

基于多级特征融合的视频火焰检测方法

严 忱¹, 严云洋^{1,2}, 高尚兵¹, 朱全银¹

(1. 淮阴工学院计算机与软件工程学院, 江苏 淮安 223003)

(2. 江苏海洋大学计算机工程学院, 江苏 连云港 225005)

[摘要] 火焰前期一般多为小目标,但一般的火焰检测方法对于小目标的检测能力较差. 为检测早期火焰,提高火灾预防能力,提出了一种融合多级特征的视频火焰检测方法,针对下采样分辨率变小导致丢失目标的问题,引入了反卷积模块,并融合深层具有较强语义信息的特征和浅层具有较强细节信息的特征,从而有效提高了火焰的检测率. 所提算法在 Bilkent 大学火灾数据库 VisiFire 数据集上进行的实验表明, mAP 相较于 YOLOv2 提高了 10.0%, 与多种经典的深度学习算法模型相比,检测率更高.

[关键词] 火焰检测, 目标检测, 特征融合, YOLOv2

[中图分类号] TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2021)03-0131-06

Video Flame Detection Based on Fusion of Multilevel Features

Yan Chen¹, Yan Yunyang^{1,2}, Gao Shangbing¹, Zhu Quanyin¹

(1. Faculty of Computer & Software Engineering, Huaiyin Institute of Technology, Huai'an 223003, China)

(2. School of Computer Engineering, Jiangsu Ocean University, Lianyungang 225005, China)

Abstract: In the early stage, the flame is usually small target, but the general flame detection method has poor detection ability for small target. To detect early flame and improve fire prevention ability, a video flame detection method with multi-stage features is proposed. Aiming at the problem of missing target caused by the lower sampling resolution, a deconvolution module is introduced, and the features with strong semantic information in the deep layer and strong detail information in the shallow layer are combined to effectively improve the detection rate of flame. The experiment of the proposed algorithm on the Fire database VisiFire data set of Bilkent University showed that mAP improved by 10.0% compared with the YOLOv2, and the detection rate was higher compared with many classic deep learning algorithm models.

Key words: flame detection, object detection, feature fusion, YOLOv2

在社会生产生活中,火灾一直是威胁公共安全,危害人们生命财产的危害之一,它给人类社会带来了重大的生命危险和财产损失. 伴随着日益加快的社会生产力发展,日益增长的社会财富,城市的发展趋于现代化,主要的特点有:生产集中、人口集中、建筑集中和财富集中等. 同时出现的危险现象有易燃易爆物品增多,火源、危险源增多等现象. 这就导致城市火灾损失呈上升趋势^[1]. 因此,火灾检测成为了必不可少的防灾措施. 伴随着深度学习、目标检测^[2]技术的发展和数字监控摄像的普及,可以在多种场合进行实时监控,于是视频火焰检测技术也成为火灾防控的一种重要手段.

Sharma 等^[3]利用改进的 Resnet50 和 VGG16 网络进行火焰检测,想法是把全连接层用作特征提取层. 该算法虽然与单一的神经卷积网络相比准确率提高了,但是速度却比之前降低了. Lee 等^[4]改进了 GoogleNet,并融合了 VGG13 网络对火焰进行检测,进行改进后,网络模型中的卷积层和全连接层的宽度只有之前的一半,该方法的模型训练时间相较于经典的卷积神经网络也大大减少,但是该模型的准确率略有下滑. Zhang 等^[5]提出一种实时火焰检测算法,使用 HSV 颜色模型对于动态的火焰区域分割,然后提取均方差、面积、边界和圆形度等 5 个特征作为输入变量输入到 BP 神经网络中,进行火焰识别. 黄文锋等^[6]以

收稿日期:2021-05-20.

基金项目:国家自然科学基金项目(61402192)、江苏省“六大人才高峰”项目(2013DZXX-023)、江苏省高校自然科学基金重大项目(18KJA52001).

通讯作者:严云洋,博士,教授,研究方向:数字图像处理、模式识别. E-mail: yuhyang@hyit.edu.cn

BN_Inception 网络为骨干网,采用多种分辨率的神经网络互补学习复杂场景中火焰的多尺度视觉特征. 该算法侧重于目标所处的背景,主要提取局部目标和整体布局等目标特征. 江洋等^[7]以 RetinaNet 网络为基础提出了一种火焰检测算法,该算法以特征金字塔网络作为骨干网,并在特征金字塔的网络结构中加入残差网络模块,通过模型计算,得到特征更加丰富的特征金字塔,最后通过模型中的分类子网络和边框预测子网络对于骨干网的输出进行分类和回归.

