

基于 Faster RCNN 的穿越围栏违规行为检测^①



王志鹏, 王 涛

(中国石油大学(华东) 计算机科学与技术学院, 青岛 266580)
通信作者: 王志鹏, E-mail: wangzp0512@163.com

摘 要: 安全围栏在电力施工现场扮演着重要的角色, 但穿越围栏的违规行为普遍存在, 给施工现场造成了极大的安全隐患. 为了实现智能化监管, 本文利用目标检测算法, 结合帧差法的思想提出了一种基于 Faster RCNN 的穿越围栏违规行为检测方法. 该方法通过读取视频监控信息, 利用目标检测方法获取围栏的位置信息以及人体关键点的信息, 通过帧差判断法识别施工现场的违规动作. 经过实验验证, 该方法可以有效的检测出施工现场的穿越围栏违规行为, 并可以满足实时性的要求.

关键词: 目标检测; 帧差法; 动作识别; Faster RCNN; 穿越围栏; 深度学习

引用格式: 王志鹏, 王涛. 基于 Faster RCNN 的穿越围栏违规行为检测. 计算机系统应用, 2022, 31(4): 346-351. <http://www.c-s-a.org.cn/1003-3254/8423.html>

Faster RCNN-based Detection Method for Violations of Crossing Fences

WANG Zhi-Peng, WANG Tao

(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: Safety fences play an important role in power construction sites. However, violations of crossing the fence are widespread, causing great safety hazards to the construction sites. To intelligently supervise, this study proposes a Faster RCNN-based detection method for crossing fence that combines the object detection and the ideas of frame difference method. The proposed method first obtains the information of the fence location and human keypoints by object detection from the captured frames in the video and then recognizes the violations at the construction site with the frame difference method. The experiment results show that the method can effectively detect violations of crossing fences at construction sites and meet the real-time requirements.

Key words: object detection; frame difference method; action recognition; Faster RCNN; fence crossing; deep learning

1 背景

电力行业是国民经济的支柱产业, 也是支撑交通、金融、供水、供气等民生基础设施安全可靠运行的关键设施. 电力系统的安全与稳定对社会的发展有很重要的影响^[1], 因此应该积极防范电力安全风险, 提前对电力运维情况分析、降低该领域潜在的风险以及积极研发防御技术. 目前, 电力维护现场都会配备监控设备实时记录作业过程, 但是对施工安全的管控仍然

主要依赖于人工监察. 这种方式会造成管控效率低下、人力资源浪费、发展进程缓慢等情况, 并会伴随着一定的经济损失.

随着社会科学技术迅速发展, 人工智能取得了突破性进展. 各行业向自动化和智能化方向发展是实现工业 4.0 智能制造目标的必然趋势^[2]. 如今, 人工智能在工业生产的许多方面都有实际应用, 这使得用智能方法对电力施工现场进行安全管控有了现实可能. 智能安全

① 基金项目: 国家自然科学基金(62072469); 国家重点研发计划(2018YFE0116700); 山东省自然科学基金(ZR2019MF049); 中央高校基础研究基金(2015020031); 西海岸人工智能技术创新中心建设专项(2019-1-5, 2019-1-6); 上海可信工业控制平台开放项目(TICPSH202003015-ZC)

收稿时间: 2021-06-28; 修改时间: 2021-07-30; 采用时间: 2021-08-12; csa 在线出版时间: 2022-03-22

管控系统利用视频设备实时检测作业过程中的违规行为,不仅可以节省人力资源,而且可以实现全时段覆盖检测,降低作业过程中的安全风险,有效提高作业效率。

安全围栏在电力运维作业中扮演着重要的角色,可以保护工作人员免受外界因素干扰,保障无关人员的安全性。现有的围栏主要有菱形围栏、单柱双带围栏、旗子围栏、网状围栏等。在变电器维修、街头环网柜维护以及吊车等作业过程中,都需要提前布置围栏。但是许多作业人员安全意识较差,出入围栏时往往不按照布置围栏时预设出入口进出,而使直接跨越围栏、挪开围栏以及钻过围栏等,这些不规范的行为给施工作业造成了很大的安全隐患。现场监管人员由于自身的作业任务,以及作业现场覆盖面积广等问题,很难监测到每个违规行为,非常容易造成安全事故。针对这种情况,设计一款高效的智能化的穿越围栏违规行为检测方法势在必行,对违规行为及时警告与制止,增强作业过程中的安全性。

