



Depth video data-enabled predictions of longitudinal dairy cow body weight using thresholding and Mask R-CNN algorithms

Ye Bi^a, Leticia M. Campos^a, Jin Wang^b, Haipeng Yu^b, Mark D. Hanigan^a, Gota Morota^{a,c,*}

^a School of Animal Sciences, Virginia Tech, Blacksburg, VA, 24061 USA

^b Department of Animal Sciences, University of Florida, Gainesville, FL, 32611 USA

^c Center for Advanced Innovation in Agriculture, Virginia Tech, Blacksburg, VA, 24061 USA

ARTICLE INFO

Editor: Stephen Symons

Dataset link: https://github.com/yebigithub/BW_dairy

Keywords:

Body weight
Computer vision
Depth camera
Prediction
Video processing

ABSTRACT

Monitoring cow body weight is crucial to support farm management decisions due to its direct relationship with the growth, nutritional status, and health of dairy cows. Cow body weight is a repeated trait, however, the majority of previous body weight prediction research only used data collected at a single point in time. Furthermore, the utility of deep learning-based segmentation for body weight prediction using videos remains unanswered. Therefore, the objectives of this study were to predict cow body weight from repeatedly measured video data, to compare the performance of the thresholding and Mask R-CNN deep learning approaches, to evaluate the predictive ability of body weight regression models, and to promote open science in the animal science community by releasing the source code for video-based body weight prediction. An Intel RealSense D435 camera was installed on Virginia Tech dairy complex to collect top-view videos of 10 lactating Holstein cows twice a day for 28 days after milking sessions and 2 non-lactating, pregnant Jersey cows once a week upon calving. At the same time, the ground truth body weight records were collected using a walk-over weighing system. A total of 40,405 depth images and depth map files were obtained. Three approaches were investigated to segment the cow's body from the background, including single thresholding, adaptive thresholding, and Mask R-CNN. Four image-derived biometric features, such as dorsal length, abdominal width, height, and volume, were estimated from the segmented images and fitted using ordinary least squares, ridge regression, least absolute shrinkage and selection operator, and linear mixed models. Two cross-validation designs, forecasting and leave-three-cows-out, were used to evaluate prediction performance. The Pearson correlation between image-derived biometric features and scale-based body weight ranged from 0.74 to 0.95. On average, the Mask-RCNN approach combined with a linear mixed model resulted in the best prediction coefficient of determination and mean absolute percentage error of 0.98 and 2.03%, respectively, in the forecasting cross-validation. The Mask-RCNN approach was also the best in the leave-three-cows-out cross-validation, followed by adaptive and single thresholding. The prediction coefficients of determination and mean absolute percentage error of the Mask-RCNN coupled with the linear mixed model were 0.90 and 4.70%, respectively. Our results suggest that deep learning-based segmentation improves the prediction performance of cow body weight from longitudinal depth video data.

1. Background

The body weight of dairy cattle varies on a daily basis. Regular monitoring of body weight is important because estimates of body weight can provide insight into management factors that may affect milk production, feed efficiency, energy balance, growth, and health status. For example, a negative energy balance can potentially indicate an inadequate diet or health and reproductive problems. For this reason, individual monitoring of body weight throughout a cow's lifespan is essential to make better management decisions. However, body weight monitoring is not performed as often as it should be because it is labor- and time-consuming when measured manually or cost-prohibitive if using an automated walk-over weighing system. In addition, the trend toward larger average farm sizes provides more opportunities for increasing the

qu Coast diet or health and reproductive problems. For this reason, individual monitoring of body weight throughout a cow's lifespan is essential to make better management decisions. However, body weight monitoring is not performed as often as it should be because it is labor- and time-consuming when measured manually or cost-prohibitive if using an automated walk-over weighing system. In addition, the trend toward larger average farm sizes provides more opportunities for increasing the

* Corresponding author at: School of Animal Sciences, Virginia Tech, Blacksburg, VA, 24061 USA.

E-mail addresses: yebi@vt.edu (Y. Bi), leticiamc@vt.edu (L.M. Campos), jin.wang@ufl.edu (J. Wang), haipengyu@ufl.edu (H. Yu), mhanigan@vt.edu (M.D. Hanigan), morota@vt.edu (G. Morota).

<https://doi.org/10.1016/j.atech.2023.100352>

Received 20 September 2023; Received in revised form 27 October 2023; Accepted 31 October 2023

Available online 10 November 2023

2772-3755/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

efficiency of identification of cow health problems. The use of a low-cost, non-contact, stress-free 3 dimensional (3D) depth sensor camera coupled with computer vision analysis has been proposed as an alternative method to address the challenges associated with assessing cow morphometric traits, including predicting body weight [21]. Computer vision is considered one of the sub-fields in a broader effort of precision livestock farming, which aims to integrate advances in engineering and technology, and monitor animals at individual levels to positively intervene in farm management [18]. Although earlier computer vision applications for dairy cows involve the simultaneous use of multiple 2D cameras, a detailed calibration process is required to synchronize them [27,13]. A 3D depth sensor camera recently used in computer vision provides distances between the camera and objects in addition to standard two-dimensional information, including body contours. Height- and volume-derived 3D images have been shown to improve the performance of body weight predictions as assessed by model goodness of fit [12,5,17,30,14] and cross-validation [20,26,13].

Body weight is a repeated trait. However, previous body weight prediction research has used data collected at a single point in time without repeated measurements from the same cow. This type of experimental design does not allow evaluation of the predictive performance of models for observations that vary over time. Previous studies in dairy cattle that collected repeated body weight measurements used limited sample sizes, limited time points, manually selected the best image from each video, or did not employ cross-validation to evaluate predictive ability [17,30,14]. Therefore, an alternative approach is needed that can better explore longitudinal records and improve body weight prediction.

Deep learning has been recognized as mainstream in image and video data analysis. Yet, the usefulness of deep learning for dairy cow body weight prediction has not been assessed. Image segmentation is often a critical first step in computer vision to separate the cow from background information. The quality of cow biometric features extracted from segmented images heavily influences the predictability of traits of interest. Among various deep learning models, Mask R-CNN is used for instance segmentation, object detection, and object classification [6]. Although Mask R-CNN has been witnessing growth in cow detection [23,32,25,4,31], research evaluating the utility of Mask R-CNN for body weight prediction is in its early stages and has not been evaluated in great detail. Additionally, applying deep learning-based computer vision to animal video data often involves extensive data analysis using custom computer codes. Open science promotes research products and processes that are accessible to all levels of the community [22]. One way to practice open science is to make the source code used in research publications available to the public. However, it has been pointed out that the adoption of open source code in animal science is lagging behind other disciplines [19]. In particular, none of the computer vision-based body weight prediction work in the published literature to date has released its source code. The lack of available open source code makes animal computer vision data analysis processes less transparent and less reliable. Breaking down these barriers requires making open source code used in analysis accessible to the animal science community without any restrictions.

Therefore, the objectives of this study were 1) to predict cow body weight from repeatedly measured video data, 2) to compare the performance of automated thresholding and Mask R-CNN approaches, 3) to evaluate the predictive ability of body weight regression models, and 4) to release open source code for depth video-based body weight prediction to the animal science community.

2. Materials and methods

2.1. Animal experiments

Animal handling and media recording were approved and conducted in accordance with the Virginia Tech Institutional Animal Care and Use

Committee. A total of 12 multiparous animals were used. Ten lactating Holstein cows (286 days in milk) with average initial body weight (SD) of 793 (148) kg and two pregnant non-lactating Jersey cows with initial body weight of 552 (33) kg at 12- and 21-days ante-parturition from Kentland Farm Dairy Complex (Virginia Tech, Blacksburg, VA) were used in the current study. Lactating Holsteins and pregnant Jerseys were chosen to ensure that our computer vision pipeline is robust across breeds and physiological stages. Cows were housed in a free-stall barn, milked twice daily, fed in *ad libitum* regime with a common diet once daily, and had free access to water. Data were collected in 2020. Lactating animal body weight were recorded daily for 28 days after cows left the milking parlor following the 12:00 am and 12:00 pm milking sessions. Non-lactating animal body weight was recorded once a week. Both animals were handled and restrained individually over the walk-through scale, where the scale-based body weight was concomitantly and automatically recorded.

2.2. Video acquisition setup

An Intel RealSense D435 depth sensor camera (Intel, Santa Clara, CA, USA) connected to a laptop via a USB 3.1 cable was used to obtain depth video data from cows in this study. The sensor camera provides 87° horizontal and 58° vertical fields of view and uses two stereo image sensors to determine depth. A single-sensor camera was placed in a temperature-controlled box to maintain camera temperatures within the normal operating range. The box was mounted 2.95 meters above a one-way exit lane equipped with a walk-over weighing system between the milking parlor and pen housing, positioned straight down to capture top views of the cows. Equipment installation and location did not interfere with daily farm activities. Each cow was manually restrained to collect depth data in a short video format of 10-12 s at 30 frames per second and a resolution of 848 × 480. Although the cow's body motion was stable throughout the videos, the head and tail were relatively active. The total size of all the video data was 50.1 GB. During the video recording, the cow's body weight was automatically recorded by a walk-through weighing system (Afimilk, Kibbutz Afikim, Israel). Lactating animals were identified via pedometers located at the animals' ankle. Due to pedometer absence in non-lactating animals body weight, identification, and time recordings using the same scale were executed after milking time and manually recorded.

