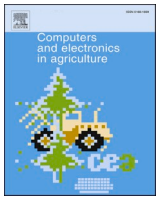
Computers and Electronics in Agriculture 209 (2023) 107857

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows

Zhiyang Zheng, Jingwen Li, Lifeng Qin *

College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China

ARTICLE INFO

Keywords:
Dairy cow
Object detection
Multi-object tracking
Kalman filter

ABSTRACT

Dairy cows tracking is an essential means to obtain their behavioral information, real-time position, activity data, and health status. A multi-object tracking method (YOLO-BYTE) is proposed to address the problem of missed detection and false detection caused by complex environments in cow individual detection and tracking. The method improves upon the YOLO v7 Backbone network feature extraction module by adding a Self-Attention and Convolution mixed module (ACmix) to account for the uneven spatial distribution and target scale variation of the cows. Additionally, in order to reduce the number of model parameters, an improved lightweight Spatial Pyramid Pooling Cross Stage Partial Connections (SPPCSPC-L) module is adopted to reduce model complexity. At the same time, the state parameters in the Kalman filter are improved by directly predicting the width and height information of the tracking boxes, so as to improve the ByteTrack algorithm to make tracking boxes matching the cows more precisely and accurately. Experimental conducted on the dairy cow object detection and multi-object tracking dataset show that the proposed YOLO-BYTE model achieves a Precision (P) of 97.3% in the dairy cow target detection dataset, with an improved Recall (R) and Average Precision (AP) by 1.1% compared to the original algorithm, and an 18% reduction in model parameters. Moreover, the proposed method demonstrated significant improvements in High Order Tracking Accuracy (HOTA), Multi-Object Tracking Accuracy (MOTA), and Identification F1 (IDF1) by 4.4%, 6.1%, and 3.8%, respectively, compared to the original model, with a decrease of 37.5% in Identity Switch (IDS). The tracker runs in a real-time manner with an average analysis speed of 47 fps. Hence, it is demonstrated that the proposed approach is capable of effective multi-object tracking of dairy cows in natural scenes and provides technical support for non-contact dairy cow automatic monitoring.

1. Introduction

Green, efficient, and precise breeding supported by information technology and intelligent technology has become an inevitable trend in the development of modern livestock husbandry (Kumar et al., 2018, Liu et al., 2021, Yang et al., 2022). Dairy cow breeding is a significant area of animal husbandry that has proliferated in recent years, as has the scope of breeding and the complexity of refined management (Jiang et al., 2019, Wu et al., 2020). The acquisition and intelligent processing of cow video information is an important technical way to realize cow target location recognition, behavior analysis, and health evaluation (Liu et al., 2020, Wang et al., 2023b). However, the Dairy cow multi-object tracking algorithm in farms is the basis for applications such as individual cow identification, behavior analysis, and individual counting, and is essential for the intelligence of dairy farming technology (Li et al., 2022b, Wang et al., 2023a, Zheng et al., 2023).

Currently, there are two categories of livestock tracking methods: contact and non-contact (Boopathi Rani et al., 2022, NOE et al., 2022, Noinan et al., 2022). Contact tracking methods are realized by putting tracking equipment on livestock (Zambelis et al., 2019, Williams et al., et al., 2022). Such methods are feasible for livestock observation in some cases, but wearable devices and sensors are costly, easily damaged, and prone to stressful behavior and even health effects when worn.

In contrast, non-contact tracking methods based on machine vision technology can effectively avoid these problems (Koniar et al., 2016, Bergamini et al., 2021, Gao et al., 2022). Sun et al. (2020) proposed a multi-channel color feature adaptive fusion algorithm, which used pig contour information to meet the requirements of tracking accuracy and

* Corresponding author at: College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China.
E-mail address: fuser@nwsuaf.edu.cn (L. Qin).

https://doi.org/10.1016/j.compag.2023.107857
Received 16 February 2023; Received in revised form 9 April 2023; Accepted 12 April 2023
Available online 21 April 2023

Fig. 1. Surveillance video footage from the active field: (a) light interference, (b) rainy weather, (c) night scene, (d) morning picture.

real-time. Xiao et al. (2019) identified pigs by color information and tracks them according to a set of association rules with constraint terms (DA-ACR). The above methods need to design the underlying features, which is not only heavy workload but also has strict requirements for the environment, making it difficult to meet the actual scene requirements. The convolution neural network, which is the mainstream method of livestock tracking (Boogaard et al., 2020, Gao et al., 2021), can automatically extract the low-level features, such as edge and texture, and high-level semantic features. Zhang et al. (2021) used the improved YOLO v3 network combined with DeepSort algorithm to achieve multi-object tracking of beef cattle. Tu et al. (2022) combined YOLO v5 with the improved DeepSort algorithm to achieve pig tracking. The above tracking methods are all based on the Tracking by Detection (TBD) paradigm, that is, the detection results, output by the detector, are then tracked by the back-end tracking optimization algorithm based on the Kalman filter and Hungarian algorithm, such as Sort (Bewley et al., 2016), DeepSort (Wojke et al., 2017). These algorithms remove the low-scored detection boxes in the current frame by a threshold, and then use the Kalman filter to predict the position of each target in the next frame. However, in realistic scenes, complex environments, such as mutually blocking of cows and uneven illumination, will generate plenty of low-scoring detection boxes. Removing all these low-scored boxes reduces the computational effort of subsequent matching, but at the same time brings about real target loss and fragmented trajectories, which affects the dairy cow tracking effect. Tassinari et al. (2021) utilized the YOLOv3 model to recognize individual cows and detect their positions and movements, only 4 cows were tested, and the number of samples was small. Guzhva et al. (2018) proposed a novel method for continuous tracking and data-marker based identification of individual cows based on convolutional neural networks (CNNs). In the aforementioned research, the tracking methods are applied to a small number of indoor cow targets, making the tracking task relatively easy. Moreover, these methods require cow targets labelling, and as the number of cows increase, it will be extremely difficult to manually identify and label specific cows in the images. This study focuses on the tracking of a large number of cows in an outdoor activity field, which further increases the difficulty of cow tracking.

In response to the above problems, this study proposes a multi-object tracking algorithm, YOLO-BYTE, for multiple cows tracking in an activity field using a single camera, which addresses the characteristics of uneven distribution of cow targets and significant variation of target scales, adds the Self-Attention and Convolution mixed module (ACmix) to improve the feature extraction module in the YOLO v7 Backbone network; at the same time, improves the Spatial Pyramid Pooling module (SPPCSPC-L) to reduce model parameters and decrease the model complexity; in the tracking stage, the ByteTrack algorithm (Zhang et al., 2022), which preserves the low score detection boxes, is adopted and improved by directly predict the width and height information of the tracking boxes in Kalman filter algorithm. The remainder of the paper is structured as follows: Section 2 introduces the materials and methods. Experiments and results are reported in Section 3. Section 4 presents the discussion and scope for future works. Section 5 presents the conclusion.

2. Materials and methods

2.1. Materials

2.1.1. Video acquisition

The subjects of the videos were Holstein dairy cows, and the experimental videos were collected from the Animal Husbandry Experimental Base of Northwest Agriculture and Forestry University located in Yangling District, Xianyang City, Shaanxi Province, China. A Hikvision camera was installed at the corner of the cow activity area; the camera was mounted on the supporting wall of the cattle shed about 3 m high, and the camera looked down diagonally to capture the whole cow activity field area.

The experimental videos were captured from June – September

Table 1
Dairy cow target tracking dataset.

No.	Weather	Period	Sparse	Dense	Interference Factors
01	Sunny	Afternoon	√	—	Camera wobbles slightly, dairy cow occlusion (Light)
02	Cloudy	Sunrise	—	√	Light interference, low light, dairy cow occlusion (Heavy)
03	Cloudy	Sunrise	—	√	Light interference, light variation, dairy cow occlusion (Heavy)
04	Cloudy	Evening	—	√	Low light, dairy cow occlusion (Medium)
05	Sunny	Morning	—	√	Sweeper, dairy cow occlusion (Medium)
06	Rainy	Evening	—	√	Light interference, low light, dairy cow occlusion (Heavy)
07	Sunny	Sunrise	—	√	Light interference, low light, dairy cow occlusion (Heavy)
08	Cloudy	Evening	√	—	Low light, leaves swaying, dairy cow occlusion (Light)
09	Cloudy	Morning	—	√	Dairy cow occlusion (Medium)
10	Cloudy	Evening	√	—	Low light, dairy cow occlusion (Light)

2022. After screening, invalid videos, such as nighttime, no target, lens pollution, etc., were excluded, and 375 valid videos were retained All the videos were in MP4 format, with a resolution of 1920 pixels (Horizontal) × 1080 pixels (Vertical), a frame rate of 25f/s, and time length of 45 min−60 min. The video scene is shown in Fig. 1.

2.1.2. Dataset construction

In this study, two data sets were constructed to train the dairy cow target detection model and verify multi-object tracking algorithms, respectively. To ensure the diversity of the dairy cow object detection dataset, this study divided the valid videos into two parts. Part 1 involved annotating the cropped video clips using DarkLabel annotation software, with every 1 out of 20 frames extracted and added to the dataset; Part 2 firstly used FFmpeg to extract key frames and then used LabelImg to mark the minimum circumscribed rectangle of each cow as the real target box, and don't mark when the cow target occlusion exceeds 80%. The two parts together constituted the dairy cow target detection dataset, consisting of 4252 images, which were divided into training set (2551 images), validation set (850 images), and test set (851 images) for model training and testing according to the ratio of 6:2:2.

In order to verify the multi-object tracking algorithm for dairy cows proposed in this study, the surveillance videos in natural group breeding scenario were used as test data. In order to compare the tracking ability of the model in different scenes, we selected 10 videos for testing, which were recorded as serial numbers 01 ~ 10, and the videos were labeled using DarkLabel software. In the meantime, based on manual observation, the videos with more cows and serious sticky obscuration were defined as dense videos and vice versa as sparse videos, and the details are shown in Table 1. The number of cows in the active field in the sunrise (5:00 ~ 7:00) and evening (18:00 ~ 21:00) was higher than that in the morning (8:00 ~ 12:00) and afternoon (13:00 ~ 17:00), resulting in more severe occlusions between cows.

