



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2021

04

VOL.26

ISSN1006-8961
CN11-3758/TB



细粒度图像分类 P847



第26卷第4期 (总第300期)

2021年4月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿, 均视为同意在本刊网站及CNKI等全文数据库出版, 所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用, 所有内容, 未经许可, 不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院空天信息创新研究院

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 吴一戎

编辑出版 《中国图象图形学报》编辑部

通信地址 北京市海淀区北四环西路19号

邮 编 100190

电子信箱 jig@radi.ac.cn

电 话 010-58887035

网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

(邮政信箱: 北京399信箱 邮编: 100048)

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Aerospace Information Research Institute, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong

Editor, Publisher Editorial Office of Journal of
Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District,
Beijing, P. R. China

Zip code 100190

E-mail jig@radi.ac.cn

Telephone 010-58887035

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

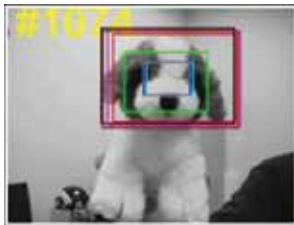
国外发行代号 M1406

国内邮发代号 82-831

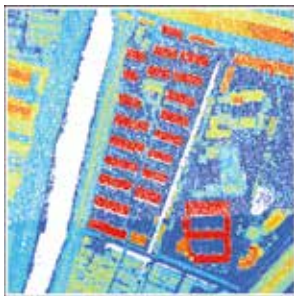
国内定价 60.00元



YOLOv3剪枝模型的多人姿态估计(第0837页)



融入时序和速度信息的自适应更新目标跟踪(第0883页)



可变半径Alpha Shapes提取机载LiDAR点云建筑物轮廓(第0910页)

综述

从图像到语言: 图像标题生成与描述

谭云兰, 汤鹏杰, 张丽, 罗玉盘 0727

CT影像肺结节分割研究进展

董婷, 魏琰, 聂生东 0751

图像处理和编码

全局注意力门控残差记忆网络的图像超分重建

王静, 宋慧慧, 张开华, 刘青山 0766

多层次感知残差卷积网络的单幅图像超分重建

何蕾, 程佳豪, 占志钰, 杨雯博, 刘沛然 0776

采用双尺度图像分解的水下彩色图像增强

李健, 张显斗, 李熹飞, 吴子朝 0787

图像分析和识别

姿态特征结合2维傅里叶变换的步态识别

王新年, 胡丹丹, 张涛, 白桂欣 0796

复杂背景下的手势识别

王银, 陈云龙, 孙前来 0815

结合形变模型与图像修复的人脸姿态矫正

吴从中, 郑荣生, 臧怀娟, 刘明威, 徐甲甲, 詹曙 0828

YOLOv3剪枝模型的多人姿态估计

蔡哲栋, 应娜, 郭春生, 郭锐, 杨鹏 0837

YOLOv3和双线性特征融合的细粒度图像分类

闫子旭, 侯志强, 熊磊, 刘晓义, 余旺盛, 马素刚 0847

深度多模态融合服装风格检索

苏卓, 柯司博, 王若梅, 周凡 0857

超像素条件随机场下的RGB-D视频显著性检测

李贝, 杨铀, 刘琼 0872

融入时序和速度信息的自适应更新目标跟踪

尹宽, 李均利, 胡凯, 李丽 0883

融合逆密度函数与关系形状卷积神经网络的点云分析

邱云飞, 朱梦影 0898

图像理解和计算机视觉

可变半径Alpha Shapes提取机载LiDAR点云建筑物轮廓

伍阳, 王丽妍, 胡春霞, 程亮 0910

3D遮挡模型引导的光场图像深度获取

吴迪, 张旭东, 张骏, 范之国, 孙锐 0924

计算机图形学

基于全正基的非均匀三次加权B样条

姬佩佩, 张贵仓, 汪凯, 孟建军 0939

遥感图像处理

LBP特征分类的极化SAR图像机场跑道检测

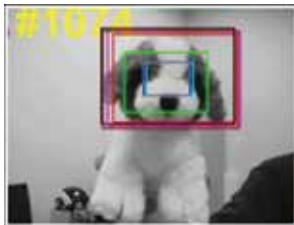
韩萍, 万义爽, 刘亚芳, 韩宾宾 0952

CONTENTS

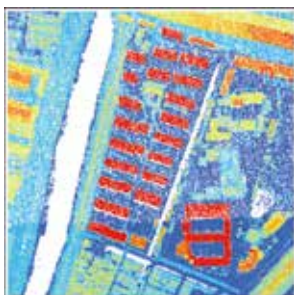
JOURNAL OF IMAGE AND GRAPHICS



Research on multiperson pose estimation combined with YOLOv3 pruning model(P0837)



Adaptive update object tracking algorithm incorporating timing and speed information(P0883)



Extraction of building contours from airborne LiDAR point cloud using variable radius Alpha Shapes method(P0910)

Review

From image to language: image captioning and description

Tan Yunlan, Tang Pengjie, Zhang Li, Luo Yupan 0727

Research progress of lung nodule segmentation based on CT images

Dong Ting, Wei Long, Nie Shengdong 0751

Image Processing and Coding

Learning global attention-gated multi-scale memory residual networks for single-image super-resolution

Wang Jing, Song Huihui, Zhang Kaihua, Liu Qingshan 0766

Single image super-resolution reconstruction based on multi-level perceptual residual convolutional network

He Lei, Cheng Jiahao, Zhan Zhiyu, Yang Wenbo, Liu Peiran 0776

Underwater color image enhancement based on two-scale image decomposition

Li Jian, Zhang Xiandou, Li Shangfei, Wu Zizhao 0787

Image Analysis and Recognition

Gait recognition using pose features and 2D Fourier transform

Wang Xinnian, Hu Dandan, Zhang Tao, Bai Guixin 0796

Hand gesture recognition in complex background

Wang Yin, Chen Yunlong, Sun Qianlai 0815

Face pose correction based on morphable model and image inpainting

Wu Congzhong, Zheng Rongsheng, Zang Huaijuan, Liu Mingwei, Xu Jiajia, Zhan Shu 0828

