

文章编号 1004-924X(2024)23-3490-14

复杂场景单目标跟踪

林慧兰¹, 赵春蕾^{1*}, 郝志成¹, 刘 时¹, 朱 明¹, 姜 鑫¹, 高 文², 张军强¹

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130000;

2. 比亚迪汽车工业有限公司, 广东 深圳 518000)

摘要:为提高目标形变、遮挡、相似干扰以及视野超出等复杂场景的目标跟踪性能,提出一种复杂场景单目标跟踪算法。基于 Staple 算法研究二维高斯函数像素权重赋值问题,优化颜色直方图统计,增强目标与背景区分度。引入基于峰值旁瓣比(Peak Side Lobe Ratio, PSR)的 HOG 特征、颜色特征的自适应融合机制,合理选择融合系数,确保混合特征更加可靠。分析目标区域中心与上一帧目标中心距离,结合最大混合响应计算最佳中心位置,解决相似目标干扰问题。采用混合响应、HOG 特征、平均峰值相关能量(Average Peak-to-Correlation Energy, APCE)判定目标丢失、遮挡情况,保持目标框位置,实现目标的及时重新跟踪。采用结合之前帧和当前帧信息的模板更新策略,进一步提升跟踪精度,并在 OTB100 数据集中涉及形变、遮挡、视野超出 3 个属性视频上测试。实验结果表明,改进算法在整体和特定属性(形变、遮挡、出视野)的成功率及形变属性的精确度上,较 Staple 算法分别提升了 1.8%, 3.3%, 2% 和 9%;在 VOT16 数据集上,改进算法在整体和遮挡属性上,重叠度较 Staple 提升了 0.022 2 和 0.019 6,满足复杂的特定场景下的目标跟踪需求。

关键词:单目标跟踪;复杂场景;背景抑制;相似目标再识别;丢失判定

中图分类号:TP391.41 **文献标识码:**A **doi:**10.37188/OPE.20243223.3490

Single target tracking in complex scenarios

LIN Huilan¹, ZHAO Chunlei^{1*}, HAO Zhicheng¹, LIU Shi¹, ZHU Ming¹,

JIANG Xin¹, GAO Wen², ZHANG Junqiang¹

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130000, China;

2. BYD Auto Industry Company Limited, Shenzhen 518000, China)

* Corresponding author, E-mail: zhaochunlei@ciomp.ac.cn

Abstract: To address challenges in single-target tracking under complex scenarios such as target deformation, occlusion, similar interference, and out-of-view situations, a novel tracking algorithm is proposed. Building on the Staple algorithm, the method optimizes pixel weight assignment using a two-dimensional Gaussian function and enhances the color histogram to improve target-background distinguishability. An adaptive fusion mechanism based on the Peak Side Lobe Ratio (PSR) is introduced to combine HOG and color features, with carefully selected fusion coefficients ensuring feature reliability. The target's optimal

收稿日期:2024-05-24;修订日期:2024-06-18.

基金项目:国家重点研发计划资助项目(No. 314);吉林省科技发展计划重点研发项目(No. 20220201146GX)

center position is determined by analyzing the distance between the current and previous frame centers, alongside the maximum composite response, effectively mitigating interference from similar targets. Target loss or occlusion is identified using composite response, HOG features, and Average Peak-to-Correlation Energy (APCE), maintaining the target frame's position and enabling timely re-tracking upon reappearance. A template update strategy combining past and current frame information further enhances tracking accuracy. Tests on the OTB100 dataset with deformation, occlusion, and out-of-view scenarios show that the improved algorithm increases overall and specific attribute success rates (deformation, occlusion, out-of-view) by 1.8%, 3.3%, 2%, and deformation precision by 9% compared to the Staple algorithm. On the VOT16 dataset, the overlap rate for overall and occlusion attributes improves by 0.022 2 and 0.019 6 respectively, meeting the demands of target tracking in complex scenarios.

Key words: single object tracking; complex scenarios; background suppression; similar target re-identification; loss determination

1 引言

目标跟踪在计算机视觉中是一个研究热点,对安全监控、军事侦察、自动驾驶等多种应用至关重要。随着应用场景的扩展,目标跟踪算法的性能要求更高,特别是在处理遮挡、目标形变、相似目标干扰以及复杂背景等复杂场景时,算法需要展现出更高的准确性和稳定性。目前,单目标跟踪算法主要分为两大类:传统方法和深度学习方法。传统方法主要包括生成式和判别式方法。

生成式方法致力于构建目标外观模型,并在视频序列后续帧中寻找与该模型最为匹配的区域。典型算法有光流法^[1-2]、卡尔曼滤波^[3-4]和Meanshift均值漂移法^[5-6]。然而,生成式方法在特征提取上过于依赖目标本身,忽视背景信息,导致在目标外观产生剧烈变化或出现遮挡时,容易出现模板漂移或目标丢失。

判别式方法,特别是相关滤波算法,通过训练滤波器来最大化目标区域与搜索区域之间的相关性,进而定位下一帧中响应最强烈的区域作为目标位置。自2010年MOSSE算法^[7]提出以来,相关滤波算法以其高效性和良好的场景适应性受到广泛的关注。随后,Henriques等提出的CSK算法^[8]在MOSSE基础上引入密集采样和核化相关滤波,KCF算法^[9-10]则融入多通道梯度的HOG(Histogram of Oriented Gradients)特征。为不断提升滤波算法性能,研究人员引入多特征融

合、尺度自适应等方法,例如SAMF算法^[11]和DSST^[12]算法,分别引入尺度因子和尺度滤波器增强了对目标尺度变化的适应性;Staple算法^[13]则在DSST基础上结合颜色特征,利用颜色直方图提高对形变目标的跟踪能力。此外,MEEM算法^[14]通过优化边缘度量提升跟踪性能,Struck算法^[15]运用支持向量机进行目标跟踪,TLD算法^[16]融合跟踪、学习和检测策略以保持长时间跟踪的稳定性,CACF算法^[17]将4个方向的背景信息加入训练中以提升目标快速移动下的跟踪准确性。尽管如此,传统方法在应对复杂场景时仍然存在局限性,例如,目标在短暂消失后再次出现,目标发生剧烈形变、快速移动等场景下,跟踪效果不尽如人意。

近年来,基于深度学习的目标跟踪方法取得了显著进展。相比传统方法,深度学习方法强大的特征提取能力能够自动学习更丰富的特征表示,更好地应对目标外观变化。一些代表性的深度学习跟踪算法,如基于CNN的CFNet^[18]和SiamFC^[19]、基于CNN-Transformer的BANDT^[20]、基于Transformer的SwinTrack^[21]、基于图神经网络(GNN)的SUSHI^[22],在多目标跟踪领域广泛应用的SORT^[23-25]系列算法,在多个标准数据集上展现出优异的性能。然而,深度学习方法在数据泛化和稳定性方面仍面临挑战,特别是在数据量有限或应用场景变化时。

为增强复杂场景中单目标跟踪的精确性与稳定性,本文对具有一定抗形变能力的Staple算

法进行改进,提出背景权重直方图、相似目标再识别以及丢失判定 3 个关键模块,并辅以特征自适应融合和模板更新策略。背景权重直方图使用二维高斯函数增强目标与背景的对比度;特征自适应融合基于峰值旁瓣比 PSR 确定 HOG 特征和颜色特征的融合比例;相似目标再识别模块优化多峰值区域的处理,提高对相似目标的辨识能力;丢失判定模块则在目标丢失或遮挡时避免错误累积,并确保目标重新出现时能被追踪。此外,采用结合之前帧和当前帧信息的模板更新策略,以适应目标外观变化,实现连续稳定跟踪。这些整合与改进不仅提升了 Stable 算法的准确性,还显著增强它在复杂场景中的适应性与稳定性。