2.2. Dairy cow target detection

2.2.1. Detector

In this paper, we used a multi-object tracking algorithm based on TBD paradigm, and the effect of target detection directly affects tracking effectiveness. The YOLO v7 algorithm of YOLO series was adopted as the target detector in this study (Wang et al., 2022a), which employs strategies such as the Extended Efficient Long-term Attention Network (ELAN), Cascade-based Model Scaling (Wang et al., 2021), Convolutional Structural Re-parameterization (Ding et al., 2021), and Dynamic Label Assignment. Within the range of 5f/s to 160f/s, the YOLO v7

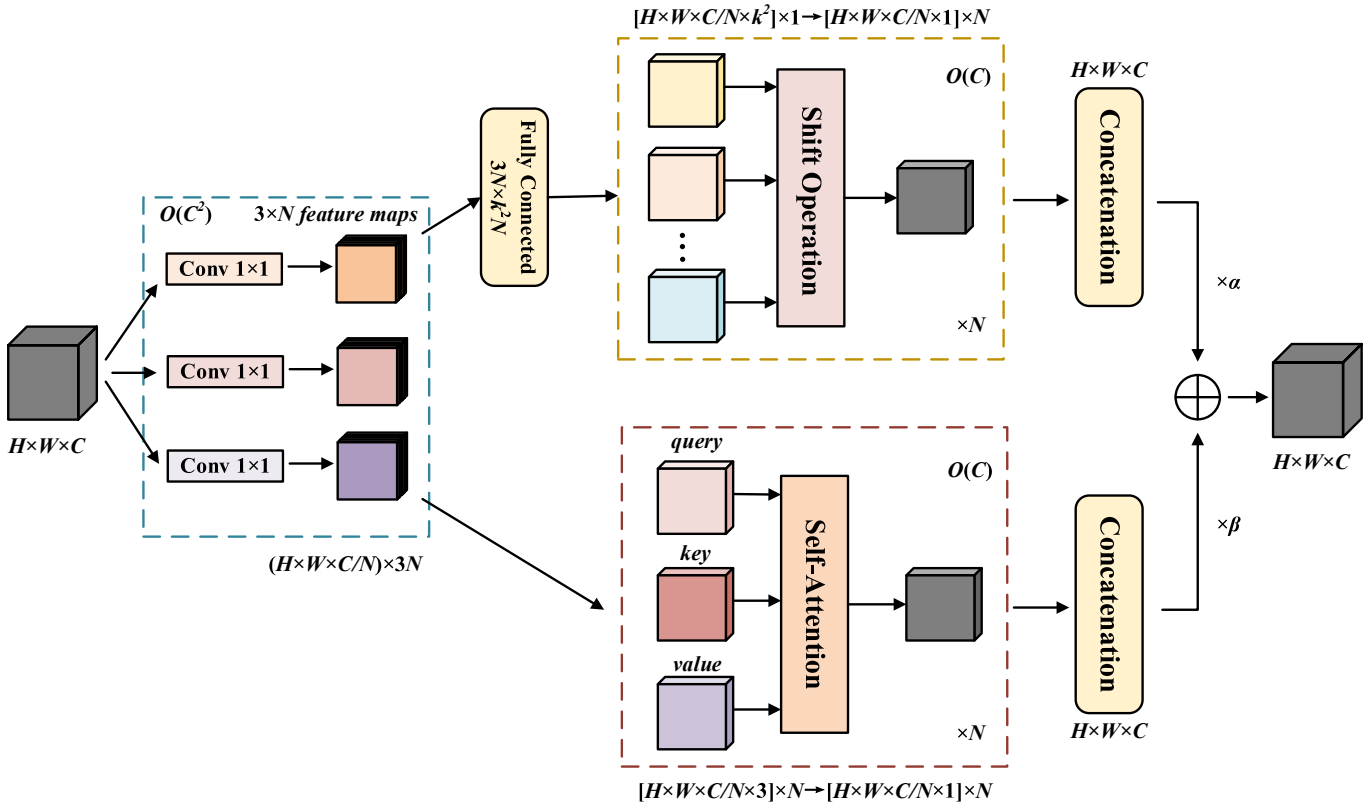


Fig. 2. Flow Chart of the ACmix Module.

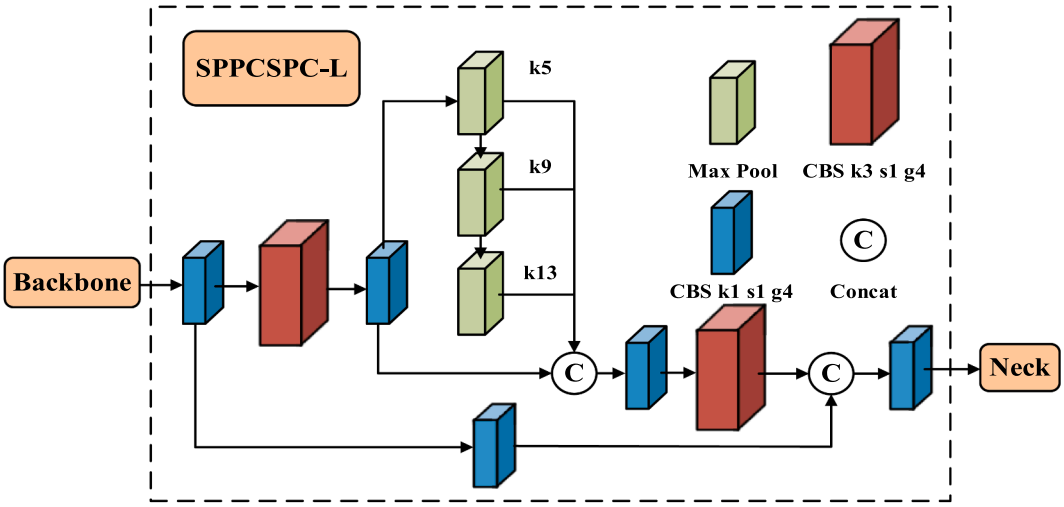


Fig. 3. SPPCSPC-L module.

algorithm outperformed most of the known target detectors in terms of both speed and accuracy, featuring faster detection speed and higher precision.

The model structure of YOLOv7 is mainly divided into three parts: (i) Backbone for extracting image features; (ii) Neck for fusing multi-scale feature information; (iii) Prediction for predicting confidence, category and anchor boxes. The Backbone of YOLO v7 is mainly consists of BConv, ELAN, and MPConv. The Neck part consists of a Path Aggregation Feature Pyramid Network (PAFPN) structure, which efficiently integrates different level features by introducing a bottom-up path to facilitate the transmission of bottom-level information to higher levels. The Prediction part adjusts the image channel numbers of three different scales of features, including P3, P4, and P5, through the REP (RepVGG Block) structure, which is finally used for the prediction of confidence, category, and anchor boxes. Fig. 4 gives a simplified overview of YOLOv7.

2.2.2. *Enhancements in the ELAN Module and Backbone Network*

In a non-restrictive environment in the cow activity area, the distance between the cows and the camera is variable, leading to different scale sizes of the dairy cow in the image. Furthermore, the movement of the dairy cows cause inter-cow occlusion, producing many small and medium-sized cow targets. Multiple dairy cow detection experiments using YOLO v7 Backbone network found that in the feature extraction process, the network can't fully extract the texture and contour information of these small and medium-sized targets in both the middle and shallow layers, leading to missed detection of dairy cow targets.

To address this issue, we integrate the ACmix module (Pan et al., 2022) into the ELAN module of YOLO v7 Backbone in this study. The ACmix module integrates self-attention with convolutional modules by sharing the same complex operations. This not only benefits from the advantages of traditional convolutional weight sharing through aggregation functions in the local receptive field, but also gains the flexibility of the self-attention module, focusing on different regions in the image adaptively and capture more features.

As shown in Fig. 2, ACmix integrates convolution and self-attention operations. In the first phase, it projects the input features through convolution to obtain a set of sub-features containing $3 \times N$ feature maps. In the second phase, it uses two different forms: the upper part processes the input features by convolution, and the output is F_{conv}; the lower part adopts the self-attention approach and aggregates the intermediate features into N groups, each group containing 3 feature maps: query, key, and value, respectively while using a multi-headed self-attention model, and the output is F_{att}. Finally, the results of the two forms are weighted and summed (Eq. (1)), with the weights controlled by two learnable scalars. In this paper, both α and β are 0.5.

$$F_{out} = \alpha F_{conv} + \beta F_{\text{att}} \tag{1}$$

In this paper, ACmix is embedded into the ELAN module by replacing the fourth CBS module, forming the ELANA module. Compared to ELAN, ELANA focuses on the extracting global features, and helps the detection model consciously extract features belonging to cows, thereby ensuring higher detection accuracy. At the same time, to improve the Backbone features extraction, the ELANA is used to replace the 3rd and 4th ELAN modules in the Backbone, enhancing the network's ability to extract features from small and medium-sized cow targets. The improved Backbone network considers both global and local features, thus improving the network's detection performance for cows in complex scenarios.

2.2.3. *Improved Spatial Pyramid Pooling module*

The Spatial Pyramid Pooling module is obtained through pooling operations to obtain different receptive fields, allowing the model to adapt to images with different resolutions. In order to enhance the deployment ability of the model on edge devices, the Spatial Pyramid Pooling module in YOLO v7 network is optimized to reduce network complexity and decrease the number of model parameters while ensuring recognition accuracy.

The SPPCSPC module in YOLO v7 has better performance than the Spatial Pyramid Pooling Fast (SPPF) module in YOLO v5, but with a significant increase in parameters and computation complexity. Therefore, in order to reduce the complexity of the model, by drawing inspiration from the SPPF module, the pooling in the SPPCSPC module is modified to run serially, and the convolution are modified to be group convolutions. This paper adopts 4-group convolution according to multiple pre-experiments. The improved SPPCSPC module is named SPPCSPC-L, as shown in Fig. 3.

