

基于YOLOV5人脸关键点检测方法的研究与改进

何翔

(南京理工大学计算机学院, 江苏南京 210000)

摘要:人脸检测作为人脸信息处理的首要且不可或缺的环节,被广泛应用到人脸对齐、活体检测、人脸识别、人机交互以及人脸超分辨率重建等研究领域,其结果直接影响后续任务的可行性,对人脸信息处理具有重要的研究价值。人脸检测的研究已经取得了突破性进展,但由于复杂环境包含光照、遮挡、面部角度以及图像分辨率等不确定因素,其仍有性能优化空间。提出一种基于YOLOV5的改进算法,为人脸检测提供面部关键点分支,该分支对面部细节的精准定位起到了重要作用。实验结果表明该算法在测试集上取得了良好的精度,并适用于实时性人脸检测。

关键词:人脸检测;YOLOV5;深度学习;面部关键点定位;RetinaFace

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2022)0031-05

Research and Improvement on Facial Landmark Localization Based on YOLOV5

HE Xiang

(School of Computer Science, Nanjing University of Technology, Nanjing 210000, China)

Abstract: As the primary and indispensable part of face information processing, face detection is widely used in research fields such as face alignment, vivo detection, face recognition, human-computer interaction, and face super-resolution reconstruction. The results of face detection can directly affect the feasibility of successor tasks, which has important research value for the processing of face information. Nowadays, the research of face detection in single environment has made breakthrough progress. However, face detection can be further improved by investigating complex environment factors, such as changes in illumination, occlusion, facial angles, and image resolution, etc. The improved algorithm based on YOLOV5 provides the branch of landmark localization for face detection, which plays an important role in accurate location of facial details. The results of the experiment show that the proposed algorithm achieves high accuracy in the test set and is suitable for real-time application.

Key words: face detection; YOLOV5; deep learning; facial landmark localization; RetinaFace

0 引言

人脸检测问题的出现源于人脸识别^[1]中对人脸定位的需求。人脸检测作为人脸信息处理的基础环节,需要对原始图像或视频是否出现人脸做出快速判断。若出现人脸,需准确地给出人脸检测框,且准确性和实时性将影响整个系统的最终性能。因此,将人脸检测作为一项独立的课题具有重要的研究意义。

随着深度学习^[2-3]的蓬勃发展,经典检测方法由于受限于人工设计特征的表现能力和分类器的预测性能,已被基于卷积神经网络的方法所超越。基于深度学习的人脸检测已取得重大科研成果,由Yang S等提出的Faceness-Net模型先将人脸划分开独立的提取局部特征,再根据打分对候选框进行排序,最后筛选候选框,提高了对有遮挡人脸的检测效果;Zhang K等

人提出的MTCNN^[4]模型,由三个级联的结构组成,能够同时完成人脸检测和人脸关键点的检测任务,级联结构的使用使检测准确度得到了进一步的提高;Deng等人提出的RetinaFace^[5]模型是单阶段(single-stage)人脸检测器,利用联合监督和自监督的多任务学习策略,在各种人脸尺度上执行像素方面的人脸定位;Libia等人提出的DBFace是轻量级的人脸检测器,其网络主干采用具有SE、Hard-Swish Activation等模块的MobilenetV3网络,使该模型能够在边缘计算上有效地使用。

对越来越高要求的人脸检测需求,研究者们纷纷致力于研究如何解决复杂背景、小尺度人脸、高密度人群等棘手的问题。现阶段提出的算法都有各自的优缺点和适用范围,仍有继续改进的空间。YOLO^[6]系列

收稿日期:2021-05-07

作者简介:何翔(1996-),女,硕士,通讯作者,研究方向为模式识别与智能系统。

是有较强实用性的单阶段目标检测框架,其中 YOLOV5 高效的性能和较好的可改造性使其能应用于人脸检测任务。将轻量级 YOLOV5s 作为目标改进人脸检测模型,增加面部关键点^[7](landmark)分支,实现对面部细节的精准定位,旨在保证检测精度的基础上,大幅减少参数量和计算量,达到实时性应用的要求。

