Computers and Electronics in Agriculture 246 (2026) 111606

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

A joint optimization method for weight estimation and re-identification based on a cattle back semantic disentanglement model

Yuqi Zhang [a], Longxiang Li [a], Hui Kang [a], Shiyuan Liang [a], Kai Niu [a], Yue Rong [b], Zhiqiang He [a,*]

[a] Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, China
[b] School of Electrical Engineering, Computing and Mathematical Sciences Curtin University, Bentley, WA 6102, Australia

ARTICLE INFO

Keywords:
Beef cattle
Weight estimation
Re-identification
Convolutional neural network

ABSTRACT

Reliable re-identification and accurate body weight estimation are fundamental for precision cattle management. This paper proposes a joint optimization method based on a semantic disentanglement model of dorsal cattle images, which improves body weight prediction accuracy while performing cattle dorsal re-identification. Specifically, the proposed model disentangles dorsal image representations into morphology, texture, and posture, and aligns each component with its corresponding downstream task, enabling more effective and independent feature learning. We further develop a method, identity-consistent multi-image aggregation, which is inspired by an information-theoretic approach to quantify the benefits of aggregating predictions across multiple images of the same individual. During inference, the proposed Weight-Constrained Identity Clustering (WCIC) method groups images by identity and filters out weight outliers to enhance cattle weight prediction accuracy. Extensive experimental results demonstrate that, compared with single-task baselines, the proposed method maintains re-identification accuracy while reducing the mean absolute error (MAE) of body-weight estimation by 4.89 kg, corresponding to an improvement of approximately 26.5%. These results highlight the effectiveness of semantic disentanglement and identity-consistent multi-image aggregation for robust, non-contact weight monitoring of cattle in real-world farm environments.

1. Introduction

In precision beef cattle management, continuous monitoring of individual body weight growth is a key indicator for evaluating production efficiency and economic profitability (da Costa Freitas and Barbosa, 2025; Ojo et al., 2024). Variations in genetic background, physiological traits, and feeding environments result in distinct growth patterns and fattening efficiencies among individual cattle (Liu et al., 2010; Cantalapiedra-Hijar et al., 2018). Consequently, modern beef cattle production is shifting from extensive, scale-oriented practices toward individual-level growth assessment, where body weight gain over time or feeding cycles serves as a critical metric closely associated with feed conversion ratio (FCR) and fattening efficiency (Tedeschi et al., 2004). This growth-based indicator integrates both inherent breed characteristics and management-induced feeding efficiency, providing essential guidance for precision feeding and individualized management.

Accurate estimation of individual growth trajectories requires the joint capability of cattle re-identification and body weight estimation, as growth rates are derived from repeated measurements of the same individual over time. Re-identification ensures identity consistency across monitoring cycles, while contactless body weight estimation enables continuous data acquisition without physical weighing. Their integration is fundamental for maintaining long-term body weight continuity and enabling reliable, data-driven evaluation of feed efficiency and individual growth performance (Merrell, 1931). With increasing emphasis on feed efficiency optimization and economic sustainability, integrated cattle identification and weight estimation have become a core requirement for digitalized, lifecycle-oriented cattle production systems (Qiao et al., 2021; Markov et al., 2022).

Currently, cattle weighing methods are generally categorized into chute-based and non-contact techniques. Chute-based weighing remains widely adopted in commercial production, where cattle are guided into weighing chutes and body weight readings are associated with individual identities through visual or electronic ear tags (Han et al., 2025; Lopes et al., 2018; Wangchuk et al., 2018). Despite its practicality, this approach suffers from low operational efficiency, safety risks, and stress-induced welfare concerns. In contrast, non-contact weighing has emerged as a promising research direction, leveraging sensors and computer vision to estimate body weight from visual inputs via regression models (Dohmen et al., 2021; Ruchay et al., 2022; Los et al., 2023; Liu et al., 2023; Xu et al., 2024; Weber

* Corresponding author.
E-mail address: hezq@bupt.edu.cn (Z. He).

https://doi.org/10.1016/j.compag.2026.111606
Received 3 December 2025; Received in revised form 21 February 2026; Accepted 25 February 2026
Available online 2 March 2026
0168-1699/© 2026 Published by Elsevier B.V.

et al., 2020; Gjergji et al., 2020; Zin et al., 2020; Imamura et al., 2017). However, existing non-contact methods still struggle to balance cost-effectiveness, estimation accuracy, and imaging quality (Tscharke and Banhazi, 2013), and more critically, often lack the capability to automatically associate weight predictions with individual identities, limiting their applicability in large-scale farming environments.

Similarly, cattle re-identification has evolved along two primary technical pathways. The first relies on manually assigned identity markers, such as visual or electronic ear tags (Williams et al., 2019; Feng et al., 2013). While simple to deploy, these approaches depend heavily on manual intervention and are prone to data loss and misidentification. The second pathway exploits biometric traits by capturing facial or dorsal images and encoding identity-discriminative features for autonomous identification (Shijun Li et al., 2021; Nasir et al., 2025; Yamamoto et al., 2025; Wang et al., 2025; Kumar et al., 2025). Biometric approaches are non-contact and non-invasive, offering greater automation potential and robustness for precision livestock management.

In computer vision, different tasks often focus on distinct semantic aspects of visual data and may exhibit semantic interference when jointly modeled. Semantic disentanglement has therefore emerged as an effective representation learning paradigm to decouple correlated yet functionally distinct semantic factors, and has been successfully applied in person re-identification, image generation, zero-shot learning, and domain adaptation (Gu et al., 2022; Xiangyu Li et al., 2021; Karras et al., 2019; Zhang et al., 2018). In the context of cattle re-identification and weight estimation, identity-discriminative texture cues and weight-related morphological cues are inherently correlated but serve different functional roles. Disentangling these semantics enables joint optimization of the two tasks while reducing negative transfer, thereby improving robustness and estimation accuracy in real-world farming scenarios.

Motivated by these observations, this study proposes an integrated computer vision framework that simultaneously performs cattle re-identification and body weight estimation from dorsal (top-view) images. Dorsal views provide stable representations of body morphology and surface texture, reducing occlusion and posture-induced variability. By disentangling identity-related and weight-related semantics embedded in dorsal images, the proposed framework enables effective joint modeling of the two tasks. Experimental results demonstrate that identity-consistent textural features substantially enhance weight estimation accuracy, supporting reliable long-term growth monitoring in practical production settings.

The main contributions of this paper are as follows:

- **Semantic disentanglement of dorsal cattle images.** We systematically perform task-specific feature separation on dorsal cattle images, decomposing them into three semantic components: morphological features (structural attributes such as body contour and trunk dimensions), textural features (identity-discriminative cues such as coat color and fur patterns), and state features (dynamic attributes such as posture and behavior). This semantic disentanglement clarifies the complementarities between body weight estimation and cattle re-identification, providing effective multi-scale prior knowledge for multi-task learning, thereby enhancing model generalization.
- **Multi-image same-identity joint estimation and quantitative information-theoretic analysis.** We propose a multi-image, same-identity joint estimation framework to leverage identity-consistent features for improved body weight prediction. Based on information-theoretic principles, the Kullback–Leibler (KL) divergence is employed to quantify information certainty, revealing how joint modeling of multiple images of the same identity refines the posterior distribution of weight estimates. This formulation theoretically substantiates the deterministic gains of multi-image joint estimation in reducing predictive uncertainty of cattle weight and guides the design of effective feature aggregation strategies.
- **Identity-consistent multi-image aggregation** We propose a dual-task framework that jointly trains cattle re-identification and body weight estimation while leveraging cross-task semantic disentanglement. The model architecture first employs a backbone to extract multi-scale dimensional features, which are then fed into task-specific multi-head learning modules for joint training of hierarchical morphological and textural features, thereby enhancing the full utilization of effective task-relevant features. During inference, an iterative refinement mechanism is introduced to perform cross-task outlier filtering, thereby mitigating the impact of erroneous predictions on each task. Experimental results demonstrate that the proposed framework effectively enhances feature representation, improves prediction robustness, and provides a practical solution for integrated identity recognition and weight monitoring in precision livestock farming, including in challenging zero-shot scenarios.

