

DOI: 10.19650/j.cnki.cjsi.J2311717

基于多阶段关联的多目标跟踪算法*

霍旭^{1,2}, 盖绍彦^{1,2}, 洪濡^{1,2}, 周伟典^{1,2}, 达飞鹏^{1,2}

(1. 东南大学自动化学院 南京 210096; 2. 东南大学复杂工程系统测量与控制教育部重点实验室 南京 210096)

摘要:针对现有多目标跟踪算法关联过程中,外观和几何信息利用不充分,同时跟踪对象的邻域间信息交互不足的问题,提出了一种基于多阶段关联的多目标跟踪算法,根据目标之间的不同关联状态,将几何信息和外观信息合理应用于不同关联阶段。算法提出了基于正则化距离交并比(DIoU-Mea)的匹配模块,仅利用几何信息快速将强关联目标匹配。同时基于稀疏图网络(GNN)的关联模块对跟踪对象的邻域建模,促进对象之间的信息交换并提高跟踪精度。基于通道注意力融合特征模型和形状交并比的双校验模块(Double-Revise)进一步细化跟踪结果。所提算法利用不同阶段匹配算法的互补优势,在各阶段合理利用外观和几何信息,过滤掉错误的匹配并识别正确的目标对应关系,在MOT17数据集上进行了验证与测试,其高阶跟踪精度(HOTA)在测试集中达到了64.8%,表明算法具有较好的性能,在密集场景下具有较好的鲁棒性。

关键词: 机器视觉; 多目标跟踪; 图网络; 注意力机制

中图分类号: TP391.4 TH865 **文献标识码:** A **国家标准学科分类代码:** 520.20

Multi-object tracking algorithm based on multi-stage association

Huo Xu^{1,2}, Gai Shaoyan^{1,2}, Hong Ru^{1,2}, Zhou Weidian^{1,2}, Da Feipeng^{1,2}

(1. School of Automation, Southeast University, Nanjing 210096, China; 2. Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: Existing multi-object tracking algorithms make insufficient use of appearance and geometric information, and the information exchange among adjacent regions of the tracked object is limited. To solve this problem, a multi-object tracking algorithm based on multi-stage association is proposed, which applies geometric and appearance information to different association stages according to the different association states among objects. Firstly, a fast matching module based on the regularized distance intersection and union ratio (DIoU-Mea) is employed to efficiently handle the matching task of strongly correlated objects only using geometric information. Secondly, an association module based on the sparse graph network (GNN) is incorporated to model the neighborhood of the tracked object, facilitate information exchange among objects, and improve tracking accuracy. Finally, a double verify module (Double-Revise) is introduced, which utilizes the channel attention fusion feature model and the shape intersection and union ratio to further refine the tracking results. By utilizing the complementary advantages of different stage matching algorithms and making reasonable use of appearance and geometric information in each stage, the proposed algorithm effectively filters out incorrect matches and accurately identifies the correct object correspondence. The proposed algorithm is evaluated and tested on the MOT17 dataset. Its high-order tracking accuracy (HOTA) reaches 64.8% on the test set. Results show its good performance and robustness in dense scenarios.

Keywords: machine vision; multi object tracking; graph network; attention mechanism

0 引言

近年来智能驾驶、视觉导航不断发展,感知算法在智

能驾驶中的重要性日益增强,而多目标跟踪算法是其中重要的部分之一^[1-3]。如今多目标跟踪算法进展迅速,但在长时跟踪下,目标密集区域的遮挡导致跟踪器不鲁棒,同时速度和精度的平衡仍然面临着挑战。

过去几年中,高性能的检测器被多次提出,因此基于检测的跟踪范式成为了有效策略。该策略依赖于检测器提供的检测结果,跟踪部分对不同帧的同一对象分配身份标签,然而由于遮挡的存在,当前帧身份标签分配错误会给跟踪器的后续分配造成严重干扰,跟踪器的鲁棒性无法得到保证。

大多数现有工作集中于改进特征生成和数据融合。文献[4]遵循基于检测的跟踪范式,将跟踪部分分为两阶段,依据得分将检测框划分为高分框和低分框两类,分别进行跟踪,利用低分检测框和跟踪轨迹之间的相似性,来恢复真实的对象,并过滤掉背景检测。不同于文献[4]使用交并比度量轨迹和检测相似性的方式,文献[5]提取一段短时帧序列的检测的特征,对所有特征进行编码,并使用轨迹查询分组为全局轨迹。而文献[6]试图从全局关联角度对多目标跟踪进行建模,构建了无向图网络,在每一帧中选择得分最高的 k 个检测,构建图节点和连接边,以边特征计算关联度并通过节点评分对候选关联对进行验证。针对外观特征的工作,文献[7]并行预测目标的中心和位置以及以像素为中心进行外观特征提取,通过改进外观特征提高了跟踪性能。文献[8]使用目标外观进行轨迹生成,利用短时接近的对象实例具有相似外观的先验,以空间和时间先验作为加权因子,将跟踪阶段分为初始轨迹,聚类 and 融合轨迹两阶段。文献[9]使用外观分支和运动分支的框架,针对外观分支,提出指数移动平均(exponential moving average, EMA)的方式更新轨迹的外观状态,对于运动分支,采用相机运动补偿,同时在匹配时使用外观和运动加权的代价矩阵。

上述方法均使用了目标外观特征和交并比度量,并试图通过更准确的特征表达和关联代价矩阵提高算法性能,但大多在单一阶段利用外观或几何信息,而在关联阶段进行外观和几何分支的矩阵融合,在进行匈牙利匹配

前仅使用轨迹预测框和检测框的交并比对融合矩阵过滤,并未考虑形状这一重要几何特征。同时模型的构建局限于单一轨迹和单一检测的特征提取。因此关联过程中,对于外观特征和几何特征的利用不够充分,同时目标邻域间信息交互不足,导致算法在密集人群和极端遮挡场景下缺乏鲁棒性。

