agriculture

MDPI

Article

Fusion of Target and Keypoint Detection for Automated Measurement of Mongolian Horse Body Measurements

Lide Su [1,2,3], Minghuang Li [1,2,3], Yong Zhang [1,2,3,*], Zheying Zong [1,2,3] and Caili Gong [4]

[1] College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China; sld@imau.edu.cn (L.S.); a1286862045@163.com (M.L.); zongzheying@imau.edu.cn (Z.Z.)
[2] Inner Mongolia Engineering Research Center of Intelligent Equipment for the Entire Process of Forage and Feed Production, Hohhot 010018, China
[3] Inner Mongolia Higher School Innovation Team of Research on Key Technologies of Dairy Cow Information Intelligent Sensing and Smart Farming, Hohhot 010018, China
[4] College of Electronic Information Engineering, Inner Mongolia University, Hohhot 010021, China; eegcl@imu.edu.cn
* Correspondence: yongz@imau.edu.cn; Tel.: +86-136-7474-8009

Abstract: Accurate and efficient access to Mongolian horse body size information is an important component in the modernization of the equine industry. Aiming at the shortcomings of manual measurement methods, such as low efficiency and high risk, this study converts the traditional horse body measure measurement problem into a measurement keypoint localization problem and proposes a top-down automatic Mongolian horse body measure measurement method by integrating the target detection algorithm and keypoint detection algorithm. Firstly, the SimAM parameter-free attention mechanism is added to the YOLOv8n backbone network to constitute the SimAM–YOLOv8n algorithm, which provides the base image for the subsequent accurate keypoint detection; secondly, the coordinate regression-based RTMPose keypoint detection algorithm is used for model training to realize the keypoint localization of the Mongolian horse. Lastly, the cosine annealing method was employed to dynamically adjust the learning rate throughout the entire training process, and subsequently conduct body measurements based on the information of each keypoint. The experimental results show that the average accuracy of the SimAM–YOLOv8n algorithm proposed in this study was 90.1%, and the average accuracy of the RTMPose algorithm was 91.4%. Compared with the manual measurements, the shoulder height, chest depth, body height, body length, croup height, angle of shoulder and angle of croup had mean relative errors (MRE) of 3.86%, 4.72%, 3.98%, 2.74%, 2.89%, 4.59% and 5.28%, respectively. The method proposed in this study can provide technical support to realize accurate and efficient Mongolian horse measurements.

Keywords: Mongolian horse; body measurements; YOLOv8n; convolutional neural network; attention mechanism

check for updates

Citation: Su, L.; Li, M.; Zhang, Y.; Zong, Z.; Gong, C. Fusion of Target and Keypoint Detection for Automated Measurement of Mongolian Horse Body Measurements. *Agriculture* **2024**, *14*, 1069. https://doi.org/10.3390/agriculture14071069

Academic Editor: Maciej Zaborowicz

Received: 4 May 2024
Revised: 18 June 2024
Accepted: 1 July 2024
Published: 3 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Inner Mongolia Autonomous Region is a significant area for horse breeding in China. It boasts a rich cultural heritage and favorable geographical and climatic conditions for the development of modern equine industry [1]. Mongolian horses, as a superior breed in the Inner Mongolia Autonomous Region, have a history of over a thousand years and have always been an important resource in China [2,3]. In recent years, the traditional horse breeding industry is gradually developing into a new modern industry, integrating sports, economy, and leisure [4]. The effective protection of local horse breed resources, conducting excellent breeding work, and fully exploiting excellent germplasm characteristics have become key aspects of the development process [5,6]. Researchers will test the horse's gene sequences to analyze the effects of different gene expressions on

various morphological traits [7]. The central task of animal breeding is the improvement of quantitative traits, among which body dimension parameters play an important role in Mongolian horse breeding [8,9]. They can directly reflect its growth and development status, as well as the effectiveness of breed improvement efforts. The differences in body dimension parameters may lead to changes in kinematic and dynamic parameters, thereby affecting the performance of horses [10,11]. However, the gene sequencing may not be applicable to the large-scale horse industry. Traditional body measurements often require direct manual contact with the horse using tools. The manual measurement method is not only inefficient and labor-intensive, but also lacks high automation and can easily cause stress reactions in horses. A rapid, accurate, and efficient measurement method is therefore crucial for overcoming this dilemma [12].

With the development of Precision Livestock Farming (PLF), researchers are increasingly turning to advanced technologies to replace manual methods, allowing them to obtain physiological indicators that reflect the health status of livestock in a precise and efficient way [13]. Pallottino et al. [14] proposed an approach that utilized a stereo vision system in combination with an image analysis algorithm to automatically extract the body information of a Lipizzan horse. Zhang et al. [15] developed a body measurement system for Yili horses using Matlab-GUI and the YOLACT algorithm to segment and extract Yili horses from the background, and combined this with manual markers to determine the measurement points. Freitag et al. [16] investigated the placement of markers on the left side of the horses and combined it with ImageJ 1.51r image analysis software to measure the corresponding body size parameters by manual labeling, and the average relative error of all body size measurements was controlled within 1.5%. Gmel et al. [17] used the tpsDig2 image analysis tool for the manual measurement and extraction of various body dimension parameters of Freiberger horses by operators. They then employed a restricted maximum likelihood model to estimate the heritability of body dimension parameters. Zhang et al. [18] employed a simple linear iterative clustering algorithm to obtain high-precision images of the target sheep, and localized the keypoint locations after calculating the maximum curvature of the image curve. GENG et al. [19] collected point cloud data from pigs through the utilization of dual KinectV2 cameras and calculated body parameters by employing curve fitting and point cloud slicing techniques. The above research indicates that it is feasible to use image processing techniques to obtain Mongolian horses' body dimension parameters and formulate breeding plans. However, the methods mentioned mainly rely on the manual selection of measurement points, which is complex, lacks automation, and requires expensive equipment. Therefore, a rapid, accurate, and efficient measurement method has become a key means to advance Mongolian horses' breeding processes.

In recent years, thanks to advancements in the field of human pose estimation, researchers have applied convolutional neural network-based keypoint detection methods to the real-time automatic detection of keypoints in livestock studies, achieving outstanding results. Li et al. [20] utilized the Hourglass model to locate measurement point positions in segmented images of the trunk of cows and goats. Du et al. [21] applied the DeepLabCut algorithm to automatically detect measurement keypoints for cattle and pigs, achieving body size measurements through keypoint location information. Wang et al. [22] achieved the detection of keypoints in standing pigs by constructing the HRNet (High-Resolution Network) model. Song et al. [23], addressing the issue of high network complexity in existing deep learning models for cow keypoint detection, proposed the SimCC-ShufleNetV2 lightweight cow keypoint detection model. This model has a floating-point operation volume of 0.15 G, a parameter quantity of 1.31×10^6 M, and a detection speed of 10.87 f/s, providing technical support for tasks such as cow body dimension measurement, behavior recognition, and weight estimation.

