高技术通讯 2018 年 第28 卷 第3 期: 207 ~213

doi: 10.3772/j.issn.1002-0470.2018.03.003

基于卷积神经网络的目标跟踪算法综述①

胡 硕② 赵银妹 孙 翔

(燕山大学电气工程学院 秦皇岛 066004)

摘 要 目标跟踪是机器视觉领域一个重要的研究方向,在军事、交通等领域有着广泛的应用。随着训练数据和硬件的发展,越来越多的学者将深度学习应用于视觉跟踪领域。近几年来,一大批基于深度学习的跟踪算法被提出,与传统的机器学习方法相比,包含多个隐含层的卷积神经网络(CNN)具有更强大的特征学习和特征表达能力。分析了目标跟踪中的难题以及用卷积神经网络解决此类问题的可能性,综述了卷积神经网络在视觉跟踪领域的发展,并对卷积神经网络在视觉目标跟踪中的最新成果进行了总结和深入分析,最后对卷积神经网络在目标跟踪领域未来的发展进行了展望。

关键词 卷积神经网络(CNN),深度学习,目标跟踪,机器视觉

0 引言

目标跟踪是计算机视觉领域的重要研究方向之一[1],在机器人、人机交互、军事侦查、智能交通、虚拟现实等军事、民用领域都有广泛的应用。近年来,许多学者在目标跟踪方面开展了大量工作,并取得了一定的进展[2]。但是,在复杂环境中仍存在目标外观变形(目标纹理、形状、姿态变化等)、光照变化、快速运动和运动模糊、背景相似干扰、平面内外旋转、尺度变化、遮挡和出视野等难题,使得复杂环境下鲁棒实时的目标跟踪仍然是一个具有挑战性的问题。一般的视觉目标跟踪系统主要包括输入图像(视频)、运动模型(均值漂移、滑动窗口、粒子滤波)、特征提取、目标外观模型以及模型更新等几个部分,其中特征的提取与表达对目标跟踪算法的性能起决定性作用[3]。

2006 年,Hinton 等提出了具有深层次特征表达能力的深度学习算法[4]。深度学习模拟人脑的视觉处理机理,可以从大量的数据中主动学习特征,并成功应用于图像分类[5]、物体检测[6]等领域。深度学习能够主动学习提取底层到高层结构性的特征,具有强大的分类功能,使得将深度学习引入到目标跟踪算法的研究具备了可行性。卷积神经网络(convolutional neural network, CNN)是深度学习常用模型之一,近年来,随着 GPU 以及图像数据的发展,深度学习尤其是卷积神经网络在国内外目标跟踪领域得到了越来越多的关注。本文旨在对基于卷积神经网络(CNN)的目标跟踪方法进行综述和分析。

1 传统的目标跟踪算法

2010 年以前目标跟踪领域的一些经典算法有 Meanshift[7]、粒子滤波[8],以及基于特征点的光流算法[9]等。2010 年出现检测与跟踪相结合的方法,于是将目标跟踪分为产生式模型和判别式模型[3]。产生式模型是将跟踪问题看作搜索与跟踪目标最为相似的图像区域,相似度最高的区域确定为目标区域,包括采用稀疏表示模型[10]、密度估计模型[11]、增量子空间模型[12]等。Mei 等人[13]将稀疏表示应

① 国家自然科学基金(61741418)资助项目。

② 男,1976 年生,博士,讲师;研究方向:模式识别与图像处理;联系人,E-mail: hus@ysu.edu.cn

(收稿日期:2017-09-14)

用于视觉跟踪，从目标模板子空间中找到具有最小重建误差的目标，取得了良好的抗遮挡性能。Ross 等人[14]采用增量子空间模型，逐步学习低维子空间表示，在跟踪过程中更新，在线适应目标外观模型的变化。

判别式方法将跟踪看作二分类问题，利用在线学习或离线训练的检测器来区分前景目标与背景，找到目标对象与背景的边界，进而将目标与背景分离，找出目标的位置。这些跟踪算法通常基于多实例学习[15]、P-N 学习[16]、在线学习[17,18]以及结构化输出支持向量机(support vector machine，SVM)[19]等分类器。Xie 等人[20]提出了多示例目标跟踪方法，将单个图像集合到样本包中，训练正负样本进行分类处理，提高了目标姿态变化等背景下的鲁棒性。

