控制与决策
Control and Decision

基于长短期超图神经网络匹配的多目标跟踪

郭文, 刘其贵, 王拓, 丁昕苗

引用本文:
郭文, 刘其贵, 王拓, 等. 基于长短期超图神经网络匹配的多目标跟踪[J]. 控制与决策, 2025, 40(3): 853-862.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2024.0011

基于长短期超图神经网络匹配的多目标跟踪

郭 文，刘其贵，王 拓，丁昕苗†

(山东工商学院 信息与电子工程学院，山东 烟台 264005)

摘 要: 针对联合检测与跟踪范式中存在的检测特征和 Re-ID 特征相互竞争的问题以及在复杂场景下难以保持被遮挡目标视觉一致性关系的问题, 提出一个端到端的超图神经网络关联的多目标跟踪方法 (HGTracker). 首先, HGTracker 设计一个增强的空间金字塔池化网络 (ESPPNet) 模块用来提高目标检测骨干网络的检测能力, 该模块通过聚合不同维度的特征来适应跟踪过程的不同任务, 有效地缓解一阶段跟踪方法中检测任务与 Re-ID 任务相互竞争的问题. 其次, 提出一个基于长短期超图神经网络的数据关联模块, 通过设计长期超图神经网络和短期超图神经网络来分别关联未被遮挡和被遮挡的检测视觉特征, 将数据关联问题转化为轨迹超图与检测超图之间的超图匹配问题, 跟踪器将轨迹片段信息与当前检测帧信息之间的关系建模为超图神经网络, 在严重遮挡的情况下保持了视觉轨迹的一致性. 通过一系列的对比实验, 所提出的 HGTracker 跟踪方法相比于 FairMOT 跟踪方法, 在 MOT17 数据集上 HOTA 值由 59.3 % 提高至 61.4 %, IDF1 值由 73.7 % 提高至 79.3 %, MOTA 值由 72.3 % 提高至 76.9 %; 在 MOT20 数据集上, HOTA 值由 54.6 % 提高至 57.9 %, IDF1 值由 61.8 % 提高至 73.1 %, MOTA 值由 67.3 % 提高至 75.1 %.

关键词: 多目标跟踪；超图神经网络匹配；视觉一致性关系；数据关联；联合检测与跟踪范式

中图分类号: TP391 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2024.0011

引用格式: 郭文, 刘其贵, 王拓, 等. 基于长短期超图神经网络匹配的多目标跟踪 [J]. 控制与决策, 2025, 40(3): 853-862.

A multi-object tracking method based on long-term and short-term hypergraph neural network matching

GUO Wen，*LIU Qi-gui*，*WANG Tuo*，*DING Xin-miao*†

(School of Information and Electronic Engineering， Shandong Technology and Business University， Yantai 264005，China)

Abstract: Addressing the issues of competition between detection features and Re-ID features in joint detection and embedding multi-object tracking methods, as well as difficulties in maintaining visual consistency for occluded targets in complex scenes, we propose an end-to-end hypergraph neural network matching tracking method, named HGTracker. Firstly, the HGTracker introduces an enhanced spatial pyramid pooling networks (ESPPNet) module to enhance the detection capability of the target detection backbone network. This module aggregates features from different dimensions to adapt to different tasks in the tracking process, effectively alleviating the issue of competition between detection and Re-ID tasks in one-stage multi-object tracking methods. Secondly, it introduces a short-term and long-term hypergraph neural network matching module, which designs long-term and short-term hypergraph neural networks to associate unoccluded and occluded detection visual features. It transforms the data association problem into a hypergraph matching problem between trajectory hypergraphs and detection hypergraphs. The tracker models the relationship between trajectory segment information and the current detection frame information as a hypergraph neural network, maintaining visual trajectory consistency under severe occlusion. Through a series of comparative experiments, compared to the FairMOT tracking method on the MOT17 dataset, the proposed tracking method increases the HOTA value from 59.3 % to 61.4 %, the IDF1 value from 73.7 % to 79.3 %, and the MOTA value from 72.3 % to 76.9 %; on the MOT20 dataset, the HOTA value is increased from 54.6 % to 57.9 %, the IDF1 value from

收稿日期: 2024-01-03；**录用日期:** 2024-07-15.

基金项目: 国家自然科学基金项目 (62072286, 61876100, 61572296)；山东省研究生教育创新计划项目 (SDYAL21211).

责任编委: 谢晖.

†通信作者. E-mail: dingxinmiao@126.com.

本文附带电子附录文件,可登录本刊官网该文“资源附件”区自行下载阅览.

61.8 % to 73.1 %, and the MOTA value from 67.3 % to 75.1 %.

Keywords: multi-object tracking； hyper-graph matching； visual consistency relationship； data association； joint detection and embedding

0 引 言

作为计算机视觉中一项基本的任务, 多目标跟踪 (MOT) 旨在预测视频序列中多个感兴趣目标的轨迹, 在自动驾驶、动作识别、移动机器人和智能监控方面得到了广泛的应用. 目前, 基于深度学习的多目标跟踪方法主要分为分离检测与嵌入跟踪范式 (SDE)[1-3] 和联合检测与嵌入跟踪范式 (JDE)[4-6]. SDE 范式将跟踪分为目标检测和数据关联两个独立的任务, 首先由检测器输出检测结果, 再送到数据关联模块, 将检测结果与上一帧轨迹进行匹配, 从而形成新的轨迹框轨迹. SDE 范式凭借通俗易懂和拥有出色跟踪精度的特点, 一度成为多目标跟踪的主流范式. SDE 范式的目标检测模块和数据关联模块是单独训练, 从而分别达到最好的性能, 但是它们通常具有无法通过整个 MOT 系统反向传播错误的缺点. 换言之, 每个模块只是对其自身进行优化, 而不是针对整个 MOT 系统进行优化. 并且, 当检测帧存在大量的物体时, 由于 SDE 范式两个模块并不共享特征, 需要为检测帧的每个边界框使用 Re-ID 模型, 从而导致其无法实现实时的推理速度.

联合检测和跟踪 (JDE) 范式以此为出发点, 将用于数据关联的外观嵌入模型合并到检测模块中, 所以 JDE 范式可以同时输出检测结果和相应的外观嵌入. JDE 范式利用一个共享权值深度学习网络完成目标检测和数据关联任务, 达到了较好的实时性. 但是, 由于目标检测模块任务和 Re-ID 任务所需特征的差异性, 导致了 JDE范式内部中的检测任务和 Re-ID 任务存在互相竞争的现象; 对于目标检测任务, 模型希望提高属于同一类别目标的外观特征相似性, 即缩小类内距离; 而对于 Re-ID 任务, 模型希望最大化不同目标之间的外观特征差异, 即扩大类内距离. 此外, 跟踪目标被相邻物体或杂乱物体遮挡时, 尤其是在严重遮挡的情况下, 目标检测到的特征可靠性会降低, 难以维持被遮挡目标的视觉一致性关系. 并且, 由于以上两种范式的关联模块大多数采用匈牙利匹配数据关联算法, 导致了在数据关联模块产生的误差不可以反传到目标检测模块和特征嵌入模块, 无法形成端到端的优化.

