基于 GANs 无监督回归三维参数化人脸模型

张星星[1],李金龙[2]

(1.中国科学技术大学 软件学院,安徽 合肥 230026;
2.中国科学技术大学 计算机科学与技术学院,安徽 合肥 230026)

摘 要:针对神经网络回归训练过程中三维人脸数据稀缺的问题,提出了基于生成对抗网络(GANs)回归三维参数化人脸模型(3DMM)的无监督学习方式。首先利用 GANs 的对抗生成训练使生成器回归的 3DMM 参数符合真实感人脸形状的参数分布。随后将生成的三维人脸网格渲染成二维图片,利用身份编码器对输入人脸及渲染人脸分别提取身份特征向量,通过不断缩小向量之间的距离使得生成的三维人脸网格靠近输入人脸的身份特征。实验结果表明,该方法在重建结果顶点位置准确性上相对于现有的方法有明显的提升,且拥有较好的 RMSE 值,能够较好应用于三维人脸重建任务。

关键词:三维人脸重建;生成对抗网络;无监督

中图分类号:TP391.41　　**文献标识码**:A　　**DOI**:10.19358/j.issn.2096-5133.2020.11.008

引用格式:张星星,李金龙. 基于 GANs 无监督回归三维参数化人脸模型[J].信息技术与网络安全,2020,39(11):50-55.

GANs-based unsupervised regression for 3D morphable model

Zhang Xingxing[1], Li Jinlong[2]

(1.School of Software Engineering, University of Science and Technology of China, Hefei 230026, China;
2.School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: Aming at the problem of scarcity of 3D face ground-truth data in the training process of neural network regressing for 3D facial mesh, an unsupervised learning method based on Generative Adversarial Networks(GANs), which regresses for 3D Morphable Model(3DMM) parameters, is proposed. Firstly, the adversarial training process of GANs is used to make the generated 3DMM parameters conform to the realistic 3DMM parameters distribution. Then, the generated 3D facial mesh is rendered into a 2D image. Later, the identity encoder is used to extract the identity feature embeddings from the input face and the rendered face respectively. By continuously minimizing the distance between the two embeddings, the generated 3D facial mesh is enforced to maintain the identical features of the input face. The experimental results show that the proposed method has a significant improvement in the accuracy of vertex position compared with the existing methods, and has a good RMSE value, which can be well applied to 3D face reconstruction tasks.

Key words: 3D face reconstruction; generative adversarial networks; unsupervised

0 引言

三维人脸重建是指通过一张或多张同一个人的照片来构建该人的三维人脸网格。该课题一直是计算机视觉和图形学的热门关注焦点,拥有广泛的应用场景,如人脸身份识别、医学方案展示、三维人脸动画等。

在三维人脸重建领域,VETTER T 和 BLANTZ V 在 1999 年提出的三维人脸参数化模型(3DMM)[1]具有重要意义。3DMM 采集了 200 位实验对象的脸部激光扫描数据集,并对该数据集进行主成分分析(PCA)。通过对 PCA 所提取的基向量进行线性组合从而构成一张新的人脸。

传统的三维人脸重建基于迭代方法[2],即针对输入人脸,利用人脸关键点,反复调整基向量的参

数使得三维人脸渲染后提取的人脸关键点与二维人脸关键点接近,以此达到具有输入人脸特征的三维人脸网格。然而,该方法较为依赖人脸关键点的检测结果,在人脸姿势较大或有遮挡物时,效果较差,迭代过程耗时也较长。

近年来,随着深度学习的不断发展,越来越多的研究开始运用基于回归的方法重建三维人脸。然而,在神经网络的训练过程中,一个亟需解决的问题便是三维人脸训练数据稀少。针对这一问题,部分研究提出利用合成数据[3-4],即先随机初始化3DMM的参数作为参照的三维人脸,而后将该三维人脸投影成的二维人脸作为输入数据,进而扩大训练数据集。因为合成数据投影形成的二维图片不能反映真实世界的复杂度,故 GENOVA K[5]提议采用真实图片及合成图片的混合数据集进行两步训练。TEWARI A[6]利用编码解码器结构直接从单张图片重建三维人脸,解码器是基于专业知识精心设计的,但可扩展性较低。TRAN A T[7]等人提议利用迭代方法形成的三维人脸作为神经网络所需的配对三维人脸数据进行训练。

