(湖南大学信息科学与工程学院 长沙 410082)  
(jiangbin@hnu.edu.cn)

## A Face Inpainting Algorithm with Local Attribute Generative Adversarial Networks

Jiang Bin, Liu Hongyu, Yang Chao, Tu Wenxuan, and Zhao Zilong  
(College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082)

**Abstract** Recent researches in neural network models have shown great potential in image inpainting task, which focuses on understanding image semantic information and reconstructs missing image content. These researches can generate visually reasonable image structures and textures, however, they usually produce distorted structures or blurry textures that are inconsistent with the surrounding areas, especially for the face inpainting task. The face inpainting task is often necessary to gene the advantages of fully connected convolution and U-net network, and the model proposes locally attributes discriminator to make the inpainted contents more innovative and is able to keep the global and local semantic consistency. The model not only improves the perception of the overall semantic information of the face image, but also restores the key parts of the face based on the local attributes. Experiments on the CelebA dataset have shown that our model can effectively deal with face image repair problems and generate novel results.

**Key words** neural network; face inpainting; local attributes discriminator; fully connected convolution; U-net

**摘 要** 最近对神经网络模型的研究在图像修复任务中显示出巨大的潜力,其核心任务是理解图像语义信息并重建缺失的图像内容.这些研究可以生成语义和内容上合理的结构和纹理,但通常会导致与孔洞周围区域不一致的扭曲结构或模糊纹理,特别是人脸图像修复问题.人脸图像修复工作经常需要为包含大量外观元素以及局部属性的缺失区域(例如眼睛或嘴巴)生成语义上的新内容,这些缺失区域往往具有独特的属性和语义信息从而导致生成内容不合理.为了解决以上问题,提出了一个有效的深度神经网络模型,模型的生成器结合全连接卷积和 U-net 网络的优越特性,同时提出局部属性辨别器使修复内容具有创新性的同时也能够使整体与局部保持语义一致性.模型不仅提升了对于人脸图像整体语义信息的感知能力,同时也基于局部属性能够有效地修复人脸关键部位,通过在 CelebA 数据集上的实验证明了该模型能够有效地修复人脸缺失部分并且能够生成新颖的修复内容.

**关键词** 神经网络;人脸修复;局部属性辨别器;全链接卷积;U-net

**中图法分类号** TP391

**收稿日期:**2018-09-18;**修回日期:**2019-05-22  
**基金项目:**国家自然科学基金项目(61702176);湖南省自然科学基金项目(2017JJ3038)

This work was supported by the National Natural Science Foundation of China (61702176) and the Natural Science Foundation of Hunan Province of China (2017JJ3038).

图像修复又被称为图像还原,主要任务是为图像缺失的区域填充合理的像素.该技术被应用于多种场景,如用于去除图像中的多余内容,或者利用一个图像中的内容和语义来填充另一个图像中的空缺区域.图像修复的核心挑战是为与上下内容和边界相关的缺失区域创建语义上合理的内容以及逼真的纹理细节.

传统的图像修复工程<sup>[1-8]</sup>尝试使用纹理合成技术<sup>[1]</sup>解决问题,利用已知区域中与未知区域最相关的部分修补孔洞或从孔洞边界开始传播并且匹配合理像素值.在文献[4]中,Barnes等人提出了 Patch-Match 算法,通过迭代搜索最相似的补丁来重建缺失的区域;文献[2]进一步精确了搜索区域并更快速地匹配最佳补丁.但是这些方法没有视觉语义的概念,只是利用已知背景的低级信号来模拟未知像素.另外一些早期工作<sup>[9-11]</sup>的重点是利用与需修补图像相似图像中的内容填补缺失区域.从相关数据集中提取相似图像,并且转移到原始的空白部分,但是这些方法通常会修补出不一致的区域并且相似图像常常难以匹配.

最近,随着神经网络的发展,修复任务被建模为条件图像生成问题,其中高级特征和语义内容被公式化为编码器-解码器网络并且总是与生成性对抗网络(generative adversarial network, GAN)相结合以激励修复区域与原始像素之间的一致性<sup>[11-15]</sup>.上下文编码器<sup>[13]</sup>(context encoder, CE)是最早利用深度神经网络进行图像修复的方法之一,它对编码器-解码器生成网络进行训练,并与对抗性网络相结合,可以大大改善完成效果,但结果充满了低频内容以及包含明显的修复痕迹.此外,为了处理高分辨率的修复问题,Yang等人<sup>[16]</sup>旨在通过提取固定区域中的高级特征来优化修复后图像结果,然后从非空洞区域传播纹理信息以填充孔区域作为后处理,然而该方法经常导致与图像边界发散并且容易被人眼检测出修复痕迹.Iizuka等人<sup>[12]</sup>使用扩张卷积和2个尺度判别器改善结果,结果比以往的方法更好,但需要后处理<sup>[16]</sup>并且不能重建大孔洞.

