

多目标跟踪的对象初始化综述

文成宇¹ 房卫东² 陈伟^{1,2}

1 中国矿业大学计算机科学与技术学院 江苏 徐州 221116

2 中国科学院上海微系统与信息技术研究所无线传感网与通信重点实验室 上海 200050

(cumtwency@163.com)

摘要 对象初始化方法决定了如何对待多目标跟踪问题,与后续的多目标跟踪效果直接相关。不同的对象初始化方法能够确定不同的多目标跟踪框架,每一种框架都提供一种解决问题的思路,使得多目标跟踪的对象初始化问题具有巨大的研究前景。目前关于多目标跟踪中的对象初始化方法的综述性文献较少或缺乏系统性的对象初始化概述,因此文中从多假设跟踪方法、网络流方法、深度学习方法与主题发现方法4个方面对多目标跟踪的对象初始化方法进行分析。系统地阐述了不同多目标跟踪框架下的任务转换和对象映射问题,汇总了多目标跟踪的对象初始化方法。

关键词: 多目标跟踪;对象初始化;多假设跟踪;网络流;深度学习;主题发现

中图法分类号 TP391

Object Initialization in Multiple Object Tracking: A Review

WEN Cheng-yu¹, FANG Wei-dong² and CHEN Wei^{1,2}

1 School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

2 Key Laboratory of Wireless Sensor Network & Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China

Abstract Object initialization method determines how to treat the multi-object tracking problem, being directly related to the subsequent tracking result. Different object initialization methods confirm different multi-object tracking frameworks and each framework provides a way to solve the problem, which makes the object initialization of multi-object tracking a huge research prospect. Currently there are few literature on object initialization methods of multi-target tracking, or lacks a systematic overview of object initialization. Therefore, we analyze the object initialization methods on four aspects: multi-hypothesis tracking, network flow, deep learning and topic discovery. We systematically expound the task conversion and object mapping problems under different multi-object tracking frameworks, and summarize the object initialization methods for the multi-object tracking.

Keywords Multiple target tracking, Object initialization, Multiple hypothesis tracking, Network flow, Deep learning, Topic discovery

1 引言

多目标跟踪问题(Multiple Object Tracking, MOT)是计算机视觉领域中一项极具挑战性的任务^[1],它的核心任务是跟踪视频序列中的不同对象,即寻找对象在视频序列中的轨迹。近年来,随着研究不断深入,多目标跟踪在智能交通^[2-3]、自动驾驶^[4]、智慧城市^[5]、智能监控^[6-7]、智能机器人^[8]、智能医疗^[9]、军事^[10-11]等领域发挥着重要作用。

多目标跟踪的基本框架如图1所示,根据是否实时或是否利用视频序列的所有帧,多目标跟踪分为在线跟踪^[12-13]、离线跟踪^[14-15]和介于两者之间的在线-离线跟踪^[16];根据是否存在多信息源头,多目标跟踪分为单摄像机跟踪和多摄像机跟踪^[17],其中多摄像机跟踪问题还涉及摄像机的时间、空间配准问题^[18];根据是否检测对象,多目标跟踪分为基于检测跟踪^[19]和基于人工标注跟踪^[20],其中基于检测跟踪(Based on Detection)是主流的多目标跟踪方法^[21]。

到稿日期:2021-02-04 返修日期:2021-06-03

基金项目:国家自然科学基金(51874300);国家自然科学基金委员会—山西省人民政府煤基低碳联合基金(U1510115);中国科学院上海微系统与信息技术研究所无线传感网与通信重点实验室开放基金(20190902,20190913)

This work was supported by the National Natural Science Foundation of China(51874300), National Natural Science Foundation of China and Shanxi Provincial People's Government Jointly Funded Project of China for Coal Base and Low Carbon(U1510115) and Open Research Fund of Key Laboratory of Wireless Sensor Network & Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences(20190902,20190913).

通信作者:陈伟(chenwdavior@163.com)

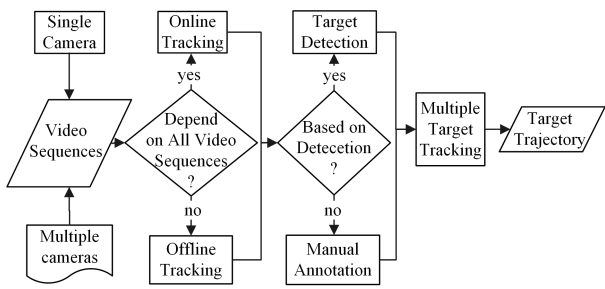


图 1 多目标跟踪的基本框架

Fig. 1 Multi-target tracking basic framework

基于检测跟踪的多目标跟踪方法可以分为 4 个主要步骤^[22-23]:1)对各个视频帧中的新对象进行建模,即对象初始化^[24];2)根据建立的模型进行对象检测,获得初始对象序列;3)在后续帧中重复寻找目标对象进行多目标跟踪;4)获得一系列的对象轨迹实现多目标跟踪。对象初始化与后续的跟踪步骤直接相关,因此合理的对象建模对跟踪结果有积极影响^[25]。

对象初始化是多目标跟踪步骤中的第一个部分,也是最重要的部分。目前关于多目标跟踪中对象初始化的综述性文献较少或者缺乏系统性概述^[26],并且大部分多目标跟踪问题的文献对对象初始化的研究层面不深^[27],或者直接使用在单目标跟踪工作中已经取得较好成绩的对象初始化方法^[28],进而将大量研究工作集中在对象关联中^[7,27-34]。虽然对象关联是多目标跟踪区别于单目标跟踪的重要特征,但是本文认为基于检测的多目标跟踪方法具有系统性、连续性等特点,因此多目标跟踪的基本框架中的每一个步骤都值得研究人员关注^[35]。对象初始化方法通过分析多目标跟踪问题的特点,对多目标跟踪问题进行转换,将对象模型映射为不同的表达方式,得到了较好的研究结果^[36]。不同的对象初始化方法提供了解决多目标跟踪问题的不同思路,表明了对象初始化方法研究拥有巨大的潜力。本文将从多假设跟踪方法、网络流方法、深度学习和主题发现方法 4 个方面对多目标跟踪的对象初始化方法进行分析,重点关注主流的、有启发性的多目标跟踪对象初始化方法,并对对象初始化未来的发展方向进行展望。

2 对象初始化

为了便于问题描述和数据处理,研究者会对视频序列中的对象进行建模,对象初始化狭义上通过符号或字母来表示视频序列中新出现的对象,完成对象建模^[26]。而实际上对象建模具有更深层的意义,为了方便与后续步骤进行对接和配合,广义上的对象表示还包括了对象的检测方法和对象的特征图描述^[24,37]。视频序列通常是稳定连续的,不存在背景和对象离散变化的情况,并且视频序列的首帧中包含最多的新对象,虽然后续视频帧中也可能出现新对象,但是其数量远远少于首帧中的新对象,因此对象初始化的大部分工作都集中在视频序列的首帧当中。

多数研究者把对象初始化和对象跟踪分为两个独立的模块^[38-40],将对象初始化模块的输出数据作为对象关联模块的输入数据;也有研究者通过结合这两个模块把多目标跟踪

问题抽象为端到端的框架^[41-43],不区分模块来进行对象建模和关联处理。不论分开处理还是整体处理,对象初始化都是为了更好地配合多目标跟踪框架,以获得更准确的轨迹结果^[44]。

对象初始化实际上是对现实世界的抽象化和数字化处理,目的是为了更方便地描述问题、分析问题和解决问题,其行为和影响贯穿了整个多目标跟踪的基本框架。反过来说,多目标跟踪问题的解决过程也需要一种方便、直观、抽象、数字化的描述符号,而对象初始化就是为了获得具有这些特征的表达方式。不同对象的初始化表达方式,代表着解决多目标跟踪问题的不同思路,而好的表达方式对解决问题起着积极的作用。

3 对象初始化方法

本节将根据不同的对象初始化方法,从多假设跟踪方法(Multiple Hypothesis Tracking, MHT)、网络流图方法(Network Flow, NF)、深度学习方法(Deep Learning, DL)和主题发现(Topic Discovery, TD)这 4 个方面,综述多目标跟踪问题在不同对象初始化方法的转换下,得到的相应的 4 种不同解决问题的思路。

