IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

Enhanced DeepSORT and StrongSORT for Multicattle Tracking with Optimized Detection and Re-identification

Hyeon-seok Sim[1], and Hyun-chong Cho[1, 2], Member, IEEE

[1]Department Graduate Program for BIT Medical Convergence, Kangwon National University, Chuncheon-si, Republic of Korea
[2]Department of Electronics Engineering and Department of Data Science, Kangwon National University, Chuncheon-si, Republic of Korea

Corresponding author: Hyun-chong Cho (hyuncho@kangwon.ac.kr)

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2022R1I1A3053872) and this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00242528).

The animal study was approved by the Gangwon State Livestock Research Institute in South Korea on 21 December 2023, with approval number 2023-6.

ABSTRACT Recently, labor shortages in the farming industry have increased the demand for automation. Object tracking technology has emerged as a critical tool for monitoring livestock through automated systems. This study focuses on tracking individual cattle using object detection and tracking algorithms. Data were collected noninvasively using cameras, and a tracking-by-detection (TBD) approach was adopted. The proposed framework introduces multiple enhancements optimized for cattle tracking. These enhancements include a comparison of five different bounding box regression losses to improve detection accuracy, modifications to the Kalman filter state vector for more accurate bounding box predictions, and adjustments to the feature vector distance metric in the re-identification algorithm. YOLOv9-t was used as the detector, whereas DeepSORT and StrongSORT served as trackers. Compared with the baseline, which uses DeepSORT, the proposed method achieved significant improvements in higher-order tracking accuracy (HOTA) by 4.1%, multiple object tracking accuracy (MOTA) by 1.08%, and identification F1 score (IDF1) by 5.12%, reaching values of 78.64%, 90.29%, and 91.41%, respectively, while reducing the number of ID switches (IDSW).

INDEX TERMS Cattle tracking, DeepSORT, Multi-object tracking, StrongSORT, YOLOv9-t

I. INTRODUCTION

Cattle tracking plays a crucial role in the livestock industry [1]. Farmers can assess cattle health by continuously monitoring their behavior and movement through object tracking [2]. This monitoring enables early disease detection, allowing for swift intervention [3], [4]. Furthermore, object tracking enhances farm efficiency [5]. Research has demonstrated that using various methods, object tracking can individually identify and track each animal within its breeding environment [6], [7]. This capability allows for the automation of livestock monitoring, which, in turn, reduces labor costs and improves the overall efficiency of farm management [8]. Additionally, real-time tracking of individual cattle locations facilitates the analysis of behavioral patterns and space utilization, further optimizing farm operations [9].

According to statistics on the number of cattle farms and cattle population released by the Korea Statistical Information Service, as shown in Fig. 1, cattle population is increasing, while the number of farms is decreasing [10]. As the number of farms decreases while cattle populations increase, farm managers face challenges in maintaining the health and productivity of larger herds. Traditional methods, such as manual observation, are time-consuming, making them less viable for large-scale operations. This underscores the importance of automated systems capable of effectively monitoring cattle health and behavior, ensuring that farms can adapt to these demographic shifts [11], [12]. Consequently, object tracking technology is gaining significance. Automated systems utilizing object tracking facilitate the monitoring of individual cattle health and location with fewer personnel, thereby enabling efficient farm management even with limited manpower [13].

Effective cattle tracking systems significantly enhance farm efficiency and animal welfare [14], [15]. For example, real-time monitoring of cattle can optimize feeding schedules by analyzing movement patterns and the time spent in feeding areas. This ensures efficient use of resources such as feed and water, thereby reducing operational costs. Such systems contribute to the long-term sustainability and profitability of livestock operations.

Several studies have applied sensor-based approaches to object tracking [16]. These studies primarily used wearable devices such as neck-mounted sensors and ear tags [17]. Wearable devices incorporate embedded technologies such as accelerometers, radio frequency identification (RFID), and global positioning systems (GPS) to monitor the activity and movement of cattle [18]-[20]. These approaches involve attaching a sensor to each cow, facilitating the high-accuracy tracking of individual animals.

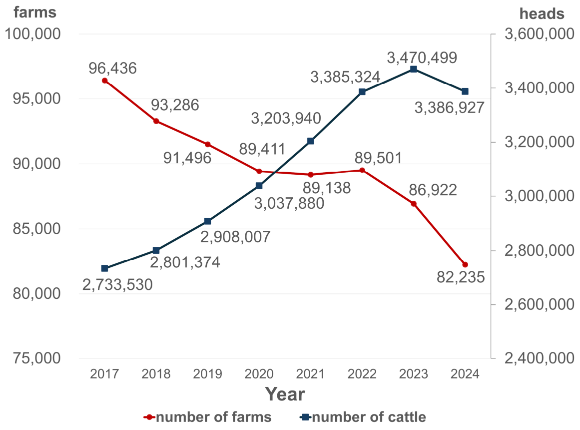

FIGURE 1. Annual trends in the number of farms and cattle in South Korea.

However, these wearable devices must be directly attached to the cattle, which can be invasive and potentially cause stress to the animals [21]. In addition, wearable devices are susceptible to damage or breakage owing to the activities of the cattle, and their battery life is limited. This situation necessitates regular replacement of both batteries and devices. On large-scale farms, managing numerous devices can significantly increase maintenance costs and labor requirements [22]-[24]. Furthermore, if a device's battery is depleted, temporary gaps in data collection may occur, disrupting the continuity of monitoring.

Recent advancements in computer vision technology have significantly increased interest in video-based object tracking [25]. Zheng et al. [26] proposed YOLO-BYTE for tracking cattle, utilizing YOLOv7 as the detector and ByteTrack as the tracker. Their approach achieved a higher-order tracking accuracy (HOTA) of 67.6%, multiple object tracking accuracy (MOTA) of 75.4%, and identification F1 score (IDF1) of 77.8%. Similarly, Li et al. [23] employed YOLOv8n and DeepSORT for cattle tracking, and their proposed framework achieved 92.1% MOTA and 81.1% IDF1. Tan et al. [27] introduced the SY-Track algorithm, employing YOLOv7-tiny and StrongSORT, and reported 55.59% HOTA, 84.33% MOTA, and 67.95% IDF1. Collectively, these studies demonstrate the increasing research efforts in video-based object tracking for livestock in recent years.

Building on this foundation, other researchers have explored various techniques. Mar et al. [28], [29] proposed a hybrid approach for cow detection and tracking, utilizing image processing techniques and deep learning methods, and later introduced a multi-feature tracking algorithm that demonstrated robust performance even with limited numbers of cows in their experimental setups. Aye et al. [30] utilized deep learning-based algorithms for tracking black cows, validating the system's performance on small datasets and demonstrating its applicability in farm environments.

