2021 年 第21卷 第35期
2021, 21(35): 14871-11

科 学 技 术 与 工 程
Science Technology and Engineering

ISSN 1671—1815
CN 11—4688/T

引用格式:彭建盛, 许恒铭, 李涛涛, 等. 生成式与判别式视觉目标跟踪算法综述[J]. 科学技术与工程, 2021, 21(35): 14871-14881.
Peng Jiansheng, Xu Hengming, Li Taotao, et al. Survey of visual target tracking algorithms based on generative and discriminant[J]. Science Technology and Engineering, 2021, 21(35): 14871-14881.

自动化技术、计算技术

生成式与判别式视觉目标跟踪算法综述

彭建盛[1,2], 许恒铭[1], 李涛涛[1], 侯雅茹[1]
(1. 广西科技大学电气与信息工程学院, 柳州 545000; 2. 河池学院人工智能与制造学院, 宜州 546300)

摘 要 视觉目标跟踪是对目标位置、速度、运动轨迹等信息检测与预测的技术。该技术融合了计算机视觉、图像处理、深度学习等众多技术。将对目标跟踪算法发展情况以及研究现状进行梳理,首先介绍目前常用的基准数据集;其次指出生成式算法与判别式算法差异;再对传统的生成式算法进行简单分析总结;随后围绕算法框架分别介绍相关滤波框架、深度学习框架、孪生网络框架、Transformer 框架的判别式算法并分析不同算法的优缺点;最后分析目前动态目标跟踪存在的问题并展望。

关键词 目标跟踪; 生成式算法; 相关滤波; 深度学习; 孪生网络
中图法分类号 TP391; **文献标志码** A

Survey of Visual Target Tracking Algorithms Based on Generative and Discriminant

PENG Jian-sheng[1,2], XU Heng-ming[1], LI Tao-tao[1], HOU Ya-ru[1]
(1. College of Electrical and Information Engineering, Guangxi University of Science and Technology, Liuzhou 545000, China;
2. Colleges of Artificial Intelligence and Smart Manufacturing, Hechi University, Yizhou 546300, China)

[**Abstract**] The position, velocity, motion trajectory and other information of the target are detected and predicted by visual target tracking. Computer vision, image processing, deep learning and many other technologies are integrated in visual target tracking. The development and research status of target tracking algorithm were introduced. Firstly, the commonly used benchmark data sets were introduced. Secondly, the difference between generative algorithm and discriminant algorithm was pointed out. In addition, the traditional generative algorithm was simply analyzed and summarized. Then, the discriminant algorithms of correlation filtering framework, deep learning framework, twin network framework and transformer framework were introduced, and the advantages and disadvantages of different algorithms were analyzed; Finally, the existing problems of dynamic target tracking were analyzed and prospected.

[**Keywords**] target tracking; generative algorithm; correlation filtering; deep learning; siamese network

20 世纪 90 年代,计算机视觉技术、图像处理技术得到了快速发展,视觉目标跟踪技术也得到广泛关注。目前该技术已经广泛应用在交通、军事、医疗以及民用设备等众多领域。实现目标跟踪具有众多难点,主要有动态目标运动过程中尺度外观发生变化、光照强度发生变化、目标遮挡或移出视野、复杂背景以及目标相似度较高等[1]。

近年来,三大顶尖会议国际计算机视觉大会(International Conference on Computer Vision, ICCV)、欧洲计算机视觉国际会议(European Conference on Computer Vision, ECCV)和国际计算机视觉与模式识别会议(Conference on Computer Vision and Pattern Recognition, CVPR)公开的目标跟踪论文数量一直居高不下,关于目标跟踪的综述众多。文献[2]从相关滤波器作为切入点围绕着核相关滤波算法(kernel correlation filter, KCF)算法进行展开,介绍了相关滤波器框架下的目标跟踪原理以及发展脉络。文献[3]从人工神经网络结构、功能、训练方式等角度重点介绍基于深度学习的目标跟踪算法,同时指出深度学习技术在目标跟踪问题上存在的

收稿日期: 2021-03-09; **修订日期**: 2021-08-26
基金项目: 国家自然科学基金(62063006);广西自然科学基金(2018GXNSFAA281164)
第一作者: 彭建盛(1979—),男,汉族,湖南茶陵人,博士,教授。研究方向:智能控制与智能自动化、嵌入式开发与应用。E-mail:sheng120410@163.com。

不足。文献[4]从目标特征提取及融合方式、外观模型创建方式、目标搜索方式等方面对目前主流的目标跟踪算法分类介绍,并且对深度学习环境下目标跟踪算法发展方向进行分析展望。文献[5]从孪生网络在目标跟踪中的应用入手,分析了基于孪生网络的目标跟踪算法比传统算法的优势以及未来该类算法的发展方向。文献[6]从尺度方向自适应角度进行总结,分析各算法跟踪过程中如何处理尺度变化问题。此外过去的目标跟踪算法综述文章均没有对最新的 Transformer 框架目标跟踪算法进行总结。

现有算法的准确性、跟踪速度等方面仍然有提升空间,视觉目标跟踪问题仍然具有挑战性。目标跟踪算法主要可以分为生成式和判别式两大类。现将从生成式动态目标跟踪算法与判别式动态目标跟踪算法分类标准作为切入点。首先主要介绍目前目标跟踪领域常用的数据标准集。其次介绍从原理的角度分析两类算法的区别。再次主要介绍两大类生成式算法,均值漂移算法和贝叶斯滤波算法。然后从主流判别式算法框架相关滤波框架、深度学习框架、孪生网络框架、Transformer 框架重点介绍判别式算法。最后结合前面所介绍的算法优缺点展望发展趋势。

1 基准数据集

基准数据集的出现,使得不同算法性能对比有了依据。目前主要使用的基准数据集有在线目标跟踪基准(object tracking benchmark, OTB)、视觉目标跟踪集(visual object tracking, VOT)等。OTB 基准数据集包含 OTB50[7]、OTB100[8]。OTB50 基准数据集由 50 段带有标记属性的视频序列组成,其中代表了现实情况中执行跟踪算法时会遇到的难题,包括光照变化、尺度变化、遮挡、形变、运动模糊、快速运动、平面内旋转、平面外旋转、移出视野范围、低分辨率以及背景杂乱等。文献[8]在 OTB50 的基础上对数据集进行了扩展,将 50 段视频序列增加到 100 段。