3.1 多假设跟踪

多假设跟踪算法是一种在关联框架下发生冲突时,不进行最优选择而形成多种假设,以推迟最终选择的一种多目标跟踪方法^[45],这种方法基于最大后验概率(Maximum A Posteriori, MAP)框架,即隐马尔可夫模型,利用当前和先前的观测来估计当前时间对象的状态^[46]。通过隐马尔可夫模型能够将多目标跟踪问题转化为计算对象状态的最大后验概率问题,而最大后验概率问题也被学者证明与最大加权独立集(Maximum Weight Independent Set, MWIS)问题等效^[47],因此多目标跟踪问题也可以被转化为最大加权独立集问题。

在实际应用中,通过对象检测器处理视频帧,利用两个连续帧中的成对检测响应构建轨迹图,之后通过迭代重复多项式时间 MWIS 算法将轨迹图中层次较小的轨道合并为较长的轨道,以实现目标的跟踪^[48];多目标跟踪问题中的 MWIS 问题需要在对象的跟踪时间内解决,而不是从整个视频序列的开始处理,因此如何解决两个相邻帧之间的 MWIS 问题非常重要^[49];通过将检测对象的步态特征和局部外观的时间一致性进行聚类,能够提高拥挤场景中多目标跟踪的性能^[28];对每个轨迹假设中的外观模型单独地进行训练,能够有效实现 MHT 算法对目标高阶信息的利用^[50]。但是 MHT 算法会保留所有正确的潜在轨迹假设,导致计算量呈指数级增长,因此在实际应用中需要限制轨迹假设的数量^[51],通过前一帧的反馈信息来制定最大加权问题(Maximum Weighted Clique Problem, MWCP),并通过求解该问题实现多假设跟踪^[52]。

MHT 算法包含以下 3 个步骤。

(1)轨迹假设构建和更新:根据视频序列构建初始化的轨迹假设集合。而更新过程为一个对象维持多个可能的轨迹,在每一帧中,处于轨迹检测阈值中的新对象被分配给现有的轨迹假设。

(2)全局假设生成:通过评估每一条轨迹假设的置信度,不断更新相邻帧之间的轨迹假设集合。假设已经获得截至当前帧的所有帧中的轨迹假设集合,则当前帧的全局假设生成就是将当前帧中所有对象分配给对应的轨迹假设。

(3)轨迹假设修剪:由于全局假设生成过程中维持多个可能的轨迹,因此随着帧数的积累,轨迹假设的数量呈指数级增长,而通过修剪轨迹假设有助于保持 MHT 算法持续稳定地运行。

MHT 算法中,用 $D=\{D^1,\cdots,D^t,\cdots,D^T\}$ 表示所有输入对象的集合,其中 T 表示视频序列的帧数。第 t 帧的对象集合 $D(t)$ 表示为 $D^t=\{d_1^t,d_2^t,\cdots,d_{b(t)}^t\}$,其中, d_k^t 表示第 t 帧中的第 k 个对象。用对象序列 $(d^1,d^2,\cdots,d^k)$ 表示第 k 帧中的轨迹假设,其中, $d^t\in D^t\cup\{\phi\}$, $d^t\in D^t$ 时,表示第 t 帧 D^t 中的对象 d^t 被连接到该轨迹假设上;而 $d^t\in\{\phi\}$ 时,表示一个虚假的对象被连接到该轨迹假设上,这个虚假对象用于处理丢失的对象。

由于同一时刻不可能出现多个同一对象,因此用 $V^t=\{V_1^t,V_2^t,\cdots,V_{N^t}^t\}$ 表示第 t 帧的轨迹假设集合,其中 N^t 表示轨迹假设的数量,用 $G^t=\{G_1^t,G_2^t,\cdots,G_{n^t}^t\}$ 表示第 t 帧的全局轨迹假设, G^t 集合的元素在当前全局假设中进行选择。表 1 列出了视频序列、对象检测和 MHT 的符号说明。

表 1 视频序列、对象检测和 MHT 的符号汇总
Table 1 Symbol summary of video sequence,detection and multiple hypothesis tracking

Symbol	Description
T	frame number of all the video
D	detection set of all frames
D^t	detection set of frame t
$D(t)$	detection number in frame t
d_k^t	the k -th tracklet hypotheses in frame t
f_k^t	frame number of detection d_k^t
V	tracklet hypothesis
V_k^t	the k -th tracklet hypothesis in frame t
N^t	number of tracklet hypothesis in frame t
G^t	global hypothesis in frame t
G_k^t	the k -th tracklet hypothesis of global hypothesis in frame t
n^t	number of tracklet hypothesis in global hypothesis in frame t

MHT 算法在全局假设更新时,构造一个无向图 $G=(V,E)$ 来建模 MWIS,每个点表示一个权重为 w_h 的轨迹假设 h ,当两个顶点 i 和 j 冲突时,用边 $(i,j)\in E$ 连接这两个点。MHT 问题如下式所示:

$$\max_x \sum_h w_h x_h \tag{1}$$

$$\text{s. t. } x_i + x_j \leq 1, (i,j) \in E \tag{2}$$

$$x_i \in \{0,1\} \tag{3}$$

其中, x_h 是一个指示符,当 $x_h=1$ 时,表示轨迹假设 h 被加入到当前的全局假设中;当 $x_h=0$ 时,表示轨迹假设 h 没有被加入到当前的全局假设中。式(2)表示两个轨迹假设不能同时被加入到全局假设中。

通过分析轨迹假设剪枝步骤能够看出 MHT 中的对象初始化方式在后续步骤中的其他变化^[53]。如图 2 所示,假设第 $k-3$ 帧中的关联假设包括 3 个检测和 2 条轨迹假设。图 2 中的方块表示轨迹 $\overline{T_1}$ 和 $\overline{T_2}$ 在上一帧中已经确定的对象,椭圆表示确定对象周围一定的范围, Z_1, Z_2 和 Z_3 分别为当前帧中已

检测到的目标。两种关联假设表示为:检测 Z_1, Z_2 和 Z_3 可能属于轨迹 $\overline{T_2}$;检测 Z_2, Z_3 可能属于轨迹 $\overline{T_1}$ 。

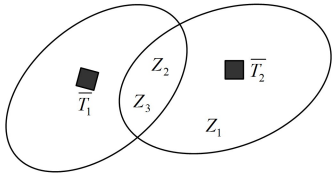


图 2 包含 3 个检测和 2 条轨迹假设的关联假设
Fig. 2 Association hypothesis including three detections and two trajectory hypotheses

如图 3 所示,基于图 2 中的关联假设可以继续向前预测两帧,得到一种 4 层结构的轨迹树。图 3 中的圆圈表示对象在不同时刻潜在的对应该检测结果,圆圈内的序号表示该检测结果所关联的轨迹号;旧根节点 A 表示一个对象,序号改变时说明出现了新对象,分支表示对象潜在在轨迹状态的假设。可以看出,随着假设层数的增多,关联假设出现组合爆炸的可能,因此,需要进行剪枝来减少假设空间的数目,例如,图 3 中第 k 帧找出每个对象可能性最大的分支假设是轨迹 2,然后回溯建立新的根节点,对不包含轨迹 2 的分支进行剪枝操作。

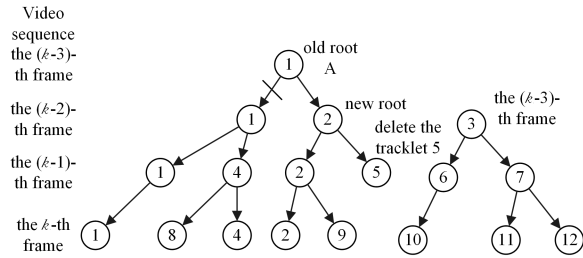


图 3 3 层假设构成的 2 棵轨迹树
Fig. 3 Two trajectory trees composed of four-level hypotheses