本文基于前人的思想,提出采用 GANs 神经网络回归 3DMM 模型参数进行三维人脸重建任务。在解决三维人脸数据稀少问题上,本文提出两种并列的监督信号:(1)二维监督信号:利用三维人脸投影后的二维人脸与输入的二维人脸身份差异及皮肤颜色差异,来提供二维层面的监督信号,使得二者相近;(2)三维监督信号:利用重构的三维人脸顶点分布与普遍三维人脸顶点分布差异,来提供三维层面的监督信号,以使得重构后的三维人脸具备真实感人脸形状。由于仅依赖二维监督信号可能会导致一些重构后三维人脸顶点离正常人脸顶点偏差较大,虽然投影结果依旧初具人脸形状,仍能被系统识别,但视觉感受却与普遍人脸形状相差较大。其原因在于缺少三维监督信号,使得重构后的三维人脸顶点分布近似于普遍三维人脸顶点分布。本文拟采用生成对抗网络(GANs)[8]来提供三维监督信号,利用生成器及判别器的对抗生成,指引人脸顶点分布接近于真实感人脸顶点分布。

1 方法

1.1 网络结构

本文所采取的网络结构如图 1 所示,共包含四个部分:(1)生成对抗网络(GANs)[8],利用生成对抗网络产生符合真实人脸分布的 398 个 3DMM 模型参数;(2)3DMM[1]模型,通过生成的 3DMM 参数重新构建三维人脸网格;(3)可微分渲染器,将重构的三维人脸网格渲染为二维渲染图片;(4)身份编码器,利用人脸身份特征识别网络提取输入及渲染身份特征向量。

1.1.1 生成对抗网络 GANs

(1)生成器

假设输入图片 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$,$x_i\in\mathbf{R}^{224\times224}$,生成器的输出为 $\hat{\boldsymbol{y}}=(\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n)$,$\hat{y}_i\in\mathbf{R}^{398}$,其中,$\hat{y}_i$ 是预测得到的 3DMM 模型参数,具体含义详见 1.1.2 节。

对于生成器,采用修改过后的 Resnet50。Resnet[9]使用的具体残差块如图 2 所示,其中,x 是输入,$F(x)$ 是经过两层卷积层学习得到的特征,最后的输出是 $F(x)$ 和 x 的叠加结果。由于梯度可以从两条支路进行传播,从而解决了随着网络的层数加深,梯度消失的现象。

