Research Article

Open Set Sheep Face Recognition Based on Euclidean Space Metric

Hongcheng Xue¹, Junping Qin¹, Chao Quan¹, Wei Ren¹, Tong Gao¹, and Jingjing Zhao²

¹College of Data Science and Application, Inner Mongolia University of Technology, Hohhot 010080, China

²Education and Training Division of Agricultural and Animal Husbandry Technology, Generalization Center of Inner Mongolia Autonomous Region, Hohhot 010080, China

Correspondence should be addressed to Junping Qin; qinjunping@imut.edu.cn

Received 13 September 2021; Revised 20 October 2021; Accepted 23 October 2021; Published 12 November 2021

Academic Editor: Pier Luigi Mazzeo

Copyright © 2021 Hongcheng Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As the essential content of intelligent animal husbandry, identifying each livestock is the only way to achieve modern and refined scientific husbandry. This paper proposes a sheep face recognition method based on European spatial metrics and realizes noncontact sheep identity recognition by training the network using sheep face image samples in the natural environment. The SheepBase data set was first proposed in this process, which contains 6559 images of Inner Mongolia fine-wool sheep and Sunite sheep. To enhance the diversity of the data, the sheep face images were data-enhanced. Secondly, to solve the problems of more redundant information in the sheep face image and the poor posture and angle of the sheep face, we propose the sheep face detection and correction (SheepFaceRepair) method. This method aims to detect the sheep face area in the image to be recognized and align the sheep face area. On this basis, we offer an open sheep facial recognition network (SheepFaceNet) based on the European spatial metric. This method incorporates the biological identity information features of the sheep face to achieve sheep identity. We also tested the effectiveness of this method in the SheepBase data set. The experimental results show that the method proposed in this paper is much higher than the other methods, and the precision of recognition reaches 89.12%. In addition, we found that integrating the biometrics of the sheep face can effectively improve the network's recognition capacity.

1. Introduction

Large-scale farming can increase production efficiencies and production levels in the meat and dairy industries. It effectively increases farmers' incomes, improves food security, improves disease prevention and control capacities, and ensures coordinated livestock and environmental development. Common contact individual identification techniques, such as hot-iron marking [1], frozen marking [2], and ear incisions [3], seriously harm the animal's body, adversely affect animal welfare, and are more limited. For this purpose, this article proposes a noncontact method of identifying the ovine identity; the structure of this method is shown in Figure 1.

The biometric-based noncontact recognition method [4] has a wide range of applications, requires low recognition distance, is convenient to operate, does not require large human consumption, and does not directly contact

individual animals, which minimizes the harm to animals. Biological characteristics [5] are also unique, not easy to forget or lose. With biological characteristics, a method for identifying noncontact animals can be made easier to use, reliable, and precise. Biological features include nasal striations, iris patterns, retinal blood vessels, animal faces, and so on and are primarily used in animal identification. Animal nasal striations are similar to human fingerprints. Each has its characteristics and is different, which can become the main feature of animal identification. In the process of exploring the problem of using nasal striations to achieve noncontact animal identification, Tharawat tried to use a feature extraction method based on Gabor filters and compared SVM classifier [6] with different kernels (Gauss, polynomial, linear); the results show that the Gaussian kernel-based SVM classifier can achieve 99.5% accuracy on the problem of nose print recognition [7]. However, due to the difficulty of collecting nasal striations and the collected

FIGURE 1: Use sheep face target detection and sheep face alignment methods to process the augmentation sheep face data, remove a large amount of redundant information, and correct the sheep face posture. The processed images are sent to SheepFaceNet to extract image features and transform them into Euclidean spatial distance. Euclidean spatial distance is used as the standard to judge the identity of sheep.

rhinoscope pictures containing more redundant information, this type of identification method is not suitable for large-scale animal identification. Like the human iris [8] also contains spots, filaments, crowns, stripes, and other shape features. The combination method will not change after birth, and it is unique. Therefore, it can become one of the essential features of individual identification. Some studies have used the two-dimensional complex wavelet transform feature (2D-CWT) [9] method to study cattle iris biometrics. Experiments on cattle iris images collected by contactless handheld devices have achieved an identification rate of 98.33% [10]. However, due to the large size of the iris image acquisition equipment, the high cost, and the image distortion due to the camera lens [11], the recognition reliability is reduced. The distance between the identification equipment and the individual animal is relatively strict, so it cannot popularize widely. Therefore, the biological feature of the animal face [12] has been paid attention to by a wide range of scholars. The face is the most direct external feature information of an individual animal. Because of the difference and uniqueness of the facial features, the animal face can turn into an identifier for individual identification. Kim et al. [13] verified that image

processing technology could be applied to cattle individual recognition without pattern features, collected the image data of 12 Japanese Wagyu cattle, and calculated the feature parameters, and the feature parameters were input into the neural network for learning by adjusting the image brightness, contrast, and noise to verify the algorithm's feasibility. However, the small amount of data makes it challenging to drive a neural network model with many parameters, and the classifier cannot converge normally. In 2018, Hansen et al. [14] used an improved convolutional neural network [15] to extract features of pig face images and identify ten individual pigs. Because the color and facial features of individual pigs are not apparent, and the overall environment is more complicated, there are many other interfering factors, and the recognition rate of individual pigs is low. The use of deep learning methods to solve noncontact sheep face recognition needs to develop further. This paper investigates the relevant theories of animal face recognition and sheep face biometrics and summarizes the following three challenges in sheep face recognition. First of all, there is no open-source sheep face data set with a large amount of data as support, so we need to collect and make a sheep face data set ourselves. Secondly, the sheep to be

identified are all in a complex environment, and the position and posture of the sheep's face will affect the recognition effect. It is necessary to extract the sheep's facial area from the image, remove the interference information in the image, and align the sheep's face to reduce the influence of the angle and posture of the sheep's face on the recognition accuracy. Finally, unlike human faces, sheep faces have unique biometric information characteristics [16–18], and sheep face recognition needs to make full use of these biological information characteristics [19]. This paper proposes a processing method of first detection, then alignment, and finally recognition, which accurately extracts the sheep face region, corrects the sheep face posture, transforms the sheep face image features into Euclidean space vector, and integrates the biological identity information features of sheep face, which greatly improves the accuracy of sheep face recognition. The main contributions of this article are as follows:

- (1) We collected sheep images in the natural environment and made SheepBase data sets, including Sunit sheep and Inner Mongolia fine-wool sheep. It is necessary to perform data enhancement on the sheep face image before training to solve the overfitting problem caused by the small amount of data. After the expansion, we obtained 6559 sheep face images.
- (2) This paper proposes the SheepFaceRepair method to solve the problems of many interference factors and improper position and posture of sheep face in complex environments. The sheep face detection neural network model based on more anchor frames, different coverage thresholds, and different detection frame merging algorithms is used to eliminate the redundant information in the picture and only keep the sheep face area, use a horizontal alignment method based on the centerline of the two eyes, and align the face of the sheep and correct the posture of the face.
- (3) In this paper, SheepFaceNet is proposed to carry out contactless identification of sheep. In this process, the characteristics of biological identity information of sheep are analysed, and the feature extraction network maps the extracted feature information to the hypersphere, which will fully integrate the biological identity information features of sheep face and transform it into Euclidean space vector, using the Euclidean spatial distance to distinguish the identity of sheep.

The structure of this paper is as follows: The second section introduces the methods of individual animal identification and expounds on the limitations and hazards of traditional methods. The third section explains the essential methods used in the paper. First, introduce the sheep face target detection data set, the sheep face key point detection data set, and the sheep face recognition data set. Then describe the sheep face target detection method, sheep face alignment method, and their improvement scheme. Finally, propose the Sheet Face Net network model. The fourth section explains the method proposed in

this paper through experiments. The fifth section is the conclusion of this paper.

2. Related Work

With the rapid development of intensive, large-scale, and intelligent breeding, the quality of animal breeding management and the requirements for healthy breeding are constantly improving, and individual animal identification is becoming more and more essential to prevent diseases and improve animal growth. At present, individual animals' identification methods can divide into contact-based identification technology and noncontact-based identification technology.

The contact animal identification method requires external tools to leave permanent marks on the animal's body or wear identity information devices. This type of identification method is harmful and causes irreparable damage to animals. It is not only time-consuming and laborious but also having a poor recognition effect. The contact identification method includes permanent identification method, temporary identification method, and electronic label method [20]. Permanent identification methods include ear prints, hot-iron branding, and freeze marking. This type of method directly causes irreversible damage to individual animals. The temporary identification method is ear tag identification technology, which needs to be read manually. The ear tag is easy to fail due to damage, stain, falling off, and loss. The piercing ear tags will cause physical damage to the sheep's body, and improper installation will even tear the eardrum. The electronic label method is a method of marking embedded microchips. Electronic ear tag based on RFID [21] has become the mainstream identification scheme. Compared with the previous methods, this method has a simple operation. It can obtain the identification information of livestock through RFID reading equipment, which effectively improves the identification efficiency. However, the cost of electronic ear tags and identification equipment is relatively high, and the electronic ear tags also have the problem of falling off and missing. Then the animal identity data is lost, and multiple electronic ear tags may conflict in the recognition process, affecting recognition accuracy. In conclusion, contact identification technology cannot provide reliable and effective technical support for animal identification, and it is easy to hurt the animal body, produce stress reactions, and lead to other adverse reactions.

The noncontact recognition method based on biometrics does not need to contact individual animals, and the damage to animals is very low. Many years of research on such problems have also been conducted in Europe and Australia, and other areas with developed animal husbandry have constructed corresponding identification systems. Compared with traditional biological features (nose print, iris pattern, and retinal vessels) [22–24], the animal's face can be recognized from a long distance without harm because of the uniqueness of facial features. Animal faces can be used as identification marks. Animal facial recognition is compared with traditional biometric methods such as iris recognition and nose print recognition. It has the characteristics of

naturalness, noncontact, nonintrusiveness, and quick and convenient collection. The identification method is the same as the biological characteristics used when the animal performs individual identification, without direct contact with animals, which is not easy to make animals uneasy. After solving the accuracy problem of animal face recognition in a complex environment, animal face recognition is expected to become the mainstream recognition technology in the market.

3. Approach

3.1. Data Processing. There are apparent facial differences during the growth of sheep. Adult breeding sheep are in estrus in autumn and winter and will be eliminated in about 3~5 years [25]. During this period, the facial features of adult sheep changed a little. Therefore, adult sheep are selected as the sheep face recognition target. Because the living environment of sheep is scattered and the posture of sheep is not easy to control, it is not easy to obtain the front face. In the actual breeding situation, the position of sheep under the same sheep pen is constantly changing, which will cause repeated shooting, and the captured image may contain multiple individuals. This type of image has a more significant impact on the subsequent recognition accuracy. The method in this paper needs to crop the sheep face from the image and remove redundant information as much as possible. Afterward, it is necessary to align the cropped sheep face to reduce facial posture and angle influence on the recognition accuracy.

In order to increase the adaptability of the model in different scenes, five methods of noise disturbance, random adjustment of brightness, horizontal flip, random adjustment of saturation, and random adjustment of contrast are used to expand the sheep face data set. Random adjustments of brightness, saturation, and contrast are all performed. The parameters are randomly selected to ensure the randomness of the generated image. The data expansion effect is shown in Figure 2.

Different data sets need to be constructed to meet the experiment's needs for different scenarios and functions. This topic is based on the research of sheep face recognition based on European spatial metrics. The SheepBase data set is made, including the following three parts: sheep face target detection data set, sheep face key point detection data set, and sheep face recognition data set.

The sheep face target detection data set format is the same as that of the Pascal VOC 2007 data set, selecting randomly 80% of the data set as the training set and 20% as the test set. The pictures used in the sheep face key point detection data set and the sheep face target detection data set are the same batch of images, and the data expansion method is also the same. Follow the dlib_faces_5points format to make the sheep face key point detection data set. It is necessary to mark the left eye's inner and outer corners of the left eye, the inner and outer corners, and the tip of the nose. The sheep face recognition data set format is the same as that of the LFW face recognition data set, and the pictures are unified as 640×480 sizes. The sheep face detection

algorithm processes all sheep face images. Only retain the sheep face information, and the key point detection algorithm aligns the sheep face.

