

基于改进 YOLOv2 网络的遗留物检测算法

张瑞林,张俊为,桂江生,高春波,包晓安

(浙江理工大学信息学院,杭州 310018)

摘要: 为了提高在复杂环境下检测遗留物体的准确度和实时性,提出了一种基于改进 YOLOv2 网络的遗留物检测算法。该算法在 YOLOv2 网络结构基础上结合残差网络,将浅层和深层特征多次融合,在基本不增加原有模型计算量和时间的情况下,提高了监控画面中检测小体积遗留物体的性能;同时以 YOLOv2 目标检测为基础,排除驻留行人和动物等非物体目标的干扰,并对目标筛选得到的可疑目标跟踪计时,停留时间超过阈值的目标标记为遗留物。以 PETS2006 和 i-LIDS 作为数据集进行实验,结果表明:该算法在提高遗留物检测准确度的同时缩短了处理时间,对人流密集的复杂环境抗干扰能力强。

关键词: YOLOv2 网络;遗留物检测;残差网络

中图分类号: TS195.644

文献标志码: A

文章编号: 1673-3851(2018)05-0325-08

0 引言

遗留物是指由运动主体携带进监控场景中,并在场景中保持静止超过一定时间阈值的物体^[1]。遗留物检测是公共安全领域中智能视频监控的重要部分,并且与我们的生活息息相关。在公共场所下,视频监控下物品遗留现象发生普遍,尤其是在商场、学校、地铁站等人流密集环境下。目前已有学者提出许多不同的遗留物检测算法。张超等^[2]通过建立不同更新频率的高斯背景模型获取场景中静止物体,根据物体的颜色轮廓以及先验知识来进行遗留物分析,但是该方法没有考虑场景中行人的干扰,误检率大。Tripathi 等^[3]使用背景差分法提取前景对象,检测连续帧前景对象的轮廓特征以得到静止目标,通过基于边缘的物体识别方法排除驻留行人得到遗留物。但是这种算法很容易将行人肢体和其他运动主体误检为遗留物目标。叶立仁等^[4]利用局部更新的混合高斯模型与改进的三帧差分法获取前景目标,通过 HOG 行人检测排除驻留行人得到遗留物。但是 HOG 行人检测的处理时间相对较长,影响了遗留物检测的实时性。Lin 等^[5]结合短期和长期背

景模型提取背景对象,采用基于代码模式的时态转换识别静态目标区域,通过分析遗留物主人的回溯轨迹判断候选区域是否含有遗留物,但是这种方法对于行人拥挤的复杂环境下的抗干扰能力不强。

卷积神经网络具有提取图像高层特征的优势^[6],在图像分类研究的基础上,该算法在目标检测领域取得了突破性进展。Redmon 等^[7]在 YOLO 算法基础上优化模型结构提出了 YOLOv2 算法,该算法的检测速度和准确率都达到了令人满意的效果。在检测速度为 67 fps 情况下,YOLOv2 在 VOC2007 数据集上平均准确率可以达到 76.8%^[8]。目标检测与遗留物检测的前景目标提取有很多相似之处,都需要在场景中快速地发现目标物体。复杂环境下背景差分提取目标会导致遗留物检测较高的误检率^[9],同时克服场景中行人干扰大大降低了实时性^[10-11]。因此,本文采用取得最佳检测效果的 YOLOv2 网络,代替传统的前景目标提取进行遗留物检测。

本文提出了一种基于改进 YOLOv2 网络的遗留物检测算法。将 YOLOv2 网络用于遗留物检测,能够排除复杂环境下驻留行人和动物等指定非物体目标的干扰,以满足智能监控系统处理的实时性和

收稿日期:2017-10-13 网络出版日期:2017-12-11

基金项目:国家自然科学基金项目(61502430,61379036,61562015);浙江理工大学 521 人才培养计划

作者简介:张瑞林(1961-),男,浙江嵊州人,教授,博士,主要从事计算机应用技术等方面的研究。

通信作者:包晓安,E-mail:baoxiaoan@zstu.edu.cn

准确性要求;在 YOLOv2 网络结构基础上引入深度残差网络,对浅层低分辨率特征和深层特征融合,在基本不增加原有模型计算量和时间情况下,提高检测小体积遗留物的性能。

