Analysis of drinking behavior of beef cattle using computer vision

M. N. Islam[1], J. Yoder[1], A. Nasiri[1] and H. Gan[1,*]
[1]*Department of Biosystems Engineering and Soil Science, University of Tennessee, Knoxville, TN 37996, USA*
[*]Corresponding author: Hao Gan, hgan1@utk.edu

Abstract

Monitoring the drinking behavior of animals is of great importance for livestock farming. Measuring drinking time is labor demanding and thus, it is still a challenge in most livestock production systems. Computer vision technology using a low-cost camera system can be useful in overcoming this issue. Therefore, the aim of this research was to develop a computer vision system for monitoring beef cattle drinking behavior. A data acquisition system, including an RGB-D camera and an ultrasonic sensor was developed to record the beef cattle drinking actions. We developed an algorithm for tracking the beef cattle key body parts, such as head-ear-neck position, using a state-of-the-art deep learning architecture DeepLabCut. The extracted key points were analyzed using a Long Short-Term Memory (LSTM) model to classify drinking and non-drinking periods. A total of 70 videos were used to train and test the model and 8 videos were used for validating purposes. During the testing, the model achieved 97.35% accuracy. The results of this study will guide us to meet the immediate needs and expand farmers' capability in monitoring animal health and well-being.

Keywords: animal behavior, drinking time, computer vision, precision livestock farming

Introduction

The US is the land of the biggest yearly beef production, the third-biggest cattle herd, and the largest fed-cattle industry globally, producing $66.2 billion in gross income in 2019 (Martinez et al., 2021). In the US, heat stress is one of the major challenges in cattle production and management. Due to heat stress, cattle experience health problems as well as lower meat production, resulting in great economic losses to cattle farmers (Tsai et al., 2020). Based on previous studies, heat stress may also affect daily activity, including drinking and feeding (Tuan et al., 2022). Detecting this problem at an initial phase provides early care for the presence of clinical signs of disease related to the particular disease. Automatic disease detection would be a promising way to detect clinical signs at an early stage. Conventionally, direct contact (i.e., attaching sensors to animal bodies) has been the primary method of monitoring animal behavior (González et al., 2015; Smith et al., 2016). However, this method faces limited sampling frequency, low accuracy, and inconsistency problems (Guo et al., 2020). To overcome those issues, computer vision techniques can be used to identify and classify specific animal behaviors (Jorquera-Chavez et al., 2019; Li et al., 2022). Therefore, in recent years, animal video analysis or phenotyping has been examined to monitor the animal behavior and health of beef cattle (Gonzalez et al., 2016; Guo et al., 2020).

Different deep-learning techniques have been developed to automate animal behavior detection methods using computer vision and video analysis technology. Tsai et al. (2020) investigated dairy cow heat stress by monitoring drinking behavior using deep learning CNN with an imaging system and found that drinking behavior reflects the effects of heat stress on dairy cows. Wu et al. (2020) investigated a method that was proposed to detect the breathing frequency of standing resting dairy cows by using computer vision and video analysis. A Deeplab V3+ semantic segmentation model was developed using the framework of ResNet-101. Li et al. (2022) studied basic motion behaviors based on cow skeletons and a hybrid convolution algorithm. The multi-resolution module was used to extract cow skeletons. In previous attempts, various hybrid deep-learning tools were used to find animal behaviors. To simplify this issue, the pose estimation technique might be a solution. Pose estimation is a basic computer vision technique that identifies the location of a series of key body parts. In 2014, Toshev and Szegedy (2014) first applied a 2D human pose estimation technique using the deep learning method. Day by day, researchers developed many updated pose estimation methods such as YOLO, DeepLabCut, LEAP, and DeepPoseKit (M. W. Mathis and Mathis, 2020). Among them, DeepLabCut is the first tool for animal pose estimation, which uses ResNets transfer learning to reduce the training times (A. Mathis et al., 2018, 2021). In this study, the DeepLabCut pose estimation technique was used to train and validate the model for tracking cattle body parts.

