# 深度卷积神经网络应用于人脸特征点检测研究

郑银环[1],王备战[2],王嘉珺[2],陈凌宇[2],洪清启[2]
1.厦门工学院 计算机科学与工程系,福建 厦门 361021
2.厦门大学 软件学院,福建 厦门 361001

**摘 要:**为解决在复杂环境下,如姿势不同、光照条件以及遮挡等因素导致传统人脸特征点检测算法的精度大幅度下降的问题,在特征点检测理论知识以及研究现状的基础上,针对传统卷积神经网络模型在处理人脸特征点检测问题时的不足之处,提出基于小滤波器的深卷积神经网络。算法引入小滤波器思想和以拓展“网络深度”优先的深层卷积神经网络模型,针对人脸特征点检测重新设计训练,提高了算法的有效性与适用性。通过将算法应用于ALFW和AFW人脸数据集上预测5点人脸特征点问题,并与其他多个经典算法进行对比分析,结果表明:基于小滤波器的深卷积神经网络在预测人脸5点特征点问题上有更好的准确性和鲁棒性。

**关键词:**小滤波器;深卷积神经网络;特征点检测;网络深度
**文献标志码:**A **中图分类号:**TP391.41 doi:10.3778/j.issn.1002-8331.1710-0280

郑银环,王备战,王嘉珺,等.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,2019,55(4):173-178.
ZHENG Yinhuan, WANG Beizhan, WANG Jiajun, et al. Research on deep convolution neural network with small filter used in facial landmark detection. Computer Engineering and Applications, 2019, 55(4):173-178.

## Research on Deep Convolution Neural Network with Small Filter Used in Facial Landmark Detection

ZHENG Yinhuan[1], WANG Beizhan[2], WANG Jiajun[2], CHEN Lingyu[2], HONG Qingqi[2]
1.Department of Computer Science and Engineering, Xiamen Institute of Technology, Xiamen, Fujian 361021, China
2.Software School, Xiamen University, Xiamen, Fujian 361001, China

**Abstract:** To solve the problem in a complex environment, such as different positions, light conditions, and barrier factors which lead to a big drop in precision by using traditional facial feature point detection algorithms, the research on the basis of theoretical knowledge is made, and the deep convolution neural network based on small filter is put forward. The algorithm introduces the small filter thought and depth-first network into deep convolution neural network model, redesigns the training in view of the facial feature points detection, and improves the effectiveness and applicability of the algorithm. By applying the algorithm in ALFW and AFW face datasets to predict five points of facial feature, and compared with several other classical algorithms analysis results, it shows that the deep convolution neural network based on the small filter in the prediction of facial feature points at five issues has better accuracy and robustness.

**Key words:** small filter; deep convolution neural network; feature point detection; depth of network

## 1 引言

随着标记数据和GPU的发展,卷积神经网络缺少训练数据和计算能力较低的问题已经不复存在,卷积神经网络在人脸识别方面的研究得到学者的广泛关注,特别是可以提取高级图像特征的卷积神经网络已经成功地应用于许多计算机视觉任务中,例如人脸验证、图像分类和对象检测。2013年,Sun等人提出了深卷积级联网络[1],北京旷视科技设计了一个初始化大型面部地标

**基金项目:**福建省中青年教师教育科研项目(No.JZ160238);国家自然科学基金(No.61502402);福建省自然科学基金(No.2015J05129)。
**作者简介:**郑银环(1984—),女,讲师,研究领域为数据挖掘、数据分析、高等教育,E-mail:yhzheng1218@gmail.com;王备战(1965—),通讯作者,男,博士,教授,研究领域为数据仓库、数据挖掘、图像处理、软件体系结构;王嘉珺(1993—),女,研究生,研究领域为数据挖掘、机器学习;陈凌宇(1993—),男,工程师,研究领域为机器学习、数据挖掘、图形图像处理;洪清启(1983—),男,博士,副教授,研究领域为医学图像处理、三维重建、人体器官建模、生物信息学等。
**收稿日期:**2017-10-27 **修回日期:**2017-12-11 **文章编号:**1002-8331(2019)04-0173-06
CNKI**网络出版:**2018-04-10, http://kns.cnki.net/kcms/detail/11.2127.TP.20180410.1055.006.html

