

控制与决策

Control and Decision

复杂背景下全景视频运动小目标检测算法

王殿伟, 杨旭, 韩鹏飞, 刘颖, 谢永军, 宋海军

引用本文:

王殿伟, 杨旭, 韩鹏飞, 等. 复杂背景下全景视频运动小目标检测算法[J]. *控制与决策*, 2021, 36(1): 249–256.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.0686>

您可能感兴趣的其他文章

Articles you may be interested in

基于操作风险的双模式传感器管理方法

A dual-mode sensor management method based on operational risk

控制与决策. 2020, 35(12): 2993–2998 <https://doi.org/10.13195/j.kzyjc.2018.1541>

基于免疫优化的平面Acrobot线性自抗扰鲁棒镇定

Robust stabilization of planar Acrobot using linear active disturbance rejection control with immune optimization

控制与决策. 2020, 35(12): 3053–3058 <https://doi.org/10.13195/j.kzyjc.2019.0289>

融合长短时记忆机制的机械臂多场景快速运动规划

Multi-scene rapid motion planning combining with long and short time memory mechanisms for manipulators

控制与决策. 2020, 35(12): 2968–2976 <https://doi.org/10.13195/j.kzyjc.2018.1387>

基于图像和高程数据的天际线定位匹配

Skyline position matching based on image and elevation data

控制与决策. 2020, 35(11): 2665–2674 <https://doi.org/10.13195/j.kzyjc.2019.0155>

一种基于稀疏系数匹配学习的图像去雾算法

An image dehazing method based on learning framework with sparse coefficient matching

控制与决策. 2020, 35(11): 2797–2802 <https://doi.org/10.13195/j.kzyjc.2018.1764>

复杂背景下全景视频运动小目标检测算法

王殿伟¹, 杨旭^{1†}, 韩鹏飞², 刘颖¹, 谢永军³, 宋海军³

(1. 西安邮电大学 通信与信息工程学院, 西安 710121; 2. 西湖大学 人工智能研究与创新中心, 杭州 310024; 3. 中国科学院西安光学精密机械研究所, 西安 710119)

摘要: 为解决复杂背景下全景视频中运动小目标检测精度低的问题, 提出一种基于复杂背景下全景视频运动小目标检测算法. 首先, 为降低复杂背景信息的干扰, 提高目标检测的精度, 采用快速鲁棒性主成分分析 (Fast RPCA) 算法将全景视频图像的前景背景信息分离, 并提取出前景信息作为有效的图像特征; 然后, 改进更快的基于区域的卷积神经网络 (Faster R-CNN) 中的区域生成网络 (RPN) 的候选框尺度大小, 使之适应全景图像中的目标尺寸, 再对前景特征图进行训练; 最后, 通过 RPN 网络和 Fast R-CNN 网络共享卷积层输出检测模型, 实现对全景视频图像中小目标的精准检测. 实验结果表明, 所提出算法可以有效抑制复杂的背景信息对目标检测精度的影响, 并对全景视频图像中的运动小目标具有较高的检测精度.

关键词: 全景图像; Fast RPCA; Faster R-CNN; 目标检测

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.0686

开放科学(资源服务)标识码(OSID):



引用格式: 王殿伟, 杨旭, 韩鹏飞, 等. 复杂背景下全景视频运动小目标检测算法[J]. 控制与决策, 2021, 36(1): 249-256.

Panoramic video motion small target detection algorithm in complex background

WANG Dian-wei¹, YANG Xu^{1†}, HAN Peng-fei², LIU Ying¹, XIE Yong-jun³, SONG Hai-jun³

(1. School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 2. Center for AI Research and Innovation, Westlake University, Hangzhou 310024, China; 3. Xi'an Institute of Optics and Precision Mechanics of CAS, Xi'an 710119, China)

Abstract: In order to solve the problem of low detection accuracy of moving small targets in the panoramic video in complex background, a small target detection algorithm based on complex background motion is proposed. Firstly, to reduce the interference of complex background information and improve the accuracy of target detection, the fast robust principal component analysis (Fast RPCA) algorithm is used to separate the foreground background information of the panoramic video image, and the foreground information is extracted as an effective image feature. Then, the candidate frame size of the region proposal network (RPN) in the faster region-convolutional neural networks (Faster R-CNN) is improved to adapt to the target size in the panoramic image, and then the foreground feature map is trained. Finally, the convolutional layer output detection model is shared by the RPN network and the Fast R-CNN network to achieve accurate detection of small targets in the panoramic video image. Experiments show that the proposed algorithm can effectively suppress the influence of complex background information on target detection accuracy, and has high detection accuracy for small moving targets in panoramic video images.