对于人脸修复工作,图像中缺失的部分不能通过图像已知区域进行推断,从而不能通过先前的方法采用复制粘贴策略或者提取语义信息进行修复.因为许多缺失部分包含独特的属性,不能与人脸中其他属性(例如眼睛或嘴巴)进行关联.同时人脸修复往往需要网络针对单属性缺失(如眼睛或嘴巴)情况进行有效的修复,然而先前方法往往只能修复图

像固定区域,对于修复图像局部信息则泛化性能不够优秀.

文献[13]中 Li 等人提出的基于生成对抗网络的人脸修复(face generation, FG)方法对人脸修复问题做出了有效的修复效果,通过局部和总体的判别器的运用,同时将人脸语义分割的图像作为辅助信息提供五官的相对位置,从而对人脸的不同属性进行有效的修复,然而对于不对称的人脸以及图像像素的连续性修复并不优秀.这主要是因为其生成器没有很好地建立深层特征之间的关系,并且其生成器采用编码器与解码器结构在通过瓶颈层时会有大量的隐含特征信息损失.

本文提出了一种新颖并且有效的深度学习修复算法,该算法采用生成对抗训练方式,首先将整体属性缺失源图像放入架构为编码器和解码器(基于 U-net<sup>[17]</sup>网络)的生成网络中得到修复结果,使生成网络学习到人脸的整体属性语义信息,再通过我们提出的局部属性判别器加强生成网络对于局部属性的语义理解,从而使修复后的图像在整体和局部上达到语义以及内容上的一致性.

相比于方法 FG,本文的模型能够有效地修复不对称人脸(非正脸),获得图像局部特征以及其对应的特征相关性,同时能够生成新颖的图像,并且在色彩重构方面达到了更好的效果,这是因为我们的生成器基于 U-net 网络进行改进,而 U-net 网络拥有很好的色彩感知性.同时,我们提出的局部属性判别器能够使模型更好地学习到局部特征.在 CelebA<sup>[18]</sup>数据集上训练以及验证本文提出的模型,从修复的结果展示和利用标准的评测指标 2 方面验证了模型的有效性.

本文的主要贡献有 3 个方面:

- 1) 提出了一个新颖的深度学习框架,从生成器到对抗器都适用于人脸修复工作;
- 2) 对于人脸修复工作中的像素连接问题以及非对称图像修复问题作了有效的改善;
- 3) 相比较于以前的方法局限于修复中心缺失区域,本文的方法能够有效地学习到人脸局部信息,更加符合实际场景.

## 1 相关工作

近年来图像修复技术的发展迅速,从深度学习到非机器学习方法都对其有研究.我们接下来将介绍最具有代表性的工作.

图像修复首先在文献[1]中引入,其利用扩散函数将来自周围已知区域的低级特征传播到未知区域.与文献[1]类似的其他方法<sup>[19]</sup>使用一些距离场将外观信息从像素传播到缺失区域.然而,这些方法仅处理小孔洞以及颜色和纹理重建的低级细节.后来的工作使用基于补丁的方法<sup>[8,20]</sup>,以迭代方式匹配非孔洞区域中最相关的内容.但它总是需要大量的时间消耗,为了应对挑战,Patch-Match<sup>[4]</sup>提出了一个快速的最近邻匹配算法,并且达到了优秀的效果,然而基于神经网络的方法无法理解图像的深层语义信息并且不能够修复较大的缺失区域.

最近,基于深度学习的方法已经成为图像修复中一个很有研究价值的技术领域,尤其是基于生成对抗网络(GAN)<sup>[21]</sup>.然而,这些方法很难保证稳定的训练并获得高分辨率的结果.随着GAN的高速发展,许多方法如DC-GAN<sup>[22]</sup>,Wasserstein GAN<sup>[23]</sup>,WGAN-GP<sup>[24]</sup>,LS-GAN<sup>[25]</sup>,infoGAN<sup>[26]</sup>,Progressive GAN<sup>[27]</sup>都能够解决训练不稳定的问题,所以基于深度学习的图像修复方法能够持续发展.上下文编码器<sup>[13]</sup>在图像修复中利用GAN思想训练深度神经网络,编码大小为 $128\times128$ 不完整图像到低维,然后解码并生成缺失区域的内容,最终使用重建损失和对抗性损失作为约束训练网络参数.在文献[15,28]中,Yeh等人搜索受损图像最相关的特征空间作

为网络的辅助信息来修复缺失内容.Yang等人<sup>[16]</sup>将上下文编码器<sup>[13]</sup>的结果作为输入,逐渐增加纹理细节,以实现高分辨率的修复.最近,Iizuka等人<sup>[12]</sup>提出了全局和局部鉴别器并在编码器-解码器中使用空洞卷积提升修复效果,这2种技术都能够增加输出神经元的感知域从而更加有效地利用已知区域的语义信息.与文献[15]一样,Yu等人<sup>[29]</sup>用语义关注层取代了后处理.在文献[30]中,Ulyanov等人认为生成网络不需要额外的训练,可以依靠自己修补不完整的图像.在文献[31]中,Song等人使用网络迭代替换来自特征空间中最接近的完整图像的补丁.在文献[32]中,Liu等人提出一种新颖的部分卷积层,同时利用自动掩模以减少伪影.但是这些工作并不能为 лица图像修复工作提供高质量的关联属性以及局部特征,同时其只能针对中心区域或随机区域进行修复,无法拓展到具体的局部属性修复工作

2 本文算法

本文提出一种基于局部和整体属性一致性的人脸图像修复算法,将整体属性遮盖的图像作为输入( $I_{in}$ ),同时联合局部属性判别器促使生成器能够更好地学习到人脸的关键属性特征.本文所提出的模型架构如图1所示.其中 $I_{gt}$ 为原始图像, $I_{re}$ 为修复结果.

