

基于YOLOv5改进模型的办公室吸烟行为检测

魏袁慧, 方睿, 石兴, 刘金智

(成都信息工程大学计算机学院, 四川成都610225)

摘要: 为解决目前吸烟行为检测中小目标检测精度低、易误检的问题, 提出一种改进YOLOv5的识别模型YOLOv5s+。该模型将YOLOv5的主干网络与BoTNet相结合, 以提高模型的特征提取能力, 使其能够检测更小的目标物体; 同时改进特征融合部分, 在网络模型的颈部应用加权双向特征金字塔BiFPN, 以高效融合浅层位置信息与深层高级语义信息, 有效提高了检测精度。将网络公开数据集与自制数据集整合成办公室吸烟实验数据集, 在该数据集上比较YOLOv5s+模型与原YOLOv5模型的检测性能。实验结果表明, 改进模型YOLOv5s+的平均精度均值(mAP)为81.8%, 精度为82.8%, 召回率为83.9%, 相较原模型分别提高了5.4%、4.1%和6.4%, 较好地实现了办公室吸烟行为检测。

关键词: 深度学习; YOLOv5; 吸烟检测; 特征融合

DOI: 10.11907/tj.dk.231791

中图分类号: TP391.4

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2024)009-0170-06



Office Smoking Behavior Detection Based on Improved YOLOv5 Model

WEI Yuanhui, FANG Rui, SHI Xing, LIU Jinzhi

(College of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: To solve the problems of low accuracy and easy false detection of small targets in current smoking behavior detection, an improved YOLOv5 recognition model YOLOv5s+ is proposed. This model combines the backbone network of YOLOv5 with BoTNet to improve the feature extraction ability of the model, enabling it to detect smaller target objects; At the same time, the feature fusion part is improved by applying a weighted bidirectional feature pyramid BiFPN in the neck of the network model to efficiently fuse shallow position information and deep high-level semantic information, effectively improving detection accuracy. Integrate publicly available online datasets and self-made datasets into an office smoking experimental dataset, and compare the detection performance of the YOLOv5s+ model with the original YOLOv5 model on this dataset. The experimental results show that the average accuracy (mAP) of the improved model YOLOv5s+ is 81.8%, with an accuracy of 82.8% and a recall rate of 83.9%. Compared with the original model, it has improved by 5.4%, 4.1%, and 6.4%, respectively, and has achieved good detection of office smoking behavior.

Key Words: deep learning; YOLOv5; smoking detection; feature fusion

0 引言

吸烟产生的烟雾中含有大量有害化学物质, 如尼古丁、一氧化碳、苯并芘等, 不仅对吸烟者本人身体造成伤害, 而且会使周围的非吸烟者被迫吸入二手烟, 尤其是在

封闭的办公室环境中, 二手烟对身体的伤害更为严重^[1]。卫生部《公共场所卫生管理条例实施细则》第十八条明确规定室内公共场所禁止吸烟。一些室内公共场所尝试使用烟雾传感器检测吸烟产生的烟雾, 从而对吸烟行为进行警报。然而, 烟雾浓度受场所通风性和空间大小的影响, 检测效果无法保证。随着计算机视觉和深度学习技术的

收稿日期: 2023-07-18

扫描二维码阅读全文:

基金项目: 国家重点研发计划项目(2020YF0608000)

作者简介: 魏袁慧(1998-), 女, 成都信息工程大学计算机学院硕士研究生, 研究方向为数字图像处理与人工智能; 方睿(1974-), 男, 成都信息工程大学计算机学院教授、硕士生导师, 研究方向为基于网络的计算机应用; 石兴(1997-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为数字图像处理与人工智能; 刘金智(1999-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为数字图像处理与人工智能。本文通讯作者: 方睿。



发展,吸烟行为检测的准确度逐步提高。然而,办公室场景光照分布差异较大,且小场景中香烟的特征并不明显,因此更易出现烟支目标与其他目标误判的情况,现有检测方法仍有提升空间。

