

基于改进YOLOv4-tiny的人脸关键点快速检测

付博闻, 李闯闯, 梁爱华

引用本文

付博闻, 李闯闯, 梁爱华. 基于改进YOLOv4-tiny的人脸关键点快速检测[J]. 计算机科学, 2022, 49(11A): 211100290-5.

FU Bo-wen, LI Chuang-chuang, LIANG Ai-hua. Facial Landmark Fast Detection Based on Improved YOLOv4-tiny [J]. Computer Science, 2022, 49(11A): 211100290-5.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于差分进化算法的字符对抗验证码生成方法](#)

Adversarial Character CAPTCHA Generation Method Based on Differential Evolution Algorithm
计算机科学, 2022, 49(11A): 211100074-5. <https://doi.org/10.11896/jsjcx.211100074>

[融合多层次视觉信息的人物交互动作识别](#)

Human-Object Interaction Recognition Integrating Multi-level Visual Features
计算机科学, 2022, 49(11A): 220700012-8. <https://doi.org/10.11896/jsjcx.220700012>

[基于注意力机制的手写体数字识别](#)

Handwritten Digit Recognition Based on Attention Mechanism
计算机科学, 2022, 49(11A): 211100009-5. <https://doi.org/10.11896/jsjcx.211100009>

[融合ViT卷积神经网络的木板表面缺陷识别](#)

Wood Surface Defect Recognition Based on ViT Convolutional Neural Network
计算机科学, 2022, 49(11A): 211100090-6. <https://doi.org/10.11896/jsjcx.211100090>

[基于双流网络结构的深度伪造人脸的检测方法](#)

Detection of Deepfakes Based on Dual-stream Network
计算机科学, 2022, 49(11A): 220100106-9. <https://doi.org/10.11896/jsjcx.220100106>

基于改进 YOLOv4-tiny 的人脸关键点快速检测

付博闻¹ 李闯闯¹ 梁爱华²

1 北京联合大学机器人学院 北京 100101

2 北京联合大学前沿智能技术研究院 北京 100101

(1171460872@qq.com)

摘要 人脸关键点检测作为人脸识别的重要环节,一直是计算机视觉领域的研究热点。为了满足高效轻量级的人脸关键点检测需求,提出了一种基于改进 YOLOv4-tiny 的人脸关键点快速检测算法。模型输入采用 $608 \times 608 \times 3$ 的彩色图像,使用 CSPDarknet53-tiny 网络对输入图像进行主干特征提取,对提取到的特征进行上采样和特征融合,在特征融合之前添加注意力机制来提高检测准确度,同时对 YOLOv4-tiny 网络的损失函数进行调整,添加人脸关键点的损失计算,实现在人脸目标检测的同时对关键点进行标定定位。模型输出包括人脸标记框和人脸 5 个关键点。实验结果表明,相比其他网络的人脸关键点检测方法,所提模型在保证识别准确度的基础上,具有更高的识别效率和更低的配置要求,可以满足快速实时检测的需求,且更易部署在边缘设备或者移动设备上。

关键词: 人脸关键点检测;YOLOv4-tiny;注意力机制;实时检测;深度学习

中图分类号 TP391

Facial Landmark Fast Detection Based on Improved YOLOv4-tiny

FU Bo-wen¹, LI Chuang-chuang¹ and LIANG Ai-hua²

1 School of Robotics, Beijing Union University, Beijing 100101, China

2 Frontier Intelligent Technology Research Institute, Beijing Union University, Beijing 100101, China

Abstract Facial landmark detection is an important part of face recognition, which has been a hot issue in the field of computer vision. In order to meet the needs of efficient and lightweight face recognition, this paper proposes a facial landmark detection algorithm based on improved YOLOv4-tiny. $608 \times 608 \times 3$ color image is used for model input. The CSPDarknet53-tiny network is adopted to extract the main features of the input image. Then the extracted features are up-sampled and fused. Attention mechanism is added before feature fusion to improve the detection accuracy. The loss function of YOLOv4-tiny target detection is reconstructed, and the loss function of facial landmark is added to realize the location of facial landmark while detecting. The model output includes face marker frame and five key points. Compared with other facial landmark detection algorithms, the proposed algorithm has higher recognition efficiency and lower configuration requirements while ensuring recognition accuracy. Therefore, it can be better deployed on edge devices or mobile devices.

