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
11
VOL.26

ISSN1006-8961
CN11-3758/TB

电力视觉前沿技术

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第26卷第11期（总第307期）

2021年11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

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》编辑出版委员会
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@aircas.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 and Publishing Board 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@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

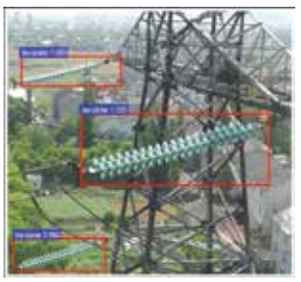
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

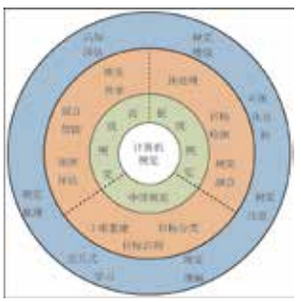
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2021年11月 第26卷 第11期（总第307期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次



面向改进尺度缩放网络的绝缘子识别(第2561页)

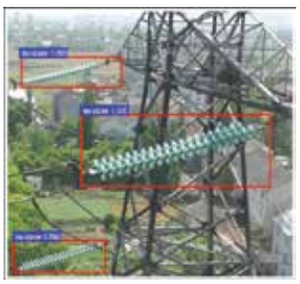


混合增强视觉认知架构及其关键技术进展(第2619页)

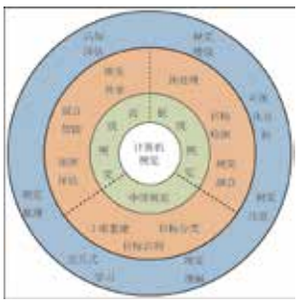
激光点云的稀疏体素金字塔邻域构建与分类(第2703页)

Volume 26, Number 11
Published November 16, 2021

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



Insulator recognition based on an improved scale-transferrable network(P2561)



Hybrid enhanced visual cognition framework and its key technologies(P2619)

Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud(P2703)

E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP399　　**文献标识码**: A　　**文章编号**: 1006-8961(2021)11-2630-15

论文引用格式: Xu Y L, Zhao J L, Lyu Z H, Zhang Z M, Li J H and Pan Z K. 2021. Automatic facial feature points location based on deep learning: a review. Journal of Image and Graphics,26(11):2630-2644(徐亚丽,赵俊莉,吕智涵,张志梅,李劲华,潘振宽. 2021. 深度学习人脸特征点自动定位综述. 中国图象图形学报,26(11):2630-2644)[DOI:10.11834/jig.200278]

深度学习人脸特征点自动定位综述

徐亚丽,赵俊莉*,吕智涵,张志梅,李劲华,潘振宽

青岛大学计算机科学技术学院,青岛　266071

摘　要:　人脸特征点定位是根据输入的人脸数据自动定位出预先按人脸生理特征定义的眼角、鼻尖、嘴角和脸部轮廓等面部关键特征点,在人脸识别和分析等系统中起着至关重要的作用。本文对基于深度学习的人脸特征点自动定位进行综述,阐释了人脸特征点自动定位的含义,归纳了目前常用的人脸公开数据集,系统阐述了针对2维和3维数据特征点的自动定位方法,总结了各方法的研究现状及其应用,分析了当前人脸特征点自动定位技术在深度学习应用中的现状、存在问题及发展趋势。在公开的2维和3维人脸数据集上对不同方法进行了比较。通过研究可以看出,基于深度学习的2维人脸特征点的自动定位方法研究相对比较深入,而3维人脸特征点定位方法的研究在模型表示、处理方法和样本数量上都存在挑战。未来基于深度学习的3维人脸特征点定位方法将成为研究趋势。

关键词:深度学习;2维人脸特征点定位;3维人脸特征点定位;卷积神经网络(CNN);配准

Automatic facial feature points location based on deep learning: a review

Xu Yali, Zhao Junli*, Lyu Zhihan, Zhang Zhimei, Li Jinhua, Pan Zhenkuan

College of Computer Science & Technology, Qingdao University, Qingdao　266071, China

Abstract: Face feature point location is to locate the predefined key facial feature points automatically according to the physiological characteristics of the human face, such as eyes, nose tip, mouth corner, and face contour. It is one of the important problems in face registration, face recognition, 3D face reconstruction, craniofacial analysis, craniofacial registration, and many other related fields. In recent years, various algorithms for facial feature point localization have emerged constantly, but several problems remain in the calibration of feature points, especially in the calibration of 3D facial feature points, such as manual intervention, low or inaccurate number of feature points, and long calibration time. In recent years, convolutional neural networks have been widely used in face feature point detection. This study focuses on the analysis of automatic feature point location methods based on deep learning for 2D and 3D facial data. Training data with real feature point labels in 2D texture image data are abundant. The research of automatic location method of 2D facial feature points based on deep learning is relatively extensive and indepth. The classical methods for 2D data include cascade convolution neural network methods, end-to-end regression methods, auto encoder network methods, different pose estimation methods, and other improved convolutional neural network (CNN) methods. In cascaded regression methods, rough detection is per-

收稿日期:2020-07-17;**修回日期**:2020-12-18;**预印本日期**:2020-12-25

*通信作者:赵俊莉　zhaojl@yeah.net

基金项目:国家自然科学基金项目(62172247,61772294,61702293,61902203);全国统计科学研究项目(2020355);山东省重点研发计划重大科技创新工程项目(2019JZZY020101);山东省自然科学基金项目(ZR2019LZH002)

Supported by: National Natural Science Foundation of China (62172247, 61772294, 61702293, 61902203)

formed first, and then the feature points are finetuned. The end-to-end method propagates the error between the real results and the predicted results until the model converges. Autoencoder methods can select features automatically through encoding and decoding. Head pose estimation has great importance for face feature point detection because image-based methods are always affected by illumination and pose. Head pose estimation and feature points detection is improved by modifying network structure and loss function. The disadvantage of cascade regression method is that it can update the regressor by independent learning, and the descent direction may cancel each other. The flexibility of the end-to-end model is low. CNN is applied to 2D training data with real feature point tags. However, in the case of a 3D, training data with rich real feature point labels are lacking. Therefore, compared with 2D facial feature points, 3D facial feature point location remains a challenge. Several automatic feature point location for 3D data are introduced. The methods for 3D data are mainly based on depth information and 3D morphable model (3DMM). In recent years, with the development of RGB + depth map (RGBD) technology, depth data have attracted more attention. Feature point detection based on depth information has become an important preprocessing step for automatic feature point detection in 3D data. Initialization is crucial for deep data, but information is easily lost. The method based on 3DMM represents 3D face data for locating feature points through deep learning. On the one hand, the shape and expression parameters of 3DMM are highly nonlinear with the image texture information, which makes image mapping difficult to estimate. Compared with 2D face data, 3D face data lack training data with remarkable changes in face shape, race, and expression. Face feature point detection still faces great challenges. In summary, this study explains the meaning of automatic location of facial feature points, summarizes the currently open and commonly used face datasets, introduces various methods of automatic location of feature points for 2D and 3D data, summarizes the research status and application of each domestic and international method, analyzes the problems and development trend of automatic location technology of face feature points in deep learning application on 2D and 3D datasets, and compares the experimental results of the latest methods. In conclusion, the research on automatic location method of 2D face feature points based on deep learning is relatively indepth. Challenges in processing 3D data remain. The current solution for locating feature points is to project 3D face data onto 2D images through cylindrical coordinates, depth maps, 3DMM, and other methods. Information loss is the main problem of these methods. The method of feature point location directly on 3D model needs further exploration and research. The accuracy and speed of feature point location also need to be improved. In the future, 3D facial feature point localization methods based on deep learning will gradually become a trend.

