doi:10.6041/j.issn.1000-1298.2025.05.045

融合深度信息与运动趋势的羊只多目标跟踪方法

王美丽　杨恩德
(西北农林科技大学信息工程学院,陕西杨凌 712100)

摘要:近年来,随着羊只养殖向大规模和精细化的方向发展,羊场对智能化管理的需求日益增加。因此,精准的个体识别和行为监测变得尤为重要,对多目标跟踪(Multiple object tracking, MOT)算法的准确性提出了更高要求。然而,现有的 MOT 算法在目标遮挡和动态场景下的性能仍不理想。本文提出两种跟踪线索:深度调制交并比(Depth modulated intersection over union, DIoU)和轨迹方向建模(Tracklet direction modeling, TDM),旨在补充交并比(Intersection over union, IoU)线索,提高多目标跟踪的精准度和鲁棒性。DIoU 线索通过引入目标的深度信息改进了传统的 IoU 计算方法。TDM 聚焦于目标的运动趋势,预测其未来的移动方向。本文将 DIoU 和 TDM 跟踪线索集成到 BoT-SORT 算法中,形成改进的多目标跟踪算法。在两个私有数据集上,改进算法相比基线方法,MOTA(Multiple object tracking accuracy)指标分别提高 1.6、1.7 个百分点,IDF1(Identification F1 score)指标分别提高 1.9、1.0 个百分点。结果显示,改进算法在复杂场景中的跟踪连续性和准确性显著提升。

关键词:多目标跟踪;识别;羊只;BoT-SORT;数据关联;目标检测
中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2025)05-0475-07　　**OSID**:

Sheep Multi-object Tracking Method Integrating Depth Information and Motion Trends

WANG Meili　YANG Ende
(*College of Information Engineering*, *Northwest A&F University*, *Yangling*, *Shaanxi* 712100, *China*)

Abstract: In recent years, the application of information technology in sheep farming has become increasingly sophisticated, necessitating more accurate individual identification and behavior monitoring. This, in turn, has placed higher demands on the accuracy of multiple object tracking (MOT) algorithms, which formed the foundation of these applications. However, existing MOT algorithms often underperformed in scenarios involving object occlusion and dynamic environments. Two novel tracking cues, depth modulated IoU (DIoU) and tracklet direction modeling (TDM), was proposed, aiming at enhancing the precision and robustness of multiple object tracking by supplementing the intersection over union (IoU) cue. DIoU improved the traditional IoU calculation by incorporating depth information of the objects. TDM focused on the movement trends of targets, predicting their future directions based on their historical movement patterns. The DIoU and TDM strategies were integrated into the BoT-SORT algorithm, resulting in an improved multiple object tracking algorithm. Evaluations on two datasets showed that the enhanced algorithm increased the multiple object tracking accuracy (MOTA) by 1.6 percentage points and 1.7 percentage points and the identification F1 score (IDF1) by 1.9 percentage points and 1.0 percentage points, respectively, compared with baseline methods. These results indicated that the improved algorithm significantly enhanced tracking continuity and accuracy in complex scenarios. This research provided insights and methods for multiple object tracking technology, holding significant implications for practical applications.

Key words: multi-object track; recognition; sheep; BoT-SORT; data association; object detection

收稿日期:2024-11-03　修回日期:2025-01-02
基金项目:陕西省秦创原"科学家+工程师"建设项目(2022KXJ-67)
作者简介:王美丽(1982—),女,教授,博士生导师,主要从事计算机图形图像和三维建模等研究,E-mail:wml@nwsuaf.edu.cn

0 引言

人工智能技术促进了养殖管理的精细化和智能化,在羊只养殖产业中发挥着至关重要的作用。养殖场可以实时收集有关羊只的各种数据[1],分析羊群的行为模式[2-3],帮助养殖者做出更加科学的决策。而多目标跟踪技术,用来实现对视频中羊只关联跟踪,是羊只个体识别、计数和行为监测等技术的基础。

近年来,高性能目标检测模型提升了多目标跟踪算法的性能,使基于检测的跟踪算法逐渐成为主流[4-6]。基于目标检测的跟踪算法一般有两个主要步骤[7]:先是逐帧检测目标图像,然后关联已知的对象历史轨迹和当前时刻检测出的对象。其中基于目标检测的跟踪算法关联步骤主要分为 3 类:通过空间位置信息和外观特征进行关联、基于图神经网络进行关联以及基于注意力机制建模的关联算法[8]。基于图神经网络的关联方法将轨迹数据关联任务建模为图论问题。MOTSolv[9]引入了一种可微分的框架,使用消息传递网络(Message passing neural networks, MPNNs)[10]在图域上操作,预测图最终划分为轨迹。该方法用独立的网络来提取特征,导致训练和推理之间不一致。GMTracker[11]提出了可学习图匹配方法来解决这个问题,将轨迹和检测的关系建模为无向图,使关联问题转变为轨迹图和检测图之间的图匹配问题。然而,该方法的主要问题在于训练成本高,求解推理速度往往难以满足实时性。

