



Research Paper

Multi-camera fusion and bird-eye view location mapping for deep learning-based cattle behavior monitoring

Muhammad Fahad Nasir^{a,b,1}, Alvaro Fuentes^{a,b,1}, Shujie Han^{a,b}, Jiaqi Liu^{a,b}, Yongchae Jeong^a, Sook Yoon^{c,**}, Dong Sun Park^{a,b,*}

^a Department of Electronics Engineering, Jeonbuk National University, Jeonju 54896, South Korea

^b Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju 54896, South Korea

^c Department of Computer Engineering, Mokpo National University, Muan 58554, South Korea

ARTICLE INFO

Article history:

Received 30 December 2024

Received in revised form 22 April 2025

Accepted 2 June 2025

Available online 6 June 2025

Keywords:

Action recognition

Bird eye view

Deep learning

Multi-camera system

Precision livestock farming

ABSTRACT

Cattle behavioral monitoring is an integral component of the modern infrastructure of the livestock industry. Ensuring cattle well-being requires precise observation, typically using wearable devices or surveillance cameras. Integrating deep learning into these systems enhances the monitoring of cattle behavior. However, challenges remain, such as occlusions, pose variations, and limited camera viewpoints, which hinder accurate detection and location mapping of individual cattle. To address these challenges, this paper proposes a multi-viewpoint surveillance system for indoor cattle barns, using footage from four cameras and deep learning-based models including action detection and pose estimation for behavior monitoring. The system accurately detects hierarchical behaviors across camera viewpoints. These results are fed into a Bird's Eye View (BEV) algorithm, producing precise cattle position maps in the barn. Despite complexities like overlapping and non-overlapping camera regions, our system, implemented on a real farm, ensures accurate cattle detection and BEV-based projections in real-time. Detailed experiments validate the system's efficiency, offering an end-to-end methodology for accurate behavior detection and location mapping of individual cattle using multi-camera data.

© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The agriculture industry plays a crucial role in feeding the global population. As concerns over overpopulation continue to grow, the demand for increased agricultural productivity has intensified, prompting the integration of advanced technologies such as robotics, automation, and artificial intelligence (AI) to optimize outputs and meet global needs (Kleanthous et al., 2022). Within this sector, livestock farming is a fundamental factor. The increasing need to maximize the productivity of each animal has made cattle monitoring and management more crucial than ever (Chung et al., 2018). In this study, we focus on behavior monitoring of cattle housed in closed barn environments.

To enhance cattle welfare and productivity, Precision Livestock Farming (PLF) technologies have emerged as promising tools for

individualized animal monitoring (Halachmi et al., 2019; Tullo et al., 2019). PLF systems collect high-resolution data on cattle behavior, enabling the early detection of stress, illness, or injury (Zheng et al., 2023). By notifying farmers in real-time, these technologies help prevent disease outbreaks and minimize animal suffering (Wisniewski et al., 2019). However, the use of wearable devices in PLF may introduce challenges such as behavioral disruptions, animal stress, and operational limitations in large-scale farms (Ayadi et al., 2020). These systems also require frequent maintenance and are susceptible to environmental wear and measurement inaccuracies (Halachmi et al., 2019; Tuytens et al., 2022). Consequently, there is growing interest in non-intrusive monitoring methods that can observe animal behavior without direct interaction (Qiao et al., 2023).

Non-intrusive approaches leverage computer vision and deep learning to analyze behavior through video surveillance systems (Mahmud et al., 2021; Guo et al., 2016; Sumi et al., 2017). A pioneering study by (Fuentes et al., 2020) introduced an action recognition framework for cattle behavior based on video data, categorizing actions into individual, group, and partial interactions. This work was later extended to include individual identification and long-term behavioral analysis through multi-object tracking and spatiotemporal statistics (Fuentes et al., 2023). These advancements pave the way for early detection of welfare

* Corresponding author at: Department of Electronics Engineering, Jeonbuk National University, Jeonju 54896, South Korea.

** Co-corresponding author at: Department of Computer Engineering, Mokpo National University, Muan 58554, South Korea.

E-mail addresses: fnasir@jbnu.ac.kr (M.F. Nasir), afuentes@jbnu.ac.kr (A. Fuentes), shujiejulie@jbnu.ac.kr (S. Han), ryugagi@jbnu.ac.kr (J. Liu), yjeong@jbnu.ac.kr (Y. Jeong), syoon@mokpo.ac.kr (S. Yoon), dspark@jbnu.ac.kr (D.S. Park).

¹ Equal contribution.

concerns, such as reduced mobility or abnormal behavioral patterns (Szenci, 2022; Kleen and Guatteo, 2023). In addition, the collected data can inform management decisions regarding breeding, feeding, and treatment schedules shifting livestock farming from experience-based to data-driven management, significantly enhancing productivity and animal welfare (Meng et al., 2023; Tzanidakis et al., 2023) .

In this study, we build on this foundation by specifically monitoring individual cattle actions. We refine the hierarchical framework introduced by Fuentes et al. (2020), expanding it to encompass 14 distinct behaviors. This structured hierarchy is explained in detail later in the paper. Our study targets Hanwoo cattle, a native Korean breed known for its uniform brown coloring and visually similar features, which makes individual identification particularly challenging. Unlike wearable-based methods, non-intrusive camera-based tracking systems must handle typical computer vision challenges, such as occlusion, pose variation, and fluctuating lighting conditions . A critical limitation of single-camera setups is the dimensional reduction of 3D animal movements into 2D projections, which increases the likelihood of occlusion (López-Cifuentes et al., 2018; Nasir et al., 2023). While Han et al. (2023) addressed some of these issues with a multi-cattle tracking system, their method is limited to short-term tracking and demands high computational resources. To overcome these limitations, we propose a multi-camera system that offers diverse perspectives, facilitating more robust behavior recognition (Guan et al., 2020).

Recent research on multi-camera tracking, particularly in pedestrian and dairy cattle monitoring, demonstrates the benefits and complexities of integrating multiple visual perspectives (Amosa et al., 2023; Lima et al., 2021; Yan et al., 2021). Yamamoto et al. (2025) tracked cows across entire barns using multiple cameras, while Myat Noe et al. (2025) conducted cattle tracking in outdoor ranches. Mar et al. (2023) focused on top-view monitoring in enclosed dairy barns. Our work differs by targeting Hanwoo cattle, whose visual similarity increases the difficulty of individual tracking, unlike dairy cattle that often display distinctive coat patterns. Furthermore, our system aims not only to track individuals but also to classify their behaviors over time as part of a broader behavioral monitoring strategy, supported by a curated dataset. The camera setup includes both overlapping and non-overlapping views of the barn, enhancing long-term monitoring accuracy.

Despite these advantages, challenges such as occlusion caused by barn structures, animal movement, and poor nighttime visibility remain (Hartley and Zisserman, 2004). Differentiating visually similar individuals across multiple views also presents difficulties (Han et al., 2023). To address these issues, we introduce mapping and fusion strategies to improve multi-camera integration. Specifically, we implement a Bird's Eye View (BEV) positioning system that projects cattle onto a unified top-down map using ground points extracted from a deep learning-based pose estimation model. Fusion algorithms are then used to merge detections from multiple cameras, resolving identity ambiguities and creating a cohesive spatiotemporal map of cattle positions and behaviors across perspectives (Hu et al., 2022). This monitoring system allows for continuous assessment of cattle movement patterns and behavioral states within the barn.

To summarize, the key contributions of this study are:

- 1. **Multi-camera setup and dataset:** We present a strategy to tackle the challenges of multi-cattle detection in real farm settings using a multi-camera setup. Using this approach, datasets were collected during both day and nighttime from four distinct camera viewpoints. Domain experts meticulously annotated the dataset, classifying individual cattle behaviors into distinct classes. Additionally, ground points were labeled for each individual, with a single key point marking the exact ground reference of each animal in every camera frame.
- 2. **Hierarchical behaviors detection and mapping:** We propose a comprehensive methodology combining object detection, pose

estimation, action classification, and position mapping into BEV maps. Multiple BEV maps are fused using novel algorithms to resolve duplicate detections.

- 3. **Scene-centric homography mapping and fusion:** To ensure accurate BEV representation, we use a scene-specific homography mapping technique and situation-adaptive fusion methods, enabling precise projection of individual cattle across views.
- 4. **Welfare monitoring:** Farm-ready system for welfare monitoring: Our system is designed for real-world deployment in cattle farms, enabling continuous behavioral analysis to support animal welfare and data-driven decision-making.

The remainder of this paper is organized as follows: Section 2 provides details about the dataset and discusses the challenges encountered in real farm scenarios. Section 3 elaborates on the proposed methodology, explaining the functionality of each module. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the paper by summarizing the research findings.

2. Problem statement and dataset

This section highlights the challenges associated with datasets collected from real farm environments and provides insights into non-intrusive cattle welfare monitoring.

2.1. Dataset challenges

2.1.1. Cattle activities

To enable precise behavioral monitoring of individual cattle, we have categorized cattle activities into distinct classifications, as outlined in Table 1. These activities are grouped into three main hierarchical categories: **Individual**—activities performed by a single cow; **Transition**—activities involving a change in the cow's state or posture; and **Group**—activities involving interactions between two or more cows. At the highest level of this hierarchy are individual actions, which are the most frequently observed, accounting for majority of a cow's daily activity and providing detailed insights into specific behaviors. Next are transitional actions, which are brief and primarily serve as inter-changing of

Table 1
Criteria for identifying cattle activity.

Category	Actions	Criteria for Identification
Individual	Walking	The cow's feet are moving, as observed from consecutive frames in motion.
	Standing	The cow is stationary, not moving, in a standing posture, and not performing other activities.
	Resting	The cow is lying on the ground with legs tucked in, eyes open, and head up.
	Eating	The cow has its head through the fence, consuming feed (may not always be visible).
	Sleeping	The cow is lying on the ground with its head resting on its body and eyes closed.
	Self-grooming	The cow licks itself, regardless of whether standing or lying down.
Transition	Standing up	A transitional action where the cow moves from resting to standing.
	Lying down	A transitional action where the cow moves from standing to resting.
Group	Fight	Cows push each other aggressively using the front of their heads, colliding or hitting bodies.
	Head-butt	A cow pushes another using its shoulder, neck, or body.
	Fight Push	A cow pushes another using its shoulder, neck, or body.
	Social licking (active)	A cow actively licks another cow.
	Social licking (passive)	A cow is being licked by another cow.
	Mounter	A cow mounting another cow (on top).
	Mountie	A cow being mounted by another cow (underneath).

different individual behaviors. Finally, group actions involve interactions between two cows and, while typically short and repetitive, offer valuable information about social dynamics and collective behavior patterns.

2.1.2. Challenges

Despite the detailed action features and characteristics for distinguishing cattle in complex scenarios, detecting and differentiating these actions remains challenging. Fig. 1 illustrates several scenarios that highlight these challenges. In Scenario (Fig. 1a), a group of cows resting together in a complex formation makes it difficult to distinguish individual cattle. Scenario (Fig. 1b) presents a similar herd formation during the day, where varying sunlight conditions result in a resting cow being undetectable, affecting surveillance accuracy. Scenario (Fig. 1c) showcases object deformation, where a cow grooming itself under low light conditions and unclear viewpoints makes it hard to identify its head and actions. Scenario (Fig. 1d) involves two Hanwoo cows moving in a circular pattern, occluding each other's bodies and complicating the tracking of their identities, a problem noted in previous research (Fuentes et al., 2023). Scenarios (Fig. 1e) and (Fig. 1f) demonstrate occlusion caused by barn structures, with Scenario (Fig. 1e) illustrating partial body occlusion by a column and Scenario (Fig. 1f) showing extensive occlusion from both barn structures and other cattle. These challenges, including issues related to object detection, occlusion, and environmental conditions, are addressed by a multi-camera setup, which improves surveillance by providing diverse viewpoints and reducing the impact of occlusion and lighting variability. This paper incorporates these real-world challenges observed in farm environments into the cattle dataset for analysis.

2.2. Data acquisition – Experimental site and camera settings

The dataset was collected from an indoor cattle farm in Imsil, South Korea. The barn, measuring 30×12 meters, features a shed roof, fencing to contain the cattle, and open areas for ventilation. For analysis, the barn was divided into five segments based on its structural layout, as shown in Fig. 2. The study monitored 17 adult Hanwoo cows, a breed renowned for its uniform brown coat and high-quality meat. The cows were fed three times daily—morning, noon, and night—through designated eating slots in the fence. Additionally, a specialized feeding machine provided supplemental nutrients to enhance their diet.

A 4 K multi-camera setup, comprising four Hikvision cameras, was strategically installed to ensure complete coverage of the barn while maintaining optimal visibility of the cattle. Our camera placement

approach deliberately includes both overlapping and non-overlapping regions. Overlapping areas help reinforce detection accuracy through multiple viewpoints, particularly in regions with high cattle activity. However, we carefully balanced this to avoid excessive overlap, which can complicate multi-view data fusion. Non-overlapping regions are primarily used to cover isolated or corner areas of the barn, ensuring that no sections are left unmonitored.

Each camera recorded timestamped footage over several months, capturing cattle engaged in a wide range of individual actions. The dataset includes footage from different times of the day: midnight to sunrise (0000–0600 h), sunrise to noon (0600–1200 h), and sunset to midnight (1800–2400 h). A total of 40 video clips were selected for this study, each containing 25 to 30 s of synchronized footage from all four cameras within each timeframe. The selected videos were converted into individual image frames at 15 frames per second (FPS) and stored as part of the dataset.

2.3. Data preparation and annotation

2.3.1. Data preparation

Monitoring Hanwoo cattle presents unique challenges due to their slow mobility and relatively passive social interactions compared to other animals (Kim et al., 2017). Hanwoo cattle primarily stand, rest, ruminate, and occasionally walk slowly, with only brief instances of social interaction. Capturing a balanced dataset that accurately reflects these behaviors was challenging. To address this, footage was carefully curated to include short clips recorded at various timeframes using all four synchronized cameras. The dataset was designed to emphasize complex actions, diverse weather conditions, and obscured movements, ensuring robust and comprehensive training data.

2.3.2. Action data annotation

Identifying cattle actions is inherently ambiguous and often requires multiple camera viewpoints for accurate recognition. Trained annotators utilized a CVAT platform-based tool to label the dataset, marking cattle with bounding boxes across 14 action classes. Each animal was assigned a unique ID that remained consistent across all camera views, ensuring synchronization throughout the scene. Heavily occluded cattle in overlapping regions were annotated with the same action class as partially visible cattle. For actions with high similarity, such as standing and walking, annotators relied on all available camera angles and specific parameters to differentiate and accurately label each action. Table 2 presents the number of instances for each cattle action in the complete dataset.