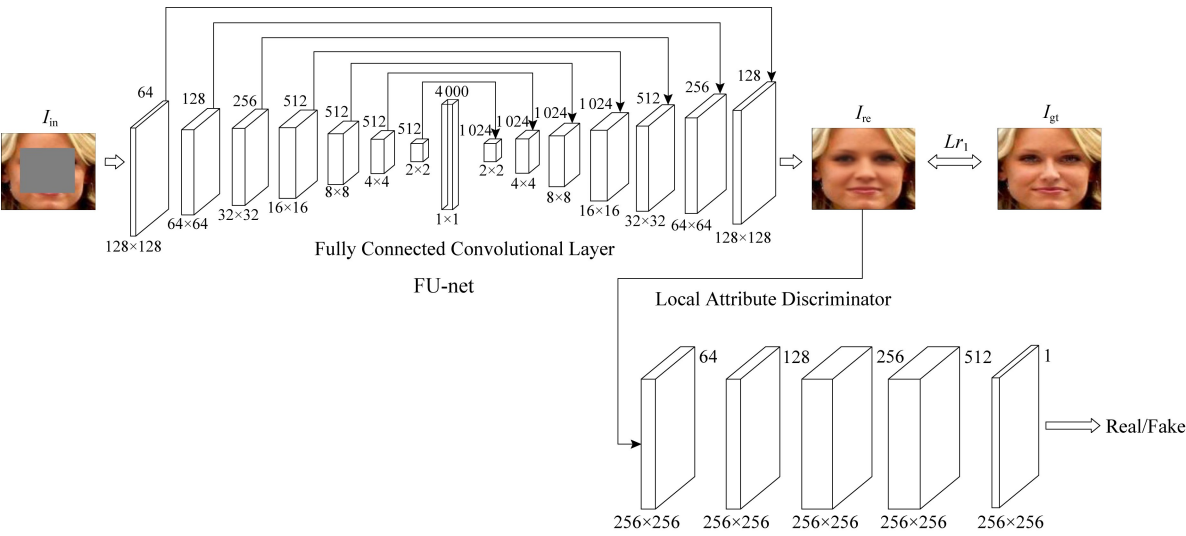


Fig. 1 Model architecture

图1 模型结构图

2.1 生成器

生成器以编码器和解码器的形式构建,编码器获得图像的未知区域和已知区域之间的变化和关系,接着将这些信息反馈到解码器中以生成图像.

在这样结构中,源图像通过逐渐下采样操作直到瓶颈层,此时该过程被反转到编码器.这种网络结构要求信息通过所有网络层,包括瓶颈层.对于图像转换问题,在输入和输出之间共享大量低级信息

能够有效地提高转换效果,并且希望直接在网络之间传送该信息而不需要多层的采样过程.图像修复问题其实也是一种图像转换问题,将有缺失的图像转换成完整的图像.所以本模型中的生成器采用了跳跃连接,编码器中卷积核大小为  $4 \times 4$ ,步长为 2,补零操作为 1,每经过一个卷积操作,图像缩小二分之一,卷积核个数翻倍,同时在经过激活函数之前将得到的特征映射图信息通过跳跃连接传递到解码器相对应的位置.输入图像的大小为  $256 \times 256 \times 3$ ,下采样到大小为  $2 \times 2 \times 512$  的特征图之后,再经过全链接卷积将特征信息传递到解码器.

2.2 全连接卷积

由于卷积只能将所有特征映射图连接在一起,不直接连接单个特征映射图内的特征值.以前提出的方法中常用全连接层进行连接,然而全连接层所用参数过多会加长训练时间以及占用更多 GPU 显存,所以我们的模型采用与上下文编码器<sup>[13]</sup>中一样的方法,使用全连接卷积.将图像下采样到  $512 \times 2 \times 2$  大小后,通过最后一层全连接卷积将图像特征压缩到  $4\,000 \times 1 \times 1$  大小模拟全连接层,使单个特征映射层达到全连接效果.如果采用全连接层进行连接,对于  $m$  个大小为  $n \times n$  的特征映射图,需要  $m^4 \times n \times n$  个参数,然而运用全连接卷积操作则只需要  $m^2 \times n \times n$  个参数.

