Computers and Electronics in Agriculture 212 (2023) 108044

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Deep learning-based multi-cattle tracking in crowded livestock farming using video

Shujie Han [a,b], Alvaro Fuentes [a,b], Sook Yoon [c], Yongchae Jeong [d], Hyongsuk Kim [a,b,*], Dong Sun Park [a,b,*]

[a] Department of Electronics Engineering, Jeonbuk National University, Jeonju, South Korea
[b] Core Institute of Intelligent Robots, Jeonbuk National University, Jeonju, South Korea
[c] Department of Computer Engineering, Mokpo National University, Muan, South Korea
[d] Division of Electronics and Information Engineering, IT Convergence Research Center, Jeonbuk National University, Jeonju, South Korea

ARTICLE INFO

Keywords:
Deep learning
Cattle tracking
Video
Crowded livestock farming
Indoor environment

ABSTRACT

Cattle monitoring is an essential aspect of precision farming, and recent advancements have greatly contributed to understanding cattle behavior using wearable devices like ear tags and collars, as well as contactless cameras for image-based detection. However, tracking multiple cattle in real farm conditions with cameras, particularly in crowded scenarios, poses significant challenges mainly due to scale variations, random motion, and occlusion. This paper proposes a deep learning-based framework with improved techniques for multi-cattle tracking using video, aiming to overcome these limitations. The proposed algorithm utilizes a detection-based tracking approach, leveraging a YOLO-v5 detector trained specifically for cattle detection to provide initial targets. The main contributions of our research primarily focus on implementing the tracking algorithm to address the aforementioned problems. Several improvements are introduced: first, to handle appearance and scale deformation, a wide residual network with SPP-Net is employed as the backbone to extract cattle appearance information. Second, an ensemble Kalman filter is utilized to adapt to unexpected movements. Additionally, the angle from the centered position of the individuals to the origin of the image is incorporated to predict their location. Third, to tackle occlusion, a novel bench-matching mechanism is designed, allowing for the retrieval of lost trajectories based on the assumption of a known number of cattle in the barn. To validate the performance of the proposed framework, experiments are conducted using video sequences from our Hanwoo cattle tracking dataset. Comparisons with other state-of-the-art trackers are also performed. Our method achieves an accuracy of 84.49 % in data association, which represents a significant improvement considering the challenges involved in precision livestock farming applications.

1. Introduction

Animal welfare is a crucial factor in precision livestock farming, as it impacts economic gains and consumer health. Ensuring a good welfare state involves maintaining animal's physical health, positive affective state, and the ability to express natural behaviors (Sih et al., 2004). Among various welfare assessment indices, behavior stands out as a readily understandable and commonly used indicator (Li et al., 2020). Studying individual animal behavior provides valuable insights for monitoring health, production, welfare assessment, and overall livestock management. In this study, our focus is on cattle, as they are a significant group of animals in this context.

Tracking plays a fundamental role in monitoring cattle behavior, enabling precise farm management for each individual animal. The objective is to track cattle individually over extended periods, allowing for in-depth behavior analysis. Wearable devices like ear tags (Dogan et al., 2019; Zin et al., 2020; Li et al., 2021) and collars (Bailey et al., 2018) equipped with Radio Frequency Identification (RFID), accelerometers, or Global Positioning System (GPS) sensors are popular methods for cattle tracking. Especially, RFID-based sensors are widely used due to their standard specifications and the provision of cattle identity information (Gillenson et al., 2019). However, this approach

* Corresponding authors at: Department of Electronics Engineering, Jeonbuk National University, Jeonju, South Korea.
E-mail addresses: shujiejulie@jbnu.ac.kr (S. Han), afuentes@jbnu.ac.kr (A. Fuentes), syoon@mokpo.ac.kr (S. Yoon), ycjeong@jbnu.ac.kr (Y. Jeong), hskim@jbnu.ac.kr (H. Kim), dspark@jbnu.ac.kr (D. Sun Park).

https://doi.org/10.1016/j.compag.2023.108044
Received 22 June 2022; Received in revised form 15 June 2023; Accepted 4 July 2023
Available online 22 July 2023

has limitations due to the distribution of tag readers and environmental influences (Dogan et al., 2019). On the other hand, GPS sensors offer more cost-effective tracking for herds, but they are primarily suitable for outdoor use (Bailey et al., 2018). Accelerometers are often combined with localization devices (Cabezas et al., 2022; Benaissa et al., 2020) to improve performance but are not commonly employed for tracking objects.

Surveillance cameras are another prevalent technique for monitoring farm conditions through smartphone apps or computer interfaces, providing video data that can be analyzed using computer vision algorithms to identify and track cattle. Several classic algorithms such as Kernelized Correlation Filter (KCF) (Henriques et al., 2015), Tracking-Learning-Detection (TLD) (Kalal et al., 2012), boosting, Multi-Instance Learning (MIL), Random Forest (Ter-Sarkisov et al., 2017), and Kalman Filter (Martinez-Ortiz et al., 2013) have been applied for this purpose. However, the effectiveness of these tracking algorithms can be influenced by the farm configuration, which affects the feature extraction process.

While the above techniques have achieved substantial results, they have certain limitations. Wearable devices may damage or alter cattle behavior, while surveillance systems with traditional algorithms require human intervention to manually observe and detect changes. To overcome these limitations, recent works have proposed automatic systems based on RGB cameras, designed to monitor cattle behavior with reduced human involvement (Meunier et al., 2018; Fuentes et al., 2020; Salau and Krieter, 2020; Zambelis et al., 2021).

With the recent advances in deep learning in computer vision tasks, cattle identification has become a popular area of study in precision livestock farming (Xu et al., 2022; Li et al., 2022; Weng et al., 2022; Qiao et al., 2021). However, tracking multiple cattle has received less attention in the literature. Common techniques involve constraining the visual space, such as creating corridors where only one animal can pass at a time (Martinez-Ortiz et al., 2013; Hu et al., 2020). Unfortunately, these limited environments are not conducive to long-term tracking. Tracking multiple cattle in crowded conditions using video data is more beneficial but poses challenges related to scale deformation, unexpected movements, and occlusion. These challenges are primarily associated with the camera's field of view in relation to the target objects.

Existing solutions have presented independent methods to address the aforementioned challenges. For instance, Dao et al. (2015) extended the Real-time Compressive Tracking (RTCT) algorithm by (Zhang et al., 2012) to handle the degradation of appearance features caused by scale deformation in cattle. Their method involved sampling data from rectangular regions within the bounding box and computing features from each sample region. Hashimoto et al. (2020) introduced a pre-processing technique called superpixels, which accurately described the appearance and centroid of cattle. However, while complex appearance models can enhance tracking performance, they are time-consuming and cannot cope with occlusion problems. To reduce unexpected movements of cattle, Martinez-Ortiz et al. (2013) implemented an exit race spanning approximately 10 m in length. They utilized an overhead-mounted standard surveillance camera system that captured a primarily rear-view perspective of the cattle before entering a large barn. This setup minimized the chances of deformation and occlusion. To overcome occlusion problems, Andrew et al. (2017; 2020) used Unmanned Aerial Vehicles (UAVs) to fix the camera's perspective, which is beneficial for

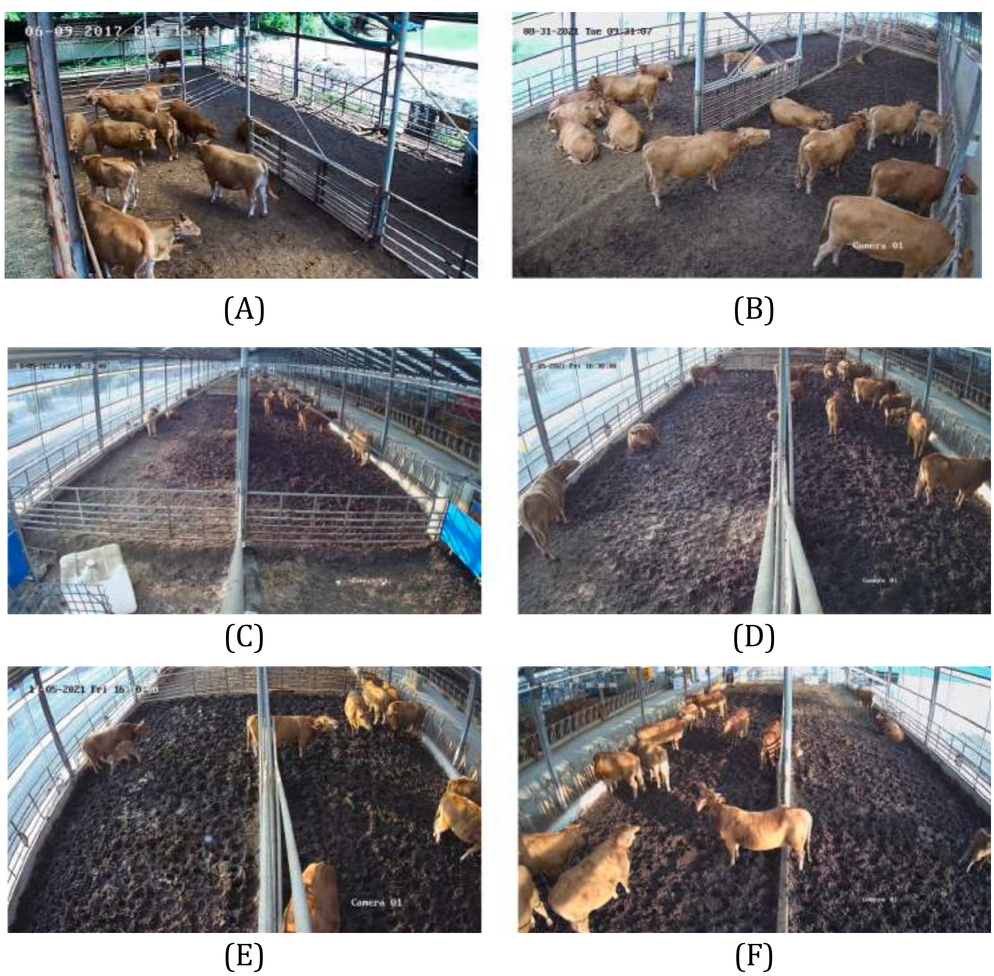


Fig. 1. Representation of a recorded video from our experimental sites. (A) Data used to train the detector. (B)-(F) Data used to test the tracker.