Research on multiperson pose estimation combined with YOLOv3 pruning model

Cai Zhedong, Ying Na, Guo Chunsheng, Guo Rui, Yang Peng 0837

Fine-grained classification based on bilinear feature fusion and YOLOv3

Yan Zixu, Hou Zhiqiang, Xiong Lei, Liu Xiaoyi, Yu Wangsheng, Ma Sugang 0847

Fashion style retrieval based on deep multimodal fusion

Su Zhuo, Ke Sibao, Wang Ruomei, Zhou Fan 0857

RGB-D video saliency detection via superpixel-level conditional random field

Li Bei, Yang You, Liu Qiong 0872

Adaptive update object tracking algorithm incorporating timing and speed information

Yin Kuan, Li Junli, Hu Kai, Li Li 0883

Point cloud analysis model combining inverse density function and relation-shape convolution neural network

Qiu Yunfei, Zhu Mengying 0898

Image Understanding and Computer Vision

Extraction of building contours from airborne LiDAR point cloud using variable radius Alpha Shapes method

Wu Yang, Wang Liyan, Hu Chunxia, Cheng Liang 0910

Light field depth estimation guided by 3D occlusion model

Wu Di, Zhang Xudong, Zhang Jun, Fan Zhiguo, Sun Rui 0924

Computer Graphics

Non uniform weighted cubic B-spline based on all positive basis

Ji Peipei, Zhang Guicang, Wang Kai, Meng Jianjun 0939

Remote Sensing Image Processing

Airport runway detection based on LBP feature classification in PolSAR images

Han Ping, Wan Yishuang, Liu Yafang, Han Binbin 0952

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)04-0828-09

论文引用格式: Wu C Z, Zheng R S, Zang H J, Liu M W, Xu J J and Zhan S. 2021. Face pose correction based on morphable model and image inpainting. Journal of Image and Graphics, 26(04): 0828-0836 (吴从中, 郑荣生, 臧怀娟, 刘明威, 徐甲甲, 詹曙. 2021. 结合形变模型与图像修复的人脸姿态矫正. 中国图象图形学报, 26(04): 0828-0836) [DOI:10.11834/jig.200011]

结合形变模型与图像修复的人脸姿态矫正

吴从中¹, 郑荣生¹, 臧怀娟¹, 刘明威¹, 徐甲甲², 詹曙^{1*}

1. 合肥工业大学计算机与信息学院, 合肥 230009; 2. 科大讯飞股份有限公司, 合肥 230009

摘要: **目的** 人脸姿态偏转是影响人脸识别准确率的一个重要因素, 本文利用3维人脸重建中常用的3维形变模型以及深度卷积神经网络, 提出一种用于多姿态人脸识别的人脸姿态矫正算法, 在一定程度上提高了大姿态下人脸识别的准确率。 **方法** 对传统的3维形变模型拟合方法进行改进, 利用人脸形状参数和表情参数对3维形变模型进行建模, 针对面部不同区域的关键点赋予不同的权值, 加权拟合3维形变模型, 使得具有不同姿态和面部表情的人脸图像拟合效果更好。然后, 对3维人脸模型进行姿态矫正并利用深度学习对人脸图像进行修复, 修复不规则的人脸空洞区域, 并使用最新的局部卷积技术同时在新的数据集上重新训练卷积神经网络, 使得网络参数达到最优。 **结果** 在LFW (labeled faces in the wild) 人脸数据库和StirlingESRC (Economic Social Research Council) 3维人脸数据库上, 将本文算法与其他方法进行比较, 实验结果表明, 本文算法的人脸识别精度有一定程度的提高。在LFW数据库上, 通过对具有任意姿态的人脸图像进行姿态矫正和修复后, 本文方法达到了96.57%的人脸识别精确度。在StirlingESRC数据库上, 本文方法在人脸姿态为 $\pm 22^\circ$ 的情况下, 人脸识别准确率分别提高5.195%和2.265%; 在人脸姿态为 $\pm 45^\circ$ 情况下, 人脸识别准确率分别提高5.875%和11.095%; 平均人脸识别率分别提高5.53%和7.13%。对比实验结果表明, 本文提出的人脸姿态矫正算法有效提高了人脸识别的准确率。 **结论** 本文提出的人脸姿态矫正算法, 综合了3维形变模型和深度学习模型的优点, 在各个人脸姿态角度下, 均能使人脸识别准确率在一定程度上有所提高。

关键词: 多姿态人脸识别; 3维形变模型(3DMM); 卷积神经网络(CNN); 图像修复; 深度学习

Face pose correction based on morphable model and image inpainting

Wu Congzhong¹, Zheng Rongsheng¹, Zang Huaijuan¹, Liu Mingwei¹, Xu Jiajia², Zhan Shu^{1*}

1. School of Computer and Information, Hefei University of Technology, Hefei 230009, China;

2. iFLYTEK Co. Ltd., Hefei 230009, China

Abstract: **Objective** Face recognition has been a widely studied topic in the field of computer vision for a long time. In the past few decades, great progress in face recognition has been achieved due to the capacity and wide application of convolutional neural networks. However, pose variations still remain a great challenge and warrant further studies. To the best of our knowledge, the existing methods that address this problem can be generally categorized into two classes: feature-based methods and deep learning-based methods. Feature-based methods attempt to obtain pose-invariant representations directly from non-frontal faces or design handcrafted local feature descriptors, which are robust to face poses. However, it is often too difficult to obtain robust representation of the face pose using these handcrafted local feature descriptors. Thus, these

收稿日期: 2020-02-15; 修回日期: 2020-06-04; 预印本日期: 2020-06-11

* 通信作者: 詹曙 shu_zhan@hfut.edu.cn

基金项目: 国家自然科学基金项目(61371156)

