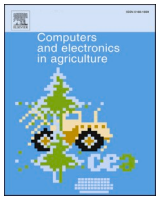
Computers and Electronics in Agriculture 251 (2026) 112047

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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



Original papers

Fusion of keypoint detection and posture-aware correction for body measurement of dairy cattle

Xingshi Xu [a,b,c], Benhai Xiong [d], Hongxing Deng [a,b], Dong Liu [c], Bo Jiang [a], Tomas Norton [c,**], Huaibo Song [a,b,*]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Shaanxi 712100, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Shaanxi 712100, China
[c] Department of Biosystems, Division M3-BIORES: Measure, Model & Manage Bioresponses, Catholic University of Leuven, Kasteelpark Arenberg 30, 3001 Heverlee, Belgium
[d] State Key Laboratory of Animal Nutrition and Feeding, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China

ARTICLE INFO

Keywords:
Precision livestock farming
Cow
Body measurement
Posture-aware correction

ABSTRACT

Accurate and efficient non-contact body measurement of dairy cattle is essential for Precision Livestock Farming (PLF), enabling data-driven monitoring, management, and genetic improvement. Keypoint-based computer vision methods offer a powerful approach to this task. However, two critical challenges hinder its effective implementation: (1) Inaccurate localization of keypoints caused by weak visual features, and (2) Measurement deviations induced by non-standard postures. To address these issues, a novel method named CowSizeNet was proposed, which integrates keypoint detection, initial size estimation, and posture-aware correction into a coherent process. In the first stage, the Cattle Keypoint Detection Network (CKDNet) was developed to achieve reliable keypoint localization. Two dedicated modules introduced in this study, the Bidirectional Feature Enhancement Module (BFEM) and the Hyper-Graph Association Enhancement Module (HGAEM), were incorporated to enrich visual representations and strengthen spatial associations, significantly improving the localization accuracy of visually ambiguous regions. In the second stage, stereo image disparity and camera parameters were utilized to obtain depth information, enabling 3D projection of detected keypoints and initial body size estimation. In the final stage, a Posture-aware Correction Graph Convolutional Network (PC-GCN) was introduced to learn the nonlinear relationships among keypoint geometry, joint angles, and body measurement deviations under varying postures. By regressing and compensating posture-related deviations, the model yielded more accurate and consistent body size estimations. Experiments verified the effectiveness and feasibility of the proposed method. For Body Length (BL), Body Height (BH), and Hip Height (HH), the Mean Absolute Percentage Errors (MAPE) were 4.26%, 1.97%, and 2.19%, respectively, satisfying the accuracy requirements for automated body measurement in PLF. Furthermore, the design and development of supporting terminal devices and visualization platforms further enhanced the practical value of the proposed method, promoting its use in dairy farm management.

Abbreviations: PLF, Precision Livestock Farming; GWAS, Genome-Wide Association Studies; CKDNet, Cattle Keypoint Detection Network; BFEM, Bidirectional Feature Enhancement Module; HGAEM, Hyper-Graph Association Enhancement Module; PC-GCN, Posture Correction Graph Convolutional Network; SSIM, Structural Similarity Index Measure; RANSAC, RANdom SAmple Consensus; DETR, DEtection Transformer; SFAM, Semantic Focus Attention Map; BGAM, Boundary Guidance Attention Map; LPF, Laplacian Pyramid Filtering; BL, Body Length; BH, Body Height; HH, Hip Height; IoU, Intersection over Union; OKS, Object Keypoint Similarity; MAE, Mean Absolute Error; MSD, Mean Squared Deviation; MAPE, Mean Absolute Percentage Error; SGD, Stochastic Gradient Descent; B/S, Browser–Server.

* Corresponding author at: College of Mechanical and Electronic Engineering, Northwest A&F University, China.
** Corresponding author.
E-mail addresses: xingshixu@nwafu.edu.cn (X. Xu), xiongbenhai@caas.cn (B. Xiong), denghongxing@nwafu.edu.cn (H. Deng), dong.liu@kuleuven.be (D. Liu), jiangbo@nwafu.edu.cn (B. Jiang), tomas.norton@kuleuven.be (T. Norton), songhuaibo@nwsuaf.edu.cn (H. Song).

https://doi.org/10.1016/j.compag.2026.112047
Received 27 December 2025; Received in revised form 27 April 2026; Accepted 13 June 2026

1. Introduction

Precision Livestock Farming (PLF) enables intelligent monitoring and regulation of both individual and herd health by collecting and analyzing large volumes of livestock growth data. This approach facilitates optimized management strategies, supports data-driven decision-making, and enhances farming efficiency, economic returns, and animal welfare (Norton and Berckmans, 2018; Qiao et al., 2021; Xu et al., 2024; Zhang et al., 2025; Xu et al., 2026). Body size of cattle serves as a key indicator in herd management, reflecting growth and development status, and is widely used in estimating body weight, assessing body condition scores, and supporting health monitoring and precision management. Moreover, when integrated with Genome-Wide Association Studies (GWAS), body size data can help identify candidate genes associated with economically important traits, advancing molecular breeding and genetic improvement (Cominotte et al., 2020; Ruchay et al., 2022; Bai et al., 2025; Huang et al., 2025a; Li et al., 2025b). Traditional body measurement methods rely on manual contact-based tools such as measuring tapes, which are time-consuming, labor-intensive, and prone to inducing stress in cattle, rendering them inadequate for modern large-scale farming (Deng et al., 2025; Jing et al., 2025). Therefore, the development of accurate and efficient non-contact body measurement is urgently needed and has attracted increasing attention.

Non-contact cattle body measurement aims to extract body size parameters using computer vision technologies without disturbing animals' normal behavior. Given that the localization accuracy of measurement keypoints directly affects the final measurement results, researchers have increasingly focused on achieving rapid and precise keypoint detection after acquiring the raw data to improve measurement accuracy (Li et al., 2022b; Li et al., 2025b; Xu et al., 2025; Yang et al., 2025). Early approaches heavily relied on manual intervention, such as attaching markers to specific anatomical sites on the cattle or manually selecting keypoint in the point cloud by experts, which rendered the overall process semi-automated (Ruchay et al., 2020; Li et al., 2022a; Yang et al., 2022). To eliminate dependence on manual assistance, some methods attempted keypoint localization based on morphological features (Zhang et al., 2018; Zhang et al., 2024; Hou et al., 2025). These approaches inferred keypoint positions through clustering, curvature estimation, or polynomial fitting; however, they tended to be sensitive to parameter settings and lacked sufficient accuracy and robustness. Several studies introduced part segmentation methods that indirectly selected measurement points by identifying different anatomical regions, thereby improving accuracy to some extent (Li et al., 2022b; Zhou et al., 2025). More recently, some scholars have devoted efforts to directly and automatically localize keypoint from data to avoid the cumbersome procedures and cumulative errors inherent in traditional methods. Among these, some regress keypoint directly in the 3D point cloud, though such approaches are highly dependent on point cloud quality and require accurate pre-segmentation of livestock point cloud from the raw data (Zhou et al., 2025). Others integrate 2D RGB images with depth data, employing deep learning models to detect keypoint in RGB images first and then project them into corresponding 3D space to complete the measurement (Du et al., 2022; Li et al., 2024; Deng et al., 2025; Yang et al., 2025; Du et al., 2026). These methods feature streamlined and stable workflows, enabling end-to-end body measurement with reliable and efficient performance.

Despite significant advances in non-contact cow body measurement techniques in recent years and many milestone achievements, the following critical issues remain unresolved:

(1) ***Keypoint localization errors caused by weak visual features.*** Inaccurate localization of the keypoint can lead to significant errors or even failure in cattle body measurement. However, critical measurement points such as the shoulder, wither, and buttock often lack distinct visual features, which constitutes a primary challenge for reliable keypoint detection (Li et al., 2024; Yang et al., 2025). Du et al. (2022) improved the saliency of key regions by applying white markings, a technique that demonstrates clear advantages under experimental conditions. Nevertheless, its dependence on preparatory marking procedures suggests that more automated solutions remain desirable for real-world applications.

(2) ***Body measurement deviations caused by non-standard postures.*** Conventional body size definitions rely on measurements obtained in standard posture, with symmetrical limb positioning, a straight back, and the head facing forward. However, in practice, cattle images are often captured while the animals are walking, feeding, or engaging in other movements. These conditions cause changes in the relative positions of skeletal joints, often resulting in the underestimation or overestimation of key body dimensions compared to standard-posture-based measurements (demonstrated in Fig. 7). The impact of non-standard postures on measurement accuracy has been widely recognized as a significant source of error (Hou et al., 2025; Lu et al., 2025; Zhou et al., 2025). However, systematic investigations into this issue and the development of effective posture correction methods remain limited (Yin et al., 2022; Li et al., 2023; Han et al., 2025). Consequently, it is crucial to develop methods that correct posture-related measurement errors, ensuring more accurate and consistent body size estimation under realistic conditions.

To address the abovementioned issues, a method named CowSizeNet was proposed in this study, which enables accurate body measurement of dairy cattle through keypoint detection and posture-aware correction. In short, the initial body size was estimated based on keypoint and depth information, and then an adjusted body size was obtained via posture-aware correction. The primary objectives were: (1) to develop a robust keypoint detection network (CKDNet) that integrates edge-sensitive and high-order semantic features for accurate localization in visually ambiguous regions, (2) to develop a learnable posture-aware correction model (PC-GCN) that mitigates measurement deviations caused by non-standard cattle postures; and (3) to evaluate the accuracy and robustness of body measurement using the partitioned test set and an independent dataset of unseen individuals, while advancing the potential application of the method through a system prototype implementation.

The main contributions of this work were summarized as follows:

(1) A novel method, CowSizeNet, was proposed to support automated non-contact cow body measurement.
(2) Measurement errors caused by inaccurate keypoint localization were effectively reduced by CKDNet, which integrates semantic and boundary cues via BFEM and captures global high-order correlations through HGAEM to enhance localization in visually ambiguous regions.
(3) Measurement deviations arising from non-standard cattle postures were analyzed and resolved using PC-GCN, which regresses compensation values to align multi-posture measurements with the standard posture.
(4) Extensive experiments and a system prototype implementation demonstrated the model's effectiveness, robustness, and potential value for practical PLF applications.

2. Materials

2.1. Data acquisition

Data collection was conducted from January to June 2025 and in April 2026 at three commercial dairy farms in Yangling District, Shaanxi Province, China. A total of 49 lactating Holstein cows were involved, with all animals having a parity ranging from 1 to 2. A ZED-2i stereo camera (Stereolabs Inc.), equipped with 2.1 mm focal length lenses, was employed for data acquisition. The camera featured a 119.73 mm

baseline, and at a resolution of 1920 × 1080 pixels, the intrinsic parameters of the left sensor were $f_x = 1067.87$, $f_y = 1067.84$, $c_x = 969.51$ and $c_y = 537.39$. Operators positioned the camera 3 to 5 m from the lateral side of the cattle at a height of 80 to 120 cm.

