

集成人脸关键点的轻量化YOLOv7人脸检测算法

曹一波, 胡伊斐, 刘 顺

(华南师范大学软件学院, 广东 佛山 538200)

摘要: 人脸检测作为计算机视觉技术的重要组成部分, 已广泛应用于欺诈防范、情感分析、年龄预测等领域。然而, 随着移动设备和智能摄像头的普及, 图像和视频数据的获取变得更加便捷, 而设备的计算能力有限, 亟需更快速、高效的人脸检测算法。为此, 提出一种集成人脸关键点的轻量化YOLOv7人脸检测算法, 旨在实现人脸的高效检测, 以适应算力有限的移动和嵌入式设备。该算法首先使用计算效率更高的GhostNet V2轻量化模块替换YOLOv7的ELAN模块, 使模型参数量下降63.9%; 其次, 将改进后的PFLD人脸关键点检测算法集成到YOLOv7的Head部分, 以适应人脸样本姿态多样、图像质量不一、数据集样本分布不均等问题, 形成一个以人脸检测为主、人脸关键点检测为辅的多任务人脸检测网络。在WIDER FACE数据集上的实验结果表明, 所提算法在Easy、Medium和Hard 3个子集上分别达到95.9%、93.5%和83.5%的平均精度, 相较原始YOLOv7模型分别提升了3.8%、2.9%和4.1%, 同时模型参数量减少了41%, 是一种快速、高效的人脸检测算法。

关键词: 人脸检测; 轻量化算法; YOLOv7; GhostNet V2; 人脸关键点

DOI: 10.11907/rjdk.241602

开放科学(资源服务)标识码(OSID):

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)008-0059-06



Lightweight YOLOv7 Face Detection Algorithm Integrated with Facial Keypoints

CAO Yibo, HU Yifei, LIU Shun

(School of Software, South China Normal University, Foshan 538200, China)

Abstract: Facial detection, as an important component of computer vision technology, has been widely applied in fields such as fraud prevention, sentiment analysis, and age prediction. However, with the popularity of mobile devices and smart cameras, the acquisition of image and video data has become more convenient, but the computing power of devices is limited, and there is an urgent need for faster and more efficient face detection algorithms. To this end, a lightweight YOLOv7 face detection algorithm integrating facial keypoints is proposed, aiming to achieve efficient face detection and adapt to mobile and embedded devices with limited computing power. The algorithm first uses the GhostNet V2 lightweight module with higher computational efficiency to replace the ELAN module of YOLOv7, reducing the model parameter count by 63.9%; Secondly, the improved PFLD facial keypoint detection algorithm will be integrated into the Head section of YOLOv7 to address issues such as diverse facial sample poses, inconsistent image quality, and uneven sample distribution in the dataset, forming a multi task facial detection network with facial detection as the main focus and facial keypoint detection as a secondary approach. The experimental results on the WIDER FACE dataset show that the proposed algorithm achieves average accuracy of 95.9%, 93.5%, and 83.5% on the Easy, Medium, and Hard subsets, respectively. Compared with the original YOLOv7 model, it improves by 3.8%, 2.9%, and 4.1%, respectively, while reducing the model parameter count by 41%. It is a fast and efficient face detection algorithm.

Key Words: face detection; lightweight algorithm; YOLOv7; GhostNet V2; facial keypoints

收稿日期: 2024-07-13

扫描二维码阅读全文:

基金项目: 宁波市重点技术研发项目(2021E007)

作者简介: 曹一波(1971-), 男, 博士, 华南师范大学软件学院副教授, 研究方向为机器视觉、人工智能导航技术; 胡伊斐(2001-), 男, 华南师范大学软件学院硕士研究生, 研究方向为人脸检测、机器视觉; 刘顺(1999-), 男, 华南师范大学软件学院硕士研究生, 研究方向为3D人脸识别。本文通讯作者: 胡伊斐。



0 引言

人脸检测作为计算机视觉的重要分支,凭借其能够从复杂图像或视频中精准定位并提取人脸信息的能力,成为众多实际场景中的核心技术。其应用范围广泛,涉及金融、医疗、安保、交通、教育、娱乐等多个行业,对于推动各行各业的智能化发展具有重要意义^[1-3]。人脸检测技术经历了传统的基于特征提取的方法,到基于机器学习的方法,再到当前基于深度学习的方法若干个发展阶段,在检测精度、速度与泛化能力等方面均取得实质性进展,但仍存在诸多局限性,亟需在未来的发展中寻求突破与平衡。