VOT 数据集是为每年 VOT 竞赛而准备的数据集,最早可以追溯到 VOT2013[9]。该数据集与 OTB 基准数据集相比采用全彩色图片并且图片清晰度也更高。随着目标跟踪算法的不断优化,该数据集的数据以及评判标准逐年更新并且逐年加大难度。目前 VOT2013[9]和 VOT2014[10]数据集以及评判标准已经被官方宣布停止使用,在此也不做讨论。VOT2015[11]在 VOT2014 数据集基础上将数据集的序列数量扩大了 1 倍,并且对目标加入旋转特性检测,在评价标准上首次提出使用平均重叠期望(expected average overlap, EAO)标准,并将该标准作为算法排名的主要依据。VOT2016[12]选用与 VOT2015 一样的样本集,并使用自动标注的方式对样本集重新进行标注,此外从重叠估计方面优化了评价体系。VOT2017[13]基准数据集除了在原来数据集的基础上进行改进,同时添加了隔离数据集用于冠军角逐。VOT2017 首次提出对跟踪器进行实时性实验,这表明目前的目标跟踪算法正在从实验阶段向实际应用阶段发展。VOT2018[14]沿用往年的数据集,在原来赛事基础上增加长时跟踪挑战项目。长时目标跟踪挑战中目标可能会出现消失或被遮挡等现象,因此评价工具需要跟踪器实时反馈目标位置以及目标情况。长时目标跟踪进一步促进了目标跟踪算法走出实验室。VOT2019[15]不仅恢复了关于热成像的目标跟踪挑战项目,还加入了融合深度特征的目标跟踪挑战,至此挑战项目多达 5 项。此外,VOT2019 对公数据集和隔离数据集进行部分更新,并根据长时跟踪和短时跟踪在目标假设上的不同优化了评价协议。VOT2020[16]使用 Python 语言重新编写了 VOT 工具包,测试跟踪器的能力得到了大幅提高。在短时跟踪挑战中使用分割的方式标记目标代替传统使用矩形框的方式标记目标。如表 1 所示为 VOT 每年排名前五的算法汇总。

2 生成式算法与判别式算法对比

生成式目标跟踪算法通过提取待跟踪目标的特征建立目标模型,利用生成的模型对待检测图像进行搜索,寻找与目标模型最匹配区域,该区域即为

表 1 VOT 每年排名前五算法汇总[9-16]

Table 1 Summary of top five algorithms in VOT of every year[9-16]

排名	VOT2013	VOT2014	VOT2015	VOT2016	VOT2017	VOT2018	VOT2019	VOT2020
1	PLT	DSST	MDNet	C-COT	LSART	LADCF	DRNet	RPT
2	FoT	SAMF	DeepSRDCF	TCNN	CFWCR	MFT	Trackyou	OceanPlus
3	EDFT	KCF	EBT	SSAT	CFCF	SiamRPN	ATP	AlphaRef
4	LGT + +	DGT	SRDCF	MLDF	ECO	UPDT	DiMP	AFOD
5	LT-FLO	PLT 14	LDP	Staple	Gnet	RCO	Cola	LWTL

目标区域。因此生成式目标跟踪算法主要框架包含四个部分,目标选择、目标特征提取、目标建模、目标定位,如图 1 所示。具体来说,人工或使用目标检测算法对第一帧图像进行处理,勾选出目标并标记,当前常用的目标检测算法有帧差法、背景差法等;然后对选中的目标的特征进行建模,通常关注目标的灰度特征、边缘特征、梯度特征、颜色特征、纹理特征等,常用的模型有混合高斯模型、贝叶斯网络模型、马尔可夫模型等;目标定位则是完成跟踪。

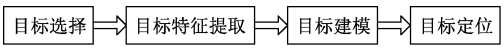

图 1　生成式目标跟踪框架

Fig. 1　Generative target tracking framework

判别式目标跟踪算法认为目标跟踪问题是关于目标和背景的分类问题。该类算法将图像中将目标区域作为正样本,背景区域作为负样本进行训练并生成分类器,生成的分类器可以在下一帧图像中找到最优区域,该区域为目标区域。目前判别式目标跟踪算法主要可以分为相关滤波类、深度学习类以及孪生神经网络类。

因此生成式算法与判别式算法最大的区别是,生成式方法仅关注目标特征信息,以目标特征建立目标模型,忽视背景信息;判别式方法采用机器学习训练分类器,训练过程中不仅包含目标信息,还兼顾背景信息。

3　生成式模型的目标跟踪算法

3.1　均值漂移算法

均值漂移(Meanshift)[17]算法是沿向量方向连续迭代候选目标帧,使其与模板的相似度最大,并收敛到目标的真实位置。该方法计算复杂度低,提取特征时会根据空间距离对中心位置周围的点进行加权。为解决光照变化导致的鬼影问题,王凯等[18]将色度、梯度以及运动矢量预测引入 Meanshift 算法中提高了监控系统场景下目标跟踪算法准确性。

针对目标运动过程中尺度变化问题对 Meanshift 算法进行改进提出 Camshif[19]。Meanshift 算法是针对单帧图像的计算,并且颜色概率分布是静态的,而 Camshift 算法对每个帧图像进行 Meanshift 运算,其颜色概率分布是动态的,Camshift 能够调节搜索窗口,解决了目标尺度变化问题,但降低了原算法的实时性。董恩增等[20]将特征匹配融入传统的 Camshift 算法,有效优化背景中与目标颜色相近导致的错误跟踪问题。

Vojir 等[21]提出尺度自适应均值漂移跟踪器(scale adaptive mean-shift tracker, ASMS)。该算法在原有的 Meanshift 算法框架下引入了尺度估计以及颜色直方图特征,同时使用尺度不巨变先验和可能偏最大先验作为正则项以及反向尺度一致性检测,在跟踪目标过程中会根据目标尺寸变化进行相应调整。ASMS 算法在保证一定准确度的同时平均帧率可以达到 125 帧/s,是 VOT2015 竞赛官方推荐的实时算法,但是当背景颜色与目标颜色相近时该算法很容易丢失目标。如表 2 所示为均值漂移算法改进方式与效果。

表 2　均值漂移算法改进与效果

Table 2　Improvement and effect of Meanshift algorithm

算法名称	改进方式	改进效果
Camshif[19]	每个帧图像进行 Meanshift、自动调节搜索窗口	解决了目标尺度变化问题
ASMS[21]	引入了尺度估计、经典颜色直方图特征、两个先验和一个检测	进一步优化目标尺度变化问题,保证速度的情况下提高了跟踪准确度。

3.2　贝叶斯滤波算法

递归贝叶斯滤波(Bayesian filtering) 算法是基于贝叶斯估计理论的基础提出的目标跟踪方案[22]。该算法包含预测和更新两个步骤,通过这两个步骤反复迭代估计图像中目标的位置。递归贝叶斯滤波概率在实际目标跟踪中很难获得最优解,为解决这一问题提出了卡尔曼滤波目标跟踪算法和粒子滤波目标跟踪算法。

