信息与控制　2022 年　第 51 卷　第 1 期：23 ~ 40

DOI：10.13976/j.cnki.xk.2022.1144　　文章编号：1002 - 0411(2022) - 01 - 0023 - 18

面向无人机的视觉目标跟踪算法：综述与展望

陈　琳，刘允刚

山东大学控制科学与工程学院，山东 济南　250061

基金项目：国家自然科学基金(62033007，61873146)；山东省重大科技创新工程项目(2019JZZY010433)；山东省泰山学者攀登计划

通信作者：刘允刚，lygfr@sdu.edu.cn　收稿/录用/修回：2021 - 04 - 25/2021 - 06 - 15/2021 - 07 - 23

摘要

近年来，无人机因其小巧灵活、智能自主等特点被广泛应用于民用和军事等领域中，特别是搜索侦察过程中首要的目标跟踪任务。无人机视觉目标跟踪场景的复杂性和运动目标的多变性，使得目标特征提取及模型建立困难，对目标跟踪性能带来巨大的挑战。本文首先介绍了无人机视觉目标跟踪的研究现状，梳理了经典和最新的目标跟踪算法，特别是基于相关滤波的跟踪算法和基于深度学习的跟踪算法，并对比了不同算法的优缺点。其次，归纳了常用的目标跟踪数据集和性能评价指标。最后，展望了无人机视觉目标跟踪算法的未来发展趋势。

关键词

无人机

目标跟踪

相关滤波

深度学习

中图法分类号：TP242.6

文献标识码：A

UAV Visual Target Tracking Algorithms: Review and Future Prospect

CHEN Lin, LIU Yungang

School of Control Science and Engineering, Shandong University, Jinan 250061, *China*

Abstract

In recent years, due to unmanned aerial vehicles (UAV) characteristics such as flexibility, intelligence, and autonomy, UAV has been widely used in military and civilian applications, especially the primary target tracking task in search and surveillance. However, due to the complexity of the scene environment and the variability of the moving targets in UAV visual target tracking, it is difficult to extract features and formulate a model for the target, bringing a significant challenge to the tracking performance. Therefore, we provide a review of UAV visual target tracking. First, we present the current state of researching visual target tracking, review the classical and latest tracking algorithms, especially those based on correlation filtering and deep learning, and compare the advantages and disadvantages of different algorithms. Then, we summarize the commonly-used UAV target tracking datasets and evaluation metrics. Finally, we present a prospect of the future directions for developing UAV visual target tracking algorithms.

Keywords

unmanned aerial vehicle;
target tracking;
correlation filtering;
deep learning

0　引言

近年来，无人机凭借其体积小、动作灵活及易于操控等特点，在民用、军事以及科学研究等多个领域得到越来越广泛的应用，例如，恶劣环境下的电力线路检测、大气环境检测、抢险救灾、侦察敌情、敌方目标跟踪、搜索战场情报等[1-6]。在无人机的诸多任务类型中，无人机目标跟踪有着重要的研究意义，并逐渐成为目前无人机领域热点研究方向之一[7-8]。

无人机的目标跟踪能够通过多种方式来实现，

比如，在跟踪目标上绑定电子标签或安装 GPS 追踪器来辅助跟踪，但该类方法需要与目标进行近距离接触，在实际跟踪任务中往往难以做到。随着计算机视觉技术的飞速发展，基于视觉的无人机目标跟踪已成为最接近人类行为且最为直观的跟踪形式。具体地，视觉目标跟踪是指在给定某视频初始帧中感兴趣区域(如人、车辆等)的情况下，对视频中的运动目标进行特征提取，根据提取到的目标特征来预测未来时间帧中运动目标的位置和大小，从而完成对目标的追踪任务[9-11]。

无人机视觉目标跟踪与地面目标跟踪相比，面临着 4 个挑战：1）由于空中视野广阔，干扰物体数量较多，目标与其他物体之间、目标与背景之间相互干扰，可区分性差，导致目标模型的可辨识性和排他性不高，建立精准的目标模型较困难。2）当无人机飞行在一定高度时，图像影幅变大，分辨率和清晰度变低，地面上的待跟踪目标尺度变得很小，目标特征和纹理变得稀少，使得目标特征提取困难，特征表示不显著，导致目标检测和跟踪难度变大。3）无人机在跟踪过程中易受到风力等外界因素的影响，导致相机抖动、视角变化、运动模糊等现象频繁，从而易产生跟踪漂移和丢失的情况，实现鲁棒、稳定、长时的无人机目标跟踪较为困难。4）由于无人机自身结构特点，大多数无人机仅有一个 CPU，计算资源有限，无法承受复杂度太高的运算，如何在保证精度的情况下开发复杂度低的跟踪算法是极具挑战的。随着无人机技术的发展和计算机信息处理能力的提升，尽管无人机视觉目标跟踪算法有了突破性进展，但由于上述难点的存在，无人机视觉目标跟踪算法仍有很大的发展空间。

视觉目标跟踪方法主要分为生成类跟踪方法[12-13]和判别类跟踪方法[14-16]。生成类跟踪方法通常忽略背景信息的影响且假设目标外观在一定时间内保持不变，故该方法无法处理和适应复杂的跟踪变化。判别类跟踪方法，尤其是基于相关滤波和基于深度学习的算法，在一定程度上解决了样本不足的问题，且能够提取目标中更多有用信息，显著提高目标跟踪准确率和速度。判别类跟踪算法出现之后，经典的生成类跟踪算法逐渐淡出，其主要原因是因为这些算法无法适应复杂多变的跟踪场景，其鲁棒性和准确性被前沿算法所超越。然而，由于上述目标跟踪挑战的存在，判别类跟踪算法仍存在一些不足。为了构建一个更精准、更高效且更鲁棒的通用跟踪器，未来研究应重点关注高效的在线训练和失跟后的重新检测机制，提高目标被完全遮挡后的跟踪效果，同时，应关注如何引入迁移学习和对抗学习等前沿方法来提高特征提取有效性，提高算法对低分辨率的小目标的跟踪性能，从而应用于机载无人机来完成实时跟踪任务。文[1]重点对无人机目标跟踪算法的共同框架进行了详细描述，分析了现有技术的不足，并提出了目标跟踪未来的发展方向。文[16]系统地介绍了基于生成类和判别类的视觉目标跟踪算法，但未涉及目标跟踪效果的相关评估标准，且未对无人机空中监视场景中的难点进行分析。为使相关研究人员对无人机视觉目标跟踪领域的研究有更加清晰深入的了解，本文结合无人机自身结构特点，首先对应用于无人机领域的目标跟踪算法进行了梳理总结，分析了各算法的核心思想及优缺点。其次，考虑到跟踪算法在无人机平台上的应用前景，重点对基于相关滤波的跟踪算法和基于深度学习的跟踪算法进行了详细介绍，阐述了各算法的特点和贡献，并对比了各算法的跟踪效果，图 1 为两类目标跟踪算法的分类框架图，图中所涉及的算法缩写及其全称如表 1 所示。接着，归纳了无人机目标跟踪领域的跟踪数据集和跟踪性能评价标准。最后，对全文进行了总结，并对无人机视觉目标跟踪的发展方向进行了展望。

表 1 算法缩写说明表

Tab.1 Algorithm abbreviation explanation table

算法缩写	算法全称
MOSSE[11]	minimum output sum of squared error
SAMF[17]	scale adaptive with multiple features tracker
CSK[18]	circulant structure of tracking-by-detection with kernels
DSST[19]	discriminative scale space tracker
KCF[20]	kernel correlation filter
SRDCF[21]	spatially regularized discriminative correlation filters
BACF[22]	background-aware correlation filter
STRCF[23]	spatial-temporal regularized correlation filter
ARCF[24]	aberrance repressed correlation filter
BiCF[25]	bidirectional incongruity-aware correlation filter
AutoTrack[26]	automatic spatio-temporal regularization tracker
MDNet[27]	multi-domain network
ADNet[28]	action-decision network
SiamFC[29]	fully-convolutional Siamese network
CFNet[30]	correlation filter networks
VITAL[31]	visual tracking via adversarial learning

表 1　(续)

Tab.1　(continued)

算法缩写	算法全称
DSiam [32]	dynamic Siamese network
SiamRPN [33]	Siamese region proposal network
DaSiamRPN [34]	distractor-aware Siamese region proposal networks
SiamDW [35]	deeper and wider Siamese networks

1　无人机视觉目标跟踪算法

1.1　生成类目标跟踪算法

生成类目标跟踪算法将注意力放于目标本身，忽略了图像的背景信息，比较经典的算法有卡尔曼滤波、粒子滤波、均值漂移等。

卡尔曼滤波(Kalman filter，KF)算法[36]是解决线性高斯问题的经典算法，而实际目标跟踪场景中多为非线性非高斯问题[37]。国内外学者为了实现卡尔曼滤波在实际场景中的应用，提出了很多改进算法，Azrad 等[38]设计了一种基于视觉的无人机目标跟踪系统，该系统利用卡尔曼滤波器对目标和无人机进行相对位姿估计，使无人机能稳定跟踪并悬停于目标上方。Chen 等[39]提出了一种基于卡尔曼滤波技术的自适应伺服控制方法，在不考虑目标三维速度的情况下，实现无人机对地面动态目标的跟踪。Hu 等[40]结合卡尔曼滤波和马尔可夫模型，在实现无人机目标状态实时更新的同时，对目标跟踪进行决策求解。Farahi 等[41]提出了一种概率卡尔曼滤波(probabilistic Kalman filter，PKF)算法，该算法能够根据所构造的概率图来跟踪有异常行为的物体，且能够解决部分遮挡问题。

