复杂场景下基于自适应特征融合的目标跟踪算法

李 博[1,2,3,4], 张心宇[1,2,3]

(1. 中国科学院光电信息处理重点实验室，辽宁 沈阳 110016；
2. 中国科学院沈阳自动化研究所，辽宁 沈阳 110169；
3. 中国科学院机器人与智能制造创新研究院，辽宁 沈阳 110169；
4. 中国科学院大学，北京 100049)

摘 要：为提升复杂场景下目标跟踪的鲁棒性,优化模型运行效率,提出一种基于自适应特征融合的相关滤波跟踪算法。该算法采用方向梯度直方图特征和卷积神经网络来对目标进行信息构建,利用特征响应的峰值旁瓣比和旁瓣值占比自适应地确定融合系数,根据融合响应来预测目标位置。为适应场景的变化,降低光照、背景和目标形变等对跟踪的影响,引入平均峰值相关能量来设计滤波器学习率调整机制,动态地进行模型更新。通过对深度特征提取网络进行轻量化设计,降低特征网络参数,提高跟踪速度。在 OTB100 通用数据集上进行测试,实验结果表明:文中所提算法有效降低了干扰对目标跟踪的影响,且跟踪精度、成功率和速度整体优于对比算法。

关键词：目标跟踪； 融合响应； 学习率调整； 轻量化

中图分类号：TP391 **文献标志码：**A **DOI：**10.3788/IRLA20220013

Target tracking algorithm based on adaptive feature fusion in complex scenes

Li Bo[1,2,3,4], Zhang Xinyu[1,2,3]

(1. Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China;
2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, China;
3. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences , Shenyang 110169, China;
4. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: In order to improve the robustness of the target tracking algorithm in complex scenes and optimize the operating efficiency of the model, a correlation filter tracking algorithm based on adaptive feature fusion was proposed. The algorithm adopts histogram of oriented gradient feature and deep feature extraction network feature to construct the target information, uses the peak to sidelobe ratio and the value of side lobe ratio of feature response to adaptively determine the fusion coefficient, and predicts the target position according to the fusion response. In order to reduce the influence of illumination variation, occlusion and target deformation on the tracking process and adapt to the change of scene, the average peak-to correlation energy was introduced to design the filter learning rate adjustment mechanism and update the model dynamically. Through the lightweight design of the deep feature extraction network, the parameters of the feature network were reduced and the tracking speed was improved. Experimental results show that the algorithm effectively reduces the influence of

收稿日期:2022−02−20； **修订日期:**2022−03−15

基金项目:装备预研重点基金项目 (JZX7 Y2019025049301)

作者简介:李博, 男, 硕士生, 主要研究方向为目标跟踪。

导师简介:张心宇, 男, 正高级实验师, 硕士生导师, 主要研究方向为制导与仿真。

interference on the tracking results, and the algorithm has better performance in tracking precision, success rate and speed compared with other tracking algorithm on the public video dataset OTB100.

Key words: target tracking; fusion response; learning rate regulation; lightweight

0 引 言

目标跟踪是计算机视觉领域的重点研究课题[1−2], 跟踪算法不仅要能够对目标进行准确的定位, 还要适应各种干扰的影响, 满足稳健性的要求。相关滤波具有速度快、效果好等优势, 已成为目标跟踪领域一大分支, 该方法通过跟踪过程的滤波响应来预测目标位置[3]。

解决相关滤波算法中跟踪器训练样本不足的问题, Henriques 等[4] 提出了经典的核相关滤波算法 (Kernel Correlation Filter, KCF), 通过循环移位补充训练样本集, 并采用多通道的方向梯度直方图特征 (HOG)[5] 对目标进行特征表达, 同时利用高斯核函数将特征映射到高维空间, 提高运算速度。针对循环采样引起的边界效应问题, Danelljan 等[6] 提出了空间正则相关滤波 (Spatially Regularized Discriminative Correlation Filters, SRDCF) 算法, 通过引入空间正则惩罚系数来抑制周围区域, 保证滤波器系数主要集中在样本中心位置, 大大降低了边界效应带来的影响。

为了准确构建目标的有效信息, 需要采用合理的特征选择和有效的融合方法。Bertinetto 等[7] 提出了 Staple(Sum of Template And Pixel-wise Learners) 算法, 将 HOG 特征与颜色特征 (CN)[8] 线性加权融合, 实现特征间的互补。房胜男等[9] 引入对冲算法自适应的调节特征响应权值, 将各响应以加权方式融合。尹宽等[10] 采用不同线性组合方式对特征进行融合, 选择置信度最高的融合特征进行目标位置预测。

随着深度学习方法在计算机视觉领域的广泛应用, 研究人员开始将其引入目标跟踪领域。神经网络具有优秀的特征提取能力, 将其与相关滤波算法结合能够显著提升算法的跟踪性能。HCF (Hierarchical Convolutional Features) 算法[11] 在 KCF算法的基础上, 采用深度特征进行特征表达, 将得到的响应图加权融合, 充分利用了图像信息, 大大提高了跟踪器的判别能力。Danelljan 等[12] 提出基于连续卷积算子的相关滤波跟踪算法 (Continuous Convolution Operators for Tracking , C-COT), 利用卷积网络进行特征提取, 通过连续空间域插值运算将离散的特征图转化到连续空间域中, 采用不同层次的深度特征训练跟踪器。Martin Danelljan 等[13] 提出了一种具有高效卷积特性的跟踪算法 (Efficient Convolution Operators, ECO), 在 C-COT 的基础上, 从模型大小、训练集大小以及更新策略三个方面进行优化改进, 取得了更加优秀的跟踪结果。Qi 等[14] 提出的 HDT(Hedged Deep Tracking) 算法使用多层卷积层特征, 对各特征图分别训练一个弱跟踪器, 利用对冲算法将所有弱跟踪器结合, 得到一个强跟踪器。