Recently, to identify and classify livestock activity, different artificial neural networks have been used to improve the performance of recognition tasks. Chen et al. (2020) investigated pig drinking and drinker-playing behavior recognition based on ResNet50 and Long short-term memory (LSTM), and the classification accuracy for the body and head regions was found as 0.87 and 0.93, respectively. Wu et al. (2021) used a fusion of convolutional neural networks and LSTM for recognizing the basic behaviors (drinking, ruminating, walking, standing, and lying) of a single cow. Nasiri et al. (2022) proposed a technique to identify pose estimation-based lameness recognition for boilers using the CNN-LSTM model. Du et al. (2022) used the Resnet50-LSTM model to investigate the Broodstock breeding behavior, and the investigated method achieved an average accuracy of 97.32% for five kinds of breeding behavior recognition.

Based on the above-mentioned automatic livestock behavior recognition works, the artificial neural network and video analysis play an important role. Although researchers tried to identify different behavior for various purposes, not sufficient methods were reported to recognize the beef cattle and/or cow drinking behavior. Identifying this behavior is crucial to detecting heat stress. Consequently, as an alternative to previous behavior recognition methods, this paper intended to evaluate beef cattle drinking behavior by considering the skeleton-based body part including a CNN-based model (DeepLabCut) and a time-series network (LSTM). Therefore, the specific objective of this research was to develop a computer vision system for monitoring beef cattle drinking behavior.

Figure 1: Structure of the vision system: (1) mainframe, (2) waterer, (3) sensing unit, (4) solar panel kit, and (5) animal enter the waterer.

Materials and methods

Experimental site

The experiment was conducted at the Middle Tennessee AgResearch and Education Center (MTREC). The vision system consisted of a mainframe, a waterer, a sensing unit, and a solar panel kit (Figure 1). The solar kit was used as the main power source for the scale and the RFID units. The sensing unit included a Raspberry Pi 4 as a microprocessor, an Intel® RealSenseTM D435 camera, and an ultrasonic sensor, was powered using Power over Ethernet (PoE). A custom-designed enclosure was 3D printed to protect the sensing unit. To optimize power usage, the ultrasonic sensor was used to trigger camera recording when an animal is in the vision system.

Data collection and annotation

In this study, the animal drinking behavior was recorded from two camera positions. A total of 78 videos, including 39 videos of each position were recorded. The videos were recorded as Bathymetric Attributed Grid (BAG) files. After that, the MPEG-4 (MP4) (RGB), MP4 (depth), and Hierarchical Data Format (HDF) files were extracted from each BAG file. The MP4 (RBG) files were used to track the key body points. a total of five key points were used to observe the pose skeleton (Figure 2). The skeleton line represents the relationship and position of the key body points by connecting the key body points. An open source 'VGG Image Annotator' online Annotator tool was used to level the drinking time for the recorded video as ground truth data(Dutta et al., 2016).

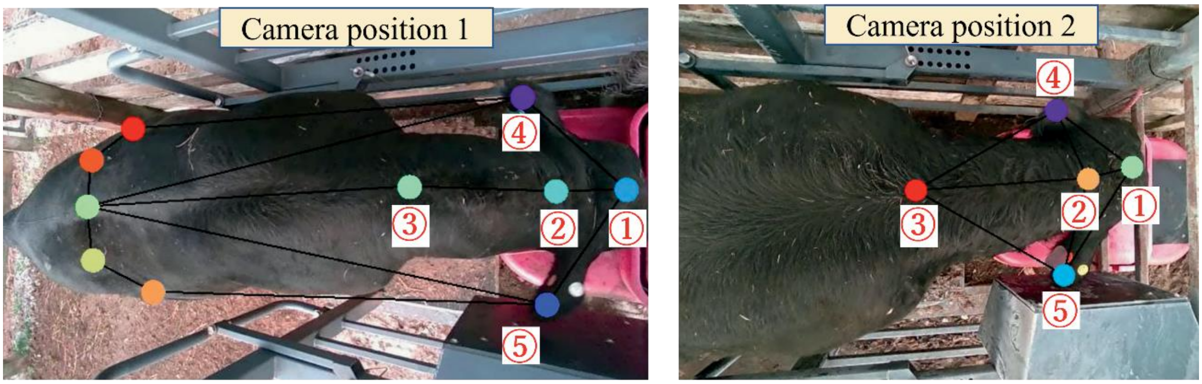

Figure 2: The annotated key points and beef cattle pose skeleton: (1) head, (2) upper neck, (3) lower neck, (4) left ear, and (5) right ear.

