轻量化对抗增强的物流违规操作识别方法

秦法波，张媛，朱磊，杨晓静，高振清
（北京印刷学院 机电工程学院，北京 102600）

摘要：目的 针对复杂强噪背景下物流违规操作难以有效识别的问题，提出一种轻量化对抗增强的物流违规操作检测方法。**方法** 以 YOLOv5 为基础框架，提出轻量化的 GhostC3 模块，运用对抗学习的思想提出轻量对抗模块，将原有结构中的 C3 模块修改为轻量化的 GhostC3 模块，Conv 模块修改为轻量对抗模块，并将定位损失修改为 CIOU 损失。**结果** 通过实验验证可知，本文方法针对复杂强噪背景下物流违规操作具有优异的检测效果，其中本文方法相较于 YOLOv5 方法的检测平均精度均值提高了 1.69%，模型参数量降低了 45.14%，检测速度提高了 2.46%。**结论** 本文提出的方法具有参数量低、检测速度快和精度高等特点，针对复杂强噪背景下物流违规操作的检测具有一定的先进性和实用性，充分满足物流违规操作检测需求。

关键词：物流；计算机视觉；目标检测；YOLOv5
中图分类号：TP391.4 **文献标识码：**A **文章编号：**1001-3563(2023)09-0265-10
DOI：10.19554/j.cnki.1001-3563.2023.09.032

Recognition Method of Improper Operation in Logistics with Lightweight and Enhanced Countermeasures

QIN Fa-bo, *ZHANG Yuan*, *ZHU Lei*, *YANG Xiao-jing*, *GAO Zhen-qing*
(School of Mechanical and Electrical Engineering, Beijing Institute of Graphic Communication, Beijing 102600, China)

ABSTRACT: The work aims to propose a method of detecting improper operations in logistics with lightweight and enhanced countermeasures in view of the difficulty in effectively identifying improper operations in logistics under the background of complex and strong noise. Based on YOLOv5, the lightweight GhostC3 module was offered, and the lightweight countermeasure module was proposed with the idea of countermeasure learning. The C3 module in the original structure was changed into the lightweight GhostC3 module, the Conv module was changed into the LAconv module, and the positioning loss was changed into CIOU loss. Finally, through experimental verification, the method proposed had an excellent detection effect against improper operations under the background of complex and strong noise. Compared with YOLOv5, the average detection accuracy of the method proposed increased by 1.69%, the number of model parameters decreased by 45.14%, and the detection speed was improved by 2.46%. The method proposed has the characteristics of a low number of parameters, fast detection speed, and high accuracy. It is advanced and practical for the detection of improper operations in logistics under the background of complex and strong noise, and fully meets the detection needs of improper operations in logistics.

KEY WORDS: logistics; computer vision; object detection; YOLOv5

收稿日期：2022−07−08
基金项目：北京市教育委员会科技/社科计划项目资助（KZ202210015020）；北京印刷学院校级项目（Ee202204）
作者简介：秦法波（1999—），男，硕士生，主攻视频图像识别。
通信作者：朱磊（1982—），男，博士，讲师，主要研究方向为物流安全检测。

2021 年全年邮政行业完成业务的总量为 13 698 亿元，比上年增长 25.1%[1]。随着物流业的发展，包裹在物流作业中损坏、破损等问题尤为严重，而造成这一现象的直接原因便是物流活动中的违规操作。物流违规操作为是指在包裹在揽收、中转、仓储以及配送等整个物流环节中，作业人员对包裹进行野蛮作业的行为，其中较为常见的物流违规操作有足踢、抛扔和踩踏。物流违规操作的发生会增加包裹内物品损坏的风险，其中违规操作为在快递方面尤为凸显，据《快递服务体验式调查报告》显示，物流违规操作在快递行业十分普遍，其中至少 16.1%的快递件出现不同程度的损伤[2]。2021 年 12 月邮政业用户申诉情况通告显示，快递服务申诉的主要问题中快件损毁占申诉总量的 21.8%[3]。

目前，物流违规操作的识别主要有 2 种技术路线：第 1 种为基于传感器集群的识别方法；第 2 种是基于视频和图像的识别方法。基于传感器集群的物流违规操作识别方法是将惯性和振动等传感器内置于包裹中，通过开发专用硬件[4-6]和云平台[7-9]对违规操作进行分析处理，具有代表性的有丁奥等[10]提出的基于加速度分布特征的快递暴力分拣识别方法，有效对包裹的违规操作进行了识别及分类，对包裹全流程监测具有重要意义。但是该类方法在进行识别时，专用的检测终端的固定安装方式相对烦琐并且需要占用包装容积，同时，由于专用检测终端一般需锂电池供电，对航空等运输方式的安全有不利影响，因此，基于传感器集群的快递异常行为识别方法存在一定弊端。近年来，基于视频和图片对目标行为进行检测的方法得到迅速发展，并被广泛应用于农业、医疗、工业制造等诸多领域。针对物流违规操作的识别，已有学者进行了深入的研究。吴鹏勃等[11]提出的基于 LSTM+Attention 和 MobileSSD 模型的快递暴力分拣检测系统可通过姿态数据进行行为识别。尚淑玲等[12]利用小波包分析方法，对采集的物流分拣的图像行为特征进行有效的提取，为物流暴力分拣识别提供依据。邓秀琴等[13]针对物流违规操作提出一种基于计算机视觉的暴力分拣行为识别方法，但以上视频和图像识别方法在复杂的物流环境中存在检测效果和稳定性较低、模型参数量高、检测速度慢等问题。因此针对上述问题，轻量且迅速的物流违规操作检测方法具有重要的理论和实际意义。

本文针对物流违规操作问题以 YOLOv5 网络为主要框架提出一种面向复杂强噪背景的轻量对抗检测方法，克服物流违规操作检测成本高和检测精度差的问题，有效降低人工干预工作量。本文首先介绍 YOLOv5 模型的网络结构；其次阐述网络的改进与优化，并且进行模型训练与实验结果分析；最后总结全文的工作内容并对取得的实验结果进行分析。

