计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190400118

基于残差生成对抗网络的人脸图像复原

李泽文　李子铭　费天禄　王瑞琳　谢在鹏
河海大学计算机与信息学院　南京 211100
(servon@hhu.edu.cn)

摘　要　得益于计算机视觉的快速发展,人脸图像复原技术可以仅利用人脸的轮廓来生成完整的人脸图像。目前已有许多基于卷积神经网络和生成对抗网络等方法的人脸复原技术被提出,它们可以利用部分破损的人脸图像进行复原或者使用人脸轮廓直接生成人脸图像。然而,使用这些技术复原后的人脸图像在定性和定量分析时效果不够理想,并且复原时存在诸多的条件限制。因此,文中提出了一种基于残差生成对抗网络的人脸图像复原(FR-RGAN)方法,该方法借助深度卷积、残差网络和更小的卷积核,提升了模型性能,利用人脸的轮廓复原面部局部细节,使其更加生动地呈现出来。实验结果表明,FR-RGAN在均方误差、峰值信噪比和结构相似度指标上比pix2pix分别提高了8.7%,2.1%和9.6%,比无残差方法分别提高了53.4%,12.6%和30.1%。

关键词:计算机视觉;人脸图像复原;生成对抗网络;残差网络
中图法分类号　TP183

Face Image Restoration Based on Residual Generative Adversarial Network

LI Ze-wen,LI Zi-ming,FEI Tian-lu,WANG Rui-lin and XIE Zai-peng
College of Computer and Information,Hohai University,Nanjing 211100,China

Abstract　Benefiting from the rapid development of computer vision,face image restoration technology can only use the contour of the face to generate a complete face image. At present,many face restoration techniques based on convolutional neural networks and generative adversarial networks have been proposed. They can restore partial damaged face images or even directly generate face images using face contours. However,the results of qualitative and quantitative analysis of face images restored by these techniques are not ideal,and there are many limitations in the restoration process. Therefore,this paper proposes a face image restoration method based on residual generative adversarial network (FR-RGAN),which improves the performance of the model by means of deep convolution,residual network and smaller convolution kernels,and restores the local details of the face by using the contour of the face,making it more vivid. Experimental results show that,compared with pix2pix,FR-RGAN has an improvement of 8.7%,2.1% and 9.6% respectively in mean square error,peak signal to noise ratio and structural similarity index,and 53.4%,12.6% and 30.1% better than non-residual method.

Keywords　Computer vision,Face image restoration,Generative adversarial networks,Residual neural networks

1　引言

近年来,随着计算机视觉的快速发展,人脸图像复原技术的应用成为具有挑战性的任务。使用人脸图像复原技术,人们可以仅用人脸的轮廓来生成完整的面部,从而较为逼真地还原人脸图像。目前已有许多基于卷积神经网络和生成对抗网络的人脸复原技术[1-6]被提出,在一定程度上其可被应用在人脸图像修复等领域。

Shen等[1]提出了一种利用全变分模型来填充图像损坏区域以复原人脸图像的方法。该方法通过从同一图像中找到相似的纹理来进行填充,当损坏区域变大,该图像中难以找到相似纹理时该方法失效。Barnes等[2]利用PatchMatch算法在图像的可用部分搜索类似的图像块,并证明了该方法的质量和效率优于全变分模型填充图像。在人脸图像损坏区域较小时,使用该方法的修复效果较好,但是当损坏区域增大、损坏区域形状不固定时,无法搜索到匹配的图像块,修复效果变差。Li等[3]提出了一种基于深度生成模型的人脸补全算法,输入图像首先被随机选择的正方形区域上的噪声像素屏蔽,然后送入自动编码器训练,使用该方法补全的人脸图像较为真实,但是在图像破损区域边缘较为刻板。Yeh等[4]使用原有图像语义信息对图像进行补全,使得生成的图像在破损区域与完好区域的过渡更加平滑,但是其局限于部分人脸图像修复,要求图像破损区域较小,在背景较为复杂或人脸在图像中的占比较小的图像中无法修复人脸。Gauthier等[5]采用将人脸条件特征和随机噪声相结合的方式来生成人脸图像,但由于随机噪声的引入,使得该人脸图像可以是真实世界中的任意人脸,并不能精确地生成唯一个体的人脸图像,且每次生成结果都会变化。Isola等[6]在其基础上优化,提出了pix2pix

基金项目:国家级大学生创新创业训练计划项目(201810294106)
This work was supported by the National College Students' Innovation and Entrepreneurship Training Program (201810294106).
通信作者:谢在鹏(zaipengxie@hhu.edu.cn)

模型,实现了仅通过将物体轮廓作为条件信息来还原物体图像,改善了生成脸部图像随机性大的问题,较为精确地还原了人脸图像,但该方法只针对物体图像的复原,并未专门对人脸图像的复原做过优化。

为解决这些问题,我们提出了基于残差生成对抗网络的人脸图像复原算法(FR-RGAN)。该方法与常见的人脸图像修复技术不同,不再需要利用原图片的完好区域来修复破损区域,而是仅利用人脸的特征轮廓对面部进行复原,因此人脸在图像中的占比不影响最终复原结果。利用卷积神经网络提取图片特征之后,再利用全连接层对所提取的特征高度提纯,使得人脸图像复原效果在 Isola 等的方法上得到进一步提升。此外,残差网络的加入解决了网络层数过深导致难以收敛的问题,并且使得模型在训练时收敛速度更快。

