第 58 卷 第 4 期/2021 年 2 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

多模型融合的孪生网络视觉跟踪

车满强*，李树斌，葛金鹏

广州海格通信集团股份有限公司无人系统技术创新中心，广东 广州 510700

摘要 为提升孪生网络视觉跟踪算法的准确性，提出一种融合多任务差异化同质型模型的孪生网络视觉跟踪算法。首先在决策层对孪生网络视觉跟踪模型与目标分割模型进行融合，然后结合多尺度搜索区域、目标上下文特征、多学习率模型更新策略进行跟踪。在标准数据集 VOT、OTB、LaSOT、UAV123 上进行算法评估。实验结果表明，所提算法在遮挡、快速运动、光照变化等干扰下可以稳定跟踪目标。

关键词 机器视觉；视觉跟踪；孪生网络；模型融合；上下文特征；多学习率

中图分类号 TP491.4 **文献标志码** A **doi**：10.3788/LOP202158.0415004

Multimodel Integrated Siamese Network Visual Tracking

Che Manqiang*, Li Shubin, Ge Jinpeng

Unmanned Systems Technology Innovation Center, Guangzhou Haige Communications Group Incorporated Company, Guangzhou, Guangdong 510700, *China*

Abstract A siamese network visual tracking algorithm based on fusion multitask differentiated homogeneous models is proposed to improve the accuracy of the algorithm. First, the siamese network visual tracking and target segmentation models are fused in the decision-making layer. Then, they are combined with multiscale search area, contextual features, and multilearning rate model updating strategy to track. Different algorithms are evaluated using standard datasets, namely, VOT, OTB, LaSOT, and UAV123. Experimental results show that the proposed algorithm can stably track the object under the interference of occlusion, fast motion, and illumination change, among others.

Key words machine vision; visual tracking; siamese network; model integration; contextual feature; multi learning rate

OCIS codes 150.1135; 100.4996; 100.2000; 100.2960

1 引 言

视觉跟踪是计算机视觉领域的一个重要分支，在视频序列的初始帧指定跟踪目标，在后续帧中即可连续地找到目标，广泛地应用于自动驾驶、车辆导航、人机交互等各项任务中。

在过去的几年，相关滤波跟踪算法[1]因快速高效性，得到广大学者的关注。多通道计算[2]，方向梯度直方图[2]、颜色属性[3]及深度卷积特征[4]等多种特征的引入，空间信息约束[5-6]和其他辅助策略[7]的应用极大地促进了相关滤波跟踪算法的发展，跟踪性能也得到显著提升。近年来，随着深度学习理论的不断发展及视觉跟踪训练数据集的不断完善，基于深度学习的视觉跟踪也得到了长足的发展。其中全卷积孪生网络(SiamFC)视觉跟踪方法[8]因具有强大的端到端学习能力和实时性，备受关注。SiamFC 视觉跟踪方法主要将视觉跟踪当作相似目标匹配任务，利用离线数据集训练的网络模型计算视频帧与初始帧目标之间的相似性响应值，通过响应值确定目标的位置。Bertinetto 等[8]通过离线训

收稿日期：2020-07-28；**修回日期**：2020-08-03；**录用日期**：2020-08-12

* **E-mail**：1229462669@qq.com

练的全卷积网络结构计算目标之间的相似度来进行跟踪，速度达 100 frame/s；Wang 等[9]在孪生网络中引入了三种注意力机制，有效地缓解了模型的过拟合现象；Li 等[10]在孪生网络中加入目标检测中的区域生成网络(RPN)[11]，提升了跟踪的精度。除此之外，还有部分改进尝试将现有的相关滤波跟踪作为神经网络结构进行端到端训练，提升跟踪的鲁棒性。Wang 等[12]将核相关滤波视为在 SiamFC 的网络层，并通过将网络输出定义为对象位置的概率热图来反向传播，充分利用了相关滤波高效的特性；Zhang 等[13]在文献[12]基础上引入空间对齐模块，进一步提升了跟踪的鲁棒性。在上述改进的基础上，仍然限制孪生网络视觉跟踪鲁棒性的其中一个原因是无法融合目标背景信息和历史帧图片信息[14]。为解决该问题，Guo 等[15]通过学习特征转换来抑制目标表观变化和背景信息的干扰；Zhu 等[16]改进训练方法，通过包含语义信息的负样本增强模型的判别力，抑制背景信息的干扰；Bhat 等[14]在进行在线跟踪时，使用多帧图片进行训练更新，融入了背景信息和目标历史信息，极大地提升了孪生网络视觉跟踪算法的鲁棒性。