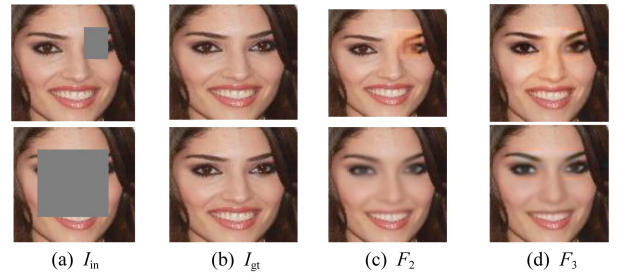
由于生成器采用全连接卷积层(fully connected convolutional layer)以及 U-net 模型架构作为基础,为我们的生成器模型称为 FU-net(fully connected convolutional U-net).

2.3 局部属性对抗网络

生成器对于修复人脸整体区域已经有着良好的效果,但是对于局部信息的学习却不够优秀.

如果缺失部分仅仅是中心的固定区域,如图 2(c) $F_2$  所示,我们的模型在图像内容以及语义上修复效果很优秀,如果缺失部分是人脸的一部分属性(眼睛或嘴巴),如图 2(c)所示,修复效果则会明显变差,这是因为生成网络 FU-net 在跳转连接各层信息的同时也会将孔洞信息传递过去,这些孔洞的像素值为 0,由于我们只用中心遮盖图像作为源图像,所以生成器只有当脸部整体被遮盖时才能将这些 0 像素值通过通过网络准换成合理的特征值.如果修复局部人脸特征,FU-net 中的编码器会将 0 像素值直接传递到解码器对应位置中,从而导致修复效果模糊并且色彩感知度低.所以,我们利用局部属性判别器对生成后的图像进行局部加强,局部属性

判别器基于补丁判别器<sup>[33]</sup>设计,将生成图像的中心区域通过多层卷积后转换成  $1 \times 256 \times 256$  大小的特征图,特征图上的值越接近 0 则与真实图像差距越小,而越接近 1 则差距越大.补丁判别器有一个良好属性,其能够适应任意大小的输入.但是原始补丁判别器并不能帮助局部修复效果更加优秀,因为其只能提高图像固定区域修复效果,所以我们提出局部属性判别器,判别器同样判别固定大小的中心区域是否为真实图像,但是生成器在与判别器对抗时并不是将修复后的中心区域与判别器对抗,我们利用补丁判别器对于输入大小没有强制性的特性,将生成的图像中的关键属性(例如眼睛或嘴巴)截取出来进行对抗训练,使生成器与判别器在人脸局部属性范围内相互对抗相互加强,这样能够保证网络不仅在整体修复上更加优秀,同时也会解决局部内容和语义修复问题.运用对抗网络之前和加入对抗网络之后的局部修复效果如图 2(d) $F_3$  所示.通过对抗网络我们明显改善了生成器的修复效果,提高了其在局部属性上的泛化性能,而不是局限于中心区域.



$F_2$  is the result using only the FU-net, and  $F_3$  is the repair after adding the local attribute discriminator. For the overall restoration of the image,  $F_3$  is slightly superior in detail. However, the difference between the local reduction effect (eyes or mouth) is obvious, which also reflects the role of the local attribute discriminator.

Fig. 2 Effect improvement display  
图 2 效果改善展示图

2.4 损失函数

本文通过将修复后的图像尽量回归到真实图像来训练我们的网络,通过联合多个损失函数来使最终的修复结果  $I_{re}$  与原始图像  $I_{gt}$  的距离减小.

2.4.1 重构损失

使用  $L_1$  损失作为图像重构损失函数  $Lr_1$ ,其定义为

$$Lr_1 = E \| I_{gt} - I_{re} \|_1, \tag{1}$$

$E$  代表求均值操作.重构损失  $Lr_1$  对于捕获修复区域的整体结构以及内容的一致性具有优越的效果,但倾向于将预测中的多种模式平均化.

### 2.4.2 对抗损失

对抗性损失试图使修复结果看起来真实,并具有从分布中选择特定模式的效果,从而弥补重构损失的不足.以往的局部判别器<sup>[12-13]</sup>虽然能够针对图像局部进行加强修复,但是对于原图像缺失部分大小和尺寸却需要固定,并不能完全适用于人脸中任意属性的修复.为了不用针对不同大小和尺寸的缺失部分去训练不同尺寸的判别器,我们基于补丁判别器设计了适用于任意输入尺寸的局部属性判别器.在训练判别器时只需判断图像中心区域是否为真实图像,当训练生成器去对抗判别器时,我们不用中心区域去对抗而是用局部区域(例如眼睛和嘴巴),所以损失函数为

$$\min_G \max_D L_{GAN}(G, D) = E[\ln(I_{gt})] + E_k[\ln(1 - D(I_{re}))], \quad (2)$$

$k$  代表不同局部属性的截取图像,  $D$  是局部属性判别器.综上所述,本文的总体函数结合对抗损失以及重构损失为

$$L_{all} = \lambda_1 \times L_{r1} + \lambda_2 \times L_{GAN}, \quad (3)$$

其中  $\lambda_1$  和  $\lambda_2$  分别为重构损失以及对抗损失的权重.

### 2.5 缺失区域

在本文的模型中,输入是 1 张人脸图像  $I_{in}$ , 灰色部分像素值为 0. 去除的区域必须对应人脸的整体大小情况以及人脸的各个属性区域分布情况.

