http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0362

显著性感知三重正则化相关滤波无人机目标跟踪算法

贺冰，王法胜*，王星，孙福明
(大连民族大学 信息与通信工程学院，大连 116600)

摘　　要：无人机(UAV)场景中的目标跟踪在很多现实任务中得到广泛应用。与一般场景中的目标跟踪任务不同，UAV目标跟踪更易受到复杂环境干扰和算力的限制。基于此，提出了一种显著性感知三重正则化相关滤波(TRCF)UAV目标跟踪算法。采用高效的显著性目标检测算法动态生成对偶空间正则化器来抑制边界效应，惩罚不相关的背景噪声系数。引入时间正则化应对目标因外观变化而导致的滤波器退化问题，提供更鲁棒的外观模型。此外，引入轻量型的深度网络CF-VGG来提取目标的深度特征，并与手工特征线性融合描述目标的语义信息，提高跟踪精度。在5个公开的UAV基准数据集上进行了充分实验，结果表明：所提算法在5个数据集上的整体性能均有不同程度提升，证明了算法的有效性和鲁棒性，且算法的实时跟踪速度约为21帧/s，能够胜任UAV的目标跟踪任务。

关 键 词：无人机目标跟踪；相关滤波；显著性特征图；时间正则化；轻量型深度网络
中图分类号：TP391
文献标志码：A　　**文章编号**：1001-5965(2025)07-2423-14

近年来，由于在救援救灾、空中监视、侦察、交通巡逻等方面的广泛应用，无人机(unmanned aerial vehicle, UAV)目标跟踪受到了广泛的关注[1-3]。在这些应用中，UAV能够通过视觉传感器感知周围环境，实现对指定目标的自主检测和跟踪。虽然近年来已经出现了许多先进的UAV目标跟踪算法，大大提升了跟踪的准确性，但由于目标的快速运动、尺度变化、完全遮挡等因素的影响，UAV目标跟踪仍然是一个极具困难和挑战性的任务[4]。

现有的主流目标跟踪算法主要有2类，即基于相关滤波(correlation filter, CF)的目标跟踪算法和基于深度学习(deep learning, DL)的目标跟踪算法。为了提高跟踪的准确性和鲁棒性，基于DL的目标跟踪算法使用1个或者多个深度网络对目标进行跟踪，但需要消耗较多的算力。由于考虑到单位成本和UAV本身的质量，大多数UAV目前没有配备图形处理单元(graphics processing unit, GPU)，导致DL目标跟踪算法在单中央处理单元(central processing unit, CPU)的UAV平台上难以满足跟踪精度的要求。基于CF的目标跟踪算法可以在跟踪精度和运行速度之间实现良好平衡，能够快速有效地部署到UAV平台，成为资源有限的UAV平台的首选[5]。

为了满足实时跟踪要求，传统的基于CF的目标跟踪算法普遍使用灰度特征[6]、方向梯度直方图及颜色名称[7]等手工特征描述目标，但手工特征并不能完整地表述目标的语义信息。由于残差网络(residual network, ResNet)、视觉几何组网络(visual geometry group network, VGG)等卷积神经网络(convolutional neural network, CNN)在图像分类和图像识别等具有挑战性的视觉任务上取得巨大成功，研究人员为提高跟踪性能，将深度特征应用到CF的

收稿日期：2023-06-15；**录用日期**：2023-09-15；**网络出版时间**：2023-12-07 07:00
网络出版地址：link.cnki.net/urlid/11.2625.V.20231206.1417.005
基金项目：国家自然科学基金(61972068,61976042)；辽宁省“兴辽英才计划”(XLYC2007023)
* **通信作者**. E-mail：wangfasheng@dlnu.edu.cn

引用格式：贺冰，王法胜，王星，等. 显著性感知三重正则化相关滤波无人机目标跟踪算法[J]. 北京航空航天大学学报，2025，51(7)：2423-2436. HE B，WANG F S，WANG X，et al. Saliency-aware triple-regularized correlation filter algorithm for UAV object tracking[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(7)：2423-2436 (in Chinese).

目标跟踪框架中[8-11]。然而，由于这些深度模型的复杂性，使用深度特征的目标跟踪算法在单CPU平台上运行速度往往只有2～3帧/s，无法满足实时跟踪的要求。受到UAV平台本身计算性能的限制，现有的面向UAV的目标跟踪算法大多使用手工特征描述目标。

CF目标跟踪算法所依赖的循环相关操作会对训练图像样本进行循环移位，导致样本在边界处出现不连续而引起边界效应问题，降低了滤波器的鉴别能力，影响跟踪性能。为了缓解边界效应，研究人员提出了2种方法：①如有限边界CF[12]和背景感知CF(background aware CF, BACF)[13]，直接从背景区域提取负训练样本用于训练相关滤波器。文献[14]在BACF算法的基础上通过在训练阶段引入2个相邻帧之间的特征残差来增强滤波器学习，提高目标跟踪算法在复杂环境下的跟踪性能。由于这类算法用于训练的所有正样本和负样本都被赋予相等的权重，使其难以适应复杂目标变化，且不能完全抑制边界效应。②包括空间正则化CF(spatially regularized discriminative CF, SRDCF)[15]和通道及空间可靠性CF[16]等，引入了空间正则化组件，对滤波器系数施加预定义的空间约束，以便滤波器可以在目标区域中学习。文献[17]在SRDCF跟踪框架中嵌入高斯形状掩模，通过增加中心样本的响应从而降低样本污染的影响。响应推理CF算法[18]在SRDCF算法基础之上引入响应推断正则化和历史响应正则化以更好地应对复杂的UAV场景。内容感知SRDCF算法[19]通过空间偏差隐式对不同的训练样本进行加权，重新构建一个能够抑制边界效应的能量函数。文献[20]通过引入上下文补丁在多个方向将响应抑制为零，以此来抑制边界效应。然而，上述算法添加的约束信息通常是固定的，并且不会根据跟踪过程中的目标变化而改变，使得模型无法适应跟踪目标的变化。同时，UAV场景目标跟踪中被跟踪目标的大小和视角易发生巨大变化，基于CF的目标跟踪算法采用基于固定速率的模型更新机制，忽略了不同帧之间的变化，当被跟踪对象外观发生显著变化时，从上一帧中学习到的滤波器就无法根据外观变化进行调整，从而导致滤波器退化。

为了解决以上问题，本文提出了一种显著性感知三重正则化CF(triple-regularized CF, TRCF)目标跟踪算法，引入一种能够有效提取目标信息的显著性目标检测算法构建对偶空间正则化因子，同时加入时间正则化与空间正则化构成三重正则化共同优化滤波器。为有效提高跟踪鲁棒性，在不额外增加算力的前提下，引入轻量型深度网络模块提取目标的深度特征。本文主要贡献如下：

1）为抑制边界效应，提高跟踪器辨别目标的能力，引入一种能够准确从背景中辨别目标的显著性目标检测算法构造动态空间正则化因子，使得相关滤波器可以对目标执行更有针对性的滤波器系数惩罚，并抑制不相关区域的背景噪声。

2）为增强目标跟踪算法在遭遇遮挡、形变等挑战因素时的抗干扰能力，在跟踪框架中引入时间正则化与空间正则化同时优化滤波器，使用交替方向乘子法(alternating direction method of multipliers, ADMM)进行优化求解，在目标发生较大外观变化时提供鲁棒的外观模型，防止滤波器退化。

3）为提高目标跟踪算法的精确度，引入一个轻量型的深度网络提取目标的深度特征，与手工特征按比例结合表示目标的语义信息。该轻量型深度网络计算量小，在单CPU支持下可以达到实时跟踪速度。

4）在5个公开的UAV基准数据集UAV123[21]、VisDrone-SOT2018[22]、UAV20L[21]、UAVTrack112[23]和UAVDT[24]上进行了充分实验，结果表明，TRCF目标跟踪算法具有较高的精确度和鲁棒性。

1 本文算法

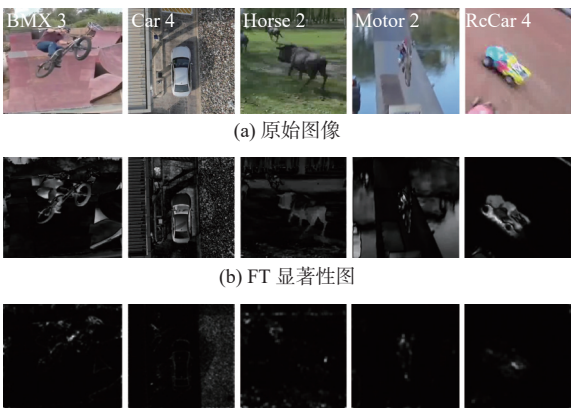
1.1 显著性目标检测模块

近年来，不断有研究者将显著性目标检测引入到CF目标跟踪算法框架中[25]，以提供更加全面的信息附加特征，帮助目标跟踪算法从背景中区分目标。此外，显著性目标检测还可以用来指示目标可能出现的位置，以缓解模型漂移问题[26]。文献[27]使用光谱残差(spectral residual, SR)[28]显著性目标检测算法构建双正则化因子，用来动态关注更多来自对象区域的信息，在相关运算中抑制边界效应。但由于SR算法使用低通滤波和下采样会生成64像素×64像素统一尺寸的显著性图，如果输入图像尺寸产生变化，在尺度空间对于图像频率产生的损失会导致模糊的显著性图。不清晰的显著性图会直接影响目标跟踪算法对目标对象的判断，进而导致滤波器性能下降。为了更加清晰地描述对象目标轮廓，产生能精准突出目标对象轮廓的显著性特征图，本文采用一种基于Lab颜色空间像素向量与平均像素向量欧氏距离的显著性检测算法，即频率协调(frequency-tuned, FT)显著性目标检测算法[29]来构建对偶正则化因子s_1和s_2，动态强调目标区域形状。首先，裁剪比目标边界框大λ倍的区域执行显著性目标检测，获得原始显著性图。然后，乘以余弦窗后，将原始映射调整为滤波器训练使用的外观模型的相应大小，通过阈值ε将其系数重新映射

