

基于深度学习的口罩佩戴识别技术研究

周家慧, 王静媛

北方工业大学理学院, 北京

收稿日期: 2023年7月24日; 录用日期: 2023年9月19日; 发布日期: 2023年9月26日

摘要

本文基于目标检测中口罩佩戴识别的相关研究, 针对人群环境下口罩佩戴检测时中小尺寸目标识别等问题, 通过原网络中用稠密模块代替原有的残差模块以及优化损失函数对YOLOv3模型进行改进。实验结果表明, 改进后的算法识别平均精度达到95.35%, 相较于原YOLOv3算法提升了4.49%, 同时该算法还具有较高的检测速度, 能够满足实时检测的需求, 在实际应用中可以将其嵌入移动设备, 能够有效解放人力, 减少管理成本。

关键词

深度学习, 目标检测, YOLOv3模型; 疫情防控

Research on Mask Wearing Recognition Technology Based on Deep Learning

Jiahui Zhou, Jingyuan Wang

College of Science, North China University of Technology, Beijing

Received: Jul. 24th, 2023; accepted: Sep. 19th, 2023; published: Sep. 26th, 2023

Abstract

This paper is based on relevant research in mask-wearing detection for object recognition. Addressing issues such as small-sized object identification in crowded settings, it improves the YOLOv3 model by replacing the original residual modules with dense modules and optimizing the loss function. Experimental results demonstrate that the improved algorithm achieves an average precision of 95.35% in mask-wearing detection, exhibiting a 4.49% enhancement compared to the original YOLOv3 algorithm. Furthermore, this algorithm maintains a high detection speed, catering to real-time detection requirements. It can be embedded in mobile devices for practical application, effectively liberating human resources and reducing management costs.

Keywords

Deep Learning, Object Detection, YOLOv3 Model, Epidemic Prevention and Control

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自 2019 年末以来, 新型冠状病毒以极快的传播速度在全球肆意蔓延, 给各地疫情防控工作带来了巨大的挑战。在这一背景下, 全球范围内推广佩戴口罩成为一项重要的防控措施, 它能有效地阻隔空气中的飞沫以及游离病毒进入呼吸道, 从而减少感染几率。如在人口密集的公共场所(机场、车站、商场、学校等)应用口罩佩戴识别系统可以帮助卫生部门和政府实时监测佩戴率, 提前发现和控制潜在的传染风险, 从而保障公众健康安全。

目前有两种主要的目标检测算法, 一种是基于图像处理的传统目标检测算法, 另一种是基于卷积神经网络的目标检测算法。传统的目标检测算法通过不同尺寸的滑动窗口或图像分割技术形成大量候选框, 在对每个区域的图像特征, 如方向梯度直方图、尺度不变特征[1]进行提取之后, 传递给分类器, 如支持向量机(SVM) [2]、K 近邻(KNN)、决策树(Decision Tree)等对该候选区域的分类。由于使用人工提取特征的方法, 传统的目标检测存在特征表达能力弱、检测准确度不高、窗口冗余等问题。随着深度学习技术的不断发展, 基于深度学习的目标检测算法在准确率和效率方面持续改进, 成为目前主流的研究方向。其中, 著名的算法包括 Faster R-CNN、YOLO (You Only Look Once)、SSD (Single Shot Multibox Detector) 等。这些方法能够高效地检测图像中的目标, 并且在复杂场景下表现较好。YOLOv3 是 YOLO 系列的第三代算法, 最早由 Joseph Redmon 等人在 2018 年提出[3]。自发布以来, 许多研究人员对 YOLOv3 进行了优化和改进。如张富凯等人在 YOLOv3 网络的基础上, 提出了改进的快速车辆检测方法[4], 孙迎春等人提出了交通灯检测算法[5], 葛雯等人提出了行人识别算法[6], 这些改进后的 YOLOv3 算法在特定场景下表现出了不错的检测效果。

本文针对人群环境下口罩佩戴检测中存在的中小尺寸目标等问题, 对基于 YOLOv3 算法的模型进行改进与优化, 以实现人群口罩佩戴检测任务。

2. 基于 YOLOv3 的改进目标检测识别算法

2.1. YOLOv3 模型简介

YOLOv3 是单阶段检测的一种经典算法, 它的优点在于简化了目标识别过程, 直接通过对视频、实时源、图像等进行卷积网络回归而实现对物体的检测, 大幅度提高了检测的效率, 具有优异的时效性。