为了验证本文方法,我们在实验中用 3 个数据集进行了对比实验。实验结果表明,本文方法与其他方法相比具有更好的性能,除了在视觉上有所提升之外,在 MSE,PSNR,SSIM 3 个指标上都表现最优。

2 相关工作

目前已有许多学者在研究人脸图像复原技术,部分学者专注于对部分破损的人脸图像进行修复,从而达到图像复原的效果,还有部分学者利用人脸的表征信息,不对原图像进行复原,而是直接重新生成人脸图像。两种技术各有优劣,前者能够通过大量样本学习到人脸的泛化特征,例如正常人脸均是由两个眼睛、一个鼻子等五官组成,因此只需要训练得到一个模型即可对所有人脸图像进行修复。但是,正因为学习泛化特征的特性使其无法精确地还原具体某个人的面部特征,如面部皱纹的多少、眼睛的大小等。后者能够充分学习具体某个人的面部特征,因此在还原时只需一个面部轮廓即可较为精确地还原整个面部;但是该技术在使用时需要对每个待还原面部图像的人都进行训练。因此,在具体应用时需要根据实际情况进行选择。我们的工作主要是基于后者,即利用人脸的轮廓信息重新生成人脸图像。由于篇幅限制,我们不能将所有的人脸图像复原方法在此一一列举,下面仅列出部分方法。

(1)人脸图像修复。Hays 等[7]提出从一个巨大的数据库中剪切粘贴一个语义相似的补丁,基于互联网的检索可以用来替换人脸图像的缺失区域,该方法需要缺失图像与数据库或互联网内容进行精确匹配,由于人脸的多样性,该方法很容易失败。Pathak 等[8]提出的上下文编码器(CE)是语义输入的一种开创性方法,在人脸图像上给定一个表示缺失区域的掩膜,训练神经网络编码上下文信息并预测缺失内容。Yeh 等[4]将语义修复视为约束图像生成问题,训练一个生成模型并搜索与潜在空间中“最接近”的损坏图像编码,训练时无需生成缺失区域的掩膜即可应用于推理过程中任意结构的缺失区域。

(2)人脸图像生成。Kingma 等[9]提出变分自动编码器(VAEs),通过在潜在单元上施加先验来调整编码器,以便通过潜在单元来采样或插入潜在单元补全人脸图像。然而,因为 VAEs 的训练目标是基于像素级高斯似然的,所以使用该方法生成的图像往往过于平滑甚至是模糊的,这在修复蓝天、白云等较为平滑的物体时具有一定优势,但是在修复人脸图像时效果不够理想。Larsen 等[10]通过添加源自生成性对抗网络的判别器[11]来改进 VAEs,并证明可以生成更真实的人脸图像,将变分自动编码器与对抗损失函数一起训练,可以防止过度平滑,但可能导致失真。Isola 等利用生成对抗网络来复原图像,生成对抗网络已经被许多学者证明可以生成高质量的图像[11-16]。条件生成对抗网络在文本生成等其他领域[17-18]已经取得了重大进展。在人脸图像复原中,其可以将人脸的特征信息作为模型输入,从而增加了额外的控制信息,使得原本难以控制的生成对抗网络容易控制,从而能够较为精确地生成确定个体的人脸图像。Zhu 等[19]提出的循环一致性生成对抗网络(CycleGAN)使用环形的网络结构,包括两个生成器与两个判别器,通过对称式的生成器环状结构来实现两个域的图像风格的转化,主要应用于图像的风格迁移领域。Berthelot 等[20]提出边界平衡生成网络(BEGAN),用一种均衡的概念,将自编码器作为分类器,通过基于 Wasserstein 距离的损失来匹配自编码器的损失分布,通过引入一个超参数在图像多样性与质量之间进行平衡,认为如果分布的误差较接近则分布也较接近。

本文方法与 Isola 等的方法最近接,除了引入条件生成对抗网络之外,还引入了带残差的深度卷积生成对抗网络,并且我们对其网络结构进行了优化,改变了卷积核的大小,增加了非线性变换的操作,使得模型的泛化能力得到进一步提高。同时对人脸特征高度提纯,使得在训练人脸图像复原模型时的收敛速度更快。在后续的实验中,我们证明了在人脸图像复原任务中 FR-RGAN 的结果优于 Isola 等提出的方法。

3 人脸图像复原算法

本节将详细地介绍本文提出的一种基于残差生成对抗式网络的人脸图像复原方法。该方法可以对人脸图像高效地进行编码和解码,从而实现人脸图像的复原。在我们的系统中需要预先训练一个可以利用人脸轮廓进行图像复原的模型,再利用该模型进行推理来复原人脸图像。

(1)模型训练。在训练模型时主要有 4 个部分:1)预处理部分;2)图像编码层;3)图像解码层;4)图像判别器。输入为训练集中的清晰人脸图像,输出为可用于生成人脸图像的模型文件。首先,将收集到的数据集通过预处理进行归一化,使图像大小适应我们的系统;其次,将处理后的图像输入至编码层进行编码得到信息矩阵;然后,利用解码层对信息矩阵进行解码得到人脸图像;最后,判别器判断该图像是否服从训练集图像的分布,从而进行优化。