除了剪枝方法以外,为了确保 MHT 算法的可行性,还可以采用聚类、假设和轨迹合并等技术来减少数据量,以达到减少 MHT 算法的计算量的目的^[51]。

MHT 算法的对象初始化方法适用于噪声环境下密集机动目标的跟踪,且实时性较好。近年来,随着计算机计算性能的提升,经典 MHT 算法在标准测试数据集上取得了不错的效果^[52]。但是对象的遮挡问题会对全局假设的生成结果产生较大的影响,导致轨迹的置信度不断降低,最终丢失跟踪目标。

3.2 网络流方法

3.1 节介绍的 MHT 方法是在一个小的时间窗口(帧数跨度很小,一般少于 5 帧)中进行处理,而网络流方法通过考虑更多的帧来获得更多的信息。网络流方法将从视频帧中检测到的对象映射到连通图中,通过这种方法把多目标跟踪问题转化为全局最优求解问题^[54];而在此基础上进一步构成本流网络,可以通过网络流算法有效解决转化后的多目标跟踪问题^[55],例如,最小成本流算法^[55]、基于多切割^[56]或多环境^[57]的子图分解法。Xu 等提出了一种具有短期存储的方法,可以纠正部分网络流图中的局部关联错误^[58]。Kumar 等通过局部线索、全局线索和中间线索来分割检测图,并通过消息传递和移动算法来优化算法性能^[59]。基于网络流方法的全局最优问题还可以利用最大流前置(push relabel)^[60]方法、

连续最短路径(successive shortest paths)^[61]方法和动态规划(dynamic programming)^[62]方法解决;此外,还能通过场景信息估计^[63]、区分性特征更新^[64]等阶段式的方法,优化网络流方法的跟踪结果。

网络流方法将多目标跟踪问题转化为全局优化问题,通过这种描述方法,将连续视频帧之间的成对匹配检测对象用最小成本流图中的节点表示,从而简化了网络流图中的路径估计步骤。用 $\chi = \{X_i\}$ 表示一个对象检测集合,其中,每一个元素 $X_i = (x_i, s_i, a_i, t_i)$ 表示一个检测响应。检测响应 X_i 中的 x_i 表示位置, s_i 表示尺寸大小, a_i 表示外观, t_i 表示对象所在的帧数;用 $T_k = \{X_{k_1}, X_{k_2}, \dots, X_{k_i}\}$ 表示一个有序的对象检测列表,其中, $X_{k_i} \in \chi$;用 $\tau = \{T_k\}$ 表示一个轨迹假设。多目标跟踪问题可以表示为在给定观测集 χ 的情况下最大化轨迹假设 τ 的后验概率:

$$\tau^* = \arg \max_{\tau} P(\tau|\chi) = \arg \max_{\tau} \prod_i P(x_i|\tau) P(\tau) \quad (4)$$

假设成本流网络为 $G(\chi)$,起始点为 s ,终点为 t ,则对于检测每一个响应 $X_{k_i} \in \chi$,创建相应的两个节点 u_i 和 v_i ,并与起点 s 和终点 t 相连,分别建立检测边(observation edges) (u_i, v_i) 、转换边(transition edges) (v_i, u_i) 、进边(enter edges) (s, u_i) 和出边(exit edges) (v_i, t) 。然后计算边的成本和流,建立一个与成本流网络对应的成本流图。

图 4 给出了 3 个连续视频帧中分别有 2, 4, 3 种检测结果时,产生的网络最小代价流的例子,目的是产生全局最优的解。

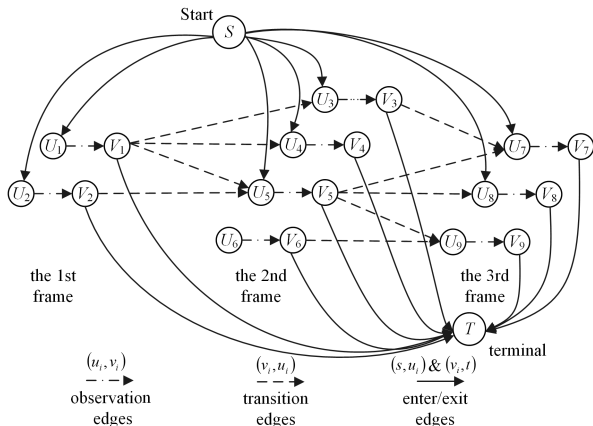


图 4 网络最小代价流的例子

Fig. 4 Example of network minimum cost flow

寻找最小成本流网络 G 的最佳关联假设等效于将流从起点 s 发送到终点 t ,使成本最小化。网络图中的每一个流(flow)表示一个对象的跟踪轨迹,从起点 s 到终点 t 的总量就等于最终对象的跟踪轨迹数量,而最小成本流网络 G 上流的总成本对应轨迹假设的对数似然函数。为了保证没有流路径共享同一条边的情况,即保证不会出现轨迹重叠的情况,通常需要流保护约束。如果最小成本流网络 G 中所有边的成本均为正,则最小成本流将是琐碎的空零成本流。但是,对于任何可能是真实的检测 x_i ,对应的检测边 (u_i, v_i) 的成本会为负,发送经过这些负成本边的流,可以使最优成本降至零以下。

网络流图中的成本计算通常是一元化的,但是在通过修改网络流的底层图形结构,理论上可以实现对象之间的多种

成本的合并,但是在实际应用中难以实现^[65];具有成对成本的网路流框架利用外观和运动信息处理跟踪过程中发生的身份切换,可以实现对丢失检测的恢复^[66];通过离散优化和连续优化之间的切换,可以获得最小化的成本函数,但是由于存在两个优化级别,因此难以保证理论或经验上的最优化^[67-68]。为了得到最优的经验保证,Chari 等提出了凸松弛方法^[69];Butt 等认为合并高阶跟踪平滑度约束能提高网络流的性能,但是该过程会产生额外的二进制约束,而通过拉格朗日松弛法能有效解决这个问题^[65];文献[69]通过在网络流图节点中添加成对成本,使基于网络流的多目标跟踪方法能有效应对复杂场景中的遮挡问题^[69];通过优化处理大量对象表示方法并采用连续最短路径算法处理节点,在有限内存和计算能力的条件下,可以有效提高算法的时间效率和空间效率^[70]。

网络流方法适用于离线视频序列的处理,相比 MHT 方法仅考虑一个窗口中的对象信息,网络流方法考虑了更多帧的对象信息,使得跟踪效果有所提升。但是网络流方法中对对象之间的成本计算方式通常是一元化的,使得对多元信息的利用被限制。

3.3 深度学习方法

深度学习算法是一种数据驱动的算法,通过学习某一特定领域样本数据的内在规律和表示层次,以达到某种具体的目的,如理解文字、声音和图像信息等。近年来,基于检测的多目标跟踪问题受益于深度学习方法,取得了显著的进步^[33]。

深度学习算法在基于检测的多目标跟踪问题中的应用大致分为 3 个部分:对象检测部分、重识别部分和对象关联部分,前两部分涉及对象初始化问题。深度排序网络(DeepSort)^[42]和基于置信度数据关联和区分度深度外观学习网络(CDA-DDAL)^[71]从人员重识别任务中学习外观特征,并用来关联检测的对象;递归自回归网络(RAN)^[44]和基于时间窗口的递归神经网络(AMIR)^[72]通过使用自动回归和匹配分类来预测对象的运动和外观特征;而使用多假设跟踪框架并且基于深度学习的 MHT-DAM 方法利用卷积神经网络来学习外观特征^[52];Li 等通过对网络进行离线预训练的方式获得目标特征,将这些特征用于在线方法的数据关联中,并且在跟踪过程中不断调整预训练卷积神经网络,以适应跟踪目标的外观变化^[73];Fan 等设计了一个卷积神经网络,并使用当前帧和前一帧的观测值来估计下一帧中对象的位置和大小^[74]。大多数深度学习方法都采用矩形框中提取的特征来度量外观相似度,当边界框彼此重叠时,辨别力较弱,因此可以基于对象掩码,仅利用真实目标区域中的特征,进行一种深成对的外观相似性度量^[75]。

通过设置网络参数和权重,利用数据来确定网络中的参数和权重,这一步骤被称作训练。一般情况下,对一定数量的训练数据,随着数据量的不断增加,训练效果也会不断提高,而如果数据量超过某个程度,那么训练效果的提升速度会变慢最后趋于稳定^[76]。深度学习方法一般需要两组数据:样本数据和待检测数据。样本数据被用于网络的训练,是一组有真实标记的数据,即一组视频帧中每一帧的对象本身和对象所属的轨迹都是已知的;待检测数据被用于检验算法功能和