针对以上问题,本文提出了基于多阶段关联的多目标跟踪模型,针对强关联目标仅使用几何信息匹配,后续使用图网络(graph neural networks, GNN)进行集群建模,图节点特征构建同时利用外观信息和几何信息,并且在对潜在错误匹配过滤时,引入外观信息和形状特征。具体为针对不同关联度的关联对设计模块,对强关联对象使用基于距离交并比(distance intersection over union, DIoU^[10])的正则化距离交并比(distance intersection over union measurement, DIoU-Mea)模块,针对难以分辨的关联对象设计稀疏图网络模块进行匹配并使用双校验模块进行错误过滤,旨在利用多段算法的优势互补,处理不同场景下的跟踪。最终,在MOT17数据集上进行验证,实验表明所提算法能取得更高的精度和更好的鲁棒性。

1 跟踪匹配算法

本文提出的多目标跟踪算法以高精度检测器的检测结果为输入,以文献[4]的级联匹配结构为基础,即第1阶段关联高置信度的检测,第2阶段关联低置信度的检测。针对第1阶段,本文提出了3个模块来优化多目标跟踪算法,如图1所示(虚线框),通过允许多阶段跟踪关联模型级联式处理不同跟踪难度的目标,来平衡多目标跟踪算法的效率与精度。

在当前 t 帧处存在一组 M 个轨迹对象,和一组 N 个检测对象。首先,基于二维检测中的DIoU提出了DIoU-

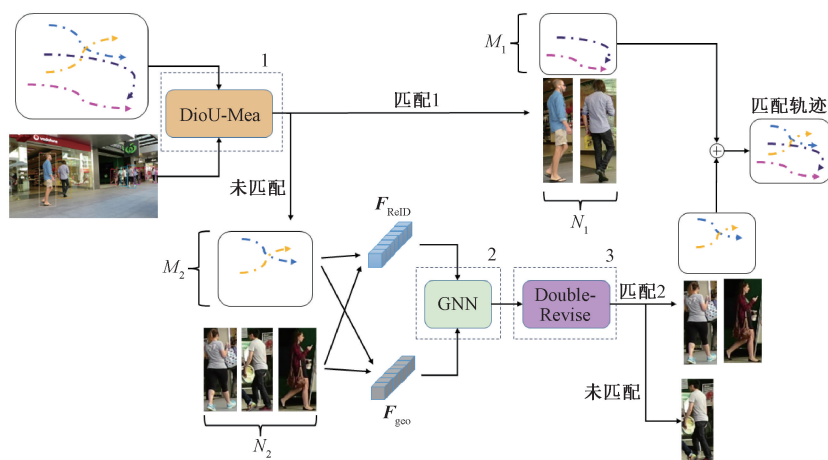


图1 整体算法流程

Fig. 1 Overall algorithm process

Mea 模块(虚线框 1),该模块同时考虑了候选项以及干扰项在轨迹预测位置中的影响,通过控制关联阈值使强几何关联的目标匹配。经过 DIoU-Mea 模块匹配后, $t-1$ 帧中的 M_1 个轨迹对象和 t 帧中的 N_1 个检测对象已匹配, $t-1$ 帧中未匹配的 M_2 个轨迹对象和 t 帧中未匹配的 N_2 个检测对象将由稀疏图网络(虚线框 2)和双校验模块(虚线框 3)匹配。

模块 2 为 GNN 的关联模型,将多目标跟踪建模为概率模型,利用目标邻域的外观信息 F_{RelD} 和几何信息 F_{geo} , 处理涉及复杂遮挡或者检测丢失的跟踪目标,增强算法在复杂场景下的准确性和鲁棒性。

模块 3 为双校验机制(double revise mechanism, Double-Revise),旨在修正模块 2 中可能存在的错误匹配。基于通道注意力的融合特征模型和引入了形状信息的形状交并比(shape intersection over union, IoU-Shape), 将两个目标是否为同一身份建模为二分类模型,输出校验矩阵对模块 2 输出的匹配矩阵过滤,降低目标关联时的歧义。

本文算法将高置信度检测目标的跟踪分为多个阶段进行,利用不同阶段匹配算法的互补优势,以过滤错误的匹配并识别正确的目标对应关系,而对低检测置信度的目标的跟踪与文献[4]相同。

1.1 基于 DIoU-Mea 矩阵的强关联匹配

同一目标在帧间空间位置的关联性使得跟踪算法使用简单的运动模型即可描述目标之间的相似度。当空间位置的唯一关联性达到较高水平时,基于几何信息的关联模型就能够实现快速高精度的跟踪。为了平衡关联速度和精度,本文首先提出一种候选检测的几何干扰度量 DIoU-Mea,对多个跟踪对象的强关联部分进行快速匹配,实现匹配降维。

DIoU-Mea 公式为:

$$\begin{cases} D_{ij}^{\text{Mea}} = D_{ij} - \max\{D_{ik}\} & k = 1, 2, \dots, N, k \neq j \\ D_{ij} = \frac{S_i}{S_u} - \frac{d^2}{c^2} & i = 1, 2, \dots, M; j = 1, 2, \dots, N \end{cases} \quad (1)$$

式中: D_{ij} 表示编号为 i 的轨迹预测框与编号为 j 的候选检测框的 DIoU^[10] 计算值; S_i 为轨迹预测框和检测框的交集面积; S_u 为并集面积; d 为两框的最小外接矩形对角线长度; c 为两框的中心点距离; M 为轨迹框的数量; N 为检测框的数量; D_{ij}^{Mea} 用于表征多目标跟踪场景中轨迹的候选检测之间的几何干扰强度。

轨迹预测框与候选检测框之间的几何相似性可用 DIoU 描述,当检测框与轨迹预测框重叠程度越大, DIoU 越大。当轨迹预测框有多个重叠的候选检测框时,依靠 DIoU 无法准确判断轨迹与候选检测是否属于同一目标。反之,轨迹预测框仅有 1 个高重叠的候选检测框时,可以

直接将轨迹与检测匹配。为此 DIoU-Mea 描述符用于计算目标的预测位置与检测框的关联性强弱。如图 2(a) 所示,当轨迹预测框仅存在一个重叠程度较大的候选检测框,直接关联,图 2(b) 表示编号为 i 的轨迹与编号为 j 的检测为不相关对,图 2(c) 表明轨迹存在多于 1 个候选检测,此时输入后续阶段进行判断。