1 相关研究

根据现有研究成果,吸烟行为检测方法主要包括烟雾检测、手势动作检测和香烟目标检测 3 类。例如,汪祖云等^[2]利用香烟烟雾在 HSV(Hue, Saturation, Value)空间下特有的颜色特征实现对疑似烟雾区域的定位,通过分析烟雾的面积变化速率特征和质心相对角度正弦变化特征,利用支持向量机(Support Vector Machine, SVM)分类器实现对吸烟行为的判定,然而香烟烟雾形态复杂、透明度高,关键检测因素采集难度大,导致检测精确度低;邢予权^[3]通过肤色分割技术对不同颜色空间进行转换,通过统计肤色在颜色空间中的分布以及对光照等的敏感程度区分吸烟的目标手势;刘婧^[4]通过分析吸烟动作特性,提出基于空间残差混合图卷积的吸烟动作识别模型与周期性检测的多人吸烟动作识别模型;赵鉴福^[5]划分手部目标区域,运用不同图像增强方法分析并提取烟支特征,提出一种结合人体运动规律与傅里叶描述子的吸烟行为检测方法。由于吸烟时的手势动作与喝水、吃饭等动作相似,靠手势动作判断吸烟的误检率较高。办公室为禁烟场所,不应出现烟支,因此一旦检测出烟支目标便可判定为存在吸烟行为。基于香烟目标的吸烟行为检测不需要复杂的烟雾形态和手势判断,只需要检测出图像中的烟支目标即可^[6]。例如,马研凯^[7]设计了一种轻量化烟支检测模型 YOLO-cigarette,通过识别室外公共场所监控摄像头所拍摄图像中的烟支目标检测吸烟行为;孙召龙等^[8]设计基于 YOLOv5 的烟支目标检测算法,对油田作业现场的吸烟行为进行检测。

由于香烟目标较小、室内背景环境复杂,以上方法

仍存在检测精度不够高、易误检的问题。为此,本文提出一种改进的 YOLOv5 算法 YOLOv5s+,用于办公室吸烟行为检测。该算法将多头自注意力机制(Multi-headed Self-attention, MHSA)加入主干网络构造 BoTNet 中,以提高模型的特征提取能力,使之能够检测更小的目标物体^[9];将颈部原有的特征融合网络改为加权双向特征金字塔 BiFPN,以高效融合浅层位置信息与深层高级语义信息^[10]。将香烟作为检测目标,待检测数据经过网络特征提取、特征图多尺度融合后可判断出是否存在吸烟行为。

2 YOLOv5 算法

YOLOv5 是一种目标检测算法,由 Ultralytics 公司开发,于 2020 年 6 月发布^[11-12]。与先前的版本相比,YOLOv5 在许多方面进行了改进,其采用了新的网络架构,主干网络采用的是跨阶段局部网络结构(Cross Stage Partial Network, CSPNet)^[13],这是一种深度可分离卷积网络(Depth-wise Separable Convolution, DSC)^[14],包含多个跨阶段部分的特征提取模块,每个跨阶段部分包含一个卷积层和一个残差块,用于提高特征表达能力和计算效率;颈部使用 YOLOv4 的 PAN(Path Aggregation Network)+FPN(Feature Pyramid Network)特征金字塔结构^[15],可以将来自不同特征层的特征信息结合到一起,以提高目标检测的准确性和鲁棒性。此外,YOLOv5 还可将来自主干网络的特征图分别送入不同的检测头中进行检测,以提高模型的检测能力^[16]。

YOLOv5 支持多种模型大小和精度级别,从小型的 YOLOv5s 到大型的 YOLOv5x 均可。本文选择 YOLOv5s 作为基础模型,其相较 YOLOv5 系列的其他模型,如 YOLOv5m、YOLOv5l 和 YOLOv5x,参数量和计算量更小,具有较快的检测速度、较好的检测精度以及较强的通用性。图 1 为 YOLOv5s 主干网络结构。

