



基于双孪生网络的自适应选择跟踪系统

张腾飞^{a,b}, 周书仁^{a,b}, 彭建^{a,b}

(长沙理工大学 a. 综合交通运输大数据智能处理湖南省重点实验室; b. 计算机与通信工程学院, 长沙 410114)

摘 要: 孪生网络在解决目标跟踪问题时具有较大的速度和精度优势, 在跟踪领域得到广泛应用。双孪生网络由独立的语义和外观 2 个分支组成, 每个分支都是一个相似学习的孪生网络, 解决了原孪生网络精度不足的问题, 但其每个分支独立训练, 导致系统速度较低。为此, 在双孪生网络的基础上提出一种自适应选择跟踪系统 ASTS。在测试过程中, 简单帧时自动停止网络向前传播, 快速判断目标所在位置, 从而提高系统的跟踪速度。复杂帧时 2 个分支相互协调以准确跟踪目标。在 OTB2013/50/100 和 VOT2017 数据集上的实验结果表明, 相对于固定的双孪生网络目标跟踪方法, ASTS 系统具有更快的速度和更高的跟踪准确率。

关键词: 卷积神经网络; 目标跟踪; 孪生网络; 语义信息; 自适应选择

开放科学(资源服务)标志码(OSID):



中文引用格式: 张腾飞, 周书仁, 彭建. 基于双孪生网络的自适应选择跟踪系统[J]. 计算机工程, 2020, 46(6): 103-107.

英文引用格式: ZHANG Tengfei, ZHOU Shuren, PENG Jian. Adaptive selective tracking system based on twofold siamese network[J]. Computer Engineering, 2020, 46(6): 103-107.

Adaptive Selective Tracking System Based on Twofold Siamese Network

ZHANG Tengfei^{a,b}, ZHOU Shuren^{a,b}, PENG Jian^{a,b}

(a. Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation;

b. Computer and Communication Engineering Institute, Changsha University of Science and Technology, Changsha 410114, China)

[Abstract] Siamese network is widely used in the field of target tracking because of its significant advantage in high speed and accuracy. The twin network is composed of two independent branches: semantic branch and appearance branch. Each branch is a twin network with similar learning, which solves the problem of insufficient accuracy of the original twin network. However, each branch is trained independently, which results in the decrease of system speed. To address this problem, this paper proposes ASTS, an adaptive selective system based on twofold siamese network. In the testing process, the network automatically stops propagating forward at the simple frame and rapidly judge the position of the target, so as to improve the tracking speed of the system. In the case of complex frames, the two branches coordinate with each other to track the target accurately. Experimental results on the OTB2013/50/100 and VOT2017 datasets show that compared with the fixed twofold siamese network object tracking method, the ASTS system has faster speed and higher tracking accuracy.

[Key words] Convolutional Neural Network (CNN); object tracking; siamese network; semantic information; adaptive selection

DOI: 10.19678/j.issn.1000-3428.0054400

0 概述

目标跟踪是计算机视觉和模式识别领域的研究热点之一, 得到了广泛关注与应用。在智能交通系统中, 相机与无人机的自动跟踪拍摄、人机智能交互系统都需要应用目标跟踪方法。虽然近年来目标跟踪方法取得了快速的发展, 但是物体被遮挡、目标发生严重形变、目标运动速度过快、光照尺度变化和背景

干扰等因素导致的目标跟踪系统鲁棒性低和实时性差等问题依然存在^[1]。

现有目标跟踪方法可以分为生成模型方法和判别模型方法两类^[2]。生成模型方法在当前帧对目标区域进行建模, 运用生成模型描述目标区域的表现特征, 在后续帧中进行目标预测, 从而寻找到与目标最为相似的区域。该类方法的典型代表有卡尔曼滤波^[3]、粒子滤波^[4]和 Mean-Shift 算法^[5]等。判别模