目前许多工作(TrackFormer[12]、MOTR[13]、MOTRv2[14])尝试使用注意力机制建模关联任务。但是自注意力和前馈神经网络的计算开销大,算法难以实时运行。

通过空间位置信息进行关联的算法 SORT[15]首次以恒速模型假设为基础,引入卡尔曼滤波器建模目标运动,然后用 IoU 指标[16]将预测结果与目标检测的结果关联起来。然而,SORT 在遮挡情况下表现不佳。为了解决这一问题,DeepSORT[17]引入了外观特征线索。外观特征的引入可以帮助轨迹重新找回,实现更稳定的关联。一些场景下相机运动使得以线性运动为假设前提的运动模型失效。BoT-SORT[18]采用基于 Lucas-Kanade 光流法[19]的运动补偿来对齐相邻帧,消除了相机运动对运动模型预测产生的影响。此外,之前的跟踪算法通常忽视了低分检测框。而低分检测框往往是部分被遮挡的对象,导致轨迹中断。ByteTrack[20]通过两阶段的关联,从低分检测框中找出正例。OC-SORT(Observation-centric SORT)[21]利用检测结果预测被遮挡的轨迹,以此修正长时间线性状态估计累积的误差。有时,空间和外观信息同时变得模糊。Hybrid-SORT[22]证明了可以通过结合其他线索来补偿空间和外观线索。

通过空间位置信息进行关联仍是数据关联的重要线索[23],但其往往会受到羊只相互遮挡的影响。同时,外观特征在羊只或相机移动使图像产生运动模糊或在弱光时图像噪点增多的情况下难以关联目标。针对上述问题,本文提出两种关联线索:DIoU 和 TDM。DIoU 是在 IoU 计算中引入深度信息。因为 IoU 方法难以区分遮挡的羊只,而利用深度信息,能够更有效地区分图像中相近的羊只。TDM 关注羊只运动趋势。SORT 类跟踪算法通常假设目标在短时间内的运动是线性的,但在实际养殖环境中,羊只的运动可能因各种因素而突然改变,相机焦距和视角也会突然改变。TDM 通过分析目标的历史运动方向,预测其未来的运动趋势,作为线性运动模型的补充。

1 实验数据

使用自行采集制作的两个私有数据集(PenSheepTrack 和 PTZSheepTrack)以及公开数据集(DanceTrack[24])来评估改进方法的性能和泛化性。

PenSheepTrack 共有视频 19 段,训练集、验证集和测试集分别有 10、3、6 段,每段视频 60 s。从视频中抽出 418 幅图像,训练集 364 幅,验证集 54 幅,用于检测模型训练验证。PTZSheepTrack 共有 20 段视频,训练集、验证集和测试集分别包含 10、4、6 段视频,视频为 25 ~ 60 s 不等。从视频中抽帧 400 幅用于检测模型训练验证,其中训练集 348 幅,验证集 52 幅。数据集涵盖了羊场多个室内圈舍的多个时间段,包括光线充足的白天和光线较弱的傍晚。两个私有数据集以摄像头运动方式为区分。PenSheepTrack 由智能监测小车在圈舍通道采集,主要涉及相机平移运动。PTZSheepTrack 由云台摄像头采集,摄像头只会旋转和调整变焦倍率。两个私有数据集都包含不同大小的羊只和数量不等的圈舍。所采集视频分辨率均为 2 560 像素 ×1 440 像素,帧率 30 f/s。数据集多目标跟踪标注制作使用 DarkLabel,此工具内置了传统跟踪算法,可以辅助标注节省人力。数据集样本如图 1、2 所示。两个私有数据集存在一些共性。在局部区域都会有羊只比较密集造成相互遮挡,这不利于空间位置线索关联羊只。光线充足时,由于相机宽容度不足导致图像中靠近门口的羊只过度曝光。光线不足时,图像噪

