

高精度轻量级的人脸关键点检测算法

徐礼淮¹, 李哲², 蒋佳佳^{1*}, 段发阶¹, 傅骁¹

¹天津大学精密测试技术及仪器国家重点实验室, 天津 300072;

²中国科学院深海科学与工程研究所, 海南 三亚 572000

摘要 针对当前人脸关键点检测算法网络模型复杂度高、在计算资源受限时不利于部署的问题, 基于知识蒸馏思想, 提出了一种高精度、轻量级的人脸关键点检测算法。通过改进残差网络(ResNet50)中的 Bottleneck 模块并引入分组反卷积, 得到轻量级的学生网络。同时提出逐像素损失函数和逐像素对损失函数, 通过对齐教师网络与学生网络的输出特征图与中间特征图, 将教师网络的先验知识迁移至学生网络, 从而提高学生网络的检测精度。实验结果表明, 本算法得到的学生网络参数量为 2.81 M, 模型大小为 10.20 MB, 在 GTX1080 显卡上的每秒传输帧数为 162 frame, 在 300 W 和 WFLW 数据集上的平均误差分别为 3.60% 和 5.50%。

关键词 图像处理; 卷积神经网络; 人脸关键点检测算法; 知识蒸馏; 模型优化; 模轻量级网络

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.241026

High-Precision and Lightweight Facial Landmark Detection Algorithm

Xu Lihuai¹, Li Zhe², Jiang Jiajia^{1*}, Duan Fajie¹, Fu Xiao¹

¹State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China;

²Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan 572000, China

Abstract In view of the high complexity of the current facial landmark detection algorithm network model, which is not conducive to deployment on devices with limited computing resources, this paper proposes a high-precision and lightweight facial landmark detection algorithm based on the idea of knowledge distillation. This algorithm improves the Bottleneck module of residual network(ResNet50) and introduces packet deconvolution to obtain a lightweight student network. At the same time, a pixel-wise loss function and a pair-wise loss function are proposed. By aligning the output feature maps and intermediate feature maps of the teacher network and the student network, the prior knowledge of the teacher network is transferred to the student network, thereby improving the detection accuracy of the student network. Experiments show that the student network obtained by this algorithm has only 2.81 M parameter amount and 10.20 MB model size, the frames per second on the GTX1080 graphics card is 162 frames and the normalized mean error on 300W and WFLW datasets are 3.60% and 5.50%, respectively.

Key words image processing; convolution neural networks; facial landmark detection algorithm; knowledge distillation; model optimization; lightweight network

OCIS codes 100.4996; 100.2960; 100.5010

1 引言

人脸关键点检测是人脸识别^[1]、人脸属性估计^[2]、人脸姿态估计^[3]等后续应用的关键步骤, 目的

是在人脸图像中定位出一组预定义的关键点坐标, 如眼睛、眉毛、鼻尖、嘴和下巴。深度学习方法的发展大幅度提升了人脸关键点检测的精度^[4-7], 但目前高精度的检测算法大多采用大型骨干网络, 如沙漏

收稿日期: 2020-01-08; 修回日期: 2020-02-28; 录用日期: 2020-03-17

基金项目: 国家自然科学基金(61971307, 61905175, 51775377)、国家重点研发计划(2017YFF0204800)、天津市自然科学基金(17JJCQNJC01100)、装备预研领域基金(61405180505, 61400040303)、水下信息与控制重点实验室开放研究项目(6142218081811)

* E-mail: jiajiajiang@tju.edu.cn

网络^[8]和残差网络(ResNet50)^[9],无法满足移动设备的需求。因此,本文采用卷积-反卷积编解码结构,设计了一种轻量级网络,用改进的 ResNet50 的 Bottleneck 模块^[9]作为编码器的基础模块,同时引入分组反卷积^[10]和卷积的上采样结构作为解码器。但该轻量级网络的参数过少,检测精度有限。

知识蒸馏(KD)^[11-13]通过高精度和高复杂度的教师网络指导学生网络进行训练,可以提升计算量小、参数较少的网络精度。因此,引入 KD 思想,用 Simple Baseline^[14]作为教师网络,将上述轻量级网络作为学生网络。为保证学生网络与教师网络的预测结果尽可能一致,通过逐像素损失函数对齐网

络的输出特征图;为从教师网络提取结构化信息,通过逐像素对损失函数对齐网络的中间特征图,使人脸关键点之间的相对空间位置更精确。与未引入 KD 时相比,引入 KD 思想的学生网络在保持参数量、计算量和网络结构不变的前提下,检测精度得到了一定的提升,为在计算资源不充足的设备上部署该网络成为可能。