经过多次覆盖实验发现,  $170 \times 170$  像素的补丁在 CelebA 数据集中能覆盖人脸的所有关键属性(如眼睛和嘴巴), 也就是人脸的中心区域, 如图 3 所示, 图 3(a) 是添加  $170 \times 170$  补丁的人脸  $I_{in}$ , 图 3(b) 是原始图像  $I_{gt}$ . 运用这种遮盖方式, 在进行对抗训练时, 能够截取到所有修复后的关键属性, 从而对这些区域进行细节加强.

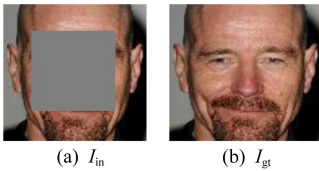


Fig. 3 Missing overall area  
图 3 整体区域缺失

## 3 实 验

### 3.1 数据集

我们的模型采用 CelebA 数据集训练以及验证

我们的模型.数据集由 202 599 幅人脸图像组成, 每张人脸图像都通过 OpenCV 自带人脸检测函数进行裁剪, 保证 2 眼的位置大致对齐, 并且被缩放成  $3 \times 256 \times 256$  的像素大小. 我们遵循该数据集标准分割, 其中 162 770 张图片作为训练集, 19 867 张图片作为验证集, 19 962 张作为测试集. 为了防止模型过拟合, 实施了随机反转、打乱数据等操作.

### 3.2 训练方法

我们通过预训练使生成器学习脸部的整体属性, 生成器中编码器激活函数采用 LeakyReLU, 解码器中的激活函数采用 ReLU 函数. 编码器与解码器中所有正则操作都采用实例正则化(instance normalization), 因为本文的模型批量大小(batch size)设置为 1, 而批量正则化(batch normalization)适合于模型批量大小(batch size)的值大于 1 的情况, 使用实例正则化更适合训练要求, 局部属性判别器同样采用实例正则化(instance normalization), 激活函数采用 ReLU.

本文采用分步训练的方法以稳定判别器的效果: 1) 利用中心缺失图像训练生成器, 使产生的图像尽量回归到真实图像; 2) 将生成器参数固定, 只训练局部属性判别器; 3) 将判别器以及生成器联合训练并利用总体函数对训练进行约束从而达到最终的效果.

本文实验使用深度学习框架 Pytorch3.0. 编程语言为 python3.5, 操作系统为 Windows10. 实验的硬件环境为 Intel® Core™ i5-8700k 处理器, 内存为 16 GB, NVIDIA 1080TI 显卡, 该配置是目前深度学习计算的主流配置.

### 3.3 人脸修复

#### 3.3.1 修复效果

如图 4 展示了在 CelebA 测试集上的修复效果, 图 4 分 3 个实验, 每个实验分 3 组, 每一组中第 1 列是真实图像, 第 2 列是需要修复的图像, 第 3 列是修复结果. 每一组需修复的图像中至少遮盖住人脸的一个重要属性(眼睛、嘴巴)甚至多个属性, 遮盖区域不能够超过  $170 \times 170$  的人脸中心区域, 需要在这个区域内修复才能得到最佳效果. 图 4 展现出本文模型生成内容的真实性、创新性以及美观性, 无论是针对人脸的整体或者局部都有良好的修复效果.

如图 4(a)(b) 的第 3 行实验结果所示, 部分原图像本身就有对人脸的重要属性进行遮盖(比如眼镜遮盖眼睛), 是否运用这种遮盖方式可以通过我们的方法控制, 如果想去除眼镜对眼睛的遮盖, 通过将