Keywords Facial landmark detection, YOLOv4-tiny, Attention mechanism, Real-time detection, Deep learning

1 引言

在计算机视觉领域,人脸检测一直被作为深入研究的经典问题。近年来,在技术不断进步的情况下,基于深度学习的人脸检测算法成为研究的热点。在对人流量较大的场景进行人员流动调查时,如地铁站口、火车站口、商场入口等,需要高效实时的人脸检测和识别,尤其在新冠疫情的当下,这种需求尤为明显。

当前的人脸检测算法使用的网络大多较为复杂,无法部署在一些低性能设备上,而且检测速度较慢,在实时检测场景下表现较差。为了满足高效轻量级的人脸关键点检测需求,

本文提出一种基于改进 YOLOv4-tiny 的人脸关键点检测算法,选择 $608 \times 608 \times 3$ 大小的图片作为模型的输入,输出的数据包括人脸标记框、关键点位置、人脸有无遮挡数据。借助 YOLOv4-tiny 简洁的网络结构,通过增加注意力机制,并针对人脸检测场景修改损失函数,实现了人脸快速检测和人脸关键点的快速实时检测和定位。该算法在检测速率上具有较大优势,更适用于需要实时检测的场景或是在边缘设备上部署。

2 相关工作

在计算机视觉发展初期,人脸检测技术作用的图像通常为受约束的特征明显的标准人脸图像,且具有清晰的五官

基金项目:国家自然科学基金(61502036);北京联合大学校级科研项目(ZK50202002);北京市高等教育学会 2021 年立项一般课题(YB202175)
This work was supported by the National Natural Science Foundation of China(61502036), Scientific Research Project of Beijing Union University (ZK50202002) and General Project of Beijing Association of Higher Education(YB202175).

通信作者:梁爱华(liangaihua@buu.edu.cn)

特征以及较小的尺度变化。在当前社会快速发展的情况下,传统约束性的人脸检测技术很难满足需求,需要在非约束的复杂环境下,对模糊、遮挡、倾斜、表情等人脸的准确识别。ASM^[1] (Active Shape Model)和 AAM^[2] (Active Appearance Models)这两种算法作为早期的人脸关键点检测算法,其特点是数据处理复杂、算法抗干扰能力差。而近年来基于深度学习的人脸检测算法越来越多。

RetinaFace^[3]网络提供了单级解决方案,对任意尺度下的图像同时进行人脸对齐和密集回归。作为一个多任务的学习算法,实现了在多尺度下的人脸检测和人脸关键点提取。主干特征网络使用了 ResNet-50^[4]网络结构,使用特征金字塔网络 (Feature Pyramid Networks, FPN^[5])来实现特征增强。通过关键点检测,对人脸实现更加精准的检测对齐,损失函数同时进行包含框体和特征点的预测。在 P2-P5 层上,分别对应每一层设置一组锚框,P2 层级选择使用较低级别的特征图和密集的锚点框来提高模型解决面积较小、质量较差的人脸识别情况的能力,但这种方式一方面会带来更多的计算开销,另一方面产生假阳性结果的风险更高^[6]。

Retinaface 使得 Arcface^[7]人脸识别模型的准确率得到提升,在 widerface 公开人脸数据集中的困难数据集上有着显著改善。但是由于网络模型整体参数量较大,对硬件要求较高,在移动和边缘设备上表现不佳或成本过高。

CascadeCNN^[8]人脸检测算法通过级联几个检测网络达到不断优化预测结果的目的,与普通级联不同的是,CascadeCNN 的几个检测网络是基于不同 IOU 阈值确定的正负样本上训练得到的,这是该算法的一大亮点。使用 6 个 CNN 的级联结构,用于人脸分类和边框矫正。最后将高度重叠窗口使用 NMS 消除。MTCNN^[9]人脸对齐网络由 P-Net, R-Net, O-Net 3 部分构成。P-Net 部分主要负责获取人脸区域的候选框和边界框的回归向量,并通过 NMS 将高度重叠的候选框合并。R-Net 将一些 false-positive 区域通过边界框和 NMS 去掉。O-Net 进一步精细化处理,同时输出人脸 5 个关键点。本文提出的模型使用 YOLO 思想,只通过一个网络输出多个结果,而不是通过多个网络的级联来输出结果。在整体上模型更加简洁,可读性更强。由于模型简单,在处理速度上有明显的优势。