2. Materials and methods

2.1. Semantic disentanglement of cattle dorsal images

In computer vision, image semantics generally refer to the abstract information conveyed by an image. In this work, we define semantics as the structured, high-level information abstracted from the entire image, moving beyond pixel-level segmentation. Such a distinction between raw representation and underlying meaning is critical for robust AI; for instance, Farquhar et al. (2024) demonstrated that focusing on "semantic" equivalence rather than "syntactic" or lexical variation is essential for accurately detecting hallucinations in generative models (Farquhar et al., 2024).

Inspired by this principle of prioritizing semantic consistency over surface-level features, we show that the dorsal view of cattle can be semantically disentangled into three principal components: morphological, textural, and postural semantics (Fig. 1). This decomposition allows for a more precise representation of both biological and biometric traits.

Formally, the semantic disentanglement of a dorsal image I is expressed as:

$$I = F_{\text{morph}} \oplus F_{\text{texture}} \oplus F_{\text{pos}} + \epsilon, \tag{1}$$

where $\oplus$ denotes the **compositional combination of semantic components**, representing the joint contribution of each semantic subspace to the overall image representation. Specifically:

- F_{morph}: **Morphological semantics**, encoding body-scale structural characteristics (e.g., body length, chest circumference);
- F_{texture}: **Textural semantics**, capturing biometric identifiers (e.g., coat color patterns, dorsal markings);
- F_{pos}: **Postural semantics**, reflecting transient behavioral or pose-related information (e.g., head orientation, hoof placement);
- ϵ: **Noise term**, encompassing external disturbances (e.g., illumination changes, occlusion).

Eq. (1) presents the disentanglement of a dorsal image into three mutually complementary semantic subspaces, each encoding distinct semantic information.

Morphological semantics form the structural basis for body weight estimation. Features such as rump fullness, abdominal tightness, and neck curvature evolve with growth, reflecting both common developmental patterns and individual differences among cattle.

Textural semantics provide discriminative biometric cues for cattle re-identification. Rather than being strictly invariant, dorsal texture patterns encode individual-specific semantic structures that remain relatively stable and comparable over time, enabling effective identity discrimination even in the presence of moderate environmental noise

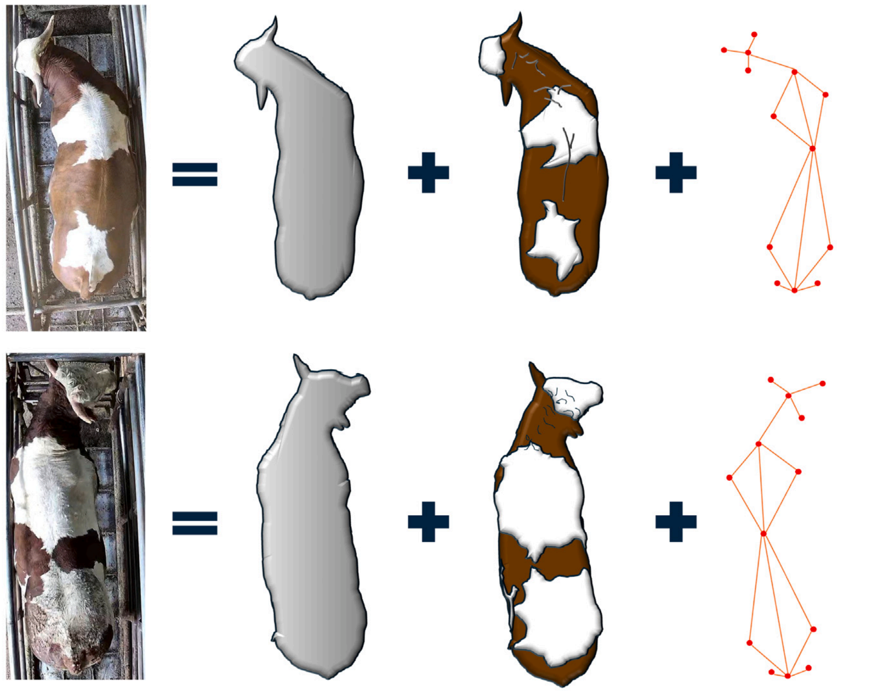

Fig. 1. Semantic disentanglement of dorsal-view cattle images into morphological, textural, and postural features.

such as dirt accumulation or illumination variations (Yuqi Zhang et al., 2025).

Postural semantics capture dynamic variations present during image acquisition, including head movements, neck rotations, hoof placement, and spinal curvature. Modeling of these variations reduces posture-induced noise, thereby improving the robustness of both weight estimation and re-identification.

2.2. Multi-image same-identity joint estimation and quantitative information-theoretic analysis

In practical cattle weight estimation, predictions derived from a single dorsal image are often biased by confounding factors such as posture variation, occlusion, and illumination. To mitigate such limitations, we introduce a semantic information theory based on the principle of synonymous mapping (Niu, 2025).

Specifically, let U denote the **set of all possible dorsal cattle images** (i.e., the syntactic space or raw pixel domain), and let $\tilde{U}$ represent the **high-level semantic attribute space** (e.g., attributes such as cattle identity and weight). The synonymous mapping $f : \tilde{U} \to 2^U$ associates each semantic symbol $\tilde{u}_{i_s}$ (e.g., a specific weight value W) with an equivalence class of syntactic representations $U_{i_s} \subseteq U$. This equivalence class comprises all image instances u that are visually distinct but correspond to the same semantic meaning (i.e., the same weight W).

For a dataset of n images belonging to the same cattle identity, $\{I_i\}_{i=1}^{n} \subseteq U$, the **true semantic probability distribution** of weight $W \in \tilde{U}$ is given by

$$P_s(W) = \sum_{u \in f^{-1}(W)} P(u), \tag{2}$$

where $f^{-1}(W)$ denotes the set of all syntactic instances mapped to W, and $P(u)$ represents a **prior probability distribution** over the syntactic space U, i.e., the probability of capturing a particular image u in the real world. This distribution serves as the theoretical foundation for our analysis, though it is typically unavailable in practice.

For a single image I_i, the **induced semantic distribution** $Q_s^{(i)}(W)$ represents the probability of each possible semantic value (e.g., body weight) inferred from that image, capturing the uncertainty due to variations in posture, illumination, or occlusion. Aggregating multiple images of the same cattle identity produces an **averaged semantic distribution**:

$$Q_s^{(\mathrm{avg})}(W) = \frac{1}{n} \sum_{i=1}^{n} Q_s^{(i)}(W). \tag{3}$$

Following the convexity property of Kullback–Leibler divergence (Niu, 2025). Formally, for any two candidate distributions $Q_s^{(1)}$ and $Q_s^{(2)}$ and $\lambda \in [0, 1]$,

$$D_{KL}\left(P_s \,\middle\|\, \lambda Q_s^{(1)} + (1-\lambda) Q_s^{(2)}\right) \le \lambda D_{KL}(P_s \parallel Q_s^{(1)}) + (1-\lambda) D_{KL}(P_s \parallel Q_s^{(2)}). \tag{4}$$

Applying this property to the multi-image case yields

$$D_{KL}\left(P_s \,\middle\|\, Q_s^{(\mathrm{avg})}\right) \le \frac{1}{n} \sum_{i=1}^{n} D_{KL}(P_s \parallel Q_s^{(i)}). \tag{5}$$

Thus, the divergence associated with the averaged estimator is always no worse than the arithmetic mean of the divergences of individual estimators. Moreover, whenever the $Q_s^{(i)}$ are non-degenerate (i.e., exhibit variance), strict inequality holds. In other words, for any single image I_i, one typically has

$$D_{KL}(P_s \parallel Q_s^{(i)}) > D_{KL}(P_s \parallel Q_s^{(\mathrm{avg})}). \tag{6}$$

From an information-theoretic perspective, this result shows that multi-image fusion provably reduces semantic uncertainty in cattle weight estimation. Equivalently, using the notation $H(P \parallel Q)$ to denote the **cross-entropy** between distributions P and Q (defined as $H(P, Q) = -\sum_x P(x) \log Q(x)$), we have:

$$\boxed{H(P_s, Q_s^{(\mathrm{avg})}) < H(P_s, Q_s^{(i)})} \tag{7}$$

The reduction in semantic uncertainty (i.e., the increase in estimation certainty) is evaluated by aggregating multiple images of the same cattle identity.