CNN-based pose recognition

Based on the literature review, key body points and pose recognition were identified using the DeepLabCut CNN-based pose estimation tool (Mathis et al., 2018). The DeepLabCut has two ResNets (50 and 101) architectures, and it helps the deconvolutional layers be replaced with dense layers to feature extraction (Russakovsky et al., 2015). The networks have the capability to learn and assign other points based on the higher or lower possibilities. In the last stage, the trained model analyzed the videos and estimated the pose for the whole dataset. In this study, the pre-trained ResNets 50 architecture was utilized as the transfer learning for pose estimation. Figure 3 demonstrates the diagram that outlines the pose estimation workflow.

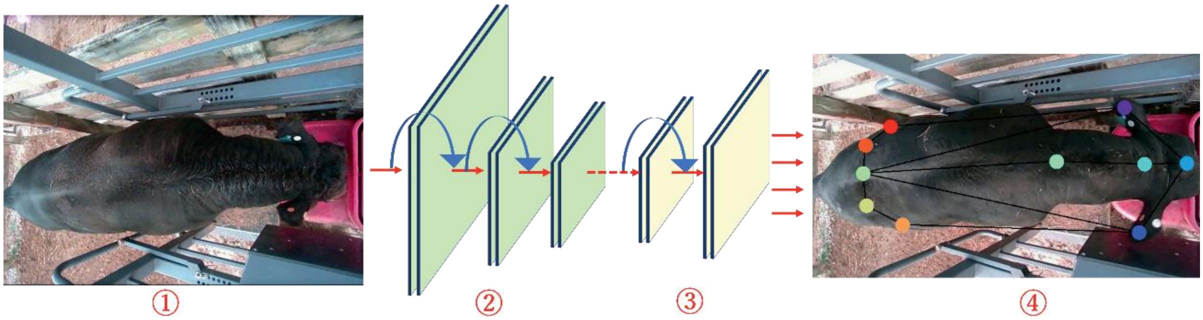

Figure 3: The diagram outlines the pose estimation workflow: (1) input, (2) pretrained model (ResNets 50), (3) deconvolutional layers, and (4) output with pose skeleton.

LSTM-based drinking behavior estimation

A total of 70 videos were used for training the model. In this study, 30-frame sequences were constantly sampled and predicted, and a 30-frame step size was set to make sure that the video sequence was repetitive each time. Therefore, setting the fixed window size as 30, sliding the window with a step size of 30, and then the proposed LSTM algorithm was used to detect the drinking behavior of small video segments obtained from the sequence. Figure 4 shows the schematic diagram of the sliding window video sequence sampling technique.

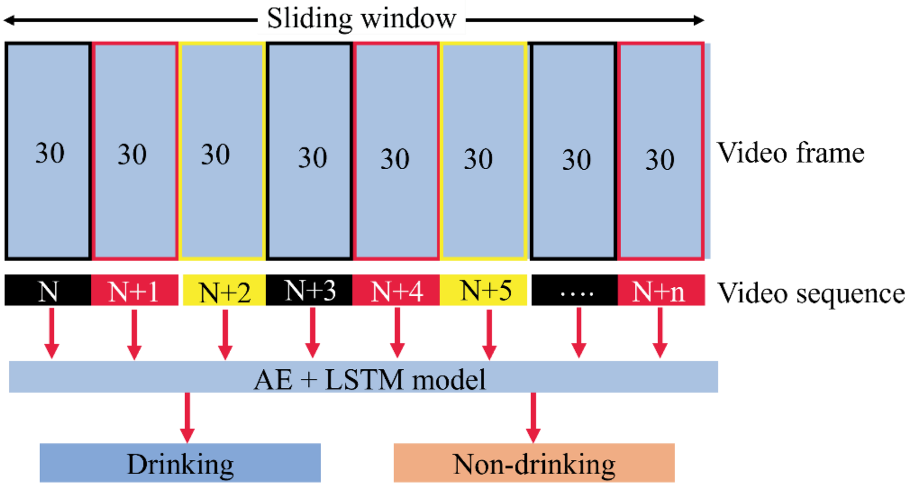

Figure 4: The schematic diagram of the sliding window sampling technique.