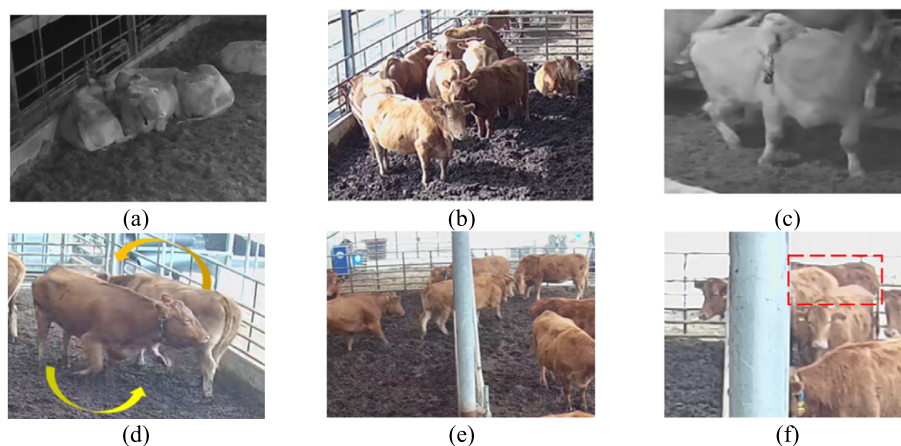


Fig. 1. Examples of challenges in cattle detection. (a) and (b) illustrate herding challenges; (c) shows object deformation; (d) represents complex actions; (e) and (f) depict different types of occlusion.

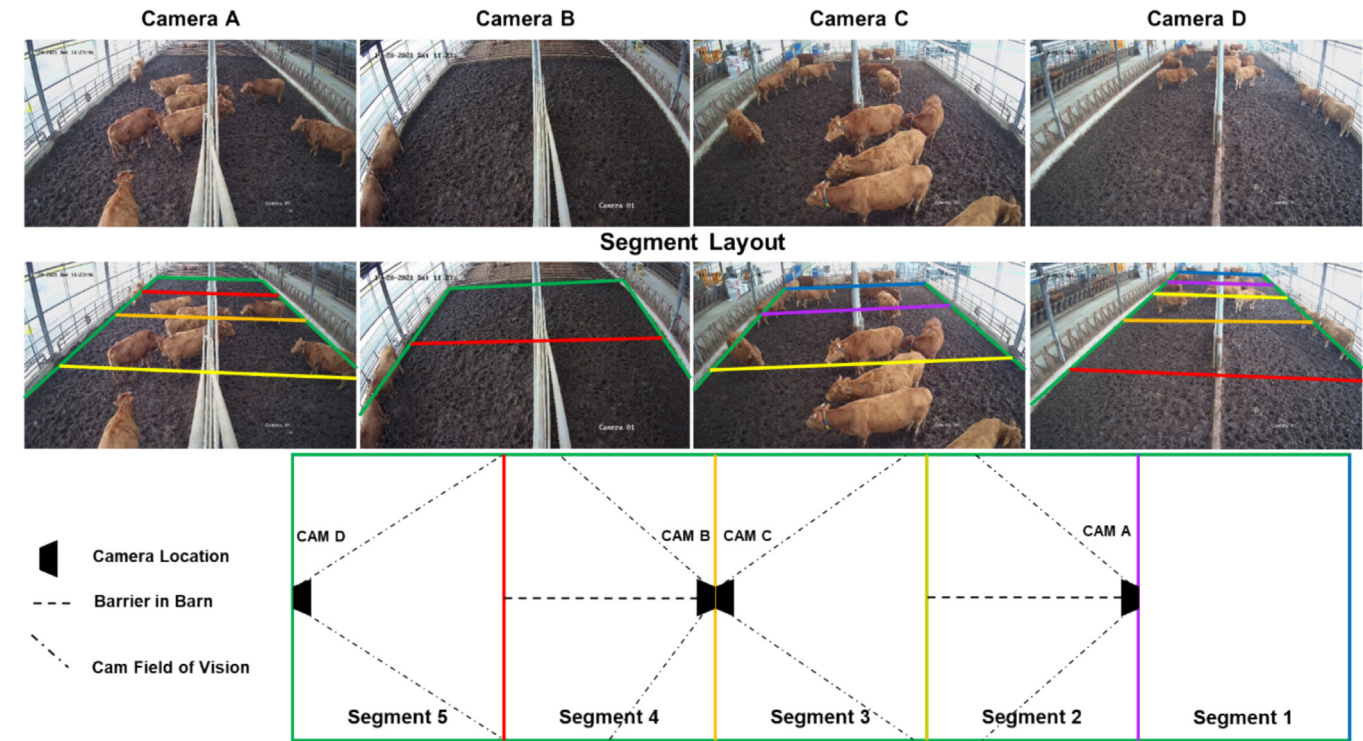


Fig. 2. Multi-Camera Setup. Camera viewpoints and fields of vision are relative to the different segments of the cattle barn, highlighting structural barriers.

2.3.3. Ground point annotation

For the BEV positioning system, a ground reference plane of the cattle barn, including its structural elements, was prepared (see Figs. 2 and 5). The barn was divided into five segments, with 18 columns spaced 6 m apart, and a fence between the columns serves as a visual reference for mapping verification. The BEV operational area aligns with the barn's architectural layout.

In the BEV cattle mapping process, each animal is represented as a single point, simplifying the three-dimensional structure of the body to a two-dimensional plane (Montero et al., 2023). Du et al. (2022) introduced a 2D-3D fusion method for livestock body measurement using depth cameras, detecting key points on RGB images and mapping them onto 3D point clouds. While effective, the method is

depth-dependent and tailored for single-animal measurement. In our approach, we aim to identify a single, precise point for each individual cattle to ensure accurate location mapping. For Hanwoo cattle, this point is defined at the center of the four limbs, aligned with the spinal vertebra. Accurate ground point estimation is critical, as bounding box-based points can vary significantly depending on the cattle's pose, leading to projection misalignments across camera views (see Fig. 3). To address these challenges, a ground point estimation model is employed to ensure consistency and reliability in the BEV mapping process.

3. Methodology

This paper presents a framework for monitoring individual cattle behavior and precisely determining their position within a barn using a multi-camera setup. An overview of the methodology is provided in Fig. 4. The system operates as follows: First, synchronized image frames are captured from each surveillance camera. These frames are then processed in parallel by two deep learning models: the Action Detector and the Ground Point Estimator. The Action Detector identifies the actions being performed and labels the bounding boxes defining each cattle's spatial position, while the Ground Point Estimator calculates the precise ground point coordinates for each animal.

To integrate the outputs from these models, we propose a Multi-Module Integration Unit (MMIU). The integrated results are then processed by the novel Enhanced BEV Algorithm, which adapts to each camera's specific viewpoint and field of vision (FOV). The individual BEV maps are merged to produce multiple projections of the same cattle. Finally, a fusion process combines these projections to ensure that each animal is accurately represented as a single entity. Each component of the process is described in detail below.

3.1. Action detection

The action detector is responsible for identifying cattle actions and bounding boxes, which define their spatial locations within the scene. In our previous work (Fuentes et al., 2020, 2023), we employed

Table 2
Action instances in the datasets.

Category	Actions	Datasets				
		Camera A	Camera B	Camera C	Camera D	Joint Cameras (A, B, C, D)
Individual	Walking	1283	414	3301	2665	7663
	Standing	11,193	3588	16,614	12,077	43,472
	Resting	2109	733	3789	2802	9472
	Eating	596	303	402	362	1663
	Sleeping	402	–	603	202	1207
	Self-grooming	–	–	207	114	321
	Transition	–	–	–	–	–
Group	Standing up	–	–	–	–	–
	Lying down	280	75	72	123	550
	Fight Head-but	33	–	490	404	927
	Fight Push	91	–	–	–	91
	Social licking (active)	–	–	102	168	270
	Social licking (passive)	–	–	102	168	270
	Mounter	175	16	415	370	976
Total	Mountie	156	–	431	370	957
		16,318	5169	26,528	19,825	67,840

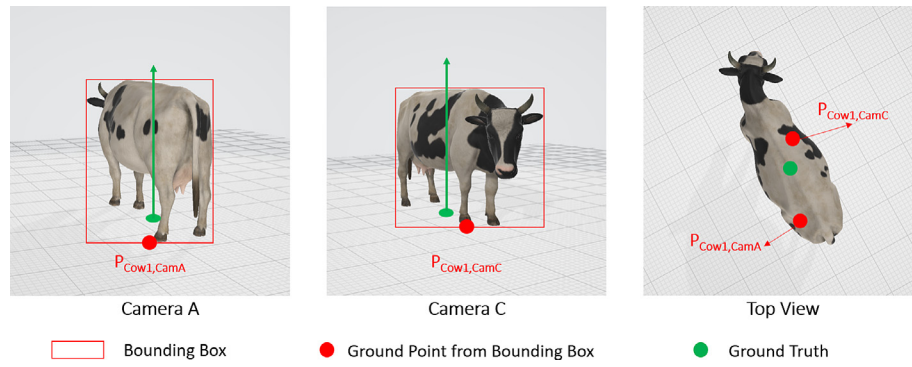


Fig. 3. Ground point calculation and accuracy. Ground points calculated from bounding boxes compared to the ground truth. Variations in cow positioning relative to the ground due to viewpoint make the bounding box center an inaccurate representation. A single, accurately estimated ground point is critical for this research.

YOLOv3 and YOLOv5, two powerful single-stage object detection algorithms widely used for their computational efficiency (Jocher, 2023). These models generate class predictions and confidence scores, optimizing their outputs using a loss function that minimizes the disparity between predicted values and ground truth annotations.

The accuracy of bounding box predictions is evaluated using the IoU (Intersection over Union) metric, which ensures precise coordinate identification. The loss function for the action detector is represented as:

$$L_{\text{Action Detector}} = L_{\text{Classification}} + L_{\text{Confidence}} + L_{\text{IoU}} \quad (1)$$

where, $L_{\text{Classification}}$, $L_{\text{Confidence}}$, and L_{IoU} represent classification loss, confidence loss, and bounding box regression loss, respectively.

Object detection experiments were conducted using the Ultralytics YOLOv5 library with default settings (Jocher, 2023). The dataset was randomly split into a 70/30 training/validation ratio.

Batch sizes alternated between 16 and 32, with the learning rate maintained at its default value. Models were trained for 300 epochs using two NVIDIA GEFORCE RTX 4090 GPUs in a distributed training setup.

3.2. Ground point estimation

To estimate the ground point of each cow from multiple camera viewpoints, a simplified pose estimation approach was adopted. The objective was to generate a single key point representing the ground point of each cow, as illustrated in Fig. 3.

The Object Keypoint Similarity (OKS) metric was employed to evaluate the keypoint accuracy (Rath, 2023). Keypoints were regressed using normalized L1 loss, with OKS integrated into the loss function:

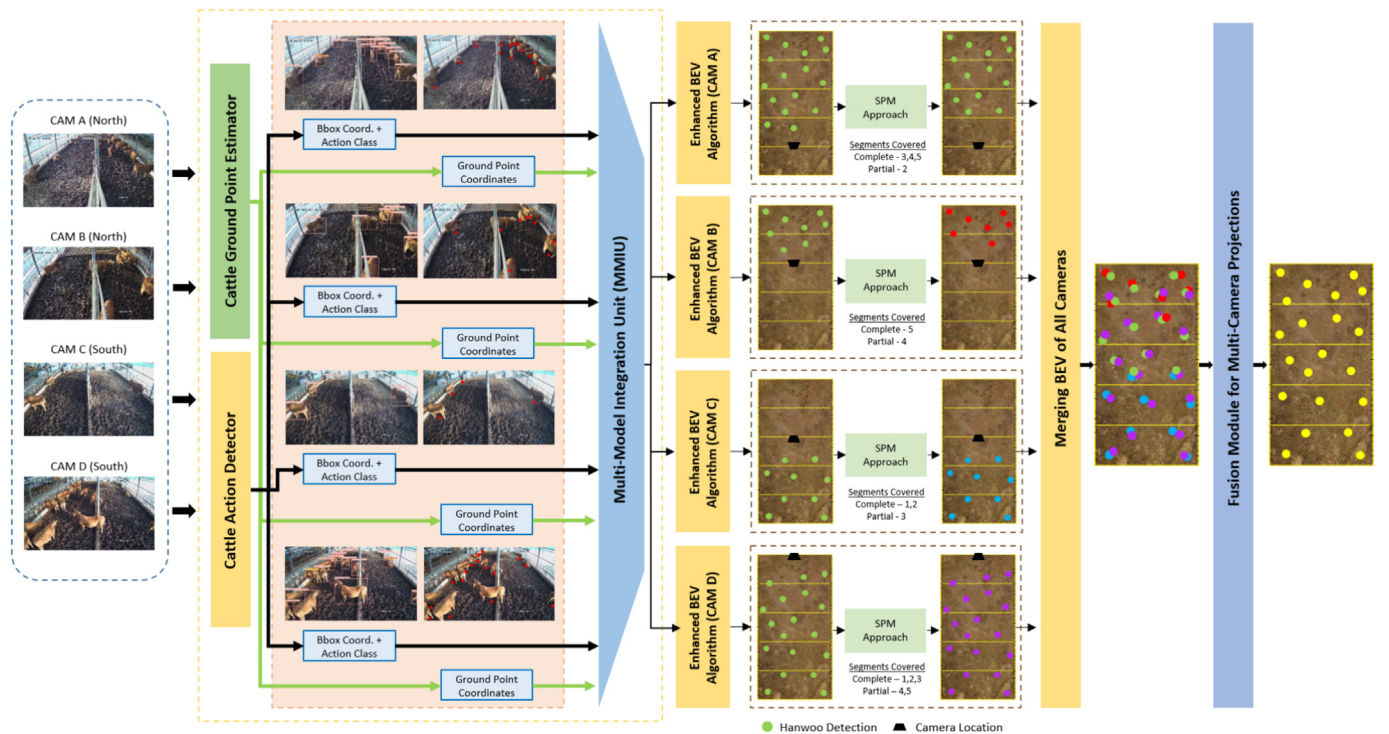


Fig. 4. Overview of the Proposed Methodology. The proposed framework for monitoring cattle behavior and determining precise positions within a barn. It integrates synchronized multi-camera image capture, deep learning-based action detection and ground point estimation, multi-module integration, and an enhanced BEV algorithm to produce accurate and unified cattle representations.

$$L_{OKS(p,g,v)} = 1 - \exp\left(\frac{-d^2}{2s^2}\right) \cdot v \quad (2)$$

where, p and g represent the predicted and ground truth keypoints respectively; v is the visibility flag, s is the scaling factor, and d is the Euclidean distance between p and g .