In summary, the application of computer vision combined with deep learning methods offers the potential to enhance the accuracy and efficiency of non-contact body measurements of livestock while reducing the need for human operators. Notably, there is a lack

of both domestic and international studies that utilize a keypoint detection method based on convolutional neural networks for the automatic measurement of Mongolian horses' body parameters. Therefore, this study focuses on Mongolian horses and proposes an end-to-end fusion approach for target and keypoint detection to automatically measure horse body parameters. The objective is to improve the accuracy of keypoint localization while obtaining the necessary measurements for breeding programs, and enable automatic measurements of Mongolian horse body parameters in a natural state.

2. Materials and Methods

2.1. Dataset Production

The experimental data for this study were collected in October 2022 and April and August 2023 at the Horse Breeding Technology Center of Inner Mongolia Agricultural University (40°35′ N, 110°34′ E). Adult Mongolian horses [aged (7 ± 1.5) years, weighing (485 ± 50) kg] in suitable breeding stages, were photographed while walking or standing in different scenes and time periods, and were filmed using an ORDRO V17 camera (1920 × 1080 pixels, 30 fps). Frames were intercepted from the captured videos as the base images of the dataset, and to improve the robustness of the model, the base images were enhanced by using geometric, luminance transformation and noise addition methods to simulate different situations that may occur in the real environment, and finally a dataset with a total number of 4000 Mongolian horse images was obtained; the specific number of images post data augmentation is detailed in Table 1. Using the location of the measurement points in Figure 1 as a reference, we employed the LabelMe image annotation tool to manually annotate the bounding boxes and keypoints of the target horses in the dataset, with different keypoints labeled as distinct categories.

Table 1. Image quantity distribution of Mongolian horse dataset.

Origin	Mirror	Rotate	Brightness	Noise
2500	375	375	375	375

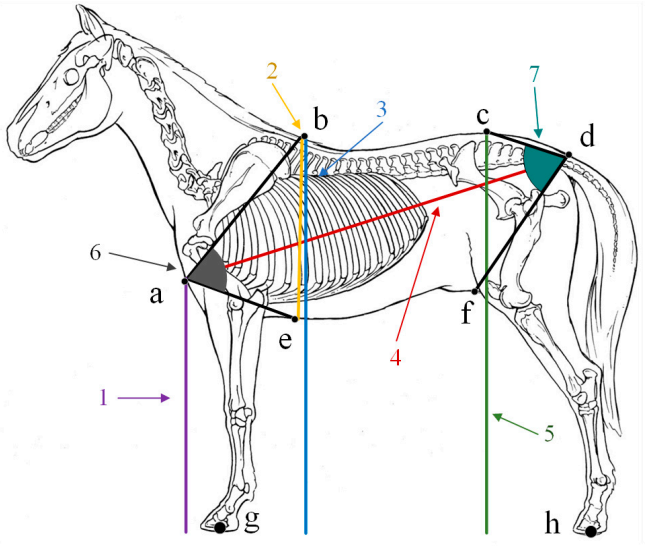
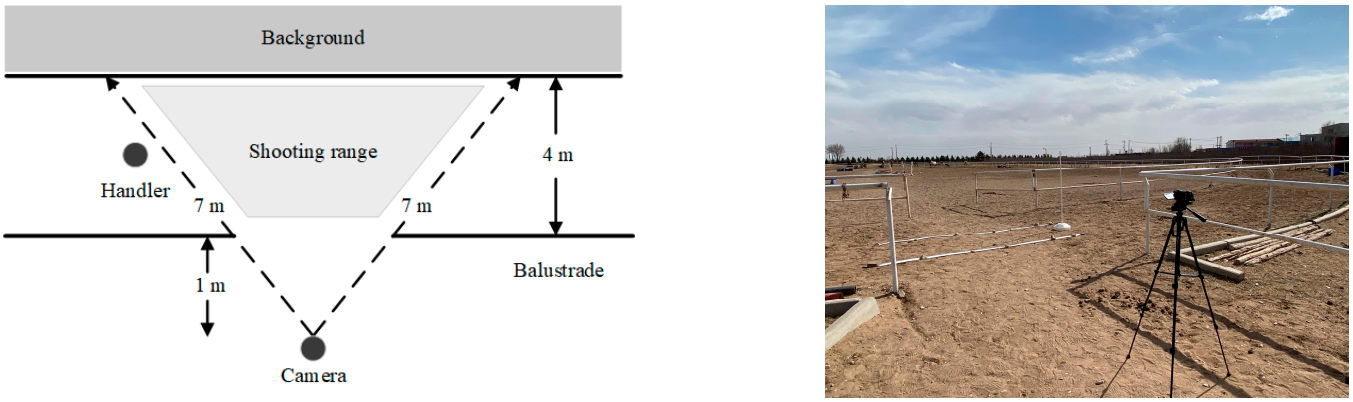


Figure 1. Illustration of measured body dimensions of Mongolian horse. 1—Shoulder height (SH); 2—Chest depth (CD); 3—Withers height (WH); 4—Body length (BL); 5—Croup height (CH); 6—Angle of shoulder (AS); 7—Angle of croup (AC).

Furthermore, we captured images of Mongolian horses as they traversed the passageway linking the barn and the training area. These images were obtained from the horses' right profiles, maintaining a standardized camera height of 110 cm and placed on a conventional tripod positioned 300 cm from the subject. To ensure the consistent

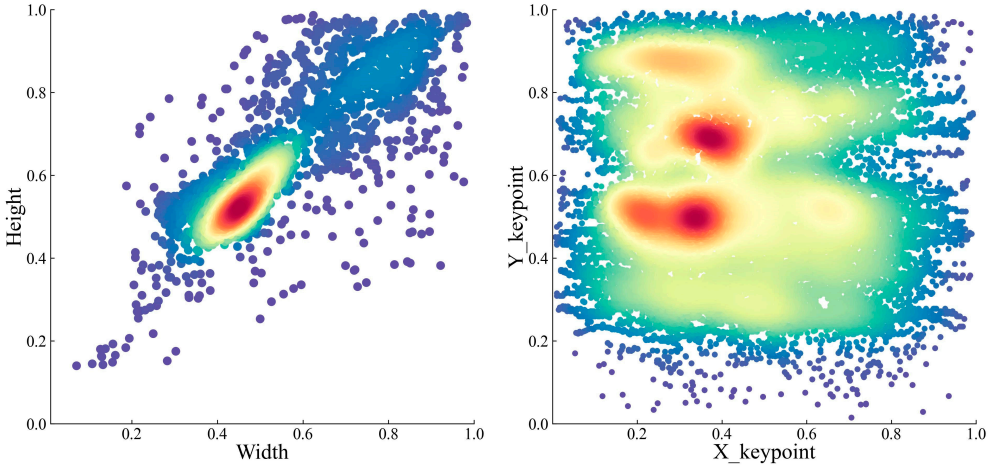
positioning of both the horses and the equipment, we marked the ground with adhesive tape. Veterinarians utilized tape measures, measuring sticks and digital goniometers to record the body size parameters of the Mongolian horses after the photographs were taken. Finally, a test set comprising complete side-view images of 20 adult Mongolian horses, along with 7 corresponding body size parameters for each horse were collected; the layout of the experiment environment is shown in Figure 2.