针对上述问题, 本文提出基于增强的空间金字塔网络和长短期超图神经网络匹配的多目标跟踪方法, 主要工作如下:

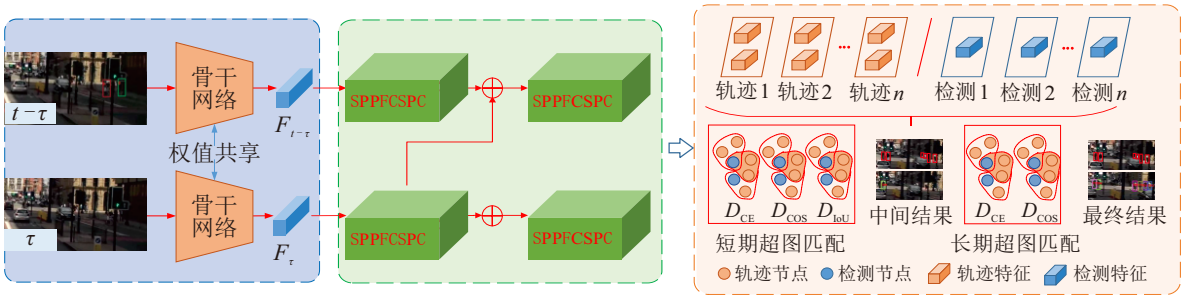
1) 设计一个增强的空间金字塔网络 (ESPPNet) 模块, 针对目标检测任务和特征提取任务的差异性问题做出改进, 缓解跟踪器中不同任务之间的特征不平衡问题.

2) 设计一个长短期超图神经网络匹配模块, 针对目标的遮挡程度来使用长期超图匹配和短期超图匹配, 提高了跟踪器在严重遮挡情况下关联的鲁棒性.

3) 通过结合以上两个模块, 本文提出的端到端的超图神经网络关联的多目标跟踪方法 (HGTracker) 不仅能够实现端到端的跟踪, 并在 MOT17 和 MOT20 等数据集的跟踪测评中在时间和准确度上获得了优良的性能.

1 方法介绍

本文提出的多目标跟踪方法整体框架如图 1 所示. 首先, 使用基于 ResNet34 改进的无锚框检测方法 DLA34 网络作为骨干网络, 用于提取输入$t-\tau$帧和t帧的原始图片特征. DLA34 网络以深度可分离卷积为基础, 能够充分利用输入图像的信息, 并具有较高的特征提取能力. 该网络具有多分辨率的特点, 可以有效地处理不同大小和尺度的目标. 得到原始图片的特征$F_{t-\tau}$和F_t后, 将其输入到增强的空间金字塔池化网络 (ESPPNet) 模块中. ESPPNet 模块结



(a) 骨干网络 (b) ESPPNET 模块 (c) 长短期超图神经网络匹配模块

图1 算法整体框架

合了空间金字塔池化的思想, 可以在不同尺度下提取特征, 并通过增强设计缓解跟踪器中不同任务之间的特征不平衡问题. 这使得 HGTracker 能够更好地适应不同尺度和复杂度的目标, 提高跟踪器的鲁棒性.

其次, HGTracker 跟踪器设计了一个长短期超图神经网络匹配模块, 该模块结合了图神经网络和长短时记忆网络的特性. 图神经网络能够有效地捕获目标之间的关系信息, 而长短期超图神经网络则可以建模目标在不同时间尺度下的运动和位置信息, 能够更好地处理目标的遮挡情况, 提高在严重遮挡情况下的关联鲁棒性.

1.1　骨干网络

基于锚框的目标检测方法需要精心调整锚框的超参数, 然而这会影响 JDE 范式中目标嵌入特征的学习. 所以, 本文基于 ResNet34 改进的无锚框检测方法 DLA34 网络作为骨干网络, 通过添加深度聚合层 (DLA) 网络, 将不同级别的语义信息和尺度特征进行融合, 获取了更加鲁棒的特征. 此外, 还将采样时所用的卷积更改为可变卷积, 以便更好地适应物体尺度和姿态的变化.

在本文中, 将$t-\tau$帧和t帧的原始图片$I_{t-\tau}$和I_t输送至骨干网络中, 得到$t-\tau$帧和t帧原始图片$I_{t-\tau}$和I_t的特征$F_{t-\tau}\in\mathbb{R}^{W\times H\times C}$和$F_t\in\mathbb{R}^{W\times H\times C}$, 其中$C$表示特征的通道数.

1.2　ESPPNet 模块

JDE 范式由单个网络获取目标检测特征和特征嵌入特征, 其中目标检测特征为低纬度的位置信息, 目标嵌入特征为高纬度的嵌入信息, 这使得目标检测任务与特征嵌入任务有所冲突. 为了缓解 JDE 范式中不同任务之间的特征冲突问题, 本文设计一个 ESPPNet 模块, 利用空间金字塔网络 (SPPFCSPC) 特征的聚合能力, 将目标检测特征和特征嵌入特征从不同尺度进行池化和聚合, 能够帮助跟踪器对目标检测特征和特征嵌入特征进行更好地区分, 从而获得更加鲁棒的特征.

SPPFCSPC 模块框架如图 2 所示. 首先将图片I_t的特征F_t经过一个大小为1×1的卷积层进行降维, 再进行一个大小为3×3的卷积层来提取更高层次的信息, 并经过一个大小为1×1的卷积层进行降维, 之后依次通过大小为5×5的池化层进行池化, 将每一次池化的特征与原始特征拼接, 再分别进行大小为1×1的卷积层和大小为3×3的卷积层, 并与特征F_t进行拼接, 最后再通过一个大小为1×1的卷积层进行降维以维持特征F_t的大小.

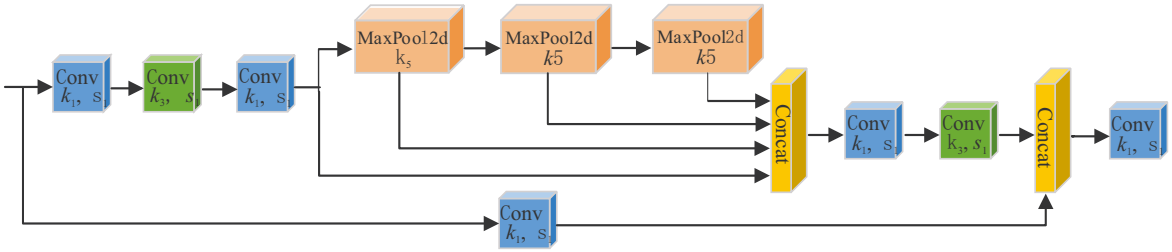


图2　SPPFCSPC 模块框架

1.3　长短期超图神经网络匹配模块

本文将多目标跟踪中的数据关联问题转换为超图神经网络匹配问题. 首先, 设计短期目标特征F_t^s和长期目标特征F_t^l来表征检测I_t帧中未被遮挡和被遮挡的目标特征. 在超图匹配模块中, 将两类目标特征应用于不同的阶段, 设计了一个两阶段的超图神经网络匹配策略.