1 遗留物检测算法

1.1 YOLOv2 网络及其改进

YOLOv2 网络结构参照 SSD 和 YOLO 的网络结构,以 Darknet-19(包含 19 个卷积层、5 个最大值池化层)为基础模型进行特征提取,YOLOv2 检测网络如图 1 所示。与 YOLO 在结构上相比,YOLOv2 将分类网络改成检测网络,去除了原网络最后一个卷积层,增加了三个尺寸为 3×3 ,通道数为 1024 的卷积层,并且在每个卷积层后面跟一个尺寸 1×1 的卷积层,输出维度是检测所需数量;YOLOv2 采用固定框(Anchor boxes)代替全连接层来预测目标框

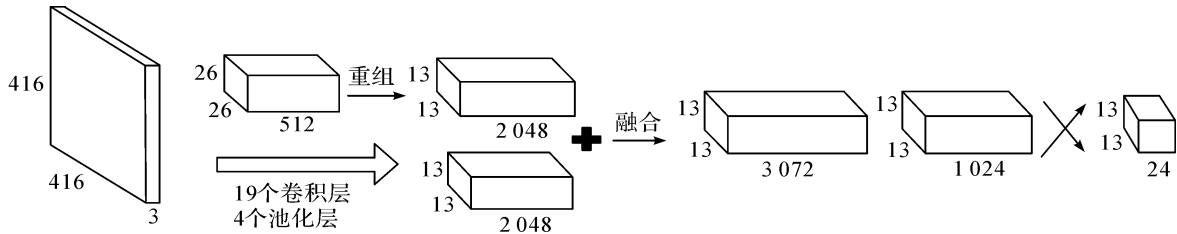


图1 YOLOv2 网络

YOLOv2 网络对 13×13 卷积特征图中的每个网格预测 5 个目标框,同时每一个目标框预测 5 个值,分别为 t_x, t_y, t_w, t_h, t_o ,其中前四个是坐标, t_o 是置信度。YOLOv2 通过相对于网格坐标来预测目标框的中心位置,如图 2 所示。如果网格距离图像左下角的边距为 (c_x, c_y) 以及该网格对应预定固定框的宽高分别为 (p_w, p_h) ,那么预测值可以表示为:

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \\ Pr(object) * IOU(b, object) = \sigma(t_o) \end{cases} \quad (1)$$

其中: σ 定义为 sigmoid 激活函数,将函数限定 $[0, 1]$,用来预测相对于网络中心的偏移; $\sigma(t_x)$ 是预测目标框的中心相对网络左上角的横坐标, $\sigma(t_y)$ 是纵坐标; $\sigma(t_o)$ 是预测目标框的置信值。

YOLOv2 网络只包括卷积层和池化层,能够随时改变输入图像分辨率。在训练时每隔十轮便改变模型的输入分辨率继续训练,使模型适应不同的输入分辨率,同时增强对不同分辨率图像的鲁棒性。

(Bounding boxes),利用 k -means 算法^[12]对 VOC2007+2012 数据集的目标框进行聚类分析,自动选择 anchor 最佳的大小和数量。同时移除一个池化层来使卷积层输出分辨率提高,将网络输入尺寸由 448×448 改为 416×416 ,使得特征图只有一个中心,较大的物品更有可能出现在图像中心。同时卷积层降采样参数 factor 设为 32,使得输入卷积网络分辨率 416×416 图像最后一层获取 13×13 特征图。YOLOv2 整个网络均为卷积操作,很好地保留了空间信息,让卷积特征图每个特征点和原图中的每个网格(cell)对应。Faster R-CNN 和 SSD 在不同层次的特征图上运行 RPN 来获得多尺度的适应性,而 YOLOv2 通过添加一个转移层(Passthrough layer)连接高低分辨率特征图,将上一层获取的 $26 \times 26 \times 512$ 的特征图转化为 $13 \times 13 \times 2048$ 的特征图,与原来的深层特征相连接。

模型对不同分辨率的处理速度不同,分辨率小的处理速度更快,因此 YOLOv2 网络能够根据需求调节准确率和处理速度。在分辨率较低情况下,YOLOv2 不仅在准确率上与 Fast R-CNN 持平,同时处理速度可以达到 90 fps。