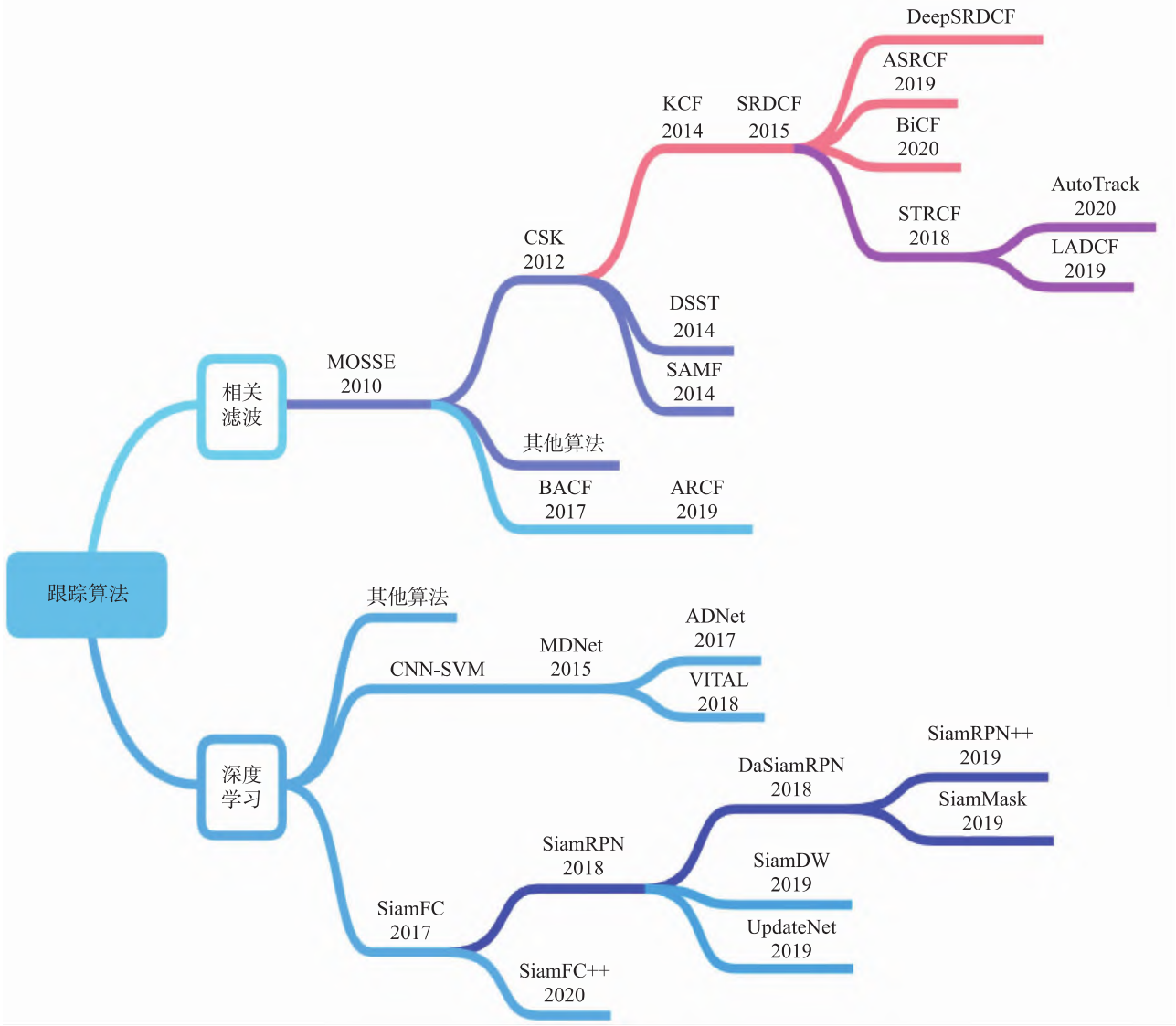

图 1　目标跟踪算法分类框架

Fig.1　Classification framework of object tracking algorithm

粒子滤波(particle filter，PF)算法[42]由 Isard 等[43]引入到目标跟踪问题中。与 KF 相关算法相比，PF 算法能够更有效地解决强非线性非高斯系统问题，但由于算法本身存在粒子退化、计算量大等问题，当跟踪目标具有较强机动性时，PF 算法跟踪效果较差。Nummiaro 等[44]将颜色特征和边缘特征引入到 PF 算法，实现了对非刚性物体的有效跟踪，但因物体的背景和颜色会随时间、光照等变化，该算法仅适用于短时稳定视频内的跟踪。Maggio 等[45]提出一种基于自适应多特征统计目标模型的跟踪算法，该算法融合颜色和方向特征，提高了目标估计的可靠性，但易受到特征退化的影响。Kamei 等[46]将自适应颜色特征融入 PF 算法来提高算法稳定性。Wang 等[47]提出一种分段多目标跟踪定位方法，利用均值漂移对当前帧的目标建模，计算其与候选目标的相似度，判断是否发生遮挡：若发生，则切换为 PF 算法进行跟踪；反之，则利用均值漂移算法。该方法能够解决无人机目标遮挡、光照变化等因素导致的失跟现象。

均值漂移(meanshift)算法[48]是经典的基于概率密度分布的方法，由 Comaniciu 等[49]引入目标跟踪领域，通过最小化目标和候选目标的概率密度函数之间的距离来跟踪目标。该算法实现简单，实时性强，具有一定的旋转不变性，但当目标被遮挡或背景杂乱时，易造成局部极大值点，使算法收敛于错误目标。Birchfield 等[50]提出了基于均值漂移的空间加权直方图算法，能够捕捉像素之间的空间关系，加强背景和目标区分度。Hwang 等[51]提出了一种分块均值漂移算法，将目标分割成多个互不遮挡的矩形块，通过跟踪每个矩形块达到对整体目标的跟踪，降低遮挡区域对跟踪的影响。Wang 等[52]在连续自适应均值漂移基础上，引入混合高斯模型背景法，将原始图像的前景和背景分割开，弱化背景色调的作用，提高了跟踪准确性和稳定性。Pareek 等[53]在融合 SURF(speeded up robust feature)和均值漂移算法的基础上，利用重投影技术来跟踪移动或静止相机拍摄的跟踪目标，该算法能够处理部分遮挡的问题，进一步降低了计算复杂度。

生成类跟踪方法通过构建当前帧目标区域的特征模型，对比目标区域与候选区域之间的误差来确定目标，该方法通常假设目标外观在一定时间内保持不变且需要大量的训练样本和计算资源，而无人机目标跟踪场景中目标外观及尺度多变，且相机抖动、运动模糊等情况时常发生，导致生成类跟踪方法在无人机目标跟踪场景中的效果不佳。

1.2 判别类目标跟踪算法

1.2.1 基于相关滤波的跟踪算法

相关滤波跟踪算法是一种在线学习方法，能够及时对模型进行更新，来适应目标的变化[54-55]。同时，为了简化运算，降低计算复杂度，相关滤波算法在滤波器求解过程中，将求解运算转换到频率域进行，显著提高了目标跟踪的速度，使算法能够达到实时跟踪的效果。在无人机空中目标跟踪中，相关滤波跟踪器的这些优点，不仅能保证跟踪模型及时进行训练和在线更新来适应物体外观变化，而且能够使无人机在单个 CPU 上实现实时跟踪功能[25-26,56-59]，节省了电能，为扩展其他辅助功能提供了可能。

基于相关滤波的跟踪算法求解过程类似，主要可分为 3 个阶段：训练阶段、模型更新阶段和检测阶段[18,20,60-64]。

首先，在训练阶段，获取目标所在的位置区域并提取目标特征，将目标特征作为训练样本对滤波器进行训练，通过求解回归方程得到当前帧图像的滤波器。大多数算法均采用岭回归方程进行求解，得到满足条件的滤波器：

$$\varepsilon(\boldsymbol{w}_f)=\left\|\sum_{c=1}^{D}\boldsymbol{w}_f^c*\boldsymbol{x}_f^c-\boldsymbol{y}\right\|^2+\lambda\sum_{c=1}^{D}\|\boldsymbol{w}_f^c\|^2 \tag{1}$$

其中，D 表示通道数，$\boldsymbol{x}_f^c$ 和 $\boldsymbol{w}_f^c$ 分别表示当前帧中第 c 通道的向量化样本信息和滤波器信息，*为循环相关操作，$\boldsymbol{y}$ 为期望的输出响应，$\boldsymbol{y}$、$\boldsymbol{x}_f^c$ 和 $\boldsymbol{w}_f^c$ 具有相同的尺寸，λ 为正则化参数。

其次，在模型更新阶段，大多数相关滤波算法会引入学习率 η，使用线性插值法对滤波器模型进行更新：

$$\boldsymbol{w}_f'=(\boldsymbol{1}-\eta)\,\boldsymbol{w}_{f-1}+\eta\boldsymbol{w}_f \tag{2}$$

最后，在检测阶段，利用得到的滤波器 $\boldsymbol{w}_f'$，计算与 $f+1$ 帧检测样本 $\boldsymbol{z}_{f+1}$ 之间的响应图，响应值越大，表示两者的相关性越大，即越接近被跟踪目标的位置。在确定目标新位置之后，再根据新的目标位置提取训练样本，接着再重复模型更新和目标检测的步骤，达到跟踪每一帧目标的效果。响应图可通过式(3)求解：

$$f(\boldsymbol{z})=\boldsymbol{F}^{-1}(\hat{\boldsymbol{w}}_f'\odot\hat{\boldsymbol{z}}_{f+1}) \tag{3}$$

其中，$\boldsymbol{F}^{-1}$ 表示傅里叶逆变换，$\odot$ 表示 Hadamard 乘积。

针对以上三步求解过程，学者们提出了很多改进的相关滤波跟踪算法，来解决目标跟踪过程中出现的由运动模糊、尺度变化、物体遮挡等因素造成

的跟踪漂移问题[17, 19, 21, 65]。基于相关滤波的跟踪算法将在第 2 节中进行详细介绍。

1.2.2　基于深度学习的跟踪算法

深度神经网络能够学习丰富的表示形式并从中提取复杂和抽象的特征，在计算机视觉、模式识别和图像处理等领域均取得了显著效果[66-71]，但由于深度网络模型的训练需要大量的数据支撑且会消耗大量内存资源，很难在单 CPU 或计算资源受限的环境中使用。随着计算机信息处理能力的提升，越来越多的学者开始关注深度网络在实时目标跟踪任务中的应用，期望同时在精度、速度和鲁棒性上获得更好的效果[27-29, 31, 72-74]。

目前，许多技术已经被用来提高深度网络跟踪器的性能(例如，空间感知采样策略、区域生成网络、数据增强策略、高效的检测和尺度估计策略等)，根据不同的硬件实现，各种各样的算法被分类为实时跟踪器。Nam 等[27]提出的 MDNet 算法将深度学习引入目标跟踪，提取目标的运动特征，能较好地区分目标和背景区域，且网络简单，能达到实时跟踪的效果。Bertinetto 等[29]提出了 SiamFC 算法，真正打破了相关滤波在跟踪领域的垄断地位，利用全卷积孪生网络对跟踪数据进行端到端的训练，结构简单。ADNet 算法[28]在 MDNet 基础上，引入深度强化学习，该算法具有较高的跟踪精度和尺度变化鲁棒性，且网络轻量，能够达到实时跟踪。CFNet 算法[30]和 DSiam 算法[32]在 SiamFC 基础上，分别结合相关滤波和动态思想，在保证跟踪准确度的同时，提高了网络实时跟踪的能力，且对目标外观变化、背景变化有较好的鲁棒性。Li 等[75]提出一种 Mate-TD3 算法，结合元学习和多任务强化学习策略，提高算法的收敛速度，实现无人机的有效控制，使无人机能够快速跟踪不确定运动的目标。精度和速度的提升，使得基于深度网络的跟踪器能够满足工业领域的应用，例如智能汽车的自主跟踪、无人机的智能跟随等。

基于 Siamese 的神经网络算法在跟踪性能和效率之间达到了很好的平衡，该网络设计的初衷就是为了使算法变得更快更准，实现算法在工业领域的应用，是目前目标跟踪领域较有前景的网络结构[33-34, 76-80]。大多基于 Siamese 网络的跟踪器均采用离线训练的方式，在跟踪过程中不对目标模型进行更新，从而提高目标跟踪速度。基于 Siamese 网络的视觉跟踪算法将在第 3 节中进行详细的介绍。

2　基于相关滤波的无人机目标跟踪

由于无人机本身特性的制约，相关滤波跟踪算法以其简单快速的优点逐渐应用于无人机领域，虽然基于相关滤波的各种跟踪算法框架类似，但每种算法均有各自的贡献和特点。本节详细介绍了 8 种具有创造性贡献的相关滤波跟踪算法，并对比了各种算法的特性及其在无人机数据集 UAV123 上的跟踪效果，如表 2 所示。

表 2　相关滤波目标跟踪算法及其在 UAV123 数据集的跟踪效果对比

Tab.2　Comparison of CF tracking algorithms and their performance conducted on UAV123

跟踪算法	出版物	特征	尺度估计	帧率 /(帧/s)	距离精度
DSST[19]	BMVC 2014	HOG	尺度滤波器	85.43	0.586
SAMF[17]	ECCV 2014	HOG + CN	尺度池	10.05	0.597
KCF[20]	TPAMI 2015	HOG	单一尺度	611.65	0.523
SRDCF[21]	ICCV 2015	HOG	尺度池	11.08	0.676
BACF[22]	ICCV 2017	HOG	尺度池	43.45	0.660
STRCF[23]	CVPR 2018	HOG + CN + Grayscale	尺度池	22.58	0.681
ARCF[24]	ICCV 2019	HOG + CN + Grayscale	尺度池	40.36	0.667
AutoTrack[26]	CVPR2020	HOG + CN + Grayscale	尺度池	48.21	0.689

由于结合深度特征的相关滤波算法跟踪速度会变慢，故表 2 中主要介绍的是使用传统特征，如 HOG(histogram of oriented gradient)、CN(color names)或灰度(Grayscale)的相关滤波算法，由表 2 可看出，SRDCF、BACF、STRCF、ARCF 和 AutoTrack 相比于以往算法，在精度上有了显著提升，但跟踪速度相对下降，其中 AutoTrack 算法在速度和精度上达到了很好的平衡。SAMF、STRCF、ARCF 和 AutoTrack 引入了多特征融合机制和尺度池估计策略，在一定程度上提高了目标的跟踪精度。所比较的相关滤波跟踪算法中，大多数算法均能达到实时跟踪(30 帧/s)的效果。