Each keypoint also includes a confidence parameter indicating its visibility, computed using Binary Cross Entropy (BCE) loss:

$$L_{BCE} = -y \cdot \log(\hat{p}) - (1-y) \cdot \log(1-\hat{p}) \quad (3)$$

where, $\hat{p}$ is the predicted probability of keypoint visibility, and y is the groundtruth (0 or 1). The BCE loss function quantifies the error between the predicted and actual visibility values, providing the model with a measure to adjust its predictions during training.

The combined loss function for ground point estimation is expressed as:

$$L_{GP} = \lambda \cdot L_{OKS} + (1-\lambda) \cdot L_{BCE} \quad (4)$$

The weighting factor λ determines the relative importance of each loss component in the overall optimization process. This combined loss function allows the model to optimize both the keypoint location accuracy (measured by OKS) and the visibility status prediction (measured by BCE).

3.3. Multi-model integration unit (MMIU)

The MMIU integrates outputs from the action detection and ground point estimation models to produce accurate BEV-based cattle positions. The action detection model identifies cattle actions, IDs, and bounding boxes, while the ground point estimation model provides precise ground point coordinates. These complementary outputs are combined to enhance positioning accuracy.

This module replaces the bounding box center coordinates from the action detection model with the more accurate ground point coordinates when available. In cases where a ground point cannot be detected due to occlusions or partial visibility, the original bounding box coordinates are retained. This adaptive approach ensures consistent positional data even in challenging conditions.

To achieve this integration, the MMIU compares each bounding box with the keypoints from the ground point estimation model. If a keypoint lies within the bounding box boundaries, the bounding box coordinates are replaced with the keypoint coordinates. If no corresponding keypoint is detected, the bounding box coordinates remain unchanged. This process, supported by a fusion module, optimally combines outputs from both models to address issues such as occlusion and partial visibility, ensuring reliable cattle positioning in the BEV framework.

3.4. Enhanced BEV algorithm

A fundamental part of this research is the BEV mapping of cattle, which provides a comprehensive view of the barn. This approach offers clearer insights into spatial relationships compared to individual camera perspectives (Myat Noe et al., 2023, 2025). The effectiveness of BEV mapping depends on factors such as the number of cameras, their placement, diverse viewpoints, and the minimization of blind spots (Hartley and Zisserman, 2004). Camera geometry principles are used to transform 3D coordinates (X, Y, Z) into 2D image coordinates (x, y), which are then mapped to the BEV coordinate system for accurate positioning.

Each camera was calibrated using feature points and a ground mesh, enabling precise BEV projection mapping and supporting the enhanced BEV algorithm. A BEV map of the barn surface was manually defined, incorporating structural elements, camera viewpoints, barriers, and

fences. Matching feature points between camera images and the BEV map reduces calibration errors and ensures projection accuracy (Saponara et al., 2022; López-Cifuentes et al., 2018). For this study, the barn's area, measuring 30 m × 12 m, was divided into five segments (Fig. 5a). The left side of each segment contains six 1-m fence slots, while the right side includes eight mechanical head gates per segment, each 0.75 m wide. These structural details are critical for calibration and mapping.

Using a multi-camera approach, the barn's four stationary cameras capture cattle as dynamic objects. A unique method was employed to calculate their projections onto the BEV map. Increasing the number of feature points enhances the accuracy of the homography matrix, leading to more precise estimations (Nguyen et al., 2018). However, marking an excessive number of points can be time consuming. To balance accuracy and efficiency, feature points were strategically distributed across the operational plane, focusing on areas prone to image distortions or outliers. This process leverages a Novel Ground Mesh-Based Approach, incorporating hundreds of corresponding points (Fig. 5b).

Using the vanishing point tool in Adobe Photoshop, the ground perspective in each camera's view was estimated. This tool facilitated the creation of detailed grids with varying densities on the ground plane (Fig. 5c). Converging lines were identified as vanishing points to define perspective, and corresponding points were marked on the BEV plane. These points were aligned with the operational area and adjusted using visual cues, serving as critical references for homography transformations.

The homography transformation relies on corresponding feature points on the image and BEV planes. Points in the image plane denoted as $p_{(n,k)}$, represent the coordinates of the n -th cattle captured by camera k , while the corresponding BEV points are denoted as $P_{(n,k)}$. This relationship is represented by the transformation matrix, that is represented as:

$$P_{n,k} = H_k \times p_{(n,k)} \quad (5)$$

Here, H_k is the homography matrix for camera k , calculated using the feature points:

$$H_k = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (6)$$

Once the homography matrix is computed, new points, such as those derived from cattle ground point estimation, can be transformed from the 2D image plane to the BEV plane and accurately projected onto the map.

3.5. Multi-camera projections

The multi-camera setup provides complete coverage of the barn structure, leaving no part of the ground plane uncovered and minimizing blind spots. The few remaining blind spots are caused by structural occlusions, such as columns. Additionally, each camera captures both overlapping and non-overlapping regions relative to the other cameras in the setup. To analyze this, we divided the barn into segments and used visual cues to assess visibility from each camera and the overall coverage of each segment, as shown in Table 2. Corner fences, as illustrated in Fig. 5(a), were used as a metric to determine visibility.

Fusion techniques were applied to merge the information from all cameras for a comprehensive understanding of the scene. Each camera focuses on specific groups of Hanwoo cattle within its field of view (FOV). A singular BEV map was used to integrate all projections, which were then identified, connected, and fused. In Table 3, the following visibility categories are defined:

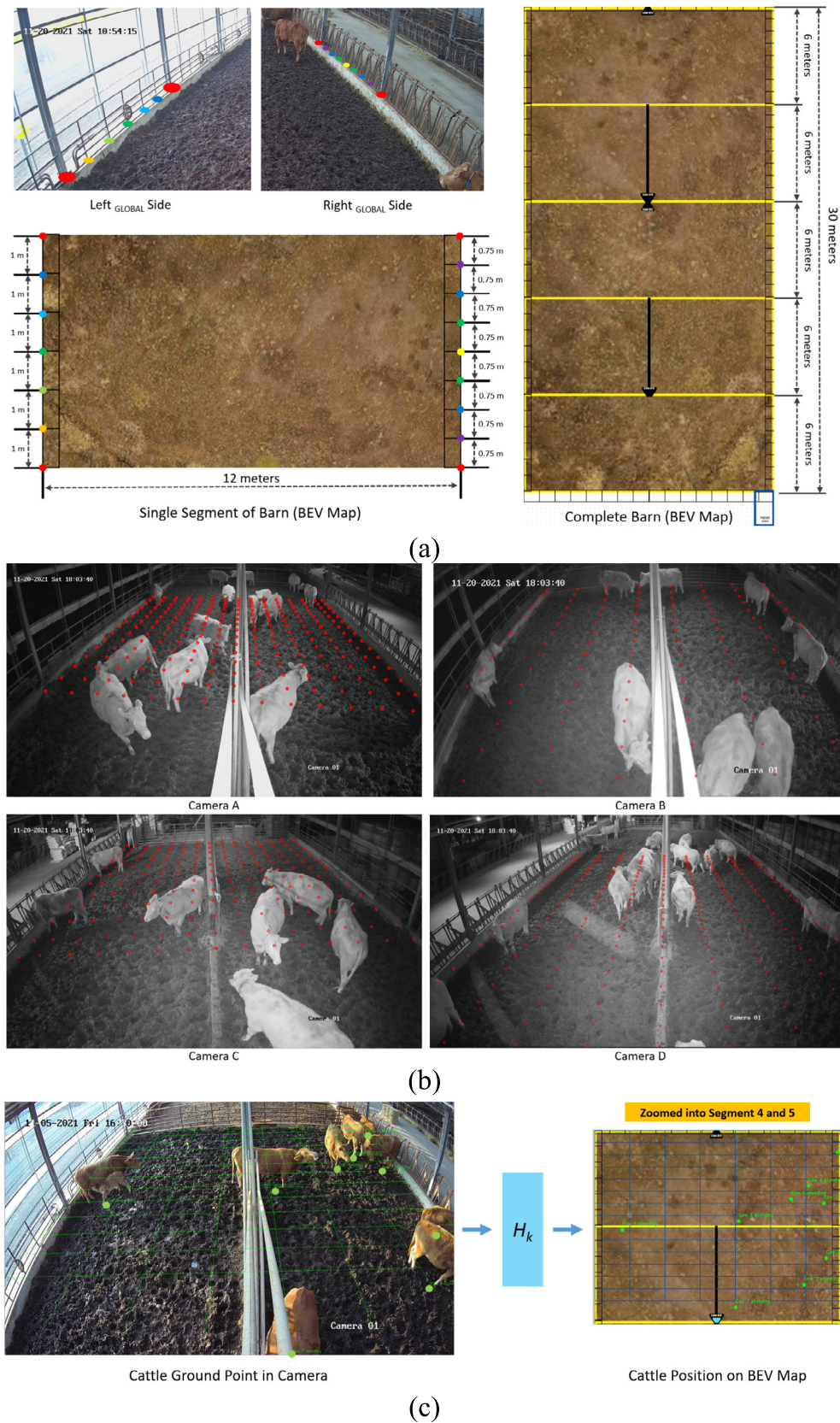


Fig. 5. Detailed Overview of the Enhanced BEV Algorithm. (a) Analysis of the operational area, including segment breakdown and structural elements, with distinctions between the global left and right sides featuring different fence types and configurations. These elements contribute to constructing a top-view BEV map of the cattle barn. (b) Visualization of the ground mesh-based approach, showing how feature points within each camera's field of view (FOV) are utilized for accurate mapping. (c) Demonstration of the enhanced BEV algorithm, illustrating the transformation of cattle detections from individual camera frames into precise projections on the BEV map.

Table 3

Segment coverage breakdown by camera fields of view (FOV). Coverage of barn segments by the fields of view (FOV) of all four cameras, determined using visual cues and boundary fences as reference points. Cameras providing the most reliable results for each segment are highlighted in bold.

Cam ID	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Camera A	–	Partial (0/8, 2/6)	Partial (7/8, 6/6)	Complete View	Complete View
Camera B	–	–	–	Partial (2/8, 4/6)	Best View
Camera C	Best View	Best View	Partial (1/8, 0.75/6)	–	–
Camera D	Complete View	Complete View	Complete View	Partial (8/8, 5/6)	Partial (3/8, 0/6)

- **Best View:** A camera that covers an entire segment without significant occlusion and can handle complex scenarios like cattle herding. These cameras are highly reliable.
- **Complete View:** A camera that provides full visibility of a segment and its neighboring areas but may lack detail due to distance.
- **Partial View:** A camera that captures part of a segment, often providing closer and clearer views of cattle, even if the overall coverage is incomplete.

Not all segments have the best view, but all segments are visible through complete views. In some cases, partial views may be more useful than complete views, as they offer unobstructed, close-up perspectives of the cattle. Conversely, complete views may show cattle farther from the camera, making them appear smaller and more prone to occlusion. Fig. A1 in the appendix illustrates these concepts in detail. Cameras A and B are north-facing, while cameras C and D are south-facing. When projections from north-facing and south-facing cameras are mapped onto the BEV plane, they may appear farther apart than projections from cameras facing the same direction. This discrepancy leads to a preference for fusing projections from similarly oriented cameras, as this approach results in higher alignment accuracy.

3.6. Fusion module

The projections mapped onto the 2D BEV plane are not merely X-Y coordinates; they also contain metadata such as the camera used for detection, the segment where the cattle are located, the action being performed, and the ID generated by each camera. This metadata is essential for associating projections of the same cattle across different cameras. To unify all projections into a single operational area, the union operator was applied:

$$\text{Operational Area} = \bigcup_{k=1}^C P_{(n,k)} \quad (7)$$

where C is the total number of cameras. A geometric method is employed to analyze the relationships between multiple projections, as illustrated in the fusion process pipeline in Fig. 6. Each component of the fusion module is detailed below.

3.6.1. Segment preference method (SPM)

Not all detections are relevant for the BEV representation. To optimize the process, we introduce the Segment Preference Method (SPM), which determines the most effective camera perspective for each segment. This method evaluates each camera's field of view (FOV) across various scenes and cattle movement patterns. Segments

with a Best View require minimal fusion and are highly reliable, while others rely on a combination of complete and partial views (Table 3).

The SPM filter is applied iteratively to refine detection accuracy and minimize fusion errors through the following steps:

1. **Initial Iteration:** All available camera views are utilized, as shown in Fig. 6. If fusion errors are significant, the SPM filter is adjusted.
2. **Second Iteration:** Relies on Best View cameras where available. For segments without a Best View, all camera views are considered.
3. **Third Iteration:** Focuses on optimal views, highlighted in bold in Table 3.
4. **Final Iteration:** Utilizes only Best View cameras for segments with them and a single camera view for remaining segments (e.g., Camera A for Segments 3 and 4).

Cameras for each segment are categorized as primary (bold in Table 3) and secondary (remaining views). If errors persist during any iteration, secondary views are filtered out, and the process is recalibrated.

Key insights from SPM highlight that partial views are effective for detecting close, unobstructed cattle, even if the full segment is not visible, while complete views reduce fusion errors but may fail to detect distant cattle accurately. Although using more camera views increases detection chances, it also introduces higher fusion errors. The iterative process balances detection and fusion challenges, with the second iteration achieving a balance between detection and fusion, the third iteration focusing on reducing fusion errors while maintaining detection rates, and the fourth iteration eliminating fusion challenges entirely, though it may negatively impact detection accuracy, particularly for segments like Segment 3. Overall, SPM has proven effective, reducing redundant projections by 60–65 % in most scenarios.

3.6.2. Basic projection fusion

Basic projection fusion involves generating a region of interest (ROI) for each projection, represented as a circle centered on the projection with a defined radius. The radius must be carefully selected: a large radius may result in overly complex connections, especially in herding scenarios, while a small radius can cause missed connections, leading to multiple projections of the same cattle being treated as separate entities. Once the radius is determined, ROIs are generated for each detection, and projections are connected based on the rules outlined in Table 4.

