计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(2):231−236 [doi: 10.15888/j.cnki.csa.007788]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于改进 YOLOv3 的自然场景人员口罩佩戴检测算法①

程可欣, 王玉德

(曲阜师范大学 物理工程学院, 曲阜 273165)
通讯作者: 王玉德, E-mail: wyude-01@163.com

摘　要: 针对新冠肺炎防控期间肉眼识别判断行人是否佩戴口罩效率低且存在较大风险的问题, 提出一种改进检测目标边框损失的自然场景下行人是否佩戴口罩的检测算法. 该算法对 YOLOv3 损失函数进行改进, 应用 GIoU 计算目标边界框损失, 完成自然场景下行人是否佩戴口罩的检测. 算法在开源的 WIDER FACE 数据集和 MAFA 数据集上训练, 采集自然场景图片进行测试, 行人是否佩戴口罩的 mAP (mean Average Precision) 达到了 88.4%, 取得了较高的检测准确率, 在自然场景视频检测中平均每秒传输帧数达到 38.69, 满足实时检测的要求.

关键词: 口罩检测; YOLOv3; DarkNet-53; GIoU; 损失函数

引用格式: 程可欣,王玉德.基于改进 YOLOv3 的自然场景人员口罩佩戴检测算法.计算机系统应用,2021,30(2):231–236. http://www.c-s-a.org.cn/1003-3254/7788.html

Algorithm of Mask Wearing Detection in Natural Scenes Based on Improved YOLOv3

CHENG Ke-Xin, WANG Yu-De

(College of Physics and Engineering, Qufu Normal University, Qufu 273165, China)

Abstract: It is inefficient and highly risky to identify whether pedestrians are wearing a mask or not through naked eyes during the prevention and control of the COrona VIrus Disease 2019 (COVID-19). To solve this, we devise an algorithm to detect whether the pedestrians are wearing masks in the natural scenes with the improvement in the loss function of bounding box regression. The algorithm improves the YOLOv3 loss function and uses GIoU to calculate the bounding box loss to detect whether pedestrians wear masks in natural scenes. The algorithm is trained on the open-source WIDER FACE dataset and MAFA dataset. When the natural scene pictures are collected for testing, the mAP (mean Average Precision) of whether pedestrians wear masks is as high as 88.4%. In the detection of natural scene videos, the average number of frames per second is 38.69, which meets the requirements of real-time detection.

Key words: mask wearing detection; YOLOv3; DarkNet-53; GIoU; loss function

佩戴口罩是一种隔离和遏制新型冠状病毒、预防新冠肺炎的有效方法. 为保护人民的身体健康与生命安全, 最大限度地降低和消除因疫情对生产经营造成的影响, 需要对复产复工人员规范佩戴口罩进行的监督和提醒. 依靠肉眼观察是否佩戴口罩, 不仅耗费人力物力, 而且有极大的漏检风险和近距离接触的感染风险, 因此, 需要一种基于图像处理的高精度高速度的口罩佩戴检测算法.

牛作东等提出了一种改进 RetinaFace 的自然场景口罩佩戴检测算法, 该算法基于 ResNet-152 网络, FPS

① 收稿时间: 2020-06-13; 修改时间: 2020-07-10, 2020-07-27; 采用时间: 2020-07-31; csa 在线出版时间: 2021-01-27

(每秒传输帧数, Frames Per Second) 较低, 不适用于实际的检测环境[1]. YOLOv3 算法利用回归思想, 通过 CNN 网络一次性生成目标位置边框和目标类别, 这种方式使得检测速度更快、模型泛化能力强, 同时可以减少背景错误产生, 因此本文选择该方法进行检测. 国内对该方法已有了成熟而广泛的应用, 如郑秋梅等在交通场景上使用该方法进行车辆检测[2], 王毅恒等使用该方法对农场环境下的奶牛进行检测[3], 孟本成等使用该方法对行人进行检测[4] 等. YOLOv3 算法虽然检测速度快, 但小目标漏检的风险相对更高, 鉴于上述问题, 本文提出改进目标边框损失的 YOLOv3算法对自然场景下人员是否佩戴口罩进行检测, 更好地做好人员防护.

1 基本原理

YOLOv3 算法是 Redmon 等在 2018 年提出的[5], 改进了网络结构、网络特征及损失计算 3 个部分, 在保持速度优势的前提下, 进一步提升了对小目标的检测能力和检测精度.

1.1 DarkNet-53 网络

YOLOv3 结构由骨架网络 DarkNet-53 和检测网络两部分组成, 用于特征提取和多尺度预测[6]. YOLOv3 网络结构如图 1.