性能,通常将检测结果与真实标记信息进行定量分析。

每一个经过训练的用于检测对象的网络对应一种对象初始化方法,不同网络代表了不同的方法。大部分情况下,网络输入的是视频序列,而输出的是对象的符号化表示,一般用多维向量表示: $\bar{\mathbf{X}}=\{x_1, x_2, \cdots, x_n\}$,其中 n 表示向量的维度,根据不同的需求可以变化,通常维度越高跟踪效果越好。如图 5 所示,全连接神经网络包含输入层、隐藏层和输出层,并最终输出二维向量。随着研究者的不断精益求精,网络的规模不断扩大,虽然跟踪效果有所提高,但是随之而来的庞大计算量限制了深度学习的进一步发展和应用。

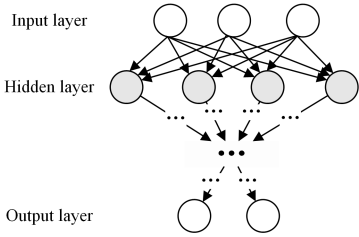


图 5 全连接神经网络

Fig. 5 Fully connected neural network

深度学习算法中对于图像对象的初始化方式经历了一个过程,早期通过建立多尺度图像金字塔,逐帧划窗遍历图像寻找对象^[77];之后通过区域生成网络^[78-82] (Region Proposal Network, RPN)寻找对象, RPN 其实就是通过锚(anchor)的方式,把划窗遍历方式的步长从 1 变成了 $n(n$ 大于 1),并通过改变划动窗口大小的方式来实现对不同尺度对象的寻找,有效减少了寻找视频序列中对象的计算量。

Faster R-CNN 在 RPN 的基础上,利用非极大抑制(NMS)方法取得了较好的效果^[83]。如图 6 所示,从锚的每个位置选取得分最高的参考检测框,并对这个参考检测框进行训练,使原始检测框更接近对象的真实位置。

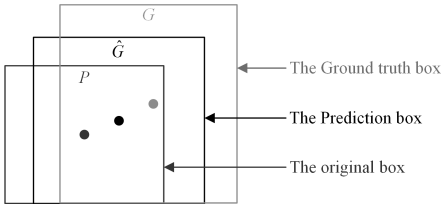


图 6 参考框训练

Fig. 6 Box training

基于区域生成网络(RPN)的方法会产生大量的潜在目标框使正负比例失衡,并且 RPN 中还引入了超参数,使得网络训练过程变得非常复杂,增加了网络的计算量。因此有研究者提出了非锚(anchor-free)方法,如角点网络(Corner-Net)^[84]和中心网络(Center-Net)^[85]。角点网络的核心观点是通过对象的左上角和右上角的两个角点来确定这个对象。由于角点网络只依赖物体的边沿,而没有利用物体内部信息,因此角点网络的全局信息获取能力较弱。针对角点网络中存在的问题,Zhou 等^[85]提出了中心网络,通过局部信息和全局信息共同确定对象的中心点来确定这个对象。Law 等提出了 CornerNet-Lite,优化了 Corner-Net 中使用的卷积网络,提高了结果的准确性和算法性^[86]。

FairMOT^[36]通过中心网络进行对象初始化,把对象初始化

和对对象关联中的结果进行整合,为了公平对待检测和关联模块,提出了端到端的多目标跟踪框架,最终获得了较好的效果,如表 2 所列。表 2 中, MOTA 表示跟踪准确度, IDF1 表示正确识别检测与平均真实数据之比, MT 表示命中轨迹与真实总轨迹数量之比, ML 表示丢失目标与真实总轨迹数量之比, IDs 表示同一对象身份的切换次数, FPS 表示算法每秒处理的帧数。其中,“ $\uparrow$ ”表示数据越大越好,“ $\downarrow$ ”表示数据越小越好。

表 2 FairMOT 的部分实验结果

Table 2 Part of experimental results of FairMOT

Dataset	MOT methods	MOTA $\uparrow$	IDF1 $\uparrow$	MT $\uparrow$
MOT16	DeepSort ^[41]	61.4	62.2	32.8%
	TAP ^[73]	64.8	73.5	38.5%
	CNNMTT ^[42]	65.2	62.2	32.4%
	POI ^[40]	66.1	65.1	34%
	FairMOT	74.9	72.8	44.7%
Dataset	MOT methods	ML $\downarrow$	IDs $\downarrow$	FPS $\uparrow$
MOT16	DeepSort ^[41]	18.2%	781	<6.4
	TAP ^[73]	21.6%	571	<8.0
	CNNMTT ^[42]	21.3%	946	<5.3
	POI ^[40]	20.8%	805	<5.0
	FairMOT	15.9%	1074	25.9

递归-长短记忆网络(RNN-LSTM)^[87]为了同时学习对象初始化、对象与轨迹之间的关联和轨迹的初始化与终止,设计了端到端的深度神经网络;时空注意机制多目标跟踪方法(STAM-MOT)利用空间和时间注意图来处理跟踪中的部分遮挡问题^[88];利用孪生卷积神经网络(SiameseCNN)^[89]和时间约束指示卷积神经网络(CNNTCM)^[90],能够实现检测对象之间和短轨道之间的相似性学习;四重卷积神经网络(Quad-CNN)^[91]考虑了时间顺序等其他信息,以获得更准确的相似性度量;通过联合多歌深度匹配网络(JointMC)^[92]中的深度流特征和解决最小成本提升多割问题实例^[93],提高了网络最优化的性能;深度网络流(DeepNetWork)^[94]用端到端的学习方式对网络流进行全局优化;基于外观模型的递归卷积神经网络(AP-RCNN)中的粒子过滤器利用端到端的深度神经网络对检测的对象、状态的更新、轨道的初始化和终止进行学习^[95]。

除此之外,深度学习方法还能实现对象的像素级分割,完全消除由于方框带来的背景噪音信息的影响^[96];对象初始化时,利用融合信息构造对象模型,能有效实现对象与背景和对象与对象之间的区分^[97];可变形深度卷积神经网络(Deform-CNN)通过学习密集的空间变换,实现了对象检测和语义分段等复杂视觉任务的有效处理^[98];Zhu 等利用特征模仿协助训练 Faster R-CNN,增强了网络的特征分类能力^[99];镜像网络(Ghost-Net)通过少量特征生成器生成典型特征,再通过训练获得转化操作对典型特征生成镜像(ghost)特征,大大减少了网络的计算量^[100];基于深度学习的多视角跟踪能从两种不同视图——顶部和水平,对数据进行获取,最终将多目标跟踪问题转化为联合优化问题^[101]。

基于深度学习的 MOT 方法与其他 MOT 方法相比,其在准确性指标方面具有显著的优势。但是大部分网络的训练过程需要耗费大量的时间资源和计算资源,并且随着具体任务的

变化,网络模型需要不断地进行调整,这同样会消耗大量资源。

3.4 主题发现

主题发现方法^[102]通过将视频序列中检测到的对象视为主题,动态逐帧检测对象集合,并且结合狄利克雷过程混合模型,将跟踪问题转换为主题发现问题,其中对视频序列的处理类似于对文档的处理。这种转换可以解决对象专有性约束之类的与跟踪相关的问题,并且不再需要启发式阈值的轨迹跟踪管理步骤。主题发现问题中,检测对象的外观变化被建模为单词共现的动力学,视频序列用于动态聚类,并通过更新聚类参数来处理视频序列中对象的外观变化。

主题发现通过聚类检测对对象进行关联,其中每个聚类对应一个唯一的对象,而一个对象被表示为一组视觉单词,并且假设这种表示会导致同一对象具有相似的模式,同时能够区分不同的对象。最终,对象对应视频序列中的语义主题,并且跟踪问题被抽象化为对主题的观察问题,主题会随着时间的变化而变化,最终消失。