基金项目: 国家自然科学基金青年基金项目“基于深度神经网络的实体关系抽取关键技术研究”(61602059)。

作者简介: 张腾飞(1993—), 男, 硕士, 主研方向为机器学习、模式识别; 周书仁, 副教授、博士; 彭建, 副教授、硕士。

收稿日期: 2019-03-27 修回日期: 2019-04-28 E-mail: z738484136@163.com

型方法通过训练分类器来区分背景和目标,这种方法也被称作检测跟踪模型。判别模型由于旨在区分一帧中的目标和背景,因此,其具有更强的鲁棒性,得到了广泛应用。经典的判别模型方法有 CT^[6] 和 TLD^[7] 等算法。文献[8]通过多次连续蒙特卡罗采样得到最优目标区域,利用子块遮挡比例自适应调节学习速率,从而解决了时空上下文跟踪易漂移和遮挡敏感的问题。目前,多数基于深度学习的方法均在判别式框架的范畴内。文献[9]提出了全卷积的孪生网络 SiamFC。SiamFC 的优点在于将跟踪任务转化为检测匹配的过程,通过比较目标帧和模板帧图片的相似度,计算出相似度最大的位置,从而得到目标在模板帧中的位置。CFNet^[10] 通过为低级别的 CNN 引入相关滤波,将相关滤波看作 CNN 网络中的一层,以提高跟踪速度并保证跟踪精度。文献[11]提出的 SINT 结合光流信息,取得了更好的跟踪性能,然而,其引入光流信息导致了跟踪速度缓慢,不能达到实时的要求。文献[12]提出的 SA-Siam 双孪生网络,在 SiamFC 的基础上加入了语义分支,其能够提高跟踪精度但降低了跟踪的速度。

为进一步提高跟踪速度,本文提出一种基于双孪生网络的自适应选择跟踪方法 ASTS。系统自动判断目标帧信息,在简单帧中只运用外观信息进行判断,复杂帧权重确定则结合语义信息和外观信息。在 OTB2013/50/100^[13] 和 VOT2017 数据集上进行实验,以验证该方法的跟踪性能与鲁棒性。

1 孪生网络

全卷积孪生网络的提出在跟踪领域具有重大意义。孪生网络在训练集 ImageNet2015 上进行离线训练,得到相似度匹配函数,在跟踪过程中,通过模板相似度比较得到相似度最大的位置。具体地,以第 1 帧为模板图像,用以在后续 255×255 的搜索图像中匹配定位 127×127 的模板图像 z 。通过离线训练出的相似度函数将模板图像 z 与搜索图像 x 中相同大小的候选区域进行比较。经过卷积得到最后的得分图,其中,目标区域会得到高分,非目标区域会得到低分。相似度函数为:

$$F_l(z, x) = \varphi_l(z) * \varphi_l(x) + v \quad (1)$$

其中, φ_l 是第 l 层的卷积特征, $v \in \mathbb{R}$ 是偏移量。式(1)中的 $F_l(\cdot, \cdot)$ 表示一个 17×17 的置信得分图^[9]。在跟踪过程中,孪生网络简单地评估模板图像与当前帧搜索区域之间的相似性,得到得分图,只要找到得分图中得分最高的区域,然后乘以相应步长即可得到当前帧与模板帧的偏移量,最终获得目标的位置。

2 自适应选择跟踪网络

ASTS 方法的总系统框图如图 1 所示。ASTS 由外观信息与语义信息 2 个分支组成。系统网络的输

入是视频第 1 帧经人工标记的目标真实位置和当前帧裁剪出的目标搜索区域。其中, z 和 z_g 分别表示目标和目标周围环境, x 表示搜索区域。 x 和 z_g 尺寸相同,都为 $W_g \times H_g$, z 的尺寸为 $W_t \times H_t \times 3$, 其中, $W_t < W_g$, $H_t < H_g$ 。每个分支都输出一个相应得分图,得到 z 和 x 的相关性。为了不相互干扰,2 个分支单独训练,直到跟踪时才叠加在一起。

