



Cattle weight estimation using 2D side-view images and estimated depth-based 3D modeling

Guilherme Botazzo Rozendo *, Maichol Dadi , Annalisa Franco , Alessandra Lumini

Department of Computer Science and Engineering, University of Bologna, via dell'Università 50, 47522, Cesena, FC, Italy

ARTICLE INFO

Keywords:
Cattle weight estimation
Segmentation
Depth estimation
Point clouds

ABSTRACT

Weighing cattle is a vital practice in livestock farming, as it provides essential data for effective herd management. Recent advancements in computer vision and machine learning have led to the development of non-invasive techniques that estimate cattle weight using images. These methods offer a way to gauge weight without needing physical scales, which helps reduce stress on the animals and minimizes labor-intensive processes. However, existing techniques often rely on dorsal (top-down) views of cattle, which can be difficult to capture in practice. In this study, we propose a method for estimating cattle weight using only side-view images, which are more accessible and easier to obtain. We utilized public datasets to extract a comprehensive set of features, including body measurements and shape descriptors from the images. We also employed advanced techniques such as cattle pose estimation, segmentation, monocular depth estimation, and point cloud generation to derive volume and area features. Our goal was to extract as much relevant information as possible from the images to accurately predict the cattle's weight. We used both linear and non-linear regression models to forecast weight based on the extracted features. Our results indicate that the proposed method can accurately predict cattle weight from side-view images, providing valuable insights for livestock management and monitoring.

1. Introduction

Weighing cattle is a crucial practice in livestock farming because it provides essential data for effective herd management. These data are important for monitoring animal health, optimizing feed strategies, and tracking growth rates to ensure productivity [1–4]. However, manual weighing can be challenging due to the size and strength of the animals, as well as the potential for stress or injury during handling [5]. Traditional methods of weighing cattle can be labor-intensive, time-consuming, and prone to inaccuracies, particularly when managing large herds [6].

Recent advances in computer vision and machine learning have led to the development of non-invasive techniques for estimating cattle weight from images. These methods allow for weight estimation without the need for physical scales, which reduces stress on the animals and minimizes labor-intensive processes. They are especially beneficial for large-scale operations, where traditional weighing methods can be impractical or time-consuming. The methods described in the literature usually rely on dorsal [5,6] or multiple views [7,8] of cattle to estimate their weight. Dorsal images can be captured using cameras mounted on chutes, allowing researchers to extract body measurements from these

images or use them as input for machine learning models to predict cattle weight.

For instance, Gjergji et al. [5] collected dorsal area images using a fixed camera positioned in a cattle chute. The images were captured as keyframes from videos taken while the cattle were moving toward the water trough. These images were then input into various models, including convolutional neural networks, recurrent convolutional neural networks, and recurrent attention models. Weber et al. [7] conducted a study to analyze the relationship between cattle weight and body measurements extracted from dorsal and lateral images. They extracted several body measurements from the images, and they found that hip width and body length had the strongest positive correlation with cattle weight. They developed an equation incorporating these measurements using stepwise regression. Lee et al. [6] also used dorsal images in a framework where authors segmented the cattle using a Mask R-CNN model and extracted features such as area, body length, width, grid lengths, perimeter, minimum bounding rectangle, convex hull area, solidity, circumscribed circle, eccentricity, and aspect ratio from the masks. They selected nine features with correlation coefficients greater than 0.5 to predict cattle weight using a ResNet model. Xu et al. [8] used

* Corresponding author.

E-mail address: guilherme.botazzo@unibo.it (G. Botazzo Rozendo).

top-view and back-view perspectives of cattle to predict their weight using neural network. To analyze the images, they first segmented them using an encoder-decoder network, employing ResNet-101-D as the encoder and the atrous spatial pyramid pooling module as the decoder. They extracted five key indicators to predict the cattle's weight: the areas of both the top view and back view, the height of the shooting distance from the top view, the shooting distance from the back view, and the age of the animal.

However, accessing the dorsal view can be challenging in practice, as it often requires specialized equipment and infrastructure. Additionally, there is a scarcity of annotated datasets of cattle that can be used for weight prediction. These factors limit the scalability and applicability of existing methods for cattle weight estimation.

Therefore, in this study we tackle the challenging task of estimating cattle weight using only side-view images, as this method is more accessible and easier to obtain. We utilized 100% public datasets available on the Kaggle platform [9] to extract and compile an extensive set of features, including body measurements and shape descriptors from the images. We also employed advanced techniques, such as cattle pose estimation, segmentation, depth estimation, and point cloud generation, to extract volume and area features from the images. Our goal was to extract as much information as possible from the images to accurately predict the cattle's weight. To achieve this, we conducted a statistical analysis to identify the most relevant features and employed linear and non-linear regression models to predict cattle weight using the extracted features. The contributions of this work are:

1. A method that extracts a comprehensive set of features related to body measurements, shape, and geometry of cattle from side-view images;
2. A strategy to generate 3D point clouds from the 2D images to extract volume features;
3. An investigation of linear and non-linear regression for cattle weight prediction using the extracted features, and;
4. An integration of techniques to predict cattle weight from side-view images, offering valuable insights for livestock management and monitoring.

2. Methodology

The proposed methodology consists of several steps, as illustrated in Fig. 1. First, we processed the public B3 and B4 datasets through pre-processing steps. Specifically, we estimated the cattle poses using the MMpose framework [10] to extract key points from the animals' bodies. We also segmented the animals in the images and applied segmentation masks to isolate them from the background. These processed images were fed into the Depth Pro model [11] to estimate the animals' depth. Based on this estimated depth, we generated point clouds for each animal.

From the key points, segmentation masks, and point clouds, we extracted 84 features to predict the weights of the animals. Using the key points, we calculated the distances between these points (KPD) and obtained various body measurements (BM). From the segmentation masks and point clouds, we derived several additional features: basic contour features (BC), Hu moments (HM) [12], statistics of distances from contour points to the centroid (SDCPC), convex hull features (CH), beam angle statistics (BAS) [13], and chain code statistics (CCS).

We conducted a correlation analysis using normalized mutual information (NMI) [14] to eliminate possible features not correlated with the weight, selecting only those with mutual information greater than zero. To predict cattle weight, we performed both linear and non-linear regression tests. We applied feature selection strategies, including step-wise regression and random forest techniques, to enhance the model's accuracy. Additionally, we evaluated the effectiveness of artificial neural networks in predicting cattle weight. The performance of the models was assessed using two metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

2.1. Preprocessing

This section describes the process of preprocessing images, which includes segmentation, keypoint estimation, depth estimation, and point cloud generation. These steps are crucial for extracting accurate features used to predict cattle weight.

2.1.1. Segmentation

To address environmental variations commonly found in farm conditions—such as changes in lighting and background clutter—we employed a robust segmentation strategy before depth estimation and point cloud generation. Specifically, we used the CAFE-Net model [15], a cross-attention and feature exploration network that incorporates a PVT encoder and a cross-attention decoder. The encoder extracts spatial and channel features, while the decoder leverages self-attention mechanisms to model long-range dependencies. Additionally, the network includes feature supplement and exploration modules composed of convolutional layers to gather local context information and bridge the semantic gap between the encoder and decoder stages.

The model was trained on the CattleSegment dataset [16], which includes a wide range of real farm environments, using the Adam optimizer with a learning rate of 0.0001 and a batch size of 16. The training process spanned 100 epochs, and structural loss [15] was used as the objective function.

The segmentation masks generated by CAFE-Net were used to isolate the animal's region in each image, reducing the influence of background and lighting variations in both the depth map estimation and point cloud generation. A detailed exploration of these environmental variations and their impact on segmentation performance can be found in [15].

2.1.2. Key points estimation

We used MMPose [10] to perform keypoint-based pose estimation on cattle images. MMPose is an open-source framework that provides a unified platform for two-dimensional and three-dimensional pose estimation applicable to humans and non-human subjects.

For our analysis, we employed the top-down model of MMPOSE with ResNet50 as the backbone. This model features a ResNet architecture with additional deconvolution layers following the last convolutional layer [17]. It was pre-trained on the AnimalPose [18] dataset, which includes annotated datasets and optimized models that account for the morphological diversity among various animal species. This model enabled us to extract the key points of the animals in the images, which were then used to calculate body measurements and distances between these key points.

2.1.3. Depth estimation and point cloud generation

We employed the Depth Pro model [11] to estimate depth information from the side-view RGB images of cattle. Depth Pro is a zero-shot monocular depth estimation method that leverages vision transformer encoders to generate high-resolution depth maps from a single RGB image. One of its main advantages is its strong generalization across domains, which enables it to produce depth maps with sharp boundaries and fine-grained details, such as hair and fur contours.

