Journal of Computer Applications
计算机应用,2020,40(11):3306 - 3313
ISSN 1001-9081
CODEN JYIIDU
2020-11-10
http://www.joca.cn

文章编号:1001-9081(2020)11-3306-08
DOI:10.11772/j.issn.1001-9081.2020030420

基于改进三维形变模型的三维人脸重建和密集人脸对齐方法

周 健,黄章进*
(中国科学技术大学 计算机科学与技术学院,合肥 230027)
(*通信作者电子邮箱zhuang@ustc.edu.cn)

摘 要:针对现在广泛使用的三维形变模型表达能力不够,导致重建出的三维人脸模型泛化性能不佳的问题,提出了一种在姿态、表情和光照未知的条件下的基于单张人脸图片的三维人脸重建和密集人脸对齐的新方法。首先,通过卷积神经网络对现有的三维形变模型进行改进,以提高三维人脸模型的表达能力;然后,基于人脸光滑性和图像相似性,在特征点和像素层面提出新的损失函数,并使用弱监督学习训练卷积神经网络模型;最后,通过训练出的网络模型进行三维人脸重建和密集人脸对齐。实验结果表明,对于三维人脸重建任务,所提模型在AFLW2000-3D上实现了2.25的归一化平均误差;对于密集人脸对齐任务,所提模型在AFLW2000-3D和AFLW-LFPA上分别实现了3.80和3.34的归一化平均误差。与原始使用三维形变模型的方法相比,所提模型在三维人脸重建和密集人脸对齐上的归一化平均误差分别降低了7.4%和7.8%。针对不同光照环境以及角度的人脸图片,该网络模型的重建准确,鲁棒性好,且具有较高的三维人脸重建和密集人脸对齐质量。

关键词:三维人脸重建;密集人脸对齐;三维形变模型;弱监督学习;卷积神经网络
中图分类号:TP391.41 **文献标志码**:A

3D face reconstruction and dense face alignment method based on improved 3D morphable model

ZHOU Jian, HUANG Zhangjin*
(*School of Computer Science and Technology, University of Science and Technology of China, Hefei Anhui 230027, China*)

Abstract: In order to solve the problem that the currently widely used 3D morphable model has insufficient expression ability, resulting in poor generalization performance of the reconstructed 3D face model, a novel method for 3D face reconstruction and dense face alignment based on a single face image under unknown pose, expression and illumination was proposed. First, the existing 3D morphable model was improved by convolutional neural network to improve the expression ability of the 3D face model. Then, based on the smoothness of the face and the similarity of the image, a new loss function was proposed at the feature point and pixel level, and the weakly-supervised learning was used to train the convolutional neural network model. Finally, the trained network model was used to perform the 3D face reconstruction and dense face alignment. Experimental results show that, for 3D face reconstruction, the proposed model has the normalized mean error on AFLW2000-3D reduced to 2.25, and for dense face alignment, the proposed model has the normalized mean errors on AFLW2000-3D and AFLW-LFPA reduced to 3.80 and 3.34 respectively. Compared with the original method using 3D morphable model, the proposed model has the normalized mean errors reduced by 7.4% and 7.8% respectively in 3D face reconstruction and dense face alignment. Therefore, for face images with different lighting environments and angles, this network model is accurate in reconstruction and robust, and has high 3D face reconstruction and dense face alignment quality.

Key words: 3D face reconstruction; dense face alignment; 3D morphable model; weakly-supervised learning; Convolutional Neural Network (CNN)

0 引言

基于单张人脸图片的三维人脸重建和密集人脸对齐是计算机视觉和计算机图形学领域一项具有挑战性的任务,在人脸识别、人脸动画、人脸表情迁移、人脸对齐等方面有着非常广泛的应用。传统的二维人脸对齐方法在遇到大姿态人脸图片和具有遮挡的人脸图片时往往会受到很大的挑战,而通过从单张人脸图片进行三维人脸几何重建可以很好地应对这些挑战。由于三维人脸网格模型强大的拓扑性质,三维人脸网格模型的数万个顶点都可以被视为密集三维人脸对齐可以使用的人脸特征点。传统的三维人脸重建方法[1-3]主要是基于优化方法,但这些方法通常优化时间复杂度较高、耗时较长,且存在局部最优解以及初始化不佳的问题。在卷积神经网络

收稿日期:2020-04-07;**修回日期**:2020-05-15;**录用日期**:2020-05-19。
基金项目:国家自然科学基金资助项目(61877056);中央高校基本科研业务费专项基金资助项目(WK6030000109)。
作者简介:周健(1995—),男,河南卫辉人,硕士研究生,主要研究方向:计算机视觉、计算机图形学、三维人脸重建; 黄章进(1980—),男,湖北天门人,副教授,博士,主要研究方向:计算机视觉、机器学习、计算机图形学。

(Convolutional Neural Network, CNN)出现之后,基于卷积神经网络的方法[4-5]在三维人脸重建以及密集人脸对齐方面取得了显著的成功。这些方法通常使用CNN来预测三维形变模型(3D Morphable Model, 3DMM)[6]的系数,显著提高了三维人脸重建的质量和效率。基于卷积神经网络的方法通常需要大量的数据集,这些数据集采集往往代价较高,甚至在很多场景下无法实现。最近一些方法[5,7-8]使用合成数据进行训练,也有例如300W-LP[4]这样的公开合成人脸数据集。但是通过合成方法生成的人脸图片通常与真实图片之间存在一定的差距,它们在表情、光照、背景方面缺乏多样性,这往往导致训练好的卷积神经网络模型泛化性能较差。

为了解决三维人脸重建数据集缺失的问题,一些最近工作[5,9-11]使用弱监督学习的方法,这些方法仅需要二维人脸图片及其对应的二维人脸特征点。使用此方法,经过训练的三维人脸重建模型可以很好地进行三维人脸重建以及密集三维人脸对齐。现阶段很容易获得带有二维人脸特征点的人脸图片数据集,因此可以建立大量训练集以满足卷积神经网络的需要。此外,这些二维人脸特征点也可以提供很有价值的人脸信息。目前使用弱监督的三维人脸重建方法的关键是使用一种可微分渲染器将重建好的三维人脸渲染到像素层面,并且比较该渲染后图片与输入图片之间的差异。例如Tewari等[5,10]使用渲染后图片与输入图片在像素颜色之间的差异建立损失函数;Genova等[11]及Tran等[9]使用人脸识别网络来建立渲染后图片与输入图片之间的损失。

