

基于深度学习的移动机器人目标跟踪系统^①



张新强, 骆 辉, 周国顺

(大连东软信息学院 智能与电子工程学院, 大连 116023)

通讯作者: 张新强, E-mail: zhangxinqiang@neusoft.edu.cn

摘 要: 针对当前智能移动机器人在跟踪过程中常因目标发生外观形态上的变化而丢失跟踪目标的问题, 利用 Caffe 深度学习框架和 ROS 机器人操作系统作为开发平台, 设计一个高准确度及高实时性的移动机器人目标跟踪系统并进行了研究. 使用对于目标形变、视角、轻微遮挡及光照变化具有鲁棒性的基于孪生卷积神经网络的 GOTURN 目标跟踪算法, 通过 ROS 系统为桥梁使离线训练的跟踪模型实时应用于 TurtleBot 移动机器人上, 并开展了详细的测试. 实验结果表明, 该目标跟踪系统不仅设计方案可行, 实现了移动机器人在各种复杂场景下有效地跟踪目标, 还具有成本低、性能高和易扩展等特点.

关键词: 目标跟踪; Caffe 深度学习框架; 孪生卷积神经网络; GOTURN 目标跟踪算法; ROS

引用格式: 张新强, 骆辉, 周国顺. 基于深度学习的移动机器人目标跟踪系统. 计算机系统应用, 2020, 29(3): 114-120. <http://www.c-s-a.org.cn/1003-3254/7317.html>

Target Tracking System for Mobile Robot Based on Deep Learning

ZHANG Xin-Qiang, LUO Hui, ZHOU Guo-Shun

(School of Intelligence and Electronic Engineering, Dalian Neusoft University of Information, Dalian 116023, China)

Abstract: In view of the current intelligent mobile robot in the tracking process due to the target shape on the changes in a loss of tracking target, using the Caffe deep learning framework and ROS robot operating system as a development platform, a high accuracy and high real-time target tracking system of mobile robots is designed for research. The GOTURN target tracking algorithm based on the twin convolutional neural network, which is robust to target deformation, viewing angle, slight occlusion and illumination changes is used, and the ROS system is used as a bridge to enable the offline training tracking model to be applied to the TurtleBot mobile robot in real time, also a detailed test is carried out. Experimental results show that the target tracking system is not only feasible in design, but also has the characteristics of low cost, high performance and easy expansion.

Key words: target tracking; Caffe deep learning framework; twin convolutional neural network; GOTURN target tracking algorithm; ROS

机器人技术涉及众多领域的研究, 目标跟踪不仅是机器人技术研究领域热点之一, 也是实现移动机器人智能化的基础. 视觉目标跟踪是一个非常具有挑战性的任务, 因为处理好实际中各种复杂多变的场景是很

困难的, 特征选择是影响目标跟踪精度的重要因素^[1]. 卷积神经网络通过对输入数据进行多层神经网络的映射, 从而得到有效的分层特征抽象表达^[2], 解决传统目标跟踪方法人工设计特征的难题, 增强机器人对环境

① 基金项目: 辽宁省自然科学基金 (20180551084)

Foundation item: Natural Science Foundation of Liaoning Province (20180551084)

收稿时间: 2019-07-30; 修改时间: 2019-09-02; 采用时间: 2019-09-18; csa 在线出版时间: 2020-02-28

的理解和建模,实现移动机器人在各种场景中对目标的实时检测与跟踪^[3-5]。

本文利用深度学习框架,训练对目标形变、视角、轻微遮挡及光照变化具有鲁棒性的跟踪模型,并根据实际应用要求进行调整,使其对特定类型的目标跟踪更加准确^[6-8],实现移动机器人在各种场景下有效地检测并跟踪目标,从而使其能高效协助其他机器人、自动化设备及相关人员完成对应复杂场景下的任务。

1 移动机器人及其控制系统

1.1 ROS 机器人操作系统

传统的机器人系统开发通常着重于整体功能的设计,使机器人系统之间每个模块的耦合能力很强,难以相互分离,极大地影响代码的可复用性和可移植性及实际开发中部署的效率。

随着机器人对代码重用和模块化的要求越来越高,使得很好地适应其要求的开源机器人操作系统 ROS 崭露头角,其由多个各自独立的节点组成,并且各个节点之间可以通过发布/订阅消息模型进行通信。

