中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-11

论文引用格式:Yang Rong, Liu Silei, Wang Han, Zhou Zhuang, Li Shengyang. XXXX. Multi-object tracking for drosophila in space science experiments under ground-matching scenarios with DN-MeMOTR. Journal of Image and Graphics, XX(XX):0001-0011(杨荣, 刘偲磊, 王涵, 周壮, 李盛阳. XXXX. 面向地面匹配场景的空间科学果蝇实验多目标跟踪方法DN-MeMOTR. 中国图象图形学报, XX(XX):0001-0011)[DOI:10.11834/jig.250589]

面向地面匹配场景的空间科学果蝇实验多目标跟踪方法DN-MeMOTR

杨荣[1,2],刘偲磊[1,2],王涵[1,2],周壮[1],李盛阳[1,2*]

1. 中国科学院空间应用工程与技术中心,北京　100094; 2. 中国科学院大学,北京　100049

摘　要:**目的**　在空间生命科学实验研究中,果蝇作为典型模式生物,常被用于探索微重力环境对生命体行为、神经和代谢的影响。果蝇体型微小、运动非线性、个体外观高度相似,密集场景下现有多目标跟踪方法易出现检测精度低、关联错误等问题,难以满足空间科学数据分析需求。精确高效地实现多个果蝇的追踪,对理解其运动模式及行为机制具有重要作用与价值。**方法**　基于SpaceAnimal数据集,本文进一步扩充并构建了高密度、高动态运动特征的果蝇跟踪数据集,包含20条视频序列,2500帧标注图像。提出了DN-MeMOTR(DeNoising training incorporated in MeMOTR)跟踪框架,通过翻转标签、偏移边界框生成带噪样本,引入去噪训练补充检测监督信号;设计了ReID特征与GIoU加权融合的联合过滤模块,结合匈牙利算法优化目标关联,剔除冗余检测结果。**结果**　验证结果表明所提方法在HOTA、MOTA、DetA、IDF1等指标表现上优于其他先进对比方法,可视化和消融实验进一步验证方法的有效性。**结论**　构建了业界专属的果蝇实验跟踪数据集,提出了DN-MeMOTR方法,通过去噪训练与联合过滤两个模块,极大地提高了跟踪的精度,为后续微重力环境下果蝇行为分析提供了技术支持。

关键词:果蝇跟踪;空间科学实验;基于查询的跟踪;去噪训练;联合过滤

Multi-object tracking for drosophila in space science experiments under ground-matching scenarios with DN-MeMOTR

Yang Rong[1,2], Liu Silei[1,2], Wang Han[1,2], Zhou Zhuang[1], Li Shengyang[1,2*]

1. Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: **Objective** In space life science research, Drosophila melanogaster (fruit fly) serves as a representative model organism for exploring the effects of microgravity environments on biological behavior, neural activity, and metabolism. Achieving precise and efficient multi-object tracking of drosophila is crucial for understanding their motion patterns and behavioral mechanisms. However, due to their small body size, nonlinear motion, and high visual similarity among individuals, especially in dense scenes, existing multi-object tracking (MOT) methods face significant challenges. To address this, we constructed a drosophila multi-object tracking dataset under ground-based experimental conditions, designed for

收稿日期:2025-11-21;**修回日期**:2026-01-21

*通信作者:李盛阳 shyli@csu.ac.cn

基金项目:中国载人航天工程空间应用系统支持/资助项目,KJZ科学数据中心(Y6140711WN);国家基础科学数据共享服务平台数据资源运行服务(Y7031511WY)

Supported by: Space Application System of China Manned Space Program, KJZ Scientific Data Center (Y6140711WN); Data Resource Operations for the National Basic Science Data-Sharing Platform(Y7031511WY).

© 中国图象图形学报版权所有

space station applications. The dataset contains high-density video sequences with complex and dynamic behaviors, providing valuable data support for intelligent behavior analysis of drosophila. In terms of methodology, existing MOTR-based approaches achieve end-to-end multi-object tracking but suffer from inherent limitations that lead to suboptimal detection performance. To overcome this issue, this paper proposes a novel end-to-end multi-object tracking framework that effectively enhances both detection accuracy and tracking robustness. **Method** A ground-based simulated space environment dataset for multi-object tracking of drosophila was constructed based on SpaceAnimal, comprising 20 video sequences (a total of 2,500 annotated frames) that capture high-density and high-dynamic motion characteristics. And we propose DN-MeMOTR, attempting to enhance the performance of detection by reconstructing ground truths with noises. Noised samples are generated from ground truth data by flipping label and shifting the bounding box, we use label noise scale and box noise scale to control the scale of noise separately, both noised samples and initial learnable detect queries serve as detect queries and are concatenated with track queries, sequentially updated via self-attention and cross-attention. It should be emphasized that noised samples are employed solely during training and play no role in the inference stage. During inference, joint-filtering module is implemented to identify and remove targets that exhibit identical or similar characteristics across detect queries and track queries. Specifically, ReID head outputs ReID features of both detect queries and track queries, first we filter out detect outputs whose scores are lower than detection score threshold to obtain high score detection outputs indexes, then calculate the ReID similarity matrix and GIoU distance between detection outputs and track outputs, two matrixes are linearly combined through a weighted sum, α and β denote the corresponding weights, and find the matched pairs via Hungarian Algorithm. Finally new tracks are initialized from unmatched detect queries. We can reduce the duplicate tracks and improve tracking performance in this way. Due to the incorporation of noisy samples, in addition to the existing losses, we introduce extra loss specifically for noised samples and ReID loss for training ReID head. Denoising loss comprises simple focal loss, L1 loss and GIoU loss. The total loss is the weighted sum of focal loss, L1 loss, GIoU loss, denoising loss and ReID loss. **Result** we conduct experiments and analysis on drosophila dataset, and compare with other prevalent MOT methods like ByteTrack, OC-SORT, MeMOTR and etc. MeMOTR demonstrates suboptimal detection performance compared to other detection-based tracking frameworks, which can be attributed to its inherent limitations. Our approach achieves the highest HOTA score of 71.8%, surpassing MeMOTR and Hybrid-SORT by 4.9% and 9.0% respectively. With respect to the MOTA metric, our DN-MeMOTR obtains 87.6%, representing a substantial improvement of 19.3% over MeMOTR and 2.9% over Hybrid-SORT. We also obtain the highest DetA score of 73.4% and the highest IDF1 score of 84.9%. To enable a more direct comparison among DN-MeMOTR and original MeMOTR, we conduct inference on video clips. According to the visualization results, MeMOTR tends to miss some targets, whereas DN-MeMOTR demonstrates better performance. The ablation study proves the effectiveness of denoising training and joint-filtering module. **Conclusion** we construct a multi-object tracking dataset of drosophila to simulate and support behavior analysis tasks under future space microgravity scenarios by extending drosophila dataset of SpaceAnimal, and propose DN-MeMOTR, which incorporates denoising training and joint-filtering module, denoising training aims to improve network's detection performance, overcoming the drawback of MOTR-like models. While ReID features and joint-filtering module attempt to alleviate mismatching, boosts performance of original MeMOTR. Experiments show that our approach achieves superior performance on drosophila dataset, ablation study demonstrate the effectiveness of our components. It lays a solid foundation for the in-depth analysis of drosophila motion patterns and behavioral characteristics.