由于现在广泛使用的3DMM只是通过200张人脸来建立,而且其中大部分都是白人。通过实验分析,发现目前的3DMM训练出的模型会使人脸丢失很多细节信息;同时对于有色人种,重建的纹理质量也会受到很大影响。为了缓解这一现象,本文设计了一个基于弱监督学习的卷积神经网络模型:将现有的3DMM作为CNN中全连接层的初始化权重,直接回归出输入人脸图片对应的三维人脸模型,从而可以从输入人脸图片进行准确的三维人脸重建和密集人脸对齐。同时,本文设计了一个新的损失函数,用于对学习到的三维人脸模型进行光滑性约束以及更好地建立输入人脸图片和渲染图片的相似性差异。本文主要有如下技术贡献:

1)提出了一个端到端的弱监督卷积神经网络,用全连接层代替3DMM,并使用3DMM模型参数作为初始化权重,对现有的3DMM进行改进,有效地重建出更精准的三维人脸模型。

2)提出了一种新的损失函数。该损失函数对学习到的三维人脸模型进行光滑性约束,并且使用输入人脸图片和渲染后图片的结构相似性(Structural SIMilarity, SSIM)作为损失。同时,该损失函数对先前训练中具有较大损失值的特征点进行强化训练。

通过在AFLW2000-3D和AFLW-LFPA数据集上和其他方法进行比较,实验结果表明本文的方法在三维人脸重建和密集人脸对齐两方面都取得了更好的结果。

1　相关研究

1.1　三维人脸模型

三维人脸模型是三维人脸重建中被广泛使用的人脸模型。与点云相比,三维人脸模型能够提供人脸形状、表情以及纹理的先验知识,而且可以将复杂的三维模型变为一组三维模型的线性组合。因此就可以使用一组系数向量来表达重建的三维人脸模型。

Cootes等[12]引入主动外观模型(Active Appearance Model, AAM)作为二维形状和纹理的统计变形模型。AAM是一种生成模型,在拟合过程中通过优化来恢复某一对象的参数描述;作为二维AAM的扩展,Blanz等[6]引入了3DMM。使用主成分分析(Principal Component Analysis, PCA)分解几何和纹理信息,这有效地减少了形状和纹理空间尺寸;后来,Gerig等[13]对3DMM进行了扩展,在3DMM中加入了表情;Booth等[14]给出了9 663个不同人脸特征的大规模人脸模型(Large Scale Face Model, LSFM),这个模型包含了来自大量不同人群的统计信息;Tran等[15]提出了一种非线性三维人脸形变模型,直接从大量的无约束人脸图像中训练出一个非线性三维形变模型,而不需要采集三维人脸扫描图像。本文使用3DMM作为初始化的人脸模型,并使用卷积神经网络来对3DMM进行改进。

1.2　三维人脸重建

单目三维人脸重建方法通常分为两大类:基于优化的方法和基于回归的方法。基于优化的方法通常建立能量函数来描述图像的一些自然过程。许多方法使用明暗恢复形状(Shape-From-Shading, SFS)[1,16]或光流(Optical Flow, OF)[2]来模拟图像信息。这类方法的主要缺点是计算复杂度高,重建速度慢;同时,优化方法对初始化非常敏感,且需要精确的二维特征点检测[17-18]。

近年来,基于回归的方法出现频率不断增加[4-5,19-20]。特别是卷积神经网络出现以后,许多方法都基于卷积神经网络来实现[4-5,7,9,21]。卷积神经网络通常需要输入大量数据来进行训练。然而,目前三维人脸数据集通常较小,因此此前的方法主要分为使用合成数据[7-8,22-23]和使用弱监督[5,10-11,15,24]来代替。但是,合成人脸图片和真实人脸图片之间通常会有很大差异。这会导致模型的泛化能力显著下降,往往在真实人脸图片上测试效果不理想。因此,当前大多数方法都采用了弱监督学习的方法,不需要人脸图片对应的三维人脸模型标签。

Richardson等[7]使用三维形变模型生成不同形状、表情以及纹理的图像,然后将其渲染成二维图像,这样就得到了真实三维人脸模型标签的图像来进行网络训练;但是这有一定局限性,主要是合成人脸图像和真实人脸图像之间存在一定差距。Tewari等[5]训练了一个自编码器网络来回归形状、表情、纹理、姿态以及光照,利用回归得到的参数生成三维人脸模型,然后利用可微分渲染器将三维人脸模型渲染到二维平面上。通过将输入人脸图像与渲染后人脸图像进行比较,建立损失函数。这种方法不需要二维图像对应的三维人脸模型标签,结果明显优于使用合成数据训练的模型。Richardson等[25]将明暗恢复形状融入学习过程中,以学习更详细的细节信息。Jackson等[21]训练了一个卷积神经网络来直接从一张二维图像中还原三维面部几何的体素表示,这是一种无模型的方法,不需要三维人脸形变模型。Feng等[26]训练卷积神经网络从单个二维图像中回归UV位置图,得到相应的三维面部结构,该方法也不依赖于任何先验人脸模型。Deng等[24]使用了同一个人的多张图片进行训练,多张图片可以从不同视角补充信息,具有防遮挡的优点。Shi等[27]通过训练卷积神经网络,从游戏中的面部图像重建出相匹配的卡通人脸模型。Chang等[28]使用了三个深层卷积神经网络,分别从一张人脸图片估计出三维人脸形状、视角和表情。这些方法大多数使

用3DMM作为三维人脸模型。但是3DMM只是建立在少量人脸图片上，且人种单一。这导致训练出来的3DMM模型易丢失细节信息，同时不易泛化到其他有色人种。基于此，本文将3DMM作为全连接层的初始化权重并设计了一个基于弱监督学习的卷积神经网络，从而可以对输入人脸图片进行准确的三维人脸重建，取得更好的效果。

1.3 密集人脸对齐

在计算机视觉领域，人脸对齐问题长期以来吸引了很多关注。传统的二维人脸对齐方法主要是定位一组稀疏的人脸关键点，比如说AAM[12]以及约束局部模型(Constrained Local Model, CLM)[29]。近年来，随着深度学习的不断发展，基于卷积神经网络的方法[30-31]在二维人脸对齐方面取得了最先进的性能。但是二维人脸对齐有一定局限性，它们通常只能检测人脸上可见的特征点，因此当人脸的姿态较大或者出现遮挡时，一些人脸特征点变得不可见，这些方法便无法处理。

