

融合更新机制的图神经网络模型在多目标跟踪中的应用研究

张鹏¹, 董宇轩²

(1. 烟台科技学院 智能工程学院; 2. 烟台科技学院 艺术设计学院, 山东 烟台 265600)

摘要: 针对现有多目标跟踪算法大多只考虑特征本身, 很少考虑帧内其他特征以及帧间特征交互的问题, 提出两种方法使特征更具有分辨性, 以实现特征间的交互, 增强特征表达, 获得更加可靠的相似度矩阵, 提高目标关联的准确度。方法一为在特征处理时引入图神经网络, 分别构建外观图神经网络和运动图神经网络, 两个网络模型中的特征相互融合、互为补充, 增强目标特征的表达, 提高特征分辨性; 方法二为在图神经网络中引入更新机制。以往图神经网络多为静态图, 不能随着应用而动态更新, 为此提出一种使图神经网络中的边模块、点模块以及全局变量模块动态更新的策略, 其中全局变量可捕获图网络中特征的全局关系, 以更好地完成目标匹配。理论分析和跟踪算法仿真结果表明, 在MOT16和MOT17两个基准测试集中, 所提多目标跟踪算法的准确度指标(MOTA)分别为49.5%和53.9%。同时通过消融实验对图神经网络的3种更新模块进行性能比较, 发现拥有完整结构模块的图神经网络比不包含点特征聚合模块和不包含全局变量模块在MOTA指标上分别提高1.9%和0.2%, 说明基于图神经网络的多目标跟踪算法在实际跟踪任务中具有较好表现。

关键词: 计算机视觉; 图卷积神经网络; 多目标跟踪; 数据关联; 外观模型; 运动模型

DOI: 10.11907/rjtk.221806

开放科学(资源服务)标识码(OSID):



中图分类号: TP399

文献标识码: A

文章编号: 1672-7800(2023)008-0059-07

Research on the Application of Graph Neural Network Model with Update Mechanism in Multiple Object Tracking

ZHANG Peng¹, DONG Yuxuan²

(1. College of Intelligent Engineering, Yantai Institute of Science and Technology;
2. College of Art and Design, Yantai Institute of Science and Technology, Yantai 265600, China)

Abstract: Most existing multi-target tracking algorithms only consider the features themselves and seldom consider other features within frames and the interaction of features between frames, two methods are proposed to make the features more discriminative in order to achieve the interaction between features, enhance the feature representation, obtain a more reliable similarity matrix, and improve the accuracy of target association. The first method is to introduce graph neural network in feature processing, and build appearance graph neural network and motion graph neural network respectively, the features in the two network models fuse and complement each other to enhance the expression of target features and improve the feature discriminative property; the second method is to introduce update mechanism in graph neural network. Previously, graph neural networks are mostly static graphs, which cannot be dynamically updated with applications. To this end, a strategy is proposed to make the edge module, point module and global variable module in the graph neural network dynamically updated, where the global variable can capture the global relationship of features in the graph network to better accomplish target matching. Theoretical analysis and tracking algorithm simulation results show that the accuracy metrics (MOTA) of the proposed multi-target tracking algorithm are 49.5% and 53.9% in two benchmark test sets, MOT16 and MOT17. Meanwhile, the performance of the three update modules of the graph neural network is compared by ablation experiments, and it is found that the graph neural network with the complete structure module improves the MOTA index by 1.9% and 0.2% than the module without the point feature aggregation module and the module without the global variables, respectively, indicating that the multi-target tracking algorithm based on the graph neural network has better performance in the actual tracking task.

Key Words: computer vision; graph neural network; multi-object tracking; data association; appearance model; motion model

收稿日期: 2022-07-13

作者简介: 张鹏(1991-), 男, 硕士, CCF会员, 烟台科技学院智能工程学院讲师, 研究方向为计算机视觉、图像处理、软件工程; 董宇轩(1990-), 女, 硕士, 烟台科技学院艺术设计学院讲师, 研究方向为视觉设计。

0 引言

多目标跟踪技术在自动驾驶、智能监控等领域应用广泛^[1-2]。然而目标间相似的外观、相互遮挡以及外观突然改变等因素均会导致跟踪任务失败。目前,多目标跟踪技术主要基于“通过检测进行跟踪”的策略进行^[3],分为目标检测、特征提取和数据关联3个子任务:首先将目标检测器^[4-5]应用于所有帧中,对视频每一帧的目标物体进行检测;然后独立地获取每一帧中检测到的每一个目标物体的特征,并计算帧与帧之间成对目标的相似度;最后通过匈牙利算法进行目标关联^[6]。多目标跟踪任务成功的关键在于从不同目标身上学习到具有分辨性的特征以提高区分度。

1 相关研究

目前,许多学者对多目标跟踪技术进行了研究。例如,Yu等^[7]基于Faster R-CNN设计检测器,在ETHZ行人数据集、Caltech行人数据集和自制监控数据集进行训练,最终获得了一个高性能检测器来执行多目标跟踪任务;Bewley等^[8]提出SORT(Simple Online and Realtime Tracking)算法,使用卡尔曼滤波跟踪目标,通过计算帧间边界框相似度的最大交并比关联检测目标;Peng等^[9]提出的Chained Tracker算法将目标检测、特征提取和数据关联3个子任务整合成一个端到端的解决方案,将重叠节点估计的成对边界框回归结果链接起来,其中每个节点覆盖两个相邻的帧,该算法简单、快速、有效,无需依赖任何额外训练数据;Feichtenhofer等^[10]提出可同时进行检测和跟踪的ConvNet架构,使用多任务目标进行基于帧的物体检测和跨帧跟踪回归,通过引入代表物体跨时间共现的相关特征,将基于跨帧跟踪的帧级检测联系起来,在视频层面产生高精度检测。