1　算法简介

YOLO 系列算法[14-17]是典型的一阶段算法，具有简便、迅速以及适用性广等特点。YOLOv5 在 YOLO 系列算法的基础上做了一系列的改进，性能得到了进一步的提升，YOLOv5 网络模型如图 1 所示，主要包含输入端（Input）、主干网络（Backbone）、特征融合端（Neck）和预测端（Head）等 4 个部分。

在输入端，YOLOv5 主要运用 Mosaic 数据增强方法、自适应锚框计算、自适应缩放图片。其中 Mosaic 数据增强是指通过随机缩放、裁剪以及排布的方法将随机选择的若干个图片组合，提高了模型的泛化能力，防止过拟合的出现。在网络模型训练过程中，网络会自动计算适合当前数据集的最佳锚框，通过聚类的锚框可以更快得到准确的预测框。为方便进行网络模型的训练和满足不同数据集图片尺寸的要求，自适应缩放图片是将原图片缩放到统一尺寸，再输入到网络中训练。

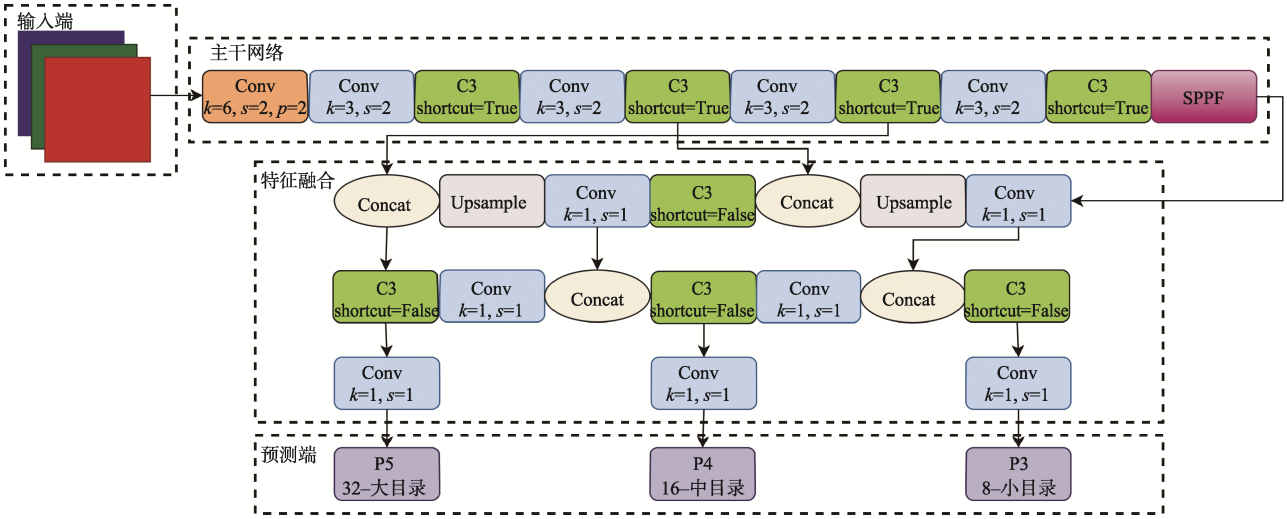


图 1　YOLOv5 网络结构
Fig.1 YOLOv5 network structure

YOLOv5 在主干网络部分应用 Conv 模块、C3 结构、SPPF 池化金字塔结构。Conv 模块是卷积层、BN（Batch Normalization）层和激活函数的组合，在整个网络中具有重要作用。Bottleneck 模块是 C3 模块中的一个重要残差部件，在一定程度上降低了模型的计算参数量并且提高了特征提取的效率。C3 模块是指含有 3 个 Conv 模块 BottleneckCSP（Bottleneck Cross Stage Partial）结构，具有加深网络结构的作用。C3 结构输入的特征图经过 2 条路线拼接（Concat）后再经过 Conv 模块得到输出的特征图。SPPF 池化金字塔结构对主干网络具有扩大网络感受野的作用，其结构是将输入特征图进行连续的最大池化下采样，最后将输出的特征图拼接。

输入的图片经过主干网络后，为满足下游任务中不同尺度目标的要求，特征融合端采用特征金字塔结构（Feature Pyramid Networks，FPN）和路径聚合网络结构（Path Aggregation Network, PAN）。FPN 是将高层特征图与低层特征图的特征进行融合，PAN 网络是在 FPN 的基础上改进而来，其增加了一条自下向上的特征融合方案，使不同层次的语义信息能够更好的融合。特征融合端对图像中不同尺度目标的检测效果具有大幅度的提升，使网络既可以学习到深层次的语义特征，又可以学习到图像整体的物体分布信息。Yolov5 在输出端采用 GIOU 函数作为定位损失的损失函数，并且在检测的后处理过程中，使用非极大值抑制的方法（Non Maximum Suppression，NMS）对目标框筛选，有效提高了模型的检测能力。

2　模型的改进与优化

2.1　C3 模块的改进

为解决传统的卷积操作得到的特征图中部分通道获取的信息冗余，参考 Ghostnet 网络[18]提出一种轻量化的 GhostC3 模块。由于在训练过程中部分通道未学习到非常有价值的特征，为解决通道信息冗余，并且尽可能使用更少的参数量获取含有更多信息的特征图。将 C3 结构中的 Conv 模块更改为 Ghostconv 模块，其结构如图 2 所示。