由粗到精的四级级联卷积神经网络[2]。相较于传统的卷积神经网络它很好地解决了传统的卷积网络在训练广泛面部标志定位任务中的问题。

2015年,香港大学Zhu等人设计了一个基于粗略到精细形状搜索的新颖的面部对齐级联回归框架[3]。不同于以初始形状开始并以级联方式细化形状的常规级联回归方法,该框架一开始在包含不同形状的空间上进行粗略搜索,并且采用粗略解来约束随后的形状搜索,其独特的逐级渐进和自适应搜索,防止由于较差的初始化导致最终解在局部最优中被捕获,这极大改善了级联线性回归在应对大姿态变化时的鲁棒性。

2015年,中山大学数据科学与计算机学院赖剑煌教授等人提出了一种基于深卷积神经网络(Convolution Neural Network,CNN)的面部对齐的级联回归框架[4]。在大多数现有的级联回归方法中,形状索引的特征通过手工制作的视觉描述符获得,或者从浅层模型中获得。此设置对于脸部对齐任务可能不是最佳的。为了解决这个问题,Lai等人提出了端到端CNN架构来学习高度辨别形状索引的特征。通过将图像编码为相同大小的图像的高级特征图,然后从这些高级描述符中提取深度特征,学习高度辨别形状索引的特征,从而形成一种新颖的“形状索引池”方法。对多个数据集进行的广泛评估实验表明,所提出的深度框架对现有技术方法有着显著的改进。但是将网络直接应用于人脸对齐效果不是特别理想,原因如下:

(1)虽然在计算机视觉中使用现有CNN架构(例如AlexNet[5]、GoogleNet[6])进行调整便能够很好地适应当前的计算机视觉任务[7-8],但是这样的策略几乎不适用于人脸对齐问题。因为现成的大型网络通常被训练用于图像分类,而人脸对齐是结构预测问题。

(2)从头开始构建一个新的卷积神经网络预测人脸对齐问题,应该考虑到过拟合的问题,因此每个阶段的网络结构都需要根据本阶段的任务进行仔细设计。

针对以上提出的问题,根据文献[9]提出的在卷积层中使用小滤波器的思想和GoogleNet网络[6]提出的用更多的卷积、更深的层次可以得到更好的结构的思想,本文提出了一种基于小滤波器的深卷积神经网络(Deep Convolution Neural Network with Small Filter,DCNNSF),并用该网络解决人脸5点特征点(眼睛、鼻子、嘴巴)预测问题。通过添加更多的卷积层稳定地增加网络的深度,并且在所有层中使用非常小(3×3)的卷积滤波器,有效减小参数,避免了过拟合的问题,更好地解决了人脸特征点检测问题。

## 2　网络描述

基于小滤波器的深卷积神经网络是通过结合最近成功的网络,采用非常深的架构[6]、低维度表示和多重损失函数[10]等许多技巧构建的。小滤波器和非常深的网络架构可以减少参数数量并增强网络的非线性。低维度表示符合面部图像通常位于低维度歧管上的假设,而低维度约束可以降低网络的复杂性。

### 2.1　网络深度的提升

从GoogleNet网络中可知,增加卷积神经网络模型的深度或宽度是得到高质量模型最保险的做法,但是这种设计方法一般会遇到如下问题:

(1)使用更多的参数是得到大尺寸网络模型的前提,但是参数越多,网络越容易过拟合;

(2)网络的计算量与网络模型的复杂度成正比,随着网络模型复杂度的增加,网络的计算量随之增大。

考虑到使用相同数量的参数时,深度网络模型比宽度网络模型有更高的准确性和鲁棒性[11]。此外,由于普通计算机处理由宽度网络模型提取的高维特征相对较困难,而深层网络模型的每一层中特征的维度都相对比较低,这使得深层模型的存储器消耗是可负担的。因此,本文最终决定通过增加卷积神经网络的深度以提升深度卷积网络性能。