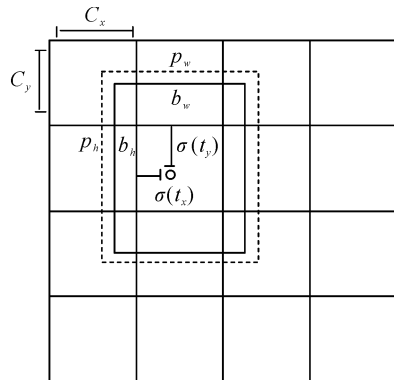


图2 目标框直接位置预测

YOLOv2 在目标检测方面取得了很好的效果,但其并不完全适用检测视频监控画面中小体积遗留物目标。对于小目标检测,往往更依赖浅层特征,但原网络结构中前面的 19 个卷积层和 4 个池化层进行特征提取后,高分辨率的浅层特征几乎不再利用,导致对应特征图上的特征往往难以得到充分训练。

深度残差网络(ResNets)采用残差块作为网络的基本组成部分,在原始卷积层外增加捷径通道构成残差块结构,残差块示意图如图 3 所示,假设残差块输入为 x ,要拟合的函数映射为 $H(x)$,定义另外一个残差映射 $F(x)=H(x)-x$,则原始映射函数 $H(x)$ 被表示为 $H(x)=F(x)+x$ 。ResNets 利用残差块将网络对 $H(x)$ 的学习变为对 $F(x)$ 的学习,同时残差映射 $F(x)$ 的优化要比原始映射 $H(x)$ 要简单。ResNets 通过捷径跳跃绕过一些层的连接,使高低层特征相加融合,将浅层的误差向上一层传播,浅层的高分辨率特征得到更充分训练,在很大程度上解决 DCNN 随着深度增加而带来的梯度消失现象^[13],提高了小目标训练精度的效果。

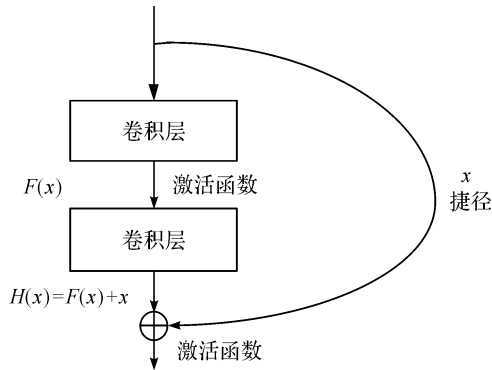


图 3 残差块示意图

对于卷积神经网络,不同深度对应着不同层次的语义特征,浅层网络分辨率高,获得更多细节特征,深度网络分辨率低,获得更多的是语义特征^[14]。本文在 YOLOv2 网络结构基础上作出改进,结合

ResNets 将网络前面的 19 个卷积层和 4 个池化层,通过增加额外的跨层连接改成残差网络,对浅层高分辨和深层低分辨率特征信息进行不断融合,构成如图 4 所示的 YOLOv2-A 网络。YOLOv2-A 网络加强了不同分辨率、不同语义强度的高低层特征融合,图 5 为 YOLOv2-A 网络中的残差网络,为减少训练过程中的计算量和资源空间,选择原网络中池化处理后得到的特征图进行融合,将第 11 层与第 14 层特征堆叠到不同通道,其中在第 11 层提取第 3 层特征图卷积池化处理后得到第 14 层 26×26 特征图,从而将浅层分辨率 104×104 特征图与低分辨率 26×26 的特征图相融合。同时增加额外的跨层连接网络第 22 层与第 25 层,其中在第 23 层提取第 8 层特征图卷积池化处理后得到第 25 层 13×13 特征图,从而将浅层分辨率 52×52 特征图与低分辨率 13×13 的特征图相融合,最后将融合后的特征图再经卷积池化处理得到 $13 \times 13 \times 1024$ 特征图。训练过程中高层的梯度通过相加融合向浅层传播,减缓了 YOLOv2 网络层数过多而造成的梯度消失现象,同时使得浅层特征图上的特征得到充分利用,在基本不增加原有模型计算量和时间情况下,提高了小目标物体检测的性能。两次浅层和深层特征融合分别表示为:

$$F(x_{15}) = F(x_{11}) + F(x_{14}) \quad (2)$$

$$F(x_{26}) = F(x_{22}) + F(x_{25}) \quad (3)$$

其中: x_i 表示 YOLOv2-A 网络中第 i 层卷积层, $\forall i \in \{0, 1, 2, \dots\}$; $F(\cdot)$ 表示网络第 x_i 层输出的特征图。

