# 面向单幅图像的逼真 3D 人脸重建方法

包永堂[1)], 周鹏飞[1)], 齐越[2,3,4)*]

[1)] (山东科技大学计算机科学与工程学院 青岛 266590)
[2)] (北京航空航天大学虚拟现实技术与系统全国重点实验室 北京 100191)
[3)] (北京航空航天大学青岛研究院 青岛 266100)
[4)] (鹏程实验室 深圳 518055)
(qy@buaa.edu.cn)

**摘 要:** 针对 3DMM 参数拟合方法生成的纹理过于粗糙、结果不够逼真的问题, 提出一种基于深度学习的单幅图像逼真 3D 人脸重建方法. 首先构建 RP-Net 回归网络和包含 5 万幅人脸图像的数据集, 从输入图像中学习参数, 并拟合人脸模型生成 3D 人脸几何; 然后通过构造多层次的损失函数进行弱监督学习, 包括低水平的像素损失、地标损失和高水平的身份损失; 最后通过纹理映射的方式生成逼真的人脸纹理. 在 2 个通用人脸数据集和 1 个人工生成的人脸数据集上与最近的 3D 人脸重建方法进行对比实验, 并对影响重建的光照、表情和转向等因素进行实验, 根据 SSIM 和 PSNR 对 3D 重建结果进行量化分析. 实验结果表明, 所提方法面向单幅图像可以生成准确的 3D 人脸形状和逼真的人脸纹理; 与最近的 3D 人脸重建方法相比, 该方法的训练时间和迭代次数分别降低了 6%和 13%, SSIM 值增加 0.005~0.010, PSNR 值平均提高 0.03~0.08 dB.

**关键词:** 3D 人脸重建; 人脸对齐; 3D 形变模型; 纹理映射; 单幅图像
**中图法分类号:** TP391.41 **DOI:** 10.3724/SP.J.1089.2022.19485

# Realistic 3D Face Reconstruction Method for Single Image

Bao Yongtang[1)], Zhou Pengfei[1)], and Qi Yue[2,3,4)*]

[1)] (*College of Compute Science and Engineering, Shandong University of Science and Technology, Qingdao* 266590)
[2)] (*State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing* 100191)
[3)] (*Qingdao Research Institute of Beihang University, Qingdao* 266100)
[4)] (*Pengcheng Laboratory, Shenzhen* 518055)

**Abstract:** To solve the problem that the 3DMM parameter fitting methods cannot generate realistic 3D face, a single-image realistic 3D face reconstruction method based on deep learning is proposed. Firstly, the RP-Net regression network is constructed, and a dataset containing 50 000 face images is constructed at the same time. The parameters are learned from the input images, and the face model is fitted to generate the 3D face geometry. Secondly, weakly supervised learning is performed by constructing a multi-level loss function, which includes low-level pixel loss, landmark loss, and high-level identity loss. Thirdly, a realistic face texture is generated by means of texture mapping. Finally, two real face data and one generated data are used to compare experiments with recent 3D face reconstruction methods. The factors affecting the reconstruction such as lighting, expression, and steering are used to test proposed method, and quantitatively evaluate the

收稿日期: 2021-11-23; 修回日期: 2021-12-03. **基金项目**: 山东省自然科学基金(ZR2020MF132); 国家自然科学基金(62072020); 国家重点研发计划(2017YFB1002602); 青岛市创新创业领军人才项目(19-3-2-21-zhc); 广东省重点研发计划(2019B010150001). **包永堂**(1983—), 男, 博士, 副教授, 硕士生导师, CCF 高级会员, 主要研究方向为人工智能、计算机图形学; **周鹏飞**(1996—), 男, 硕士研究生, CCF 学生会员, 主要研究方向为 3D 人脸重建; **齐越**(1969—), 男, 博士, 教授, 博士生导师, 论文通信作者, 主要研究方向为虚拟现实、计算机视觉.

reconstruction by SSIM and PSNR. These results show that proposed method can generate accurate face shapes and realistic face textures. Compared with the recent 3D face reconstruction method, the training time and number of iterations of the proposed method are reduced by 6% and 13%, respectively, the SSIM value is increased by 0.005–0.010, and the PSNR value is increased by 0.03–0.08 dB on average.

**Key words:** 3D face reconstruction; face alignment; 3D morphable model; texture mapping; single image

3D 人脸重建是计算机视觉和计算机图形学领域重要的研究内容之一, 在人脸识别、人脸动画和辅助医疗等领域均有广泛的应用. 近年来, 基于单幅图像的 3D 人脸重建越来越受到研究者的重视, 其中, 3D 形变模型(3D morphable model, 3DMM)的应用最为广泛. 随着深度学习技术的兴起, 通过网络估计人脸图像对应的 3DMM 参数越来越受到关注. 基于 3DMM 的方法可以实现人脸的稠密对齐, 降低人脸重建的难度; 但是, 基于参数拟合得到的模型不能体现人脸的高频信息, 重建后的 3D 人脸不能表达出褶皱、酒窝等高频细节信息. 3DMM 方法采用主成分分析(principal component analysis, PCA)的方式生成纹理, 因其精度不够准确, 生成的 3D 人脸模型缺乏逼真性; 此外, 由于 3D 人脸数据集相对较少, 3D 数据不易在训练网络中直接表示, 难以获得真实的 3D 地标用于训练网络, 因此, 无法直接通过深度学习的方式从 3D 人脸数据集中生成逼真的人脸模型.