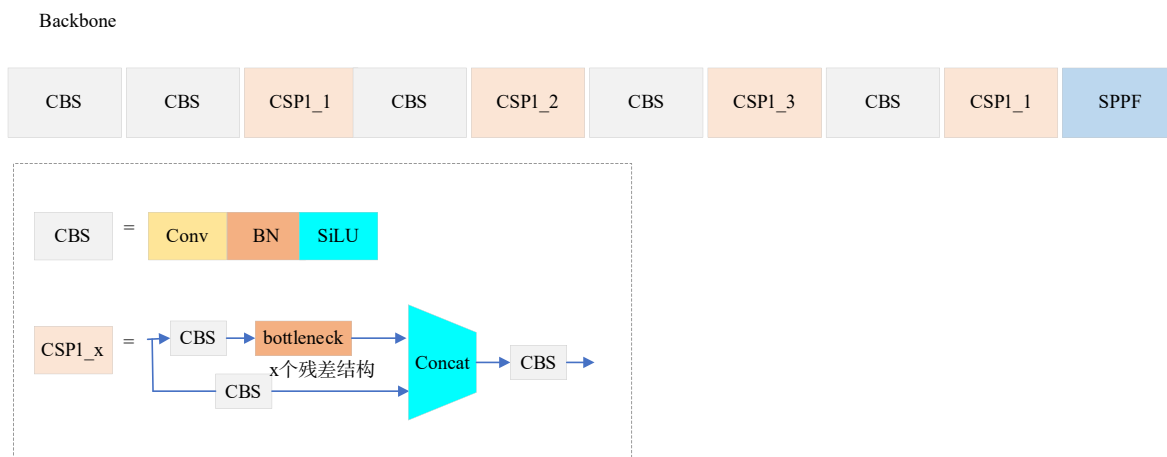


Fig. 1 YOLOv5s backbone network structure

图 1 YOLOv5s 主干网络结构

3 YOLOv5s+算法

3.1 BoTNet

文献[17]证实了Self-attention的有效性和实用性。在计算机视觉中使用Self-attention的简单做法就是直接采用

多头自注意力机制MHSA替换卷积层。本文将ResNet第4个block中BottleNeck的3×3卷积层替换为MHSA模块,形成新的模块,取名为BottleNeck Transformer (BoT)。BoT-Net由多个BoT block构成。如图2所示,BottleNeck与BoT两者唯一的不同在于3×3的卷积层和MHSA,其他没有任何区别。

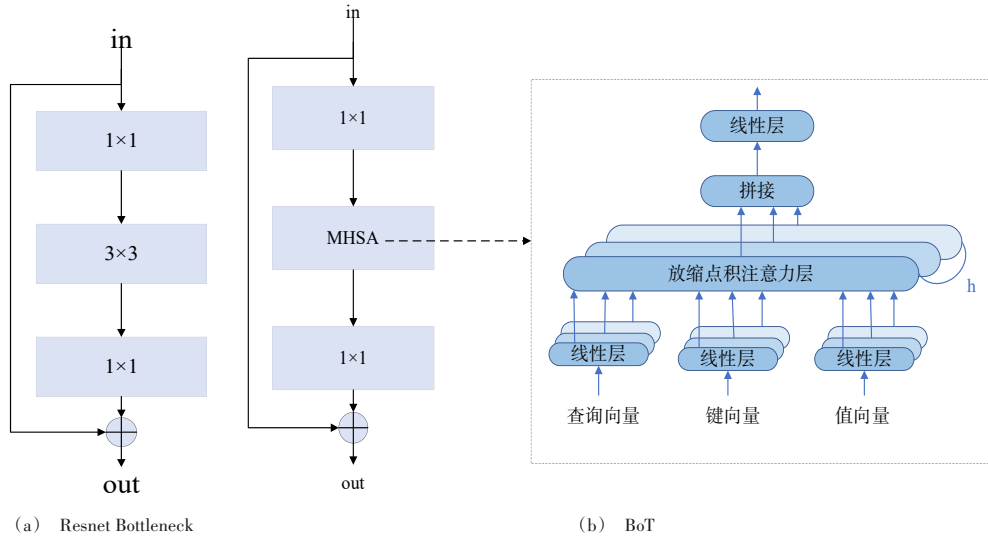


Fig. 2 BottleNeck and BoT
图2 BottleNeck与BoT

改进YOLOv5s的主干网络,即将主干网络第4个CSP1_X网络中的bottleneck按照图2所示方法构造BoT结构。图3为改进后的CSP1_X结构。通过改进,模型不仅可以更好地捕捉特征的复杂性和多样性,而且能减少特征冗余,提高特征的可靠性和有效性。