点会增多，且由于相机视角移动带来的运动模糊现象更加明显。这两种情况均影响了检测器的精度和外观线索发挥帮助跟踪关联的作用，给跟踪算法带来挑战。



图 1　PenSheepTrack 样本

Fig. 1　PenSheepTrack sample



图 2　PTZSheepTrack 样本

Fig. 2　PTZSheepTrack sample

DanceTrack 数据集视频内容是多人舞蹈表演，所需跟踪的目标频繁移动、相互遮挡。是一个对跟踪算法有很大挑战的数据集，尤其对于小参数规模的检测器。

2　方法

2.1　算法框架

本文方法遵循通过检测进行跟踪（Tracking by detection，TBD）的范式[25]，逐帧检测跟踪。TBD 框架大体分为两个阶段：第一阶段先使用目标检测算法在每个独立帧中定位对象。综合模型精度和速度后，本文选择 YOLO v8n[26] 作为第一阶段的目标检测算法。检测之后，第二阶段涉及跨帧关联相同目标的检测结果，以维持每个跟踪对象的身份 ID 和轨迹。在有一个性能良好的检测器前提下，如何准确关联检测器输出和已知轨迹就成为跟踪算法的关键，跟踪算法常见的错误，如 ID 跳变就是由关联错误导致。所以在第二阶段，为了将检测器的输出与现有轨迹准确关联，提高跟踪性能，本文将 DIoU 和 TDM 引入到 BoT－SORT 中，通过结合多种线索得到更好的关联性能。这里的线索是历史轨迹的预测与当前检测器输出目标的相似度，历史轨迹是一系列已经被成功关联的目标检测结果，跟踪算法根据多种线索计算相似度以关联不同帧之间相同目标的检测结果。在检测器和轨迹预测性能一定的情况下，算法最终的跟踪性能很大程度被相似度计算的方式影响。而现有方法关联线索单一，通过引入多种有效的关联线索，有利于解决跟踪 ID 匹配错误，跟踪性能将得到提高。

具体来说，算法在每一帧的处理流程如图 3 所示。输入为前一帧的图像、轨迹和当前帧的图像，输出为当前帧的轨迹。一共有两个阶段的关联，第一阶段关注高置信度检测框的关联，第二阶段意在挖掘低置信度检测框中潜在的轨迹。匹配算法使用匈牙利算法。第一次匈牙利算法输入的代价是由 DIoU 代价和 TDM 代价组成，即

$$C = \lambda_1 C_{DIoU} + \lambda_2 C_{direction} \tag{1}$$

其中

$$C_{DIoU} = 1 - DIoU \tag{2}$$

$$C_{direction} = |\theta_t - \theta_t^{predict}| \tag{3}$$

式中　C_{DIoU}——DIoU 代价

$C_{direction}$——TDM 代价

λ_1、λ_2——超参数，用于调节权重

θ_t——第 t 帧轨迹观测方向

$\theta_t^{predict}$——第 t 帧轨迹预测方向

其中 $DIoU$、θ_t 和 $\theta_t^{predict}$ 的计算在 2.2 节和 2.3 节详细说明。

轨迹的当前帧状态是前一帧的轨迹经由基于线性假设模型的卡尔曼滤波器预测得到的。预测出的轨迹的当前帧状态还需由一个仿射变化矩阵校准。这个用于校准轨迹的矩阵由来自 BoT－SORT 的全局运动补偿（Global motion compensation，GMC）模块估计相机运动，配准前后两帧图像而来。最后，轨迹管理器综合前面的输出，删除连续多帧未成功匹配的轨迹，用未成功匹配的高分检测框新建轨迹，再加上两次匈牙利算法匹配中成功匹配的轨迹，得到当前帧的轨迹。视频所有帧循环重复以上流程。

轨迹管理还涉及以下方面：①ID 初始化。新建轨迹时，分配当前未使用的最小整数作为轨迹 ID。维护一个全局 ID 计数器，确保 ID 不重复使用。对于置信度高于阈值的检测框才会初始化新轨迹，降低检测器误检导致的 ID 混乱。②ID 保持。通过提出的 DIoU 和 TDM 双重线索提高关联准确性，减少 ID 切换。设置最大容忍帧数 N（本文取 30），若轨迹 N 帧内未匹配则删除，提供因短暂遮挡而丢失跟踪的目标被重新关联的机会，同时避免长期保留错误轨迹。③ID 恢复。轨迹中断后重新出现时，优先考虑与历史轨迹关联。基于卡尔曼滤波的位置预测和本文提出的两种新线索多重验证，确保 ID 恢复的准确性，避免出现 ID 跳变。

2.2　深度调制 IoU

现有方法大多只利用 IoU 线索关联检测框和轨迹，而与某一检测框具有相同或近似 IoU 的检测框可以有完全不同的位置。如果仅仅使用 IoU 线索进行关联，跟踪算法将无法有效区分这些检测框。这可能造成错误的匹配，导致跟踪 ID 变化。这些方法没有完全利用检测器输出的目标空间位置信息，比

