

基于注意力机制的在线自适应孪生网络跟踪算法

董吉富¹, 刘畅^{1*}, 曹方伟¹, 凌源², 高翔¹

¹大连海事大学信息科学技术学院, 辽宁 大连 116026;

²华录智达科技有限公司, 辽宁 大连 116023

摘要 针对全卷积孪生(SiamFC)网络算法在相似目标共存和目标外观发生显著变化时跟踪失败的问题,提出一种基于注意力机制的在线自适应孪生网络跟踪算法(AAM-Siam)来增强网络模型的判别能力,实现在线学习目标外观变化并抑制背景。首先,分别在模板分支和搜索分支中加入前一帧跟踪所得到的结果,弥补网络在应对目标外观变化的不足;然后通过孪生网络中加入空间注意力模块和通道注意力模块实现不同帧之间的特征融合,从而实现在线学习目标形变并抑制背景,进一步提升模型的特征表达能力;最后,在 OTB 和 VOT2016 跟踪基准库上进行实验。实验结果表明,本文算法在 OTB50 数据集上的精确度和平均成功率比基础算法 SiamFC 分别高出了 4.3 个百分点和 3.6 个百分点。

关键词 机器视觉; 孪生网络; 注意力机制; 卷积神经网络; 视觉跟踪

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.021510

Online Adaptive Siamese Network Tracking Algorithm Based on Attention Mechanism

Dong Jifu¹, Liu Chang^{1*}, Cao Fangwei¹, Ling Yuan², Gao Xiang¹

¹College of Information Sciences and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China;

²Hualu Zhida Technology Co., Ltd., Dalian, Liaoning 116023, China

Abstract This study aims at resolving the tracking failure caused by the coexistence of similar targets and the significant change in the appearance of a target based on the full convolution siamese (SiamFC) network algorithm. An online adaptive siamese network tracking algorithm (AAM-Siam) based on attention mechanism is proposed to enhance the discriminative ability of the network model and achieve the online learning target appearance change and suppress background. Firstly, the results obtained by tracking the previous frame are added into the template branch and the search branch respectively to compensate for the shortcomings of the network by responding to the changes in the appearance of the target. Secondly, the spatial attention module and the channel attention module are employed into the siamese network to achieve the feature fusion among various frames, learn the target deformation online and suppress background, as well as enhancing the model's ability to express features. Finally, detailed experiments are conducted on the online tracking benchmark (OTB) and visual object tracking 2016 (VOT2016) benchmark. The experimental results indicate that the accuracy and average success rate of the proposed algorithm on the OTB50 dataset are 4.3 and 3.6 percentage points higher than those obtained using the basic SiamFC network algorithm, respectively.

Key words machine vision; siamese network; attention mechanism; convolutional neural network; visual tracking

OCIS codes 150.0155; 150.1135; 330.1400

1 引言

视觉目标跟踪在许多领域中都有重要的应用,

比如人机交互、智能交通、视频监控等。但是,对于运动目标而言,在运动过程中自身会不断变化,其运动场景也常伴随着光照变化、剧烈形变、相似背景等

收稿日期: 2019-05-20; 修回日期: 2019-06-22; 录用日期: 2019-07-15

基金项目: 国家科技支撑计划子课题(2015BAG20B02)、辽宁省博士启动基金(201601065)

* E-mail: liuchang@dlnu.edu.cn

干扰,加上跟踪目标事先未知,不同跟踪序列跟踪目标完全不同。到目前为止目标跟踪仍然是一个具有挑战性的任务。

随着卷积神经网络(CNN)在图像处理和目标检测中应用广泛^[1-3],越来越多的学者也将其应用在目标跟踪任务中^[4-9],其优点在于深度特征能够编码高级信息,而且能不受复杂外观变化和混乱的影响,已成为近几年研究的热点方向。其中取得开创性工作的是 Bertinetto 等^[10]提出的全卷积孪生网络(SiamFC)单目标跟踪算法。SiamFC将跟踪视为相似性学习,通过学习匹配函数适应目标变化,候选样本和目标模板通过计算网络所提取特征的互相关得到最终的相似性响应图,预测目标位置。GOTURN^[11]基于上一帧的跟踪结果找到待搜寻目标的区域,利用大量离线数据训练网络,能获得较快的跟踪速度。SANet^[12]提出结构感知网络架构,在学习过程中通过使用循环神经网络(RNN)对对象的自身结构进行编码,进一步提升跟踪精度,但速度只有1 frame/s,很难应用于实际场景。SA-Siam^[13]在SiamFC的基础上增加了语义分支,将语义特征和外观特征互补结合,提升了跟踪性能。StructSiam^[14]提出一种局部结构学习方法用于解决非刚性外观变化和部分遮挡,该方法同时考虑了目标的局部模式及其结构关系,能实现更准确的目标跟踪。

