

基于 YOLOv7 的人体关联实时吸烟目标检测方法

孙冰¹, 李好¹, 黄鑫凯¹, 任长宁², 邹启杰¹

(1.大连大学, 辽宁 大连 116622;

2.大连东软信息学院, 辽宁 大连 116023)

✉ 15762467351@163.com; lihao@s.dlu.edu.cn; 1432796737@qq.com;

renchangning@neusoft.edu.cn; 55440414@qq.com



摘要: 聚焦以智能安防系统为基础的吸烟行为自动检测问题, 解决复杂背景影响下吸烟检测的误检漏报缺陷, 提出了基于 YOLOv7 的人体关联实时吸烟目标检测方法。通过定位香烟和场地内人员, 利用目标关联的方式解决吸烟行为的实时检测问题。实验测得该方法的 mAP 值为 90%, 帧率为 130.1 FPS, 结果说明基于 YOLOv7 的人体关联实时吸烟目标检测方法可适用于高精度的实时目标检测系统。

关键词: 吸烟检测; 目标关联; YOLOv7; 目标检测

中图分类号: TP391.4 **文献标志码:** A

Real-Time Smoking Target Detection Method for Human Body Based on YOLOv7

SUN Bing¹, LI Hao¹, HUANG Xinkai¹, REN Changning², ZOU Qijie¹

(1.Dalian University, Dalian 116622, China;

2.Dalian Neusoft University of Information, Dalian 116023, China)

✉ 15762467351@163.com; lihao@s.dlu.edu.cn; 1432796737@qq.com; renchangning@neusoft.edu.cn; 55440414@qq.com

Abstract: Focusing on the problem of automatic detection of smoking behavior based on intelligent security systems, this paper proposes a real-time smoking target detection method based on YOLOv7 to solve the problems of false detection and missed reporting defects of smoking detection under complex circumstances. By locating cigarettes and people in the venue, real-time detection of smoking behavior can be solved through target association. The experimental results show that the mAP value of the proposed method is 90% and the frame rate is 130.1FPS, indicating that the real-time smoking target detection method for human body based on YOLOv7 is more suitable for high-precision real-time target detection systems.

Key words: smoking detection; target association; YOLOv7; target detection

0 引言 (Introduction)

多数作业环境明令禁止吸烟, 诸如物流仓储、化工工厂、供电单位等, 虽然可以通过中央监控系统实时监控各种作业场景, 但是主要依赖人工监查, 存在漏报情况, 不但不能节省人力成本, 而且会对企业和工作人员的生命财产安全造成威胁。随着计算机视觉技术的发展, 目标检测技术日渐成熟, 智能检测得到广泛的研究和应用, 由此本文提出一项基于 YOLOv7^[1] 的

人体关联实时吸烟目标检测方法。

当下不乏实时吸烟行为检测的相关研究, 但是对于实际应用仍有可提升、待完善的地方, 具体如下: 第一, 对于复杂场景小目标(烟)的检测精度尚有提升空间^[2-3]; 第二, 仅对香烟进行检测, 对香烟相似物存在漏检误报的情况^[4-5]; 第三, 部分研究虽然通过增加姿态检测、特征检测、面部检测等方式进行优化, 但是复杂的算法导致模型检测速度降低^[3,6]。基于以上问题,

本文提出改进方法。首先,通过数据增强提升算法小目标检测能力,从而提高香烟检测的准确率,并解决过拟合问题;其次,基于当前比较先进的 YOLOv7 算法同时检测人和烟,通过设置人和烟的目标关联阈值,增加检测条件的限制,降低误检率;最后,经过对比实验和消融实验验证方法的有效性和检测方法性能的提升程度,并用工作现场视频进行算法的验证。

1 目标检测(Object detection)

现有典型的目标检测方法可以分为单阶段目标检测(YOLO、SSD、RetinaNet 等)和多阶段目标检测(R-CNN、Fast R-CNN、Mask R-CNN 等)(图 1)^[7-8]。

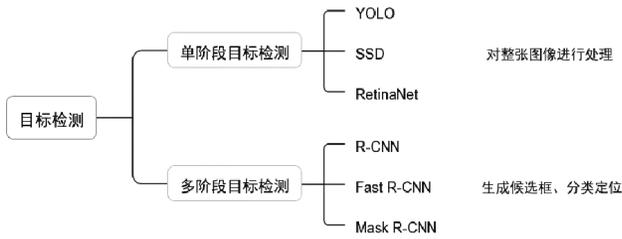


图 1 目标检测方法分类

Fig. 1 Classification of target detection methods

1.1 多阶段目标检测算法

多阶段目标检测算法通常包含两个阶段,即生成候选框和分类定位。在第一个阶段,算法使用候选框生成器生成多个候选框,每个候选框都与某个物体相对应。在第二阶段,候选框中的特征图将被送入一个分类器和回归器中,以进一步提取物体的位置和类别信息。常见的多阶段目标检测算法有 R-CNN 及其变种 Fast R-CNN 和 Mask R-CNN 等^[8]。多阶段目标检测算法通过使用候选框定位和识别物体,具有更高的准确率和定位精度。但是,与单阶段目标检测算法相比,多阶段目标检测的计算复杂度更高,计算速度较慢,不适用于对实时检测要求高的作业环境。

