基于 YOLO-CARAFE 的人员异常行为识别方法

李　嘎[1],加云岗[1],王志晓[2],张九龙[2],闫文耀[3],高　昂[4],薛　尧[5]
(1. 西安工程大学 计算机科学学院,陕西 西安 710600;
2. 西安理工大学 计算机科学与工程学院,陕西 西安 710048;
3. 延安大学西安创新学院,陕西 西安 710100;
4. 国家卫星气象中心,北京 100081;
5. 西安交通大学,陕西 西安 710049)

摘　要:智能监控中,由于存在环境复杂、监控目标多、画质质量差、人员尺寸不同等因素,从而给人体异常行为识别带来很多挑战。为了提高视频中人员异常行为识别的准确率和识别效率,提出了人员异常行为识别方法 YOLO-CARAFE。该方法首先利用轻量级上采样算子 CARAFE 代替最近邻插值上采样算子,CARAFE 不仅利用相邻像素进行工作,还会对相邻像素进行加权融合,可以在大感受野中聚合上下文信息,从而提高在复杂场景下人体异常行为识别时神经网络的特征提取和融合能力;其次,利用 Focal-EIOU 损失函数的难易样本学习策略,使得模型更加关注难以分类的目标对象,有效减小预测框与真实框之间的差异,提高人体异常行为识别的准确度,有效解决异常行为样本数据量少的问题。通过在自建数据集上的实验表明,YOLO-CARAFE 在人体异常行为识别上具有良好的识别效果,提出的 YOLO-CARAFE 算法在 R 不变的情况下 mAP@0.5, P 分别为 96.9%,97.6%,提高了 1.9 百分点,7.4 百分点,能够满足监控视频中人员异常行为识别对于准确度的需求。

关键词:YOLOV7;行为识别;损失函数;CARAFE;深度学习
中图分类号:TP391　　**文献标识码**:A　　**文章编号**:1673-629X(2024)06-0185-07
doi:10.20165/j.cnki.ISSN1673-629X.2024.0093

Human Abnormal Behavior Recognition Method Based on YOLO-CARAFE

LI Ga[1],JIA Yun-gang[1],WANG Zhi-xiao[2],ZHANG Jiu-long[2],YAN Wen-yao[3],GAO Ang[4],XUE Yao[5]
(1. School of Computer Science,Xi'an Polytechnic University,Xi'an 710600,China;
2. School of Computer Science and Engineering,Xi'an University of Technology,Xi'an 710048,China;
3. Xi'an Innovation College,Yan'an University,Xi'an 710100,China;
4. National Satellite Meteorological Center,Beijing 100081,China;
5. Xi'an Jiaotong University,Xi'an 710049,China)

Abstract:In intelligent monitoring,there are many factors such as complex environment,multiple monitoring targets,poor picture quality,and different personnel sizes,which bring many challenges to human abnormal behavior recognition. In order to improve the accuracy and efficiency of human abnormal behavior recognition in video,we propose YOLO-CARAFE for human abnormal behavior recognition. In this method,CARAFE,a lightweight up-sampling operator,is first used to replace the nearest neighbor interpolation up-sampling operator. CARAFE not only works with adjacent pixels,but also performs weighted fusion of adjacent pixels,which can aggregate context information in the large sensing field,thereby improving the feature extraction and fusion capability of neural networks in recognizing human abnormal behaviors in complex scenes. Secondly,using the difficulty sample learning strategy of Focal-EIOU loss function,the model pays more attention to the target objects that are difficult to classify,effectively reduces the difference between the

收稿日期:2023-10-12　　**修回日期**:2024-02-19
基金项目:陕西省重点研发区域创新引领计划(2022QFY01-17);教育部人文社会科学研究青年基金(16YJCZH109);西安市科技计划(2023JH-RGZNGG-0011);陕西省科技厅重点研发计划项目(2023-YBGY-217);咸阳市科技局"揭榜挂帅"科技项目(L2022-JBGS-GY-07)
作者简介:李　嘎(1995-),女,硕士研究生,研究方向为深度学习、行为识别;加云岗(1966-),男,高级工程师,研究方向为嵌入式、计算机网络等;通讯作者:王志晓(1976-),男,博士,副教授,CCF 会员(42177M),研究方向为大数据分析与理解等。

prediction frame and the real frame, improves the accuracy of human abnormal behavior identification, and effectively solves the characteristics of small amount of abnormal behavior sample data. Experiments on self-built data sets show that YOLO-CARAFE has a good recognition effect on human abnormal behavior recognition. When R is unchanged, mAP@0.5 and P of the proposed YOLO-CARAFE algorithm are 96.9% and 97.6% respectively, increasing by 1.9 percentage points and 7.4 percentage points. It can meet the accuracy and real-time requirements of abnormal behavior identification in surveillance video.