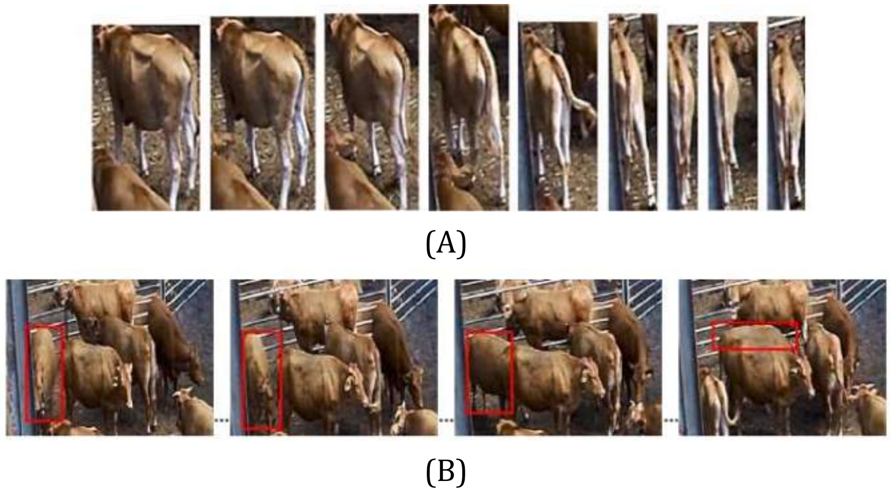


Fig. 2. Challenges on video-based multi-cattle tracking. (A) Scale deformation. (B) Unexpected movements and occlusion.

cattle tracking through a deep learning-based appearance model. However, this method is not applicable to more complex scenarios encountered in real farm setups, where freely moving cattle are monitored using cameras. Occlusion can lead to the loss of trajectories, particularly when individuals obstruct the camera's view. Ter-Sarkisov et al. (2017)devised a technique that involved extracting the contours of the tracked objects rather than relying on the background to retrieve trajectories. This approach enabled accurate tracking of cattle while minimizing confusion in trajectory reconstruction following occlusion events. However, it is important to note that this method may not consistently generalize to all cattle angles, making it less suitable for long-term tracking.

To tackle the limitations of existing techniques, this paper proposes a deep learning-based framework for multi-cattle tracking, incorporating improved techniques to address scale deformation, unexpected movements, and occlusion. Drawing inspiration from pedestrian (Sundararaman et al., 2021; Zou et al., 2022) and vehicle (Kocur and Ftacnik, 2021; Wang et al., 2021) tracking methods, we adapted and extensively explored the concepts for cattle tracking. Our framework comprises two main components: an object detector based on YOLOv5 (Jocher et al., 2022) to obtain the initial regions of the target cattle, and a tracker based on DeepSORT (Wojke et al., 2017) to track the identified individuals in the scene. DeepSORT is a well-established algorithm that leverages both motion and appearance information to achieve real-time tracking. It consists of three modules: an independent detector, a feature extractor, and a data association component operating under a tracking-by-detection paradigm. The detector operates independently from the tracker, providing detection results as input to the tracker. The feature extractor incorporates both appearance and motion models, employing a Convolutional Neural Network (CNN) model for appearance and a linear Kalman filter for motion. Finally, the Hungarian algorithm is utilized to associate the data, taking advantage of the appearance and motion model outputs.

In this paper, we introduce key contributions that focus on improving the performance of the DeepSORT-based tracking algorithm for cattle tracking, specifically addressing the challenges discussed earlier. These contributions encompass the introduction of a fixed appearance model (AM) capable of accommodating scale deformation, the integration of a 5-dimensional ensemble Kalman filter as a motion model (MM) to effectively adapt to unexpected cattle movements and the development of an innovative bench-matching mechanism (BM) designed to handle occlusion. By integrating these proposed techniques into our framework, we achieved accurate and efficient multi-cattle tracking, successfully overcoming the hurdles associated with scale deformation, unexpected movements, and occlusion.

To demonstrate the effectiveness of our work, extensive experiments were conducted using video sequences from our cattle tracking dataset, focusing specifically on Korean Hanwoo cattle. The proposed model achieved satisfactory performance in data association by employing the implemented strategies, surpassing the original DeepSORT algorithm in the same task. The advancements presented in this research provide an efficient strategy to address the challenges of tracking multiple cattle for precision livestock farming in indoor environments.

The rest of this paper is organized as follows: Section 2 introduces our datasets and elaborates on our proposed framework. Section 3 shows the experimental results and ablation study. Section 4 discusses the limitations and deployment requirements of our framework. Finally, Section 5 concludes the paper.

2. Materials and methods

In this section, we first present the dataset employed in this study. Subsequently, we examine the challenges associated with our approach and introduce the proposed method for mitigating these challenges.

2.1. Dataset

In order to conduct this study, we deployed a CCTV surveillance camera system within a cattle farm situated in Imsil, South Korea. The primary objective of this setup was to capture the natural behavior of the cattle without inducing any disturbances or modifications. Examples of the recorded videos from the two sites are presented in Fig. 1. Our experiments encompassed the utilization of a dataset specifically of Hanwoo cattle, which included both training and testing videos.

For the training of the detector, a total of 2,250 images were extracted from the video footage captured at the first farm (Fig. 1A) with a frame rate of 15 frames per second (fps). Initially, the annotated dataset was designed for cattle activity recognition (Fuentes et al., 2020), encompassing behaviors such as walking, resting, standing, etc. However, for the purpose of this study, only the location information with bounding boxes was utilized to generate a dataset specifically tailored for individual cattle detection. To train the appearance model, 570 images from the same video of the first farm were selected, comprising a total of 9,690 instances within the same video setting.

To evaluate the performance of the tracker, various videos captured from a different farm were utilized. These videos featured different viewpoints and encompassed a total of 17 cattle (Fig. 1B-1F). The recordings were obtained from a closed barn measuring 30 × 12 m. The videos included a side view (Fig. 1B), long-distance views (Fig. 1C, Fig. 1D, Fig. 1F), and a close-up view (Fig. 1E), all in the north–south

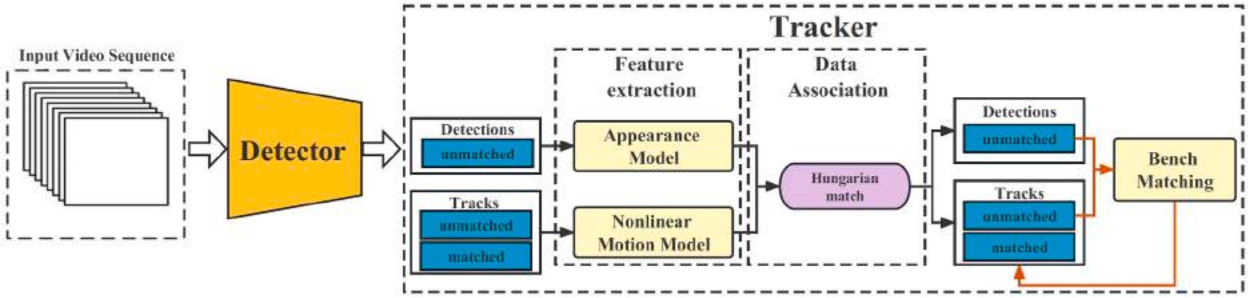


Fig. 3. General overview of our model. The detector has the function of detecting cattle targets from a video sequence. Then, the obtained detections and tracks are used for feature extraction except in the initial frame, where only detections are employed. In the data association stage, a Hungarian algorithm matches detections and tracks. Finally, the bench-matching mechanism retrieves unmatched tracks.

direction. From these videos, 2,000 images were utilized for fine-tuning the detector. Finally, the test set consisted of six video clips, totaling 3,800 images. Specifically, the side view data consisted of two clips recorded at 9:31 A.M., while the north–south direction views comprised four clips captured at 4:30P.M. on a different day.