ROS 支持多种编译语言协作实现功能,使用标准的 TCP/IP 协议实现了系统内部各个节点之间的通信,接口可以将第三方组件认为是其操作系统的一部分,这些功能可满足大部分研发人员的需求。如今 ROS 的应用范围已涵盖机械臂抓取、导航机器人、人形机器人、移动机器人底盘、无人机器船、无人飞行器等高新技术领域^[9,10]。

1.2 TurtleBot 移动机器人

TurtleBot 是一个基于 ROS 机器人操作系统的低成本开源移动机器人,可在复杂环境下实现许多功能,充分地满足大多数科研工作的研发需要。

Kinect 视觉深度传感器将特定图案的光投射到物体表面,然后用相机接收物体表面反射的结构光图案的变形,从而测量物体与物体之间的距离。Kinect 能够同时提供彩色信息和深度信息,因此非常适用于移动机器人跟踪目标,且可以避免单一信息不足带来的困扰^[11]。

2 Caffe 深度学习框架

2.1 卷积神经网络

传统的目标跟踪针对不同的样本通过人工设计获

取样本的特征,不仅需要专业的知识字段,而且实际应用中因为其特征缺乏泛化性,效果并不令人满意。因此传统的目标跟踪算法实现的目标跟踪准确度无法满足移动机器人工程化应用的需求。

卷积神经网络是近年发展起来,并在计算机视觉领域广泛应用并已取得丰硕成果的高效识别方法,其能够学习到原始输入图像的不同层次的特征抽象,对各种计算机视觉任务具有优良的学习能力和泛化能力,为实现移动机器人目标跟踪的功能奠定基础。

如图 1 所示,卷积神经网络是一种前馈神经网络结构,其主体通常是由卷积层和池化层组成的模块级联构成,然后通过全连接层或者分类器,对提取到的特征向量进行分类,输出分类结果。其采用局部感知和参数共享的技术手段,降低卷积神经网络中参数的规模,从而获得针对不同检测任务有更强鲁棒性的网络模型^[1,12-14]。

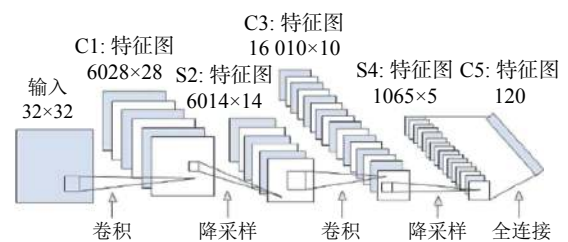


图 1 卷积神经网络结构图

2.2 Caffe 深度学习框架

伴随着近年深度学习技术的快速发展,许多大学和科研机构都推出了深度学习开发框架,推动了深度学习工程化应用的进程,极大地增强深度学习工作的可复用性及实际从事深度学习的研究人员在开发中部署的效率。与 Torch7、Theano 和 Pylearn2 等框架相比,Caffe 具有以下独特的优势:

(1) Caffe 内置 Python 和 Matlab 两种外部接口,方便对实验数据的后续分析。

(2) Caffe 有多个基本数据集的培训配置文档,包括快速培训模型和完整培训模型。

(3) Caffe 采用 GPU 并行技术,增强代码的执行效率,大大地缩短模型的训练时间。

(4) Caffe 基于尽可能多的模块化原则,这使得扩展新的数据格式、网络层和丢失功能变得很容易。

使用能从数据中学习深层本质特征的卷积神经网络,利用 Caffe 深度学习框架对模型进行大规模数据训

练,从而获得大量具有代表性的功能信息,以实现移动机器人目标跟踪的功能^[15].

3 基于卷积神经网络的目标跟踪算法实现

3.1 GOTURN 目标跟踪算法

众多研究人员往往尝试使用神经网络以在线训练的方式进行跟踪,然而,神经网络训练是一个缓慢的过程,导致跟踪速度非常慢,其中性能最好的跟踪器在 GPU 上以 100 FPS 的速度运行^[8].因此,这些使用在线训练方式的跟踪器不适用于需要以实时速度进行目标跟踪的移动机器人上.

GOTURN (Generic Object Tracking Using Regression Networks) 是基于回归网络的通用对象跟踪算法,其能通过大量数据进行离线训练,并采用单次回归到目标对象位置的方法,成为第一个能够以 100 FPS 的速度跟踪目标的深度学习跟踪算法,对未见过的类别样例进行鲁棒性地跟踪,对特定类别样例的跟踪效果更好.

GOTURN 的网络结构是将跟踪目标区域与搜索区域作为输入,同时输入至两个卷积神经网络中,再经过全连接层返回目标的位置,如图 2 所示.因其两个卷积层配置和参数是相同的,故也称为孪生神经网络.