1 相关研究

1.1 基于特征工程的方法

在人脸检测算法的发展进程中,Viola等^[4]于2001年提出的Viola-Jones算法是最具里程碑意义的早期方法之一。该算法巧妙结合了Haar-like特征和级联分类器,通过在训练过程中使用积分图像技术实现了特征的快速计算,从而达到实时人脸检测的目的。其核心创新包括应用积分图像以及利用AdaBoost算法进行特征选择与分类器构建。尽管Viola-Jones算法因高效性而广受欢迎,但是其在处理复杂姿态变化和光照场景时表现不佳。随后,基于HOG(Histograms of Oriented Gradients)特征和支持向量机(Support Vector Machine, SVM)分类器的方法成为研究热点。Dalal等^[5]于2005年首次将HOG描述子应用于行人检测,该方法随后被成功迁移至人脸检测领域。HOG特征通过捕捉局部梯度方向信息,能够有效描述图像中的形状特征,因此在处理姿态和光照变化场景时表现出色。尽管HOG+SVM方法在检测性能上优于Viola-Jones算法,但较高的计算复杂度限制了其在实时检测中的应用。以上两种方法代表了早期人脸检测算法的两个重要发展方向:一是追求实时性,二是提高检测准确度。它们为后续深度学习时代的人脸检测算法奠定了重要基础。

1.2 基于深度学习的方法

近年来,目标检测作为计算机视觉领域的核心研究方向之一,持续吸引着学术界和工业界的广泛关注。作为目标检测的重要分支,人脸检测的发展轨迹与通用目标检测技术的进步密切相关。根据生成最终检测结果前是否需要先产生物体候选区域,基于深度学习的目标检测算法可分为单阶段(One-stage)和二阶段(Two-stage)两大类。这些算法最初并非专门为人脸检测而设计,而是面向更广泛的通用目标检测任务。通用目标检测算法的优势在于其考虑了物体的多样化特征和广泛属性,这种全面性使其在检测各类目标时展现出卓越性能。基于这一优势,学者们开始探索将这些算法应用于人脸检测任务。例如,Zhang

等^[6]基于SSD(Single Shot MultiBox Detector)结构设计了一个轻量级网络FaceBoxes,通过快速缩小特征尺寸和使用多尺度网络模块来提高人脸检测的效率和精度,但在复杂场景和姿态变化情况下的检测准确性较低,同时模块化设计的优化空间有限,难以进一步提升性能。Ke等^[7]在RefineDet和RetinaNet的基础上进行改进,提出SRN(Side-output Residual Network)模型。该模型引入两阶段分类和回归,并设计多分支模块来增大感受野,有助于捕捉更多特征,显著提高了检测精度,但增加了网络的复杂性,训练和推理时间较长,同时需要更多计算资源,不适合嵌入式和移动设备。Deng等^[8]提出的RetinaFace模型在多个特征金字塔层级应用独立上下文模块,显著提高了检测精度,且对光照变化和部分遮挡场景具有较强的鲁棒性,但增加了模型复杂度,推理速度较慢,同时特征金字塔的多层处理机制导致内存占用较高。

Redmon等^[9]提出的YOLO(You Only Look Once)模型是目标检测领域的一个重要里程碑。该模型的核心思想为将目标检测问题转化为回归问题,通过单次前向传播便能同时预测多个目标的位置和类别,大大提高了检测速度。随后迭代的YOLO系列模型均在取得优良性能的同时保持了较低的计算复杂度。例如,陈泽等^[10]提出一种基于YOLO剪枝的设备端人脸检测方法,首先通过改进YOLO网络实现人脸检测与面部关键点回归;然后对模型进行剪枝,以降低其复杂度;最后综合利用多任务损失函数与训练策略增强模型的鲁棒性。

随着移动设备和智能摄像头的普及,实时高效的人脸检测算法需求日益增长。然而,这些设备的计算能力相对有限,给高精度人脸检测带来了巨大挑战。现有人脸检测算法通常面临以下问题:①模型参数量大,难以在资源受限的设备上部署;②对姿态变化、光照条件复杂、部分遮挡等场景下的人脸检测效果不佳;③难以在检测精度与计算效率之间取得良好平衡。为此,本文对检测精度和速度均表现良好的单阶段通用目标检测算法YOLOv7进行改进,提出一种轻量化人脸检测算法^[11]。

