

# 目标跟踪算法综述

卢湖川<sup>1</sup> 李佩霞<sup>1</sup> 王 栋<sup>1</sup>

**摘 要** 作为计算机视觉领域的一个重要课题,在线目标跟踪在导弹制导、视频监控、无人机跟踪等众多领域中具有重要作用. 尽管现在已有大量研究,但是仍然存在很多问题亟待解决,如光照变化、尺度变化、形变、遮挡和相机移动等. 为了更清楚地梳理现存的算法,文中对典型的目标跟踪算法进行分析总结. 首先,简单介绍研究意义及相关工作. 然后,从传统算法和深度学习算法两方面对经典算法进行概述和分析. 最后,讨论算法目前存在的问题,给出未来的研究趋势.

**关键词** 目标跟踪, 相关滤波, 深度学习

**引用格式** 卢湖川,李佩霞,王 栋.目标跟踪算法综述. 模式识别与人工智能, 2018, 31(1): 61–76.

**DOI** 10.16451/j.cnki.issn1003-6059.201801006

## Visual Object Tracking: A Survey

LU Huchuan<sup>1</sup>, LI Peixia<sup>1</sup>, WANG Dong<sup>1</sup>

**ABSTRACT** Online object tracking is a fundamental problem in computer vision and it is crucial to application in numerous fields such as guided missile, video surveillance and unmanned aerial vehicle. Despite many studies on visual tracking, there are still many challenges during the tracking process including illumination variation, rotation, scale change, deformation, occlusion and camera motion. To make a clear understanding of visual tracking, visual tracking algorithms are summarized in this paper. Firstly, the meaning and the related work are briefly introduced. Secondly, the typical algorithms are classified, summarized and analyzed from two aspects: traditional algorithms and deep learning algorithms. Finally, the problems and the prediction of the future of visual tracking are discussed.

**Key Words** Object Tracking, Correlation Filter, Deep Learning

**Citation** LU H C, LI P X, WANG D. Visual Object Tracking: A Survey. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 61–76.

计算机视觉是研究如何让计算机从图像或视频中获取信息并进行“感知”的科学,它可以看作是人类视觉和感知的延伸. 近二十年来,随着计算机硬件水平、图像视频处理技术及人工智能技术的迅猛发展,计算机视觉技术在智能控制(如机器人、无人机等)、决策系统(如监控、检测和识别系统等)、信息检索(如多媒体数据查询、图像搜索引擎等)、人机交互等许多关系到人类活动和生活领域均得到广泛

应用.

在计算机视觉领域中,基于视频的目标跟踪(也称为视觉跟踪)一直都是一个重要课题和研究热点. 视觉跟踪通过在连续的视频图像序列中估计跟踪目标的位置、形状或所占区域,确定目标的运动速度、方向及轨迹等运动信息,实现对运动目标行为的分析和理解,以便完成更高级的任务. 目标跟踪无论在军事国防还是民用安全方面都具有重要的研究意义和广阔的应用前景. 具体研究和应用包括:

1) 智能视频监控. 目前,视频监控是目标跟踪技术最活跃的应用领域<sup>[1-6]</sup>. 视频监控系统通过对视频中的目标进行自动识别、跟踪及更高级的语义分析处理,从而实现对特定场所(如军事部门、政府机关、车站、机场、银行、校园、住宅区等)进行监控和预警. 通过对视频的分析 and 理解,智能视频监控能自动提取对监控和预警有用的关键信息,尽可能地

收稿日期:2017-09-19;录用日期:2017-12-05

Manuscript received September 19, 2017;

accepted December 5, 2017

本文责任编辑 高新波

Recommended by Associate Editor GAO Xinbo

1. 大连理工大学 信息与通信工程学院 大连 116024

1. School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024

减少人为干预,提高监控效率,同时减轻人的工作负担。作为智能视频监控的关键技术,视觉跟踪算法的研究具有巨大的产业价值。

2)现代化军事。视觉跟踪技术是当今以及未来现代化军事的研究热点<sup>[7-8]</sup>,主要集中在精确制导、飞行控制、无人机侦察、靶场测量及武器观测瞄准等方面。如在精确制导方面,早在 20 世纪 70 年代之前,美国就开始研制视频制导导弹和炸弹,并在越南战争和第四次中东战争中用于实战,取得引人注目的作战效果。现在大多数制导导弹均采用可见光、红外线及其它射线相结合的视觉跟踪技术。因此,对于应对现代化局部战争,视觉跟踪技术的研究也具有重要意义。

3)基于视频的人机交互。基于视频的人机交互技术通过对由摄像头或其它采集设备拍摄的视频进行分析,实现人与计算机之间的“对话”。计算机软件可以对人的眼神<sup>[9]</sup>、表情<sup>[10-11]</sup>、手势<sup>[12-13]</sup>、姿态及身体动作<sup>[14-15]</sup>等进行分析和理解,从而完成目标跟踪。一个完整的目标跟踪系统通常包括目标的检测、提取、识别和跟踪 4 个环节。计算机想要获得如表情、手势、姿态等高级语义信息,必须首先实现对人脸、人眼及人体的定位和跟踪。因此,视觉跟踪也是人机交互的重要环节之一。

4)智能交通系统。随着目前城市机动车辆的逐渐增多和车流密度的逐渐增大,智能交通系统成为辅助交通调度不可或缺的工具。智能交通系统利用目标检测和跟踪技术<sup>[16-18]</sup>,对车辆进行实时检测和跟踪,进一步自动获取车辆的流量、车速、车流密度、道路拥塞状况等信息,辅助交通调度。所以,在视频中快速定位和跟踪目标是智能交通的关键。

5)智能视觉导航。随着摄像机硬件设备和视频处理技术的快速发展,基于计算机视觉的智能导航技术具有较广泛的应用(如智能机器人、无人驾驶汽车及无人驾驶飞机等)。智能导航技术利用摄像头或其它传感器对周围环境和运动物体进行检测和跟踪<sup>[19-20]</sup>,使机器能安全运转及获取有用信息进行特殊的作业或任务。如何在摄像机运动时快速精确地实现检测和跟踪是智能视觉导航的关键。

6)三维重建。基于视频图像序列的三维重建(如人体三维运动和姿态重构<sup>[21-22]</sup>、城市建模<sup>[23-24]</sup>等)一直是计算机视觉领域的一个研究热点,是数字博物馆、智能导航、计算机动画、人机交互等诸多应用的关键技术。视觉跟踪技术能够计算摄像头或人体各个关键部分的运动参数,为后续的三维重构提供支持。除此之外,视觉跟踪技术还在航空航

天<sup>[25]</sup>、医疗诊断<sup>[26-27]</sup>、虚拟现实<sup>[28]</sup>、增强现实<sup>[29]</sup>等方面应用广泛。

1 目标跟踪

在线视觉跟踪的基本框架如图 1 所示。一般在线视觉跟踪具有 4 个基本组成部分:特征提取、运动模型、外观模型、在线更新机制。

特征提取。在线视觉跟踪算法要求提取的视觉特征既能较好地描述跟踪目标又能快速计算。常见的图像特征有灰度特征<sup>[30]</sup>、颜色特征<sup>[31]</sup>、纹理特征<sup>[32]</sup>、Haar-like 矩形特征<sup>[33]</sup>、兴趣点特征<sup>[34]</sup>、超像素特征<sup>[35]</sup>等。

运动模型 (Motion Model)。运动模型旨在描述帧与帧目标运动状态之间的关系,显式或隐式地在视频帧中预测目标图像区域,并给出一组可能的候选区域。经典的运动模型有卡尔曼滤波 (Kalman Filtering)<sup>[36]</sup>、粒子滤波 (Particle Filtering)<sup>[37-38]</sup>等。

外观模型。外观模型的作用是在当前帧中判决候选图像区域是被跟踪目标的可能性。提取图像区域的视觉特征,输入外观模型进行匹配或决策,最终确定被跟踪目标的空间位置。在视觉跟踪的 4 个基本组成中,外观模型处于核心地位,如何设计一个鲁棒的外观模型是在线视觉跟踪算法的关键。

在线更新机制。为了捕捉目标(和背景)在跟踪过程中的变化,在线视觉跟踪需要包含一个在线更新机制,在跟踪过程中不断更新外观模型。常见的外观模型更新方式有模板更新、增量子空间学习算法及在线分类器等。如何设计一个合理的在线更新机制,既能捕捉目标(和背景)的变化又不会导致模型退化,也是在线视觉跟踪研究的一个关键问题。

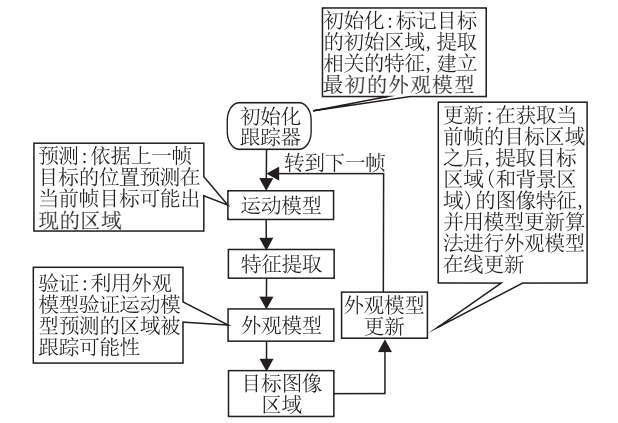


图 1 在线视觉跟踪的基本框架图

Fig. 1 Basic framework of online visual tracking

以前目标跟踪的研究更多局限于在可控条件(如静态背景、光照均匀等)下对特定目标(如人脸、行人、车辆等)实施跟踪,而在线视觉跟踪旨在对现实场景中的任意目标进行跟踪。由于现实场景的复杂性和目标物的不确定性,实现在线视觉跟踪存在诸多挑战,部分在线视觉跟踪的挑战如图 2 所示。



