

基于YOLOX-BT的奶牛多目标跟踪方法

王克俭¹ 刘玲玲¹ 司永胜¹ 何振学¹ 马亚宾² 袁明²

(1. 河北农业大学信息科学与技术学院, 保定 071001; 2. 河北省畜牧良种工作站, 石家庄 050061)

摘要: 奶牛跟踪是获取奶牛位置信息和活动数据的重要方式,对奶牛的智能管理具有重要意义。针对视频中部分奶牛目标较小、存在重叠与环境遮挡造成的漏检、误检和轨迹断裂的问题,提出了YOLOX-BT算法,该算法将YOLOX-s的CSP2_1结构替换成了CFP_EVCBlock结构,用于提升小目标检测性能;在CFP_EVCBlock结构中,通过整合多种特征增强机制设计了特征重标定模块(Feature refinement module, FRM),解决重叠与遮挡问题,实现了对重要特征的有效提取;为了改善轨迹断裂情况,在数据关联部分,引入外观相似度改进相似度计算方法,并且在匹配过程中,动态调整外观相似度匹配和IoU匹配的所占权重,提升了奶牛的匹配精度。实验结果表明,改进后的检测器平均检测精度、精确率、召回率、帧速率分别达到了94.9%、94.2%、92.7%和22.9 f/s,较原模型相比分别提高了1.2、0.8、2.3个百分点和29.4%,参数量下降了9.3%,浮点运算量减少了16.7%。YOLOX-BT方法在奶牛多目标跟踪任务上,MOTA为93.2%,IDF1为95.2%,IDs为13,MOTP为5.3%,与目前常用方法相比,在MOTA和IDF1上均有显著提升,并有效减少了ID切换。YOLOX-BT算法能实现养殖场奶牛的准确且持续有效跟踪,能够为奶牛智能管理提供技术支持。

关键词: 奶牛; 多目标跟踪; ByteTrack; FRM; CFP-EVCBlock

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2026)11-0364-11

OSID:



Multi-object Tracking Method for Cows Based on YOLOX-BT

WANG Kejian¹ LIU Lingling¹ SI Yongsheng¹ HE Zhenxue¹ MA Yabin² YUAN Ming²

(1. College of Information Science and Technology, Hebei Agricultural University, Baoding 071001, China

2. Hebei Provincial Station for Livestock Varieties Producing and Spreading, Shijiazhuang 050061, China)

Abstract: Cow tracking is an essential technique for obtaining positional and behavioral data of cows, playing a vital role in intelligent farm management. To tackle challenges such as missed and false detections, as well as track fragmentation caused by small object sizes, occlusion, and overlap in farm videos, the YOLOX-BT algorithm was proposed. This method enhanced both detection and tracking performance. Specifically, the YOLOX-s model was improved by replacing the original CSP2_1 structure with the CFP_EVCBlock, boosting small object detection capability. Within CFP_EVCBlock, a feature refinement module (FRM) was designed by integrating multiple feature enhancement mechanisms, enabling prioritized extraction of important features and effectively mitigating occlusion and overlap issues. In the tracking phase, appearance similarity was introduced into the data association process, and the weights between appearance and IoU matching were dynamically adjusted, improving cow identity matching accuracy. Experimental results showed that the improved detector achieves an average precision of 94.9%, precision of 94.2%, recall of 92.7%, and an inference speed of 22.9 f/s, representing gains of 1.2, 0.8, 2.3 percentage points, and 29.4%, respectively, over the baseline. Additionally, the number of parameters and FLOPs were reduced by 9.3% and 16.7%, respectively. In the cow multi-object tracking task, YOLOX-BT achieved a MOTA of 93.2%, IDF1 of 95.2%, 13 ID switches, and a MOTP of 5.3%. Compared with commonly used methods, the proposed algorithm showed significant improvements in both tracking accuracy and identity consistency, providing robust technical support for intelligent cow management.

Key words: cows; multi-object tracking; ByteTrack; FRM; CFP-EVCBlock

收稿日期: 2025-01-06 修回日期: 2025-04-28

基金项目: 河北省重大科技支撑计划项目(252N7403D)和国家自然科学基金项目(62102130)

作者简介: 王克俭(1971—),女,教授,主要从事农业领域图像识别、视频处理及智能算法研究,E-mail: wkj71@163.com

通信作者: 马亚宾(1975—),男,研究员,主要从事奶牛生产性能测定和高效养殖研究,E-mail: dhimyb@163.com

0 引言

奶牛的健康和生产状况直接影响到奶制品行业的经济效益和可持续发展^[1]。利用先进的信息技术实现智能养殖,保障奶牛身体健康、提高奶牛生活福利和养殖场经济效益已经成为奶牛养殖研究领域的重要目标和任务^[2]。在实现智能养殖的众多技术手段中,计算机视觉技术凭借其非接触和自动化等特点,在奶牛多目标跟踪方面展现出了独特优势^[3]。

目前,国内外针对行人、车辆等多目标跟踪任务已经进行了很多研究,并已取得了很多成果。关于动物类多目标跟踪方法,目前可分为接触式^[4-5]和非接触式^[6]两大类。接触式跟踪方法主要是通过 在动物身上安装跟踪设备来实现,例如项圈、耳标等,通过读取设备内数值,来确定时间、目标 ID 以及动物位置等信息^[7-8]。非接触式方法可分为基于检测的跟踪 (Tracking by detection, TBD) 和检测跟踪一体 (Joint tracking and detection, JDT) 两大类^[9]。在 TBD 模式中,通过目标检测算法识别视频帧中的目标,之后利用检测结果进行后续跟踪^[10-15]。TBD 方法具有更高的灵活性,能够利用高性能检测器进行目标检测,并通过独立的跟踪算法优化轨迹管理。在处理复杂场景时,能够通过调整检测器和跟踪器的参数来适应不同的环境需求。JDT 模式则将跟踪和检测过程结合在一起^[16]。JDT 方法将检测和跟踪的子模块融合在一个单一的框架中,虽然在理论上可以提高推理速度,但在实际实现中,各个子模块之间的协同训练互相干扰,不易平衡。在处理大型场景或高帧率视频时,模型的计算量和内存占用量可能会显著增加,实现成本较高。

虽然接触式方法具有较高的跟踪精度,但是价格昂贵,设备维护较为困难,甚至会影响奶牛健康。非接触式方法避免了可穿戴设备老化、损坏等硬件问题,但仍存在一些难点,由于奶牛养殖场景复杂,目标遮挡、光照变化等情况时有发生,进而导致跟踪性能下降。

针对上述问题,TBD 方法训练时具有更优的平衡性,跟踪效率与性能较好。因此,本文采用 TBD 框架,并通过改进跟踪算法进一步提升奶牛多目标跟踪的准确性与稳定性,设计一种 YOLOX - BT 算法,以牧场自然环境下的群体养殖奶牛作为监测目标,监控视频作为数据源,通过改进的 YOLOX 检测器和改进的 ByteTrack 跟踪策略的有效结合,既避免传统可穿戴设备的局限性,又缓解传统方法在处理复杂场景时的性能下降问题,以期 为研究奶牛多目