Fig. 7 illustrates the process for connecting and fusing projection points. Each projection point is represented as $P_{(cow_{id}, camera_{id}, action_{id}, segmentation_{id}, coordinates(x,y))}$. For simplicity, they are

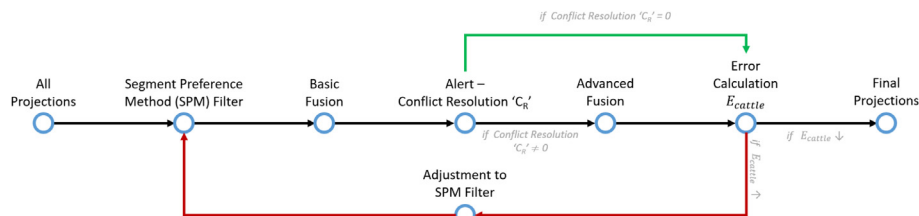


Fig. 6. Fusion module pipeline for multi-camera projections. Diagram illustrating the process of unifying multi-camera projections into a single operational area.

Table 4
Ruleset for connectivity of cattle projections.

No.	Rules of connection between projections
1	Distance (D): Each point must lie in the ROI of the other point, meaning distance between each point must be less than or equal to ROI's radius (R_{ROI}). If $D(P_{(x,k)}, P_{(y,k+1)}) \leq R_{ROI}$, then $P_{(x,k)}$ and $P_{(y,k+1)}$ are connected.
2	Camera (k): Connections can only be formulated between projections originating from different cameras. If $ROI_{(x,k)} \cap ROI_{(y,k)} = \emptyset$, then $P_{(x,k)}$ and $P_{(y,k+1)}$ are not connected.
3	Segment (S): Connections are allowed within segments and between neighboring segments due to the dynamic nature of cattle. If $ S_{(x,k)} - S_{(y,k+1)} \leq 1$, then $P_{(x,k)}$ and $P_{(y,k+1)}$ are connected.
4	Action (A): Connected points must belong to the same action category (e.g., Standing Actions, Grounded Actions, Transitional Actions, or Group Actions). If $A_{(x,k)} = A_{(y,k+1)}$, then $P_{(x,k)}$ and $P_{(y,k+1)}$ are connected.

denoted as $P_{(n,k)}$, where n represents cow ID number and k indicates the camera.

Connections are formed between two projection points and are categorized into groups:

1. **No Connections** (P_{No_Conn}): Projections that do not establish any connections.
2. **Connections** (T_{Conn}): Projections that are linked to others detailing their associations.

Within the Connections category, further subcategories are defined:

- **Single Connections** (T_{S_Conn}): Projections with one connection.
- **Multi-Connections** (T_{M_Conn}): Projections with multiple connections.

Once connections are authenticated, the finalized connections are merged by calculating the mean position among all linked points. New fused projection points are created and assigned unique cattle IDs. The majority action across the connected projections is selected as the new action, and the mean position is designated as the final location of the fused projection point.

3.6.3. Conflict resolution in complex scenarios

As shown in Fig. 6, after basic fusion, a conflict resolution alert stage is triggered to identify any connection vertex comprising more than three projections. Conflict Resolution (C_R) is determined by verifying the connections between projections:

$$C_R = \begin{cases} 1 & \text{if Connection Vertex Detected} > 3 \text{ Connections} \\ 0 & \text{Otherwise} \end{cases} \quad (8)$$

If an alert is raised ($C_R = 1$), the system proceeds to the advanced fusion module, which resolves detected conflicts. The approach for resolving such conflicts is outlined in Methodology 1 (Fig. 8).

In complex scenarios, such as herding cattle performing the same actions (Fig. 1(a) and (b)), incorrect connections may form despite the rules of authentication. These errors are challenging to resolve but are addressed by the advanced fusion module, as detailed in Methodology 1 and illustrated step-by-step in Fig. A2 (Appendix). Each projection is represented as $P_{(i,k,act)}$, where i is a sequential ID generated by camera k , and act is the detected action.

3.6.4. Feedback by error

To ensure accuracy, the remaining detected cattle projections (DET_{cattle}) are compared with the ground truth cattle count (GT_{cattle}).

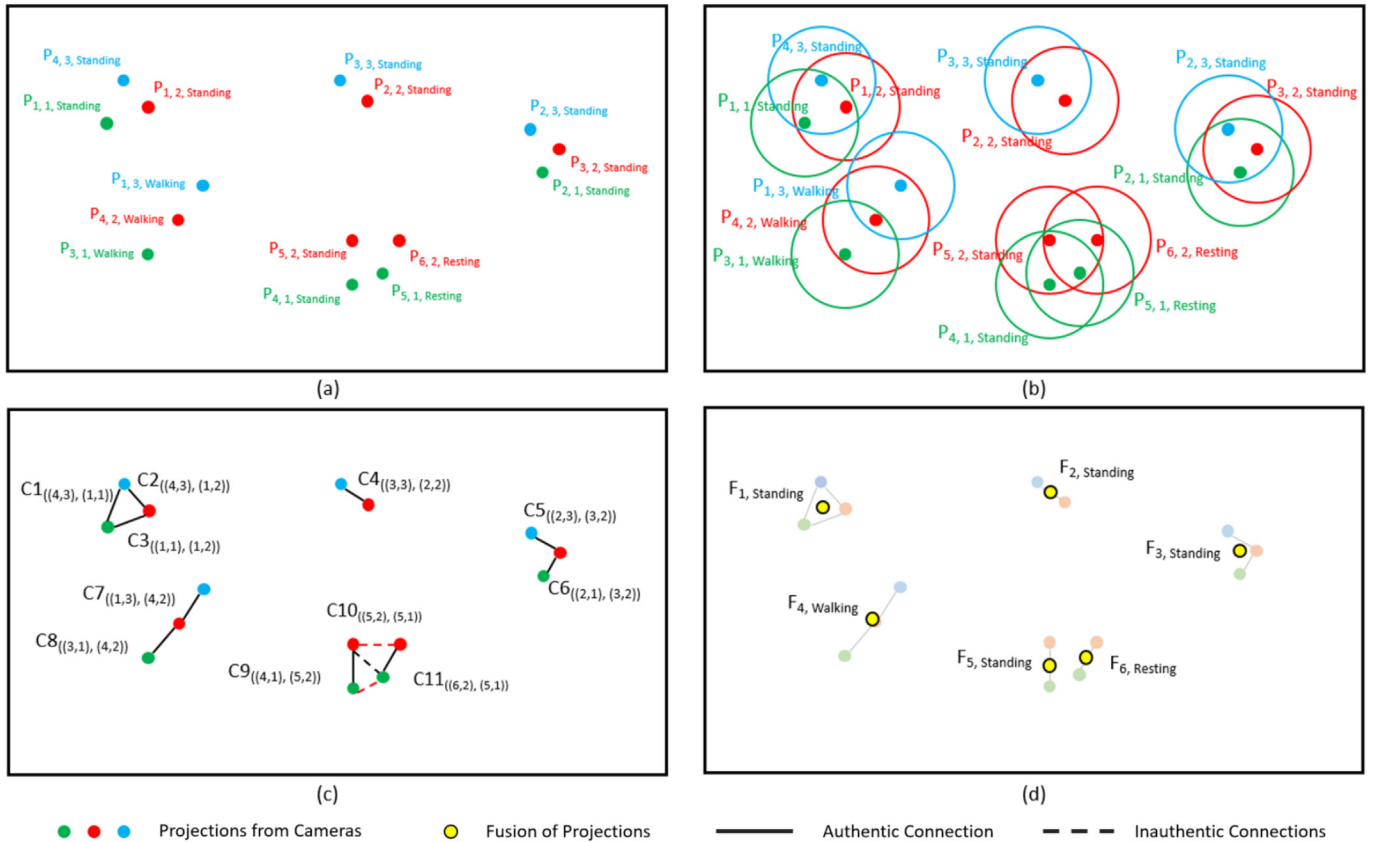


Fig. 7. Fusion Process for Cattle Projections onto the BEV Map. (a) Initial mapping of all projections onto the BEV plane. (b) Creation of Regions of Interest (ROIs) with a defined radius around each projection. (c) Establishment of connections based on authentication rules. (d) Generation of final fused projection points by merging authenticated connections.

Methodology 1: Advanced Fusion - Conflict Resolution Module

Input: BEV Projections of each cattle from each Camera $A = \{A_1, \dots, A_{Ta}\}$, $B = \{B_1, \dots, B_{Tb}\}$, $C = \{C_1, \dots, C_{Tc}\}$ and $D = \{D_1, \dots, D_{Td}\}$. Each projection includes metadata such as cow_{id} , $camera_{id}$, $action_{id}$, $segmentation_{id}$, $coordinates(x,y)$.

Stage A – Identifying Conflicted Vertex

- 1: Detect the number of conflicts (C_R).
- 2: Identify the segment causing the conflict.
- 3: Use the SPM filter to determine the Best View camera and secondary camera for the segment.
- 4: Exclude projections with single connections or no conflicts.
- 5: **Return** the conflict vertexes.

Stage B – Resolving Conflicted Vertex

- 6: Identify the number of connections for each projection in the conflict vertex.
- 7: Display connections, e.g. (A_1, B_3) , (B_3, A_2) , (A_2, B_4) , (B_4, A_3) , (A_3, B_5) .
- 8: Prioritize projections with single connections for merging (e.g., (A_1, B_3) is fused as F_1 and (A_3, B_5) is also fused as F_2).
- 9: Recheck the number of connections for the remaining projection points.
- 10: If no remaining conflicts and only single connections exist:
Fuse the remaining projections into F_n .
- 11: Otherwise:
- 12: Go to Step 8 to continue resolving conflicts.
- 13: **Return** fused projections: $F = \{F_1, \dots, F_n\}$

Fig. 8. Conflict resolution module methodology.

An 18 % fluctuation in the final cattle count is tolerated: higher deviations require adjustments to the SPM filter.

$$\hat{T} = Abs\left(\frac{DET_{cattle} - GT_{cattle}}{GT_{cattle}}\right) \times 100 \quad (9)$$

$$E_{cattle} = \begin{cases} 1 & \text{if } \hat{T} > 18\% \\ 0 & \text{if } \hat{T} \leq 18\% \end{cases} \quad (10)$$

Based on the results obtained from E_{cattle} , the flow of the fusion module is determined. If E_{cattle} value is high, we shift to the reset module and re-adjust the SPM filter along with reinitiating the pipeline. This iterative adjustment improves all elements of the fusion module and leads to improved results.

4. Experiments and results

This section details the experimental process, performance evaluation of the deep learning models, and the analysis of the proposed mapping and fusion techniques. These techniques aim to create a highly precise BEV map of the barn capable of handling complex cattle scenarios.

4.1. Implementation

4.1.1. Dataset

The proposed methodology was evaluated using a dataset specifically collected as part of this research. The dataset comprises footage converted into images at 15 FPS, captured by four HD cameras strategically positioned to monitor 17 Hanwoo cattle.

The dataset includes approximately 67,840 bounding box annotations across 6398 image frames, along with cattle ground keypoint annotations for a subset of 4188 image frames (see Table 2 for dataset details). The multi-camera setup provides both overlapping and non-overlapping fields of view, with temporally synchronized footage from all four cameras for each video sequence.

Given the central role of BEV mapping in this research, each camera was carefully calibrated with the BEV map to ensure precise cattle projections. The methodology's success relies on the seamless integration

of all proposed modules, with optimal results achievable only when each module performs at its highest efficiency.

4.1.2. Evaluation metrics

To evaluate the performance of the methodology, specific metrics were used for each module:

- **Action Detection:** The action detector was evaluated using the Mean Average Precision (mAP). A higher mAP indicates better performance of the detector.
- **Pose Estimation:**
 - **Keypoint Error (K_p Loss):** Measures the difference between predicted key points and the ground truth keypoint.
 - **Keypoint Classification Error (K_{cls} Loss):** A binary classification loss that determines whether a keypoint is present.
- **BEV Mapping:**
 - **Qualitative Assessment:** Visual inspection of the projected BEV map.
 - **Quantitative Assessment:** The re-projection error (E_{RP}), which measures the deviation of projected pixel points from their actual values caused during geometric transformation (e.g., homography).
- **Fusion of BEV-Based Projections:** The fusion process was evaluated using two key metrics:
 - **Missed Connections Error (E_{Miss}):** Refers to projections of the same detection not fused.
 - **Incorrect Multi-Connection Error (E_{IMC}):** Refers to cases where extra projections not belonging to the same object were fused.

These metrics serve as primary indicators of the fusion process. E_{IMC} leads to E_{Miss} . An ideal fusion scenario is when all cattle are detected, projected, and fused without introducing errors. Failure of cattle detection is detrimental to the fusion process, as it can lead to multiple errors.

4.2. Results

4.2.1. Action detector

The action detector was trained using the multi-camera cattle action dataset, comprising footage captured during both day and night from all four synchronized cameras. The dataset was split into training and

validation sets with a 70/30 ratio. To ensure diverse scenario representation, temporally adjacent frames were excluded between subsets. The dataset includes 6398 images containing a total of 67,840 annotations, averaging 11 annotations per image.

Table 5 shows the quantitative results for action detection, evaluated in terms of mean Average Precision (mAP) for individual actions across all cameras. The model was trained for 300 epochs, with the training curves depicted in Fig. 9 showing significant improvements: (a) bounding box loss decreases steadily over epochs, indicating enhanced spatial accuracy; (b) classification loss reduces consistently, reflecting precise action detection; (c) the F1-confidence curve reaches an optimal score of 98 %; and (d) precision-recall (PR) curves demonstrate near-ideal precision for individual actions, culminating in an overall mAP of 99.2 %.

4.2.2. Ground point estimation

The integration of the YOLOv8 Pose Estimation model significantly enhances the precision of determining the optimal ground position for cattle. A subset of the cattle action dataset, comprising 4188 images, was labeled with ground points. However, limitations in labeling primarily stemmed from the extensive coverage provided by Camera D and the complexity of annotating its data.

The YOLOv8l model was trained for 200 epochs using a dataset annotated with a single class, focusing solely on ground point estimation without multi-class elements. The model detects cattle and estimates a single keypoint for each animal. The pose estimation loss consists of two components: Keypoint Loss (K_p), which calculates the distance between predicted ground points and ground truth values (Eq. (2)), and Keypoint Classification Loss (K_{cls}), which evaluates whether a keypoint is correctly generated. Missed detections contribute to the error value. Fig. 10 illustrates the reduction of both losses across epochs, indicating improved accuracy.

The Multi-Model Integration Unit (MMIU) integrates outputs from the action detector and ground point estimator, ensuring the optimal extraction of cattle ground points from either model. This seamless combination improves the reliability and precision of the overall system.

4.2.3. BEV algorithm

The precise positioning of cattle using the BEV algorithm relies on two key elements: mapping and fusion, each thoroughly analyzed in this section.