虽然SiamFC在单目标跟踪领域取得了很好的结果,但是和现有的基于深度特征的跟踪器相比仍具有很大的精度差距。由于SiamFC的目标模板始终使用第一帧特征,在应对背景或目标环境的变化时,来自第一帧的特征不足以稳健地表示目标的特征,因此SiamFC跟踪缺乏重要的在线适应性。

针对上述问题,本文在SiamFC基础上分别对孪生网络模板分支和搜索分支加入前一帧的跟踪结果,并引入一种新的注意力机制,以实现更好的特征融合,进一步增强网络模型的判别能力,实现在线学习目标外观变化并抑制背景。最后与几个具有代表性的跟踪算法进行比较,取得了不错的跟踪效果。近年来,基于孪生网络的目标跟踪算法更多集中于优化模板帧稳健性的问题^[15],第一帧的目标信息已知,孪生网络使用其作为模板帧,能够得到最稳健的表达,但是随着时间的推移,目标在形变变化较大或存在相似物体时,后续帧的相关信息发生较大变化,导致模板分支不够稳健。文献^[15]在序列特定信息的更新中结合了孪生网络和加权CNN,而本文算法

(AAM-Siam)首先是基于SiamFC网络结构,其本质是基于模板匹配的跟踪;其次,本文算法拥有更加轻量级的网络结构,降低了模型运算的复杂度;然后,文献^[15]利用加权CNN使用序列特定信息自适应分配权重,以应对目标外观变化,而本文算法则是在模板分支和搜索分支中结合了序列的历史信息,从而实现模板匹配的自适应更新,以应对目标形状的变化;最后,为了解决相似背景存在时导致的模型漂移,本文算法引入注意力机制构建信息特征,以提升网络的判别力。

2 基于注意力机制的在线自适应孪生网络跟踪

2.1 基于相似学习的全卷积孪生网络跟踪算法

基于SiamFC的跟踪算法采用两个参数共享的CNN组成孪生结构。其中一支为模板分支,以初始帧作为模板图像输入,得到模板特征图;另外一支为搜索分支,以当前帧作为搜索图像输入,得到搜索特征图。SiamFC将跟踪视为模板匹配的过程,模板图像 z 和搜索图像 x 通过两个分支所提取的特征进行卷积运算,得到最终的相似性响应图,整个过程可以表示为

$$f(z_t, x_t) = \varphi_t(z_t) * \varphi_t(x_t) + b \cdot \mathbf{I}, \quad (1)$$

式中: b 为偏置项; \mathbf{I} 为单位矩阵; $\varphi_t(z_t)$ 和 $\varphi_t(x_t)$ 为模板图像和搜索图像的第 t 帧特征图; $*$ 为卷积运算; $f(z_t, x_t)$ 为两者相似性得分响应图,响应图得分最高的位置即为目标在搜索图像中的位置。

目标位置与模板之间的距离小于阈值 d 时,标记为正,否则标记为负。对于得分图上的每一个位置 $\epsilon \in D \rightarrow \mathbf{R}(D)$ (D 表示相似度得分图),有

$$g_\epsilon = \begin{cases} +1, & |\epsilon - c| \leq d \\ -1, & |\epsilon - c| > d \end{cases}, \quad (2)$$

式中: c 为模板图像的特征图的中心, g 为样本标签。

应用Logistic损失来定义训练的损失函数,离线训练时,通过最小化损失函数优化孪生网络参数 θ :

$$\arg\min_{\theta} E L [g, f(z_t, x_t; \theta)], \quad (3)$$

式中: E 表示期望; L 表示Logistic损失。

2.2 基于注意力机制的在线自适应孪生网络框架

SiamFC的模板分支始终使用第一帧作为模板帧,不能自适应地应对目标变化和环境影响。如图1所示,为了有效地学习目标外观变化和排除相似背景的干扰,本文算法在SiamFC的基础上分别在模板分支和搜索分支上加入了上一帧的跟踪结果,通过注意力机制实现特征融合。假设第 $t-1$ 帧跟