标跟踪任务提供参考和借鉴。

1 实验数据

1.1 视频数据采集

实验视频采集自河北省保定市满城区宏达牧场,利用安装在牛棚棚顶并通过互联网控制的 4 个广角摄像头 (DH - IPC - HFW8449K - ZFR 型,大华) 进行采集,牛舍总长 50 m、宽 30 m,摄像头安装在牛舍中部的左右两侧,相邻 2 台摄像机之间距离为 12 m,高度 6 m,俯视角约为 15°,摄像机焦距为 3.6 mm,传感器分辨率为 2 688 像素 × 1 520 像素。图像采集环境如图 1 所示。

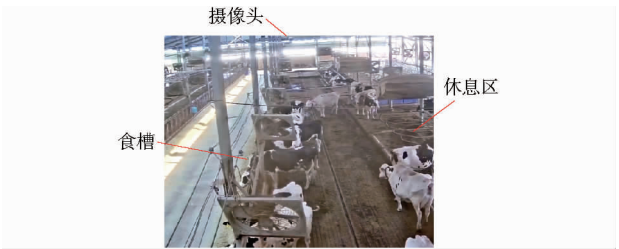


图 1 图像采集现场环境

Fig. 1 Image acquisition site environment

1.2 数据集构建

为了考虑光照的影响,视频数据采集时间包括白天 (11:00—14:00)、傍晚 (16:00—19:00) 和夜间 (20:00—21:00),连续采集 20 d,筛选保留 131 段奶牛移动较多且同类视频中奶牛数量相同的有效视频,其中包含 43 段白天视频,43 段夜间视频以及 45 段傍晚视频。在人工观察的基础上,将 45 段傍晚视频分为 2 类,第 1 类视频奶牛活动较为频繁,共 24 段;第 2 类视频奶牛活动较少,共 21 段。每段视频 30 s,每秒 10 帧。依据照明情况和活动程度将视频分为 4 类,依次为:夜间环境奶牛活动情况适中、傍晚环境奶牛活动情况频繁、白天环境奶牛活动情况适中、傍晚环境奶牛活动情况较少,分别对应视频类别号 A、B、C、D。视频处理后分辨率为 1 920 像素 × 1 080 像素,使用 cvat 软件对视频段进行标注,构建奶牛多目标跟踪数据集,在每种类别当中,按照比例 1:1 分为训练集和测试集。在视频数据采集期间,光照条件为自然光,养殖环境温度适宜,通风良好。具体信息如表 1 所示。

表 1 奶牛多目标跟踪数据集样本数

Tab. 1 Multi-object tracking dataset for dairy cows

视频类别号	视频段数量	照明情况	活动程度	牛只数量
A	43	夜间	适中	23
B	24	傍晚	频繁	12
C	43	白天	适中	18
D	21	傍晚	较少	9

2 研究方法

本文提出的 YOLOX-BT 算法,通过对目标检测和数据关联过程的优化,实现了对奶牛目标的稳定跟踪,跟踪流程图如图 2 所示。采用改进的 YOLOX-s 模型作为目标检测器,对输入的视频序列进行逐帧目标检测,获取每帧中奶牛的位置和外观特征。计算当前帧与历史帧之间的外观相似度和 IoU 相似度,并将两者进行加权融合得到融合相似度。基于该融合相似度,采用匹配算法实现当前检测结果与已有轨迹的关联,从而完成对视频序列中奶牛个体的持续跟踪。

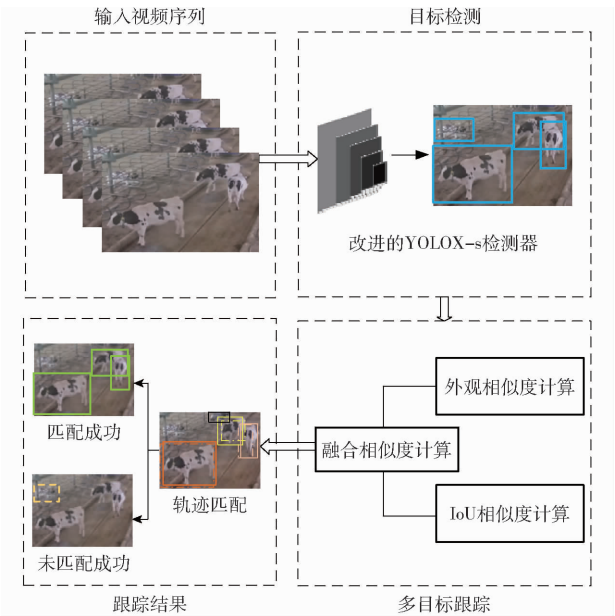


图2 奶牛多目标跟踪过程图

Fig.2 Flow chart of multi-object tracking process for cows

2.1 改进的 YOLOX-s 目标检测器

YOLOX-s 模型保持了 YOLOX 系列核心的无锚框设计和解耦头结构,SimOTA 标签分配策略能够使检测结果更准确,网络架构更精简,在保证基本检测精度的同时大幅降低了计算开销,适用于奶牛检测,改进后的 YOLOX-s 模型结构如图 3 所示。

为了增强 YOLOX-s 检测网络的特征提取能力,本文在检测网络的 Neck 部分做出改进,将原有的 CSP2_1 特征融合模块替换为 CFP-EVCBlock^[17] 结构,并在其中引入了特征重标定模块 (Feature refinement module, FRM)^[18-20]。CFP-EVCBlock 通过结合固有字典编码器 (LVCBlock) 和轻量级 MLP 来捕获信息,FRM 则利用通道注意力和空间注意力机制,自适应地调整特征权重分配,从通道维度和空间维度对特征进行重标定,提升了模型对关键视觉特征的感知能力。

2.1.1 CFP-EVCBlock

通过视频数据可知,大多数奶牛距离摄像头较远,在画面中所占比例较小,因此本文选择在检测器的颈部引入 CFP-EVCBlock 模块来增强对小目标的检测能力。CFP-EVCBlock 主要由 2 个关键分支组成:LVCBlock 和 LightMLPBlock,并引入 FRM 增强特征表示能力,如图 4 所示。

CFP-EVCBlock 的输入经过 7×7 的卷积层,结合批归一化和 ReLU 激活函数,对特征进行初步处理,经过最大池化后,进入 LVCBlock 和 LightMLPBlock 分支,得出的特征图在通道维度进行拼接,再经过卷积层、归一化层和 ReLU 激活函数后,加入 FRM,输出最后的特征图。

