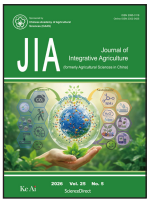
Journal of Integrative Agriculture 2026, 25(5): 2028–2040

KeAi
CHINESE ROOTS
GLOBAL IMPACT

Available online at ScienceDirect

Journal of Integrative Agriculture



RESEARCH ARTICLE

Multi-scale keypoints detection and motion features extraction in dairy cows using ResNet101-ASPP network

Saisai Wu[1, 2*], Shuqing Han[1, 2*], Jing Zhang[1, 2], Guodong Cheng[1, 2], Yali Wang[1, 2], Kai Zhang[1, 2], Mingming Han[3, 4], Jianzhai Wu[1, 2#]

[1] *Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[2] *Key Laboratory of Agricultural Blockchain Application, Ministry of Agriculture and Rural Affairs, Beijing 100081, China*
[3] *Zibo Agricultural Science Research Institute, Zibo 255020, China*
[4] *Zibo Institute for Digital Agriculture and Rural Research, Zibo 255051, China*

Highlights

- A ResNet101-ASPP network enhances multi-scale dairy cow keypoints detection in complex farming scenarios.
- Distinct keypoint trajectories were tracked to simultaneously extract seven motion features for gait analysis and lameness characterization in dairy cows.
- The proposed framework improves recognition of occluded keypoints and supports discrimination of health, mild, and severe lameness states.

Abstract

Detecting keypoints in dairy cows aims to locate and track the motion trajectories of the body's joints, which plays a crucial role in behavior analysis and lameness detection. However, real farming scenarios, characterized by occlusions and large variations in object scale may result in poor detection results. Therefore, we introduce the atrous spatial pyramid pooling (ASPP) module into the shallow layers network of ResNet101, designed to improve the multi-scale feature extraction capability of the model. The ASPP module enhances the robustness of recognition for different dimensional sizes and occluded keypoints using different dilatation rates in the parallel atrous convolutional layers to expand the model's receptive field. Furthermore, seven types of motion features, including tracking up, gait symmetry, step height balance, motion speed variability, head swing amplitude, head-neck slope and back curvature are extracted simultaneously by monitoring and tracking the motion trajectory of distinct keypoints. Several of these features represent innovative extraction models and attributes, first proposed in this study. Multiple models are trained and tested on datasets containing 2,385 frames for ablation experiments. The experiments show that, in comparison with the ResNet50, MobileNet_v2_1.0, and EfficientNet-b0 backbone networks, the training error and test error of ResNet101 are reduced by 4.04–30.12 pixels and 3.81–28.14 pixels. Therefore, ResNet101 is used as the benchmark for subsequent model improvement by adding the ASPP module. The training error and test error of the ResNet101-ASPP network are reduced by 0.27 and 0.24 pixels, respectively, compared to the benchmark network. The prediction confidence improves by 1.65–2.50% at three different dairy cow object scales. In addition, the keypoints under different occlusion conditions improve considerably, especially for small-scale keypoints, demonstrating the capability of the ASPP module for multi-scale feature extraction. By analyzing the distribution of the seven features and health, mild lameness, and severe lameness in dairy cows, it is shown that all the different features play an important role in distinguishing between different levels of lameness.

Keywords: dairy cows, multi-scale, keypoints detection, ResNet101-ASPP network, motion features

Received 6 February 2024; Received in revised form 30 April 2024; Accepted 1 June 2024; Available online 19 July 2024
Saisai Wu, E-mail: wusaisai0324@163.com; Shuqing Han, E-mail: hanshuqing@caas.cn; #Correspondence Jianzhai Wu, E-mail: wujianzhai@caas.cn
*These authors contributed equally to this study.

© 2026 CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer review under responsibility of Editorial Board of *Journal of Integrative Agriculture*.
doi: 10.1016/j.jia.2024.07.023

1. Introduction

Keypoints detection is a specific implementation of pose estimation, observing fine-grained behavioral changes through the detection and tracking of multiple joint points. It is vital for cow gait recognition, lameness detection, estrus behavior detection, and locomotion estimation. It has been widely studied and applied in species such as fruit flies (Pereira *et al.* 2019), broiler chickens (Fang *et al.* 2021; Nasiri *et al.* 2022), pigs (Alameer *et al.* 2020), and dairy goats (Chen *et al.* 2024). Recent advances in computer vision and human pose estimation provide new directions and tools for animal pose estimation. A significant amount of research is directed toward using deep learning algorithms for this task, demonstrating impressive feature extraction capabilities (Zimmermann *et al.* 2020). Current research on animal pose estimation methods focuses on three main aspects: (1) building lightweight algorithms to facilitate subsequent deployment efforts (Fan *et al.* 2023; Shuai *et al.* 2023); (2) working on keypoints localization accuracy in complex farming environments (Helena *et al.* 2022); (3) adopting an unsupervised approach to reduce dependence on manual annotation data (Li and Lee 2021).

Keypoints detection provides crucial data support for cow trajectory tracking and motion feature extraction. While some studies mainly utilized various electronic sensors attached to different cow body parts to record motion information (Hoang *et al.* 2018; Taneja *et al.* 2020; Cheng *et al.* 2022; Zhang K *et al.* 2022; Bovo *et al.* 2024), these devices are not only costly but also potentially stressful for cows. The development of computer vision technology led researchers to use images and videos for acquiring cow motion data, thereby circumventing the limitations of wearable sensors. Initially, specific body regions and visual features of cows are identified from images, followed by the construction of motion feature extraction models (Flower *et al.* 2005; Poursaberi *et al.* 2010; Zhao and He 2014; Chen Y *et al.* 2017; Kang *et al.* 2019; Jiang *et al.* 2022). However, these methods typically extract only one type of motion feature per model (Song *et al.* 2018b; Guo *et al.* 2019; Wu *et al.* 2020; Kang *et al.* 2022). The integration of diverse motion features can yield detailed insights into cow behavior and health, as lameness symptoms are manifested in various body parts during movement. Higaki *et al.* (2024) tracked 25 keypoints and measured 17 mobility variables using a software for cattle pose estimation, demonstrating the usefulness of different variables for lameness recognition.

However, in real farming scenarios, challenges such as differences in object size or inconsistent distances from the camera in individual cows can increase detection errors (Barney *et al.* 2023). Accurate keypoints detection requires integrating information at different scales, where local information detects different body components, and global information models the relationship between these components. This approach is crucial when dealing with large-scale changes and occlusions. For instance, Stacked Hourglass Networks (Newell *et al.* 2016) incorporate scale information by adding feature maps at different scales to the decoding stage through a short circuit layer, predicting the heatmap of the keypoints location from coarse to fine. HRNet (Wang *et al.* 2020) maintains high-resolution and spatial location information of the network's full-process feature map by preserving original resolution branches during downsampling, uniquely fusing multi-scale features at different resolutions. Designs that expand the model's receptive field have also proven effective in occlusion and multi-scale detection. For example, the RFB module expands the model's receptive field by simulating the receptive field of human vision in order to improve the detection accuracy of partially occluded and larger size objects (Zhang H M *et al.* 2022). Atrous convolution adds "holes" to the filter, which enlarges the receptive field of the convolution kernel, thus improving the ability of feature extraction (Gupta *et al.* 2021).

DeepLabCut with a pre-trained ResNet stands out as a precise and efficient tool for animal keypoints detection, and has been applied in real dairy cow farming scenarios (Zhao *et al.* 2023). Nonetheless, it lacks enhancements for addressing issues of occlusion and variation in object scale. To address these challenges, we designed a ResNet101-ASPP network, which combines a DeepLabCut's residual network and atrous convolution. This innovation incorporates the atrous spatial pyramid pooling (ASPP) module into shallow residual blocks, featuring both serial and parallel atrous convolution modules, and uses diverse dilation rates to expand the model's receptive field. It enables the capture of multi-scale features and the encoding of global context information. Based on the detection of the keypoints by the ResNet101-ASPP network, we extract seven types of motion features simultaneously by monitoring and tracking the trajectory coordinates of keypoints. Gait features (tracking up, gait symmetry, step height balance, motion speed variability) are obtained from the hooves, head bob features (head swing amplitude, head-neck slope) are obtained from the head and neck, and back curvature is obtained from the fitted curves of the neck, back and hip. The combination of different features offers a more comprehensive and fine-grained database for behavior recognition and lameness detection in dairy cows.