卡尔曼滤波(Kalman filtering)算法[23]是一种以状态方程和观测方程为基础,运用递归的方式来预测目标在下一帧的位置。在跟踪过程中该算法可以通过卡尔曼增益变化规避传感器噪声、观测值不准确等问题,并实现对状态真实值的最优估计。卡尔曼滤波算法作为一种预测机制,当目标发生遮挡时可能会出现丢失目标现象。杨鹏生等[24]提出了一种基于改进扩展卡尔曼滤波的目标跟踪算法。该算法通过构建时间差和信号到达方向的观测方程,利用几何和代数关系化简得到伪线性模型提高了目标跟踪精度,而且使目标跟踪结果更加稳定。基于卡尔曼滤波的目标跟踪算法仅对线性目标模型有较好的效果,然而在实际应用中大多数模型都是非线性的。许红香等[25]将无损变换引入传统的卡尔曼滤波算法,并将采集到的图像分为多簇处理后使用协方差交叉融合,改善非线性目标跟踪。对于非线性目标模型,通常适用粒子滤波方法实现目标跟踪。

粒子滤波算法[26]将蒙特卡洛思想引入贝叶斯滤波中。该算法核心思想是将随机采样与重要性重采样相结合。通过对图像随机散布粒子并采样特征,将采样结果与目标特征对比,计算出每个粒子的相似度,对相似度高的区域投入更多的粒子,迭代操作最终确定目标位置。粒子滤波从一定程度上,属于卡尔曼滤波的拓展,解决了卡尔曼滤波只适用于线性高斯分布概率问题,为分析非线性模型提供了一种有效的解决方案。赵宗超等[27]利用引导图像滤波(guided image filer, GIF)对待检测图像滤波处理增强目标区域,使编码器增加训练样本,提高粒子置信度准确性,实现在线跟踪。如表3所示为贝叶斯滤波算法改进方式与效果。

表3 贝叶斯滤波算法改进与效果

Table 3 Improvement and effect of Bayesian filtering algorithm

算法名称	改进方式	改进效果
卡尔曼滤波[23]	采用最小均方误差的最优线性递归滤波方法	实现观测值不准确情况下对状态真实值的最优估计,只适用于高斯模型
粒子滤波[26]	蒙特卡洛思想引入贝叶斯滤波	适用于非线性模型,提高跟踪精度

4 判别式模型的目标跟踪算法

4.1 相关滤波目标跟踪算法

基于相关滤波的目标跟踪算法可以近似看成两个信号寻找最大相关值。通过对第一帧样本图片进行训练,输出一个具有区分背景和目标能力的滤波器,使用该滤波器对后面的每一帧图片进行运算获取相关值,根据运算后相关值的大小判断目标位置,相关值越大,说明该区域与目标的相似度越高,同时将每一回合响应结果返回滤波器对滤波器进行更新以提高下回合跟踪的准确性。如图2所示为相关滤波结构框图。

基于相关滤波器目标跟踪算法最早可以追溯到由Bolme等[28]提出的最小化输出误差平方和滤波器(minimum output tracking using adaptive correlation filters, MOSSE)。MOSSE能够在初始化单个帧的时候产生稳健的滤波器,并用卷积定理将输入图像和滤波器转换到频域相乘,再返回时域得到响应图。根据响应值大小来实现目标跟踪。由于MOSSE算法仅提取灰度特征,因此具有较快的运行速度可以达到669帧/s。后续的基于相关滤波的目标跟踪算法大多是建立在MOSSE算法框架的基础上围绕特征表达、尺度自适应、边界效应等方面进行改进。

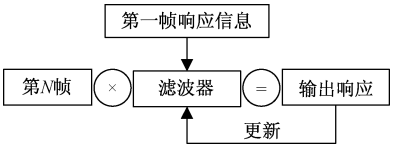

图2 相关滤波结构框图

Fig. 2 Structure block diagram of correlation filter

Henriques等[29]首次将核函数引入目标跟踪算法提出循环结构核跟踪器(circulant structure kernels, CSK),使用了一种基于循环矩阵的采样方法代替传统的粒子采样方法。该跟踪器与MOSSE相同仅采集单通道灰度特征,灰度特征的表征能力不足以处理背景复杂或者目标与背景颜色相似的情况。为丰富目标特征,Danelljan等[30]以CSK算法为基础,将原本的灰度特征换成多通道颜色特征提出自适应颜色属性(color name, CN)算法,同时结合自适应降维策略将主成分分析(principal components analysis, PCA)从11维降到2维,改进了模型跟踪方案在降低计算成本的同时提升了跟踪性能。除了颜色特征,方向梯度直方图(histogram of oriented gradients, HOG)特征在描述目标方面也显示出了明显的优势。Henriques等[31]在CSK基础上提出了核相关滤波算法(kernel correlation filter, KCF)和判别式相关滤波器(discriminative correlation filter, DCF),该算法将单通道的灰度特征拓展到了HOG特征。其采用岭回归与循环移位进行密集采样,并通过核函数将低维线性空间映射到高维空间,提高了相关滤波跟踪算法的鲁棒性。Li等[32]对DCF算法进行改进,提出了AutoTrack算法。该算法引入空间局部相应作为正则项,实现了在线自适应时空正则项优化,在保证跟踪精度的同时实现了60帧/s。AutoTrack算法定位用于无人机系统,是目前最好的基于中央处理器(central processing unit, CPU)跟踪器。

为优化尺度变化问题,Li等[33]提出了集成尺度自适应和核相关滤波器的跟踪器(a kernel correlation filter tracker with scale adaptive and feature integration, SAMF)算法。该算法在KCF框架基础上,将KCF单一特征换成对Gray、Color、HOG三种特征矢量叠加,同时创建拥有7个粗略尺度的尺度池通。该算法对预选区域目标按照尺度池做七次尺度计算并于上一帧图片目标进行对比,响应值最大的尺度则为当前目标,由此可知SAMF在实际使用中速度非常慢。Danelljan等[34]提出的精确尺度目标跟踪(accurate scale estimation for robust visual tracking, DSST)算法,该算法采用先平移跟踪后尺度跟踪策略实现局部寻优过程,同时将尺度池扩展到33个精

细尺度,有效解决了 SAMF 最高峰处不一定是尺度最优解问题,并在 VOT2014 竞赛上取得冠军成绩。为提高效率,该团队在 2017 年提出快速精确尺度目标跟踪(fast DSST, FDSST)算法[35],采用降维以及正交三角分解减低了计算量,有效的解决复杂度问题。以上几种算法均为穷举尺度池的方法,除此之外还有分块处理方法。宸梦楠等[36]将 DSST 算法中的手动标记换成检测算法,提高了指定场景下目标跟踪便捷度。Li 等[37]提出的可信块跟踪器(reliable patch trackers, RPT),由部件间的相对位置关系推断目标的尺度变换情况。Akin 等[38]从平衡准确性与实时性的角度出发提出了通过全局和局部耦合相关滤波对部分形变目标跟踪的跟踪器(deformable part-based tracking by coupled global and local correlation filters, DPCF),通过一个全局滤波器和多个部件滤波器间的耦合作用协同处理局部遮挡和尺度变化问题。除了尺度池穷举法和分块模型法可以实现尺度自适应外,还可以通过特征点模型实现尺度自适应。Montero 等[39]提出的 sKCF 算法在相邻图像帧中以目标为中心提取特征点并对按照距离对特征点赋予不同的权值,离目标中心距离越近权值越大,反之权值越小。通过对比前后图像帧中的特征点的位置变化实现对目标尺度自适应变化。相比之下分块模型法和特征点模型法作为自适应定义尺度方法是对尺度池穷举法的优化,具有更好的效果。