In addition to depth estimation, Depth Pro includes a focal length estimation head, which predicts the camera's horizontal angular field of view. This module takes frozen features from the depth network and task-specific features from a ViT-based encoder to infer the focal length required for geometric transformations.

We used the pre-trained Depth Pro to estimate both the depth map and the focal length for each image. To convert the resulting depth map into a 3D point cloud, we applied a standard pinhole camera projection model based on the intrinsic matrix K :

$$K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

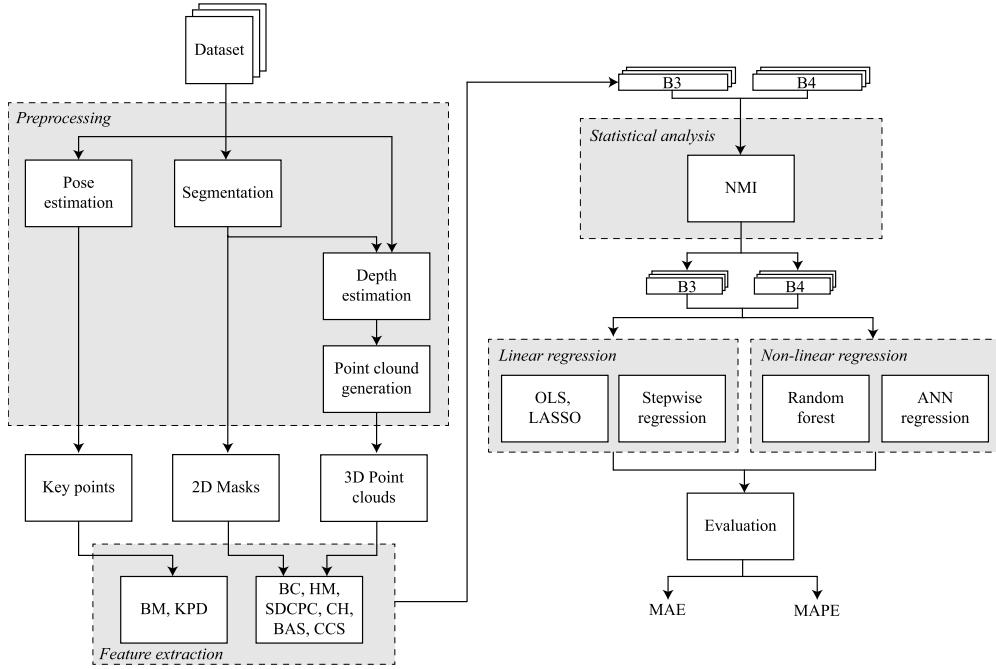


Fig. 1. Overview of the proposed methodology.

where f is the focal length estimated by Depth Pro, and (c_x, c_y) are the coordinates of the optical center, assumed to lie at the center of the image. This assumption is common when working with monocular depth models and known image dimensions in controlled acquisition conditions.

The intrinsic matrix encodes the internal parameters of the camera and is essential for projecting 2D image coordinates and their associated depth values into 3D space. Using this matrix, we converted each pixel in the depth map into a 3D point via:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \times d, \quad (2)$$

where (u, v) are the pixel coordinates in the image and d is the estimated depth at that location. This transformation results in a point cloud that spatially represents the surface geometry of the cattle.

2.2. Feature extraction

We extracted 84 features to predict cattle weight, categorizing these features into eight groups: body measurements (BM), basic contour features (BC), Hu moments (HM), statistics of the distance of contour points from the centroid (SDCPC), convex hull features (CH), beam angle statistics (BAS), chain code statistics (CCS), and MMpose distances.

The selection of these feature categories was based on their potential to reflect the geometric, structural, and volumetric properties associated with cattle body mass. BM and BC features capture direct dimensions of the animal, such as height and length. Hu moments, which are invariant to rotation and scale, summarize the overall shape of the silhouette, relating to the animal's body form and size. SDCPC features characterize the spatial distribution and symmetry of the body by analyzing the dispersion of contour points around the centroid. CH and BAS provide insights into body spread and angular complexity, both of which can vary with body condition. CCS describe the frequency and orientation of boundary transitions, which can indicate local contour irregularities. Finally, MMpose distances encode skeletal proportions that can assist in assessing body development.

The following sections provide a detailed description of each feature category.

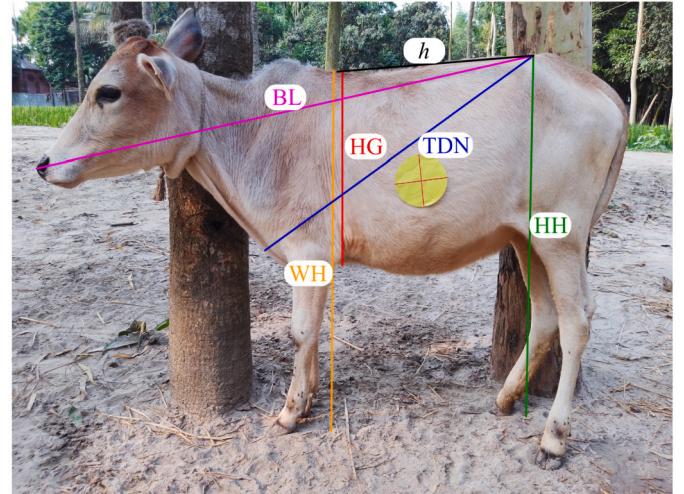


Fig. 2. Schematic illustration of BM extracted from the MMpose key points.

2.2.1. Body measurements (BM)

We have used the key points extracted by the MMpose framework [10] to determine the measurements of the cattle's body from the images. The following measurements were extracted from the key points: heart girth (HG), withers height (WH), hip height (HH), tail distance to the neck (TDN), and body length (BL). The measurements were calculated by taking the Euclidean distance in pixels between the key points. Fig. 2 shows a visual representation of each measurement extracted.

We also estimated the body volume (V) of the cattle by approximating the body as a cylinder. We calculated V using the formula:

$$V = \pi \left(\frac{HG}{2} \right)^2 h, \quad (3)$$

where h was the distance between the tail base and withers.

2.2.2. Basic contour features (BC)

We extracted the BC to provide foundational information about the size and spatial extent of the cattle side shape. From the segmentation

Table 1
Description of each HM.

Hu Moment	Description
Hu Moment 1	Represents the overall variance of the shape.
Hu Moment 2	Measures the difference in the spread along the principal axes, indicating elongation.
Hu Moment 3	Captures the skewness of the shape, indicating asymmetry.
Hu Moment 4	Measures the roundness of the shape, capturing higher-order asymmetry.
Hu Moment 5	Sensitive to the orientation of the shape, capturing rotational characteristics.
Hu Moment 6	Measures the spread of the shape along the diagonal axes, capturing diagonal asymmetry.
Hu Moment 7	Captures the overall complexity of the shape, indicating irregularities.

masks, we calculated the width and height of the bounding box enclosing the cattle contour and its perimeter.

2.2.3. Hu moments (HM)

The HM [12] is a set of seven statistical measures used to describe the shape of an object in an image, regardless of its scale, rotation, or translation. These moments are derived from central moments of the object's pixel intensity distribution, capturing unique geometric characteristics of the shape. Table 1 provides a description of each HM.

2.2.4. Statistics of the distance of contour points from the centroid (SDCPC)

The SDCPC features [13] provide information about the distribution of contour points in relation to the centroid. In the segmentation masks, we calculated the statistics: mean, standard deviation, 25th and 75th percentiles, minimum, maximum, and median distances from the contour points to the centroid. These features offer insights into the shape's symmetry and the distribution of contour points. For the point clouds, we computed the same statistics for the distances between the points and the centroid. Together, these features enhance our understanding of the spatial distribution of points within the point cloud.

2.2.5. Convex hull (CH)

The CH features [13] describe the solidity and convexity of a shape. From the segmentation masks, we extracted three SC features: area, perimeter, and solidity. These features offer insights into the shape's compactness and the ratio of the contour area to the area of the convex hull.

Additionally, we calculated the convex hull for the point clouds. The convex hull is the smallest convex polygon that encloses all the points in the point cloud. We also extracted the area, perimeter, and solidity of the convex hull to provide further information about the shape's compactness and convexity.

