

基于信息熵的级联 Siamese 网络目标跟踪

赵钦炎¹ 李宗民¹ 刘玉杰¹ 李 华²

1 中国石油大学(华东)计算机与通信工程学院 山东 青岛 266580

2 中国科学院计算技术研究所智能信息处理重点实验室 北京 100190

(zhaoqinyan@s.upc.edu.cn)



摘 要 目标跟踪是计算机视觉领域的一个重要研究方向,针对目前算法对于目标外观变化的鲁棒性较差等问题,提出了一种基于信息熵的级联 Siamese 网络目标跟踪方法。首先利用孪生神经网络(Siamese network)对第一帧目标模板和当前帧待检测区域提取深度卷积特征,并通过相关性计算响应图;然后根据定义的信息熵和平均峰值系数评价响应图质量,针对质量差的响应图对卷积特征进行模型因子更新;最后利用最终的响应图确定目标位置并计算最佳尺度。在 VOT2016, VOT2017 数据集上进行实验,结果证明在保证实时运行的基础上所提算法的精度优于其他算法。

关键词: 目标跟踪;信息熵;神经网络;尺度估计

中图法分类号 TP391.41

Cascaded Siamese Network Visual Tracking Based on Information Entropy

ZHAO Qin-yan¹, LI Zong-min¹, LIU Yu-jie¹ and LI Hua²

1 College of Computer & Communication Engineering, China University of Petroleum, Qingdao, Shandong 266580, China

2 Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Abstract Visual tracking is an important research direction in the field of computer vision. In view of the problems such as poor robustness of the current algorithms to object appearance changes, this paper proposes a cascaded Siamese network visual tracking method based on information entropy. Firstly, the deep convolution feature is extracted from the first frame target template and the area to be detected of the current frame by using the Siamese network, and the response map is calculated by correlation. Then, the quality of the response map is evaluated according to the defined information entropy and the average peak coefficient, and for the response map with poor quality, the model factor of convolution feature is updated. Finally, the final response map is used to determine the target position and calculate the optimal scale. The experimental results on VOT2016 and VOT2017 datasets show that the proposed method is superior to other algorithms on the basis of ensuring real-time operation.

Keywords Visual tracking, Information entropy, Neural networks, Scale estimation

1 引言

随着信息技术的发展和深度学习的进一步探索,图形图像技术迎来了新的高峰。目标跟踪是计算机视觉领域的一个重要研究方向,其目的是对于给定视频序列第一帧中目标的位置和大小,在整个视频序列中估计目标位置并定位目标。目标跟踪在视频监控、人机交互、医学图像处理、智慧交通等领域已被广泛应用,但是仍然面临诸如目标快速运动、目标遮挡、光照变化、非刚性运动、背景复杂和目标出视野等问题。

目前,在该研究领域内,国内外学者已开展了相关的工作,并取得了较大的进展。相关滤波器最早应用于信号处理领域,它的思想在近年才被用于计算机视觉领域,并取得了

较好的效果。Bolme 等^[1]首次将相关滤波器的思想应用在目标跟踪领域,提出了最小输出误差平方和算法(Minimum Output Sum of Squared Error, MOSSE),通过快速傅里叶变换将时域中复杂的运算转化为频域中的点乘运算来进行滤波器模板的求解,因此可以快速地跟踪到目标,但是其对灰度特征的特征能力限制了其性能。Henriques 等^[2]在 MOSSE 算法的基础上提出了 KCF(Kernel Correlation Filter)算法,利用循环矩阵采集正负样本,使用核函数提高分类器性能,并定义了一种能融合多通道特征的方法,极大地提高了跟踪器的性能。Wang 等^[3]从相关滤波响应图入手,对响应图进行多峰目标检测并提出高置信度的模型更新策略,避免了模型频繁更新导致跟踪速度变慢。

到稿日期:2019-08-30 返修日期:2019-10-30 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61379106,61379082,61227802);山东省自然科学基金(ZR2013FM036, ZR2015FM011);中央高校基本科研业务费专项资金资助(18CX06046A)

This work was supported by the National Natural Science Foundation of China (61379106,61379082,61227802), Shandong Provincial Natural Science Foundation (ZR2013FM036, ZR2015FM011) and Fundamental Research Funds for the Central Universities (18CX06046A).