4.2.3.1. Mapping and projection analysis. The ground mesh-based approach was employed for mapping, producing the most accurate results.

Table 5
Quantitative results of action detection (mAP) across all cameras.

Category	Actions	Action Detection Join Cameras (A,B,C,D) YOLOv5l mAP(%)
Individual	Walking	98.0
	Standing	99.5
	Resting	99.5
	Eating	98.7
	Sleeping	99.5
	Self-grooming	99.3
Transition	Lying Down	99.1
Group	Fight Head-butt	99.5
	Fight push	99.5
	Social licking active	99.5
	Social licking passive	99.4
	Mounter	99.3
	Mountie	98.8
Total (mAP)		99.2

A detailed analysis was conducted to examine the relationship between time to label points, the number of feature points, and projection error (E_{Proj}). The projection error, calculated using Eq. (11), measures the deviation between ground truth coordinates (x_i, y_i) and projected points (x'_i, y'_i) obtained from different homography techniques:

$$E_{Proj} = \frac{1}{n} \sum_{i=1}^n ((x_i - x'_i)^2 + (y_i - y'_i)^2) \quad (11)$$

where n is the number of cattle in a scene.

Table 6 shows that increasing the number of feature points reduces projection error, with the ground mesh technique serving as the ground truth. However, Camera B exhibits higher error due to its wide field of view, capturing curved or hemispherical perspectives with inherent distortions. Fig. A3 (See Appendix) illustrates these findings using footage from Camera D.

4.2.3.2. Re-projection error analysis. An important aspect of assessing inaccuracies in homography is the homography error, often measured as the re-projection error. This metric quantifies the disparity between transformed points and their actual positions after applying a geometric transformation (i.e., homography) between images or perspectives.

To calculate the re-projection error, the source feature points from the camera footage were transformed using the estimated homography matrix. These transformed projection points were then compared to their corresponding actual positions in the top-view image. Both sets of points were manually annotated, as shown in the Mesh Homography technique in Table 6.

Table 7 presents the re-projection error for each camera. The Euclidean distance between the transformed and actual points determines the re-projection error, which indicates how well the transformation aligns the camera view with the top-view image. The re-projection error is calculated using the following formula:

$$E_{Re-proj} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_{Transf_i} - P_{Dest_i})^2} \quad (12)$$

where P_{Transf_i} represents the transformed points while P_{Dest_i} refers to the actual destination point on the BEV map, and n is the number of projections in a particular scene.

The re-projection error reflects the accuracy of the corresponding annotations on the camera images and BEV map. A lower error indicates precise annotation, while a higher error suggests less accurate annotation. In Table 7, the average re-projection error is between 0 and 3 pixels, which is satisfactory, meaning any projections deviate only slightly. Cameras A and C exhibit slightly higher re-projection errors than cameras B and D, due to minor deviations in pixel annotation. Overall, the annotations of corresponding points on the camera images and BEV map are highly accurate.

Additionally, we found that projections of the same cattle from different cameras often show discrepancies due to varying viewpoints, making the fusion process challenging. Cameras with similar orientations (e.g., north-facing or south-facing) are generally more effective at fusing projections compared to combinations of north- and south-facing cameras. However, fusing projections from north- and south-facing cameras is prone to failure due to the limited size of the ROI and variations in the cattle's ground points when viewed from different perspectives.

To quantify these discrepancies, Table 8 presents the average Euclidean distance between the projections of the same cattle as captured by different cameras. Projection filters are applied in this analysis: "ALL" includes results from all overlapping cameras, while "SPM" highlights only the optimal cameras selected for fusion in specific segments. It is worth noting that the average Hanwoo cattle spans approximately 80 pixels on the BEV map. Therefore, a

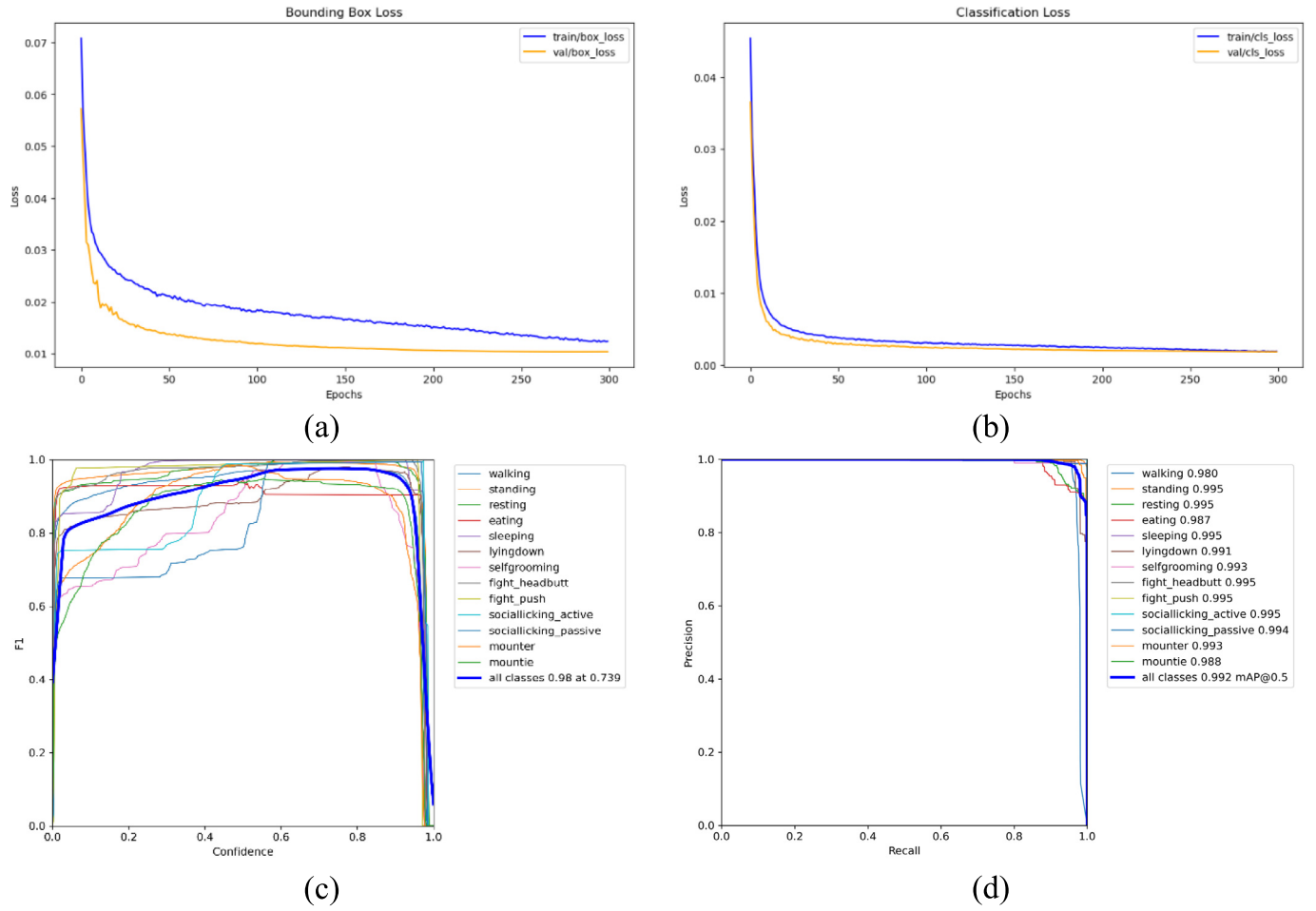


Fig. 9. Training curves of the action detector for the multi-camera dataset. (a) Bounding box loss, (b) Classification loss, (c) F1-confidence curve, and (d) Precision-recall (PR) curve for individual actions.

higher distance often indicates that the ground point is predicted near the front or back limbs, with this specific factor being the primary source of the discrepancy.

4.2.3.3. Fusion and projection scenarios. After precise mapping, the fusion process is required to combine multiple projections into finalized points. This process involves several stages. Fig. 11 shows the cattle detections obtained from the MMIU across all cameras in Scenario 1. The

Enhanced BEV model, with its homography matrices prepared, maps all detections onto the BEV plane, as depicted in Fig. 12(a). This process initially resulted in 44 projections, far exceeding the ground truth of 17 cattle.

To filter out redundant projections, the SPM filter was applied in its 2nd iteration, reducing the number of projections, as shown in Fig. 12 (b). Next, regions of interest (ROIs) with a radius of 30 pixels were generated for the remaining projections, depicted in Fig. 12(c). Overlapping

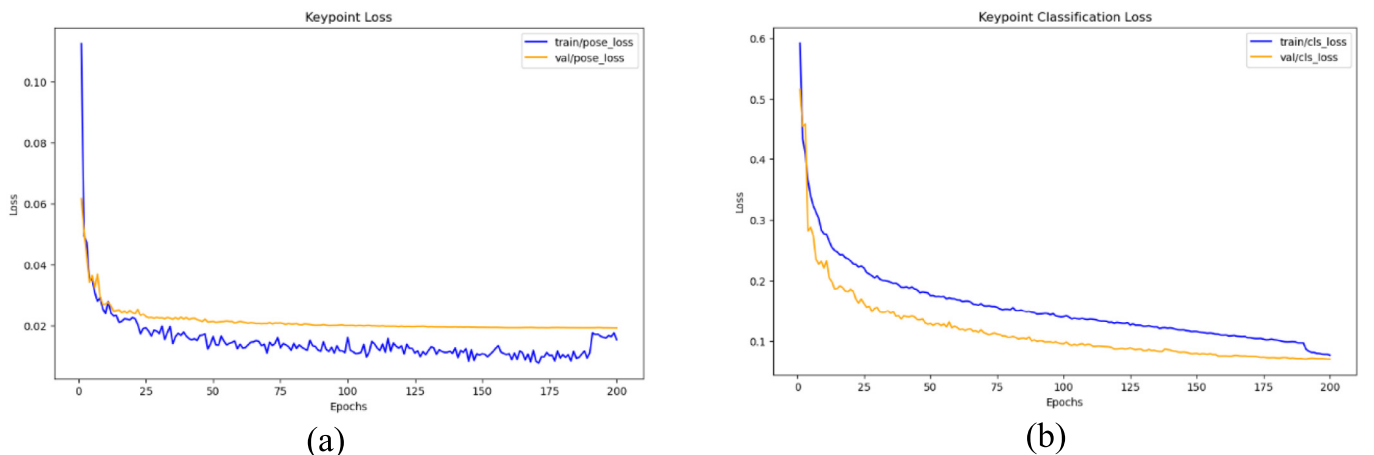


Fig. 10. Pose estimation loss analysis. (a) Keypoint loss (K_p), (b) Keypoint classification loss (K_{ck}).

Table 6
Comparison of homography estimation techniques. Evaluation of various homography estimation techniques based on labeling time, number of feature points, and projection error. Datasets A, B, C, and D correspond to different experimental setups.

Technique	Time to label feature points (s)				Number of Feature Points				Projection Error (Mean Square Error, px)			
	A	B	C	D	A	B	C	D	A	B	C	D
Perspective Transformation (Corner)	30	20	30	30	8	8	8	8	121	462.3	82.50	315.5
Columns Only	60	30	60	60	14	8	12	18	92.6	435.1	61.33	224.8
All Corner Columns + Fence	180	150	180	360	90	44	62	70	77.1	372.5	41.09	81.54
Mesh (SPM based)	900	800	900	950	448	226	360	422	–	–	–	–

cases were analyzed using authentication rules to establish connections, as shown in Fig. 12(d). Fused projection points, represented in yellow and labeled as “Fused,” were generated, while independent projections without overlapping ROIs were retained in green and labeled as “Ind” (independent). Unfortunately, basic fusion introduced several errors, detailed in Table 9 and visualized in Fig. 12(e).

Following the fusion pipeline outlined in Fig. 6, no conflict resolution alert (C_R) was triggered. However, the number of detected cattle (DET_{cattle}) was 24, exceeding the tolerance threshold ($\hat{T}$), resulting in a high value. This required a reset of the SPM filter. Adjustments were made to reduce reliance on partial and complete view cameras, focusing instead on Best View cameras. The 3rd iteration of the SPM filter was implemented, producing the results shown in Fig. 12(f).

After these modifications, the process was repeated, resulting in 18 projections. However, due to occlusion and ROI limitations, two projections could not form connections and remained unfused, leading to one incorrect independent cattle projection, as depicted in Fig. 12(h). These results represent a significant improvement over the initial outcomes. A detailed breakdown of Scenario 1 is provided in Table 9.

Table 9 provides a detailed breakdown of various scenarios. In Scenario 1, depicted in Fig. 12, the initial 44 projections from the first SPM iteration were reduced to 30 following the second iteration. After applying basic fusion, this number was further refined to 24 projection points, compared to the actual ground truth of 17 cattle. Analysis of the basic fusion results revealed an E_{Miss} value of 5, indicating that five connections were missed during the ROI connectivity process, leading to independent projection points. Additionally, the E_{IMC} value was 2, reflecting two erroneous multi-connections.

To address these discrepancies, feedback-driven adjustments were made, and the third SPM iteration filter was applied. This enhanced fusion process resulted in 18 projection points, with E_{IMC} reduced to 0 and E_{Miss} minimized to 1. As shown in Fig. 12 (g-h), two projections, despite intersecting each other's ROIs, failed to connect due to ROI shortfall. While minor challenges persisted, this example demonstrates the effectiveness of the fusion module pipeline.

In real-world scenarios, a variety of challenges arise, requiring additional processing to resolve conflicts and generate precise projections. Scenarios 1–7 in Table 9 represent typical challenges, while Scenarios 8–10 involve extraordinary circumstances with heavy occlusion, which degrade detector performance, resulting in fewer detected cattle and fewer projections than the ground truth. The challenges faced in all these scenarios are summarized below:

- Detection Issues:** Incorrect detections by the action detector and false negatives lead to detection errors. Blind spots, particularly in Scenarios 8–10, often result in undetected cattle.
- SPM Limitations:** Multiple cattle transitioning between segments can strain the SPM, increasing E_{IMC} and generating missed connections E_{Miss} due to partial views and occlusion.

Table 7
Re-projection error for individual cameras in the multi-camera setup.

