Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Yolo11-Based Pose Estimation and Geometric Locomotion Trait Extraction in Beef Cattle: A Proof-Of-Concept Study

Luana Maria Benicio
University of Illinois Urbana-Champaign
Cody Ryan Dawson
University of Illinois Urbana-Champaign
Mamunur Rahman
University of Illinois Urbana-Champaign
Daniel Shike
University of Illinois Urbana-Champaign
Joshua C. McCann
University of Illinois Urbana-Champaign
Isabella C. F. S. Condotta
icfsc@illinois.edu

University of Illinois Urbana-Champaign

Article

Keywords: computer vision, pose estimation, locomotion, beef cattle, precision livestock farming

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11117323/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

YOLO11-BASED POSE ESTIMATION AND GEOMETRIC LOCOMOTION TRAIT EXTRACTION IN BEEF CATTLE: A PROOF-OF-CONCEPT STUDY

Luana Maria Benicio[1], Cody Ryan Dawson[2], Mamunur Rahman[2], Joshua C. McCann[2], Daniel Shike[2], Isabella C. F. S. Condotta[1,2*]

[1] *Agricultural & Biological Engineering, University of Illinois Urbana-Champaign, Urbana, IL, United States 61801*

[2] *Department of Animal Sciences, University of Illinois Urbana-Champaign, Urbana, IL, United States 61801*

*Corresponding author: Isabella C. F. S. Condotta, icfsc@illinois.edu

Abstract

Objective assessment of locomotion-related traits could improve repeatability in beef cattle structural phenotyping. This proof-of-concept study evaluated a YOLO11-based workflow for cattle detection, localization of five limb keypoints, and extraction of forelimb step length, hock angle, and hind-limb angle from controlled side-view RGB images. Detection and pose datasets contained 1,000 images from 138 steers and 1,200 images from 120 steers, respectively, with animal-level training, validation, and test partitions. On held-out animals, detection mAP@50-95 was 0.9678 and pose mAP@50-95 was 0.9950. Mean keypoint localization error was 9.86 ± 5.76 pixels, or 2.46% of the larger cattle bounding-box dimension, with percentage of correct keypoints (PCK) values of 0.943, 0.998, and 1.000 at thresholds of 0.05, 0.10, and 0.20. Forelimb step length showed good agreement between manual and model-derived measurements (MAE 4.68 cm; Lin's CCC 0.903). Hock and hind-limb angles showed weak individual-animal concordance (CCC 0.146 and 0.317).

Keywords: computer vision, pose estimation, locomotion, beef cattle, precision livestock farming.

Introduction

Structural soundness and mobility influence the welfare, productive life, and management of beef cattle. Feet and leg traits are commonly evaluated using visual scoring systems, but scoring depends on evaluator training and can show limited repeatability across observers and settings [1–

3]. Image-based methods offer a way to quantify limb position and movement using repeatable geometric measurements.

Computer vision methods have increasingly been used to characterize cattle posture and locomotion. Pose-estimation approaches identify anatomical landmarks that can be converted into joint angles, stride-related measures, and other locomotion features. Most published work has focused on dairy cattle, including keypoint detection, temporal pose estimation, and automated lameness assessment [4–8]. Beef cattle applications remain less developed, particularly for objective extraction of feet and leg traits that can complement conventional conformation assessment. We evaluated both keypoint localization and the agreement between measurements calculated from predicted and manually annotated keypoints.

YOLO11 includes object-detection and pose-estimation model variants [9]. We used these models with controlled side-view RGB images of walking steers and calculated forelimb step length, hock angle, and hind-limb angle from five predicted limb keypoints.

The objective was to develop and evaluate a YOLO11-based pipeline for beef cattle pose estimation and geometric locomotion trait extraction. Model performance was evaluated with animal-level separation of training, validation, and test data, and the derived traits were compared with measurements calculated from manually annotated keypoints on held-out animals. The analysis was limited to geometric trait extraction and did not evaluate lameness diagnosis, biological associations with performance, or autonomous gait-cycle detection.

Materials and Methods

Experimental Setup

The source data were collected at the Beef Cattle and Sheep Laboratory, University of Illinois Urbana-Champaign, Urbana, IL. The source study recorded 189 Angus × Simmental steers. The present gait-feature analysis included 100 steers with selected side-view gait images from June 13 (n = 73) and June 14 (n = 27), 2019 (BW = 352 ± 43 kg). Images used for object-detection and pose-estimation model development were drawn from the broader source dataset, with 140 of the 189 steers represented in the computer-vision datasets used in the present study. The experimental procedures were approved by the University of Illinois Institutional Animal Care and Use Committee (IACUC protocol #19020), and all procedures were performed in accordance with the approved protocol and relevant institutional guidelines and regulations. No anesthesia or

euthanasia was used during the video-recording procedures. Following recording, the steers were returned to their normal housing at the Beef Cattle and Sheep Laboratory and resumed routine management.

Image Acquisition

Side-view videos were recorded using an Intel RealSense D435i stereo depth camera. RGB video was recorded at 30 frames/s with a resolution of 1920 × 1080 pixels; depth video was also collected in the source study, but only RGB data were used in the present analysis. Animals walked through a 10.7-m compacted-surface pathway. A 6.1-m green-painted wooden panel provided a consistent background, and five strands of non-electrified poly-wire formed a 1.8-m-wide guide path. The camera was mounted 1.05 m above the ground and positioned 6.1 m from the wooden background (Figure 1).