用来调节权重。最后, 将规则映射$\boldsymbol{s}_{\mathrm{f}}$与另外预定义的正则化因子$\boldsymbol{s}_{\mathrm{r}}$相交, 得到最终的正则化因子, 如下:

$$\boldsymbol{s}_i = \boldsymbol{s}_{\mathrm{f}} \cap \boldsymbol{s}_{\mathrm{r}} \qquad i = 1,2 \tag{1}$$

为了展示 FT 显著性目标检测算法的效果, 选取 UAV 基准数据集中部分图像对 SR 与 FT 显著性目标检测算法进行测试, 实验对比结果如图 1 所示。可以看出, FT 显著性目标检测算法相较于 SR 显著性目标检测算法能够产生对于目标描述更加清晰的显著性图, 帮助跟踪器更加关注目区域, 使滤波器受到更少背景噪声的影响, 在检测目标时获得更强的响应。



(a) 原始图像

(b) FT 显著性图

(c) SR 显著性图

图 1　FT 与 SR 显著性目标检测结果图对比

Fig. 1　Comparison of saliency object detection results of FT and SR methods

1.2　轻量型深度网络

近年来, 同时使用深度特征和手工特征结合的 CF 目标跟踪算法表现出了较强的竞争力, 但多数跟踪器采用结构复杂的大型深度网络提取深度特征, 大大增加计算成本, 只有在搭载 GPU 的跟踪平台上才能够达到实时的跟踪效率。由于 UAV 的主要优势就在于体积小、灵活度高, 导致其仅能搭载 CPU。为了满足实时速率, 目前面向 UAV 的跟踪器大都使用手工特征进行特征提取, 这对跟踪器的精度造成一定影响。文献 [30] 采用知识蒸馏技术设计了一个轻量型的深度网络 CF-VGG 提取目标的深度特征, 并且该网络在单个 CPU 平台上就能达到实时的跟踪速率, 适合 UAV 目标跟踪的应用场景。

CF-VGG 将 VGG-M[31] 作为教师网络, 使用知识蒸馏[32-33] 生成轻量型的学生网络。CF-VGG 的网络架构如图 2 所示。

本文将 CF-VGG 引入到算法的跟踪框架中来提取目标的深度特征, 与手工提取的特征一起描述目标特征。具体而言, 对于第$t+1$帧的搜索区域, CF-VGG 使用 stride = 2的步长进行下采样, 并用

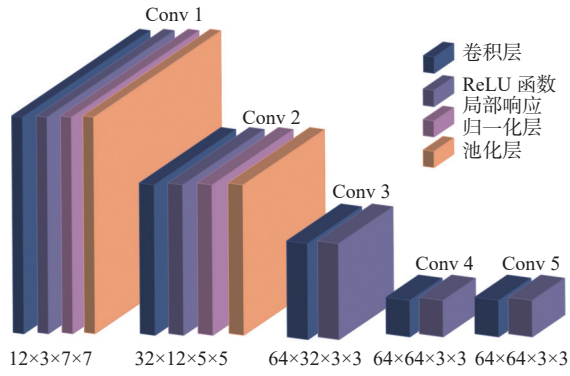


图 2　CF-VGG 网络架构

Fig. 2　Architecture of CF-VGG

ReLU函数对更有助于精确定位的第 3 卷积层上的 CNN 特征进行处理, 以描述目标的语义信息。

1.3　时间正则化模块

CF 目标跟踪算法中使用循环移位操作增加样本, 从而引起边界效应, 降低了跟踪性能。Danelljan 等[15] 提出空间正则化项, 根据相关滤波系数的空间位置对其进行惩罚, 可以减轻边界效应并提高相关滤波器的性能。但面对复杂的 UAV 跟踪场景, 目标更容易因为遮挡、形变、尺度变化等挑战因素导致目标外观发生显著变化, 如果上一帧中的滤波器无法根据外观变化进行调整, 就会导致过滤器退化从而影响跟踪器性能。针对该问题, 本文将时间正则化[34] 引入到基线跟踪框架中构建三重正则化器。由于时间正则化是对多个样本的合理近似, 几乎不额外增加计算复杂度, 能够使算法在不损失效率的同时更好处理边界效应; 并且时间正则化项可以自适应平衡主动和被动模型学习, 对于遭遇遮挡、形变和尺度变化等会产生较大外观变化的序列提供更加稳健的模型, 防止滤波器退化。

本文算法的整体跟踪流程如图 3 所示。

1.4　算法的建模和优化

1.4.1　TRCF 目标跟踪算法的建模和优化

本文算法的目标函数为

$$\arg\min_{\boldsymbol{w}} \frac{1}{2}\left\|\sum_{d=1}^{D}\boldsymbol{x}_d \star (\boldsymbol{s}_1 \odot \boldsymbol{w}_d) - \boldsymbol{y}\right\|_2^2 + \frac{1}{2}\sum_{d=1}^{D}\|\boldsymbol{s}_2 \odot \boldsymbol{w}_d\|_2^2 + \frac{\mu}{2}\|\boldsymbol{w} - \boldsymbol{w}_{t-1}\|_2^2 \tag{2}$$

式中: $\boldsymbol{x} = \{\boldsymbol{x}_d\}_{d=1:D}$ 和 $\boldsymbol{w} = \{\boldsymbol{w}_d\}_{d=1:D}$分别为第$d$个特征通道中的训练样本和滤波器; D为特征通道数; $\boldsymbol{y} \in \mathbf{R}^{M\times N}$为预定义的高斯形状期望响应, 其空间大小尺寸为$M \times N$; $\star$表示循环相关运算; $\odot$表示逐元素相乘; $\boldsymbol{s}_1$和$\boldsymbol{s}_2 \in \mathbf{R}^{M\times N}$为由显著性特征图生成的动态双正则化器; t为第t帧; $\boldsymbol{w}_{t-1}$为$t-1$帧滤波器; μ为正则化参数; $\|\boldsymbol{w} - \boldsymbol{w}_{t-1}\|_2^2$定义为时间正则化因子。

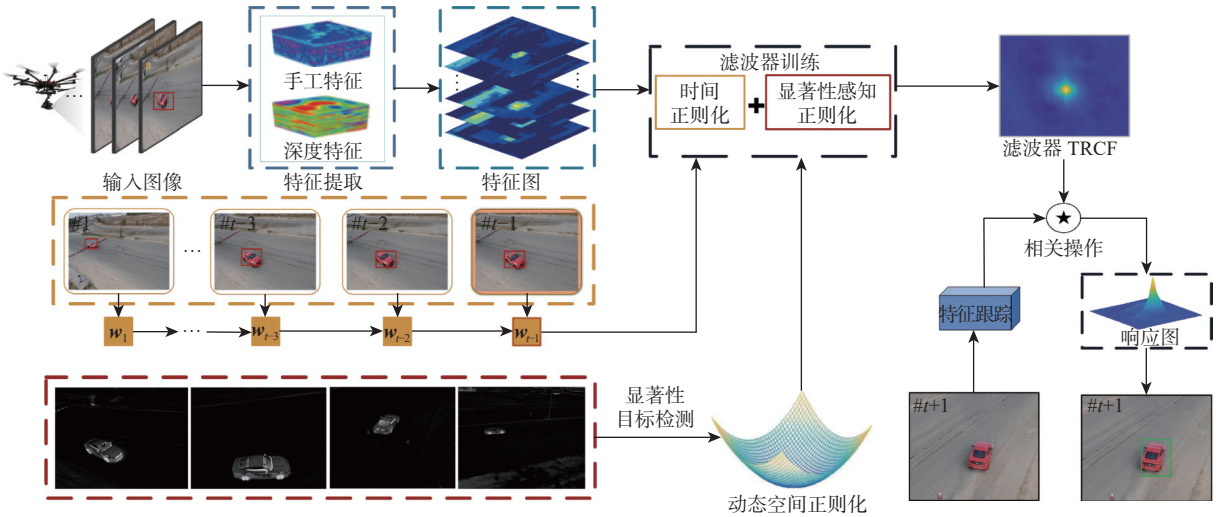


图3 TRCF目标跟踪算法框架

Fig. 3 Framework of TRCF object tracking algorithm

通过引入约束$\boldsymbol{g}=\boldsymbol{s}_1\odot\boldsymbol{w}_d$、拉格朗日乘数$\boldsymbol{r}$及相应的步长参数$\rho\in\mathbf{R}$，对式(2)使用增广拉格朗日方法求解。式(2)利用增广拉格朗日法转化为

$$\begin{aligned}L(\boldsymbol{w},\boldsymbol{g},\boldsymbol{r})=&\frac{1}{2}\left\|\sum_{d=1}^{D}\boldsymbol{x}_d\star\boldsymbol{g}_d-\boldsymbol{y}\right\|_2^2+\frac{1}{2}\sum_{d=1}^{D}\|\boldsymbol{s}_2\odot\boldsymbol{w}_d\|_2^2+\\&\sum_{d=1}^{D}\operatorname{trace}\left[(\boldsymbol{g}_d-\boldsymbol{s}_1\odot\boldsymbol{w}_d)^{\mathrm{T}}\boldsymbol{r}_d\right]+\\&\frac{\rho}{2}\sum_{d=1}^{D}\|\boldsymbol{g}_d-\boldsymbol{s}_1\odot\boldsymbol{w}_d\|_2^2+\frac{\mu}{2}\|\boldsymbol{w}-\boldsymbol{w}_{t-1}\|_2^2\end{aligned}\tag{3}$$