(**a**) Layout diagram (**b**) Real scene diagram

Figure 2. The layout of experiment environment test set filming diagram.

After the labeling was completed, the dataset was then divided into training and validation sets at a 8:2 distribution ratio. As shown in Figure 3, a visual analysis of the labeled JSON files reveals that the homemade dataset in this study primarily consists of medium- and large-sized targets, and the distribution of keypoints is relatively even, indicating a rich variety of Mongolian horse poses in the dataset.



(**a**) Distribution of target frame size (**b**) Distribution of keypoint locations

Figure 3. Results of dataset visualization.

2.2. Construction of a Model for Measuring Body Size Parameters in Mongolian Horses

The study utilized YOLOv8n as the foundational model for Mongolian horse target detection and improved both the object detection and keypoint detection models, drawing inspiration from human pose estimation principles, to excel in the task of Mongolian horse body measurement keypoint localization. The improvements were as follows:

(1) SimAM, a parameterless attention mechanism, was introduced into the original YOLOv8n backbone network. This enhancement dynamically allocated different weight co-

efficients to convolutional neural networks based on the importance of features, effectively filtering out irrelevant features, enhancing the algorithm's feature extraction capability, and improving the detection accuracy of Mongolian horses in real environments without increasing network complexity or computational resource requirements.

(2) The RTMPose keypoint detection algorithm, based on coordinate regression, was selected. This approach transformed the traditional keypoint localization problem into two classification problems for horizontal and vertical coordinates, achieving the accurate localization of Mongolian horse keypoints and facilitating subsequent body dimension measurement tasks.

(3) The cosine annealing method was employed to dynamically adjust the learning rate throughout the entire training process. This ensured more the consistent and effective convergence of the model, improving its performance.

These enhancements enabled the YOLOv8n-based model to better adapt to the task of Mongolian horse body dimension measurement, resulting in significant performance improvements.

2.3. Target Detection Algorithm Construction

A high-precision base image can reduce the probability of missing keypoint detections while minimizing the impact of environmental noise and other factors on keypoint localization accuracy. The YOLO series has been widely used in target detection tasks for its speed, low computational effort, and high accuracy. Until January 2023 the YOLO series has been updated to v8 version [24]. YOLOv8 has gained popularity among AI researchers due to its straightforward deployment process and low computational requirements. It has enjoyed widespread adoption and validation for its efficacy, thereby diminishing uncertainties during the developmental phase to a considerable extent. Furthermore, YOLOv8 is crafted to offer modularity, making it convenient for customization and extension. Considering the detection efficiency and subsequent deployment on devices with limited computational resources, YOLOv8n, which has the smallest network depth and width in YOLOv8, is chosen as the base algorithm for Mongolian horse target detection in this study. The network structure of YOLOv8n mainly contains inputs, backbone network, neck network, detect head and loss function.

To enhance the detection accuracy of the YOLOv8n algorithm, this study incorporates SimAM into the YOLOv8 backbone. Introducing an attention mechanism dynamically assigns varying weight coefficients to the convolutional neural network based on feature importance, effectively filtering out irrelevant features and enhancing target detection accuracy. Traditional attention mechanisms encompass the Spatial Attention Mechanism, Channel Attention Mechanism, and SENet. Each of these attention mechanisms brings in new parameters to the network, elevating its complexity and thereby impacting the speed and performance of the model operation. The simAM [25] Attention Module stands out as a straightforward, effective, and lightweight 3D attention module. Notably, it requires no extra parameters to be integrated into the network. Its essence is to simulate the activation method of neurons in the brain, using a three-dimensional attention module to design the energy function and enabling the network to infer the weight coefficients of the feature mapping with a small number of additional parameters added, thus assigning a higher priority to the active neurons. The schematic diagram of the SimAM attention mechanism is shown in Figure 4.

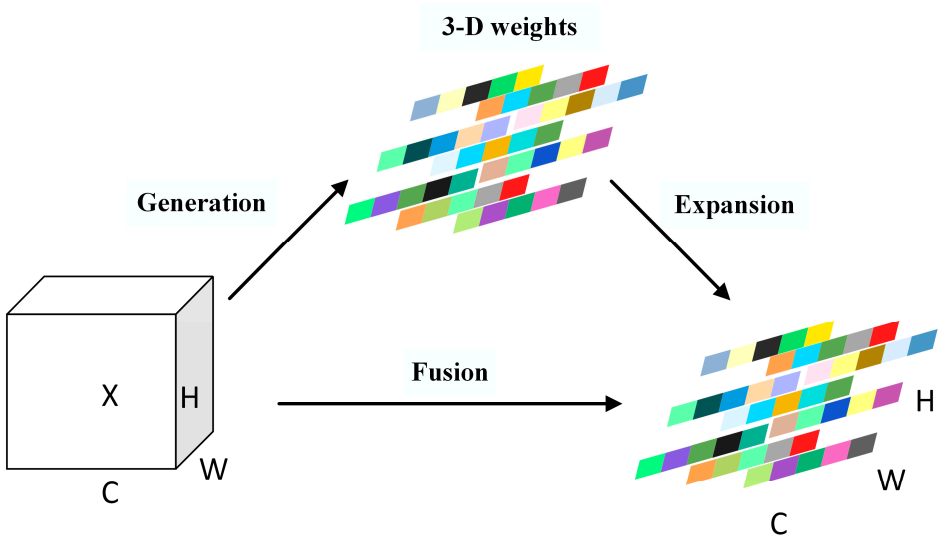


Figure 4. Schematic diagram of SimAM.