While Eq. (7) demonstrates the reduction of uncertainty in the semantic probability space, implementing full distributional fusion is computationally intensive. We therefore adopt a practical approximation: assuming the output distribution of the weight regression network approximates a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, the fusion of semantic distributions $Q_s^{(avg)}$ can be effectively approximated by the arithmetic mean of the individual weight predictions. This approximation is justified because the average of Gaussian distributions remains Gaussian, with its mean equal to the average of the individual means. Therefore, in our practical implementation, we utilize the averaged weight value to realize this theoretical gain in certainty while maintaining computational efficiency.

2.3. Identity-consistent multi-image aggregation

2.3.1. Data preparation

To construct a dataset of multiple dorsal images of cattle along with their corresponding body weights, we designed the experimental setup shown in Fig. 2. The dataset includes a diverse range of cattle breeds, including Simentaler (Simmental) cattle, Angus cattle, various crossbreeds, and indigenous local breeds, ensuring a broad representation of morphological and textural variations. The weighing chute was modified by installing two cameras at different heights, both connected to a central host via Bluetooth. The electronic scale was interfaced with the host computer through an RS232 serial connection, and a custom software system was developed to monitor the real-time scale waveform and synchronize image capture with weight measurements. This system is only used to provide the ground truth of model training. Once the model is trained, there is no need to use the chute.

When a cow passed through the chute, the system activated a locking mechanism on the scale waveform and applied a sliding-window filter to track the stability of the weight readings. Once the weight signal reached a relatively stable state, the cameras were triggered to capture images at 2-second intervals, with each camera storing up to six frames.

The recorded weights were calibrated against manual readings to ensure measurement accuracy. Based on this procedure, we constructed a dataset of dorsal-view cattle images paired with ground-truth weights. The final dataset comprises 15,161 images in the training set and 1494 images in the validation set, with a weight range from 500 kg to 930 kg, an average weight of 714.78 kg, and a standard deviation of 81.43 kg. Importantly, the training and validation sets were split such that cattle identities do not overlap, ensuring that models are evaluated on entirely unseen individuals. Fig. 2 illustrates the experimental setup for data acquisition, while Fig. 3 presents the statistical distribution of cattle weights in the training and validation subsets.

Before training the model on the collected cattle images, a series of preprocessing steps were applied to enhance robustness against spatial and illumination variations (Fig. 4).

Specifically, to address viewpoint misalignment across different cameras, peripheral regions introduced during alignment were first cropped and padded with black pixels while maintaining the original canvas size. This cropping–resizing strategy effectively reduced artifacts caused by cross-camera inconsistencies. Subsequently, all processed images were uniformly scaled to a resolution of 128 × 256, which preserves the statistical aspect ratio of dorsal cattle views. These steps encourage the model to focus on the central semantic regions of the cattle while reducing reliance on background information, thereby improving spatial feature robustness.

From a spatial perspective, different image transformations affect the two tasks in distinct ways due to their underlying semantic dependencies. We observed that image flipping has a significant impact on re-identification tasks, which rely heavily on texture semantics.

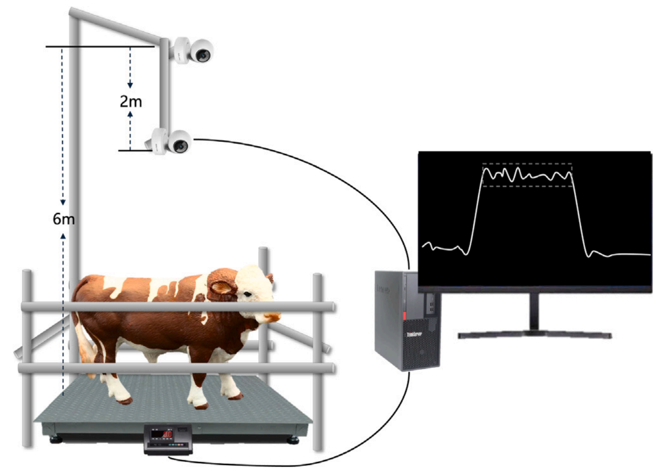

Fig. 2. Experimental setup for dorsal image and weight data acquisition.

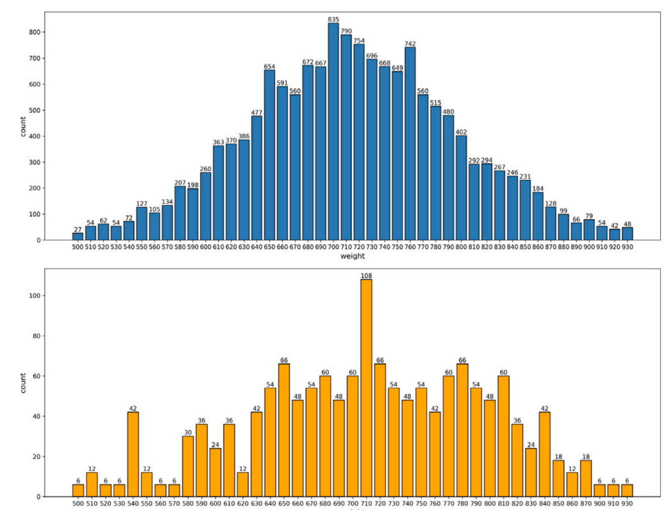

Fig. 3. Weight distribution of the collected dataset. Blue: training set (15,161 samples); Yellow: validation set (1494 samples).

Re-identification depends on specific, location-dependent visual cues such as coat color patterns and dorsal markings, whose relative spatial configurations are crucial for consistent individual matching. Flipping alters the orientation of these key identifiers, potentially disrupting the model's discriminative capability.

In contrast, body weight estimation is primarily driven by morphological semantics, focusing on structural attributes such as body contour, projected area, and dorsal shape fluctuations. Consequently, image scaling — either enlargement or shrinkage — can substantially influence the representation of these morphological features, leading to variations in weight estimation accuracy. Overall, re-identification is more sensitive to flipping due to its dependence on fine-grained texture features, whereas body weight estimation is more affected by scaling because of the proportional and structural nature of the morphological cues involved.

2.3.2. Semantic-disentanglement based dual-task training framework

Building on the proposed theory of semantic disentanglement in cattle dorsal imagery, our method aims to achieve simultaneous yet disentangled modeling of identity and body weight. Specifically, the dorsal view is leveraged to separate texture semantics, which serve as discriminative features for re-identification, from morphological semantics, which capture body-scale characteristics crucial for weight estimation.

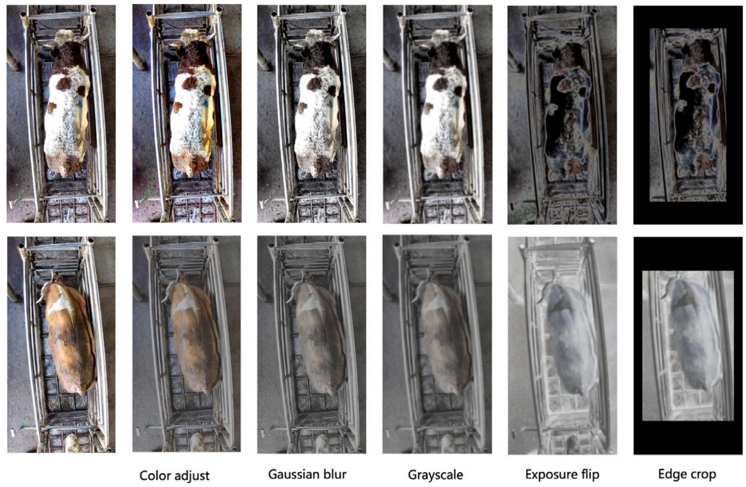

Fig. 4. Overview of preprocessing transformations for dorsal cattle images.