(2)模型推理。在利用模型对人脸图像进行复原时主要有 3 个部分:1)预处理部分;2)图像编码层;3)图像解码层。输入需要复原的人脸轮廓,利用训练得到的模型进行推理,输出为还原后的人脸图像。首先,将需要还原的人脸图像进行预处理,使图像大小适应我们的系统;其次,将其输入至图像编码层得到信息矩阵;最后,进行解码,得到复原后的人脸图像。

3.1 模型训练

模型训练的伪代码描述如下:

利用伪代码描述本网络的训练算法。其中,k 是一个超参数,表示更新 k 次判别模型后再更新一次生成模型;$D(x)$ 表示判别模型对图像 x 判断后的输出,范围为(0,1),越接近

1 表示越接近真实图像；$G(s^{(i)})$表示生成模型利用人脸特征点生成的图像。

for 迭代次数 do：

for k do：

以 m 为小批量，从源图像分布 $p_{\text{data}}(x)$中选取 m 张需要复原人脸的清晰视频数据集$\{x^{(1)},x^{(2)},\cdots,x^{(m)}\}$。

以 m 为小批量，利用 dlib 库从 $p_{\text{data_feature}}(x)$中选取上述 m 张图像的各 68 个人脸特征点集合$\{s^{(1)},s^{(2)},\cdots,s^{(m)}\}$。

通过下列公式增加判别模型的随机梯度：

$$\nabla_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}[\log D(x^{(i)})+\log(1-D(G(s^{(i)})))]$$

end for

以 m 为小批量，利用 dlib 库从 $p_{data_feature}(x)$中选取 m 张图像的各 68 个人脸特征点集合$\{s^{(1)},s^{(2)},\cdots,s^{(m)}\}$。

通过以下公式减小生成模型的随机梯度：

$$\nabla_{\theta_g}\frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(s^{(i)})))$$

end for

当达到最优时，$G(\mathrm{p}_{\mathrm{data}_{\mathrm{feature}}})(x))=p_{\text{data}}(x)$，此时生成器利用人脸特征点生成的图像符合源图像分布。

目前，随着人工神经网络的发展，基于神经网络的算法已被广泛应用于各行各业。在应用的过程中，如何构建更加高效的模型，如何通过恰当的训练来得到优良的参数是能否具有实用性的关键所在。如图 1 所示，本文算法分为以下几个步骤。

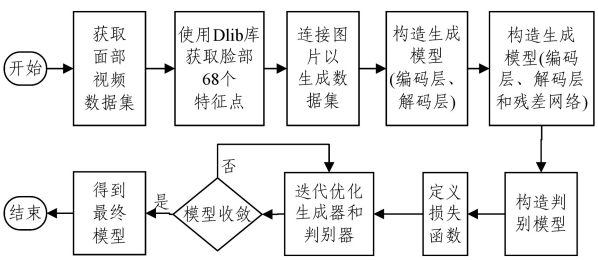

图 1 FR-RGAN 算法的模型训练流程图

Fig. 1 Flow chart of model training for FR-RGAN algorithm

步骤 1 获取需要复原人脸的清晰视频数据集。捕获数据集中的每一帧，并将它们转换为 256×256×3 标准大小的三通道 RGB 图像用作输入。

步骤 2 利用 Dlib 库[21]来获取图像中人脸的 68 个特征点。Dlib 库是一个包含了广泛的机器学习算法的库，其中包括人脸识别算法。这 68 个特征点表示为特征集合 s：

$$s=\{P_i\mid P_i=(x,y),x\in[0,H-1],y\in[0,W-1],i\in[0,67]\} \tag{1}$$

基于 68 个特征点生成人脸轮廓图。人脸轮廓图中，我们使用白色像素值来表示面部表情线，其余像素值均为黑色。具体到本文的系统中，像素矩阵将表示为：

$$I_{(i,j)}=\begin{cases}(255,255,255), & \text{if } (i,j)\in s\\ (0,0,0), & \text{if (i,j)}\notin \text{s}\end{cases} \tag{2}$$

其中，I 和 s 分别表示为像素值矩阵和面部特征点的集合。

在此之后，为了改变原始图像的比例，我们需要对图像进行裁剪。为了获得输入图像的标准大小，我们根据以下公式裁剪出轮廓图并调整轮廓图的大小来适应我们的系统。

$$\begin{cases}S=\min(H,W)\\ \hat{H}=\dfrac{(H-S)}{2}\\ \hat{W}=\dfrac{(W-S)}{2}\end{cases} \tag{3}$$

$$\hat{M}=F(G(M))=F(\sum_{i=\hat{H}}^{H}\sum_{j=\hat{W}}^{W}M[i-\hat{H},j-\hat{W}]) \tag{4}$$

其中，S 表示原图像高度 H 和原图像宽度 W 的最小值，M 表示原始图像，F 表示调整图像大小的操作，G 表示对图像的裁剪操作，将图像现有大小调整为高度为 $\hat{H}$、宽度为 $\hat{W}$ 的标准大小图像，其结果为一个新矩阵，表示为 $\hat{\boldsymbol{M}}$。