主题模型通常采用单词、主题和文档的概念,其最早出现在文本分析中^[103]。Blei等提出了一种具有较强泛化性能的主题模型,由于该模型基于狄利克雷过程并且含有隐变量,因此被称为潜在狄利克雷分配模型(Latent Dirichlet Allocation, LDA),该模型的主题分布具有狄利克雷先验的特征^[104]。狄利克雷过程混合模型(Dirichlet Process Mixture Model, DPMM)也被称为无限混合模型,实现了狄利克雷过程的特别应用^[105]。

如图7(a)所示,标准狄利克雷过程混合模型是一个非参数模型,它假定数据是由无数种混合物控制的,而数据只激活其中的少数混合物;而图7(b)是根据时间划分周期得到的动态狄利克雷过程混合模型,其中CNL和ML分别表示不能连接和必须连接两种约束。不能连接约束用于表示每一条轨迹假设在同一时刻不能占据一个以上的检测对象,其公式表示为:

CNL_t = { (x_{t,i}, x_{t,j}) | z_{t,i} ≠ z_{t,j} } (5)

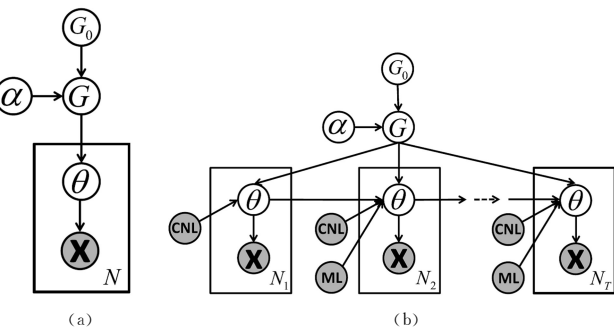


图7 标准狄利克雷过程混合模型和动态狄利克雷过程混合模型
Fig. 7 DPMM and dynamic DPMM

其中, x_{t,i}和 x_{t,j}表示在时间间隔 t 中重叠的轨迹假设,而 z_{t,i}和 z_{t,j}分别表示这两条轨迹假设的聚类成员指示符。而必须连接约束表示每一个检测对象在同一时刻只能被分配给不超过一个的轨迹假设,其公式表示为:

ML_t = { (x_{t,i}, x_{t-1,j}) | z_{t,i} ≠ z_{t-1,j} } (6)

假设标准狄利克雷过程中的第 k 个混合物用 θ_k 表示,那么每个样本将由下式生成:

G | α, G₀ ~ DP(α, G₀) (7)

θ_k | G ~ G (8)

x_i | θ_{Zi} ~ F(θ_{Zi}) (9)

其中, DP 是狄利克雷过程, G₀ 是基础分布, α 是集中参数, G 是从狄利克雷过程中得到的分布,而 θ_k 是从 G 中得到的第 k 个混合物; F(θ_{Zi}) 表示给定 θ_{Zi} 的观测值 x_i 的分布,其中, Zi 是 x_i 的混合表示,应用于聚类时, Zi 是聚类索引。最终的混合物数量由数据确定,即群集数量是自动学习的。

连续狄利克雷过程混合模型之间的关系被建模为一阶马尔可夫过程,除了第一个时期以外,其他每个时期都与其前一个时期紧密相关。假设整个视频序列中存在 N 个小轨迹和 T 个时期,用 x_{1:N} 表示第 t 个时期中的对象,其相应的估计用 θ_{1:N_t} 表示。主题发现转换为求解后验概率^[102]的公式如式(10)所示:

P(θ_{1:N_t} | x_{1:N_t}, α, G₀, CNL, ML) (10)

最终通过求解后验概率解决主题发现问题,可以映射回多目标跟踪问题。

基于狄利克雷过程混合模型的多目标跟踪框架可以不使用检测器;而基于跨相邻帧计算的差异图像或背景的高斯混合模型(GMM),实现检测对象和跟踪对象的结合,不需要任何初始标记,且对象的数量可以未知^[106-107];通过狄利克雷过程来跟踪运动对象,有效地从有噪音数据中得到了典型的运动对象,因此不需要考虑对象的外观建模^[108];对于特定的复杂集群模型,可以不进行初始化和更新,利用颜色直方图来表示对象的外观,通过计算对象之间的相似性实现对复杂小轨迹进行聚类^[109]。

此外,潜在主题模型被用于图像中对目标的分割和分类^[110]。将狄利克雷分配模型与图像中的空间信息相结合,能够实现对图像中目标的分割^[111];在区域分类^[112]、轨迹分析^[113]、图像注释^[114]和图像场景分类^[115]中也有主题模型的应用。

表3列出了主题模型与对象跟踪问题之间的对应关系。在实际应用中,由于对象的外观会发生变化(单词出现的时间动态),并且对象(主题)中视觉单词的分布在整个视频序列中是动态的,因此将视频序列根据时间划分为多个时期,每个时期都利用上述的狄利克雷过程混合模型来进行建模,并且聚类的状态会在各个时期之间进行更新。视频序列中的对象出现和消失问题对应主题发现过程中主题的诞生和死亡。

表3 主题发现任务与多目标跟踪问题之间的对应关系
Table 3 Correspondences between topic discovery and multiple object tracking problem

Multiple target tracking	Topic Discovery
detections	word occurrence
trajectory	topic
video	document
video segment	epoch
exclusivity constrains	cluster membership exclusivity &. cannot link
data association	dynamic clustering

主题发现任务将视频序列从时间上进行分段,通过动力学对对象的外观进行建模,可以用来跟踪刚性和非刚性目标,而对象的动态聚类,可解决跟踪轨迹的管理问题。但是,主题发现任务对非刚性物体(如行人)的跟踪会受到背景的噪声

影响,从而影响跟踪结果。

4 未来展望

表 4 列出了多目标跟踪方法的转化思路:多假设跟踪方法将 MOT 转化为最大后验概率问题和最大加权独立集问题;MHT 方法的优点是可以实现在线跟踪,并对每一个对象都能关注其整个生命周期的具体细节,缺点是在遮挡的情况下会丢失目标,并且如果窗口设置过大,计算量也会非常庞大;网络流图方法将 MOT 转化为全局优化问题,其优点相比 MHT 方法拓展了时间窗口,考虑了更多帧中的对象信息,提高了跟踪准确度,缺点是网络流方法更适合离线跟踪,达不到在线跟踪的要求,并且网络流中的成本计算一元化,多元成本的融合存在难度;深度学习方法通过合理设计网络,利用大量数据进行训练,能够取得较好的跟踪结果,并且能有效处理遮挡问题,缺点是训练过程需要耗费大量时间和计算资源;主题发现利用主题模型,将 MOT 转化为主题发现任务,通过动态聚类和时间分段,解决轨迹的管理问题,缺点是对于非刚性物体(如行人)的跟踪易受背景噪声干扰。

表 4 多目标跟踪方法的转化

Multiple target tracking	problem formulation
Multiple Hypothesis Tracking	Maximum A Posteriori Model and Maximum Weight Independent Set
Network Flow	Global Optimal Solution Problem
Deep Learning	Data-Driven Benchmark
Topic Discovery	Topic Model and Clustering

除了要解决跟踪的准确性问题以外,多目标跟踪算法的时间和空间性能也很重要。离线跟踪方法由于利用了整个视频序列信息,因此其准确性通常高于在线跟踪方法,但处理速度低于在线跟踪方法。多目标跟踪问题是一个复杂的计算机视觉任务,其在效果和性能方面还有较大的提升空间,并且存在巨大的研究潜力,寻找又好又快的方法是研究人员的目标。未来多目标跟踪的对象初始化方法可以从以下角度进行进一步研究。

(1)多目标跟踪问题根据对象初始化方法的不同,可以抽象成为不同的解决框架,而这些转换后的框架仍然有较大的改进空间,如多假设跟踪方法中对于遮挡问题的处理、网络流方法中的多元成本融合问题。

(2)每一种对象初始化方法对应了一种问题的解决思路,而不同的对象初始化方法各有优劣。本文从 4 个方面综述了多目标跟踪的不同对象初始化方法和问题转化,从经验上来说,仍然存在新的对象初始化方法,因此未来可以探索新的对象初始化方法,即新的多目标跟踪问题框架。

