Article

Goat-Face Recognition in Natural Environments Using the Improved YOLOv4 Algorithm

Fu Zhang [1,2], Shunqing Wang [1], Xiahua Cui [1], Xinyue Wang [1], Weihua Cao [1], Huang Yu [1], Sanling Fu [3,*] and Xiaoqing Pan [4,*]

1   College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China
2   Collaborative Innovation Center of Machinery Equipment Advanced Manufacturing of Henan Province, Luoyang 471003, China
3   College of Physical Engineering, Henan University of Science and Technology, Luoyang 471023, China
4   Institute of Animal Science, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China
*   Correspondence: 9903401@haust.edu.cn (S.F.); pxq1611@jaas.ac.cn (X.P.)

Abstract: In view of the low accuracy and slow speed of goat-face recognition in real breeding environments, dairy goats were taken as the research objects, and video frames were used as the data sources. An improved YOLOv4 goat-face-recognition model was proposed to improve the detection accuracy; the original backbone network was replaced by a lightweight GhostNet feature extraction network. The pyramid network of the model was improved to a channel management mechanism with a spatial pyramid structure. The path aggregation network of the model was improved into a fusion network with residual structure in the form of double parameters, in order to improve the model's ability to detect fine-grained features and distinguish differences between similar faces. The transfer learning pre-training weight loading method was adopted, and the detection speed, the model weight, and the mean average precision (mAP) were used as the main evaluation indicators of the network model. A total of 2522 images from 30 dairy goats were augmented, and the training set, validation set, and test set were divided according to 7:1:2. The test results of the improved YOLOv4 model showed that the mAP reached 96.7%, and the average frame rate reached 28 frames/s in the frontal face detection. Compared with the traditional YOLOv4, the mAP improved by 2.1%, and the average frame rate improved by 2 frames/s. The new model can effectively extract the facial features of dairy goats, which improves the detection accuracy and speed. In terms of profile face detection, the average detection accuracy of the improved YOLOv4 goat-face-recognition network can reach 78%. Compared with the traditional YOLOv4 model, the mAP increased by 7%, which effectively demonstrated the improved profile recognition accuracy of the model. In addition, the improved model is conducive to improving the recognition accuracy of the facial poses of goats from different angles, and provides a technical basis and reference for establishing a goat-face-recognition model in complex situations.

Keywords: channel management mechanism; YOLOv4; fused residuals; transfer learning

1. Introduction

In recent years, machine-vision technology has developed rapidly in the field of target-individual identification [1–4]. With the further development of precision agriculture, deep-learning methods have been widely used in the fields of the agricultural pest identification [5–7] and biometric identification [8–10], and corresponding progress has been made. The precise identification of individual livestock has become a pressing problem to be solved [11–14]. The common livestock-identification methods are mainly divided

into contact and non-contact kinds [15,16]. At present, non-contact identification technology mainly includes individual identification methods for livestock breeding based on physiological characteristics, such as irises and retinal vessels [17,18]. In this non-contact identification method, data collection is complicated, since herd animals do not readily cooperate, resulting in poor practical applicability.

Face-recognition technology offers advantages in that it is natural and intuitive, does not involve contact, and does not require the cooperation of livestock in a fixed posture; in addition, face-recognition technology has strong anti-interference ability, and has wide practical-application prospects. Therefore, contactless identification based on visual biometrics has become a new trend in individual-livestock identification [19–21]. Chen et al. [22] proposed a lightweight convolutional-neural-network cow-facial-recognition algorithm suitable for edge-computing applications, with an average detection accuracy of 90%. A neural-network model based on a single bovine nose-tip-texture feature by Kumar et al. [23] achieved discrimination between different individuals with 98.99% accuracy. Huang et al. [24] used a multiscale local-differential-direction-number (MLDDN) model for the facial recognition of pigs. Weng et al. [25] proposed a cow-face-recognition model based on a double-branch convolutional neural network (TB-CNN), which has a good detection accuracy. Yang et al. [26] proposed a YOLOv4 detection network incorporating coordinate information to achieve the accurate identification of individual cows, with an average recognition accuracy of 93.4%. Yan et al. [27] proposed an FPA-Tiny-YOLO model combining pyramid attention and Tiny-YOLO to enhance feature-extraction ability and target detection accuracy to solve the problem of individual pig adhesion and obscuration. Hu et al. [28] introduced the dual-attentional-feature mechanism into the Mask-RCNN network structure, which can achieve individual pig segmentation in complex environments. He et al. [29] improved detection accuracy by introducing a dense connection-block structure in the YOLOV3 backbone network to achieve the detection of small targets with occlusion at long distances. Wang et al. [30] proposed a multi-scale convolutional-neural-network-based individual-pig-identification model with 92% accuracy to perform contactless individual-pig identification in complex and variable environments. Yang et al. [31] used a full convolutional-network structure for image segmentation to perform the fast and accurate identification of lactating sows in a piggery environment, with better detection results.

Currently, scholars at home and abroad rarely pay attention to the identification of individual herding goats. Han [32] proposed an improved VGGNet pain-expression recognition algorithm with a recognition accuracy of up to 96.06%, which solved the problems of high experience requirements and low recognition accuracy in current manual pain-recognition processes for individual goats. Zhang et al. [33] proposed an improved MobileFaceNet goat-face-recognition network with an accuracy of 97.91% The above study improves the model's feature-extraction effect by introducing a spatial-attention mechanism and a spatial-transformation module, but the recognition accuracy decreases when encountering situations in which the difference between goat facial-texture features becomes smaller and the similarity increases. In addition, the two-stage target algorithm is limited by the large amounts of computational resources, hardware, and software, which are difficult to apply in practical conditions.

To achieve high accuracy, low cost, and high efficiency in non-contact goat recognition, the main contributions of this paper are as follows: (1) A YOLOv4 goat-face-recognition network based on GhostNet is proposed to reduce the number of model parameters and computational effort. (2) Combined with the small differences and high similarity of goat facial-texture features, a channel-management mechanism with a pyramid structure is introduced to improve the detection capability and accuracy of the model for fine-grained features. (3) The original path-aggregation network (PANet) is changed to a two-parameter PANet structure to improve the generalization performance of the model. (4) In order to comprehensively evaluate the improved goat-face-recognition model, in this experiment, a goat-face training set, validation-verification set, and test set were produced and compared with the traditional YOLOv4. The results show that the improved model

helps to improve the recognition accuracy of different facial-angle postures of dairy goats, which provides a technical basis and reference for establishing a goat-face-recognition model in complex situations.