### 2.2　小型卷积滤波器

采用增加卷积神经网络的深度方法后,新的卷积神经网络模型要解决的主要问题是如何有效地减少网络的参数,避免过拟合,同时减少网络的计算量。本文使用多个小型卷积滤波器(卷积层)近似大型卷积滤波器。

虽然大尺寸的卷积核可以带来更大的感受野,但也意味着更多的参数,比如5×5卷积核参数是3×3卷积核的(5×5+1)/(3×3+1)=2.6倍。如图1所示,网络可以用两个连续的3×3卷积层(stride步长为1)组成的小网络代替单个的5×5卷积层,而三个连续的3×3叠加卷积层就能拥有7×7卷积层的感受野,这样在保持感受野范围的同时又减少了参数量。

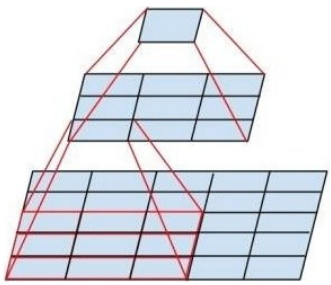

**图1　3×3卷积层替代5×5卷积层**

使用三个连续的3×3叠加卷积层替代7×7卷积层的优势体现在:

(1)网络的决策功能分辨性更高。新的网络结合了三个非线性卷积层,而不是一个单一的卷积层。

(2)有效减少参数的数量。假设三个3×3卷积层堆叠的输入和输出都有 $C$ 个通道,三个卷积层需要的参数为 $3\times(3^2\times C^2)=27C^2$ 位;同样情况下,单个7×7卷积层则需要 $7^2\times C^2=49C^2$ 位参数。

### 2.3　基于小滤波器的深卷积神经网络具体架构

在保证图片清晰度的情况下，将图片缩减为128×128×3大小的彩色图像，以达到缩减图片大小的目的。本文提出的DCNNSF网络包括11个卷积层、5个池层和3个全连接层，其详细结构如图2所示。网络中所有卷积层的大小都为3×3的，并且所有池化层都采用最大池化。DCNNSF在全局神经网络中使用3个全连接层，虽然这样增加了模型的参数，但是使得模型表达能力和非线性表达能力更强，卷积神经网络更加有效与鲁棒。

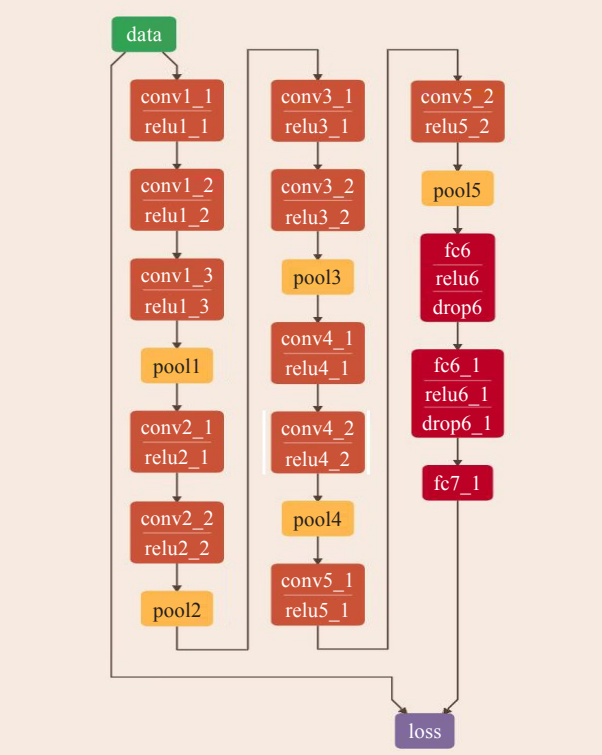


**图2　DCNNSF的网络结构图示**

其中DCNNSF网络参数配置如表1所示。

网络的I(Data)层为网络输入层，该层为整个网络输入一张128×128的三通道彩色图片。

网络的Conv1_1层为第一个卷积层(Convolution Layer)，该层由32个特征图谱(Feature Map)构成。特征图中的每个单元与输入层的一个3×3的相邻区域相连，卷积的输入区域大小是3×3，因此特征图是由32个大小为3×3的卷积核从输入图片中卷积得到，每个特征图谱内参数共享。其输出32个大小为128×128的特征图。