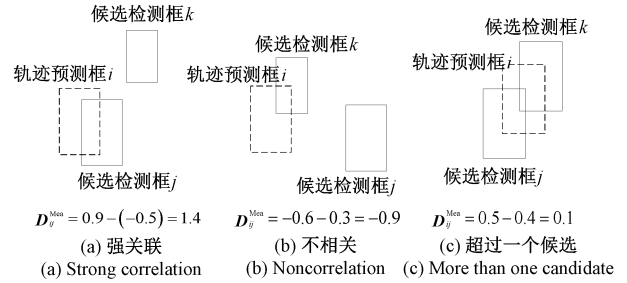


图 2 DIoU-Mea 示例说明

Fig. 2 Explanation of the example of DIoU-Mea

如图 3 步骤(1)~(3)所示,图块颜色越深表示数值越大,首先通过设定阈值 δ_1 和 δ_2 以区分强关联对和强不相关对,具体为在 DIoU-Mea 矩阵中大于 δ_2 的视为关联正例,设置为 1(图 3 步骤(2)中虚线),小于 δ_1 的视为关联负例,设置为 0,在 δ_1 和 δ_2 之间的视为不确定关联对。如图 3 步骤(2)所示(虚线框),预处理之后对矩阵中被设置为 1 的关联对依次搜索,当且仅当所在行与列只存在唯一的大于 0 项,即认为轨迹只存在唯一的候选检测,该对匹配是强关联匹配对,直接对其进行关联。如图 3 步骤(3)所示,经过 DIoU-Mea 模块之后,关联矩阵由 $M \times N$ 维降维为 $M_2 \times N_2$ 维。对于未匹配的 $M_2 \times N_2$ 不确定关联对,输入模块 2 稀疏图网络进行关联,同时使用模块 3 双校验机制对模块 2 结果进行修正。

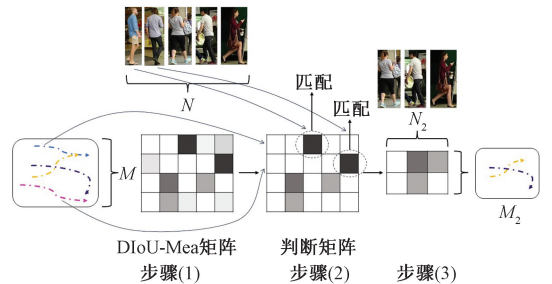


图 3 DIoU-Mea 模块匹配流程

Fig. 3 DIoU-Mea module matching process

1.2 稀疏图网络

对于密集区域的目标互相遮挡和干扰而导致无法有效跟踪的问题,尽管现有的多目标跟踪器应用运动预测器(例如卡尔曼滤波器)来预测丢失轨迹的可能位置,但在长期跟踪下,此类运动预测器的预测结果由于误差累

积而不准确,导致基于几何位置的关联表现为假阳性或假阴性。传统的方法对图像进行裁剪并使用卷积神经网络提取特征,计算特征欧氏距离来描述目标的相似性,此类方法仅仅涉及单个轨迹对象和检测对象,不具有集群对象的建模能力。

为处理此问题,本文采用了图网络对轨迹和检测对象的交互关系进行构建,以获取目标邻域间的关联特征,将邻域目标之间的关联视为概率模型,利用被遮挡目标和其邻域对象间潜在的关联信息来恢复因遮挡而丢失的被跟踪对象。然而,全连接图模型的数据处理量过大,尽管在模块①中已经对匹配进行降维,但此类计算开销仍然影响多目标跟踪器的速度,因此本文首先使用 DIoU-Mea 对关联预选,得到稀疏图模型减少计算开销。

1) 图构建

如图 4 所示,模块 1 对匹配维度降维后, $t-1$ 帧仍然有 M_2 个未匹配跟踪对象, t 帧有 N_2 个未匹配检测对象。首先,本文对未匹配的轨迹对象和检测对象构建图节点,图 4 中共有 M_2 个轨迹节点和 N_2 个检测节点,图的特征包括几何信息和外观信息。接着,本文针对在不同帧的节点对构建无向边,且为了减少计算开销,本文构建稀疏图而非全连接图。对于 $t-1$ 帧已经跟踪的轨迹对象,其在 t 帧的候选检测对象通常位于其临近区域,因此本文设置节点对间的 DIoU-Mea 阈值,对跟踪对象的候选检测对象过滤,仅满足阈值的节点对才被视为潜在关联对,构建关联边。这种关联是双向的,因此轨迹对象和检测对象的特征都会更新,而不是单一地更新当前帧的检测对象特征。相比于仅利用单一匹配对之间的特征进行概率估计,这种方式可以构建集群间的关系,更新丢失对象的周围信息,以对抗因密集干扰产生的错误匹配。最后,经过两层节点特征聚合更新,计算节点间的 $\cos$ 相似度并归一化,以获得节点间的相似度矩阵。

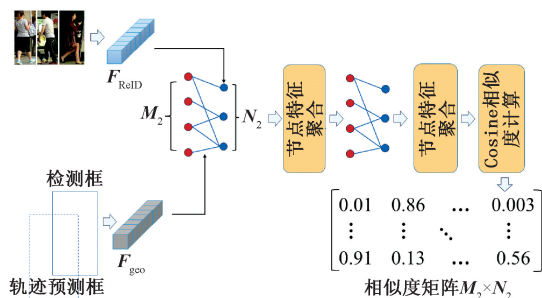


图 4 稀疏图网络概览

Fig. 4 Overview of the sparse graph network

2) 节点聚合与边回归

为了应对遮挡导致的轨迹丢失,本文对于图节点特征和边特征的定义同时考虑了几何信息和外观信息。其中外观特征 $F_{\text{ReID}} \in \mathbf{R}^{n \times 2.048}$ 使用 FastReID^[11] 提取, $t-1$