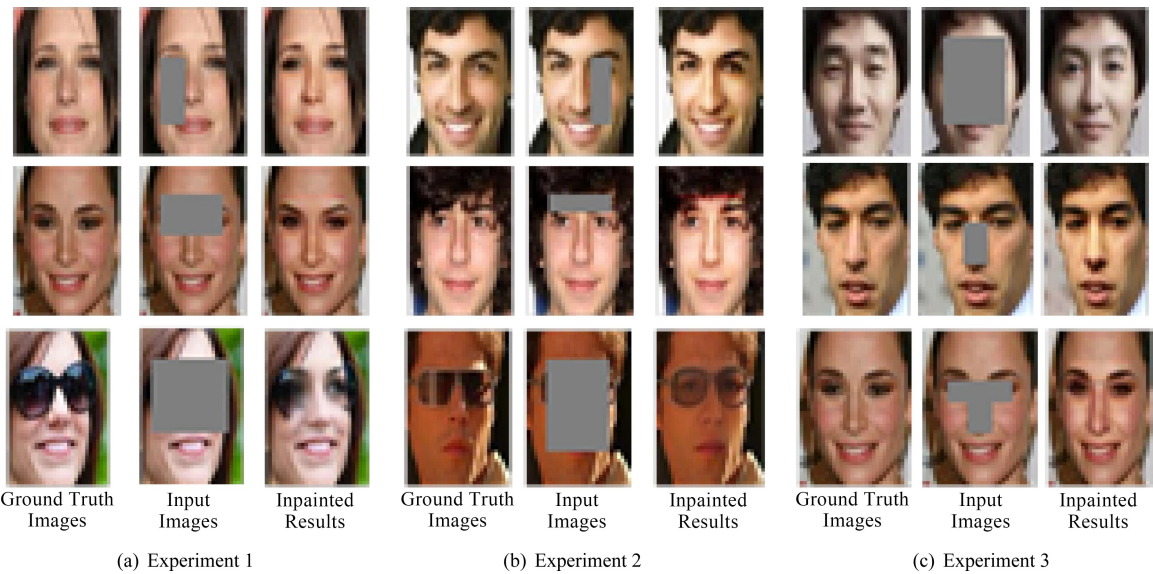


Fig. 4 Inpainted result  
图4 修复结果展示图

眼镜完全覆盖就能够还原出眼睛.因为数据集中大部分数据都是没有使用眼镜进行遮挡,我们的方法能够有效地学习到人脸的眼睛属性.

3.3.2 算法对比

我们将CE和FG作为基准方法用于验证模型有效性.在测试集上运用2个评价指标对本文的方法进行评价,这2个指标也是图像修复中的通用指标,在FG与CE中都采用了这2个标准.第1个指标是峰值信噪比(PSNR),直接比较图之间像素值的差距.第2个是结构相似性指数(SSIM)<sup>[34]</sup>,这是一种全参考的图像质量评价指标,它分别从亮度、对比度、结构3方面度量图像的相似程度.对于测试用例,本文采用与FG中同一张图片并且采用同样的6种遮盖方式,如图5所示.FG中采用的是128×128像素图片,而本文的模型采用的图像大小是256×256,所以将修复后的图像压缩成128×128进行对比.

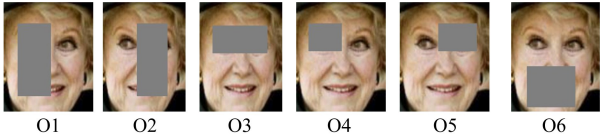


Fig. 5 Test case  
图5 测试用例

表1和表2分别展示了本文模型与基准模型在PSNR以及SSIM指标上的实验比较结果.其中FU表示只有生成器进行图像修复没有局部属性判别器

对其局部属性修复效果进行加强;而Ours表示本文的完整方法;列4和列5分别表示上下文编码器CE和人脸生成方法FG.

Table 1 Comparison Results Based on PSNR  
表1 基于PSNR评价指标的对比结果

Input	FU	Ours	CE	FG
O1	19.90	<b>20.04</b>	18.6	20.0
O2	21.39	<b>21.41</b>	18.4	19.8
O3	19.87	<b>20.35</b>	17.9	18.8
O4	20.92	<b>20.94</b>	19.0	19.7
O5	20.57	<b>20.92</b>	19.1	19.5
O6	18.70	<b>20.02</b>	19.3	20.2

Note: The best results are bold.

Table 2 Comparison Results Based on SSIM  
表2 基于SSIM评价指标的对比结果

Input	FU	Ours	CE	FG
O1	0.730	0.740	0.772	<b>0.824</b>
O2	0.730	0.797	0.774	<b>0.826</b>
O3	0.777	<b>0.789</b>	0.719	0.759
O4	0.817	<b>0.823</b>	0.754	0.789
O5	0.813	<b>0.820</b>	0.757	0.784
O6	0.707	0.717	0.818	<b>0.841</b>

Note: The best results are bold.

如表1以及表2所示,PSNR以及SSIM评测指标都测定出本文的完整方法效果更优秀,这证明

本文的模型在经过局部属性判别器进行对抗训练后,模型整体对于图像深层语义的理解能力有了显著的提高.同时在 PSNR 数值上本文模型针对 CE 和 FG 都有明显的提高,这也证明了本模型的优越效果.

本文的模型虽然在 PSNR 上对比结果比较优秀,但与 FG 在 SSIM 数值方面进行比较发现:仅仅在 O3,O4,O5 上比 FG 高,这是因为除了 O3,O4,O5 以外的遮盖区域都有大部分超过了图像中心遮盖(170×170)范围,这些区域会模糊并且变暗,存在色彩差距,图像整体性被破坏.但是另一小部分在中心区域内,这一部分的修复效果会比较优秀,像素值得到了还原.PSNR 检测的是图像像素值的差距,而 SSIM 从亮度、对比度、结构 3 方面衡量,在 O1,O2,O4,O5,O6 区域还原像素值差距并不大,而亮度和对比度差距相对较大,所以会出现这种 PSNR 值较高而 SSIM 值较小的反差结果.