通信作者:李宗民(lizongmin@upc.edu.cn)

随着深度学习在视频领域中的深入应用,目标跟踪领域涌现出了一批优秀的深度学习跟踪方法。Ma 等^[4]提出了层次化卷积特征跟踪方法(Hierarchical Convolutional Features, HCF),放弃使用传统手工特征,使用 VGG 网络提取卷积特征,并结合了深层卷积特征语义信息和浅层卷积特征细节信息,但是深度网络限制了跟踪器的速度。Li 等^[5]在 HCF 的基础上,使用单层卷积特征和 CN 特征,融合多模板结果,使得跟踪速度得到提高。Nam 等^[6]提出一个端到端的多域网络(Multi-Domain Convolutional Neural Networks, MDNet),从多个标注的视频序列中学习物体共享的表示,在测试时将最后的多个全连接层替换为一个全连接层并在线微调。该方法使用浅层网络,但得到了较好的效果。

近年来,孪生神经网络由于拥有简单的网络结构和强大的相似性度量能力,得到了国内外学者的重视。Bertinetto 等^[7]使用 Siamese 网络进行目标跟踪,提出了 SiamFC 方法,并使用 ILSVRC 数据集训练两个相同的网络分支,在测试阶段输入模板分支和检测区域分支,通过计算相关性得到响应图。SiamFC 在跟踪过程中平衡了速度和精度,但其本身不能更新模板,导致在跟踪过程中不能很好地捕捉外观随时间变化的目标。Guo 等^[8]针对 SiamFC 不能更新模板的问题,引入更新因子和变换因子模拟更新模板,根据上一帧目标外观计算更新因子,有效解决了目标外观变化的问题,但是逐帧求解更新因子会导致跟踪速度大幅下降。He 等^[9]为了解决 SiamFC 方法泛化能力差的问题,提出了一个结合语义分支和外观分支的网络结构(Semantic Branch and Appearance Branch Siamese Network, SA-Siam),他们发现相似性学习的外观特征和分类任务的语义特征具有互补性质,因此将两个分支的响应图加权融合在一起得到最终的响应图来定位目标位置。

现有的目标跟踪方法主要存在以下问题:1)不能在跟踪过程中很好地应对目标外观变化;2)难以平衡跟踪过程中的速度和精度;3)不能有效应对跟踪过程中目标尺度发生的变化。针对以上问题,本文提出了一种基于信息熵的 Siamese 网络目标跟踪算法。该方法采用级联分类的思想,首先使用 Siamese 网络卷积得到响应图,对得到的响应图使用本文提出的信息熵和平均峰值系数联合判断响应图质量,若单峰较突出且无旁瓣峰值,则认为响应图的质量较高,直接根据响应图定位目标;当响应图纯净度较低时,放弃使用当前响应图,对两输入分支分别计算更新因子和变换因子,使跟踪器适应跟踪过程中目标外观的变化,重新计算响应图并定位目标。最后,该方法通过贝叶斯统计最大化后验概率,得到最佳尺度。所提算法的具体实现流程如图 1 所示。

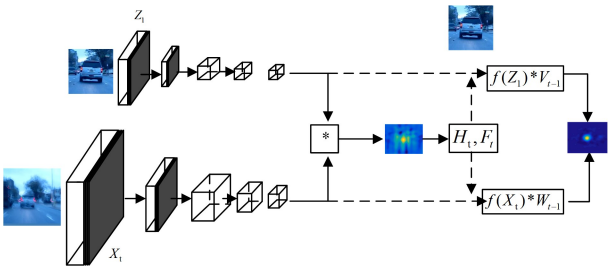


图 1 基于信息熵的级联 Siamese 网络目标跟踪算法的流程图

Fig. 1 Flow chart of Siamese network visual tracking algorithm based on information entropy

本文提出的响应图检测策略避免了在每一帧中重复更新模型造成计算复杂度较高的问题,能在保证精度的同时提高运行速度。本文在 VOT 公共数据集上同近年来的一些优秀方法进行了对比,验证了本文方法的精确度和高效性。

2 Siamese 网络跟踪算法

孪生神经网络指两个分支网络共享权值,将输入映射到新的空间,通过计算损失函数,最终衡量两个分支网络输入的相似度。在计算机视觉中,其常用于人脸识别^[10]和图像检索等任务。