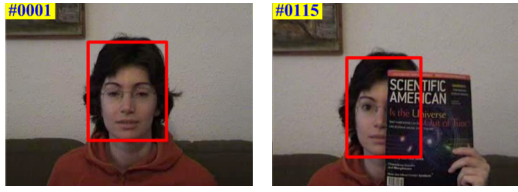
(a) 形状变化  
(a) Deformation



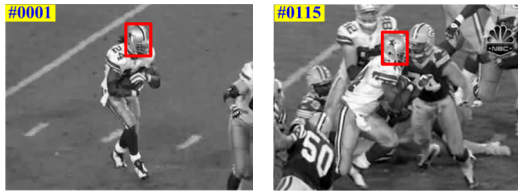
(b) 姿态和光照变化  
(b) Gesture and illumination variation



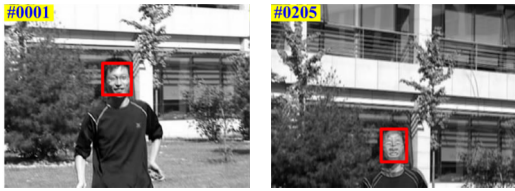
(c) 光照和尺度变化  
(c) Illumination and scale variation



(d) 遮挡  
(d) Occlusion



(e) 背景杂乱  
(e) Background cluttered



(f) 运动模糊  
(f) Motion blur

图 2 在线视觉跟踪的挑战  
Fig. 2 Challenge for online visual tracking

这些挑战主要可分为由目标内在变化引发的挑战(如姿态变化、形状变化等)和由外在因素引发的挑战(如光照变化、运动模糊、遮挡、背景杂乱等),它们为实现在线视觉跟踪造成很多困难:如姿态和形状变化会造成被跟踪目标外观在不同帧之间存在较大的差异;光照通常会引起被跟踪目标外观的明暗变化并带来不可预知的阴影;遮挡会造成目标外观的不完整或给目标区域引入背景干扰;目标的快速运动会造成运动模糊,使被跟踪目标变得不易分辨;在杂乱的背景中,跟踪算法难以较好地鉴别目标和周围的背景,容易发生跟踪漂移。

## 2 在线目标跟踪算法

在线目标跟踪算法一般由人工设计特征和浅层外观模型构成,旨在利用简单有效的视觉特征和浅层匹配或分类模型设计快速鲁棒的跟踪算法,如早期均值偏移算法、子块跟踪算法、子空间算法等。这些算法能在有效计算资源情况下取得一定的跟踪效果和速度,但通常不够鲁棒。稀疏表示理论在计算机视觉方面的成果应用使研究学者不断尝试利用稀疏表示模型和稀疏特征编码解决在线目标跟踪问题,取得非常卓越的效果。但稀疏表示模型计算复杂度较高,难以实时处理。鉴于外观模型在传统视觉跟踪的核心地位,下面从外观模型建立和决策过程的角度对传统视觉跟踪算法进行概述。

### 2.1 基于均值偏移的目标跟踪

作为一种典型的非参数统计工具,均值偏移(Mean Shift)<sup>[39]</sup>算法本质上是基于梯度上升的局部寻优算法,广泛应用于模式识别、数字图像处理和计算机视觉等领域。Mean Shift 跟踪算法一直是目标跟踪领域研究的一个热点,很多研究学者不断进行改进,使其更加稳定和鲁棒。Comaniciu 等<sup>[31,36]</sup>最先



将 Mean Shift 算法用于跟踪领域,实现简单,速度较快,对目标的非刚性变化、旋转等挑战都具有较好的鲁棒性,但是不能较好地解决遮挡、背景杂乱、尺度变化等挑战.为了解决局部遮挡问题,Wang 等<sup>[44]</sup>提出基于分块 Mean Shift 的跟踪算法,通过不同分块对中心位置的加权投票,降低被遮挡的目标区域对跟踪结果的影响. Collins 等<sup>[40]</sup>结合尺度空间 (Scale Space) 理论和 Mean Shift 算法,在一定程度上提升 Mean Shift 算法对目标尺度变化的稳定性.考虑到 Mean Shift 是一个局部寻优算法,容易陷入局部最小值, Hager 等<sup>[41]</sup>通过在多个尺度上选择最稳定的核宽,减少局部最小值的影响.此外,也有学者结合局部寻优的 Mean Shift 算法和全局寻优的粒子滤波 (Particle Filtering) 算法<sup>[42]</sup>,快速寻找全局最优值.

基于 Mean Shift 的跟踪算法属于**基于模板匹配的跟踪算法**,主要通过 Mean Shift 非参数局部寻优工具搜索和目标模板相似度最大的候选区域.在目标跟踪领域,基于模板匹配的跟踪算法在当前仍较活跃,如何设计有效的目标描述机制和鲁棒的相似度函数或距离函数是此类算法的关键.

2.2 基于分块的目标跟踪

早期的目标跟踪算法常采用简单有效的全局直方图,但这会丢失特征的空间位置信息,导致跟踪算法在遮挡或背景杂乱等情况下容易失败. Adam 等<sup>[43]</sup>提出子块跟踪 (Frag Track),利用分块-联合机制克服传统模板跟踪算法 (如 Mean Shift) 对遮挡敏感的问题.受到子块跟踪算法的启发,在视频跟踪领域,越来越多的学者考虑采用分块的方式,**引入特征的空间位置信息**. Wang 等<sup>[44]</sup>将分块的思想引入 Mean Shift 跟踪框架,提出基于分块 Mean Shift 的跟踪算法.通过不同分块对中心位置的加权投票降低被遮挡的目标区域对跟踪结果的影响. Jia 等<sup>[45]</sup>提出基于分块多核学习的跟踪算法,利用多个分块灰度直方图描述目标,并利用在线多核学习选择具有判决性的分块. Yang 等<sup>[46]</sup>采用由随机分块组成的词包 (Bag-of-Words) 描述被跟踪目标,提出基于词包学习的视觉跟踪算法.除此之外,基于局部图像块稀疏编码特征的跟踪算法<sup>[47-48]</sup>目前也取得较好的效果.在具体跟踪算法设计中,如何选择分块的方式和分块的大小仍需研究和探索.

2.3 基于子空间的目标跟踪

线性子空间学习 (简称子空间学习) 算法是模式识别和机器学习领域的一个研究热点,基于线性子空间的目标跟踪可看作是子空间学习算法在视觉跟踪领域的延伸和扩展.通过学习一组线性投影,实

现高维特征空间到低维空间的映射,保持或获得某种属性 (如保留整体或局部距离、提取判决性的特征等). 子空间学习算法有基于重构准则的主成分分析 (Principal Component Analysis, PCA)<sup>[49]</sup>、基于判决准则的线性判决分析 (Linear Discriminant Analysis, LDA)<sup>[50]</sup>、具有流形和图论背景的局部保持映射 (Local Preserving Projection, LPP)<sup>[51]</sup>、基于向量子空间算法的图嵌入 (Graph Embedding, GE) 框架<sup>[52]</sup>. 由于子空间学习算法的输入特征都是向量形式,所以称其为一维 (1D) 子空间算法.除此之外,子空间算法还包括多维子空间算法.

目标在跟踪过程中可能出现较大的变化,单一 PCA 子空间可能不足以描述所有的变化.为此, Yu 等<sup>[53]</sup>提出增量多流形子空间模型,相比增量视觉跟踪器 (Incremental Visual Tracking, IVT) 中仅使用一个 PCA 子空间在线学习外观模型,该算法可利用若干个 PCA 子空间,用于分段线性近似跟踪过程中目标的非线性变化.尽管该模型在处理目标的非线性变化上具有一定的优势,但由于算法复杂,导致跟踪的速度较慢.

通常子空间算法都会假设数据服从高斯分布或局部高斯分布,但在跟踪问题中背景分布较混乱,难以满足子空间算法的假设.所以基于判决性子空间的跟踪算法往往结果不稳定,未达到很好的效果.

未来基于子空间的跟踪算法的研究主要集中在如下 2 个方面:如何构造一个子空间使之能更好地描述被跟踪目标的外观变化;如何处理遮挡等异常噪声情况.

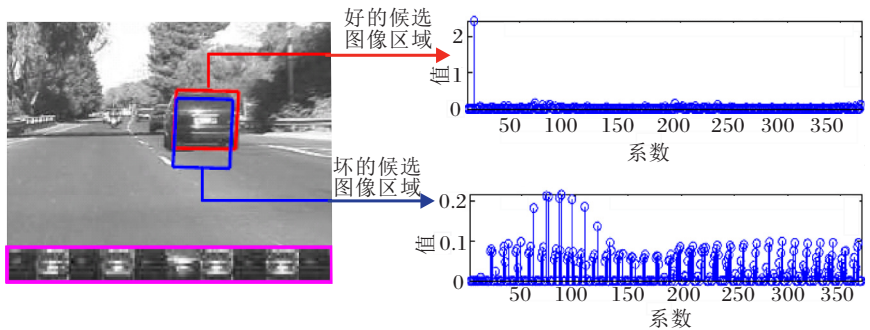
2.4 基于稀疏表示的目标跟踪

自 Donoho 等<sup>[54]</sup>完成较严密的理论工作后,压缩感知和稀疏表示算法在很多计算机视觉应用领域都得到广泛研究,包括图像去噪、图像修复、图像去模糊、超分辨重建等<sup>[55]</sup>.

Mei 等<sup>[56]</sup>提出基于  $l_1$  最小化的目标跟踪算法,首次将稀疏表示理论用于解决跟踪问题.假设被跟踪的目标能由目标模板和遮挡模板组成的字典稀疏表示,其中目标模板动态更新以捕捉目标在跟踪过程中的变化,而遮挡模板用于描述目标可能遭遇的遮挡,如图 3 所示,其中遮挡模板由正负琐碎模板组成. (b) 为好的和差的跟踪候选的稀疏表示系数的差别,左侧显示当前视频帧的图像,紫红色曲线框内显示当前存储的目标模板,红色和蓝色框分别表示一个好的和一个差的目标候选,右侧显示好的和差的候选图像区域的稀疏表示系数.可看出,好的候选表示系数明显稀疏于差的候选表示系数,这说明稀



(a) 目标模板和琐碎模板  
(a) Target template and trivial templates



(b) 候选图像区域的稀疏表示系数  
(b) Good and bad sparse representation coefficients for candidate image area

图 3 好的和差的候选图像的区别

Fig. 3 Differences between good and bad candidate images

疏限制有利于定位好的目标候选,获取精确的跟踪结果.