Key words: drosophila tracking; space scientific experiments; tracking-by-query; denoising training; joint-filtering module

0 引 言

2024 年 11 月,天舟八号飞船为中国空间站运送了多项科学实验项目,其中包括果蝇空间生物学实验。中国科学家首次在空间站环境下,系统研究亚磁-微重力条件对果蝇基因表达、行为模式及生存繁衍能力的影响。为有效区分空间环境因素与其他实验条件的干扰,在开展在轨果蝇实验的同时,同步设计并实施了地面匹配果蝇实验。该类实验在地面

© 中国图象图形学报版权所有

环境中尽可能复现实验装置、饲养条件和观测流程，仅改变重力与磁环境因素，从而为在轨实验结果提供可对照的基线数据。需要指出的是，由于在轨实验受到载荷体积限制、成像设备性能、空间光照条件及数据传输等因素的综合影响，其获取的图像在清晰度、分辨率和稳定性方面通常不及地面匹配实验影像，这也为后续的数据分析与算法设计带来了额外挑战。

通过对在轨实验与地面匹配实验的对比分析，可以系统揭示特殊空间环境对生物个体行为、生理状态及其遗传机制的影响，为探索生命过程中的关键科学规律，以及支撑人类长期在轨驻留和深空探索提供重要理论依据。在这一研究过程中，对果蝇个体行为的精细化量化分析是核心环节之一，而稳定、准确的个体身份识别与轨迹获取则是行为分析的前提条件。

随着深度学习技术的快速发展，将先进的人工智能算法引入空间生物学实验，利用自动化手段提升数据处理效率与分析精度，已成为重要发展方向。其中，多目标跟踪（Multi-Object Tracking，MOT）技术在果蝇实验数据分析中发挥着关键作用。多目标跟踪旨在在视频序列中对多个动态目标进行持续定位与身份保持，已广泛应用于视频监控、智能交通等领域。在果蝇空间实验中，受限于成像质量和环境复杂性，目标外观特征更加模糊且噪声更为显著；而地面匹配实验在成像条件上更为稳定，为算法研究和模型训练提供了更理想的数据基础。因此，研究适用于果蝇实验场景的多目标跟踪方法，对于实现果蝇行为的自动化、客观化分析具有重要意义。

基于上述背景，本文以地面匹配果蝇实验数据为研究对象，开展果蝇多目标跟踪方法研究。一方面，地面匹配实验具备可控性强、数据获取稳定、标注条件相对完善等优势，适合作为果蝇目标跟踪算法研究与验证的基础平台；另一方面，相关方法和经验可为后续在轨果蝇实验中复杂成像条件下的目标跟踪提供技术储备与方法参考。通过构建高质量的数据集并探索具有针对性的跟踪算法，本文旨在为果蝇空间实验中精细化、自动化的信息提取提供关键技术支撑。

主流多目标跟踪数据集如MOT17（Milan等，2016），MOT20（Dendorfer等，2020）以行人跟踪为核心，主要应对遮挡、视角变换等挑战；DanceTrack（Sun等，2022）则通过引入外观高度相似、运动行为多样的目标，进一步提升跟踪难度。但现有以人为中心的数据集，所涵盖的行为模式多为常规连续动作，难以覆盖自然场景中常见的突发性、高速、多变运动特征。相比之下，果蝇（Drosophila melanogaster）作为神经科学和遗传学研究的经典模式生物，具有体型微小、运动速度快、个体视觉特征高度相似等特点，是研究高动态复杂场景下多目标跟踪问题的理想对象。然而，目前缺乏针对这类密集、小型高动态目标的专用数据集，也制约了相关多目标跟踪算法的探索。

本文基于SpaceAnimal（Li等，2025）扩展并构建了地面匹配实验场景下的果蝇多目标跟踪数据集。该数据集具备两大特征与挑战：（1）目标体型微小且外观高度相似，易出现遮挡与身份混淆；（2）目标运动轨迹独特、复杂且非典型，难以建模预测。

另外，当前多目标跟踪（MOT）的常用范式仍是“基于检测的跟踪”（tracking-by-detection），通常分为目标检测与目标关联两个阶段。得益于目标检测技术的快速发展，多数此类方法采用YOLOX（Ge等，2021）等高性能检测器实现逐帧目标定位，随后融合外观特征、运动信息构建代价矩阵，通过匈牙利算法（Kuhn，1955）完成目标与轨迹的最优匹配。但在复杂密集或高速运动场景中，传统方法难以应对非线性运动与高度相似视觉外观的挑战，易出现跟踪漂移与身份切换问题。

Transformer架构（Vaswani等，2017）的提出，为深度学习领域带来里程碑式突破，也重塑了多目标跟踪的研究方向。在目标检测任务中，主流范式已从卷积神经网络（CNN）逐步转向Transformer结构，DETR（Carion等，2020）、Deformable-DETR（Zhu等，2020）、DN-DETR（Li等，2022）等代表性工作不断推动检测性能提升。在此基础上，MOTR（Zeng等，2022）等“基于查询的跟踪”方法以DETR变体为框架，通过迭代回归检测查询（detect queries）与跟踪查询（track queries），实现检测与关联的端到端联合建模，展现出了较好的鲁棒数据关联的潜力。

然而，MOTR及其变体的局限性日益凸显，核心问题在于检测性能显著退化。这一问题的根源是标签分配机制设计欠佳：随着已跟踪目标数量增加，真实标注（ground truth）中需与检测查询匹配的未跟踪目标数量减少，导致训练过程中检测查询的监督

© 中国图象图形学报版权所有

信号逐渐减弱。

本研究提出 DN-MeMOTR（DeNoising training incorporated in MeMOTR）方法，在 MeMOTR 框架中引入去噪训练机制，缓解检测监督信号稀疏问题，以显著提升检测性能。具体而言，本文在检测查询中主动注入带噪声的边界框与标签作为训练样本，将其与跟踪查询拼接后输入网络；同时设计融合外观信息的联合过滤模块，通过广义交并比（GIoU）（Rezatofighi 等，2019）与重识别（ReID）（He 等，2023）特征构建代价矩阵，剔除检测查询中与跟踪查询高度相似的冗余目标。与传统方法依赖独立外部 ReID 模块不同，本研究的 ReID 特征由网络内部独立 ReID 分支提取，参考 FairMOT（Zhang 等，2021）设计思路实现端到端优化。