网络的Conv1_2层为第二个相连3×3卷积层，该层输入信息为Conv1_1层中输出的32个大小为128×128特征图信息，其输出64个大小为128×128的特征图。

网络的Conv1_3层为第三个相连3×3卷积层，该层输入信息为Conv1_2层中输出的64个大小为128×128特征图信息。它的作用是与Conv1_1层、Conv1_2层叠加，从而替代大型7×7卷积层，减少参数，并达到相同的卷积效果。该层最终将96个大小为128×128的特征图输出到下一层中。

**表1　DCNNSF网络参数配置**

| 层名称 | 类型 | 核大小/步长 | 深度 | 输出维度 |
|---|---|---|---|---|
| I(Data) | Input | | | |
| Conv1_1 | Convolution | 3×3/1 | 1 | 128×128×32 |
| Conv1_2 | Convolution | 3×3/1 | 1 | 128×128×64 |
| Conv1_3 | Convolution | 3×3/1 | 1 | 128×128×96 |
| Pool1 | Max pooling | 2×2/2 | 0 | 64×64×96 |
| Conv2_1 | Convolution | 3×3/1 | 1 | 64×64×64 |
| Conv2_2 | Convolution | 3×3/1 | 1 | 64×64×128 |
| Pool2 | Max pooling | 2×2/2 | 0 | 32×32×128 |
| Conv3_1 | Convolution | 3×3/1 | 1 | 32×32×96 |
| Conv3_2 | Convolution | 3×3/1 | 1 | 32×32×192 |
| Pool3 | Max pooling | 2×2/2 | 0 | 16×16×192 |
| Conv4_1 | Convolution | 3×3/1 | 1 | 16×16×128 |
| Conv4_2 | Convolution | 3×3/1 | 1 | 16×16×256 |
| Pool4 | Max pooling | 2×2/2 | 0 | 8×8×256 |
| Conv5_1 | Convolution | 3×3/1 | 1 | 8×8×160 |
| Conv5_2 | Convolution | 3×3/1 | 1 | 8×8×320 |
| Pool5 | Max pooling | 2×2/2 | 0 | 4×4×320 |
| Conv6 | Fully connection | | 1 | 320 |
| Conv6_1 | Fully connection | | 1 | 256 |
| Conv7 | Fully connection | | 1 | 10 |

网络的Pool1层为第一个池化层(Pooling Layer)，该层对输入的特征图进行压缩，首先可以使特征图变小，有效简化网络计算复杂度；其次对特征进行压缩，以便于提取主要特征。Pool1接收输入为Conv1_3层64个大小为128×128的特征图信息，使用最大池化(Max Pooling)，池化领域2×2最大领域，由于步长为2，所输出的特征图为$128/2=64$个，数据规模为原来的1/4，最终输出96个大小为128×128的特征图。

其他卷积层与池化层的具体操作同Conv1_1、Conv1_2、Conv1_3、Pool1层类似，每个大型的卷积层之间初始参数以32、64、96、128、160的方式递增，除了第一个大型的卷积层，其他大型卷积层中的3×3小卷积层之间的参数都以倍增的形式增加。因此Pool5层最终输出320个4×4的特征图。

Conv6_1、Conv6_2与Conv7层都为全连接层。Conv6_1层接收来自Pool5输出的特征图信息，将其转化为一维向量，用来表示整个面部点，并将结果输出到Conv6_2层中，使模型表达能力和非线性表达能力得到增强，最终通过Conv7层输出一维向量(如式(1)所示)，其中$x$、$y$代表10个人脸坐标信息。

$$L_{\mathrm{ALL}}=\{x_1,y_1,x_2,y_2,\cdots,x_5,y_5\} \tag{1}$$

## 3　实验

### 3.1　模型训练与评估指标

实验使用的大型面部属性数据集CelebA[12]拥有20多万张名人图像，每张图片均有5个手动地标位置，40个二进制属性。为了增强训练样本，进一步增加样本数

量,作者在实验时通过小的改动(包括平移、面内旋转)保存成新的图片。

评估指标:人脸特征点算法检测性能用每个面部点的平均检测误差和累积误差曲线(Cumulative Error Distribution,CED)进行测量。它们能清晰直观地表示算法的准确性和可靠性。

