



Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score

Sun Yukun,¹ Huo Pengju,¹ Wang Yujie,¹ Cui Ziqi,¹ Li Yang,¹ Dai Baisheng,^{2,3} Li Runze,² and Zhang Yonggen^{1*}

¹College of Animal Sciences and Technology, Northeast Agriculture University, Harbin 150030, PR China

²College of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, PR China

³Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling 712100, PR China

ABSTRACT

Body condition score (BCS) is a common tool for indirectly estimating the mobilization of energy reserves in the fat and muscle of cattle that meets the requirements of animal welfare and precision livestock farming for the effective monitoring of individual animals. However, previous studies on automatic BCS systems have used manual scoring for data collection, and traditional image extraction methods have limited model performance accuracy. In addition, the radio frequency identification device system commonly used in ranching has the disadvantages of misreadings and damage to bovine bodies. Therefore, the aim of this research was to develop and validate an automatic system for identifying individuals and assessing BCS using a deep learning framework. This work developed a linear regression model of BCS using ultrasound backfat thickness to determine BCS for training sets and tested a system based on convolutional neural networks with 3 channels, including depth, gray, and phase congruency, to analyze the back images of 686 cows. After we performed an analysis of image model performance, online verification was used to evaluate the accuracy and precision of the system. The results showed that the selected linear regression model had a high coefficient of determination value (0.976), and the correlation coefficient between manual BCS and ultrasonic BCS was 0.94. Although the overall accuracy of the BCS estimations was high (0.45, 0.77, and 0.98 within 0, 0.25, and 0.5 unit, respectively), the validation for actual BCS ranging from 3.25 to 3.5 was weak (the F1 scores were only 0.6 and 0.57, respectively, within the 0.25-unit range). Overall, individual identification and BCS assessment performed well in the online measure-

ment, with accuracies of 0.937 and 0.409, respectively. A system for individual identification and BCS assessment was developed, and a convolutional neural network using depth, gray, and phase congruency channels to interpret image features exhibited advantages for monitoring thin cows.

Key words: body condition score, backfat thickness, convolutional neural network, individual identification

INTRODUCTION

Animal welfare is defined on the basis of animal behavior, physiology, clinical state, and performance and it is attracting increasing attention from the government and the public (Støier et al., 2016). The influence of process-based animal welfare on technical efficiency in the dairy industry is neither reliably increasing nor decreasing the percentage of cow losses, replacement rate, age of first calving, milk performance, or SCC, all of which can affect farm efficiency (Allendorf and Wettemann, 2015). Furthermore, along with the development of electronic technology, precision livestock farming, also known as integrated management systems, is persistently concerned with being smart and non-artificial technology for continuous automatic monitoring and improvement of animal health, animal welfare, quality assurance at the farm and chain level, and improved risk analysis and risk management (Wathes et al., 2008). One strategy for high efficiency and precision management relies on the automatic monitoring of livestock, in which related physical processes such as animal growth, the output of milk, some endemic diseases, and aspects of animal behavior are regarded as a complex network in livestock production (Berckmans, 2014). A substantial increase in the application of artificial intelligence on dairy farms is anticipated if electronic information technology is continuously improved and perfected; therefore, new technologies that meet the requirements for both animal welfare and improvement in livestock production have been widely studied and reviewed.

Received December 14, 2018.

Accepted July 8, 2018.

*Corresponding author: zhangyonggen@sina.com

One approach to dairy farming management is the use of BCS to indirectly estimate the mobilization of energy reserves in the fat and muscle of cattle; BCS has been widely applied to monitor feed intake and health status and to optimize management of the lactation curve (Roche et al., 2009; Vasseur et al., 2013). Accurate and repeatable evaluation of BCS requires personal experience in visual and tactile methods based on the shape of the cow at the tailhead and the hooks. However, this procedure is time consuming and subjective, and it can result in erroneous classifications due to the potential for bias by comparison with previously observed cows. Recently, electronic visual systems have attempted to automate the BCS process by using different types of cameras to analyze the cow's contours and have produced a strong correlation between manual observations and technological representations (Fischer et al., 2015; Spoliansky et al., 2016; Alvarez et al., 2018).

Although an automatic BCS system can eliminate observer error, there are still cases where the validation accuracy is low (Alvarez et al., 2018). One reason for poor validation is that the primary way to collect data for the system is through manual scoring. For example, if the score that reflects the actual state of the animal should be 3 points but is labeled as 3.25, there is error in the process of collecting the raw data; in this way, the model may show that the correctly recorded score is inaccurate. Although there is a standard method for assessing BCS (Edmonson et al., 1989), manual assessment inevitably entails subjective error because the process of evaluation is based on visual and tactile impressions; even experienced observers agreed only 58.1% of the time with a modal BCS from 4 observers (Ferguson et al., 1994). Therefore, an automatic BCS system developed from manual evaluation will represent only the opinions of the observers who collected the training data.

Image processing is the key step for training and testing data sets to extract the available features that will be used to automatically assess the BCS. Establishing a predictive BCS model with traditional manual extraction of features is less efficient and less accurate, primarily due to the small number of samples (related to the time-consuming nature of manual evaluation) and the low number of feature points, which results in a poor fit (Fischer et al., 2015). To address these challenges, deep learning based on artificial neural networks has been shown to be a powerful tool for predictive modeling; popular techniques used for analyzing images include convolutional neural networks (CNN). The CNN structure mimics a biological neural network and is readily applied to visual image identification (Singh et al., 2016). In livestock agriculture, CNN have been

widely applied to image classification and identification in areas such as recognition of animal behaviors, monitoring disease status, and identification of individuals (Kamilaris and Prenafeta-Boldú, 2018).

To address the needs of each cow, the system needs to include a function for individual identification. The main method currently in use is radio frequency identification (RFID), but it was found that production of the signal is affected by different electromagnetic environments, reducing the effectiveness of the application (Ruiz-Garcia and Lunadei, 2011). For example, tag power, iron fences, cooling fans, circuit quality, and so on affect the transponder receiving the electromagnetic waves, resulting in a reduction in the reading distance and thus affecting data collection. Image biometrics have advanced in recent years and compensate for the shortcomings of traditional methods (Awad, 2016). However, combining the functions of body condition scoring and individual recognition in a single camera has not yet been accomplished. This research focused on the development and validation of an automated system for individual identification and BCS assessment, with the goal of objective evaluation using ultrasonic measurements and the creation of an algorithm to generate a predictive model using CNN to accurately assess BCS.