In this study, the ELAN model in the P4 and P5 layers of the original Backbone network is replaced with our improved ELANA module, and the Spatial Pyramid Pooling module in the original network is replaced with the proposed SPPCSPC-L module. The overall structure of the improved YOLO v7 model is shown in Fig. 4.

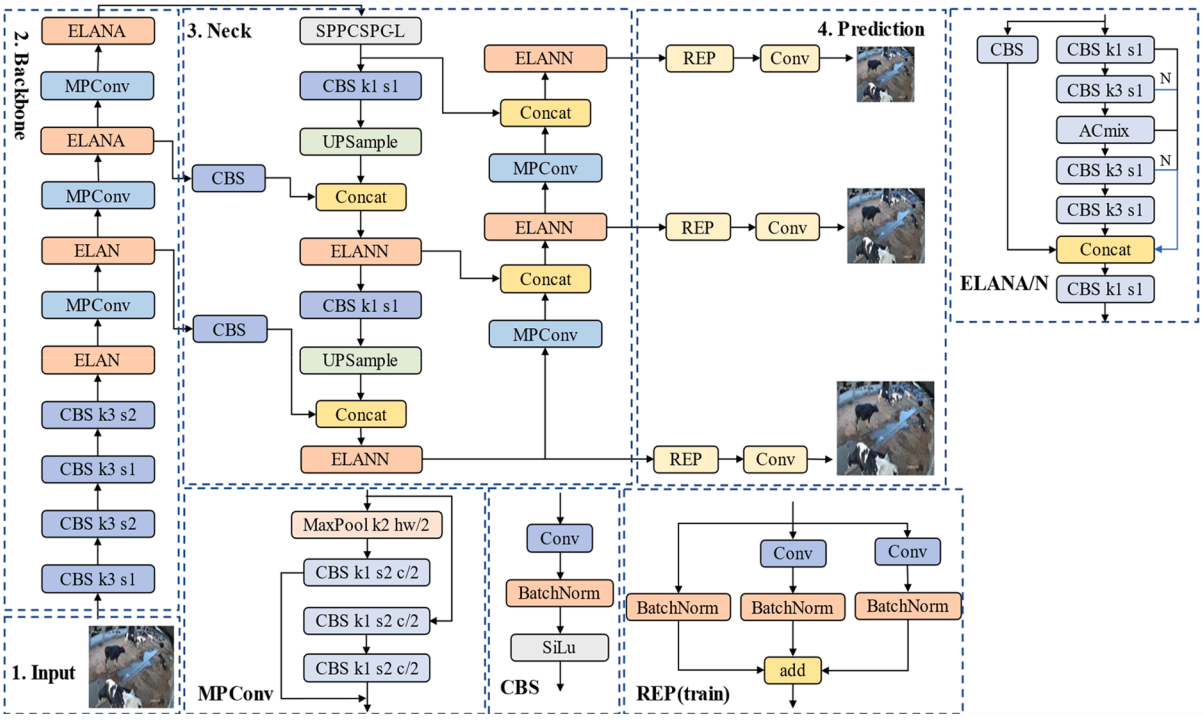


Fig. 4. The improved YOLO v7 network architecture.

2.2.4. Object detection evaluation metrics

In this paper, four evaluation metrics were used for object detection evaluation, including Precision (P), Recall (R), Average Precision (AP), defined as shown in Eqs. (2)~(4), and the number of model Parameters (Params). Intersection over Union (IoU) $\geq$ 0.5 was set as the correct detection of cows during all the tests.

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 PRdR \tag{4}$$

Among them, *TP* (True Positives) refers to the number of positive samples; *FN* (False Negatives) refers to the positive samples predicted by the model to be negative; *FP* (False Positives) represents the negative samples predicted by the model as positive samples.

2.3. Dairy cow multi-object tracking

2.3.1. Dairy cow multi-object tracking algorithm

This study improved the ByteTrack algorithm for implementing multi-object tracking of dairy cows, and the tracking process is shown in Fig. 5. Firstly, the detection results output from the improved YOLO v7 model are divided into high and low-score detection boxes based on two confidence intervals $[\theta_{High}, 1]$ and $[\theta_{Low}, \theta_{High}]$. Then, different matching strategies are used to first match the trajectories T with high-score detection boxes D_{High}, and the unmatched high-score detection boxes generate new tracks T_{New}, and the remaining tracks T_{remain} are matched with low-score detection boxesD_{Low}; finally, the track set T is updated using Kalman filter. The method utilizes the similarity between detection boxes and tracks to remove the background from low-score detection results while preserving high-score detections, thereby uncovering the actual targets (obscured, blurred, and other challenging cow targets), reducing missed detections and improving the coherence of the tracks. In the dairy cow multi-object tracking experiment, the high-score detection boxes threshold θ_{High} is 0.6, and the low-score detection boxes threshold θ_{Low} is 0.3.

2.3.2. Improved tracking algorithm

In this section, we provide a detailed description of the improvements made to the ByteTrack algorithm, and the specific experimental results and ablation experiments are shown in section 3.3.1.

The Kalman filter algorithm is commonly used for modeling object motion in a plane (Tan et al., 2022), and the algorithm employs a constant velocity and linear observation model for predicting and updating the target trajectory (Bewley et al., 2016). In the prediction stage, the Kalman filter in the ByteTrack algorithm uses the estimation from the previous frame to predict the current frame. In this paper, the position of the cow target and its covariance matrix is estimated based on the position of the target detected by the YOLO v7 network, as shown in Eqs. (5)~(6).

$$\widehat{x}_{t|t-1} = F_t \widehat{x}_{t-1|t-1} + B_t u_{t-1} \tag{5}$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q \tag{6}$$

There, $\widehat{x}_{t|t-1}$ is the prior state estimate of frame t and is the result of the prediction equation; F_t and B_t are the state transition matrix and measurement matrix, respectively; $\widehat{x}_{t-1|t-1}$ is the posterior state estimate of frame t-1; u_{t-1} is the noise of frame t-1; $P_{t|t-1}$ is the prior estimate covariance of frame t; $P_{t-1|t-1}$ is the posterior estimate covariance of frame t-1; Q is the covariance of the system noise, representing the reliability of the entire system.

In the update stage, the Kalman filter algorithm utilizes the observation value of the current frame to optimize the predicted value obtained in the prediction stage, thus getting a more accurate estimate. The tracking position and covariance matrix are updated by checking the match between the observed and predicted cows, as shown in Eqs. (7)~(9).

$$K_t = P_{t|t-1} H^T (H P_{t|t-1} H^T + R)^{-1} \tag{7}$$

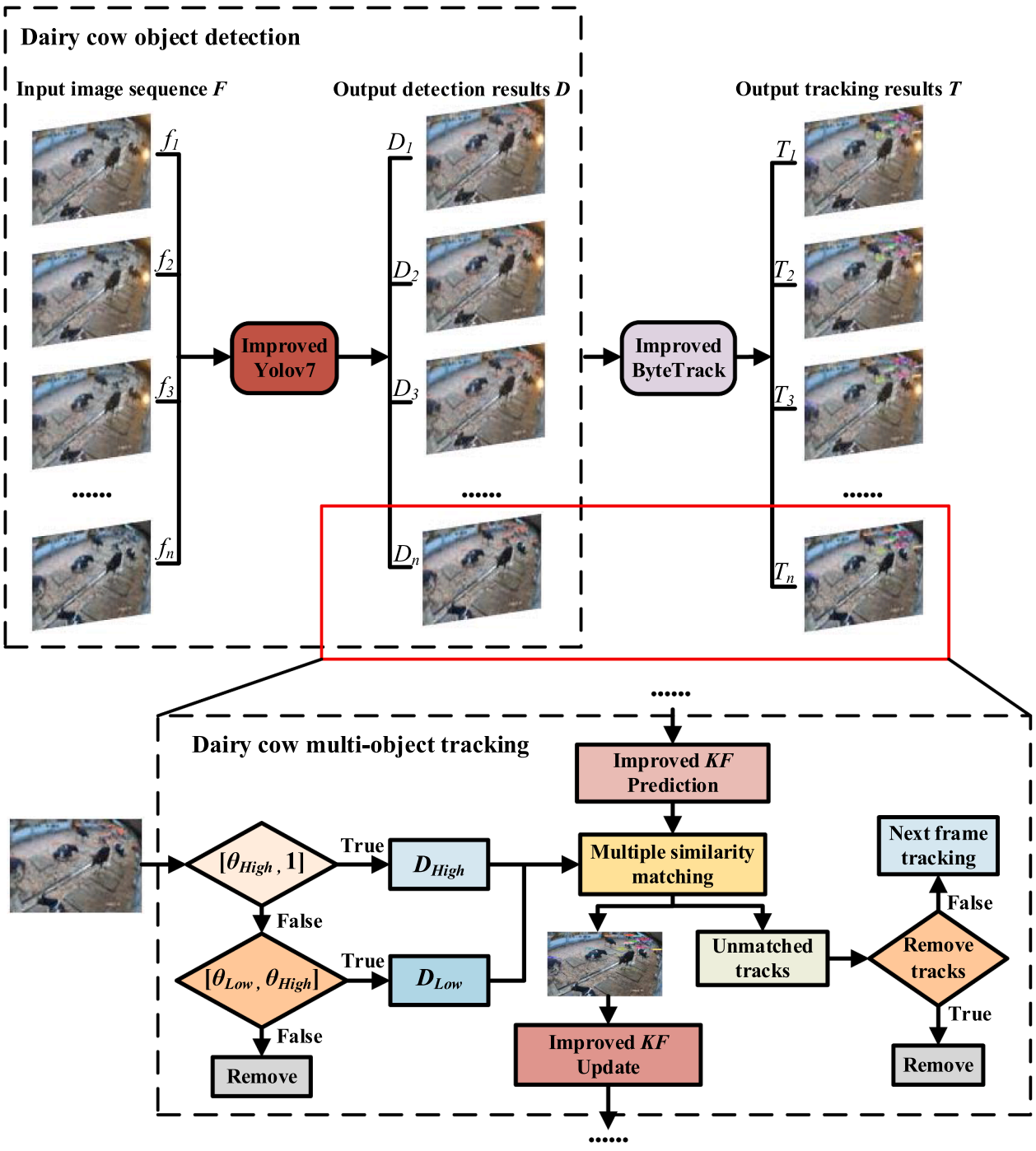


Fig. 5. Flow chart of multi-object tracking process for cows.