随着视觉跟踪理论的不断完善，视觉跟踪算法得到了长足的发展，但仍然存在定位不够准确、特殊目标难以跟踪的现象。为进一步提升视觉跟踪算法的精度，本文以 DiMP 视觉跟踪算法[14]为基础，进行几点改进。1)多模型融合。通过分析视觉跟踪与目标分割的优势，设计差异化同质型模型融合框架，融合多模型预测目标位置。2)多尺度搜索区域。根据目标相对于整幅图像的大小，设计多尺度搜索区域，提升算法对小目标的跟踪性能。3)上下文特征。根据初始目标的形状，在目标区域融入部分的背景信息，使用带有背景信息的上下文特征训练跟踪分类器，提升算法对特殊形状目标的跟踪性能。4)多学习率融合。判断算法定位的可靠性，根据可靠性设置多个学习率，自适应选择学习率更新定位模型，防止定位不准确导致模型漂移。

2 差异化同质型模型融合跟踪

2.1 差异化同质型跟踪模型

基于孪生网络视觉跟踪分割方法 SiamMask[17]，首先采用大规模的离线数据集训练卷积神经网络参数，学习同类相似物体之间的通用共性特征，同时获取相同目标匹配函数；然后在进行在线跟踪时，输入从初始帧和后续帧中裁剪的搜索区域图像，通过离线训练的匹配函数对初始帧模板与搜索区域图像进行相关运算，获取相似性响应得分图，从得分图中获取最大值所在位置并将其作为预测的目标位置；在对跟踪目标进行分割时，在跟踪模型后面添加反卷积运算(离线训练)，通过反卷积操作获取像素级分类，得到目标前景分割结果。上述孪生网络视觉跟踪算法只用到了目标区域内提取的特征，未使用到视频帧背景区域的特征，在一些特别复杂场景的跟踪中，存在判别力不够的问题，难以处理复杂场景(相似物体干扰)下的跟踪任务；但其在大规模数据集上预训练得到的目标分割模型，在背景与目标区别较大时，具有较高的定位准确性，通过目标分割图像拟合出的目标框更加贴合目标的真实尺寸。

为解决孪生网络视觉跟踪算法不够鲁棒的问题，Bhat 等[14]提出一种改进的孪生网络视觉跟踪算法 DiMP。该算法除采用大规模数据集离线训练的骨干网络参数外，在进行在线跟踪时，首先对截取的初始帧目标区域图像进行数据增强，通过提取多张图像的特征来训练初始化分类器；然后利用初始化分类器和结合背景信息的特征获取优化后的目标分类器，在后续帧跟踪时，计算目标搜索区域特征与分类器之间的相似性响应图；再通过牛顿迭代法获取目标精确的位置；最后将视频序列的某一帧及前面抽取的三帧作为训练集，后面抽取的三帧作为测试集，对目标分类器进行训练更新，以使分类器模型适应目标和环境的变化。虽然 DiMP 具有较强的判别性，但其采用 IOU-Net[18]预测目标的尺度大小，在视频的部分帧处会出现目标框未完全贴合目标的现象。在进行连续跟踪时，DiMP 主要通过上一帧目标的位置和尺度确定搜索区域，且根据预测的结果更新分类器，所以累计多帧时会出现目标框过大或过小现象，导致模型发生漂移，从而跟踪失败。

2.2 融合跟踪算法

单独使用孪生网络目标跟踪与分割模型时，在目标特征明显的场景下，目标分割较为准确，拟合的目标框更加贴合目标，但难以处理具有相似物干扰的复杂环境下的跟踪和分割。改进的 DiMP 模型对复杂场景下的跟踪具有较强的鲁棒性，但跟踪中会出现预测目标框未完全贴合目标的现象，累计多帧时容易导致跟踪失败。孪生网络目标跟踪与分割模型可弥补 DiMP 跟踪模型的不足，因此可通过集成学习构建融合跟踪模型，提升跟踪的准确性。具体的融合跟踪模型如图 1 所示。

