文章编号：2095-2163(2024)03-0159-04 中图分类号：TP391.41 文献标志码：A

基于条件生成对抗网络的大视角单图像人脸纹理重建

孙 进，周 威，谢文涛
（扬州大学 机械工程学院，江苏 扬州 225100）

摘 要：针对当前的人脸重建方法、尤其是纹理重建部分，应用于自遮挡或物体遮挡区域表现不佳，使得人脸纹理重建后的结果不真实的问题，本文提出了基于条件生成对抗网络的大视角单图像人脸纹理重建的方法，实现人脸图像的纹理补全。补全网络基于改进的条件生成对抗网络，包括编码器和解码器的粗层之间的跳跃连接来保存高频细节；每个卷积层的输出上叠加了高斯噪声映射；将 U-V 纹理映射与其翻转版本共同连接输入的方法来提高纹理重建的质量以及真实性。使用 Multi-PIE 数据集与 CFP 数据集进行评估，整体网络能够实现更高的纹理重建精度，尤其在±90°图像重建上，能获得更为完整的纹理图像。

关键词：纹理补全；单图像；大视角；生成对抗网络；跳跃连接

Large view single image face texture reconstruction based on conditional generation adversarial network

SUN Jin, ZHOU Wei, XIE Wentao
(School of Mechanical Engineering, Yangzhou University, Yangzhou 225100, Jiangsu, China)

Abstract: In view of the problem that the current face reconstruction methods, especially the texture reconstruction part, often perform poorly when applied to self-occlusion or object occlusion area, which makes the result of face texture reconstruction unreal, this paper proposes the method of face texture reconstruction based on conditional generation adversarial network for large view single image to achieve texture completion of face images. The completion network based on the improved conditional generation adversarial network, including the skip connection between the coarse layers of encoder and decoder to preserve the high frequency details, the superimposed Gaussian noise map on the output of each convolutional layer, and the method of connecting the U-V texture map with its flipped version to the input to improve the quality and authenticity of texture reconstruction. Experimental results: Using Multi-PIE data set and CFP data set respectively for comparative experimental evaluation, the overall network can achieve higher texture reconstruction accuracy, especially in ±90° image reconstruction, and can obtain more complete texture images.

Key words: completion of texture; single image; the big view; generation adversarial network; skip connection

0 引 言

人脸作为人体最重要的生物特征，含有丰富的特征信息，对于年龄、性别的判断、喜怒的识别以及真实身份的确认等方面有很大的作用[1-3]。基于图像的三维人脸重建成为计算机视觉研究的基础性问题。目前，由于图像拍摄角度的苛刻性，重建方法在物体遮挡或自遮挡区域表现不佳，使得纹理重建后的结果并不真实可靠，导致单图片输入的三维人脸模型的重建不准确，影响在实际需求场景中的使用。

1 人脸纹理补全相关工作

目前基于单张图像的三维人脸重建方法大致可以分为 3 类：基于单一模板的三维人脸重构、基于统计模型的三维人脸重构以及基于深度学习的三维人脸重构[4-7]。

生成对抗网络在深度学习中主要体现在网络中的生成模型（generative model）和判别模型（discriminative model）间的相互博弈学习[8-9]，其中生成模型的目标是不断学习并生成与真实数据分布相

基金项目：扬州市市校合作专项（YZ2022195）；扬州市产业前瞻与共性关键技术-产业前瞻技术研发项目（YZ2022022，YZ2021020）。
作者简介：周 威（1998-），男，硕士研究生，主要研究方向：机器视觉、人脸纹理重建。
通讯作者：孙 进（1973-），男，博士，副教授，硕士生导师，主要研究方向：机器视觉、人脸三维建模与分析以及医疗辅助机械。Email：sunjin@yzu.edu.cn
收稿日期：2023-03-28