2.2. Challenges

The task of tracking multiple cattle in crowded farm scenarios using video requires addressing the following key challenges:

- **Scale deformation**. Cattle can undergo significant shape changes relative to the camera, even with slight movements from a stable viewpoint, as shown in Fig. 2A. This deformation degrades the appearance model, as a fixed-size input image can distort the cattle's features. To address scale deformation, we employed a wide-residual network (Zagoruyko and Komodakis, 2017) with SPP-Net (He et al., 2014) as the backbone, serving as the appearance model (AM). This model effectively extracts the appearance information of cattle, considering their varying scales.
- **Unexpected movements**. Cattle movements are irregular and unpredictable, as illustrated in Fig. 2B. Conventional linear uniform motion models fail to accurately predict their movements. To account for unexpected movements, an ensemble Kalman filter was used to implement a nonlinear motion model (MM). Additionally, we incorporated the angle from the centered position of the cattle to the image origin to enhance the prediction of their corresponding locations. This approach enabled the filter to utilize a 5-dimensional prediction vector encompassing position, size, and angle.
- **Occlusion**. The camera installation setup in the farm often leads to severe occlusion as cattle move around the barn, resulting in the loss of trajectories, as shown in Fig. 2B. To tackle occlusion, we propose a

Algorithm 1: Our framework workflow.

```
/* D_bboxes,D_features: Bboxes and features of detections. */
/* D_bboxes*, D_features*: matched Bboxes and features of detections. */
/* D_bboxes', D_features': unmatched Bboxes and features of detections. */
/* T_bboxes, T_features: Bboxes and features of tracks. */
/* NNDist: Nearest Neighbor Distance. */
/* MahaDist: Mahalanobis Distance. */
input  : Images(imgs) of consecutive frames
output: The bboxes and IDs of the objects

for i ← 1 to len(imgs) do
    D_bboxes, crops ← Detector(i);
    D_features ← AppearanceModel(crops);
    appearance_cost_matrix ← NNDist(D_features, T_features);
    motion_cost_matrix ← MahaDist(D_bboxes,
     MotionModel(T_bboxes));
    cost_matrix = 0.3appearance_cost_matrix + 0.7motion_cost_matrix;
    D_bboxes*, D_features*, D_bboxes', D_features' ←
     Hungarian_match(cost_matrix);
    D_bboxes*, D_features*, D_bboxes', D_features' ←
     BenchMatching(D_bboxes', D_features', appearance_cost_matrix);
    T_bboxes, T_features ← update(D_bboxes*, D_features*);
    T_bboxes, T_features ← create(D_bboxes', D_features');
```

Fig. 4. Pseudocode of the cattle tracking framework.

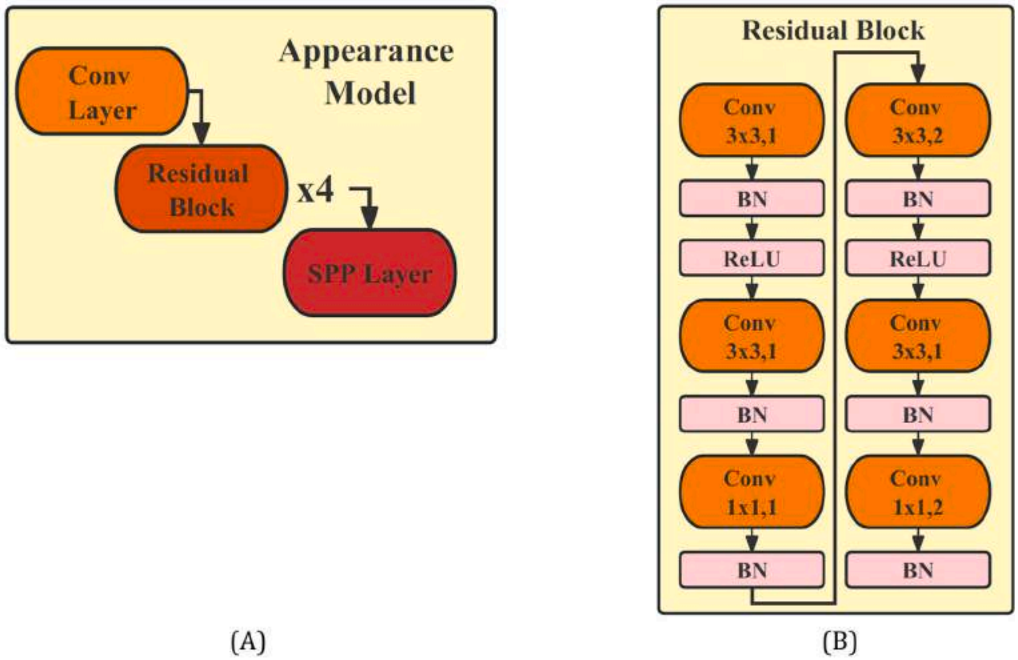


Fig. 5. Architecture of the appearance model. (A) Overview of the model with a convolution layer, four residual blocks, and an SPP-Net layer. (B) A residual block from the appearance model. 1 and 2 represent the stride size.

novel bench-matching mechanism (BM) aimed at recovering lost trajectories. This mechanism relies on the assumption that the number of cattle in the barn remains constant and they never disappear from the scene. When a new trajectory is created, it is considered a "bench player" temporarily listed on the bench for the subsequent few frames. Subsequently, this new trajectory is compared with the lost trajectories to determine whether it should replace a lost trajectory or be retained as a new one.

2.3. Proposed method for multi-cattle tracking

This part introduces the proposed method, starting with a general overview of the architecture. Subsequently, we provide a comprehensive description of each component within the system, which includes the appearance model (AM) to adapt cattle features effectively, the utilization of a nonlinear motion model (MM) that incorporates 5-dimensional motion information, and the implementation of a bench-matching mechanism (BM) designed to recover lost trajectories.

2.3.1. General overview

The proposed cattle tracking framework consists of the following modules: the detector, feature extraction, data association method, and bench-matching mechanism. These modules are depicted in Fig. 3. Our primary contributions lie in the feature extraction component and the novel bench-matching mechanism (BM) to address the challenges mentioned in Section 2.2. Within the feature extraction part, we employed the appearance model (AM) and non-linear motion model (MM) to extract target features in diverse scenarios. Additionally, we introduce the innovative bench-matching mechanism (BM) to iteratively match and recover lost trajectories within the barn. Subsequent subsections provide detailed explanations of each module.

The process starts with an image sequence extracted from a video used as input to the detector. The object detector was trained separately on our cattle dataset, and its final weights were used to provide the initial detection targets with their corresponding location within an image. Specifically, we utilized a one-stage detector, YOLOv5 (Jocher et al., 2022), due to its robustness in dealing with scale variations and faster processing speed. YOLOv5 creates features from the input sequence and feeds them through the prediction head to draw boxes around targets and predict their classes.

Once the initial targets are obtained, they are initially marked as unmatched detections and randomly assigned ID numbers to enter the tracker. Then, the feature extraction module, serving as the first part of the tracker, processes the AM and MM. The data association then matches the detections and tracks. As a result, matched detections are marked as matched tracks. Finally, the BM uses the remaining unmatched detections to retrieve unmatched tracks.

The algorithm's pseudocode is presented in Fig. 4. The appearance cost was evaluated using the nearest neighbor distance, and the motion cost was described using the Mahalanobis distance. We set the appearance weight as 0.3 and the motion weight as 0.7 to obtain the cost matrix to match tracks, considering that cattle movements were easier to predict in the short term. Finally, the system creates new tracks if there are still unmatched detections after going through all the modules.

2.3.2. Appearance model (AM)

The appearance model involves extracting features from the detections, but it does not apply to tracks. For tracks, we store the feature vectors obtained from the previous frame. To address this, we designed a wide-residual network as the appearance model. The appearance model consists of a convolution layer, followed by four residual blocks in the backbone, and an SPP layer, as shown in Fig. 5A.

In the case of our detection problem, the shape of the cattle in the image sequence undergoes significant deformation due to changes in perspective, primarily because of their large and non-rigid body. While a typical solution would involve normalizing the size of detections, this approach introduces feature distortion, preventing the model from learning essential features. To overcome this issue, we incorporated an SPP layer into our architecture. The SPP layer utilizes spatial pyramid pooling to remove the fixed-size constraint of the network instead of normalization, allowing for more effective feature extraction.

Fig. 5B illustrates the architecture of a residual block used in the appearance model. Each residual block consists of convolution layers with kernel sizes of 3×3 and 1×1, except the first block, which solely contains convolution layers with a 3×3 kernel size. The channel configurations for the blocks were set as [64, 128, 256, 512]. Additionally,

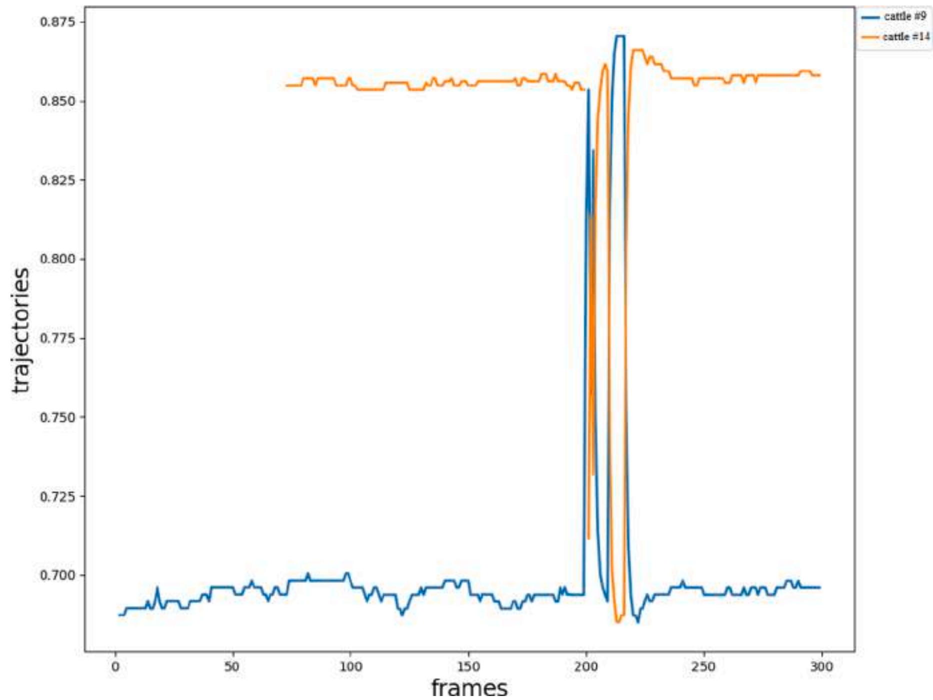


Fig. 6. The *sinθ* value of different trajectories. The value changed drastically when an ID transfer happened.

we introduced an SPP layer with kernel sizes of 3×3, 2×2, and 1×1.