最近，学者们开始研究三维人脸对齐，主要是使用3DMM进行拟合[4,32-33]或者是将2D人脸图像和3D人脸模板进行匹配[34-35]。基于模型的三维重建方法可以通过选择三维重建模型中的x、y坐标来实现二维的人脸对齐任务。文献[4,36]均使用特定的方法来对3DMM进行拟合来完成人脸对齐任务。文献[37-38]使用深度神经网络直接预测热力图来获得三维人脸特征点，并且实现了较为先进的性能。文献[20,39-40]对于3DMM的系数进行估计，然后将估计出的三维人脸特征点投影到二维空间中，这样可以显著提高效率。近些年来，密集人脸对齐任务开始受到越来越多的关注，目标是实现大姿态人脸图像的密集三维对齐。Liu等[41]使用多重约束来训练CNN模型，估计3DMM的系数，可以实现密集的三维估计。文献[36,42]利用深度卷积神经网络在只考虑可见区域的情况下学习二维人脸图像和三维模板之间的关系。

这些方法通常仅在特征点处建立损失函数，而忽略像素层面的信息。本文的方法综合了特征点和像素层面的信息，并且使用了SSIM作为损失函数来衡量输入人脸图片和渲染后图片的相似性。实验表明，本文方法取得了更好的效果。

2 算法设计

本章主要介绍提出的三维人脸重建算法模型。整体流程如图1所示，它的目的是训练一个卷积神经网络回归模型。回归模型输入一张224×224×3大小的人脸图片以及其对应的二维人脸特征点信息，经过VGG-16[43]回归出人脸的3DMM系数、相机参数以及球谐光照系数，随后通过两个全连接层对三维人脸模型的形状和纹理进行改进。通过加入球谐光照重建出对应的三维人脸模型，随后使用全透视投影经过一个可微分渲染器将重建出的三维人脸模型渲染到二维平面上，建立特征点、像素空间和人脸光滑性损失函数。最后通过反向传播训练整体网络。

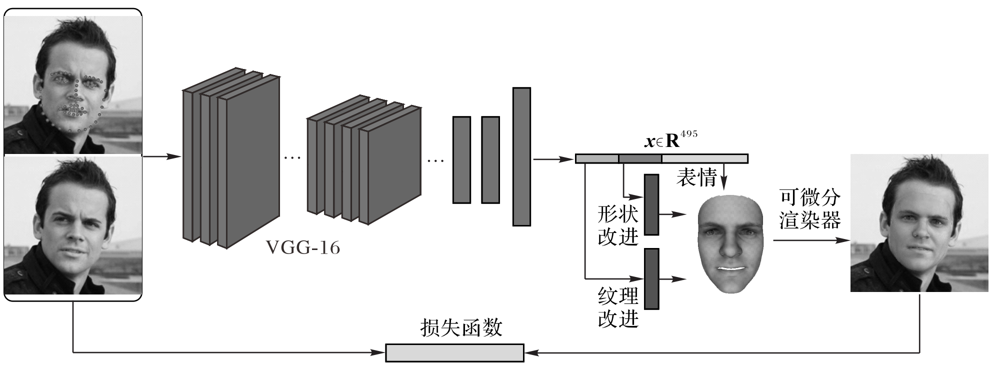


图1 三维人脸重建算法流程

Fig. 1 Flowchart of 3D face reconstruction algorithm

2.1 人脸模型

由于人脸比较复杂，很难直接回归出三维人脸模型，因此现阶段学者们广泛使用3DMM来进行三维人脸重建。3DMM不仅能够保证人脸重建中不会出现非人脸的情况，也可以建立不同脸型之间的对应关系。但是目前通常使用的三维形变模型BFM09(2009 Basel Face Model)[6]存在训练数据少、数据集中人脸多样性不足等缺点，导致重建质量不佳、人脸易丢失细节。

为解决人脸模型不能反映足够的人脸细节这一问题，本文对现有的3DMM进行改进。由于三维形变模型的实质是一个多维的三维形变函数，它是基于大量的三维人脸的扫描模型的线性组合，因此三维形变模型可以看作一个线性模型，这和神经网络中全连接层的结构是一致的。本文将参数与3DMM进行结合、重建出三维人脸这一过程看作一个全连接层操作，通过在全连接层中引入更多参数，并且使用已有的3DMM模型作为初始化参数，进而使得人脸模型更加符合真实人脸的分布。

本方法采用参数化的三维人脸几何模型$\boldsymbol{S}=\{\boldsymbol{s}_i \in \mathbf{R}^3 | 1 \leqslant i \leqslant N\}$作为初始的人脸几何模型，其中$N$=35 709为顶点个数。同时采用参数化的人脸纹理模型$\boldsymbol{T}=\{\boldsymbol{t}_i \in \mathbf{R}^3 | 1 \leqslant i \leqslant N\}$作为初始的人脸纹理模型。如式(1)所示：

$$\begin{cases} \boldsymbol{S}=\boldsymbol{S}(\boldsymbol{\alpha},\boldsymbol{\delta})=\overline{\boldsymbol{S}}+\boldsymbol{E}_{\text{shape}}\boldsymbol{\alpha}+\boldsymbol{E}_{\exp}\boldsymbol{\delta} \\ \boldsymbol{T}=\boldsymbol{T}(\boldsymbol{\beta})=\overline{\boldsymbol{T}}+\boldsymbol{E}_{\text{tex}}\boldsymbol{\beta} \end{cases} \tag{1}$$

其中$\overline{\boldsymbol{S}}$是平均人脸几何模型，$\overline{\boldsymbol{T}}$是平均人脸纹理模型。$\boldsymbol{E}_{\text{shape}} \in \mathbf{R}^{3N \times 199}$、$\boldsymbol{E}_{\text{tex}} \in \mathbf{R}^{3N \times 199}$和$\boldsymbol{E}_{\exp} \in \mathbf{R}^{3N \times 64}$分别是人脸形状、纹理和表情的PCA基。$\boldsymbol{\alpha} \in \mathbf{R}^{199}$、$\boldsymbol{\beta} \in \mathbf{R}^{199}$和$\boldsymbol{\delta} \in \mathbf{R}^{64}$是三维人脸模型对应的形状、纹理和表情的系数。在本文中，$\overline{\boldsymbol{S}}$、$\overline{\boldsymbol{T}}$、$\boldsymbol{E}_{\text{shape}}$和$\boldsymbol{E}_{\text{tex}}$使用BFM09，$\boldsymbol{E}_{\exp}$是从FaceWarehourse[44]中提取。

通过建立两个全连接层FC_{shape}和FC_{texture}来对现有的人脸形状和纹理的PCA基进行改进。全连接层FC_{shape}的大小为199×107 127，输入为199和3DMM中人脸形状的PCA基的个数一致，输出107 127为顶点的个数35 709乘以3，并使用