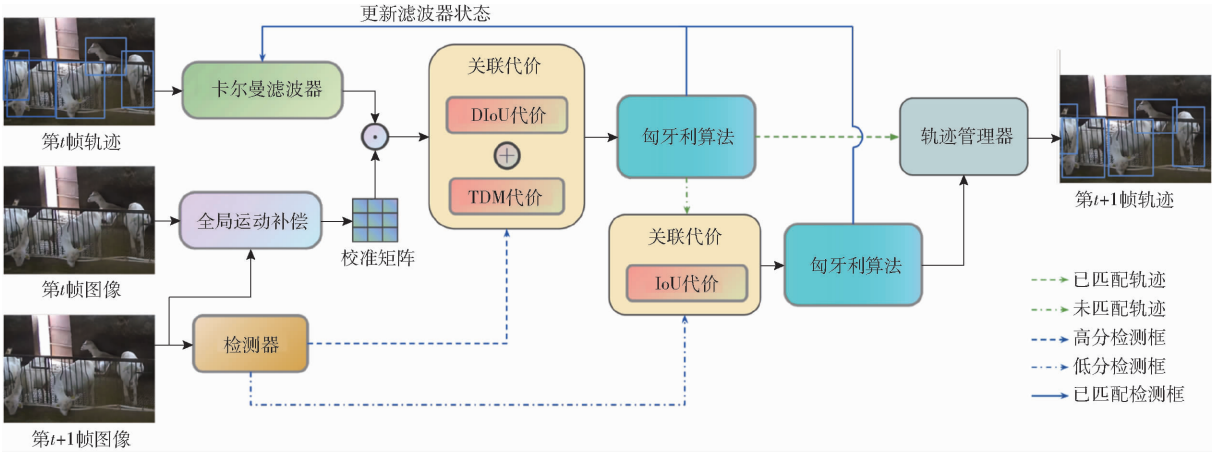

图 3 跟踪算法流程图

Fig. 3 Flowchart of tracking algorithm

如忽略了跟踪对象的深度信息。而本文数据集跟踪的羊只在场景中的深度变化是相对稳定的。深度信息对数据关联有帮助的原因是当 IoU 线索失败时(比如不同检测框与同一轨迹有相同或近似的 IoU),跟踪目标的深度仍然有区分度,利用深度信息可以帮助区分这些目标检测框,并且深度信息的稳定性相比 IoU 受对象姿势变化造成的检测框大小变化的影响更小,这使跟踪器更容易做出正确的关联。

由于数据集中缺少直接的深度信息,一个高效的伪深度估计方法就显得尤为重要。本文使用 SparseTrack[27] 提出的伪深度估计方法。该方法以检测框底边到整个相平面底边的距离作为检测框的伪深度。因为考虑到羊场的一般特性:要跟踪的羊只都在空间中一个平面上,即地面;拍摄所用摄像头一般安置在平面侧上方。所以,羊只在空间中的深度和检测框底边与图像底边的距离呈正相关。这一方法在提供可用深度信息的同时兼顾了极高的计算效率。虽然这一伪深度估计方法的结果无法代表真实三维空间内的深度值,但完全可以反映出羊只之间的相对深度关系,这有利于区分 IoU 相近的检测框。

为了更好地利用深度信息,将深度信息引入 IoU 中,提出 DIoU,即

$$DIoU = IoU\left(1 - \frac{|d_a - d_b|}{h_a}\right) \tag{4}$$

其中

$$IoU = \frac{S(A \cap B)}{S(A \cup B)} \tag{5}$$

式中 A——检测框 B——预测框

S——计算图形面积的函数

$A \cap B$——A 与 B 重叠部分构成的几何图形

$A \cup B$——A 与 B 合并所构成的几何图形

d_a、d_b——伪深度 h_a——高度

d_a、d_b和 h_a如图 4 所示。由于透视效应,在三维空间中同样的深度差在越深的位置越小,同样高度的检测框在越深的位置越小。故而将检测框的高度考虑其中,使用式(4)的计算方式可以消除这种效应。

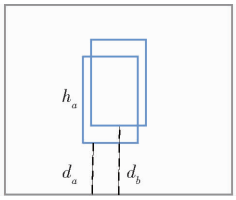

图 4 DIoU 计算示意图

Fig. 4 DIoU calculation illustration

2.3 轨迹方向建模

轨迹方向建模尝试从轨迹运动方向的角度为检测框和轨迹的关联补充更多线索。

OC - SORT 首次将以观测为中心的动量(Observation - centric momentum)引入 SORT 类跟踪算法中,实际上此处的动量与物理学中的动量不同,其仅是速度的方向。这一方法假设轨迹在短的间隔内可以近似为线性运动,而线性运动具有不变的速度方向。由于视频数据是离散的,计算速度方向需要取两个时刻的状态,其中时间差 Δt 的选取尤为重要。在线性运动模型下,Δt 越大速度方向估计的噪声越小。而 Δt 越小越能满足线性运动的假设前提,如此才有速度方向不变的推论,这其中存在天然的矛盾[21]。

