



# Article Division of Cow Production Groups Based on SOLOv2 and Improved CNN-LSTM

Guanying Cui<sup>1,†</sup>, Lulu Qiao<sup>1,†</sup>, Yuhua Li<sup>1,\*</sup>, Zhilong Chen<sup>1</sup>, Zhenyu Liang<sup>1</sup>, Chengrui Xin<sup>2</sup>, Maohua Xiao<sup>2</sup> and Xiuguo Zou<sup>1,3</sup>

- <sup>1</sup> College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China; 9203020903@stu.njau.edu.cn (G.C.); 19220117@stu.njau.edu.cn (L.Q.); 2021819069@stu.njau.edu.cn (Z.C.); 2022819080@stu.njau.edu.cn (Z.L.); xzou05@mail.ubc.ca (X.Z.)
- <sup>2</sup> College of Engineering, Nanjing Agricultural University, Nanjing 210031, China; 2021112042@stu.njau.edu.cn (C.X.); xiaomaohua@njau.edu.cn (M.X.)
- <sup>3</sup> Faculty of Applied Science, University of British Columbia, Kelowna, BC V1V 1V7, Canada
- \* Correspondence: lyhresearch@njau.edu.cn; Tel.: +86-25-58606585
- <sup>+</sup> These authors contributed equally to this work.

Abstract: Udder conformation traits interact with cow milk yield, and it is essential to study the udder characteristics at different levels of production to predict milk yield for managing cows on farms. This study aims to develop an effective method based on instance segmentation and an improved neural network to divide cow production groups according to udders of high- and low-yielding cows. Firstly, the SOLOv2 (Segmenting Objects by LOcations) method was utilized to finely segment the cow udders. Secondly, feature extraction and data processing were conducted to define several cow udder features. Finally, the improved CNN-LSTM (Convolution Neural Network-Long Short-Term Memory) neural network was adopted to classify high- and low-yielding udders. The research compared the improved CNN-LSTM model and the other five classifiers, and the results show that CNN-LSTM achieved an overall accuracy of 96.44%. The proposed method indicates that the SOLOv2 and CNN-LSTM methods combined with analysis of udder traits have the potential for assigning cows to different production groups.

**Keywords:** cow udder classification; udder features; instance segmentation; CNN-LSTM; udder conformation

# 1. Introduction

Milk and cow products are essential foods for daily life, which provide vital proteins required by humans [1]. Dairy is integral to China's modern agriculture and food industry and indispensable for a healthy China. With the improvement in quality of life, the scale of the dairy industry, milk production, and milk consumption are increasing. Cow breeding is the first step in the milk industry chain and is a prerequisite for obtaining high-quality milk production. In recent years, with economic growth, cow farming in China has gradually shifted from a traditional family-based mode to an intensive, large-scale, and facility-based mode [2]. However, there are still areas for improvement in farming management techniques because cows with different levels of milk production are often managed similarly by farmers, who are unable to manage high-yielding cows based on their characteristics, which affects milk yield and quality. Therefore, a reasonable grouping of cows in production areas based on milk production areas, such as forage to concentrate ratios and exercise levels, are important to promote the development of the milk industry in China.

Numerous studies have found a correlation between milk production and udder traits in cows. Pawlina et al. [3] found an increase in udder and teat size and a decrease in



Citation: Cui, G.; Qiao, L.; Li, Y.; Chen, Z.; Liang, Z.; Xin, C.; Xiao, M.; Zou, X. Division of Cow Production Groups Based on SOLOv2 and Improved CNN-LSTM. *Agriculture* 2023, *13*, 1562. https://doi.org/ 10.3390/agriculture13081562

Academic Editor: Michael Schutz

Received: 6 June 2023 Revised: 30 July 2023 Accepted: 1 August 2023 Published: 4 August 2023



**Copyright:** © 2023 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/). udder distance from the floor between the first and third lactation in high-yielding cows. Okkema et al. [4] found that swollen teats in cows with edematous udders reduced milk production. Juozaitiene et al. [5] evaluated morphological indicators of cow udders and measured an increase in milk production of 2.72–3.01 kg in cows with a pelvic shape compared to cows with a round udder under the action of the milking machine, indicating that milk production was associated with cow udder shape. Miseikiene et al. [6] analyzed cows' milk production in different lactation zones. After measuring, cows produced about 4.6 kg (42.2%) of milk in the anterior lactation area and 6.32 kg (57.8%) in the posterior lactation area, indicating a correlation between relative udder capacity and milk production. The research problem is whether the cow production groups could be assigned according to udder characteristics.

Feature extraction from the udder is vital for the analysis of udder traits. Recent domestic and foreign researches have divided udder measurement methods into two main categories. The first category uses manual measurement methods, and the second category uses computer vision techniques to extract cow udder traits. The first category method usually uses tools such as a body ruler [7], aluminum foil [8], and a dynamometer [9,10] for udder traits extraction. However, it is time-consuming. The second is the extraction of cow feature points, which can be realized in several ways. For example, feature point labeling is performed manually [11,12], template matching with images with standard feature points is utilized to obtain feature points [13], and contour maps are obtained from 3D point clouds to compute feature points [14]. Finally, it calculates the cow eigenvalues from the eigen points. Contrasted with manual measurement, it is more automatic and effective [15]. However, the selection and number of feature points greatly impact the calculation of feature values, so there is a certain error between those points and the true feature values. Therefore, the aim of this study is to automate the extraction and analysis of udder features using computer vision, deep learning, and other technologies, and to explore the efficiency of udder features in classifying high- and low-yielding cows.

Nowadays, with the continuous innovation and development of artificial intelligence technology, instance segmentation algorithms can achieve mask segmentation of target objects [16–19]. The neural network can fit nonlinear relationships to analyze and predict unknown data attributes [20] of production groups. Therefore, we discuss the importance of instance segmentation algorithms (SOLOv2, Mask R-CNN (Region-based Convolutional Neural Network) [21,22]), as well as neural network algorithms (CNN-LSTM, BPNN (Back Propagation Neural Network)), for the division of high- and low-yielding cows.

Our objectives were to construct an udder segmentation model to extract targets from the image; realize udder feature extraction, analyze high- and low-yielding udder features, and explore the most suitable classification features; select appropriate classification methods to explore the effectiveness of udder features in cow classification; and apply the constructed scheme to the dairy farm to achieve the division of production groups and provide support for zoning management.

#### 2. Materials and Methods

#### 2.1. Cow Video Acquisition

The data for this study were collected in February 2023 at a 1000-cow farm owned by Jiangsu Yuhang Food Technology Co., Ltd., Yancheng, China, a large modern cow farm, in Bailin Village, a southwest suburb of Dongtai City, Yancheng City, Jiangsu Province. There were several passages inside the experimental site with a width of about 2.5 m and cow living areas on both sides of the passages. The cow farming areas were divided into high- and low-yielding areas based on agricultural experts, considering the factors of milk production, parity, and cow condition. The reference standard for milk production is that a cow producing more than 9000 kg of milk in a lactation (305 d) is a high-yielding cow and the rest are low-yielding cows (excluding unproductive cows). The cows' age ranged from around two to eight years old (excluding unproductive cows), in height from 130 to 145 cm and in weight from 550 to 750 kg. The bedding is sorted daily and changed monthly.