为解决上述问题, 研究人员提出了一些弱监督学习的方法[1-4]. 在训练网络中, 弱监督学习的方法一般通过构造损失函数来提高人脸重建的精度. 为生成高质量的纹理, Deng 等[1]提出专注肤色的人脸重建方法, 可以恢复人脸的肤色, 但是生成的纹理无法表达褶皱等细节信息; Lin 等[2]使用图卷积网络(graph convolutional network, GCN)对 3DMM 生成的粗糙纹理进行纹理细化, 细化后的纹理能够实现高保真. 当前, 通过生成对抗网络(generative adversarial networks, GANs)生成高质量的 UV 贴图, 并用于 3D 人脸重建的工作开始引起研究者的关注. Deng 等[3]提出的 UV-GAN 框架可以获得大姿态下完整的 UV 贴图, 能够在 UV 空间中训练生成完整的 UV 贴图. Gecer 等[4]利用 GANs 在 UV 空间中训练人脸纹理生成器, 使用纹理生成器生成人脸纹理, 但通过该方法生成的纹理在人脸肤色上与输入图像有较大差异; 此外, 该方法需要大量的 UV 图像来训练 GANs, 时间消耗非常大.

本文提出一种基于深度学习的人脸重建的方法, 从单幅图像生成逼真的 3D 人脸. 基于残差卷积神经网络(convolutional neural network, CNN)构建参数回归网络(regression parameter network, RP-Net)用于回归人脸参数, 采用 3DMM 拟合人脸形状和表情参数生成 3D 人脸. 由于通过 3DMM 生成的纹理过于粗糙, 不能体现人脸的高频细节信息, 本文使用纹理映射的方式直接生成逼真人脸纹理. 本文方法不需要在网络中优化训练生成纹理, 减少网络计算的开销. 此外, 由于本文构建的 RP-Net 是基于 ResNet-50 网络提取的特征来回归不同的参数, 不需要使用多种网络获得参数, 因此能够降低训练的时间和空间复杂度.

## 1　相关工作

Blanz 等[5]最先提出 3DMM, 此后基于 3DMM 的变形算法不断被提出, 这些方法均采用 PCA 的方式对面部进行扫描, 得到低维的人脸形状和纹理. 虽然这类方法生成人脸的基本特征效果较好, 但是 PCA 方法无法获得人脸的高频信息, 拟合的人脸不能表现纹理的细节, 无法保证重建人脸的质量.

随着深度学习技术的发展, CNN 被广泛应用于 3D 人脸重建. CNN 从图像中学习生成人脸的 3DMM 参数, 并拟合 3DMM 生成人脸, 这种回归方法当前得到快速发展. Genova 等[6]采用无监督学习的方式, 通过改进的 FaceNet[7]回归 3DMM 参数等, 提高重建人脸的质量. Deng 等[1]采用微调的 ResNet-50 回归人脸的参数, 并通过多视图聚合的方式实现信息互补, 提高人脸重建的准确性. 可微分渲染[1-4,6]在 3D 人脸重建的方法中被广泛使用, 其将 3D 人脸几何、人脸纹理、光照参数和位姿参数共同渲染到二维图像上, 通过渲染图像和输入图像计算损失, 并通过弱监督学习的方式训练网络模型. 由于在训练网络模型时不需要使用 3D 人脸的地标进行强监督学习, 因此训练的成本大大减小. 此外, Wu 等[8]通过自动编码器-解码器网络

结构预测人脸的深度映射和人脸的纹理反照率等信息，能够重建出高质量的 3D 人脸. 目前，GCN 在人脸重建中发展迅速，Lee 等[9]提出使用 GCN 生成准确的人脸几何；Lin 等[2]将 GCN 作为纹理的细化器，将粗糙的纹理输入到 GCN 中，结合人脸的纹理特征生成逼真的纹理；Chen 等[10]采用 UNet 生成人脸的深度图和颜色距离图，其中颜色距离图用来表示当前顶点与模板人脸的距离，该方法通过局部信息迭代生成详细的人脸几何信息；Liu 等[11]也提出一个编码器-解码器组合的网络结构，在 3D 人脸重建中对人脸特征进行解耦合，完成精准的人脸形状重建和判别人脸形状的任务.

目前，多特征融合的方法被引入到 3D 人脸重建中. Dou 等[12]提出端到端的深度学习方法，使用多任务损失函数和融合 CNN 改进了面部表情的重建，提高人脸重建的细节和准确度；Sanyal 等[13]联合 RingNet 和耦合字典模型进行 3D 人脸的重建，能够增强脸部细节；Zhu 等[14]将 CNN 用于大姿态下人脸对齐，提高了人脸重建的精度；Richardson 等[15]使用 2 个 CNN 从粗到细地重建人脸，生成更准确的人脸几何；Zhao 等[16]在遮挡场景下生成人脸的解析图来构建 3D 纹理，生成了更具真实性的人脸纹理.

## 2 本文方法

为了从无约束的二维图像中重建逼真的 3D 人脸，本文构建使用 CNN 的 RP-Net，该网络基于 3DMM，从输入图像中回归 3DMM 参数等，如图 1 所示. 其中，网络骨干采用 ResNet-50，输出层使用残差块生成 512 维向量，通过全连接层输出 3DMM 参数等. 由于 3DMM 纹理过于粗糙，无法实现纹理的高保真，本文使用纹理映射的方式生成纹理来解决此问题. 为了提高人脸重建的质量，利用多层次的损失函数通过弱监督学习的方式训练网络. 多层次的损失函数包括低层次的像素损失、地标损失和高层次的身份损失.

