BECA: A Computer Vision Dataset for Long-Term Recognition In Beef Cattle

Yuqi Zhang[1,†], Longxiang Li[1,†], Chunyang Li[1], Sen Wang[1], Yue Rong[2], Kai Niu[1], and Zhiqiang He[1,*]

[1]Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China
[2]School of Electrical Engineering, Computing and Mathematical Sciences Curtin University, Bentley, WA 6102, Australia
[*]Corresponding author: Zhiqiang He (hezq@bupt.edu.cn)
[†]These authors contributed equally to this work.

Abstract

Despite significant progress in computer vision for precision livestock farming, a critical challenge persists: the lack of effective long-term cattle re-identification datasets under real-world conditions, primarily due to substantial phenotypic changes. To address this challenge, we present the Beef Cattle dataset (BECA), a novel, large-scale dataset specifically designed to support long-term and diverse cattle recognition from dorsal views. BECA encompasses two sub-datasets: the Beef Cattle dataset for Diversity (BECA-D), containing 16,889 images from 5,661 beef cattle across multiple breeds, designed to capture visual diversity for recognition; and the Beef Cattle dataset for Long-term recognition (BECA-L), comprising 12,172 annotated images from 103 cattle tracked over a continuous period of up to five months—representing a notably long duration for cattle recognition datasets. Additionally, we provide annotation subsets for key pipeline tasks such as object detection and pose estimation, supporting the development of diverse models for long-term recognition. Collectively, BECA establishes a comprehensive benchmark for vision-based livestock management, facilitating research in cattle identification, behavior analysis, and welfare monitoring.

Background & Summary

Large-scale and centralized growing and fattening of beef cattle is essential for societal development [1]. Large-scale farming can reduce costs and save space while meeting the growing demands for increased production [2]. Precision livestock farming, which leverages modern technologies to improve efficiency, animal welfare, and sustainability, is crucial in this context. As the dairy industry becomes more intelligent and streamlined [3, 4, 5, 6], similar advancements in the beef cattle industry deserve attention.

The dairy industry offers valuable reference and comparison points. Large-scale dairy farming is characterized by high profit, numerous metrics, and uniform breeds. The main product is milk, and by-products include dairy bulls and culled cows. The focus in dairy farming is on managing the estrus and lactation periods [9, 10], with dairy cows requiring nearly 300 days of milking each year, using specialized lanes. These specialized facilities also enable systematic dorsal data collection, as evidenced by datasets like OpenCows2020[21] and Cows2021[22]. This is significantly different from the focus of the beef cattle industry. Beef cattle breeds mainly include Simmental and Angus [11], and fattening primarily involves bulls. Once grouped, the cattle are fed and fattened in confined areas under group management, with the main by-product being the fattened beef cattle [12, 13].

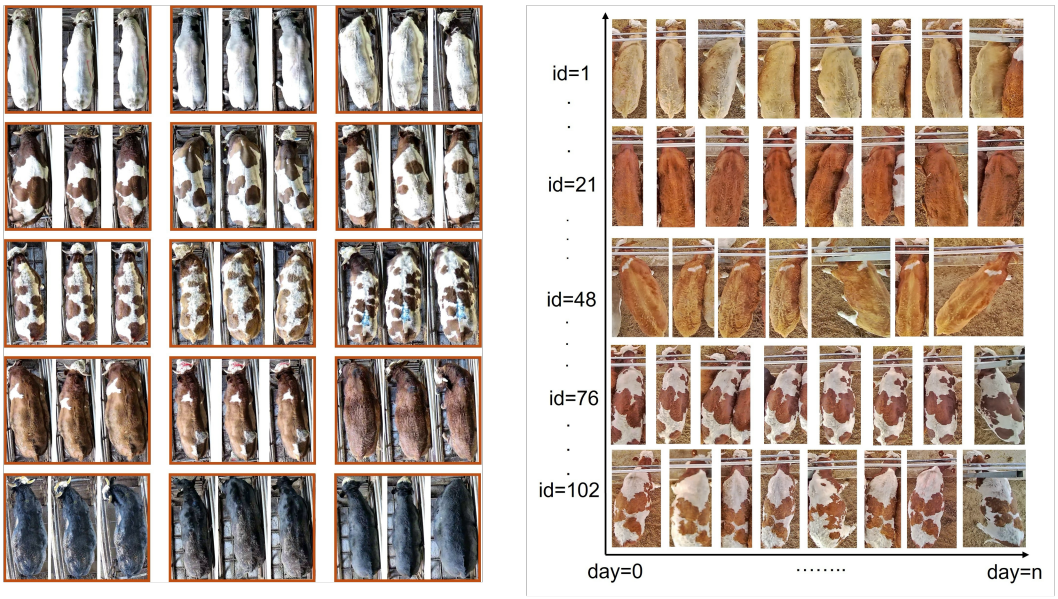
The beef cattle industry presents several challenges for implementation of precision livestock farming implementation. These challenges include longer profit cycles with relatively lower returns, fewer metrics, and a high proportion of crossbred cattle. This leads to several issues during the fattening process: Firstly, beef cattle breeds are predominantly hybrid, with a variety of common breeds including Simmental and Angus [14, 15], with some cattle showing minimal weight gain despite consistent feeding. Secondly, the strong aggression and fighting tendencies of bulls increase the difficulty of external tagging [16, 17], such as higher probabilities of ear tag loss and increased difficulty in re-tagging. Thirdly, the stress response caused by ear tags is amplified. Dairy cattle have a lifespan of 3-5 years, whereas beef cattle are only kept for 18 months, magnifying

the stress response from pain. As a result, ear tags are not ideal for beef cattle management, which demands lower-cost and less intrusive solutions.

Recent advancements in artificial intelligence have propelled bovine recognition technologies to the forefront of precision livestock research, primarily through sophisticated computer vision techniques encompassing cattle detection[18, 19], tracking[20], facial recognition[16, 17], individual identification[21, 22, 23, 24, 25], and coat pattern analysis[26, 17, 27]. Current methodologies demonstrate specialized progress across domains: In cattle detection, Wenjie Yang et al. achieved state-of-the-art performance through deformable convolutions and coordinated attention mechanisms in YOLOv8[27]. In cattle tracking, Shujie Han et al. implemented trajectory tracking via modified YOLOv5 architectures[20], while Su Myat Noe et al. recently optimized a multi-camera YOLOv8 system for tracking black cattle in complex open ranch environments[43]. In cattle keypoint estimation, Qingcheng Fan et al. combined optimized YOLO with HRNet for comprehensive skeletal analysis[7]. In re-identification, Xingshi Xu et al. pioneered few-shot identification through meta-learning frameworks[26]. Notably, W. H. E. Mg et al. leveraged trajectory data and time-series analysis for automated calving time prediction, highlighting the growing demand for long-term behavioral monitoring[44].

While recent advances have improved cattle recognition performance, two limitations persistent challenges for real-world deployment and model generalization. Firstly, most studies rely on viewpoint-specific datasets, where cattle images are captured from limited or non-overlapping angles, preventing effective cross-view integration. Secondly, there is a notable absence of multi-task benchmark datasets and standardized evaluation metrics, which limits reproducibility and hampers real-world deployment of these technologies.

Considering the aforementioned limitations, the dataset developed in this paper focuses on the dorsal view of cattle. This choice leverages the rich and stable texture patterns on the back, which provide strong visual cues for individual identification. The dorsal view also enables non-contact image capture, particularly when cattle are feeding, allowing consistent viewpoint alignment under practical farm conditions. Furthermore, it reduces perspective distortion, scale variation, and occlusion compared to other viewpoints, resulting in higher image consistency and improved suitability for long-term recognition. Figure 1 presents an overview of the BECA-D and BECA-L, illustrating the diversity and long-term characteristics of beef cattle images. Computer vision-based beef cattle recognition systems can generally be categorized into two stages: upstream tasks focused on identity image acquisition and downstream tasks aimed at recognition and matching. Upstream tasks typically involves object detection and tracking, where the goal is to accurately localize and identify individual cattle within the visual field. Downstream tasks then build upon this foundation and include pose estimation, re-identification, and other analyses performed on the detected targets.