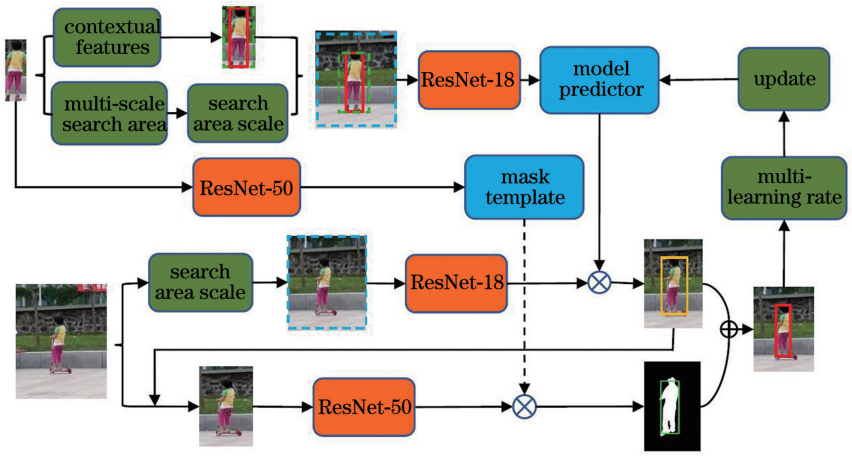

图 1　所提算法的总体框架

Fig. 1　Overall framework of the proposed algorithm

设计的融合跟踪算法主要包含两条线路，第一条为核心目标跟踪定位模块(DiMP)，第二条为目标分割辅助定位模块(SiamMask)。首先，在初始帧中选取目标区域，确定包含上下文特征的新目标区域和目标搜索区域大小，使用预训练好的 ResNet 模型提取深度卷积特征，训练目标预测模型和目标分割模板；然后，在后续视频序列中，先根据初始帧确定的目标搜索区域大小获取搜索区域图片，提取卷积特征，并对其与目标预测模型进行相关计算，计算目标的位置和尺度大小；再根据预测的目标位置和尺度大小获取预测目标区域附近的图片，通过目标分割模块获取目标分割结果并拟合目标区域框；再在决策层融合目标跟踪框与目标分割拟合框得到最终目标的位置和大小，并作为最后的预测结果输出；最后，完成目标跟踪后，根据跟踪结果使用多学习率融合方式更新目标预测模型。

目标跟踪定位与目标分割辅助定位得到目标位置和大小的融合方法为

$$p=\alpha p_{\text{track}}+(1-\alpha)p_{\text{mask}},\tag{1}$$

式中：p_{track} 与 p_{mask} 分别为跟踪预测的目标位置框和分割获得的目标位置框；α 为融合系数，是经验值，通过实验验证确定。图 1 中其他相关计算过程与选取的基础算法 DiMP[14] 和 SiamMask[17] 相同。

3　其他改进

3.1　多尺度目标搜索区域

视觉跟踪主要在连续的运动过程中确定目标的位置和大小，根据目标连续运动的特点，目标在相邻帧之间运动的范围有限，所以在下一帧图片中确定目标时，只需要在上一帧目标所在位置周围一定范围内寻找即可。判别式视觉跟踪主要在初始帧处通过目标区域特征训练分类器，然后在后续视频帧中根据上一帧图像中目标所在位置确定目标搜索区域，提取搜索区域特征，计算特征与分类器之间的相关响应，根据响应值确定目标的位置。

在视觉发展的初期，很多算法在整张图中寻找目标，由于过多背景区域的存在，目标定位容易出现漂移现象，而且计算量过大，影响跟踪的实时性。随着视觉跟踪的不断发展，出现多种确定目标搜索区域的方法，典型的有：1)根据初始帧目标长宽的大小，直接将长和宽加固定值作为搜索区域的大小[19]；2)将目标尺度的大小放大固定的倍数作为搜索区域的大小[4]；3)首先将目标尺度的大小放大固定的倍数，然后将该放大区域变为面积相同的正方形区域作为搜索区域[14,18]。这些方法能够较好地适应目标形态和尺度的变化，也是目前较为常用的方法。

基础目标定位算法 DiMP[14] 使用方法 3)，在所有跟踪过程中均采用同一固定倍数确定搜索区域大小。不同大小的目标在运动过程中呈现出的特点不同，小目标包含目标自身的特点和运动信息较少，在使用正常或者较小的搜索区域时，所在位置容易超出搜索区域；而较大的目标包含更多自身的表征信息，在正常或者较小范围内搜索时即可定位到目标。例如，较小的球类目标运动时，运动速度快，目标较模糊，易超出搜索区域范围；普通车辆等物体特征较明显，较容易定位到目标，不易超出搜索区域。