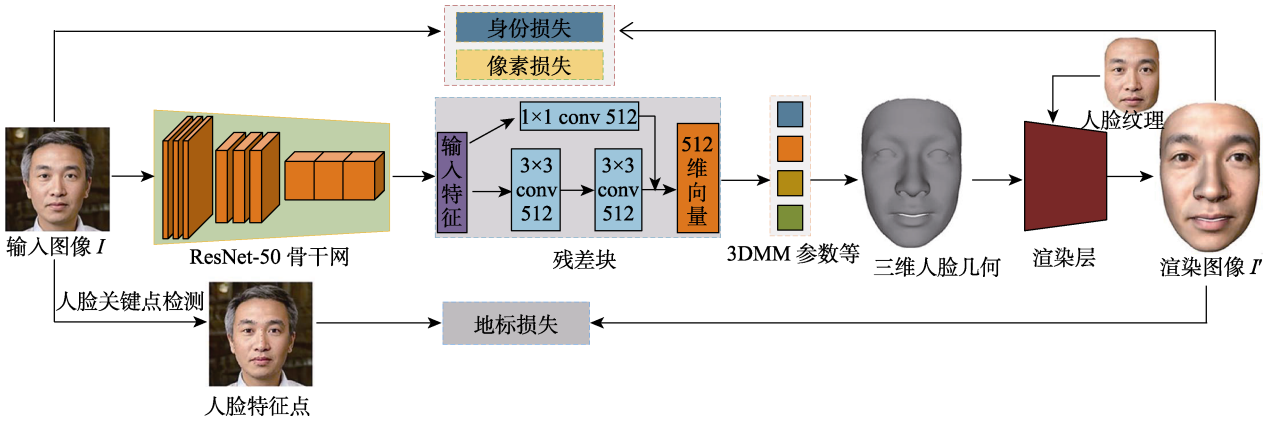



图 1 训练流程图

### 2.1 模型参数

#### 2.1.1 3DMM

本文采用 3DMM 恢复 3D 人脸，使用 BFM (basel face model)数据集[17]生成 3D 人脸的形状，利用 FaceWarehouse 数据集[18]生成 3D 人脸的表情. 3D 人脸模型线性表示为

$$S=\overline{S}+A_{\mathrm{id}}\boldsymbol{\alpha}+A_{\mathrm{exp}}\boldsymbol{\beta} \tag{1}$$

其中，$\overline{S}$ 表示 3D 人脸的平均形状；$A_{\mathrm{id}}$ 和 $A_{\mathrm{exp}}$ 分别表示形状和表情的主成分基；$\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 分别表示形状和表情的参数，用于拟合生成 3D 人脸. 通过 CNN 回归 3DMM 参数向量，其中，人脸形状参数 $\boldsymbol{\alpha}\in\mathbb{R}^{80}$，人脸表情参数 $\boldsymbol{\beta}\in\mathbb{R}^{64}$. 本文对 BFM 进行调整，调整后得到的 3DMM 不包括颈部和耳朵，只保留人脸区域.

#### 2.1.2 相机模型

为了计算地标损失，在生成 3D 人脸后，本文采用透视投影将 3D 人脸投影到原视角的图像中，其过程可以表示为

$$M=f*\boldsymbol{P}*\boldsymbol{R}*\boldsymbol{V}+\boldsymbol{T} \tag{2}$$

其中，$f$ 表示缩放因子；$\boldsymbol{P}$ 表示投影矩阵，采用正交矩阵；$\boldsymbol{R}$ 表示旋转矩阵；$\boldsymbol{T}$ 表示平移向量；$\boldsymbol{V}$ 表示 3D 人脸的顶点向量. 本文通过 CNN 回归相机参数和位姿参数，位姿参数 $\boldsymbol{\rho}\in\mathbb{R}^3$.

综上所述，为了从图像中预测 3DMM 的参数等，本文构建了 RP-Net 回归 3DMM 参数、位姿参数和相机参数，其中，3DMM 参数是 239 维的向量，

位姿参数和相机参数是 18 维的向量.

### 2.2　纹理映射

基于 3DMM 生成的纹理过分依赖于先验知识, 不能捕获人脸的高频细节信息. 因此, 本文采用纹理映射的方式生成纹理, 纹理映射是将 3D 顶点与二维图像的像素点建立起映射关系, 将二维图像上的纹理值赋给对应 3D 人脸顶点的过程. 图 2 所示为采用纹理映射方法与直接采用 3DMM 生成纹理的对比, 图 2a 所示为 CelebA 人脸数据集[19]中的 4 幅图像. 由图 2b 和图 2c 可以看出, 本文方法生成的结果在光照、高频几何细节信息等方面更加逼近于输入图像, 特别是在皱纹和胡须等纹理方面, 比 3DMM 方法生成的纹理更加逼近于原始图像. 3D 人脸模型到二维人脸图像的投影映射采用改进的黄金标准算法, 该算法确定仿射摄像机投影矩阵的最大似然估计, 本文改进后的黄金标准算法步骤如下.

输入. 3D 人脸 68 个特征点 $\boldsymbol{X}_i$ 和二维图像的 68 个特征点 $\boldsymbol{x}_i$.

输出. 仿射摄像机的投影变换矩阵 $\boldsymbol{P}$.

Step1. 用第 1 个相似变换 $\boldsymbol{T}$ 归一化图像点 $\boldsymbol{x}_i$, 假设归一化后得到的图像点是 $\tilde{\boldsymbol{x}}_i = \boldsymbol{T}\boldsymbol{x}_i$.

Step2. 用第 2 个相似变换 $\boldsymbol{U}$ 归一化 3D 点 $\boldsymbol{X}_i$, 归一化后的 3D 点是 $\tilde{\boldsymbol{X}}_i = \boldsymbol{U}\boldsymbol{X}_i$, 其最后一个元素是 1.