平均检测误差是预测的特征点坐标位置与通过眼间距离归一化的人脸特征点坐标实际位置之间的平均距离:

$$error=\frac{1}{N}\sum_{i=1}^{N}\frac{\frac{1}{M}\sum_{j=1}^{M}\left|P_{i,j}-g_{i,j}\right|_2}{\left|l_i-r_i\right|_2} \tag{2}$$

其中,$i$、$j$ 分别表示第 $i$ 张图片和 $j$ 个人脸特征点;$N$ 表示输入图片数量;$M$ 是人脸特征点总数;$P_{i,j}$ 是预测的人脸特征点坐标;$g_{i,j}$ 是真实特征点坐标;$l_i$ 和 $r_i$ 是左右眼瞳孔中心的位置。平均误差由估计的坐标与面部特征点坐标真实值之间的距离来测量,大于10%的平均误差报告为失败。

### 3.2　测试数据集

AFLW(Labeled Face Parts in the Wild 2011)数据库包含大量多姿态、多视角的大规模人脸,总共有25 993张已手工标注的人脸图片,每张图片标有21个地标。在ALFW数据库中,59%的图片为女性,剩下的为男性,并且拥有少部分的灰色图片。标注5点的ALFW数据集部分图示如图3所示。

图3　标注5点的ALFW数据集部分图示

AFW(Annotated Faces in the Wild 2012)数据库包含205个具有高度杂乱的背景和大面积尺寸和姿态的图像。每个图像都标有6个地标和相应面的边界框。包含复杂背景的AFW数据集图示如图4所示。

图4　包含复杂背景的AFW数据集图示

### 3.3　实验环境与参数设置

实验采用的参数设置如表2所示,实验的预测样本共160 000张图片,设置batch_size为100时一次处理完所有样本需要1 600次迭代,因此test_interval必须大于等于1 600(实验设置为2 000)。同理,实验有40 000张测试样本,batch_size为100,则test_iter值必须大于等于400(实验设置为500)。实验中设置学习率将随着迭代次数的增加慢慢地减少,学习率最初设定为0.01,当验证集精度停止改善时,gamma值设置为10,stepsize设置为100 000,因此,迭代次数达到100 000次的时候,学习率减少10倍。

表2　实验参数设置

| 参数 | 描述 | 值 |
|---|---|---|
| test_interval | 训练一遍迭代数 | 2 000 |
| test_iter | 测试一遍迭代数 | 500 |
| batch_size1 | 单次处理训练样本数 | 100 |
| batch_size2 | 单次处理测试样本数 | 100 |
| base_lr | 基础学习率 | 0.01 |
| lr_policy | 学习率变化规律 | step |
| gamma | 学习率变化指数 | 0.1 |
| stepsize | 学习率变化频率 | 100 000 |
| display | 屏幕显示间隔 | 100 |
| max_iter | 最大迭代次数 | 400 000 |
| momentum | 动量 | 0.9 |
| weight_decay | 权重衰减 | 0.000 5 |
| snapshot | 保存模型间隔 | 5 000 |
| solver_mode | Mode选择 | GPU |

## 4　实验结果与分析

因为AFLW数据库收集的是具有挑战性的姿势变化和遮挡的图片,DCNNSF方法与级联CNN方法选取相同的10 000个训练面进行实验更能体现效果的差别。实验表明,虽然DCNNSF和级联CNN[1]都是在CNN的基础上建立的,但是DCNNSF与联级CNN相比效果相近甚至部分位置有更好的检测精度,并且DCNNSF明显降低了计算成本。从图5可以看出,DCNNSF方法在两个地标中表现更好,两个地标中表现相近时,DCNNSF整体精度相对优于级联CNN。