步骤 3 将步骤 1 数据集中的面部图像和步骤 2 得到的标准化特征图像相连接，得到大小为 256×512×3 的图像构成训练集。

步骤 4 构造编码层。编码层由 8 个卷积层和 1 个全连接层组成。为方便起见，我们使用符号 $\varepsilon(n)$来表示编码层的第 n 个编码器。编码网络的结构如图 2 所示。8 个编码器的卷积核心大小均为 3×3，使用 SAME 方式[22]填充，滑动步长为 2。$\varepsilon(1)$的卷积核心数为 64，$\varepsilon(2)$的为 128，$\varepsilon(3)$的为 256，$\varepsilon(4)$和 $\varepsilon(5)$均为 512，$\varepsilon(6)$、$\varepsilon(7)$和 $\varepsilon(8)$均为 1024。最后一层是全连接层，包含 100 个神经元，可以将输出转换为 100 维的信息矩阵。因为该 100 维信息矩阵在解码器中被用作恢复人脸图像，所以在训练时步骤 7 中的判别模型会让该 100 维矩阵尽可能多地表示人脸特征，因为人脸特征越多，生成的人脸图像质量越好，损失函数就越小，即该 100 维矩阵表示的人脸特征的纯度随着训练过程逐渐增加。

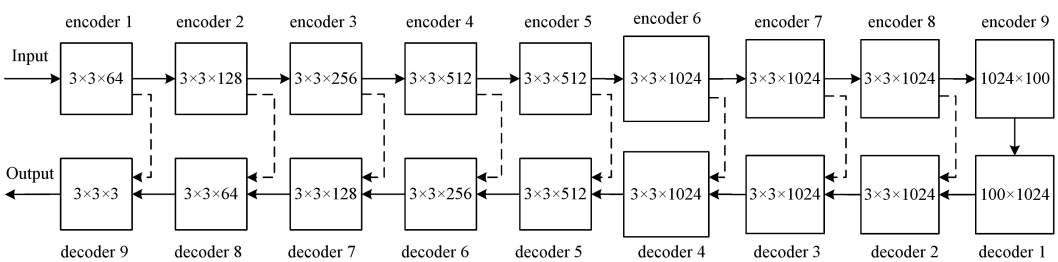

图 2 生成模型结构

Fig. 2 Generation model structure

步骤 5 构造解码层。解码层由 1 个全连接层和 8 个转置卷积层组成，它们与编码层的 8 个卷积层对称。为方便起见，我们使用符号来表示解码层的第 n 个解码器。解码网络的结构如图 2 所示。第一层是全连接层，包含 1024 个神经元，其输入是编码层最后的输出，即 100 维信息矩阵。8 个后续解码器的转置卷积核心大小均为 3×3，使用 SAME 方式填充，滑动步长为 2。需要指出的是，每层解码器的配置与编码器是对称的，换言之即 $D(n)$的输出等同于 $D(10-n)$的输入。

经过上述处理，面部图像得以还原。

步骤 6　利用残差网络将编码层与解码层相连。实验结果表明，直接编码和解码面部图像获得的结果差强人意，因此我们借鉴了残差网络[23]的思想。如图 2 所示，虚线表示残差连接。我们将 $D(n)$ 的输入记为 $I(D(n))$，其由两部分组成：$D(n-1)$ 的输出，记为 $O(D(n-1))$，以及 $E(10-n)$ 的输出，记为 $O(E(10-n))$，表示为：

$$I(D(n))=Q(O(D(n-1)),O(E(10-n))) \tag{5}$$

其中，Q 表示连接操作，$O(D(n-1))\in\mathbb{R}^{H\times W\times C}$，$O(E(10-n))\in\mathbb{R}^{H\times W\times C}$，$I(D(n))\in\mathbb{R}^{H\times W\times(2\times C)}$。

步骤 7　构建判别模型。为方便起见，我们使用符号 $L(n)$ 来表示判别模型的第 n 层。如图 3 所示，判别模型共有 5 层。前四个卷积核心的大小为 3×3，使用 VALID 方式[22]填充，批量归一化[24]，使用泄漏整流线性单元(Leaky ReLU)[25]激活函数，滑动步长为 2。它们的卷积核的数量分别为 64，128，256 和 512。第五个卷积核的大小为 3×3，使用 VALID 填充，sigmoid 激活函数，滑动步长为 1，卷积核的数量为 1。

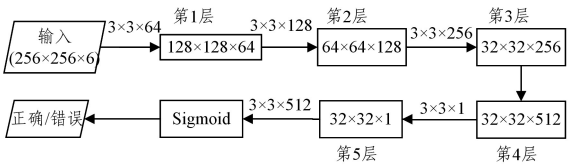

图 3　判别模型结构

Fig. 3　Discrimination model structure

步骤 8　定义损失函数，用于模型优化。对于生成模型，为了使模型在复原人脸图像时尽可能真实，使得判别模型难以区分真实图像和生成图像，为其定义损失函数，并期望其函数值最小。生成模型的输出图像 $I_1\in\mathbb{R}^{H\times W\times C1}$ 与输入的人脸轮廓图像 $I_2\in\mathbb{R}^{H\times W\times C2}$ 进行矩阵拼接，得到 $I_3\in\mathbb{R}^{H\times W\times(C1+C2)}$，将 I_3 输入至判别网络得到 $pred_{fake}$；将 I_2 与真实图像 $I_4\in\mathbb{R}^{H\times W\times C3}$ 进行矩阵拼接，得到 $I_5\in\mathbb{R}^{H\times W\times(C2+C3)}$，将 I_5 输入判别模型得到 $pred_{real}$。