Keywords: panoramic image; Fast RPCA; Faster R-CNN; target detection

0 引言

目标检测技术是在给定图像的情况下找到一个或多个对象类的所有实例, 通过检测网络减少对人力资本的消耗. 全景视觉系统最大限度地保留了场景

的真实性, 在全景视频图像下进行目标检测, 不会因为视角盲区造成目标遗漏. 目前, 全景视频图像目标检测技术已广泛地应用于智能视频监控、无人驾驶、生物特征识别等诸多领域, 具有重要的现实意义.

收稿日期: 2019-05-17; 修回日期: 2019-09-09.

基金项目: 陕西省自然科学基金基础研究计划科技创新“双导师制”项目(2018JM6118); 西安邮电大学创新创业基金项目(2018SC-08); 西安邮电大学研究生创新基金项目(CXJJLY2018038).

责任编辑: 张海涛.

†通讯作者. E-mail: 735583415@qq.com.

近几年来,随着计算机性能和硬件的快速发展,深度学习方法也随之崛起。现有的目标检测算法可以分为两级检测算法^[1-5]和单级检测算法^[6-9]两大类^[10]。两者的不同之处是:两级检测算法在第1阶段生成对象提议,在第2阶段分类提议并重新确定坐标,这类方法的侧重点是提升检测精度,但检测速度较低;而单级目标检测算法跳过对象提议生成,在一次评估中预测边界框和类别分数,检测精度比两级检测算法略低,但检测速度却有了长足进步。

两级检测算法一直是目标检测的主流算法,包括基于候选区域(region proposal)的R-CNN^[1-2]、SPP-NET(spatial pyramid pooling net)^[3]、Fast R-CNN^[4]、Faster R-CNN^[5]等两级目标检测算法。在早期的研究^[1-2,4]中,对象提议由外部模块生成^[11-12],后来经过发展,两级检测算法变得更加准确和高效。R-CNN先提取每个候选区域,然后输入到卷积神经网络(CNN)分类器中以实现特征提取,重复工作较多。此外,由于网络中全连接层^[13]的存在,要求输入的候选区域为固定尺寸大小,容易造成目标畸变,影响检测结果。SPP-NET和Fast R-CNN引入区域特征提取的思想,显著加快了整个探测器的速度,同时调整候选框输入可大可小,从而避免了R-CNN目标形变等问题。2017年提出的Faster R-CNN检测算法,针对候选区域计算量大的问题,引入了区域提议网络RPN(region proposal network)^[5]。它使用完全卷积网络将每个滑动窗口的anchor映射到一个低维特征,这一改进解决了CNN系列算法在区域推荐费时的难题,在后来的两级检测算法中被广泛采用^[14-16]。

以YOLO(you only look once)^[6]、SSD(single shot multi-box detector)^[7]为代表的单级目标检测算法省略了提议步骤,但仍使用滑动窗口产生anchor。SSD和DenseBox^[8]从特征映射中密集生成anchor,并将其评估为多类RPN;RetinaNet^[9]与SSD有许多相似之处,增加了Focal Loss^[9]和特征金字塔网络(feature pyramid networks, FPN)^[16]来解决前景-背景不平衡的问题;YOLO v2采用滑动窗生成anchor来进行分类和位置预测。这些检测算法在单目图像上的效果越来越好,但是应用到全景图像时往往效果不佳。

全景图像具有视角广、包含信息较多等优点,但是单张图像分辨率较大,背景复杂,而且待检测目标在图像中大多为小目标,容易与背景信息混淆,给检测带来难度,而现有深度学习算法不能有效地把目标与背景区分开,容易产生漏检和误检现象。为解决这一问题,本文提出一种改进的Faster R-CNN的全景视

频图像运动小目标检测算法。

本文的贡献主要包括以下几个方面:

1) 为了降低背景信息对目标检测的干扰,提高检测精度,采用Fast RPCA(fast robust principal component analysis)^[17]算法分离全景视频图像的前景背景信息,有助于后续的目标检测与分类;

2) 针对全景图像中目标容易发生形变、失真、小目标较多的情况,改进Faster R-CNN中RPN网络的候选框尺度大小,使之适应全景图像中的目标尺寸,再对前景特征图进行训练;

3) 为了克服普通图像目标检测算法不适用于全景图像的问题,把Fast RPCA与Faster R-CNN两种方法相结合,有效解决全景视频图像背景复杂、小目标较多、难以检测的问题。