文献[7]是将孪生神经网络应用于目标跟踪任务的一个成功案例,其在视频数据集中采集图像对,离线学习匹配函数,用于衡量输入的相似性。SiamFC 网络拥有两个输入,其中模板分支为 $Z_t \in R^{m \times n \times 3}$,其输入是视频序列第一帧的目标模板,在跟踪过程中,在待检测区域分支的输入为当前帧的待检测区域 $X_t = R^{m_t \times n_t \times 3}$ 。SiamFC 方法使用全卷积网络的方式计算响应图,其每帧响应图的计算方式如式(1)所示:

$$S_t = f(Z_t) \otimes f(X_t) \tag{1}$$

其中, $f(\cdot)$ 代表通过神经网络提取到全卷积特征; $\otimes$ 代表进行互相关计算,计算两个特征图的相关性从而得到响应图,将响应图插值到待检测区域大小,响应图的最高峰值所在位置即为当前帧目标位置。

本文方法采用 SiamFC 框架初步得到响应图,但是由于得到的响应图在目标跟踪过程中的外观变化鲁棒性较差,因此本文在得到响应图后通过结合信息熵检测策略来检测响应图质量,对于质量较差的响应图,根据上一帧目标外观来训练更新因子和变换因子,拟合目标外观变化以重新定位目标。

3 基于信息熵的级联 Siamese 网络目标跟踪

通过双分支 Siamese 网络得到视频序列中每一帧的响应图^[7],使用信息熵和峰值来判断响应图的质量,结合更新因子和变换因子对质量较差的响应图进行二次检测和重定位,最后利用贝叶斯统计得到拥有最大后验概率的尺度,并将其作为当前帧的最佳尺度。

3.1 结合信息熵检测响应图质量的方法

理想情况下的响应图应该只有一个峰值且峰值周围较为平滑,在目标外观发生变化、遮挡或者出现相似物体的情况下响应图会产生多峰或者不平滑现象。评价孪生神经网络响应图的质量使用了两个指标:1)响应图信息熵;2)响应图的平均峰值系数。我们称该方式为结合信息熵的评价方式。

信息熵于 1948 年由 Shannon^[11]首次从热力学中借鉴来解决信息度量问题,其表达式如式(2)所示:

$$H(X) = -\sum p(x_i) \log(p(x_i)), i = 1, 2, \dots, n \tag{2}$$

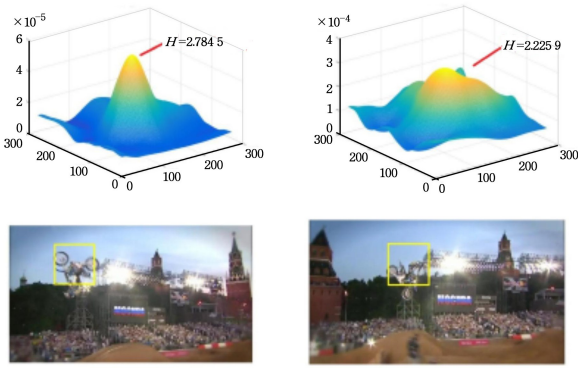
其中, X 表示随机变量,取值为 $(x_1, x_2, \dots, x_n)$; $p(x_i)$ 表示事件 x_i 发生的概率。信息熵在图像处理中常用来计算灰度图的灰度分布的聚集特征所包含的信息量。

首先使用信息熵得分来判断响应图的质量,计算公式为:

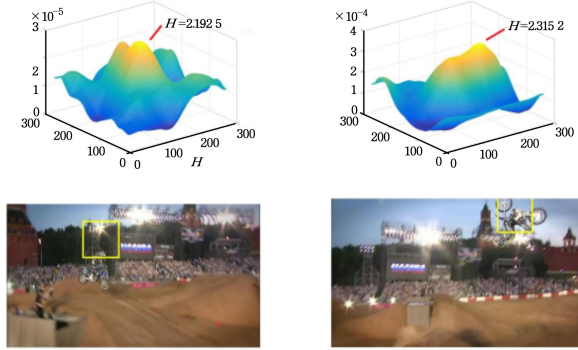
$$H_t = \frac{1}{-\sum_{i=0}^k p(x_i) \log(p(x_i))} \tag{3}$$