MATERIALS AND METHODS

Animal Management and Routine Care

The experiment was performed under an experimental license from Northeast Agricultural University, Harbin, China. All experimental procedures were conducted in accordance with the principles and responsibilities outlined in the university's guidelines for animal research. The experiment was performed at a commercial farm in Harbin, China, from March to August 2018. Multiparous lactating Holstein-Friesian dairy cows (~1,000 dairy cows) were grouped according to lactation stage (early or late lactation), and the herds were divided into high yield and low yield based on the threshold of 25 kg of milk/d. Mean parity was 2.36 (± 0.48 ; \pm SD), with a range of 2 to 3. The cows were housed in freestall barns with poudrette bedding that were fitted with self-locking neck clips at each feed line. There were also automatic sprinklers, ventilation systems, and natural light, although artificial lighting was provided all day. A TMR mainly comprising corn and grass silage was supplied in 2 equal portions after the milking process was completed using a TMR agitating lorry that allowed ad libitum food; water was freely available at all times. The cows were monitored for feed intake, which was

approximately 43 kg/d. Milk production for cows that had calved during the study period averaged (\pm SD) 32.67 ± 6.39 kg, with a range of 15.2 to 43.1. The cows were milked twice per day (0730 to 1030 h and 1430 to 1700 h) in a 40-stand DeLaval (Tumba, Sweden) rotary milking parlor. Before milking, the cows were brought to the waiting area and then entered the milking parlor one at a time. After milking, the cows were returned to livestock housing and were fed a TMR that allowed them to replenish their energy.

Acquisition of BCS and Individual Information

The acquisition of images of the cows' backs from a dolly allowed the collection of image data to model the BCS estimation and individual identification. Manual BCS had low repeatability and inconsistent standards due to the different perceptions of the observers. An objective method, therefore, was used to obtain the BCS required for developing the model. First, the manual BCS and backfat thickness (**BFT**) were measured to generate the regression model to acquire the objective BCS. Images were then collected to analyze cattle back characteristics. These different methods of data acquisition are explained in more detail below (Figure 1).

Manual BCS scoring was evaluated simultaneously by 2 experienced observers using a 5-point scale with 0.25-point increments, where 1 represents thin and 5 represents fat. The observers first scored from a flank perspective that evaluated the visibility of the rib and the shape from the hip bone to the ischial tuberosity ($\leq 3.25 = V$; 3.25 to $4 = U$; $\geq 4 =$ straight line). The observers were then positioned behind the cows to check the height of the sacral ligament and evaluate the tailhead, from which the final score was determined by touch.

Simultaneously, ultrasound measurements of BFT were conducted by an experienced ultrasonographer using a portable real-time ultrasound (BW-2017, Beishiwu Technology Ltd., Shenzhen, Guangdong Province, China), which corrected the manual score to avoid multiple-score instability. The maximum penetration depth of the scanner was 10 cm with a 7.5-MHz probe, which produced almost instantaneous images with an adjustable digital caliper. The transducer was positioned vertical to the sacral region between the caudal one-quarter and one-fifth line connecting the dorsal portion of the tuber ischia (pins) to the tuber coxae (hooks); the right side was used in this study. There were 3 clear dividing lines in the ultrasonic frozen images, including the skin, superficial fascia, and deep fascia, and the data standard was based on the distance

between the superficial fascia and the deep fascia. Finally, the ultrasound BCS developed from ultrasound BFT was treated as the actual BCS.

A dolly with a computer and a series of fixed aluminum stents was positioned to create a database of the image data from the cows' backs (Figure 2). The camera (Intel RealSense D435, Integrated Electronics Corporation, Santa Clara, CA) was placed 3.2 m above the concrete platform and fixed on an aluminum tube that was connected to the computer by a 3-m-long line. The images were captured while the cows were locked in by neck clips. The technician pushed the dolly behind the cows to capture the images automatically, one cow at a time. Meanwhile, 2 experienced observers recorded the BCS in a video of the cows, and the average score was treated as the final BCS for the training model; ear tags were recorded at the same time. A 30-s video clip of the 3-dimensional (**3D**) computer image from the dorsal view of each cow was made that contained between 150 and 180 still frames. It took 1 mo to collect the images 5 times, and the system was installed and adjusted between May and August.

Depth Image Processing and Model Training

The data set comprised 3,430 images (5 images/cow) and was divided into a training set of 2,400 images (480 cows) and a test set of 1,030 images (206 cows) based on the 7:3 principle of data splitting. The distribution of BCS in the training and test sets is presented in Table 1. In this system, there were 3 types of training models: the automatic detection model, the bovine recognition model for the cow's back, and the BCS evaluation model. All the models developed in this experiment were based on a deep learning method and were constructed to apply a CNN network structure of multifunctional image identification.

Image preprocessing was conducted before generating the training model. Three channels of data were taken from the cow images, including depth, gray, and phase congruency, to extract the effective features for the next phase. Initially, the color images of the cows' backs were converted to gray channel. Then, depth images were reserved as depth channel. Due to uneven illumination and object surface properties, the depth images collected by the camera usually contained holes in both the smooth and edge regions of the cow. To avoid the distortions and artifacts introduced by these holes, a depth map enhancement operation (Shen et al., 2014) using the default parameters was applied to all images during training and testing. Ultimately, each depth image obtained was transformed by generating