Step3. 每个特征点执行变换 $\tilde{\boldsymbol{X}}_i \leftrightarrow \tilde{\boldsymbol{x}}_i$, 产生方程

$$\begin{bmatrix} \tilde{\boldsymbol{X}}_i^{\mathrm{T}} & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0}^{\mathrm{T}} & \tilde{\boldsymbol{X}}_i^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{y}}_i \end{bmatrix};$$

然后将其垒成一个 $2n\times 8$ 的矩阵方程 $\boldsymbol{A}_8\boldsymbol{p}_8 = \boldsymbol{b}$, 其中, $\boldsymbol{p}_8$ 表示估计投影矩阵 $\tilde{\boldsymbol{P}}_A$ 前 2 行的 8 维矢量; $\boldsymbol{b}$ 表示 $\tilde{\boldsymbol{x}}$ 的一维矢量; $\boldsymbol{A}_8$ 表示 $\tilde{\boldsymbol{X}}^{\mathrm{T}}$ 的对称矩阵.

Step4. 通过 $\boldsymbol{A}_8$ 的伪逆进行求解: $\boldsymbol{p}_8 = \boldsymbol{A}_8^{-1}\times\boldsymbol{b}$, 并且仿射约束 $\tilde{\boldsymbol{p}}^{3\mathrm{T}} = (0,0,0,1)$.

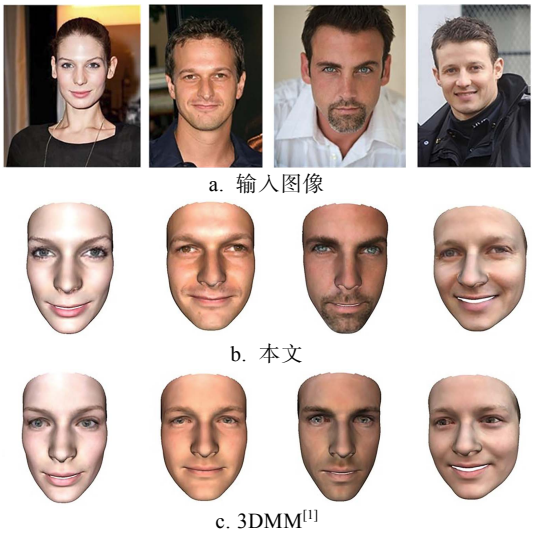
Step5. 去除归一化, 求出投影矩阵 $\boldsymbol{P}$, 即 $\boldsymbol{P} = \boldsymbol{T}^{-1}\tilde{\boldsymbol{P}}_A\boldsymbol{U}$.

Step6. 算法结束.

通过上述黄金标准算法可以获得 3D 顶点到二维图像的映射关系, 从而通过映射关系生成 3D 人脸纹理. 实验结果表明, 通过纹理映射可以生成逼真的人脸纹理, 能够保持人脸纹理的高频细节信息.

### 2.3　可微渲染层

为了建立 3D 人脸和二维渲染图像之间的弱监督学习信息, 本文对 Genova 等[6]提出的可微渲染层进行改进, 实现 3D 人脸渲染生成二维图像, 并采用弱监督学习的方式训练网络模型. 可微渲染层采用基于延迟着色的光栅化器, 光栅化器使用



a. 输入图像

b. 本文

c. 3DMM[1]

图 2　2 种方法生成纹理对比

三角形面片的序号与像素的重心坐标计算屏幕空间缓冲区. 网格的颜色和法线在像素处通过插值得到. 可微渲染层将 3D 人脸渲染到二维的图像上, 采用网络预测的参数计算损失函数, 通过反向传播来提高重建的质量. 本文通过渲染层获得渲染图像, 通过计算损失进行弱监督学习, 同时训练网络模型, 以提高网络的准确性. 可微渲染层使用透视投影和球谐函数(spherical harmonics, SH)的光照模型, 提高渲染图像的质量. 通过可微渲染层, 将 3D 人脸网格渲染到二维图像上训练网络模型, 以提高人脸的生成能力.

### 2.4　损失函数

在训练过程中, 由于在人工构造的数据集中没有 3D 人脸的真实地标信息, 无法进行强监督学习, 因此, 本文构造损失函数进行弱监督学习, 以提高人脸重建的质量. 为了对人脸进行有效地重建, 本文采用多层次的损失函数约束 3D 人脸重建的几何形状. 损失函数包括像素损失 $L_{\mathrm{pixel}}$、地标损失 $L_{\mathrm{lm}}$、身份损失 $L_{\mathrm{id}}$ 和 3DMM 参数的正则化 $L_{\mathrm{reg}}$. 损失函数定义为

$$L = \lambda_{\mathrm{pixel}}L_{\mathrm{pixel}} + \lambda_{\mathrm{lm}}L_{\mathrm{lm}} + \lambda_{\mathrm{id}}L_{\mathrm{id}} + \lambda_{\mathrm{reg}}L_{\mathrm{reg}} \tag{3}$$

其中, $\lambda_{\mathrm{pixel}}, \lambda_{\mathrm{lm}}, \lambda_{\mathrm{id}}, \lambda_{\mathrm{reg}}$ 分别表示控制各自损失的权值.

#### 2.4.1　$L_{\mathrm{pixel}}$

为了使重建的 3D 人脸模型更加逼近于真实的人脸, 本文采用重建渲染图像与输入图像之间的像素差异计算损失. 定义像素损失函数为

$$L_{\mathrm{pixel}} = \frac{\sum_{i\in M}\left\|\boldsymbol{I}_i - \boldsymbol{I}_i'\right\|_2}{\left|M_{2\mathrm{d}}\right|} \tag{4}$$

其中，$i$ 表示像素点，$M$ 表示人脸的投影区域，$M_{2\mathrm{d}}$ 表示二维图像的人脸皮肤区域．$\boldsymbol{I}$ 表示输入图像，$\boldsymbol{I}'$ 表示重建的渲染图像，$\|\cdot\|_2$ 表示 $L_2$ 范数．