YOLOv3 主干网络采用了 darknet-53 网络, 一共提取了三个特征层进行目标检测, 分别位于主干特征提取网络的中间层、中下层和底层, 用于对大、中、小目标物体进行多尺度的检测。为了获得更强的特征表现力, 将 3 个不同尺寸的特征图进行特征融合, 中间经过上采样以及下采样, 使特征图大小相同, 然后进行堆叠、融合及相应的卷积等操作后, 得到最终的 3 个特征层。结构图 1 中, DBL 为 YOLOv3 的基本组件, 在卷积部分使用了特有的 DarkNetconv2D 结构, 卷积后进行了标准化与 Leaky ReLU 激活。另外, YOLOv3 最重要的改进为引用了 ResNet 深度残差网络, 通过使用这种改进措施可以增加网络结构

层数, 利用恒等映射的优势, 对模型训练和图像特征提取效果起到质的突破。

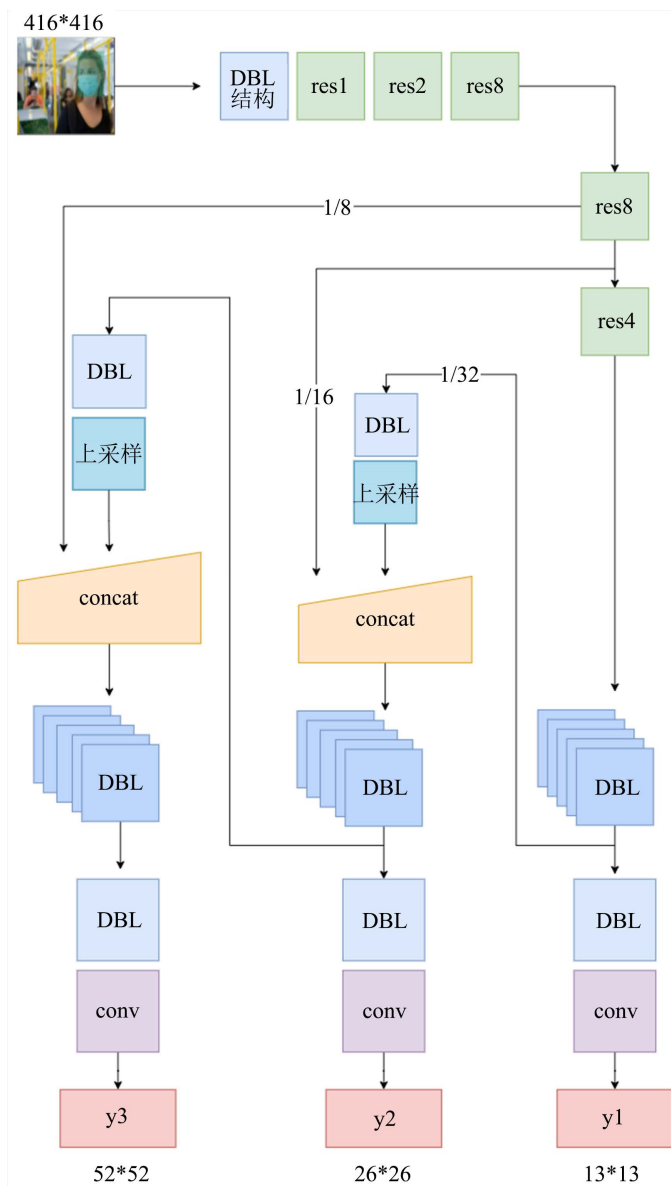


Figure 1. Diagram of YOLOv3 network architecture

图 1. YOLOv3 网络结构图

YOLOv3 相比于初代 YOLO 网络在精度和速度上有了很大程度的提高, 但针对输入的图像目标尺寸过小、具体应用场景复杂等情况, 该算法仍然存在敏感程度一般、精确度较低、达不到实时性要求等问题。针对以上问题, 本文从网络结构和损失函数入手进行改进, 以提高检测系统识别中小尺寸目标时的检测精度。

2.2. 模型改进

2.2.1. 基于稠密网络的 YOLOv3 检测模型

口罩作为小目标物体在目标检测中常常存在轮廓模糊, 不易识别的问题, 在实际应用中很容易被错

检或误检。本文提出一种融合 YOLOv3 和 DenseNet 的改进网络。在该网络中, 用稠密模块取代传统的残差模块, 以获得更加丰富的低层细节特征信息。通过对该尺度目标进行集中优化训练, 能够有效提升对中小尺度目标的敏感程度。