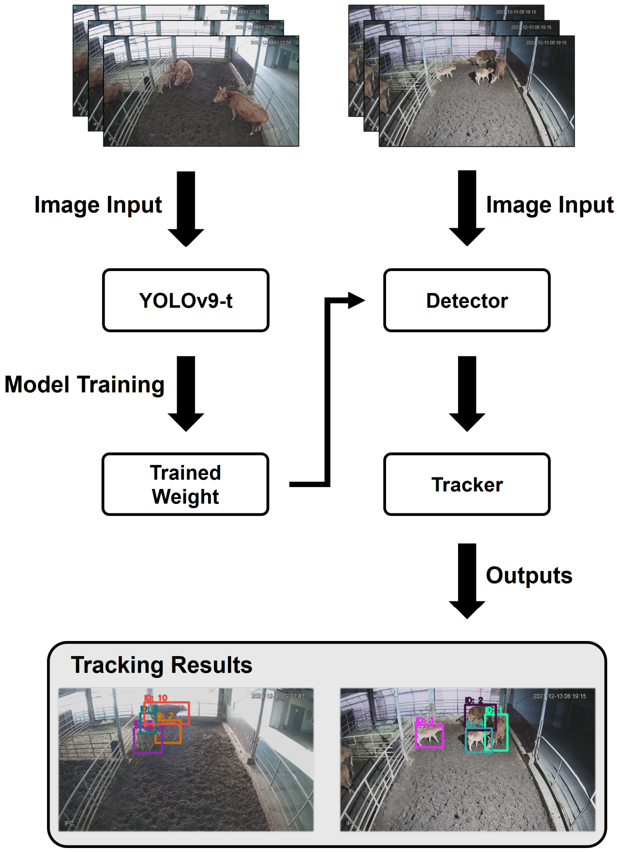

FIGURE 2. Workflow of the proposed cattle tracking system.

Video-based object tracking uses data collected from cameras to analyze the movements of individual cattle. This method offers the advantage of real-time monitoring of cattle movement without the need for wearable devices. Additionally, because it relies solely on video footage for tracking, it is noninvasive [31].

Object tracking algorithms primarily utilize a tracking-by-detection (TBD) approach and can be categorized into online and offline tracking methods [32]. Offline tracking processes an entire video dataset by simultaneously assigning IDs to all frames. This approach offers high accuracy and demonstrates robust performance, even in scenarios involving occlusion or complex environments. However, because it processes all the frames simultaneously, achieving real-time tracking can be challenging. In contrast, online tracking detects objects on a frame-by-frame basis and assigns matching IDs to the same object across consecutive frames. This enables real-time object tracking and facilitates rapid tracking, even in complex environments.

In this study, we adopted an online tracking approach that facilitates real-time cattle tracking using video data collected from RGB cameras. However, a drawback of the online tracking method is that its performance may deteriorate if missed detections or false positives occur during the detection stage [33]. To mitigate this issue, we propose a modified detection model and tracker specifically optimized for cattle tracking. The overall flow of the cattle tracking process is illustrated in Fig. 2.

II. RESEARCH METHOD

The aim of this study was to track cattle using the TBD approach. The tracking process consisted of three main stages. First, a cattle dataset was collected and organized to train the object detection model for tracking purposes. Next, a deep learning-based object detection model was employed to detect the cattle. Finally, based on the detection results, object tracking was performed by assigning IDs to the detected objects within the images. In this research, to enhance the performance of the DeepSORT and StrongSORT frameworks for multi-cattle tracking, we introduced several key modifications. First, the bounding box regression loss was refined to improve detection accuracy, particularly for objects of varying scales such as cows and calves. Second, we adjusted the state vector of the Kalman filter to more directly model the motion of cattle based on bounding box information. Finally, we customized the feature vector distance metric in the re-identification (Re-ID) algorithm to improve differentiation between cattle with similar visual appearances, thereby enhancing ID consistency. The following sections describe the research process and propose an optimized method for cattle tracking.

A. DATA COLLECTION AND DATASET COMPOSITION

CCTV cameras were installed in a research pen at the Gangwon State Livestock Research Institute in Hoengseong, Gangwon State, South Korea. The pen housed two cows and two calves. The data used in this study were collected over a 13-day period, from December 2 to December 14, 2021. Network IP cameras (GB-CDX04, GASI) were used for data collection, continuously recording at HD resolution (1280 × 720) for 24 h. The cameras were positioned 3 m high at a 45-degree angle to ensure comprehensive coverage of the 4.8 m × 9.6 m pen accommodating the four cattle. The camera setup was designed such that the center of the field of view aligned with the center of the research pen, ensuring accurate tracking and modeling of cattle movements within the pen. A schematic of the pen and sample images from the collected data are presented in Fig. 3.

To enhance the reliability of the experiment, the training, validation, and test sets were composed of data collected on different days. Data collected from December 2 to 7 were allocated to the training set, data from December 8 to 9 were used for the validation set, and data from December 10 to 11 were assigned to the test set. A total of 3,126 images were used to train and evaluate the object detection model. The detailed composition of the dataset is presented in Table I.

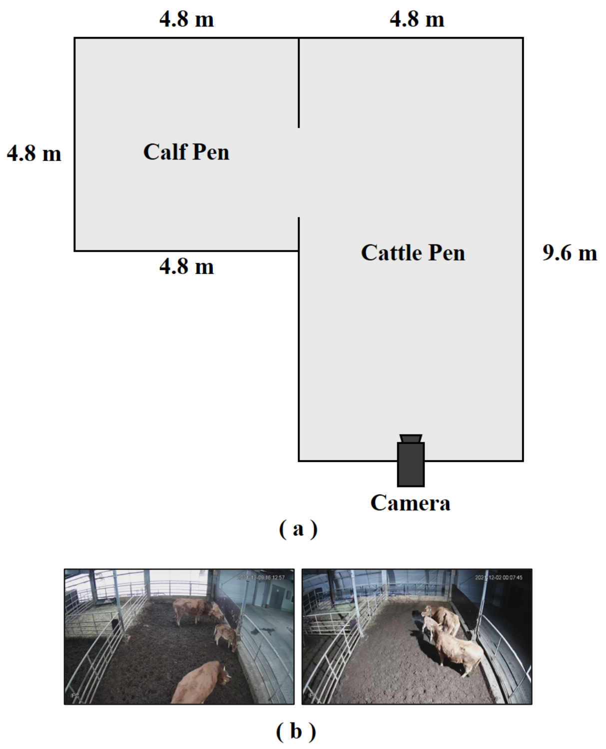

FIGURE 3. (a) Layout of the research pen used for data collection. (b) Sample images from the collected cattle tracking dataset.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3535092

TABLE I
DATASET COMPOSITION FOR CATTLE DETECTION

Task	Images	Instances
Train	2,056	8,125
Validation	559	2,524
Test	511	2,133
Total	3,126	12,782

The clips used to evaluate the performance of the object tracker were collected from December 13 to 14, using data from dates different from those in the detection phase. To enhance the reliability of the tracker performance evaluation, eight clips featuring frequent object occlusions and significant movement during the collection period were selected. Each clip was recorded at a rate of 15 frames per second for a duration of 10 s.

B. CATTLE DETECTION USING YOLOv9-t

YOLOv9 [34] is a real-time object detection model designed to balance lightweight architecture with high performance. It introduces programmable gradient information (PGI) and a generalized efficient layer aggregation network (GELAN) to enhance performance compared with previous models. The PGI, introduced in YOLOv9, employs an auxiliary reversible branch to mitigate information loss caused by bottlenecks during training. This approach enables the model to effectively learn multi-scale features without sacrificing critical information from the input data. Because PGI is used only during the training phase, it does not incur any overhead during inference, thus maintaining a high inference speed.