帧的第 i 个轨迹对象和 t 帧的第 j 个候选检测的几何特征 $F_{\text{geo}} \in \mathbf{R}^{n \times 8}$ 定义如下:

$$F_{\text{geo}} = \left[\frac{x_j^{(t)} - x_i^{(t-1)}}{(w_i^{(t-1)} + w_j^{(t)})/2}, \frac{y_j^{(t)} - y_i^{(t-1)}}{(h_i^{(t-1)} + h_j^{(t)})/2}, \log_2 \left(\frac{w_j^{(t)}}{w_i^{(t-1)}} \right), \log_2 \left(\frac{h_j^{(t)}}{h_i^{(t-1)}} \right), \frac{x_j^{(t)} - x_i^{(t-1)}}{w^{(t)}}, \frac{y_j^{(t)} - y_i^{(t-1)}}{h^{(t)}}, \frac{w_j^{(t)} - w_i^{(t-1)}}{w^{(t)}}, \frac{h_j^{(t)} - h_i^{(t-1)}}{h^{(t)}} \right] \quad (2)$$

式中: $x_j^{(t)}$ 和 $y_j^{(t)}$ 分别是 t 帧第 j 个候选检测框中心点横坐标和纵坐标; $w_j^{(t)}$ 和 $h_j^{(t)}$ 分别是 t 帧第 j 个候选检测框的宽度和高度; $w^{(t)}$ 和 $h^{(t)}$ 分别是 t 帧图片的宽度和高度。

边特征 F_{edge} 定义为:

$$F_{\text{edge}} = f_{\text{MLP}}[f_{\text{MLP}}(|F_{\text{ReID}}^{(t-1)}, F_{\text{ReID}}^{(t)}|), f_{\text{MLP}}(F_{\text{geo}})] \quad (3)$$

式中: $| \cdot, \cdot |$ 表示特征拼接, $F_{\text{ReID}}^{(t-1)}$ 和 $F_{\text{ReID}}^{(t)}$ 分别表示 $t-1$ 帧轨迹目标和 t 帧候选检测目标的外观特征, f_{MLP} 为多层感知机层。

为了在 GNN 中对特征交互建模,本文输入以 F_{ReID} 构建的节点特征,以 F_{edge} 构建的边特征,通过两层图网络卷积层 (graph convolution network convolution layer, GCNConv^[12]) 迭代更新相连的节点及边的特征,如图 4 所示。GNN 通过消息传递过程更新图节点和边的特征,在迭代更新过程中,当前节点包含了相邻节点和边的特征,而边则间接获得了与同一节点相连的其他边的信息,经过几轮迭代,集群信息得到广泛传播,即:

$$F_{\text{node}} = f_{\text{conv}}[f_{\text{conv}}(F_{\text{ReID}}, F_{\text{edge}}), F_{\text{edge}}] \quad (4)$$

式中: f_{conv} 表示 GCNConv 层; F_{node} 表示经过 GCNConv 更新后的节点特征。

最终,使用节点特征计算 $\cos$ 相似度并进行 Sinkhorn 归一化,对两节点是否为同一对象进行概率预测,输出概率矩阵。模型的输出概率 P 为:

$$P = f_{\text{S}}[f_{\text{cos}}(F_{\text{node}}^{\text{tra}}, F_{\text{node}}^{\text{det}})] \quad (5)$$

式中: f_{cos} 表示特征的 $\cos$ 相似度计算; f_{S} 表示 Sinkhorn 归一化; $F_{\text{node}}^{\text{tra}}$ 和 $F_{\text{node}}^{\text{det}}$ 分别表示两层 GCNConv 更新后的轨迹节点特征和检测节点特征。最后使用模块 3 得出的双屏蔽矩阵对概率矩阵进行校验,以过滤可能的假阳性和假阴性结果,获得更准确的概率矩阵,屏蔽后的概率值将与归一化的 DIoU-Mea 值相乘,以进行数据融合,并使用匈牙利算法输出匹配对。

在模型训练阶段,本文将概率矩阵直接提供给稀疏图模型的目标函数,这涉及最小化以下加权二元交叉熵损失:

$$L = - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n w \cdot G_{ij} \cdot \log_2(P_{ij}) + (1 - G_{ij}) \cdot \log_2(1 - P_{ij}) \quad (6)$$

式中: P_{ij} 表示由式(5) 计算的检测 D_i 与轨迹 T_j 的匹配概率; G_{ij} 表示第 i 个检测是否属于第 j 个轨迹的真值; w 是平衡正例和负例之间损失的权重超参数。

1.3 基于通道注意力融合特征模型和形状交并比的双校验机制

由于多目标跟踪中的遮挡问题,且运动模型在长时跟踪下存在误差累积而使轨迹对象的预测结果不准确时,可能导致跟踪器产生假阳性或假阴性匹配对。在遮挡后,由于运动预测不准确,即使轨迹预测框和候选目标检测框之间的交并比(intersection over union, IoU)为低值,关联对仍然可能为关联正例,此时 IoU 的结果置信度偏低,因此本文首先构建通道注意力的融合特征分类模型,同时考虑几何和外观特征,模型侧重于对不同通道的

特征进行注意力融合,以提高重点区域的特征权重。

同时,远距离目标和被遮挡的目标提取的外观特征并不可靠,基于融合特征模型的结果仍然需要改进,因此本文使用双校验机制,有助于改善仅置信单一结果导致的错误屏蔽,减少错误结果对跟踪精度的影响。

由于在该阶段进行匹配的对象通常存在遮挡或者与相机距离过远,导致提取的外观特征并不可靠,因此本文对于特征的定义同时考虑了几何信息和外观信息。构建的高效注意力(efficient channel attention, ECA^[13])融合特征模型如图 5 中①所示,输入为外观特征 $F_{\text{ReID}} \in \mathbf{R}^{n \times 2048}$,以及第 i 个轨迹对象和其第 j 个候选检测的几何特征 $F_{\text{geo}} \in \mathbf{R}^{n \times 8}$,定义与式(2)相同。图 5 中 BC 为二分类网络,MAX 为 max 运算。