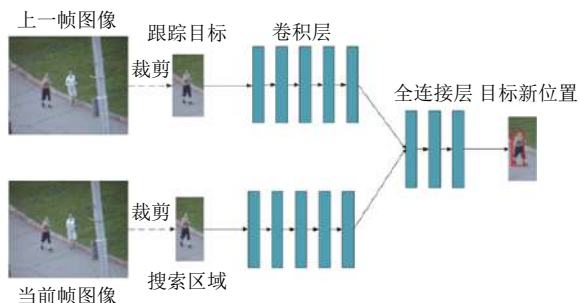


图 2 运动目标误判效果

GOTURN 网络的卷积层是由 Caffe Net 的前五层卷积层及池化层依次排列组成,其输出被连接成完整的向量输入到 3 个有 4096 个节点并完全连接的全连接层中,最后连接到拥有 4 个节点的输出层输出矩形返回目标的运动位置,表示目标左上角和右下角的坐标.

如图 3 所示,假设在坐标系 $t-1$ 中,目标被标记为中心位置 $C(c_x, c_y)$, w 为矩形框的宽度, h 为矩形框的高度.然后在 t 时刻,取以 $C(c_x, c_y)$ 为中心的图像块对坐标系 t 当前帧的要跟踪的目标进行标定.剪裁出来的框大小为 $k_1 \times w, k_1 \times h$, k_1 决定接受多少背景信息.

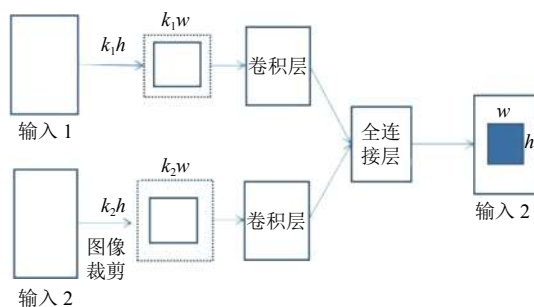


图 3 GOTURN 输入输出网络

为了知道目标在当前帧中的位置,跟踪首先基于上一帧的位置,找到待搜寻目标的区域,并确定目标在当前帧的近似位置.网络的目的是要回归目标在当前搜索区域中的位置,在帧中选择跟踪器的搜索框,确定搜索字段后,留下搜索字段对应的图像块.

具体来说,从帧图像中截取的搜索字段 P 以 $C(c_x, c_y)$ 为中心,即 C 作为目标的平均位置.假设 $C'=C$,和前一帧框出来的区域是一样的,搜索区域的大小为 $k \times w, k_2 \times h$, w 和 h 为上一帧目标标记矩形框的宽度和高度, k_2 为目标搜索半径.事实上,让 $k_1=k_2=2$.只要目标没有被遮挡或移动得太快,通常就会在搜索字段中找到并定位目标.对于快速移动的目标, k_1, k_2 就需要增大.

3.2 GOTURN 网络离线训练实现

本文设计的目标跟踪模型在 ILSVRC Challenge^[16] 及 VOT Challenge^[17] 数据集上进行训练和评估,如图 4 和图 5 所示,它由美国斯坦福大学的李飞飞等所创立,包含近 1400 万的图像帧.视频的每一帧都注释有遮挡、光照变化、运动、大小变化、机位移动等变化形式.本文改进目标跟踪算法的训练集由 ILSVRC2017 训练集、验证集和 ALOV 训练集组成,改进检测模型的测试集在 VOT2014 测试集上进行评估.

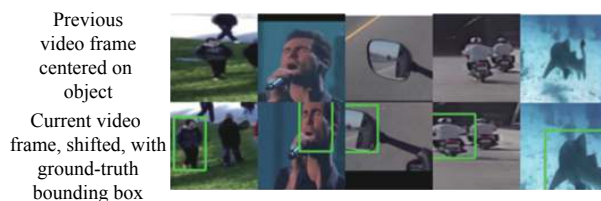


图 4 VOT challenge 数据集

深度卷积神经网络需要海量的有标注数据进行训练,从而使得训练出的模型能从训练图像数据集中获

得较好的特征表达. 当数据集数据不足时, 深度卷积神经网络训练出的模型往往表现不佳. 对于训练样本, 现有的带有真实标签的数据集包含的视频序列较少, 为了使深度学习训练出的模型针对不同的任务有更强的泛化能力, 需要使用更大的数据集进行训练. 目前, 改善这个问题的方法是数据集增强, 即对已有数据进行处理, 创建新数据对训练集进行扩充.