Key words: YOLOV7; behavior recognition; loss function; CARAFE; deep learning

0 引 言

随着经济社会的发展,在公共区域以及家庭私人住宅区域,时有异常行为的发生,然而普通监控视频只是以存储记录的方式进行监控,无法做到实时报警提示异常的发生,具有滞后性[1]。并且对于人体异常行为的识别也会受到背景复杂度、光线明亮、摄像头视角以及遮挡等原因的密切影响[2],因此需要根据异常行为的类别研究出特定场景下的异常行为识别方法。目前对于异常行为的识别主要有基于传统手工提取特征的方法,基于穿戴式[3]、地毯式传感器的方法,基于深度学习的方法[4]。传统人员异常行为识别方法主要有基于局部特征提取的时空兴趣点、轨迹[5]以及基于全局特征提取的人体骨骼[6]和光流等方法,但基于手工提取特征的方法泛化性较弱,无法有效捕捉目标的动作信息。基于穿戴式和传感式的方法限定了人体异常行为识别的范围并且需要大量的人力、物力去采集所需数据集,随着监控摄像头的普及,覆盖范围越来越宽泛,可以适用于复杂环境。

随着深度学习的兴起,神经网络在分类领域取得了很好的效果,其精度大大提升,准确率更高,实时性更强。张红民[7]利用残差网络构建更深的特征提取网络,在网络中融入注意力机制,提高了人体异常行为的检测能力,但是不利于不同尺寸目标的识别;王雪[8]将从 YOLOV5 上提取的特征送入时间注意块,赋予各个特征不同的贡献值,用 LSTM 搭建的预测网络对历史特征进行解码,以达到预测当前行为特征识别异常行为的目的,但是对于小目标识别准确度不高。虽然上述方法在人体异常行为识别上具有重要的意义,但是在异常行为数据量少,异常行为背景复杂的现状下仍然存在不足之处。

针对基于监控视频的人体异常行为识别方法[9],为了提高人体异常行为识别的准确度,解决人体异常行为场景泛化性差和数据样本少的问题,文中实验基于 YOLOV7[10]算法,相较于以往的 YOLO 系列算法,可以达到更高的检测精度。针对特征融合网络和预测网络加以改进,在特征融合网络将最近邻上采样算子改为 CARAFE 轻量级上采样算子[11],提高在复杂环境下对于特征的提取与融合能力,提高检测性能;同时对损失函数进行改进,将 CIOU 损失函数[12]改为 Focal-EIOU 损失函数[13],将 Focal loss 损失函数与 EIOU 损失函数结合,加大难学习样本的权重,加强网络模型对于困难样本的关注,使真实框和预测框的宽高差异最小化,从而获得更加精确的检测框,提高网络检测能力。

1 相关工作

YOLOV7 是一阶段目标检测算法,是在 YOLO 系列算法[14-16]的基础上改进过来的,其检测速度和检测准确率都高于已有的目标检测算法。YOLOV7 分为输入侧、主干网络、输出侧。YOLOV7 的网络结构如图 1 所示。

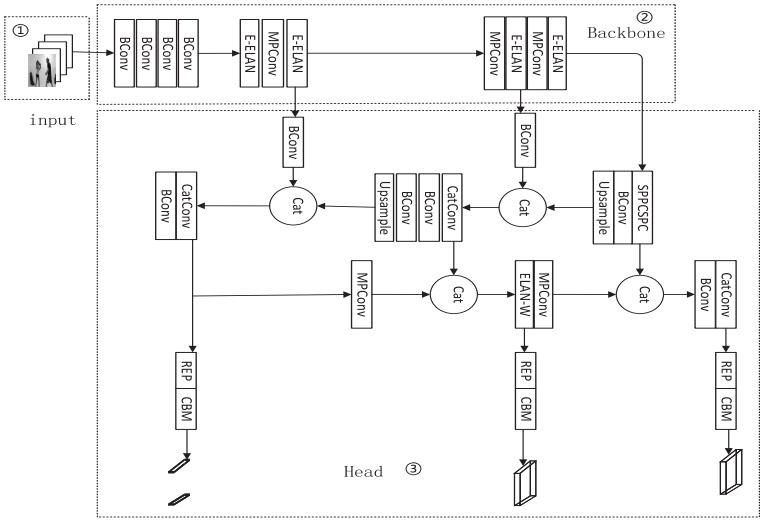

图 1 YOLOV7 结构

YOLOV7 输入侧使用了 Mosaic、Mixup[17] 等数据增强方法来通过马赛克技术拼接多张图片或者将多张图像按比例混合从而增加难例扩充数据集,提高模型的泛化力,使图片满足特征提取网络的需求。用自适应锚框计算策略对不同尺寸的目标进行自适应锚框的计算来获得最佳的锚框值。

