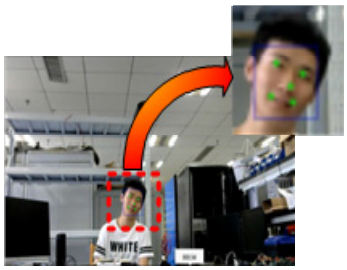
Opto-Electronic Engineering
光 电 工 程

Article

2020 年，第 47 卷，第 1 期

DOI: 10.12086/oee.2020.190299

基于深度学习检测器的多角度人脸关键点检测



赵兴文，杭丽君*，宫恩来，叶　锋，丁明旭

杭州电子科技大学自动化学院，浙江 杭州 310018

摘要：针对人脸关键点检测(人脸对齐)在应用场景下的速度和精度需求，首先在 SSD 基础之上融合更多分布均匀的特征层，对人脸框坐标进行级联预测，形成对于多尺度人脸信息均具有更加鲁棒响应的深度学习检测器 MR-SSD。其次在局部二值特征 LBF 的级联形状回归方法基础上，提出了基于面部像素差值的多角度初始化算法。采用端正人脸正负 90°倾斜范围内的五组特征点形状进行初始化，求取每组回归后形状的眼部特征点像素均方差值并以最大者对应方案作为最终回归形状，从而实现对多角度倾斜人脸优异的拟合效果。本文所提出的最优架构可以实时获得极具鲁棒性的人脸框坐标并且可实现对于多角度倾斜人脸的关键点检测。

关键词：深度学习；机器学习；人脸关键点检测；人脸对齐；像素差值

中图分类号：TP391　　**文献标志码：**A

引用格式：赵兴文，杭丽君，宫恩来，等. 基于深度学习检测器的多角度人脸关键点检测[J]. 光电工程，2020，**47**(1): 190299

Multi-angle key point detection of face based on deep learning detector

Zhao Xingwen, Hang Lijun*, Gong Enlai, Ye Feng, Ding Mingxu

College of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China

Abstract: In order to meet the speed and accuracy requirements of face key point detection (face alignment) in application scenarios, firstly, cascaded prediction is carried out on the basis of SSD (single shot multibox detector), which combines more uniformly distributed feature layers to form MR-SSD (more robust SSD), a deep learning detector with more robust response to multi-scale faces. Secondly, based on the cascade shape regression method of local binary feature (LBF), a multi-angle initialization algorithm based on the difference between the facial pixels is proposed. Five groups of feature points in the 90 degree inclination range of positive and negative face are initialized to achieve excellent fitting effect for inclined face under multi angles. The mean square deviation of each group of feature points after regression is calculated and the maximum corresponding shape is used as the final regression shape. The optimal architecture proposed in this paper can obtain robust face bounding box and face alignment schemes against multi-angle tilt in real time.

Keywords: deep learning; machine learning; face keypoint detection; face alignment; pixel difference

Citation: Zhao X W, Hang L J, Gong E L, *et al*. Multi-angle key point detection of face based on deep learning detector[J]. *Opto-Electronic Engineering*, 2020, **47**(1): 190299

收稿日期：2019-05-31；　收到修改稿日期：2019-09-17

基金项目：国家自然科学基金资助项目(51777049)；青年科学基金资助项目(51707051)

作者简介：赵兴文(1995-)，男，硕士，主要从事深度学习、计算机视觉的研究。E-mail：18405813956@163.com

通信作者：杭丽君(1979-)，女，博士，教授，主要从事电力电子与电力传动、深度学习的研究。E-mail：ljhang@hdu.edu.cn

1 引 言

深度学习技术的引进和机器学习技术的成熟使得计算机视觉相关任务取得巨大进展，从而促进了众多检测以及定位领域的应用技术趋于完善。人脸关键点检测，即人脸对齐，在这众多应用领域中与目标检测、回归定位息息相关[1]。作为人脸检测任务的延伸和拓展，同时又作为人脸校准和人脸识别任务的基础，关键点检测具有举足轻重的意义。除去人脸识别范畴的探究，人脸对齐技术在多个领域皆有建树。如在表情识别中，人脸对齐为情绪识别的研究提供了可能性[2]。又如具备美化图片功能的应用，包括人脸磨皮美化功能、动态变脸特效等等，需要以人脸对齐技术得到面部特征点或者特征区域从而进行相关操作，这就要求人脸对齐技术能够快速且精准地实现特征点定位[3]。