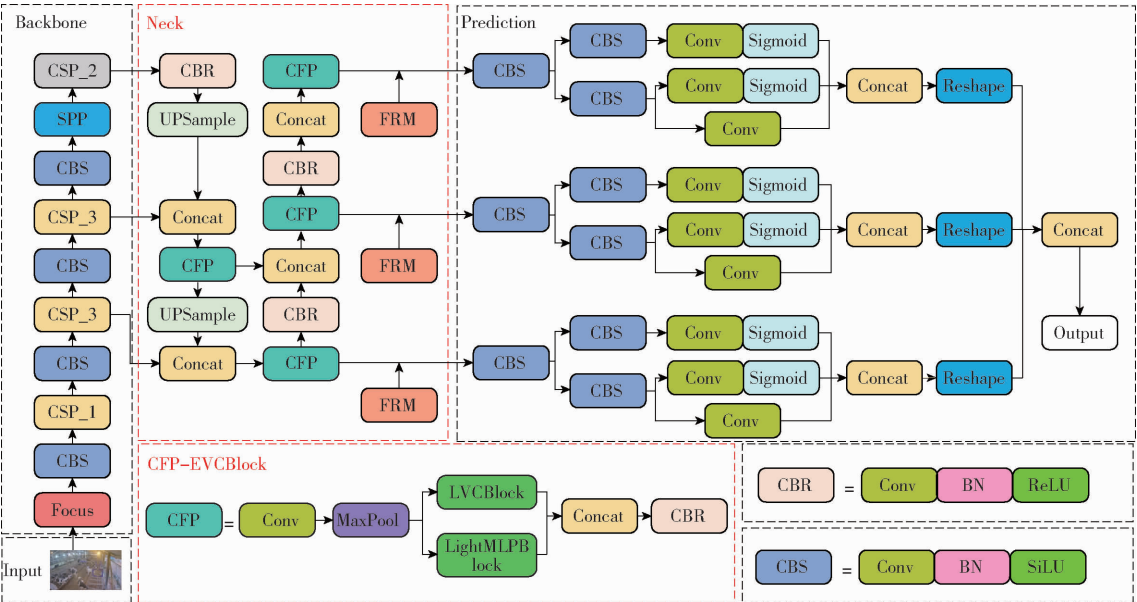


图3 改进后的 YOLOX-s 网络模型

Fig.3 The improved YOLOX-s network model

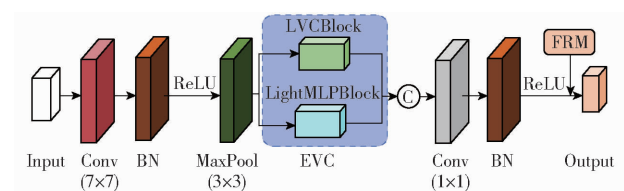


图 4 CFP-EVCBlock 模型结构图

Fig. 4 CFP-EVCBlock model structure diagram

LVCBlock 通过融合卷积和编码机制来增强特征表达能力。如图 5 所示,首先使用带有残差连接的 ConvBlock,结合 BatchNorm 和 ReLU 激活函数进行特征重构,进入 LVC 编码分支,计算输入与编码词之间缩放的 L2 范数,利用 softmax 函数生成分配权重并加权求和以获得编码特征。编码特征经过归一化、非线性变换和平均池化后,通过全连接层和 Sigmoid 函数生成通道注意力权重,与原始特征相乘并结合残差连接得到输出。这种设计能够有效捕获

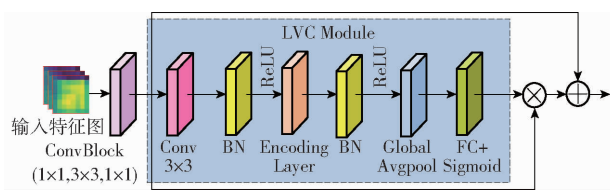


图 5 LVCBlock 模型结构图

Fig. 5 LVCBlock model structure diagram

特征的全局上下文信息,同时保持计算效率。

如图 6 所示,LightMLPBlock 采用轻量级混合架构设计:在深度卷积 (DWConv) 分支中,输入特征经 GroupNorm 后通过深度可分离卷积^[21]处理,结合 Layer Scale 和 DropPath 后与输入特征残差连接,得到的特征进入 MLP 模块,经 GroupNorm 和 1×1 卷积,通过 Layer Scale 和 DropPath 后再次残差连接。这种串联结构有效结合了 DWConv 和 MLP 的优势,并通过 Layer Scale 和 DropPath 增强了模型泛化能力。

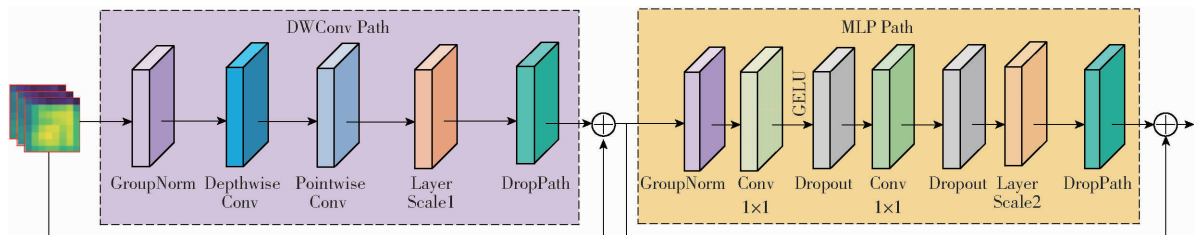


图 6 LightMLPBlock 模型结构图

Fig. 6 LightMLPBlock model structure diagram

2.1.2 特征重标定模块

本文在检测器部分设计了特征重标定模块 (FRM),该模块通过结合全局特征提取、局部依赖建模和空间特征增强机制,提升了模型对关键特征的感知能力,如图 7 所示。其中,通道全局特征提取能够有效捕捉奶牛的整体特征和花纹分布,在不同光照条件下保持稳定的检测效果;通道间局部依赖

建模则专注于处理奶牛的局部特征关联,主要用于在密集场景中区分个体边界和处理部分遮挡情况;空间特征增强机制则提升了模型对不同距离的奶牛的检测能力。这 3 个组件协同工作,共同增强了特征表示能力,使模型能够更好地应对牧场环境中光照变化、拥挤和遮挡等复杂情况。

(1) 全局特征提取。对于输入 $X \in \mathbf{R}^{C' \times H' \times W'}$,通

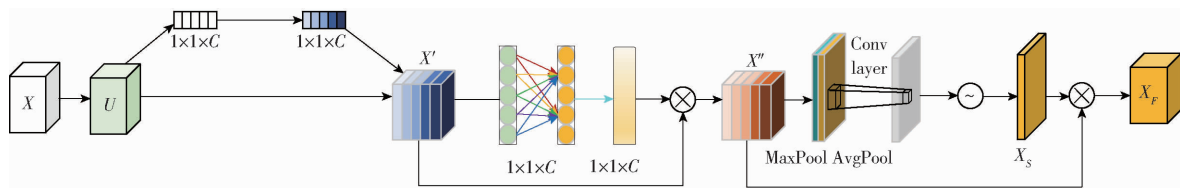


图 7 FRM 模块结构图

Fig. 7 FRM module structure diagram

过卷积变换进行特征映射,得到特征图 $U \in \mathbf{R}^{C \times H \times W}$,进行全局平均池化后,利用 2 个全连接层构建通道注意力机制,其中第 1 个全连接层用于降维以减少计算量并增加非线性,第 2 个全连接层则将通道数恢复,最后通过 Sigmoid 激活函数生成通道权重并与原始特征图相乘,实现通道维度的特征增强。此过程数学表达式为

$$U_c = v_c * X = \sum_{s=1}^{C'} v_c^s * X^s \quad (1)$$

$$Z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W U_{c(i,j)} \quad (2)$$

$$S_c = \sigma(W_2 \delta(W_1 Z_c)) \quad (3)$$

$$X'_{c(i,j)} = U_{c(i,j)} S_c \quad (4)$$

式中 v_c ——第 c 个滤波器的参数

v_c^s ——二维空间核

C' ——输入特征图的总通道数

$*$ ——卷积操作

Z_c ——通道 C 的全局特征向量

(i, j) ——特征图上的空间位置坐标
 S_c ——根据全局特征计算出的通道权重
 H, W ——特征图的高度、宽度
 σ ——Sigmoid 激活函数
 δ ——ReLU 函数
 W_1, W_2 ——全连接的权重矩阵
 $X'_{c(i, j)}$ ——全局特征提取后的特征图