Camera	A	B	C	D
Re-projection Error	2.386	0.771	2.776	0.93

- Segment-Specific Challenges:** In Segment 1, occlusions often cause ground points to be slightly elevated on the cattle's body, leading projections to extend beyond the operational area of the BEV map. These issues are particularly evident during herding or mounting activities near fences.
- Fusion Limitations:** Complex cases may lead to incorrect connections due to conflicts or proximity issues, resulting in both missed and erroneous projections. Additionally, projections from north- and south-facing cameras often exhibit bias, as documented in Table 8.
- ROI Shortfall:** ROI shortfall occurs when projections of the same cattle from different cameras fail to overlap, preventing connections. While increasing the ROI radius can mitigate this, it also raises E_{IMC} and E_{Miss} values for other projections, requiring careful balance to optimize performance.

To address these challenges, the fusion pipeline modifies the SPM filter, reducing projections and relying more on cameras with the best views for each segment. This strategy significantly reduces E_{IMC} and E_{Miss} values, as shown in Table 9. For complex connections, Methodology 1 (Fig. 8) is employed, effectively resolving conflicts one by one, as illustrated in Fig. A2 (Appendix). In highly challenging cases, the optimal approach involves selecting the best view cameras for each segment and minimizing reliance on fusion. While this approach may result in slight deviations from the ground truth, it avoids additional errors. This final iteration of the SPM filter, however, relies on single view-points or cameras, which introduces limitations.

The best results were achieved when all modules worked harmoniously. Deep learning models must accurately detect cattle and generate precise ground points, the MMIU must forward reliable results, and the Enhanced BEV algorithm must produce accurate cattle positions. Finally, basic and advanced fusion techniques should generate projection points that closely match the ground truth, with minimal errors.

4.3. Discussion

The proposed study presents a robust framework for continuous cattle monitoring and surveillance, enabling detailed analysis of individual trajectories and behavior, as illustrated in Fig. A4 (Appendix). By integrating Bird's Eye View (BEV) techniques, the system effectively mitigates occlusion challenges inherent to conventional camera setups. BEV provides a comprehensive top-down perspective, ensuring unobstructed visibility of all animals within the barn and supporting uninterrupted behavior tracking. This facilitates the development of a rich behavioral dataset, capturing cattle positions, movement patterns, and activity levels over time.

While multi-camera systems have been widely explored in the context of human surveillance and public datasets (Amosa et al., 2023), their application in livestock environments remains limited. (Amosa et al., 2023) emphasize persistent challenges in Multi-Camera Multi-Target Tracking (MCMTT), including cross-camera re-identification, occlusion handling, and trajectory continuity. Unlike human surveillance studies, typically conducted in controlled environments with well-annotated datasets, real-world livestock settings are characterized by uncontrolled lighting conditions, dynamic social interactions, and

Table 8
Analysis of cattle projections relative to North- and South-facing cameras. A, B, C, and D represent the corresponding datasets. The notations X and Y in the column are replaced by the respective cameras.

Camera Bias	Projection Filter		Overlapping Camera Bias (Euclidean Distance, px / Camera X / Camera Y)				
	ALL	SPM	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
North – North (A, B)	✓	✓	–	–	–	28.7/A/B 28.7/A/B	59.94/A/B –
North – South (A, B, C, D)	✓	–	–	–	–	100.25/B/D	83.55/B/D
	✓	–	–	58.13/A/D	69.42/A/D	66.86/A/D	46.51/A/D
	✓	✓	–	31.31/A/C	61.21/A/C	–	–
	–	–	–	–	61.21/A/C	–	–
South – South (C, D)	✓	–	47.4/C/D	18.9/C/D	–	–	–
	–	✓	–	–	–	–	–

irregular movement patterns. These complexities limit the direct transferability of existing MCMT techniques. Our approach addresses these challenges by employing a multi-camera BEV projection strategy that minimizes the need for re-identification across views, focusing instead on continuity in ground-projected tracking.

Another important consideration in multi-camera systems is the evaluation methodology. Metrics used in prior studies are often inconsistent and tailored to specific benchmark conditions, failing to capture the nuanced performance demands of real-world agricultural environments. In contrast, our evaluation adopts a more application-oriented perspective, focusing on practical system performance indicators (Table 9). These include re-projection error, fusion ID mismatches, missed connection errors, incorrect multi-connection errors, and overall fusion consistency. Although not yet standardized across the broader MCMT field, these metrics are better suited for assessing effectiveness in real-time, open-set scenarios such as livestock barns. We prioritize usability and interpretability in operational settings over optimization for benchmark-specific metrics with limited real-world relevance.

Through statistical analysis of tracking data, the system also enables the identification of anomalous behaviors, supporting timely interventions, enhanced welfare strategies, and early disease detection. Nonetheless, limitations persist. As reported in broader MCMT research

(Amosa et al., 2023; Xu et al., 2016; Zhang et al., 2020) fusion algorithms can face difficulties in complex settings, occasionally resulting in tracking inconsistencies. These challenges become more pronounced during rapid or intricate cattle movements, where maintaining identity continuity can be difficult. Additionally, novel or unforeseen behaviors not represented in the training data may impair the detector's performance, underscoring the complexity of real-world deployment. While our multi-camera BEV-based approach addresses many of these issues, completely eliminating them remains a challenge.

Future work will focus on refining fusion algorithms and expanding behavioral monitoring capabilities. Our long-term objective is to establish reliable behavioral patterns that serve as early indicators of stress or illness, contributing to smarter, welfare-oriented farm management.

5. Conclusion

This paper introduced a comprehensive methodology utilizing deep learning models and fusion techniques to analyze Hanwoo cattle behaviors. The system integrated an action detector, which achieved a detection accuracy of 99.2 %, and a pose estimation model that enhanced the identification of cattle ground points. The

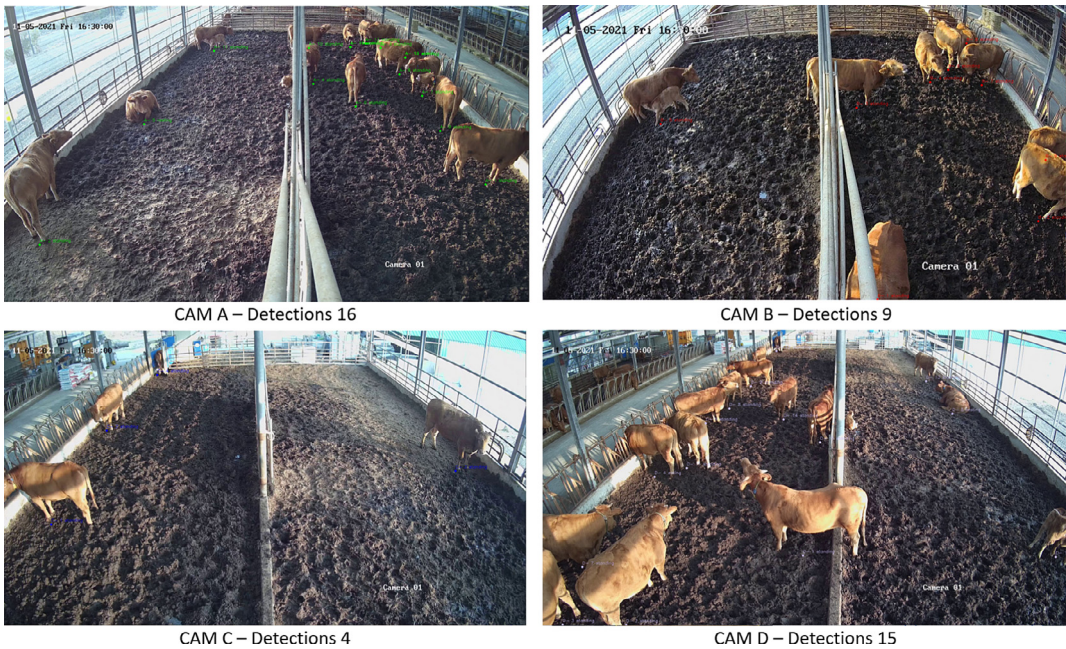


Fig. 11. Cattle detections on camera image frames. Visualization of cattle detections on individual camera image frames, including the number of detections per frame. Best viewed in color; zoom in for detailed observations.

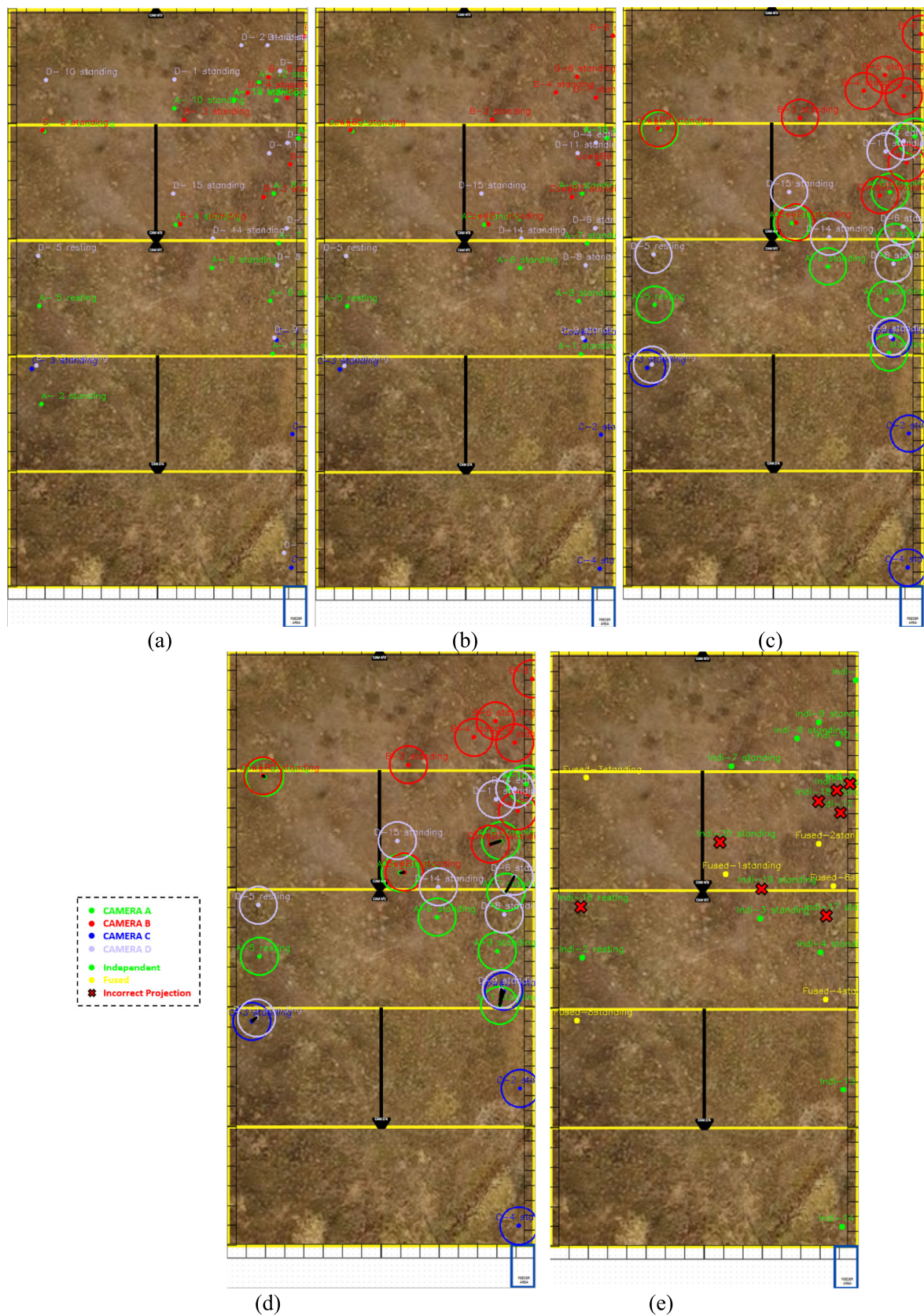


Fig. 12. BEV map showing the step-by-step fusion process. Illustration of the step-by-step process from basic to advanced fusion on the BEV map: (a) Initial mapping of all projections, (b) Application of the second SPM iteration filter, (c) ROI generation, (d) Establishment of connections using authentication rules, (e) Merging of projection points to produce basic fusion results, (f) Error mitigation through the third SPM iteration filter, (g) ROI generation and connection establishment via authentication rules, and (h) Finalized BEV projection points after advanced fusion.

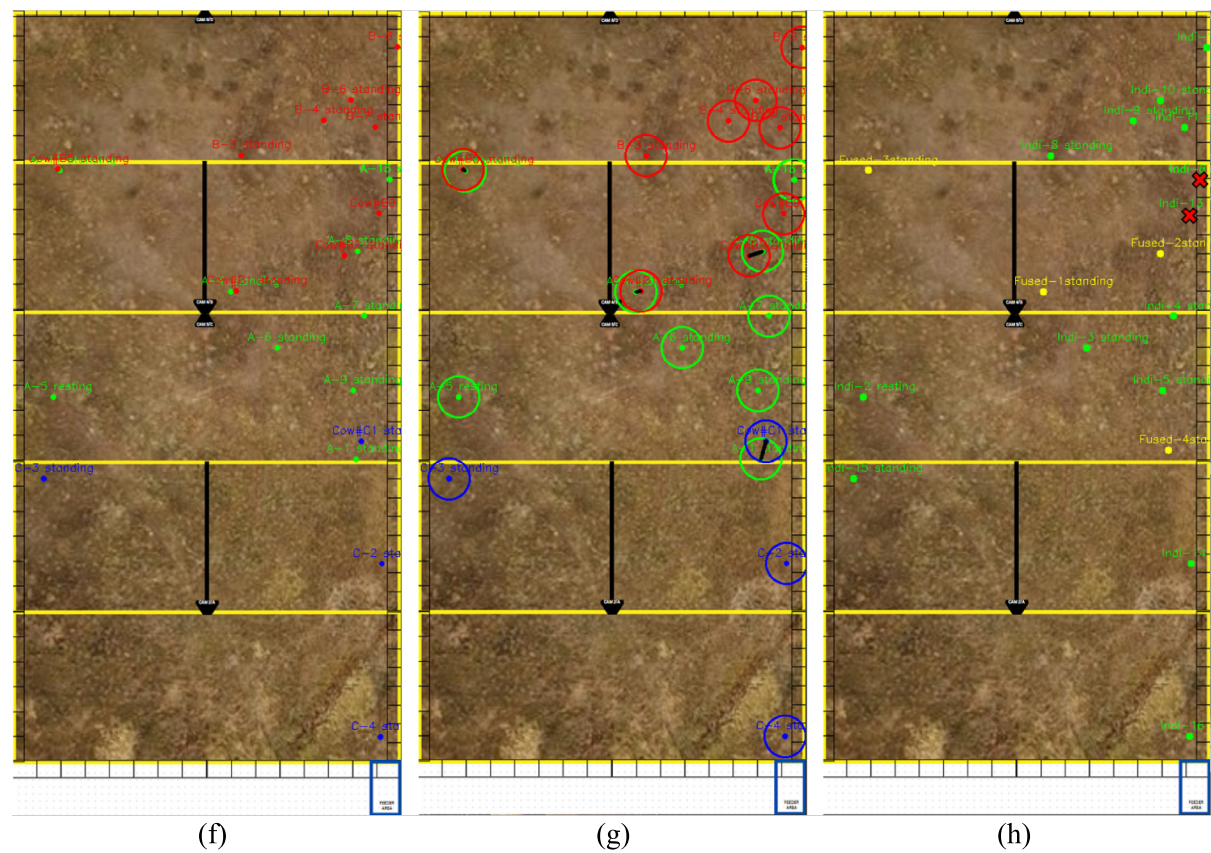


Fig. 12 (continued).