信息熵在计算灰度图时取值是离散的,深度卷积响应图为连续取值,本文将响应图连续值离散为 k 个区间, k 为常

数,其取值的可视化效果如图 2(a)所示。



(a) 单峰值和平滑情况下的响应图



(b) 多峰值和山脊情况下的响应图

图 2 信息熵可视化图

Fig. 2 Visualization of information entropy

一般情况下,信息熵得分越大,得到的响应图质量越高,但是在高取值区间内数据增多会对信息熵得分产生负影响,即响应图出现类似山脊峰值时仍然会取得高值,如图 2(b)所示。因此,本文对响应图分布本身进行评价,提出结合平均峰值系数来辅助判断响应图质量,判断公式如式(4)所示:

$$F_t = \frac{\max(S_t) - \text{mean}(S_t)}{\text{std}(S_t)} \quad (4)$$

其中, $\max(\cdot)$ 和 $\text{mean}(\cdot)$ 函数分别用于计算响应图的最大值和均值; $\text{std}(\cdot)$ 函数用于计算标准差,其效果图如图 3 所示。当 H_t 和 F_t 均大于历史均值时,认为当前帧的响应图质量较好,直接通过响应图定位目标;当 H_t 和 F_t 至少有一个小于其历史均值时认为目标外观变化导致响应图质量下降,放弃使用当前响应图,根据上一帧目标外观求解更新因子和变换因子,重新计算新的响应图。

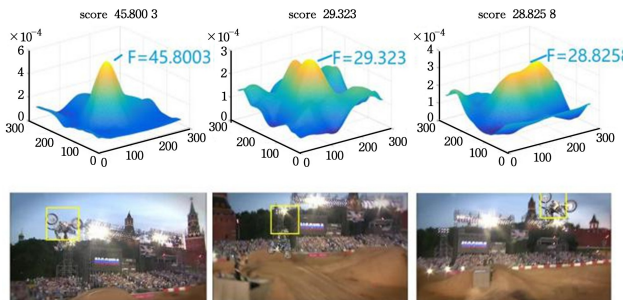


图 3 平均峰值系数

Fig. 3 Average peak coefficient

3.2 模型更新方法

根据文献[8]分别计算模板和待检测区域分支的更新因

子和变换因子,来应对目标外观变化。更新因子作用于 Siamese 网络的模板分支,使模板适应于跟踪过程中的外观变化;变换因子作用于待检测区域分支,旨在突出当前帧目标的特征,并减轻背景对跟踪目标的影响。

对于质量较差的响应图,使用式(5)重新计算响应图。

$$S_t = [f(Z_1) * V_{t-1}] \otimes [f(X_t) * W_{t-1}] \quad (5)$$

其中, V_{t-1} 是更新因子, W_{t-1} 是变换因子, $*$ 代表循环卷积操作。求解过程使用线性回归正则化 (Regularized Linear Regression, RLR) 方法。

$$V_{t-1} = \arg \min_V \| V * f(Z_1) - f(Z_{t-1}) \|^2 + \lambda_v \| V \|^2 \quad (6)$$

$$V_{t-1} = F^{-1} \left(\frac{F^*(f(Z_1)) \cdot F(f(Z_{t-1}))}{F^*(f(Z_1)) \cdot F(f(Z_1)) + \lambda_v} \right) \quad (7)$$

当前帧更新因子 V_{t-1} 由第一帧目标模板图像 Z_1 和上一帧目标外观图像 Z_{t-1} 共同计算得到,目的是学习一个变换矩阵,使模板外观更接近上一帧目标的外观,以应对跟踪过程中目标外观的突然变化。由于循环卷积的理想特性,更新因子的求解可以通过离散傅里叶变换加速进行。式(7)中, F 代表离散傅里叶变换, F^{-1} 代表逆变换。

同理,变换因子 W_{t-1} 也使用 RLR 方法,考虑到目标在帧与帧之间的相对位置不会变化太大,变换因子根据上一帧目标中心位置,使用一个高斯函数来突出目标外观,同时减轻候选区域周围背景的干扰。

$$W_{t-1} = \arg \min_W \| W * f(G_{t-1}) - f(\bar{G}_{t-1}) \|^2 + \lambda_w \| W \|^2 \quad (8)$$

$$W_{t-1} = F^{-1} \left(\frac{F^*(f(G_{t-1})) \cdot F(f(\bar{G}_{t-1}))}{F^*(f(G_{t-1})) \cdot F(f(G_{t-1})) + \lambda_w} \right) \quad (9)$$