首先从生成模型的角度定义损失函数：

$$Loss_1(pred_{fake},x_{out},x_{tar})=w_d\times Mean(-\log(pred_{fake}))+w_g\times Mean(abs(x_{out}-x_{tar})) \tag{6}$$

记真实图片为 x_{tar}，生成模型生成的图像为 x_{out}，用判别器判断生成图像的真假，判断值为 $pred_{fake}$，其范围为(0，1)。$pred_{fake}$ 越接近于 1，表示生成的图像越像真实图像，越接近 0 表示生成的图像与真实图像相差越大。w_d 和 w_g 为权重系数，用于平衡结果。用 $Loss_1(pred_{fake},x_{out},x_{tar})$ 来优化生成模型。

另外，对于判别模型，为了从输入图像中区分真实图像和生成图像，同样需要为其定义损失函数，并期望其函数值最小。

$$Loss_2(pred_{fake},pred_{real})=Mean(-\log(pred_{real})-\log(1-pred_{fake})) \tag{7}$$

其中，$pred_{fake}$ 为对生成图片的预测结果，$pred_{real}$ 为对真实图片预测的置信度，使用 $Loss_2(pred_{fake},pred_{real})$ 来优化判别模型。

步骤 9　利用训练集和损失函数迭代优化已构造的生成模型(生成器)和判别模型(判别器)。其中，生成器包含对称的编码层和解码层，编码层由 8 个卷积层和 1 个全连接层构成，使用整流线性单元(ReLU)[26]激活函数和 SAME 方式填充；解码层由 1 个全连接层和 8 个转置卷积层构成，使用 Leaky ReLU 激活函数和 SAME 方式填充。如图 4 所示，编码层对称地与解码层连接。判别器包含 5 个卷积层，使用 Leaky ReLU 激活函数和 VALID 方式填充。最后利用随机梯度下降算法训练模型，直至模型收敛。

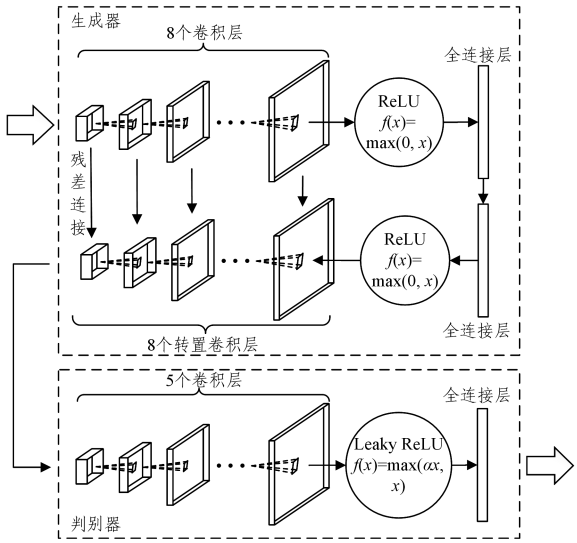

图 4　生成器和判别器的结构

Fig. 4　Generator and discriminator structure

3.2　模型推理

如图 5 所示，在训练得到神经网络模型后，可利用该模型复原人脸图像，步骤如下：

步骤 1　获得所需复原人脸的轮廓图。将输入图像的大小调整为 256×256×3，并使用 Dlib 库获取图像中的 68 个特征点，该过程与 3.1 节中的模型训练步骤 2 相同。

步骤 2　将轮廓图输入至神经网络模型，编码层将提取图像的高维特征，提取过程与训练过程相同，得到 100 维的信息矩阵，以便直接在后续步骤中传输或使用。

步骤 3　神经网络模型通过解码层将 100 维信息矩阵复原为 256×256×3 大小的图像，图像复原过程与训练过程中的解码层解码过程相同。

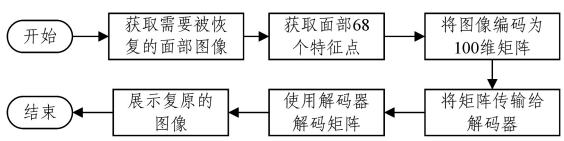

图 5　FR-RGAN 模型推理流程图

Fig. 5　Flow chart of FR-RGAN model inference

4　实验

本文采用的实验环境为：64 GB 内存，Intel® Xeon® E5-2630v3 处理器，2.40 GHz；GPU 为 NVIDIA GeForce GTX 1080Ti；操作系统为 Ubuntu 14.04；实现语言为 Python。在实验部分，使用 Tensorflow 来实现本文方法。为加快模型的训练速度，使用 GPU 进行高速计算。同时使用 cuDNN 加速库，进一步实现高性能并行计算。

4.1　数据集

本节使用了 3 个不同的包含人脸的视频来测试本文算法。因为该方法实现的是通过人脸轮廓来复原人脸，因此将

所提出的 FR-RGAN 算法仅与同类算法相比较，包括使用 CycleGAN、EBGAN 和条件对抗网络的图像到图像转换(pix2pix)[6]以及没有添加残差网络的原始模型。在下一节中会定性和定量地评估实验结果。

