



基于异步相关判别性学习的孪生网络目标跟踪算法

许龙 魏颖 商圣行 张皓云 边杰 徐楚翹

Design of Asynchronous Correlation Discriminant Single Object Tracker Based on Siamese Network

XU Long, WEI Ying, SHANG Sheng-Xing, ZHANG Hao-Yun, BIAN Jie, XU Chu-Qiao

在线阅读 View online: <https://doi.org/10.16383/j.aas.c200237>

您可能感兴趣的其他文章

面向边缘计算应用的宽度孪生网络

Broad Siamese Network for Edge Computing Applications

自动化学报. 2020, 46(10): 2060–2071 <https://doi.org/10.16383/j.aas.c200555>

基于自更新混合分类模型的肌电运动识别方法

An EMG-motion Recognition Method With Self-update Hybrid Classification Model

自动化学报. 2019, 45(8): 1464–1474 <https://doi.org/10.16383/j.aas.2018.c170301>

基于异步IMM融合滤波的网络化系统故障诊断

Fault Diagnosis for Networked Systems By Asynchronous IMM Fusion Filtering

自动化学报. 2017, 43(8): 1329–1338 <https://doi.org/10.16383/j.aas.2017.c160768>

异步时钟下基于信息物理融合的水下潜器协同定位算法

Cyber-Physical Cooperative Localization Algorithms for Underwater Vehicle With Asynchronous Time Clock

自动化学报. 2019, 45(4): 739–748 <https://doi.org/10.16383/j.aas.c180377>

基于相关滤波器的视频跟踪方法研究进展

Research Progress of Visual Tracking Methods Based on Correlation Filter

自动化学报. 2019, 45(2): 265–275 <https://doi.org/10.16383/j.aas.2018.c170394>

一种基于语义关系图的词语语义相关度计算模型

A Model for Calculating Semantic Relatedness of Words Considering Semantic Relationship Graph

自动化学报. 2018, 44(1): 87–98 <https://doi.org/10.16383/j.aas.2018.c170002>

基于异步相关判别性学习的孪生网络目标跟踪算法

许龙¹ 魏颖¹ 商圣行¹ 张皓云¹ 边杰¹ 徐楚翘¹

摘要 现有基于孪生网络的单目标跟踪算法能够实现很高的跟踪精度,但是这些跟踪器不具备在线更新的能力,而且其在跟踪时很依赖目标的语义信息,这导致基于孪生网络的单目标跟踪算法在面对具有相似语义信息的干扰物时会跟踪失败.为了解决这个问题,提出了一种异步相关响应的计算模型,并提出一种高效利用不同帧间目标语义信息的方法.在此基础上,提出了一种新的具有判别性的跟踪算法.同时为了解决判别模型使用一阶优化算法收敛慢的问题,使用近似二阶优化的方法更新判别模型.为验证所提算法的有效性,分别在 Got-10k、TC128、OTB 和 VOT2018 数据集上做了对比实验,实验结果表明,该方法可以明显地改进基准算法的性能.

关键词 孪生网络, 语义信息, 异步相关, 判别性, 在线更新

引用格式 许龙, 魏颖, 商圣行, 张皓云, 边杰, 徐楚翘. 基于异步相关判别性学习的孪生网络目标跟踪算法. 自动化学报, 2023, 49(2): 366–382

DOI 10.16383/j.aas.c200237

Design of Asynchronous Correlation Discriminant Single Object Tracker Based on Siamese Network

XU Long¹ WEI Ying¹ SHANG Sheng-Xing¹ ZHANG Hao-Yun¹ BIAN Jie¹ XU Chu-Qiao¹

Abstract The existing single target object tracking algorithms based on the siamese network can achieve very high tracking performance, but these trackers can not update online, and they heavily rely on the semantic information of the target in tracking. It caused the trackers, which based on the siamese network, fail when facing the disruptor who has similar semantic information. To address this issue, this paper proposes an asynchronous correlation response calculation model and an efficient method of using the target's semantic information in different frames. Based on this, a new discriminative siamese network-based tracker is proposed. To address the convergence speed issue in the traditional first-order optimization algorithm, an approximate second-order optimization method is introduced to update the discriminant model online. To evaluate the effectiveness of the proposed method, comparison experiments on Got-10k, TC128, OTB, and VOT2018 between the proposed tracker and other latest state-of-the-art trackers are adopted. The experimental results demonstrate that the proposed method can significantly improve the performance of the baseline.

Key words Siamese network, semantic information, asynchronous correlation, discriminative, update online

Citation Xu Long, Wei Ying, Shang Sheng-Xing, Zhang Hao-Yun, Bian Jie, Xu Chu-Qiao. Design of asynchronous correlation discriminant single object tracker based on siamese network. *Acta Automatica Sinica*, 2023, 49(2): 366–382

视觉目标跟踪算法广泛应用于水下机器人, 无人机协同, 机器人设计等诸多领域^[1–3], 得到一个跟踪精度高, 速度快的跟踪器面临着各种各样的挑战. 近些年来, 大量的研究成果涌现出来, 极大地推进了该领域的发展, 其中又以基于孪生网络的跟踪算

法性能最好^[4–8].

目标跟踪的核心问题是目标与背景的分类问题, 在第 1 帧中给定任意要跟踪的目标, 目标跟踪算法都能在接下来的帧中给出该目标的准确位置. 近年来, 以基于粒子滤波加分类为代表的多域卷积神经网络 (Multi-domain convolutional neural network, MDNet)^[9] 和基于相关滤波为代表的核化相关滤波 (Kernel correlation filter, KCF)^[10] 等判别类跟踪方法受到了研究人员的广泛重视, 以这两类跟踪框架为基础, 又衍生出了大量的跟踪算法^[11–15]. Wang 等^[13] 首先将编码器特征引入到目标跟踪的任务中, 取得了不错的跟踪性能, 但是由于该方法得

收稿日期 2020-04-21 录用日期 2020-09-07

Manuscript received April 21, 2020; accepted September 7, 2020

国家自然科学基金 (61871106) 资助

Supported by National Natural Science Foundation of China (61871106)

本文责任编辑 赖剑煌

Recommended by Associate Editor LAI Jian-Huang

1. 东北大学信息科学与工程学院 沈阳 110819

1. College of Information Science and Engineering, Northeastern University, Shenyang 110819

到的特征比较简单, 该模型很难应付目标发生较大变化时的情景. 为解决这个问题, Ma 等^[12] 充分利用卷积神经网络不同层之间的卷积特征进行跟踪, 利用更高效的图像特征提高了跟踪的性能. 进一步地, Nam 等^[9] 提出了一种多域学习的算法 MDNet, 并引入了一个更大的卷积神经网络用于提取目标特征, 在当时的基准上取得了最好的性能. 但是, 由于 MDNet 基于粒子滤波算法, 速度较慢, 因此越来越多的研究转移到了相关滤波的框架下. Bolme 等^[16] 首先将相关滤波器引入到了目标跟踪的任务中, 将目标跟踪由分类任务变成了相似度计算的任务, 通过求解最小二乘问题得出可能是目标位置的最高响应. 为了解决 Bolme 等^[16] 的算法在跟踪过程中训练样本过少的问题, Henriques 等^[10] 将循环矩阵的思想引入到训练样本的生成过程中, 同时充分利用了循环矩阵的特点, 从而能以很快的速度优化一个非闭合的二次优化问题. 出于计算速度的考虑, 相关滤波的相关性计算都在频域中进行^[16], 在实际的应用中会遇到截断误差 (边界效应) 的问题, 为了解决这一问题, Danelljan 等^[11] 在求解滤波器的过程中引入了一个空间正则化项, 并使用 Gauss-Siedel^[11] 求解滤波器, 同时将原始的图像特征^[8, 14] 替换为对应图像的深度特征, 进行相关滤波器的计算, 进一步提高了跟踪的性能. 为了解决滤波器在离散空间计算导致的跟踪精度损失问题, Danelljan 等^[14] 使用插值的方法, 将滤波器的计算转移到了连续域空间, 并取得了很好的跟踪效果, 但是由于需要计算的滤波器数量比较多, 这导致算法速度较慢. 为了解决这一问题, Danelljan 等^[14] 提出了一种降维的方法, 求解出对响应贡献最大的几组滤波器进行跟踪, 同时利用高斯混合模型进行样本空间的管理, 实现了速度与精度的提升.

随着相关滤波算法的发展和完善, 同为相似度计算的孪生网络模型进入了人们的视野. Held 等^[4] 将孪生网络的结构引入到了目标跟踪的相似度计算中. Bertinetto 等^[5] 进一步设计了训练数据集的结构, 以此为基础衍生出了大量的基于孪生网络的跟踪算法. Li 等^[6] 将检测中的 RPN (Region proposal network)^[17] 结构引入到了跟踪中, 利用边框回归以及锚点的方法, 缓解了边界框结果精度低的问题. 为进一步解决边界框定位精度的问题, Danelljan 等^[18] 将目标检测算法中的 IOU-Net (Intersection over union-network)^[19] 边框回归组件引入到了跟踪中, 提出了一个非孪生网络结构的跟踪器, 相比于 RPN 结构, IOU-Net 有着更高的边界框回归的精度, 这帮助该算法获得了更高的跟踪性能, 同时由于该

算法可以在线更新, 因此其具有更高的判别性能. Zhu 等^[7] 和 Li 等^[8] 提出了更深网络结构的 Siam-RPN++^[8]、DaSiamRPN^[7] 和将分割引入到孪生网络跟踪器的 SiamMask^[20], 进一步提高了跟踪器的精度. 但是这些基于孪生网络的跟踪算法只通过离线训练获得一组网络参数, 其相似度的计算很大程度上依赖于目标的语义信息, 这导致其在在线跟踪时很难处理具有相同语义干扰物的识别问题.

为解决基于孪生网络的跟踪器对于相似目标判别能力弱的问题, 本文提出了一种异步相关的理论模型, 并基于此提出了一种新的具有判别性的跟踪方法. 在进行在线更新的过程中, 本文使用了二阶优化的方法对所提出的模型进行更新, 相比于传统的一阶优化, 本文使用的方法可以在更少的迭代次数下实现更快的收敛. 为了验证本文提出算法的有效性, 本文在 Got-10k^[21]、TC128^[22]、OTB^[23] 和 VOT2018^[24] 上分别进行了对比实验, 实验结果表明本文所提出的方法可以有效地提升在线跟踪器的判别能力以及鲁棒性, 同时还能保证较高的跟踪速度.