通过引入$\boldsymbol{h}=\dfrac{1}{\rho}\boldsymbol{r}$，式(3)可以重新表述为

$$\begin{aligned}L(\boldsymbol{w},\boldsymbol{g},\boldsymbol{r})=&\frac{1}{2}\left\|\sum_{d=1}^{D}\boldsymbol{x}_d\star\boldsymbol{g}_d-\boldsymbol{y}\right\|_2^2+\frac{1}{2}\sum_{d=1}^{D}\|\boldsymbol{s}_2\odot\boldsymbol{w}_d\|_2^2+\\&\frac{\rho}{2}\sum_{d=1}^{D}\|\boldsymbol{g}_d-\boldsymbol{s}_1\odot\boldsymbol{w}_d+\boldsymbol{h}_d\|_2^2-\frac{\rho}{2}\sum_{d=1}^{D}\|\boldsymbol{h}_d\|_2^2+\\&\frac{\mu}{2}\|\boldsymbol{w}-\boldsymbol{w}_{t-1}\|_2^2\end{aligned}\tag{4}$$

采用ADMM方法对算法进行优化，将原问题转化为关于滤波器和辅助变量求解的子问题。式(4)可以被分解为3个子问题进行迭代求解，且所有子问题都有封闭解：

$$\begin{cases}\boldsymbol{g}^{k+1}=\arg\min\limits_{\boldsymbol{g}}\left\{\dfrac{1}{2}\left\|\sum\limits_{d=1}^{D}\boldsymbol{x}_d\star\boldsymbol{g}_d-\boldsymbol{y}\right\|_2^2+\dfrac{\rho}{2}\left\|\boldsymbol{g}-\boldsymbol{s}_1\odot\boldsymbol{w}^k+\boldsymbol{h}^k\right\|_2^2\right\}\\\boldsymbol{w}^{k+1}=\arg\min\limits_{\boldsymbol{w}}\left\{\dfrac{1}{2}\|\boldsymbol{s}_2\odot\boldsymbol{w}\|_2^2+\dfrac{\rho}{2}\left\|\boldsymbol{g}-\boldsymbol{s}_1\odot\boldsymbol{w}+\boldsymbol{h}^k\right\|_2^2+\dfrac{\mu}{2}\|\boldsymbol{w}-\boldsymbol{w}_{t-1}\|_2^2\right\}\\\boldsymbol{h}^{k+1}=\boldsymbol{h}^k+\boldsymbol{g}^{k+1}-\boldsymbol{s}_1\odot\boldsymbol{w}^{k+1}\end{cases}\tag{5}$$

式中：k为迭代次数。

1)子问题$\boldsymbol{g}^{k+1}$

利用帕塞瓦尔定理和卷积定理，式(5)中的第1个子问题可以转换到傅里叶域中，如下：

$$\begin{aligned}\hat{\boldsymbol{g}}^{k+1}=&\arg\min_{\hat{\boldsymbol{g}}}\left\{\frac{1}{2}\left\|\sum_{d=1}^{D}\hat{\boldsymbol{x}}_d^*\odot\hat{\boldsymbol{g}}_d-\hat{\boldsymbol{y}}\right\|_2^2+\right.\\&\left.\frac{\rho}{2}\left\|\hat{\boldsymbol{g}}-\left(\widehat{\boldsymbol{s}_1\odot\boldsymbol{w}}\right)+\hat{\boldsymbol{h}}\right\|_2^2\right\}\end{aligned}\tag{6}$$

式中：$*$表示复共扼；“^”表示离散傅里叶变换，$\hat{\boldsymbol{a}}=\boldsymbol{F}\cdot\boldsymbol{a}$，$\boldsymbol{F}$为离散傅里叶变换矩阵。由于$\hat{\boldsymbol{y}}$仅依赖于所有通道上的$\hat{\boldsymbol{g}}_d(i,j)$和$\hat{\boldsymbol{x}}_d^*(i,j)$，$\boldsymbol{e}_{ij}\in\mathbf{C}^{D\times1}$表示覆盖所有$D$个通道上的$\hat{\boldsymbol{x}}_d^*(i,j)$元素。式(6)可以分解为$M\times N$个子问题，每个子问题的形式如下：

$$\begin{aligned}&\arg\min_{\boldsymbol{e}_{ij}(\hat{\boldsymbol{g}})}\left\{\frac{1}{2}\left\|\sum_{d=1}^{D}\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^{\mathrm{H}}\boldsymbol{e}_{ij}(\hat{\boldsymbol{g}})-\hat{\boldsymbol{y}}_{ij}\right\|_2^2+\right.\\&\left.\frac{\rho}{2}\left\|\boldsymbol{e}_{ij}(\hat{\boldsymbol{g}})-\boldsymbol{e}_{ij}\left(\widehat{\boldsymbol{s}_1\odot\boldsymbol{w}}\right)+\boldsymbol{e}_{ij}\left(\hat{\boldsymbol{h}}\right)\right\|_2^2\right\}\end{aligned}\tag{7}$$

其中，$(\hat{\boldsymbol{x}})^{\mathrm{H}}$表示Hermitian转置，$(\hat{\boldsymbol{x}})^{\mathrm{H}}=(\hat{\boldsymbol{x}}^*)^{\mathrm{T}}$。将式(7)关于$\boldsymbol{e}_{ij}(\hat{\boldsymbol{g}})$的导数设为零，可得

$$\boldsymbol{q}\cdot\boldsymbol{e}_{ij}(\hat{\boldsymbol{g}})=\rho\left[\boldsymbol{e}_{ij}(\boldsymbol{s}_1\odot\boldsymbol{w})-\boldsymbol{e}_{ij}\left(\hat{\boldsymbol{h}}\right)\right]+\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^*\hat{\boldsymbol{y}}_{ij}\tag{8}$$

式中：$\boldsymbol{q}=\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^*\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^{\mathrm{H}}+\rho\boldsymbol{I}$，$\boldsymbol{I}$为单位矩阵。式(8)中含有矩阵的求逆过程会影响计算速度，为了减少计算量，采用Sherman-Morrison公式，可以获得$\boldsymbol{e}_{ij}(\hat{\boldsymbol{g}})$的封闭解：

$$\begin{cases}\boldsymbol{e}_{ij}(\hat{\boldsymbol{g}})=\dfrac{1}{\rho}\left[\boldsymbol{I}-\dfrac{\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^*\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^{\mathrm{H}}}{\rho+\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^{\mathrm{H}}\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^*}\right]\boldsymbol{\psi}\\\boldsymbol{\psi}=\rho\left[\boldsymbol{e}_{ij}\left(\widehat{\boldsymbol{s}_1\odot\boldsymbol{w}}\right)-\boldsymbol{e}_{ij}\left(\hat{\boldsymbol{h}}\right)\right]+\boldsymbol{e}_{ij}(\hat{\boldsymbol{x}})^*\hat{\boldsymbol{y}}_{ij}\end{cases}\tag{9}$$

2) 子问题$\boldsymbol{w}^{k+1}$

第2个子问题可以通过将导数设为零在时域中直接求解，即

$$\frac{\partial}{\partial \boldsymbol{w}}\left\{\frac{1}{2}\|\boldsymbol{s}_2\odot\boldsymbol{w}\|_2^2+\frac{\rho}{2}\|\boldsymbol{g}-\boldsymbol{s}_1\odot\boldsymbol{w}+\boldsymbol{h}\|_2^2+\frac{\mu}{2}\|\boldsymbol{w}-\boldsymbol{w}_{t-1}\|_2^2\right\}=0 \tag{10}$$

得到$\boldsymbol{w}$的解如下:

$$\boldsymbol{w}=\frac{\rho\cdot\boldsymbol{s}_1\odot(\boldsymbol{g}+\boldsymbol{h})+\mu\boldsymbol{w}_{t-1}}{\rho(\boldsymbol{s}_1\odot\boldsymbol{s}_1)+(\boldsymbol{s}_2\odot\boldsymbol{s}_2)+\mu} \tag{11}$$

3) 迭代更新

在每次的迭代更新中, $\boldsymbol{h}$使用式(5)的最后 1 行进行更新。步长参数ρ使用文献 [15] 中相同的方案用尺度β进行更新, 即

$$\rho^{k+1}=\min\left(\rho_{\max},\beta\rho^k\right) \tag{12}$$

1.4.2　TRCF 目标跟踪算法描述

输入: 第t帧图像、目标位置$\boldsymbol{p}_{t-1}$、前一帧$t-1$上的尺度$\boldsymbol{\iota}_{t-1}$、外观模型$\boldsymbol{x}_{t-1}^{\text{model}}$、正则化器$\boldsymbol{s}_1$、$\boldsymbol{s}_2$和滤波器$\boldsymbol{w}_{t-1}$。