During training, we employed two fully-connected layers and a SoftMax classifier to detect cattle. In the tracking stage, we only utilized the feature vectors from the backbone.

2.3.3. Nonlinear motion model (MM)

In addition to the appearance features, motion features were also extracted in the feature extraction module. While an original Kalman filter relies on detections as measures to predict tracks, we found that the random nature of cattle movements in our application requires a nonlinear movement model for more accurate predictions. Therefore, we employed an ensemble Kalman filter approach, which utilizes an ensemble of hundreds to thousands of state vectors randomly sampled around the estimate and adds perturbations at each update and prediction step, resulting in improved predictions that better align with cattle motion patterns.

Additionally, to capture essential information about cattle movement, we utilized a 5-dimensional vector. This vector includes the center position, height, width of the cattle, and the angle from the center position to the image origin. We observed that the absolute position of the cattle within the barn serves as a reliable indicator of their location, assuming they never leave the designated space.

In Fig. 6, we present an example of the trajectories of the *sine* value of the angle for two annotated cattle, identified as No.9 and No.14. We noticed significant changes in these values when the cattle's trajectory deviated. Hence, the accurate angle feature played a crucial role in data association. To predict the 5-dimensional vector $[x, y, w, h, \theta]$ as the motion features, we utilized the ensemble Kalman filter. Here, x and y

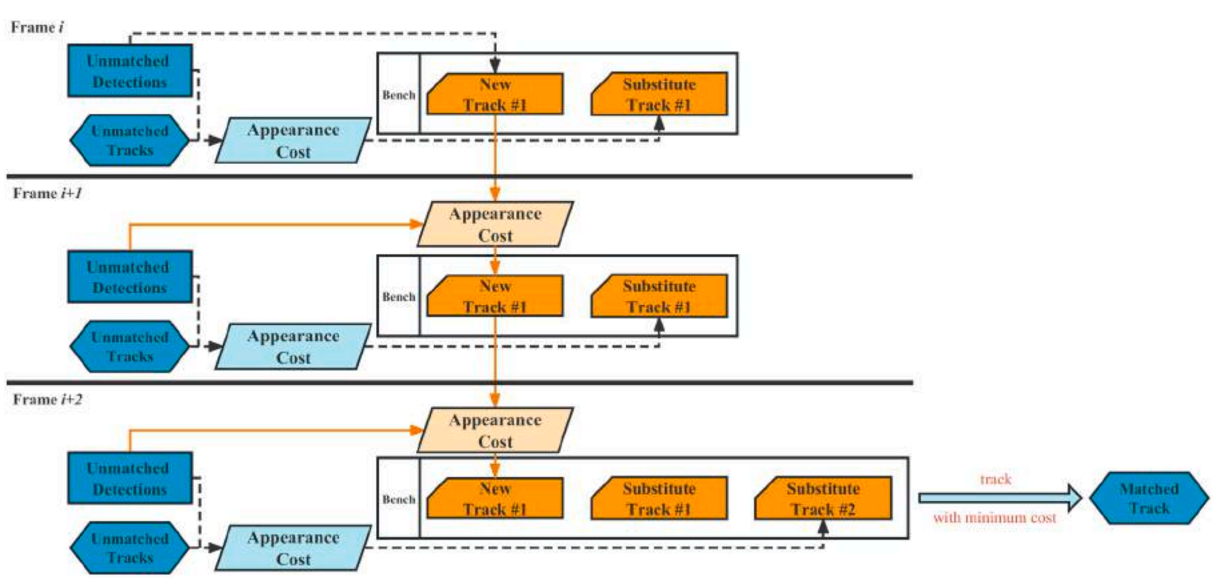
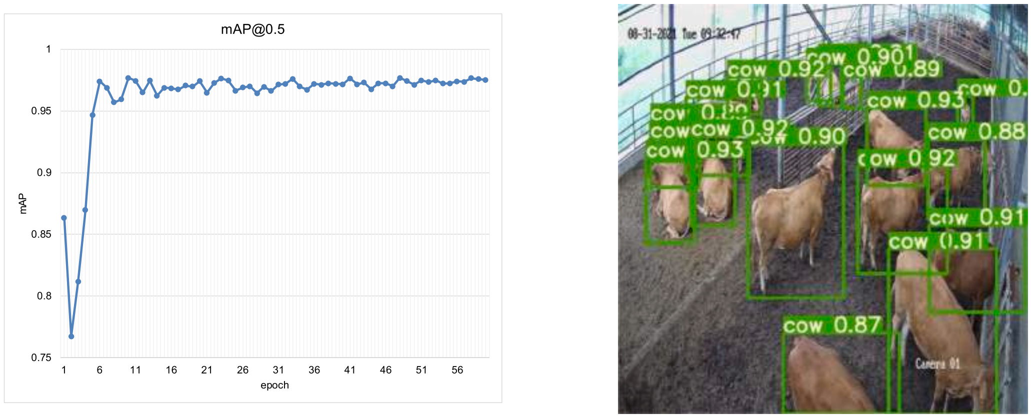


Fig. 7. Flow of the bench matching mechanism. Starting from the *i*-th frame, an unmatched detection is processed in two ways, created as a new track and re-matched with a substitute track. After 3 frames, the minimum cost bench is selected as matched track.



(A) (B)

Fig. 8. Results of cattle detection based on the YOLOv5 architecture. (A) mAP curve trained with an IoU threshold of 0.5. (B) Example results of the detector on a test image.

represent the center position coordinates, *w* and *h* denote the width and height of the cattle, and θ indicates the angle from the centered position to the image origin.

2.3.4. Bench-matching mechanism (BM)

In the final step of tracking, we applied our novel bench-matching mechanism to handle any remaining unmatched detections, typically objects that reappear after being occluded. Unmatched detections often occur during occlusion when the location of the cattle changes from its previous position. Consequently, retrieving the trajectory when the same object is detected again becomes challenging for the original DeepSORT algorithm. Instead, it creates a new trajectory, resulting in an ID transfer.

To address this issue, we propose a bench-matching mechanism specifically designed for a cattle barn, which is a closed environment where individuals consistently remain in place. Additionally, we assumed that cattle will never disappear within the barn, enabling us to retrieve lost trajectories through long-term recoveries.

Fig. 7 illustrates the concept of a bench for a trajectory. The appearance cost value is used to create a substitute track on the bench for unmatched tracks. Simultaneously, an unmatched detection is initiated as a new track on the bench. It is important to note that the bench merely represents a potential trajectory. The confirmed trajectory is determined by selecting the substitute or new track with the minimum appearance cost. In this process, motion information is not utilized as the location of the redetected object is lost. Only substitutes that match for three consecutive frames can be selected.

3. Experimental results

This section provides an overview of the implementation details for both the detector and the tracker. Subsequently, we present experimental results to showcase the performance of our proposed techniques when applied to the tracker.

3.1. Implementation details

Detector. To facilitate the training of the YOLO-v5 detector, all images were resized to dimensions of 608 $\times$ 608. Data augmentation techniques were applied to enhance the diversity of the training data, including methods such as mosaic, mixup, shear, rotation, pepper, and salt.

The detector was trained with Stochastic Gradient Descent (SGD) with a learning rate of 0.01. The training was conducted for a total of 60 epochs, with 3 warm-up epochs and a weight drop of 0.0005. Fig. 8 shows the training curves, evaluated using the mean Average Precision (mAP). Our model achieved an mAP of 97 % mAP with an IoU (Intersection over Union) threshold of 0.5. To generate more candidate detections and avoid misdetection, we adjusted the confidence and IoU thresholds. Specifically, we set the thresholds to 0.3 to increase the number of candidates while maintaining a reasonable level of accuracy.

Tracking. The tracker was trained using SGD with a learning rate of 0.01 but for 30 epochs. We employed a drop weight of 0.0005 and did not include a warm-up phase in the training process.

During the subsequent stage, candidates were first evaluated for non-maximal suppression (NMS). We implemented candidate pools with three different NMS thresholds: 0.9, 0.7, and 0.5. This approach allowed us to add more candidates to the pools when the number of candidates does not match all the paths. In the data association stage, we assigned an appearance weight of 0.3 and a motion weight of 0.7. The confidence and IoU thresholds were also set to 0.55 and 0.5, respectively. In the bench-matching stage, only the appearance model is utilized for data association.