检测器质量和具有分辨性的特征学习这两个因素对多目标跟踪算法的性能影响最大。在众多特征中,运动特征和外观特征被证实是具有分辨性的特征。早期研究一般采用手工特征作为运动特征,如空间距离和交并比^[11]等;使用颜色直方图^[12]作为外观特征,目前研究常采用卷积神经网络(Convolutional Neural Network, CNN)提取外观特征。然而很多研究工作并没有考虑所有跟踪目标与检测目标的交互^[13],即只是从每一个物体中独立获得特征,而未考虑物体之间的特征。这种独立的特征提取并不是获得具有分辨性特征的最优解,因为在多目标跟踪任务中,不同物体间的特征相似度具有一定的依赖性,即在当前帧的物体应能最多匹配一个前一帧的物体。为解决这个问题,图神经网络模型被引入到多目标跟踪模型中^[14-15],并取得了不错效果。图卷积神经网络使用神经网络

处理基于图结构的数据,即将目标与检测作为图网络的点,目标与检测之间的关联性作为图网络的边。例如,Shen等^[16]提出通过图网络解决行人重识别问题,使用图网络表示每一对待认证图片与其所属图片库的相似度;Braso等^[17]将特征信息传入到消息传递网络中进行解析,并利用图网络中的边对检测对象进行分类关联;Dai等^[18]提出的多目标跟踪算法首先对每一帧中的目标生成候选框,然后导入图网络中进行精炼并输出可靠的质量分数,以完成精确的目标关联。

本文在以上研究基础上设计了一种新的基于多目标跟踪的图神经网络更新机制,在完成点聚合和边更新的同时引入全局变量模块,以保存点与边变化的全局信息,增强特征的稳定性和分辨性。然而,如果只是直接将图神经网络应用于多目标跟踪任务中,而不对图网络进行恰当设计,则很有可能导致多目标跟踪任务失败^[19-20]。前文提到,在处理多目标跟踪任务时,外观特征和运动特征均为必要因素,不能仅依靠某一种。当跟踪目标间外观具有很强的相似性、跟踪目标外观突然改变或被频繁遮挡时,基于运动特征的跟踪器比基于外观特征的跟踪器表现更优;而当摄像机出现大幅度移动,或由于检测器不准确而导致错误运动估计时,基于外观特征的模型比基于运动特征的模型性能更优。为使图网络中的特征点更加稳定,本文设计外观图网络和运动图网络两种网络,使每一个节点分别拥有外观属性和运动属性。同样的,每一个节点的外观特征和运动特征也需要进行交互融合,以增强物体特征的分辨度。在图神经网络的每一层,每一个节点可以通过与其他点聚合进行更新,这种点聚合的更新机制使得点之间不再孤立,并产生相互关联。经过几层图神经网络后,计算获得的相似矩阵会比未经过特征融合的模式更加具有分辨性。

2 多目标跟踪算法

多目标跟踪任务是将在之前帧中通过跟踪算法已经检测到并成功确定身份的目标与当前帧中新检测到但未确认身份的检测进行匹配关联,从而形成一个连续帧之间的跟踪轨迹。目标为已经在跟踪轨迹中成功确定身份的对象,轨迹中的所有目标用 O_i 表示;检测为在当前帧 t 被检测到但没有识别身份的边界框,检测集用 D_t 表示。在一次完整的跟踪任务中,第 t 帧共有 M 个已经确定身份的目标,且 $i \in \{1, 2, \dots, M\}$, $o_i \in O$,其中 i 表示每一个已确定身份的跟踪目标。同样的,在 $t+1$ 帧中共检测到 N 个未确定身份的对象,且在第 $t+1$ 帧中 $j \in \{1, 2, \dots, N\}$, $d_j \in D$,其中 j 表示当前帧中每一个检测到的待确定身份的对象。多目标跟踪算法的主要任务是从 O 和 D 中形成的所有边界框中学习具有分辨性的特征,然后根据所有目标和检测的特征相似度对当前帧中所有待确定身份的检测与前一帧已确

定身份的目标进行身份识别以及对应的匹配关联,实现 M 到 N 中对应身份的识别与关联,从而完成相邻帧之间的跟踪任务。

2.1 算法流程

本文算法分为特征提取、图神经网络构建、数据关联以及跟踪目标管理策略 4 个阶段,具体流程见图 1。算法输入为第 t 帧和第 t+1 帧,首先经过两个分支,第一个分支

通过 CNN 提取外观特征,第二个分支通过长短期记忆网络运动模型(Long Short-Term Memory, LSTM)提取运动特征。提取后的特征分别进入外观神经网络和运动神经网络中,通过图神经网络的点特征聚合、边回归等模块增强特征间的交互,获得可靠的相似度矩阵。根据输出的相似度矩阵,使用匈牙利算法进行数据关联、实现多目标跟踪。