人脸对齐算法实现方案众多，设计思路由人脸检测架构与人脸对齐技术两部分组合而成。目前人脸检测的主流实现方案集中在深度学习领域中，如 two stage 网络的经典架构 RCNN[4]，Fast R-CNN[5]，Faster R-CNN[6]，通过多种候选框生成方案获得深度模型，比如 SS[7]、RPN 等得到感兴趣区域，然后再将其投入分类网络进行打分；另一类是实时速度较快的 one-stage 网络，如 SSD[8]、YOLO 系列[9]省去了候选框生成步骤，对全图提取特征完成分类和坐标框的回归，在保证目标框的坐标回归精度的同时也满足了在实时应用场景下的速度要求，从而使该类模型拥有较高的速度-精度平衡。深度网络在检测领域取得的精度极高，在目标检测应用领域有不可撼动的地位。

人脸对齐算法实现方案也多样化，Cao 等[10]提出 ESR(explicit shape regression)方案，对显示形状进行回归。采用 SDM 算法[11](supervised descent method)实现非线性最小二乘(non-linear least squares)目标函数，使其以非常快的速度收敛到最小值。LBF[12]方案采用提取局部二值特征进行回归，大幅提升了关键点位置定位速度。PNMS 方案[13]引入非连续的线性函数和基于高斯分布的连续函数，改进非极大值抑制算法，对候选窗口重打分，获取精度与速度的提升。以上方案均采用相比深度学习模型较小的轻量级模型，在深度学习架构方案中，Zhang 等[14]提出代表性的 MTCNN (multi-task convolutional neural network)架构采用深度级联网络，利用人脸检测和人脸对齐的内在联系同时提升这两个任务的性能，其采用统一的三阶段级联 CNN 由粗粒度到细粒度逐步预测人脸和关键点坐标。其后 DAN[15](deep alignment network)使用深度学习方案进行人脸关键点提取，DAN 包含多个阶段，每个阶段都对上一阶段估计的人脸关键点位置进行修正。相比其他方法，DAN 采用人脸关键点热度图，每一阶段都使用完整的人脸图像而非局部图像块进行特征点估计。上述深度学习方案可以获得具有竞争力的人脸对齐精度，但深度模型受限于庞大的参数量和繁重的深层次结构，即使通过模型压缩和小型化处理也同样为后期集成在硬件造成了极大的阻碍[16]。

随着便携式设备的智能化发展，人脸对齐应用领域大部分需要移植在基础硬件之上，甚至是便携式设备和小型化设备，兼备精度和速度需求的人脸对齐模型日益成为主要需求。人脸检测器对人脸对齐的意义弥足重要，没有一个良好的检测架构提供精确的人脸坐标框的回归，那么众多人脸对齐算法即使再精妙也只是空中楼阁。因此本文基于速度和精度的需求，使用深度学习架构来提供精准的人脸坐标框回归，在传统机器学习算法的基础上，提出多角度初始化算法方案来获得快速的人脸关键点定位。本文做出以下两项工作：1) 以深度学习 one-stage 网络 SSD 作为基础，融合分布均匀的共八个特征层进行级联回归预测，选择精准的符合人脸比例的预测框生成尺度形成鲁棒性模型 MR-SSD，对多尺度人脸信息做出较好响应的同时节省了时间。2) 基于 LBF 二值特征的级联回归方案，提出一种基于像素差值的多角度初始化算法，对每一张图片采取五组均匀分隔角度的初始化形状送入模型回归，其后对眼部关键区域计算像素均方差，获得抖动最大的回归形状作为最终特征点的回归形状。本文架构相比于传统的机器学习人脸对齐方案可获得更加精准的面部特征点的回归以及较快的实时速度。

2 人脸检测和人脸对齐架构

2.1 MR-SSD 鲁棒模型

SSD 架构用于 20 个类别的目标检测任务，它从不同层次的网络融合多尺度的特征进行级联预测。相比仅仅使用某个层的预测更为精准和完善。同时它又是一个可进行端对端训练且易整合在其他检测系统之中的统一架构，在速度和精度之间有一个较好的平衡。但实际的人脸对齐应用场景要求模型具有较高人脸坐标框的回归精度和实时的检测速度。在面对这样的挑战下 SSD 性能不足[17]。本文希望通过融合更多层次更均匀尺度的信息来获得更加鲁棒的响应。

如图 1(a)所示，SSD 中选择的六个层融合的底层特征较少，其中最接近底层的特征层已是第四个卷积块的 conv4_3 层，而图片的细节信息此时已经在多次卷积过程中逐步转化为语义特征，其选择的六个特征层不能较高程度地涵盖各个尺度的人脸信息[18]。因此本文以特征提取的 VGG16 的基础网络中增加更底层的 con3_3 层，以及第五个卷积块的 conv5_3 层。附加层中选择来自 fc7 的 conv8_2 和 conv9_2、conv10_2、conv11_2 这些均匀层次并保留足够丰富的各个尺度的特征进行融合。其余遵循 SSD 的处理，裁剪掉 VGG 架构末端的全连接层，并将最后的池化层改为卷积层。同时选择 1: 1，1: 1.3，1: 1.5 共三种不同比例生成预测框，可较好地拟合人脸形状。相比 SSD 的六种尺度，本文节省了一半用于预测框生成的时间成本，同时以更加精准地生成方式为后续坐标框的回归提供了良好的基础。本文的整体架构如图 1(b)所示。