2 算法原理

本算法的整体框架如图 1 所示,主要包括网络结构、损失函数及训练策略,其中, MSE 表示均方误差。

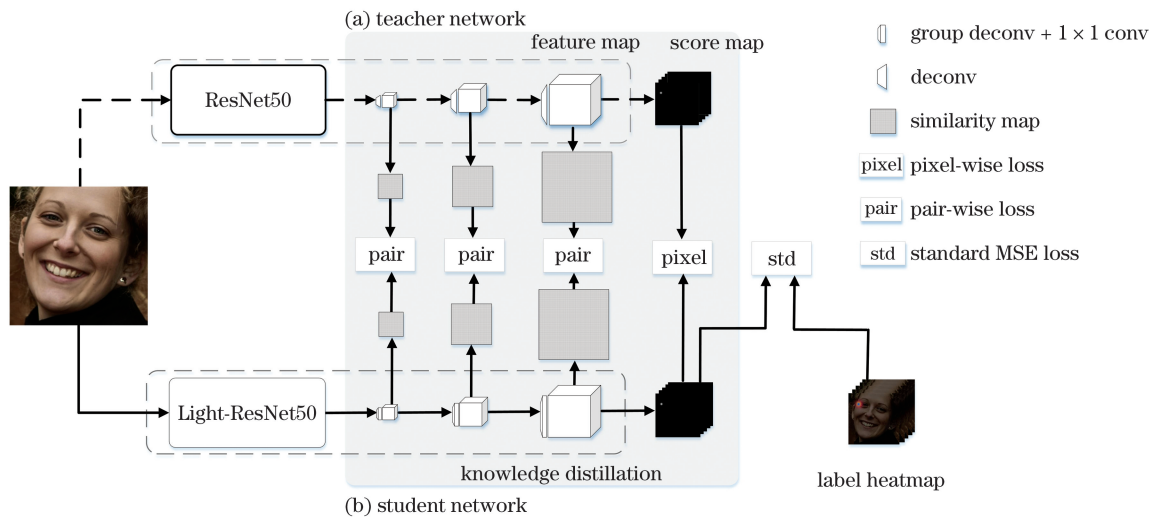


图 1 本算法的整体框架图。(a)教师网络;(b)学生网络

Fig. 1 Overall framework of our algorithm. (a) Teacher network; (b) student network

2.1 网络结构

如图 1(a)所示,教师网络 Simple Baseline 由编码器与解码器构成。其中,编码器为移除自适应全局池化层的 ResNet50,可用于提取人脸图像的特征。解码器采用三层反卷积层,可将编码器得到的低分辨率特征恢复至高分辨率特征,每层反卷积层有 256 个卷积核,每个卷积核的大小为 2×2 ,步长为 2,每次反卷积操作均会使特征图放大 4 倍。为了使预测的热图通道数与关键点个数一致,在教师网络的最后一层反卷积层后添加尺寸为 1×1 的卷积层。

编码器的基础模块如图 2 所示,其中,ReLU 为线性整流函数, M 和 N 均为卷积核的个数。为了增强本框架的重构性,学生网络采用与教师网络一样的编解码结构。同时将教师网络编码器 Bottleneck 模块中的标准 3×3 卷积替换为 3×3 深度卷积,如图 2(a)所示。标准卷积的参数量为

$D_K \times D_K \times C_{in} \times C_{out}$,深度卷积的参数量为 $D_K \times D_K \times C_{in}$,其中, D_K 为卷积核的大小, C_{in} 和 C_{out} 分别为卷积操作的输入、输出通道数。将 Bottleneck 模块输出通道的倍乘因子 expansion 由 4 调整为 1,