图5 ILSVRC challenge 数据集

创建新数据的方法是将输入数据进行随机平移或旋转等操作扩充数据集, 并且对标定的真实数据保持相同的变换. 假设 (c_x, c_y) 和 (c'_x, c'_y) 分别为变换前后的目标中心点坐标, 通过预测其下一步移动位置来生成样本, 如式(1), 式(2):

$$c'_x = c_x + \omega \cdot \Delta x \tag{1}$$

$$c'_y = c_y + \omega \cdot \Delta y \tag{2}$$

其中, 横纵向的位移量 Δx 和 Δy 服从Laplace分布, 其概率分布如式(3):

$$f(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right) \tag{3}$$

其中, μ 为位置参数, b 为尺度参数, 其函数形式类似于双指数分布, 如图6所示.

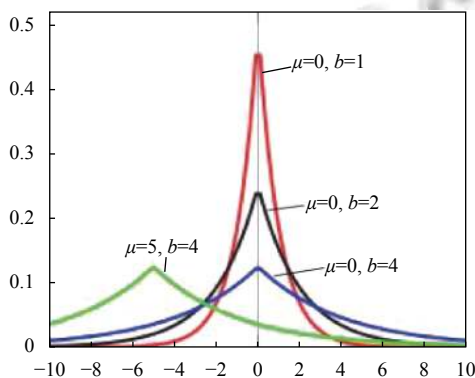


图6 横纵向的位移量 Δx 和 Δy 服从Laplace分布

μ 决定函数中心位置, 而 b 越小函数越陡, 且中间区域面积越大, 因此目标中心位置主要是小范围的变动.

通过交叉验证, 该实验最终选取的分布参数如式(4):

$$\Delta \sim \text{Laplace}\left(0, \frac{1}{5}\right), \gamma \sim \text{Laplace}\left(1, \frac{1}{15}\right) \tag{4}$$

利用以上方法, 对训练集中的图像随机切块, 然后将样本做小批的梯度下降处理, 以此保证跟踪窗口必须至少包含目标的一半, 并通过约束来防止目标窗口过度拉伸和变形. 通过这个方法实现对训练样本的扩充, 大大改善模型的泛化能力.

GOTURN 算法离线训练网络的过程如图7所示, 将耗时的计算过程离线进行, 使跟踪时不需要进行在线更新, 是算法速度足够快的一个重要原因. 由于能够获得到测试数据中任意时刻的信息, 因此可以在测试数据上同时行前向和后向搜索以进行全局的优化.

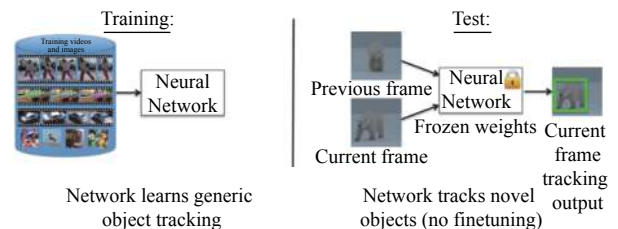


图7 GOTURN 离线训练和目标跟踪测试

为了更直观地认识 GOTURN 算法的跟踪结果, 用视频序列对算法进行了测试. 每个序列的属性如表1所示.

表1 测试视频序列分析

序列名称	帧数	干扰性因数
Boy	891	尺度变化, 运动模糊, 快速运动, 旋转
Caescaler	251	尺度变化, 遮挡, 快速运动, 旋转
Couple	139	尺度变化, 形变, 快速运动, 旋转, 背景杂乱
Dog1t	1349	尺度变化, 旋转
Doll	3871	光照变化, 尺度变化, 遮挡, 旋转
Walking2	499	尺度变化, 遮挡, 低分辨率

GOTURN 算法的视频序列跟踪结果如图8所示. 通过实验数据分析, GOTURN 算法具有良好的跟踪能力, 能在简单的场景中实现目标跟踪, 对尺度变化、变形等具有较强的鲁棒性. 在跟踪目标有类似的目时 GOTURN 算法可能会导致跟踪目标丢失而跟踪其他对象. 分析原因发现 GOTURN 算法使用的数据集是一个分类数据集, 其离线训练的跟踪模型更适合数据集中存在的类别的目标, 对于不存在的类别其识别能力较差. 因此可以利用离线海量数据集, 根据某一场景, 训

练专门跟踪某一类别的跟踪器, 以充分利用 GOTURN 算法的优势.