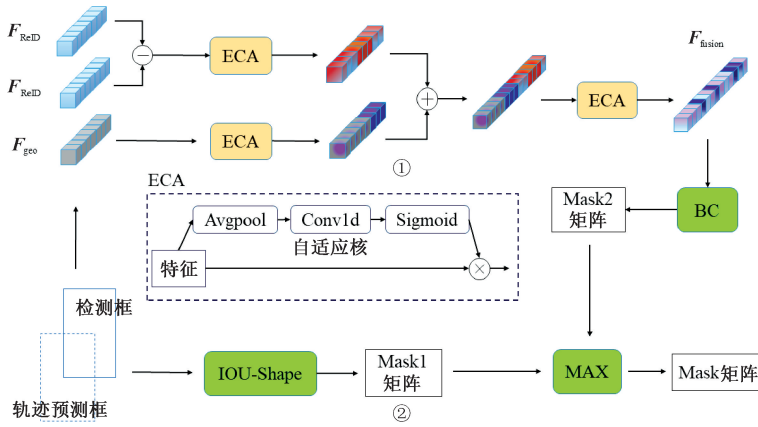


图 5 双校验模块总览

Fig. 5 Overview of the double-reverse module

注意力融合特征模型为:

$$\mathbf{A}_{\text{fusion}} = f_{\text{BC}} \{ f_{\text{ECA}} \{ | f_{\text{ECA}}(\mathbf{F}_{\text{ReID}}), f_{\text{ECA}}[f_{\text{MLP}}(\mathbf{F}_{\text{geo}})] | \} \} \quad (7)$$

式中: f_{MLP} 对 $\mathbf{F}_{\text{geo}}$ 升维为 $\mathbf{R}^{n \times 512}$; f_{ECA} 为 ECA 层,对通道权重调整后输出融合特征 $\mathbf{F}_{\text{fusion}} \in \mathbf{R}^{n \times 2560}$; f_{BC} 为二分类网络层, $\mathbf{A}_{\text{fusion}}$ 为轨迹目标和候选检测目标为同一目标的概率矩阵。

ECA 模块用于调整提升关键通道的特征权重,具体如图 5 所示(虚线框),公式为:

$$\mathbf{F} = \sigma \{ f_{\text{conv1}} [f_{\text{gavg}}(\mathbf{F}_{\text{ori}})] \} \cdot \mathbf{F}_{\text{ori}} \quad (8)$$

式中: $\mathbf{F}_{\text{ori}}$ 为原始特征; f_{gavg} 为全局局池化; σ 为 Sigmoid 激活函数; f_{conv1} 为自适应卷积核的一维卷积。其中卷积核 k 表示为:

$$k = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor \quad (9)$$

式中: C 为输入的通道数; $b = 1$; $\gamma = 2$ 。

在行人跟踪中,根据经验,处于跟踪状态的目标在相邻两帧的检测框形状接近,传统的 IoU 的定义为两个边

界框相交面积和并集面积的比值,并未包含形状的相似计算,因此,本文使用引入形状属性的 IoU-Shape,以提高模型对 IoU 较低但形状接近目标的辨识度。IoU-Shape 的定义如下:

$$S_{\text{IoU}} = \frac{S_I}{S_U} + \frac{\min(w_{t-1}, h_t, w_t, h_{t-1})}{\max(w_{t-1}, h_t, w_t, h_{t-1}) + \varepsilon} \quad (10)$$

式中: S_I 为轨迹框和检测框的交集面积; S_U 为轨迹框和检测框的并集面积; w_{t-1} 和 w_t 分别为 $t-1$ 帧轨迹框和 t 帧检测框的宽度; h_{t-1} 和 h_t 分别为 $t-1$ 帧轨迹框和 t 帧中检测框的高度; ε 为防止分母为 0 的常数,取 10^{-7} 。

如图 5 中②所示,注意力模型输出的二分类屏蔽矩阵 $\mathbf{A}_{ij}^{\text{Mask2}}$ 将与按 IoU-Shape 获得的屏蔽矩阵 $\mathbf{A}_{ij}^{\text{Mask1}}$ 进行双过滤运算。双过滤机制可以表示为:

$$\mathbf{A}_{ij}^{\text{Mask}} = \max(\mathbf{A}_{ij}^{\text{Mask1}}, \mathbf{A}_{ij}^{\text{Mask2}}) \quad (11)$$

式中: $\mathbf{A}_{ij}^{\text{Mask}}$ 矩阵中元素仅为 1 或 0,用于对 1.2 节中图网络的输出矩阵进行校验,0 表示屏蔽,1 表示不屏蔽(图 6 虚线)。

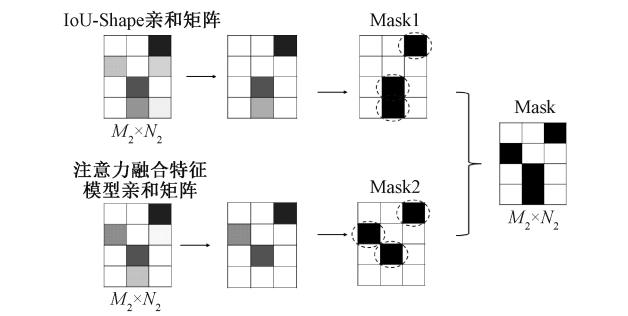


图 6 双校验机制算例

Fig. 6 Example of double-reverse mechanism calculation

如图 6 所示,图块颜色越深表示数值越大,首先对 IoU-Shape 矩阵中低于阈值 δ_3 的进行屏蔽(设置为 0),高于 δ_3 设置为 1,接着对融合模型所得矩阵中低于阈值 δ_4 的进行屏蔽,当且仅当两个屏蔽矩阵都判断该匹配对是负例时(即同时为 0),才将其屏蔽,最终得到该校验矩阵将用于屏蔽稀疏图模型得出的概率结果,以减少错误匹配。同时,本文在消融实验中验证了该双屏蔽策略对跟踪性能的提升。

2 实验结果及分析

2.1 数据集与评价指标

模型的训练和测试在 MOT17 数据集上进行。MOT17 数据集包含了用静态和移动相机拍摄的视频序列,分为训练集和测试集,无验证集,训练集包含 02、04、05、09、10、11、13 号共 7 个序列,测试集包含训练集

包含 01、03、06、07、08、12、14 号共 7 个序列。由于 MOT17 数据集未提供验证集,为了进行消融实验,将训练集重新划分,其中每个序列的前半部分用于训练,后半部分用于验证。

评估指标采用综合评价指标高阶跟踪精度 (higher order tracking accuracy, HOTA^[14]) 以及 CLEAR^[15] 指标,包括多目标跟踪精度 (multi-object tracking accuracy, MOTA)、身份切换数目 (identity switches, IDSW) 和 IDF1 得分。