2 相关工作

2.1 电力安全管控

电力施工现场的安全管控涉及到方方面面^[3],但大多需要监管人员的现场监管,针对施工现场检查表逐一排查,效率较低且时效性较差。围栏是重要的安全工具,为了实现高效智能化管控,基于红外线的电子围栏应用于实际^[4],同时相对应的监管系统也逐步完善^[5]。随着人工智能高速发展,房凯也提出了基于深度学习的跨越围栏违规行为检测方法^[6],促进了监管方法的多样性。

2.2 目标检测

随着卷积神经网络的盛行,目标检测取得了突破性的进展。目前目标检测主要分为两大流派: anchor-based 和 anchor-free。Anchor-based 方法会在特征图上生成一系列的人为定义尺寸和宽高比的矩形框作为候选框,然后对这些候选框进行分类和回归进而得到目标区域。Ren 等人^[7]提出的 RPN (区域建议网络)改进了传统的 SS (selective search) 方法,极大地推进了 anchor-based 方法的发展,YOLO 也由 v1^[8] 的版本 anchor-free 改为了 anchor-based 方法。随后许多方法提出解决 anchor-based 方法中的缺点,如使用 nms 删除候选框和 Mask RCNN^[9] 提出的 ROI Align 解决 ROI pooling 中的量化取整问题。Anchor-based 方法还包括: SSD^[10], FPN^[11], RetinaNet^[12] 等。Anchor-free 方法则是摒弃了 anchor 的思想,利用关键点进行回归获取目标

框。2018 年, Law 等人^[13] 提出 CornerNet 通过预测左上角和右下角的角点检测目标,随后 ExtremeNet^[14] 改进 CornerNet 通过预测极值点和中心点来预测目标,类似的还有 CenterNet^[15] 预测角点以及中心点。同时 FPN 网络也被应用于 anchor-free 方法中,包括 FSAF^[16]、FCOS^[17]、FoveaBox^[18] 等。虽然目标检测取得了很大成功,但无法有效的检测出运动物体,在动作识别方面有很大的局限性。

2.3 帧差法

帧差法是通过计算相邻帧图像的灰度差与设定阈值的大小关系来判断图像中是否存在运动目标的检测方法,主要分为两帧差法和三帧差法。两帧差法适合运动目标较缓慢的场景,当运动目标移动速度较快时,三帧差法的检测结果更为准确。帧差法被广泛应用于运动目标的检测,周越等人^[19] 利用帧差法检测运动车辆,也有研究将帧差法应用于滤波跟踪^[20]。帧差法也会与其他技术进行结合,例如背景差分法^[21,22]。欧先锋等人^[23] 将帧差法融入卷积神经网络中,提升了复杂场景以及小目标的检测效果。

2.4 动作识别

为了有效地捕获视频中的外观和运动信息, Simonyan 等人^[24] 基于 CNN 设计了一种双流卷积网络分别处理图片的 RGB 图像信息和光流信息。双流网络也是动作识别领域的重要方法。之后 Wang 等人^[25] 提出了 TSN (temporal segment networks) 解决长视频的行为判断和小数据的问题。除了双流网络,2015 年 Tran 等人^[26] 提出 C3D 卷积网络,将卷积网络的卷积核结构从二维扩展为三维来捕获视频中的时空特征,从此三维卷积网络获得快速发展。Hara 等人^[27] 将三维卷积与 ResNet^[28] 结合,增加了网络的深度,有效地提升了在动作识别领域的准确率。Carreira^[29] 将 2D 网络的与训练权重应用与 3D 网络,提升了模型的效果。虽然视频动作识别已经取得了很好的效果,但是其主要用于短视频的行为判断,无法满足安全管控的实时性的要求。

3 整体思路

为了有效地解决电力运维作业现场的穿越围栏监控问题,本文提出了一种基于 Faster RCNN 的智能检测方法。首先对现场监控视频进行取帧,利用改进的 Faster RCNN 算法检测围栏位置以及使用关键点算法检测人的关键点位置,通过设计的算法判断人员是否

存在违规行为. 穿越围栏是一个连续的行为, 仅凭单张图像的检测结果容易造成误检, 所以我们借助了帧差法的思想, 将多张图像的检测结果进行连续性判断. 最后我们将目标检测、关键点检测算法以及判断算法结合, 提出一种端到端的穿越围栏检测算法.