(3)近年来,深度学习方法在计算机视觉任务中被广泛关注和应用。基于深度学习的多目标跟踪算法在准确性上超过了传统方法,并且能够处理对象丢失和遮挡问题,但是该网络的训练和调整会耗费大量的计算资源,对于某些极端场景或者网络本身庞大的规模,其训练和调整过程会耗费几天甚至一个月的时间,因此基于深度学习的 MOT 框架可以从网络规模、时间与准确性的平衡方面进行改进和优化。

(4)除了单摄像头跟踪以外,多摄像头、多视角模式的

多目标跟踪算法也有较大潜力,能弥补单摄像头对目标外观和运动模型信息缺失的缺点。直观上,多摄像头是利用多个单摄像头对目标进行跟踪,但是,实际上除了利用摄像头以外,还可以利用其他信息采集器采集目标信息。当前已有的多摄像头下的多目标跟踪方法没有统一的标准,因此未来对于多目标跟踪问题的研究 可以结合多摄像头和其他信息采集器,如利用各种传感器来进行信息辅助,使其在多目标跟踪问题的对象初始化中获得更多、更全面的信息。

结束语 对象初始化方法不同,处理多目标跟踪问题的思路也不同,本文从多假设跟踪、网络流方法、深度学习方法和主题发现 4 个方面综述了多目标跟踪的对象初始化方法。多目标跟踪根据不同的对象初始化方法被建模为不同的问题框架,每个框架下的任务转换和对对象映射各有特点,而处理流程的不同也使得这些框架各有优劣,合理转换多目标跟踪中的对象模型是这些框架的共同点。当前多目标跟踪准确性仍然存在较大的提升空间,而相对准确的跟踪方法又需要大量的时间和计算成本,因此需要实现效果和性能的平衡。未来可以根据现有的问题框架,优化或改进这些框架中未解决的问题;此外,也可以根据多目标跟踪问题的特点,提出新的对象初始化方法,从而建立新的问题框架。

参 考 文 献

[1] LEAL-TAIXÉ L,MILAN A,REID I,et al. MOT Challenge 2015:Towards a Benchmark for Multi-Target Tracking[EB/OL]. (2015-04-01)[2020-11-27]. <https://ui.adsabs.harvard.edu/abs/2015arXiv150401942L>.

[2] KHALKHALI M B,VAHEDIAN A,YAZDI H S. Multi-Target State Estimation Using Interactive Kalman Filter for Multi-Vehicle Tracking[J]. IEEE Transactions on Intelligent Transportation Systems,2020,21(3):1131-1144.

[3] LUO Z Z,ATTARI M,HABIBI S,et al. Online Multiple Maneuvering Vehicle Tracking System Based on Multi-Model Smooth Variable Structure Filter[J]. IEEE Transactions on Intelligent Transportation Systems,2020,21(2):603-616.

[4] DU Y,WEI S G,CHAI L G. Particle Filter Based Object Tracking of 3D Sparse Point Clouds for Autopilot[C]//2018 Chinese Automation Congress. 2018:1102-1107.

[5] LIU Y,ZHANG W X,YANG Y,et al. RAMTEL:Robust Acoustic Motion Tracking Using Extreme Learning Machine for Smart Cities[J]. IEEE Internet Things,2019,6(5):7555-7569.

[6] BLOISI D,IOCCHI L. Argos-a Video Surveillance System for Boat Traffic Monitoring in Venice[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23 (7): 1477-1502.

[7] NANAWARE V S,NERKAR M H,PATIL C M. A Review of the Detection Methodologies of Multiple Human Tracking & Action Recognition in a Real Time Video Surveillance[C]//2017 IEEE International Conference on Power,Control,Signals and Instrumentation Engineering. 2017:2484-2489.

[8] ZHOU L F,TOKEKAR P. Active Target Tracking With Self-Triggered Communications in Multi-Robot Teams[J]. IEEE

- Transactions on Automation Science and Engineering, 2019, 16(3):1085-1096.
- [9] VALDIVIA-GRANDA W A. Biosurveillance enterprise for operational awareness, a genomic-based approach for tracking pathogen virulence[J]. Virulence, 2013, 4(8):745-751.
 - [10] SHAH P, FAZA A, NIMMALA R, et al. Infrared and Inertial Tracking in the Immersive Audio Environment for Enhanced Military Training[C]//2012 IEEE International Conference on Multimedia and Expo Workshops. 2012:181-186.
 - [11] SENG K Y, CHEN Y, CHAI K M A, et al. Tracking Body Core Temperature in Military Thermal Environments: An Extended Kalman Filter Approach[C]//IEEE International Conference on Wearable and Implantable Body Sensor Networks. 2016:296-299.
 - [12] ZHANG J M, PRESTI L L, SCLAROFF S. Online Multi-Person Tracking by Tracker Hierarchy[C]//2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance. 2012:379-385.
 - [13] RISTANI E, TOMASI C. Tracking Multiple People Online and in Real Time[J]. Lecture Notes in Computer Science, 2015, 9007:444-459.
 - [14] QIN Z, SHELTON C R. Improving Multi-target Tracking via Social Grouping[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2012:1972-1978.
 - [15] BOSE B, WANG X, GRIMSON E. Multi-class object tracking algorithm that handles fragmentation and grouping[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2007:1567-1574.
 - [16] CHOI W G. Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor[C]//IEEE Conference on Computer Vision. 2015:3029-3037.
 - [17] SONG J, CHO H, YOON S M. Target Object Tracking-Based 3D Object Reconstruction in a Multiple Camera Environment in Real Time[C]//Asian Conference on Intelligent Information and Database Systems. 2017.
 - [18] TESFAYE Y T, ZEMENE E, PRATI A, et al. Multi-target Tracking in Multiple Non-overlapping Cameras Using Fast-Constrained Dominant Sets[J]. International Journal of Computer Vision, 2019, 127(9):1303-1320.
 - [19] WU B, NEVATIA R. Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors[J]. International Journal of Computer Vision, 2007, 75(2):247-266.
 - [20] SVENSSON D, ULMKE M, DANIELSSON L. Joint probabilistic data association filter for partially unresolved target groups [C]//Information Fusion. 2011.
 - [21] LEAL-TAIXÉ L, MILAN A, SCHINDLER K, et al. Tracking the Trackers: An Analysis of the State of the Art in Multiple Object Tracking[EB/OL]. (2017-04-01)[2020-11-27]. <https://ui.adsabs.harvard.edu/abs/2017arXiv170402781L>.
 - [22] DENDORFER P, REZATOFIGHI H, MILAN A, et al. CVPR19 Tracking and Detection Challenge: How crowded can it get[EB/OL]. (2019-06-01)[2020-11-27]. <https://ui.adsabs.harvard.edu/abs/2019arXiv190604567D>.
 - [23] SUGIMURA D, KITANI K M, OKABE T, et al. Using Individuality to Track Individuals: Clustering Individual Trajectories in Crowds using Local Appearance and Frequency Trait[C]//IEEE International Conference on Computer Vision. 2009:1467-1474.
 - [24] YOON J H, LEE C R, YANG M H, et al. Structural Constraint Data Association for Online Multi-object Tracking[J]. International Journal of Computer Vision, 2019, 127(1):1-21.
 - [25] KUO C H, HUANG C, NEVATIA R. Inter-camera Association of Multi-target Tracks by On-Line Learned Appearance Affinity Models[J]. Lecture Notes in Computer Science, 2010, 6311:383-396.
 - [26] NITTI D, CHLIVEROS G, PATERAKI M, et al. Application of Dynamic Distributional Clauses for Multi-hypothesis Initialization in Model-based Object Tracking[C]//Proceedings of the 2014 9th International Conference on Computer Vision, Theory and Applications. 2014:256-261.
 - [27] LUO W, XING J, ZHANG X, et al. Multiple Object Tracking: A Literature Review [EB/OL]. (2017-05-22) [2020-11-27]. <https://arxiv.org/abs/1409.7618>.
 - [28] XU Y K, ZHOU X L, CHEN S Y, et al. Deep learning for multiple object tracking: a survey[J]. IET Computer Vision, 2019, 13(4):355-368.
 - [29] OH S, RUSSELL S, SASTRY S. Markov Chain Monte Carlo Data Association for Multi-Target Tracking[J]. IEEE Transactions on Automatic Control, 2009, 54(3):481-497.
 - [30] FAN L T, WANG Z L, CAI B G, et al. A Survey on Multiple Object Tracking Algorithm[C]//2016 IEEE International Conference on Information and Automation. 2016:1855-1862.
 - [31] BAE S H. Survey of amplitude-aided multi-target tracking methods[J]. IET Radar Sonar and Navigation, 2019, 13(2):243-253.
 - [32] FENG X Y, MEI W, HU D S. A Review of Visual Tracking with Deep Learning[C]//International Conference on Artificial Intelligence and Industrial Engineering. 2016.
 - [33] CIAPARRONE G, SÁNCHEZ F L, TABIK S, et al. Deep Learning in Video Multi-Object Tracking: A Survey 2019[EB/OL]. (2019-06-01)[2020-11-27]. <https://ui.adsabs.harvard.edu/abs/2019arXiv190712740C>.
 - [34] CAMPLANI M, PAIEMENT A, MIRMEHDI M, et al. Multiple human tracking in RGB-depth data: a survey[J]. IET Computer Vision, 2017, 11(4):265-285.
 - [35] ATH G D, EVERSON R. Visual Object Tracking: The Initialisation Problem[C]//2018 15th Conference on Computer and Robot Vision. 2018:142-149.
 - [36] ULLAH M, CHEIKH F A. A Directed Sparse Graphical Model for Multi-Target Tracking[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2018:1897-1904.
 - [37] KIERITZ H, HUBNER W, ARENS M. Joint detection and on-line multi-object tracking[C]//IEEE Computer Society Conference. 2018:1540-1548.
 - [38] YU F W, LI W B, LI Q Q, et al. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature[J].