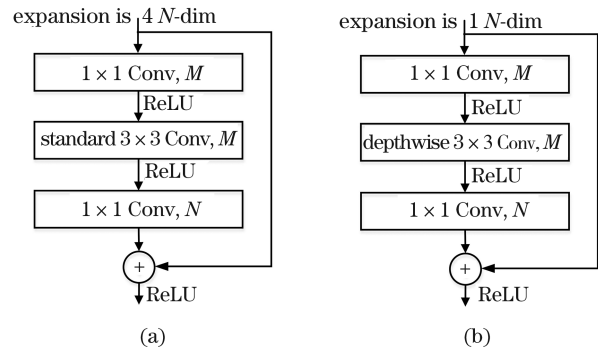


图 2 编码器的基础模块。(a) Bottleneck;(b)轻量级 Bottleneck

Fig. 2 Basic module of the encoder. (a) Bottleneck; (b) lightweight Bottleneck

并将改进后的轻量级 Bottleneck 模块作为学生网络编码器的基础模块,如图 2(b)所示。为了减少解码器的参数量,将教师网络解码器的反卷积层替换为分组反卷积层,使参数量减少为原来的 $\frac{1}{G}$,其中, G 为分组反卷积的组数。最后在每层分组反卷积层后添加尺寸为 1×1 的卷积,以增加通道间的信息融合。

2.2 损失函数

在深度学习中,主要利用损失函数衡量网络预测结果的好坏,即预测结果与实际数据的差异程度。损失函数越小,网络的鲁棒性就越好。因此,设计合理的损失函数至关重要。与文献[5]类似,实验用 MSE 损失函数作为标准损失函数,用来对比学生网络预测的热图 \mathbf{H}^s 与真实关键点生成的热图 \mathbf{H} 之间的偏差,可将每个样本的 MSE 表示为

$$L_{\text{std}}(\mathbf{H}^s, \mathbf{H}) = \frac{1}{L} \sum_{l=1}^L \|\mathbf{H}_l^s - \mathbf{H}_l\|_2^2, \quad (1)$$

式中, L 为关键点的个数, \mathbf{H}_l^s 为学生网络预测的第 l 个关键点热图, $\|\cdot\|_2$ 为 2 范数。用 \mathbf{x}_l 表示第 l 个真实人脸关键点的坐标, \mathbf{H}_l 表示以 \mathbf{x}_l 为中心, σ 为方差生成的二维高斯概率分布的热图,如图 3 所示,其中,各个像素点 \mathbf{p} 的概率可表示为

$$\mathbf{H}_l(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p} - \mathbf{x}_l\|^2}{2\sigma^2}\right). \quad (2)$$

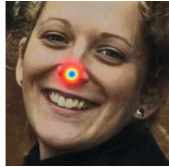


图 3 真实图像关键点的热图

Fig. 3 Heatmap of key points of real image

为了将教师网络学到的知识迁移到学生网络中,将关键点检测问题看作一系列像素分类问题的集合,提出了一种逐像素损失(pixel-wise loss)函数。将教师网络的输出 \mathbf{H}^t 作为软目标,与学生网络中产生的像素概率对齐,逐像素损失函数可表示为

$$L_{\text{pixel}}(\mathbf{H}^s, \mathbf{H}^t) = \frac{1}{L} \sum_{l=1}^L \|\mathbf{H}_l^s - \mathbf{H}_l^t\|_2^2, \quad (3)$$

式中, \mathbf{H}_l^t 为教师网络预测的第 l 个热度图。通过最小化 \mathbf{H}^s 与 \mathbf{H}^t 之间的偏差,可使学生网络的输出结果更接近教师网络。

为将人脸结构化信息从教师网络迁移到学生网络,使人脸关键点之间的相对空间位置更精确,

在文献[13]的基础上,提出了一种逐像素对损失函数。对于维度为 $H \times W$ 的空间,第 i 个与第 j 个空间位置的结构相似度 a_{ij} 可表示为

$$a_{ij} = \frac{\mathbf{f}_i \cdot \mathbf{f}_j}{\|\mathbf{f}_i\| \|\mathbf{f}_j\|}, \quad (4)$$

式中, \mathbf{f}_i 为特征图上第 i 个位置的特征矢量。分别计算学生网络与教师网络中间特征图上各个位置的结构相似度,得到对应的结构相似度矩阵;再通过逐像素对损失函数最小化两结构相似度矩阵之间的偏差,实现人脸结构信息的传递。设 a_{ij}^s 和 a_{ij}^t 分别为学生网络和教师网络在中间特征图上第 i 和第 j 个位置的结构相似度,则逐像素对损失函数可表示为

$$L_{\text{pair}} = \frac{\sum_{i \in K} \sum_{j \in K} (a_{ij}^s - a_{ij}^t)^2}{(H \times W)^2}, \quad (5)$$