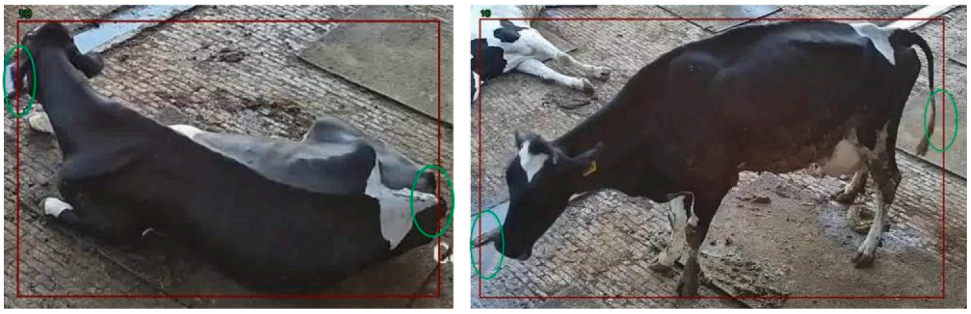
$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(Z_t - H\hat{x}_{t|t-1}) \tag{8}$$

$$P_{t|t} = P_{t|t-1} - K_t H P_{t|t-1} \tag{9}$$

There, K_t is the gain matrix of Kalman filter; H is the transformation matrix from the state variable to the observation variable, establishing the relationship between them; R is the measurement noise matrix; $\hat{x}_{t|t}$ is the posterior state estimate; Z_t is the measurement value; $P_{t|t}$ is the posterior estimate covariance for frame t.

In the ByteTrack algorithm, the Kalman filter state vector is used for predicting and associating the tracking boxes. Since the Kalman filter adopts a constant velocity model assumption, its state estimation can lead to sub-optimal bounding box shapes compared to the object-detector driven detections. When using the ByteTrack algorithm to track multiple target cows, the problem of the tracking box not completely and inadequately fitting the cow may occur, as shown in Fig. 6. The Kalman filter state vector used in the ByteTrack algorithm estimates the aspect ratio of the tracking boxes, leading to an inaccurate estimation of the width of the tracking boxes.

To address this issue, this paper modifies the state vector in the ByteTrack algorithm to be$x = (x_c, y_c, w, h, \dot{x}_c, \dot{y}_c, \dot{w}, \dot{h})$, where (x_c, y_c) represents the center position coordinates of the target, w is the width of the bounding box, h is the height of the bounding box, and $(\dot{x}_c, \dot{y}_c, \dot{w}, \dot{h})$ is the velocity of the corresponding parameters in the image coordinates for (x_c, y_c, w, h). The width and height of the tracking box are directly



(a) (b)

Fig. 6. The tracking problem in the ByteTrack algorithm: (a) The tracking box isn't surrounding the cow. (b) The tracking box encompasses extra background information.

predicted and updated through the Kalman filter. After the modification of the state parameters, the state vector x_t, measurement vector z_t, state covariance matrix Q, and measurement noise matrix R in the Kalman filter algorithm are modified in Eqs. (10)~(13).

$$x_t = [x_c(t), y_c(t), w(t), h(t), \dot{x}_c(t), \dot{y}_c(t), \dot{w}(t), \dot{h}(t)]^T \tag{10}$$

$$z_t = [\dot{x}_c(t), \dot{y}_c(t), \dot{w}(t), \dot{h}(t)]^T \tag{11}$$

$$Q = diag((\delta_p \widehat{w}_{t-1|t-1})^2, (\delta_p \widehat{h}_{t-1|t-1})^2, (\delta_p \widehat{w}_{t-1|t-1})^2, (\delta_p \widehat{h}_{t-1|t-1})^2, (\delta_v \widehat{w}_{t-1|t-1})^2, (\delta_v \widehat{h}_{t-1|t-1})^2, (\delta_v \widehat{w}_{t-1|t-1})^2, (\delta_v \widehat{h}_{t-1|t-1})^2) \tag{12}$$

$$R = diag((\delta_m \widehat{w}_{t|t-1})^2, (\delta_m \widehat{h}_{t|t-1})^2, (\delta_m \widehat{w}_{t|t-1})^2, (\delta_m \widehat{h}_{t|t-1})^2) \tag{13}$$

There, $[x_c(t), y_c(t)]$ represents the center position coordinates of the target, $w(t)$ is the width of the bounding box, $h(t)$ is the height of the bounding box, and $[\dot{x}_c(t), \dot{y}_c(t), \dot{w}(t), \dot{h}(t)]$ is the velocity of the corresponding parameters in the image coordinates for $[x_c(t), y_c(t), w(t), h(t)]$; $\widehat{w}_{t-1|t-1}$ and $\widehat{h}_{t-1|t-1}$ are the posterior state estimates of the width and height of the tracking boxes in frame t-1; $\widehat{w}_{t|t-1}$ and $\widehat{h}_{t|t-1}$ are the prior estimates of the width and height of the tracking boxes in frame t; δ_p, δ_v and δ_m are noise factors. After modifying the state parameters, the ByteTrack algorithm better matches the tracking boxes for dairy cows.

2.3.3. Multi-object tracking evaluation metrics

We select High Order Tracking Accuracy (HOTA) (Luiten et al., 2021), Multi-Object Tracking Accuracy (MOTA), Multi-Object Tracking Precision (MOTP), and Identification F1 (IDF1) as the evaluation metrics for cow multi-object tracking (Wang et al., 2022b). HOTA introduces a higher dimensional tracking accuracy index that can measure the performance of multi-object trackers more comprehensively and evenly. MOTA measures the performance of the tracker in detecting targets and maintaining tracks. MOTP measures the localization accuracy of the detector. IDF1 measures the stability of the tracker, with a higher value indicating that the algorithm can track a target accurately for a more extended period (Li et al., 2022a).

HOTA calculation formula is shown as Eq. (14), where *DetA* represents the detection accuracy score and *AssA* represents the association accuracy score. *TP* refers to the number of positive samples; *FN* refers to the number of the positive samples predicted by the model to be negative; *FP* represents the number of the negative samples predicted by the model as positive and C is a point belonging to *TP*, according to which we can always determine a unique GT trajectory. $A(c)$ represents the association accuracy, defined by Eq. (15), where $TPA(c)$ represents the accuracy of correct association, $FPA(c)$ represents the accuracy of predicted trajectory without association or incorrect association and $FNA(c)$ represents the accuracy of predicted trajectory without association or association error prediction.

$$\text{HOTA} = \sqrt{\text{DetA}\cdot\text{AssA}} = \sqrt{\frac{\sum_{C\in TP} A(c)}{TP + FN + FP}} \tag{14}$$

$$A(c) = \frac{TPA(c)}{TPA(c) + FPA(c) + FNA(c)} \tag{15}$$

MOTA calculation is shown as Eq. (16), where *FP* represents the total number of false detections in frame t; *FN* represents the total number of missed detections in frame t; *IDS* represents the number of times the target label ID switched during tracking in frame t; g_t represents the number of targets observed at time t.

$$MOTA = 1 - \frac{\sum_t (FP + FN + IDS)}{\sum_t g_t} \tag{16}$$

MOTP calculation is shown as Eq. (17), where $d_{i,t}$ represents the distance between the given and its paired hypothetical position in frame t; c_t represents the number of matches between targets and hypothesis positions in frame t, i represents the current detection target.

$$MOTP = \frac{\sum_{t,i} d_{i,t}}{\sum_t c_t} \times 100\% \tag{17}$$

IDF1 calculation is shown as Eq. (18), where *IDTP* represents the total number of targets correctly tracked with unchanged ID; *IDFP* represents the total number of targets incorrectly tracked with unchanged ID; *IDFN* represents the total number of targets lost in tracking with unchanged ID.

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \tag{18}$$

Additionally, the model performance is evaluated with other 2 metrics in this paper, including the number of Identity Switches (IDS) and the average Frames Per Second (FPS). Higher values of MOTA, MOTP, IDF1, and FPS, and a lower value of IDS indicate better model performance.

3. Results and analysis

3.1. Experimental platform and parameter settings

To examine the performance of the proposed YOLO-BYTE algorithm in natural scene activity fields for cow target tracking, two experiments are conducted: (1) dairy cow detection experiment to validate the performance of the improved YOLO v7 detector and (2) dairy cow multi-object tracking experiment to analyze the performance of the improved ByteTrack algorithm. In the dairy cow detection experiment, the training image size is set to 640 pixels × 640 pixels, the number of iterations (Epoch) is set to 150, and the training batch size is set to 8.

Table 2
Comparison of target detection results for different algorithms.

Models	P/%	R/%	AP/%	Params
YOLO v5-l	97.0	94.5	95.0	4.66×10^8
YOLO v7	97.1	94.9	95.1	3.64×10^8
YOLO v7-x	97.4	95.9	96.1	7.07×10^8
Improved YOLO v7	97.3	96.0	96.2	3.00×10^8

Notes: The best performance values for each indicator are shown in bold in the table.

The test platform is a computer running the Ubuntu 18.04 system with a 3.60 GHz Intel Xeon(R) W-2123 processor, 32 GB of memory, and a 1 TB hard drive. The GPU is NVIDIA GeForce RTX 2080 Ti. The algorithm development platform is python 3.8, and the deep learning framework is Pytorch 1.8.1.

3.2. Analysis of dairy cow detection results and accuracy evaluation

3.2.1. Comparison experiments of different models

In order to verify the detection performance of the improved YOLO v7 model in this paper, the YOLO v5 model and the YOLO v7 model are trained using the data set in this paper, and the performance of the models is evaluated using the test set data, the experimental results are shown in Table 2.