采用 Resnet50 作为生成器的目的是为了提升深

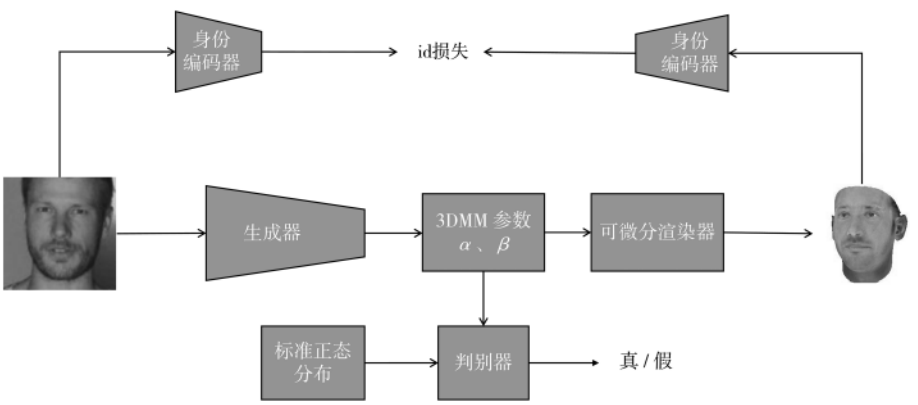

图 1 三维人脸重建网络结构

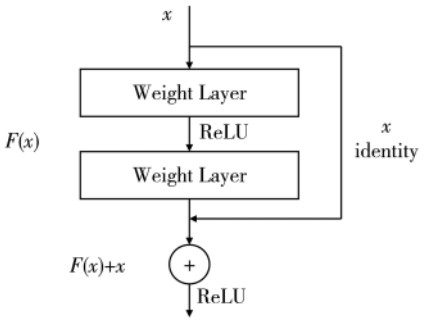

图 2 Resnet 残差块

度学习的训练效果。为了回归 3DMM 模型 398 个参数,将 Resnet50 中最后一层 1 024-D 全连接层改为 398-D。

(2)判别器

对于判别器来说,输入共两类,均为 398 维向量 $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)$, $\hat{\boldsymbol{y}}=(\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n)$, $y_i,\hat{y}_i\in\mathbf{R}^{398}$。其中,$y_i$ 是从符合真实人脸 3DMM 参数分布 $p_{3DMM}(\boldsymbol{y})$(详见 1.1.2 节)中取样而得,而 $\hat{y}_i$ 为生成器回归而得的 3DMM 模型参数。

由于判别器的输入仅一维,故而采用 3 层全连接层对生成器的回归结果是否符合真实人脸 3DMM 参数分布进行判断。判别器的网络结构如图3所示。

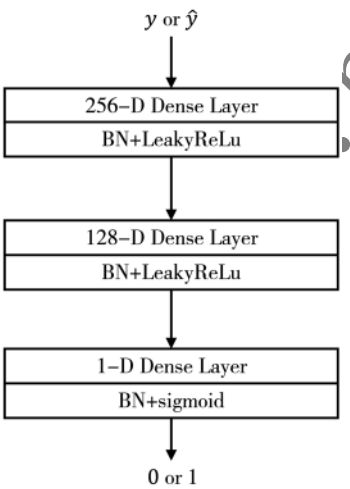

图 3 判别器网络结构

1.1.2 3DMM 模型

3DMM[1]人脸参数化模型是 VETTER T 和 BLANTZ V 基于 100 位年轻男性及 100 位年轻女性的脸部激光扫描数据集而形成的一个数理统计模型。具体来说,首先假设对于每一个脸部激光扫描样本的顶点均是按一致顺序进行排列的,提取包含所有点位置信息的位置向量及颜色信息的颜色向量。之后,利用主成分分析(PCA)技术分别作用于位置向量及颜色向量,形成新的相互正交的位置特征基向量及颜色特征基向量。于是,一张新的人脸就可以由式(1)获得。

$$\begin{cases} S=\bar{S}+\sum_{i=1}^{199}\alpha_i s_i \\ T=\bar{T}+\sum_{i=1}^{199}\beta_i t_i \end{cases} \tag{1}$$

其中,$\bar{S},\bar{T}\in\mathbf{R}^{3N}$ 是 200 个样本对 $N=52\ 149$ 个顶点的位置及颜色求平均值;$s_i,t_i\in\mathbf{R}^{3N}$ 是 PCA 作用后形成的正交位置特征基向量及颜色特征基向量;α_i、β_i 则是位置特征基向量及颜色特征基向量的系数。一张新的人脸就可通过通过调节系数 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_{199})$,$\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_{199})$ 获得。

然而,系数并非任意数值皆可。VETTER T 和 BLANTZ V 发现,若令 α_i、β_i 服从标准正态分布,可使得生成二维人脸网格符合真实人脸形状。在此背景下,用 $p_{3DMM}(\boldsymbol{y})$,$\boldsymbol{y}=(\boldsymbol{\alpha},\boldsymbol{\beta})$ 来表示符合真实人脸 3DMM 参数分布,即标准正态分布。

Vetter 和 Blantz 采集的人脸数据库于 2009 年在 Basel Face Model[10]中公开,称为 BFM2009。本文所采用的 3DMM 模型便是 BFM2009。