考虑到本模型的训练需要大量同一个人面部的特写图像，我们从互联网上选择了 3 个开源视频，这些视频长时间保持相似的姿势，便于获取训练数据集。在这 3 个数据集上分别测试了 FR-RGAN 的效果。

(1)A 数据集[27]包含 343 张图像，该图像取自个人演讲视频，在删除受干扰的帧之后，仅保留具有人体上半身并且面部正对相机的图像，保证面部可以在图像中被识别出来。

(2)B 数据集[28]包含 326 张人物语音视频的截图，所有图像都是纯蓝色背景，以减少背景内容变化对神经网络训练的干扰。所有图像按动作顺序排列，并根据我们系统的需要将图像大小调整为 256×256。

(3)C 数据集[29]包含共 343 张图像，这些图像来自语音视频的屏幕截图，背景中带有深绿色大理石线条。我们从最初捕获的大约 2 000 张图像中移除不包含人脸面部的图像，将保留的图像大小调整为 256×256。

4.2 视觉评估

首先尝试将 CycleGAN 与 BEGAN 模型引入到本实验的对比环节，在 A 数据集上进行训练生成的图像如图 6 和图 7 中的右图所示。

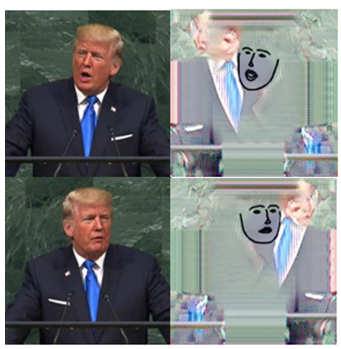

图 6 CycleGAN 复原效果图

Fig. 6 Restoration effect of CycleGAN

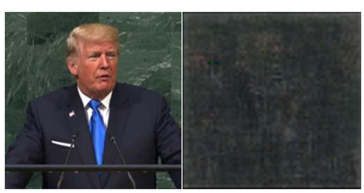

图 7 BEGAN 复原效果图

Fig. 7 Restoration effect of BEGAN

正如第 2 节所述，BEGAN 模型引入随机噪声参与生成图像的过程，因此生成的图像具有随机性，并不能很好地用于实现精确人脸图像复原；而 CycleGAN 主要应用于图像的风格迁移领域，并不能将人脸的轮廓完全迁移至目标生成图像，且实验证明其效果未达到预期目标，因此这两种方法将不参与进一步的对比讨论。

我们将 FR-RGAN 生成的图像与 pix2pix 生成的其他图像以及没有残差网络的原始模型进行比较。如图 8 和图 9 所示，非残差产生的图像显然最差。pix2pix 的结果更容易发生面部遮蔽或颈部和身体的不对称等问题。在 FR-RGAN 中，细节上的表现要好于其他方法的结果。在实验中，我们使用人脸轮廓的信息矩阵来复原面部。随着网络的深化和剩余网络的加入，我们可以看到结果的视觉效果得到了显著改善。

图 8 人脸图像复原效果对比图

Fig. 8 Comparison of face images restoration

图 9 人脸图像复原放大效果对比图

Fig. 9 Comparison of enlarged face images restoration

4.3 定量评估指标

在图像像素统计的基础上，峰值信噪比(PSNR)和均方误差(MSE)是两种常见的质量评估方法[30]。通过计算要评估的图像与参考图像之间的对应像素的灰度值之间的差异，从统计的角度来测量要评估的图像质量。如果要评估的图像是 F 并且参考图像是 R，Q 是图像量化的灰度级数，并且它们的尺寸是 $M\times N$，则 MSE 和 PSNR 可用如下公示表示：

$$PSNR=10\times\lg\frac{Q^2}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{i=1}^{N}\|R(i,j)-F(i,j)\|^2} \tag{8}$$

$$MSE=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\|R(i,j)-F(i,j)\|^2 \tag{9}$$

除了基于图像的像素统计方法，2002 年，Wang[31]首次在图像处理、图像压缩和图像视觉质量评价的研究工作的基础上，提出了结构信息的概念[32]。他们声称人类视觉系统的主要功能是在后台提取结构信息，而人类视觉系统可以实现这一目标的高度适应性。因此，图像结构失真的测量应该是图像感知质量的最佳近似。在此基础上，提出了符合人类视觉系统特征的图像质量客观评价标准——结构相似度(SSIM)。

SSIM 基于图像像素之间的相关性构建结构相似性。假设给出尺寸为 $M\times N$ 的两个图像 X 和 Y，$u(X)$ 及 $u(Y)$、$\sigma(X)$ 及 $\sigma(Y)$分别表示 X 和 Y 的平均值和标准差，$\sigma(XY)$表示 X 和 Y 的协方差。亮度 $l(X,Y)$、对比度 $c(X,Y)$和结构 $s(X,Y)$的比较功能定义如下：

$$\begin{cases} l(X,Y)=\dfrac{2u(X)u(Y)+c1}{u(X)^2+u(Y)^2+c1} \\ c(X,Y)=\dfrac{2\sigma(X)\sigma(Y)+c2}{\sigma(X)^2+\sigma(Y)^2+c2} \\ s(X,Y)=\dfrac{\sigma(XY)+c3}{\sigma(X)\sigma(Y)+c3} \end{cases} \tag{10}$$

