计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201100186

视觉目标跟踪十年研究进展

张开华　樊佳庆　刘青山
南京信息工程大学江苏省大数据分析技术重点实验室　南京 210044

摘　要　视觉目标跟踪指在一个视频序列中,给定第一帧目标区域,在后续帧中自动匹配到该目标区域的任务。通常来说,由于场景遮挡、光照变化、物体本身形变等复杂因素,目标与场景的表观会发生剧烈的变化,这使得跟踪任务本身面临极大的挑战。在过去的十年中,随着深度学习在计算机视觉领域的广泛应用,目标跟踪领域也迅速发展,研究人员提出了一系列优秀算法。鉴于该领域处于快速发展的阶段,文中对视觉目标跟踪研究进行了综述,内容主要包括跟踪的基本框架改进、目标表示改进、空间上下文改进、时序上下文改进、数据集和评价指标改进等;另外,还综合分析了这些改进方法各自的优缺点,并提出了可能的未来的研究趋势。

关键词:视觉目标跟踪;深度学习;计算机视觉
中图法分类号　TP391

Advances on Visual Object Tracking in Past Decade

ZHANG Kai-hua,FAN Jia-qing and LIU Qing-shan
Jiangsu Key Laboratory of Big Data Analysis Technology,Nanjing University of Information Science and Technology,Nanjing 210044,China

Abstract　Visual object tracking is a task in which the target region of the first frame in a video sequence is given,and then the target area is automatically matched in subsequent frames. Generally speaking,due to the complex factors such as scene occlusion,illumination change and object deformation,the appearance of the target and scene will change dramatically,which makes the tracking task itself is extremely challenging. In the past decade,with the extensive application of deep learning in the field of computer vision,the field of target tracking has also developed rapidly,resulting in a series of excellent algorithms. In view of this rapid development stage,this paper aims to provide a comprehensive review of visual object tracking research,mainly including the following aspects:the improvement of the basic framework of tracking,the improvement of target representation,the improvement of spatial context,the improvement of temporal context,the improvement of data sets and evaluation indicators. This paper also analyzes the advantages and disadvantages of these methods,and puts forward the possible future research trends.

Keywords　Visual object tracking,Deep learning,Computer vision

1　引言

目标跟踪是计算机视觉领域的一项经典研究课题,目的是在给定第一帧初始目标边界框的情况下,在后续视频序列中准确定位目标(见图1)。随着高性能移动设备与高配置摄像机的爆炸式增长,以及新一代5G网络的逐步应用,人们对自动视频分析的需求日益增长。自动视频分析中有3个关键步骤:自动检测感兴趣的运动物体、逐帧跟踪这些物体、通过分析物体的轨迹来进行行为识别。目标跟踪作为其中的一项重要技术,引起了相关学者的极大关注[1-2]。然而,视觉目标跟踪是一项极具挑战性的任务,因为有一系列不同的问题需要在单个跟踪算法中解决。例如,跟踪算法能很好地处理光照变化,但是难以应对因相机角度变化而带来的物体表观的变化;跟踪算法擅长准确预测物体运动,但是难以跟踪快速弹跳的物体;跟踪算法能对外观做出详细假设,但是不能处理有关节的物体。

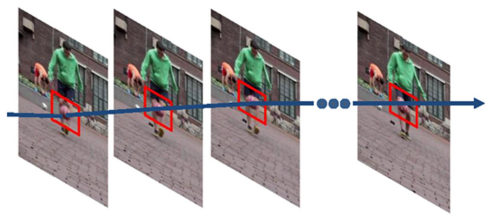

图1　在一个视频序列中的物体跟踪
Fig. 1　Object tracking in a video sequence

为了解决上述跟踪领域的一系列挑战性问题,近十年来,

到稿日期:2020-11-26　返修日期:2021-01-02
基金项目:国家新一代人工智能重大项目(2018AAA0100400);国家自然科学基金(61872189);江苏省333工程人才项目(BRA2020291)
This work was supported by the National Major Project of China for New Generation of AI(2018AAA0100400),National Natural Science Foundation of China(61872189) and 333 High-level Talents Cultivation Project of Jiangsu Province(BRA2020291).
通信作者:张开华(zhkhua@gmail.com)

目标跟踪领域涌现出了大批经典算法[3-21]，具体如图 2 所示。本文分 4 个阶段对目标跟踪的发展进行综述，即早期的目标跟踪探索阶段、稀疏表示阶段、相关滤波阶段和孪生网络阶段，主要介绍的跟踪算法包括 Histogram[3]，Ensemble[4]，IVT[5]，MIL[6]，L1 Tracker[7]，TLD[8]，MOSSE[9]，Struck[10]，ASLA[11]，CT[12]，CSK（KCF）[13]，CN[14]，STC[15]，CF2[16]，ECO[17]，SiamFC[18]，SiamRPN[19]，ATOM[20]，SiamRCNN[21]等。

本文详细梳理了最近几年目标跟踪领域的相关工作，并将其分为了五大类：数据集和评价标准的改进、目标跟踪基本框架改进、目标表示的改进、空间上下文方面的改进和时序上下文方面的改进。对上述 5 类工作分别进行介绍和分析之后得出本文的结论，并提出未来目标跟踪领域可能的发展趋势。