1.1.3 可微分渲染器

可微分渲染器是将 3DMM 生成的三维人脸网格渲染成二维图片 $\hat{\boldsymbol{x}}=(\hat{x}_1,\hat{x}_2,\cdots,\hat{x}_n)$,$\hat{x}_i\in\mathbf{R}^{160\times160}$。由于神经网络的训练过程中涉及反向传播,故本文采纳 GENOVA K[5]等人的提议使用了一个基于延迟阴影模型的可微光栅化器。光栅化器在每个像素处生成包含三角形 id 和重心坐标的屏幕空间缓冲区。光栅化过后,使用重心坐标和 id 在像素处插值每个顶点的属性,如颜色和法线。光栅化导数是根据重心坐标计算的,而不是三角形 id。

因为可微分渲染器使用延迟着色,照明是用一组顶点属性插值后形成的缓冲区作为图像,独立计算每个像素,所以本文采用 Phong 反射光照模型进行阴影处理,Phong 反射光照模型比漫反射模型更有真实感,且既有效又可微。

1.1.4 身份编码器

身份编码器输入一共有两类,均为 160×160 的图片 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$,$\hat{\boldsymbol{x}}=(\hat{x}_1,\hat{x}_2,\cdots,\hat{x}_n)$,$x_i,\hat{x}_i\in\mathbf{R}^{160\times160}$,其中,$x_i$ 是将预处理过的 224×224 人脸图片

进一步裁剪后的图片,$\hat{x}_i$ 是经过可微分渲染器渲染后的人脸图片。对应于每类输入,身份编码器的输出也是两类,均为对人脸图片提取的 128-D 身份特征向量 $\boldsymbol{f}=(f_1,f_2,\cdots,f_n)$,$\hat{\boldsymbol{f}}=(\hat{f}_1,\hat{f}_2,\cdots,\hat{f}_n)$,$f_i,\hat{f}_i\in\mathbf{R}^{128}$,其中,$f_i$ 是对 x_i 提取的身份特征向量,$\hat{f}_i$ 而是对 $\hat{x}_i$ 提取的身份特征向量。

本文采用预训练好的 FaceNet[11]作为身份编码器来提取输入人脸及渲染人脸的身份特征向量。FaceNet 主要利用 DNN,从原始图片学习到欧氏距离空间的映射,故而图像在欧式空间里的距离与人脸相似度关联。本文通过最小化两者身份特征向量,以使得生成的三维人脸接近输入的二维人脸,详见 1.2.2 节。

1.2 损失函数

为了使得回归得到的三维人脸网格能够接近输入人脸图像,本文提出了三项损失函数提供监督信号:GANs 损失函数、id 损失函数和皮肤颜色损失函数。

1.2.1 GANs 损失函数

为了简化 GANs 的目标函数,将生成器的回归过程表示为 $G(\boldsymbol{x})$,$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$,代表生成器根据输入的人脸图片 $\boldsymbol{x}$ 回归得到的 3DMM 参数 $\hat{\boldsymbol{y}}$。而判别器的判别学习过程被表示为 $D(\boldsymbol{Y})$,$\boldsymbol{Y}=\{\boldsymbol{y},\hat{\boldsymbol{y}}\}=\{y_1,y_2,\cdots,y_n,\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n\}$,代表判别器给出的对于某样本 Y_i 的结果集 $\hat{\boldsymbol{y}}$ 的概率来源于真实人脸 3DMM 参数分布 $p_{3DMM}(\boldsymbol{y})$而不是生成器生成。GANs 的目标函数可以由式(2)表示:

$$\min_G \max_D \mathcal{L}_{GANs}(G,D)=$$
$$\mathrm{E}_{\boldsymbol{y}\sim p_{3DMM}(\boldsymbol{y})}[\log D(\boldsymbol{y})]+\mathrm{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log(1-D(G(\boldsymbol{x}))] \quad (2)$$

其中,$\boldsymbol{x}\sim p_{data}(\boldsymbol{x})$ 表示 $\boldsymbol{x}$ 取样于人脸图像数据分布,而 $\boldsymbol{y}\sim p_{data}(\boldsymbol{y})$ 表示 $\boldsymbol{y}$ 取样于真实人脸 3DMM 参数分布 $p_{3DMM}(\boldsymbol{y})$。