通过实验表明,所提出算法对全景视频图像中运动小目标进行检测时,可以较好地解决现有基于Faster R-CNN算法存在的漏检误检问题,降低全景视频图像中背景干扰信息对检测精度的影响,从而提高全景视频图像中运动小目标的检测精度。

1 Fast RPCA+改进Faster R-CNN的全景视频运动小目标检测算法

普通的单摄像机获取的视频中的目标在图像中占据较大的比例,而且待检测的目标尺寸比较符合anchor候选框的大小,因此,现有的Faster R-CNN算法能够在单目图像中取得较高的检测精度。但应用于全景视频图像时,因为图像分辨率过大,且全景视频图像相比普通视频图像背景更加复杂,所以检测效果就会变差;并且待检测目标大多是小于anchor尺寸范围,经常出现误检、漏检现象。针对以上问题,本文提出一种Fast RPCA+改进Faster R-CNN的全景视频运动小目标检测算法。首先采用Fast RPCA算法将全景视频图像的前景背景信息分离,并提取出前景信息作为有效的图像特征;然后改进Faster R-CNN中RPN网络的候选框尺度大小,使之适应全景图像中的目标尺寸,再对前景特征图进行训练;最后,通过RPN网络和Fast R-CNN网络共享卷积层输出检测模型,实现对全景视频图像中小目标的精准检测。该模型对全景视频图像有着较高的检测精度,并且对全景视频图像的固有失真、待检测目标较小等缺点有较好的鲁棒性。

本文算法由特征提取模块和目标检测模块构成,所提出算法流程如图1所示。首先采用Fast RPCA算法将全景视频图像的前景与背景信息进行分离,提取出前景信息,从而减少背景等无关信息对全景视频图

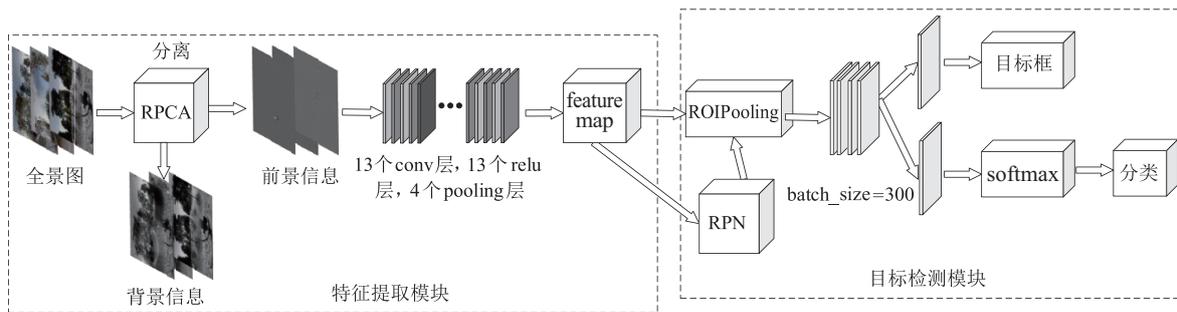


图1 本文算法网络模型

像中运动小目标检测效果的影响;然后将提取的前景信息经过卷积层的处理,输出图像的特征图;最后,针对待检测目标尺寸与anchor尺寸不匹配的问题,通过改进Faster R-CNN中的RPN网络,得到最适合全景视频图像目标尺寸的检测框,从而实现对全景视频图像中小目标的检测。

1.1 前景信息提取

由于原始全景视频图像容易受到光照影响,背景信息较为复杂,导致全景视频图像中小目标检测精度过低.为了更好地获得全景视频图像中待检测目标的有效特征信息,减少复杂场景下的环境信息对检测精度的影响,准确地提取图像的特征信息尤为重要,本文采用Fast RPCA对全景视频进行前景背景分离。

RPCA (robust principal component analysis)^[18]的基本理论是,一组图像的观测矩阵 D 可以被分解为包含背景信息的低秩矩阵 L 和前景信息的稀疏矩阵 S ,即 $D = L + S$,稀疏矩阵提取出的特征就是原图像的显著信息. RPCA算法最初可以表示为

$$\begin{aligned} \arg \min_{L,S} \|L\|_* + \lambda \|S\|_1; \\ \text{s.t. } D = L + S. \end{aligned} \tag{1}$$

其中: $D \in R^{m \times n}$ (m 为帧行数乘以帧列数的积, n 为视频帧数) 是原始图像矩阵, $\|L\|_*$ 是矩阵 L 的核范数,参数 $\lambda = 1/\sqrt{\max(m,n)}$, $\|S\|_1$ 是矩阵 S 的 ℓ^1 范数。