Key words: deep learning; 2D facial feature point location; 3D facial feature point location; convolutional neural network (CNN); registration

0 引 言

人脸特征点定位是根据输入的人脸数据自动定位出预先按人脸生理特征定义的面部关键特征点，如眼角、鼻尖、嘴角和脸部轮廓等，在人脸配准、人脸识别、颅面分析和 3 维人脸重建等领域具有重要作用。

传统的人脸特征点定位方法包括主动形状模型(active shape models, ASM)(Cootes 和 Taylor, 1992)、主动外观模型(active appearance models, AAM)(Cootes 等，2001)、约束局部模型(constrained local models，CLM)(Cristinacce 和 Cootes，2008)和基于回归的方法(Cao 等，2014)等。ASM 和 AAM 采用局部纹理模型和全局统计模型寻找最优形状，可分别应用于实时性和精度要求较高的场合，但均依赖于初始形状。CLM 采用判别性局部纹理模型对特征点位置进行正则化，具有更好的泛化能力和鲁棒性。基于回归的方法直接使用特征点周围的判别特征来估计特征点位置，但大多数不考虑不同视角下人脸特征点的可见性，所以对于具有大姿态的人脸图像，其性能会显著降低。

随着基于深度学习的人脸识别研究的开展，卷积神经网络(convolutional neural networks, CNN)(Lecun 等，1998)广泛应用于人脸特征点检测。本文重点分析针对 2 维和 3 维人脸数据基于深度学习的特征点自动定位方法。2 维纹理图像数据中具有真实特征点标签的训练数据非常丰富，对其特征点自动定位方法较多，包括基于级联卷积神经网络的方法、基于深度端到端回归的方法、基于自动编码器

网络的方法、基于不同姿态估计的方法以及其他改进 CNN 的方法等。级联回归先进行粗检测,然后再微调特征点。端到端的方法将真实结果和预测结果之间的误差反向传播直至模型收敛。2 维人脸的特征点自动定位方法取得了较为理想的效果,并在人脸识别等实际任务中广泛应用。

随着 RGBD(RGB + depth map)技术的发展,深度数据越来越受到关注,基于深度信息的特征点检测成为 3 维数据自动检测特征点的重要预处理步骤。Gilani 等人(2017)提出深度特征点识别网络(deep landmark identification network, DLIN),针对 3 维数据研究基于深度信息的 3 维人脸特征点自动标定方法。Zhu 等人(2016)提出跨大姿态人脸对齐的 3D 解决方案(3D dense face alignment,3DDFA),在 3D 人脸空间的 3 维可变形模型(3D morphable model,3DMM)(Blanz 和 Vetter,2002)等的基础上通过深度学习研究自动定位特征点的方法。这些研究对 3 维人脸特征点的自动标定方法进行了有益探索,但与 2 维人脸数据相比,3 维人脸数据缺乏人脸形状、种族和表情等显著变化的训练数据。3 维人脸特征点的标定依然存在手工干预、标定特征点数量少或不准确和标定时间长等问题。3 维数据在模型表示、数据处理和样本数量上都存在挑战,直接在 3 维模型上进行特征点定位的方法还有待进一步探索和研究,3 维人脸模型特征点定位的精度和速度也有待提高。

1 人脸数据集与预处理

1.1 数据集

2 维人脸图像特征点检测常用的数据库较多,且包含头部姿态、人脸表情和照明等的适度变化。主要有 300W(300 faces in the wild)(Sagonas 等,2016)、AFLW (annotated facial landmarks in the wild)(Köstinger 等,2011)、LFW (labled faces in the wild)(Belhumeur 等,2013)、AFW(annotated datasets in the wild)(Zhu 和 Ramanan,2012)、CelebA(Large-scale CelebFaces Attributes)(Liu 等,2015)和 Helen(Le 等,2012),如表 1 所示。常用的 3 维人脸数据集不多,每个数据集中的 3 维人脸模型数量也较少,且大部分没有特征点信息,需进行预处理后才能使用,主要有 CASIA(Institute of Automation, Chinese Academy of Sciences)(Yi 等, 2014)、FRGCv2.0 (Face Recognition Grand Challenge Biometrics Database (v2.0))(Kelkboom 等,2007)、BU-3DFE(Binghamton University 3D Facial Expression)(Zheng 等,2009)和 BJUT(Beijing University of Technology)(尹宝才 等,2009),如表 2 所示。

表 1 2 维人脸图像数据集

Table 1 2D face dataset

数据集	图像数量	特征点数	条件变化	简介
300W (Sagonas 等,2016)	3 837 幅,训练集 3 148 幅,测试集 689 幅	68	姿态、表情、光照	每幅图像包含不止 1 张人脸,但仅标注 1 张
AFLW (Köstinger 等,2011)	25 993 幅,女性 59%,男性 41%	21	姿态、表情、光照	有少部分灰度图像
LFW (Belhumeur 等,2013)	13 233 幅,5 749 张人脸	未公开	非受限	每张脸标记人名,其中 1 680 人有两张或以上不同照片
AFW (Zhu 和 Ramanan, 2012)	205 幅,473 张标记的人脸	6	背景、姿态、尺寸	每张人脸包含 1 个长方形边界框、相关的姿势角度
CelebA (Liu 等,2015)	202 000 幅	5	–	大规模人脸数据库,包含 10 000 个身份
Helen (Le 等,2012)	2 330 幅,训练集 2 000 幅,测试集 330 幅	194	–	提供标记特征点数目最多的公开图像库

注:“–”表示官方未介绍。网址:300W:https://ibug.doc.ic.ac.uk/resources/300-W/;AFLW:http://lrs.icg.tugraz.at/research/aflw/; LFW:http://vis-www.cs.umass.edu/lfw/index.html#download;AFW:https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/; CelebA:http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html; HELEN:http://www.ifp.illinois.edu/~vuongle2/helen/。

表 2　3 维人脸图像数据集

Table 2　3D face dataset

名称	数量	类型	三角网络数	纹理信息	条件变化	简介
CASIA (Yi 等,2014)	123 人	三角网格	10 000～20 000	有	姿态、表情、光照	每人 27 或 38 幅 3 维人脸数据
FRGCv2. 0 (Kelkboom 等,2007)	466 人, 4 007 个扫描	–	–	–	表情、种族、年龄	主要是正面扫描,姿态变化较小,表情从中性到极端
BU-3DFE (Zheng 等,2009)	100 人,女性 56%，男性 44%	三角网格	130 000	有	表情、种族、年龄	每个对象扫描 7 个表情,除中性表达,6 个原型表达都含 4 个强度
BJUT (尹宝才 等,2009)	男女各 250 人	三角网格	130 000	有	中性、表情、年龄	每人 1 个 3 维模型,去噪并切除多余部分,部分人有 3 个样本

注：“–”表示官方未介绍。网址：CASIA：https://paperswithcode. com/dataset/casia-webface；FRGCv2. 0：https://ibug. doc. ic. ac. uk/resources/facial-point-annotations/；BU-3DFE:https://www. cs. binghamton. edu/～lijun/Research/3DFE/3DFE_Analysis. html；BJUT:http://www. bjut. edu. cn/sci/multimedia/mul-lab/3 d face/face_database. htm。

1.2　数据预处理

数据预处理是在提取特征前排除与人脸特征点定位无关的干扰,如光照、背景等。主要方法有人脸归一化(章柏幸和苏光大,2013)和人脸边界框检测(Sun 等,2013;李启运 等,2019)。同时,为确保特征点定位任务的普遍性,需要对数据进行增强,以保证有足够有效的训练数据。

1.2.1　人脸归一化

人脸归一化可以预先消除坐标变换的影响,常用的方法有 min-max 标准化和 Z-score 标准化。min-max 标准化对图像数据进行线性变换,将结果值映射到 0～1 之间。Z-score 标准化是根据均值(mean)和标准差(standard deviation)将图像数据标准化。经过处理的图像符合标准正态分布,即均值为 0,标准差为 1。

1.2.2　数据增强

数据增强的方法有缩放、平移、对比度变换、噪声扰动(张明 等,2018)、颜色变化、水平和垂直翻转等。同时,还有其他方法如生成对抗网络(generative adversarial networks, GAN)(Karras 等, 2019; Goodfellow 等,2014)等生成人脸数据。

1.2.3　人脸检测

通过人脸边界框(bounding box)检测裁剪的图像范围变小,外界因素(背景、头发等)的干扰减小,有利于提高特征点定位精度。经典方法采用人脸检测器识别边界框,而采用 CNN 定位的边界框(Ren 等,2017)精度比直接采用人脸检测器定位高。人脸边界框如图 1 所示,且目前尚未制定统一的人脸特征点标准。

(a) 300W数据集边界框　(b) Dlib边界框

图 1　人脸边界框(Xu 和 Kakadiaris,2017)

Fig. 1　Face bounding box(Xu and Kakadiaris,2017)

((a)300W dataset bounding box;(b)Dlib bounding box)