1.2 单阶段目标检测算法

单阶段目标检测算法通过处理整张图像预测物体的位置和类别。这种算法速度较快,适合实时应用场景。常见的单阶段目标检测算法有(Single Shot MultiBox Detector, SSD)、RetinaNet、YOLO 等^[3]。这些算法通常将物体位置和类别信息结合起来作为网络输出,使用较少的候选框定位物体。SSD 是以单个 CNN 为基础的目标检测算法,可以同时检测不同大小和比例的物体。RetinaNet 将分类和回归任务分别交给两个并行的子网络来解决分类精度和定位精度不平衡的问题,同时引入损失函数来调整难易样本的权重,故此能够获得更好的目标检测性能。

YOLO 是端到端的单阶段目标检测算法,具有快速和高准确率的特点。YOLO 将输入图像划分成网格,并对各个网格进行分类、定位,生成每个物体的边界锚框和概率。YOLO 采用卷积神经网络处理整个图像,以便于实时应用。2020 年发布

的 YOLOv5 具有更小的模型体积和更快的推理速度,同时在目标检测精度方面也有一定的提升。2022 年发表的 YOLOv7 算法集成以往 YOLO 系列的优点,并不断推陈出新,尽管在准确率和运算速度上较以往 YOLO 系列都有所提升,但存在对小目标检测精度不够的问题。在本实验测试中发现,YOLOv7 比以往 YOLO 系列更容易出现过拟合现象。相比多阶段目标检测算法,单阶段目标检测算法的处理速度更快、计算复杂度更低,适合实时应用场景。但是,单阶段目标检测算法在一些复杂场景下的准确率可能不如多阶段目标检测算法,因此本文在 YOLOv7 算法的基础上做了一些调整,以适应极小目标(香烟)的检测。

2 基于 YOLOv7 的吸烟检测方法(A smoking detection method based on YOLOv7)

2.1 YOLOv7 模型

YOLOv7 算法主要由输入端(Input)、主干网络(Backbone)和头部网络(Head)构成,采用扩展高效长程注意力网络(E-ELAN)、基于级联模型的模型缩放、卷积重参数化等策略,在检测效率与精度之间取得了非常好的平衡^[9-10]。输入端由数据增强、自适应锚框计算和自适应图片放缩构成,将输入图像缩放至固定的尺寸,实现数据增强。主干网络主要由多个 CBS、ELAN、MPConv 等模块组成,用于图像特征的提取^[11]。预测端用于预测,采用聚合特征金字塔网络结构,将底层信息通过自底向上的路径传递到高层,实现差别层次特征的融合,借助 REPcon 结构对不同尺度的特性进行通道数调整。

如图 2 所示,YOLOv7 首先对输入的图片进行预处理,处理为 640×640 像素大小的 RGB 图片,其次输入主干网络,通过主干网络的三层高效长程注意力网络进行输出,并继续在头部网络层输出三个不同大小的特征图,经过重参数化和卷积进行图像分类、图像前后背景分类以及边框预测,输出最后的结果。

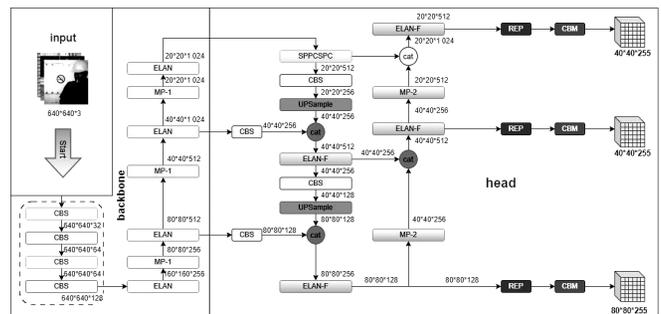


图 2 YOLOv7 整体结构图

Fig. 2 Overall structure of YOLOv7

2.2 数据集增强

本实验通过互联网收集了来自安防场景、公共场所、危险场所等现实场景的共计 3 628 张吸烟行为的图像,并通过旋转方式进行数据集增强,将抽烟行为数据集扩充到 14 512 张,结合 14 508 张人体数据集,共计 29 020 张图像,并按 70%、15% 和 15% 的占比分成训练集、测试集和验证集,数据集划分见表 1。