本文在构建的果蝇多目标跟踪数据集上对 DN-MeMOTR 进行了实验评估。实验结果表明，本方法在跟踪精度上显著优于现有主流方法，消融实验进一步验证了去噪训练与联合过滤模块的有效性。综上，本文的主要贡献如下：

1）构建了空间科学地面匹配场景的果蝇跟踪数据集，为果蝇行为分析提供了数据支撑；

2）提出 DN-MeMOTR 方法，通过重构带噪样本，利用去噪训练增强检测的监督信息，缓解 MOTR 类模型因标签分配局限导致的检测性能退化；

3）设计融合 GIoU 与 ReID 特征的联合过滤模块，有效过滤冗余目标，提升模型轨迹关联能力；

4）在果蝇数据集上完成充分的实验与消融分析，验证了方法的性能优势与核心模块有效性，为后续空间科学实验中果蝇行为分析奠定技术基础。

1 相关工作

1.1 基于检测的跟踪

基于检测的跟踪（tracking-by-detection）是多目标跟踪领域应用最广泛的范式，其核心思路是将跟踪任务拆解为“目标检测”与“目标关联”两个独立子任务，通过分步优化实现对目标的连续追踪。随着目标检测技术的快速迭代，近年来涌现出 YOLOX（Ge 等，2021）、YOLOv8（Jocher 等，2023）、YOLOv11（Jocher 等，2024）等一系列高性能检测器，为该类跟踪方法提供了精准的逐帧目标定位基础。典型的“基于检测的跟踪”框架采用逐帧处理逻辑：先利用检测器获取当前帧的目标位置与类别信息，再以该检测结果为基础，通过融合外观特征、运动模型或交并比（IoU）等信息构建代价矩阵，最终借助匈牙利算法完成当前检测目标与历史轨迹的最优匹配，实现身份延续。SORT（Bewley 等，2016）采用基于 CNN 的检测器，并结合卡尔曼滤波器与匈牙利算法进行关联；MOTDT（Chen 等，2018）通过在检测和轨迹输出中选择候选项来应对不可靠的检测；Deep-SORT（Sun 等，2022）在 SORT 的基础上引入外观信息；ByteTrack（Zhang 等，2022）利用低置信度检测结果，并在第一次关联后将其与未匹配的轨迹进行匹配；BoT-SORT（Aharon 等，2022）将外观模型、运动模型和相机运动补偿整合到 ByteTrack 框架中；OC-SORT（Cao 等，2023）在遮挡情况下，通过强调观测而非线性状态估计，避免了卡尔曼滤波误差的累积；Hybrid-SORT（Yang 等，2024）利用置信度状态、高度状态和速度方向等弱线索对强线索进行补偿；FairMOT（Zhang 等，2021）通过双分支结构实现检测与重识别的联合训练；DeepOCSort（Maggiolino 等，2023）提出了动态外观与自适应加权机制，使其在特征退化情况下依然具有鲁棒性。

尽管这些方法依托高性能检测器实现了较高的检测精度，但在目标运动非线性强、个体外观高度相似的复杂场景（如密集果蝇群体跟踪）中，手工设计的关联策略往往难以灵活适配目标动态变化，易出现轨迹断裂或身份混淆问题。

1.2 基于查询的跟踪

“基于检测的跟踪”因采用两阶段分步优化模式，难以实现检测与关联的全局最优，因此“基于查询的跟踪”（tracking-by-query）范式应运而生，其核心是通过端到端训练，将检测与关联任务统一建模，实现网络整体优化。该类方法的设计源于基于 Transformer 的 DETR 系列目标检测器。近年来，DETR 类目标检测器取得了显著进展。DETR 首次实现无需手工联合过滤的端到端目标检测，通过 Transformer 编码器 - 解码器结构与可学习查询（query）完成目标定位与分类；DN-DETR 则通过在训练中注入带噪声的查询样本，增强了模型泛化能力并加速收敛。与传统 DETR 类检测器仅依赖“检测查询”（detect queries）不同，“基于查询的跟踪”方法额外引入独立的“跟踪查询”（track queries），通过跟踪查询隐式存

© 中国图象图形学报版权所有

储目标历史信息，再将两类查询拼接后输入 Transformer 网络，实现检测与关联的联合建模。

在具体方法方面，TransTrack（Sun 等，2020）为检测查询与跟踪查询设计并行解码器，分别完成目标检测与轨迹更新，最后通过边界框 IoU 匹配筛选最终结果；TrackFormer（Meinhardt 等，2022）借助自注意力机制动态更新跟踪查询，无需额外的匹配操作；MOTR 将多目标跟踪（MOT）表述为序列预测集合问题，迭代更新跟踪查询；MOTRv2（Zhang 等，2023）引入 YOLOX 检测器生成高质量候选框，作为跟踪查询的初始化依据，提升了目标定位精度；MeMOTR（Gao 等，2023）为跟踪查询引入长期记忆机制，通过存储目标多帧历史特征，增强了长序列跟踪的稳定性；CO-MOT（Yan 等，2025）提出了一个新的标签分配机制，提升了目标的建模能力。

尽管 MeMOTR 在目标关联任务中展现出潜力，但其性能仍受限于不合理的标签分配策略：随着训练过程中已跟踪目标数量增加，真实标注（ground truth）中可与检测查询匹配的未跟踪目标数量逐渐减少，导致检测查询的监督信号持续弱化，最终造成检测性能显著下降。这一问题也成为制约多数“基于查询的跟踪”方法应用于小目标、高密集场景（如果蝇跟踪）的核心瓶颈。

2 方 法

2.1 概述

本节详细介绍所提出的 DN-MeMOTR 方法，其整体架构如图 1 所示。该方法的核心设计逻辑为：在训练阶段，通过真实标注数据生成带噪声的训练样本，引入额外的监督信息，将这些噪声样本与初始可学习的检测查询（detect queries）共同作为输入检测查询，与跟踪查询（track queries）拼接后送入 Transformer 网络，再通过自注意力与交叉注意力机制完成特征迭代更新；为适配噪声样本训练，除 MeMOTR 原有的损失函数外，额外引入针对噪声样本的去噪损失，并为 ReID 分支设计 ReID 损失以优化外观特征提取。在推理阶段，通过联合过滤模块识别并剔除检测查询与跟踪查询中特征高度相似的冗余目标，进一步提升跟踪准确性。