由于相关滤波器的计算效率以及优异性能，使其在视觉跟踪领域备受关注，Bolme 等人[21]提出了一种具有平均误差的滤波器，也叫最小输出和快速相关滤波器，跟踪速度为数百帧每秒。2012 年 Henriques 等人提出了一种基于相关滤波的 CSK[22]跟踪方法，解决了密集采样的问题，并利用傅立叶变换快速实现了检测过程，完成跟踪。后续又提出了基于方向梯度直方图(histogram of oriented gradient，HOG)特征[23]的核相关滤波(KCF)跟踪方法[24]。

随着深度卷积网络在目标跟踪领域的发展，其效果已经逐渐超越传统的跟踪算法，表 1 是部分传统的目标跟踪算法与深度卷积神经跟踪算法在 OTB2013 测试集上的测试效果对比。

表 1 传统目标跟踪与卷积神经跟踪算法比较

算法	CF2	CNN	KCF	SCM	CSK
成功率	0.605	0.597	0.514	0.499	0.398
精确率	0.891	0.852	0.740	0.656	0.545

CF2 和 CNN 是基于卷积神经网络的目标跟踪算法，其他是传统的跟踪算法，从表格中可以看出前者有更好的跟踪性能，主要原因是传统的目标跟踪算法采用手工特征，选取时需要经验并且对目标的表达能力不足，跟踪准确率很难得到进一步提高，而深度卷积神经网络提取的特征比传统的手工特征更加丰富，具有更强的表达能力，因此为目标跟踪提供了新思路。

2 卷积神经网络

卷积神经网络模拟人大脑的学习过程，构建多层的神经网络，随着层数的增加，获得的特征也越来越抽象。经典的卷积神经网络由输入层、卷积层、池化层、全连接层和输出层，如图 1 所示。其中输入层、卷积层和池化层组成特征提取层，全连接层和输出层组成分类层。

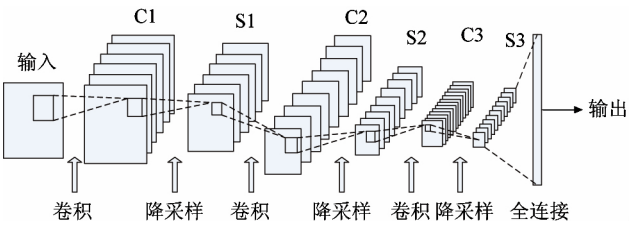

图 1 卷积神经网络的基本结构图

卷积层为特征提取层，每个神经元的输入与前一层的局部感受野相连，并提取该局部的特征，一旦该局部特征被提取后，它与其它特征间的位置关系也随之确定下来。卷积层的主要作用是根据卷积核对图像进行卷积操作，卷积运算一个重要的特点是通过卷积运算，可以使原信号特征增强，并且降低噪音。

池化层是特征映射层，通过对每个特征映射图的局部区域进行加权求和，增加偏置后通过一个非线性函数在池化层得到新的特征图。池化的作用是：(1)对特征图进行降维，避免过拟合；(2)可以一定程度上缓解形变等问题。

全连接层用于连接所有的特征，将输出值送给分类器(如 Softmax 分类器)，起到一个分类的作用。

卷积神经网络同时具有感受野和权重共享的特性，可以使网络输入多维的图像，降低特征提取过程中的数据重建的复杂度，抑制平移、缩放带来的影响。相比于浅层学习模型，卷积神经网络的这种深度分层架构，可以不依赖外界条件学习数据特征，对于类似图像这种高度非结构化、分布复杂的数据具有很强的刻画能力和泛化性能。因此在计算机视觉

领域,利用卷积神经网络提取特征的方法正超越传统的手工提取特征方法,成为主流方法。

3　基于卷积神经网络的目标跟踪

近几年,学者们提出了一大批优秀的基于卷积神经网络目标跟踪算法[25-39],在 VOT(visual object tracking) 数据集每年举办的跟踪算法竞赛中,基于卷积神经网络的目标跟踪算法取得了优异的结果。

3.1　基于分类的卷积神经网络目标跟踪

传统跟踪算法在跟踪时只给定第一帧的信息作为训练数据,在这种情况下,在跟踪过程中针对当前目标从头训练一个深度模型面临巨大挑战。针对上述问题,Wang 等人[25] 在 2013 年的深度学习跟踪器(deep learning tracker,DLT) 算法中首次提出"离线训练,在线微调"的训练模式,即利用大规模数据集离线训练网络,获得通用特征的表示能力,再利用第一帧目标位置周围的正负样本进行在线微调,这种迁移学习的思路,大大减少了跟踪对于样本的需求,使得深度学习开始应用于跟踪领域。