2. Materials and methods

2.1. Dataset acquisition

The video dataset was acquired at the dairy farm of the Beijing Agricultural Machinery Experimental Station of the China Academy of Agricultural Mechanization Science. Filming took place from March 2023 to September 2023, focusing on Holstein cows in lactation. This study involved capturing side-view videos of cows in their natural walking state. The filming location was the passageway connected from the exit of the milking parlor to the barn. A digital camera (Canon EOS 5D Mark III) was mounted on a tripod, positioned perpendicular to the passageway, at a height of 1.5 m above the ground and a distance of 5–9 m from the passageway. Care was taken to ensure that each video captured at least three full gait motion cycles of the cow. The videos were recorded with a resolution of 1,920×1,080 pixels at a frame rate of 50 frames per second. The experimental environment for video acquisition, including the filming

equipment, is illustrated in Appendix A.

This study collected 267 videos of 108 cows. The video dataset was designed to assess the model's robustness and generalization across various conditions, including a range of cow sizes, weather and lighting conditions, background complexities, scenes, filming distances, and environmental disturbances such as flying birds. Fig. 1 illustrates selected samples of these filming scenarios. Videos featuring stationary, running, or unclear cow visibility were excluded, resulting in a final dataset of 238 videos. These included 20 videos shot at a distance of 5 m, 198 at 7 m, and 20 at 9 m. From the captured videos, 159 videos were selected for model training and testing, and all of these videos were shot at a distance of 7 m, where the training set and test set contained 1,908 and 477 images, respectively.

After consulting with three experts in the field of precision livestock farming and considering that the positions of the limbs, head, and spine are crucial for pose analysis, we focused on labeling keypoints on these three body parts. As shown in Appendix B, a total of 16 keypoints were defined. During the labeling process, keypoints that are not directly visible but can be inferred are annotated, whereas completely invisible keypoints are not annotated.

2.2. Keypoints detection model

DeepLabCut toolbox DeepLabCut (Mathis *et al.* 2018), a toolbox for markerless pose estimation of animals, was the first to apply CNNs to animal pose estimation, achieving good performance with minimal training data through transfer learning (approximately 200 frames, in controlled laboratory environment). Despite improvements by other researchers (Graving *et al.* 2019; Pereira *et al.* 2019), DeepLabCut remains unparalleled in balancing accuracy and processing speed.

Considering the high costs associated with dataset acquisition and manual labeling, as well as the challenges in deployment due to the extensive param of network models, our keypoints detection model design emphasizes minimizing data labeling requirements, reducing network model param, while maintaining high accuracy and generalization capability on unseen data. Based on these considerations, the DeepLabCut toolbox was selected for its comprehensive facilitation of the keypoints detection process. The DeepLabCut toolbox simplifies the project creation, data labeling, training, evaluation, and automatic pose estimation through an integrated GUI interface. It efficiently tackles complex pose changes and image background interferences with minimal training data. DeepLabCut incorporates two key components: pre-trained ResNets and a deconvolutional layer (Insafutdinov *et al.* 2016), which are essential for object recognition and semantic segmentation algorithms. The backbone network employs a variant of ResNet pre-trained on the ImageNet dataset, a large-scale repository used for object recognition. The pre-training eliminates the need for extensive training samples to achieve human-level accuracy. The deconvolutional layer upsamples the visual information to output spatial probability density. Furthermore, it modifies the standard ResNet network by limiting its downsampling multiplicity to 16 times, resulting in larger feature maps. Additionally, it removes superfluous param and features from ResNet to accelerate inference.

The process of cow keypoints detection is detailed in Fig. 2. First, select the videos that fulfill specified criteria from the collected cow data. Second, use OpenCV's K-means clustering algorithm to randomly extract frames from these videos. Third, 16 keypoints of the cow are annotated according to Appendix B. Then, the annotated images are divided into the train set and test set in an 8:2 ratio, and then input to the improved network for training. The optimal param are saved during the training epochs to evaluate the model's performance. Finally, the trained model can also process the unlabeled dataset, predicting keypoint coordinates and skeleton features.

ResNet101-ASPP network The pose estimation algorithm and its innovative improvements in this study are mainly based on the DeepLabCut toolbox. First, the best network,



Fig. 1 Samples of collected videos. A, good lighting conditions, shooting distances of 5, 7, and 9 m respectively. B, poor lighting conditions, shooting distances of 5, 7, and 9 m respectively.

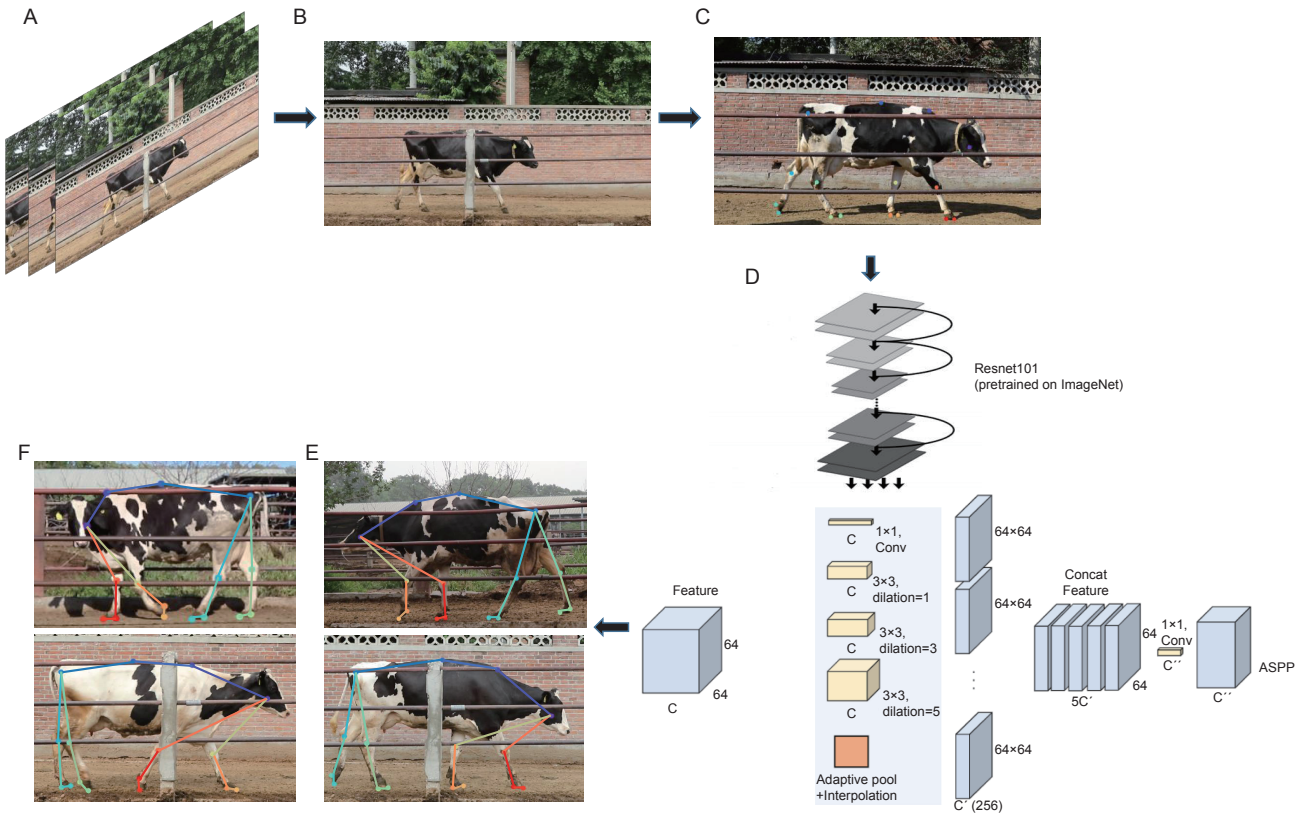


Fig. 2 Processing framework for keypoints detection. A, all frames in the video. B, randomly select key frames from the videos. C, label keypoints in frames, according to the keypoints setting in Appendix B. D, train the ResNet101-ASPP network, a distinct readout layer per body part is generated to predict the probability that a body part is in a particular pixel. Training adjusts both the readout and network weights. Post-training, the weights are stored. E, evaluate the performance of the trained network. F, apply the trained network to the unlabeled dataset.