1 问题描述

在目标跟踪中, 跟踪任意目标的任务可以看作是相似性学习的问题. 基于孪生网络的跟踪器利用孪生网络可以学习相似性的特点实现跟踪. 假设要学习的相似性函数由孪生网络 $f(x, z|\theta)$ 表示, 其中 x 表示搜索区域, z 表示样本图像, θ 表示该网络的参数. 经过 $f(x, z)$ 的计算, 得到一个目标样本 z 在搜索区域 x 中不同位置的打分, 当前位置得分越高, 说明当前的目标越有可能处于这个位置. 在实际应用中, x 和 z 并不直接参与相似性的计算, 而是先经过一个特征提取器 ϕ 得到相应的特征 $\phi(x)$ 和 $\phi(z)$, 然后将这些特征送入相似度计算函数 g 来进行打分, 得到相应的相似性度量函数如式 (1) 所示:

$$f(x, z|\theta) = g(\phi(x), \phi(z)) \quad (1)$$

式中, ϕ 是一个基于卷积的特征提取器. Li 等^[6] 通过提升 ϕ 的结构, 并引入检测中的边框回归提升了性能. Zhu 等^[7] 和 Li 等^[8] 通过将 ϕ 做的更深更大, 进一步提升了跟踪器的性能. 相似性函数 g 是一个简单的距离或者相似性度量函数, 在基于孪生网络的跟踪器中, g 的形式与高效卷积算子 (Efficient convolution operators, ECO)^[22] 相似, 都是使用相关滤波的操作进行相似度计算.

式 (1) 得到的孪生网络 f 通过在通用数据集上构造特定形式的数据集进行离线训练, 在训练

的过程中, 利用判别的方法对 f 进行训练. 设定正样本的概率为 $1/(1 + e^{-v})$, 负样本的概率为 $1 - 1/(1 + e^{-v})$, 将其代入交叉熵的计算公式, 得到相应的损失函数定义为:

$$l(y, v) = \ln(1 + \exp(-yv)) \quad (2)$$

式中, $y \in [-1, 1]$, 表示是否是目标, v 表示相似性计算后的实际得分. 在训练过程中, 利用所有候选位置的平均损失来表示最终的训练损失函数:

$$L(y, v) = \frac{1}{\mathcal{D}} \sum_{u \in \mathcal{D}} l(y[u], v[u]) \quad (3)$$

式中, \mathcal{D} 表示最后得到的相关性计算得分图, u 表示 \mathcal{D} 中的所有位置. 通过随机梯度下降的方法最小化式 (4) 所示的损失函数, 从而得到网络的参数 θ :

$$\arg \min_{\theta} = \mathbb{E}_{(z, x, y)} L(y, f(z, x|\theta)) \quad (4)$$

式中, 训练样本对 (z, x) 从精心标注好的视频数据集采样, 搜索区域 x 以目标区域 z 为中心, 保持目标宽高比例不变, 裁剪固定大小的区域, 超出区域的部分用图像像素平均值填充. 所有操作都是在离线状态下进行, 即此时跟踪器不进行跟踪, 只进行训练.

2 判别性孪生网络跟踪器

为了给目标跟踪器引入判别性, 本文以 SiamRPN^[6] 为基准算法做了大量的实验工作, 发现了孪生网络跟踪器的一些异步相关特性可以对相同语义特征的目标进行有效的抑制, 同时可以缓解跟踪过程中目标发生形变的问题.

2.1 置信度时间衰减假设

在传统跟踪器工作的过程中, 在第 1 帧时, 跟踪器利用第 1 帧目标的样本 z_0 计算得到当前要跟踪目标的滤波器 k_0 , 然后在接下来的第 t 帧, 利用 x_t 与 k_0 , 通过互相关计算, 得到目标的响应图 \mathcal{D} , 相应的计算过程如式 (5) 所示:

$$f(x_t|\theta, k_0) = g(\phi(x_t), k_0) \quad (5)$$

在整个跟踪的过程中, k_0 保持不变, 这意味着如果目标的状态发生了很大的变化 (包括形状、大小和颜色等), 跟踪器的识别能力就会降低.

假设被跟踪目标随着时间 t 的变化而变化, 当前滤波器 k_0 与时刻 t 下的 x_{mt} 计算得到的得分会发生衰减, 设这个衰减因子为 η , 时刻 t 后, 得分强度会变为 $v_t(u) = \eta \cdot v_0(u)$, $u \in \mathcal{D}$. 同时, 由于在第 1 节中提到的孪生网络 $f(z, x|\theta)$ 并不是为当前跟踪的类别特定训练的, 而是对尽可能多类的目标进行训

练, 从而实现一定的通用性. 因此假设搜索区域中存在同类目标 z_{fake} (例如人类、汽车、同类的动物和物体等). 经过卷积以后, 在 $t = 0$ 时刻, 跟踪目标与同类目标得分应满足:

$$v_0(u) > v_0(u_{fake}) \quad (6)$$

然而, 因为有衰减因子 η 的存在, 会在某一时刻 t 有 $v_t(u) \leq v_t(u_{fake})$, 此时, 跟踪器会错误地判断目标的位置, 将 u_{fake} 当做是真正的目标, 从而导致跟踪失败. 图 1 为分别采用初始的 k_0 与当前的 k_t 所得到的目标响应得分图.

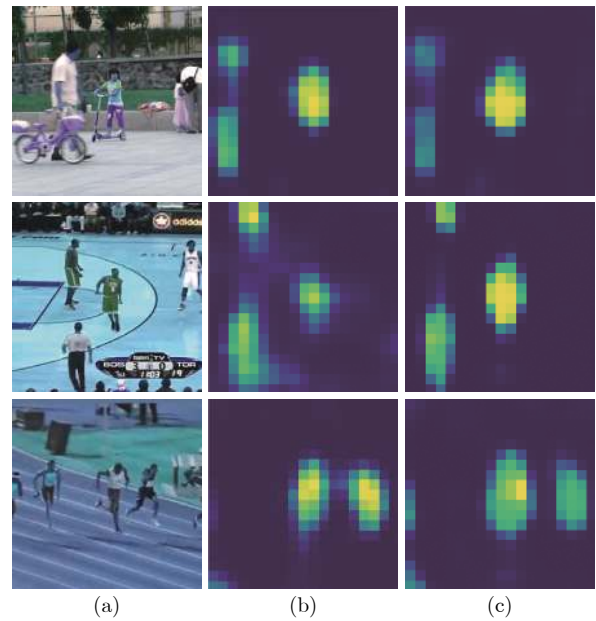


图 1 不同滤波器下响应结果对比

Fig.1 Comparison of response results under different filters

图 1(b) 和图 1(c) 分别表示滤波器 k_0 与滤波器 $k_t = \phi(z_t)$ 计算得到的响应得分图. 由图 1 可以看出, 本文提出的置信度时间衰减假设是合理的, 因此可以利用 k_t 所具备的目标判别能力对基准算法 SiamRPN 进行改进.

2.2 判别性模型的建立

基于以上假设, 本文提出了一种异步相关的打分策略, 从而为跟踪器引入了对于真实目标与虚假同类目标的可判别性.

由于目标的形变会导致 k_0 以 η 的衰减速率计算真实的目标得分, 为此本文直接将 t 时刻的样本图像 z_t 引入到跟踪的过程中, 得到在线打分公式如下:

$$f(x_t, z_t|\theta) = g(\phi(x_t), \phi(z_t)) \quad (7)$$

式 (7) 虽然引入了当前时刻 t 下的样本图像 z_t , 在一定程度上消除了目标形变带来的打分衰减效应, 但是由于跟踪器在跟踪过程中存在着一定的噪声以及跟踪结果上的偏差 (例如跟踪位置的错误, 跟踪得到的目标尺寸的误差), 如果直接采用式 (7) 的打分方式, 这些误差会随着时间 t 逐渐累积, k_t 会被污染, 从而导致跟踪器无法找到目标。

为解决上述问题, 本文将第 1 帧计算得到的 k_0 作为监督信息, 引入一个可在线学习的判别性模型 Φ , 得到新的打分公式:

$$f(x_t, z_t | \theta, w) = g(\phi(x_t), \Phi(\phi(z_t))) \quad (8)$$

式中, θ 是离线训练好的一组参数, 在跟踪的过程中不发生变化, w 是需要在线更新的参数. 令 $k_t = \phi(z_t)$, 代入式 (8) 有:

$$f(x_t, k_t | w) = (a(w * k_t)) * \phi(x_t) \quad (9)$$

式中, $\phi(\cdot)$ 表示离线训练好的特征提取器, a 表示激活函数, w 表示在线判别模型 Φ 的权重, $*$ 表示卷积操作. 为得到一个较为合理的映射, 令 $\hat{k}_t = \Phi(k_t) = a(w * k_t)$, 构建用于在线更新的相似度计算损失函数如式 (10) 所示:

$$L_{up}(w) = \sum_{j=t-n}^t \gamma_j (\hat{k}_j - k_0)^2 + \lambda \|w\|_2 \quad (10)$$

式中, $\|\cdot\|$ 表示 L_2 范数, 对神经网络权重 w 进行正则化, λ 是正则化系数, n 表示样本的个数, $\gamma_j = 1/C\alpha^{t-j}$, $C = \sum_{j=t-n}^t \alpha^{t-j}$ 表示第 j 个样本的权重, $0 < \alpha < 1$ 表示样本权重的衰减参数。

由式 (10) 可知, 在实际跟踪的过程中, 本文算法不仅利用了 t 时刻的样本图像 x_t , 同时还能保证由 x_t 得到的滤波器 k_t 不会偏离真实的滤波器 k_0 太远. 由于滤波器 $\Phi(\phi(z_t))$ 与 k_0 存在较小的差距, 同时 $\Phi(\phi(z_t))$ 还包含了时刻 t 目标的信息, 因此这样的滤波器在面对相同类别的目标时, 具有更高的判别性。

2.3 异步相关响应计算

除了利用当前时刻 t 的 k_t 计算目标的响应外, 同时还可以充分利用时刻 t 之前特定的 m 个判别性网络 $\Phi_{\{1, \dots, m\}} \in \mathcal{S}$ 来计算最终的得分响应:

$$s_t = g(\phi(x_t), \frac{1}{m} \sum_{i=1}^m \Phi_i(\phi(z_t))) \quad (11)$$

式中, s_t 表示当前时刻 t 的得分响应图. Φ_i 表示不同时刻得到的判别性模型。

在实际的应用中, 假设相邻帧之间的判别性模

型更为相似, 而距离较远的帧之间的判别性模型差异更大. 这样的假设与实际情况相符合, 因为随着时间 t 的增加, 不同帧下相同目标之间的差异会变得更大. 根据这个假设, 本文提出了一种利用不同的 Φ_i 网络参数 w_i 之间的 Kullback-Leibler (KL) 散度对 Φ_i 管理的策略, 计算不同网络参数分布之间的 KL 散度矩阵 D_{KL} , 如下式所示:

$$D_{KL}^{(i,j)}(P(w_i) | Q(w_j)) = \sum_{t=1}^N P(x_t) \ln \left(\frac{P(x_t)}{Q(x_t)} \right) \quad (12)$$