- Lecture Notes in Computer Science, 2016, 9914: 36-42.
- [39] WOJKE N, BEWLEY A, PAULUS D. Simple Online and Real-time Tracking with a Deep Association Metric[J]. IEEE Transactions on Image Processing, 2017, 3645-3649.
 - [40] MAHMOUDI N, AHADI S M, RAHMATI M. Multi-target tracking using CNN-based features: CNNMTT [J]. Multimed Tools Application, 2019, 78(6): 7077-7096.
 - [41] ZHANG Y, WANG C, WANG X, et al. FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking2020[EB/OL]. (2020-04-01) [2020-11-27]. <https://ui.adsabs.harvard.edu/abs/2020arXiv200401888Z>.
 - [42] WANG Y, WENG X, KITANI K. Joint Detection and Multi-Object Tracking with Graph Neural Networks[EB/OL]. (2020-06-23) [2020-11-27]. <https://arxiv.org/abs/2006.13164v2>.
 - [43] WAN X, CAO J, ZHOU S, et al. End-to-End Multi-Object Tracking with Global Response Map2020[EB/OL]. (2020-07-01) [2020-11-27]. <https://ui.adsabs.harvard.edu/abs/2020arXiv200706344W>.
 - [44] FANG K, XIANG Y, LI X C, et al. Recurrent Autoregressive Networks for Online Multi-Object Tracking[C] // 2018 IEEE Winter Conference on Applications of Computer Vision. 2018: 466-475.
 - [45] REID D. An Algorithm for Tracking Multiple Targets[J]. IEEE Transaction on Automatics Control, 1979, 24(6): 1202-1211.
 - [46] HAN M, XU W, TAO H, et al. An algorithm for multiple object trajectory tracking[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004: 864-871.
 - [47] PAPAGEORGIOU D J, SAPUKAS M R. The Maximum Weight Independent Set Problem for Data Association in Multiple Hypothesis Tracking[J]. Lecture Notes in Control and Information Sciences, 2009, 381: 235-255.
 - [48] BRENDDEL W, AMER M, TODOROVIC S. Multiobject Tracking as Maximum Weight Independent Set [C] // 2011 IEEE Conference on Computer Vision and Pattern Recognition. 2011: 1273-1280.
 - [49] SHENG H, CHEN J H, ZHANG Y, et al. Iterative Multiple Hypothesis Tracking With Tracklet-Level Association[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(12): 3660-3672.
 - [50] KIM C, LI F X, CIPTADI A, et al. Multiple Hypothesis Tracking Revisited[C] // IEEE International Conference on Computer Vision. 2015: 4696-4704.
 - [51] KOJIMA M, KAMEDA H, TSUJIMICHI S, et al. A study of target tracking using track-oriented multiple hypothesis tracking [C] // Sice '98-Proceedings of the 37th Sice Annual Conference. 1998: 933-938.
 - [52] YOO H, KIM K, BYEON M, et al. Online Scheme for Multiple Camera Multiple Target Tracking Based on Multiple Hypothesis Tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(3): 454-469.
 - [53] AHMADI K, SALARI E. A novel Multiple Hypothesis Testing (MHT) scheme for tracking of dim objects[C] // IEEE International Conference on Electro/information Technology. 2015.
 - [54] WANG X C, TURETKEN E, FLEURET F, et al. Tracking Interacting Objects Using Intertwined Flows[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11): 2312-2326.
 - [55] PIRSAIVASH H, RAMANAN D, FOWLKES C C. Globally-Optimal Greedy Algorithms for Tracking a Variable Number of Objects[C] // 2011 IEEE Conference on Computer Vision and Pattern Recognition. 2011: 1201-1208.
 - [56] TANG S Y, ANDRES B, ANDRILUKA M, et al. Subgraph Decomposition for Multi-Target Tracking[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5033-5041.
 - [57] DEGHAN A, ASSARI S M, SHAH M. GMMCP Tracker: Globally Optimal Generalized Maximum Multi Clique Problem for Multiple Object Tracking[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4091-4099.
 - [58] XU Y K, QIN L, HUANG Q M. Coupling Reranking and Structured Output SVM Co-Train for Multitarget Tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(6): 1084-1098.
 - [59] KUMAR R, CHARPIAT G, THONNAT M. Multiple Object Tracking by Efficient Graph Partitioning [C] // Asian Conference on Computer Vision. 2014.
 - [60] ZHANG L, LI Y, NEVATIA R. Global data association for multi-object tracking using network flows [C] // 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2008: 1881-1888.
 - [61] BERCLAZ J, FLEURET F, TURETKEN E, et al. Multiple Object Tracking Using K-Shortest Paths Optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(9): 1806-1819.
 - [62] UZUN C, DEMIREKLER M. Stochastic Dynamic Programming Based Resource Allocation for Multi Target Tracking for Electronically Steered Antenna Radar[C] // 2015 23rd Signal Processing and Communications Applications Conference. 2015: 867-870.
 - [63] HUANG C, WU B, NEVATIA R. Robust Object Tracking by Hierarchical Association of Detection Responses[C] // European Conference on Computer Vision. 2008: 788-801.
 - [64] KUO C H, HUANG C, NEVATIA R. Multi-Target Tracking by On-Line Learned Discriminative Appearance Models [C] // 2010 IEEE Conference on Computer Vision and Pattern Recognition. 2010: 685-692.
 - [65] BUTT A A, COLLINS R T. Multi-target Tracking by Lagrangian Relaxation to Min-Cost Network Flow[C] // IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1846-1853.
 - [66] WANG B, WANG G, CHAN K L, et al. Tracklet Association by Online Target-Specific Metric Learning and Coherent Dynamics Estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(3): 589-602.
 - [67] MILAN A, ROTH S, SCHINDLER K. Continuous Energy Minimization for Multitarget Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1): 58-72.
 - [68] MILAN A, SCHINDLER K, ROTH S. Detection and Trajectory