2.2 实验细节

本文实验在 CPU 为 Intel(R) Core(TM) i5-10 600KF CPU @ 4.10 GHz,以及 GPU 为 NVIDIA GeForce RTX 3 070 的桌面端运行。对于检测,本文使用了一阶段目标检测模型 (you only look once X, YOLOX^[16]) 的检测结果,本文选择的最佳参数为模块 1 的 DIoU-Mea 强关联阈值 δ_2 设置 0.7,无关联阈值 δ_1 设置为 0,模块 2 中图卷积为两层,模块 3 中 IoU-Shape 矩阵的屏蔽阈值 δ_3 为 0.3,特征融合矩阵的屏蔽阈值 δ_4 为 0.5。

2.3 消融实验

1) 组件分析

消融实验旨在验证 3 个模块对算法性能的提升,以及量化各个模块对结果的影响,所有实验在 MOT17 划分的验证集上进行,使用相同的检测结果和相同的参数设置。

MOT17 验证集上从基线模型到完整模型的跟踪评估结果如表 1 所示,并以帧率说明各个模块对于模型运行速度的影响,√表示使用了该模块,基线模型未受到 3 个附加模块的影响。

表 1 在 MOT17 验证集上不同模块的消融实验结果
Table 1 Results of ablation experiments on different modules on the MOT17 validation set

模块 1	模块 2	模块 3	HOTA/%	MOTA/%	IDF1/%	IDSW	帧率/fps
√			76.909	90.085	83.791	367	31.5
	√		67.526	74.390	72.168	2 371	2.0
√	√		78.247	90.169	86.287	268	9.7
√	√	√	78.919	90.501	87.114	232	7.9

由表 1 可知,仅使用模块 1 (DIOU-Mea) 时,算法能够快速对强关联的对象匹配,但由于无外观信息,模型在密集和遮挡场景下跟踪精度不佳。仅使用模块 2 (GNN) 时,由于模型完全依赖于特征的有效性,此时计算开销增加且缺乏鲁棒性,模型精度和速度均下降。在模块 1 基础上增加模块 2 时,模型精度和运行速度均提升,说明稀疏图网络模块对于遮挡问题具有较好的身份恢复和保持能力 (IDF1 提升, IDSW 下降), 增加了模型的鲁棒性,同时模块 1 可有效提高模型效率,相比于仅使用模块 2, 帧

率提高到 9.7 fps (+7.7 fps)。使用模块 1+2+3 时,由于模块 3 (Double-Reverse) 使用的通道注意力能够提升关键特征的权重,使模型能够处理部分弱关联对象,同时加入的形状特征提升了模型对于相邻帧遮挡的处理能力,将精度进一步提高,并且在运行速度小幅度下降的情况下将 IDSW 减少到 232,表明轨迹的稳定性增加。

模块 3 中不同的校验方式对于性能的影响,如图 7 所示。使用通道注意力融合特征模型校验将 HOTA 提高 0.126%, MOTA 提高 0.178%, IDF1 提高 0.037%, 减少了

错误匹配。相比于仅使用通道注意力融合特征模型校验,加入 IoU 校验将 HOTA 提高 0.511%, MOTA 提高 0.16%, IDF1 提高 0.597%, 将 IoU 替换为 IoU-Shape 之后, HOTA 提高 0.035%, IDF1 提高 0.193%, 表明形状特征减少了密集区域的错误匹配, 增强了模型的识别能力。

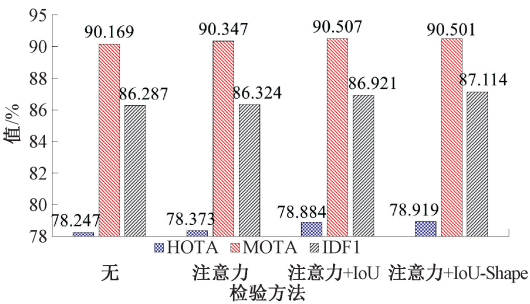


图 7 不同校验方式在 MOT17 验证集中的结果

Fig. 7 Results of different verification methods on the MOT17 validation set

2) 模型超参数分析

实验使用的模型为添加 3 个附加模块的完整模型, 超参数包括 DIOU-Mea 模块的强关联判别阈值 δ_2 , 图网络的 GCNConv 层数, 以及双校验模块中的 δ_3 和 δ_4 。不同超参数对结果的影响如表 2 所示, 在实验中调整一种超参数时保持其余超参数不变, 实验同样在 MOT17 验证集上进行。

表 2 超参数对模型性能的影响

Table 2 The influence of hyperparameters on model performance

超参数	数值	HOTA/%	MOTA/%	IDF1/%	IDSW
DIOU-Mea 阈值 δ_2	0.65	78.783	90.495	86.859	235
	0.70	78.919	90.501	87.114	232
	0.75	78.620	90.481	86.572	235
GCNConv 层数	1	78.602	90.455	87.033	244
	2	78.919	90.501	87.114	232
	3	78.600	90.514	86.856	241
δ_3	0.25	78.146	88.347	86.327	244
	0.30	78.919	90.501	87.114	232
	0.35	78.932	90.523	87.069	233
δ_4	0.45	78.919	90.501	87.114	232
	0.50	78.919	90.501	87.114	232
	0.55	78.751	90.345	86.736	237

δ_2 为区分强相关对的阈值, 只有大于 δ_2 的才被视为强相关, 该阈值决定了 DIOU-Mea 模块匹配的严格程度, 提高 δ_2 表明严格程度提升, 因此进入图网络和双校验模

块的轨迹和检测数量增加。当阈值为 0.65, 在 DIOU-Mea 模块中匹配的轨迹和检测较多, 但在复杂场景下错误跟踪增加, 当阈值为 0.7, HOTA 提升 0.136%, IDF1 提升 0.255%, 表明部分复杂场景下的轨迹和检测由图网络 and 校验模型解决, 提升了模型的性能, 而阈值提升到 0.75 时, 进入模块 2 和模块 3 的目标增多, 目标间外观特征相互干扰, 提升了模型的匹配难度, 模型性能下降。