(a) Beef Cattle Dataset for Diversity (BECA-D), showing three images for each of 5,661 individual cattle.

(b) Beef Cattle Dataset for Long-term Recognition (BECA-L), featuring images of 103 cattle tracked over 5 months.

Figure 1. Overview of the BECA-D and BECA-L

In the upstream identity image acquisition task: BECA-L supports rotated object detection through Oriented Bounding Box (OBB) annotations. Compared to axis-aligned bounding boxes, OBB detection provides a significant advantage for

dorsal cattle images, where targets often exhibit elongated shapes and tend to overlap due to camera perspectives. The use of oriented bounding boxes enhances the coverage ratio of the target within the detection region, while preserving the natural aspect ratio of the image. Furthermore, BECA-L incorporates keypoint annotations that facilitate precise image alignment. Through multi-keypoint localization tailored for dorsal cattle recognition, the dataset enables more accurate image registration, thereby lowering the requirements for image acquisition consistency and enhancing detection performance in subsequent re-identification tasks.

In the downstream recognition and matching task: BECA-L captures growth variations in beef cattle across growth cycles, with images collected at daily or near-daily intervals. This temporal richness addresses a critical gap in existing cattle recognition datasets, particularly in real-world scenarios. Moreover, the dataset facilitates research into zero-shot and few-shot learning paradigms for cattle identification. Pretraining models on the BECA-D enables transfer learning that achieves competitive performance in open-set cattle recognition tasks. Additionally, by utilizing a limited number of multi-angle samples from BECA-L, effective few-shot learning strategies can be developed, yielding promising recognition accuracy with minimal labeled data.

Methods

The BECA was developed to bridge the gap between AI model validation in controlled environments and its deployment in precision livestock farming. It comprises multiple beef cattle breeds and a systematically collected set of long-term dorsal-view images. As shown in Figure 2, the construction process of the BECA involves the following steps:

1. Deploy the camera and collect the data.
2. Annotate the cattle dorsal object detection dataset and train an object detection model.
3. Annotate the Oriented Bounding Box based on the trained object detection model.
4. Define keypoints for the cattle dorsal view, annotate a posture estimation dataset, and train a keypoint detection model.
5. Annotate long-term re-identification data and train the Re-ID model.

The two sub datasets of BECA differ in their annotation strategies and task-specific design choices, which will be elaborated separately. Each step in the pipline will be further detailed in the following sections.

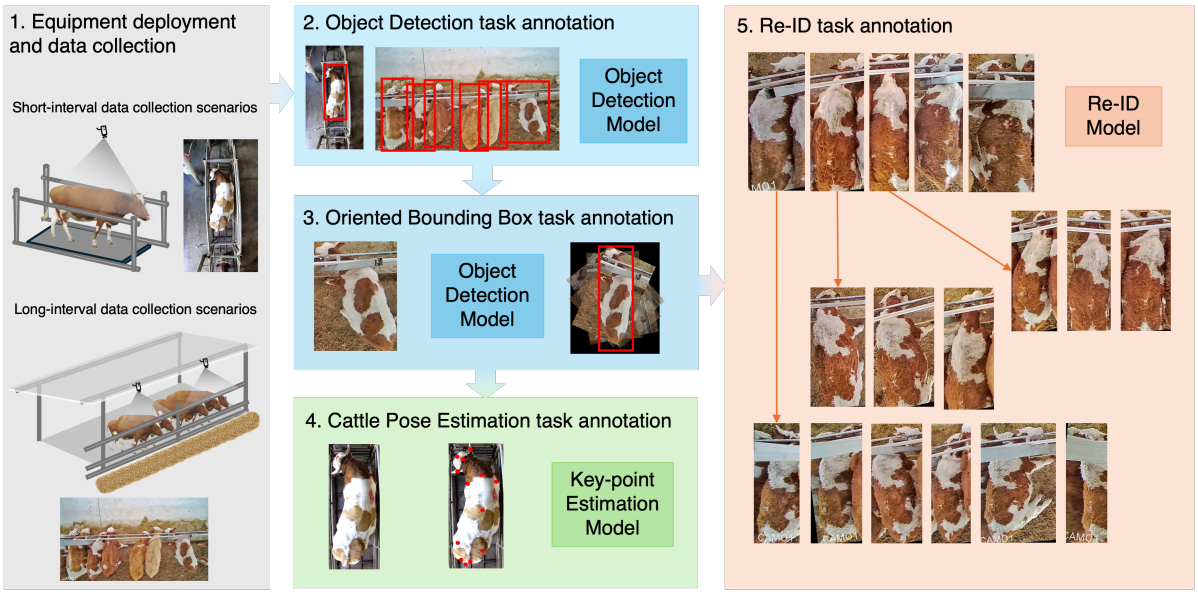


Figure 2. Diagram of data collection and processing

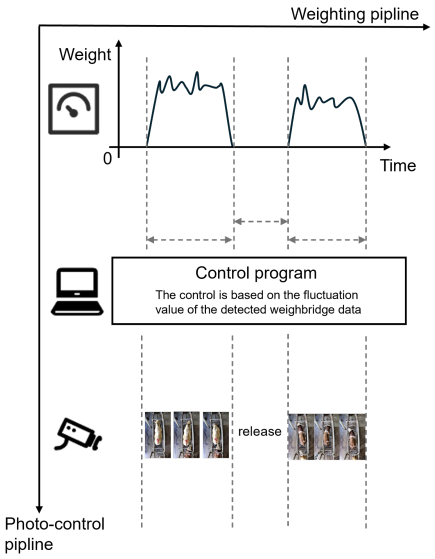
1. Deploy the camera and collect the data

Proper camera deployment is crucial for collecting image data of beef cattle in farm scenes. Cameras are installed 3 or 6 meters above the cattle weighing area. The deployment height is mainly determined by two factors: the metal framework of the roof and whether the camera's focal length can cover the cattle. When the cattle pass through the weighing passage, they

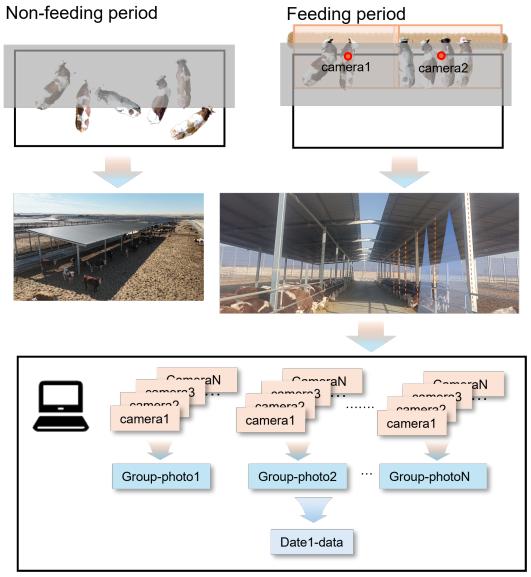
pause briefly, allowing the cameras to take and save photographs. Since the cameras do not have exposure devices, the data collection process does not cause additional stress to the cattle, thereby not affecting their welfare.

The design details of the data collection system for the BECA-D are shown in Figure 3(a). The system consists of terminals, a local area network (LAN), cameras, and scales. When a cow steps onto the scale, the weight reading rapidly increases and stabilizes within a few seconds. Once stabilized, the LAN-connected camera is automatically triggered to capture images at 2-second intervals, with a maximum of three images recorded per session. This 6-second capture window was determined based on real-world observations, where the average time a cow spends on the scale is approximately 30 seconds. The resulting images typically capture variations in the cow's behavior and posture, such as lowering its head, turning its head sideways, looking backward, or taking a step forward.

The data collection system design for the BECA-L is shown in Figure 3(b). It mainly consists of terminal control and cameras. The terminal controls all cameras to take simultaneous photos while the cattle are feeding, and the images are saved. Each day, 18 sets of images are collected, with a 3-minute interval between each set. During the data collection period for the BECA-L, the beef cattle were maintained in fixed groups without any changes to their pens or social configurations, ensuring consistent sampling of the same individuals over time.