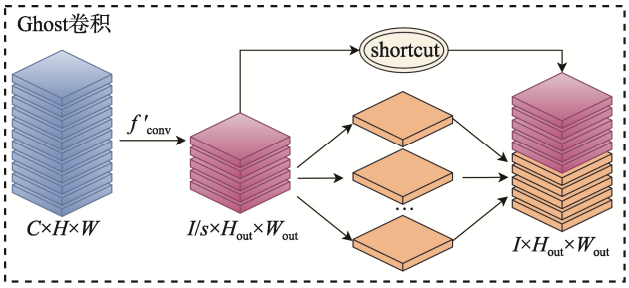


图 2　Ghost 卷积模块
Fig.2 Ghostconv module

给定输入特征图 $\boldsymbol{X}\in R^{C\times H\times W}$，Conv 模块如式（1）所示，将输入特征图通过函数 f_{conv} 得到输出特征图；而 Ghostconv 模块如式（2）所示，将输入特征图通过函数 f'_{conv} 经过缩放因子 s 降低通道数去除冗余通道，再通过式（3）的线性运算得到全新特征图。其中 y_i 是 Y'中第 i 个特征图，函数 $\varphi_{i,j}$ 是指第 j 个线性运算，Y_{ij} 是输出特征集合的元素。将降低通道数的捷径（Shortcut）特征图和线性运算得到的特征图拼接得到最终的特征图。

$$Y = X\times f_{conv} \tag{1}$$

$$Y' = X\times f'_{conv} \tag{2}$$

$$Y_{ij} = \varphi_{i,j}(y_i), \forall i = 1,\ldots,\frac{I}{s}, j = 1,\ldots,s \tag{3}$$

如图 3 所示，将 Bottleneck 模块中的 Conv 模块更改为 Ghostconv 模块，并应用于 GhostC3 模块。其中，Bottleneck 模块步距（Stride）设置为 1，输入的特征图经过 2 个 Ghostconv 模块与其自身进行相加得到输出特征图。同时，将 C3 模块更改为 GhostC3 模块，并且将输出时的 Conv 模块更改为 Ghostconv 模块，其中更改后的 GhostC3 模块 Backbone 中 GhostC3 使用捷径（Shortcut），Neck 中 GhostC3 不使用捷径（Shortcut）。

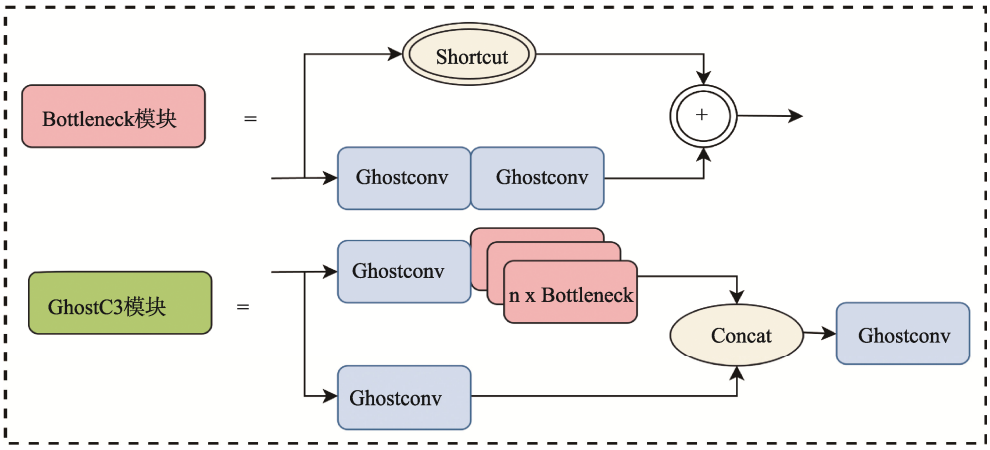


图 3　Bottleneck 模块和 GhostC3 模块
Fig.3 Bottleneck module & GhostC3 module

2.2　Conv 模块的改进

为提高网络的强噪声数据下的检测能力，引入对抗学习的思想，其中生成的对抗网络[19]（Generative Adversarial Network，GAN）主要包含如图 4 所示的生成器（Generator，G）和鉴别器（Discriminator，D）两部分。生成器和鉴别器之间形成对抗，并且生成器和鉴别器可以是任何非线性映射的函数。

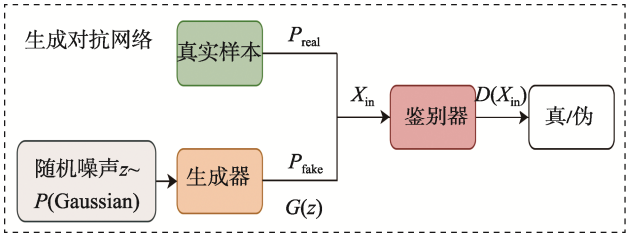


图 4　生成对抗网络结构
Fig.4 GAN structure

生成对抗网络生成器负责将随机服从高斯分布的噪声 z 生成为分布近似于实际样本的数据 $G(z)$；鉴别器负责判断输入的特征 X_{in} 是真实样本还是生成器生成的样本。其中生成器要不断及时优化生成的数据，使真实数据分布 P_{real} 和生成数据分布 P_{fake} 尽可能相似，以达到鉴别器无法鉴别出真实数据和生成数据的目的。鉴别器目的是提升自身的鉴别能力，更好地识别真实和虚假数据的区别，因此其优化公式见式（4）。

$$\min_G \max_D L(D,G) = E_{X_{in}\sim P_{real}}[\log D(X_{in})] + E_z[\log(1-D(G(z)))] \tag{4}$$

为提高网络的在高噪声图片的检测效果引入通道注意力模块[20]，通道注意力模块可以获取特征图不同通道间的重要程度。在网络训练的过程中，通道注意力模块会增大存在重要语义信息的通道权重，降低冗余信息通道的权重。通道注意力模块如图 5 所示，由全局平均池化（Global Average Pooling，GAP）、若干个全连接层（Fully Connected Layer，FC）和激活函数组成。