2.2.6. Beam angle statistics (BAS)

The BAS features [13] refer to the angles formed between contour points and the centroid of a shape. These features include the mean, standard deviation, 25th and 75th percentiles, minimum, maximum, and median angles. They allow for the analysis of the directional, rotational, and asymmetrical characteristics of shapes. Additionally, we calculated the same statistics for the angles between the points in the point cloud and the centroid.

2.2.7. Chain code statistics (CCS)

The CCS allowed us to quantify local transitions along the contour by calculating the frequency of steps in each of the eight possible directions. The features we analyzed included the mean, standard deviation, lower and upper quartiles, minimum, maximum, and median values. These statistics offer insights into the complexity of the shape and the distribution of transitions along the contour. Additionally, we calculated the same statistical measures for the chain codes of the points in the point cloud.

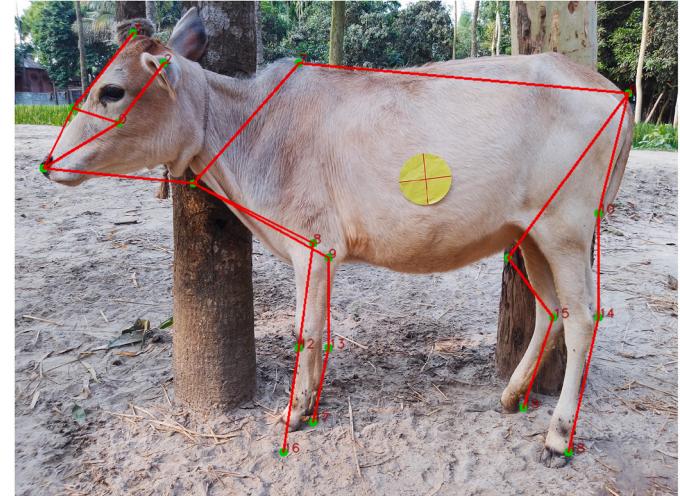


Fig. 3. Illustration of the distances extracted from the key points.

2.2.8. Key points distances (KPD)

Lastly, we calculated the distances between the key points links extracted by the MMpose framework. These distances provided supplementary information about the spatial relationships between the key points of the cattle's body. We extracted 20 distances between the key points, such as the distance between the eyes, the distance between the elbows and knees, and the distance between the tail and withers. Fig. 3 illustrates all the distances extracted from the key points.

2.3. Statistical analysis

We employed Normalized Mutual Information (NMI) [14] to evaluate the relationship between the features extracted and cattle weight in our regression analysis and eliminate uncorrelated features. NMI is a key measure from information theory [19] that quantifies the amount of information shared between two random variables. One of the main advantages of using NMI is its ability to capture both linear and nonlinear dependencies, making it especially suitable for complex data distributions. NMI is defined as:

$$\text{NMI}(X, Y) = \frac{I(X; Y)}{H(X) + H(Y)}, \quad (4)$$

where $I(X; Y)$ represents the mutual information between the input variable X and the target variable Y , while $H(X)$ and $H(Y)$ denote their respective entropies.

2.4. Regression

We used linear and non-linear regression models to predict cattle weight based on the extracted features. The regression models were trained on the features with NMI greater than zero. The following sections describe the regression models used in our experiments.

2.4.1. Linear regression

To predict the cattle weight, we perform linear regression tests. The linear regression [20] is a fundamental statistical and machine learning method used to model the relationship between a dependent variable Y and one or more independent variables X . The goal is to find a linear function that best describes this relationship. For multiple variables, the linear regression model can be expressed as:

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i + \epsilon, \quad (5)$$

where β_0 is the intercept, β_i are the coefficients of the independent variables, X_i are the independent variables, and ϵ is the error term. The coefficients were estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of the squared differences between the observed and predicted values. The OLS method is defined as:

$$\hat{\beta} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (6)$$

where y_i is the actual value and \hat{y}_i is the predicted value.

We also used stepwise regression [21] to select the most relevant features for composing an equation that predicts the cattle weight. The stepwise regression algorithm iteratively selects a subset of relevant variables for a regression model by adding and removing predictors based on specific criteria. In this paper, we performed tests using the p-value and the Akaike Information Criterion (AIC) as criteria. The AIC allows quantifying the trade-off between the goodness of fit and the complexity of the model. It is calculated as follows:

$$AIC = 2k - 2 \ln(\hat{L}), \quad (7)$$

where k is the number of parameters in the model and \hat{L} is the maximum value of the likelihood function for the model. The AIC penalizes the number of parameters in the model, favoring simpler models with fewer predictors.

To build a model with the most statistically significant predictors, we combined forward selection and backward elimination. Thus, starting with no features, we added variables one at a time and evaluated the statistical significance in the regression model. In the p-value test, we added the feature with the smallest p-value if it was less than 0.005, and in the AIC test, we added the feature if the AIC was improved. After each addition, we performed backward elimination, removing any feature in the model with a p-value exceeding 0.10 or that did not improve the AIC. This process continued until no features could be added or removed.

2.4.2. Non-linear regression

For the non-linear task, we employed the Random Forest (RF) algorithm [22] to predict cattle weight since it is less affected by multicollinearity between features and can handle a large number of features effectively. RF is an ensemble method that constructs multiple decision trees using bootstrapping and random feature selection. Each decision tree is trained on a bootstrap sample, which is created by sampling the original dataset with replacement. At each node of the tree, a random subset of features is evaluated through the mean squared error (MSE) to determine the best split that minimizes the variance of the target variable. In this study, we used a model consisting of 400 trees with a maximum depth of 20. These parameters were determined through grid search.

We also evaluated the effectiveness of artificial neural networks (ANN) [23] in predicting cattle weight using the extracted features. We employed a feedforward neural network with three hidden layers, each containing 64 neurons. We used the ReLU activation function for the hidden layers and a linear activation function for the output layer. The model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. We trained the model for 50 epochs and used MSE as the loss function.

2.5. Performance metrics

To evaluate the performance of the models, we used MAE and MAPE metrics. These metrics provide insights into the accuracy and precision of the predictions, allowing us to compare the models' performance and identify the most effective approach for cattle weight prediction.

The MAE is a metric that calculates the average absolute difference between the predicted and actual values. It is calculated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (8)$$

where n is the number of samples, y_i is the actual value, and \hat{y}_i is the predicted value.

The MAPE is a metric that calculates the average percentage difference between the predicted and actual values. It is calculated as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100, \quad (9)$$

where n is the number of samples, y_i is the actual value, and \hat{y}_i is the predicted value.

2.6. Dataset description

In our experiments, we used two public image datasets, referred to as B3 and B4, which are available on the Kaggle platform as part of the publication “Cattle Weight Detection Model + Dataset (12k)” [9]. Although these datasets do not have specific formal names, we provide a direct link to their source¹ to ensure transparency and reproducibility.

The B3 and B4 datasets contain side-view images of cattle, with the respective weight of each animal encoded in the image filenames. The images were captured under varying conditions, including different angles, distances, and lighting. Each animal was marked with a circular sticker of known and uniform size, which is also annotated in the segmentation masks provided.

Figs. 4 and 5 illustrate samples from the B3 and B4 datasets, respectively. The first row in each figure shows the original images, while the second row presents the corresponding segmentation masks, including the circular sticker.

To reduce the impact of scale variations due to distance between the camera and the cattle, we used the circular stickers as a reference for normalization. We first applied a combination of thresholding and edge detection to detect the sticker in each image. Then, we computed its radius and resized all images so that the sticker radius is fixed at 100 pixels. This normalization ensures consistent scale across the dataset, enabling more accurate extraction of body measurements and features.

The B3 dataset contains 2,588 images of cattle with weights ranging from 36 kg to 621 kg, while the B4 dataset comprises 1,934 images of cattle weighing between 69 kg and 307 kg. Fig. 6 presents the weight distribution histograms for both datasets.

2.7. Experimental setup

The proposed method was implemented using Python 3.9.16 and the Pytorch 1.13.1 API. The experiments were performed on a computer with an 12th Generation Intel® Core™ i7-12700, 2.10 GHz, NVIDIA® GeForce RTX™ 3090 card, 64 GB of RAM and Windows operating system with 64-bit architecture.

¹ <https://www.kaggle.com/dsv/8858637>.

