

基于BiFPN改进的深度学习口罩人脸检测方法

于 晓, 张茂松, 周子杰

(天津理工大学, 天津 300384)

✉yx_ustb@163.com; 1798588432@qq.com; Zhou_Zijie@outlook.com



摘 要: 准确检测人员是否佩戴口罩对于保证食品生产环境的卫生、预防疾病的传播等具有重要意义。文章改进了YOLOv5算法的网络结构, 首先基于双向特征金字塔网络(BiFPN)改进了YOLOv5网络的颈部(Neck)结构, 然后使用K-means聚类算法对数据集进行聚类分析, 提出YOLOv5_BM口罩人脸检测算法。在自制人脸口罩数据集上的测试结果表明, YOLOv5_BM算法的平均精度高达95.3%, 相比YOLOv5网络提升了3.8%。在公开数据集上与其他经典的目标检测算法相比, YOLOv5_BM算法在性能方面也取得了提升, 相比SSD算法, YOLOv5_BM算法的平均精度提高了4.4%; 相比YOLOv3算法, YOLOv5_BM算法的平均精度提高了2.9%。

关键词: 口罩人脸检测; YOLOv5; 双向特征金字塔网络; K-means聚类算法

中图分类号: TP751.1 **文献标识码:** A

An Mask Face Detection Method of Improved Deep Learning based on BiFPN

YU Xiao, ZHANG Maosong, ZHOU Zijie

(Tianjin University of Technology, Tianjin 300384, China)

✉yx_ustb@163.com; 1798588432@qq.com; Zhou_Zijie@outlook.com

Abstract: Accurately detecting whether staff wear masks or not is of great significance for ensuring the sanitation of food production environment and preventing the spread of diseases. This paper proposes a YOLOv5_BM mask face detection algorithm by improving the network structure of YOLOv5 algorithm. Firstly, the neck part structure of YOLOv5 network is improved based on Bi-directional Pyramid Network (BiFPN); then K-means clustering algorithm is used to cluster the data set. The test results on the self-made face mask dataset show that the average accuracy of YOLOv5_BM algorithm is 95.3%, which is 3.8% higher than YOLOv5 network. Compared with other classical target detection algorithms on the public dataset, YOLOv5_BM algorithm has also improved its performance. Compared with SSD (Single Shot MultiBox Detector) algorithm, the average accuracy of YOLOv5_BM algorithm has been improved by 4.4%; compared with YOLOv3 algorithm, the average accuracy of YOLOv5_BM algorithm is improved by 2.9%.

Keywords: mask face detection; YOLOv5; BiFPN; K-means clustering algorithm

1 引言(Introduction)

准确检测人员是否佩戴口罩对于保证食品生产环境的卫生、预防疾病的传播、减少安全事故的发生^[1]等具有关键作用。因此, 设计高精度的、可用于人群场景的口罩人脸目标检测算法具有重要意义。

研究者在口罩人脸检测领域已进行了大量研究。牛作东等人提出了一种改进过的RetinaFace算法^[2]; 邓黄潇将迁移学习与RetinaNet模型结合, 在口罩人脸检测的任务中表现出

良好的性能^[3]。随着目标检测技术的蓬勃发展, 基于深度学习的检测方法相继出现^[4], 其可以分为两大类: 两阶段目标检测算法(Two-Stage)^[5]和单阶段目标检测算法(One-Stage)^[6]。研究人员基于深度学习算法提出了新的口罩人脸检测模型, 王艺皓等人通过改进YOLOv3完成了复杂场景下的口罩人脸检测任务^[7]。管军霖等人证明了YOLOv4算法可以应用在口罩人脸检测任务中^[8]。王沅改进了YOLOv5算法, 并将提出的算法应用于口罩及安全帽的检测^[9]。

在口罩人脸检测任务中,检测精度极其重要,在现实场景中,行人密集且存在被物体遮挡的现象,使得口罩人脸检测难度增大。当前,存在的口罩人脸检测算法的检测精度仍需进一步提高。