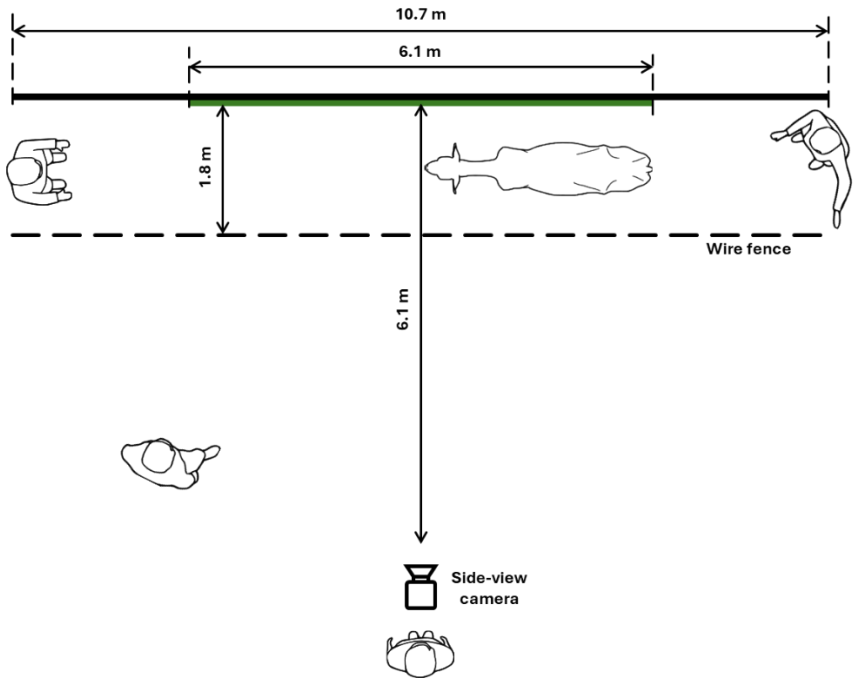


Figure 1. Recording area and camera configuration. Animals walked individually through a fenced pathway and were guided by two handlers. Side-view video was recorded with an Intel RealSense D435i camera using Intel RealSense SDK software.

Steers were guided individually through the recording area by two trained handlers. A recording was considered acceptable when the animal completed at least two relaxed walking

strides approximately perpendicular to the camera. Recordings were repeated when the animal did not walk continuously through the field of view, moved faster than a walking pace, or did not provide a suitable side view. Across the original 189-animal data collection, acceptable recordings were obtained for 94% of steers across two recording sessions. The present gait-feature analysis used 100 steers with selected gait images from June 13 (n = 73) and June 14 (n = 27), 2019.

Data Processing

RGB frames were extracted from the recorded videos using custom Python code. Near-duplicate frames were screened with the structural similarity index measure (SSIM) using a 0.85 threshold before dataset partitioning. A total of 49,920 frames were extracted. To prevent information leakage arising from multiple frames of the same animal, partitioning was performed at the animal level. The 140 steers were assigned once to training, validation, or test partitions using a 60:20:20 target split stratified by availability of detection, pose, and gait data, with a random seed of 42, resulting in 84 training, 28 validation, and 28 test animals. All images originating from a given steer were retained within the same partition throughout model development. Because annotation availability differed among computer-vision tasks, the number of animals contributing to each model differed from the full accounting cohort. The detection dataset contained 1,000 annotated images from 138 steers, of which 82, 28, and 28 animals contributed to the training, validation, and test partitions, respectively, corresponding to 518, 241, and 241 images. The pose-estimation dataset contained 1,200 annotated images from 120 steers, with 72 training, 24 validation, and 24 test animals, corresponding to 633, 267, and 300 images, respectively.

A subset of 100 steers meeting the gait-image selection criteria was retained for geometric trait extraction; this subset comprised 60 training, 20 validation, and 20 test animals. Each gait animal contributed one image for forelimb step-length extraction and one independently selected image for hock and hind-limb angle extraction, resulting in 200 gait images. All model training and data processing were conducted using Python 3.12.3 and Ultralytics 8.4.102 on a workstation equipped with an NVIDIA RTX PRO 6000 Blackwell GPU, AMD Ryzen Threadripper 7970X processor, 128 GB RAM, and Ubuntu 24.04.3 LTS.

Computer Vision Model Development

The workflow consisted of four steps: (1) cattle detection, (2) image cropping, (3) pose estimation, and (4) geometric trait extraction (Figure 2). Model evaluation was structured to prevent animal-level leakage. All images from a given steer were assigned to a single training, validation, or test partition in the final analysis. Hyperparameter selection was based on validation data only, and the held-out test partition was evaluated once after model selection. The final animal-level split was 60:20:20, with a random seed of 42.