Yang 等<sup>[57]</sup>开始探索基于稀疏表示的目标跟踪算法,最常用的是利用稀疏表示建模目标的外观.如何合理地选择目标模板和遮挡模板建模跟踪目标及如何设计快速有效的跟踪算法是这类算法的关键.

在视觉跟踪中,稀疏表示算法还用于特征提取<sup>[48]</sup>、特征选择<sup>[47]</sup>及运动状态建模<sup>[58]</sup>等方面.

基于稀疏表示的跟踪算法是未来在线视觉跟踪领域的一个研究热点,如何更好地结合稀疏表示等理论与跟踪问题的先验知识仍是未来研究趋势.

2.5 基于判别学习的目标跟踪

基于判别学习的目标跟踪算法将目标跟踪看作一个二分类问题,试图从周围的背景中鉴别跟踪目标,因此目标跟踪的准确性和稳定性很大程度依赖于在特征空间上目标与背景的可分性.在跟踪过程中,被跟踪的目标和周围背景的外观一般都会发生变化,导致判别性的特征集合也会发生变化.因此,在线建立能适应目标和背景外观变化的判别模型是这类算法的关键.基于判别学习的目标跟踪算法基本包含基于像素或超像素级分类的跟踪和基于检测的目标跟踪这两大类.

基于像素或超像素级分类的跟踪算法通过分类当前图像帧中搜索区域内的像素或超像素,形成一个关于跟踪目标的置信图,然后利用 Mean Shift 等

模态搜索算法精确定位目标.基于像素分类的跟踪算法的流程图如图 4 所示.基于像素分类的跟踪算法速度较快,可以提供对目标的软分割(Soft Segmentation),但是不能处理遮挡和背景杂乱的情况.

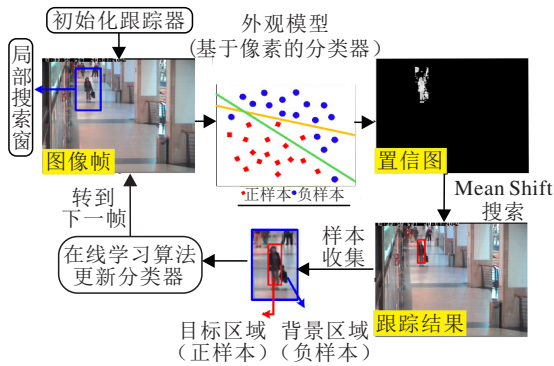


图 4 基于像素分类的跟踪算法流程图

Fig. 4 Flowchart of tracking algorithm based on pixel classification

基于检测的目标跟踪算法将目标跟踪当作一个局部搜索区域内的检测问题,通过在线学习分类器实现目标跟踪.受到人脸检测和使用检测实现跟踪的启发,学者们开始使用 Boosting 算法建立外观模型,用于视觉跟踪. Grabner 等<sup>[59]</sup>提出在线的 Adaboost 算法( Online Adaboost, OAB)用于视觉跟踪.随后又将半监督学习( Semi-supervised Learning)

算法与 OAB 相结合,提出 Semi-Boost 跟踪算法<sup>[60]</sup>,极大地提高跟踪算法的鲁棒性和适应性.

总体上,基于检测的跟踪算法和机器学习的分类器算法联系紧密.未来基于判决学习的目标跟踪算法将是在线视觉跟踪研究的一个热点方向.机器学习算法的发展仍会成为该方向发展的重要推动力.

除此之外,比较活跃的在线视觉跟踪算法研究方向还有:基于主动轮廓 (Active Contour) 或水平集 (Level Set) 的目标跟踪<sup>[61]</sup>、基于图割 (Graph Cuts) 的目标跟踪<sup>[62]</sup>、基于兴趣点 (Interest Point) 的目标跟踪<sup>[63]</sup>、基于度量学习 (Metric Learning) 的目标跟踪<sup>[64]</sup>等.

3 相关滤波算法

目标跟踪运用各种方法根据已知观测图像估计当前图像中目标物体的位置.使用目标图像训练的滤波器对图像进行滤波处理,在响应图像中寻找最大值位置,即图像中对应的目标位置.在这种情况下可把目标跟踪的过程近似地看成对搜索区域图像进行相关滤波 (Correlation Filter) 的过程,寻找目标位置也就是寻找滤波器响应图像的最大值位置.下面从基础算法以及对基础算法改进的角度概述相关滤波算法,并给出简要分析.

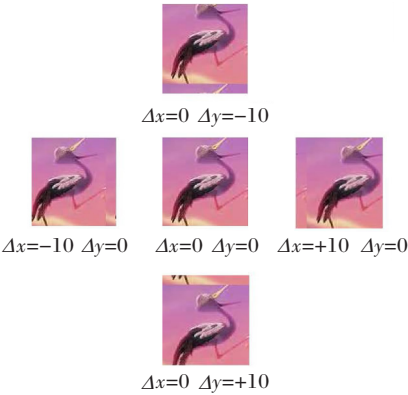
3.1 基础算法

Bolme 等<sup>[65]</sup>提出相关滤波 (Correlation Filter) 用于目标跟踪.基于输出结果的最小均方误差训练滤波器,作用于搜索区域,可以得到响应图像,值越大说明该位置处的图像与初始化目标相关性越大,提高滤波器的准确度和稳定性.与此同时,算法采用灰度特征,速度超过 600 帧/s.这种相关滤波跟踪算法即是常说的平方误差最小滤波器 (Minimum Output Sum of Squared Error, MOSSE)<sup>[65]</sup>,可以算得上最早关于相关滤波的算法.

随着基于检测跟踪算法 (Tracking by Detection) 的广泛应用,相关滤波得到进一步发展.在相关滤波跟踪算法中,关键部分是训练一个具有判别性的相关滤波器,用于区分前景和背景. Henriques 等<sup>[66]</sup>针对相关滤波中样本数量不足对分类器结果的影响,创新循环密集采样方法.利用中心图像块循环移位近似窗口移位,如图 5 所示.

图 5(a) 中间为中心图像块,位于它左右的图像块是中心图像块沿  $x$  轴分别左右移动 10 个像素得到的图像块,位于它上下的图像块是中心图像块沿

$y$  轴分别移动 10 个像素得到的图像块. (b) 是中心矩阵循环移位后对应的图像相应地上下左右偏移,循环采样粒子可组成循环矩阵.在训练滤波器时,利用循环矩阵时域和频域的特殊性质,将优化求解中的复杂矩阵求逆转变为简单矩阵点除.在进行目标跟踪时,将滤波器与粒子的相关性操作变成频域内的点乘操作,大幅减少计算量,显著提高训练速度.算法采用单通道灰度特征,可以实现简单跟踪,但当遇到目标遮挡、背景复杂、目标旋转及尺度变化等情况时,容易发生跟踪漂移.



(a) 中心图像及移位后图像  
(a) Central image and images after shift

22	23	24	21	21	22	23	24	24	21	22	23
32	33	34	31	31	32	33	34	34	31	32	33
42	43	44	41	41	42	43	44	44	41	42	43
12	13	14	11	11	12	13	14	14	11	12	13
12	13	14	11	11	12	13	14	14	11	12	13
22	23	24	21	21	22	23	24	24	21	22	23
32	33	34	31	31	32	33	34	34	31	32	33
42	43	44	41	41	42	43	44	44	41	42	43
42	43	44	41	41	42	43	44	44	41	42	43
12	13	14	11	11	12	13	14	14	11	12	13
22	23	24	21	21	22	23	24	24	21	22	23
32	33	34	31	31	32	33	34	34	31	32	33

(b) 二维循环矩阵示例  
(b) Examples of two-dimensional circular matrices

图 5 循环样本及循环矩阵示例

Fig. 5 Examples of circulate samples and circulate matrices

3.2 特征改进

为了提升算法性能, Henriques 等优化 CSK<sup>[66]</sup>的特征部分,提出基于颜色空间的算法 (Color Name, CN)<sup>[67]</sup>.对 RGB 颜色空间进行扩展,提出 CN 空间,该空间通道数为 11 (包括黑、蓝、棕、灰、绿、橙、粉、紫、红、白和黄).将 CN 空间内提取的图像特征代入原训练公式,对每个通道进行快速傅里



叶变换,并进行特征加和.虽然对算法精度有所帮助,但多通道特征使运算量加大,影响算法速度.为了降低运算量,作者提出使用 PCA 降维的方法提取更有效的特征.除了对颜色空间进行改进之外, Henriques 等<sup>[68]</sup>在相关滤波中引入核空间,同时提出将多通道特征融入相关滤波的方法.采用方向梯度直方图(Histogram of Oriented Gradient, HOG)特征,通过核函数将原来线性空间中的脊回归映射到非线性空间,在非线性空间中求解一个对偶问题和某些常见的约束,同样可以利用循环矩阵傅里叶空间对角化简化计算.

HOG 特征对运动模糊、光照变化及颜色变化等很鲁棒,但对形变的鲁棒性较差.颜色特征对形变鲁棒性较好,但对光照变化不够鲁棒.如果将 2 种特征进行有效结合,将会提升跟踪算法的精度. Bertinetto 等<sup>[69]</sup>提出 STAPLE(Complementary Learners for Real-Time Tracking),从特征融合方面入手,改进跟踪算法.利用 HOG 特征训练相关滤波器,得到跟踪得分,并融合使用颜色直方图得到的统计得分进行,得到最后的响应图像并估计目标位置.两种特征的融合提高跟踪算法的准确度,但也使得计算略微复杂.

### 3.3 多核算法

除了采用多种颜色特征之外,还有一些算法将多核方法(Multi-kernel)应用到相关滤波中,取得不错的效果.核函数可以将原本线性空间内不可分的样本映射到高维空间,增加样本的可分性.单核学习中,核函数的选择通常根据经验人为设定,这样的核函数不一定最优,而多核学习旨在通过多个核函数按权重融合或自适应的选择解决人工设定带来的不合理性. Tang 等<sup>[70]</sup>提出多核相关滤波器(Multi Kernelized Correlation Filter, MKCF),在核化相关滤波器(Kernelized Correlation Filter, KCF)<sup>[68]</sup>中引入多核,并提出优化求解方法. Zhang 等<sup>[71]</sup>融合多特征与多核,将多个滤波器进行有效整合,并通过最后的跟踪结果判断特征与核的权重,使算法具有自适应选择特征及核的能力.继而, Choi 等<sup>[72]</sup>对其进行改进,提出注意力机制相关滤波网络(Attentional Correlation Filter Network, ACFN),使用注意力网络(Attention Network)选择合适的相关滤波器,使算法能根据跟踪情况自适应调整选用的特征与核函数.同特征改进相同,单一特征或单一的核函数很难适应所有的跟踪情况,因此多特征及多核之间的联合作用通常起到较好的效果.