4 实验方法与结果分析

4.1 数据集采集与制作

本文使用的办公室吸烟检测数据集由Kaggle网络中竞赛数据集Cigarette Smoker Detection与自制数据集组成。由于Cigarette Smoker Detection数据集中存在部分图片数据重复以及标签缺失的问题,本文从中提取了部分数据并进行了整理,得到了一部分可用数据。同时在Label Img上对自己在办公室内拍摄图片上的香烟目标进行标注,统一标签为“cigarette”。整合所有数据,最终形成包含5 000张图片数据的办公室吸烟检测数据集。将该数据集按8:1:1的比例划分为训练集、验证集和测试集。部分办公室吸烟图像如图6所示。

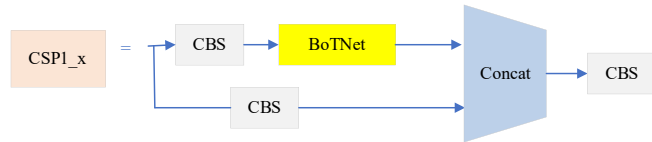


Fig. 3 Structure of improved CSP1_X
图3 改进后的CSP1_X结构

3.2 加权双向特征金字塔网络BiFPN

在多尺度特征融合中,不同的输入特征具有不同的分辨率,它们通常对融合的输出特征作出不平等的贡献。针对这一现象,本文将YOLOv5s中原有的特征融合网络替换为高效的加权双向特征金字塔网络BiFPN,该网络引入可学习的权重来学习不同输入特征的重要性,可以在不同特征层之间进行信息传递,从而提高目标检测的准确性。与只有一条自顶向下和一条自底向上路径的FPN+PAN结构不同,BiFPN将每个双向(自顶向下和自底向上)路径视为一个特征网络层,并多次重复同一层以启用更高级的特征融合^[18]。PANet与BiFPN的结构比较见图4。

4.2 实验环境与参数设置

本文实验在AMD Ryzen 5 5600H with Radeon Graphics处理器,主频为3.30 GHz,内存为16 GB,显卡型号为NVIDIA GeForce RTX 3050的PC机上进行,在CUDA11.8环境下利用深度学习框架PyTorch对模型进行训练与测试。模型训练次数epoch设为100,batch size设为8,输入分辨率均采用模型原始初始值640×640。