本文对轨迹方向的建模中放弃方向不变的假设,利用线性模型从轨迹历史观测的速度方向估计下一帧的速度方向。线性建模的预测方法为

$$\theta_t^{predict} = \theta_{t-1} + \frac{\sum_{i=1}^{n}(\theta_{t-i} - \theta_{t-i-1})}{n} \tag{6}$$

其中

$$n = \min(K, M-1) \tag{7}$$

式中 K——超参数,至多取 K 个历史观测

M——历史观测数量

n——历史观测速度方向数量

轨迹的观测速度方向计算方式为

$$\theta_t = \arctan\frac{y_t - y_{t-k}}{x_t - x_{t-k}} \quad (k>1) \tag{8}$$

式中　(x_t, y_t)——轨迹第 t 帧检测框的中心点坐标

一般取 $k>1$ 表示使用非相邻两帧计算速度方向，这是因为相邻两帧的速度方向受检测器误差抖动影响较大。即使对于一个静止物体的相邻两帧检测器也有极大可能得出位置不完全相同的两个检测框，大于 1 的步长可以降低这种误差的影响。

3　实验

3.1　评价指标

选择多目标跟踪领域常用评价指标全面评估算法性能，包括 MOTA、MOTP(Multiple object tracking precision)[28] 和 IDF1、IDP(Identification precision)、IDR(Identification recall)[29]。MOTP 评估目标位置预测的准确性，反映了算法对目标位置坐标的精度。MOTA 综合考虑了漏检、误检和身份切换 3 方面，以量化算法的整体跟踪性能。IDF1 则侧重于身份保留的效果，衡量算法在跟踪过程中保持目标身份一致性的能力。IDP 和 IDR 是 IDF1 的组成部分，分别衡量跟踪算法的身份精度和身份召回率。以上评价指标中，MOTA、IDF1、IDP 和 IDR 越高说明算法性能越好，MOTP 越低说明算法性能越好。通过结合这些指标，可以全面分析和比较不同 MOT 算法在定位精度、整体准确性和身份一致性方面的优劣。

3.2　实现细节

本文研究集中在跟踪关联的部分。为了公平地比较并体现改进算法的优越性，选择公开可用的目标检测模型 YOLO v8n 作为检测器，并以其在 COCO[30] 数据集上的预训练权重作为初始权重。目标检测模型训练时输入图像尺寸为 640 像素×640 像素，是在预处理阶段由原始视频帧缩放并填充而来，批量大小为 16，共训练 100 轮。其他设置与 YOLO v8n 公布代码的默认设置相同。跟踪部分设置 τ_{high} 和 τ_{low} 分别为 0.7 和 0.1，GMC 使用 Lucas－Kanade 光流法，表示 DIoU 和 TDM 在总代价所占比重的 λ_1 和 λ_2 均设置为 0.2。评价指标计算使用 py-motmetrics 项目。实验软硬件平台配置如表 1 所示。

3.3　对比实验

对于所有的算法，本文使用相同的检测器和检测模型权重做对比。在以下评测的两个私有数据集和一个公开数据集上，本文算法均优于基线方法(BoT－SORT)，且额外的计算可忽略不计。没有相机运动补偿的方法如 SORT、OC－SORT 和 ByteTrack，IDF1 参数明显落后于其他方法。这说明了相机运动补偿对于跟踪身份保持的重要性。

表 1　软硬件平台配置

Tab.1　Hardware and software platform

软硬件类型	配置
中央处理器	Intel(R) Xeon(R) Platinum 8160T
图形处理器	NVIDIA RTX A5000
操作系统	Ubuntu 18.04
开发语言	Python 3.11.5
深度学习库	Pytorch 2.1.0

改进算法对比其他算法在 PenSheepTrack 数据集的评估结果如表 2 所示。在 PenSheepTrack 数据集上，改进算法的 IDF1、MOTA 和 MOTP 为 95.8%、92.1% 和 17.3%。对比基线方法(BoT－SORT)，IDF1 和 MOTA 分别高出 1.9、1.6 个百分点，MOTP 持平。对比 SparseTrack，在 MOTP 相近的情况下，IDF1 和 MOTA 分别高出 1.8、3.8 个百分点。