In deep learning studies, the augmentation of data is a very useful method to enhance the efficiency of the trained model by reducing over-sampling and augmenting random transformations (Nasiri et al., 2022). Therefore, before starting the LSTM training, a convolutional AutoEncoder (AE) was applied to augment the training dataset. Figure 5 shows the architecture of the proposed AE and LSTM model. For the AE, the training set was fully randomly split into two sets (train and test) with a proportion of 8:2. The selected AE model was trained for 300 epochs. The mean absolute error loss function and Adam optimizer were used for the AE model. After applying the augmentation process, the proposed LSTM model was trained for 1000 epochs including a cross-entropy loss function and Adam optimizer. The learning rate and learning rate decay were considered to train the LSTM model as $1e^{-5}$ and $1e^{-7}$, respectively. The input dimension of the LSTM was 30 × 10.

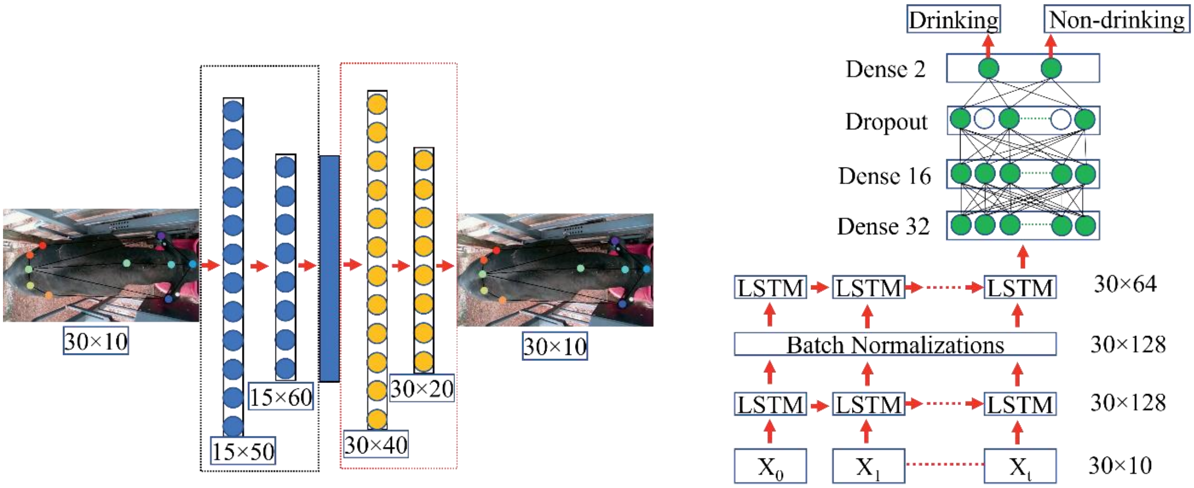

Figure 5: The architecture of the proposed AE (left) and LSTM model (right).

Results and discussion

Pose estimation

Figure 6 shows the DeepLabCut-based model loss values to identify key points. The model training procedure took nearly five days for each model, and the selected weights of the DeepLabCut were attained at 1030000 iterations along with a loss value of 0.0033 and 0.0023 at the learning rate of 0.001 for camera position 1 and 2, respectively.

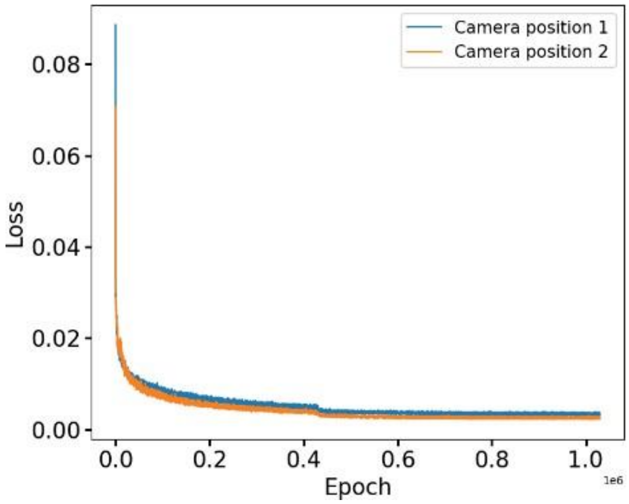

Figure 6: The loss values throughout the DeepLabCut model training for two camera positions