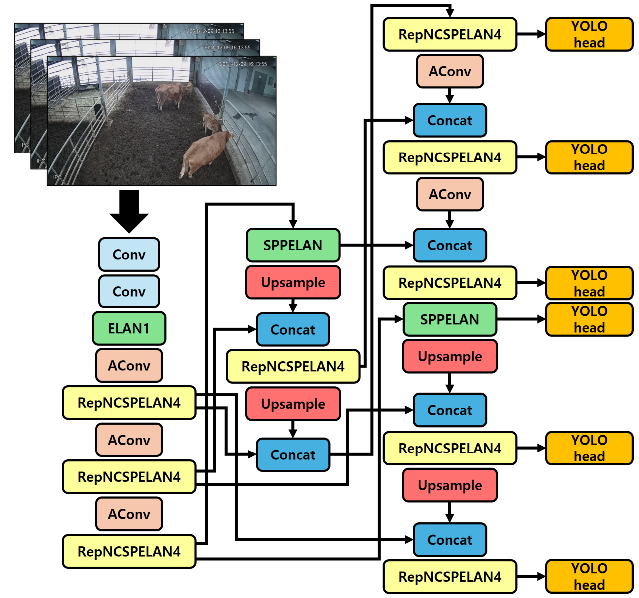

FIGURE 4. Architecture of the YOLOv9-t model used for cattle detection.

GELAN integrates elements from CSPNet [35] and ELAN [36], resulting in a more compact model that delivers rapid inference while maintaining high detection accuracy.

YOLOv9 is available in several versions based on model size, including t, s, m, and e, with smaller models offering faster inference speeds. In this study, a high detection speed is essential for real-time cattle object tracking. Therefore, the smallest and fastest version of YOLOv9, specifically YOLOv9-t, was used. The structure of YOLOv9-t is illustrated in Fig. 4.

1) BOUNDING BOX REGRESSION LOSS

In this study, we compared the loss functions used for bounding box regression in the YOLOv9-t model. Five intersection over union (IoU)-based loss functions were applied to the YOLOv9-t model: generalized-IoU (GIoU) [37], distance-IoU (DIoU) [38], complete-IoU (CIoU) [38], Scylla-IoU (SIoU) [39], and wise-IoU (WIoU) [40].

First, GIoU extends the traditional IoU by introducing a penalty term for instances where the two bounding boxes do not overlap. This penalty term is derived from the unoccupied area of a global box 'C' that encompasses both bounding boxes.

DIoU builds upon GIoU by taking into account the distance between the centers of the two bounding boxes. It incorporates a penalty term based on the squared distance between the centers ρ^2, divided by the squared diagonal length of the global box C (c^2).

CIoU, which is the default loss function used in YOLOv9, further improves DIoU by incorporating the aspect ratio of the bounding boxes. It evaluates the differences in width and height ratios between the predicted box and the ground truth (gt) box, adding a penalty term αv that reflects this difference. In this penalty term αv, v represents the aspect ratio difference between the predicted box and ground truth box, and α is a weighting factor that regulates the influence of this difference.

SIoU introduces additional penalty terms for angle, distance, and shape costs, in conjunction with the standard IoU loss. It sums the average of the distance cost Δ and the shape cost Ω, with the distance cost Δ also accounting for the angle cost.

WIoU is a dynamic loss function based on IoU that adjusts the IoU loss by weighting it according to the distance between the centers of the predicted and target bounding boxes, as well as the size of the global box. It applies an exponential function to the squared distance between the center points, divided by the sum of the squared width (W_g) and height (H_g) of the global box. The following section presents the formulae for each loss function.

$$L_{IoU} = 1 - IoU \tag{1}$$

$$L_{GIoU} = 1 - IoU + \frac{|C - (B \cup B^{gt})|}{|C|} \tag{2}$$

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3535092

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \tag{3}$$

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{4}$$

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{5}$$

$$L_{WIoU} = \exp\left(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{W_g^2 + H_g^2}\right) L_{IoU} \tag{6}$$

C. MULTI OBJECT TRACKING

In this study, the multi-object tracking of cattle was conducted using object detection techniques. The tracking algorithms employed were the online trackers DeepSORT [41] and StrongSORT [42]. DeepSORT and StrongSORT algorithms track object motion by modeling object positions based on bounding boxes and assigning IDs within the angle. To further enhance tracking performance, we proposed modifying the state vector of the Kalman filter and adjusting the distance metric used for comparing feature vectors during the matching stage. The following section provides a detailed explanation of the methodology implemented to improve performance.

1) OBJECT TRACKER FOR CATTLE TRACKING

DeepSORT [41] and StrongSORT [42] are extensions of the simple online and realtime tracking (SORT) [43] algorithm. SORT employs a Kalman filter to predict object positions and the Hungarian algorithm to associate objects across frames. Although SORT provides high tracking speed, it has limitations in handling occlusions, making it susceptible to errors when objects are temporarily obstructed or overlapped.

DeepSORT addresses the limitations of SORT by incorporating a deep-learning-based re-identification model. This re-identification model extracts features from the objects detected in each frame, generating feature vectors that facilitate the comparison of object similarities across frames. When a high degree of similarity is identified, the same ID is assigned to the object, thereby enhancing the continuity of ID tracking. This capability allows DeepSORT to accurately match objects between frames, even in instances of occlusion or when an object temporarily disappears and reappears. These improvements render DeepSORT more robust against occlusions and significantly enhance tracking performance.

StrongSORT is a further advancement of DeepSORT. Whereas DeepSORT relies solely on appearance information for matching, StrongSORT incorporates both appearance and motion information. Additionally, DeepSORT uses a matching cascade algorithm, whereas StrongSORT improves the matching accuracy by employing a simpler global linear assignment during the matching phase. Moreover, StrongSORT incorporates the NSA Kalman filter introduced in GIAOTracker [44], thereby improving the original Kalman filter used in DeepSORT. The NSA Kalman filter adaptively calculates the noise covariance $\widetilde{R_k}$, making it more resilient to low-quality detections and noise. By employing these enhancements, StrongSORT achieves superior tracking performance.

In this study, both DeepSORT and StrongSORT were used as tracker algorithms to evaluate the proposed methods for improving cattle tracking performance. Fig. 5 illustrates the operational process of DeepSORT, which served as the baseline tracker in this study.

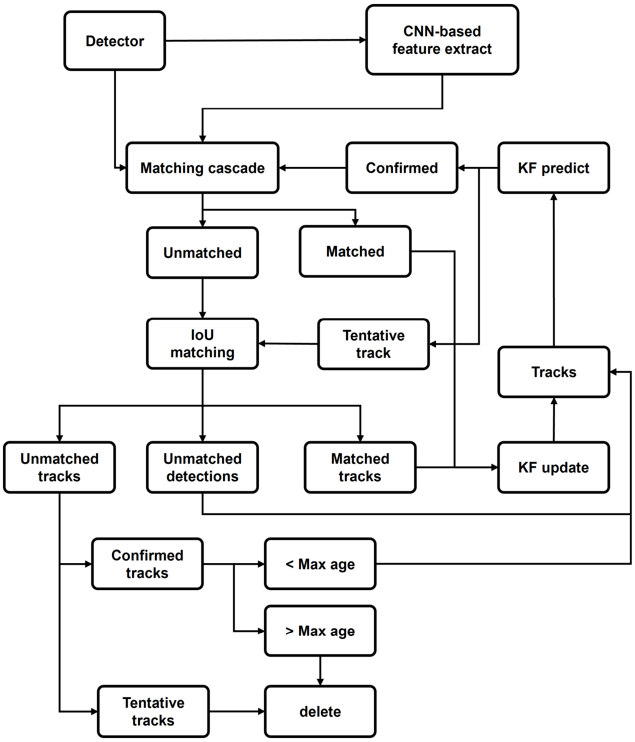

FIGURE 5. Architecture of the DeepSORT Object Tracking Algorithm.