本文的方法不仅仅在评测指标上体现出优秀的结果,同时也解决了 FG 中 2 个局限,第 1 个局限是不对称图像修复效果不够优秀,这是因为其生成网络所采用的编码器-解码器结构虽然能够有效地还原图像,但是由于编码器与解码器之间的特征没有直接的关联,在特征信息传递通过瓶颈层时会出现不可避免的信息损失,从而 FG 中的生成器并不能充分理解语义.第 2 个局限是该模型没有充分理解相邻像素之间的空间相关性,无法通过已知的上下文信息推断出要生成的信息,特别是针对颜色的连续性修复.对于以上 2 个问题,本文都进行了有效的解决.由于本文的生成器 FU-net 是基于 U-net 网络

进行改进,利用特征图跳跃传播的特性有效地将相邻像素连接起来,学习到像素点的空间相关性.同时由于采用局部属性判别器进行对抗训练,在学习到人脸的整体属性后再去学习关键属性,在非对称图中能够有效地进行局部修复.如图 6 所示.图 6 包含 3 组实验结果,图 6(a)是真实图像,图 6(b)是需修复的图像,图 6(c)是修复结果.如行 1、行 2 所示,我们的模型能够很好地修复不对称人脸中的局部属性;如行 3、行 4 所示,本文方法通过下嘴唇的色彩特征完好地修复了上嘴唇色,解决了相邻像素的空间不相关问题.

3.4 局限性

本文的方法仍然有着局限性,在训练过程中我们采用局部属性判别器进行加强,然而局部属性判别器只能针对截取出来的人脸的重要属性特征(眼睛、鼻子)加强生成器修复效果,当我们的遮盖范围不是 170×170 的中心区域或者不是局部属性区域时修复效果就会不够真实,如图 7 所示,图 7 有 (a) (b) 两组实验结果,每一组实验结果包括最左侧的真实图像、中间的输入图像和最右侧的修复后的图像.这 2 组实验结果中的遮盖区域都超过了 170×170 的中心区域,其中第 1 组实验遮盖区域与表 1 和表 2 所示的 O6 区域相近,这也是本文方法在 SSIM 数值上相对于其他方法较低的原因.

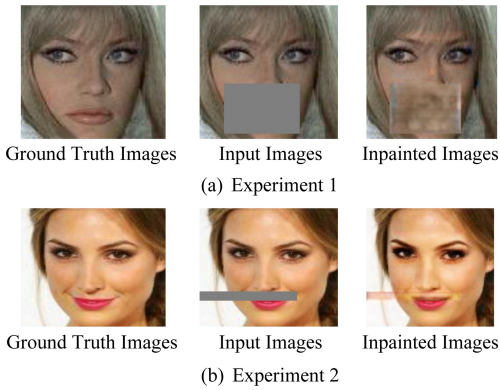


Fig. 7 Model limitations

图 7 模型局限性

5 总 结

本文提出了一种新颖的模型用于解决人脸图像修复问题,该模型从人脸总体以及局部属性出发,基于局部属性对抗网络和 FU-net 生成器,从图像语义到图像视觉都达到优秀的效果,并且能够有效地修复人脸的重要属性.未来我们会改进我们的方法,

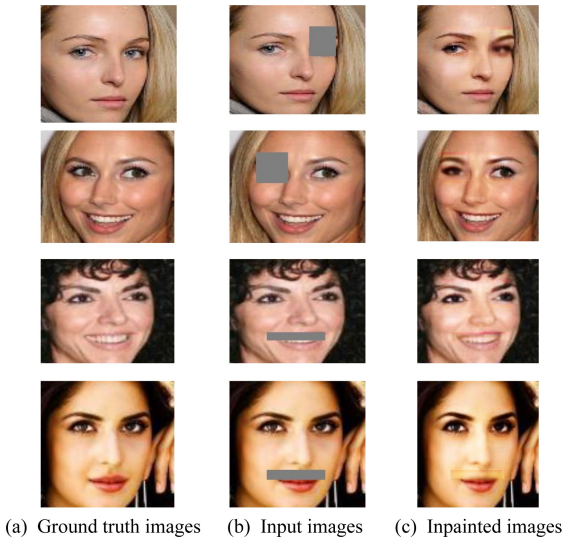


Fig. 6 The display of defect improvement

图 6 缺陷的改善展示

在输入时覆盖任意随机区域都能够达到优秀的修复效果,我们会构建图像模具,基于模具进行精确修复,并且改善 FU-net 网络,使我们的修复效果进一步精进.