G_{t-1} 是以上一帧目标为中心截取的与 X_{t-1} 大小相同的区域, $\bar{G}_{t-1}$ 是 G_{t-1} 乘一个高斯函数得到的图像。

通过计算更新因子和变换因子,使当前帧的模板可以学习到上一帧的目标外观,也使得目标区域得到更大的权重,让静态的 Siamese 网络对目标外观的变化有了更好的辨别能力,也有效提升了跟踪框架的鲁棒性。

3.3 尺度估计

目标的大小在跟踪过程中常常发生变化,我们称之为尺度变化。除了 Siamese RPN^[12-13] 使用 RPN 网络回归目标大小,大部分 Siamese 网络仍然使用尺度池划分多尺度区域的方法计算目标尺度,提高了计算时间复杂度。本文虽然也使用尺度池的方法,但是利用高斯函数作为先验概率,使用当前帧不同尺度的观测值,通过贝叶斯统计最大后验概率来计算最佳的目标尺度。

出于对算法时效性的考虑,本文采用三尺度的方法,即尺度池 $S = \{s_1, s_2, s_3\}$ 包含 3 个尺度系数,其中 $s_i = \gamma_n s$, s 是上一帧的尺度系数。根据贝叶斯统计,后验概率的最大值为当前帧的最佳尺度,如式(10)所示:

$$s = \arg \max P(s_i | y) = P(y | s_i) P(s_i) \quad (10)$$

其中, $s_i \in S$, 观测值 $P(s_i | y)$ 为检测目标的最大响应值,其计算方法为 $P(s_i | y) = \arg \max(S_{s_i})$, 先验项 $P(s_i)$ 定义为一个高斯函数 $P(s_i) = N(\mu, \sigma^2)$, 高斯函数的中心为上一帧最优尺度位置的中心。

3.4 基于信息熵的级联 Siamese 网络目标跟踪算法

基于信息熵的级联 Siamese 网络目标跟踪算法的具体描述如算法 1 所示。

算法 1 基于信息熵的级联 Siamese 网络目标跟踪

输入: 第一帧目标模板 Z_1 和位置 P_1 , 孪生神经网络 N

输出: 第 t 帧目标位置 P_t

- Step1 使用网络 N 对模板 Z_1 提取特征 $f(Z_1)$;
- Step2 进入下一帧并确定当前帧的待检测区域 X_t , 使用网络 N 提取特征 $f(X_t)$;
- Step3 利用 $f(X_t)$ 和 $f(Z_1)$ 进行相关计算, 并根据式 (3)、式 (4) 计算信息熵得分 H_t 和平均峰值系数 F_t ;
- Step4 如果 H_t 和 F_t 至少有一个小于历史均值, 则进入 Step5, 否则直接进入 Step6;
- Step5 根据式 (7)、式 (9) 计算 V_{t-1} 和 W_{t-1} , 通过式 (5) 计算最终响应图;
- Step6 根据响应图确定目标位置 P_t , 并根据最大后验概率计算最佳尺度, 重复 Step2—Step6 直到视频序列最后一帧结束。

4 实验与分析

4.1 实验设置

本文方法使用的硬件平台为 Intel(R) Xeon(R) CPU e5-2675 @ 1.80GHz, 1.80GHz (2 处理器), 128 GB RAM, GTX TITAN 显卡的服务器, 软件平台为 MATLA B2017a, 使用 MATCONVET 工具箱。为验证本文算法的有效性, 本文使用两个基准数据集 VOT2016^[14] 与 VOT2017^[15]。VOT2016 和 VOT2017 分别拥有 60 个精确标注的视频序列, 并对序列进行了相机移动 (camera_motion)、光照变化 (illum_change)、移动变化 (motion_change)、闭塞 (occlusion)、尺度变化 (size_change)、目标出视野 (empty) 6 种视觉标注。跟踪器使用精度 (accuracy)、鲁棒性 (robustness) 和期望平均重叠率 (EAO) 作为主要评价标准, 并在目标丢失时被初始化。

此外, 本文实验的对比算法分为两组, 第一组能够实时运行的算法, 包括 HCF^[4], siamFC^[7], colorKCF^[2], ASMS^[16], DAT^[18], DSS-T^[20]; 第二组是近期提出的优秀跟踪器, 包括 Dsiam^[8], PBTS^[17], DensSiam^[19], DCFNet^[21], MEEM^[22]。