2.3. Image segmentation

After the completion of on-farm video acquisition, off-farm data process was initiated. The rs-convert program was used to convert .bag video files into PNG depth images and comma-separated value (CSV) depth map files [9]. The size of the video data increased to 103.2 GB after converting to PNG and CSV. Different colors in each depth image are proportional to distances from the camera to the object. The corresponding CSV file of the same size contains the distance in meters for each pixel. Three image segmentation techniques, single-thresholding, adaptive-thresholding, and Mask R-CNN, were used to segment the cow from its background. The background was considered extraneous information in the images that does not represent the cows' body (i.e., constraint fences).

2.3.1. Single-thresholding

The OpenCV library (version 4.7.0) was used in Python (version 3.10.9) for both thresholding methods. First, all depth images were cropped while maintaining the same ratio to remove the unnecessary surrounding area (i.e., background) while keeping the body part of the cow intact. The cropped images were converted from RGB (red, green, and blue) to HSV (hue, saturation, and value) color models (Fig. 1). The hue channel was extracted to segment the cow's body [10]. Hue values below the threshold were converted to black and those above were converted to white. Single-thresholding is often used in the literature for

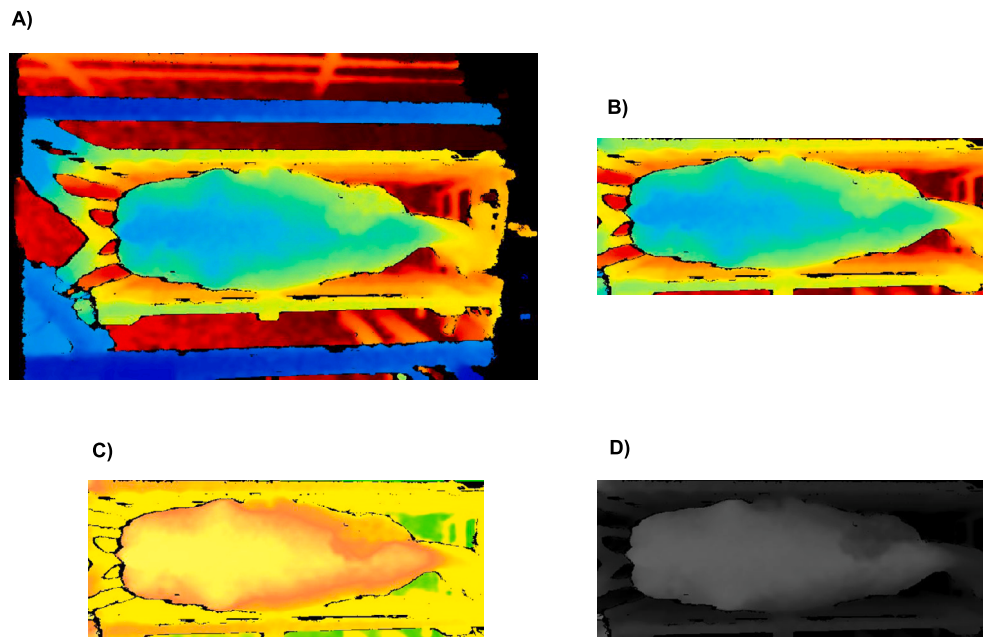


Fig. 1. Image conversion process. A) The original depth image. B) Cropped depth image. C) The RGB color model was converted to the HSV color model. D) Hue was extracted and used for subsequent analyses.

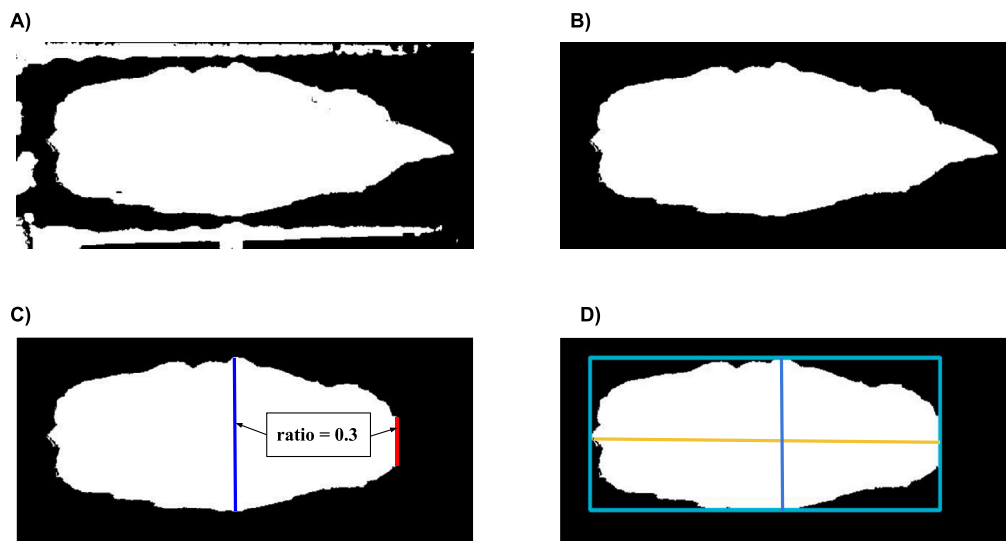


Fig. 2. Neck removal process. A) Binary image after thresholding. B) The initial segmentation that includes the neck. C) The ratio of body width was used to remove the neck. Blue line represents the column location that returns the maximum number of white pixels across all columns. Red line represents the column location that provides the ratio of 0.3 relative to the blue line. D) The final cow body area with the bounding box. The cyan box, yellow line, and blue line denote the bounding box, dorsal length, and abdomen width, respectively.

image-based body weight prediction [17,33]. In single-thresholding, the mean of all hue values was applied to all images as a fixed threshold. After thresholding, the OpenCV findcontour function was used to identify and isolate the largest area connecting all contiguous points of equal value, commonly known as the contour. The contour was then drawn on a black mask to remove depth-blank areas (i.e., missing data chunks) in the segmented cow binary image. The body width ratio was used to remove the neck in each image (Fig. 2) because previous studies have shown that removing the neck improves the prediction of body mass [5,17]. Specifically, the number of white pixels in each column was calculated and the maximum number of white pixels across all columns was kept. We then calculated the ratio of the total number of white pixels in each column to the maximum value, starting from the center to the head side of the image. If the ratio was less than 0.3 (according to

a preliminary analysis), we added several columns of black pixels to remove neck and also ensured that the image size remained the same. Finally, the contour was identified and the bounding box was drawn on the binary image. The bounding box and depth information associated with the identified contour were retained for body weight regression.

2.3.2. Adaptive-thresholding

A potential drawback of the single-thresholding approach is that the best threshold can vary dramatically from image to image due to differences in cow position and lighting conditions. For example, an inadequate threshold will result in incomplete background removal, often resulting in the cow's body being connected to the fence rail, while over-removal of the background will reduce the area of the cow's body part. Lighting conditions at noon and midnight also affect the pixel in-