2.1 最小输出误差平方和(MOSSE)跟踪算法

相关滤波跟踪算法的核心问题就是求解满足条件的滤波器，2010 年 Bolme 等[11] 首次将相关滤波算法引入到目标跟踪领域，提出最小输出误差平方和(MOSSE)跟踪算法。该算法的目标是训练一个滤波器，使循环相关输出与期望标签之间的平方误差最小，该最小化问题可表示成:

$$\varepsilon(w) = \|w*x - y\|^2 \tag{4}$$

其中，w 为滤波器，x 为训练样本，y 为期望输出响应。

Bolme 等[11] 将该问题转化到频率域，将卷积运算转化成点积运算，在降低计算复杂度的同时大大提高了计算速度。在频率域对 w 求偏导，求得式(4)的闭式解:

$$\hat{w} = \frac{\hat{x}^* \odot \hat{y}}{\hat{x}^* \odot \hat{x}} \tag{5}$$

在模型更新阶段，该文令 $\hat{w} = \frac{A}{B}$，即将滤波器拆分成分子 A 和分母 B，两部分分别进行迭代与更新，实现更鲁棒的位置估计，η 为学习率参数，更新公式为

$$\begin{cases} A_{f,\,\mathrm{model}} = \eta A_f + (1-\eta) A_{f-1,\,\mathrm{model}} \\ B_{f,\,\mathrm{model}} = \eta B_f + (1-\eta) B_{f-1,\,\mathrm{model}} \end{cases} \tag{6}$$

当相关参数更新之后，用滤波器对新一帧中的样本 z 进行检测，得到响应图，其中响应最大值的位置即为当前帧中目标的新位置，检测阶段的响应函数表示为

$$f(z) = \boldsymbol{F}^{-1}\left(\sum_c \hat{w}_c \odot \hat{z}_c\right) \tag{7}$$

值得一提的是，MOSSE 算法在目标跟踪领域中具有里程碑的意义，它大大提高了目标跟踪的速度，能够达到每秒 600 多帧，且在光照和目标姿态的变化上具有较强的鲁棒性，减少了跟踪过程中的目标漂移现象。

2.2 核相关滤波(KCF)跟踪算法

核相关滤波跟踪算法(KCF)[44] 由 Henriques 等在 2014 年提出，算法提出后轰动一时，并为后续许多基于判别相关滤波器的跟踪算法提供了基础框架。KCF 算法引入了 HOG 特征代替原有特征，并提出了多通道样本的训练策略，将样本划分为若干个区域，并对每个区域提取了 31 维特征。该算法对光照变化等有较强的鲁棒性，跟踪速度快。

首先，在训练阶段，KCF 算法将非线性回归问题引入到高维空间，转化为线性可分问题，利用岭回归模型对滤波器进行训练，寻找一个函数 $f(x) = w*\varphi(x)$，使样本与目标标签之间的平方误差最小，用于滤波器训练的回归方程为

$$\varepsilon(w) = \left\| \sum_{c=1}^{D} w^c * \varphi(x^c) - y \right\|^2 + \lambda \sum_{c=1}^{D} \|w^c\|^2 \tag{8}$$

其中，λ 为用于控制过拟合的正则化参数。

该算法利用核技巧将 w 映射到高维空间，假设 $w = \sum_i \alpha_i \varphi(x_i)$，求得 $\boldsymbol{\alpha}$ 解的形式为 $\boldsymbol{\alpha} = (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} y$，引入循环矩阵特性，简化 $\boldsymbol{\alpha}$ 的解为

$$\hat{\boldsymbol{\alpha}} = \frac{\hat{y}}{\hat{\boldsymbol{k}}^{xx} + \lambda \boldsymbol{I}_N} \tag{9}$$

其中，$\boldsymbol{\alpha}$ 为系数 α_i 的向量形式，$\boldsymbol{K}$ 为核函数的矩阵表示，$\boldsymbol{I}$ 和 $\boldsymbol{I}_N$ 均为单位矩阵，$\hat{k}^{xx}$ 指的是两个向量之间的核相关。

KCF 算法讨论了两种常用的核相关函数，即多项式核函数和高斯核函数，假设两个单特征通道向量为 x 和 x'，他们的两种核相关矩阵计算方式分别为

$$\begin{cases} \boldsymbol{k}_{\mathrm{d}}^{xx'} = \left(\boldsymbol{F}^{-1}\left(\sum \hat{\boldsymbol{x}}^* \odot \hat{\boldsymbol{x}}'\right) + \boldsymbol{a}\right)^b \\ \boldsymbol{k}_{\mathrm{g}}^{xx'} = \exp\left(-\frac{1}{\sigma^2}\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{x}'\|^2 - 2\boldsymbol{F}^{-1}\left(\sum \hat{\boldsymbol{x}}^* \odot \hat{\boldsymbol{x}}'\right)\right)\right) \end{cases} \tag{10}$$

当训练样本为多通道时，式(9)、式(10)可推广为

$$\begin{cases} \boldsymbol{k}_{\mathrm{d}}^{xx'} = \left(\boldsymbol{F}^{-1}\left(\sum_c \hat{\boldsymbol{x}}_c^* \odot \hat{\boldsymbol{x}}_c'\right) + \boldsymbol{a}\right)^b \\ \boldsymbol{k}_{\mathrm{g}}^{xx'} = \exp\left(-\frac{1}{\sigma^2}\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{x}'\|^2 - 2\boldsymbol{F}^{-1}\left(\sum_c \hat{\boldsymbol{x}}_c^* \odot \hat{\boldsymbol{x}}_c'\right)\right)\right) \end{cases} \tag{11}$$

对于线性核，其表达式可表示成

$$\boldsymbol{k}_1^{xx'} = \boldsymbol{F}^{-1}\left(\sum_c \hat{\boldsymbol{x}}_c^* \odot \hat{\boldsymbol{x}}_c'\right) \tag{12}$$

接着，在模型更新阶段，通过更新滤波器参数 $\boldsymbol{\alpha}$ 和样本 x 对当前帧模型进行更新。更新表达式:

$$\begin{cases} \hat{\boldsymbol{\alpha}}_{f,\,\mathrm{model}} = (1-\eta)\,\hat{\boldsymbol{\alpha}}_{f-1,\,\mathrm{model}} + \eta \hat{\boldsymbol{\alpha}}_f \\ \boldsymbol{x}_{f,\,\mathrm{model}} = (1-\eta)\,\boldsymbol{x}_{f-1,\,\mathrm{model}} + \eta \boldsymbol{x}_f \end{cases} \tag{13}$$

其中，η 为学习率参数。

最后，利用更新后的参数进行检测，得到响应函数，根据得到的响应函数，推导出响应最大值点，再根据响应最大值点的信息重新训练滤波器，更新模型，从而达到准确追踪目标的效果。其响应函数为

$$f(z) = \boldsymbol{F}^{-1}\left(\sum_c \hat{\boldsymbol{k}}^{xz} \odot \hat{\boldsymbol{\alpha}}\right) \tag{14}$$

相比其他算法，KCF 在精度和速度上均得到了

较大的提升，该算法利用核技巧和循环矩阵对角化等特性，使计算过程和实现变得简单，大大降低了计算复杂度，促进了相关滤波跟踪算法的发展。然而，该算法依赖于循环矩阵，对于多尺度的目标跟踪效果并不理想。

2.3　判别尺度空间(DSST)跟踪算法

为了解决复杂场景中因目标尺度变化而导致的跟踪漂移现象，Danelljan 等在 2014 年基于 MOSSE 算法提出了判别尺度空间(DSST)算法[19]。该算法在引入多特征融合机制的同时，增加了尺度滤波器，通过分别训练位置和尺度两个不同的滤波器对目标进行准确跟踪。具体地，在新一帧中，首先利用位置滤波器搜索目标位置，使用以预测位置为中心的样本训练位置滤波器，然后以目标位置为中心提取不同尺度的样本用于训练尺度滤波器，从而确定目标的位置和尺度大小。

DSST 算法同样使用岭回归模型对滤波器进行训练，其回归方程与 KCF 算法相似:

$$\varepsilon=\left\|\sum_{l=1}^{d}\boldsymbol{f}^{l}*\boldsymbol{x}^{l}-\boldsymbol{g}\right\|^{2}+\lambda\sum_{l=1}^{d}\|\boldsymbol{f}^{l}\|^{2} \tag{15}$$

其中，$\boldsymbol{f}^{l}$、$\boldsymbol{x}^{l}$ 分别表示第 l 个通道的滤波器和样本，$\boldsymbol{g}$ 表示期望的相关输出，$\boldsymbol{f}^{l}$、$\boldsymbol{x}^{l}$ 和 $\boldsymbol{g}$ 具有相同的尺寸和大小。

将问题转换到傅里叶域中求解，可得到使式(15)最小的滤波器为

$$\boldsymbol{F}^{l}=\frac{\overline{\boldsymbol{G}}\boldsymbol{X}^{l}}{\sum_{k=1}^{d}\overline{\boldsymbol{X}}^{k}\boldsymbol{X}^{k}+\lambda},\quad l=1,\cdots,d \tag{16}$$

其中，$\boldsymbol{F}$、$\boldsymbol{G}$ 和 $\boldsymbol{X}$ 分别为 f、g 和 x 向量在傅里叶域中对应的表示形式；$\overline{\boldsymbol{G}}$ 为 $\boldsymbol{G}$ 的共轭形式。

为了降低在线学习任务的计算复杂度，该算法采用将相关滤波器 $\boldsymbol{F}^{l}$ 的分子(设为 $\boldsymbol{A}_{t}^{l}$)和分母(设为 $\boldsymbol{B}_{t}$)分别更新的方式，来获得一个较鲁棒的近似值。更新方程为

$$\begin{cases}\boldsymbol{A}_{t}^{l}=(1-\eta)\boldsymbol{A}_{t-1}^{l}+\eta\overline{\boldsymbol{G}}_{t}\boldsymbol{X}_{t}^{l}\\ \boldsymbol{B}_{t}=(1-\eta)\boldsymbol{B}_{t-1}+\eta\sum_{k=1}^{d}\overline{\boldsymbol{X}}_{t}^{k}\boldsymbol{X}_{t}^{k}\end{cases} \tag{17}$$

其中，t 表示第 t 帧，$t-1$ 表示第 $t-1$ 帧.

最后，对样本 $\boldsymbol{Z}$ 进行检测，得到响应输出，再根据响应最大值点的信息找到新的目标状态，响应函数为

$$\boldsymbol{y}=\boldsymbol{F}^{-1}\left(\frac{\sum_{l=1}^{d}\overline{\boldsymbol{A}}^{l}\boldsymbol{Z}^{l}}{\boldsymbol{B}+\lambda}\right) \tag{18}$$

在目标跟踪领域，目标尺度估计是最具挑战的问题之一。DSST 算法能较好地处理尺度变化问题，算法简洁，可移植性高。另外，SAMF 算法[17]也是解决尺度问题中较成功的跟踪算法，该算法结合 HOG、CN 和灰度特征，利用尺度池实现了目标的自适应跟踪。

2.4　空间正则相关滤波(SRDCF)跟踪算法

为了简化计算过程，上述目标函数的求解均通过傅里叶变换转换到频域进行，但在变换过程中，需对图像窗口进行循环拼接，这将导致拼接图像在边缘处不连续，从而产生边缘效应，影响目标跟踪的质量。

为了更好地解决边缘效应，Danelljan 等提出了空间正则相关滤波(SRDCF)算法[21]，引入空间正则化分量，利用正则化权重对训练过程中的滤波器系数进行惩罚，从而生成一个判别性更强的模型。其改进后的优化问题可表示为

$$\varepsilon=\sum_{k=1}^{t}\alpha_{k}\left\|\sum_{l=1}^{d}\boldsymbol{x}_{k}^{l}*\boldsymbol{f}_{k}^{l}-\boldsymbol{g}_{k}\right\|^{2}+\sum_{l=1}^{d}\|w\cdot\boldsymbol{f}^{d}\|^{2} \tag{19}$$

其中，$\alpha_{k}\geqslant 0$ 为每个训练样本的权重系数，w 为正则化权重系数，t 表示样本数量，d 为特征图维数。式(19)的第 2 项为空间正则化项。

利用帕萨瓦尔定理与循环矩阵性质，简化式(19):

$$\tilde{\varepsilon}(\tilde{\boldsymbol{f}})=\sum_{k=1}^{t}\alpha_{k}\|\boldsymbol{D}_{k}\tilde{\boldsymbol{f}}-\tilde{\boldsymbol{g}}_{k}\|^{2}+\|\boldsymbol{W}\tilde{\boldsymbol{f}}\|^{2} \tag{20}$$

其中，$\boldsymbol{W}$ 为块对角矩阵，$\boldsymbol{D}_{k}=[\boldsymbol{D}_{k}^{1},\cdots,\boldsymbol{D}_{k}^{d}]$，$\boldsymbol{D}_{k}^{l}$ 为对角矩阵，其对角线元素由向量 $\boldsymbol{x}_{k}^{l}$ 组成.