(2)局部依赖建模。对于 X' , 同样经过全局平均池化后, 通过基于通道数的自适应卷积核大小计算方法, 使用一维卷积操作捕获局部通道间的关系, 利用 Sigmoid 函数生成权重并作用于原始特征图。此过程数学表达式为

$$k = \left\lfloor \frac{\text{lb}(c) + b}{\gamma} \right\rfloor_{\text{odd}} \quad (5)$$

$$y_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W X'_{c(i, j)} \quad (6)$$

$$\alpha_c = \sigma(\text{Conv1D}(y_c, k)) \quad (7)$$

$$X''_{c(i, j)} = X'_{c(i, j)} \alpha_c \quad (8)$$

式中 k ——一维卷积核大小

b ——偏置项

γ ——控制因子

y_c ——第 c 个通道的平均池化结果

Conv1D——一维卷积操作

α_c ——第 c 个通道的权重

$X''_{c(i, j)}$ ——局部依赖建模后的特征图

(3)空间特征增强机制。对于 X'' , 空间特征增强模块通过空间维度的自适应增强来处理通道加权特征。模块采用全局最大池化与全局平均池化操作, 生成 2 个维度为 $H \times W \times 1$ 的空间特征图, 将特征图在通道维度上进行拼接, 并通过一个 7×7 卷积层进行特征融合, 最后经过 Sigmoid 函数归一化, 得到空间注意力权重图 X_s 。该权重图通过逐元素相乘的方式作用于输入特征图, 从而生成最终的增强特征图 X_F 。

2.2 目标匹配

为了实现对奶牛的稳定跟踪, 本文对 ByteTrack 跟踪算法中数据关联部分的相似度融合方法做了改进。在相似度计算中引入了外观特征匹配, 并将其与传统的 IoU 匹配相结合。修改了目标检测模型的检测头, 使其不仅输出检测框, 同时在回归检测框之前输出外观特征张量, 在关联时使用该特征张量进行匹配, 提升了匹配精度。

2.2.1 SimilarityFuser 模块

与传统的静态融合策略不同, 本文设计了一个名为 SimilarityFuser 的模块, 将交并比 (IoU) 匹配与外观匹配结合起来, 该模块通过使用可训练的偏移

参数 O_x , 根据每帧中的目标变化情况, 自动调整这两者的权重。在奶牛的多目标跟踪任务中, 个体的运动模式差异较小, 特别是在目标重叠的牛场环境中, 单一的 IoU 匹配无法精准跟踪, 而奶牛外观特征明显, 所以本文选择将外观特征引入, 以优化匹配算法的性能, 并选用成本矩阵衡量匹配效果。数据关联部分采用匈牙利算法, 成本矩阵 C 作为其输入, 是一个 $M \times N$ 的矩阵, 其中 M 是当前帧中的检测框数量, N 是已有轨迹的数量, 矩阵中的每一个元素 C_{ij} 表示第 i 个检测框与第 j 个轨迹之间的匹配成本, 计算公式为

$$C_{ij} = 1 - \text{IoU}_{ij} \quad (9)$$

在此模块中, 基于每帧目标成本矩阵的均值和标准差动态调整权重比例, 从而适应不同环境下的跟踪需求。通过 SimilarityFuser 模块, 系统能够动态调整 IoU 和外观特征之间的权重, 使其在不同的跟踪条件下保持较好的适应性。成本矩阵的均值 C_{mean} 反映了目标之间的整体匹配情况, 匹配度越高, 成本越低; 匹配度越低, 成本越高。通过降低匹配成本, 提高非匹配成本来减少误匹配, 而标准差 C_{std} 则表示匹配差异的离散程度。通过调整 O_x , 可以进一步优化不同条件下的匹配效果。其数学表达式为

$$W_{\text{iou}} = \frac{1}{1 + \exp\left(-\frac{C_{\text{mean}} - O_x}{C_{\text{std}}}\right)} \quad (10)$$

$$W_{\text{reid}} = 1 - W_{\text{iou}} \quad (11)$$

式中 O_x ——偏移参数

W_{iou} ——IoU 匹配所占权重

W_{reid} ——外观特征匹配所占权重

W_{iou} 的计算公式基于成本矩阵的均值和标准差, 通过 Sigmoid 函数平滑地调节 IoU 相似度在融合相似度中的影响。当 C_{mean} 较高时, 意味着目标之间的差异较大, 此时 IoU 相似度的权重会降低, 而更多地依赖外观相似度; 反之, 当 C_{mean} 较低时, 系统会更倾向于依赖 IoU 信息, 以避免因外观相似导致的匹配混淆。融合相似度矩阵 F_{sim} 的计算公式为

$$F_{\text{sim}(ij)} = W_{\text{reid}} R_{\text{sim}(ij)} + W_{\text{iou}} I_{\text{sim}(ij)} \quad (12)$$

式中 $R_{\text{sim}(ij)}$ ——外观特征张量的余弦相似度

$I_{\text{sim}(ij)}$ ——基于交并比计算的相似度

相应的融合成本矩阵 C_{fuse} 为

$$C_{\text{fuse}} = 1 - F_{\text{sim}} \quad (13)$$

在匹配算法中, 网络的前向传播通过计算外观相似度和 IoU 相似度的加权融合来生成最终的融合

成本矩阵。训练过程中,将融合成本矩阵的均值作为损失函数 L 进行最小化,以优化参数 $\mathbf{O}_x$ 。其数学表达式为

$$L = \text{mean}(\mathbf{C}_{\text{fuse}}) \tag{14}$$

式中 $\text{mean}()$ ——计算融合成本矩阵均值
这种优化方法的核心思想是通过降低所有可能匹配对的平均不匹配程度,来提升整体的跟踪性能。

本文在 TBD 框架下引入可训练的损失函数,其通过偏移参数 $\mathbf{O}_x$ 优化匹配策略,该损失函数独立于检测器训练过程,仅作用于数据关联部分,约束融合成本矩阵的分布,提升匈牙利算法分配最优性。

2.2.2 奶牛跟踪流程

经过目标检测阶段得到检测结果后,进行数据关联,匹配算法流程如图 8 所示,具体流程如下:

- (1)根据置信度阈值,将检测结果分为高置信度检测框和低置信度检测框,并放入相应集合。
- (2)利用匈牙利算法将已有轨迹与高置信度框进行匹配,匹配成功后更新轨迹,并放入当前的轨迹集合中,未能够匹配的高置信度检测框判断是否高于跟踪阈值,若高于跟踪阈值,则生成一条新的轨迹,否则不做处理。
- (3)剩余未匹配的跟踪轨迹与低置信度检测框进行匹配,未能匹配的低置信度检测框删除,匹配成功则更新轨迹,未能匹配的轨迹则判断是否满足删除条件,若满足,则删除。
- (4)返回当前帧的所有轨迹集合,将其作为下一帧图像的已有轨迹集合,之后再次重复上述步骤,直至最后一帧。

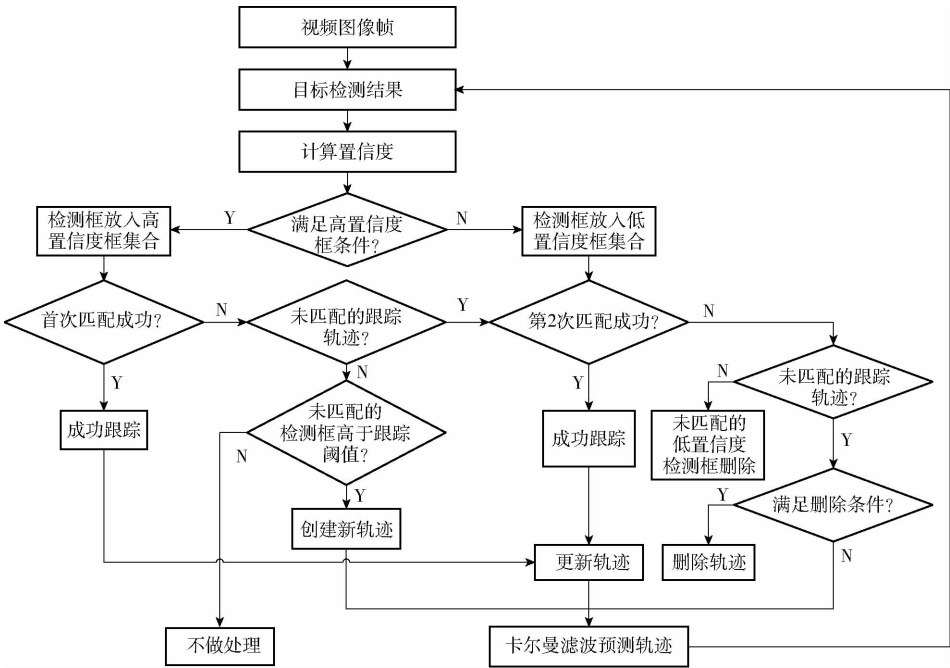


图 8 YOLO - BT 匹配算法流程图
Fig. 8 Flow chart of YOLOX - BT matching algorithm

3 实验与结果分析

3.1 实验设置

实验采用基于 ByteTrack 的深度学习多目标跟踪模型,在 Linux 系统下,编程语言为 Python 3.8,基于 PyTorch 1.10.0 + CUDA 11.4 实现。中央处理器为 Intel(R) Xeon(R) Gold 5220,使用的 GPU 显卡为 NVIDIA Quadro RTX5000,运行内存为 16 GB,批处理量设置为 8。为了更好地拟合数据,采用余弦退火策略来优化模型训练,初始学习率设为 0.01,学习率在训练的每个迭代步骤上根据余弦函数平滑调整,逐步减小到最小值,这种调整方式有助于在训练初期快速收敛,训练后期参数更新更加稳定。

3.2 评价指标

(1)目标检测评价指标
本文选用在 IoU 阈值为 0.5 时的平均精度 (mAP50)、精确率 (Precision, P)、召回率 (Recall, R)、参数量 (Parameters)、浮点运算量 (Floating point operations, FLOPs) 和帧速率 (Frames per second, FPS) 作为目标检测性能的评价指标。mAP50 用于评估模型的整体检测性能,较高的 mAP50 表明模型在目标定位上具有良好的综合表现,较低则说明模型的检测结果与真实标注存在较大偏差。精确率反映了模型预测结果的准确性。较高的精确率表明模型的误检率低,预测框大多数都准确对应真实目标;较低的精确率则意味着模型产生了大量虚假检测,即将背景误判为目标的情况较多。召回率衡量模型

检出目标的完整性。较高的召回率表明模型能够成功检测到大部分真实目标,漏检率低;较低的召回率则说明模型存在显著的漏检问题,许多真实目标未被识别。这 3 种指标数值越高,表明模型的检测性能越好。参数量表示模型中可训练参数的总数量,浮点运算量衡量模型总计算量,这 2 种指标越小代表模型的计算负担越小。帧速率表示模型每秒可处理的图像帧数,数值越大,代表推理速度越快。

(2)多目标跟踪评价指标

选取识别平均数比率(Identification F1 ,IDF1)、召回率、精确率、大多数跟踪个数(Mostly tracked, MT)、主要丢失目标个数(Mostly lost, ML)、身份切换次数(ID switch, IDs)、多目标跟踪准确度(Multiple object tracking accuracy, MOTA)和多目标

跟踪精确度(Multiple object tracking precision, MOTP)作为评价指标衡量模型性能,其中,IDF1、IDs和 MOTA 主要反映跟踪器的性能,IDF1 和 MOTA 数值越高,性能越好,IDs 值越小,性能越好;辅助以 R、P、MT、ML、MOTP 等指标进行跟踪模型的性能评估,R、P、MT、MOTP 值越高,性能越好,ML 值越小,性能越好。

3.3 结果分析

3.3.1 不同检测模型对比

为了选出在 Bytetrack 算法中性能优异的检测器,在自建数据集上设计了检测器对比实验,分别为 YOLO v7、YOLO v8 - s、YOLO v9 - s、YOLO v10、YOLO 11 - s、YOLO 12 检测器,比较结果如表 2 所示。

表 2 不同检测模型的实验对比结果

Tab.2 Experimental comparison results of different detection models

模型	mAP50/%	精确率/%	召回率/%	参数量	浮点运算量	帧速率/(f·s ⁻¹)
YOLO v7	92.7	92.2	89.6	1.02 × 10 ⁷	2.79 × 10 ¹⁰	16.2
YOLO v8 - s	93.8	92.7	88.9	1.11 × 10 ⁷	2.84 × 10 ¹⁰	19.1
YOLO v9 - s	93.5	92.1	88.8	7.77 × 10 ⁶	2.67 × 10 ¹⁰	17.5
YOLO v10	90.1	90.2	87.6	8.04 × 10 ⁶	2.44 × 10 ¹⁰	13.4
YOLO 11 - s	92.6	91.1	89.2	9.41 × 10 ⁶	2.13 × 10 ¹⁰	17.9
YOLO 12	89.6	89.6	85.4	9.23 × 10 ⁶	2.12 × 10 ¹⁰	15.8
YOLOX - s	93.7	93.4	90.4	7.96 × 10 ⁶	2.33 × 10 ¹⁰	17.7
改进 YOLOX - s	94.9	94.2	92.7	7.22 × 10 ⁶	1.94 × 10 ¹⁰	22.9