### 3.4 尺度估计

尺度变化是目标跟踪中一个较常见的问题,但

前述算法中都无法进行尺度估计.当目标尺度缩小时,会造成选取图像块中包含大量背景信息,当目标尺度扩大时,会造成选取图像块中只包含目标的局部信息,这两种情况都会引起跟踪漂移.为了解决尺度变化对跟踪的影响,自适应尺度变化的相关滤波跟踪器(Scale Adaptive Kernel Correlation Filter Tracker, SAMF)<sup>[73]</sup>和判断尺度空间跟踪器(Discriminative Scale Space Tracker, DSST)<sup>[74]</sup>在 KCF<sup>[68]</sup>的基础上引入尺度估计. SAMF 使用 7 个较粗的尺度,使用平移滤波器在多尺度图像块上进行检测,选取响应值最大处对应的平移位置和目标尺度. DSST 分别训练平移滤波器和尺度滤波器,使用 33 个较精细的尺度,先使用平移滤波器进行位置估计,然后在该位置处使用尺度滤波器进行尺度估计.以后的算法中经常使用这两种尺度估计的算法.

### 3.5 分块算法

此外,跟踪过程中的物体通常使用一个矩形框选取,由于跟踪的物体一般不为矩形,目标图像块不可避免地引入背景信息,导致跟踪不准确.为了解决这一问题,可以采用分割算法(Segmentation)或关键点算法(Key Point)表示目标.基于分割的物体表示虽然能较好地表示目标形状,但计算量过大,影响跟踪算法的速度.而基于关键点的目标表示难以获取跟踪目标的整体特征,也不是一个很好的方法.为了解决这一问题, Li 等<sup>[75]</sup>提出选取有效局部图像块表示物体结构,利用每个局部图像块的响应图像计算其置信度.如图 6 所示,这些响应图像按一定方法组合后可大致表示图中物体的形状.算法采用霍夫投票的方法融合多个响应图像,估计目标位置和尺度.

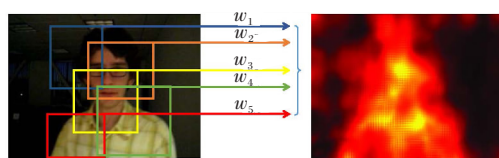


图6 RPT 局部图像块表示目标物体示意图

Fig. 6 Schematic diagram of representing target object via PRT local image block

Liu 等<sup>[76]</sup>提出使用分块的相关滤波算法增强跟踪器对遮挡的鲁棒性.将目标区域分为 5 部分,每部分分别训练相关滤波器,将得到的响应图像(Response Map)按权重加和,得到最终的响应图,由此估计目标状态.根据每个滤波器的权重判断跟踪好坏,决定是否进行模板更新. Fan 等<sup>[77]</sup>将整体与局部相结合进行目标跟踪,提出局部全局相结合的

相关滤波器 (Local-Global Correlation Filter, LGCF). 将整个目标区域平分成  $2 \times 2$  的图像块, 在相关滤波优化方程中加入整体和局部的优化方程, 使用交替方向乘子算法 (Alternating Direction Method of Multipliers, ADMM) 进行优化求解. Lukežić 等<sup>[78]</sup> 利用整体与局部相结合的思想, 提出可形变分块相关滤波器 (Deformable Parts Correlation Filter, DPT), 与 LGCF<sup>[77]</sup> 不同的是, 按照图像关键部分选取该算法的 4 个局部图像块, 不是平均分成, 在新一帧的估计中按照系统能量优化的原则重新寻找关键部分, 并与整体结合计算新位置.

### 3.6 样本标签改进

在相关滤波算法中, 通过样本循环移位构造密集采样, 这样样本对应的二分类标签一般采用高斯分布的标签. 当上一帧目标跟踪不准时, 由于当前帧的标签是以上一帧中心点为中心产生, 会导致错误的累积, 发生跟踪漂移现象. 其次, 这种标签帧间独立, 忽略跟踪过程中物体时间上的关联, 并不可靠. 为了克服上述问题, Bibi 等<sup>[79]</sup> 提出目标响应自适应相关滤波跟踪算法 (Target Response Adaptation for Correlation Filter Tracking, CF+AT), 将原来固定的高斯型标签变为自适应的样本标签, 在 KCF<sup>[68]</sup> 的优化方程中同时加入滤波器学习和样本标签先验的学习, 根据样本标签先验信息, 计算每帧的自适应标签, 并提出快速求解的方法. 这种自适应样本标签用于相关滤波算法中, 可以提升算法的抗干扰能力, 具有理论上的合理性和实践上的有效性. 相同的思想也出现在鲁棒相关滤波器 (Robust Correlation Filter, RCF)<sup>[80]</sup> 中, 但做法有所不同. RCF 提出现有的相关滤波优化方程中的  $L_2$  损失项对应的响应为各向同性, 较易受遮挡或光照等因素的影响, 通过引入稀疏性的损失项探索各向异性的响应对跟踪结果的影响, 算法也通过实验证明这种响应更准确鲁棒.

### 3.7 边界效应

边界效应也是影响滤波器性能提升的一个主要原因. 在训练阶段, 由于密集样本是经过中心图像块循环移位得到, 如图 5 所示, 只有中心样本准确, 其它的样本都会存在位移边界, 导致训练的分类器在物体快速移动时不能准确跟踪. 大部分算法的解决方案是在图像上加上余弦窗, 弱化图像边界对结果的影响, 这样只要保证移位后图像中心部分合理即可. 虽然增加合理样本的数量, 但仍不能保证所有训练样本的有效性. 另外, 加入余弦窗也会使跟踪器屏蔽背景信息, 只接受部分有效信息, 降低分类器的判别能力. 在目标检测阶段, 当物体处于搜索区域的边

界时, 余弦窗的加入使物体信息弱化甚至清除, 导致跟踪失败. 为了克服边界效应, Galoogahi 等<sup>[81]</sup> 采用较大尺寸的检测图像块和较小尺寸的滤波器, 提高真实样本的比例, 并采用 ADMM 迭代求解. Danelljan 等<sup>[82]</sup> 提出空间约束相关滤波器 (Spatially Regularized Correlation Filter, SRDCF), 采用更大的检测区域, 在滤波器系数上加入权重约束, 越靠近边缘权重越大, 越靠近中心权重越小, 使滤波器系数主要集中在中心区域. 由于滤波器在整个搜索区域内移动检测图像块的相关性, 因此在克服边界效应的同时不会忽略边缘物体的检测.

在 SRDCF 的基础上, Danelljan 等<sup>[83]</sup> 提出连续域卷积相关滤波算法 (Continuous Convolution Operators for Visual Tracking, CCOT). 将原来时域离散的位置估计转换为时域连续的位置估计, 使位置估计更准确. 此外, 算法还在相关滤波算法中融合分辨率不同的图像特征, 将多种传统特征与深度特征相结合, 提升算法精度. CCOT 中对待高维特征的方法是每个通道训练一个相关滤波器, 使算法存在大量的冗余, 也影响算法速度. 继而, Danelljan 等<sup>[84]</sup> 改进 CCOT, 从训练样本、滤波器系数和模板更新 3 个方面去除冗余, 提出高效卷积操作 (Efficient Convolution Operators, ECO). 算法指出, 相邻帧之间的样本具有很多冗余信息, 影响样本多样性, 限制分类器性能. 使用高斯混合模型对样本建模, 可以有效增强算法性能. 同时, 训练代表性滤波器, 并用线性组合表示其它滤波器, 每隔 6 帧更新一次, 大幅降低计算量. Lukežić 等<sup>[85]</sup> 提出可靠通道与空间的批判性相关滤波器 (Discriminative Correlation Filter with Channel and Spatial Reliability, CSR-DCF), 利用空域可靠性得到二值掩膜, 作用于滤波器上, 自适应地选择更易跟踪的目标区域, 减少边界效应. 通道间采用加权求和的方法进行融合. CF+CA<sup>[86]</sup> 指出用于相关滤波训练的负样本只通过正样本循环移位得到, 训练的分类器对背景的判别能力有限, 因此在训练时引入正样本周围采集的负样本, 提高跟踪准确度.

### 3.8 结合其它算法

一些算法结合循环矩阵与其它方法, 提高速度和精度. 支持向量机 (Support Vector Machine, SVM) 能准确区分目标和背景, 但粒子滤波 (Particle Filter) 的框架导致算法速度较慢. 相关滤波算法得益于循环矩阵的特殊性质, 能将运算转换到频域进行简化. Zuo 等<sup>[87]</sup> 结合 SVM 与循环矩阵, 提出支持相关滤波器 (Support Correlation Filter, SCF), 并给出优化方程在频域中的简化计算方法. Wang 等<sup>[88]</sup>



结合结构化 SVM 与相关滤波算法,提出 LMCFT (Large Margin Object Tracking with Circulant Feature Maps),分类器使用结构化 SVM,循环采样目标框作为训练样本.对响应图像进行多峰检测,并自适应进行模板更新,提高定位精度.此外,Zhang 等<sup>[89]</sup>结合粒子滤波与相关滤波算法,提出多任务相关粒子滤波器 (Multi-task Correlation Particle Filter, MCPF),利用相关滤波的响应图像引导采样粒子趋近于样本的真实分布,同时利用循环密集采样减少粒子个数,提升算法速度.随着深度学习的流行,有些算法也结合相关滤波与深度学习.

3.9 相关滤波算法总结

在现阶段,相关滤波算法凭借良好的精度与速度综合性能得到广泛应用,也发展出一系列基于相关滤波改进的目标跟踪方法,从特征、核、尺度、分块、样本标签、边界效应及结合其它算法等方面进行改进,使算法精度得到大幅度提升,但仍存在一些问题需要解决.