表 2　不同算法在 PenSheepTrack 上的对比结果

Tab.2　Comparative results of different algorithms on PenSheepTrack

算法	MOTA/%	MOTP/%	IDF1/%	IDP/%	IDR/%	IDs
SORT	88.6	19.4	79.6	82.5	77.0	91
OC－SORT	88.3	17.4	85.0	86.0	84.0	55
ByteTrack	88.9	20.4	83.8	84.2	83.5	63
SparseTrack	88.3	17.0	94.0	96.9	91.2	26
BoT－SORT	90.5	17.3	93.9	93.7	94.1	43
本文	92.1	17.3	95.8	96.9	94.7	27

改进算法与其他算法在 PTZSheepTrack 数据集上的对比实验结果如表 3 所示。改进算法的 IDF1、MOTA 和 MOTP 分别为 82.4%、71.9% 和 20.0%。相比 BoT－SORT，IDF1 和 MOTA 提升 1.0、1.7 个百分点，MOTP 相近。对比 SparseTrack，MOTP 基本持平，而 IDF1 和 MOTA 分别高出 1.7、1.1 个百分点。由表 2、3 可知，所有对比的算法在

表 3　不同算法在 PTZSheepTrack 上的对比结果

Tab.3　Comparative results of different algorithms on PTZSheepTrack

算法	MOTA/%	MOTP/%	IDF1/%	IDP/%	IDR/%	IDs
SORT	62.6	20.6	72.5	86.9	62.3	104
OC－SORT	66.2	19.1	73.7	87.0	63.9	59
ByteTrack	66.9	22.0	74.8	86.4	65.9	93
SparseTrack	70.8	19.8	80.7	95.4	70.0	15
BoT－SORT	70.2	19.6	81.4	93.1	72.4	15
本文	71.9	20.0	82.4	96.8	71.8	14

PTZSheepTrack 数据集上的表现均不如 PenSheepTrack 数据集,原因是 PTZSheepTrack 涉及更复杂的摄像头运动,包括旋转和焦距变化。焦距变化使得目标羊只检测框大小变化,而摄像头旋转给目标羊只带来了比 PenSheepTrack 中摄像头平移更快速的相对移动。

为说明改进算法的泛化性,本文在公开数据集 DanceTrack 上对比了其他算法,结果如表 4 所示。改进算法的 IDF1、MOTA、MOTP 分别为 39.6%、74.3% 和 17.3%。与基线方法(BoT-SORT)相比,IDF1 和 MOTA 分别提高 1.2、0.5 个百分点。相比 SparseTrack,MOTA 高出 2.4 个百分点,但 MOTP 和 IDF1 略低。因为 DanceTrack 数据集视频拍摄角度多为平视,且跟踪目标在空间中多集中在相差不大的深度范围内,这阻碍了 DIoU 利用深度信息区分目标的效果。结果表明,改进算法的泛化性良好,即使 DanceTrack 数据集并非本文所针对优化的场景,效果也不落后于现有方法。

表 4 不同算法在 DanceTrack 上的对比结果

Tab. 4 Comparative results of different algorithms on DanceTrack

算法	MOTA/%	MOTP/%	IDF1/%	IDP/%	IDR/%	IDs
SORT	70.3	17.4	33.8	37.9	30.4	3 808
OC-SORT	72.0	19.0	39.3	40.6	38.0	4 308
ByteTrack	72.4	21.5	37.5	39.1	36.0	3 411
SparseTrack	71.9	16.5	39.9	44.6	36.1	2 509
BoT-SORT	73.8	17.7	38.4	39.7	37.1	3 597
本文	74.3	17.3	39.6	42.4	37.2	2 840

图 5、6 分别为两个算法跟踪可视化示例,左右两幅图像截取自视频中的前后两帧。图 5 为 BoT-SORT 结果,图 6 为本文改进算法结果。由图 5 可知,BoT-SORT 发生了 ID 切换,图中 BoT-SORT 所跟踪的一个目标 ID 由 7 变为 17,而 7 号轨迹被错误关联到之前的 10 号轨迹。但本文算法在此例中则没有发生错误的 ID 切换和跟踪丢失。



图 5 BoT-SORT 跟踪结果示例

Fig. 5 Examples of BoT-SORT tracking results

结果表明,改进算法性能优于其他对比算法,且额外的计算成本可以忽略。DIoU 有效利用了轨迹的深度信息,而轨迹运动方向的预测解决了复杂场景下线性运动假设失效的问题。这说明,在因遮挡