Table 9 Projections for multiple scenarios across fusion steps. Details of projections for multiple scenarios during different steps of fusion. ↓ denotes that a lower value is better.										
Scenarios in Barn	All Cattle Projections	Segment Preference Method (SPM)	Basic Fusion Results	Groundtruth (Number of Cattle)	Alerts for Conflict Resolution	Basic Fusion		Advance Final Fusion (Number of cattle)	Final Fusion	
						E _{IMC} ↓	E _{Miss} ↓		E _{IMC} ↓	E _{Miss} ↓
1	44	30	24	17	–	0	7	18	0	1
2	41	25	21	17	–	0	2	18	0	0
3	44	30	23	17	2	1	3	17	1	0
4	39	23	19	17	–	0	2	17	0	0
5	43	17	17	17	–	0	1	16	0	0
6	42	21	18	17	2	2	0	16	0	0
7	31	23	16	17	–	1	1	16	0	0
8	33	18	15	17	–	0	1	14	0	1
9	45	18	15	17	–	0	2	14	0	0
10	31	16	14	17	–	0	0	14	0	0

outputs from these models were fused and processed through the Enhanced Bird's Eye View algorithm, which accurately mapped cattle positions from the 3D environment onto a 2D BEV plane. A significant number of projections were successfully mapped onto the BEV plane. To address complex scenarios and obstructions, novel methods such as the Segment Preference Method (SPM) and innovative fusion techniques were employed, ensuring precise cattle positioning. Despite the inherent challenges of real-farm environments, which are prone to inaccuracies, this research demonstrated a robust fusion procedure capable of handling diverse scenarios. This methodology offers a non-intrusive, scalable solution for precision livestock farming, enabling continuous cattle monitoring, individual behavioral analysis, and improved farm management. Furthermore, the proposed system has

potential applications beyond Hanwoo cattle, as it can be adapted to monitor and map the behavior and positions of various livestock in enclosed environments.

CRediT authorship contribution statement

Muhammad Fahad Nasir: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Alvaro Fuentes:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shujie Han:** Investigation, Data curation. **Jiaqi Liu:** Investigation, Data curation. **Yongchae Jeong:** Supervision, Methodology. **Sook Yoon:** Writing – review & editing, Funding acquisition, Formal analysis,

Conceptualization. **Dong Sun Park:** Supervision, Methodology, Funding acquisition, Formal analysis.

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.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. RS-2019-NR040079); and by the NRF grants funded by the Korea Government (MSIT) (2020R1A2C2013060) and (RS-2024-00392406).

Appendix A

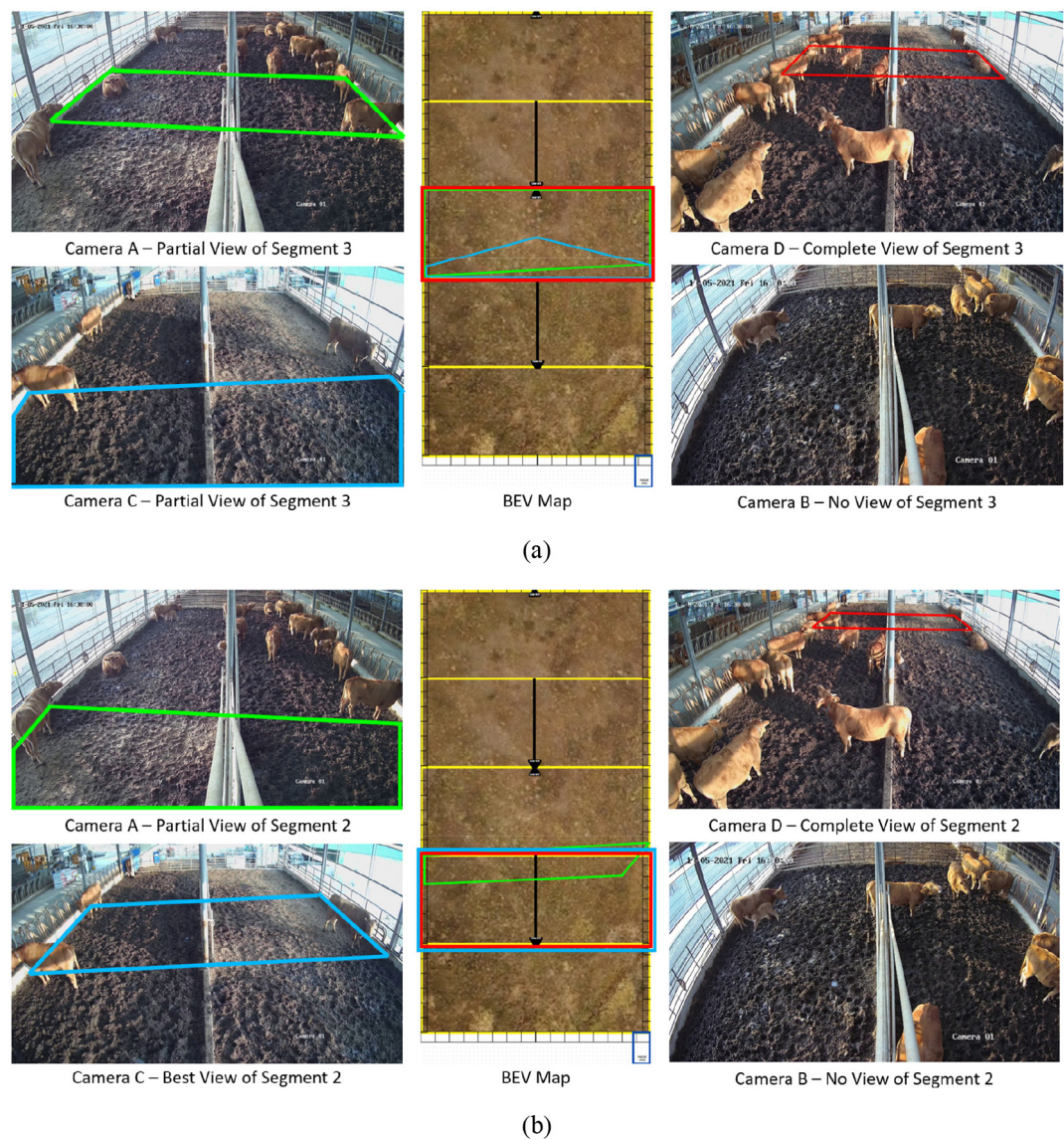


Fig. A1. Segment coverage from all camera FOVs onto the BEV Map. (a) Partial view focuses on unoccluded and clear cattle projections, offering closer perspectives compared to the complete view. (b) Complete view spans the entire segment and qualifies as the best view only when it provides the closest and least occluded perspective. For example, cameras C and D both offer complete views; however, camera C provides a clearer, unobstructed view of the segment compared to camera D. (Best viewed in color).

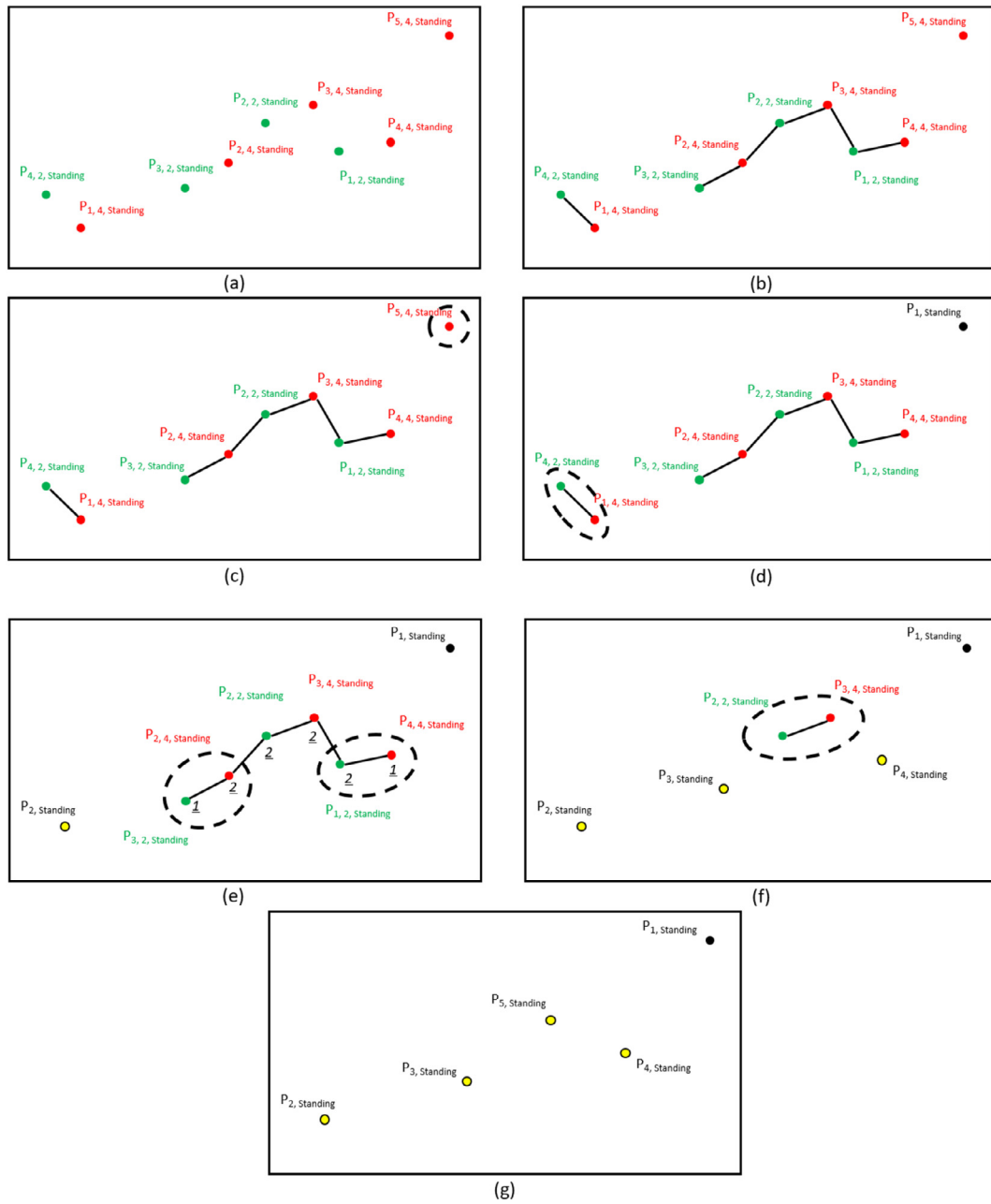


Fig. A2. Advanced fusion techniques for challenging scenarios. The BEV map illustrates a challenging real-world scenario from segment 5 encountered during the research: (a) Initial mapping of all projections, (b) Formation of connections based on authentication rules, (c) Identification of projection points without connections, (d) Detection and fusion of non-conflicting connections, (e) Evaluation of remaining connections, with the minimum sum prioritized for fusion, (f) Generation of final fused projection points, and (g) Fusion of all connections with newly assigned IDs.

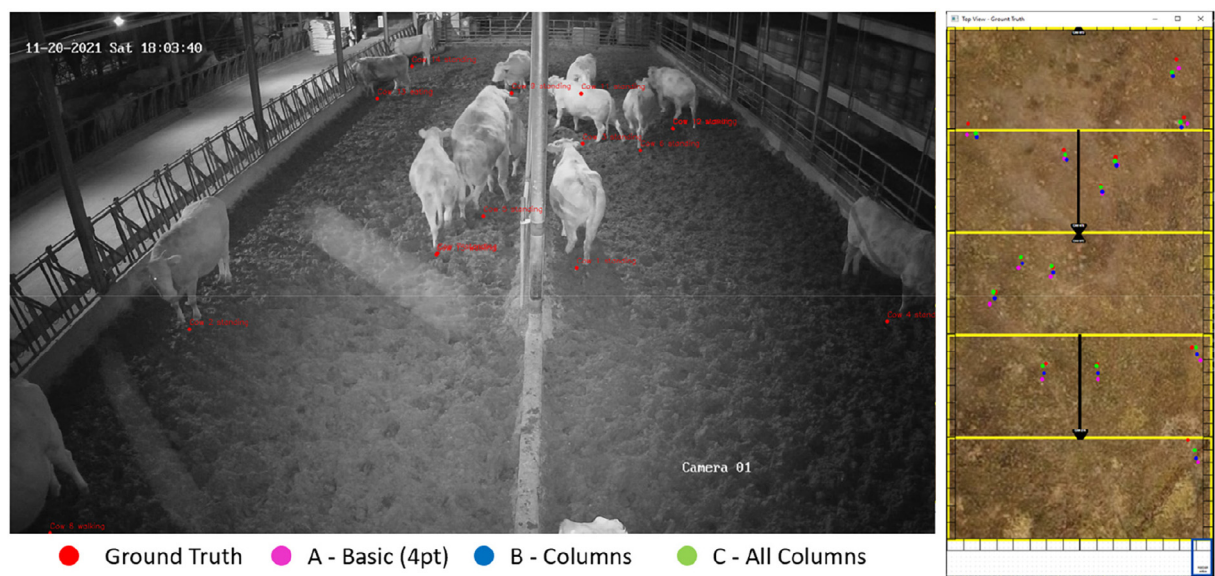


Fig. A3. BEV map of cattle projections using different homography techniques. The analysis is based on nighttime footage captured by camera D.