由表 2 可知, GCNConv 为 1 层时, 模型的性能最低, 这种情况下, 模型对集群信息的特征表达不足, 无法准确进行二分类估计, 当层数增加到 2 层时, 4 项指标均提高, 说明模型对特征的表达能力提升, 而层数增加到 3 层时, 除 MOTA 略微提升外, 其余指标均下降, 此时模型参数数量增加, 增加了对训练数据的细节和噪声过拟合的风险, 泛化能力下降, 同时信息在较深的网络中传播时, 可能会逐渐丢失或变得模糊, 导致模型对信息的理解能力下降。 δ_3 和 δ_4 决定了双校验模块进行屏蔽时的严格程度, 在验证集中的最佳实验结果将被用于测试集。

2.4 可视化分析

本文对完整模型在 MOT17 部分验证集上的表现进行可视化, 更直观地分析模型对于遮挡问题的鲁棒性。

选取验证集中 09 和 10 序列下的部分跟踪场景如图 8 所示, 其中数字代表该目标被分配的身份标签。图 8(a) 中, 2 号目标在 421 帧开始被目标 8 和目标 14 遮挡, 期间轨迹丢失, 直到 453 帧模型将轨迹重新恢复, 且目标 8 和目标 14 的身份标签未发生切换, 图 8(b) 中 10 号目标在 264 帧开始被周围目标遮挡, 直到 288 帧目标仍然未发生身份切换, 表明模型在遮挡场景下能较好利用邻域信息, 保持轨迹身份标签的稳定。

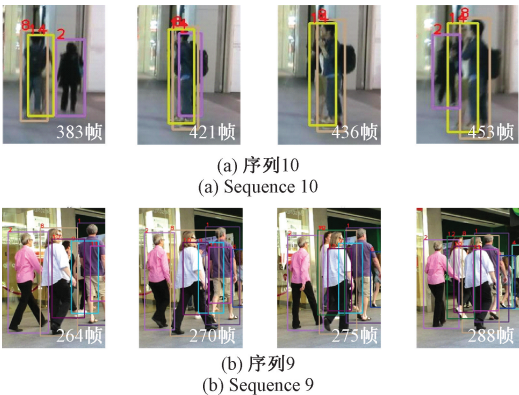


图 8 验证集中部分遮挡场景

Fig. 8 Partial occlusion scenarios on the validation set

3 个模块在整体流程中的匹配次数和匹配时间如图 9 所示, 分别为 02、04、05、09、10、11、13 号序列的结果。除 02 和 09 序列外, 其余序列中 DIOU-Mea 模块的匹配次数均高于 GNN 模块, 各序列中 GNN 模块的匹配时

间占比均为最高。所有序列的总体统计结果中,DIoU-Mea 模块的匹配次数占比接近 60%,同时 GNN 模块的匹配时间占比超过 60%,这是由于模型需要利用邻域的外

观和几何信息进行计算,因此效率低于 DIoU-Mea 模块,同时 Double-Revise 模块的运行时间占比低于 20%,模块在提升模型性能的同时平衡了效率。

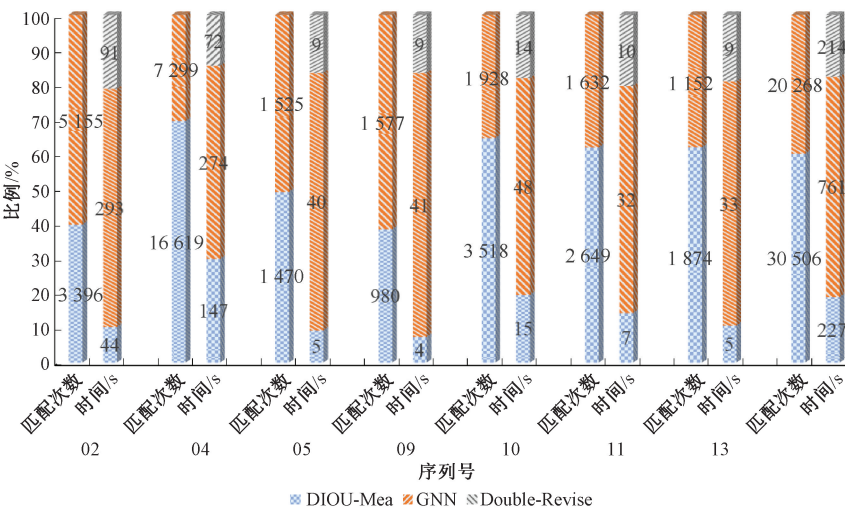


图 9 各模块在 MOT17 验证集中的匹配次数和时间

Fig. 9 The number and time of matching for each module on the MOT17 validation set

选取验证集中 10 序列下的夜晚街道的部分跟踪场景如图 10 所示。图 10 中,409 帧中 36 号目标在 410 帧未能被检测器成功检测,轨迹丢失,在 411 帧由于相机抖动导致图像模糊,目标发生了身份切换,589 帧中的目标在 590 帧图像严重模糊的情况下发生身份切换。原因可能为该类目标因夜晚光照不足,相机发生抖动导致图像严重模糊时,提取的外观特征不准确,而几何形状的相似也导致根据几何信息无法区分目标,从而匹配失败。



图 10 验证集中部分失败跟踪

Fig. 10 Partial failed tracking on the validation set

2.5 测试集对比结果

如表 3 所示,本文比较了近年来在 MOT17 测试集上执行的跟踪器,所有结果直接从官方的 MOT 挑战评估服务器上获得,黑色粗体是每项指标的最佳值。相比最近

的一些公开算法,本文模型在 4 项关键性能指标上都具有较好的表现,说明本文提出的算法的稳健性和全面的跟踪能力。本文方法侧重于解决在密集人群场景下因遮挡导致的身份切换问题,使跟踪器具有更强大的身份保持能力。因此,它会产生更准确的关联,并且在反映关联能力的指标(即 IDF1 达到 80.0%)中优于第二性能跟踪器,说明模型匹配关联组件的强大能力。