ResNet101, is selected from several backbone networks integrated by DeepLabCut. To address the problem of accurately localizing of object multi-scale and occluded keypoints, the ASPP module is introduced into the shallow network of ResNet101 to enhance the multi-scale feature extraction capability of the algorithm.

(1) ResNet101 backbone. In DeepLabCut, there are various backbone networks, such as ResNet50, ResNet101, and MobileNet. After conducting experimental comparisons on the performance of each network, we ultimately select ResNet101 as the backbone network for keypoint extraction in dairy cows. Specific experimental details are presented in Section 3.1. (Experimental settings). The schematic of the ResNet101 framework is shown in Fig. 3, comprising four residual blocks with residual units in each block totaling 3, 4, 23, and 3. This structure ensures that the network effectively captures key information from images of varying scales and complexities. Following the post-processing by all blocks, the feature map is processed using the ReLU activation function, followed by 7×7 average pooling to average the feature information of each channel, resulting in a more compact feature representation. Then, the 2,048-dimensional feature vector is then input into the fully connected layer, followed by the Softmax layer to transform the features into relevant outputs such as score map and Part Affinity Fields (PAF). The score map is a probability distribution map indicating the location of the predicted keypoint, encoding the likelihood of the 16 keypoint at a specific pixel. Visualizing score maps aids in debugging the prediction process by highlighting which image parts the model focuses on to predict body parts. Additionally, location refinement maps predict offsets to counteract quantization errors due to the reduced sampling rate of the score map.

(2) ASPP module. In the unconstrained natural farming environment of dairy cows, where cows may occlude each other, be obscured by obstacles, or be captured at varying camera angles and distances, there are challenges in extracting effective features. Keypoints detection, a pixel-level output, requires up-sampling to restore the size of the original image. However, significant information from the original image can be lost during pooling and up-sampling, which can impact semantic segmentation outcomes. Inspired by the great segmentation performance of DeepLabv3 (Chen L C *et al.* 2017), we added the ASPP module into the residual block of ResNet101 network, which captures multi-scale features and encodes global context information, improving the accuracy of segmentation and vector-field prediction.

ASPP captures different levels of contextual information by introducing multiple parallel branches in the network, each of which uses different scales of atrous convolution and pooling operations. The receptive field can be enlarged using different dilation rates for atrous convolution to obtain a wider

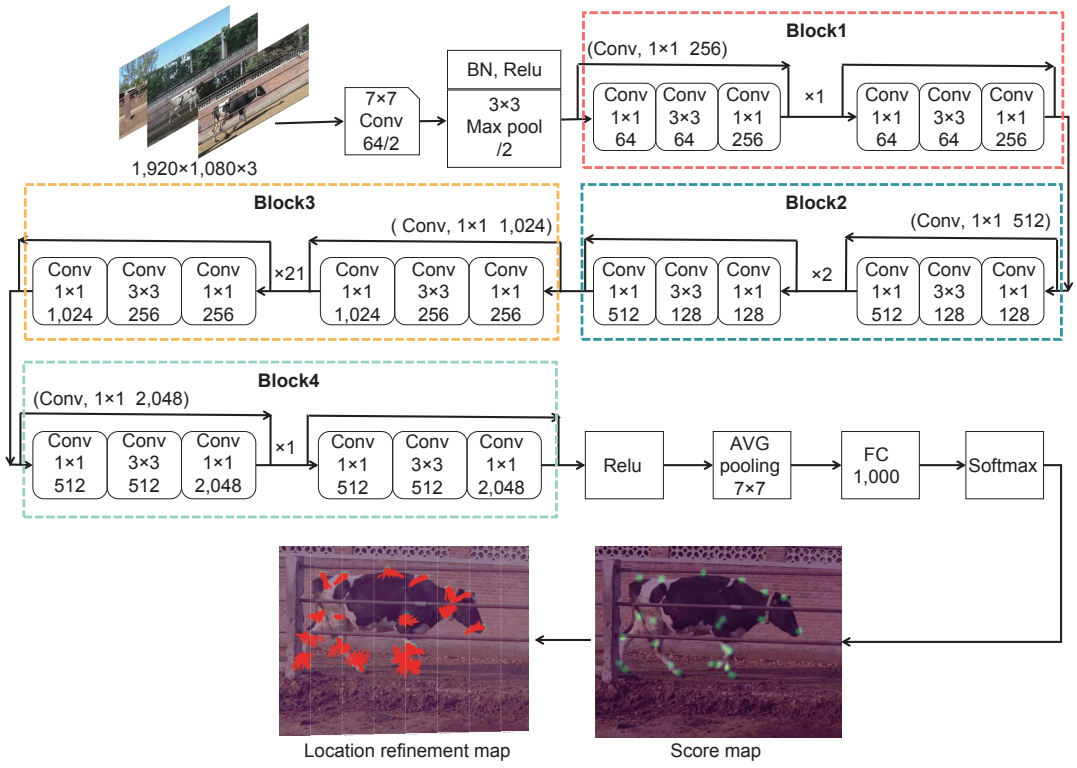


Fig. 3 Schematic of the ResNet101 framework.

range of contextual information. Simultaneously, pooling operations can aggregate information at different scales, thereby increasing the network's ability to perceive objects and structures at different scales. Compared with ordinary convolution, atrous convolution expands the receptive field without losing the resolution of the feature map (Fig. 4), and it does not increase the number of param, ensuring efficient access to global image information. When the convolution kernel size is k and the dilation rate is r, the atrous convolution introduces $r-1$ zeros between consecutive filter values, and the size of the receptive field is $k+(k-1)\times(r-1)$. Therefore, setting different dilation rates alters the receptive field, allowing the acquisition of multi-scale information.

The integration and structure of the ASPP module within the ResNet101 network are shown in Fig. 5. The core enhancement of the model is situated in the first residual block of ResNet101, where the intermediate convolutional step of the 2nd residual unit is replaced by the ASPP module, capitalizing on the abundant feature information present in the shallow residual block. The output of the first convolutional step of unit 2 first passed undergoes a 1×1 convolutional layer for dimensionality reduction. Then, a pooling pyramid is constructed. With specified dilation rates (1, 3, 5), the image features extracted from the shallow ResNet101 network are passed through three parallel atrous convolutional layers of the ASPP module to obtain the receptive fields at different scales, which facilitates the extraction of multi-scale information. Among them, the convolutional layer with a dilation rate of 1 is the standard convolution operation with the smallest receptive field to capture the detailed information. Convolutional layers with dilation rates of 3 and 5 expand the model's receptive field to varying degrees to extract image features in larger regions. Finally, the outputs of the three null convolutional layers are fused. Global average pooling then compresses the feature maps of each channel to 1×1, allowing for the extraction of channel-specific features and the subsequent acquisition of global features.

Evaluation metric In this study, the performance of the keypoints detection model is evaluated by calculating the average Euclidean distance error between manually labeled keypoints and those predicted by the model. The root mean square error (RMSE) is employed as the primary evaluation metric. The formula for calculating RMSE is:

$$\text{RMSE}=\sqrt{\frac{1}{m}\sum_i^m (x_i-y_i)^2} \quad (1)$$

where m is the number of keypoints, x_i is the predicted value of ith keypoint, and y_i is the ground truth of ith keypoint.

2.3. Motion features extraction model

After locating and tracking the 16 body joint keypoints of cow using the ResNet101-ASPP network, three types of cow motion features are extracted based on the continuous motion trajectories of several body parts, including gait, head bob, and back arch, which involved a total of seven distinct motion parameters.