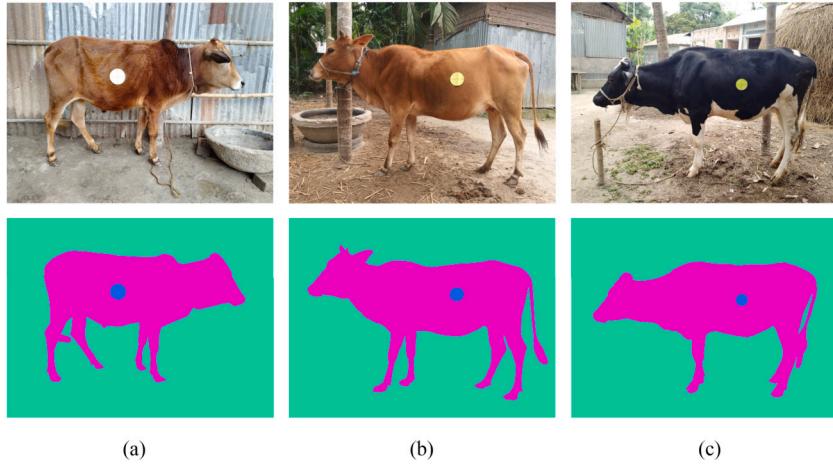


Fig. 4. Samples from the B3 dataset alongside its annotations from cattle weighting 109 (a), 256 (b), and 405 (c) kilograms.

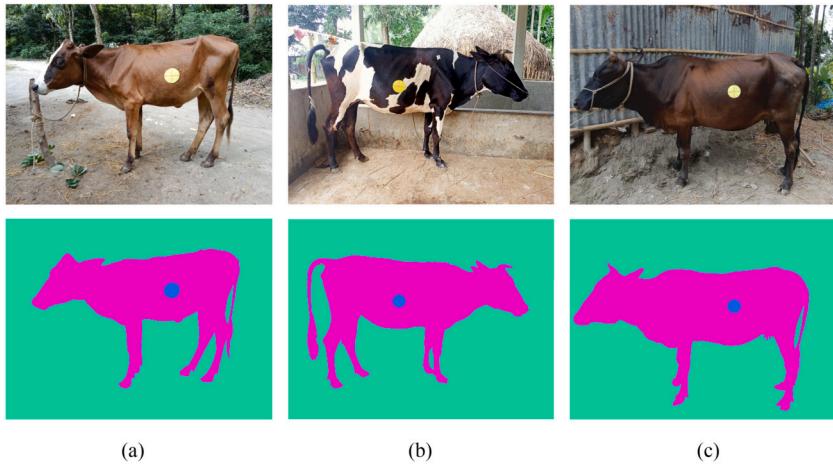


Fig. 5. Samples from the B4 dataset alongside its annotations from cattle weighting 114 (a), 274 (b), and 307 (c) kilograms.

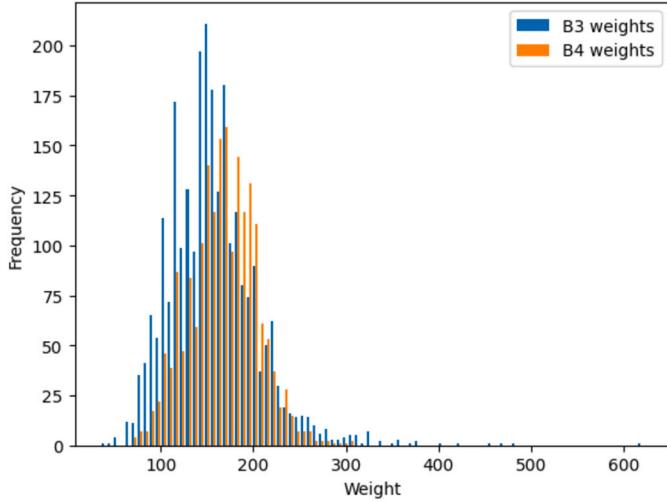


Fig. 6. Histogram of the cattle weights in the B3 dataset and the B4 dataset.

3. Results

3.1. Preprocessing results

We utilized the CAFE-Net model to generate segmentation masks for the B3 and B4 datasets. These masks were subsequently used to extract

features and aid in generating point clouds. Some examples of the segmentation results are displayed in Fig. 7, which shows two images from the B3 dataset (Figs. 7a and 7b) and two images from the B4 dataset (Figs. 7c and 7d). In our evaluation, the CAFE-Net model achieved an Intersection over Union (IoU) of 0.9738 and a Dice coefficient of 0.9497 for the B3 dataset. For the B4 dataset, the model attained an IoU of 0.9532 and a Dice coefficient of 0.9274.

Next, we generated the depth images, with three examples of the results presented in Fig. 8. The first column displays the original images, the second column shows the depth maps generated from these images, and the third column illustrates the depth maps produced from the segmented images. In this context, colors closer to red indicate regions nearer to the camera, while colors closer to blue represent areas farther away.

In Fig. 8a, it is evident that the depth map generated from the original image (in the second column) is of high quality, effectively capturing the volume of the animal with distinct details. For instance, the left legs, chest, and neck are depicted as being farther from the camera compared to the right legs, head, and body of the animal.

However, for the images in Figs. 8a and 8b, objects positioned in front of the animal interfered with the depth calculation, making the animal appear on a more distant plane from the camera and obscuring its details. Additionally, these objects disrupted the normalization of the images, resulting in discrepancies between the images. This issue was addressed by utilizing segmented images of the animal, as shown in the third column of Fig. 8. This approach enabled the depth model to

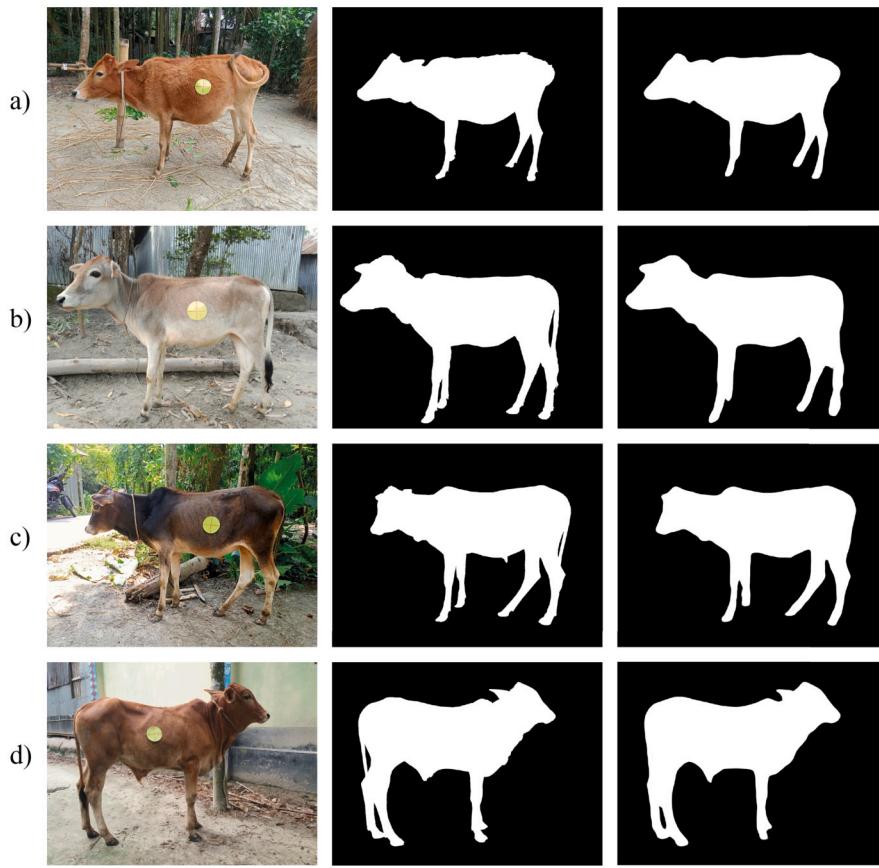


Fig. 7. Examples of segmentation results. The first column displays the original images, the second column shows the ground truth masks, and the third column illustrates the segmentation masks generated by the CAFE-Net model.