As shown in Table 2, compared with YOLO v5-l and the original YOLO v7, the P, R, and AP values of the improved YOLO v7 model in this paper are slightly improved, with a 0.2%, 1.1%, and 1.1% improvement over the original YOLO v7 and a 0.3%, 1.5%, and 1.2% improvement over YOLO v5-l, respectively. Compared with YOLO v7-x, R and AP are promoted, and only the P value is slightly lower. However, it is worth noting that the number of parameters of the improved model in this paper is the lowest among the four models, and it is 1.66×10^8 and 0.64×10^8 less than YOLO v5-l and YOLO v7, and only 40% of YOLO v7-x. This is due to the use of grouped convolution and serial pooling operations in the Spatial Pyramid Pooling module of the model in this study, which reduces the number of parameters while ensuring the invariance of the receptive field, indicating the effectiveness of the improved Spatial Pyramid Pooling module in this paper.

In conclusion, the proposed model in this paper has the highest Recall (R) value, Average Precision (AP) among the four models and the least number of model parameters. This result demonstrates that the proposed model in this paper has certain advantages in accurately recognizing cow targets in complex environments.

3.2.2. Network improvement ablation experiments

In order to further verify the effectiveness of the optimization method used in this study and the performance improvement of the fusion model in dairy cow detection, an ablation experiment was conducted on the improvement methods in sections 2.2.2 and 2.2.3, training the network in 4 different situations, i.e., using the original ELAN and SPPCSPC, only improving one of ELAN and SPPCSPC, and improving both ELAN and SPPCSPC at the same time. The detection performance of the network in 4 situations is verified, and the experimental results are shown in Table 3.

As shown in Table 3, in experiment (2), by using the SPPCSPC-L module, the number of parameters is reduced by 16%, compared to the original YOLO v7 model in experiment (1). P, R and AP are improved by 0.2%, 0.6%, and 0.7%, respectively. In experiment (3), ELANA is used to improve the Backbone network; compared to experiment (1), the model parameters slightly decrease with P, R and AP promoted by 0.2%, 1.0%, and 1.1%, respectively. Experiment (4) is the improved model in this study; compared to the original YOLO v7 model, the number of parameters decreases by 18%, P improves by 0.2%, R improves by 1.1%, and AP improves by 1.1%. The results of this experiment show that the improved Spatial Pyramid Pooling module (SPPCSPC-L) and the improved Backbone network (ELANA) play an important role in improving object detection performance and the model simplification.

Table 3
Ablation experiment results.

Index	ELANA	SPPCSPC-L	P/%	R/%	AP/%	Params
1	×	×	97.1	94.9	95.1	3.64×10^8
2	×	√	97.3	95.5	95.8	3.07×10^8
3	√	×	97.3	95.9	96.2	3.57×10^8
4	√	√	97.3	96.0	96.2	3.00×10^8

The above-mentioned optimization algorithms, along with the original YOLO v7 model, are tested on cow image instances, as shown in Fig. 7. It can be seen in Fig. 7a and 7b that the YOLO v7 model shows cow misses due to the tiny target caused by the occlusion between cows. In Fig. 7c, YOLO v7 misclassifies workers further away from the camera position as cows. In both cases, our improved model can correctly identify all the cows. This is because in cases where the targets are small, or there is occlusion, our improved method has incorporated the ACmix module which better balances the attention given to small targets by shallow networks and large targets by deep networks, effectively reducing false positives and false negatives.

3.3. Experiments and results analysis of multi-object tracking for dairy cow

3.3.1. Comparison of effects before and after improvement of the algorithm

To verify the performance of the multi-object algorithm, 10 videos of different scenes are used as test videos, and the tracking experiment results on these 10 videos are shown in Table 4. Also, we tested the performance of the model on these 10 videos before and after the improvement, and the results are shown in Table 5.

As shown in Table 4, the disparities in accuracy indicators are mainly due to the video environments discrepancies, such as background, weather, day or night, sparse or dense, and occlusion conditions. Comparing the performance on different videos, it can be noted that the tracking indicators on video 06 are worse than others due to the complex environment in video 06. The video is captured on a dark rainy evening, 8p.m., and there is severe occlusion between cows, which leads to poor tracking accuracy. In video 01, cows are sparse, with few occlusions' interference, so the tracking effect is best.

In Table 5, the tracking inference speed decreases slightly using the improved YOLO v7 algorithm as a detector compared with the original YOLO v7 model, but all other accuracy indicators improve significantly, which proves the effectiveness of our improved YOLO v7 model. After adopting the improved ByteTrack algorithm in this paper, HOTA, MOTA, and MOTP all have some accuracy improvement, which shows that the tracking performance can be effectively improved by directly predicting the width and height of the tracking boxes.

The average HOTA of our algorithm is 67.6%, which is an improvement of 4.4% compared to the original model, indicating that the performance of the trajectory tracking method in this paper is superior overall. MOTA and MOTP, compared to the original model, improves by 6.1% and 0.8%, respectively, indicating that the method in this study has higher accuracy and more minor positional error for tracking targets. The observed 3.8% increase in IDF1 implying that the algorithm in this paper can better adapt to different scenarios and has higher tracking stability. The YOLO-BYTE model has an ID value of 85, which is 37.5% less than the original model. Additionally, the ID values of videos 08 and 10 are 0, and the ID values of videos 01, 02, 05, 06, and 09 are all less than 10, indicating that that our tracking model can continuously track targets in a variety of environments without losing track of the targets, resulting in stable tracking performance. In terms of tracking algorithm speed, the average FPS of YOLO-BYTE is 47f/s, which is 9.4f/s slower compared to the original model. In conclusion, the proposed model in this study significantly improves the tracking

Fig. 7. Comparison of cow instance detection after algorithm optimization. In Scenes 1 and 2, the targets in the black boxes are missed cows; in Scene 3, the target in the black box is a false positive, misclassifying the worker as a cow.

Table 4
Experiment results on the cow tracking dataset.

No.	HOTA/%	MOTA/%	MOTP/%	IDF1/%	IDS	FPS/(f/s)
01	87.1	94.6	88.1	97.2	1	47.4
02	69.7	83.9	83.3	85.2	7	47.2
03	62.1	68.7	83.9	67.7	27	46.8
04	68.6	81.8	82.6	78.3	14	46.4
05	68.9	74.3	79.2	84.1	6	46.7
06	55.8	59.0	80.1	69.8	3	47.5
07	68.6	73.4	84.3	76.4	22	46.3
08	83.7	96.5	85.3	98.2	0	46.8
09	59.0	68.1	75.1	79.6	5	47.3
10	71.4	79.7	79.4	89.7	0	47.1
Overall	67.6	75.4	83.0	77.8	85	47.0

accuracy and stability compared to the original model. It also achieves high accuracy and speed in tracking the ID of cows, providing technical support for precise management in dairy farming.

The improved ByteTrack algorithm's tracking results are shown in Fig. 8, where the green boxes represent the tracking results of cows using the improved ByteTrack algorithm, and the red boxes represent the original ByteTrack algorithm's tracking performance. As shown in Fig. 8a and 8c, the red boxes put some areas of the cow's mouth, belly, and tail outside the tracking boxes, while the green tracking boxes can better encompass the cow and the tracking boxes are more accurate. In Fig. 8b and 8d, the red boxes contain more background while the green tracking boxes fit the cow more closely and include less background, indicating that the improved ByteTrack algorithm better matches the tracking boxes to the cow targets.

As shown in Fig. 9, it represents the performance of the proposed model and the original model in video 07. It can be seen from Fig. 9a that the original model missed the cow during tracking, such as the black dotted line boxes in frames 48, 450, and 711. It can be seen from Fig. 9b that the YOLO-BYTE algorithm has a good performance in daytime and cow occlusion (Medium), which can accurately track each cow and significantly reduce missed detection. From Fig. 9a, it can be seen that there is a false detection, on the upper right corner of the video frame 450, where a worker is wrongly detected in the tracking box 249 as a cow. However, our proposed YOLO-BYTE algorithm doesn't have any false detection and has better detection and tracking performance.

The video 06 is shot on a rainy night with partial lighting, dark illumination, heavy obstructing, and other disturbances. The tracking results of YOLO-BYTE and the original model on video 06 are shown in Fig. 10. As can be seen from Fig. 10, due to the dark lighting and serious obstruction between the cows, as well as the interference of lights, some cows got similar color to the video background, causing difficulties in detecting and tracking of these cow targets, resulting in missed tracking

Table 5
Tracking results before and after improvement of algorithm.

Models	HOTA/%	MOTA/%	MOTP/%	IDF1/%	IDS	FPS/(f•s^{-1})
YOLOv7+ByteTrack (Original)	63.2	69.3	82.2	74.0	136	56.4
Improved YOLOv7+ByteTrack	67.0	74.7	82.2	78.3	86	47.9
YOLO-BYTE	67.6	75.4	83.0	77.8	85	47.0