2　基于深度学习的 2 维人脸特征点自动检测

2.1　基于级联卷积神经网络的方法

如果只采用一个网络做回归,训练 CNN 检测特征点,会发现得到的特征点坐标不准确;如果采用更大的网络,特征点的预测会准确,但耗时会增加。为了在速度和性能上找到一个平衡点,采用级联回归。现有方法多是大致估计人脸特征点初始位置,而级联回归先进行粗检测,然后再微调特征点。

Sun 等人(2013)提出的 DCNN(deep convolutional neural network)将整张脸作为输入,用三级卷积网络检测人脸特征点,可以很好地利用上下文信

息。由于同时预测多个点,关键点的约束也隐含其中,且在第1级就做出准确估计,有效避免了局部最小值问题。

DCNN特征点定位结果如图2所示。在DCNN的基础上,Zhou等人(2013)设计了一个4级级联DCNN,在输入方面,采用CNN分别预测内点和边界点的最小人脸边界框,并分开预测特征点。预测内点的部分从第2层开始将五官裁剪进行精定位。这个改进提高了初始层的定位精度,相较于传统CNN,很好地解决了传统网络在训练广泛人脸特征点定位任务中的问题,实现了68个人脸特征点的高精度定位。Xiao等人(2016)提出递归注意细化网络(recurrent attentive refinement, RAR),按顺序细化特征点位置并引入长短期记忆网络(long short-term memory, LSTM)模型(Karim等,2017),在多个基准测试中展示了优异性能。

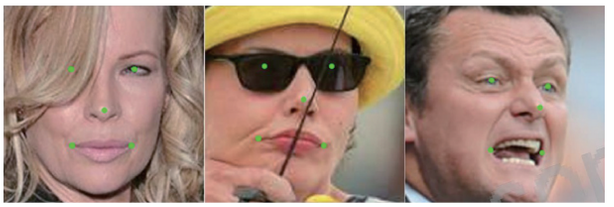


图2 DCNN特征点定位结果(Sun等,2013)

Fig. 2 Locating results of DCNN feature points (Sun et al., 2013)

级联回归方法的缺点是以独立学习的方式更新回归器、下降方向可以互相抵消、手工提取的特征,如方向梯度直方图(histogram of oriented gridients, HOG)、尺度不变特征变换(scale invariant feature transform, SIFT)(Lowe, 2004)等主要用于驱动级联,但不能保证学习的特征点是最优的。

2.2 基于深度端到端回归的方法

端到端的方法是从输入到输出端得到一个预测结果,与真实结果相比得到一个误差,这个误差在模型中的每一层反向传播,每一层的表示都会根据这个误差进行调整,直到模型收敛。端到端的学习省去了每一个独立学习任务执行之前的数据标注,节省了时间和人力成本。

Kumar等人(2019)使用高斯对数似然损失来联合估计特征点位置及其不确定性。在Du-Net(Tang等,2020)瓶颈层中添加两个分支:1)平均值估计器,计算每个特征点的估计位置;2)cholesky估计器网络(cholesky estimator network, CEN),估计每个特征点位置的2维高斯概率分布协方差矩阵的cholesky系数,网络结构如图3所示,图3中GLL为高斯对数似然(Gaussian log-likelihod), UGLLI为不确定性高斯对数似然(uncertainty with Gaussian log-likelihood)。

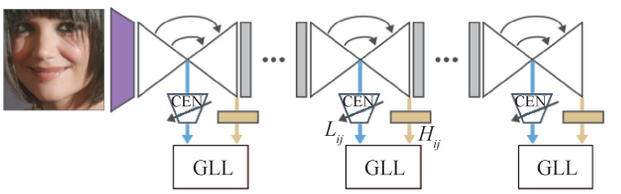


图3 UGLLI网络结构(Kumar等,2019)

Fig. 3 Network structure of UGLLI (Kumar et al., 2019)

随后,Kumar等人(2020)提出LUVLi(location, uncertainty and visibility likelihood)损失函数,在瓶颈层添加CEN和可见性估计器网络(visibility estimator network, VEN),并对热图(heatmap)应用均值估计来优化深度网络,从而很好地估计特征点。Dapogny等人(2019)提出基于特征点注意图(attention map)的深度卷积级联结构(deep convolutional cascade for face alignment, DeCaFA),为每一个特征点定位任务生成特征点注意图,加权中间监督以及各阶段之间的有效特征融合,允许学习以端到端的方式逐步完善注意图。DeCaFA可以从使用粗注释数据的极少数图像中学习到具有合理精度的精细定位。Ranjan等人(2019)提出HyperFace,对融合特征采用多任务学习算法,首先在检测到的人脸区域上对全局进行训练,获得对姿态的粗略估计,然后对7个主要特征点进行定位,该方法利用任务之间的协同作用,提高了各任务的能力。Cong等人(2017)提出硬示例建议网络(hard example proposal network, HEPN),使用3个独立CNN,产生3种不同输出。12net产生人脸与非人脸分类结果、24net产生候选窗口校准结果、48net产生特征点检测结果,实现端到端的交替训练方法,网络结构如图4所示。图4中,Fe为facial estimation。Kumar和Chellappa(2018)提出姿势条件树突状卷积神经网络(pose conditioned dendritic convolution neural network, PCD-CNN),第1个分类网络之后是第2个模块化的分类网络,以端到端方式训练,以获得准确的特征点。在贝叶斯公式基础上,通过对姿态的特征点估计进行

条件化处理，使其不同于多任务处理方法，从而明确分离出人脸图像的 3 维姿态。该方法对姿态的调节可以使人脸姿态不可知，从而减少定位误差。提出的模型可以扩展到生成可变数目的特征点，从而扩大了对其他数据集的适用性。

端到端模型的灵活性较低，例如，原本多个模块中数据获取难度不一样时，可能不得不依靠额外的模型来协助训练。

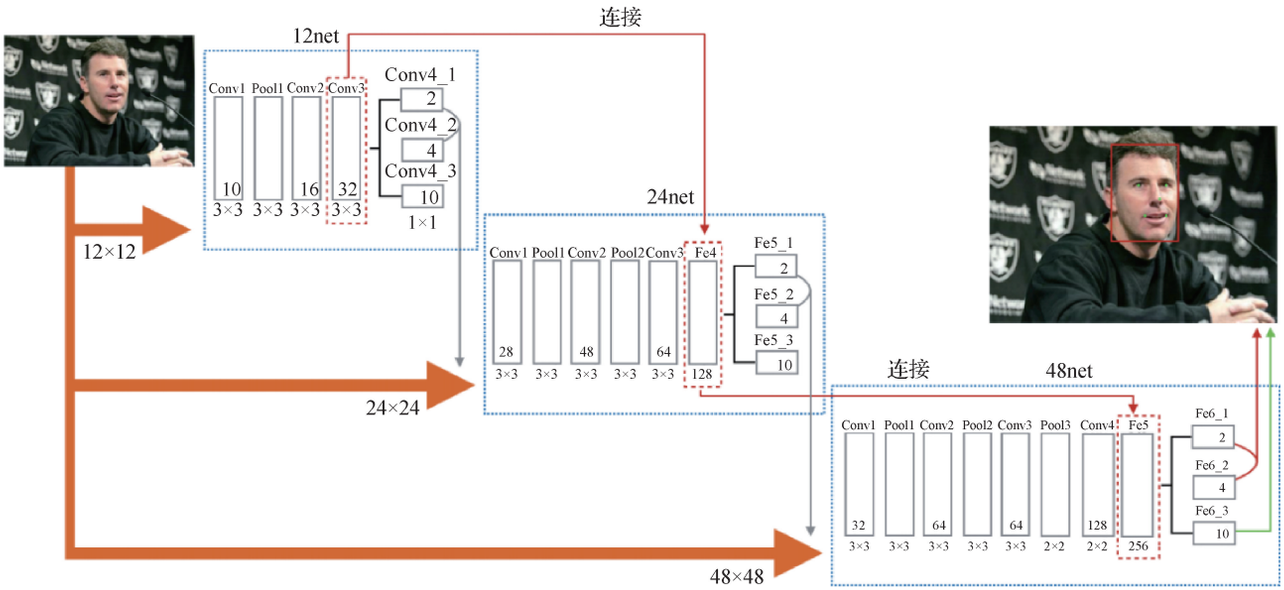


图 4　硬示例建议网络（Cong 等，2017）

Fig. 4　Hard example proposal network（Cong et al. ，2017）

2. 3　基于自动编码器网络的方法

自动编码器（auto-encoder）（Zhang 等，2020）是一种神经网络，属于无监督学习算法，用于降维或学习特征，利用反向传播使目标值等于输入值，对从人脸图像到人脸形状的非线性映射十分有用，可使计算量降低，可解释性提高，实现特征的自动选择。