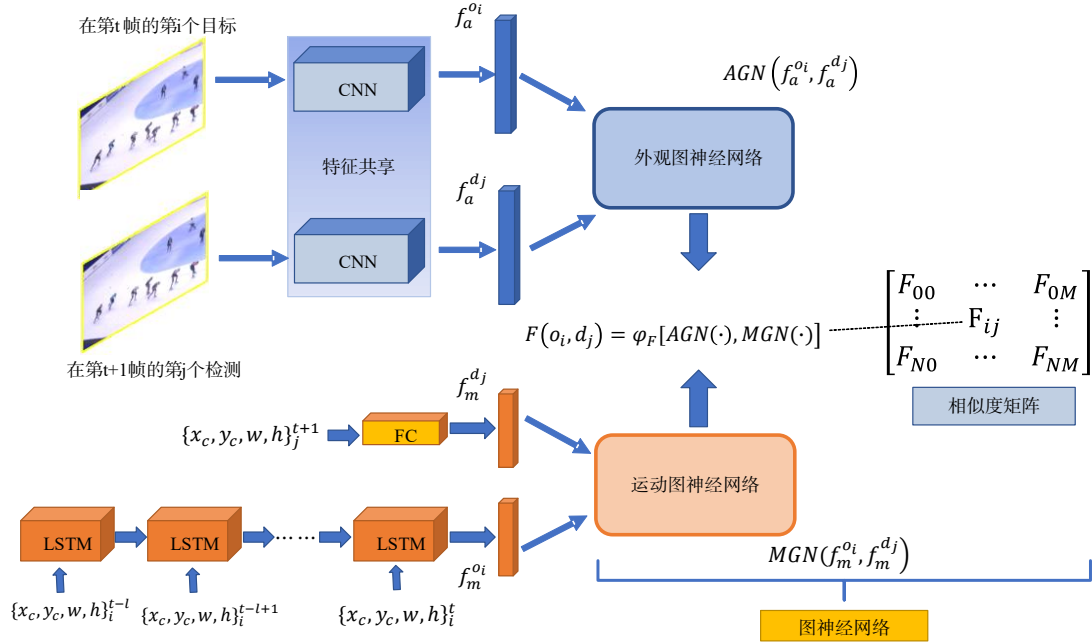


Fig. 1 Multiple object tracking algorithm flow based on the graph neural network

图 1 基于图神经网络的多目标跟踪算法流程

(1)特征提取阶段。该阶段从目标物体和检测物体中分别提取外观特征和运动特征。输入到 LSTM 网络中的参数通过一个数组表示,即 $\{x_c, y_c, w, h\}$,其中 (x_c, y_c) 表示物体中心点坐标; (w, h) 表示长与宽。

(2)图神经网络构建阶段。提取特征后使用图神经网络表示每一个目标与每一个检测的相似度。图网络中的每一个节点表示目标或检测特征,每一条边表示目标与检测之间的关联相似度。

(3)数据关联阶段。该阶段主要用于检测输出目标与检测之间是否有关联性。当图网络输出物体与检测的关联相似度后,利用匈牙利算法找到目标与检测之间的最优匹配。

(4)跟踪目标管理策略。对于每一个没有匹配到任何跟踪轨迹的检测对象来说,当该检测的置信分数高于 0.4 时,初始化一条新的跟踪轨迹,根据式(1)计算得到置信分数。未匹配到任何检测目标,表明该轨迹中的目标可能已经离开当前场景,或仍然在场景中但未被检测到。为避免仍然在场景中的轨迹目标跟踪终止,使用卡尔曼滤波^[21]在 T 帧内延长未检测到的轨迹,尽可能在接下来的 T 帧中使检测对象与目标轨迹匹配。如果经过 T 帧后

该目标轨迹仍然没有匹配到任何新的检测,则终止该目标轨迹。根据历史经验,设置 T 帧的值为 30。置信分数计算公式为:

$$\Pr(\text{Object}) * IOU_{pred}^{truth} \quad (1)$$

式中: $\Pr(\text{Object})$ 表示当前候选框中是否包含物体, $IOU(\text{pred}|\text{truth})$ 表示当前预测的候选框位置的准确程度; IOU 为交并比,即预测值与真实值的重合程度。

目标 o_i 与检测 d_j 之间的代价函数可表示为 $F(o_i, d_j)$ 。计算公式为:

$$F(o_i, d_j) = \alpha AGN(f_a^{o_i}, f_a^{d_j}) + (1 - \alpha) MGN(f_m^{o_i}, f_m^{d_j}) \quad (2)$$

式中: $AGN(\cdot)$ 和 $MGN(\cdot)$ 分别表示外观神经网络和运动图网络, α 为一个超参数, $f_a^{o_i}$ 和 $f_m^{d_j}$ 分别表示目标 o_i 与检测 d_j 的外观特征, $f_m^{o_i}$ 和 $f_m^{d_j}$ 分别表示目标 o_i 与检测 d_j 的运动特征。根据训练数据, α 设置为 0.3。

2.2 融合更新机制的图神经网络结构

图神经网络由节点、边和全局变量组成。节点分为目标节点和检测节点两种,其中目标节点为在第 t 帧提取 M 个已经确定身份的目标特征,形成 M 个节点;检测节点为在第 t+1 帧提取 N 个已检测到但需要确定身份的检测特