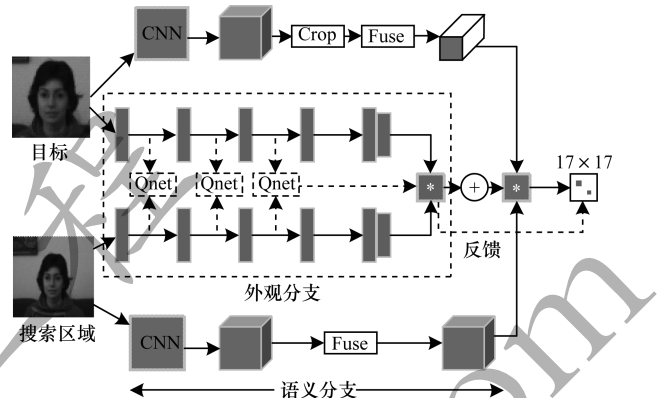


图 1 基于双孪生网络的自适应选择跟踪系统

Fig. 1 Adaptive selective tracking system based on twofold siamese network

2.1 系统外观分支

系统外观分支的输入为目标区域 z 和搜索区域 x 。系统外观分支并非一个简单的孪生网络,而是加入了深度 Q 学习网络^[14]。和 EAST 不同的是,外观分支 P 中最后 2 层卷积层 convn4 和 convn5 没有 Q 网络则不会提前停止,原因是 convn4 和 convn5 层属于深层的网络信息,语义分支会较好地处理,因此,网络不会在最后 2 层提前停止。

在外观分支 P 中执行提前停止的过程被认为是一个马尔可夫决策过程(Markov Decision Process, MDP)。本文通过深度强化学习训练一个有效的决策网络(Agent)^[15]。通过训练决策网络能够学习动作(Action)和判断状态(State),得到提前停止标准从而提前停止网络。决策网络可以跨过特征层进行一系列的操作,比如判断将何时执行停止或者进入下一层,以及如何有效地对边界框进行变形。

在强化学习过程中,马尔可夫决策过程分为一组动作 A 、一组状态 S 和奖励函数 R 。在第 $n(n < 4)$ 层,决策网络检查当前状态 S_n ,然后决定动作 A_n 是停止并输出还是对边界框进行移动变形以进入下一层,同时获得正面或负面的反馈奖励并反映当前框对目标的覆盖程度,以及动作停止前所执行的步骤。

1) 动作:动作集 A 通过验证设置为 6 个不同的缩放动作和一个停止动作,如图 2 所示。缩放动作包括整体缩小和整体放大 2 个全局动作变换以及 4 个改变宽高的局部动作变换。每个边界框由坐标 $b = [x_1, x_2, y_1, y_2]$ 表示,每次转换动作都会通过式(2)对边界框进行离散变换。



图2 马尔可夫决策中的动作说明

Fig. 2 Movements description in Markov decision process

$$\begin{aligned}\alpha_w &= \alpha * (x_2 - x_1) \\ \alpha_h &= \alpha * (y_2 - y_1)\end{aligned}\quad (2)$$

通过对 x 坐标(y 坐标)加上或者减去 α_w (α_h) 来进行变换,与文献[15]相同,本文取 $\alpha = 0.2$ 。

2) 状态:状态是当前层的得分图和历史层得分图的平均值 F_n 和采取动作的历史向量 h_n 组成的二元组,这种结构将会使系统更加鲁棒。历史向量跟踪 h_n 包含了3次历史动作,每个动作又是7维的矢量,则 $h \in R^{21}$ 。

3) 奖励:奖励函数 R 在采取特定动作后,该机制定位物体的提升为正反馈。所设定的提升标准通过计算预测的目标矩形框与手动标记的目标矩形框的交叉联合 (Intersection-over-Union, IoU) 来衡量。IoU 定义为:

$$\text{IoU} = \frac{\text{area}(b \cap R_g)}{\text{area}(b \cup R_g)} \quad (3)$$