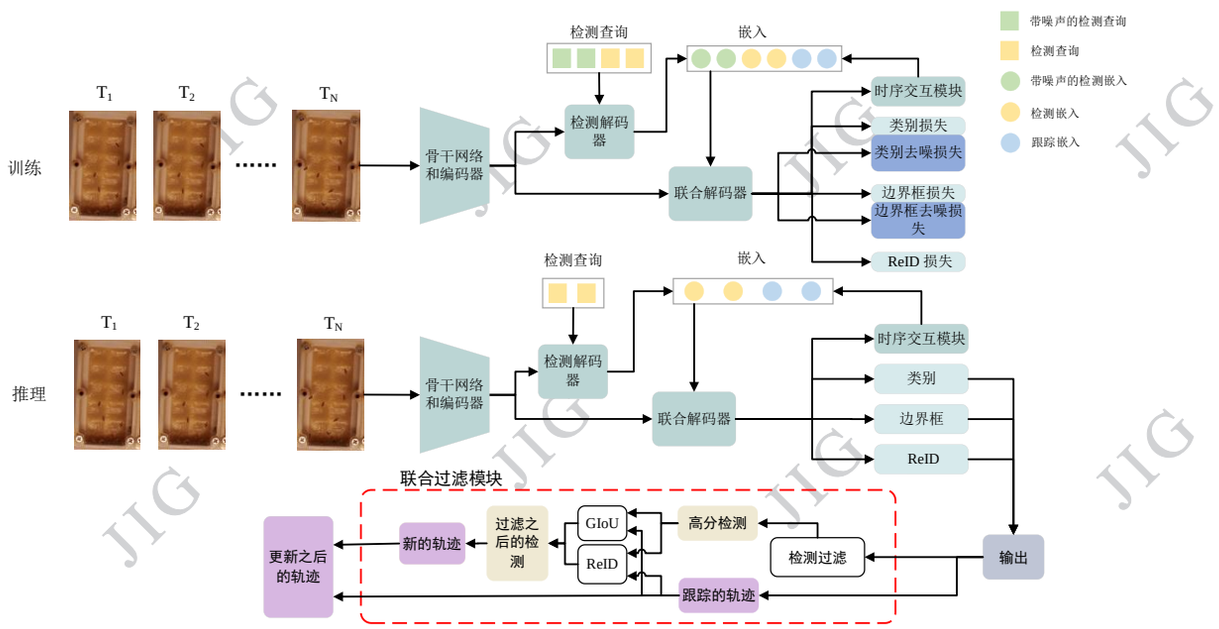


图 1 DN-MeMOTR 网络结构图

Fig. 1 Framework of DN-MeMOTR

2.2 去噪训练

为解决 MOTR 类跟踪方法普遍存在的检测性能退化问题，本研究借鉴 DN-DETR 的噪声注入思路，将去噪训练机制融入 MeMOTR 框架，通过补充检测查询的监督信号增强模型检测能力。具体来说，将真值目标记为 $\boldsymbol{T}_i \in \boldsymbol{T}_1, \boldsymbol{T}_2, \cdots, \boldsymbol{T}_n$，n 是批次的大小，每个目标可能包含多个目标类别和边界框，表示为 $\boldsymbol{T}^i_{\text{label}}, \boldsymbol{T}^i_{\text{box}}$。所有的类别标签和边界框分别拼接到

© 中国图象图形学报版权所有

一起，得到 $\boldsymbol{T}_{\text{label}}$，$\boldsymbol{T}_{\text{box}}$。*scalar* 是噪声查询的组数，$\boldsymbol{T}_{\text{label}}$，$\boldsymbol{T}_{\text{box}}$ 还有对应的批次索引 $\boldsymbol{T}_{\text{index}}$ 将被重复 *scalar* 次。

对于标签噪声，用 *label noise scale* 来控制噪声的大小，具体地，产生一个随机张量 $\boldsymbol{P}$，它和 $\boldsymbol{T}_{\text{label}}$ 的长度相同，它的值在[0, 1)区间内。这个 $\boldsymbol{P}$ 的值中低于 *label noise scale* 的对应的索引表示位置，$\boldsymbol{T}_{\text{label}}$ 在这选定的位置的值将会被随机的改变。对 $\boldsymbol{T}_{\text{label}}$ 施加一个嵌入函数即可得到标签嵌入 $\boldsymbol{E}_{\text{label}}$。

对于边界框噪声，根据原始的 $\boldsymbol{T}_{\text{box}}$，计算一个张量 ***Diff***，每一行表示 $\frac{w}{2}$，$\frac{h}{2}$，w，h，然后计算另一个随机的张量 $\boldsymbol{V}_{\text{rand}}$，其长度和 $\boldsymbol{T}_{\text{box}}$ 一样，它的值是在[−1, 1)区间，噪声边界框张量 $\boldsymbol{T}_{\text{noise}}$ 定义为 $\boldsymbol{V}_{\text{rand}}$、***Diff*** 和 *box noise scale* 的乘积。最后的噪声的边界框张量 $\boldsymbol{T}_{\text{noised\_box}}$ 是 $\boldsymbol{T}_{\text{box}}$ 和 $\boldsymbol{T}_{\text{noise}}$ 的和。所有的值是在[0, 1]区间中，通过对 $\boldsymbol{T}_{\text{noised\_box}}$ 施加一个逆 Sigmoid 运算得到输入的边界框嵌入 $\boldsymbol{E}_{\text{bbox}}$。

为了得到最后的检测查询，它包含噪声查询和可学习的查询。首先，计算不同批次中目标的最大数量 *single pad*，*padding label* 和 *padding bbox* 用零进行初始化，然后第一个维度的大小是 *single pad* × *scalar*。然后 *padding label* 和 *query embed*，*padding box* 和 *reference point embed* 分别进行拼接。在重复 *batch size* 次之后，得到了输入标签查询 $\boldsymbol{Q}_{\text{label}}$ 和输入边界框查询 $\boldsymbol{Q}_{\text{bbox}}$。最后一步是，用 $\boldsymbol{E}_{\text{label}}$ 和 $\boldsymbol{E}_{\text{bbox}}$ 分别更新 $\boldsymbol{Q}_{\text{label}}$ 和 $\boldsymbol{Q}_{\text{bbox}}$；另外，$\boldsymbol{M}_{\text{attn}}$ 是一个矩阵，防止不同组的查询之间互相干涉。

网络的输入查询嵌入 $\boldsymbol{E}_{\text{query}}$ 是 $\boldsymbol{Q}_{\text{label}}$ 和 $\boldsymbol{Q}_{\text{bbox}}$ 的拼接，原始的查询掩码 $\boldsymbol{M}_{\text{query}}$ 包含检测查询掩码 $\boldsymbol{M}_{\text{det\_query}}$ 和跟踪查询掩码 $\boldsymbol{M}_{\text{trk\_query}}$，需要更新来适应新增的检测查询，确保与扩展了的检测查询对齐，同时保持 $\boldsymbol{M}_{\text{trk\_query}}$ 不变，需要强调的是：噪声样本仅在训练期间使用，在推理阶段没有使用。

2.3 联合过滤模块

传统 MOTR 类方法依赖跟踪查询的迭代回归完成目标关联，未设计显式的冗余目标过滤机制，易出现同一目标被重复检测的问题。为此，本研究设计融合 ReID 特征与 GIoU（广义交并比）的联合过滤模块，在推理阶段剔除检测查询与跟踪查询中的相似冗余目标，具体流程如下：