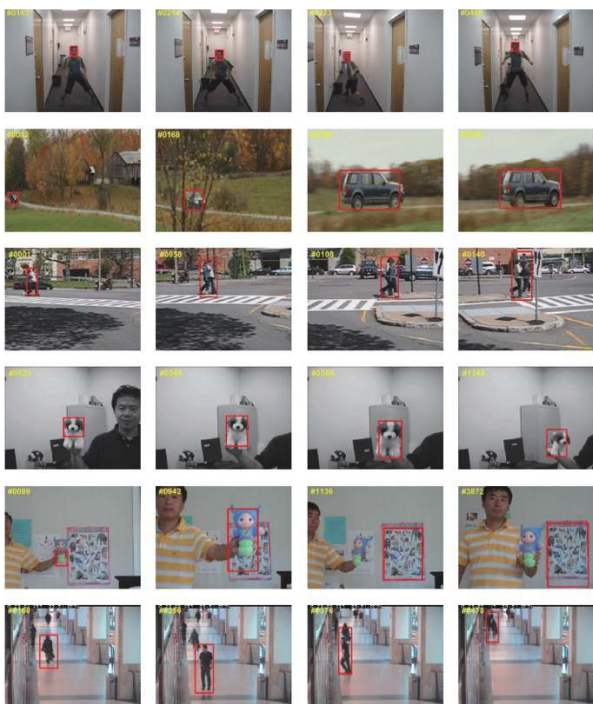


图 8 GOTURN 算法的视频序列跟踪结果

分别更换数据集、损失函数, 对 GOTURN 的跟踪结果进行严谨的实验, 如图 9 所示, 随着训练视频的增加, GOTURN 算法的泛化性能越好. 因此, 如果通过标记更多的视频来增加训练集的大小, 则可以获得进一步的收益.

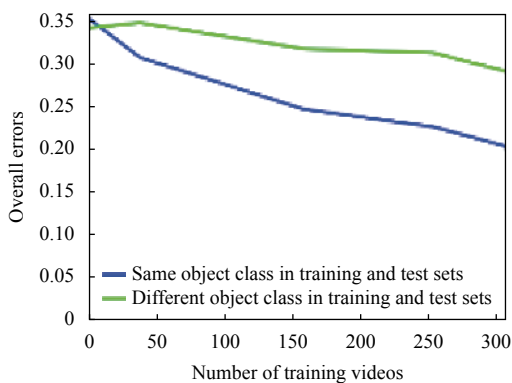


图 9 泛化性能实验

通过对大量标记图像的训练, 跟踪器可以跟踪各种不同类型的物体; 通过对视频和图像的训练, 跟踪器学会在不同的条件下跟踪各种对象类型. 跟踪器对背

景运动、平面外旋转、变形、光照变化和轻微遮挡保持不变, 达到最大的性能.

4 移动机器人目标跟踪系统实现

4.1 目标跟踪程序实现

本文系统采用 Ubuntu14.04 操作系统、ROS 系统 (版本为 Indigo)、CPU 为 Intel(R)Core(TM)i7-4720HQ CPU @ 2.60 GHz、内存为 16 GB, 在 Caffe 深度学习框架的开发环境下编码设计实现的. 使用 Caffe 深度学习框架实现目标检测与跟踪系统, 并将其应用到移动机器人上, 从而高效协助其他相关机器人、自动化设备及人员等完成对应复杂场景下的任务.

系统由各个功能程序的节点组成, 使用 ROS 系统中的主题方式来实现各个程序之间的互相通信. 程序主要部分是获取视觉信息节点、跟踪节点、移动机器人目标跟踪运行程序, 传输图像在 ROS 中同样是以消息格式进行, 节点之间的通信状态如图 10 所示.

对于目标跟随来说, 一个关键的问题是当被跟随目标经过遮挡后再次出现在摄像头画面时, 跟随系统能再次检测到它, 并保持对该目标跟踪. 传统的跟随算法在运行时, 目标容易受到外界环境, 如光照或者目标形变等外界因素影响, 导致跟随失败.

利用具备深度感知能力的 Kinect 传感器获取跟踪目标的视觉信息, 并且以 OpenCV 的 CvBridge 库为桥梁, 将目标的视觉信息转换为 ROS 机器人操作系统可以使用的数据格式, 使得 GOTURN 算法离线训练出的目标跟踪模型, 可应用于 TurtleBot 移动机器人上, 通过 CvBridge 库传递视觉信息的结构如图 11 所示.