## 参 考 文 献

- [1] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting [C] //Proc of the 27th Annual Conf on Computer Graphics and Interactive Techniques. New York: ACM, 2000: 417-424
- [2] Komodakis N. Image completion using global optimization [C] //Proc of 2006 IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2006: 442-452
- [3] Wexler Y, Shechtman E, Irani M. Spacetime video completion [C] //Proc of the 2004 IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2004: 100-111
- [4] Barnes C, Shechtman E, Finkelstein A, et al. Patch match: A randomized correspondence algorithm for structural image editing [J]. ACM Transactions on Graphics, 2009, 28(3): 21-28
- [5] Bertalmio M, Vese L, Sapiro G, et al. Simultaneous structure and texture image inpainting [J]. IEEE Transactions on Image Processing, 2003, 12(8): 882-889
- [6] Wilczkowiak M, Brostow G J, Tordoff B, et al. Hole filling through photomontage [C] //Proc of the 2005 Conf of the British Machine Vision. Piscataway, NJ: IEEE, 2015: 30-41
- [7] Efros A A, Leung T K. Texture synthesis by nonparametric sampling [C] //Proc of the 7th IEEE Int Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 1999: 1033-1038
- [8] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer [C] //Proc of the 28th Conf on Computer Graphics and Interactive Techniques. New York: ACM, 2001: 341-346
- [9] Hays J, Efros A A. Scene completion using millions of photographs [J]. ACM Transactions on Graphics, 2007, 26(4): 36-40
- [10] Whyte O, Sivic J, Zisserman A. Get out of my picture! [C] //Proc of 2009 British Machine Vision. Piscataway, NJ: IEEE, 2009: 126-138
- [11] Kwatra V, Essa I, Bobick A, et al. Texture optimization for example based synthesis [J]. ACM Transactions on Graphics, 2005, 24: 795-802
- [12] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion [J]. ACM Transactions on Graphics, 2017, 36(4): 107-114
- [13] Li Yijun, Liu Sifei, Yang Jimei, et al. Generative face completion [C] //Proc of the 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR 2017). Piscataway, NJ: IEEE, 2017: 5892-5900
- [14] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition (CVPR 2016). Piscataway, NJ: IEEE, 2016: 2536-2544
- [15] Yeh R, Chen Chen, Lim T Y, et al. Semantic image inpainting with perceptual and contextual losses [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition (CVPR 2017). Piscataway, NJ: IEEE, 2017: 5009-5018
- [16] Yang Chao, Lu Xin, Lin Zhe, et al. High-resolution image inpainting using multi-scale neural patch synthesis [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition (CVPR 2017). Piscataway, NJ: IEEE, 2017: 4076-4084
- [17] Ronneberger O, Fischer P, Brox P. U-net: Convolutional networks for biomedical image segmentation [C] //Proc of the 18th Int Conf on Medical Image Computing and Computer Assisted Intervention (MICCAI 2015). Piscataway, NJ: IEEE, 2015: 3157-3165
- [18] Liu Ziwei, Luo Ping, Wang Xiaogang, et al. Deep learning face attributes in the wild [C] //Proc of 2015 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3730-3738
- [19] Coloma B, Marcelo B, Vicent C et al. Filling-in by joint interpolation of vector fields and black levels [J]. IEEE Transactions on Image Processing, 2001, 10(8): 1200-1211
- [20] Telea A. An image inpainting technique based on the fast marching method [J]. Journal of Graphics Tools, 2004, 9(1): 23-34
- [21] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [G] //Advances in Neural Information Processing Systems. New York: ACM, 2014: 2672-2680
- [22] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. arXiv preprint arXiv: 1511.06434, 2015
- [23] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN [J]. arXiv preprint arXiv: 1701.07875, 2017
- [24] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs [J]. arXiv preprint arXiv: 1704.00028, 2017
- [25] Mao Xudong, Li Qing, Xie Haoran, et al. Least squares generative adversarial networks [J]. arXiv preprint ArXiv: 1611.04076, 2016
- [26] Chen Xi, Duan Yan, Houthoofd R, et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets [C] //Proc of the 30th Conf on Neural Information Processing Systems (NIPS 2016). New York: ACM, 2016: 5600-5611
- [27] Tero K, Timo A, Samuli L, et al. Progressive growing of GANs for improved quality, stability, and variation [J]. arXiv preprint arXiv: 1710.10196, 2017
- [28] Raymond A Y, Chen Chen, Teck Y L, et al. Semantic image inpainting with deep generative models [C] //Proc of the 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR2017). Piscataway, NJ: IEEE, 2017: 5485-5493

[29] Yu Jiahui, Lin Zhe, Yang Jimei, et al. Generative image inpainting with contextual attention [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition (CVPR 2018). Piscataway, NJ: IEEE, 2018: 4076-4084

[30] Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior [J]. arXiv preprint arXiv: 1711.10925, 2017

[31] Song Yuhang, Yang Chao, Lin Zhe, et al. Image inpainting using multi-scale feature image translation [J]. arXiv preprint arXiv: 1711.08590, 2017

[32] Liu Guilin, Fitsum A, Kevin J, et al. Image inpainting for irregular holes using partial convolutions [J]. arXiv preprint arXiv: 1804.07723, 2018

[33] Isola P, Zhu Junyan, Zhou Tinghui, et al. Image-to-image translation with conditional adversarial networks [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition (CVPR 2017). Piscataway, NJ: IEEE, 2017: 6822-6834

[34] Wang Zhou, Bovik A C, Sheikh H R, et al. Image quality assessment: From error vsibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13 (4): 32-40



**Jiang Bin**, born in 1971. PhD. Associate professor. His main research interests include big data technology, computer vision and machine learning.



**Liu Hongyu**, born in 1994. Master candidate. His main research interests include computer vision and machine learning.



**Yang Chao**, born in 1978. PhD. Associate professor. Her main research interests include big data technology, social network computing and intelligent information processing.



**Tu Wenxuan**, born in 1995. Master candidate. His main research interests include computer vision and machine learning.



**Zhao Zilong**, born in 1995. Master candidate. His main research interests include computer vision, and machine learning.