The energy function for each neuron based on neuroscience theory can be calculated using Equation (1):

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(w_t x_i + b_t))^2 + (1-(w_t t + b_t))^2 + \lambda w_t^2 \quad (1)$$

where M denotes the number of neurons in the channel of the neural network; w_t and b_t are the weights and biases of the linear transformation of t; and t and x_i denote the target neuron and other neurons in a single channel of the input feature, respectively. The minimum energy formula can be computed with:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t-\hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (2)$$

where $\hat{\mu}$ and $\hat{\sigma}$ denote the mean and variance of all neurons in a single channel in the feature mapping, respectively. As e_t^* decreases, it signifies that the neuron has a higher importance, with a larger disparity from its surrounding neurons. The final weight coefficients can be calculated as $1/e_t^*$. The entire attention module is guided by this energy function, eliminating the need for excessive heuristics and adjustments. The result of the enhanced output feature map is shown in Equation (3):

$$\widetilde{X} = sigmoid\left(\frac{1}{E}\right) \otimes X \quad (3)$$

where E represents the set of all e_t^* in the channel and spatial dimensions; the *sigmoid* function is added to restrict values too large in E; and $\otimes$ represents Hadamard's product.

In this study, the SimAM was integrated into the YOLOv8n backbone network to enhance the feature extraction capability of the algorithm, which improves the detection accuracy of Mongolian horses in real environments without increasing network complexity or computational resource demands [24]. The network structure of SimAM–YOLOv8n is depicted in Figure 5.

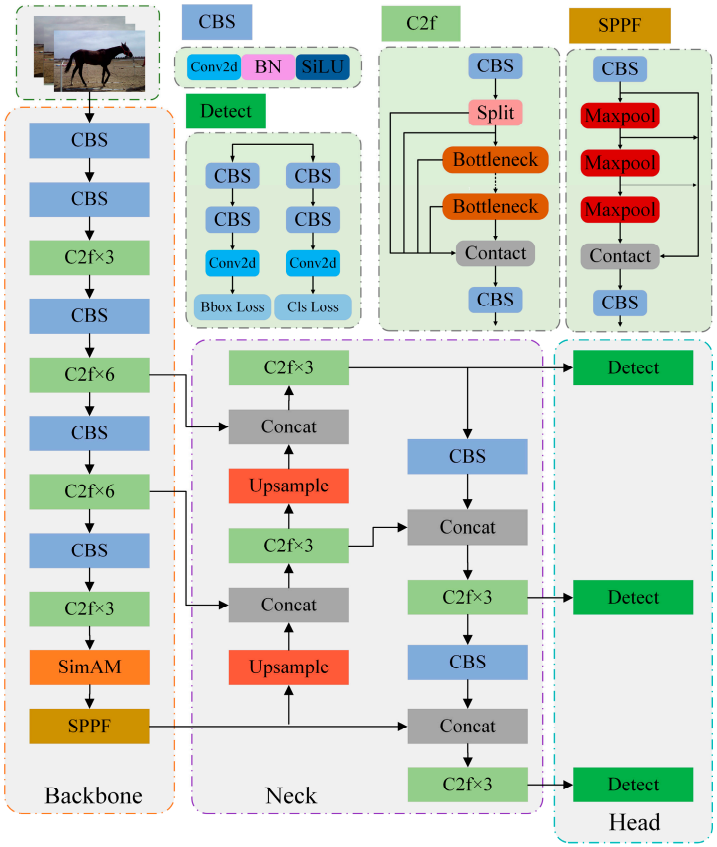


Figure 5. SimAM–YOLOv8n network structure.

2.4. Keypoint Detection Algorithm Construction

Keypoint detection is a significantly important research area in image processing, primarily employed to locate critical positions in images or videos. This study draws on the idea of human pose estimation to improve the human keypoint detection model so that it can performs equally well in the task of keypoint localization for Mongolian horse body measurements.

RTMPose, proposed by Jiang et al. [26], utilized the idea of coordinate classification to achieve human keypoint detection [27]. The keypoint detection algorithm requires multiple convolutions or downsampling operations, which inevitably leads to a smaller size of the generated feature image, and a certain degree of quantization error occurs when the coordinates are finally mapped back to the original image. The algorithm transforms the traditional keypoint localization task problem into two classification problems of horizontal and vertical coordinates. One-dimensional vectors in the x and y directions of the image, with a length greater than the original image size, are used to characterize the keypoints, i.e., n one-dimensional vectors (corresponding to n keypoints) are output, which are then transformed into n coordinate representations by linear projection. Thus, subpixel localization accuracy is achieved, and quantization error is effectively avoided.

The structure of the RTMPose network is shown in Figure 6. After expanding the feature information extracted from the backbone network from (n, H', W') to $(n, H' \times W')$, the fully connected (FC) layer is utilized to expand the representation form of the one-dimensional keypoint to the desired dimensions controlled by hyperparameters, and the Gated Attention Unit (GAU) is used to further improve the model's perception of the feature information. The realization process is shown below:

$$A = \frac{1}{n} relu^2\left(\frac{Q(X)K(Z)^\top}{\sqrt{s}}\right), Z = \phi_z(XW_z) \tag{4}$$

where Q and K are simple linear transformations, s = 128.

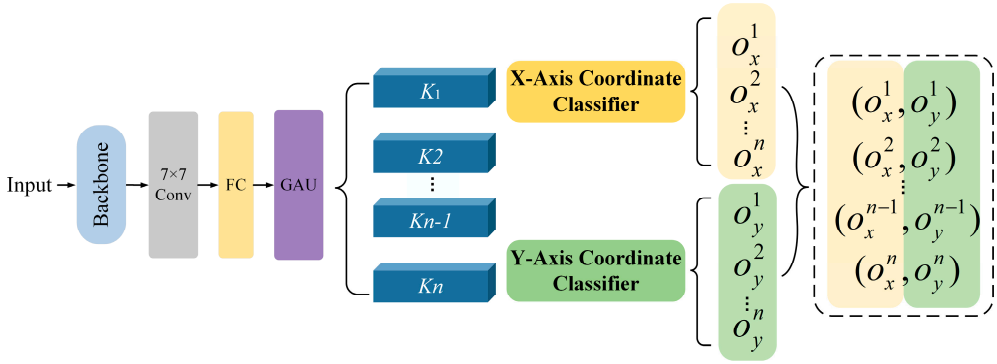


Figure 6. RTMPose network structure.

Then, the keypoint information is classified using a X-axis coordinate classifier and a Y-axis coordinate classifier. To implement the process, the coordinate values are unified as integers as class labels for model training. The process is shown below:

$$c_x \in [1, N_x], c_y \in [1, N_y] \tag{5}$$

where $N_x = W \cdot k$, $N_y = H \cdot k$ are the number of bins on the X and Y axis, respectively, and k is the scaling factor ($k \geq 1$). Its function is to reduce the quantization error to achieve subpixel level positioning accuracy of the coordinates.