To capture a diverse range of postures such as standing, walking, and feeding, multiple videos were recorded for each animal while they moved naturally without human intervention. These recordings typically ranged from 1 to 7 clips per individual, with an average of 3 clips. Crucially, all video data for a specific cow were captured within a single day to ensure strict synchronization with manual ground truth measurements of body size and to avoid potential deviations caused by temporal changes such as animal growth.

The recorded ".svo" files were decoded using the official tool ZED_SVO_Export.exe to extract synchronized left and right image pairs. Frames were extracted at a 2-second interval to capture meaningful postural shifts based on preliminary observations of cattle behavior, thus reducing data redundancy. Subsequently, the dataset was further refined by applying the Structural Similarity Index Measure (SSIM) (Wang et al., 2004) to remove visually similar frames. After manually removing invalid samples, such as frames without cattle, the remaining data were used for subsequent processing.

2.2. Data construction

To support the development and evaluation of the proposed method, two datasets were constructed: (1) a cow keypoint detection dataset for the CKDNet (Fig. 1(a)), and (2) a measurement deviation regression dataset for the PC-GCN (Fig. 1(b)).

For cow keypoint detection dataset, images captured by the left lens of the camera were used. Right-view images (i.e., those captured by the right lens) were not used in this stage, as they offered redundant information for the 2D detection task due to their high similarity to the left-view images. Specifically, the small baseline (119.73 mm) relative to the shooting distance (3–5 m) results in a horizontal disparity of only approximately 1.5% of the image width. Furthermore, as the sensors are hardware-synchronized, the cow's posture remains virtually identical in both views; consequently, including right-view images would significantly increase manual annotation and computational overhead without providing meaningful perspective diversity. Moreover, considering that the data collection took place in semi-open and open-air farming environments, characterized by complex backgrounds and variable lighting, additional data from (Sheng et al., 2025) were incorporated to supplement indoor scenes, which feature simpler backgrounds but lower-light conditions, thereby enhancing overall environmental diversity. Under the guidance of animal science experts (two professors specializing in animal anatomy and livestock phenotyping) and based on bovine anatomical structures, all images were manually annotated using Labelme software. Specifically, each cow was labeled with 21 anatomical keypoints and one bounding box. The keypoint definitions and their skeleton structure are illustrated in Fig. 1(a). The 21 keypoints correspond to the: (1) nose, (2) head, (3) neck, (4) wither, (5) back, (6) buttock, (7) ischium, (8) left shoulder, (9) right shoulder, (10) left elbow, (11) right elbow, (12) left wrist, (13) right wrist, (14) left front hoof, (15) right front hoof, (16) left stifle, (17) right stifle, (18) left hock, (19) right hock, (20) left rear hoof, and (21) right rear hoof. These keypoints were subsequently used to compute the initial body size when combined with depth information, and also served as input to the PC-GCN model to estimate compensation values, thereby obtaining the adjusted body size.

For the measurement deviation regression dataset, the data structure is illustrated in Fig. 1(b) and consists of five components: *File_Name*, *Initial_Body_Size*, *Ground_Truth*, *Keypoint_Location*, and *Joint_Angle*. The *File_Name* serves to identify each data entry and link it to the corresponding image. The *Initial_Body_Size* refers to body measurement parameters obtained by projecting keypoints into real 3D space, typically collected when the cattle are in non-standard postures. The *Ground_Truth* represents the actual body measurements acquired through manual measurement while the cattle are in a standard posture. To ensure the reliability of the *Ground_Truth* values, all manual measurements were conducted by two experienced technicians (farm technical staff with over two years of practical experience in body measurement) using a calibrated measuring stick. During measurement, each cow was gently restrained in a headlock system, and its limbs were positioned symmetrically to maintain a standard upright posture. These procedures effectively minimized animal movement and operator-induced variability, thereby ensuring the accuracy and consistency of the manual reference measurements. The *Keypoint_Location* corresponds to the normalized 2D coordinates of 21 keypoints within the bounding box. All data are stored in a.csv file, where each row represents a complete data record. For each data instance, the *Keypoint_Location* serves as the input to the model, while the difference between the *Ground_Truth* and *Initial_Body_Size* is used as the target output for the regression model, representing the measurement deviations caused by non-standard postures.

The data were categorized into an internal dataset involving 39 Holstein cows from the 2025 sessions and an external test set featuring 10 additional cows from the April 2026 session. The internal dataset was random partitioned into training, validation, and test sets at a ratio of 7:2:1. Specifically, the keypoint detection dataset comprised 3414 internal images, with 2390 used for training, 682 for validation, and 342 for testing. The measurement deviation regression dataset included 1398 samples, which were split into 978 for training, 280 for validation, and 140 for testing. Furthermore, the external test set consisting of 400 images and corresponding manual measurements from 10 cows was utilized to further evaluate the model's generalization to unseen animals.

3. Methods

In this study, a novel method for dairy cattle body measurement was

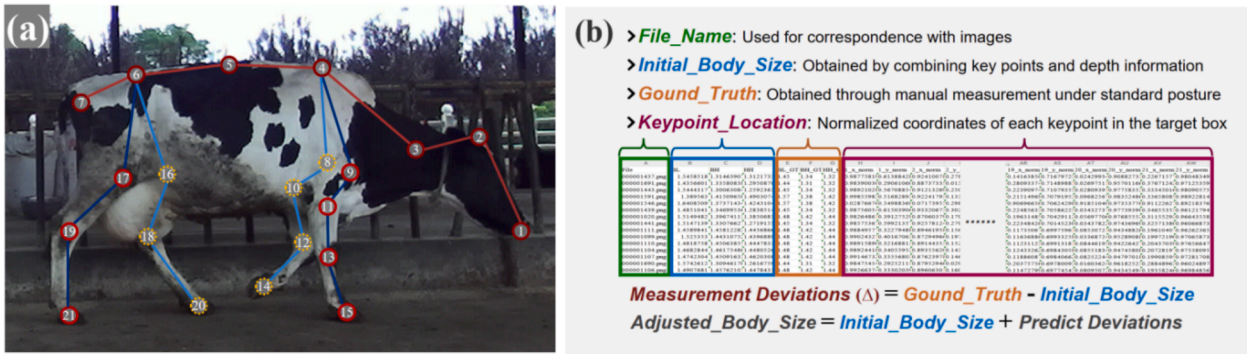


Fig. 1. Organizational form of data. (a) keypoint annotation protocol; (b) organizational structure of the measurement deviation regression dataset.

proposed. As illustrated in Fig. 2, the method consists of three primary components: keypoint detection, initial body size estimation, and posture-aware correction.

In the first stage, images captured by the left lens of a ZED stereo camera were processed using the proposed keypoint detection model CKDNet. This model automatically identifies the bounding box of each cow and localizes 21 anatomical keypoints, which are used in the subsequent two stages.

In the second stage, disparity between the left and right views was estimated using the DEFOM-Stereo (Jiang et al., 2025). Combined with the camera's intrinsic parameters, the depth map of the left view was obtained, allowing projection of each visible keypoint into the world coordinate system. RANdom SAmple Consensus (RANSAC) (Fischler and Bolles, 1981) was then employed to fit the ground plane. Initial body measurements were calculated as Euclidean distances either between pairs of keypoints or between keypoints and the ground plane.

In the third stage, the previously detected keypoints were used to compute normalized coordinates relative to the bounding box, along with a set of joint angle features. These features were fed into the proposed PC-GCN model, which regresses the deviations in body size caused by non-standard postures. By correcting the initial measurements with these predicted deviations, a final adjusted body size was obtained, yielding more accurate results.

3.1. Keypoint detection

DEtection TRansformer (DETR) and its variants have attracted increasing attention from the research community due to their promising performance and end-to-end operation paradigm (Carion et al., 2020; Zhao et al., 2024; Huang et al., 2025b). Recent studies have explored the adoption of DETR's encoder-decoder architecture for keypoint detection tasks, demonstrating superior performance in terms of localization accuracy and robustness (Shi et al., 2022; Yang et al., 2023b; Dai and Ling, 2025; Yang et al., 2025). Compared with conventional top-down approaches, the single-stage method eliminates intermediate steps, thereby enabling a more streamlined and efficient detection pipeline.

In light of these advantages, a single-stage model tailored for cattle keypoint detection was proposed. As depicted in Fig. 3, the model comprises three major components: a backbone, an efficient encoder, and a transformer decoder. The backbone employed ResNet-50 (He et al., 2016) as the feature extractor to capture the initial visual features of the input image. The efficient encoder was designed to further refine the multi-scale feature maps extracted from the backbone, with its carefully designed DFEM and HGAEM modules aimed at enhancing the representation of visual cues and the modeling of semantic relationships. The transformer decoder was based on the design of ED-Pose (Yang et al., 2023b) and was responsible for predicting the cattle bounding box locations and keypoint coordinates.

Specifically, an input image $I \in \mathbb{R}^{H_I \times W_I \times 3}$ was processed by the backbone to generate a pyramid of multi-scale feature maps. The feature maps from the last three stages $\{C_3, C_4, C_5\}$ were selected, and 1×1 convolutions were applied to unify their channel dimensions. This process yielded the input set for the efficient encoder, denoted as $\{P_3, P_4, P_5\}$, where $P_3 \in \mathbb{R}^{H \times W \times C}$, $P_4 \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times C}$, $P_5 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C}$. In this work, the spatial and channel dimensions were set as $H = 92, W = 164, C = 256$, respectively. Within the efficient encoder, the features were refined through a hybrid path-aggregation structure. In addition to the BFEM and HGAEM modules (detailed in subsequent sections), the "Up Sampling" operation utilized Nearest-Neighbor Interpolation with scale of 2, while the "Conv" operation was executed via 3×3 convolutions with a stride of 2. Feature fusion was achieved by concatenating maps along the channel dimension followed by a RepC3 block. This architecture ensured that the final output features $\{F_3', F_4', F_5'\}$ preserve the original spatial resolutions of $\{P3, P4, P5\}$, respectively. Finally, the refined output features $\{F_3', F_4', F_5'\}$ were flattened and concatenated to form the visual feature sequence V_F. Following the ED-Pose (Yang et al., 2023b) architecture, a set of object queries Q was employed to extract target-related features from V_F. Instance queries undergone query expansion into collaborative groups to capture both global and local features. Bounding boxes and keypoints were then jointly regressed via Feed-Forward Networks (FFNs) from these expanded embeddings, ensuring inherent spatial consistency and alignment between the two tasks.

3.1.1. Bidirectional feature enhancement module

In the task of cattle keypoint detection, certain keypoint regions such as the shoulder, back, and rump exhibit relatively weak visual features, making accurate localization challenging. To enhance the localization precision, the BFEM was proposed, which focused on optimizing the shallow feature map P_3 with high spatial resolution. The structure of BFEM was illustrated in Fig. 4(a).