2.4.2 $L_{\mathrm{lm}}$

为了提高人脸对齐的准确性，本文使用地标损失进行弱监督学习．使用人脸检测算法[20]检测训练图像的 68 个特征点 $\{q_n\}$，将重建的 3D 人脸投影到二维图像上，得到 68 个特征点 $\{q'_n\}$．为降低二维人脸和重建人脸的距离，提高人脸对齐的准确性，定义地标损失函数为

$$L_{\mathrm{lm}}=\frac{\sum_{i=1}^{N}\omega_i\left\|q_i-q'_i\right\|^2}{N} \tag{5}$$

其中，$\omega_i$ 表示地标的权重，设在内嘴边地标点的权重 $\omega_i=20$，其他区域权重 $\omega_i=1$；取 $N=68$．

2.4.3 $L_{\mathrm{id}}$

基于像素的损失有时会导致局部最小值问题，本文从高维的角度使用人脸识别网络，提取人脸图像的深层特征．由于 ArcFace 网络[21]在人脸数据集上能够获得更高的人脸识别准确率，本文采用 ArcFace 网络作为人脸深度特征提取器．在 VGGFace2 数据集[22]上训练 ArcFace 网络，然后将训练好的网络用于提取输入图像和渲染图像的 512 维人脸的深度特征，最后计算深度特征，对应的余弦距离为

$$L_{\mathrm{id}}=1-\frac{\left\langle F(\boldsymbol{I}),F(\boldsymbol{I}')\right\rangle}{\|F(\boldsymbol{I})\|\cdot\|F(\boldsymbol{I}')\|} \tag{6}$$

其中，$\boldsymbol{I}$ 表示输入图像，$\boldsymbol{I}'$ 表示重建渲染图像．$F(\cdot)$ 表示人脸的深度特征编码，$\langle\cdot,\cdot\rangle$ 表示向量内积，$\|\cdot\|$ 表示 $L_1$ 正则化．

2.4.4 $L_{\mathrm{reg}}$

为了防止 3D 人脸形状退化，本文将针对回归网络得到的人脸形状参数和表情参数添加受约束的正则化．实验结果表明，正则化可以有效地防止人脸形状参数和表情参数退化．定义正则化损失为

$$L_{\mathrm{reg}}=\omega_{\boldsymbol{\alpha}}\|\boldsymbol{\alpha}\|^2+\omega_{\boldsymbol{\beta}}\|\boldsymbol{\beta}\|^2 \tag{7}$$

其中，$\boldsymbol{\alpha}$ 表示形状参数；$\boldsymbol{\beta}$ 表示表情参数；在训练时将其对应权重分别设置为 $\omega_{\boldsymbol{\alpha}}$=1.0，$\omega_{\boldsymbol{\beta}}$=0.8．

## 3 实验及结果分析

### 3.1 实现细节

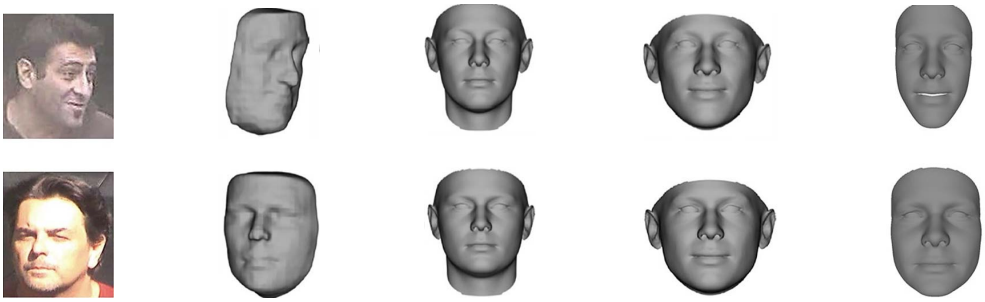
本文使用 TensorFlow 深度学习框架设计 RP-Net，该网络基于 ResNet-50 对输出层进行改进：首先将 ResNet-50 输出的特征向量输入到残差块中生成 512 维的向量，残差块是将 ResNet-50 输出的特征向量进行 3×3 的卷积和 ReLU 激活函数；然后做相同的 3×3 卷积和 ReLU 激活函数，将 ResNet-50 输出的特征向量进行 1×1 的卷积，并将通道数降至 512 维；再将上述 2 个 512 维的向量相加，构成残差块；最后全连接输出 257 维参数向量．实验结果表明，残差块可以提高训练收敛的速度，减少训练时间．采用基于延迟着色的渲染函数对重建的 3D 人脸进行渲染，用渲染图像和输入图像计算损失，提高人脸重建的精度．

本文构造了一个人脸数据集，该数据集是由从网站上收集的高分辨的人脸图像组成，包含 5 万幅图像，图像分辨率大小为 1 024×1 024；包括亚洲和欧洲人种，成人和小孩的人脸图像．本文对该数据集进行人脸特征点标注，采用 MTCNN 人脸检测算法[20]对输入图像检测 5 个特征点，并标记这 5 个特征点的位置．在图像预处理阶段，本文根据标记的 5 个特征点对图像进行裁剪，裁剪后的图像分辨率为 224×224 像素．根据 MTCNN 人脸检测算法计算输入图像的 68 个特征点，用于计算地标损失．采用 Adam 优化方法对 RP-Net 进行训练，训练学习率设为 $10^{-4}$．设置训练的 batch size 为 16，共训练 13 万次．RP-Net 初始参数设置为 ResNet-50 在 ImageNet 上训练好的参数，以提高训练的速度和精度．实验中，本文设置损失函数的训练权重分别为 $\lambda_{\mathrm{pixel}}=1.92$，$\lambda_{\mathrm{lm}}=1.6\times10^{-3}$，$\lambda_{\mathrm{id}}=0.2$ 和 $\lambda_{\mathrm{reg}}=3\times10^{-3}$．