Backbone 结构主要作为算法模型的特征提取器和特征融合器。如图 1②所示,YOLOV7 主要采用高效聚合网络,在 CSPNet[18] 网络基础上添加 Catconv 和 MPConv 结构,Catconv 通过高效长程注意力网络来控制梯度的最短最长路径,使得更深的网络也可以更加有效地学习和收敛,增加模型对特征的提取融合能力。MPConv 模型通过两个分支进行下采样,第一个分支先进行最大池化下采样,然后经过 1 * 1 的卷积改变通道数;第二个分支先经过 1 * 1 的卷积改变通道数,再经过 3 * 3 的卷积进行下采样,最后将两个张量拼接在一起,得到下采样结果。

YOLOV7 中的 Head 结构使用了正负样本分配策略[19] 来分配正样本,提供了更加精确的先验知识,主要用于预测分类。整个 Head 层主要包含 SPPCSPC 模块、BConv 模块、Catconv 模块以及 Rep VGG block 模块。SPPCSPC 模块在 SPP 模块基础上添加了 Catconv 模块,与 SPP 模块之前的特征图进行融合,更加丰富了特征信息,可以将高层语义特征与底层语义特征进行融合,进行多尺度特征的独立预测,获得 4 种不同的感受野,用以区分大小目标。在 SPP 模块处理完之后将另一部分进行常规 BConv 处理的张量拼接在一起。在经过特征提取层提取特征后,分别通过 3 个 REP 和 Conv 层输出 3 个不同尺寸大小的预测结果。

2　YOLO-CARAFE 方法

该文基于 YOLOV7 算法进行改进,提出了一种检测精度高的人员异常行为识别模型,进而实现正常行为和异常行为的识别任务。在人体异常行为识别中,可能存在背景噪声、遮挡、光照变化等因素的影响,并且人体异常行为种类繁多,各个种类的异常行为出现的频率也有所差异。因此,需要提取更高层的上下文语义信息来解决异常行为背景复杂的问题,通过改进损失函数的方式来增加权重损失,从而增加难学习样本的权重,解决异常行为数据量少的问题。实验的改进点主要在训练过程和 Head 部分。在训练过程中,针对 YOLOV7 的 CIOU 损失函数会阻碍模型减小真实框的宽高和预测框宽高差异的问题,将 CIOU 损失函数改为 Focal-EIOU 损失函数,加快收敛效果,提高检测精度,解决数据集类别不平衡的问题。对于 YOLOV7 上采样部分的最近邻插值上采样无法有效提取上下文信息,减少环境干扰的问题,将 YOLOV7 的 Head 部分最近邻插值上采样替换为轻量级上采样算子 CARAFE。CARAFE 可以更好地融合高层语义特征,使得网络学习更加细化的特征,解决最近邻插值算法对于特征的上下文语义信息提取不足的问题。YOLO-CARAFE 的网络结构如图 2 所示。

当输入 640 * 640 * 3 的图片时,Backbone 部分的 E-ELAN 经过 BConv 的 1 * 1 卷积分别输出 80 * 80 * 128 和 40 * 40 * 256,与经过 CARAFE 上采样的部分进行 Catconv 操作,提高特征融合能力,最后输出预测结果。在训练部分将 CIOU 损失函数改为 Focal EIOU 损失函数,缩小检测框与预测框之间的差距,提高锚框检测的精度。