Zhang 等人（2014a）提出一种粗到细的自动编码器网络（coarse-to-fine auto-encoder networks，CFAN），将几个连续的堆叠式自动编码器网络（stackable automatic encoder netwrok，SAN）（Mirjalili 等，2018）层叠起来，如图 5 所示。其中，H_1，H_2 是隐藏层。通过函数 F_ϕ，在当前形状 S_i 的人脸特征点周围提取局部特征 $\phi(S_i)$。全局 SAN 将检测到的低分辨率人脸作为输入，初步预测特征点。随后局部 SAN 通过以越来越高的分辨率将从当前特征点周围提取的局部特征（之前 SAN 的输出）作为输入，逐步完善特征点。每个 SAN 都可以解决部分非线性问题。

Browatzki 和 Wallraven（2020）介绍了一种半监督方法，核心思想是首先从现有的大量未标记人脸图像中生成隐式人脸知识，然后在第 1 个完全无监

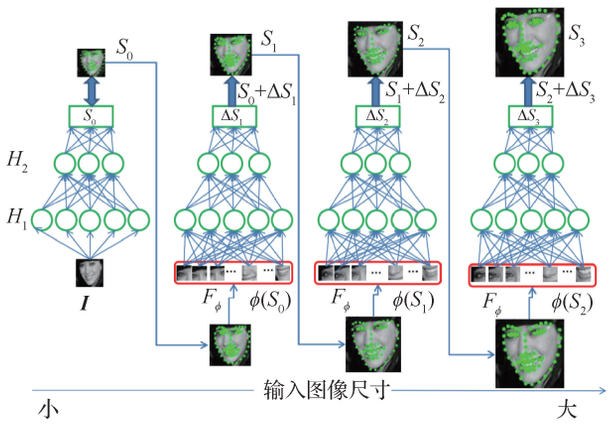


图 5　堆叠式自动编码器网络（Zhang 等，2014a）

Fig. 5　Stackable automatic encoder network（Zhang et al. ，2014a）

督阶段训练一个对抗性的自动编码器，通过低维人脸嵌入来重建人脸，在第 2 个有监督阶段将解码器与传输层交错，重新分配彩色图像的生成任务，以预测特征点热图。Browatzki 和 Wallraven（2020）提出的 3FabRec（face alignment by reconstruction）框架能够在非常小的训练集上保持极高的精确度。由于高度非线性的优化，该方法的主要问题是时间复杂度高和易于陷入局部最小值。

2.4 基于不同姿态估计的方法

头部姿态估计对人脸特征点检测具有重要意义,但基于图像的方法总是受到光照和姿态等的影响,早期的大姿态人脸对齐工作对不同视图使用不同的特征模板(Yu 等,2013),且会产生较高的计算成本。

Zhang 等人(2014b)提出任务约束的深度卷积网络(tasks-constrained deep convolutional network, TCDCN),通过提前停止任务来促进学习收敛,大幅降低了处理严重遮挡和姿态变化的人脸模型的复杂度。与孤立学习人脸特征点检测不同的是,通过与不同但微妙相关的任务(如外观属性、表情、头部姿态)的联合学习,可以实现更健壮的特征点检测。TCDCN 允许相关任务的错误在深隐藏层中反向传播,以构建与主任务相关的共享表示。由于头部姿态与特征点位置具有高度相关性,Xu 和 Kakadiaris(2017)利用 CNN 获得的全局和局部特征来联合估计头部姿态和定位特征点,首先在整个人脸图像上应用 GNet(global net)估计头部姿态和特征点,提供很好的初始化,然后将 LNet(local net)应用于从当前形状裁剪的面片中学习局部特征,局部 CNN 为级联回归提供了判别特征。Bulat 和 Tzimiropoulos(2017)基于 Newell 等人(2016)提出的沙漏(hour glass, HG)网络构建人脸配准网络(face alignment network, FAN),使用由 4 个 HG 网络组成的堆栈,在 HG 基础上引入分层、并行和多尺度块代替 HG 中的瓶颈块,当使用相同数量的网络参数时,该方法的性能优于 HG,可以很好地适用于人体姿态估计和人脸对齐数据集。Wu 和 Yang(2017)利用不同数据集内部及其之间的丰富变化提出深度变异杠杆网络(deep variation leveraging network, DVLN),该网络由跨数据集网络(dataset across network, DA-Net)和候选决策网络(candidate decision network, CD-Net)两个强耦合子网络组成,其中 DA-Net 利用不同数据集的不同特征和分布,而 CD-Net 则对 DA-Net 给出的候选假设做出最终决策,以利用某一特定数据集中的变化。Kowalski 等人(2017)提出基于深度神经网络结构的深度对齐网络(deep alignment network, DAN),该方法由多阶段组成,每个阶段改进前一阶段估计的特征点位置,并引入特征点热图来提供有关特征点位置的视觉信息,可使 DAN 处理头部姿态变化大、初始化困难的人脸图像。与基于局部面片的方法不同,该方法通过特征点热图和特征图像在各个阶段之间传递当前特征点的位置信息,从而能够利用整个人脸图像,避免陷入局部极小值。如图 6 所示,$\boldsymbol{I}$ 为输入的图像,网络的每个阶段从初始估计 S_0 开始,细化前一阶段产生的特征点位置,连接层通过生成特征点热图 $\boldsymbol{H}_t$、特征图像 $\boldsymbol{F}_t$ 和用于将输入图像扭曲为标准姿态的变换 $\boldsymbol{T}_t$ 在网络的连续阶段之间形成连接。Feng 等人(2018)为了解决数据集头部大范围外旋转时样本的欠表达问题,提出了基于姿势的数据平衡,通过复制少数训练样本并注入随机图像旋转、边界盒平移和其他数据增强方法来处理数据不平衡问题,同时提出新的损失函数,即机翼损失(wing loss),新的损失通过从 L1 损失切换到修正的对数函数,放大了区间误差的影响,最后建立一个两阶段的人脸特征点定位架构。

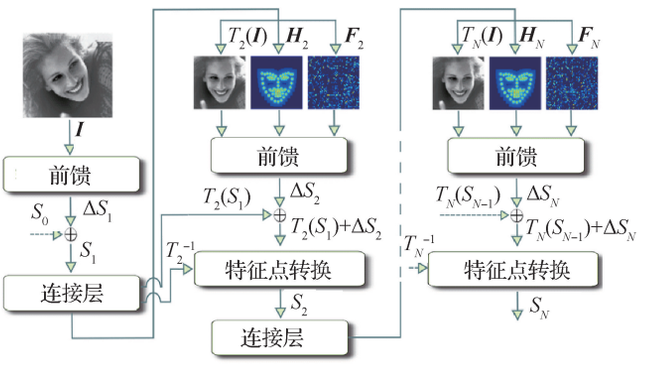


图 6 DAN 网络结构(Kowalski 等,2017)

Fig. 6 Network structure of DAN(Kowalski et al., 2017)

2.5 其他改进网络的方法

GoogLeNet(Szegedy 等,2015)的核心思想是 Inception 模块,主要贡献是使用 1 × 1 卷积进行升降维,在多个尺寸上同时卷积再聚合。Inception 能更高效地使用计算资源,在相同计算量下提取更多的特征,从而提升训练成果。郑银环等人(2019)提出了一种基于小滤波器的深度卷积神经网络(deep convolution neural network with small filter, DCNNSF),引入小滤波器思想有效减少参数,避免过拟合问题,拓展网络深度以提升网络性能,提高了算法的有效性与适用性。

现有的用于训练人脸关键点检测算法的数据集主要由低分辨率图像组成,现有算法进行高分辨率人脸特征点检测的方法是降低图像样本,牺牲分辨率和质量。Chandran 等人(2020)基于注意力机制

提出完全卷积的区域架构，专门设计用于预测高分辨率人脸图像上的特征点，而不需要向下采样，结合注意力驱动的裁剪人脸区域，引入可微的 soft-argmax 操作，搭建一个基于人脸区域的全卷积网络（fully convolutional networks，FCN）（Long 等，2015）特征点检测器。