当目标发生快速运动或者形变时将会导致边界效应发生。单使用 HOG 特征 CSK、KCF 等算法将无法应对。为应对边界效应问题,较早是由 Danelljan 等[40]提出的空间正则化鉴别相关滤波器跟踪器(spatially regularized discriminative correlation filter tracker, SRDCF)。该算法在传统的 KCF 算法基础上扩大了搜索区域,并采用空域正则化对滤波器边界函数加大权重约束,并使用 Gauss-Seidel 函数来求解滤波器,使得分类器能够更准确地进行追踪。SRDCFdecon[41]在原算法的基础上对样本问题和学习率问题进行优化。Kiani 等[42]提出有限边界的相关滤波器(correlation filters with limited boundaries, CFLB)算法。该算法通过使用大于目标区域的图片进行训练,然后使用二值矩阵对目标区域进行提取。由于选择图片过大会导致训练出的滤波器不准算法,因此该算法在鲁棒性方面并不理想。

表 4 为部分相关滤波目标跟踪算法分析对比。通过对比发现每一种算法都有独特的优点,因此将多种算法合理的融合,实现优势互补也是较好的研究方向。随着硬件的高速发展,计算机计算能力不断提高实时性的问题可能会得到一定的解决。当前众多学者致力于深度特征与相关滤波算法相结合,进一步优化相关滤波目标跟踪算法的性能。

4.2　深度学习目标跟踪算法

从 VOT 目标跟踪大赛比赛结果可以看出,竞赛的前五名大多使用了深度学习算法。目前主流的深度学习算法在目标跟踪领域仍然存在两个明显的缺陷:①目标跟踪中正样本通常只有初始帧中的目标,没有大量的数据支撑很难训练出良好的分类器;②为改善目标跟踪效果,通常需要提高网络的复杂度,这导致了算法的复杂度提高,实时性大幅下降。当前基于深度学习的目标跟踪算法主要可以分为两大类,一类是将深度学习技术引入相关滤波目标跟踪算法,对原有的算法进行改进;另一类是直接使用深度学习技术对目标进行端到端的跟踪。

深度学习与相关滤波相结合的目标跟踪算法主要是使用深度学习技术对目标进行特征提取,并将提取的特征代替原相关滤波算法中的特征。Ma

表 4　部分相关滤波目标跟踪算法

Table 4　Target tracking algorithm based on partial correlation filtering

算法名称	采用特征	特点	缺点
MOSSE[28]	Gray	利用傅里叶变换使滤波过程在频域执行跟踪速度快	采用的图像特征均为单通道灰度特征,因此在很大程度上限制了这两种跟踪方法在面对复杂背景或目标与背景颜色相近时的跟踪能力,无法应对目标尺度变化
CSK[29]	Gray	利用核函数和傅里叶变换计算相邻两帧图像的相关性,在提高跟踪效果的同时保证了较快的跟踪速度	
KCF[31] DCF	HOG	提升了跟踪器在复杂环境下对运动模糊和光照变化的鲁棒性	对于目标本身尺度变化导致的跟踪漂移问题的处理能力仍然有待加强
SAMF[33]	Gray、Color、HOG	使用 7 个粗略尺度的尺度池,单个滤波器	全局寻优的过程,响应最高的全局最优不能保证是局部最优
DSST[34]	HOG、Intensity	分别训练平移滤波器和尺度滤波器,特征选择上保持高度的灵活性,其对运动模 糊、光照变化具有很强的鲁棒性	算法复杂度高,计算量大,实时性差,对于目标发生形 变时跟踪效果差
SRDCF[40]	HOG	用空域正则化对滤波器边界函数加大权重约束,并进行迭代优化,分类器能够更准确地进行追踪	算法复杂度高,计算量大,实时性差

等[43]提出分层卷积特征目标跟踪算法(hierarchical convolutional features for visual tracking, HCF),该算法用 VGGNet-19 模型作为特征提取器,经过三层网络分别训练滤波器,并通过滤波器得到的响应图进行加权,得到最终的响应图,并通过最大响应定位目标中心。但该算法在应为目标尺寸变化问题鲁棒性较差,主要是因为该算法仅关注目标的中心位置,并默认被跟踪物体尺寸不发生变化。Danelljan 等[44]在 SRDCF 的基础上将原特征替换成深度特征提出 DeepSRDCF 算法。DeepSRDCF 探索了不同卷积层的深度特征对跟踪效果的不同影响,实验表明使用第一卷积层所提取的特征的效果优于使用其他层特征的效果,同时相对于深层特征,浅层特征拥有更加快速的提取速度,保证了算法的运行速度。该算法证明了浅层卷积层比深层卷积层更适用于目标跟踪并且拿下了 VOT2015 大赛第二名,传统的 SRDCF 算法名次为第四名。Danelljan 等[45]根据 DCF 算法改进提出了连续卷积算子跟踪器(continuous convolution operator tracker, C-COT)。将原算法中使用单一分辨率的特征改为使用多分辨率特征并隐式插值模型实现多特征自然整合,改进后算法还运用了连续空间域中的区分卷积算子理论框架,提高跟踪精度并取得了 VOT2016 大赛排名第一,但是由于使用高维特征需要训练大量滤波器,使得算法复杂度高,导致算法速度较慢。之后 Danelljan 等从样本集大小、滤波器构造和模型更新策略对 C-COT 进行改进提出了高效卷积算子跟踪器(efficient convolution operator tracker, ECO)算法[46]。该算法使用因式分解的卷积操作对滤波器进行降维,在样本集方面利用混合高斯模型对具有类内相似性和类间差异性的样本组合,在更新策略方面采用间断式模型更新策略间隔多帧完成一次更新,有效地避免模型漂移,同时降低了算法复杂度,提高了跟踪精度的同时也大幅提高了算法效率,达到了同时期、同类算法的顶峰。如表 5 为部分算法在 OTB50 和 OTB100 标准集上的表现情况[47]。

端到端的分类深度网络跟踪最早由 Nam 等[48]提出多域卷积神经网络跟踪器(multi-domain convolutional neural network tracker, MDnet)算法并取得了 VOT2015 大赛冠军。该网络由特征提取通用的共享层和多分支检测的全连接层两部分组成。MDnet 算法包括两个阶段,训练阶段和目标跟踪阶段。训练阶段通过对数据集训练得到全连接层,共享层对所有数据通用以便获得动态目标通用性深度特征;目标跟踪保留固定共享层,并根据新的数据建立新的全连接层,二者结合组成新的端到端网络。MDnet 利用卷积操作将目标跟踪作为目标与背景的二分类问题处理,导致其在跟踪过程中易受目标相似物的干扰,并且对目标遮挡的鲁棒性较差。