Seunghoon 等人[26] 通过使用卷积神经网络学习辨别显著性图来提出在线视觉跟踪算法,如图 2 所示,首先借助大型图片数据集离线训练卷积神经网络 RCNN,然后将最后卷积层的输出作为对象的通用特征描述符,使用在线支持向量机(SVM) 来学习目标外观模型。最后由显著性图显示目标的空间配置,提高目标定位精度,并显示了目标分割的能力。缺点: RCNN 模型是一种用来图像分类的模型,重点是区分不同类别的物体,在训练过程中,对于每张图像都产生 2000 个候选区域,并对每个候选区域都进行特征提取,使得消耗的时间非常长。

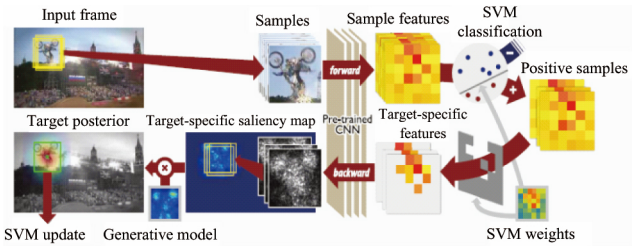

图 2　Seunghoon 等人算法流程图[26]

Nam 等人[27] 提出了创新的多域训练方法和训练数据交叉运用的新思路,采用视频序列作为训练集,将每一个序列当成单独的区域,每个区域都有一个针对它的二分类层,用于区分当前序列的背景和前景,算法流程如图 3 所示,网络设计为共享层和特定领域层,并且借鉴检测任务中的难例挖掘和边框回归。

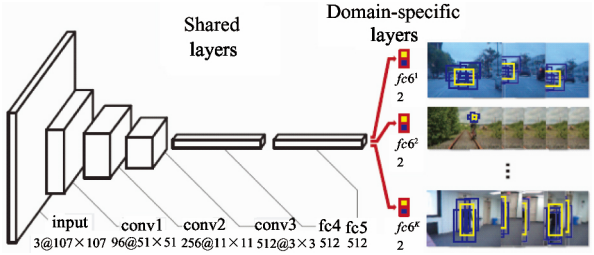

图 3　Nam 等人算法流程图[27]

Heng 等人[28] 在此基础上提出改进,同样采用多域的思路,不同的是利用循环神经网络(RNNs) 对目标结构进行建模,采用级联策略融合 CNN 与 RNN,提高模型在类似干扰物存在下的鲁棒性,两种跟踪算法的总体思路和目标检测中的 RCNN 模型(regions with CNN) 比较类似,虽然网络已经很小,但是速度仍然是一个难题。

Nam 等人[29] 又从模型的可靠性出发提出了树形结构的 CNN 目标跟踪算法,旨在解决跟踪问题中的目标遮挡以及目标丢失问题。首先根据第一帧的信息初始化网络模型,然后利用加权检测方式计算每个 CNN 模型的可靠性,找出分数最大的 CNN 模型作为跟踪的网络,最后结合分类器和边框回归模型定位出目标的位置。该算法夺得 VOT2016 视觉跟踪比赛的冠军。

3.2　基于回归的卷积神经网络目标跟踪

基于分类的方法充分利用卷积神经网络优秀的特征提取与分类能力,对每一个正负样本进行二分类,其缺点是依赖于分类器的性能。与基于分类的目标跟踪不同,基于回归的目标跟踪算法是通过网络输出的热度图或概率图直接回归出目标的位置。

Wang 等人[30] 在文献 [24] 的基础上将自编码器替换为深度卷积神经网络使用 CNN 作为获取特征的网络模型,充分利用图片本身结构化信息生成概

率图，从概率图中可以直接回归获得目标的最终位置，并且在卷积层与全连接层中间采用空间金字塔采样(spatial pyramid pooling，SPP-NET) 来提高最终的定位准确度。测试准确率和成功率都有很大提高。

Wang 等人[31] 通过分析卷积层的特征图谱，提出采用卷积神经网络层级间特征来进行目标跟踪的新方法，如图 4。首先将通过构建两个互补的热度图预测网络 GNet，SNet，其中 GNet 捕获目标的类别信息，SNet 将目标与背景进行相似外观分离。然后 GNet 和 SNet 都在第一帧进行初始化，为目标进行前景热图回归，最后回归出目标的位置。能达到有效防止跟踪器漂移，具有对目标本身的形变更加鲁棒的效果。但在实验中，网络的复杂性使得实时性不能保证，大约只有 1 帧/秒，同时对于遮挡处理效果不好。