尽管以上基于相关滤波的目标跟踪算法已经取得了出色的跟踪效果, 但当受到目标遮挡、运动模糊等干扰时, 容易出现不同程度的漂移。而对于采用了深度特征的跟踪算法, 跟踪模型的参数量较大, 跟踪速度受限。文中在 ECO 算法的基础上, 提出一种复杂场景下基于自适应特征融合的相关滤波跟踪算法。该算法采用多层深度特征和 HOG 特征对目标进行表征, 并通过自适应地对各特征进行融合, 以及对跟踪过程中滤波器学习率进行动态调整, 以适应目标及场景的变化。为降低模型的复杂度和参数量, 文中采用通道注意力机制和逆残差模块, 设计了一种轻量级神经网络进行特征提取, 实现了对目标的精确、稳定跟踪。

1 相关滤波目标跟踪原理

1.1 滤波器训练及更新

相关滤波算法的原理是对输入图像进行特征提取, 初始化相关滤波器, 将图像特征与滤波器进行卷积运算, 得到响应分布, 响应的峰值位置就是预测的目标中心, 训练及更新如下:

若输入图像为f, 对应的滤波器为h, 对两者做相关操作可以表示为:

$$g = f \otimes h \tag{1}$$

将训练样本与其对应标签的误差平方和作为损失函数, 可表示为:

$$\delta = \sum_{j=1}^{t} \left\| h * f_j - g_j \right\|^2 = \frac{1}{MN} \sum_{j=1}^{t} \left\| H^* \odot F_j - G_j \right\|^2 \tag{2}$$

为将整体损失降为最低, 对H求偏导数可得:

$$0=\frac{\partial}{\partial H^*}\sum_{j=1}^{t}\left|H^*\odot F_j-G_j\right|^2 \tag{3}$$

求解公式 (3) 可得滤波器的计算公式:

$$H^*=\frac{\sum_{j=1}^{t}F_j\odot G_j^*}{\sum_{j=1}^{t}F_j\odot F_j^*} \tag{4}$$

由于在跟踪过程中会遇到目标及场景的变化, 所以需要对滤波器进行更新, 可表示为:

$$H_j^*=\frac{A_j}{B_j} \tag{5}$$

$$A_j=\eta G_j\odot F_j^*+(1-\eta)A_{j-1} \tag{6}$$

$$B_j=\eta F_j\odot F_j^*+(1-\eta)B_{j-1} \tag{7}$$

1.2 高效卷积操作

为提高位置估计的精确度, 引入连续卷积算子。在训练样本中引入一个插值算子, 将离散特征图通过三次样条插值转化为连续特征图。假设有M个训练样本$x=\{x_1,x_2,...,x_j,...,x_M\}$, 第$j$个样本$x_j$包含 D 个特征通道$\{x_j^1,x_j^2,...,x_j^d,...,x_j^D\}$, 其中每个通道$x_j^d\in R^{N_d}$对应的特征图有$N$种分辨率$N_d$。定义函数$J_d$来整合不同分辨率的特征图:

$$J_d\left\{x^d\right\}(t)=\sum_{n=0}^{N_d-1}x^d[n]\,b_d\left(t-\frac{T}{N_d}n\right) \tag{8}$$

式中: b_d为周期$T>0$的插值函数, 这里用的是三次样条插值; $J_d\{x^d\}$由b_d的各个平移叠加构造而来。

最终的置信函数$S_f\{x_h\}$通过寻找最大置信分数来估计下一帧目标的位置, 它通过一个连续的多通道卷积滤波器$f=\{f^1,f^2,...,f^d,...,f^D\}$与插值后的各特征通道卷积而得, 可表示为:

$$S_f\{x\}=f\times J\{x\}=\sum_{d=1}^{D}f^d\times J_d\{x^d\} \tag{9}$$

式中: $J\{x\}$表示整个插值特征图。

各个滤波器在跟踪过程中的作用不同, 为降低算法的复杂度, 假设其中C个特征对图像的表达起到了关键作用, 定义一个$D\times C$的矩阵P, 对其进行降维操作, 可表示为:

$$J_d\{f\}=P^{\mathrm{T}}J_d\left\{f^d\right\} \tag{10}$$

此外, 为减少训练过程的过拟合, 采用高斯混合模型建模, 对样本进行多样化处理, 增加样本间差异。

2 复杂场景下自适应特征融合的跟踪算法

文中基于 ECO 算法的框架进行改进, 提取目标的 HOG 特征和浅、中、深三层深度特征, 通过自适应特征融合机制, 根据不同特征的响应情况得到最佳融合结果, 完成目标位置的预测。

设计学习率动态调整策略, 根据当前跟踪状态, 实时地对跟踪器进行自适应更新, 减少跟踪过程中的模型漂移; 改进深度特征提取网络, 大大降低了网络的参数量和计算量, 提高了跟踪速度。算法框架见图 1。