征,形成N个节点。因此,共在图神经网络中形成M+N个节点。边表示目标节点与检测节点之间的关联关系分配。最简单的方法是在每一对节点上都建立一条边,但这会导致一个全连接图出现,非常耗费算力。节点之间的匹配只能发生在相邻帧的相近节点,同一帧中或距离较远的节点间不会发生匹配。因此,对于任意一个在第t帧的目标物体 o_i ,在第t+1帧中检测 d_j 的附近区域匹配到的可能性最大。算法中仅对两个节点距离小于 $Dist_{max}$ (最大曼哈顿距

离)的点才构建边,因而在最终网络中,帧之间获得的边是稀疏的。本文将 $Dist_{max}$ 设置为200。

如图2所示,本文图神经网络包含4个模块,即边回归模块1、点特征聚合模块、全局变量更新模块、边回归模块2。图神经网络中的每一层都会循环经历这4个模块,并获得相应的相似度矩阵。为保证良好的训练效果,在图神经网络的每一层都引入点损失函数和相似度损失函数,分别对特征节点和相似度进行评估。

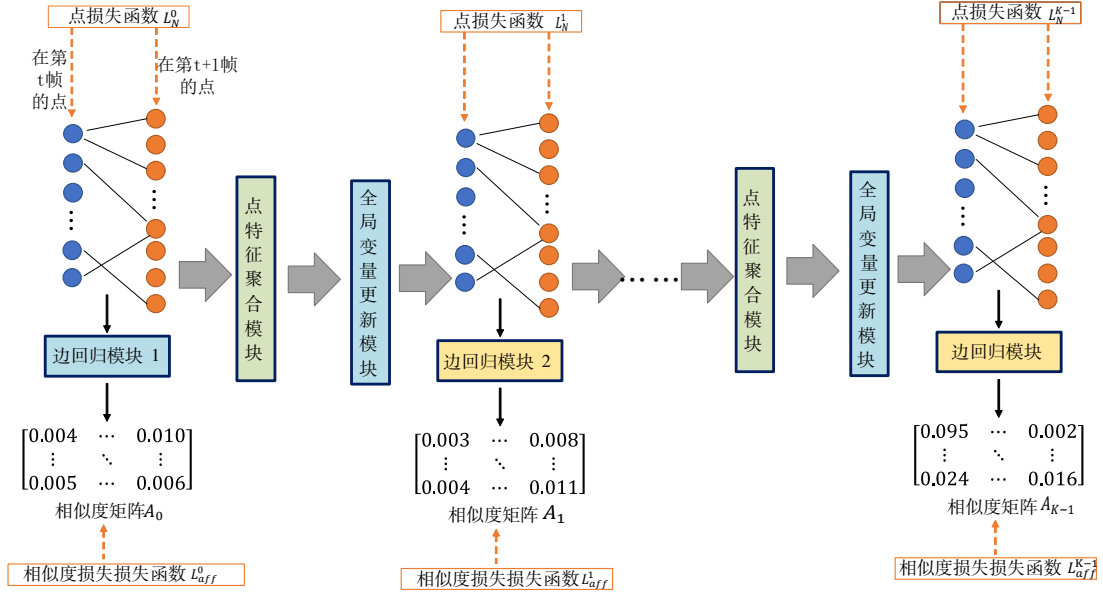


Fig. 2 Structure of the proposed graph neural network

图2 本文图神经网络结构

在多目标跟踪任务中,为了成功进行数据关联,需要在第t帧中M个跟踪目标和第t+1帧中N个检测目标之间通过计算获得一个M×N尺寸的相似度矩阵A,图神经网络将这一过程称为边回归。计算相似度的方法包括余弦相似度、欧氏距离相似度等,多为传统相似度度量方式。本文通过一个2层的多层感知机(Multilayer Perceptrons, MLP)计算相似度,其原理是将两个差异节点的特征作为输入进入到MLP中,然后输出一个取值位于0~1之间的标量值作为配对相似性分数。计算公式为:

$$A_{ij} = \text{Sigmoid}(\sigma_2(\text{RELU}(\sigma_1(n_i^t - n_j^{t+1})))) \quad (3)$$

式中, σ_1 和 σ_2 为两个不同的线性层; n_i^t 和 n_j^{t+1} 为不同帧的两个节点特征, $i \in \{1, 2, \dots, M\}$, $j \in \{1, 2, \dots, N\}$ 。

为构建图神经网络中的特征交互,在网络中的每一层聚合当前节点与邻近点,同时引入全局变量 μ ,输出的 $n_j^{t'}$ 为更新后的检测节点。计算公式为:

$$n_j^{t'} = \sigma_4(n_i^t) + \sum_{j \in \mathcal{N}(i)} \sigma_3(n_j^{t+1}) + \mu \quad (4)$$

式中, $\mathcal{N}(i)$ 表示在t帧时i节点在t+1帧中已经匹配到的一系列邻接点; σ_3 和 σ_4 表示线性层在图神经网络的邻近层中具有不同权重。