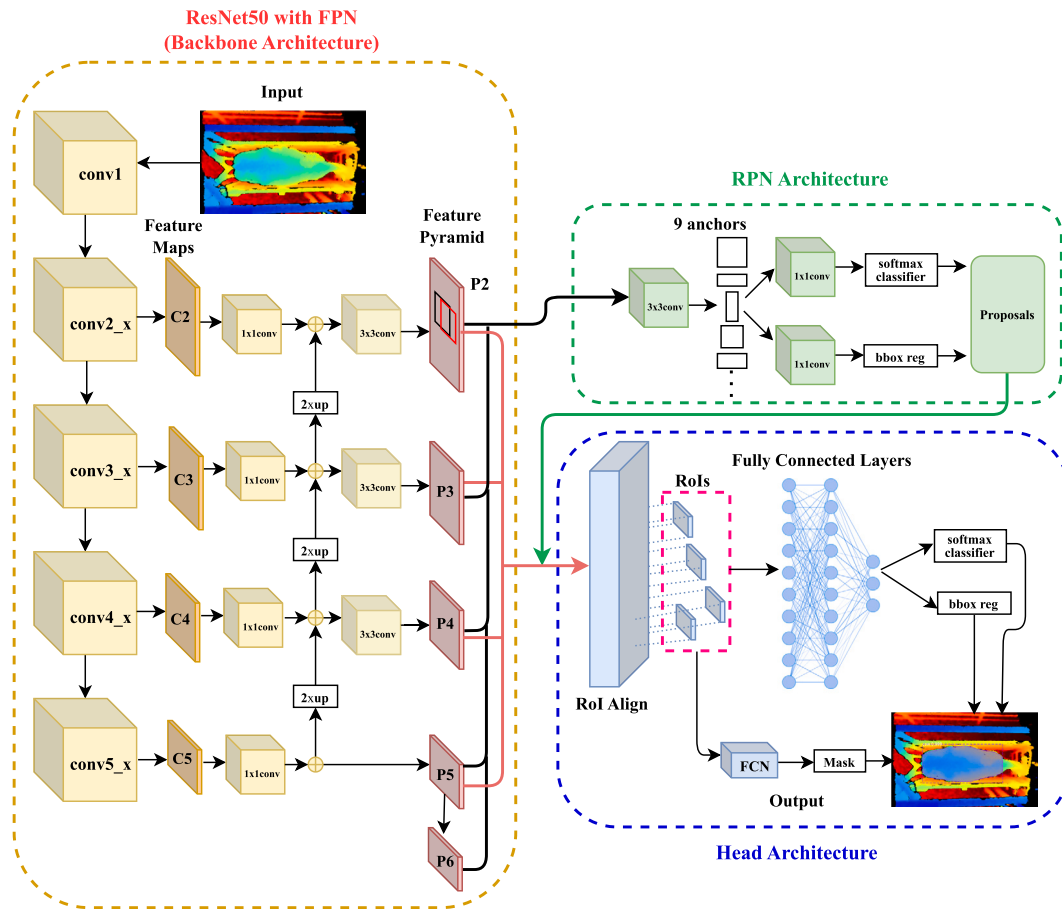


Fig. 3. Mask R-CNN architecture used in the current study. conv1: initial convolutional block. conv_x: multiple residual convolutional blocks. C2-C5: feature maps generated from each residual block. P1-P5: feature maps generated after the feature pyramid network (FPN). RPN: regional proposal network. RoI: region of interest. FCN: fully connected neural network. Bbox reg: bounding box regression.

tensity of an image. To address this challenge, single-thresholding was extended by exploring the optimal hue threshold from minimum to maximum in the adaptive-thresholding approach rather than keeping a fixed threshold. We started with the lowest hue threshold for each image to create a binary image. We drew the bounding box using the OpenCV boundingRect function around the largest contour to determine an appropriate threshold. If the coordinates of four corners were less than or equal to five pixels from the edge, the hue threshold was increased by one unit until the four corners were greater than five pixels from the edge of image, because in this case the entire cow body contour is obtained without being connected to the fence rails. The OpenCV findcontour function was then used to find the contour with the largest area, and the contour was drawn on a black mask to remove empty areas in the segmented cow binary image. Finally, we used the OpenCV minAreaRect function to draw the minimum bounding rectangle of the cow's body contour. The same image processing step was used to remove the neck in each image, identify the contour, and draw the final bounding box (Fig. 2).

2.3.3. Mask R-CNN

Mask R-CNN is developed on top of Faster R-CNN [24] by adding a new branch that generates a binary mask on each region of interest (RoI). In addition, Mask R-CNN also incorporates RoIAlign, which enhances feature alignment and aids in accurate mask generation. The overall architecture of Mask R-CNN is shown in Fig. 3. It includes the backbone network to extract features from the input image, followed by a region proposal network (RPN) that generates object proposals, and a head network that produces the final outputs of the model, in-

cluding object class labels, bounding box regression, and binary mask predictions for object instances in the input image.

Backbone: All depth images were provided to the backbone of Mask R-CNN after being resized to $1024 \times 1024 \times 3$ with zero padding. The backbone of Mask R-CNN was implemented using ResNet50, a 50-layer residual convolutional neural network model, to perform feature extraction over an entire image [7]. Its architecture includes five stages that extract features from low level (edges and corners) to high level (rich semantics). The first stage includes a convolutional layer and a max-pooling layer. The second through fifth stages contain three, four, six, and three iterations of three-layer convolutional blocks with different designs, respectively, followed by an average pooling layer. In total, there are 48 convolutional layers, one max pooling layer, and one average pooling layer. Stages 2 to 4 in ResNet50 include skip connections, which allow information to be passed directly from one stage to another without going through the intermediate layers, to avoid a vanishing gradient problem, improve prediction, and facilitate convergence. The bottom-up pathway in ResNet50, coupled with the rectified linear unit activation function from the last residual block of each stage, produces a pyramid of feature maps. They were further processed via the top-down pathway to develop the feature pyramid network (FPN) with the aim of improving segmentation performance [16].

Region proposal network: The feature maps generated from the backbone and FPN are fed into a region proposal network (RPN), which is a fully convolutional network, to generate the region proposals [24]. It uses sliding windows over the feature maps to detect whether the target object (i.e., cow) is present within the region proposals to define ROIs and determine their coordinates. In particular, anchors (rectan-

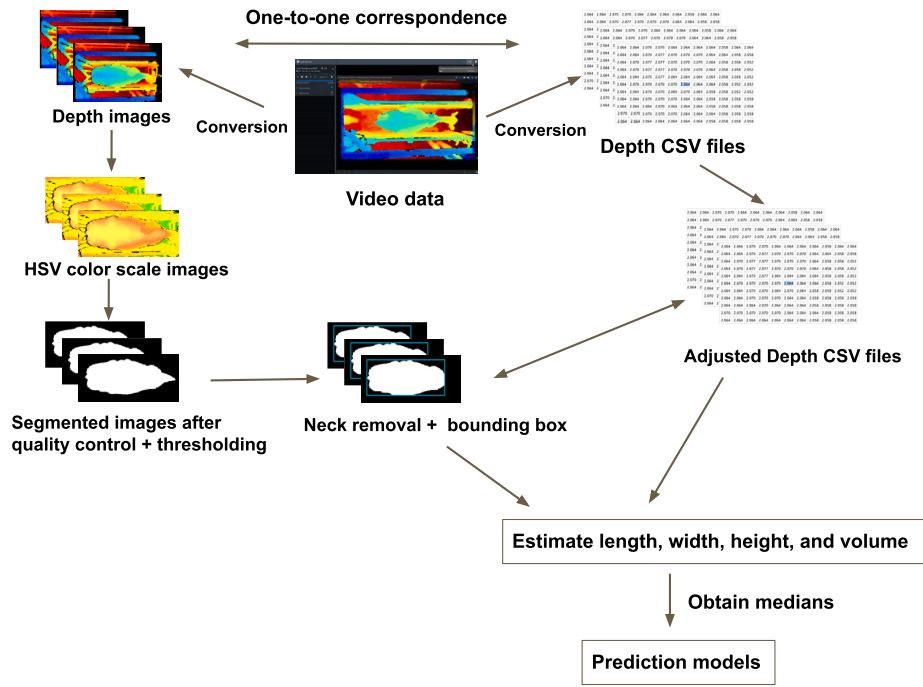


Fig. 4. Overview of the biometric approach using thresholding.

gles) with nine different sizes, three scales, and three aspect ratios were used. In total, about 21,000 anchors were used per image, since the feature map size after feature extraction was 40×60 . The softmax function was used for each anchor to classify the cow as foreground or background. Here, foreground means that the cow is located in that particular anchor. For anchors with the most possibilities to contain the cow, the bounding box size and location were corrected to get a better fit. Non-maximum suppression was applied to remove bounding boxes that overlapped with other bounding boxes with higher foreground scores. The final set of RoIs are regions in the feature map that contain the cow.

Region of interest align: Region of interest align (RoIAlign) was used to properly project the RoIs onto the extracted features to produce better pixel segmentation results [6]. It adjusts the size of the bounding box to a fixed size of 7×7 and splits the proposal into 7×7 sub-regions within each bin. The bilinear interpolated value of four top points from the image grid was used to keep a single value per bin.

Mask head: After the RoIAlign step, the fixed size RoIs were fed into a fully connected layer for object classification and bounding box regression. The RoIs were also fed to a small fully convolutional network for cattle body segmentation via mask prediction. Specifically, at each pixel location in the RoI, the fully convolutional neural network produced a mask prediction indicating the probability of that pixel belonging to the object class. The mask predictions are then post-processed to obtain the final cow segmentation by thresholding the mask probabilities and assigning each connected component to a unique object instance (i.e., cow). This process is equivalent to generating a mask for each RoI, allowing for pixel-level image segmentation.

Model training: A multi-task loss function was used to determine whether the training process of bounding box regression, classification, and mask prediction in a Mask R-CNN model was complete. It was calculated as $L = L_c + L_b + L_m$, where L_c is the classification loss, L_b is the bounding box regression loss, and L_m is the mask loss. The training process of Mask R-CNN was considered complete when the loss converged. The Mask R-CNN model was pre-trained on the COCO dataset, which is widely used for object detection, image segmentation, and image annotation [15]. It contains more than 300,000 annotated im-

ages. In addition to the COCO dataset, we randomly selected 300 depth images per cow per time point per day to include more specific situations for cow identification. The use of a pre-trained model is known as transfer learning, where we transferred the parameters from the previous training as initial values. We used Labelme to manually label the ground truth for the cow contour from the background [29]. These labeled cow images were added to the training dataset to fit the Mask R-CNN model. Segmentation performance was evaluated using a confidence score, which ranges from 0 to 1, with higher values indicating better segmentation accuracy [1]. The Mask R-CNN was implemented using Keras [1]. We used a batch size of 1, a step number of 300, an epoch of 42, learning rate 0.001. Since the original version was based on Python 3.4, TensorFlow 1.3, and Keras 2.0.8, we manually upgraded to Python 3.9, TensorFlow 2.9, and Keras 2.9, and updated the Mask R-CNN source code.

