文章编号:1671-6833(2021)01-0028-07

改进 YOLOv3 算法与人体信息数据融合的视频监控检测方法

张 震,李浩方,李孟洲,马军强

(郑州大学 电气工程学院,河南 郑州 450001)

摘 要:针对目前社区视频监控使用人脸相机仅采集出入口人脸数据,而缺失有数据价值的人体其他属性信息的问题,提出一种将改进 YOLOv3 网络和调用人体信息识别模块相结合的人体信息检测方法。采用 *K*-means++算法获取数据集的先验框,选用新的边界框回归损失函数 *GIoU* 提高检测精度,再进行多尺度训练得到人体检测网络模型,最后利用人体检测模型在检测到人体目标后调用人体信息识别模块对人体信息进行分析和保存。实验结果表明:该方法既能快速检测人体目标,还能准确获取人体目标的各种属性信息,其中人体检测模型在测试集上的 *mAP* 达到 91.8%,识别速度为 45 f/s。

关键词:视频监控;*K*-means++; *GIoU*; 多尺度训练; 改进 YOLOv3; 人体信息

中图分类号:TP391.41 **文献标志码**:A **doi**:10.13705/j.issn.1671-6833.2021.01.005

0 引言

随着城市社区的快速发展,社区视频监控管理系统对于提升社区管理效率和保障社区安全起着重要的作用[1]。然而现有社区视频监控仅采用人脸相机在特定的角度对小区出入口抓拍人脸数据,忽略了具有潜在价值的人体其他信息。因此,若能利用人脸相机准确快速地获取其他人体信息,并能与现有视频管理系统对接,就可以提升社区公共安全管理能力和精细化管理水平。

随着深度学习技术在目标检测领域的快速发展,现有目标检测算法可以分为两类。第一类是双阶段模型,该类模型首先用窗口标定算法生成一系列待筛选目标区域,然后通过深度神经网络对目标区域进行特征训练,最后用所训练出来的模型选出最优的目标边界框。主要代表网络有 R-CNN[2]、Fast R-CNN[3]、Faster R-CNN[4]、R-FCN[5]等。这些算法准确率较高,但提取出大量冗余特征,算法比较耗时。第二类是单阶段模型,该类模型不经过窗口预标定,而是直接利用整张图像一次性预测出目标的位置并标价边框,代表网络有 YOLO[6]、SSD[7]等。这些检测算法采用端到端的目标检测,具有检测效率高、原理简单和背景误检率低等特点。其中具有代表性的是 Redmon 等[8]在 2018 年提出的 YOLOv3 检测算法,其在 COCO 数据集上51 ms 内 *mAP* 为 57.9%[9]。

为利用人脸相机准确检测人体信息,笔者先采用 *K*-means++[10]算法获得适应于自制数据集的先验框;再选用一种新的边界框回归损失函数 *GIoU*[11]提高检测性能;然后,使用多尺度方式[12]进行模型训练;最后,通过调用人体属性识别模块实现对人体信息准确检测。经实验验证,该方法在快速检测人体的同时,可以准确识别人体其他属性信息。

1 YOLOv3 原理与分析

1.1 目标检测原理

YOLOv3 算法通过特征提取网络对输入的图像提取特征,得到一定大小的特征图。然后将输入的图片分割成 $S\times S$ 个网格,其中每个网格中预测 B 个边界框,对 C 类目标进行检测。网格中边界框不仅要确定自身位置,还要预测一个置信度,置信度由每个网格中包含检测目标概率和输出边界框准确度共同确定。若预测目标中心落在该网格中,则该网格负责预测目标。整张图像目标位置类别预测如式(1)所示:

$$Y = A \times A \times B \times (5 + C)。 \tag{1}$$

式中:Y 表示图像目标位置类别预测张量;A 表示

收稿日期:2020-06-20;**修订日期**:2020-07-29

基金项目:国家重点研发计划公共安全风险防控与应急技术装备专项(2018YFC0824XXX)

作者简介:张震(1966—),男,河南郑州人,郑州大学教授,博士,博士生导师,主要从事多媒体信息安全、图像处理与模式识别研究,E-mail:zhangzhen66@126.com。