2. Experimental Data

2.1. Experimental Data Sources

The test goat-face video was taken in a standardized indoor goat factory in Li Zhuang Village, Yichuan County, Luoyang City, Henan Province, China. Thirty adult (35–45 kg) Saanen-breed dairy goats were selected as the test subjects and marked in advance, as shown in Figure 1. In the experiment, a Canon camera was used to track a single dairy goat at a frame rate of 30 fps, and the length of each video recording was between 15 and 30 min to ensure the effectiveness of the recorded video.
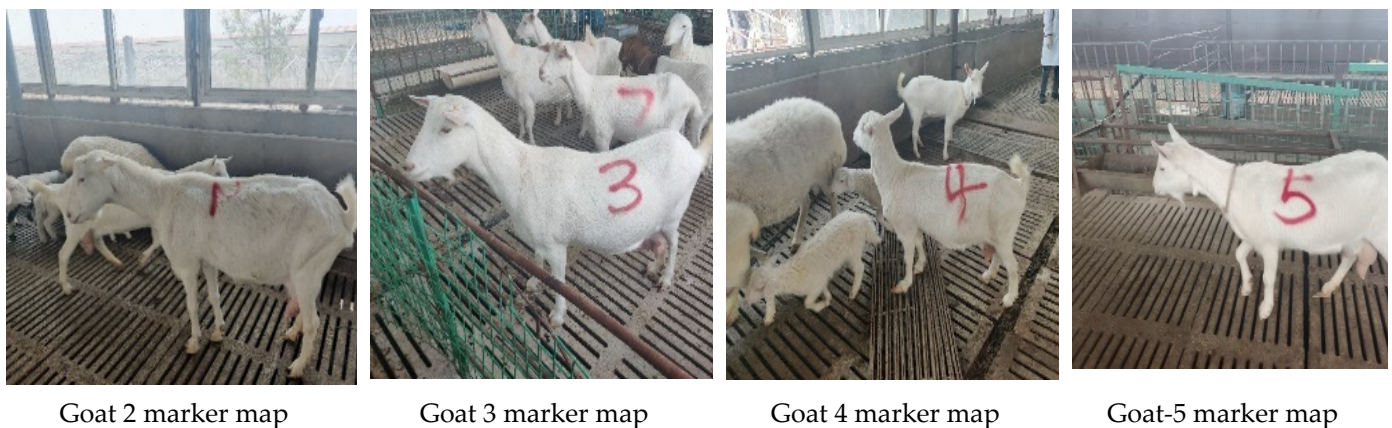

| Goat 2 marker map | Goat 3 marker map | Goat 4 marker map | Goat-5 marker map |

Figure 1. Schematic diagram of milk-goat marking.

2.2. Data Pre-Processing and Labeling

In this study, images were intercepted by 25 frames of the collected video, and the effective images with large similarity differences were selected from the retained images as the sample data for the experiment. Eventually, a total of 3428 valid images were screened. The labeling was used to annotate the images according to annotation format of Pascal VOC dataset, and to generate an annotation file of xml type. The whole dataset was divided according to the ratio of 7:1:2. In order to improve the generalization performance of the model, different-scale images were used to enhance the data in four ways: the random rotation (−15°~15°), the mirror flip, the horizontal flip, and the brightness change. At the same time, the corresponding annotation files of t each image were transformed simultaneously to generate training set (9598 images), validation set (1371 images), and test set (2742 images). The overall process was divided into four parts: video image processing, image augmentation, data-set division and model training, and model validation, as shown Figure 2.
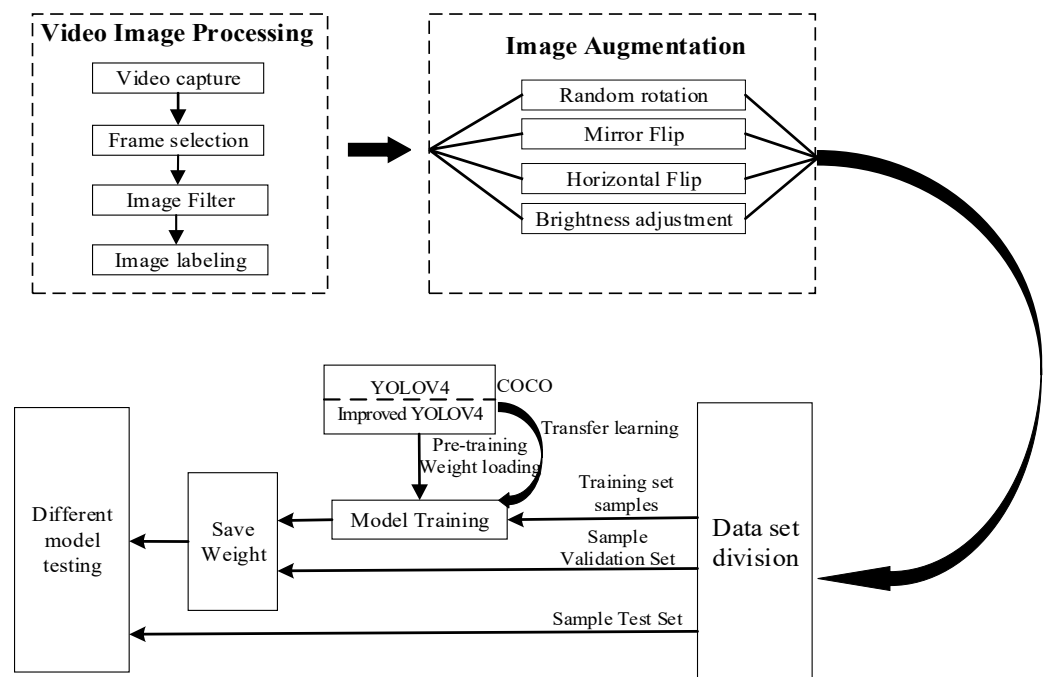
Figure 2. Flow chart of goat-face recognition.

3. YOLOv4 Recognition Algorithm

3.1. YOLOv4 Algorithm

YOLOv4 target-recognition network makes a series of improvements to YOLOv3. It has a backbone-feature-extraction network (CSPDarkNet53), a spatial pyramid (SPP), a path-aggregation network (PANet), a head network (YOLOhead), and four other components. The specific structure is shown in Figure 3.
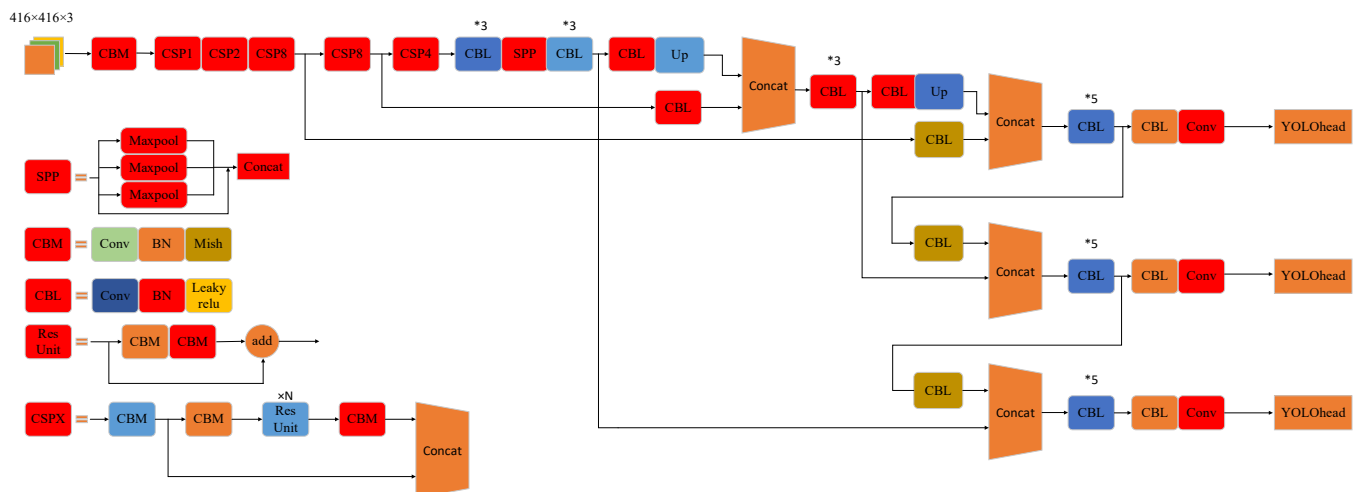