全局变量 μ 用于聚合所有目标节点和检测节点,以及所有相关的边。式(5)中的 \bar{v} 和式(6)中的 \bar{E} 分别表示聚合

的点和聚合的边。式(5)中, $\sum_j^N n_j^{t'}$ 表示对在点特征聚合模块中输出的更新后的检测节点 $n_j^{t'}$ 进行汇总, σ_5 和 σ_6 为两个线性层。因此,全局变量可表示为:

$$\bar{v} = \frac{1}{2N} (\sum_j^N n_j^{t'} + \sum_{j \in \mathcal{N}(i)} n_j^{t+1}) \quad (5)$$

$$\bar{E} = \frac{1}{MN} \sum_i^M \sum_j^N A_{ij} \quad (6)$$

$$\mu = \sigma_6(\text{RELU}(\sigma_5([\bar{v}, \bar{E}]))) \quad (7)$$

边回归2的公式与边回归1相似,只是将点聚合后的新节点引入到边回归中,用于不断更新边,实现相似度的更新。表示为:

$$A'_{ij} = \text{Sigmoid}(\sigma_2(\text{RELU}(\sigma_1(n_i^{t'} - n_j^{t+1})))) \quad (8)$$

本文更新策略基于图神经网络的4个模块进行。首先进行边更新,输入不同帧间的两个节点(目标节点和检测节点)特征,输出更新后的边;然后进行点更新,输入为上一帧的目标节点以及该节点在当前帧的一系列邻接点(检测节点)和全局变量,输出为更新后的检测节点。在全局变量更新中,输入为聚合后的节点以及边,输出为更新后的全局变量;最后进行边更新,输出相似度量。

2.3 损失函数

在训练期间,图神经网络的每一层都有两个损失函数,分别为点损失函数 L_N 和相似度损失函数 L_{aff} 。表示为:

$$L = L_N + \lambda L_{aff} \quad (9)$$

式中: λ 用于调整权重,当前设置为 1。如果目标与检测的配对属于同一个人,则检测节点进行更新,否则检测节点不进行更新。

采用相似度损失 L_{aff} 直接监督网络的最终结果,该损失由一个交叉熵损失函数构成。表示为:

$$L_{aff} = \frac{-1}{MN} \sum_i^M \sum_j^N A_{ij}^g \log A_{ij} + (1 - A_{ij}^g) \log(1 - A_{ij}) \quad (10)$$

式中: A_g 表示目标与检测之间正确关联的相似矩阵, A_g 内元素的取值只能是 1 或 0,取 1 表示检测关联到目标,取 0 表示未关联到目标。相似度损失 L_{aff} 主要作用于边回归模块 1 和边回归模块 2。

为学习到具有分辨性的特征用于匹配,图神经网络的每一层均使用了点损失函数。表示为:

$$\sum_i^M \sum_j^N (1 - A_{ij}^g) \times MSE(n_j^{i+1}, n_i^{j'}) \quad (11)$$

式中: n_j^{i+1} 为检测物体的原始特征, $n_i^{j'}$ 为点聚合更新后的特征, $MSE(\cdot, \cdot)$ 为均方差损失函数。

3 实验方法与结果分析

3.1 数据集

MOTChallenge 是一个公开的多目标跟踪基准测试平台,许多前沿方法都在该平台进行评估。MOT16 和 MOT17 是该平台最流行的两个基准测试集。之前提到影响多目标跟踪任务性能的因素为检测器质量和具有分辨性的特征学习能力,本次主要讨论后者。为保证比较的公平性,本文实验使用 MOT16 和 MOT17 平台提供的公共检测器。不同之处在于,MOT16 提供的检测结果仅来自于 DPM(Deformable Part-based Model)^[22] 算法,而 MOT17 提供的检测结果来自于 DPM^[22]、Faster R-CNN^[23] 和 SDP(Semidefinite Programming)^[24] 3 种不同算法。同时,为了防止过拟合问题出现,算法采用早停法进行训练,即将数据分为两部分:A 部分占数据集的前 4/5 帧,B 部分占后 1/5 帧;A 部分用于训练集,B 部分用于验证集。验证集主要用于监督训练过程、修正超参数,在退化前提前停止效果不佳的训练程序,以提高算法泛化能力,防止过拟合。

3.2 评估度量

实验使用的度量指标包括:①多目标跟踪准确率(Multiple Object Tracking Accuracy, MOTA)。该指标表示多目标跟踪的准确度;②身份的 F-score 值(ID F1 Score, IDF1)。该指标表示在 ID 保持相同的情况下跟踪准确率与召回率的 F-score 值;③最有可能跟踪率(Mostly Tracked, MT)。该指标表示最有可能跟踪到的百分比;④ML(Mostly Lost)最有可能丢失率。该指标表示最有可能丢失目标的百分比;⑤错误的正样本(False Positive, FP);⑥错误的负样本(False Negative, FN);⑦身份切换次数(ID Switches, IDS)。其中 MOTA 是最重要的度量指标,

因为其结合了 FP、FN 和 IDS 3 种指标。

3.3 模型性能实验

实验硬件环境为 Linux 操作系统, intel Core i7-8700, 3.20 GHZ, 模型训练与测试均通过 GPU 加速。本文在孪生网络中构建一个 CNN, 包含 4 个卷积层, 内嵌 ReLU(Rectified Linear Units)非线性激活函数。对于 LSTM 模块, 设置跟踪片段的长度为 $L=5$ 。对于被关联的检测为 R (小于 L) 的情况, 通过重复最早的第 $L-R$ 长度检测来凑够长度为 5 的跟踪片段。模型采用 Adam optimizer 函数进行训练, 权重衰减设置为 0.000 5; 初始学习率设置为 $lr=0.001$, 每 10 000 次迭代后学习率除以 10, 共进行 40 000 次迭代。