式中, $K = \{1, 2, \dots, H \times W\}$ 为特征图中所有的空间位置。

为将不同尺度的结构信息从教师网络传输到学生网络,将逐像素对损失函数扩展到三个反卷积层,从而使学生网络学习到更丰富的面部细节。学生网络训练优化的总损失函数可表示为

$$L_{\text{total}} = L_{\text{std}}(\mathbf{H}^s, \mathbf{H}) + \lambda_1 L_{\text{pixel}}(\mathbf{H}^s, \mathbf{H}^t) + \lambda_2 \sum_{r=1}^3 L_{\text{pair}}^r, \quad (6)$$

式中, L_{pair}^r 为第 r 层反卷积层的逐像素对损失, λ_1, λ_2 为可调参数,用于平衡标准损失、逐像素损失和逐像素对损失,实验将两参数均设置为 0.1。

2.3 训练策略

为了更好地训练学生网络,首先,使用标准损失函数对教师网络进行优化,得到高精度的教师网络。然后,固定教师网络的参数,利用标准损失函数、逐像素损失函数和逐像素对损失函数进行监督,在教师网络的指导下,对学生网络进行训练,使总损失最小。

3 实验与结果分析

3.1 实验数据集

为验证本算法的有效性,在主流的 300W^[15] 和 WFLW^[16] 数据集上进行实验。300W 数据集由 HELEN、LFPW、AFW 和 IBUG 数据集组成,广泛应用于人脸关键点检测中。由于 HELEN、LFPW 和 AFW 数据集的图像在野外环境中采集,可能存在姿态变化、表情变化和部分遮挡的情况。在 300W 数据集中,每张人脸图像有 68 个标注的人脸

关键点。实验按照文献[4]的规定,采用 3148 张图像进行训练。其中,2000 张图像来自 HELEN 数据集的训练子集,811 张图像来自 LFPW 数据集的训练子集,337 张图像来自 AFW 数据集。完整的测试集 Full 有 689 张图像,包括 Common 和 Challenging 两个子集。其中 Common 子集由 HELEN 测试子集和 LFPW 测试子集组成,Challenging 子集则为 IBUG 数据集。

WFLW 数据集是在非受限条件下采集的数据集,图像中存在较大的姿态变化、夸张的表情和严重的遮挡。包括 7500 张训练图像和 2500 张测试图像,并标注有 98 个人脸关键点。为验证不同情况下算法的鲁棒性,将 WFLW 数据集的测试集细分为 large pose、expression、illumination、make-up、occlusion 和 blur 数据集。

3.2 评价指标

归一化平均误差(NME)是人脸关键点检测任

务中常用的评价指标,可表示为

$$X_{\text{NME}} = \frac{1}{L} \sum_{i=1}^L \frac{\|p_i - g_i\|_2}{d}, \quad (7)$$

式中, p_i 和 g_i 分别为第 i 个预测关键点和真实关键点的坐标, d 为两个外眼角关键点的欧氏距离。 X_{NME} 越小,表明算法的检测效果越好。

3.3 实验设置

实验环境:软件框架为 Pytorch,处理器为 6 核 Intel Core i5-9600,显卡为 GTX 1080(进行加速计算)。

为了扩充样本数据的规模,防止过拟合,对训练样本进行数据增强。首先根据标注的人脸边界框对人脸进行裁剪并缩放成分辨率为 256 pixel × 256 pixel 的图像,再将给定的训练图像在 $-30^\circ \sim 30^\circ$ 范围内随机旋转,在缩放因子为 0.8~1.2 范围内进行随机缩放,最后以 50% 的概率对图像进行随机水平镜像,如图 4 所示。



图 4 数据增强的效果图。(a)水平镜像;(b)缩放;(c)旋转

Fig. 4 Effect image of data augmentation. (a) Horizontal mirror; (b) scale; (c) rotation

训练过程中,采用 Adam 优化器更新参数,设置批量大小为 8,初始学习率为 0.001,一共训练 80 个 epoch,当训练到第 50 个 epoch 和第 70 个 epoch 时,学习率均衰减为初始学习率的 10 倍。

3.4 实验结果分析

表 1、表 2 分别为本算法与现有人脸关键点检测算法在 300W 和 WFLW 数据集上的评估结果。其中,RCPR 为鲁棒级联姿态回归算法^[16],DSRN 为直接形状回归网络^[17],DVLN 为利用数据变化实现的算法^[18],PCD-CNN 为姿态条件树突算法^[5],CPM 为监督配准算法^[7],SAN 为风格迁移算法^[19],LAB 为引入边界信息的算法^[20],CFSS 为由粗至细的搜索算法^[21],MANet 为基于多注意力网