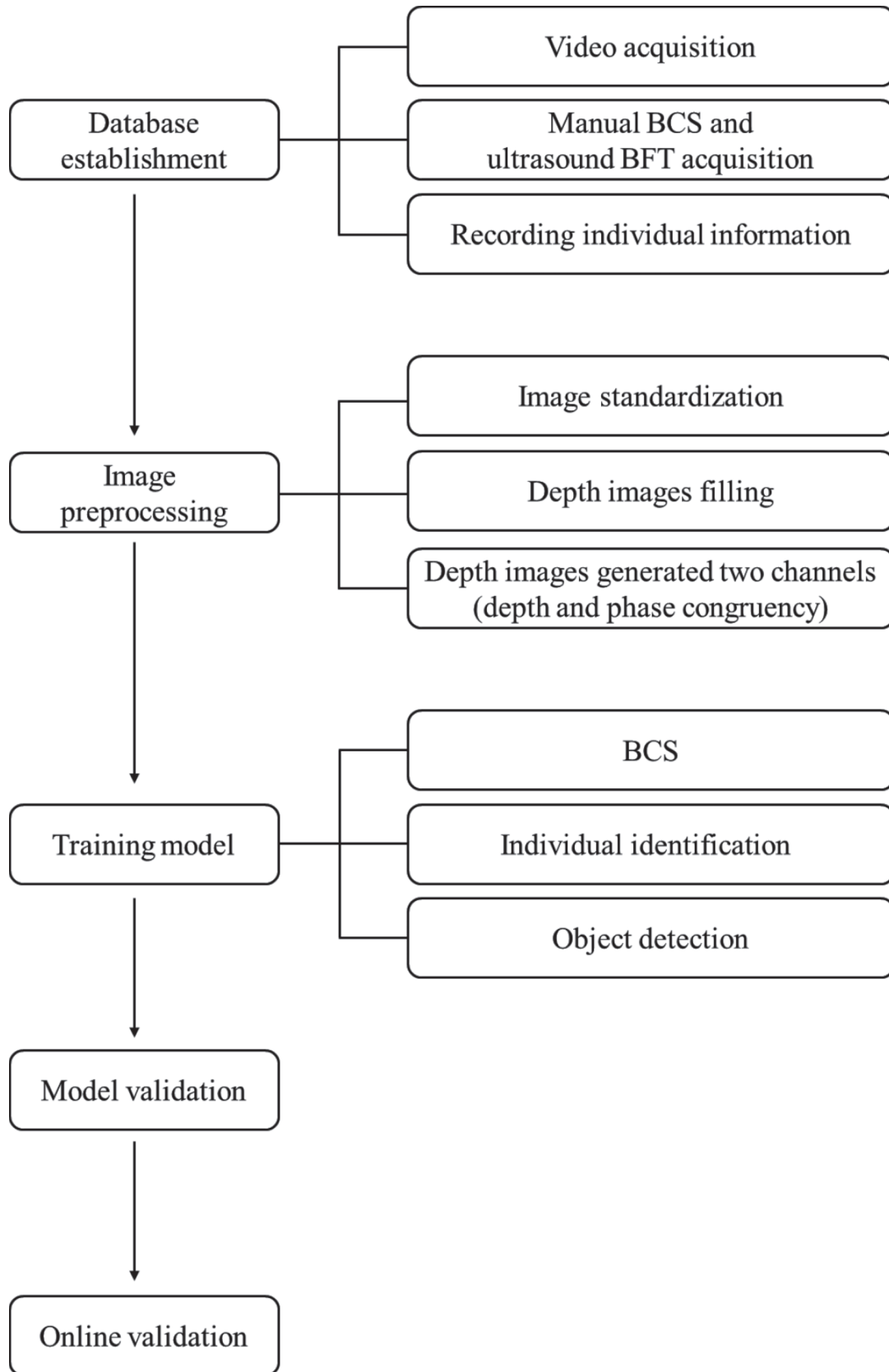


Figure 1. Workflow chart of image processing and data validation in this study. BFT = backfat thickness.



Figure 2. Image acquisition.

an additional channel, and phase congruency was used to preserve additional edge information within the frequency domain of the depth image.

Construction of the model was completed through the CNN, which could automate the processing of large numbers of training images more efficiently than traditional manual methods. To achieve an automatic system of cow BCS assessment and individual identification, there were 3 modules—cow object detection, individual identification, and evaluation for BCS—in the proposed framework. For object detection and individual identification, based on findings of our previous research (Shen et al., 2019), YOLO and AlexNet (Krizhevsky et al., 2012; Redmon et al., 2016) have proven to be effective and were specifically selected for this work. Models of cow object detection and identification were trained herein with cow back-view images instead of cow side-view images. In addition, unlike the background subtraction applied by Alvarez et al. (2018), the YOLO object detector is more suitable for locating each cow in one crowded scene in our investigations. With respect to cow BCS assessment, the biggest problem in our study was the overfitting of model training due to the small scale of cow images. To reduce the effect of overfitting, we needed a shallow (fewer parameters) but powerful model for BCS classification. Therefore, DenseNet (Huang et al., 2017) was specifically considered in this work. Its dense connection scheme not only fused the information from both the bottom and top layers to provide more discriminable features but also reduced the number of parameters, which can be regarded as a regularization effect to avoid overfitting, as described in the original work with DenseNet (Huang et al., 2017).

Overfitting of image data occurs during the model training phase, and it is necessary to randomly rotate and flip each image during image preprocessing to re-

duce this phenomenon. In addition, data augmentation techniques were applied by shifting each image left and right to increase the variation of image samples. In the process of constructing the neural network, the CNN mainly consists of 3 types of layers: convolution (to generate the feature map to locate specific simple features), pooling (to reduce the feature mapping dimensions), and full connection (to classify the characteristics and the connections of nodes within each layer). In the CNN training process, the convolution layer (model parameters) and pooling were used to extract features, and classification results were automatically determined by the full connection. In this case, the CNN transformed the preprocessed images layer by layer to extract features to finally determine each classification score, and the obtained feature parameters were used to update the prediction model function to constantly improve the accuracy of identification.

To adapt the CNN model to BCS classification, some modifications were performed in our investigation. First, given that different regions of a cow's back make different contributions to BCS classification and that different channels of the feature map in the CNN indicate the differential importance of the information at different spatial locations, a visual attention mechanism, the squeeze-and-excitation module, was introduced to the model (i.e., DenseNet) to exploit the importance of different regions of a cow's back by enhancing the contribution of important channels of the feature map while suppressing the contribution of trivial channels. Second, to reduce the total number of model parameters and avoid overfitting during model training, we used a relatively shallow DenseNet and set a smaller depth ($L = 24$) and growth rate ($k = 12$) within the DenseNet-BC architecture. In addition, we used the SGD algorithm to update model parameters during the training stage with a batch size of 8, a learning rate of 0.01, and an epoch number of 300.