本文提出的YOLOv5s+模型在主干网络部分引入BoT-Net结构,颈部采用加权双向特征金字塔网络BiFPN。结构如图5所示。

4.3 评价指标

使用精确率P、召回率R和平均精确率mAP作为模型性能评价指标。其中,精确率用于衡量模型检测的准确性,召回率用于衡量检测的全面性,平均精度均值是由单

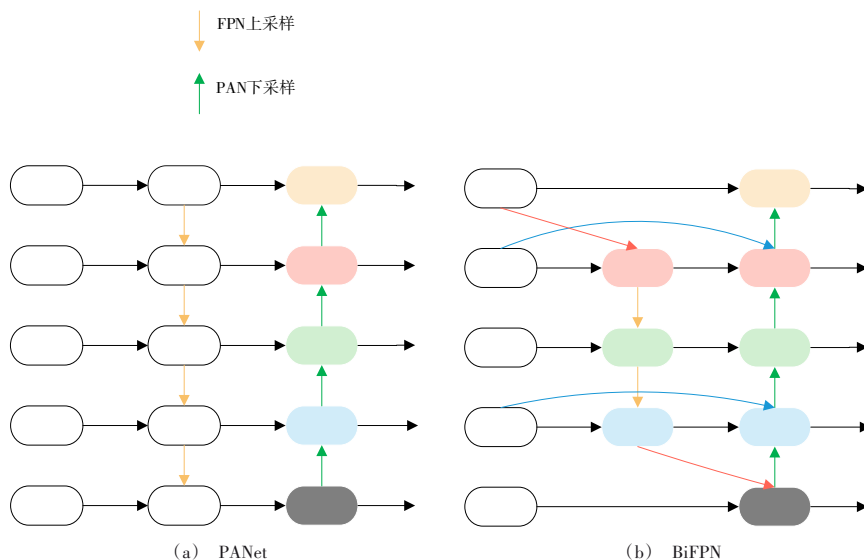


Fig. 4 Comparison of PANet and BiFPN Structures

图 4 PANet 与 BiFPN 结构比较

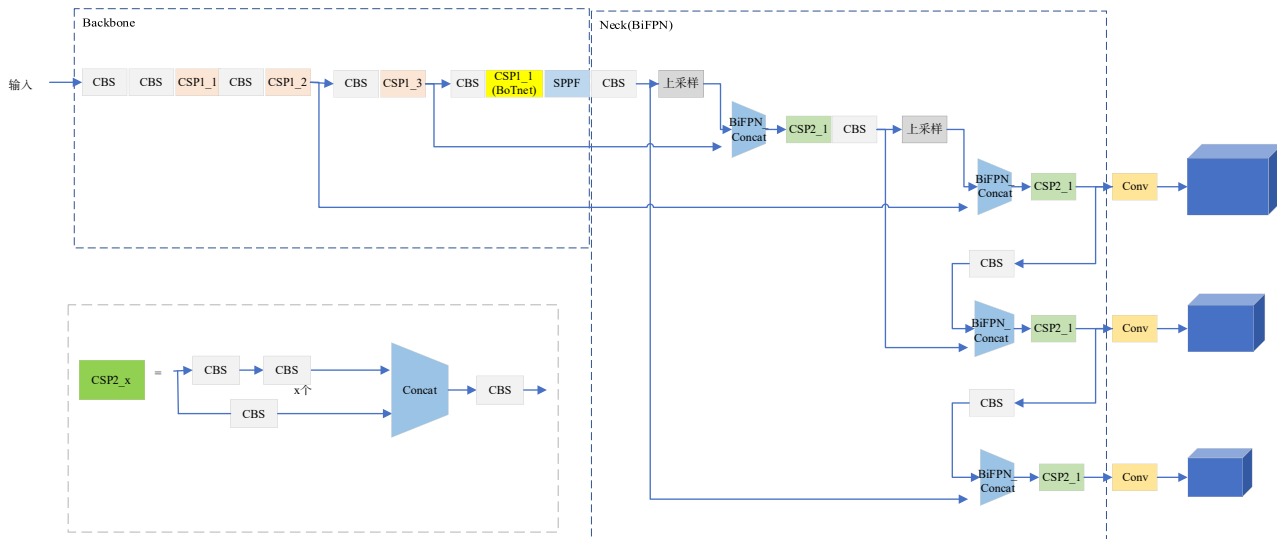


Fig. 5 YOLOv5s+ network structure

图 5 YOLOv5s+网络结构



Fig. 6 Partial office smoking image

图 6 部分办公室吸烟图像

类别精确率 AP 求和之后再求平均。具体计算公式分别为:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$AP = \int_0^1 P(r) dr \quad (3)$$

$$mAP = \frac{\sum_{i=1}^C AP_i}{C} \quad (4)$$

式(1)和式(2)中, T_p 表示将正类别预测为正确类别的个数, F_p 表示将负类别预测为正确类别的个数, F_n 表示将正类别预测为负类别的个数; 式(3)中, $P(r)$ 为精确率与召回率的平滑曲线; 式(4)中, i 表示第 i 个类别, C 表示总类

别数。

4.4 消融实验

在相同实验条件下分析比较以下4种模型的性能:①YOLOv5s模型;②改进模型1。向YOLOv5s模型主干网络中的残差结构中加入MHSA,构造BoTNet;③改进模型2。将YOLOv5s特征融合部分的FPN+PAN结构改为加权双向特征金字塔BiFPN;④本文模型YOLOv5s+。具体结果如表1所示。可以看出,单独将BoTNet和BiFPN加入YOLOv5s模型中均可以提高目标检测的精度和召回率,且同时加入两种结构的YOLOv5s+模型性能最佳,说明改进策略均有效。