似的样本;判别模型的目标是尽可能地将输入的真实数据与生成模型生成的伪数据区分开。在不断迭代的过程中,当生成模型和鉴别模型达到均衡的状态时,可以认为生成模型学习到了真实数据的近似分布,即达到生成对抗网络的理想效果。为了重建出具有高频细节的三维人脸,Lattas 等学者[10]提出了 AvatarMe 算法,首先对可变形的参数化模型(3D Morphable Model,3DMM)与输入图像相匹配形成的 UV 纹理(UV 表示二维坐标系中的 U 和 V 轴)进行多次上采样操作,得到确切的高频人脸细节;其次,通过去除 UV 图上的纹理信息以获取具有高频细节的漫反射率;最后,根据漫反射率和 3DMM 的法线来计算出图像的镜面反照率、漫反射法线等。该算法将上述特征分别映射回 UV 图像上,然后将该 UV 图像投影到重建的三维人脸模型,完成高频细节的三维人脸的重建。Jabbar[11]使用堆叠生成对抗网络实现人脸重建,利用生成粗略面部图像的第一阶段网络和生成真实逼真图像的第二阶段网络,引入多种损失函数实现人脸重建。Gecer 等学者[12]利用深度卷积神经网络作为面部纹理预处理,并结合自回归监督的生成对抗网络实现三维的逼真面部重建。Gecer 等学者[13]利用图卷积神经网络和生成对抗网络分别对人脸的三维形状和纹理进行重建,采用不确定性感知编码器以及非线性解码器模型实现三维面部重建。

现有的三维人脸重建模型中的关于纹理补全方法并不能保证补全细节信息的同时,拥有较高的补全相似率,因此,本文提出了基于条件生成对抗网络的大视角单图像人脸纹理重建的方法。

2 本文方法

2.1 整体网络结构

整体网络由 2 个模块组成:分割网络与补全网络,如图 1 所示。

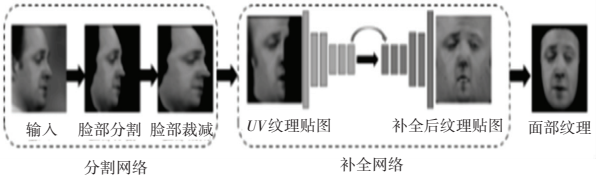


图 1 整体网络结构

Fig. 1 Global network structure

2.2 分割网络

分割网络由 2 个像素级预测算法组成,分别是:脸部分割算法和纹理贴图算法。基于像素的预测需要适当的网络架构,在编码器和解码器的相应层之间使用跳跃连接。这里,脸部分割方法接收人脸图像并输出掩码,纹理贴图方法接收分割后的人脸作为输入输出纹理映射。输出的纹理映射用于将人脸图像映射到预定义的 $U-V$ 映射中。

使用二元交叉熵损失函数来训练脸部分割方法数学公式可表示为:

$$L_c = -\frac{1}{n}\sum_{i=1}^{n}(C_{g,i}\log C_{p,i} + (1-C_{g,i})\log(1-C_{p,i})) \quad (1)$$

其中,$C_{p,i}$ 表示面部图像中第 i 个像素的输出;$C_{g,i}$ 表示对应像素的真实值;n 表示面部图像的像素数。

使用逐像素 $L1$ 损失函数来训练纹理贴图方法,推得的数学公式为:

$$L_p = \| R_g - R_p \|_1 \quad (2)$$

其中,R_p 是预测的纹理映射图,R_g 是对应像素的真实值。

2.3 补全网络

将人脸图像与对应的纹理贴图进行 $U-V$ 空间弯曲,得到 $U-V$ 纹理映射。$U-V$ 纹理图由于自聚焦而不完整,因此使用补全网络填补缺失的区域,在这个任务中引入对抗性训练。