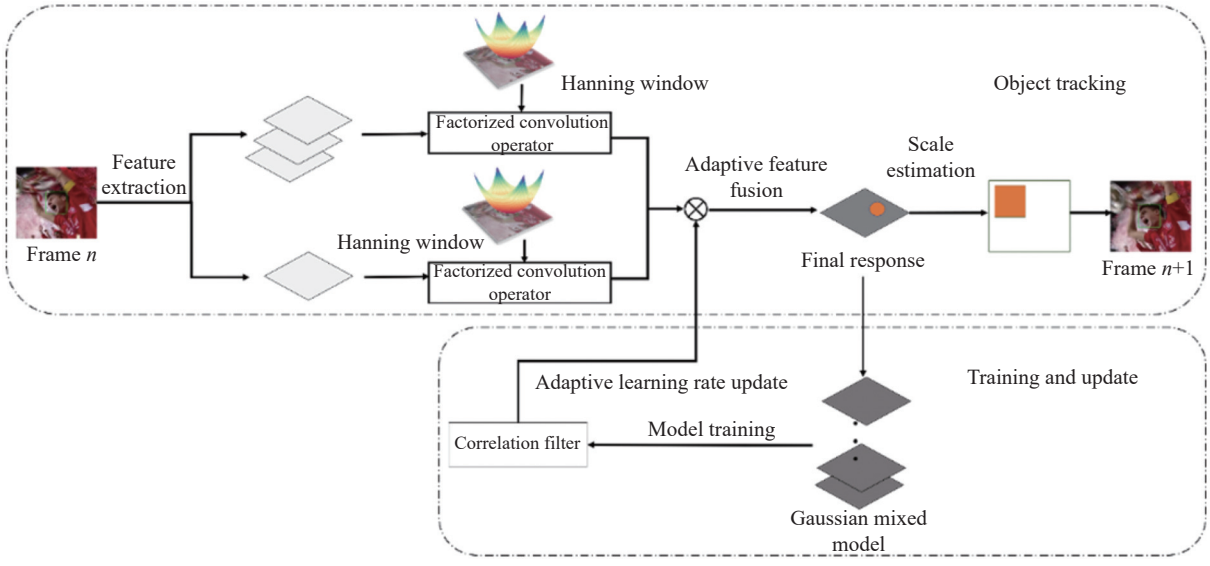


图 1 文中算法框架

Fig.1 Framework of proposed algorithm

2.1 特征选择

不同的特征包含目标不同维度的信息，合理地特征选择对目标跟踪的结果有着重要的影响。HOG 特征通过计算图像的方向梯度直方图来提取目标的多通道梯度信息，其反映了图像局部区域内像素之间的关联性，具有良好的几何和光学不变性。

深度特征通过卷积网络进行提取，不同深度网络层特征包含的图像信息并不相同。浅层深度特征具有较高的分辨率，主要构建目标的纹理和边缘等细节信息，注重跟踪的精度。随着层数的加深，深度特征的分辨率会逐层降低，中、深层深度特征主要用来构建目标的语义信息，其不易受到目标尺度和形状变化的干扰，具有较好的鲁棒性。

深度特征和 HOG 特征在不同的场景下具有不同的表达能力，因此对跟踪结果的预测有一定差异。深度特征和 HOG 特征能够从不同角度来对目标进行特征提取，但是单一特征的信息表达具有局限性，在复杂场景下目标及背景的特性发生变化时单一特征不能很好地描述，影响跟踪的结果。因此，文中提取目标的 HOG 特征和浅、中、深三层深度特征并融合，以丰富目标的特征信息。

2.2 自适应多特征融合

为尽可能利用多特征信息的互补性，文中采用一种自适应特征融合方法，根据当前跟踪场景及目标的情况自适应地进行特征融合，动态地为不同响应分配权值，最终融合得到更准确地滤波响应。定义R_i为第i个特征的滤波响应，最终的滤波响根据各响应的权值w_i加权得到，可表示为：

$$R_{final}=\sum_{i=1}^{N}w_iR_i \tag{11}$$

峰值旁瓣比PSR可以表示一个滤波响应的尖锐度，可以衡量跟踪的置信情况，反映了对目标和背景的区分程度，其可表示为：

$$PSR=\frac{R_{\max,i}-\mu_i}{\sigma_i} \tag{12}$$

式中：$R_{\max,i}$、μ_i和σ_i分别代表滤波响应的最大值、均值和标准差。

PSR值越大，代表该特征下跟踪器对目标与背景区分的越好，置信度越高。但是由于涉及到多特征融合，为了保证融合之后的准确性，设计了旁瓣值占比SLR作为融合置信度评价指标。

旁瓣值占比SLR计算当前响应图中低于峰值一定阈值的响应值的比例，它反映特征融合之后结果的置信程度，可表示为：

$$SLR=\frac{sum(R_i<\alpha R_{\max,i})}{m\times n} \tag{13}$$

式中：R_i为特征i的响应；α为用来调节响应阈值的系数；m,n各为特征图的大小。SLR反映了多特征融合操作对目标定位的影响程度，SLR越大，表示融合之后响应图的置信程度越高。

由此，结合PSR和SLR这两个置信度指标，得到特征融合权值的方法：

$$w_i=\frac{PSR_i\times SLR_i}{\sum_j PSR_j\times SLR_j} \tag{14}$$

式中：w_i为特征响应i的融合权重。

上述特征融合方法考虑到各特征在不同场景和时刻表达能力的差异，合理地加大了表达能力更好特征的作用，充分利用到目标有限的信息，并且在多特征融合的过程中抑制较强的干扰对融合后响应图峰值的影响，突出了目标真实响应。相较固定权值融合的方法，上述方法针对不同滤波响应可以根据场景及目标变化自适应调节融合权值，得到更可靠的融合响应，对跟踪过程中的常见噪声干扰更具鲁棒性，能够提高对目标位置估计的精度，更适合复杂场景的跟踪。

2.3 动态学习率调整

考虑到跟踪在连续视频帧内会出现一系列的变化，包括目标的表观变化和场景的变化，这是一个渐变的过程。而在跟踪过程中，当前帧的跟踪结果会作为样本在后续帧中对模型进行训练更新，当出现与目标相似度较高的干扰区域时，跟踪质量下降，模型会学习到错误信息，随着每一帧跟踪误差的积累，导致跟踪漂移的现象。因此文中在跟踪过程中对跟踪器的学习率进行动态调整，以适应跟踪场景的变化。

引入平均峰值相关能量 $APCE$ 作为干扰判断的指标。当目标出现背景杂乱等干扰时，响应图会出现多个峰值，$APCE$ 也会随之变化，其定义可表示为：

$$APCE=\frac{|F_{\max}-F_{\min}|^2}{mean\left(\sum_{w,h}(F_{w,h}-F_{\min})^2\right)} \tag{15}$$