3.2. Sheep Face Target Detection. This paper uses the self-made sheep face target detection data set to train the improved Faster-RCNN [26] to complete the sheep face target detection task and extract the sheep face region. Given the particularity of sheep face data, if the original Faster-RCNN model directly applies to sheep face detection, there may be three disadvantages: (a) The original fast RCNN uses few training image size types, poor detection performance for targets with large differences in size; (b) the size of the sheep's face in different scenes is quite different, and it may not be possible to detect smaller targets; (c) the two targets to be detected are close in the image, and there is partial overlap, which is prone to missed detection and false detection.

Based on the above reasons, this paper proposes a Faster-RCNN-based model, which combines multiscale training [27], increases the number of anchors, and introduces the Soft-NMS [28] algorithm to optimize the original model, which improves the situation of significant differences in target sizes. It overcomes the possibility of missed target detection with less confidence when two targets are relatively close and partially overlapped. The problem is that the bounding box with a higher confidence score is not always more reliable than the bounding box with lower confidence. The accuracy of sheep face detection is improved. The improved Faster-RCNN network model is shown in Figure 3.

3.2.1. Multiscale Training. This paper uses multiscale images for training, which mainly solves the problem of poor generalization performance of original Faster-RCNN for target detection of different sizes. Before sending the image to the neural network model, the image size is randomly adjusted on the premise of ensuring the proportion of the original image. Experiments have proved that the participation of multiscale images in training enables the network to learn the characteristics of targets of various sizes, making the network have stronger generalization capabilities for targets of different sizes.

3.2.2. Increase the Number of Anchors. The number of anchors in RPN is a very important hyperparameter in the network, which directly affects the generation of subsequent candidate regions. Increasing the number of anchors generates regional candidate boxes for smaller targets to avoid missing detection as much as possible. In the enclosure environment, sheep are dense, and some sheep faces in the pictures are small, so a smaller anchor is needed to generate a regional candidate box for them. The original Faster-RCNN uses nine different anchors, and each sliding window generates nine candidate regions of different sizes and different aspect ratios. Finally, a nonmaximum suppression algorithm eliminates the redundant candidate regions generated in the

FIGURE 2: Data expansion.

FIGURE 3: Improved Faster-RCNN network model.

entire image. However, the anchor parameter set by default in the network cannot recall targets with small areas. Therefore, based on the default parameters, adding a group of 64×64 anchor (smaller than default) allows the network

to detect more targets of smaller size. During the training process, the RPN part uses 12 different anchors. The dimensions are 64×64 , 128×128 , 256×256 , and 512×512 . The three aspect ratios are 1:1, 1:2, and 2:1. The

experiment proves that the increased size of 64×64 in this paper can detect more objects of smaller size.

3.2.3. Soft-NMS Algorithm. NMS [29], as a necessary processing step in the target detection algorithm, aims to eliminate the situation where the same target is detected multiple times. However, the NMS algorithm has some problems. When two objects of the same category in the picture are close and partially overlapped, the detection algorithm should output the boxes of two objects. However, the traditional NMS algorithm will directly eliminate the boxes with low scores, resulting in only one object being detected. Compared with the violent frame removal method of the conventional NMS algorithm, the Soft-NMS algorithm will be much softer. When the Soft-NMS algorithm detects that most of the detection frames overlap, it will reduce the score of the detection frame according to the confidence, rather than directly set the score of the detection frame with low confidence to zero. In addition, the Soft-NMS algorithm has the same complexity as the traditional NMS. It can easily introduce into the target detection algorithm, a significant advantage of the Soft-NMS algorithm.

The traditional NMS processing method can be expressed by the following Rescoring Function; see the following equation:

$$s_i = \begin{cases} s_i, & \text{iou}(M, b_i) < N_t, \\ 0, & \text{iou}(M, b_i) \geq N_t. \end{cases} \quad (1)$$

In the above equation, NMS uses a clear threshold to determine whether to retain adjacent detection frames. Soft-NMS improves the original score reset function

$$s_i = \begin{cases} s_i, & \text{iou}(M, b_i) < N_t, \\ s_i, & (1 - \text{iou}(M, b_i)) \geq N_t. \end{cases} \quad (2)$$

When the overlap between the adjacent detection frame and M exceeds the overlap threshold N_t , the detection score of the detection frame decays linearly. In this case, the detection frame close to M decays greatly, but the frame far away from M will not be affected. However, this score reset function is not continuous. When the overlap degree exceeds the overlap threshold N_t , this score reset function will greatly change the detection result sequence. Therefore, it is necessary to use a continuous score reset function, which has no attenuation effect on the original detection score of the detection frame without overlap. At the same time, it can attenuate highly overlapping detection frames, so the score reset function of Soft-NMS is modified. See the following equation:

$$s_i = s_i e^{-(\text{iou}(M, b_i))^2 / \sigma}, \quad (\text{Does not belong to } D \text{ for any } b_i). \quad (3)$$

Although the Soft-NMS algorithm is also greedy, it always makes the best choice that seems to be the best at the moment. It does not guarantee that the optimal global detection frame score will be reset. The result is local optimal

in a certain sense. But the Soft-NMS algorithm is a more general nonmaximum suppression algorithm.