表 1 不同算法在 300W 数据集的 NME

Table 1 NME of different algorithms in 300W dataset unit: %

Method	Common	Challenging	Full
RCPR	6.18	17.26	8.35
DSRN	4.12	9.68	5.21
DVLN	3.79	7.15	4.45
PCD-CNN	3.67	7.62	4.44
CPM	3.39	8.14	4.36
SAN	3.34	6.60	3.98
LAB	2.98	5.19	3.49
Teacher(our)	3.03	5.26	3.51
Student(our)	3.35	5.78	3.79
Student+KD (our)	3.21	5.41	3.60

络的算法^[22]。可以看出,本算法在两个公开数据集上的检测效果均优于表中绝大部分算法,原因是相比基于卷积神经网络直接回归关键点坐标的方法和传统建模方法,本算法采用基于热图间接回归关键点坐标的方法为每个人脸关键点生成所有位置的概率热图,空间泛化能力更好。对学生网络进行

KD后,本算法在 300W 的 Common、Challenging 及 Full 三个测试集上的精度分别提升了 4.18%、6.40%和 5.01%,在 WFLW 测试集上的 X_{NME} 为 5.50%,与未采用 KD 的方法相比,精度提升了 8.94%。

表 2 不同算法在 WFLW 数据集的 NME
Table 2 NME of different algorithms in WFLW dataset

Method	Test	Pose	Expression	Illumination	Make-up	Occlusion	Blur
CFSS	9.07	21.36	10.09	8.30	8.74	11.76	9.96
DVLN	6.08	11.54	6.78	5.73	5.98	7.33	6.88
MANet	5.41	10.03	5.56	5.54	6.03	7.00	6.25
LAB	5.27	10.24	5.51	5.23	5.15	6.79	6.32
Teacher(our)	5.20	9.01	5.36	5.27	5.45	6.37	6.22
Student(our)	6.04	11.52	6.76	5.70	5.94	7.30	6.87
Student+KD (our)	5.50	9.83	5.89	5.40	5.48	6.88	6.81

表 3 为不同算法的网络结构、参数量和速度,其中,FPS 为每秒传输帧数,M 表示百万个参数。可以发现,虽然本算法与加入边界信息的 LAB 算法和教师网络相比,精度略低,但本算法在学生网络中引入深度卷积和分组反卷积,并调整其卷积的输出通道数倍乘因子,相比 LAB 算法和教师网络,参数

量分别减少了 88.84%和 88.80%,处理速度更快。综合分析算法的精度与网络的模型大小,基于 KD 的人脸关键点检测算法在保持高精度的同时,参数量更少,运行速度更快,更适用于在移动设备上部署。图 5 为本文算法的可视化结果,可以看出,本算法能准确定位到人脸的关键点。

表 3 不同算法的网络结构、参数量和速度

Table 3 Network structure, parameter amount and speed of different algorithms

Method	Backbone	Params /M	FPS/frame
LAB	Houglass	25.1	6
Teacher(our)	ResNet50	25	126
Student+KD (our)	Light-ResNet50	2.8	162

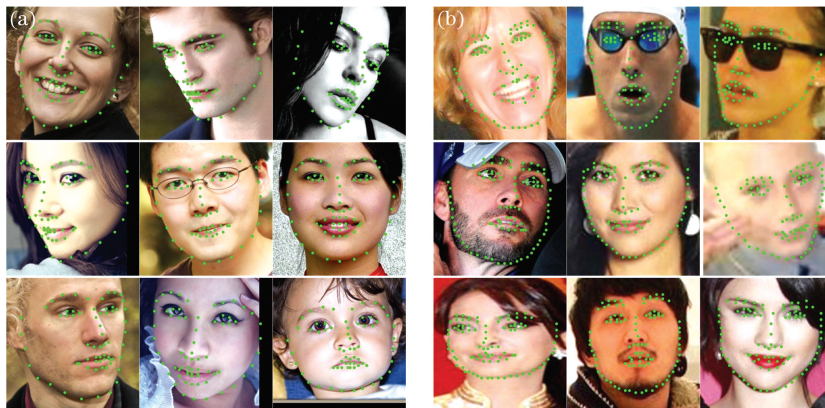


图 5 不同数据集的可视化结果。(a) 300W 数据集;(b) WFLW 数据集

Fig. 5 Visualization results of different datasets. (a) 300W dataset; (b) WFLW dataset