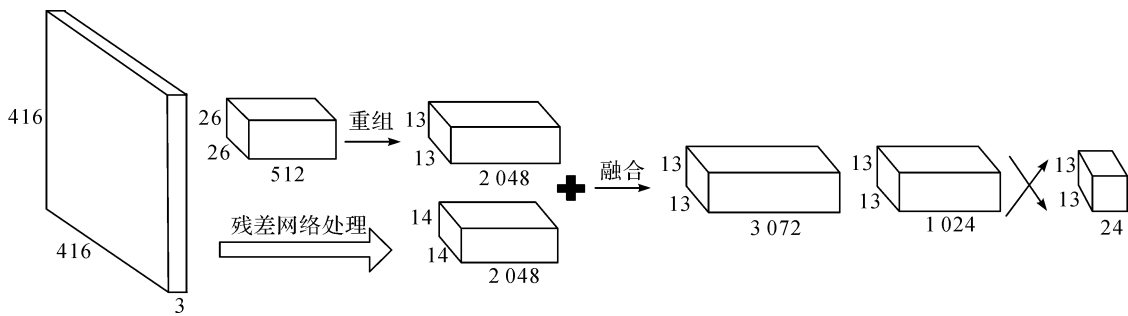


图 4 YOLOv2-A 网络

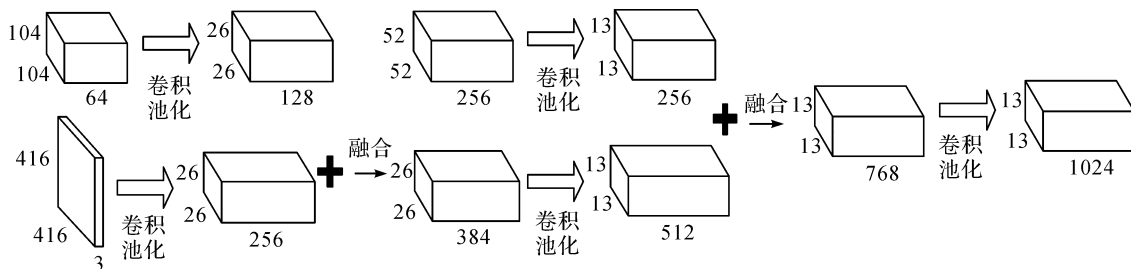


图 5 YOLOv2-A 网络中残差网络

1.2 遗留物分析

视频监控画面经过 YOLOv2-A 实时检测后,获取每帧图像检测到的目标坐标 (x, y, w, h) 和类别值 $type$ 。复杂场景中经常有行人、动物等运动目标驻留和静止行人肢体摆动,检测时容易将其误检为遗留物,因此对于检测到的目标,根据目标类别值过滤运动目标,排除运动目标对检测的干扰。本文采用比较两个目标的类别值和重叠度 IOU ,将每帧过滤后的目标依次筛选划分到背景、可疑、遗留物队列,并定义队列 $deque<objInfo, pair<time, objStatus>>$ 其中: $objInfo$ 表示目标信息,即目标坐标和类别; $time$ 表示对目标统计的时间,即目标静止时间 $timeStatic$ 和目标消失时间 $timeLost$; $objStatus$ 表示目标状态,即背景、可疑、遗留物状态。

图 6 中: n_{pre} 表示当前帧检测到的目标数量, n_{brg} 和 n_{sus} 分别表示上一帧背景、可疑队列中目标的数量。如图 6 所示,在目标类别值相同情况下,依次比较当前帧目标分别与背景、可疑队列中的目标是否重合,即比较两个目标重心在 x 轴或 y 轴方向上的距离与两个目标的长或宽一半相加的距离。若两个目标不重合,则重叠度为 0,判断当前帧目标为新可疑目标,状态值记为可疑并存入可疑队列。两个目标重叠情况下按式(4)计算重叠度:

$$IOU/\% = \frac{S_{overlap}}{S_{union}} \times 100 \quad (4)$$

其中: $S_{overlap}$ 表示两个目标的交集面积; S_{union} 表示两个目标的并集面积。

$$F(IOU, T) = \begin{cases} 0, IOU_{(i,y)} < T_{IOU} \\ 1, IOU_{(i,y)} < T_{IOU}, IOU_{(k,j)} > T_{IOU} \\ 2, IOU_{(i,y)} < T_{IOU}, IOU_{(k,j)} < T_{IOU} \end{cases} \quad (5)$$