Figure 3. Structure diagram of YOLOv4 model. Note: CSPX stands for cross-stage partial structure, Conv stands for convolutional, BN stands for batch norm, CBL stands for Conv+Batch BN+Leaky-relu-activation-function-synthesis module, CBM stands for Conv+BN+Mish-activation-function-synthesis module, ResUnit stands for the residual connection module, Concat stands for the feature-concatenation operation, Up stands for upsampling operation, Maxpool stands for the pooling operation, *3 and *5 stand for the number of repetitions of the CBL module.

YOLOv4 combines the advantages of the CSPNet and DarkNet53 feature-extraction networks, replacing the DarkNet53 backbone network in the original YOLOV3 with the CSPDarkNet53 backbone network. CSPDarkNet53 feature-extraction network consists of the five residual modules from CSP1 to CSP5; each residual module consists of small residual structures (ResUnit) and CBM modules stacked together, as shown in Figure 3. The SPP structure is located between the backbone network and the neck network. It uses three sizes of pooling kernel, 13×13, 9×9, and 5×5, and then splices the feature maps of different scales with the original feature maps for output, which can improve the receptive field of the network and facilitate subsequent path aggregation of network-feature-information fusion. The YOLOv4 elicits three different-sized feature maps from CSP3~CSP5, 52×52, 26×26, and 13×13, with the aim of detecting objects of different sizes in the image more comprehensively. The three different-sized feature maps are fused with bottom-up and top-down features using PANet, which enhances utilization of effective features and prevents loss of low-order features in the feature-extraction process.

3.2. Improved Goat-Face-Recognition-Algorithm Construction

The original CSPDarkNet53 backbone-feature-extraction network was replaced by a lightweight GhostNet feature-extraction network to reduce the number of parameters and amount of computation of model to solve the problems of huge parameters of the YOLOv4 algorithm-backbone network, increased computation, and poor goat-face recognition in complex environments. GhostNet is a more efficient generation method proposed for the phenomenon of feature redundancy in feature-extraction networks, as shown in Figure 4. GhostNet generates a large number of redundant feature maps through linear operations in the Ghost module to reduce the network computation and increase the speed of the model. The GhostNet feature-extraction network consists of a series of Ghost Bottlenecks modules stacked on top of each other. Ghost Bottlenecks are divided into two types, as shown in Figure 4a,b. Figure 4b contains the separable convolution (DWConv) structure used to reduce the number of parameters of the model and improve the efficiency of the model's operation. Figure 4c,d shows this separable convolutional structure and Ghost module.
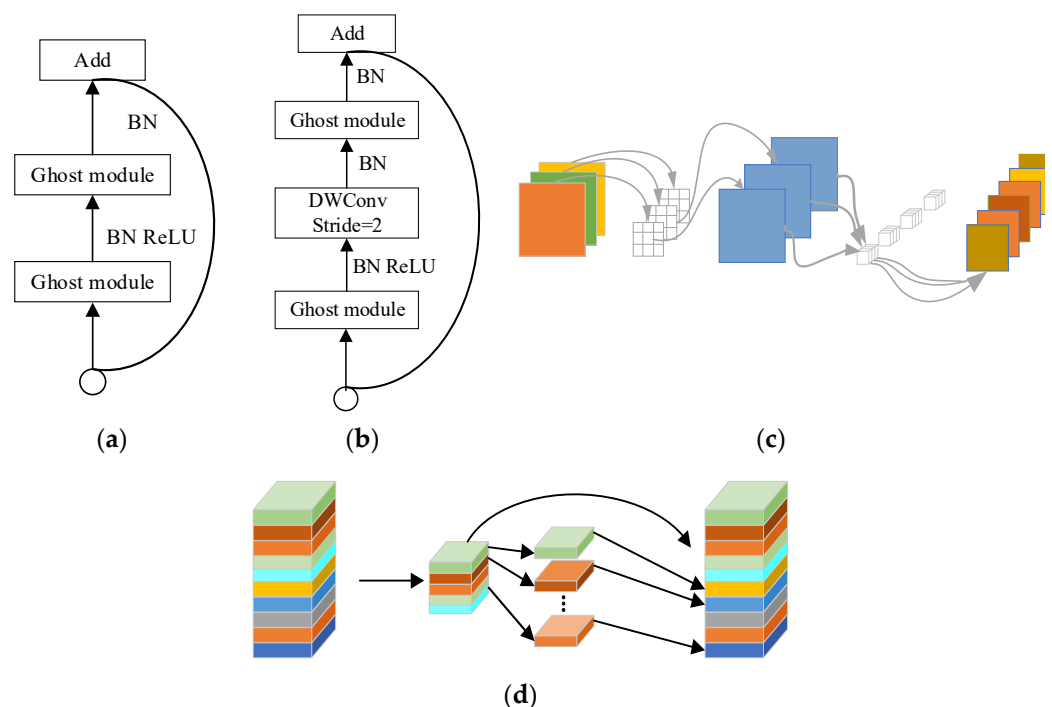

Figure 4. Structure diagram of Ghost Bottlenecks. (**a**) stride = 1 Bottleneck, (**b**) stride = 2 Bottlenecks, (**c**) DWConv, (**d**) Ghost module.