Based on the location of the cow's udder in the body region and the measurement method of udder characteristics, this study used a self-designed dairy farm inspection robot to collect images of different cows in the high- and low-production groups to reduce cow stress and improve image quality.

The cow farm inspection robot comprises a mobile chassis, a lifting bar, an industrial camera, a Jestson Nano, and corresponding control components. The mobile chassis refers to a modern automobile drive and steering structure, with DC (direct current) brush motors providing the driving force and digital servos controlling the steering. The wheels are 180 mm solid rubber wheels, of which the two rear wheels are the driving wheels to drive the chassis movement, and the two front wheels are the driven wheels to control the chassis steering. The mobile chassis adopts the SLAM (Simultaneous Localization and Mapping) algorithm, which can realize laser map building and autonomous navigation. The lift rod is a DC electric actuator with a stroke of 500 mm and a maximum height of 1160 mm. Relays control the direction of lift rod movement, which can meet the demand for udder height shooting. An industrial camera is mounted on top of the lift rod to capture the side udder image of the cow, with an image size of  $640 \times 640$  and a frame rate of 30 fps. Then, the image is transmitted to the cloud platform via Jetson Nano. The robot body structure is based on a modern car body and was produced using 3D printing technology. During image acquisition, the robot inspects the passage, keeping the same distance from the cow and moving in the direction parallel to the cow's side, continuously captures the cow's udder side image, and uploads the video to the AliCloud OSS (Object Storage Service) object storage platform for data cloud transmission and storage. The actual view of the device on the cow farm is shown in Figure 1.



Figure 1. View of inspection robot in operation.

## 2.2. Keyframe Extraction

Since the cow images are acquired by intercepting the video taken by the inspection robot at a specific frame rate, considering the slow movement of the cows and the inspection robot, it may result in a certain amount of duplicate images. Therefore, this study proposes a method to extract keyframes from the video, which segments the video sequence with shots to obtain the distinct features of the images, and then extracts the critical information from the video to increase the amount of information in the dataset and reduce the redundancy.

This study uses the inter-frame difference method based on local maxima to extract keyframes, judging the changing size between adjacent images by differencing two adjacent frames according to the average pixel intensity. Then, the image with a large change compared to the previous image is extracted, which is the keyframe. The extracted before and after keyframes are shown in Figure 2. Based on a reasonable threshold, cows with different features can be obtained.



**Figure 2.** Adjacent keyframe images and origin images: (**a**) the previous keyframe; (**b**) the next keyframe; (**c**) the previous origin frame; (**d**) the next origin keyframe.

# 2.3. Image Augmentation

The randomness of cow movement and the instability of the inspection robot camera tracking led to insufficient initially acquired datasets, category imbalance, and problems in image quality. This study used image augmentation methods such as panning, mirroring, brightness adjustment, and contrast transformation to increase the diversity of the dataset, improve the image quality, meet the higher requirements of the deep learning algorithm model for the dataset, improve the accuracy of mask extraction, and lay the foundation for the subsequent neural network to classify the production groups.

Image augmentation is one of the data augmentation techniques used to address the problem of insufficient data required in deep neural network training in this study. Image augmentation can expand the dataset without collecting new samples [23]. Panning and mirroring are image augmentation methods based on geometric transformations. Panning is achieved by setting a threshold value to move the cow in a specific range along a random distance horizontally or vertically, in which the pixel size of the cow does not change, but only the filling of its background edges. Figure 3a shows where the edges after panning are filled with zero-pixel values. Mirroring refers to flipping the cow image left and right or up and down. This study mainly used left and right flipping to change the object's center position in the image to reduce the influence of the target object's position when taking pictures. Figure 3b shows that the cow image is flipped left and right. Luminance and contrast are image augmentation methods based on image color channel adjustment. The luminance adjustment can reduce the sensitivity of the model to color and reduce the influence of the light intensity of the cow farm on the shooting by setting a reasonable

threshold value. Figure 3c shows that the cow image is darkened after the luminance adjustment. Adjusting the image contrast can make a particular area in the image with a noticeable color difference more prominent. Combined with the cow's physical signs, the udder area will be more protuberant and facilitate feature extraction. Figure 3d shows that the cow udder outline is more transparent. The dataset was increased from 503 images to 1307 images by image augmentation, which enhances the diversity of samples.



**Figure 3.** Image augmentation: (**a**) image after panning; (**b**) image after mirroring; (**c**) image after adjusting brightness; (**d**) image after adjusting contrast.

## 2.4. Udder Segmentation Model

Instance segmentation combines object detection and semantic segmentation to achieve pixel-level individual segmentation and classification. Mask R-CNN and SOLOv2 are typical two-stage and one-stage models in instance segmentation, respectively. Mask R-CNN separates detection from segmentation and uses a top–down idea to predict the bounding box first and then segment individuals from each bounding box. SOLOv2 is an anchor-free instance segmentation model, which defines instance segmentation as a simultaneous detection task and segmentation task [24]. The two-stage detection model detects first and then segments, which has poor real-time performance, and the segmentation results correlate with excellent or low-quality bounding box localization. The one-stage model parallels detection with classification and has the characteristics of fast speed and high accuracy. However, such a model is strongly influenced by the detection accuracy. If the individuals have overlapping phenomena, the segmentation effect will be poor. Therefore, this study compared the effects of two segmentation models applied to cow udders, and selected a more suitable model. The parameter settings of the two segmentation models are shown in Table 1.

Instance Segmentation Algorithms	Parameter				
	Max_iter = 60,000				
	Solver.Gamma = 0.1				
SOL Or 2	Solver.Warmup_Factor = $1.0/100$				
SOLOV2	Solver.Warmup_Iters = 10				
	Base_Lr: 0.0001				
	Batch size = $1$				
	Epoches = 600				
	Steps per epoch = $100$				
Mask R-CNN	First 300 Epoches, Learning rate = 0.001, layers = 'heads'				
	After 300 Epoches, Learning rate = 0.0001, layers = 'all'				
	Batch size = $1$				

Table 1. Segmentation model parameter settings.

# 2.4.1. SOLOv2

The cow images were fed into the backbone network Res-101-FPN (Resnet-101-FeaturePyramidNetwork). Resnet ensures the correlation of gradients in the deep network during learning and avoids network degradation due to the increasing number of layers. FPN uses image pyramids to solve the multi-scale problem, fuses features from different convolutional layers during feature extraction to ensure the efficiency of detection of different-size cow udders, and obtains deeper semantic information, which in turn connects prediction of semantic categories and the instance mask of subsequent dynamic heads.