### 3.2 重建质量

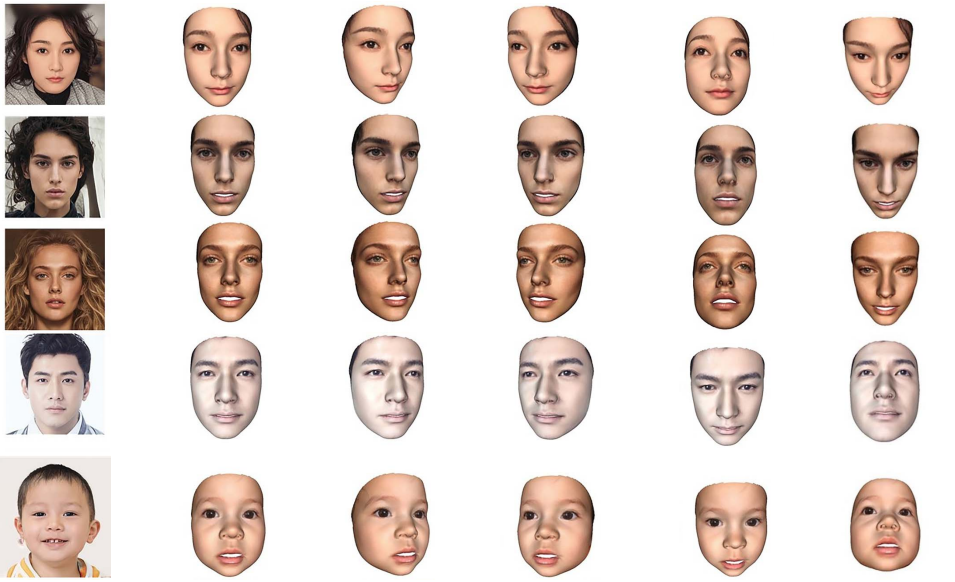
为了定性评价重建质量，将本文方法结果与最新方法生成的结果进行对比，结果如图 3 所示．图 3a 所示为 MICC 数据集[23]中的 2 幅图像；从图 3b~图 3e 可以看出，图 3b 生成的结果表面网格较粗糙，无法重建出完整的几何形状；图 3d 虽然能够重建出完整的几何形状，但是人脸几何身份和输入图像差距较大；图 3e 重建的结果在人脸几何身份上与输入图像更加逼近．

与已有方法[24-27]不同，为了进一步测试本文方法重建 3D 人脸的能力，选择不同年龄和肤色的人脸图像进行重建测试，结果如图 4 所示．图 4a 所示为本文构造的人脸数据集中的 5 幅图像；从图 4b~图 4f 可以看出，图 4b 在年龄和肤色上均与输入图像非常接近；另外，从图 4c~图 4f 可以看出，重建结果的身份特征信息均未发生变化，具有较好的鲁棒性．



a. 输入图像　b. 文献[28]　c. 文献[14]　d. 文献[29]　e. 本文

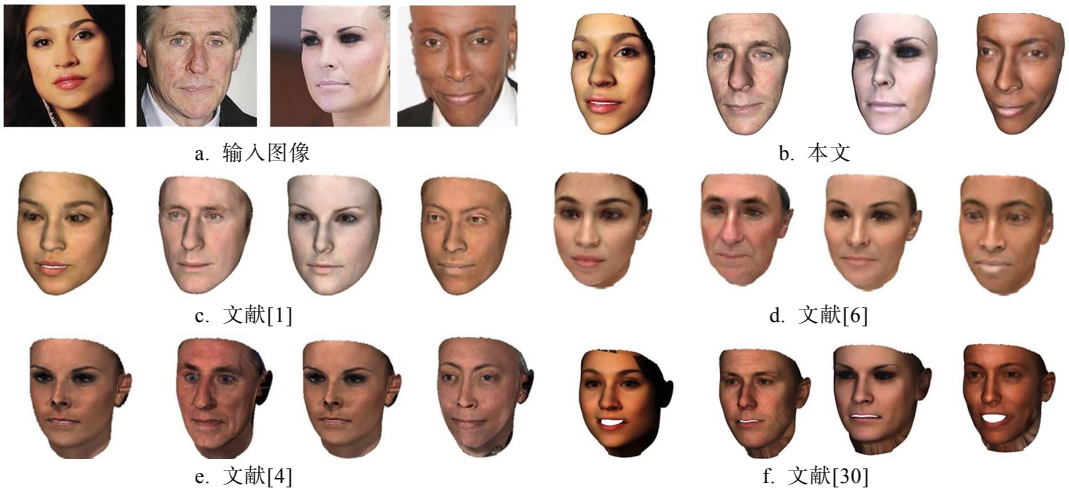
图 3　4 种方法生成的 3D 几何对比



a. 输入图像　b. 正面重建结果　c. 左转向 15°　d. 右转向 15°　e. 上转向 15°　f. 下转向 15°

图 4　5 种转向生成结果对比

在 MoFA-test 数据集上选择合适的人脸图像进行重建，将本文方法与当前最先进方法的重建结果进行对比，结果如图 5 所示. 尽管图 5c~图 5f 均采用可微渲染层，但只有图 5b 可以重建出逼真的人脸，其他方法重建出的人脸均不够逼真. 此外，本文方法只需要输入二维的人脸图像，不需要进行复杂的图像处理；而文献[30]需要从 122 个对象中捕获 366 幅高清扫描图像，文献[4]采用 1 000 幅未包裹的纹理图像在 UV 空间上采用渐进 GAN 训练一个纹理生成器，Deng 等[3]对大姿态不完整的



a. 输入图像　b. 本文

c. 文献[1]　d. 文献[6]

e. 文献[4]　f. 文献[30]