网格数;B 表示边界框数量;5 表示 4 个边框坐标数值和 1 个边框置信度数值;C 表示对象类别。

YOLOv3 不仅借鉴了 FPN[13] 架构,采用 3 个尺度对不同大小的目标进行预测,提升了小物体的检测效果,还采用多个独立的逻辑 logistic 分类器替换 *softmax*[14] 函数,以计算输入属于特定标签的可能性,每个标签使用二元交叉熵损失降低了计算复杂度。

1.2　Darknet-53 网络

YOLOv3 算法采用 Darknet-53 作为主干网络。该网络主要是由一系列的 1×1 和 3×3 卷积层组合而成的,并且每个卷积层后增加了批次归一化层,可以有效防止过拟合现象。其次,网络借鉴 ResNet[15] 残差网络结构,通过残差层实现跨层数据更快地向前传播。最后,网络使用 5 个步长为 2 的 3×3 卷积层替换上代网络中的最大池化层实现下采样。该主干网络在 ImageNet 数据集中进行测试,测试结果如表 1 所示。表中 $A_{\text{Top-1}}$ 和 $A_{\text{Top-5}}$ 分别表示模型在图片识别时前 1 个和前 5 个结果中有一个正确的准确率,计算量表示浮点运算的次数,运算速度是每秒浮点运算的次数,帧速度为每秒刷新图片的帧数。

表 1　特征提取网络

Table 1　Feature extraction network

主干网络	$A_{\text{Top-1}}$/%	$A_{\text{Top-5}}$/%	计算量/次	运算速度/(次·s^{-1})	帧速率/(f·s^{-1})
Darknet-19	74.1	91.8	7.29×10^{9}	1.246×10^{12}	171
ResNet-101	77.1	93.7	1.97×10^{10}	1.039×10^{12}	53
ResNet-152	77.6	93.8	2.94×10^{10}	1.090×10^{12}	37
Darknet-53	77.2	93.8	1.87×10^{10}	1.457×10^{12}	78

由表 1 可知,Darknet-53 相比 Darknet-19[6] 检测的准确率有了进一步的提升,但是运算速度有所降低,与 ResNet-152 的网络性能基本一致,并且目标检测速度提升到 78 f/s,满足目标检测实时性要求。

2　分析与讨论

现有人体目标检测存在人体信息获取不完整和检测速度较慢等问题,因此,为得到更适合人体目标检测的方法并且提升运算速度,对人体目标检测网络重新进行设计,整体检测框架如图 1 所示。

本文使用新的边界框回归损失函数 *GIoU* 和 *K*-means++算法聚类算法,经过多尺度训练得到人体检测模块。视频监控再通过人体检测模块与智能应用平台信息交互,实现人体目标属性信息的检测。

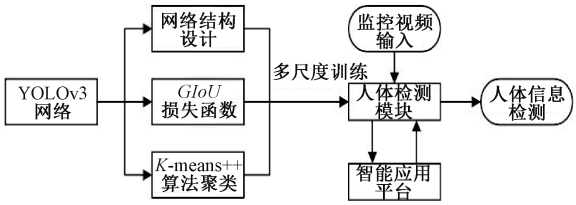

图 1　整体框架图

Figure 1　Overall frame diagram

2.1　网络结构设计

训练深度神经网络模型时,批次归一化利用小批量上的均值和标准差,不断地调整神经网络中间输出,从而使整个神经网络在各层的中间输出的数值更加稳定,同时可以加速网络的收敛并控制过拟合产生。批次归一化虽然可以使神经网络的训练更容易,但是在网络向前推理时会增加一些运算,一定程度上占用了更多显存。因此,将批次归一化层与原有卷积层相整合构建新的卷积层,这样有利于提升模型向前推理的速度。

由于 YOLOv3 算法存在网络层数的加深使模型复杂和特征消失等问题,因此笔者在 YOLOv3 网络结构上适当减少 YOLO 层的卷积次数,得到新的网络结构。改进网络先将输入图像缩放为通道为 3、长和宽均为 416 的统一参数,然后通过 Darknet-53 特征提取网络结构提取特征,对采集到的特征采用 1×1 和 3×3 的卷积进行卷积操作,降低计算量以及融合特征之间的通道数,得到一个小尺度 YOLO 层一个 13×13×255 维的输出量;然后对小尺度 YOLO 层进行上采样,与 Darknet-53 中的卷积第 45 层进行拼接,再进行 2 组 1×1 和 3×3 的卷积进行卷积操作,得到一个中尺度 YOLO 层一个 26×26×255 维的输出量;接着将得到的中尺度 YOLO 层进行上采样,与 Darknet-53 中的卷积第 29 层进行拼接,再进行 2 组 1×1 和 3×3的卷积进行卷积操作,得到一个大尺度 YOLO 层的一个52×52×255维输出量;最后,将已得到的 3 个尺度 YOLO 层进行边界框和类别的预测。改进后的网络一共有 102 层,由 70 层卷积层、23 层残差层、4 层特征层、2 层上采样层和 3 层 YOLO 层组成,改进的网络结构如图 2 所示。