Table 1 Ablation experimental results

表1 消融实验结果		%			
模型	修改主干网络	修改特征金字塔	P	R	mAP
YOLOv5s	×	×	78.7	77.5	76.4
改进模型1	√	×	81.8	80.6	79.7
改进模型2	×	√	80.2	80.1	78.9
YOLOv5s+	√	√	82.8	83.9	81.8

4.5 改进前后检测效果比较

改进前后网络模型在办公室场景下的实际检测效果比较如图7所示。可以看出,YOLOv5将图片中非烟支的小目标误检为烟支,而YOLOv5s+并未出现该误检现象,且YOLOv5s+对烟支的检测精度高于YOLOv5模型。



Fig. 7 Comparison of model detection effects before and after improvement

图7 改进前后模型检测效果比较

4.6 改进后模型训练过程

图8为改进后模型训练时的精度和训练损失随迭代次数的变化值。从训练次数Epochs的值为0开始,精度从0开始快速上升,训练损失值迅速下降,标志着网络模型开始对图片特征进行学习。当训练次数Epochs为60左右时,精度曲线趋于稳定,波动减少;损失值降至0.042,下降速度减缓。当训练次数Epochs为80左右时,精确率曲线趋向为0.81的稳定值,训练损失曲线趋向为0.02的稳定值。综上所述,改进后模型训练过程稳定,性能优良。

4.7 不同算法检测效果比较

为证明YOLOv5s+算法的检测性能,在办公室吸烟行为数据集上将其与主流YOLO系列模型YOLOv6^[19]、YO-

LOX^[10]、YOLOv7^[21]、YOLOv8^[22]进行比较,结果如表2所示。可以看出,YOLOv5s+算法表现最佳。

5 结语

本文对经典的一阶段目标检测网络模型YOLOv5s的主干网络和颈部网络进行改进,将其应用于办公室吸烟行为检测中,取得了较高的检测精度和召回率。然而改进模型YOLOv5s+仍有提升空间,例如本研究的数据集将卷烟和电子烟标记为相同的标签,后续计划进一步丰富数据集,将卷烟与电子烟区分开来,提高模型对不同吸烟行为的识别能力;还考虑将改进模型嵌入监控系统,对摄像头

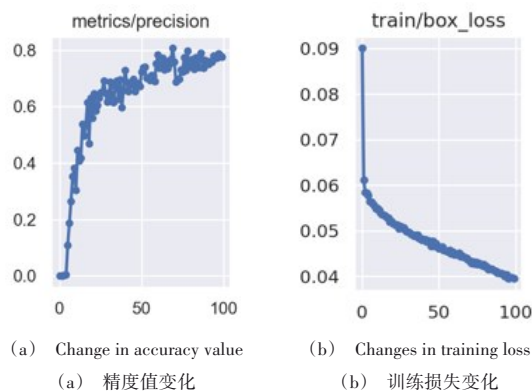


Fig. 8 Changes in the data during the model training process
图 8 模型训练过程数据变化

Table 2 Comparative experimental results of YOLOv5+ model and mainstream YOLO model

表 2 YOLOv5+模型与主流 YOLO 模型检测效果比较 %

算法	P	R	mAP
YOLOv5s	77.5	80.1	76.4
YOLOv6	69.8	64.6	76.2
YOLOX	76.4	82.8	74.1
YOLOv7	61.1	58.6	60.1
YOLOv8	75.4	72.0	77.8
YOLOv5s+	80.6	83.9	81.8

拍摄的视频流输入帧进行处理,通过图像特征提取、特征融合、目标分类以及目标定位等步骤识别并定位烟支位置,提高模型的实际应用价值。

参考文献:

[1] WANG C, XIAO D, CHI H. Summary of China's smoking health hazards report 2020 [J]. Chinese Journal of Circulation, 2021, 36(10):937-952. 王辰,肖丹,池慧. 中国吸烟危害健康报告 2020 概要[J]. 中国循环杂志, 2021, 36(10):937-952.

[2] WANG Z Y, LIAO H M, ZHANG R D, et al. Detection of smoking behavior of taxi drivers based on smoke characteristics [C]//Proceedings of the 12th National Conference on Signal and Intelligent Information Processing and Application, 2018: 368-373. 汪祖云,廖惠敏,张日东,等. 结合烟雾多特征的出租车司机吸烟行为检测[C]//第十二届全国信号和智能信息处理与应用学术会议论文集, 2018:368-373.