由表 2 可知,本文研究采用的基线模型 YOLOX - s 的各项指标均优于 YOLO v7、YOLO v10,与 YOLO v8 - s 相比,YOLOX - s 的计算量和参数量均低于 YOLO v8 - s,且在 P、R 指标方面优于 YOLO v8 - s。在帧速率相差不大的情况下,YOLO v9 - s 模型虽然在参数量方面优于 YOLOX - s,但其他指标,YOLOX - s 的表现更为优越。YOLO 11 - s 以及 YOLO 12 模型虽然在计算量方面优于 YOLOX - s,但 YOLOX - s 模型 mAP50、P、R 和参数量更优,所以本文选用 YOLOX - s 模型作为检测器基线模型进行改进。改进后的 YOLOX - s 模型与原模型相比,mAP50、P、R 分别提高了 1.2、0.8、2.3 个百分点,每秒帧数由 17.7 f/s 上升到了 22.9 f/s,参数量下降了 9.3%,运算量减少了 16.7%,表明改进后的 YOLOX - s 检测

器在保持较高精度的同时,具备更高的计算速度和更低的模型复杂度。

为检验 ByteTrack 多目标跟踪模型性能,将 4 类不同环境场景下的奶牛视频作为测试视频,将轨迹分数阈值设为 0.75,匹配阈值为 0.95,ByteTrack 在测试集上的跟踪结果如表 3 所示,其中真实奶牛数量(Ground truth,GT)代表每个视频段中共有几头奶牛。该方法的 MOTA 为 88.1%,IDF1 为 89.2%,说明该算法在牛舍场景中,对于奶牛的多目标跟踪具有良好的效果,IDs 为 29,说明模型性能较为稳定,基本能够实现持续跟踪的效果。结合 R、P 等指标结果,认为该跟踪方法适用于牛舍中奶牛多目标跟踪任务,所以本文选用 ByteTrack 算法作为基础算法,并加以改进。

表 3 基于 ByteTrack 模型的奶牛多目标跟踪结果

Tab.3 Cow MOT results based on ByteTrack

视频类别	IDF1/%	召回率/%	精确率/%	GT	MT	ML	IDs	MOTA/%	MOTP/%
A	95.7	91.8	97.9	23	20	2	7	91.8	2.6
B	68.1	82.5	80.5	12	7	0	9	72.1	9.2
C	96.1	99.5	97.7	18	18	0	6	98.1	6.3
D	90.6	84.4	95.7	9	8	1	7	82.4	3.2
整体	89.2	90.7	95.4	62	53	3	29	88.1	5.0

3.3.2 多目标跟踪实验结果

基于 YOLOX - BT 的多目标跟踪测试结果如表 4 所示,4 种类别的 IDF1 值全部在 90% 以上,说明在这 4 种类别当中对目标的跟踪表现稳定,其中类别 A 和类别 C 的 IDF1 值达到了 96% ~ 99%,说明保持目标跟踪效果良好。并且所有测试视频的整体 IDF1 值达到了 95.2%,表明整体上模型算法的跟踪性能较优。所有测试视频召回率整体数值都在

85% 及以上,所有测试视频的识别个体的精确率达到了 95% 以上,且相较于原模型提高了 2.4 个百分点,MOTA 和 IDF1 值分别提升了 5.1、6.0 个百分点,ID 切换次数减少了 16 次,MOTP 提升了 0.3 个百分点。说明模型能够更加准确地识别出奶牛目标。相比于未改进算法 MT 和 ML 均有提升,表明有更多的轨迹在大多数时间都能被正确跟踪,跟踪持续性相较于未改进算法有所提升。

表 4 基于 YOLOX - BT 模型的奶牛多目标跟踪结果
Tab.4 Cow MOT results based on YOLOX - BT

视频类别	IDF1/%	召回率/%	精确率/%	GT	MT	ML	IDs	MOTA/%	MOTP/%
A	96.3	92.9	98.9	23	21	1	2	92.8	3.0
B	90.1	93.2	95.6	12	11	0	6	92.7	8.4
C	98.9	98.0	97.9	18	18	0	1	97.9	7.0
D	92.6	86.3	96.9	9	8	1	5	86.3	3.8
整体	95.2	92.1	97.8	62	58	2	13	93.2	5.3

所有测试视频的 IDs 总数只有 13,表明 YOLOX - BT 跟踪模型在密集和遮挡复杂环境下,能对目标进行持续跟踪,跟踪性能稳定。同时,4 种测试视频的整体 MOTA 均在 85% 以上,整体数值为 93.2%,表明模型对检测目标的准确性和保持轨迹稳定性较优。整体的 MOTP 相较于之前也有提升,表明位置误差有所减小。综上所述,YOLOX - BT 多目标跟踪模型无论是在识别目标的精确度,还是在跟踪轨迹方面,都取得了更好的效果。

YOLOX - BT 模型的奶牛多目标跟踪部分效果如图 9 所示,其中,类别 A 中 005 号视频在夜晚密集状态下,在第 40、140、240 帧都保持稳定的跟踪。类别 B 中 007 号视频在傍晚条件下,在第 40、140、240 帧都保持稳定的跟踪,最大 ID 号为 12,与表中类别 B 视频的真实牛只数量相同。类别 C 中 018 号视频在白天条件下,在第 40、140、240 帧都保持稳定的跟踪,最大 ID 号为 18,同样与表中 GT 值对应,说明改进算法在奶牛多目标跟踪中取得了良好的效果。

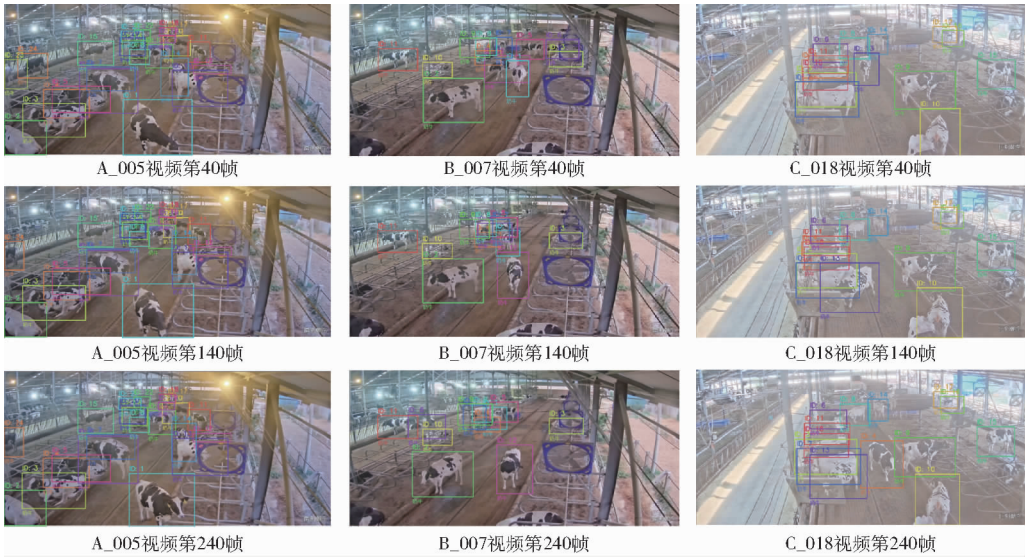


图 9 YOLOX - BT 算法部分跟踪结果

Fig.9 Partial tracking results of the YOLOX - BT algorithm

3.3.3 不同跟踪模型对比

为了验证本文所提出的 YOLOX - BT 模型在奶牛多目标跟踪任务上具有优势,设计了以下对比实验,在相同条件下,将本文模型与 DeepSORT^[22]、GTR (Global Tracking Transformer)^[23]、