踪器所预测到的边界框为 $[c_x, c_y, w, h]$ (c_x, c_y 表示目标中心位置, w, h 表示目标框大小), 以此将第 $t-1$ 帧图像分别裁剪到 127×127 和 255×255 , 如图 2 所示, 分别输入到搜索分支和模板分支, 相比于模板分支, 搜索分支因为输入的尺寸更大而会保留更多的目标信息。将模板分支和搜索分支得到的

特征图分别通过注意力机制模块进行特征融合, 实现在线学习目标外观变化的目的。上一帧目标信息和当前帧最为相似, 将二者共同作为搜索分支的输入, 将提取到的特征图通过空间注意力机制使得当前帧的背景区域得到抑制。表 1 为具体的网络参数。

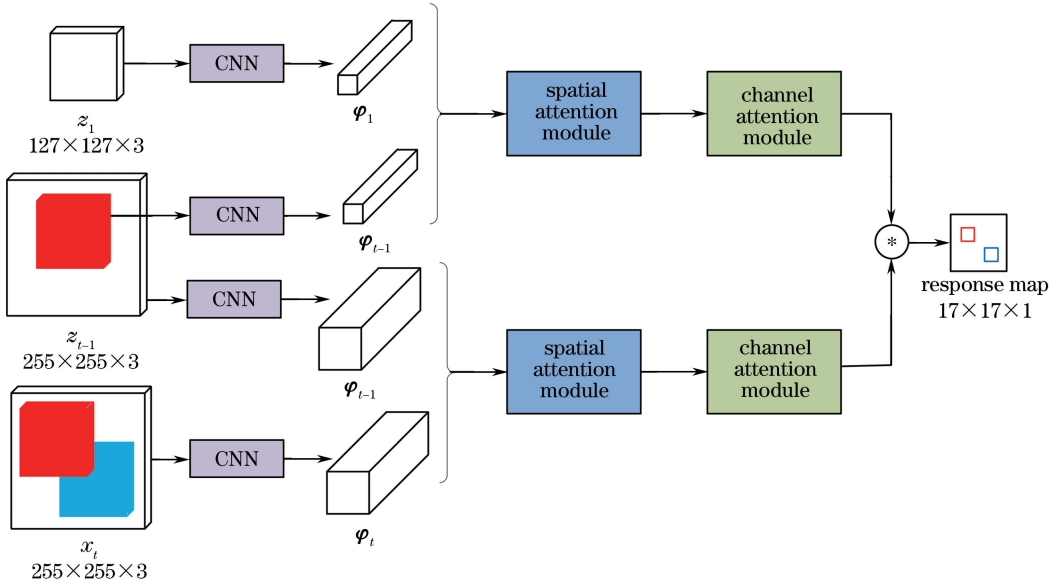


图 1 基于注意力机制的在线自适应孪生网络框架

Fig. 1 Framework of online adaptive siamese network based on attention mechanism



图 2 经过裁剪的第 $t-1$ 帧

Fig. 2 Cropped $(t-1)$ th frame

表 1 SiamFC 网络参数

Table 1 Parameters of SiamFC network

Type	Size /Stride	Number of filters	Size of exemplar	Search size
Conv1	11×11/2	96	59×59	123×123
Pool1	3×3/2		29×29	61×61
Conv2	5×5	256	25×25	57×57
Pool2	3×3/2		12×12	28×28
Conv3	3×3	384	10×10	26×26
Conv4	3×3	384	8×8	24×24
Conv5	3×3	256	6×6	22×22

2.3 注意力机制算法

深度 CNN 所提取的特征含有丰富的高级语义信息^[16], 但是当出现相似性目标时, 高级语义信息缺乏判别性, 容易导致模型漂移。因此为了增

强网络判别能力, 本文设计一种适用于目标跟踪任务的注意力机制, 并将其分为空间注意力模块和通道注意力模块, 通过局部感知域内融合空间和通道方式来构建信息特征, 提高跟踪器的识别