式中, $P(w_i)$ 、 $Q(w_j) \in \mathbf{R}^{1 \times N}$ 表示 w_i 、 w_j 相应的概率分布 (直方图分布), (i, j) 表示 D_{KL} 中第 i 行, 第 j 列的元素。

对于每一帧得到的新的 Φ_t , 计算其与 m 个现有网络的 KL 散度向量 d_t , 找到最小 KL 散度距离所对应的 Φ_k , $k \in \{1, 2, \dots, m\}$ 和距离 $d_t(i)$. 利用如式 (13) 所示的策略 u , 对 m 个判别性模块进行更新。

$$u = \begin{cases} \mathcal{S}(k) = \Phi_t, & d_t(i) > \min(D_{KL}) \\ \text{none}, & \text{否则} \end{cases} \quad (13)$$

由式 (13) 可知, 本文利用判别模型参数间 KL 散度的不同, 保留差异最大的 m 个判别器作为异步相关响应打分的依据, 这样的做法可以最大程度上保留目标在不同形态时的语义信息, 在遇到具有相似语义目标的时候, 不同帧中目标的历史信息会帮助跟踪器做出很好的判断, 从而增强了跟踪器的判别能力。

2.4 近似二阶优化算法下的在线更新策略设计

传统的一阶优化算法 (如随机梯度下降、自适应矩估计等)^[25] 在优化过程中比二阶的优化方法有更快的速度, 因为二阶优化算法涉及到二阶信息的计算, 这些信息有助于找到一个网络泛化能力最强的解, 但是高昂的计算代价让其在对海量的数据进行训练时处于劣势。

在本文提出的在线更新策略中, 因为要学习的参数和样本少, 因此近似二阶的优化算法更适用于本任务, 可以同时兼顾速度与精度。

在进行二阶优化的过程中, Hessian 矩阵的计算是影响速度的关键因素. 本文简化计算形式, 根据式 (10) 定义残差为: $r_j(w) = \sqrt{\gamma_j}(\hat{k}_j - k_0)$, $r_{n+1} = \sqrt{\lambda}w$. 将 $r_j(w)$ 和 r_{n+1} 拼接起来, 构成 $r(w)$. 此时, 式 (10) 可等价于:

$$L_{up}(w) = \|r(w)\|^2 \quad (14)$$

对式 (14) 进行二阶泰勒展开, 有:

$$\begin{aligned} \hat{L}_{up}(w) &\approx L_{up}(w + \Delta w) = L_{up}(w) + \frac{\partial L_{up}(w)}{\partial w} \Delta w + \\ &\frac{1}{2} \Delta w^T \frac{\partial^2 L_{up}(w)}{\partial w^2} \Delta w + O(\|\Delta w\|^2) = \\ &r^T(w)r(w) + 2\Delta w^T \nabla^T r_w + \Delta w^T H_w \Delta w \end{aligned} \quad (15)$$

式中, ∇^T 表示 $\partial L_{up}/\partial w$, 利用 PyTorch 的自动求导机制求解. H_w 表示 Hessian 矩阵, $O(\|\Delta w\|^2)$ 是一个极小量, 可忽略不计. 优化的目的是寻找一个合适的 Δw , 使 $\hat{L}_{up}(w)$ 最小.

具体优化流程步骤如下:

步骤 1. 初始化网络权重 w 、残差 $r(w)$ 、 N_{CG} 和 N .

步骤 2. for $i = 1, \dots, N$ do.

步骤 3. 计算 w 下 $\nabla L_{up}(w)$, $H_w v = \mathcal{R}_v\{\nabla_w(w)\}$, 代入式 (15), $\Delta w \leftarrow 0$.

步骤 4. for $j = 1, \dots, N_{CG}$, $\Delta w = \Delta w_0$, $d_i = d_0 = -\nabla L_{up}(w)$ do.

步骤 5. 计算步长. 计算 α 来最小化式 (15):

$$\alpha = -\frac{d_j^T (H_w \Delta w + \nabla^T)}{d_j^T H_w d_j}$$

步骤 6. 更新权重. $\Delta w = \Delta w + \alpha d_j$.

步骤 7. 更新方向. Let $d_{j+1} = -\nabla^T + \beta_j d_j$, 其中:

$$\beta_j = \frac{\nabla L_{up}(\Delta w)^T H_w d_j}{d_j^T H_w d_j}$$

步骤 8. end for.

步骤 9. $w = w + \Delta w$.

步骤 10. end for.

不同于基于最大化交并比的精确跟踪算法 (Accurate tracking by overlap maximization, AT-OM)^[18] 利用雅可比矩阵近似计算 Hessian 矩阵的方式, 为避免求解 H_w 的逆矩阵, 本文利用 Pearlmutter 等^[26] 提出的 $\mathcal{R}\{\cdot\}$ 直接求解 $H_w v$, 进一步提高了优化过程中的精度与速度. 令 $\Delta w = rv$, 形式为:

$$H_w v = \lim_{r \rightarrow 0} \frac{\nabla(w + rv) - \nabla(w)}{r} = \frac{\partial}{\partial r} \nabla(w + rv)|_{r=0} \quad (16)$$

3 实验

3.1 实验设置

本文使用基于 Python 的 PyTorch 作为实验平台, CPU 为 I5 8400, 内存为 24 GB, GPU 为 2060 Super. 分别在 Got-10k^[21] 的 180 个视频序列, TC-128^[22] 的 128 个视频序列, OTB 的 100 个视频序列, 以及 VOT2018^[24] 的 60 个视频序列上进行实验. 利

用 m 个判别模块 Φ 进行异步相关性计算, 其中每个判别模块有 256 个卷积核, 每个卷积核的大小为 1×1 . 在优化的过程中, $N = 5$, $\alpha = 0.3$, 共轭梯度下降迭代次数 $N_{CG} = 100$, 实际计算过程中, 由于网络很快就收敛, 因此共轭梯度优化部分只迭代很少的次数. 本文算法在 4 个评测数据集上采用同一套参数进行测试.

在对比实验部分, 对比了本文算法与最新的基于孪生网络的目标跟踪算法, 包括 SiamFC^[5]、SiamRPN++^[8] 和 DaSiamRPN^[7] 等; 并且对比了其他的跟踪算法, 包括基于相关滤波算法的 ECO^[15] 和判别性跟踪器 ATOM^[18] 等, 由于不同跟踪器源码, 评估数据集及其结果公开情况的不同, 因此在第 3.2 ~ 3.4 节的对比分析中, 参与对比的跟踪器会略有不同.

3.2 与基准算法的对比分析

为验证本文算法的有效性, 本文在 Got-10k 和 TC128 上进行了自身对比实验, 包括消融实验和量化分析. 同时与一些最先进或者经典的跟踪器进行对比. 表 1 为基准算法与本文算法在 Got-10k 上的性能对比情况.

表 1 本文方法与基准算法的消融实验
Table 1 Ablation studies between the proposed algorithm and baseline

算法	AO	SR _{0.5}	SR _{0.75}	FPS
Baseline	0.445	0.539	0.208	21.95
Baseline + AC	0.445	0.539	0.211	20.03
Baseline + AC + S	0.447	0.542	0.211	19.63
Baseline + AC + S + $D_{KL}^m = 3$	0.442	0.537	0.209	18.72
Baseline + AC + S + $D_{KL}^m = 6$	0.457	0.553	0.215	18.60
Baseline + AC + S + $D_{KL}^m = 9$	0.440	0.532	0.211	18.49

表 1 中, Baseline 表示基准模式, AC (Asynchronous correlation) 表示本文提出的异步相关模块, D_{KL} 表示判别性模块管理算法, 该算法利用 KL 散度对 m 个异步相关判别模型进行管理. FPS 表示算法每秒可处理的帧数, 值越大表示性能越好. S 表示本文使用的二阶优化算法. AO 表示所有帧上跟踪的结果和真实标签之间重叠率的平均值, 值越大表示性能越好. SR 表示跟踪的成功率, 值越大表示性能越好, SR_{0.75} 下标表示成功率高于 0.75%. 由表 1 可以看出, 本文算法比基准算法在 AO、SR_{0.5} 和 SR_{0.75} 三个指标下分别提升 1.2%、1.4% 和

0.7%. 因此使用本文提出的异步相关判别模型 Φ 可明显改善基准算法在 Got-10k 上的性能.

本文同时对比了 $m = 3, 6, 9$ 时算法的性能表现. 可以看出, 当 $m = 6$ 时, 算法性能最好; 在只使用 AC 的情况下, $SR_{0.75}$ 比基准算法高出 0.3%; 当引入二阶优化算法 S 后, 平均重叠率 AO 和 $SR_{0.5}$ 分别提升 0.2% 和 0.3%; 而当 $m = 3, 9$ 时, 跟踪器的性能出现下降. 在模型的优化方面, 将一阶的 Adam 优化算法替换为本文所使用的二阶优化方法 S 后, AO 提升了 0.2%, 可以看出, 本文使用的二阶优化方法可以有效地提高跟踪的精度, 同时对于跟踪速度的影响较小.

图 2 为本文方法在 $m = 6$ 时与其他先进的跟踪器在 Got-10k 上的对比情况.

由图 2 可以看出, 在平均成功率的性能表现上, 本文方法分别比基准算法 SiamRPN 的改进算法 SiamRPN++ 和 SiamMask 高出 0.4% 和 0.5%. 而 SiamRPN++ 和 SiamMask 分别比基准算法提升了 0.8% 和 0.7%, 相比之下, 本文方法相对于基准算法提升了 1.2%, 可见本文算法在 Got-10k 的评价标准下对于基准算法的提升较大, 同时本文算法在面对最新的一些跟踪器时, 也具有较大的优势. 分析原因发现, Got-10k 中存在着大量相似语义信息的跟踪情景, 这正是本文提出异步相关判别模型所要解决的问题, 通过利用帧间被跟踪样本信息的共享以及初始目标提供的监督信息, 本文方法在一定程度上缓解了相似语义目标所导致的跟踪失败问题, 相比于 SiamRPN++ 和 SiamMask, 本文算法在应对这类情景时有着较为明显的优势.

为评估跟踪器在实际跟踪中的性能表现, 本文在 Got-10k 测试集的 8 个具有代表性序列上进行了

改进前后跟踪结果的对比分析, 对比结果如图 3 所示. 图 3 中虚线框表示本文算法的跟踪结果, 实线框表示基准算法的跟踪结果.