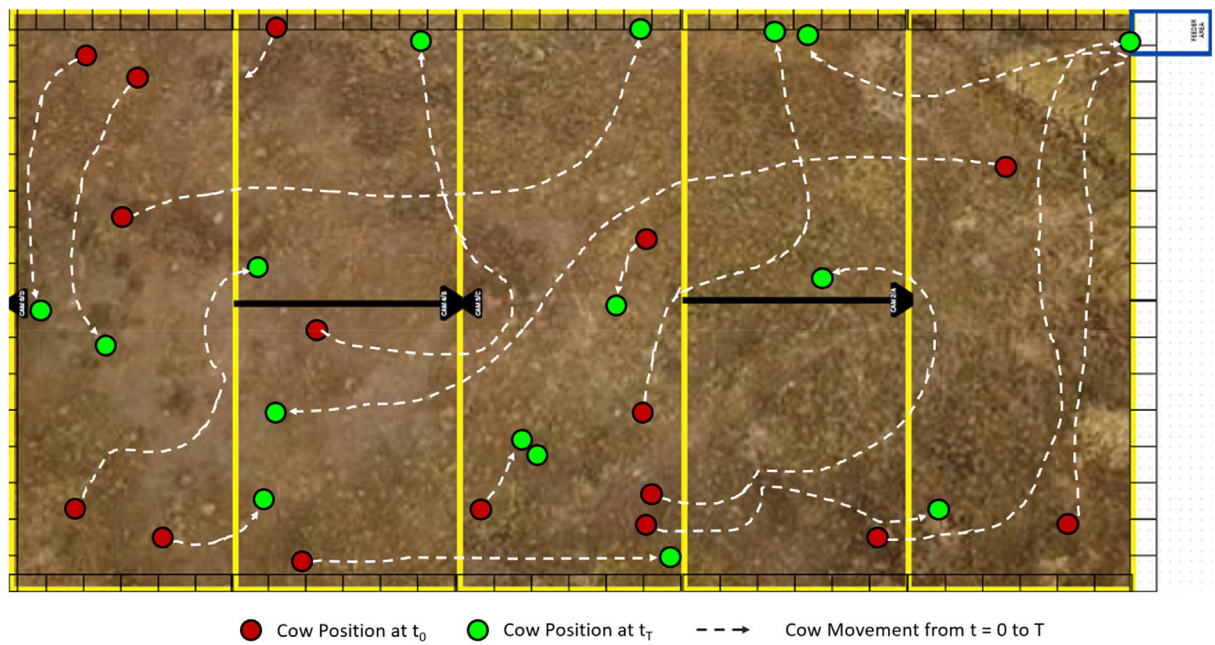


Fig. A4. Cattle BEV positioning system. This system tracks individual cattle positions over time while the action detector provides simultaneous information on cattle behaviors. The primary objective is to monitor cattle movements and behaviors within the enclosure, enabling a precise assessment of their health and well-being.

Data availability

The datasets used in this study are not publicly available as the research is part of an ongoing project funded by the Korean Government. In accordance with the project agreement, the dataset will remain restricted until the project's conclusion. For inquiries regarding data access, please contact Alvaro Fuentes, afuentes@jbnu.ac.kr.

References

- Amosa, T.I., Sebastian, P., Izhar, L.I., Ibrahim, O., Ayinla, L.S., Bahashwan, A.A., Bala, A., Samaila, Y.A., 2023. Multi-camera multi-object tracking: A review of current trends and future advances. *Neurocomputing* 552, 126558. doi:[10.1016/j.neucom.2023.126558](https://doi.org/10.1016/j.neucom.2023.126558).
- Ayadi, S., Ben Said, A., Jabbar, R., Aloulou, C., Chabbouh, A., Achballah, A. Ben, 2020. Dairy cow rumination detection: A deep learning approach. *Commun. Comput. Inf. Sci.* 1348. doi:[10.1007/978-3-030-65810-6_7](https://doi.org/10.1007/978-3-030-65810-6_7).
- Chung, K.Y., Lee, S.H., Cho, S.H., Kwon, E.G., Lee, J.H., 2018. Current situation and future prospects for beef production in South Korea – A review. *Asian Australas. J. Anim. Sci.* 31 (7). doi:[10.5713/ajas.18.0187](https://doi.org/10.5713/ajas.18.0187).
- Du, A., Guo, H., Lu, J., Su, Y., Ma, Q., Ruchay, A., Marinello, F., Pezzuolo, A., 2022. Automatic livestock body measurement based on keypoint detection with multiple depth cameras. *Comput. Electron. Agric.* 198, 107059. doi:[10.1016/j.compag.2022.107059](https://doi.org/10.1016/j.compag.2022.107059).
- 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. doi:[10.1016/j.compag.2020.105627](https://doi.org/10.1016/j.compag.2020.105627).
- Fuentes, A., Han, S., Nasir, M.F., Park, J., Yoon, S., Park, D.S., 2023. Multiview monitoring of individual cattle behavior based on action recognition in closed barns using deep learning. *Animals* 13 (12). doi:[10.3390/ani13122020](https://doi.org/10.3390/ani13122020).
- Guan, H., Motohashi, N., Maki, T., Yamaai, T., 2020. Cattle identification and activity recognition by surveillance camera. *IS&T Int. Symp. Electron. Imaging Sci. Technol.* 2020 (12). doi:[10.2352/ISSN.2470-1173.2020.12.FAIS-174](https://doi.org/10.2352/ISSN.2470-1173.2020.12.FAIS-174).
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S., 2016. Deep learning for visual understanding: a review. *Neurocomputing* 187. doi:[10.1016/j.neucom.2015.09.116](https://doi.org/10.1016/j.neucom.2015.09.116).
- Halachmi, I., Guarino, M., Bewley, J., Pastell, M., 2019. Smart animal agriculture: application of real-time sensors to improve animal well-being and production. *Annu. Rev. Anim. Biosci.* 7. doi:[10.1146/annurev-animal-020518-114851](https://doi.org/10.1146/annurev-animal-020518-114851).
- Han, S., Fuentes, A., Yoon, S., Jeong, Y., Kim, H., Sun Park, D., 2023. Deep learning-based multi-cattle tracking in crowded livestock farming using video. *Comput. Electron. Agric.* 212. doi:[10.1016/j.compag.2023.108044](https://doi.org/10.1016/j.compag.2023.108044).
- Hartley, R., Zisserman, A., 2004. Multiple view geometry in computer vision. *Multiple View Geometry in Computer Vision*. doi:[10.1017/cbo9780511811685](https://doi.org/10.1017/cbo9780511811685).
- Hu, H., Hachiuma, R., Saito, H., Takatsume, Y., Kajita, H., 2022. Multi-camera multi-person tracking and re-identification in an operating room. *J. Imaging* 8 (8). doi:[10.3390/jimaging8080219](https://doi.org/10.3390/jimaging8080219).
- Jocher, G., 2023. *Ultralytics/yolov5: Yolov5 in PyTorch > ONNX > CoreML > TFLite*. GitHub.
- Kim, N.Y., Kim, S.J., Jang, S.Y., Oh, M.R., Tang, Y.J., Seong, H.J., Yun, Y.S., Moon, S.H., 2017. Behavioral characteristics of Hanwoo (*Bos taurus coreanae*) steers at different growth stages and seasons. *Asian Australas. J. Anim. Sci.* 30 (10). doi:[10.5713/ajas.16.0992](https://doi.org/10.5713/ajas.16.0992).
- Kleanthous, N., Hussain, A.J., Khan, W., Sneddon, J., Al-Shamma'a, A., Liatsis, P., 2022. A survey of machine learning approaches in animal behaviour. *Neurocomputing* 491. doi:[10.1016/j.neucom.2021.10.126](https://doi.org/10.1016/j.neucom.2021.10.126).
- Kleen, J.L., Guatteo, R., 2023. Precision livestock farming: what does it contain and what are the perspectives? *Animals* 13 (5). doi:[10.3390/ani13050779](https://doi.org/10.3390/ani13050779).
- Lima, J.P., Roberto, R., Figueiredo, L., Simões, F., Teichrieb, V., 2021. Generalizable multi-camera 3D pedestrian detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- López-Cifuentes, A., Escudero-Viñolo, M., Bescós, J., Carballeira, P., 2018. Semantic Driven Multi-Camera Pedestrian Detection. <http://arxiv.org/abs/1812.10779>.
- Mahmud, M.S., Zahid, A., Das, A.K., Muzammil, M., Khan, M.U., 2021. A systematic literature review on deep learning applications for precision cattle farming. *Comput. Electron. Agric.* 187. doi:[10.1016/j.compag.2021.106313](https://doi.org/10.1016/j.compag.2021.106313).
- Mar, C.C., Zin, T.T., Tin, P., Honkawa, K., Kobayashi, I., Horii, Y., 2023. Cow detection and tracking system utilizing multi-feature tracking algorithm. *Sci. Rep.* 13 (1), 1–17. doi:[10.1038/s41598-023-44669-4](https://doi.org/10.1038/s41598-023-44669-4).
- Meng, Y., Yoon, S., Han, S., Fuentes, A., Park, J., Jeong, Y., Park, D.S., 2023. Improving known-unknown cattle's face recognition for smart livestock farm management. *Animals* 13 (22), 3588. doi:[10.3390/ANI13223588](https://doi.org/10.3390/ANI13223588).
- Montero, D., Aranjuelo, N., Leskovsky, P., Loyo, E., Nieto, M., Aginako, N., 2023. Multi-camera BEV video-surveillance system for efficient monitoring of social distancing. *Multimed. Tools Appl.* 82 (22). doi:[10.1007/s11042-023-14416-y](https://doi.org/10.1007/s11042-023-14416-y).
- Myat Noe, S., Zin, T.T., Tin, P., Kobayashi, I., 2023. Comparing state-of-the-art deep learning algorithms for the automated detection and tracking of black cattle. *Sensors* 23 (1), 532. doi:[10.3390/S23010532](https://doi.org/10.3390/S23010532).
- Myat Noe, S., Zin, T.T., Kobayashi, I., Tin, P., 2025. Optimizing black cattle tracking in complex open ranch environments using YOLOv8 embedded multi-camera system. *Sci. Rep.* 15 (1), 1–24. doi:[10.1038/s41598-025-91553-4](https://doi.org/10.1038/s41598-025-91553-4).
- Nasir, M.F., Fuentes, A., Han, S., Park, J., Yoon, S., Park, D.S., 2023, February 20. *Cattle Action Recognition with Multi-Viewpoint Cameras based on Deep Learning*. *International Workshop on Frontiers of Computer Vision*.
- Nguyen, T., Chen, S.W., Shivakumar, S.S., Taylor, C.J., Kumar, V., 2018. Unsupervised deep Homography: A fast and robust Homography estimation model. *IEEE Robot. Autom. Lett.* 3 (3). doi:[10.1109/LRA.2018.2809549](https://doi.org/10.1109/LRA.2018.2809549).
- Qiao, Y., Guo, Y., He, D., 2023. Cattle body detection based on YOLOv5-ASFF for precision livestock farming. *Comput. Electron. Agric.* 204. doi:[10.1016/j.compag.2022.107579](https://doi.org/10.1016/j.compag.2022.107579).
- Rath, S., 2023. YOLOv8 Ultralytics: State-of-the-Art YOLO Models. <https://Learnopencv.Com/Ultralytics-Yolov8/>.
- Saponara, S., Elhanashi, A., Zheng, Q., 2022. Developing a real-time social distancing detection system based on YOLOv4-tiny and bird-eye view for COVID-19. *J. Real-Time Image Proc.* 19 (3). doi:[10.1007/s11554-022-01203-5](https://doi.org/10.1007/s11554-022-01203-5).
- Sumi, K., Zin, T.T., Kobayashi, I., Horii, Y., 2017. A study on cow monitoring system for calving process. 2017 IEEE 6th global conference on consumer electronics, GCCE 2017. doi:[10.1109/GCCE.2017.8229442](https://doi.org/10.1109/GCCE.2017.8229442) 2017-January, 1–2.
- Szenci, O., 2022. Accuracy to predict the onset of calving in dairy farms by using different precision livestock farming devices. *Animals* 12 (15). doi:[10.3390/ani12152006](https://doi.org/10.3390/ani12152006).
- Tullo, E., Finzi, A., Guarino, M., 2019. Review: Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy. *Science of The Total Environment* 650, 2751–2760. doi:[10.1016/j.scitotenv.2018.10.018](https://doi.org/10.1016/j.scitotenv.2018.10.018).
- Tuytens, F.A.M., Molento, C.F.M., Benaissa, S., 2022. Twelve threats of precision livestock farming (PLF) for animal welfare. *Front. Vet. Sci.* 9. doi:[10.3389/fvets.2022.889623](https://doi.org/10.3389/fvets.2022.889623).
- Tzanidakis, C., Tzamaloukas, O., Simitzis, P., Panagakos, P., 2023. Precision livestock farming applications (PLF) for grazing animals. *Agriculture (Switzerland)* 13 (2). doi:[10.3390/agriculture13020288](https://doi.org/10.3390/agriculture13020288).
- Wisniewski, L., Norby, B., Pierce, S.J., Becker, T., Sordillo, L.M., 2019. Prospects for predictive modeling of transition cow diseases. *Anim. Health Res. Rev.* doi:[10.1017/S1466252319000112](https://doi.org/10.1017/S1466252319000112).
- Xu, Y., Liu, X., Liu, Y., Zhu, S.C., 2016. Multi-view people tracking via hierarchical trajectory composition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, pp. 4256–4265. doi:[10.1109/CVPR.2016.461](https://doi.org/10.1109/CVPR.2016.461).
- Yamamoto, Y., Akizawa, K., Aou, S., Taniguchi, Y., 2025. Entire-barn dairy cow tracking framework for multi-camera systems. *Comput. Electron. Agric.* 229, 109668. doi:[10.1016/j.compag.2024.109668](https://doi.org/10.1016/j.compag.2024.109668).
- Yan, Y., Xu, M., Smith, J.S., Shen, M., Xi, J., 2021. Multicamera pedestrian detection using logic minimization. *Pattern Recogn.* 112, 107703. doi:[10.1016/j.patcog.2020.107703](https://doi.org/10.1016/j.patcog.2020.107703).
- Zhang, R., Wu, L., Yang, Y., Wu, W., Chen, Y., Xu, M., 2020. Multi-camera multi-player tracking with deep player identification in sports video. *Pattern Recogn.* 102, 107260. doi:[10.1016/j.patcog.2020.107260](https://doi.org/10.1016/j.patcog.2020.107260).
- Zheng, Z., Zhang, X., Qin, L., Yue, S., Zeng, P., 2023. Cows' legs tracking and lameness detection in dairy cattle using video analysis and Siamese neural networks. *Comput. Electron. Agric.* 205. doi:[10.1016/j.compag.2023.107618](https://doi.org/10.1016/j.compag.2023.107618).