FaceBoxes^[10]是一个高轻量的人脸检测算法,该算法的核心思想是实现 CPU 下的实时人脸检测。通过使用 RDCL 模块快速降低特征图片大小,提高检测速度,使用 anchor 稠密化策略提高召回率,有效地提高了模型在 CPU 上的运行效率和准确度。相比 FaceBoxes 模型,本文提出的模型不仅能快速检测出人脸,而且能够标记出人脸关键点,且人脸检测和人脸关键点检测同时进行,这种检测方式更有利于通过关键点调整人脸,提高人脸检测模型的准确率与效率。

YOLO 系列算法是目标检测中使用单阶段 (One-stage) 检测的方法,凭借其简单的网络模型和高准确率,在目标检测方面有着诸多应用。经过多年的发展,YOLO 经历了多个版本的迭代,YOLO 算法模型做到了目标检测的轻量化和高效,并针对不同的应用场景衍生出许多版本。

传统 YOLO 算法采用单阶段结构的目的是适用于一般

物体的目标检测,并没有对特殊物体或特殊情况做针对性的适配,因此 YOLO 算法并不支持人脸关键点的定位。YOLOv4-tiny^[11]作为 YOLOv4^[12]的一个衍生版本,对 YOLOv4 模型进行了改进压缩,在保证检测准确率的同时大幅提高了检测效率,在同等情况下比 YOLOv4 模型的处理速度快 10 倍左右。于是我们借鉴 YOLOv4-tiny 网络的模型结构进行改进,保留其准确高效的特点,实现在人脸检测的同时进行人脸关键点检测。

3 改进 YOLOv4-tiny 的人脸关键点检测

3.1 主干特征提取网络

主干特征提取网络沿用 YOLOv4-tiny 的 CSPDarknet53-tiny,网络中使用的基本卷积块 (BasicConv)是由二维卷积层 Conv2d+标准化 BatchNorm2d+LeakyReLU 激活函数叠加组成。CSPnet 作为 CSPDarknet53-tiny 的主要结构,区别于普通 ResNet 残差网络网络的直接相连,CSPnet 是将 5 个基本卷积块堆叠而成的残差块拆分成两个部分,其中一部分继续进行原来残差块的堆叠,另一部分则经过少量处理直接连接到最后。CSPnet 网络结构如图 1 所示。CSPnet 网络能够保持残差网络的特性,解决梯度弥散,加深了网络深度。设计 CSPnet 的目的是最大化梯度联合的差异。其使用梯度流截断的手段避免不同的层学习到重复的梯度信息。如果能够有效地减少重复的梯度学习,那么网络的学习能力将大大提升^[3]。CSPnet 作为 CSPDarknet53-tiny 网络的重要结构,被运用在中间层作为残差层。在 CSPDarknet53-tiny 特征提取网络中 P3, P4 和 P5 都使用了 CSPnet 残差网络结构。

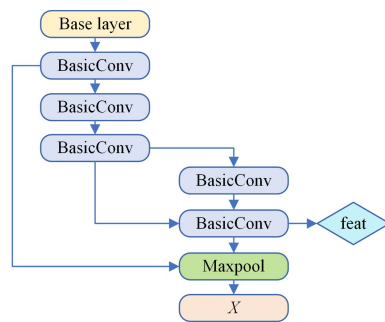


图 1 CSPNet 结构

Fig. 1 Architecture of CSPnet network

在 CSPDarknet53-tiny 网络中,将输入的图片首先经过两次 BasicConv 卷积,卷积后的图像通道数为 $152 * 152 * 64$,卷积后的图像通过 3 次 CSPnet 残差结构,进一步压缩图片,增加图片通道数,加深网络深度。在 P4 层中的 CSPnet 结构,输出图像不经过 Maxpool 处理,直接从 feat 结构输出,向下传递的图像经过 Maxpool 处理后传入下一层。在 P6 层中经过一次基本卷积后,构成完整的 CSPDarknet53-tiny 主干特征提取网络。

在 YOLOv4-tiny 中使用的是 $416 * 416 * 3$ 大小的图片输入,但我们的模型需要处理更高分辨率的图片和更多的细节特征,过度压缩图片会导致细节特征丢失,不利于关键点的定位,于是选择 $608 * 608 * 3$ 大小的图片作为模型的输入。