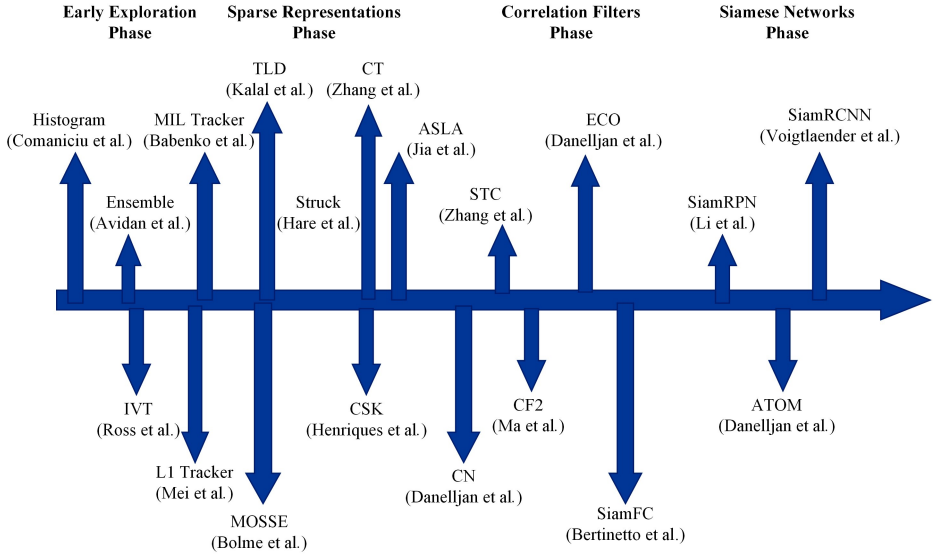

图 2 目标跟踪发展阶段

Fig. 2 Development phases of object tracking

2 数据集和评价标准的改进

考虑到目标跟踪环境的复杂性和跟踪方法的多样性，用于评估算法性能的视频序列也需要尽可能多地具备各种复杂的场景属性，如图 3 所示。但是，在早期的工作中，用于测试跟踪算法性能的视频数量通常是极其有限的，甚至只有 5～10 个[2,22]。考虑到视觉目标跟踪的重要性，如此少的测试视频难以有效评测跟踪算法的性能。另一方面，在几乎所有的视频分析任务中，跟踪都会起到一定的作用，换言之，目标追踪已经发展到了令人印象深刻的地步，甚至能在尘土弥漫的环境中追踪到摩托车或汽车[8,23]。因此，急需大规模跟踪评测数据集。

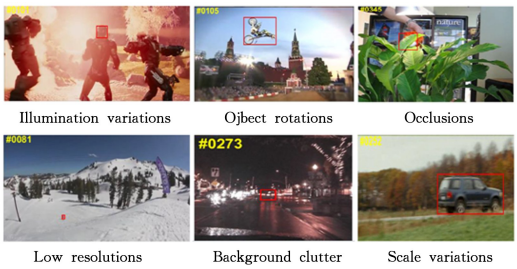

图 3 OTB 数据集中的复杂场景

Fig. 3 Complex scenarios in OTB dataset

为此，大量大规模的目标跟踪数据集[24-30]被提出，用于评估跟踪算法在各种复杂环境中的性能。下文将介绍一些代表性的数据集。

2.1 目标跟踪基准测试集（Object Tracking Benchmark）

评估跟踪算法的一个常见问题是，结果报告仅基于几个具有不同初始条件或参数的序列。因此，这些视频序列并没有公平和全面地评估这些算法的整体性能及优缺点。为了解决这一问题，Wu 等[6]构建了首个用于目标跟踪性能测试的代码库 Object Tracking Benchmark（OTB），其中包括大多数公开可用的跟踪算法和带有 50 个视频 ground-truth 注释的测试数据集，以便评估任务。数据集中的每个序列都标注了常见的影响跟踪性能的不同属性，如遮挡、快速运动和光照变化。

此外，OTB 基准集更是提出从 ground-truth 目标位置对初始状态进行时空扰动。虽然初始化的鲁棒性是该领域的一个众所周知的问题，但在已发表文献中很少涉及。这是解决和分析目标跟踪初始化问题的第一个综合性工作。OTB 采用基于定位误差度量的精确度图和基于重叠度量的成功率图，分别分析了各算法的性能。

2.2 视觉目标跟踪挑战赛数据集（The Visual Object Tracking VOT Challenge）

目标跟踪数据集的性能度量方式过于复杂、繁多，从中心误差、区域重叠、跟踪长度和失败率到更复杂的度量指标，以及将多个度量指标组合成一个单一的指标。为了简化度量方法，Kristan 等[25]提出了一个更好的策略，即应用一些不太相关的度量，通过排名将它们结合起来，并且首次组织了视觉目标跟踪（VOT2013）挑战赛，其目的是提供一个超越当前技术水平的评估平台。特别地，VOT 汇编了一个从广泛使用的序列收集而来的数据集，显示了各种对象和场景的平衡性。序列中的每帧图像都用不同的视觉属性进行标记，以减小跟踪结果的分析误差。进一步地，VOT 组织者利用 Matlab 创建了一个评估工具包，其可以使用用户自己提出的跟踪器在官

方提供的数据集上自动执行基本实验；同时提出了一种基于基本性能度量的比较协议，该协议具有明显的新颖性，即明确考虑到了结果的统计意义，并考虑了跟踪器的等价性问题。

在具体评价指标方面，基于最近对广泛使用的性能指标的分析，VOT 选择了两个正交的指标：准确度和鲁棒性。准确度衡量跟踪预测的边界框与 ground truth 边界框的重叠程度；鲁棒性度量跟踪过程中丢失目标的次数。VOT 的目标是比较每个实验中跟踪器分别在 6 个不同属性（摄像机运动、光照变化、物体大小变化、物体运动变化和非退化）的视频序列下的性能。为了实现一个跟踪器与其他跟踪器的性能比较，本文提出了一种基于排名的方法。简而言之，在一个实验中，VOT 根据每个属性序列上的得分分别对跟踪器进行排序。通过每个跟踪器在不同属性上的平均排名，VOT 获得了与性能度量相关的排名。对所有的绩效指标给予同等的权重，通过对相应的两个指标（准确度和鲁棒性）的排名进行平均，得到一个选定实验的最终排名。

在 VOT 后续的版本中[31-34]，开发方不断改进自己的数据标注方式（见图 4）和评价标准。文献[31]用旋转的边界框标注每帧数据，以更真实地表示目标位置。启用一个新的评估系统，使得跟踪器和 VOT 能更快地执行实验的直接通信，并兼容之前的 VOT 版本。文献[32-33]引入了一种易于解释的性能评价方法——预计的平均重叠率（Expected Average Overlap，EAO），扩展了评价方法，跟踪器将根据这个指标排名。其引入了子挑战 VOT-TIR，它也在 VOT 的框架下进行，主要用于处理红外和热成像的跟踪。与之前的 VOT 挑战一样，最新的 VOT2020 中[34]仍然使用准确性和鲁棒性来衡量目标跟踪的总体性能，最终汇总 EAO。通过更新数据集，VOT2020 超越了以往的挑战赛，提出了旨在解决短期跟踪问题的 VOT-ST 挑战赛、实时挑战赛 VOT-RT、长时跟踪挑战赛 VOT-LT、热成像图目标跟踪挑战赛 VOT-RGBT 和深度图 RGBD 挑战赛，并且采用了最新的绩效评估方案和最新的 VOT2020 Python 工具包[34]。