Gait feature In studies of behavioral perception, gait detection of locomotion plays a pivotal role in assessing limb and hoof diseases and the overall health status. The gait features extracted in this study include tracking up, gait symmetry, step height balance and motion speed variability. Notably, gait symmetry is inspired by the feature extraction

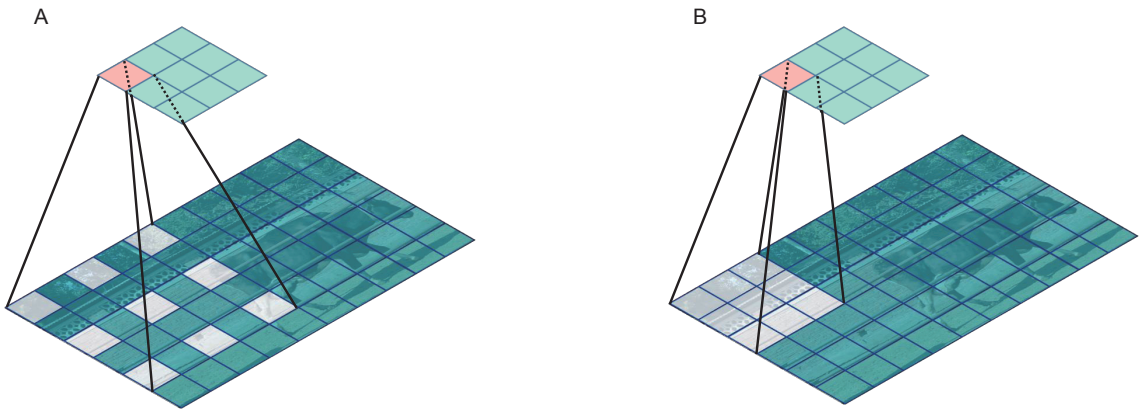


Fig. 4 Atrous convolution with different dilation rates. A and B, dilation rate is 2 (A) and 1 (B).

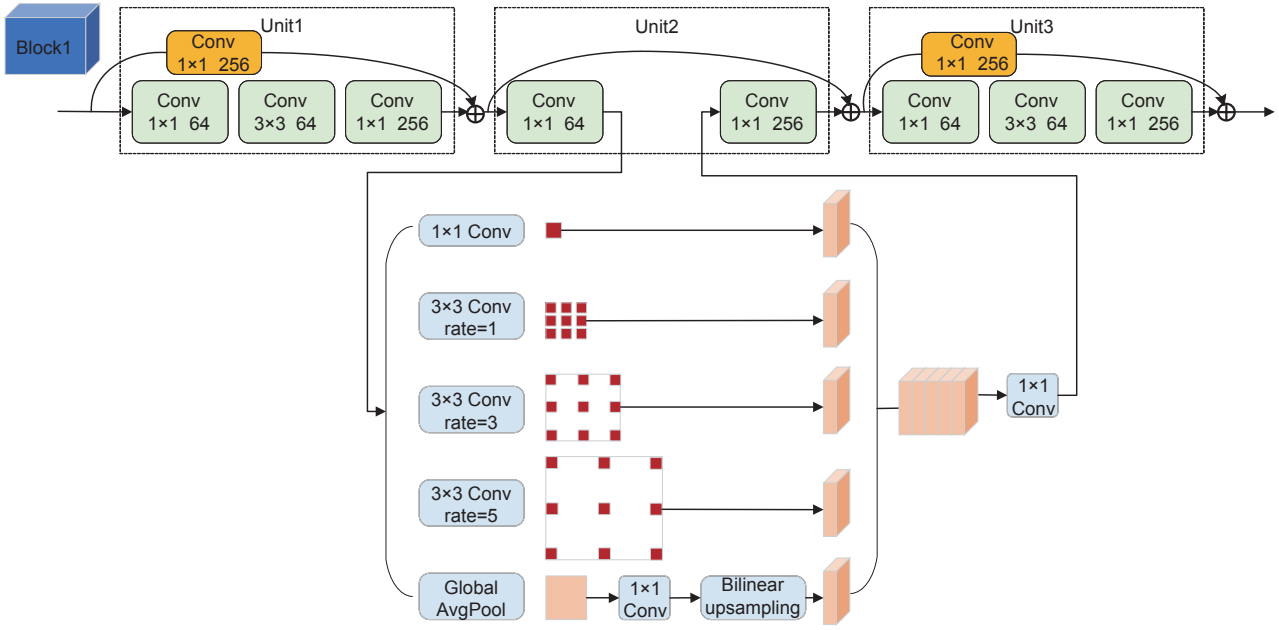


Fig. 5 The integration and structure of atrous spatial pyramid pooling (ASPP) module within the ResNet101 network.

methodology employed in human rehabilitation practices. Furthermore, the concept of step height balance represents a novel motion feature introduced for the first time in this research.

For gait feature extraction, the crucial aspect is identifying the time point and location of hooves starting and landing. During a cow's walking sequence in the video, hoof movements are primarily reflected in the changes of the *x*-axis coordinates. Hence, the *x*-coordinate data of hooves serve as the foundation for extracting gait features. As illustrated in Appendix C, the cow's gait is segmented into supporting and swing phases based on the gradient change of the x-coordinate trajectory. The supporting phase indicates when the hoof is at rest, while the swing phase denotes its motion.

(1) Tracking up. Tracking up primarily focuses on the evenness of stride lengths of the limbs during stepping to determine the presence of lameness or other diseases. Tracking up is calculated by measuring the positional difference between the front and hind hooves on the same side during a step. Assuming the sequence of steps for the four legs is: left front, right hind, right front, and left hind, with the starting positions of the left front hoof and right front hoof being A1 and A2, and the landing positions of the left hind hoof and right hind hoof being B1 and B2, respectively. The formula for calculating tracking up can be given as:

$$TU=\max(\Delta LP, \Delta RP) \quad (2)$$

$$\Delta LP=|x_{A1}-x_{B1}| \quad (3)$$

$$\Delta RP=|x_{A2}-x_{B2}| \quad (4)$$

where TU is tracking up, ΔLP and ΔRP are the step position differences of the left and right hooves, respectively. x_{A1} and x_{A2} are the x-coordinate values of the starting positions of the left front hoof and right front hoof, respectively. x_{B1} and x_{B2} are the x-coordinate values of the landing positions of the left hind hoof and right hind hoof, respectively. The lower TU value indicate the better tracking up performance in cows.

(2) Gait symmetry. Gait symmetry, involving both temporal

and spatial walking param in dairy cows, is indicative of potential gait abnormalities, including reduced speed, balance disorders, and uncoordinated limb movement. Gait asymmetry in dairy cows manifests not only in lameness but also in conditions such as udder fill, predominantly in the hind limbs. Thus, our focus is on the gait symmetry of the hind limbs. Unlike most current studies on gait symmetry that consider only a single aspect, like support time or stride length, we draw on the approach used to calculate human gait symmetry, which involves assessing the temporal and spatial param of both sides' limbs during stepping, providing a more comprehensive understanding of gait abnormalities. The formula for cow gait symmetry is:

$$GS=||GST_L-GST_R|\div 0.5\times|GST_L+GST_R||\times 100\% \quad (5)$$

$$GST_L=\frac{SP_L}{SW_L}\times\frac{1}{ST_L} \quad (6)$$

$$GST_R=\frac{SP_R}{SW_R}\times\frac{1}{ST_R} \quad (7)$$

where GS is the gait symmetry, GST_L and GST_R are the spatio-temporal information of the gait of the left and right hind limbs, respectively. SP_L and SW_L are the time of the left hind limb's supporting phase and swing phase, respectively. ST_L is the stride length of left hind limb. SP_R and SW_R are the time of right hind limb's supporting phase and swing phase, respectively. ST_R is the stride length of right hind limb. The gait symmetry value closer to 0 indicates better symmetry in the cow's gait.

(3) Step height balance. Prolonged observation and analysis of cow locomotion showed a significant correlation between the hoof tip to ground distance and lameness. Typically, cows raise their right and left hooves in a height balance. However, a hoof affected by stress or injury is often raised lower than its healthy counterpart, resulting in noticeable changes in the swing phase. Because the step height pertains to an individual hoof rather than the entire body, it mitigates the impact of individual differences and facilitates the direct identification of the specific lame hoof. The step height balance is calculated by comparing the difference in height between hoof tip and ground when all four hooves are stepping. The formula for step height balance is:

$$SHB=\max(FHV, RHV) \quad (8)$$

$$FHV=HV_{LF}-HV_{RF} \quad (9)$$

$$RHV=HV_{LR}-HV_{RR} \quad (10)$$

$$HV_i=\frac{\sum FF_{y,i}-G_{y,i}}{SF_i} \quad (11)$$

where SHB is the step height balance, FHV and RHV are the step height variability of front and hind hooves, respectively, HV_i is the average value of the lifting height of the ith hoof during stepping, $FF_{y,i}$ is the y-coordinate of the ith hoof tip during stepping, $G_{y,i}$ is the y-coordinate at the landing point of the ith hoof, and SF_i is the number of stepping frames of the ith hoof during the motion cycle, i can take the LF, RF, LR and RR. A higher SHB value indicates a higher probability of lameness in the cow.