3 基于深度学习的 3 维人脸特征点自动检测

与 2 维人脸特征点定位相比，3 维数据在模型表示、数据处理和样本数量上都存在挑战，适用于深度学习 3 维人脸特征点标定的 3 维模型的有效表示方法还有待进一步深入研究和探讨。目前针对 3 维数据的研究包括将 3 维模型转化为 2 维图像和基于 3 维形变模型的两类特征点自动定位方法。

3.1 将 3 维模型转化为 2 维图像的方法

由于 2 维人脸图像的特征点标定方法研究较多，因而将 3 维人脸模型转化为 2 维图像数据，就可以利用较成熟的 2 维图像数据来辅助检测 3 维人脸数据特征点，但此类算法无法只用 3 维人脸数据信息实现，要求存在 2 维辅助图像。深度图像中包含丰富的几何信息，基于深度信息的特征点检测已成为 3 维数据自动检测特征点的重要预处理步骤。

Gilani 等人（2017）提出了一种全自动的多线性算法，可以自动跨身份、人脸表情和姿态等在任意数量的特定人群的 3D 面孔上建立密集的点对点对应关系。设计了一个深度特征点识别网络（deep landmark identification network，DLIN），首先生成 3D 数据深度图像，然后计算点云中每个顶点的笛卡儿表面法线（$\boldsymbol{n}_x$，$\boldsymbol{n}_y$，$\boldsymbol{n}_z$）并将其转换为球坐标（$\boldsymbol{n}_\theta$，$\boldsymbol{n}_\varphi$，$\boldsymbol{n}_r$），其中 θ 是方位角，φ 是仰角，将深度图、方位图和仰角图作为 3 个通道代替 RGB 通道作为 DLIN 的输入。DLIN 架构改进 FCN 以更好地适合 3D 深度数据而不是 RGB 数据，网络结构如图 7 所示。Terada 等人（2018）在 ResNet（He 等，2016）的基础上改进 CNN，提出利用柱坐标系的方法将 3 维人脸图像中的数据转换为 2 维图像。最后使用圆柱投影法将识别出的人脸特征点从 2 维图像转换为 3 维图像。

深度图像中的主要问题是噪声甚至数据丢失。其中，初始化是关键因素，好的初始值对于人脸特征点的检测是非常重要的，如果初始值离实际形状较远，则最终的回归结果将不够精确。

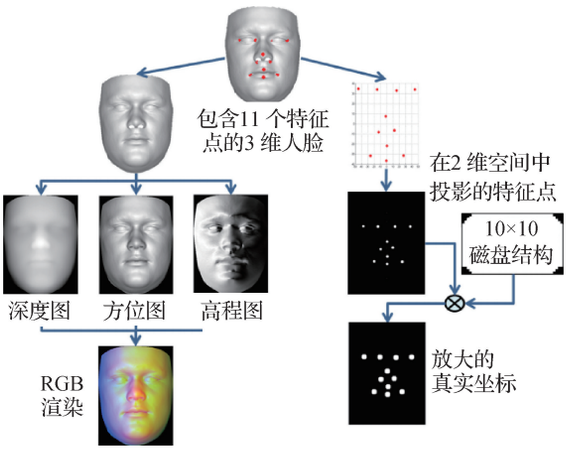


图 7 DLIN 网络结构（Gilani 等，2017）

Fig. 7 Network structure of DLIN（Gilani 等，2017）

3.2 基于 3 维形变模型的方法

3DMM（3D morphable models）（Blanz 和 Vetter，2002）是 3 维人脸形状和纹理的有力统计模型，可以基于一组人脸形状和纹理的统计模型来表示任意一张人脸，是联系 2 维人脸图像与 3 维人脸模型的重要工具。

Zhu 等人（2016）提出面向多姿态的人脸对齐的 3D 解决方案（3D dense face alignment，3DDFA），根据 3DMM，投影 3D 人脸到 2 维图像平面上，构造投影的标准化坐标代码（projected normalized coordinate code，PNCC），将标准化的平均面（normalized coordinate code，NCC）作为投影的 3 维人脸的颜色映射，如图 8 所示。NCC_x 是 NCC 在 x 轴的值，表示 R；$RNCC_y$ 为 RNCC 在 y 轴的值，表示 G；NCC_z 为 NCC 在 z 轴的值，表示 B。PNCC 提供了图像平面上可见 3D 顶点的 2D 位置，将其与输入图像叠加，很好地解决了大姿态下人脸对齐问题。与传统的特征点检测框架不同，3DDFA 拟合了具有级联 CNN 的密集 3D 变形模型，以解决建模中的遮挡和大型姿态拟合中的高非线性问题。同时，提出了人脸分析算法在轮廓视图中合成脸部外观，可提供丰富的训练样本。采用 Z-缓冲区来渲染由 NCC 着色的投影 3D 面。

利用 3DMM 处理自遮挡和大姿态特征点检测，可以利用 3 维可变形模型计算 2 维特征点的可见性和位置。Jiang 等人（2018）提出了一种利用 CNN 对人脸图像进行特征提取的模型 DFF（deep face feature），用神经网络将每个人脸图像像素映射到一个高

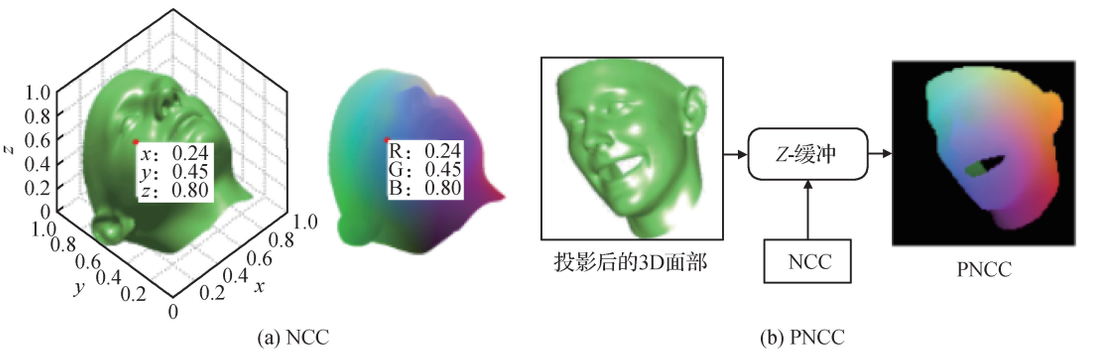


图 8　NCC 和 PNCC(Zhu 等,2016)

Fig. 8　NCC and PNCC(Zhu et al. ,2016)((a)NCC;(b) PNCC)

维点,然后将其归一化为单位长度。为了有效地表示和区分人脸特征,规范化的 DFF 描述符必须保留 3 维人脸表面的度量结构,如图 9 所示,如果两个像素对应于 3 维表面上的附近点,则它们的标准化 DFF 应在超球面上接近;否则,它们应彼此足够远。

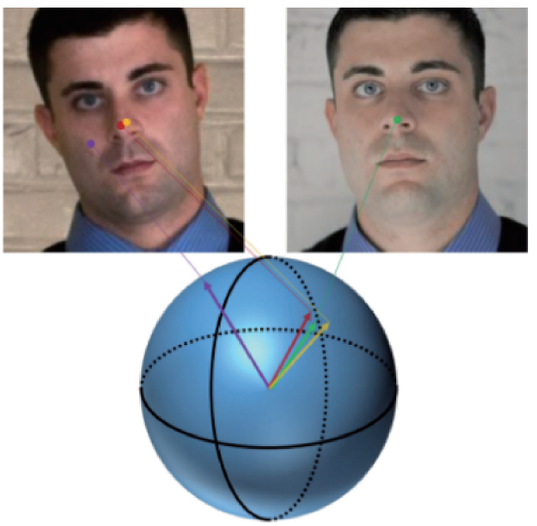


图 9　DFF 示例(Jiang 等,2018)

Fig. 9　Example for DFF(Jiang et al. ,2018)

Guo 等人(2019)利用 DFF 描述符和参数化人脸模型来估计人脸姿态和特征点。输入是有边界框的面部图像,从最初的 3 维人脸模型及其投影的 2 维特征点位置开始,根据 DFF 描述符计算目标特征点位置,根据从训练数据中学习到的一般下降方向更新相机参数,从而更新 3 维人脸模型,与目标特征点位置对齐,并重新计算其投影的 2 维特征点及其可见性,这个过程一直迭代到收敛。