图 4 VOT 数据集中的多边形标注

Fig. 4 Polygon labeling in VOT datasets

2.3 其他视觉目标跟踪数据集

除了上述两个主流目标跟踪数据集外，目标跟踪社区也提出了大量针对各种目标特性而设计的数据集[26-30]。TC-128 由 128 个视频组成，这些视频被特别指定用于评估色彩增强的跟踪器。与 OTB 中定义的视频属性一样，这些视频也被划分为 11 种不同的属性[26]。TrackingNet[27]是首个大规模的目标跟踪数据集，介绍了从粗注释生成密集注释的标注技术。NfS[28]提供了 100 个快速运动的视频序列，其帧率高达 240 帧每秒，旨在分析外观变化对视觉跟踪性能的影响。LaSOT[29]属于密集跟踪数据集的范畴，有 352 万帧，平均序列长度为 2512 帧。此外，LaSOT 为每个视频提供了额外的语言描述，其他数据集则没有。为了放宽短期跟踪的强假设，文献[30]提出了一个拥有时长为 1 h 的视频（150 万帧）的跟踪数据集，用于评估长期跟踪场景中的算法性能。

3 目标跟踪基本框架的改进

早期的跟踪框架大多是基于粒子滤波和均值漂移实现的，这些早期的目标跟踪算法的探索为这一领域的后继研究奠定了基础[3-5]。近年来主要的跟踪框架是基于相关滤波（Correlation Filters，CF）[13]和孪生网络（Siamese Nets）[18]的改进。本文主要介绍这两方面的工作。

3.1 相关滤波 KCF 框架

大多数现代跟踪算法的核心部分是设计判别分类器，其任务是区分目标和周围环境。为了应对图像的自然变化，该分类器通常使用经过转换和缩放的样本图像块来进行训练，而这样的样本集是冗余且低效的。为了增广样本数量，CSK[13]利用循环矩阵构建了数千个转换过的图像块（见图 5），用于训练分类器。因为得到的数据矩阵是循环的，所以可以用离散傅里叶变换来对其进行对角化，从而极大地减少了存储量和计算量。有趣的是，对于线性回归，最终的公式相当于一个相关滤波器。文献[13]进一步提出了一种新的核化相关滤波器（Kernelized Correlation Filter，KCF），实现了快速准确的跟踪。

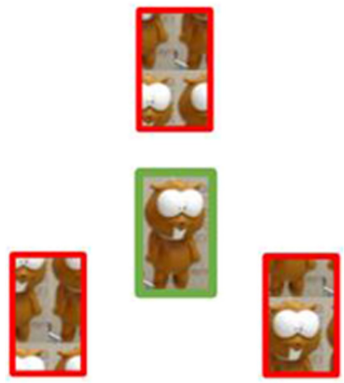

图 5 循环采样构建训练样本

Fig. 5 Circulant sampling used to construct training samples

具体来说，KCF 利用标准的相关滤波框架，训练出了一个岭回归分类器。目标是找到一个函数 $f(\boldsymbol{z})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{z}$，使得在循环样本$\{\boldsymbol{x}_i\}$上的检测结果和回归目标$\{y_i\}$之间的最小平方误差最小，即：

$$\min_{\boldsymbol{w}}\sum_i(f(\boldsymbol{x}_i)-y_i)^2+\lambda\|\boldsymbol{w}\|_2^2 \tag{1}$$

进一步地，利用核技巧，本文直接得出式(1)的闭式解：

$$\tilde{\boldsymbol{\alpha}}=F^{-1}\left(\frac{\hat{\boldsymbol{y}}}{\hat{\boldsymbol{k}}^{xx}+\lambda}\right) \tag{2}$$

其中，$\boldsymbol{k}^{xx}$表示 $\boldsymbol{x}$ 与它自己的核相关，$\wedge$ 表示离散傅里叶变换，而 F^{-1}表示离散的快速傅里叶逆变换。

此外，采用一种在线更新的策略来更新学到的参数 $\boldsymbol{\alpha}^t$，即：

$$\boldsymbol{\alpha}^t=(1-\eta_{cf})\alpha^{t-1}+\eta_{cf}\tilde{\boldsymbol{\alpha}}^t \tag{3}$$

其中，η_{cf} 表示相关滤波分类器的学习率，$\tilde{\boldsymbol{\alpha}}^t$ 利用当前 t 帧的跟踪结果通过式(2)计算得到。最终，当输入新一帧即 $t+1$ 帧图片 $\boldsymbol{z}^{t+1}$ 时，每层的检测响应结果为：

$$\boldsymbol{r}_{channel}=F^{-1}(\hat{\boldsymbol{\alpha}}^t\odot\hat{k}^{\boldsymbol{x}^t\boldsymbol{z}^{t+1}}) \tag{4}$$

每层响应相加之后，便可得到最终的相关滤波响应。

$$\boldsymbol{r}_{cf}=\sum_i w^i \tag{5}$$

3.2 孪生网络目标跟踪架构

传统上，对任意目标的跟踪是通过在线学习目标的表观模型来实现的，而唯一的训练数据则是视频本身。尽管这些方法取得了一些成功，但是这类在线学习的方法本质上都受限于可学习模式的丰富程度。最近，相关学者做了一些尝试来提升用于目标跟踪的深度卷积网络的表达能力。然而，当需要跟踪的对象事先未知时，需要在线进行随机梯度下降来适应网络的权重，从而严重影响了跟踪速度。SiamFC[18]在 ILSVRC15 数据集中的视频检测子集上训练了一种新的基于全卷积 Siamese 网络的跟踪算法。尽管该跟踪算法极其简单，但是在多个基准测试集中都达到了最优的性能，并且它能实时运行。

图 6 给出了孪生网络跟踪架构，对于搜索图像 x 和模板图像 $\boldsymbol{z}$，首先利用共享参数的卷积网络提取图像的特征；然后经过一个匹配模块，匹配之后得到一个相似度响应图，搜索图像 x 中的所有的区域都会和模板图像 $\boldsymbol{z}$ 计算一个相似度值，x 中和模板图像越相似的区域，相应得分也会越高。