其中, b 为预测的目标框面积, R_g 为目标实际所在的位置。奖励函数通过一个状态到另一个状态的 IoU 差别来估计,即当决策网络执行动作 A 、状态从 S_n 转到 S_{n+1} 时,每个状态 S 都有一个相关的矩形框 b ,则奖励函数为:

$$R(S_n, S_{n+1}) = \text{sign}(\text{IoU}(b_{n+1}, R_g) - \text{IoU}(b_n, R_g)) \quad (4)$$

从式(4)可以看出,若 IoU 变大,则奖励为正 (+1);反之,奖励就为负 (-1)。式(4)适用于所有转换矩形框的动作,通过这种方式奖励正向的变化,直到没有更好的动作来使定位更精确或者到达卷积层第3层。停止动作拥有异于其他动作的奖励函数。根据文献[14]可得:

$$R(S_n, S_{n+1}) = \begin{cases} 3, \text{IoU}(b_{n+1}, R_g) \geq 0.6 \\ -3, \text{IoU}(b_{n+1}, R_g) < 0.6 \end{cases} \quad (5)$$

最后,本文应用文献[14]的深度 Q 强化学习网络来学习行动值函数。

2.2 系统语义分支

系统语义分支的输入为目标周围环境 z_g 和搜索区域 x ,本文直接使用在图像分类任务中已经训练好的 AlexNet^[16] 作为语义分支,在训练和测试期间确定所有参数。网络中用 conv4 和 conv5 最后2个卷积层的特征作为输出,并在特征提取后插入一个 1×1 的卷积层进行特征融合,这样做的目的是使语义分支网络能够更好地进行相关操作,并且提高跟踪精度。外观分支 G 的输出表示为:

$$F_g(z_g, x) = \text{corr}(f(\varphi_g(z_g)), f(\varphi_g(x))) \quad (6)$$

其中, $\text{corr}(\cdot, \cdot)$ 表示相关操作, $f(\cdot)$ 表示特征融合, $\varphi(\cdot)$ 表示级联的多层特征。

2.3 双孪生自适应网络

训练期间2个网络完全单独分开训练,互不干扰,跟踪时才对2个网络进行选择性地叠加。跟踪期间,在一串连续的跟踪序列中,帧与帧之间存在大量的相似帧,相比目标帧,这些帧图片的目标形变较小、周围环境语义信息变换不明显。这些帧只利用外观分支较浅层的特征信息跟踪器就能很好地对目标进行跟踪,这时如果完全考虑2个分支,则会使跟踪速度减慢,因此,针对变换不明显语义信息的简单帧,语义分支完全可以忽略。同时在较浅层的网络中,空间的分辨率较高,但特征的语义信息较少,随着网络的加深,从深层网络中提取到的特征语义信息会比较丰富,但是会导致空间的分辨率降低,不利于目标定位与跟踪。因此,在外观分支上浅层的信息能够更好地跟踪目标,定位出目标所在位置。

在外观分支中,让网络通过训练好的深度强化学习 Q 网络来选择合适的停止层,既能够增加跟踪器的跟踪速度,又能很好地利用浅层网络空间分辨率高的特性定位出目标,提高跟踪性能。在变化较大的复杂帧中,外观分支不会提前停止,能够提取到目标更丰富的特征信息,得到的特征与语义分支提取到的特征进行叠加能够更准确地定位出目标的位置,使跟踪器在速度与性能之间得到平衡。当外观网络提前停止时,则外观分支对整体网络作反馈,语义分支的占比为0,完全由外观分支输出;当外观网络没有提前停止时,将上述2个网络得到的相关系数得分图按一定比例进行叠加,即:

$$F(z_g, x) = \begin{cases} F_p(z, x), \tau = 1 \\ \lambda F_p(z, x) + (1 - \lambda) F_g(z_g, x), \tau \neq 1 \end{cases} \quad (7)$$