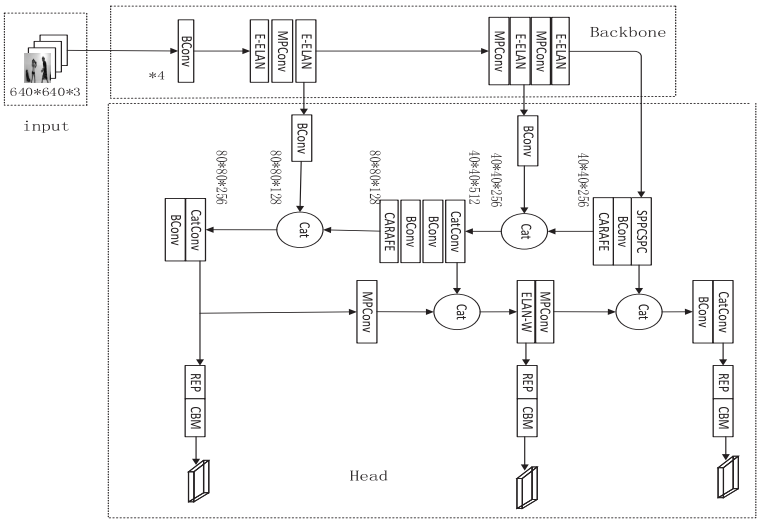

图 2　YOLO-CARAFE 结构

鉴于轻量级上采样算子 CARAFE 和 Focal-EIOU 损失函数的特性,为提高网络对于异常行为识别的精确度,通过对 YOLOV7 网络模型的分析,提出了一种人员异常行为识别方法(YOLO-CARAFE)。

2.1　轻量级上采样算子 CARAFE 改进

在 YOLOV7 中经过 Backbone 部分提取特征之

后,需要将提取到的高层语义特征进行放大然后融合。Backbone 提取的特征经过 SPPCSPC 模块的多次最大池化操作进行下采样,用来减小神经网络特征图的尺寸,再与正常卷积处理操作的特征信息进行 Catconv 拼接,经过一个 CBS 模块用 1 * 1 的卷积来改变通道数。然后在上采样部分将特征图扩大一倍,提高图像的分辨率,通道数不变。

YOLOV7 采用的是最近邻插值方法进行上采样,最近邻插值采用像素之间的最相邻距离来表示上采样,即将变换后像素的灰度值等于距它最近的原始特征图像素的灰度值,但是这种方法仅仅考虑了相邻像素之间的关系,而忽略了特征图的上下文语义信息,这就导致在计算最近邻值的时候在原图像的同一点上生成的图像像素灰度值不同,图像出现锯齿状的边缘,造成图像的细节信息丢失,缺乏细粒度的特征表达能力。而且它的感知域也很小,没有考虑到目标尺寸的变化,导致上采样之后的特征图对于目标的尺寸变化较为敏感。因此,在上采样部分用轻量级上采样算子 CARAFE 来替代原先的最近邻插值上采样,可以提高特征融合能力,获得更加丰富的上下文细节信息。

CARAFE 不仅仅只利用相邻像素进行工作,它会对周围像素进行加权融合,减少上采样过程中的锯齿效应,还可以在一个大感受野中聚合上下文信息,从而获得丰富的上下文语义信息。CARAFE 具有内容感知处理模块,可以动态地生成自适应内核,进而对局部区域特征进行重新组装融合,灵活适应不同目标尺寸变化,可以很好地弥补最近邻插值的缺点。CARAFE 可以提供更加丰富的细节特征,对于细粒度的特征具有更好的感知能力,可以提高目标检测边框和定位的精度。

YOLOV7 将最近邻插值上采样改为轻量级上采样算子 CARAFE 图像如图 3 所示。

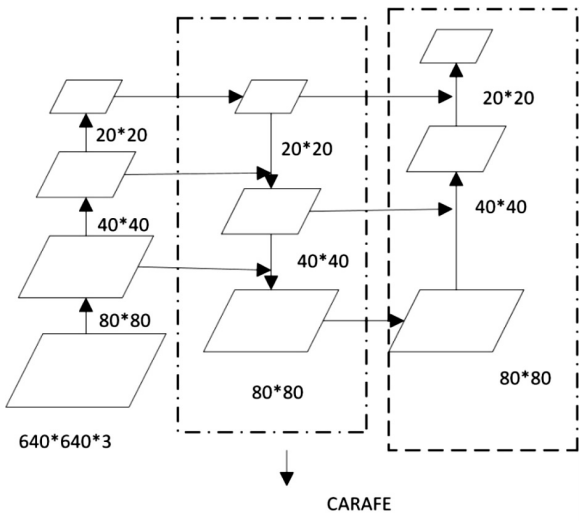

图 3 CARAFE 特征融合结构

改进后的 CARAFE 上采样扩大了感受野,从而可以获取更多的上下文细节信息,提高了检测精度。

2.2 损失函数的改进

分类损失、定位损失和置信度损失组成了目标检测任务的损失函数。损失函数是用来衡量模型预测效果的模块,旨在通过优化预测边界框与真实边界框之间的交并比获得更好的检测效果。损失函数可以根据训练结果反向传播调整神经网络的权重,最小化损失函数,加快网络的收敛速度。YOLOV7 的损失函数是由 CIOU 损失函数构成的,从公式上看,CIOU 损失函数考虑到了预测框与真实框重叠的面积、预测框和真实框中心点的距离以及长宽比的损失这几个参数。CIOU 的公式如公式 1 ~ 3。

$$L_{\mathrm{CIOULoss}} = 1 - \mathrm{IOU} + \frac{r^2(b,b^{gt})}{c^2} + av \tag{1}$$