随着YOLO系列算法的不断发展,YOLOv5已具备很高的精度和速度^[10],具有较好的工程实用性。考虑到终端的计算能力,本文以YOLOv5为基础,基于双向特征金字塔网络(BiFPN)结构修改YOLOv5原有的FPN(特征金字塔网络)结构。此外,本文以现有的口罩人脸数据集为基础,加入互联网搜集的口罩人脸图片,得到自制口罩数据集。基于以上模型和数据集,本文提出YOLOv5_BM口罩人脸检测算法。

2 改进后的YOLOv5_BM网络(Improved YOLOv5_BM network)

YOLOv5网络结构可以分为输入端、Backbone(主干网络)、Neck(颈部网络)、Prediction(预测层)四个部分。输入端有Mosaic数据增强、自适应图片缩放及自适应锚框计算,Backbone包含Focus(切片)结构、CSP(跨阶段局部单元)结构,Neck包括FPN+PAN(路径聚合网络)结构,Prediction中有GIOU_Loss(广义交并比损失函数)结构。

YOLOv5网络的输入端采用Mosaic数据增强方法,Mosaic数据增强过程使用到四张图片,每次数据增强过程中先读取四张图片,然后按照顺序对四张图片进行翻转、缩放等变化,接着按照四个方向把处理后的图片放置好,最后依照图片位置的区别将四张图片进行组合。此外,YOLOv5网络为了将不同大小的输入图片变为相同尺寸的图片,采用自适应图片缩放的方法。图片两边填充的黑边较多时会影响YOLOv5网络的检测速度,为了解决这个问题,YOLOv5网络对letterbox函数进行改进,使输入图片能够自适应减少信息冗余,从而使黑边的数量减少。自适应缩放的步骤可以分为三步:首先计算缩放比例,然后计算输入图片缩放后的尺寸大小,最后计算需要填充的数值。除了Mosaic数据增强和自适应图片缩放,YOLOv5网络在输入端还使用了自适应锚框计算方法,该方法可以为不同的数据集设置初始长宽的锚框。YOLOv5网络在训练过程中,会在设定的初始锚框基础上得到输出预测锚框,并且计算预测锚框和Ground Truth(真实框)的误差,进而通过反向传播更新数据,通过多次迭代网络参数达到最优的锚框计算水平。

Backbone用于提取图像的识别特征,包括边缘特征、纹理特征和位置信息等,YOLOv5网络的Backbone层包含Focus结构、CSP结构、SPP(空间金字塔池化)结构。Focus结构可以先对输入信息进行切片操作,然后进行拼接,这样能够在进行下采样的同时保留更多的特征信息。CSP结构可以减少神经网络的计算量,对网络进行局部特征的跨层融合,进而增强网络的特征提取能力。SPP模块对输入的图片先进行卷积操作,然后执行三个不同尺寸的池化操作,最后将输出特征进

行特征融合,增大了感受野,提取出最重要的上下文的特征。

Neck结构可以对不同阶段提取的特征图进行再处理,从而更好地使用在主干网络中提取的特征。YOLOv5网络的Neck结构和YOLOv4网络相同,仍然沿用了FPN+PAN的结构,但是YOLOv5网络在YOLOv4网络的基础上做了一些调整,YOLOv5网络在FPN层的后面还添加了一个自底向上的特征金字塔,提高了YOLOv5网络的特征提取能力。

输出端部分用来检测目标对象的类别和位置,采用GIOU_Loss函数作为损失函数,并通过非极大值抑制筛选目标框。

虽然YOLOv5网络具备良好的检测性能,但它是以COCO(上下文中的公共对象)数据集为基础,如果应用于口罩人脸检测领域,存在对小目标的检测能力有限的问题,并且当目标存在密集、模糊等情况时,该问题更突出。此外,YOLOv5网络的特征增强模块PANet更侧重于深度特征图的融合,影响了对小目标的检测能力,增加了干扰噪声,影响了检测精度。