3.1 整体结构

图1描述了本文设计的基于Faster RCNN的穿越围栏违规行为检测方法的整体流程图. 首先获取视频流, 对视频取帧之后, 将图像分别输入到Faster RCNN网络以及人体关键点检测网络获取到围栏的位置信息以及人体信息暂存至队列当中. 然后利用帧差判断法判断该张图片是否存在违规行为, 如果存在, 就对该帧图像中的任务用红色框表示并将检测结果保存至本地.

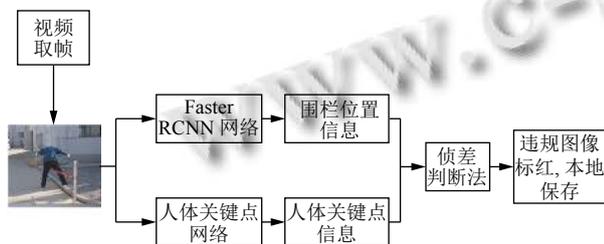


图1 整体流程图

3.2 对于Faster RCNN的改进

Faster RCNN网络结构如图2所示. 数据经过预处理后, 送入到backbone网络中提取特征得到feature map. Faster RCNN提出的RPN (region proposal networks)网络首先生成anchor boxes并通过Softmax分类判断属于positive或者negative. 得到的anchors通过ROI pooling层与feature map结合提取proposal feature, 然后通过全连接和Softmax进行分类, 并对proposals进行回归修正. 本文依据围栏目标的数据特征, 对Faster RCNN进行了如下调整:

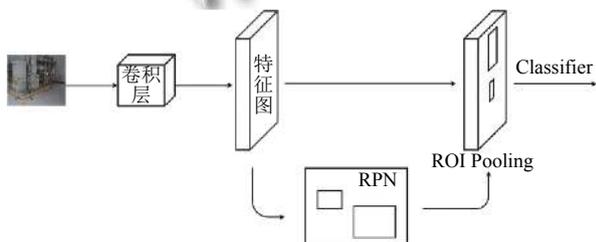


图2 Faster RCNN网络结构

(1) 选用Resnet101+FPN作为Backbone: Faster RCNN采用的是VGG网络的最后一层特征作为feature

map, 为了提升网络的特征提取能力, 本文采用ResNet101网络结合FPN网络作为特征提取网络. ResNet网络是一个深度卷积神经网络, 针对深度神经网络可能出现的随着网络深度加深导致训练集准确率下降的问题, ResNet设计了残差模块提出了深度残差网络, 提升了网络抗过拟合能力. 图3展示了两种残差结构.

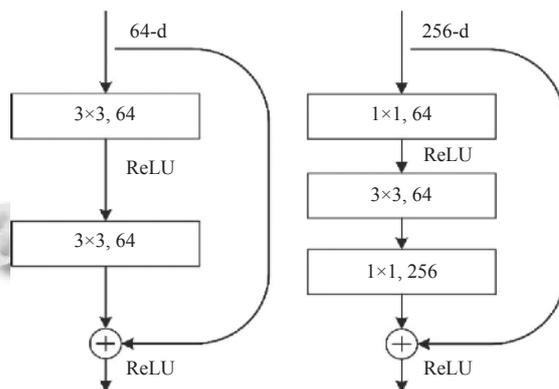


图3 两种残差结构

低层的特征图语义信息少但是目标位置准确, 高层的语义信息丰富但目标位置粗略, 为了解决物体检测中的多尺度问题, FPN设计了一种金字塔式的网络, 可以在不增加原有模型计算量的情况下提升物体检测的性能. ResNet101结合FPN的具体网络结构如图4.

(2) 修改anchors比例. 在Faster RCNN中, RPN会对特征图中的每个特征点提取9个anchors, 其中anchors的长宽比设置为{0.5, 1.0, 2.0}. 本文考虑到围栏数据往往是长条形状的, 所以将长宽比设置为{0.5, 1.0, 2.0, 3.0}, 这样使产生的anchors更容易匹配到原图中围栏的目标区域. 产生的候选anchors通过Softmax网络二分类为positive与negative, 挑选2000个positive anchors进行后续的计算.