4.2 实验细节

为了确保实验的公平性, 分支网络同文献[8]一样使用浅层 Alexnet, 且使用最后一层特征 conv5 求解响应图。在训练过程中使用 logistics 损失。

$$\ell(y, v) = \log(1 + \exp(-yv)) \tag{11}$$

y 是响应图的 groundtruth 取值 $\{-1, 1\}$, v 是真实得分。使用随机梯度下降的方法来求解参数, 值得注意的是, 在训练过程中 2.1 节中的评价方法。训练数据集使用 ILSVC-2015 数据集, 其包括 4 500 个视频和 30 个目标类, 具体参数设置与文献[8]中的初始化方法相同。其他一些具体的参数设置如表 1 所列。

表 1 参数设置表

Table 1 Parameter settings table

参数名称	取值	参数说明
m, n	127	模板分支输入
m_z, n_z	255	待检测分支输入
σ	0.1	高斯核函数带宽
μ	0.01	学习率
S	{0.975, 1, 1.025}	尺度缩放系数
k	30	信息熵区间数

4.3 性能评估指标

Accuracy 用来评价 tracker 跟踪目标的准确度, 它借助了 IoU (Intersection-over-Union) 的方法, 其计算方式为预测区域和真值区域的交集与并集之比, 数值越大, 准确度越高。

Robustness 用来评价 tracker 跟踪目标的稳定性, 数值越大, 稳定性越强。

EAO (Expected Average Overlap) 首先按照视频长度对所有序列进行分类, 求得每一帧的 Accuracy, 并对所有序列的所有帧求平均, 数值越高代表跟踪器的效果越好。

EFO (Equivalent Filter Operations) 是用来衡量跟踪速度的指标, 先测量在一个 $600 * 600$ 的灰度图像上用 $30 * 30$ 最大值滤波器进行滤波的时间, 以此得出一个基准单位, 再以这个基础单位衡量 tracker 的速度, 以减少硬件平台和编程语言等外在因素对 tracker 速度的影响。

4.4 定量分析

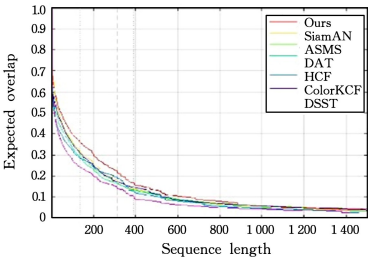
在 VOT2016 数据集上将本文算法与能够实时运行的方法进行对比, 结果如表 2 和图 4 所示。

表 2 VOT2016 上的实验结果

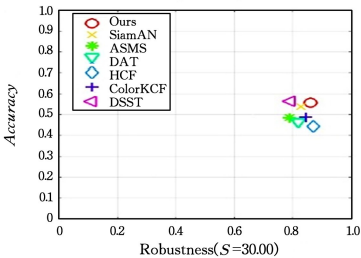
Table 2 Experimental results on VOT2016

tracker	Accuracy	Robustness	EAO	EFO
Ours	0.56	13.67	0.25	8.23
SiamAN	0.54	18.17	0.22	9.12
ASMS	0.49	18.67	0.20	82.58
DAT	0.46	18.00	0.19	18.98
HCF	0.44	13.67	0.21	1.06
ColorKCF	0.49	14.67	0.21	91.46
DSST	0.56	25.00	0.17	12.75

注: 加粗数据为最优结果, 斜体数据为次优结果



(a)



(b)

图 4 VOT2016 上的期望重叠曲线和 A-R 图

Fig. 4 Expected overlap curve (top) and A-R map (bottom) on VOT2016

可以看到, 本文方法取得了最佳的精度, 且能保证实时运行。同基准方法 SiamFC 相比, EAO 上升了 0.03。本文方法由于使用深度学习, 在跟踪速度上比 DAT, DSST, ColorKCF 等使用手工特征的要慢, 但是精度分别提高了 0.06, 0.08,

0.04。与同样使用深度学习的 HCF 算法相比,本文方法的 EAO 上升了 0.04,且能以多倍于 HCF 算法的速度实施。通过图 4 所示的期望重叠曲线可以看到,在 400 帧以内的视频序列中本文方法保证了较高的期望重叠率,说明本文的判断方法和更新方案能够有效提高短时视频序列的精度,但是对于长时跟踪其优势逐渐降低。