YOLOv4 feature-fusion phase consists of two components: SPP and PANet. Spatialpyramid structure can extract different-scale features from the pixel level and consider multiple receptive-field data in parallel, which has a strong recognition effect on targets of large and small size. However, the fusion of feature information between feature maps of different scales in the traditional pyramid structure is completed only by linear superposition, which tends to ignore detailed features and lacks further extraction of the important features. Therefore, fusion of features directly through path-aggregation networks may lose important location information. In this study, an improved pyramid structure was used, as shown in Figure 5. A channel-management mechanism was added to the SPP to achieve the effect of improving the screening of the important feature layers and increasing the utilization of effective feature layers by introducing the SE channel-management mechanism between different-sized pooling kernels. It can achieve effective features that facilitate target recognition by introducing channel-management mechanism to achieve weight distribution among feature maps. This operation can effectively focus on the features that contain objects and suppress secondary information to improve the effect of model detection.
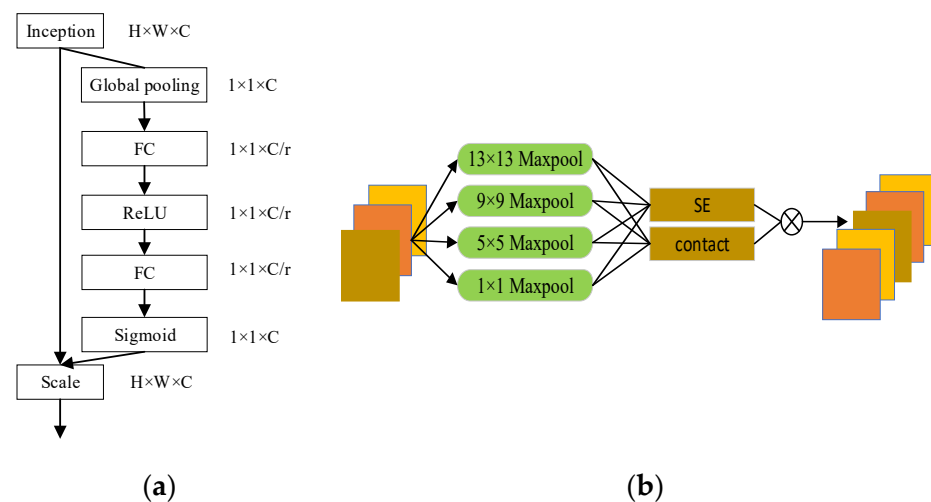

Figure 5. Improved pyramid structure. (**a**) SE spatial-attention-mechanism structure diagram, (**b**) improved SPPnet space diagram.

The PANet structure in the YOLOv4 has multi-port feature-fusion effect, which can perform bottom-up and top-down feature fusion from shallow features to deep features and improve the detection capability of large, medium, and small objects. However, the transfer path from shallow features to deep features is long, and its important feature and localization information are easily lost, which causes problems such as low data utilization and unsatisfactory detection accuracy. To address these problems, the PANet is replaced by a PANet structure with a double-parameter residual structure. It reduces the network-model size, number of algorithmic parameters, and amount of computation by introducing the trainable parameter W_i for focusing on important features and using deep separable convolution instead of part of the normal convolution in the PANet. At the same time, it improves the feature-fusion capability of the network by increasing the output ports of backbone network ($104 \times 104 \times 64$). This operation preserves the location information of the lower-order feature maps and adds the higher-level abstract semantic information to improve the recognition accuracy and feature-extraction capability of the network in complex situations. The improved YOLOv4 goat-face-recognition algorithm is shown in Figure 6. As can be seen from the figure, the network as a whole is divided into four parts: ① represents the GhostNet backbone network structure; ② represents the improved spatial pyramid structure; ③ represents the improved PANet structure; and ④ represents the head network (YOLO Head). The combined convolution block in Figure 6 contains a DWConv and a Ghost module. The DWConv reduces the number of parameters in the goat-face

model and improves the network's recognition speed, while the Ghost module reduces the redundancy of the features in the feature-fusion process and improves the utilization of effective features. W1~W4 denote the trainable parameters added in this experiment, which were used to achieve the focus of the residual network structure on the effective features and enhance the recognition effect of the network model in complex environments.
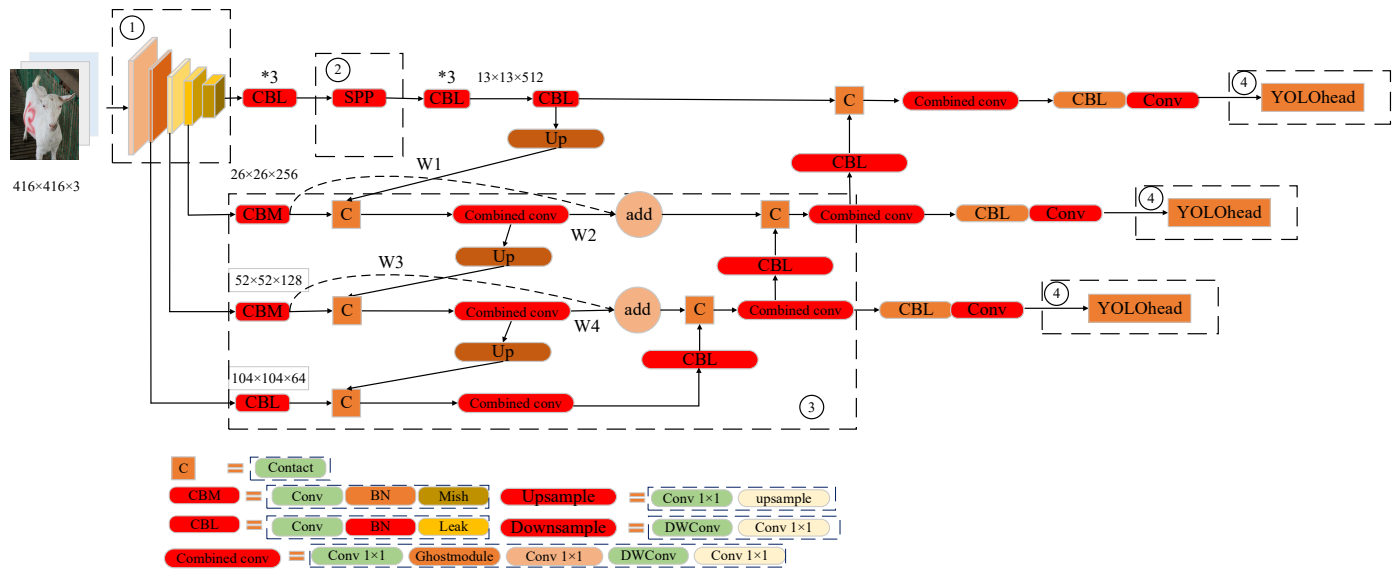

Figure 6. Goat-face-recognition network. Note: 416 × 416 × 3 stands for the input size of image, and 104 × 104 × 64, 52 × 52 × 128, 26 × 26 × 256, and 13 × 13 × 512 stand for four different sizes of feature maps induced from the GhostNet backbone-feature-extraction network, and *3 stands for the number of repetitions of the CBL module.