2 基本原理

Staple 算法的核心在于互补,将相关滤波器和颜色直方图模型两种不同的跟踪模型结合起来,以提高跟踪的准确性和稳定性。为了更好地适应目标尺寸的变化,Staple 算法继承了 DSST 算法的尺度自适应,但是本文不涉及尺度上的改进,因此不过多介绍尺度。

2.1 位置估计

Staple 算法结合 HOG 特征的相关滤波响应和颜色直方图信息,分别计算前景概率分布图并线性加权,得到混合响应图,从而确定混合响应最大值对应的像素位置,即目标中心位置。混合响应如下:

$$\begin{cases} f(x) = \gamma_{\text{tmpl}} f_{\text{tmpl}}(x) + \gamma_{\text{hist}} f_{\text{hist}}(x) \\ \gamma_{\text{tmpl}} + \gamma_{\text{hist}} = 1 \end{cases}, \quad (1)$$

式中: γ_{tmpl} 是 HOG 特征的相关滤波响应权重, γ_{hist} 是前景概率分布权重; $f_{\text{tmpl}}(x)$ 是在图像区域 x 上由相关滤波算法获得的 HOG 特征响应,输入为上一帧图像获得的目标中心位置和尺寸在该帧中的区域块,输出为通过岭回归求解相关滤波得到的响应; $f_{\text{hist}}(x)$ 是在图像区域 x 上由颜色直方图的贝叶斯方法求取的前景概率分布图,通过计算该帧图像中每种颜色在上一帧图像目标区域前景和背景颜色直方图上的频率,获得该帧图像目标区域的前景概率分布。

2.2 在线更新

完成一帧目标定位后,需要在新目标位置提取目标特征用以更新目标模型,以适应目标跟踪过程目标出现的外观变化。Staple 算法的模型更新包括相关滤波更新和颜色直方图更新,通过结合当前帧信息以及之前帧信息更新模型。

相关滤波 h_t 更新为滤波器参数,其计算如下:

$$h_t = \frac{r_t}{d_t + \lambda}, \quad (2)$$

式中: t 为当前滤波器更新的帧号, λ 为正则化系数,用于防止过拟合; d_t 和 r_t 是当前帧相关滤波器 h_t 的参数,其计算公式为:

$$\begin{cases} d_t = (1 - \eta_{\text{tmpl}}) d_{t-1} + \eta_{\text{tmpl}} d_{t,\text{new}} \\ r_t = (1 - \eta_{\text{tmpl}}) r_{t-1} + \eta_{\text{tmpl}} r_{t,\text{new}} \end{cases}, \quad (3)$$

式中: $d_{t,\text{new}}$ 为当前第 t 帧内目标框区域在频域内通过循环移位的期望值的归一化值,该变量用于捕捉目标的运动信息; $r_{t,\text{new}}$ 是将当前第 t 帧内目标框区域在频域内特征与其自身的共轭进行逐元素相乘并归一化来获得,代表能量谱密度或者自相关性,反映目标区域内特征稳定性; d_{t-1} 和 r_{t-1} 为 $t-1$ 帧更新的相关滤波器 h_{t-1} 的参数; η_{tmpl} 为固定学习率,以固定步长更新 d_t 和 r_t 。

利用当前帧与前一帧更新的目标框区域内的概率信息线性加权更新颜色直方图,令 $\rho_t(O)$ 和 $\rho_t(B)$ 为第 t 帧更新的前景和背景直方图,有:

$$\begin{cases} \rho_t(O) = (1 - \eta_{\text{hist}}) \rho_{t-1}(O) + \eta_{\text{hist}} \rho_{t,\text{new}}(O) \\ \rho_t(B) = (1 - \eta_{\text{hist}}) \rho_{t-1}(B) + \eta_{\text{hist}} \rho_{t,\text{new}}(B) \end{cases}, \quad (4)$$

其中: $\rho_{t-1}(O)$ 和 $\rho_{t-1}(B)$ 为第 $t-1$ 帧的前景和背景直方图, $\rho_{t,\text{new}}(O)$ 和 $\rho_{t,\text{new}}(B)$ 为第 t 帧中搜索区域的前景和背景直方图分布, η_{hist} 为固定权重,用于更新前 t 帧的目标和背景直方图。

3 算法改进

本文在权重直方图、相似目标特征再识别以及丢失判定这 3 个关键环节进行优化,辅以特征自适应融合与模板更新策略,实现复杂场景下目标的精准跟踪。算法流程详见图 1,每个模块的作用如图 2 所示。改进算法的具体步骤如下:

Step1: 在起始帧选中跟踪目标,并初始化 HOG 特征、尺度相关滤波器模型以及前景直方

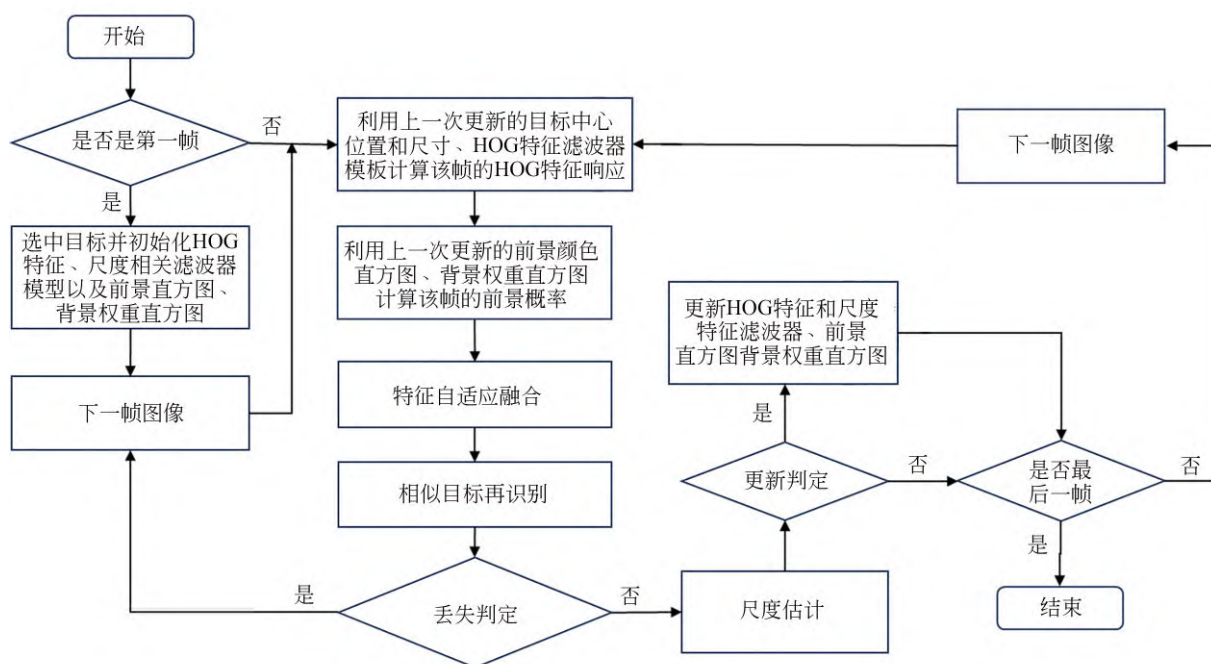


图1 改进 Stable 算法流程

Fig. 1 Flow chart of improved Stable algorithm

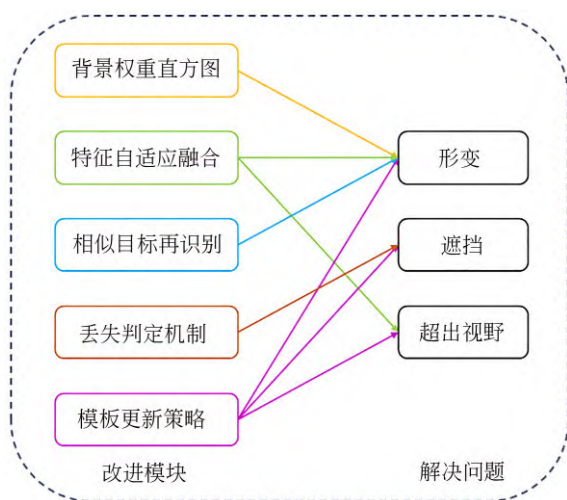


图2 不同模块针对性问题

Fig. 2 Targeted issues for different modules

图、背景权重直方图。

Step2: 读取新一帧图像,利用上一次更新的目标中心位置、尺寸和HOG特征滤波器模板计算该帧的HOG特征响应。

Step3: 利用上一次更新的前、背景权重直方图计算该帧的前景概率。

Step4: 对该帧的HOG特征响应和前景概率进行特征自适应融合。

Step5: 对融合后的混合特征进行相似目标再识别,排除相似目标干扰,确定真正的目标中心位置。

Step6: 对检测到的目标进行丢失判定,确保模型能够准确识别真正目标,而非随着错误目标不断更新。若目标丢失,返回Step2;反之进行Step7。

Step7: 对检测到的目标进行尺度估计,以获得与目标尺寸相匹配的边框。

Step8: 基于当前帧检测结果,按照固定权重,采用新的更新策略对3个模型进行更新。

Step9: 如果当前帧不是最后一帧,则回到Step2继续执行,直至当前帧为最后一帧图像。

3.1 背景权重直方图

在目标跟踪领域,准确区分目标与背景是提高跟踪性能的关键之一。Staple算法在基于图像统计理论获取目标区域的颜色直方图时,只考虑像素点的颜色信息,忽略了位置信息,进而导致在目标发生显著形变或背景与目标颜色相近的情况下,前景概率的区分度不大,难以准确判定目标位置。为解决这一问题,本文提出了一种背景直方图方法,该方法基于二维高斯函数对像素点进行权重分配,以模拟人类视觉系统对信息的

聚焦和抑制机制。

该方法基于高斯函数在图像模糊和平滑等操作中的广泛应用。采用二维高斯函数对背景像素点进行动态权重分配,其中,权重与像素点到目标中心的距离是高斯分布关系。这种权重分配机制有效地增强了目标和背景之间的对比度,尤其是在目标发生显著形变或背景与目标颜色相近的复杂场景下,显著提高了目标跟踪的准确性和稳定性。具体步骤如下:

Step1: 以前一帧检出的目标中心为原点,以搜索区域长宽为尺寸,建立二维高斯函数:

$$W_{x,y} = \exp \left\{ - \left(\frac{(x-x_0)^2}{2\delta_w^2} + \frac{(y-y_0)^2}{2\delta_h^2} \right) \right\}, \quad (5)$$

式中: (x,y) 是背景像素点位置, $W_{x,y}$ 是 (x,y) 处像素值的权重, (x_0,y_0) 是上一帧检出的目标中心, δ_w 和 δ_h 是二维高斯函数的标准差。其计算公式如下:

$$\begin{cases} \delta_w = \frac{w}{2\alpha_w} \\ \delta_h = \frac{h}{2\alpha_h} \end{cases}, \quad (6)$$

式中: w, h 分别是搜索区域的长和宽, α_w, α_h 分别是用于限制 δ_w 和 δ_h 的系数, 根据实际应用时目标的移动速度合理调整搜索区域边缘像素点权重。

Step2: 将所得背景区域权重进行归一化处理, 归一化公式如下:

$$\hat{W}_{x,y} = \frac{W_{x,y}}{\sum \sum W_{x,y}}, \quad (7)$$

式中: $\hat{W}_{x,y}$ 是 (x,y) 位置归一化后的权重。权重归一化用于平衡不同区域的权重贡献, 确保在进行目标与背景分割时, 模型能够更加关注于目标区域, 同时抑制背景噪声的影响。

Step3: 计算背景区域权重直方图 $H(b)$:

$$H(b) = \sum_{x=1}^w \sum_{y=1}^h Q(z_{x,y}, \hat{W}_{x,y}, b), \quad b = 1, \dots, nbins, \quad (8)$$

式中: b 为不同颜色值对应的直方图通道数, $Q(\cdot)$ 表示统计颜色直方图, $z_{x,y}$ 表示 (x,y) 位置上的像素值。

3.2 特征自适应融合

在复杂场景下的目标跟踪中, 单一特征往往

难以应对目标的多样性和环境的不确定性。因此, 融合多种特征成为提高跟踪性能的有效手段。然而, 不同特征在不同场景下的可靠性会有所变化, 这就需要根据相应特征动态调整融合比例的策略。本文基于这一需求提出特征自适应融合机制, 通过引入峰值旁瓣比(PSR)作为特征可靠性的量化指标, 实现了对HOG特征和颜色特征的动态融合。

PSR量化了特征响应的可靠性, 使算法能够根据特征在当前帧的响应强度动态调整其权重, 通过线性融合得到自适应融合的混合响应 Mix_res , 其融合公式如下:

$$Mix_res = (1 - \sigma) \times cf + \sigma \times pwp, \quad (9)$$

式中: cf 为HOG特征响应, pwp 为颜色特征响应, σ 是融合系数。通过计算 cf 和 pwp 的PSR值, 能够确定不同特征的可信度, 并据此构建融合系数 σ , 其计算方式如下:

$$\sigma = \exp \left\{ - \frac{PSR_cf}{PSR_pwp} \right\}, \quad (10)$$

式中: PSR_cf 为HOG特征响应 cf 的PSR值, PSR_pwp 为颜色特征响应 pwp 的PSR值, 其计算公式如下:

$$\begin{cases} PSR_cf = \frac{cf_{\max} - \text{mean}(R_{cf})}{\text{std}(R_{cf})} \\ PSR_pwp = \frac{pwp_{\max} - \text{mean}(R_{pwp})}{\text{std}(R_{pwp})} \end{cases}, \quad (11)$$

式中: $cf_{\max}$ 是HOG特征响应的最大值, R_{cf} 是HOG特征响应 cf 中以 $cf_{\max}$ 对应像素为中心 11×11 邻域以外的区域内响应值的集合, $\text{std}(R_{cf})$ 是 R_{cf} 集合中响应值的标准差; $pwp_{\max}$ 是颜色特征的最大值, R_{pwp} 是颜色特征响应 pwp 中以 $pwp_{\max}$ 对应像素为中心 11×11 邻域以外的区域内响应值的集合, $\text{std}(R_{pwp})$ 是计算 R_{pwp} 集合中响应值的标准差。

3.3 相似目标再识别

在复杂视觉场景中, 目标的多样性和动态变化对跟踪算法提出了更高的要求。特别是当场景中不存在与目标外观相似的干扰物或目标形变后与周围相似时, 传统的目标跟踪方法往往难以准确跟踪特定目标。针对这一问题, 本文提出了相似目标再识别模块, 显著提升了跟踪算法的辨识力和稳定性。

相似目标再识别建立在目标表征和模式识别的理论基础上,利用目标与背景、目标与干扰物之间的视觉差异性以及目标在视频序列中的时空连贯性,增强对相似目标的辨识能力。具体步骤如下:

Step1: 从混合特征响应图中提取局部最大值 $\max L_i, i=1, \dots, I$ (I 表示局部最大值个数), 并计算这些局部最大值与全局最大值 $\max G$ 的比值。