表1 数据集划分
Tab.1 Dataset division

参数	图像数量/张
总数据集	29 020
训练集	20 398
测试集	4 382
验证集	4 240

对数据集进行如图3所示的标签标注,采用线上标注工具makesense进行目标标注,分别标记为“person”“smoke”,获得以文本文件形式存储的标注结果。



图3 目标标注示例

Fig. 3 Example of target annotation

2.3 目标关联

吸烟行为的出现一般需要同时具备两个基本条件——人和烟,故本研究通过YOLOv7模型同时定位人和烟,并进行目标关联,在算法中计算人和烟的中心点距离,当二者距离小于设定阈值后,发出吸烟告警,实现吸烟检测。

场地人员检测锚框 $(x_{2p}-x_{1p}) \times (y_{2p}-y_{1p}) < 50\ 000$ 时,人的坐标 (x_p, y_p) 如下:

$$\begin{cases} x_p = \frac{|x_{2p}-x_{1p}|}{2} + x_{1p} \\ y_p = \frac{|y_{2p}-y_{1p}|}{2} + y_{1p} \end{cases} \quad (1)$$

香烟检测锚框 $(x_{2s}-x_{1s}) \times (y_{2s}-y_{1s}) < 1\ 000$ 时,香烟坐标 (x_s, y_s) 如下:

$$\begin{cases} x_s = \frac{|x_{2s}-x_{1s}|}{2} + x_{1s} \\ y_s = \frac{|y_{2s}-y_{1s}|}{2} + y_{1s} \end{cases} \quad (2)$$

同时检测到人和烟,判断是否出现吸烟行为,当 $\sqrt{(y_s-y_p)^2+(x_s-x_p)^2} < 200$ 时,表明人与烟存在关联性,经推理满足时空关系,打出危险警告标签“Is smoking”,表示该人员可能存在吸烟行为,系统发出告警。

3 实验与分析(Experiment and analysis)

3.1 实验环境

实验环境使用Windows 10操作系统、NVIDIA GeForce RTX 3080显卡进行运算。具体实验配置见表2。网络模型训练阶段,训练迭代次数设置为150次,Batch size设置为8,Img size设置为[640,640],设置好限定条件后进行训练。

表2 实验环境配置

Tab.2 Experimental environment configuration

参数	配置
CPU	Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 GHz
GPU	NVIDIA GeForce 3080 * 2
系统环境	Windows 10
语言	Python3.8

3.2 性能分析

3.2.1 分析指标

本实验将查准率(Precision)、召回率(Recall)、平均精度均值(mAP)作为评价指标,生成了三者随迭代次数变化的折线图,并生成Precision-recall曲线图像作为分析评价的辅助参考。

精确率也称查准率,该指标用来判断模型检测是否准确,是在识别出的物体中正确的正向预测所占的比率。在公式(3)中,TP表示真的正样本,FP表示假的正样本。

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

召回率也称查全率,是指正确识别出的物体占总物体数的比率,该指标用来判断模型检测是否全面,在公式(4)中,FN表示假的负样本。

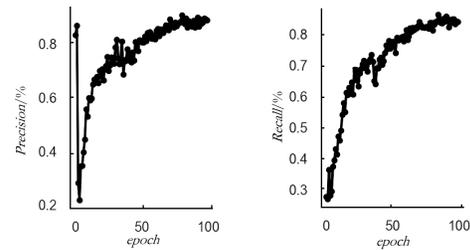
$$Recall = \frac{TP}{TP+FN} \quad (4)$$

ΔP 代表Precision-recall曲线下方面积,分类器越好,AP值越高。如公式(5)所示,目标检测算法中最重要的指标之一mAP代表多个类别AP的平均值,大小处于[0,1]内,越接近1,表明该目标检测模型在给定的数据集上的检测效果越好。

$$mAP = \frac{1}{N} \sum_{k=i}^N P(k) \Delta R(k) \quad (5)$$

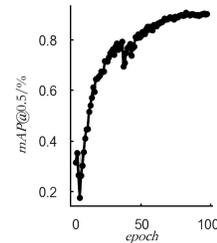
3.2.2 结果分析

将实验数据进行可视化处理,绘制查准率、召回率和平均精度均值的折线图,如图4(a)、图4(b)所示,查准率和召回率均达到90%且图4(c)中的平均精度均值达到90%以上,说明模型在检测精度上表现出色。



(a)查准率折线图

(b)召回率折线图



(c)平均精度均值折线图

图4 模型测试结果

Fig. 4 Model test results

通过不断改变识别阈值,使得系统能够依次识别前 N 张图片,阈值变化的同时会导致 $Recall$ 与 $Precision$ 值的变化,从而得到 $Precision-recall$ 曲线^[12]。如图 5 所示,本模型测试所得曲线下方的面积较大,并且在 $Recall$ 值增长的同时, $Precision$ 的值能保持在一个很高的水平,在 $Precision$ 和 $Recall$ 之间实现了较好的平衡^[12]。