现存的算法大部分是取一定面积的搜索区域用于图像滤波操作.如果搜索区域过小,在目标快速移动时会检测不到目标.如果搜索区域过大,引入过多背景信息,当背景复杂时,容易造成跟踪漂移,大部分算法一般取目标大小的 2~3 倍作为搜索区域.这种限定区域的搜索方法具有一定的局限性,不能保证完全准确的跟踪.同时,用于训练分类器的负样本均为中心样本循环移位得到,负样本的局限性不能保证分类器的判别性能.

4 深度学习算法

近些年,深度学习方法 (Deep Learning) 借助其优秀的特征建模能力,在目标跟踪领域取得巨大成功.在基础测试数据库 (OTB, VOT 等) 上,准确度排名前几的算法都是基于深度学习的目标跟踪算法.采用粒子滤波的深度目标跟踪算法一般需要上百个粒子训练分类器或匹配器,每个粒子都要经过复杂的卷积层以提取特征,增加算法复杂度,限制算法速度的提升.为了同时保证算法精度与速度,很多算法结合深度学习方法与相关滤波算法,取得不错效果.

4.1 基于相关滤波的深度目标跟踪

相关滤波算法的核心思想是将目标模板与搜索区域内滑动窗口取得的图像块进行相关性匹配,响应最大位置处对应的图像块为目标图像块.为了加快相关性对比的速度,利用中心样本循环移位,近似滑动窗口取出的图像块,并利用循环样本构建二维

循环矩阵,将计算转换至频域加速.相比传统特征,深度神经网络提取的图像卷积特征具有良好的抗干扰能力,在大规模图像分类比赛中取得巨大的成功.但是,多层卷积计算使特征提取部分耗时较多,对多粒子图像块的算法不利.因此,相关滤波在深度目标跟踪中具备很好的应用前景.

Ma 等<sup>[90]</sup>提出基于多层卷积特征的跟踪器 (Hierarchical Convolutional Features for Visual Tracking, HCFT),结合深度特征与相关滤波算法,取得很好的效果.算法指出深度神经网络不同层的特征具有不同的特点,浅层特征包含更多的位置信息,但语义信息不明显.深层特征包含更多的语义信息,抗干扰能力较强,但位置信息弱化,因此不同层特征结合有助于效果的提升.如图 7 所示,使用 VGG16 分别提取 Conv3-4, Conv4-4 和 Conv5-4 层的输出特征,训练各自的相关滤波器.在目标检测时,将搜索区域的三层特征输入对应的相关滤波器,得到响应图像,按权重相加得到最终响应图像,通过最大值位置定位目标.实验表明,相比传统特征,深度特征具有良好的鲁棒性,相关滤波算法使得多层特征的利用变得可行.

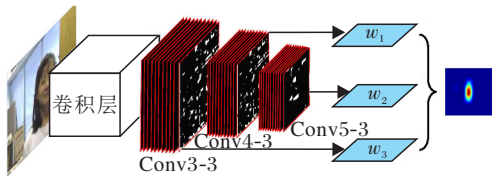


图 7 HCFT<sup>[90]</sup> 网络结构示意图

Fig. 7 Schematic diagram of HCFT algorithm<sup>[90]</sup> network architecture

与 HCFT<sup>[90]</sup> 类似,强化深度跟踪 (Hedged Deep Tracking, HDT)<sup>[91]</sup> 也采用多层深度特征与相关滤波相结合的方法.由原来的三层深度特征变为六层深度特征,将原来固定权重相加的方法变为自适应权重相加.Danelljan 等<sup>[82]</sup>在 SRDCF 的基础上引入深度特征,构成 DeepSRDCF (Deep Features for SRDCF),与文献[90]和文献[91]不同,DeepSRDCF 提出使用 VGG 网络的第一层卷积特征效果优于深层特征,同时浅层特征提取速度更快,有利于算法速度提升.CCOT<sup>[83]</sup>在特征融合上进行改进,创建时域连续的相关滤波器进行更精确的定位,使滤波器输入可以是不同分辨率的特征图,实现传统特征与深度特征的融合,高维特征训练多个滤波器,增加算法复杂度,算法速度较慢.为了加快算法速度,ECO<sup>[84]</sup>从滤波器系数、样本和模板更新 3 个方面进行去冗余操作.除了采用大规模图像分类网络提取的特征

之外, 基于深度运动特征的 SRDCF (Deep Motion Features for SRDCF, DMSRDCF)<sup>[92]</sup> 采用深度运动特征进行目标跟踪。

现今深度学习与相关滤波的结合不仅局限于使用深度网络提取特征, 一些算法结合特征提取与分类器, 使用神经网络模拟相关滤波的整个过程。在相关滤波中, 需要保存模板信息并提取搜索区域特征, 因此基于相关滤波思想的网络一般都采用孪生网络 (Siamese Network) 结构, 其中一条支路保存目标模板信息, 另一条支路用于搜索区域特征提取, 最后将两部分特征进行相关操作, 得到响应图像 (Response Map), 根据响应图像最大值位置判断目标状态。Bertinetto 等<sup>[93]</sup> 提出 Siamese-FC, 构造孪生网络结构, 分别提取目标特征和搜索区域特征进行相关操作, 使用大规模视觉图像识别比赛 (ImageNet Large Scale Visual Recognition Competition, ILSVRC) 中的视频对网络进行训练。使用模板支路信息引导跟踪, 取代网络的在线更新, 算法精度与速度都能得以保障。基于相关滤波的网络 (Correlation Filter Based Network, CFNet)<sup>[94]</sup> 和判别相关滤波网络 (Discriminant Correlation Filter Based Network, DCFNet)<sup>[95]</sup> 将滤波器系数转换为神经网络的一层, 推导前向与后向传播的公式, 实现网络的端到端训练, 算法速度可达每秒几十帧甚至上百帧。

神经网络功能多种多样, 除负责相关滤波跟踪算法中的特征提取和相关性操作之外, 还可以作为辅助网络进行特定区域显著性增强或自适应选择。Cui 等<sup>[96]</sup> 提出循环目标强化跟踪算法 (Recurrently Target-Attending Tracking, RTT), 使用 RNN 网络提取显著性图, 显著性图在目标部分具有较高的值, 将它加到相关滤波器系数上, 增强相关滤波器对于背景噪声信息的抗干扰能力, 也使得滤波器能更好地关注对结果有帮助的部分。注意力机制相关滤波网络 (Attentional Correlation Filter Network, ACFN)<sup>[72]</sup> 训练注意力机制的神经网络 (Attention Network), 对多个特征及滤波器核进行选择, 主体部分由不同特征和核函数组成的多个相关滤波器构成, 将多个滤波器的输出送到注意力机制的神经网络中, 判断结果可靠部分, 引导下一帧目标定位及相应的滤波器更新, 取得不错的效果。神经网络灵活多样, 相信在未来的相关滤波算法中能够得到更广泛的应用。

#### 4.2 其它深度目标跟踪

目前, 大部分深度学习算法都使用 CNN 网络结构, 小部分使用 RNN 或自己设计的网络结构。与传统算法类似, 也可分为判别式算法与生成式算法。判

别式算法基于粒子滤波或相关滤波框架 (上面已讨论基于相关滤波的深度目标跟踪算法, 这里不重复讨论), 通过训练分类器区分前景和背景。生成式算法运用生成模型描述目标的表现特征, 之后搜索候选目标最小化重构误差, 采用孪生网络结构。

Wang 等<sup>[97]</sup> 提出深度目标跟踪 (Deep Learning Tracking, DLT), 使用 4 个堆叠的栈式降噪自编码器 (Stacked Denoising Autoencoder, SDAE) 在大规模自然图像数据集上进行无监督训练, 获得物体表征能力, 然后在解码器后面加入 sigmoid 分类层用于区分目标和背景图像块。根据分类最高得分判断是否进行模型更新, 当得分高于一定阈值时进行模型更新。由于训练数据不足, 网络并没有获得很好的表征能力, 算法精度不高, 但它提出的“离线预训练+在线微调”的方法为深度学习在目标跟踪中的应用提供一个可行的方向, 之后很多算法都采用这种方法进行目标跟踪。

特征强化深度目标跟踪 (Transferring Rich Feature Hierarchies for Robust Visual Tracking, SO-D-LT)<sup>[98]</sup> 延续 DLT<sup>[97]</sup> “离线预训练+在线微调”的方法, 将原来的全连接网络 (Fully Connected Network) 变为卷积神经网络 (Convolutional Network), 在最后一层卷积后面连接不同尺度的池化层, 输出特征经过全连接层得到目标位置概率图, 用于目标定位。为了解决训练样本不足的问题, 在网络离线训练中使用 ImageNet 2014 检测数据集。模型更新采用长期与短期 CNN 相结合的方式, 使算法既能适应目标外观的变化, 又能具备一定的抗干扰能力。该算法成为大规模图像数据集训练 CNN 模型在目标跟踪领域的一次成功应用。

近几年, 随着大规模图像分类比赛的流行, 很多典型卷积网络被应用到图像处理领域, 如 AlexNet<sup>[99]</sup>, VGGNet<sup>[100]</sup>, GoogleNet<sup>[101]</sup> 和残差网络 (Residual Network, ResNet)<sup>[102]</sup> 等, 也出现很多基于这些网络的目标跟踪算法。Wang 等<sup>[103]</sup> 提出基于全卷积网络的跟踪算法 (Fully Convolutional Network Based Tracking, FCNT), 利用 VGG-16 网络, 提出深度神经网络不同层的特征具有不同的特点, 浅层特征含有较多位置信息, 深层特征含有更多语义信息, 而且深度特征存在大量冗余。因此, 算法针对 Conv4-3 和 Conv5-3 两层输出的特征图谱, 训练特征选择网络分别提取有效的特征, 降低特征维度。然后将选好的特征输送到各自的定位网络中, 得到热力图, 综合 2 个定位网络的热力图得到最终的热力图, 用于目标定位。算法利用不同层特征相互补充, 达到



有效抑制跟踪器漂移,同时对目标本身形变更鲁棒的效果.多特征补充可以弥补单一特征的不足,多分类器也能弥补单一分类器的不足.基于序列训练卷积网络的跟踪算法 (Sequentially Training Convolutional Networks for Visual Tracking, STCT)<sup>[104]</sup>在特征提取网络的后面连接多个卷积层训练弱分类器,综合弱分类器结果得到强分类器,提升算法精度.