- ry-Level Exclusion in Multiple Object Tracking[C]// Computer Vision and Pattern Recognition. 2013.
- [69] CHARI V, LACOSTE-JULIEN S, LAPTEV I, et al. On Pairwise Costs for Network Flow Multi-Object Tracking[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015;5537-5545.
- [70] LENZ P, GEIGER A, URTASUN R. FollowMe: Efficient On-line Min-Cost Flow Tracking with Bounded Memory and Computation[C]// IEEE International Conference on Computer Vision. 2015;4364-4372.
- [71] BAE S H, YOON K J. Confidence-Based Data Association and Discriminative Deep Appearance Learning for Robust Online Multi-Object Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3):595-610.
- [72] SADEGHIAN A, ALAHI A, SAVARESE S. Tracking The Untrackable: Learning to Track Multiple Cues with Long-Term Dependencies[C]// 2017 IEEE International Conference on Computer Vision. 2017;300-311.
- [73] LI H X, LI Y, PORIKLI F. DeepTrack: Learning Discriminative Feature Representations Online for Robust Visual Tracking[J]. IEEE Transactions on Image Processing, 2016, 25(4):1834-1848.
- [74] FAN J L, XU W, WU Y, et al. Human Tracking Using Convolutional Neural Networks[J]. IEEE Transactions on Neural Networks, 2010, 21(10):1610-1623.
- [75] ZHOU Z W, XING J L, ZHANG M D, et al. Online Multi-Target Tracking with Tensor-Based High-Order Graph Matching[C]// 2018 24th International Conference on Pattern Recognition (ICPR). 2018;1809-1814.
- [76] ZENG X Y, OUYANG W L, WANG M, et al. Deep Learning of Scene-Specific Classifier for Pedestrian Detection[J]. Lecture Notes in Computer Science, 2014, 8691:472-487.
- [77] BAE S H, YOON K J. Robust Online Multi-Object Tracking based on Tracklet Confidence and Online Discriminative Appearance Learning[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2014;1218-1225.
- [78] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2014;580-587.
- [79] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.
- [80] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017;2980-2988.
- [81] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2017;6517-6525.
- [82] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[EB/OL]. (2018-04-08)[2020-11-27]. <https://arxiv.org/abs/1804.02767>.
- [83] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS Improving Object Detection With One Line of Code[C]// IEEE International Conference on Computer Vision. 2017;5562-5570.
- [84] LAW H, DENG J. CornerNet: Detecting Objects as Paired Key-points[J]. International Journal of Computer Vision, 2020;642-656.
- [85] ZHOU X, WANG D, KRHENBÜHIP. Objects as Points[EB/OL]. (2019-04-16)[2020-11-27]. <https://arxiv.org/abs/1904.07850>.
- [86] LAW H, TENG Y, RUSSAKOVSKY O, et al. CornerNet-Lite: Efficient Keypoint Based Object Detection[EB/OL]. (2019-04-18)[2020-11-27]. <https://arxiv.org/abs/1904.08900v2>.
- [87] MILAN A, REZATOFIGHI S H, DICK A, et al. Online Multi-Target Tracking Using Recurrent Neural Networks[C]// Thirty-First Aaai Conference on Artificial Intelligence. 2017;4225-4232.
- [88] QI C, OUYANG W L, LI H S, et al. Online Multi-Object Tracking Using CNN-based Single Object Tracker with Spatial-Temporal Attention Mechanism[C]// IEEE International Conference on Computer Vision. 2017;4846-4855.
- [89] LEAL-TAIXE L, CANTON-FERRER C, SCHINDLER K. Learning by tracking: Siamese CNN for robust target association[C]// IEEE Computer Society Conference. 2016;418-425.
- [90] WANG B, WANG L, SHUAI B, et al. Joint Learning of Convolutional Neural Networks and Temporally Constrained Metrics for Tracklet Association[C]// IEEE Computer Society Conference. 2016;386-393.
- [91] SON J, BAEK M, CHO M, et al. Multi-Object Tracking with Quadruplet Convolutional Neural Networks[C]// 30th IEEE Conference on Computer Vision and Pattern Recognition. 2017;3786-3795.
- [92] TANG S Y, ANDRES B, ANDRILUKA M, et al. Multi-person Tracking by Multicut and Deep Matching[C]// 2016 European Conference on Computer Vision Workshops. 2016;100-111.
- [93] TANG S Y, ANDRILUKA M, ANDRES B, et al. Multiple People Tracking by Lifted Multicut and Person Re-identification[C]// 30th IEEE Conference on Computer Vision and Pattern Recognition. 2017;3701-3710.
- [94] SCHULTER S, VERNAZA P, CHOI W, et al. Deep Network Flow for Multi-Object Tracking[C]// 30th IEEE Conference on Computer Vision and Pattern Recognition. 2017;2730-2739.
- [95] CHEN L, AI H Z, SHANG C, et al. Online Multi-Object Tracking with Convolutional Neural Networks[C]// IEEE International Conference on Image Processing. 2017;645-649.
- [96] TIAN Y C, DEGHAN A, SHAH M. On Detection, Data Association and Segmentation for Multi-Target Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9):2146-2160.
- [97] CHU P, FAN H, TAN C C, et al. Online Multi-Object Tracking With Instance-Aware Tracker and Dynamic Model Refreshment[C]// IEEE Winter Conference on Applications of Computer Vision. 2019.
- [98] DAI J F, QI H Z, XIONG Y W, et al. Deformable Convolutional Networks[C]// IEEE International Conference on Computer Vision. 2017;764-773.

[99] ZHU X Z, HU H, LIN S, et al. Deformable ConvNets v2: More Deformable, Better Results[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (Cvpr 2019). 2019:9300-9308.

[100] HAN K, WANG Y, TIAN Q, et al. GhostNet: More Features From Cheap Operations[C]// CVF Conference on Computer Vision and Pattern Recognition. 2020.

[101] HAN R, FENG W, ZHAO J, et al. Complementary-View Multiple Human Tracking[C]// Thirty-Fourth AAAI Conference on Artificial Intelligence. 2020.

[102] LUO W H, STENGER B, ZHAO X W, et al. Trajectories as Topics: Multi-Object Tracking by Topic Discovery[J]. IEEE Transactions on Image Processing, 2019, 28(1): 240-252.

[103] PAPADIMITRIOU C H, RAGHAVAN P, TAMAKI H, et al. Latent semantic indexing: A probabilistic analysis[J]. Journal of Computer and System Sciences, 1998, 61(2): 217-235.

[104] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation [C]// Advances in Neural Information Processing Systems. 2001.

[105] JBABDI S, WOOLRICH M W, BEHRENS T E J. Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models[J]. Neuroimage, 2009, 44(2): 373-384.

[106] TOPKAYA I S, ERDOGAN H, PORIKLI F. Detecting and Tracking Unknown Number of Objects with Dirichlet Process Mixture Models and Markov Random Fields[J]. Advances in Visual Computing, 2013, 8034: 178-188.

[107] NEISWANGER W, WOOD F, XING E. The Dependent Dirichlet Process Mixture of Objects for Detection-free Tracking and Object Modeling[J]. Journal of Machine Learning Research, 2014, 33: 660-668.

[108] FOX E, SUDDERTH E B, WILLSKY A S. Hierarchical Dirichlet processes for tracking maneuvering targets[C]// 2007 Proceedings of the 10th International Conference on Information Fusion. 2007.

[109] TOPKAYA I S, ERDOGAN H, PORIKLI F. Tracklet clustering for robust multiple object tracking using distance dependent Chinese restaurant processes [J]. Signal Image and Video

Processing, 2015, 10(5): 795-802.

[110] CAO L L, LI F F. Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes [C]// 2007 IEEE 11th International Conference on Computer Vision. 2007: 1080-1087.

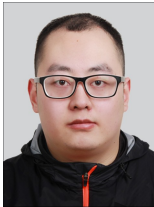
[111] WANG X, GRIMSON E. Spatial Latent Dirichlet Allocation [C]// Conference on Advances in Neural Information Processing Systems. 2007.

[112] VERBEEK J, TRIGGS W. Region Classification with Markov Field Aspect Models[C]// 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2007.

[113] WANG X G, MA K T, NG G W, et al. Trajectory Analysis and Semantic Region Modeling Using Nonparametric Hierarchical Bayesian Models[J]. International Journal of Computer Vision, 2011, 95(3): 287-312.

[114] WANG C, BLEI D, LI F F. Simultaneous Image Classification and Annotation[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009.

[115] FEI-FEI L, PERONA P. A Bayesian hierarchical model for learning natural scene categories[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005: 524-531.



WEN Cheng-yu, born in 1995, postgraduate. His main research interests include machine learning and image processing.



CHEN Wei, born in 1978, Ph.D, professor, is a member of IEEE. His main research interests include machine learning, image processing, and computer networks.

(责任编辑:柯颖)