接着，将式(20)化解为 $\boldsymbol{A}_{t}\tilde{\boldsymbol{f}}=\tilde{\boldsymbol{b}}_{t}$ 的形式，从而求得满足条件的最小值，其中:

$$\begin{cases}\boldsymbol{A}_{t}^{l}=\sum_{k=1}^{t}\alpha_{k}\boldsymbol{D}_{k}^{\mathrm{T}}\boldsymbol{D}_{k}+\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\\ \tilde{\boldsymbol{b}}_{t}=\sum_{k=1}^{t}\alpha_{k}\boldsymbol{D}_{k}^{\mathrm{T}}\tilde{\boldsymbol{g}}_{k}\end{cases} \tag{21}$$

在模型更新阶段，通过更新 $\boldsymbol{A}_{t}$ 和 $\tilde{\boldsymbol{b}}_{t}$ 两项对滤波器进行更新，该算法引入了学习率 $\eta\geqslant 0$，更新公式为

$$\begin{cases}\boldsymbol{A}_{t}=(1-\eta)\boldsymbol{A}_{t-1}+\eta(\boldsymbol{D}_{t}^{\mathrm{T}}\boldsymbol{D}_{t}+\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})\\ \tilde{\boldsymbol{b}}_{t}=(1-\eta)\tilde{\boldsymbol{b}}_{t-1}+\eta\boldsymbol{D}_{t}^{\mathrm{T}}\tilde{\boldsymbol{g}}_{t}\end{cases} \tag{22}$$

在检测阶段，求取更新后的滤波器参数与样本之间的响应，响应的最大值点即为新的目标位置，该算法的响应函数与前述算法相似。

之后，Danelljan 等[63]又在 SRDCF 算法上进行了改进，提出了 SRDCFdecon 算法，该算法针对在线跟踪时训练样本易损坏的问题，利用目标模型与训练样本权值联合学习的方式，使用于滤波器训练的样本权值不再是固定的，而是根据回归方程动态

求解，保证了高置信度帧的影响，也减少了污染样本的可能性。

2.5 时空正则相关滤波(STRCF)跟踪算法

SRDCF 在引入空间正则的同时利用多个训练样本对滤波器进行训练，这不可避免地大大增加了计算复杂度，影响了实时跟踪效果。鉴于此，Li 等[23]将时间正则化引入单样本 SRDCF 算法中，提出了时空正则相关滤波(STRCF)算法，该算法提出利用交替方向乘子法(ADMM)求解方程组，在提高算法跟踪精度的同时也显著提升了算法的跟踪速度。

STRCF 算法在 SRDCF 基础上引入时间正则化，并采用与其相似的回归方程:

$$\varepsilon=\frac{1}{2}\left\|\sum_{d=1}^{D}\boldsymbol{x}_t^d*\boldsymbol{f}^d-\boldsymbol{g}\right\|^2+\frac{1}{2}\sum_{d=1}^{D}\|w\cdot\boldsymbol{f}^d\|^2+\frac{\mu}{2}\|\boldsymbol{f}-\boldsymbol{f}_{t-1}\|^2 \tag{23}$$

其中，$\|\boldsymbol{f}-\boldsymbol{f}_{t-1}\|^2$ 为时间正则化项，$\boldsymbol{f}_{t-1}$为 $t-1$ 帧时的滤波器信息，μ 为时间正则化系数。

为了提高目标跟踪效率，加快问题求解速度，该文将原问题分解为多个子问题，假设 $\boldsymbol{f}=\boldsymbol{q}$，并引入步长参数 γ 和拉格朗日乘子 s，令 $\boldsymbol{h}=\frac{1}{\gamma}s$，利用 ADMM 算法进行求解，子问题求解公式为

$$\begin{cases}\boldsymbol{f}^{(i+1)}=\operatorname{argmin}\left\|\sum_{d=1}^{D}\boldsymbol{x}_t^d*\boldsymbol{f}^d-\boldsymbol{g}\right\|^2+\gamma\|\boldsymbol{f}-\boldsymbol{q}+\boldsymbol{h}\|+\mu\|\boldsymbol{f}-\boldsymbol{f}_{t-1}\|^2\\ \boldsymbol{q}^{(i+1)}=\operatorname{argmin}\sum_{d=1}^{D}\|w\cdot\boldsymbol{q}^d\|^2+\gamma\|\boldsymbol{f}-\boldsymbol{q}+\boldsymbol{h}\|^2\\ \boldsymbol{h}^{(i+1)}=\boldsymbol{h}^{(i)}+\boldsymbol{f}^{(i+1)}-\boldsymbol{q}^{(i+1)}\end{cases} \tag{24}$$

上述每个子问题都有封闭形式的解，且能够在较少的迭代次数中快速收敛。相比 SRDCF 算法，由于引入了时间正则项，在训练阶段不再需要之前所有 T 帧的信息，从而大大减少了训练所需的内存，提高了计算速度，达到实时跟踪的效果(30 帧/s)。值得一提的是，STRCF 算法具有较强的时间稳定性，在目标快速运动和较大外观变化情况下，能够提供更鲁棒的外观模型。

2.6 背景感知相关滤波(BACF)跟踪算法

由于此前的相关滤波算法只对目标进行建模，忽略了目标背景随时间变化的情况，这将会影响目标跟踪效果。为了解决该问题，Galoogahi 等[22]提出了背景感知相关滤波(BACF)算法，该算法利用 HOG 特征，随时间变化动态地对跟踪目标的前景和背景进行建模，同时提出了一种基于 ADMM 的优化方法对滤波器进行求解。最终通过实验，证明了算法具有更高的准确性和实时性。

该算法中多通道背景感知相关滤波器的训练公式为

$$\varepsilon(\boldsymbol{f})=\frac{1}{2}\sum_{j=1}^{T}\left\|\boldsymbol{g}(j)-\sum_{k=1}^{K}\boldsymbol{f}_k^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x}_k\left[\Delta\tau_j\right]\right\|^2+\frac{\lambda}{2}\sum_{k=1}^{K}\|\boldsymbol{f}_k\|^2 \tag{25}$$

其中，$[\Delta\tau_j]$ 为循环移位算子，$\boldsymbol{x}_k[\Delta\tau_j]$ 表示对 $\boldsymbol{x}_k$ 的 j 阶离散循环移位，λ 为正则化系数，T 是 $\boldsymbol{x}_k$ 的长度，$\boldsymbol{P}$ 是裁剪矩阵，用来选取 $\boldsymbol{x}_k$ 的中间固定长度的元素。

同样，为了提高计算效率，作者将滤波器的训练学习放在频率域进行，即可将式(25)化成:

$$\begin{aligned}&\varepsilon(\boldsymbol{f},\hat{\boldsymbol{q}})=\frac{1}{2}\|\hat{\boldsymbol{g}}-\hat{\boldsymbol{X}}\hat{\boldsymbol{q}}\|^2+\frac{\lambda}{2}\|\boldsymbol{f}\|^2\\ &\text{s.t.}\quad\hat{\boldsymbol{q}}=\sqrt{T}(\boldsymbol{F}\boldsymbol{P}^{\mathrm{T}}\otimes\boldsymbol{I}_K)\boldsymbol{f}\end{aligned} \tag{26}$$

其中，$\hat{\boldsymbol{q}}$ 为辅助变量，$\hat{\boldsymbol{X}}$ 矩阵为分块对角矩阵，被定义为 $\hat{\boldsymbol{X}}=[\operatorname{diag}^{\mathrm{T}}(\hat{\boldsymbol{x}}_1),\cdots,\operatorname{diag}^{\mathrm{T}}(\hat{\boldsymbol{x}}_K)]$，$\boldsymbol{F}$ 为标准正交矩阵，用于将其他向量化信号映射到傅里叶域，⊗指的是克罗内克积。

为了使用 ADMM 算法技巧对问题进行迭代求解，该文利用增广拉格朗日算法来变形式(26)，然后分解成多个子问题进行求解:

$$\begin{aligned}\boldsymbol{L}(\boldsymbol{f},\hat{\boldsymbol{q}},\hat{\boldsymbol{\zeta}})=&\frac{1}{2}\|\hat{\boldsymbol{g}}-\hat{\boldsymbol{X}}\hat{\boldsymbol{q}}\|^2+\frac{\lambda}{2}\|\boldsymbol{f}\|^2+\\ &\hat{\boldsymbol{\zeta}}^{\mathrm{T}}(\hat{\boldsymbol{q}}-\sqrt{T}(\boldsymbol{F}\boldsymbol{P}^{\mathrm{T}}\otimes\boldsymbol{I}_K)\boldsymbol{f})+\\ &\frac{\mu}{2}\left\|\hat{\boldsymbol{q}}-\sqrt{T}(\boldsymbol{F}\boldsymbol{P}^{\mathrm{T}}\otimes\boldsymbol{I}_K)\boldsymbol{f}\right\|^2\end{aligned} \tag{27}$$

其中，$\hat{\boldsymbol{\zeta}}=[\hat{\boldsymbol{\zeta}}_1^{\mathrm{T}},\hat{\boldsymbol{\zeta}}_2^{\mathrm{T}},\cdots,\hat{\boldsymbol{\zeta}}_K^{\mathrm{T}}]^{\mathrm{T}}$ 为傅里叶域的拉格朗日矢量，$\boldsymbol{I}_K$ 为单位矩阵，μ 为惩罚因子。

BACF 算法在之前相关滤波框架上扩大了循环矩阵的采样区域，并利用裁剪矩阵将样本裁剪为多个与目标大小相同的小样本，即这些小样本均是通过循环移位和裁剪生成，包括更大范围的搜索区域和真实背景，显著提高了用来训练滤波器的样本的数量和质量，且减弱了边缘效应的影响，BACF 算法框架成为相关滤波跟踪算法历史上的又一个里程碑.