本文主要创新点为:①使用GhostNet V2轻量化模块替换YOLOv7的ELAN(Effective Long-Range Aggregation Network)模块,显著降低模型参数量和计算复杂度,同时保持较强的特征提取能力^[12];②设计多任务学习框架,将改进后的PFLD(A Practical Facial Landmark Detector)人脸关键点检测算法集成到YOLOv7的Head部分,提高模型对复杂场景的适应能力^[13];③借鉴TCDCN(Tasks-Constrained Deep Convolutional Network)网络模型解决多任务学习中不同任务收敛速度不一致的问题,使模型在速度与精度之间达到平衡^[14]。实验结果表明,本文算法不仅计算复杂度显著降低,而且在WIDER FACE数据集上取得了优于多个最新人脸检测算法的性能^[15]。

2 算法构建

本文算法整体结构如图 1 所示。首先, 在 YOLOv7 模

型中使用 GhostNet V2 模块替换 ELAN 模块以降低模型复杂度; 其次, 在 YOLOv7 的 Head 中加入改进后的 PFLD 人脸关键点检测算法, 形成以人脸检测为主、人脸关键点检测为辅的多任务人脸检测框架。

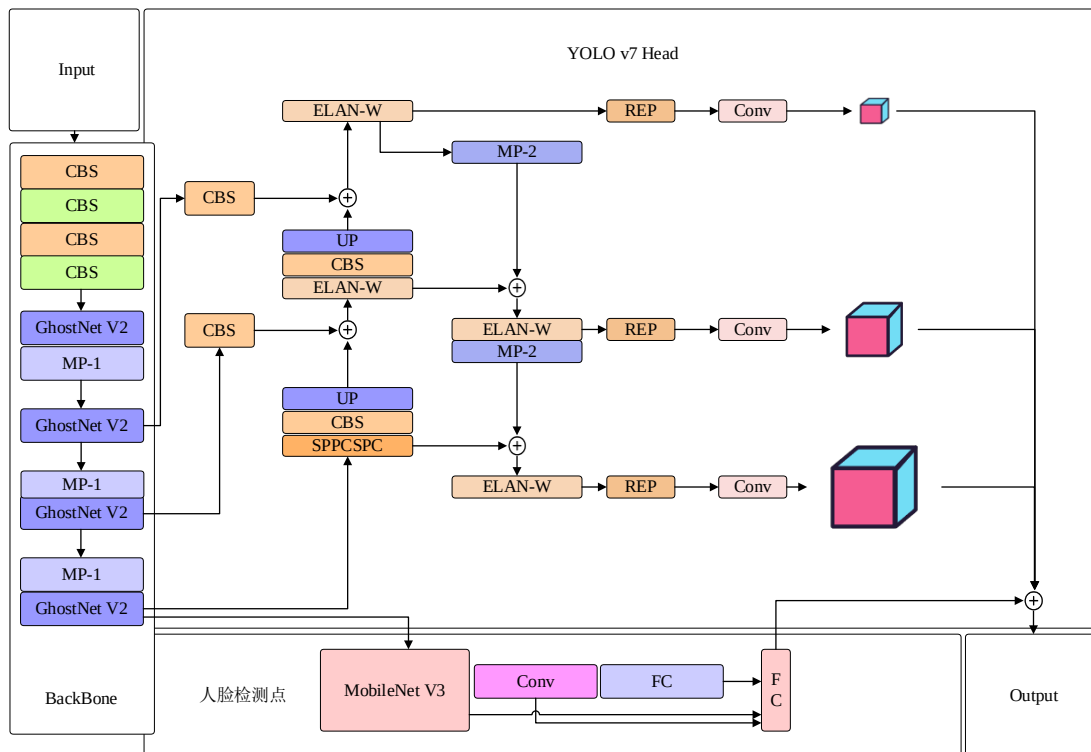


Fig. 1 Overall structure of the algorithm

图 1 算法整体结构

2.1 GhostNet v2 模块

GhostNet V2 是一种高效的轻量级卷积神经网络架构, 其核心思想是通过廉价操作生成更多特征图, 从而在保持模型表达能力的同时大幅度降低计算成本。本文选择 GhostNet V2 替换 YOLOv7 中的 ELAN 模块主要基于以下考虑: ①并行计算效率。GhostNet V2 采用并行的 Ghost 模块结构, 能够更高效地利用现代硬件的并行计算能力, 特别适合在移动设备上运行; ②注意力机制。GhostNet V2 引入了解耦全连接注意力机制 (Decoupled Fully Connected Attention, DFCattention), 能有效捕捉长距离空间信息, 提高特征表达能力, 这对于人脸检测任务尤为重要; ③动态特征融合。GhostNet V2 的反向瓶颈结构和动态特征融合机制使模型能够更好地适应不同尺度和复杂度的人脸样本; ④低秩分解。GhostNet V2 采用低秩分解技术, 在保证模型精度的同时进一步减少冗余参数量, 这对于模型轻量化至关重要。

2.2 人脸关键点检测