式中：F_{max}、F_{min}和$F_{w,h}$分别为响应最大值、最小值和对应位置处的响应值。

文中设计了新的学习率更新机制，结合*APCE*标准值和历史平均值，具体更新公式如下：

$$\varepsilon=\lambda\frac{APCE_t}{APCE_0}+(1-\lambda)\frac{APCE_t}{\text{mean}\left(\sum_{i=1}^{t}APCE_i\right)} \tag{16}$$

$$lr_t=lr\times\varepsilon \tag{17}$$

式中：ε为学习率调节系数；$APCE_t$为当前帧的平均峰值相关能量；$APCE_0$为设定的标准值；lr_t为当前帧学习率；lr为初始学习率。

当$APCE_t$值较小时，说明滤波响应出现了较大的波动，目标受到了干扰的影响，此时学习率调节系数ε减小，可以降低学习率，以减少错误信息的学习。其中权重调节系数λ可以改变标准值和历史平均值的贡献，若λ>0.5，则标准值在学习率更新中占主导地位，反之则表明历史多帧平均值占主导地位。

2.4 特征提取网络轻量化改进

2.4.1 通道注意力机制

通道注意力机制[15]通过学习各个通道的权重，使网络模型对不同通道的作用具有区分能力，以提升网络性能。通道注意力机制的本质是让模型多关注信息量更大、重要性更强的通道特征，抑制不重要的通道特征，如图 2 所示。

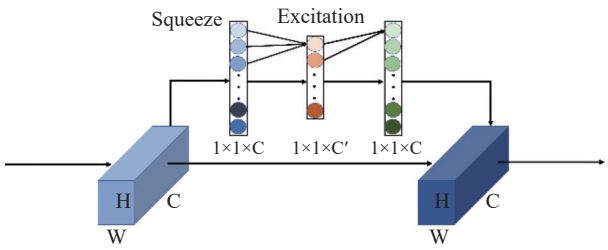


图 2 通道注意力机制模块

Fig.2 Channel attention module

该模块通过全连接层建立各通道之间的关系，起到通道交互的作用。首先对输入特征图进行 Squeeze 操作，经全局池化，得到全局特征；之后进行 Excitation 操作，这里采用两层全连接层，来学习各个通道间的关系，也即各通道所占权重，将其与原特征图相乘，即可得到最终输出的特征图。

2.4.2 轻量级网络改进

为降低网络参数量，提高跟踪速度，需要对提取深度特征的神经网络进行轻量化改进。在卷积神经网络中，卷积层的参数占比较大，其主要影响着网络的推理速度和模型大小，因此对网络的改进主要在卷积单元的优化上。

深度可分离卷积可以有效降低卷积操作的参数量，其由深度卷积和点卷积两部分组成。深度卷积操作在二维平面上进行，对输入层的每个通道独立进行卷积运算；点卷积将深度卷积的结果在通道维度上进行加权相加，综合各通道的特征，形成新的特征图。

一次常规卷积的参数量P_c可表示为：

$$P_c=k\times k\times m\times n \tag{18}$$

一次深度可分离卷积的参数量P_d为深度卷积和点卷积的参数量之和，可表示为：

$$P_d=k\times k\times m+m\times n \tag{19}$$

式中：k 为卷积核的尺寸大小；m、n 为输入和输出通道数。

卷积核的尺寸设置为3×3，深度可分离卷积和常规卷积的参数量的比例ρ可表示为：

$$\rho=\frac{P_d}{P_c}=\frac{k\times k\times m+m\times n}{k\times k\times m\times n}=\frac{1}{n}+\frac{1}{9} \tag{20}$$

分析可知，理论上深度可分离卷积可以将参数量降低至常规卷积的1/9左右。

逆残差模块 SandGlass[16]中采用深度可分离卷积思想设计，对输入先进行深度卷积，对各通道内的信息进行交互，再利用一次点卷积对通道压缩，降低计算量的同时减少特征图的冗余。之后利用点卷积将输出通道数提升至预期数量，最后接一层深度卷积后输出。

改进网络基于通道注意力机制和逆残差模块 SandGlass[16]，改进后网络基本模块如图 3 所示。

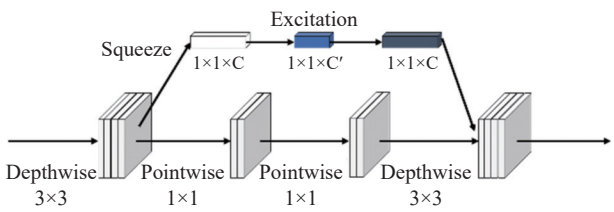


图 3 轻量级网络模块

Fig.3 Lightweight network module

改进网络结合通道注意力机制和逆残差模块，对目标特征进行优化，突出重要特征信息，通过计算不同通道特征的贡献程度对特征进行筛选，使得目标信息得到充分利用，同时卷积单元的优化实现了特征在低维度下的交互和传递，具有更少的参数量的同时保持了优越的特征表达能力。相较主流特征提取网络，改进网络具有更小的体积、更高的精度以及更低的计算代价，应用到复杂跟踪场景中，可以有效区分目标和场景信息，大大提高跟踪的性能和速度。

网络结构如表 1 所示。

表 1 改进后的网络结构

Tab.1 Improved network structure

Layer	Output size	Stride	Num	Input size
Conv2 *d* 3×3	112×112×32	2	1	224×224×3
Block1	112×112×96	1	1	112×112×32
Block2	56×56×144	2	3	112×112×96
Block3	28×28×192	2	3	56×56×144
Block4	14×14×288	2	4	28×28×192
Block5	14×14×384	2	4	14×14×288
Block6	7×7×576	2	3	14×14×384
Block7	7×7×1280	1	2	7×7×576
Avgpooling	1×1×1280	-	1	7×7×1 280
Conv2 *d* 1×1	*k*	-	1	1×1×1 280

3 实验结果分析

3.1 实验环境

文中实验硬件平台配置如下：CPU 为 i7-7800 X，内存为 32 GB，GPU 为 NVIDIA GeForce RTX 2080 TI。跟踪器初始学习率为 0.01，学习率权重调节系数 λ 为 0.5，空间正则化系数为 1.2，SLR 阈值调节系数 α 为 0.2，尺度估计的个数为 5，尺度变化步长为 1.02。