其中, τ 代表外观分支对整体网络的反馈, λ 是平衡2个分支重要性的加权参数,其可以通过实验来取值, $F(z_g, x)$ 表示被跟踪的目标位置。

3 实验结果与分析

本文在 MatConvNet 库^[17] 上进行仿真,实验环境为 Ubuntu 4.8.2 系统, Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.3 GHz 四核处理器,配备有 NVIDIA GeForce GTX TITAN X GPU,在 OTB50、OTB100、OTB2013 和 VOT2017 基准上分别进行实验。

采用2015年版 Imagenet 大规模视频识别挑战 (ILSVRC)^[18] 的视频数据集进行训练,该数据集包含约4500个视频,接近一百万个注释帧。具体地,在训练过程中,随机地从数据集同一个视频中选取两帧,对其中一帧裁剪出以 z 为中心的 z_g ,从另一帧中裁剪出以人工标注目标为中心的 x 。目标图像 z 大小为 $127 \times 127 \times 3$,对大小为 $255 \times 255 \times 3$ 像素的搜索区域图像 x 进行搜索,并且外观分支网络的 z_g 与 x 具有相同的大小,最终的输出都为 17×17 维。

学习率设定为 10^{-4} 。经过实验得出,当外观网络没有提前停止,即返回值 τ 为 1 时,当 λ 为 0.36 时系统性能最佳。

3.1 OTB 基准实验

OTB 包含 OTB50、OTB100、OTB2013 3 个数据集^[13]。OTB 数据集中的序列分为遮挡、比例变化、快速运动和平面内旋转等 11 个不同的注释属性,OTB 一般有 2 个评估标准,分别是成功率和精确度。对于每一帧,计算跟踪矩形框与人工标注的目标框边界的 IoU 以及它们中心位置的距离,采用跟踪成功率与精确度来评估跟踪器。

本文在 OTB50、OTB100、OTB2013 3 个基准数据集上对 SiamFC^[9]、CFNet^[10]、SINT^[19]、Staple^[20]、EAST^[21] 及本文系统 6 个跟踪器进行评估,结果如

表 1 OTB 基准下的评估结果

Table 1 Assessment results under OTB standards

跟踪器	跟踪速度/FPS	OTB2013		OTB50		OTB100	
		AUC	精确度	AUC	精确度	AUC	精确度
SiamFC	86.0	0.607	0.807	0.516	0.692	0.582	0.771
CFNet	78.4	0.618	0.805	0.503	0.702	0.568	0.748
SINT	4.0	0.655	—	0.633	—	0.572	—
Staple	80.0	0.602	0.793	0.507	0.684	0.578	0.784
EAST	148.0	0.638	—	0.638	—	0.629	—
ASTS	97.0	0.657	0.847	0.639	0.801	0.644	0.835

3.2 VOT 基准实验

VOT 测试基准拥有多个不同的版本,最新的版本有 VOT2015^[22]、VOT2016^[23] 和 VOT2107^[24]。VOT2015 和 VOT2016 拥有相同的序列,但是 VOT2016 中的人工标注标签比 VOT2015 更加准确。由于 VOT2016 中的部分标签已经能够被多数跟踪器准确跟踪,因此 VOT2017 将 VOT2016 中的 10 个序列替换为新的序列,但依然保持总体序列属性分布不变。本文应用 VOT2017 作为评测基准。VOT 基准主要的评测指标为平均重叠期望(Expected Average Overlap, EAO)、准确率(Accuracy, A)、鲁棒性(Robustness, R)。一个性能良好的跟踪器应该有较高的准确率和平均重叠期望分数,但鲁棒性较低。

在 VOT2017 基准下对 ECOhc^[25]、Staple^[20]、SiamFC^[9]、SA-Siam^[12] 和 ASTS 进行比较,结果如表 2 所示,其中量化展示了 5 个跟踪器的平均重叠期望、准确率、鲁棒性和跟踪速度。从表 2 可以看出,ASTS 的平均重叠期望为 0.227,略低于 ECOhc,但 ASTS 具有速度优势,准确率