输出: t帧处的预测位置$\boldsymbol{p}_t$和尺度$\boldsymbol{\iota}_t$。

在跟踪过程中, 对于每一帧, 以前一帧中的目标位置为中心裁剪相应尺度的样本区域来提取其手工特征和深度特征, 并采用式(13)将 2 种特征进行线性组合以得到最终的目标特征。

$$\boldsymbol{\chi}_{\text{total}}=v\boldsymbol{\chi}_{\text{m}}+(1-v)\boldsymbol{\chi}_{\text{n}} \tag{13}$$

式中: $\boldsymbol{\chi}_{\text{m}}$为手工特征; $\boldsymbol{\chi}_{\text{n}}$为深度特征; $\boldsymbol{\chi}_{\text{total}}$为总特征; v为平衡 2 种特征的参数。

本文采用文献 [26] 中提出的线性插值方法更新外观模型$\boldsymbol{\chi}_t^{\text{model}}$和滤波器$\boldsymbol{w}_t$, 学习率为$\eta$, $\boldsymbol{\chi}_t$为从第t帧提取的特征。如下:

$$\boldsymbol{\chi}_t^{\text{model}}=(1-\eta)\boldsymbol{\chi}_{t-1}^{\text{model}}+\eta\boldsymbol{\chi}_t \tag{14}$$

$$\boldsymbol{w}_t=(1-\eta)\boldsymbol{w}_{t-1}+\eta\boldsymbol{w}_t \tag{15}$$

将相关滤波器的响应函数定义为$\dot{y}$, 如下:

$$\dot{y}=\boldsymbol{F}^{-1}\left\{\sum_{d=1}^{D}\hat{\boldsymbol{x}}_d\odot\hat{\boldsymbol{w}}_d\right\} \tag{16}$$

响应的最大值位置即为目标的估计位置。

对于序列的第 2 帧至最后 1 帧:

1) 在以目标位置$\boldsymbol{p}_{t-1}$为中心的第t帧上裁剪图像补丁, 确定搜索区域。

2) 提取手工特征及深度特征使用式(14)进行线性融合, 构建目标描述。

3) 搜索式(16)$\dot{y}$中的最高值, 作为目标位置估计$\boldsymbol{p}_t$。

4) 使用文献 [35] 中的尺度空间相关滤波器来估计尺度$\boldsymbol{\iota}_t$。

5) 生成新空间正则化器$\boldsymbol{s}_1$、$\boldsymbol{s}_2$, 并放弃旧的空间正则化器。

6) 提取$\boldsymbol{p}_t$和$\boldsymbol{\iota}_t$处的特征$\boldsymbol{\chi}_t$, 并使用式(14)将外观模型更新为$\boldsymbol{\chi}_t^{\text{model}}$。

7) 使用外观模型$\boldsymbol{\chi}_t^{\text{model}}$和正则化器$\boldsymbol{s}_1$、$\boldsymbol{s}_2$及第$t-1$帧滤波器$\boldsymbol{w}_{t-1}$优化滤波器。

8) 使用式(15)将滤波器更新为$\boldsymbol{w}_t$。

2　实验结果与分析

2.1　实验细节

本文使用 MATLAB 2018a 实现算法, 所有实验均在 64 位 Windows10 平台下完成。计算机处理器参数为 Intel(R) Core (TM) i7-9700 3.00 GHz CPU, 内存为 48 GB。跟踪器的搜索区域被设置为大小为$5XY/2$的正方形, 其中, X和Y分别为目标的高度和宽度。使用$v=0.85$来平衡手工特征与深度特征。在使用显著性目标检测算法构建正则化因子过程中, 尺度$\lambda=2$, 重映射阈值$\varepsilon=0.1$。对于 ADMM 迭代, $\boldsymbol{g}$、$\boldsymbol{w}$和 $\boldsymbol{h}$的初始值都是空矩阵。ADMM 迭代次数、初始步长ρ、超参数μ、最大值$\rho_{\max}$和尺度β分别为 2、1、16、100 和 10。用于模型更新的学习率设置为$\eta=0.0192$。

将本文算法与其他 CF 目标跟踪算法采用一遍评估法(one pass evaluation, OPE)进行评估, 即使用精确度和成功率[36] 对算法进行测评。

在每一帧中, 精确度由跟踪目标物体的中心位置与手动标记的目标物体真实值之间的中心位置误差(center location error, CLE)确定, 用精确度图中 CLE 低于 20 像素的帧的百分比对跟踪器进行排名。成功率由边界框和真实边界框之间的交并比来衡量, 定义为

$$I_{\text{oU}}=\frac{\boldsymbol{C}_p\cap\boldsymbol{C}_q}{\boldsymbol{C}_p\cup\boldsymbol{C}_q} \tag{17}$$

式中: $\boldsymbol{C}_p$为跟踪预测的边界框; $\boldsymbol{C}_q$为标定的真实边界框; $\cap$和$\cup$分别表示这 2 个区域的交集和并集。当交并比大于给定阈值时, 定义为跟踪成功, 跟踪成功的帧数占视频序列总帧数的百分比即为成功率。使用成功图的曲线下面积(area under curve, AUC)对跟踪器进行排名。

2.2　定量分析

本节对提出的跟踪器进行全面的实验评估, 采用 5 个公开的 UAV 目标数据集 UAV123[21]、VisDrone-SOT2018[22]、 UAV20L[21]、 UAVTrack112[23]、UAVDT[24] 与多个先进的 CF 目标跟踪算法进行对比实验, 包括 AutoTrack[37]、ADTrack_v2[38]、MRCF[39]、MSCF[40]、ARCF[41]、SRECF[42]、DRCF[27]、SRDCF[15]、STRCF[35]、Staple[43]、BACF[13]、DSST[34]、fDSST[44]、SAMF[45]。

UAV123数据集中包含UAV拍摄的123个彩色序列。VisDrone-SOT2018数据集由35个专注于复杂场景下面临众多挑战的视频序列组成。UAV20L包含20个长时跟踪序列，由超过54 000帧的各种具有挑战性的场景序列组成。UAVDT是采用UAV拍摄的车辆交通内容的数据集，共50个视频序列，包含天气条件、飞行高度、车辆类别等14种属性。UAVTrack112是基于真实世界生活场景构建的空中跟踪的数据集，共有112个视频序列。

2.2.1 UAV123数据集

OPE实验结果如图4所示。在精确度方面，TRCF(70.3%)目标跟踪算法排名第1，比第2名的ADTrack_v2算法高出0.2%，比排名第3的基准算法DRCF性能提升0.7%。在成功率方面，TRCF目标跟踪算法以48.8%的得分排名第1，比第2名的MRCF算法高0.3%，比第3名的MSCF算法高0.5%。ADTrack_v2算法和基准算法DRCF以相同的分数分别排在第5位和第6位。

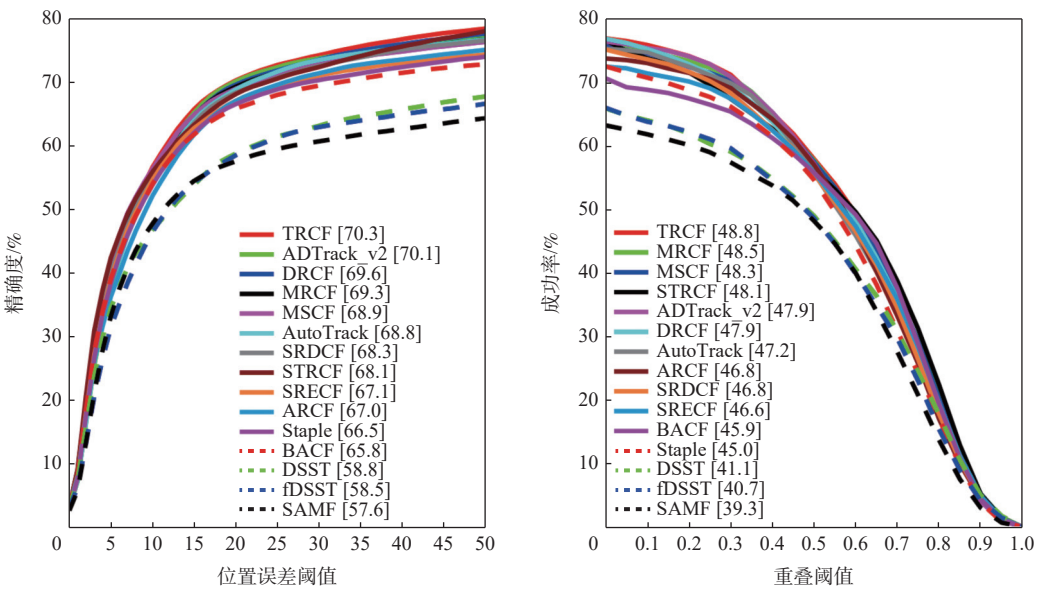


图4 UAV123数据集上的精确度和成功率曲线

Fig. 4 Precision and success rate plots on UAV123 dataset

为评估TRCF目标跟踪算法在不同挑战属性上的表现，将15种CF目标跟踪算法在UAV123数据集对于不同挑战属性上的性能进行汇总。图5为12种挑战属性上的成功率曲线。TRCF目标跟踪算法在6个挑战属性达到了最好性能，分别为纵横比变换(41.6%)、相机运动(49.0%)、超出视野(42.0%)、尺度变化(45.7%)、部分遮挡(41.0%)及视角变化(43.5%)。在小物体和快速运动2种挑战属性中，TRCF目标跟踪算法分别以46.6%和37.0%排进了前3。上述结果表明，TRCF目标跟踪算法对于目标发生较大外观变化和位置变化的挑战属性性能表现优异，加入时间正则化模块与高效的显著性目标检测算法模块相结合共同优化滤波器后，可以更加精准地跟踪目标，并构建更加鲁棒的外观模型，从而提高跟踪性能。