能力。

空间注意力模块通过计算不同特征图中相同位置之间的相似程度来赋予不同的权重,相似程度越大则给予更大的权重,反之,给予更小的权重。空间注意力模块通过建立不同特征图之间的空间信息关

系来增强网络的特征表达能力。如图 3 所示,从网络输出的特征 $\varphi_{t-1} \in \mathbf{R}^{w \times h \times d}$ 和 $\varphi_t \in \mathbf{R}^{w \times h \times d}$ 中取得 ν 位置的特征,并对取出的位置特征计算 Tanimoto 系数相似度,并进行 Softmax 函数运算,得到在 ν 位置上各帧特征图的归一化权重 $\eta_t(\nu)$ 。

$$\eta_t(\nu) = \text{Softmax} \cdot \left[\frac{\varphi_t(\nu)\varphi_{t-1}(\nu)}{\|\varphi_{t-1}(\nu)\|^2 + \|\varphi_t(\nu)\|^2 - \varphi_{t-1}(\nu)\varphi_t(\nu)} \right] \quad (4)$$

将各个位置的特征图进行元素项相加,得到空间注意力输出的特征 φ_{spa} ,即

$$\varphi_{\text{spa}} = \sum_{\nu} [\eta_t(\nu)\varphi_t(\nu) + \eta_{t-1}(\nu)\varphi_{t-1}(\nu)] \quad (5)$$

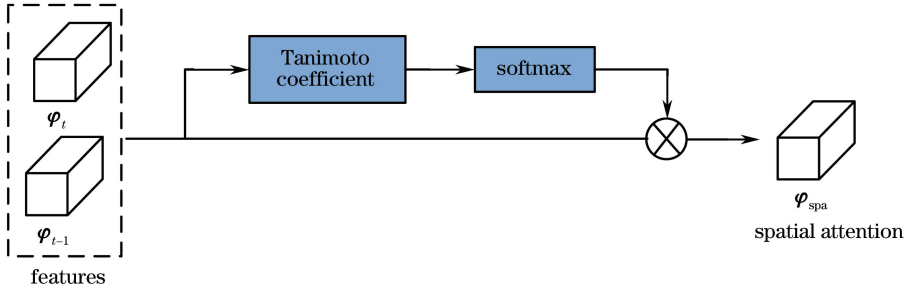


图 3 空间注意力模块

Fig. 3 Spatial attention module

对于通道注意模块,特征图的每个通道都可以视为一种特殊的特征检测器^[17],可通过建模通道之间的相互依赖性来自适应地校准通道方面的特征响应,同时为了能够学习通道之间的非线性相互作用,选用 Sigmoid 函数作为非线性激活函数。如图 4 所示,通道包含全局平均池化层(G-avg-pooling)和两层全连接层(FC),其中 ReLU 为激活函数。为了汇总通道信

息,保持深度网络对目标外观变化的适应能力,将空间注意力模块的输出 $\varphi_{\text{spa}} \in \mathbf{R}^{w \times h \times d}$ 作为通道注意模块的输入,可以得到各个通道 l 的权重参数 $\mathbf{s} = [s_1, s_2, \dots, s_n]$, n 为通道总数。权重参数与输入特征图元素项相乘得到注意力模块的输出特征 φ_{cha} ,即

$$\varphi_{\text{cha}} = \sum_{l=1}^n \mathbf{s} \varphi_{\text{spa}}^{(l)} \quad (6)$$

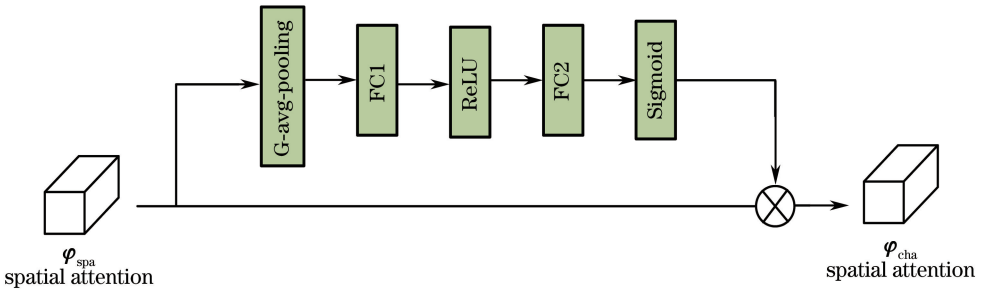


图 4 通道注意力模块

Fig. 4 Channel attention module

3 实验结果及分析

3.1 参数设置

为了检验本文目标跟踪算法 AAM-Siam 的有效性,使用 OTB50^[18]、OTB100^[19]和 VOT2016^[20],将 AAM-Siam 与近几年公认效果较好的算法进行对比实验,用于对比的算法包括 SiamFC、Staple^[21]、