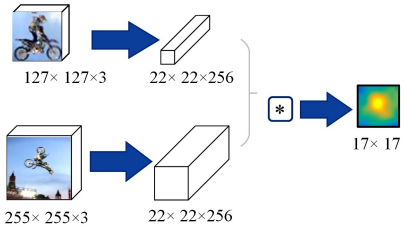

图 6 孪生网络跟踪框架

Fig. 6 Framework of siamese network tracking

4 目标表示的改进

4.1 传统的判别式跟踪方法

4.1.1 通过自适应结构化局部稀疏表观模型进行视觉跟踪 ASLA

稀疏表示方法应用于视觉跟踪，一般都会利用最小化重构误差来寻找最佳的候选目标。但是，多数基于稀疏表示的跟踪器只考虑整体表示，没有充分利用稀疏系数对目标和背景进行区分，因此场景中存在相似目标或遮挡的情况下更有可能失败。在 ASLA[11]中，Jia 等开发了一种简单且鲁棒的跟踪方法，即基于结构化的局部稀疏表观模型。该表示方法利用了目标的局部信息和空间信息，并采用了一种新的对准池方法。通过局部块间的相似度池化得到的相似度不仅有助于更准确地定位目标，而且可以处理遮挡问题。此外，文献[11]还采用了一种增量子空间学习和稀疏表示相结合的模板更新策略。该策略使模板能够适应目标的外观变化，减小了漂移的可能性，同时减小了被遮挡目标对模板的不良影响。在基准图像序列上的定性和定量评估结果表明，该跟踪算法优于当时的大部分算法。

4.1.2 实时压缩跟踪算法 CT

由于姿态变化、光照变化、遮挡和运动模糊等因素的影响，如何建立高性能、高效率的表观模型来实现鲁棒目标跟踪是一项具有挑战性的任务。现有的在线跟踪算法经常使用最近帧中的观测样本来更新模型，尽管取得了很大的成功，但仍有许多问题有待解决。首先，虽然这些自适应外观模型是依赖于数据的，但是没有足够的数据可供在线算法从一开始就进行学习；其次，在线跟踪算法经常遇到漂移问题。由于自学的结果，有可能增加未对齐的样本，导致外观模型的质量降低。Zhang 等[12]提出一种简单有效的基于多尺度图像特征空间提取特征的跟踪算法(Compressive Tracking，CT)。该方法中的表观模型采用非自适应随机投影，保持了目标的图像特征空间结构。为了有效地提取外观模型的特征，CT 构造了一个非常稀疏的测量矩阵，并使用相同的稀疏测量矩阵压缩前景目标和背景的样本图像(见图 7)。通过在压缩域内进行在线更新的朴素贝叶斯分类器，将跟踪任务表述为二值分类。本文提出的压缩跟踪算法可以实时运行，并且在效率、准确性和鲁棒性方面在当时都达到了领先水平。

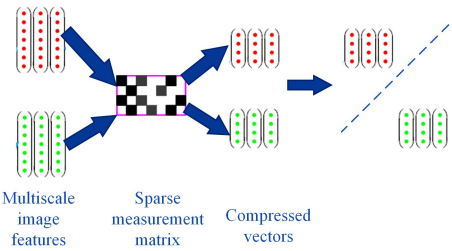

图 7 压缩跟踪示意图[26]

Fig. 7 Diagram of compressive tracking[26]

4.2 相关滤波类跟踪方法

在相关滤波类跟踪算法中，也有很多代表性的工作是基于目标表示方面的改进，如 CN[14]，CF2[16]，Staple[35]等。

4.2.1 自适应颜色属性的实时视觉跟踪

大多数最先进的视觉跟踪器要么依赖亮度信息，要么使用简单的颜色表示来描述图像。与视觉跟踪相反，在目标识别和检测方面，复杂的颜色特征与亮度相结合可以提供优异的性能。由于跟踪问题的复杂性，所要求的颜色特征必须具有较高的计算效率，并具有一定的光照不变性，同时保持较高的识别能力。CN[14]研究了颜色在检测跟踪框架中的贡献，结果表明，颜色属性提供了优越的视觉跟踪性能。Danelljan 等[14]进一步提出了一种自适应的颜色属性的低维变种形式。该方法在 41 个具有挑战性的颜色视频序列上进行了定量和基于属性的评估，证实了该方法的有效性。

4.2.2 多层卷积特征视觉跟踪 CF2

视觉目标跟踪是一个极具挑战性的问题，因为目标物体经常会由于变形、突然运动、背景杂乱和遮挡等因素而导致外观发生显著的变化。在 CF2[16]中，Ma 等利用从目标识别数据集中训练得到的深度卷积神经网络特征，来提高跟踪精度和鲁棒性。最后一个卷积层输出编码目标的语义信息，这种表示对剧烈的外观变化具有鲁棒性。然而，深层特征的空间分辨率粗糙，无法精确定位目标。相比之下，早期的卷积层能

够提供更精确的定位，但是对外观变化的不确定性更小。接着，将卷积层的层次解释为一个对应的非线性图像金字塔表示，并利用这些多层次的抽象信息来进行视觉跟踪。具体来说，该方法自适应地学习每个卷积层上的相关滤波以编码目标外观，并通过分层推断每一层的最大响应来定位目标。在大规模基准测试数据集上的大量实验结果表明，该算法具有较好的性能。

4.2.3 实时跟踪的互补学习 Staple

最近，基于相关滤波器的跟踪取得了优异的性能，对运动模糊和光照变化等挑战具有极大的鲁棒性。然而，由于相关滤波学习的模型严重依赖于被跟踪对象的空间布局，因此对变形非常敏感。而基于颜色统计的模型具有互补的特点：它们能很好地处理形状的变化，但当整个序列的光照不一致时，就会受到严重影响。在 Staple[35] 中，作者展示了一个在岭回归框架中结合多种线索的简单跟踪器（见图 8），其运行速度可以超过 80 帧每秒，不仅在主流的 VOT14 竞赛中优于所有参赛算法，而且在多个基准测试集上也优于最近的更复杂的跟踪器。

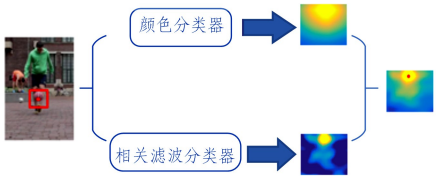

图 8 互补学习示意图

Fig. 8 Diagram of complementary learning

4.3 孪生网络类跟踪方法

在孪生网络框架下，目标跟踪社区涌现出了一系列快速且优秀的算法，如 SiamRPN[19]，Siam R-CNN[21] 等。

4.3.1 孪生区域推荐网络 SiamRPN

在 SiamRPN 中，Li 等[19] 提出了基于大规模图像对端到端训练的孪生区域推荐网络（Siamese Region Proposal Network，Siamese-RPN）。该网络由 Siamese 特征提取子网络和区域推荐子网络组成，包括分类分支和回归分支。在推理阶段，提出的框架被设定为一次局部检测任务。可以预先计算孪生子网的模板分支，并将相关层表示为普通的卷积层来进行在线跟踪。由于推荐（proposal）的不断精细化，传统的多尺度测试和在线微调可以直接省略。Siam-RPN 的运行速度为 160 帧每秒，同时在 VOT2015，VOT2016 和 VOT2017 的实时挑战中取得了领先的性能。