Supported by: National Natural Science Foundation of China (61371156)

methods cannot produce satisfactory results, especially when the face pose is too large. In recent years, convolutional neural networks have been introduced in face recognition problems due to their outstanding performance in image classification tasks. Different from traditional methods, convolutional neural networks do not require the manual extraction of local feature descriptors. They try to directly rotate the face image of arbitrary pose and illuminate into the target pose, which maintains the face identity feature well. In addition, due to the powerful ability of image generation, generative adversarial network is also used in frontal face image synthesis and has achieved great progress. Compared with traditional methods, deep learning-based methods can obtain a higher face recognition rate. However, the disadvantage of deep learning-based methods is that the face images synthesized from the large face pose have low credibility, which lead to poor face recognition accuracy. To deal with the limitations of these two kinds of methods, we present a face pose correction algorithm based on 3D morphable model (3DMM) and image inpainting. **Method** In this study, we propose a face frontalization method by combining deep learning model and a 3DMM, which can generate a photorealistic frontal view of the face image. In detail, we first detect facial landmarks by using a well-known facial landmark detector, which is robust to large pose variations. We detect a total of 68 facial landmarks to fit the face image more accurately. Then, we perform accurate 3DMM fitting for face image with facial landmark weighting. Next, we estimate the depth information of the face image and rotate the 3D face model into frontal view using 3D transformation. Finally, we employ image inpainting for the irregular facial invisible region caused by self-occlusion by utilizing deep learning model. We fine-tune the pre-trained model to train our image inpainting model. In the training process, all of the convolutional layers are replaced with partial convolutional layers. Our training set consists of 13 223 face images that are selected from the labeled faces in the wild (LFW) dataset. Our image inpainting network is implemented in Keras. The batch size is set to 4, the learning rate is set to 10^{-4} , and the weight decay is 0.000 5. The network training procedure is accelerated using NVIDIA GTX 1080 Ti GPU devices, which takes approximately 10 days in total. **Result** We compare our method with state-of-the-art methods, including the traditional method and deep learning method, on two public face datasets, namely, LFW dataset and StirlingESRC 3D face dataset. The quantitative evaluation metric is face recognition rate under different face poses, and we provide several synthesized frontal face images by our method. The synthesized frontal face images show that our method can produce more photorealistic results than other methods in the LFW dataset. We achieve 96.57% face recognition accuracy on the LFW face dataset. In addition, the quantitative experiment results show that our method outperforms all other methods in StirlingESRC 3D face dataset. The experimental results show that the face recognition accuracy of our method is improved under different face poses. Compared with the other two methods in the StirlingESRC 3D face dataset, the face recognition accuracy increased by 5.195% and 2.265% under the face pose of 22° and by 5.875% and 11.095% under the face pose of 45° , respectively. Moreover, the average face recognition rate increased by 5.53% and 7.13%, respectively. The experimental results show that the proposed multi-pose face recognition algorithm improves the accuracy of face recognition. **Conclusion** In this study, we propose a face pose correction algorithm for multi-pose face recognition by combining 3DMM with deep learning model. The qualitative and quantitative experiment results show that our method can synthesize a more photorealistic frontal face image than other methods and can improve the accuracy performance of multi-pose face recognition.

Key words: multi-pose face recognition; 3D morphable model (3DMM); convolutional neural network (CNN); image inpainting; deep learning

0 引言

人脸识别一直是计算机视觉领域的研究热点。得益于卷积神经网络强大的计算能力和广泛应用, 人脸识别取得了很大进展(Wang等, 2018; Deng等, 2018)。然而受姿态变化的影响, 任意姿态条件下的人脸识别仍然是一项非常具有挑战性的工作。对

此, 现有的解决方法大致分为特征级正面化和图像级正面化两类。前者直接从非正面人脸中提取对姿态变化鲁棒的特征表示。例如, Schroff等人(2015)使用联合模型来提取面部特征, Ding和Tao(2015)以及Masi等人(2016)使用多个基于特定姿态的模型来获得鲁棒的特征表示。后者基于深度学习的方法(Zhu等, 2014; Sagonas等, 2015; Zhu等, 2013; Cole等, 2017)将侧面人脸图像旋转为正面人脸图

像。例如, Yim 等人(2015)使用多任务的卷积神经网络,在保证人脸身份特征不变的前提下,将任意姿态和光照条件下的人脸图像旋转为目标姿态的人脸图像。此外,生成对抗网络(generative adversarial network, GAN)的迅速发展和成功应用推动了基于 GAN 的正面人脸合成方法的研究。例如, Tran 等人(2017a)提出通过 DR-GAN(disentangled representation GAN)来学习一种生成式表示和判别式表示,用于姿态不变的人脸识别。但这些方法都无法合成逼真的正面人脸图像,导致人脸识别准确率较低。

针对大姿态下的人脸矫正和人脸识别,本文提出一种结合 3 维形变模型和图像修复技术的方法。3 维形变模型是用于 3 维人脸重建的经典方法,本文将 3 维形变模型用于多姿态人脸的拟合,并只使用人脸形状参数和表情参数的线性组合表示 3 维形变模型。本文的主要工作如下:1)通过更新 3 维人

脸模型中轮廓处的关键点使得 3 维形变模型可以拟合较大姿态下的人脸图像。另外考虑到人脸具有不同的表情变化,在拟合 3 维形变模型的过程中,对人脸不同区域的关键点赋予不同的权值,使得人脸拟合对表情变化更鲁棒。2)将基于局部卷积层的图像修复网络(Liu 等,2018)用于修复带有不规则空洞区域的人脸,并在本文带空洞区域的人脸图像和掩膜数据集中重新训练网络,更新网络参数和损失函数加权参数,使得修复后的人脸图像更接近于真实正面人脸图像,识别率更高。尽管深度学习在图像修复领域取得了很大进展(Hong 等,2019; Yeh 等,2017),但大都是修复规则空洞区域的图像,不适用于本文修复带有不规则空洞区域的人脸图像。定量和定性实验表明,本文方法合成的正面人脸图像在视觉上更逼真,并且提高了多姿态人脸识别的准确率。本文方法的框架如图 1 所示。