近年来,深度学习领域发展迅速,有许多优秀的深度学习模型,如 SSD^[8] 系列、Yolo^[9] 系列等,都被应用于目标检测. 得益于它的“沙漏”结构,SSD 系列中的 DSSD^[10] 模型在小目标检测方面表现非常优秀. Yolo 系列中,最新的是 YOLOv3^[11] 和 YOLOv4^[12] 模型,检测率和检测速度表现依旧稳定. 但是由于从 YOLOv3 开始,作者加入了大量的残差结构,并融入了特征金字塔结构,对于不同大小的图像进行 3 次预测,大幅度提升了模型复杂度,加深了网络结构,对于设备的要求较高,不利于应用. 模型复杂度对比如表 1 所示.

表 1 模型复杂度对比
Table 1 Comparison of the model complexity

Model	Yolov2	Yolov3	Yolov4	Proposed
FLOPs(BFLOPs)	29.33	65.28	59.55	51.33

因此从应用性更强的角度出发,本文在 YOLOv2^[12] 算法的基础上,融合多级特征,充分利用高层和低层卷积层的语义信息,生成特征信息更丰富的 feature map,提高模型对于小目标的检测能力和整体的检测精度.

1 基于多级特征融合的视频火焰检测方法

1.1 YOLOv2^[13]

YOLOv2 是在 YOLO 的基础上进行的改进,优化了对物体的定位(localization)和召回率(recall),从而大大提高了目标检测的精度. 同时,YOLOv2 也保留了单阶段(one-stage)检测器检测速度快的优点.

YOLOv2 将 Darknet-19 作为主干网,包含了 19 个卷积层和 5 个最大池化层^[14],如表 2. Darknet-19 进行了 5 次下采样,为了得到更高分辨率的特征图(feature map),最后一层卷积没有添加池化层.

YOLOv2 在每层卷积后面都添加了 Batch Normalization 层,以 Batch Normalization 层来代替 dropout,虽然 Batch Normalization 层和 dropout 都有一定的正则化效果,但是 Batch Normalization 可以提升模型的收敛速度,一定程度上可以降低模型的过拟合. 在 YOLOv2 中还加入 Faster R-CNN^[15] 中的 anchor 机制,就是利用 K-Means 聚类的方式在训练集中聚类计算出更好的候选框的值,用于更好地预测 bounding box. 在引入 anchor 机制后,虽然 YOLOv2 的精度有大幅度下降,但是 recall 上升了,说明 YOLOv2 还有更多可提升的空间.

表 2 Darknet-19 网络结构
Table 2 Network structure of Darknet-19

Type	Filters	Size/Stride	Output	Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224	Convolutional	256	1×1	14×14
Maxpool		2×2/2	112×112	Convolutional	512	3×3	14×14
Convolutional	64	3×3	112×112	Convolutional	256	1×1	14×14
Maxpool		2×2/2	56×56	Convolutional	512	3×3	14×14
Convolutional	128	3×3	56×56	Maxpool		2×2/2	7×7
Convolutional	64	1×1	56×56	Convolutional	1024	3×3	7×7
Convolutional	128	3×3	56×56	Convolutional	512	1×1	7×7
Maxpool		2×2/2	28×28	Convolutional	1024	3×3	7×7
Convolutional	256	3×3	28×28	Convolutional	512	1×1	7×7
Convolutional	128	1×1	28×28	Convolutional	1024	3×3	7×7
Convolutional	256	3×3	28×28	Convolutional	1000	1×1	7×7
Maxpool		2×2/2	14×14	Avgpool		Global	1000
Convolutional	512	3×3	14×14	Softmax			

1.2 基于多级特征融合的检测模型

火焰扩散速度非常快,因此必须在前期尽早检测出火焰,防止扩散酿成火灾. 火焰前期多为小目标,因此

需要加强模型对于小目标的检测能力. 而卷积神经网络普遍存在的问题是:高层网络感受野较大,语义信息表征能力强,但分辨率低,几何细节信息表征能力较弱;而低层网络感受野比较小,几何细节信息表征能力强,虽然分辨率高,但语义信息表征能力弱^[16]. 在 YOLOv2 模型中,输入 416×416 的图像后经过 5 个最大池化层下采样,最终输出的图像大小仅为 13×13 ,许多小目标的特性信息和位置信息在下采样时容易丢失.