表 3 MOT17 测试集上“私有检测器”协议下方法的比较

Table 3 Comparison of the methods under the “private detector” protocol on the MOT17 test set					
方法	HOTA/%	MOTA/%	IDF1/%	IDSW	帧率/fps
文献[17]	54.5	73.2	62.2	4 616	1.0
文献[5]	59.1	75.3	71.5	2 859	19.6
文献[6]	60.6	76.3	72.4	4 578	62.5
文献[8]	62.6	76.7	77.7	1 737	4.9
文献[18]	63.2	78.0	77.5	1 950	29.0
文献[4]	63.1	80.3	77.3	2 196	29.6
文献[19]	64.3	80.0	79.5	1 272	35.5
文献[9]	64.4	79.6	79.5	1 194	7.1
文献[20]	64.1	81.1	79.7	1 455	1 109.8
文献[21]	64.4	80.0	79.8	1 356	142.1
文献[22]	64.0	81.8	78.7	1 431	1 355.5
本文	64.8	81.0	80.0	1 293	3.7

3 结 论

本文提出一种基于多阶段关联的 2 D 多目标跟踪算法,分析并验证了所提出的 DIoU-Mea 模块对于模型速度的提升,以及图网络匹配模块和双校验模块在密集遮挡场景下的有效性。在 MOT17 数据集上的实验结果表明,基于多阶段的关联方式能针对不同关联度的对象进行更加准确的匹配,所构建的 DIoU-Mea 模块能快速完成强关联目标的匹配,图网络模块能够较好地利用目标的邻域信息解决遮挡导致的身价切换问题,采用注意力机制能通过调整特征通道权重增强表达能力,且与增加形状描述的 IoU-Shape 组成双校验机制更能有效修正图网络模块匹配结果。

参考文献

- [1] 朱 浩, 李 鑫. 基于一致性点漂移的智能车视觉目标跟踪方法 [J]. 仪器仪表学报, 2022, 43 (10): 195-204.
ZHU H, LI X. Visual multi-object tracking method for intelligent vehicle based on coherent point drift [J]. Chinese Journal of Scientific Instrument, 2022, 43(10): 195-204.
- [2] 杨 梅, 贾 旭, 殷浩东, 等. 基于联合注意力孪生网络目标跟踪算法 [J]. 仪器仪表学报, 2021, 42 (1): 127-136.
YANG M, JIA X, YIN H D, et al. Object tracking algorithm based on siamese network with combined attention [J]. Chinese Journal of Scientific Instrument, 2021, 42(1): 127-136.
- [3] 余仁伟, 朱 浩, 蔡昌恺. 基于薄板样条函数的无人机多目标跟踪算法 [J]. 仪器仪表学报, 2021, 42 (3): 168-176.
YU R W, ZHU H, CAI CH K. Multi-object tracking algorithm for UAV based on the thin plate spline function [J]. Chinese Journal of Scientific Instrument, 2021, 42(3): 168-176.
- [4] ZHANG Y F, SUN P Z, JIANG Y, et al. Bytetrack: Multi-object tracking by associating every detection box [C]. Computer Vision-ECCV 2022: 17th European Conference, 2022: 1-21.
- [5] ZHOU X Y, YIN T W, KOLTUN V, et al. Global tracking transformers [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8771-8780.
- [6] HYUN J, KANG M, WEE D, et al. Detection recovery in online multi-object tracking with sparse graph tracker [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 4850-4859.
- [7] 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.
- [8] GIRBAU A, MARQUÉS F, SATOH S. Multiple object tracking from appearance by hierarchically clustering tracklets [C]. 33rd British Machine Vision Conference, 2022: 1-15.
- [9] DU Y H, ZHAO Z C, SONG Y, et al. Strongsort: Make deepsort great again [J]. IEEE Transactions on Multimedia, 2023: 1-14.
- [10] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [11] WANYONYI D, CELIK T. Open-source face recognition frameworks: A review of the landscape [J]. IEEE Access, 2022, 10(1): 50601-50623.
- [12] MA X X, WW J, XUE S, et al. A comprehensive survey on graph anomaly detection with deep learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2021(1): 1-32.
- [13] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [14] LUITEN J, OSEP A, DENDORFER P, et al. Hota: A higher order metric for evaluating multi-object tracking [J]. International Journal of Computer Vision, 2021, 129(2): 548-578.
- [15] BERNARDIN K, STIEFELHAGEN R. Evaluating multiple object tracking performance: The clear mot metrics [J]. EURASIP Journal on Image and Video Processing, 2008(1): 1-10.
- [16] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [17] XU Y H, BAN Y T, DELORME G, et al. TransCenter:

Transformers with dense representations for multiple-object tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(6): 7820-7835.

[18] CAO J K, PANG J M, WENG X S, et al. Observation-centric sort: Rethinking sort for robust multi-object tracking[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 9686-9696.

[19] REN H, HAN S D, DING H L, et al. Focus on details: Online multi-object tracking with diverse fine-grained representation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 11289-11298.

[20] YANG F, ODASHIMA S, MASUI S, et al. Hard to track objects with irregular motions and similar appearances? Make it easier by buffering the matching space [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 4799-4808.

[21] WANG S, YANG D, WU Y B, et al. Tracking game: Self-adaptative agent based multi-object tracking [C]. Proceedings of the 30th ACM International Conference on Multimedia, 2022:1964-1972.

[22] YOU S S, YAO H T, BAO B K, et al. UTM: A unified

multiple object tracking model with identity-aware feature enhancement [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 21876-21886.

作者简介



霍旭,2021 年于重庆大学获得学士学位,现为东南大学硕士研究生,主要研究方向为目标跟踪。
E-mail:huoxu@ seu. edu. cn

Huo Xu received his B.Sc. degree from Chongqing University in 2021. He is currently a master student at Southeast University. His main research interest is object tracking.



达飞鹏(通信作者),1998 年于东南大学获得博士学位,现为东南大学教授、博士生导师,主要研究方向为三维精密测量、三维精确识别、三维优化控制理论与技术。
E-mail:dafp@ seu. edu. cn

Da Feipeng (Corresponding author) received his Ph. D. degree from Southeast University in 1998. He is currently a professor and a Ph. D. advisor at Southeast University. His main research interests include 3D precision measurement, 3D accurate recognition, and 3D optimal control theory and technology.