Step2: 判定是否需要进行相似目标再识别。设定阈值 β , 如果局部最大值与全局最大值的比值不超过这个阈值, 则全局最大值对应的位置为目标中心位置, 不再进行相似目标再识别; 若超过阈值, 则认为存在相似目标干扰, 进行 Step3。

Step3: 选择以阈值筛选出的局部最大值 $\max l_p, p=1, \dots, P$ (P 表示筛选后的局部最大值个数) 对应位置为中心, η 倍目标区域大小的区域, 计算该区域内的特征响应、全局最大值 $\max g_p$ 及其位置 (x_p, y_p) 。

Step4: 为保证目标响应强度和运动空间的一致性, 设计了目标中心距离和最大响应值的加权融合策略, 获得最终响应值 $score_p$ 。如公式 (12) 所示:

$$score_p = \lambda \times e^{-d_p} + (1 - \lambda) \times e^{\max g_p}, \quad (12)$$

式中: λ 为加权融合的权重因子, d_p 为局部响应最大值 $\max l_p$ 对应位置 (x_p, y_p) 与上一帧目标中心位置 (x_0, y_0) 的欧氏距离, 其计算公式如下:

$$d_p = \sqrt{(x_p - x_0)^2 + (y_p - y_0)^2}. \quad (13)$$

Step5: 取 $score_p, p=1, \dots, P$ 的最大值 $\max g_p$ 对应的位置 (x_p, y_p) 为跟踪目标的中心位置。

3.4 丢失判定机制

传统的跟踪算法在目标丢失时往往无法有效维持跟踪性能, 导致错误累积和跟踪失效。为了解决这一问题, 本文提出了一种基于连续性分析的丢失判断机制, 根据目标跟踪的连续性和稳定性理论, 分析目标在视频序列中的运动模式和外观变化, 从而建立一套评判标准, 用以评估目标丢失的风险。该机制综合利用混合响应最大值 $\max G$, HOG 特征响应最大值 $\max CF$ 以及混合响应的平均峰值相关能量 (APCE) 的多模态信息, 充分考虑帧与帧之间的连续性, 设计合理的

评判阈值, 使算法能够准确捕获目标跟踪状态的变化, 即使在目标被遮挡或丢失的情况下也能保持跟踪器的稳定性。具体评判方法如下:

$$\begin{cases} \text{mean}(\sum_{i=f-4}^f \max G_i) < \mu_1 \times \text{mean}(\sum_{i=f_s}^f \max G_i) \\ \text{mean}(\sum_{i=f-4}^{\text{frame}} \max CF_i) < \mu_2 \times \text{mean}(\sum_{i=f_s}^{\text{frame}} \max CF_i), \\ \text{mean}(\sum_{i=f-4}^{\text{frame}} apce_i) < \mu_3 \times \text{mean}(\sum_{i=f_s}^{\text{frame}} apce_i) \end{cases} \quad (14)$$

式中: f 是当前帧, f_s 是从第一帧开始准确跟踪目标的帧号, μ_1, μ_2, μ_3 是根据实际应用场景设定的权重系数, $apce$ 是尺寸为 $M \times N$ 的区域响应图的平均峰值相关能量, 其计算公式如下:

$$apce = \frac{(\text{Mix_res}_{\max} - \text{Mix_res}_{\min})^2}{\sum_{m,n} (\text{Mix_res}_{m,n} - \text{Mix_res}_{\min})^2 / MN}, \quad (15)$$

式中: Mix_res 为混合响应图, $\text{Mix_res}_{m,n}$ 为 (m, n) 这一像素点的混合响应值。 $apce$ 代表响应图的波动程度, 值越大, 波动越大, 是目标的概率越大。

根据上述评判方法判断目标处于丢失或遮挡状态时, 采取的策略是保持目标框位置和尺寸与上一帧一致, 直至重新检测到目标。然后, 暂停跟踪器模板的更新操作, 以排除任何可能干扰跟踪准确性的无关信息。

3.5 模板更新策略

在目标跟踪领域, 模板更新是确保跟踪算法适应目标外观变化和环境干扰的关键环节。传统的模板更新方法在目标快速运动或显著形变时跟踪性能会下降, 在目标丢失时还会引起模板漂移。为了解决这些问题, 本文提出一种模板更新策略, 该策略基于混合响应和平均峰值能量来智能地调整更新时机。

模板更新策略的理论基础建立在目标跟踪器的稳定性和适应性理论上。跟踪算法应能够根据目标的外观变化动态调整其模型, 以准确捕捉目标。该策略借鉴控制理论中的自适应控制概念, 通过实时反馈调整系统参数应对外部扰动。本文在丢失判定基础上, 加上混合响应最大值和平均峰值能量作为更新依据。混合响应最大值反映了目标在当前帧中的显著性, 平均峰值能量则量化了目标响应的一致性。通过设定阈

值,算法能够判断当前帧是否需要更新模板,从而避免在目标丢失或遮挡时的错误累积。具体公式如下:

$$\begin{cases} A: \exists i \in \{0, 1, \dots, n-1\}, f_s[i] = f \\ B: Mix_res_{\max} \geq \varphi_1 \times mean(\sum_{j=f_s} Mix_res_j) \\ C: apce \geq \varphi_2 \times mean(\sum_{j=f_s} apce) \end{cases} (16)$$

式中: f 是当前帧, f_s 是从第一帧开始准确跟踪目标的帧号, φ_1, φ_2 是根据实际应用场景设定的权重系数。当满足 $A \wedge (B \vee C)$ 时,该帧的跟踪效果比较好,可以更新模板。更新时采用 Staple 算法的更新策略,其中 HOG 特征按 0.01 的学习率更新,颜色特征按照 0.04 的学习率更新。但在更新背景颜色直方图时,采用背景权重直方图,以便进一步区分背景和目标。

4 实验与结果

实验使用 Matlab2018a 作为软件测试平台。采用 OTB100^[26]中涉及遮挡、形变、超出视野等属性的 64 个视频序列以及 VOT16^[27]共两个公开数据集作为测试视频,比较改进方法和 KCF、Staple 等竞争方法。表 1 展示了改进方法的主要参数设置,未提及参数均与 Staple 算法一致。

表 1 实验参数

Tab. 1 Experimental parameters

参 数	数值
二维高斯函数标准差因子 α_w, α_h	2.8, 2.8
有无相似目标干扰阈值 β	0.7
相似目标再识别区域缩放因子 η	0.8
丢失判定的权重系数 μ_1, μ_2, μ_3	0.28, 0.37, 0.18
模板更新阈值 φ_1, φ_2	0.88, 0.7

4.1 定量分析

4.1.1 OTB64 数据集

OTB64 数据集利用成功率和精确度作为跟踪算法的定量分析指标。其中,成功率(R_s)是指目标真实框与目标跟踪框的 IOU 大于给定阈值的帧数占总帧数的比例;精确度(P)是指算法跟踪到的目标中心位置和目标实际的中心位置差值小于某个阈值占总帧数的比例。采用一次通过评估(One-Pass Evaluation, OPE)方法,对相关跟踪算法进行实验评估。将改进算法与所选数据集整体以及形变、遮挡、出视野 3 个属性下的成功率和精度进行对比分析,结果如表 2 和图 3~图 6 所示。

表 2 不同算法在 OTB64 数据集上的跟踪结果

Tab. 2 Tracking results of different algorithms on OTB64 dataset (%)