为解决上述问题，采用多尺度搜索区域方法确定目标搜索区域。首先在第一帧，计算目标区域大小与图像大小之间的比值 γ，当 $\gamma\leqslant\gamma_1$ 时，判定目标为小目标，此时设定较大的搜索区域，即将目标区域放大 ζ_1 倍作为搜索区域，同时为更好地表征目标，

将目标区域图像通过双线性插值放大 τ 倍，并将结果输入到网络提取特征；当 $\gamma>\gamma_1$ 时，判定目标为常规目标，采用普通的搜索区域，即将目标区域放大 ζ_2 倍作为搜索区域。具体可表示为

$$s_{\text{object}}=\begin{cases}w_{\text{resize}}\times h_{\text{resize}}, & \gamma\leqslant\gamma_1\\ w_{\text{object}}\times h_{\text{object}}, & \gamma>\gamma_1\end{cases},\tag{2}$$

$$s_{\text{search}}=\begin{cases}s_{\text{object}}\times\zeta_1, & \gamma\leqslant\gamma_1\\ s_{\text{object}}\times\zeta_2, & \gamma>\gamma_1\end{cases},\tag{3}$$

式中：γ_1 为判定阈值；s_{object} 为输入网络的目标区域大小；w_{object}，h_{object} 分别为原图像目标区域的宽和高；w_{resize}，h_{resize} 分别为通过双线性插值将原目标区域图像放大 τ 倍的宽和高；s_{search} 为搜索区域大小。

3.2　上下文特征

采用判别式视觉跟踪算法对初始帧图像进行训练，得到的分类器的判别性对目标定位的准确度有重要的影响。所提融合跟踪算法选用的主干定位模块 DiMP 在训练分类器时，首先截取目标区域的图像，然后进行旋转、平移、翻转、模糊等数据增强处理，提取增强数据的卷积特征，训练分类器用于后序视频帧目标的定位。从初始帧提取到的特征的表征能力对分类器的判别力有重大影响。

视觉跟踪中有部分长条形目标的长宽比例较大，这类目标在发生形变、快速运动等情形时，易出现与分类器之间的响应值不明显现象，导致定位不够准确。根据文献[20]，图像中目标的少数背景信息可以提升分类器的泛化能力，因此在跟踪时，首先在初始帧计算目标的长宽比，将长宽比大于 2 的长条形目标的长和宽放大相应的倍数，引入部分的背景信息，通过提取带有少量背景信息区域的上下文特征来训练分类器。上下文特征的引入，一方面使得目标样本包含更加完整的目标特征，更加适应于自身长宽大幅度变化的形变，另一方面，使得目标样本包含自身运动的边界信息，目标在发生快速运动时，也能完整地定位到目标。

3.3　多学习率融合

所提多模型融合跟踪算法属于在线跟踪模型，需要根据最终预测结果对目标预测定位模型进行更新。现有的在线跟踪模型更新策略主要有：1）每一帧更新一次模型[4]；2）间隔相同帧更新一次模型[19]；3）跟踪到目标时选用同一学习率更新模型，未定位到目标时不更新模型[14,18]。目标在运动过程中是多变的，快速运动、形变等会导致某些视频帧中目标特征不明显，从而定位不够准确。在定位不够准确时如果过度更新模型，会导致模型发生偏差，影响后续定位的准确性。因此根据目标定位的准确性，采用多学习率融合方式更新主干定位模型。

所提视觉跟踪算法主干定位模型主要通过计算分类器与搜索区域特征之间的响应图，寻找响应图最大值所在位置以确定目标中心位置，响应图峰值在一定程度上可以反映目标定位的准确度。在目标定位准确时，响应图应具有明显的峰值，其他位置值应较为平缓。响应图的平均峰值相关能量比(APCE)[21]可以很好地反映响应图峰值与周围值之间的关系，本文通过联合响应图最大值和 APCE 来判断跟踪的置信度，根据置信度采取多个学习率更新模型。获取到响应图 $\boldsymbol{F}$ 后，计算 APCE：

$$R_{\text{APCE}}=\frac{|F_{\max}-F_{\min}|^2}{\text{mean}\left[\sum\limits_{w_F,h_F}(F_{w_F,h_F}-F_{\min})^2\right]},\tag{4}$$