Table 1. Distribution of the number of images of different BCS in the training and test sets

BCS ¹	Training	Test
2.00	95	40
2.25	235	100
2.50	465	200
2.75	560	240
3.00	465	200
3.25	325	140
3.50	140	60
3.75	45	20
4.00	70	30

¹Five-point scale with 0.25-point increments, where 1 represents thin and 5 represents fat.

Statistical Analysis

Model Performance Evaluation and Validation. Validation of the performance of the model was conducted to confirm bovine identity, the automatic detection of cows, and their BCS assessments. A confusion matrix for evaluating the model of BCS was constructed to compare the predicted and actual values of the BCS in each tuple of different classes, showing a detailed breakdown of correct and incorrect classifications for each class. The main diagonal of the confusion matrix shows the number of observations that were correctly classified, whereas the off-diagonal elements indicate the number of observations that were incorrectly classified. The information from the confusion matrix was used to derive the following measurements.

- **Accuracy:** The proportion of samples that are classified correctly out of all the samples. Calculated as $\text{accuracy} = (\text{true positive} + \text{true negative}) / \text{all events}$.
- **Precision:** The proportion of correctly identified positive samples out of all correctly classified samples. Calculated as $\text{precision} = \text{true positive} / (\text{true positive} + \text{false positive})$.
- **Recall:** The proportion of correctly classified positive samples out of all positive samples. Calculated as $\text{recall} = \text{true positive} / (\text{true positive} + \text{false negative})$.
- **F₁ score:** The harmonic mean of precision and recall. Calculated as
$$\text{F}_1 \text{ score} = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}.$$

Online Estimation. Online automatic warning systems were used to evaluate performance, including accuracy and precision. Every online predicted datum collected was compared with the actual value, which included the transformed BCS from the ultrasonic measurement and the ear tag number of the cow. The calculated accuracy was consistent with the accuracy description of the above model.

The 2 main factors affecting the precision of the online measurement system are repeatability and reproducibility. Repeatability is estimated based on the dispersion of results obtained from 6 measurements of the same cow BCS and identified by the same camera at the same time. After the cattle enter the video and their identities are confirmed, 6 frames of images can be analyzed within 1 s. Reproducibility is estimated by scanning the cows 6 times for 3 consecutive days to assess the error in the average value of repeated measurements of BCS and recognition (ignoring changes in body condition over the 3 d). Repeatability or reproducibility are typically equal to

$$\sigma_{\text{repeatability/reproducibility}} = \frac{\bar{R}}{d_2},$$

where $\bar{R}_{\text{repeatability}}$ is the average of the range differences between 2 measurements on an individual cow on the same day and $\bar{R}_{\text{reproducibility}}$ is the range of the means of different conditions—that is, $\bar{R}_{\text{repeatability}} = |x_{\text{max}} - x_{\text{min}}|$ and $\bar{R}_{\text{reproducibility}} = |\bar{x}_{\text{max}} - \bar{x}_{\text{min}}|$. The parameter $d_{2\text{repeatability}}$ is based on the sample size; for the sample number when the calculating range is 1, $d_{2\text{repeatability}} = 2.534$ ($m = 6$, $g > 15$) and $d_{2\text{reproducibility}} = 2.67$ ($m = 6$, $g = 1$), where m is the number of ranges calculated and g is the number of calculations of the mean of the range.

Online precision was calculated using the following formula:

$$\sigma_{\text{system}} = \sqrt{\sigma_{\text{repeatability}}^2 + \sigma_{\text{reproducibility}}^2}.$$

RESULTS

Individual Identification

The automatic detection function begins when the caudal end of a cow is completely within the lens, and the online test starts. In this process, the accuracy of the automatic detection when a cow passes is 100%; the images of 2 cows adjacent to one another (with the cow's posterior completely visible or a following cow's head hiding the tailhead area) occurred in 100% of selected videos, whereas no detection results occurred when the backs or tails of the 2 cows were obscured in the images. The accuracy of the individual recognition model reached 0.985 after 50 iterations, but the accuracy was decreased by 4.9% (to 0.937) in the online test.

Acquisition of Ultrasound BCS

The linear regression between the manual BCS and the ultrasound BFT had a coefficient of determination (R^2) of 0.903 (Figure 3). The initial BCS calculated from the linear regression was a continuous rather than a categorical variable, and therefore the original ultrasound BCS was converted through the CEILING function in Excel (Microsoft Corp., Redmond, WA) to a scale with units of 0.125. Because the CEILING function is a rule of progression, the BCS value was taken to be a value less by 1 unit of 0.125 when there was a value other than a multiple of 0.25 unit; for example, an ultrasound BCS of 2.625 was assigned a score of 2.5. The resulting R^2 between the manual BCS and ultrasound BCS was 0.882, as shown in Figure 4. Figure 5

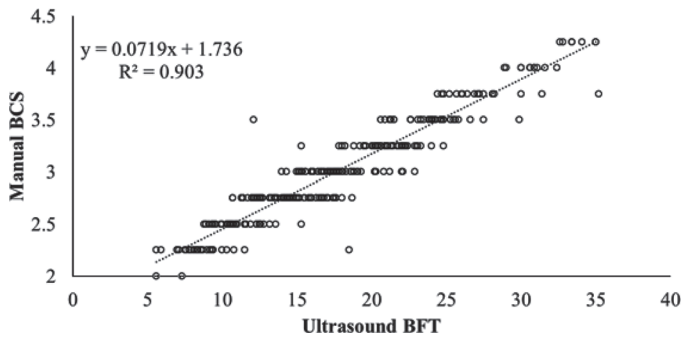


Figure 3. The link between manual BCS and ultrasound backfat thickness (BFT). BCS scale = 5-point scale with 0.25-point increments, where 1 represents thin and 5 represents fat.

shows the linear regression between ultrasound BFT and ultrasound BCS with an R^2 of 0.976; the regression formula was as follows:

$$\text{ultrasound BCS} = 0.0723 \times \text{ultrasound BFT} + 1.7327.$$

Model Performance

The details of the distribution of differences between the model prediction of BCS and actual BCS are shown in Table 2. A close ideal tendency in the images and the information associated with the change in BCS was performed within 0.5 range error along the diagonal of the confusion matrix. The different pictures of each cow did not necessarily show the same result, but the average accuracy of the model within 0, 0.25, and 0.5 error ranges was 0.45, 0.77, and 0.98, respectively. The weakness of the model performance is that the outputs of the predictive BCS were more dispersed than the actual BCS from 3.25 to 3.5 compared with below 2.75; this range (3.25–3.5) accounts for only 19.4% of total test set samples.