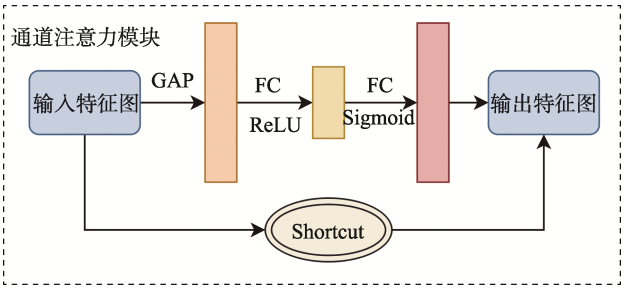


图 5　通道注意力模块
Fig.5 SE module

给定输入特征图 $\boldsymbol{X}\in R^{C\times H\times W}$，通过全局平均池化压缩特征矩阵的高和宽，输出特征集合 $\boldsymbol{u}=[u_1, u_2, u_3, \ldots, u_C]$。全连接层和激活函数的计算过程可表示为式（5），先降低通道维度再升高通道维度。其中 σ_1 为 ReLU 激活函数，σ_2 为 Sigmoid 激活函数，计算式见式（6）—（7），最终输出通道权重与原特征图的乘积。

$$V = \sigma_2(FC_2(\sigma_1(FC_1(u)))) \tag{5}$$

$$\sigma_1(x) = \max(0, x) \tag{6}$$

$$\sigma_2(x) = \frac{1}{1+e^{-x}} \tag{7}$$

为增强网络的特征学习能力并提升网络的抗噪能力，结合对抗学习思想和通道注意力模块，提出一种如图 6 所示的轻量化对抗模块（Lightweight Adversarial Conv，LAconv）。将 Backbone 和 Neck 中的 Conv 模块修改为 LAconv 模块，将 Ghostconv 模块中的线性变换作为生成器 G，将通道注意力模块作为鉴别器 D。轻量化对抗模块中运用生成器 G 的线性变换生成类似于真实特征图的虚假特征图，然后与降低通道的真实特征图拼接，输入至通道注意力模块，将通道注意力模块作为鉴别器来鉴别生成数据的真伪，二者形成对抗关系。

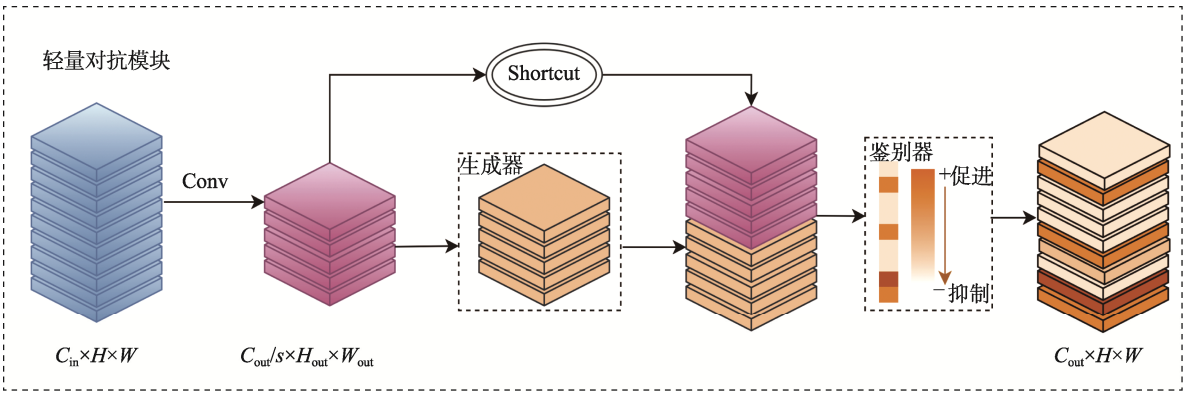


图 6　轻量对抗模块
Fig.6 LAConv module

2.3　损失函数的改进

YOLOv5 的损失函数由分类损失（Classification Loss）、定位损失（Localization Loss）和置信度损失（Confidence Loss）3 个部分组成。其中分类损失和置信度损失使用二值交叉熵损失（Binary CrossEntropy Loss，BCELoss），定位损失 L_G 使用 GIOU（Generalized Intersection Over Union，G）损失[21]。GIOU 损失(见式(8)—(9))取代了 IOU（Intersection Over Union，I）损失，解决了预测框和真实框不相交时损失无法回传的问题，不仅关注重叠区域，还关注其他非重合区域，能更好地反映两者的重合度。但是，GIOU 损失在预测框和真实框没有很好的对齐时，会导致最小外接框的面积增大，从而使 GIOU 的值变小，不利于模型训练。

$$G = I - \frac{A^{C} - U}{A^{C}} \tag{8}$$

$$L_G = 1 - G \tag{9}$$

为解决上述定位损失的缺点，将 GIOU 损失 L_G 调整为 CIOU（Complete Intersection Over Union，C）损失 L_C[22]（见式（10）—（13））。具体调整如下，首先将 GIOU 中的惩罚项最小外接矩形修改为如式（10）所示的最小化真实框和预测框中心点的欧式距离 $\rho(b,b^{gt})$，并加入最小包围 2 个边界框的对角线长度 c 的平方进行修正。同时，加入能够同时包含预测框和真实框的最小闭包区域的对角线距离。并且，为了充分地考虑横纵比对预测的影响，加入如式（13）—（14）所示的惩罚项，其中 w^{gt}、h^{gt} 分别代表预测框的宽和高，w、h 分别代表目标框的宽和高。通过上述的调整，定位损失充分考虑了横纵比与预测框和真实框的关系，并且提高了网络模型的收敛速度和平均精度。

$$C = I - \frac{\rho^2(b,b^{gt})}{c^2} - \alpha v \tag{10}$$

$$L_C = 1 - C \tag{11}$$

$$\alpha = \frac{v}{1 - I + v} \tag{12}$$

$$v = \frac{4}{\pi}\left(\arctan\frac{\omega^{gt}}{h^{gt}} - \arctan\frac{\omega}{h}\right)^2 \tag{13}$$

2.4　改进后的网络结构

YOLOv5 网络本身具有检测速度快、适用性广的特点，以 YOLOv5 网络为主要框架，如图 7 所示将 C3 模块修改为 GhostC3 模块、Conv 模块修改为 LAconv 模块，并且修改定位损失函数为 CIOU 损失函数。改进后的网络具有检测效果优异、推理迅速、轻量化的特点，满足复杂高噪物流环境下违规操作的检测。