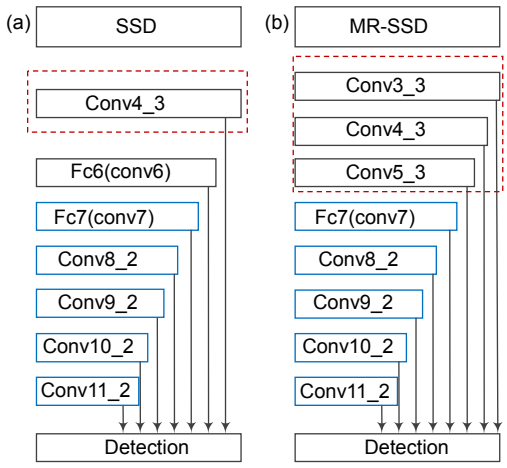


图 1 SSD 和 MR-SSD 整体架构

Fig. 1 Framework of SSD and MR-SSD

2.2 人脸对齐架构

2.2.1 局部二值特征(LBF)浅析

LBF 仅仅提取局部二值特征，这意味着采取该种特征提取方式提升了总体模型的实时性能。该算法由随机森林和全局线性回归相结合而成，对于每一张图采用训练集关键点均值形状作为初始化值，并将其定义为初始形状；在初始形状关键点不同半径范围内，随机生成预测特征点。其后训练随机森林，以随机森林输出的稀疏的“ 0，1 ”二值特征向量作为最终的特征点向量。随机森林的应用特点在于对于每一张图的每一个关键点建立多个决策树，每一棵决策树最后都会输出一个值。由图 2 可知对第一棵决策树执行遍历，到了输出层最左边的子节点，所以该树的输出记为[1，0，0，0]。依此类推，访问到的叶子节点记为 1，其他的节点记为 0，那么每一个关键点都会拥有一个由决策树组成的森林。对于图 2 中所有决策树的输出为[1，0，0，0，0，1，0，0，0，0，1，0，…]。

LBF 算法将所有的关键点特征级联起来，再用稀疏特征向量训练一个回归器。线性回归是一个不断迭代的过程，即采用上一个阶段的状态作为输入来更新，从而产生下一个阶段的状态，以此类推直到最后阶段：

$$\Delta S^t = \boldsymbol{W}^t \varphi^t (I, S^{t-1}) \ , \tag{1}$$

其中：I 为输入图像，S^{t-1} 为第 $t-1$ 阶段的人脸形状，φ^t 为该阶段的特征匹配函数，$\boldsymbol{W}^t$ 为线性回归矩阵。每一个阶段的人脸关键点形状均采用上一个阶段的人脸形状作为输入，经过回归之后得到下一个阶段的人脸形状，逐级回归直到获取最终的人脸形状。线性回归以 LBF 提取的人脸特征关键点作为输入，通过训练学习到线性回归矩阵 $\boldsymbol{W}^t$ 以及特征匹配函数。

2.2.2 基于像素差值的多角度初始化算法

在 LBF 算法中，人脸关键点标志位预先采用训练集人脸关键点均值进行初始化，其后对该关键点形状进行回归。而在实际应用中，人脸倾斜角度往往不是固定不变的，时常会出现侧倾或者左右倾斜情况。同时总体的形状均值往往维持在端正人脸的正负几度之内。因此对于这种多角度倾斜的人脸，统一采用整体均值形状进行初始化，对于正脸来说效果比较精准，而对于具有一定倾斜角度效果欠佳。对后期的回归造成了极大的困难，这就需要应用模型具有抗倾斜度的鲁棒性。

为此，针对不同角度的人脸倾斜，本文提出基于像素差值的多角度初始化算法来实现精确鲁棒的特征点的回归。根据像素差异可将人脸划分为不同的区域。额头、腮部区域内均由皮肤覆盖，因此额头以及腮部区域内像素差异微小；但眼部区域有黑色瞳孔和白色眼白，同时还有睫毛和眼袋，这决定了眼部区域内像素差异较大。因此，本文在五大关键点区域中选择像素差异抖动最大的眼部区域作为评判最优回归形状的关键区域。