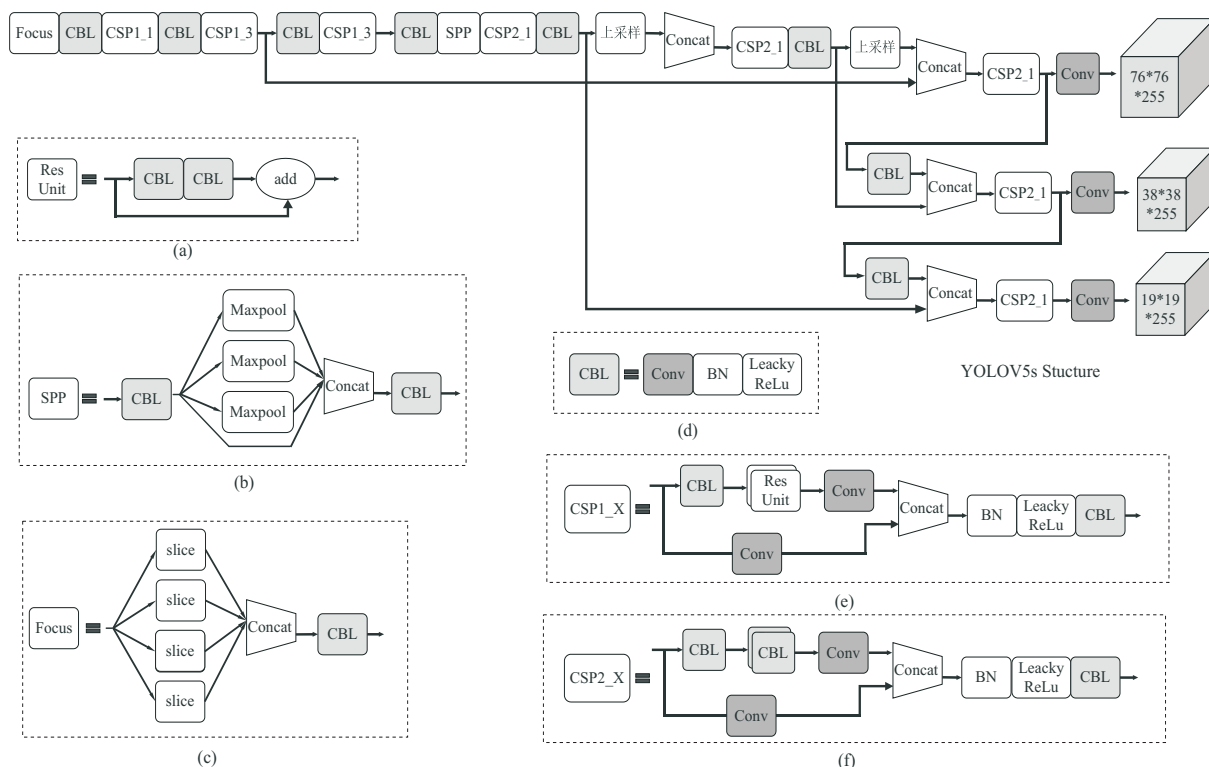


图 1 YOLOV5s 网络结构

YOLOV5 使用 CSP-Darknet 作为 Backbone,从输入图像中提取丰富的信息特征。

其中的 CSPNet^[8](跨阶段局部网络)基于 Densnet^[9]的思想,复制基础层的特征映射图,通过密集块(dense-block)发送副本到下一个阶段,从而将基础层的特征映射图分离出来,有效地解决了其他大型卷积神经网络框架 Backbone 中网络优化的梯度信息重复的问题。YOLOV5 中设计了两种 CSP 结构,如图 1(e)、(f)所示,其中 CSP1_X 结构应用于 Backbone 中,而 CSP2_X 结构则应用于 Neck 中。

Neck 包括两个部分, PANet^[10](路径聚合网络)和 SPP^[11](空间金字塔池化)。

PANet 基于 Mask R-CNN^[12]和 FPN 框架,其作用是加强信息之间的流通。PANet 结构如图 2 所示,左侧卷积结果从低到高输出为 C3, C4, C5, C6, C7, 右侧从低到高记为 P3, P4, P5, P6, P7; 以 P3 为例, P4 将 P3 的特征映射作为输入,并用 3×3 卷积层处理它们。因此最底层的特征流动到最上层只需要经过很少的层,改善了低层特征的传播。PANet 还使用了自适应特征

1 基于 YOLOV5 的人脸检测算法

1.1 网络结构

YOLO 系列网络主要由三个组件组成,分别为 Backbone、Neck 和 Head。YOLOV5 模型由小到大包括了 YOLOV5s、YOLOV5m、YOLOV5l、YOLOV5x。以 YOLO5s 为例,其网络结构如图 1 所示。