如图 3 所示, 权重值 $\alpha=0.3$ 时模型准确率最高。如图 4 所示, 图神经网络层数为 3 时模型准确率最高, 这是由于随着图神经网络层数的增加, 特征点间的融合以及不同特征间的交互愈加充分且深入, 特征表达能力也随之增强。而当图神经网络层数超过 3 层后, 网络的特征融合与交互能力趋于饱和, 更多层数只会导致模型越来越复杂, 而不会对跟踪准确率产生积极影响。

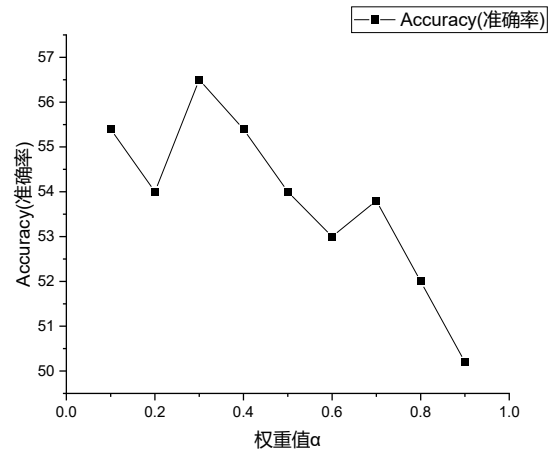


Fig. 3 Relationship between weight and accuracy

图 3 权重与准确率的关系

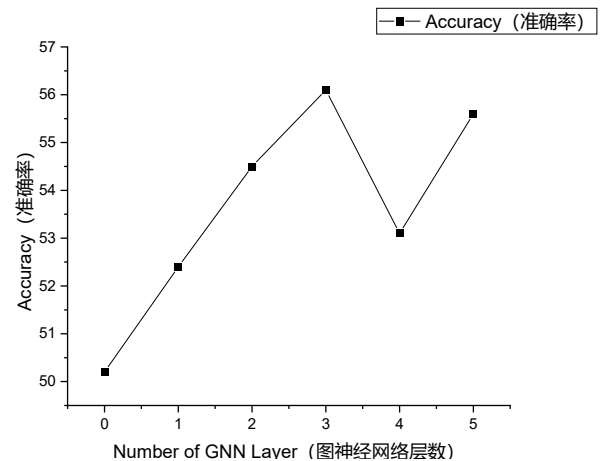


Fig. 4 Relationship between number of neural network layers and accuracy

图 4 图神经网络层数与准确率的关系

基于“通过检测进行跟踪”策略的多目标跟踪算法可分为离线跟踪算法和在线跟踪算法两种,其中离线跟踪算法的检测器作用于整个视频帧,最终获得一个全局最优解;在线跟踪算法的检测器仅作用于当前帧以及历史帧。本文算法为在线跟踪算法。

在 MOT16 数据集上将本文算法与离线跟踪算法 EDMT (Enhancing Detection Model Tracking) 算法^[25]、MHT-bLSTM (Multi-object Tracking Bilinear LSTM) 算法^[26]、Quad-CNN (Quadruplet Convolutional Neural Networks)^[27], 在线跟踪算法 STAM (Spatial-temporal Attention Mechanism) 算法^[28]、MOTDT (Multiple Object Tracking Deeply Tracking) 算法^[29]、KCF (Kernel Correlation Filter) 算法^[30]进行了比较,结果如表 1 所示。可以看出,STAM 算法在 IDS 指标上表现最佳;MOTDT 算法在 IDF1 指标上表现最佳;KCF 算法在 MT 和 FP 指标上表现最佳;本文算法在 MOTA、ML、FN 3 个指标方面表现最佳。这是由于改进图神经网络的模块更新机制可对获得的目标特征进行更加深入的融合与交互,使特征表达更加稳定,从而有效应对环境变化和目標重叠等问题。

Table 1 Comparison of performance of various algorithms on the MOT16 dataset

表 1 各算法在 MOT16 数据集上的表现比较

方式	方法	MOTA/%	IDF1/%	MT/%	ML/%	FP	FN	IDS
离线	EDMT	45.3	47.9	17.00	39.90	11 122	87 890	639
	MHT-bLSTM	42.1	47.8	14.90	44.40	11 637	93 172	753
	Quad-CNN	44.1	38.3	14.60	44.90	6 388	94 775	745
在线	STAM	46.0	50.0	14.60	43.60	6 895	91 117	473
	MOTDT	47.6	50.9	15.20	38.30	9 253	85 431	792
	KCF	48.8	47.2	15.80	38.10	5 875	86 567	906
	本文算法	49.5	49.0	15.40	38.00	6 073	81 149	673

在 MOT17 数据集上将本文算法与离线跟踪算法 EDMT^[25]、MHT-bLSTM^[26], 在线跟踪算法 MOTDT^[29]、DMAN (Dual Matching Attention Networks)^[31]进行比较,结果见表 2。可以看出,MOTDT 算法在 ML 指标上表现最佳;DMAN 算法在 IDF1 和 IDS 两个指标上表现最佳;本文算法在 MOTA、MT、FN、FP 4 个指标上表现最佳。