3.3. Sheep Face Alignment. After completing the sheep face detection, we need to use the self-made sheep face key point detection data set to train the key point detection algorithm. Firstly, initialize the sheep face feature points, extract the corresponding HOG [30] (histogram of gradient) features and LBP [31] (local binary patterns) features according to the current feature point distribution, and extract the features of the image in combination with hog and LBP features, to ensure the stability of image features to geometric deformation and optical deformation and reduce the computational complexity. It can adapt to sheep face feature extraction tasks from different angles. Input the extracted features into the trained weak regressor to obtain the residual distribution of feature points, and update the distribution of current feature points according to the residual distribution of feature points. We can realize the distribution of feature points close to the real sheep face by repeating the above process. The iterative process of the algorithm is

$$S^{\wedge(t+1)} = S^{\wedge(t)} + R_t(h(I, S^{\wedge(t)})), \quad t = 1, \dots, T. \quad (4)$$

$S^{\wedge(t)}$ represents the distribution of feature points after T -stage regression, R_t represents the updated amount of T -stage regressor, I represents the face image to be detected, and H is the hog feature extraction operator.

The key point detection algorithm can obtain the positions of the left and right inner and outer corners of the eyes and the tip of the nose in the picture. Use the self-made sheep face key point detection data set to train the Dlib key point detection algorithm to obtain the left and right positions of the inner and outer corners of the eyes and the tip of the nose. Calculate the center of each eye according to the position information of the left and right inner and outer corners of the eyes in the sheep's face and then calculate the angle between the line of the eye center and the horizontal line of the image, with the intersection of the line between the center of the two eyes and the horizontal line of the image as the origin and angle as the rotation angle to get the aligned sheep face. The schematic diagram of sheep face alignment is shown in Figure 4.

As shown in the figure above, the angle between the centerline of the two eyes and the horizontal coordinate axis is Angle, and the rotation angle is α . The point (x, y) rotates clockwise around the origin by an angle α , and the point (x', y') is obtained. The matrix explanation is shown in the following equation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}. \quad (5)$$

Use matrix multiplication to expand. See the following equation:

$$\begin{cases} x' = x \cdot \cos \alpha + y \cdot \sin \alpha, \\ y' = y \cdot \cos \alpha - x \cdot \sin \alpha. \end{cases} \quad (6)$$

FIGURE 4: Sheep face alignment.

The polar coordinates are explained in the following equation:

$$\begin{aligned}
 (x, y) \text{Coordinate s: } & \begin{cases} r \cdot \cos(\text{Angle}), \\ r \cdot \sin(\text{Angle}), \end{cases} \\
 (x', y') \text{Coordinate s: } & \begin{cases} x' = r \cdot \cos(\text{Angle} - \alpha) = r \cdot \cos(\text{Angle}) \cdot \cos \alpha + r \cdot \sin(\text{Angle}) \cdot \sin \alpha \\ \quad = x \cdot \cos \alpha + y \cdot \sin \alpha, \\ y' = r \cdot \sin(\text{Angle} - \alpha) = r \cdot \sin(\text{Angle}) \cos \alpha - r \cdot \cos(\text{Angle}) \sin \alpha \\ \quad = y \cdot \cos \alpha - x \cdot \sin \alpha. \end{cases} \quad (7)
 \end{aligned}$$

All sheep face images are processed by sheep face detection. Only retained is the sheep face facial information. The key point detection and alignment algorithms process the sheep face images to complete the alignment. So far, all data preprocessing tasks have been completed.

3.4. SheepFaceNet. In actual scenes, sheep face recognition is an open set problem. In a breeding factory, population reproduction, purchase, and sale of individual sheep often occur. The flock is dynamically changing, and the training set cannot include all individual sheep. Retraining the neural network is not feasible when new sheep are added to the population, so sheep face recognition is not a simple classification problem. The method used for the open set problem of sheep face recognition is to map the feature vector of the sheep face extracted by the neural network to

the Euclidean space. The distance of the space vector directly corresponds to the similarity of the sheep face. The biological identity information characteristics of sheep face are innate, mainly including eyes, nose, mouth, ears, and horns. The combination of these parts realizes the uniqueness of each sheep's face. In sheep face recognition, the process of transforming the extracted feature vectors into Euclidean space vectors incorporates biometric identity information features, which can more accurately describe the differences and uniqueness of each sheep face and improve the recognition accuracy. The biological identity information characteristics of the sheep's face are shown in Figure 5.

This article uses an end-to-end encoding method for learning an image to Euclidean space. The distance in Euclidean space is directly related to the similarity of the category. Unlike the traditional method, the distance between points on the feature space indicates whether the two

FIGURE 5: Features of biometric identification information of sheep face.

images are of the same type. Set a threshold T . When the spatial distance is less than T , it belongs to the same sheep, and when the space is greater than T , judge it as a different sheep. The structure diagram of the network model used is shown in Figure 6.

The part of the interest in different images is quite different, and it is more challenging to select the appropriate convolution kernel size for the convolution operation. A larger convolution kernel is more suitable for targets with more global information distribution, and a smaller convolution kernel is ideal for targets with more local information distribution. If a very deep network structure is used, overfitting will occur, and all gradients cannot be effectively transmitted and consume computational resources. Therefore, the inception network with multiple size convolution kernels on the same layer came into being, making the network broader rather than deeper. Choose Inception-V1 [33] and Inception-V2 [34] as the feature extraction network of SheepFaceNet to extract the features of sheep face. First, download the two feature extraction network model parameters pretrained on the CASIA-WebFace data set. On this basis, use the self-made sheep face recognition data set for retraining. The Inception-V2 network uses two superimposed 3×3 convolution kernels to replace the 5×5 convolution kernel in the Inception-V1 network. Because the computational cost of a 5×5 convolution kernel is 2.78 times that of 3×3 , the performance of superimposing two 3×3 convolutions will be improved. The use of large-size convolution kernels to extract features may cause the loss of

FIGURE 6: SheepFaceNet network structure: deep architecture refers to feature extraction networks, such as inception, ZFNet; L2 refers to feature normalization (make the length of its eigenvector 1, so that all feature points will be mapped to Euclidean space vector); embedding refers to the feature vector after L2 normalization, and this feature vector represents a sheep face picture; center loss refers to the center loss function [32], which specifies a category center for each class (in the sheep face recognition model, one class corresponds to a sheep). The corresponding features of the same type of image should be as close as possible to their category center, and the category centers of different categories should be as far away as possible. The center loss can make the training data have “cohesion.”

apparent features such as eyes, mouth, and ears of the sheep's face. In addition, the Inception-V2 network decomposes the size of the $n \times n$ convolution kernel into two convolutions of $1 \times n$ and $n \times 1$. This method can effectively reduce the cost. In the Inception-V2 network, a 3×3 convolution is equivalent to performing a 1×3 convolution and then a 3×1 convolution. Such a convolution method will reduce the computational cost by 33% compared to a simple 3×3 convolution. Before the feature extraction network, batch normalization (BN) [35] is added. The BN structure makes the change of all weights consistent, enhances the model's generalization ability, prevents overfitting, and reduces the amount of model calculation.