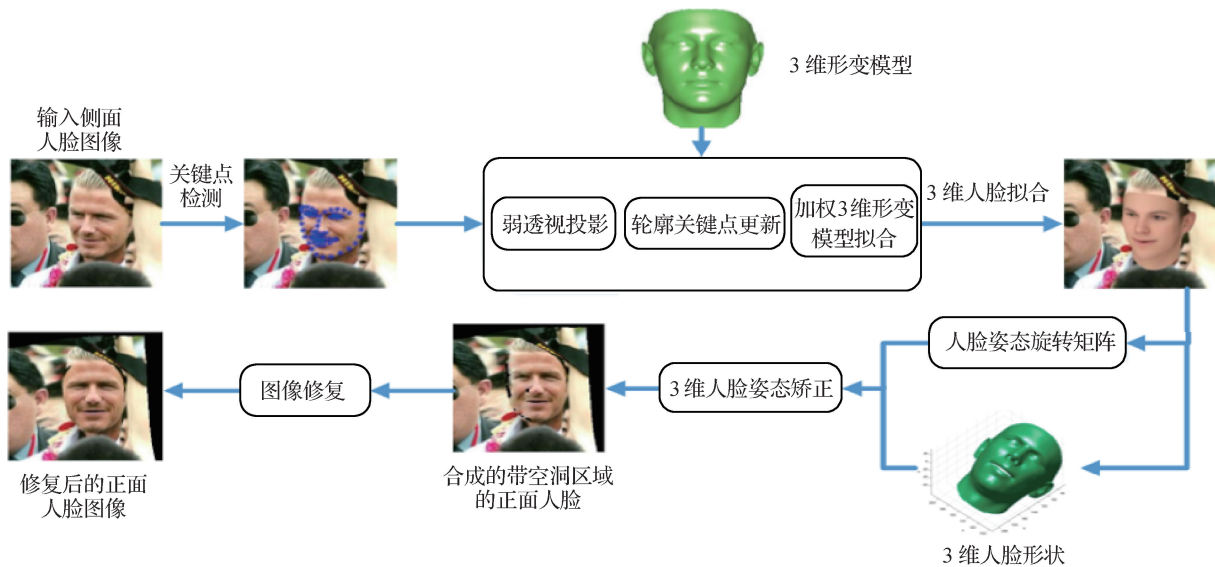


图 1 结合 3 维形变模型和图像修复的正面人脸合成系统框架图

Fig. 1 The structure of frontal face synthesis system combining 3D morphable model with image inpainting

1 本文算法

1.1 3 维形变模型

Blanz 和 Vetter(1999)提出使用 3 维形变模型(3D morphable model, 3DMM)来生成 3 维人脸模型,并广泛用于人脸识别(Blanz 和 Vetter, 2003; Heisele 等, 2007)和人脸重建(Tran 等, 2017b)。3DMM 通常表示为人脸的形状参数和纹理参数的线性组合,具体为

$$\begin{aligned} S &= \bar{S} + \lambda_{id} S_{id} \\ T &= \bar{T} + \lambda_{tex} S_{tex} \end{aligned} \quad (1)$$

式中, $\bar{S}$ 和 $\bar{T}$ 分别代表平均人脸形状和平均人脸纹理, S_{id} 和 S_{tex} 分别代表人脸形状特征向量和人脸纹理特征向量, λ_{id} 是人脸形状权重系数, λ_{tex} 是人脸纹理权重系数。本文使用人脸形状参数和表情参数的线性组合来表示 3DMM, 即

$$S = \bar{S} + \sum_{i=1}^m \lambda_{id}^i S_{id}^i + \sum_{i=1}^n \lambda_{exp}^i S_{exp}^i \quad (2)$$

式中, S_{exp} 代表人脸表情的特征向量, λ_{exp} 代表人脸

表情权重系数。人脸形状特征向量 S_{id} 来自 BFM (basel face model) (Paysan 等, 2009), 人脸表情特征向量 S_{exp} 来自 Face Warehouse (Cao 等, 2014)。本文将这两种常用的人脸模型与非刚性的迭代最近点 (iterative closest point, ICP) 算法 (Amberg 等, 2007) 相结合, 用于表示 3 维人脸模型。

对于给定的人脸图像, 本文通过弱透视投影来拟合 3DMM, 即

$$S_{2d} = f \cdot P \cdot R(\alpha, \beta, \chi) S(:, X) + t \quad (3)$$

式中, S_{2d} 是 3 维坐标点在图像平面上的 2 维投影, f 是尺度参数, P 是正交投影 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, R 是由人脸在 3 维空间中的旋转角度俯仰角度 (α)、翻滚角度 (β) 和偏航角度 (χ) 组成的 3×3 的旋转矩阵, X 是索引向量, 代表 3 维关键点与 2 维关键点之间的对应关系, t 是平移向量。假设所有参数组成一个向量, 记作 $v = (f, R, X, t)$ 。拟合过程是搜索 3 维坐标点的真实 2 维坐标点 S_{2d} , 并通过最小化 S_{2d} 与 S_{2d} 之间的距离来估计模型参数 $v = (f, R, X, t)$, 即

$$\underset{f, R, X, t, \lambda_{id}, \lambda_{exp}}{\operatorname{argmin}} \|f \cdot P \cdot R(\alpha, \beta, \chi) S(:, X) + t - S_{2d}\| \quad (4)$$

1.2 基于关键点加权的 3DMM 拟合

人脸关键点检测是对 3 维形变模型进行拟合的重要步骤之一。由于头部姿态和面部表情的变化, 人脸关键点检测一直是具有挑战性的问题。大姿态下人脸关键点检测的难点是图像中存在不可见的面部区域。本文采用 CE-CLM (convolutional experts-constrained local model) 算法 (Zadeh 等, 2017) 对人脸关键点进行检测, 该算法对头部旋转角度较大的人脸图像具有很好的鲁棒性, 关键点检测精度高。本文共检测 68 个面部关键点。