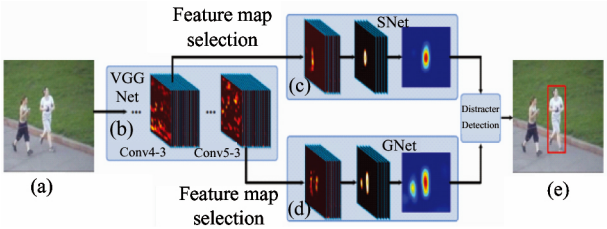

图 4 Wang 等人算法流程图[31]

Ma 等人[32] 也采用层级间的特征进行目标跟踪，与上一篇文献不同的是作者首先在第一帧的时候，在 VGG19 网络的 Conv3 _4，Conv4 _4，Conv5 _4 三层训练三个相关滤波器; 之后的每一帧，从上一帧的预测跟踪区域中提取出三个层的特征图进行双线性差值运算，最后得到相关响应图定位出位置，完成粗粒度到细粒度的跟踪。但是作者在进行跟踪过程时一直假设目标的尺度是不变的，因此在应对尺度变化时缺乏鲁棒性。

Qi 等人[33] 利用卷积神经网络每个层后面的卷积图进行训练，将每一层特征训练出来的相关滤波器(correlation filter) 作为一个弱分类器，同时利用对冲算法对弱跟踪器进行线性融合，从而得到目标的位置。

Danelljan 等人[34] 提出了一种使用连续卷积滤波的目标跟踪方法，利用内插值法将卷积神经网络不同分辨率的特征图插值到连续空间域，再应用 Hessian 矩阵求得亚像素精度的目标位置。随后 Danelljan 等人[35] 又在此基础上进行改进，将特征提取部分进行筛选降维，去除冗余，并且利用高斯混合模型简化了训练集，提高了算法的时间效率和空间效率，降低模型的更新频率，有效地防止模型漂移。

3.3 基于相似度匹配的卷积神经网络目标跟踪

基于相似度匹配的目标跟踪算法是将第一帧或者上一帧看成是模板，模板与候选区域图像进行相似度计算，从而得到最优解。一些学者受到传统相似度匹配算法的启发，提出了基于相似度匹配的卷积神经网络目标跟踪算法并取得了理想的效果。

Held 等人[36] 提出一种以 100 帧/秒速度学习卷积网络跟踪方法，作者在分析视觉跟踪前后两帧的关系时发现，其符合拉普拉斯分布。于是作者从这两帧出发提出了使用一个简单的不需要训练的前馈网络来进行跟踪。首先使用有边框标签(但没有类信息) 的 ImageNet 数据集和视频序列离线训练两个具有 5 个卷积层和 3 个全连接层的网络进行特征学习，然后将目标对象的特征与当前帧中的要素进行相似度匹配，找到相似度最大的位置，最后输出当前帧中目标位置左上角和右下角的坐标，从而定位出目标的位置。但当光线、遮挡等问题发生时，模型更新不及时会使跟踪目标丢失。

Tao 等人[37] 提出了一种利用双网络搜索机制进行目标跟踪的方法，首先应用 ImageNet 大规模数据集训练两个网络模型 VGG 网络(或 Alexnet[38] 网络) 得到匹配函数，然后在线跟踪时使用外部视频进行微调，其中一个网络只输入第一帧，另一个网络输入当前帧，根据匹配函数选择与初始帧目标最为匹配的区域作为跟踪的结果。这种网络的好处是对于搜索图像的大小不做规定。

Bertinetto 等人[39] 提出了一种全卷积孪生网络跟踪算法，如图 5 所示。该方法利用计算机视觉竞赛组织 ILSVRC 提供的用于目标检测的视频集来离线训练一个解决相似性学习的模型，其中一个网络输入视频的第一帧，另一个网络输入待搜索目标的图像，然后输出一个响应图，预测对象在特定位置出现的可能性。与文献 [37] 不同的是，该方法的双流

网络是共享全卷积层的,此外,网络可以端对端训练,提取专门用于视觉跟踪的浅和深卷积特征。由于此方法是从一次性模型中获益,跟踪速度快。

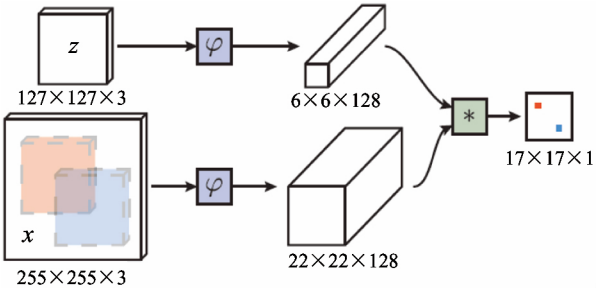

图 5　Bertinetto 等人算法流程图[39]