Transtrack^[24]、 FairMOT^[25]、 CenterTrack^[26]、 QDTrack^[27]、MotionTrack^[28]、BoT - SORT^[29] 以及原始 ByteTrack 模型进行对比,对比结果如表 5 所示。

由表 5 可知,在基线模型当中,从 4 个指标上综合来看,原始 ByteTrack 模型性能更优,所以选取

表 5 不同跟踪模型的实验对比结果

Tab.5 Experimental comparison results of different tracking models				
多目标跟踪模型	IDF1/%	IDs	MOTA/%	MOTP/%
DeepSORT	75.4	44	86.2	5.1
GTR	73.7	42	85.7	5.1
Transtrack	72.2	61	83.5	4.6
FairMOT	75.1	53	84.4	4.6
CenterTrack	76.4	66	85.6	5.3
QDTrack	73.5	54	82.3	4.4
MotionTrack	77.1	76	77.9	3.9
BoT - SORT	88.4	28	87.6	5.0
ByteTrack	89.2	29	88.1	5.0
YOLOX - BT	95.2	13	93.2	5.3

ByteTrack 作为本文的跟踪器。

在对比实验中,DeepSORT、GTR、Transtrack、FairMOT、CenterTrack、QDTrack、MotionTrack、BoT - SORT 以及原始 ByteTrack 模型的 IDF1 值均低于本

文提出的 YOLOX - BT 模型(95.2%),表明本文算法对于目标能够实现更长时间的跟踪。本文模型的 IDs 仅为 13,在跟踪过程中,本文模型 ID 切换次数更少,性能更加稳定。本文模型 MOTA 值为 93.2%,说明本文模型在奶牛多目标跟踪任务中,检测精度和维持轨迹的能力均高于其他模型。本文模型 MOTP 值依然最优,表明本文模型的定位能力强于其他模型。

综上所述,本文提出的 YOLOX - BT 模型的所有性能指标都高于其他算法模型,表明其拥有更精准的检测功能和更强大的跟踪能力。

3.3.4 消融实验

为明确在 YOLOX - s 目标检测模型中引入 FRM 结构并将 CSP2_1 替换为 CFP - EVCBlock 结构对模型整体性能的影响,验证各改进点的贡献,在上述相同实验条件下,设计了以 YOLOX - s 为基础网络的消融实验,结果如表 6 所示。

表 6 YOLOX - s 模型消融实验结果对比
Tab.6 Comparison of ablation results of YOLOX - s model

YOLOX - s	FRM	CFP - EVCBlock	mAP50/%	精确率/%	召回率/%	参数量	浮点运算量	帧速率/(f·s ⁻¹)
√			93.7	93.4	90.4	7.96 × 10 ⁶	2.33 × 10 ¹⁰	17.7
√	√		94.1	93.9	91.5	8.11 × 10 ⁶	2.44 × 10 ¹⁰	16.1
√		√	93.5	91.1	90.5	7.05 × 10 ⁶	1.86 × 10 ¹⁰	19.2
√	√	√	94.9	94.2	92.7	7.22 × 10 ⁶	1.94 × 10 ¹⁰	22.9

注:“√”表示使用该模块,下同。

根据表 6 结果可知,与原始 YOLOX - s 模型相比,加入 FRM 后,mAP50、P、R 分别提升 0.4、0.5、1.1 个百分点,但参数量和运算量稍有增加,运算速率稍有下降;将 CSP2_1 替换为 CFP - EVCBlock 结构后,对比原模型,参数量和运算量分别下降 11%、20%,帧速率也稍有提升;将两部分组合起来之后,相较于原模型,mAP50、P、R 分别提升 1.2、0.8、2.3 个百分点,参数量和运算量分别减少 9.3%、

16.7%,运算速率也有所提升,更好地平衡了检测精度与计算复杂度之间的关系。

为说明 SimilarityFuser 模块中外观特征匹配与 IoU 信息匹配融合策略的作用,设计了如下消融实验,在原始 ByteTrack 中分别加入外观特征以及融合策略进行匹配,实验结果如表 7 所示,由表 7 可知,相比于单纯使用外观特征和 IoU 信息进行匹配,本文所提出的融合策略在奶牛多目标跟踪任务中效果更佳。

表 7 SimilarityFuser 模块融合策略消融实验结果对比
Tab.7 Comparison of ablation experimental results of SimilarityFuser module fusion strategy

ByteTrack	外观特征	SimilarityFuser 融合策略	IDF1/%	召回率/%	精确率/%	IDs	MOTA/%
√			89.2	90.7	95.4	29	88.1
√	√		91.2	90.9	94.7	25	88.8
√		√	94.5	91.1	95.3	16	90.9

同时为验证在 YOLOX - BT 多目标跟踪算法中各部分改进的独立贡献以及组合效果,以 ByteTrack 算法为基准进行消融实验,在上述相同实验条件下,参照控制变量法的思想,保证学习率、优化器等参数一致,通过消融实验验证本文模型的效果,结果如

表 8 所示。

相较于原模型,对比引入 FRM 前后,检测精度提升 0.5 个百分点,ID 切换次数减少了 16,MOTA 提升 2.6 个百分点,这是由于在处理数据时,为输入序列分配了不同的权重,更加关注到了关键信息。

表 8 YOLOX – BT 模型消融实验结果对比

Tab. 8 Comparison of ablation results of YOLOX – BT model

ByteTrack	FRM	CFP	SimilarityFuser	IDF1/%	召回率/%	精确率/%	IDs	MOTA/%
✓				89. 2	90. 7	95. 4	29	88. 1
✓	✓			92. 8	91. 0	95. 9	13	90. 7
✓		✓		92. 3	93. 9	95. 1	11	91. 0
✓			✓	94. 5	91. 1	95. 3	16	90. 9
✓	✓	✓		93. 6	92. 4	96. 7	13	92. 0
✓	✓		✓	92. 9	93. 3	96. 2	13	92. 5
✓		✓	✓	93. 3	93. 0	95. 2	14	92. 8
✓	✓	✓	✓	95. 2	92. 1	97. 8	13	93. 2

在引入 CFP 模块前后,对比原模型,召回率提升 3. 2 个百分点,这是由于 CFP 模块强调中小目标的检测,尤其是在此基础上引入 FRM 模块之后,检测精度较原模型相比提升 1. 3 个百分点,这表明模型在检测方面更加优秀,对比 MOTA 和 IDs 可以看出,这种改进方式对跟踪性能的提升也效果明显。在原模型中引入 SimilarityFuser 模块之后, MOTA 提升 2. 8 个百分点,ID 切换次数减少 13,表明将外观匹配特征利用起来后,进行数据关联时,匹配得更加精准,保持轨迹的能力有所加强,这是由于,奶牛是一种外观特征强烈的目标生物,外观特征起到了重要作用。