针对该问题,Qi 等[49]在 MDnet 的基础上进行改进,提出规模和状态感知跟踪器(scale-and-state aware tracker, SSAT)算法。该算法包含三个主要组件:使用基础跟踪器进行跟踪、边界框细化和目标状态分类,针对目标旋转、变形问题,利用水平集图像分割或边界框回归技术实现了边界框自适应缩紧。针对目标遮挡问题设计单独的卷积神经网络(convolution neural network, CNN)对目标遮挡状态和非遮挡状态进行分类,当目标被遮挡时该帧图像将不会被用于跟踪器更新。如图 3 所示为 SSAT 网络结构图。

Nam 等[50]还提出了利用多个 CNN 构成树结构并协作以完成目标跟踪的树结构卷积神经网络跟踪器(tree-structured convolutional neural network tracker,

表 5 部分算法在 OTB50 和 OTB100 标准集上的表现[47]

Table 5 Performance of some algorithms on OTB50 and OTB100 standard set[47]

算法	OTB100		OTB50	
	成功率	精准率	成功率	精准率
KCF	0.514	0.740	0.477	0.696
SAMF	0.579	0.785	0.541	0.752
DSST	0.503	0.737	0.470	0.693
FDSST	0.530	0.713	0.479	0.644
SRDCF	0.626	0.838	0.598	0.789
Deep-SRDCF	0.641	0.849	0.635	0.851
C-COT	0.672	0.899	0.671	0.896
ECO	0.709	0.928	0.689	0.910

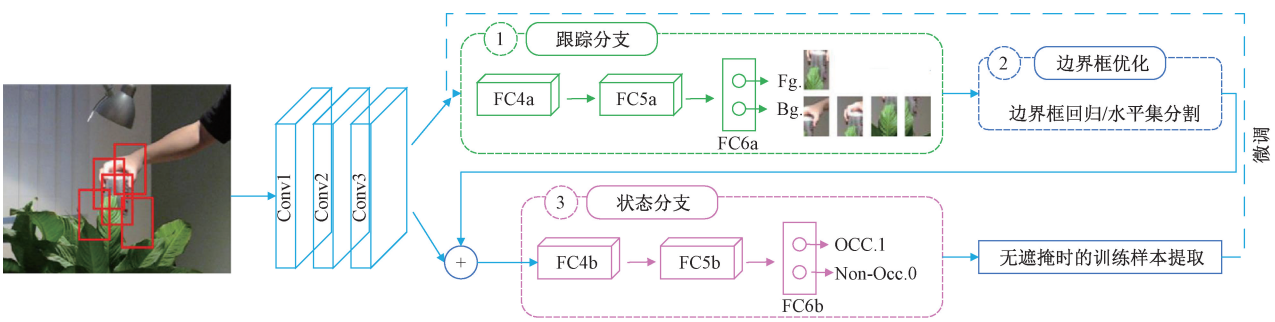

图 3 SSAT[49]网络结构图

Fig. 3 SSAT[49] network structure

TCNN)。TCNN 对每个 CNN 进行可靠性评估以确定权重,CNN 模型可靠性越高权重约大,并利用所有 CNN 进行加权检测,充分发挥每个 CNN 的作用并以十帧图像创建一个新节点,同时删除一个距离创建时刻最久的旧节点,并选择使新节点可靠性最高的节点作为父节点,完成在线微调,避免邻近帧之间的过拟合,以此保证 TCNN 在整个跟踪过程中的可靠性。

为解决尺度变化问题,Wang 等提出了多级深度特征跟踪器(multi-level deep feature tracker, MLDF)算法。该算法通过融合多级特征预训练一个多级网络(multi-level network, MLN)。该网络用于确定目标中心位置,然后使用尺度预测网络(scale prediction network, SPN)处理尺度变化问题。如表 6 为 MDnet、SAnet 和 TCNN 在 VOT2016 中的测试情况[12]。

表 6　MDnet、SSAT、MLDF 和 TCNN 性能对比[12]

Table 6　Performance comparison of MDnet, SSAT, MLDF and TCNN[12]

算法	EAO	*A*	*R*	EFO
TCNN	0.325	0.554	0.268	1.049
SSAT	0.321	0.577	0.291	0.475
MLDF	0.311	0.490	0.233	1.483
MDnet	0.257	0.541	0.337	0.534

注:EAO 为平均重叠期望;*A* 为平均重叠率;*R* 为鲁棒性;EFO 为等效滤波运算速率。

4.3　孪生网络目标跟踪算法

孪生网络结构是由两个网络结构或多个网络结构共同组成,并且该两个网络结构参数共享,一种特殊的神经网络架构。在目标跟踪领域,孪生网络算法同时接收两个图片并分别输入两个子神经网络进行训练,同时两个子神经网络权值共享,通过对不同子神经网络获得的图像特征用于获取相关响应图像进行分析计算,以此判断目标位置并完成目标跟踪。

Tao 等[52]提出的孪生网络目标跟踪(siamese instance search tracking, SINT)算法是首次将孪生网络思想应用到目标跟踪中。该算法包含两个步骤,离线训练和在线跟踪。离线训练使用阿姆斯特丹视频跟踪训练集(Amsterdam library of ordin-ary videos for tracking, ALOV)算法,采用最小化损失函数的方式进行。在线跟踪过程中分别将初始帧目标图像和当前帧图像分别输入两个神经网络,并返回最大响应区域作为目标区域。Bertinetto 等[53]提出全卷积孪生网络(fully-convolutional siamese network, SiamFC),该算法与 SINT 算法基本原理类似,都是首先通过离线训练然后进行在线跟踪。图 4 为 SiamFC 网络结构图,其中 127×127×3 为目标图像、255×255×3 为当前图像,神经网络采用 AlexNet 结构,通过对响应图进行处理则可以实现目标跟踪。以 SiamFC 为代表的早期的孪生网络算法在跟踪速度上表现良好,但是应对尺度变化等问题还有待提高。

为了解决 SiamFC 在面对目标尺度变化时鲁棒性较差的问题,Li 等[54]借鉴了目标检测 Faster R-CNN 算法[55]将区域生成网络(region proposal network, RPN)加入 SiamFC 框架提出了 SiamRPN。将孪生网络结构分为一个分类分支和一个检测分支,其中分类分支用来判断是否为目标,检测分支用边界回归来替代多尺度检测,得到了更好的目标尺度。如表 7 所示为 VOT2018 大赛前五算法。该算法取得综合排名第三名好成绩,并且在实时跟踪赛项中远高于相近排名的算法[14]。Zhu 等[56]为提高 SiamRPN 跟踪器抗干扰能力对其进行进一步优化,提出了 DaSiamRPN 算法。解决了传统孪生网络目标跟踪过程中只能区分前景和非语义背景导致的复杂背景下性能无法保证问题。