(a) The procedure of image capturing for the BECA-D. The fluctuations of the scale are used as a trigger to control the imaging process. When the scale is activated and enters a stable period, the control program initiates the photo-taking operation.



(b) The procedure of image capturing for the BECA-L. The feeding period is used as the trigger condition. When the cattle shed enters the feeding period, the control program initiates the imaging process.

Figure 3. Data Collection process for cowshed equipment

2. Annotate the cattle dorsal object detection dataset and train an object detection model.

We selected a subset of images for object detection annotation. In total, 874 overhead images of cowsheds were manually annotated, yielding 4,178 individual cattle instances. The annotation process followed two key principles o ensure high-quality and reliable data for cattle recognition tasks:

- **Preservation of complete cattle information:** Images were selected to include as much of the entire cattle body as possible, particularly focusing on maintaining full visibility of key features such as the head, tail, and body. This helps in supporting a wide range of downstream tasks, including detection, pose estimation, and re-identification.
- **Exclusion of partially cropped cattle:** To avoid ambiguity and reduce noise in the training and evaluation processes, images in which cattle were partially cut off by the image boundaries were excluded. Such cropping can lead to incomplete feature representation and may negatively impact model performance.

For model training, we split the dataset into training and validation sets at an 80:20 ratio. And we employed YOLOv11 [8] as the backbone network. Given the relatively uniform size of targets in the overhead view, only the detection heads for medium and large objects were retained in the model's output layer. Based on the trained model, we then performed inference on the remaining data. A manual review process was employed, where detection boxes were drawn on the original images

for screening and adjustment, ultimately resulting in the final Axis-Aligned Bounding Box (AABB) annotations for object detection.

It is worth noting that cattle dorsal targets tend to be elongated and often appear at rotational angles, resulting in aspect ratios significantly greater than one. In dense or tilted arrangements, this can cause bounding boxes to be suppressed during non-maximum suppression. To address this issue, oriented object detection annotations will be incorporated in the subsequent steps of dataset development.

3. Annotate the Oriented Bounding Box based on the trained object detection model.

Due to the misalignment between the bounding boxes and the actual orientation of cattle, challenges arise when cattle gather closely during feeding. In such high-density scenarios, any tilt in the cattle's posture may cause bounding boxes to overlap with neighboring individuals, leading to inclusion errors and a reduction in the clarity of identity-related features captured for each cow. This can hinder subsequent re-identification tasks. Furthermore, since the dorsal view of cattle is typically an elongated target with a height-to-width ratio significantly greater than one, missed detections become more likely when cattle are densely packed. It can be observed that the oriented detection box captures the target cow with less interference from surrounding cows or the environment. This is further exacerbated by the non-maximum suppression algorithm, which may remove valid results under such conditions.

To address this, we developed an approach for capturing oriented bounding boxes (OBBs) from the dorsal view of cattle, as detailed in Algorithm 1. This method ensures that the cattle are consistently aligned with the head oriented upward and the tail downward. By doing so, it helps reduce image noise and misalignment, thereby improving the accuracy and reliability of downstream tasks such as recognition and pose estimation.

Algorithm 1 Oriented Bounding Box Generation via Rotational Scan

Require: Original image I, detection model M, rotation step $\Delta\theta$
Ensure: OBB annotations O, orientation angles Θ

Phase 1: Canvas Preparation

$h, w = \text{height}(I), \text{width}(I)$
$C = \text{CreateCanvas}(1.4h, 1.4w)$
$\text{CenterCopy}(I, C)$ ▷ Copy I to center of C

Phase 2: Rotational Scanning

$\theta = -45^\circ$, $T = \emptyset$ ▷ Initialize angle and target array
while $\theta \leq +45^\circ$ **do**
  $C_\theta = \text{Rotate}(C, \theta, \text{pivot} = \text{center}(C))$
  $B_\theta = M(C_\theta)$ ▷ Detect bounding boxes

Phase 3: Temporal Matching

  if $\theta \neq -45^\circ$ **then**
    $H = \text{HungarianMatch}(B_{\theta - \Delta\theta}, B_\theta)$
    for $(b_{prev}, b_{curr}) \in H$ **do**
      $T = \text{UpdateTarget}(T, b_{curr}, \theta)$
    end for
  end if
  $\theta = \theta + \Delta\theta$
end while

Phase 4: OBB Selection

for target $\in T$ **do**
  $b^* = argmax_{b \in \text{target}} \left(\frac{\text{width}(b)}{\text{height}(b)} \right)$
  $O = O \cup \{\text{ConvertToOBB}(b^*)\}$, $\Theta = \Theta \cup \{\theta_{b^*}\}$
end for

The proposed method consists of four coordinated phases:

1. **Canvas Preparation**: Creates an extended canvas C with dimensions calculated as:

$$\begin{cases} \text{height}(C) = n \times \text{height}(I) \\ \text{width}(C) = n \times \text{width}(I) \end{cases} \quad \text{where } n > 0 \tag{1}$$

The original image I is centered via CenterCopy preserving aspect ratio.

2. **Rotational Scanning**: Performs discrete rotation scan with:

$$\theta \in \{-45^\circ, -45^\circ + \Delta\theta, \ldots, +45^\circ\} \tag{2}$$

Recommended step size $\Delta\theta = 1^\circ$ balances precision and computation.

3. **Temporal Matching**: Hungarian algorithm minimizes matching cost:

$$H = argmin_{\text{matching } \pi} \sum_i \|b_i - \pi(b_i')\|^2 \tag{3}$$

where c_i and c_i' denote the centers of bounding boxes in consecutive frames, and π denotes the permutation of boxes. We use the L2 distance between bounding box centers instead of IoU as the cost metric. This choice is motivated by the fact that in densely packed dorsal images of tilted cattle, elongated boxes often overlap, leading to unstable IoU values. Additionally, the rotation of boxes introduces orientation differences, further complicating IoU computation. In contrast, the L2 distance between centers provides a more stable and reliable matching measure, especially in the context of iterative rotations.

4. **OBB Selection**: Maximizes aspect ratio for rotation robustness:

$$\text{OBB} = \text{ConvertToOBB}(b^*, \theta_{b^*}) \tag{4}$$

The conversion considers both rotation angle and projective transformation.

After the OBB annotation, a manual review was conducted to eliminate abnormal matching results generated during the rotation process and to remove incomplete detection boxes. This ensures that the dorsal view of cattle is fully contained within the oriented detection box. The advantage of OBBs over AABBs is illustrated in Figure 4.

Figure 4. Comparison between axis-aligned and oriented object detection boxes: the red box represents the axis-aligned detection box and the green box denotes the oriented object detection box.

4. Define keypoints for the cattle dorsal view, annotate a posture estimation dataset, and train a keypoint detection model.

In face recognition systems, images typically undergo an alignment process after detection, which significantly improves the accuracy of subsequent matching. Similarly, in cattle recognition, pose estimation annotations can provide valuable pose information that benefits various downstream tasks, such as re-identification and behavior analysis. Inspired by this concept, we designed 13 skeletal key points for annotation based on the joint degrees of freedom of cattle, as shown in the Figure 5. The 13 key points, numbered from 1 to 13, are as follows: 1.Poll 2.Cervical Vertebrae Crest (Near the Atlas Region) 3.Withers 4.Thoracolumbar Junction 5.Tail Head 6.Left Tuber Coxae (Left Hook Bone) 7.Right Tuber Coxae (Right Hook Bone) 8.Left Patella (Stifle Region) 9.Right Patella (Stifle Region) 10.Left Scapulohumeral Joint (Shoulder Joint) 11.Right

Scapulohumeral Joint (Shoulder Joint) 12.Left Ear Base 13.Right Ear Base. We believe these 13 points exhibit uniform and symmetrical distribution across the dorsal view of cattle, which is beneficial for describing the structural angles caused by the twisting of the cervical and lumbar vertebrae.