2.2.2 VisDrone-SOT2018数据集

图6显示了在VisDrone-SOT2018数据集上的评估结果。在精确度方面，TRCF目标跟踪算法以79.9%排名第2，比排名第3的ARCF算法高0.2%，SRECF算法则以79.1%排名第4，与本文算法相差0.8%。在成功率方面，TRCF目标跟踪算法与第1名相差1.6%排在第2位，比SRECF算法(57.4%)、AutoTrack算法(57.3%)、DRCF算法(57.3%)、MSCF算法(57.0%)分别提高了1%、1.1%、1.1%、1.4%。

2.2.3 UAV20L数据集

目标跟踪算法在UAV20L数据集上的测评结果如图7所示。在精确度方面，本文算法以60.6%的得分排在第1名，比第2名的STRCF算法高3.1%。MSCF算法和MRCF算法分别以57.0%和56.8%排在第3位和第4位。本文算法比排在第8位的基准算法DRCF提高了6.4%。在成功率方面，TRCF目标跟踪算法以42.9%的得分领先其他先进算法排在第1名，与排名第2～3名的STRCF算法和MRCF算法相比，分别提高了1.9%和3.2%，而与BACF算法、MSCF算法、ADTrack_v2算法、ARCF算法及本文基准算法DRCF相比，TRCF目标跟踪算法在成功率上分别提高了3.2%、3.6%、4.4%、4.8%和5.2%。实验结果表明，本文设计的CF目标跟踪框架提升了CF目标跟踪算法对长时UAV目标跟踪的处理能力。