文中在 OTB100 数据集[17] 上测试跟踪算法的性能，该数据集共 100 组视频序列，涵盖 11 种目标跟踪常见的干扰挑战，包括光照变化 (IV)、目标形变 (DEF)、尺度变化 (SV)、遮挡 (OCC)、运动模糊 (MB)、快速运动 (FM)、平面旋转 (IPR)、非平面旋转 (OPR)、目标消失 (OV)、背景杂乱 (BC) 和低分辨率 (LR) 等，以一次通过评估 OPE 作为算法性能的评价标准。

采用 ImageNet2012 数据集[18] 对神经网络进行训练，该数据集包含 1 000 类对象，分为训练集、验证集和测试集，三者之间没有重叠。训练集共有 128 116 7 张训练图像，每个类别的训练图像数量从 732~1 300 不等；验证集共有 50 000 张验证图像，每个类别有 50 个验证图像；测试集共有 100 000 张测试图像，每个类别有 100 个测试图像。

跟踪算法的性能采用精确度 (precision) 和成功率 (success rate) 这两个指标进行评价。其中精确度为视频帧中跟踪结果与真实位置之间的欧氏距离小于一定阈值的百分比，成功率为预测目标框与真实的目标框之间的重合率大于某一阈值的帧数占总视频帧数的百分比。

3.2 消融实验

评价文中改进策略对跟踪结果的影响，在 ECO 算法的基础上，分别做了以下几组实验，包括：实验 1，添加自适应特征融合机制 (ECO-feature 表示)；实验 2，添加动态学习率调整机制 (ECO-learning_rate 表示)；实验 3，同时添加自适应特征融合机制和动态学习率调整机制 (ECO-feature+learning_rate表示)；实验 4，在实验 3 的基础上，将特征提取网络换成改进后的轻量级网络，即文中所提算法 (Proposed 表示)。实验结果如表 2 所示。

跟踪过程中对目标的估计误差包含两方面，定位误差和尺度误差，分别体现在精确度和成功率上。精确度反映了特征响应与真实响应的差异，两者越接近，定位误差越小。成功率在定位误差的基础上，添

加了尺度误差的信息，更注重预测跟踪框的重叠率。在保证定位精度的基础上，尺度估计越准确，成功率越高。

表 2 各改进策略下的跟踪性能对比

Tab.2 Comparison of tracking performance under each improvement strategy

Methods	Precision	Success rate	Network params
Proposed	92.53	67.67	3.7×10^6
ECO-feature+learning_rate	92.30	67.15	138.4×10^6
ECO-feature	92.14	67.05	138.4×10^6
ECO-learning_rate	90.91	66.95	138.4×10^6
ECO	89.90	66.02	138.4×10^6

从表 2 可以看出，各改进策略对跟踪精确度和成功率均有很大提升，大大降低了跟踪过程中的误差。跟踪误差产生的原因包括特征表达的不充分以及干扰的影响，加入自适应特征融合机制的 ECO-feature 算法构建了合理的目标特征信息，并且对多特征进行自适应的在线融合，调节不同特征在跟踪过程中的贡献程度，增强了对目标的辨别能力，这使得特征响应更接近真实响应，降低了对目标位置估计和尺度估计的误差，相较 ECO 算法跟踪性能有了较大提升；动态学习率调整机制可以避免跟踪过程中误差的累积，降低对跟踪质量低的样本信息的学习，在样本更新和抑制干扰之间达到了平衡，实验中 ECO-learning_rate 算法的精确度和成功率均有明显提升。同时加入自适应特征融合和动态学习率调整机制对算法跟踪性能的提升，要高于每一个改进策略单独添加时所带来的性能提升，ECO-feature+learning_rate 算法相较 ECO 算法的精确度提升了 2.40%，成功率提升了 1.13%，说明各改进策略之间存在互补和促进作用。

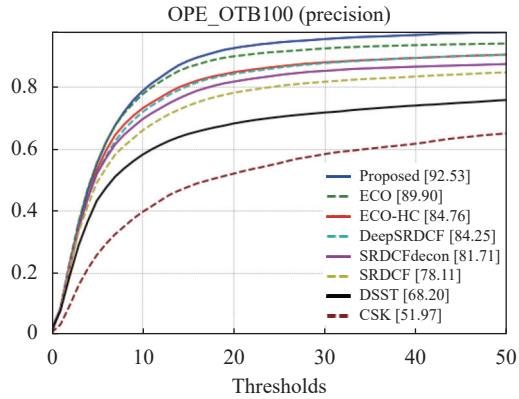
ECO 算法采用 VGG16 网络来对目标进行深度特征提取，参数量为138.4×10^6，导致跟踪模型较为复杂。文中所提 Proposed 算法将 VGG16 网络替换为改进的轻量级网络，参数量仅3.7×10^6，大幅降低了参数量。同时 Proposed 算法的跟踪性能并未因网络参数量的降低而下降，精确度和成功率相较 ECO-feature+learning_rate 算法分别提升了 0.23% 和 0.52%，相较 ECO 算法分别提升了 2.63% 和 1.65%，说明改进轻量级网络对目标的特征提取能力出色，提取的深度特征能够加强对目标的鉴别能力，更适用于跟踪任务。

3.3 整体性能综合对比

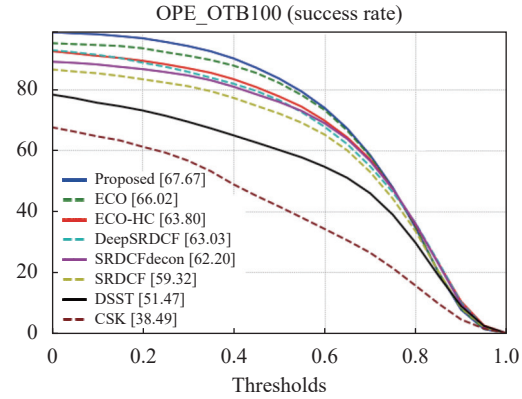
为了对比各跟踪算法的整体性能，将 ECO、ECO-HC[13]、DeepSRDCF[19]、SRDCFdecon[20]、SRDCF、DSST[21]、CSK[22] 等算法作为对比算法，与文中所提 Proposed 算法在 OTB100 数据集上进行跟踪结果的性能对比。图 4 为不同算法的性能对比图。