图 6 改进算法跟踪结果示例

Fig. 6 Examples of method tracking results

导致 IoU 线索失效时,因光线和相机运动使外观线索失效时,本文引入的关联线索有助于保持跟踪器正确跟踪,提高羊只多目标跟踪性能。

3.4 消融实验

表 5、6 分别为在两个数据集上的消融实验结果,其中 3 行数据分别代表原始 BoT-SORT 算法、引入 DIoU 的算法和同时引入 DIoU 和 TDM 的算法。在 PenSheepTrack 上,DIoU 使 MOTA 和 IDF1 提高 1.1、0.6 个百分点,TDM 进一步使 MOTA 和 IDF1 提高 0.5、1.3 个百分点。在 PTZSheepTrack 上,DIoU 使 IDF1 提高 1.3 个百分点,但略微降低了 MOTA。TDM 的引入在几乎不影响其他指标的基础上,抵消了 DIoU 对 MOTA 指标的负面影响,并相对原始方法提高 1.7 个百分点。这表明,DIoU 和 TDM 在某些情况下可以相互补充,共同弥补单一 IoU 线索不总是有效的问题。结果表明了本文所提关联线索的有效性。

表 5 在 PenSheepTrack 上的消融实验结果

Tab. 5 Ablation experiment results on PenSheepTrack

%

DIoU	TDM	MOTA	MOTP	IDF1
×	×	90.5	17.3	93.9
√	×	91.6	17.4	94.5
√	√	92.1	17.3	95.8

注:√表示使用此算法,×表示不使用此算法。下同。

表 6 在 PTZSheepTrack 上的消融实验结果

Tab. 6 Ablation experiment results on PTZSheepTrack

%

DIoU	TDM	MOTA	MOTP	IDF1
×	×	70.2	19.6	81.4
√	×	69.4	19.3	82.7
√	√	71.9	20.0	82.4

4 结论

(1)提出了一种基于 BoT-SORT 的改进多目标跟踪算法,融合了两种新的关联线索:DIoU 和 TDM。DIoU 通过引入伪深度信息,改进传统的 IoU 计算方法,有效区分图像中位置接近的羊只,从而提高跟踪精度。TDM 通过分析羊只的历史移动方向,结合线性预测方法,提供更为多样的数据关联线索,

可作为因跟踪羊只外观近似和低图像质量而失效的外观线索的补充。将两种线索集成到 BoT－SORT 算法中，形成了一种改进的多目标跟踪算法。

（2）结果表明，改进算法在羊场养殖场景中的跟踪连续性和准确性有所提升。对比基线方法，在 PenSheepTrack 数据集上 MOTA 和 IDF1 分别提高 1.6、1.9 个百分点，在 PTZSheepTrack 数据集上 MOTA 和 IDF1 分别提高 1.7、1.0 个百分点。

参 考 文 献

[1] 宁纪锋，张静，杨蜀秦，等. 基于 DiMP 的类特定养殖奶山羊跟踪方法[J]. 农业机械学报，2023，54(6)：280－286.
NING Jifeng，ZHANG Jing，YANG Shuqin，et al. Class-special real-time dairy goat tracking method based on DiMP[J]. Transactions of the Chinese Society for Agricultural Machinery，2023，54(6)：280－286.（in Chinese）

[2] 袁洪波，曹润柳，程曼. 融合 Res3D、BiLSTM 和注意力机制的羊只行为识别方法[J]. 农业机械学报，2024，55(4)：221－230.
YUAN Hongbo，CAO Runliu，CHENG Man. Fusion of Res3D，BiLSTM and attention mechanism for sheep behavior recognition method[J]. Transactions of the Chinese Society for Agricultural Machinery，2024，55(4)：221－230.（in Chinese）

[3] 王旺，王福顺，张伟进，等. 基于改进 YOLO v8s 的羊只行为识别方法[J]. 农业机械学报，2024，55(7)：325－335.
WANG Wang，WANG Fushun，ZHANG Weijin，et al. Sheep behavior recognition method based on improved YOLO v8s[J]. Transactions of the Chinese Society for Agricultural Machinery，2024，55(7)：325－335.（in Chinese）

[4] ZHANG Y，WANG C，WANG X，et al. Fairmot：on the fairness of detection and re-identification in multiple object tracking[J]. International Journal of Computer Vision，2021，129：3069－3087.

[5] QIN Z，ZHOU S，WANG L，et al. Motiontrack：learning robust short-term and long-term motions for multi-object tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2023：17939－17948.

[6] DU Y，ZHAO Z，SONG Y，et al. Strongsort：make deepsort great again[J]. IEEE Transactions on Multimedia，2023，25：8725－8737.

[7] 张瑶，卢焕章，张路平，等. 基于深度学习的视觉多目标跟踪算法综述[J]. 计算机工程与应用，2021，57(13)：55－66.
ZHANG Yao，LU Huanzhang，ZHANG Luping，et al. Overview of visual multi-object tracking algorithms with deep learning[J]. Computer Engineering and Applications，2021，57(13)：55－66.（in Chinese）

[8] LUO W，XING J，MILAN A，et al. Multiple object tracking：a literature review[J]. Artificial Intelligence，2021，293：103448.