图像分类中的图像分辨率较高,主要目的是判断图像中物体所属类别.目标跟踪中图像分辨率一般较低,只要目的是对图像中的目标进行定位,可能会遇到背景杂乱和目标太小等情况.虽然在大规模图像分类比赛中训练的网络在目标跟踪任务中的成绩不错,但仍不能完全适用于目标跟踪任务.目标跟踪中可能出现图像分类任务中未包含的物体种类,随着网络的深入,目标位置信息的削弱不利于目标定位.为了扩展 CNN 在目标跟踪领域的的能力,需要大量的训练数据,但这在目标跟踪中很难做到. Nam 等<sup>[105]</sup>提出多域网络 (Multi-domain Network, MDNet),解决上述问题.如图 8 所示,算法采用在图像分类任务中预训练的模型 VGG-M 作为网络初始化模型,后接多个全连接层用于分类器.训练时,每个跟踪视频对应一个全连接层,用于学习普遍的特征表示跟踪.跟踪时,去掉训练时的全连接层,使用第一帧样本初始化一个全连接层,新的全连接层在跟踪过程中继续微调,适应新的目标变化.这种方法使特征更适合于目标跟踪,效果大幅提升.由此可看出,通过视频训练的网络更适合目标跟踪这一任务.后来,结构关注网络 (Structure-Aware Network, SANet)<sup>[106]</sup>针对 MDNet<sup>[105]</sup>进行改进,将 RNN 网络得到的目标显著性图融入到原网络中,使网络能更关注对结果有用的部分.树状卷积神经网络 (CNNs in a Tree Structure, TCNN)<sup>[107]</sup>采用未经视频训练的 VGG-M 网络提取特征,改进后面全连接层结构为树状结构,采用多个可信度高的全连接层共同决定最终结果.

基于粒子滤波框架的深度学习算法由于需要提取多个图像块的深度特征,速度难以达到实时性要求.为了提升算法速度,Held 等<sup>[108]</sup>提出利用回归网络的通用目标跟踪框架 (Generic Object Tracking Using Regression Networks, GOTURN).如图 9 所示,将上一帧的目标和当前帧的搜索区域同时经过 CNN 的卷积层,级联特征输出通过全连接层,回归当前帧目标的位置,由于没有模板更新,算法速度可超过 100 帧/s. Redmon 等<sup>[109]</sup>提出 YOLO,构建孪生网络输出目标概率图,得到目标状态.这种基于模板

匹配的目标跟踪算法使用一条支路保存模板信息,为目标跟踪提供先验信息,取代全连接层在线更新,算法速度一般较快.

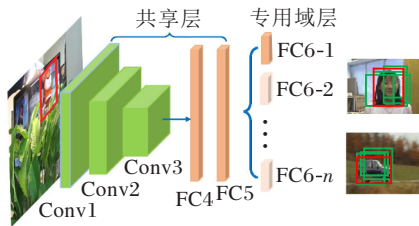


图 8 MDNet<sup>[105]</sup>网络结构示意图

Fig. 8 Schematic diagram of MDNet algorithm<sup>[105]</sup> network architecture

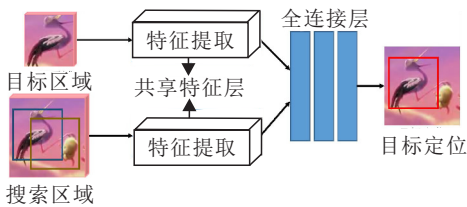


图 9 GOTURN<sup>[108]</sup>网络结构示意图

Fig. 9 Schematic diagram of GOTURN algorithm<sup>[108]</sup> network architecture

与分类图像不同,视频图像的帧与帧之间具有很强的时间及空间连续性,这种视频信息对于跟踪结果的准确性及算法速度的提升都有较大帮助.虽然 CNN 网络的特征提取能力很好,但没有办法建立视频帧之间的时间连续性.循环 YOLO (Recurrent YOLO, ROLO)<sup>[110]</sup>采用 YOLO 进行候选样本检测,将检测图像块输送到长短期记忆模块 (Long Short-Term Memory, LSTM) 输出目标框信息,利用 LSTM 保存帧间连续性信息.循环神经网络在文本和语义等方面得到广泛应用,能建模前后文之间的关系,在目标跟踪领域具有广阔的发展前景.

近几年,强化学习 (Reinforcement Learning) 受到广泛关注,作为一种介于半监督与无监督之间的训练方法,适用于目标跟踪这种缺乏训练样本的领域.动作决策网络 (Action-Decision Network, AD-Net)<sup>[111]</sup>是强化学习在目标跟踪领域的一个成功应用.通过强化学习得到一个智能体 (Agent),预测目标框的移动方位及尺度变化,在当前帧中,以前一帧的目标位置为初始点,经过多次方位估计、位移和尺度变化,得到最后的目标位置.与此不同,Choi 等<sup>[112]</sup>提出强化决策跟踪算法 (Reinforcement Decision Making Tracker, RDMT),利用策略学习 (Policy Gradient) 得到智能体用于模板选择.无监督或弱监



督学习是目标跟踪领域的一个新兴方向,具有巨大的潜力。

### 4.3 深度算法总结

综合现存深度目标跟踪算法的结果,判别式模型的效果优于生成式模型。在判别式模型中,训练分类器区分目标物体和背景;而在生成式模型中,训练匹配器用于模板匹配。由于目标跟踪任务中训练数据的缺乏,很难训练一个代表性的网络模型,因此很多跟踪算法采用已有的网络模型作为算法预训练模型,这些网络在大规模图像分类数据集上训练,更适合进行分类任务,所以在判别式网络模型中表现更优。在目标跟踪领域,分类任务只是单纯的二分类,当样本分值达到阈值后就可作为正样本,而在回归任务中,只有当输出结果与标签完全吻合时,输出损失才为 0,因此相比回归任务,区分前景与背景对于网络来说是一个更简单的任务。

特征好坏是影响跟踪器效果的一个重要因素。目标跟踪任务中的视频图像不同于图像分类和目标检测,像素值较低。随着网络深度的增加,会对图像信息造成较大的损失。很多优秀的目标跟踪算法采用小型神经网络(如 VGG-M)进行特征提取,可以有效防止信息损失。除此之外,这种小型网络提取特征时间较短,有助于算法速度的提升。浅层特征可适用于大部分跟踪情况,但当目标发生较严重形变或外界环境变化较复杂时,抗干扰能力不够。将浅层特征与深层特征进行结合往往对算法精度具有较大帮助,但同时深度特征的提取也会影响算法速度。根据不同层特征特点,自适应选择特征是一种同时保证算法精度和速度的好方法。

## 5 结束语

发展至今,目标跟踪算法已取得很大的进步,特别是深度神经网络的引入,使算法精度得到大幅提升,但仍存在一些需要解决的问题。

速度是评价在线目标跟踪算法的一个重要指标。由于深度神经网络复杂的计算及模型更新时繁琐的系数,现存大部分深度目标跟踪算法速度都较慢。目标跟踪中大部分为低像素图像,不适合使用层数太多的卷积网络,是因为信息损失严重,也影响计算速度。很多深度目标跟踪算法采用小型神经网络(如 VGG-M)提取特征。另外,跟踪中只给定第一帧目标位置,缺少跟踪物体先验信息,这就要求模型实时更新以确保跟踪精度,而这在深度目标跟踪算法中非常耗时。一些算法采用孪生网络结构保存先验

信息,代替模型在线更新,算法速度得以提高。深度特征的高维度也会影响跟踪算法的速度,如果能提出有效的特征压缩方法,不管对算法速度还是精度都会有所帮助。只有高速且有效的算法才具有实际的应用价值。

影响深度神经网络效果的两个主要因素是网络结构和训练数据。现存大部分深度目标跟踪算法均采用 CNN 结构,虽然 CNN 结构的特征提取能力很好,但是难以建模视频帧中的时间连续性信息。一些算法采用 RNN 结构构建目标跟踪模型,但效果并不是很突出,仍需探索发展。还有一些新型的网络架构(如 ResNet 和 DenseNet 等)在图像分类领域取得很好的效果,这些网络架构是否能在目标跟踪领域成功应用也是令人期待。另外,限制深度目标跟踪算法发展的一个主要原因是训练样本的缺乏,没有如同图像分类任务那样的大规模训练样本,在线跟踪时只能给定第一帧的信息,难以训练一个适合当前跟踪物体的网络模型。近几年,无监督或弱监督方法受到人们的广泛关注。也有一些算法开始尝试将强化学习应用到目标跟踪领域。对抗网络可以生成迷惑机器的负样本,增强分类器判别能力。这些无监督和弱监督的方法可有效解决目标跟踪领域样本不足的问题。

除此之外,将视觉注意力机制引入目标跟踪的方向也很有前途。在原来的跟踪模型中加入特定网络生成目标区域显著性图,弱化背景突出前景。也可将注意力机制融入网络内部结构,引导跟踪器关注对结果有帮助的部分。现在已有一些方法将相关滤波当作特定层嵌入到网络实现端到端训练,但只是简单的核化相关滤波器,如何能将传统算法中的一些方法嵌入到深度神经网络中也值得研究。

## 参 考 文 献

- [1] COLLINS R T, BIERNACKI C, CELEUX G, *et al.* Introduction to the Special Section on Video Surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 745–746.
- [2] HARITAOGLU I, HARWOOD D, DAVIS L S. W4: Real-Time Surveillance of People and Their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 809–830.
- [3] SIEBEL N T, MAYBANK S J. The Advisor Visual Surveillance System[C/OL]. [2017–10–25]. <https://www.informatik.uni-kiel.de//inf/Sommer/doc/Publications/nts/SiebelMaybank2004-acv.pdf>.
- [4] 孔晓东. 智能视频监控技术研究. 博士学位论文. 上海: 上海交通大学, 2008.