DenseNet 是由 Gao Huang [7] 等人于 2016 年提出, 其主要特点是通过在网络层之间构建紧密的连接, 从而使其在训练过程中更加紧凑和稠密, 同时还能有效解决梯度消失问题。DenseNet 的核心思想是在每一层将前面所有层的特征图连接在一起。具体来说, DenseNet 由多个密集块(Dense Block)和过渡层(Transition Layer)组成, 其结构图如图 2 所示。

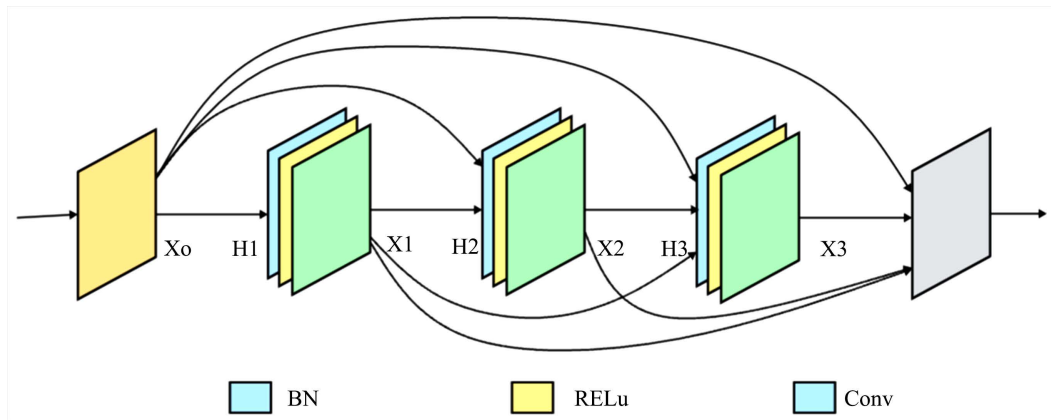


Figure 2. Schematic diagram of Dense Block Structure

图 2. Dense Block 结构示意图

图 2 中, X_0 是 H_1 的输入, H_2 的输入为 X_0 和 X_1 , H_3 的输入为 X_0 、 X_1 、 X_2 , H_3 的输出即为 H_3 , 则第 P 层的输出可表示为式 1。

$$X_P = H_P([X_0, X_1, \dots, X_{P-1}]) \tag{1}$$

其中, H_P 是一个非线性组合函数, $[X_0, X_1, \dots, X_{P-1}]$ 是表示一个通道合并运算, 用于表示前 $P-1$ 层特征图的输出。

密集块是网络中的基本模块, 它由多个卷积层组成, 每个卷积层之后都紧密连接着前面所有层的特征图。这样的连接方式有助于信息的充分传递, 同时减少了特征图的重复计算, 使得网络更加高效, 具备抗拟合性和参数效率高的优势, 为提升口罩佩戴检测模型的性能提供了新思路。

将 YOLOv3 网络结构中的 Residual Block 替换为 Dense Block, 并更改其连接方式, 即可完成替代工作。将 Darknet-53 网络中用于提取 52×52 、 26×26 、 13×13 三种尺度特征图的残差块替换成密集连接块, 即密集连接块是由多个结构相同的 Resnet 残差单元组成的密集连接网络, 其结构如图 3 所示。

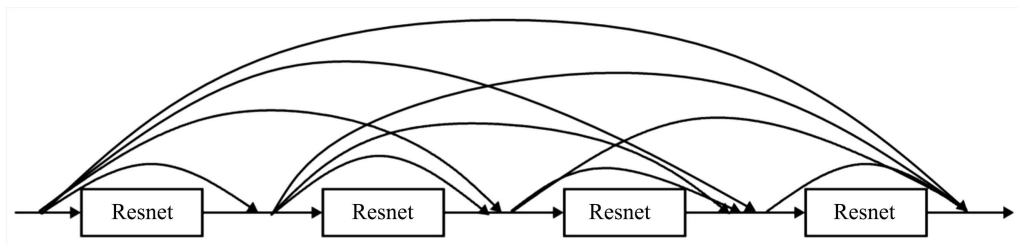


Figure 3. Diagram of Dense connection block with Resnet

图 3. 含有 Resnet 的密集连接块

具体来说, 即用 4 个改进后的 Resnet 残差块将 13×13 特征图密集连接块以稠密连接方式连接在一起; 而用于提取 52×52 、 26×26 特征图的密集连接块由 8 个改进后的 Resnet 残差块以密集连接方式连接在一起, 将该模型记为 DS-YOLOv3 模型, 其网络结构如图 4 所示。