3　实验与分析

3.1　实验环境

实验所使用的处理环境的显卡为 NVDIA Tesla K80，CPU 为 3 个 Xeon E5–2678 v3 处理器，操作系统为 Ubuntu 18.04，深度学习框架为 Pytorch 架构。具体配置见表 1。

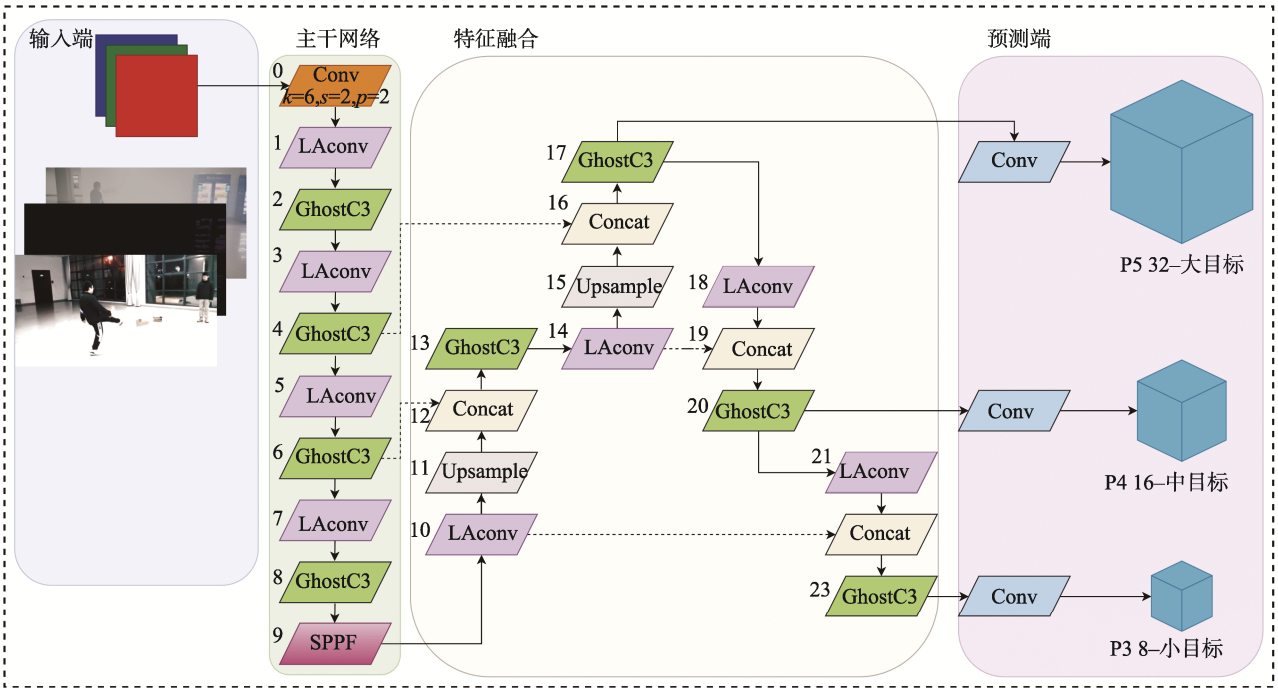


图 7　改进后的 YOLOv5 结构
Fig.7 Improved YOLOv5 structure

表 1　硬件环境配置
Tab.1 Experimental environment configuration

硬件	配置信息
操作系统	Ubuntu 18.04
Python 版本	Python3.8
CPU	3×Xeon E5–2678 v3
GPU	NVIDIA Tesla K80
加速环境	CUDA11.3 Cudnn8.0

训练时采用 SGD 优化器优化网络，使用 Mosaic 数据增强，调整批量大小为 64，初始学习率设置为 0.001，动量为 0.937，权重衰减系数为 0.000 5，迭代次数设置为 300，模型深度系数为 0.33，宽度系数为 0.5。

3.2　数据预处理

3.2.1　数据集采集

为验证模型效果，使用物流违规操作数据进行验证。因物流违规操作数据在真实物流环境下采集困难，实验所需数据集[23]是由本文作者使用 TP–LINK 高清摄像头模拟物流违规操作行为，包含足踢、抛扔和踩踏 3 类常见的物流违规操作行为，其中足踢指用脚踢包裹以至导致包裹发生位移或变形；抛扔是指包裹从高于地面 30 cm 处被人为抛落；踩踏是指包裹被人为踩踏并发生形变。数据集共包含图片 1 215 张，其中足踢 344 张，抛扔 531 张，踩踏 340 张。为尽可能模拟真实物流环境，调用 Opencv 库将获取的数据集处理为如图 8 所示的雾天、高曝光、黑夜场景，其中每个动作下的每类场景数量占比相同。

图 8　数据集示意图
Fig.8 Schematic diagram of data set

3.2.2　数据标注

不同类别的图片数据使用 Labelimg 数据标注工具进行标注，其中物流违规操作分为 3 类，分别是足踢、抛扔、踩踏。其中类别 0 代表足踢，标签为 Kick；类别 1 代表抛扔，标签为 Throw；类别 2 代表踩踏，标签为 Trample。按照表 2 示例标注后的文件分别代表标注的类别、经过规一化操作的人工标注框的中心点坐标、标注框宽度和高度，最终将数据集按 8∶2 随机分配为训练集和测试集，得到实验所需数据集。

表 2　数据标注示例
Tab.2 Example of data annotation

标注名称	中心点坐标		宽度	高度
	x 轴	y 轴		
Kick（0）	0.370 258	0.495 199	0.084 977	0.159 122
Throw（1）	0.410 091	0.508 880	0.170 713	0.217 213
Trample（2）	0.492 413	0.850 949	0.151 745	0.219 512