Metrics. To evaluate the effectiveness of our framework, we employed several standard metrics commonly used in tracking evaluations. These include Higher-Order Tracking Accuracy (HOTA) (Luiten et al., 2021), Association Accuracy (AssA) (Luiten et al., 2021), Identification metric (IDF1) (Ristan et al., 2016), Multi-Object Tracking Accuracy (MOTA) (Bernardin and Stiefelhagen, 2008), and Multi-Object Tracking Precision (MOTP) (Bernardin and Stiefelhagen, 2008). MOTA and MOTP are metrics that accumulate the accuracy per frame and the precision of the bounding boxes. However, they do not count errors where the same predicted ID is changed to a different ground truth ID (ID transfer). It is then suggested that they are more likely to measure detection performance than the association. To measure ID transfer, IDF1 calculates a bijective mapping between the ground truth sets and the trajectory predictions, unlike MOTA, which matches at a detection level. Similarly, AssA measures the accuracy of the data association directly. To fairly combine all the different aspects of the tracking evaluation, HOTA is the geometric mean of a detection score and an association score. Further details regarding the metrics and their corresponding equations can be found in the Appendix.

3.2. Quantitative results

We conducted a comparative analysis of our model, which incorporates the Appearance Model (AM), Motion Model (MM), and Bench-Matching (BM) techniques to address challenges related to scale deformation, unexpected movements, and occlusion. We evaluated our model against several state-of-the-art trackers including DeepSORT (Wojke et al., 2017), SORT (Bewley et al., 2016), OC-SORT (Cao et al., 2022), ByteTrack (Zhang et al., 2022), and Tracktor (Bergmann et al.,

Table 1
Tracking results of the applied models.

Model	IDF1 (%)	AssA (%)	HOTA (%)	MOTA (%)	MOTP (%)	fps
SORT	83.59	75.4	73.55	83.08	**85.51**	**28**
DeepSORT	88.96	81.17	76.80	**84.23**	85.43	11
OC-SORT	67.97	70.97	59.39	54.59	84.39	5
ByteTrack	64.42	68.44	56.76	52.78	84.88	6
Tracktor	84.65	80.63	73.33	77.85	82.42	4
Ours	**90.22**	**84.49**	**77.64**	82.75	85.36	3

*The numbers in bold represent the best scores in each case.

2019) using our test set comprising six video clips. The tracking results averaged across the predictions from these six video clips are presented in Table 1.

In the context of cattle tracking, it is crucial to minimize ID transfer and accurately maintain the trajectories of target candidates. Therefore, metrics such as IDF1, AssA, and HOTA hold particular significance to assess cattle tracks. As presented in Table 1, our model outperformed the others in terms of IDF1, AssA, and HOTA, indicating superior data association capabilities. These results suggest that our model is better suited for effective cattle tracking, as it demonstrates enhanced ability in maintaining accurate track associations.

Our framework performed best in terms of IDF1 due to its discreet creation of new IDs through the Bench-Matching (BM) mechanism. New IDs were generated only when none of the candidates on the bench matched with unmatched detections. On the other hand, our model sacrificed detection performance by predicting absolute positions and creating bench candidates to improve data association. In contrast, DeepSORT and SORT algorithms showcased better in MOTA and MOTP, indicating that both methods focus more on detection performance than ours.

In terms of processing time, SORT utilized the IoU overlapping instead of CNN as the appearance model, resulting in the fastest inference speed. SORT achieved a rate of 28 fps, DeepSORT at 11 fps, and our proposed approach operated at 3 fps.

3.3. Ablation study

To provide a more comprehensive analysis of our proposed improvements, we conducted an ablation study that decomposed the three modules: Appearance Model (AM), Motion Model (MM), and Bench-Matching (BM). This study aims to showcase their individual effects on the final tracking results using two side-view videos, as summarized in Table 2.

Starting by implementing each model independently (top of Table 2), the results revealed that AM served as the base module for improving the baseline performance. It enhanced the AssA score by 6.99 % and achieved at least a 3 % improvement in HOTA and IDF1. This improvement can be attributed to AM effectively addressing the challenge of scale deformation by extracting more informative features. Notably, BM played a significant role in achieving these outcomes. The best results were obtained when AM, MM, and BM were used in combination. However, a decrease in performance was observed when using only two of the three models (AM, MM, BM). This can be attributed to the fact that AM focuses on appearance, while MM focuses on motion. Overemphasizing AM may undermine the effectiveness of MM when matching objects, and vice versa. It is important to note that our BM solely utilizes the appearance feature for data association. Thus, to resolve the conflict between the three models, it is essential to balance their weights based on different test data.

In summary, our results indicate that AM and MM contributed to improved tracking precision, while BM enhanced the data association capabilities of the framework. However, it is crucial to carefully balance the weights assigned to each model. Furthermore, we analyzed the inference speed of different modules and found that the AM module consumed the most time, resulting in a rate of 3 frames per second (fps). This is because inputs of varying sizes require more time to extract features compared to batch processing with uniform sizes. Nevertheless, the computational cost of the complete model depends on the specific application and the utilization of the AM. In certain cases where cattle appearance is difficult to distinguish, the AM module can be excluded from the test data. Consequently, the results of the ablation study highlight the advantages of our model in closed barns with crowded scenarios.

3.4. Qualitative results

Fig. 9 and Fig. 10 depict the qualitative results of our model. In particular, Fig. 9 showcases the outcomes obtained from various viewpoint videos recorded simultaneously. To demonstrate the significance of our study, we present the results for a continuous duration of 3 s from each video. It is important to note that each video has a duration of 46 s and was captured at a frame rate of 15 frames per second (fps). The testing was conducted independently for each video, which means the IDs assigned to the cattle in different videos do not need to be aligned.

While our model exhibited stable performance on most clear objects, it missed detecting some cattle in instances where they were located too far from the viewpoint. Nonetheless, the model's performance remained consistent on the majority of visible objects.

Fig. 10 displays the outcomes obtained from different models using the same views. The video used for evaluation has a duration of 33 s and was recorded at a frame rate of 15 frames per second (fps). In this scenario, our model successfully retrieved some cattle that were initially lost during detection, as evident from the trajectories of No. 3 and No. 18.

When comparing our model with other methods, it becomes apparent that OC-SORT and ByteTrack exhibited noticeable missing trajectories and experienced ID transfer issues, particularly with half-exposed cattle located at the bottom of the images. Additionally, Tracktor encountered detection errors, as evidenced by instances like No. 16, No. 21, and No. 23, where only the heads of the cattle were detected.

To provide a clear complement to the qualitative results, we utilized

Table 2
Ablation study on our three proposed techniques.

DeepSORT	AM	MM	BM	IDF1(%)	AssA(%)	HOTA(%)	MOTA(%)	MOTP(%)	fps
✓				80.78	73.22	68.84	**78.06**	83.98	**11**
✓	✓			84.18	80.21	72.02	74.45	84.13	3
✓		✓		84.62	81.11	72.41	77.35	84.13	5
✓			✓	85.45	81.49	72.57	77.37	84.17	5
✓	✓	✓		84.37	80.29	72.10	77.40	84.17	3
✓	✓		✓	84.44	80.12	71.97	77.28	**84.18**	3
✓		✓	✓	83.67	79.70	71.66	77.26	84.11	5
✓	✓	✓	✓	**85.48**	**81.50**	**72.58**	77.36	84.14	3

AM: Appearance model; MM: Motion model; BM: Bench matching mechanism.
*The numbers in bold represent the best scores in each case.