本文首先从Neck部分对YOLOv5网络进行改进,基于BiFPN结构修改YOLOv5网络的Neck部分,将YOLOv5网络Neck层中的节点连接方式做出部分改变,减少了对网络特征融合贡献度小的不必要连接,增加了输入节点和输出节点处于同一层时二者的连接,并且为输入特征增加了可更新的权重,使网络可以不断调整权重,确定每个输入特征对输出特征的重要性,增强了网络的特征提取能力;其次使用K-means聚类算法对训练数据进行聚类分析,提高模型的泛化能力和定位精度,进而提升口罩人脸检测算法的性能。改进后的YOLOv5_BM网络结构图如图1所示。

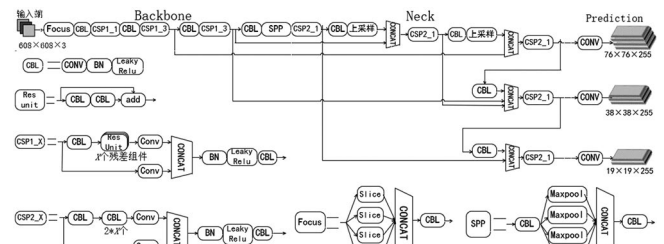


图1 YOLOv5_BM结构

Fig.1 YOLOv5_BM structure

2.1 FPN

YOLOv5网络的Neck结构中,采用了FPN+PAN的结构,其结构图如图2所示,前两列自底向上路线和自顶向下路线为FPN结构^[11],最右侧一列自底向上的结构为PAN结构。FPN结构自上而下建立了一条通路进行特征融合,在特征融合之后,使用更高语义特征的特征层进行预测,但是FPN结构受制于单向信息流。在FPN的基础上,PAN结构被提出,该结构基于FPN结构再建立了一条自下而上的通路,将底层的位置信息传到预测特征层中,使得预测特征层中同时包含

来自顶层的语义信息和来自底层的位置信息,从而大大提升了算法的目标检测精度。

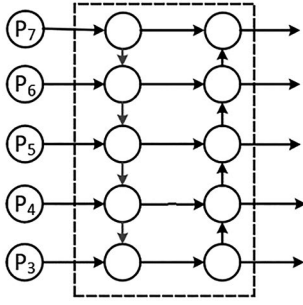


图2 FPN+PAN 结构

Fig.2 FPN+PAN structure

2.2 BiFPN

BiFPN是在PAN的基础上进行改进的。双向特征金字塔结构(BiFPN)运用双向融合思想,重新构造了自顶向下和自底向上的路线,对不同尺度的特征信息进行融合,通过上采样和下采样统一特征分辨率尺度,并且在同一尺度的特征图之间建立了双向连接,在一定程度上解决了特征信息遗失的问题^[12],能够更好地融合不同尺寸特征图的特征信息。BiFPN的结构图如图3所示,由于只有一个输入的节点对特征融合的贡献度较低,所以为了简化网络BiFPN将其删去;在原始输入节点和输出节点之间增加了一条边,旨在不消耗成本的情况下进行更多的特征融合;将自上向下和自底向上的路径融合到了一个模块,以便实现更高层次的特征融合。

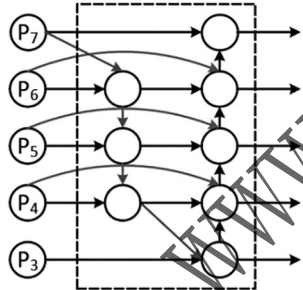


图3 BiFPN 结构

Fig.3 BiFPN structure

不同于传统的特征融合, BiFPN为了学习不同的输入特征的重要性,还对不同输入特征进行了区分的融合,是一种加权融合的机制,对此BiFPN使用了快速归一化方法,精度与softmax-based fusion(基于Softmax的融合)类似但速度更快,快速归一化方法如式(1)所示。

$$Out = \sum_i \frac{w_i}{\varepsilon + \sum_i w_i} \cdot In_i \quad (1)$$

式(1)中, w_i 代表权重,在激活函数ReLU的作用下可以保证 $w_i \geq 0$, ε 用来避免数值不稳定,是一个很小的值, In_i 表示输入的特征, Out 表示加权特征融合的结果。