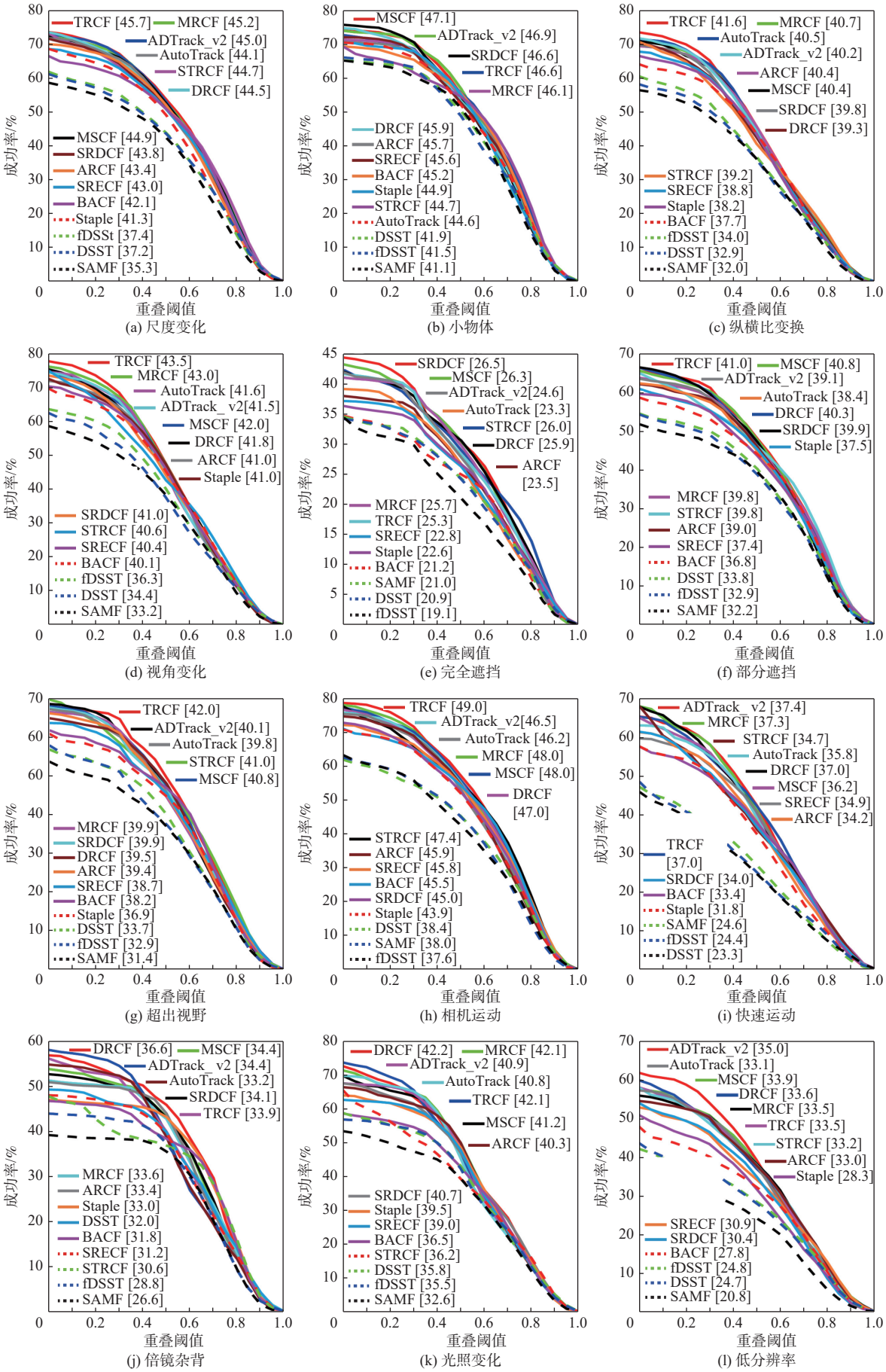


图 5　UAV123 数据集上的 12 个挑战属性下的成功率曲线

Fig. 5　Success rate plots of 12 challenging attributes on UAV123 dataset

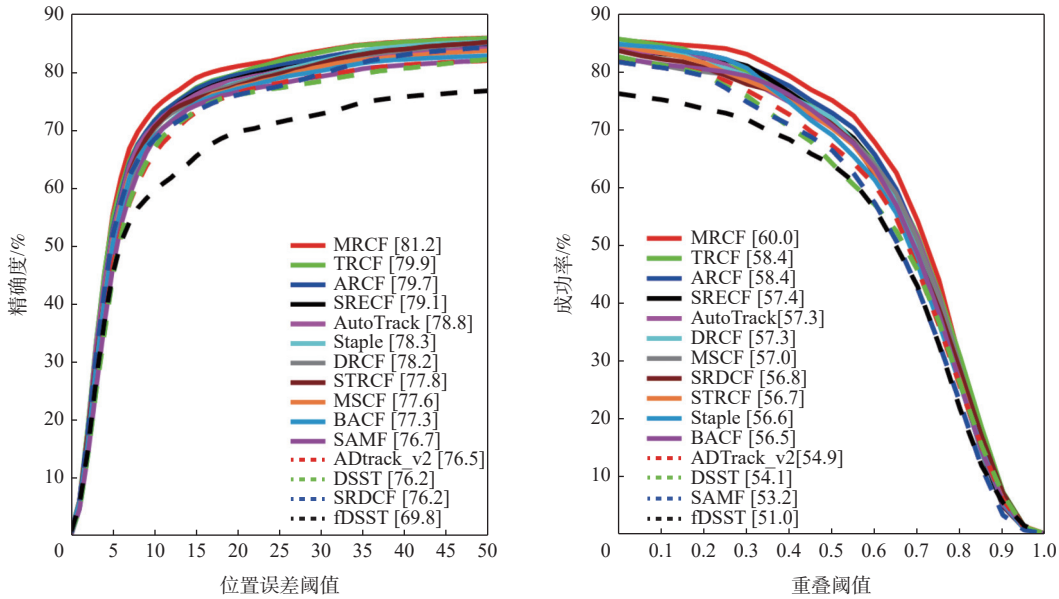


图 6 VisDrone-SOT2018 数据集上的精确度和成功率曲线

Fig. 6 Precision and success rate plots on VisDrone-SOT2018 dataset

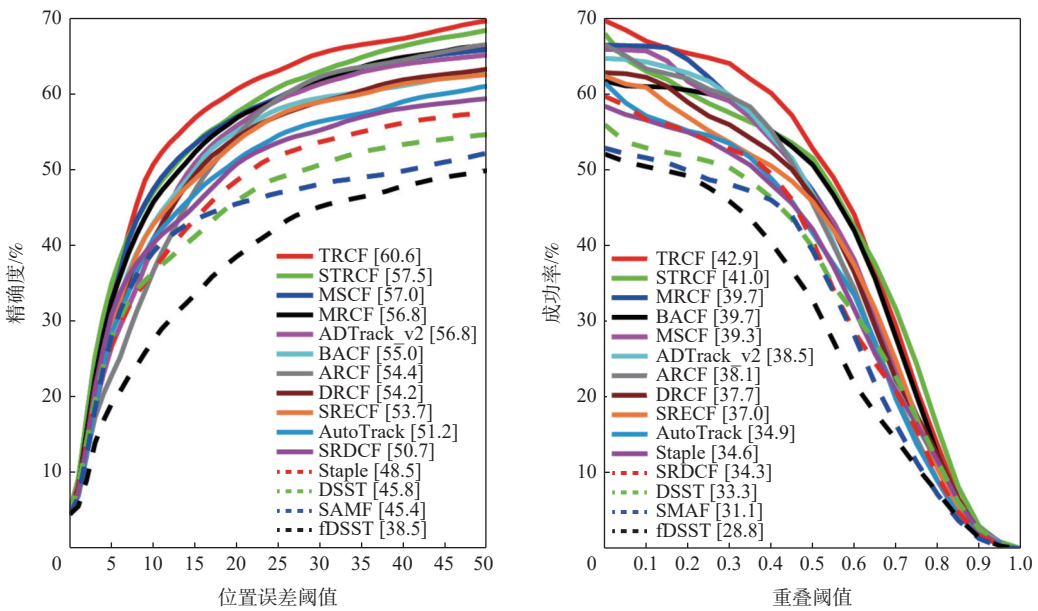


图 7 UAV20L 数据集上的精确度和成功率曲线

Fig. 7 Precision and success rate plots on UAV20L dataset

2.2.4 UAVTrack112 数据集

图 8 为算法在 UAVTrack112 数据集上的评测结果。TRCF 目标跟踪算法的精确度为 68.0%，排在第 5 位。TRCF 目标跟踪算法的成功率为 46.1%，比 ARCF 算法高出 0.4%，比 SRECF 算法高出 0.9%，与第 1 名算法差距仅为 0.7%。由于 UAVTrack112 数据集主要关注小物体、光照变化及快速运动等挑战因素，是面向全天场景构建的 UAV 目标序列。对于目标占比极小的小物体，深度网络能提取到的语义信息有限，显著性特征图的提取也易受到周围较大物体的影响而导致不精准显著性图的生成，使跟踪结果受到影响。但 TRCF 目标跟踪算法相比于基准算法 DRCF 精确度高出 0.5%，成功率高出 0.3%，改进后的本文算法跟踪性能整体有明显提升。

2.2.5 UAVDT 数据集

各算法在 UAVDT 数据集上的评估结果如图 9 所示。TRCF 目标跟踪算法的精确度和成功率分别为 73.2% 和 45.4%，排在第 6 名。在精确度方面，TRCF 目标跟踪算法与排名第 1～5 位的 ARCF 算法、MRCF 算法、AutoTrack 算法、MSCF 算法、SRECF 算法的精确度相差均未超过 1%，但比基准算法 DRCF 的精确度提高了 1.3%。成功率方面，TRCF 目标跟踪算法比基准算法 DRCF 的成功率高出 0.6%。UAVDT 数据集是面向 UAV 拍摄的车辆

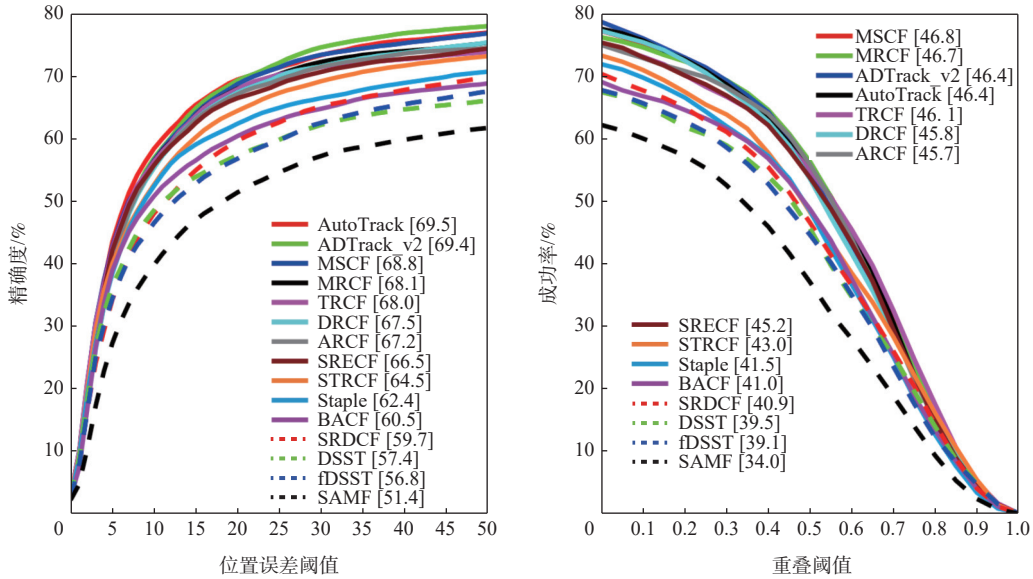


图 8　UAVTrack112 数据集上的精确度和成功率曲线

Fig. 8　Precision and success rate plots on UAVTrack112 dataset

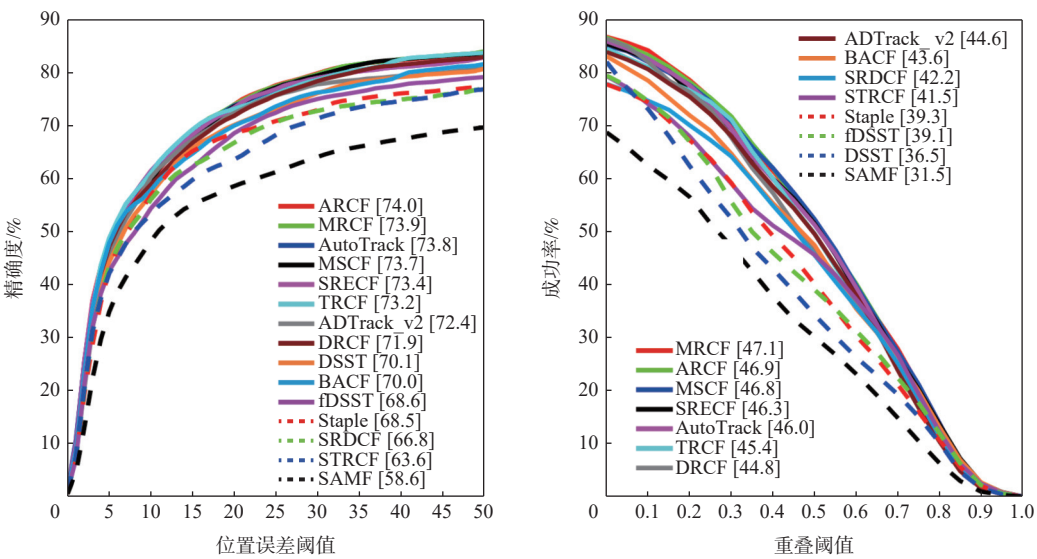


图 9　UAVDT 数据集上的精确度和成功率曲线

Fig. 9　Precision and success rate plots on UAVDT dataset

目标, 多数为距离较远的小目标车辆及包含数个夜间和雾天场景的目标序列。在照明条件有利的情况下, 显著性目标检测算法能够较好地检测到目标轮廓, 但在暗夜场景目标不清晰, 影响显著性特征图的结果, 使跟踪性能受到影响。相比于基准算法 DRCF, 本文算法无论是精确度还是成功率都有明显提升, 证明本文算法能起到提升算法整体跟踪性能的作用。

为直观地展示本文算法相较于基准算法在不同数据集上的提升幅度, 表 1 列出了本文算法与基准算法在 5 个不同基准数据集上的性能数值量化。

表 1　本文算法与基准算法的性能比较

Table 1　Comparisons between the proposed method and baseline method　%

基准数据集	精确度		成功率	
	DRCF	TRCF	DRCF	TRCF
UAV123[21]	69.6	**70.3**	47.9	**48.8**
VisDrone-SOT2018[22]	78.2	**79.9**	57.3	**58.4**
UAVTrack112[23]	67.5	**68.0**	45.8	**46.1**
UAV20L[21]	54.2	**60.6**	37.7	**42.9**
UAVDT[24]	71.9	**73.2**	44.8	**45.4**

注: 粗体表示与基准算法相比性能增强。

2.2.6　时间正则化参数的影响

进一步分析不同时间正则化参数对滤波器跟踪性能的影响, 正则化参数 μ 决定了用当前帧中的新样本替换从前一帧中学习到的滤波器的速率。

从图 10 可以看出，μ对 TRCF 目标跟踪器的精度有明显影响，当$\mu=16$时，TRCF 目标跟踪器的性能达到最佳。

2.3 消融实验

2.3.1 跟踪性能对比

为了验证 TRCF 目标跟踪算法中显著性特征检测模块、轻量级深度特征和时间正则化对本文算法性能的影响，在 3 个 UAV 基准数据集上进行了消融实验，结果如表 2 所示。

在UAV123 数据集中，成功率方面，加入 CF-VGG 和加入 CF-VGG 与时间正则化的改进算法高于基准算法，加入 3 个模块后，本文算法的成功率比基准算法提高 0.9%。在精确度方面，加入 CF-VGG、加入 CF-VGG 与时间正则化及本文算法得分均高于基准算法，分别提高了 0.5%、0.2% 和 0.7%。