2.2　*GIoU* 边界框回归损失函数

目标检测任务中,预测框与真实框之间的交

	Type	Filters	Size	Input	Output
	卷积层	32	3×3/1	416×416×3	416×416×32
	卷积层	64	3×3/2	416×416×32	208×208×64
	卷积层	32	1×1/1	208×208×64	208×208×32
	卷积层	64	3×3/1	208×208×32	208×208×64
1×	残差			208×208×64	208×208×64
	卷积层	128	3×3/2	208×208×64	104×104×128
	卷积层	64	1×1/1	104×104×128	104×104×64
	卷积层	128	3×3/1	104×104×64	104×104×128
2×	残差			104×104×128	104×104×128
	卷积层	256	3×3/2	104×104×128	52×52×256
	卷积层	128	1×1/1	52×52×256	52×52×128
	卷积层	256	3×3/1	52×52×128	52×52×256
8×	残差			52×52×256	52×52×256
	卷积层	512	3×3/2	52×52×256	26×26×512
	卷积层	256	1×1/1	26×26×512	26×26×256
	卷积层	512	3×3/1	26×26×256	26×26×512
8×	残差			26×26×512	26×26×512
	卷积层	1 024	3×3/2	26×26×512	13×13×1 024
	卷积层	512	1×1/1	13×13×1 024	13×13×512
	卷积层	1 024	3×3/1	13×13×512	13×13×1 024
4×	残差			13×13×1 024	13×13×1 024

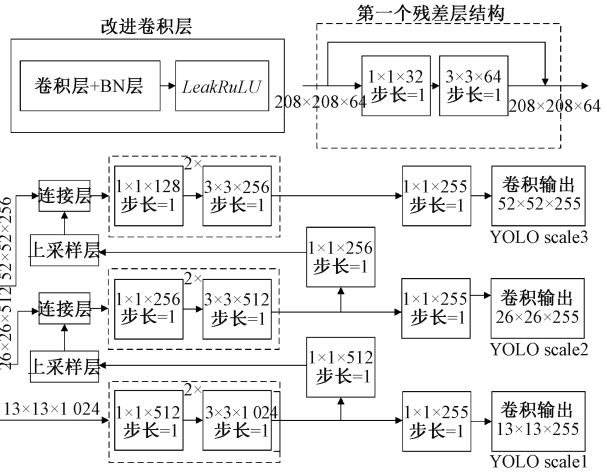

图 2 改进的网络结构

Figure 2 Improved network structure

并比 *IoU* 不仅可以反映预测检测框与真实检测框的检测效果,还是评价网络性能指标的重要参数。*IoU*[16] 定义如下:

$$IoU = \frac{|A \cap B|}{|A \cup B|}。 \tag{2}$$

式中:*A* 为目标的预测框;*B* 为目标的真实框;*IoU* 是真实框和预测框面积的交并比。

现有网络常采用 *IoU* 作为损失函数,遇到轴对齐的二维边界框不相交情况,依据 *IoU* 计算公式,此时 *IoU* 为零,无法进行模型训练。因此 Rezatofighi 等[11] 提出了一种既能维持 *IoU* 尺度不变性,还能在目标重叠时更好地反映预测框和真实框的重合度的评价指标 *GIoU*,其定义公式如下:

$$GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|}。 \tag{3}$$

式中:*A* 为目标的预测框;*B* 为目标的真实框;*C* 为预测框和真实框的最小框面积。

由式(3)可知,*GIoU* 引入了包含 *A*、*B* 两个形状的 *C*,所以当 *A*、*B* 不重合时,依然可以进行边界框回归优化,因此采用 *GIoU* 构造边界框回归损失函数。若已知预测框和真实框的坐标如下:

$$\begin{cases} B^p = (x_1^p, y_1^p, x_2^p, y_2^p), B^g = (x_1^g, y_1^g, x_2^g, y_2^g) \\ \text{且 } x_2^p > x_1^p, y_2^p > y_1^p; \\ \hat{x}_1^p = \min(x_1^p, x_2^p), \hat{x}_2^p = \max(x_1^p, x_2^p); \\ \hat{y}_1^p = \min(y_1^p, y_2^p), \hat{y}_2^p = \max(y_1^p, y_2^p)。 \end{cases} \tag{4}$$