DarkNet-53 网络共有 53 层卷积层, 最后一层为 1×1 卷积实现全连接, 主体网络共有 52 个卷积. 52 个卷积层中, 第一层由一个 32 个 3×3 卷积核组成的过滤器进行卷积, 后面的卷积层是由 5 组重复的残差单元 (resblock body) 构成的, 这 5 组残差单元每个单元由一个单独的卷积层与一组重复执行的卷积层构成, 重复执行的卷积层分别重复 1、2、8、8、4 次; 在每个重复执行的卷积层中, 先执行 1×1 的卷积操作, 再执行 3×3 的卷积操作, 过滤器数量先减半, 再恢复, 共 1+(1+1×2)+ (1+2×2)+ (1+8×2)+ (1+8×2)+ (1+4×2)=52 层.

YOLOv3 模型的输出为 3 个不同尺度的特征层, 分别位于 DarkNet-53 网络的中间层、中下层和底层, 用于检测不同大小的物体. 如图 1 所示, 对 3 个特征层进行 5 组卷积处理, 可输出该特征层对应的预测结果.

1.2 网络性能分析

DarkNet-53 特征提取网络通过大量的 3×3 和 1×1 卷积层构成, 该网络在 ImageNet 数据集下测试, 网络性能比 ResNet 网络更好[3], 结果如表 1.

通过表 1 我们可以看到, 在图像分类的准确率以及检测速度等方面, DarkNet-53 网络与其他 3 种网络模型相比, 表现更加优越. DarkNet-53 网络在满足检测实时性的同时比 DarkNet-19 具有更高的精度, 并且在网络性能相差无几的情况下, 网络的速度约是 ResNet-152 网络的 2 倍.

2 损失函数改进

2.1 GIoU 损失函数

在原始 YOLOv3 算法中, 使用均方误差作为目标定位损失函数来进行目标框的回归, 均方误差函数对尺度较为敏感, 并且无法反应不同质量的预测结果, 故大量使用 YOLOv3 算法的工作中, 常使用预测框和真实目标框的 IoU 值来衡量两个边界框之间的相似性, 虽然改善了这两个问题, 却也带来了新问题. 首先, 当预测框和真实框之间没有重合时, IoU 的值为 0, 导致优化损失函数时梯度也为 0, 意味着无法优化. 其次, 即使预测框和真实框之间相重合且具有相同的 IoU 值时, 检测的效果也具有较大差异.

Rezatofighi 等于 CVPR2019 上提出了 GIoU (Generalized IoU, 广义 IoU) 目标边界框优化方法, GIoU 针对 IoU 无法反应不重叠的两个框之间距离和重叠框对齐方式的问题进行了优化, 图 2 中 3 幅图的 IoU 均为 0.33, GIoU 的值分别是 0.33, 0.24 和−0.1, 这表明如果两个边界框重叠和对齐得越好, 那么得到的 GIoU 值就会越高[7].

图 2 中黑色框为真实框 A, 灰色为预测框 B, 虚线框为最小可包含 A、B 的框 C. 假设有框 A 和 B, 总可以找到一个最小的封闭矩形 C, 将 A 和 B 包含在内, 然后计算 C 中除了 A 和 B 外的部分的面积占 C 总面积的比值, 再用 A 与 B 的 IoU 减去这个比值, IoU 计算公式和 GIoU 计算公式如式 (1)、式 (2) 所示.

$$R_{\mathrm{IoU}} = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$R_{\mathrm{GIoU}} = R_{\mathrm{IoU}} - \frac{|C - (A \cup B)|}{|C|} \tag{2}$$

$$L_{\mathrm{GIoU}} = 1 - R_{\mathrm{GIoU}} \tag{3}$$

GIoU 与 IoU 类似, 可以作为一种距离度量, 损失可以由式 (3) 计算. GIoU 对物体的大小并不敏感, 其值总是小于等于 IoU, 是 IoU 的下界. 在两个形状完全重合时, GIoU 和 IoU 大小均为 1. GIoU 引入了包含框 A 和 B 两个形状的 C, 解决了 IoU 不能反映重叠方式, 无法优化 IoU 为 0 的预测框的问题.