由个体图构建亲和度张量K, 匹配问题等价于图上的顶点分类. 对关联超图邻接矩阵进行匹配感知嵌入和顶点分类, 生成顶点得分, 然后进行 reshape 和 Sinkhorn 归一化, 得到双随机矩阵, 进而得到置换矩阵, 最后得到数据关联的结果. 从个体图到关联结果示例如图 3 所示.

1.3.1　超图神经网络的构建

设超图神经网络$H=(V,E)$, 其中轨迹超图神经网络表示为$H^T=(V^T,E^T)$, 当前检测帧超图神经网络表示为$H^D=(V^D,E^D)$. 通常, V^T表示从第一帧到当前帧前一帧中存在的检测节点, 其中$|V^T|=n_T$; V^D表示当前帧中检测到的节点个数 (包括上一帧中存在的节点、新增的节点和移动出摄像头的节点), 其中$|V^D|=n_D$. E^T、E^D为图中具有二阶特征属性边的集合, 其中$|E^T|=n_{e_T}$, $|E^D|=n_{e_D}$.

在本文的超图神经网络匹配中采用三阶亲和度张量. 通过考虑三元节点之间的相似性和几何一致性, 该方法具备了缩放和旋转的不变性, 超图神经网络的三阶亲和力匹配示意图如图 4 所示. 与其他超

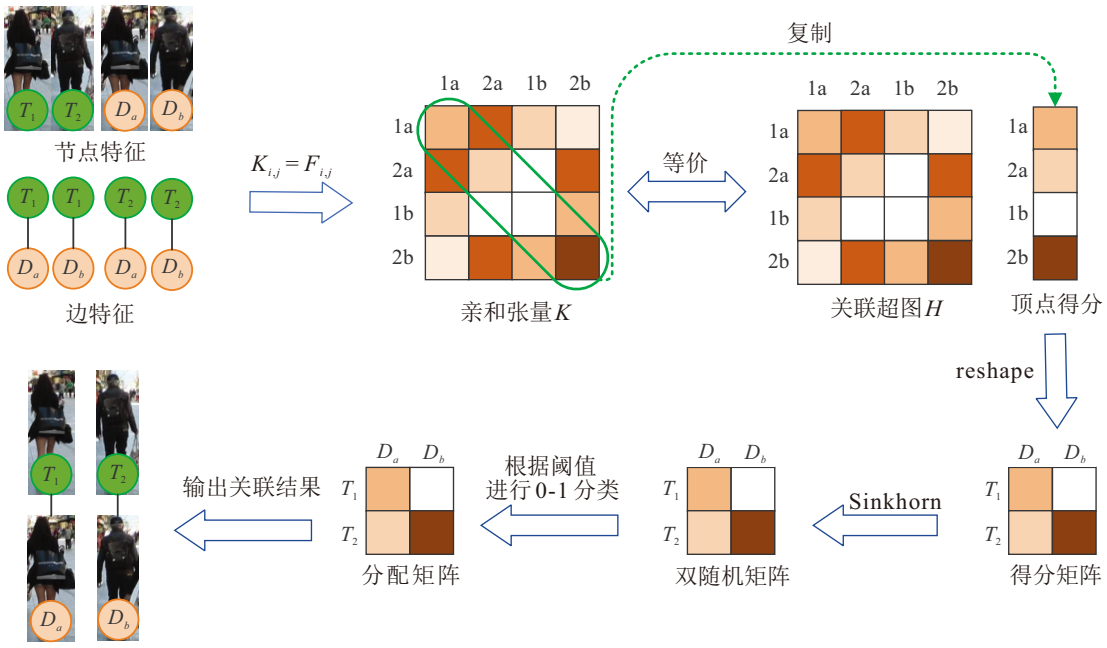


图3 从个体图到关联结果示例

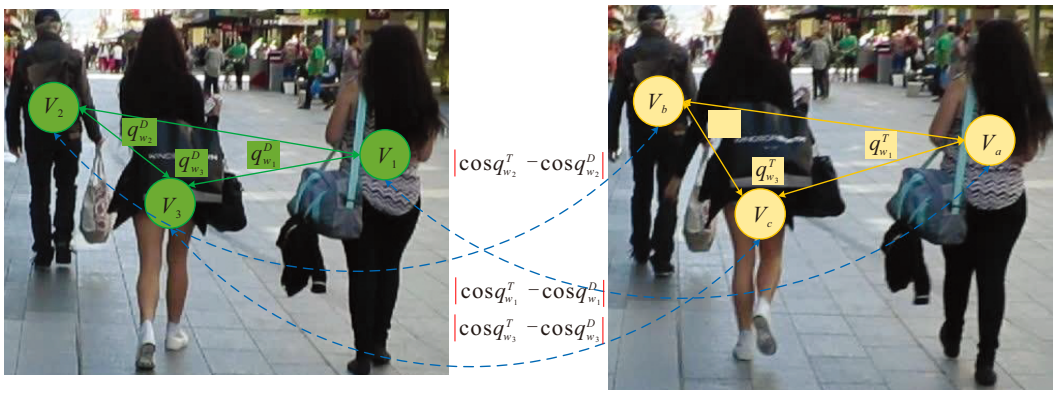


图4 超图神经网络的三阶亲和力超图匹配示意

图匹配文献[7-8]保持一致, 三阶亲和度张量定义为

$$\boldsymbol{H}^{(3)}{}_{\omega_1,\omega_2,\omega_3}=\exp\left(-\frac{\sum_{q=1}^{3}|\cos\theta^{T}_{\omega_q}-\cos\theta^{D}_{\omega_q}|}{\sigma_3}\right). \tag{1}$$

其中$\theta^{T}_{\omega_q}$和$\theta^{D}_{\omega_q}$分别表示H^T和H^D中的夹角.