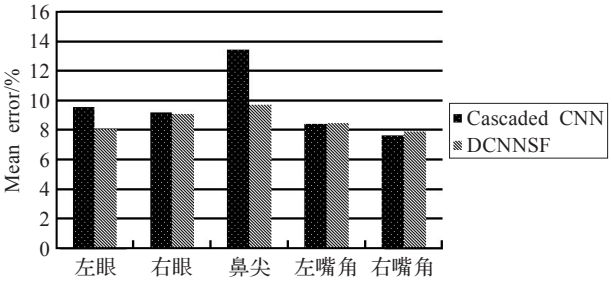


图5　DCNNSF同级联CNN的比较

为了进一步证明DCNNSF算法在人脸5点特征点预测上的有效性和鲁棒性,实验将DCNNSF算法分别与经典算法在ALFW和AFW数据集上的预测效果进行对比。这些经典算法包括树结构部分模型(TSPM)[13]、显式形状回归(ESR)[14]、级联变形模型(CDM)[15]、商业软件Luxand face SDK[16]、使用公开的实现和参数设置的鲁棒级联姿态回归(RCPR)[17]、监督下降法(SDM)[18]。在ALFW数据集上各算法的人脸5点特征点预测平均误差值如表3、图6、图7所示。

表4、图8、图9是各算法在图像数据集AFW中预测

表3 各算法在ALFW数据集上的平均误差

| 算法 | 左眼 | 右眼 | 鼻尖 | 左嘴角 | 右嘴角 | 总体 |
|---|---|---|---|---|---|---|
| TSPM | 14.46 | 9.49 | 22.39 | 19.12 | 19.27 | 15.93 |
| ESR | 11.59 | 11.06 | 12.13 | 12.47 | 15.01 | 13.01 |
| CDM | 10.01 | 10.83 | 14.91 | 19.65 | 14.97 | 13.09 |
| Luxand | 12.95 | 12.51 | 15.02 | 12.29 | 12.61 | 12.43 |
| RCPR | 11.83 | 10.64 | 18.36 | 11.23 | 11.93 | 11.61 |
| SDM | 8.14 | 9.12 | 11.21 | 6.84 | 7.22 | 8.53 |
| DCNNSF | 9.04 | 10.08 | 10.75 | 8.51 | 7.93 | 8.40 |

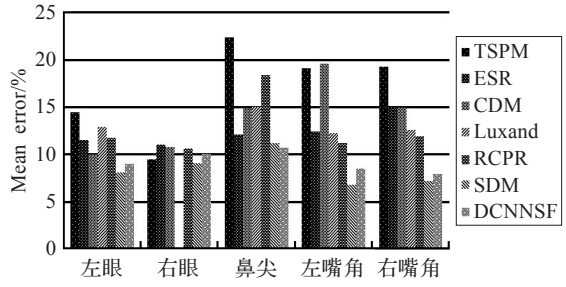


图6 各算法在ALFW数据集上5个特征点的平均误差

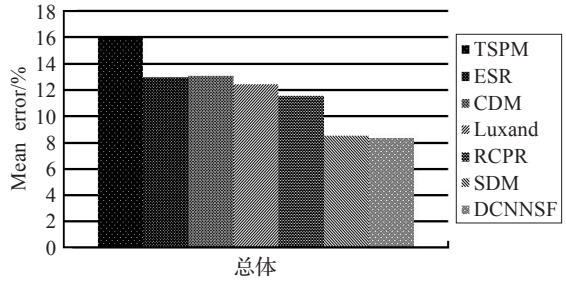


图7 ALFW数据集上各算法整体的平均误差

表4 各算法在AFW数据集上的平均误差

| 算法 | 左眼 | 右眼 | 鼻尖 | 左嘴角 | 右嘴角 | 总体 |
|---|---|---|---|---|---|---|
| TSPM | 10.49 | 9.24 | 19.27 | 15.17 | 17.22 | 14.28 |
| ESR | 11.01 | 12.91 | 16.98 | 10.26 | 10.17 | 12.27 |
| CDM | 8.47 | 8.57 | 14.79 | 11.01 | 12.83 | 11.13 |
| Luxand | 7.62 | 7.75 | 12.26 | 12.03 | 12.09 | 10.35 |
| RCPR | 7.93 | 7.59 | 12.72 | 8.09 | 10.31 | 9.33 |
| SDM | 7.89 | 8.23 | 11.49 | 7.12 | 9.27 | 8.80 |
| DCNNSF | 7.97 | 7.13 | 9.21 | 6.97 | 10.01 | 8.26 |