The classification measures show the performance of the model within different error ranges (Table 3). The performance parameter for when the real BCS was less than or equal to 3 points within the absolute accuracy error range was significantly higher than for when the real BCS was greater than 3 points. As the error range increased, the F1 score (the harmonic mean of precision and recall) narrowed the gap between thin (BCS ≤ 3) and fat (BCS > 3) cows, with average differences of 0.26, 0.2, and 0.1 between the fat and thin cows within the error ranges of 0, 0.25, and 0.5, respectively. However, the precision values within the 0.5 error range were not significantly higher because the number of samples classified as BCS = 4 was small, limiting the F1 score within the range of 0.5 point.

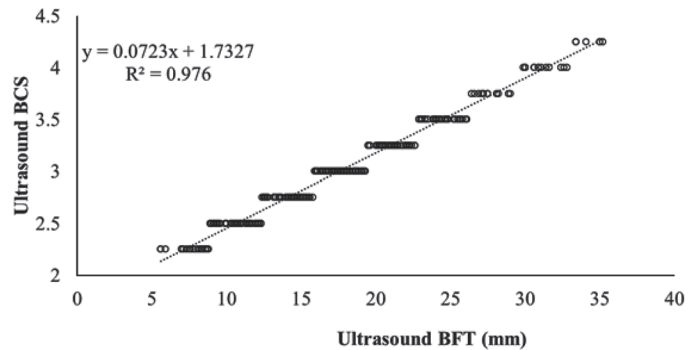


Figure 5. The link between ultrasound BCS and ultrasound backfat thickness (BFT). For example, if BFT is 16 mm, BCS will be 2.89 after calculation by regression formula, and the final BCS will be 3 after taking 0.125 as the unit according to the CEILING function in Excel (Microsoft Corp., Redmond, WA). BCS scale = 5-point scale with 0.25-point increments, where 1 represents thin and 5 represents fat.

Online Automatic Measurements

This step described the performance of the system in the actual application within the error associated with the acquisition of each image of a cow's back. The repeatability and reproducibility were calculated from the range difference, such that a value closer to zero indicated better performance of the system. The repeatability was 0.065, and the reproducibility was 0.019. Further analysis showed that the system precision was 0.068; the system accuracy represents the proportion of all the images correctly classified, and the accuracy of the image samples obtained in the online process was 0.409.

DISCUSSION

Although the accuracy of the model in evaluating BCS ranging from 3.25 to 3 was low, we found that

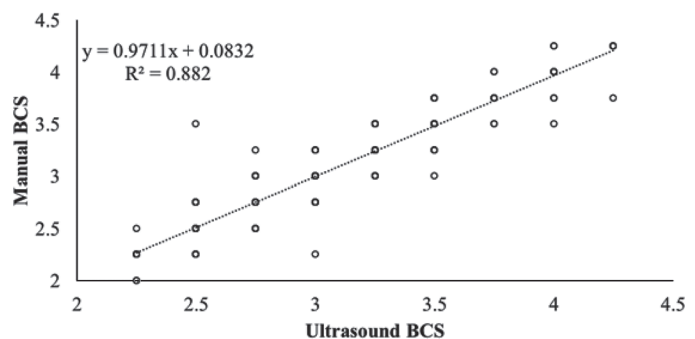


Figure 4. Scatter plot of manual BCS and ultrasound BCS. BCS scale = 5-point scale with 0.25-point increments, where 1 represents thin and 5 represents fat.

Table 2. Confusion matrix of test set sample classification

Real BCS ¹	Test predictive BCS								
	2	2.25	2.5	2.75	3	3.25	3.5	3.75	4
2	25	6	4	5	0	0	0	0	0
2.25	12	52	32	4	0	0	0	0	0
2.5	0	38	77	23	15	30	17	0	0
2.75	0	13	19	149	18	41	0	0	0
3	0	0	20	42	89	27	22	0	0
3.25	0	0	1	23	50	35	24	2	5
3.5	0	0	0	0	20	15	18	6	1
3.75	0	0	0	0	0	6	3	10	1
4	0	0	0	0	0	4	6	15	5

¹Five-point scale with 0.25-point increments, where 1 represents thin and 5 represents fat.

the model achieved an overall accuracy level similar to recent related efforts with a repeatability of 0.065 and a reproducibility of 0.019. In another study using repeatability (extracting the same landmarks 8 times on the same day from the same 3D scan) and reproducibility (scanned 8 times on the same day with the same landmarks extracted once per 3D surface) to evaluate BCS, the results were 0.075 and 0.103, respectively (Fischer et al., 2015). Additionally, continuous improvement in measures of performance for accuracy and automation has been reported for the prediction of BCS in previous similar studies. Manual marking of 23 anatomical features around the rear contour of each animal to assess BCS has resulted in estimations with an accuracy of 92.79%, within 0.25 unit of difference between the true and predicted BCS values (Bewley et al., 2008). As with this method, Azzaro et al. (2011) used shape descriptors based on principal component analysis to build a robust descriptor for automatic BCS estimation, and the error was 0.31. Some reports suggested that accuracy could be improved by changing the type of camera. For example, Halachmi et al. (2013) obtained a Pearson correlation coefficient of 0.94 by using a thermal imaging camera to improve the accuracy in extracting cow body shape from its background. Sev-

eral attempts with iterative updates of hardware have been published to optimize the process of automatic measurement of BCS, mainly using principal component analysis ($R = 0.98$, $RMSE = 0.31$; Fischer et al., 2015) and Fourier descriptors ($R = 0.64$, accuracy = 43% within 0.25 unit; Bercovich et al., 2013) to analyze the contour and shape of cows. Recent observations using a low-cost camera (Spoliansky et al., 2016) and an image analysis model for CNN (Alvarez et al., 2018) have achieved ideal overall accuracy (71.35 and 78%, respectively, within the 0.25 range). These studies suggest that both methods of data collection and methods of extracting features for image analysis can influence the accuracy of predicting cow BCS.