2) KALMAN FILTER FOR CATTLE TRACKING

In this study, both DeepSORT and StrongSORT use a Kalman filter to predict the positions of objects in the subsequent frame. The Kalman filter functions by alternating between "state prediction" and "state update," continuously estimating the object's position and velocity. The basic operation of the Kalman filter in DeepSORT and StrongSORT is outlined as follows.

First, state vector $\mathbf{x_0}$ and covariance matrix P_0 are defined. The state vector ($\mathbf{x_0}$) includes the center coordinates of the bounding box (x, y), aspect ratio and height (a, h), as well as their respective velocity components (vx, vy, va, vh), which represent the rate of change over time for each element.

$$\mathbf{x_0} = [x, y, a, h, vx, vy, va, vh]^T \tag{7}$$

$$P_0 = diag\left([\sigma_x^2, \sigma_y^2, \sigma_a^2, \sigma_h^2, \sigma_{vx}^2, \sigma_{vy}^2, \sigma_{va}^2, \sigma_{vh}^2]\right) \tag{8}$$

In the prediction stage, the next state of the object is predicted based on its current state, and the covariance matrix is updated accordingly. The prediction is carried out using the state transition matrix F, which models the evolution of the state of the object over time. This matrix is applied to the current state vector to estimate the position and velocity of the object in the subsequent frame. In addition, the covariance matrix is updated to account for the uncertainty in the predicted state.

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & dt & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & dt & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & dt & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & dt \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

$$x_{k|k-1} = F x_{k-1|k-1} + B u_k \tag{10}$$

$$P_{k|k-1} = F P_{k-1|k-1} F^T + Q \tag{11}$$

In the update stage, the predicted state is used to estimate the observed values, and the Kalman gain K_k is calculated. Subsequently, both the state vector and the covariance matrix are updated. The Kalman gain determines the weight assigned to the observed values relative to the predicted state. The observation matrix H maps the predicted state to the observed measurements.

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{12}$$

$$z_{k|k-1} = H x_{k|k-1} \tag{13}$$

$$K_k = P_{k|k-1} H^T \left(H P_{k|k-1} H^T + R \right)^{-1} \tag{14}$$

$$x_{k|k} = x_{k|k-1} + K_k \left(z_k - H x_{k|k-1} \right) \tag{15}$$

$$P_{k|k} = (I - K_k H) P_{k|k-1} \tag{16}$$

In the Kalman filter of the original tracker, the aspect ratio (a) is used to account for changes in the shape of the bounding box. However, in real farm environments, when the cattle rotate, the aspect ratio may not accurately reflect changes in the state of the object. To address this limitation, we replaced the aspect ratio with the bounding box width (w). The updated state vector $\mathbf{x}_0'$ and covariance matrix P_0' are defined as follows.

$$\mathbf{x}_0' = [x, y, w, h, vx, vy, vw, vh]^T \tag{17}$$

$$P_0' = diag\left(\left[\sigma_x^2, \sigma_y^2, \sigma_w^2, \sigma_h^2, \sigma_{vx}^2, \sigma_{vy}^2, \sigma_{vw}^2, \sigma_{vh}^2 \right] \right) \tag{18}$$

By incorporating the bounding box width (w) into the state vector, the Kalman filter can more intuitively reflect the bounding box elements during the prediction process. This integration facilitates a more direct modeling of the effects associated with changes in the state of cattle. Consequently, this approach can improve the performance of tracking algorithms. The effectiveness of this approach was validated through experiments, and the results are discussed in Section III.C.

3) IMPROVED FEATURE VECTOR DISTANCE METRIC

The re-identification algorithms in DeepSORT and StrongSORT use a convolutional neural network (CNN) model to extract feature vectors for the objects in each frame. The similarity between these feature vectors is then assessed to maintain the object's ID even when it is temporarily occluded or moves offscreen and subsequently reappears. In this process, the cosine distance metric is typically employed to compare the similarities between the feature vectors. The cosine distance metric is calculated as follows:

$$\text{cosine similarity}(\mathbf{A}, \boldsymbol{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \, \|\mathbf{B}\|} \tag{19}$$

$$\text{cosine distance}(\mathbf{A}, \mathbf{B}) = 1 - \text{cosine similarity}(\boldsymbol{A}, \boldsymbol{B}) \tag{20}$$

A and **B** represent the two feature vectors being compared, and $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ denote the norms of these vectors. The cosine distance metric considers only the directions of the two feature vectors. Thus, even if the magnitudes of the feature vectors differ, they can still be evaluated as highly similar, which may not accurately reflect the true characteristics of the objects. Significant noise can occur in actual farm environments due to variations in lighting and environmental conditions. Under such noisy conditions, cosine distance may not function effectively.

To overcome these limitations, we propose using the Pearson correlation as the distance metric for comparing re-identification feature vectors. The Pearson correlation measures the linear correlation between two vectors and evaluates their variability relative to their mean values. The Pearson correlation coefficient and the corresponding distance are calculated as follows:

$$\text{Pearson correlation}(\mathbf{A}, \mathbf{B}) = \frac{\sum (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum (a_i - \bar{a})^2} \sqrt{\sum (b_i - \bar{b})}^2} \tag{21}$$

$$\text{Pearson distance}(\mathbf{A}, \mathbf{B}) = 1 - \text{pearson correlation}(\mathbf{A}, \mathbf{B}) \tag{22}$$

a_i and b_i represent the individual elements of each vector, and $\bar{a}$ and $\bar{b}$ denote the mean values of the vectors. The Pearson correlation evaluates the variability of each vector's elements relative to their mean, focusing on changes in patterns and correlations between objects. Therefore, it is better suited for tracking changes in object patterns, even when the size or shape of the cattle changes drastically in real-world environments. Additionally, the Pearson correlation is less sensitive to noise in images, enabling more stable similarity evaluations. This means that even if the size or shape of the cattle is distorted owing to occlusion or other noise, the overall correlation remains relatively unaffected. Therefore, the Pearson distance metric can assess correlations more accurately than the cosine distance metric, thus enhancing the continuity of object ID assignment. In addition, the use of Pearson correlation improves the robustness of re-identification particularly in scenarios where visual similarity between objects like cattle poses a significant challenge.

III. RESULTS

A. EVALUATION METRIC

Performance evaluations were conducted at each stage to demonstrate the effectiveness of the proposed method. In the object detection phase, the evaluation focused on the accuracy of the model in detecting cattle. In the object tracking phase, the evaluation focused on the consistency and accuracy of object tracking. The following sections describe the evaluation metrics employed at each stage.