3DMM 中人脸形状的 PCA 基 $\boldsymbol{E}_{\text{shape}}$ 作为初始化权重，从而得到改进的人脸形状 $\boldsymbol{S}_{\text{new\_shape}}$。同理，全连接层 FC_{texture} 的大小为 199×107 127，使用 3DMM 中人脸纹理的 PCA 基 $\boldsymbol{E}_{\text{tex}}$ 作为初始化权重，得到改进的人脸纹理模型 $\boldsymbol{T}_{\text{new\_texture}}$。最终的三维人脸重建结果如式(2)所示：

$$\begin{cases} \boldsymbol{S} = \boldsymbol{S}(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \overline{\boldsymbol{S}} + \boldsymbol{S}_{\text{new\_shape}} + \boldsymbol{E}_{\text{exp}}\boldsymbol{\delta} \\ \boldsymbol{T} = \boldsymbol{T}(\boldsymbol{\beta}) = \overline{\boldsymbol{T}} + \boldsymbol{T}_{\text{new\_texture}} \end{cases} \tag{2}$$

2.2　相机模型

相机模型用于将人脸网格模型从三维空间转化到二维平面。和文献[11]相同，本文使用全透视投影模型。相机在世界坐标系中的位置和方向由旋转矩阵 $\boldsymbol{R} \in SO(3)$（三维旋转群）和平移向量 $\boldsymbol{m} \in \mathbf{R}^3$ 来表示，如式(3)所示：

$$\boldsymbol{q} = \Pi(\boldsymbol{R}\boldsymbol{p} + \boldsymbol{m}) \tag{3}$$

其中：$\boldsymbol{p} \in \mathbf{R}^3$ 为顶点在世界坐标系下的坐标，$\boldsymbol{q} \in \mathbf{R}^2$ 为顶点在图像平面的坐标，$\Pi: \mathbf{R}^3 \to \mathbf{R}^2$ 为全透视投影模型将相机坐标系变为像素平面坐标系。

2.3　光照模型

本方法假设光照是低频的，并且将人脸表面近似看作一个兰伯特曲面。基于这两个假设，本文使用球谐函数[45]来表示光照。顶点颜色 $\boldsymbol{C}(\boldsymbol{t}_i, \boldsymbol{n}_i, \boldsymbol{\gamma})$ 由网格顶点纹理 $\boldsymbol{t}_i \in \mathbf{R}^3$、网格顶点法向量 $\boldsymbol{n}_i \in \mathbf{R}^3$ 和光照系数 $\boldsymbol{\gamma} \in \mathbf{R}^{27}$ 来计算，如式(4)所示：

$$\boldsymbol{C}(\boldsymbol{t}_i, \boldsymbol{n}_i, \boldsymbol{\gamma}) = \boldsymbol{t}_i \cdot \sum_{b=1}^{B^2} \boldsymbol{r}_b \mathrm{H}_b(\boldsymbol{n}_i) \tag{4}$$

其中 $\mathrm{H}_b: \mathbf{R}^3 \to \mathbf{R}$ 是球谐基函数，$\boldsymbol{\gamma} = \{\boldsymbol{r}_b \in \mathbf{R}^3 | 1 \leqslant b \leqslant B^2\}$ 是对应的光照系数。本文使用前三阶 B=3。

2.4　网络模型

本文提出的三维人脸重建模型主要包括两个模块：一个是基于卷积神经网络的回归模块，将一张人脸图片输入卷积神经网络，回归出对应的三维人脸的 3DMM 系数、球谐光照参数以及相机模型参数；另一个是三维形变模型改进模块，以现有的 3DMM 作为全连接层的初始化权重，直接回归出三维人脸模型的形状和纹理。

本方法采用 VGG-16[43]作为回归模块。回归模块的输入为一张 224×224 大小的 RGB 人脸图片。要回归的三维人脸参数 $\boldsymbol{x} \in \mathbf{R}^{495}$ 包含 3DMM 形状参数 $\boldsymbol{\alpha} \in \mathbf{R}^{199}$、3DMM 纹理参数 $\boldsymbol{\beta} \in \mathbf{R}^{199}$、3DMM 表情参数 $\boldsymbol{\delta} \in \mathbf{R}^{64}$、相机旋转 $\boldsymbol{R} \in SO(3)$、相机平移 $\boldsymbol{m} \in \mathbf{R}^3$ 以及球谐光照参数 $\boldsymbol{\gamma} \in \mathbf{R}^{27}$ 如式(5)所示：

$$\boldsymbol{x} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{R}, \boldsymbol{m}, \boldsymbol{\gamma}) \tag{5}$$

编码器包含 13 个卷积层，每个卷积层后都接着一个 ReLU（Rectified Linear Unit）激活层。在第 2、4、7、10 以及 13 卷积层后接着最大池化层。卷积层后面跟着三个全连接层，大小分别为 4 096、4 096 和 495。

三维形变模型改进模块包含两个全连接层，分别连接在 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 参数的后面，大小分别为 199×107 127 和 199×107 127。分别使用 3DMM 中人脸形状和纹理的 PCA 基作为初始化权重。

2.5　损失函数

本文设计了一种新的损失函数。该损失函数不仅考虑到特征点和像素空间的信息，而且对于学习到的三维人脸模型进行了光滑性约束，有效提升了重建的质量。新的损失函数包含六项，如式(6)所示：

$$\begin{aligned} L_{\text{loss}}(\boldsymbol{x}) = {} & \omega_{\text{land}} L_{\text{land}}(\boldsymbol{x}) + \omega_{\text{land\_error}} L_{\text{land\_error}}(\boldsymbol{x}) + \\ & \omega_{\text{photo}} L_{\text{photo}}(\boldsymbol{x}) + \omega_{\text{ssim}} L_{\text{ssim}}(\boldsymbol{x}) + \\ & \omega_{\text{smooth}} L_{\text{smooth}}(\boldsymbol{x}) + \omega_{\text{reg}} L_{\text{reg}}(\boldsymbol{x}) \end{aligned} \tag{6}$$

其中：$L_{\text{land}}(\boldsymbol{x})$ 和 $L_{\text{land\_error}}(\boldsymbol{x})$ 为特征点对齐的损失函数和特征点增强训练的损失函数，$L_{\text{photo}}(\boldsymbol{x})$ 为原图和三维人脸渲染图片像素之间差异的损失函数，$L_{\text{ssim}}(\boldsymbol{x})$ 为原图和三维人脸渲染图片之间 SSIM 的损失函数，$L_{\text{smooth}}(\boldsymbol{x})$ 为三维人脸模型光滑性约束损失函数，$L_{\text{reg}}(\boldsymbol{x})$ 是正则化项损失函数。为了平衡各个部分的损失函数，本方法将权重设置为 ω_{land}=400、$\omega_{\text{land\_error}}$=2 000、$\omega_{\text{photo}}$=100、$\omega_{\text{ssim}}$=2、$\omega_{\text{smooth}}$=50 以及 ω_{reg}=1。