式中：$F_{\max}$，$F_{\min}$ 和 F_{w_F,h_F} 分别为 $\boldsymbol{F}$ 的最大值，最小值和第 w_F 行，h_F 列的元素。APCE 值越大，表示预测越准确。则可根据预测的准确性，通过设置多学习率 η 来更新主干预测模型：

$$\eta=\begin{cases}\eta_1, & F_{\max}\geqslant\kappa_1\ \text{and}\ R_{\text{APCE}}\geqslant\kappa_2\\ \eta_2, & F_{\max}<\kappa_1\ \text{or}\ R_{\text{APCE}}<\kappa_2\end{cases},\tag{5}$$

式中：η_1 与 η_2 为不同条件下的学习率值；κ_1，κ_2 为用于预测准确性的判断阈值。

4　实验结果与分析

为验证所提算法的性能，使用标准数据集 OTB[22]、LaSOT[23]、UAV123[24] 及 VOT[25-26] 对所提算法进行测试，评价标准均使用各数据集官网中的评价方法。所有实验均在配置为 Intel® Xeon(R) W-2135 CPU @ 3.70 GHz×12，内存 32 GB，配有 NVIDIA RTX 2080 GPU 的 ubuntu16.04 系统下完成。

所提算法在 DiMP[14] 和 SiamMask[17] 的基础下改进得到，与 DiMP 相似，具有两个版本。第一个使用 ResNet-18 提取特征(Ours_Res18)，跟踪速度达 42 frame/s，第二个使用 ResNet-50 提取特征(Ours_Res50)，跟踪速度达 31 frame/s。算法基础参数与原算法相同，添加及优化的改进参数为：位置融合系数 $\alpha=0.8$；目标大小判断阈值 $\gamma_1=0.005$；小目标搜索区域放大倍数 $\zeta_1=6$，正常目标搜索区域放大倍数 $\zeta_1=4.5$，小目标区域缩放为 22×16；学习率调整阈值 $\kappa_1=0.8$，$\kappa_2=0.75$；跟踪稳定时学习率为 0.01，不稳定时为 0.0075。

4.1　算法改进实验

表 1 列出了加入各种改进策略后算法在

VOT2019 数据集上跟踪的结果。可以看出：在 DiMP50 跟踪算法[14]的基础上，融入 Mask 目标分割进行位置融合后，目标的尺度预测更加准确，准确度上升最明显；多尺度搜索区域、上下文特征、多学习率策略的引入，使得目标的定位更加准确，跟踪的失败数明显降低；非重置重叠期望（EAO）整体呈上升趋势，说明各种策略的不断加入，所提算法的稳定性不断增强。

表 1 加入各项策略后，所提算法的跟踪结果

Table 1 Tracking results of proposed algorithm after adding various strategies

DiMP50	Mask	Multi-scale search area	Contextual feature	Multi-learning rate	EAO	Number of failures	Accuracy
√					0.379	13.583	0.574
√	√				0.386	13.245	0.596
√	√	√			0.404	12.336	0.603
√	√	√	√		0.407	12.570	0.603
√	√	√	√	√	0.414	11.245	0.605

4.2 与其他算法的对比实验

为进一步验证所提算法的整体性能，在标准数据集 VOT2018、VOT2019、LaSOT、OTB100、UAV123 上进行测试，并与现有的优秀跟踪算法进行对比。

VOT2018 和 VOT2019：这两组数据集为 2018 年和 2019 年视觉跟踪比赛（VOT）的短时比赛数据集，各包含 60 组视频，其中有 20 组为不同视频，VOT2019 比 VOT2018 难度更大。使用比赛排名标准 EAO、Robustness、Accuracy 进行算法评估与对比。表 2 为在 VOT2018 数据集上所提算法与 SimMask[17]、DaSiam-RPN[16]、LADCF[27]、ATOM[18]、SiamRPN++[28]、DiMP18[14]之间的对比结果。表 3 为在 VOT2019 数据集上所提算法与 SimMask[17]、cola[26]、ATOM[18]、DRNet[26]、trackyou[26]、ATP[26]、DiMP[14]之间的对比结果。可以看出：在 VOT2018 数据集上，Ours_Res50 的 EAO 比原 DiMP50 高 0.011，Ours_Res18 的 EAO 比原 DiMP18 高 0.008；在 VOT2019 数据集上，Ours_Res50 的 EAO 比原 DiMP50 高 0.035，鲁棒性也都优于原 DiMP 跟踪算法。说明多种策略的使用，提升了算法跟踪的整体稳定性。在两种数据集上，Ours_Res50 整体性能也优于选取的对比算法，其中 ATP 算法为 VOT2019 短时跟踪比赛的冠军。