DeepSRDCF^[22]和 KCF^[23]。实验参数设置:采用动量为 0.9 的随机梯度下降法来训练网络;设置权重衰减为 0.0005;学习率以指数衰减方式从 10^{-2} 减少到 10^{-5} ;训练周期为 50 个周期,每个周期包括 53200 组模板图像和搜索图像;每次小批量训练样本数为 8。同时,使用视频目标检测数据集 ILSVRC2015-VID^[24]中的 4417 个视频序列对网络

进行离线训练,训练时忽略目标对象的类别。本次实验使用 Matlab2014a 和 MatConvNet 深度学习框架^[25]实现整个网络框架的跟踪、训练过程。实验的评估都是在一台配置为 Intel i5 CPU 和 NVIDIA1060 GPU 电脑上完成的。

3.2 数据集和评价标准

对于 OTB 数据集,使用跟踪精度(Precision plots)和平均跟踪成功率(Success plots)作为跟踪算法的性能评价标准,并绘制精确度和成功率曲线图。精确度中一个广泛使用的评估指标是中心位置误差(center location error),其定义是跟踪算法所估计的目标边界框的中心点与人工标注边界框的中心点的平均欧几里德距离。然后,使用该序列所有帧上的平均中心位置误差来评估该序列的整体性能,但是,当跟踪器丢失目标时,输出位置可能是随机的,平均误差值可能无法正确测量跟踪性能。为了更加准确地评价跟踪算法,选用对应阈值为 20 pixel 的精度值来衡量算法的跟踪精度,该阈值表征跟踪器边界框与人工标注边界框之间至少重叠 50%^[26],实验结果图中显示了不同跟踪算法在阈值为 20 pixel 处的跟踪精度。成功率使用边界框重合率(bounding box overlap)来进行评估,边界框重合率可以表示为

$$S = \frac{|r_t \cap r_g|}{|r_t \cup r_g|}, \quad (7)$$

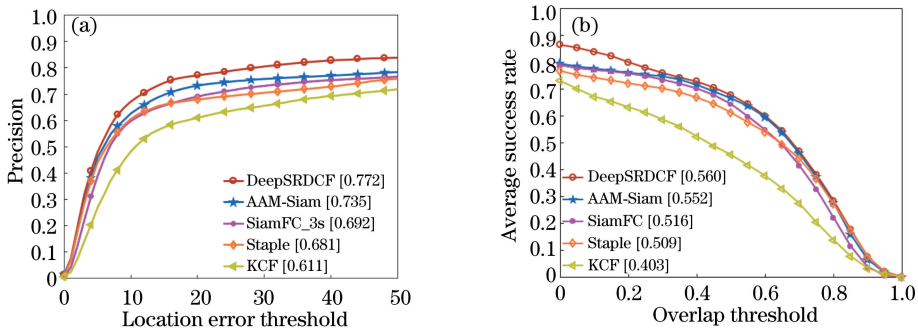


图 5 OTB50 精度图和平均成功率图。(a)精度图;(b)平均成功率图

Fig. 5 Plots of accuracy and average success rate on OTB50 dataset. (a) Plot of accuracy; (b) plot of average success rate

本文在 OTB100 数据集上对包括本文算法在内的 5 个跟踪算法进行了 11 种不同属性的对比分析实验。图 7 和图 8 分别为当跟踪算法在经历相似背景和物体发生形变时两种属性的成功率图,这两种属性表明跟踪遇到了相似目标和物体发生了外观变化。在相似背景的属性下本文算法的跟踪平均成功率为 58.3%,比基础算法 SiamFC 提高了 6.0 个

式中: r_t 为跟踪目标的边框区域; r_g 为目标真实边框区域; \cap 和 \cup 分别为面积交集和并集运算,通常用成功率曲线下的面积来计算实际成功率的得分。