2.7 抑制畸变相关滤波(ARCF)跟踪算法

相关滤波跟踪算法的研究为视觉跟踪在无人机上的应用做出了巨大贡献，但由于无人机跟踪场景的复杂性和不确定性，通常在飞行过程中会遇到较大的环境干扰，加上边缘效应的存在，这将会导致在跟踪过程中融入更多的背景噪声，从而影响目标

跟踪的效果。

针对这一问题，Huang 等[24]提出了抑制畸变相关滤波(ARCF)算法，作者认为在目标检测过程中，由于背景环境、边缘效应和目标外观变化等噪声的影响，会导致生成的响应图在某些位置产生畸变或异常，而响应图中的信息在一定程度上反映了外观模型与当前帧实际目标之间的相似性，畸变的响应图会影响目标准确位置的判断。该算法在训练过程中的损失函数方程为

$$\varepsilon(\boldsymbol{f}_k)=\frac{1}{2}\Big\|\boldsymbol{g}-\sum_{d=1}^{D}\boldsymbol{B}\boldsymbol{x}_k^d*\boldsymbol{f}_k^d\Big\|^2+\frac{\lambda}{2}\sum_{d=1}^{D}\|\boldsymbol{f}_k^d\|^2+\frac{\gamma}{2}\Big\|\sum_{d=1}^{D}(\boldsymbol{B}\boldsymbol{x}_{k-1}^d*\boldsymbol{f}_{k-1}^d)\quad[\psi_p,\psi_q]-\sum_{d=1}^{D}\boldsymbol{B}\boldsymbol{x}_k^d*\boldsymbol{f}_k^d\Big\|^2 \tag{28}$$

其中，$\boldsymbol{B}$ 为裁剪矩阵，D 表示通道数，k 和 $k-1$ 表示第 k 帧和第 $k-1$ 帧，公式第 3 项为限制响应图变化率的正则化项，λ 和 γ 均为正则化参数，p 和 q 表示二维空间响应图中两个峰的位置差，$[\psi_p,\psi_q]$ 表示两个峰重合的平移过程。通常当畸变发生时，相邻两帧响应图的相似度会下降，第 3 项的值会变高，通过判断该值的大小，可以容易地识别畸变的发生。

在最小化整体目标的过程中，为了提高计算效率，将式(28)转化到频率域，简化运算复杂度:

$$\hat{\varepsilon}(\boldsymbol{f}_k,\hat{\boldsymbol{q}}_k)=\frac{1}{2}\|\hat{\boldsymbol{g}}-\hat{\boldsymbol{X}}_k\hat{\boldsymbol{q}}_k\|^2+\frac{\lambda}{2}\|\boldsymbol{f}_k\|^2+\frac{\gamma}{2}\|\hat{\boldsymbol{M}}_{k-1}^s-\hat{\boldsymbol{X}}_k\hat{\boldsymbol{q}}_k\|^2 \tag{29}$$

$$\text{s.t.}\quad\hat{\boldsymbol{q}}_k=\sqrt{N}(\boldsymbol{I}_D\otimes\boldsymbol{F}\boldsymbol{B}^{\mathrm{T}})\boldsymbol{f}_k$$

其中，$\hat{\boldsymbol{q}}_k$ 为辅助变量，N 为 $\boldsymbol{x}_k$ 的长度，$\hat{\boldsymbol{M}}_{k-1}^s$ 表示移位信号的离散傅里叶变换，由于已经生成了前一帧的响应图，因此该量可视为常数信号.

与 BACF 算法相似，首先将式(29)写成增广拉格朗日形式，再采用 ADMM 将问题分解成多个子问题进行迭代求解，其增广拉格朗日形式为

$$\boldsymbol{L}(\boldsymbol{f}_k,\hat{\boldsymbol{q}}_k,\hat{\boldsymbol{\zeta}})=\frac{1}{2}\|\hat{\boldsymbol{g}}-\hat{\boldsymbol{X}}_k\hat{\boldsymbol{q}}_k\|^2+\frac{\lambda}{2}\|\boldsymbol{f}_k\|^2+\frac{\gamma}{2}\|\hat{\boldsymbol{M}}_{k-1}^s-\hat{\boldsymbol{X}}_k\hat{\boldsymbol{q}}_k\|^2+\hat{\boldsymbol{\zeta}}^{\mathrm{T}}(\hat{\boldsymbol{q}}_k-\sqrt{N}(\boldsymbol{I}_D\otimes\boldsymbol{F}\boldsymbol{P}^{\mathrm{T}})\boldsymbol{f}_k)+\frac{\mu}{2}\|\hat{\boldsymbol{q}}_k-\sqrt{N}(\boldsymbol{I}_D\otimes\boldsymbol{F}\boldsymbol{P}^{\mathrm{T}})\boldsymbol{f}_k\|^2 \tag{30}$$

其中，$\hat{\zeta}$ 为拉格朗日乘子，μ 为惩罚因子，$\boldsymbol{I}_D$ 为单位矩阵。

实验证明，ARCF 算法在大多数无人机数据集上的效果都非常显著，它将提取的背景样本当作负样本用于模型训练和检测阶段，同时引入了限制响应图变化率的正则化项，避免了响应图的突然变动，提高了跟踪算法的鲁棒性和准确性。

2.8　自动时空正则化(AutoTrack)跟踪算法

自动时空正则化(Automatic spatio-temporal regularization Tracker，AutoTrack)跟踪算法[26]也是在无人机场景下改进的目标跟踪算法。Li 等[26]在 STRCF 算法的基础上，提出了一种在线自动和自适应学习相关参数的方法，能够自动地对预定义参数进行学习调整，同时引入了局部响应图和全局响应图，分别动态地对空间权值和时间权值进行调整，能够有效利用隐藏在响应图中的局部和全局信息。AutoTrack 跟踪算法联合优化滤波器和时间正则化项的目标函数为

$$\varepsilon(\boldsymbol{f}_t,\theta_t)=\frac{1}{2}\Big\|\boldsymbol{g}-\sum_{k=1}^{K}\boldsymbol{x}_t^k*\boldsymbol{f}_t^k\Big\|^2+\frac{1}{2}\sum_{k=1}^{K}\|\tilde{u}\odot\boldsymbol{f}_t^k\|^2+\frac{\theta_t}{2}\sum_{k=1}^{K}\|\boldsymbol{f}_t^k-\boldsymbol{f}_{t-1}^k\|^2+\frac{1}{2}\|\theta_t-\tilde{\theta}\|^2 \tag{31}$$

其中，$\tilde{u}$ 表示自动空间正则化参数，$\tilde{\theta}$ 和 θ_t 分别表示期望的时间正则化参数和待优化的时间正则化参数。

与前述相关滤波算法求解过程相似，文[26]引入了一个辅助变量 $\hat{q}$，将式(31)转化到频率域进行求解，并表示成增广拉格朗日形式，再通过 ADMM 进行求解:

$$\begin{aligned}&\varepsilon(\boldsymbol{f}_t,\theta_t,\hat{\boldsymbol{q}},\hat{\zeta}_t)\\&=\frac{1}{2}\Big\|\hat{\boldsymbol{g}}-\sum_{k=1}^{K}\boldsymbol{x}_t^k\odot\hat{\boldsymbol{q}}_t^k\Big\|^2+\frac{1}{2}\sum_{k=1}^{K}\|\tilde{u}\odot\boldsymbol{f}_t^k\|^2+\frac{\theta_t}{2}\sum_{k=1}^{K}\|\hat{\boldsymbol{q}}_t^k-\hat{\boldsymbol{q}}_{t-1}^k\|^2+\frac{1}{2}\|\theta_t-\tilde{\theta}\|^2+\\&\frac{\mu}{2}\sum_{k=1}^{K}\|\hat{\boldsymbol{q}}_t^k-\sqrt{N}\boldsymbol{F}\boldsymbol{f}_t^k\|^2+\\&\sum_{k=1}^{K}(\hat{\boldsymbol{q}}_t^k-\sqrt{N}\boldsymbol{F}\boldsymbol{f}_t^k)^{\mathrm{T}}\hat{\zeta}_t^k\end{aligned} \tag{32}$$

其中，$\hat{\zeta}$ 为拉格朗日乘子，N 为 $\boldsymbol{x}_k$ 的长度，μ 为惩罚因子，$\boldsymbol{F}$ 为标准正交矩阵。

通过利用 ADMM 将增广拉格朗日形式的问题分解成多个子问题进行迭代求解，由于该问题是凸的，故可以通过 ADMM 最小化来得到全局最优解。

该文提出了一种自动和自适应学习时空正则化参数的方法，引入空间局部响应变化作为空间正则化，使滤波器专注于学习目标的可信部分，同时利用全局响应变化来决定滤波器的更新速率，提高了算法的稳定性。通过大量对比实验证明了所提算法对复杂多变的无人机场景有非常强的鲁棒性。另外，文[25]还提出一种双向不一致感知相关滤波器

(BiCF)，通过将基于响应的双向不一致误差集成到CF中，达到有效学习目标外观变化的目的，使无人机能够在跟踪过程中解决目标和背景外观变化导致的失跟现象。

2.9 算法总结

上述介绍的基于相关滤波的目标跟踪算法中，虽然每种算法均有各自的贡献和特点，但均为在线学习算法，能够及时对模型进行更新，来适应目标的变化，且该类算法将求解运算转换到频率域进行，显著地降低了计算复杂度，提高了目标跟踪实时性。相关滤波跟踪算法的这些优点能够保证无人机跟踪模型及时进行训练和在线更新，提取目标有效特征，同时，能够使无人机在单个CPU上实现实时跟踪功能。

然而，基于相关滤波的目标跟踪算法很难有效处理快速尺度变化、运动物体遮挡、低分辨率、小目标等跟踪问题，当目标经历旋转引起快速尺度变化时，算法无法及时适应外观变化，从而做出错误的尺度估计；当运动目标被完全遮挡，再次出现在跟踪搜索区域之外时，算法很难准确预测目标的位置，最终导致跟踪失败；低分辨率和小目标的物体会导致训练样本不足，降低滤波器将目标从背景中区分出来的能力，从而容易导致跟踪失败。

3 基于深度学习的无人机目标跟踪

近年来，随着深度神经网络的发展，越来越多的学者开始关注深度网络在实时目标跟踪任务中的应用，并且取得了一定的研究成果。本节对Siamese系列的目标跟踪算法进行了详细介绍，并对比了各算法的网络配置及速度等特性，如表3所示。

表3 深度学习目标跟踪算法及其配置对比

Tab.3 Comparison of tracking algorithms based on deep learning

跟踪算法	出版物	主干网	帧率 /(帧/s)	CPU/GPU	编程语言
SiamFC [29]	ECCVW 2016	AlexNet	58	Intel I7-4790K 4.00GHz CPU/GTX Tian X GPU	Matlab
SiamRPN [33]	CVPR 2018	AlexNet	160	Intel I7 CPU，12GB RAM/GTX 1060 GPU	Python
DaSiamRPN [34]	ECCV 2018	AlexNet	160	Intel I7 CPU，48GB RAM/GTX Tian X GPU	Python
SiamRPN ++ [30]	CVPR 2019	ResNet-50	35	N/A/Titan Xp Pascal GPU	Python
SiamDW [35]	CVPR 2019	ResNet，ResNextInception	–	Intel Xeon 2.40GHz CPU/GTX 1080 GPU	Python
SiamFC ++ [31]	AAAI 2020	GoogleNet	90	N/A/NVIDIA RTX 2080Ti GPU	Python

由表3可看出，深度目标跟踪算法对硬件配置是有要求的，需要有GPU的支持，且在有GPU的环境下，各算法均能达到实时跟踪的效果。为了提高跟踪精度，各算法的主干网络逐渐由简单的AlexNet向复杂的ResNet、Inception等网络过渡，在提高精度的同时也增加了网络复杂度，对算力有了更高的要求，目前，深度目标跟踪算法的编程语言主要以Python为主。接下来，对算法进行详细介绍。

3.1 全卷积孪生网络(SiamFC)跟踪算法

全卷积孪生网络(Fully-Convolutional Siamese network，SiamFC)算法可以看作是Siamese系列跟踪论文的开篇之作。为了解决深度学习在目标跟踪领域中的数据缺乏和实时性差的问题，Bertinetto等提出了SiamFC算法[29]，该算法真正打破了相关滤波在跟踪领域的垄断地位，利用全卷积孪生网络对跟踪数据进行端到端的训练，结构简单，实时性强，其跟踪算法框架如图2所示。

SiamFC网络框架中包含两个权重共享的分支，其中，$z \in \mathrm{R}^{127\times127\times3}$为样本图像，相当于要跟踪的目标，$x \in \mathrm{R}^{255\times255\times3}$为搜索图像，需要做的是在搜索图像中找到目标所在位置。具体地，首先分别用相同的网络φ对样本图像和搜索图像进行特征提取，然后对两者的特征图进行互相关操作，生成响应图，即利用卷积运算提取两种图像中最相似的部位，响应最高的位置就最有可能是跟踪目标的位置。

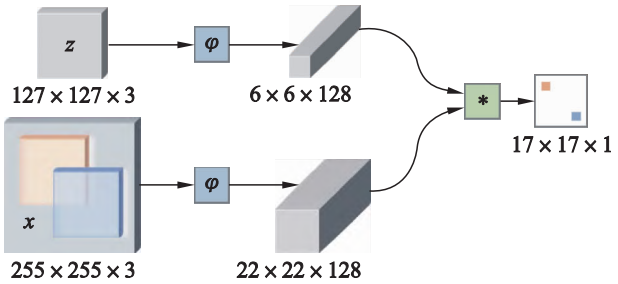

图2 SiamFC跟踪算法框架

Fig.2 SiamFC tracking algorithm framework

该算法选用简单的AlexNet网络进行特征提取，网络没有加入padding操作，并且只有5层卷积，每层卷积之后(除了第5层卷积)都使用了ReLU层。

SiamFC 算法采用离线训练方式对深度网络进行训练学习，并且在跟踪过程中不对模型进行更新也不保留历史外观信息，尽管该算法很简单，但它兼容了精度和速度，取得了较为满意的结果。