(4) Motion speed variability. The motion speed of dairy cows serves as a significant indicator of their health and nutritional status. A notable discrepancy in the motion speed of one leg compared to the other three can suggest conditions such as arthritis or other diseases in the affected leg. Based on comparing the differences in motion speeds of the four legs during a motion cycle, the motion speed variability can be calculated as:

$$MSV=\max(MS_i)-\min(MS_i) \quad (12)$$

$$MS_i=\frac{MD_i}{SW_i} \quad (13)$$

where MSV is the motion speed variability, MS_i is the motion speed of ith leg, MD_i is the motion distance of ith leg, and SW_i is the swing phase time of ith leg. A larger MSV value indicates a greater likelihood of an issue with one of the legs.

Head bob feature Head bobbing behavior in cows during motion may indicate lameness, with nodding as an unconscious response to alleviate hoof pressure. It can also be a symptom of neck or spinal problems, or certain respiratory diseases where nodding aids in breathing by facilitating increased oxygen intake. This study observes the head bob feature of cows by extracting the head swing amplitude and head-neck slope during their walking motion.

(1) Head swing amplitude. The process involves fitting the y-coordinate data of the cow's head in consecutive video frames. Subsequently, the highest and lowest points across multiple head movement cycles are determined, and their differences are calculated, where the maximum difference obtained is considered the head swing amplitude of the cow. As shown in Appendix D, a swing cycle of the head is defined from one highest point to the next. The head swing amplitude is the maximum value among the differences in these swing cycles, and the swing cycle difference is computed as the disparity between the highest and lowest points of the cow's head movement within a swing cycle.

The formula for calculating head swing amplitude is:

$$HSA=\Delta SCD_i \quad (14)$$

$$SCD_i=y_{h,i}-y_{l,i} \quad (15)$$

where HSA is head swing amplitude, SCD_i is head swing difference of ith cycle, and $y_{h,i}$ and $y_{l,i}$ are the highest and lowest y-coordinate of head during the ith swing cycle, respectively. A lager HSA value indicates a greater likelihood of abnormality in the cow's condition.

(2) Head-neck slope. When cows experience problems such as lameness, there is also a significant change in the characteristics of the connection between the head, neck and back of the cow. Consequently, some studies have utilized this feature to detect lameness in cows (Song *et al.* 2018a). The formula for calculating the head-neck slope is:

$$HNS=\left|\frac{(y_1-y_2)}{(x_1-x_2)}\right| \quad (16)$$

where HNS is the head-neck slope; (x_1, y_1) and (x_2, y_2) are the position coordinates of the head and neck at the same moment, respectively.

Back arch feature Vertebral deformity or lameness in cows often results in an arched back as a compensatory mechanism for weight distribution and pain alleviation. The severity of lameness is typically correlated with the degree of back curvature. Following the methodologies outlined by Poursaberi *et al.* (2010) and Jiang *et al.* (2022), the extraction of back curvature involves three main keypoints located in the neck, back, and hip regions. As illustrated in Appendix E,

the radius of the fitted circle and the back curvature are calculated using the coordinates of these three points. The detailed derivation process is given as:

Suppose a circle passes through 3 points A(x_1,y_1), B(x_2,y_2), C(x_3,y_3) and the equation of the circle is given by:

$$Mx^2+My^2+Nx+Qy+P=0 \quad (17)$$

The values of M, N, Q, and P are solved in the following way:

$$M=x_1(y_2-y_3)-y_1(x_2-x_3)+x_2y_3-x_3y_2 \quad (18)$$

$$N=(x_1^2+y_1^2)(y_3-y_2)+(x_2^2+y_2^2)(y_1-y_3)+(x_3^2+y_3^2)(y_2-y_1) \quad (19)$$

$$Q=(x_1^2+y_1^2)(x_2-x_3)+(x_2^2+y_2^2)(x_3-x_1)+(x_3^2+y_3^2)(x_1-x_2) \quad (20)$$

$$P=(x_1^2+y_1^2)(x_3y_2-x_2y_3)+(x_2^2+y_2^2)(x_1y_3-x_3y_1)+(x_3^2+y_3^2)(x_2y_1-x_1y_2) \quad (21)$$

The radius r of a circle is calculated as follows:

$$r=\sqrt{\frac{N^2+Q^2-4MP}{4M^2}} \quad (22)$$

Given the relatively minor curvature involved, and to accentuate individual differentiation and to ease subsequent processing and application, we apply the negative logarithm to the calculated curvature based on the circle's radius. The computational equation for this process is:

$$K=-\log(\frac{1}{r}) \quad (23)$$

Analysis of association between motion features and lameness Cows with different levels of lameness exhibit variations in their motion characteristics, which serve as critical information for lameness classification. To validate and analyze the effectiveness and accuracy of the motion features extracted for lameness recognition, we first utilized the improved ResNet101-ASPP model to track the motion trajectories of 16 keypoints of cows. Then, we extracted a total of seven motion features, including gait, head swing and back arch, based on the above feature extraction model. We subsequently analyzed the distributions and correlations between cows with different lameness scores and different motion features.

A 5-point numerical rating system for cow motion was used as a standard for scoring cow lameness. A score of 1 indicates a normal cow and a score of 5 indicates a cow with a very high degree of lameness. Generally, when the score is greater than or equal to 3, the lame cow should be treated promptly. Three veterinarians experienced in lameness scoring were arranged to score cows for lameness by observing their walking status, and the average value was used as the gold standard for cow motion scoring. In this study, $1<x\leq3$ represents mild lameness and $3<x\leq5$ score represents severe lameness, where x denotes the average cow lameness score.

Box plots were used to visualize the lameness features and analyze the values of their quartiles to study the distribution of the features at different levels of lameness. To analyze the distribution of features at three degrees of lameness, the outliers, upper and lower bounds, mean, median, 25th, and 75th quartiles were calculated and compared for each feature. This approach clearly represents the distribution pattern between the feature and the degree of lameness, determining whether the feature is a key feature for lameness identification. Due to the different dimensions and large differences between features, all feature classes are normalized to avoid errors caused by this for feature analysis and lameness recognition.

3. Results and discussion

3.1. Keypoints detection

Experimental settings The experimental setup was configured using Python 3.7 within the Tensorflow 2.10.0 framework. The computational hardware comprised an Intel(R) Xeon(R) Gold 6132 CPU@2.60GHz, an NVIDIA GeForce RTX 2080 Ti GPU (11GB), and 256GB of memory.

For the experiment, 2,385 frames were selected from 159 videos of cows walking at a distance of 7 m for keypoints annotation. These frames were then randomly split into train and test sets in an 8:2 ratio, with the test set images used to assess the performance of the network model. The data were augmented during the training period by rotating image angles, zooming in and out images, and applying motion blur techniques. The training included 7 rotated frames per training image as well as 9 rotated and 14 partial images per training image, with the image adjustment range set to (50%,150%) and the motion blur angle between (–90, 90). The Stochastic Gradient Descent (SGD) optimizer was chosen for training, with a batch size of 1. The learning rate was set using a stepwise approach as detailed in Table 1, and the model was trained for a total of 600,000 epochs.