1) CATTLE DETECTION EVALUATION METRIC

The cattle detection performance of the proposed model was evaluated using precision, recall, and mean average precision (mAP). Precision refers to the proportion of correct predictions out of all predictions made by the model, and recall indicates the proportion of actual positive instances that were correctly predicted by the model. The curve that illustrates the relationship between precision values and varying recall is known as the precision-recall (P-R) curve, and the area under this curve is defined as the average precision (AP). mAP is the average AP across all the classes. mAP is calculated for objects with an IoU of 0.5 or higher. The formulas for these metrics are as follows:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{23}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{24}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{25}$$

2) OBJECT TRACKING EVALUATION METRIC

The evaluation metrics commonly used for object tracking include HOTA [45], MOTA [46], and IDF1 [47]. These metrics assess different aspects of the tracking performance. In the formulas for the object tracking evaluation metrics, false negatives (FN) indicate missed detections where the model failed to predict a bounding box for an object, false positives (FP) indicate incorrect bounding box predictions, and true positives (TP) indicate correctly detected objects. HOTA is a comprehensive metric that evaluates both detection and association accuracy. It combines detection accuracy (DetA) and association accuracy (AssA) using harmonic means. HOTA focuses on balancing correct object detection with the accurate maintenance of object identities across frames. In HOTA, true positives for associations (TPA) indicate correctly associated object identities across frames, false negatives for associations (FNA) denote cases where an object's identity is not correctly maintained, and false positives for associations (FPA) represent instances where an identity is incorrectly assigned to an object.

$$HOTA = \sqrt{DetA \cdot AssA} \tag{26}$$

$$DetA = \frac{TP}{TP + FN + FP} \tag{27}$$

$$AssA = \frac{1}{TP}\sum_{c\in\{TP\}} A(c) \tag{28}$$

$$A(c) = \frac{TPA(c)}{TPA(c) + FNA(c) + FPA(c)} \tag{29}$$

MOTA is a comprehensive metric used to evaluate overall tracking performance, taking into account both object detection and ID consistency. In MOTA, ground-truth detection (gtDet) refers to the number of actual objects. ID switches (IDSW) represent the number of times an object's ID changes during tracking, indicating errors in maintaining consistent ID assignments.

$$MOTA = 1 - \frac{FN + FP + IDSW}{gtDet} \tag{30}$$

IDF1 is a metric used to evaluate the consistency of ID assignment during the tracking process. It measures how effectively the tracking algorithm maintains the correct identity of objects over time. In IDF1, ID true positive (IDTP) refers to the number of instances in which a correct ID is consistently assigned to an object. ID false positive (IDFP) represents the number of times an incorrect ID is assigned to an object. ID false negative (IDFN) refers to instances in which an object is not assigned an ID when it should have been.

$$DF1 = 2 \times \frac{DP \times DR}{DP + DR} \tag{31}$$

$$DP = \frac{DTP}{DTP + DFP} \tag{32}$$

$$DR = \frac{DTP}{DTP + DFN} \tag{33}$$

B. OBJECT DETECTION RESULTS

In this study, cattle detection was performed using the YOLOv9-t object detection model. YOLOv9-t was trained on 2,056 images with 8,125 labeled instances. To enhance the cattle detection performance, five different bounding box regression loss functions were applied, and the detection performance of each loss function was compared. The results of the detection performance comparison for each loss function are presented in Table II.

TABLE II
PERFORMANCE COMPARISON OF YOLOV9-T CATTLE DETECTION USING DIFFERENT BOUNDING BOX REGRESSION LOSS FUNCTIONS

Loss	Precision	Recall	mAP
GIoU	95.8	87.7	94.1
DIoU	95.8	90.7	95.3
CIoU	95.9	89.5	94.9
SIoU	95.6	89.3	94.7
WIoU	95.6	88.1	94.6

In the performance evaluation of 511 test images, DIoU outperformed other more advanced loss functions, including CIoU, which incorporates a penalty term for aspect ratio. DIoU achieved an mAP of 95.3%. This result is presumed to be due to the relatively small variation in the aspect ratio of cattle, compared to other objects. Since CIoU penalizes changes in aspect ratio, it may not significantly improve detection accuracy in cattle detection, where the shape and size of the cattle are relatively uniform. In contrast, DIoU focuses on minimizing the Euclidean distance between the centers of the predicted and ground truth bounding boxes. This approach appears to enhance detection performance in cattle detection, where objects have similar shapes.

The effect of modifying the bounding box regression loss for cattle detection on the object tracking process was assessed by comparing CIoU, the original loss function used in the YOLOv9-t model, with DIoU, which demonstrated the highest cattle detection performance. This comparison aimed to ascertain how the improved detection performance of DIoU influenced the overall tracking accuracy and consistency during the object tracking phase.

C. OBJECT TRACKING RESULTS

In this study, object tracking was performed using DeepSORT and StrongSORT algorithms. Performance evaluation was conducted on eight sequences, each consisting of a 10-second video with 150 frames. The tracker was evaluated across four scenarios to validate the effectiveness of the proposed method. All performance evaluations were conducted using the same dataset across all scenarios to ensure consistency and reliability in comparative analyses.

Original: Tracking performance without any modifications to the tracker.
Width: The Kalman filter state vector was modified by substituting the aspect ratio component of the bounding box with the width component.
Pearson: The feature vector distance metric in the re-identification algorithm was changed from the original cosine distance metric to the Pearson distance metric.

TABLE III
CATTLE TRACKING PERFORMANCE OF DEEPSORT
(↑ INDICATES THAT HIGHER VALUES REPRESENT BETTER PERFORMANCE, ↓ INDICATES THAT LOWER VALUES REPRESENT BETTER PERFORMANCE)

Detector Loss	Methods	HOTA (↑)	MOTA (↑)	DetA (↑)	AssA (↑)	IDSW (↓)	IDF1 (↑)
CIoU	original	74.54	89.21	73.92	76.61	18	86.29
	width	76.93	89.88	76.15	78.94	13	88.52
	Pearson	76.88	89.24	73.35	81.97	6	91.15
	width + Pearson	78.64	90.29	76.31	82.22	5	91.41
DIoU	original	74.77	89.15	74.15	77.13	19	86.92
	width	76.93	89.88	76.15	78.94	13	88.52
	Pearson	78.07	89.11	74.12	83.82	3	93.50
	width + Pearson	78.64	90.29	76.31	82.22	5	91.41

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3535092

Width + Pearson: Both proposed modifications (the Kalman filter state vector adjustment and the Pearson distance metric) were combined and implemented.

Each of these four scenarios was compared using two bounding box regression losses: the default CIoU loss of the YOLOv9-t model, and the DIoU loss, which demonstrated the highest performance in Section III.B. The test results for each scenario and loss function are listed in Tables III and IV, respectively.

1) DEEPSORT RESULTS

In the results presented in Table III for DeepSORT, the “width” setting using the CIoU loss demonstrated performance improvements across all metrics, including HOTA, MOTA, and IDF1. The “width” setting directly influences the positioning of the bounding box during the prediction process. This adjustment resulted in a 2.23% improvement in the DetA metric, and the overall performance metric HOTA increased by 2.39%, reaching 76.93%.