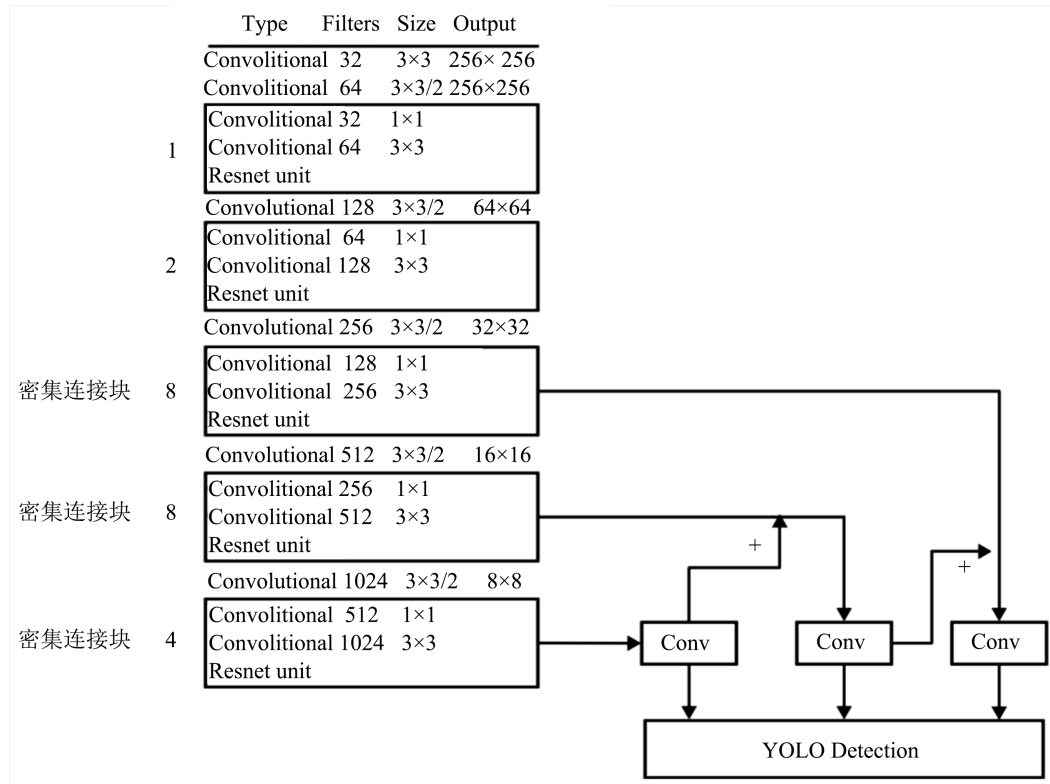


Figure 4. Diagram of DS-YOLOv3 network architecture
图 4. DS-YOLOv3 网络结构图

2.2.2. 改进损失函数

在 YOLOv3 中, 损失函数包含三个主要组成部分: 边界框回归损失、目标置信度损失以及分类损失。计算公式如式 2。

$$Loss = Loss_{loc} + Loss_{conf} + Loss_{class} \tag{2}$$

其中, 边界框回归损失衡量了预测边界框和真实边界框之间的位置误差, 计算公式如式 3。

$$Loss_{loc} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \tag{3}$$

在置信度和分类损失上采用的是二元交叉熵损失函数, 如式 4 和式 5 所示。

$$Loss_{conf} = \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{obj} \left[C_i - \hat{C}_i \right]^2 \tag{4}$$

$$Loss_{class} = \lambda_{class} \sum_{i=0}^{S^2} 1_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \tag{5}$$