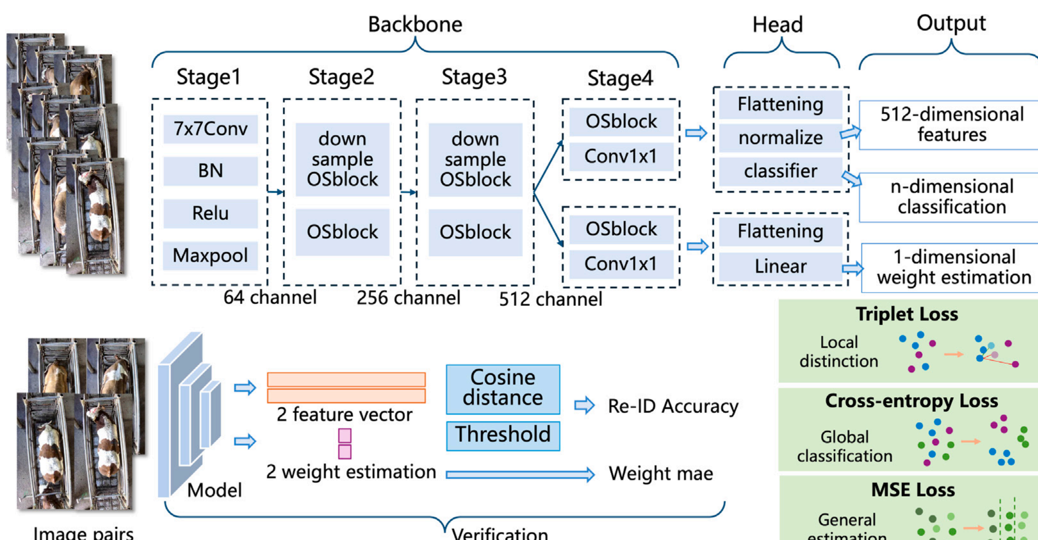

Fig. 5. Overview of the proposed network.

Due to the acquisition setting, where the cattle torso remains nearly parallel to the imaging plane, this viewpoint minimizes the impact of behavioral variations on semantic disentanglement, enabling the neural network to independently learn identity-related features from texture and weight-related features from morphology.

We propose a semantic disentanglement based dual-task training framework that effectively integrates cattle re-identification and body weight estimation. The network consists of four main components: a multi-scale shared feature extraction module, a texture semantic disentanglement module, a morphological semantic disentanglement module, and the loss functions (Fig. 5). The backbone is adapted from an enhanced OSNet architecture (Zhou et al., 2019), which is well-known for its omni-scale feature learning capability, enabling effective capture of both fine-grained texture semantics and large-scale morphological information. OSNet's dynamic feature aggregation mechanism supports multi-scale perception and fusion, improving spatial representation and making it highly suitable for semantic disentanglement tasks.

In our design, the multi-scale shared feature extraction module retains the original OSBlock modules. For further semantic disentanglement, two parallel OSBlock-based branches are introduced to handle different feature semantics. The texture semantic branch explicitly supports the re-identification task, while the morphological semantic branch explicitly supports body weight estimation. In the output heads, texture features are processed through fully connected layers for re-identification, whereas morphological features are first reduced in dimensionality before being fed into a regression head for weight prediction.

To jointly optimize the two tasks of cattle re-identification and body weight estimation, we adopt a hybrid loss function that integrates both classification, metric learning, and regression objectives. The overall optimization objective is formulated as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_t \mathcal{L}_{\text{Triplet}} + \lambda_w \mathcal{L}_{\text{MSE}}, \tag{8}$$

where $\mathcal{L}_{\text{CE}}$ denotes the cross-entropy loss for global re-identification, $\mathcal{L}_{\text{Triplet}}$ is the triplet loss used to enhance intra-batch embedding discrimination for re-identification, and $\mathcal{L}_{\text{MSE}}$ represents the mean squared error loss for body weight regression. The weighting coefficients λ_t and λ_w balance the contributions of metric learning and regression objectives during joint training.

Although the two disentanglement modules share similar convolutional designs, their distinct loss function combinations serve two purposes: on one hand, the task-specific supervision signals effectively guide the semantic separation. The Re-ID loss enforces the texture branch to focus on identity-invariant patterns (coat color), ignoring structural changes; simultaneously, the Weight loss constrains the morphological branch to prioritize shape indicators correlated with weight, disregarding pixel-level identity textures. on the other hand, through backpropagation, the two branches collaboratively enhance the effectiveness and accuracy of multi-scale shared feature extraction, thereby improving the performance of both disentanglement tasks. This mutual benefit has been validated in our experiments.

2.3.3. Cross-task outlier filtering for inference

During inference, we adopt a cross-task outlier filtering strategy to enhance robustness. Initially, the model generates both identity embeddings and weight estimates for all images. The identity embeddings are clustered to group samples of the same individual. Within each identity cluster, weight predictions are further refined by clustering to filter outliers caused by misidentification or large estimation errors. Clusters exhibiting excessive weight variation are discarded, and the remaining clusters yield the final inferred weight by averaging the predictions within each cluster. This twice iterative refinement ensures consistency across both identity and weight dimensions, suppressing error propagation and significantly improving prediction accuracy.

Since the per-identity inference Mean Absolute Error (MAE) on the validation set is observed to fall within the range of 20–25 kg, and many samples are affected by environmental factors such as camera viewpoint variations and animal posture, we adopt a conservative outlier threshold of $\tau = 15$ kg in the proposed WCIC scheme. This threshold represents a practical trade-off between measurement uncertainty and biological variability: dynamic weight fluctuations caused by respiration and subtle posture changes are typically within 5–10 kg, while 15 kg corresponds to approximately 2.1% of the average body weight in our dataset (714.78 kg), which is consistent with the normal intra-day physiological variation (approximately 2%–3%) in beef cattle. Although a tolerance of up to 25 kg is commonly acceptable in production-oriented growth monitoring (Liu et al., 2023), we deliberately adopt a stricter threshold to suppress outliers and improve the robustness of identity clustering and weight inference.

As described in Algorithm 1, the inference proceeds in two stages: identity clustering and weight-based outlier filtering.

Algorithm 1 Cross-Task outlier filtering inference

Require: Image set $\mathcal{I}$, trained model M, outlier threshold τ
Ensure: Cluster mapping C, inferred weight per cluster $\hat{W}_c$

```
Model inference:
for each i ∈ I do
    (f_i, ŵ_i) ← M(i)          ▷ Where f_i: identity feature, ŵ_i: weight estimate
end for
Normalize {f_i}
ε* ← optimal DBSCAN threshold
C ← DBSCAN({f_i}, ε*)          ▷ DBSCAN: a density-based clustering algorithm
Split large clusters by K-Means on ŵ_i
for each cluster c in C do
    if max(ŵ_i) − min(ŵ_i) > τ kg then
        Exclude c
    else
        Ŵ_c ← mean(ŵ_i | i ∈ c)
    end if
end for
return C, {Ŵ_c}
```

2.4. Evaluation indicators

For validation, re-identification is assessed using positive–negative image pairs. Two randomly sampled images are encoded into identity embeddings, and their cosine distance is computed. Standard re-identification metrics, such as **mAP**, **Rank-1**, and **Rank-3**, are employed for quantitative evaluation.