对给定的正面视图的人脸图像, 使用式 (3) 来拟合 3 维形变模型。然而当人脸姿态偏转时, 原来在 3 维人脸模型轮廓处的 3 维关键点与 2 维人脸图像轮廓处的 2 维关键点就会失去对应关系, 导致 3 维形变模型拟合不准确。为了保留式 (3) 中 2D 到 3D 关键点的对应关系, 需要重新检测人脸轮廓处的关键点。考虑到姿态偏转较大时轮廓处关键点与鼻子、嘴巴处关键点有重合现象, 本文只对姿态偏转角度小于等于 60° 的人脸模型进行关键点更新。

首先, 利用检测到的 α 和 β 角度在平面内对

3 维人脸模型进行矫正, 即

$$S_{new} = R(\alpha, \beta, 0) \cdot S \quad (5)$$

式中, S_{new} 是姿态矫正后的 3 维人脸模型。

其次, 在人脸模型上定义 8 组路径 (包含轮廓关键点), 在每条路径上寻找最大或最小的横坐标值作为新的人脸左 (右) 轮廓处的关键点。显然, 每条路径两端分别有最大和最小的两个横坐标, 由此构成了人脸轮廓上的 16 个关键点。3 维人脸轮廓处关键点的更新算法具体步骤如下:

输入: 3 维人脸模型 S 和估计参数 v 。

输出: 索引向量 X 。

1) 旋转 3 维人脸模型

$$S_{new} = R(\alpha, \beta, 0) \cdot S$$

2) if $0^\circ < \beta < 60^\circ$ then

for i in range (1, 4) do

$V_{cheek}(i) = \operatorname{argmax}(S_{new}(1, path_{cheek}(i)))$ // 遍历变量 i , 寻找轮廓处关键点。

3) if $-60^\circ < \beta < 0^\circ$ then

for i in range (5, 8) do

$$V_{cheek}(i) = \operatorname{argmax}(S_{new}(1, path_{cheek}(i)))$$

4) 更新索引向量 X 中的 8 个元素值。

考虑到鼻子以下的人脸区域上的关键点位置受表情影响较大, 在进行 3 维形变模型拟合时, 定义了一个权重矩阵 ω , 对位于人脸不同区域的关键点赋予不同的权值, 以此增强对表情变化的拟合精度, 具体为

$$\underset{f, \omega, R, X, t, \lambda_{id}, \lambda_{exp}}{\operatorname{argmin}} \|\omega(f \cdot P \cdot R(\alpha, \beta, \chi) S(:, X) + t - S_{2d})\| \quad (6)$$

式中, ω 是一个对角矩阵, ω_i 表示第 i 个 3 维关键点和第 i 个 2 维关键点之间的权重。定义向量 $U = f \cdot R(\alpha, \beta, \chi) P \cdot S(:, X) + t - S_{2d}$, 本文假设 U_i 和 ω_i 存在正比例关系, 具体为

$$\omega_i = x_1 \times U_i + x_2 \quad (7)$$

式中, 常数 x_1, x_2 是为了确保避免权重系数过大或过小。基于关键点加权的 3DMM 拟合的权重更新算法的具体步骤如下:

输入: 输入图像的 2 维关键点 S_{2d} 。

输出: 拟合后的 3 维人脸模型 S 。

While ($D < \text{阈值 } d$) do

1) 估计尺度参数 f , 角度参数 R , 平移参数 t ;

2) 估计形状参数 λ_{id} 和表情参数 λ_{exp} ;

3) 计算 3 维形变模型拟合误差 D , 即

$$\operatorname{argmin}_{f, \omega, R, t, \lambda_{\text{id}}, \lambda_{\text{exp}}} \|\omega(f \cdot R(\alpha, \beta, \chi)P \cdot S(:, X) + t - S_{2d})\|$$

4) 通过式(7)更新权重参数 ω 。

得到拟合的 3 维人脸模型后, 先估计 3 维人脸模型的深度信息, 并利用 delaunay 算法对人脸模型进行三角剖分, 得到 3 维网格化的人脸头部, 再用逆旋转矩阵 R^{-1} 对 3 维人脸进行姿态矫正, 具体为

$$I_{\text{syn}} = R^{-1}I_{\text{mesh}} \quad (8)$$

式中, I_{mesh} 是 3 维网状人脸, I_{syn} 是合成的正面 3 维网状人脸。通过姿态矫正后, 得到的正面人脸图像中, 脸部存在不规则的空洞区域, 这是由原始人脸图像的自遮挡所导致的。

1.3 人脸图像修复网络

得到带空洞区域的人脸图像后, 需要对人脸图像进行修复, 修复算法包括基于传统方法的图像修复算法和基于深度学习的图像修复算法。

基于传统方法的图像修复算法取得的人脸图像从视觉上看一般都不真实, 与原始的正面人脸图像相差较大, 无法直接用于人脸识别系统。例如, Ding 等人(2012)根据人脸面部对称, 使用人脸空洞区域的对称区域像素来修复空洞区域。Asthana 等人(2011)不使用任何图像修复算法来修复空洞区域。因此, 使用这类方法得到的人脸图像会在一定程度上降低人脸识别的精度。

基于深度学习的图像修复算法的主要原理是在语义层次上搜索与空洞区域的像素最接近的图像像素来对空洞区域的图像进行修复。深度学习通过多层网络模型提取看不到的抽象高层语义图像特征, 自动对图像特征进行学习, 从而取得最佳的网络模型。基于深度学习的图像修复技术取得了很好效果, 但不足之处是修复的区域限制是规则区域(例如: 正方形)。而本文需要修复的是脸部不规则的空洞区域, 难度比修复规则空洞区域高。Liu 等人(2018)提出的图像修复方法达到了当时的前沿水平, 与以往图像修复技术不同, 该方法制作了掩膜数据集, 提出了掩膜更新机制和局部卷积层技术, 可以修复图像中任意形状的空洞区域。网络的输入是一幅带不规则空洞区域的图像和一幅掩膜图像, 输出是一幅修复后的图像。与传统神经网络不同, 网络中所有的卷积层都用局部卷积层代替, 即将输入图像中像素点不为 0 的输入到网络中。局部卷积层定

义为

$$x' = \begin{cases} W^T(X \odot M) \frac{\text{sum}(1)}{\text{sum}(M)} + b & \text{sum}(M) > 0 \\ 0 & \text{其他} \end{cases} \quad (9)$$