Valmadre[40] 等人提出了采用不对称的孪生网络,将相关滤波器作为微 CNN 层,通过在线学习算法进行反向传播梯度,以优化基础特征表示,从而形成一个端到端的跟踪。在训练卷积神经网络时采用数百万个从视频集合中获取的随机对,很大程度上接近真实跟踪,但这种数据也存在很大的弊端,当视频中目标发生很大的遮挡、形变时,跟踪结果会出现误差。文章作者通过实验证明,在比较深的卷积神经网络里,相关滤波器作用不明显,在相对比较浅的网络里能实现比较好的效果。

4　对比分析

表 2 展示了一份基于卷积神经网络的部分最新成功的跟踪方法在 OTB2013 测试数据集上的测试结果对比,表中分别包含测试的精确率、成功率与跟踪速度。分数越高代表算法的跟踪性能越好。

表 2　卷积神经网络最新成功的跟踪方法对比

跟踪算法法	精确率	成功率	跟踪速度
SANet	0.950	0.686	1 帧/秒
MDNet	0.948	0.708	1 帧/秒
ECO	0.930	0.709	6 帧/秒
CCOT	0.899	0.672	0.3 帧/秒
CF2	0.891	0.605	11 帧/秒
HDT	0.889	0.603	10 帧/秒
FCNT	0.856	0.599	1 帧/秒
CFNet	0.822	0.610	75 帧/秒
SiamFC	0.809	0.608	86 帧/秒
GOTURN	0.620	0.444	165 帧/秒

卷积神经网络在目标跟踪领域取得了一些不错的成果,但仍然存在一些困难和问题,例如: (1) 训练一个深度神经网络是复杂、耗时的,如何保持跟踪精度的情况下提高速度是卷积神经网络跟踪中的一大难题; (2) 选取哪种训练集能够使跟踪任务更稳定。

5　发展趋势展望

目标跟踪是机器视觉领域的一个重要研究方向,经历了数十年的发展后,传统跟踪算法在复杂场景下仍存在大量问题。深度学习的出现及其在目标检测领域的优良表现为目标跟踪提供了新思路,许多学者对基于深度学习的目标跟踪展开研究。目前研究重点和发展趋势主要为以下几点:

(1) 构建合适的网络。目标跟踪问题和物体检测分类问题的差异性使得运用检测模型来解决跟踪问题,效果不够理想。而且池化操作还会降低图像的分辨率,从而损失空间位置信息。因此单纯的套用卷积神经网络并不是很可取,还需要考虑与其他算法相结合来弥补损失的空间位置信息。

(2) 视频数据集的使用。使用更贴合跟踪任务的视频数据集来训练网络。近日,谷歌发布了一个大型视频数据集 YouTube-8M,其中包含 800 万个带标注的 YouTube 视频的 URL。使用大型视频训练网络还需要兼顾训练速度与准确率。

6　结　论

本文对基于卷积神经网络的目标跟踪进行了详细地介绍。首先介绍了目标跟踪的发展历程以及跟踪中的难题。其次介绍了卷积神经网络的基本结构。然后对基于卷积神经网络目标跟踪的方法进行介绍对比,最后对本文进行了总结。纵观各种大型的跟踪竞赛,基于卷积神经网络的跟踪正在逐步替代传统的跟踪,成为跟踪领域的佼佼者,但是这些跟踪算法准确性的提升都是以牺牲速度为前提,因此在今后的发展中,如何充分发挥卷积神经网络端到端的能力,是大家值得思考的问题。

参考文献

[1] Smeulders A W, Chu D M, Cucchiara R, et al. Visual tracking: an eperimental survey [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 36(7): 1442-1468

[2] Yang H, Shao L, Zheng F, et al. Recent advances and trends in visual tracking: a review [J]. *Neurocomputing*, 2011, 74(18): 3823-3831

[3] Wang N, Shi J, Yeung D Y, et al. Understanding and diagnosing visual tracking systems [C]. In: Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015. 3101-3109

[4] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. *Neural Computation*, 2006, 18(7): 1527-1554

[5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. In: Proceedings of Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014. 580-587