由图 3 可以看出, 在第 1 个目标发生遮挡的情景下, 本文方法通过不同帧间的信息共享, 准确地在遮挡发生时跟踪到了目标; 在第 2 个水下相似背景跟踪过程中, 本文算法通过利用最新的当前样本来生成核, 从而及时地响应了目标的形状变化, 跟踪到了目标; 在第 3 个有相似语义目标的情景下, 本文算法同样利用及时更新的样本核准确地跟踪到了目标; 在第 4 个目标发生尺度变化的跟踪情景中, 通过异步信息的计算以及当前样本核的帮助, 本文算法准确地跟踪到了目标; 在第 5 到第 8 的 4 个序列中, 本文算法在处理相应的复杂环境进行跟踪时同样表现出了较基准算法更强的性能优势.

Liang 等^[22] 指出, 利用不同的颜色模型对提升跟踪器性能具有很大帮助, 不同于 OTB20-15 中包含一些灰度图像的情景, TC128 中的 128 个序列均为彩色序列, 其中 70 个序列为新增的, Liang 等^[22] 认为这些彩色图像在理论上会更能充分评估跟踪器的性能. 为此本文在 TC128 上对跟踪器的精度和成功率进行自身对比实验, 实验中默认使用二阶优化算法 S 来优化 AC 模型, 并选择了跟踪器高效卷积算子、空间正则化的相关滤波算法 (Spatially regularized correlation filters, SRDCF)、多专家跟踪 (Multiple experts using entropy minimization, MEEM)^[27]、Struck^[28]、KCF、稀疏跟踪算法 (AS-LA)^[29]、半监督跟踪算法 (SemiT)^[29] 和整数直方图跟踪 (Frag)^[30] 作为对比, 得到精度和成功率图见图 4.

由图 4 可知, 本文算法在 $m = 3$ 时性能最好. 与基准算法相比, 在加入 AC 与 D_{KL}^m 后, 本文方

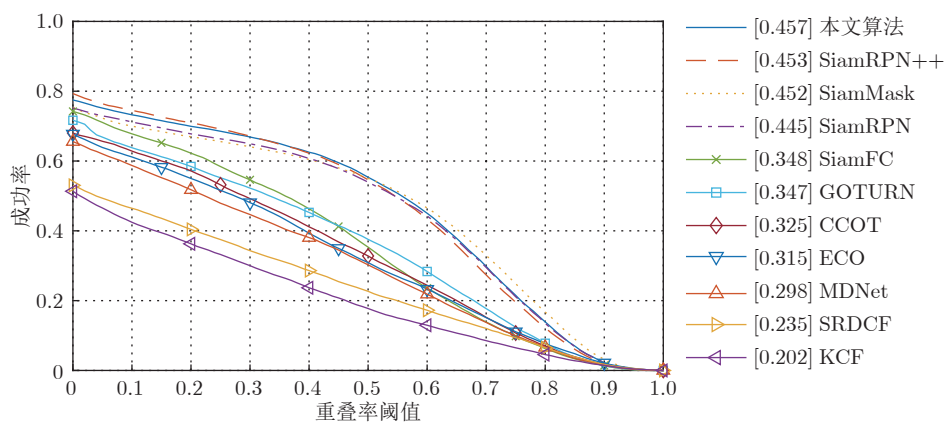


图 2 本文算法与其他先进跟踪器在 Got-10k 上的对比情况

Fig. 2 Comparison between the proposed method with other advanced trackers on Got-10k

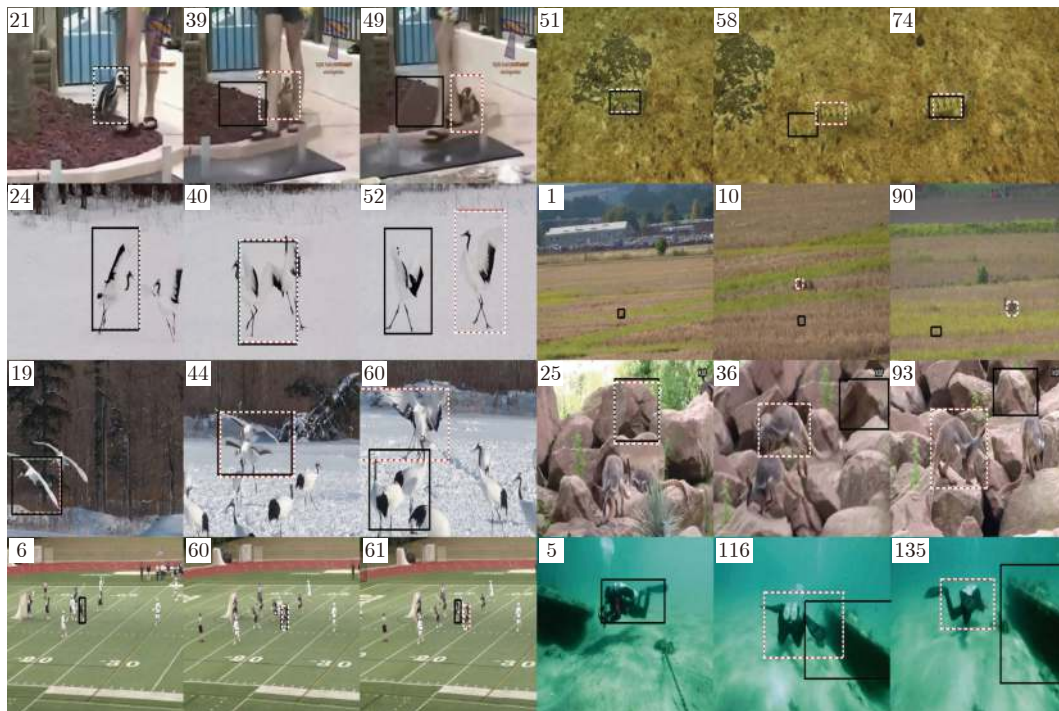


图 3 Got-10k 上跟踪结果对比实验

Fig.3 Comparison of tracking results on Got-10k

法的精度和成功率分别提升 1.6% 和 1.0%。而 SRDCF 的改进算法 ECO-HC 相较于 SRDCF 分别提高 0.8% 和 1.5%，因此本文方法相比于基准算法有较大提升。值得注意的是，在 Got-10k 上第 8 名的 ECO，在 TC128 上是第 1 名，类似现象也发生在 VOT2018 的评估结果中。本文算法比采用多专家模型的 MEEM 在精度和成功率上分别提升 9.5% 和 7.8%。本文算法比基于相关滤波的 KCF 算法在精度和成功率上分别提升了 17.8% 和 15.2%。比其他的跟踪器（如 Struck、ASLA 等）算法，本文算法具有较大的性能优势。同时由图 4 可以看出，当 $m = 6$ 时，本文算法的性能较 $m = 9$ 时在精度和成功率上分别提升 0.1% 和 0.1%，在精度与成功率上均要好于基准算法。

3.3 跟踪精度与成功率

本节进一步分析本文算法在 OTB2015/2013 上的性能表现，包括成功率图、精度图，以及在不同的跟踪情景下，不同跟踪器的性能对比情况。

首先，在 OTB2015 上对比本文算法与最先进的跟踪器，包括 DaSiamRPN、ATOM、DIMP^[31] 和 CF2^[2] 等，以及具有代表性的算法，包括 ECO、MDNet 和 SiamFC^[6] 等，结果如图 5 所示。

由图 5 可以看出，本文方法比基准算法在平均

精确度性能上提升 1.7%，在平均成功率性能上比基准算法提升 1.2%。当 $m = 3$ 和 $m = 6$ 时的精度性能表现几乎相同。当 $m = 6$ 时，本文算法达到相较于基准算法的最好性能。同时可以看出，SiamRPN 的另一个改进算法 DaSiamRPN 较基准算法分别提升 2.5% 和 2.0%，稍高于本文算法对于基准算法的提升。分析原因发现，在 SiamRPN 的基础上，DaSiamRPN 在离线训练阶段通过数据增强的技术来生成用于网络训练的数据样本对，替换了基准算法的训练数据来对 SiamRPN 重新训练，通过扩充训练数据样本对的多样性，来增强原始 SiamRPN 的泛化能力，使之在面对更复杂的跟踪情景或者快速变化的目标时有着更好的跟踪性能，同时 DaSiamRPN 也采用一种利用得分图进行干扰物感知的算法来提升性能。相比之下，本文算法只需要更新一个简单的 AC 模型，且无需修改主干网络参数。

为对本文方法在不同跟踪情景下进行综合评估，选取了 OTB2013^[23] 的 50 个序列中 11 种跟踪情景对本文算法进行评估，同时为对比的公平，本文方法将与基准算法和一些最新的方法（包括 ATOM、DaSiamRPN 和 DIMP 等）做对比实验，实验结果如表 2 ~ 表 4 所示。

表 2 ~ 表 4 中，下划线表示当前跟踪器在所有参与对比的跟踪器中是第 1 名，加粗字体表示当前

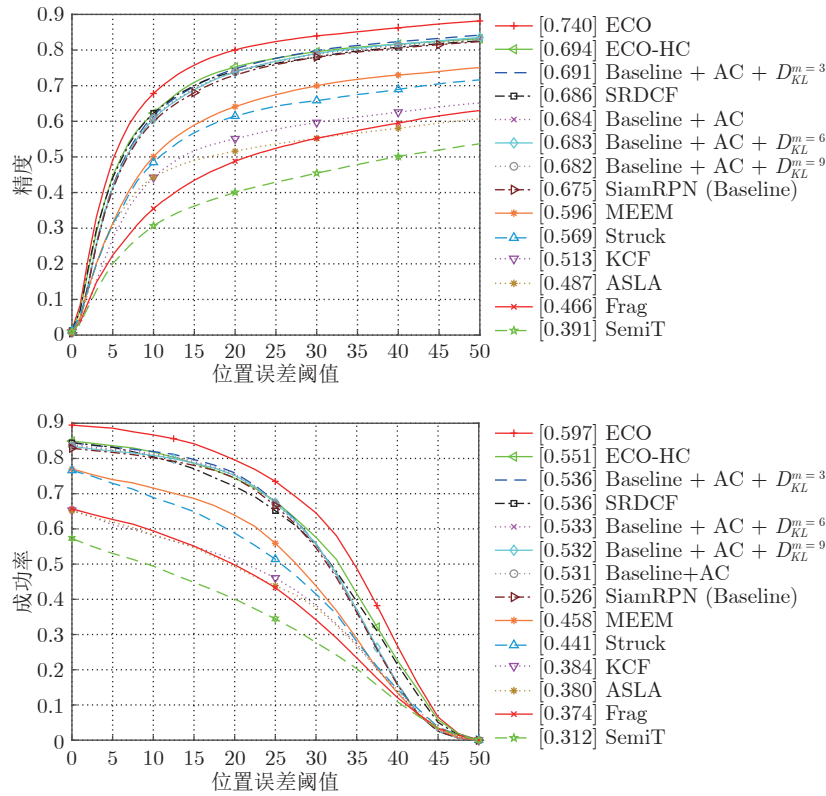


图 4 本文算法在 TC128 上的精度-成功率对比实验结果

Fig. 4 The accuracy-success rate comparison experiment results of the proposed algorithm on TC128

跟踪器在与基准算法对比过程中是第 1 名. 当 $m = 3$ 时, 本文算法在形变、快速运动、平面内旋转、光照变化、低分辨率、运动模糊、遮挡、平面外旋转、视野外、尺度变化共 10 种跟踪情景下的精度和成功率上, 分别比基准算法提升 (1.2%, 2.5%)、(5.6%, 7.3%)、(2.4%, 3.8%)、(1.5%, 2.6%)、(3.5%, 4.4%)、(7.8%, 9.8%)、(0.7%, 1.4%)、(1.3%, 2.4%)、(4.4%, 5.3%) 和 (4.8%, 6.8%). 而在背景干扰的跟踪情景下, 本文算法的性能与基准算法相近. 从以上分析可以看出, 本文算法较为全面地提升了基准算法性能.