(3) 采用Softnms: NMS (non-maximum suppression)会对候选框进行排序, 选中分数最高的候选框, 最后抑制与选中框有明显重叠的候选框. 为了优化NMS这种略显暴力的方法, Softnms应运而生. 和NMS相同, Softnms首先也是对候选框排序, 选择分数最高的框作为选中框, 不同的是, 对于其他有明显重叠的候选框的处理不是暴力的将其置信度置为0, 而是会降低其置信度, 进行下一次的选择.

(4) 使用GIoU Loss: 对于许多基于anchor的目标检测算法, 计算预测框与Ground Truth Bbox的损失

时候往往使用 $L1$ 和 $L2$ 范数,但是评价预测框置信度的时候却使用的 IoU (交并比),显然存在一定的问题.所以本文选择了 $GIoU$ Loss 作为损失函数. $GIoU$ 的做法是:假设现在有两个预测框 A 和 B ,需要找到最小的一个可以将 A 和 B 包围的封闭形状 C ,然后计算 C 中不包含 A 、 B 的部分占 C 的比值. IoU 以及 $GIoU$ 的公式如下所示:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|} \quad (2)$$

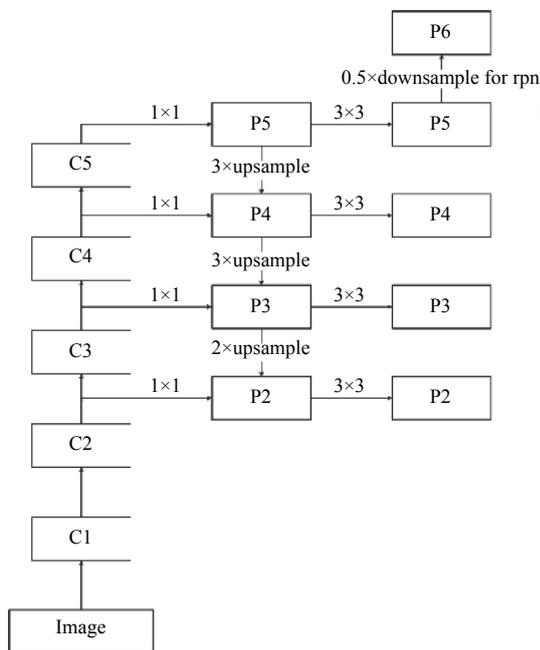


图4 ResNet101+FPN 网络结构

3.3 帧差判断法

帧差法是根据物体运动时目标在相邻帧图像上的差异进行运动目标检测的算法,通常采用灰度图的方式判断运动目标.本文借鉴帧差法的思想,将帧差法与目标检测结合进行穿越围栏违规行为判断.首先利用改进的 Faster RCNN 以及人体关键点检测算法获取图像中围栏坐标以及人体 17 个关键点的坐标并存储在判断队列中,判断队列的大小设置为 5,即每次存储 5 张图像的信息.如果某帧图像被认为可能存在违规行为,就会继续判断该帧图像与接下来几帧图像的运动关系,具体是当人体的关键点位于异常区域,就会继续侦察后几帧图像相关关键点的运动趋势,综合判断该帧图像的违规行为.算法流程图如图 5 所示.

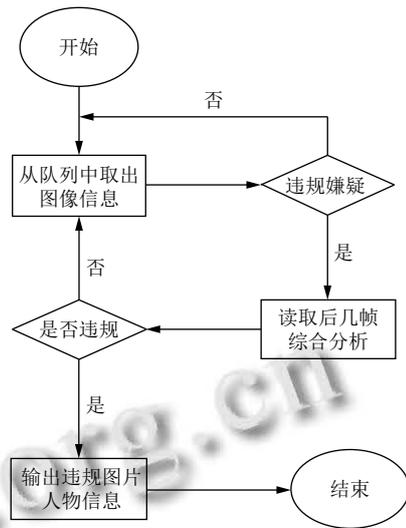


图5 帧差判断法流程图

4 实验分析

本实验的硬件条件为:显卡 NVIDIA GeForce RTX 2080Ti,处理器为 Inter i7 9900k.选择 MMDetection 框架,使用 PyTorch 进行深度学习网络搭建,语言为 Python 3.6.8.