算 法	整体		形变		遮挡		出视野	
	R_s	P	R_s	P	R_s	P	R_s	P
Our work	66.4	73.0	68.8	76.4	64.4	68.6	63.8	68.0
SiamFC	65.4	70.3	63.1	69.0	68.3	72.2	63.0	67.3
CFNet	64.3	69.9	65.5	71.4	64.5	69.3	51.1	60.4
STAPLE	64.6	74.3	65.5	74.6	62.4	70.3	54.8	66.8
SAMF	60.2	70.3	61.1	70.4	61.5	70.4	50.3	58.4
MEEM	58.2	74.5	54.5	73.0	59.7	76.2	60.2	73.9
KCF	50.8	62.7	49.5	61.1	48.9	60.0	45.7	49.9
DSST	47.1	61.0	45.2	56.8	44.1	58.2	41.5	48.0
ASLA	43.5	47.2	43.1	46.6	43.0	46.1	37.2	40.9
Struck	43.0	52.3	40.5	51.6	42.0	52.0	48.4	55.9
TLD	40.9	48.8	38.6	46.6	43.3	50.6	41.9	50.1
CSK	35.9	44.8	35.8	45.7	32.9	40.0	27.4	27.6

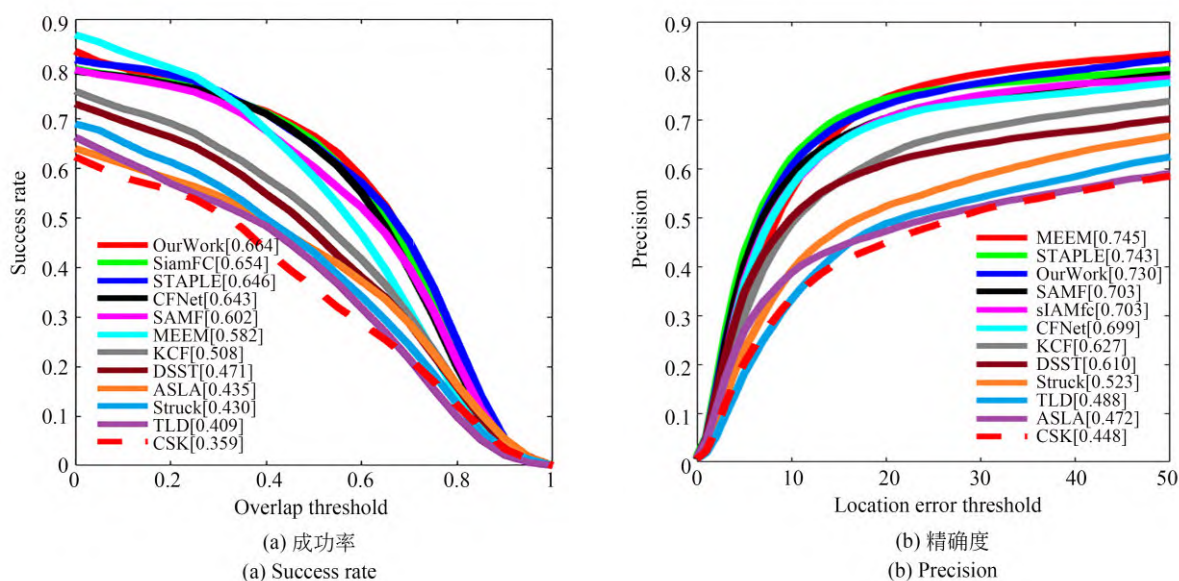


图 3 不同种算法在 OTB64 数据集上的成功率和精确度曲线

Fig. 3 Success rate and precision curves of different algorithms on OTB64 dataset

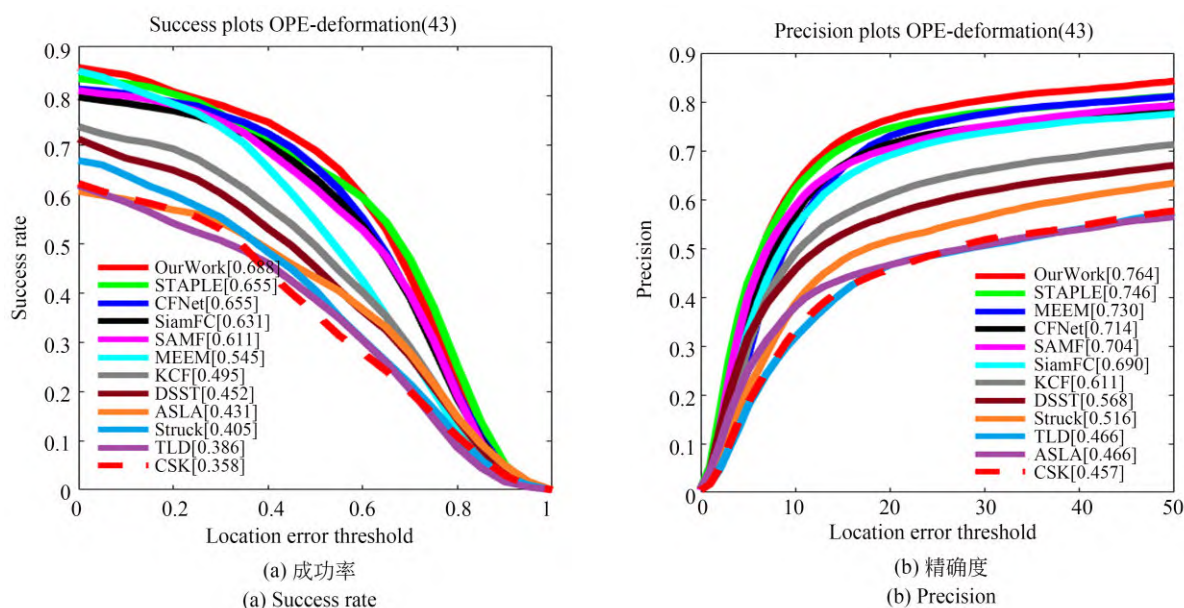


图 4 基于形变属性的成功率和精确度曲线

Fig. 4 Curves of success rate and precision based on deformation attributes

本文所提的算法在 OTB64 数据集上,其成功率相较于 Staple 算法均有所提升,具体提升幅度分别为 1.8%, 3.3%, 2% 和 9%。此外,与 SiamFC 和 CFNet 等基于深度学习的方法相比,本文算法的成功率也有一定程度的提升。在所有算法中,本文算法在整体、形变和超出视野属性

的视频上成功率都是第 1,在遮挡属性上成功率第 3;在精确度方面,本文算法在形变和超出视野两个属性上较 Staple 算法分别提升了 1.8%, 1.2%,在其他属性的视频上效果略有下降。经分析,尤其是目标丢失或遮挡时,由于目标框保持静止不动,当移动目标再次出现时,与框的位