2.3 K-means聚类优化锚框尺寸

K-means是一种聚类算法,适用于数据量较大的数据

集。K-means算法的复杂程度低,聚类效果好,比较容易实现,因此其应用领域非常广泛。YOLOv5网络模型的初始锚框是由COCO数据集设计生成的九个锚框,涵盖了大小不同的目标且目标的大小差距较大。本实验训练的数据集大多数都是小目标,部分锚框的设计合理性有待提高,因此需要聚类优化锚框尺寸,使其与训练的数据集更加匹配,进而提升模型的检测精度。

K-means算法的具体步骤有五步:(1)随机设置类簇的个数 K ;(2)在数据集中随机选取 K 个默认宽高比,将它们作为类簇的初始质心;(3)计算所有样本与类簇中心的距离,再将所有样本划分到离它最近的类簇中;(4)更新类簇的中心,计算每个类簇中所有样本的均值,并将其作为新的聚类中心;(5)重复步骤(4)(5)直到类簇中心不再变化。

3 实验结果及分析(Experimental results and analysis)

3.1 实验环境与数据集

本实验平台为Intel Core i5-6300HQ处理器,实验的配置环境为Windows 10操作系统。在Anaconda中,基于Python 3.8搭建了虚拟环境,深度学习框架为PyTorch 1.7.0。

本实验基于RMFD(真实口罩人脸数据集)、MAFA(遮挡人脸检测数据集)、WIDER Face(人脸识别网络图像数据集)三个人脸数据集,同时加入了互联网搜集的图片,对数据集图片进行数据清洗,并通过LabelImg(图片标注工具)对数据集进行了标注,将图片中佩戴口罩的人脸标注为mask,未佩戴口罩的人脸标注为nomask,得到了自制口罩数据集。自制口罩人脸数据集共包括11,000余张口罩人脸图片,每张图片有对应的标签文件。

3.2 评价指标

为了验证本文提出的改进算法的性能,本文选择在目标检测领域中常用的 mAP_{50} 和 $mAP_{0.5:0.95}$ 指标,以及准确率 P 和召回率 R 。准确率和召回率的公式如式(2)、式(3)所示。

$$P = TP / (TP + FP) \quad (2)$$

$$R = TP / (TP + FN) \quad (3)$$

其中, TP 表示样本中被正确划分为正例的个数, FP 是样本中被错误分类为正例的负例的个数。 mAP 为平均的精度均值,其计算公式如式(4):

$$mAP = \frac{\sum_{c=1}^n AP_i}{n} \quad (4)$$

式(4)中, n 表示数据集中类别的数量; AP_i 代表第 i 类的平均精度(AP),即 $P-R$ 曲线的曲线下的面积。基于所有预测框的精确率和召回率,可以绘制出 $P-R$ 曲线。 mAP_{50} 是指在 IOU 阈值取0.5时的 mAP 值。然后将 IOU 阈值从0.5开始以0.05为步长逐渐增加到0.95,获得不同阈值下的 mAP 值,最后求取它们的平均值即可获得 $mAP_{0.5:0.95}$ 。

3.3 实验结果比较

本文将数据集分为训练集和数据集，其中9,000 张用于训练，2,000 张用于测试，在YOLOv5_BM网络的训练过程中，参数值的设定和损失函数系数设置采用YOLOv5网络所提供的基本设定，输入图片的分辨率为640×640 像素。模型初始学习率为0.01，权重衰减系数为0.0005，*epoch*设为300。

为了验证本文提出的YOLOv5_BM网络的口罩人脸检测效果，将原始YOLOv5算法与本文提出的YOLOv5_BM算法在本文自制数据集上进行口罩人脸检测的性能对比，实验的结果如表1所示。