The application of the “Pearson” setting to the re-identification algorithm improves the ID continuity between frames. Consequently, the value of IDSW was significantly reduced to just 6, and AssA and IDF1 improved by 5.36% and 4.86%, respectively, achieving 81.97% and 91.45%, respectively. As a result, the HOTA levels increased by 2.34%, reaching 76.88%.

Finally, the “width + Pearson” setting integrated both proposed methods, resulting in improvements in both ID continuity and bounding box prediction performance. As a result, all major metrics demonstrated improvements, with HOTA, MOTA, and IDF1 reaching 78.64%, 90.29%, and 91.41%, respectively. These results underscore the effectiveness of the combined approach in improving the tracking performance.

In addition, when DIoU was used as the bounding box regression loss for the detector, performance improvements were observed in most metrics, with MOTA remaining consistent with that in the original setting. Similar to the results obtained with CIoU, applying the proposed methods to DIoU yielded a comparable pattern of performance improvement. The highest scores were recorded across all major metrics, reaching 78.64%, 90.29%, and 91.41%, respectively. This demonstrates that the proposed modifications effectively improve tracking performance, irrespective of the bounding box regression loss applied.

2) STRONGSORT RESULTS

The same scenarios evaluated in the DeepSORT tests were also assessed in the StrongSORT tests. As presented in Table IV, StrongSORT demonstrated superior baseline performance compared with DeepSORT. This improvement can be attributed to StrongSORT being an enhanced version of DeepSORT, featuring improvements in the re-identification algorithm and modifications to the Kalman filter.

In the StrongSORT results, for the “width” setting with CIoU loss, although the MOTA slightly decreased, performance improvements were observed in other metrics, including HOTA and IDF1. HOTA increased by 0.99%, reaching 79.52%, whereas IDF1 improved by 1.23%, reaching an accuracy of 92.60%.

For the “Pearson” setting, AssA improved, resulting in a 0.48% increase in HOTA to 79.01%. Additionally, the improved ID continuity contributed to a 1.41% increase in IDF1, reaching a value of 92.78%.

Lastly, for the “width + Pearson” setting, HOTA experienced the largest increase of 1.03%, reaching 79.56%, while IDF1 improved by 1.23%, reaching 92.60%, despite a slight decrease in MOTA.

When the DIoU loss was applied, the “original” setting maintained the same MOTA value while demonstrating improvements in other metrics. The results obtained with

TABLE IV
CATTLE TRACKING PERFORMANCE OF STRONGSORT
(↑ INDICATES THAT HIGHER VALUES REPRESENT BETTER PERFORMANCE, ↓ INDICATES THAT LOWER VALUES REPRESENT BETTER PERFORMANCE)

Detector Loss	Methods	HOTA (↑)	MOTA (↑)	DetA (↑)	AssA (↑)	IDSW (↓)	IDF1 (↑)
CIoU	original	78.53	89.56	75.02	83.38	8	91.37
	width	79.52	89.00	75.18	85.22	8	92.60
	Pearson	79.01	89.54	74.91	84.49	9	92.78
	width + Pearson	79.56	89.09	75.24	85.22	8	92.60
DIoU	original	78.58	89.52	75.38	83.15	5	91.37
	width	78.60	89.52	75.42	83.11	5	91.37
	Pearson	78.59	89.52	75.20	83.36	5	91.76
	width + Pearson	78.62	89.52	75.25	83.33	5	91.76

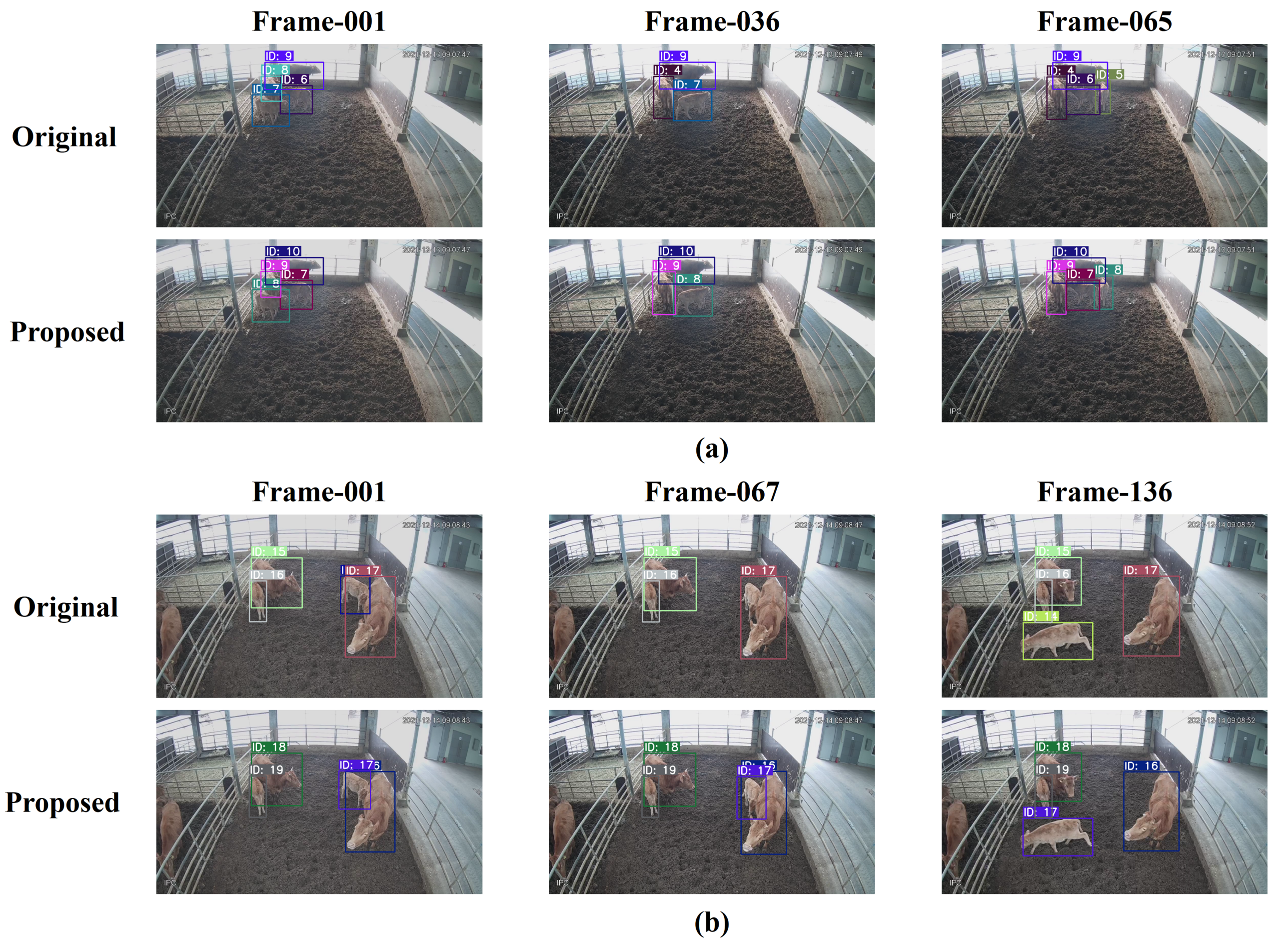

FIGURE 6. (a) Tracking results using the original and proposed cattle tracking system with DeepSORT for Clip 1. (b) Tracking results using the original and proposed cattle tracking system with DeepSORT for Clip 2.