3. Experiments and Results

3.1. Test Environment and Experimental Configuration

All the experiments in this study were conducted under a unified environment. The test environment was based on Ubuntu 20.04 operating system equipped with Intel Core i7-9700K and Nvidia GeForce RTX 2080ti (USA). The acceleration environment was CUDA 11.3, CUDNN 8.2.1, the deep learning framework was PyTorch 1.11.0, and Python 3.8 was selected as programming language.

In the training process of SimAM–YOLOv8n, the size of input image was 640 × 640, the batch size was 16, the initial learning rate was 0.01, the learning rate momentum was 0.937, the weight decay coefficient was 0.0005, the SGD was used as the optimizer, and the training epoch was set to 150. The RTMPose training parameters are set as follows: the training image size was 256 × 256, the batch size was set to 16, the initial learning rate was 0.004, the weight decay was 0.05, AdamW was used as the optimizer, and the training epoch was set to 200.

Owing to the substantial volume of training iterations, the cosine annealing method is employed to dynamically modulate the learning rate throughout the training regimen, thereby ensuring a more consistent and efficient convergence of the model [28]. The computational procedure for the cosine annealing method is shown in Equation (6).

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})[1 + \cos(\frac{T_{cur}}{T_i}\pi)] \tag{6}$$

where η_{max} and η_{min} represents the maximum and minimum values of the learning rate, respectively; T_{cur} represents the number of epochs that have been executed; and T_i represents the total number of epochs in the ith run.

3.2. Evaluation Indicators

This study comprises two models: one for target detection and one for keypoint detection, each with different computational foundations. To distinguish between them,

AP-obj was set as the evaluation metric for the target detection experiment, as illustrated in Equations (7)–(9).

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$AP = \int_0^1 P(R)dR \tag{9}$$

where TP is the number of correct positive samples detected; FP is the number of false positive samples detected; and FN is the number of false negative samples detected. AP-obj is the AP value at different Intersection over Union (IoU) thresholds, spanning from 0.5 to 0.95 in increments of 0.05.

The evaluation metric chosen for the keypoint identification experiment was AP-kp. Its computation is grounded in Object Keypoint Similarity (OKS):

$$OKS_p = \frac{\sum_i exp\left\{-d_{pi}^2/(2S_p^2\sigma_{pi}^2)\right\}\delta(v_{pi} > 0)}{\sum_i \delta(v_{pi} > 0)} \tag{10}$$

where pi is the ID of the keypoint; v_{pi} is the visibility of the keypoint; d_{pi}^2 is the Euclidean distance between the detected keypoint and the corresponding labeled keypoint; S_p is the scale of the target bounding box; and σ_i is the normalization factor of the ith keypoint.

In this phase, OKS is equivalent to the IoU value in the target detection experiment, and a higher threshold represents more accuracy and consistency between the predicted keypoints and the labeled keypoints [29]. In this study, AP-kp values are calculated when the threshold value ranges from 0.5 to 0.95 with a step size of 0.05. Additionally, the number of parameters, floating-point operations per second (FLOPs), and model size are used for evaluating the algorithm computational requirements and complexity.

3.3. SimAM–YOLOv8n Performance Validation

The effectiveness and superiority of the enhanced algorithm proposed in this study were verified and a performance analysis of various algorithms was conducted. The selected algorithms for comparison encompass the prevailing target detection methods: Faster RCNN, SSD, YOLOv5n, YOLOv7-tiny, and YOLOv8n. The experimental results are presented in Table 2.

Table 2. Comparison of performance for different target detection algorithms.

Model	Parameter ($\times 10^6$)	FLOPs (G)	Model Size (MB)	AP-obj (%)
Faster RCNN	41.36	172.3	159.6	85.8
SSD	26.17	88.2	262.3	85.4
YOLOv5n	2.66	7.8	5.9	84.6
YOLOv7-tiny	6.01	13.2	23.9	85.7
YOLOv8n	3.01	8.1	6.2	86.6
SimAM–YOLOv8n	3.01	8.1	6.2	90.1

The results demonstrate that SimAM–YOLOv8n achieves the highest AP-obj value. Moreover, in comparison to the other models, SimAM–YOLOv8n boasts a more compact set of parameters, reduced computational demands, and lighter model weights. These results suggest that the algorithm excels in detection efficiency and is well-suited for real-time detection applications.

3.4. RTMPose Performance Validation

Based on the self-constructed dataset, RTMPose and three mainstream keypoint detection models were trained and validated, and the accuracy curves of the four models during the training process are shown in Figure 7, which shows that RTMPose achieves convergence earlier than other algorithms. The algorithm performance validation results are shown in Table 3.

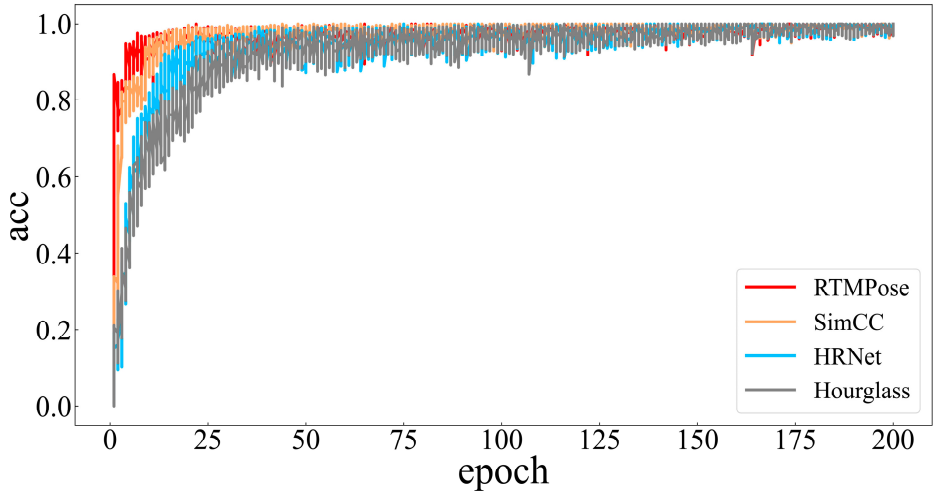


Figure 7. Training process for different keypoint detection algorithms.

Table 3. Comparison of performance for different keypoint detection algorithms.