1.3.2 长期和短期超图神经网络的构建

首先, 将置信度小于等于α的检测目标建模为长期超图神经网, 通过骨干网络提取长期轨迹超图H^{LT}中的节点特征 (表示为$\overline{F_V^{\mathrm{LT}}}\in\mathbb{R}^{n_{\mathrm{LT}}\times d}$) 和边特征 (表示为$\overline{F_E^{\mathrm{LT}}}\in\mathbb{R}^{n_{\mathrm{LT}}\times d}$), 其中, $\overline{F_V^{\mathrm{LT}}}$和$\overline{F_E^{\mathrm{LT}}}$由前$t-1$帧中的特征取平均值得到, 当前长期检测帧超图$H^{\mathrm{LD}}$中节点特征表示为$\overline{F_V^{\mathrm{LD}}}\in\mathbb{R}^{n_{\mathrm{LD}}\times d}$, 边特征表示为$\overline{F_E^{\mathrm{LD}}}\in\mathbb{R}^{n_{\mathrm{LD}}\times d}$, d为特征维度大小, 特征都经过双线性插值处理. 然后, 将置信度大于α的检测目标建模为短期超图神经网络, 通过骨干网络提取短期轨迹超图H^{ST}中的节点特征 (表示为$\overline{F_V^{\mathrm{ST}}}\in\mathbb{R}^{n_{\mathrm{ST}}\times d}$) 和边特征 (表示为$\overline{F_E^{\mathrm{ST}}}\in\mathbb{R}^{n_{\mathrm{ST}}\times d}$), 其中, $\overline{F_V^{\mathrm{ST}}}$和$\overline{F_E^{\mathrm{ST}}}$由前$t-1$帧中的特征取平均值得到, 当前短期检测帧超图$H^{\mathrm{SD}}$中节点特征$\overline{F_V^{\mathrm{SD}}}\in\mathbb{R}^{n_S D\times d}$和边特征$\overline{F_E^{\mathrm{SD}}}\in\mathbb{R}^{n_S D\times d}$, d为特征维度大小, 特征都经过双线性插值处理. 由于对长期超图和短期超图采用了相同的求解过程, 在下文求解过程中将省略长期超图和短期超图的标识.

轨迹超图和检测超图的连通性分别用$\overline{G_{\mathrm{in}}^T}$、$\overline{G_{\mathrm{out}}^T}\in\{0,1\}^{n_T\times n_{eT}}$和$\overline{G_{\mathrm{in}}^D}$、$\overline{G_{\mathrm{out}}^D}\in\{0,1\}^{n_D\times n_{eD}}$表示, 轨迹超图的邻接张量$\overline{A^T}=\overline{G_{\mathrm{in}}^T}\overline{G_{\mathrm{out}}^T}^{\mathrm{T}}$, $\overline{G_{\mathrm{in}}^T}\overline{G_{\mathrm{out}}^T}^{\mathrm{T}}$代表对$\overline{G_{\mathrm{in}}^T}\overline{G_{\mathrm{out}}^T}$求转置操作, 下文操作与此相同; 当前检测帧超图的邻接张量$\overline{A^D}=\overline{G_{\mathrm{in}}^D}\overline{G_{\mathrm{out}}^D}^{\mathrm{T}}$, $\overline{G}_{\mathrm{in}}(i,k)=\overline{G}_{\mathrm{out}}(j,k)=1$表示边$k$为节点$i$到节点$j$所连接的边. 通过在边两端串联节点特征来构建边的表示, 即

$$\begin{aligned}\overline{X}&=[\overline{G_{\mathrm{in}}^T}^{\mathrm{T}}\overline{F_E^T}\quad\overline{G_{\mathrm{out}}^T}^{\mathrm{T}}\overline{F_E^T}],\\ \overline{Y}&=[\overline{G_{\mathrm{in}}^D}^{\mathrm{T}}\overline{F_E^D}\quad\overline{G_{\mathrm{out}}^D}^{\mathrm{T}}\overline{F_E^D}],\end{aligned} \tag{2}$$

其中$[\cdot\ \cdot]$表示沿列连接两个张量. 通过构建节点到节点的相似度张量$K_v \in \mathbb{R}^{n_T \times n_D}$和边与边的相似度张量$K_e \in \mathbb{R}^{n_{eT} \times n_{eD}}$, 有

$$K_v = \overline{F_V^T} \overline{F_V^D}^{\mathrm{T}},\ K_e = \overline{X} \Lambda \overline{Y}^{\mathrm{T}}, \tag{3}$$

其中$\Lambda \in \mathbb{R}^{2d \times 2d}$为亲和度度量的可学习参数. 亲和矩阵是根据$K$的分解公式构建, 有

$$\begin{aligned} K = &\operatorname{diag}(\operatorname{vec}(K_v))(\overline{G_{\text{in}}^T} \otimes_K \overline{G_{\text{in}}^D}) + \\ &\operatorname{diag}(\operatorname{vec}(K_e))(\overline{G_{\text{out}}^T} \otimes_K \overline{G_{\text{out}}^D})^{\mathrm{T}}. \end{aligned} \tag{4}$$

其中: $\operatorname{diag}(\cdot)$为从输入向量构建对角矩阵, $\otimes_K$为Kronecker乘积. 上述所有的操作都可以进行反向传播操作.

1.3.3　关联图的构建和匹配感知嵌入

从亲和张量K推导出关联超图张量H, 张量H同时包含了连通性和权重信息. 关联超图张量H的一阶邻接权重来自K的对角元素, 关联超图张量H的二阶邻接权重来自K的非对角元素. 关联超图张量H的一阶邻接权重和二阶邻接权重分别为

$$\boldsymbol{H}_{i,i}^{(1)} = K_{i,i},\ \boldsymbol{H}_{i,j}^{(2)} = K_{i,j}. \tag{5}$$

三阶关联超图张量H的邻接权重计算方式如式 (1) 所示. 当一阶相似度$K_{i,i}$不存在时, 可以给所有的$\boldsymbol{H}^{(1)}$分配一个常数 (例如 1). $\boldsymbol{H}^{(k)} \in \mathbb{R}^{n_T \times n_D \times l_k}$为$k$层上 (从$k=1$开始) 的$l_k$维顶点嵌入.

匹配问题可以转化为选择关联图中编码轨迹超图H^T与检测超图H^D两个超图之间节点对应关系的顶点. 定义关联超图$\boldsymbol{A} \in \{0,1\}^{n_T n_D \times n_T n_D}$的邻接张量: 如果$K_{i,j} > 0$, 则$\boldsymbol{A}_{i,j} = 1$, $\boldsymbol{A}_{i,j}$作为关联图中顶点i和j 是否存在超边的指标. 在关联超图中, 一条存在于i与j之间的边当且仅当节点i与节点j存在边关系, 由此定义存在节点i到节点j之间边的亲和度分数. 从关联超图中由超边连接的所有更新顶点嵌入, 再计算 3 阶的归一化度量张量. 与超图匹配顶点更新文献 [7] 保持一致, 其计算过程为

$$\boldsymbol{D}_{\omega1,\omega2,\omega3}^{(t)\ -1} = \frac{\boldsymbol{A}_{\omega1,\omega2,\omega3}^{(t)}}{(\boldsymbol{A}^{(t)} \otimes_2 \ldots \otimes_t 1)_{\omega1}}. \tag{6}$$

顶点聚合步骤如下:

$$\begin{aligned} &p^{(k)} = f_m^{(t)}(\boldsymbol{v}^{(k-1)}), \\ &\boldsymbol{H}^{(t)\prime} = (\boldsymbol{D}^{(t)\ -1} \odot \boldsymbol{H}^{(t)})_{t+1}, \\ &\boldsymbol{m}^{(k)} = \sum_t \lambda_t \boldsymbol{H}^{(t)\prime} \otimes_t p^{(k)} \ldots \otimes_2 p^{(k)} + f_v, \\ &\boldsymbol{v}^{(k)} = [\boldsymbol{m}^{(k)} \quad \operatorname{vec}(\text{Classifier}(\boldsymbol{m}^{(k)}))]. \end{aligned} \tag{7}$$

其中: f_v为顶点自更新函数$f_v(\boldsymbol{v}^{(k-1)})$的简写; $\otimes_i$表示i维向量乘积; $\odot$表示逐元素相乘; $(\cdot)_{t+1}$表示沿着维度$(t+1)$展开; $f_m^{\langle t \rangle} : \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{(l_k - 1)}$为$t$ 阶的消息传递函数, 不同阶数的特征以λ_t加权求和的方式进行融合; $[\cdot\ \cdot]$表示串联, 所提出的带有 Sinkhorn 网络的顶点分类器被表示为 Classifier : $\mathbb{R}^{n_T \times n_D \times l_k} \to [0,1]^{n_T \times n_D}$; 消息传递函数$f_m : \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_k}$和顶点自更新函数$f_v : \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_k}$通过两个全连接层和 ReLU 激活的网络实现.