Prediction The estimated mask R-CNN parameters were used to segment cow contours in the remaining depth images. To reduce computational time, we chose to predict a subset of depth images for contour segmentation. Specifically, we excluded the first two frames in each video and retained one frame every 15 frames, resulting in predictions of 2 frames per second. The total number of newly predicted frames was $(40405 - 300)/15 = 2674$.

2.4. Biometric features

Instead of selecting the single best image from each video, all images were processed to estimate biometric features. The biometric features of the cows were estimated from the single-thresholding, adaptive-thresholding, and Mask R-CNN segmented contours. Overviews of the biometric approaches using thresholding and Mask R-CNN are shown in Figs. 4 and 5, respectively. The following four biometric features were extracted from each image.

Dorsal length and abdominal width: The dorsal length and abdominal width of the cow in each image were estimated by drawing the bounding box with minimum area using the OpenCV minAreaRect function.

Height: Two methods were evaluated to measure cow height: centroid-based and average-based, to measure cow height. In the first method,

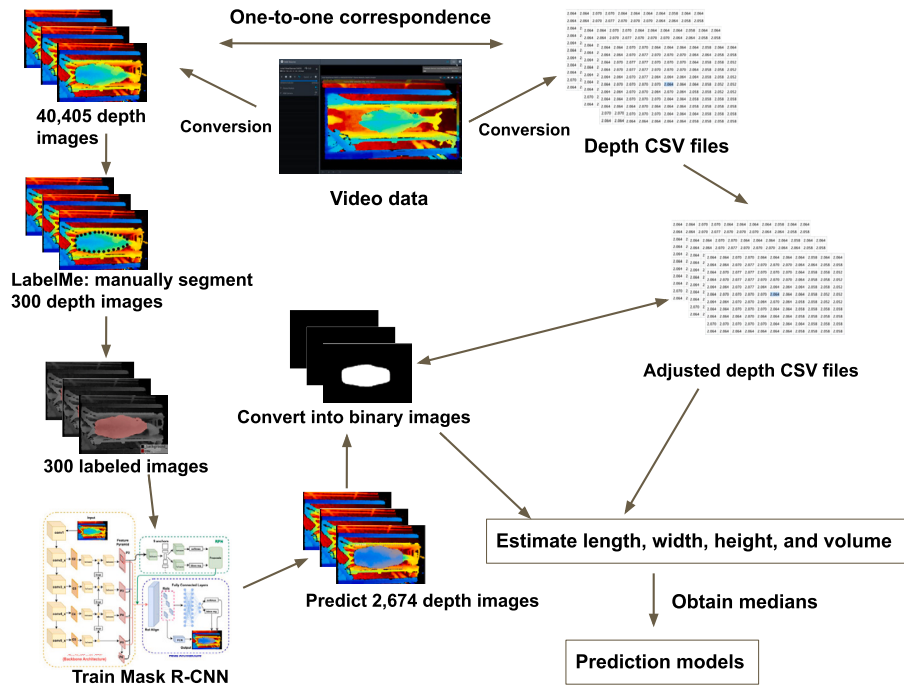


Fig. 5. Overview of the biometric approach using Mask R-CNN.

the centroid of the cow's contour was determined using the OpenCV moments function. The height of the cow was defined as 2.95 m (height of the camera) minus the distance from the camera to the centroid point using the corresponding CSV depth map file. In the second method, we replaced all zeros (missing data chunks) within contour pixels in the CSV file with the average distance, and the average of all distances between the camera and elements within the entire contour was obtained from the corresponding CSV file. The average cow height was defined as 2.95 m minus the average distance from the camera to the cow contour.

Volume: Because a 3D depth image captures the curvature of the back, the volume of the cow was obtained by summing the height pixels within the cow contour [20]. In mathematics, this is analogous to obtaining the volume by integrating the pixels of a cow area over its depth. An example of the estimated cow volume is shown in Fig. 6.

Once all images were processed and biometric features were estimated, the medians of dorsal length, abdominal width, height, and volume were calculated from all images to derive the single final estimates for each video. The median, rather than the mean, was chosen to make the final biometric estimates robust to outliers.

2.5. Body weight regression model

After obtaining all four biometric parameters (length, width, height, and volume), we used them as predictors to build regression models to predict dairy cow body weights. We used ordinary least squares (OLS), ridge regression (RR), and least absolute shrinkage and selection operator (LASSO) to estimate model parameters.

Ordinary least squares: OLS was constructed using width, length, height, and volume as predictors.

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{w} + \mathbf{l} + \mathbf{h} + \mathbf{v} + \epsilon,$$

where \mathbf{y} is a vector of body weight (kilograms), μ is the intercept, \mathbf{w} is a vector of width (pixels), \mathbf{l} is a vector of length (pixels), \mathbf{h} is a vector of height (meters), \mathbf{v} is a vector of volume (cubic meters), ϵ is a vector of residuals. The residual sum of squares (RSS) to be minimized for ordinary least square model is

$$RSS_{OLS} = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$$

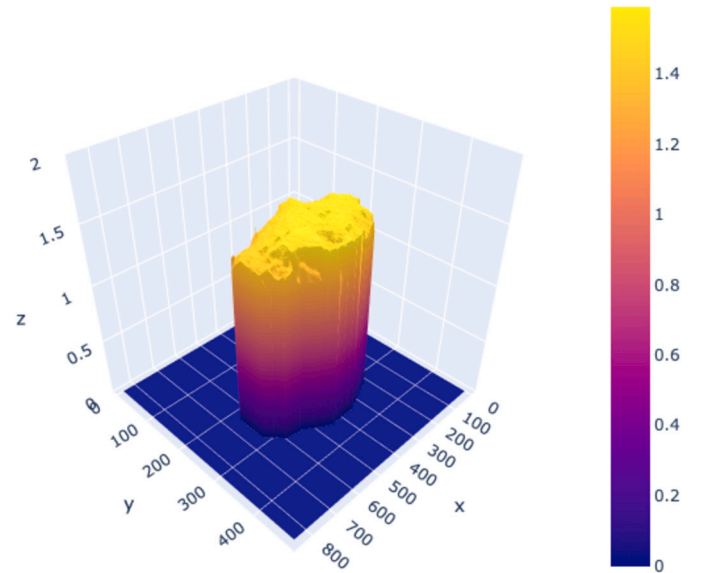


Fig. 6. Three-dimensional visualization of the cow volume. The volume of the cow was obtained by integrating the pixels of the cow area over its depth.

where \mathbf{X} is a matrix including the intercept, length, width, height, and volume; and β is a vector of regression coefficients for biometric predictors.

Penalized regression: Two types of penalized regression, RR [8] and LASSO [28], add a penalty function to the OLS objective function. Specifically, an L2 penalty was used for RR to induce shrinkage, while LASSO adds an L1 penalty to have shrinkage and sparsity (variable selection) properties. The RSS for RR and LASSO are

$$RSS_{RR} = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) + \lambda\hat{\beta}'\hat{\beta}$$

$$RSS_{LASSO} = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) + \lambda|\hat{\beta}|$$

where λ is the regularization parameter, which was fine-tuned using five-fold cross-validation. We implemented OLS, RR, and LASSO using the caret package of R [11].

Linear mixed model: A linear mixed model (LMM) including random intercepts and random slopes was applied to better capture cow variability or cow longitudinal trajectories of body weight records. The OLS was extended by considering individual cow differences as a random effect (random intercepts) and further assuming that the slopes between body weight and time points vary between cows (random slopes). The model used was the following.

$$y_{ij} = \alpha_{0j} + \alpha_{1j}t_{ij} + \epsilon_{ij}$$

$$\alpha_{0j} = \gamma_{00} + \sum_{\theta=1}^4 x_{j\theta} \beta_{\theta} + u_{0j}$$

$$\alpha_{1j} = \gamma_{10} + u_{1j},$$

where y_{ij} is the i th observation from the j th animal, t_{ij} is the animal centered time points, γ_{00} is the grand mean, $\sum_{\theta=1}^4 x_{j\theta} \beta_{\theta}$ is the 4 biometric predictors described earlier, u_{0j} is the random intercept of the j th animal, γ_{10} is the fixed effect of time points, and u_{1j} is the random slope of the j th animal associated with time points. The following variance-covariance matrix of the random effects were assumed

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{10} \\ \tau_{10} & \tau_1^2 \end{pmatrix} \right],$$

where τ_0^2 is the variance in intercepts between animals, τ_1^2 is the variance in slopes between animals, and τ_{10} is the covariance between intercepts and slopes. LMM was implemented using the lme4 package of R [3].