Model	Parameter ($\times 10^6$)	FLOPs (G)	Model Size (MB)	AP-obj (%)
Hourglass	94.84	28.65	363.5	89.1
HRNet	28.52	16.78	110.8	89.6
SimCC	26.74	6.76	142.2	91.8
RTMPose	5.33	0.89	26.5	91.4

From the comparison results, it can be seen that the AP-kp value of RTMPose was 91.4%, which was 2.3% and 1.8% higher compared to Hourglass and HRNet, respectively, and although it was 0.4% lower compared to SimCC, the parameters, FLOPs, and the size of the model weights are reduced by 21.41 M, 5.87 G, and 115.7 MB, respectively, which comprehensively shows that the RTMPose required fewer computational resources and the model complexity is low, and it achieves a good balance between accuracy and speed.

3.5. Body Measurements Accuracy Verification

The detection results of each keypoint and the heatmap representation results after inputting the horse image to be measured into the algorithmic model proposed in this study are shown in Figure 8, from which the algorithm proposed in this study can effectively determine and classify the location and category of each measurement keypoint. Finally, the automated calculation of individual body measurements is predicated on the coordinate data associated with keypoints across distinct categories.

The conversion factor (CF) was determined through the analysis of the height contrast stick positioned within the passageway using image processing software (ImageJ 1.51 r, National Institute of Mental Health, USA). The calculation procedure is elucidated in Equations (11).

$$CF = \frac{real_{stick}}{pixel_{stick}} \tag{11}$$

where $real_{stick}$ represents the true length of the stick; and $pixel_{stick}$ represents the pixel length of the stick in the figure. A comparison of the manual and modeled measurements of body parameters is shown in Figure 9. The MRE of shoulder height, chest depth, body height, body length, croup height, shoulder angle of shoulder and croup angle of croup were determined to be 3.86%, 4.72%, 3.98%, 2.74%, 2.89%, 4.59% and 5.28%, respectively. The results show that the method proposed in this study can be used as a non-contact method for the automatic measurement of equine body size.

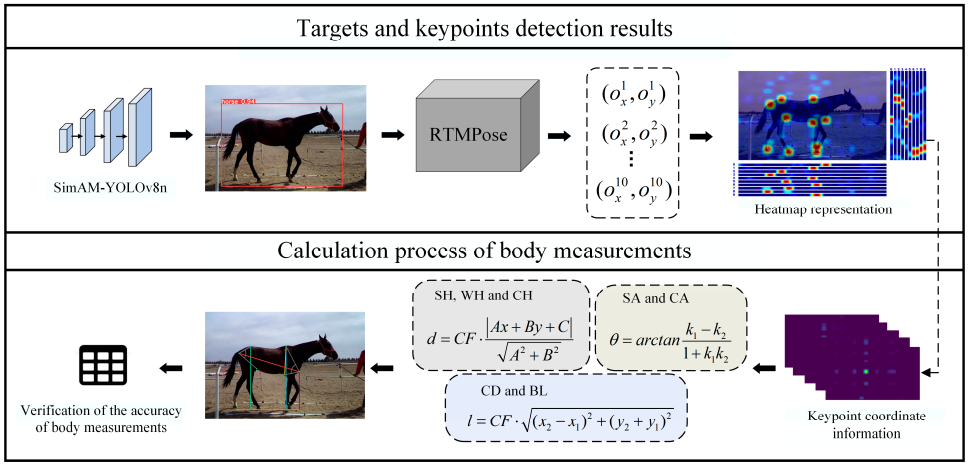
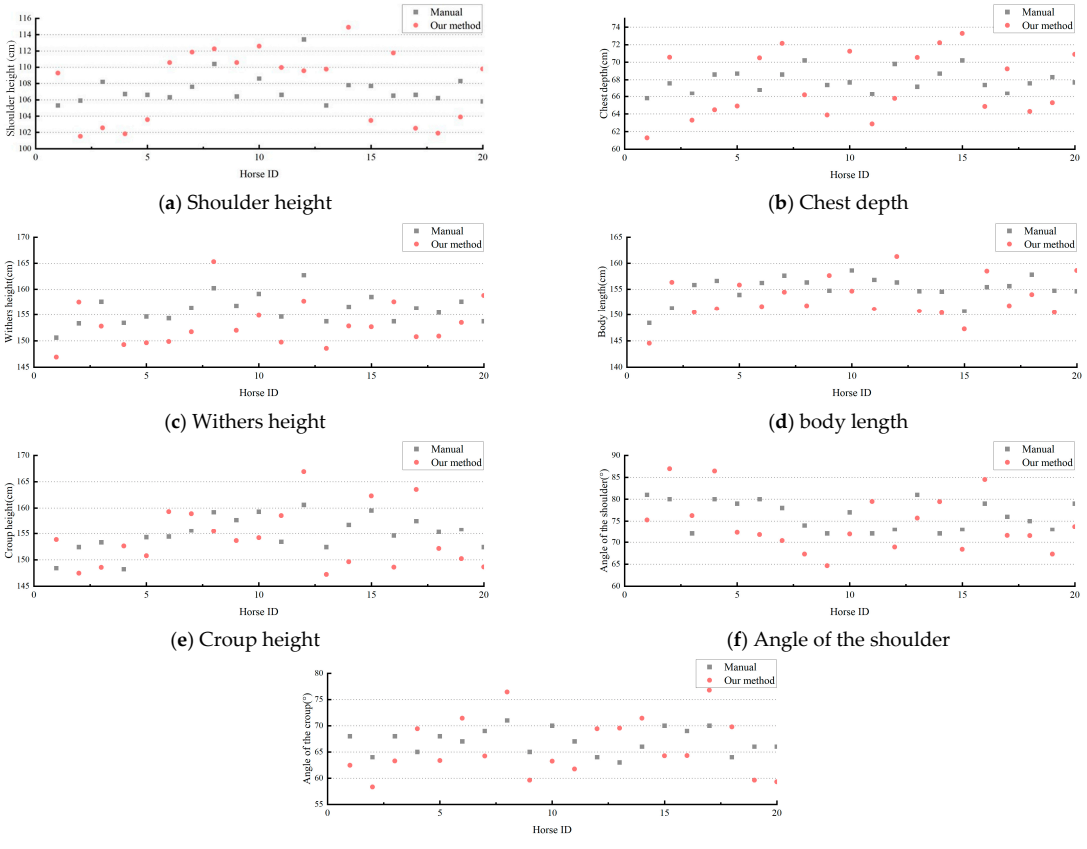


Figure 8. Technological roadmap.



(**a**) Shoulder height

(**b**) Chest depth

(**c**) Withers height

(**d**) body length

(**e**) Croup height

(**f**) Angle of the shoulder

(**g**) Angle of the croup

Figure 9. Comparison of manual measurement and our method for the seven different body parameters.