2.3.1 生成器模块

Iizuka 等学者[14]提出了生成局部和整体图像的补全方法,可以通过填充任何形状的确实区域来完成目标图像的生成。本方法使用该方法的生成器作为初始网络,并做出一些修改与改进,在编码器和解码器的粗层之间使用跳跃式连接来保存高频细节。初始网络的训练数据的 10%是具有 90°偏航的半闭塞式图像,$U-V$ 纹理图需要接近一半的区域内嵌,在粗分辨率层中采用空洞卷积来扩大接受野并提供更好的补全。为了使内嵌的 $U-V$ 纹理映射在视觉上更加逼真,本文在每个卷积层的输出上叠加了高斯噪声映射,进一步将 $U-V$ 纹理映射与其翻转版本共同连接输入,形成一个六通道输入,翻转的 $U-V$ 纹理映射相较于随机噪声能更好地初始化缺失的纹理区域,生成器结构如图 2 所示。

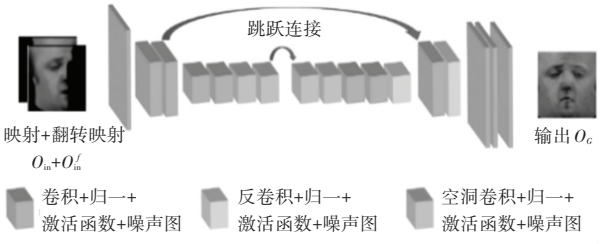


图 2 补全网络生成器结构

Fig. 2 Complete the network generator structure

使用 $L1$ 损失函数来训练补全网络,数学表达如式(3)所示:

$$L_{mae} = \| O_{gt} - O_G \|_1 \tag{3}$$

其中，$O_G = G(O_{in}, O_{in}^f)$；$G$ 为生成器；O_{gt}、O_{in}、O_{in}^f 分别为 $U-V$ 纹理映射的真实值、输入和翻转后的输入。

为了保留细节内容，采用广泛使用的感知损失，此处的数学公式可写为：

$$L_{per} = \| V(O_G) - V(O_{gt}) \|_1 \tag{4}$$

其中，V 为预训练深度模型的激活映射。

2.3.2　鉴别器模块

通过生成器模块虽然可以得到一个完成的 $U-V$ 纹理映射，但是得到的纹理缺乏细节。为了提高内嵌纹理的质量，本文引入了对抗学习，鉴别器的损失函数见式(5)：

$$L_d = -\underset{O_{gt}\in M}{E}[\log(D(O_{gt}))] - \underset{O_G\in N}{E}[\log(1 - D(O_G))] \tag{5}$$

其中，M 和 N 分别表示真实值 $U-V$ 纹理图的分布和生成器的输出。

为了训练鉴别器，使用的目标函数具体如下：

$$L_g = -\underset{O_G\in N}{E}[\log D(O_G)] \tag{6}$$

为了使纹理网络能够稳定地训练，在鉴别器中加入一个梯度惩罚正则化项[15]，其表达式为：

$$L_{st} = E[\| \nabla D(O_G) \|^2] \tag{7}$$

其中，$\nabla D(O_G)$ 为鉴别器相对于输入 O_{in} 的梯度。

3　实验评估

本文网络使用 Python 语言，在 Windows11 系统上训练。实验硬件环境：CPU Intel(R) core(TM) i5-11260H，主频 2.60 GHz；GPU NVIDA GeForce RTX 3050；深度学习环境为 Python 3.6.13，Pytorch 1.12.1。

使用 Multi-PIE 数据集[16]与 CFP 数据集的对应图像数据评估分割网络的可行性与准确性，以部分大角度人脸图像作为原始数据，在已有的面部特征提取器的基础上，利用补全网络实现对面部特征的正面化重建。

3.1　分割网络评估

面部轮廓上的可见标志是 $U-V$ 纹理映射边界的重要标志信息之一。Multi-PIE 数据集[16]包含大量的多姿态的面部图像。本文选择具有极端姿态[±60°，±90°]的面部图像，作为分割网络评估的测试图像集。将本文方法与 3DDFA 方法[17]和 Deep3DFace 方法[18]进行比较，采用 *NME*(Normalized Mean Error)作为评估度量，即鼻子长度归一化后的平均误差，对比结果见表 1。