分析算法性能提升原因可以发现, 在性能提升较大的 10 种跟踪情景下, 本文方法在面对这些挑战时, 相比基准算法有更强的鲁棒性. 这 10 种情景大多对应的是目标在被跟踪的过程中其外观所发生的几类变化, 而本文正是通过在线更新参与计算样本核的多个 AC 模块, 来部分地解决目标在跟踪过程中发生形变, 导致语义信息发生较大偏差的问题. 通过利用具有判别性的异步相关策略, 本文方法可通过实时更新的方式, 将这些扰动对性能的影响降到最低.

为进一步验证本文所提异步相关响应模型带给

跟踪器的判别性, 在 OTB2015 中选择了具有代表性的 6 个序列进行对比实验, 实验结果如图 6 所示. 其中初始样本表示第 1 帧目标计算得到的 k_0 , 当前样本表示当前帧目标计算得到的 k_t , 优化后样本表示对当前 k_t 优化的结果.

由图 6 可以看出, 对有相同语义信息的目标, 当前样本计算得到的结果较初始样本结果有更准确响应, 这是因为在加入异步相关响应计算后, 由于不同帧间的目标信息被充分利用, 使其对于具有相同语义信息的目标有明显抑制作用. 同时可以看出, 由于利用目标语义信息对其相应位置进行打分, 因此目标外观在跟踪过程中发生多次变化后, 如图 1 和图 6 所示, 采用当前样本计算得到的响应结果比在第 1 帧计算结果有更强响应, 同时在进行多帧 AC 平均后, 得到的响应更为集中, 在应对相似语义背景以及目标形变上更具鲁棒性.

3.4 跟踪器鲁棒性能对比分析

为进一步评估本文算法的性能与最先进算法在精确度和鲁棒性上的对比情况, 在 VOT2018 上对本文算法做了相应的对比实验, Li 等^[8]指出 SiamRPN++ 在 VOT2018 上的性能要好于 DaSiam-

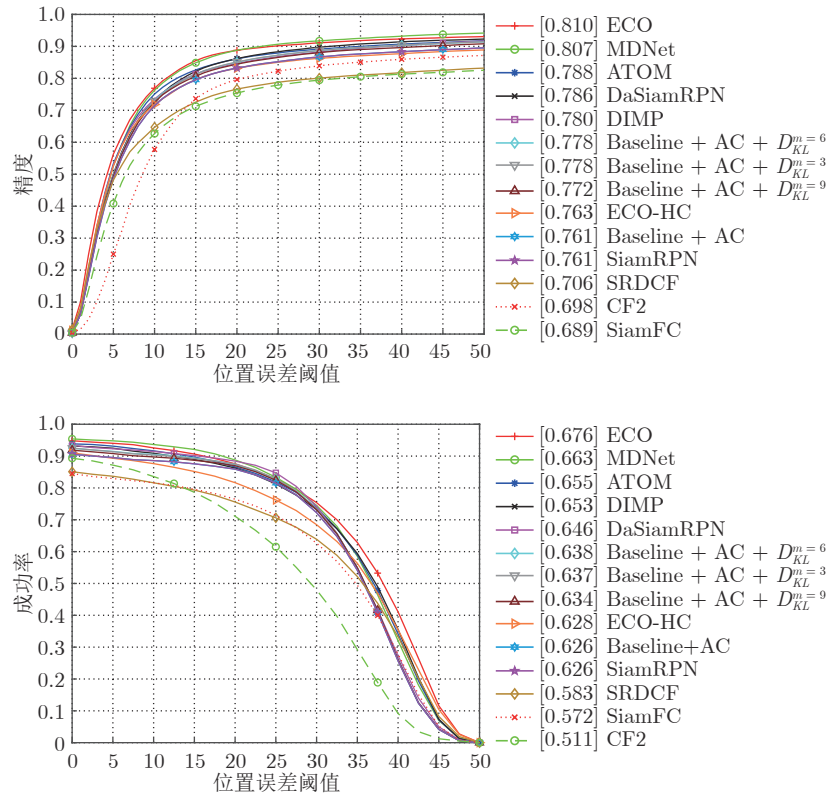


图 5 本文算法在 OTB2015 上的精度-成功率对比实验结果

Fig. 5 The accuracy-success rate comparison experiment results of the proposed algorithm on OTB2015

表 2 OTB2013 上的背景干扰、形变等情景下的跟踪性能对比

Table 2 Tracking performance comparisons among trackers on OTB2013 in terms of background clusters and deformation

算法	背景干扰		形变		快速运动		平面内转动	
	成功率	精度	成功率	精度	成功率	精度	成功	精度
ECO-HC	0.700	0.559	0.567	0.719	0.570	0.697	0.517	0.648
ECO	0.776	0.619	0.613	0.772	<u>0.655</u>	0.783	0.630	0.764
ATOM	0.733	0.598	<u>0.623</u>	0.771	0.595	0.709	0.579	0.714
DIMP	0.749	0.607	0.602	0.740	0.618	0.739	0.561	0.685
MDNet	<u>0.777</u>	0.621	0.620	<u>0.780</u>	0.652	0.796	<u>0.658</u>	<u>0.822</u>
SiamFC	0.605	0.494	0.487	0.608	0.509	0.618	0.483	0.583
DaSiamRPN	0.728	0.592	0.609	0.761	0.565	0.702	0.625	0.780
SiamRPN (Baseline)	0.605	<u>0.745</u>	0.591	0.724	0.589	0.724	0.627	0.770
Baseline + AC	0.605	0.745	0.591	0.724	0.589	0.724	0.627	0.770
Baseline + AC + $D_{KL}^m = 3$	0.599	0.741	0.603	0.749	0.645	0.797	0.651	0.808
Baseline + AC + $D_{KL}^m = 6$	0.592	0.733	0.597	0.742	0.636	0.787	0.650	0.807
Baseline + AC + $D_{KL}^m = 9$	0.598	0.736	0.586	0.725	0.587	0.723	0.654	0.809

RPN, 因此为了让更多的跟踪器参与比较, 本文选用 DaSiamRPN 在 VOT2018 上的实验结果与本文

算法进行比较. 表 5 为本文算法与先进跟踪器在 3 种评估框架基准、非监督和实时性能下的性能对比

表 3 OTB2013 上的光照变化、低分辨率等情景下的跟踪性能对比

Table 3 Tracking performance comparisons among trackers on OTB2013 in terms of illumination change and low resolution

算法	光照变化		低分辨率		运动模糊		遮挡	
	成功率	精度	成功率	精度	成功率	精度	成功率	精度
ECO-HC	0.556	0.690	0.536	0.619	0.566	0.685	0.586	0.749
ECO	0.616	0.766	0.569	0.677	0.659	0.786	<u>0.636</u>	<u>0.800</u>
ATOM	0.604	0.749	0.554	0.654	0.529	0.665	0.617	0.762
DIMP	0.606	0.749	0.485	0.571	0.564	0.695	0.610	0.750
MDNet	<u>0.619</u>	<u>0.780</u>	<u>0.644</u>	<u>0.804</u>	<u>0.662</u>	<u>0.813</u>	0.623	0.777
SiamFC	0.479	0.593	0.499	0.600	0.485	0.617	0.512	0.635
DaSiamRPN	0.589	0.736	0.490	0.618	0.533	0.688	0.583	0.726
SiamRPN (Baseline)	0.585	0.723	0.519	0.653	0.532	0.684	0.586	0.726
Baseline + AC	0.585	0.723	0.519	0.653	0.532	0.684	0.586	0.726
Baseline + AC + $D_{KL}^{m=3}$	0.600	0.749	0.554	0.697	0.610	0.785	0.593	0.740
Baseline + AC + $D_{KL}^{m=6}$	0.592	0.741	0.546	0.688	0.596	0.770	0.586	0.732
Baseline + AC + $D_{KL}^{m=9}$	0.581	0.724	0.549	0.689	0.533	0.687	0.576	0.716

表 4 OTB2013 上的平面外旋转、视野外等情景下的跟踪性能对比

Table 4 Tracking performance comparisons among trackers on OTB2013 in terms of out-of-plane rotation and out of view

算法	平面外旋转		视野外		尺度变化	
	成功率	精度	成功率	精度	成功率	精度
ECO-HC	0.563	0.718	0.549	0.763	0.587	0.740
ECO	<u>0.628</u>	<u>0.787</u>	<u>0.733</u>	<u>0.827</u>	0.651	0.793
ATOM	0.607	0.751	0.522	0.563	0.654	0.792
DIMP	0.596	0.737	0.549	0.593	0.636	0.767
MDNet	<u>0.628</u>	<u>0.787</u>	0.698	0.769	<u>0.675</u>	<u>0.842</u>
SiamFC	0.500	0.620	0.574	0.642	0.542	0.665
DaSiamRPN	0.599	0.750	0.570	0.633	0.587	0.740
SiamRPN (Baseline)	0.598	0.736	0.658	0.725	0.608	0.751
Baseline + AC	0.598	0.736	0.658	0.725	0.608	0.751
Baseline + AC + $D_{KL}^{m=3}$	0.611	0.760	0.702	0.778	0.656	0.819
Baseline + AC + $D_{KL}^{m=6}$	0.604	0.752	0.659	0.733	0.631	0.791
Baseline + AC + $D_{KL}^{m=9}$	0.597	0.740	0.660	0.735	0.603	0.755

结果. 由于 ATOM 未提供 VOT2018 下非监督模式和实时模式下的实验结果, 因此与这些指标相关的值均设置为 0.