Results of different backbone networks The DeepLabCut toolbox offers a variety of backbone networks. In this study, four distinct networks were trained to identify the most effective one, which would then serve as the foundation for further model enhancements. The keypoints detection results of four backbone networks, namely MobileNet_v2_1.0, EfficientNet-b0, ResNet50, and ResNet101, are presented in Table 2. The results clearly demonstrate that ResNet101 outperforms the other three networks, with the train error and test error reduced by 4.04–30.12 pixels and 3.81–28.14 pixels, respectively. Overall, ResNet101 has an advantage in understanding and processing complex image tasks due to its depth and complexity, compared to MobileNet_v2_1.0 and EfficientNet-b0, which are more suitable in situations requiring lower computational resources, but may not be adequate in tasks requiring high accuracy. Consequently, ResNet101 is selected as the baseline network to add the

Table 1 Relationship between the number of iterations and the learning rate

Iteration number (≤)	Learning rate
10,000	0.005
430,000	0.02
600,000	0.002

Table 2 Performance comparison of different backbone networks

Network	Train error/pixel	Test error/pixel
MobileNet_v2_1.0	33.37	35.36
EfficientNet-b0	7.52	11.50
ResNet50	7.29	11.03
ResNet101	3.25	7.22

ASPP module for model improvement and subsequent experimental comparison.

Results of ResNet101-ASPP network To evaluate the effectiveness of the ASPP module in the optimized network, six sets of ablation experiments were conducted. There were five videos for each of the three shooting distances, each containing more than 100 images,which were independent of the train and test sets. The percentage of the cow's pixel area within the image was used as a metric to gauge the cow's scale size.

Since more feature information can be extracted from the shallow layers of the convolutional network, the ASPP module was primarily inserted into the shallow residual block of ResNet101. Experiments were conducted to observe the impact of different dilation rates at various locations on keypoints detection. Specifically, ASPP modules with dilation rates of (1, 3, 5) and (2, 4, 6) were added to several shallow residual blocks of ResNet101. For instance, 'ResNet101_b2u2_r1' indicates the addition of an ASPP module with a dilation rate of (1, 3, 5) in block2 unit2 of ResNet101, while 'ResNet101_b1u2_r2' refers to an ASPP module with a dilation rate of (2, 4, 6) in block1 unit2 of ResNet101.

Table 3 displays the performance outcomes following the integration of ASPP modules with varying dilation rates into different residual blocks of the ResNet101 network. It is observed that some improved networks demonstrate enhanced performance, while others exhibited a decline in performance. Notably, across three scales in dairy cows of prediction confidence, the confidence levels of all the improved networks are higher than the benchmark network. This phenomenon is likely attributable to the training and test datasets comprising images captured from a 5 m distance, where the benchmark network had already acquired enough image features and keypoints location information. However, when the prediction set includes images from 5, 7, and 9 m distances, the networks augmented with ASPP demonstrate its capability in extracting features at various scales and generalization potential.

The best train error result in the improved network is 2.98 pixels (an improvement of 0.27 pixels), and the test error is 6.8 pixels (an improvement of 0.42 pixels). Considering the model's input image resolution of 1,920×1,080, a test error of less than 0.63% is remarkably low, especially given the substantial size of the cows. The prediction confidence improved by 0.03 to 3.09% at 5 m, 0.65 to 1.65% at 7 m, and 0.15 to 2.50% at 9 m. Among all improved networks, ResNet101_b1u2_r1 achieves the best overall performance, excelling in train error, prediction confidence at 7 and 9 m scales, with its other results also ranking high. Its train error and test error were 2.98 and 6.98 pixels, respectively, and the prediction confidence at 5, 7, and 9 m scales were 84.17, 96.79, and 85.60%, respectively. These figures mark significant improvements over the benchmark network, with respective enhancements of 2.35, 1.65, and 2.50%.

Fig. 6 presents the keypoints prediction results of ResNet101_b1u2_r1 on unknown cows and its accuracy comparison with the benchmark network in different scenarios, encompassing various filming distances, occlusion levels, and lighting conditions. A, illustrates the keypoints detection results corresponding to each confidence score in the score map. In the detection process, a "+" signifies that the confidence level of the detection result exceeds the set threshold, while "×" indicates a confidence level below the threshold, set at 0.6 for this study. B, displays the skeletal structure of a cow, connected by various body joints. The posture of the cow can be analyzed using these skeleton features and the connection angles between joints. C, represents the prediction results of the benchmark and ResNet101_b1u2_r1 at filming distances of 5, 7 and 9 m respectively. In the Fig. 6, it can be observed that the benchmark is prone to keypoints miss-detection at 5 and 9 m, and it inaccurately detects of keypoints at small scales at 7 m. The improved network has higher prediction accuracy at all three different scales, showing a more complete and accurate keypoints detection and skeleton extraction, except that it exhibits a fuzzy keypoint miss-detection at 5 m. D, shows that the benchmark is more frequently subject to keypoint drift or miss-detection when the cow is self-obscured or obscured by an obstacle. The ResNet101_b1u2_r1 network benefits from the powerful ability of multi-scale feature extraction of the ASPP module to obtain accurate predictions from the spatial relationships of neighboring visible keypoints even when some keypoints are occluded. E, demonstrates that lighting conditions do not impact keypoints detection, likely due to the data augmentation strategy employed by DeepLabCut during training.

3.2. Motion features extraction

Motion trajectory Tracking the motion trajectories of cows is a fundamental step in the objective, accurate, and automated identification of abnormal health states in behavioral analysis. Based on keypoints detection, the positions of the same

Table 3 Performance comparison of different networks that add the ASPP module into ResNet101[1)]

Model	Train error/pixel	Test error/pixel	Prediction confidence (%)		
			5 m	7 m	9 m
ResNet101	3.25	7.22	81.82	95.14	83.10
ResNet101_b2u2_r1	3.31	6.81	84.24	96.48	83.65
ResNet101_b2u2_r2	3.36	7.34	**84.91**	95.79	85.12
ResNet101_b3u2_r2	3.25	7.30	81.85	96.52	83.38
ResNet101_b2u1u2_r2	3.31	**6.80**	81.87	96.11	85.38
ResNet101_b1u2_r2	3.67	7.90	83.10	95.93	83.25
ResNet101_b1u2_r1	**2.98**	6.98	84.17	**96.79**	**85.60**

[1)] Bold numbers are the optimal value for each evaluation metric.



Fig. 6 Keypoints and skeleton prediction results of unknown cows at different scales, obscuration levels and light conditions. A, keypoint detection confidence. B, skeleton. C, benchmark and ResNet101-ASPP prediction results for 5, 7, and 9 m filming distances. The top row represents the benchmark, while the bottom row represents ResNet101-ASPP. D, benchmark and ResNet101-ASPP prediction results for different levels of obscuration. The top row represents the benchmark, while the bottom row represents ResNet101-ASPP. E, ResNet101_b1u2_r1 network prediction results for different light conditions.

body parts across the entire video frame sequence are connected to form the motion trajectory. Given that cows move in a high-dimensional space, the keypoints annotation or prediction process is prone to distortions or jumps, which introduce noise and errors. To address this, we employ the Savitzky-Golay (SG) filter (Savitzky and Golay, 1964) for noise removal and data smoothing.

The SG filter is a method of local polynomial least squares fitting that smooths 2m+1 discrete data points within a window using the least squares technique. This method directly smooths data in the time domain, eliminating the need for conversions between frequency and time domains. This not only reduces computational memory and data processing demands but also ensures the shape and width of the signal remain consistent while filtering out noise. In our study, the SG filter is applied to process the coordinates of keypoints detected by the ResNet101-ASPP network, filtering out data noise and thus enhancing the accuracy of motion feature extraction and the model's robustness. The principle of the SG filter is as follows.

As shown in Appendix F, x_i $(i=-m, ..., 0, ..., m)$ is a set of continuous data in a moving window with the following nth-order fitted polynomial.

$$y(n)=\sum_{k=0}^{n} b_k i^k=b_0+b_1 i+b_2 i^2+\cdots+b_n i^n \quad (24)$$

where b_k is the coefficient of the polynomial, which can be obtained by querying the table of convolution coefficients produced by Savitzky and Golay (1964).

Since this study focuses on cows in a normal walking state, characterized by relatively uniform motion, the error in the predicted keypoints is minimal. As shown in Fig. 7, the hoof motion trajectory formed by the filtered data is smoother, and the motion peaks and valleys of each motion cycle can be clearly identified. Therefore, we will utilize this noise reduced data as the basis for motion trajectory analysis and subsequent motion feature extraction.

Distribution of features among different lameness levels

A total of 15 cows were used to analyze the distribution and association between motion features and lameness scores, with five cows in each of the health, mild and severe lameness. Fig. 8 shows the box plots for seven categories of motion features across three distinct lameness levels, where each feature category has been normalized.