DIoU exhibited a pattern similar to those observed with CIoU. In the “width + Pearson” setting with DIoU, HOTA increased by 0.04%, reaching 78.62%, MOTA remained consistent at 89.52%, and IDF1 improved by 0.39%, achieving 91.76%.

From Tables III and IV, the DIoU loss, which demonstrated high performance in YOLOv9-t, also contributed to overall improvements in DeepSORT and significantly lower IDSW in StrongSORT, thereby proving its effectiveness in cattle tracking. Furthermore, the proposed “width” and “Pearson” settings enhanced both detection and re-identification accuracy in DeepSORT and StrongSORT. The combination of these methods in the “width + Pearson” setting resulted in substantial improvements in cattle tracking performance for both algorithms. These results confirm that the proposed method yields significant gains in tracking performance for both DeepSORT and StrongSORT.

Fig. 6 presents the qualitative evaluation results of cattle tracking performed using the DeepSORT algorithm without any modifications to its parameters and settings, compared to the results obtained after applying the proposed method. In Fig. 6 (a), the original method exhibits ID switching when the object reappears in frame-036. In contrast, the proposed method maintains the same ID without any switching. Similarly, in Fig. 6 (b), the original method fails to detect objects during occlusion, resulting in ID switching, whereas the proposed method demonstrates improved detection and tracking accuracy under the same conditions. These results indicate that the proposed method enhances object detection performance and improves ID continuity between objects.

IV. CONCLUSIONS

In this study, cattle tracking was conducted based on the results of cattle detection. YOLOv9-t was used as the detector, and five different bounding box regression loss functions were compared to improve the cattle detection performance. Both DeepSORT and StrongSORT algorithms were employed for tracking. To optimize the cattle tracking performance of DeepSORT and StrongSORT, modifications were made to the Kalman filter state vector, and the feature vector distance metric was adjusted.

In this study, the "width" setting demonstrated significant performance improvements, particularly in bounding box prediction accuracy, while the "Pearson" setting notably enhanced ID continuity. The combination of these two approaches in the "width + Pearson" setting leveraged both advantages, resulting in a balanced improvement in bounding box prediction accuracy and ID consistency, ultimately achieving high overall performance. The proposed methods yielded substantial performance gains in both DeepSORT and StrongSORT, and the results of this study confirmed the effectiveness of these methods for cattle tracking.

The modifications to the DeepSORT and StrongSORT frameworks address critical challenges in multi-cattle tracking, such as handling size variability, accounting for the unique motion dynamics of cattle, and mitigating issues of visual similarity among individuals. The modification of the detector's bounding box regression loss has improved object detection performance, which, in turn, has enhanced overall tracking performance. Additionally, the adjustment of the Kalman filter's state vector in the tracker stage enables a more intuitive modeling of the movement of object bounding boxes, further boosting tracking performance. Lastly, the modification of the feature vector distance metric in the re-identification algorithm has improved ID continuity and re-identification performance, particularly for cattle with similar visual features. These advancements significantly enhance the applicability of our system to precision farming, where tracking accuracy and robustness are crucial for managing livestock in real-world environments.

This study has some limitations owing to the environment and dataset used in the experiments. Data were collected from only four cattle housed in a research pen under controlled conditions. Further research is necessary to evaluate the performance of the proposed cattle tracking system in diverse farm environments, particularly in larger pens with more cattle. Future studies should collect additional data from various farm environments to improve the generalization performance of the system.

Additionally, although this study improved re-identification performance to mitigate occlusion issues, relying on data from a single camera angle did not completely eliminate occlusions. Future research will implement a multi-camera system to associate cattle IDs detected from multiple angles, further enhancing tracking accuracy.

REFERENCES

[1] A. I. Awad, "From classical methods to animal biometrics: A review on cattle identification and tracking," *Comput. Electron. Agric.*, vol. 123, pp. 423-435, 2016.
[2] A. Huhtala, K. Suhonen, P. Mäkelä, M. Hakojärvi, and J. Ahokas, "Evaluation of instrumentation for cow positioning and tracking indoors," *Biosyst. Eng.*, vol. 96, no. 3, pp. 399-405, 2007.
[3] D. Y. Pavkin, A. S. Dorokhov, F. E. Vladimirov, I. M. Dovlatov, and K. S. Lyalin, "Algorithms for detecting cattle diseases at early stages and for making diagnoses and related recommendations," *Appl. Sci.*, vol. 11, no. 23, Art. no. 11148, 2021.
[4] C. C. Mar, T. T. Zin, P. Tin, K. Honkawa, I. Kobayashi, and Y. Horii, "Cow detection and tracking system utilizing multi-feature tracking algorithm," *Sci. Rep.*, vol. 13, no. 1, Art. no. 17423, 2023.
[5] W. Andrew, C. Greatwood, and T. Burghardt, "Visual localisation and individual identification of holstein friesian cattle via deep learning," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 2850-2859.
[6] O. Guzhva, H. Ardö, M. Nilsson, A. Herlin, and L. Tufvesson, "Now you see me: Convolutional neural network based tracker for dairy cows," *Front. Robot. AI*, vol. 5, Art. no. 107, 2018.
[7] S. Benaissa, F. A. M. Tuyttens, D. Plets, L. Martens, L. Vandaele, W. Joseph, and B. Sonck, "Improved cattle behaviour monitoring by combining ultra-wideband location and accelerometer data," *Animal*, vol. 17, no. 4, Art. no. 100730, 2023.
[8] B. Jiang, W. Tang, L. Cui, and X. Deng, "Precision livestock farming research: A global scientometric review," *Animals*, vol. 13, no. 13, Art. no. 2096, 2023.
[9] J. A. Hassan-Vásquez, F. Maroto-Molina, and J. E. Guerrero-Ginel, "GPS tracking to monitor the spatiotemporal dynamics of cattle behavior and their relationship with feces distribution," *Animals*, vol. 12, no. 18, Art. no. 2383, 2022.
[10] Statistics Korea, Livestock Products Quality Evaluation, "Survey on Livestock Trends," 2024. [Online]. Available: http://kostat.go.kr/
[11] F. M. de Oliveira, G. A. E. S. Ferraz, A. L. G. André, L. S. Santana, T. Norton, and P. F. P. Ferraz, "Digital and precision technologies in dairy cattle farming: A bibliometric analysis," *Animals*, vol. 14, no. 12, Art. no. 1832, 2024.
[12] M. Pastell, M. Kujala, A. M. Aisla, M. Hautala, V. Poikalainen, J. Praks, and J. Ahokas, "Detecting cow's lameness using force sensors," *Comput. Electron. Agric.*, vol. 64, no. 1, pp. 34-38, 2008.
[13] W. Luo, G. Zhang, Q. Yuan, Y. Zhao, H. Chen, J. Zhou, and Z. Yu, "High-precision tracking and positioning for monitoring holstein cattle," *PLoS One*, vol. 19, no. 5, Art. no. e0302277, 2024.
[14] L. Tong, J. Fang, X. Wang, and Y. Zhao, "Research on cattle behavior recognition and multi-object tracking algorithm based on YOLO-BoT," *Animals*, vol. 14, no. 20, Art. no. 2993, 2024.
[15] W. H. E. Mg, P. Tin, M. Aikawa, I. Kobayashi, Y. Horii, K. Honkawa, and T. T. Zin, "Customized tracking algorithm for robust cattle detection and tracking in occlusion environments," *Sensors (Basel, Switzerland)*, vol. 24, no. 4, 2024.
[16] A. da Silva Santos, V. W. C. de Medeiros, and G. E. Gonçalves, "Monitoring and classification of cattle behavior: A survey," *Smart Agric. Technol.*, vol. 3, Art. no. 100091, 2023.
[17] M. Lee and S. Seo, "Wearable wireless biosensor technology for monitoring cattle: A review," *Animals*, vol. 11, no. 10, Art. no. 2779, 2021.
[18] G. M. Pereira, B. J. Heins, and M. I. Endres, "Validation of an ear-tag accelerometer sensor to determine rumination, eating, and activity behaviors of grazing dairy cattle," *J. Dairy Sci.*, vol. 101, no. 3, pp. 2492-2495, 2018.
[19] L. R. Williams, D. R. Fox, G. J. Bishop-Hurley, and D. L. Swain, "Use of radio frequency identification (RFID) technology to record grazing beef cattle water point use," *Comput. Electron. Agric.*, vol. 156, pp. 193-202, 2019.
[20] L. A. González, G. J. Bishop-Hurley, R. N. Handcock, and C. Crossman, "Behavioral classification of data from collars containing motion sensors in grazing cattle," *Comput. Electron. Agric.*, vol. 110, pp. 91-102, 2015.
[21] A. Fuentes, S. Han, M. F. Nasir, J. Park, S. Yoon, and D. S. Park, "Multiview monitoring of individual cattle behavior based on action recognition in closed barns using deep learning," *Animals*, vol. 13, no. 12, Art. no. 2020, 2023.
[22] J. Li, Y. Liu, W. Zheng, X. Chen, Y. Ma, and L. Guo, "Monitoring cattle ruminating behavior based on an improved keypoint detection model," *Animals*, vol. 14, no. 12, Art. no. 1791, 2024.
[23] G. Li, J. Sun, M. Guan, S. Sun, G. Shi, and C. Zhu, "A new method for non-destructive identification and tracking of multi-object behaviors in beef cattle based on deep learning," *Animals*, vol. 14, no. 17, Art. no. 2464, 2024.
[24] H. Hu, B. Dai, W. Shen, X. Wei, J. Sun, R. Li, and Y. Zhang, "Cow identification based on fusion of deep parts features," *Biosyst. Eng.*, vol. 192, pp. 245-256, 2020.