2.5.1　特征点损失函数

本方法将二维人脸图片的特征点作为一种弱监督信息来训练神经网络，同时使用现阶段比较先进的人脸特征点检测算法[38]来检测训练集中人脸图片的 68 个关键点。损失函数 $L_{\text{land}}(\boldsymbol{x})$ 如式(7)所示：

$$L_{\text{land}}(\boldsymbol{x}) = \sum_{i=1}^{68} w_i \times \| \boldsymbol{v}_{k_i} - \boldsymbol{v}_i' \|_2^2 \tag{7}$$

其中 w_i 为特征点对应的权重。固定的 52 个特征点权重为 1，16 个轮廓特征点的权重为 0.5，固定特征点和轮廓特征点的选取如图 2 所示，其中人脸中间显示实心菱形的 52 个点为固定特征点，人脸边界显示空心菱形的 16 个点为轮廓特征点。$\boldsymbol{v}_i' \in \mathbf{R}^2$ 是人脸二维特征点的真实标签，$k_i \in \{1, 2, \cdots, N\}$ 是对应的三维人脸模型顶点索引，$\boldsymbol{v}_{k_i}$ 是重建出的三维人脸模型投影到像素平面的坐标。

图 2　人脸特征点示意图

Fig. 2　Schematic diagram of facial feature points

为了加强训练那些误差相对较大的特征点，在第 5 次迭代后，加入损失函数 $L_{\text{land\_error}}(\boldsymbol{x})$ 如式(8)所示：

$$L_{\text{land\_error}}(\boldsymbol{x}) = \sum_{i=1}^{52} e_i \times \| \boldsymbol{v}_{k_i} - \boldsymbol{v}_i' \|_2^2 \tag{8}$$

其中 e_i 为固定的 52 个特征点分别在上一个迭代中训练中的平均误差。

2.5.2　像素损失函数

像素损失函数 $L_{\text{photo}}(\boldsymbol{x})$ 的目标是使渲染后的图像和输入图像尽可能相近，将重建好的三维人脸模型渲染到像素空间上，和输入单目人脸图片进行对齐。为了将三维人脸模型渲染到二维平面上，本方法使用了可微分渲染器[11]。将渲染图片与输入单目人脸图片进行匹配，比较它们在像素空间的相似性。损失函数 $L_{\text{photo}}(\boldsymbol{x})$ 如式(9)所示：

$$L_{\text{photo}}(\boldsymbol{x}) = \frac{1}{n} \sum_{i \in \boldsymbol{V}} \| \boldsymbol{I}_i - \boldsymbol{I}_i' \|_2 \tag{9}$$

其中：$\boldsymbol{V}$ 为像素平面上所有经过投影得到的人脸区域像素点的集合，n 为 $\boldsymbol{V}$ 中像素点的个数，$\boldsymbol{I}_i$ 为输入的单目人脸图片在 i

位置的颜色，I_i' 为将三维人脸模型渲染到像素空间后得到的图像在 i 位置的颜色。

2.5.3 SSIM损失函数

SSIM损失函数的目标是保障输入图像和渲染图像的结构相似性。通过添加SSIM损失函数，可以更好地重建三维人脸模型的纹理。损失函数 $L_{\mathrm{ssim}}(\boldsymbol{x})$ 如式(10)所示：

$$L_{\mathrm{ssim}}(\boldsymbol{x}) = 1 - \frac{(2\mu_I \mu_{I'} + c_1)(2\sigma_{II'} + c_2)}{(\mu_I^2 + \mu_{I'}^2 + c_1)(\sigma_I^2 + \sigma_{I'}^2 + c_2)} \tag{10}$$

其中 $\mu_I = \frac{1}{n} \cdot \sum_{i \in \boldsymbol{V}} \boldsymbol{I}_i$ 为输入图片中可见区域的平均纹理，$\mu_{I'} = \frac{1}{n} \cdot \sum_{i \in \boldsymbol{V}} \boldsymbol{I}_i'$ 为渲染图片中可见区域的平均纹理，$\boldsymbol{V}$ 为像素平面上所有经过投影得到的人脸区域像素点集合，n 为 $\boldsymbol{V}$ 中像素点的个数，$\sigma_{II'}$ 为输入图片和渲染图片可见区域纹理的协方差，σ_I^2 为输入图片中可见区域纹理的方差，$\sigma_{I'}^2$ 为渲染图片中可见区域纹理的方差，$c_1 = 0.012$，$c_2 = 0.032$。

2.5.4 三维人脸光滑性约束

为了防止三维人脸重建的结果出现面片翻转、表面粗糙等情况，本文引入了三维人脸的光滑性约束，来保证重建出的三维人脸模型形状光滑。损失函数 $L_{\mathrm{smooth}}(\boldsymbol{x})$ 如式(11)所示：

$$L_{\mathrm{smooth}} = \frac{1}{N}\sum_{i=1}^{N}\left\| \frac{1}{d_i}\sum_{j \in Adj_j}(\boldsymbol{v}_i - \boldsymbol{v}_j)\right\|_2^2 \tag{11}$$

其中：N 为三维人脸模型的顶点个数，d_i 为三维人脸模型第 i 个顶点的度，Adj_j 为三维人脸模型第 j 个顶点的邻居索引的集合，$\boldsymbol{v}_i$ 为三维人脸模型第 i 个顶点的坐标。

2.5.5 正则化项

在训练过程中，为了防止训练过程中出现三维人脸形状和表情变形，本方法在回归的3DMM系数上加入正则化项的损失，它能够让预测得到3DMM的系数强制满足先验的正态分布。损失函数 $L_{\mathrm{reg}}(\boldsymbol{x})$ 如式(12)所示：

$$L_{\mathrm{reg}} = \omega_\alpha \|\boldsymbol{\alpha}\|^2 + \omega_\beta \|\boldsymbol{\beta}\|^2 + \omega_\delta \|\boldsymbol{\delta}\|^2 \tag{12}$$

其中：$\omega_\alpha = 2 \times 10^{-5}$，$\omega_\beta = 2 \times 10^{-2}$，$\omega_\delta = 4 \times 10^{-4}$。

3 实验结果与分析

在一些真实的人脸图片和数据集上测试本方法在三维人脸重建和密集人脸对齐上的性能。首先，介绍实验的训练细节和测试数据集；然后，展示本方法和其他方法进行比较的实验结果；最后，给出是否使用3DMM改进模块和是否在损失函数中使用三维人脸光滑性约束的消融实验。