另一种是由 Wang 等[57]提出的可同时实现对目标分割与目标跟踪的 SiamMask,该算法在 SiamFC 基础上增加多种功能模块进行改进。SiamMask 利用掩膜分支实现目标二值化分割,并将目标分割任务中的细节性特征用于目标跟踪任务;利用边界分支得到目标候选区域;利用得分分支完成候选区域的筛选,最终可实现实时跟踪。SiamMask 首次提出使用掩膜的方式标记目标。He 等[58]提出的 SA-Siam 首次将语义信息分支和外观信息分支加入到目标跟踪框架,并且每个分支都是独立的孪生网络,所以该算法结构又叫做双重孪生的网络结构。两个

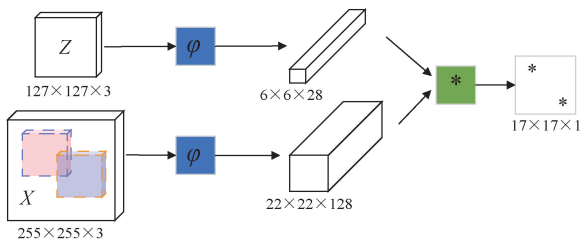

图 4　SiamFC[53]网络结构图

Fig. 4　Network structure of SiamFC[53]

表 7　VOT2018 前五名算法[14]

Table 7　Top five algorithm in VOT2018[14]

算法	短时跟踪			实时跟踪		
	EAO	*A*	*R*	EAO	*A*	*R*
LADCF	0.39	0.50	0.15	0.06	0.31	1.36
MFT	0.39	0.50	0.14	0.06	0.33	1.59
SiamRPN	0.38	0.59	0.27	0.38	0.34	0.28
UPDT	0.38	0.54	0.28	0.38	0.59	1.36
RCO	0.38	0.51	0.16	0.06	0.33	1.70

分支在训练过程中遵循特征互异性原则,通道加权机制可以充分利用并融合具有互补性质的语义特征和外观特征使得算法在保证实时性的基础上可以有效地判断目标外观变化,相比于 SiamFC 算法跟踪精度和鲁棒性都得到了显著提高。Li 等[59]认为视觉跟踪中的目标是任意形式任意类别的,因此预先训练的深度特征在目标跟踪领域贡献并不大。针对这一观点提出了目标感知深度跟踪算法(target-aware deep tracking, TADT)。该算法使用回归损失和排名损失生成目标运动特征和目标尺度感知特征,并根据反向传播梯度特征和目标感知特征对每一个神经网络赋权值从而确定目标。

Zhang 等[60]对孪生网络的骨干网络进行探索并提出更广更深的孪生网络(deeper and wider siamese networks, SiamDW)算法。作者总结当前基于孪生网络的目标跟踪算法均是使用 AlexNet 网络框架,并通过实验更换网络框架对比现有网络框架下目标跟踪效果,发现效果均不如 AlexNet 网络框架,且随着网络深度增加效果越差。随后作者从步长、填充、感受野等角度分析原因,并发现各影响因素之间并不完全独立,且对于全卷积的孪生网络,填充操作会导致跟踪性能退化。SiamDW 使用新的残差模块(cropping-inside residual, CIR)来控制接收域的大小和网络步长,削弱由填充操作引起的偏差对目标跟踪的负面影响,这种方法改进下的 SiamFC 算法和 SiamRPN 算法在性能上都有所提高。Li 等[61]通过在 SiamRPN 的基础上改进骨干网络,得到使用多层特征融合的深层孪生网络模算法 SiamRPN + +算法。该算法使用了多级级联方式完成特征的逐层聚合,有助于目标跟踪性能的提升。在 VOT2017 性能测评中对 SiamRPN 骨干网络进行改进后得到的 SiamRPN + +,其各项性能均得到优化。SiamMargin 算法在 SiamRPN + +基础上将判别式特征嵌入到孪生神经网络中,并用于离线训练,使得算法在保证速度的同时提高了准确率。在线更新方面选用基于 ROIAlign 的更新方式,通过 ROIAlign 对当前帧特征进行建模,使用移动平均策略更新模板,有效地应对了目标外观变化问题[15]。

除了具有对称结构的孪生网络完成目标跟踪的方法外,还有部分方法虽然借鉴了孪生网络的并行框架,但是结构并不严格对称的迭代最大化跟踪(accurate tracking by overlap maximization, ATOM)算法[62]。ATOM 主要包括目标估计模块和目标分类模块,并借鉴目标检测领域的 IoU-Net[63]设计了用于目标估计模块的 IoU-Predictor 网络。如图 5 为 ATOM 网络模型。Lukezic 等[64]将在 ATOM 基础上引入相关滤波器模型,提高了跟踪速度。

4.4 Transformer 框架目标跟踪算法

Transformer[65]框架不同于传统深度学习框架,是一种基于注意力机制的框架。该框架早期服务于自然语言,近年才被应用于计算机视觉。Transformer 框架本质是一个编码解码的结构。在计算机视觉中主要被用于捕捉图像上的目标感受野。

Wang 等[66]提出 TrDiMP,将 Transformer 作为中间模块用于特征提取,并有效地提升了特征质量。同时对传统的 Transformer 进行改良,将编码和解码两部分成两个并行分支。编码部分对模板部分进行提取并使用注意力机制进行增强;搜索部分有解码进行处理。该算法虽然在检测精度方面取得较好效果,但是对于目标遮挡、目标消失仍需要优化。Chen 等[67]提出 Transt 算法,在孪生网络的基础上引入 Transformer 框架,利用该框架中的注意力机制用于避免目标跟踪过程中的语义丢失问题。该算法虽然已经取得较好效果,但是并没有充分利用背景信息,还具有较大的提升空间。

目前基于 Transformer 框架的目标跟踪算法尚

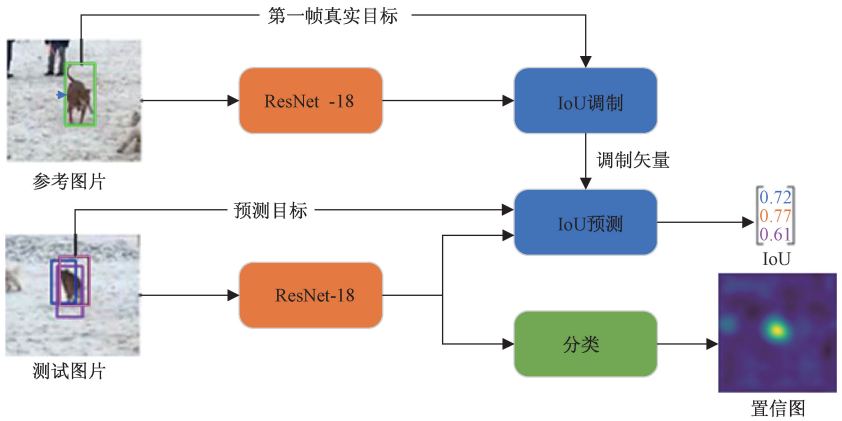

图 5 ATOM[62]网络模型

Fig. 5 Atom[62] network mode

处于起步阶段,但是已经在现有的数据集上取得了较好的效果,潜力巨大。

5　总结与展望

从生成式目标跟踪算法和判别式目标跟踪算法作为切入点对目标跟踪算法进行了梳理和分析。由上文可知,对于生成式模型来说主要缺点表现在忽略背景信息仅仅关注目标信息,导致往往背景存在与目标相似情况下无法正确跟踪目标,并且生成式模型建模使用特征相对单一,对于光照变化、运动模糊、分辨率低、目标旋转等问题不能很好地兼顾,从而影响跟踪的准确性。由于原理的限制,生成式算法的综合性能很难超越判别式算法的综合性能,因此目前大部分科研人员将精力投入到判别式模型研究中。基于判别式模型的目标跟踪算法也将是未来目标跟踪算法的主力军。结合本文中不同算法对比分析可得未来研究方向包括以下几个方面。