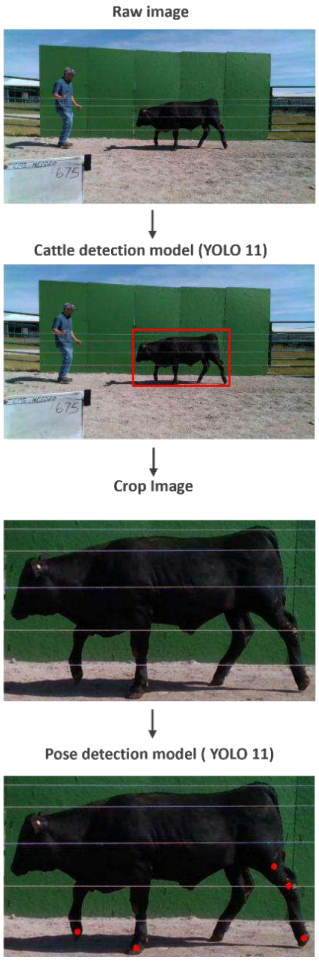


Figure 2. Computer-vision workflow. Raw images were processed by YOLO11 cattle detection, cropped using a custom Python script, processed by YOLO11 pose estimation, and used for geometric trait extraction.

Step 1: Object Detection Model

A YOLO11n object-detection model was trained to identify the cattle body region in side-view images. A total of 1,000 images were manually annotated with cattle bounding boxes in CVAT. Following the predefined animal-level partition, 82 animals were included in training, 28 in validation, and 28 in the test set, corresponding to 518, 241, and 241 images, respectively. All images from the same steer were maintained within a single partition to prevent animal-level data leakage. Images were resized to 640 × 640 pixels, and the model was initialized with pretrained YOLO11n weights. The evaluated hyperparameters included batch sizes of 16 and 32, 50 and 100 training epochs, learning rates of 0.0001, 0.0005, 0.001, 0.005, and 0.01, and Adam or SGD optimization; momentum was set to 0.9 for SGD. Early stopping used a patience of 25 epochs, and a fixed random seed of 42 was used for reproducibility. The training code also enabled deterministic execution during model development.

The model was initialized with pretrained weights and fine-tuned on the custom cattle dataset. Hyperparameter tuning was performed using only the training and validation partitions, and the test set was not used during model selection. Validation performance was evaluated using mAP@50 and mAP@50-95, and model configurations were ranked according to validation mAP@50-95. The configuration with the highest validation mAP@50-95 was selected as the final detection model and was subsequently evaluated once on the animal-level test set using mAP@50 and mAP@50-95. Training time was retained as descriptive computational information and was not used as a model-selection criterion.

The selected configuration was then evaluated once on the animal-level test set using mAP@50 and mAP@50-95. Across the evaluated configurations, the selected model used a batch size of 16, 50 epochs, a learning rate of 0.001, and SGD optimization with momentum = 0.9, achieving a validation mAP@50 of 0.995 and mAP@50-95 of 0.9603. The selected model was then evaluated once on the test set, which contained 241 images from 28 steers, achieving a test mAP@50 of 0.995 and mAP@50-95 of 0.9678.

Step 2: Image Cropping and Keypoint Annotation

Detected cattle regions were cropped before pose annotation. A total of 1,200 cropped images were annotated in CVAT with five limb keypoints. One annotator completed all annotations and reviewed the labels for obvious errors. Keypoints were annotated from individual frames; adjacent frames were not used during annotation. Frames were selected to avoid occlusion,

and poorly visible landmarks were placed using anatomical judgment. The five landmarks were the right front hoof, left front hoof, left tibial landmark, left hock, and left hind hoof. Left and right were defined from the anatomical perspective of the animal..

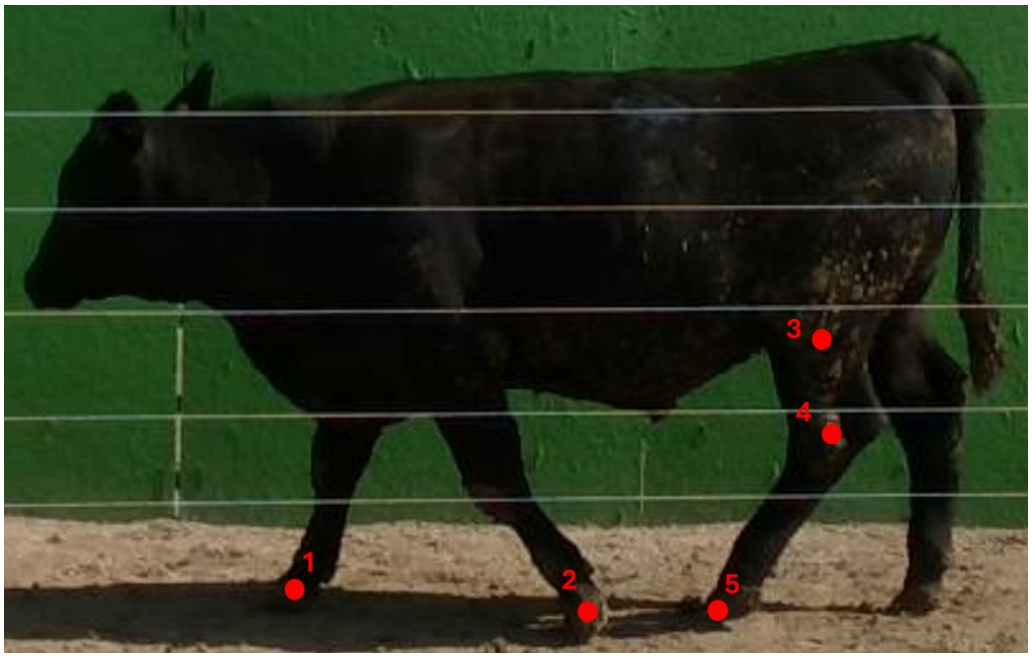


Figure 3. Annotated limb keypoints. Keypoints were defined as (1) right front hoof, (2) left front hoof, (3) left tibial landmark, (4) left hock, and (5) left hind hoof. Left and right are defined from the anatomical perspective of the animal.