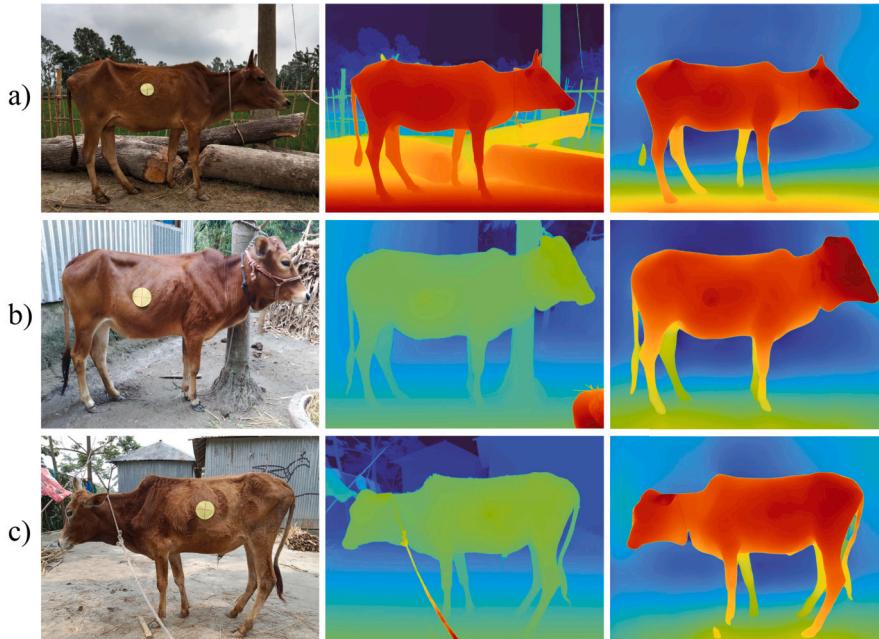


Fig. 8. Examples of depth estimation results. The first column displays the original images, while the second and the third columns showcase the depth maps generated by the Depth Pro model from the original and the masked images, respectively.

more accurately estimate the animal's depth, normalize the images, and eliminate unwanted objects from the scene.

Using Equation (2), we generated point clouds for the animals from the depth images. Fig. 9 presents an example of a point cloud created

from an image in the B3 dataset. The points in the cloud along the z-axis represent the distances from the points in the image to the camera plane. Additionally, the cloud displays noise and imperfections, particularly around the animal's contours and legs.

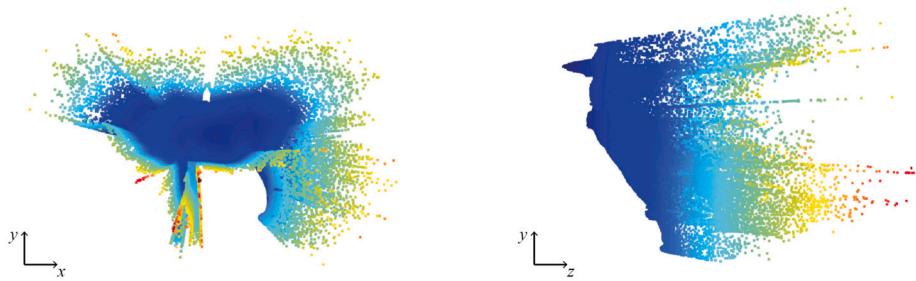


Fig. 9. Example of a point cloud generated for an image from the B3 dataset.

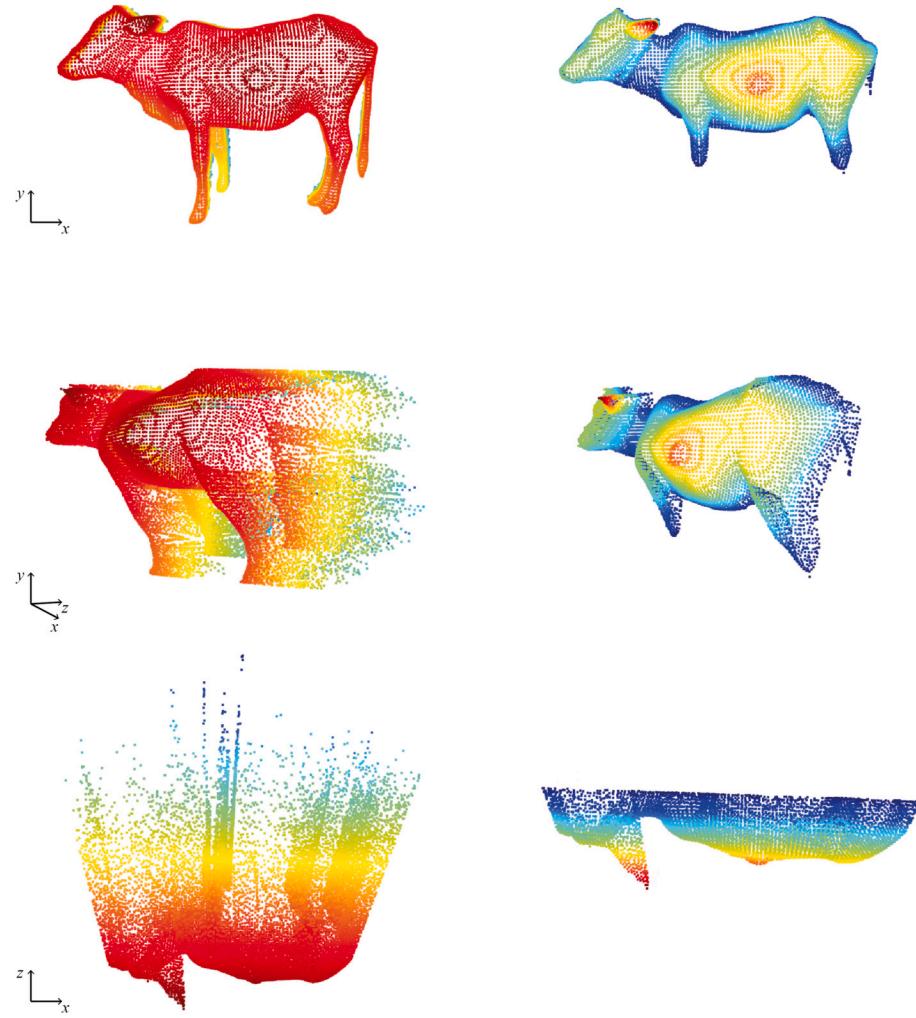


Fig. 10. Example of a point cloud generated for an image from the B3 dataset after preprocessing.

We performed several preprocessing steps to create a cleaner point cloud and extract more accurate features. First, we inverted the values along the z -axis and normalized them so that the outermost point of the animal had a z -value of 0. Next, we downsampled the point cloud, retaining only 2% of the original points. The resulting point cloud after preprocessing is shown in the first column of Fig. 10.

The next step involved removing noise from the z -axis. To do this, we analyzed the histogram of the z -values of the point cloud. The histogram for the point cloud depicted in Fig. 10 is shown in Fig. 11. From this histogram, we observed that the points belonging to the animal concentrated within a specific range of z -values, while the noise was spread over a larger range. Based on this observation, we defined a threshold using Otsu's method to eliminate the noise. The result of the point cloud after noise removal is shown in the second column of Fig. 10.

3.2. Baseline

The first test we conducted to calculate the cattle weight involved using the traditional Heart Girth Method (HGM) [24] to estimate weight based on our collected measurements. This method is commonly used to estimate the weight of cattle by taking their heart girth and body length measurements. The formula to estimate the weight is:

$$\text{Weight (kg)} = \left(\frac{\text{Heart Girth}^2 \times \text{Body Length}}{300} \right). \quad (10)$$

The formula takes measurements in inches and returns the weight in pounds; therefore, a multiplicative correction factor is needed to adjust the result. Using this formula, the MAE and MAPE were calculated

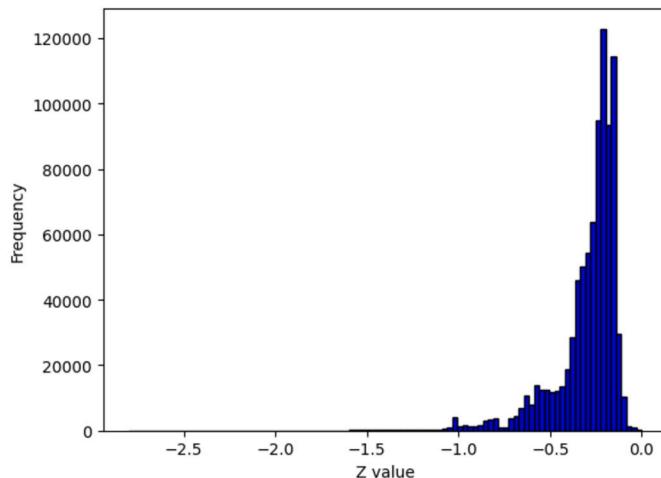


Fig. 11. Histogram of the z values of the point cloud from Fig. 10.

Table 2
Results of the Heart Girth Method for estimating cattle weight.

	B3	B4
MAE	40.69 kg	53.65 kg
MAPE	24.63%	31.36%

for the B3 and B4 datasets. The results are shown in Table 2. We used these results as a baseline to compare the performance of the models we developed.