式 (7) 介绍了超图神经网络顶点嵌入过程, 通过在每一层中添加一个软置换 (即双随机矩阵), 利用 Sinkhorn 网络分类器进行评分: $\mathbb{R}^{n_T \times n_D \times l_k} \to [0,1]^{n_T \times n_D}$, 通过矢量化操作符$\operatorname{vec}(\cdot)$进行计算, 将预置的软置换被级联到顶点嵌入中, 从而更好地在嵌入层中考虑匹配信息.

1.3.4　利用 Sinkhorn 网络进行顶点分类

因为图匹配等价于关联图上的顶点分配, 由此采用带 Sinkhorn 网络的顶点分配器对匹配结果进行预测. 本文使用一个单层的全连接分类器$f_c : \mathbb{R}^{l_k} \to \mathbb{R}$, 再使用指数函数进行激活. 其计算过程为

$$\boldsymbol{s}_{ij}^{(k)} = \exp(f_c(\boldsymbol{v}_{ij}^{(k)})). \tag{8}$$

在将分类器分数重新构造为$\mathbb{R}^{n_T \times n_D}$后, Sinkhorn 网络对$s$添加一对一的分配约束. 它将一个非负方阵作为输入, 并通过反复运行得到一个双随机矩阵. 其计算过程为

$$R_i = \sum_{j=1}^{n_T} s_{ij},\ C_j = \sum_{i=1}^{n_D} s_{ij},\ s_{ij} = \frac{s_{ij}}{R_i + C_j}. \tag{9}$$

其中:R_i表示对每一行元素求和, C_j表示对每一列元素求和, 再重复更新矩阵s中的元素, 直到s逐渐收敛为一个双随机矩阵, 其行和列之和均为 1. 再得到分配矩阵, 最后输出关联结果.

1.4　损失函数

针对高置信度的短期超图神经网络匹配, 通过计算轨迹超图H^{ST}和检测超图H^{SD}之间的交叉熵损失L_{CE}、余弦损失L_{COS}和 IoU 损失L_{IoU}来完成短期超图神经网络匹配, 计算过程如下所示:

$$L_S = \lambda_1 L_{\text{CE}} + \lambda_2 L_{\text{COS}} + \lambda_3 L_{\text{IoU}}. \tag{10}$$

针对低置信度的长期超图神经网络匹配, 通过计算轨迹超图H^{LT}和检测超图H^{LD}之间的交叉熵损失L_{CE}和余弦损失L_{COS}来完成长期超图神经网络匹配, 即

$$L_L = \lambda_4 L_{\text{CE}} + \lambda_5 L_{\text{COS}}. \tag{11}$$

最终得到的预测矩阵S是一个双随机矩阵, 每个元素可以看成一个二分类, 所以使用交叉熵损失. 由于在本文超图神经网络中使用了角度超边, 可使用余弦损失来度量特征向量之间的相似性. 此外, 由于

短期超图神经网络中的置信度比较高, 目标检测框和跟踪检测框效果更加显著, 使用 IoU 损失来帮助检测框可以进行更好更快地回归; 但是,在长期图神经网络中, 大多数目标已经被遮挡, 如果对其进行 IoU 度量,则很容易会使得检测框发生偏移、回归不准确等现象, 所以在长期超图神经网络中并没有使用 IoU 损失.

2 实验结果及分析

2.1 数据集简介

本文在 MOTChallenge 基准[9] 上进行评估, 主要是 MOT17[10] 数据集和 MOT20 数据集[11]. MOT17 数据集包含 14 段视频序列, MOT20 数据集包含 8 段视频序列, 密度程度极为复杂. 由于数据集并未提供验证分割, 在消融实验中采用文献 [6] 的做法, 将训练序列分为两半, 使用一半数据集训练, 另一半用于验证.

2.2 参数设置及评价指标

本文使用 DLA34 网络变体结构作为主干网络, 在 COCO 数据集上与训练的模型参数用于初始化模型, 本文实验的运行环境为 Ubuntu 20.04 系统, 内存 64 GB, GPU 为 TITAN RTX. 软件配置为 CUDA11.3, 使用 Adam 优化器训练 30 个 epochs, 初始学习率为 10^{-4}, 学习速率在 20 个 epoch 时衰减到10^{-5}, 批处理大小设置为 16. 输入图像大小调整为 1088×608, 特征图分辨率为 272×152.

为了评估跟踪性能指标, 本文使用清晰的度量标准, 包括 Higher Order Tracking Accuracy(HOTA↑)、Multi-Object Tracking Accuracy(MOTA↑)、IDF1 Score (IDF1↑)、False Positive(FP↓)、False Negative(FN↓) 和 Number of Identity Switches(IDs↓) 等. MOTA 比较关注检测分支性能; IDF1 评估身份保持能力, 更关注关联性能; HOTA 可以综合地评估检测分支和数据关联的性能.

MOTA 可以表示为

$$\mathrm{MOTA}=1-\frac{\sum_t(\mathrm{FN}_t+\mathrm{FP}_t+\mathrm{IDS}_t)}{\sum_t \mathrm{GT}_t}. \tag{12}$$

其中: t表示第t帧; GT_t表示第t帧对象的真实值; FN_t表示第t帧漏检的数量, 即未被该方法检测到真实值的数量; FP_t表示第t帧误检的数量, 即被该方法错误检测到但不存在真实值的数量; IDS_t表示第t帧的目标身份跳变的次数, 即给定轨迹从一个真实值改变到另一个物体的次数.

IDF1 可以表示为

$$\mathrm{IDF1}=\frac{2\mathrm{IDTP}}{2\mathrm{IDTP}+\mathrm{IDFP}+\mathrm{IDFN}}. \tag{13}$$

其中: IDTP 表示真的正身份数, 即正确跟踪正确目标的数目; IDFP 表示假的正身份数, 即错误跟踪正确目标的数目; IDFN 表示假的负身份数, 即错误跟踪非目标的数目.