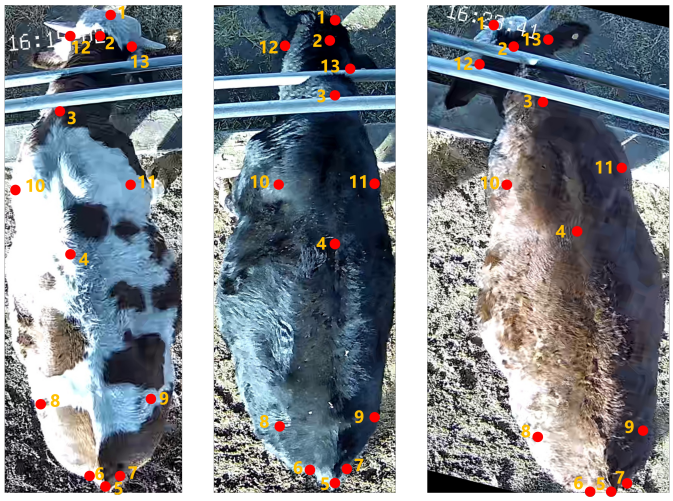


Figure 5. Keypoints annotation example

We annotated the 4178 images from results obtained in the third step. During the annotation process, we invited 3 professionals in beef cattle breeding and 3 experts in beef cattle slaughtering, relying on their familiarity with cattle anatomy to annotate the key points.

In particular, certain skeletal landmarks (points 9, 10, 11, and 12) are partially occluded by muscle tissue from the top-view perspective. These points, however, are essential for assessing spinal tilt and evaluating cattle posture and behavior. Breeding experts contributed knowledge of muscle distribution, while slaughterhouse experts provided insights into the underlying skeletal structure. This complementary expertise allowed both groups to accurately infer the positions of occluded landmarks. Each image was annotated independently by both expert types, and the final keypoint coordinates were determined as the midpoint of the two annotations to ensure high precision and consistency.

After splitting the resulting cattle keypoint dataset into training and testing sets in an 8:2 ratio, we trained a pose estimation model. The trained model was then used to perform inference on all images in the BECA-L dataset, yielding the cattle keypoint annotations.

5. Annotate long-term re-identification data and train the Re-ID model.

In this step, we used the OBB dataset obtained from the traversal process described in Step 3 for training and annotation. Since each cow in the BECA-D appears relatively independently within the field of view and images are captured at 3-second intervals, it is possible to associate each image group with a specific cattle identity, ensuring that each image carries an ID attribute. Based on the BECA-D, we conducted re-identification training. During training, 16,083 images from 5,361 cattle were used for training, and 806 images from 300 cattle were used for testing. For the evaluation protocol, we constructed an equal number of positive and negative sample pairs, computed their cosine similarity, and then searched for the optimal threshold to separate positive from negative samples. The model was trained using a combination of cross-entropy loss and triplet loss functions.

Given the prolonged data acquisition period of the BECA-L, where individual cattle reappear over extended time spans—manual annotation of identity labels becomes prohibitively expensive. To alleviate this challenge, we propose a similarity-driven simulation-optimization heuristic based on the zero-shot identification performance of the OSNet[30] trained on the BECA-D. The core procedure is formalized in Algorithm 2.

Algorithm 2 Identity-Aware Clustering with Conflict Constraints

Require: Image dataset D, pre-trained OSNet M, initial threshold τ_0
Ensure: Cluster assignments A, conflict set C

Phase 1: Threshold Initialization
$\tau \leftarrow \tau_0$ ▷ Optimal ZSL threshold from OSNet training
Phase 2: Negative Sample Construction
for all $(x_i, x_j) \in D \times D$ **do**
  $f_i \leftarrow M(x_i)$, $f_j \leftarrow M(x_j)$ ▷ Feature extraction
  if $\text{sim}(f_i, f_j) < \tau$ **then**
    $C \leftarrow C \cup \{(x_i, x_j)\}$ ▷ Conflict pair registration
  end if
end for
Phase 3: Hierarchical Clustering
$P \leftarrow \text{AllPairs}(D)$ ▷ Generate $n(n-1)/2$ pairs
$P \leftarrow \text{Sort}(P, \text{sim}(f_i, f_j))$ ▷ Ascending order by distance
Phase 4: Conflict-Aware Aggregation
while $P \neq \emptyset$ **do**
  $(x_i, x_j) \leftarrow \text{PopMin}(P)$ ▷ Closest remaining pair
  if $\nexists c \in C$ s.t. $(x_i, x_j) \subseteq c$ **then**
    $A \leftarrow \text{MergeClusters}(x_i, x_j)$ ▷ Agglomerative merging
  end if
end while

The proposed method consists of four coordinated phases:

1. **Threshold Initialization**: Uses the optimal cosine similarity threshold (τ) determined during OSNet's zero-shot learning phase, where $\tau_0 = \mu - 2\sigma$ (mean minus two standard deviations of positive pair similarities).
2. **Negative Sample Construction**: Constructs conflict set C containing all image pairs with similarity below τ. This implements:

$$C = \{(x_i, x_j) | \forall x_i, x_j \in D, \text{sim}(M(x_i), M(x_j)) < \tau\} \tag{5}$$

3. **Hierarchical Clustering**: Generates and sorts all possible feature vector pairs by ascending cosine distance using:

$$\text{sim}(f_i, f_j) = 1 - \frac{f_i \cdot f_j}{\|f_i\| \|f_j\|} \tag{6}$$

4. **Conflict-Aware Aggregation**: Performs agglomerative clustering while respecting conflict constraints through:

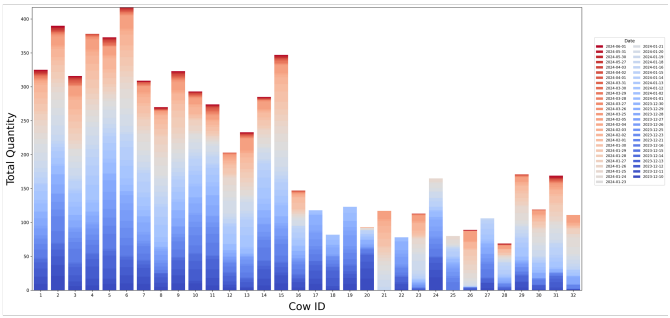
$$A_{new} = \begin{cases} A \cup \{x_i \cup x_j\}, & \text{if } \forall x_k \in A, (x_k, x_j) \notin C \\ A, & \text{otherwise} \end{cases} \tag{7}$$

Following the initial grouping, a second round of manual annotation was performed. All images in the BECA-L were independently reviewed and labeled by at least two annotators. Any images with inconsistent labels were discarded to maintain a high level of identity consistency throughout the dataset. By integrating automated clustering with rigorous human validation, we obtained the final, refined version of the BECA-L.

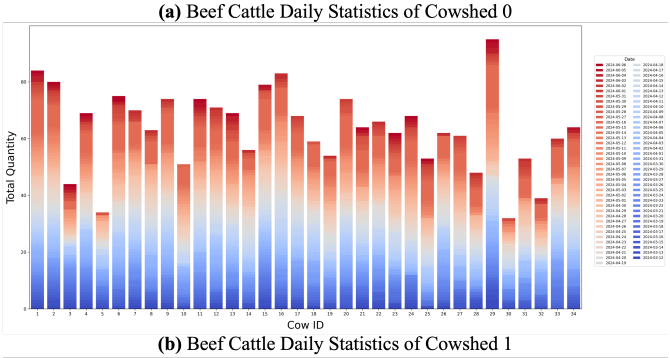
Data Records

The data is available at Figshare[45]. The directory is divided into two subfolders: "BECA-D" and "BECA-L", with no overlapping cattle individuals between the two folders, and each can be used independently.