SOLOv2 continues the design of SOLOv1 but further improves the extraction efficiency and accuracy of the mask. Its network structure is shown in Figure 4. SOLOv2 is based on object detection and semantic segmentation. It transforms the segmentation problem into a location division problem by matching the target object's category to the instance's center. It divides the image into a grid of  $s \times s$ . If the target object falls in the center of the grid, the grid performs semantic category prediction on the one hand and instance mask prediction on the other. When the overlap between the center region of the object and the grid is detected to be greater than a threshold, it is considered a positive sample, i.e., there is a category output. Accordingly, an instance mask corresponding to this output is generated. However, since there are often not many instances in the image so that the objects are sparsely distributed, there will be a channel (classifier) redundancy. SOLOv2 solves the output channel redundancy problem by decoupling the mask branch into the kernel branch and feature branch directly into convolutional kernel learning. For the post-processing step of repeated prediction, a matrix NMS (Non-Maximum Suppression) is proposed to accelerate the processing speed of the mask, and the generation of the target mask is more efficient and flexible compared with SOLOv1.



Figure 4. SOLOv2 network structure.

Mask R-CNN also uses Resnet-FPN as the backbone network for feature extraction. Its network structure is shown in Figure 5. The model retains the RPN (Region Proposal Network) in Faster R-CNN for generating region proposals. The RPN input is the feature map generated in the feature extraction stage. To adapt to different target sizes, it generates nine anchor boxes of three scales and three aspect ratios for each point of the feature map. The obtained anchor boxes are processed in two ways: one is to perform foreground and background classification, i.e., to discriminate whether there is a target object in the anchor box and to score the likelihood; and the other is to perform regression to make the anchor frame closer to the ground truth box. Finally, the inaccurate anchor boxes are filtered to obtain the final RoI (Region of Interest). Then, the RoIAlign (Region of Interest Align) is used to adjust the feature map obtained by RPN to the same size. RoIAlign removes the quantization operation and instead uses bilinear interpolation for feature map reduction to avoid losing the information of the original feature map in the process. The feature map obtained by RoIAlign is input into the three-branch structure of Mask R-CNN to complete classification, bounding box regression, and segmentation mask prediction.



**Region Proposal Network** 

Figure 5. Mask R-CNN network structure.

The loss function equation of Mask R-CNN is shown in Equation (1).

$$L = L_{\rm cls} + L_{\rm box} + L_{\rm mask} \tag{1}$$

where  $L_{cls}$  represents the classification loss,  $L_{box}$  represents the bounding-box loss, and  $L_{mask}$  represents the mask loss.

#### 2.4.3. Comparison of Segmentation Effects

Figure 6 shows the comparison of SOLOv2 and Mask R-CNN segmentation results in the same environment, from which it can be seen that both algorithms segment well and the masks are close to the natural contours of the cow udder.

## 2.5. Udder Feature Extraction, Cleaning and Selection

#### 2.5.1. Udder Feature Extraction

In this study, 10 features were initially selected as neural network inputs: circumscribed regular rectangle width and height (max-width, max-height), minimum circumscribed rectangle width and height, aspect ratio (min-width, min-height, rect rate), circumcircle radius (radius), circumcircle area to contour area ratio (circle/contour), fitted elliptical length of major axis and minor axis, and major and minor axis ratio (elliptical a, elliptical b, elliptical rate). The feature values were extracted from the binary mask map extracted by the segmentation model, and its schematical map is shown in Figure 7.



Figure 6. Comparison of SOLOv2 and Mask R-CNN segmentation effects: (a) SOLOv2; (b) Mask R-CNN.



**Figure 7.** Contour feature extraction: (**a**) circumscribed regular rectangle contour; (**b**) minimum circumscribed rectangle contour; (**c**) circumcircle contour; (**d**) fitted elliptical contour.

# 2.5.2. Data Cleaning of Udder Features

In the acquisition of cow udder mask features, NaN (Not a Number) values and outliers with large deviations occurred due to the error of extracting the mask by instance segmentation and the influence of external environmental factors such as shooting angle and cows walking during the acquisition of cow images. In order to ensure the quality of the data, improve the accuracy of neural network prediction, and retain valuable data, this study conducted the mean replacement of missing values and outliers. It consisted of reading the CSV (Comma Separated Values) file through the Pandas, performing a lookup judgment, and applying the mean value of this feature data to replace the NaN, as shown in Table 2, where label 0 represents the actual low-yielding cows on the dairy farm and label 1 represents the high-yielding cows. In this study, based on the distribution of the data, the probability that the values are distributed in ( $\mu - 2\sigma$ ,  $\mu + 2\sigma$ ) was 95.44% based on the 2 $\sigma$  principle. Considering the probability of falling outside  $\pm 2\sigma$  was 4.56%, due to the influence of environmental errors and the sufficient data samples, the mean value replaces

the data with absolute values of errors  $vi > 2\sigma$ , and those with significant deviations from the mean are excluded.

**Table 2.** Comparison of example data before and after cleaning: (a) original data; (b) data after replacing null and outliers by mean values.

(a)										
Max- Width	Max- Height	Rect Rate	Min- Width	Min- Height	Radius	Circle/ Contour	Elliptical Rate	Elliptical a	Elliptical b	Production group
35 23 21 26 25 26 37 36	33 23 28 19 28 35 38	$\begin{array}{c} 1.2381 \\ 1.1304 \\ 1.2381 \\ 2.3684 \\ 1.3333 \\ 1.4000 \\ 1.0167 \\ 1.0267 \end{array}$	$\begin{array}{c} 0.3473\\ 23.2551\\ 23.2551\\ 31.8198\\ 24.0000\\ 29.6985\\ 33.8367\\ 34.4354 \end{array}$	19.6373 20.5718 18.7830 13.4350 18.0000 21.2132 33.2820 33.5410	$\begin{array}{c} 20.3040\\ 12.4308\\ 11.9509\\ 16.1371\\ 12.5507\\ 15.1163\\ 18.9607\\ 19.5209 \end{array}$	0.6562 0.6417 0.6474 0.6562 0.6830 0.6491 NaN 0.7581	$\begin{array}{c} 0.4521 \\ 0.8726 \\ 0.7938 \\ 0.7349 \\ 0.7253 \\ 0.6565 \\ 0.9455 \\ 0.9622 \end{array}$	$\begin{array}{c} 18.9112 \\ 19.6495 \\ 18.1305 \\ 13.1578 \\ 17.9755 \\ 20.4981 \\ 33.3850 \\ 34.0023 \end{array}$	41.8268 22.5194 22.8401 33.0540 24.7833 31.2251 35.3099 35.3373	$egin{array}{c} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array}$
						(b)				
Max- Width	Max- Height	Rect Rate	Min- Width	Min- Height	Radius	Circle/ Contour	Elliptical Rate	Elliptical a	Elliptical b	Production group
35 23 21 26 25 26 37 36	33 23 23 28 32 28 35 38	2.0546 1.1304 1.2381 1.2381 1.3333 1.4000 1.0167 1.0267	$\begin{array}{c} 0.3473\\ 23.2551\\ 23.2551\\ 31.8198\\ 24.0000\\ 29.6985\\ 33.8367\\ 34.4354\end{array}$	19.6373 20.5718 18.7830 13.4350 18.0000 21.2132 33.2820 33.5410	20.3040 12.4308 11.9509 16.1371 12.5507 15.1163 18.9607 19.5209	$\begin{array}{c} 0.3702 \\ 0.6417 \\ 0.6474 \\ 0.6562 \\ 0.6830 \\ 0.6491 \\ 0.7840 \\ 0.7581 \end{array}$	$\begin{array}{c} 0.4521\\ 0.8726\\ 0.7938\\ 0.7349\\ 0.7253\\ 0.6565\\ 0.9455\\ 0.9622\end{array}$	18.9112 19.6495 18.1305 13.1578 17.9755 20.4981 33.3850 34.0023	41.8268 22.5194 22.8401 33.0540 24.7833 31.2251 35.3099 35.3373	0 0 0 1 1 1 1