4.1 Faster R-CNN 实验

本次实验共有围栏数据 1601 张,其中分为训练集 1387 张和测试集 214 张.围栏共分为 7 类:菱形围栏,横条围栏,旗子围栏,网状围栏,栅栏,单柱双带围栏,黄黑菱形围栏.使用 UPCLabel 框出目标之后转化为 COCO 格式.

为了尽量的减少背景噪声,对数据首先裁剪至大小为 1300×600,然后为了增强模型的泛化性能,对数据以 0.5 的概率进行随机翻转.本次实验的评价指标为 bbox_mAP 以及 bbox_mAP_50.实验共训练 50 个 epochs,初始学习率设置为 0.001,并在第 20 次以及第 40 次 epoch 将学习率乘 0.1.实验效果如表 1 所示.

表1 Faster RCNN 实验设置及结果 (%)

组数	设置	bbox_mAP_50	bbox_mAP
1	R50+FPN	68.8	43.7
2	R101+FPN	74.9	44.5
3	R101+FPN+Softnms_0.5	73.6	46.6
4	R101+FPN+Softnms_0.6	75.8	48.0
5	R101+FPN+Softnms_0.7	73.0	46.3
6	R101+FPN+Softnms_0.6+GIoU	76.5	48.3
7	R101+FPN+Softnms_0.6+GIoU +Anchor_resize	76.9	48.8

综合以上结果,本文选用了 ResNet101+FPN 作为网络的 backbone, 并对 anchor 重新设计比例, 使用 Softnms 和 GIoU 优化, 最终模型可以达到 76.9 的 mAP₅₀ 以及 48.8 的 mAP. 此外研究还发现, 在本文的数据集上, ResNet101 的效果要好于 ResNet50, 使用 Softnms 以及 GIoU 都可以提高模型的效果. 另外, Softnms 的阈值需要根据数据集的情况来设定, 本实验尝试了阈值选择为 0.5, 0.6, 0.7 分别进行实验, 最终选择了 0.6 作为最终的阈值.

4.2 性能测试

训练得到模型后, 将模型与帧差算法等其他部分结合, 构建基于 Faster RCNN 的穿越围栏检测方法. 为了验证该方法的性能, 本文选取了 5 段现场场景下模拟的违规视频, 经过视频取帧、分类之后, 共获取非法图片 168 张, 合法图片 132 张. 由于要求图片的连续性, 所以我们将图片按照视频取帧顺序依次输入检测方法中, 利用 txt 文件存储图片的标签. 本次实验采用准确率作为评价指标, 其计算公式如式 (3):

$$ACC = \frac{\sum_{i=1}^N (y = y_i)}{N} \quad (3)$$

其中, N 表示图片的总数量, 本实验中为 300, y 表示该图片是否合法, y_i 表示该图片的标签. 经过一轮实验, 共正确检测图片 276 张, 漏检和误检图片共 24 张, 准确率达到 92%.

为了验证本方法的稳定性以及帧差判断方法的有效性, 本文将上述实验进行多次, 然后将其结果与未使用帧差判断方法的结果进行对比, 结果取多次实验结果的平均值, 实验结果如表 2.

表 2 帧差判断法验证实验结果

方法	正确数量	错误数量	误检数量	漏检数量	准确率 (%)
帧差判断法	276.3	23.7	5.3	18.3	92.1
未使用帧差判断法	227	73	40	33	75.7

从表 2 结果中可以看出, 本文提出的帧差判断方法可以有效的提升穿越围栏检测方法的准确性, 准确率提升 16.4%. 误检率和漏检率都大幅度减小, 特别是误检数量, 采用帧差判断方法之后从 40 张减少到 5.3 张. 部分实验效果如图 6 所示.

如果图 6 中检测到存在违规的工作人员, 我们会发出违规警告信息, 然后将违规人员信息保存并通过

OpenCV 使用红色警戒框表示并保存该帧图像至本地. 并且该方法的处理速度可以达到 32 fps, 完全可以达到实际应用中的实时性的要求.