- (KONG X D. Research on Intelligent Video Surveillance Technology. Ph. D. Dissertation. Shanghai, China; Shanghai Jiao Tong University, 2008. )
- [5] 贾春华. 智能监控中的行人检测与运动分析研究. 硕士学位论文. 大连:大连理工大学, 2008.  
(JIA C H. Research on Real-Time Human Detection and Motion Analysis of Intelligent Surveillance. Master Dissertation. Dalian, China; Dalian University of Technology, 2008. )
- [6] 焦波. 面向智能视频监控的运动目标检测与跟踪方法研究. 博士学位论文. 长沙:国防科学技术大学, 2009.  
(JIAO B. Research on Moving Object Detection and Tracking Methods for Intelligent Video Surveillance. Ph. D. Dissertation. Changsha, China; National University of Defense Technology, 2009. )
- [7] 赵久奋,王明海. 红外探测阵列对固体导弹尾焰跟踪定位的研究. 固体火箭技术, 2000, 23(4): 64-68.  
(ZHAO J F, WANG M H. A Study on Missile Plume Tracking and Localizing by Means of Forward Looking Infra-Red (FLIR). Journal of Solid Rocket Technology, 2000, 23(4): 64-68. )
- [8] 卢莉萍. 目标跟踪算法与检测处理技术研究. 博士学位论文. 南京:南京理工大学, 2012.  
(LU L P. Study on Targets Tracking Algorithm and Detection Processing Technology. Ph. D. Dissertation. Nanjing, China; Nanjing University of Science and Technology, 2012. )
- [9] LU H C, FANG G L, WANG C, *et al.* A Novel Method for Gaze Tracking by Local Pattern Model and Support Vector Regress. Signal Processing, 2010, 90(4): 1290-1299.
- [10] PRENDINGER H, DESCAMPS S, ISHIZUKA M. Scripting Affective Communication with Life-Like Characters in Web-Based Interaction Systems. Applied Artificial Intelligence, 2002, 16(7/8): 519-553.
- [11] LU H C, HUANG Y J, CHEN Y W. Automatic Facial Expression Recognition Based on Pixel-Pattern-Based Texture Feature. International Journal of Image Systems and Technology, 2010, 20(3): 253-260.
- [12] MITRA S, ACHARYA T. Gesture Recognition: A Survey. IEEE Transactions on Systems, Man, and Cybernetics (Applications and Reviews), 2007, 37(3): 311-324.
- [13] EROL A, BEBIS G, NICOLESCU M, *et al.* Vision-Based Hand Pose Estimation: A Review. Computer Vision and Image Understanding, 2007, 108(1/2): 52-73.
- [14] 王亮,胡卫明,谭铁牛. 人运动的视觉分析综述. 计算机学报, 2002, 25(3): 225-237.  
(WANG L, HU W M, TAN T N. A Survey of Visual Analysis of Human Motion. Chinese Journal of Computer, 2002, 25(3): 225-237. )
- [15] GANAPATHI V, PLAGEMANN C, KOLLER D, *et al.* Real Time Motion Capture Using a Single Time-of-Flight Camera // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA; IEEE, 2010: 755-762.
- [16] KASTRINAKI V, ZERVAKIS M E, KALAITZAKIS K. A Survey of Video Processing Techniques for Traffic Applications. Image and Vision Computing, 2003, 21(4): 359-381.
- [17] 崔雨勇. 智能交通监控中运动目标检测与跟踪算法研究. 博士学位论文. 武汉:华中科技大学, 2012.  
(CUI Y Y. A Dissertation Submitted on Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Science. Ph. D. Dissertation. Wuhan, China; Huazhong University of Science and Technology, 2012. )
- [18] RIOS-CABRERA R, TUYTELAARS T, VAN GOOL L. Efficient Multi-camera Vehicle Detection, Tracking, and Identification in a Tunnel Surveillance Application. Computer Vision and Image Understanding, 2012, 116(6): 742-753.
- [19] DESOUZA G N, KAK A C. Vision for Mobile Robot Navigation: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(2): 237-267.
- [20] BONIN-FONT F, ORTIZ A, OLIVER G. Visual Navigation for Mobile Robots: A Survey. Journal of Intelligent and Robotic Systems, 2008, 53(3): 263-296.
- [21] ANDRILUKA M, ROTH S, SCHIELE B. Pictorial Structures Revisited: People Detection and Articulated Pose Estimation // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA; IEEE, 2009: 1014-1021.
- [22] BOURDEV L D, MALIK J. Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations // Proc of the 12th IEEE International Conference on Computer Vision. Washington, USA; IEEE, 2009: 1365-1372.
- [23] AGARWAL S, FURUKAWA Y, SNAVELY N, *et al.* Building Rome in a Day. Communications of the ACM, 2011, 54(10): 105-112.
- [24] XIAO J X, FURUKAWA Y. Reconstructing the World's Museums. International Journal of Computer Vision, 2014, 110(3): 243-258.
- [25] 乐宁. 基于视觉的直升机飞行模拟及跟踪系统研究. 硕士学位论文. 南京:南京理工大学, 2010.  
(LE N. Vision-Based Helicopter Simulation and Tracking System Research. Master Dissertation. Nanjing, China; Nanjing University of Science and Technology, 2010. )
- [26] STEFANSIC J D, BASS W A, HARTMANN S L, *et al.* Design and Implementation of a PC-Based Image-Guided Surgical System. Computer Methods and Programs in Biomedicine, 2002, 69(3): 211-224.
- [27] REVELL J, MIRMEHDI M, MCNALLY D. Computer Vision Elastography: Speckle Adaptive Motion Estimation for Elastography Using Ultrasound Sequences. IEEE Transactions on Medical Imaging, 2005, 24(6): 755-766.
- [28] ZHAO Q P. A Survey on Virtual Reality. Science in China (Information Sciences), 2009, 52(3): 348-400.
- [29] BERRYMAN D R. Augmented Reality: A Review. Medical Reference Services Quarterly, 2012, 31(2): 212-218.
- [30] ROSS D A, LIM J, LIN R S, *et al.* Incremental Learning for Robust Visual Tracking. International Journal of Computer Vision, 2008, 77(1/2/3): 125-141.
- [31] COMANICIU D, RAMESH V, MEER P. Real-Time Tracking of Non-rigid Objects Using Mean Shift // Proc of the IEEE Conf-

- erence on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2000, II; DOI:10.1109/CVPR.2000.854761.
- [32] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2005: 886–893.
- [33] VIOLA P, JONES M J. Robust Real-Time Face Detection. International Journal of Computer Vision, 2004, 57(2): 137–154.
- [34] TUYTELAARS T, MIKOLAJCZYK K. Local Invariant Feature Detectors: A Survey. Foundations and Trends in Computer Graphics and Vision, 2007, 3(3): 177–280.
- [35] WANG S, LU H, YANG F, *et al.* Super Pixel Tracking // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2011: 1323–1330.
- [36] COMANICIU D, RAMESH V, MEER P. Kernel-Based Object Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564–575.
- [37] PÉREZ P, HUE C, VERMAAK J, *et al.* Color-Based Probabilistic Tracking // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2002: 661–675.
- [38] LI Y, AI H Z, YAMASHITA T, *et al.* Tracking in Low Frame Rate Video: A Cascade Particle Filter with Discriminative Observers of Different Life Spans. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(10): 1728–1740.
- [39] CHENG Y Z. Mean Shift, Mode Seeking, and Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790–799.
- [40] COLLINS R T. Mean-Shift Blob Tracking through Scale Space // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2003: 234–240.
- [41] HAGER G D, DEWAN M, STEWART C V. Multiple Kernel Tracking with SSD // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2004: 790–797.
- [42] YAO A B, LIN X G, WANG G J, *et al.* A Compact Association of Particle Filtering and Kernel Based Object Tracking. Pattern Recognition, 2012, 45(7): 2584–2597.
- [43] ADAM A, RIVLIN E, SHIMSHONI I. Robust Fragments-Based Tracking Using the Integral Histogram // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2006: 798–805.
- [44] WANG F L, YU S Y, YANG J. Robust and Efficient Fragments-Based Tracking Using Mean Shift. AEU – International Journal of Electronics and Communications, 2010, 64(7): 614–623.
- [45] JIA X, WANG D, LU H C. Fragment-Based Tracking Using On-line Multiple Kernel Learning // Proc of the IEEE International Conference on Image Processing. Washington, USA: IEEE, 2012: 393–396.
- [46] YANG F, LU H C, ZHANG W H, *et al.* Visual Tracking via Bag of Features. IET Image Processing, 2012, 6(2): 115–128.
- [47] ZHONG W, LU H C, YANG M H. Robust Object Tracking via Sparsity-Based Collaborative Model // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2012: 1838–1845.
- [48] LIU B Y, HUANG J Z, YANG L, *et al.* Robust Tracking Using Local Sparse Appearance Model and  $k$ -Selection // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2011: 1313–1320.
- [49] TURK M, PENTLAND A. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, 1991, 3(1): 71–86.
- [50] FRIEDMAN J H. Regularized Discriminant Analysis. Journal of the American Statistical Association, 1989, 84: 165–175.
- [51] HE X F, YAN S C, HU Y X, *et al.* Face Recognition Using Laplacianfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328–340.
- [52] YAN S C, XU D, ZHANG B Y, *et al.* Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40–51.
- [53] YU Q, DINH T B, MEDIONI G G. Online Tracking and Reacquisition Using Co-trained Generative and Discriminative Trackers // Proc of the 10th European Conference on Computer Vision. Berlin, Germany: Springer, 2008, II: 678–691.
- [54] DONOHO D L. Compressed Sensing. IEEE Transactions on Information Theory, 2006, 52(4): 1289–1306.
- [55] WRIGHT J, MA Y, MARAL J, *et al.* Sparse Representation for Computer Vision and Pattern Recognition. Proceedings of the IEEE, 2010, 98(6): 1031–1044.
- [56] MEI X, LING H B. Robust Visual Tracking Using  $l_1$  Minimization // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2009: 1436–1443.
- [57] YANG H X, SHAO L, ZHENG F, *et al.* Recent Advances and Trends in Visual Tracking: A Review. Neurocomputing, 2011, 74(18): 3823–3831.
- [58] LIU H P, SUN F C. Visual Tracking Using Sparsity Induced Similarity // Proc of the 20th IEEE International Conference on Pattern Recognition. Washington, USA: IEEE, 2010: 1702–1705.
- [59] GRABNER H, GRABNER M, BISCHOF H. Real-Time Tracking via On-line Boosting // Proc of the British Machine Vision Conference. Berlin, Germany: Springer, 2006: 47–56.
- [60] GRABNER H, LEISTNER C, BISCHOF H. Semi-supervised On-line Boosting for Robust Tracking // Proc of the 10th European Conference on Computer Vision. Berlin, Germany: Springer, 2008: 234–247.
- [61] SHI Y, KARL W C. Real-Time Tracking Using Level Sets // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2005: 34–41.
- [62] MALCOLM J G, RATHI Y, TANNENBAUM A. Multi-object Tracking through Clutter Using Graph Cuts // Proc of the 11th IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2007. DOI: 10.1109/ICCV.2007.4409178.
- [63] TA D N, CHEN W C, GELFAND N, *et al.* SURFTrack: Efficient Tracking and Continuous Object Recognition Using Local Feature