目标在于使得生成器回归得出的 3DMM 参数服从真实人脸 3DMM 参数分布 $p_{3DMM}(\boldsymbol{y})$,即强迫 $\hat{\boldsymbol{y}}\sim p_{3DMM}(\boldsymbol{y})$,该目标是由 GANs 的 min-max 对抗过程实现的。GANs 一共包含两个交叉训练阶段:第一阶段固定生成器 G,训练判别器 D,该阶段中 D 的目标是最大化 $\mathcal{L}_{GANs}(G,D)$,如此,D 可以在样本来源于真实人脸 3DMM 参数分布 $p_{3DMM}(\boldsymbol{y})$时给出一个较大概率,在样本来源于生成器 G 时给出一个较小的概率;第二阶段固定判别器 D,训练生成器 G,该阶段中 G 的目标是最小化 $\mathcal{L}_{GANs}(G,D)$,以使得当样本来源于生成器 G 时 D 能给出一个较大的概率。

GANs 的损失函数对三维人脸重建任务提供了三维监督信号。基于 3DMM 模型本身的假设,利用生成器及判别器的对抗生成过程,指引生成器回归的 3DMM 参数接近于真实人脸 3DMM 参数分布 $p_{3DMM}(\boldsymbol{y})$,进而生成的三维人脸网格中人脸的顶点分布能够接近于真实人脸顶点分布。

1.2.2 id 损失函数

在欧式空间中,无论表情、姿势或者照明条件如何,FaceNet 对于同一人的两张照片提取的身份特征向量之间的距离要比从两个不同的人的照片提取的身份特征向量之间的距离更加接近。采用式(3)来衡量两张人脸的相似度:

$$\mathcal{L}_{id}=1-\cos(f_i,\hat{f}_i)=1-\frac{f_i\cdot\hat{f}_i}{|f_i||\hat{f}_i|} \quad (3)$$

其中,$f_i,\hat{f}_i\in\mathbf{R}^{128}$ 分别表示 FaceNet 从输入人脸照片 x_i 及其对应的三维人脸网格渲染后的人脸照片 $\hat{x}_i$ 提取出来的身份特征向量。当 x_i 与 $\hat{x}_i$ 越相似,则 FaceNet 提取出来的两个身份特征向量 f_i 及 $\hat{f}_i$ 的 cosine 分数就越接近 1。而目标是最小化 f_i 及 $\hat{f}_i$ 之间的距离,故而将 id 损失函数设计为 $1-\cos(f_i,\hat{f}_i)$。

id 损失函数对三维人脸重建任务提供了二维监督信号。由于三维人脸数据集稀少的原因,评判三维人脸重建结果的好坏便可迁移到二维空间。首先将三维人脸重建网格渲染成二维人脸图片,再利用 FaceNet 对二维输入人脸图片及渲染后的人脸图片进行身份特征向量提取,通过最小化两个身份特征向量之间的距离,使得渲染后的人脸身份特征接近输入图片的人脸特征,进而迫使重建三维人脸网格具有输入图片的人脸特征。

1.2.3 皮肤颜色损失函数

由于 FaceNet 可以忽略照明等因素对人脸特征进行身份特征提取,仅依靠 FaceNet 提供的二维监督信号导致重建的三维人脸皮肤不能很好地反映输入图片人脸皮肤颜色。故而采用式(4)来进一步衡量两张人脸的皮肤颜色损失。

$$\mathcal{L}_{tex}=||\psi(\hat{x}_i)-x_i|| \quad (4)$$

其中,$\psi(\hat{x}_i)$表示将三维人脸网格渲染后的二维照片 $\hat{x}_i$ 按照 68 个人脸关键点进行人脸提取并嵌入至输入图片 x_i 中。

由于嵌入后背景像素一致,通过对两张图片逐像素求 l_2 损失,可以使得渲染后的人脸皮肤颜色逼近输入人脸皮肤颜色。因此皮肤颜色损失函数对三维人脸重建任务提供了二维监督信号。