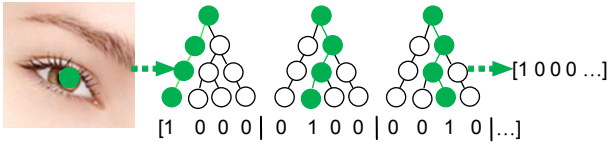


图 2 随机森林

Fig. 2 Random forest

如上所述，以左眼、右眼、鼻尖、左嘴角和右嘴角五个关键点作为标准形状。因为人脸倾斜角度大有不同，使用统一的均值初始化形状会造成极大的不精准拟合，例如用一个居中的人脸初始化形状去拟合一个右倾 45°的人脸会对后期人脸关键点形状的回归过程造成困难。而使用一个右倾 30°的人脸初始化形状去拟合这种人脸将会获得一个更好的回归起点。由图 3 所示，定义相对于基准线向右倾斜方向为正向，向左倾斜方向为负向。采取负 30°(粉色组别)，负 60°(白色组别)，正 30°(浅蓝色组别)，正 60°(深蓝色组别)，以及一组居中人脸(红色)形状共五组人脸关键点形状进行初始化。使用该五种初始化方案可对多种不同倾斜角度的人脸进行覆盖度比较高的关键点拟合，可看出每组初始化方案关键点可覆盖面部不同区域。本文相应地从五组不同初始化形状中选择眼部半径为 N 个像素点区域作为关键区域统计像素，分别得出均值 μ 和均方差 σ。公式如下:

$$\mu=\frac{1}{N}\sum_{i=1}^{N}x_i \quad , \tag{2}$$

$$\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2} \quad , \tag{3}$$

其中：μ 是每一种初始化方案的两只眼部区域指定范围内所有像素点 x_i 的像素均值，σ 是均方差目标函数。与以往不同，在该算法内是要使得目标函数最大化，得到像素抖动最剧烈的区域，从而获得更贴近眼部区域的最优回归形状。它代表该区域内所有像素点到均值的偏离程度。如眼部周围像素均方差明显大于其余几组关键点的像素差值，因此可将眼部区域作为区分脸部不同区域的特征。将五组不同倾斜角度的初始化方案进行训练得到五组经过算法回归过后的预测结果。此时的五种初始化方案关键点的回归较之前更加精准。但因为每一种方案以不同的角度覆盖脸部的区域，需要根据本文基于像素的多角度初始化算法计算出均方差最大者代表与人脸最契合的回归形状方案。

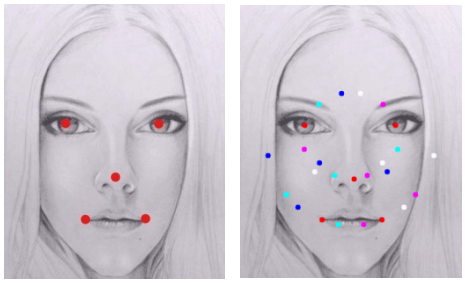


图 3 不同角度初始化算法

Fig. 3 Multi-angle algorithm

在预测阶段，本文提出统一架构如图 4 所示。首先对于每一张输入的预测图片使用 MR-SSD 检测器获得人脸坐标框。其次针对检测出的人脸使用如上所述的包括正负 30°、正负 60°以及基准形状五种人脸特征点形状对人脸关键点进行多角度初始化。其次将检测得到的人脸以及五组初始化形状送入 LBF 训练，对每一组初始化形状进行回归，可获得五组回归后的人脸关键点形状。此时每一组人脸形状相比初始化状态均已有了向真值形状靠近的关键点位置。

为了得到五组预测所得人脸关键点形状中最为精准的一组，采用如上所述在左眼、右眼为中心确立固定半径的区域，计算该区域内各组预测形状的像素差，像素抖动最为明显的一组被认定为更好地贴近了眼部真值区域。因此对每一种回归后形状的两个眼部关键点区域计算均方差目标函数，选出最大值，即代表该种形状的两个预测眼部关键点位于人脸像素抖动差异最大的眼部区域，因此也被认定为最终的预测结果。该方案相当于对倾斜的人脸引入一定的修正。

3 实验结果

本文训练环境搭建在 ubuntu14.04 系统下进行开发，采用 cuda8.0 版本、cudnn5.0 版本进行 GPU 加速，使用 opencv3.1 版本库以及 caffe 深度学习框架，8G NVIDIA GTX 1080 型号显卡，不同组别对比实验显卡数目存在区别。测试环境使用 8G NVIDIA GTX 1060