PFLD 是一个简单、快速、高精度的人脸关键点检测算法, 主要由预测人脸关键点坐标的主干网络 (Backbone) 和预测几何信息的辅助网络 (Auxiliary) 组成。作为一个轻量级网络模型, PFLD 可以布置在手机终端上。本文使用 MobileNet V3 改进原 PFLD 算法中的 MobileNet V2 模块。Mo-

bileNet V3 更新了倒残差结构, 加入了 SE (Squeeze-and-Excitation) 模块, 更新了激活函数, 使用了 NAS (Neural Architecture Search) 搜索参数, 重新设计了耗时层结构, 进一步降低了延迟和计算量^[16-17]。

Auxiliary 模块预测的几何信息实际上是人脸姿态, 通过欧拉角来表示人脸的三维旋转信息, 包括偏航角、俯仰角和滚动角。具体如图 2 所示。这些信息辅助 Backbone 进行人脸关键点定位任务, 从而提高整个网络的精确性和鲁棒性。欧拉角的计算步骤为: ①预先定义标准人脸。在一组正面人脸图像上取平均值, 并在标准人脸平面上选取 11 个标志点作为参考; ②估计旋转矩阵。利用每个人脸的 11 个对应标志点与参考标志点进行比对, 从而估算出旋转矩阵; ③计算欧拉角。根据得到的旋转矩阵计算出相应的欧拉角。

PFLD 算法巧妙地设计了一个损失函数来防止训练数据规模不足、训练样本类别不均衡等问题。该损失函数表示为:

$$\ell = \frac{1}{A} \sum_{a=1}^A \sum_{b=1}^B \mu_b \|d_b^a\| \quad (1)$$

式中: A 表示每个训练过程中的图像样本数; B 表示每张人脸上的关键点个数; $\|d_b^a\|$ 表示第 a 个训练图像样本上第 b 个关键点的距离或误差度量; μ_b 表示不同的权重, 考虑

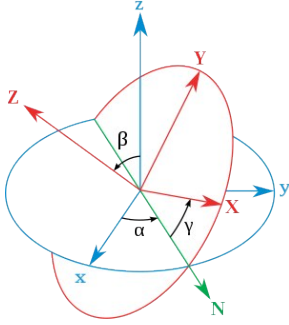


Fig. 2 Expressions of vectors in Euler angles

图2 向量在欧拉角中的表达

到几何约束和数据不平衡的问题,该参数的计算公式为:

$$\mu_b = \sum_{c=1}^C \omega_b^c \sum_{k=1}^K (1 - \cos \theta_b^k) \quad (2)$$

将式(2)代入式(1)后可得:

$$\ell = \frac{1}{A} \sum_{a=1}^A \sum_{b=1}^B \sum_{c=1}^C \omega_b^c \sum_{k=1}^K (1 - \cos \theta_b^k) \|d_b^a\| \quad (3)$$

式中: ω_b^c 为加权参数,表示第**b**个人脸关键点在第**c**类中所对应的权重; C 表示类别数量; θ 表示地面真实值与所估计的欧拉角之间的偏差,在PFLD算法中**k**=3,角度越大,cos值越小,权重就越大,说明大姿态的人脸权重更大。