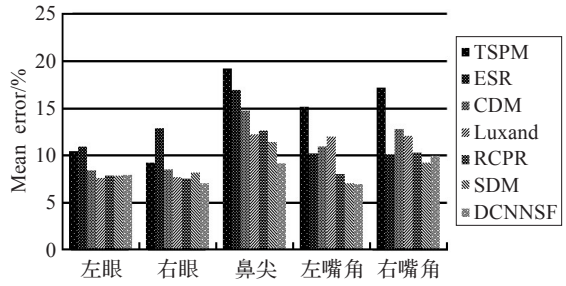


图8 各算法在AFW数据集上5个特征点的平均误差

5点特征点的平均误差值。由图表可知DCNNSF算法相较于其他经典算法的优势。

实验结果表明,在具有多姿态、多视角、不同光照以及高度杂乱背景的情况下,相较于大部分传统的经典算法,本文提出的DCNNSF算法预测效果有明显优势,并

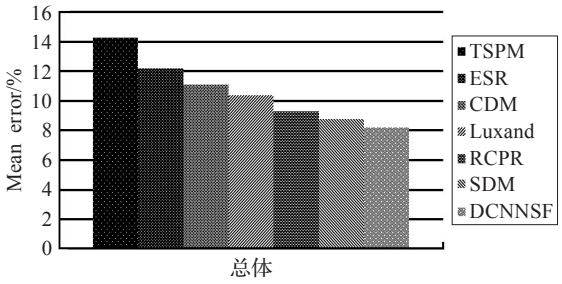


图9 AFW数据集上各算法整体的平均误差

且在计算量减少的基础上DCNNSF算法在5点预测上效果与级联CNN相近甚至更好。

图10和图11显示了DCNNSF算法在ALFW数据集和AFW数据集上检测的几个示例。实验选用的是具有较大面部姿态变化,照明和严重闭塞的面部的实例对模型进行检测。

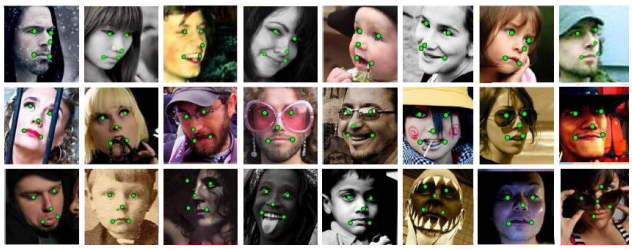

图10 ALFW数据集上的实例效果

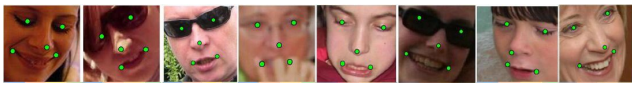

图11 AFW数据集上的实例效果

从实例图片中不难看出,在面对抬头、低头、侧脸等姿势变化,极度光照导致的人脸背光,以及戴墨镜或者被衣物遮挡等情况下,DCNNSF模型依然是有效和鲁棒的。

## 5 结束语

本文首先讲述了传统卷积神经网络在处理人脸对齐特征点定位问题时的不足,针对这些问题提出了一种新的基于小滤波器的深卷积神经网络(DCNNSF),并用该网络解决人脸5点特征点预测问题。通过添加更多的卷积层稳定地增加网络深度的同时有效减少参数,避免了过拟合的问题,使模型达到更高精度和鲁棒性。

## 参考文献:

[1] Sun Y,Wang X,Tang X.Deep convolutional network cascade for facial point detection[C]//Computer Vision and Pattern Recognition,2013:3476-3483.