HOTA 指标可以对多目标跟踪过程中的检测精度和关联精度进行统一衡量, 并且还可以评估长期高阶跟踪关联; 而且, HOTA 也可以分解为子指标, 从而允许分析跟踪器性能的不同组成部分. HOTA 可以表示为

$$\mathrm{HOTA}=\int_0^1 \mathrm{HOTA}_\alpha \mathrm{d}\alpha \approx \frac{1}{19}\sum_{\alpha\in\{0.05,0.1,\ldots,0.95\}}\mathrm{HOTA}_\alpha, \tag{14}$$

HOTA_α为 HOTA 分解的子指标, 可以表示为

$$\mathrm{HOTA}_\alpha=\sqrt{\frac{\sum_{c\in\{\mathrm{TP}\}}A_c}{|\mathrm{TP}|+|\mathrm{FN}|+|\mathrm{FP}|}}. \tag{15}$$

其中: TP 为检测器预测为正样本的数量; FN 为检测器漏检的数量; FP 为检测器误检的数量; A_c可以表示为

$$A_c=\frac{|\mathrm{TPA}(c)|}{|\mathrm{TPA}(c)|+|\mathrm{FNA}(c)|+|\mathrm{FPA}(c)|}, \tag{16}$$

$\mathrm{TPA}(c)$为真正关联的数量, $\mathrm{FNA}(c)$为假负关联的数量, $\mathrm{FPA}(c)$为假正关联的数量. $\mathrm{TPA}(c)$、$\mathrm{FNA}(c)$和$\mathrm{FPA}(c)$分别为

$$\begin{aligned}&\mathrm{TPA}(c)=\{k\},\\&\quad k\in\{\mathrm{TP}|\mathrm{prID}(k)=\mathrm{prID}(c)\bigwedge\mathrm{gtID}(k)=\\&\quad \mathrm{gtID}(c)\};\end{aligned} \tag{17}$$

$$\begin{aligned}&\mathrm{FNA}(c)=\{k\},\\&\quad k\in\{\mathrm{TP}|\mathrm{prID}(k)\neq\mathrm{prID}(c)\bigwedge\mathrm{gtID}(k)=\\&\quad \mathrm{gtID}(c)\}\bigcup\{\mathrm{FN}|\mathrm{gtID}(k)=\mathrm{gtID}(c)\};\end{aligned} \tag{18}$$

$$\begin{aligned}&\mathrm{FPA}(c)=\{k\},\\&\quad k\in\{\mathrm{TP}|\mathrm{prID}(k)=\mathrm{prID}(c)\bigwedge\mathrm{gtID}(k)\neq\\&\quad \mathrm{gtID}(c)\}\bigcup\{\mathrm{FP}|\mathrm{gtID}(k)=\mathrm{gtID}(c)\}.\end{aligned} \tag{19}$$

2.3 实验结果分析

为了验证超图神经网络的高阶结构信息阶数t对跟踪效果的影响, 对超图神经网络的不同高阶结构信息阶数进行了消融实验, 实验结果如表 1 所示. 随着超图神经网络的高阶信息阶数增多, MOTA 和 IDF1 也呈现出增高的趋势. 这一趋势表明, 在学习时空一致性时, 高阶结构信息关系特征比成对关系特征更为有效. 然而, 由于高阶结构中爆炸的计算代

价 ($O((n_D n_T)^t)$, t为阶数), 在本文中没有探索超过3阶结构信息的超图神经网络.

表1　超图神经网络中高阶结构信息阶数 t对跟踪效果的影响　　%

t	HOTA↑	MOTA↑	IDF1↑	FP↓	FN↓	IDs↓	Hz↑
1	57.7	70.4	71.3	1367	**11762**	382	**25.9**
2	58.6	71.3	73.3	1257	13762	282	21.3
3	**58.9**	**71.5**	**73.6**	**1081**	12219	**215**	18.8

为了验证本文所提出的模块对跟踪效果的影响, 对本文框架中的模块进行了消融实验, 实验结果如表2所示. 从MOTA指标的角度来看, ESPPNet模块显著提高了跟踪器的跟踪准确度, MOTA提高了3.3个百分点(从69.3 %提升至72.6 %);而从IDF1指标的角度来看, 超图匹配模块明显提高了关联精度, IDF1提高了2.5个百分点(从72.1 %提升至74.6 %). 并且, 由于本文属于端到端学习方法, 当两个模块相加时, 每个模块的性能亦得到了极大的提升, 其中HOTA提高了2.1个百分点(从56.5 %提升至58.6 %), MOTA提升了4个百分点(从69.3 %提升至73.3 %), IDF1提升了2.8个百分点(从72.1 %提升至74.9 %).

表2　不同模块对跟踪效果的影响　　%

Baseline	ESPPNet	Hypergraph	HOTA↑	MOTA↑	IDF1↑	IDs↓
✓			56.5	69.3	72.1	239
✓	✓		57.0	72.6	72.4	193
✓		✓	57.8	70.5	74.6	207
✓	✓	✓	**58.6**	**73.3**	**74.9**	**194**

表3展示了短期超图神经网络匹配中不同参数比例对跟踪效果的影响, 表4展示了长期超图神经网络匹配中不同参数比例对跟踪效果的影响. 可以看出, 无论是长期超图匹配模块还是短期超图匹配模块, L_{CE}损失对跟踪效果的影响都较为重要. 此外,

表3　短期超图神经网络匹配中不同参数比例对跟踪效果的影响　　%

λ_1	λ_2	λ_3	HOTA↑	MOTA↑	IDF1↑	IDs↓
1	0	0	56.3	69.2	72.1	239
1	1	1	56.9	69.9	72.7	208
1	1	0.1	56.5	70.3	72.1	239
1	0.1	1	59.5	71.3	74.1	256
1	1	10	57.0	71.6	71.7	188
1	10	1	57.3	70.7	72.4	193
0.1	1	1	57.8	70.5	71.1	206
0.1	1	10	57.2	71.5	71.6	239
0.1	10	1	57.9	71.2	73.3	224
10	1	0.1	58.8	71.3	74.2	211
10	0.1	1	58.1	71.6	72.6	206
10	1	1	**59.6**	**71.9**	**75.9**	**194**

表4　长期超图神经网络匹配中不同参数比例对跟踪效果的影响　　%

λ_4	λ_5	HOTA↑	MOTA↑	IDF1↑	IDs↓
1	0	56.3	69.2	72.1	239
1	1	57.0	69.6	72.9	203
1	10	58.5	70.9	73.1	239
10	0.1	57.5	71.3	72.0	210
10	1	**59.1**	**73.3**	**75.6**	**178**

L_{COS}和L_{IoU}可以帮助跟踪器进行更鲁棒的关联. 由表3和表4可以得出, 当$\lambda_1=10$、$\lambda_2=1$、$\lambda_3=1$时, 短期超图神经网络模块跟踪性能能够达到最优; 当$\lambda_4=10$和$\lambda_1=1$时, 长期超图神经网络模块跟踪性能能够达到最优.