1.2.4 总体损失函数

综上,总体损失函数可由式(5)表示:

$$\mathcal{L}=\mathcal{L}_{\text{GANs}}(G,D)+\omega_{\text{id}}\mathcal{L}_{\text{id}}+\omega_{\text{tex}}\mathcal{L}_{\text{tex}} \tag{5}$$

其中,$\omega_{\text{id}}=5$,$\omega_{\text{tex}}=0.15$ 是训练时控制 id 损失和皮肤颜色损失比例的超参数。$\mathcal{L}_{\text{GANs}}(G,D)$为三维人脸重建任务提供了三维监督信号,使得生成器回归得到的 3DMM 参数符合 3DMM 模型假设的真实人脸 3DMM 参数分布 $p_{\text{3DMM}}(\boldsymbol{y})$,进而迫使重建的三维人脸网格顶点分布接近于真实人脸顶点分布。而 $\mathcal{L}_{\text{id}}$,$\mathcal{L}_{\text{tex}}$ 为三维人脸重建任务提供了二维监督信号,$\mathcal{L}_{\text{id}}$ 将三维人脸重建结果的评估移到二维空间,其中使得渲染后的二维人脸接近输入图片的人脸,进而迫使重建三维人脸网格具有输入图片的人脸特征,$\mathcal{L}_{\text{tex}}$ 使得渲染后的皮肤颜色更加接近输入图片的人脸皮肤颜色。通过这三项损失函数,可以使得重构后的人脸既符合真实三维人脸网格顶点的分布,又具备输入图片人脸的身份特征及皮肤颜色,故而使得人脸重建结果具有说服力。

2 实验结果及分析

2.1 训练细节

本文实验是在 VGGFace2[12]数据集上进行训练,该数据集包含了 9 131 个人物的 3.31M 张图片,图片来源于不同的年龄、种族的人物及各个人物的不同姿势。

由于冗余的背景信息往往对人脸重构任务无用且降低网络收敛速度,因而使用 MTCNN[13]对每张图片提取 224×224 像素的人脸。

在训练过程中使用了 Adam 优化算法来有效地更新网络权重,学习率设置为 0.001,batch size 设置为 5,整个网络共训练了 500k 次 iteration。

2.2 重建效果比较

因为三维人脸重建的重建效果判别较为主观,故而主要将本文重建的三维人脸网格与文献[5]~[7]重建结果进行比较。比较结果如图4 所示,其中,输入图片来自文献[6]提供的 MoFA 测试图像数据集,该数据集共包含 84 位实验对象。

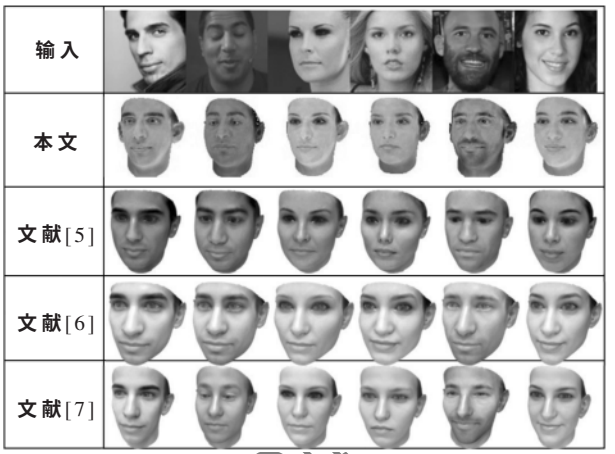

图 4 重建效果比较

从视觉效果上来说,相对于文献[6]、文献[7],本文重建结果在皮肤及人脸形状真实感上更有说服力;相对于文献[5],本文重建结果更能反映输入人脸的身份特征。这说明,从渲染得到的二维图片来说,本文重建结果渲染后更加接近真实人脸效果并保持输入人脸的身份特征。

2.3 重建结果准确性比较

为了衡量重建的三维人脸网格的准确性,在 MICC Florence[14]人脸数据库中与文献[5]、文献[7]方法进行比较。

MICC Florence 人脸数据库一共包含 53 位白种人的无表情人脸扫描,同时,数据集还提供了每位实验对象的分别在三种条件下的三个视频:采集人脸扫描环境、室内环境、室外环境,可以看出,环境复杂度在依次增加。其中,文献[7]对于视屏中每帧图像的重建三维网格结果进行逐顶点求平均值,文献[5]及本文对于视屏中每帧图像回归的 3DMM 参数求平均值。