首先，利用模型的 ReID 分支提取检测查询与跟踪查询的 ReID 特征，设置检测分数阈值 τ_{high}，过滤掉检测置信度低于 τ_{high} 的检测输出，仅保留高分检测输出及其索引 $\boldsymbol{I}_{\text{high}}$，减少低质量检测结果对关联的干扰。随后计算两类相似度矩阵：一是高分检测输出与跟踪输出之间的 ReID 特征相似度矩阵 $\boldsymbol{M}_{\text{reid}}$；二是两者之间的 GIoU 距离矩阵 $\boldsymbol{M}_{\text{giou}}$。为平衡外观特征与空间位置的权重，将 $\boldsymbol{M}_{\text{reid}}$ 与 $\boldsymbol{M}_{\text{giou}}$ 加权求和（其中 α、β 分别为两者的权重系数），得到最终代价矩阵；采用匈牙利算法对代价矩阵进行最优匹配，确定高分检测输出与跟踪输出的匹配对。最后，对未匹配成功的高分检测输出初始化新的跟踪轨迹，通过这一过程有效减少重复轨迹，提升跟踪的准确性与一致性。

2.4 损失函数

由于引入噪声样本训练与 ReID 特征，需在 MeMOTR 原有损失函数基础上补充新的去噪损失项和 ReID 损失，构建更全面的损失函数以优化模型性能。去噪损失包含 focal 损失 $L_{\text{dn\_focal}}$（Lin 等，2017），L1 损失 $L_{\text{dn\_L1}}$ 和边界框的 GIoU 损失 $L_{\text{dn\_giou}}$。将噪声的输出分隔开，包括噪声的边界框输出 $\boldsymbol{O}_{\text{noise\_bbox}}$ 和噪声标签输出 $\boldsymbol{O}_{\text{noise\_label}}$，用于计算 L_{dn}。总的去噪损失 L_{dn} 的定义如下：

$$L_{\text{dn}} = \lambda_{\text{focal}} \cdot L_{\text{dn\_focal}} + \lambda_{\text{L1}} \cdot L_{\text{dn\_L1}} + \lambda_{\text{giou}} \cdot L_{\text{dn\_giou}} \tag{1}$$

为优化 ReID 分支提取的外观特征区分度，引入 L_{reid} 作为 ReID 损失，监督模型学习同一目标不同帧间的特征一致性与不同目标间的特征差异性，确保外观特征能有效支撑目标关联。因此，总的损失 L_{total} 包含 focal 损失 L_{focal}，L1 损失 L_{L1}，GIoU 损失 L_{giou}，去噪损失 L_{dn} 和 ReID 损失 L_{reid}，公式如下：

$$L_{\text{total}} = \lambda_{\text{focal}} \cdot L_{\text{focal}} + \lambda_{\text{L1}} \cdot L_{\text{L1}} + \lambda_{\text{giou}} \cdot L_{\text{giou}} + \lambda_{\text{dn}} \cdot L_{\text{dn}} + \lambda_{\text{reid}} \cdot L_{\text{reid}} \tag{2}$$

$\text{i} \in [\text{focal}, \text{L1}, \text{giou}, \text{dn}, \text{reid}]$，$\lambda_{\text{i}}$ 是对应项的权重。

3 数据集构建和评价指标

本研究在 SpaceAnimal 数据集的基础上，扩展并构建了空间科学地面匹配实验的果蝇跟踪数据集：共包含 20 条视频序列，每条序列均由连续 125 帧标注图像组成，图像分辨率覆盖 1200×1200 至 3840×3840 像素范围，数据集总标注量达 2500 帧图

© 中国图象图形学报版权所有

像；为满足模型训练与测试的合理性，将数据集按7∶3比例划分，训练集包含14条视频序列，涵盖158个果蝇目标，对应19729个精准标注的边界框；测试集包含6条视频序列，涵盖66个果蝇目标，对应8195个边界框标注。在数据标注过程中，为每一只果蝇目标分配唯一且固定的身份标识（ID），并通过逐帧的目标匹配与人工校对机制，确保同一果蝇在整个视频序列中始终对应一致的身份标签。最大程度地保证了身份关联（ReID）的准确性，为后续ReID网络的稳定训练与性能提升奠定了可靠的数据基础。

与现有动物跟踪数据集相比，本文数据集对果蝇跟踪算法检测精度与关联鲁棒性提出了更高的挑战。单帧图像中平均包含10个果蝇目标，个体尺寸微小，且不同个体的外观特征高度相似（体色、体型差异极小），易因遮挡导致身份混淆；果蝇运动模式呈现显著的非线性特征（包含快速转向、短距离跳跃等突发动作），轨迹难以预测。表1将其与两个具有代表性的动物多目标跟踪数据集GMOT-40-Animal（Bai等，2021）和TAO-Animal（Dave等，2020）进行了对比，本文数据集在总体规模上与上述数据集处于同一量级，具备支持深度学习模型训练与性能评估的基本条件。

数据集遵循DanceTrack的格式，数据集样例及其标注示例如图2所示。数据集采用通用的MOT的评价指标，常用的指标包括HOTA（Luiten等，2021）、MOTA（Bernardin等，2008）、IDF1（Ristani等，2016）。

表1　果蝇数据集与典型动物多目标跟踪数据集的对比

Table 1　Comparisions between Drosophila Dataset and other typical multi-object tracking datasets for animals

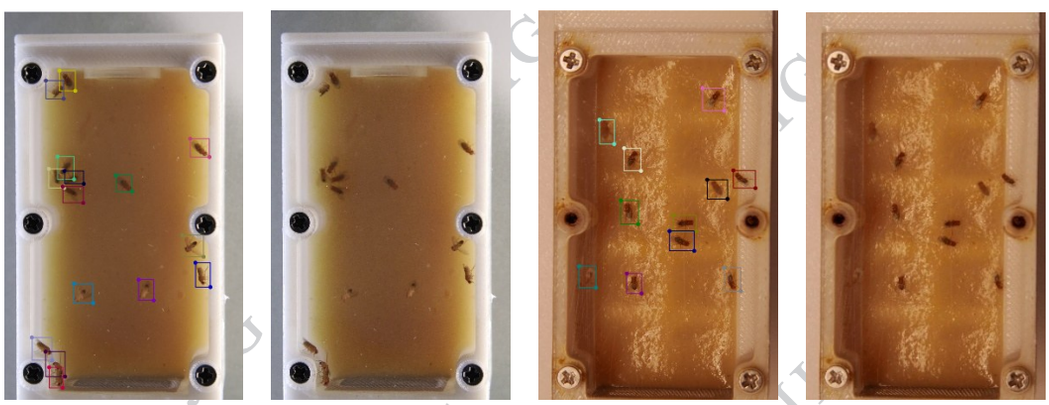
数据集	GMOT-40-Animal	TAO-Animal	果蝇数据集
视频序列数	12	39	20
总帧数	2.6k	2.5k	2.5k
总边界框数	63k	3.4k	27.9k
总轨迹数	837	250	224