BFEM consists of two independent branches, each designed to generate a Semantic Focus Attention Map (SFAM) and a Boundary Guidance Attention Map (BGAM), respectively, thereby enhancing the P_3 feature map by modulating it from both semantic and boundary perspectives. Specifically, the first branch leverages the rich contextual and semantic information embedded in deeper features (i.e., P_5) to guide an attention that focuses on regions more likely to contain keypoints. This process involves a convolution with a 1×1 kernel to integrate cross-channel information, followed by an up-sampling operation to match the spatial resolution of P_3. A sigmoid activation then generated the SFAM, providing preliminary semantic localization cues for keypoints. This procedure could be formalized using Eq. (1):

$$\mathrm{SFAM} = \sigma(U(Conv(\mathrm{P}_5))) \tag{1}$$

where, $Conv(\cdot)$ denotes the 1×1 convolutional operation followed by Batch Normalization and ReLU activation, maintaining the channel dimension ($C = 256$), $U(\cdot)$ represents the up-sampling operation executed via corner-aligned bilinear interpolation to specific spatial dimensions ($H \times W$), and $\sigma(\cdot)$ refers to the Sigmoid function.

The second branch focused on enhancing the role of boundary information in keypoint localization. Compared to deeper features where edge details tend to be lost, shallow features are better at preserving image structure and fine details. Therefore, this branch utilized P_3 as input and employed the proposed Laplacian Pyramid Filtering (LPF) to explicitly extract edge-aware structural information, thereby enhancing the boundary awareness in the vicinity of keypoints. The resulting features were then passed through a Sigmoid function to generate the BGAM, which refines the boundary-relevant response regions in P_3,

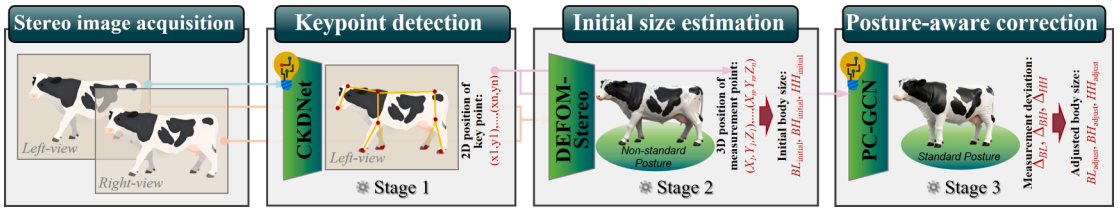


Fig. 2. Overview of the CowSizeNet pipeline.

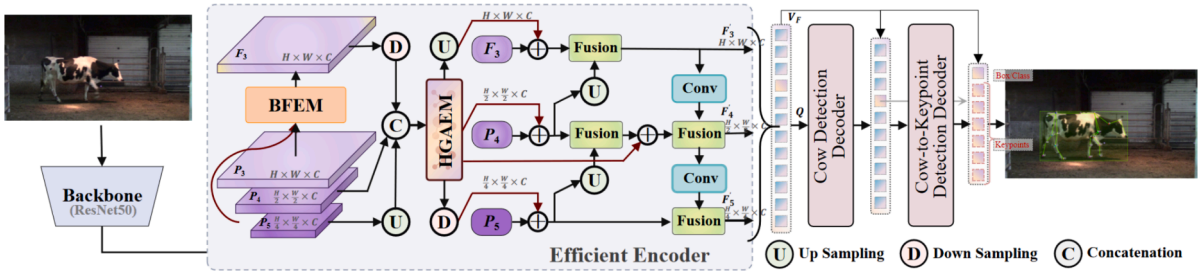


Fig. 3. Structure of the CKDNet.

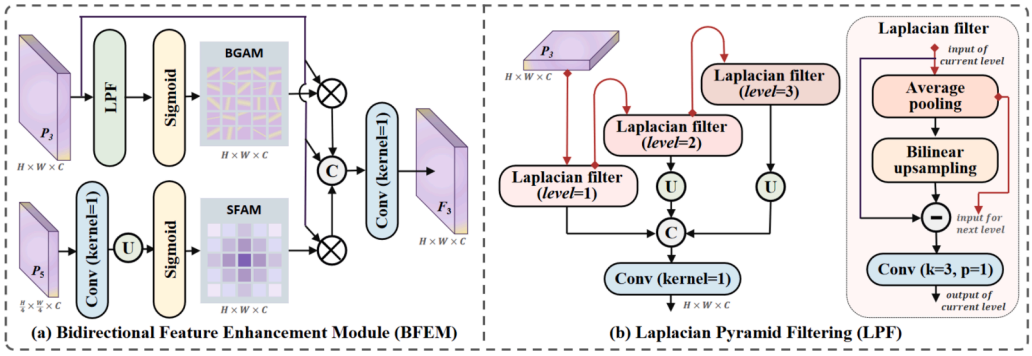


Fig. 4. Structure of BFEM.

ultimately improving the accuracy of keypoint localization. The above process can formalized as Eq. (2):

$$\mathrm{BGAM} = \sigma(f_{LPF}(\mathrm{P}_3)) \tag{2}$$

where, $f_{LPF}(\cdot)$ denotes LPF operation, which was detailed in Fig. 4(b). This module efficiently extracted multi-scale boundary features by implementing a 3-level cascaded structure. Specifically, at each level, the input first undergone 2×2 average pooling. This downsampled feature map served as the input for the subsequent cascade level, while concurrently undergoing bilinear interpolation. The interpolated map was then subtracted from the original input to compute the Laplacian residual, which was subsequently processed by a learnable 3×3 convolutional block (padding = 1). All feature maps were then up-sampled to a uniform spatial resolution, concatenated along the channel dimension, and passed through a learnable 1×1 convolution to fuse the multi-scale features and adjust the channel dimensionality to match that of P_3. Formally, the LPF can be expressed as Eq.(3):

$$f_{LPF}(\mathrm{P}_3) = Conv(\mathrm{Concat}[\mathscr{L}_1(\mathrm{P}_3), U(\mathscr{L}_2(\mathrm{P}_3)), U(\mathscr{L}_3(\mathrm{P}_3))]) \tag{3}$$

where, $\mathrm{Concat}[\cdot,\cdot]$ refers to channel-wise concatenation, $\mathscr{L}_i(\cdot)$ denotes the output of the i-th level Laplacian filter.

Finally, the generated SFAM and BGAM are respectively applied to weight the shallow feature map P_3 via element-wise multiplication. The two enhanced results are concatenated with the original P_3 along the channel dimension and integrated through a 1×1 convolution to produce the final enhanced feature map, which is then fed into subsequent modules. This fusion process was expressed as Eq. (4):

$$F_3 = Conv(\mathrm{Concat}[(\mathrm{SFAM} \otimes \mathrm{P}_3), (\mathrm{BGAM} \otimes \mathrm{P}_3), \mathrm{P}_3]) \tag{4}$$

where, $\otimes$ denotes element-wise multiplication, and $Conv(\cdot)$ represents a 1×1 convolutional operation that reduces the concatenated channel dimension from $3C$ back to C, matching the original channels of P_3.

3.1.2. Hyper-graph association enhance module

Accurate localization of cattle keypoints typically requires comprehensive reasoning based on prior knowledge and semantic cues, such as global posture and anatomical structure. Humans, benefiting from such prior knowledge of animal morphology, can accurately infer keypoint locations even under challenging conditions like self-occlusion (Chen et al., 2024). However, most existing models are built upon convolutional neural architecture or self-attention mechanism, which focus on local features or pairwise correlations, thus falling short in capturing complex, global, high-order semantic relationships. Recent studies have demonstrated that hypergraphs offer unique advantages in modeling high-order correlations among multiple semantic components (Feng et al., 2025; Fixelle, 2025; Wang et al., 2025). Unlike traditional graph, where each edge connects only two nodes, a hypergraph allows each hyperedge to connect multiple nodes simultaneously, thereby enabling more flexible representation of complex semantic associations across spatial regions. Motivated by this, HGAEM was proposed to efficiently capture high-order semantic interactions across the entire feature.

The overall architecture of HGAEM was illustrated in Fig. 5(a). This input feature, denoted as X_{in}, was first processed by two parallel 1×1 convolutional layers to generate $X_{enhance}$ and $X_{lateral}$. Among them, $X_{enhance}$ was used for hyper-graph computation. The resulting feature was then connected with $X_{lateral}$ followed by a 1×1 convolutional layer to produce the final output. The above process can be formally expressed as Eqs. (5) and (6):

$$X_{enhance} = Conv(X_{in}) X_{lateral} = Conv(X_{in}) \tag{5}$$

$$X_{out} = Conv(Concat[f_{HC}(X_{enhance}), X_{lateral}]) \tag{6}$$

where, $f_{HC}(\cdot)$ denotes the hypergraph computation process, which consists of two stages: hyper-edge generation and hyper-graph convolution.

In this study, hypergraph $G=\{V, E\}$ was constructed to capture and model high-order semantic dependencies among different spatial regions within the image. The hypergraph consists of N vertices (V) and M hyper-edges (E), where each hyper-edge was connected to all vertices with continuously differentiable connection strengths. Accordingly, the connectivity of the hypergraph is represented by a matrix $\mathscr{A} \in [0,1]^{N \times M}$, where, $A_{i,j}$ denotes the association degree between the i-th vertex and the j-th hyper-edge.

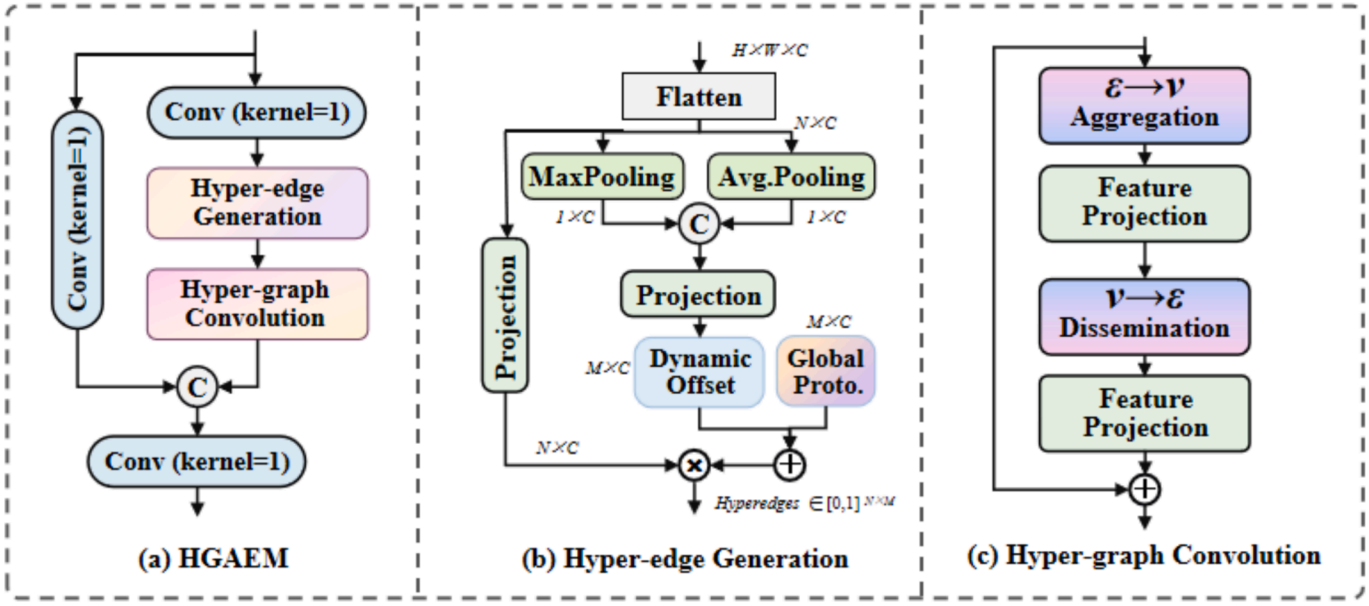


Fig. 5. Structure of Hyper-Graph Association Enhancement Module.