其边界框回归损失函数的计算过程如下。

步骤 1 计算 B^g 的面积 $B^g = (x_2^g - x_1^g) \times (y_2^g - y_1^g)$。

步骤 2 计算 B^p 的面积 $B^p = (x_2^p - x_1^p) \times (y_2^p - y_1^p)$。

步骤 3 计算 B^p 和 B^g 的重叠面积:

$$x_1^I = \max(\hat{x}_1^p, x_1^g), x_2^I = \max(\hat{x}_2^p, x_2^g);$$

$$y_1^I = \max(\hat{y}_1^p, y_1^g), y_2^I = \max(\hat{y}_2^p, y_2^g);$$

$$I = \begin{cases} (x_2^I - x_1^I) \times (y_2^I - y_1^I), x_2^I > x_1^I, y_2^I > y_1^I, \\ 0, \quad \text{其他}。 \end{cases}$$

步骤 4 找到可以包含 B^p 和 B^g 的最小框 B^C:

$$x_1^C = \max(x_1^p, x_1^g), x_2^C = \max(x_2^p, x_2^g),$$

$$y_1^C = \max(y_1^p, y_1^g), y_2^C = \max(y_2^p, y_2^g)。$$

步骤 5 计算 B^C 的面积 $B^C = (x_2^C - x_1^C) \times (y_2^C - y_1^C)$。

步骤 6 计算 *IoU*。$IoU = \frac{I}{U} = \frac{I}{B^p + B^g - I}$。

步骤 7 计算 *GIoU*。$GIoU = IoU - \frac{B^C - U}{B^C}$。

步骤 8 计算损失。$L_{GIoU} = 1 - GIoU$。

2.3 *K*-means++算法聚类

YOLOv3 网络中的 9 个先验框是采用 *K*-means 算法在 COCO 数据集下聚类产生的,不能应用于本文的数据集。并且由于 *K*-means[10] 算法在运算过程中,初始聚类中心是随机产生的,因此存在聚类中心不断变化,导致每次运行获得不同的聚类效果,从而影响模型的检测效果。为解决初始聚类中心不断变化的问题,笔者采用 *K*-means++算法进行先验框的聚类。聚类过程如下。

步骤 1 随机选取数据集中的一个锚定框的宽和高作为第一个聚类中心。

步骤 2 计算数据集中每个锚定框坐标与已知聚类中心的距离,再根据概率重新选择下一个聚类中心。

步骤 3　重复第 2 步的计算,直到选出 K 个聚类中心。

步骤 4　重新计算数据集中的每个锚定框坐标与聚类中心距离,并根据最小距离重新进行分类划分。

步骤 5　计算每个分类的中心值,直到聚类中心的位置不再变化时结束。

对自制数据集重新聚类依次获取的 9 组先验框为(76,23),(81,37),(89,46),(102,61),(113,68),(119,71),(150,83),(164,89)和(170,95)。

2.4　人体目标特征属性识别

海康威视萤石开放平台一方面能提供人体目标属性图像识别技术,并且对分析任务提供同步和异步两种接入方式。同步分析任务的接入方式较为简便,但是整体的性能比较受限,适用于小批量的图片检测;异步分析任务则可以提供更大的分析吞吐量,适用于大批量的任务提交。另外一方面,开放平台对人体目标属性有较成熟的识别方案,且对人体目标属性结构化数据如人脸目标位置、年龄段、性别、是否戴眼镜、是否背包、是否拎东西、发型、上衣类型、下衣类型、上衣颜色、下衣颜色和是否骑车等,有较高的检测速度和识别率。