由于RPCA是通过复杂的优化来计算的,对于大小为 2048×1024 的全景视频中的210个帧序列,大约需要40min来处理,速度过慢,因此,本文采用Fast RPCA算法来对全景视频图像进行前背景分离. Fast RPCA算法可以表示为

$$\begin{aligned} \arg \min_{L,S} \frac{1}{2} \|L + S - D\|_F + \lambda \|S\|_1; \\ \text{s.t. } \|L\|_* \leq t, \end{aligned} \tag{2}$$

其中 t 是低秩矩阵 L 的秩。

Fast RPCA可以显著加快全景视频前背景分离的时间,对于大小为 2048×1024 的全景视频中的210

个帧序列,大约需要178s来处理。

如图2所示,采用Fast RPCA算法对全景视频图像进行处理后,分离出不包含待检测目标的背景信息,留下的前景信息包含了车辆、行人等感兴趣的待检测目标,把提取出的前景信息输入到下一层网络中,有效解决了全景视频图像包含信息量过大、容易受环境干扰的问题,进而对检测精度有所提高。

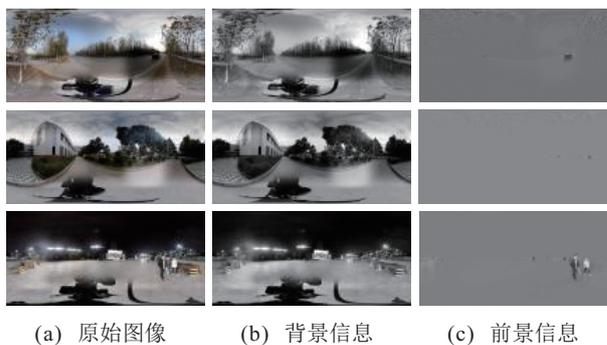


图2 Fast RPCA实现全景视频图像前景背景分离结果

1.2 Faster R-CNN算法

Fast RPCA输出的前景信息依然包含着干扰信息,为了更好地提升检测精度,本文采用改进的Faster R-CNN算法对前景图像进行训练. Faster R-CNN在卷积层中进一步进行特征提取,通过RPN输出的候选区域和卷积层的特征图一起进入感兴趣区域池化层。

Faster R-CNN算法改进了Fast R-CNN算法利用选择性搜索(selective search, SS)来提取候选框,针对候选区域计算复杂的问题,使用更先进的RPN网络来提取候选框,RPN网络提取的候选框更少,效率更高.对于输入进来的图像,首先进行一系列卷积层和池化层提取出样本的特征图,然后检测网络和RPN网络共用全图像的特征图,有效解决了候选区域计算量大的问题。

Faster R-CNN算法的损失函数为

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) +$$

$$\lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (3)$$

$$L_{\text{cls}}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)], \quad (4)$$

$$L_{\text{reg}}(t_i, t_i^*) = R(t_i - t_i^*), \quad (5)$$

$$\text{IoU} = \frac{(A \cap B)}{(A \cup B)}. \quad (6)$$

其中:IoU(intersection-over-union)为两个检测框A、B的重叠度, i 表示anchor的索引值, p_i 表示预测anchor为目标的概率, p_i^* 表示对应的真实值(ground truth)的预测概率, t_i 代表预测边界框, t_i^* 代表对应的前景anchor所对应的真实值的标记框, L_{cls} 是RPN分类损失层计算的平滑损失, L_{reg} 是RPN边界框回归计算的正则化 L_1 损失, N_{cls} 和 N_{reg} 是两个归一化值, λ 为平衡权重。

相比较 L_1 损失函数, smooth_{L_1} 损失函数对离群点、异常值不敏感,鲁棒性更好,可以使训练过程

更加稳定,所以在Faster R-CNN算法的损失函数中采用了 smooth_{L_1} ,其中式(5)中的 R 是鲁棒损失函数 smooth_{L_1} ,即

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1; \\ |x| - 0.5, & \text{otherwise.} \end{cases} \quad (7)$$

RPN网络可以把任意分辨率大小的图片作为输入,这样就解决了全景视频图像因分辨率较大而难以输入的问题.如图3所示,对于卷积层输出到RPN网络的特征图,使用 3×3 的卷积核对特征图进行处理,这个区域进行卷积计算后可以生成一个三维矩阵的数据结构.每一个滑动窗口都对应一个256维的特征,可以预测得到多个候选区域,其中每个目标位置的候选区域数量用 k 表示.在回归层有 $4k$ 个输出出来表示 k 个预测框的坐标,在分类层输出 $2k$ 个得分来估计每个候选区域是否为目标的概率,本文把这些区域称为anchor.

