Article

Pig Face Recognition Based on Metric Learning by Combining a Residual Network and Attention Mechanism

Rong Wang [1,2] , Ronghua Gao [1,3,*], Qifeng Li [1,3] and Jiabin Dong [1,3]

1   Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
2   College of Information Engineering, Northwest A&F University, Yangling, Xianyang 712100, China
3   National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China
*   Correspondence: gaorh@nercita.org.cn

Abstract: As machine vision technology has advanced, pig face recognition has gained wide attention as an individual pig identification method. This study establishes an improved ResNAM network as a backbone network for pig face image feature extraction by combining an NAM (normalization-based attention module) attention mechanism and a ResNet model to probe non-contact open-set pig face recognition. Then, an open-set pig face recognition framework is designed by integrating three loss functions and two metrics to finish the task with no crossover of individuals in the training and test sets. The SphereFace loss function with the cosine distance as a metric and ResNAM are combined in the framework to obtain the optimal open-set pig face recognition model. To train our model, 37 pigs with a total of 12,993 images were randomly selected from the collected pig face images, and 9 pigs with a total of 3431 images were set as a test set. 900 pairs of positive sample pairs and 900 pairs of negative pairs were obtained from the images in the test set. A series of experimental results show that our accuracy reached 95.28%, which was 2.61% higher than that of a human face recognition model. NAM was more effective in improving the performance of the pig face recognition model than the mainstream BAM (bottleneck attention module) and CBAM (convolutional block attention module). The research results can provide technological support for non-contact open-set individual recognition for intelligent farming processes.

Keywords: pig face recognition; open-set recognition; ResNet; attention mechanism; metric learning

1. Introduction

Intensive pig farms are replacing small-scale breeding models, such as individual breeding, as livestock farming moves toward scale, informatization, and refinement. Technology that identifies individual pigs is important for the daily fine management of individual pigs on large-scale pig farms. In the traditional pig industry, physical tags and radio-frequency identification (RFID) chips are usually applied to identify pigs. Methods based on physical labels, such as pattern marking and ear cutting, can cause stress in pigs and affect animal welfare [1]. With RFID chips of different frequencies, the identification range is different for each chip, which can cause the problem of false identification. There are a variety of non-contact recognition methods based on computer vision that have recently been applied to a variety of animal recognition tasks. These methods require the assistance of only cameras and computing equipment, without additional personnel or other equipment, and they are able to quickly and accurately identify individual animals.

Prior human knowledge was needed to extract image features in the early development of computer vision technology. Kashiha et al. [2] used Fourier descriptors with rotation and translation invariance to preserve the pattern features of 10 pigs, and the accuracy of pig pattern recognition was 88.7%. Zhao et al. [3] proposed a vision system for Holstein cow body image extraction and identity recognition. Side-view images of 66 cows were collected. The methods of features from an accelerated segment test (FAST),

scale-invariant feature transform (SIFT), and fast library for approximate nearest neighbors (FLANN) were used for feature extraction, descriptors, and matching, respectively. The highest recognition accuracy was 96.72%. However, the traditional methods rely on manually selecting features, and the results of feature extraction directly affect the performance of individual cow recognition. With the development of deep learning and computer vision technology, convolutional neural networks (CNNs) have shown strong feature extraction abilities, and they have been applied to the field of livestock by many researchers [4–8]. Hu et al. [9] proposed a novel non-contact cow identification method based on the fusion of deep parts of features by making good use of the YOLO (You Only Look Once) method to detect the cows' heads, trunks, and legs; they built three CNN models to recognize the identities of cows together, and they achieved an accuracy of 98%. Shen et al. [10] segmented side-view images of cows into three parts—head, trunk, and legs—and then they used a YOLO model to detect cow objects in the side-view images and input them into a CNN model to identify each individual; the accuracy was 96.65%. These deep-learning-based recognition methods provided good solutions for pig face recognition. Deep learning algorithms have good feature extraction abilities and can fully extract image features. With the development of parallel GPU (graphics processing unit) acceleration hardware, a variety of CNNs aiming to improve the image feature extraction ability and reduce the number of parameters have been proposed one after another. Two methods—increasing the network depth and decreasing the training parameters—have been used to optimize CNNs. Their unique network structure can improve image the information extraction and training speeds, which has driven the development of the image recognition field. The eight-layer Alexnet [11] network was proposed in the 2012 ILSVRC competition (ImageNet Large Scale Visual Recognition Competition). The network contains five convolutional layers, three fully connected layers, 60 million parameters, and 65,000 neurons in order to achieve an error rate of 15.3%, which far exceeds that of traditional recognition methods. To solve the problem that large convolutional kernels tend to lose image information, the VGG (visual geometry group) [12] network was proposed by the Computer Vision Geometry Group at the University of Oxford. The method of using small convolutional kernels instead of large convolutional kernels increases the depth of a network and improves the feature extraction ability of the network. To achieve a deeper network structure and fewer parameters, the deep residual network (ResNet) [13] was made up to 152 layers deep by using the jump connection method, which solved the problem of deep networks being difficult to train, and it became the mainstream network in the image field.