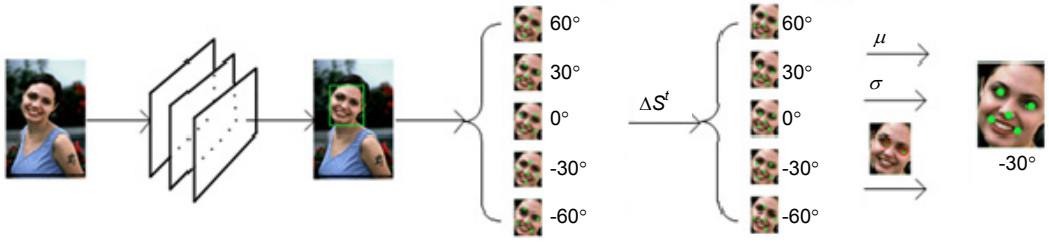


图 4 整体架构

Fig. 4 Architecture

显卡进行实验，其余实验条件与上述一致。

3.1 人脸检测器

将 MR-SSD 与 SSD 分别在 FDDB，Wider Face 两个数据集进行测试，使用 2 张 8G NVIDIA GTX1080 计算。测试集使用 FDDB 全部测试集，以及 Wider Face 的 easy、medium 和 hard 的全体测试集。所得 PR 曲线如图 5 所示。由图 5(a)可知，在 FDDB 数据集上，当召回率为 0.8 时，MR-SSD 可获得 0.89 的预测精度，相比传统 SSD 的 0.81 高出 0.08。当召回率为 0.9 时，MR-SSD 仍然可以维持 0.8 的预测精度，而 SSD 的预测精度此时快速下降至 0.69，远远低于常规任务人脸检测的期望效果。在 Wider Face 数据上由图 5(b)可知，当召回率为 0.8 时，MR-SSD 达到 0.86 的预测精度，SSD 只能达到 0.75。当召回率为 0.9 时，MR-SSD 可获得 0.86 的预测精度，相比 SSD 的 0.75 高出 11%精度。由该组对比实验可知，本文通过增加融合不同尺度的特征层，修改适合人脸比例的预测框生成尺度，使得模型在 FDDB 以及 Wider Face 数据集上获得了较之传统算法 10%提升的平均检测精度。

由表 1 可知，在 NVIDIA GTX 1060 显卡下，对 FDDB 的全体测试集进行实验，可看出 MR-SSD 不但在平均精度上获得 10%左右的提升，同时可获得 41 f/s 的实时速度，也维持在与 SSD 相同的水准。可知尽管在级联预测结构中，本文相比 SSD 多融合了两个特征层用以训练更加鲁棒的模型，选择了更适合人脸尺度的比例，比 SSD 的六种生成方式降低了一半的无用比例生成成本，因此网络达到较快的实时速度，MR-SSD 获得了一个更好的速度-精度的平衡点。

表 1 精度与速度的比较

Table 1 Comparison between accuracy and speed

Network	Mean accuracy		Speed/(f/s)
	FDDB	Wide Face	
SSD	0.812	0.71	42
MR-SSD	0.907	0.824	41

3.2 基于像素差异的多角度初始化算法

3.2.1 人脸对齐性能分析

使用经典的人脸对齐算法 ESR、SDM、PNMS 三种惩罚措施架构，与深度学习方案 MTCNN、DAN 在 AFLW 数据集上进行测试获得平均错误率如图 6。平均错误率(mean error)计算方式是由左眼、右眼、鼻子、左嘴角和右嘴角五个关键点真值与各个架构所得预测值求误差，并将 AFLW 所有测试图片求平均数而得。

由图 6 可知，本文架构在 AFLW 人脸测试集上对左眼、右眼、鼻子、左嘴角和右嘴角五个关键点可得 5%的平均误差率，相比于 ESR 的 13%误差率、SDM8.5%的误差率，以及 PNMS 系列算法获得约 6.9%的误差率大幅度降低。同时相比于深度学习方案中 MTCNN 架构 6.9%的平均误差率降低 1.9%，仅比 DAN 网络高 0.4%。本文架构在五个关键点定位精度方面优于多种传统机器学习算法以及深度学习经典架构 MTCNN。

表 2 中数据由文献[13]中获取，本文算法在显卡性能远低于文献[13]中显卡的条件下，仍可获得优于

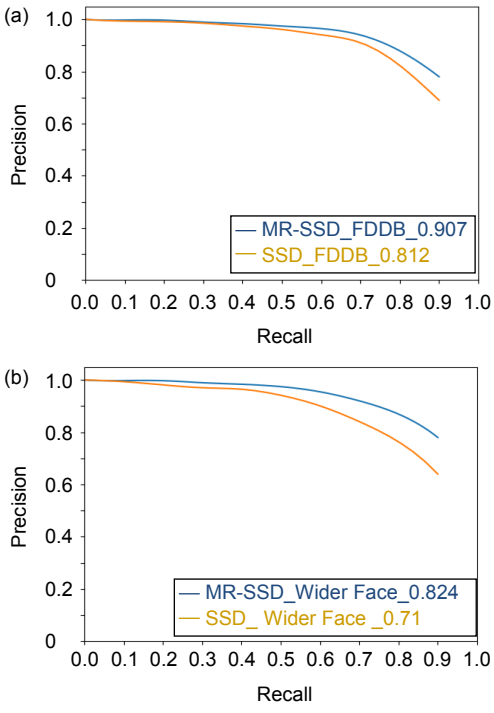
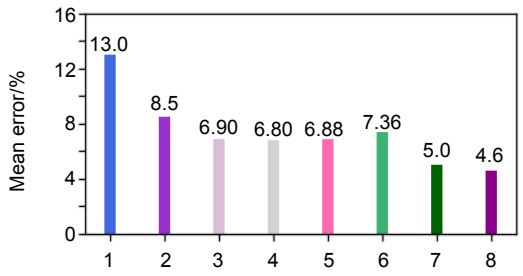