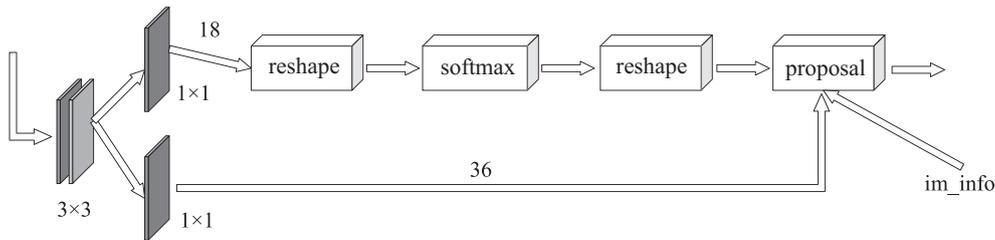


图3 RPN网络流程

原始的Faster R-CNN对anchor范围内的目标检测效果较好,当待检测目标小于anchor范围时,检测效果就会变差.如图4所示,在夜间、白天、傍晚3种光照条件下,远景小目标均不能有效检测到,只有远景尺寸稍大的目标可以被检测到。



图4 原始Faster R-CNN检测效果

1.3 区域候选框

anchor是现代物体检测技术重要的组成,在空间域上均匀采样,具有预定义的尺度和纵横比.anchor用于预测建议或回归最终边界框,目前比较先进的目标检测算法主要依赖于生成密集anchor.合理的anchor设计有两个规则:对齐和一致性^[19].首先,要使用卷积特征作为anchor表示,anchor中心需要与特征映射像素很好地对齐;其次,不同位置的anchor的尺度和形状应该是一致的.滑动窗口是应用广泛的anchor生成方案,对于大多数检测方法,特征图中的

每个位置与 k 个anchor相关联。

Faster R-CNN是从一组密集的anchor生成区域提议,然后将它们分类为特定的类别,并通过边界框回归来确定它们的位置.因为Faster R-CNN对anchor尺寸范围内的目标检测效果较好,范围之外则较差,全景视频图像尺寸较大,但单一目标所占像素点较少,所以本文把原网络中的anchor尺寸进行改进:首先,每个anchor对应的每个输入图像为3种尺度(64、128、256),每个尺度也会产生3种长宽比(1:1、1:2、2:1),即每个anchor可以产生9个目标候选框;然后,将产生的所有anchor输入到RPN网络的分类层和回归层,分别用于目标分类和回归;最后,依据候选区域的得分高低,把前256个候选区域作为Fast R-CNN网络的输入进行全景目标检测。

单一地降低anchor尺寸来检测原始全景图,不但不能有效检测出远景小目标,而且会把大量anchor范围内的背景干扰误检为目标,如远处的树木、建筑等,所以在缩小anchor尺寸的同时,运用Fast RPCA算法对全景视频图像进行前背景分离,直接在前景图像上进行训练检测.如图5所示,对全景视频图像进行前

背景分离后,在提取出的前景图像上训练检测,因为没有背景信息的干扰,降低了目标检测的难度,所以缩小 anchor 尺寸后,本文算法不但可以检测到近景大目标,而且可以有效地对远景小目标进行检测。

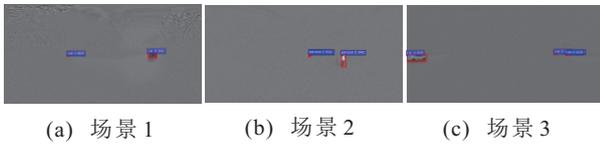


图5 改进的Faster R-CNN检测前景图像效果

2 实验结果及分析

文中所有训练及测试数据均为泰科易720 Pro七目全景相机采集所得,数据库共有不同场景全景图像2654张,图像分辨率大小为2048×1024。实验平台为Tensorflow,计算使用CPU(Intel Xeon E5-2620v4×2)和GPU(Nvidia Titan XP×4)。对全景图和特征图采用VOC 2007格式进行标注。其中66%的图像用于训练,34%用于测试。

2.1 损失函数

损失函数是评估模型稳定性的重要指标,用来表示预测与实际数据的差距程度。当训练次数达到一定次数时,损失函数值如果变得稳定,不再发生剧烈

变化,则所训练的模型已经趋于稳定。

图6为检测网络训练过程中损失函数的收敛曲线。其中:横坐标表示训练迭代次数,最大训练次数为70000;纵坐标表示总体的损失值。当网络训练次数超过60000时,各项参数变化趋于平稳,最终的损失值稳定在0.36左右。从损失函数收敛情况来看,训练结果较为理想。