现有的方法往往使用通用特征,如 SIFT 通过级联回归来确定 3DMM 的参数,然而,这些方法的重建精度通常不足。一方面,3DMM 的形状和表达式参数与图像纹理信息高度非线性,使得图像的映射难以估计;另一方面,SIFT 类型的类型描述符是基于局部面片的颜色信息设计的,对于人脸图像,SIFT 类型的类型描述符并没有利用特定的先验知识。因此,设计一个适合人脸图像的特征类型描述符,可以获得更好的对齐性能。

4　典型方法实验对比分析

4.1　评价指标

人脸特征点自动检测结果的评价指标包括平均误差(mean error)、标准化平均误差(normalized mean error, NME)(Kumar 等,2019)和曲线下面积(area under the curve, AUC)(Kumar 等,2019)。

平均误差指在等精度测量中,测得的所有测量值的随机误差的算术平均值,定义为

$$\eta = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| \tag{1}$$

式中,η 表示平均误差,x_i 和 $\bar{x}$ 分别表示特征点的预测值和真实值,n 表示测量次数,平均误差 η 即 $|x_i - \bar{x}|$ 的平均值。

单个人脸图像的标准化平均误差定义为

$$NME = \frac{1}{N}\sum_{j=1}^{N}\frac{\|p_j - \mu_j\|_2}{d} \times 100\% \tag{2}$$

式中,N 表示图片数量,p_j 和 μ_j 分别表示特征点真值和预测位置。根据下标不同,当计算 $NME_{\text{inter-ocular}}$ 时,设置 d 为外眼角之间的距离。当计算 NME_{box} 时,将 d 设置为提供的数据真值边界框的宽度 W 和高度 H 的几何平均值:$\sqrt{W_{\text{box}} \cdot H_{\text{box}}}$。如果没有提供真值边界框,则使用特征点的紧边界框。

曲线下面积 AUC 定义为

$$AUC_a = \int_0^a f(e)\,\mathrm{d}e \tag{3}$$

式中，a 表示所取上限，e 表示归一化误差，$f(e)$ 表示累计误差分布函数，AUC_{box} 将 d 设置为图片边界框的宽度和高度。要计算 AUC，首先绘制 ROC（receiver operating characteristic）曲线，AUC 就是通过计算该曲线下的面积进行评估，该方法不像平方误差一样受单个点误差较大而发生较大变化。

4.2 常见方法定位精度对比

不同实验之间特征点标定的位置和个数不尽相同，经典方法如 DCNN 主要对左眼、右眼、鼻尖以及左右嘴角进行实验，不同方法在 5 个特征点间的平均误差对比如图 10 所示。不同方法在 300W-Common、300W-Challenge、300W-Full、CelebA 和 Menpo 等数据库上的 $NME_{\text{inter-ocular}}$、$NME_{\mathrm{box}}$ 和 AUC_{box} 的对比如表 3 所示。从表 3 可以看出，级联的方法在速度和性能上找到一个平衡点，可以实行由粗到细的训练，然而级联的方法较大的缺点是依赖初始化。端

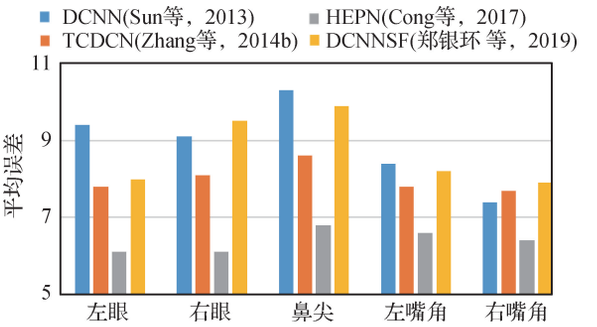


图 10 不同方法在 5 个特征点上的平均误差对比

Fig. 10 Comparison of mean errors of five feature points among different methods

表 3 不同方法的 $NME_{\text{inter-ocular}}$、$AUC_{\mathrm{box}}$ 和 NME_{box} 比较

Table 3 Comparison of $NME_{\text{inter-ocular}}$, AUC_{box} and NME_{box} among different methods

/%

方法	$NME_{\text{inter-ocular}}$				AUC_{box}		NME_{box}	
	常用数据集（common）	挑战数据集（challenge）	全数据集（full）	300W-All	300W	Menpo	300W	Menpo
DAN（Kowalski 等，2017）	3.19	5.24	3.59	–	55.33	57.07	–	–
DU-Net（Tang 等，2020）	2.97	5.53	3.47	–	–	–	–	–
PCD-CNN（Kumar 和 Chellappa，2018）	3.67	7.62	4.44	–	–	–	–	–
DeCaFA（Dapogny 等，2019）	2.93	5.26	3.39	4.62/2.10（CelebA）	66.1	–	–	–
DVLN（Wu 和 Yang，2017）	3.79	7.15	4.45	10.84	–	–	–	–
Wing（Feng 等，2018）	–	–	–	–	53.5	–	3.56	–
UGLLI（Kumar 等，2019）	2.78	5.08	3.23	–	68.27	69.85	2.24	2.20
FAN（Bulat 和 Tzimiropoulos，2017）	–	–	–	–	66.9	67.4	2.56	2.32
LUVLi（Kumar 等，2020）	2.76	5.16	3.23	–	68.3	70.1	2.24	2.28
CFAN（Zhang 等，2014a）	5.50	16.78	7.69	–	–	–	–	–
RAR（Xiao 等，2016）	4.12	8.35	4.94	–	–	–	–	–
TCDCN（Zhang 等，2014b）	4.80	8.60	5.54	–	–	–	–	–
Chandran 等人（2020）	2.83	7.04	4.23	–	–	–	–	–
3DDFA（Zhu 等，2016）	6.15	10.59	7.01	10.59（IBUG）	–	–	–	–
3FabRec（Browatzki 和 Wallraven，2020）	3.36	5.74	3.82	–	–	–	–	–
HyperFace（Ranjan 等，2019）	–	–	–	10.88（IBUG）	–	–	–	–
DLIN（Gilani 等，2017）	–	–	–	3.0（FRGCv2）	–	–	–	–
DFF（Jiang 等，2018）	–	–	–	3.14（AFLW）	–	–	–	–

注：“–”表示原文中并未给出实验结果。IBUG 为 Intelligent Behaviour Understanding group。

到端的方法节省时间和人力,UGLLI 加入特征点的不确定性分析,LUVLi 在 UGLLI 基础上改进网络,增加可见性估计器网络,使整个算法更加鲁棒精确。这两种方法较其他方法更为精确。RAR 算法作为一种端到端的可训练模型,用于无约束条件下进行特征点检测。PCD-CNN 则更能适应不同数据集。HyperFace 利用任务之间的协同作用提高任务能力。DeCaFA 通过生成特征点注意图,加权中间监督,取得了比较高的性能。在处理头部姿态变化大的人脸图像的方法中,FAN 引入分层、并行和多尺度块,TCDCN 通过停止任务和联合各个任务来学习,Wing(Feng 等,2018)方法则是提出机翼损失来提高精确度,DVLN 则更能适应不同数据集的不同特征和分布,而引入特征点热图的 DAN 和 DU-NET 更加鲁棒。3FabRec 的优点是能够在非常小的训练集上保持极高的精确度,较同类型的 CFAN 精确度更高。Chandran 等人(2020)在 FAN 的基础上提出的方法更适合高分辨率的图像且精确率较高。针对 3 维数据的方法中,DLIN 基于深度信息,在 FRGCv2 数据集的 $NME_{\text{inter-ocular}}$ 为 3.0%,能较准确地找到特征点位置,只是基于深度信息的方法依赖于初始化且丢失大量信息。在基于 3 维形变模型的方法中,DFF 方法的特点是使用 DFF 将每个像素映射到一个高维点,利用 DFF 描述符代替 SIFT 特征,在 AFLW 数据集上取得了较为鲁棒的效果。3DDFA 利用 3D 可变形模型,不仅解决大型姿态拟合中的高非线性问题,还可以提供丰富的训练样本。

5 存在的问题及发展趋势

通过对各种方法的分析与比较,对深度学习技术在人脸特征点自动定位应用中的现状、存在问题及发展趋势从 2 维和 3 维两方面进行系统归纳。

5.1 在 2 维人脸特征点自动标定方面

在 2 维图像数据中,具有真实特征点标签的训练数据非常丰富。针对 2 维数据的特征点自动定位方法的研究已比较深入,包括级联卷积神经网络、深度端到端回归网络、自动编码器网络及其他改进 CNN 的网络,已取得较为理想的效果,并在人脸识别和人脸编辑等实际任务中广泛应用。存在的主要问题和未来发展趋势包括以下方面:

1)从特征提取角度入手,使用自动学习更符合人脸图像的特征来优化特征点定位性能。

2)端到端的学习不需要手动标注但需要依靠大量数据,适用于大规模的数据集。所以端到端的方法可以从数据增强、数据生成以及如何构建小样本深度学习网络结构方面继续研究。

3)自动编码器依据特定的样本进行训练,因此其适用性很大程度上局限于与训练样本相似的数据且很容易过拟合,如何降低时间复杂度、避免陷入局部最小值是亟待解决的问题。

4)对于姿态变化较大的图像,如何更好地处理姿态、初始化以及减少计算成本仍然值得探讨。

综上所述,针对 2 维数据的人脸特征点自动标定研究,需要从网络结构、数据增强、特征选取和初始化等方面创新。同时,减少数据集的局限性、避免过拟合和减少时间复杂度也是需要考虑的问题。

5.2 在 3 维人脸特征点自动标定方面

目前,3 维数据在深度学习网络中的表示方法主要有两类。一类方法主要的解决思路是将 3 维数据转换成 2 维数据表示,如深度图投影、柱状图投影等,但这些方法都不可避免存在信息丢失问题,采用此种方法自动标定的 3 维人脸特征点的精度还有待提高。另一类方法是基于 3 维形变模型 3DMM 的方法,但 3DMM 的表示能力受数据的影响较大,在很多情况下表示能力受限。因而,对于 3 维人脸特征点自动标定问题,总体来说,目前研究尚处于探索阶段,未来主要需要解决的问题和研究方向主要包括以下方面:

1)探索适用于深度学习 3 维人脸特征点标定的 3 维模型的有效表示方法。

2)设计适用于 3 维人脸特征的深度学习网络结构和优化方法。

3)充分利用 3 维人脸数据的先验特征,研究适合 3 维人脸数据特征的类型描述符以获得更好的标定结果。

4)研究 3 维人脸数据的数据增强方法、3 维人脸数据生成以及 3 维人脸数据预处理方法等,为基于深度学习的 3 维人脸特征点标定提供充足有效的数据。

5)研究基于小样本以及无监督的深度学习3 维人脸特征点自动标定方法,从方法上提供对 3 维人脸数据缺乏问题的应对方案。

综上所述，针对 3 维数据的研究可以继续从模型表示、网络结构优化和特征描述符等方面创新，同时，数据预处理和数据增强同样也是需要继续考虑的问题。

6 结 语

针对基于深度学习的人脸特征点自动定位问题，本文在分析人脸特征点自动定位的含义和 2 维、3 维数据特征点自动标定的各种方法优劣的基础上，对比各种方法在公开人脸数据集上的性能表现，归纳出深度学习技术在人脸特征点自动定位应用中的现状、存在问题及发展趋势。针对 2 维数据的特征点自动定位方法的研究已比较深入，后续可以继续从网络结构、数据增强、特征选取和初始化等方面创新，同时减少数据集的局限性、避免过拟合和减少时间复杂度也是需要继续考虑的问题。与 2 维人脸数据相比，3 维人脸数据缺乏人脸形状、种族和表情等显著变化的训练数据。3 维人脸特征点的标定依然存在手工干预、标定特征点个数少或不准确以及标定时间长等问题，3 维数据在模型表示、数据处理和样本数量上都存在挑战，直接在 3 维模型上进行特征点定位的方法还有待进一步探索和研究，3 维人脸模型上特征点定位的精度和速度方面也有待提高。未来基于深度学习的 3 维人脸特征点定位的研究将逐渐成为研究趋势。

参考文献(References)

Belhumeur P N, Jacobs D W, Kriegman D J and Kumar N. 2013. Localizing parts of faces using a consensus of exemplars. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(12): 2930-2940 [DOI: 10.1109/TPAMI.2013.23]

Blanz V and Vetter T. 2002. A morphable model for the synthesis of 3d faces//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. San Antonio, USA: ACM Press/Addison-Wesley Publishing Co.: 187-194 [DOI: 10.1145/311535.311556]

Browatzki B and Wallraven C. 2020. 3FabRec: fast few-shot face alignment by reconstruction//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 6109-6119 [DOI: 10.1109/CVPR42600.2020.00615]

Bulat A and Tzimiropoulos G. 2017. Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 3726-3734 [DOI: 10.1109/ICCV.2017.400]

Cao X D, Wei Y C, Wen F and Sun J. 2014. Face alignment by explicit shape regression. International Journal of Computer Vision, 107(2): 177-190 [DOI: 10.1007/s11263-013-0667-3]

Chandran P, Bradley D, Gross M and Beeler T. 2020. Attention-driven cropping for very high resolution facial landmark detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 5860-5869 [DOI: 10.1109/CVPR42600.2020.00590]

Cong W L, Zhao S Y, Tian H and Shen J B. 2017. Improved face detection and alignment using cascade deep convolutional network [EB/OL]. [2020-07-17]. https://ui.adsabs.harvard.edu/abs/2017arXiv170709364C

Cootes T F, Edwards G J and Taylor C J. 2001. Active appearance models. IEEE Transactions on pattern analysis and machine intelligence, 23(6): 681-685 [DOI: 10.1109/34.927467]

Cootes T F and Taylor C J. 1992. Active shape models-smart snakes//Hogg D and Boyle R, eds. British Machine Vision Conference. Edinburgh, UK: Springer: 266-275 [DOI: 10.1007/978-1-4471-3201-1_28]

Cristinacce D and Cootes T. 2008. Automatic feature localisation with constrained local models. Pattern Recognition, 41(10): 3054-3067 [DOI: 10.1016/j.patcog.2008.01.024]

Dapogny A, Cord M and Bailly K. 2019. DeCaFA: deep convolutional cascade for face alignment in the wild//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6892-6900 [DOI: 10.1109/ICCV.2019.00699]

Feng Z H, Kittler J, Awais M, Huber P and Wu X J. 2018. Wing loss for robust facial landmark localization with convolutional neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2235-2245 [DOI: 10.1109/CVPR.2018.00238]

Gilani S Z, Mian A and Eastwood P. 2017. Deep, dense and accurate 3d face correspondence for generating population specific deformable models. Pattern Recognition, 69: 238-250 [DOI: 10.1016/j.patcog.2017.04.013]

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative Adversarial Networks. Advances in Neural Information Processing Systems. 3: 2672-2680 [EB/OL]. [2020-07-17]. https://arxiv.org/pdf/1406.2661v1.pdf

Guo Y D, Zhang J Y, Cai J F, Jiang B Y and Zheng J M. 2019. CNN-Based real-time dense face reconstruction with inverse-rendered photo-realistic face images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(6): 1294-1307 [DOI: 10.1109/TPAMI.2018.2837742]

He K, Zhang X, Ren S and Sun J. 2016. Deep residual learning for image recognition//Proceedings of Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Jiang B Y, Zhang J Y, Deng B L, Guo Y D and Liu L G. 2018. Deep face feature for face alignment. Computer Vision and Pattern Recognition [EB/OL]. [2020-07-17]. https://arxiv.org/pdf/1708.02721.pdf

Karim F, Majumdar S, Darabi H and Chen S. 2017. LSTM fully convolutional networks for time series classification. IEEE Access, 6: 1662-1669 [DOI: 10.1109/ACCESS.2017.2779939]

Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 4396-4405 [DOI: 10.1109/CVPR.2019.00453]

Kelkboom E J C, Gökberk B, Kevenaar T A M, Akkermans A H M and van der Veen M. 2007. 3d face: biometric template protection for 3d face recognition//Proceedings of International Conference on Biometrics. Seoul: Korea (South): Springer: 4642: 566-573 [DOI: 10.1007/978-3-540-74549-5_60]

Köstinger M, Wohlhart P, Roth P M and Bischof H. 2011. Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). Barcelona, Spain: IEEE: 2144-2151 [DOI: 10.1109/ICCVW.2011.6130513]

Kowalski M, Naruniec J and Trzcinski T. 2017. Deep alignment network: a convolutional neural network for robust face alignment//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, USA: IEEE: 2034-2043 [DOI: 10.1109/CVPRW.2017.254]