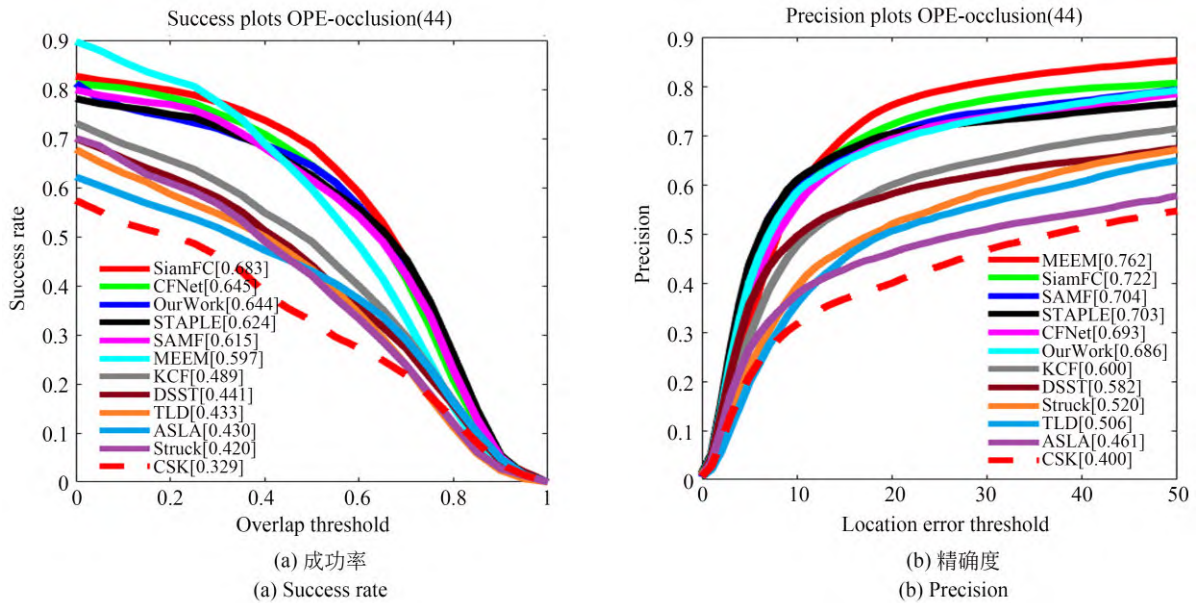


图 5 基于遮挡属性的成功率和精确度曲线

Fig. 5 Curves of success rate and precision based on occlusion attributes

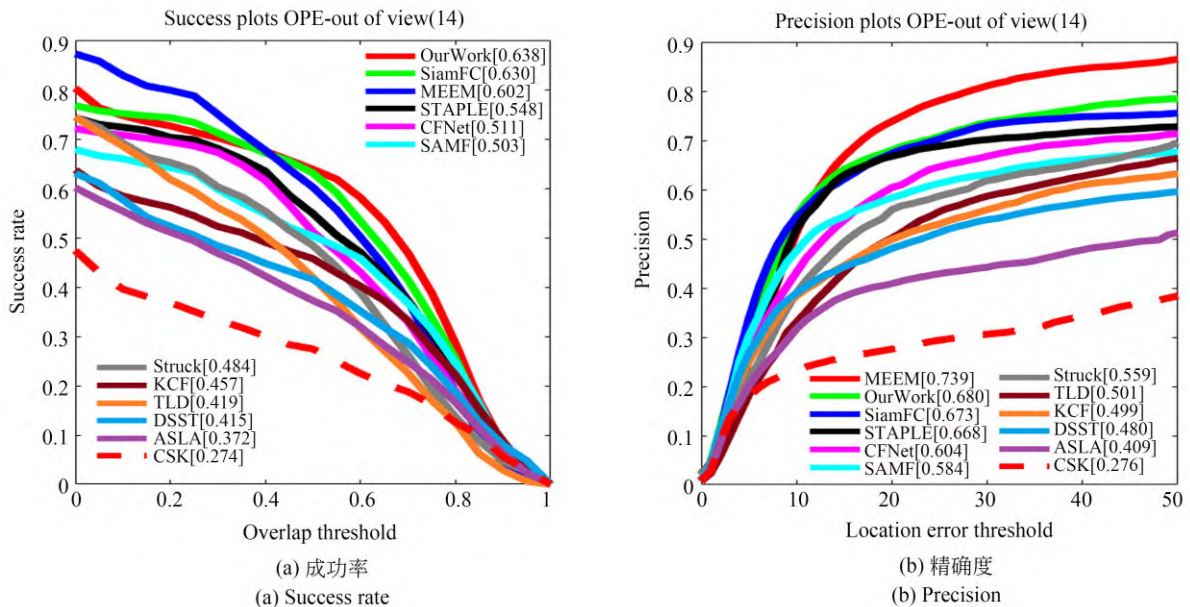


图 6 基于出视野属性的成功率和精确度曲线

Fig. 6 Curves of success rate and precision based on out-of-view attributes

置距离较大,从而影响精确度。针对这一问题,会在后续工作中进一步提升算法的精确度和整体性能。

4.1.2 VOT16 数据集

VOT16 数据集包含 60 个视频序列,分别代表遮挡、光照变化、运动变化、尺寸变化和相机移

动五类属性。考虑到本文改进算法只需指定初始帧目标位置,后续跟踪完全依赖算法本身,采用无监督的方式,因此,将跟踪框与真值框的重叠度(Overlap)作为评价指标,重叠度越高表明跟踪越精确。表 3 展示了本文改进算法与基础算法 Staple、相关滤波算法中的 DAT,DSST,KCF,

表 3 不同算法在 VOT16 数据集上的跟踪结果

Tab.3 Tracking results of different algorithms on VOT16 dataset

算 法	整体	遮挡
Our work	0.409 3	0.254 1
Staple	0.387 1	0.234 5
DAT	0.309 1	0.208 3
DSST	0.325 4	0.229 9
KCF	0.303 2	0.229 5
SAMF	0.349 6	0.268 8
SiamRN	0.422 3	0.261 5

SAMF 以及深度学习算法 SiamRN^[28] 的跟踪结果对比。改进算法在 VOT16 整体上跟踪框的重叠度仅次于 SiamRN,较基础算法 Staple 提高了 0.022 2;在遮挡属性视频上的重叠度排名第 3,

但仍较 Staple 提升了 0.019 6。由此可见,改进算法的参数具有一定的通用性。

4.2 定性分析

为了更直观地对比本文算法与其他方法的效果,挑选 OTB64 中具有挑战性的视频序列进行分析。这些视频在目标形变、遮挡以及出视野等属性上具有代表性,能够全面考验跟踪算法的稳定性和适应性。图 7 中,Bird1 视频代表形变属性;图 8 中,Girl2 视频代表遮挡属性;图 9 中,Tiger2 视频代表目标超出视野属性。

图 7 中,Bird1 视频第 214 帧中的鸟在不断地形变,在后续帧中,画面中还出现与目标相似的干扰物。在处理这一复杂场景时,本文算法能够持续稳定跟随目标,虽然与真值还有一定的差距,但较其他算法,跟踪效果明显增强,从而验证了本文算法的抗形变性及抗干扰性。