表 5 中, Baseline 指 VOT2018 中基于复位的

监督实验^[24]. 而非监督原理与 OTB 的评估策略一致, 即被评估的跟踪器仅用第 1 帧给定的目标信息来初始化跟踪器, 然后记录其在后续帧中的跟踪结果, 最后计算平均的跟踪重叠率 (Average overlap,

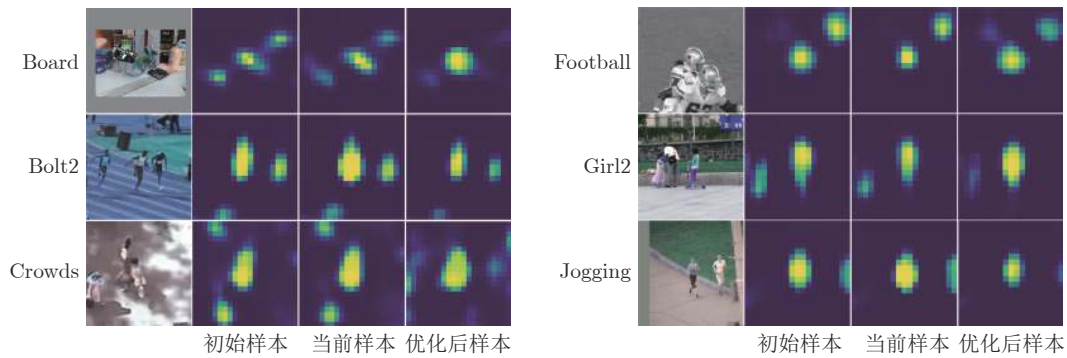


图 6 OTB50 中 6 个序列的响应对比结果

Fig.6 The response comparisons of 6 different sequences on OTB50

表 5 VOT2018 上的实验结果
Table 5 Experimental results on VOT2018

算法	Baseline				非监督	实时性能	
	精度-鲁棒性	失败率	EAO	FPS	AO	FPS	EAO
KCF	0.4441	50.0994	0.1349	<u>60.0053</u>	0.2667	<u>63.9847</u>	0.1336
SRDCF	0.4801	64.1136	0.1189	2.4624	0.2465	2.7379	0.0583
ECO	0.4757	17.6628	0.2804	3.7056	0.4020	4.5321	0.0775
ATOM	<u>0.5853</u>	<u>12.3591</u>	<u>0.4011</u>	5.2061	0	0	0
SiamFC	0.5002	34.0259	0.188	31.889	0.3445	35.2402	0.182
DaSiamRPN	0.5779	17.6608	0.3826	58.854	<u>0.4722</u>	64.4143	<u>0.3826</u>
SiamRPN (Baseline)	0.5746	23.5694	0.2941	14.3760	0.4355	14.4187	0.0559
Baseline + AC	0.5825	27.0794	0.2710	13.7907	0.4431	13.8772	0.0539
Baseline + AC + $D_{KL}^m = 3$	0.5789	14.8312	0.2865	13.6035	0.4537	13.4039	0.0536
Baseline + AC + $D_{KL}^m = 6$	0.5722	22.6765	0.2992	13.5359	0.4430	12.4383	0.0531
Baseline + AC + $D_{KL}^m = 9$	0.5699	22.9148	0.2927	13.5046	0.4539	12.1159	0.0519

AO). VOT-2018 中的实时性能实验部分, 被用来评估跟踪器的实时性能和限定跟踪器的响应时间. 精度-鲁棒性评估的是跟踪器在每一帧中预测结果与实际状态的重叠率和每个序列的平均失败次数. 失败率表示当重叠率低于某一阈值时视为失败时的统计结果. EAO (Expected average overlap) 是对每个跟踪器在一个短时图像序列上未发生重置的平均重叠率期望值, 表示期望平均重叠率, 这个值越大, 表示跟踪器精确度越高.

由表 5 可以看出, 本文算法在 3 种评测方案中均好于基准算法. 在 Baseline 精度-鲁棒性指标下, 本文方法比基准算法提升了 0.79%. 当 $m = 3$, 本文方法在 Baseline 失败率指标下比基准算法降低 8.7382%. 当 $m = 6$ 时, 在 Baseline EAO 指标下, 本文方法比基准算法提升了 0.51%.

在基准算法与 Baseline 的 FPS 指标的对比中

可以发现, 本文方法对于速度的影响很小, 在最坏的情况下, 跟踪器的速度只降低了 0.8714 帧/秒.

在非监督的 AO 指标对比中, 当 $m = 9$ 时, 本文方法比基准算法提升 1.84%. 分析原因发现, 本文方法在跟踪过程中可视为一个弱监督跟踪算法, 其监督信息由第 1 帧指定, 为此相比于基准算法和 SiamFC, 本文方法在非监督评估模式下具有较大优势.

在实时性能对比中, 由于本文算法采用了附加的模块, 同时实时性能的评估也受到了实验平台硬件性能的影响, 因此本文算法在该指标下的性能整体上较基准算法要弱, 最好的情况下 EAO 性能比基准算法降低 0.2%.

图 7 为参与对比的不同跟踪器在 Baseline 下的精度-鲁棒性和跟踪失败率的对比情况.

由图 7 可看出, 当 $m = 3$ 时, 对应的精度-鲁棒

性是最高的,可以看出,本文算法在只采用 AC 的情况下,相应的精度较基准算法有所提高,但是鲁棒性则较基准算法有所下降,分析原因可以发现,这是因为单一的 AC 可能会带来未知的噪声,而在使用了多个 AC 的情况下,由于利用了不同帧间的目标信息,不同 AC 模块之间的噪声可以通过均值滤波的方式进行抑制. 在 AC 内部噪声被抑制的同时,多个异步相关模块的引入也带来对于目标在多

种状态(包括形状、色彩等)的鲁棒性,这使得跟踪器具有了目标在不同时间维度上的信息,在利用这些信息处理后续目标帧时,跟踪器可充分利用前几帧目标信息来计算当前帧目标响应. 通过利用不同时刻目标差异信息来增强跟踪器对目标各种变化的鲁棒性,使算法可以在精度与鲁棒性上都有较好表现.

图 8 为不同跟踪器在 VOT2018, 包含光照变化、相机运动、运动变化、遮挡、尺度变化等,情景

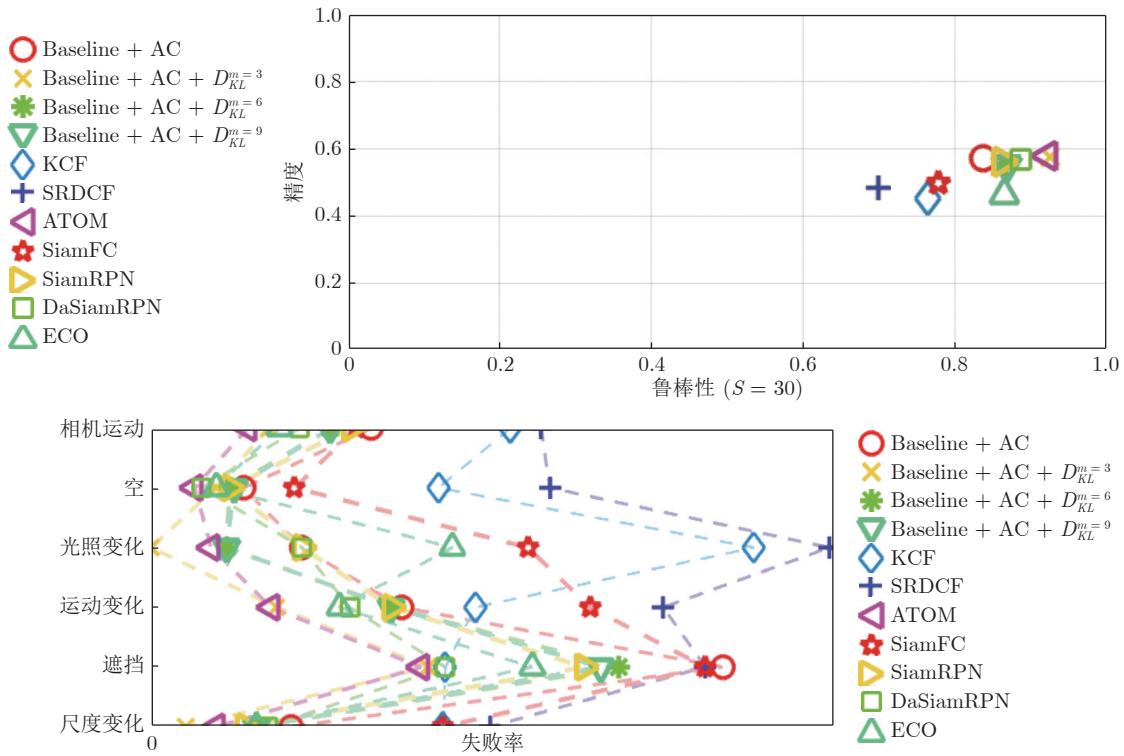


图 7 精度-鲁棒性跟踪失败情况对比图

Fig.7 Comparison of accuracy robustness and tracking failure

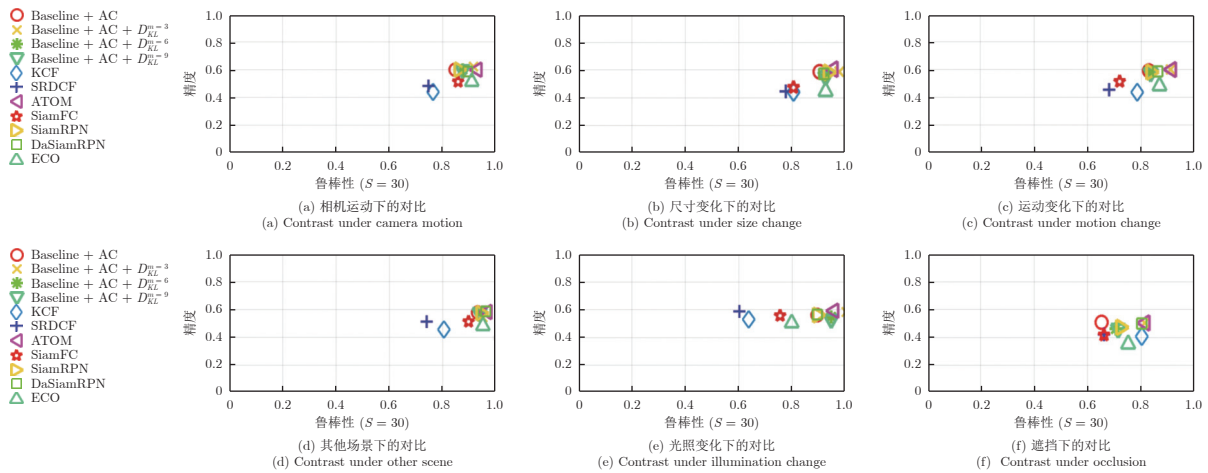


图 8 在 VOT2018 序列的不同情景下精度-鲁棒性对比情况

Fig.8 Comparison of accuracy robustness performance under different attributes on VOT2018

的精度-鲁棒性对比, 这些情景与 OTB2013 类似, 不同的是 OTB 中包含 11 种情景, VOT2018 包含 6 种.