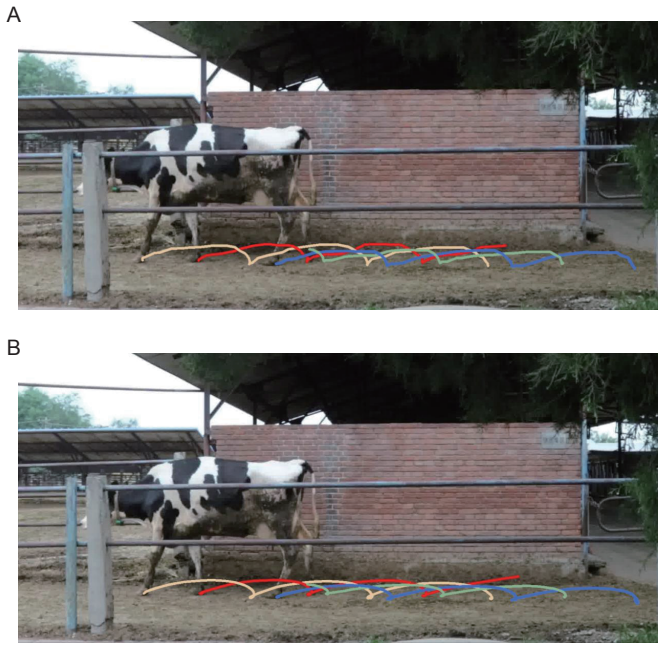


Fig. 7 Comparison of motion trajectory smoothness before and after data filtering. A and B, motion trajectory before (A) and after (B) data filtering.

Overall, a positive correlation is observed between the level of lameness and the distribution of most features. The tracking up feature (Fig. 8-A) exhibits a clear upward trend across the three groups, with the severe group showing a significantly wider range of values compared to the other two groups. This indicates that as the level of lameness in cows intensifies, noticeable issues arise in their tracking up during locomotion. The gait symmetry feature (Fig. 8-B) with the health group shows a concentrated distribution, with values closely approximating zero, suggesting excellent gait symmetry in healthy cows. Conversely, the severe group displays a much larger distribution range. Fig. 8-C illustrates that while there is a gradual upward trend in the distribution of the step height balance feature across the three groups, the boundaries of distribution are somewhat blurred. The motion speed variability feature (Fig. 8-D) shows dense distributions with similar ranges for both the health and mild groups, but a significantly larger range for the severe group, indicating a noticeable difference in motion speed between the affected leg and the other three legs in cows with severe lameness. Fig. 8-E reveals a higher incidence of outliers in the head swing amplitude, especially at the lower range of the health and mild groups, indicating a high degree of variation between individual cows. The distribution of the head-neck slope feature (Fig. 8-F) generally shows a marked upward trend, with the mild group's distribution being quite dense and the severe group exhibiting the largest distribution range. The back curvature feature (Fig. 8-G) demonstrates a smaller distribution range for the health group, a slight increase in back curvature for the mild group, with a slightly wider interquartile range and increased variability among individuals, and the highest median for the severe group, indicating a significant increase in back curvature.

The results indicate that relying on a single feature for the identification of lameness is not sufficiently significant, particularly for mild cases of lameness, where a combination of multiple features is required to provide more comprehensive information. By analyzing the data distribution and patterns of different features between the health group and the other groups, key features critical for lameness detection can be identified. When comparing the health group with the other two groups of lame cows, it is found that the maximum values for tracking up, gait symmetry, and head-neck slope are less than the 25th percentile of the other two groups, highlighting significant differences in the distribution of these features between healthy and lame cows. Therefore, these three features are important for distinguishing between healthy and lame cows.

In the health group, the 75th percentile values for tracking up, gait symmetry, head swing amplitude, and head-neck slope are all below the median of the mild group, making these features crucial in differentiating mild lame cows from healthy ones.

When cows exhibit severe lameness, the medians for tracking up, gait symmetry, step height balance, motion speed variability, head-neck slope, and back curvature show significant increases compared to the mild group, indicating that these features are crucial for distinguishing between severe and mild cases of lameness.

4. Conclusion

In this study, we developed a ResNet101-ASPP network for dairy cow keypoint and skeleton detection in real farm environments. By incorporating ASPP into shallow ResNet101 layers, the model enhances multi-scale feature extraction and global context modeling, improving robustness to scale variation and occlusion. Based on keypoint trajectories, seven motion features were extracted simultaneously to characterize gait, head movement, and back posture. Tracking up and gait symmetry were consistently informative for lameness assessment, while head-neck slope, head swing amplitude, step height balance, motion speed variability, and back curvature further improved discrimination across lameness levels. Overall, this study provides a practical basis for automated lameness detection and behavior analysis in dairy cows. Future work should prioritize lightweight deployment and reducing the effects of individual variability.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (32102600), the Central Public-interest Scientific Institution Basal Research Fund, China (Y2023XK13, JBYW-AII-2024-28/40, and JBYW-AII-2023-33/37/42), Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2021-AII), the Wuhu Science and Technology Bureau Two Strong One Increase Project, China (2023ly12).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used

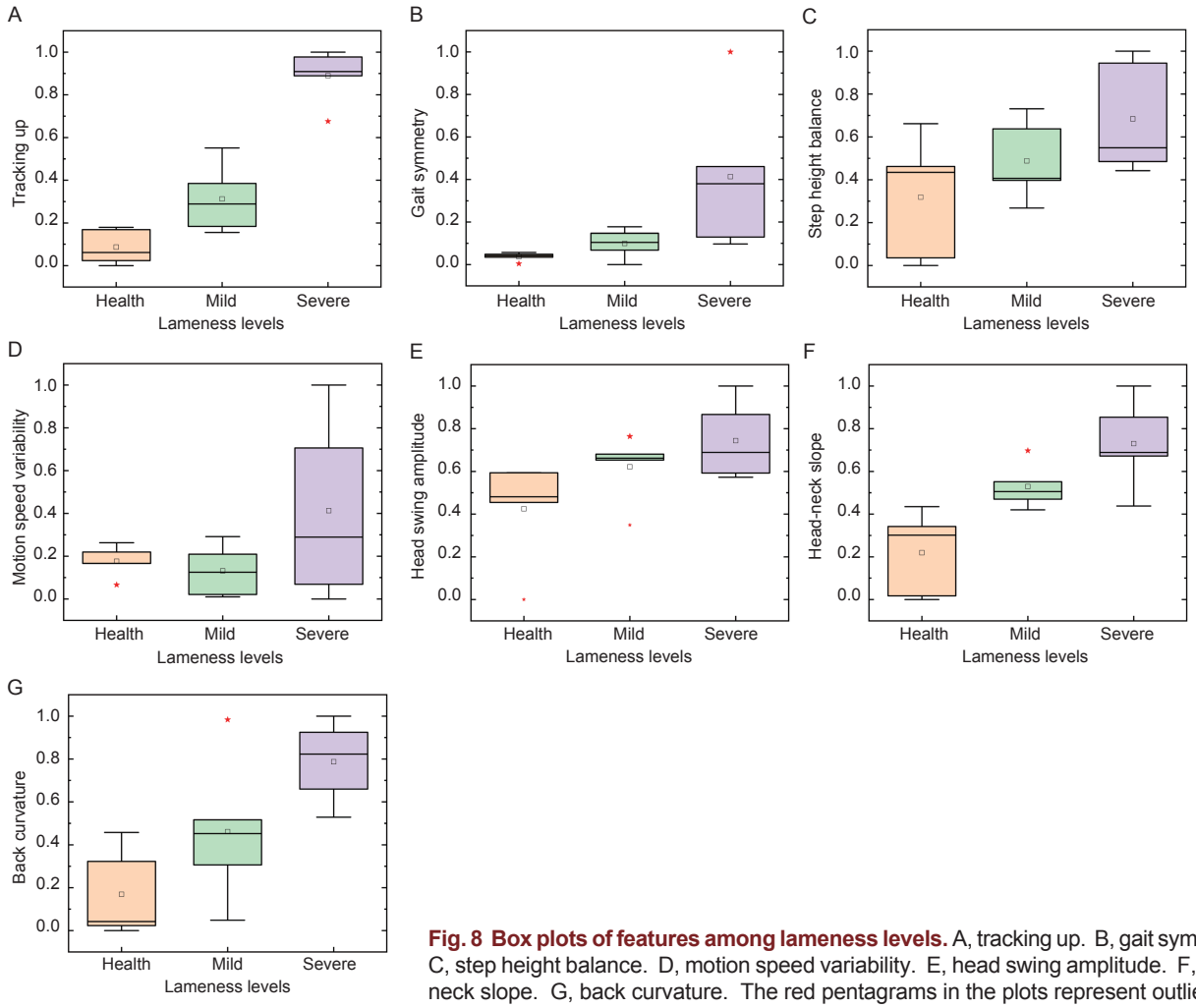


Fig. 8 Box plots of features among lameness levels. A, tracking up. B, gait symmetry. C, step height balance. D, motion speed variability. E, head swing amplitude. F, head-neck slope. G, back curvature. The red pentagrams in the plots represent outliers.