4. Discussion and Conclusions

Given the significance of obtaining the morphological parameters of horses and the limited research on automating this process using deep learning methods, this paper proposes a deep learning-based method for the automatic measurement of Mongolian horse body size and conformation, aiming to be able to obtain equine morphometric measurements in an efficient and accurate way. However, we also noticed the shortcoming in our method. Additionally, only one side of the Mongolian horse is assessed, which means that traits such as body width, heart girth and cannon bone girth, cannot be evaluated with a 2D camera. Some 3D equipment is already available in the context of equine performance laboratories; however, the equipment is challenging to operate, and the manual placing of anatomical landmarks is time-consuming, which therefore limits implementation in a routine measurement. Although animal body measurements based on 3D reconstruction have been researched, the real-time performance is poor and there is a high demand for operating equipment, and it mainly relies on the horse's side image in the horse body measurement task. Additionally, since the dataset of this study was only for Mongolian horses, its applicability to the other horse breeds remains to be verified.

This study proposes an end-to-end fusion of target and keypoint detection for body measurements, which can realize the measurement task with low computational cost and provides technical support and a theoretical basis for the development of subsequent mobile devices for horse body measurements. In the original YOLOv8n backbone network, the SimAM parameterless attention mechanism is introduced, the coordinate regression-based RTMPose keypoint detection algorithm is selected, and the cosine annealing method is used to dynamically adjust the learning rate so as to ensure a more consistent and effective convergence of the model. The experimental results show that the parameters of SimAM–YOLOv8n model are reduced by 38.35, 23.16, and 3 M compared to Faster RCNN, SSD, and YOLOv7-tiny, respectively; the FLOPs of the SimAM–YOLOv8n model are reduced by 164.2, 80.1, and 5.1 G compared to Faster RCNN, SSD, and YOLOv7-tiny, respectively; the model size of SimAM–YOLOv8n model is reduced by 153.4, 256.1, and 17.7 MB compared to Faster RCNN, SSD, and YOLOv7-tiny, respectively; and the AP-obj of the SimAM–YOLOv8n model increases by 4.3, 4.7, 5.5, 4.4, and 3.5 percentage points compared to Faster RCNN, SSD, YOLOv5n, YOLOv7-tiny, and YOLOv8n, respectively. The parameters of the keypoint detection model RTMPose are reduced by 89.51, 23.19, and 21.41 M compared to Hourglass, HRNet, and SimCC, respectively; the FLOPs of the keypoint detection model RTMPose are reduced by 27.76, 15.89, and 5.87 G compared to Hourglass, HRNet, and SimCC, respectively; the model size of the keypoint detection model RTMPose is reduced by 337, 84.3, and 115.7 MB compared to Hourglass, HRNet, and SimCC, respectively; and the AP-kp value of RTMPose was 91.4%, which was 2.3% and 1.8% higher compared to Hourglass and HRNet, respectively, and it was 0.4% lower compared to SimCC. Compared with the manual measurements the shoulder height, chest depth, body height, body length, croup height, angle of shoulder and angle of croup had mean relative errors (MRE) of 3.86%, 4.72%, 3.98%, 2.74%, 2.89%, 4.59% and 5.28%, respectively.

The measurement of horse body parameters through machine vision technology has been corroborated by relevant studies. In one instance, researchers utilized dual cameras for linear and angular measurements on Lipizzan horses. The manual and visual measurements displayed a strong overall correlation among operators ($r = 0.998$) with an average error rate of less than 3%. Nonetheless, this study was limited to a small sample size of only 10 horses [14]. To enhance the precision of body measurements, researchers adhered high-contrast stickers at designated points on horses for measurements. Subsequently, the side images of the horses were captured with cameras, and the body size parameters were determined through image processing and related procedures. The Pearson correlation coefficient between manual and automated systems reached 0.999. Nonetheless, the manual placement of stickers in this method was prone to human factors, and the markers were susceptible to dislodging when the horses were in motion, which could adversely affect body measurements [16]. In another study, a digital 3D modeling-

based method was employed to measure the body parameters of five Pura Raza Española horses. A comparison with human measurements revealed that 88% of the system's samples had an average relative error of less than 20% [30]. By comparing the non-contact measurement methods of different horse breeds, it can be found that the method proposed in this paper has the advantage of automatic measurement of parameters, even though it is slightly lower than manual measurement in terms of measurement accuracy. Therefore, future research should focus on expanding the variety of body parameters obtained from automated measurements and further improving the measurement accuracy. This will help to enhance the reliability and accuracy of automated measurement techniques in practical applications.

Furthermore, to identify where the model underperforms in specific scenarios or keypoints, a detailed error analysis was conducted. This analysis aids in targeting improvements and understanding the limitations of the current method. By examining the error distribution across different keypoints, we found that certain keypoints exhibit higher error rates. Specifically, in dynamic scenarios (e.g., when the horse is walking or moving), shoulder height and hip height errors are relatively higher. This may be due to significant posture changes during movement, which increases the difficulty of accurately localizing these keypoints. Additionally, shoulder and hip angles are particularly sensitive to variations in lighting and background clutter. In images with inconsistent lighting or more complex backgrounds, these keypoints tend to show greater errors. Simultaneously, in manual measurements, reference points are susceptible to disturbances from both horses and testers, potentially leading to measurement inaccuracies [31]. This suggests that the model needs further optimization to improve robustness under such varying conditions.

The deep learning-based automatic measurement of equine morphological parameters proposed in this paper is an effective and valid tool for farmers, breeding enterprises and research scholars. Additionally, our method can be carried out in the daily life of horses, which greatly safeguards the welfare of the horses and reduces the risks of injury for the evaluator; with the development trend of precision livestock farming (PFL), we can apply the method to the establish the database to allow better Mongolian horse breeding programs.

Author Contributions: Conceptualization, methodology, writing—original draft preparation, and investigation, L.S.; validation, formal analysis, and writing—review and editing, M.L., Z.Z. and C.G.; and writing—review and editing, supervision, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Natural Science Foundation of China (32360856), and The Natural Science Foundation of Inner Mongolia Autonomous Region (2022QN03019), and The Scientific Research Program of Higher Education Institutions in Inner Mongolia Autonomous Region (NJZY22516), and The Innovation Team of Higher Education Institutions in Inner Mongolia Autonomous Region (NMGIRT2312).