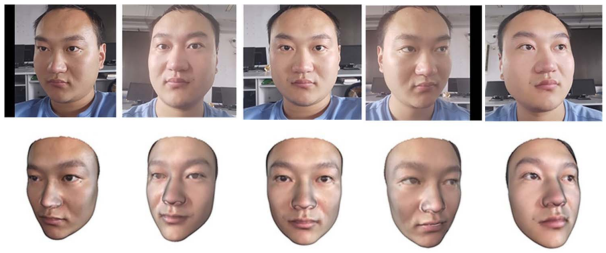
图 5　5 种方法生成结果对比

UV 贴图采用 GANs 训练生成完整的 UV 图. 从本文方法与这些方法的重建结果可以看出, 这些方法无法保留人脸的高频细节信息, 而本文方法能够突出人脸的高频细节. 从图 5a 第 2 幅图像可以看出, 本文方法可以重建出人脸的皱纹等高频信息, 而其他方法却无法表现这些高频细节信息. 此外, 图 5c~图 5f 也无法体现光照信息, 而图 5b 却可以最大限度地体现出输入图像的光照信息.

为了验证本文方法的鲁棒性, 对同一人脸在不同表情、不同朝向和不同光照条件下采集的图像进行人脸重建, 结果如图 6~图 8 所示.

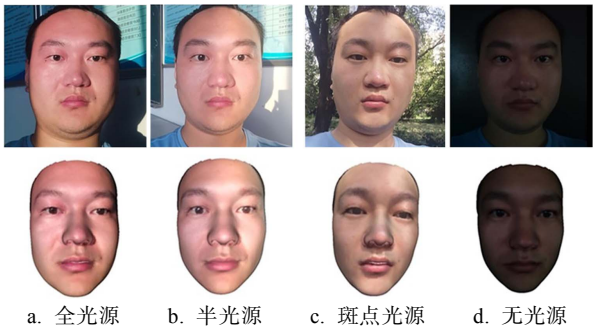
a. 中性 b. 开心 c. 痛苦 d. 惊讶 e. 生气

图 6 5 种表情重建结果对比



a. 左转 30° b. 左转 20° c. 正面 d. 右转 20° e. 右转 30°

图 7 5 种朝向重建结果对比



a. 全光源 b. 半光源 c. 斑点光源 d. 无光源

图 8 4 种光照重建结果对比

从图 6 可以看出, 本文方法可以恢复同一人脸的不同表情, 且可以恢复不同表情中复杂的几何细节.

从图 7 可以看出, 本文方法可以重建出与输入图像相近的 3D 人脸; 可以恢复不同朝向下的人脸几何信息, 并且也可以正确地估计人脸转向的角度.

从图 8 可以看出, 本文方法可以恢复不同光照条件下的 3D 人脸. 图 8c 所示为室外斑点光源条件下的人脸图像, 本文方法可以完全恢复这种光照复杂的 3D 人脸信息.

综上所述, 本文方法在不同表情、朝向和光照条件下均可以重建出逼真的 3D 人脸几何模型, 具有较高的鲁棒性.

### 3.3 重建精度

为了定量评价重建的精度, 将本文方法与文献[1]方法在时间性能上进行对比, 结果如表 1 所示. 可以看出, 与文献[1]方法相比, 本文方法训练时间短; 训练迭代收敛次数少, 只需要较少的训练次数就能够达到文献[1]方法的重建精度. 但是, 由于本文方法在生成纹理模块的过程中使用的是纹理映射的方式, 而该方式需要大量点到点的映射计算, 因此在测试阶段, 与文献[1]方法直接使用训练好的参数生成 3D 人脸所需要的时间相比, 本文方法耗时更多. 在训练阶段, 本文方法不需要回归纹理参数和光照参数, 且不需要在 3D 顶点上将光照信息添加到生成纹理的计算过程, 因此在时间性能上略高于文献[1]方法.

**表 1 2 种方法时间性能比较**

| 方法 | 测试时间/s | 训练时间/s | 训练迭代次数 |
|---|---|---|---|
| 文献[1] | 0.0748 | 1.23 | 15000 |
| 本文 | 1.6900 | 1.15 | 13000 |

为了测试多层次损失函数对重建质量的影响, 本文使用均方根均方误差(mean root mean squared error, RMSE)对重建误差进行度量. 本文使用表 2 中不同组合的损失来训练 RP-Net, 计算其生成的 3D 人脸与本文方法生成的 3D 人脸的 RMSE. 从表 2 可以看出, $L_{id}$ 和 $L_{lm}$ 对人脸重建的质量影响较大, $L_{pixel}$ 对人脸重建的质量影响相对较小.

**表 2 5 种损失组合比较**

| 损失 | | | RMSE |
|---|---|---|---|
| $L_{pixel}$ | $L_{lm}$ | $L_{id}$ | |
| | | √ | 0.089331 |
| | √ | √ | 0.469641 |
| √ | | √ | 1.082232 |
| √ | √ | | 0.372699 |
| √ | √ | √ | 0.081014 |

注. √表示在训练中使用的不同损失.

为了定量地比较重建图像的质量, 本文引入结构相似性评价(structure similarity index measure,

SSIM)和峰值信噪比(peak signal to noise ratio, PSNR)这 2 个通用图像评价指标进行度量. 从表 3 可以看出, 本文方法的 SSIM 和 PSNR 均比文献[1]方法要高, 说明本文方法重建的图像质量更好.

**表 3　2 种图像质量评价方法比较**

| 方法 | SSIM | PSNR/dB |
|---|---|---|
| 文献[1] | 0.6096 | 10.21 |
| 本文 | 0.6115 | 10.24 |

## 4　结　语

本文提出一种基于深度学习的单幅图像逼真 3D 人脸重建方法. 首先构建 RP-Net, 采用 3DMM 重建出准确的人脸几何形状. 基于 3DMM 生成的纹理过于依赖先验知识, 导致生成的纹理不够逼真. 为了达到逼真的效果, 本文采用纹理映射的方式获取逼真的纹理, 能够达到较好的效果. 此外, 本文还设计多层次的损失函数, 包括低层次的像素损失和高层次的身份损失. 同时, 本文还将多层次的损失作为弱监督信号, 对重建的人脸进行监督学习, 提高重建的质量和精度.