4　实验结果与分析

4.1　实验细节

关于对比方法，本文训练了一个YOLOX检测器，共80个epochs，对于所有的实验除了MeMOTR，保持他们的参数和原始论文中的方法一致。

当训练DN-MeMOTR和MeMOTR的时候，按照MeMOTR方法，用预训练的DAB-Deformable-DETR（Liu等，2022）权重来初始化模型，以保证公平比较。学习率和epochs分别设为0.0001和40，学习率会在第24个epoch的时候会缩小10倍。批次大小设为1，每个批次包含一个视频片段，帧数从1到10不等。初始的帧数设为2，这个数量会在第12、20、28个epoch的时候分别变为3，4，5。其它的超参数的设定和原始的MeMOTR一样，λ_{focal}，λ_{giou}，λ_{L1}，λ_{dn}，λ_{reid}分别设为2，2，5，1，1。实验是基于pytorch和英伟达的RTX 3090显卡。



（a）　（b）　（c）　（d）

图2　数据集图片及其标注样例，（a）和（c）分别是（b）和（d）的标注

Fig. 2　Sample images and their annotations from the dataset.（a）and（c）are annotations of（b）and（d）separately

© 中国图象图形学报版权所有

4.2 对比实验

本文将提出的DN-MeMOTR方法和多个先进的多目标跟踪方法进行了对比，如表2所示，DN-MeMOTR实现了最高的HOTA指标，达到了71.8%，比MeMOTR高4.9%，比Hybrid-SORT高9.0%。如表格所示，MeMOTR相比于其它的基于检测的跟踪方法，它的检测性能较弱。这主要是由于它内在的局限性。在MOTA指标上，DN-MeMOTR得到了87.6%的精度，比MeMOTR高19.3%，比Hybrid-SORT高2.9%。另外，本文的方法也实现了最高的DetA分数，高达73.4%，IDF1分数是84.9%。这都显著地超过了现有的方法。至于AssA指标，也能够得到最高的分数73.9%。这个结果是通过适当调整权重得到的。为了得到更直接的DN-MeMOTR和原始的MeMOTR之间的对比，对测试集中的视频片段进行了推理，可视化结果如图3所示，MeMOTR更容易丢失一些目标，然而DN-MeMOTR展示出了更好的性能，实验表明了本文方法的有效性。

表2 和其它方法的对比结果，DN-MeMOTR*表示$\alpha=1,\beta=0$的结果

Table 2 Comparisons with other methods, DN-MeMOTR* denotes the result when $\alpha=1,\beta=0$

Methods	HOTA	DetA	AssA	MOTA	IDF1
MOTDT	42.8	56.0	33.5	62.5	46.3
SORT	54.6	58.6	51.5	71.4	64.4
DeepSORT	52.6	56.0	50.1	67.0	64.0
OC-SORT	62.3	68.1	57.4	83.1	73.1
DeepOCSORT	61.0	64.3	58.3	76.6	71.5
Hybrid-SORT	62.8	69.2	57.4	84.7	73.8
BoT-SORT	59.9	61.5	58.8	74.8	72.6
ByteTrack	58.8	58.9	59.4	73.6	75.9
MOTRv2	63.4	57.7	70.1	61.9	75.4
MeMOTR	66.9	61.4	73.3	68.3	79.3
CO-MOT	66.4	61.3	72.3	70.5	79.6
DN-MeMOTR*	70.2	67.0	**73.9**	79.8	84.4
DN-MeMOTR	**71.8**	**73.4**	70.7	**87.6**	**84.9**

4.3 消融实验

在本节，对DN-MeMOTR中各个组件的有效性进行了分析。包括去噪训练和联合过滤模块。在表3中，当引入去噪训练的时候，所有的指标除了AssA都有提高，DetA实现了10.1%的提升，MOTA实现了16.5%的提升。这验证了去噪训练能够提高MOTR类模型的检测性能。

表3 去噪训练和联合过滤模块的消融实验

Table 3 Ablation studies on denoising training and joint-filtering module

Denosing	Joint-Filtering Module	HOTA	DetA	AssA	MOTA	IDF1
■	■	66.9	61.4	73.3	68.3	79.3
■	■	69.5	71.5	67.9	84.8	81.5
■	■	**71.4**	**72.0**	71.3	**85.0**	**84.6**

表4 成本矩阵的不同权重的消融实验

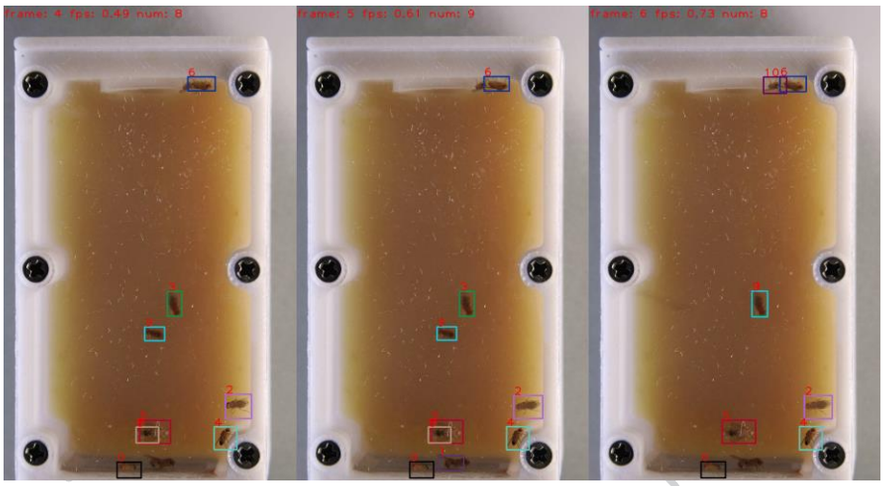
Table 4 Ablation study on different weights of cost matrix

α	β	Denoising	HOTA	DetA	AssA	MOTA	IDF1
0	1	■	71.5	**74.1**	69.4	**88.2**	83.7
0.3	0.7	■	**71.9**	73.9	70.4	88.2	84.6
0.5	0.5	■	71.8	73.4	70.7	87.6	**84.9**
0.7	0.3	■	70.5	69.0	72.4	82.4	84.8
1	0	■	70.2	67.0	**73.9**	79.8	84.4