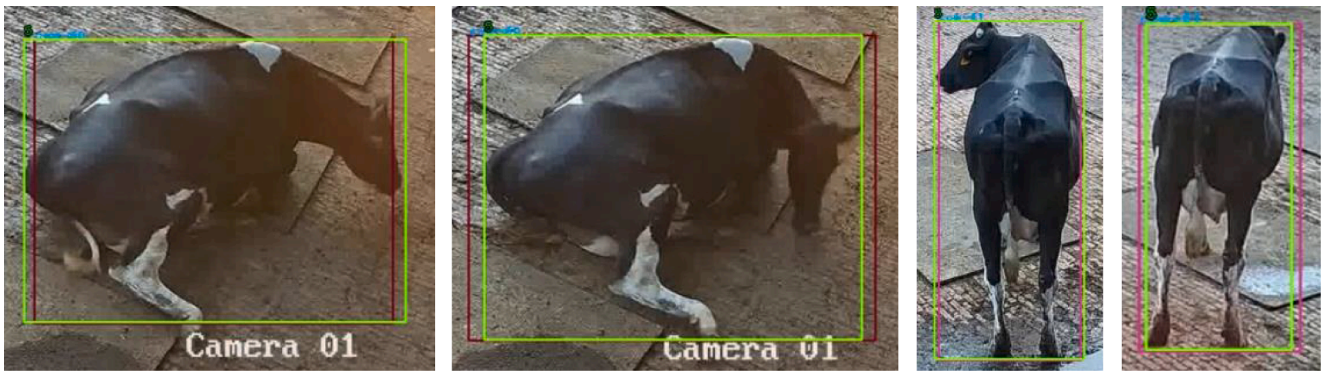
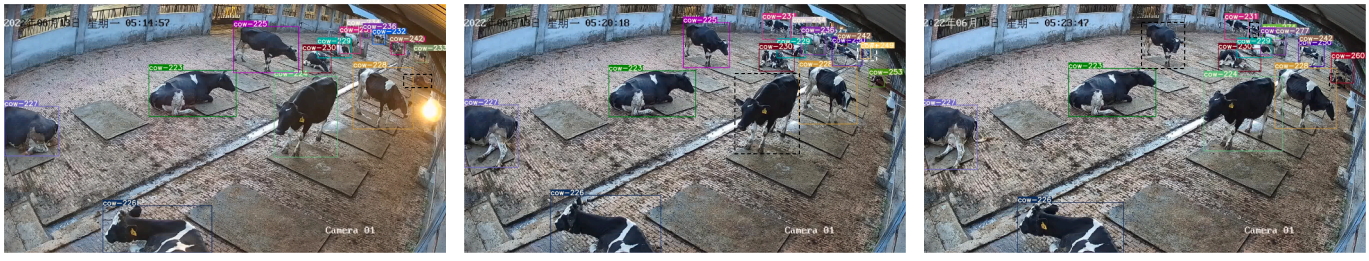


Fig. 8. Comparison of Kalman filter improved tracking results. The green boxes represent the improved ByteTrack algorithm for tracking the cow, while the red boxes represent the tracking results of the original algorithm.



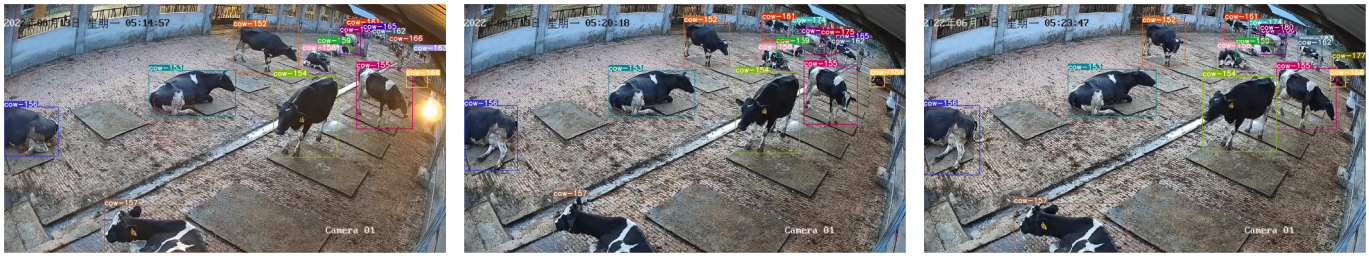


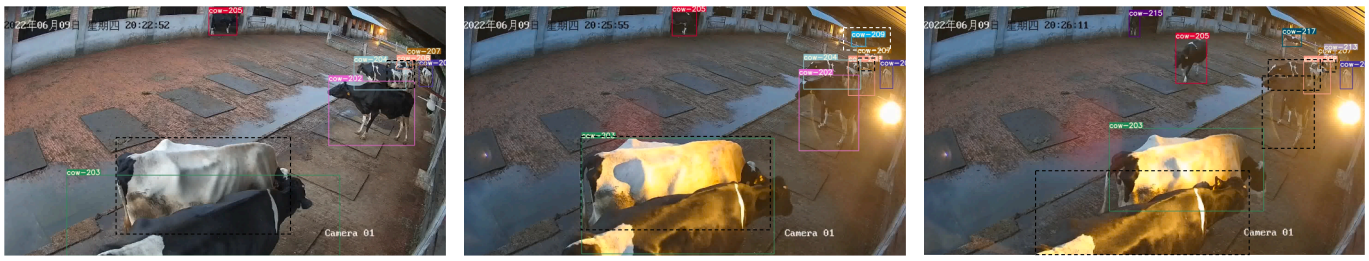
Fig. 9. The tracking results of the proposed model compared to the original model. The black dotted line in the image represents a target that is not tracked, while the white dotted line represents a target that is mistakenly tracked.

which are the black dotted in Fig. 10a, while the YOLO-BYTE model in Fig. 10b successfully tracks the corresponding target. Due to the dark background of the video at night, both the proposed algorithm and the original model have encountered false tracking cases, mistaking the distant drinking pool for a cow, such as the white dotted line targets in frame 161 of Fig. 10a and Fig. 10b. At the same time, due to the complicated environment, some cows have similar appearances, causing ID switching during tracking, such as cow 203 in Fig. 10a and cow 150 in Fig. 10b. Overall, the proposed method still has good tracking results in complex environments compared to the original model.

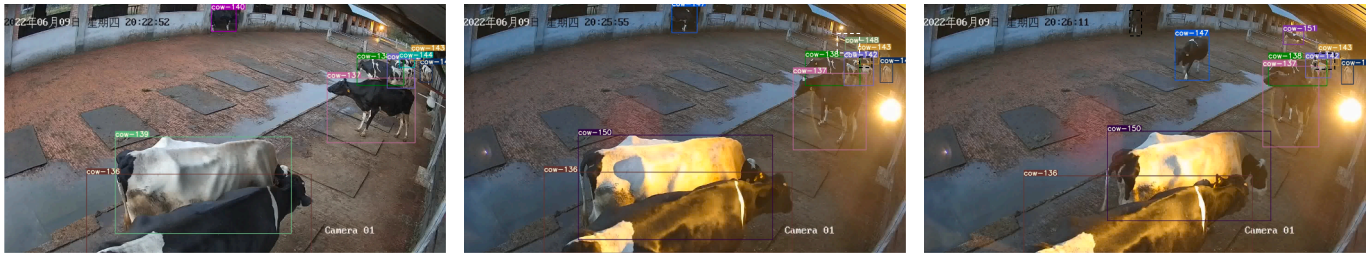
3.3.2. *Comparison of different multi-object tracking algorithms*

To verify the performance of the cow multi-object tracking method in this study, it is tested on the cow multi-object tracking dataset and compared with 6 state-of-art models in object tracking field, and the results are shown in Table 6 and Fig. 11.

As can be seen from Fig. 11, the YOLO-BYTE algorithm has the highest HOTA metric, indicating that our tracking algorithm has a better tracking performance. In Table 6, although the inference speed of the tracking algorithm using YOLO v5 as the detector is close to that of the algorithm using YOLO v7, each accuracy index is significantly higher, indicating that YOLO v7 is more suitable as the detector for the multi-object cows tracking method. Comparing the same type of tracking algorithm, we can see that DeepSort has the worst performance. Compared with Sort, DeepSort adds appearance feature associations to cascade matching. Still, the feature extraction of appearance feature associations uses only less computational cost to ensure inference speed. Thus, it is unable to effectively discriminate targets with high similarity in appearance like dairy cows. By comparing Sort and ByteTrack algorithms, it can be seen that the FPS decreases slightly (1 ~ 2FPS) as the complexity of the algorithm gradually increases, but the tracking accuracy has improved significantly. In terms of accuracy indexes such as



(a) YOLO v7+ByteTrack



(b) YOLO-BYTE

Fig. 10. The tracking results of the proposed model compared to the original model. The black dotted line in the image represents a target that is not tracked, while the white dotted line represents a mistakenly tracked target.

Table 6
Multi-object tracking results.

Models	HOTA/%	MOTA/%	MOTP/%	IDF1/%	IDS	FPS(f•s^{-1})/
YOLOv5-l+ByteTrack	63.6	67.3	81.8	72.4	246	55.6
YOLOv5-l+Sort	60.7	64.3	82.7	67.4	303	56.3
YOLOv5-l+DeepSort	58.9	61.1	81.9	67.4	164	40.7
YOLOv7+ByteTrack (Original)	63.2	69.3	82.2	74.0	136	56.4
YOLOv7+Sort	62.3	67.5	82.9	69.7	248	56.7
YOLOv7+DeepSort	59.5	63.3	82.4	68.1	151	41.6
YOLO-BYTE	67.6	75.4	83.0	77.8	85	47.0

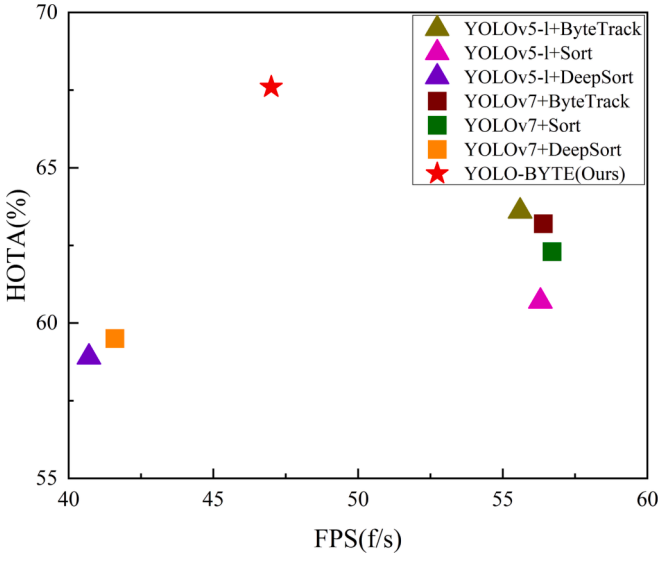


Fig. 11. Performance comparison of tracking algorithms.