表1改进前后的YOLOv5网络的性能对比
Tab.1 Performance comparison of YOLOv5 network before and after improvement

算法	<i>P</i>	<i>R</i>	<i>mAP50</i>	<i>mAP0.5:0.95</i>
YOLOv5	0.932	0.862	0.915	0.607
YOLOv5_BM	0.958	0.921	0.953	0.656

由表1可知，改进的YOLOv5_BM算法和原始YOLOv5算法在口罩人脸检测中均取得了较好的效果。相比原始YOLOv5网络，YOLOv5_BM网络的准确率、召回率分别上升了2.6%、5.9%，*mAP50*指标提升了3.8%，*mAP0.5:0.95*指标则提升了4.9%。这说明相比原始YOLOv5网络，加入YOLOv5_BM网络确实能够提高检测精度，并且算法性能得到了提升。

在实际人群场景中，口罩人脸目标存在密集、侧脸等情况，导致算法检测难度增大。因此，本文设计了在此情况下的算法检测结果对比。

如图4和图5所示，测试图片中的人脸目标交错且存在口罩人脸目标为侧脸的情况，口罩人脸目标的检测难度相对较大。如图4中(a)、(b)所示，原始YOLOv5算法存在漏检、误检的情况，未检测到测试图片1中的侧脸口罩人脸目标，并且将测试图片2中佩戴口罩的口罩人脸目标标记成nomask。相比之下，检测效果如图5中(a)、(b)所示的YOLOv5_BM算法的检测性能更好，检测结果均正确。相比YOLOv5算法，本文改进的YOLOv5_BM算法具备更高的检测精度，即使在密集、侧脸目标的情况下，YOLOv5_BM算法仍然具备良好的检测效果，不易出现误检、漏检的情况。



(a)YOLOv5测试图片1检测效果 (b)YOLOv5测试图片2检测效果
图4 YOLOv5网络口罩人脸检测效果图
Fig.4 Effect picture of YOLOv5 network mask face detection



(a)YOLOv5_BM测试图片1检测效果 (b)YOLOv5_BM测试图片2检测效果
图5 YOLOv5_BM网络口罩人脸检测效果图
Fig.5 Effect picture of YOLOv5_BM network mask face detection

此外，为了验证本文提出的YOLOv5_BM算法在检测精度方面的优越性，在自建人脸口罩数据集上将提出的YOLOv5_BM算法与其他经典目标检测算法的检测效果进行比较，检测结果如表2所示，相比于SSD算法，本文算法的*mAP50*指标提升了4.4%，相比于YOLOv3算法，本文提出的算法的*mAP50*指标提升了2.9%。在AIZOO数据集上将本文所提出的算法与其他经典目标检测算法进行对比试验，各类算法的*mAP50*和*mAP0.5:0.95*指标如表3所示。由表2和表3知，YOLOv5_BM算法在检测性能方面均保持着较大的领先优势。

表2 基于RMFD数据集的算法性能比较
Tab.2 Comparison of algorithm performance based on RMFD dataset

算法	<i>mAP50</i>	<i>mAP0.5:0.95</i>
SSD	0.897	0.586
YOLOv4	0.888	0.574
YOLOv3	0.912	0.595
YOLOv5	0.915	0.607
YOLOv5_BM	0.941	0.638

表3 基于AIZOO数据集的算法性能比较
Tab.3 Comparison of algorithm performance based on AIZOO dataset

算法	<i>mAP50</i>	<i>mAP0.5:0.95</i>
NanoDet-320	0.913	0.60
YOLOv4-tiny	0.925	/
MSAF R-CNN	0.904	/
YOLOv5	0.949	0.645
RetinaFaceMask	0.948	/
RetinaFace	0.93	/
SL-FMDet	0.938	/
YOLOv5_BM	0.953	0.659

4 结论(Conclusion)
本文提出了一种基于YOLOv5网络改进的口罩人脸检测算法，首先基于BiFPN结构在Neck部分对原始网络进行优化
(下转第14页)