(1)有效的特征选择以及多特征融合。在讨论相关滤波算法时发现,目标跟踪的特征选择经历了由简单到复杂,由单一到多特征融合的过程,如表5所示多特征的SAMF算法性能上优于单一特征KCF算法。不同特征在目标跟踪过程中均可以发挥不同的优势,因此后续研究过程中可以尝试更多的特征组合方式,探究不同特征对于不同场景下目标跟踪的影响,从而寻找较优的特征组合方式。

(2)相关滤波算法与深度学习算法相融合。深度学习算法具有跟踪准确、精度高等优点,相关滤波算法精度不及深度学习算法,但是运算速度较快。将两种类型算法相互结合取长补短,C-COT、ECO等算法为融合算法均取得良好效果(C-COT在VOT2016中第一名、ECO在VOT2017中第四名)。因此相关滤波算法与深度学习算法相融合具有巨大的潜力,如何合理融合两种算法发挥各自优势是值得探索的问题。

(3)从近年来VOT[14-16]大赛情况来看,基于孪生网络的目标跟踪算法比例在不断提高,并且取得骄人的成绩。基于孪生网络的目标跟踪算法是未来目标跟踪算法主要发展之一。目前随着深度学习技术发展提出的网络结构众多,但是应用于孪生网络结构却比较单一。Zhan等[60]进行过探索并总结出一些规律,但是并没有找到合适的方法将新的神经网络结构引入孪生网络目标跟踪算法中。如何将新型网络结构合理的加入孪生网络模型中,在保证实时性的同时尽可能地提高跟踪进度,是值得深入研究和探索的。

(4)Transformer框架作为最新的网络框架近年被应用于目标跟踪领域,该框架的引入注重了连续帧图像中的时序关系,突出了目标跟踪算法与目标检测算法本质区别。目前该框架下的目标跟踪算法刚刚开始发展,在短时跟踪下表现效果良好,在长时跟踪、目标丢失重检仍然存在巨大的提升空间。

参　考　文　献

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

[2] 徐　宁, 王娟娟, 郭晓雨, 等. 判别式相关滤波器的目标跟踪综述[J]. 小型微型计算机系统, 2020, 41(12): 2484-2493.
Xu Ning, Wang Juanjuan, Guo Xiaoyu, et al. Survey of target tracking based on discriminant correlation filter[J]. Journal of Chinese Mini-Micro Computer Systems, 2020, 41(12): 2484-2493.

[3] 李　玺, 查宇飞, 张天柱, 等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报, 2019, 24(12): 2057-2080.
Li Xi, Zha Yufei, Zhang Tianzhu, et al. Survey of object tracking algorithms based on deep learning[J]. Journal of Image and Graphics, 2019, 24(12): 2057-2080.

[4] 王海涛, 王荣耀, 王文嶂, 等. 目标跟踪综述[J]. 计算机测量与控制, 2020, 28(4): 1-6, 21.
Wang Haitao, Wang Rongyao, Wang Wenhao, et al. Survey of target tracking[J]. Computer Measurement and Control, 2020, 28(4): 1-6, 21.

[5] 陈云芳, 吴　懿, 张　伟. 基于孪生网络结构的目标跟踪算法综述[J]. 计算机工程与应用, 2020, 56(6): 10-18.
Chen Yunfang, Wu Yi, Zhang Wei. Survey of target tracking algorithms based on twin network structure[J]. Computer Engineering and Applications, 2020, 56(6): 10-18.

[6] 单玉刚, 胡卫国. 尺度方向自适应视觉目标跟踪方法综述[J]. 计算机工程与应用, 2020, 56(9): 13-23.
Shan Yugang, Hu Weiguo. Survey of scale direction adaptive visual object tracking methods[J]. Computer Engineering and Applications, 2020, 56(9): 13-23.

[7] 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.

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

[9] Kristan M, Matas J, Leonardis A, et al. The visual object trackingVOT2013 challenge results[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2013: 98-111.

[10] Kristan M, Leonardis A, Matas J, et al. The visual object trackingVOT2014 challenge results[C]// Proceedings of the European Conference on Computer Vision Workshops. Cham: Springer, 2014: 191-217.

[11] Kristan M, Matas J, Leonardis A, et al. The visual object trackingVOT2015 challenge results[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2015: 564-586.

[12] Kristan M, Leonardis A, Matas J, et al. The visual object trackingVOT2016 challenge results[C]//Proceedings of the European Conference on Computer Vision Workshops. Cham: Springer, 2016: 777-823.

[13] Kristan M, Leonardis A, Matas J, et al. The visual object tracking VOT2017 challenge results[C] //Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 1949-1972.

[14] Kristan M, Leonardis A, Matas J, et al. The sixth visual object trackingVOT2018 challenge results[C]//Proceedings of the European Conference on Computer Vision Workshops. Cham: Springer, 2018: 3-53.

[15] Kristan M, Matas J, Leonardis A, et al. The seventh visual object trackingVOT2019 challenge results [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2019: 2206-2241.

[16] Kristan M, Leonardis A, Matas J, et al. The eighth visual object tracking VOT2020 challenge results [C]//European Conference on Computer Vision. Cham: Springer, 2020: 547-601.

[17] Dorin C, Visvanathan R, Peter M. Real-time tracking of non-rigid objects using mean shift [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2000: 142-149.

[18] 王 凯, 高 勇, 吴 敏, 等. 改进 ViBe 算法与 Meanshift 结合的目标跟踪方法[J]. 科学技术与工程, 2020, 20(21): 8663-8670.
Wang Kai, Gao Yong, Wu Min, et al. An moving target tracking method based on improved ViBe algorithm and meanshift[J]. Science Technology and Engineering, 2020, 20(21): 8663-8670.

[19] Allen J G, Xu R Y D, Jin J S. Tracking using camshift algorithm and multiple quantized feature spaces[C]//Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing. New York: Australian Computer Society, 2003: 3-7.

[20] 董恩增, 陈津宇, 焦迎杰, 等. 一种改进的 Camshift 算法及其 ARM 实现[J]. 科学技术与工程, 2018, 18(6): 279-284.
Dong Enzeng, Chen Jinyu, Jiao Yingjie, et al. An improved Camshift algorithm and its ARM implementation[J]. Science Technology and Engineering, 2018, 18(6): 279-284.