本文以 Darknet-19 网络为骨干网,去掉所有的最大池化层,使用步长为 2 的跨步卷积层代替池化层对图像进行下采样,防止最大池化层丢失大量图像信息,尽量最大化保留图像的特征信息. 同时引入 DSSD 的“沙漏”结构思想,将模型设计成“宽-窄-宽”的形式,目的就是将高层和低层的语义信息进行融合,得到更丰富的预测回归位置框和分类任务输入的多尺度特征图,以此来提高检测精度. 即在 5 次下采样结束后,在模块后添加反卷积模块,生成对应的 26×26 、 52×52 、 104×104 共 3 种大小的卷积层,而每个卷积层的实现最小单位为 DBL (卷积+Batch Normalization+Leaky relu),并在卷积层中加入 1×1 的小卷积核来减少模型的计算量,算法模型结构如图 1 所示,算法流程图如图 2 所示. 在反卷积模块中穿插进行张量拼接 (Concat),每次进行反卷积,得到更大的 feature map 后,Concat 可以将输出的大小相同的 feature map 进行合并,扩充张量的维度,将高层和低层的语义信息融合在一起. 去掉用于融合 13×13 和 26×26 大小卷积层的 reorg 层,使上下层之间相同大小图像中的特征信息更好地融合,以便获取更丰富的 feature map,提高模型检测率.

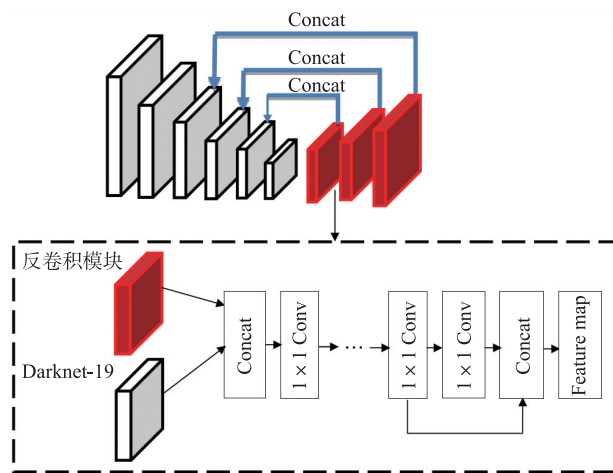


图 1 算法模型结构

Fig. 1 Model structure

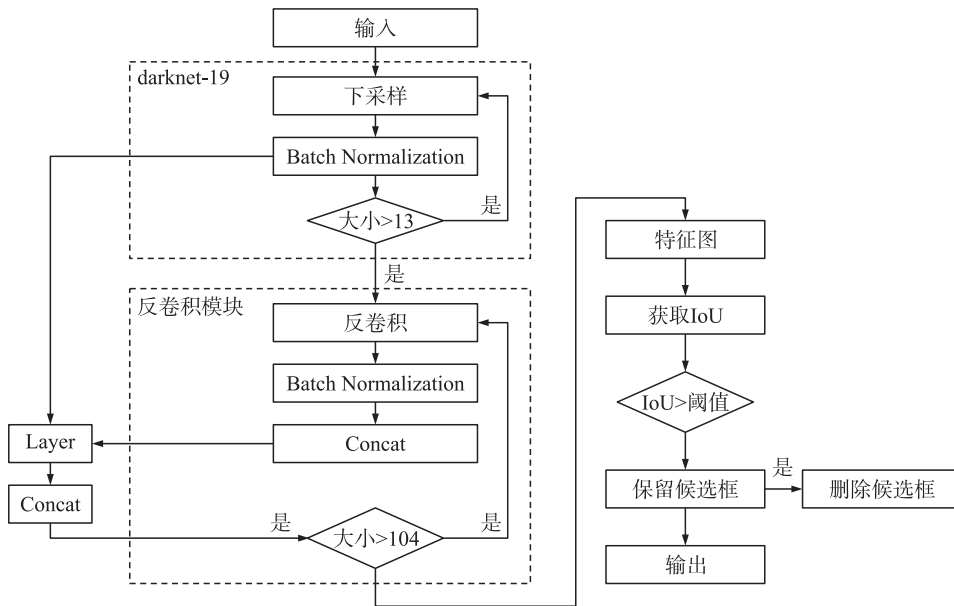


图 2 算法流程图

Fig. 2 Flow chart of the model

基于多级特征融合的视频火焰检测算法过程描述如下:

- (1) 输入待检测图片;
- (2) 通过步长为 2 的跨步卷积对图像进行下采样;
- (3) 将下采样得到的图像通过 Batch Normalization 层进行正则化处理;
- (4) 使用 3×3 的卷积核处理 feature map, 得到候选框位置、置信度和候选框中类别分数;
- (5) 若图像大小大于 13×13 时, 则跳转步骤 2, 否则, 进行反卷积;

- (6)将反卷积后得到的图像通过 Batch Normalization 层进行正则化处理;
- (7)通过 concat,将步骤 4 和步骤 6 大小相同的卷积层输出的 feature map 融合起来;
- (8)通过反卷积模块中的卷积层处理步骤 7 生成的 feature map,再次通过 concat 拼接,生成新的 feature map;
- (9)若图像大小小于 104×104 时,则跳转步骤 3;否则,得到输出的 feature map,更新候选框位置、置信度和候选框中类别分数;
- (10)计算出 feature map 上的候选框和真实边框的 IoU 值;
- (11)若 IoU 值大于阈值,则删除该候选框;否则,保留该候选框;
- (12)输出检测图片、候选框和候选框中的分类.

2 实验结果及分析

2.1 实验环境

实验环境:CPU: Intel® Core™ i5-7500 3.40GHz×4, GPU: 华硕 GeForce GTX 1080-8G, 内存:16GB, 操作系统: Ubuntu16.04.

2.2 实验数据集

本文的实验数据集分为两部分:一部分是通过网络搜集和整理所得的火焰视频和图片,其中包括室内、室外等不同场景,整理为 FireData 数据集;另一部分是 Bilkent 大学火灾数据库 VisiFire 的测试视频. 测试视频首先要逐帧抽取成图片,再对图片中的检测对象打标处理,最后整理为 VisiFire 数据集. 数据集具体内容如表 3 所示.

表 3 数据集描述

Table 3 Dataset description

Item	FireData	VisiFire	Total
Pictures	5267	1163	6430
Objects	9274	3740	13014

2.3 对比试验

本文将对对比算法模型与本文模型在公开的数据集 ImageNet 上进行预训练,然后在 FireData 数据集上进行微调,最后在建立的 VisiFire 数据集上进行检测. 选取各模型性能最佳的一轮进行比较. 以 mAP (mean Average Precision)作为评价标准,从表 4 可以看出将 Yolov2 的最大池化层全部替换为跨步卷积后, mAP 相较于原模型提高了 2.3%.

从表 5 可以看出在替换最大池化层并添加反卷积模块后,相同的场景下,本文的模型检测率更高,改进后模型的 mAP 相较原模型提高了 10%,虽然比 Yolov4 低了 0.7%,但是模型复杂度远低于 Yolov4.

表 4 模型性能比较

Table 4 Comparison of model performance

Model	Yolov2	Proposed
mAP/%	70.2	72.3

表 5 模型性能比较

Table 5 Comparison of model performance

Model	Yolov2	Yolov3	Yolov4	Proposed
mAP/%	70.2	73.6	77.8	77.2

2.4 视频实验结果

实验测试集取自 Bilkent 大学火灾数据库 VisiFire 中的火灾视频,本文模型对于视频火焰的检测结果如图 3 所示. 将文献[10]、文献[12]、文献[13]和文献[15]的算法模型与本文模型进行对比,对比试验结果如表 6 和表 7 所示. 其中 TP 代表火焰检测率,FP 代表漏检率,TN 代表非火焰视频正确率,FN 代表误检率.

表 6 火焰视频的检测结果

Table 6 Detection results on flame videos

视频	总帧数	文献[10]		文献[12]		文献[13]		文献[15]		本文	
		TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
Video1	134	98.0	2.0	99.6	0.4	97.8	2.2	100.0	0.0	99.3	0.7
Video2	170	97.8	2.2	98.0	2.0	96.2	3.8	99.2	0.8	98.1	1.9
Video3	264	97.6	2.4	98.8	1.2	97.5	2.5	98.1	1.9	98.0	2.0
Video4	595	95.5	4.5	97.8	2.2	95.9	4.1	97.5	2.5	97.9	2.1
平均值	290.8	97.2	2.8	98.5	1.5	96.9	3.1	98.7	1.3	98.3	1.7