4.3.2 Siam R-CNN：重新检测的视觉跟踪

Siam R-CNN 孪生的再检测架构释放出了两阶段目标检测方法在视觉目标跟踪领域的全部力量。文献[19]提出了一种新的基于轨迹的动态规划算法，该算法利用第一帧模板和前一帧预测的重新检测，来建模被跟踪对象和潜在干扰对象的完整历史信息，使得该方法能够做出更好的跟踪决策，以及在长时间遮挡后重新检测出被跟踪的对象。最后，文献[19]提出了一种新的难样本挖掘策略来提高 Siam R-CNN 对相似外观目标的鲁棒性。Siam R-CNN 在 10 个跟踪基准集上达到了目前的最佳性能，对于长时跟踪有着极佳的性能表现。

5 时空上下文方面的改进

5.1 空间上下文类改进

5.1.1 使用自适应相关滤波器的目标跟踪 MOSSE

在传统方法中，相关滤波虽然不常用，但它可以在旋转、遮挡和其他干扰的情况下跟踪复杂的目标，而且其速度是最先进技术的 20 倍以上。最古老和最简单的相关滤波器使用简单的模板，应用于跟踪通常会导致失败，更现代的方法（如 ASEF[36]）表现更好，但它们的训练需求不适合视觉目标跟踪。视觉目标跟踪要求从单一帧训练鲁棒的滤波器，并随着目标对象的外观变化而动态调整。

文献[9]提出一种新的相关滤波器，即最小输出平方误差（Minimum Output Sum of Squared Error，MOSSE）滤波器，其在仅使用一帧初始化时可以稳定地跟踪目标物体。基于 MOSSE 滤波器的跟踪算法在以每秒 669 帧的速度运行的情况下，对光线、尺度、姿态和非刚性形变都是鲁棒的。MOSSE 算法充分利用了空间上下文信息，构造了大量循环矩阵来训练目标跟踪分类器。所构造的循环矩阵本质上包含了复杂的空间上下文信息，保证了 MOSSE 算法的有效性。

5.1.2 通过密集时空上下文学习的快速视觉跟踪 STC

文献[15]提出了一种简单、快速、鲁棒的基于密集时空上下文的视觉跟踪算法 STC。该方法在贝叶斯框架中建立了感兴趣的目标及其局部密集上下文之间的时空关系，该框架则对目标及其周边区域的简单低层特征（如图像强度和位置）之间的统计相关性建模。然后在计算中考虑了目标位置先验信息的置信度图来处理跟踪问题，从而有效地缓解了目标位置模糊不清的问题。进一步地，Zhang 等[15] 提出了一种新的尺度自适应方案，该方案能够有效地处理目标尺度的变化。该方案采用快速傅里叶变换（Fast Fourier Transform，FFT）进行快速学习和检测，只需要 4 次 FFT 操作，因此提出的跟踪算法能在 i7 机器上以每秒 350 帧的速度运行。大量的实验结果表明，该算法在效率、准确性和鲁棒性方面都优于已有的算法。

5.2 时序上下文类改进

5.2.1 模板更新

视觉跟踪本质上是处理随时间变化的非平稳图像流。虽然大多数现有的算法能够在受控环境下很好地跟踪目标，但它们通常会在目标外观或周围光照发生显著变化时失败。造成这种失败的一个原因是许多算法使用了目标的固定外观模型。固定外观模型在跟踪开始之前只使用可用的外观数据进行训练，这实际上限制了被建模的外观的范围，并且忽略了跟踪过程中可用的大量信息（如形状变化或特定的光照条件）。在 IVT[5] 算法中，Ross 等[5] 提出了一种增量学习低维子空间表示的跟踪方法，有效地在线适应目标的外观变化。基于主成分分析的增量算法的模型更新包括两个重要的特性：1）正确更新样本均值的方法；2）设置一个遗忘因子，以确保更少的建模能力会被用到拟合旧的观测上。这两种特性对于提高整体跟踪性能都有着显著的贡献。大量的实验证明了该跟踪算法在室内和室外环境下的有效性，特别是当目标的姿态、尺度和光照都发生了很大的变化时，也能成功处理。

5.2.2 帧间信息关联

近年来,基于判别式相关滤波器(Discriminative Correlation Filter,DCF)的方法极大地促进了目标跟踪技术的发展。然而,随着跟踪性能的不断提高,其速度和实时性逐渐下降。此外,越来越复杂的模型和大量的可训练参数,引入了严重的过拟合。在ECO的工作中,Danelljan等[17]找到了计算复杂度和过拟合问题背后的关键原因,可同时提高速度和性能。并且,ECO引入了如下策略:1)分解的卷积算子,极大地减少了模型中的参数数量;2)具有样本分布特性的生成模型,显著降低了记忆和时间复杂度,同时提供了更好的样本多样性;3)具有提高鲁棒性并降低复杂性的模型更新策略。在VOT2016,UAV123,OTB-2015和Temple-Color这4个基准上进行了全面的实验,结果表明该方法实现了20倍的加速性能。

在此基础上,Zhang等[37]把帧间信息关联的思想应用到了Siamese网络中。Siamese方法通过从当前帧中提取一个外观模板来定位下一帧中的目标,以解决视觉跟踪问题。通常,此模板与前一帧中积累的模板线性组合,导致信息随时间呈指数衰减。虽然这种更新方法使结果变得更好,但由于它过于简单,限制了学习更新可以获得的潜在收益。因此,文献[37]提出用一种学习更新的方法来代替手工制作的更新函数。具体来说,使用一个名为UpdateNet的卷积神经网络,给定初始模板、累积模板和当前帧的模板,以估计下一帧的最优模板。UpdateNet非常紧凑,可以很容易地集成到现有的Siamese跟踪器中。通过将该方法应用于SiamFC[18]和DaSiamRPN[38]两种Siamese跟踪器,证明了该方法的通用性。

6 最新的进展

6.1 应用新的神经网络

随着深度神经网络技术的发展[39-46],目标检测[47-51]、语义分割[52-56]等领域已在大量使用最新的深度神经网络技术,目标跟踪领域也成功地应用了一系列最新的神经网络技术[57-86]。