#### 2.5.3. Udder Feature Selection

Based on the data-cleaned cow udder trait dataset, correlation analysis was performed on the initially selected 10 traits. In this study, the Pearson correlation coefficient was used to analyze the correlation between the 10 features and the production group. The equation for calculating the Pearson correlation coefficient is shown in Equation (2).

$$\mathbf{r} = \frac{\sum_{i=1}^{n} (x_i - \overline{X}) (y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{X})^2 \sum_{i=1}^{n} (y_i - \overline{Y})^2}}$$
(2)

where r represents the Pearson correlation coefficient,  $x_i$  represents the *i*-th value in the sample of variable X,  $\overline{X}$  represents the mean value of the sample of variable X,  $y_i$  represents the *i*-th value in the sample of variable Y, and  $\overline{Y}$  represents the mean value in the sample of variable Y.

Figure 8a shows the correlation heat map of the data extracted based on the SOLOv2 mask, and Figure 8b shows the correlation heat map of the data extracted based on the Mask R-CNN mask. The color from dark to light indicates the correlation from low to high. The analysis shows that the Pearson correlation coefficients of the circumscribed regular rectangle width and height (max-width, max-height), the minimum circumscribed rectangle width and height (min-width, min-height), the circumcircle radius (radius), the fitted elliptical length of major axis and minor axis (elliptical a, elliptical b), and the production group are 0.49/0.51, 0.52/0.47, 0.52/0.60, 0.49/0.45, 0.57/0.58, 0.47/0.43, and 0.52/0.60, respectively, which are correlated between 0.4 and 0.6. The Pearson correlation coefficients of the minimum circumscribed rectangle aspect ratio (rect rate), the circumscribed circle area to contour area ratio (circle/contour), and the fitted elliptical major-to-minor axis ratio (elliptical rate) with the production group are 0.09/0.21, 0.00/-0.16, and -0.07/-0.27, respectively, with absolute values in the range 0.0–0.4. The absolute values of the correlation coefficients were found to be 0–0.3 (0 is not included) for weak correlation, 0.3–0.5 (0.3 is not included) for low correlation, 0.5–0.8 (0.5 is not included) for moderate correlation, and 0.8–1.0 (0.8 is not included) for high correlation. Therefore, to reduce the complexity of the neural network algorithm and improve the classification accuracy and efficiency, the weakly correlated features were excluded.

														1 00
production group	1.00							0.47		0.00	- 0.07			
radius	0.57	1.00	0.93	0.86	0.93	0.74	0.81	0.72		- 0.18	- 0.15		- (	0.75
elliptical b	0.52	0.93	1.00	0.79	0.97	0.68	0.82	0.65		- 0.20	- 0.20			0.50
max-height	0.52	0.86	0.79	1.00	0.78	0.79	0.64	0.77	0.02	- 0.02	0.03			0.50
min-width	0.52	0.93	0.97	0.78	1.00	0.67	0.81	0.65		- 0.19	- 0.20		- (	0.25
min-height	0.49	0.74	0.68	0.79	0.67	1.00	0.75	0.97	- 0.27	0.30	0.33			
max-width	0.49	0.81	0.82	0.64	0.81	0.75	1.00	0.74	0.04	- 0.02	0.03			0.00
elliptical a		0.72	0.65	0.77	0.65	0.97	0.74	1.00	- 0.25	0.32	0.36			-0.25
rect rate	0.09	0.18	0.20	0.02	0.27	- 0.27	0.04	- 0.25	1.00	- 0.72	- 0.84			0.50
dircle/contour	0.00	- 0.18	- 0.20	- 0.02	- 0.19	0.30	- 0.02	0.32	- 0.72	1.00	0.81			-0.50
elliptical rate	- 0.07	- 0.15	- 0.20	0.03	- 0.20		0.03	0.36	- 0.84	0.81	1.00			-0.75
	oduction group	radius	elliptical b	max-height	min-width	min-height	max-width	elliptical a	rect rate	arcle/contour	elliptical rate	-		
	īd					(a	.)							
production group	1.00	0.60	0.60	0.58	0.51	0.47	0.45	0.43	0.21	- 0.16	- 0.27			- 1.00
elliptical b	0.60	1.00	0.98	0.99	0.88	0.88	0.81	0.80	0.30	- 0.30	- 0.32			- 0.75
min-width	0.60	0.98	1.00	0.99	0.88	0.87	0.79	0.79	0.37	- 0.31	- 0.32			
radius	0.58	0.99	0.99	1.00	0.87	0.91	0.85	0.83	0.26	- 0.28	- 0.26			- 0.50
max-width	0.51	0.88	0.88	0.87	1.00	0.68	0.83	0.84	0.12	- 0.02	- 0.08			- 0.25
max-height	0.47	0.88	0.87	0.91	0.68	1.00	0.84	0.83	0.11	- 0.18	- 0.10			0.00
min-height	0.45	0.81	0.79	0.85	0.83	0.84	1.00	0.99	- 0.26					- 0.00
elliptical a	0.43	0.80	0.79	0.83	0.84	0.83	0.99	1.00	- 0.24					0.25
rect rate	0.21	0.30	0.37	0.26	0.12	0.11	- 0.26	- 0.24	1.00	- 0.79	- 0.87			0 50
dircle/contour	- 0.16	- 0.30	- 0.31	- 0.28	- 0.02	- 0.18			- 0.79	1.00	0.87			0.50
elliptical rate	- 0.27	- 0.32	- 0.32	- 0.26	- 0.08	- 0.10			- 0.87	0.87	1.00			0.75
	roduction group	elliptical b	min-width	radius	max-width	max-height	min-height	elliptical a	rect rate	drcle/contour	elliptical rate			•
• (b)														

Figure 8. Heat map of correlation coefficients between udder features: (a) correlation coefficients of SOLOv2 mask map extracted data; (b) correlation coefficients of Mask R-CNN mask map extracted data.