Step 3: Pose Estimation Model

A YOLO11n-pose model was trained on the 1,200 annotated images using the predefined animal-level training, validation, and test partitions. The pose-estimation dataset included 120 steers, with 72 animals represented in the training set, 24 in the validation set, and 24 in the test set, corresponding to 633, 267, and 300 images, respectively. Five anatomical keypoints were defined per image, and all images were resized to 640 × 640 pixels. The model was initialized using pretrained YOLO11n-pose weights. The hyperparameter search included batch sizes of 16 and 32, 100, 150, and 200 training epochs, learning rates of 0.001, 0.01, and 0.1, and Adam or SGD optimization, with momentum = 0.9 for SGD. Early stopping was applied with a patience of 50 epochs, and a random seed of 42 was used with deterministic execution for reproducibility.

Model selection was based only on validation Pose mAP@50-95, and the test partition was not used during hyperparameter tuning. Four configurations achieved the same maximum reported validation Pose mAP@50-95 of 0.995. The tuning procedure ranked configurations according to validation Pose mAP@50-95 and retained the first top-ranked configuration, which used a batch size of 32, 200 epochs, a learning rate of 0.1, and SGD optimization with momentum = 0.9. No secondary selection criterion or test-set performance was used to resolve the tie. After selection of the final configuration, the model was evaluated once on the held-out animal-level test set. Pose performance was evaluated using Pose mAP@50 and Pose mAP@50-95, while the bounding-box component generated by the YOLO11n-pose model was also evaluated using Box mAP@50 and Box mAP@50-95.

Keypoint localization was evaluated using only the animal-level test set and the final YOLO11n pose model selected. The test set contained 300 images from 24 steers. For each image, the highest-confidence pose prediction was selected and compared with the manually annotated ground-truth keypoints. Euclidean localization error was calculated in pixels as the distance between each predicted keypoint and its corresponding ground-truth location. To account for differences in animal size within the image, the Euclidean error was normalized by the larger dimension of the ground-truth cattle bounding box, defined as the maximum between bounding-box width and height. The normalized localization error was also expressed as a percentage of this dimension. Percentage of Correct Keypoints (PCK) was calculated at thresholds of 0.05, 0.10, and 0.20, with a predicted keypoint considered correct when its normalized localization error was less than or equal to the specified threshold. Keypoint localization was summarized separately for each of the five anatomical keypoints and across all keypoints using mean, standard deviation, median, minimum, and maximum pixel error, mean normalized error, and PCK at each threshold. Distance-based localization metrics were used because they provide interpretable positional errors for downstream trait calculations.

Because the number of test images differed among animals, a sensitivity analysis calculated mean Euclidean error, mean normalized error, and PCK within each of the 24 test animals and then averaged the animal-level values with equal weight.

Computational performance was evaluated separately after model selection and final accuracy assessment using only the final detection and pose-estimation models. Inference latency was not used as a model-selection criterion. Benchmarking was performed on the NVIDIA RTX

PRO 6000 Blackwell Workstation Edition GPU using CUDA 12.8 and PyTorch 2.9.0+cu128. The models were warmed up for 10 inference passes before timing, and each held-out test image was processed five times. GPU synchronization was performed immediately before and after each timed GPU operation, and latency was measured using time.perf_counter().

Detection-only latency was benchmarked using the 241 held-out detection test images, resulting in 1,205 timed measurements. Pose-only latency was benchmarked using the 300 held-out pose test images, resulting in 1,500 timed measurements. The complete end-to-end pipeline was benchmarked using the 241 held-out detection test images, with five repetitions per image, resulting in 1,205 measurements. The end-to-end measurement included image loading, cattle detection, cropping of the detected animal, and pose estimation. Gait-trait calculation and video decoding were not included in the timed end-to-end benchmark.

Latency was summarized as mean ± standard deviation, median, minimum, and maximum milliseconds per image, and throughput was reported as frames per second (FPS), calculated from mean latency. Image loading, cattle detection, cropping, and pose estimation were also timed separately within the end-to-end pipeline to characterize the computational contribution of each processing stage.

Step 4: Geometric Locomotion Trait Extraction

A custom Python algorithm calculated three geometric traits from predicted keypoints: forelimb step length, hock angle, and hind-limb angle (Figure 4). The gait-feature dataset included 100 animals from the previously established animal-level partition, of which 60 belonged to the training partition, 20 to the validation partition, and 20 to the test partition. The original animal-level assignments were retained throughout gait-feature extraction, and images were not repartitioned during this stage. The 20 test animals remained unavailable during hyperparameter selection and were used exclusively for final model and measurement validation.

Two predefined side-view images were retained for each of the 100 animals included in the gait analysis, resulting in a total of 200 selected images. One image per animal was selected for forelimb step-length extraction, resulting in 100 forelimb step-length images. A second image was selected independently for evaluation of hind-limb geometry, resulting in 100 hock- and hind-limb-angle images. For forelimb step length, the selected frame showed the right and left

forehooves at visible ground-contact positions. For angular measurements, the second frame was selected when the left hind hoof was fully in contact with the ground. Manual frame selection was retained in this proof-of-concept study and was therefore not considered part of an autonomous end-to-end pipeline.



a **b**

Figure 4. Geometric trait definitions. (a) Forelimb step length was calculated from points 1 and 2. (b) Hock angle (A) was calculated from the segments formed by points 3–4 and 4–5, and hind-limb angle (B) was calculated from the segment formed by points 4–5 relative to the image horizontal axis.