从图 4 可知，文中算法跟踪的精确度为 92.53%，相较 ECO 算法提升了 2.63%，相较未使用深度特征的 ECO-HC 算法提升了 7.77%；成功率为 67.67%，相较 ECO 算法提升了 1.65%，相较 ECO-HC 算法提升了 3.87%。表 3 为各跟踪器性能和跟踪速度的综合对比。

由于提取深度特征的神经网络进行了轻量化的改进，文中算法跟踪速度为 15.1 FPS，同样基于深度特征的 ECO 算法跟踪速度为 7.1 FPS，文中跟踪速度



(a) 不同算法精确度对比曲线
(a) Tracking precision plots of different algorithms



(b) 不同算法成功率对比曲线
(b) Success plots of different algorithms

图 4 不同算法的跟踪性能对比

Fig.4 Comparison of tracking performance of different algorithms

为其 2.13 倍，跟踪速度上有较大的提升，大大降低了计算代价。

相较其他算法，文中算法在跟踪过程中表现稳定，通过加入自适应特征融合机制和动态学习率调整机制，替换更轻量级的深度特征提取网络，算法取得了更好的跟踪效果，在提高跟踪精确度和成功率的同时，还大大提升了跟踪的速度。

表 3 不同跟踪器综合性能对比

Tab.3 Comparison of comprehensive performance of different trackers

Results	Proposed	ECO	ECO-HC	DeepSRDCF	SRDCFdecon	SRDCF	DSST	CSK
Precision	92.53	89.90	84.76	84.25	81.71	78.11	68.20	51.97
Overlap	67.67	66.02	63.80	63.03	62.20	59.32	51.47	38.49
Mean FPS	15.1	7.1	61.3	1.1	3.7	4.4	50.4	346.9

3.4 不同干扰下的性能对比

针对 OTB100 数据集中包含的 11 种不同干扰场景，分别评估并对比各跟踪算法的性能，实验结果如表 4 所示，其中最优结果加粗处理。

表 4 不同干扰下各算法的精确度对比

Tab.4 Comparison of precision of each algorithm under different interference

Algorithms	BC	DEF	FM	IPR	IV	LR	MB	OCC	OPR	OV	SV
Proposed	**95.35**	**89.02**	**88.35**	**90.17**	**91.35**	**93.03**	**90.26**	**89.84**	**92.05**	**87.37**	**91.57**
ECO	86.46	84.93	87.20	87.86	87.46	79.22	86.60	89.61	90.06	83.20	89.53
ECO-HC	84.16	78.20	80.41	78.65	77.38	80.47	78.79	83.05	83.02	81.83	82.45
DeepSRDCF	83.24	75.78	78.89	79.24	74.05	70.22	80.61	80.72	82.11	78.06	82.20
SRDCFdecon	84.12	72.84	75.39	74.66	79.14	67.20	79.92	75.14	78.10	64.07	80.81
SRDCF	76.15	70.84	74.82	71.28	74.29	63.09	75.47	71.82	72.56	59.71	74.93
DSST	69.09	52.88	57.82	68.35	67.51	58.08	58.04	59.48	65.22	47.78	65.42
CSK	57.42	43.62	42.01	53.14	45.07	36.72	38.85	43.08	49.94	31.52	46.33

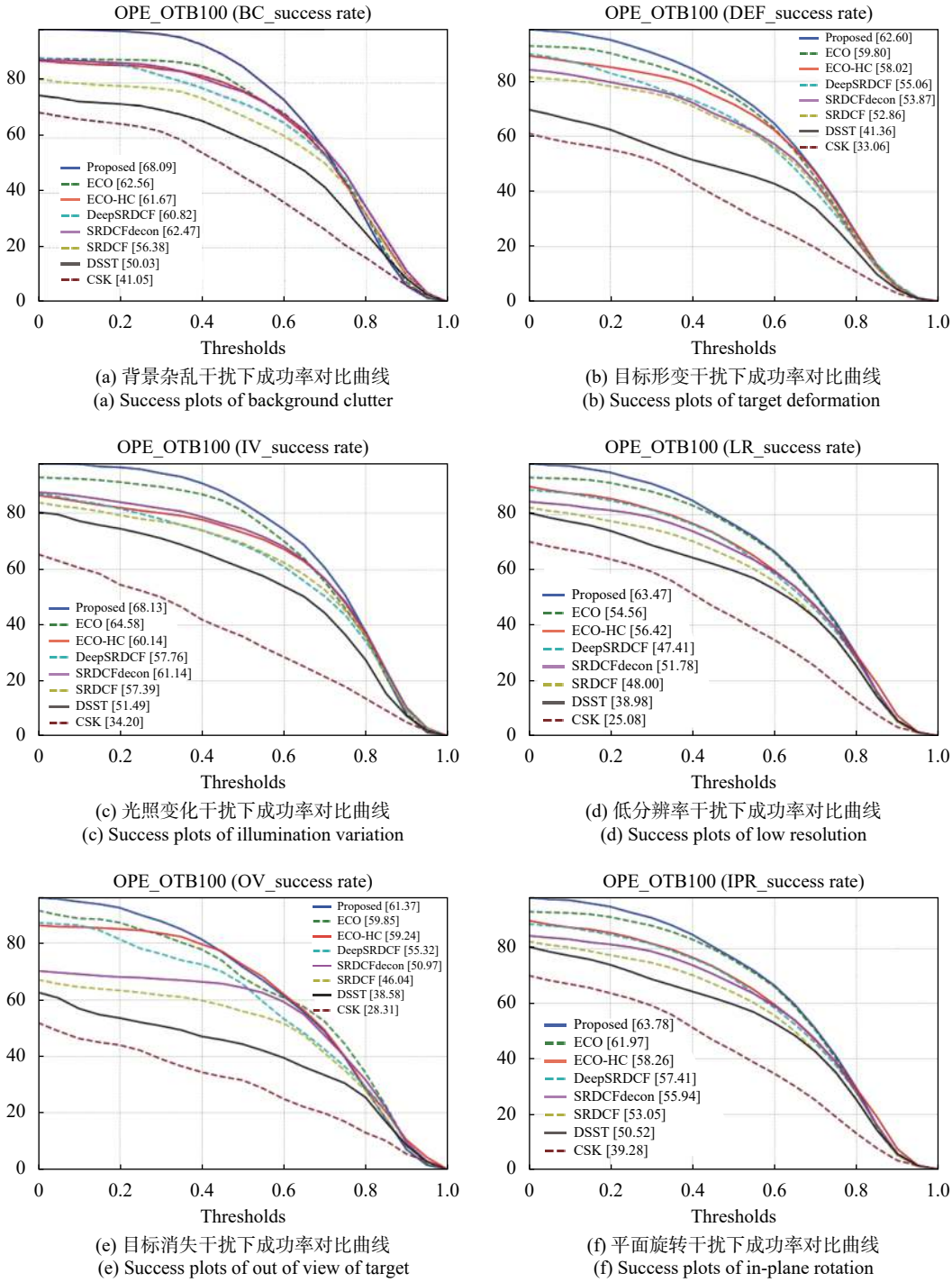
在各种干扰下，文中算法均有较优的跟踪性能，相较排名第二的算法，在低分辨率、背景杂乱、目标形变和光照变化等干扰下的精确度分别提升了 13.81%、6.33%、4.09% 和 3.89% 等。