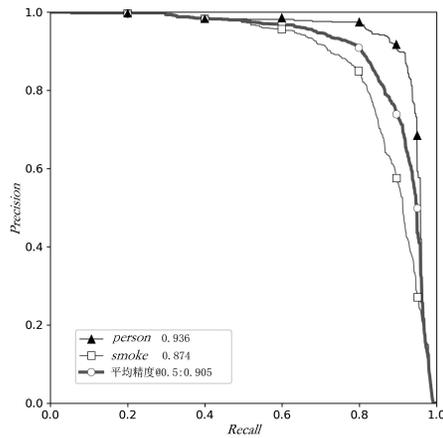


图 5 Precision-recall 曲线图
Fig. 5 Precision-recall curve

为了验证本模型性能,将原 YOLOv7 作为基线模型,通过对 YOLOv7 模型定位内容的修改对比进行如表 3 所示的消融实验,分别取实验中最优的模型在制作好的测试集进行验证。在基于 YOLOv7 分别进行仅定位香烟和人体关联检测两种方法中,虽然人体关联的方法牺牲了一定的训练时长,但是将误检率降低到了 0.001%,较仅定位香烟的方法有 6% 的性能提升,基本解决了香烟检测的误判问题。

表 3 不同改进方法的检测结果

Tab.3 Test results of different improvement methods

改进方法	训练时间/h	误判率/%
仅定位香烟	24.8	6.687 50
人体关联	30.0	0.001 25

3.2.3 性能对比

为了证明本文方法对比于其他方法在检测准确率、模型精简度和检测速度上具有一定优势,选取 RetinaNet 和 YOLOv5 两种算法在同一数据集上进行测试,以 mAP 、训练时间和帧率作为评价指标进行 3 种算法的性能对比判断。

由表 4 可知,与其他两种算法相比,本文所提方法的 mAP 值较 RetinaNet 算法提升了 94.6%,较 YOLOv5 算法提升了 31.5%,训练时间较 RetinaNet 算法和 YOLOv5 算法分别缩短了 51.2%、33.9%,帧率分别提高了 210.5%、95.1%,性能提升效果显著。

表 4 不同算法平均测试评估指标

Tab.4 Average test evaluation indicators for different algorithms

算法	mAP /%	训练时间/h	帧率/FPS
RetinaNet	46.3	50.8	41.9
YOLOv5	68.5	37.5	66.7
本文算法	90.1	24.8	130.1

综合对比实验结果,基于 YOLOv7 的人体关联实时吸烟目标检测方法较 RetinaNet 和 YOLOv5 算法大大提升了吸烟行为的检测准确率,降低了训练时间,提高了检测帧率。

3.3 检测结果可视化

使用现场拍摄的安防场景吸烟行为和人员作业行为的视频做测试,在低像素、远距离的情况下,图 6(a)中安防器械场地准确定位了场地作业人员。图 6(b)中安防办公场地准确定位了人和烟,判断满足距离关系后,以人和烟中心为对角线框出吸烟行为,实现抽烟行为的准确定位,并对吸烟行为发出告警。



(a) 安防器械场地 (b) 安防办公场地

图 6 安防场地吸烟行为和人员作业行为检测

Fig. 6 Detection of smoking and work behavior in security venue

分别在不同光线、不同角度、不同像素情况下进行视频检测验证,结果如图 7 所示,图 7(a)有照明高像素正面视角和图 7(b)无照明低像素侧面视角都成功定位了人和烟,并判断存在抽烟行为,说明方法具有一定的可靠性和可行性。



(a) 有照明高像素正面视角 (b) 无照明低像素侧面视角

图 7 不同光线、不同角度、不同像素情况下的视频检测

Fig. 7 Video detection under different lighting conditions, angles, and pixels

4 结论 (Conclusion)

本文以复杂背景下小目标检测为基础,针对安全要求较高的环境下的吸烟行为检测进行研究,提出了基于 YOLOv7 的人体关联的实时吸烟目标检测方法。通过同时关联定位香烟和场地内人员,提升吸烟行为检测的准确度。通过以相同数据集为基础进行验证发现,人体关联检测较仅定位香烟检测,虽然在训练时间上有一定的牺牲,但是显著提升了检测准确度。此外在同一数据集验证发现, YOLOv7 算法与 RetinaNet、YOLOv5 算法相比,大大缩短了检测时间,并且检测准确度和帧率都显著提升,说明本方法能够有效提升检测率,降低漏检率,具备实时性和高效性。

(下转第 78 页)