其中: j 表示当前帧第 j 个目标, $j \in \{0, 1, \dots, n_{pre-1}\}$; i 表示背景队列第 i 个目标 $i \in \{0, 1, \dots, n_{brg-1}\}$; k 表示上一帧可疑队列第 k 个目标, $k \in \{0, 1, \dots, n_{sus-1}\}$; $IOU_{()}$ 表示两个目标的重叠度; T_{iou} 表示重叠度阈值; 0 代表目标为背景目标; 1 代表目标保持静止; 2 代表目标为新可疑目标。

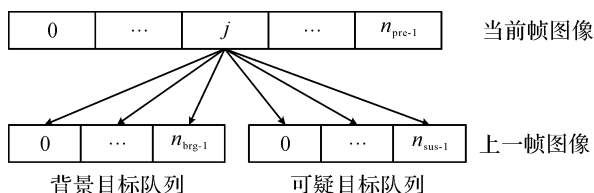


图 6 当前帧目标与背景、可疑队列目标比较

当前帧目标与背景、可疑队列目标重叠度比较结果按照式(5)进行筛选。若 $IOU_{i,j} < T_{iou}$,则表示

当前帧的第 j 个目标是背景队列的目标,继续比较当前帧第 $j+1$ 个目标。若 $IOU_{i,j} < T_{iou}, IOU_{k,j} > T_{iou}$,说明当前帧第 j 个目标与上一帧可疑队列的第 k 个目标对应,且认为该目标依然保持静止,对 $timeStatic$ 值加 1, $timeLost$ 值清零;若 $IOU_{i,j} < T_{iou}, IOU_{k,j} < T_{iou}$,则认为当前帧第 j 个目标是新检测到的目标,存入可疑队列,状态值记为可疑。同时也说明上一帧可疑队列第 k 个目标物体发生移动或被遮挡,对 $timeLost$ 值加 1, $timeStatic$ 值清零。

得到划分后的背景、可疑队列,将可疑队列中 $timeStatic$ 值大于静止时间阈值的目标,对其再次目标类别值过滤,以防漏检现象。若非运动目标,将其状态值记为遗留物存入遗留物队列。同时对遗留物队列中的目标消失计时,若 $timeLost$ 值超过消失时间阈值 T_l ,删除其目标数据。

1.3 遗留物检测算法与实现

遗留物检测算法的步骤为:

- 对 720 p 视频帧进行 YOLOv2-A 目标检测,检测得到当前帧 num 个目标的坐标和类别值。
- 根据目标类别值过滤行人、动物等指定非物体目标。计算当前帧过滤后目标与背景、可疑队列目标的重叠度 IOU 。
- 比较两个目标的类别值和重叠度,将过滤后的目标筛选划分到背景、可疑队列。
- 对可疑队列中目标跟踪计时,若目标 $timeStatic$ 值大于阈值 T_s ,则将其状态值记为遗留物,存入遗留队列中;否则转到步骤 e)。

e) 对遗留物队列中的目标消失计时,若目标 $timeLost$ 值超过 T_l ,则将删除其目标数据。

实现本文算法的伪代码为:

- Begin:
- capture image //获取 720p 视频帧
- detect object //YOLOv2-A 检测 num 个目标
- for $m=1:num$
- if ($type \neq peopleType \ \& \ \& \ type \neq animalType$) //排除类别值为行人、动物的目标
- if ($type_i = type_j$)
- count $IOU_{k,j}$ //计算与背景目标的重叠度
- end
- if ($type_k = type_j$)
- count $IOU_{k,j}$ //计算与可疑目标的重叠度
- end
- compare($IOU_{i,j}, T_{iou}$)
- compare($IOU_{k,j}, T_{iou}$) //目标筛选


```

14. push backdeque, extradeque //目标存入背景、可疑队列
15. if(timeStatic > Ts)
16. push abdeque //目标存入遗留物队列
17. end
18. if(timeLost > Tl)
19. delete object //删除消失遗留物
20. end
21. End

```

2 实验结果与分析

本文实验环境为: Intel Core CPU i7-4790k 4.00 GHz, 32 G 内存, Nvidia Geforce GTX1080, Ubuntu16.04, 64 位操作系统, 使用 Darknet 框架。