Wang等[58]提出一种基于全卷积神经网络的通用目标跟踪新方法,融合了高层语义信息和底层位置信息。在文献[61]中,STCT使用了一种卷积神经网络(Convolutional Neural Network,CNN)的序列训练方法,来有效地转移预先训练好的深度特征,用于在线应用。Qi等[64]提出一种新的基于CNN的跟踪框架,该框架充分利用了不同CNN层的特征,并利用自适应对冲方法将多个CNN跟踪器对冲为一个更强的跟踪器。Zhang等[59]提出,即使在没有使用大量辅助数据离线训练的情况下,简单的两层卷积网络也足以强大到学习视觉跟踪的鲁棒表示。Guo等[65]提出了动态变化的孪生网络,通过一个快速转换的学习模型,该方法能够有效地从先前帧在线学习目标的外观变化和背景抑制。Song等[68]提出了一种基于对抗学习的关键算法来解决外观变化和正负样本间的不平衡问题。

最近的工作中,Zhu等[69]利用连续帧中丰富的光流信息[87-88]来提高特征表示和跟踪精度;Dong等[72]提出了一种新的三元组损失算法,通过将表达性深度特征加入到Siamese网络框架中来代替成对损失进行训练,从而提取目标跟踪过程中有表达力的深度特征;Zhang等[74]为了解决卷积网络填充在目标跟踪过程中产生位置偏差这一问题,设计了更深和更宽的残差网络来消除填充的负面影响;Gao等[75]引入了图卷积神经网络[89-90],从历史目标样本的时空结构中获取上下文信息;Wang等[91-92]利用无监督深度训练技术,提出了无监督深度跟踪算法[76],达到了与有监督方法同等的跟踪精度,展示了无监督类方法的潜力;基于孪生网络,Wang等[77]在两个独立的阶段分别增强跟踪器的鲁棒性和判别性,把两个独立的阶段串联起来得到最后的跟踪器;Wang等[86]把目标跟踪建模为一个特殊的目标检测问题(实例检测),先训练出一个目标检测器,再利用模型不可知的元学习策略[93]来初始化检测器,得到了非常惊人的性能;Yang等[83]使用了可调整大小的卷积来适应目标物体的形状变化,并提出离线训练一个递归神经优化器[94],采用元学习的思路来更新跟踪模型,使之快速收敛;Chen等[85]及Yu等[86]则直接把目标检测领域中的边界框回归策略[95]和交叉注意力机制[96]应用到孪生网络目标跟踪框架中。

这些新的神经网络应用到目标跟踪领域之后,显著地提升了跟踪的鲁棒性和精度,从而推动了目标跟踪的发展。

6.2 解决目标跟踪领域的特定问题

虽然更深的神经网络能提升目标跟踪领域的特征表示能力,但视觉跟踪任务经常遭遇各种独特而严苛的挑战。为了解决跟踪任务中的特定问题,研究者们提供了许多有效的解决方案[57-86]。

为了解决目标对象由于变形、突然运动、严重遮挡和目标超出视线外而导致表观发生显著变化的视觉跟踪问题,Ma等[57]提出了长时相关滤波跟踪算法,该算法能够在跟踪失败的情况下重新检测目标。为了加快之前的神经网络类跟踪方法的速度,Held等[62]提出了一种无需在线训练的跟踪方法,其跟踪速度能达到100帧每秒。为了处理大规模的表观变化问题,Zhang等[63]提出了多任务相关滤波器,考虑了不同特征之间的相互依赖性。Gao等[64]提出了相对型跟踪器,它可以有效地利用图像中前景和背景之间的相对关系来进行物体外观建模。为了在跟踪时能够更好地检查和修正跟踪结果,Fan等[66]利用多线程并行技术,提出了并行跟踪与验证框架来实现有效跟踪。Song等[67]应用残差学习来考虑外观的变化,并将相关滤波重构为一层卷积神经网络。该方法将特征提取、响应图生成和模型更新集成到神经网络中进行端到端训练。为了消除背景中干扰物的影响,Zhu等[70]在离线训练阶段引入了一种有效的采样策略来控制训练数据分布,使模型关注语义干扰。

近年来,为了突破极端的前-背景数据不平衡这一网络训练时的瓶颈,Lu等[71]提出了一种新的收缩损失来惩罚简单训练的数据。Zhang等[73]引入了一种新的空间对准模块,该模块可提供连续反馈,使目标以标准化的高宽比从边界向中心转换,从而使相关滤波器能够在对齐良好的样本上工作,以便更好地进行跟踪。为了更好地利用梯度信息和避免过拟合,Li[78]等设计了一种新的梯度引导网络[45,97],利用梯度中的判别信息,通过前馈和后向操作更新孪生网络中的模板。为了使短时跟踪更接近实际应用,达到长时跟踪的目的,Yan

等[79]在略读和精读的思想[98-99]下，提出了一种新的具有鲁棒性和实时性的跟踪算法。为了去除跟踪过程中边界效应的影响，Huang 等[80]通过对检测阶段产生的响应图的变化率进行限制，来抑制极端变化的发生，从而得到更加鲁棒和准确的目标跟踪器。Huang 等[81]直接在目标检测器上构建跟踪算法，并引入了锚点更新策略来避免过拟合问题，缩小了与目标检测算法之间的差距。为了适应无人机定位问题，Li 等[84]提出了一种在线自动自适应学习时空正则化项的新方法，引入空间局部响应图变化作为空间正则化，使相关滤波器专注于可信度较高部分的学习。

上述针对目标跟踪领域的特定问题提出的算法均有效地解决了某个特定问题，在大型数据集上也都获得了非常高的准确率。

6.3 讨论与展望

近年来，目标跟踪领域取得了令人瞩目的成就，但是跟踪算法也变得越来越耗时，表 1 列出了近年来较有代表性的快速跟踪算法[100-102]。特别地，MOSSE 在只使用原始灰度特征的情况下可实现超高速的目标跟踪，取得了 669 帧每秒的惊人的速度。然而，随着目标跟踪算法研究的深入，实时性方面却越来越差，如 CRPN 级联了多个 RPN 网络，使得区域推荐网络具有更强的判别性，但是仅仅取得了 32 帧每秒的速度。希望未来有更多的工作能关注目标跟踪算法的速度。