常数 $c1$，$c2$ 和 $c3$ 用于调节接近零的分母的不稳定性。这 3 个组成部分合并为 SSIM 指标。

$$SSIM(X,Y)=[l(X,Y)]^{\alpha}[c(X,Y)]^{\beta}[s(X,Y)]^{\gamma} \quad (11)$$

其中，α，β 和 γ 为大于 0 的常数。SSIM 根据图像像素之间的相关性构建参考图像和待评估图像之间的结构相似性。SSIM 值越大，图像质量就越好。

4.4 定量比较

在 3 个数据集上测试不同的模型，并计算评估指标 MSE，PSNR 和 SSIM，结果如图 10 所示。

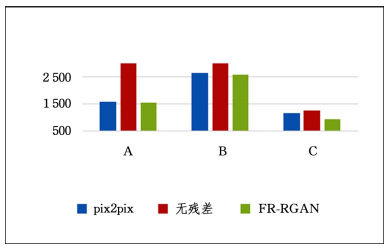

图 10 3 种算法在 3 个数据集上的 MSE 性能比较

Fig. 10 MSE performance comparison of three algorithms on three data sets

从图 10 可以看出，在 3 个测试集上残差模型 FR-RGAN 的性能明显优于 pix2pix 和非残差模型。FR-RGAN 在 3 个数据集上的 MSE 值为 1553.80/2565.66/948.09；pix2pix 和非残差模型的 MSE 值分别为 1568.60/2625.53/1163.72 和 3188.72/3121.02/1265.52。

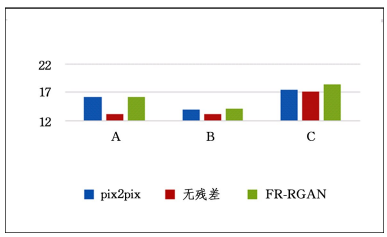

图 11 3 种算法在 3 个数据集上的 PSNR 性能比较

Fig. 11 PSNR performance comparison of three algorithms on three data sets

为了进一步评估模型的质量，我们引入了 PSNR，这是评估图像质量的最常用和最广泛使用的客观测量方法。所提出的 FR-RGAN 模型在 PSNR 上表现最佳，其 PSNR 值为 16.25/14.04/18.44。此外，pix2pix 模型的性能为 16.21/13.95/17.50，优于非残差模型 13.10/13.22/17.14。

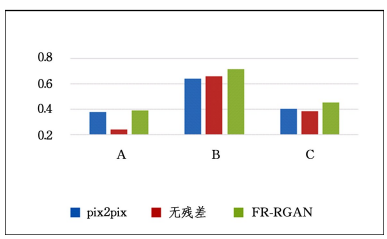

图 12 3 种算法在 3 个数据集上的 SSIM 性能比较

Fig. 12 SSIM performance comparison of three algorithms on three data sets

除基于图像的像素统计方法之外引入了 SSIM 来表征结构相似性。经过比较，FR-RGAN 的 SSIM 值仍然最高，为 0.39/0.71/0.45。非残差模型具有最差的性能，其 SSIM 值为 0.24/0.65/0.38。pix2pix 模型的 SSIM 值为 0.37/0.64/0.40。

表 1 3 个测试集上的 MSE 值、PSNR 值(dB)、SSIM 值

Table 1 MSE, PSNR(dB) and SSIM values on three test sets

	MSE	PSNR	SSIM
A	1568.60/3188.72/**1553.80**	16.21/13.10/**16.25**	0.37/0.24/**0.39**
B	2625.53/3121.02/**2565.66**	13.95/13.22/**14.04**	0.64/0.65/**0.71**
C	1163.72/1265.52/**948.09**	17.50/17.14/**18.44**	0.40/0.38/**0.45**

注：左/中/右的结果分别是 pix2pix/无残差/FR-RGAN

总而言之，通过在 3 个测试集上的验证，在定量分析后可知，我们提出的 FR-RGAN 模型性能最优，在 MSE，PSNR 和 SSIM 指标上比 pix2pix 分别提高了 8.7%，2.1%和 9.6%，比无残差方法分别提高了 53.4%，12.6%和 30.1%。我们将原因归结为以下几点：1）深度卷积神经网络的应用。使用更深层次的卷积网络结构，使其在原有基础上能够提取图像中更加高维的信息，该信息可能表示更加完整的人脸信息。2）残差思想的应用。当网络层数加深时，反向传播时的梯度消失导致神经网络的表现越来越差，很难训练到前层的网络，残差思想的应用使得梯度更容易流向前层网络，使得梯度消失问题得以缓解，同时为训练更深层的网络提供了条件。3）卷积核大小的减小。其增加了非线性变换的操作，使得模型的泛化能力进一步得到提高，同时也使得在处理相同大小图片时的计算量减小，有利于模型深度的增加和计算性能的提高。

结束语 本文讨论了常见的人脸修复和复原技术，并提出了一种新的残差生成对抗网络人脸复原方法 FR-RGAN，该方法利用人脸的轮廓来复原面部局部细节，使其生动地呈现出来，可用于多种场合，具有广阔的研究和应用前景。FR-RGAN 方法的优化可以从定性和定量角度改善性能。在 3 个不同的数据集上使用 3 种不同的算法进行的对比实验可以看到，FR-RGAN 算法在 MSE，PSNR 和 SSIM 性能指标方面均表现优异。结果表明，所提出的算法为需要面部复原的应用提供了良好的选择。在今后的工作中，我们计划将其应用在视频通话领域，以实现在视频聊天中发送方将人脸轮廓图像编码传输，接受方对其解码复原。这样，一方面可节约流量带宽，降低延迟，另一方面在网络环境拥塞的情况下也能正常视频通话。

参 考 文 献

[1] SHEN J, CHAN T F. Mathematical models for local nontexture inpaintings[J]. SIAM Journal on Applied Mathematics, 2002, 62(3): 1019-1043.