由于 MICC Florence 人脸数据库给出的人脸扫描顶点数与 3DMM 模型的顶点数不对应且数目不一致,故首先将回归得到的三维人脸网格及人脸数据库提供的人脸扫描均裁剪到以鼻头为中心,95 mm 为半径的范围内,然后采用了各项同性的 ICP 算法来将两个三维点云进行对齐。实验结果将 53 位实验对象的 point-to-plane 的距离求均方差。最终的准确性比较结果如表 1 所示。可以看出,本文的重建方法在三种环境下均优于文献[5]和文献[7]方法。这

表 1 重建结果准确性比较比较

(mm)

方法	扫描环境	室内环境	室外环境
文献[7]	1.93±0.27	2.02±0.25	1.86±0.23
文献[5]	1.50±0.13	1.50±0.11	1.48±0.11
本文	1.48±0.17	1.48±0.14	1.47±0.12

得益于三维监督信号使得三维人脸网格更具有真实人脸顶点分布。

3 结论

本文基于 GANs 提出了一种无监督回归输入图片对应的三维人脸网格的方法。具体来说,提出了三项并行的损失函数,GANs 损失函数使得回归的 3DMM 参数符合真实人脸参数分布,重构的三维人脸网格顶点符合真实人脸顶点分布;id 损失函数使得重构的三维人脸具有输入人脸的身份特征;皮肤颜色损失使得重构的三维人脸更能反映输入人脸的皮肤颜色。在 MoFA 数据集及 MICC Florence 数据集上的实验效果表明,重构结果不仅保持了输入人脸的身份特征,也在顶点位置上具有更小的误差。

参考文献

[1] BLANZ V, VETTER T, ROCKWOOD A. A morphable model for the synthesis of 3D faces[J]. ACM Siggraph, 2002: 187-194.

[2] BLANZ V. Face recognition based on a 3D morphable model[C]. International Conference on Automatic Face & Gesture Recognition. IEEE, 2006.

[3] RICHARDSON E, SELA M, KIMMEL R. 3D face reconstruction by learning from synthetic data[C]. 2016 Fourth International Conference on 3D Vision(3DV), USA, 2016.

[4] RICHARDSON E, SELA M, OR-EL R, et al. Learning detailed face reconstruction from a single image[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), USA, 2017.

[5] GENOVA K, COLE F, MASCHINOT A, et al. Unsupervised training for 3d morphable model regression[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2018.

[6] TEWARI A, ZOLLHFER M, KIM H, et al. MoFA: model-based deep convolutional face autoencoder for unsupervised monocular reconstruction[C]. 2017 IEEE International Conference on Computer Vision(ICCV). IEEE, 2017.

[7] TRAN A T, HASSNER T, MASI I, et al. Regressing robust and discriminative 3D morphable models with a very deep neural network[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2017.

[8] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Advances in Neural Information Processing Systems, 2014(3): 2672-2680.

[9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), USA, 2016.

[10] PAYSAN P, KNOTHE R, AMBERG B, et al. A 3D face model for pose and illumination invariant face recognition[C]. 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, Italy, 2009.

[11] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), USA, 2015.

[12] CAO Q, SHEN L, XIE W, et al. Vggface2: a dataset for recognising faces across pose and age[C]. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition(FG 2018), China, 2018.

[13] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[14] BAGDANOV A D, DEL BIMBO A, MASI I. The florence 2D/3D hybrid face dataset[J]. Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, 2011(11): 79-80.

(收稿日期:2020-08-05)

作者简介:

张星星(1994-),通信作者,女,硕士研究生,主要研究方向:三维人脸重建、对抗生成网络。E-mail:345651730@qq.com。

李金龙(1975-),男,博士,副教授,主要研究方向:深度神经网络、自然语言处理。

版权声明