2.3 多任务学习框架的损失函数

本文算法是一个多任务算法,其中各任务的复杂度可能不同,如果对所有任务使用统一的损失权重可能会阻碍其收敛。为解决权重分配和任务收敛速度不一致的问题,本文借鉴TCDCN网络模型策略,尤其是任务约束深度模型和任务级提前停止方法。具体来说,主要的人脸检测任务采用最小二乘损失,而辅助的人脸关键点检测任务则采用交叉熵损失。该框架的损失函数定义为:

$$\ell = \arg \min_{W^r, \{W^a\}} \frac{1}{2} \sum_{i=1}^N \|y_i^r - f(x_i; W^r)\|^2 - \sum_{i=1}^N \sum_{a \in A} \lambda^a y_i^a \log(p(y_i^a | x_i; W^a)) + \sum_{i=1}^T \|W\|_2^2 \quad (4)$$

式中: $f(x_i; W^r) = (W^r)^\top x_i$ 为线性函数; $p(y_i = m | x_i) = \frac{\exp\{(W_m^a)^\top x_i\}}{\sum_j \exp\{(W_j^a)^\top x_i\}}$ 为softmax函数,表示类后验概率;第三项训练大权重($W = \{W^r, \{W^a\}\}$)。

为解决收敛速度不一致的问题,当辅助任务不再对主任务有帮助时便停止训练,即引入一个标准:设 E_{val}^a 和 E_{tr}^a 分别为任务**a**在验证集和训练集上损失函数的值,如果乘积超过阈值 ϵ 便停止该任务。表示为:

$$\frac{k \cdot \text{med}_{j=t-k}^t E_{tr}^a(j)}{\sum_{j=t-k}^t E_{tr}^a(j) - k \cdot \text{med}_{j=t-k}^t E_{tr}^a(j)} \cdot \frac{E_{val}^a(t) - \min_{j=1..t} E_{tr}^a(j)}{\lambda^a \cdot \min_{j=1..t} E_{tr}^a(j)} > \epsilon \quad (5)$$

式中: t 表示当前迭代轮次; k 表示训练长度;med表示

计算中位数的函数。式(5)中的第一项表示训练误差的趋势。如果在长度为**k**的周期内训练误差迅速下降,第一项的值较小,说明任务仍然有价值,训练可以继续,反之则停止。第二项表示泛化误差与训练误差的差距,其中 λ^a 为第**a**个任务误差的重要性系数,可通过梯度下降学习得到,越重要的任务会产生越长时间的影响。

3 实验方法与结果分析

3.1 实验环境与参数设置

本实验使用Python作为编程语言,PyCharm作为Windows10操作系统中的实验调试平台,使用PyTorch来训练网络模型参数。具体配置如表1所示。

Table 1 Experimental environment

表1 实验环境

配置	参数
编程语言	Python 3.8.5
深度学习框架	Pytorch 1.9
CPU	Intel i7 12700
显卡	RTX 4060Ti 16 GB
内存	32 GB

输入图像大小为640×640像素。优化算法选择随机梯度下降(Stochastic Gradient Descent, SGD),权重衰减参数设为0.000 5,批处理大小设为128。训练过程分为预热和正式训练两个阶段,在前3个epoch中,动量参数设置为0.8,学习率从初始值0.001线性增加至0.01。在正式训练阶段,动量参数调整为0.937,学习率采用余弦退火策略,从0.01逐渐降低至0.000 1。IoU阈值设为0.5,共训练300个epoch。

3.2 数据集及预处理方法

WIDER FACE数据集是于2016年提出的公开人脸检测数据集,包含32 203张图像,标注了393 703个人脸,涵盖了各种真实场景下的人脸检测情况。该数据集的一个显著特点为高度多样化的人脸尺度分布。图像中的人脸大小从几个像素到数千像素不等,为检测算法提出严峻挑战。此外,WIDER FACE数据集还包含大量具有遮挡、表情变化、姿势变化和光照变化的人脸样本,有效模拟了实际应用中可能遇到的复杂情况。

WIDER FACE数据集被分为训练集(40%)、验证集(10%)和测试集(50%)。基于人脸检测难度,数据集进一步被划分为容易(EASY)、中等(MEDIUM)和困难(HARD)3个子集。这种划分方式使研究者们能够全面评估算法在不同难度级别上的性能。