# 2.5.4. Data Distribution

The data were analyzed to obtain the data distribution of samples with different characteristics of high- and low-yielding cows processed by the two algorithms. The

characteristic kernel density of high- and low-yielding cows was plotted by selecting the characteristic variables for the four different calculation methods, with the horizontal coordinates indicating the range of values taken and the vertical coordinates indicating the probability density of the occurrence of data points. Figure 9 shows that high-yielding cows have greater values than low-yielding cows in the max-width, the min-width, the radius, and the elliptical b, where the distribution is dense. A shaded variogram was used to visualize the relationship between the two characteristic variables, and the shading indicates the density of the data points, which can be used to visualize the distribution between the characteristic variables and the difference in the distribution of the characteristics of high-and low-yielding cows.



**Figure 9.** The high- and low-yielding cows are distributed in different feature variables, with the high yield on the left and low yield on the right: (a) SOLOv2 mask feature density; (b) Mask R-CNN mask feature density.

#### 2.6. Production Groups Classification Model

This study focuses on improving the neural network model for the production groups cows. The neural network is the core of deep learning, which is connected by several neurons. Its elemental composition is the input, hidden, and output layers. The neurons in the hidden layer refine the input features to enhance the model training effect. The

in the hidden layer refine the input features to enhance the model training effect. The neurons in the network adjust the weights and biases corresponding to different features by continuous learning, constantly normalize the input of the lower layers by using the activation function for nonlinear transformation, and connect different layers. The model parameters are updated by backpropagating the loss function to close the predicted value to the actual value and improve the classifier simultaneously. Finally, the neural network classifies the udder dataset based on the weight vector.

(1) CNN-LSTM

Convolutional neural networks have superior performance, and their application areas include image and data classification, object detection, video processing, natural language processing, speech recognition, etc. [25,26]. LSTM is a long short-term memory network, capable of handling sequential and textual problems [27], a variant of RNN (Recurrent Neural Network). It combines short-term memory with long-term memory through exquisite gate control. It solves the problem of gradient disappearance [28]. LSTM can learn long-term dependent information and generally targets back-and-forth logic, sequence problems with temporal concepts, and text problems. This study explored the effect of binary classification of the high- and low-yielding cow dataset with a certain temporal nature by improving the CNN-LSTM deep learning model. The convolution extracts deep features of the cow mask, and then adding LSTM further processes the output features of the convolution layer. The input layer is set as a sequence input layer with size  $7 \times 1 \times 1$  (the dataset has seven features). The folding sequence layer converts the sequence data into the vector, then puts it into the convolutional network with two convolutional layers, which respectively have 16 and 32 convolutional kernels, both with sizes  $2 \times 1$ . Furthermore, a batch normalization layer is added before the activation function to speed up the model convergence and alleviate the gradient dispersion. The max pooling layer is chosen (i.e., downsampling, to compress the multiple features after convolution and filter out the unimportant features), and then the deep features are obtained after convolution, which are sequence unfolded and input into the LSTM layer. Some inconsequential features are discarded using the dropout layer to prevent the occurrence of the overfitting phenomenon. Finally, the output size of the fully connected layer is 2 for two classifications. The softmax activation function is employed to connect the classification layer; the network model is shown in Figure 10.



Figure 10. CNN-LSTM improvement model.

(2) BPNN

A BP neural network is a multilayer feedforward network using a backpropagation algorithm, and its basic idea is gradient descent. The BP neural network includes two processes: forward propagation of signals and backward propagation of errors. The sample data are input into the neural network through the input layer, and the hidden layer calculates the prediction result to complete the forward propagation. Then, according to the error between the prediction result and the actual result, the chain rule is used to calculate the error of each layer and to calculate the gradient according to the error to update the weights and biases of each layer to complete the backward propagation. The neural network has a strong nonlinear mapping ability and can establish relationships between various udder characteristics and the production area. Therefore, based on the idea of the BP neural network, this study improved the primary BP neural network to make it suitable for classifying production groups. The neural network structure diagram is shown in Figure 11. There are seven feature values for the input data, and thus seven nodes were selected for the input layer. In order to ensure the low complexity of the network parameters and better map the relationship between the features and the production area, two hidden layers with six nodes were constructed. Due to the small number of classification samples, Bayesian regularization was selected as the training function to improve the model's generalization ability. The activation function of the hidden layer uses tansig; the equation is as in Equation (3) and the activation function of the output layer uses softmax to achieve classification; the equation is as in Equation (4). The number of nodes in the output layer was two with the same classification category.

$$tansig = \frac{2}{1 + e^{-2x}} - 1 \tag{3}$$

where *x* represents the output value of the node.

$$softmax = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}}$$
(4)

where  $z_i$  represents the *i*-th node's output value, and C represents the number of output nodes.



Input Layer  $\in \mathbb{R}^7$ 

Hidden Layer ∈ ℝ<sup>6</sup>

Hidden Layer ∈ ℝ<sup>6</sup>

Output Layer  $\in \mathbb{R}^2$ 

Figure 11. BP neural network structure.

#### 2.7. Classification Assessment Indicators

In the classification task, classification results were classified into four categories: true positives (*TP*), false positives (*FP*), false negatives (*FN*), and true negatives (*TN*). In this study, accuracy, precision, recall, and F1-score metrics were chosen to assess the model classification performance.

The accuracy indicates the accuracy of the model prediction, i.e., the proportion of correctly predicted samples to the overall samples, and is calculated as in Equation (5).

$$Accuracy = \frac{n_{correct}}{n_{total}} = \frac{TP + TN}{TP + FP + FN + TN}$$
(5)

The precision reflects the ability of the model to discriminate negative samples, which is the proportion of samples predicted to be positives out of samples that are true positives, is calculated as in Equation (6).

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

The recall reflects the ability of the model to identify positive samples, which is the proportion of true positives predicted to be positives and is calculated as in Equation (7).

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

The F1-score is the summed average of precision and recall, calculated as in Equation (8).

$$F1 = \frac{2 \times Presion \times Recall}{Presion + Recall}$$
(8)

## 2.8. Experimental Design and Setup

## 2.8.1. Experimental Environment

The software platforms used in this study are Labelme 5.1.1 (MIT, Cambridge, MA, USA) for image annotation, PyCharm 2022.1 Community Edition (JetBrains, Prague, Czech Republic) and Python 3.7 (Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands) for image augmentation and feature extraction, IBM SPSS Statistics 26 (IBM Corp, Armonk, NY, USA) for correlation analysis, and Matlab 2021b for neural network construction.

The deep learning network was GPU parallel-accelerated by CUDA 11.6, and cuDNN 8.8.1 was used as the acceleration library for deep convolutional neural networks. SOLOv2 was built based on Detectron2 and AdelaiDet, which are deep learning frameworks. Mask R-CNN was built based on the TensorFlow and Keras frameworks.