式中, W 是卷积核的权重, b 是相应的偏差, X 是当前卷积滑动窗口的像素值, M 为相应的二进制掩膜, $\odot$ 是卷积, sum 是求和。网络结构与 U-Net(Ronneberger 等, 2015) 类似, 是一个编码器—解码器结构。编码器经过一系列卷积操作将图像进行编码, 再经过一系列反卷积, 恢复到输入图像大小。网络中卷积核尺寸分别为 7×7 , 5×5 和 3×3 , 网络编码模块激活函数采用 ReLU, 网络解码模块激活函数采用 LeakyReLU。本文使用该网络结构在带空洞区域人脸数据集中重新训练网络参数, 使网络参数达到最优。网络输出就是一幅修复后的正面人脸图像, 并且可以直接用于后续人脸识别任务。

2 实验

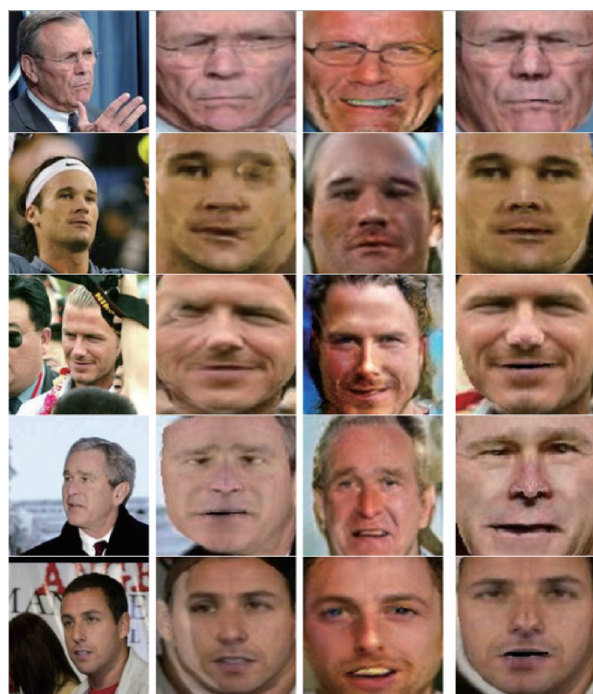
2.1 实验数据集

LFW(labeled faces in the wild) 数据库是应用最广、无约束环境下的人脸识别数据库, 共包含 5 729 个人的 13 223 幅人脸图像, 这些人脸图像具有任意的姿态角度、遮挡情况和表情变化等。在 LFW 数据库上进行多姿态人脸识别具有很大的挑战。

StirlingESRC 数据库是 3 维人脸数据库, 数据库中的每个个体都包含极其丰富和多样化的人脸图像, 主要用于人脸感知和识别的研究。为了验证本文提出的正面人脸合成方法在不同人脸姿态下的有效性, 本文从 StirlingESRC 数据库中选择 4 个不同人脸姿态($\pm 22^\circ$ 和 $\pm 45^\circ$) 的图像进行实验。

2.2 实验结果及分析

通过将不同姿态下的人脸图像合成为一幅正面的人脸图像, 再对人脸图像进行修复, 最后与对比方法进行定性和定量比较, 验证本文方法的有效性。图 2 和图 3 分别展示了本文方法在 LFW 和 StirlingESRC 数据库上与 Hassner 等人(2015) 的算法以及 Tran 等人(2017a) 的算法的定性实验比较。表 1 显示了在 StirlingESRC 数据库上, 本文方法与两种对比算法在人脸识别率上的对比。

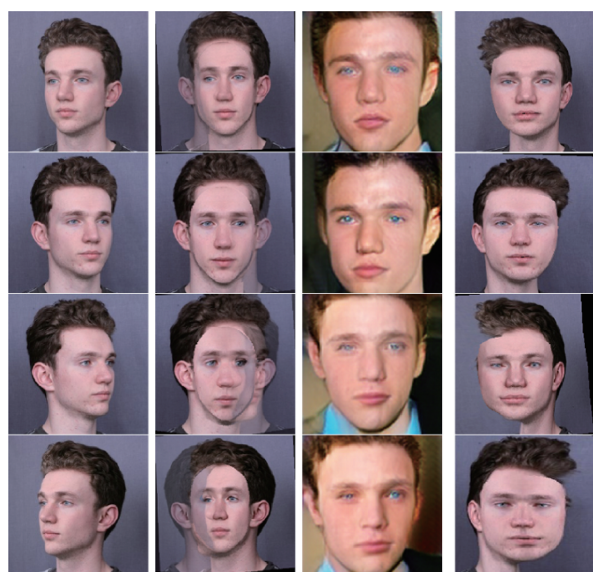


(a) 原始图像 (b) Hassner等人 (2015) (c) Tran等人 (2017a) (d) 本文

图2 不同方法在 LFW 数据库上的人脸矫正结果比较

Fig. 2 Comparisons of face frontalization results of different algorithms on LFW database

((a) original images; (b) Hassner et al. (2015); (c) Tran et al. (2017a); (d) ours)



(a) 原始图像 (b) Hassner等人 (2015) (c) Tran等人 (2017a) (d) 本文

图3 不同方法在 StirlingESRC 数据库上的人脸矫正结果比较

Fig. 3 Comparisons of face frontalization results of different algorithms on StirlingESRC database

database ((a) original images; (b) Hassner et al. (2015); (c) Tran et al. (2017a); (d) ours)

表1 不同方法在 StirlingESRC 数据库上的人脸识别率比较

Table 1 Comparison of face recognition rates of different methods on StirlingESRC database

方法	人脸姿态				/%
					平均
	- 22°	+ 22°	- 45°	+ 45°	
Tran 等人(2017a)	90.71	88.11	76.67	84.34	84.96
Hassner 等人(2015)	92.47	92.21	76.31	72.46	83.36
本文	94.69	94.52	85.29	87.47	90.49