4. Experiment

4.1. Experimental Analysis of Sheep Face Target Detection

4.1.1. Comparison of Detection Effects under Different Coverage Thresholds. The impact of four sets of different coverage thresholds on the detection results is compared, the foreground threshold is 0.7, the background threshold is 0.3; the foreground threshold is 0.7, the background threshold is 0.4; the foreground threshold is 0.6, the background threshold is 0.3; the foreground threshold is 0.6, the background threshold is 0.4. The parameter settings and training steps are the same to ensure the fairness of the comparative experiments in each group, except for the different coverage thresholds. The experimental results are shown in Table 1.

It can be seen from Table 1 that, by changing different coverage thresholds, the foreground threshold is 0.6, and the background threshold is 0.3. Compared with the other three sets of coverage thresholds, the mAP results are increased by 0.0353, 0.0081, and 0.0037. It is fully proved that the detection effect of sheep face is the best when the current scene coverage threshold is 0.6, and the background coverage threshold is 0.3.

4.1.2. Comparison of Detection Results after Increasing the Anchor Size. The preset anchor size used in the original fast RCNN is $\{128 \times 128, 256 \times 256, 512 \times 512\}$; this size can not well cover the sheep face scale range in the self-made sheep

TABLE 1: Comparison of detection results of different coverage area thresholds.

Coverage threshold	Pre (%)	Rec (%)	mAP (%)
0.7~0.3	0.8802	0.8916	0.8719
0.7~0.4	0.8856	0.8991	0.8991
0.6~0.3	0.8963	0.9127	0.9072
0.6~0.4	0.8887	0.9001	0.9035

face target detection data set. Therefore, the study compared the detection performance results of different anchor sizes in the sheep face data set. The anchor size distribution is group A $\{128 \times 128, 256 \times 256, 512 \times 512\}$, group B $\{64 \times 64, 128 \times 128, 256 \times 256\}$, and group C $\{64 \times 64, 128 \times 128, 256 \times 256, 512 \times 512\}$; the experimental results are shown in Table 2.

It can be seen from Table 2 that, by adding a group of 64×64 anchors, the size distribution can better cover the size range of the objects in the data set. After removing an anchor with a size of 512×512 in group B, the mAP is lower than that of the original Faster-RCNN because the proportion of training images with normal-sized targets in the data set is larger than that of smaller-sized training images. After adding a group of anchors, the map increased by 0.0097 compared with the original fast RCNN. Experiments show that adding a smaller anchor can effectively improve the detection accuracy of small-size targets and the regression performance of detection frames. Compared with fast RCNN, the time spent training and testing increased slightly after adding a group of anchors.

4.1.3. Comparison of Detection Results after Introducing the Soft-NMS Algorithm. Finally, using the model shown in the last row of Table 2 as a benchmark, the Soft-NMS algorithm is introduced to optimize the network model. The experimental results are shown in Table 3.

As shown from Table 3, after introducing the Soft-NMS algorithm, the mAP increases by 0.0104. Finally, combining three methods to improve the target detection results effectively, the average accuracy of the target detection category of the method in the sheep face reached 0.9273, which enhances the mAP value by 0.0554 compared to the original Faster-RCNN. Finally, Figure 7 shows the change of loss value between the proposed method (Figure 7(a)) and the original Faster-RCNN (Figure 7(b)).

4.2. Experimental Analysis of Sheep Face Recognition. This chapter designs four groups of comparative experiments. The first group uses Inception-V1 and Inception-V2 as feature extraction networks to verify the impact of different feature extraction networks on recognition accuracy. The second group of feature extraction networks uses the center loss function and the triple loss function [36] to join the training process to explore whether the triple loss function is better than the central loss function under small-batch data. In the third set of experiments, use sheep face detection, sheep face key point detection, and sheep face alignment to process data, the data that has only been processed by sheep

TABLE 2: Comparison of test results of different anchor sizes.

Anchor size	Pre (%)	Rec (%)	mAP (%)
Group A	0.8963	0.9127	0.9072
Group B	0.8876	0.8881	0.8918
Group C	0.9097	0.9231	0.9169

TABLE 3: Influence of Soft-NMS algorithm on detection results.

	Pre (%)	Rec (%)	mAP (%)
NMS	0.9097	0.9231	0.9169
Soft-NMS	0.9218	0.9410	0.9273

face target detection, and the data that has not been processed are used as training data. Explore the necessity of sheep face target detection and sheep face key point detection and alignment. The fourth group compares the traditional classification method supervised by the central loss function with the European spatial measurement method proposed in this paper to explore the impact on the recognition accuracy after fusing the biometric features of sheep face.

4.2.1. SheepFaceNet Based on Different Feature Extraction Networks. The position of the part of interest is quite different in different images, and it is more challenging to select the appropriate convolution kernel size for the convolution operation. The larger convolution kernel is more suitable for global targets, and the smaller convolution kernel is ideal for local targets. If a very deep network structure is used, overfitting will occur, all gradients cannot be effectively transmitted, and computational resources are consumed. Therefore, the Inception network with multiple size convolution kernels on the same layer came into being, making the network wider rather than deeper. In this paper, Inception-V1 and Inception-V2 are selected as the feature extraction network of SheepFaceNet to extract sheep face features. First, download the two feature extraction network model parameters pretrained on the CASIA-WebFace data set. On this basis, use the self-made sheep face recognition data set for retraining. In the network training process, the model learning rate is 0.001, the weight attenuation term is 0.005, the neuron retention rate is 0.8, and the batch_size is 32. The final experimental results show that, using Inception-V2 as a feature extraction network, the recognition accuracy is better than Inception-V1. The experimental results are shown in Figure 8.