For body weight estimation, predicted values are compared against ground-truth measurements. The primary evaluation metric is the MAE:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right|, \tag{9}$$

where $\hat{y}_i$ and y_i denote the predicted and ground-truth weights of the ith sample, respectively, and N is the total number of samples. Additionally, the Root Mean Squared Error (RMSE) is used to quantify the sensitivity to larger deviations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2}. \tag{10}$$

To leverage the benefits provided by multi-image same-ID joint estimation and quantitative information-theoretic analysis, we define the Multi-image Mean Absolute Error (**MMAE**) and Multi-image Root Mean Squared Error (**MRMSE**). For an individual j with M_j images, we first compute the average predicted weight across all images of that individual:

$$\bar{\hat{y}}_j = \frac{1}{M_j}\sum_{k=1}^{M_j}\hat{y}_{jk}, \tag{11}$$

where $\hat{y}_{jk}$ represents the predicted weight of the kth image of individual j. Since all images of the same individual share the same ground-truth weight y_j, the absolute error and squared error for individual j are then computed as:

$$\mathrm{AE}_j = \left|\bar{\hat{y}}_j - y_j\right|, \tag{12}$$

$$\mathrm{SE}_j = \left(\bar{\hat{y}}_j - y_j\right)^2. \tag{13}$$

The overall MMAE and MRMSE are obtained by averaging these errors across all individuals in the dataset:

$$\mathrm{MMAE} = \frac{1}{N_{\mathrm{ID}}}\sum_{j=1}^{N_{\mathrm{ID}}}\mathrm{AE}_j, \quad \mathrm{MRMSE} = \sqrt{\frac{1}{N_{\mathrm{ID}}}\sum_{j=1}^{N_{\mathrm{ID}}}\mathrm{SE}_j}, \tag{14}$$

where N_{ID} denotes the total number of individuals. These metrics provide a robust evaluation of weight estimation performance when multiple views per individual are available, capturing the accuracy of the fused estimate derived from multiple observations.

3. Results

3.1. Multi-task joint training

In this study, we propose a lightweight dual-task semantic disentanglement framework built upon classical backbone architectures, including OSNet, ResNeXt (Xie et al., 2017), and EfficientNet (Koonce, 2021). Notably, both ResNeXt and EfficientNet have been demonstrated to achieve advanced performance in cattle body weight estimation tasks, showing strong regression accuracy and robustness in practical livestock scenarios (Haoyu Zhang et al., 2025; Gjergji et al., 2020). Each model was modified by incorporating a parallel grouping module in the final stage, enabling two decoupled branches for simultaneous cattle re-identification and weight estimation. This design allows shared feature extraction in earlier layers while maintaining task-specific representations in the output stage.

As presented in Tables 1 and 2, all dual-task models outperform their single-task counterparts in both re-identification accuracy and weight estimation precision. These results validate the effectiveness of the proposed multi-task joint optimization strategy, which leverages shared semantic information between identity and body-weight information.

Among all evaluated architectures, **OSNet** achieved the best overall performance in both tasks, benefiting from its strong capacity to

Table 1
Comparison of model performance under different task configurations based on re-identification metrics.

Model	ReID-task	Weight-task	mAP ↑	Rank-1 ↑	Rank-3 ↑
PCB (Sun et al., 2018)	✓	–	0.904	0.972	0.992
MGN (Wang et al., 2018)	✓	–	0.958	0.984	0.993
CACE-Net (Yu et al., 2022)	✓	–	0.951	0.979	0.996
EfficientNet	✓	–	0.799 ± 0.026	0.921 ± 0.017	0.965 ± 0.006
ResNeXt	✓	–	0.962 ± 0.004	0.989 ± 0.005	0.997 ± 0.003
TransReID	✓	–	0.962 ± 0.004	0.989 ± 0.005	0.997 ± 0.003
OSNet	✓	–	0.970 ± 0.003	0.992 ± 0.002	**1.000 ± 0.001**
EfficientNet	✓	✓	0.887 ± 0.012	0.954 ± 0.008	0.982 ± 0.006
ResNeXt	✓	✓	0.972 ± 0.006	**0.995 ± 0.003**	0.998 ± 0.002
OSNet	✓	✓	**0.972 ± 0.003**	**0.995 ± 0.003**	0.999 ± 0.001

Table 2
Comparison of model performance under different task configurations based on Weight estimation metrics.

Model	ReID-task	Weight-task	MAE (kg) ↓	MSE ↓	RMSE (kg) ↓
SE-ResNeXt (Haoyu Zhang et al., 2025)	–	✓	25.217	1070.198	32.673
EfficientNet (Gjergji et al., 2020)	–	✓	25.217 ± 0.984	1070.198 ± 105.896	32.673 ± 1.625
ResNeXt	–	✓	28.329 ± 1.904	1265.540 ± 147.634	35.574 ± 2.184
OSNet	–	✓	23.340 ± 1.618	895.465 ± 130.250	29.856 ± 2.009
EfficientNet	✓	✓	23.914 ± 0.921	1042.610 ± 66.409	30.309 ± 1.092
ResNeXt	✓	✓	25.786 ± 1.221	1065.143 ± 102.105	32.600 ± 1.525
OSNet	✓	✓	**21.936 ± 0.414**	**779.840 ± 30.823**	**27.920 ± 0.550**

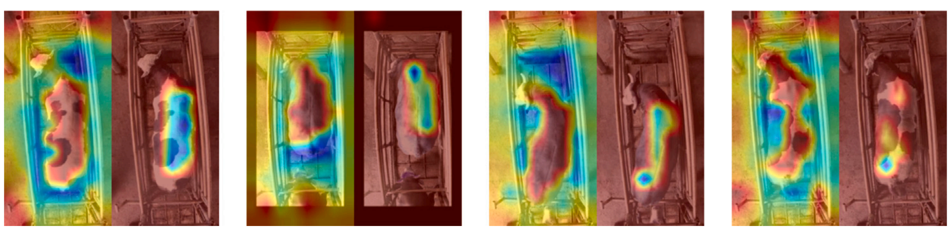

Fig. 6. Grad-CAM visualizations of the Re-ID branch and the weight estimation branch. For each pair, the left image corresponds to the Re-ID branch, while the right image corresponds to the weight estimation branch. It can be observed that the Re-ID branch is more sensitive to distinctive texture patterns and coat color variations, whereas the weight estimation branch focuses more on spatial structure and overall body shape.

capture global context and fine-grained local structures. After disentanglement modification, the dual-task OSNet achieved the highest re-identification accuracy (**mAP** = 0.972, **Rank-1** = 0.995, **Rank-3** = 0.999) and the lowest weight estimation errors (**MAE** = 21.94 kg, **MSE** = 779.84, **RMSE** = 27.92 kg). Compared with its single-task counterparts, the dual-task configuration significantly improved both ReID precision and weight prediction stability. Furthermore, the effectiveness of the proposed semantic disentanglement is qualitatively supported by Grad-CAM visualizations (Fig. 6), which show that the Re-ID branch primarily attends to fine-grained texture and coat pattern cues, while the weight estimation branch focuses on global spatial structure and body morphology.

These results indicate that the re-identification and weight-estimation tasks are semantically correlated, and that incorporating a semantic-decoupling-based approach into their joint learning enables more effective mutual regularization. Across all tested backbones, multi-task training almost consistently outperformed single-task training, underscoring the general advantage of semantically decoupled joint feature learning for integrated cattle perception systems.

3.2. Results of multi-image same-ID joint estimation

Table 3 summarizes the results of multi-image weight estimation using the proposed multi-task models based on identity-consistent multi-image aggregation. In this experiment, the models were evaluated on sets of 1 to 6 images per individual, with errors aggregated across all images of the same ID. Specifically, the **MMAE** and **MRMSE** are computed by first evaluating the MAE and RMSE for each image of an individual, and then averaging these errors across all images belonging to the same ID, as defined in Eq. (14). These metrics reflect the overall estimation performance for each individual when multiple observations are available.

As the number of images per individual increases from one to six, both **MMAE** and **MRMSE** steadily decrease, with the most notable improvement observed when moving from one to two images. This trend demonstrates the deterministic information gain obtained from aggregating multiple views of the same individual. Among all tested architectures, **OSNet** consistently achieved the best results, reaching an **MMAE** of **18.45 kg** and an **MRMSE** of **23.55 kg** when using six images, indicating that multi-image fusion substantially improves both accuracy and robustness.