从 CSPDarknet53-tiny 主干特征网络中提取 P4 层和 P6 层的输出,作为通道注意力网络 ECA-Net^[13] (Efficient Chan-

nel Attention for Deep Convolutional Neural Networks) 的输入,来提高网络的检测精度。ECA-Net 的结构如图 2 所示。对主干特征网络的 P_6 层输出进行上采样后再经过一个 ECA-Net 通道注意力网络,与主干特征网络 P_4 层经过上采样的输出进行特征融合。

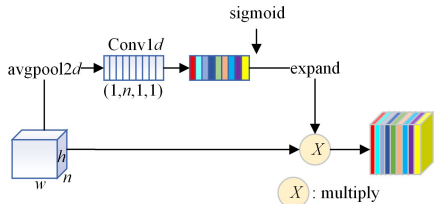


图 2 ECA-Net 结构

Fig. 2 Architecture of ECA-Net network

对处理后的 P_4 和 P_6 层使用 3×3 的 BasicConv 卷积块进行特征整合,由 1×1 的 Conv 卷积块对最后的通道数进行调整,调整为 $(n, 256, 38, 38)$ 和 $(n, 256, 19, 19)$,作为模型最后的输出大小。改进后的整体网络模型如图 3 所示。

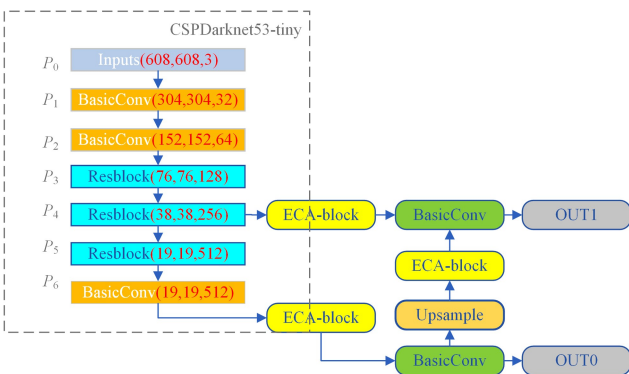


图 3 改进后的 YOLOv4-Tiny 网络结构

Fig. 3 Architecture of improved YOLOv4-Tiny network

3.2 锚框设计

在 RetinaFace 中 4 层特征层对应 4 种锚框大小,每层对应一种锚框大小。而在 YOLOv4 中针对目标不同的形状和大小设计多种锚框,3 层特征层对应 9 种锚框大小。综合分析考虑两种模型情况,使用基于 YOLOv4-tiny 的模型,输出为 2 层特征层,分别对应于原图划分 19×19 和 38×38 ,将划分框的左上角作为锚点,对每个特征层设计了两种锚框尺寸,如表 1 所列。

表 1 各层对应锚框大小
Table 1 Anchor frame size

特征层	步长	锚框
$P_4(n, 38, 38)$	16	(33, 39)
		(99, 102)
$P_6(n, 19, 19)$	32	(175, 226)
		(488, 466)

3.3 预测结果与编码

对特征层进一步提取,分别获得 BboxHead, ClassHead 和 LandmarkHead 3 种预测结果。在 BboxHead 中是人脸预测框的预测结果,结果表示每个锚点所对应锚框的调整参数,调整参数为锚点在 x, y 轴上的移动距离 (t_x, t_y) ,和锚框的调整比例 (t_w, t_h) 。BboxHead 的大小为 $(n, 7220, 4)$,其中 7220 表示 19×19 的锚点和 38×38 的锚点分别对应的两种锚框,

4 代表锚框调整的 4 个参数。锚框调整方式如图 4 所示。

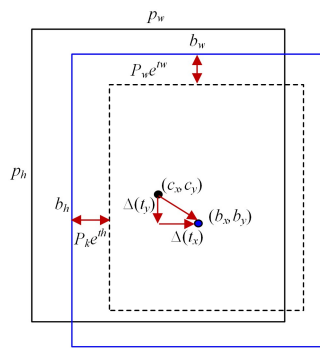


图 4 锚框调整图

Fig. 4 Adjusted anchor frame

ClassHead 是对应锚框是否包含人脸的分类预测结果,其大小为 $(n, 7220, 2)$ 。预测结果表示锚框中包含目标的概率和不包含目标的概率,最后使用 SoftMax 函数作为最终框的置信度。