During the hyper-edge generation stage, hyper-edges were dynamically constructed based on $X_{enhance}$. As shown in Fig. 5(b), $X_{enhance} \in \mathbb{R}^{H\times W\times C}$ was first flattened into a set of vertices $X_{ver} = \{x_{ver_1}, x_{ver_2}, \ldots, x_{ver_N}\} \in \mathbb{R}^{N\times C}$, where each spatial location in the feature map is regarded as a vertex and $N = H * W$. On the right branch, MaxPooling and Avg-Pooling operations were applied to extract global features, which are then concatenated and passed through a projection to generate dynamic offsets. These offsets were added to a set of learnable global prototypes, forming the final M hyper-edge prototypes. On the left branch, vertex features were projected to obtain query vectors, which were then compared with the M hyperedge prototypes to compute similarity scores. These scores indicate the association degree to which each vertex participates in each hyper-edge, thereby constructing the matrix A.

During the hyper-graph convolution stage, information is aggregated across hyperedges to update vertex features based on high-order interactions. As shown in Fig. 5(c), each hyper-edge aggregated contributions from all vertices, weighted by their association degrees, and these undergo a linear projection to generate a representation for each hyper-edge. This representation was then disseminated back to the vertices, enhancing their semantic expressiveness and enabling cross-region information fusion and feature refinement. This process can be formally described by Eqs. (7) and (8).

$$f_j = \sigma\left(W_e \sum_{i=1}^{N} \mathscr{A}_{i,j} x_i\right) \tag{7}$$

$$\widetilde{x}_i = \sigma\left(W_v \sum_{j=1}^{M} \mathscr{A}_{i,j} f_j\right) \tag{8}$$

where, W_e and W_v denote the hyperedge and vertex projection weights, respectively.

3.2. Initial body size estimation

The detailed workflow of this stage was illustrated in Fig. 6. To obtain depth information, the DEFOM-Stereo method was employed to estimate the disparity between the left and right images (Jiang et al., 2025). As one of the most accurate stereo matching models currently available, this neural network–based approach is more robust and precise than traditional stereo matching methods, and its effectiveness in livestock body measurement tasks has been demonstrated in our previous work (Deng et al., 2025). Based on the intrinsic parameters of the stereo camera, the computed disparity map was converted into a corresponding depth map, where the depth of each pixel was calculated as Eq. (9)

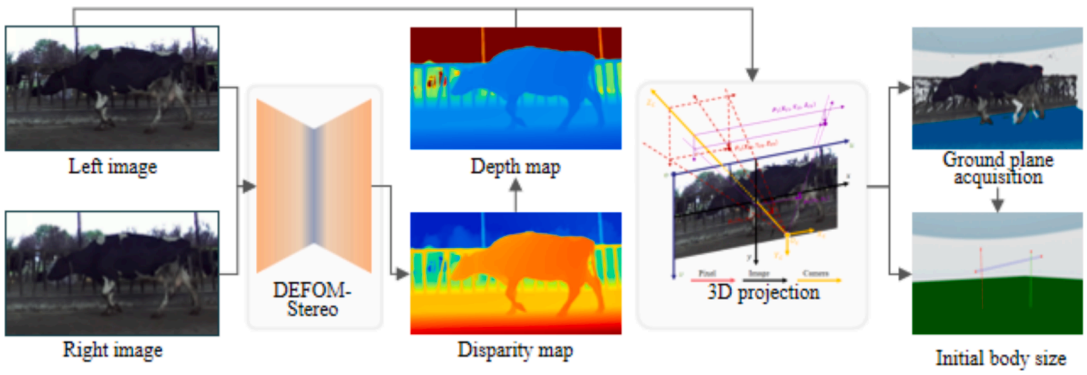


Fig. 6. Schematic diagram of initial body size estimation.

$$Z_{i,j} = \frac{f \cdot B}{D_{i,j}} \tag{9}$$

where, $D_{i,j}$ denotes the disparity value at a given pixel, B is the baseline distance between the two cameras, f is the focal length of the camera, and Z represents the obtained depth map.

The RGB-D image was transformed from the pixel coordinate system to the 3D camera coordinate system using the camera's intrinsic parameters. The ground plane was extracted using the RANSAC (Fischler and Bolles, 1981) method, and initial body size parameters were computed based on the Euclidean distances between keypoints and between keypoints and the ground plane. Specifically, Body Length (BL) was obtained as the distance from the ischium (Point 7) to either the left shoulder (Point 8) or right shoulder (Point 9); Body Height (BH) was defined as the vertical distance from the withers (Point 4) to the ground plane; and Hip Height (HH) was defined as the distance from the buttock (Point 6) to the ground plane.

3.3. Posture-aware correction

As shown in Fig. 7, the body size parameters of dairy cattle are partly affected by posture variations under unconstrained conditions. For example, when a cow is walking, the limb movement can cause observable fluctuations in the measurement of body length BL_{mea}. When a cow lowers its head to feed, the measured body height BH_{mea} is typically lower compared with the standing posture. In traditional manual measurement, operators use professional equipment to fix the cow's head and align the four legs in parallel, so as to minimize the influence of posture differences. However, such physical constraint is nearly infeasible for non-contact automated measurement methods.

To address the above challenge, PC-GCN was proposed to compensate for the deviations in body size estimation caused by posture variations. As shown in Fig. 7, the normalized coordinates of 21 keypoints, together with the key joint angles calculated from their positions, were utilized to represent the overall posture of the cow. In the node-skeleton feature branch, each keypoint was regarded as a node in the graph structure, with its coordinates used as node features. These features were aggregated through three layers of graph convolution along the skeletal adjacency structure to capture spatial dependencies among keypoints, and a global pooling layer was applied to obtain the skeletal feature vector F_{ske}. In the joint-angle feature branch, the key joint angles were encoded by a multilayer perceptron to produce the angle feature vector F_{ang}. Finally, the two feature vectors were fused and fed into a linear regression module to predict deviations in body length, body height, and hip height, thereby enabling automatic correction of body measurement deviations under non-standard postures.

In this study, ten joint angles were selected as global posture descriptors, including the left and right fore-wrist (Points 10–12-14 and 11–13-15 in Fig. 1 (a), respectively), fore-elbow (Points 8–10–12 and 9–11–13), hind-hock (Points 16–18-20 and 17–19-21), and hind-knee angles (Points 6–16–18 and 6–17–19), as well as the neck–shoulder (Points 3–4–5) and back angles (Points 4–5-6). Each angle was defined by three keypoints and computed from their two-dimensional coordinates, as formulated in Eq. (10).

$$\theta = \arccos \frac{(a-b)\cdot(c-b)}{\|a-b\|\cdot\|c-b\|} \tag{10}$$

where, a, b, and c represent the 2D coordinates of the corresponding keypoints, with b acting as the joint center. To ensure that the calculated angles are consistent with the cow's head orientation and the actual direction of joint opening, a correction strategy was introduced. Specifically, the head orientation (left or right) was determined according to the relative horizontal positions of the wither (Point 4) and the buttock (Point 6). If the calculated joint opening direction was inconsistent with the expected one, the complementary angle was adopted as the final value. The above process can be described as Eqs. (11)–(13).

$$H = sgn(x_6 - x_4) \tag{11}$$

$$CP = u_x v_y - u_y v_x \tag{12}$$

$$\theta = \begin{cases} \theta & , H \cdot CP < 0 \\ 360^\circ - \theta & , H \cdot CP > 0 \end{cases} \tag{13}$$

where, x_4 and x_6 denote the horizontal coordinates of the wither and the buttock, respectively; $sgn(\cdot)$ is the sign function; u_x, u_y and v_x, v_y represent the 2D components of the vectors $u = a - b$ and $v = c - b$; H is the head orientation indicator; and CP is the 2D cross-product.

3.4. Evaluation metric

The accuracy of body measurement in dairy cattle is closely related to the detection precision of both the target box and the keypoints. To evaluate the performance of the proposed CKDNet model, the Average Precision of detection boxes AP_{box} and Average Precision of keypoints AP_{kp} were adopted as evaluation metrics, following the COCO standard. In the calculation of AP_{box}, the Intersection over Union (IoU) was used as a quantitative measure to evaluate the overlap degree between the predicted and ground-truth boxes. In the calculation of AP_{kp}, the Object Keypoint Similarity (OKS) was adopted to represent the closeness between the predicted and true keypoints. The IoU and OKS metrics used in this study are defined as formulated in Eqs. (14)–(15).

$$\text{IOU} = \frac{\text{Area}(B_p \cap B_g)}{\text{Area}(B_p \cup B_g)} \tag{14}$$

$$\text{OKS} = \frac{\sum_{k=1}^{K} \exp(-d_k^2 / 2s^2\sigma_k^2) \cdot \delta_k}{\sum_{k=1}^{K} \delta_k} \tag{15}$$

where, B_p and B_g represent the predicted and ground-truth bounding boxes, and $\text{Area}(\cdot)$ denotes the area of the corresponding region. K is the total number of keypoints, d_k is the Euclidean distance between the k-th predicted keypoint and its ground-truth counterpart, s indicates the object scale, σ is a constant, and δ_k denotes the visibility flag of the keypoint.

To evaluate the performance of the proposed cattle body measurement method, CowSizeNet, the manually measured body size parameters were used as the gold standard. Three quantitative metrics were employed, including Mean Absolute Error (MAE), Mean Squared Deviation (MSD), and Mean Absolute Percentage Error (MAPE). MAE and MAPE were used to assess the deviation between the estimated and reference values, whereas MSD quantified the overall variability of the measurement errors. Lower values of these metrics indicate superior measurement performance.

3.5. Experimental environment and implementation details

The experiments in this study were conducted on the Windows 11 operating system. The hardware configuration included a 12th Gen Intel (R) Core(TM) i5-12600KF CPU (10 cores, 16 threads, base frequency 3.69 GHz), a 24 GB NVIDIA GeForce RTX 3090 GPU, 128 GB of RAM, and an 8 TB hard drive. Python 3.8 served as the programming language, while PyTorch 2.1 was utilized as the deep learning framework, with CUDA 11.8 providing the acceleration environment. PyCharm was employed as the programming platform. All comparison algorithms were executed under the same environment (see Fig. 8).