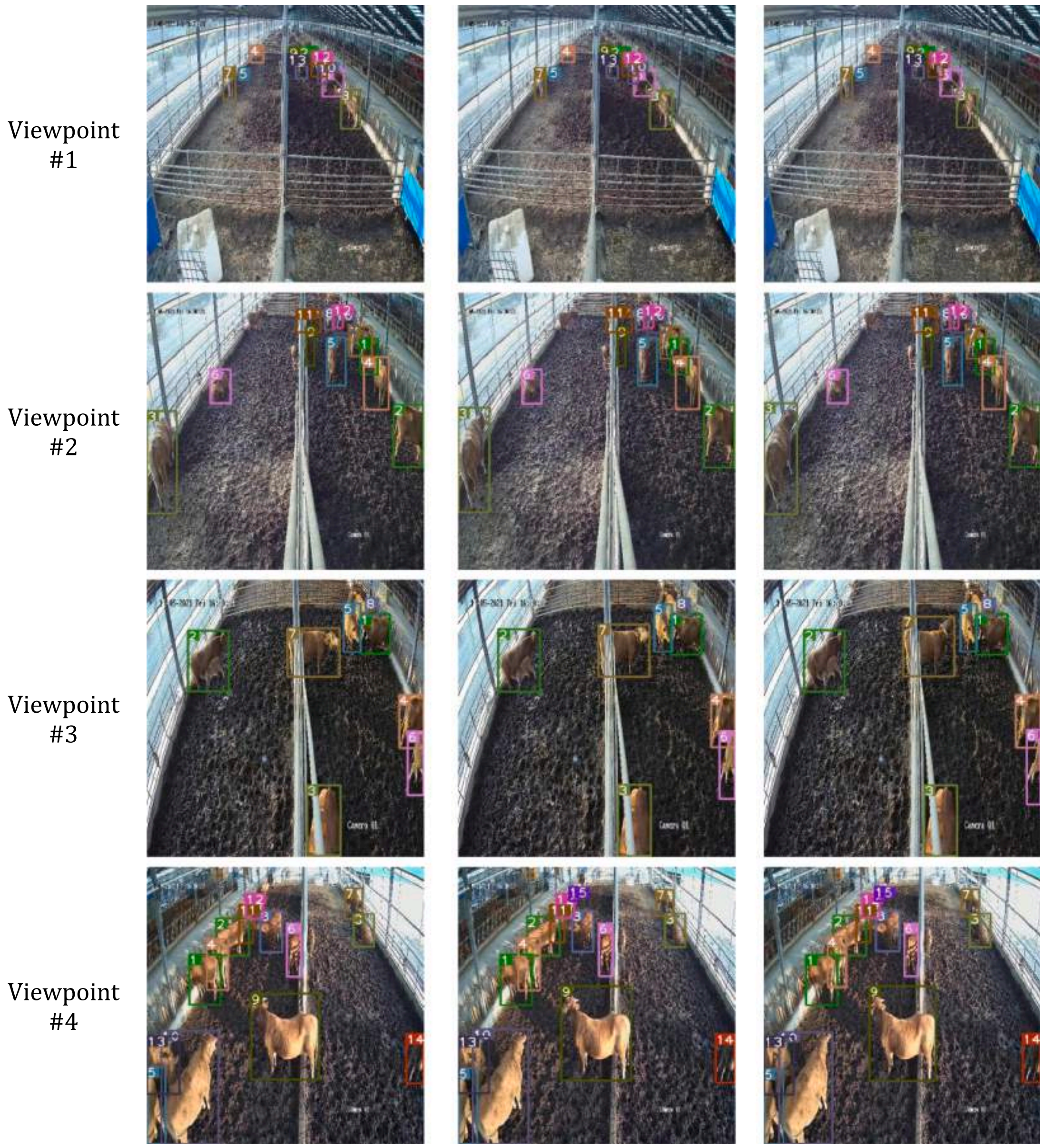


Fig. 9. Qualitative tracking results selected in 3 consecutive seconds using different viewpoints at same time.

the *sinθ* of the trajectories to visualize their positional changes. We specifically compared the trajectories generated by our proposed model with those produced by the DeepSORT model. In Fig. 11 and Fig. 12, each curve corresponds to a trajectory generated by the respective tracking model. Note that each line in the figure represents a specific object only when the model's tracking is completely accurate. Hence, an ideal result would consist of a continuous curve without any abrupt changes. A broken curve indicates that no object was detected, while a curve with sudden changes can indicate an ID transfer.

In Fig. 11, we present the tracking results of our model on a test sequence consisting of 500 frames and 17 cattle. Our model generated a total of 22 trajectories, most of which are continuous and exhibit smooth transitions. Note that the area where trajectory No. 14 is located includes multiple broken and overlapping paths. This indicates the occurrence of frequent occlusions among the cattle in that specific area, resulting in some prediction errors.

Similarly, Fig. 12 shows the results of the DeepSORT model on the same sequence. DeepSORT generated 43 trajectories, which is 21 more than our model. The trajectories produced by DeepSORT exhibit higher discontinuity and overlap. Additionally, the visualizations clearly demonstrate that the trajectories predicted by our model maintain greater consistency over time and remain in space in the long term.

In summary, the qualitative results indicate that our proposed method exhibits strong data association capabilities, although it may face challenges with distant targets or the movement of calves. This observation is consistent with the *sinθ* trajectories, where our method generates more stable and continuous trajectories compared to the original method. Even in areas with frequent occlusions, where the trajectories overlap, our method demonstrates clearer tracking results.

4. Discussion

Cattle tracking poses unique challenges due to factors such as occlusion, scale deformation, and unexpected movements. Cattle can be heavily occluded by other cattle due to their large body and their collective behaviors. And it is often observed that occludes keep moving in the state of occlusion and reappear at a different location. And also, they can be shown in various areas in an image plane according to distance

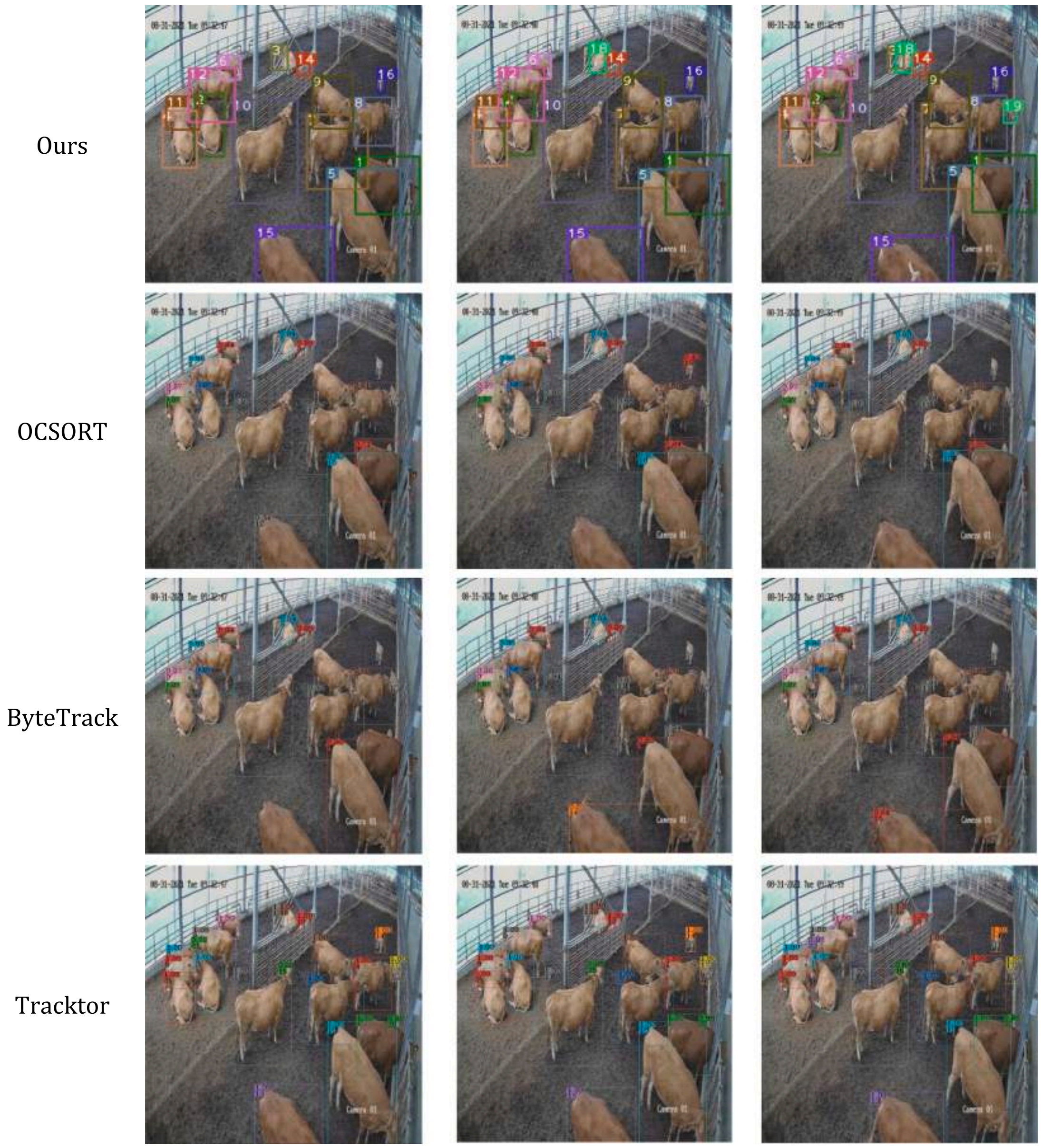


Fig. 10. Qualitative tracking results selected in 3 consecutive seconds by different models.

and angle with a camera and their postures.

Our proposed framework addressed these challenges through the integration of AM, MM, and BM techniques, enabling efficient tracking in crowded barn scenarios. However, it is important to acknowledge the strengths and weaknesses of our algorithm to guide future improvements and replication of our work. One major limitation of our framework is that the number of cattle must remain unchanged or the maximum number of tracks is required for effective bench-matching.

In comparison to other algorithms like SORT (Bewley et al., 2016), our framework does have a processing time limitation. As demonstrated in the ablation study, the AM component is the main contributor to reduced speed. This issue arises because the appearance model extracts features on a per-sample basis rather than in batches. Addressing this challenge could involve exploring techniques such as parallel computing, presenting an opportunity for improvement in future studies.

For broader application, it is recommended to implement the proposed algorithm in closed indoor barns, where a camera can be positioned at the top of the barn with a high viewing angle to ensure a clear view of all cattle. In this setup, camera resolution becomes less critical as our algorithm downscales the original image to a standardized size of 608 × 608. However, it is important to have sufficient infrastructure, including computing and networking equipment, to support the algorithm's processing requirements and achieve the desired tracking results. Adequate resources will ensure smooth operation and accurate tracking performance in real-time scenarios.