The hardware platform for this study was 11th Gen Intel<sup>®</sup> Core(TM) i5-11400H @ 2.70 GHz, 16 G RAM, and NVIDIA GeForce RTX 3050 Laptop GPU.

#### 2.8.2. Instance Segmentation Dataset

After crucial frame extraction and image augmentation, 1093 cow udder images were gained, and of these 449 images were of high-yielding cows and 644 images of low-yielding cows. The training set and test set were divided according to the ratio of 7:3 to obtain 766 images in the training set and 327 images in the test set.

## 2.8.3. Classification Dataset

Two datasets, both of size 1307, were constructed by extracting the mask features of SOLOv2 and Mask R-CNN segmentation separately and randomly dividing the training and test sets according to the ratio of 7:3.

# 2.9. Cow Farm Management

# 2.9.1. Animal Welfare

The average weight of cows in this cattle farm was 660 kg, and the average age was 5. Their living conditions were good. In terms of diet, based on the physiological differences between high- and low-yielding cows, high-yielding cows have a high feed intake and high cow metabolism compared to low-yielding cows, and therefore need to be supplied with more feed and drinking water for the maintenance of physiological needs and metabolism, with high-yielding cows having up to  $90 \pm 10$  kg of daily feed intake. According to the weather one must provide a reasonable amount of feeding water, when the weather is hot in summer, the water should be increased by five to six times; in terms of living environment, to ensure the cleanliness and comfort of the cows' living environment, milking aisles are cleaned up two times a day, lying feces are cleaned up three times a day, lying beds are tidied up at least one time a day, and the depth of plowing is more than 15 centimeters. This fully guarantees animal welfare.

### 2.9.2. Practice and Production

Based on the results of our classification result, high- and low-yielding cows can be categorized for zonal management. In actual management, feeding management is mainly focused on high-yielding cows to improve milk production. Compared with low-yielding cows, high-yielding cows have many unique physiological characteristics. Firstly, high-yielding cows have high nutrient requirements and high daily feed intake. Secondly, their basal metabolic rate is higher, and their respiratory and heart rate are higher than those of low-yielding cows. Therefore, when feeding, attention was given to the structure of the diet with a moderate forage to concentrate ratio, adopting a scientific feeding method, and controlling the amount and frequency of feeding. At the same time, cows were provided with a suitable barn environment and were cleaned regularly. Our classification of high-and low-yielding cows provides support for zoning and fine management of dairy farms and provides a boost to improve cow production.

## 3. Results and Discussions

#### 3.1. Segmentation Model Evaluation

3.1.1. Loss Function

The loss functions of optimal models of SOLOv2 and Mask R-CNN are shown below, and both types of algorithms use weights that have been trained on the MSCOCO (Microsoft Common Objects in Context) dataset as pre-training weights. Utilizing the weights attained from training on the large-scale dataset to initialize the network model allows transferring the learned generic features to the new task, thus improving the performance and generalization of the model. As shown in Figure 12, both algorithms converge after a small number of iterations, taking low loss values.



Figure 12. Segmentation model loss function.

#### 3.1.2. Segmentation Accuracy

This study first exploits the idea of segmentation and then classification for dividing production groups. In the instance segmentation stage, the two-stage and one-stage segmentation models Mask R-CNN and SOLOv2, respectively, were compared. The mAP (mean Average Precision), AP50 (Average Precision), and AP75 were used as metrics to measure the performance of the two algorithms. As can be seen from Table 3, SOLOv2 outperformed Mask R-CNN in all three metrics, and mAP was 7.12% higher than Mask R-CNN. Extended analysis of the AP50 index in Table 3 shows that SOLOv2 and Mask R-CNN were 98.87% and 95.03%, respectively, implying that the vast majority of extracted masks of both algorithms was above 50% of the actual cow udder IoU (Intersection over Union) ratio, which can achieve complete cow udder segmentation more accurately. Since SOLOv2 outperformed Mask R-CNN for object edge segmentation, SOLOv2 performed better for targets with distinct edge features such as cow udders.

Table 3. Segmentation model accuracy.

Instance Segmentation Algorithms	mAP	AP50	AP75
SOLOv2	74.09%	98.87%	92.49%
Mask R-CNN	66.97%	95.03%	70.48%

Since the metrics selected cannot fully evaluate the performance of the segmentation model, this study extracted features from the mask maps segmented by both algorithms. The features were input into the classification algorithm to further analyze the performance of the segmentation model through the classification effect.

#### 3.2. Classification Model Evaluation

### 3.2.1. Effect of Neural Network Model on Test Results

Based on the cow udder mask features datasets, two neural network models were improved in this study. The first one was because the udder mask feature dispersion had certain temporal nature characteristics. A variant LSTM of the recurrent neural network was introduced and convolutional layer and max pooling layer were added to optimize the network and boost the model performance. The second model was employed that improves the basic BP neural network, builds two hidden layers, and uses a backpropagation algorithm to reduce the prediction error. As can be seen from Table 4, the accuracy of the testing sets of the two neural network models is relatively ideal, and the accuracy of CNN-LSTM is superior to BPNN no matter whether for the dataset segmented by SOLOv2 or in the dataset segmented by Mask R-CNN. This is because the CNN-LSTM neural network, compared with the BP neural network, has added convolution layers and increased the number of neurons, making the network structure more complex. Additionally, the performance of CNN-LSTM and BPNN on the dataset segmented by SOLOv2 is superior to that of Mask R-CNN, with the highest accuracy of 96.44% (SOLOv2 + CNN - LSTM), which further indicates that the segmentation effect of SOLOv2 is better than that of Mask R-CNN. The loss function curves corresponding to the two segmentation models based on the CNN-LSTM neural network are shown in Figure 13, and the cross-entropy loss functions corresponding to the two segmentation models based on the BPNN neural network are shown in Figure 14.

Table 4. Improvement of neural network evaluation metrics.

Classification Algorithms	Instance Segmentation Algorithms	Accuracy	Precision	Recall	F1 Score
CNN-LSTM	SOLOv2	96.44%	98.00%	96.47%	97.23%
	Mask R-CNN	90.49%	92.40%	91.88%	92.14%
BPNN	SOLOv2	93.13%	88.65%	91.91%	90.25%
	Mask R-CNN	90.19%	87.70%	90.68%	89.17%



Figure 13. Improved neural network loss function: (a) SOLOv2; (b) Mask R-CNN.



Figure 14. Cross-entropy loss functions: (a) SOLOv2; (b) Mask R-CNN.

## 3.2.2. Comparison of Test Results

In this study, four commonly used machine learning algorithms, namely naive Bayes, K-nearest neighbor, support vector machine, and random forest, were used to classify production groups and compare the effect with neural network classification. Performance metrics are shown in Table 5. The confusion matrix is shown in Figure 15. After analysis and comparison, K-nearest neighbor and random forest performed better among the four algorithms, with the accuracy of SOLOv2 reaching 92.62%/92.74% and Mask R-CNN reaching 85.93%/89.77%. However, both are lower than the two types of neural networks, reflecting the unique advantages of neural networks in multi-feature classification problems.