池化(Adaptive feature pooling)恢复每个候选区域和所有特征层次之间被破坏的信息路径,聚合每个特征层次上的每个候选区域,避免被任意分配。

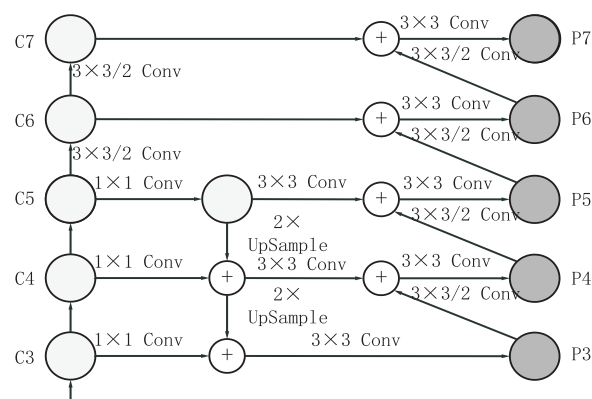


图 2 PANet 结构

SPP(spatial pyramid pooling)操作可以增大感受野,有助于解决 anchor 和 feature map 的对齐问题。如图 1(c)所示,特征图经过三个 pooling 窗口进行池化,然后将得到的结果分别在 channel 维度进行拼接。

Head 部分主要用于最终检测,在特征图上应用锚

定框,并生成包含类概率、对象得分和检测框的最终输出向量。YOLOV5的Head共包含三个输出头,对应stride分别是8,16,32的特征图,大特征图检测小目标,小特征图检测大尺度物体。另外,与经典的YOLOV3^[13]不同的是,在anchor上YOLOv5以跨网格匹配规则的方式来区分anchor的正负样本。

1.2 损失函数

YOLOV5的损失函数设计和前YOLO系列最主要的区别是正样本锚定框区域的计算。YOLOV5的分类分支和置信度分支(confidence)采用的loss是二分类交叉熵损失函数(BCE loss),如式(1)所示。

$$\begin{aligned} \text{loss}(z, y) &= \text{mean}\{l_0, \dots, l_{N-1}\} \\ l_n &= -(y_n * \log(\delta(z_n)) + (1 - y_n) * \\ &\quad \log(1 - \delta(z_n))) \end{aligned} \quad (1)$$

其中, N 表示样本数量, z_n 表示预测第 n 个样本为正例的得分, y_n 表示第 n 个样本的标签, δ 表示 sigmoid 函数。

在检测框(bounding box)损失函数方面, IoU 作为损失函数无法分辨不同方式的对齐, 并且若预测框和 ground truth 没有重叠, IoU(intersection over union, 介于 0 到 1 之间)始终等于 0 并且损失函数变为不可导, 导致 loss 无法优化。为避免出现以上情况, 检测框分支采用 GIoU^[14] 损失函数, 如式(2)所示。

$$\text{GIoU} = \text{IoU} - \frac{|B^c \setminus B^p \cup B^g|}{|B^c|} \quad (2)$$

该损失函数的执行过程如下所示:

算法: GIoU 损失函数。

输入: 预测框 $B^p(x_1, y_1, x_2, y_2)$ 和真值(ground truth) $B^g(x_1, y_1, x_2, y_2)$ 。

输出: $L_{\text{IoU}}, L_{\text{GIoU}}$ 。

步骤 1: 分别计算预测框的面积 A^p 、ground truth 的面积 A^g ;

步骤 2: 找出最小的封闭矩形框 B^c , 且 B^c 要将 A^p 、 A^g 包含在内;

步骤 3: 计算封闭面积 B^c 中未覆盖 A^p 和 A^g 的面积部分占 B^c 总面积的比值;

步骤 4: 计算 IoU, GIoU 以及损失函数 $L_{\text{IoU}}, L_{\text{GIoU}}$;

步骤 5: 输出 $L_{\text{IoU}}, L_{\text{GIoU}}$ 。

2 改进的 YOLOV5 的面部关键点检测

2.1 面部关键点定位

面部关键点定位(facial landmark localization)是指在人脸检测的基础上自动定位出面部的关键特征点, 该检测分支对面部细节的精准定位提供了先决条件。

文章提出的关键点检测方法是在 YOLOV5 的基础上增加了 landmark 分支, 实现对左眼中心、右眼中心、鼻头、左边嘴角、右边嘴角五个面部关键点的定位。

如图 3 所示, 左边为未改进的图片标注, 右边为改进后的。



图 3 训练标注图片

2.2 损失函数

在分类任务中, 训练数据分布的不均衡会导致模型性能不佳。在面部关键点分支任务中, 若正脸多, 侧脸或有角度的脸太少都会导致模型对有旋转人脸的关键点检测精度低。

新增的面部关键点回归分支采用的损失函数为 Wing Loss^[15], 特别针对不同姿态角度的人脸检测, 其公式如式(3)所示。

$$\text{Wing}(x) = \begin{cases} w \ln(1 + \frac{|x|}{\varepsilon}) & |x| < w \\ |x| - C & |x| \geq w \end{cases} \quad (3)$$