为了进一步分析不同α取值对跟踪效果的影响, 本文在MOT17数据集上和MOT20数据集上对不同α取值进行实验验证, 结果如表5和表6所示. 由表5可以看出, 在MOT17数据集上当α取0.4时可以达到最好的跟踪性能; 由表6可以看出, 在MOT20数据集上当α取0.3时可以达到最好的跟踪性能. 由于MOT20数据集人流过于密集, 需要比MOT17数据集更低的阈值.

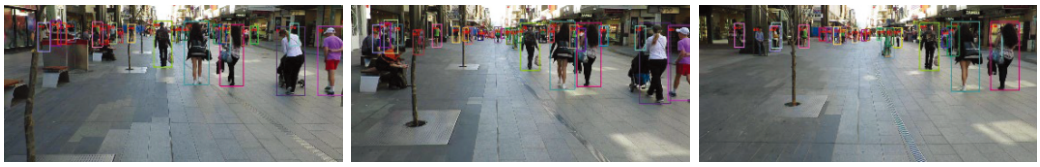
表5　MOT17数据集上不同α取值对跟踪效果的影响　　%

α	HOTA↑	MOTA↑	IDF1↑	IDs↓
0.6	59.3	75.9	71.5	5018
0.5	60.5	75.6	77.1	5239
0.4	**61.4**	**76.9**	**79.3**	**4062**
0.3	59.9	76.3	76.9	4203

表6　MOT20数据集上不同α取值对跟踪效果的影响　　%

α	HOTA↑	MOTA↑	IDF1↑	IDs↓
0.6	56.2	72.1	70.1	3063
0.5	56.4	70.9	73.1	2939
0.4	57.5	71.3	72.0	2910
0.3	**57.9**	**75.0**	**73.1**	**2280**
0.2	57.1	74.0	71.1	2828

为了验证输入两帧图像之间的间隔对实验结果的影响, 对输入间隔τ的不同取值做了消融实验, 结果如表7所示. 当$\tau=0$时代表仅仅使用了当前帧图像信息, 并未使用多帧图像信息. 可以看出, 当$\tau=1$时可以达到最好的跟踪性能,可能是由于本文的设计架构是针对相邻帧之间进行超图神经网络建模, 通过超图神经网络匹配完成了多目标跟踪的数据关联过程; 如果是相隔两帧或者三帧情况下, 由于长时间的等待(两帧或者三帧), 跟踪目标的外观信息会发生突变、遮挡等状况, 影响目标的相似度计算.



(a) MOT17

(b) MOT20

图5 HGTracker 跟踪器结果可视化

表7 不同输入时间间隔τ对跟踪效果的影响 %

τ	本文模块	HOTA↑	MOTA↑	IDF1↑
0	无	56.5	69.3	72.1
0	有	57.9	73.0	73.9
1	有	**58.6**	**73.3**	**74.9**
2	有	58.0	73.1	73.6
3	有	57.9	72.6	72.7

图 5 展示了 HGTracker 跟踪器在 MOT 数据集上的跟踪结果可视化, 可以看出本文跟踪器具有较好的鲁棒性.

2.4 与现有方法对比

为了进一步验证本文 HGTracker 方法的有效性, 在表 8 给出了本文算法 HGTracker 与当前主流多目标跟踪算法在 MOT17 测试集上的结果对比. 如表 8 所示, HGTracker 在 HOTA、MOTA 和 IDF1 指标上优于大多数主流一阶段跟踪方法. 在 MOTA 评价指标方面, 本文在 FairMOT 的基础上使用了 ESPPNet 特征增强模块,有效缓解了 FairMOT[6] 中存在的检测特征和 ID 嵌入特征相互竞争的问题, 提高了 HGTracker 的检测性能. 在 IDF1 评价指标方面, 由于本文 HGTracker 跟踪器超图神经网络匹配模块相对于匈牙利匹配算法更优, 并且本文 HGTracker 跟踪器的长短期超图神经网络匹配模块更加关注视觉信息的一致性, 从而在严重遮挡的情况下帮助跟踪器进行更好地关联. 此外, 在 HOTA 评价指标方面, 相对于其他基于图神经网络的多目标跟踪方法使用的两阶信息特征, 本文算法基于长短期超图神经网络不仅可以提取更高阶的信息特征, 还更加关注被遮挡跟踪目标的视觉一致性关系. 从 HOTA 指标可以看出,本文 HGTracker 跟踪器优于绝大多数基于图神经网络的一阶段多目标跟踪方法.

表8 本文方法在 MOT17 数据集上与其他方法结果对比 %

方法	发表	HOTA	MOTA	IDF1	IDs	FP	FN	FPS
CenterTrack[12]	CVPR2021	48.2	59.6	61.5	2 583	**14076**	200 672	17.0
SOTMOT[13]	CVPR2021	–	71.9	71.0	5 184	39 537	118 983	16.0
MOTR[14]	ECCV2022	57.8	68.6	73.4	2 439	20 268	133 440	4.5
MAT[15]	Neurocomputing2022	56.0	69.2	67.1	**1 279**	22 756	161 547	11.5
GSDT[16]	ICRA2021	55.5	68.7	66.2	3 318	43 368	144 261	4.9
MeMOT[17]	CVPR2022	56.9	69.0	72.5	2 724	37 221	115 248	–
YOLOTracker[18]	PR2022	53.5	65.1	67.1	4 983	37 701	142 914	24.9
GMTracker[19]	CVPR2021	54.0	68.7	65.0	2 200	18 213	177 058	–
FairMOT[6]	IJCV2021	59.3	72.3	73.7	3 303	27 507	117 477	18.9
CSTrack[20]	TIP2022	59.3	72.6	74.9	3 567	23 847	114 303	15.8
CTracker[21]	ECCV2020	49.0	61.2	66.6	5 529	22 284	160 491	**34.4**
Trackformer[22]	CVPR2022	57.3	68.0	74.1	2 829	34 602	108 777	5.7
PermaTrack[23]	ICCV2021	55.5	68.9	73.8	3 699	28 998	115 104	11.9
SGT[24]	WACV2023	60.6	72.8	76.4	4 578	25 983	102 984	–
CorrTracker[25]	CVPR2021	60.7	73.6	76.5	3 369	29 808	**99 510**	15.6
ours	–	**61.4**	**76.9**	**79.3**	4 062	25 515	100 539	18.8

表9　本文方法在 MOT20 数据集上与其他方法结果对比　　%

方法	发表	HOTA	MOTA	IDF1	IDs	FP	FN	FPS
FairMOT[6]	IJCV2021	54.6	67.3	61.8	7 874	103 440	88 901	13.2
CSTrack[20]	TIP2022	54.0	68.6	66.6	3 196	**25 404**	144 358	4.5
RelationTrack[26]	TMM2023	55.1	67.9	60.6	5 686	112 927	**85 062**	3.0
SGT[24]	WACV2023	56.9	70.5	72.8	2 853	33 204	93 612	1.8
GSDT[16]	ICRA2021	53.6	67.5	67.1	3 131	31 913	135 409	0.9
Trackformer[22]	CVPR2022	54.7	65.7	68.6	2 474	20 348	140 373	5.7
MAA[27]	WACV2022	57.3	73.9	71.2	2 331	24 942	108 744	**14.7**
DecodeMOT[28]	TIP2023	54.5	67.2	69.0	2 805	35 217	131 502	12.2
ours	–	**57.9**	**75.0**	**73.1**	**2 280**	26 759	100 594	9.6

表 9 给出了本文算法 HGTracker 与当前主流多目标跟踪算法在 MOT20 测试集上的结果对比. 可以看出, 本文算法在 HOTA、MOTA 和 IDF1 指标上优于大多数主流跟踪算法方法. 另外一点值得说明的是, 本文算法的 IDS 优于大多数主流方法, IDS 越小, 说明跟踪轨迹的目标身份跳变次数越少, 跟踪结果的可靠性越高.