Overall, these results demonstrate that the proposed dual-task semantic disentanglement framework not only enhances single-image weight estimation but also significantly improves estimation accuracy and stability under multi-image conditions. The performance gains observed in **MMAE** and **MRMSE** highlight the benefits of leveraging multiple views of the same individual, suggesting that identity-consistent multi-image aggregation is an effective strategy for robust cattle weight monitoring in practical applications. From the downward trend, it can be observed that using four images of the same individual within the same time period yields relatively accurate joint weight estimation, while the performance gain begins to plateau as the number of images continues to increase.

Table 3
Performance comparison of different algorithms under varying numbers of images per individual. MMAE and MRMSE are the average MAE and RMSE across all images of the same ID.

Algorithm	1-image		2-image		3-image		4-image		5-image		6-image	
	MAE	RMSE	MMAE	MRMSE	MMAE	MRMSE	MMAE	MRMSE	MMAE	MRMSE	MMAE	MRMSE
OSNet	**21.06**	**27.01**	**19.91**	**25.28**	**19.21**	**24.43**	**18.81**	**24.00**	**18.58**	**23.73**	**18.45**	**23.55**
ResNeXt	23.62	30.15	21.53	27.33	20.74	26.32	20.33	25.80	20.08	25.49	19.94	25.27
EfficientNet	23.99	30.57	22.25	28.18	21.61	27.34	21.29	26.91	21.11	26.65	21.01	26.48

Table 4
Cluster analysis summary of cross-task outlier filtering inference results.

Metric	Value
Total Samples	1488
Original Classes	248
Clustered Classes	253
Excluded Clusters (>15 kg)	42
Valid Clusters	211
Correct Clusters	211
Correct Cluster Ratio	1.00
MMAE	18.80 kg
Valid Clusters Sample MAE	21.65 kg
MAE	21.91 kg

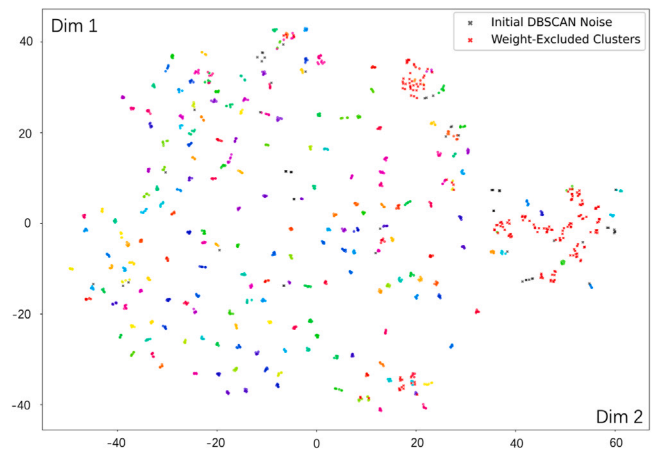

Fig. 7. DBSCAN clustering visualization. The embeddings are projected into a two-dimensional space using t-SNE. Black crosses denote feature-level outliers identified during identity clustering, typically caused by image blur or extreme body posture. Red crosses indicate clusters excluded by the weight-consistency constraint, highlighting cases where visually similar samples can be effectively distinguished by weight estimation.

3.3. Inference results

This subsection presents the inference results of the proposed cross-task outlier filtering framework on the validation set, highlighting its ability to improve clustering reliability and weight estimation accuracy under unseen identities. Since the validation identities and weight distributions are completely disjoint from those used during training, the results provide a rigorous evaluation of the framework's generalization capability. During inference, the re-identification branch first groups images into identity-consistent clusters based on feature similarity, enabling multiple observations of the same individual to be aggregated. The weight estimation branch then provides complementary feedback by identifying clusters with abnormal intra-cluster weight variation, allowing unreliable clusters to be excluded and improving the consistency of the remaining predictions.

Table 4 summarizes the inference statistics obtained from 1488 validation images, corresponding to 248 cattle identities with six images per individual. The DBSCAN-based clustering process produced 253 clusters in total. Among them, 42 clusters were identified as abnormal due to excessive intra-cluster weight variation exceeding 15 kg and were excluded. After filtering, 211 valid clusters remained, all of which were correctly grouped, yielding a perfect clustering performance (Correct Cluster Ratio = 1.0000).

To provide an intuitive understanding of the clustering and filtering behavior, Fig. 7 visualizes the similarity embeddings using t-SNE projection. Each point represents an identity embedding extracted from the re-identification branch. Black crosses denote feature-level outliers directly identified by DBSCAN, typically caused by degraded visual quality such as motion blur, partial occlusion, or extreme body posture. Red crosses indicate clusters that initially appear consistent in the embedding space but are subsequently excluded due to excessive weight inconsistency. This visualization demonstrates that feature-based clustering alone may be insufficient to guarantee semantic consistency, and that weight estimation provides an effective complementary signal for identifying ambiguous or unreliable clusters.

Fig. 8 presents a qualitative analysis of the filtered feature-level outliers. The representative samples exhibit degraded visual characteristics, including blur, occlusion, and significant appearance distortion, which lead to unreliable identity embeddings. In contrast, samples from valid clusters maintain consistent visual structure and produce stable predictions in both re-identification and weight estimation tasks. This comparison confirms that the proposed filtering mechanism effectively removes visually unreliable samples while preserving high-quality data.

Further analysis is provided in Fig. 9, which illustrates sample pairs extracted from clusters excluded by the weight-consistency constraint. Although these samples exhibit high similarity in the identity embedding space, their predicted weights differ substantially, revealing underlying semantic inconsistency. This observation highlights the limitation of relying solely on visual similarity and demonstrates the importance of weight estimation as an auxiliary task for improving clustering reliability, particularly in scenarios where individuals share highly similar visual appearance.

After applying both identity-based clustering and weight-based filtering, the mean cluster-level MAE (MMAE) is reduced to 18.80 kg, which is significantly lower than both the sample-level MAE within valid clusters (Valid Clusters Sample MAE = 21.65 kg) and the overall MAE across all samples (MAE = 21.91 kg). This result demonstrates that aggregating predictions at the cluster level after cross-task filtering improves estimation accuracy and stability by reducing the influence of noisy or inconsistent samples.

Although individual identities are not explicitly available during inference, the proposed dual-task reasoning framework effectively exploits the mutual constraints between re-identification and weight estimation. By leveraging cross-task consistency, the framework improves clustering robustness and enhances prediction reliability. These results demonstrate the effectiveness and practical applicability of the proposed framework for real-world cattle monitoring scenarios.

It is worth noting that the current implementation of the semantic disentanglement framework is primarily based on CNN architectures, which provide strong local feature representations and stable optimization behavior. However, recent advances in Transformer-based models suggest that their superior capability for modeling long-range dependencies may further enhance semantic disentanglement and multi-task

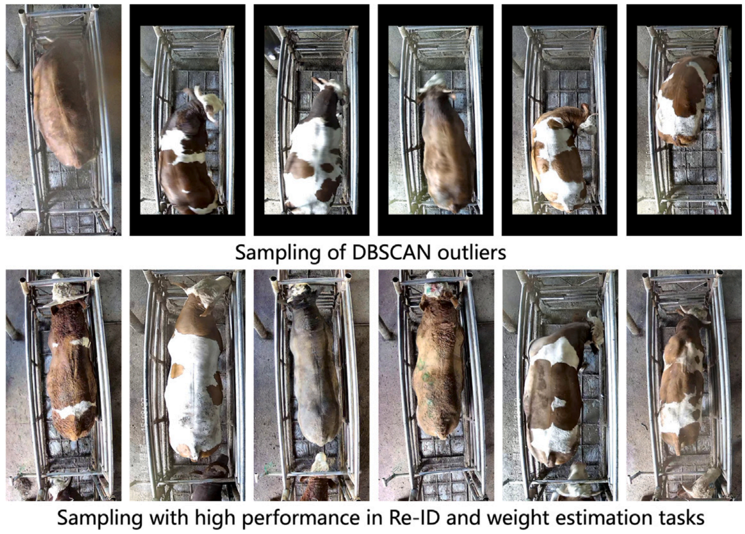

Fig. 8. Qualitative analysis of feature-level outliers filtered by DBSCAN. The first row shows representative samples corresponding to the black-cross outliers in Fig. 7, which commonly suffer from image blur or severe posture deformation. The second row presents typical samples from valid clusters that achieve reliable performance in both re-identification and weight estimation tasks.