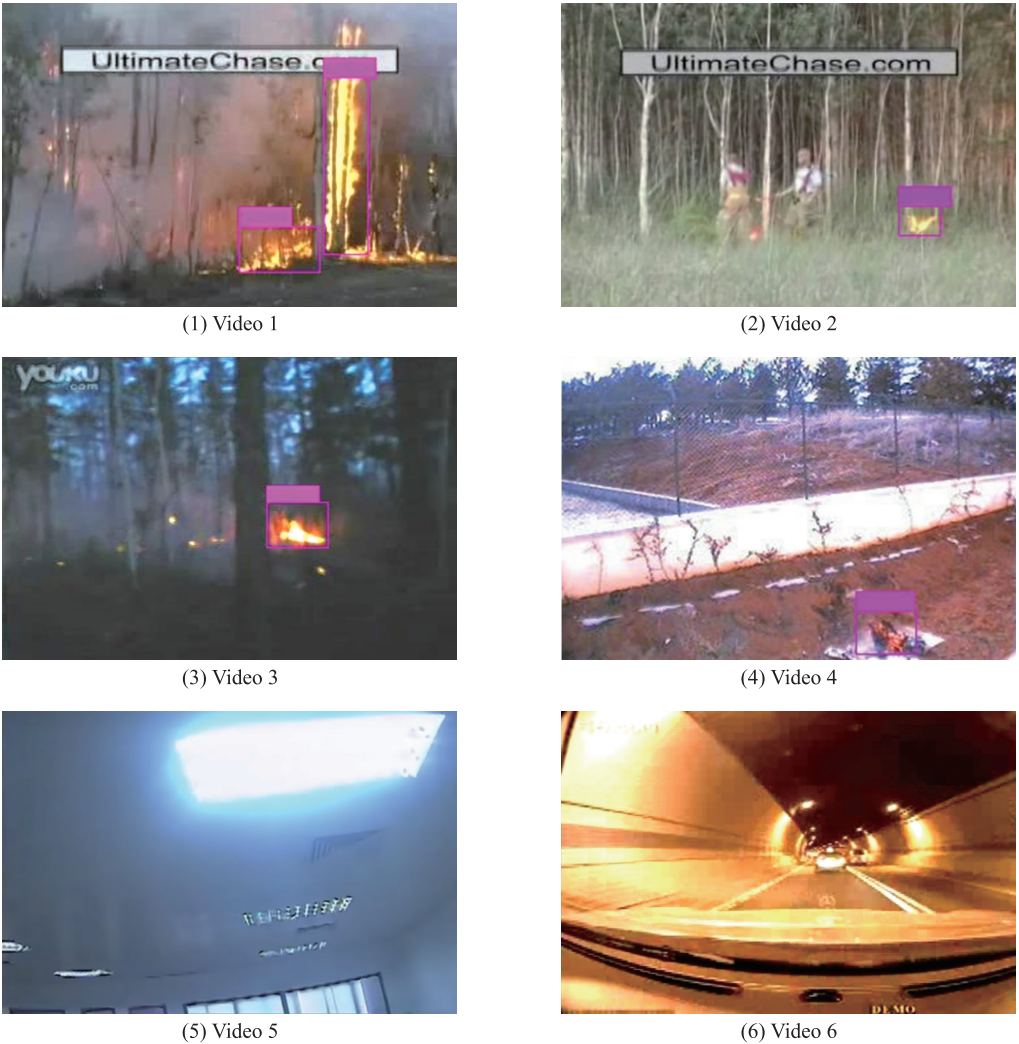


图 3 检测结果

Fig. 3 Detection results

表 7 非火焰视频的检测结果

Table 7 Detection results on non-flame videos

视频	总帧数	文献[10]		文献[12]		文献[13]		文献[15]		本文	
		TN	FN	TN	FN	TN	FN	TN	FN	TN	FN
Video5	600	96.7	3.3	98.0	2.0	95.3	4.7	97.1	2.9	97.9	2.1
Video6	100	96.0	4.0	97.2	2.8	96.1	3.9	97.4	2.6	98.3	1.7
平均值	350	96.3	3.7	97.6	2.4	95.7	4.3	97.3	2.7	98.1	1.9