3.3. Feature relevance analysis using normalized mutual information

We extracted a total of 84 features from the original side-view images, the segmentation masks, and the point clouds, as detailed in Section 2.2. These features included 2D body measurements, shape descriptors, and 3D metrics derived from the depth-based point clouds.

To evaluate the relevance of these features for cattle weight estimation, we conducted a correlation analysis using NMI. This metric was chosen because it captures both linear and non-linear dependencies between features and the target variable (cattle weight). Our feature selection criterion was to retain all features with an NMI greater than zero, indicating a meaningful relationship with weight.

The results showed that the vast majority of extracted features, including those based on the estimated 3D point clouds, had positive NMI scores with cattle weight. Only three features—the maximum beam angle, the minimum chain code, and the maximum chain code—exhibited zero mutual information and were therefore excluded from further analysis.

These findings indirectly validate the usefulness of the 3D features generated through monocular depth estimation. Although direct geometric validation of the point clouds was not feasible due to the lack of ground-truth 3D measurements, the positive NMI scores confirm that the depth-based descriptors contributed valuable information to the weight prediction task.

3.4. Regression results

We conducted experiments to evaluate the effectiveness of our proposed methodology for predicting cattle weight using 2D side-view images and 3D point clouds generated from these images. We employed both linear and non-linear regression models to predict the weights of the animals based on the extracted features.

For these tests, we divided the B3 and B4 datasets into training and test sets, allocating 80% of the data for training and 20% for testing. In

the experiments with the B3 dataset, we used 2,080 images for training and 520 images for testing. For the B4 dataset experiments, we utilized 1,547 images for training and 387 images for testing. This data split was consistently applied across all regression tests. The training set was used to fit the models, while the test set was utilized to evaluate their performance.

Additionally, we performed cross-dataset evaluations to assess the models' generalization capabilities. This involved training the models on one dataset and testing them on the other, which allowed us to evaluate their ability to predict cattle weight across different datasets. The results of these experiments are presented in the following sections.

3.4.1. Linear regression results

We applied OLS and LASSO methods to develop a linear regression model for predicting cattle weights, focusing on features that have mutual information greater than zero. Table 3 presents the results for both training and test sets, along with evaluations across different datasets.

When fitting the OLS method with the B3 training set, we achieved an MAE of 25.70 kg and a MAPE of 17.02% on the B3 test set. For the cross-dataset evaluation using the B4 test set, we attained an MAE of 24.78 kg and a MAPE of 14.41%. In contrast, when using LASSO with the B3 training set, we recorded an MAE of 25.45 kg and a MAPE of 16.81% on the B3 test set. On the B4 test set, we achieved an MAE of 24.56 kg and a MAPE of 14.23%.

Regarding the B4 dataset, the OLS method resulted in an MAE of 28.96 kg and a MAPE of 20.14% on the B3 test set. In the B4 test set, we achieved an MAE of 20.54 kg and a MAPE of 13.02%. For the LASSO method applied to the B4 dataset, we obtained an MAE of 29.01 kg and a MAPE of 20.28% on the B3 test set. In the B4 test set, we recorded an MAE of 20.46 kg and a MAPE of 12.91%.

These results indicate that the LASSO method generally outperforms the OLS method, particularly in cross-dataset evaluations. This suggests that LASSO effectively reduces the impact of irrelevant features on the model's performance, leading to more accurate predictions.

Next, we used SWLR to select a subset of features that would enhance the linear regression model's performance. The feature subsets were formed using p-value and AIC criteria, and the results are summarized in Table 3. For the B3 dataset, the p-value test yielded an MAE of 25.79 kg and a MAPE of 17.08% on the B3 test set. In comparison, the AIC test produced an MAE of 25.96 kg and a MAPE of 17.22%. For the B4 dataset, the p-value test showed an MAE of 29.55 kg and a MAPE of 20.86% on the B3 test set, while the AIC test resulted in an MAE of 29.46 kg and a MAPE of 20.72%.

The results indicate that using the AIC criterion slightly improved model performance compared to the p-value criterion. This observation is evident in the cross-dataset test (values marked with *), where the errors from models adjusted with subgroups based on AIC were slightly smaller than those adjusted using p-values. However, the stepwise regression did not enhance the performance of the linear regression model when compared to the original LR model.

3.4.2. Non-linear regression results

To examine potential non-linear relationships between the features and the weight of animals, we conducted tests using random forest and ANN models. The results of these experiments are summarized in Table 4. For the random forest model, we obtained an MAE of 25.98 kg and a MAPE of 17.21% on the B3 test set. On the B4 test set, we achieved an MAE of 25.41 kg and a MAPE of 14.85%. In the case of the ANN model, we recorded an MAE of 25.57 kg and a MAPE of 17.01% on the B3 test set, while on the B4 test set, we obtained an MAE of 25.04 kg and a MAPE of 14.41%. These results indicate that the random forest model outperformed the ANN. This suggests that the random forest was more effective at capturing the non-linear relationships between the features and the weight of the animals, leading to more accurate predictions.

Table 3

Results of the linear regression and stepwise regression models for predicting cattle weight. Values marked with * represent the results of the cross-dataset test.

Dataset	Method	Train set		Test set (B3)		Test set (B4)		
		MAE	MAPE	MAE	MAPE	MAE	MAPE	
LR	B3	OLS	24.05 kg	16.60%	25.70 kg	17.02%	24.78 kg *	14.41% *
		LASSO	24.13 kg	16.63%	25.45 kg	16.81%	24.56 kg *	14.23% *
	B4	OLS	19.40 kg	12.38%	28.96 kg *	20.14% *	20.54 kg	13.02%
		LASSO	19.72 kg	12.57%	29.01 kg *	20.28% *	20.46 kg	12.91%
SWLR	B3	p-value	24.26 kg	16.70%	25.79 kg	17.08%	25.34 kg *	14.62% *
		AIC	24.12 kg	16.59%	25.96 kg	17.22%	25.25 kg *	14.60% *
	B4	p-value	19.97 kg	12.72%	29.55 kg *	20.86% *	20.57 kg	13.05%
		AIC	19.69 kg	12.57%	29.46 kg *	20.72% *	20.43 kg	12.94%

Table 4

Results of the random forest and ANN regression models for predicting cattle weight. Values marked with * represent the results of the cross-dataset test.

Dataset	Train set		Test set (B3)		Test set (B4)		
	MAE	MAPE	MAE	MAPE	MAE	MAPE	
Random forest	B3	9.28 kg	6.46%	25.89 kg	17.01%	24.96 kg *	14.41% *
	B4	7.68 kg	4.92%	30.12 kg *	21.47% *	20.85 kg	13.28%
ANN	B3	17.87 kg	12.16%	26.07 kg	17.13%	27.22 kg *	15.92% *
	B4	18.17 kg	11.61%	29.36 kg *	18.86% *	21.72 kg	13.80%

4. Discussion

The results of the experiments indicate that the proposed methodology effectively predicts cattle weights using features extracted from side-view images and 3D point clouds derived from these images. This is particularly evident when comparing the performance of the HGM method (see Table 2), which yielded a MAPE of 24.63% for dataset B3 and 31.36% for dataset B4. In contrast, our method achieved MAPE values ranging from 12% to 19%, representing a reduction of up to 58.83% in the MAPE value.

Both linear and non-linear regression analyses demonstrated that the B3 dataset provided better generalization for the models, particularly evident in cross-dataset evaluations. The performance metrics achieved by the models trained with the B3 dataset on the B4 test set showed only a small difference when compared to models trained with the B4 dataset and evaluated on the same B4 test set. Conversely, the performance of the models trained with B4 and evaluated using B3 exhibited significant differences compared to models trained with B3 and evaluated on B3. This discrepancy can be attributed to the wider weight range present in the B3 dataset (see Fig. 11). Specifically, the B3 dataset has a weight range from 36 kg to 621 kg, while the B4 dataset's weights vary from 69 kg to 307 kg.

For instance, the LASSO model trained with the B3 dataset achieved a MAPE of 14.23% on the B4 test set. In comparison, the same model trained with the B4 dataset achieved a MAPE of 12.91% on the B4 test set, resulting in a difference of 1.32%. Alternatively, this model trained with B3 and evaluated on B3 recorded a MAPE of 16.81%, while this model trained with B4 and evaluated on B3 had a MAPE of 20.28%, leading to a difference of 3.47%.

These findings suggest that the B3 dataset contained more representative samples and features that more effectively captured the underlying relationships between the extracted features and the weights of the animals. Additionally, the B3 dataset included a greater number of images and animals than the B4 dataset, along with a wider range of weights, which likely contributed to the improved generalization of the models.