Table 2 Comparison of performance of various algorithms on the MOT17 dataset

表 2 各算法在 MOT17 数据集上的表现比较

方式	方法	MOTA/%	IDF1/%	MT/%	ML/%	FP	FN	IDS
离线	EDMT	50.0	51.3	21.60	36.30	32 279	247 297	2 264
	MHT-bLSTM	47.5	51.9	18.20	41.70	25 981	268 042	2 069
	MOTDT	50.9	52.7	17.50	35.70	24 069	250 768	2 474
在线	DMAN	48.2	55.7	19.30	38.30	26 218	263 608	2 194
	本文算法	53.9	54.9	19.60	37.00	21 009	249 327	2 268

3.4 消融实验

表 3 为图神经网络中各个模块消融实验结果,主要比较了算法同时采用外观模型和运动模型的前提下,图神经网络结构在 4 个模块均完整、缺少点特征聚合模块、缺少全局变量模块 3 种不同条件下对算法跟踪性能的影响。可以看出,当图神经网络中缺少点聚合模块时整体性能下

降明显,说明点特征聚合模块长时间、持续地更新模型起到了关键作用,也说明模块更新时捕捉跟踪目标间全局关系的能力下降。在缺少全局变量模块后,图神经网络在 MOTA、IDF1 等指标方面均有所下降。因此,本文模型性能最优。

Table 3 Results of ablation experiments for each module in the neural network

表 3 图神经网络中各个模块消融实验结果

方法	MOTA/%	IDF1/%	MT/%	ML/%	FP	FN	IDS
CNN+LSTM+GNN*	52.6	54.8	31.6	32.2	1 529	28 765	1 226
CNN+LSTM+GNN#	54.3	62.3	32.9	32.0	1 622	28 247	517
CNN+LSTM+GNN~	54.5	63.7	33.2	32.3	1 525	28 210	511

注:CNN+LSTM+GNN*表示基于外观模型与运动模型的不包含点特征聚合模块的图神经网络结构;CNN+LSTM+GNN#表示基于外观模型与运动模型的不包含全局变量模块的图神经网络结构;CNN+LSTM+GNN~表示基于外观模型与运动模型的 4 个模块均完整的图神经网络结构,即本文模型。

本文对外观模型和运动模型对跟踪算法性能的影响进行了比较试验,在保证图神经网络同时具备 4 个更新模块的情况下通过控制模型属性进行消融实验,结果如表 4 所示。可以看出,运动模型的整体性能优于外观模型;而同时具备运动模型和运动模型的本文模型 MOTA 值高于单纯的外观模型和运动模型,说明两个模型可以在跟踪任务中相互补充。例如当两个人的外观相似且又间隔很近时,仅使用外观模型很难区分,此时可通过运动模型根据运动轨迹进行有效区分;而当两人运动轨迹相同时可采用外观模型进行有效区分。此外,本文模型对外观属性和运动属性使用权重策略,权重系数 α 为 0.3,说明运动属性的重要性高于外观属性。

Table 4 Impact of appearance model and motion model on tracking algorithm performance

表 4 外观模型与运动模型对跟踪算法性能的影响

方法	MOTA/%	IDF1/%	MT/%	ML/%	FP	FN	IDS
CNN+GNN~	52.7	56.3	31.5	33.0	1455	28882	1161
LSTM+GNN~	53.9	61.4	31.9	32.2	1390	28570	690
CNN+LSTM+GNN~	54.5	63.7	33.2	32.3	1525	28210	511

注:CNN+GNN~表示基于外观模型的包含完整 4 个模块的图神经网络结构;LSTM+GNN~表示基于运动模型的包含完整 4 个模块的图神经网络结构。

4 结语

为增强特征表达、提高目标关联的准确度,本文提出一个基于图神经网络的多目标跟踪算法,其针对提取的目标特征分别构建外观图神经网络和运动图神经网络,以增强物体间特征的交互性;同时提出图神经网络模块更新机制,对网络中的节点进行聚合,并实现边更新以及节点与边的全局变量更新,从而输出更加可靠的相似度矩阵,使多目标跟踪模型更加持久、稳定。与其他前沿跟踪算法的比较实验结果表明,本文算法能对帧间多目标的相似度进行准确判断,并成功实现数据关联。然而,本算法虽然增强了特征分辨能力,使多目标跟踪模型更加持久、稳定,但

并没有过多考虑前期目标间相互遮挡的问题,而目标间的遮挡会导致目标特征被污染,进而影响图神经网络输出结果的可靠性。后续会通过引入注意力机制处理目标遮挡问题,以增强模型的鲁棒性和可靠性。

参考文献:

- [1] BEYER L, BREUERS S, KURIN V, et al. Towards a principled integration of multi-camera re-identification and tracking through optimal Bayes filters[C]// Computer Vision and Pattern Recognition Workshops, 2017: 1444-1453.
- [2] JIN L S, HUA Q, GUO B C, et al. Multi target tracking of forward vehicle based on optimized deepsort[J]. Journal of Zhejiang University (Engineering Edition), 2021, 55(6): 1056-1064.
金立生, 华强, 郭柏苍, 等. 基于优化 DeepSort 的前方车辆多目标跟踪[J]. 浙江大学学报(工学版), 2021, 55(6): 1056-1064.
- [3] LI J H, GAO X, JIANG T. Graph networks for multiple object tracking [C]// IEEE Winter Conference on Applications of Computer Vision, 2020: 708-717.
- [4] BOCHINSKI E, EISELEIN V, SIKORA T. High-speed tracking-by-detection without using image information[C]// 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2017: 1-6.
- [5] KANG X M, MU K N, KANG X, et al. Moving target detection and tracking method based on SIFT feature matching[J]. Electronic Design Engineering, 2018, 26(1): 174-177.
康晓梅, 穆柯楠, 康贤. 基于 SIFT 特征匹配的运动目标检测及跟踪方法[J]. 电子设计工程, 2018, 26(1): 174-177.
- [6] LI J, LI J S, FENG M Y, et al. Research on multi-target tracking algorithm based on improved evolutionary Hungary [J]. Journal of Military Communications College, 2019, 21(6): 78-85.
李炯, 李建市, 冯明月, 等. 基于改进进化匈牙利的多目标跟踪算法研究[J]. 军事交通学院学报, 2019, 21(6): 78-85.
- [7] YU F, LI W, LI Q, et al. POI: multiple object tracking with high performance detection and appearance feature [C]// European Conference on Computer Vision, 2016: 36-42.
- [8] BEWLEY A, GE Z Y, OTT L, et al. Simple online and real time tracking [C]// IEEE International Conference on Image Processing, 2016: 3464-3468.
- [9] PENG J L, WANG C, WAN F B. Chained-tracker: chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking [C]// Proceedings of the European Conference on Computer Vision, 2020: 1-16.
- [10] FEICHTENHOFER C, PINZ A, ZISSERMAN A. Detect to track and track to detect [C]// IEEE International Conference on Computer Vision, 2017: 3057-3065.
- [11] HUANG Z H, ZHAO H M, ZHAN J, et al. Twin network target tracking algorithm based on distance intersection ratio regression [J]. Journal of Yangzhou University (Natural Science Edition), 2021, 24(3): 48-54.
黄智慧, 赵慧民, 詹瑾, 等. 基于距离交并比回归的孪生网络目标跟踪算法[J]. 扬州大学学报(自然科学版), 2021, 24(3): 48-54.
- [12] JIANG L C, SHEN G Q, ZHANG G X. Image retrieval algorithm based on HSV block color histogram [J]. Electromechanical Engineering, 2009, 26(11): 54-57.
姜兰池, 沈国强, 张国焯. 基于 HSV 分块颜色直方图的图像检索算法[J]. 机电工程, 2009, 26(11): 54-57.
- [13] DEHGHAN A, ASSARI S, SHAH M. Gmmcp tracker: globally optimal generalized maximum multi clique problem for multiple object tracking [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 4091-4099.
- [14] LIU Q K, CHU Q, LIU B, et al. Graph similarity model for multi-object tracking [C]// International Joint Conference on Artificial Intelligence, 2020: 530-536.
- [15] CHARI V, LACOSTE-JULIEN S, LAPTEV I, et al. On pairwise costs for network flow multi-object tracking [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5537-5545.
- [16] SHEN Y, LI H, YI S, et al. Person reidentification with deep similarity-guided graph neural network [C]// 15th European Conference on Computer Vision, 2018: 508-526.
- [17] BRASO G, LEAL-TAIXE L. Learning a neural solver for multiple object tracking [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6247-6257.
- [18] DAI P, WENG R L, CHOI W G, et al. Learning a proposal classifier for multiple object tracking [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2021: 2443-2452.
- [19] LI N, ZHAO X M, ZHAO F, et al. Research progress of object tracking algorithm based on appearance model [J]. Computer Engineering and Science, 2017, 39(3): 524-533.
李娜, 赵祥模, 赵凤, 等. 基于外观模型的目标跟踪算法研究进展[J]. 计算机工程与科学, 2017, 39(3): 524-533.
- [20] LIAO J C, CAO L B, XIA J H, et al. Multi target trajectory tracking based on kernel correlation filtering and motion model [J]. Automotive Engineering, 2021, 6: 2443-2452.
- [21] WANG G Y, DOU L, DOU J. Multi target tracking algorithm based on adaptive Kalman filter [J]. Computer Applications, 2022, 42 (S1): 271-275.
王广玉, 窦磊, 窦杰. 基于自适应卡尔曼滤波的多目标跟踪算法[J]. 计算机应用, 2022, 42(S1): 271-275.
- [22] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [23] 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.
- [24] YANG F, CHOI W, LIN Y. Exploit all the layers: fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers [C]// Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [25] CHEN J, SHENG H, ZHANG Y, et al. Enhancing detection model for multiple hypothesis tracking [C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 2143-2152.
- [26] KIM C, LI F, REHG J M. Multi-object tracking with neural gating using bilinear LSTM [C]// 15th European Conference on Computer Vision, 2018: 208-224.
- [27] SON J, BAEK M, CHO M, et al. Multi-object tracking with quadruplet convolutional neural networks [C]// Honolulu: IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [28] CHU Q, OUYANG W, LI H, 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.
- [29] CHEN L, AI H, ZHUANG Z, et al. Real-time multiple people tracking with deeply learned candidate selection and person re-identification [C]// San Diego: International Conference on Multimedia and Expo, 2018.
- [30] CHU P, CHU H, FAN 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: 161-170.
- [31] ZHU J, YANG H, LIU N, et al. Online multi-object tracking with dual matching attention networks [C]// 15th European Conference on Computer Vision, 2018: 379-396.