Forelimb step length was defined operationally as the horizontal image-plane distance between the right and left forehoof ground-contact positions in the selected frame. Only horizontal displacement between points 1 and 2 was included in the calculation:

$$D_{px} = | x_2 - x_1 |$$

where D_{px} represents the horizontal distance between the two forehoof landmarks in pixels. Vertical displacement between the two landmarks was not incorporated into the step-length calculation.

Pixel distance was converted to centimeters using the RGB calibration relationship described by Condotta et al. [10]:

$$L = D_{px} \times 0.166 \times d \times 0.983$$

where L represents forelimb step length in centimeters and d represents camera-to-animal distance in meters. A fixed value of d = 4.3 m was used for all animals, calculated from the recording

geometry as the 6.1-m camera-to-background distance minus the 1.8-m pathway width. The same calibration procedure was applied to measurements derived from manually annotated and model-predicted landmarks.

Hock angle, designated angle A, was derived from points 3, 4, and 5. Point 3 represented the left tibial landmark, point 4 represented the left hock, and point 5 represented the left hind hoof. Point 4 served as the angle vertex. Two vectors were defined as:

$$u = P_3 - P_4$$

and

$$v = P_5 - P_4$$

The internal hock angle was calculated as:

$$A = \cos^{-1}\left(\frac{u \cdot v}{\| u \| \| v \|}\right) \times \frac{180}{\pi}$$

and was expressed in degrees.

Hind-limb angle, designated angle B, was calculated from the segment connecting points 4 and 5, corresponding to the left hock and left hind hoof, relative to the horizontal image axis. Horizontal and vertical segment displacement were defined as:

$$\Delta x = | x_5 - x_4 |$$

and

$$\Delta y = | y_5 - y_4 |$$

The hind-limb angle was then calculated as:

$$B = \text{atan2}(\Delta, y\Delta x) \times \frac{180}{\pi}$$

using a two-argument arctangent implementation to obtain the acute angle relative to the horizontal image axis. Both hock angle and hind-limb angle were calculated from the frame selected when the left hind hoof was fully in contact with the ground.

Manual Reference Annotation for Trait Validation

Formal validation of the extracted gait traits was restricted to the 20 animals belonging to the held-out test partition. The same predefined images subsequently used for automated trait extraction were manually annotated to establish reference measurements independent of the model predictions.

Manual landmark annotation was performed directly on the original images using an interactive Python interface. For forelimb step-length images, points 1 and 2 were manually identified. For hind-limb images, points 3, 4, and 5 were manually identified. Model-predicted landmarks were not displayed during manual annotation. Manual landmark placement was therefore completed without visualization of model predictions to reduce observer bias toward the automated landmark locations.

The manually selected coordinates were stored before automated model inference and retained as the reference annotations. The same geometric equations were then independently applied to the manually annotated coordinates and to coordinates generated by the frozen final pose-estimation model. Thus, each held-out test animal contributed paired manual and automated measurements for forelimb step length, hock angle A, and hind-limb angle B.

Validation of Extracted Traits

Agreement between manually derived and model-derived measurements was evaluated exclusively using the 20 held-out test animals. For each trait, mean absolute error (MAE) was calculated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | P_i - M_i |$$

and root mean squared error (RMSE) was calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (P_i - M_i)^2}$$

where M_i represents the manually derived measurement, P_i represents the corresponding automated measurement, and n represents the number of paired observations.

Mean signed bias was calculated as:

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)$$

such that positive values represented overestimation by the automated method and negative values represented underestimation.

Agreement between manual and automated measurements was further summarized using Lin's concordance correlation coefficient (CCC). Bland–Altman analysis was performed using paired differences calculated as:

$$d_i = P_i - M_i$$

The mean paired difference represented measurement bias, and the 95% limits of agreement were calculated as:

$$Bias \pm 1.96 \times SD_d$$

where SD_d represents the standard deviation of the paired differences. Step-length MAE, RMSE, bias, and limits of agreement were expressed in centimeters, whereas the corresponding angular statistics were expressed in degrees. Lin's CCC was dimensionless.

Descriptive Analysis of the 100-Animal Gait

After model development and formal measurement validation were completed, the frozen final pose-estimation model was applied to the complete gait dataset from all 100 animals. For each animal, one predefined forelimb image was processed to obtain forelimb step length, and one predefined hind-limb image was processed to obtain hock angle A and hind-limb angle B. Thus, each animal contributed one final value for each of the three geometric traits.

Formal agreement validation was performed exclusively on held-out test animals. Measurements from animals belonging to the training and validation partitions were included only for descriptive characterization of the complete 100-animal study population and were not included in estimates of model accuracy.

For the complete 100-animal gait cohort, descriptive statistics were calculated separately for forelimb step length, hock angle A, and hind-limb angle B. These included the number of

observations, mean, standard deviation, minimum, first quartile (Q1), median, third quartile (Q3), and maximum. Trait-to-trait correlations were not analyzed because the study focused on model performance, measurement agreement, and descriptive characterization of the extracted gait traits.