基于以上考虑,采用开放平台中的异步分析任务方式进行人体目标属性的识别。首先,使用改进的网络配合人脸相机截取到监控区域的人体图像;其次,将截取图像使用基于 ISAPI 协议的图片任务分析接口,提交异步图片分析任务请求中截取图片的发送地址;最后,开放平台进行人体目标属性的分析任务。当平台分析完成后,分析结果会通过 TCP 的方式发送格式为 ISAPI 协议所规定的 JSON 报文;接着使用 Python 将 JSON 报文中人体目标检测属性的信息进行提取和展示;最后存储到自建的 ACCESS 社区人体信息数据库中。调用人体特征识别流程图如图 3 所示。

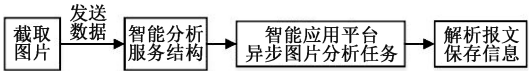

图 3　人体特征调用流程图

Figure 3　Flow chart of human character recognition

3　实验结果与分析

3.1　实验数据集

为获得准确的人体信息,将 3 个数据集中的人体图片整合为标准 PASCAL VOC 数据集格式。其中第一部分采用的是 PASCAL VOC2012 数据集中人体的图片共 4 015 张,标注框数量 5 717 个;第二部分采用的是 COCO 数据集中人体的图片,采用了 2 693 张图片,标注框数量 11 004 个;第三部分数据采用郑州市某视频中心监控视频,采用 labelImg 工具对自制人体数据集进行人工标注,人工标注照片3 292张,标注框数量 4 326 个。总共 10 000 张图片,标注框 21 047 个。

在训练和测试前,先将数据集中图片随机分为 8 000 张训练集和 2 000 张测试集。然后,再把训练集和测试集中照片统一缩放分成 4 个不同的尺寸组。其中 A 组 320×320、B 组 416×416、C 组 512×512 和 D 组 608×608。数据集图片如下图 4 所示。

(1) COCO数据集　(2) VOC2012数据集　(3) 自制人体数据集

图 4　实验数据集

Figure 4　Experimental dataset

3.2　实验环境及参数

本实验在 PC 端完成,实验平台使用操作系统为 Ubuntu 16.04,显卡为 NVIDIA GeForce 2080 Ti。

训练过程中采取多尺度训练的策略,每 10 个批次随机挑选训练集中的一组尺寸进行迭代,共进行 50 000 次迭代。其中学习率为 0.001 4,在迭代到 25 000、40 000 和 45 000 次时,学习率变为之前的十分之一。其中训练参数批量大小(*batch*)为 64,动量参数(*momentum*)为 0.9。

3.3　性能对比分析

为验证改进网络和改进目标损失函数对目标检测算法的影响,笔者采用对比实验进行验证。第一组为 YOLOv3 网络分别采用 *MSE*(*mean squared error*)、*IoU* 和 *GIoU* 边界框回归损失函数进行对比;第二组为均采用 *IoU* 损失函数的改进网络与 YOLOv3 网络进行对比;第三组为改进网络采用 *GIoU* 边界损失函数与 YOLOv3 网络采用 *IoU* 进行对比,对比结果如表 2 所示。

由表 2 可知,第一组 YOLOv3 网络采用 *GIoU* 边界框回归损失函数相比采用 *MSE* 和 *IoU* 作为损失函数,检测准确率 *mAP* 分别提升了 7.1% 和 1.7%,说明采用 *GIoU* 损失函数可以提升网络检

表 2 边界框回归损失函数

Table 2 Bounding-box regression loss function

优化方案	检测准确率 *mAP*/%
YOLOv3+ *MSE*	83.4
YOLOv3+ *IoU*	87.8
YOLOv3+ *GIoU*	89.3
改进网络+ *IoU*	89.7
改进网络+ *GIoU*	91.8

测性能。第二组对比实验中,改进网络的 *mAP* 提升了 0.4%,说明改进网络也能提升网络性能。第三组实验将改进网络采用 *GIoU* 边界框回归损失函数与 YOLOv3 采用 *IoU* 边界框回归损失函数相比,*mAP* 提升了 4.6%。说明改进网络结构与改进边界框回归损失函数可以进一步提升网络的性能。不同网络与不同边界框损失函数训练过程如图 5 所示。