5. Conclusion

Automatic cattle monitoring using video in indoor precision farming poses significant challenges, necessitating effective solutions to address issues like scale deformation, unexpected movement, and occlusion. This paper presented a deep learning-based framework for multi-cattle tracking in video, aiming to overcome these challenges. We modified the DeepSORT algorithm to accommodate scale deformation and unexpected cattle movement. Additionally, we introduced a novel bench-matching mechanism to alleviate the problem of long-term lost trajectories. By prioritizing data association performance over detection accuracy, our method aligns with the specific requirements of cattle farms. The experimental results demonstrated the effectiveness of our

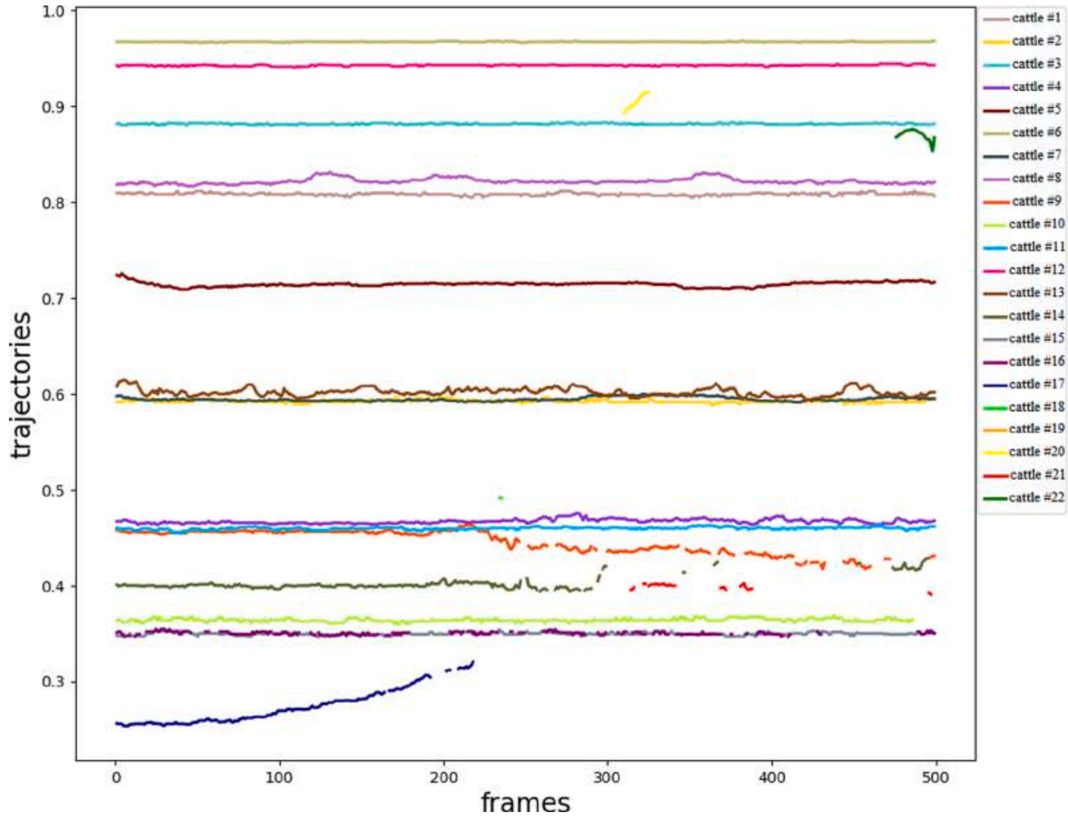


Fig. 11. *sinθ* value of all trajectories in the test sequence by our tracking model. The test sequence consisted of 17 ground truths, while our model generated 22 predicted trajectories.

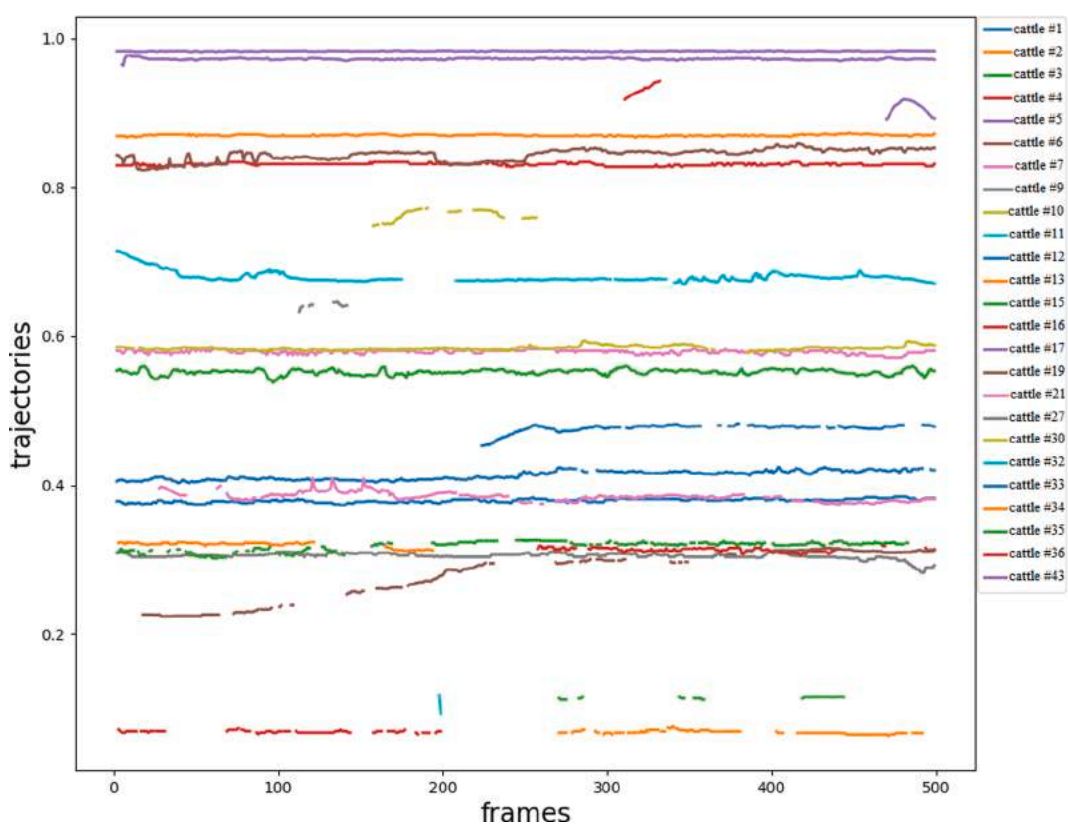


Fig. 12. *sinθ* value of all trajectories in the test sequence by DeepSORT. The test sequence consisted of 17 ground truths, while DeepSORT generated 43 predicted trajectories.

Table A1
shows the metrics and equations for multi-object tracking, IDF1 = I.

Metric	Equation	Notation
IDF1	$IDF1 = \frac{\lvert IDTP\rvert}{\lvert IDTP\rvert + 0.5\lvert IDFN\rvert + 0.5\lvert IDFP\rvert}$	IDTP: Identity True Positive IDFN: Identity False Negative IDFP: Identity False Positive
MOTA	$MOTA = 1 - \frac{\lvert FN\rvert + \lvert FP\rvert + \lvert IDSW\rvert}{\lvert gtDet\rvert}$	FP: False Positive FN: False Negative IDSW: ID Switch gtDet: Groundtruth
MOTP	$MOTP = \frac{1}{\lvert TP\rvert}\sum_{TP} S$	S: Similarity Score TP: True Positive
AssA	$AssA = \frac{1}{\lvert TP\rvert}\sum_{c\in TP}\frac{\lvert TPA(c)\rvert}{\lvert TPA(c)\rvert + \lvert FNA(c)\rvert + \lvert FPA(c)\rvert}$	TPA: True Positive Associations FNA: False Negative Association FPA: False Positive Association
HOTA	$HOTA_\alpha = \sqrt{DetA_\alpha \bullet AssA_\alpha} = \sqrt{\frac{\sum_{c\in TP_\alpha} AssIoU_\alpha(c)}{\lvert TP_\alpha\rvert + \lvert FN_\alpha\rvert + \lvert FP_\alpha\rvert}}; DetA_\alpha = \frac{\lvert TP\rvert}{\lvert TP\rvert + \lvert FN\rvert + \lvert FP\rvert}$ $HOTA = \int_{0<\alpha\leq 1} HOTA_\alpha \approx \frac{1}{19}\sum_{\alpha=0.05,\ \alpha+=0.05}^{0.95} HOTA_\alpha$	DetA: Detection Association

framework, showcasing its adaptability to real-world conditions and successful retrieval of lost trajectories in crowded scenarios. Both qualitative and quantitative results obtained from our dedicated cattle tracking dataset validate the efficiency of our proposed approach.

CRediT authorship contribution statement

Shujie Han: Conceptualization, Methodology, Software, Writing - review & editing. **Alvaro Fuentes:** Conceptualization, Methodology, Writing - review & editing. **Sook Yoon:** Supervision, Conceptualization, Methodology, Funding acquisition. **Yongchae Jeong:** Conceptualization, Methodology. **Hyongsuk Kim:** Investigation, Methodology, Review and Editing. **Dong Sun Park:** Supervision, Conceptualization, Methodology, 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

The authors do not have permission to share data.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. 2019R1A6A1A09031717); by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Korea Smart Farm R&D Foundation (KosFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Ministry of Science and ICT (MSIT), Rural Development Administration (RDA)(1545027423); and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2020R1A2C2013060).

Appendix A

See Table A1.

References

Andrew, W., Greatwood, C., and Burghardt, T. (2017). Visual Localization and Individual Identification of Holstein Friesian Cattle via Deep Learning. IN: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2850–2859. doi: 10.1109/ICCVW.2017.336.

Andrew, W., Greatwood, C., and Burghardt, T. (2020). Fusing Animal Biometrics with Autonomous Robotics: Drone-based Search and Individual ID of Friesian Cattle *(Extended Abstract). 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW),* 38–43. doi: 10.1109/WACVW50321.2020.9096949.

Bailey, D.W., Trotter, M.G., Knight, C.W., Thomas, M.G., 2018. Use of GPS tracking collars and accelerometers for rangeland livestock production research. Transl. Anim. Sci. 2 (1), 81–88. https://doi.org/10.1093/tas/txx006.