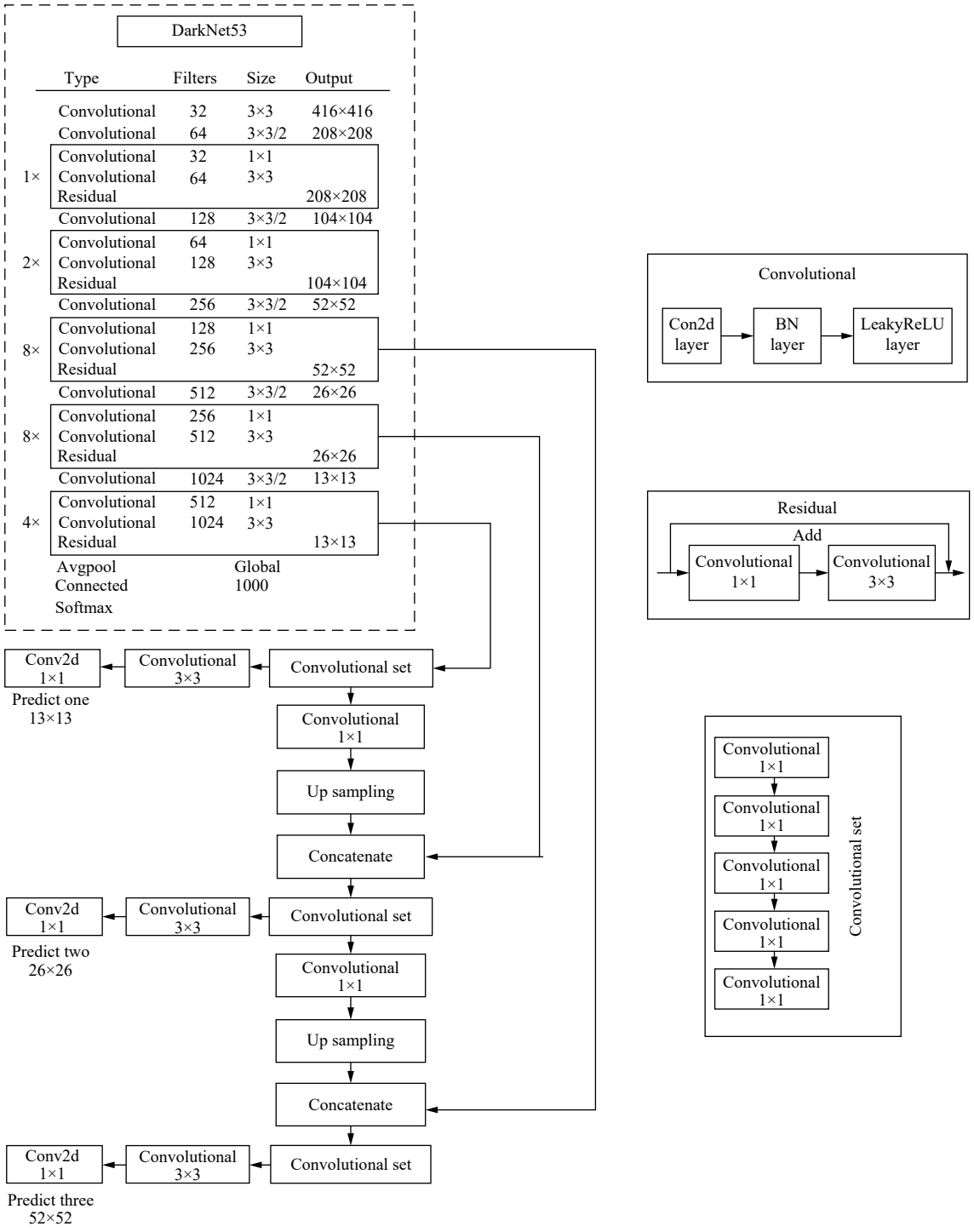

图 1 YOLOv3 网络结构 (含 DarkNet-53)

注: 图 1 中所示 DarkNet-53 网络中的最后 3 层在 YOLOv3 网络结构中不存在

表 1 4 种网络框架性能对比

Backbone	Top-1 accuracy	Top-5 accuracy	BFLOPS	FPS
Darknet-19	74.1	91.8	1246	171
ResNet-101	77.1	93.7	1039	53
ResNet-152	77.6	93.8	1090	37
DarkNet-53	77.2	93.8	1457	78

注: Top-1 accuracy是指ImageNet排名第一的类别与实际结果相符的准确率; Top-5 accuracy是指排名前五的类别包含实际结果的准确率; BFLOPS为每秒十亿次浮点数运算次数.