Based on the mainstream ResNet network, an open-set pig face recognition framework and backbone network are designed in this paper. The rest of this paper is organized as follows. A brief overview of the related work on pig face recognition is presented in Section 2. The acquisition and processing of our data are presented in Section 3. Section 4 describes the open-set pig face recognition framework and methods. Section 5 discusses the training results of the models, the comparison results of different models, and the results of the ablation experiments. Section 6 summarizes the important conclusions about the open-set pig face recognition model.

2. Related Work

Closed-set recognition can only recognize livestock individuals who have appeared in the training set, whereas open-set recognition can recognize livestock individuals that have never been seen in the model according to an assessment of the two types of recognition by Andrew et al. [14]. Therefore, there are two main types of livestock individual recognition: open-set recognition and closed-set recognition.

Closed-set recognition has been studied more often. Hansen et al. [15] designed a nine-layer CNN and SVM (support vector machine) to realize pig face recognition. Marsot et al. [16] proposed a novel framework composed of computer vision algorithms, machine learning, and deep learning techniques with an accuracy of 83% on 320 test-

ing images to provide a relatively low-cost and scalable solution for pig recognition. Salama et al. [8] used Bayesian optimization to find the best CNN for sheep face recognition, with an accuracy of 98%. Wang et al. [17] introduced a Keras convolutional-neural-network-based pig facial recognition model. The model's recognition accuracy was the best, reaching 97.6%. Wang et al. [18] introduced a ternary-loss-function-based pig face recognition approach that lowered intra-class differences while increasing the distance between classes, giving a novel research idea for raising pig face recognition accuracy. The aforementioned deep-learning-based livestock individual recognition algorithms achieved good recognition results, thus proving the viability of utilizing deep learning to identify pig individuals and providing an important reference for future research on pig face recognition. The general drawback with these methods is that the training and test sets contain the same individuals, and the models can only detect livestock individuals that exist in the training set.

There are few studies on open-set recognition of livestock. Andrew et al. proposed a method for recognizing Holstein–Friesian cattle through metric learning and achieved open-set recognition for the first time in individual cow identification. The accuracy of a test set containing seven cows was 93.8%, which laid the foundation for the open-set recognition of cows [14]. Then, Guo et al. achieved 57.0% accuracy in open-set cow recognition in a test set containing 186 Holstein cows [19]. To further improve the accuracy of open-set cow recognition, Xu et al. proposed the CattleFaceNet model; a total of 72 cows were collected as the training set and 9 were used cows as the test set, with an accuracy of 91.35% [20]. These methods provide new ideas for the open-set recognition of livestock, but there is no relevant research on the open-set recognition of pig faces. Therefore, this paper proposes an open-set pig face recognition method with an improved backbone network and a metric method.

The contributions of this study can be summarized as follows.

1. To improve feature extraction from pig face photos, this research offers a feature extraction backbone network (ResNAM) based on a normalized attention mechanism.

2. To increase the accuracy of open-set pig face recognition, a framework for open-set pig face recognition was created by integrating three loss functions and two measurement techniques. The best open-set pig face recognition method was then obtained by combining the ResNAM network, SphereFace loss function [21], and cosine distance.

3. An open-set pig face recognition model based on the BAM and CBAM attention mechanisms was constructed, and ablation experiments were designed to verify the effectiveness of ResNAM.