ChatGPT 3.5 to assist in improving the clarity and readability of the language. All content generated was carefully reviewed and revised by the authors as necessary, and the authors take full responsibility for the final content of the article.

Ethical approval

All animal procedures followed protocols approved by the Chinese Academy of Agricultural Sciences and Institutional Animal Care and Use Committee.

Appendices associated with this study are available at https://doi.org/10.1016/j.jia.2024.07.023

References

Alameer A, Kyriazakis I, Bacardit J. 2020. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs. *Scientific Reports*, **10**, 13665.

Barney S, Dlay S, Crowe A, Kyriazakis L, Leach M. 2023. Deep learning pose estimation for multi-cattle lameness detection. *Scientific Reports*, **13**, 4499.

Bovo M, Ceccarelli M, Agrusti M, Torreggiani D, Tassinari P. 2024. DAIRY CHAOS: Data driven approach identifying dairy cows affected by HeAt lOad Stress. *Computers and Electronics in Agriculture*, **218**, 108729.

Chen L, Zhang L, Tang J, Tang C, An R, Han R, Zhang Y. 2024. GRMPose: GCN-based real-time dairy goat pose estimation. *Computers and Electronics in Agriculture*, **218**, 108662.

Chen L C, Papandreou G, Schroff F, Adam H. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv*, **1706**, 05587.

Chen Y, He D, Fu Y, Song H. 2017. Intelligent monitoring method of cow ruminant behavior based on video analysis technology. *International Journal of Agricultural and Biological Engineering*, **10**, 194–202.

Cheng G, Wu J, Xing L, Zhu M, Zhang J, Han, S. 2022. Fine-grained cow's behavior classification method based on IMU. *Journal of China Agricultural University*. **27**, 179–186. (in Chinese)

Fan Q C, Liu S C, Li S Q, Zhao C J. 2023. Bottom-up cattle pose estimation *via* concise multi-branch network. *Computers and Electronics in Agriculture*, **211**, 107945.

Fang C, Zhang T M, Zheng H K, Huang J D, Cuan K X. 2021. Pose estimation and behavior classification of broiler chickens based on deep neural networks. *Computers and Electronics in Agriculture*, **180**, 105863.

Flower F, Sanderson D, Weary D. 2005. Hoof pathologies influence kinematic measures of dairy cow gait. *Journal of Dairy Science*, **88**, 3166–3173.

Graving J M, Chae D, Naik H, Li L, Koger B, Costelloe B R, Couzin I D. 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife*, **8**, e47994.

Guo Y, Zhang Z, He D, Niu J, Tan Y. 2019. Detection of cow mounting behavior using region geometry and optical flow characteristics. *Computers and Electronics in Agriculture*, **163**, 104828.

Gupta D, Artacho B, Savakis A. 2021. VehiPose: A multi-scale framework for vehicle pose estimation. In: *Applications of Digital Image Processing XLIV. SPIE* (*The international society for optics and photonics*), United States. pp. 517–523.

Helena R, Rik T, Gert K. 2022. T-LEAP: Occlusion-robust pose estimation of walking cows using temporal information. *Computers and Electronics in Agriculture*, **192**, 106559.

Higaki S, Matsui Y, Miwa M, Yamamura T, Hojo T, Yoshioka K, Vang A, Negreiro A, Dórea J. 2024. Leveraging computer vision-based pose

estimation technique in dairy cows for objective mobility analysis and scoring system. *Computers and Electronics in Agriculture*, **217**, 108573.

Hoang Q T, Phung C P K, Bui T N, Chu T P D, Tran D T. 2018. Cow behavior monitoring using a multidimensional acceleration sensor and multiclass SVM. *International Journal of Machine Learning and Networked Collaborative Engineering*, **2**, 110–118.

Insafutdinov E, Pishchulin L, Andres B, Mykhaylo A, Bernt S. 2016. *Deepercut*: *A Deeper*, *Stronger*, *and Faster Multi-Person Pose Estimation Model*. Springer International Publishing, The Netherlands. pp. 34–50.

Jiang B, Song H, Wang H, Li C. 2022. Dairy cow lameness detection using a back curvature feature. *Computers and Electronics in Agriculture*, **194**, 106729.

Kang X, Li S, Li Q, Liu G. 2022. Dimension-reduced spatiotemporal network for lameness detection in dairy cows. *Computers and Electronics in Agriculture*, **197**, 106922.

Kang X, Zhang X, Liu G, Ma L. 2019. Hoof location method of lame dairy cows based on machine vision. *Transactions of the Chinese Society for Agricultural Machinery*, **50**, 276–282. (in Chinese)

Li C, Lee G H. 2021. From synthetic to real: Unsupervised domain adaptation for animal pose estimation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1482–1491.

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

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

Newell A, Yang K, Deng J. 2016. *Stacked Hourglass Networks for Human Pose Estimation*. Springer International Publishing, The Netherlands. pp. 483–499.

Pereira T D, Aldarondo D E, Willmore L, Kislin M, Wang S S H, Murthy M, Shaevitz J W. 2019. Fast animal pose estimation using deep neural networks. *Nature Methods*, **16**, 117–125.

Poursaberi A, Bahr C, Pluk A, Van N A, Berckmans D. 2010. Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques. *Computers and Electronics in Agriculture*, **74**, 110–119.

Shuai L Y, Mu J, Jiang X Q, Chen P, Zhang B D, Li H D, Wang Y C, Li Z Y. 2023. An improved YOLOv5-based method for multi-species tea shoot detection and picking point location in complex backgrounds. *Biosystems Engineering*, **231**, 117–132.

Song H, Jiang B, Wu Q, Li T, He D. 2018. Detection of dairy cow lameness based on fitting line slope feature of head and neck outline. *Transactions of the Chinese Society of Agricultural Engineering*, **34**, 190–199. (in Chinese)

Taneja M, Byabazaire J, Jalodia N, Davy A, Olariu C, Malone P. 2020. Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle. *Computers and Electronics in Agriculture*, **171**, 105286.

Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X. 2020. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 3349–3364.

Wu D, Wu Q, Yin X, Jiang B, Wang H, He D, 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.

Zhang H M, Li Y H, Zhou L X, Wang R, Li S Q, Wang H Y. 2022. Multi-objective skeleton extraction method of beef cattle based on improved YOLO v3. *Transactions of the Chinese Society for Agricultural Machinery*, **53**, 285–293. (in Chinese)

Zhang K, Han S, Cheng G, Wu S, Liu J. 2022. Gait phase recognition of dairy cows based on Gaussian Mixture model and Hidden Markov model. *Smart Agriculture*, **4**, 53–63. (in Chinese)

Zhao K, He D. 2014. *Real-Time Automatic Classification of Lameness in Dairy Cattle Based on Movement Analysis with Image Processing Technique*. American Society of Agricultural and Biological Engineers, Canada.

Zhao K, Zhang M, Ji J, Zhang R, Bewley J. 2023. Automatic lameness scoring of dairy cows based on the analysis of head-and back-hoof linkage features using machine learning methods. *Biosystems Engineering*, **230**, 424–441.

Zimmermann C, Schneider A, Alyahyay M, Brox T, Diester I. 2020. FreiPose: A deep learning framework for precise animal motion capture in 3D spaces. *BioRxiv*, **27**, 967620.

Executive Editor-in-Chief Kui Li
Managing Editor Han Zhang