2.6. Model goodness of fit

Goodness of fit of OLS, RR, LASSO, and LLM was evaluated using the full data. This measure determines how well a model fits the current data (inference). In other words, goodness of fit assesses how well the model can predict body weight records that have already been used to estimate model parameters. Coefficient of determination (R^2) and mean absolute percentage error (MAPE) were used as goodness of fit evaluation metrics.

2.7. Cross-validation strategies

Cross-validation evaluates how well a model can predict new body weight records that were not used to estimate model parameters. Two cross-validation design strategies were used to test model prediction performance: time series forecasting and leave-three-cows-out (Fig. 7). In time series forecasting, data from earlier time points were trained to predict the body weight of later time points. All data were divided into training and testing sets according to time points. We varied the ratio of training and testing sets to have five scenarios: 90%:10%, 80%:20%, 70%:30%, 60%:40%, and 50%:50%. Here, 50%:50% indicates that the first half of the time points were used to predict the remaining half. In leave-three-cows-out, nine and three cows were used as the training and testing sets, respectively. This corresponds to partitioning the total data into 75% and 25%. All possible combinations of training and testing set partitioning were considered ($\binom{12}{3} = 220$). Prediction R^2 and MAPE were used to evaluate the model.

3. Results

Scale-based body weight measurements automatically collected via the walking-through weighting system across time are shown in Fig. 8. The record taken on Day 5 am was removed because of a systematic error due to the scale. The bottom two cows were non-lactating Jerseys, which had a lower body weight than the Holsteins.

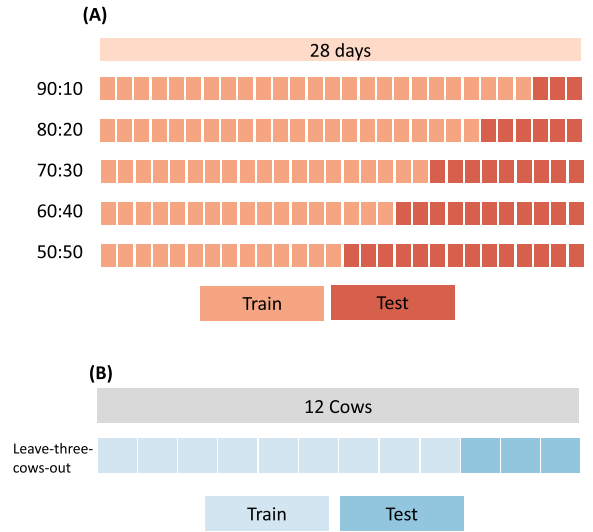


Fig. 7. Cross-validation design for body weight prediction. Time series forecasting (A). 90:10: the first 90% of time points were used to predict the remaining 10%. 80:20: the first 80% of time points were used to predict the remaining 20%. 70:30: the first 70% of time points were used to predict the remaining 30%. 60:40: the first 60% of time points were used to predict the remaining 40%. 50:50: the first 50% of time points were used to predict the remaining 50%. leave-three-cows-out (B). Of 12 cows, 9 and 3 cows were used as the training and testing sets, respectively.

3.1. Pearson correlation analysis

The confidence score of the trained Mask-RCNN model using 300 images was 0.99, indicating that the model training process was successful. Pearson correlation coefficients between scale-based body weight and five biometric features obtained from the image analysis averaged across time points ranged from 0.74-0.93, 0.80-0.93, and 0.80-0.95 for single-thresholding, adaptive-thresholding, and Mask R-CNN, respectively (Fig. 9). Volume was the most highly correlated with body weight (≥ 0.93) for all single-thresholding and Mask R-CNN and the second most correlated for adaptive thresholding. Width was the most correlated biometric feature for adaptive-thresholding and the second most correlated for Mask R-CNN. Length was the least correlated for single-thresholding. Average heights were consistently more correlated than centroid heights. Overall, Mask R-CNN, which provides pixel-level segmentation, resulted in the highest correlation with biometric features and body weight. The patterns of Pearson correlations between scale-based body weight and image-based biometric features were stable over time (Fig. S1). Volume and width showed high correlations in all the methods. Again length from single-thresholding was not strongly correlated with body weight. We did not observe major differences in correlations when comparing midnight and noon body weights (Fig. S2).

3.2. Model goodness of fit

Based on the correlation analysis, average height rather than centroid height was selected to build regression models for goodness of fit analysis and cross-validation. The extent to which the inferred image-based biometric features explained the observed body weight was evaluated by goodness of fit using the full data. Regardless of the regression models used, the Mask R-CNN segmentation yielded the best goodness of fit according to R^2 and MAPE (Table 1). Adaptive-thresholding consistently outperformed single-thresholding when OLS, RR, and LASSO were used as regression models. In contrast, the three segmentation methods performed equally well when LMM was fitted ($R^2 = 0.98$ and MAPE = 1.80%), which was the best of all. Regarding the regression models, LMM was always the best, followed by OLS, LASSO, and RR.

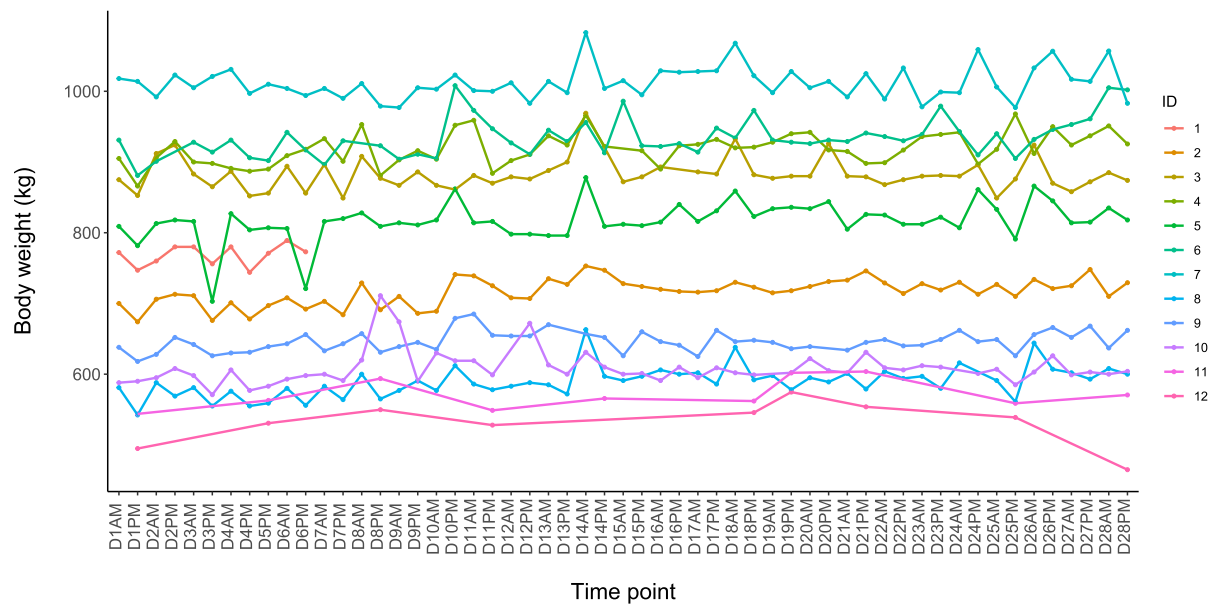


Fig. 8. Scale-based cow body weight records across time points. ID 1-10: lactating Holstein cows. ID 11-12: non-lactating Jersey cows.