2.1 YOLOv2-A 网络实验结果与分析

实验采用 VOC 数据集 (VOC2007 + 2012) 和 COCO 数据集来训练 YOLOv2-A 网络, 在初始学习率为 0.001 条件下训练 160 次, 训练至 60 次和 90 次的时候, 学习率变为原来的 1/10。网络调整学习率的策略 policy 为 STEPS, 冲量 momentum 系数设置为 0.9; 每次迭代 (batch) 训练的图像数量为 64, 最大迭代次数为 210000; 在迭代次数为 100、80000、120000 次时学习率发生变化 ($steps = 100, 80000, 120000$), 学习率变化的比率为 10.0、0.1、0.1, 与 $steps$ 中的参数对应。

图 7 为选取迭代次数 0 到 9×10^3 , YOLOv2 与 YOLOv2-A 网络的 loss 值随着迭代次数变化。loss 值表示网络计算预测的目标框 boxes 与标定框 ground truth 之间的误差, loss 值越低, 表明该网络检测目标标准度越高。在两种网络网络迭代 200 次内, loss 值快速大幅度下降, 之后随着迭代次数增加缓慢下降趋于平稳。总体上 YOLOv2-A 网络的 loss 值低于 YOLOv2 网络, 表明本文提出的 YOLOv2-A 网络的检测目标标准度要高于原网络。图 8、图 9 分别表示 YOLOv2 和 YOLOv2-A 网络对存在小目标的图像检测效果, YOLOv2-A 网络基本检测出图像中存在的小目标物体, YOLOv2 网络对检测小目标物体漏检率较大, 表明本文提出的 YOLOv2-A 网络通过将浅层高分辨率特征与深层特征融合, 提高了检测小目标物体的准确度。

2.2 遗留物算法实验结果与分析

为了验证本文算法在实际复杂环境下的效果, 本文采用 PETS2006 数据集^[15]、i-LIDS 数据集^[16]和自制数据集进行测试。由于本文算法采用单摄像

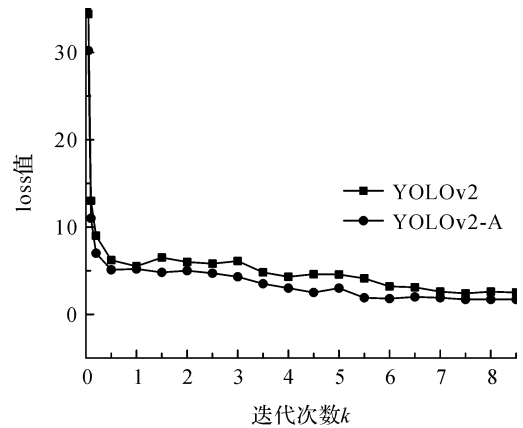


图7 两种网络的 loss 值随迭代次数变化



图8 YOLOv2 网络检测结果

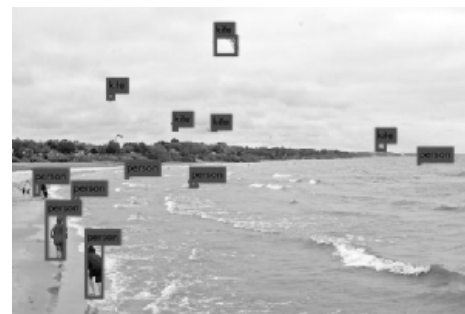


图9 YOLOv2-A 网络检测结果

头进行实验测试, 实验选取数据集中同一个相机的视图序列帧进行测试。将重叠率阈值设置为 70%, 即目标类别值相同情况下, 重叠率大于 70% 的两个目标判定为同一目标。将可疑队列中静止时间达到 4 s 的可疑目标判断为遗留物, 对消失时间超过 3 s 的物体作出消失处理。由于视频帧率为 25 fps, 因此静止时间阈值 T_s 为 100 帧, 消失时间阈值 T_l 为 75 帧。

图 10、图 11 分别是本文算法对 PETS2006 和 i-LIDS 序列帧的检测结果。在人流拥挤的公共场所, 本文算法能有效的排除场景中行人带来的干扰, 场景中遗留的行李箱和无人看管的包都能被准确检测出, 并且被检测出的遗留物通过矩形框标记出作为报警的依据。



图10 PETS2006 序列帧遗留物检测结果



图11 iLids 序列帧遗留物检测结果

本文算法对 PETS2006 和 i-LIDS 数据集不同序列的检测结果如表 1 所示。PETS2006 序列中的场景相对比较简单,没有太多的遮挡和和行人驻留,算法对序列 S1-S4 检测的误检率为 0。i-LIDS 数据集序列场景复杂度由简单到难,算法通过 YOLOv2-A 检测有效排除了场景中静坐行人肢体移动和部分遮挡产生的干扰,依次检测出场景中放置的多个遗留物。实验结果充分表明了本文算法对复杂环境的