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

$$\alpha = \frac{v}{(1 - \mathrm{IOU}) + v} \tag{3}$$

v 关于 w 和 h 的梯度计算如下:

$$\frac{\partial v}{\partial w} = \frac{8}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right) * \frac{h}{w^2 + h^2} \tag{4}$$

$$\frac{\partial v}{\partial h} = -\frac{8}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right) * \frac{w}{w^2 + h^2} \tag{5}$$

式中,IOU 表示检测物体预测框与真实框的交并比,参数 $\rho^2(b,b^{gt})$ 表示预测框与真实框中心点的距离,c 为预测框与真实框最小外接封闭区域的对角线距离,参数 w^{gt} ,h^{gt} ,w ,h 分别表示检测物体真实框的宽、高和预测框的宽、高。

由公式 4 和公式 5 可得:

$$\frac{\partial v}{\partial w} = -\frac{h}{w}\frac{\partial v}{\partial h} \tag{6}$$

由式 6 可得 w 与 h 是逆相关的关系,即变量 w 和 h 一个增加一个就减少,这就导致了模型无法有效地减小预测框以及真实框宽高之间的差异。针对此问题,该文用 Focal-EIOU 损失函数替代 CIOU 损失函数,Focal-EIOU 损失函数公式为:

$$L_{\mathrm{EIOU}} = L_{\mathrm{IOU}} + L_{\mathrm{dis}} + L_{\mathrm{asp}} = 1 - \mathrm{IOU} + \frac{\rho^2(b,b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w,w^{gt})}{(w^c)^2} + \frac{\rho^2(h,h^{gt})}{(h^c)^2} \tag{7}$$

$$L_{\mathrm{FocalE-IOU}} = \mathrm{IOU}^{\gamma} L_{\mathrm{EIOU}} \tag{8}$$

Focal-EIOU 损失函数由 IOU 损失 L_{IOU} 、距离损失 L_{dis} 、方位损失 L_{asp} 组成。引入 Focal loss 优化了边界框回归中的样本不平衡问题,即减少了与目标框重叠度较低的锚框,使回归过程专注于与目标框重叠较多

的锚框。Focal-EIOU 损失函数具有难易样本学习策略,更加关注难以分类的目标对象。Focal-EIOU 损失函数在保留 CIOU 损失函数优点的同时又可以使真实框和预测框的宽高差异最小化,使收敛速度更快。在异常行为识别中会存在遮挡、噪声等因素,会产生一些异常预测框,对于行为的识别造成困难,Focal-EIOU 损失函数可以消除异常框对于行为识别造成的干扰。

3　实验分析

3.1　实验环境

实验采用云服务器作为 YOLOV7 的训练机器以及 YOLOV7 的检测机器。采用的服务器 GPU 型号为 NVIDIA RTX 3080Ti,显存 12G,内存大小 15G,使用 Python 语言,Pytorch2.0.1 框架,YOLOV7 迭代次数为 100,batchsize 为 8。

3.2　实验数据集预处理

数据集通过处理公共异常行为视频数据集 Fighting UCF-Crime、HMDB、Violence LAD、UBI-Fights 和从网络爬取的方式来获取,自制人体异常行为数据集。该数据集总共分为 6 种类别,总计约 3 000 张图片,按照 6∶2∶2 的比例划分训练集、验证集、测试集。将踢腿和拳击作为人体异常行为,将走路、跑步、站立、坐等作为人体正常行为。从视频数据集中获取这些人体行为的视频,进而通过筛选得到高质量的人体行为视频数据。

用 ffmpeg 从视频中截取视频帧图片,将模糊以及噪声大的图像剔除掉,并将这些数据图片进行重命名。用 labelimg 来标注图片,获取到人体异常行为的坐标值,实现人体异常行为的精确定位。在标注过程中将踢腿、拳击标为异常行为,将行走、坐等动作标注为正常动作。数据集划分为两类别以及 6 种不同的人体行为,在获取到的 txt 文本中,“1”代表正常行为,“2”代表异常行为,其余的 4 个数字分别代表标注框中心点的 x 轴、标注框中心点的 y 轴、图片的宽度、图片的高度,获取到的值均为归一化的值。

