

DOI:10.16644/j.cnki.cn33-1094/tp.2023.07.004

# YOLOv5 算法的人脸识别检测方法研究

宋传旗

(山东建筑大学信息与电气工程学院, 山东 济南 250000)

**摘要:** 以 YOLOv5 算法为核心, 针对小尺度人脸的识别精准度不高问题进行研究。分析 YOLOv5 系统结构, 研究加权双向特征金字塔网络(BiFPN)的特征融合思想, 提出了权重的 BiFPN, 充分利用深层、浅层以及原始的特征信息, 加强了特征融合, 减少了卷积过程中特征信息的丢失, 提高了检测精度。采用 WiderFace 人脸数据集进行改进前后对比训练, 得出通过改进 YOLOv5 结构中的 Neck 部分, 使得算法在少量增加计算量和参数数量的情况下, 人脸识别精准度(Precision)达到 91.2%, 召回率(Recall)达到 83.4%, 检测速度也有所加快。具有较好的检测度与实时性。

**关键词:** 人脸识别; 双向特征金字塔网络 BiFPN; YOLOv5; 检测精度

中图分类号: TP391.4

文献标识码: A

文章编号: 1006-8228(2023)07-15-05

## Research on face recognition detection method of YOLOv5 algorithm

Song Chuanqi

(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, Shandong 250000, China)

**Abstract:** The YOLOv5 algorithm is used as the core to study the problem of low recognition accuracy of small-scale faces. The YOLOv5 system structure is analyzed, the feature fusion idea of weighted BiFPN is studied, and a weighted BiFPN is proposed, which makes full use of deep and shallow layers as well as the original feature information, strengthens the feature fusion, reduces the loss of feature information in the convolution process, and improves the detection accuracy. The WiderFace face dataset is used to compare the training before and after improvement, and it is concluded that by improving the Neck part of the YOLOv5 structure, the algorithm achieves 91.2% Precision and 83.4% Recall for face recognition with a small increase in computation and number of parameters, and the detection speed is also accelerated. It has good detection degree and real-time performance.

**Key words:** face recognition; bi-directional feature pyramid network (BiFPN); YOLOv5; detection accuracy

## 0 引言

深度学习在人工智能领域获得发展与广泛应用, 如自动翻译、人脸识别等智能化技术都离不开深度学习。基于深度学习的目标检测划分为单阶段目标检测和两阶段目标检测。两阶段目标检测步骤通过特征提取与目标检测两个部分来完成; 而单阶段目标检测则是将上述步骤合二为一来完成。两阶段目标检测算法以 R-CNN<sup>[1]</sup>、FastR-CNN<sup>[2]</sup>为代表, 需要先进行区域选择, 再进行分类; 单阶段目标检测以 YOLO<sup>[3]</sup>、SSD<sup>[4]</sup>算法为代表, 是将区域选择和分类合并到同一个网络结构中, 共同构建出以“分类+回归”为基础框架的多任务模型。

本文利用开源人脸数据集 WiderFace<sup>[5]</sup>作为训练数据集, 通过改进 YOLOv5 网络结构中的 Neck 部分, 弥补原始 YOLOv5 算法中特征金字塔网络(feature pyramid networks)FPN+PAN<sup>[6]</sup>特征融合结构缺少原始信息的缺点, 利用 BiFPN<sup>[7]</sup>的特征融合思想, 减少卷积过程中特征信息的丢失, 提升算法的检测精度。