目标跟踪程序通过获取目标视觉信息模块、测距模块、目标跟踪模块 3 种相互配合, 获取目标视觉信息模块与测距模块交互进行, 获取目标视觉信息模块通过 CvBridge 将 Kinect 传感器获得的目标视觉信息由 ROS 数据格式转换为 GOTURN 可用的数据格式, 目标跟踪模块使跟踪目标被出现在画面中时, 会被程序检测到并对其进行跟踪, 测距模块保证了该目标被系统锁定并始终保持对该目标的跟随并使跟随效果不易受到外界环境影响.

跟踪目标时, 机器人记录目标物体的坐标, 利用 GOTURN 算法标定的跟踪目标矩形框回归输出的目标左上角和右下角的坐标, 使用控制基于 ROS 的

法是实时性和精度要求都很高的目标跟踪算法,其离线训练和单次回归目标位置的方式使其能够以 100 FPS 以上的速度跟踪目标,使基于深度学习的目标跟踪应用在移动机器人上成为可能,并满足移动机器人目标跟踪的实时性及鲁棒性,这是我们选择使用 GOTURN 目标跟踪算法的重要原因。

GOTURN 算法具有良好的跟踪能力,对尺度变化、形变、光照变化、平面外旋转和轻微遮挡等具有较强的鲁棒性,但在有类似跟踪目标时 GOTURN 算法可能会导致目标丢失而跟踪其他对象,而在线更新机制能可以很好地适应目标自身和周围背景的变化,同时也能区分同一类别的不同目标。因此,将离线训练和在线更新机制有机结合起来能弥补两者直接的劣势,同时又能发展各自的优势。

GOTURN 算法使用的数据集是一个分类数据集,其离线训练的跟踪模型更适合数据集中存在的类别的目标,因此可以利用离线海量数据集,根据某一场景,训练专门跟踪某一类别的跟踪器,以充分利用 GOTURN 算法的优势。GOTURN 目标跟踪算法离线训练和单次回归目标位置的方式使基于深度学习的目标跟踪应用在移动机器人上成为可能,并满足移动机器人目标跟踪的实时性及鲁棒性,从而使其能高效协助其他相关机器人、自动化设备及人员等完成对应复杂场景下的任务,在军事和民用领域都具备十分宽泛的应用价值。

参考文献

- 1 卢湖川,李佩霞,王栋.目标跟踪算法综述.模式识别与人工智能,2018,31(1):61-76.
- 2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097-1105.
- 3 刘甜甜.核相关滤波目标跟踪算法研究[硕士学位论文].武汉:华中科技大学,2017.
- 4 李鹏.基于全卷积孪生网络的目标跟踪算法研究[硕士学位论文].武汉:华中科技大学,2018.
- 5 曾钰廷.基于深度学习的物体检测与跟踪方法的研究[硕士学位论文].南昌:东华理工大学,2018.
- 6 史璐璐,张索非,吴晓富.基于 Tiny Darknet 全卷积孪生网络的目标跟踪.南京邮电大学学报(自然科学版),2018,38(4):89-95.
- 7 何信华,赵龙.基于改进高斯混合模型的实时运动目标检测与跟踪.计算机应用研究,2010,27(12):4768-4771.
- 8 Held D, Thrun S, Savarese S. Learning to track at 100 FPS with deep regression networks. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 749-765.
- 9 王兴伟.基于 ROS 的临场移动机器人控制系统设计与研究[硕士学位论文].合肥:合肥工业大学,2017.
- 10 Quigley M, Gerkey B, Conley K, et al. ROS: An open-source robot operating system. ICRA Workshop on Open Source Software. 2009. 1-6.
- 11 Willow Garage. Turtlebot. <http://www.willowgarage.com/turtlebot>. [2018-04-04/2019-07-30].
- 12 Valmadre J, Bertinetto L, Henriques J, et al. End-to-end representation learning for correlation filter based tracking. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 5000-5008.
- 13 Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2261-2269.
- 14 Wang LJ, Ouyang WL, Wang XG, et al. Visual tracking with fully convolutional networks. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 3119-3127.
- 15 Jia YQ, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding. arXiv: 1408. 5093, 2014.
- 16 Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 2015, 115(3): 211-252. [doi: 10.1007/s11263-015-0816-y]
- 17 Kristan M, Matas J, Leonardis A, et al. A novel performance evaluation methodology for single-target trackers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11): 2137-2155. [doi: 10.1109/TPAMI.2016.2516982]