Based on the overall performance of the linear and non-linear regression models, it is clear that the LASSO linear regression model trained with the B3 dataset was the best performer. This model achieved a MAPE of 16.81% on the B3 test set and 14.23% on the B4 test set. The stepwise

regression models did not demonstrate any improvements over the original linear regression models, indicating that feature selection did not significantly enhance the model's performance. Furthermore, the non-linear regression models did not surpass the linear regression models, suggesting that the relationships between the extracted features and the weights of the animals were primarily linear. We report that the performance of the LASSO model using features extracted from CAFE-Net masks was 0.11 higher than that using ground-truth masks.

Figs. 12 and 13 display the predicted versus actual cattle weights from the B3 and B4 test sets in the best-case scenario: the linear regression LASSO model fitted with the B3 training dataset using all extracted features. The red line represents the ideal prediction, where the predicted and actual weights are equal. Notably, in both cases, the predicted values align closely with the ideal regression line, indicating that the model effectively captures the overall trend in the data. However, the observed points are dispersed around the line, leading to a broader spread of predictions. This behavior is particularly pronounced in the B4 test set (Fig. 13) within the 150 to 275 kg range. This variability suggests that while the model provides a reasonable approximation, its predictive accuracy is influenced by underlying factors.

One possible reason is that variations in image capture conditions contributed to the observed dispersion. While distance variations were addressed during preprocessing, differences in the angle from which the images were taken were not corrected. These angular variations may have distorted key measurements, leading to inconsistencies in extracted features and, as a result, greater prediction errors.

Another potential source of error arises from the MMpose model. Since this model was not fine-tuned due to a lack of ground truth key point annotations, its predictions may have introduced inaccuracies in the extracted measurements. Small errors in key point localization could accumulate, affecting the precision of the derived features and ultimately resulting in greater variability in the weight predictions. This issue likely contributed to the increased dispersion observed.

Finally, during our analysis of the data, we identified inaccuracies in some cattle weight annotations within datasets B3 and B4. Specifically, several samples indicated weights significantly lower than those expected for this type of animal. For instance, Figs. 14a and 14b show two cows weighing 36 kg and 73 kg, respectively. In contrast, Fig. 14c displays a cow weighing 176 kg. It is evident that these values are unrealistic. This inconsistency may have undermined the model's robust-

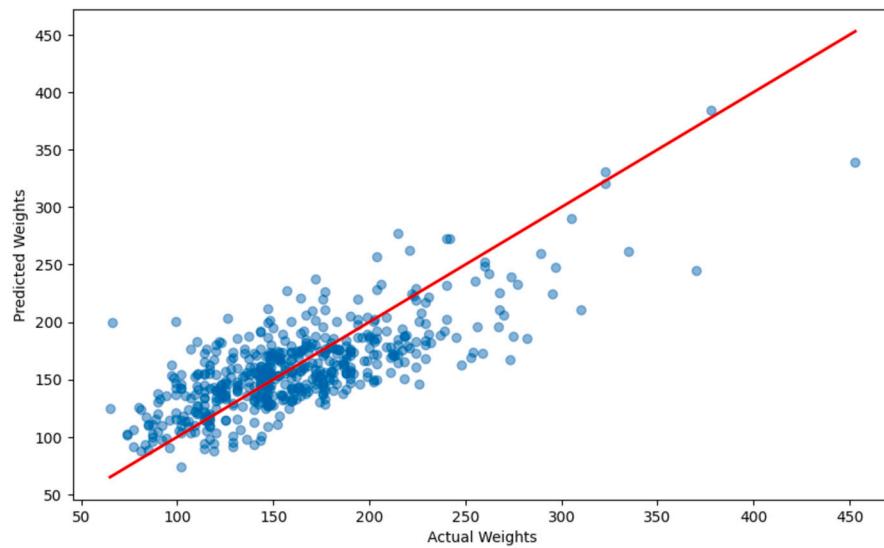


Fig. 12. Predicted vs. actual cattle weights for the B3 test set using the linear regression model.

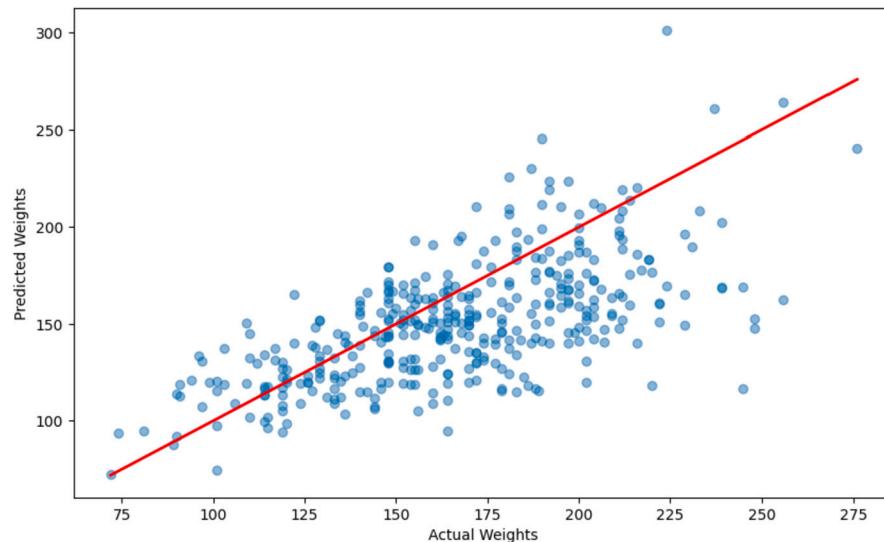


Fig. 13. Predicted vs. actual cattle weights for the B4 test set using the linear regression model.

ness by associating the visual features of adult cows with incompatible weights, thereby impairing both generalization and prediction accuracy.

To further investigate the influence of potential outliers on model performance, we conducted an additional experiment using a filtered version of the B3 dataset. Specifically, we selected a subset of samples whose weights fell within one standard deviation from the mean, thereby reducing the impact of extreme values. Using the LASSO regression model—previously identified as the best-performing approach—we retrained and evaluated the model on this filtered subset. The results showed an improvement in prediction accuracy: a MAPE of 11.62% and MAE of 17.63 kg on the B3 test set, and a MAPE of 10.02% and MAE of 17.55 kg on the B4 test set. These findings reinforce the hypothesis that outliers and annotation inconsistencies are significant contributors to prediction error and that the proposed methodology can achieve higher reliability when such noise is mitigated.

5. Conclusions

This paper presented a methodology for predicting cattle weights using only side-view images and 3D point clouds generated from these images. We extracted an extensive set of features from the images, in-

cluding body measurements, geometric, contour, chain code, and key point distances, to characterize the cattle's body shape, volume, and spatial relationships between key points. We conducted experiments using linear and non-linear regression models to predict cattle weights based on these features. The results demonstrated that the proposed methodology can effectively predict cattle weights, with the LASSO linear regression model providing the best performance. The experiments further highlighted the importance of the dataset used for training, with the B3 dataset providing better generalization for the models. The cross-dataset evaluations revealed that models trained on the B3 dataset performed better on the cross-dataset test set than those trained on the B4. The results suggest that the proposed methodology can be a valuable tool for estimating cattle weights using side-view images, with potential applications in precision livestock farming and animal health monitoring.

Future work can explore several directions to enhance the predictive capabilities of the proposed model. Improving the quality of key point annotations may reduce feature extraction errors and lead to more accurate predictions. Additionally, more complex regression techniques, such as ensemble methods, could be investigated to further improve performance. Addressing dataset limitations—such as angular variations and image quality—may also contribute to better generalization.