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the fact that our study did not cause any stress to the horses and there was no contact with the horses during the collection process, which would not affect the normal activities of the horses.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, X.; Zhu, X.; Liu, R. Inner Mongolia Horse Industry Development Path Analysis. *Mod. Anim. Husb. Sci. Technol.* **2022**, *12*, 126–128.
2. Wang, Q.; Zou, Y. China's Equine Industries in a Transitional Economy: Development, Trends, Challenges, and Opportunities. *Sustainability* **2020**, *12*, 5135. [CrossRef]

3. Mang, L.; Bai, D.Y. Analysis of the current situation of the horse industry in Inner Mongolia autonomous region. *North. Econ.* **2019**, *11*, 20–25.
4. Huang, W.; Li, Y.; Zhang, Z. Research on the Development Path of China's Horse Industry from the Perspective of Supply Side Structural Reform. *Contemp. Sports Technol.* **2021**, *11*, 191–197.
5. Cao, X.J.; Wang, H.D.; Wang, Y. Countermeasures for the development of China's horse industry based on SWOT analysis. *Heilongjiang Anim. Sci. Vet. Med.* **2020**, *10*, 23–28.
6. Rosengren, M.K.; Sigurðardóttir, H.; Eriksson, S.; Naboulsi, R.; Jouni, A.; Novoa-Bravo, M.; Albertsdóttir, E.; Kristjánsson, Þ.; Rhodin, M.; Viklund, Å. A QTL for conformation of back and croup influences lateral gait quality in Icelandic horses. *BMC Genom.* **2021**, *22*, 267. [CrossRef]
7. Meira, C.T.; Curi, R.A.; Silva, J.A.I.V. Morphological and genomic differences between cutting and racing lines of Quarter Horses. *J. Equine Vet. Sci.* **2013**, *33*, 244–249. [CrossRef]
8. Zheng, H.; Fang, C.; Zhang, T.; Zhao, H.; Yang, J.; Ma, C.J.C. Shank length and circumference measurement algorithm of breeder chickens based on extraction of regional key points. *Comput. Electron. Agric.* **2022**, *197*, 106989. [CrossRef]
9. Ghezelsoflou, H.; Hamidi, P.; Gharahveysi, S. Study of factors affecting the body conformation traits of Iranian Turkoman horses. *J. Equine Sci.* **2018**, *29*, 91–96. [CrossRef]
10. Paksoy, Y.; Ünal, N. Multivariate analysis of morphometry effect on race performance in Thoroughbred horses. *Rev. Bras. De Zootec.* **2019**, *48*, e20180030. [CrossRef]
11. Meira, C.; Fortes, M.M.; Farah, M.R.S.; Porto-Neto, L.; Curi, R.; Moore, S.; Mota, M. A genome-wide association study for height at withers in racing quarter horse. *J. Equine Vet. Sci.* **2013**, *20*, 420–423.
12. Minero, M.; Canali, E. Welfare issues of horses: An overview and practical recommendations. *Ital. J. Anim. Sci.* **2009**, *8*, 219–230. [CrossRef]
13. Berckmans, D. Precision livestock farming technologies for welfare management in intensive livestock systems. *Rev. Sci. Tech.* **2014**, *33*, 189–196. [CrossRef] [PubMed]
14. Pallottino, F.; Steri, R.; Menesatti, P.; Antonucci, F.; Costa, C.; Figorilli, S.; Catillo, G. Comparison between manual and stereovision body traits measurements of Lipizzan horses. *Comput. Electron. Agric.* **2015**, *118*, 408–413. [CrossRef]
15. Zhang, J. Measurement and Design of Horse's Body Size Based on Deep Learning and Image Processing Technology. *Comput. Technol. Dev.* **2020**, *30*, 180–184+189.
16. Freitag, G.P.; de Lima, L.G.F.; Jacomini, J.A.; Kozicki, L.E.; Ribeiro, L.B. An accurate image analysis method for estimating body measurements in horses. *J. Equine Vet. Sci.* **2021**, *101*, 103418. [CrossRef] [PubMed]
17. Gmel, A.I.; Burren, A.; Neuditschko, M. Estimates of Genetic Parameters for Shape Space Data in Franches-Montagnes Horses. *Animals* **2022**, *12*, 2186. [CrossRef]
18. Zhang, A.L.; Wu, B.P.; Wuyun, C.T.; Jiang, D.X.; Xuan, E.C.; Ma, F.Y. Algorithm of sheep body dimension measurement and its applications based on image analysis. *Comput. Electron. Agric.* **2018**, *153*, 33–45. [CrossRef]
19. Geng, Y.; Xiaodong, Y.; Yankai, J.; Yanbo, L.; Yanfang, F.; Shucai, Y. Research on pig body size measurement system based on stereo vision. *INMATEH-Agric. Eng.* **2023**, *70*, 76. [CrossRef]
20. Li, K.; Teng, G. Study on body size measurement method of goat and cattle under different background based on deep learning. *Electronics* **2022**, *11*, 993. [CrossRef]
21. Du, A.; Guo, H.; Lu, J.; Su, Y.; Ma, Q.; Ruchay, A.; Marinello, F.; Pezzuolo, A. Automatic livestock body measurement based on keypoint detection with multiple depth cameras. *Comput. Electron. Agric.* **2022**, *198*, 107059. [CrossRef]
22. Wang, X.; Wang, W.; Lu, J.; Wang, H. HRST: An Improved HRNet for Detecting Joint Points of Pigs. *Sensors* **2022**, *22*, 7215. [CrossRef]
23. Song, H.; Hua, Z.; Ma, B.; Wen, Y.; Kong, X.; Xu, X. Lightweight Keypoint Detection Method of Dairy Cow Based on SimCC-ShuffleNetV2. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 275–281+363.
24. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [CrossRef]
25. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR 2021, Virtual, 18–24 June 2021; pp. 11863–11874.
26. Jiang, T.; Lu, P.; Zhang, L.; Ma, N.; Han, R.; Lyu, C.; Li, Y.; Chen, K. RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose. *arXiv* **2023**, arXiv:2303.07399.
27. Li, Y.; Yang, S.; Liu, P.; Zhang, S.; Wang, Y.; Wang, Z.; Yang, W.; Xia, S.-T. Simcc: A simple coordinate classification perspective for human pose estimation. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 89–106.
28. Loshchilov, I.; Hutter, F. Stochastic gradient descent with warm restarts. In Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
29. Wu, Z.; Xia, F.; Zhou, S. A method for identifying grape stems using keypoints. *Comput. Electron. Agric.* **2023**, *209*, 107825. [CrossRef]

30. Pérez-Ruiz, M.; Tarrat-Martín, D.; Sánchez-Guerrero, M.J.; Valera, M. Advances in horse morphometric measurements using LiDAR. *Comput. Electron. Agric.* **2023**, *174*, 105510. [CrossRef]
31. Santos, M.R.; Freiberger, G.; Bottin, F.; Chiocca, M.; Zampar, A.; Cucco, D.C. Evaluation of methodologies for equine biometry. *Livest. Sci.* **2017**, *206*, 24–27. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.