Application of a New Individual Identification Method

The results of this study showed that 2 automatic functions, individual identification and BCS assessment, could be simultaneously achieved by one camera. The identification accuracy during model validation (98.5%) was similar to that observed by Li et al. (2017) but decreased in the online test (93.7%). One reason for this decrease discussed by Li et al. (2017) was that in their experimental scenario, the white pattern was

Table 3. Classification measurement within different range errors

BCS ¹	0 range error			0.25 range error			0.5 range error		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
2.00	0.68	0.63	0.65	1.00	0.78	0.88	1.00	0.88	0.94
2.25	0.48	0.52	0.50	0.88	0.96	0.92	1.00	1.00	1.00
2.50	0.50	0.39	0.44	0.84	0.69	0.76	0.99	0.77	0.87
2.75	0.61	0.62	0.61	0.87	0.78	0.82	0.98	1.00	0.99
3.00	0.46	0.45	0.45	0.82	0.79	0.80	1.00	1.00	1.00
3.25	0.22	0.25	0.23	0.49	0.78	0.60	0.78	0.96	0.86
3.50	0.20	0.30	0.24	0.50	0.65	0.57	0.81	1.00	0.90
3.75	0.30	0.50	0.38	0.94	0.70	0.80	1.00	1.00	1.00
4.00	0.42	0.17	0.24	0.50	0.67	0.57	0.58	0.87	0.70
Average	0.43	0.42	0.42	0.76	0.76	0.75	0.90	0.94	0.92

¹Five-point scale with 0.25-point increments, where 1 represents thin and 5 represents fat.

compressed when it was rotated, resulting in image downsampling. Identification is based on the analysis of the characteristic shape of the black and white contour on the back of a cow. Therefore, when an animal is in an active state, the white portion is squeezed, which deforms the feature contour and thus causes errors in identification. However, although we developed an automatic detection algorithm for the back of a cow such that 2 cows could be tracked simultaneously, 2 targets cannot be identified in the same image in our method. On a farm, there are often situations in which 2 cows are next to each other in the same image. Accordingly, it might be possible to lose information of animals that pass through the camera due to crowding at the exit.

We assumed that automatic individual identification of cows using backside images was capable of replacing RFID-based systems for animal identification that are commonly used on farms. The main disadvantage of RFID registrations is that RFID modules require labor to be configured as identification, and a common way to identify external tags requires the installation of ear tagging with permanent damage to animals (Awad, 2016). Meanwhile, the costs of purchasing a system consisting of an RFID reader, tags, and a management server and of replacing RFID tags are expensive. By contrast, we used the biological characteristics of animal body parts to identify individuals without additional equipment placed on the body and at a lower cost than an RFID system. However, the actual performance of backside image identification was not better than RFID in this study. This shortcoming is mainly due to the ability of the RFID system to read up to 1,000 tags/s with a reading error of less than 0.1% (Rathgeb and Uhl, 2011), whereas this image identification system can record only 9 frames/s, and the identification error reached 6.3%. As a result, in the future, cameras with higher frame rates should be used to increase the chance that the back pattern is in its natural state in an image and ultimately improve the accuracy of identification.

Objective Acquisition Method

To solve the problem of subjective errors in the assessment of energy reserves, ultrasound measurements for BFT were used in the data acquisition process of this study because of their uniqueness and convenience. Compared with a manual BCS, using a linear regression model to determine true BCS increased the reliability of BCS assessment for the same body fat content, as shown in Figures 3 and 5. The use of ultrasound to calibrate 3D surfaces in a more precise and repeatable manner should reduce the error in measurements of body energy reserve (Weber et al., 2014) because the change in BFT represents an augmentation or loss of body fat energy;

this is because body energy reserves are predominantly stored in the adipose tissue (Schröder and Staufenbiel, 2006). Another study determined that the correlation coefficients between BCS and BFT range from 0.82 to 0.93 (Ayres et al., 2009). Similarly, a previous study that developed a 3D optical system to objectively assess the fat thickness in the back of cows suggested that 3D imaging was highly sensitive to changes in body fat (Weber et al., 2014). The correlation coefficient between the estimated value obtained by the algorithm and the actual value obtained by the ultrasound measurement was 0.96. Therefore, it is feasible to use the ultrasound of BFT to accurately characterize BCS and reflect energy storage in training sets.

Halachmi et al. (2008), however, found that an ultrasound operator can influence the results due to variability in measurement positions and different interpretation of the ultrasound picture. An experiment by Schröder and Staufenbiel (2006) suggested that the measured position falls on a point in the same direction. Moreover, the BFT is theoretically constant no matter who measures it. Although some reports have shown no significant differences in measurements on the left and right sides, only one side was chosen for consistency (Hussein et al., 2013). As consequence, ultrasound measurements can reduce alarm errors caused by data recording errors. However, the technical progress prompt to make the ultrasonic measurement has a clear scale to address the error resulting from operator variability. The key to transforming the BCS with ultrasound measurements of BFT was the choice of regression models. Although Hussein et al. (2013) obtained a linear regression with a higher correlation coefficient between BCS and BFT for separate lactation periods, under the assumption that differences in genetic background and energy levels associated with feeding would influence the rules of BCS in different lactation periods (Roche et al., 2009), a significant difference in the linear regression between BCS and BFT between lactation stages was not found in our previous studies. Thus, we calculated a new regression model instead of selecting from a series of previous models with higher correlation coefficients between manual BCS and BFT (R^2 between 0.924 and 0.963; Hussein et al., 2013).

Improvement in the Accuracy of the Model Using CNN

The difference between the findings of a previous study (Alvarez et al., 2018) and the present results can be attributed to the use of DenseNet and the types of image channels used for feature extraction before feeding an image to the network. DenseNet was used

to implement the CNN model with fewer parameters than the Visual Geometry Group used by Alvarez et al. (2019). The lightweight deep models (e.g., SqueezeNet and MobileNet; Iandola et al., 2016; Zhang et al., 2017) have great potential for most practical applications, but due to the small scale of our data, these lightweight models did not achieve expected performance in our preliminary trials. SqueezeNet with 9 fire modules and SqueezeNet with 7 fire modules are both difficult to converge in the training phase with our data. Theoretically, compared with small CNN architectures such as SqueezeNet, DenseNet has a convolutional layer with a short connection layer that is closer to the network input, which is more attentive to details such as the edges and corners of extracted images; however, the features extracted from the layer close to the network output are more abstract and tend to have high-level semantic information. The final classification of this dense connection is based on a decision using multilevel feature information. In other words, the individual DenseNet framework tends to extract information on the edge contour of a cow to improve the classification performance of low-scoring cows with more angular features.