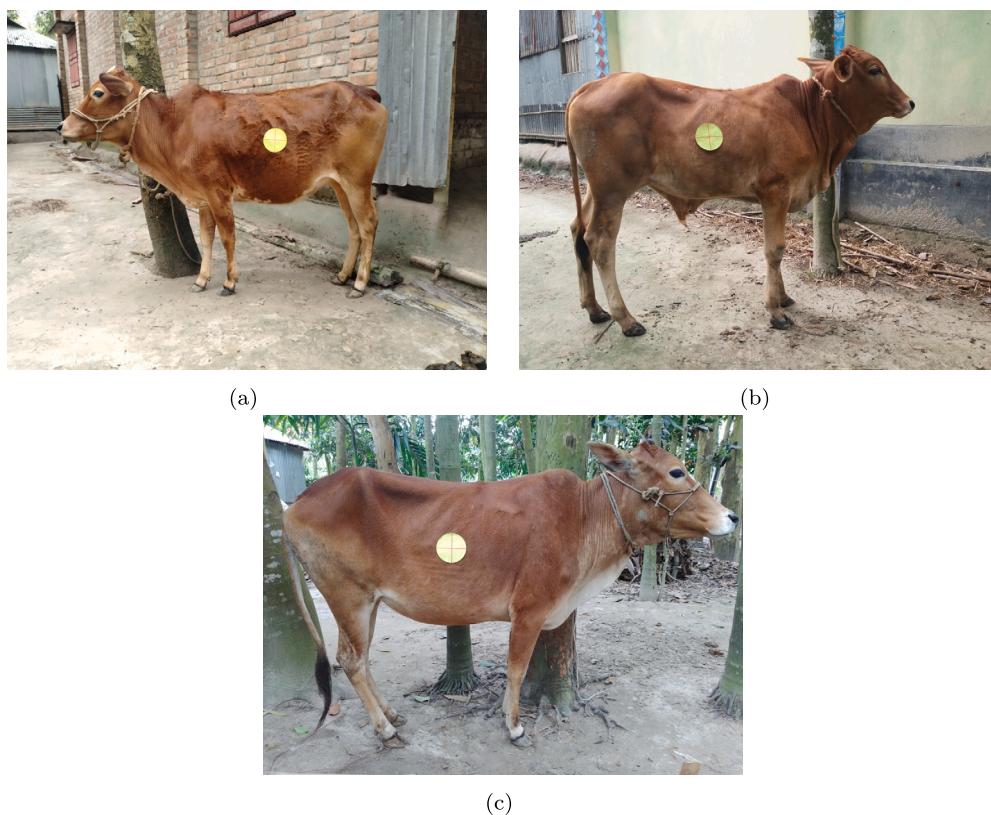


Fig. 14. Examples of cattle images labeled with 36 kg (a), 73 kg (b), and 176 kg (c).

Another promising avenue involves integrating the weight estimation system with farm management data, including feeding schedules and health records, to enable a more comprehensive and automated approach to herd monitoring. Moreover, breed-specific differences in body conformation may impact weight prediction accuracy. Future studies could explore breed-aware models or incorporate breed information as an input feature. Finally, an important direction is the analysis of the relationship between cattle weight and Body Condition Score. Since both metrics are vital indicators of animal health and nutrition, jointly modeling them or extracting features relevant to both tasks could offer a more complete and insightful evaluation framework.

CRediT authorship contribution statement

Guilherme Botazzo Rozendo: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation. **Maichol Dadi:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation. **Annalisa Franco:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Investigation, Conceptualization. **Alessandra Lumini:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Investigation, Conceptualization.

Funding

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

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.

Data availability

Data will be made available on request.

References

- [1] S. Immura, T.T. Zin, I. Kobayashi, Y. Horii, Automatic evaluation of cow's body-condition-score using 3d camera, in: 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), 2017, pp. 1–2.
- [2] A. Ruchay, V. Kober, K. Dorofeev, V. Kolpakov, K. Dzhulamanov, V. Kalschikov, H. Guo, Comparative analysis of machine learning algorithms for predicting live weight of Hereford cows, Comput. Electron. Agric. 195 (2022) 106837, <https://doi.org/10.1016/j.compag.2022.106837>.
- [3] Y. Zhao, Q. Xiao, J. Li, K. Tian, L. Yang, P. Shan, X. Lv, L. Li, Z. Zhan, Review on image-based animals weight weighing, Comput. Electron. Agric. 215 (2023) 108456, <https://doi.org/10.1016/j.compag.2023.108456>.
- [4] P.P. Kyaw, P. Tin, M. Aikawa, I. Kobayashi, T.T. Zin, Cow's back surface segmentation of point-cloud imaging using pointnet++ for individual identification, in: J.-S. Pan, T.T. Zin, T.-W. Sung, J.C.-W. Lin (Eds.), Genetic and Evolutionary Computing, Springer Nature, Singapore, Singapore, 2025, pp. 199–209.
- [5] M. Gjergji, V. de Moraes Weber, L. Otávio Campos Silva, R. da Costa Gomes, T. Luís Alves Campos de Araújo, H. Pistori, M. Alvarez, Deep learning techniques for beef cattle body weight prediction, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–8.
- [6] C.-b. Lee, H.-s. Lee, H.-c. Cho, Cattle weight estimation using fully and weakly supervised segmentation from 2d images, Appl. Sci. 13 (5) (2023), <https://doi.org/10.3390/app13052896>, <https://www.mdpi.com/2076-3417/13/5/2896>.
- [7] V.A.d.M. Weber, F.d.L. Weber, R.d.C. Gomes, A. d. S. Oliveira Junior, G.V. Menezes, U.G.P.d. Abreu, N.A. d. S. Belete, H. Pistori, Prediction of Girolando cattle weight by means of body measurements extracted from images, Rev. Bras. Zootec. 49 (2020) e20190110, <https://doi.org/10.37496/rbz4920190110>.

- [8] B. Xu, Y. Mao, W. Wang, G. Chen, Intelligent weight prediction of cows based on semantic segmentation and back propagation neural network, *Front. Artif. Intell.* 7 (2024), <https://doi.org/10.3389/frai.2024.1299169>.
- [9] A.A. Ltd, S. Roomy, A.B.S. Nayem, A.M. Tonmoy, S.M.S. Islam, M.M. Islam, Cattle weight detection model + dataset (12k), <https://doi.org/10.34740/KAGGLE/DSV/8858637>, <https://www.kaggle.com/dsv/8858637>, 2024.
- [10] M. Contributors, Openmmlab pose estimation toolbox and benchmark, <https://github.com/open-mmlab/mmpose>, 2020.
- [11] A. Bochkovskii, A. Delaunoy, H. Germain, M. Santos, Y. Zhou, S.R. Richter, V. Koltun, Depth pro: sharp monocular metric depth in less than a second, arXiv:2410.02073, <https://arxiv.org/abs/2410.02073>, 2024.
- [12] M.-K. Hu, Visual pattern recognition by moment invariants, *IRE Trans. Inf. Theory* 8 (2) (1962) 179–187, <https://doi.org/10.1109/TIT.1962.1057692>.
- [13] Y. Mingqiang, K. Kidiyo, R. Joseph, et al., A survey of shape feature extraction techniques, *Pattern Recognit.* 15 (7) (2008) 43–90.
- [14] P.A. Estevez, M. Tesmer, C.A. Perez, J.M. Zurada, Normalized mutual information feature selection, *IEEE Trans. Neural Netw.* 20 (2) (2009) 189–201, <https://doi.org/10.1109/TNN.2008.2005601>.
- [15] A. Lumini, G.B. Rozendo, M. Dadi, A. Franco, Comparison of cnn and transformer architectures for robust cattle segmentation in complex farm environments, in: 14th International Conference on Pattern Recognition Applications and Methods (ICPRAM), 2025, pp. 91–102.
- [16] CattleDetector, Cattlesegment dataset, universe.roboflow.com/cattledetector/cattlesegment-60nea, visited on 2024-09-30, <https://universe.roboflow.com/cattledetector/cattlesegment-60nea>, Aug 2023.
- [17] B. Xiao, H. Wu, Y. Wei, Simple baselines for human pose estimation and tracking, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 466–481.
- [18] J. Cao, H. Tang, H.-S. Fang, X. Shen, C. Lu, Y.-W. Tai, Cross-domain adaptation for animal pose estimation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9498–9507.
- [19] T.M. Cover, *Elements of Information Theory*, John Wiley & Sons, 1999.
- [20] D.C. Montgomery, E.A. Peck, G.G. Vining, *Introduction to Linear Regression Analysis*, John Wiley & Sons, 2021.
- [21] N. Draper, *Applied Regression Analysis*, McGraw-Hill, Inc, 1998.
- [22] L. Breiman, *Random forests*, *Mach. Learn.* 45 (2001) 5–32.
- [23] D.F. Specht, et al., A general regression neural network, *IEEE Trans. Neural Netw.* 2 (6) (1991) 568–576.
- [24] A. Heinrichs, H. Erb, G. Rogers, J. Cooper, C. Jones, Variability in Holstein heifer heart-girth measurements and comparison of prediction equations for live weight, *Prev. Vet. Med.* 78 (3) (2007) 333–338, <https://doi.org/10.1016/j.prevetmed.2006.11.002>, <https://www.sciencedirect.com/science/article/pii/S0167587706002455>.