表 4 结果表明，文中算法通过构建合理的目标特征，采用自适应特征融合机制融合多信息，在跟踪中即使出现了目标形变、背景杂乱等干扰，也能够自适应地调节融合权值，加强对目标的辨别能力。同时滤波器学习率动态更新机制可以找到跟踪结果不佳的视频帧并调节学习率，避免误差累积导致后续帧的跟踪出现性能下降甚至漂移，在视频帧的跟踪质量较好时，会主动提高学习率以恢复对模型及样本的更新，这使得跟踪算法在面对目标消失、遮挡等场景时，精确度也有了明显的提升。

相较其他算法，在常见干扰场景下文中算法同样具有更高的跟踪成功率。图 5 为部分干扰场景下各算法成功率对比。

为了更直观地分析文中算法的跟踪性能，选择具有代表性的 6 种跟踪算法和 6 组包含不同干扰挑战的视频序列进行定性评估，如图 6 所示。

在 Bolt2 序列中，视频分辨率较低，目标周围有许多相似干扰，SRDCF 等算法跟踪出现错误，文中算法采用深度特征和手工特征相结合，特征信息更为丰富，可以实现连续稳定跟踪。Box 序列存在部分遮挡和非平面旋转等干扰，文中算法结合了目标特征的语义信息，可以很好地适应这种场景，表现最好。当目标出现大面积遮挡，如 Girl2 序列，部分算法更新到错误的目标信息，从而导致跟踪产生漂移，文中算法加入的动态学习



(a) 背景杂乱干扰下成功率对比曲线
(a) Success plots of background clutter

(b) 目标形变干扰下成功率对比曲线
(b) Success plots of target deformation

(c) 光照变化干扰下成功率对比曲线
(c) Success plots of illumination variation

(d) 低分辨率干扰下成功率对比曲线
(d) Success plots of low resolution

(e) 目标消失干扰下成功率对比曲线
(e) Success plots of out of view of target

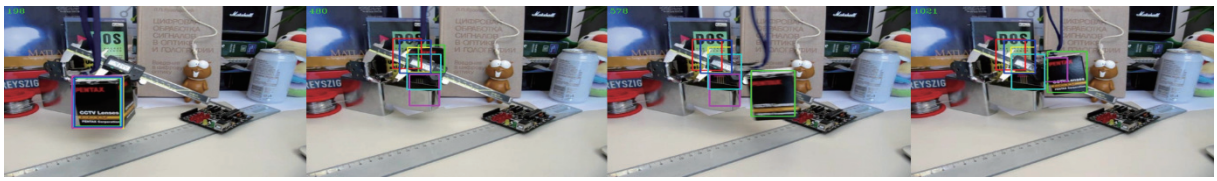
(f) 平面旋转干扰下成功率对比曲线
(f) Success plots of in-plane rotation

图 5 部分干扰场景下 OTB100 数据集上各算法成功率对比曲线

Fig.5 Success plots of partial interference scenes on OTB100 datasets

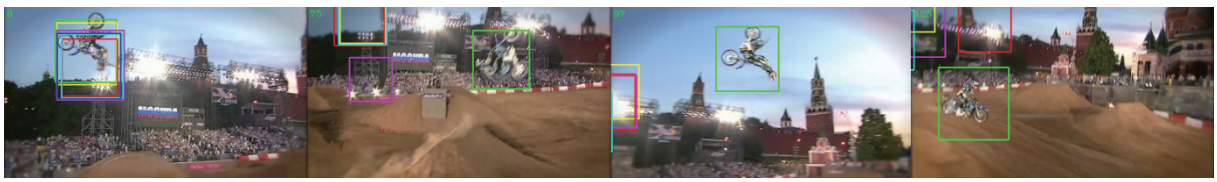
率更新, 能够及时调整模型更新速率, 在遮挡消失之后还能重新定位到目标, 抗干扰效果较好。在 MotorRolling 序列中, 目标受到快速平面旋转和光照变化等干扰, 文中算法将目标特征进行自适应地融合, 融合后特征响应可以更准确反映目标位置信息, 跟踪更为精确, 其他跟踪算法均丢失目标。Skiing序列中的目标尺度变化较大, 且存在快速运动, 文中算法同样具有较好的跟踪效果。Soccer 序列存在背景杂乱、目标被遮挡等干扰, 改进策略同样带来了优秀的跟踪效果, 文中算法在跟踪过程中一直保持精准的位置定位和尺度估计。