Code Availability

The custom Python code used for frame processing, model training and evaluation, keypoint-localization analysis, geometric trait extraction, measurement validation, and computational benchmarking is available from the corresponding author upon reasonable request.

Use of generative artificial intelligence

OpenAI ChatGPT was used during manuscript preparation to assist with language editing, organization, and consistency checks. All analyses, scientific interpretations, and final manuscript content were reviewed and approved by the authors, who take responsibility for the work.

Results

YOLO Detection and Pose Estimation Performance

The YOLO11n cattle detection model configuration with the highest validation mAP@50-95 used a batch size of 16, 50 training epochs, a learning rate of 0.001, and SGD optimization with momentum = 0.9. This configuration achieved a validation mAP@50 of 0.995 and mAP@50-95 of 0.9603. When evaluated once on the test set, the final model achieved a mAP@50 of 0.995 and mAP@50-95 of 0.9678 (Table 1). Detection performance was similar between the validation set and the held-out test animals.

The YOLO11n-pose model configuration with the highest validation Pose mAP@50-95 used a batch size of 32, 200 epochs, a learning rate of 0.1, and SGD optimization with momentum = 0.9. The selected configuration achieved a validation Pose mAP@50 of 0.995 and Pose mAP@50-95 of 0.9950. On the test set, Pose mAP@50 remained 0.995 and Pose mAP@50-95 was 0.9950 (Table 1). Pose-estimation performance was similar between the validation and held-out test datasets.

Table 1. Final animal-level performance of YOLO11n cattle detection and pose estimation models.

Model	Dataset	Animals	Images	mAP@50	mAP@50-95
YOLO11n Detection	Validation	28	241	0.995	0.9603
YOLO11n Detection	Test	28	241	0.995	0.9678
YOLO11n Pose	Validation	24	267	0.995	0.9950
YOLO11n Pose	Test	24	300	0.995	0.9950

The full hyperparameter search and validation results are provided in Supplementary Tables S1 and S2.

Keypoint Localization Accuracy and Computational Performance

Across all keypoints (Table 2), mean Euclidean localization error was 9.86 ± 5.76 pixels, with a mean normalized localization error of 2.46% of the larger ground-truth cattle bounding-box dimension. Overall PCK was 0.943 at a threshold of 0.05, 0.998 at 0.10, and 1.000 at 0.20. Among individual keypoints, the left hock had the lowest mean localization error (8.35 ± 4.95 pixels) and mean normalized error (2.07%). The left front hoof had a mean error of 9.49 ± 5.90 pixels, followed by point 3 at 9.91 ± 5.06 pixels and the right front hoof at 10.30 ± 5.64 pixels. The left hind hoof had the highest mean localization error (11.23 ± 6.73 pixels) and mean normalized error (2.80%). PCK@0.05 ranged from 0.897 for the left hind hoof to 0.980 for the left hock, while PCK@0.10 ranged from 0.993 to 1.000. All five keypoints reached PCK@0.20 = 1.000.

With each of the 24 test animals weighted equally, mean Euclidean error was 9.80 pixels, mean normalized error was 2.47%, and PCK was 0.945, 0.997, and 1.000 at thresholds of 0.05, 0.10, and 0.20, respectively, closely matching the image-weighted estimates.

Table 2. Keypoint localization accuracy on test-set animals.

Keypoint	N	Mean Euclidean error (px)	SD (px)	Mean normalized error (%)	PCK@0.05	PCK@0.10	PCK@0.20
(1) Right front hoof	300	10.30	5.64	2.58	0.947	1.000	1.000
(2) Left front hoof	300	9.49	5.90	2.38	0.960	0.993	1.000
(3) Left tibial landmark	300	9.91	5.06	2.49	0.930	1.000	1.000
(4) Left hock	300	8.35	4.95	2.07	0.980	1.000	1.000
(5) Left hind hoof	300	11.23	6.73	2.80	0.897	0.997	1.000
Overall	1,500	9.86	5.76	2.46	0.943	0.998	1.000

Computational benchmarking showed low inference latency for both final models. The cattle detection model required 5.68 ± 1.35 ms/image, corresponding to 175.92 frames/s, while the pose estimation model required 4.95 ± 1.46 ms/image, corresponding to 202.12 frames/s. The complete end-to-end pipeline, including image loading, cattle detection, cropping, and pose estimation, required 9.77 ± 1.62 ms/image, corresponding to 102.37 frames/s. Median latency was 5.06 ms/image for detection, 4.31 ms/image for pose estimation, and 9.12 ms/image for the end-to-end pipeline.

Agreement Between Manual and Model-Derived Gait Traits

Table 3 shows forelimb step length; the manual measurements averaged 79.01 cm, whereas the model-derived measurements averaged 83.37 cm. The mean absolute error was 4.68 cm and the RMSE was 5.52 cm. Mean signed bias was +4.36 cm, indicating that the automated method tended to overestimate forelimb step length relative to the manual reference. Lin's concordance correlation coefficient was 0.903, and the Bland–Altman 95% limits of agreement ranged from −2.45 to 11.17 cm. For hock angle A, the manual mean was 145.67°, compared with a model-derived mean of 144.76°.