[25] M. E. Hossain, M. A. Kabir, L. Zheng, D. L. Swain, S. McGrath, and J. Medway, “A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions,” *Artif. Intell. Agric.*, vol. 6, pp. 138-155, 2022.

[26] Z. Zheng, J. Li, and L. Qin, “YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows,” *Comput. Electron. Agric.*, vol. 209, Art. no. 107857, 2023.

[27] X. Tan, C. Yin, X. Li, M. Cai, W. Chen, Z. Liu, and Y. Han, “SY-Track: A tracking tool for measuring chicken flock activity level,” *Comput. Electron. Agric.*, vol. 217, Art. no. 108603, 2024.

[28] C. C. Mar, T. T. Zin, P. Tin, K. Honkawa, I. Kobayashi, and Y. Horii, “Cow detection and tracking system utilizing multi-feature tracking algorithm,” *Sci. Rep.*, vol. 13, no. 1, Art. no. 17423, 2023.

[29] C. C. Mar, T. T. Zin, I. Kobayashi, and Y. Horii, “A hybrid approach: Image processing techniques and deep learning method for cow detection and tracking system,” in *Proc. IEEE 4th Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2022, pp. 566–567.

[30] C. Aye, T. Zin, and I. Kobayashi, “Black cow tracking by using deep learning-based algorithms,” *ICIC Express Lett. Part B Appl.*, vol. 13, pp. 1313–1319, 2022.

[31] S. Myat Noe, T. T. Zin, P. Tin, and I. Kobayashi, “Comparing state-of-the-art deep learning algorithms for the automated detection and tracking of black cattle,” *Sensors*, vol. 23, no. 1, Art. no. 532, 2023.

[32] K. Ren, C. Hu, and H. Xi, “Rlm-tracking: Online multi-pedestrian tracking supported by relative location mapping,” *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 7, pp. 2881-2897, 2024.

[33] L. Vaquero, Y. Xu, X. Alameda-Pineda, V. M. Brea, and M. Mucientes, “Lost and found: Overcoming detector failures in online multi-object tracking,” *arXiv preprint arXiv:2407.10151*, 2024.

[34] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, “YOLOv9: Learning what you want to learn using programmable gradient information,” *arXiv preprint arXiv:2402.13616*, 2024.

[35] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, “CSPNet: A new backbone that can enhance learning capability of CNN,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 390-391.

[36] C. Y. Wang, H. Y. M. Liao, and I. H. Yeh, “Designing network design strategies through gradient path analysis,” *arXiv preprint arXiv:2211.04800*, 2022.

[37] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 658-666.

[38] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU loss: Faster and better learning for bounding box regression,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, pp. 12993-13000, 2020.

[39] Z. Gevorgyan, “SIoU loss: More powerful learning for bounding box regression,” *arXiv preprint arXiv:2205.12740*, 2022.

[40] Z. Tong, Y. Chen, Z. Xu, and R. Yu, “Wise-IoU: Bounding box regression loss with dynamic focusing mechanism,” *arXiv preprint arXiv:2301.10051*, 2023.

[41] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3645-3649.

[42] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong, and H. Meng, “Strongsort: Make deepsort great again,” *IEEE Trans. Multimedia*, vol. 25, pp. 8725-8737, 2023.

[43] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 3464-3468.

[44] Y. Du, J. Wan, Y. Zhao, B. Zhang, Z. Tong, and J. Dong, “Giaotracker: A comprehensive framework for MCMOT with global information and optimizing strategies in visdrone 2021,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2809-2819.

[45] J. Luiten et al., “HOTA: A higher order metric for evaluating multi-object tracking,” *Int. J. Comput. Vis.*, vol. 129, pp. 548-578, 2021.

[46] K. Bernardin and R. Stiefelhagen, “Evaluating multiple object tracking performance: The clear mot metrics,” *EURASIP J. Image Video Process.*, vol. 2008, Art. no. 1, 2008.

[47] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 17-35.

Hyeon-seok Sim is currently working towards his B.S. and M.S. degree at the Department of Electronics Engineering and Interdisciplinary Graduate Program for BIT Medical Convergence from Kangwon National University, South Korea.

Hyun-chong Cho (Member, IEEE) received his M.S. and Ph.D. degrees in electrical and computer engineering from the University of Florida, USA, in 2009. During 2010–2011, he was a Research Fellow at the University of Michigan, Ann Arbor, USA. From 2012 to 2013, he was a Chief Research Engineer at LG Electronics, South Korea. He is currently a Professor at the Department of Electronics Engineering, the Department of Data Science, and Interdisciplinary Graduate Program for BIT Medical, Kangwon National University, South Korea.