3.3. YOLOv4 Objective Loss Function

The YOLOv4 objective loss function consists of four parts, namely, positive sample-coordinate loss, positive sample-confidence loss, negative sample-confidence loss, and positive sample-classification loss. The loss function is calculated as shown in Equation (1).

$$
\begin{aligned}
Loss = \lambda_{coord} \sum_{i=0}^{K \times K} \sum_{j=0}^{M} I_{ij}^{obj} (2 - w_i \times h_i)(1 - CIOU) &- \sum_{i=0}^{K \times K} \sum_{j=0}^{M} I_{ij}^{obj} \left[\hat{C}_l \log(C_i) + (1 - \hat{C}_l) \log(1 - C_i) \right] \\
&- \lambda_{noobj} \sum_{i=0}^{K \times K} \sum_{j=0}^{M} I_{ij}^{noobj} \left[\hat{C}_l \log(C_i) + (1 - \hat{C}_l) \log(1 - C_i) \right] \\
&- \sum_{i=0}^{K \times K} \sum_{j=0}^{M} I_{ij}^{obj} \sum_{c \in classes} [\hat{p}_l \log(p_i(c)) + (1 - p_i(c))] p
\end{aligned}
\tag{1}
$$

where λ_{coord} and λ_{noobj} represent positive sample-weight coefficients and negative sample coefficients, respectively; $\sum_{i=0}^{K \times K} \sum_{j=0}^{M}$ represents traversal all prediction boxes; I_{ij}^{obj} and I_{ij}^{noobj} represent the presence or absence of an object, i.e., 1 for presence of an object and 0 for absence of an object, respectively; $\hat{C}_l$, C_i represent the predicted and true values of the sample, respectively; and p_i represents the predicted probability for a category. Complete intersection of union loss (CIOU) represents the loss function used between the prediction frame and the true frame in this experiment.

CIOU loss function is calculated as follows.

$$
CIOU = 1 - IOU + \frac{\rho^2 \left(b - b^{gt} \right)}{c^2} + \alpha v
\tag{2}
$$

$$
v = \frac{4}{\pi^2} \left(arctan \frac{w^{gt}}{h^{gt}} - arctan \frac{w}{h} \right)^2
\tag{3}
$$

where $\rho^2\left(b - b^{gt}\right)$ represents the diagonal distance of the minimum closure region between the prediction frame and the real frame; α is used to measure the consistency parameter between the prediction frame and the true frame and represents a trade-off parameter.

3.4. Training of Models

3.4.1. Model Training and Parameters

The hardware platform was Intel(R) Xeon(R) Silver4210R with 3.5 GHz, 32 GB memory, and NIVIDIA GeForce RTX 2080Ti GPU with 16 GB video memory. The software platform used was Pycharm2020.2+ CUDNN7.4.1.5+ Python3.8+ pytorch1.2. In this experiment, the transfer-learning training method was used to train improved goat-face-recognition network on COCO dataset. Next, the pre-COCO trained network weights were used as initialization, which can accelerate the model convergence and improve the generalization performance of goat-face-recognition network. In terms of network-parameter settings, this experiment uniformly set the training-image size to 416 × 416 size, the training batch size (Batchsize) to 16, and the network-training-period size (epoch) to 100. It automatically saved the weights once for each epoch training completed by the model. The backbone layer in the first 50 epochs of goat-face-recognition network was trained by freezing, and the learning rate (lr) was initially set to 0.001. The backbone network was trained by thawing for the last 50 epochs. To enhance the extraction of the goat-face-recognition network features by the network, the lr was set to 0.0001. In order to enhance the generalization and recognition accuracy while the model was training, employing training techniques were used to make YOLOv4 more versatile and robust in terms of detection, such as Mosic data-enhancement method, label-smoothing algorithm, and cosine-annealing algorithm.

3.4.2. Model-Evaluation Indicators

The average precision (AP), average precision mean (mAP), Accuracy, Precision, Recall, Frames Per Second (FPS), model weight size, model parametric number (Params), model computation (FLOPS), and memory required for model-network node inference (Memory) were used as network-model-evaluation metrics in this study. AP ($AP = \int_0^1 Precision \cdot Recall$) is obtained by plotting the P-R curve with Recall (R) as the horizontal axis and Precision (P) as the vertical axis and integrating it to find the area under the curve. The mAP ($mAP = \frac{\sum_{i=1}^N AP_i}{N}$) is obtained by summing the AP values for the each category and then averaging them, as follows.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$AP = \int_0^1 Precision \cdot Recall \tag{7}$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \tag{8}$$

where TP represents the number of positive samples that the model predicts to be consistent with the true label; FP represents the number of samples in which the model prediction does not match actual positive sample; FN represents the number of samples in which the model prediction does not match the actual negative sample; TN represents the number of samples in which the model prediction is consistent with the actual negative sample.

4. Results and Discussion

4.1. Comparison of Frontal Face Results of Different Models

This experiment used a series of improved YOLOv4 goat face recognition model and the YOLOv4 model to detect the positive faces of the goats, respectively, and the results are shown in Table 1. The mAPs in the table were all obtained at IOU = 0.5. In Table 1, ① represents the replacement of the original YOLOv4 backbone network with a lightweight GhostNet structure, ② represents the replacement of the pyramidal network with a network structure that adds an attention mechanism, and ③ represents the replacement of the original path-aggregation-network structure with a path-aggregation network in the form of residual structured double parameters.

Table 1. Face-detection results of goats with different models under IOU = 0.5.

Model	FPS	mAP	Weight/M	Params	Flops/G	Memory/M
YOLOv4	26	94.6	244.0	64,093,851	29.98	606.95
YOLOv4+①	35	85.8	152.0	39,982,331	13.00	266.69
YOLOv4+①+②	31	89.9	153.0	40,015,643	13.00	266.70
YOLOv4+①+③	30	93.4	57.0	11,440,293	4.62	428.61
YOLOv4+①+②+③	28	96.7	57.6	11,473,605	4.62	428.62

As shown in Table 1, the mAP detection accuracy decreased by 8.8%, after replacing the YOLOv4 backbone network with GhostNet. However, the frame rate reached 35/s. To verify the effectiveness of the network structure, the combination of the replaced backbone target-recognition network with ② and ③ approaches, respectively, yielded significant improvements compared to replacing only the YOLOv4 backbone structure. The mAPs were 89.9%, 93.4%, and 96.7%, respectively; the goat-face-recognition network was improved by 2.1% compared to the YOLOv4 recognition network after introducing the operations ①②③. In terms of the detection speed and model size, the improved YOLOv4 goat-face-recognition network recognized the animals faster than the YOLOv4, with a frame rate of up to 28/s and a model weight reduced to one-fourth of the YOLOv4 weight. This study also shows the results of the goat-face-recognition model in terms of the model parameters, the memory required for the model node inference, and the model computation. As shown in Table 1, this improved goat-face-recognition model reduced the model parameters and model computation significantly. However, there was less change in the memory required for the model-node inference, and the inference speed was delayed compared to the modification of the backbone. Nevertheless, is the model demonstrated an improvement to the original YOLOv4, which affects the time required for the model frame rate to some extent. Figure 7 shows the positive face-recognition results of each model for goats 9, 11, 13, and 22, from which it can be seen that each model can accurately recognize the corresponding goats without omission or misrecognition However, in this study, the improved YOLOv4+①+②+③ goat-face-recognition model demonstrated the best results and had a higher detection accuracy.