文献[10]、文献[12]和文献[13]都是单阶段检测器,文献[15]是经典的双阶段检测器. 文献[10]的DSSD检测器采用了“沙漏”结构,以此来提高模型对于小目标的检测能力,虽然利用了侧连接将高层和低层语义信息做了融合,但是这种连接只存在于模型后半部分,并没有充分地利用模型的低层语义信息. 文献[12]是Yolo的最新版本Yolov4,Yolov4的精度远高于Yolov2,但是Yolov4的模型复杂度较高,对于设备的要求较高,一些实际场景难以应用. 文献[13]中的Yolov2检测器相比于Yolo加入了passthrough layer,简单地说,就是将 26×26 的feature map和本层 13×13 的feature map进行连接,目的是为了提高小目标检测率,但是经过了多次的下采样,更多的特征信息极有可能在最大池化层就已经丢失了,这一层并不具有代表性,所以Yolov2的小目标检测并不那么理想,如Video2和Video4. 文献[14]中的Faster R-CNN有较高的正检率,这是双阶段检测器最大的优点,尤其是在Video4这种检测目标和背景差异较小的场景下,Faster R-CNN的检测率远高于其他单步检测器,但是由于模型中包括特征金字塔和全连接层等复杂结构,计算量较大,对于设备性能要求特别高,并且很难满足实时检测要求.

本文基于 YOLOv2 改进的网络模型,通过去掉大量的最大池化层,尽可能地保留了原图的特征信息,并且通过反卷积模块和 concat 将高层和低层的特征信息融合起来,在 Video2、Video3 中对于小目标的检测效果较好,并且在 Video4 中目标与背景差异较小的情况下,也有较好的检测率。

3 结论

本文以 Darknet-19 为骨干网,提出引入反卷积模块,通过连接层融合多级特征信息,一定程度上提高了模型对于小目标的检测率;通过以跨步卷积代替最大池化层对图像进行采样,最大程度上避免了特征信息的丢失,提高了模型整体的检测精度。该模型对于设备配置要求较低,实用性强,可以满足多场合火焰检测需求。

[参考文献]

- [1] 洪柳. 建筑火灾事故损失特性研究[J]. 建筑安全,2020,35(1):59-64.
- [2] HOWARD A G,ZHU M,CHEN B,et al. Mobilenets:efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861,2017.<https://arxiv.org/abs/1704.04861>.
- [3] SHARMA J,GRANMO O C,GOODWIN M. Deep CNN-ELM hybrid models for fire detection in images[C]//International Conference on Artificial Neural Networks. Cham:Springer,2018:245-259.
- [4] LEE W,KIM S,LEE Y T,et al. Deep neural networks for wild fire detection with unmanned aerial vehicle[C]//IEEE International Conference on Consumer Electronics. Las Vegas,USA:IEEE,2017.
- [5] ZHANG D,HAN S,ZHAO J,et al. Image based forest fire detection using dynamic characteristics with artificial neural networks[C]//Iita International Joint Conference on Artificial Intelligence. Hainan Island,China,2009:290-293.
- [6] 黄文锋,徐珊珊,孙焱等. 基于多分辨率卷积神经网络的火焰检测[J]. 郑州大学学报(工学版),2019,40(5):80-84.
- [7] 江洋,白勇. 基于 RetinaNet 深度学习模型的火焰检测研究[J/OL]. 海南大学学报(自然科学版):1-7[2019-12-12]. <http://kns.cnki.net/kcms/detail/46.1013.N.20191119.1528.010.html>.
- [8] LIU W,ANGUELOV D,ERHAN D,et al. SSD:single shot multibox detector[C]//European Conference on Computer Vision. Cham:Springer,2016:21-37.
- [9] 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. Las Vegas,USA:IEEE,2016:779-788.
- [10] FU C Y,LIU W,RANGA A,et al. Dssd:deconvolutional single shot detector[J]. arXiv preprint arXiv:1701.06659,2017.
- [11] REDMON J,FARHADI A. YOLOv3:an incremental improvement[J]. arXiv preprint arXiv:1804.02767,2018.
- [12] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. YOLOv4:optimal speed and accuracy of object detection[J]. Computer vision and pattern recognition,2020,17(9):198-215.
- [13] REDMON J,FARHADI A. YOLO9000;better,faster,stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu,USA:IEEE,2017:7263-7271.
- [14] 张旭,李建胜,郝向阳,等. 基于差分筛选的 YOLOv2 监控视频目标检测方法[J]. 测绘科学技术学报,2018,35(6):616-621.
- [15] REN S,HE K,GIRSHICK R,et al. Faster R-CNN:towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems. Montreal,2015:91-99.
- [16] 黄同愿,向国徽,杨雪姣. 基于深度学习的行人检测技术研究进展[J]. 重庆理工大学学报(自然科学),2019,33(4):98-109.

[责任编辑:陆炳新]