表 2 不同算法在 VOT2018 数据集上的对比结果

Table 2 Comparison results of different algorithms on VOT2018 dataset

Parameter	SiamMask	DaSiam-RPN	LADCF	ATOM	SiamRPN++	DiMP18	DiMP50	Ours_Res18	Ours_Res50
EAO	0.380	0.383	0.389	0.401	0.414	0.402	0.440	0.410	0.451
Robustness	0.276	0.155	0.159	0.204	0.234	0.182	0.153	0.176	0.160
Accuracy	0.609	0.507	0.503	0.590	0.600	0.594	0.597	0.597	0.608

表 3 不同算法在 VOT2019 数据集上的对比结果

Table 3 Comparison results of different algorithms on VOT2019 dataset

Parameter	SiamMask	cola	ATOM	DRNet	trackyou	DiMP50	ATP	Ours_Res18	Ours_Res50
EAO	0.287	0.371	0.292	0.395	0.395	0.379	0.394	0.350	0.414
Robustness	0.405	0.277	0.361	0.229	0.237	0.243	0.255	0.277	0.208
Accuracy	0.575	0.589	0.582	0.580	0.586	0.574	0.619	0.603	0.605

LaSOT：选取 LASOT 的测试数据集对算法进行测试，该测试数据集总共包含 280 个视频，涵盖了视觉跟踪可能遇到的各种挑战，并包含多种长时跟踪视频，是现有最大难度数据集之一。图 2 为所提算法与 DiMP[14]、ATOM[18]、SiamRPN++[28]、MDNet[29]、VITAL[30]、SiamFC[8]、Dsiam[15]、CFNet[31]、BACF[32]、PTAV[33]的对比结果。可以看出：Ours_Res50 的距离精度比原 DiMP50 高 0.9 个百分点，Ours_Res18 的距离精度比原 DiMP18 高 0.8 个百分点；Ours_Res50 的成功率比原 DiMP50 高 0.2 个百分点，Ours_Res18 的成功率比原 DiMP18 高 0.3 个百分点，且优于选取的对比算法。

OTB100：OTB100 数据集总共包括 100 组短视频，是视觉跟踪测试的经典数据集之一。对所

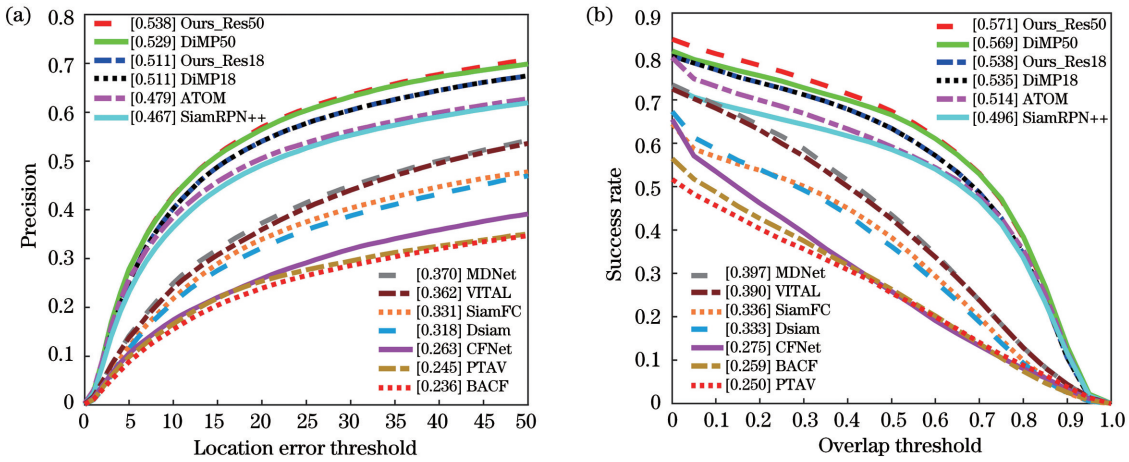

图 2　LaSOT 数据集上各算法的距离精度和成功率曲线。(a)精度;(b)成功率

Fig. 2　Distance precision and success rate of the algorithms on LaSOT dataset. (a) Precision; (b) success rate

提算法、ECO[19]、DiMP[14]、SiamRPN++[28]、DaSiamRPN[16]、SiamRPN[10]进行对比,对比结果如图 3 所示。可以看出:Ours_Res50 精度和成功率值分别为 0.909 和 0.698,精度与 ECO、SiamRPN++相等,但是 Ours_Res50 的成功率值高于这两种算法;Ours_Res18 精度和成功率值分别为 0.885 和 0.676;与原 DiMP 算法相比,Ours_Res18 与 Ours_Res50 均有一定的提升。