[2] Zhou E,Fan H,Cao Z,et al.Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops,2013:386-391.

[3] Zhu S,Li C,Loy C C,et al.Face alignment by coarse-to-fine shape searching[C]//Computer Vision and Pattern Recognition,2015:4998-5006.
[4] Lai H,Xiao S,Cui Z,et al.Deep cascaded regression for face alignment[J].arXiv:1510.09083,2015.
[5] Krizhevsky A,Sutskever I,Hinton G E.Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems,2012:1097-1105.
[6] Szegedy C,Liu W,Jia Y,et al.Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015:1-9.
[7] Girshick R,Donahue J,Darrell T,et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2014:580-587.
[8] Zhang N,Donahue J,Girshick R,et al.Part-based R-CNNs for fine-grained category detection[C]//Computer Vision.Berlin:Springer,2014:834-849.
[9] Simonyan K,Zisserman A.Very deep convolutional networks for large-scale image recognition[J].arXiv:1409.1556,2014.
[10] Sun Y,Wang X,Tang X.Deep learning face representation by joint identification-verification[J].arXiv:1406.4773,2014.
[11] Sagonas C,Tzimiropoulos G,Zafeiriou S,et al.300 faces in-the-wild challenge:the first facial landmark localization challenge[C]//2013 IEEE International Conference on Computer Vision Workshops,2013:397-403.
[12] Celeb.Faces attributes dataset[EB/OL].http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html/.
[13] Zhu X,Ramanan D.Face detection,pose estimation,and landmark localization in the wild[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition,2012:2879-2886.
[14] Cao X,Wei Y,Wen F,et al.Face alignment by explicit shape regression[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition,2012:2887-2894.
[15] Yu X,Huang J,Zhang S,et al.Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model[C]//2013 IEEE International Conference on Computer Vision,2013:1944-1951.
[16] Luxand Incorporated.Luxand face SDK[EB/OL].http://www.luxand.com/.
[17] Burgos-Artizzu X P,Perona P,Dollar P.Robust face landmark estimation under occlusion[C]//2013 IEEE International Conference on Computer Vision,2013:1513-1520.
[18] Xiong X,De la Torre F.Supervised descent method and its applications to face alignment[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition,2013:532-539.

(上接第141页)

[4] Bhuyan M K,Neog D R,Kar M K.Hand pose recognition using geometric features[C]//2011 National Conference on Communications,2011:1-5.
[5] 陈一民,张云华.基于手势识别的机器人人机交互技术研究[J].机器人,2009,31(4):351-356.
[6] 黄振翔,彭波,吴娟,等.基于DTW与混合判别特征检测器的手势识别[J].计算机工程,2014,40(5):216-218.
[7] Krizhevsky A,Sutskever I,Hinton G E.ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems,2012:1097-1105.
[8] Girshick R,Donahue J,Darrell T,et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition,2014:580-587.
[9] Ren S,He K,Girshick R,et al.Faster R-CNN:towards real-time object detection with region proposal networks[C]//International Conference on Neural Information Processing Systems.[S.l.]:MIT Press,2015:91-99.
[10] Redmon J,Divvala S,Girshick R,et al.You only look once:unified,real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition,2016:779-788.
[11] Liu W,Anguelov D,Erhan D,et al.SSD:single shot multibox detector[C]//European Conference on Computer Vision.Springer International Publishing,2016:21-37.
[12] Howard A G,Zhu M,Chen B,et al.MobileNets:efficient convolutional neural networks for mobile vision applications[J].arXiv:1704.04861,2017.
[13] Zhang X,Zhou X,Lin M,et al.ShuffleNet:an extremely efficient convolutional neural network for mobile devices[C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:6848-6856.
[14] Chen S,Shrivastava A,Singh S,et al.Revisiting unreasonable effectiveness of data in deep learning era[C]//IEEE International Conference on Computer Vision,2017:843-852.
[15] Huang J,Rathod V,Sun C,et al.Speed/accuracy trade-offs for modern convolutional object detectors[C]//IEEE Conference on Computer Vision and Pattern Recognition,2017:7310-7311.