## 1 YOLOv5 目标检测算法

### 1.1 YOLOv5 系列算法概述

YOLO 系列算法是应用于目标检测的深度学习神经网络模型。该系列算法模型可划分为两个任务部分: 一是对检测目标分类, 二是对检测目标的位置确认。

收稿日期: 2023-01-20

作者简介: 宋传旗(1995-), 男, 山东德州人, 大学本科, 工程师, 主要研究方向: 智能环境与网络化控制。

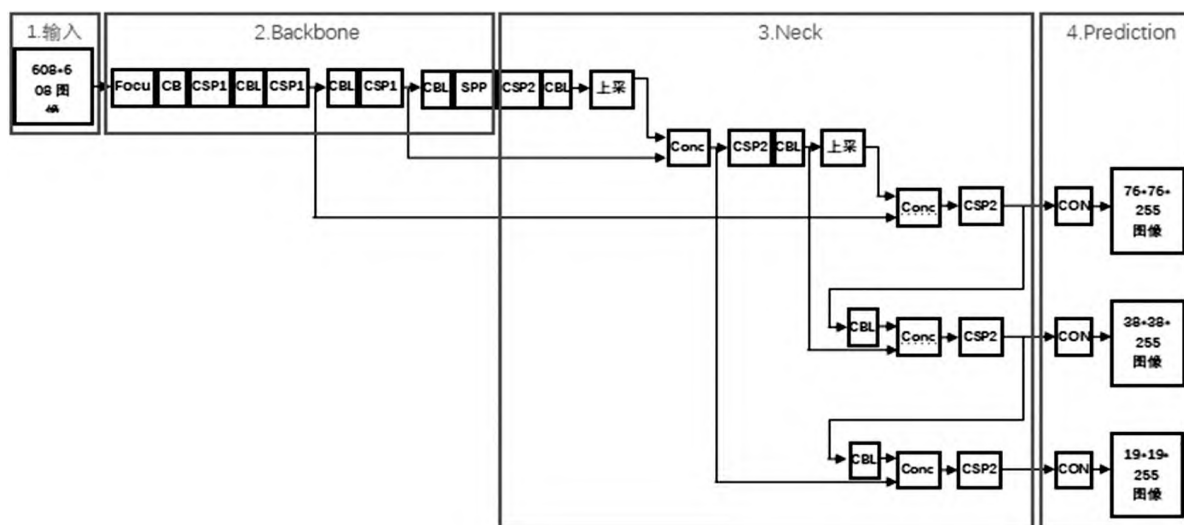


图1 YOLOv5网络结构图

YOLOv5的高性能且易于使用的特点,使其在目标检测领域运用更加广泛。本文以YOLOv5s模型来展开人脸检测算法研究,图1为YOLOv5的网络结构示意图。从图1可以看出YOLOv5的网络结构主要由四部分组成,分别是Input(输入端)、Backbone(基准网络)、Neck、Prediction(输出网络)。

## 1.2 YOLOv5算法的网络结构

首先Input(输入端)是输入原始数据图像,图像大小为608\*608,在该阶段使用Mosaic数据增强的方式以及自适应锚框、自适应的图片缩放对原始图像进行预处理,这种方式丰富了数据,提升了对小目标检测能力。

Backbone(基准网络)使用Focus模块实现了快速下采样,该模块能够在没有信息丢失的情况下将信息集中到通道上,使接下来的特征提取更加充分。SPP模块的使用则提高了图像的尺度不变性,有效增加了主干特征的接收范围,更容易使网络收敛,提高了准确率。

Neck网络结构采用FPN+PAN结构,FPN层向下采样,不断缩小图像,通过融合高低层特征提升目标检测的效果,后面添加PAN结构向上采样,充分利用网络浅层特征进行分割,顶层也可以接收底层带来的丰富位置信息。

Prediction(输出网络)用于完成目标检测结果的输出,为评价对象定位算法是否精准,YOLOv5采用损失函数GIOU<sup>[8]</sup>计算,以及采用非最大抑制算法把GIOU\_Loss贡献值不是最大的其他预测结果去掉,输出概率最大的分类结果通过分析YOLOv5算法模型的原理以及网络结构,发现在针对人脸检测识别过程中,

对小模型和超小模型的数据检测结果不精确。因此我们将BiFPN(Bidirectional Feature Pyramid Network,双向加权特征金字塔网络)特征融合结构引入YOLOv5算法,在特征提取网络中加入小目标检测层,提取浅层的物理信息来加强对小目标的检测能力。