[21] Vojir T, Noskova J, Matas J. Robust scale-adaptive mean-shift for tracking[J]. Pattern Recognition Letters, 2014, 49: 250-258.

[22] StengerB, Thayananthan A, Torr P H, et al. Model-based hand tracking using a hierarchical Bayesian filter[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(9): 1372-1384.

[23] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.

[24] 杨鹏生, 吴晓军, 张玉梅. 改进扩展卡尔曼滤波算法的目标跟踪算法[J]. 计算机工程与应用, 2016, 52(5): 71-74.
Yang Pengsheng, Wu Xiaojun, Zhang Yumei. Target tracking algorithm based on improved extended Kalman filter algorithm[J]. Computer Engineering and Applications, 2016, 52(5): 71-74.

[25] 许红香, 白星振, 董礼廷, 等. 基于无迹卡尔曼滤波和协方差交叉融合的分层多簇无线传感器网络多速率跟踪算法[J]. 科学技术与工程, 2020, 20(27): 11149-11154.
Xu Hongxiang, Bai Xingzhen, Dong Liting, et al. Multi-rate tracking algorithm for hierarchical multi-cluster wireless sensor networks based on unscented Kalman filter and covariance intersection fusion[J]. Science Technology and Engineering, 2020, 20(27): 11149-11154.

[26] Arulampalam M S, Maskell S, Gordon N, et al. A tutorialon particle filters for online nonlinear/non-Gaussian Bayesian tracking [J]. IEEE Transactions on Signal Processing, 2002, 50(2): 174-188.

[27] 赵宗超, 李东兴, 赵蒙娜. 基于改进的引导图像滤波和深度去噪自编码器的微弱目标跟踪算法[J]. 科学技术与工程, 2020, 20(14): 5696-5701.
Zhao Zongchao, Li Dongxing, Zhao Mengna. A tracking algorithm for dim target based on guided image filter and deep denoising autoencoder[J]. Science Technology and Engineering, 2020, 20(14): 5696-5701.

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

[29] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detectionwith kernels[C]//European Conference on Computer Vision. Berlin: Springer, 2012: 702-715.

[30] 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. Portland: IEEE Press, 2014: 1090-1097.

[31] 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, 2014, 37(3): 583-596.

[32] Li Y, Fu C, Ding F, et al. AutoTrack:towards high-performance visual tracking for UAV with automatic spatio-temporal regularization[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11920-11929.

[33] Li Y, Zhu J. A scale adaptive kernel correlation filtertracker with feature integration[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2014: 254-265.

[34] Danelljan M, Hager G, Khan F, et al. Accurate scale estimation for robust visual tracking[C]//British Machine Vision Conference. Nottingham: BMVA Press, 2014: 1-11.

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

[36] 宸梦楠, 白一帆, 邓红霞, 等. 变电站场景下高置信度策略的 DSST 目标检测及跟踪[J]. 科学技术与工程, 2018, 18(23): 257-264.
Yi Mengnan, Bai Yifan, Deng Hongxia, et al. DSST target detection and tracking for high confidence strategy in substation scene [J]. Science Technology and Engineering, 2018, 18(23): 257-264.

[37] Li Y, Zhu J, Hoi S C H. Reliable patch trackers: robust visual tracking by exploiting reliable patches[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 353-361.

[38] Akin O, Erdem E, Erdem A, et al. Deformable part-based tracking by coupled global and local correlationfilters[J]. Journal of

Visual Communication and Image Representation, 2016, 38: 763-774.

[39] Montero A S, Lang J, Laganiere R. Scalable kernel correlation filter with sparse feature integration [C]//Proceedings of the IEEE International Conference on Computer Vision Workshop. Piscataway: IEEE Press, 2015: 587-594.

[40] Danelljan M, Hager G, Khan F, et al. Learning spatially regularized correlation filters for visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 4310-4318.

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

[42] Kiani H G, 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.

[43] Ma C, Huang J B, Yang X, et al. Hierarchical convolutional features for visual tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE Press, 2015: 3074-3082.

[44] 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. Santiago: ICCVW, 2015: 621-629.

[45] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking [C]// Proceedings of the European Conference on Computer Vision. Springer: ECCV, 2016: 472-488.

[46] 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: 6638-6646.

[47] 罗 元, 王薄宇, 陈 旭. 基于深度学习的目标检测技术的研究综述[J]. 半导体光电, 2020, 41(1): 1-10.
Luo Yuan, Wang Boyu, Chen Xu, Research progresses of target detection technology based on deep learning [J]. Semiconductor Optoectronics, 2020, 41(1): 1-10.

[48] Nam H, Han B. Learning multidomain convolu-tional neural networks for visual tracking [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 4293-4302.

[49] Qi Y, Qin L, Zhang S, et al. Robust visual tracking via scale-and-state-awareness[J]. Neurocomputing, 2019, 329: 75-85.

[50] Nam H, Baek M, Han B. Modeling and prop-agating CNNs in a tree structure for visual tracking[J]. ArXiv, 2016: 1608.07242.

[51] Wang L, Ouyang W, Wang X, et al. Visual tracking with fully convolutional networks[C]// Proceedings of the IEEE international conference on computer vision. Piscataway: IEEE Press, 2015: 3119-3127.

[52] Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1420-1429.

[53] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]//Proceedings of the European Conference on Computer Vision. Munich: ECCV Press, 2016: 850-865.

[54] Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8971-8980.

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

[56] Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking[C]// Proceedings of the European Con-ference on Computer Vision. Munich: ECCV Press, 2018: 103-119.

[57] Wang Q, Zhang L, Bertinetto L, et al. Fast online object tracking and segmentation: a unifying approach[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 1328-1338.

[58] He A F, Luo C, Tian X M, et al. A twofold siamese network for real-time object tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4834-4843.

[59] Li X, Ma C, Wu B, et al. Target-aware deep tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 1369-1378.

[60] Zhang Z P, Peng H W. Deeper and wider Siamese networks for real-time visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4591-4600.

[61] Li B, Wu W, Wang Q, et al. SiamRPN++: evolution of siamese visual tracking with very deep networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4277-4286.

[62] Danelljan M, Bhat G, Khan F S, et al. ATOM: Accurate tracking by overlap maximization[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4655-4664.

[63] Jiang B, Luo R, Mao J, et al. Acquisition of localization confidence for accurate object detection[C]//European Conference on Computer Vision. Spring:ECCV,2018: 816-832.

[64] Lukezic A, Matas J, Kristan M. D3s-adiscriminative single shot segmentation tracker[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7131-7140.

[65] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Proceedings of the 31 st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc.,2017: 6000-6010.

[66] Wang N, Zhou W, Wang J, et al. Transformer meets tracker: exploiting temporal context for robust visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021:2103.11681.

[67] Chen X, Yan B, Zhu J, et al. Transformer Tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press,2021:2103.15436.