表 1 快速目标跟踪算法的性能对比

Table 1 Comparison of fast object tracking algorithm

跟踪器名称	OTB 结果	VOT 结果	速度(FPS)	GPU	发表	源代码	亮点
MOSSE[9]	—	—	**669**	No	CVPR2010	Matlab	只使用原始灰度特征的情况下实现超高速的目标跟踪
KCF[13]	0.477	**VOT2014 第 3**	**172**	No	TPAMI2014	Matlab	运用循环矩阵理论解释了相关滤波目标跟踪，并使用 HOG 特征[40]，显著提升了跟踪性能
Staple[35]	0.579	**VOT2015 第 5**	80	No	CVPR2016	Matlab	在 HOG 特征之外加入了颜色特征，形成优势互补
ECO[17]	**0.630**	VOT2017 第 12	60	No	CVPR2017	Matlab	使用各种加速策略，把使用深度特征的相关滤波方法提速 20 倍
SiamFC[18]	0.582	VOT2017 第 22	86	Yes	ECCV2016	MatConvNet	利用孪生网络，在深度学习框架下近乎完美地模拟了相关滤波类目标跟踪方法
SiamRPN[19]	**0.637**	**VOT2018 第 3**	**160**	Yes	CVPR2018	Pytorch	引入区域推荐网络(RPN)来代替传统目标跟踪中的多尺度搜索框
C-RPN[101]	**0.663**	—	32	Yes	CVPR2019	MatConvNet	级联了多个 RPN 网络，使得区域推荐网络具有更强的判别性
SiamMask[102]	—	VOT2017 第 17	60	Yes	CVPR2019	Pytorch	把快速的视频目标跟踪和半监督视频分割融合在一个框架中，取得了较高的准确度

注：前三名用加粗字体显示

结束语 在复杂的现实场景中，计算机目标跟踪系统和人类视觉系统相比仍有巨大差距，因此真正意义上通用且快速准确的目标跟踪研究还远未完成。但是，基于过去十几年中目标跟踪社区取得的突破性进展，我们相信，在研究者们的共同努力下，未来目标跟踪领域一定会取得更大的成就。

参 考 文 献

[1] LI X, HU W, SHEN C, et al. A survey of appearance models in visual object tracking[J]. ACM transactions on Intelligent Systems and Technology (TIST), 2013, 4(4): 1-48.

[2] SMEULDERS A W M, CHU D M, CUCCHIARA R, et al. Visual tracking: An experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(7): 1442-1468.

[3] COMANICIU D, RAMESH V, MEER P. Kernel-based object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.

[4] AVIDAN S. Ensemble tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261-271.

[5] ROSS D A, LIM J, LIN R S, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1-3): 125-141.

[6] BABENKO B, YANG M H, BELONGIE S. Visual tracking with online multiple instance learning[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 983-990.

[7] MEI X, LING H. Robust visual tracking using ℓ1 minimization [C]// 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009: 1436-1443.

[8] KALAL Z, MATAS J, MIKOLAJCZYK K. Pn learning: Bootstrapping binary classifiers by structural constraints[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 49-56.

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

[10] HARE S, GOLODETZ S, SAFFARI A, et al. Struck: Structured output tracking with kernels[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(10): 2096-2109.

[11] JIA X, LU H, YANG M H. Visual tracking via adaptive structural local sparse appearance model[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 1822-1829.

[12] ZHANG K,ZHANG L,YANG M H. Fast compressive tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2014,36(10):2002-2015.

[13] 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.

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

[15] ZHANG K,ZHANG L,LIU Q,et al. Fast visual tracking via dense spatio-temporal context learning[C]// European Conference on Computer Vision. Springer,Cham,2014:127-141.

[16] 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. 2015:3074-3082.

[17] DANELLJAN M,BHAT G,SHAHBAZ K F,et al. Eco:Efficient convolution operators for tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:6638-6646.

[18] BERTINETTO L,VALMADRE J,HENRIQUES J F,et al. Fully-convolutional siamese networks for object tracking[C]// European Conference on Computer Vision. Springer, Cham, 2016:850-865.

[19] 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. 2018:8971-8980.

[20] 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. 2019:4660-4669.

[21] VOIGTLAENDER P,LUITEN J,TORR P H S,et al. Siam r-cnn:Visual tracking by re-detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:6578-6588.

[22] YILMAZ A,JAVED O,SHAH M. Object tracking:A survey[J]. Acm computing surveys (CSUR),2006,38(4):13.

[23] HU W,ZHOU X,LI W,et al. Active contour-based visual tracking by integrating colors, shapes, and motions[J]. IEEE Transactions on Image Processing,2012,22(5):1778-1792.

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

[25] KRISTAN M,MATAS J,LEONARDIS A,et al. The visual object tracking vot2015 challenge results[C]// Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015:1-23.

[26] LIANG P,BLASCH E,LING H. Encoding color information for visual tracking:Algorithms and benchmark[J]. IEEE Transactions on Image Processing,2015,24(12):5630-5644.

[27] MULLER M,BIBI A,GIANCOLA S,et al. Trackingnet:A large-scale dataset and benchmark for object tracking in the wild[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018:300-317.

[28] KIANI G H,FAGG A,HUANG C,et al. Need for speed:A benchmark for higher frame rate object tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017:1125-1134.

[29] FAN H,LIN L,YANG F,et al. Lasot:A high-quality benchmark for large-scale single object tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019:5374-5383.

[30] VALMADRE J,BERTINETTO L,HENRIQUES J F,et al. Long-term tracking in the wild:A benchmark[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018:670-685.

[31] 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 Workshops. 2017:1949-1972.

[32] KRISTAN M,LEONARDIS A,MATAS J,et al. The sixth visual object tracking vot2018 challenge results[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[33] KRISTAN M,MATAS J,LEONARDIS A,et al. The seventh visual object tracking vot2019 challenge results[C]// Proceedings of the IEEE International Conference on Computer Vision Workshops. 2019:2206-2241.

[34] KRISTAN M,LUKEZIC A,DANELLJAN M,et al. The new VOT2020 short-term tracking performance evaluation protocol and measures[J/OL]. https://data.votchallenge.net/vot2020/vot-2020-protocol.pdf.

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

[36] BOLME D S,DRAPER B A,BEVERIDGE J R. Average of synthetic exact filters[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2009:2105-2112.

[37] ZHANG L,GONZALEZ-GARCIA A,WEIJER J,et al. Learning the model update for siamese trackers[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019:4010-4019.

[38] ZHU Z,WANG Q,LI B,et al. Distractor-aware siamese networks for visual object tracking[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018:101-117.

[39] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM,2017,60(6):84-90.

[40] GOODFELLOW I,BENGIO Y,COURVILLE A,et al. Deep learning[M]. Cambridge:MIT press,2016.