The Inception-V2 network uses two superimposed 3×3 convolution kernels to replace the 5×5 convolution kernel in the Inception-V1 network. The 5×5 convolution kernel is 2.78 times the computational cost of the 3×3 convolution kernel, which will improve the performance of superimposing two 3×3 convolutions. In addition, the Inception-V2 network decomposes the size of the $n \times n$ convolution kernel into two convolutions of $1 \times n$ and $n \times 1$. This method can effectively reduce the cost. In the Inception-V2 network, 3×3 convolution is equivalent to performing 1×3

FIGURE 7: Loss value change graph. (a) The method in this paper. (b) Original Faster-RCNN.

FIGURE 8: Recognition accuracy.

convolution first and then performing 3×1 convolution. This convolution method will reduce the cost by 33% compared to 3×3 convolution.

It can be seen from the experimental results that the recognition accuracy of using Inception-V2 as the feature extraction network on the self-made sheep face data set is 85.71%, which is 2.55% higher than that of using Inception-V1 as the feature extraction network. Although the improvement effect is not apparent, the training time will be shortened by about 0.0005 seconds in each epoch, and the overall performance is better than that of Inception-V1.

4.2.2. SheepFaceNet Based on Different Loss Functions.

The objective of triplet loss function optimization is the distance between features, which needs to design for this distance. Specifically, three sheep face images are extracted from the training data each time. The first image is marked as x_i^a (I represent the photo of sheep selected for the I -th time), the second image is marked as x_i^p , and the third image is marked as x_i^n . In such a triplet, x_i^a and x_i^p correspond to the

image of the same sheep, and x_i^n is the sheep's face image of another sheep. So the length $\|f(x_i^a) - f(x_i^p)\|_2$ should be closer, and the length $\|f(x_i^a) - f(x_i^n)\|_2$ should be further. Strictly speaking, the triplet loss requirement satisfies the following inequality (8):

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2. \quad (8)$$

The spatial distance of facial images of the same sheep is at least shorter than that of different sheep α (square is mainly for the convenience of derivation). Therefore, the design loss function is shown in formula (9):

$$L_i = \left[\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha - \|f(x_i^a) - f(x_i^n)\|_2^2 \right]. \quad (9)$$

In the process of training networks, it is challenging to select appropriate triples. If A , P , and N (A , P , and N replace $f(x_i^a)$, $f(x_i^p)$, and $f(x_i^n)$) are randomly chosen to form a training set, the principle is that A and P are the same sheep and A and N are different sheep. Suppose triples are selected according to this principle. In that case, this constraint $(d(A, P) + \alpha \leq d(A, N))$ ($d(A, P)$ replace $\|f(x_i^a) - f(x_i^p)\|_2$, $d(A, N)$ replace $\|f(x_i^a) - f(x_i^n)\|_2$) will be easy to achieve because the probability of A and N being very different from A and P is enormous. The robustness of the network will be very poor according to this principle. If the most challenging triples are selected as the training data each time, it is difficult for the model to converge correctly.

Therefore, we should choose the "semihard" triples A , P , and N as much as possible. The "semihard" data training triplet is that the selection of A , P , and N will make $d(A, P)$ very close to $d(A, N)$, i.e., $d(A, P) \approx d(A, N)$. The triplet selected will try its best to make the formula on the right $d(A, N)$ larger or the formula on the left $d(A, P)$ smaller. In this way, there is at least a distance α on the left and right sides. Many sheep face data are required to generate a sufficient number of "semihard" triples to make the model converge successfully; the triplet loss learning process is shown in Figure 9.

FIGURE 9: Triplet loss: negative refers to the negative sample most similar to the sample, and positive refers to the positive sample least identical to the sample. Through the supervision of the triple loss function and combined with the “semihard” triple data training network, the neural network can learn the ability to distinguish positive and negative samples better. (a) Before learning, the model cannot correctly measure the spatial distance between the standard sample and the positive sample and the negative sample. (b) The model’s learning process of “semihard” data optimizes the model’s ability to measure the spatial distance between the sample and the positive and negative samples. (c) It means that after learning the triple loss function supervised model, the model can correctly measure the spatial distance between standard samples and positive and negative samples and accurately distinguish sample categories.

Different from triplet loss, the center loss does not directly optimize the distance. It retains the original classification model but specifies a category center for each class (in the sheep face recognition model, a class corresponds to a sheep). The corresponding features of the same class of images should be as close to their category centers as possible. The category centers of different classes should be as far away as possible. The center loss can make the training data “cohesive.” Compared with triplet loss, using center loss to train the sheep face recognition model does not need a unique sampling method. Moreover, it can achieve an effect similar to triplet loss by using fewer images; the center loss learning process is shown in Figure 10.

Due to the small number of sheep face images collected by oneself, the number of images in the large public data set cannot be reached. Therefore, this article will compare and verify the three-tuple loss function and the center loss function, which are two loss functions that have better spatial distance optimization effects, and study the convergence of small-batch self-made sheep face data sets on different loss functions. According to the results of the first set of comparative experiments, select Inception-V2 as the feature extraction network, and choose different loss functions, respectively. The change of loss value during training is shown in Figure 11. Center loss is on the left, and triplet loss is on the right.

It can see from Figure 11 that the convergence of the central loss function on the self-made sheep face recognition data set is significantly better than the triplet loss function.

Therefore, this article will use the SheepFaceNet neural network model based on the central loss function and the experimental results are shown in Figure 12.

It can be seen from the above figure that after replacing the triple loss function with the central loss function, the recognition rate on the self-made sheep face recognition data set reaches 89.12%. The accuracy is improved by 3.41% compared with the original SheepFaceNet neural network model using the triple loss function, which fully proves that the triple loss function is more suitable for large-scale data sets and has a poor convergence effect on small-batch data.