3. Materials

3.1. Data Acquisition

Pig facial images were collected in August 2018 at Hui Kang Breeding Farm, Tianjin, China. A system for capturing images of pigs' faces that used a positioning pen to fix the camera on a tripod at a height of 50 cm from the positioning pen—at the same height as the pig's face—was designed. Videos of pigs under natural breeding conditions were collected with an industrial camera with a resolution of 1920 pixels × 1080 pixels (HD1080). Pig face images were collected between the hours of 6:00 a.m. and 8:00 a.m., 10:00 a.m. and 12:00 a.m., and 2:00 p.m. and 6:00 p.m., for a total of 46 pigs.

3.2. Data Preprocessing

Pig face images were extracted from each pig face video. The pig face images collected under unconstrained conditions contained large amounts of background noise, and information such as pig pens and windows appearing in the images could affect the effect of recognition of pig faces. Therefore, this paper used Faster RCNN to crop the pig face images out of the images, and then the pig face data were screened by manually selecting the pig-face-positive images. The images of the pig face data after screening are shown in Figure 1, and the processing results for the complete pig face dataset are shown in Table 1.

As can be seen in Table 1, to achieve open-set pig face recognition, the number of pigs was divided according to the ratio of 8:2, and the facial images of 37 pigs were randomly selected for model training. The accuracy of the open-set recognition model was tested by using the facial images of nine pigs that never appeared in the training set, and the distribution of the number of images in the pig face dataset is shown in Figure 2. Thus, the training set sample contained a total of 12,993 images, and the test set contained 3431 images. Two images from each pig in the test set were randomly selected to form image pairs as positive samples for testing pig face recognition, and 200 pairs of positive sample images were reserved for each pig; thus, a total of 1800 pairs of positive sample image pairs were generated in the test set. Two images of different pigs were randomly selected to form image pairs as negative sample pairs for testing pig face recognition, and 1800 negative sample image pairs were randomly retained to ensure that there was the same number of positive and negative image sample pairs.
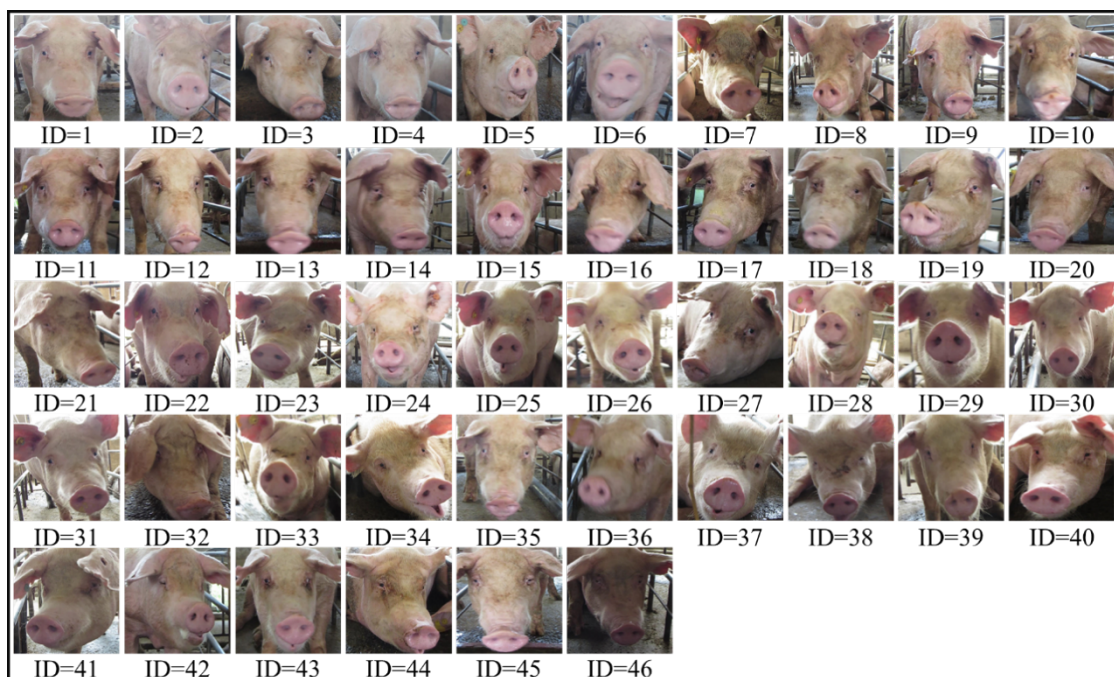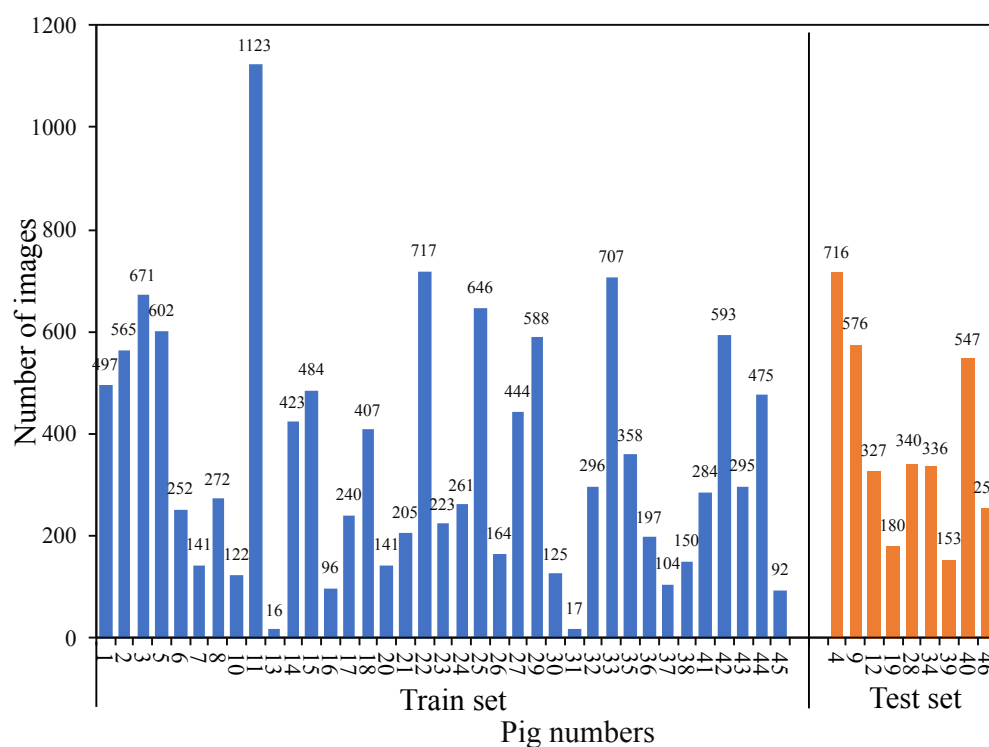