其中, w 为正数, 将非线性部分限制在区间内, ε 为了防止误差导致梯度爆炸而设置为一个小的数值, 用来约束非线性区域的曲率。计算损失时, 特征图输出参数维度为 (batch_size, grid_size * grid_size * num_anchors, 5+class_num)。文中方法仅需要输出人脸类别, 因而 class_num=1。在输出的最后一个维度, 需要增加各关键点与预测框左上角顶点距离的偏移量, 共为 10 个参数。因此, 改进后的关键点分支中计算损失时, 特征图输出参数维度为 (batch_size, grid_size * grid_size * num_anchors, 16)。

2.3 模型设计

YOLOV5 的灵活性还体现在可以通过调整 depth_multiple 和 width_multiple 参数, 实现不同大小、复杂度的模型设计。depth_multiple 表示通道(channel)的缩放系数, 能够将配置里的 backbone 和 head 部分有关通道的设置全部乘以相应系数。而 width_multiple 表示 BottleneckCSP 模块的层缩放系数, 将所有的 BottleneckCSP 模块的 number 系数乘上该参数, 得到最终的层个数。

在实际的工业应用中, 人脸检测往往更注重实时性^[16]。YOLOV5 虽然相较于 YOLOV4^[17] 有一定精度的损失, 但大幅提升了检测速度, 这也是文章选择 YOLOV5 进行人脸检测的主要原因。特别地, 若取 YOLOV5 的最小模型 YOLOV5s, 处理速度高达 140FPS, 且训练收敛快, 具备较好的实用性。

3 测试实验

3.1 实验环境

所有实验均在 Genuine Intel(R) CPU i7-7700 @ 3.60 GHz 上进行,使用 GeForce RTX 2070 GPU,8 GB 内存。编程语言采用的是 Python 3.7,深度学习框架为 PyTorch 1.8.0,并另外使用 GPU 加速工具 CUDA 11.1.1。

3.2 实验数据集

实验数据集为 WIDER FACE^[18] 人脸数据集,其数据集详细分配情况依据表 1。该数据集一共有 32 203 张图片,标记了 393 703 个具有遮挡、模糊、尺度变化、光照变化和姿态变化等因素的人脸图像。各个子集中的数据根据人脸检测的难易程度还被划分为简单、中等和困难 3 个级别,分别对应 Easy、Medium 和 Hard。

表 1 WIDER FACE 数据集的分配情况

| 集合类型 | Train | Test | Val |
|-------|--------|--------|-------|
| 图片(张) | 12 880 | 16 097 | 3 226 |

3.3 结果分析

实验选取 YOLOV5s 在数据集 WIDER FACE 上训练 300 个 epoch,并与经典人脸检测模型 RetinaFace 做结果对比。改进的面部关键点的 Wing Loss 中参数设置为: $\varepsilon = 2, w = 10$ 。模型的人脸检测结果如表 2 所示。

结果表明,文中提出的 YOLOV5s+landmark 检测精度在验证集上的表现均超于 RetinaFace^[19] 的 MobileNet 版本;在 Easy 和 Medium 验证集上与 Retinaface 的 ResNet152 版本结果相近,但 YOLOV5s 的模型大小和计算量更小,因而更加高效;在 Hard 验证集上的平均检测精度(mAP)比 RetinaFace 高 0.6%,这体现出了该方法在人脸检测上的强大性能。

表 2 改进后的 YOLOV5s 和 RetinaFace 实验对比

| Backbone | Easy | Medium | Hard |
|------------------------|------|--------|------|
| RetinaFace-mnt | 90.4 | 88.1 | 73.8 |
| RetinaFace-ResNet50 | 95.4 | 94.0 | 84.4 |
| RetinaFace-ResNet152 | 95.7 | 95.2 | 87.5 |
| YOLOV5s+landmark(ours) | 95.1 | 94.3 | 88.1 |