## 2 YOLOv5改进

### 2.1 BiFPN头部特征融合网络

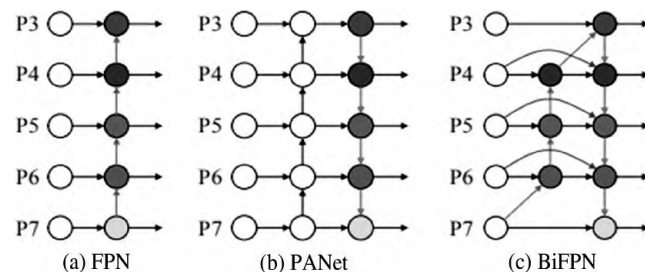


图2 不同特征融合结构示意图

YOLOv3采用的如图2(a)所示,FPN特征融合结构本质上受到单向信息流的限制。从上到下传输特征信息,并通过上采样和低级别特征融合高级别特征,以获得用于预测的特征图。YOLOv4和YOLOv5采用的如图2(b)所示结构,PAN特征融合结构增加了一个额外的自下而上的路径聚合网络,它将浅层定位信息传输到深层,并在多个尺度上增强定位能力。谷歌大脑团队提出的BiFPN特征融合结构如图2(c)所示。他们将双向(自下而上和自上而下)路径视为特征网络层,并多次重复同一层以实现更高级别的特征融合。他们在此结构的基础上提出了名为EfficientDet的目标检测算法,和同时期的其他目标检

测算法如 YOLOv3、MaskRCNN、NAS-FPN 等相比,在准确率、检测速度和运算量上 EfficientDet 均是一枝独秀,从图 3 可以看到详细的实验数据对比图。

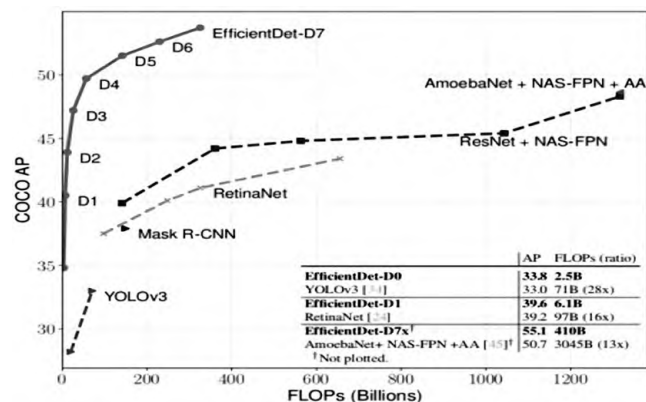


图3 EfficientDet与其他算法的性能对比图

## 2.2 Neck部分的改进

随着 YOLO 算法的发展, EfficientDet 算法的性能已经比不上 YOLOv4 和 YOLOv5,但是它提出的 BiFPN 特征融合结构却值得我们参考,所以我们将 BiFPN 特征融合结构融入到 YOLOv5 算法,改进后的 YOLOv5 结构中的 Neck 层的网络结构如图 5 所示。

YOLOv5 中的 PANet 通过自顶向下和自底向上的路径聚合实现了较好的多尺度特征融合结果,但计算量较大,且自底向上特征融合阶段的输入特征中并没有主干网络生成的原始输出特征,如图 4 所示。BiFPN 采用跨连接去除 PANet 中对特征融合贡献度较小的节点,在同一尺度的输入节点和输出节点间增加一个跳跃连接,在不增加较多成本的同时,融合了更多的特征。在同一特征尺度上,把每一个双向路径看作一个特征网络层,并多次反复利用同一层,以实现更高层次的特征融合。

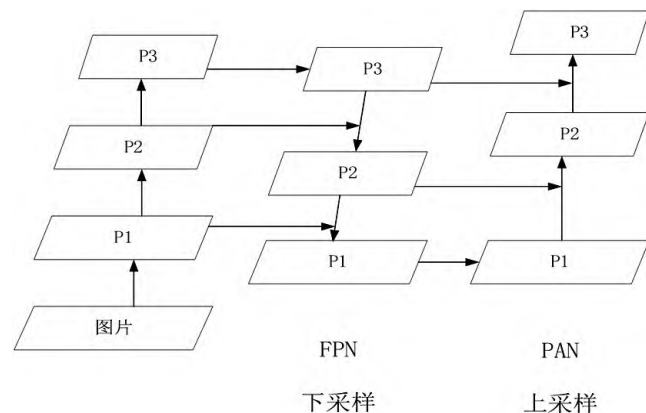


图4 改进前的 Neck 网络架构

算法改进后, PAN 结构每个结点的输入边都有一

条是来自主干特征提取网络部分的原始特征层,训练过程有了原始特征信息的持续参与,就可以有效避免训练学习过程偏离人们的期望,特征融合也更加合理有效,从而使得检测精度有所提升。