3.1 实验细节

Bulat等[38]的方法可以检测出较为粗糙的68个完整特征点，但是当人脸图片中人脸角度较大时，可能会出现特征点检测误差较大的情况。因此本文对人脸角度超过45°的人脸图片手动进行筛选，过滤掉特征点误差较大的图片，本方法从CelebA[46]和300W-LP中选取了共约250 000张人脸图片（如图3所示）用于训练卷积神经网络，并对这些图片进行数据增强。数据增强包括图片翻转、图片随机旋转以及模拟光照。图片翻转为对图片进行水平翻转，在训练过程中对于图片进行随机选择是否进行翻转操作，是否翻转操作的概率均为50%。图片旋转操作是对人脸图片基于中心点按照顺时针方向旋转 −30°~30°、旋转角度为 −30°~30° 的均匀分布。此外，为了降低光照对于人脸图片的影响，本方法对于训练集中的人脸图片进行模拟光照操作。模拟光照操作对人脸图片的颜色通道RGB随机乘上0.7~1.3，三个通道独立操作，概率分布为满足从0.7~1.3的均匀分布。

卷积神经网络的输入图片大小为224×224×3。在卷积神经网络的训练过程中，批量大小为16，学习器为Adam优化器[47]，初始学习率为 1×10^{-5}，在5个迭代之后衰减到 1×10^{-6}。

对于损失函数 $L_{\mathrm{loss}}(\boldsymbol{x})$，特征点损失项、像素损失项以及SSIM损失项提供了弱监督的信息来帮助神经网络进行训练，它们是较强的损失项，这三个损失项的数值越小，说明现有的卷积神经网络模型在训练集上的表现越好。经过实验研究发现：特征点、像素以及SSIM损失项的数值比为2：2：1时能保证三项损失之间的平衡，效果最好，因此本文选择 ω_{land}=400、$\omega_{\mathrm{land\_error}}$=2 000、$\omega_{\mathrm{photo}}$=100以及 ω_{ssim}=2。但这三个损失项的数值越小并不代表训练出的卷积神经网络模型表达能力越好，因为可能导致过拟合的问题，因此本文在损失函数中加入惩罚项来进行约束。惩罚项过大，会导致卷积神经网络模型陷入欠拟合；而惩罚项过小，会导致卷积神经网络陷入过拟合。因此经过实验分析，本文在选择 ω_{smooth}=50以及 ω_{reg}=1时效果最好。

图3 训练集中部分真实场景人脸图片

Fig. 3 Some real scene face images in the training set

3.2 测试数据集

AFLW2000-3D是通过选择AFLW数据集中前2 000张图片构建的。每张人脸图片有对应的3DMM系数以及68个三维人脸特征点位置。本文使用这个数据集来验证三维人脸重建以及密集人脸对齐的效果。

AFLW-LFPA[48]是AFLW数据集的另外一个扩展，它是根据人脸姿态从AFLW中选择图像构建的。它包含1 299张测试图像，具有平衡的偏航角分布。每张人脸图片都有34个人脸关键点。本文使用这个数据集来评估模型在密集人脸对齐任务上的性能。

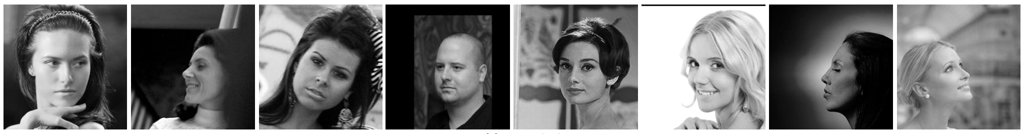
3.3 密集人脸对齐结果

作为三维人脸重建的应用，本实验首先验证本方法在ALFW2000-3D上的效果，效果如图4所示，其中第一行为输入图片，第二行为密集人脸对齐效果。为了将本方法和其他相关方法进行比较，本实验采用归一化平均误差（Normalized Mean Error, NME）[4]来作为评价算法性能的指标。归一化平均误差是根据人脸包围盒的大小进行归一化，如式(13)所示：

$$I_{\mathrm{NME}} = \frac{1}{T}\sum_{k=1}^{N}\frac{\|\boldsymbol{u}_k - \boldsymbol{y}_k\|_2}{d} \tag{13}$$

其中：T 为点的个数，d 为人脸真实的包围盒长宽之积的平方根，计算公式为 $d=\sqrt{w_{\mathrm{bbox}} * h_{\mathrm{bbox}}}$，$\boldsymbol{u}_k \in \mathbf{R}^2$ 和 $\boldsymbol{y}_k \in \mathbf{R}^2$ 分别为预测的点坐标和测试集上的标签。

本实验采用68个稀疏的特征点来进行衡量密集人脸对齐效果，68个稀疏的特征点可以看作是密集人脸特征点的一个采样。本实验采用归一化平均误差分别在AFLW2000-3D和AFLW-LFPA[48]上和其他方法进行比较。由于AFLW2000-3D中人脸偏航角的分布不均匀，因此为了使得偏航角的绝对值在[0°，30°]、(30°，60°]以及(60°，90°]上均匀分布，本文和文献[4]一样，随机从AFLW2000-3D选择696张人脸图片进行测试，使得各个角度的人脸图片比例为1：1：1。



(a) 输入图片

(b) 密集人脸对齐效果

图4　密集人脸对齐效果

Fig. 4　Results of dense face alignment

表1展示了本方法和其他方法在AFLW2000-3D(68个特征点)和AFLW-LFPA(34个特征点)上的比较结果。评判标准使用归一化平均误差(%)。在AFLW2000-3D中，分别计算偏航角的绝对值在[0°，30°]、(30°，60°]和(60°，90°]度的归一化平均误差以及所有图片上的归一化平均误差。在AFLW-LFPA[48]上使用所有图片上的归一化平均误差。归一化平均误差越小，人脸对齐的效果越好。相关方法的数据信息都从相关论文中直接获取，"—"表示没有相应的数据。从表中可以看出，本方法在ALFW2000-3D和AFLW-LFPA这两个数据集上都展现出了比其他方法更低的归一化平均误差，同时在不同的人脸姿态中也展现了很好的鲁棒性，这表明本方法可以很好地对大姿态的人脸图片进行密集人脸对齐。