The MAE was 3.93° and RMSE was 4.69°. Mean signed bias was −0.91°, indicating a small average underestimation by the automated method. Lin's concordance correlation coefficient was 0.146. Bland–Altman analysis resulted in limits of agreement from −10.17° to 8.34°.

For hind-limb angle B, the manual measurements averaged 51.60°, whereas the model-derived measurements averaged 52.57°. The MAE was 2.65° and RMSE was 3.14°. Mean signed bias was +0.97°, indicating a small average overestimation by the automated method. Lin's concordance correlation coefficient was 0.317, with Bland–Altman 95% limits of agreement ranging from −5.04° to 6.97°. Forelimb step length had the highest concordance (CCC 0.903); CCC was 0.146 for hock angle A and 0.317 for hind-limb angle B, despite small mean signed biases for both angular traits.

Table 3. Agreement between traits calculated from manual and model-predicted keypoints in the held-out test animals.

Trait	n	Manual mean	Predicted mean	MAE	RMSE	Mean bias	Lin's CCC	Bland–Altman 95% LoA
Forelimb step length (cm)	20	79.01	83.37	4.68	5.52	4.36	0.903	−2.45 to 11.17
Hock angle A (°)	20	145.67	144.76	3.93	4.69	−0.91	0.146	−10.17 to 8.34
Hind-limb angle B (°)	20	51.6	52.57	2.65	3.14	0.97	0.317	−5.04 to 6.97

Descriptive Geometric Traits

After formal validation was completed, the frozen final pose-estimation model was applied to the complete 100-animal gait cohort to characterize the distribution of the three extracted geometric traits. Measurements from training and validation animals were included only in this descriptive analysis and were not used in the agreement assessment reported in Table 3.

Mean hock angle A was 144.70 ± 1.16°, with values ranging from 141.34° to 149.23°. The first quartile, median, and third quartile were 144.00°, 144.86°, and 145.42°, respectively. Mean hind-limb angle B was 52.72 ± 2.54°, ranging from 47.24° to 61.66°. The first quartile was 51.45°, the median was 52.76°, and the third quartile was 54.08°. Forelimb step length showed greater variability across the study population than either of the angular measurements (Table 4).

Table 4. Descriptive statistics for the final extracted gait traits in the 100 animals.

Trait	n	Mean	SD	Minimum	Q1	Median	Q3	Maximum
Forelimb step length (cm)	100	82.06	12.35	47.47	74.52	83.70	89.42	104.80
Hock angle A (°)	100	144.70	1.16	141.34	144.00	144.86	145.42	149.23
Hind-limb angle B (°)	100	52.72	2.54	47.24	51.45	52.76	54.08	61.66

Discussion

Animal-level partitioning kept all frames from each steer within a single training, validation, or test set. On held-out animals, the YOLO11n cattle-detection model achieved mAP@50 = 0.995 and mAP@50-95 = 0.9678, and the YOLO11n-pose model achieved Pose mAP@50 = 0.995 and Pose mAP@50-95 = 0.9950. Previous cattle pose-estimation studies have focused primarily on dairy-cow gait and lameness assessment [4–8]. Reported performance is not directly comparable across studies because cattle populations, landmark definitions, recording conditions, model architectures, and evaluation metrics differ. Under the hardware used here, detection and pose-estimation latency averaged 5.68 ± 1.35 and 4.95 ± 1.46 ms/image, respectively, and the detection-cropping-pose pipeline averaged 9.77 ± 1.62 ms/image.

Across the held-out pose test set, mean Euclidean keypoint localization error was 9.86 ± 5.76 pixels, corresponding to 2.46% of the larger cattle bounding-box dimension. Overall PCK was 0.943, 0.998, and 1.000 at thresholds of 0.05, 0.10, and 0.20, respectively. The left hind hoof had the greatest localization error (11.23 ± 6.73 pixels) and the lowest PCK@0.05 (0.897), whereas

the left hock had the lowest localization error (8.35 ± 4.95 pixels). Agreement differed more substantially among the derived traits. Forelimb step length had a MAE of 4.68 cm, RMSE of 5.52 cm, mean bias of +4.36 cm, and Lin's CCC of 0.903. Hock angle had a MAE of 3.93°, RMSE of 4.69°, bias of −0.91°, and CCC of 0.146; hind-limb angle had a MAE of 2.65°, RMSE of 3.14°, bias of +0.97°, and CCC of 0.317. The Bland–Altman 95% limits of agreement were −2.45 to 11.17 cm for forelimb step length, −10.17° to 8.34° for hock angle, and −5.04° to 6.97° for hind-limb angle. Although mean bias was small for both angular traits, their low concordance indicates limited agreement among individual animals. Forelimb step length showed considerably better individual-animal agreement with the manual reference.

With equal weighting across the 24 test animals, mean Euclidean error was 9.80 pixels, mean normalized error was 2.47%, and PCK was 0.945, 0.997, and 1.000 at thresholds of 0.05, 0.10, and 0.20, respectively, closely matching the image-weighted estimates. Previous cattle studies have used stride-related measures, joint or back posture, and temporal locomotion features, primarily for dairy-cow lameness assessment [6,11,12]. The present study focused on establishing the measurement pipeline and quantifying the three geometric traits under controlled recording conditions. These results provide the basis for a follow-up study in which we will evaluate their relationships with locomotion and structural assessments and determine how the measurements vary across cattle populations and recording conditions.