The validation-set-loss (val_loss) variation curves of each model for the 100-epoch training cycles were plotted, as shown in Figure 8. From the figure, it can be seen that the val_loss variation curves for the different models all tended to converge steadily with the training period. However, after the introduction of the ①②③ structure in this study, the optimal smoothness of the goat-face-recognition network further demonstrated the network's effectiveness and stability.
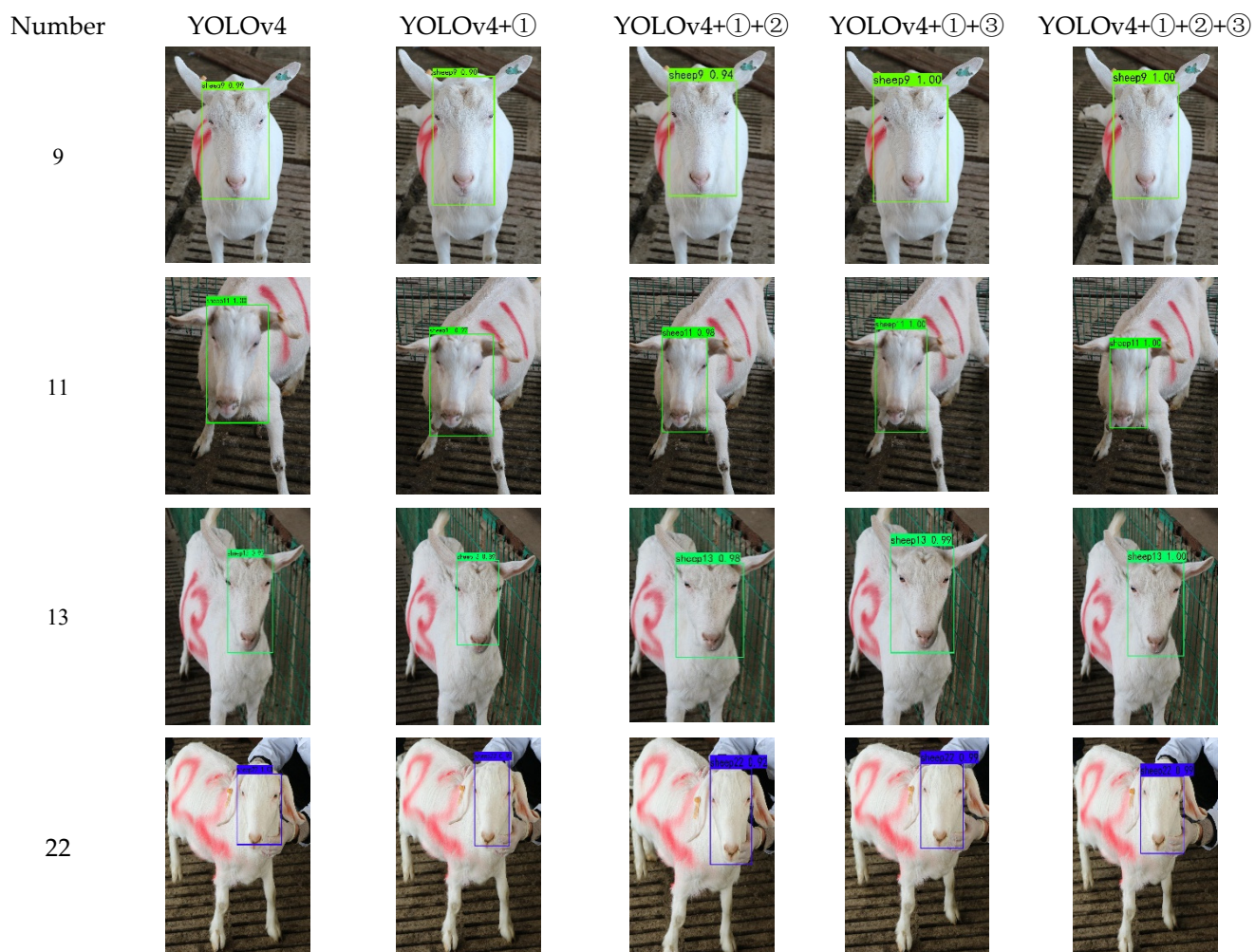
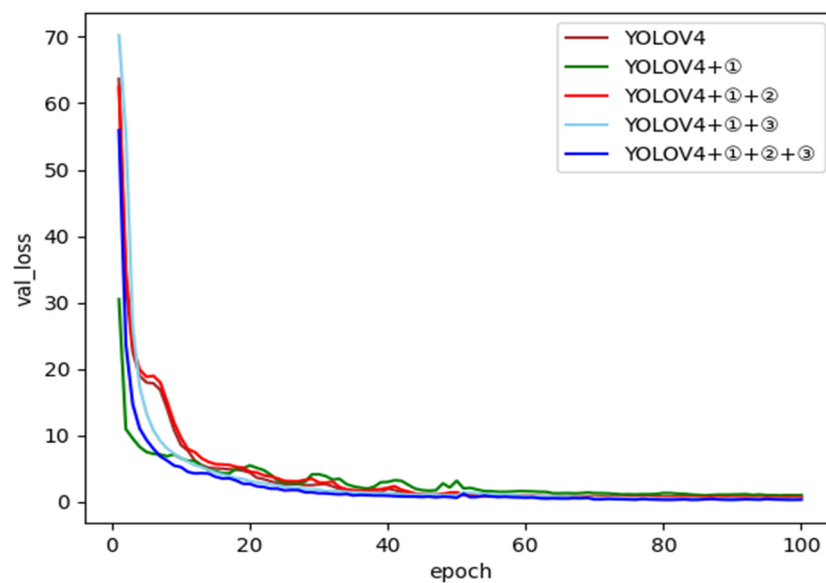**Figure 7.** Different models of face-recognition results (IOU = 0.5).

Figure 8. Loss curve of each model test set.

4.2. Recognition Results of Different Models for Side-Facing Dairy Goats

Dairy goat side face recognition is unavoidable in the face recognition process, with physical occlusions such as fences and dairy goat behavior where they are. Therefore, the side-face recognition of goats is of great importance for their identity verification. To some extent, it represents the quality of goat-face-recognition networks and their resistance to external influences. In this experiment, 225 side-face photographs from outside the dataset of five dairy goats were selected to test the built goat-face-recognition model, and the results are shown in Table 2.

Table 2. Side-face-recognition results of goats with different side faces.

Model	Goat6	Goat9	Goat13	Goat17	Goat21	mAP
YOLOv4	38	42	24	32	24	71
YOLOv4+①	38	42	24	32	24	58
YOLOv4+①+②	38	42	24	32	24	69
YOLOv4+①+③	38	42	24	32	24	72
YOLOv4+①+②+③	38	42	24	32	24	78

The distribution of the side-face photographs of the five goats is shown in Table 2. For the side-face photographs, the amount of data was relatively small. Since the tests were completed on the same side-face photographs between different models, the test data of each goat in the different models in Table 2 were consistent. From Table 2, it can be observed that GhostNet has a smaller network structure compared to CSPDarknt53, which is a lightweight network structure and is less effective at the side-face recognition of goats. Therefore, the mAP of the goat-face-recognition network after replacing the backbone was significantly decreased, by 13%, compared to the YOLOv4. From Table 2, it can be observed that the mAP of the test was significantly improved by combining the recognition network with the improved ② and ③ structures, respectively, after replacing the backbone. The map in the goat face recognition network increases to 69% in side-face recognition after introducing the ② structure, 72% after the introducing the ③ structure, and 78% after adding both ② and ③ improved structures. The goat-face-recognition network in this study improved the side-face recognition by 7% compared to the YOLOv4, indicating the effectiveness of the goat-face-recognition network built for side-face recognition.