During the training of CKDNet, the learning rate was set to 0.01, the batch size to 4, and the number of training epochs to 100. The Stochastic Gradient Descent (SGD) optimizer was employed, with a weight decay of 5×10^{-4} and a momentum factor of 0.9. For PC-GCN, the learning rate was set to 1×10^{-5}, the batch size to 32, and the training was conducted for 1000 epochs. The Adam optimizer was used, with a weight decay of

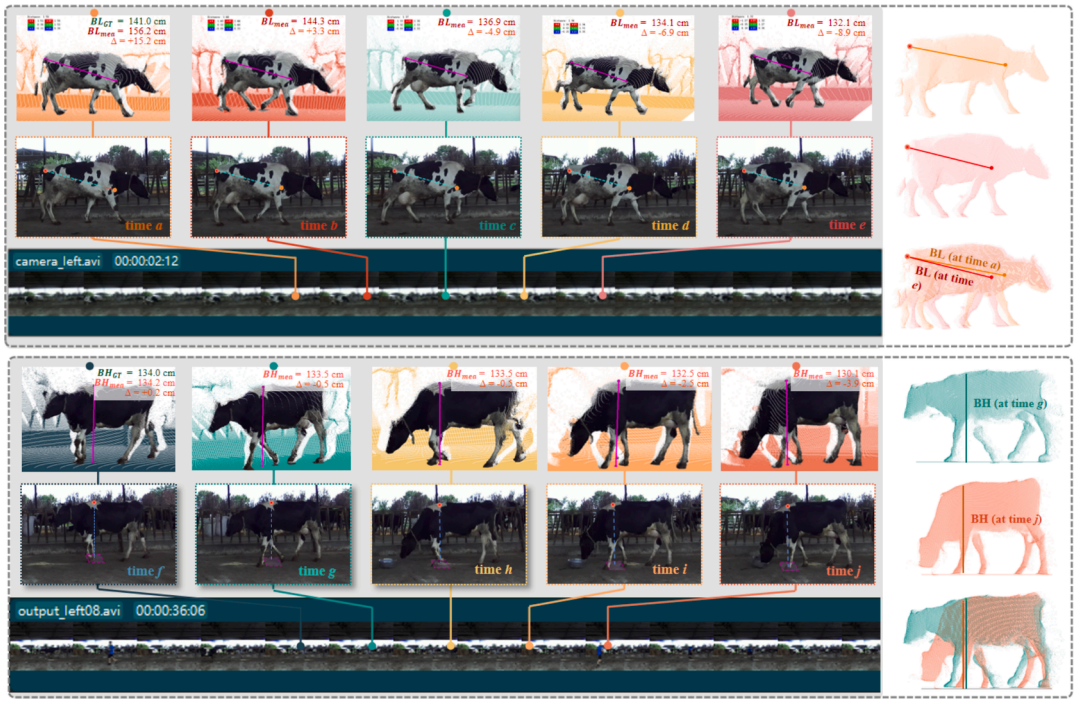


Fig. 7. Variations in body measurements due to different non-standard postures of dairy cattle.

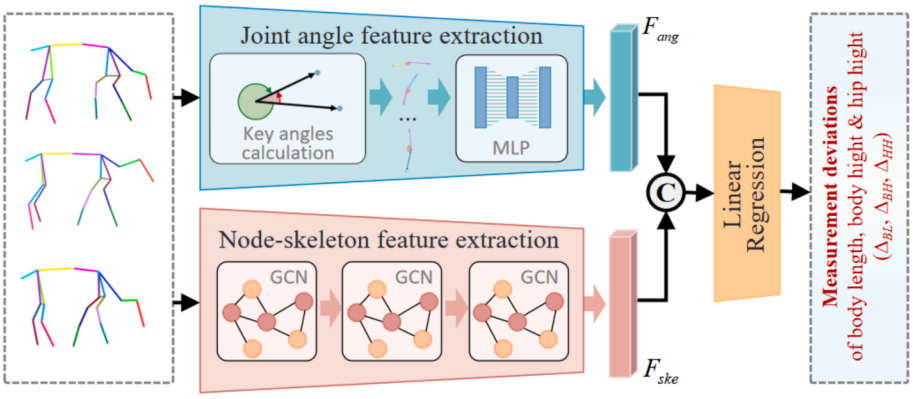


Fig. 8. Structure of PC-GCN.

1×10^{-4}.

4. Results and analysis

4.1. Performance of the CKDNet

4.1.1. Comparison experiments of different keypoint detection models

To comprehensively evaluate the performance of the proposed CKDNet, it was compared against several state-of-the-art models. All

Table 1
Comparison with different keypoint detection models.

Method	Backbone	Params (M)	Internal test set		External test set	
			AP_{box} (%)	AP_{kp} (%)	AP_{box} (%)	AP_{kp} (%)
HRNet	HRNet-w32	25.9 + 28.5	–	91.12	–	89.47
HRFormer	HRFormer-s	25.9 + 43.2	–	93.21	–	91.70
ViTPose	ViT-Base	25.9 + 89.9	–	93.06	–	91.15
RTMPose	CSPNeXt	25.9 + 34.05	–	92.65	–	91.75
DEKR	HRNet-w32	**31.5**	–	90.72	–	88.64
RTMO	CSPDarknet	44.8	–	93.09	–	92.04
EDPose	ResNet-50	50.6	98.56	94.17	97.81	93.09
Ours (CKDNet)	ResNet-50	43.16	**98.65**$_{\pm 0.07}$	**95.10**$_{\pm 0.13}$	**98.36**$_{\pm 0.10}$	**94.38**$_{\pm 0.17}$

models, including our CKDNet and the selected competitors, were trained on the training set and subsequently tested on both the internal test set and the independent external test set. The competing models included HRNet (Sun et al., 2019), HRFormer (Yuan et al., 2021), ViTPose (Xu et al., 2022), DEKR (Geng et al., 2021), RTMPose (Jiang et al., 2023), RTMO (Lu et al., 2024), and EDPose (Yang et al., 2023b). The comparison results were shown in Table 1. It should be noted that the top-down methods require an additional object detection stage before performing keypoint localization. Therefore, YOLOv8m (Params: 25.9 M) was first employed to detect the cattle bounding boxes, and the detection results were then used as inputs for evaluating their keypoint detection accuracy.

Performance metrics for the proposed CKDNet were reported as Mean ± SD over three independent runs with different random seeds. On the internal test set, the proposed CKDNet model achieved the highest AP_{box} and AP_{kp} of $98.65_{\pm 0.07}$% and $95.10_{\pm 0.13}$%, respectively. Furthermore, it maintained its superior performance on the external test set with an AP_{box} of $98.36_{\pm 0.10}$% and an AP_{kp} of $94.38_{\pm 0.17}$%, demonstrating stable generalization to new animals with only slight performance degradation. The high-accuracy keypoint localization establishes a robust basis for subsequent initial body measurement and posture deviation regression. In comparison with top-down methods, such as HRNet and HRFormer, the proposed CKDNet eliminated the need for an additional object detection stage, thereby simplifying the overall pipeline and reducing cumulative errors and computational overhead. Furthermore, compared with representative bottom-up and one-stage frameworks including DEKR and EDPose, the proposed CKDNet maintains superior keypoint localization accuracy on both test sets while exhibiting strong competitiveness in model compactness and parameter efficiency.

4.1.2. Ablation experiment on the CKDNet

The ablation experiments were carefully designed to emphasize the contribution of each component to the performance enhancement of the CKDNet model. A baseline model without the BFEM and HGAEM modules was established, and the influence of adding BFEM, HGAEM, and their combination on model performance was comprehensively evaluated. The results of the ablation experiment were shown in Table 2.

4.1.2.1. Impact of BFEM. A comparison between the Baseline and Variant-1 (w/o and w/ BFEM) reveals that BFEM improved AP_{box} and AP_{kp} by 0.35% and 0.74% on the internal test set, and by 0.26% and 0.99% on the external test set. This result demonstrates its effectiveness in enhancing keypoint detection performance. This conclusion could also be further verified by comparing Variant-2 and the proposed model.

Fig. 9 illustrates a visualization of BFEM's working mechanism. Column 2 and Column 5 display the feature maps of the input image before and after BFEM, respectively. Column 3 and Column 4 show the SFAM and BGAM. In these visualizations, red regions indicate areas of high attention, whereas blue regions indicate lower response. It can be observed that, SFAM exhibited high response mainly over the cow's main body, highlighting its role in focusing on semantically relevant regions. In contrast, BGAM concentrated on areas with contours or boundary changes, enhancing the model's sensitivity to edge and detail information. The combined effect of the two modules strengthens feature representation in the cow region, enabling accurate keypoint localization even in areas with weak visual cues.

4.1.2.2. Impact of HGAEM. A comparison between the Baseline and Variant-2 (w/o and w/ HGAEM) reveals that the introduction of HGAEM improved AP_{box} and AP_{kp} by 0.24% and 1.65% on the internal test set, and by 0.35% and 1.50% on the external test set. This improvement indicates that the module effectively enhances the model's perception ability and structural understanding. Although the model's parameter count slightly increased, the task primarily involves single-image analysis rather than real-time video processing. Therefore, the additional computation has limited impact on system efficiency and user experience, while the resulting accuracy improvement has practical significance. The performance gains brought by HGAEM are also reflected in the comparison between Variant-1 and the proposed model, further validating the effectiveness and necessity of the HGAEM module.

To further confirm that the observed performance improvement originates from HGAEM's ability to model high-order visual correlations, representative hyperedges generated by the module were visualized, as shown in Fig. 10. It can be observed that, different hyperedges focus on distinct semantic regions. For instance, Hyperedge 1 mainly attends to the cow's legs and head, Hyperedge 2 primarily highlights the hooves and neck, and Hyperedge 3 responds more to background information. These visualizations demonstrate that HGAEM effectively captures potential high-order semantic correlations in images, enabling the model to learn many-to-many relationships rather than only low-order pairwise correlations, thereby improving both the accuracy and robustness of keypoint detection.

4.2. Result of bodysize measurements

To evaluate the measurement accuracy of the proposed method, manually measured body sizes were used as ground truth. The results are summarized in Table 3. On the internal test set (140 samples), the proposed method achieved MAEs of 6.05 cm, 2.71 cm, and 2.95 cm for BL, BH, and HH, respectively. The corresponding MAPEs were 4.26%, 1.97%, and 2.19%, while the MSDs were 61.28 cm^2, 12.34 cm^2, and 13.94 cm^2, respectively. Furthermore, when evaluated on the independent external test set (400 samples) comprising unseen individuals, the model maintained reliable performance, yielding MAEs of 7.87 cm, 4.81 cm, and 5.25 cm for BL, BH, and HH, respectively. The corresponding MAPEs were 5.47%, 3.59%, and 3.89%, while the MSDs were 102.38 cm^2, 33.87 cm^2, and 41.74 cm^2, respectively. These results demonstrated that the proposed method achieves high measurement accuracy, with the maximum MAPE below 10%, meeting the precision requirements for automated body measurement in precision livestock farming.

Further examination of Table 3 reveals that the measurement accuracy was improved after applying the PC-GCN to the initial estimates. On the internal test set, compared with the initial body size (w/o PC-GCN), the adjusted results (w/ PC-GCN) showed decreases in MAE by 0.47 cm, 0.06 cm, and 0.10 cm for BL, BH, and HH, respectively. Similar improvements were consistently observed on the external test set, where the MAE decreased by 0.49 cm, 0.08 cm, and 0.24 cm for BL, BH, and HH, respectively. Furthermore, statistical analysis confirmed the reliability of these improvements. All paired comparisons yielded statistically significant differences ($P < 0.05$). These statistical metrics confirm

Table 2
Contribution of different improvements.