[6] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015. 2892-2900

[7] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, USA, 2002. 142-149

[8] Djuic P M, Kotecha J H, Zhang J, et al. Particle filtering [J]. *Signal Processing Magazine IEEE*, 2003, 20(5): 19-38

[9] Dowson N, Bowden R. Mutual information for Lucas-Kanade tracking (MILK): an inverse compositional formulation [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2008, 30(1): 180-185

[10] Mei X, Ling H. Robust visual tracking using l1 minimization [C]. In: Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 2009. 1436-1443

[11] Han B, Comaniciu D, Zhu Y, et al. Sequential kernel density approximation and its application to real-time visual tracking [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2008, 30(7): 1186-1197

[12] Jepson A D, Fleet D J, El-Maraghi T F. Robust online appearance models for visual tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(10): 1296-1311

[13] Mei X, Ling H, Wu Y, et al. Efficient minimum error bounded particle resampling L1 tracker with occlusion detection [J]. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 2013, 22(7): 2661

[14] 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

[15] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2011, 33(8): 1619-32

[16] Kalal Z, Mikolajczyk K, Matas J. Tracking learning detection [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(7): 1409-1422

[17] Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting [C]. In: Proceedings of the British Machine Vision Conference, Edinburgh, UK, 2013. 47-56

[18] Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking [C]. In: Proceedings of the European Conference on Computer Vision, Marseille, France, 2008. 234-247

[19] Hare S, Saffari A, Torr P H S. Struck: structured output tracking with kernels [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(10): 2096-2109

[20] Xie Y, Qu Y, Li C, et al. Online multiple instance gradient feature selection for robust visual tracking [J]. *Pattern Recognition Letters*, 2012, 33(9): 1075-1082

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

[22] Henriques J, o F, Caseiro R, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]. In: Proceedings of the 12th European Conference on Computer Vision. Florence, Italy, 2012: 702-715

[23] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005. 886-893

[24] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 37(3): 583-596

[25] Wang N, Yeung D Y. Learning a deep compact image representation for visual tracking [C]. In: Proceedings of

the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2013. 809-817

[26] Hong S, You T, Kwak S, et al. Online tracking by learning discriminative saliency map with convolutional neural network [J]. *Computer Science*, 2015: 597-606

[27] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking [C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016. 4293-4302

[28] Fan H, Ling H. SANet: structure-aware network for visual tracking [EB/OL]. arXiv. org, arXiv: 1611. 06878

[29] Nam H, Baek M, Han B. Modeling and propagating CNNs in a tree structure for visual tracking [EB/OL]. arXiv. org, arXiv: 1608. 07242

[30] Wang N, Li S, Gupta A, et al. Transferring rich feature hierarchies for robust visual tracking [EB/OL]. arXiv. org, arXiv: 1501. 04587

[31] Wang L, Ouyang W, Wang X, et al. Visual tracking with fully convolutional networks [C]. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015. 3119-3127

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

[33] Qi Y, Zhang S, Qin L, et al. Hedged deep tracking [C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016. 4303-4311

[34] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual gracking [C]. In: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016. 472-488

[35] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking [C]. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA , 2017. 6931-6939

[36] Held D, Thrun S, Savarese S. Learning to track at 100 FPS with deep regression networks [C]. In: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016. 749-765

[37] Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking [C]. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016. 1420-1429

[38] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. arXiv. org, arXiv: 1409. 1556

[39] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking [C]. In: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016. 850-865

[40] Valmadre J, Bertinetto L, Henriques J F, et al. End-to-end representation learning for Correlation Filter based tracking [C]. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017. 5000-5008

Review of object tracking based on convolutional neural networks

Hu Shuo, Zhao Yinmei, Sun Xiang

(* School of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004)

Abstract

The article points that object tracking is an important research topic in the field of machine vision, and it is widely used in national defense, transportation and other fields. With the development of training data and hardware, more and more researchers are applying the technique of deep learning to visual tracking. Recently, a large number of tracking algorithms based on deep learning are proposed. Compared with the traditional machine learning methods, the techniques using convolution neural networks with multiple hidden layers have more powerful capacities of feature learning and feature expression. Then, it analyzes the difficult problems in object tracking and the possibility of using convolution neural networks to solve object-tracking problems. Furthermore, the development of convolution neural networks in visual tracking is reviewed, and the latest results of applying convolution neural networks to visual target tracking are summarized and analyzed. Finally, the future development of convolutional neural networks in object tracking is discussed.

Key words: convolutional neural networks (CNN) , deep learning, object tracking, machine vision