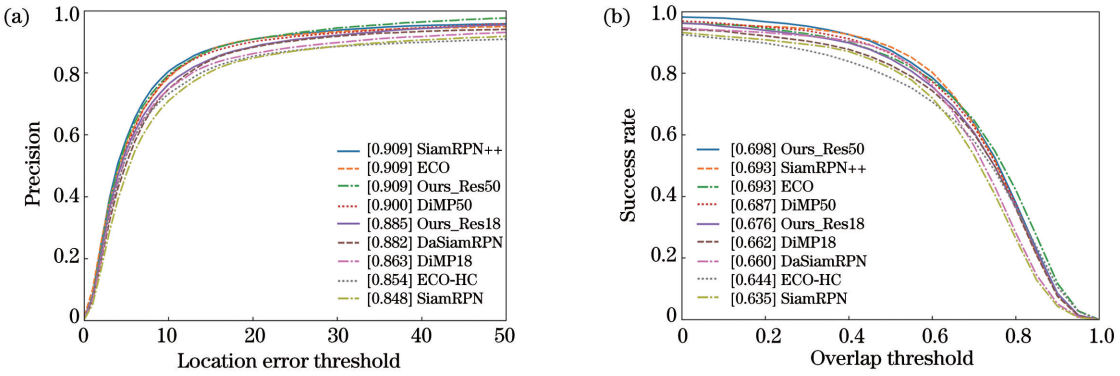

图 3　OTB100 数据集上各算法的距离精度和成功率。(a)精度;(b)成功率

Fig. 3　Distance precision and success rate of the algorithms on OTB100 dataset. (a) Precision; (b) success rate

UAV123:该数据集包含由无人机低空拍摄的 123 个视频,包含了各种俯视跟踪的挑战。对所提算法、DiMP[14]、ATOM[18]、DaSiamRPN[16]、SiamRPN[10]、ECO[19]进行对比,对比结果如图 4 所示。可以看出:Ours_Res18 的精度和成功率值分别为 0.834 和 0.649,与原 DiMP18 相比,精度低 0.2 个百分点,但成功率高 0.2 个百分点;Ours_Res50 的精度和成功率值分别为 0.849 和 0.662,均

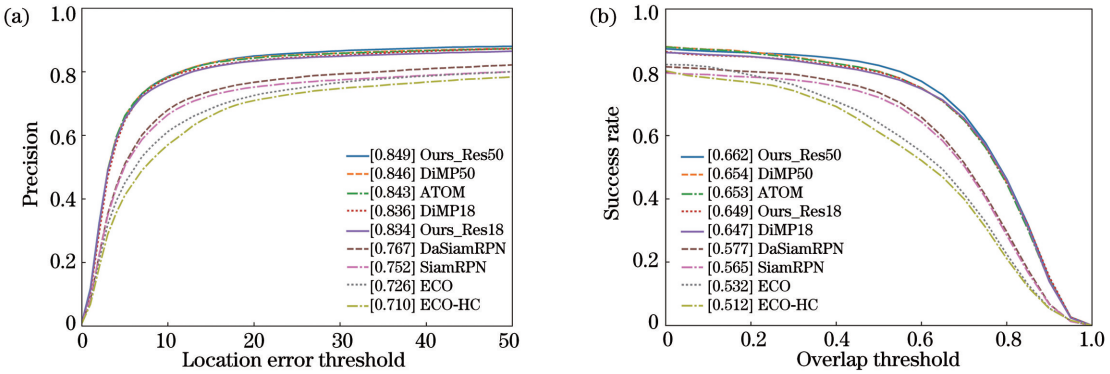

图 4　UAV123 数据集上各算法的距离精度和成功率。(a)精度;(b)成功率

Fig. 4　Distance precision and success rate of the algorithms on UAV123 dataset. (a) Precision; (b) success rate