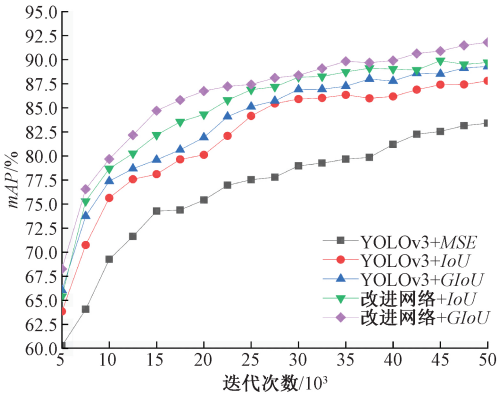

图 5 对比实验变化

Figure 5 Comparative experimental change

改进网络采用 *GIoU* 和 YOLOv3 网络采用 *IoU* 两种边界框回归损失函数对自制数据集的检测效果如图 6 所示。其中蓝色为改进网络采用 *GIoU* 边界框回归损失函数,红色为 YOLOv3 网络采用 *IoU* 损失函数。由图 6 可知,蓝色框不仅检测精度略高于红色框,还能更准确地框住待检测的人体目标,一定程度上减少因框住人体目标不完整而造成待检测人体信息缺失的问题。

3.4 多尺度训练分析

训练过程中采用多尺度训练的方法,可以增强模型对不同分辨率检测的鲁棒性。笔者使用 YOLOv3 网络采用 *IoU* 损失函数和改进网络采用 *GIoU* 损失函数对测试集中 4 组不同尺寸图片进行对比实验,实验结果以平均 *mAP* 为指标,对比结果如表 3 所示。

由表 3 可知,首先从不同尺寸图像上的检测指标上看,改进网络使用 *GIoU* 损失函数比 YOLOv3 使用 *IoU* 损失函数要高。其次,采取增大输入照片尺度能够提升人体目标检测性能。最后采用多尺度训练能够使模型对不同尺寸图像具有鲁棒性。改进网络使用 *GIoU* 回归损失函数检测效果如图 7 所示。

(a) A社区入口 (b) B社区出口 (c) C社区入口

图 6 边界框检测对比效果

Figure 6 Bounding-box contrast effect

表 3 目标敏感性分析

Table 3 Performance comparison of different algorithms

优化方案	检测准确率 *mAP*/%			
	A 组	B 组	C 组	D 组
YOLOv3+*IoU*	84.58	86.73	90.69	93.08
改进网络+*GIoU*	87.74	90.01	92.47	96.23

(a) D社区入口 (b) A社区入口 (c) B社区入口

图 7 改进网络检测效果

Figure 7 Improved network detection effect

3.5 人体目标属性调用处理

开放平台对人体目标属性检测的过程为:先检测图片中的人体目标再检测人体属性,平均一张监控图片的检测时间在 2.9 s,不能满足检测实时性要求。因此,为满足检测任务实时性要求,采用改进网络先检测监控视频中的人体目标并进行截图后,再使用平台对图片中的人体目标进行检测,最后再对报文进行解析并保存到自制数据库中。当有正面遮挡时,以截取人脸为主要存储信息,同时其他人体信息捕捉超过检测属性的 40% 即为有效信息,避免冗余数据,易于数据管理。本文方法对监控视频中一个人体目标进行人体属性检测并显示的平均时间为 1.15 s,不仅检测效率大幅提升了 60.34%,还在一定程度上减少了计算资源的浪费。通过 MATLAB 构建 GUI 界面进

行展示的检测效果如图 8 所示。

图 8　调用分析显示效果

Figure 8　The show of call analysis

由显示效果可知,在实时监控下借用人脸相机应用本文方法,既能使用人脸相机获取人脸照片数据,还能获取目标人体的其他属性数据(如:性别、年龄、衣服类型、有无眼镜、有无电话等)。并且本文方法对采集到人体数据直接解析到自建的 ACCESS 社区人体信息数据库中,解决了实验平台中数据库资源不能直接调用的问题,真正地实现了视频数据的结构化描述。自建的社区人体信息数据库不仅可以实现大规模数据实时存储和查询的要求,还能提升社区对于日常监控管理的水平。

3.6　改进网络分析

将改进后的网络与 Faster R-CNN[4]、SSD[7] 和 YOLOv3[8] 网络在自制数据集上进行性能实验对比,以 *mAP* 以及每秒刷新图片的帧数作为检测评价指标,对比结果如表 4 所示。