对于 VOT2016 数据集,该数据集包含 60 个人工标注的具有挑战性的视频序列。使用该数据集中定义的期望平均重叠率(EAO)、准确率值(A_v)、稳健性值(R_v)来比较跟踪算法的性能。EAO 是每个跟踪器在视频序列中的非重叠期望, A_v 是统计每帧的预测值与真实值之间的重合率, R_v 是计算每个视频跟踪的平均失败次数。

3.3 定量分析

3.3.1 OTB 数据集结果分析

图 5 为在 OTB50 数据集中包括本文算法在内的 5 种跟踪算法的结果曲线图。由图 5 可以看出,本文算法的跟踪精度为 73.5%,平均成功率为 55.2%,分别比基于模板帧匹配的 SiamFC 算法提高 4.3 个百分点和 3.6 个百分点,分别比使用人工特征的 KCF 算法提高 12.4 个百分点和 14.9 个百分点。图 6 为各个跟踪算法在 OTB100 数据集上的跟踪结果对比。由图 6 可以看出,本文算法的跟踪精度为 80.9%,平均成功率为 61.3%,分别比基础 SiamFC 算法高出 3.8 个百分点和 3.1 个百分点。本文算法在融入了上一帧跟踪结果的同时引入注意力机制来实现特征的实时融合和更新,因此能够提高跟踪的稳健性。

百分点。在面对形变属性时本文算法也取得了较好的表现。SiamFC 在面对这两种属性时得分较低,说明该算法的稳健性较差。因为本文算法将稳健的第一帧与上一帧通过注意力机制进行特征融合,能够通过局部感知域内融合空间和通道方式来构建稳健的信息特征,从而提升算法的稳健性。

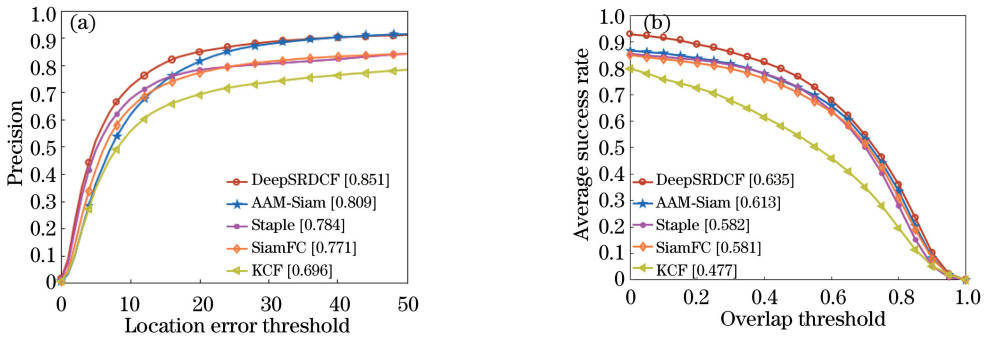


图 6 OTB100 精度图和平均成功率图。(a)精度图;(b)平均成功率图

Fig. 6 Plots of accuracy and average success rate on OTB100 dataset. (a) Plot of accuracy; (b) plot of average success rate