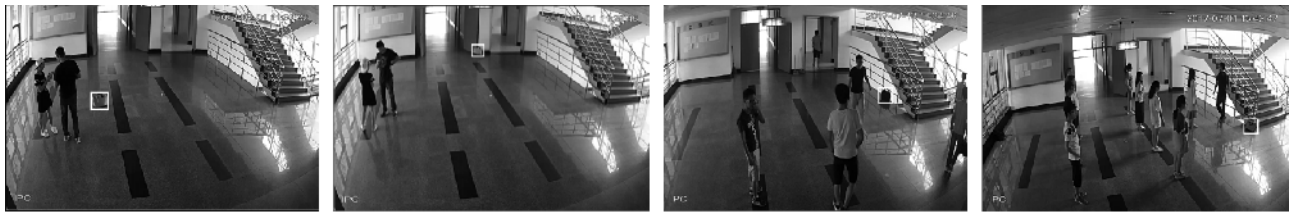
抗干扰能力强。

表 1 PETS2006 和 i-LIDS 数据集序列遗留物检测结果

数据集	序列	遗留物数量	检测数量	静止人数	误检数量
PETS2006	S1	1	1	1	0
	S2	1	1	1	0
	S3	1	1	1	0
	S4	1	1	1	0
i-LIDS	AB Easy	1	1	1	0
	AB Medium	2	2	3	0
	AB Hard	4	4	4	0

本文还对自制数据集的视频序列进行测试,测试结果如图 12。图 12(a)演示了本文算法将放置场景中不同位置的遗留物依次检测出。图 12(b)演示了从存在少数行人驻留到人员流动密集的场景下,本文算法均能克服存在的干扰影响,在保证准确率的同时,快速标记出被放置的遗留物。

为了进一步验证本文算法在复杂环境下的准确度和实时性,选取 DAOCS(Detecting abandoned objects in crowded scenes)算法^[17]、YOLOv2-AOD(YOLOv2-abandoned detecting objects)算法,与本文提出的算法进行对比,实验结果如图 13 和图 14 所示。



(a) 场景中不同位置的遗留物检测

(b) 不同复杂度场景下遗留物检测

图 12 本文算法对自制数据集视频序列检测结果



(a) DAOCS算法

(b) YOLOv2-AOD算法

(c) 本文算法

图 13 三种算法序列帧第一组检测结果



(a) DAOCS算法

(b) YOLOv2-AOD算法

(c) 本文算法

图 14 三种算法序列帧第二组检测结果

从图13和图14可以看出,三种算法对i-LIDS两组序列帧检测结果对比。在人流拥挤、行人驻留较多的复杂场景下,DAOCS算法能够检测出场景中放置的遗留物,但是采用的HOG行人检测对场景中静坐行人的肢体移动和部分遮挡产生的干扰难以排除,导致了遗留物较高的误检率。YOLOv2-AOD算法跟本文算法相比较,YOLOv2-AOD算法在两组检测中都漏检了场景中体积较小的遗留物。本文算法提高了对场景中物体检测准确度,从而降低了对遗留物检测的误检率。

DAOCS算法和本文算法在视频帧位置上的处理时间如图15所示。由图15可知,本文算法对测试视频每帧图像的处理时间基本上小于38ms。同时相比DAOCS算法,处理时间明显减少。本文i-LIDS测试视频的帧率为25fps,采用本文算法对每帧视频图像的平均处理时间为23ms,可以满足智能监控系统处理的实时性。