Table 5. Evaluation metrics of the machine learning algorithm.

Classification Algorithms	Instance Segmentation Algorithms	Accuracy	Precision	Recall	F1 Score
Naive Bayes	SOLOv2	75.72%	79.28%	84.66%	81.88%
	Mask R-CNN	76.35%	75.69%	84.62%	79.90%
K-Nearest Neighbor	SOLOv2	<b>92.62%</b>	<b>94.82%</b>	<b>93.70%</b>	<b>94.62%</b>
	Mask R-CNN	85.93%	86.79%	89.61%	88.18%
Support Vector Machines Random Forest	SOLOv2 Mask R-CNN SOLOv2 Mask R-CNN	68.45% 66.16% <b>92.74%</b> 89.77%	67.02% 61.97% <b>91.37%</b> 89.22%	100% 100% <b>97.55%</b> 92.86%	80.25% 76.52% <b>94.36%</b> 91.00%



**Figure 15.** Confusion matrix obtained from 4 classification models: (**a**) input is the feature data extracted by SOLOv2 mask; (**b**) input is the feature data extracted by Mask R-CNN mask.

In this study, we introduced a method to divide high- and low-yielding cows already in their own pens according to their production levels based on the SOLOv2 and CNN-LSTM models. The main objectives were to investigate the potential of instance segmentation to extract the cow udders and establish a classification model for high- and low-production groups based on neural network. The segmentation effect of SOLOv2 and Mask R-CNN was evaluated; features that can well characterize cow udder traits were explored; and the effectiveness of the improved CNN-LSTM classifier for high- and low-yielding area division was verified.

The technology in the study allows for adjustments to be made to cows after they have been grouped. For example, if some cows in the high-yielding group have entered the low-yielding threshold, for large farms with many cows, it is labor-intensive to rely on manual labor to identify which cows need to be adjusted to the low-yielding group on a regular basis, whereas the technology in the study can be used to realize automatic and convenient identification and adjustment. Meanwhile, cows in the low-yielding group whose milk production capacity has been improved through effective management can also be adjusted to the high-yielding group by identification. This will help farmers to make a decision.

The technology used in this study has certain value and significance compared with grouping high- and low-yielding cows directly according to their actual milk production. Firstly, if the cows are divided by 305 d milk production, the statistical time is long, and it cannot divide the cows quickly and conveniently. In this study, the technology can directly realize the grouping of cows by obtaining cow images and recognizing cow udders under the condition of unknown milk production. Secondly, when the actual milk production is recorded manually, the workload is larger. However, the technology in the study does not need a large amount of data when classifying new cows, which reduces the labor cost and can directly obtain the grouping results. Thirdly, for some cows in the high-yielding group that enter the threshold of the low-yielding group, the techniques in this study allow for quick batch screening and then adjusting cows from the high-yielding group to the low-yielding group.

We compared our technique with several similar studies, and found that there were a few limitations of our technique's employment. A previous study [29] used multiple cameras simultaneously to obtain the depth maps of the cow's body in different directions, artificially labeled the different body parts of the cow, and classified body parts by pixels. The method can alleviate cow fences occlusions to a certain extent, which may seriously influence cow udders segmentation and classification results. The problem of cow fence occlusions also appeared in our research and should be well-handled in the follow-up work.

The environment of a cow barn is complex and weather causes vast variations in illumination, which greatly challenged the subsequent image processing procedure. Thus, the results and reliability of image-processing-based methods may decrease significantly when the conditions covered by training samples are insufficient. Bobbo et al. [30] compared multiple machine learning methods to predict udder health status based on somatic cell counts in dairy cows. Another study [31] utilized ultrasound echotexture analysis of the mammary gland and a deep learning algorithm to predict milk yield. Methodology in [32] proposed a Rfine mask two-stage instance segmentation, a combination of the convolutional neural network ConvNeXt and ECA modules. Inspired by these studies, division of high- and lowproduction groups by fusing multimodal data should be considered, such as physical and chemical data, visible light data, and ultrasound images. Moreover, attention modules can be integrated into CNN-LSTM to deal with small-target and multi-scale-target problems.

#### 4. Conclusions

Based on the relationship between udder properties and milk production, this study proposed a method to divide production groups by segmentation first and then classification. In the segmentation stage, a self-designed inspection robot acquired the video of the cow's udder. Then, for the problem of many duplicated images but low diversity, keyframe extraction and image augmentation were used to expand the dataset. After image preprocessing, to compare the performance of one-stage and two-stage segmentation models in this task, SOLOv2 and Mask R-CNN were selected to segment the images and extract the binary mask images. In the classification stage, 10 feature values were extracted from the mask images. Afterward, the data were cleaned, and features were selected to make the classification model training more efficient and accurate. The results show that the segmentation effect of SOLOv2 was better than Mask R-CNN with mAP up to 74.09%, and the classification using CNN-LSTM was better than BPNN. The segmentation using SOLOv2 and classification using that the proposed method based on the segmentation model and the neural network has effective results in cow production groups.

Author Contributions: Conceptualization, Y.L., M.X. and X.Z.; data curation, G.C., L.Q., Z.C., Z.L. and C.X.; formal analysis, G.C., L.Q., Z.C., Z.L. and C.X.; funding acquisition, M.X. and X.Z.; methodology, G.C., L.Q., Y.L. and X.Z.; project administration, Y.L., M.X. and X.Z.; supervision, Y.L., M.X. and X.Z.; visualization, G.C. and L.Q.; writing—original draft, G.C. and L.Q.; writing—review and editing, Y.L., M.X. and X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Program for International S&T Cooperation Projects of Jiangsu, China (BZ2021022), and the Student Innovative Training Program of Nanjing Agricultural University (202219XX476).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

**Acknowledgments:** We are thankful to Yungang Bai, Sunyuan Wang and Hengtai Li, who have contributed to our field data collection and primary data analysis.