图 5 MR-SSD 与 SSD 比较。(a) FDDB; (b) Wider Face

Fig. 5 Comparison between MR-SSD and SSD. (a) FDDB; (b) Wider Face



1. ESR; 2. SDM; 3. MTCNN; 4. PNMS-linear; 5. PNMS-nonlinear; 6. PNMS-Gaussian; 7. Ours; 8. DAN

图 6 各算法在 AFLW 上的表现

Fig. 6 Performance of algorithms on AFLW

MTCNN、PNMS-linear 的平均检测速度，并与实时性最好的 PNMS-Gaussian 方法仅存微小差别。由表 3 可知本文方案在对齐速度优于传统算法 SDM 和深度学习架构 DAN，虽远低于 LBF 实时速度，但精度优于 LBF。本文在速度与精度平衡方面有一个较好的提高。

表 2 平均人脸检测速度的比较

Table 2 Comparison of detection speed

Architecture	GPU	Real-time speed/(f/s)
MTCNN	GeForce GTX TITAN X	50
PNMS-linear	GeForce GTX TITAN X	40
PNMS-nonlinear	GeForce GTX TITAN X	71
PNMS-Gaussian	GeForce GTX TITAN X	83
Ours	NVIDIA 1060Ti	66

表 3 平均对齐速度的比较

Table 3 Comparison of alignment speed

Architecture	Real-time speed/(f/s)
SDM	40
LBF	3000
DAN	73
Ours	76

3.2.2 图片分析

使用本文的多角度初始化算法与 LBF 级联回归模型在 AFLW 测试集与 Helen 测试集抽样 500 张测试图片进行对比，LBF 采用 OpencV3.1 自带检测器。由表 4 可知，LBF 架构在左眼、右眼、鼻尖、左嘴角以及右嘴角五个关键点定位获得 7.9%的平均误差率，本文使用 MR-SSD 检测器并采用多角度初始化算法可获得 5.4%的平均误差率。本文人脸对齐方案相比 LBF 可在五个关键点定位表现取得优势。抽样测试图片如 7。由图 7(a)可看出，LBF 架构在人脸具有不同角度倾斜时均出现特征点偏移现象，定位与真值相去甚远。由图 7(b)可知本文方案在面对多种倾斜侧脸时均有精确的特征点回归。

表 4 测试图片关键点检测平均误差

Table 4 Mean error of key point detection on test image

Architecture	Mean error/%
LBF	7.9
Multi-angle key point detection algorithm	5.4

3.2.3 实时 USB 摄像头下分析

由图 8(a)可以看出，在 USB 摄像头下进行实时检测以及特征点回归时，无论左倾或右倾都能得到非常精确的人脸框的回归和特征点定位。图中展示了稳定的正负 45°范围内的实时检测效果。

在此基础之上，进一步验证 MR-SSD 结合基于像素差值的多角度初始化算法的整体架构在远距离实时场景下的应用。由图 8(b)可以看出，在远距离摄像头下，人脸信息已极其微小，并且在背景复杂的情况下，对于向左和向右倾斜的人脸，MR-SSD 检测器仍可以获得精准的坐标框的回归。在高精度人脸框的基础之上，多角度初始化算法对于正负大角度倾斜的人脸同样取得了精确的关键点定位。

将本文所提 MR-SSD 结合多角度初始化算法架构与 LBF 对比如表 5。本文算法在 HELEN 数据集上得到 3.61 的瞳孔间距损失，在 LFPW 数据集上可取得 2.17 的损失。在两个主流数据集上均获得比 LBF 较大的提高。同时从倾斜角可容错程度对比可得 LBF 只能针对端正人脸正负 10°左右范围内进行较为精准的回归。而本文最优架构可获得实时场景下正负 50°稳定的人脸框的坐标回归以及非常精准的特征点的回归。

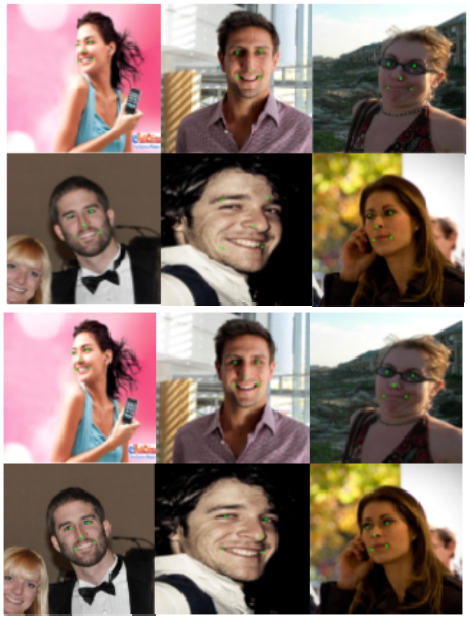


图 7 关键点定位对比。

(a) LBF 回归结果；(b) 本文架构回归结果

Fig. 7 Comparison of key-point location.