为验证逐像素损失函数和逐像素对损失函数的有效性,利用两种损失函数在 300 W 数据集的 Full 验证集上进行对比实验,结果如表 4 所示。可以发现,相比未使用 KD 的方法,在仅采用逐像素损失函数、仅采用逐像素对损失函数和两者相结合三种情况下进行 KD,算法的精度分别提升了 3.2%、4.5%和 5.0%。这表明逐像素损失函数和逐像素对损失函数都能将教师网络学到的知识迁移到学生网络中,而将两种损失函数相结合时,教师网络能提供更多的监督信息,最大程度地提高本算法的性能。

表 4 不同损失函数在 300W 数据集上的精度
Table 4 Accuracy of different loss functions on 300W dataset

Method	NME
Student	3.79
Student+Pixel	3.67
Student+Pair	3.62
Student+Pixel+Pair	3.60

4 结 论

以改进的 ResNet50 为编码器,以三层分组反卷积为解码器,提出了一种轻量级人脸关键点检测算法。为提高算法的检测精度,引入 KD 思想并设计了两种损失函数,将教师网络编码的精细结构信息迁移到轻量级的学生网络中,在不增加计算量的前提下,进一步提高了本算法的性能。实验结果表明,本算法在 300 W 数据集和 WFLW 数据集的平均误差分别为 3.60%和 5.50%,且参数量仅有 2.81 M,在实现高精度的同时,极大减少了计算资源的消耗,具有很大的实用价值。

参 考 文 献

- [1] Masi I, Rawls S, Medioni G, et al. Pose-aware face recognition in the wild [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 4838-4846.
- [2] Zhuang N, Yan Y, Chen S, et al. Multi-label learning based deep transfer neural network for facial attribute classification [J]. Pattern Recognition, 2018, 80: 225-240.
- [3] Hong C Q, Yu J. Multi-modal face pose estimation with multi-task manifold deep learning [EB/OL]. [2019-12-28]. <http://arxiv.org/abs/1712.06467>.
- [4] Kowalski M, Naruniec J, Trzcinski T. Deep alignment network: a convolutional neural network for robust face alignment [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2034-2043.
- [5] Kumar A, Chellappa R. Disentangling 3D pose in a dendritic CNN for unconstrained 2D face alignment [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 430-439.
- [6] Wu Y, Ji Q. Facial landmark detection: a literature survey [J]. International Journal of Computer Vision, 2019, 127(2): 115-142.
- [7] Dong X Y, Yu S I, Weng X S, et al. Supervision-by-registration: an unsupervised approach to improve the precision of facial landmark detectors [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 360-368.
- [8] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation [M]//Leibe B, Matas J, Sebe N, et al. Computer Vision-ECCV 2016. Lecture Notes in Computer Science. Cham: Springer, 2016, 9912: 483-499.
- [9] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [10] Zhang T, Qi G J, Xiao B, et al. Interleaved group convolutions [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 4383-4392.
- [11] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network [EB/OL]. [2019-12-25]. <http://arxiv.org/abs/1503.02531>.
- [12] Zagoruyko S, Komodakis N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [EB/OL]. [2019-12-25]. <https://arxiv.org/abs/1612.03928>.
- [13] Liu Y F, Chen K, Liu C, et al. Structured knowledge distillation for semantic segmentation

- [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 2599-2608.
- [14] Xiao B, Wu H P, Wei Y C. Simple baselines for human pose estimation and tracking[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2018, 11210: 472-487.
- [15] 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, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE, 2013: 397-403.
- [16] Burgos-Artizzu X P, Perona P, Dollár P. Robust face landmark estimation under occlusion [C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE, 2013: 1513-1520.
- [17] Miao X, Zhen X T, Liu X L, et al. Direct shape regression networks for end-to-end face alignment [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 5040-5049.
- [18] Wu W Y, Yang S. Leveraging intra and inter-dataset variations for robust face alignment [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2096-2105.
- [19] Dong X Y, Yan Y, Ouyang W L, et al. Style aggregated network for facial landmark detection [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 379-388.
- [20] Wu W Y, Qian C, Yang S, et al. Look at boundary: a boundary-aware face alignment algorithm [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2129-2138.
- [21] Zhu S, Li C, Loy C C, et al. Face alignment by coarse-to-fine shape searching [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 4998-5006.
- [22] Liu X, Wang H B, Cheng R, et al. Multi-attention network for 2D face alignment in the wild [M]//Wang Y, Huang Q, Peng Y, et al. Image and Graphics Technologies and Applications. IGTA 2019. Communications in Computer and Information Science. Singapore: Springer, 2019, 1043: 246-255.