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

## References

- Feil, A.A.; Schreiber, D.; Haetinger, C.; Haberkamp, A.M.; Kist, J.I.; Rempel, C.; Maehler, A.E.; Gomes, M.C.; Silva, G.R. Sustainability in the dairy industry: A systematic literature review. *Environ. Sci. Pollut. Res. Int.* 2020, 27, 33527–33542. [CrossRef] [PubMed]
- Teng, G.H. Information sensing and environment control of precision facility livestock and poultry farming. Smart Agric. 2019, 1, 1–12. [CrossRef]
- 3. Pawlina, E.; Wojciech, K.; Marian, K. The changes in udder size of Red & White breed cows in first and third lactation. *Med. Weter.* **2000**, *56*, 672–674.
- 4. Okkema, C.; Grandin, T. Graduate Student Literature Review: Udder edema in dairy cattle—A possible emerging animal welfare issue. *J. Dairy Sci.* 2021, 104, 7334–7341. [CrossRef]
- 5. Juozaitienė, V.; Saulius, T.; Evaldas, S. The correlation between cows udders morphology and milking characteristics. *Vet. Ir Zootech.* (*Vet. Med. Zoot*) **2007**, *38*, 17–21.
- Mišeikienė, R.; Tušas, S.; Matusevičius, P.; Kerzienė, S. Quarter milking parameters by lactation in dairy cows. *Mljekarstvo Časopis* Za Unaprjeđenje Proizv. I Prerade Mlijeka 2019, 69, 108–115. [CrossRef]
- Lin, C.Y.; Lee, A.J.; McAllister, A.J.; Batra, T.R.; Roy, G.L.; Vesely, J.A.; Wauthy, J.M.; Winter, K.A. Intercorrelations Among Milk Production Traits and Body and Udder Measurements in Holstein Heifers. J. Dairy Sci. 1987, 70, 2385–2393. [CrossRef]
- Magaña-Sevilla, H.; Sandoval-Castro, C.A. Technical Note: Calibration of a Simple Udder Volume Measurement Technique. J. Dairy Sci. 2003, 86, 1985–1986. [CrossRef]
- Franchi, G.A.; Jensen, M.B.; Foldager, L.; Larsen, M.; Herskin, M.S. Effects of dietary and milking frequency changes and administration of cabergoline on clinical udder characteristics in dairy cows during dry-off. *Res. Vet. Sci.* 2022, 143, 88–98. [CrossRef]
- 10. Bertulat, S.; Fischer-Tenhagen, C.; Werner, A.; Heuwieser, W. Technical note: Validating a dynamometer for noninvasive measuring of udder firmness in dairy cows. J. Dairy Sci. 2012, 95, 6550–6556. [CrossRef]
- 11. Chen, S.S.; Wang, M.H. Linearized Appraisal of Dairy Cow's Conformation Using Image Measurement Technique. *J. China Agric. Univ.* **1996**, *1*, 93–98.
- 12. Guo, H.; Wang, P.; Ma, Q.; Zhu, D.H.; Zhang, S.L.; Gao, Y.B. Acquisition of Appraisal Traits for Dairy Cow Based on Depth Image. *Trans. Chin. Soc. Agric. Mach.* 2013, 44, 273–276+229. [CrossRef]
- 13. Huang, J.R.; Qian, D.P.; Wang, W.D.; Chen, X.H. Developing Linear Appraisal of Dairy Cow Conformation System with Image Processing Technique. *Trans. Chin. Soc. Agric. Mach.* **2007**, *38*, 111–113+171. [CrossRef]

- 14. Hu, X.T.; Zhang, Y.C. Cow Breast Shape Features Analysis Method Based on Three-Dimensional Point Cloud. J. Tianjin Univ. Sci. Technol. 2012, 27, 61–64. [CrossRef]
- Xie, Q.J.; Zhou, H.; Bao, J.; Li, Q.D. Review on Machine Vision-based Weight Assessment for livestock and Poultry. *Trans. Chin. Soc. Agric. Mach.* 2022, 53, 1–15. [CrossRef]
- 16. Gao, Y.; Guo, J.L.; Li, X.; Lei, M.G.; Lu, J.; Tong, Y. Instance-level Segmentation Method for Group Pig Images Based on Deep Learning. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 179–187. [CrossRef]
- Rossi, L.; Valenti, M.; Legler, S.E.; Prati, A. LDD: A Grape Diseases Dataset Detection and Instance Segmentation. *Image Anal. Process. –ICIAP* 2022, 13232, 383–393. [CrossRef]
- Sun, X.M.; Fang, W.T.; Gao, C.Q.; Fu, L.S.; Majeed, Y.; Liu, X.J.; Gao, F.F.; Yang, R.Z.; Li, R. Remote estimation of grafted apple tree trunk diameter in modern orchard with RGB and point cloud based on SOLOv2. *Comput. Electron. Agric.* 2022, 199, 107209. [CrossRef]
- 19. Qiao, Y.L.; Truman, M.; Sukkarieh, S. Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. *Comput. Electron. Agric.* 2019, 165, 104958. [CrossRef]
- 20. Ma, L.; Xie, F.; Liu, D.; Wang, X.; Zhang, Z. An Application of Artificial Neural Network for Predicting Threshing Performance in a Flexible Threshing Device. *Agriculture* **2023**, *13*, 788. [CrossRef]
- Kumar, G.; Bhatia, P.K. A detailed review of feature extraction in image processing systems. In Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, India, 8–9 February 2014; pp. 5–12. [CrossRef]
- 22. Duan, E.; Hao, H.; Zhao, S.; Wang, H.; Bai, Z. Estimating Body Weight in Captive Rabbits Based on Improved Mask RCNN. *Agriculture* **2023**, *13*, 791. [CrossRef]
- 23. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. J. Big Data 2019, 6, 1-48. [CrossRef]
- Huang, T.; Li, H.; Zhou, G.; Li, S.B.; Wang, Y. Survey of Research on Instance Segmentation Methods. J. Front. Comput. Sci. Technol. 2023, 17, 810–825. [CrossRef]
- 25. Alghamdi, H.; Turki, T. PDD-Net: Plant Disease Diagnoses Using Multilevel and Multiscale Convolutional Neural Network Features. *Agriculture* **2023**, *13*, 1072. [CrossRef]
- Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* 2020, 53, 5455–5516. [CrossRef]
- 27. Kim, J.G.; Lee, S.Y.; Lee, I.B. The Development of an LSTM Model to Predict Time Series Missing Data of Air Temperature inside Fattening Pig Houses. *Agriculture* **2023**, *13*, 795. [CrossRef]
- 28. Shi, X.J.; Chen, Z.R.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 802–810.
- 29. Salau, J.; Haas, J.H.; Junge, W.; Thaller, G. Determination of Body Parts in Holstein Friesian Cows Comparing Neural Networks and k Nearest Neighbour Classification. *Animals* **2021**, *11*, 50. [CrossRef]
- 30. Bobbo, T.; Biffani, S.; Taccioli, C.; Taccioli, M.; Cassandro, M. Comparison of machine learning methods to predict udder health status based on somatic cell counts in dairy cows. *Sci. Rep.* **2021**, *11*, 13642. [CrossRef]
- Themistokleous, K.S.; Sakellariou, N.; Kiossis, E. A deep learning algorithm predicts milk yield and production stage of dairy cows utilizing ultrasound echotexture analysis of the mammary gland. *Comput. Electron. Agric.* 2022, 198, 106992. [CrossRef]
- Zhao, H.; Mao, R.; Li, M.; Li, B.; Wang, M. SheepInst: A High-Performance Instance Segmentation of Sheep Images Based on Deep Learning. *Animals* 2023, 13, 1338. [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.