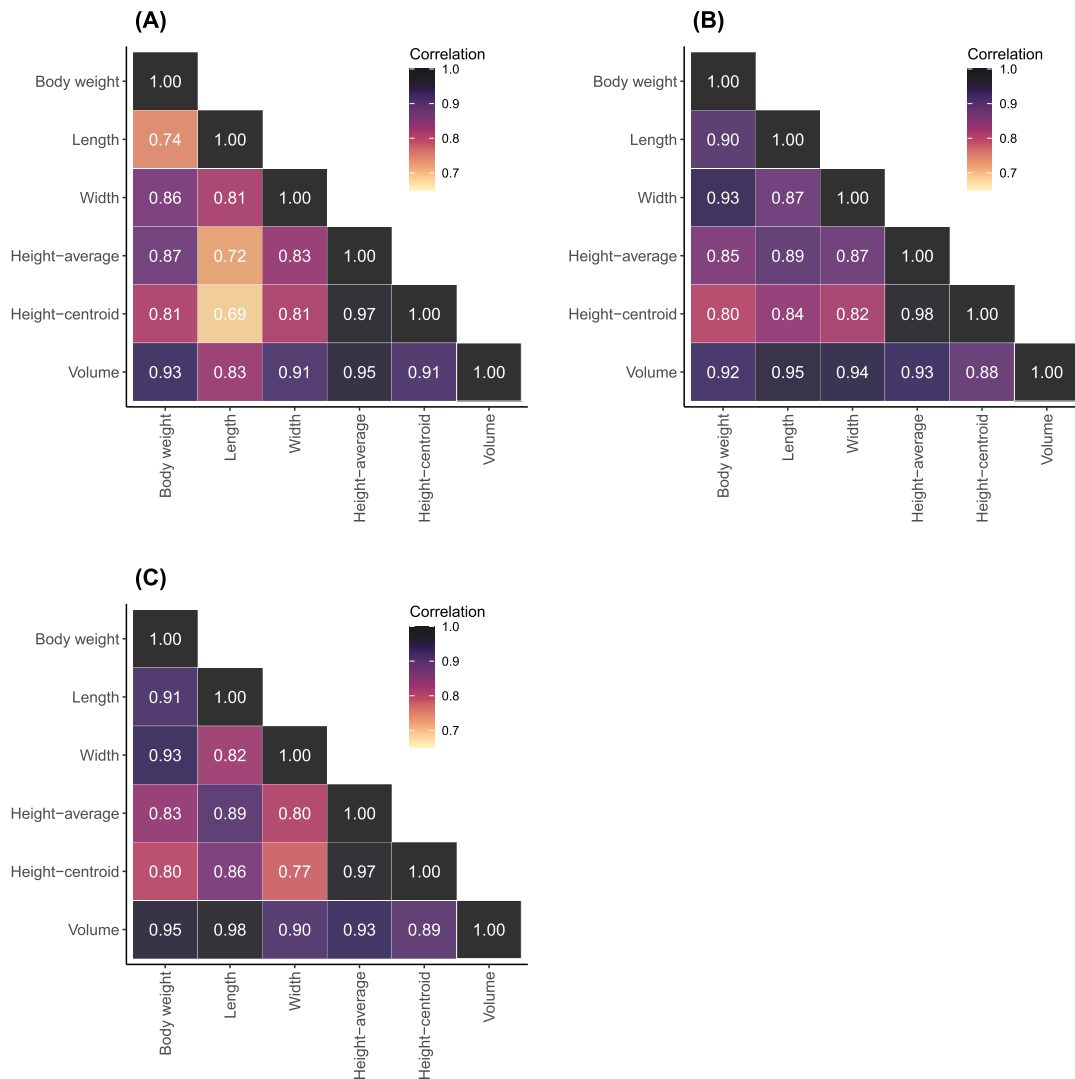


Fig. 9. Pearson correlation heat map between scaled-based body weight, length, width, centroid height, average height, and volume averaged across time points. A) single-thresholding, B) adaptive-thresholding, and C) Mask R-CNN.

Table 1
Model goodness of fit.

Method	R ² ¹				MAPE (%) ²			
	OLS ³	RR ⁴	LASSO ⁵	LMM ⁶	OLS	RR	LASSO	LMM
Single-thresholding	0.88	0.85	0.87	0.98	5.93	6.42	6.04	1.80
Adaptive-thresholding	0.91	0.89	0.91	0.98	4.72	5.40	4.84	1.80
Mask R-CNN	0.95	0.94	0.95	0.98	3.57	4.36	3.71	1.80

¹ Coefficients of determination² Mean absolute percentage error³ Ordinary least squares⁴ Ridge regression⁵ Least absolute shrinkage and selection operator⁶ Linear mixed model**Table 2**
Forecasting cross-validation results.

Method	Scenario ¹	R ² ²				MAPE (%) ³			
		OLS ⁴	RR ⁵	LASSO ⁶	LMM ⁷	OLS	RR	LASSO	LMM
Single-thresholding	90:10	0.89	0.86	0.88	0.98	5.94	6.91	6.29	2.10
	80:20	0.89	0.86	0.88	0.98	5.82	6.51	5.99	2.17
	70:30	0.87	0.86	0.87	0.98	6.22	6.57	6.29	2.06
	60:40	0.88	0.86	0.87	0.98	6.07	6.49	6.13	2.14
	50:50	0.88	0.87	0.88	0.98	6.02	6.40	6.03	3.28
Adaptive-thresholding	90:10	0.93	0.92	0.93	0.98	4.74	5.30	4.75	2.12
	80:20	0.91	0.89	0.91	0.98	5.16	5.72	5.22	2.22
	70:30	0.91	0.89	0.91	0.98	5.05	5.69	5.14	2.14
	60:40	0.91	0.89	0.90	0.98	5.00	5.72	5.16	2.39
	50:50	0.91	0.89	0.90	0.98	4.96	5.68	5.10	3.44
Mask R-CNN	90:10	0.96	0.94	0.95	0.98	3.41	4.30	3.53	2.03
	80:20	0.96	0.94	0.95	0.98	3.45	4.40	3.66	2.17
	70:30	0.96	0.93	0.95	0.98	3.51	4.58	3.87	2.18
	60:40	0.96	0.93	0.95	0.98	3.54	4.58	3.84	2.33
	50:50	0.96	0.93	0.95	0.98	3.60	4.59	3.90	3.59

¹ 90:10: 90% training and 10% testing; 80:20: 80% training and 20% testing; 70:30: 70% training and 30% testing; 60:40: 60% training and 40% testing; and 50:50: 50% training and 50% testing² Coefficients of determination³ Mean absolute percentage error⁴ Ordinary least squares⁵ Ridge regression⁶ Least absolute shrinkage and selection operator⁷ Linear mixed model

3.3. Cross-validation

Forecasting cross-validation results are shown in Table 2. Three regression models, OLS, RR, and LASSO, included only the image-based biometric features as predictors, while LMM included individual cow variability in the model. When the biometric features alone were used to predict body weight, the larger the training data, R² and MAPE were slightly larger and smaller, respectively, suggesting that having more past records is useful for predicting future body weight. However, this trend was not observed when single-thresholding biometric features were fitted with RR or LASSO. Mask R-CNN-based segmentation consistently produced the best prediction results. The next best segmentation method was adaptive-thresholding followed by single-thresholding. When R² was used to evaluate the prediction performance, OLS and LASSO performed equally well. The prediction performance of RR was slightly lower than OLS and LASSO. The highest R² of 0.96 was achieved by the combination of Mask R-CNN and OLS. According to MAPE, OLS performed best, followed by LASSO and RR. Similar to R², the lowest MAPE of 3.41% was obtained by the combination of Mask R-CNN and OLS. When individual cow variability was added to the prediction model using LMM, the prediction performance increased. Similar to the models including biometric features alone (OLS, RR, and LASSO), lower MAPE was observed when historical records were included in the training data. However, there was no obvious difference between the three segmentation methods used. Overall, LMM achieved the highest prediction R² of 0.98 and the lowest MAPE of 2.03.

The results of the leave-three-cows-out cross-validation are shown in Fig. 10. Within each regression model, segmentation-derived parameters from Mask R-CNN consistently gave the best prediction according to R² and MAPE. Adaptive thresholding was the second best, followed by single-thresholding. Although there was not much difference between OLS, RR, and LASSO when evaluated by R², OLS produced the best predictions based on MAPE. The combination of Mask R-CNN and OLS produced the highest average R² of 0.87 and the lowest average MAPE of 4.53.

4. Discussion

In this study, we evaluated three segmentation methods for their preciseness in segmenting body images of dairy cows from background. This segmentation process is the important first step for the application of computer vision in precision livestock farming, because the quality of the biometric features (length, width, height, and volume) extracted after the image segmentation step greatly influences the subsequent data analysis. We fed the biometric features obtained from the three segmentation methods into four different regression models to evaluate the predictability of longitudinal body weight. The main contribution of the current work was to evaluate the utility of Mask R-CNN instance segmentation and contour extraction for cow body weight prediction in great detail and to compare it with two types of thresholding approaches.

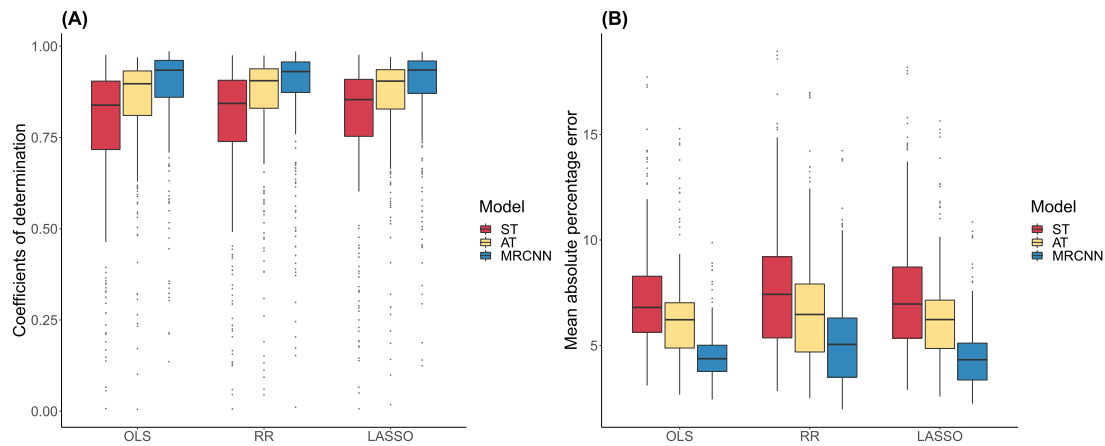


Fig. 10. Leave-three-cows-out cross-validation results. A) Prediction coefficient of determination. B) Mean absolute percentage error. ST: single-thresholding. AT: adaptive-thresholding. MRCNN: Mask R-CNN. OLS: Ordinary least squares. RR: Ridge regression. LASSO: Least absolute shrinkage and selection operator.