对于人体动作的精确划分有利于更好地区分以及识别各类人体行为。从这些公共数据集中获取到的图片包含不同的场景、不同身高体型的人,同时场景中存在遮挡、光线变化以及人员尺寸不同的情景,因此训练的人体异常行为识别模型具有泛化性,可以满足人体异常行为在特定场景下识别的需求。

3.3　评价指标

在神经网络模型中,为了验证 YOLO-CARAFE 算法对于人员异常行为识别的精确度,该文选取了 mAP(平均准确率)、P(精确率)、R(查全率)作为评价指标。P 表示在所有识别出来的异常行为图像中,该图像确实是人体异常行为图像所占的比率。P 越高,表示预测为异常行为的类别中是真正的异常行为的比例越高。

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

mAP 值指的是平均精度均值,即 AP 的平均值。

$$\mathrm{AP} = \int_0^1 P \mathrm{d}R \tag{11}$$

$$\mathrm{mAP} = \frac{1}{c}\sum_{i=1}^{c} \mathrm{AP}_i \tag{12}$$

式中,c 为检测图像类别数,i 为检测次数。AP 代表 P、R 曲线下面积,为一个类别的识别平均准确率,mAP@0.5 为所有类别的平均识别准确率相加之后取平均值。mAP 值越高,算法性能越好。

R(查全率)指的是所有异常行为被正确预测的比例。R 越高说明模型对于人体异常行为的区分能力越好。

其中,FN 指的是预测错误,将人体异常行为预测为正常行为所占的比例;FP 指的是预测错误,将人体正常行为预测为异常行为所占的比例;TP 指的是预测正确,将人体异常行为预测为异常行为所占的比例。

置信度指的是用来判断边界框内的物体是正样本时,模型认为它属于哪一个类别的概率。置信度越高,说明模型认为这个物体是某一类对象的概率越高。

3.4　模型异常行为识别结果分析

3.4.1　对比实验

使用相同数据集对 YOLOV3、YOLOV5 进行对比实验,并与改进的 YOLOV5 算法以及 C2D-YOLO 神经网络模型进行对比验证,对比得到各个算法模型的评价指标如表 1 所示。

表 1　模型对比结果

算法模型	mAP@0.5/%	P	R
YOLOV3	0.880	0.88	0.86
YOLOV5	0.903	0.904	0.98
改进的 YOLOV5L[8]	0.658	/	/
改进的 YOLOV5[20]	0.917	/	/
YOLOV3-MSSE[7]	0.781	0.899	0.794
YOLOV7	0.950	0.902	1
YOLO-CARAFE	0.969	0.976	1

由实验数据可知,YOLO-CARAFE 网络相较于 YOLOV5、YOLOV3、改进的 YOLOV5 网络和 C2D-YOLO 在检测平均精度、精确率、召回率上提升了很多,YOLOV7 的检测精确率相较于 YOLOV5 略低,但是模型效果差距较小,可能因为人体异常行为相似性

较高，存在误判，但 YOLOV7 的 mAP、*R* 高于 YOLOV5，YOLOV7 具有很大的改进潜力，因此该文选用 YOLOV7 网络进行改进。实验结果表明，对于人体异常行为的识别问题，改进 YOLOV7 神经网络可以取得更好的效果，并且该算法也适用于人员异常行为的识别。

3.4.2 消融实验

为验证每个改进点对于神经网络模型的提升效果，进行了 4 组消融实验。实验在相同环境下进行，消融实验的结果见表 2，由表可得：

(1)第一组实验为改进前 YOLOV7 算法实验结果，同时作为后面各组实验的对比对象，检测 mAP 值为 95.0%，*P* 为 90.2%，*R* 为 100%。

(2)第二组实验为将最近邻插值上采样替换为 CARAFE 上采样算子，检测 mAP 提高了 1.5 百分点，*P* 提高了 7 百分点，*R* 为 99%，下降了 0.01 百分点，可能因为某一类异常行为与正常行为差异较小，存在误判。

(3)第三组实验为将 CIOU 损失函数改为 Focal-EIOU 损失函数，检测 mAP 提高了 1.4 百分点，*P* 提高了 6.5 百分点，*R* 为 99%，下降了 0.01 百分点，可能因为某一类异常行为与正常行为差异较小，存在误判。