3.2　孪生候选区域生成网络(SiamRPN)跟踪算法

由于深度学习跟踪器往往具有较复杂的网络结构，导致模型在训练和检测过程中需要消耗大量的内存资源，很难达到实时跟踪的效果。尽管离线训练提高了目标跟踪的速度，但跟踪阶段模型的不更新在一定程度上影响了跟踪的精度，不能达到速度和精度兼备。

为了解决该问题，Li 等[33]提出了孪生候选区域生成网络(Siamese Region Proposal Network，SiamRPN)跟踪算法，利用大规模的图像对进行端到端的离线训练，并且于在线跟踪阶段使用单样本检测方式，避免了在线微调和多尺度测试，在提高精度的同时提高了算法速度。该跟踪算法框架如图 3 所示。

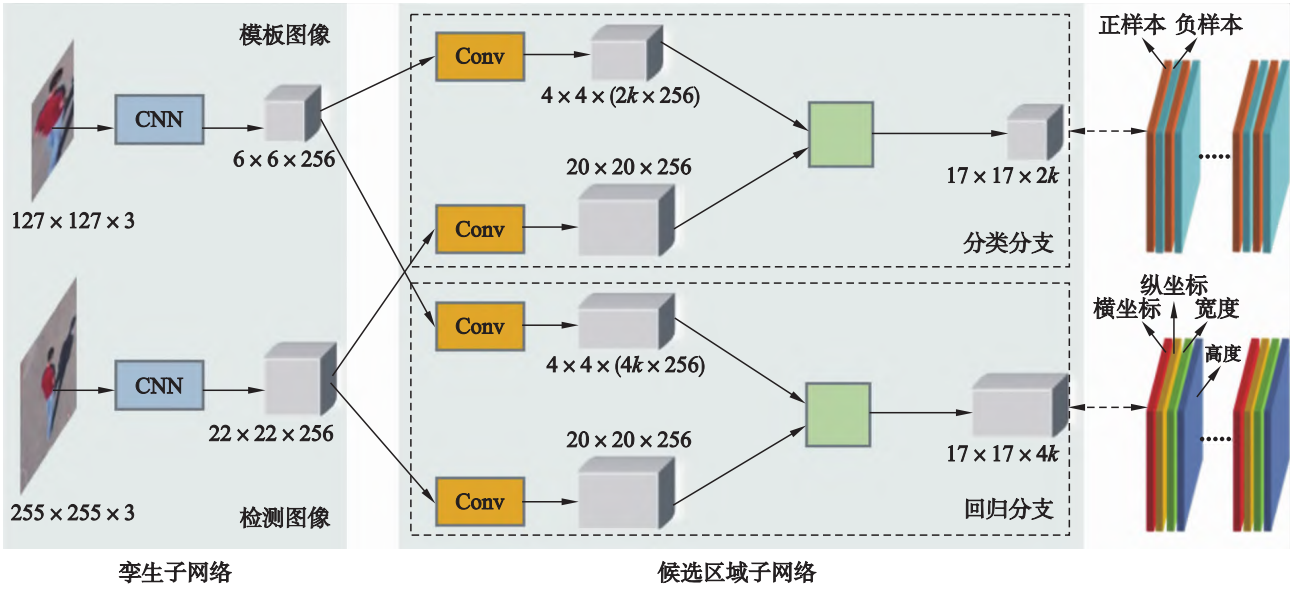

图 3　SiamRPN 跟踪算法框架

Fig.3　SiamRPN tracking algorithm framework

SiamRPN 网络包含孪生子网络部分(Siamese network)和候选区域子网络部分(region proposal network)。其中，孪生子网络部分与 SiamFC 网络结构类似，均采用不含 padding 的全卷积网络对样本图像和搜索图像进行特征提取。候选区域子网络包含两个分支，一个负责前景—背景分类，另一个用于边界框回归，通过对两种结果进行融合，能够同时得到跟踪目标当前的位置和尺寸。

尽管 SiamRPN 算法在速度和精度上均有了一定的提升，但由于训练阶段使用的数据集种类比较少，不足以训练出高质量的 SiamRPN 网络，该算法在无人机数据集 UAV20L 中，测试的 AUC(area under curve)结果仅为 0.454，可见，算法在性能上还有很大提升空间。

3.3　干扰物感知孪生网络(DaSiamRPN)跟踪算法

针对 SiamRPN 算法存在的不足，Zhu 等[34]提出了 DaSiamRPN 跟踪算法。该算法在 SiamRPN 基础上进行了3 个部分的改进：1) 在模型训练阶段引入了大规模数据集(ImageNet Detection 和 COCO Detection)，用于增加训练中正样本对的数量，从而提高跟踪器的泛化能力。2) 在训练过程中引入了足够多的同类别和不同类别的负样本对，来消除因样本不平衡而导致的跟踪不准确问题。该操作不仅能够避免跟踪器在目标遮挡等情况下发生漂移现象跟踪到其他物体上，而且能够使跟踪器专注于细粒度的目标跟踪。3) 除了引入平移、尺度变化等数据增强策略外，还首次引入了运动模糊数据增强来充分发挥孪生网络的潜力，同时提出了局部到全局的搜索策略，将跟踪扩展到长时间领域。

DaSiamRPN 跟踪算法在一定程度上提高了跟踪器的泛化能力和判别能力，使跟踪器不会轻易地被视频中其他物体所干扰，能够较好地处理跟踪目标部分或全部遮挡、目标出画面等情况，且能够达到 160 帧/s 的实时跟踪速度。

为了验证 DaSiamRPN 跟踪算法在无人机场景中的应用，文[34]分别在短时和长时无人机数据集

UAV123 与 UAV20L 上进行了实验，评估了该算法在两个数据集上的精度图和成功率图，并与当前最先进的跟踪算法进行了比较。结果表明，DaSiamRPN 算法能够有效地对无人机视频中的目标进行跟踪，并且能够较好地处理全遮挡和背景杂乱等挑战性的情况。

3.4 更深更宽孪生网络(SiamDW)跟踪算法

由于当前的 Siamese 系列网络都只采用较简单的网络(例如 AlexNet 网络)作为主干网，没有充分发挥深度神经网络的巨大潜力。为了更有效地利用更深更宽的网络对目标进行跟踪，Zhang 等[35]提出了更深更宽孪生网络框架(Deeper and Wider Siamese networks，SiamDW)。具体地，文[35]详细分析了使用深度网络(如 ResNet、Inception 等网络)代替浅层网络时性能下降的原因，并基于这些深度网络提出了新的残差模块(Cropping-Inside Residual unit，CIR)，再用这些模块设计新架构来解决深层网络中填充等因素的影响。

文[35]在设计 CIR 模块时，提出了 4 项基本准则来消除网络框架内部因结构因素带来的影响：1)由于网络步长会影响两个相邻输出特征的感受野重叠率，建议使用一个相对较小的网络步长，如 4 或 8；2)输出特征的感受野建议设置为样本图像的 60%～80%；3)设计网络时需要综合考虑网络步长、感受野和输出特征大小等因素的影响；4)对于孪生网络结构，去除掉主干网的填充操作是很有必要的。

文[35]利用所设计的 CIR 模块思想构建了 6 种不同的主干网络，并将其中的 CIResNet-22 网络应用于 SiamFC 和 SiamRPN 网络架构，通过充分的实验验证了所提结构的有效性，与现有先进算法相比，获得了最好的性能，且能达到实时跟踪效果。

3.5 改进全卷积孪生网络(SiamFC ++)跟踪算法

当前的跟踪算法虽然使用各种技巧提高了目标跟踪的精度和速度，但很少考虑到跟踪问题本身的特性。为了更深入地研究跟踪问题本身，Xu 等[81]提出了改进版全卷积孪生网络(Fully Convolutional Siamese tracker ++，SiamFC ++)算法，该算法在 SiamFC 框架基础上，结合当前跟踪算法的不足，做了 4 个方面的改进：1)同时引入了分类和目标状态估计分支，使跟踪器能够准确地从背景和干扰物中区分出目标；2)引入无歧义的分类得分，来代表目标存在区域的置信度，避免了 SiamRPN 系列方法中容易出现的假阳性结果，减少了跟踪失败的情况；3)不再考虑先验知识的学习，认为当前算法普遍依赖于数据分布的先验知识，影响了跟踪器的泛化能力；4)引入了评估质量分数分支，该分支独立于分类分支，共同对边界框进行选择，进一步提高了跟踪准确度。

此外，文[81]详细地分析了当前孪生网络跟踪器存在的不合理问题，为了验证所提准则的有效性，在 5 个具有挑战性的数据集上进行了测试，SiamFC ++算法均表现出更加出色的跟踪效果，在无人机数据集上具有更大的跟踪潜力。

3.6 算法总结

近年来，基于深度学习的目标跟踪算法在网络架构、骨干网络、特征融合、算法更新方式、目标搜索策略等方面均进行了大量研究，以使跟踪器学习丰富的表示形式并能有效地从目标中提取复杂和抽象的特征。相比于相关滤波跟踪算法，该算法能够获得更高的跟踪准确度，且研究者们期望能够提供通用的深度学习跟踪算法，同时在精度、速度和鲁棒性上获得更好的效果。

虽然深度学习跟踪算法取得了突破性进展，但大多数深度学习跟踪算法均采用离线训练的方式，无法快速适应完全遮挡甚至出视野的情况，故探索高效的神经网络在线训练方式是至关重要的。同时，训练样本缺乏问题导致难以训练出高质量的深度网络且易造成过拟合现象，影响跟踪效果。另外，现有深度目标跟踪算法对硬件配置是有要求的，需要有 GPU 的支持，这在一定程度上也影响了该算法在无人机上的普遍使用。

4 目标跟踪数据集和评价指标

4.1 目标跟踪数据集

构建目标跟踪数据集和建立相应的评价指标是目标跟踪任务中的重要部分。其中，目标跟踪数据集为模型训练和算法验证提供了强有力的数据支撑。

目前，通用的跟踪领域数据集主要有：Tracking-Net 数据集[82]、OTB 数据集[83-84]、VOT 竞赛数据库[85]、NUS-PRO 数据集[86]和 OxUvA 数据集[87]等。常用的无人机拍摄的空中场景类数据集有 VIVID 数据集[88]、VisDrone-2019 数据集[89]、UCLA Aerial Event 数据集[90]、UAV123 数据集[91]、UAVDT 数据集[92]和 DTB70 数据集[93]等，该类数据集能够用于无人机目标跟踪任务，增强无人机对地面目标物体的搜索和跟踪能力，表 4 为无人机目标跟踪数据集对比表。

表 4　无人机目标跟踪数据集对比

Tab.4　Comparison of UAV target tracking datasets

数据集	提出年份	数据集描述	视频数量	视频总帧数	属性数量
VIVID [88]	2005	相对时间较长，均从高空拍摄，被跟踪目标为地面上的运动车辆，遮挡情况较频繁	9	–	–
UAV123 [91]	2016	背景干净，视角变化较多，为高空拍摄的短时彩色视频	123	113 k	12
UAV20L [91]	2016	UAV123 数据集的父集，场景复杂，为长时彩色视频	20	59 k	12
DTB70 [93]	2017	平均时长 7 s，高空环境复杂多样，为短时彩色视频	70	15 k	11
BUAA-PRO [94]	2018	像素级标注的大规模数据集，利用掩模形式分割目标	150	8.7 k	12
UAVDT [92]	2018	环境复杂，目标小且密度高，摄像机运动剧烈的短时视频	50	80 k	9
VisDrone2018—2020 [95]	2018—2020	源于“Vision Meets Drone：A Challenge”挑战赛，为精细标注的短时彩色序列	132	106.4 k	12
VisDrone2019—2020L [89]	2019—2020	源于“Vision Meets Drone：A Challenge”挑战赛，无人机长时跟踪代表数据集之一	25	82.6 k	12
Small-90 [96]	2020	该数据集航拍视频大多来自于其他数据集，为短时彩色视频	90	–	11
Small-112 [96]	2020	在 Small-90 数据集的基础上，增加了 22 个具有挑战性的序列	112	–	11

表 4 中，所有的数据集均为无人机在高空对目标进行拍摄的视频序列，其中，数据集 VIVID、UAV20L 和 VisDrone2019—2020L 为长时跟踪数据集，其余数据集为短时跟踪数据集。除了 VIVID 数据集外，其他数据集均将视频序列分为多种属性，如光照变化、尺度变化、快速运动、视角变化和遮挡等。UAV123 数据集共包含 123 段视频序列和超过 11 万帧的图像，是无人机目标跟踪领域的代表性数据集之一。BUAA-PRO 是基于像素级标注的数据集，用于解决跟踪框中不可避免的非目标元素问题。VisDrone 数据集来源于“Vision Meets Drone：A Challenge”挑战赛，分为短时和长时视频两种，视频序列均为精细标注的彩色序列。Small-90 和 Small-112 为针对小目标进行跟踪的数据集，且所包含的视频序列大部分来自其他数据集的航拍视频.