[2] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[C]// ACM Transactions on Graphics (ToG). ACM, 2009, 28(3): 24.

[3] LI Y, LIU S, YANG J, et al. Generative face completion[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3911-3919.

[4] YEH R A, CHEN C, YIAN LIM T, et al. Semantic image inpainting with deep generative models[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5485-5493.

[5] GAUTHIER J. Conditional generative adversarial nets for convolutional face generation [J]. Class Project for Stanford

CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester, 2014, 2014(5): 2.
[6] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1125-1134.
[7] HAYS J, EFROS A A. Scene completion using millions of photographs[J]. ACM Transactions on Graphics (TOG), 2007, 26(3): 4.
[8] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2536-2544.
[9] KINGMA D P, WELLING M. Auto-encoding variational bayes [J]. arXiv: 1312. 6114, 2013.
[10] LARSEN A B L, SØNDERBY S K, LAROCHELLE H, et al. Autoencoding beyond pixels using a learned similarity metric [J]. arXiv: 1512. 09300, 2015.
[11] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. 2014: 2672-2680.
[12] DENTON E L, CHINTALA S, FERGUS R. Deep generative image models using a laplacian pyramid of adversarial networks [C]// Advances in Neural Information Processing Systems. 2015: 1486-1494.
[13] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv: 1511. 06434, 2015.
[14] SUN Q, ZENG X Q. Image Inpainting Based on Generative Adversarial Networks [J]. Computer Science, 2018, 45(12): 229-234, 261.
[15] CHENG X Y, XIE L, ZHU J X, et al. Review of Generative Adversarial Network [J]. Computer Science, 2019, 46(3): 74-81.
[16] XU Q, ZHONG S P, CHEN K Z, et al. Optimized Selection Method of Cycle-consistent Loss Coefficient of CycleGAN in Image Generation with Different Texture Complexity[J]. Computer Science, 2019, 46(1): 100-106.
[17] LIU F, LI Z W, et al. A Text-Based CAPTCHA Cracking System with Generative Adversarial Networks[C]// 2018 IEEE International Symposium on Multimedia (ISM). IEEE. 2018.
[18] REED S, AKATA Z, YAN X, et al. Generative adversarial text to image synthesis[J]. arXiv: 1605. 05396.
[19] ZHU J Y, PARK T, ISOLA P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. International Conference on Computer Vision, 2017: 2242-2251.
[20] BERTHELOT D, SCHUMM T, Metz L, et al. BEGAN: Boundary Equilibrium Generative Adversarial Networks[J]. arXiv: Learning, 2017.
[21] KING D E. Dlib-ml: A machine learning toolkit [J]. Journal of Machine Learning Research, 2009: 1755-1758.
[22] KRIZHEVSKY A, ILYA S, GEOFFREY E H. Imagenet classification with deep convolutional neural networks[C]// Advances in neural information processing systems. 2012.
[23] HE K M, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
[24] IOFFE S, CHRISTIAN S. Batch normalization: Accelerating deep network training by reducing internal covariate shift [J]. arXiv: 1502. 03167.
[25] XU B, WANG N, CHEN T, et al. Empirical evaluation of rectified activations in convolutional network [J]. arXiv: 1505. 00853, 2015.
[26] NAIR V, GEOFFREY E H. Rectified linear units improve restricted boltzmann machines[C]// Proceedings of the 27th International Conference on Machine Learning. 2010.
[27] man-talking [EB/OL]. [2019-04-03]. https://www.pexels.com/video/man-talking-1769632/.
[28] Putin delivers annual address to Russia's Federal Assembly [EB/OL]. [2019-04-03]. https://www.youtube.com/watch?v=P6HM9pKrxqE.
[29] CHANNEL 90 seconds TV. Official Channel, President Trump speech to the 72nd Session of the UN General Assembly [EB/OL]. [2019-04-03]. https://www.youtube.com/watch?v=pyZ965-3qP4&t=3957s.
[30] HUYNH-THU Q, MOHAMMED G. Scope of validity of PSNR in image/video quality assessment [J]. Electronics Letters, 2008: 800-801.
[31] WANG Z. The SSIM index for image quality assessment[OL]. https://ece.uwaterloo.ca/~z70wang/research/ssim.
[32] HORE A, DIEMEL Z. Image quality metrics: PSNR vs. SSIM [C]// 2010 20th International Conference on Pattern Recognition. IEEE, 2010.

LI Ze-wen, born in 1998, senior undergraduate. His main research interests include computer vision and distributed machine learning.

XIE Zai-peng, born in 1982, associate professor. His main research interests include distributed and embedded machine learning.