4.2.3. The Impact of Sheep Face Key Point Detection and Alignment on the Recognition Rate. In the third chapter, introduce the content of sheep face key point detection and alignment in detail. After detecting the face area, modify the sheep face image to reduce the impact of posture on recognition accuracy. In this part, divide the data sets into three categories. Group A is the sheep face recognition data set after sheep face target detection, key point detection, and alignment. Group B is the sheep face recognition data set after sheep face target detection without key point detection and alignment. Group C is the sheep face recognition data set without any processing. The number of pictures in the three types of data sets is the same. The data expansion method is the same. The three types of data sets have the same number of pictures, and the data expansion method is the same. The feature extraction network and loss function

FIGURE 10: Center loss: (a) For image features extracted by the model, the intraclass distance is even greater than the interclass distance (b). (c) represents the feature description learned under the center loss supervised training. The sample features are close to the center of the class, compact within the class, and separated between classes.

FIGURE 11: Loss value change graph. (a) Center loss. (b) Triplet loss.

FIGURE 12: Recognition accuracy.

FIGURE 13: Recognition accuracy.

of the three data sets are the same during the training process. The recognition accuracy of the three data sets is shown in Figure 13.

As can be seen from the above figure, the recognition accuracy of group A is 89.12%, that of group B is 87.06%, and that of group C is 85.93%. The recognition accuracy of the

data set after sheep face target detection, sheep face key point detection, and sheep face alignment processing is 2.06% higher than that of the data set without sheep face key point detection and sheep face alignment. The recognition accuracy of the data set after sheep face target detection, sheep face key point detection, and sheep face alignment processing is 3.19% higher than that of the data set without any processing. The improvement rates of the two operations on the experimental accuracy are different. The recognition accuracy of sheep faces target detection is greater than that of sheep face key point detection and alignment.

After the sheep face target detection and cropping operation, the redundant information can be effectively removed except for the sheep face in the picture. Only the sheep face area is retained, which significantly improves the efficiency of feature extraction. The key point detection and alignment operation of the sheep face can eliminate the errors caused by the posture and angle of the sheep face under certain circumstances and “rectify” the sheep face. In terms of improved accuracy, sheep face target detection plays a significant role in this paper’s sheep face recognition task. There are many interferences in the original sheep face image, and the feature extraction network extracts many useless feature vectors. The subsequent measurement of distance in Euclidean space will affect the accuracy. All sheep face data need this step for processing. Secondly, after extracting the sheep face region, carry out the sheep face alignment operation to correct the posture and position of the sheep face improving recognition accuracy.

4.2.4. The Impact on the Recognition Rate after Fusing the Features of the Biometric Identification Information of the Sheep Face. In this part, compare center loss function and softmax classifier with the method in this paper. Group A used the central loss function and softmax classifier to classify the feature vectors extracted by the feature extraction network to distinguish the identity of sheep. Group B performs L2 normalization after obtaining the feature vector of the sheep face, then maps it to the Euclidean space distance, and integrates the biological identity information characteristics of the sheep’s face, using the center loss function to optimize the distance to complete the sheep face recognition; the recognition accuracy of group A and group B is shown in Figure 14.

As can be seen from the above figure, under the condition that the data processing method is consistent, group A is the method of this paper, and the recognition accuracy rate is 89.12%. The recognition accuracy rate of group B is 76.96%, which is significantly lower than the SheepFaceNet model proposed in this paper. The feature vector is mapped into Euclidean space vector after extracting the sheep face feature vector through the feature extraction network. This process will fully integrate the biological identity information features of the sheep face. At this time, some facial feature points with clear semantic information, such as nose tip, eyes, ears, and so on, will be transformed into Euclidean spatial distance. Finally, use the Euclidean space distance to express the similarity of the sheep’s face. The spatial distance of the same

FIGURE 14: Recognition accuracy.

sheep is small, and the spatial distance of different sheep face images is relatively large. After fully integrating the biometric features of sheep face, this method can more accurately transform the sheep face feature vector into Euclidean space vector and significantly improve the recognition accuracy.

5. Conclusion

Given the poor recognition effect of the existing open set sheep identification methods, this paper uses the European spatial distance to correlate the sheep face similarity directly. It proposes a sheep face recognition model based on Euclidean spatial measurement. Using sheep face target detection and sheep face alignment methods to refine the collected data set, filter out the interference information in the picture, and reduce the influence of the sheep face pose and angle on the recognition accuracy. Improve the size and combination of convolution kernels in the feature extraction network, reduce training costs, and better extract sheep face features. The central loss function replaces the triple loss function that has a better effect on the open set problem to solve the problem that the network cannot converge due to the small amount of sheep face data, and the recognition effect is poor.

To further improve the accuracy of the sheep face recognition model, improving the network structure and expanding the data set can be used to improve the recognition ability of the network in harsh environments. At the same time, most recognition networks use end-to-end recognition models, which can not focus on recognizing prominent features. Therefore, increasing the model’s ability to recognize local features can be further studied.

Data Availability

The data set used in the research is uploaded to the Baidu network disk. Please click the link to view or download the data set (extraction code: zrcp) <https://pan.baidu.com/s/1HgNdEYqAz2SXpEbrmEb8UA>.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Doctoral Foundation of the Inner Mongolia University of Technology with Grant no. BS201935, the Natural Science Foundation of Inner Mongolia of China with Grant no. 2019MS06005, and China's National Natural Science Foundation (no. 61962044).