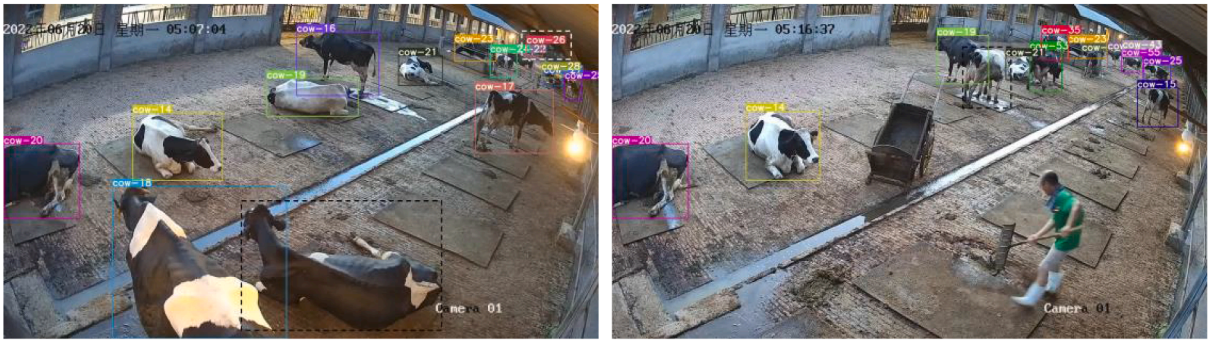
HOTA, MOTA, MOTP, and IDF1, YOLO-BYTE is optimal, which is because the multi-object tracking method proposed in this paper improves the target tracking performance by removing the background and other interfering information from the low score detection results while retaining the high score detection boxes, and effectively using the low-score detection boxes. Our method's total number of IDS is only 85, far fewer than the 6 comparison algorithms, indicating that YOLO-BYTE has the least number of ID switching and the most stable performance in multi-object tracking under different environments. The results suggest that the proposed method is more suitable for multi-object tracking dairy cows in complex scenarios.

YOLO-BYTE, DeepSort, Sort, and ByteTrack algorithms in video 02 visualization tracking results are shown in Fig. 12. Video 02 is shot on a cloudy morning, with light interference, low light, and heavy dairy cows' occlusion. Compared to the tracking results of the algorithms in Fig. 12, Sort, DeepSort, and ByteTrack all have cases of missed detection in tracking (the targets in the black dotted line frames in Fig. 12a, 12b, and 12c). However, our method has no missed tracking targets and still has good detection and tracking ability for cows with higher degrees of occlusion. In terms of false detection, Sort and DeepSort eliminate low-score detection boxes, so there are no false detection cases. Since ByteTrack algorithm retains low-score detection boxes when similar detection targets appear, false detection cases may occur, such as the 26th target in frame 26 of video 02 (Fig. 10c), which falsely detects the background as a cow. However, our method improves the YOLO v7

Video 02 frame 026 Video 02 frame 742

(a) YOLO v7+Sort

(b) YOLO v7+DeepSort



(c) YOLO v7+ByteTrack

(d) YOLO-BYTE

Fig. 12. The tracking results of multiple tracking algorithm. The black dotted line in the image represents a target that is not tracked, while the white dotted line represents a target that is mistakenly tracked.

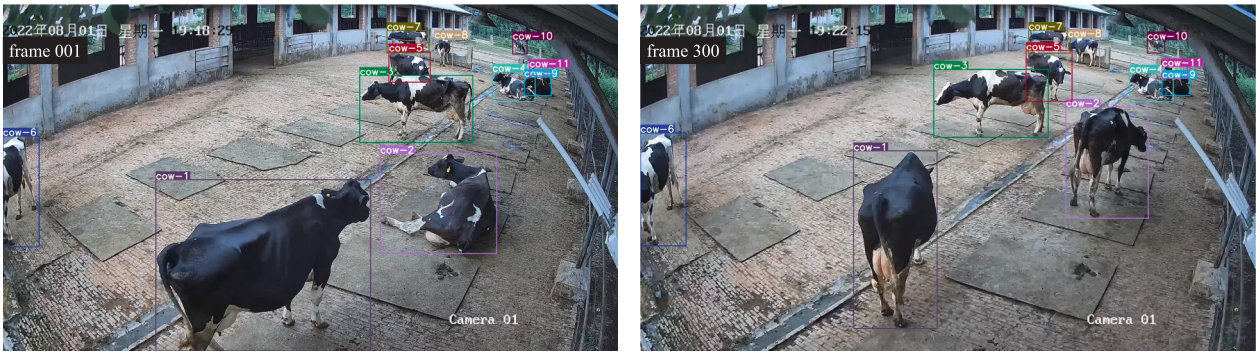


Fig. 13. Tracking results under different camera in this article cow farm.

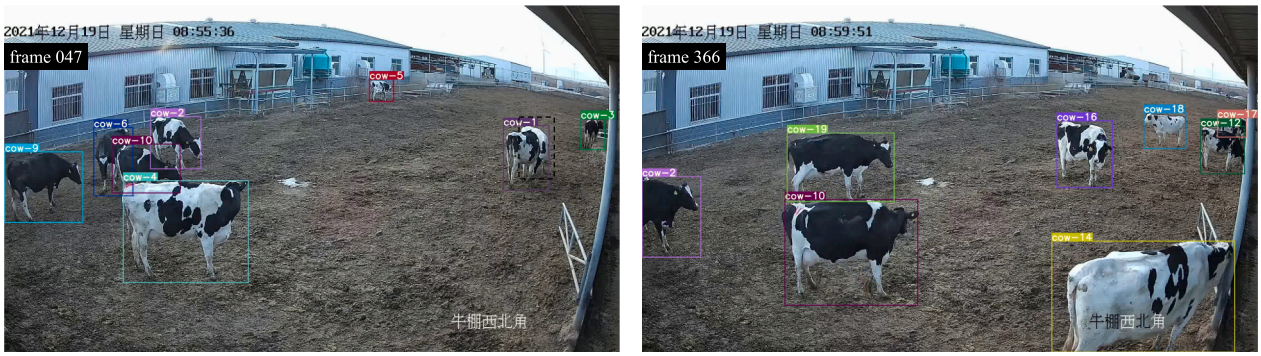


Fig. 14. Tracking results at another dairy farm. The black dotted line in the image represents a target that is not tracked.

algorithm and integrates it with ByteTrack, so it also has stable tracking performance in complex environments with low light and light interference.

In conclusion, compared to other multi-object tracking methods, our algorithm demonstrates superior tracking performance in complex scenes, providing good technical support for the precise management of dairy farming by achieving good multi-object tracking performance for dairy cows under different scenarios.

3.3.3. Analysis of tracking performance in different dairy farm environments

To further evaluate the effectiveness of the multi-target tracking algorithm proposed in this study in other scenarios, we carried out two extra experiments. For the first one, the video was captured at the cattle farm under another camera angle, and the tracking performance of YOLO-BYTE is shown in Fig. 13. As shown in Fig. 13, YOLO-BYTE can completely track the targets in the dairy farm and there is no cow target tracking ID switching.

For the second, the video was obtained from the standardized core production base in Wuzhong City, the Ningxia Hui Autonomous Region, China, and YOLO-BYTE tracking results are shown in Fig. 14. The scope of this dairy farm is broader, and YOLO-BYTE can track almost all cows. In Frame 47, due to the overlapping of Cow 1 and another cow, they are identified as a single target, resulting in a missed detection. Fig. 15 shows the results of another camera tracking under this cow field. In frame 325, the cow at the corner is far away from the camera and the target is small, so there is a missed detection. In summary, YOLO-BYTE still has good tracking performance in different cow farms and with different cameras.

4. Discussion

This study presents a multi-object tracking method for Holstein dairy cows in activity fields with a single camera, which shows higher tracking accuracy than various existing methods under real-time tracking conditions. However, as with any research, there are limitations to our work that need to be considered. One major limitation is the potential impact of environmental factors on our model's accuracy. For example, the placement of the camera and the size and shape of the activity field may affect the accuracy of our tracking algorithm. And also, the algorithm can benefit from further examination on diverse datasets to improve the robustness and generalizability. In addition, in order to record the behavior trajectory of cows 24/7 and serve the breeding information system, the next step we will explore Multi-Target Multi-Camera Tracking (MTMCT) of dairy cows across different scenarios, such as grazing or resting areas. Such research can help to extend the applicability of our method and contribute to the development of more effective and robust tracking algorithms for dairy cows in various settings.

5. Conclusion

(1) The improved YOLOv7 algorithm is proposed to address the problem of low tracking accuracy caused by the uneven distribution of cows and large changes in target scale. The ACmix module is introduced into the model, improves the design of the ELANA feature extraction module, and adopts the improved lightweight Spatial Pyramid Pooling module (SPPCSPC-L) to reduce the number of model parameters.

(2) The improved ByteTrack algorithm is proposed to overcome information loss such as cow occlusion and deformation by utilizing ByteTrack tracking to retain more low-score detection boxes and thereby preserve more tracking information. In order to ensure that the cow matching tracking boxes fit well with the cow targets, the improved ByteTrack algorithm directly predicts the width and height information of the tracking boxes in the Kalman filter.

(3) Experimental results show that the proposed model has a P of 97.3% on the dairy cow target detection dataset, with an improvement of 1.1% in R and AP compared to the original

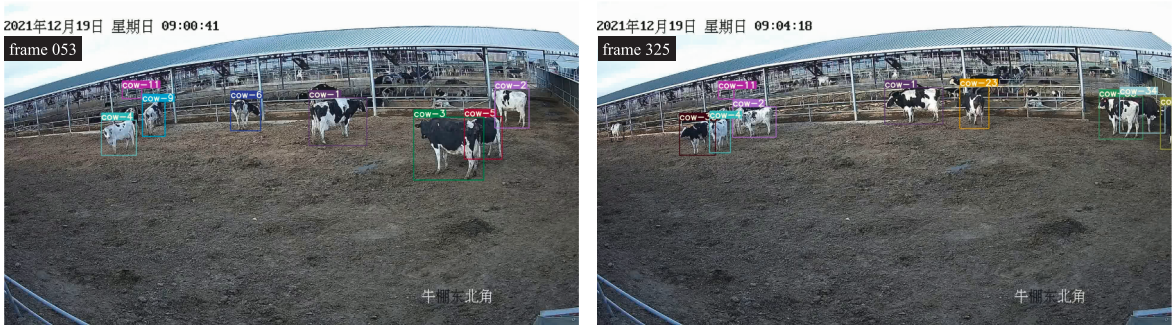


Fig. 15. Tracking results at another camera. The black dotted line in the image represents a target that is not tracked.

algorithm and a reduction of 18% in the number of model parameters. In terms of tracking performance, compared to the original model, HOTA increased by 4.4%, MOTA increased by 6.1%, IDF1 increased by 3.8%, ID decreased by 37.5%, and FPS is 47f/s. These results indicate that the proposed dairy cow multi-object tracking method can achieve real-time tracking of multiple dairy cows in natural scenes and provide technical support for non-contact automatic monitoring of dairy cows.