在 VOT2017 数据集上将本文方法与当下最优秀的几种算法进行了比较,虽然 VOT2017 视频序列数量与 VOT2016 相同,但是极大地增加了标注的难度。实验对比结果如图 5 和表 3 所示。

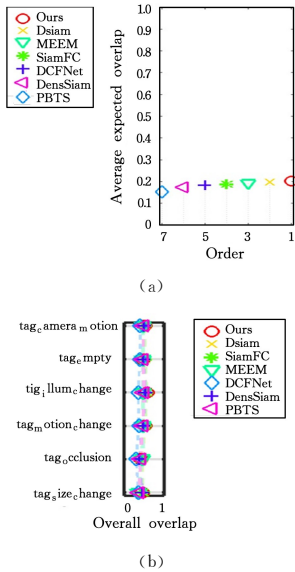


图 5 VOT2017 上的期望重叠率排序和挑战标注排序
Fig. 5 Expected overlap sort,challenge ordering on VOT2017

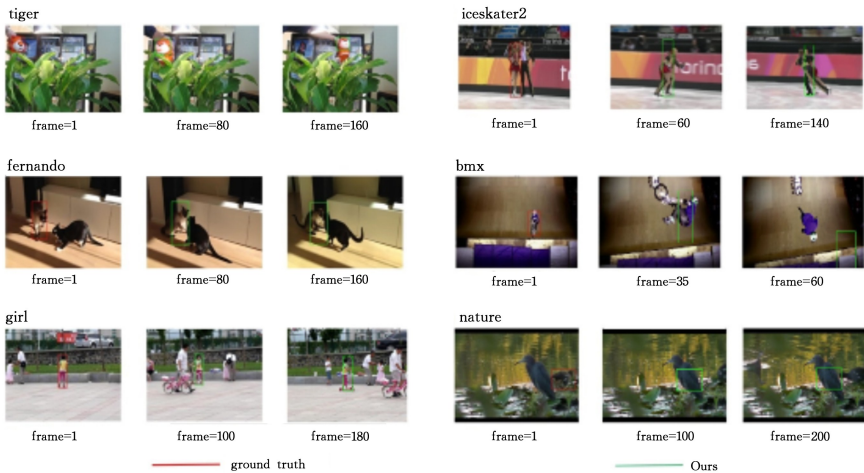


图 6 本文方法在部分序列上的跟踪结果(电子版为彩色)
Fig. 6 Tracking results of method in this paper on partial sequences

结束语 本文在 Siamese 网络框架的基础上,对模板和当前帧的待搜索区域进行计算得到相关性响应图,通过信息熵和平均峰值系数评价响应图的质量和信信息量。我们认为拥有单峰值且旁瓣平滑的响应图质量较好,目标外观变化可能导致响应图的质量下降。对于质量差的响应图,本文方法根据目标上一帧外观计算更新因子和变换因子,并重新构建响应图,通过响应图确定目标位置,使用贝叶斯统计方法计算最

表 3 VOT2017 上的实验结果

Table 3 Experimental results on VOT2017

tracker	EAO	FPS
Ours	0.20	25.17
Dsiam	0.20	9.64
SiamFC	0.19	31.89
MEEM	0.19	4.12
PBTS	0.15	13.04
DCFNet	0.18	27.26
DensSiam	0.17	19.02

注:粗体数据为最优结果,斜体数据为次优结果

本文方法在 EAO 得分上与 Dsiam 并列第一,但是由于本文方法不是每帧更新因子,速度上比 Dsiam 每秒提高了 10 多帧的处理速度。与同样使用 Siamese 网络作为基本网络框架的 SiamFC 和 DCFNet 相比,本文方法在速度在略低于 SiamFC 的情况下,EAO 分别提高了 0.01 和 0.02。与 DensSiam 算法和 PBTS 算法相比,本文方法的 EAO 分别提高了 0.03 和 0.05。在视频序列的各项挑战中,其效果接近专注于解决样本污染问题的 MEEM 算法,尤其在尺度变化、光照变化、移动变化等条件下取得了较好的效果。