References

- [1] J. Scordino, "Steller sea lions (*Eumetopias jubatus*) of Oregon and Northern California: seasonal haulout abundance patterns, movements of marked juveniles, and effects of hot-iron branding on apparent survival of pups at Rogue Reef," *Steller Sea Lion*, vol. 109, no. 4, pp. 369–381, 2006.
- [2] M. K. Vestal, C. Ward, D. Doye, and D. L. Lalman, "Beef cattle production and management practices and implications for educators," *Research Papers in Economics*, vol. 6, pp. 1636–1681, 2006.
- [3] E. Leslie, M. Hernández-Jover, R. Newman, and P. Holyoake, "Assessment of acute pain experienced by piglets from ear tagging, ear notching and intraperitoneal injectable transponders," *Applied Animal Behaviour Science*, vol. 127, no. 3–4, pp. 86–95, 2010.
- [4] C. W. Pobanz and T. Itoh, "A microwave non-contact identification transponder using subharmonic interrogation," in *Proceedings of the 1994 IEEE MTT-S International Microwave Symposium Digest (Cat. No.94CH3389-4)*, vol. 43, pp. 1673–1679, San Diego, CA, USA, May 1994.
- [5] J. R. Orozco-Arroyave, J. C. Vdsquez-Correa, F. Honig et al., "Towards an automatic monitoring of the neurological state of parkinson's patients from speech," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6490–6494, Shanghai, China, March 2016.
- [6] P.-H. Chen, C.-J. Lin, and B. Schölkopf, "A tutorial on ν -support vector machines," *Applied Stochastic Models in Business and Industry*, vol. 21, no. 2, pp. 111–136, 2005.
- [7] A. Tharwat, T. Gaber, and A. E. Hassanien, "Cattle identification based on muzzle images using gabor features and SVM classifier," in *Proceedings of the 2nd International Conference on Advanced Machine Learning Technologies and Applications*, Cairo, Egypt, November 2014.
- [8] C. Musgrave and J. L. Cambier, "System and method of animal identification and animal transaction authorization using iris patterns," 2002.
- [9] I. W. Selesnick and Y. L. Ke, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Proceedings of the Wavelets: Application in Signal & Image Processing X International Society for Optics and Photonics*, San Diego, CA, USA, 2003.
- [10] A. C. Bovik, "The essential guide to image processing," *Journal of Electronic Imaging*, vol. 19, no. 2, Article ID 29901, 2009.
- [11] A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," in *Proceedings of the Seventh International Symposium on IEEE Signal Processing and its Applications*, White Plains, NY, USA, 2003.
- [12] G. A. Rousselet, M. J.-M. Macé, and M. Fabre-Thorpe, "Comparing animal and face processing in the context of natural scenes using a fast categorization task," *Neurocomputing*, vol. 58–60, pp. 783–791, 2004.
- [13] H. T. Kim, Y. Ikeda, and H. L. Choi, "The identification of Japanese black cattle by their faces," *Asian-Australasian Journal of Animal Sciences*, vol. 18, no. 6, pp. 868–872, 2005.
- [14] M. F. Hansen, M. L. Smith, L. N. Smith et al., "Towards on-farm pig face recognition using convolutional neural networks," *Computers in Industry*, vol. 98, pp. 145–152, 2018.
- [15] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the Computer Vision-ECCV 2014*, pp. 818–833, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Zurich, Switzerland, September 2014.
- [16] J. W. Peirce, A. E. Leigh, and K. M. Kendrick, "Configurational coding, familiarity and the right hemisphere advantage for face recognition in sheep," *Neuropsychologia*, vol. 38, no. 4, pp. 475–483, 2000.
- [17] A. Towler, R. I. Kemp, V. Bruce, A. M. Burton, J. D. Dunn, and D. White, "Are face recognition abilities in humans and sheep really 'comparable'?" *Royal Society Open Science*, vol. 6, p. 1, 2019.
- [18] F. Knolle, R. P. Goncalves, and A. J. Morton, "Sheep recognize familiar and unfamiliar human faces from two-dimensional images," *Royal Society Open Science*, vol. 4, Article ID 171228, 2017.
- [19] J. W. Peirce, A. E. Leigh, A. P. C. daCosta, and K. M. Kendrick, "Human face recognition in sheep: lack of configurational coding and right hemisphere advantage," *Behavioural Processes*, vol. 55, no. 1, pp. 13–26, 2001.
- [20] S. Kumar, S. Tiwari, and S. K. Singh, "Face recognition of cattle: can it be done?" *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, vol. 86, no. 2, pp. 137–148, 2016.
- [21] C. M. Roberts, "Radio frequency identification (RFID)," *Computers & Security*, vol. 25, no. 1, pp. 18–26, 2006.
- [22] W. E. Petersen, "The identification of the bovine by means of nose-prints," *Journal of Dairy Science*, vol. 5, no. 3, pp. 249–258, 1922.
- [23] A. M. Johnston and D. S. Edwards, "Welfare implications of identification of cattle by ear tags," *The Veterinary Record*, vol. 138, no. 25, pp. 612–614, 1996.
- [24] U. Gonzales Barron, G. Corkery, B. Barry, F. Butler, K. McDonnell, and S. Ward, "Assessment of retinal recognition technology as a biometric method for sheep identification," *Computers and Electronics in Agriculture*, vol. 60, no. 2, pp. 156–166, 2008.
- [25] W. Ryan, I. Williams, and R. Moir, "Compensatory growth in sheep and cattle. 1. Growth pattern and feed intake," *Australian Journal of Agricultural Research*, vol. 44, no. 7, pp. 1609–1621, 1993.
- [26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [27] V. M. Ganapathy and K. L. Liew, "Handwritten character recognition using multiscale neural network training technique," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 2, no. 3, pp. 638–643, 2008.
- [28] W. Liu, "SSD: single shot MultiBox detector," *European Conference on Computer Vision*, Springer, Cham, Switzerland, 2016.
- [29] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 850–855, 2006.
- [30] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer*

Society Conference on Computer Vision & Pattern Recognition IEEE, San Diego, CA, USA, 2005.

- [31] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [32] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1735–1742, New York, NY, USA, 2006.
- [33] Q. Cui, S. Lu, B. Ni et al., "Improved prediction of aqueous solubility of novel compounds by going deeper with deep learning," *Frontiers in Oncology*, vol. 10, p. 121, 2020.
- [34] C. Szegedy, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 4278–4284, Phoenix, AZ, USA, 2016.
- [35] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," *Proceedings of the 32nd International Conference on Machine Learning*, vol. 1, pp. 448–456, 2015.
- [36] S. Hongxin, "Triple loss-based improved neural network pedestrian re-identification method," 2017.