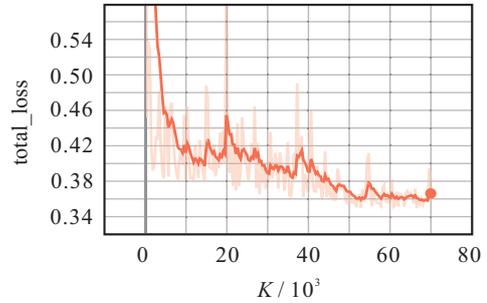


图6 损失函数曲线

2.2 算法结果对比

为了更好地评价本文所提出算法对于全景视频图像中运动小目标的检测效果,将其与现有的Faster R-CNN^[5]算法进行比较。图7为Faster R-CNN^[5]算法和本文算法对部分场景的检测结果。

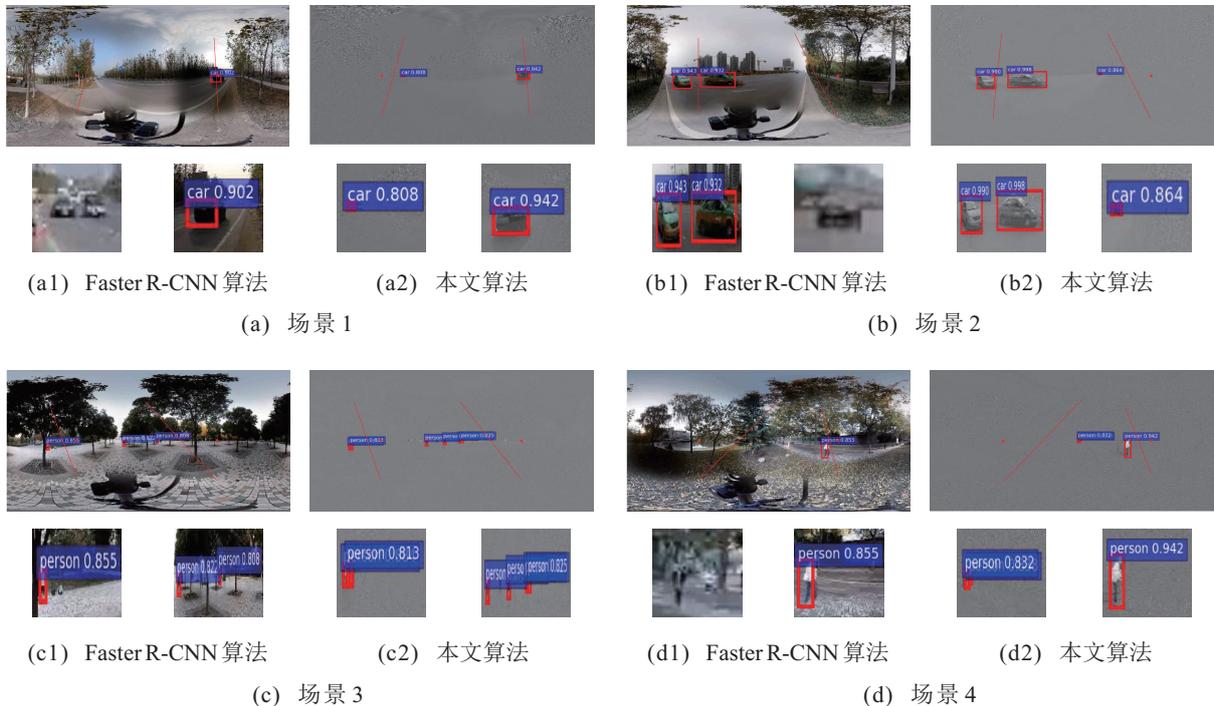


图7 正常光照全景图像目标检测

对所选出的256个 anchor 进行分类和回归时,将每一个真实值与它重叠度超过0.7的 anchor 作为正样本,与真实值重叠度小于0.3的 anchor 作为负样本,0.3~0.7之间的 anchor 不参与训练,采用查准率、

查全率计算出的平均准确率来进行评价,即

$$P_{re} = \frac{T_P}{T_P + F_P}, R_{ec} = \frac{T_P}{T_P + F_N}; \quad (8)$$

$$mAP = \int_0^1 P_{re}(R_{ec})d(R_{ec}). \quad (9)$$

其中: T_P 是正确的正样本, F_P 是错误的正样本, F_N 是错误的负样本, P_{re} 是查准率, R_{ec} 是查全率, mAP 是平均准确率. 实验中每个输出框都与一个类别签和 $[0, 1]$ 之间的 softmax 分数相关联, 使用 0.6 的分数阈值来显示这些图像.