LandmarkHead 是对应锚框中人脸关键点的调整参数, LandmarkHead 大小为 $(n, 7220, 10)$,其中 10 个参数为 5 个关键点相对于锚点 (C_x, C_y) 的调整参数。人脸关键点的调整策略使用 RetinaFace 网络中人脸关键点的调整方法,调整模式如图 5 所示。

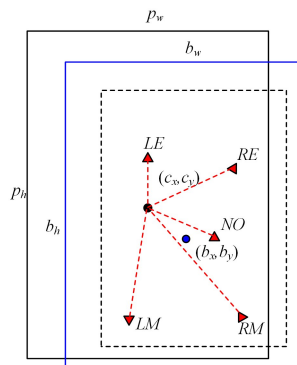


图 5 关键点调整

Fig. 5 Landmark adjustment

基于锚点偏移的人脸各关键点的定位。其中 LE, RE, NO, LM, RM 分别表示人脸 5 个关键点左眼、右眼、鼻子、左嘴角、右嘴角到预测框左上角顶点的偏移量。在对人脸关键点的训练中,通过标签数据中提供的人脸关键点的位置来计算各个关键点与预测框左上角顶点的偏移量进行回归。在检测时,网络在输出 4 个预测框参数的同时,输出人脸各个关键点相对于锚点的 5 组偏移量。通过这些偏移量计算得出人脸关键点。计算如式(1)所示:

$$\begin{cases} X_n = C_x + t_{x_n} \times 0.1 \times P_w \\ Y_n = C_y + t_{y_n} \times 0.1 \times P_h \end{cases} \quad (1)$$

其中, (C_x, C_y) 为原始锚点的坐标。 (t_x, t_y) 为人脸关键点的调整参数,与 5 个人脸关键点相对应。考虑到人脸关键点必须在人脸预测框内,且关键点与人脸大小比例相关。因此引入 P_w 和 P_h 对应锚框的宽和高,作为关键点的调整参数。

3.4 Loss 函数设计

根据网络预测结果,损失函数主要分为 3 部分,分别是预测框损失 r_{loss} 、置信度损失 c_{loss} 、特征点损失 $landm_{\text{loss}}$,损失

函数如式(2)所示。其中的 α, β 和 δ 分别为各损失函数系数,本模型首先需要准确的检测出人脸,在此基础上再进行关键点的检测。于是提高预测框损失(r_{loss})系数的值,提高该损失值在方向传播中的比重,降低其他损失值对其的影响,来提高人人脸框的定位准确。各损失函数比例为 $\alpha:\beta:\delta=2:1:1$ 。

$$L = \alpha \times r_{\text{loss}} + \beta \times c_{\text{loss}} + \delta \times \text{landm}_{\text{loss}} \quad (2)$$

其中, c_{loss} 先通过网络对锚点置信度预测,与真实数据进行交叉熵(cross-entropy)运算,然后对锚点求均值得到最后的锚点置信度损失,如式(3)所示:

$$c_{\text{loss}} = \frac{1}{n} \sum_i (P^* \ln(P) + (1-P^*) \ln(1-P)) \quad (3)$$

其中,有 n 个锚点, P^* 是数据集中的真实数据, P 为预测数据。

借鉴fasterRCNN^[14]中的RPN的loss计算方法,使用Smooth L1损失函数。Smooth L1的计算如式(4)所示。Smooth L1函数就是光滑之后的L1损失函数,基础L1损失函数的缺点就是有折点、不光滑,容易导致梯度下降不稳定。在式(4)中,当 $x_i - y_i$ 的值范围在 $-1 \sim 1$ 之间时,用其平方的0.5倍替换基础L1损失函数来消除折点平滑函数。预测框损失 r_{loss} 的计算如式(5)所示,使用式(4)Smooth L1损失函数计算每个预测框的损失值,再求所有框损失值的均值作为当前 r_{loss} 的损失值。置信度损失 $\text{landm}_{\text{loss}}$ 的计算如式(6)所示,原理与式(5)同理。

$$\text{Smooth}_{L1}(x, y) = \begin{cases} 0.5(x_i - y_i)^2, & \text{if } |x_i - y_i| < 1 \\ |x_i - y_i| - 0.5, & \text{otherwise} \end{cases} \quad (4)$$

$$r_{\text{loss}} = 1/n \sum_i \text{Smooth}(R, R^*)_{L1} \quad (5)$$

$$\text{landm}_{\text{loss}} = 1/n \sum_i \text{Smooth}(L, L^*)_{L1} \quad (6)$$

4 实验与结果分析

在对视频流进行实时检测的过程中,本文提出的模型以平均40fps的流畅速度,远远领先Retinaface-resnet50和Retinaface-mobilenet两种模型,结果如图6所示。