CRediT authorship contribution statement

Zhiyang Zheng: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Jingwen Li:** Investigation, Validation, Methodology. **Lifeng Qin:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by the Shaanxi Provincial Technology Innovation Guidance Planned Program (No. 2022QFY11-02).

References

Bergamini, L., Pini, S., Simoni, A., Vezzani, R., Calderara, S., Eath, R.B.D., Fisher, R.B., 2021. Extracting accurate long-term behavior changes from a large pig dataset. In: 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. https://doi.org/10.5220/0010288405240533.

Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. Simple online and realtime tracking. In: 2016 IEEE International Conference on Image Processing (ICIP), DOI: https://doi.org/10.1109/ICIP.2016.7533003.

Boogaard, F.P., Rongen, K.S.A.H., Kootstra, G.W., 2020. Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging. Biosyst. Eng. 192, 117–132. https://doi.org/10.1016/j.biosystemseng.2020.01.023.

Boopathi Rani, R., Wahab, D., Dung, G.B.D., Seshadri, M.R.S., 2022. Cattle Health Monitoring and Tracking System. In: 3rd International Conference on VLSI, Communication and Signal processing, VCAS 2020, October 9, 2020 - October 11, 2020, Prayagraj, India, Springer Science and Business Media Deutschland GmbH, DOI: https://doi.org/10.1007/978-981-16-2761-3_69.

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J., 2021. RepVGG: Making VGG-style ConvNets Great Again. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR46437.2021.01352.

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., Li, R., Fu, L., Zhang, Q., 2022. A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. Comput. Electron. Agric. 197, 107000 https://doi.org/10.1016/j.compag.2022.107000.

Gao, F., Wu, Z., Suo, R., Zhou, Z., Li, R., Fu, L., Zhang, Z., 2021. Apple detection and counting using real-time video based on deep learning and object tracking. Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, 37(21), 217-224, DOI: https://doi.org/10.11975/j.issn.1002-6819.2021.21.025.

Guzhva, O., Ardö, H., Nilsson, M., Herlin, A., Tufvesson, L., 2018. Now You See Me: Convolutional Neural Network Based Tracker for Dairy Cows. Front. Robotics and AI, 5. https://doi.org/10.3389/frobt.2018.00107.

Jiang, B., Wu, Q., Yin, X., Wu, D., Song, H., He, D., 2019. FLYOLOv3 deep learning for key parts of dairy cow body detection. Comput. Electron. Agric. 166, 104982 https://doi.org/10.1016/j.compag.2019.104982.

Koniar, D., Harga, L., Loncova, Z., Ducho, F., Beo, P., 2016. Machine vision application in animal trajectory tracking. Comput. Methods Programs Biomed. 127, 258–272. https://doi.org/10.1016/j.cmpb.2015.12.009.

Kumar, S., Singh, S.K., Abidi, A.I., Datta, D., Sangaiah, A.K., 2018. Group Sparse Representation Approach for Recognition of Cattle on Muzzle Point Images. Int. J. Parallel Prog. 46 (5), 812–837. https://doi.org/10.1007/s10766-017-0550-x.

Li, W., Li, F., Li, Z., 2022a. CMFTNet: Multiple fish tracking based on counterpoised JointNet. Comput. Electron. Agric. 198, 107018 https://doi.org/10.1016/j.compag.2022.107018.

Li, Z., Song, L., Duan, Y., Wang, Y., Song, H., 2022b. Basic motion behaviour recognition of dairy cows based on skeleton and hybrid convolution algorithms. Comput. Electron. Agric. 196, 106889 https://doi.org/10.1016/j.compag.2022.106889.

Liu, C., Jian, Z., Xie, M., Cheng, I., 2021. A Real-Time Mobile Application for Cattle Tracking using Video Captured from a Drone. In: 2021 International Symposium on Networks, Computers and Communications, ISNCC 2021, October 31, 2021 - November 2, 2021, Dubai, United arab emirates, Institute of Electrical and Electronics Engineers Inc., DOI: https://doi.org/10.1109/ISNCC52172.2021.9615648.

Liu, H., Reibman, A.R., Boerman, J.P., 2020. Video analytic system for detecting cow structure. Comput. Electron. Agric. 178, 105761 https://doi.org/10.1016/j.compag.2020.105761.

Luiten, J., Ošep, Aljoša, Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., Leibe, B., 2021. HOTA: A Higher Order Metric for Evaluating Multi-object Tracking. Int. J. Comput. Vis. 129 (2), 548–578.

Noe, S.M., Zin, T.T., Tin, P., Kobayashi, I., 2022. Automatic detection and tracking of mounting behavior in cattle using a deep learning-based instance segmentation model. Int. J. Innovative Computing, Inform. Control, 18(1): 211-220, DOI: https://doi.org/10.24507/ijicic.18.01.211.

Noinan, K., Wicha, S., Chaisricharoen, R., 2022. In: The IoT-based weighing system for growth monitoring and evaluation of fattening process in beef cattle farm. Institute of Electrical and Electronics Engineers Inc., Chiang Rai, Thailand https://doi.org/10.1109/ECTIDAMTNCON53731.2022.9720346.

Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z., Huang, G., 2022. On the Integration of Self-Attention and Convolution. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR52688.2022.00089.

Sun, L., Chen, S., Liu, T., Liu, C., Liu, Y., 2020. Pig target tracking algorithm based on multi-channel color feature fusion. Int. J. Agric. Biol. Eng., 13, 180-185, DOI: https://doi.org/10.25165/j.ijabe.20201303.5346.

Tan, C., Li, C., He, D., Song, H., 2022. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric. 193, 106683 https://doi.org/10.1016/j.compag.2021.106683.

Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L.M.E., Mattoccia, S., Di Stefano, L., Bonora, F., Barbaresi, A., Santolini, E., Torreggiani, D., 2021. A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. Comput. Electron. Agric. 182, 106030 https://doi.org/10.1016/j.compag.2021.106030.

Tu, S., Liu, X., Liang, Y., Zhang, Y., Huang, L., Tang, Y., 2022. Behavior Recognition and Tracking Method of Group housed Pigs Based on Improved DeepSORT Algorithm. Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery, 53(8): 345-352, DOI: https://doi.org/710.6041/j.issn.1000-1298.2022.08.037.

Wang, C.-Y., Bochkovskiy, A., Liao, H.-y., 2022a. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), DOI: https://doi.org/10.48550/arXiv.2207.02696.

Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2021. Scaled-YOLOv4: Scaling Cross Stage Partial Network. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR46437.2021.01283.

Wang, M., Larsen, M.L.V., Liu, D., Winters, J.F.M., Rault, J.-L., Norton, T., 2022b. Towards re-identification for long-term tracking of group housed pigs. Biosyst. Eng. 222, 71–81. https://doi.org/10.1016/j.biosystemseng.2022.07.017.

Wang, Y., Li, R., Wang, Z., Hua, Z., Jiao, Y., Duan, Y., Song, H., 2023a. E3D: An efficient 3D CNN for the recognition of dairy cow's basic motion behavior. Comput. Electron. Agric. 205, 107607 https://doi.org/10.1016/j.compag.2022.107607.

Wang, Y., Xu, X., Wang, Z., Li, R., Hua, Z., Song, H., 2023b. ShuffleNet-Triplet: A lightweight RE-identification network for dairy cows in natural scenes. Comput. Electron. Agric. 205, 107632 https://doi.org/10.1016/j.compag.2023.107632.

Williams, M., Zhan Lai, S., 2022. Classification of dairy cow excretory events using a tail-mounted accelerometer. Comput. Electron. Agric. 199, 107187 https://doi.org/10.1016/j.compag.2022.107187.

Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric. In: 2017 IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/ICIP.2017.8296962.

Wu, D., Wu, Q., Yin, X., Jiang, B., Wang, H., He, D., Song, H., 2020. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. Biosyst. Eng. 189, 150–163. https://doi.org/10.1016/j.biosystemseng.2019.11.017.

Xiao, D., Feng, A., Liu, J., 2019. Detection and tracking of pigs in natural environments based on video analysis. Int. J. Agric. Biol. Eng., 12, 116-126, DOI: https://doi.org/10.25165/j.ijabe.20191204.4591.

Yang, G., Xu, X., Song, L., Zhang, Q., Duan, Y., Song, H., 2022. Automated measurement of dairy cows body size via 3D point cloud data analysis. Comput. Electron. Agric. 200, 107218 https://doi.org/10.1016/j.compag.2022.107218.

Zambelis, A., Wolfe, T., Vasseur, E., 2019. Technical note: Validation of an ear-tag accelerometer to identify feeding and activity behaviors of tiestall-housed dairy cattle. J. Dairy Sci. 102 (5), 4536–4540. https://doi.org/10.3168/jds.2018-15766.

Zhang, H., Wang, R., Dong, P., Sun, H., Li, S., Wang, H., 2021. Beef Cattle Multi-target Tracking Based on DeepSORT Algorithm. Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery 52 (4), 248–256. https://doi.org/10.6041/j.issn.1000-1298.2021.04.026.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X., 2022. ByteTrack: Multi-object Tracking by Associating Every Detection Box. Computer Vision – ECCV 2022, Cham, Springer Nature Switzerland, DOI: https://doi.org/10.48550/arXiv.2110.06864.

Zheng, Z., Zhang, X., Qin, L., Yue, S., Zeng, P., 2023. Cows' legs tracking and lameness detection in dairy cattle using video analysis and Siamese neural networks. Comput. Electron. Agric. 205, 107618 https://doi.org/10.1016/j.compag.2023.107618.