In this work, inspired by the work of Alvarez et al. (2018), 3 types of images were specifically selected to generate the input for the CNN model to classify cow back images with different BCS. First, similar to the work of Alvarez et al. (2018), a depth image of a cow back was used to capture the fluctuations in the surface of the cow's back. The visual characteristics seen in the evaluations above 3.25 were relatively subtle, and it was necessary to add the tactile evaluation of fat thickness to the subjective evaluation. Compared with 2D, 3D is capable of depicting the most complete information available to analyze a shape's variability (Fischer et al., 2015). Second, important confirmation was the observation of Alvarez et al. (2018) that the edge information from the depth image could capture the contour information from a cow's back. In our experiment, phase congruency was used to replace edge channel because some cows appeared to be very close to each other during the image capture procedure in our experiments (Figure 2), the boundaries of those cows in the depth images were too weak (low contrast) to be extracted, and phase congruency was an illumination-invariant image and thus was more sensitive to such weak boundaries. Third, to improve the classification performance of the cow back images of BCS within the range of 2 and 2.75, a gray image generated from each color image was used to provide supplemental information regarding the visual information of the cow back. This aspect of information is more suitable for the feature content observed by the human eye; in particular, the angular information at the center of the back is more sensi-

tive, where 2.75 to 2 points visually corresponds to the number of sacral edges and ribs and the protrusion of the spine. The inverse Fourier-transformed image used in the work of Alvarez et al. (2018) was not adopted in our study because it did not improve the classification performance in our preliminary trials; this point was also demonstrated in the work of Alvarez et al. (2019). In addition, the combination in our investigation of 3 channels, including depth, phase congruency, and gray, achieved superior performance compared with other combinations and with the combinations in the work of Alvarez et al. (2018), although the improvement was not remarkable. A possible reason for the performance degradation of the combination (depth, edge, and Fourier transform) is the complexity of the data collected in our experiments (e.g., cows adjacent to one another; Figure 2). Similar to the results in previous work, high model accuracy was achieved in this study, where it was more important to emphasize how to extract information more effectively to improve the results of the model and less attention was paid to the actual change in energy storage. This model was more effective in monitoring the animals with lower body condition; specifically, we observed that the metabolic disorder caused by a negative energy balance caused a rapid decline in body condition.

One of the reasons why using CNN to analyze the images produced a highly accurate model was the removal of subjectivity through a completely automated system of image processing. Subjective marking errors would directly lead to a bias in the model training process. Compared with traditional methods to calculate sample regression formulae, the CNN were treated as complicated functions with the continual addition of new characteristic values based on the repeated extraction of anatomical landmarks and automatic modification of model parameters to improve the accuracy of target identification (Song et al., 2017; Kamilaris and Prenafeta-Boldú, 2018). In this respect, the process of non-artificial feature extraction was considered a completely objective method to eliminate the resulting repeatability error. Past studies required the manual extraction of characteristics to develop the estimation models, which were the basis of the analysis of the rear contour used to identify the feature points of image interpretation. Fischer et al. (2015) spent an average of 3 min per 3D surface; in comparison, our processing time was approximately 5 s per image to box out the back of the dairy cows. Hence, our method solved the issue of time and labor for high-throughput sample manipulation in situ.

An interesting report by Bewley et al. (2008) showed better accuracy (92.79% with 0.25 unit) than our proposed method (75% with 0.25 unit). The features based

on anatomical points played an important role in the assessment of BCS, which was indirectly validated by the importance of edge information indicated in the work of Alvarez et al. (2018). Bewley et al. (2008) extracted more accurate contour features through 23 anatomical points identified manually. Although the proposed method used the phase congruency to enhance low-contrast edges of the object to be fed into the CNN model in our study, some edges could not be extracted or some noise edges were not introduced by the algorithm. In addition, according to the work of Bewley et al. (2008), the cow skeleton became more prominent at larger parity. Therefore, more robust contour features can be extracted from the experimental cows in the study of Bewley et al. (2008), considering that the mean parity (2.46) of their experimental cows was larger than that (2.36) of our experimental cows. However, the main advantage of our approach compared with previous work was to improve training model efficiency through the application of CNN to automatically extract features. Moreover, using BFT to train the model for BCS estimation was effective for training the model with a highly repeatable measure.

CONCLUSIONS

In this study, a system using a single camera achieved simultaneous automatic assessment of BCS and individual identification. The finding of improved model performance for thin cows (BCS <3) with the addition of phase congruency and gray channels in a CNN suggests that a minor improvement in the average accuracy may be achievable within the absolute accuracy error range. Importantly, the method of obtaining an ultrasound BCS from an ultrasound BFT could mitigate the errors in the assessment of energy reserves by manual BCS. Consequently, this system could help production decision-makers reduce negative energy balance in early lactation through the accurate observation of individuals experiencing rapid declines in body condition.

ACKNOWLEDGMENTS

This study was financially supported by the China Agriculture Research System (Beijing, China, CARS-36), the Young Talents Project of the Northeast Agricultural University (Harbin, China, 18QC35), the China Postdoctoral Science Foundation Grant for the 65nd batch (Beijing, China, 2019M651252), and the Postdoctoral Foundation in Heilongjiang Province (Harbin, China, LBH-Z17035), the Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs under a grant (Beijing, China,

2018AIOT-02). The authors also thank the Nestle dairy cow training center (Harbin, China) for its support.