表 1 所示,最好的结果用加粗表示。从表 1 可以看出,在 OTB2013 基准下,ASTS 具有最佳的性能,其 AUC(Area-Under-Curve)达到了 0.657,超出孪生网络 SiamFC 跟踪器 0.050。虽然 SINT 的 AUC 也达到了 0.655,但是 SINT 并非一个实时的跟踪器,其跟踪速度只有 4.0 FPS。在 OTB50 基准下,EAST 跟踪器虽然达到了高速的 148 FPS,ASTS 的 AUC 也只比其高出 0.001,但在 OTB2013 和 OTB100 中,ASTS 跟踪器的 AUC 分别高出 EAST 约 0.019 和 0.013。OTB100 是 OTB50 的扩充,因此,其更具有挑战性。本文 ASTS 跟踪器在 OTB100 基准中 AUC 依然保持在 0.644,比 OTB50 基准中更高。而在 OTB2013 中表现良好的 SINT 跟踪器,在更多的测试中其 AUC 不够稳定。

达到 0.527,高于 ECOhc 跟踪器。在准确率方面,ASTS 跟踪器表现最优异,高于 SA-Siam 约 0.02。在跟踪速度方面,ASTS 最高达到了 97.0 FPS。在鲁棒性方面,ASTS 表现不如 ECOhc,同样是因为 ECOhc 在速度方面做出了巨大牺牲,但本文方法的鲁棒性均优于其他跟踪器。

表 2 VOT2017 基准下的评估结果

Table 2 Assessment results under VOT2017 standards

跟踪器	EAO	A	R	跟踪速度/FPS
ECOhc	0.238	0.494	0.435	60.0
Staple	0.169	0.525	0.688	80.0
SiamFC	0.188	0.502	0.585	85.6
SA-Siam	0.236	0.501	0.597	50.0
ASTS	0.227	0.527	0.563	97.0

图 3 所示为均值漂移算法^[5]、SiamFC、CT、Staple 和 ASTS 的跟踪实验结果,可以看出,除本文 ASTS 方法外,其他方法都发生了不同程度的漂移现象。

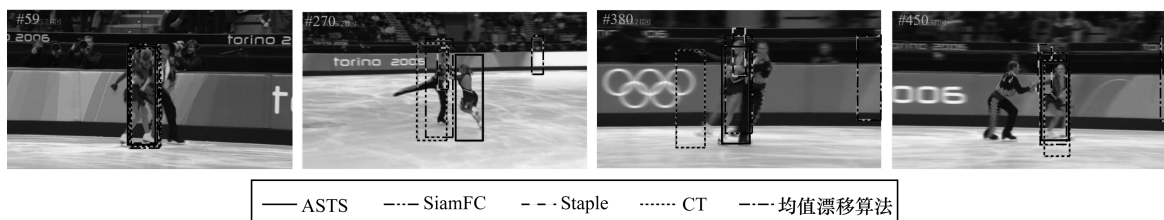


图 3 5 种跟踪器的跟踪结果比较

Fig. 3 Comparison of tracking results of five trackers

4 结束语

本文提出一种基于双孪生网络的自适应选择跟踪方法 ASTS。2个孪生网络分别负责语义信息和外观信息,在外观分支上加入自动停止操作,当在简单帧时自动停止网络向前传播,此时不再与语义信息相结合从而提高跟踪速度,在复杂帧时,孪生网络的速度优势使得 ASTS 方法同样取得了较高的跟踪速度。实验结果验证了 ASTS 方法的高效性与高准确率。下一步将探究更好的注意力机制,并将深度特征与 HOG 特征进行融合,以提高本文方法的跟踪性能。