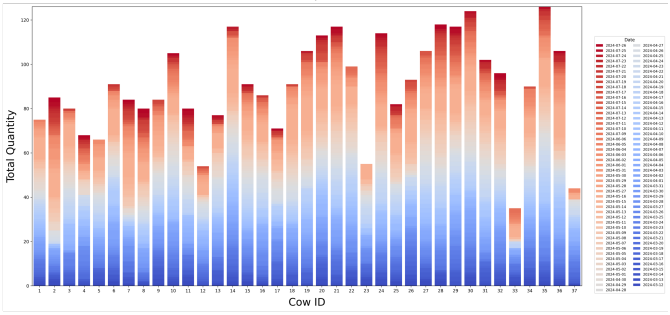
The BECA-D folder includes three components: the "train" folder containing 16,083 images from 5,361 individual cattle for training, the "val" folder with 806 images from 300 identities for validation, and the "pairs.txt" file, which provides 1,424 labeled image pairs for similarity evaluation.



(a) Beef Cattle Daily Statistics of Cowshed 0



(b) Beef Cattle Daily Statistics of Cowshed 1



(c) Beef Cattle Daily Statistics of Cowshed 2

Figure 6. Statistics of Cattle IDs, Number of Collected Images, and Collection Dates in the BECA-L

The BECA-L folder consists of two subfolders: "obj_detection" and "reid". The "obj_detection" folder provides object detection annotations and consists of "image", "obb_output", and "annotations.json". The "image" folder contains raw images collected from cowsheds. The "annotations.json" file follows the COCO format and includes image information and annotation information, while the "annotations" section includes object-level annotations, such as AABBs, OBBs, and keypoints. The object detection annotations adopt the 'xyxy' format. For OBB annotations, we follow the 'xyxyangle' format as defined in the HRSC dataset specification [38]. The "obb_output" folder stores results where the OBB annotations are overlaid on the original images and cropped accordingly. The "reid" folder contains re-identification annotations based on OBB-cropped images, organized into four subfolders: "cowshed0", "cowshed2", "cowshed3", and "cowshed _sum". The first three subfolders include identity-labeled images from individual cowsheds. The "cowshed_sum" folder aggregates all images from the three sheds, with identities relabeled for consistency, totaling 12,172 images from 103 cattle. Long-term statistical distributions across the sheds are shown in Figure 6.

It is important to note that the number of images per cattle varies due to two primary factors. The frequency with which each animal appears during the fixed data collection periods is inherently unpredictable. Additionally, pen adjustments occasionally occur during the beef cattle rearing process. For animal welfare reasons, some cattle were relocated out of the collection area before the completion of the data acquisition period.

Technical Validation

The dataset development process encompasses four main tasks: object detection, oriented object detection, keypoint detection, and re-identification. At each task, specific strategies were employed to ensure data accuracy and consistency. The

evaluation of model performance is based on several standard metrics commonly used in object detection, classification, and pose estimation. The definitions and formulas of these metrics are given below:

- **Precision**:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{8}$$

where TP is the number of true positives and FP is the number of false positives.

- **Recall**:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{9}$$

where FN is the number of false negatives.

- **Accuracy**:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

where TN is the number of true negatives.

- **F1 Score**:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

- **Object Keypoint Similarity (OKS)** for pose estimation:

$$\text{OKS} = \frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right) \cdot \delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \tag{12}$$

where d_i is the Euclidean distance between the predicted and ground-truth keypoint i, v_i is the visibility flag (1 if visible, 0 otherwise), s is the object scale (e.g., square root of the bounding box area), and k_i is a keypoint-specific constant that controls the decay rate.

- **Mean Average Precision at matching threshold = 0.5 (mAP@0.5)**:

$$\text{mAP@0.5} = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad \text{with matching score} \geq 0.5 \tag{13}$$

where AP_i is the average precision for the i-th class, and the matching threshold of mAP can be based on either IoU (for detection) or OKS (for pose estimation).

- **Mean Average Precision from matching threshold = 0.5 to 0.95 (mAP@0.5:0.95)**:

$$\text{mAP@0.5:0.95} = \frac{1}{10N}\sum_{j=0}^{9}\sum_{i=1}^{N} AP_i^{\text{threshold}=0.5+0.05j} \tag{14}$$

where the matching threshold varies between 0.5 and 0.95 in steps of 0.05.

- **Mean Average Recall at matching threshold = 0.5 (mAR@0.5)**:

$$\text{mAR@0.5} = \frac{1}{N}\sum_{i=1}^{N} \text{AR}_i^{\text{threshold}=0.5} \tag{15}$$

where AR_i denotes the average recall for class i, computed using a matching threshold of 0.5.

- **Mean Average Recall across matching thresholds (mAR@0.5:0.95)**:

$$\text{mAR@0.5:0.95} = \frac{1}{10N} \sum_{i=1}^{N} \sum_{j=0}^{9} \text{AR}_i^{\text{threshold}=0.5+0.05j} \tag{16}$$

where the matching threshold again spans from 0.5 to 0.95.

The TP, FP, TN, and FN values used in Formula 8–11 are calculated based on the confusion matrix derived from the similarity threshold that achieves the highest accuracy in the current round of image pair classification. This approach is widely used in face recognition tasks [42]. Additionally, for each metric, we report both the average value and its variation after the model has converged.

The accuracy of the object detection

In Step 2, training and validation are performed using the manually annotated data with the model. The training results are shown in Table 1. Precision reflects the reliability of the model in detecting cattle. Precision of 97.38% indicates an extremely low false detection rate, effectively avoiding the misclassification of the background as cattle. Recall measures the model's ability to identify all actual cattle targets. The experiment achieved a high Recall of 97.9%, indicating a low rate of missed detections. This result demonstrates that the model can reliably detect cattle even in semi-outdoor scenes with high lighting contrast, where images are captured using standard cameras. The mAP@0.5 calculated using an Intersection over Union (IoU) threshold of 0.5, reached 99.34%. This suggests that under relaxed localization conditions, where a detected bounding box overlaps with the ground truth by at least half, the model performs exceptionally well, fulfilling the accuracy requirements of most agricultural monitoring applications. A mAP@0.5:0.95 score of 91.62% demonstrates that the model sustains high accuracy even when stricter localization criteria are applied.

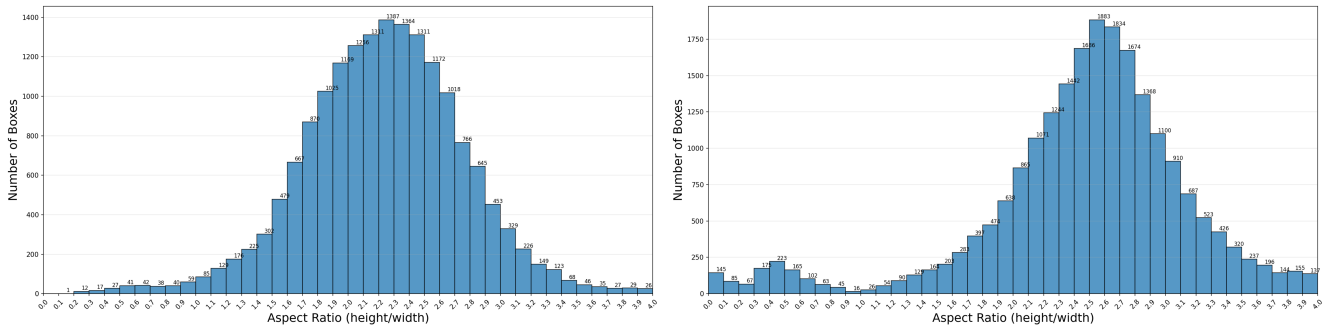
Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
97.4	97.9	99.3	91.6

Table 1. YOLOv11 training performance results.

In real-world scenarios, partial visibility of cattle can occur due to camera placement or animal posture, leading to missing regions such as the head or tail. However, in practical recognition tasks, these missing parts have minimal impact on overall identification performance.

Oriented bounding boxes are used to optimize the object detection boxes.