最后,将 3 种改进融合,与原模型对比,所有指标均有正向提升,表明 YOLOX – BT 算法模型更加适用于奶牛多目标跟踪任务。

4 结论

(1)为实现养殖场的奶牛多目标跟踪,本文提出了一种 YOLOX – BT 多目标跟踪模型,在目标检测和跟踪阶段,均作出改进。在目标检测时,将原模型中 Neck 部分的 CSP2_1 结构替换成了 CFP –

EVCBlock,旨在提升对中小目标的检测性能,并引入 FRM 模块;在多目标跟踪阶段,为了解决轨迹断裂问题,提出了一种改进的匹配算法,通过计算成本矩阵的均值和标准差来动态调整外观相似度和 IoU 相似度的融合比例,以适应不同环境下的跟踪需求。

(2)实验结果表明,改进后的检测器在平均检测精度、精确率、召回率分别达到了 94. 9%、94. 2%、92. 7%,较原模型相比分别提高了 1. 2、0. 8、2. 3 个百分点,帧速率为 22. 9 f/s,较原模型增加 29. 4%,参数量下降 9. 3%,浮点运算量减少 16. 7%。本文提出的 YOLOX – BT 模型 MOTA 为 93. 2%,IDF1 为 95. 2%,IDs 为 13,MOTP 为 5. 3%,对比改进前的 ByteTrack 算法模型,MOTA 和 IDF1 值分别提升了 5. 1、6. 0 个百分点,ID 切换次数减少了 16 次,MOTP 提升 0. 3 个百分点。

(3)本文提出的 YOLOX – BT 奶牛多目标跟踪算法能够有效应对复杂环境下的遮挡和小目标检测问题,在稳定性、准确性方面相较原模型表现出色,为奶牛智能管理领域的研究和实际应用提供了技术支持,具有良好的应用前景。

参 考 文 献

[1] 李琦,尚绛岚,李宝山. 基于 YOLOv3 和 Deep SORT 的草原牛跟踪系统[J]. 传感器与微系统,2021, 40(6):83 – 85,88. LI Qi, SHANG Jianglan, LI Baoshan. Grassland cattle tracking system based on YOLOv3 and Deep SORT[J]. Transducer and Microsystem Technologies, 2021, 40(6):83 – 85, 88. (in Chinese)

[2] 李恩泽,王克俭,司永胜,等. 基于改进 ConvNeXt 的奶牛行为识别方法[J]. 农业机械学报,2024,55(5):282 – 289,404. LI Enze, WANG Kejian, SI Yongsheng, et al. Cow behavior recognition method based on improved ConvNeXt[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(5):282 – 289,404. (in Chinese)

[3] 涂淑琴,汤寅杰,李承桀,等. 基于改进 ByteTrack 算法的群养生猪行为识别与跟踪技术[J]. 农业机械学报,2022, 53(12):264 – 272. TU Shuqin, TANG Yinjie, LI Chengjie, et al. Behavior recognition and tracking of group-housed pigs based on improved ByteTrack algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12):264 – 272. (in Chinese)

[4] PANG Z, LI Z, WANG N. Simpletrack: understanding and rethinking 3D multi-object tracking[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 680 – 696.

[5] PENG Y, KONDO N, FUJIURA T, et al. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units[J]. Computers and Electronics in Agriculture, 2019, 157: 247 – 253.

[6] 许贝贝,王文生,郭雷风,等. 基于非接触式的牛只身份识别研究进展与展望[J]. 中国农业科技导报, 2020,22(7):79 – 89.

- XV Beibei, WANG Wensheng, GUO Leifeng, et al. A review and future prospects on cattle recognition based on non-contact identification[J]. *Journal of Agricultural Science & Technology*, 2020, 22(7): 79–89. (in Chinese)
- [7] REN K, KARLSSON J, LIUSKA M, et al. A sensor-fusion-system for tracking sheep location and behaviour[J]. *International Journal of Distributed Sensor Networks*, 2020, 16(5): 1550147720921776.
- [8] ALQARALLEH B A Y, MOHANTY S N, GUPTA D, et al. Reliable multi-object tracking model using deep learning and energy efficient wireless multimedia sensor networks[J]. *IEEE Access*, 2020, 8: 213426–213436.
- [9] WANG Y, CHEN T, LI B, et al. Automatic identification and analysis of multi-object cattle rumination based on computer vision[J]. *Journal of Animal Science and Technology*, 2023, 65(3): 519.
- [10] ZHENG Z, QIN L. PrunedYOLO – Tracker: an efficient multi-cows basic behavior recognition and tracking technique[J]. *Computers and Electronics in Agriculture*, 2023, 213: 108172.
- [11] 张宏鸣,汪润,董佩杰,等. 基于 DeepSORT 算法的肉牛多目标跟踪方法[J]. *农业机械学报*, 2021, 52(4): 248–256.
- ZHANG Hongming, WANG Run, DONG Peijie, et al. Beef cattle multi-target tracking based on DeepSORT algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(4): 248–256. (in Chinese)
- [12] 涂淑琴,刘晓龙,梁云,等. 基于改进 DeepSORT 的群养生猪行为识别与跟踪方法[J]. *农业机械学报*, 2022, 53(8): 345–352.
- TU Shuqin, LIU Xiaolong, LIANG Yun, et al. Behavior recognition and tracking method of group-housed pigs based on improved DeepSORT algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(8): 345–352. (in Chinese)
- [13] ZHANG Y, SUN P, JIANG Y, et al. Bytetrack: multi-object tracking by associating every detection box [C] // *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 1–21.
- [14] MILAN A. MOT16: a benchmark for multi-object tracking[J]. *arXiv Preprint*, arXiv:1603.00831, 2016.
- [15] DENDORFER P. Mot20: a benchmark for multi object tracking in crowded scenes[J]. *arXiv preprint arXiv:2003.09003*, 2020.
- [16] KIERITZ H, HUBNER W, ARENS M. Joint detection and online multi-object tracking [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018: 1459–1467.
- [17] QUAN Y, ZHANG D, ZHANG L, et al. Centralized feature pyramid for object detection[J]. *IEEE Transactions on Image Processing*, 2023: 32: 4341–4354.
- [18] 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.
- [19] WANG Q, WU B, ZHU P, 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.
- [20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3–19.
- [21] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1251–1258.
- [22] 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.
- [23] ZHOU X, YIN T, KOLTUN V, et al. Global tracking transformers [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 8771–8780.
- [24] SUN P, CAO J, JIANG Y, et al. Transtrack: multiple object tracking with transformer [J]. *arXiv Preprint*, arXiv:2012.15460, 2020.
- [25] 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.
- [26] ZHOU X, KOLTUN V, KRAHENBUHL P. Tracking objects as points [C] // *European Conference on Computer Vision*. Cham: Springer International Publishing, 2020: 474–490.
- [27] FISCHER T, HUANG T E, PANG J, et al. Qdtrack: quasi-dense similarity learning for appearance-only multiple object tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(12): 15380–15393.
- [28] 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.
- [29] AHARON N, ORFAIG R, BOBROVSKY B Z. Bot-sort: robust associations multi-pedestrian tracking [J]. *arXiv Preprint*, arXiv:2206.14651, 2022.