本文使用随机水平翻转(概率为0.5)和颜色失真等传统图像增强方法。颜色失真包括对图像亮度、对比度、色调、饱和度进行调整,以及添加随机噪声,以尽可能模拟真实场景中可能出现的各种光照和成像条件。数据增强策略不仅平衡了不同尺度人脸的分布,而且显著增加了训练

数据的多样性, 可以提高模型对各种复杂场景的适应能力, 从而在实际应用中实现更高的检测精度和鲁棒性。

3.3 评价指标

采用平均精度 (Average Precision, AP)、参数量、浮点运算总数 (Floating Point Operations, FLOPs) 评价模型性能。AP 为综合考虑精确率 (Precision) 和召回率 (Recall) 的评价指标。具体来说, 模型根据对样本的置信度得分可以计算出一系列不同阈值下的 Precision 和 Recall, 绘制出 PR (Precision-Recall) 曲线, AP 即为 PR 曲线下方的面积。在计算 AP 时, 通常对 PR 曲线进行插值平滑, 然后积分得到面积值。对于单类别检测任务, AP 即为该类别的 AP。例如, 在 PASCAL VOC 挑战中, AP 通常指在 IoU 阈值 0.5 下该类别 PR 曲线下的面积值。参数量反映了模型的轻量化程度。FLOPs 用于衡量模型的计算复杂度。

3.4 实验结果与分析

3.4.1 比较实验

为验证本文算法的有效性, 将其与 RetinaFace^[8]、DSFD^[18]、TinaFace^[19]、YOLO5Face-n^[20]、HAMBox^[21]、MTCNN^[22]等具有影响力的算法进行比较, 结果如表 2 所示。可以看出, 在模型复杂度方面, 本文算法的参数量仅为 3.60 M、FLOPs 为 4.20 G, 与 YOLO5Face-n 处于同一轻量化级别, 远低于 DSFD、RetinaFace、TinaFace、HAMBox 等算法。MTCNN 虽然在参数量和 FLOPs 上极具优势, 但其检测精度远低于本文算法, 尤其是在 Medium 和 Hard 难度样本上 AP 差距显著。与 YOLO5Face-n 相比, 本文算法的 AP 在参数量和 FLOPs 略有增加的情况下取得较为明显的提升, 尤其是在 Hard 难度上提升了 3.18%。表明本文算法在轻量化与检测精度之间取得了很好的平衡。

Table 2 Comparison of performance indicators of various algorithms
表 2 各算法性能指标比较

算法	EASY/%	MEDIUM/%	HARD/%	参数量/ M	FLOPs/G
DSFD	94.29	91.47	71.39	120.06	259.55
RetinaFace	94.92	91.90	64.17	29.50	37.59
TinaFace	95.61	94.25	81.43	37.98	172.85
HAMBox	95.27	93.76	76.75	30.24	172.85
YOLO5Face-n	93.74	91.54	80.32	1.73	2.11
MTCNN	85.10	82.00	60.70	0.12	0.01
本文算法	95.90	93.50	83.50	3.60	4.20

图 3 为原始 YOLOv7 模型和本文方法在 2014 年世界最大自拍图片上的检测效果比较。可以看出, 面对人数众多、不同姿势、不同遮挡情况的人脸样本时, 本文方法能检测出更多人脸。

3.4.2 消融实验

为验证本文所提改进措施的有效性, 进行 GhostNet v2 和人脸关键点检测辅任务两个模块的消融实验, 结果见表 3。可以看出, 在使用 GhostNet v2 替换 YOLOv7 中的 ELAN 模块后, 虽然在 EASY、MEDIUM、HARD 3 个子集上的 AP 有 1%~2% 的下降, 但是参数量也明显下降, 说明 GhostNet



(a) Original YOLOv7

(a) 原始 YOLOv7



(b) The proposed method

(b) 本文算法

Fig. 3 Comparison of detection performance between the original YOLOv7 and the algorithm proposed

图 3 原始 YOLOv7 与本文算法的检测效果比较

v2 的并行计算与注意力机制可以在保持模型轻量化的同时很好地获取图像中的人脸特征, 降低模型复杂度, 适用于简单设备。在加入人脸关键点检测辅任务后, 3 种难度子集的 AP 均显著提升, 特别是 Hard 子集的 AP 提升了 4.1%, 说明人脸关键点的权重以及姿态估计策略对于复杂情况下的人脸识别有较为明显的改善效果。