As the color of goat faces is mainly pure white, some goat faces have high similarity, which increases the difficulty of identifying the side faces of goats, leading to misidentification and omission. This experiment demonstrates the occurrence of misidentified and omitted goat-face measurements in the five categories of images containing side-faced goats mentioned above, as shown in Figure 9. Since the side-face images contain limited features of goat faces, it is difficult for the recognition network to capture important goat-face features in terms of feature extraction. As can be observed in Figure 9, YOLOv4 was weak at side-face recognition, misidentifying goat13 as goat17. YOLOv4 missed the recognition of goat 21. The AP was only 67% for both goat 13 and goat 17, respectively. The YOLOv4+① network structure misidentified goat 21 as goat 20, and the mAP was only 43%. Since the goat's side-face recognition contained fewer important features, it can be observed in Figure 9 that the improved goat-face-recognition structure of this experiment did not show this misrecognition, but the omitted recognition was not resolved. The improved network structure still failed to detect goat21, although the AP improved to 75%.

This experiment further demonstrates the effectiveness of the present network at identifying individual goats and the robustness effect of the model by detecting photo graphs of five side facing dairy goats. In this study, the introduction of the attention mechanism in the pyramid structure can enhance the fine-grained feature extraction of goat-face-recognition networks and the detection of differences between similar faces. By introducing the residual path structure of trainable parameters, the screening of effective features can be enhanced and the recognition accuracy of the model can be further improved. In this study, the path network structure in the original YOLOv4 was improved to a residual path structure with trainable parameters. The trainable parameters can further enhance the

extraction of important features of goat faces and improve the detection accuracy of the model. Although the goat-face-recognition network based on the YOLOv4+①+②+③ has high accuracy in frontal face recognition, it still needs further improvements in its side-face recognition to improve the accuracy with which it identifies individual goats.
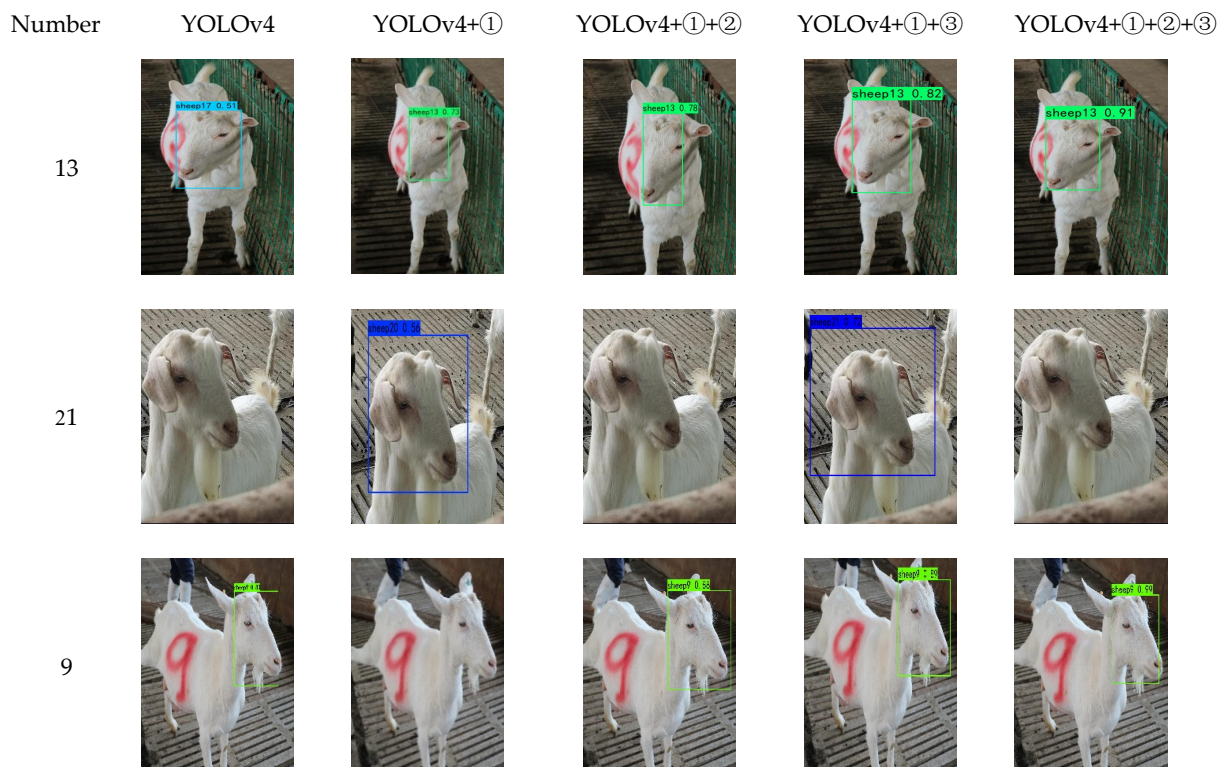

Figure 9. Side-face-recognition results of different models (IOU = 0.5).

5. Conclusions

(1) The backbone network in YOLOv4 was replaced by a GhostNet lightweight network structure to address the problems of the large number of YOLOv4 network parameters, low accuracy of goat-face-recognition, and slow recognition speed. After replacing the backbone, the goat-face-recognition network can reduce the number of network parameters and improve the operation speed and detection efficiency of the model.

(2) The SPP and PANet structure in YOLOv4 was changed to a pyramid structure with a spatial attention mechanism and a fusion network with a residual structure in the form of double parameters. The improved goat-face-recognition network enhances the detectability of fine-grained features and improves the detection of similar faces. The improved goat-face-recognition network improved on the frontal face recognition of the YOLOv4 by 2.1%, and the mAP reached 96.7%. In terms of the side-face detection, the improved goat-face-recognition model improved on the YOLOv4 by 7% compared. The model's detection speed was up to 28 frames/s to meet the needs of real-time monitoring. However, the network still needs to be improved in terms of side-face recognition to improve the accuracy with which it identifies individual goats.

(3) This study mainly focuses on the characteristics of goats' facial texture features, which become less different and difficult to recognize. Furthermore, it proposes a low-cost and high-efficiency improved lightweight YOLOv4 face-recognition model. In order to further achieve individual-goat recognition in flock scenarios, future research will be carried out on flock goats on large-scale farms. By constructing a goat-face-detection network, the interception of goat faces will be achieved. The data will be transmitted to the improved YOLOv4 model to achieve the recognition of goats in multiple situations.

References

1. Kukreja, V.; Kumar, D.; Bansal, A.; Solanki, V. Recognizing wheat aphid disease using a novel parallel real-time technique based on mask scoring RCNN. In Proceedings of the International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Noida, India, 28–29 April 2022.