(a) Result of LBF; (b) Result of the paper

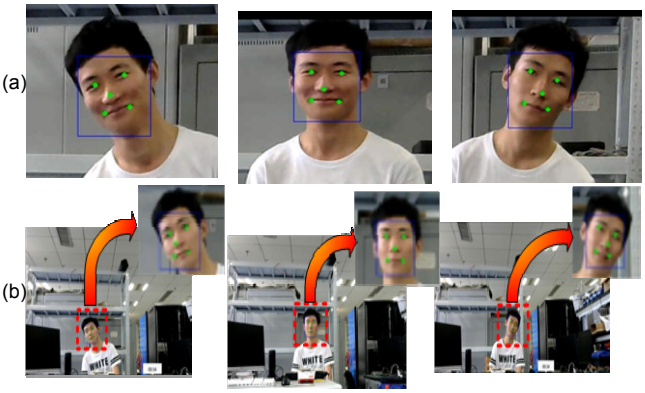


图 8 不同距离关键点检测效果对比。

(a) 近距离角度; (b) 远距离倾斜角度

Fig. 8 Comparison about key point upon variable distance.

(a) Close-distance angle; (b) Remote-distance angle

表 5 变化倾斜角下不同算法的效果对比

Table 5 Different algorithm comparison upon variable angle

Architecture	Error/%		Inclination angle/(°)
	HELEN	LFPW	
LBF	5.41	3.35	±10
MR-SSD+multi-angle key point detection algorithm	3.61	2.17	±50

4 结 论

本文采用改进后的 one-stage 人脸检测 MR-SSD 获得精确的人脸框坐标回归，并且在轻量级机器学习方案 LBF 基础之上，提出基于像素差异的多角度初始化算法，在预测阶段获得了对于多角度倾斜人脸关键点的精确回归。本文从小型化快速化设计理念出发， 在充分考虑实时应用场景的速度与精度要求下设计整体网络架构。最终为将该模型应用于便携式设备以及可移动设备提供可能并取得鲁棒的表现。

本文认为在基于像素差值的初始化算法基础上，进一步细分角度范围将会带来更加优异的精度提升，但同时也不可避免地增加计算负担。因此，如何在不增加时间成本和模型量级的前提下，实现更加精准的特征点的回归方案，将有必要进行深入的探索。

参考文献

[1] Wang Y M, Pan G, Wu Z H. A survey of 3D face recognition[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2008, **20**(7): 819–829.

[2] Peng M C, Bao J, Ye M, *et al*. Face alignment algorithm based on shape parameter regression[J]. *Pattern Recognition and Artificial Intelligence*, 2016, **29**(1): 63–71.

[3] Zhu C R, Wang R S. Adaptive facial feature selection algorithm[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2002, **14**(1): 26–30.

[4] Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580–587.

[5] Girshick R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision* (*ICCV*), 2015: 1440–1448.

[6] Ren S Q, He K M, Girshick R, *et al*. Faster R-CNN: towards real-time object detection with region proposal networks[C]//*Advances in Neural Information Processing Systems*, 2015: 91–99.

[7] Uijlings J R R, van de Sande K E A, Gevers T, *et al*. Selective search for object recognition[J]. *International Journal of Computer Vision*, 2013, **104**(2): 154–171.

[8] Liu W, Anguelov D, Erhan D, *et al*. SSD: single shot multibox detector[C]//*Proceedings of the 14th European Conference on Computer Vision,* 2016: 21–37.

[9] Redmon J, Divvala S, Girshick R, *et al*. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2016: 779–788.

[10] Cao X D, Wei Y C, Wen F, *et al*. Face alignment by explicit shape regression[J]. *International Journal of Computer Vision*, 2014, **107**(2): 177–190.

[11] Xiong X H, De la Torre F. Supervised descent method and its applications to face alignment[C]//*Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 532–539.

[12] Ren S Q, Cao X D, Wei Y C, *et al*. Face alignment at 3000 FPS via regressing local binary features[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 1685–1692.