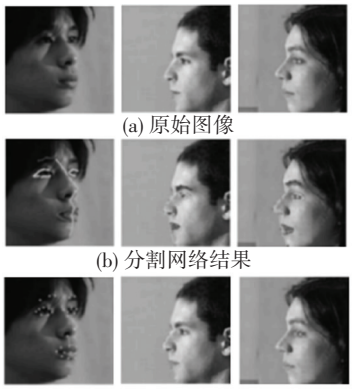
表 1　具有大姿态[60°，90°]图像的轮廓定位结果

Table 1　Contour location results of the image with large attitude [60°，90°]

方法	3DDFA[17]	Deep3DFace[18]	本文方法
NME/ %	19.13	17.41	13.22

由表 1 可知，本文的分割网络的 *NME* 较 3DDFA 与 Deep3DFace 有了很大的下降，表明分割网络优于其他方法，即使在极端姿势下，也能捕捉到 2D 人脸图像和 3D 模型之间的精确对应关系。

Multi-PIE 数据集不同角度的剖面标志定位结果如图 3 所示。由图 3 可以看出通过分割网络标记出来的各个特征线与实际的轮廓标记是非常接近的。



(a) 原始图像

(b) 分割网络结果

(c) 真实值轮廓标记

图 3　分割网络的定性评估

Fig. 3　Qualitative evaluation of split networks

3.2　补全网络评估

通过低维度 3D 面部重建建立 2D 面部图像和 3D 模型之间的精确对应关系，较其他方法可以产生更好的纹理推断。为了评估补全网络的有效性，将其应用于姿态不变人脸识别，将重建的 3D 纹理人脸投影到具有正面视图的图像平面中。将从人脸网络中提取的探针特征向量及其前端版本进行融合，计算融合后的探针特征与图库特征之间的余弦距离作为相似度度量。以 CFP 作为多视图人脸数据集，选取数据集中姿态大于 45°的人脸作为评估图像，使用预训练的 Light CNN 模型[6]作为面部特征的提取器，进而通过补全网络实现对非极端姿态的重建。

c-CNN 方法[18]、Light-CNN 方法[19]和本文方法在 CFP 数据集上，对不同视角与相同光照下人脸图像重建的结果对比见表 2。由表 2 可知，本文方法的补全网络对不同视角的人脸重建效果较其他方法都有一定的提升。视角为±45°、±60°以及±75°的人脸图像的重建相似度有较小的提高，但对视角为±90°人脸图像的重建相似度有超过 8.51%的提升。

表 2 CFP 不同视角与相同光照下的重建相似度

Table 2 Similarity of reconstruction under different view angles and the same illumination

方法	±90°	±75°	±60°	±45°
c-CNN[19]	47.31	59.70	76.60	89.10
Light CNN[20]	54.80	89.60	98.66	99.77
本文	63.31	91.18	99.21	99.92

CFP 数据集的人脸正面化结果如图 4 所示。由图 4 可见,该补全网络可以较好地完成对于大视角图像的纹理补全任务。

(a) 原始图像

(b) 正面图像

(c) 真实值人脸

图 4 CFP 数据集的正面化结果

Fig. 4 Positive results of the CFP data set

4 结束语

本文提出了基于条件生成对抗网络的大视角单图像人脸纹理重建方法,用来解决人脸重建中单个图像难以推断人脸纹理的问题。整体网络分为 2 个子网络,即分割网络与补全网络。分割网络建立 2D 人脸图像与 3D 模型之间的对应关系,方便补全网络对纹理进行更好的推断,在 Multi-PIE 数据集上的特征线标记优于其他方法。补全网络基于条件生成网络的方法,利用现有的网络进行生成器上的改进,将编码器和解码器的粗层之间使用跳跃式连接来保存高频细节,并且使每个卷积层的输出上叠加了高斯噪声映射,以及将 $U-V$ 纹理映射与其翻转版本共同连接输入,在 CFP 数据集上的正面化重建相似度,较 c-CNN 以及 Light-CNN 在小角度图像有较小的提高,但在视角为±90°的人脸图像有超过8.51%的提升。虽然对称性对于面部纹理贴图是较为合理的,但该方法并不适合某些面部细节,例如嘟嘴或歪笑便会破坏面部纹理的对称性,另外整个框架是串联进行的,意味着某一步的错误会累积叠加到最后的结果,该整体框架鲁棒性有待提高。

参考文献

[1] XIAO H, LONG W, LI Z, et al. The progress of face recognition on artificial intelligence [C]//ISHC. Shanghai: University of Toronto, 2022: 220-226.