Model	BFEM	HGAEM	Params (M)	Internal test set		External test set	
				AP_{box} (%)	AP_{kp} (%)	AP_{box} (%)	AP_{kp} (%)
Baseline	×	×	**38.27**	98.07	92.89	97.72	91.84
Variant-1	√	×	42.08	98.42	93.63	97.98	92.83
Variant-2	×	√	39.55	98.31	94.54	98.07	93.34
Ours	√	√	43.16	**98.72**	**95.23**	**98.46**	**94.55**

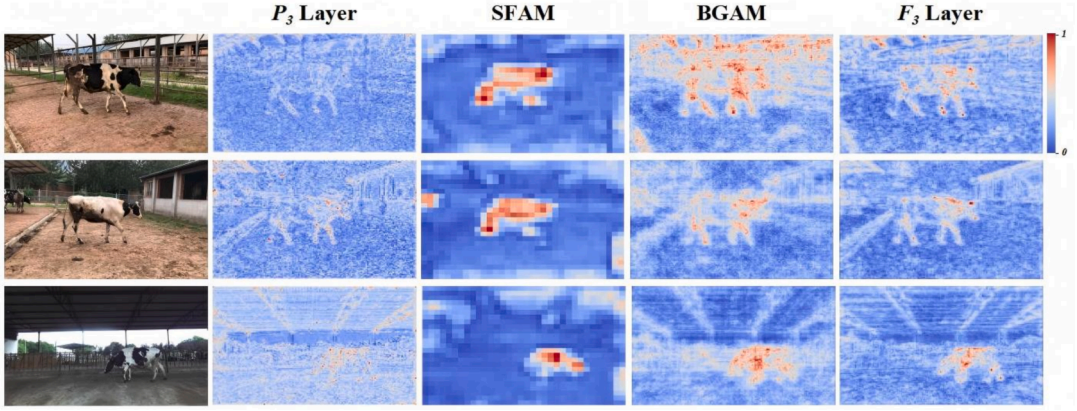


Fig. 9. Visualization of BFEM with feature maps and attention maps.

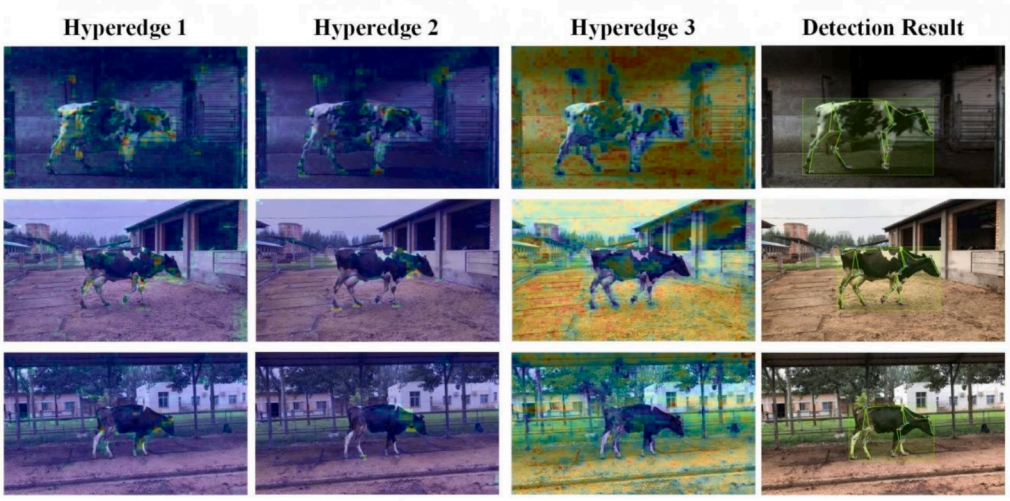


Fig. 10. Visualization of hyperedges in HGAEM.

Table 3
Measurement results of cow body size.

Dataset	Traits	Initial Body Size (w/o PC-GCN)			Adjusted body size (w/ PC-GCN)			Cohen's d	P-value
		MAE (cm)	MAPE (%)	MSD (cm^2)	MAE (cm)	MAPE (%)	MSD (cm^2)		
Internal test set	BL	6.52	4.58	66.2011	6.05	4.26	61.2781	0.0709	0.0015*
	BH	2.77	2.12	13.3292	2.71	1.97	12.3363	0.0335	0.0325*
	HH	3.05	2.27	14.6943	2.95	2.19	13.9441	0.0616	0.0102*
External test set	BL	8.36	5.80	107.40	7.87	5.47	102.38	0.0776	0.0143*
	BH	4.89	3.64	36.64	4.81	3.59	33.87	0.0234	0.0387*
	HH	5.49	4.06	44.95	5.25	3.89	41.74	0.0619	0.0461*

* indicates statistical significance at $P < 0.05$.

that the performance gains provided by the PC-GCN module were systematic rather than random. Collectively, these results indicate that PC-GCN effectively mitigates the influence of posture variation on body size estimation.

To further demonstrate the correction capability of PC-GCN, 20 image samples of the same cow were randomly selected for visual analysis of body measurement deviations. As shown in Fig. 11, the results of each measurement exhibited certain fluctuations, which can be partly attributed to posture variations. After the correction by PC-GCN, the measurement errors caused by non-standard postures were partially suppressed. The fluctuations in BL, BH, and HH were slightly reduced, improving the overall stability and consistency of the measurements.

It should be noted that errors in manual measurement and 3D imaging are inevitable, leading to a certain level of inherent noise in the ground-truth data. This inherent noise fundamentally restricts the theoretical ceiling for error reduction, which explains why the calculated effect sizes are numerically small and why the model may occasionally amplify prediction errors for individual samples. Nevertheless, the overall results demonstrate that PC-GCN plays a role in mitigating the influence of posture variation and enhancing the stability of body size estimation.

5. Prototype system implementation

To further facilitate the practical application of the proposed model, a supporting data acquisition terminal device, database system, and

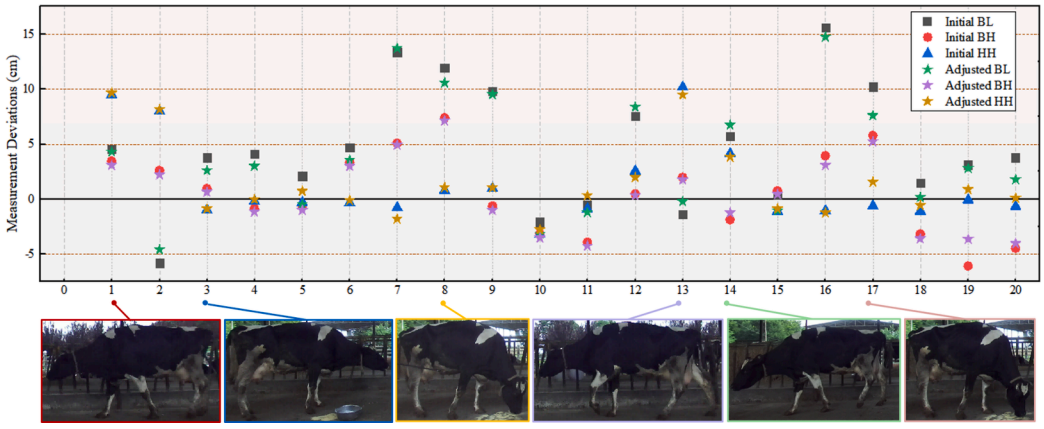


Fig. 11. Visualization of the PC-GCN performance in mitigating posture effects.

information management platform were developed. The model was deployed on a terminal device, as shown in Fig. 12 (a). The device design was inspired by previous work (Xu et al., 2024). During operation, it can be either carried by farmers (Fig. 12(a-1)) or mounted on a tripod (Fig. 12 (a-2)). The system comprises a Nvidia Jetson AGX Orin, a stereo camera, a power supply, a touchscreen display, and several auxiliary modules. Users can conveniently perform non-contact body measurements through the graphical user interface. For a single-instance test, the 2D keypoint localization (CKDNet) required 1.85 s, the initial body measurement process including depth estimation (DEFOM-Stereo) took 7.47 s, and the posture-adjusted refinement (PC-GCN) required only 0.0002 s. Including data I/O and storage, the total end-to-end processing time was controlled within 10 s per cow.

The acquired body size data are automatically transmitted to a MySQL database (as shown in Fig. 12 (c)) and visualized through a self-developed cattle information management platform (as shown in Fig. 12 (b)). The platform was built on a Browser–Server (B/S) architecture, with a Java-based backend integrating the Spring Boot and MyBatis-Plus frameworks, and a JavaScript-based frontend developed with Vue 2.0 and ECharts for dynamic interaction and visualization. The platform supports not only the query of historical body size records for individual cattle, but also statistical and analytical functions for herd-level status. Through this intuitive and comprehensive digital management approach, farmers can monitor herd dynamics in real time and make informed decisions, thereby improving the efficiency and intelligence of farm management. Further details on the system's implementation, including its workflow, current limitations, user feedback, and plans for subsequent updates, are documented and will be continuously updated at the project repository: https://github.com/XingshiXu/Cow_Inf_Platform.

6. Discussion

6.1. Discussion with similar research

In recent years, accurately localizing the keypoints used for non-contact body measurement has remained a major research focus in this field (Xu et al., 2026). Early measurements primarily relied on manually selecting keypoints within 3D point clouds (Salau et al., 2017; Le Cozler et al., 2019). Although these methods eliminate the need for large-scale annotated (Yang et al., 2023a). While intuitive and capable of providing high baseline accuracy, this approach is labor-intensive and difficult to automate. Subsequently, several studies proposed localization methods based on 3D morphological features (such as curvature analysis, clustering, or polynomial fitting) data, their stability and robustness are often difficult to guarantee. More recently, many studies have tended to utilize deep learning to acquire measurement points directly through point cloud segmentation or regression (Hou et al., 2025; Zhou et al., 2025). While these methods have been proven to achieve excellent performance, they typically impose high demands on the quality of the acquired point cloud data because they ignore the rich texture and high-level semantic information available in 2D RGB images. In contrast, similar to the works of Li et al. (2024) and Yang et al. (2025), the proposed method adopts a fusion strategy combining RGB images with depth data. This approach fully leverages the superior semantic feature extraction capabilities of deep learning models on RGB images, subsequently mapping the detected keypoints precisely into the corresponding 3D space. This method significantly improves the robustness of the system while ensuring measurement accuracy.