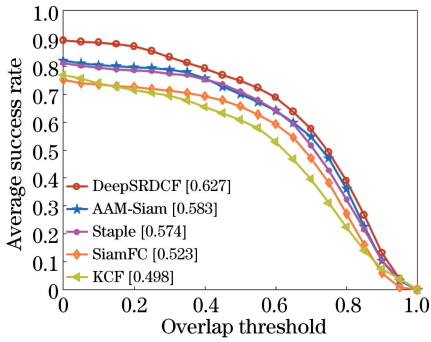


图 7 OTB100 中相似背景属性的平均成功率

Fig. 7 Average success rates of similar background attributes on OTB100 dataset

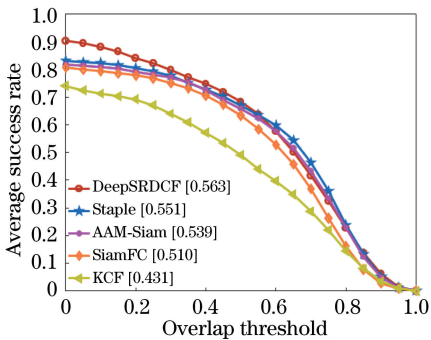


图 8 OTB100 中形变属性的平均成功率

Fig. 8 Average success rates of deformation properties on OTB100 dataset

3.3.2 VOT2016 数据集结果分析

为了进一步验证本文算法的稳健性,将本文算法和上述算法在 VOT2016 测试集中进行测试,实验结果如图 9 和表 2 所示。由表 2 可以看出,在期望平均重叠率(EAO)上,本文算法比基准算法 SiamFC 提高 4 个百分点,从而进一步验证本文算法具有较强的稳健性,可以实现很好的跟踪效果。

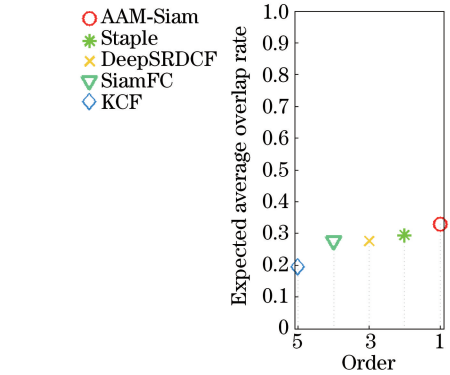


图 9 在 VOT2016 数据集上的期望平均重叠率

Fig. 9 Expected average overlap rates on VOT2016 dataset

表 2 在 VOT2016 数据集上的精度和稳健性评估

Table 2 Evaluation of accuracy and robustness on VOT2016 dataset

Tracker	EAO	Accuracy	Robustness
AAM-Siam	0.32	0.56	0.32
Staple	0.30	0.54	0.38
DeepSRDCF	0.28	0.52	0.33
SiamFC	0.28	0.55	0.38
KCF	0.19	0.49	0.57

3.4 定性分析

为了检验本文算法,在 OTB 数据集上选取 7 个具有挑战性的视频序列属性进行测试,这 7 个视频序列分别是 Coke、Football、Lemming、Shaking、Singer2、Walking2、Ironman,这些视频序列包含了不同的属性,如光照变化(IV)、遮挡(OCC)、目标形变(DEF)、运动模糊(MB)、相似背景(BC)、图像低分辨率(LR)。图 10 为各个跟踪算法在 7 个挑战序列上的跟踪结果。

相似背景:在序列 Football 中,本文算法由于

保留了稳健的第一帧的同时融合前一帧的历史信息,并且利用了注意力机制,能很好地建立空间和通道的联系,能够使模板分支有效抑制相似背景之间的混杂,从而对当前目标位置作出有效判断。

目标形变:在序列 Singer2 中,SiamFC 在前期能够跟踪目标,但在后期跟踪失败,原因是相机出现了快速移动,导致图像出现模糊现象,序列后续帧的目标信息相比于第一帧已经发生了很大的变化,使用第一帧作为模板匹配从而影响了跟踪效果;而本文算法则能够进行稳健的跟踪。在序列 Ironman 中,由于目标在每一帧都发生剧烈形变,利用第一帧

的信息难以确定目标位置,从而导致 SiamFC 跟踪失败。

光照变化:在序列 Lemming 中,视频开始阶段本文算法能够对目标进行有效跟踪,但在第 61 帧时由于光照发生显著变化,跟踪器无法处理如此大的突变导致跟踪失败。

遮挡和运动模糊:在序列 Coke 中,本文算法和其他跟踪算法都可以得到较好的跟踪效果。在序列 Walking2 中,在跟踪后期 SiamFC 算法出现了跟踪失败的现象,而本文算法由于融合了上一帧的历史信息,能够对目标进行稳健的跟踪。



图 10 不同算法在 7 个富有挑战性的视频序列上的定性评估

Fig. 10 Qualitative evaluation of different algorithms on seven challenging video sequences

4 结 论

针对全卷积孪生网络在相似目标共存和目标外观发生显著变化时的跟踪失败问题,提出一种基于注意力机制的在线自适应孪生网络跟踪算法(AAM-Siam)。该算法将前一帧跟踪所得到的结果输入到网络,再通过空间注意力模块和通道注意力模块在线学习目标形变并抑制背景,弥补应对目标外观变化的不足。通过在 OTB 和 VOT2016 数据集集中的对比实验可以看出,本文算法比基准算法

SiamFC 跟踪更加精确,可以实现很好的跟踪效果。但是当强光变化等其他干扰因素出现时,本文算法跟踪效果不太理想,未来将着重在此方面进行改进,以达到更加稳健的跟踪效果。

参 考 文 献

- [1] Li P, Zhang Y. Video smoke detection based on Gaussian mixture model and convolutional neural network [J]. Laser & Optoelectronics Progress, 2019, 56(21): 211502.