(4)第四组实验为将 CIOU 损失函数改为 Focal-EIOU 损失函数，将最近邻插值上采样替换为 CARAFE 上采样算子，检测 mAP 提升了 1.9 百分点，*P* 提高了 7.4 百分点，*R* 保持不变。

表 2 消融实验

算法模型	mAP@0.5/%	*P*	*R*
YOLOV7	0.950	0.902	1
YOLOV7+CARAFE	0.965	0.972	0.99
YOLOV7+Focal-EIOU loss	0.964	0.967	0.99
YOLO-CARAFE	0.969	0.976	1

如表 2 所示，可以得出同时改进上采样算子 CARAFE 与改进 Focal-EIOU 损失函数的算法模型在 mAP，*P*，*R* 值上都高于仅改进上采样算子与仅改进 Focal-EIOU 损失函数的算法模型，*R* 值稳定，模型能够正确识别异常行为，具有稳定性，说明 YOLO-CARAFE 可以在异常行为数据集上很好地提高模型的行为识别能力。

3.4.3 实验结果分析

实验在自制异常行为数据集上进行，改进后模型与原模型 mAP 值对比如图 4 所示。可以看到改进后的 YOLOV7 算法模型的 mAP 值相较于原 YOLOV7 的 mAP 值提升了，并且改进后的模型收敛速度更快。图 5 为原模型与改进后模型 Loss 值的对比。可以看到改进后的 YOLOV7 模型使得损失值更低，但没有影响到模型最终的收敛情况，Loss 值逐渐趋于稳定，说明改进后的模型可以更好地定位和分类识别对象。

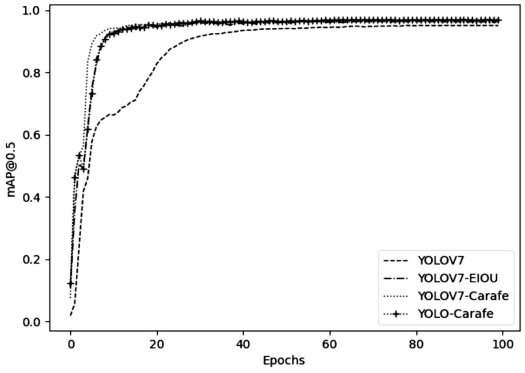

图 4 改进后模型与原模型 mAP 值对比

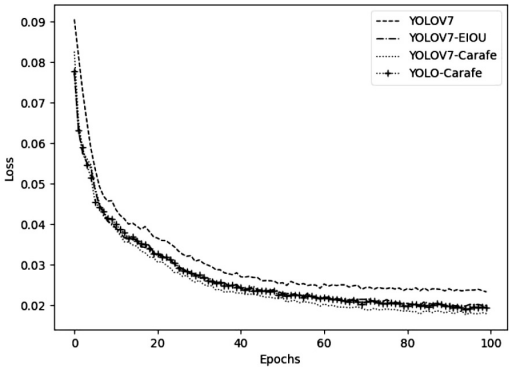

图 5 改进后模型与原模型 Loss 对比

3.4.4 检测结果

随机选用 2 张检测图片进行算法模型效果的检测，如图 6 所示，对比图为 YOLOV7 神经网络对于人体异常行为的识别效果以及改进后的 YOLO-CARAFE 模型对于人体异常行为的识别效果。

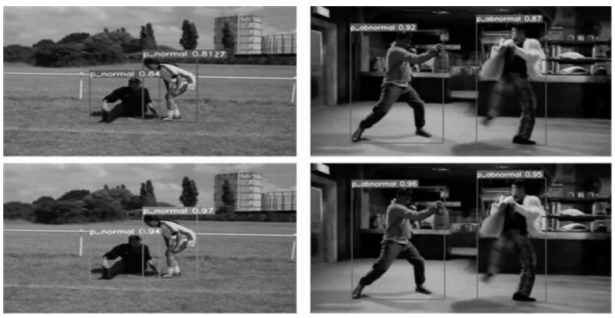

图 6 模型对于人体行为的检测效果

由图 6 可以看到，第一行为未改进 YOLOV7 的识别结果，第二行为改进后的 YOLO-CAREFE 神经网络的检测结果。改进后的 YOLOV7 对于人体异常行为的识别精度较高，置信度高于未改进 YOLOV7 网络。根据实验，原模型和改进后的模型均能很好地检测出人体异常行为。表 3 列出了图片检测置信度，可以得出改进后 YOLOV7 的置信度最高，依次为 0.97 和 0.95，高于 YOLOV7 算法的识别效果，改进后模型的