[2] MUSTAFA A. Evaluation of gender bias in masked face recognition with deep learning models[C]//2023 IEEE SSCI. Mexico: IEEE, 2023: 829-835.

[3] FAN Yu, LUO Yiyue, CHEN Xianjun. Research and design of face and expression recognition system based on convolutional neural network[C]//2021 IEEE AEECA. Dalian, China: IEEE, 2021: 459-462.

[4] KHAN A, HAYAT S, AHMAD M, et al. Learning-detailed 3D face reconstruction based on convolutional neural networks from a single image[J]. Neural Computing & Applications, 2021(11):33.

[5] ZHAO D, CAI J, Qi Y. Convincing 3D face reconstruction from a single color image under occluded scenes[J]. Electronics, 2022, 11(4):543-557.

[6] ZHANG Jian, LI Ke, LIANG Yun, et al. Learning 3D faces from 2D images via stacked contractive autoencoder [J]. Neuro-Computing, 2017,257:67-78.

[7] ZHANG Jian, ZHU Chaoyang. Approach to 3D face reconstruction through local deep feature alignment[J]. IET Computer Vision, 2019,13(2):213-223.

[8] GOODFELLOW I, POUGET-ABADIE J, et al. Generative Adversarial Networks [J]. Communications of The ACM, 2020, 63(11): 139-144.

[9] MIRZA M, OSINDERO S. Conditional Generative Adversarial Nets [J]. arXiv preprint arXiv: 1411.1784, 2014.

[10] LATTAS A, MOSCHOGLOU S, GECER B, et al. Avatarme: Realistically renderable 3d facial reconstruction ´in-The-wild´[J]. arXiv preprint arXiv:2003.13845v1,2020.

[11] JABBAR A. FD-StackGAN: Face de-occlusion using stacked Generative Adversarial Networks[J]. KSII Transactions on Internet and Information Systems, 2021, 15(7): 2547-2567.

[12] GECER B, PLOUMPIS S, KOTSIA I, et al. Fast-GANFIT: Generative Adversarial Network for high fidelity 3D face reconstruction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44: 4879-4893.

[13] GECER B, PLOUMPIS S, KOTSIA I , et al. Fast-GANFIT: Generative Adversarial Network for high fidelity 3D face reconstruction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(9): 4879-4893.

[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] SENGUPTA S, LICHY D, KANAZAWA A, et al. SfSNet: Learning shape, reflectance and illuminance of faces in the wild [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 3272-3284.

[16] GROSS R, MATTHEWS I. Multi-PIE [J]. Image and Vision Computing, 2010, 28(5): 807-813.

[17] ZHU Xiangyu, LIU Xiaoming, LEI Zhen, et al. Face alignment In full pose range: A 3D total solution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,41(1): 167-178.

[18] DENG Yu, YANG Jiaolong, XU Sicheng, et al. Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set[J]. CoRR, 2019, 42(2): 213-224.

[19] WU Xiang, HE Ran, SUN Zhenan, et al. A light CNN for deep face representation with noisy labels[J]. IEEE Transactions on Information Forensics and Security,2018,13(11): 1-13.

[20] XIONG Chao, ZHAO Xiaowei, TANG Danhang, et al. Conditional Convolutional Neural Network for modality-aware face recognition [C]//IEEE International Conference on Computer Vision (ICCV). Santiago, Chile :IEEE, 2016: 1-9.