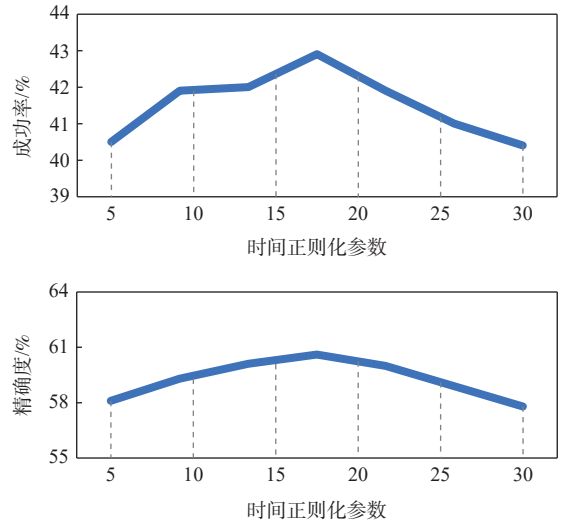


图 10 UAV20L 数据集上超参数分析

Fig. 10 Hyperparameter analysis on UAV20L dataset

表 2 消融实验结果对比

Table 2 Ablation study results %

基准数据集	精确度						成功率					
	DRCF	DRCF+Tr	DRCF+FT	DRCF+CF-VGG	DRCF+CF-VGG+Tr	DRCF+CF-VGG+FT+Tr	DRCF	DRCF+Tr	DRCF+FT	DRCF+CF-VGG	DRCF+CF-VGG+Tr	DRCF+CF-VGG+FT +Tr
UAV123[21]	69.6	66.3	67.9	70.1	69.8	**70.3**	47.9	46.4	47.7	48.3	48.4	**48.8**
UAVDT[24]	71.9	71.0	73.2	70.2	73.3	**73.2**	44.8	44.3	46.0	44.7	45.4	**45.4**
VisDrone-SOT2018[22]	78.2	80.6	78.3	76.2	79.9	**79.9**	57.3	57.6	56.8	55.6	58.5	**58.4**

注：粗体表示与基准算法相比性能增强；Base+FT、Base+CF-VGG、Base+Tr、Base+CF-VGG+Tr分别表示在基准算法中加入FT显著性目标检测模块、轻量级深度特征、时间正则化、同时加入轻量级深度特征和时间正则化；Base+CF-VGG+FT+Tr为加入所有模块的TRCF目标跟踪算法。

在 UAVDT 数据集中，包含较多飞行高度较高拍摄的小目标物体及夜间、雾天等光线原因使得目标外观模糊的序列。结果显示，只加入时间正则化和只加入 CF-VGG 的算法性能有轻微下降，但其他改进算法的跟踪性能均高于基准算法。在精确度方面，本文算法提高了 1.3%，成功率方面提高了 0.6%。

在 VisDrone-SOT2018 数据集中，依次加入时间正则化、CF-VGG 和 FT 显著性目标检测模块后，相比于基准算法，在跟踪性能上都得到了不同程度的提升。加入时间正则化模块在精确度方面提升最明显，提高了 2.4%，加入 CF-VGG 和时间正则化在成功率方面提升显著，提高了 1.2%。本文算法相较于基准算法，精确度和成功率分别提高了 1.7% 和 1.1%。

由于 UAV 数据集构建时就有针对性的集中面向某几种主要挑战或是面向某种特殊场景，例如，UAVDT 是基于 UAV 摄制的车辆交通内容的数据集，其中由于拍摄高度较高小目标对象序列较多，导致单独加入轻量级深度特征模块时，能提取到的目标语义信息有限，跟踪性能有所下降。加入 FT 显著性目标检测模块后在 UAV123 数据集上跟踪性能略有下降。这是由于显著性目标检测算法对于分辨率高、轮廓清楚的目标具有较好的提取效果，对于 UAV123 数据集存在的因为像素等原因导致的分辨率低的目标，检测能力会相对较弱。但从表 2 可知，当各模块逐层递增加入基准算法后，相较于基准算法，无论是成功率还是精确度都有不同程度的提升。不同 UAV 目标跟踪数据集所面向的主要挑战在 3 个模块的协同处理下都得到了相较于基准算法更好的解决。

2.3.2 跟踪速度分析

对于 UAV 目标跟踪问题，帧率（frames per second，FPS）是至关重要的评价指标。表 3 为各算法在 UAV 基准数据集上的帧率对比。为分析不同模块对计算复杂度及跟踪速度的影响，在 3 个 UAV 目标跟踪数据集上加入不同模块的算法进行帧率对比，实验结果如表 4 所示。由于时间正则化是对多个样本的合理近似，几乎不额外增加计算复杂度，因此单独加入时间正则化模块，帧率几乎不受影响。使用显著性目标检测算法构建正则化器后，算法整体复杂度略有上升，在 3 个数据集上平均下降了 4.3 帧/s，对于算法的整体速率影响相对较小。使用 CF-VGG 第 10 层的输出作为深度特征来描述

表 3　各算法在 UAV 数据集上的帧率比较

Table 3　FPS comparison of different algorithms on UAV datasets　　帧/s

算法	帧率			平均值
	UAVDT	VisDrone-SOT2018	UAV123	
MRCF	46.931 6	33.364 1	35.062 8	38.452 8
MSCF	28.691	24.818 9	23.393 7	25.634 5
BACF	58.749	37.088 9	39.502 7	45.113 5
DSST	*140.946*	*61.745 7*	*84.639 1*	*95.776 9*
fDSST	**201.106 6**	**146.093**	**164.268 4**	**170.489**
SAMF	14.722	7.553 3	9.955 9	10.743 7
SRDCF	15.505 8	8.074 5	10.292 3	11.290 9
STRCF	28.712	20.912 9	21.209 5	23.611 5
AutoTrack	35.213 8	41.812 4	40.865 2	39.297 1
ARCF	22.156 7	21.146 3	22.44	21.914 3
DRCF	38.362 7	30.208	31.977 4	33.516 0
TRCF	24.115 7	16.790 8	23.022 3	21.309 6

注: 粗体、斜体、下划线分别表示第1、第2、第3名。

表 4　消融实验帧率对比

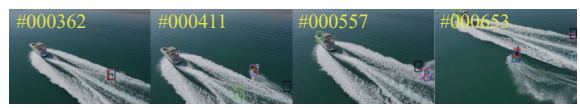
Table 4　FPS results of ablation study　　帧/s

模块	帧率			平均值
	VisDrone-SOT2018	UAV123	UAVDT	
DRCF	30.208	39.078 3	38.362 7	35.883
DRCF+Tr	27.464 7	33.701 1	41.445 6	34.203 8
DRCF+FT	25.242 4	30.748 7	38.730 3	31.573 8
DRCF+CF-VGG	19.629 8	23.057 5	29.153 9	23.947 1
DRCF+CF-VGG+FT+Tr	16.790 8	23.022 3	24.115 7	21.309 6

目标的语义信息, 该层数据输出的计算量大, 耗费时间长, 对帧率影响明显, 在 5 个数据集上的平均帧率下降了 11.935 7 帧/s。引入 3 个模块的 TRCF 目标跟踪算法计算复杂度最高, 在 VisDrone-SOT-2018 数据集上的帧率为 16.790 8 帧/s, 在 UAV123 数据集上的帧率为 23.022 3 帧/s, 在 UAVDT 数据集上的帧率为 24.115 7 帧/s, 平均帧率约为 21 帧/s, 整体仍可满足实时性的要求。需要强调的是, 本文实验均是在仅配备单 CPU 的计算机上进行, 以最大程度地模拟 UAV 运行环境。

2.4　定性分析

为了更加直观地展现 TRCF 目标跟踪算法相对于其他 UAV 目标跟踪算法的优势, 在 UAV123 数据集上进行可视化的定性分析, 结果如图 11 所示。图 11(a)为 wakeboard5 序列, 其中包括形变、部分遮挡、平面外旋转的挑战因素, 从第 362～411 帧目标被海浪部分遮挡并且伴有平面外旋转, ADTrack_v2、MSCF、MRCF 算法完全偏离目标, STRCF 及 AutoTrack 算法也发生不同程度的漂移。第 557～653 帧, 目标在运动的过程中发生旋转和形变, 只有 TRCF 目标跟踪算法能够精确跟踪目标, MRCF 算法相比于本文算法更加偏离中心, 其他 4 种算法已经远离目标中心。图 11(b)为 bird1_1 序列。从第 92～108 帧, 由于视频中数字和画质波动导致目标外观被严重干扰, 并且在目标自身也发生了平面外旋转和形变, 明显看到只有 TRCF 目标跟踪算法能够持续跟踪目标, 其他算法均发生偏离。在 uav5 序列中(见图 11(c)), 主要的挑战因素是小物体、低分辨率、尺度变化、光照变化及视角变化。从第 42 帧开始, 其他算法均已偏离目标, 并且在后续帧中目标受到剧烈的光照变化影响, 自身尺度也发生变化, TRCF 目标跟踪算法能够持续稳定地跟踪目标。图 11(d)为 person19_2 序列, 从 1 890～1 942 帧目标从视野中消失。在第 1 988 帧, 目标的一部分重新出现在视野中, 只有 TRCF 目标跟踪算法能够继续对目标进行跟踪。直到第 2001 帧, 目标再次完整出现在视野中, 其他 5 种算法已经完全丢失跟踪目标。



(a) wakeboard5 序列



(b) bird1_1 序列



(c) uav5 序列



(d) person19_2 序列

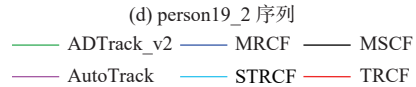


图 11　6 个先进的目标跟踪算法的跟踪结果在部分序列上的可视化对比

Fig. 11　Visualization of tracking results of six advanced object tracking algorithms on selected sequences

3　结　论

1) 引入的高效显著性目标检测算法, 有效提高了滤波器描述目标对象轮廓的能力。

2) 本文算法使用轻量级深度网络 CF-VGG 进行特征提取, 有效提高了跟踪精确度。

3) 本文算法将自适应空间正则化与时间正则化融合, 在不提高算法复杂度的前提下提高了应对复杂跟踪环境的能力。

4) 本文算法在 5 个 UAV 数据集上实现了较为优异的跟踪性能。在 UAV20L 数据集上成功率为

42.9%, 精确度为 60.6%; 在 UAV123 数据集上成功率为 48.8%, 精确度为 70.3%; 在 UAVDT 数据集上成功率为 45.4%, 精确度为 73.2%; 在 VisDrone-SOT-2018 数据集上成功率为 58.4%, 精确度为 79.9%; 在 UAVTrack112 数据集上成功率为 46.1%, 精确度为 68.0%。

在下一步工作中, 将考虑在本文所设计的框架中加入图像增强模块, 增强算法在黑暗环境跟踪目标的能力。

参考文献（References）

[1] TAHIR A, BÖLING J, HAGHBAYAN M H, et al. Swarms of unmanned aerial vehicles: a survey[J]. Journal of Industrial Information Integration, 2019, 16: 100106.