Benaissa, S., Tuyttens, F.A.M., Plets, D., Trogh, J., Martens, L., Vandaele, L., Joseph, W., Sonck, B., 2020. Calving and estrus detection in dairy cattle using a combination of indoor localization and accelerometer sensors. Comput. Electron. Agric. 168, 105153 https://doi.org/10.1016/j.compag.2019.105153.

Bergmann, P., Meinhardt, T., Leal-Taixe, L., 2019. Tracking without bells and whistles. In: In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 941–951. https://doi.org/10.1109/ICCV.2019.00103.

Bernardin, K., Stiefelhagen, R., 2008. Evaluating multiple object tracking performance: the clear mot metrics. EURASIP Journal on Image and Video Processing 2008, 1–10.

Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. Simple online and real-time tracking. IEEE International Conference on Image Processing (ICIP) 2016, 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003.

Cabezas, J., Yubero, R., Visitación, B., Navarro-García, J., Algar, M.J., Cano, E.L., Ortega, F., 2022. Analysis of accelerometer and GPS data for cattle behaviour identification and anomalous events detection. Entropy 24 (3), 336. https://doi.org/10.3390/e24030336.

Cao, J., Weng, X., Khirodkar, R., Pang, J., & Kitani, K. (2022). Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking. *arXiv preprint arXiv,* 2203-14360. doi: 10.48550/ARXIV.2203.14360.

Dao, T.-K., Le, T.-L., Harle, D., Murray, P., Tachtatzis, C., Marshall, S., Michie, C., Andonovic, I., 2015. Automatic cattle location tracking using image processing. In: 2015 23rd European Signal Processing Conference (EUSIPCO), pp. 2636–2640. https://doi.org/10.1109/EUSIPCO.2015.7362862.

Dogan, H., Basyigit, I.B., Yavuz, M., Helhel, S., 2019. Signal level performance variation of radio frequency identification tags used in cow body. Int. J. RF Microwave Comput. Aided Eng. 29 (7), e21674.

Fuentes, A., Yoon, S., Park, J., Park, D.S., 2020. Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. Comput. Electron. Agric. 177, 105627 https://doi.org/10.1016/j.compag.2020.105627.

Gillenson, M., Zhang, X., Muthitacharoen, A., Prasarnphanich, P., 2019. I've Got You Under My Skin: The Past, Present, and Future Use of RFID Technology in People and Animals. J. Inf. Technol. Manag 30 (2), 19–29.

Hashimoto, Y., Hama, H., Zin, T.T., 2020. Robust Tracking of Cattle Using Super Pixels and Local Graph Cut for Monitoring Systems. Int. J. Innovative Comput., Inform. Control 16 (4), 1469–1475. https://doi.org/10.24507/ijicic.16.04.1469.

He, K., Zhang, X., Ren, S., and Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9), 1904-1916. doi: 10.1109 / TPAMI.2015.2389824.

Henriques, J.F., Caseiro, R., Martins, P., Batista, J., 2015. High-Speed Tracking with Kernelized Correlation Filters. IEEE Trans. Pattern Anal. Mach. Intell. 37 (3), 583–596. https://doi.org/10.1109/TPAMI.2014.2345390.

Hu, H., Dai, B., Shen, W., Wei, X., Sun, J., Li, R., Zhang, Y., 2020. Cow identification based on fusion of deep parts features. Biosyst. Eng. 192, 245–256. https://doi.org/10.1016/j.biosystemseng.2020.02.001.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., TaoXie, Fang, J., imyhxy, Michael, K., Lorna, V, A., Montes, D., Nadar, J., Laughing, tkianai, yxNONG, Skalski, P., Wang, Z., … Minh, M. T. (2022). *ultralytics/yolov5: V6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference*. Zenodo. doi: 10.5281/zenodo.6222936.

Kalal, Z., Mikolajczyk, K., Matas, J., 2012. Tracking-Learning-Detection. IEEE Trans. Pattern Anal. Mach. Intell. 34 (07), 1409–1422. https://doi.org/10.1109/TPAMI.2011.239.

Kocur, V., Ftacnik, M., 2021. Multi-Class Multi-Movement Vehicle Counting Based on CenterTrack. In: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4009–4015. https://doi.org/10.1109/CVPRW53098.2021.00452.

Li, W., Bakker, J.D., Li, Y., Zheng, S., Li, F.Y., 2021. Applying a high-precision tracking system to distinguish the spatiotemporal patterns of animal movement in grassland ecology. Biol. Conserv. 255, 109016 https://doi.org/10.1016/j.biocon.2021.109016.

Li, Z., Lei, X., Liu, S., 2022. A lightweight deep learning model for cattle face recognition. Comput. Electron. Agric. 195, 106848 https://doi.org/10.1016/j.compag.2022.106848.

Li, N., Ren, Z., Li, D., Zeng, L., 2020. Review: Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming. Animal 14 (3), 617–625. https://doi.org/10.1017/S1751731119002155.

Luiten, J., Osep, A., 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. https://doi.org/10.1007/s11263-020-01375-2.

Martinez-Ortiz, C. A., Everson, R. M., and Mottram, T. (2013). Video tracking of dairy cows for assessing mobility scores. https://ore.exeter.ac.uk/repository/handle/10871/13481.

Meunier, B., Pradel, P., Sloth, K.H., Cirié, C., Delval, E., Mialon, M.M., Veissier, I., 2018. Image analysis to refine measurements of dairy cow behaviour from a real-time location system. Biosyst. Eng. 173, 32–44. https://doi.org/10.1016/j.biosystemseng.2017.08.019.

Qiao, Y., Kong, H., Clark, C., Lomax, S., Su, D., Eiffert, S., Sukkarieh, S., 2021. Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation. Comput. Electron. Agric. 185, 106143 https://doi.org/10.1016/j.compag.2021.106143.

Salau, J., Krieter, J., 2020. Analysing the space-usage-pattern of a cow herd using video surveillance and automated motion detection. Biosyst. Eng. 197, 122–134. https://doi.org/10.1016/j.biosystemseng.2020.06.015.

Sih, A., Bell, A.M., Johnson, J.C., Ziemba, R.E., 2004. Behavioral Syndromes: An Integrative Overview. Q. Rev. Biol. 79 (3), 241–277. https://doi.org/10.1086/422893.

Sundararaman, R., De Almeida Braga, C., Marchand, E., Pettre, J., 2021. Tracking Pedestrian Heads in Dense Crowd. In: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3865–3875. https://doi.org/10.1109/CVPR46437.2021.00386.

Ter-Sarkisov, A., Ross, R., Kelleher, J., 2017. Bootstrapping Labelled Dataset Construction for Cow Tracking and Behavior Analysis. In: 2017 14th Conference on Computer and Robot Vision (CRV), pp. 277–284. https://doi.org/10.1109/CRV.2017.25.

Wang, G., Gu, R., Liu, Z., Hu, W., Song, M., Hwang, J.-N., 2021. Track Without Appearance: Learn Box and Tracklet Embedding With Local and Global Motion Patterns for Vehicle Tracking. In: In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9876–9886. https://doi.org/10.1109/ICCV48922.2021.00973.

Weng, Z., Meng, F., Liu, S., Zhang, Y., Zheng, Z., Gong, C., 2022. Cattle face recognition based on a Two-Branch convolutional neural network. Comput. Electron. Agric. 196, 106871 https://doi.org/10.1016/j.compag.2022.106871.

Wojke, N., Bewley, A., and Paulus, D. (2017). Simple online and real-time tracking with a deep association metric. *2017 IEEE International Conference on Image Processing (ICIP)*, 3645–3649. doi: 10.1109/ICIP.2017.8296962.

Xu, B., Wang, W., Guo, L., Chen, G., Li, Y., Cao, Z., Wu, S., 2022. CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss. Comput. Electron. Agric. 193, 106675 https://doi.org/10.1016/j.compag.2021.106675.

Zagoruyko, S., and Komodakis, N. (2017). Wide Residual Networks. *arHiv preprint arHiv: 1605.07146.* doi: 10.5244/C.30.87.

Zambelis, A., Saadati, M., Dallago, G.M., Stecko, P., Boyer, V., Parent, J.-P., Pedersoli, M., Vasseur, E., 2021. Automation of video-based location tracking tool for dairy cows in their housing stalls using deep learning. Smart Agricultural Technology 1, 100015. https://doi.org/10.1016/j.atech.2021.100015.

Zhang, K., Zhang, L., and Yang, M.-H. (2012). Real-Time Compressive Tracking. *Computer Vision – ECCV 2012*, pp. 864–877. doi: 10.1007/978-3-642-33712-3_62.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., ... & Wang, X. (2022). Bytetrack: Multi-object tracking by associating every detection box. *In European Conference on Computer Vision, Springer, Cham,* pp. 1-21. doi: 10.1007/978-3-031-20047-2_1.

Zin, T.T., Pwint, M.Z., Seint, P.T., Thant, S., Misawa, S., Sumi, K., Yoshida, K., 2020. Automatic Cow Location Tracking System Using Ear Tag Visual Analysis. Sensors 20 (12), 3564. https://doi.org/10.3390/s20123564.

Zou, Z., Huang, J., Luo, P., 2022. Compensation Tracker: Reprocessing Lost Object for Multi-Object Tracking. In: In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 307–317. https://doi.org/10.1109/WACV51458.2022.00273.