注:加粗字体表示各列最优结果。

从图2和图3可以看出,使用 Hassner 等人(2015)方法得到的正面人脸图像存在明显伪影,导致人脸图像不真实,从而影响最终的人脸识别准确率。使用 Tran 等人(2017a)方法得到的正面人脸图像中整个图像的像素发生了变化,导致人脸面部纹理发生明显改变,降低了人脸识别的精度。而使用本文算法得到的正面人脸图像更具有真实性,本文提出的基于关键点加权的人脸拟合算法能最大程度地模拟人脸面部的表情变化,基于深度学习的人脸图像修复算法可以最大程度地维持原始人脸的面部纹理细节不变,从而可以在一定程度上提高人脸识别的准确率。从表1可以看出,本文方法可以在较大的姿态范围内提高人脸识别的准确率。与 Tran 等人(2017a)和 Hassner 等人(2015)的方法相比,在人脸姿态为 $\pm 22^\circ$ 情况下,本文方法的人脸识别准确率分别提高了 5.195% 和 2.265%;在人脸姿态为 $\pm 45^\circ$ 情况下,本文方法的人脸识别准确率分别提高了 5.875% 和 11.095%;平均人脸识别率分别提高了 5.53% 和 7.13%。对比实验结果表明,本文提出的人脸姿态矫正算法可以有效提高人脸识别的准确率。

图4是本文方法在 LFW 数据库上与 HF-PIM (high fidelity pose invariant model) (Cao 等,2018)以及 CAPG-GAN (couple-agent pose-guided generative adversarial network) (Hu 等,2018)的定性比较结果。可以看出,本文方法取得的正面人脸图像与对比算法取得的正面人脸图像很接近,面部纹理和表情变化都基本一致。表2是在 LFW 数据库上,本文方法与其他方法的人脸识别精度的定量比较。可以看出,尽管本文方法的人脸识别精确度达到了 96.57%,但由于 HF-PIM 和 CAPG-GAN 使用深度学

习的方法进行人脸图像的正面合成,不受人脸关键点的限制,所以这两种方法的人脸识别精确度要高于本文方法。而本文方法结合了人脸重建中的3 维形变模型的优点和深度学习模型的优点,得到的正面人脸图像的面部纹理细节和原始图像更接近,所以人脸识别精确度高于传统方法,如 Hassner 等人 (2015)的方法和 FF-GAN (Yin 等,2017)取得的人脸识别精度。

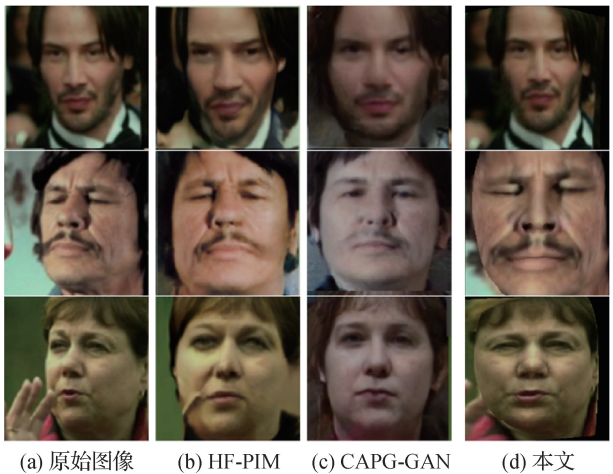


图4 不同方法在 LFW 数据库上的人脸矫正结果比较

Fig. 4 Comparisons of face frontalization results of different algorithms on LFW database
((a) original faces; (b) HF-PIM; (c) CAPG-GAN; (d) ours)

表2 不同方法在 LFW 人脸数据库上的人脸识别精度比较
Table 2 Comparison of face recognition accuracies of different methods on LFW database

方法	人脸识别精度 /%
Hassner 等人 (2015)	91. 65
FF-GAN (Yin 等, 2017)	96. 42
HF-PIM (Cao 等, 2018)	99. 41
CAPG-GAN (Hu 等, 2018)	99. 37
本文	96. 57

注:加粗字体表示最优结果。

3 结 论

针对人脸姿态偏转导致的人脸识别率下降问题,本文提出了一种结合3 维形变模型和图像修复技术的人脸姿态矫正方法。其中,基于关键点加权

的3 维形变模型拟合和基于深度学习的图像修复技术使得最终得到的正面人脸图像更逼真,在一定程度上提高了人脸识别精度。在 LFW 和 StirlingESRC 数据库上的定性实验表明,本文方法能够在较大的人脸姿态下合成真实的正面人脸图像,并且在 LFW 数据库上取得了 96. 57% 的人脸识别精度,在 StirlingESRC 数据库上取得了 90. 49% 的平均人脸识别率,验证了本文算法的有效性。

然而,本文方法也存在以下两点不足,有待改进:1)本文方法依赖于人脸关键点的检测,所以在极端的人脸姿态下无法较好地拟合人脸,在后续的研究中,需要研究更好的人脸拟合方法。例如使用深度学习方法直接从侧面人脸图像中重建出正面人脸图像,不依赖于人脸关键点检测。2)随着深度学习的迅速发展,越来越多基于深度学习的图像修复方法陆续提出,因此,在未来的工作中,需要继续研究如何在深度学习的基础上,提出更好的修复不规则人脸空洞区域的方法,从而不断提高多姿态人脸识别的准确率。

参考文献 (References)

Amberg B, Romdhani S and Vetter T. 2007. Optimal step nonrigid ICP algorithms for surface registration//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA; IEEE: 1-8 [DOI: 10. 1109/CVPR. 2007. 383165]

Asthana A, Marks T K, Jones M J, Tieu K H and Rohith M V. 2011. Fully automatic pose-invariant face recognition via 3D pose normalization//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain; IEEE: 937-944 [DOI: 10. 1109/ICCV. 2011. 6126336]

Blanz V and Vetter T. 1999. A morphable model for the synthesis of 3D faces//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA; ACM: 187-194 [DOI: 10. 1145/311535. 311556]

Blanz V and Vetter T. 2003. Face recognition based on fitting a 3D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25 (9): 1063-1074 [DOI: 10. 1109/TPAMI. 2003. 1227983]