[2] YUAN C, LIU Z X, ZHANG Y M. Aerial images-based forest fire detection for firefighting using optical remote sensing techniques and unmanned aerial vehicles[J]. Journal of Intelligent & Robotic Systems, 2017, 88(2): 635-654.

[3] KANELLAKIS C, NIKOLAKOPOULOS G. Survey on computer vision for UAVs: current developments and trends[J]. Journal of Intelligent & Robotic Systems, 2017, 87(1): 141-168.

[4] JIAO L C, WANG D, BAI Y D, et al. Deep learning in visual tracking: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(9): 5497-5516.

[5] FU C H, LI B W, DING F Q, et al. Correlation filters for unmanned aerial vehicle-based aerial tracking: a review and experimental evaluation[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(1): 125-160.

[6] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2010: 2544-2550.

[7] DANELLJAN M, KHAN F S, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 1090-1097.

[8] MA C, HUANG J B, YANG X K, et al. Hierarchical convolutional features for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 3074-3082.

[9] DANELLJAN M, ROBINSON A, SHAHBAZ KHAN F, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[C]//European Conference on Computer Vision. Berlin: Springer, 2016: 472-488.

[10] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6931-6939.

[11] BHAT G, JOHNANDER J, DANELLJAN M, et al. Unveiling the power of deep tracking[C]//European Conference on Computer Vision. Berlin: Springer , 2018: 493-509.

[12] GALOOGAHI H K, SIM T, LUCEY S. Correlation filters with limited boundaries[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 4630-4638.

[13] GALOOGAHI H K, FAGG A, LUCEY S. Learning background-aware correlation filters for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 1144-1152.

[14] YUAN Y, CHEN Y M, JING Y Q, et al. FRATCF: feature-residue real-time UAV tracking based on automatic spatio-temporal regularization correlation filter[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2022: 1-6.

[15] DANELLJAN M, HÄGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 4310-4318.

[16] LUKEŽIC A, VOJÍR T, ZAJC L C, et al. Discriminative correlation filter with channel and spatial reliability[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 4847-4856.

[17] 王法胜, 尹双双, 贺冰, 等. 引入高斯掩膜的相关滤波目标跟踪算法[J]. 中国图象图形学报, 2023, 28(10): 3092-3106.
WANG F S, YIN S S, HE B, et al. A Gaussian mask-based correlation filter tracking algorithm[J]. Journal of Image and Graphics, 2023, 28(10): 3092-3106(in Chinese).

[18] LIN F L, FU C H, HE Y J, et al. ReCF: exploiting response reasoning for correlation filters in real-time UAV tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8): 10469-10480.

[19] HAN R Z, FENG W, WANG S. Fast learning of spatially regularized and content aware correlation filter for visual tracking[J]. IEEE Transactions on Image Processing, 2020, 29: 7128-7140.

[20] LI Y M, FU C H, HUANG Z Y, et al. Intermittent contextual learning for keyfilter-aware UAV object tracking using deep convolutional feature[J]. IEEE Transactions on Multimedia, 2020, 23: 810-822.

[21] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[C]//European Conference on Computer Vision. Berlin: Springer, 2016: 445-461.

[22] WEN L Y, ZHU P F, DU D W, et al. VisDrone-SOT2018: the vision meets drone single-object tracking challenge results[C]//European Conference on Computer Vision. Berlin: Springer, 2019: 469-495.

[23] FU C H, CAO Z A, LI Y M, et al. Onboard real-time aerial tracking with efficient Siamese anchor proposal network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5606913.

[24] DU D W, QI Y K, YU H Y, et al. The unmanned aerial vehicle benchmark: object detection and tracking[C]//European Conference on Computer Vision. Berlin: Springer, 2018: 375-391.

[25] WANG Y B, WANG F S, WANG C, et al. Learning saliency-aware correlation filters for visual tracking[J]. The Computer Journal, 2022, 65(7): 1846-1859.

[26] ZHU G B, WANG J Q, WU Y, et al. MC-HOG correlation tracking with saliency proposal[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016, 30(1): 3690-3696.

[27] FU C H, XU J T, LIN F L, et al. Object saliency-aware dual regularized correlation filter for real-time aerial tracking[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(12): 8940-8951.

[28] HOU X D, ZHANG L Q. Saliency detection: a spectral residual approach[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2007: 1-8.

[29] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 1597-1604.

[30] WANG N, ZHOU W G, SONG Y B, et al. Real-time correlation tracking via joint model compression and transfer[J]. IEEE Transactions on Image Processing, 2020, 29: 6123-6135.

[31] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the devil in the details: delving deep into convolutional nets[C]//Proceedings of British Machine Vision Conference. Nottingham: BMVA Press, 2014: 1-11.

[32] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 10387-10396.

[33] ROMERO A, BALLAS N, KAHOU S E, et al. FitNets: hints for thin deep nets[EB/OL].(2015-03-27)[2023-06-01]. http://arxiv.org/abs/1412.6550.

[34] DANELLJAN M, HÄGER G, KHAN F S, et al. Accurate scale estimation for robust visual tracking[C]//Proceedings of the British Machine Vision Conference. Nottingham: BMVA Press, 2014: 1-11.

[35] LI F, TIAN C, ZUO W M, et al. Learning spatial-temporal regularized correlation filters for visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4904-4913.

[36] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2013: 2411-2418.

[37] LI Y M, FU C H, DING F Q, et al. AutoTrack: towards high-performance visual tracking for UAV with automatic spatio-temporal regularization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11920-11929.

[38] LI B W, FU C H, DING F Q, et al. All-day object tracking for unmanned aerial vehicle[J]. IEEE Transactions on Mobile Computing, 2023, 22(8): 4515-4529.

[39] YE J J, FU C H, LIN F L, et al. Multi-regularized correlation filter for UAV tracking and self-localization[J]. IEEE Transactions on Industrial Electronics, 2022, 69(6): 6004-6014.

[40] ZHENG G Z, FU C H, YE J J, et al. Mutation sensitive correlation filter for real-time UAV tracking with adaptive hybrid label[C]//Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2021: 503-509.

[41] DAI K N, WANG D, LU H C, et al. Visual tracking via adaptive spatially-regularized correlation filters[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4665-4674.

[42] FU C H, JIN J, DING F Q, et al. Spatial reliability enhanced correlation filter: an efficient approach for real-time UAV tracking[J]. IEEE Transactions on Multimedia, 2021, 26: 4123-4137.

[43] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1401-1409.

[44] 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.

[45] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[C]//European Conference on Computer Vision. Berlin: Springer, 2015: 254-265.

Saliency-aware triple-regularized correlation filter algorithm for UAV object tracking

HE Bing，WANG Fasheng*，WANG Xing，SUN Fuming

(School of Information and Communication Engineering，Dalian Minzu University，Dalian 116600，China)

Abstract：Object tracking in unmanned aerial vehicle (UAV) scenarios has been widely applied in many real-world tasks. Different from general object tracking, UAV object tracking is more easily affected by complex environmental interferences and computational limitations. In this paper, a saliency-aware triple-regularized correlation filter (TRCF) for UAV object tracking was proposed. An efficient saliency detection algorithm was used to dynamically generate a dual-spatial regularizer to suppress boundary effects and apply a penalty to irrelevant background noise coefficients. Temporal regularization was introduced to address the filter degradation problem caused by appearance variation, providing a more robust appearance model. In addition, a lightweight deep network CF-VGG was employed to extract deep features, which were linearly fused with hand-crafted features to describe the object's semantic information with improved tracking accuracy. Experiments were conducted on five publicly available UAV benchmark datasets. Results show that compared to the baseline methods, the proposed method demonstrates improvements on the five datasets, proving its effectiveness and robustness. The method also shows a real-time tracking speed of about 21 frames per second, making it suitable for UAV object-tracking tasks.

Keywords：unmanned aerial vehicle object tracking；correlation filter；saliency feature map；temporal regularization；lightweight deep network

Received：2023-06-15；**Accepted**：2023-09-15；**Published Online**：2023-12-07 07：00

URL：link.cnki.net/urlid/11.2625.V.20231206.1417.005

Foundation items：National Natural Science Foundation of China (61972068,61976042); Liaoning Revitalization Talents Program (XLYC2007023)

* **Corresponding author**. E-mail：wangfasheng@dlnu.edu.cn