4.2　评价指标

为了说明跟踪算法效果的好坏，需要对结果进行定性或定量评价，并对算法鲁棒性进行分析。目标视觉跟踪性能的优良主要通过性能度量值和性能对比图两种类型进行评价。性能度量值能够直观地根据准确性、鲁棒性的定量值来说明性能好坏，主要包括距离精度(distance precision，DP)，重叠精度(overlap precision，OP)，时序鲁棒性评估(temporal robustness evaluation，TRE)和空间鲁棒性评估(spatial robustness evaluation，SRE)。性能对比图能够根据不同的阈值进行分析，提供直观的不同算法之间的性能对比图，主要包括精度图(precision plot)和成功图(success plot)。

距离精度　通过计算中心定位误差(center location error，CLE)来实现。具体地，首先计算序列所有帧中算法估计的目标中心位置与人工标记的真实中心位置之间的欧氏距离，然后计算该距离小于给定阈值的视频帧占视频总帧的百分比。

$$\mathrm{DP}=\frac{N_{\mathrm{CLE}\leqslant \mathrm{th}_0}}{N}\times 100\% \tag{33}$$

其中，DP 为距离精度值；N 表示视频的总帧数；$N_{\mathrm{CLE}\leqslant \mathrm{th}_0}$ 表示中心定位误差小于给定阈值的视频帧总帧数，设 $\mathrm{CLE}=\sqrt{(x-x_0)^2+(y-y_0)^2}$ 为中心定位误差，(x,y) 为估计中心位置的坐标，(x_0,y_0) 为真实中心位置的坐标；th_0 为给定阈值。

重叠精度　通过计算重叠得分(overlap score，OS)来实现。具体地，首先计算序列所有帧中算法估计的目标区域与人工标记的目标区域重叠得分，然后计算该得分大于给定阈值的视频帧占视频所有帧的百分比。

$$\mathrm{OP}=\frac{N_{\mathrm{OS}\geqslant \mathrm{th}_1}}{N}\times 100\% \tag{34}$$

其中，OP 为重叠精度值；N 表示视频的总帧数；$N_{\mathrm{OS}\geqslant \mathrm{th}_1}$ 表示重叠得分大于给定阈值的视频帧数，$\mathrm{OS}=\frac{(r_{\mathrm{t}}\cap r_{\mathrm{g}})}{(r_{\mathrm{t}}\cup r_{\mathrm{g}})}$，$r_{\mathrm{t}}$ 为算法估计的目标区域，r_{g} 为人工标注的目标区域，th_1 为给定阈值。

时序鲁棒性评估　指通过选择不同的帧作为起始帧进行初始化，来对算法性能进行评估。具体地，首先选不同的帧作为起始帧进行跟踪，并利用起始帧对应的人工标注帧进行初始化，验证算法的

跟踪效果，再对所有结果取平均值，最终得到时序鲁棒性评估结果。

空间鲁棒性评估　指在起始帧中，标定不同的目标边界框进行初始化，对算法性能进行评估。具体地，首先将初始帧中的目标区域进行轻微平移或尺度缩放，再根据变动后的初始帧进行初始化，验证算法跟踪效果，最后取所有结果的平均值作为空间鲁棒性评估结果。

精度图　精度图(precision plot)显示了中心定位误差在不同给定阈值下，满足条件的视频帧占视频总帧的百分比，由于不同阈值所得百分比不同，最终得到精度曲线图。该方法通过中心定位误差来反映跟踪目标的位置准确性，但无法体现目标尺度和大小的变化。

成功图　成功图(success plot)显示了重叠得分在不同给定阈值(0 到 1 之间)下，满足条件的视频帧占视频总帧的百分比，由于不同阈值所得百分比不同，最终得到成功曲线图。该方法根据预测目标框与真实目标框之间的重叠比例来反映目标尺度和大小的变化，评估跟踪准确性。

5　未来研究展望

近年来，虽然无人机视觉目标跟踪算法有了突破性进展，鲁棒性越来越高，但仍受到真实空中场景中快速尺度变化、完全遮挡、低分辨率等各种因素的影响，导致跟踪漂移或丢失，针对无人机目标跟踪算法面临的问题，提出以下发展趋势:

1) 现有无人机目标跟踪算法更多解决的是视野之内的跟踪问题，但在无人机实际飞行环境中，目标被部分或完全遮挡甚至出视野的情况时常发生，目标跟踪器需要快速适应看不见的目标，因此，高效的在线训练和失跟后的重新检测是至关重要的。由于目标样本有限，如何有效引入小样本或单样本学习方法以及重新检测机制，针对目标被完全遮挡甚至出视野的情况构建更为鲁棒、稳定的跟踪器仍需要更深入的研究。

2) 针对无人机空中跟踪时，目标尺度小、干扰物数量多、特征提取困难的问题，深度网络算法能够通过预训练对小目标特征进行提取，可考虑引入增量学习、迁移学习和对抗学习等技巧来提高特征提取有效性，从而提高目标可辨识性和可区分性。同时，由于深度网络算法的结构复杂且会占用大量内存资源，会影响无人机实时跟踪的效果，可考虑引入知识蒸馏、网络剪枝等技术，构建结构精简、复杂度低的目标跟踪网络模型。实现无人机高精度、高实时性的目标跟踪是未来发展方向之一。

3) 目前无人机视觉目标跟踪算法多解决的是单一挑战下的短时跟踪(例如，只对光照变化、尺度变化或运动模糊中一种效果明显)，很少有算法能够同时满足多挑战下的长时目标跟踪。而无人机目标跟踪中，视角变化、光照变化、外观及尺度变化等现象时常发生，且易受到多目标干扰，因此，开发对多种影响因素鲁棒的长时通用目标跟踪算法是亟需研究者们关注的问题。

4) 现有的研究多关注的是单架无人机目标跟踪算法，而由于单架无人机在检测范围、观察视角等方面的局限性，其目标跟踪在场景理解方面往往存在不足，从而无法在复杂动态环境中对运动目标进行持续准确的跟踪。多无人机协同目标跟踪能够弥补单无人机局部信息提取不显著、全局信息识别不完整等缺陷，从而提高目标跟踪准确性和高效性。因此，研究如何有效融合多无人机目标观测数据，开发多无人机协同目标跟踪算法，值得研究者们重点关注。

6　结论

无人机目标跟踪任务是民用和军事领域中执行搜索、侦察、救援等任务的首要问题，有着重要的研究意义。本文首先综述了无人机视觉目标跟踪领域的经典跟踪算法，尤其是基于相关滤波的跟踪算法和基于深度学习的跟踪算法，详细介绍了两类算法的模型框架，阐述了各算法的特点和贡献，并对比了各算法的跟踪效果。其次，归纳了无人机目标跟踪相关的跟踪数据集和评价指标。最后，展望了无人机目标跟踪任务未来可能的发展趋势，为相关研究人员提供参考。

参考文献

[1] Zhao J H, Xiao G, Zhang X, et al. A survey on object tracking in aerial surveillance [C] //The International Conference on Aerospace System Science and Engineering. Berlin, Germany: Springer, 2018: 53 - 68.

[2] Ciaparrone G, Sánchez F L, Tabik S, et al. Deep learning in video multi-object tracking: A survey [J]. Neurocomputing, 2019, 381: 61 - 88.

[3] Lee K, Hwang J, Okopal G, et al. Ground-moving-platform based human tracking using visual SLAM and constrained multiple

kernels [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(12): 3602 – 3612.

[4] Uzkent B, Rangnekar A, Hoffman M J. Tracking in aerial hyperspectral videos using deep kernelized correlation filters [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(1): 449 – 461.

[5] Yuan C, Liu Z X, Zhang Y. 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(2/3/4): 635 – 654.

[6] Vanegas F, Campbell D, Roy N, et al. UAV tracking and following a ground target under motion and localisation uncertainty [C] //IEEE Aerospace Conference. Piscataway, USA: IEEE, 2017. DOI: 10.1109/AERO.2017.7943775.

[7] Xiong J, Niu Y F. Guidance law for multi-UAVs collaborative ground target tracking under obstacle environment [C] //29th Chinese Control and Decision Conference. Piscataway, USA: IEEE, 2017: 7219 – 7223.

[8] Hao J, Zhou Y, Zhang G, et al. A review of target tracking algorithm based on UAV [C] //IEEE International Conference on Cyborg and Bionic Systems. Piscataway, USA: IEEE, 2019: 328 – 333.

[9] Liu S, Liu D, Srivastava G, et al. Overview and methods of correlation filter algorithms in object tracking [J/OL] //Complex & Intelligent Systems. [2021 – 06 – 06]. https://doi.org/10.1007/s40747 – 020 – 00161 – 4.

[10] Zhao Z Q, Feng P, Guo J J, et al. A hybrid tracking framework based on kernel correlation filtering and particle filtering [J]. Neurocomputing, 2018, 297: 40 – 49.

[11] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2010: 2544 – 2550.

[12] Mei X, Ling H. Robust visual tracking and vehicle classification via sparse representation [J] //IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2259 – 2272.

[13] Wang N, Wang J, Yeung D. Online robust non-negative dictionary learning for visual tracking [C] //IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2013: 657 – 664.

[14] Held D, Thrun S, Sebastian S. Learning to track at 100 FPS with deep regression networks [C] //European Conference on Computer Vision, Berlin, Germany: Springer, 2016: 749 – 765.

[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] 葛宝义，左宪章，胡永江. 视觉目标跟踪方法研究综述 [J]. 中国图象图形学报, 2018, 23(8): 1091 – 1107.
Ge B Y, Zuo X Z, Hu Y J. Review of visual object tracking technology [J]. Journal of Image and Graphics, 2018, 23(8): 1091 – 1107.

[17] Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration [C] //European Conference on Computer Vision, Berlin, Germany: Springer, 2014: 254 – 265.

[18] 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, Germany: Springer, 2012: 702 – 715.

[19] Danelljan M, Häger G, Khan F S, et al. Accurate scale estimation for robust visual tracking [C/OL] //British Machine Vision Conference. London, UK: BMVA. [2021 – 06 – 06]. http://www.bmva.org/bmvc/2014/files/paper038.pdf.

[20] 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.

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

[22] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking [C] //IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2017: 1135 – 1143.

[23] Li F, Tian C, Zuo W M, et al. Learning spatial-temporal regularized correlation filters for visual tracking [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 4904 – 4913.

[24] Huang Z, Fu C H, Li Y, et al. Learning aberrance repressed correlation filters for real-time UAV tracking [C] //IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2019: 2891 – 2900.

[25] Lin F, Fu C H, He Y, et al. BiCF: Learning bidirectional incongruity-aware correlation filter for efficient UAV object tracking [C] //IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2020: 2365 – 2371.

[26] 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] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020: 11920 – 11929.

[27] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking [C] //IEEE Conference on Computer

Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016: 4293 – 4302.