参考文献

- [1] FAN Heng, XIANG Jinhai. Robust visual tracking with multitask joint dictionary learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(5): 1018-1030.
- [2] ZHANG T, GHANEM B, LIU S, et al. Robust visual tracking via structured multi-task sparse learning[J]. International Journal of Computer Vision, 2013, 101(2): 367-383.
- [3] WELCH G, BISHOP G. An introduction to the Kalman filter[EB/OL]. [2019-03-10]. http://www.cs.unc.edu/~tracker/media/pdf/SIGGRAPH2001_Slides_08.pdf.
- [4] WANG Fasheng, LU Mingyu, ZHAO Qingjie, et al. Particle filter algorithm[J]. Chinese Journal of Computers, 2014, 37(8): 1679-1694. (in Chinese)
王法胜, 鲁明羽, 赵清杰, 等. 粒子滤波算法[J]. 计算机学报, 2014, 37(8): 1679-1694.
- [5] LI Xiangru, WU Fuchao, HU Zhanyi. Convergence of a mean shift algorithm[J]. Journal of Software, 2005, 16(3): 365-374. (in Chinese)
李乡儒, 吴福朝, 胡占义. 均值漂移算法的收敛性[J]. 软件学报, 2005, 16(3): 365-374.
- [6] ZHANG K, ZHANG L, YANG M H. Real-time compressive tracking[C]//Proceedings of European Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2012: 52-63.
- [7] WANG N, YEUNG D Y. Learning a deep compact image representation for visual tracking[C]//Proceedings of International Conference on Neural Information Processing Systems. Washington D. C., USA: IEEE Press, 2013: 809-817.
- [8] LI Long, LIU Kai, LI Ling. Spatial-temporal context tracking algorithm based on target detection[J]. Computer Engineering, 2018, 44(9): 263-268, 273. (in Chinese)
李珑, 刘凯, 李玲. 基于目标检测的时空上下文跟踪算法[J]. 计算机工程, 2018, 44(9): 263-268, 273.
- [9] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 52-68.
- [10] XU H, GAO Y, YU F, et al. End-to-end learning of driving models from large-scale video datasets[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 3530-3538.
- [11] TAO R, GAVVES E, SMEULDERS A W M. Siamese instance search for tracking[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 1420-1429.
- [12] HE Anfeng, LUO Chong, TIAN Xinmei, et al. A two-fold siamese network for real-time object tracking[EB/OL]. [2019-03-10]. <https://arxiv.org/pdf/1802.08817.pdf>.
- [13] WU Y, LIM J, YANG M H. Online object tracking: a benchmark[C]//Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013: 23-28.
- [14] VOLODYMYR M, KORAY K, DAVID S, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [15] CAICEDO J C, LAZEBNIK S. Active object localization with deep reinforcement learning[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 230-256.
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of International Conference on Neural Information Processing Systems. Washington D. C., USA: IEEE Press, 2012: 1097-1105.
- [17] VEDALDI A, LENC K. MatConvNet: convolutional neural networks for Matlab[C]//Proceedings of the 23rd ACM International Conference on Multimedia. New York, USA: ACM Press, 2015: 689-692.
- [18] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2014, 115(3): 211-252.
- [19] TAO R, GAVVES E, SMEULDERS A W M. Siamese instance search for tracking[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 230-238.
- [20] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 52-68.
- [21] HUANG C, LUCEY S, RAMANAN D. Learning policies for adaptive tracking with deep feature cascades[C]//Proceedings of International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 36-48.
- [22] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2017 challenge results[C]//Proceedings of 2017 IEEE International Conference on Computer Vision Workshop. Washington D. C., USA: IEEE Press, 2017: 368-392.
- [23] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2016 challenge results[C]//Proceedings of 2016 IEEE International Conference on Computer Vision Workshop. Washington D. C., USA: IEEE Press, 2016: 96-113.
- [24] KRISTAN M, MATAS J, ALEŠ L, et al. The visual object tracking VOT2015 challenge results[C]//Proceedings of 2015 IEEE International Conference on Computer Vision Workshop. Washington D. C., USA: IEEE Press, 2015: 1130-1138.
- [25] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 59-68.