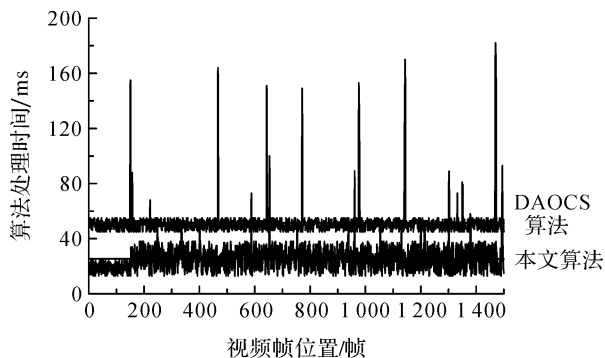


图15 两种遗留物检测算法的处理时间

3 结 语

本文构建了YOLOv2-A网络结构,该网络可以更好地适应遗留物检测的特点,提高检测体积较小遗留物性能。同时以YOLOv2目标检测为基础,对目标筛选得到的可疑目标跟踪计时后,及时检测出场景中放置的遗留物,对场景中存在的多个遗留物体进行标记。通过以PETS2006和i-LIDS作为数据集实验证明,本文算法能很好地适用于不同复杂度的场景,同时能有效克服复杂环境中存在的干扰影响,在很大程度上保证了遗留物检测的实时性和准确性。但是目前本文算法的部分阈值参数需人工对复杂度不同的环境做出调整,今后会对算法阈值参数的自适应取值问题深入研究,以实现对不同复杂度环境的自适应能力。

参考文献:

[1] Muchtar K, Lin C Y, Kang L W, et al. Abandoned

object detection in complicated environments[C]//Signal and Information Processing Association Summit and Conference. IEEE,2014:1-6.

[2] 张超,吴小培,周建英,等. 基于改进混合高斯建模和短时稳定度的遗留物检测算法[J]. 信号处理,2012,28(8):1101-1111.

[3] Tripathi R K, Jalal A S, Bhatnagar C. A framework for abandoned object detection from video surveillance[C]//Computer Vision, Pattern Recognition, Image Processing and Graphics. IEEE,2014:1-4.

[4] 叶立仁,何盛鸿,赵连超. 复杂环境下的遗留物检测算法[J]. 计算机工程与科学,2015,37(5):986-992.

[5] Lin K, Chen S C, Chen C S, et al. Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance [J]. IEEE Transactions on Information Forensics & Security,2015,10(7):1359-1370.

[6] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature,2015,521(7553):436-444.

[7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2016:779-788.

[8] 魏湧明,全吉成,侯宇青阳. 基于YOLOv2的无人机航拍图像定位研究[J]. 激光与光电子学进展,2017,54(11):111002.

[9] Huang W, Jin L, Luo Y, et al. A novel algorithm for abandoned object detection [C]//IEEE International Conference on Information and Automation. IEEE, 2017:1583-1587.

[10] Bhargava M, Chen C C, Ryoo M S, et al. Detection of abandoned objects in crowded environments [C]//Advanced Video and Signal Based Surveillance, 2007, IEEE Conference on. IEEE,2007:271-276.

[11] Singh A, Sawan S, Hanmandlu M, et al. An abandoned object detection system based on dual background segmentation[C]//Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE Computer Society,2009:352-357.

[12] 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.

[13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Computer Vision and Pattern Recognition. IEEE,2016:770-778.

[14] Bell S, Zitnick C L, Bala K, et al. Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks[C]//Computer Vision and

- Pattern Recognition, IEEE, 2016:2874-2883.
- [15] PETS 2006 Dataset[DS/OL]. (2006-06-18)[2017-10-13]. <ftp://ftp.cs.rdg.ac.uk/pub/PETS2006/>.
- [16] i-LIDS Dataset for AVSS 2007[DS/OL]. (2007-02-23)[2017-10-13]. <ftp://motinas.elec.qmul.ac.uk/pub/iLids>.
- [17] Wahyono, Filonenko A, Jo K H. Detecting abandoned objects in crowded scenes of surveillance videos using adaptive dual background model [C]//International Conference on Human System Interactions. IEEE, 2015: 224-227.

Abandoned object detection algorithm based on improved of YOLOv2 network

ZHANG Ruilin, ZHANG Junwei, GUI Jiangsheng, GAO Chunbo, BAO Xiaolan

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: To improve the accuracy and real time of detecting abandoned objects in complex environments, an abandoned object detection algorithm based on improved of YOLOv2 network was proposed. Based on YOLOv2 network structure, the algorithm combined the residual network and integrated shallow-layer and deep-layer features. In the case of no increase of the calculation amount and time of the original model, the performance of small-volume abandoned object detection in the monitoring screen was improved. Meanwhile, based on YOLOv2 target detection, the interference of non-objects such as pedestrians and animals was excluded, and the suspicious target gained by target screening was tracked and timed. The target with the time of stay exceeding the threshold target was marked as an abandoned object. PETS2006 and i-LIDS were used as data set for experiments. The results show that the algorithm improves the accuracy of the residue detection and reduces processing time, with strong anti-interference ability in the complex environment.

Key words: YOLOv2 network; abandoned object detection; residual network

(责任编辑: 康 锋)