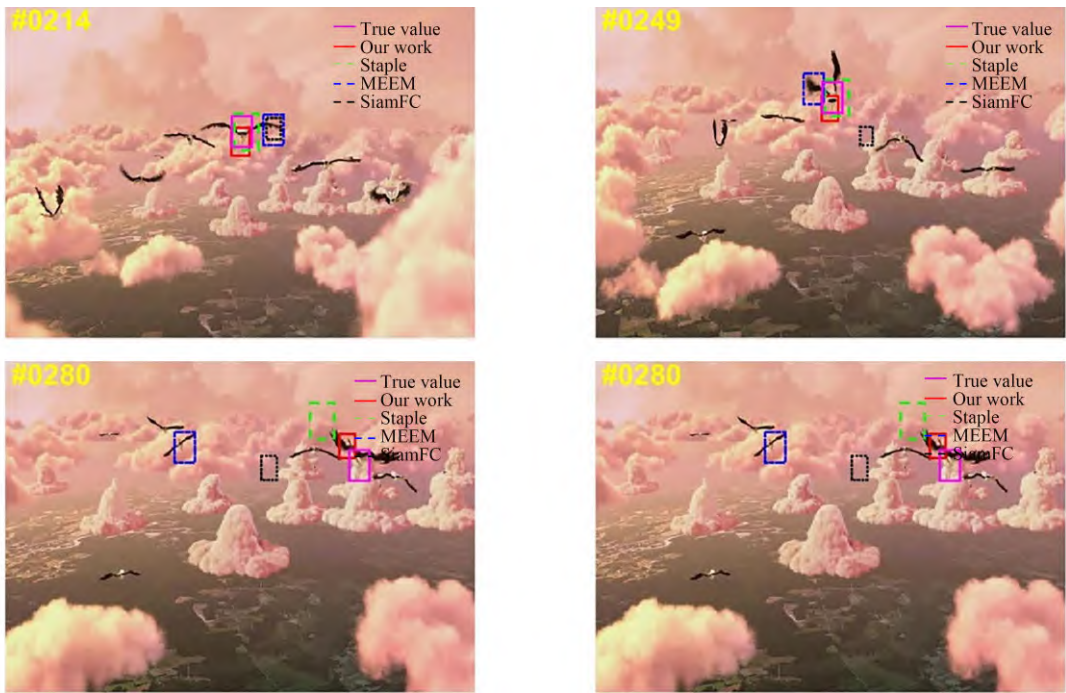


图 7 不同算法在 Bird1 视频序列上的效果对比

Fig. 7 Effectiveness comparison of different algorithms on Bird1 video sequence

图 8 中,目标女孩在视频的第 105 帧及后续帧中被男子遮挡,导致目标在视频中临时不可见。在第 105 帧中所有算法均可以准确跟踪目标,但随着男子的出现,除所提算法之外其他算法均会将跟踪目标转移到男子上,女孩再次出现时检测框也没有重新回到目标上。这一场景验

证了遮挡条件下本文所提算法的有效性,与其他算法相比,本文方法在目标被遮挡和重新出现时表现出更高的跟踪精度和稳定性。

如图 9 所示,布偶老虎在第 86 帧出现部分超出视野。当目标还在视野内时,所有算法的跟踪效果差距不大,但是在第 86 帧本文算法对

于目标超出视野这一动作反应过度,目标框几乎也要超出视野,但在第 92 帧目标再次出现时,相较于其他算法能够快速适应并持续跟踪。这

一场景验证了本文所提算法对于超出视野范围再次出现的目标具有很好的适应性和恢复能力。

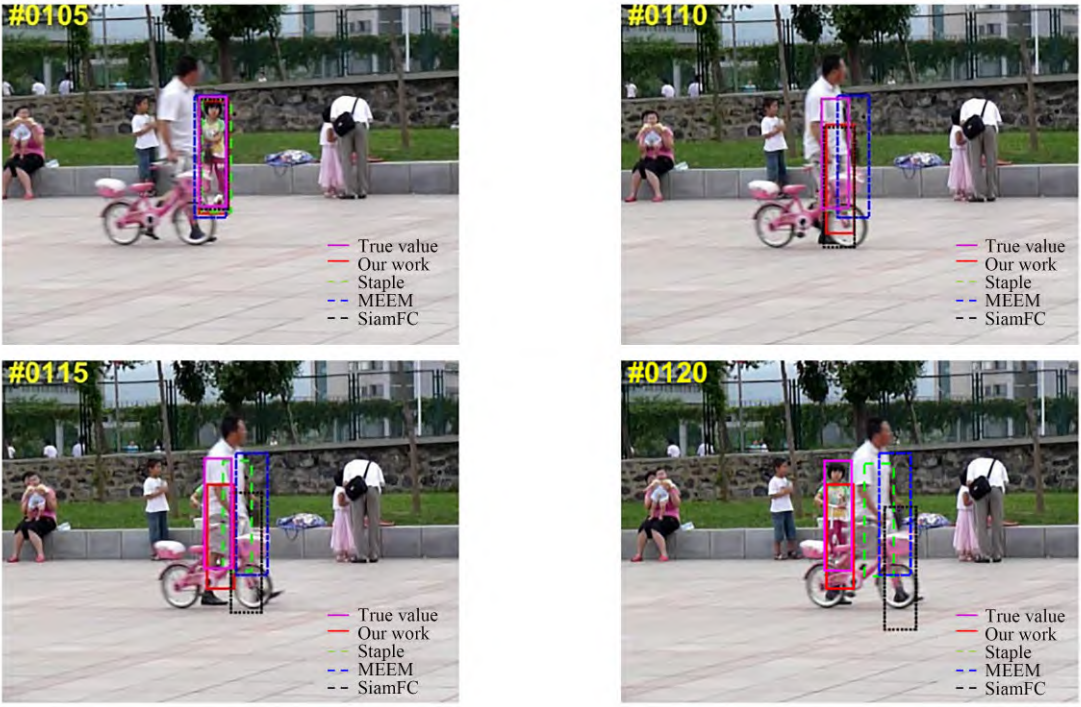


图 8 不同算法在 Girl2 视频序列上的效果对比

Fig. 8 Effectiveness comparison of different algorithms on Girl2 video sequence



图 9 不同算法在 Tiger2 视频序列上的效果对比

Fig. 9 Effectiveness comparison of different algorithms on Tiger2 video sequence

4.3 消融实验

通过探索性实验,在 OTB64 数据集上评估改进算法的不同模块的有效性。公平起见,这里使用的所有实验数据和参数设置都是绝对一致的。实验旨在评估背景权重直方图、特征自适应融合、相似目标再识别、丢失判定机制以及模板更新策略这 5 个关键模块对跟踪性能的影响,基础算法是 Staple 算法,结果如表 4 所示。

从表 4 中可以看出,每个模块都对整体跟踪性能有积极的影响,模板更新策略和背景权重直方图对于提高跟踪的成功率和精确度尤为关键。在特定挑战性场景下,如目标形变和超出视野;相似目标再识别和丢失判定机制显著提高了跟踪性能。

值得注意的是,虽然丢失判定机制在 OTB64 数据集上的效果不是很明显,但在更复杂的场景中,该机制的作用可能更有效。消融实验结果表明,改进的算法在复杂场景下具有较好的跟踪性能,具有更广泛的应用前景。

新增模块虽然增加了算法的复杂性,但是在目标形变、遮挡等复杂情况下,不仅没有增加跟踪时间,而且显著优化跟踪效果。以 Girl2 视频为例,该视频由 1 500 帧图片构成,在基础算法下,系统平均每秒处理 61.4 帧,每帧平均耗时为 16.2 ms;而在改进后的算法中,系统将处理能力提升至每秒 63.3 帧,每帧平均耗时降低至 15.8 ms。如图 8 所示,在基础算法中,在第 110 帧目标被完全遮挡的情况下,模板仍然进行了更新,每次更新耗时平均 7.1 ms,且模板发生漂移;相比之下,改进算法在目标连续被遮挡的几帧中均未更新模板,每次更新的平均耗时为 8.2 ms,遮挡期间跟踪耗时更短,效率更高。此外,Girl2 视频的跟踪精度和成功率也得到了显著提升,基础算法分别为 7.7% 和 7.4%,而改进算法则跃升至 87.7% 和 44.1%。这些数据表明,改进后的算法在处理复杂场景中的目标跟踪时,速度和精度上均超过基础算法。

表 4 不同模块在 OTB64 数据集上的消融实验结果

Tab. 4 Ablation results of different modules on OTB64 dataset (%)

基础 算法	背景权重 直方图	特征自适 应融合	相似目标 再识别	丢失判 定制	模板更 新策略	整体		形变		遮挡		出视野	
						R_s	P	R_s	P	R_s	P	R_s	P
✓						64.6	74.3	65.5	74.6	62.4	70.3	54.8	66.8
✓	✓					65.8	73.2	66.7	74.8	63.7	68.4	60.6	66.6
✓	✓	✓				65.6	74.1	65.3	74.6	63.1	69.8	61.1	69.0
✓	✓	✓	✓			65.9	73.0	65.8	74.9	62.3	67.9	60.3	68.7
✓	✓	✓	✓	✓		66.1	72.9	66.8	74.5	63.1	68.3	60.9	67.3
✓	✓	✓	✓	✓	✓	66.4	73.0	68.8	76.4	64.4	68.6	63.8	68.0