Evaluating the AE and LSTM model performance

Figure 7 (left) indicates the AE model loss values. The loss values for the best model were 17.90 and 11.77 for training and testing, respectively. There was a decreasing movement in the loss values with increasing epoch numbers. After epoch 150, training and testing loss curves were stabilized, which indicates that the proposed AE model has gained sufficient convergence domain. Figure 7 (right) shows the performance of the proposed LSTM model. After 200 epochs, the accuracy and loss curve increased and decreased simultaneously. The accuracy values for the best LSTM model were achieved at 97.35% and 97.37% for training and testing, respectively.

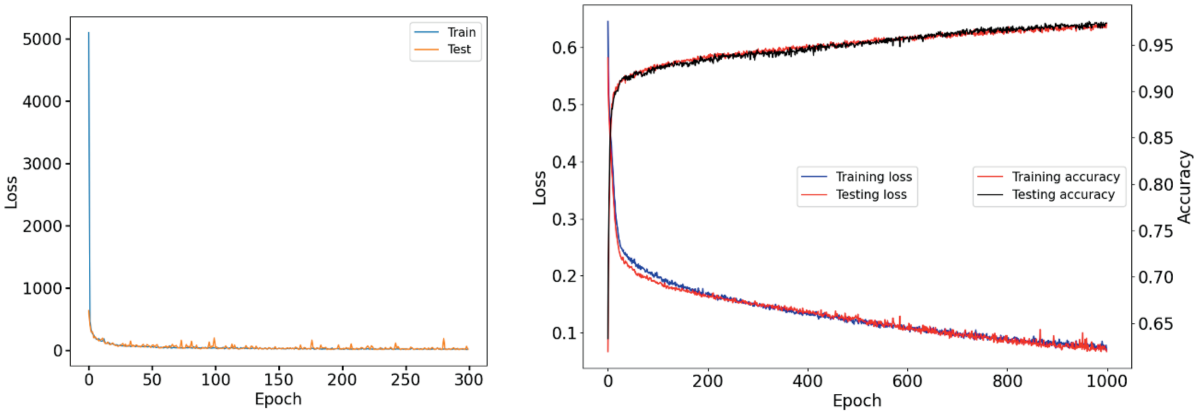

Figure 7: The changes in loss values during the AE model (left), and loss and accuracy values during the LSTM model (right)

Eight individual videos were used as validation of the proposed LSTM model. Table 1 shows the validation results of the proposed LSTM algorithm. The efficiency and confusion matrix were analyzed to evaluate the model. The highest accuracy was obtained at 98% for video numbers 1, 4, 5, 6, and 7. The lowest accuracy was obtained at 95% for video number 2. In addition, this proposed algorithm was able to calculate the drinking and non-drinking time.

Table 1: Validation results of the proposed LSTM algorithm

Video Number	Accuracy (%)	TP: Non-drinking (s)	TN: Drinking (s)	FP: Non-drinking (s)	FN: Drinking (s)	Drinking time (s)	Non-drinking time (s)
1	98	36	20	0	1	21	36
2	96	16	39	1	1	40	17
3	95	26	28	0	3	21	36
4	98	15	41	1	0	41	16
5	98	21	36	0	1	37	21
6	98	40	17	0	1	18	40
7	98	44	13	0	1	14	44
8	97	17	39	2	0	39	19

** Video numbers 1 to 4 were recorded as camera position 1, and 5 to 8 were recorded as camera position 2
** TP = True positive, TN = True negative, FP = False positive, and FN = False negative

Conclusions

This study proposed a skeleton-based computer vision drinking behavior recognition in beef cattle. The study established the first different camera position beef cattle pose estimation using DeepLabCut with ResNet50 backbone. A dataset containing 70 videos was evaluated using a DeepLabCut-based model to create the sequential key body points data. After that, an LSTM was classifying the drinking and non-drinking period by using the sequential key body points data. Currently, we are continuing to collect our video data from different farms using different camera positions. More collected data will further increase the accuracy of the LSTM as well as different classification model and prepare the vision system for profitable beef farms worldwide.