[13] Li Z D, Zhong Y, Chen M, *et al*. PNMS algorithm based on penalty factors for face detection and alignment[J]. *Advanced Engineering Sciences*, 2018, **50**(6): 225–231.
李振东, 钟勇, 陈蔓, 等. 基于惩罚因子的 PNMS 算法的人脸检测和对齐[J]. 工程科学与技术, 2018, **50**(6): 225–231.

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

[15] Jiao F, Shan S G, Cui G Q, *et al*. Face recognition based on local feature analysis[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2003, **15**(1): 53–58.

[16] Song H, Shi F. Multi-view face detection and pose discrimination in video[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2007, **19**(1): 90–95.

[17] Zhang S F, Zhu X Y, Lei Z, *et al*. S^3FD: single shot scale-invariant face detector[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*, 2017: 192–201.

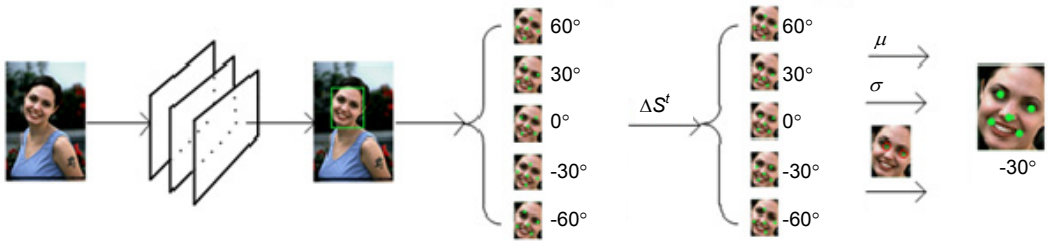
[18] Wan J, Li J, Chang J, *et al*. Face alignment on local-shape-based combined model[J]. *Chinese Journal of Computers*, 2018, **41**(9): 2162–2174.

[19] Bodini M. A review of facial landmark extraction in 2D images and videos using deep learning[J]. *Big Data and Cognitive Computing*, 2019, **3**(1): 14.

Multi-angle key point detection of face based on deep learning detector

Zhao Xingwen, Hang Lijun*, Gong Enlai, Ye Feng, Ding Mingxu

College of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China



Architecture

Overview: The introduction and maturity of deep learning technology greatly promote the development of object detection and key point detection technology. Face alignment, as an extension of the task of face detection, as well as the basis of face calibration and face recognition, is of great significance. For example, in expression recognition, face alignment provides possibilities for the research of emotion recognition. In addition, many applications with the function of beautifying pictures, including face polishing, dynamic face changing effects and so on, need face alignment technology to get facial feature points or feature areas for related operations. There are many methods for realizing face alignment algorithm. Cao et al. put forward ESR (explicit shape regression) scheme to regress the display shape. SDM algorithm uses supervised descent method to achieve the objective function of non-linear least squares, so that it converges to the minimum at a very fast speed. The LBF scheme uses the method of extracting local binary features for regression, which greatly improves the speed of location of key points. In the PNMS scheme, discontinuous linear functions and continuous functions based on Gauss distribution are introduced to improve the non-maximum suppression algorithm, and the candidate windows are re-scored to improve the accuracy and speed. In the scheme of deep learning architecture, Zhang et al. proposed the representative MTCNN (multi-task convolutional neural network) architecture using the deep cascade network, which improves the performance of tasks by utilizing the intrinsic relationship between face detection and face alignment. The unified three-stage cascade CNN is used to advance from coarse-grained to fine-grained step by step. Later, DAN (deep alignment network) used in-depth learning scheme to extract key points of human face. DAN contains many stages, each stage is to modify the position of key points of human face estimated in the previous stage. Based on the requirement of speed and accuracy, the paper uses deep learning architecture to provide accurate regression of face bounding box, and then a multi-angle initialization algorithm is proposed to achieve fast face key point location. This paper makes the following two tasks: 1) On the basis of one-stage network SSD, cascaded regression prediction is carried out by fusing eight feature layers with uniform distribution, and a robust model MR-SSD is formed by choosing the scale of accurate prediction which accords with the proportion of faces, and can make better response to multi-scale face information and save time. 2) A cascade regression scheme based on LBF binary feature is proposed, and a multi-angle initialization algorithm based on pixel difference is proposed. Five groups of uniformly separated initial shapes are used for each image to be fed into the model regression. Then the mean square deviation of the pixels is calculated for the key areas of the eye, and the regression shape with the largest jitter is obtained as the final regression shape of points. Compared with the traditional face alignment scheme based on machine learning, the architecture can obtain more accurate facial feature points regression and faster real-time speed.

Citation: Zhao X W, Hang L J, Gong E L, *et al*. Multi-angle key point detection of face based on deep learning detector[J]. *Opto-Electronic Engineering*, 2020, **47**(1): 190299

Supported by National Natural Science Foundation of China (51777049) and Youth Science Foundation (51707051)
* E-mail: ljhang@hdu.edu.cn