3.3　模型评价指标

准确率（Precision，P）、召回率（Recall，R）是衡量模型训练结果的重要指标。准确率为真实存在违规操作且被预测出存在异常行为违规操作的预测框占所有预测框的比例，召回率为真实存在违规操作且被预测出存在违规操作的预测框占人工标注框的比例。P 和 R 的计算式分别见式（14）—（15），其中 T_P 表示正类被识别为正类的样本个数，F_P 表示正类被识别为负类的样本个数，F_N 表示负类被识别为负类的样本个数。

$$P=\frac{T_P}{T_P+F_P} \tag{14}$$

$$R=\frac{T_P}{T_P+F_N} \tag{15}$$

平均精度（Average Precision，A_P）、平均精度均值（Mean Average Precision，A_{mAP}）是衡量目标检测模型优劣的重要指标，其计算式分别见式（16）—（17）。式中 A_P 为 PR 曲线围成的面积，其中 PR 曲线代表的是精准率与召回率的关系，A_{mAP} 为足踢、抛扔和踩踏 3 类平均精度的均值。

$$A_p=\int_0^1 PR\mathrm{d}R \tag{16}$$

$$A_{mAP}=\frac{\sum A_p}{n} \tag{17}$$

3.4　实验结果与分析

3.4.1　消融实验

为验证网络修改的有效性，将网络的改进依次对比，见表 3。其中，Base 为未改进的 YOLOv5 基础

表 3　消融实验对比
Tab.3 Comparison of ablation experiment

模型	GhostC3	LAconv	CIOU	平均精度均值/%	参数量/10^6	检测时间/ms
Base				94.5	6.69	20.3
改进 1	√			92.9	4.71	16.7
改进 2		√		96.6	5.67	22.9
改进 3			√	95.1	6.69	20.5
本文方法	√	√	√	96.1	3.67	19.8

版本；改进 1 为将 C3 模块修改为 GhostC3 模块；改进 2 为将 Conv 模块修改为 LAconv 模块；改进 3 为将 GIOU 损失函数修改为 CIOU 损失函数。由表 4 可知，本文方法的平均精度均值较高且值为 96.1%，参数量最小为 3.67×10^6，检测时间较短为 19.8 ms。改进 1 相较于 Base 版本的平均精度均值降低了 1.6%，参数量降低了 1.98×10^6，检测时间降低了 3.6 ms；改进 2 相较于 Base 版本的平均精度均值提高了 2.1%，参数量降低了 1.02×10^6，检测时间增加了 2.6 ms；改进 3 相较于 Base 版本的平均精度均值提高了 0.6%，参数量降低相同，检测时间增加了 0.3 ms；本文方法相较于 Base 版本的平均精度均值提高 1.6%，参数量降低了 3.02×10^6，检测时间降低了 0.5 ms。由此综合可见，本文所提出的检测方法明显优于 Base 版本和其他改进方法，本文方法针对复杂高噪声图片有效提高了算法效果。

3.4.2　对比实验

为更好地展示改进后网络模型的优势，采用相同的违规操作数据集在相同的实验环境下进行训练，训练结果如图 9 和表 4 所示。

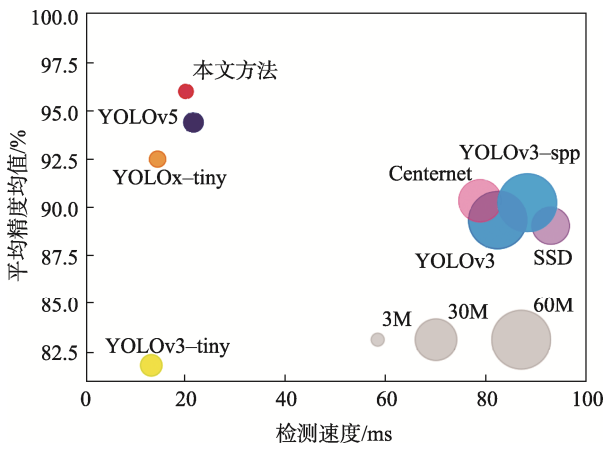


图 9　实验结果对比
Fig.9 Comparison of experimental results

1）在识别精度方面。本文方法检测的平均精度均值为 96.1%，比 YOLOv3–tiny[17]、YOLOx–tiny[24]、YOLOv3[17]、YOLOv3–spp[17,25]、SSD[26]、Centernet[27]、YOLOv5 的平均精度均值分别高出 14.3%、3.5%、6.7%、5.8%、7%、5.7%、1.6%。

表 4　实验结果对比
Tab.4 Comparison of experimental results

模型	A_{mAP}/%	A_P/%		
		足踢	抛扔	踩踏
YOLOv3–tiny	81.8	73.9	78.7	92.8
YOLOx–tiny	92.6	84	94.3	99.5
YOLOv3	89.4	80.7	92.9	94.5
YOLOv3–SPP	90.3	82.7	93.2	95
SSD	89.1	80	88.2	99.2
Centernet	90.4	84.2	90.6	96.5
YOLOv5	94.5	91.9	92.1	99.5
本文方法	96.1	89.6	97.5	99.5

2）在参数量大小方面。本文方法的参数量为 3.67×10^6，比 YOLOv3–tiny、YOLOx–tiny、YOLOv3、YOLOv3–spp、SSD、Centernet、YOLOv5 的参数量分别低了 1.17×10^6、4.6×10^6、54.67×10^6、55.45×10^6、20.86×10^6、28.04×10^6、3.02×10^6。

3）在检测时间方面。本文方法单张图片的检测速度为 19.8 ms，比 YOLOv3、YOLOv3–spp、SSD、Centernet、YOLOv5 的检测时间分别低了 62.5、68.5、73.2、59、0.5 ms，比 YOLOv3–tiny、YOLOx–tiny 的检测时间略高了 6.9、5.7 ms。