Cao C, Weng Y L, Zhou S, Tong Y Y and Zhou K. 2014. FaceWarehouse: a 3D facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics, 20(3): 413-425 [DOI: 10. 1109/TVCG. 2013. 249]

Cao J, Hu Y B, Zhang H W, He R and Sun Z. 2018. Learning a high fidelity pose invariant model for high-resolution face frontalization//

- Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, USA; Curran Associates Inc.: 2872-2882
- Cole F, Belanger D, Krishnan D, Sarna A, Mosseri I and Freeman W T. 2017. Synthesizing normalized faces from facial identity features//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 3386-3395 [DOI: 10.1109/CVPR.2017.361]
- Deng J K, Guo J, Xue N N and Zafeiriou S. 2018. ArcFace: additive angular margin loss for deep face recognition//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 4685-4694 [DOI: 10.1109/CVPR.2019.00482]
- Ding C X and Tao D C. 2015. Robust face recognition via multimodal deep face representation. IEEE Transactions on Multimedia, 17(11): 2049-2058 [DOI: 10.1109/TMM.2015.2477042]
- Ding L, Ding X Q and Fang C. 2012. Continuous pose normalization for pose-robust face recognition. IEEE Signal Processing Letters, 19(11): 721-724 [DOI: 10.1109/LSP.2012.2215586]
- Hassner T, Harel S, Paz E and Enbar R. 2015. Effective face frontalization in unconstrained images//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 4295-4304 [DOI: 10.1109/CVPR.2015.7299058]
- Heisele B, Serre T and Poggio T. 2007. A component-based framework for face detection and identification. International Journal of Computer Vision, 74(2): 167-181 [DOI: 10.1007/s11263-006-0006-z]
- Hong X, Xiong P F, Ji R H and Fan H Q. 2019. Deep fusion network for image completion//Proceedings of the 27th ACM International Conference on Multimedia. New York, USA; Association for Computing Machinery: 2033-2042 [DOI: 10.1145/3343031.3351002]
- Hu Y B, Wu X, Yu B, He R and Sun Z. 2018. Pose-guided photorealistic face rotation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 8398-8406 [DOI: 10.1109/CVPR.2018.00876]
- Liu G L, Reda F A, Shih K J, Wang T C, Tao A and Catanzaro B. 2018. Image inpainting for irregular holes using partial convolutions//Proceedings of the European Conference on Computer Vision. Munich, Germany; Springer: 89-105 [DOI: 10.1007/978-3-030-01252-6_6]
- Masi I, Rawls S, Medioni G and Natarajan P. 2016. Pose-aware face recognition in the wild//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 4838-4846 [DOI: 10.1109/CVPR.2016.523]
- Paysan P, Knothe R, Amberg B, Romdhani S and Vetter T. 2009. A 3D face model for pose and illumination invariant face recognition//Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance. Genova, Italy; IEEE: 296-301 [DOI: 10.1109/AVSS.2009.58]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany; Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sagonas C, Panagakis Y, Zafeiriou S and Pantic M. 2015. Robust statistical face frontalization//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 3871-3879 [DOI: 10.1109/ICCV.2015.441]
- Schroff F, Kalenichenko D and Philbin J. 2015. FaceNet: a unified embedding for face recognition and clustering//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 815-823 [DOI: 10.1109/CVPR.2015.7298682]
- Tran A T, Hassner T, Masi I and Medioni G. 2017b. Regressing robust and discriminative 3D Morphable models with a very deep neural network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 1493-1502 [DOI: 10.1109/CVPR.2017.163]
- Tran L, Yin X and Liu X M. 2017a. Disentangled representation learning GAN for pose-invariant face recognition//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 1283-1292 [DOI: 10.1109/CVPR.2017.141]
- Wang H, Wang Y T, Zhou Z, Ji X, Gong D H, Zhou J C, Li Z F and Liu W. 2018. CosFace: large margin cosine loss for deep face recognition//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 5265-5274 [DOI: 10.1109/CVPR.2018.00552]
- Yeh R A, Chen C, Lim T Y, Schwing A G, Hasegawa-Johnson M and Do M N. 2017. Semantic image inpainting with deep generative models//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 6882-6890 [DOI: 10.1109/CVPR.2017.728]
- Yim J, Jung H, Yoo B, Choi C, Park D and Kim J. 2015. Rotating your face using multi-task deep neural network//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 676-684 [DOI: 10.1109/CVPR.2015.7298667]
- Yin X, Yu X, Sohn K, Liu X M and Chandraker M. 2017. Towards large-pose face frontalization in the wild//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 4010-4019 [DOI: 10.1109/ICCV.2017.430]
- Zadeh A, Lim Y C, Baltrušaitis T and Morency L P. 2017. Convolutional experts constrained local model for 3D facial landmark detection//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. Venice, Italy: 2519-2528 [DOI: 10.1109/ICCVW.2017.296]
- Zhu Z, Luo P, Wang X and Tang X. 2014. Multi-View perceptron: a

deep model for learning face identity and view representations//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press: 217-225

Zhu Z Y, Luo P, Wang X G and Tang X O. 2013. Deep learning identity-preserving face space//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE: 113-120 [DOI: 10.1109/ICCV.2013.21]

作者简介



吴从中, 1968年生, 男, 副教授, 主要研究方向为人工智能。

E-mail: 329161005@qq.com



詹曙, 通信作者, 男, 教授, 博士生导师, 主要研究方向为人工智能。

E-mail: shu_zhan@hfut.edu.cn

郑荣生, 男, 硕士研究生, 主要研究方向为人脸识别。

E-mail: zhengrs@mail.hfut.edu.cn

臧怀娟, 女, 博士研究生, 主要研究方向为图像超分辨率重建。E-mail: zanghj@mail.hfut.edu.cn

刘明威, 男, 硕士研究生, 主要研究方向为目标检测。

E-mail: mwliu@mail.hfut.edu.cn

徐甲甲, 男, 研究员, 主要研究方向为人脸识别。

E-mail: 2310798020@qq.com