Fig. 9. Visualization of sample pairs extracted from clusters excluded by the weight-consistency constraint. Although these samples exhibit high similarity in the identity embedding space, their weight estimation results provide strong discriminative signals, demonstrating the complementary role of weight information in visually ambiguous Re-ID scenarios.

reasoning. Extending the proposed framework to Transformer-based architectures therefore represents a promising direction for future research.

Furthermore, although the current implementation adopts a conservative strategy that discards clusters with inconsistent weight estimates, practical deployment scenarios may alternatively incorporate intra-cluster decision mechanisms to selectively retain reliable samples, enabling more flexible and adaptive handling of uncertain observations.

4. Conclusion

In this paper, we present a method for cattle weight estimation and re-identification based on a dorsal-image semantic disentanglement model. By disentangling morphological, textural, and postural features, the proposed network extracts task-specific information for accurate weight estimation and reliable re-identification. The integration of multi-image fusion and information-theoretic guidance enables robust estimation under intra-animal variations and reduces prediction error. During inference, a cross-task error-filtering clustering mechanism effectively consolidates identity information while filtering abnormal weight predictions, further enhancing accuracy. Experimental results on a large-scale cattle dataset demonstrate that the proposed approach outperforms single-task methods in both re-identification and weight estimation metrics. Overall, this study provides a practical and scalable solution for non-contact monitoring of individual cattle, contributing to precision livestock management and data-driven decision-making in beef production. As part of future work, data collection will be extended to fully open environments, such as watering and feeding zones, to further improve the robustness and real-world applicability of the proposed framework.

CRediT authorship contribution statement

Yuqi Zhang: Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Longxiang Li:** Writing – review & editing, Validation, Formal analysis. **Hui Kang:** Validation, Software, Conceptualization. **Shiyuan Liang:** Writing – review & editing, Software, Resources. **Kai Niu:** Supervision, Resources, Methodology, Conceptualization. **Yue Rong:** Writing – review & editing, Resources.

Zhiqiang He: Writing – review & editing, Visualization, Resources, Conceptualization, Project administration.

Declaration of competing interest

The authors declare that there are no conflicts of interest related to the content of this manuscript.

Acknowledgments

This paper was supported by the BUPT innovation and entrepreneurship support program (project funding number: 2025-YC-S004).

Data availability

The authors do not have permission to share data.

References

Cantalapiedra-Hijar, Gonzalo, Abo-Ismail, M., Carstens, G.E., Guan, L.L., Hegarty, R., Kenny, David A., McGee, Mark, Plastow, G., Relling, A., Ortigues-Marty, Isabelle, 2018. Biological determinants of between-animal variation in feed efficiency of growing beef cattle. Animal 12 (s2), s321–s335. http://dx.doi.org/10.1017/s1751731118001489.

da Costa Freitas, Joana, Barbosa, Maria Joana Baptista, 2025. Effective strategies in modern livestock management to increase productivity and profitability. J. Acad. Sci. 2 (5), 1488–1495. http://dx.doi.org/10.37676/jmea.v4i2.893.

Dohmen, Roel, Catal, Cagatay, Liu, Qingzhi, 2021. Image-based body mass prediction of heifers using deep neural networks. Biosyst. Eng. 204, 283–293. http://dx.doi.org/10.1016/j.biosystemseng.2021.02.001.

Farquhar, Sebastian, Kossen, Jannik, Kuhn, Lorenz, Gal, Yarin, 2024. Detecting hallucinations in large language models using semantic entropy. Nature 630 (8017), 625–630. http://dx.doi.org/10.1038/s41586-024-07421-0.

Feng, Jianying, Fu, Zetian, Wang, Zaiqiong, Xu, Mark, Zhang, Xiaoshuan, 2013. Development and evaluation on a RFID-based traceability system for cattle/beef quality safety in China. Food Control 31 (2), 314–325. http://dx.doi.org/10.1016/j.foodcont.2012.10.016.

Gjergji, Mikel, de Moraes Weber, Vanessa, Silva, Luiz Otávio Campos, da Costa Gomes, Rodrigo, De Araújo, Thiago Luís Alves Campos, Pistori, Hemerson, Alvarez, Marco, 2020. Deep learning techniques for beef cattle body weight prediction. In: 2020 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8. http://dx.doi.org/10.1109/ijcnn48605.2020.9207624.

Gu, Xinqian, Chang, Hong, Ma, Bingpeng, Bai, Shutao, Shan, Shiguang, Chen, Xilin, 2022. Clothes-changing person re-identification with rgb modality only. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1060–1069. http://dx.doi.org/10.1109/cvpr52688.2022.00113.

Han, Ying, Si, Yongsheng, He, Zhijiang, Li, Qian, Li, Zhiruo, Zhang, Man, Liu, Gang, 2025. Dynamic weighing system for dairy cows based on arrayed weighing platforms. Comput. Electron. Agric. 230, 109943. http://dx.doi.org/10.2139/ssrn.4789229.

Imamura, Sosuke, Zin, Thi Thi, Kobayashi, Ikuo, Horii, Yoichiro, 2017. Automatic evaluation of Cow's body-condition-score using 3D camera. In: 2017 IEEE 6th Global Conference on Consumer Electronics. GCCE, IEEE, pp. 1–2. http://dx.doi.org/10.1109/gcce.2017.8229435.

Karras, Tero, Laine, Samuli, Aila, Timo, 2019. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4401–4410. http://dx.doi.org/10.1109/cvpr.2019.00453.

Koonce, Brett, 2021. EfficientNet. In: Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization. Springer, pp. 109–123. http://dx.doi.org/10.7717/peerjcs.2968/fig-3.

Kumar, Niraj, Sharma, Anshul, Kumar, Abhinav, Singh, Rishav, Singh, Sanjay Kumar, 2025. Cattle verification with YOLO and cross-attention encoder-based pairwise triplet loss. Comput. Electron. Agric. 234, 110223. http://dx.doi.org/10.1016/j.compag.2025.110223.

Li, Shijun, Fu, Lili, Sun, Yu, Mu, Ye, Chen, Lin, Li, Ji, Gong, He, 2021. Individual dairy cow identification based on lightweight convolutional neural network. Plos One 16 (11), e0260510. http://dx.doi.org/10.1371/journal.pone.0260510.

Li, Xiangyu, Xu, Zhe, Wei, Kun, Deng, Cheng, 2021. Generalized zero-shot learning via disentangled representation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (3), pp. 1966–1974. http://dx.doi.org/10.1609/aaai.v35i3.16292.

Liu, George E., Hou, Yali, Zhu, Bin, Cardone, Maria Francesca, Jiang, Lu, Cellamare, Angelo, Mitra, Apratim, Alexander, Leeson J., Coutinho, Luiz L., Dell'Aquila, Maria Elena, et al., 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20 (5), 693–703. http://dx.doi.org/10.1101/gr.105403.110.

Liu, He, Reibman, Amy R., Boerman, Jacquelyn P., 2023. Feature extraction using multi-view video analytics for dairy cattle body weight estimation. Smart Agric. Technol. 6, 100359. http://dx.doi.org/10.1016/j.atech.2023.100359.