[9] BRASÓ G，LEAL-TAIXÉ 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.

[10] GILMER J，SCHOENHOLZ S S，RILEY P F，et al. Neural message passing for quantum chemistry[C]// International Conference on Machine Learning. PMLR，2017：1263－1272.

[11] HE J，HUANG Z，WANG N，et al. Learnable graph matching：incorporating graph partitioning with deep feature learning for multiple object tracking[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：5299－5309.

[12] MEINHARDT T，KIRILLOV A，LEAL-TAIXE L，et al. Trackformer：multi-object tracking with transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022：8844－8854.

[13] ZENG F，DONG B，ZHANG Y，et al. Motr：end-to-end multiple-object tracking with transformer[C]// European Conference on Computer Vision. Springer，2022：659－675.

[14] ZHANG Y，WANG T，ZHANG X. Motrv2：bootstrapping end-to-end multi-object tracking by pretrained object detectors[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2023：22056－22065.

[15] BEWLEY A，GE Z，OTT L，et al. Simple online and realtime tracking[C]//2016 IEEE International Conference on Image Processing（ICIP）. IEEE，2016：3464－3468.

[16] YU J，JIANG Y，WANG Z，et al. Unitbox：an advanced object detection network[C]// Proceedings of the 24th ACM International Conference on Multimedia，2016：516－520.

[17] WOJKE N，BEWLEY A，PAULUS D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing（ICIP）. IEEE，2017：3645－3649.

[18] AHARON N，ORFAIG R，BOBROVSKY B-Z. BoT－SORT：robust associations multi-pedestrian tracking[J]. arXiv Preprint，arXiv：2206.14651，2022.

[19] LUCAS B D，KANADE T. An iterative image registration technique with an application to stereo vision[C]// IJCAI'81：7th International Joint Conference on Artificial Intelligence，1981：674－679.

[20] ZHANG Y，SUN P，JIANG Y，et al. Bytetrack：multi-object tracking by associating every detection box[C]// European Conference on Computer Vision. Springer，2022：1－21.

[21] CAO J，PANG J，WENG X，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.

（下转第 491 页）

[23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.
[24] BOCHKOVSKIY, ALEXEY, WANG Chien-Yao, et al. Yolov4: optimal speed and accuracy of object detection[J]. arXiv Preprint, arXiv:2004. 10934,2020.
[25] ZHENG Z, 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.
[26] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146 – 157.
[27] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658 – 666.
[28] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[J]. arXiv Preprint, arXiv:2205. 12740, 2022.
[29] CAI Z, VASCONCELOS N. Cascade R – CNN: delving into high quality object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154 – 6162.
[30] BOLYA D, ZHOU C, XIAO F, et al. Yolact: real-time instance segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9157 – 9166.
[31] WANG X, KONG T, SHEN C, et al. Solo: segmenting objects by locations[C]// Computer Vision-ECCV 2020: 16th European Conference, 2020: 649 – 665.

(上接第 **481** 页)

[22] YANG M, HAN G, YAN B, et al. Hybrid-sort: weak cues matter for online multi-object tracking[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2024: 6504 – 6512.
[23] 龚轩, 乐孜纯, 王慧, 等. 多目标跟踪中的数据关联技术综述[J]. 计算机科学, 2020, 47(10): 136 – 144
GONG Xuan, LE Zichun, WANG Hui, et al. Survey of data association technology in multi-target tracking[J]. Computer Science, 2020, 47(10): 136 – 144. (in Chinese)
[24] SUN P, CAO J, JIANG Y, et al. Dancetrack: multi-object tracking in uniform appearance and diverse motion[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 20993 – 21002.
[25] SUN Z, CHEN J, CHAO L, et al. A survey of multiple pedestrian tracking based on tracking-by-detection framework[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(5): 1819 – 1833.
[26] JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLOv8[Z]. 2023. https://github. com/ultralytics/ultralytics.
[27] LIU Z, WANG X, WANG C, et al. Sparsetrack: multi-object tracking by performing scene decomposition based on pseudo-depth[J]. arXiv Preprint, arXiv:2306. 05238, 2023.
[28] BERNARDIN K, STIEFELHAGEN R. Evaluating multiple object tracking performance: the clear mot metrics[J]. EURASIP Journal on Image and Video Processing, 2008,2008: 1 – 10.
[29] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]// European Conference on Computer Vision, 2016: 17 – 35.
[30] LIN T-Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context[C]// Computer Vision-ECCV 2014: 13th European Conference, 2014: 740 – 755.