图 6 实验效果图

5 结论与展望

本文将深度学习目标检测与帧差法的思想结合提出了一种基于 Faster RCNN 的穿越围栏违规行为检测方法. 首先基于数据的特点, 改进了 Faster RCNN 网络, 包括选用 ResNet101+FPN 的骨干网络, 修改的 anchor 的比例, 选用 Softnms 和 GIoU. 然后借助围栏目标与人体关键点信息, 结合帧差判断法进行穿越围栏违规行为检测. 实验结果证明, 改进的 Faster RCNN 网络在围栏数据集取得了更好的效果, 提出的帧差判断法也能够更好的检测出视频中的违规行为, 检测速度达到 32 fps, 可以达到实际应用中的实时性要求, 并具有较高的准确性, 可以有效的提升电力场景安全管控的效率.

实验结果也显示方法仍然存在一定的漏检数量, 未来将继续考虑实际复杂的场景, 更加充分利用视频的光流信息, 提升穿越围栏检测的准确性.

参考文献

1 叶帆, 王娜. 电力系统信息安全的重要性及防护措施. 通信

- 电源技术, 2019, 36(12): 248–249.
- 2 金小谷. 中国电力装备研究应对中国工业 4.0 的创新动能. 企业科技与发展, 2015, (23): 1–3.
 - 3 陈俊勇. 浅谈电力工程施工现场安全管控. 科技经济导刊, 2021, 29(9): 229–230.
 - 4 张昭, 刘洋, 吕梁辉, 等. 基于红外对射和视频监控的变电站电子围栏设计. 国网技术学院学报, 2020, 23(6): 31–33, 51. [doi: 10.3969/j.issn.1008-3162.2020.06.010]
 - 5 黄纯, 陈国伟, 毛新健. 洋山港水域电子围栏系统的应用. 水运管理, 2020, 42(9): 3–5. [doi: 10.3969/j.issn.1000-8799.2020.09.002]
 - 6 房凯. 基于深度学习的围栏跨越行为检测方法. 计算机系统应用, 2021, 30(2): 147–153. [doi: 10.15888/j.cnki.csa.007770]
 - 7 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
 - 8 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
 - 9 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2980–2988.
 - 10 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
 - 11 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 936–944.
 - 12 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2999–3007.
 - 13 Law H, Deng J. CornerNet: Detecting objects as paired keypoints. *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018. 765–781.
 - 14 Zhou XY, Zhuo JC, Krähenbühl P. Bottom-up object detection by grouping extreme and center points. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 850–859.
 - 15 Duan KW, Bai S, Xie LX, *et al.* CenterNet: Keypoint triplets for object detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 6568–6577.
 - 16 Zhu CC, He YH, Savvides M. Feature selective anchor-free module for single-shot object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 840–849.
 - 17 Tian Z, Shen CH, Chen H, *et al.* FCOS: Fully convolutional one-stage object detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 9626–9635.
 - 18 Kong T, Sun FC, Liu HP, *et al.* FoveaBox: Beyond anchor-based object detection. *IEEE Transactions on Image Processing*, 2020, 29: 7389–7398. [doi: 10.1109/TIP.2020.3002345]
 - 19 周越, 李硕. 基于帧差法的运动车辆检测算法研究. 电子世界, 2021, (3): 35–36.
 - 20 刘志峰, 陈姚节, 程杰. 结合帧差法的尺度自适应核相关滤波跟踪. 计算机技术与发展, 2021, 31(2): 70–74.
 - 21 何银飞. 基于改进的帧差法和背景差法实现运动目标检测 [硕士学位论文]. 秦皇岛: 燕山大学, 2016.
 - 22 张应辉, 刘养硕. 基于帧差法和背景差法的运动目标检测. 计算机技术与发展, 2017, 27(2): 25–28. [doi: 10.3969/j.issn.1673-629X.2017.02.006]
 - 23 欧先锋, 晏鹏程, 王汉谱, 等. 基于深度帧差卷积神经网络的运动目标检测方法研究. 电子学报, 2020, 48(12): 2384–2393.
 - 24 Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos. arXiv: 1406.2199, 2014.
 - 25 Wang LM, Xiong YJ, Wang Z, *et al.* Temporal segment networks: Towards good practices for deep action recognition. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 20–36.
 - 26 Tran D, Bourdev L, Fergus R, *et al.* Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 4489–4497.
 - 27 Hara K, Kataoka H, Satoh Y. Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet? *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 6546–6555.
 - 28 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770–778.
 - 29 Carreira J, Zisserman A. Quo vadis, action recognition? A new model and the kinetics dataset. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 4724–4733.