5 结 论

本文根据目标跟踪应用场景复杂性和跟踪准确性的需求,提出一种单目标跟踪算法。该算法采用二维高斯函数构建背景权重直方图加强目标和背景区分;采用基于峰值旁瓣比 PSR 的特征自适应融合机制动态调整不同特征的融合比例;通过目标中心距离和最大混合响应值构成的相似目标再识模块剔除相似目标干扰;结合混合响应、HOG 特征响应和平均峰值能量 APCE 的丢失判定机制有效避免跟踪器模板错误累积,提升目标再次出现时被重新检测

到的概率;最后,利用模板更新策略调整模板更新时机。实验结果表明,本文所提方法在 OTB64 数据集中涉及形变、遮挡和超出视野三个属性的视频序列上目标跟踪的成功率较 Staple 算法分别提升了 3.3%,2% 和 9%,整体上提升了 1.8%,在 VOT16 数据集上,改进算法在整体和遮挡属性上跟踪框和真值的重叠度较基础算法 Staple 分别提高了 0.022 2 和 0.019 6。此外,改进算法在处理复杂场景目标跟踪时,通过一系列判定机制减少不必要的计算,在提升跟踪效率的同时,显著提升了目标的跟踪精度和成功率。考虑到改进算法中一些阈值的设定

依赖经验和应用场景,未来的工作将集中在进一步优化算法的参数设定,提高对极端场景的

适应能力,并探索算法在多目标跟踪等更广泛领域的应用。

参考文献:

- [1] 赵恩喆. 基于光流法与目标跟踪网络结合的视频行人跟踪研究[D]. 哈尔滨: 哈尔滨工业大学, 2022.
ZHAO E ZH. *Research on Video Pedestrian Tracking Based on Optical Flow Method and Target Tracking Network*[D]. Harbin: Harbin Institute of Technology, 2022. (in Chinese)
- [2] 徐萌, 路稳, 方澄, 等. 融合光流特征和显著性检测的目标跟踪算法[J]. 计算机应用与软件, 2024, 41(2): 164-171, 187.
XU M, LU W, FANG CH, *et al.* Target tracking algorithm combining optical flow features and saliency detection[J]. *Computer Applications and Software*, 2024, 41(2): 164-171, 187. (in Chinese)
- [3] 张红颖, 贺鹏艺, 彭晓雯. 基于改进高分辨率神经网络的多目标行人跟踪[J]. 光学精密工程, 2023, 31(6): 860-871.
ZHANG H Y, HE P Y, PENG X W. Multi-object pedestrian tracking method based on improved high resolution neural network [J]. *Opt. Precision Eng.*, 2023, 31(6): 860-871. (in Chinese)
- [4] ALI A, TERADA K. A general framework for Multi-Human tracking using Kalman filter and fast mean shift algorithms [J]. *J. Univers. Comput. Sci.*, 2010, 16: 921-937.
- [5] VOJIR T, NOSKOVA J, MATAS J. Robust scale-adaptive mean-shift for tracking [J]. *Pattern Recognition Letters*, 2014, 49: 250-258.
- [6] 胡亮, 杨德贵, 王行, 等. 基于改进MEANSHIFT的可见光低小慢目标跟踪算法[J]. 信号处理, 2022, 38(4): 824-834.
HU L, YANG D G, WANG X, *et al.* Visible light low-small-slow-target tracking algorithm based on improved MEANSHIFT[J]. *Journal of Signal Processing*, 2022, 38(4): 824-834. (in Chinese)
- [7] BOLME D S, BEVERIDGE J R, DRAPER B A, *et al.* Visual object tracking using adaptive correlation filters[C]. 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA. IEEE, 2010: 2544-2550.
- [8] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels[C]. *European Conference on Computer Vision*. Berlin, Heidelberg: Springer, 2012: 702-715.
- [9] 董超, 郑兵, 李彬, 等. 改进核相关滤波器的海上船只目标跟踪[J]. 光学精密工程, 2019, 27(4): 911-921.
DONG CH, ZHENG B, LI B, *et al.* Shiptarget tracking with improved kernelized correlation filters [J]. *Opt. Precision Eng.*, 2019, 27(4): 911-921. (in Chinese)
- [10] 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.
- [11] LI Y, ZHU J K. A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration [M]. *Computer Vision-ECCV 2014 Workshops*. Cham: Springer International Publishing, 2015: 254-265.
- [12] DANELLJAN M, HAGER G, KHAN F S, *et al.* Discriminative scale space tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(8): 1561-1575.
- [13] 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)*. Las Vegas, NV, USA. IEEE, 2016: 1401-1409.
- [14] ZHANG J, MA S, SCLAROFF S. MEEM: robust tracking via multiple experts using entropy minimization[C]. *Proceedings of the Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12*, Springer, 2014: Part VI 13.
- [15] HARE S, GOLODETZ S, SAFFARI A, *et al.* Struck: structured output tracking with kernels[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(10): 2096-2109.
- [16] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection [J]. *IEEE Transac-*

- tions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [17] GALOOGAHI H K, FAGG A, LUCEY S. Learning background-aware correlation filters for visual tracking[C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy. IEEE, 2017: 1144-1152.
- [18] VALMADRE J, BERTINETTO L, HENRIQUES J, *et al.* End-to-end representation learning for correlation filter based tracking[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA. IEEE, 2017: 5000-5008.
- [19] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al.* Fully-convolutional siamese networks for object tracking[C]. *Proceedings of the Computer Vision-ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, Springer, 2016:Part II* 14, F.
- [20] YANG K, ZHANG H J, SHI J Y, *et al.* BANDT: a border-aware network with deformable transformers for visual tracking[J]. *IEEE Transactions on Consumer Electronics*, 2023, 69 (3) : 377-390.
- [21] LIN L T, FAN H, XU Y, *et al.* SwinTrack: a simple and strong baseline for transformer tracking [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 16743-16754.
- [22] CETINTAS O, BRASÓ G, LEAL-TAIXÉ L. Unifying short and long-term tracking with graph hierarchies [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada. IEEE, 2023: 22877-22887.
- [23] MAGGIOLINO G, AHMAD A, CAO J K, *et al.* Deep OC-sort: multi-pedestrian tracking by adaptive re-identification[C]. 2023 *IEEE International Conference on Image Processing (ICIP)*. Kuala Lumpur, Malaysia. IEEE, 2023: 3025-3029.
- [24] DU Y H, ZHAO Z C, SONG Y, *et al.* StrongSORT: make DeepSORT great again[J]. *IEEE Transactions on Multimedia*, 1809, 25: 8725-8737.
- [25] KIM V, JUNG G, LEE S W. AM-SORT: adaptable motion predictor with historical trajectory embedding for multi-object tracking [J]. *arXiv preprint arXiv:240113950*, 2024.
- [26] WU Y, LIM J, YANG M H. Online object tracking: a benchmark[C]. 2013 *IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA. IEEE, 2013: 2411-2418.
- [27] HUA G, JÉGOU H. *Computer Vision-ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II*[M]. Cham: Springer International Publishing, 2016.
- [28] CHENG S Y, ZHONG B N, LI G R, *et al.* Learning to filter: Siamese relation network for robust tracking[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 4419-4429.

作者简介:



林慧兰(1999—),女,河南信阳人,博士研究生,主要从事计算机视觉方向的研究。E-mail: linhuilan21@mails.ucas.ac.cn

通讯作者:



赵春蕾(1989—),女,吉林通化人,博士,助理研究员,2011年于长春理工大学获得学士学位,2016年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事视频压缩传输、计算机视觉方向等方面的研究。E-mail: zhaochunlei@ciomp.ac.cn