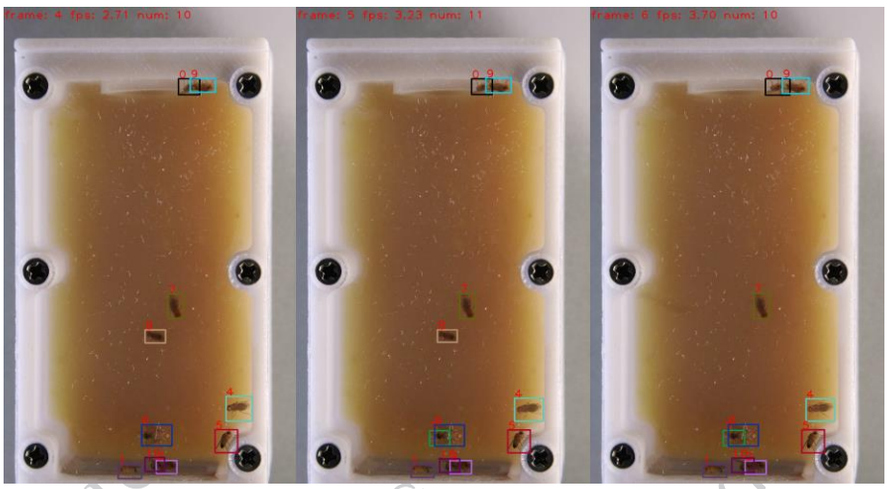
在表4中，评估了ReID特征和GIoU距离结合的有效性，表格中清晰地表明了用相似度矩阵过滤有干扰的查询对于整体的性能是有帮助的。AssA比原始的方法低，是由于ReID和GIoU之间的权衡。

另外，尝试了不同的ReID特征和GIoU距离之间的结合方式，在表4中，α和β表示不同的权重，当ReID的权重增加时，AssA指标有提升，然而其它的指标例如HOTA、DetA、MOTA是下降。相反，当GIoU的权重增加时，这个趋势是相反的。AssA指标下降，其它的指标HOTA、DetA和MOTA是增加的。实验表明，以不同的权重结合，会导致不同的结果，将$\alpha=0.5,\beta=0.5$作为默认的设置。

© 中国图象图形学报版权所有



(a)



(b)

图3 (a)和(b)分别表示MeMOTR和DN-MeMOTR推理的结果

Fig. 3 (a) and (b) show the inference results of MeMOTR and DN-MeMOTR

5 结 论

为支撑空间科学实验中微重力环境下果蝇运动模式与行为机制的精细化研究，本文围绕果蝇小目标、高动态、多遮挡等典型特征，构建了地面实验场景下的小型高动态果蝇多目标跟踪数据集，为相关算法研究提供了可靠的数据基础。在此基础上，提出了一种DN-MeMOTR多目标跟踪框架。该方法通过“翻转标签—偏移边界框”的策略生成带噪样本，在训练阶段为检测查询引入额外监督信号，有效缓解了由标签分配机制引起的监督稀疏问题；同时设计了融合ReID特征与GIoU约束的联合过滤模块，并结合匈牙利算法进行目标关联，在抑制冗余检测结果的同时显著提升了复杂场景下的关联稳定性，突破了传统方法在小尺度、高密度目标跟踪任务中的性能瓶颈。

实验结果表明，所提出的DN-MeMOTR在HOTA、MOTA、DetA与IDF1等多目标跟踪的核心评价指标上均优于主流方法；可视化结果显示，其在密集遮挡与快速运动场景下，性能优于对比方法。消融实验进一步验证了去噪训练和联合过滤模块两者协同作用可实现“检测-关联”性能的均衡优化。

本文工作仍存在进一步拓展空间。一方面，当前模型的推理效率尚难以满足实时分析需求，后续可通过网络结构轻量化与模型量化等手段提升推理速度；另一方面，现有数据集尚未覆盖空间微重力环境下的在轨果蝇实验数据，未来可结合真实在轨实

© 中国图象图形学报版权所有

验样本对模型进行针对性优化,并进一步开展天地环境下果蝇行为差异的对比分析研究。

参考文献(References)

Aharon N, Orfaig R. and Bobrovsky B.Z.2022. Bot-sort: Robust associations multi pedestrian tracking[EB/OL].[2025-09-27]. https://arxiv.org/pdf/2206.14651

Bai H X, Cheng W S, Chu P, Liu J H, Zhang K and Ling H B.2021. Gmot-40: A benchmark for generic multiple object tracking//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, IEEE: 6719-6728 [DOI: 10.1109/CVPR46437.2021.00665]

Bernardin K and Stiefelhagen R. 2008. Evaluating multiple object tracking performance: the clear mot metrics. EURASIP Journal on Image and Video Processing. 2008: 1 - 10 [DOI: 10.1155/2008/246309]

Bewley A, Ge Z, Ott L, Ramos F and Upcroft B. 2016. Simple online and realtime tracking//2016 IEEE International Conference on Image Processing. Phoenix, IEEE: 3464 - 3468 [DOI: 10.1109/ICIP.2016.7533003]

Cao J K, Pang J M, Weng X S, Khirodkar R and Kitani K. 2023. Observation-centric sort: Rethinking sort for robust multi-object tracking// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, IEEE: 9686 - 9696 [DOI: 10.1109/CVPR52729.2023.00934]

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End to-end object detection with transformers//European Conference on Computer Vision. Glasgow, Springer: 213 - 229 [DOI: 10.1007/978-3-030-58452-8_13].

Chen L, Ai, H Z, Zhuang Z J and Shang C.2018. Real-time multiple people tracking with deeply learned candidate selection and person re-identification//2018 IEEE International Conference on Multimedia and Expo. San Diego, IEEE: 1-6 [DOI: 10.1109/ICME.2018.8486597].

Dave A, Khurana T, Tokmakov P, Schmid C and Ramanan D. Tao. 2020. A large-scale benchmark for tracking any object//European conference on computer vision. Glasgow, Springer: 436-454 [DOI: 10.1007/978-3-030-58558-7_26]

Dendorfer P, Rezatofighi H, Milan A, Shi J, Cremers D, Reid I, Roth S, Schindler K and Leal-Taix′e L. 2020. Mot20: A benchmark for multi object tracking in crowded scenes [EB/OL]. [2025-09-27]. https://arxiv.org/pdf/2003.09003.

Gao R P and Wang L M. 2023. Memotr: long-term memory-augmented transformer for multi object tracking //Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, IEEE: 9901 - 9910 [DOI:10.1109/ICCV51070.2023.00908].

Ge Z, Liu S T, Wang F, Li Z M and Sun J. 2021. Yolox: Exceeding yolo series in 2021 [EB/OL]. [2025-09-27]. https://arxiv.org/pdf/2107.08430.

He L X, Liao X Y, Liu W, Liu X C, Cheng P and Mei T. 2023. Fastreid: a pytorch toolbox for general instance re-identification//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, ACM: 9664 - 9667 [DOI: 10.1145/3581783.3613460].

Jocher G, Chaurasia A and Qiu J.2023. Ultralytics YOLOv8[EB/OL]. [2025-09-27]. https://github.com/ultralytics/ultralytics.