(a) New YOLO v4-tiny (b) Retinaface-resnet50 (c) Retinaface-mobilenet

图6 视频检测

Fig. 6 Video detection

4.1 数据集及实验环境

实验使用widerface公开数据集^[15]。widerface是最广泛应用于人脸检测测试和训练的数据集,其中包含人脸标记框、关键点位置、人脸有无遮挡数据。

使用widerface公开人脸数据集对模型进行训练,对可分辨人脸特征点的人脸进行特征点的标注,对不可分辨人脸特征点的图片只进行人脸标注。在读取数据集时,按照图片最长边进行等比例缩放,为防止图片变形,给图片添加灰条,使图片大小统一为 $608 \times 608 \times 3$ 。将图片的人脸标注点转化到 19×19 和 38×38 的锚点中,解码出锚框是否包含物体和包含物体的锚点到特征点和预测框中心点的距离。

实验环境为 windows10 操作系统, CPU 为 i5-9300H,

GPU 为 NVIDIA GTX 1650, 显存容量为 4069Mb, 使用 python3.7 和 pytorch1.2 进行网络搭建和模型训练。

4.2 实验结果与分析

模型训练采用 Adam 优化算法。该算法是一种对随机梯度下降法(SGD)的扩展,模型训练从 Widerface 数据集中随机抽取 4 张图片为一组做梯度下降,共完成 98426 次抽取训练,为模型完成一次完整训练。学习率(Learning Rate)初始值为 0.001, 权重衰减参数为 0.0005。学习率调整策略为等间隔调整,调整倍数 gamma 为 0.92,调整间隔 step_size 为 1。调整间隔单位是进行一次完整数据集训练。

默认模型训练 100 次,为了加速模型训练过程,在前 50 次训练中冻结模型 backbone 层 50% 的参数,使模型的损失快速下降,对 50 次后的 backbone 层进行解冻,进行完整模型的训练。训练过程中 loss 下降结果如图 7 所示。

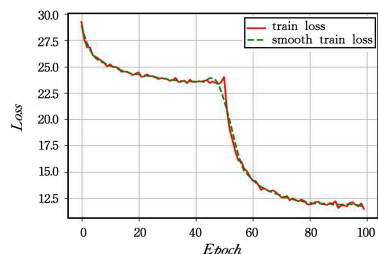


图7 loss下降图

Fig. 7 Loss dropping

多种模型的FPS测试结果如表2所列。所提模型(New YOLO v4-tiny)远远领先于其他几种模型,检测速度是Retinaface-resnet50的近15倍,稍低于同为轻量网络模型中的MTCNN。该模型在检测速率上具有较大优势,更适用于需要实时检测的场景,适合在边缘设备上部署。

表2 模型帧率对比

Table 2 Frame rate comparison

模型	GPU	FPS
New YOLO v4-tiny	GTX1650	70.1FPS
Retinaface-resnet50	GTX1650	4.7FPS
Retinaface-mobilenet	GTX1650	38.2FPS
MTCNN	Nvidia Titans Black	99FPS
Faceness	Nvidia Titans Black	20FPS

在widerface数据集中进行准确度测试,使用IOU阈值为0.4的标准来对预测脸部边框进行测试。在测试结果中对应Easy, Medium和Hard3种难度数据集。在Easy图片中,人物数量少,人脸清晰,遮挡较少,但存在较小人脸;Medium数据集中,大多数为多人照片且人脸有遮挡,存在分辨率较低的模糊照片;Hard数据集中为复杂场景下的照片,如球场、街道和聚会等,人脸数量多且杂乱,模糊和清晰的人脸共存,人脸姿态不一且有严重遮挡,对人脸及关键点检测较难。测试准确率对比如表3所列。对于开放场景下的简单数据集,本文模型准确度略高于MTCNN模型准确度,相比Faceboxes有所提升。我们的模型采用的是CSPDarknet53-tiny网络作为主干特征提取网络,相比MTCNN使用全连接神经网络和Faceboxes经过十几层卷积层,模型参数量远远小于上述模型。在整体的设计上使用的是单阶段模式,在人脸检测的同时进行人脸关键点定位,势必会导致人脸检测的准确率降低,因此在困难数据集下表现一般。而对于实际生活中更需要关注非约束场景下的简单样本的情况,本模型在同时进行人脸检测和定位方面有一定优势。