Of the four biometric measurements, volume was found to correlate most closely with body weight. Although volume is not always used to predict body weight in dairy cows, the importance of volume is consistent with the recent literature [13,30,14]. For instance, [13] developed a 3D reconstruction system using cloud points to perform body weight prediction on 64 lactating Holstein cows. They found that body weight and volume were most strongly correlated. Theoretically, volume simultaneously accounts for all other biometric features (i.e., length, width, and height). Width was the most correlated and second most correlated with body weight in adaptive-thresholding and Mask R-CNN, respectively. The importance of width is consistent with [26] reporting among rump length, hip height, and hip width, hip width had the largest correlation with body weight. Height estimates obtained from averages were more correlated with body weight than height estimates obtained from centroids of body regions. This is because, unlike pigs, the height of a cow is often not flat with a curved surface. For this reason, average-based height estimates were used in the regression analyses in our study.

The quality of biometric features was evaluated from two aspects: goodness of fit and predictability. In the current study, the combination of Mask R-CNN and LMM provided the best model fitness of R^2 of 0.95 and MAPE of 3.57% when cow variability was not considered. When cow variability was included in the prediction model, R^2 and MAPE improved to 0.98 and 1.80%, respectively. These values obtained from without or with the inclusion of cow variability were similar or better than previous studies that reported R^2 of 0.80 using six different types of body length and width measurements from eight cows [12], MAPE of 6.1% using the volume of 185 cows [5], R^2 of 0.96 using the rump width, thorax width, and dorsal area of 28 Holstein cows and 27 Holstein heifers [17], R^2 of 0.92 using the volume of 16 cows [30], R^2 of 0.98 using the volume of 16 cows coupled with LMM including cow variability [30], R^2 of 0.94 using the volume, partial surface area, hip width, chest depth, diagonal length, and heart girth, and body condition score of 16 cows [30], R^2 of 0.98 using the volume, partial surface area, hip width, chest depth, diagonal length, and heart girth and body condition score of 16 cows coupled with cow variability [30], and R^2 of 0.98 using the volume and hip width of 5 heifers [14]. Although the degree of goodness of fit varied among the three segmentation methods, the choice of segmentation method was no longer relevant when LMM was used as the regression model. This suggests that the inclusion of cow variability can compensate for the disadvantages of thresholding compared to Mask R-CNN.

The first type of prediction scenario was evaluated using forecasting cross-validation. Overall, better prediction results were obtained when a larger number of past records were included in the training sets, suggesting the importance of training data size in forecasting. The best R^2 of 0.96 and MAPE of 3.41% was achieved by the combination of Mask

R-CNN and OLS when cow variability was not included. When cow variability was added to a prediction model, R^2 and MAPE improved to 0.98 and 2.03%, respectively. Similar to the goodness of fit results, adding cow variability increased predictions and obscured the differences between the segmentation methods. Because body weight is a repeatedly measured trait, the random effect of the animals contains information about the body mass of each animal. Thus, this information is something we can use when predicting the same cow. We found that the performance of OLS was better than RR or LASSO. This is probably because shrinkage or variable selection was not strongly needed since the total number of predictors used was only four. Our results show that it is possible to predict future body weight of cows from their past biometric features. One potential application is when ground truth body weight records can only be sparsely collected manually on different days for each cow due to the lack of a walk-over weighing system. In such a case, forecasting can be used to predict body weight on specific days from image data alone.

Predictability of new cows not used in the training data was evaluated using leave-three-cows-out cross-validation in the second type of a prediction scenario. Since new cows are predicted in the testing sets, the use of LMM was not feasible in this scenario unless there is genetic information (pedigree or genomics) that allows the cows in the training and testing sets to be connected [2]. The combination of Mask R-CNN and OLS achieved the highest average R^2 of 0.86 and the lowest average MAPE of 4.59%. Overall, our models were able to deliver acceptable predictive performance, but we observed some negative outliers where the predictions were not as successful (Fig. 10). These outliers (low R^2 and high MAPE) were further investigated by taking a closer look at the specific partitioning of the training and testing sets that delivered the top 4 predictions and the bottom 4 predictions (outliers) based on R^2 and MAPE (Fig. 11). We found that the bottom 4 predictions always occurred when extrapolation was performed. In such a case, all three cows in the testing sets were in either the upper tail or lower tail of the body weight records. For example, the testing set contained two Jersey cows and one small Holstein cow. In contrast, the top 4 predictions were observed when interpolation was performed where all three cows in the testing sets were within the body weight range of the training set. This contrast occurred because of the relatively small sample size used in this study. It is likely that predictions will be more stable as more data are added to leave-several-cow-out cross-validation. To our knowledge, no previous prediction studies employing cross-validation similar to leave-three-cows-out cross-validation with repeated body weight records were found. Although this prevents a direct comparison of the predictive performance obtained in this study and previous work, our results were within the range or slightly better than previous studies that used larger numbers of cows. For example, R^2 of 0.95 and MAPE of 5.60% were reported with a hold out method using length, width, area,

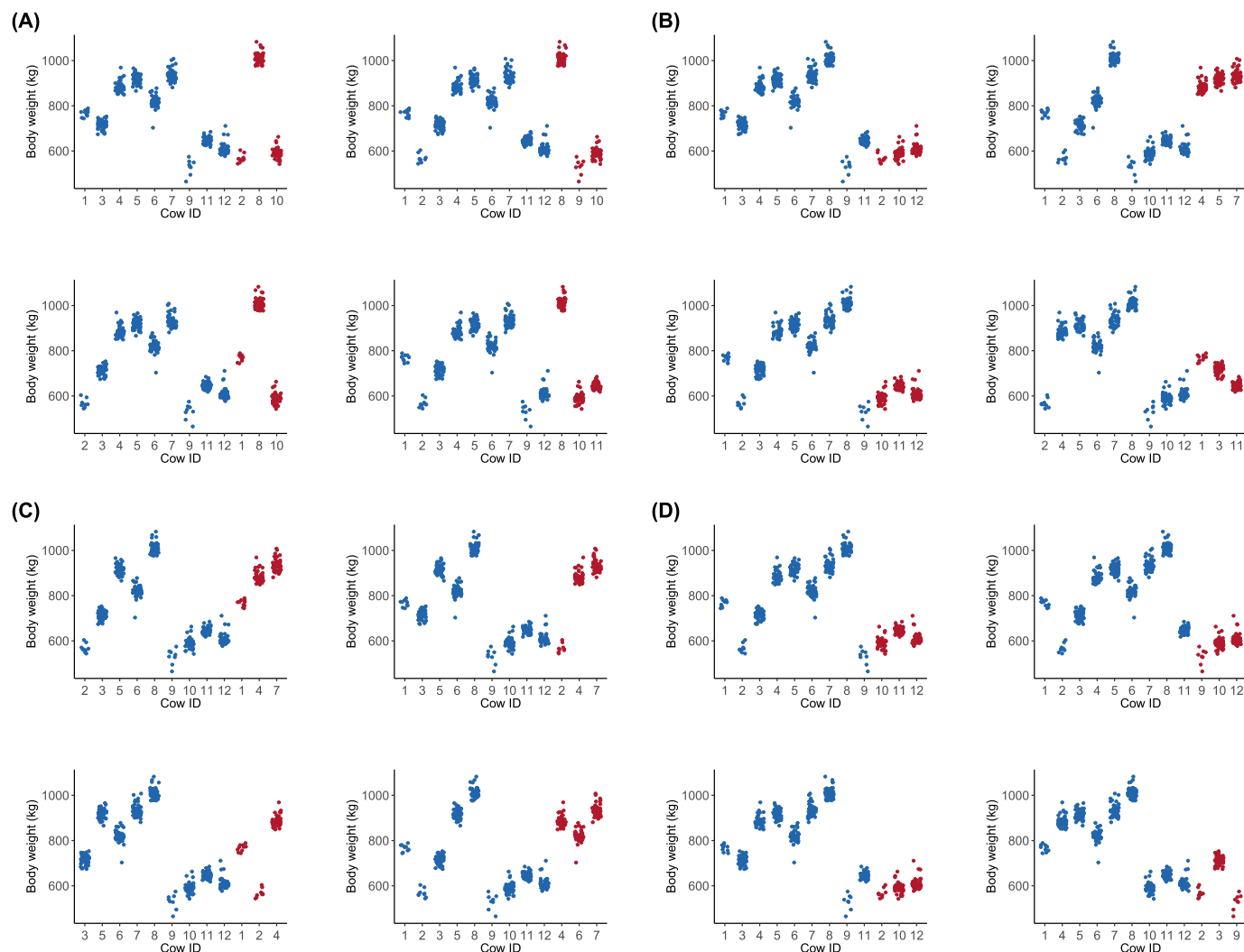


Fig. 11. Range of body weight records observed in training and testing set partitioning for leave-three-cows-out cross-validation using Mask R-CNN and ordinary least squares. A) top 4 predictions based on the coefficient of determination. B) bottom 4 predictions based on coefficient of determination. C) top 4 predictions based on mean absolute percentage error. D) bottom 4 predictions based on mean absolute percentage error.

volume, withers height, hip height, and the additional non-biometric variable age of 107 cows [20]. Additionally, a MAPE of 7.4% was obtained in leave-cow-out cross-validation including the hip width of 30 cows [26] and R^2 of 0.93 and a MAPE of 2.72% was obtained in repeated random sub-sampling validation replicated 100 times including volume, area, backside width, hip width, heart girth, and wither height of 64 cows [13].