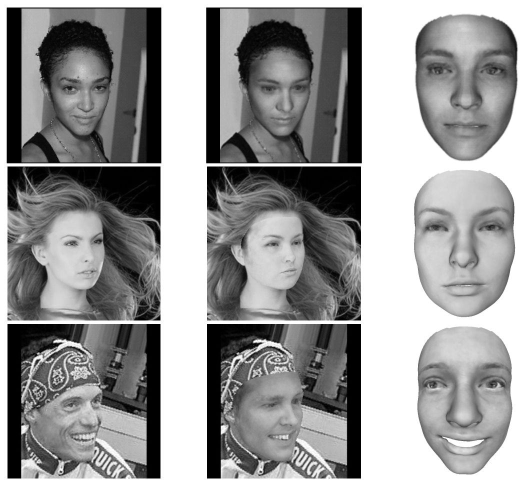
表1　密集人脸对齐归一化平均误差结果　　单位：%

Tab. 1　NME results of dense face alignment　　unit：%

方法	AFLW2000-3D				AFLW-LFPA
	[0°，30°]	(30°，60°]	(60°，90°]	平均	平均
SDM[32]	3. 67	4. 94	9. 67	6. 12	—
3DDFA[4]	3. 78	4. 54	7. 93	5. 42	—
3DDFA+SDM[4]	3. 43	4. 24	7. 17	5. 42	—
PAWF[48]	—	—	—	—	4. 72
Yu等方法[36]	3. 62	6. 06	9. 56	—	—
3DSTN[49]	3. 15	4. 33	5. 98	4. 49	—
DeFA[41]	—	—	—	4. 50	3. 86
Chang等方法[28]	3. 11	3. 84	6. 60	4. 52	—
Tran等方法[15]	—	—	—	4. 12	—
本文方法	**2. 88**	**3. 75**	**4. 77**	**3. 80**	**3. 34**

3.4　三维人脸重建结果

本实验在AFLW2000-3D上测试了所提方法的三维人脸重建效果，并将结果与3DDFA[4]以及DeFA[41]进行比较。如图5所示。首先，使用迭代最近点(Iterative Closest Point，ICP)来将预测的三维人脸和真实的三维点云标签进行对齐；然后，计算点云之间的均方误差(Mean Square Error，MSE)；随后，使用包围盒大小进行归一化，来得到归一化平均误差。表2展示了本文方法和其他方法的比较结果，评判标准使用归一化平均误差(%)，其中，归一化平均误差越小，三维人脸重建效果越好。由表中可以得出，本方法的归一化平均误差比3DDFA低0.18，比DeFA低2.08。因此本方法展示出了更好的效果。



(a) 原图　　(b) 渲染图片　　(c) 重建模型

图5　三维人脸重建效果

Fig. 5　Results of 3D face reconstruction

表2　三维人脸重建归一化平均误差结果　　单位：%

Tab. 2　NME results of 3D face reconstruction　　unit：%

方法	NME	方法	NME
3DDFA[4]	2. 43	本文方法	**2. 25**
DeFA[41]	4. 33		

3.5　消融实验

为了验证不同模块对实验效果的影响，本文分为两个部分进行消融实验：第一个部分为是否添加3DMM改进模块；第二个部分为是否在损失函数中使用三维人脸光滑性约束的损失项。

3.5.1　*是否改进3DMM*

表3展示了是否添加3DMM改进模块在AFLW2000-3D数据集上的比较结果，其中"不改进3DMM"表示不使用3DMM改进模块，"改进3DMM"表示使用3DMM改进模块，"对齐NME"表示在密集人脸对齐上的NME，"重建NME"表示在三维人脸重建上的NME。从表3中可以看出，对于密集人脸对齐，添加了3DMM改进模块之后，在AFLW2000-3D数据集上归一化平均误差降低了0.39，这说明3DMM改进模块起到了明显的效果。在三维人脸重建方面，添加了3DMM改进模块之后归一化平均误差降低了0.06。这说明改进3DMM在三维人脸重建和密集人脸对齐两方面都起到了关键的作

用,改进了3DMM之后,效果有了明显的提升。

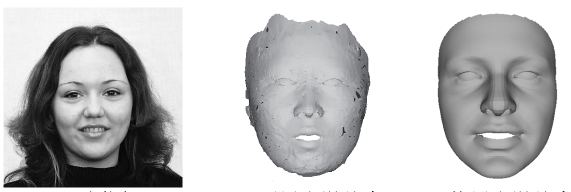
表3 是否改进3DMM的归一化平均误差结果 单位:%

Tab. 3 NME results before and after 3DMM improvement unit:%

方法	对齐NME	重建NME
不改进3DMM	4.19	2.31
改进3DMM	**3.80**	**2.25**

3.5.2 *是否使用三维人脸光滑性约束*

在损失函数中添加三维人脸光滑性约束保障了重建出的三维模型局部光滑,这会对三维人脸重建产生较大影响,而对密集人脸对齐影响较小,因此本实验主要分析是否添加光滑性约束对三维人脸重建的影响。图6展示了是否添加三维人脸光滑性约束的效果。可以看出,当未使用三维人脸光滑性约束时,重建出的三维人脸模型出现了面片翻转、表面粗糙的情况,重建的质量不佳;在添加光滑性约束后,效果有了明显提升。这说明三维人脸光滑性约束起到了至关重要的作用。



(a) 原图 (b) 不用光滑约束 (c) 使用光滑约束

图6 是否使用三维人脸光滑性约束重建效果

Fig. 6 3D face reconstruction results with and without 3D face smoothness constraint

4 结语

本文提出了一个基于弱监督学习的端到端的卷积神经网络模型,用于从单张人脸图片进行三维人脸重建和密集人脸对齐。通过将现有的3DMM模型作为卷积神经网络全连接层的一部分,本文对现有的3DMM模型进行改进,有效地重建出更精准的三维人脸模型,提高了模型的泛化能力。本文设计了一个新的损失函数,通过对学习到的三维人脸模型加入光滑性约束,并且使用输入人脸图片和渲染后图片的SSIM作为损失,有效地提升了重建质量。同时本文的方法在不同的人脸姿态中都展现了很好的鲁棒性,可以对大姿态的人脸图片有效地进行三维人脸重建和密集人脸对齐。在AFLW2000-3D和AFLW-LFPA数据集上的实验结果表明,本文的方法在三维人脸重建和密集人脸对齐方面均优于以前的方法。在未来工作中,将继续改进三维人脸模型以重建出更详细的面部信息,例如胡须、痣和皱纹等。

参考文献(References)

[1] KEMELMACHER-SHLIZERMAN I, BASRI R. 3D face reconstruction from a single image using a single reference face shape[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(2): 394-405.

[2] FYFFE G, JONES A, ALEXANDER O, et al. Driving high-resolution facial scans with video performance capture[J]. ACM Transactions on Graphics, 2014, 34(1): No. 8.

[3] 邓秋平,赵宇明. 基于单幅正面照片的三维人脸重建方法[J]. 计算机工程, 2010, 36(20): 176-178.(DENG Q P, ZHAO Y M. 3D face reconstruction method based on single frontal photo[J]. Computer Engineering, 2010, 36(20): 176-178.)

[4] ZHU X, LEI Z, LIU X, et al. Face alignment across large poses: a 3D solution[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 146-155.

[5] TEWARI A, ZOLLHÖFER M, KIM H, et al. MoFA: model-based deep convolutional face autoencoder for unsupervised monocular reconstruction[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3735-3744.