李鹏,张炎.基于高斯混合模型和卷积神经网络的视

- 频烟雾检测[J]. 激光与光电子学进展, 2019, 56(21): 211502.
- [2] Li Z R, Wang K X, He X L, et al. Heel-strike event detection algorithm based on convolutional neural networks [J]. *Laser & Optoelectronics Progress*, 2019, 56(21): 211503.
李卓容, 王凯旋, 何欣龙, 等. 基于卷积神经网络的足跟着地事件检测算法[J]. 激光与光电子学进展, 2019, 56(21): 211503.
- [3] Wang D C, Chen X N, Zhao F, et al. Vehicle detection algorithm based on convolutional neural network and RGB-D images [J]. *Laser & Optoelectronics Progress*, 2019, 56(18): 181003.
王得成, 陈向宁, 赵峰, 等. 基于卷积神经网络和RGB-D图像的车辆检测算法[J]. 激光与光电子学进展, 2019, 56(18): 181003.
- [4] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 4293-4302.
- [5] Song Y B, Ma C, Gong L J, et al. CREST: convolutional residual learning for visual tracking[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2574-2583.
- [6] Wang Q, Teng Z, Xing J L, et al. Learning attentions: residual attentional Siamese network for high performance online visual tracking [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 4854-4863.
- [7] Chu Q, Ouyang W L, Li H S, et al. Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 4846-4855.
- [8] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR), June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 7132-7141.
- [9] Yang D W, Gong X F, Mao L, et al. Multi-domain convolutional neural network tracking algorithm based on reconstructed feature combination[J]. *Laser & Optoelectronics Progress*, 2019, 56(19): 191501.
杨大伟, 巩欣飞, 毛琳, 等. 重构特征联合的多域卷积神经网络跟踪算法[J]. 激光与光电子学进展, 2019, 56(19): 191501.
- [10] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking[M]//Hua G, Jégou H. *Computer vision-ECCV 2016 workshops. Lecture notes in computer science*. Cham: Springer, 2016, 9914: 850-865.
- [11] Held D, Thrun S, Savarese S. Learning to track at 100 FPS with deep regression networks[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 749-765.
- [12] Fan H, Ling H B. SANet: structure-aware network for visual tracking [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2217-2224.
- [13] He A F, Luo C, Tian X M, et al. A twofold Siamese network for real-time object tracking [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 4834-4843.
- [14] Zhang Y H, Wang L J, Qi J Q, et al. Structured Siamese network for real-time visual tracking[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11213: 355-370.
- [15] Kashiani H, Shokouhi S B. Visual object tracking based on adaptive Siamese and motion estimation network[J]. *Image and Vision Computing*, 2019, 83/84: 17-28.
- [16] Li L X. Research on visual tracking based on deep learning [D]. Harbin: Harbin Institute of Technology, 2018: 31-32.
李蓝星. 基于深度学习的目标跟踪算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2018: 31-32.
- [17] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [M] // Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8689: 818-833.
- [18] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013,

- Portland, OR, USA. New York: IEEE, 2013: 2411-2418.
- [19] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [20] Kristan M, Leonardis A, Matas J, et al. The visual object tracking vot2016 challenge results [M] // Hua G, Jégou H. Computer vision-ECCV 2016 Workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 777-823.
- [21] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1401-1409.
- [22] Danelljan M, Hager G, Khan F S, et al. Convolutional features for correlation filter based visual tracking [C] // 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 621-629.
- [23] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.
- [24] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [25] Vedaldi A, Lenc K. MatConvNet: convolutional neural networks for MATLAB [C] // Proceedings of the 23rd ACM international conference on Multimedia-MM'15, October 26-30, 2015, Brisbane, Australia. New York: ACM, 2015: 689-692.
- [26] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632.