表 3 不同测试集下的准确率

Table 3 Accuracy of different testsets

Model	Easy	Medium	Hard
New YOLO v4-tiny	0.854	0.762	0.463
MTCNN	0.851	0.820	0.608
Faceboxes	0.840	0.766	0.395

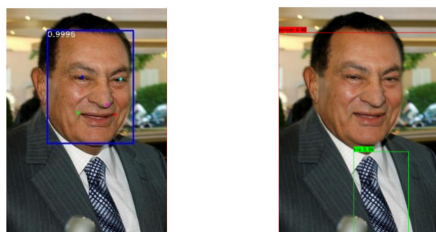
本文提出的模型使用 ECA-Net 结构,与不使用 ECA-Net 结构而用一般卷积模型在 widerface 数据集中进行人脸检测训练和测试。测试结果如表 4 所列。使用 ECA-Net 结构后,模型对整个预测的准确度都有所上升,在困难数据集中相较于一般数据集和简单数据集上升幅度更大。

表 4 注意力结构对准确度的影响

Table 4 Impact of ECA

注意力结构	Easy	Medium	Hard
ECA-Net	0.824	0.717	0.409
—	0.801	0.689	0.315

对比原始 YOLOv4-tiny 的人脸识别情况,本文模型不但能够准确地预测出人脸而且还能准确定位到 5 个人脸特征点,如图 8 所示。非约束场景简单样本下的检测准确情况和 Retinaface 模型检测结果几乎没有差别。而在检测速度上本文模型表现明显优于 Retinaface-mobilenet 模型。本文模型在检测速度和检测精度上做到了一个很好的平衡。



(a) New YOLO v4-tiny

(b) YOLOv4-tiny

图 8 与传统 YOLO 检测对比

Fig. 8 Comparison before and after improvement

结束语 随着神经网络深度学习的快速发展,基于深度学习的脸部关键点检测网络不断出现并不断优化,大多数网络模型都专注于提高网络的准确度,而忽略了模型的复杂度。本文提出了一种基于改进 YOLOv4-tiny 的人脸关键点快速检测模型,经过充分测试验证,得出以下结论:1) CSPDarknet53-tiny 主干特征提取网络可用于人脸关键点检测,在简单场景下对人脸的检测的速度远远优于 mobilenet 特征提取网络;2)使用 ECA-Net 特征提取网络能够有效地优化网络,提高检测准确度,针对 CSPDarknet53-tiny 主干特征提取实现人脸特征点检测的锚框设计,也能有效提高检测精度。

本文方法对图片处理速度进行了极大的优化,能够满足对视频监控进行实时处理的需求。对一般尺度的人脸识别效果很好,对极小尺度模糊人脸识别效果不佳。尤其是复杂场景下的模糊低分辨率人脸,还需在后续研究中继续优化。

参考文献

- [1] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models—their training and application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.
- [2] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis and

Machine Intelligence, 2001, 23(6): 681-685.

- [3] DENG J, GUO J, ZHOU Y, et al. Retinaface: Single-stage dense face localisation in the wild[J]. arXiv:1905.00641, 2019.
- [4] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [5] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.
- [6] SUI Y T, YAN Z Y, DAI L L, et al. Research on face multi-attribute detection algorithm based on RetinaFace[J]. Railway Computer Applications, 2021, 30(3): 1-4.
- [7] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4690-4699.
- [8] LI H, LIN Z, SHEN X, et al. A convolutional neural network cascade for face detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5325-5334.
- [9] 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.
- [10] ZHANG S, ZHU X, LEI Z, et al. Faceboxes: A CPU real-time face detector with high accuracy[C]// 2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2017: 1-9.
- [11] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-yolov4: Scaling cross stage partial network[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13029-13038.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [13] WANG Q L, WU B G, ZHU P F, et al. ECT-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[J]. arXiv:1910.03151, 2019.
- [14] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [15] YANG S, LUO P, LOY C C, et al. WIDER FACE: A Face Detection Benchmark[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016.



FU Bo-wen, born in 2000, undergraduate. His main research interests include computer vision and so on.



LIANG Ai-hua, born in 1979, Ph.D., associate professor. Her main research interests include biometric recognition and image processing.