Beyond keypoint localization, a few notable studies have addressed body measurement deviations in cattle caused by non-standard postures during free walking. Luo et al. (2023) employed a statistical shape model to standardize livestock poses, which effectively reduced measurement inconsistencies arising from posture variations. However, the model fitting process required approximately 20 min per sample, making real-time or near real-time detection infeasible. Li et al. (2023) proposed a

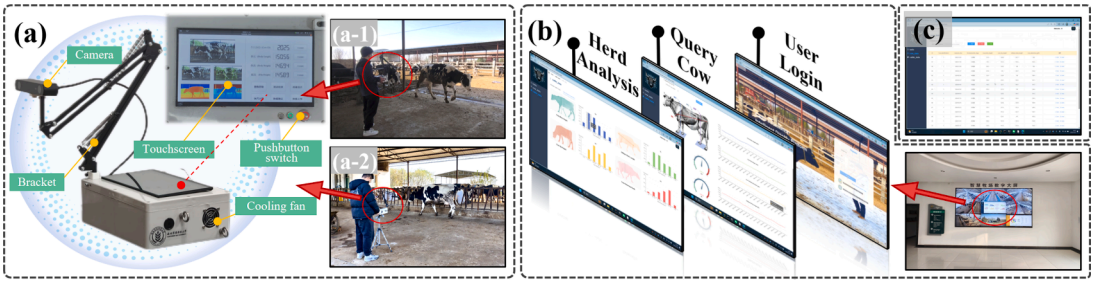


Fig. 12. Practical application. (a) Body measurement device; (b) Information management platform; (c) Information database.

posture-based measurement adjustment method that refined body size estimation by extracting twelve micro-posture features from 3D point cloud data. This approach demonstrated the feasibility of mitigating measurement deviations through posture-related features. Nevertheless, its reliance on high-precision point clouds and complex data processing procedures limits its applicability in practice. Similar to these efforts, our work also addresses the influence of posture variation on body size estimation in freely walking cattle. Unlike previous approaches, the proposed method avoids the need for multi-camera systems and hyperparameter-sensitive algorithms. The method provides accurate, efficient, and streamlined measurements, supporting the advancement of non-contact body size estimation toward automation and practical deployment.

6.2. Robustness evaluation under environmental degradation

Considering that the complex and volatile conditions of farms may interfere with the keypoint detection precision of CKDNet, the external test set was utilized to simulate three typical environmental degradations, including rainfall, fog, and low-light conditions to systematically evaluate the robustness of the model. Each degradation type was categorized into four intensity levels, ranging from light to severe. The representative samples of the degraded images are displayed in Fig. 13 (a).

In the simulation of rainfall scenarios, a multi-factor coupled model was constructed to account for both rain streak interference and atmospheric scattering effects. Rain streak noise was simulated by generating linear structures with consistent directional distributions, where the quantity and length increased nonlinearly with rainfall intensity. Simultaneously, global brightness attenuation and a light misting term were introduced to characterize the reduced illumination and air scattering typical of rainy environments. The continuous degradation process, ranging from light rain to torrential rain, was jointly controlled by the brightness attenuation coefficient $\alpha \in \{0.92, 0.84, 0.76, 0.68\}$, the rain streak count $N \in \{134, 379, 697, 1073\}$, and the fogging coefficient $\eta \in \{0.04, 0.08, 0.12, 0.16\}$. For fog scenarios, the degradation was implemented based on the Koschmieder atmospheric scattering model, which describes the scattering and attenuation of light in a medium via a transmission function and an atmospheric light term. The atmospheric scattering coefficient was set to $\beta \in \{0.5, 1.0, 1.5, 2.0\}$ to simulate varying fog intensities under different visibility conditions. Regarding low-light scenarios, a combination of Gamma transformation and global brightness attenuation was employed to simulate diverse illumination levels. Continuous modeling from slight underexposure to severe low-light environments was achieved by adjusting the parameter pairs $(\gamma, \alpha) \in \{(1.2, 0.9), (1.5, 0.7), (2.0, 0.5), (2.5, 0.4)\}$, with the inclusion of subtle noise and blurring to approximate the characteristics of low-light imaging. The implementation details of these environmental degradation algorithms are available at: https://github.com/XingshiXu/Agri-Env-Degradation-Toolkit.

The test results for CKDNet across these degraded environments were illustrated in Fig. 13 (b). The results indicated that the proposed model exhibits significant and satisfactory robustness when the degradation was at a low intensity (Disturbing degree $\leq$ 2), with the keypoint localization precision AP_{kp} remaining above 90%. However, a more pronounced decline in model precision was observed when the degradation reaches the extreme level (Disturbing degree = 4). This suggests that extreme visual interference leads to a severe loss of spatial information, causing the feature response maps to become diffused. Consequently, future research may consider incorporating image enhancement algorithms as a pre-processing module to further improve system reliability in harsh environments.

6.3. Limitations and future work

Despite the promising performance of the CowSizeNet, several limitations remain to be addressed in future research to enhance its practical utility:

(1) ***Generalizability across breeds, ages, and species*:** The proposed method is theoretically applicable to various livestock such as sheep, horses, and diverse cattle breeds. However, the current model cannot be directly deployed across different species or growth stages due to significant variations in skeletal structures and morphological proportions. Specifically, the CKDNet must learn new visual features and anatomical landmarks, while the PC-GCN needs to adapt to unique geometric constraints and mappings specific to the target breed or age group. Future research will explore transfer learning and domain adaptation to extend the framework's versatility to a broader range of livestock populations with minimal retraining costs.

(2) ***Robustness against external occlusions*:** A fundamental limitation of the proposed method is its sensitivity to external occlusions, such as the railings in milking parlors or sorting alleys. When a keypoint is physically obstructed by a railing (i.e., the target anatomical point is situated behind the railing relative to the camera), the system suffers from significant measurement errors due to the inability to accurately localize the 3D coordinates of the occluded point. This failure stems from the direct loss or corruption of depth information in the occluded regions. Future research will explore geometric completion algorithms that leverage geometric priors of the animal body to estimate the missing depth data for obstructed regions (Li et al., 2025a).

(3) ***Computational efficiency and lightweight design*:** Current efforts primarily focus on enhancing the accuracy of body size estimation rather than optimizing the model's lightweight architecture.

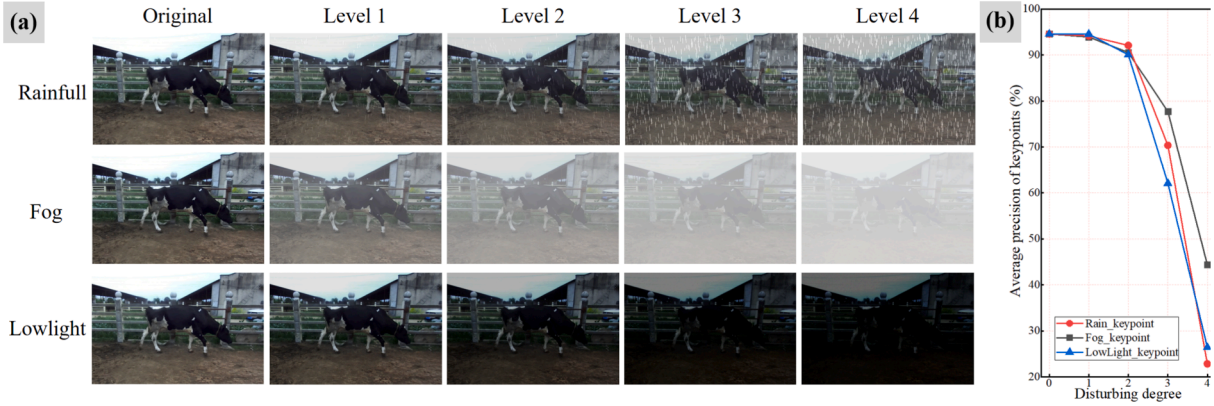


Fig. 13. Environmental degradation. (a) Simulated samples of rainfall, fog, and low-light; (b)AP_{kp} across four degradation levels.

Further optimization is necessary to enable deployment on edge devices with lower computational power. Future research will investigate model compression techniques and the design of efficient, lightweight network architectures to reduce the computational complexity without compromising precision.

(4) ***Mitigation of inherent data noise*:** Inherent errors in manual measurements and fluctuations in 3D imaging sensors introduce non-negligible noise into the training data. Such noise makes it challenging for the PC-GCN to capture highly significant corrective features during the learning process. To address this, future research will aim to incorporate high-fidelity synthetic datasets generated from 3D modeling and skeletal rigging (Okour et al., 2024; Iwami et al., 2026). These "virtual cows" provide noise-free ground truth along with greater data diversity and volume, serving as a robust pre-training foundation before the model is fine-tuned on real-world experimental data to achieve higher precision.

7. Conclusion

This study addressed two major challenges in non-contact cattle body measurement, namely keypoint localization errors caused by weak visual features and measurement deviations induced by non-standard postures, by proposing a novel method called CowSizeNet. To alleviate the first issue, the CKDNet was developed, which enhances feature representation through the BFEM and strengthens spatial associations via the HGAEM. This design effectively improved the localization accuracy of visually ambiguous regions such as the withers, back, and buttocks. To overcome the second challenge, the PC-GCN was introduced to learn the nonlinear relationship between keypoint geometry, joint angles, and body measurement deviations under different postures. By regressing posture-related deviations, PC-GCN effectively compensated for the measurement deviations caused by non-standard postures, yielding more accurate and consistent results. Extensive experiments and a system prototype implementation demonstrated the model's effectiveness, robustness, and potential value for practical PLF applications.

CRediT authorship contribution statement

Xingshi Xu: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Benhai Xiong:** Writing – review & editing, Supervision, Investigation. **Hongxing Deng:** Writing – review & editing, Software, Investigation. **Dong Liu:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Bo Jiang:** Supervision, Funding acquisition. **Tomas Norton:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition. **Huaibo Song:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

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 work was supported by the National Key R&D Program of China, China (No. 2024YFD1300600), the National Natural Science Foundation of China, China (No. 32272931, No. 32472964 and No. 62401476), the Shaanxi Province Agricultural Key Core Technology Project, China (No. 2025NYGG005), the Shaanxi Province Key R&D Program, China (No. 2024NC-ZDCYL-05-12), and the Technology Development, China (No. 2025ZY-QYCXYL-04), Northwest A&F University PhD Research Startup Fund Project, China (No. 2452024115). The authors appreciate the funding organization for their financial supports. The first author (X. Xu) acknowledges the support of the China Scholarship Council program, China (No. 202506300020) and the Doctoral Student Program of the Youth S&T Talents Cultivation Project, CAST, China.

The authors would also like to thank the helpful comments and suggestions provided by all the authors cited in this article and the anonymous editors and reviewers.

Data availability

Data will be made available on request.

References

Bai, L., Guo, C., Song, J., 2025. Cattle weight estimation model through readily photos. Eng. Appl. Artif. Intel. 143, 109976.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. European Conference on Computer Vision (ECCV), Vancouver Convention Center, Vancouver, Canada.

Chen, L., Zhang, L.Y., Tang, J.L., Tang, C., An, R., Han, R.Z., Zhang, Y.Y., 2024. GRMPose: GCN-based real-time dairy goat pose estimation Chen Chao Rui an Ruizi Han. Comput. Electron. Agric. 218, 108662.

Cominotte, A., Fernandes, A., Dorea, J., Rosa, G., Ladeira, M., van Cleef, E., Pereira, G., Baldassini, W., Neto, O., 2020. Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. Livest. Sci. 232, 103904.