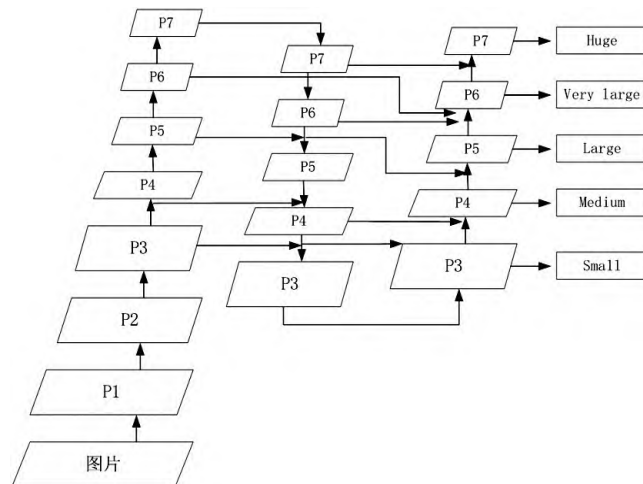


图5 改进后的 Neck 网络架构

## 3 实验与结果分析

### 3.1 实验环境

为了验证 YOLOv5Face 算法在人脸检测和识别中的效果,本次实验在 Ubuntu16.04 操作系统下完成软件平台搭建和实验, GPU 选用 NVIDIA GeForce GTX1660Ti, 软件工具: prcharm2022.1 社区版、Anaconda3、python3.7、Pytorch1.11.0、tensorflow-gpu1.8.1、cuda10.2。

### 3.2 实验数据集

本文的实验研究数据选用了 WiderFace 的公开人脸数据集,其中包含了 393703 个人脸图像,为了更好的检测识别功能,有利于对模型准确率的检测,其数据集中的所有图像均具备明显的差别。本次实验选用的 WiderFace 的公开人脸数据集由 61 个事件进行类别划分,其中每组事件选取其中 80% 图像为训练集,剩余 20% 为测试集。算法网络进行了 30 次迭代训练,为了加快训练速度并防止过拟合,初始学习率设置为 0.001,学习率衰减系数设置为 0.0001,权值衰减系数设置为 0.0005。

### 3.3 实验评判标准

在目标检测当中,通常使用精确率(precision)、召回率(recall)、PR 曲线、平均精度(average precision, AP)和多个类平均精度(mean average precision, mAP)来作为评价指标,平均精度与精确率和召回率有关。式(1)为精确率公式,式(2)为召回率计算公式。



$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

其中, TP代表被正确检测出来的目标数量; FP代表检测错误的目标数。TP和FP作为分母说明准确度在主观上评判预测的效果好坏。FN则代表未被检测出来的目标数。TP和FN作为分母说明召回率在客观程度评判预测效果好坏<sup>[9]</sup>。Precision代表预测正确的正样本个数与预测为正样本个数的比值, Recall则代表获得判定样本占总样本数的比重。

### 3.4 实验结果与分析

本文首先使用改进前的YOLOv5对WiderFace数据集的训练集和验证集上进行30个迭代周期的训练和验证, 训练结果曲线图中可见, 其中横向坐标的迭

代次数为30, 则表示其中有30个训练迭代周期, 纵坐标其中Box是BoundingBox的GIoU损失函数均值, 值越小边界框就越准; Precision代表分类器所衡量的正类; Recall则是正类中被正确预测的正确率; Objectness是目标检测loss均值, 越小代表目标检测的正确率越高。一般主要观察Recall和Precision的波动情况判断网络的训练效果, 通过改进后与改进前的Recall和Precision的数值通过图6与图7对比, 可知改进后检测准确度更高, 网络的训练效果更好。图8与图9是WiderFace数据集的训练结果对比。