Table 3 Results of ablation experiments

表 3 消融实验结果

消融模块	EASY/%	MEDIUM/%	HARD/%	参数量/M
YOLOv7	92.1	90.6	79.4	6.1
+GhostNet v2	91.2	88.5	77.1	2.2
+人脸关键点	95.9	93.5	83.5	3.6

4 结语

本文提出一种集成人脸关键点的轻量化 YOLOv7 人脸检测算法, 旨在使资源受限的设备 (如移动和嵌入式设备) 实现高精度且实时的人脸检测。该算法使用 GhostNet V2 模块替换 YOLOv7 的 ELAN 模块, 在保证特征提取能力的同时降低了模型的参数量和计算复杂度; 同时改进了 YOLOv7 的 Head 部分, 形成以人脸检测为主、人脸关键点检测

为辅的多任务人脸检测框架。实验结果表明,本文方法在 WIDER FACE 数据集上表现优异,不仅参数数量和计算复杂度大幅度下降,而且检测精度显著提升。

本文算法在轻量化方面尚有改进空间,未来将考虑使用更多模型压缩技术,如知识蒸馏、剪枝和量化等,进一步降低模型的计算复杂度和存储需求,提高其在资源受限设备上的运行效率。

参考文献:

- [1] SHAKER H, AL-DARRAJI S. A face anti-spoofing detection with multi-modal CNN enhanced by ResNet: face anti-spoofing detection[J]. Basrah Researches Sciences, 2024, 50(1): 12.
- [2] ABDELLAOUI B, MOUMEN A, IDRISSE Y E B, et al. Face detection to recognize students' emotion and their engagement: a systematic review [C]//IEEE International Conference on Electronics, Control, Optimization and Computer Science, 2020: 1-6.
- [3] ABIRAMI B, SUBASHINI T S, MAHAVALSHNAVI V. Gender and age prediction from real time facial images using CNN[J]. Materials Today: Proceedings, 2020, 33: 4708-4712.
- [4] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]// Kauai: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005: 886-893.
- [6] ZHANG S, ZHU X, LEI Z, et al. Faceboxes: a CPU real-time face detector with high accuracy[C]// IEEE International Joint Conference on Biometrics, 2017: 1-9.
- [7] KE W, CHEN J, JIAO J, et al. SRN: side-output residual network for object symmetry detection in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1068-1076.
- [8] DENG J, GUO J, VERVERAS E, et al. Retinaface: single-shot multi-level face localisation in the wild[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5203-5212.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [10] CHEN Z, ZHU F B. A device-side face detection method based on YOLO pruning[J]. Software Guide, 2023, 22(3): 196-200.
陈泽, 朱范炳. 一种基于 YOLO 剪枝的设备端人脸检测方法[J]. 软件导刊, 2023, 22(3): 196-200.
- [11] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [12] TANG Y, HAN K, GUO J, et al. GhostNetv2: enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9969-9982.
- [13] GUO X, LI S, YU J, et al. PFLD: a practical facial landmark detector[DB/OL]. <https://arxiv.org/abs/1902.10859>.
- [14] ZHANG Z, LUO P, LOY C C, et al. Facial landmark detection by deep multi-task learning [C]// European Conference on Computer Vision, 2014: 94-108.
- [15] YANG S, LUO P, LOY C C, et al. Wider face: a face detection benchmark [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 5525-5533.
- [16] KOONCE B, KOONCE B. Mobilenetv3[M]. Berkeley: Apress, 2021: 125-144.
- [17] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [18] LI J, WANG Y, WANG C, et al. DSFD: dual shot face detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5060-5069.
- [19] ZHU Y J, CAI H X, ZHANG S H, et al. Tinaface: strong but simple baseline for face detection[DB/OL]. <https://arxiv.org/abs/2011.13183>.
- [20] QI D, TAN W, YAO Q, et al. YOLO5Face: why reinventing a face detector[C]//European Conference on Computer Vision, 2022: 228-244.
- [21] LIU Y, TANG X, WU X, et al. Hambox: delving into online high-quality anchors mining for detecting outer faces [DB/OL]. <https://arxiv.org/abs/1912.09231>.
- [22] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

(责任编辑:尹晨茹)