(a) Bolt2 视频序列
(a) The video sequence of Bolt2

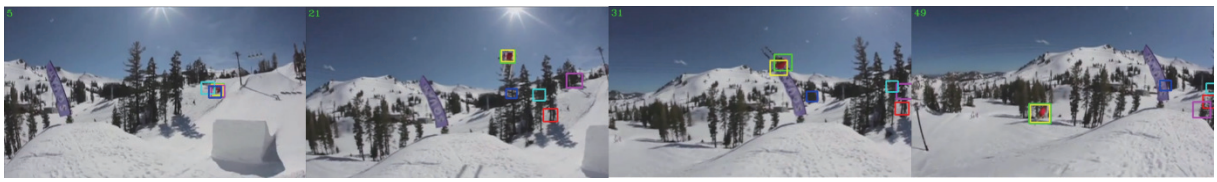


(b) Box 视频序列
(a) The video sequence of Box

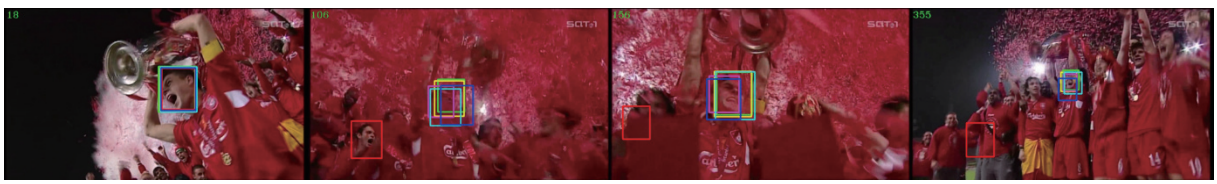
(c) Girl2 视频序列
(c) The video sequence of Gril2



(d) MotorRolling 视频序列
(d) The video sequence of MotorRolling



(e) Skiing 视频序列
(e) The video sequence of Skiing



(f) Soccer 视频序列
(f) The video sequence of Soccer

图 6 6 种跟踪算法对部分视频序列的定性评估结果

Fig.6 Qualitative evaluation results of 6 tracking algorithms on partial video sequences

4 结 论

针对目标跟踪算法在光照、背景杂乱等干扰时容易导致跟踪失败的问题，文中提出一种基于自适应特征融合的目标跟踪算法。通过 HOG 特征和深度特征对目标进行表征，考虑到不同特征对跟踪结果的贡献，采用自适应特征融合机制来完成目标的位置估

计;加入动态学习率调整机制,避免了误差的累计,有效降低了干扰对跟踪的影响,防止了模型漂移;轻量化改进了深度特征提取网络,降低了采用深度特征所带来的计算代价,在一定程度上提高了跟踪速度。在OTB100数据集上的实验结果表明,对于不同干扰和不同场景下的目标,文中算法具有更高的精度和成功率,整体跟踪性能优于对比算法。下一步工作将优化跟踪算法的运行效率,在保证跟踪性能的同时,进一步加快跟踪的速度。

参考文献:

[1] Yilmaz A, Javed O, Shah M. Object tracking: A survey [J]. ***Acm Computing Surveys***, 2006, 38(4): 13.

[2] Li Peixia, Wang Dong, Wang Lijun, et al. Deep visual tracking: Review and experimental comparison [J]. ***Pattern Recognition***, 2018, 76: 323-338.

[3] Bolmeds 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, 2010: 2544-2550.

[4] Henriques J F, Caseiro R, Martins P, et al. Highspeed tracking with kernelized correlation filters [J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2015, 37(3): 583-596.

[5] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005: 886-893.

[6] Danelljan M, Hager G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//2015 IEEE International Conference on Computer Vision, 2015: 4310-4318.

[7] 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, 2016: 1401-1409.

[8] 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. 2014: 1090-1097.

[9] Fang Shengnan, Gu Xiaojing, Gu Xingsheng. Infrared object tracking with correlation filtering based on adaptive response fusion [J]. ***Infrared and Laser Engineering***, 2019, 48(6): 0626003. (in Chinese)

[10] Yin Kuan, Li Junli, Li Li, et al. Adaptive feature update object tracking algorithm under complex conditions [J]. ***Acta Optica Sinica***, 2019, 39(11): 235-250. (in Chinese)

[11] Ma C, Huang J, Yang X, et al. Hierarchical convolutional features for visual tracking[C]//2015 IEEE International Conference on Computer Vision, 2015: 3074-3082.

[12] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation Filters: Learning continuous convolution operators for visual tracking [C]//European Conference on Computer Vision, 2016: 472-488.

[13] 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, 2017: 6931-6939.

[14] Qi Yuankai, Zhang Shengping, Qin Lei, et al. Hedged deep tracking [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4303-4311.

[15] Jie Hu, Li Shen, Samuel Albanie, et al. Squeeze-and-excitation networks [J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2017, 42(8): 2011-2023.

[16] Zhou Daquan, Hou Qibin, Chen Yunpeng, et al. Rethinking bottleneck structure for efficient mobile network design[C]// European Conference on Computer Vision, 2020: 680-697.

[17] Wu Yi, Lim Joowoo, Yang Ming-Hsuan. Object tracking benchmark [J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2015, 37(9): 1834-1848.

[18] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.

[19] Danelljan M, Hager G, Khan F S, et al. Convolutional features for correlation filter based visual tracking[C]//2015 IEEE International Conference on Computer Vision Workshop, 2016: 621-629.

[20] Danelljan M, Häger G, Khan F S, et al. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1430-1438.

[21] Danelljan M, Häger G, Khan F S, et al. Discriminative scale space tracking. [J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2017, 39(8): 1561-1575.

[22] Henriques J F, Rui C, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]//European Conference on Computer Vision, 2012: 702-715.