为了验证Neck部分改进的有效性, 将原始YOLOv5算法中的FPN+PAN结构改为本文提出的权重的BiFPN结构, 其余部分保持不变, 然后将其与原始YOLOv5算法在WiderFace数据集上进行实验对比, 实验结果统计如表1所示。

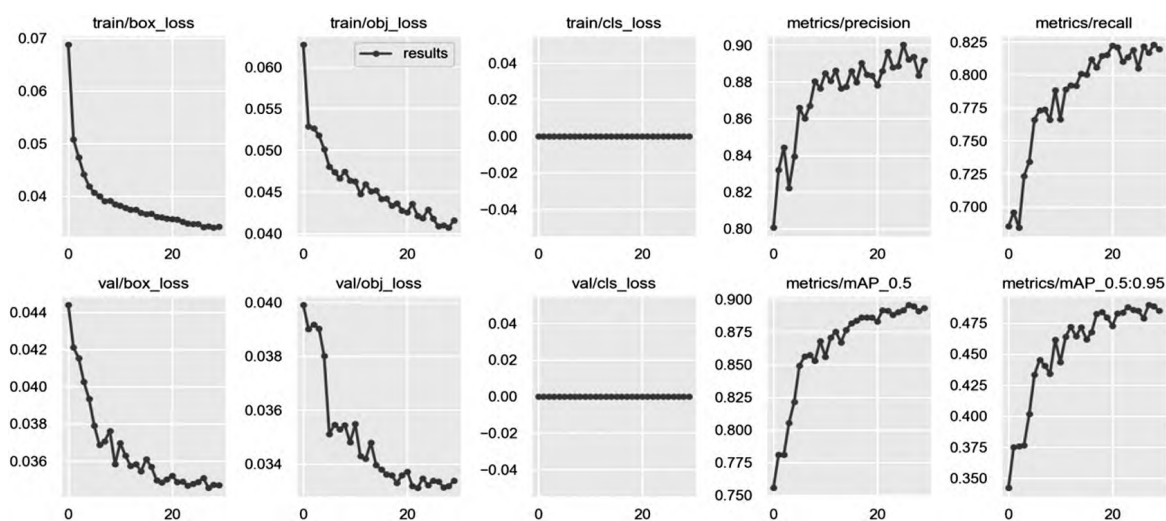


图6 改进前YOLOv5训练结果曲线

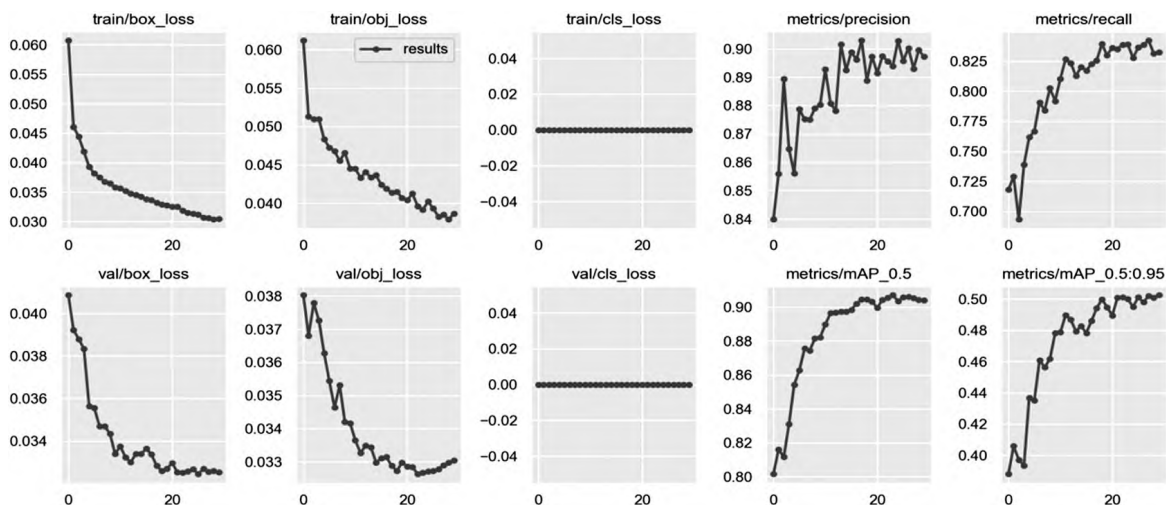


图7 改进后YOLOv5训练结果曲线



图8 改进前YOLOv5检测效果图

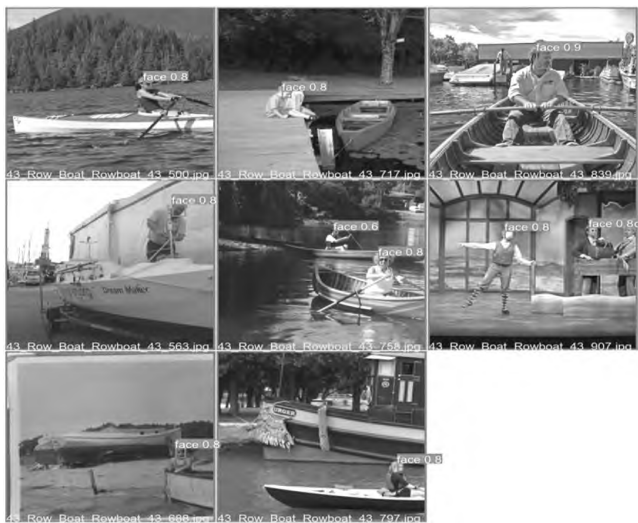


图9 改进后YOLOv5检测效果图

表1 人脸识别测试结果

算法模型	体积/ MB	计算量/ GFLOPS	参数量 /10 <sup>6</sup>	Precision	Recall	FPS
original_YOLOv5	14.4	16.3	7.06	90	82.5	82
improved_YOLOv5	16.6	17.8	8.16	91.2(+1.2)	83.4(+0.9)	84

基于上述图9中可以看出,通过改进Neck部分,使得算法在少量增加计算量和参数量的情况下,YOLOv5训练算法的人脸识别精准度达到91.2%,比改进前增加1.2个百分点,召回率达到83.5%,比改进前增加0.9个百分点,检测速度也有所加快,所进行的训练集试验中均可实现目标的有效识别,说明该改进在一定程度上加强了Neck部分的特征融合,证明了该改进方法的有效性。

4 结束语

本文通过改进YOLOv5的Neck部分网络结构,将

BiFPN特征融合结构引入YOLOv5算法,采用使PAN结构每个结点的输入边都有一条是来自主干特征提取网络部分的原始特征层,训练过程有了原始特征信息的持续参与特征融合也更加合理有效,从而使得检测精度有所提升。基于文中的实验检测分析结果来看,改进后的YOLOv5能够增强小尺度人脸的快速精准识别的准确度,也因此判定在司机疲劳检测中该算法具有极高的可行性。后续将结合车载摄像头进行人脸疲劳检测,更精确的识别人脸面部器官的变化来实现对列车司机的疲劳监控。

参考文献(References):

[1] GIRSHICKR, DONAHUEJ, DARRELLT, etal. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587

[2] GIRSHICKR. Fast R-CNN[C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448

[3] REDMONJ, DIVVALAS, GIRSHICKR, etal. You only look once: unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:779-788

[4] LIUW, ANGUELOVD, ERHAND, etal. SSD:single shot multi box detector [C]//European Conference on Computer Vision. Cham:Springer,2016:21-37

[5] YangS, LuoP, LoyCC, etal. WIDERFACE: a face detection benchmark [C]//IEEE Conference on Computer Vision & Pattern Recognition.[S.l.]:IEEE,2016(2)

[6] LIUS, QIL, QINH, etal. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:8759-8768

[7] TANM, PANGR, LEQV. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020:10781-10790

[8] 吕永丰,陆华才.基于YOLOv5算法的交通标志识别技术研究[J].电子测量与仪器学报,2021,35(10):137-144

[9] Z. Feng, J. Kittler, M. Awais, P. Huber, and X. Wu", Wing loss for robust faciall and mark localization with convolution alneural networks,"CVPR,2018