Kumar A and Chellappa R. 2018. Disentangling 3D pose in a dendritic CNN for unconstrained 2D face alignment//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 430-439 [DOI: 10.1109/CVPR.2018.00052]

Kumar A, Marks T K, Mou W X, Feng C and Liu X M. 2019. UGLLI face alignment: estimating uncertainty with Gaussian log-likelihood loss//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea (South): IEEE: 778-782 [DOI: 10.1109/iccvw.2019.00103]

Kumar A, Marks T K, Mou W X, Wang Y, Jones M, Cherian A, Koike-Akino T, Liu X M and Feng C. 2020. LUVLi face alignment: estimating landmarks' location, uncertainty, and visibility likelihood//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 8233-8243 [DOI: 10.1109/CVPR42600.2020.00826]

Le V, Brandt J, Lin Z, Bourdev L D and Huang T S. 2012. Interactive facial feature localization//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer: 679-692 [DOI: 10.1007/978-3-642-33712-3_49]

Lecun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]

Li Q Y, Ji Q G and Hong S D. 2019. FastFace: a real-time robust algorithm for face detection. Journal of Image and Graphics, 24(10): 1761-1771 (李启运, 纪庆革, 洪赛丁. 2019. FastFace: 实时鲁棒的人脸检测算法. 中国图象图形学报, 24(10): 1761-1771) [DOI: 10.11834/jig.180662]

Liu Z W, Luo P, Wang X G and Tang X O. 2015. Deep learning face attributes in the wild//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 3730-3738 [DOI: 10.1109/ICCV.2015.425]

Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]

Lowe D G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2): 91-110 [DOI: 10.1023/B:VISI.0000029664.99615.94]

Mirjalili V, Raschka S, Namboodiri A and Ross A. 2018. Semi-adversarial networks: convolutional autoencoders for imparting privacy to face images//Proceedings of 2018 International Conference on Biometrics (ICB). Gold Coast, Australia: IEEE: 82-89 [DOI: 10.1109/ICB2018.2018.00023]

Newell A, Yang K Y and Deng J. 2016. Stacked hourglass networks for human pose estimation//Proceedings of European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 483-499 [DOI: 10.1007/978-3-319-46484-8_29]

Ranjan R, Patel V M and Chellappa R. 2019. Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1): 121-135 [DOI: 10.1109/TPAMI.2017.2781233]

Ren S, He K, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]

Sagonas C, Antonakos E, Tzimiropoulos G, Zafeiriou S and Pantic M. 2016. 300 faces in-the-wild challenge: database and results. Image and Vision Computing, 47: 3-18 [DOI: 10.1016/j.imavis.2016.01.002]

Sun Y, Wang X G and Tang X O. 2013. Deep convolutional network cascade for facial point detection//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 3476-3483 [DOI: 10.1109/CVPR.2013.446]

Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan

D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]

Tang Z Q, Peng X, Li K and Metaxas D N. 2020. Towards efficient U-nets: a coupled and quantized approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(8): 2038-2050 [DOI: 10.1109/TPAMI.2019.2907634]

Terada T, Chen Y W and Kimura R. 2018. 3D Facial Landmark Detection Using Deep Convolutional Neural Networks//Proceedings of the 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery. Huangshan, China: IEEE: 390-393 [DOI:10.1109/FSKD.2018.8687254]

Wu W Y and Yang S. 2017. Leveraging intra and inter-dataset variations for robust face alignment//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, USA: IEEE: 2096-2105 [DOI: 10.1109/CVPRW.2017.261]

Xiao S T, Feng J S, Xing J L, Lai H J, Yan S C and Kassim A. 2016. Robust facial landmark detection via recurrent attentive-refinement networks//Proceedings of European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 57-72 [DOI: 10.1007/978-3-319-46448-0_4]

Xu X and Kakadiaris I A. 2017. Joint head pose estimation and face alignment framework using global and local CNN features//Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017). Washington, USA: IEEE: 642-649 [DOI: 10.1109/FG.2017.81]

Yi D, Lei Z, Liao S and Li S Z. 2014. Learning face representation from scratch[EB/OL]. [2020-07-17]. https://arxiv.org/pdf/1411.7923.pdf

Yin B C, Sun Y F, Wang C Z and Gai Y. 2009. BJUT-3D large scale 3D face database and information processing. Journal of Computer Research and Development, 46(6): 1009-1018 (尹宝才, 孙艳丰, 王成章, 盖赟. 2009. BJUT-3D 三维人脸数据库及其处理技术. 计算机研究与发展, 46(6): 1009-1018)

Yu X, Huang J Z, Zhang S T, Yan W and Metaxas D N. 2013. Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE: 1944-1951 [DOI: 10.1109/ICCV.2013.244]

Zhang B X and Su G D. 2013. Studies on human face imaging properties and the goals of face normalization. Journal of Optoelectronics · Laser, 14(4): 406-410 (章柏幸, 苏光大. 2013. 人脸成像特性研究及人脸归一化的目标. 光电子 · 激光, 14(4): 406-410) [DOI: 10.3321/j.issn:1005-0086.2003.04.020]

Zhang C Q, Liu Y Q and Fu H Z. 2020. AE2-nets: autoencoder in autoencoder networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 2572-2580 [DOI: 10.1109/CVPR.2019.00268]

Zhang J, Shan S G, Kan M N and Chen X L. 2014a. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment//Proceedings of European Conference on Computer Vision. Zurich, Switzerland: Springer: 1-16 [DOI: 10.1007/978-3-319-10605-2_1]

Zhang M, Lyu X Q, Wu L and Yu D H. 2018. Multiplicative denoising method based on deep residual learning. Laser and Optoelectronics Progress, 55(3): 197-203 (张明, 吕晓琪, 吴凉, 喻大华. 2018. 基于深度残差学习的乘性噪声去噪方法. 激光与光电子学进展, 55(3): 197-203) [DOI: 10.3788/LOP55.031004]

Zhang Z P, Luo P, Loy C C and Tang X O. 2014b. Facial landmark detection by deep multi-task learning//Proceedings of European Conference on Computer Vision. Zurich, Switzerland: Springer: 94-108 [DOI: 10.1007/978-3-319-10599-4_7]

Zheng W M, Tang H, Lin Z C and Huang T S. 2009. A novel approach to expression recognition from non-frontal face images//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE: 1901-1908 [DOI: 10.1109/ICCV.2009.5459421]

Zheng Y H, Wang B Z, Wang J J, Chen L Y and Hong Q Q. 2019. Research on deep convolution neural network with small filter used in facial landmark detection. Computer Engineering and Applications, 55(4): 173-178 (郑银环, 王备战, 王嘉珺, 陈凌宇, 洪清启. 2019. 深度卷积神经网络应用于人脸特征点检测研究. 计算机工程与应用, 55(4): 173-178) [DOI: 10.3778/j.issn.1002-8331.1710-0280]

Zhou E J, Fan H Q, Cao Z M, Jiang Y N and Yin Q. 2013. Extensive facial landmark localization with coarse-to-fine convolutional network cascade//Proceedings of 2013 IEEE International Conference on Computer Vision Workshops. Sydney, Australia: IEEE: 386-391 [DOI: 10.1109/ICCVW.2013.58]

Zhu X X and Ramanan D. 2012. Face detection, pose estimation, and landmark localization in the wild//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 2879-2886 [DOI: 10.1109/CVPR.2012.6248014]

Zhu X Y, Lei Z, Liu X M, Shi H L and Li S Z. 2016. Face alignment across large poses: a 3D solution//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 146-155 [DOI: 10.1109/CVPR.2016.23]

作者简介

徐亚丽，1996 年生，女，硕士研究生，主要研究方向为计算机图形学。
E-mail: 2398914081@qq.com

赵俊莉,通信作者,女,副教授,硕士生导师,主要研究方向为计算机图形学、计算机视觉、虚拟现实。
E-mail: zhaojl@yeah.net

吕智涵,男,副教授,主要研究方向为虚拟现实与增强现实。
E-mail: lvzhihan@gmail.com
张志梅,女,副教授,主要研究方向为深度学习、计算机视觉。
E-mail: zhangzm@qdu.edu.cn
李劲华,男,教授,硕士生导师,主要研究方向为软件工程。
E-mail: lijh@qdu.edu.cn
潘振宽,男,教授,博士生导师,主要研究方向为多体系统动力学与控制,虚拟现实技术,计算机视觉,图像科学。
E-mail:zkpan@qdu.edu.cn