5. Conclusions

The current study evaluated model goodness of fit and cross-validation predictability of longitudinal body weight records using automatically extracted image-based biometric features. We found that volume followed by width, correlated best with body weight. The combination of Mask R-CNN and OLS or Mask R-CNN and LMM provided the best predictive performance when the model did not include cow variability or when the model included cow variability, respectively. We believe that the availability of the source code used in the current analyses will promote open science in the animal science community.

CRediT authorship contribution statement

Ye Bi: Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Leticia M. Campos:** Data curation, Investigation,

Writing – review & editing. **Jin Wang:** Methodology, Writing – review & editing. **Haipeng Yu:** Methodology, Writing – review & editing. **Mark D. Hanigan:** Funding acquisition, Writing – review & editing. **Gota Morota:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft.

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

The Python and R code used to perform the video processing and regression analyses are available on GitHub https://github.com/yebigithub/BW_dairy.

Acknowledgements

This work was funded, in part, by the United States Department of Agriculture National Institute of Food and Agriculture, Hatch project 7000564 and the College of Agriculture and Life Sciences Pratt Endow-

ment at Virginia Tech. We thank Tara Pilonero for making the mounting box to store the depth camera.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.atech.2023.100352>.

References

- [1] W. Abdulla, Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow, <https://github.com/matterport/MaskRCNN>, 2017.
- [2] T. Baba, M. Momen, M.T. Campbell, H. Walia, G. Morota, Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping, *PLoS ONE* 15 (2) (2020) e0228.
- [3] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, *J. Stat. Softw.* 67 (1) (2015) 1–48, <https://doi.org/10.18637/jss.v067.i01>.
- [4] R.W. Bello, A.S.A. Mohamed, A.Z. Talib, Enhanced mask r-cnn for herd segmentation, *Int. J. Agric. Biol. Eng.* 14 (4) (2021) 238–244.
- [5] M.F. Hansen, M.L. Smith, L.N. Smith, K.A. Jabbar, D. Forbes, Automated monitoring of dairy cow body condition, mobility and weight using a single 3D video capture device, *Comput. Ind.* 98 (2018) 14–22.
- [6] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.
- [7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [8] A.E. Hoerl, R.W. Kennard, Ridge regression: applications to nonorthogonal problems, *Technometrics* 12 (1) (1970) 69–82.
- [9] Intel RealSense, Intel realsense github, <https://github.com/IntelRealSense>. (Accessed 1 March 2023).
- [10] R. Kadlec, S. Indest, K. Castro, S. Waqar, L.M. Campos, S.T. Amorim, Y. Bi, M.D. Hanigan, G. Morota, Automated acquisition of top-view dairy cow depth image data using an rgb-d sensor camera, *Transl. Anim. Sci.* 6 (4) (2022), txac163.
- [11] M. Kuhn, Caret: classification and regression training, *Astrophysics Source Code Library*, 2015, ascl–1505 pp.
- [12] Y. Kuzuhara, K. Kawamura, R. Yoshitoshi, T. Tamaki, S. Sugai, M. Ikegami, Y. Kurokawa, T. Obitsu, M. Okita, T. Sugino, et al., A preliminary study for predicting body weight and milk properties in lactating Holstein cows using a three-dimensional camera system, *Comput. Electron. Agric.* 111 (2015) 186–193.
- [13] Y. Le Cozler, C. Allain, C. Xavier, L. Depuille, A. Caillot, J. Delouard, L. Delattre, T. Luginbuhl, P. Faverdin, Volume and surface area of Holstein dairy cows calculated from complete 3D shapes acquired using a high-precision scanning system: interest for body weight estimation, *Comput. Electron. Agric.* 165 (2019) 104977.
- [14] Y. Le Cozler, E. Brachet, L. Bourguignon, L. Delattre, T. Luginbuhl, P. Faverdin, Three-dimensional (3D) imaging technology to monitor growth and development of Holstein heifers and estimate body weight, a preliminary study, *Sensors* 22 (12) (2022) 4635.
- [15] T. Lin, M. Maire, S.J. Belongie, L.D. Bourdev, R.B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: common objects in context, *CoRR*, arXiv:1405.0312 [abs], 2014.
- [16] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.
- [17] B. Martins, A. Mendes, L. Silva, T. Moreira, J. Costa, P. Rotta, M. Chizzotti, M. Marcondes, Estimating body weight, body condition score, and type traits in dairy cows using three dimensional cameras and manual body measurements, *Livest. Sci.* 236 (2020) 104054.
- [18] G. Morota, R.V. Ventura, F.F. Silva, M. Koyama, S.C. Fernando, Big data analytics and precision animal agriculture symposium: machine learning and data mining advance predictive big data analysis in precision animal agriculture, *J. Anim. Sci.* 96 (4) (2018) 1540–1550.
- [19] R. Muñoz-Tamayo, B.L. Nielsen, M. Gagaoua, F. Gondret, E.T. Krause, D.P. Morgavi, I.A.S. Olsson, M. Pastell, M. Taghipoor, L. Tedeschi, et al., Seven steps to enhance open science practices in animal science, *PNAS Nexus* 1 (3) (2022) pgac106.
- [20] O. Nir, Y. Parmet, D. Werner, G. Adin, I. Halachmi, 3D computer-vision system for automatically estimating heifer height and body mass, *Biosyst. Eng.* 173 (2018) 4–10.
- [21] T. Norton, D. Berckmans, Developing precision livestock farming tools for precision dairy farming, *Anim. Front.* 7 (1) (2017) 18–23.
- [22] B.A. Nosek, G. Alter, G.C. Banks, D. Borsboom, S.D. Bowman, S.J. Breckler, S. Buck, C.D. Chambers, G. Chin, G. Christensen, et al., Promoting an open research culture, *Science* 348 (6242) (2015) 1422–1425.
- [23] Y. Qiao, M. Truman, S. Sukkarieh, Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming, *Comput. Electron. Agric.* 165 (2019) 104958.
- [24] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *Adv. Neural Inf. Process. Syst.* 28 (2015).
- [25] J. Salau, J. Krieter, Instance segmentation with Mask R-CNN applied to loose-housed dairy cows in a multi-camera setting, *Animals* 10 (12) (2020) 2402.
- [26] X. Song, E. Bokkers, P. Van der Tol, P.G. Koerkamp, S. Van Mourik, Automated body weight prediction of dairy cows using 3-dimensional vision, *J. Dairy Sci.* 101 (5) (2018) 4448–4459.
- [27] S. Tasdemir, A. Urkmez, S. Inal, Determination of body measurements on the Holstein cows using digital image analysis and estimation of live weight with regression analysis, *Comput. Electron. Agric.* 76 (2) (2011) 189–197.
- [28] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc., Ser. B, Methodol.* 58 (1) (1996) 267–288.
- [29] K. Wada, labelme: image polygonal annotation with python, <https://github.com/wkentaro/labelme>, 2018.
- [30] C. Xavier, Y. Le Cozler, L. Depuille, A. Caillot, A. Lebreton, C. Allain, J. Delouard, L. Delattre, T. Luginbuhl, P. Faverdin, et al., The use of 3-dimensional imaging of Holstein cows to estimate body weight and monitor the composition of body weight change throughout lactation, *J. Dairy Sci.* 105 (5) (2022) 4508–4519.
- [31] J. Xiao, G. Liu, K. Wang, Y. Si, Cow identification in free-stall barns based on an improved Mask R-CNN and an SVM, *Comput. Electron. Agric.* 194 (2022) 106738.
- [32] B. Xu, W. Wang, G. Falzon, P. Kwan, L. Guo, G. Chen, A. Tait, D. Schneider, Automated cattle counting using Mask R-CNN in quadcopter vision system, *Comput. Electron. Agric.* 171 (2020) 105300.
- [33] H. Yu, K. Lee, G. Morota, Forecasting dynamic body weight of nonrestrained pigs from images using an rgb-d sensor camera, *Transl. Anim. Sci.* 5 (1) (2021), txab006.