Oriented object detection was introduced for dorsal view annotation, with the primary goal of ensuring that the detected subject occupies as much of the image as possible. This addresses the issue of "long targets" such as dorsal view of cattle, where adjacent targets may intrude into the bounding box, thereby degrading re-identification accuracy. We use an iterative traversal approach to select the most suitable oriented detection box for each individual cow.



(a) Bounding Box Aspect Ratio Distribution.

(b) Oriented Bounding Box Aspect Ratio Distribution.

Figure 7. Comparison of different between bounding box annotation and oriented bounding box annotation aspect ratio distribution

Following the iterative process, we analyzed the distribution of aspect ratios for both the axis-aligned and rotated bounding boxes, as shown in Figure 7. It can be observed that the number of OBBs with aspect ratios close to 1 has significantly decreased, with these cases now distributed across higher or lower aspect ratios after rotation. To further ensure annotation accuracy, the OBBs are mapped back to the original image coordinates for inspection, and any abnormal annotations are manually corrected.

In the experiments shown in Table 2 and Table 5,We observed that image quality within the BECA-L varies significantly, with some images exhibiting pattern distortions caused by differing dorsal view angles, which adversely affect recognition performance. Due to the consistency between the upstream and downstream datasets, we divided the images into five sub-regions based on the x-coordinate of the bounding box center obtained from object detection. For each identity, only one image per region was used for training, while the remaining images were reserved for validation.

We conducted experiments on the BECA-L to demonstrate the necessity of using OBB instead of AABB. The following experimental settings were evaluated:

- Zero-shot: inference using a model trained solely on the BECA-D, without any additional training on BECA-L data.
- Few-shot: training the Re-ID model directly on BECA-L, with no more than 5 images per individual used for learning.
- Few-shot transfer learning: fine-tuning a model pre-trained on BECA-D by using up to 5 images per individual from BECA-L for few-shot transfer learning.

As shown in Table 2, the use of aligned OBBs results in a significant improvement in recognition performance compared to the AABBs obtained from object detection. This clearly highlights the importance of annotating OBBs, which refines the bounding box orientation and contributes to better identification accuracy.

Subset	Box Type	Setting	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
BECA-L	AABB	Few-shot	83.10 ± 3.25	79.53 ± 3.68	88.00 ± 3.10	83.94 ± 3.03
BECA-L	AABB	Few-shot transfer learning	84.35 ± 1.45	94.03 ± 3.11	73.50 ± 2.10	82.55 ± 1.51
BECA-L	OBB	Few-shot	90.15 ± 1.90	89.17 ± 3.88	91.30 ± 4.30	90.03 ± 1.93
BECA-L	OBB	Few-shot transfer learning	93.45 ± 0.70	93.39 ± 0.97	93.85 ± 1.35	93.44 ± 0.71

Table 2. Performance metrics for BECA-L across different settings and bounding box types.

The accuracy of the Keypoint Annotation in Pose Estimation

The purpose of pose estimation is to assist researchers in aligning the target within the image via keypoint estimation. Among the 13 defined keypoints, two groups can be distinguished: Central dorsal keypoints, including the Poll, Cervical Vertebrae Crest (near the Atlas Region), Withers, Thoracolumbar Junction, Tail Head, Left Tuber Coxae (Left Hook Bone), and Right Tuber Coxae (Right Hook Bone). These points are located along the spine, easily identifiable from the dorsal view, and based on subcutaneous anatomical landmarks. Lateral keypoint, including the Left and Right Patella (Stifle Region), Left and Right Scapulohumeral Joint (Shoulder Joints), and Left and Right Ear Bases. These are located symmetrically on both sides of the animal and are less visually prominent from the dorsal view, often inferred from muscle contour protrusions. To ensure annotation consistency, professionals with expertise in beef cattle anatomy and slaughtering were recruited for this task.

In step 3 of the methods, training and validation methods are performed using the manually annotated data with the pose estimation model, the results were as follows in Table 3 ,

mAR@0.5	mAR@0.5:0.95	mAP@0.5	mAP@0.5:0.95
98.9	83.6	97.9	79.5

Table 3. HR-Net model training performance results

The evaluation metrics employed in this study are based on keypoint detection accuracy and recall. The results indicate that the observed performance differences between mAR@0.5:0.95 and mAR@0.5, as well as between mAP@0.5:0.95 and mAP@0.5, are mainly due to the fact that keypoint estimation achieves relatively good predictions under lower localization requirements, but generally lacks sufficient accuracy under stricter criteria.

Finally, we overlaid the annotations from the first three steps onto the original images. An example of the annotated samples is shown in the Figure 8.

Figure 8. The red bounding boxes, green bounding boxes, and red dots represent the annotations for object detection, oriented object detection, and pose estimation keypoints, respectively.

Verification of the re-identification annotation results.

In the BECA-D, cattle typically pass through a chute, placing their backs near the center of the camera's field of view. In contrast, the BECA-L lacks this structured setup—dorsal views may be captured from the upper-left or upper-right angles. Since the dorsal view of cattle should be modeled as a cylindrical surface, the patterns vary significantly depending on the viewing angle.

To account for this, we introduced relative coordinates to indicate the cattle's position in the image with respect to the camera's viewpoint. These values range from 0 to 1 and are divided into five intervals. During re-identification, cattle are compared only within the same interval, enabling more accurate self-matching. Using a model trained on BECA-D, we performed fine-tuned inference on BECA-L by selecting five sample angles per class. This approach significantly improved accuracy, and the relative coordinate division also aided manual filtering during annotation. The validation results are shown in the table below. The Table 4 indicate that OSNet demonstrates strong performance in the beef cattle dorsal view re-identification task.

Backbone	Accuracy (%)	Precision (%)	Recall (%)	F1 Score
Resnext50	97.75 ± 0.28	98.10 ± 0.76	97.33 ± 0.56	97.75 ± 0.28
Efficientnet-b2	97.89 ± 0.42	97.16 ± 0.46	98.73 ± 0.71	97.91 ± 0.42
MaxVit	98.31 ± 0.56	97.98 ± 1.02	98.10 ± 0.91	98.32 ± 0.55
OSNet	98.91 ± 0.31	99.16 ± 0.56	98.52 ± 0.50	98.90 ± 0.31

Table 4. Performance comparison between ResNeXt50[39], Efficientnet-b2[40], MaxVit[41] and OSNet models across evaluation metrics

Owing to differences in cameras and acquisition durations, the recognition performance across sub-datasets from different cowsheds may vary. To investigate this, we conducted an experiment. The experimental results are shown in Table 5.

We observed a performance drop in cowshed0, which has the longest data collection period. This is likely due to long-term morphological changes in cattle that affect feature stability over time. In the three subsets, the proportion of solid-color cattle images, where dorsal features are difficult to distinguish, is 27.00%, 22.47%, and 20.01%, respectively. The relatively high proportion in cowshed0 contributes to its lower recognition accuracy. Notably, the diversity in BECA-D enables effective dorsal view modeling, and transfer learning using a model pre-trained on BECA-D has proven to be effective.

To evaluate the generalization capability of the BECA in breeding scenarios, we conducted a cross-dataset experiment using the same model architecture, training hyperparameters, number of training epochs, and data augmentation strategies. The model was pre-trained on the BECA-D and OpenCows2020, and a unified public validation set was constructed from their

Subset	Setting	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
cowshed0	Zero-shot	67.60	69.64	62.40	65.82
cowshed0	Few-shot	81.27 ± 0.52	79.19 ± 1.86	84.75 ± 2.25	81.84 ± 0.36
cowshed0	Few-shot transfer learning	90.25 ± 0.50	92.85 ± 3.04	87.55 ± 4.15	89.99 ± 0.85
cowshed2	Zero-shot	83.70	85.18	81.60	83.35
cowshed2	Few-shot	97.08 ± 0.82	95.80 ± 1.54	97.95 ± 1.25	97.12 ± 0.79
cowshed2	Few-shot transfer learning	98.18 ± 0.18	98.89 ± 0.30	97.50 ± 0.20	98.17 ± 0.18
cowshed3	Zero-shot	86.90	86.90	84.50	86.58
cowshed3	Few-shot	91.03 ± 1.32	91.94 ± 2.00	90.65 ± 3.15	90.87 ± 1.36
cowshed3	Few-shot transfer learning	96.72 ± 0.93	96.43 ± 0.98	97.05 ± 0.85	96.74 ± 0.92

Table 5. Performance metrics for individual cowsheds across different settings.