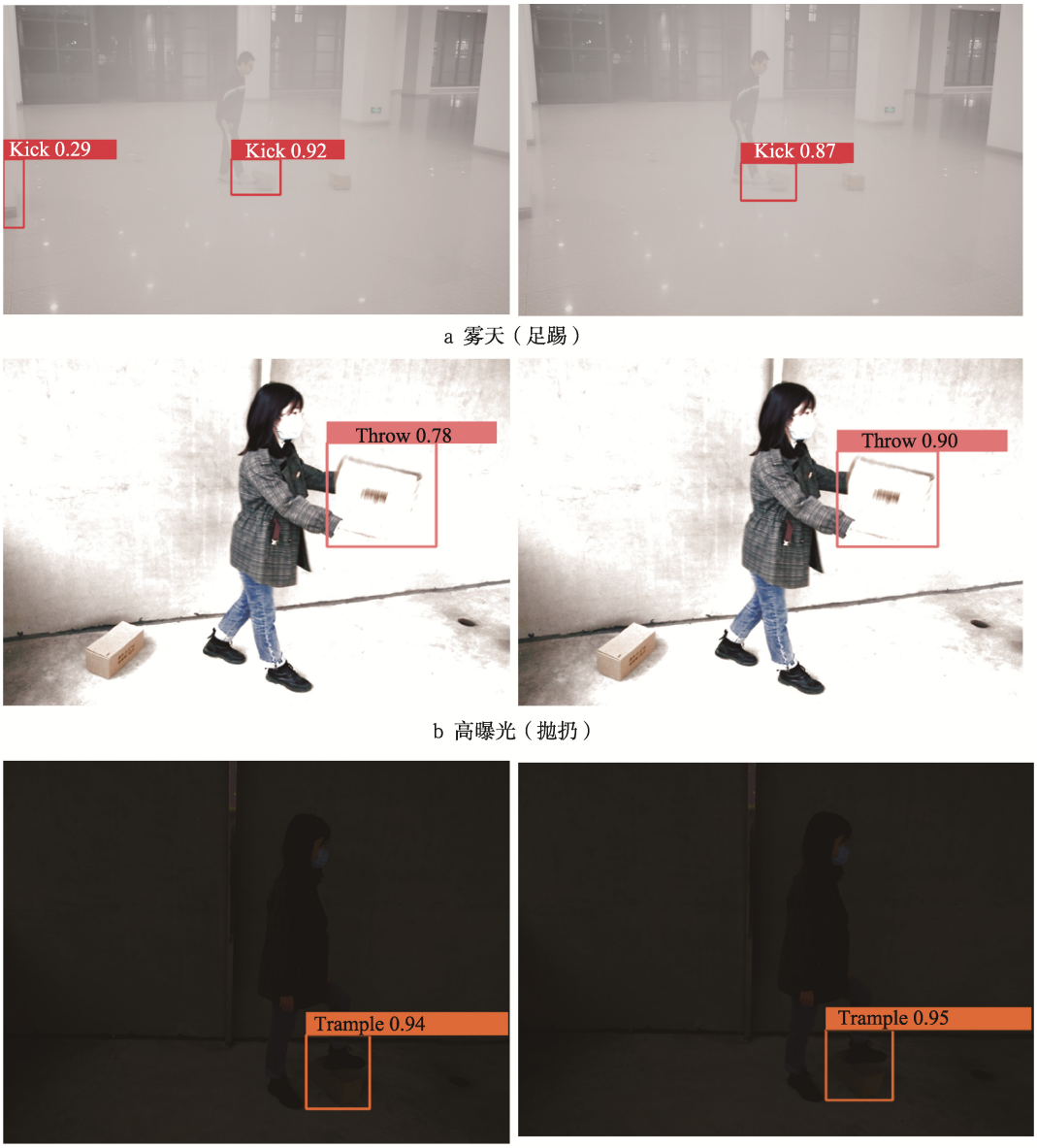
4）在单类违规操作的识别精度方面。如表 4 所示，本文方法足踢检测的平均精度均值比 YOLOv3–tiny、YOLOx–tiny、YOLOv3、YOLOv3–spp、SSD、Centernet 的平均精度均值分别高了 15.7%、5.6%、8.9%、6.9%、9.6%、5.4%，但比 YOLOv5 的平均精度均值略低了 2.3%。本文方法抛扔检测的平均精度均值比 YOLOv3–tiny、YOLOx–tiny、YOLOv3、YOLOv3–spp、SSD、Centernet、YOLOv5 的平均精度均值分别高了 18.8%、3.2%、4.6%、4.3%、9.3%、

6.9%、5.4%。本文方法踩踏检测的平均精度均值比 YOLOv3–tiny、YOLOv3、YOLOv3–spp、SSD、Centernet 的平均精度均值分别高了 6.7%、5%、4.5%、0.3%、3%。

经过本文方法与 YOLOv3–tiny、YOLOx–tiny、YOLOv3、YOLOv3–spp、SSD、Centernet、YOLOv5 等方法在平均精度均值、单张图片检测时间、模型参数量和单类违规操作识别精度方面的对比，得出本文方法的平均精度均值最高，为 96.1%，单类违规操作识别精度除足踢外普遍高于其他算法，参数量最小且值为 3.67×10^6，且单张图片检测速度为 19.8 ms，略高于 YOLOv3–tiny、YOLOx–tiny 的检测速度。因此，本文的网络结构在面向复杂高噪图片取得了优异的检测效果，在保证平均精度均值提高的前提下，有效降低了模型参数量和检测时间。

3.4.3　检测结果分析

为更好地验证模型的效果，如图 10 选取测试集中不同复杂环境图片进行测试。其中图 10a 为雾天异常行为的检测效果，左图为 YOLOv5 算法检测时出现误检现象，将墙角检测为异常行为，右图为本文算法有效检测并定位出异常操作的位置。图 10b 为高曝光下异常行为的检测效果，本文算法（右图）准确识别出异常行为，并且预测框的位置更加准确，检测效果明显高于 YOLOv5 算法（左图）的。图 10c 是黑夜下异常行为的检测效果，本文算法检测精度略高于 YOLOv5 算法的。故本文提出的轻量对抗的方法减少了误检，提高了检测能力。



a 雾天（足踢）

b 高曝光（抛扔）

c 黑夜（踩踏）

图 10　检测结果

Fig.10 Detection results

4　结语

物流违规操作检测是物流环境安全监测领域重要的研究课题。本文以 YOLOv5 为主要框架，提出了 GhostC3 模块和 LAconv 模块，修改定位损失函数为 CIOU 损失函数，应用于物流违规操作检测。经过实验验证，在提出轻量化模块和对抗学习的思想加持下，本文方法在复杂高噪的环境下仍具有优异的检测效果。本文方法的平均精度均值、参数量和推理速度均优于 YOLOv5 的，能够有效解决物流违规操作行为无法识别、识别困难的问题，有效提高了检测效果，减少了人工干预检测的成本。

参考文献：

[1] 中华人民共和国中央人民政府官网. 中华人民共和国 2021 年国民经济和社会发展统计公报[EB/OL]. (2022-02-28)[2022-07-08]. http://www.gov.cn/index. htm.
The Official Website of the State Council of the People's Republic of China. Announcement of National Economic and Social Development of the People's Republic of China in 2021[EB/OL]. (2022-02-28)[2022-07-08]. http://www.gov.cn/index.htm.