虽然本文方法可以恢复纹理的细节, 但是仍不能处理人脸遮挡的问题. 人脸遮挡一般分为自遮挡和外部遮挡. 自遮挡是由于人脸在不同的姿态下拍照时导致部分人脸纹理被自身的其他纹理遮挡, 无法表示遮挡部分的纹理信息. 本文方法无法映射遮挡部分的纹理信息, 故无法处理自遮挡问题. 外部遮挡是由于眼镜或头发等非人脸区域物体遮挡人脸部分区域, 本文的方法也无法处理此类问题.

在下一步的工作中, 本文将采用人脸正面化的方法获得人脸的正面信息, 解决因遮挡导致纹理缺失的问题. 本文拟采用带有自注意力机制的 GANs 来生成无遮挡的正面人脸纹理图像, 在不丢失人脸高频细节的情况下生成带有姿态和光照的人脸正面图像. 此外, 本文还将引入对抗损失函数和在 RP-Net 中引入通道自注意力机制, 加强对人脸特征的学习来进行细粒度的重建, 进一步提高人脸重建的逼真性.

## 参考文献(References):


[1] Deng Y, Yang J L, Xu S C, *et al*. Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2019: 285-295

[2] Lin J K, Yuan Y, Shao T J, *et al*. Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 5890-5899

[3] Deng J K, Cheng S Y, Xue N N, *et al*. UV-GAN: adversarial facial UV map completion for pose-invariant face recognition[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 7093-7102

[4] Gecer B, Ploumpis S, Kotsia I, *et al*. GANFIT: generative adversarial network fitting for high fidelity 3D face reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 1155-1164

[5] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[C] //Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1999: 187-194

[6] Genova K, Cole F, Maschinot A, *et al*. Unsupervised training for 3D morphable model regression[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8377-8386

[7] Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 815-823

[8] Wu S Z, Rupprecht C, Vedaldi A. Unsupervised learning of probably symmetric deformable 3D objects from images in the wild[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 1-10

[9] Lee G H, Lee S W. Uncertainty-aware mesh decoder for high fidelity 3D face reconstruction[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 6099-6108

[10] Chen D, Hua G, Wen F, Sun J, *et al*. Supervised transformer network for efficient face detection[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 122-138

[11] Liu F, Zhu R H, Zeng D, *et al*. Disentangling features in 3D face shapes for joint face reconstruction and recognition[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 5216-5225

[12] Dou P F, Shah S K, Kakadiaris I A. End-to-end 3D face reconstruction with deep neural networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1503-1512

[13] Sanyal S, Bolkart T, Feng H W, *et al*. Learning to regress 3D face shape and expression from an image without 3D supervi-

sion[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 7755-7764

[14] Zhu X Y, Lei Z, Liu X M, *et al*. Face alignment across large poses: a 3D solution[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 146-155

[15] Richardson E, Sela M, Or-El R, *et al*. Learning detailed face reconstruction from a single image[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5553-5562

[16] Zhao D P, Qi Y. Generative face parsing map guided 3D face reconstruction under occluded scenes[C] //Proceedings of the 38th Computer Graphics International Conference. Heidelberg: Springer, 2021: 252-263

[17] Paysan P, Knothe R, Amberg B, *et al*. A 3D face model for pose and illumination invariant face recognition[C] //Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance. Los Alamitos: IEEE Computer Society Press, 2009: 296-301

[18] Cao C, Weng Y L, Zhou S, *et al*. FaceWarehouse: a 3D facial expression database for visual computing[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(3): 413-425

[19] Liu Z W, Luo P, Wang X G, *et al*. Deep learning face attributes in the wild[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 3730-3738

[20] Bulat A, Tzimiropoulos G. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230, 000 3D facial landmarks)[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1021-1030

[21] Deng J K, Guo J, Xue N N, *et al*. ArcFace: additive angular margin loss for deep face recognition[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 4685-4694

[22] Cao Q, Shen L, Xie W D, *et al*. VGGFace2: a dataset for recognising faces across pose and age[C] //Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 67-74

[23] Bagdanov A D, del Bimbo A, Masi I. The Florence 2D/3D hybrid face dataset[C] //Proceedings of the Joint ACM Workshop on Human Gesture and Behavior Understanding. New York: ACM Press, 2011: 79-80

[24] Zhang K P, Zhang Z P, Li Z F, *et al*. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503

[25] Bai Z Q, Cui Z P, Liu X M, *et al*. Riggable 3D face reconstruction via in-network optimization[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 6212-6221

[26] Cai Lin, Guo Yudong, Zhang Juyong. High-quality 3D face reconstruction from multi-view images[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(2): 305-314(in Chinese)
(蔡麟, 郭玉东, 张举勇. 基于多视角的高精度三维人脸重建[J]. 计算机辅助设计与图形学学报, 2020, 32(2): 305-314)

[27] Yang M X, Guo J W, Ye J T, *et al*. Detailed 3D face reconstruction from single images via self-supervised attribute learning[C] //Proceedings of ACM Special Interest Group for Computer Graphics and Interactive Techniques Asia Posters. New York: ACM Press, 2020: Article No.26

[28] Jackson A S, Bulat A, Argyriou V, *et al*. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1031-1039

[29] Tran L, Liu X M. Nonlinear 3D face morphable model[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 7346-7355

[30] Chen A P, Chen Z, Zhang G L, *et al*. Photo-realistic facial details synthesis from single image[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 9428-9438