2. Kumar, D.; Kukreja, V. Quantifying the severity of loose smut in wheat using MRCNN. In Proceedings of the International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 23–25 March 2022.

3. Kumar, D.; Kukreja, V. N-CNN based transfer learning method for classification of powdery mildew wheat disease. In Proceedings of the International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 5–7 March 2021.

4. Kumar, D.; Kukreja, V.; Kadyan, V.; Mittal, M. Detection of DoS attacks using machine learning techniques. *Int. J. Veh. Auton. Syst.* **2020**, *15*, 256–270. [CrossRef]

5. Kukreja, V.; Kumar, D. Automatic classification of wheat rust diseases using deep convolutional neural networks. In Proceedings of the International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 3–4 September 2021.

6. Kumar, D.; Vinay, K. An instance segmentation approach for wheat yellow rust disease recognition. In Proceedings of the International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 7–8 December 2021.

7. Kumar, D.; Kukreja, V. Image-based wheat mosaic virus detection with Mask-RCNN model. In Proceedings of the International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 23–25 March 2022.

8. Gu, J.Q.; Wang, Z.H.; Gao, R.H.; Wu, H.R. Recognition method of cow behavior based on combination of image and activities. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 141–150.

9. Wang, K.; Liu, C.H.; Duan, Q.L. Identification of sow oestrus behavior based on MFO-LSTM. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 211–219.

10. Song, H.B.; Ning, M.T.; Ji, C.H.; Li, Z.Y.; Zhu, Q.M. Monitoring of multi-target cow ruminant behavior based on video analysis technology. *Trans. Chin. Soc. Agric. Eng.* **2017**, *34*, 219–225.

11. Tsai, D.M.; Huang, C.Y. A motion and image analysis method for automatic detection of estrus and mating behavior in cattle. *Comput. Electron. Agric. Vol.* **2014**, *104*, 25–31. [CrossRef]

12. Liu, Z.C.; Zhai, T.S.; He, D.J. Research status and progress of individual information monitoring of dairy cows in precision breeding. *Heilongjiang Anim. Sci. Vet. Med.* **2019**, 30–33+38.

13. Xu, B.B.; Wang, W.S.; Guo, L.F.; Chen, G.P. A review and future prospects on cattle recognition based on non-contact identification. *J. Agric. Sci. Technol.* **2020**, *22*, 79–89.

14. Yao, Z.; Tan, H.; Tian, F.; Zhou, Y.; Zhang, C. Research progress of computer vision technology in wisdom goat farm. *China Feed* **2021**, 7–12.

15. Leslie, E.; Hernandez-Jover, M.; Newman, R.; Holyoake, P. Assessment of acute pain experienced by piglets from ear tagging, ear notching and intraperitoneal injectable transponders. *Appl. Anim. Behav. Sci.* **2010**, *127*, 86–95. [CrossRef]

16. Gonzales Barron, U.; Corkey, G.; Barry, B.; Butler, F.; McDonnell, K.; Ward, S. Assessment of retinal recognition technology as a biometric method for goat identification. *Comput. Electron. Agric.* **2008**, *60*, 156–166. [CrossRef]

17. Adell, N.; Puig, P.; Rojas-Olivares, A.; Caja, G.; Carne, S.; Salama, A.A.K. A bivariate model for retinal image identification in lamb. *Comput. Electron. Agric.* **2012**, *87*, 108–112. [CrossRef]

18. Wei, Z. Research on Imperfect Bivine Iris Recognition Based on Combination of Local Features and Gloabal Features. Master's Thesis, Southeast University, Nanjing, China, 2017.

19. Xie, L.L. Design and research of convolutional neural networks based intelligent campus face recognition system. *Comput. Era* **2021**, 72–74.

20. Wang, B.; Le, H.X.; Li, W.J.; Zhang, M.H. Mask detection algorithm based on improved YOLO lightweight network. *Comput. Eng. Appl.* **2021**, *57*, 62–69.

21. Tang, F.G.; Wu, X.D.; Zhu, Z.Y.; Wan, Z.G.; Chang, Y.C.; Du, Z.P.; Gu, L.L. An end-to-end face recognition method with alignment learning. *Optik* **2020**, *205*, 164238. [CrossRef]

22. Chen, Y.S.; Kuan, C.Y.; Hsu, J.T.; Lin, T.T. Lightweight Cow Face Recognition Algorithm based on Few-Shot Learning for Edge Computing Application. In Proceedings of the American Society of Agricultural and Biological Engineers (ASABE), Anaheim, CA, USA, 12–16 July 2021.

23. Kumar, S.; Pandey, A.; Satwik, K.S.R.; Kumar, S.; Singh, S.K.; Singh, A.K.; Mohan, A. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* **2018**, *116*, 1–17. [CrossRef]

24. Huang, W.J.; Zhu, W.X.; Ma, C.H.; Guo, Y.Z. Weber Texture local descriptor for identification of group-housed pigs. *Sensors* **2020**, *20*, 4649. [CrossRef]

25. Weng, Z.; Meng, F.; Liu, S.Q.; Zhang, Y.S.; Zheng, Z.Q. Cattle face recognition based on a Two-Branch convolutional neural network. *Comput. Electron. Agric.* **2022**, *196*, 1–9. [CrossRef]

26. Yang, S.Q.; Liu, Y.Q.H.; Wang, Z.; Han, Y.Y.; Wang, Y.S.; Lan, X.Y. Improved YOLO V4 model for face recognition of diary cow by fusing coordinate information. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 129–135.

27. Yan, H.W.; Liu, Z.Y.; Cui, Q.L.; Hu, Z.W. Multi-object pig detection based on feature pyramid attention and deep convolutional network. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 193–202.

28. Hu, Z.W.; Yang, H.; Lou, T.T. Detection of herd Pigs using double attention feature pyramid Network. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 166–174.

29. He, Y.T.; Li, B.; Zhang, F.; Tao, H.B.; Gu, L.C.; Jiao, J. Pig face recognition based on improved YOLOv3. *J. China Agric. Univ.* **2021**, *26*, 53–62.

30. Wang, R.; Shi, Z.F.; Gao, R.H.; Li, Q.F. Pig individual recognition based on multi-scale convolutional network in variable environment. *Acta Agric. Univ. Jiangxiensis* **2020**, *42*, 391–400.

31. Yang, A.Q.; Xue, Y.J.; Huang, H.S.; Huang, N.; Tong, X.X.; Zhu, X.M.; Yang, X.F.; Mao, L.; Zheng, C. Lactating sow image segmentation based on fully convolutional networks. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 219–225.

32. Han, D.; Wang, B.; Wang, L.; Hou, Y.C.; Tian, H.Q.; Zhang, S.L. Individual Pain Recognition Method of Goat Based on Improved VGGNet. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 311–317.

33. Zhang, H.M.; Zhou, L.X.; Li, Y.H.; Hao, J.Y.; Sun, Y.; Li, S.Q. Sheep face recognition method based on improved MobileFaceNet. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 267–274.