由图 8 可看出, 当 $m = 3$ 时, 跟踪器在相机运动、光照变化、运动变化、遮挡和尺度变化 5 个方面的精度与鲁棒性上要优于基准算法. 在相机运动的情景下, 本文算法的精度达到了与第 1 名的 ATOM 算法相近的性能, 鲁棒性也要好于 ECO, 仅次于 ATOM. 在鲁棒性和精度方面均好于 DaSiamRPN. 在尺度变换、遮挡、运动变化、光照变化 4 个方面的情景下, 本文算法在鲁棒性上达到了第 1 名的成绩. 因此本文算法在不损失算法精确度的情况下, 提升了算法在应对尺度变换时的鲁棒性.

在运动变化的情景下, 本文算法的精确度与第 1 名算法 ATOM 基本一致, 这一实验结果验证了, 目标在发生形变后, 本文算法可以有效地提升跟踪的性能的假设.

在其他跟踪情景下, 本文算法在精确度-鲁棒性上与基准算法基本保持一致, 同时与 DaSiamRPN 以及第 1 名的 ATOM 差距不大. 在光照变化情景下, 本文算法在精度-鲁棒性上高于基准算法和 DaSiamRPN. 原因依然是光照变化导致被跟踪的目标发生了剧烈的外观变化, 而本文算法在解决

这些外观变化方面具备一定优势.

在应对遮挡的情景时, 本文算法在精度-鲁棒性上均与第 1 名的 ATOM 保持一致, 同时在鲁棒性上高于基准算法与 DaSiamRPN, 这一现象的原因可以归结为多个 AC 模块的引入, 利用第 1 帧选定的未被遮挡的目标作为监督, 在线更新当前采样得到的目标样本, 从而使其在兼顾多帧目标和遮挡语义的同时, 也能保持与初始选定的目标相近的语义信息, 保证了跟踪性能.

由图 9 可以看出, 基于相关滤波的 KCF 和 SRDCF 在序列长度增加到 200 帧后, 其对应的期望重叠率性能低于 0.2, 这说明 KCF 和 SRDCF 的跟踪性能对序列的长度更敏感. SiamFC 在序列长度超过 200 帧时性能出现了明显下降, 对应的 EAO 只有 0.2. 本文算法在 $m = 6$ 时, 比基准算法有较大的提升.

本文在 VOT2018 的非监督实验模式下与基准算法以及其他最新的跟踪器进行对比, 得到非监督模式下的期望重叠率对比曲线如图 10 所示. 可以看出, 本文算法在 4 个情景下对基准算法有明显提升. 当 $m = 9$ 时, 本文算法在所有 6 种情况下的性能都优于基准算法. 在光照变化情景下, 好于

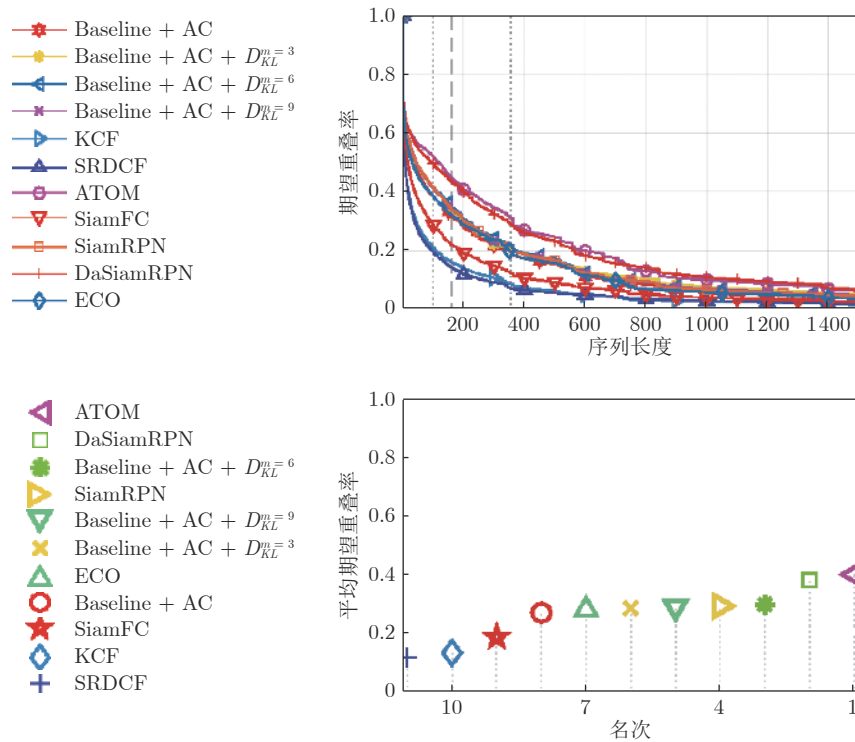


图 9 跟踪器在 VOT2018 基准模式下的期望重叠率性能对比

Fig. 9 Trackers' expected overlap performance comparisons on VOT2018

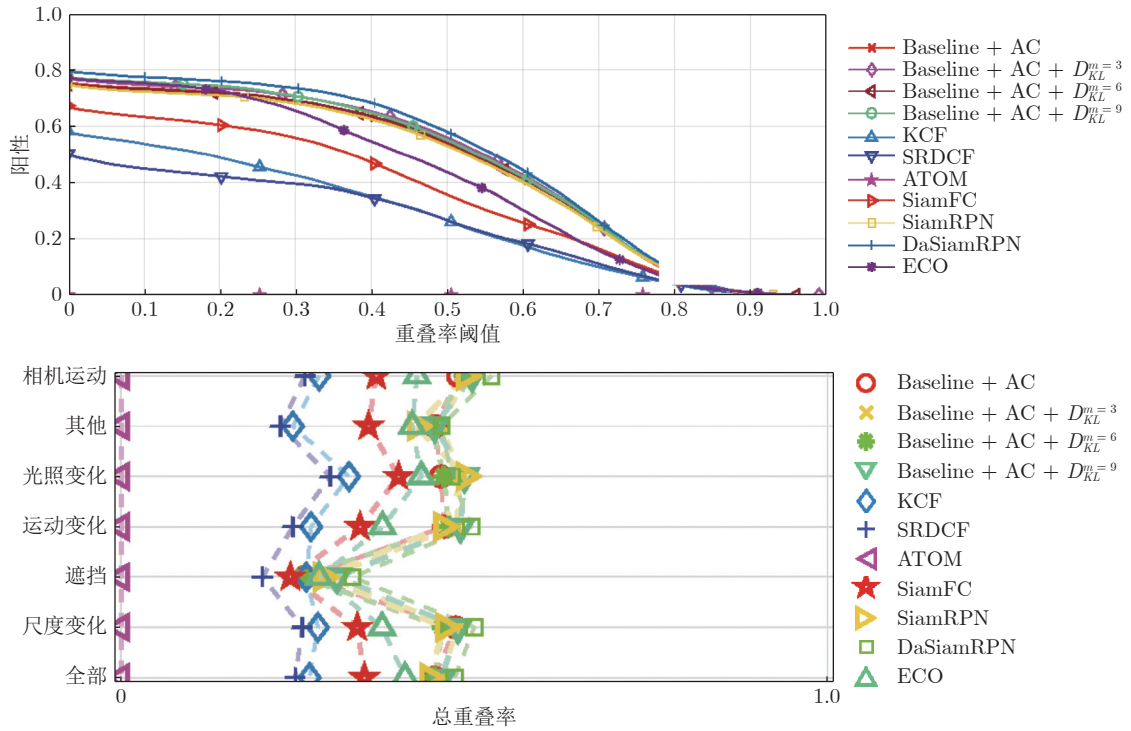


图 10 在 VOT2018 的非监督模式下的 EOA 对比曲线

Fig.10 EOA comparison curve of unsupervised training on VOT2018

DaSiamRPN. 在所有 7 种情况下, 本文算法均好于 OTB 中第 1 名的算法 ECO. 当 $m = 9$ 时, 跟踪器在所有 6 种情景下表现最好. 在整体对比环节, $m = 9$ 时有更好性能. 同时, 与最先进的单目标跟踪器相比, 本文算法也有较强竞争力. 由于 ATOM 未给出在非监督模式下的评测结果, 因此本文显示的都为 0.

在 VOT2018 的实时性能对比中, 得到的期望平均重叠率曲线如图 11 所示. 可以看出, 本文算法与基准算法的实时性能相比较差. 因为本文算法使用异步互相关模型增加了额外开销, 因而实时性能较基准算法差, 但损失的实时性能要远小于跟踪精度与鲁棒性的提升.