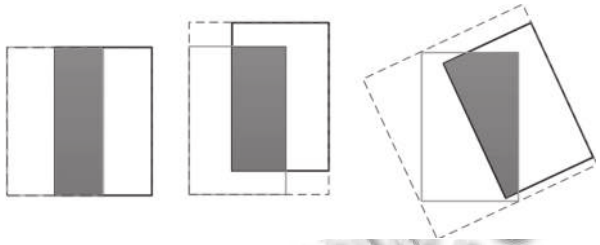

图 2 IoU 均为 0.33 时 3 种不同的重叠情况

2.2 Loss 损失计算

本文将L_{GIoU}损失函数应用于 YOLOv3 目标检测算法中, 以L_{GIoU}直接作为边界框回归损失函数代替原来的均方差和损失函数.

损失函数的公式包含目标定位损失、目标置信度损失和目标类别损失 3 个部分, 分别对应式 (4) 中的第一项、第二、三项和第四项.

$$\begin{aligned} Loss =& \lambda_{\mathrm{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\mathrm{obj}} L_{\mathrm{GIoU}} \\ & - \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\mathrm{obj}} \left[\hat{C}_i^j \log\left(C_i^j\right) + \left(1-\hat{C}_i^j\right) \log\left(1-C_i^j\right) \right] \\ & - \lambda_{\mathrm{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\mathrm{noobj}} \left[\hat{C}_i^j \log\left(C_i^j\right) + \left(1-\hat{C}_i^j\right) \log\left(1-C_i^j\right) \right] \\ & - \sum_{i=0}^{S^2} I_{ij}^{\mathrm{obj}} \sum_{c \in classes} \left(\left[\hat{P}_i^j \log\left(P_i^j\right) + \left(1-\hat{P}_i^j\right) \log\left(1-P_i^j\right) \right] \right. \end{aligned} \tag{4}$$

3 算法流程

该算法流程主要分为特征提取和多尺度预测两部分, 图片经过 DarkNet-53 网络, 生成 3 种尺度的特征图, 每种尺度的特征图划分为大小不同的网格, 每个网格预测 3 个先验框, 经过非极大抑制等, 得到预测框, 在训练过程中, 还会进行损失函数计算, 更新权重.

4 实验与结果分析

实验用计算机配置为 Intel Corei5-5200 2.20 GHz CPU、Tesla V100-SXM2 GPU, 显存 16 GB. 软件环境为 Linux 操作系统、PyCharm2019.3.3、PyTorch 框架.

文中使用了 AIZOO 团队公开的人脸口罩佩戴数据集, 该数据集中的图片来源于 WIDER FACE 数据集[8]和中科院信工所葛仕明老师开源的 MAFA 数据集[9], 图片示例如图 3, 对应的标注数据如表 2. 训练样本与测试样本按 0.7: 0.3 划分, 训练集共 5566 张图片, 来自 MAFA 的图片 2873 张 (基本都是戴口罩的图片)、WIDER Face 图片 2693 张 (基本都是不戴口罩的图片). 验证集共 2385 张图片, 取自 MAFA 1188 张、WIDER Face 1197 张.

图 3 数据集示例图片

表 2 示例图片标注数据

Object	Bounding-box
face_mask	(169, 42, 226, 113, 0)
face_mask	(66, 75, 123, 149, 0)

采用迁移学习的方式, 采用了 ImageNet 预训练好的模型参数, 通过初始化模型前 47 层卷积层参数、微调末端参数的方式对模型进行训练, 将检测类别按是否佩戴口罩调整为 2 种、初始学习率设置为 0.01、batch-size 为 16.

从图 4–图 6 可以看出, 使用 GIoU 作为目标边界框损失函数总是小于 IoU, 该结果符合 GIoU 的特点, 平均检测精度 (mAP) 上升速度更快且有小幅度的提高. 在使用 GIoU 的条件下, 训练迭代次数到达 50 次后, mAP 曲线渐趋平缓, 最后达到 88.4% 左右不再增加, 而 GIoU 一直平稳下降至 0.93, 目标定位损失下降至 0.315, 目标分类损失下降至 0.0361. 验证集数量少后期易出现过拟合现象, 故迭代次数达到 100 次时停止训练.

GIoU 作为目标边界框损失函数的训练模型, 多目标检测的平均精度达到 88.4%, 佩戴口罩的类别达到 96.5%, 检测结果如表 3, 与 IoU 相比均有提高.