[41] RUSSAKOVSKY O,DENG J,SU H,et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision,2015,115(3):211-252.

[42] KINGMA D P,BA J. Adam:A method for stochastic optimization[J]. arXiv:1412.6980,2014.

[43] WITTEN I H, FRANK E. Data mining: practical machine learning tools and techniques with Java implementations[J]. Acm Sigmod Record, 2002, 31(1): 76-77.

[44] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.

[45] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. 2014: 2672-2680.

[46] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9729-9738.

[47] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.

[48] GIRSHICK R. Fast r-cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.

[49] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]// Advances in Neural Information Processing Systems. 2015: 91-99.

[50] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[51] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// European Conference on Computer Vision. Springer, Cham, 2016: 21-37.

[52] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.

[53] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[54] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, Cham, 2015: 234-241.

[55] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2961-2969.

[56] HE K, GIRSHICK R, DOLLÁR P. Rethinking imagenet pre-training[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019: 4918-4927.

[57] MA C, YANG X, ZHANG C, et al. Long-term correlation tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5388-5396.

[58] WANG L, OUYANG W, WANG X, et al. Visual tracking with fully convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 3119-3127.

[59] ZHANG K, LIU Q, WU Y, et al. Robust visual tracking via convolutional networks without training[J]. IEEE Transactions on Image Processing, 2016, 25(4): 1779-1792.

[60] QI Y, ZHANG S, QIN L, et al. Hedged deep tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4303-4311.

[61] WANG L, OUYANG W, WANG X, et al. Stct: Sequentially training convolutional networks for visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1373-1381.

[62] HELD D, THRUN S, SAVARESE S. Learning to track at 100 fps with deep regression networks[C]// European Conference on Computer Vision. Springer, Cham, 2016: 749-765.

[63] ZHANG T, XU C, YANG M H. Multi-task correlation particle filter for robust object tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4335-4343.

[64] GAO J, ZHANG T, YANG X, et al. Deep relative tracking[J]. IEEE Transactions on Image Processing, 2017, 26(4): 1845-1858.

[65] GUO Q, FENG W, ZHOU C, et al. Learning dynamic siamese network for visual object tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 1763-1771.

[66] FAN H, LING H. Parallel tracking and verifying: A framework for real-time and high accuracy visual tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 5486-5494.

[67] SONG Y, MA C, GONG L, et al. Crest: Convolutional residual learning for visual tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2555-2564.

[68] SONG Y, MA C, WU X, et al. Vital: Visual tracking via adversarial learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8990-8999.

[69] ZHU Z, WU W, ZOU W, et al. End-to-end flow correlation tracking with spatial-temporal attention[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 548-557.

[70] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 101-117.

[71] LU X, MA C, NI B, et al. Deep regression tracking with shrinkage loss[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 353-369.

[72] DONG X, SHEN J. Triplet loss in siamese network for object tracking[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 459-474.

[73] ZHANG M, WANG Q, XING J, et al. Visual tracking via spatially aligned correlation filters network[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 469-485.

[74] ZHANG Z, PENG H. Deeper and wider siamese networks for real-time visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 4591-4600.

[75] GAO J, ZHANG T, XU C. Graph convolutional tracking[C]// Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition. 2019：4649-4659.

[76] WANG N，SONG Y，MA C，et al. Unsupervised deep tracking [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019：1308-1317.

[77] WANG G，LUO C，XIONG Z，et al. Spm-tracker：Series-parallel matching for real-time visual object tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019：3643-3652.

[78] LI P，CHEN B，OUYANG W，et al. Gradnet：Gradient-guided network for visual object tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019：6162-6171.

[79] YAN B，ZHAO H，WANG D，et al. 'Skimming-Perusal' Tracking：A Framework for Real-Time and Robust Long-term Tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019：2385-2393.

[80] HUANG Z，FU C，LI Y，et al. Learning aberrance repressed correlation filters for real-time uav tracking[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019：2891-2900.

[81] HUANG L，ZHAO X，HUANG K. Bridging the gap between detection and tracking：A unified approach[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019：3999-4009.

[82] WANG G，LUO C，SUN X，et al. Tracking by instance detection：A meta-learning approach[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：6288-6297.

[83] YANG T，XU P，HU R，et al. ROAM：Recurrently Optimizing Tracking Model[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：6718-6727.

[84] 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/CVF Conference on Computer Vision and Pattern Recognition. 2020：11923-11932.

[85] CHEN Z，ZHONG B，LI G，et al. Siamese Box Adaptive Network for Visual Tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：6668-6677.

[86] YU Y，XIONG Y，HUANG W，et al. Deformable Siamese Attention Networks for Visual Object Tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：6728-6737.

[87] DOSOVITSKIY A，FISCHER P，ILG E，et al. Flownet：Learning optical flow with convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015：2758-2766.

[88] ILG E，MAYER N，SAIKIA T，et al. Flownet 2.0：Evolution of optical flow estimation with deep networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017：2462-2470.

[89] KIPF T N，WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv：1609.02907，2016.

[90] SCHLICHTKRULL M，KIPF T N，BLOEM P，et al. Modeling relational data with graph convolutional networks[C]// European Semantic Web Conference. Springer，Cham，2018：593-607.

[91] RADFORD A，METZ L，CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv：1511.06434，2015.

[92] DOSOVITSKIY A，SPRINGENBERG J T，RIEDMILLER M，et al. Discriminative unsupervised feature learning with convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2014：766-774.

[93] FINN C，ABBEEL P，LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[J]. arXiv：1703.03400，2017.

[94] YANG T，CHAN A B. Learning dynamic memory networks for object tracking[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018：152-167.

[95] ZHU C，HE Y，SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019：840-849.

[96] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017：5998-6008.

[97] DUCHI J，HAZAN E，SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of machine learning research，2011，12(7)：2121-2159.

[98] NEBEHAY G，PFLUGFELDER R. Clustering of static-adaptive correspondences for deformable object tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015：2784-2791.

[99] HONG Z，CHEN Z，WANG C，et al. Multi-store tracker (muster)：A cognitive psychology inspired approach to object tracking [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015：749-758.

[100] DALAL N，TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE，2005，1：886-893.

[101] FAN H，LING H. Siamese cascaded region proposal networks for real-time visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019：7952-7961.

[102] 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. 2019：1328-1338.

ZHANG Kai-hua，born in 1983，Ph.D，professor. His main research interests include image segmentation，level sets and visual tracking.