Lopes, Marcos Aurélio, Vieira, Juliana Aparecida, de Lima, Francisco Helton Sá, Demeu, Fabiana Alves, Bruhn, Fábio Raphael Pascoti, Pereira, Adriano, Vicente, Fabio Henrique, Casas, Pedro Salermo, 2018. Technical and economic efficiency of bovine weighing methods. Semin.: Ciências Agrar. 39 (3), 1167–1180. http://dx.doi.org/10.5433/1679-0359.2018v39n3p1167.

Los, Stan, Mücher, Caspar Alexander, Kramer, Henk, Franke, Gerrit Jappe, Kamphuis, Claudia, 2023. Estimating body dimensions and weight of cattle on pasture with 3D models from UAV imagery. Smart Agric. Technol. 4, 100167. http://dx.doi.org/10.2139/ssrn.4253378.

Markov, Nikolay, Stoycheva, Svetoslava, Hristov, Miroslav, Mondeshka, Lora, et al., 2022. Digital management of technological processes in cattle farms: a review. J. Central Eur. Agric. 23 (3), 486–495. http://dx.doi.org/10.5513/jcea01/23.3.3543.

Merrell, Margaret, 1931. The relationship of individual growth to average growth. Hum. Biol. 3 (1), 37. http://dx.doi.org/10.2307/41447065.

Nasir, Muhammad Fahad, Fuentes, Alvaro, Han, Shujie, Liu, Jiaqi, Jeong, Yongchae, Yoon, Sook, Park, Dong Sun, 2025. Multi-camera fusion and bird-eye view location mapping for deep learning-based cattle behavior monitoring. Artif. Intell. Agric. http://dx.doi.org/10.1016/j.aiia.2025.06.001.

Niu, Kai, 2025. Semantic Information Theory. Longmen Press, Beijing, http://dx.doi.org/10.3390/books978-3-7258-5808-8.

Ojo, Ayooluwa O., Mulim, Henrique A., Campos, Gabriel S., Junqueira, Vinícius Silva, Lemenager, Ronald P., Schoonmaker, Jon Patrick, Oliveira, Hinayah Rojas, 2024. Exploring feed efficiency in beef cattle: From data collection to genetic and nutritional modeling. Animals 14 (24), 3633. http://dx.doi.org/10.20944/preprints202411.1926.v1.

Qiao, Yongliang, Kong, He, Clark, Cameron, Lomax, Sabrina, Su, Daobilige, Eiffert, Stuart, Sukkarieh, Salah, 2021. Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation. Comput. Electron. Agric. 185, 106143. http://dx.doi.org/10.1016/j.compag.2021.106143.

Ruchay, Alexey, Kober, Vitaly, Dorofeev, Konstantin, Kolpakov, Vladimir, Dzhulamanov, Kinispay, Kalschikov, Vsevolod, Guo, Hao, 2022. Comparative analysis of machine learning algorithms for predicting live weight of Hereford cows. Comput. Electron. Agric. 195, 106837. http://dx.doi.org/10.1016/j.compag.2022.106837.

Sun, Yifan, Zheng, Liang, Yang, Yi, Tian, Qi, Wang, Shengjin, 2018. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 480–496. http://dx.doi.org/10.1007/978-3-030-01225-0_30.

Tedeschi, Luis Orlindo, Fox, Danny G., Guiroy, Pablo J., 2004. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. Agricult. Sys. 79 (2), 171–204. http://dx.doi.org/10.1016/s0308-521x(03)00070-2.

Tscharke, M., Banhazi, T.M., 2013. Review of methods to determine weight and size of livestock from images. Aust. J. Multi-Disciplinary Eng. 10 (1), 1–17. http://dx.doi.org/10.7158/n12-ae12.2013.10.1.

Wang, Yaowu, Mücher, Sander, Wang, Kaiwen, Wang, Wensheng, Kooistra, Lammert, 2025. Deep metric learning for individual cattle identification using coat patterns: Proposal for a best practice. Comput. Electron. Agric. 238 (11075), 4. http://dx.doi.org/10.1016/j.compag.2025.110754.

Wang, Guanshuo, Yuan, Yufeng, Chen, Xiong, Li, Jiwei, Zhou, Xi, 2018. Learning discriminative features with multiple granularities for person re-identification. In: Proceedings of the 26th ACM International Conference on Multimedia. pp. 274–282. http://dx.doi.org/10.1145/3240508.3240552.

Wangchuk, Kesang, Wangdi, Jigme, Mindu, Mindu, 2018. Comparison and reliability of techniques to estimate live cattle body weight. J. Appl. Anim. Res. 46 (1), 349–352. http://dx.doi.org/10.1080/09712119.2017.1302876.

Weber, Vanessa Aparecida Moraes, de Lima Weber, Fabricio, da Silva Oliveira, Adair, Astolfi, Gilberto, Menezes, Geazy Vilharva, de Andrade Porto, João Vitor, Rezende, Fábio Prestes Cesar, de Moraes, Pedro Henrique, Matsubara, Edson Takashi, Mateus, Rodrigo Gonçalves, et al., 2020. Cattle weight estimation using active contour models and regression trees bagging. Comput. Electron. Agric. 179, 105804. http://dx.doi.org/10.1016/j.compag.2020.105804.

Williams, Lauren R., Fox, David R., Bishop-Hurley, Greg J., Swain, Dave L., 2019. Use of radio frequency identification (RFID) technology to record grazing beef cattle water point use. Comput. Electron. Agric. 156, 193–202. http://dx.doi.org/10.1016/j.compag.2018.11.025.

Xie, Saining, Girshick, Ross, Dollár, Piotr, Tu, Zhuowen, He, Kaiming, 2017. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1492–1500. http://dx.doi.org/10.1109/cvpr.2017.634.

Xu, Beibei, Mao, Yifan, Wang, Wensheng, Chen, Guipeng, 2024. Intelligent weight prediction of cows based on semantic segmentation and back propagation neural network. Front. Artif. Intell. 7, 1299169. http://dx.doi.org/10.3389/frai.2024.1299169.

Yamamoto, Yota, Akizawa, Kazuhiro, Aou, Shunpei, Taniguchi, Yukinobu, 2025. Entire-barn dairy cow tracking framework for multi-camera systems. Comput. Electron. Agric. 229, 109668. http://dx.doi.org/10.1016/j.compag.2024.109668.

Yu, Fufu, Jiang, Xinyang, Gong, Yifei, Zhao, Shizhen, Guo, Xiaowei, Zheng, Wei-Shi, Zheng, Feng, Sun, Xing, 2022. Devil's in the details: Aligning visual clues for conditional embedding in person re-identification. IEEE Trans. Image Process. 31, 6188–6199. http://dx.doi.org/10.1109/TIP.2022.3204481.

Zhang, Yuqi, Li, Longxiang, Li, Chunyang, Wang, Sen, Rong, Yue, Niu, Kai, He, Zhiqiang, 2025. BECA: A computer vision dataset for long-term recognition in beef cattle. Sci. Data 13, 7. http://dx.doi.org/10.1038/s41597-025-06326-5.

Zhang, Quanshi, Wu, Ying Nian, Zhu, Song-Chun, 2018. Interpretable convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8827–8836. http://dx.doi.org/10.1109/CVPR.2018.00920.

Zhang, Haoyu, Zhang, Yuqi, Niu, Kai, He, Zhiqiang, 2025. Neural network-based method for contactless estimation of carcass weight from live beef images. Comput. Electron. Agric. 229, 109830. http://dx.doi.org/10.1016/j.compag.2024.109830.

Zhou, Kaiyang, Yang, Yongxin, Cavallaro, Andrea, Xiang, Tao, 2019. Omni-scale feature learning for person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3702–3712. http://dx.doi.org/10.1109/iccv.2019.00380.

Zin, Thi Thi, Seint, Pann Thinzar, Tin, Pyke, Horii, Yoichiro, Kobayashi, Ikuo, 2020. Body condition score estimation based on regression analysis using a 3D camera. Sensors 20 (13), 3705. http://dx.doi.org/10.3390/s20133705.