[6] 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, 1999: 187-194.

[7] RICHARDSON E, SELA M, KIMMEL R. 3D face reconstruction by learning from synthetic data[C]// Proceedings of the 4th International Conference on 3D Vision. Piscataway: IEEE, 2016: 460-469.

[8] LIU F, ZHU R, ZENG D, et al. Disentangling features in 3D face shapes for joint face reconstruction and recognition[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 5216-5225.

[9] TRAN A T, HASSNER T, MASI I, et al. Regressing robust and discriminative 3D morphable models with a very deep neural network[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5163-5172.

[10] TEWARI A, ZOLLHÖFER M, GARRIDO P, et al. Self-supervised multi-level face model learning for monocular reconstruction at over 250 Hz[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2549-2559.

[11] GENOVA K, COLE F, MASCHINOT A, et al. Unsupervised training for 3D morphable model regression[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8377-8386.

[12] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681-685.

[13] GERIG T, MOREL-FORSTER A, BLUMER C, et al. Morphable face models — an open framework[C]// Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition. Piscataway: IEEE, 2018: 75-82.

[14] BOOTH J, ROUSSOS A, ZAFEIRIOU S, et al. A 3D morphable model learnt from 10000 faces[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 5543-5552.

[15] TRAN L, LIU X. On learning 3D face morphable model from in-the-wild images[EB/OL]. [2020-04-02]. https://arxiv. org/pdf/1808. 09560v1. pdf.

[16] ROTH J, TONG Y, LIU X. Adaptive 3D face reconstruction from unconstrained photo collections[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4197-4206.

[17] JIN X, TAN X. Face alignment in-the-wild: a survey[J]. Computer Vision and Image Understanding, 2017, 162: 1-22.

[18] WANG N, GAO X, TAO D, et al. Facial feature point detection: a comprehensive survey[J]. Neurocomputing, 2018, 275: 50-65.

[19] ZHU X, YAN J, YI D, et al. Discriminative 3D morphable model fitting[C]// Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Piscataway: IEEE, 2015: 1-8.

[20] JENI L A, COHN J F, KANADE T. Dense 3D face alignment from 2D videos in real-time[C]// Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and

Gesture Recognition. Piscataway: IEEE, 2015: 1-8.

[21] 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 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1031-1039.

[22] JIANG L, ZHANG J, DENG B, et al. 3D face reconstruction with geometry details from a single image [J]. IEEE Transactions on Image Processing, 2018, 27(10): 4756-4770.

[23] GUO Y, ZHANG J CAI J, et al. CNN-based real-time dense face reconstruction with inverse-rendered photo-realistic face images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(6): 1294-1307.

[24] DENG Y, YANG J, XU S, et al. Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set [EB/OL]. [2020-04-02]. https://arxiv. org/pdf/1903. 08527. pdf.

[25] RICHARDSON E, SELA M, OR-EL R, et al. Learning detailed face reconstruction from a single image [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1259-1268.

[26] FENG Y, WU F, SHAO X, et al. Joint 3D face reconstruction and dense alignment with position map regression network [C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11218. Cham: Springer, 2018: 557-574.

[27] SHI T, YUAN Y, FAN C, et al. Face-to-parameter translation for game character auto-creation [C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 161-170.

[28] CHANG F, TRAN A T, HASSNER T, et al. Deep, landmark-free FAME: face alignment, modeling, and expression estimation [J]. International Journal of Computer Vision, 2019, 127(6/7): 930-956.

[29] ASTHANA A, ZAFEIRIOU S, CHENG S, et al. Robust discriminative response map fitting with constrained local models [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3444-3451.

[30] LIANG Z, DING S, LIN L. Unconstrained facial landmark localization with backbone-branches fully-convolutional networks [EB/OL]. [2020-04-02]. https://arxiv. org/abs/1507. 03409.

[31] PENG X, FERIS R S, WANG X, et al. A recurrent encoder-decoder network for sequential face alignment [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9905. Cham: Springer, 2016: 38-56.

[32] MCDONAGH J, TZIMIROPOULOS G. Joint face detection and alignment with a deformable Hough transform model [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 569-580.

[33] GOU C, WU Y, WANG F, et al. Shape augmented regression for 3D face alignment [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 604-615.

[34] SÁNTA Z, KATO Z. 3D face alignment without correspondences [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 521-535.

[35] DE BITTENCOURT ZAVAN F H, NASCIMENTO A C P, E SILVA L P, et al. 3D face alignment in the wild: a landmark-free, nose-based approach [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 581-589.

[36] YU R, SAITO S, LI H, et al. Learning dense facial correspondences in unconstrained images [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 4733-4742.

[37] BULAT A, TZIMIROPOULOS G. Two-stage convolutional part heatmap regression for the 1st 3D face alignment in the wild (3DFAW) challenge [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 616-624.

[38] BULAT A, TZIMIROPOULOS G. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230000 3D facial landmarks) [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1021-1030.

[39] JOURABLOO A, LIU X. Pose-invariant 3D face alignment [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 3694-3702.

[40] CAO C, HOU Q, ZHOU K. Displaced dynamic expression regression for real-time facial tracking and animation [J]. ACM Transactions on Graphics, 2014, 33(4): No. 43.

[41] LIU Y, JOURABLOO A, REN W, et al. Dense face alignment [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2017: 1619-1628.

[42] GÜLER R A, TRIGEORGIS G, ANTONAKOS E, et al. DenseReG: fully convolutional dense shape regression in-the-wild [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6799-6808.

[43] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-04-02]. https://arxiv. org/pdf/1409. 1556v1. pdf.

[44] CAO C, WENG Y, 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.

[45] RAMAMOORTHI R, HANRAHAN P. A signal-processing framework for inverse rendering [C]// Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 2001: 117-128.

[46] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 3730-3738.

[47] KINGMA D P, BA J L. Adam: a method for stochastic optimization [EB/OL]. [2020-04-02]. https://arxiv. org/pdf/1412. 6980. pdf.

[48] JOURABLOO A, LIU X. Large-pose face alignment via CNN-based dense 3D model fitting [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4188-4196.

[49] BHAGAVATULA C, ZHU C, LUU K, et al. Faster than real-time facial alignment: a 3D spatial transformer network approach in unconstrained poses [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 4000-4009.

This work is partially supported by the National Natural Science Foundation of China (61877056), the Fundamental Research Funds for the Central Universities (WK6030000109).

ZHOU Jian, born in 1995, M. S. candidate. His research interests include computer vision, computer graphics, 3D face reconstruction.

HUANG Zhangjin, born in 1980, Ph. D., associate professor. His research interests include computer vision, machine learning, computer graphics.