Acknowledgments

This project was funded by USDA NIFA (Award No. 2022-67021-37863) and the University of Tennessee AgResearch. Project title: PLF SPRINT: Precision Livestock Farming: Investing for Future Success

References

Chen, C., Zhu, W., Steibel, J., Siegford, J., Han, J., and Norton, T. (2020) Classification of drinking and drinker-playing in pigs by a video-based deep learning method. *Biosystems Engineering* 196, 1-14.

Du, L., Lu, Z., and Li, D. (2022) Broodstock breeding behaviour recognition based on Resnet50-LSTM with CBAM attention mechanism. *Computers and Electronics in Agriculture* 202, 107404.

Dutta, A., Ankush G., and Andrew Z. (2016) VGG image annotator (VIA). *https://www.robots.ox.ac.uk/~vgg/software/via/.*

González, L.A., Bishop-Hurley, G.J., Handcock, R.N., and Crossman, C. (2015) Behavioral classification of data from collars containing motion sensors in grazing cattle. *Computers and Electronics in Agriculture* 110, 91-102.

Gonzalez, L.F., Montes, G.A., Puig, E., Johnson, S., Mengersen, K., and Gaston, K.J. (2016) Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation. *Sensors* 16(1), 97.

Guo, Y., He, D., and Chai, L. (2020) A Machine Vision-Based Method for Monitoring Scene-Interactive Behaviors of Dairy Calf. *Animals* 10(2), 190.

Jorquera-Chavez, M., Fuentes, S., Dunshea, F.R., Warner, R.D., Poblete, T., and Jongman, E.C. (2019) Modelling and Validation of Computer Vision Techniques to Assess Heart Rate, Eye Temperature, Ear-Base Temperature and Respiration Rate in Cattle. *Animals* 9(12), 1089.

Li, Z., Song, L., Duan, Y., Wang, Y., and Song, H. (2022) Basic motion behaviour recognition of dairy cows based on skeleton and hybrid convolution algorithms. *Computers and Electronics in Agriculture* 196, 106889.

Martinez, C.C., Maples, J.G., and Benavidez, J. (2021) Beef Cattle Markets and COVID-19. Applied *Economic Perspectives and Policy* 43(1), 304-314.

Mathis, A., Biasi, T., Schneider, S., Yuksekgonul, M., Rogers, B., Bethge, M., and Mathis, M. (2021) Pretraining Boosts Out-of-Domain Robustness for Pose Estimation. In: *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1859-1868.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* 21(9), 1281-1289.

Mathis, M.W., and Mathis, A. (2020) Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology* 60, 1-11.

Nasiri, A., Yoder, J., Zhao, Y., Hawkins, S., Prado, M., and Gan, H. (2022) Pose estimation-based lameness recognition in broiler using CNN-LSTM network. *Computers and Electronics in Agriculture* 197, 106931.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., and Fei-Fei, L. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115(3), 211-252.

Smith, D., Rahman, A., Bishop-Hurley, G.J., Hills, J., Shahriar, S., Henry, D., and Rawnsley, R. (2016) Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems. *Computers and Electronics in Agriculture* 131, 40-50.

Toshev Alexander, and Christian Szegedy. (2014) DeepPose: Human Pose Estimation via Deep Neural Networks. In: *Proc. EEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1653-1660.

Tsai, Y.C., Hsu, J.T., Ding, S.T., Rustia, D.J.A., and Lin, T. (2020) Assessment of dairy cow heat stress by monitoring drinking behaviour using an embedded imaging system. *Biosystems Engineering* 199, 97-108.

Wu, D., Wu, Q., Yin, X., Jiang, B., Wang, H., He, D., and Song, H. (2020) Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. *Biosystems Engineering* 189, 150–163.