Dai, Q., Ling, Q., 2025. Hybrid representation learning for end-to-end multi-person pose estimation. IEEE Trans. Circuits Syst. Video Technol. 35 (7), 6437–6451.

Deng, H., Yang, G., Xu, X., Hua, Z., Liu, J., Song, H., 2025. Fusion of CREStereo and MobileViT-Pose for rapid measurement of cattle body size. Comput. Electron. Agric. 232, 110103.

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.

Du, S., Zou, Y., Wang, Z., Li, X., Li, Y., Shang, C., Shen, Q., 2026. Unsupervised hyperspectral image super-resolution via self-supervised modality decoupling. Int. J. Comput. Vis. 134, 152.

Feng, Y., Huang, J., Du, S., Ying, S., Yong, J., Li, Y., Ding, G., Ji, R., Gao, Y., 2025. Hyper-YOLO: when visual object detection meets hypergraph computation. IEEE Trans. Pattern Anal. Mach. Intell. 47 (4), 2388–2401.

Fischler, M., Bolles, R., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.

Fixelle, J. 2025. Hypergraph vision transformers: Images are more than nodes, more than edges. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Music City Center, Nashville TN.

Geng, Z., Sun, K., Xiao, B., Zhang, Z., Wang, J., 2021. Bottom-up human pose estimation via disentangled keypoint regression. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA.

Han, B., Zhang, J., Xiang, Y., Li, X., Li, S., 2025. Shapewarp: a "global-to-local"non-rigid sheep point cloud posture rectification method. Expert Syst. Appl. 270, 126524.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA.

Hou, Z., Zhang, Q., Zhang, B., Zhang, H., Huang, L., Wang, M., 2025. CattlePartNet: an identification approach for key region of body size and its application on body measurement of beef cattle. Comput. Electron. Agric. 232, 110013.

Huang, J., Ma, Z., Wu, Y., Bao, Y., Wang, Y., Su, Z., Guo, L., 2025a. YOLOv8-DDS: a lightweight model based on pruning and distillation for early detection of root mold in barley seedling. Inform. Process. Agricult. 12 (4), 581–594.

Huang, S., Lu, Z., Cun, X., Yu, Y., Zhou, X., Shen, X., 2025b. DEIM: DETR with improved matching for fast convergence. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Music City Center, Nashville TN.

Iwami, R., Yamamoto, Y., Nakamura, K., Taniguchi, Y., 2026. Generating textured 3D dairy cow models from a few images. International Workshop on Advanced Imaging Technology (IWAIT), Kaohsiung, Taiwan.

Jiang, H., Lou, Z., Ding, L., Xu, R., Tan, M., Jiang, W., Huang, R., 2025. DEFOM-Stereo: Depth foundation model based stereo matching. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Jiang, T., Lu, P., Zhang, L., Ma, N., Han, R., Lyu, C., Li, Y., Chen, K., 2023. RTMPose: real-time multi-person pose estimation based on MMPose. arXiv:2303.07399v2.

Jing, X., Wu, T., Shen, P., Chen, Z., Jia, H., Song, H., 2025. In situ volume measurement of dairy cattle via neural radiance fields-based 3D reconstruction. Biosyst. Eng. 250, 105–116.

Le Cozler, Y., Allain, C., Caillot, A., Delouard, J.M., Delattre, L., Luginbuhl, T., Faverdin, P., 2019. High-precision scanning system for complete 3D cow body shape imaging and analysis of morphological traits. Comput. Electron. Agric. 157, 447–453.

Li, J., Ma, W., Bai, Q., Tulpan, D., Gong, M., Sun, Y., Xue, X., Zhao, C., Li, Q., 2023. A posture-based measurement adjustment method for improving the accuracy of beef cattle body size measurement based on point cloud data. Biosyst. Eng. 230, 171–190.

Li, J.W., Ma, W.H., Li, Q.F., Zhao, C.J., Tulpan, D., Yang, S.M., Ding, L.Y., Gao, R.H., Yu, L.G., Wang, Z.Q., 2022a. Multi-view real-time acquisition and 3D reconstruction of point clouds for beef cattle. Comput. Electron. Agric. 197, 106987.

Li, J.W., Ma, W.H., Zhao, C.J., Li, Q.F., Tulpan, D., Wang, Z.Q., Yang, S.X., Ding, L.Y., Gao, R.H., Yu, L.G., 2022b. Extraction of key regions of beef cattle based on bidirectional tomographic slice features from point cloud data. Comput. Electron. Agric. 199.

Li, R., Wen, Y., Zhang, S., Xu, X., Ma, B., Song, H., 2024. Automated measurement of beef cattle body size via key point detection and monocular depth estimation. Expert Syst. Appl. 244, 123042.

Li, Y., Dai, X., Dai, B., Song, P., Wang, X., Chen, X., Li, Y., Shen, W., 2025a. Cow depth image restoration method based on RGB guided network with modulation branch in the cowshed environment. Comput. Electron. Agric. 229, 109773.

Li, Z., Sun, H., Guo, J., Zhang, H., Liu, H., 2025b. Research progress on machine vision technology for non-contact body measurement of large livestock. Trans. Chinese Soc. Agricult. Eng. (Transactions of the CSAE). 41 (7), 1–12.

Lu, H., Zhang, J., Yuan, X., Lv, J., Zeng, Z., Guo, H., Ruchay, A., 2025. Automatic coarse-to-fine method for cattle body measurement based on improved GCN and 3D parametric model. Comput. Electron. Agric. 231, 110017.

Lu, P., Jiang, T., Li, Y., Li, X., Chen, K., Yang, W., (2024). RTMO: Towards high-performance one-stage real-time multi-person pose estimation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle,WA, USA.

Luo, X., Hu, Y., Gao, Z., Guo, H., Su, Y., 2023. Automated measurement of livestock body based on pose normalisation using statistical shape model. Biosyst. Eng. 227, 36–51.

Norton, T., Berckmans, D., 2018. Engineering advances in Precision Livestock Farming. Biosyst. Eng. 173, 1–3.

Okour, M., Falque, R., Alempijevic, A., 2024. Sim2real cattle joint estimation in 3D point clouds. arXiv:2410.14419v14411.

Qiao, Y., Kong, H., Clark, C., Lomax, S., Su, D., Eiffert, S., Sukkarieh, S., 2021. Intelligent perception for cattle monitoring: a review for cattle identification, body condition score evaluation, and weight estimation. Comput. Electron. Agric. 185, 106143.

Ruchay, A., Kober, V., Dorofeev, K., Kolpakov, V., Miroshnikov, S., 2020. Accurate body measurement of live cattle using three depth cameras and non-rigid 3-D shape recovery. Comput. Electron. Agric. 179, 105821.

Ruchay, A., Kolpakov, V., Kosyan, D., Rusakova, E., Dorofeev, K., Guo, H., Ferrari, G., Pezzuolo, A., 2022. Genome-Wide associative study of phenotypic parameters of the 3D body model of aberdeen angus cattle with multiple depth cameras. Animals 12 (16), 2128.

Salau, J., Haas, J., Junge, W., Thaller, G., 2017. A multi-Kinect cow scanning system: calculating linear traits from manually marked recordings of Holstein-Friesian dairy cows. Biosyst. Eng. 157, 92–98.

Sheng, K., Foris, B., von Keyserlingk, M., Timbers, T., Cabrera, V., Weary, D., 2025. Redefining lameness assessment: constructing lameness hierarchy using crowd-sourced data. Comput. Electron. Agric. 234, 110206.

Shi, D., Wei, X., Li, L., Ren, Y., Tan, W., 2022. End-to-end multi-person pose estimation with transformers. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA.

Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach,CA, USA.

Wang, H., Zhang, S., Leng, B., 2025. HGFormer: Topology-aware vision transformer with HyperGraph learning. IEEE Trans. Multimedia. 3542958.

Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Xu, X., Deng, H., Wang, Y., Zhang, S., Song, H., 2024. Boosting cattle face recognition under uncontrolled scenes by embedding enhancement and optimization. Appl. Soft Comput. 164, 111951.

Xu, X., Xiong, B., Liu, D., Norton, T., Song, H., 2026. Non-contact technologies for cattle monitoring: advances and challenges towards precision livestock farming. Comput. Electron. Agric. 248, 111785.

Xu, Y., Zhang, J., Zhang, Q., Tao, D., 2022. ViTPose: simple vision transformer baselines for human pose estimation. Advances in Neural Information Processing Systems (NeurIPS), New Orleans, Louisiana, United States of America.

Xu, Z., Li, Q., Ma, W., Li, M., Morris, D., Ren, Z., Zhao, C., 2025. A geodesic distance regression-based semantic keypoints detection method for pig point clouds and body size measurement. Comput. Electron. Agric. 234, 110285.

Yang, G., Li, R., Zhang, S., Wen, Y., Xu, X., Song, H., 2023a. Extracting cow point clouds from multi-view RGB images with an improved YOLACT plus plus instance segmentation. Expert Syst. Appl. 230, 120730.

Yang, G., Qiao, Y., Deng, H., Shi, J., Song, H., 2025. One-stage keypoint detection network for end-to-end cow body measurement. Eng. Appl. Artif. Intel. 146, 110333.

Yang, G., Xu, X., Song, L., Zhang, Q., Duan, Y., Song, H., 2022. Automated measurement of dairy cows body size via 3D point cloud data analysis. Comput. Electron. Agric. 200, 107218.

Yang, J., Zeng, A., Liu, S., Li, F., Zhang, R., Zhang, L., 2023b. Explicit box detection unifies end-to-end multi-person pose estimation. 2023 International Conference on Learning Representations (ICLR), Kigali, Rwanda.

Yin, L., Zhu, J., Liu, C., Tian, X., Zhang, S., 2022. Point cloud-based pig body size measurement featured by standard and non-standard postures. Comput. Electron. Agric. 199, 107135.

Yuan, Y., Fu, R., Huang, L., Lin, W., Zhang, C., Chen, X., Wang, J., 2021. Hrformer: high-resolution vision transformer for dense predict. Advances in Neural Information Processing Systems (NeurIPS), Virtual-only.

Zhang, A., Wu, B., Wuyun, C., Jiang, D., Xuan, E., Ma, F., 2018. Algorithm of sheep body dimension measurement and its applications based on image analysis. Comput. Electron. Agric. 153, 33–45.

Zhang, Q., Hou, Z., Huang, L., Wang, F., Meng, H., 2024. Reparation with moving least squares sampling and extraction of body sizes of beef cattle from unilateral point clouds. Comput. Electron. Agric. 224, 109208.

Zhang, Y., Zhang, Y., Ga, M., Wan, X., Dai, B., Shen, W., 2025. Multimodal behavior recognition for dairy cow digital twin construction under incomplete modalities: a modality mapping completion network approach. Artif. Intell. Agric. 15 (3), 459–469.

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J., 2024. DETRs beat YOLOs on real-time object detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle Convention Center, Seattle WA, USA.

Zhou, G., Ye, W., Li, S., Wang, Z., Li, G., Li, J., 2025. FGPointKAN++ point cloud segmentation and adaptive key cutting plane recognition for cow body size measurement. Artif. Intell. Agric. 15 (4), 783–801.