选取一张包含大小不一人脸的图片作可视化对比,图 4 展示了各个模型的人脸检测结果。由图 4(a)可知,RetinaFace 同时给出较高置信度和面部关键点(landmark),但漏检了几张海报里的较小人脸;图 4(b)为未改进的 YOLOV5s 检测结果,虽然检测到了较多的人脸,但缺少了面部关键点,得到的面部信息较少;而图 4(c)为改进后的 YOLOV5s 人脸检测,既有较高的人脸检测精度,同时检测出面部关键点。



图 4 人脸检测结果

4 结束语

该文研究了 YOLOV5 人脸检测方法,并通过增加面部关键点(landmark)的检测对其进行了改进。Landmark 提供的辅助信息使模型能够对面部进行更

细致的分析,进一步提升了算法的性能。实验结果显示,基于 YOLOV5s 的人脸关键点检测精度在 WIDER FACE 的 Hard 验证集上高于经典人脸检测模型 RetinaFace 的 Resnet 版本,充分证明改进的 YOLOV5 在人脸检测任务上有较高的性能和实用性,有助于人

脸检测技术的推广和发展。

参考文献:

- [1] 王鑫,王忠举,李锐. 基于神经网络的人脸识别研究综述[J]. 信息与电脑,2020(23):56-58.
- [2] 陈科圻,朱志亮,邓小明,等. 多尺度目标检测的深度学习研究综述[J]. 软件学报,2021,32(4):1201-1227.
- [3] 吴雪,宋晓茹,高嵩,等. 基于深度学习的目标检测算法综述[J]. 传感器与微系统,2021,40(2):4-18.
- [4] 蓝雯飞,张盛兰,朱容波,等. 基于改进MTCNN的人脸检测算法[J]. 中南民族大学学报,2020,39(6):637-641.
- [5] 牛作东,覃涛,李捍东,等. 改进RetinaFace的自然场景口罩佩戴检测算法[J]. 计算机工程与应用,2020,56(12):1-7.
- [6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//IEEE conference on computer vision and pattern recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 6517-6525.
- [7] 杜虓龙,余华平. 基于深度学习的面部动作检测[J]. 软件导刊,2021,20(5):29-33.
- [8] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//IEEE conference on computer vision and pattern recognition (CVPR). Long Beach: IEEE, 2019.
- [9] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//IEEE conference on computer vision and pattern recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 2261-2269.
- [10] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//IEEE conference on computer vision and pattern recognition (CVPR). Salt Lake City: IEEE, 2018: 8759-8768.
- [11] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//European conference on computer vision. [s. l.]: Springer, 2015: 346-361.
- [12] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//IEEE international conference on computer vision (ICCV). Venice: IEEE, 2017.
- [13] REDMON J, FARHADI A. Yolov3: an incremental improvement[C]//IEEE conference on computer vision and pattern recognition (CVPR). Honolulu, HI, USA: IEEE, 2018.
- [14] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//IEEE conference on computer vision and pattern recognition (CVPR). Long Beach: IEEE, 2019.
- [15] FENG Z H, KITTLER J, AWAIS M, et al. Wing loss for robust facial landmark localisation with convolutional neural networks[C]//IEEE conference on computer vision and pattern recognition (CVPR). Honolulu, HI, USA: IEEE, 2018.
- [16] 蒋纪威,何明祥,孙凯. 基于改进YOLOv3的人脸实时检测方法[J]. 计算机应用与软件,2020,37(5):200-204.
- [17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[C]//IEEE conference on computer vision and pattern recognition (CVPR). Seattle: IEEE, 2020.
- [18] YANG S, LUO P, LOY C C, et al. WIDER FACE: a face detection benchmark[C]//IEEE conference on computer vision and pattern recognition (CVPR). Boston: IEEE, 2015.
- [19] DENG J K, GUO J, VERVERAS E, et al. RetinaFace: single-shot multi-level face localisation in the wild[C]//IEEE conference on computer vision and pattern recognition (CVPR). Long Beach: IEEE, 2019.
- [20] ZHU Y, CAI H, ZHANG S, et al. TinaFace: strong but simple baseline for face detection[C]//IEEE conference on computer vision and pattern recognition (CVPR). Seattle: IEEE, 2020.
- [21] PRASAD S, LI Y, LIN D, et al. MaskedFaceNet: a progressive semi-supervised masked face detector[C]//IEEE conference on computer vision and pattern recognition (CVPR). [s. l.]: IEEE, 2021.
- [22] LIU Y, TANG X, WU X, et al. HAMBox: delving into online high-quality anchors mining for detection[C]//IEEE conference on computer vision and pattern recognition (CVPR). Seattle: IEEE, 2020.