由图7可见: 采用 Faster R-CNN 算法对正常光照图像的检测结果及其检测目标的放大图表明, 测试数据中的近景目标均被检测到, 但远景中的小目标无法准确识别和检测, 出现了漏检现象; 而采用本文算法

的检测结果及其检测目标的放大图表明, 本文算法对全景视频图像中小目标检测效果较好, 测试所用的4张前景图像中的所有目标均被检测到.

表1给出了对同一图像分别采用两种算法的结果比较. 表1中结果为“未检出的”表示该目标检测失败, 其余数字代表该目标的检测准确率. 由表1可以看出, 对于采用 Faster R-CNN 算法检测到的目标, 在采用本文算法后, 对应目标的检测准确率有了小幅度提高.

表1 Faster R-CNN算法和本文算法结果对比

场景	Faster R-CNN 算法 mAP				本文算法 mAP			
	目标 a	目标 b	目标 c	目标 d	目标 a	目标 b	目标 c	目标 d
1	未检出	0.902	—	—	0.808	0.942	—	—
2	0.943	0.932	未检出	—	0.990	0.998	0.864	—
3	0.855	未检出	0.822	0.808	0.884	0.813	0.879	0.826
4	未检出	未检出	0.855	—	0.832	0.838	0.942	—



图8 低照度全景图像目标检测结果

图8为Faster R-CNN算法和本文算法对低照度全景图像目标的检测结果. 可以看出: 在低照度全景图像中, 待检测目标容易与背景融为一体; 而经过Fast RPCA算法前背景分离后的前景图像中, 只留下了待检测的运动目标, 这些目标在检测过程中将不再受不良光照的影响. 对比两种算法检测低照度全景图像的结果可以看出: Faster R-CNN算法检测低照度图像时只能检测到部分颜色谱段与背景差异较大的目标, 对于目标与背景相似度较高的颜色谱段, 容易出现漏检; 而本文算法因为是在Fast RPCA分离出的前景图像上检测, 目标较为明显, 所以本文算法可以准确地检测到各颜色谱段目标.

为了更加客观地对检测效果进行评价, 选用检测准确率作为评价指标, 所用数据集均为2 654张全景图像, 检测类别分为3类, 分别为机动车、行人、非机动车, 以下准确率为3类目标的平均准确率.

如表2所示: 当输入为原始全景图像时, 由于其较大的分辨率和复杂的背景干扰, 导致检测准确率比输入为Fast RPCA前景图低了3.8%; 本文算法适应全景图像多为小目标的特点, 通过采用Fast RPCA算法减少背景信息的影响, 从而提高算法检测率. 最终, 本文算法的平均检测率为80.6%.

表2 Faster R-CNN算法和本文算法结果对比

输入图像 (2048×1024)	Faster R-CNN	本文算法
原始图像 mAP	0.752	0.768
Fast RPCA 前景图 mAP	0.790	0.806

以开源深度学习框架ResNet 101为基础, 改进的Faster R-CNN算法为模型, 训练基于Faster R-CNN的全景图像运动目标检测器. 由表3可得, 相比较VGG 16框架, ResNet 101精度提升近2%, 但单张照片测试时间增加了0.04s.

表3 ResNet101和VGG16网络运行结果及时间对比

指标	ResNet101	VGG16
输入图像 (2048×1024) mAP	0.806	0.788
单张图片测试时间/s	0.169	0.130

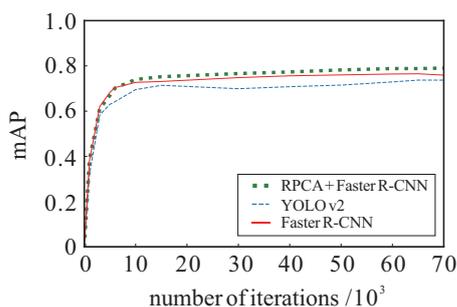


图9 Faster R-CNN、YOLO v2和本文算法对比

将本文算法与Faster R-CNN^[5]算法、YOLO v2^[6]算法的检测准确率进行比较分析, 结果如图9所示. 可以看出, 本文算法相比较其他两种算法, 在全景图像上的检测准确率超过80%, 检测效果有明显提升.

3 结论

本文提出了一种改进的Faster R-CNN全景视频图像运动小目标检测算法. 首先, 采用Fast RPCA算法将全景视频图像的前背景信息进行分离, 提取出有效的图像特征; 然后, 利用改进的Faster R-CNN对含有图像有效信息的前景特征图进行训练; 最后, 通过RPN网络和Fast R-CNN网络共享卷积层输出检测模型, 实现对全景图像中小目标的精准检测.