[3] XING Y Q. Smoking detection in surveillance scenes based on gesture recognition [D]. Hangzhou: Zhejiang University of Technology, 2020. 邢予权. 基于手势识别的监控场景下抽烟检测[D]. 杭州:浙江工业大学, 2020.

[4] LIU J. Research on smoking action recognition based on skeleton information [D]. Xuzhou: China University of Mining and Technology, 2021. 刘婧. 基于骨骼信息的吸烟动作识别方法研究[D]. 徐州:中国矿业大学, 2021.

[5] ZHAO J F. Research on smoking behavior detection method based on human key point recognition [D]. Tianjin: Hebei University of Technology, 2022. 赵鉴福. 基于人体关键点识别的吸烟行为检测方法研究[D]. 天津:河北工业大学, 2022.

[6] WEI Y H, WU S J, LIU T K, et al. Office smoking behavior detection

based on deep learning [J]. Artificial Intelligence and Robotics Research, 2023, 12(2):55-61.

魏袁慧,吴四九,刘天锴,等. 基于深度学习的办公室吸烟行为检测 [J]. 人工智能与机器人研究, 2023, 12(2):55-61.

[7] MA Y K. Research and application of outdoor cigarette detection algorithm based on YOLOv5 [D]. Yinchuan: Ningxia University, 2022. 马研凯. 基于 YOLOv5 的室外烟支检测算法研究与应用 [D]. 银川:宁夏大学, 2022.

[8] SUN Z L, XU X, ZHU Y L, et al. Smoking detection at oilfield work sites based on YOLOv5 method [J]. System Simulation Technology, 2021, 17(2):89-93. 孙召龙,徐昕,朱云龙,等. 基于 YOLOv5 的油田作业现场吸烟检测方法 [J]. 系统仿真技术, 2021, 17(2):89-93.

[9] SRINIVAS A, LIN Y L, PARMAR N, et al. Bottleneck Transformers for visual recognition [C]// Seattle: Conference on Computer Vision and Pattern Recognition, 2021.

[10] TAN M, PANG R, LE V. EfficientDet: scalable and efficient object detection [C]// Seattle: Conference on Computer Vision and Pattern Recognition, 2020.

[11] ZHU X, LU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on Transformer prediction head for object detection on drone-captured scenarios [C]// Montreal: IEEE/CVF International Conference on Computer Vision Workshops, 2021.

[12] ZHU D, DAI L, DU P. CCE-YOLOv5s: an improved YOLOv5 model for UAV small target detection [C]// Dali: IEEE 5th International Conference on Civil Aviation Safety and Information Technology, 2023.

[13] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]// Seattle: Conference on Computer Vision and Pattern Recognition Workshops, 2020.

[14] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]// Honolulu: Conference on Computer Vision and Pattern Recognition, 2017.

[15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [C]// Harbin: 5th International Conference on Mechanical, Control and Computer Engineering, 2020.

[16] LIN T Y, DOLLAROLL P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// Honolulu: IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[17] ASHISH V, NOAN S, NIKI P, et al. Attention is all you need [C]// Toronto: IEEE International Conference on Acoustics, Speech and Signal Processing, 2021.

[18] LI R, ZHANG B P, LIU W, et al. PANet: an end-to-end network based on relative motion for online multi-object tracking [J]. IEEE Robotics and Automation Letters, 2022, 7(2):1840-1847.

[19] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications [DB/OL]. <https://arxiv.org/abs/2209.02976>.

[20] GE Z, LIU S T, WANG F, et al. Yolox: exceeding yolo series in 2021 [DB/OL]. <https://arxiv.org/abs/2107.08430>.

[21] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new states-of-the-art for real-time object detectors [C]// Vancouver: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

[22] TERVEN J, CORDOVA-ESPARZA D. A comprehensive review of YOLO: from YOLOv1 to YOLOv8 and beyond [DB/OL]. <https://arxiv.org/pdf/2304.00501v1>.

(责任编辑:尹晨茹)