[28] Yun S, Choi J, Yoo Y, et al. Action-decision networks for visual tracking with deep reinforcement learning [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 2711 – 2720.

[29] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking [C] //European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 850 – 865.

[30] Valmadre J, Bertinetto L, Henriques J, et al. End-to-end representation learning for correlation filter based tracking [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 2805 – 2813.

[31] Song Y, Ma C, Wu X H, et al. VITAL: Visual tracking via adversarial learning [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 8990 – 8999.

[32] Guo Q, Feng W, Zhou C, et al. Learning dynamic Siamese network for visual object tracking [C] //IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2017: 1763 – 1771.

[33] Li B, Yan J, Wu W, et al. High performance visual tracking with Siamese region proposal network [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 8971 – 8980.

[34] Zhu Z, Wang Q, Li B, et al. Distractor-aware Siamese networks for visual object tracking [C] //European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 103 – 119.

[35] Zhang Z P, Peng H W. Deeper and wider Siamese networks for real-time visual tracking [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2019: 4586 – 4595.

[36] Kalman R E. A new approach to linear filtering and prediction problems [J]. Journal of Basic Engineering, 1960, 82: 35 – 45.

[37] Pallavi S, Laxmi K R, Ramya N, et al. Study and analysis of modified mean shift method and Kalman filter for moving object detection and tracking [C] //The Third International Conference on Computational Intelligence and Informatics, Berlin, Germany: Springer, 2020: 821 – 828.

[38] Azrad S, Kendoul F, Nonami K. Visual servoing of quadrotor micro-air vehicle using color-based tracking algorithm [J]. Journal of System Design and Dynamics, 2010, 4(2): 255 – 268.

[39] Chen L, Jiang Y, Wang C. Adaptive visual servo control of UAV ground-target-autonomous-tracking system [C] //10th World Congress on Intelligent Control and Automation. Piscataway, USA: IEEE, 2012: 133 – 137.

[40] Hu X X, Chen Y, Luo H. Robust decision making for UAV air-to-ground attack under severe uncertainty [J]. Journal of Central South University, 2015, 22(11): 4263 – 4273.

[41] Farahi F, Yazdi H S. Probabilistic Kalman filter for moving object tracking [J]. Signal Processing: Image Communication, 2020. DOI: 10.1016/j.image.2019.115751.

[42] Metropolis N, Ulam S. The Monte Carlo method [J]. Journal of the American Statistical Association, 1949, 44(247): 335 – 341.

[43] Isard M, Blake A. Condensation-conditional density propagation for visual tracking [J]. International Journal of Computer Vision, 1998, 29(1): 5 – 28.

[44] Nummiaro K, Koller-Meier E, Gool L V. An adaptive color-based particle filter [J]. Image and Vision Computing, 2003, 21(1): 99 – 110.

[45] Maggio E, Smerladi F, Cavallaro A. Adaptive multi-feature tracking in a particle filtering framework [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(10): 1348 – 1359.

[46] Kamei K, Narumi K, Shuto D. Adaptive color tracking for particle filter to control a drone in a forest [C] //International Conference on Soft Computing and Intelligent Systems. Piscataway, USA: IEEE, 2016: 512 – 517.

[47] 王耀南，罗琼华，陈彦杰，等. 旋翼飞行机器人多机视觉跟踪定位系统及方法 [J]. 仪器仪表学报，2018，39(2)：1 – 10.
Wang Y N, Luo Q H, Chen Y J, et al. Visual tracking and positioning system and method of multiple rotorcrafts [J]. Chinese Journal of Scientific Instrument, 2018, 39(2): 1 – 10.

[48] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition [J]. IEEE Transactions on Information Theory, 1975, 21(1): 32 – 40.

[49] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2000: 142 – 149.

[50] Birchfield S T, Rangarajan S. Spatiograms versus histograms for region-based tracking [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2005: 20 – 26.

[51] Hwang J P, Baek J, Choi B, et al. A novel part-based approach to mean-shift algorithm for visual tracking [J]. International

Journal of Control，Automation，and Systems，2015，13(2)：443－453.

[52] 王旭东，王屹炜，闫贺. 背景抑制直方图模型的连续自适应均值漂移跟踪算法[J]. 电子与信息学报，2019，41(6)：1480－1487.
Wang X D，Wang Y W，Wang H. Continuously adaptive mean-shift tracking algorithm with suppressed background histogram model[J]. Journal of Electronics & Information Technology，2019，41(6)：1480－1487.

[53] Pareek A，Arora N. Re-projected SURF features based mean-shift algorithm for visual tracking[J]. Procedia Computer Science，2020，167：1553－1560.

[54] Dai K，Wang D，Lu H，et al. Visual tracking via adaptive spatially-regularized correlation filters[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，USA：IEEE，2019：4665－4674.

[55] Liu X，Sun M，Meng L. Object tracking algorithm of UAV based on fast kernel correlation filter[C]//Chinese Intelligent Systems Conference. Berlin，Germany：Springer，2020：584－593.

[56] Fu C H，Xu J，Lin F，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.

[57] Li F，Fu C H，Lin F，et al. Training-set distillation for real-time UAV object tracking[C]//IEEE International Conference on Robotics and Automation. Piscataway，USA：IEEE，2020：9715－9721.

[58] Fu C H，Zhang Y，Duan R，et al. Robust scalable part-based visual tracking for UAV with background-aware correlation filter[C]//IEEE International Conference on Robotics and Biomimetics. Piscataway，USA：IEEE，2018：2245－2252.

[59] Fu C H，He Y，Lin F，et al. Robust multi-kernelized correlators for UAV tracking with adaptive context analysis and dynamic weighted filters[J]. Neural Computing and Applications，2020，32：12591－12607.

[60] Danelljan M，Khan F S，Felsberg M，et al. Adaptive color attributes for real-time visual tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，USA：IEEE，2014：1090－1097.

[61] Li F，Yao Y J，Li P H，et al. Integrating boundary and center correlation filters for visual tracking with aspect ratio variation[C]//IEEE International Conference on Computer Vision Workshop. Piscataway，USA：IEEE，2017：2001－2009.

[62] Zhang T Z，Xu C S，Yang M H. Multi-task correlation particle filter for robust object tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，USA：IEEE，2017：4335－4343.

[63] 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. Berlin，Germany：Springer，2016：472－488.

[64] Danelljan M，Bhat G，Khan F S，et al. ECO：Efficient convolution operators for tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，USA：IEEE，2017：6931－6939.

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

[66] Krizhevsky A，Sutskever I，Hinton G E. ImageNet classification with deep convolutional neural networks[C]//25th International Conference on Neural Information Processing Systems. Cambridge，USA：MIT Press，2012：1097－1105.

[67] Szegedy C，Liu W，Jia Y，et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，USA：IEEE，2015：1－9.

[68] Simonyan K，Zisserman A. Very deep convolutional networks for large-scale image recognition[C/OL]//International Conference on Learning Representations.[2021－06－10]. https://arxiv.org/pdf/1409.1556.pdf.

[69] He K，Zhang X，Ren S，et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，USA：IEEE，2016：770－778.

[70] Chatfield K，Simonyan K，Vedaldi A，et al. Return of the devil in the details：Delving deep into convolutional nets[C/OL]//British Machine Vision Conference. London，UK：BMVA.[2021－06－06]. http://www.bmva.org/bmvc/2014/files/paper054.pdf.

[71] Chen L，Liu Y G，Man Y C. Spatial-temporal channel-wise attention network for action recognition[J/OL]. Multimedia Tools and Application.[2021－06－06]. https://doi.org/10.1007/s11042－021－10752－z.

[72] Song Y B，Ma C，Gong L J，et al. Crest：Convolutional residual learning for visual tracking[C]//IEEE International Conference on Computer Vision. Piscataway，USA：IEEE，2017：2555－2564.

[73] Ma C，Xu Y，Ni B，et al. When correlation filters meet convolutional neural networks for visual tracking[J]. IEEE Signal Processing Letters，2016，23(10)：1454－1458.

[74] Jung I，Son J，Baek M，et al. Real-time MDNet[C]//European Conference on Computer Vision. Berlin，Germany：Springer，

2018: 89 – 104.

[75] Li B, Gan Z, Chen D, et al. UAV maneuvering target tracking in uncertain environments based on deep reinforcement learning and meta-learning [J]. Remote Sensing, 2020, 12(22). DOI: 10.3390/rs12223789.

[76] He A, Luo C, Tian X, et al. Towards a better match in Siamese network based visual object tracker [C] //European Conference on Computer Vision Workshop, Berlin, Germany: Springer, 2019: 132 – 147.

[77] Yu Y, Xiong Y, Huang W, et al. Deformable Siamese attention networks for visual object tracking [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020: 6727 – 6736.

[78] Guo D, Wang J, Cui Y, et al. SiamCAR: Siamese fully convolutional classification and regression for visual tracking [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020: 6268 – 6276.

[79] Chen Z, Zhong B, Li G, et al. Siamese box adaptive network for visual tracking [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020: 6667 – 6676.

[80] Li B, Wu W, Wang Q, et al. SiamRPN ++: Evolution of Siamese visual tracking with very deep networks [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2019: 4282 – 4291.

[81] Xu Y, Wang Z, Li Z, et al. SiamFC ++: Towards robust and accurate visual tracking with target estimation guidelines [C] //AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020: 12549 – 12556.

[82] Müller M, Bibi A, Giancola S, et al. TrackingNet: A large-scale dataset and benchmark for object tracking in the wild [C] //European Conference on Computer Vision, Berlin, Germany: Springer, 2018: 310 – 327.

[83] Wu Y, Lim J, Yang M. Online object tracking: A benchmark [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2013: 2411 – 2418.

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

[85] Kristan M, Leonardis A, Matas J, et al. The eighth visual object tracking VOT2020 challenge results [C/OL] //European Conference on Computer Vision Workshops. Berlin, Germany: Springer. [2021 – 06 – 06]. http: //prints.vicos.si/publications/384.

[86] Li A, Lin M, Wu Y, et al. NUS-PRO: A new visual tracking challenge [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 335 – 349.

[87] Valmadre J, Bertinetto L, Henriques J, et al. Long-term tracking in the wild: A benchmark [C] //European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 670 – 685.

[88] Collins R, Zhou X, Teh S K. An open source tracking testbed and evaluation web site [C] //IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Piscataway, USA: IEEE, 2005: 17 – 24.

[89] Du D, Zhu P, Wen L, et al. VisDrone-SOT2019: The vision meets drone single object tracking challenge results [C] //IEEE International Conference on Computer Vision Workshop. Piscataway, USA: IEEE, 2019: 199 – 212.

[90] Shu T, Xie D, Rothrock B, et al. Joint inference of groups, events and human roles in aerial videos [C] //IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2015: 4576 – 4584.

[91] Mueller M, Smith N, Ghanem B. A benchmark and simulator for UAV tracking [C] //European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 445 – 461.

[92] Yu H, Li G, Zhang W, et al. The unmanned aerial vehicle benchmark: Object detection, tracking and baseline [J]. International Journal of Computer Vision, 2020, 128: 1141 – 1159.

[93] Li S Y, Yeung D. Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models [C] //31st AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 4140 – 4146.

[94] Li A, Chen Z, Wang Y. BUAA-PRO: A tracking dataset with pixel-level annotation [C/OL] //British Machine Vision Conference. London, UK: BMVA. [2021 – 06 – 06]. http: //bmvc2018.org/contents/papers/0851.pdf.

[95] Zhu P, Wen L, Du D, et al. VisDrone-VDT2018: The vision meets drone video detection and tracking challenge results [C] //European Conference on Computer Vision Workshop. Berlin, Germany: Springer, 2018: 496 – 518.

[96] Liu C, Ding W, Yang J, et al. Aggregation signature for small object tracking [J]. IEEE Transactions on Image Processing, 2020, 29: 1738 – 1747.

作者简介

陈　琳(1990 –), **女, 博士生。研究领域为视觉目标跟踪与行为识别。**

刘允刚(1970 –), **男, 教授, 博士生导师。研究领域为非线性控制, 自适应控制, 协同控制理论及其应用, 智能系统调控。**