其中, S 是每个边界框预测的网格单元数, B 是每个网格单元预测的边界框数量, λ_{coord} 是边界框回归损

失的权重系数, i 表示第 i 个网格单元, j 表示第 j 个边界框, 1_{ij}^{obj} 是一个指示函数, 当网格单元 i 的第 j 个边界框包含对象时为 1, 否则为 0。 C 是类别数, $p_i(c)$ 是预测网格单元 i 的第 j 个边界框属于类别 c 的置信度, $p_i(c)$ 是真实标签中网格单元 i 的第 j 个边界框属于类别 c 的标签值。

IOU 表示检测结果的矩形框与样本标注的矩形框之间的交并比, 用于反映预测框和真实框之间的关联程度。

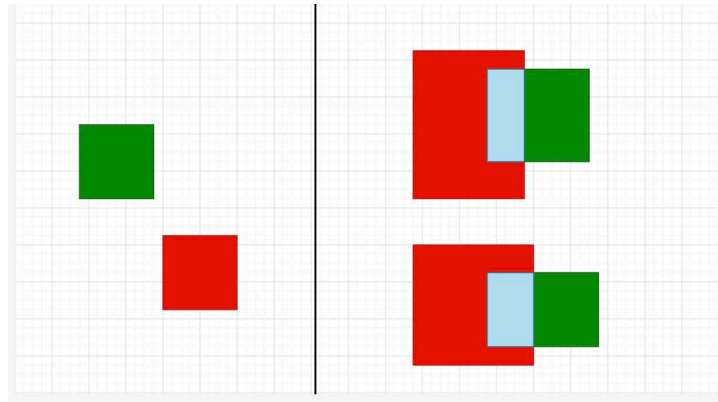


Figure 5. Diagram of intersection-over-union
图 5. 交并比 IOU

如图 5 所示, 红框为预测框, 记为 B^l , 绿框为目标框, 记为 B^{gt} , B^l B^{gt} 相交区域为 $B^l \cap B^{gt}$, B^l B^{gt} 合并区域为 $B^l \cup B^{gt}$, 则 IOU 的计算公式如式 6。

$$IOU = \frac{B^l \cap B^{gt}}{B^l \cup B^{gt}} \quad (6)$$

IOU 作为评价预测准确性的指标, 在对检测框进行判定时, 通常设定一个阈值, 当 IOU 大于阈值时, 则认为检测到口罩。在 YOLOv3 中, 目标物体边界框的回归损失用均方误差 MSE 来表示, 但这种方法没有关注到预测框与真实框位置和宽高比的问题。为解决这一问题, 本文引入 CIOU 作为损失函数[8], 在原有的基础上, 考虑预测框和真实框重合率及两者中心点之间的距离和宽高比, 可有效地提高算法的收敛速度和检测的准确度, 其计算公式如下。

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b_l, b_{gt})}{\rho^2} + \alpha v \quad (7)$$

$$v = \frac{4}{\pi} \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \quad (8)$$

$$\alpha = \frac{v}{(1 - IOU) + v} \quad (9)$$

式中, α 为加权系数, b_l 、 b_{gt} 分别表示预测框和真实框的中心点, ρ^2 表示 b_l 与 b_{gt} 之间的欧几里得距离 [9], w 、 h 分别为预测框的宽和高, v 表示宽高比的相似性。

3. 数据集

3.1. 数据收集与处理

口罩佩戴识别训练数据集主要来源于网络爬虫, 部分数据由手机拍摄获得, 剔除过于模糊、口罩被

遮盖、角度较偏、距离过远等图像,共收集图像 2199 张。按照随机划分的方式将其分为训练集和验证集,数量比例为 9:1。

为了增强数据的多样性,在采集数据时尽可能多地获取了不同环境、不同角度下的样本数据,并利用数据增强的手段,对图像进行了平移、翻转、裁剪等操作,以提高检测模型的鲁棒性和准确性。

3.2. 数据集的标注

利用 Labelimg 软件对样本数据采样框及类别进行标注时,标签文件格式为 Pascal Voc,提取出数据中佩戴口罩的样本区域并标注为(mask)类,结合(EYES)类对于样本识别准确性有较大提升。对图片进行标注时,用矩形框框选眼部及佩戴口罩的部分,并保存为与图像对应的 xml 文件,如图 6 所示。



Figure 6. Diagram of dataset annotation
图 6. 数据集标注

4. 实验

4.1. 实验环境

本文实验的具体配置如表 1 所示。

Table 1. Experimental environment configuration data
表 1. 实验环境配置

配置	环境
GPU	Tesla T4*4
CPU	Xeon 8279w
Python	3.6
Keras	2.2.4
Tensor Flow-GPU	1.12.0
CUDA	9.0
OS	Ubuntu 18.04