REFERENCES

- Allendorf, J. J., and P. J. C. Wettemann. 2015. Does animal welfare influence dairy farm efficiency? A two-stage approach. *J. Dairy Sci.* 98:7730–7740.
- Alvarez, J. R., M. Arroqui, P. Mangudo, J. Toloza, D. Jatip, J. M. Rodríguez, A. Teyseyre, C. Sanz, A. Zunino, and C. Machado. 2018. Body condition estimation on cows from depth images using convolutional neural networks. *Comput. Electron. Agric.* 155:12–22.
- Alvarez, J. R., M. Arroqui, P. Mangudo, J. Toloza, D. Jatip, J. M. Rodríguez, A. Teyseyre, C. Sanz, A. Zunino, and C. Machado. 2019. Estimating body condition score in dairy cows from depth images using convolutional neural networks, transfer learning and model ensembling techniques. *Agronomy (Basel)* 9:90.
- Awad, A. I. 2016. From classical methods to animal biometrics: A review on cattle identification and tracking. *Comput. Electron. Agric.* 123:423–435.
- Ayres, H., R. M. Ferreira, and P. S. Baruselli. 2009. Validation of body condition score as a predictor of subcutaneous fat in Nelore (*Bos indicus*) cows. *Livest. Sci.* 123:175–179.
- Azzaro, G., M. Caccamo, J. D. Ferguson, S. Battiato, G. M. Farinella, G. C. Guarnera, G. Puglisi, R. Petriglieri, and G. Licitra. 2011. Objective estimation of body condition score by modeling cow body shape from digital images. *J. Dairy Sci.* 94:2126–2137.
- Berckmans, D. 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. *Rev. Sci. Tech.* 33:189–196.
- Bercovich, A., Y. Edan, V. Alchanatis, U. Moallem, Y. Parmet, H. Honig, E. Maltz, A. Antler, and I. Halachmi. 2013. Development of an automatic cow body condition scoring using body shape signature and Fourier descriptors. *J. Dairy Sci.* 96:8047–8059.
- Bewley, J. M., A. M. Peacock, O. Lewis, R. E. Boyce, D. J. Roberts, M. P. Coffey, S. J. Kenyon, and M. M. Schutz. 2008. Potential for estimation of body condition scores in dairy cattle from digital images. *J. Dairy Sci.* 91:3439–3453.
- Edmonson, A. J., I. J. Lean, L. D. Weaver, T. Farver, and G. Webster. 1989. A body condition scoring chart for Holstein dairy cows. *J. Dairy Sci.* 72:68–78.
- Ferguson, J. D., D. T. Galligan, and N. Thomsen. 1994. Principal descriptors of body condition score in Holstein cows. *J. Dairy Sci.* 77:2695–2703.
- Fischer, A., T. Luginbühl, L. Delattre, J. M. Delouard, and P. Faverdin. 2015. Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in Holstein dairy cows. *J. Dairy Sci.* 98:4465–4476.
- Halachmi, I., M. Klopčič, P. Polak, D. J. Roberts, and J. M. Bewley. 2013. Automatic assessment of dairy cattle body condition score using thermal imaging. *Comput. Electron. Agric.* 99:35–40.
- Halachmi, I., P. Polak, D. J. Roberts, and M. Klopčič. 2008. Cow body shape and automation of condition scoring. *J. Dairy Sci.* 91:4444–4451.
- Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. 2017. Densely connected convolutional networks. Pages 4700–4708 in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Honolulu, HI*. Institute of Electrical and Electronics Engineers, New York, NY.
- Hussein, H. A., A. Westphal, and R. Staufenbiel. 2013. Relationship between body condition score and ultrasound measurement of backfat thickness in multiparous Holstein dairy cows at different production phases. *Aust. Vet. J.* 91:185–189.
- Iandola, F. N., S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv* 1602.07360.
- Kamilaris, A., and F. X. Prenafeta-Boldú. 2018. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147:70–90.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. Pages 1097–1105

- in *Advances in Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY.
- Li, W., Z. Ji, L. Wang, C. Sun, and X. Yang. 2017. Automatic individual identification of Holstein dairy cows using tailhead images. *Comput. Electron. Agric.* 142:622–631.
- Rathgeb, C., and A. Uhl. 2011. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP J. Inform. Security* 2011:3.
- Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV.
- Roche, J. R., N. C. Friggens, J. K. Kay, M. W. Fisher, K. J. Stafford, and D. P. Berry. 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. *J. Dairy Sci.* 92:5769–5801.
- Ruiz-Garcia, L., and L. Lunadei. 2011. The role of RFID in agriculture: Applications, limitations and challenges. *Comput. Electron. Agric.* 79:42–50.
- Schröder, U. J., and R. Staufenbiel. 2006. Invited review: Methods to determine body fat reserves in the dairy cow with special regard to ultrasonographic measurement of backfat thickness. *J. Dairy Sci.* 89:1–14.
- Shen, W., H. Hu, B. Dai, X. Wei, J. Sun, L. Jiang, and Y. Sun. 2019. Individual identification of dairy cows based on convolutional neural networks. *Multimedia Tools Appl.* <https://doi.org/10.1007/s11042-019-7344-7>.
- Shen, Y., J. Li, and C. Lü. 2014. Depth map enhancement method based on joint bilateral filter. Pages 153–158 in *Proc. 7th International Congress on Image and Signal Processing*, Dalian, China. Institute of Electrical and Electronics Engineers, New York, NY.
- Singh, A., B. Ganapathysubramanian, A. K. Singh, and S. Sarkar. 2016. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21:110–124.
- Song, Q., M.-R. Zhao, X.-H. Zhou, Y. Xue, and Y.-J. Zheng. 2017. Predicting gastrointestinal infection morbidity based on environmental pollutants: Deep learning versus traditional models. *Ecol. Indic.* 82:76–81.
- Spoliansky, R., Y. Edan, Y. Parmet, and I. Halachmi. 2016. Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera. *J. Dairy Sci.* 99:7714–7725.
- Støier, S., H. D. Larsen, M. D. Aaslyng, and L. Lykke. 2016. Improved animal welfare, the right technology and increased business. *Meat Sci.* 120:71–77.
- Vasseur, E., J. Gibbons, J. Rushen, and A. M. D. Passillé. 2013. Development and implementation of a training program to ensure high repeatability of body condition scoring of dairy cows. *J. Dairy Sci.* 96:4725–4737.
- Wathes, C. M., H. H. Kristensen, J. M. Aerts, and D. Berckmans. 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? *Comput. Electron. Agric.* 64:2–10.
- Weber, A., J. Salau, J. H. Haas, W. Junge, U. Bauer, J. Harms, O. Suhr, K. Schönrock, H. Rothfuß, and S. Bielecki. 2014. Estimation of backfat thickness using extracted traits from an automatic 3D optical system in lactating Holstein-Friesian cows. *Livest. Sci.* 165:129–137.
- Zhang, X., X. Zhou, M. Lin, and J. Sun. 2017. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. *arXiv 1707.01083v2*.