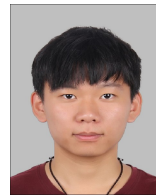
4.5 定性分析

我们在测试序列中选择了 6 个具有挑战性的序列进行定性分析,如图 6 所示。其中,红色区域标注的是第一帧给定的真值,绿色区域是本文方法的结果。由图 6 可知,在 tiger 序列中,本文方法在应对杂乱环境下的目标遮挡问题时能够精确跟踪到目标,这得益于本文方法没有逐帧更新导致模型漂移。在 iceskater2 和 farnando 序列中,本文方法能够较好地应对目标外观变化和光照变化。对于 bmx 序列,本文方法虽然能够有效应对尺度变化,但是对于目标外观剧烈变化的情况效果不佳。对于 nature 序列的出视野情况,本文算法也会产生模型漂移。

佳尺度。本文通过实验证明了本文方法的可行性,使用模板更新能够有效应对目标外观变化以提高跟踪精度,同时避免了逐帧计算导致的计算复杂度增加,保证了跟踪器的实时运行。另外,本文的信息熵评价标准还可以轻松扩充到其他的跟踪器,且不仅用于评价响应图质量,还可用于为特征通道特征赋予权重等。但是本文方法难以对目标精确跟踪,今后将尝试结合目标框回归的方法,使跟踪精度得到进一步的提高。

参 考 文 献

- [1] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010:2544-2550.
- [2] 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.
- [3] WANG M, LIU Y, HUANG Z. Large Margin Object Tracking with Circulant Feature Maps[C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017:4021-4029.
- [4] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015:3074-3082.
- [5] LI Z M, FU H J, LIU Y J, et al. Multi-Template Correlation Filter Tracking Based on Deep Feature[C]//conference of Computer-Aided Design & Computer Graphics. 2018.
- [6] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:4293-4302.
- [7] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C]//European Conference on Computer Vision. Springer, Cham, 2016:850-865.
- [8] GUO Q, FENG W, ZHOU C, et al. Learning dynamic siamese network for visual object tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017:1763-1771.
- [9] HE A, LUO C, TIAN X, et al. A twofold siamese network for real-time object tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:4834-4843.
- [10] HINTON G E. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair[C]//International Conference on International Conference on Machine Learning. Omnipress, 2010.
- [11] SHANNON C E. A mathematical theory of communication[J]. Bell Labs Technical Journal, 1948, 27(4):379-423.
- [12] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018:101-117.
- [13] LI B, YAN J, WU W, et al. High performance visual tracking with siamese region proposal network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:8971-8980.
- [14] KRISTAN M, LEONARDIS A, MATAS J, et al. The Visual Object Tracking VOT2016 Challenge Results[C]//IEEE International Conference on Computer Vision Workshops. IEEE Computer Society, 2016.
- [15] KRISTAN M, LEONARDIS A, MATAS J, et al. The Visual Object Tracking VOT2017 Challenge Results[C]//2017 IEEE International Conference on Computer Vision Workshop (ICCVW). IEEE Computer Society, 2017.
- [16] VOJIR T, NOSKOVA J, MATAS J. Robust scale-adaptive mean-shift for tracking[J]. Pattern Recognition Letters, 2014, 49:250-258.
- [17] DE ATH G, EVERSON R. Part-Based Tracking by Sampling [J]. arXiv:1805.08511, 2018.
- [18] POSSEGGGER H, MAUTHNER T, BISCHOF H. In defense of color-based model-free tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:2113-2120.
- [19] ABDELPAKEY M H, SHEHATA M S, MOHAMED M M. Denssiam: End-to-end densely-siamese network with self-attention model for object tracking[C]//International Symposium on Visual Computing. Springer, Cham, 2018:463-473.
- [20] DANELLJAN M, HÄGER G, KHAN F, et al. Accurate scale estimation for robust visual tracking[C]//British Machine Vision Conference, Nottingham. BMVA Press, 2014:1-5.
- [21] WANG Q, GAO J, XING J, et al. Defnet: Discriminant correlation filters network for visual tracking[J]. arXiv:1704.04057, 2017.
- [22] ZHANG J, MA S, SCLAROFF S. MEEM: robust tracking via multiple experts using entropy minimization [C]//European Conference on Computer Vision. Springer, Cham, 2014:188-203.



ZHAO Qin-yan, born in 1994, postgraduate, is a member of China Computer Federation. His main research interests include computer vision, picture processing and machine learning.



LI Zong-min, born in 1965, professor, is a member of China Computer Federation. His main research interests include computer graphics, picture processing and visualization in scientific computing.