表 4　不同算法的性能对比

Table 4　Performance comparison of different algorithms

网络框架	主干网络	*mAP*/%	帧速率/($f \cdot s^{-1}$)
Faster R-CNN[4]	ResNet-50	90.3	6
SSD-300[7]	VGG16	86.1	51
YOLOv3[8]	Darknet-53	87.8	42
改进网络	Darknet-53	91.8	45

由表 4 可知,改进网络 *mAP* 数值不但略高于 Faster R-CNN,识别帧率还是其 7.5 倍。其次,改进网络与 SSD 算法相比,检测速度略低,但 *mAP* 高于后者。最后,改进网络相比 YOLOv3 在准确率和识别帧速率上都有一定的提升。综上所述,改进网络不仅兼顾了检测准确率和检测速度,还可以较好地完成人体目标检测任务。

4　结论

提出一种将改进 YOLOv3 网络和调用人体信息识别模块相结合的人体信息检测方法。先使用自制数据集进行人体目标信息的检测实验,通过改进网络结构,使用新的边界框回归损失函数 *GIoU*、*K*-means++算法进行目标框维度聚类以及多尺度训练方式改进 YOLOv3 网络,再采用改进后的网络实现监控视频下人体目标的截取,最后调用人体目标属性检测模块,检测视频监控下的人体信息并存储到自建数据库中。本文方法能够利用现有人脸相机实现快速、准确地检测人体信息,并与视频监控管理系统进行对接,可显著提升社区视频监控管理系统的精细化管理能力。

参考文献:

[1] 赵思阳. 社区治理现代化视域下智慧社区建设研究——以洛阳市涧西区天津路街道为例[D].郑州:郑州大学, 2018.

[2] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE,2014:580-587.

[3] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision. New York: IEEE, 2015: 1440-1448.

[4] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017,39(6): 1137-1149.

[5] DAI J F,LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks[J]. Computer vision and pattern recognition, 2016, 29: 379-387.

[6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, realtime object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE,2016: 779-788.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [M]//LEIBE B, MATAS J, SEBE N, et al. Computer Vision-ECCV 2016. Cham: Springer International Publishing,2016: 21-37.

[8] REDMON J, FARHADI A. YOLOv3: an incremental improvement [C]// 2017 IEEE Conference on

Computer Vision and Pattern Recognition (CVPR). New York:IEEE,2017:6517-6525.
[9] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE transactions on pattern analysis & machine intelligence, 2010, 32(9): 1627-1645.
[10] 张素洁,赵怀慈. 最优聚类个数和初始聚类中心点选取算法研究[J]. 计算机应用研究, 2017, 34(6): 1617-1620.
[11] REZATOFIGHI H,TSOI N,GWAK J,et al.Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).New York:IEEE,2019:658-666.
[12] 施辉,陈先桥,杨英. 改进YOLOv3的安全帽佩戴检测方法[J]. 计算机工程与应用, 2019, 55(11): 213-220.
[13] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. New York:IEEE,2017:936-944.
[14] LIU W, WEN Y, YU Z, et al. Large-margin softmax loss for convolutional neural networks[C]//Proceedings of the 33rd International Conference on Machine Learning. Washington DC:IMLS, 2016: 507-516.
[15] HE K M,ZHANG X Y,REN S Q,et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770- 778.
[16] 魏宏彬,张端金,杜广明,等. 基于改进型YOLO v3的蔬菜识别算法[J].郑州大学学报(工学版),2020,41(2):7-12.

Video Surveillance Detection Method Based on Improved YOLOv3 algorithm and Human Body Information Data Fusion

ZHANG Zhen, LI Haofang, LI Mengzhou, MA Junqiang

(School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: In the current community video surveillance system, only a face camera was used to collect the entrance and exit face data, other valuable human information was negelected. In this paper, a human information detection method that combined improved YOLOv3 network and calling human information recognition module was proposed. The *K*-means++ algorithm was used to obtain the prior frame of the data set; the new bounding box regression loss function *GIoU* was used to improve the detection accuracy, and then multi-scale training was performed to obtain the human detection network model. Finally, the human detection model was used to detect human targets; and the human body information recognition module was used to analyze and save human body information. The experimental results showed that the method could detect human targets quickly, and accurately obtain various attribute information of human targets. Among them, the *mAP* of human detection model on the test set reached 91.8%, and the recognition speed was 45 f/s.

Key words: video surveillance; *K*-means++; *GIoU*; multi-scale training; improved YOLOv3; human information