为更加直观地看出本文算法与其他算法在实时

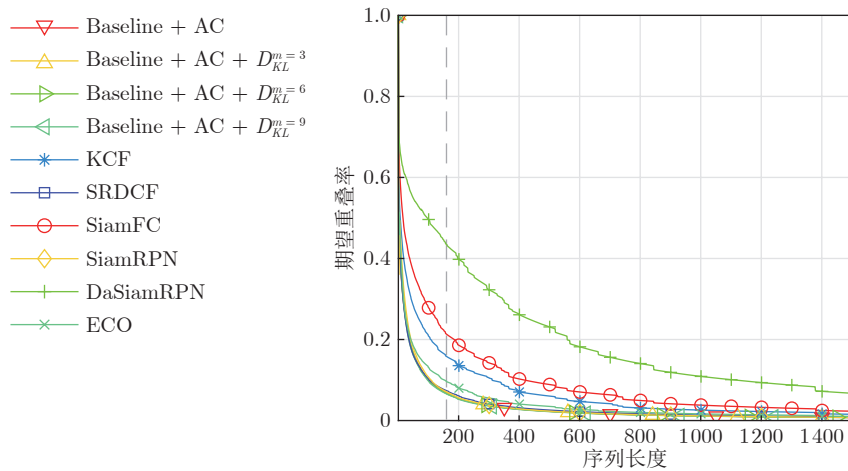


图 11 在 VOT2018 的实时性能对比下的 EOA 对比曲线

Fig.11 EOA comparison curve in realtime on VOT2018

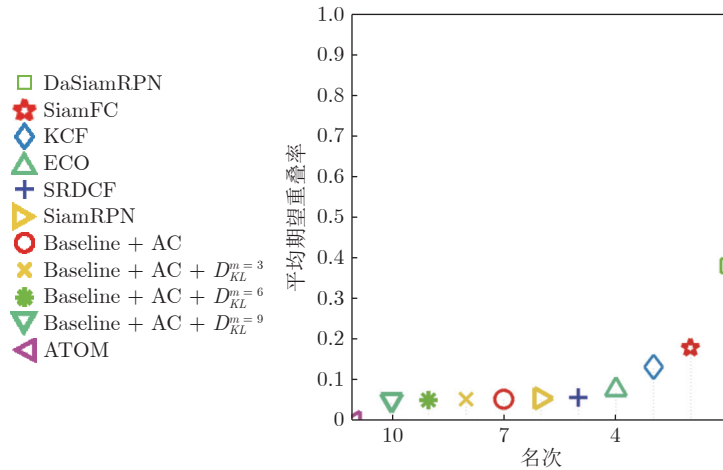


图 12 在 VOT2018 的实时性能对比下不同跟踪器的期望重叠率性能排名情况对比

Fig. 12 Ranking of different trackers' expected overlap ratio in realtime on VOT2018

性能上的对比情况, 本文将不同跟踪器在实时性能上的期望重叠率排名情况进行可视化, 如图 12 所示. 可以看出, 虽然所提算法在实时性能上有所牺牲, 但该指标下的跟踪精度损失很小. 同时, 本文算法在监督实验和非监督实验上的性能都要普遍优于基准算法.

4 结束语

针对基于孪生网络的单目标跟踪器在面对相似语义目标时会发生跟踪失败的情况, 本文提出了一种异步相关的判别性学习模型, 在 Got-10k、TC128、OTB 和 VOT2018 数据集上的实验结果表明, 本文算法可显著提升跟踪器鲁棒性和精度. 在 Got-10k 上的消融实验表明, 本文提出的异步相关判别模型、二阶优化方法和基于 KL 散度的多模型融合管理算法, 可有效提升跟踪性能, 并在 TC128 上做了进一步的验证. 在 OTB 上对跟踪器在不同跟踪条件下的跟踪性能进行对比发现, 本文算法可有效改善基准算法在 11 种跟踪情景下的性能. 并在 Got-10k 上进行了实验结果的可视化, 验证了本文方法带来的判别性. 同时在 OTB 上对这一判别性做了进一步验证. 最后本文在 VOT2018 中验证了本文方法可以有效提升基准算法的精度和鲁棒性. 通过引入异步相关模型, 本文算法在牺牲较少实时性能的情况下提升了准确度. 在未来的工作中, 将探究一种自适应使用异步相关模型个数的方法, 以增强跟踪器的泛化性能.

References

- 1 Liu Qiao-Yuan, Wang Yu-Ru, Zhang Jin-Ling, Yin Ming-Hao. Research progress of visual tracking methods based on correlation filter. *Acta Automatica Sinica*, 2019, **45**(2): 265–275 (刘巧元, 王玉茹, 张金玲, 殷明浩. 基于相关滤波器的视频跟踪方法研究进展. *自动化学报*, 2019, **45**(2): 265–275)

- 2 Liu Chang, Zhao Wei, Liu Peng, Tang Xiang-Long. Auxiliary objects selecting, tracking and updating in target tracking. *Acta Automatica Sinica*, 2018, **44**(7): 1195–1211 (刘畅, 赵巍, 刘鹏, 唐降龙. 目标跟踪中辅助目标的选择、跟踪与更新. *自动化学报*, 2018, **44**(7): 1195–1211)
- 3 Lin Hai-Feng, Ma Yu-Feng, Song Tao. Research on object tracking algorithm based on SIFT. *Acta Automatica Sinica*, 2010, **36**(8): 1204–1208 (蔺海峰, 马宇峰, 宋涛. 基于 SIFT 特征目标跟踪算法研究. *自动化学报*, 2010, **36**(8): 1204–1208)
- 4 Held D, Thrun S, Savarese S. Learning to track at 100 FPS with deep regression networks. In: *Proceedings of the European Conference on Computer Vision*. Amsterdam, Netherlands: 2016. 749–765
- 5 Bertinetto L, Valmadre J, Henriques J F, Vedaldi A, Torr P H S. Fully-convolutional siamese networks for object tracking. In: *Proceedings of the European Conference on Computer Vision*. Amsterdam, Netherlands: 2016. 850–865
- 6 Li B, Yan J J, Wu W, Zhu Z, Hu X L. High performance visual tracking with Siamese region proposal network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE, 2018. 8971–8980
- 7 Zhu Z, Wang Q, Li B, Wu W, Yan J J, Hu W M. Distractor-aware Siamese networks for visual object tracking. In: *Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: 2018. 103–119
- 8 Li B, Wu W, Wang Q, Zhang F Y, Xing J L, Yan J J. SiamRPN++: Evolution of Siamese visual tracking with very deep networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE, 2019. 4277–4286
- 9 Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE, 2016. 4293–4302
- 10 Henriques J F, Caseiro R, Martins P, Batista J. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(3):

- 583–596
- 11 Danelljan M, Hager G, Khan F S, Felsberg M. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 4310–4318
 - 12 Ma C, Huang J B, Yang X K, Yang M H. Hierarchical convolutional features for visual tracking. In: Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 3074–3082
 - 13 Wang N Y, Yeung D Y. Learning a deep compact image representation for visual tracking. In: Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, USA: 2013. 809–817
 - 14 Danelljan M, Robinson A, Khan F S, Felsberg M. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: 2016. 472–488
 - 15 Danelljan M, Bhat G, Khan F S, Felsberg M. ECO: Efficient convolution operators for tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 6931–6939
 - 16 Bolme D S, Beveridge J R, Draper B A, Lui Y M. Visual object tracking using adaptive correlation filters. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: 2010. 2544–2550
 - 17 Ren S Q, He K M, Girshick R, Sun J. 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
 - 18 Danelljan M, Bhat G, Khan F S, Felsberg M. ATOM: Accurate tracking by overlap maximization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 4655–4664
 - 19 Jiang B R, Luo R X, Mao J Y, Xiao T T, Jiang Y N. Acquisition of localization confidence for accurate object detection. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: 2018. 816–832
 - 20 Wang Q, Zhang L, Bertinetto L, Hu W M, Torr P H S. Fast online object tracking and segmentation: A unifying approach. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 1328–1338
 - 21 Huang L H, Zhao X, Huang K Q. Got-10k: A large high-diversity benchmark for generic object tracking in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, **43**(5): 1562–1577
 - 22 Liang P P, Blasch E, Ling H B. Encoding color information for visual tracking: Algorithms and benchmark. *IEEE Transactions on Image Processing*, 2015, **24**(12): 5630–5644
 - 23 Wu Y, Lim J, Yang M H. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1834–1848
 - 24 Kristan M, Matas J, Leonardis A, Vojir T, Pflugfelder R, Fernandez G, et al. A novel performance evaluation methodology for single-target trackers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(11): 2137–2155
 - 25 Ramasubramanian K, Singh A. *Machine Learning Using R: With Time Series and Industry-Based Use Cases in R*. Berkeley: Springer, 2017. 219–424
 - 26 Pearlmutter B A. Fast exact multiplication by the Hessian. *Neural Computation*, 1994, **6**(1): 147–160
 - 27 Zhang J M, Ma S G, Sclaroff S. MEEM: Robust tracking via multiple experts using entropy minimization. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: 2014. 188–203
 - 28 Hare S, Golodetz S, Saffari A, Vineet V, Cheng M M, Hicks S L, et al. Struck: Structured output tracking with kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(10): 2096–2109
 - 29 Jia X, Lu H C, Yang M H. Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 1822–1829
 - 30 Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 798–805
 - 31 Bhat G, Danelljan M, Van Gool L, Timofte R. Learning discriminative model prediction for tracking. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 6181–6190



许龙 东北大学信息科学与工程学院博士研究生. 2016 年获得内蒙古大学学士学位. 主要研究方向为机器学习与视觉目标跟踪.

E-mail: wahaha4ever@163.com

(XU Long Ph.D. candidate at the College of Information Science and Engineering, Northeastern University. He received his bachelor degree from Inner Mongolia University in 2016. His research interest covers machine learning and visual object tracking.)



魏颖 东北大学信息科学与工程学院教授. 分别于 1997 年和 2001 年获得东北大学硕士学位和博士学位. 主要研究方向为图像处理与模式识别, 医学图像计算和分析, 计算机辅助诊断. 本文通信作者.

E-mail: weiying@ise.neu.edu.cn

(WEI Ying Professor at the College of Information Science and Engineering, Northeastern University. She received her master and Ph.D. degrees from Northeastern University in 1997 and 2001, respectively. Her research interest covers image processing & pattern recognition, medical image computation and analysis, and computer-aided diagnosis. Corresponding author of this paper.)



商圣行 东北大学信息科学与工程学院硕士研究生. 主要研究方向为模式识别, 计算机视觉和深度学习.

E-mail: ssh3108@163.com

(SHANG Sheng-Xing Master student at the College of Information Science and Engineering, Northeast-

ern University. His research interest covers pattern recognition, computer vision, and deep learning.)



张皓云 东北大学信息科学与工程学院硕士研究生. 2019 年获得东北大学学士学位. 主要研究方向为目标跟踪与目标检测.

E-mail: nicolascloud@163.com

(ZHANG Hao-Yun Master student at the College of Information

Science and Engineering, Northeastern University. He received his bachelor degree from Northeastern University in 2019. His research interest covers visual object tracking and detection.)



边杰 东北大学信息科学与工程学院硕士研究生. 2017 年获得东北大学学士学位. 主要研究方向为视觉目标跟踪. E-mail: qbzxjb@163.com

(BIAN Jie Master student at the College of Information Science and Engineering, Northeastern Uni-

versity. He received his bachelor degree from Northeastern University in 2017. His main research interest is visual object tracking.)



徐楚翘 东北大学信息科学与工程学院硕士研究生. 主要研究方向为视觉目标跟踪.

E-mail: xuchuqiao@mail.neu.edu.cn

(XU Chu-Qiao Master student at the College of Information Science and Engineering, Northeastern Uni-

versity. Her main research interest is visual object tracking.)