Jocher G and Qiu J. Ultralystics yolo11[EB/OL]. [2025-09-27]. https://github.com/ultralytics/ultralytics.

Kalman R.E. 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME--Journal of Basic Engineering. 82(1):35-45[DOI: 10.1115/1.3662552]

Kuhn H. W. 1955. The hungarian method for the assignment problem. Naval Research Logistics Quarterly, 2(1 - 2): 83 - 97 [DOI: 10.1002/nav.3800020109].

Li F, Zhang H, Liu S L, Guo J, Ni L M and Zhang L.2022. Dn-detr: accelerate detr training by introducing query denoising//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, IEEE: 13619 - 13627 [DOI: 10.1109/CVPR52688.2022.01325].

Li S Y, Liu K, Wang H, Yang R, Li X Z, Sun Y Q, Zhong R T, Wang W, Li Y, Sun Y J and Wang G H. 2025. Pose estimation and tracking dataset for multi-animal behavior analysis on the china space station. Scientific Data, 12(1): 1 - 9 [DOI: 10.1038/s41597-025-01234-5].

Liu S L, Li F, Zhang H, Yang X, Qi X B, Su H, Zhu J and Zhang L. 2022. Dab-detr: dynamic anchor boxes are better queries for detr. [EB/OL].[2025-09-27]. https://arxiv.org/pdf/2201.12329.

Lin T Y, Goyal P, Girshick R, He K M and Dollar P. 2017. Focal loss for dense object detection//Proceedings of the IEEE International Conference on Computer Vision. Venice, IEEE:2999-3007[DOI: 10.1109/ICCV.2017.324].

Luiten J, Osep A, Dendorfer P, Torr P, Geiger A, Leal-Taix′e L and Leibe B. 2021. Hota: a higher order metric for evaluating multi-object tracking. International Journal of Computer Vision, 129: 548 - 578 [DOI: 10.1007/s11263-020-01411-w].

Maggiolino G, Ahmad A, Cao J K and Kitani K. 2023. Deep oc-sort: multi-pedestrian tracking by adaptive re-identification//2023 IEEE International Conference on Image Processing. Malaysia, IEEE: 3025 - 3029[DOI: 10.1109/ICIP49359.2023.10222576].

Meinhardt T, Kirillov A, Leal-Taixe L and Feichtenhofer C.2022. Trackformer: multi object tracking with transformers//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, IEEE: 8844 - 8854 [DOI: 10.1109/CVPR52688.2022.00884].

Milan A, Leal-Taix′e L, Reid I, Roth S and Schindler K. 2016. Mot16:

© 中国图象图形学报版权所有

A benchmark for multi-object tracking[EB/OL]. [2025-09-27]. https://arxiv.org/pdf/1603.00831.

Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I and Savarese S. 2019. Generalized intersection over union: a metric and a loss for bounding box regression // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, IEEE:658-666[DOI:10.1109/CVPR.2019.00075]

Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C. 2016. Performance measures and a data set for multi-target, multi-camera tracking//European Conference on Computer Vision, Amsterdam, Springer: 17 - 35[DOI: 10.1007/978-3-319-48881-3_2]

Sun P Z, Cao J K, Jiang Y, Yuan Z H, Bai S, Kitani K and Luo P. 2022. Dancetrack: multi-object tracking in uniform appearance and diverse motion//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, IEEE: 20993 - 21002 [DOI: 10.1109/CVPR52688.2022.02032]

Sun P Z, Cao J K, Jiang Y, Zhang R F, Xie E Z, Yuan Z H, Wang C H and Luo P. 2020. Transtrack: multiple object tracking with transformer [EB/OL].[2025-09-27]. https://arxiv.org/pdf/2012.15460.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A.N, Kaiser L and Polosukhin I. 2017. Attention is all you need. Advances in Neural Information Processing Systems. Long Beach, MIT Press.

Wojke N, Bewley A and Paulus D. 2017. Simple online and realtime tracking with a deep association metric//2017 IEEE International Conference on Image Processing. Beijing, IEEE: 3645 - 3649 [DOI: 10.1109/ICIP.2017.8296962].

Yan F, Luo W X, Zhong Y J, Gan Y Y and Ma L. 2025. CO-MOT: Boosting End-to-end Transformer-based Multi-Object Tracking via Coopetition Label Assignment and Shadow Sets//The Thirteenth International Conference on Learning Representations. Singapore.

Yang M Z, Han G X, Yan B, Zhang W H, Qi J Q, Lu H C and Wang D. 2024. Hybrid-sort: weak cues matter for online multi-object tracking// Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver. AAAI: 38(7): 6504 - 6512 [DOI: 10.1609/aaai.v38i7.28471].

Zeng F A, Dong B, Zhang Y A, Wang T C, Zhang X Y and Wei Y C. 2022. Motr: end-to-end multiple-object tracking with transformer// European Conference on Computer Vision. Tel Aviv. Springer: 659 - 675 [DOI:10.1007/978-3-031-19812-0__38].

Zhang Y F, Sun P Z, Jiang Y, Yu D D, Weng F C, Yuan Z H, Luo P, Liu W Y and Wang X G. 2022. Bytetrack: multi-object tracking by associating every detection box//European Conference on Computer Vision. Tel Aviv. Springer: 1 - 21 [DOI: 10.1007/978-3-031-200047-2_1].

Zhang Y F, Wang C Y, Wang X G, Zeng W J and Liu W Y.2021. Fairmot: on the fairness of detection and re-identification in multiple object tracking. International Journal of Computer Vision, 129: 3069 - 3087 [DOI:10.1007/s11263-021-01513-4].

Zhang Y, Wang T C and Zhang X Y. 2023. Motrv2: bootstrapping end-to-end multi-object tracking by pretrained object detectors//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, IEEE: 22056 - 22065 [DOI: 10.1109/CVPR52729.2023.02112].

Zhu X Z, Su W J, Lu L W, Li B, Wang X G and Dai J F. 2020. Deformable detr: deformable transformers for end-to-end object detection [EB/OL]. [2025-09-27]. https://arxiv.org/pdf/2010.04159

作者简介

杨荣，男，硕士研究生，主要研究方向为多目标跟踪和行为检测。E-mail：yangrong23@csu.ac.cn

刘偲碹，男，硕士研究生，主要研究方向为多模态大模型。E-mail：liusilei23@csu.ac.cn

王涵，男，博士研究生，主要研究方向为计算机视觉与多模态。E-mail：wanghan221@csu.ac.cn

周壮，男，助理工程师，主要研究方向为遥感图像分类与目标检测。E-mail：zhouzhuang@csu.ac.cn

李盛阳，通信作者，男，研究员，主要研究方向为图像智能处理和卫星视频目标检测和跟踪。E-mail：shyli@csu.ac.cn

© 中国图象图形学报版权所有