- Descriptors // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2009: 2937–2944.
- [64] WANG X Y, HUA G, HAN T X. Discriminative Tracking by Metric Learning // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2010: 200–214.
- [65] BOLME D S, BEVERIDE J R, DRAPER B A, *et al.* Visual Object Tracking Using Adaptive Correlation Filter // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2010: 2544–2550.
- [66] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.* Exploiting the Circulant Structure of Tracking-by-Detection with Kernels // Proc of the 12th European Conference on Computer Vision. Berlin, Germany: Springer, 2012: 702–715.
- [67] DANELLJAN M, KHAN F S, FELSBERG M, *et al.* Adaptive Color Attributes for Real-Time Visual Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2014: 1090–1097.
- [68] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.* High-Speed Tracking with Kernelized Correlation Filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(3): 583–596.
- [69] BERTINETTO L, VALMADRE J, GOLODETZ S, *et al.* Staple; Complementary Learners for Real-Time Tracking // Proc of the IEEE International Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 1401–1409.
- [70] TANG M, FENG J. Multi-kernel Correlation Filter for Visual Tracking // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2016: 3038–3046.
- [71] ZHANG K H, ZHANG L, YANG M H, *et al.* Fast Tracking via Spatio-Temporal Context Learning [J/OL]. [2017–10–21]. <https://arxiv.org/pdf/1311.1939v1.pdf>.
- [72] CHOI J, CHANG H J, YUN S, *et al.*, eds. Attentional Correlation Filter Network for Adaptive Visual Tracking // Proc of the IEEE International Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2017: 4807–4816.
- [73] LI Y, ZHU J K. A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 254–265.
- [74] DANELLJAN M, HÄGER G, KHAN F S. Accurate Scale Estimation for Robust Visual Tracking [C/OL]. [2017–10–21]. [http://www.cvl.isy.liu.se/research/objrec/visualtracking/scal\\_vistrack/ScaleTracking\\_BMVC14.pdf](http://www.cvl.isy.liu.se/research/objrec/visualtracking/scal_vistrack/ScaleTracking_BMVC14.pdf).
- [75] LI Y, ZHU J K, HOI S C H. Reliable Patch Trackers: Robust Visual Tracking by Exploiting Reliable Patches // Proc of the IEEE International Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2015: 353–361.
- [76] LIU T, WANG G, YANG Q X. Real-Time Part-Based Visual Tracking via Adaptive Correlation Filters // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2015: 4902–4912.
- [77] FAN H, XIANG J H. Robust Visual Tracking via Local-Global Correlation Filter [C/OL]. [2017–10–21]. <https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14351/14276>.
- [78] LUKEŽIĆ A, ČEHOVIN L. Deformable Parts Correlation Filters for Robust Visual Tracking [J/OL]. [2017–10–21]. <https://arxiv.org/pdf/1605.03720v1.pdf>.
- [79] BIBI A, MUELLER M, GHANEM B. Target Response Adaptation for Correlation Filter Tracking // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 419–433.
- [80] SUI Y, ZHANG Z M, WANG G H, *et al.* Real-Time Visual Tracking: Promoting the Robustness of Correlation Filter Learning // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 662–678.
- [81] GALOOGAHI H K, SIM T, LUCEY S. Correlation Filters with Limited Boundaries // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2015: 4630–4638.
- [82] DANELLJAN M, HÄGER G, KHAN F S, *et al.* Learning Spatially Regularized Correlation Filters for Visual Tracking // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2015: 4310–4318.
- [83] DANELLJAN M, ROBINSON A, KHAN F S, *et al.* Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 472–488.
- [84] DANELLJAN M, BHAT G, KHAN F S, *et al.* ECO: Efficient Convolution Operators for Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1611.09224.pdf>.
- [85] LUKEŽIĆ A, VOJIR T, ZAJC L Č, *et al.* Discriminative Correlation Filter with Channel and Spatial Reliability // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 4847–4856.
- [86] MUELLER M, SMITH N, GHANEM B. Context-Aware Correlation Filter Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2017: 1396–1404.
- [87] ZUO W M, WU X H, LIN L, *et al.* Learning Support Correlation Filters for Visual Tracking [J/OL]. [2017–10–21]. <https://arxiv.org/pdf/1601.06032v1.pdf>.
- [88] WANG M M, LIU Y, HUANG Z Y. Large Margin Object Tracking with Circulant Feature Maps [J/OL]. [2017–10–21]. <https://arxiv.org/pdf/1703.05020.pdf>.
- [89] ZHANG T Z, XU C S, YANG M H. Multi-task Correlation Particle Filter for Robust Visual Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2017: 4335–4343.
- [90] MA C, HUANG J B, YANG X K, *et al.* Hierarchical Convolutional Features for Visual Tracking // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2015: 3074–3082.
- [91] QI Y K, ZHANG S P, QIN L, *et al.* Hedged Deep Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Re-

- cognition. Washington, USA: IEEE, 2016: 4303–4311.
- [92] GLADH S, DANELLJAN M, KHAN F S, *et al.* Deep Motion Features for Visual Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1612.06615v1.pdf>.
- [93] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al.* Fully-Convolutional Siamese Networks for Object Tracking // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 850–865.
- [94] VALMADRE J, BERTINETTO L, HENRIQUES J F, *et al.* End-to-End Representation Learning for Correlation Filter Based Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1704.06036.pdf>.
- [95] WANG Q, GAO J, XING J L, *et al.* DCFNET: Discriminant Correlation Filters Network for Visual Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1704.04057v1.pdf>.
- [96] CUI Z, XIAO S T, FENG J S, *et al.* Recurrently Target-Attending Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 1449–1458.
- [97] WANG N Y, YEUNG D. Learning a Deep Compact Image Representation for Visual Tracking // BURGESS C J C, BOTTOU L, WELLING M, *et al.*, eds. Advances in Neural Information Processing Systems 26. Cambridge, USA: The MIT Press, 2013: 809–817.
- [98] WANG N Y, LI S Y, GUPTA A, *et al.* Transferring Rich Feature Hierarchies for Robust Visual Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1501.04587.pdf>.
- [99] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks // PEREIRA F, BURGESS C J C, BOTTOU L, *et al.*, eds. Advances in Neural Information Processing Systems 22. Cambridge, USA: The MIT Press, 2012: 1097–1105.
- [100] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1409.1556v6.pdf>.
- [101] SZEGEDY C, LIU W, JIA Y Q, *et al.* Going Deeper with Convolutions [C/OL]. [2017–10–21]. <https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf>.
- [102] HE K M, ZHANG X Y, RE S Q, *et al.* Deep Residual Learning for Image Recognition // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 770–778.
- [103] WANG L J, OUYANG W L, WANG X G, *et al.* Visual Tracking with Fully Convolutional Networks // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2015: 3119–3127.
- [104] WANG L J, OUYANG W L, WANG X G, *et al.* STCT: Sequentially Training Convolutional Networks for Visual Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 1373–1381.
- [105] NAM H, HAN B. Learning Multi-domain Convolutional Neural Networks for Visual Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 4293–4302.
- [106] FAN H, LING H B. SANET: Structure-Aware Network for Visual Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1611.06878v1.pdf>.
- [107] NAM H, BAEK M, HAN B. Modeling and Propagating CNNs in a Tree Structure for Visual Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1608.07242.pdf>.
- [108] HELD D, THRUN S, SAVARESE S. Learning to Track at 100 FPS with Deep Regression Networks // Proc of the IEEE Conference on Computer Vision. Washington, USA: IEEE, 2016: 749–765.
- [109] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You Only Look Once: Unified, Real-Time Object Detection // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 779–788.
- [110] NING G H, ZHANG Z, HUANG C, *et al.* Spatially Supervised Recurrent Convolutional Neural Networks for Visual Object Tracking [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1607.05781.pdf>.
- [111] YUN S, CHOI J, YOO Y, *et al.* Action-Decision Networks for Visual Tracking with Deep Reinforcement Learning // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2017: 2711–2720.
- [112] CHOI J, KWON J, LEE K M. Visual Tracking by Reinforced Decision Making [C/OL]. [2017–10–21]. <https://arxiv.org/pdf/1702.06291.pdf>.

## 作者简介



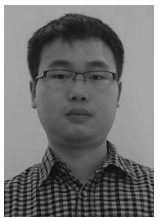
卢湖川, 博士, 教授, 主要研究方向为行人再识别、图像显著性检测、在线目标跟踪。E-mail: lhchuan@dlut.edu.cn.

(**LU Huchuan**, Ph. D., professor. His research interests include person re-identification, saliency detection and online visual tracking.)



李佩霞, 硕士研究生, 主要研究方向为视觉跟踪、深度学习。E-mail: pxli@mail.dlut.edu.cn.

(**LI Peixia**, master student. Her research interests include visual tracking and deep learning.)



王栋, 博士, 副教授, 主要研究方向为人脸识别、交互式图像分割、目标跟踪。E-mail: wdice@dlut.edu.cn.

(**WANG Dong**, Ph. D., associate professor. His research interests include face recognition, interactive image segmentation and object tracking.)