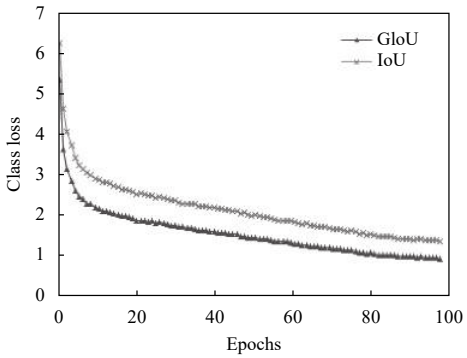

图 4 GIoU 与 IoU 训练过程曲线

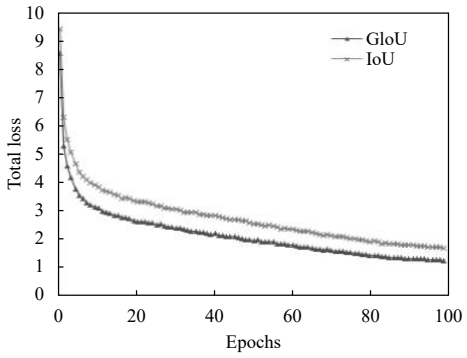

图 5 使用 GIoU 与 IoU 的 YOLOv3 总损失函数训练过程曲线

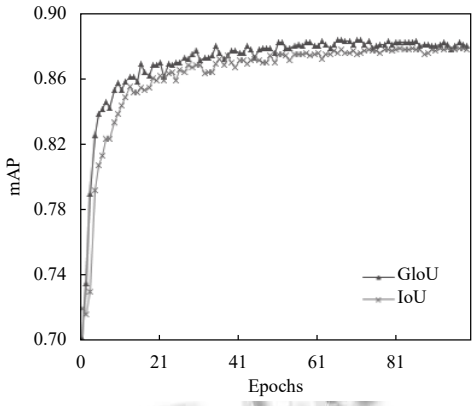

图 6 GIoU-YOLOv3 与 IoU-YOLOv3 mAP 训练过程曲线

注: epoch 为 0 时, GIoU 的 mAP 为 0.465, IoU 的 mAP 为 0.312.

表 3 不同类别使用 GIoU 和 IoU 的 mAP 对比

class	GIoU-mAP@0.5	IoU-mAP@0.5
all	0.884	0.879
face	0.803	0.796
face_mask	0.965	0.963

图 7(a) 为 GIoU-YOLOv3 测试结果, 图 7(b) 组图片为 IoU-YOLOv3 测试结果. 从测试结果可见 GIoU-YOLOv3 算法针对小目标的漏检率有明显降低.

(a) GIoU-YOLOv3 检测结果

(b) IoU-YOLOv3 检测结果

图 7 未戴口罩和佩戴口罩的检测图像

图 8 为从网络随机爬取的 512×320 大小 416 帧的视频的测试结果, 共用时 10.751 s, 平均每帧用时 0.026 s, FPS 达到 38.69, 满足实时检测的要求.

5 结论

论文提出改进检测目标边框损失的自然场景人员口罩佩戴检测算法, 在一定程度上减小了漏检率, mAP 也有一定的提高. 该方法平均每秒检测约 38 张图片, 可以实现实时检测. 同时, 对佩戴口罩类别的检测准确率可达 96.5%, 行人是否佩戴口罩的 mAP 达到了 88.4%. 在多目标检测上有较好的表现, 检测速度更快、成本更低、准确率更高.

图 8 视频测试结果部分截取

参考文献

1 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. 计算机工程与应用, 2020, 56(12): 1–7. [doi: 10.3778/j.issn.1002-8331.2002-0402]

2 郑秋梅, 王璐璐, 王风华. 基于改进卷积神经网络的交通场景小目标检测. 计算机工程, 2020, 46(6): 26–33.

3 王毅恒, 许德章. 基于 YOLOv3 算法的农场环境下奶牛目标识别. 广东石油化工学院学报, 2019, 29(4): 31–35. [doi: 10.3969/j.issn.2095-2562.2019.04.007]

4 孟本成. 基于 YOLOV3 算法的行人检测方法. 电视技术, 2019, 43(9): 6–9, 46.

5 范丽, 苏兵, 王洪元. 基于 YOLOv3 模型的实时行人检测改进算法. 山西大学学报 (自然科学版), 2019, 42(4): 709–717.

6 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv preprint arXiv: 1804.02767, 2018.

7 Rezatofighi H, Tsoi N, Gwak J, *et al*. Generalized intersection over union: A metric and a loss for bounding box regression. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 1–9.

8 Yang S, Luo P, Loy CC, *et al*. WIDER FACE: A face detection benchmark. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. 2016. 5525–5533. [doi: 10.1109/CVPR.2016.596]

9 Ge SM, Li J, Ye QT, *et al*. Detecting masked faces in the wild with LLE-CNNs. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 426–434. [doi: 10.1109/CVPR.2017.53]