3　结　论

本文旨在设计多目标跟踪框架的特征增强模块和数据关联模块, 提出了一种 ESPPNet 特征增强模块和基于长短期超图神经网络匹配的端到端多目标跟踪方法. 通过聚合低纬度的检测特征和高纬度的 Re-ID 嵌入特征, 缓解了目标检测任务和 Re-ID 特征存在的不平衡问题; 通过超图神经网络对数据关联过程进行建模, 利用长期超图匹配模块和短期超图匹配模块分别关联被遮挡的目标和未被遮挡的目标, 较好地保持了跟踪目标的视觉一致性关系, 并实现了端到端的多目标跟踪. 实验结果表明, 本文所提出的 HGTracker 跟踪器在 MOT17 和 MOT20 两个主流数据集上, 与当前相关的主流方法相比具有较好的鲁棒性.

参考文献 (References)

[1] Bewley A, Ge Z Y, Ott L, et al. Simple online and realtime tracking[C]. 2016 IEEE International Conference on Image Processing. Phoenix, 2016: 3464-3468.

[2] Bochinski E, Eiselein V, Sikora T. High-speed tracking-by-detection without using image information[C]. The14th IEEE International Conference on Advanced Video and Signal Based Surveillance. Lecce, 2017: 1-6.

[3] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]. 2017 IEEE International Conference on Image Processing. Beijing, 2017: 3645-3649.

[4] Voigtlaender P, Krause M, Osep A, et al. MOTS: Multi-object tracking and segmentation[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 7942-7951.

[5] 朱姝姝, 王欢, 严慧. 基于帧内关系建模和自注意力融合的多目标跟踪方法[J]. 控制与决策, 2023, 38(2): 335-344.
(Zhu S S, Wang H, Yan H. Multi-object tracking based on intra-frame relationship modeling and self-attention fusion mechanism[J]. Control and Decision, 2023, 38(2): 335-344.)

[6] Zhang Y F, Wang C Y, Wang X G, et al. FairMOT: On the fairness of detection and re-identification in multiple object tracking[J]. International Journal of Computer Vision, 2021, 129(11): 3069-3087.

[7] Wang R Z, Yan J C, Yang X K. Neural graph matching network: Learning lawler's quadratic assignment problem with extension to hypergraph and multiple graph matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(9): 5261-5279.

[8] Zass R, Shashua A. Probabilistic graph and hypergraph matching[C]. 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, 2008: 1-8.

[9] Dendorfer P, Ošep A, Milan A, et al. MOTChallenge: A benchmark for single-camera multiple target tracking[J]. International Journal of Computer Vision, 2021, 129(4): 845-881.

[10] Milan A, Leal-Taixé L, Reid I, et al. MOT16: A benchmark for multi-object tracking[J/OL]. 2016, arXiv: 1603.00831.

[11] Dendorfer P, Rezatofighi H, Milan A, et al. MOT20: A benchmark for multi object tracking in crowded scenes[J/OL]. 2020, arXiv: 2003.09003.

[12] Zhou X Y, Koltun V, Krähenbühl P. Tracking objects as points[C]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 474-490.

[13] Zheng L Y, Tang M, Chen Y Y, et al. Improving multiple object tracking with single object tracking[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 2453-2462.

[14] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 213-229.

[15] Han S D, Huang P, Wang H W, et al. MAT: Motion-aware multi-object tracking[J]. Neurocomputing, 2022,

476: 75-86.

[16] Wang Y X, Kitani K, Weng X S. Joint object detection and multi-object tracking with graph neural networks[C]. 2021 IEEE International Conference on Robotics and Automation. Xi'an, 2021: 13708-13715.

[17] Cai J R, Xu M Z, Li W, et al. MeMOT: Multi-object tracking with memory[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 8090-8100.

[18] Chan S X, Jia Y W, Zhou X L, et al. Online multiple object tracking using joint detection and embedding network[J]. Pattern Recognition, 2022, 130: 108793.

[19] He J W, Huang Z H, Wang N Y, et al. Learnable graph matching: Incorporating graph partitioning with deep feature learning for multiple object tracking[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 5299-5309.

[20] Liang C, Zhang Z, Zhou X, et al. Rethinking the competition between detection and ReID in multiobject tracking[J]. IEEE Transactions on Image Processing, 2022, 31: 3182-3196.

[21] Peng J L, Wang C G, Wan F B, et al. Chained-tracker: Chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking[C]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 145-161.

[22] Meinhardt T, Kirillov A, Leal-Taixe L, et al. TrackFormer: multi-object tracking with transformers[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 8844-8854.

[23] Tokmakov P, Li J, Burgard W, et al. Learning to track with object permanence[C]. 2021 IEEE/CVF International Conference on Computer Vision. Montreal, 2021: 10860-10869.

[24] Hyun J, Kang M, Wee D, et al. Detection recovery in online multi-object tracking with sparse graph tracker[C]. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, 2023: 4850-4859.

[25] Wang Q, Zheng Y, Pan P, et al. Multiple object tracking with correlation learning[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 3876-3886.

[26] Yu E, Li Z L, Han S D, et al. RelationTrack: Relation-aware multiple object tracking with decoupled representation[J]. IEEE Transactions on Multimedia, 2023, 25: 2686-2697.

[27] Stadler D, Beyerer J. Modelling ambiguous assignments for multi-person tracking in crowds[C]. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. Waikoloa, 2022: 133-142.

[28] Lee S H, Park D H, Bae S H. Decode-MOT: How can we hurdle frames to go beyond tracking-by-detection?[J]. IEEE Transactions on Image Processing, 2023, 32: 4378-4392.

作者简介

郭文 (1978−), 男, 教授, 博士, 主要研究方向为计算机视觉、多媒体计算, E-mail: wguo@sdtbu.edu.cn;

刘其贵 (1998−), 男, 硕士生, 主要研究方向为计算机视觉, E-mail: 2021420049@sdtbu.edu.cn;

王拓 (1996−), 男, 硕士生, 主要研究方向为计算机视觉, E-mail: 2023410093@sdtbu.edu.cn;

丁昕苗 (1979−), 女, 教授, 博士, 主要研究方向为计算机视觉、视频理解, E-mail: dingxinmiao@126.com.