Train Dataset	Train/Val Ratio	Accuracy (%)	Precision (%)	Recall (%)	F1 Score
BECA-D	1:9	91.84	91.58	92.58	92.08
BECA-D	2:8	97.63	97.77	97.61	97.69
OpenCows2020	1:9	88.38	86.63	91.15	88.83
OpenCows2020	2:8	92.17	92.88	91.74	92.31

Table 6. Comparison of performance on the BECA-D, and OpenCows2020 datasets under different train/validation split ratios.

respective validation sets. The final validation set contains 10,000 sample pairs from different datasets. Experimental results, as shown in Table 6, indicate that under various train/validation split ratios, the model pre-trained on BECA-D consistently outperforms that pre-trained on OpenCows2020. These findings demonstrate that BECA has strong adaptability to diverse breeding scenarios and exhibits sufficient generalization ability.

Usage Notes

The BECA was captured using multi-camera setups with lenses of three different focal lengths: 4mm (Cowshed0), 3.6mm (Cowshed2), and 2.8mm (Cowshed3). All cameras operate at a resolution of 2304××1296 pixels (2K), ensuring high-quality image capture while enhancing the robustness of the dataset through diverse viewpoints and focal conditions. Given that most cattle sheds are designed to balance functionality and cost-efficiency, the supporting columns in these structures often double as load-bearing elements within the feeding passages. To optimize field-of-view and spatial coverage, cameras are mounted on horizontal beams that are perpendicular to these columns, with an average spacing of 6 meters between adjacent cameras. These cameras also feature built-in distortion correction to minimize visual distortion of the structural columns and maintain image fidelity. As all deployments rely solely on natural lighting without additional artificial illumination, the data collection process does not introduce any extra stress to the animals, thereby ensuring their welfare remains unaffected.

All annotations related to object detection and pose estimation were created using the VGG Image Annotator (VIA) [35]. The images capture beef cattle of various breeds, colors, and genders housed in a fattening farm located in Gansu Province, China. Annotations were jointly provided by domain experts and farm personnel, ensuring accurate and realistic ground truth labels for cattle identification in practical breeding environments. This dataset can be used to train and evaluate computer vision algorithms for cattle recognition, with significant applications in precision livestock farming. Both the datasets and corresponding annotations are publicly available, with the aim of fostering collaboration and accelerating advancements in the beef cattle industry. It is hoped that these resources will facilitate further research into animal recognition, behavior analysis, and intelligent monitoring systems in agricultural settings.

The innovative provision of long-term identification data within the BECA enables a closer integration of individual cattle identities with their corresponding images and related information. This feature enhances the potential of the BECA for applications in precision livestock farming, such as monitoring animal welfare, behavior and health[37]. In addition, the datasets facilitate the evaluation and prediction of animal growth trends, as dorsal view images of cattle can offer substantial information related to weight[36, 37]. Furthermore, they offer a valuable resource for researchers studying individual animal behavior, cognition, and group dynamics.

It should be noted that, although the developed BECA offers broad applicability in beef cattle breeding, it is subject to two main limitations. Firstly, the number of images per individual in BECA-L varies across cattle. This inconsistency stems

from the data collection protocol, which was designed to prioritize animal welfare and minimize interference with routine farm management. As part of standard breeding practices, cattle undergo periodic selection and cowshed reassignment over the growth cycle. Consequently, individuals that are moved out of monitored cowsheds are no longer captured by the fixed-position cameras. Secondly, as the dataset is captured from the dorsal view, it may not be directly applicable to breeding scenarios that rely on face or lateral views.

Code availability

The source code and scripts utilized in this study are publicly available on GitHub. The repository provides all the necessary code to reproduce the results, including scripts for data preprocessing, with a focus on object detection and ReID (Re-identification) tasks. The code is available at https://github.com/xxzyq/Two-Cattle-Datasets-in-Dorsal-View-for-Cattle-Recognition.

Data availability

The dataset is available at https://doi.org/10.6084/m9.figshare.29425316[45].

References

[1] EFSA Panel on Animal Health and Welfare (AHAW). (2012). *Scientific Opinion on the welfare of cattle kept for beef production and the welfare in intensive calf farming systems*. EFSA Journal, 10(5), 2669. Wiley Online Library.

[2] Terry, S. A., Basarab, J. A., Guan, L. L., McAllister, T. A. (2020). *Strategies to improve the efficiency of beef cattle production*. Canadian Journal of Animal Science, 101(1), 1-19. Canadian Science Publishing.

[3] Alonso, R. S., Sittón-Candanedo, I., García, Ó., Prieto, J., & Rodríguez-González, S. (2020). *An intelligent Edge-IoT platform for monitoring livestock and crops in a dairy farming scenario*. Ad Hoc Networks, 98, 102047. Elsevier.

[4] Nleya, S. M., & Ndlovu, S. (2021). *Smart dairy farming overview: innovation, algorithms and challenges*. Smart Agriculture Automation Using Advanced Technologies: Data Analytics and Machine Learning, Cloud Architecture, Automation and IoT, 35-59. Springer.

[5] Akbar, M. O., Khan, M. S., Ali, M. J., Hussain, A., Qaiser, G., Pasha, M., et al. (2020). *IoT for development of smart dairy farming*. Journal of Food Quality, 2020(1), 4242805. Wiley Online Library.

[6] GB, A. K., KN, S. K., Prasad, R., Gatti, R., Kumar, S., et al. (2021). *Design and Development of Intelligent System for Dairy Farming*. In Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) (pp. 960-964). IEEE.

[7] Fan, Q., Liu, S., Li, S., & Zhao, C. (2023). *Bottom-up cattle pose estimation via concise multi-branch network*. Computers and Electronics in Agriculture, 211, 107945.

[8] Khanam, R., & Hussain, M. (2024). *Yolov11: An overview of the key architectural enhancements*. arXiv preprint arXiv:2410.17725.

[9] Carpinelli, N. A., Halfen, J., Trevisi, E., Chapman, J. D., Sharman, E. D., Anderson, J. L., & Osorio, J. S. (2021). *Effects of peripartal yeast culture supplementation on lactation performance, blood biomarkers, rumen fermentation, and rumen bacteria species in dairy cows*. Journal of Dairy Science, 104(10), 10727-10743. Elsevier.

[10] Dos Santos Daltro, D., Padilha, A. H., Da Silva, M. V. G. B., Kern, E. L., et al. (2019). *Heterosis in the lactation curves of Girolando cows with emphasis on variations of the individual curves*. Journal of Applied Animal Research. Taylor & Francis.

[11] Zheng, Y., Chen, J., Wang, X., Han, L., Yang, Y., Wang, Q., & Yu, Q. (2022). *Metagenomic and transcriptomic analyses reveal the differences and associations between the gut microbiome and muscular genes in Angus and Chinese Simmental cattle*. Frontiers in Microbiology, 13, 815915. Frontiers Media SA.

[12] Isyanto, A. Y., & Dehen, Y. A. (2015). *Sustainability analysis of beef cattle fattening in Ciamis regency, West Java Province, Indonesia*. Journal of Economics and Sustainable Development, 6(20), 148-154.

[13] Wen-jun, Z. (2011). *Several Key Measures on Improving Beef Cattle Fattening Effect*. China Cattle Science.

[14] Berry, D. P., & Crowley, J. J. (2012). *Residual intake and body weight gain: a new measure of efficiency in growing cattle*. Journal of Animal Science, 90(1), 109-115. Oxford University Press.

[15] Hales, K. E., Galyean, M., & Smith, Z. K. (2022). *Evaluating the Difference between Formulated Dietary Net Energy Values and Net Energy Values Determined from Growth Performance and Estimates of Shrunk Body Weight Gain in Finishing Beef Cattle*. Journal of Animal Science, 100.