优于对比算法。

5 结 论

提出一种融合差异化同质性模型的视觉跟踪算法。所提算法对视觉跟踪模型与目标分割模型进行融合，它们相互弥补存在的不足；然后将跟踪目标划分为小目标与常规目标，根据目标大小划分不同的搜索区域；再在长条形目标中融入上下文信息，提升目标特征的表征能力；最后根据跟踪的可靠性，设计多学习率更新跟踪模型，提升跟踪的鲁棒性。在 VOT、LaSOT、OTB100、UAV123 四种数据集上对不同算法进行比较。结果表明，所提算法的 ResNet50 版本在 VOT2018 数据集上的 EAO 值为 0.451，在 VOT2019 数据集上的 EAO 值为 0.414，在 LaSOT 数据集上的成功率值为 0.571，在 OTB100 数据集上的成功率值为 0.698，在 UAV123 数据集上的成功率值为 0.662，与原 DiMP50 算法相比，均有一定提升。所提算法优于对比的跟踪算法，跟踪速度为 31 frame/s，满足实时要求，在遇到遮挡、形变等各种干扰时可以稳定跟踪目标。

参 考 文 献

[1] 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, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 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] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 1090-1097.

[4] Ma C, Huang J B, Yang X K, et al. Hierarchical convolutional features for visual tracking[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 3074-3082.

[5] 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 Press, 2016: 1401-1409.

[6] Lukežic A, Vojír T, Zajc L C, et al. Discriminative correlation filter with channel and spatial reliability [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4847-4856.

[7] Xiong C Z, Che M Q, Wang R L, et al. Robust real-time visual tracking via dual model adaptive switching [J]. Acta Optica Sinica, 2018, 38(10): 1015002.
熊昌镇, 车满强, 王润玲, 等. 稳健的双模型自适应切换实时跟踪算法[J]. 光学学报, 2018, 38(10): 1015002.

[8] 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.

[9] 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 Press, 2018: 4854-4863.

[10] Li B, Yan J J, Wu W, et al. High performance visual tracking with siamese region proposal network [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8971-8980.

[11] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[12] Wang Q, Gao J, Xing J L, et al. DCFNet: discriminant correlation filters network for visual tracking[EB/OL]. [2020-05-03]. https://arxiv.org/pdf/1704.04057.pdf.

[13] Zhang M, Wang Q, Xing J L, et al. Visual tracking via spatially aligned correlation filters network[EB/OL]. [2020-05-03]. https://www.researchgate.net/publication/333660310_Visual_tracking_via_spatially_aligned_correlation_filters_network.

[14] Bhat G, Danelljan M, van Gool L, et al. Learning discriminative model prediction for tracking[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6181-6190.

[15] Guo Q, Feng W, Zhou C, et al. Learning dynamic Siamese network for visual object tracking[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 1781-1789.

[16] Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11213: 103-119.

[17] Wang Q, Zhang L, Bertinetto L, et al. Fast online object tracking and segmentation: a unifying approach [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1328-1338.

[18] Danelljan M, Bhat G, Khan F S, et al. ATOM: accurate tracking by overlap maximization[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 4655-4664.

[19] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6931-6939.

[20] XiongC Z, Lu Y, Yan J Q. Weighted correlation filter tracking algorithm based on context and relocation[J]. Acta Optica Sinica, 2019, 39(4): 0415004.
熊昌镇, 卢颜, 闫佳庆. 融合上下文和重定位的加权相关滤波跟踪算法[J]. 光学学报, 2019, 39(4): 0415004.

[21] Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4800-4808.

[22] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.

[23] Fan H, Lin L T, Yang F, et al. LaSOT: a high-quality benchmark for large-scale single object tracking [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 5369-5378.

[24] Mueller M, Smit N, Ghanem B. A benchmark and simulator for UAV tracking[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 445-461.

[25] Kristan M, Leonardis, Matas J, et al. The sixth visual object tracking VOT2018 challenge results [M] // Laura L T, Stefan R. Computer vision-ECCV 2018 workshops. Lecture notes in computer science. Amsterdam: Springer, 2018, 11129: 3-53.

[26] Lukezic A, Kart U, Käpylä J, et al. CDTB: a color and depth visual object tracking dataset and benchmark [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 10012-10021.

[27] Xu T Y, Feng Z H, Wu X J, et al. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking[J]. IEEE Transactions on Image Processing, 2019, 28(11): 5596-5609.

[28] Li B, Wu W, Wang Q, et al. SiamRPN++: evolution of Siamese visual tracking with very deep networks [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 4277-4286.

[29] 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 Press, 2016: 4293-4302.

[30] Song Y B, Ma C, Wu X H, et al. VITAL: VIsual tracking via adversarial learning [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT. New York: IEEE Press, 2018: 8990-8999.

[31] 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), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5000-5008.

[32] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 1144-1152.

[33] Fan H, Ling H B. Parallel tracking and verifying[J]. IEEE Transactions on Image Processing, 2019, 28(8): 4130-4144.