4.2. 实验参数

实验中,各参数量的设置如表 2 所示。

在训练过程中,采用了动态调整学习率的策略,以便更好地收敛到最优解。设定初始学习率为 0.01,

Table 2. Parameters data of network training
表 2. 网络训练主要参数设置

主要参数	参数值
batch size	256
优化器	Adam_v2
epoch	3000
moduel_image_size_1	416*416
moduel_image_size_2	1080*1080

当模型连续 2 个 Epoch 的损失不再下降时, 将学习率减少为原来的一半。batchsize 为同时代入训练模型的样本集, 其大小会影响模型的优化程度和速度。综合考虑到硬件设施水平和网络的收敛速度, 最终将 batchsize 确定为 256。将训练所得的最优权重进行测试, 第一次采用默认值 moduel_image_size_1 = 416*416。在该参数下, 部分结果中能准确识别已佩戴口罩的人员, 但并不能完全识别出所有对象。可能的原因是初次采样比例过高造成的要素丢失, 考虑到设备算力较高, 故将 moduel_image_size_2 = 960*960 作为测试参数。

4.3. 评价指标与实验结果分析

4.3.1. 评价指标

本文采用三个常见的性能评价指标: 使用平均精度(AP)、平均精度均值(mAP)以及每秒检测图片的帧数(FPS)来评估改进算法的性能。AP 和 mAP 计算公式如式 10 和式 11。

$$AP = \int_0^1 P(R) \quad (10)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n \int_0^1 P(R) dR \quad (11)$$

在目标检测中, 对样本数据的分类判断有四种可能的结果, 分别是正确地识别正样本(True positives), 记为 TP; 正确识别出负样本(True negatives), 记为 TN; 原为负样本, 被误判成了正样本(False positives), 用 FP 表示; 原为正样本, 将其错判成了负样本(False negatives), 用 FN 表示, 如表 3 所示。

Table 3. Classification results of sample data

表 3. 样本数据分类结果

Hypothesized class	True class		
	P		N
Y	True Positives		False Positives
N	False Negatives		True Negatives

Precision (准确率), 是指正样本占有所有样本中预测正确的百分比; Recall (召回率), 是指被正确识别的正样本占有所有正样本的百分比[10], 计算公式如下。

$$Pecision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

当改变阈值时, Precision 和 Recall 的值也随之改变, 以召回率为横坐标, 精确率为纵坐标, 可得到 PR 曲线, 用于描述 Precision 和 Recall 之间的关系。

4.3.2. 实验结果分析

在测试集上进行测试, 得到 DS-YOLOv3 算法 PR 曲线如图 7 所示。结果表明, 两类样本均具有较高的识别准确率。对于 have_mask 目标类别, AP 值为 96.13%, 而对于 no_mask 目标类别, AP 值为 94.57%。