检测置信度有所提高。

表 3　检测图片置信度

Confidence	YOLOV7	YOLOV7-CARAFE
Picture(a)	0.81	0.97
Picture(b)	0.87	0.95

4　结束语

YOLO-CARAFE 方法通过改变上采样方式来加强图像的特征融合和提取能力,提高了模型对于人体异常行为检测的精度。同时,又改进了 YOLOV7 网络的 CIOU 损失函数,使得获得的预测框更加精确,解决了人员异常行为识别中可能出现的遮挡等因素对于检测框的定位所造成的困难。使得网络模型提升了检测的准确度,对人员异常行为的识别具有积极的意义。

但是,YOLO-CARAFE 模型仍然存在不足之处。(1)实验中所用的数据集少,使得神经网络对于人员异常行为的训练不是很充分,应该扩大数据集,增加在不同场景中包含不同的人体行为动作的数据集数量;(2)该模型对于人员异常行为的识别仍然存在误判,有待进一步完善。

参考文献:

[1] MA N,WU Z,CHEUNG Y,et al. A survey of human action recognition and posture prediction[J]. Tsinghua Science and Technology,2022,27(6):973-1001.

[2] 施新凯,张雅丽,李御瑾,等. 基于 YOLOv4 改进算法的人群异常行为检测研究[J]. 现代计算机,2022,28(7):29-34.

[3] 郭毅博,孟文化,范一鸣,等. 基于可穿戴传感器数据的人体行为识别数据特征提取方法[J]. 计算机辅助设计与图形学学报,2021,33(8):1246-1253.

[4] 康　帅,章坚武,朱尊杰,等. 改进 YOLOv4 算法的复杂视觉场景行人检测方法[J]. 电信科学,2021,37(8):46-56.

[5] HUYNH M,ALAGHBAND G. GPRAR:graph convolutional network based pose reconstruction and action recognition for human trajectory prediction[J]. arXiv:2103.14113,2021.

[6] QIU H,HOU B,REN B,et al. Spatio-temporal tuples transformer for skeleton-based action recognition[J]. arXiv:2201.02849,2022.

[7] 张红民,李萍萍,房晓冰,等. 改进 YOLOV3 网络模型的人体异常行为检测方法[J]. 计算机科学,2022,49(4):233-238.

[8] 王　雪,程换新,骆晓玲,等. 基于改进的 YOLOv5 网络的异常行为检测算法研究[J]. 电子测量技术,2022,45(16):137-141.

[9] 詹健浩,吴鸿伟,周成祖,等. 基于深度学习的行为识别多模态融合方法综述[J]. 计算机系统应用,2023,32(1):41-49.

[10] WANG C Y,BOCHKOVSKIY A,LIAO H Y M. YOLOv7:trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Vancouver:IEEE 2023:7464-7475.

[11] WANG J,CHEN K,XU R,et al. Carafe:content-aware reassembly of features[C]//Proceedings of the IEEE/CVF international conference on computer vision. California:IEEE 2019:3007-3016.

[12] ZHENG Z,WANG P,LIU W,et al. Distance-IoU loss:faster and better learning for bounding box regression[J]. arXiv:1911.08287v1,2019.

[13] ZHANG Y F,REN W,ZHANG Z,et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing,2022,506:146-157.

[14] REDMON J,FARHADI A. YOLOv3:an incremental improvement[J]. arXiv:1804.02767.2018.

[15] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. YOLOv4:optimal speed and accuracy of object detection[J]. arXiv:2004.10934,2020.

[16] LI C,LI L,JIANG H,et al. YOLOv6:a single-stage object detection framework for industrial applications[J]. arXiv:2209.02976,2022.

[17] ZHANG H,CISSE M,DAUPHIN Y N,et al. Mixup:beyond empirical risk minimization[J]. arXiv:1710.09412,2017.

[18] WANG C Y,LIAO H Y M,WU Y H,et al. CSPNet:a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. Seattle:IEEE,2020:390-391.

[19] ZHANG S,CHI C,YAO Y,et al. Bridging the gap between anchor-based and anchor-free recognizevia adaptive training sample selection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle:IEEE,2020:9759-9768.

[20] 林宝华,刘　坤,朱一帆,等. 基于 YOLOv5 的工人玩手机行为检测方法研究[J]. 南京工程学院学报:自然科学版,2023,21(1):39-44.