[2] 中国消费者协会官网. 快递服务体验式调查报告[EB/OL]. (2017-06-08)[2022-07-08]. https://www.cca.cn/.
The Official Website of the China Consumers Association. Investigation Report on Package Service[EB/OL]. (2017-06-08)[2022-07-08]. https://www.cca.cn/.

[3] 国家邮政业安全中心官网. 国家邮政局关于 2021 年 12 月邮政业用户申诉情况的通告[EB/OL]. (2022-01-30)[2022-07-08]. http://www.yzyaqzx.cn/.
The Official Website of the State Post Bureau Safety Supervision Center. Notice of the State Post Bureau on the Appeal of the Users of the Post Industry in December 2021[EB/OL]. (2022-01-30)[2022-07-08]. http://www.yzyaqzx.cn/.

[4] 李立鹏, 田园, 董志祥, 等. 基于 MCU 的物流状态监测装置的研究与设计[J]. 天津科技, 2017, 44(2): 62-65.
LI Li-peng, TIAN Yuan, DONG Zhi-xiang, et al. Research and Design of Logistics Condition Monitoring Device Based on MCU[J]. Tianjin Science & Technology, 2017, 44(2): 62-65.

[5] WANG F, LOU W, GUO M, LU Y. Intelligent Logistics Monitoring Microsystem Based on STM32[J]. Key Engineering Materials, 2015, 645/646: 896-899.

[6] 张小红, 罗科干, 陶贤露, 等. 一种基于穿戴式 MEMS 传感器状态识别的多部位 PDR 算法[J]. 武汉大学学报(信息科学版), 2021, 46(12): 1791-1801.
ZHANG Xiao-hong, LUO Ke-gan, TAO Xian-lu, et al. A Multi-position PDR Algorithm Based on Wearable MEMS Sensors State Recognition[J]. Geomatics and Information Science of Wuhan University, 2021, 46(12): 1791-1801.

[7] 王以忠, 王明银, 张锐, 等. 用于物流运输过程监测的振动测量系统[J]. 现代电子技术, 2011, 34(13): 133-135.
WANG Yi-zhong, WANG Ming-yin, ZHANG Rui, et al. Vibration Measuring System for Monitoring Transportation Process in Logistics[J]. Modern Electronics Technique, 2011, 34(13): 133-135.

[8] CHUANG C H, LEE D H, CHANG W J, et al. Real-time Monitoring via Patch-type Piezoelectric Force Sensors for Internet of Things Based Logistics[J]. IEEE Sensors Journal, 2017, 17(8): 2498-2506.

[9] ANG H, YAN F, LIU T, et al. Design of Multi-sensor Monitoring System for Logistics Yard[J]. Wireless Personal Communications, 2017, 97: 5483-5494.

[10] 丁奥, 张媛, 朱磊, 等. 基于加速度分布特征的快递暴力分拣识别方法[J]. 包装工程, 2020, 41(23): 162-171.
DING Ao, ZHANG Yuan, ZHU Lei, et al. Recognition Method for Rough Handling of Express Parcels Based on Acceleration Distribution Features[J]. Packaging Engineering, 2020, 41(23): 162-171.

[11] 吴蓬勃, 张金燕, 王帆, 等. 快递暴力分拣行为视觉识别系统[J]. 包装工程, 2021, 42(15): 245-252.
WU Peng-bo, ZHANG Jin-yan, WANG Fan, et al. Visual Recognition System of Violent Sorting Behavior in Express Delivery[J]. Packaging Engineering, 2021, 42(15): 245-252.

[12] 尚淑玲. 基于计算机视觉的物流暴力分拣行为识别[J]. 计算机仿真, 2013, 30(12): 430-433.
SHANG Shu-ling. Logistics Sorting Violence Behavior Recognition Based on Computer Vision[J]. Computer Simulation, 2013, 30(12): 430-433.

[13] 邓秀琴, 何鹏志, 倪卫红, 等. 一种基于计算机视觉的暴力分拣行为识别方法[J]. 供应链管理, 2021, 2(6): 109-116.
DENG Xiu-qin, HE Peng-zhi, NI Wei-hong, et al. An Identification Method of Violent Sorting Behavior in the Express Delivery Industry Based on Computer Vision[J]. Supply Chain Management, 2021, 2(6): 109-116.

[14] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6527.

[15] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition, 2017: 6517-652.

[16] REDMON J, FARHADI A. Yolov3: An Incremental Improvement[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[18] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More Features from cheap operations[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 658-666.

[19] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[20] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[21] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.

[22] ZHENG Zhao-hui, WANG Ping, LIU Wei, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI conference on artificial intelligence, 2020, 34(7): 12993-13000.

[23] BIGC229. Beijing Institute of Graphic Communication[EB/OL].(2017-06-08)[2022-07-08].https://github.com/BIGC-229/.

[24] GE Zheng, LIU Song-tao, WANG Feng, et al. YOLOX: Exceeding YOLO Series in 2021[EB/OL]. 2021: arXiv: 2107.08430. https://arxiv.org/abs/2107.08430

[25] HE Kai-ming, ZHANG Xiang-yu, REN Shao-qing, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[26] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2016: 21-37.

[27] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint Triplets for Object Detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.

责任编辑：曾钰婵