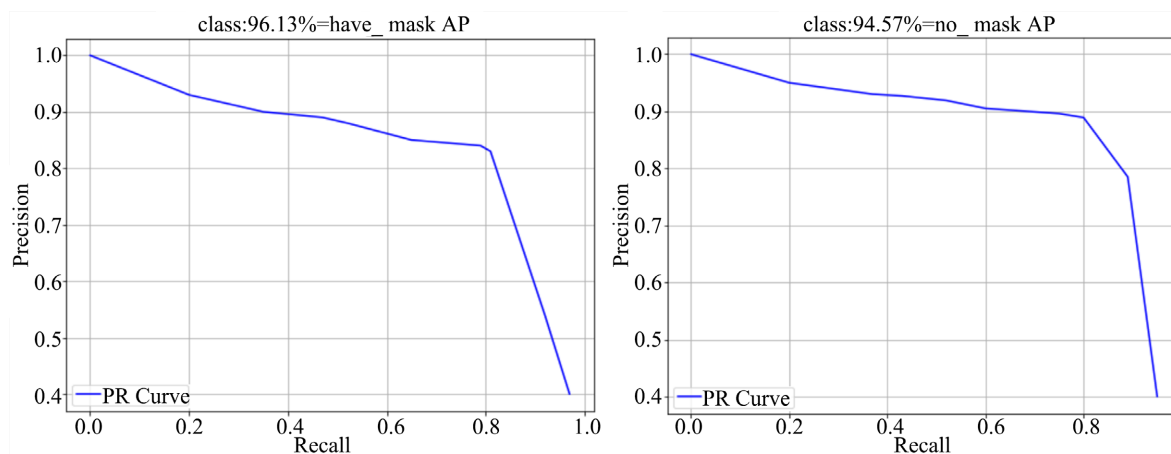


Figure 7. Improved YOLOv3 Algorithm PR curve

图 7. 改进的 YOLOv3 算法 PR 曲线

将改进的 DS-YOLOv3 算法与 YOLOv3 算法性能指标进行比较, 对比结果如表 4 所示。

Table 4. The detection speed of YOLOv3

表 4. YOLOv3 的检测速度

算法	AP/%		mAP/%	单张耗时/秒	FPS
	have mask	no mask			
YOLOv3	92.04	89.68	90.86	0.331	32.6
DS-YOLOv3	96.13	94.57	95.35	0.682	25.4

与原始 YOLOv3 算法相比, 佩戴口罩的一类 AP 值提高 4.08%, 未佩戴口罩的一类 AP 值提高了 4.89%, mAP 值提升了 4.49%。在单张图片检测速率上, 检测每张图片时间增加了 0.351 秒, 速度略慢于原 YOLOv3。在 FPS 上, DS-YOLOv3 算法 FPS 为 25.4, 与原 YOLOv3 相比, 帧率降低了 7.2, 但仍大于 25, 可以达到实时性要求。综上所述, 该改进的算法具有较高的检测精度和时效性, 且在小目标物体的检测效果上有显著的提升。

5. 结语

本文基于目标检测中口罩佩戴识别的相关研究, 针对人群环境下口罩佩戴检测中小尺寸目标识别等问题, 在 YOLOv3 算法模型的基础上进行改进与优化。首先, 对 YOLOv3 的网络结构进行改进, 通过原网络中用稠密模块代替原有的残差模块, 有效解决了对小尺寸目标识别度低的问题。接着, 对损失函数进行了改进。利用 CIUO 损失函数进行边框回归, 提高模型的定位精度。最后, 在构建的数据集上进行训练和测试, 对改进前后算法的检测效果进行比较。实验结果显示, 改进后的算法在对口罩进行识别时平均精度达到了 95.35%, 相较于原 YOLOv3 算法提升了 4.49%, 同时该算法还保持了较高的检测速

度, 能够满足实时检测的需求。此外, 该系统具有参数少, 量级小的优点, 在实际应用中可以将其嵌入移动式设备, 能够有效解放人力, 减少管理成本。

参考文献

- [1] Lowe, D.G. (2004) Distinctive Image Features from Scale Invariant Key Points. *International Journal of Computer Vision*, **60**, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [2] Hearst, M.A., Dumais, S.T., Osuna, E., *et al.* (1998) Support Vector Machines. *IEEE Intelligent Systems and Their Applications*, **13**, 18-28. <https://doi.org/10.1109/5254.708428>
- [3] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767
- [4] 张富凯, 杨峰, 李策. 基于改进 YOLOv3 的快速车辆检测方法[J]. 计算机工程与应用, 2019, 55(2): 12-20.
- [5] 孙迎春, 潘树国, 赵涛, 等. 基于优化 YOLOv3 算法的交通灯检测[J]. 光学学报, 2020, 40(12): 143-151.
- [6] 葛雯, 史正伟. 改进 YOLOv3 算法在行人识别中的应用[J]. 计算机工程与应用, 2019, 55(20): 128-133.
- [7] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2016) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 21-26 July 2017, 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>
- [8] Zheng, Z., Wang, P., Liu, W., *et al.* (2020) Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 12993-13000. <https://doi.org/10.1609/aaai.v34i07.6999>
- [9] 朱杰, 王建立, 王斌. 基于 YOLOv4-tiny 改进的轻量级口罩检测算法[J]. 液晶与显示, 2021, 36(11): 1525-1534.
- [10] Cottam, J.A., Heller, N.C., Ebsch, C.L., *et al.* (2020) Evaluation of Alignment: Precision, Recall, Weighting and Limitations. 2020 *IEEE International Conference on Big Data*, Atlanta, GA, 10-13 December 2020, 2513-2519. <https://doi.org/10.1109/BigData50022.2020.9378064>