实验结果表明, 所提出算法可以减少背景信息对检测目标的干扰, 实现对全景图像中小目标的精确检测. 但由于算法复杂度太高, 导致运算时间较长, 无法实现对全景图像中小目标的实时检测, 因此, 如何实现实时检测将是今后的研究重点.

参考文献 (References)

- [1] 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. Columbus: IEEE, 2014: 580-587.
- [2] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(1): 142-158.
- [3] He K M, Zhang X Y, Ren S R, 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.
- [4] Girshick R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1440-1448.
- [5] 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.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779-788.
- [7] Wei Liu, Dragomir Anguelov, Dumitru Erhan. SSD: Single shot multibox detector[C]. European Conference on Computer Vision (ECCV). Amsterdam: Springer International Publishing, 2016: 21-37.
- [8] Huang L C, Yang Y, Deng Y F, et al. Densebox: Unifying landmark localization with end to end object

- detection[J]. Computer Science, 2015, arXiv preprint arXiv: 1509.04874.
- [9] Lin T, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2999-3007.
- [10] 李会军, 王瀚洋, 李杨, 等. 一种基于视觉特征区域建议的目标检测方法[J]. 控制与决策, 2020, 35(6): 1323-1328.
(Li H J, Wang H Y, Li Y, et al. Target detection method based on visual feature area suggestion[J]. Control and Decision, 2020, 35(6): 1323-1328.)
- [11] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [12] Zitnick C L, Dollar P. Edge boxes: Locating object proposals from edges[C]. European Conference on Computer Vision. Zurich: IEEE, 2014: 391-405.
- [13] Agrawal P, Girshick R, Malik J. Analyzing the performance of multilayer neural networks for object recognition[C]. European Conference on Computer Vision. Zurich: Springer, 2014: 329-344.
- [14] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]. Conference on Neural Information Processing Systems. Barcelona, 2016: 379-387.
- [15] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2980-2988.
- [16] 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 (CVPR). Honolulu: IEEE, 2017: 936-944.
- [17] Rodríguez P, Wohlberg B. Fast principal component pursuit via alternating minimization[C]. 2013 IEEE International Conference on Image Processing (ICIP). Melbourne: IEEE, 2013: 69-73.
- [18] Candès E J, Li X D, Ma Y, et al. Robust principal component analysis?[J]. Journal of the ACM, 2011, 58(3): 1-37.
- [19] Wang J, Chen K, Yang S, et al. Region proposal by guided anchoring[C]. 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Angeles: IEEE, 2019: 2960-2969.

作者简介

王殿伟(1978—), 男, 副教授, 博士, 从事图像清晰化、目标检测、识别等研究, E-mail: wangdianwei@126.com;

杨旭(1995—), 男, 硕士生, 从事基于深度学习的目标检测的研究, E-mail: 735583415@qq.com;

韩鹏飞(1994—), 男, 硕士, 从事图像处理、人脸识别的研究, E-mail: hanpengfei@westlake.edu.cn;

刘颖(1972—), 女, 教授级高工, 博士, 从事图像检索、视频语义分析等研究, E-mail: ly_yolanda@sina.com;

谢永军(1974—), 男, 研究员, 博士, 从事衍射光学和遥感图像等研究, E-mail: yjxie@opt.ac.cn;

宋海军(1976—), 男, 副研究员, 博士, 从事光纤通信、量子光学等研究, E-mail: shjwl123@sina.com.

(责任编辑: 李君玲)

下 期 要 目

- 机器视觉在轨道交通系统状态检测中的应用综述 魏秀琨, 等
- 数据驱动的综合能源系统运行优化方法研究综述 陈 龙, 等
- Anchor-free的尺度自适应行人检测算法 邹逸群, 等
- 基于多班教学优化的多目标分布式混合流水车间调度 雷德明, 等
- 求解约束优化问题的改进果蝇优化算法及其工程应用 石建平, 等
- 标签Petri网的路径信息在故障诊断中的应用 叶丹丹, 等
- 基于Stacking的高炉铁水质量区间预测 周 平, 等
- 基于多维泰勒网的超前 d 步预测模型 李晨龙, 等
- 空间机器人捕获航天器操作的避撞柔顺复合自抗扰控制 艾海平, 等
- 纵向速度和艏向角受限的水面艇有限时间协同路径跟踪 李莉莉, 等
- 基于混合差分遗传算法的Bouc-Wen迟滞模型辨识策略 李自成, 等
- 四旋翼无人机抗干扰轨迹跟踪控制 高俊山, 等