The current workflow used manually selected frames corresponding to defined gait positions. In the next phase, we will extend the method to consecutive video frames and automate gait-phase identification, allowing repeated measurements across multiple strides for each animal. This will also allow us to evaluate within-animal repeatability and determine whether stride-level aggregation improves measurement precision.

The controlled side-view configuration used here provided a consistent setting for developing and validating the keypoint and measurement methods. A follow-up study will expand the approach to more variable environments and cattle populations, including additional breeds and recording conditions, and will incorporate independent structural or locomotion assessments as reference measurements. We also plan to evaluate annotation repeatability with multiple observers and, where feasible, compare image-derived measurements with three-dimensional reference measurements. These additions will allow us to determine how well the approach generalizes beyond the controlled conditions evaluated in the present study.

Conclusion

This proof-of-concept study established a YOLO11-based workflow for cattle detection, limb-keypoint localization, and extraction of geometric locomotion traits from controlled side-view RGB images. The models maintained high performance on held-out animals, and forelimb step length showed good agreement with measurements derived from manually annotated keypoints. Agreement was lower for the hock and hind-limb angles, indicating that these measurements require further refinement before they are used for individual-animal phenotyping. The next stage will focus on extending the workflow to repeated measurements across video sequences and evaluating its performance under more variable recording conditions.

References

1. Giess, L. K., Jensen, B. R., Weaber, R. L., Bormann, J. M. & Fiske, W. A. Feet and leg traits are moderately to lowly heritable in Red Angus cattle. *Kansas Agricultural Experiment Station Research Reports* **4**, 1 (2018). https://doi.org/10.4148/2378-5977.7533
2. Gardenier, J., Underwood, J., Weary, D. M. & Clark, C. E. F. Pairwise comparison locomotion scoring for dairy cattle. *J. Dairy Sci.* **104**, 6185–6193 (2021). https://doi.org/10.3168/jds.2020-19356
3. Tunstall, J., Mueller, K., Sinfield, O. & Higgins, H. M. Reliability of a beef cattle locomotion scoring system for use in clinical practice. *Vet. Rec.* **187**, 319 (2020). https://doi.org/10.1136/vr.105781
4. Gong, C. et al. Multicow pose estimation based on keypoint extraction. *PLoS One* **17**, e0269259 (2022). https://doi.org/10.1371/journal.pone.0269259
5. Russello, H., van der Tol, R. & Kootstra, G. T-LEAP: Occlusion-robust pose estimation of walking cows using temporal information. *Comput. Electron. Agric.* **192**, 106559 (2022). https://doi.org/10.1016/j.compag.2021.106559
6. Russello, H., van der Tol, R., Holzhauer, M., van Henten, E. J. & Kootstra, G. Video-based automatic lameness detection of dairy cows using pose estimation and multiple locomotion traits. *Comput. Electron. Agric.* **223**, 109040 (2024). https://doi.org/10.1016/j.compag.2024.109040

7. Taghavi, M., Russello, H., Ouweltjes, W., Kamphuis, C. & Adriaens, I. Cow keypoint detection in indoor housing conditions with a deep learning model. *J. Dairy Sci.* **107**, 2374–2389 (2024). https://doi.org/10.3168/jds.2023-23680
8. Wu, S. et al. Multi-scale keypoints detection and motion features extraction in dairy cows using a ResNet101-ASPP network. *J. Integr. Agric.* **25**, 2028–2040 (2026). https://doi.org/10.1016/j.jia.2024.07.023
9. Ultralytics. Ultralytics YOLO11 documentation. https://docs.ultralytics.com/models/yolo11/ (2024).
10. Condotta, I. C. F. S., Brown-Brandl, T. M., Pitla, S. K., Stinn, J. P. & Silva-Miranda, K. O. Evaluation of low-cost depth cameras for agricultural applications. *Comput. Electron. Agric.* **173**, 105394 (2020). https://doi.org/10.1016/j.compag.2020.105394
11. Telezhenko, E. & Bergsten, C. Influence of floor type on the locomotion of dairy cows. *Appl. Anim. Behav. Sci.* **93**, 183–197 (2005). https://doi.org/10.1016/j.applanim.2004.11.021
12. Van Nuffel, A. et al. Exploration of measurement variation of gait variables for early lameness detection in cattle using the GAITWISE. *Livest. Sci.* **156**, 88–95 (2013). https://doi.org/10.1016/j.livsci.2013.06.013

Acknowledgements

The authors acknowledge the staff of the University of Illinois Beef Cattle and Sheep Laboratory for their assistance with animal handling, experimental setup, data collection, and general support during the experiment.

Funding

This work was supported in part by the Illinois Beef Association, United States Department of Agriculture's National Institute of Food and Agriculture (USDA NIFA) through grants numbers 2020-67021-32799 and 2022-77038-37306 and the Hatch Act of 1887 (Accession no. 7004080).

Author contributions

L.M.B.: Conceptualization, Methodology, Software, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. C.R.D.: Investigation, Data curation. M.R.: Investigation, Data curation, Formal analysis. J.C.M.: Methodology, Supervision, Resources,

Writing – review & editing. D.S.: Conceptualization, Writing – review & editing. I.C.F.S.C.: Conceptualization, Methodology, Supervision, Project administration, Resources, Writing – review & editing. All authors reviewed and approved the final manuscript.

Data availability

The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BENICIOETAL2026SUPPLEMENTARYINFORMATION.docx