[16] Weng, Z., Meng, F., Liu, S., Zhang, Y., Zheng, Z., & Gong, C. (2022). *Cattle face recognition based on a Two-Branch convolutional neural network*. Computers and Electronics in Agriculture, 196, 106871. Elsevier.

[17] Li, Z., Lei, X., & Liu, S. (2022). *A lightweight deep learning model for cattle face recognition*. Computers and Electronics in Agriculture, 195, 106848. Elsevier.

[18] Qiao, Y., Guo, Y., & He, D. (2023). *Cattle body detection based on YOLOv5-ASFF for precision livestock farming*. Computers and Electronics in Agriculture, 204, 107579. https://doi.org/10.1016/j.compag.2022.107579

[19] Yang, W., Wu, J., Zhang, J., Gao, K., Du, R., Wu, Z., Firkat, E., & Li, D. (2023). *Deformable convolution and coordinate attention for fast cattle detection*. Computers and Electronics in Agriculture, 211, 108006. https://doi.org/10.1016/j.compag.2023.108006

[20] Han, S., Fuentes, A., Yoon, S., Jeong, Y., Kim, H., & Park, D. S. (2023). *Deep learning-based multi-cattle tracking in crowded livestock farming using video*. Computers and Electronics in Agriculture, 212, 108044. doi: https://doi.org/10.1016/j.compag.2023.108044

[21] Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A. W., & Burghardt, T. (2021). *Visual identification of individual Holstein-Friesian cattle via deep metric learning*. Computers and Electronics in Agriculture, 185, 106133. doi: 10.1016/j.compag.2021.106133

[22] Gao, J., Burghardt, T., Andrew, W., Dowsey, A. W., & Campbell, N. W. (2021). *Towards Self-Supervision for Video Identification of Individual Holstein-Friesian Cattle: The Cows2021 Dataset*. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop on Computer Vision for Animal Behavior Tracking and Modeling (CV4Animals).

[23] Qiao, Y., Su, D., Kong, H., Sukkarieh, S., Lomax, S., & Clark, C. (2019). *Individual cattle identification using a deep learning based framework*. IFAC-PapersOnLine, 52(30), 318-323. Elsevier.

[24] Xu, X., Wang, Y., Shang, Y., Yang, G., Hua, Z., Wang, Z., & Song, H. (2024). *Few-shot cow identification via meta-learning*. Information Processing in Agriculture. doi: https://doi.org/10.1016/j.inpa.2024.04.001

[25] Lu, Y., Weng, Z., Zheng, Z., Zhang, Y., & Gong, C. (2023). *Algorithm for cattle identification based on locating key area*. Expert Systems with Applications, 228, 120365. https://doi.org/10.1016/j.eswa.2023.120365

[26] Xu, Z., Zhao, Y., Yin, Z., & Yu, Q. (2024). *Optimized BottleNet Transformer model with Graph Sampling and Counterfactual Attention for cow individual identification*. Computers and Electronics in Agriculture, 218, 108703. Elsevier.

[27] Yang, L., Xu, X., Zhao, J., & Song, H. (2023). *Fusion of RetinaFace and improved FaceNet for individual cow identification in natural scenes*. Information Processing in Agriculture. doi: https://doi.org/10.1016/j.inpa.2023.09.001

[28] Weng, Z., Hu, R., & Zheng, Z. (2023). *Study on individual identification method of cow based on CD-YOLOv7*. In Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing (pp. 169-175).

[29] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7464-7475).

[30] Zhou, K., Yang, Y., Cavallaro, A., & Xiang, T. (2019). *Omni-scale feature learning for person re-identification*. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3702-3712).

[31] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., & Zitnick, C. L. (2014). *Microsoft COCO: Common Objects in Context*. In Computer Vision - ECCV 2014 (pp. 740-755). Springer.

[32] Dong, Z., Xu, K., Yang, Y., Bao, H., Xu, W., & Lau, R. W. H. (2021). *Location-Aware Single Image Reflection Removal*. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 5017-5026

[33] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[34] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248-255). doi: 10.1109/CVPR.2009.5206848

[35] Dutta, A., & Zisserman, A. (2019). *The VIA Annotation Software for Images, Audio and Video*. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2276-2279). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3343031.3350535.

[36] de Oliveira Bezerra, A., Ap. de Moraes Weber, V., de Lima Weber, F., Alves de Arruda, Y., da Costa Gomes, R., Hirokawa Higa, G. T., Pistori, H., & Gonçalves Mateus, R. (2024). *Exploring cluster analysis in Nelore cattle visual score attribution*. Smart Agricultural Technology, 8, 100489. doi: https://doi.org/10.1016/j.atech.2024.100489

[37] García, R., & Aguilar, J. (2024). *A meta-learning approach in a cattle weight identification system for anomaly detection*. Computers and Electronics in Agriculture, 217, 108572. doi: https://doi.org/10.1016/j.compag.2023.108572

[38] Neukum, G., Jaumann, R. (2004). *HRSC: The high resolution stereo camera of Mars Express. In: Mars Express: the scientific payload.* Ed. by Andrew Wilson, scientific coordination: Agustin Chicarro. ESA SP-1240, Noordwijk, Netherlands: ESA Publications Division, ISBN 92-9092-556-6, 2004, p. 17-35, 1240, 17-35.

[39] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). *Aggregated residual transformations for deep neural networks*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).

[40] Tan, M., & Le, Q. (2019, May). *Efficientnet: Rethinking model scaling for convolutional neural networks.* In International conference on machine learning (pp. 6105-6114). PMLR.

[41] Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., & Li, Y. (2022, October). *Maxvit: Multi-axis vision transformer.* In European conference on computer vision (pp. 459-479). Cham: Springer Nature Switzerland.

[42] Wang, M., Deng, W. (2021). *Deep face recognition: A survey.* Neurocomputing, 429, 215-244.

[43] S. Myat Noe, T. T. Zin, I. Kobayashi, and P. Tin, "Optimizing black cattle tracking in complex open ranch environments using YOLOv8 embedded multi-camera system," *Scientific Reports*, vol. 15, no. 1, p. 6820, 2025, doi: 10.1038/s41598-024-80742-2.

[44] W. H. E. Mg, T. T. Zin, P. Tin, M. Aikawa, K. Honkawa, and Y. Horii, "Automated cattle monitoring system for calving time prediction using trajectory data embedded time series analysis," *IEEE Open Journal of the Industrial Electronics Society*, vol. 6, pp. 216–234, 2025, doi: 10.1109/OJIES.2025.3533663.

[45] Zhang, Y., Li, L., Li, C., Wang, S., Rong, Y., Niu, K. BECA: A Dataset for Long-Term Recognition In Beef Cattle. figshare https://doi.org/10.6084/m9.figshare.29425316 (2025).

Acknowledgements

Here, we would like to thank Gansu Yimulan Meat Solutions Co., Ltd. and Zhengning Dingwang Meat Solutions Co., Ltd. for their valued support in data collection and annotation for the BECA-D. We are also grateful to Wuwei Xindingle Beef Cattle Breeding Co., Ltd. for their contributions to data collection and annotation for the BECA-L.

Author information

Contributions

Z.Y. is responsible for developing the data collection methodology, deploying cameras on-site, developing collection programs, and researching object detection model and feature extraction model for preliminary classification. L.L. is responsible for optimizing the cattle detection algorithm, collecting training datasets for cattle detection, performing manual fine classification based on preliminary classification of feature extraction model, and annotating the features of each cow. L.C. is responsible for annotating the cattle detection training set and organizing the training set. W.S was responsible for the annotation and validation of the BECA-L. H.Z. is responsible for funding the project, participating in providing algorithms for process optimization, including cattle classification algorithms and data transmission methods. N.K. is involved in the overall process design, including using unsupervised methods to assist in manual preliminary classification and is responsible for ensuring the quality of the paper. Y.R. contributed by providing guidance on algorithm and verifying datasets to ensure accuracy and consistency. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.