Figure 1. Pig face dataset.

Table 1. Results of the processed pig face dataset.

Number of Pigs	Number of Images	Number of Positive Sample	Number of Negative Sample	Number of Generated Image Pairs
37	12,993			
9	3431	900	900	1800

Figure 2. Image quantity distribution of the pig face dataset.

4. Methods

4.1. MobileFaceNet

Recently, lightweight networks such as MobilenetV1 and MobileNetV2 [22] have been used for visual recognition tasks in mobile terminals, but due to the specificities of facial structures, these networks have not obtained satisfactory results for facial recognition tasks. Chen et al. [23] specifically proposed a lightweight network for facial recognition—MobileFaceNet. The model used a global depth-wise convolution (GDConv) layer instead of a 7×7 global average pooling layer, which assigned weight coefficients to the importance of different positions. In addition, the PRelu activation layer is used instead of Relu, and a smaller expansion factor than that of MobileNetV2 was selected to make the model lighter. The Arcface loss function was used to increase the inter-class distance and decrease the intra-class distance during training.

4.2. Feature Extraction Backbone Network with a Normalized Attention Mechanism

Pig face images contain more levels of semantic information, and extracting the low-level and high-level semantic information of images is a key and difficult point in improving the recognition rate. An attention mechanism can weight the semantic information in an image through autonomous learning and filter the image features that are beneficial for the recognition result. In this study, a new feature extraction backbone network, which was called the ResNAM network, was made by combining the NAM [24] and a residual module to get the shallow semantic information of pig face images. The structure of the ResNAM network is shown in Figure 3. The feature extraction process is as follows: A 224 pixel $\times$ 224 pixel image of a pig face was fed into the backbone network to extract the facial features. The feature extraction backbone network consisted of a CBR module, four ResNAM modules, and a dropout layer. The CBR module included a 3×3 convolutional layer, a BN layer, and a Relu activation function, which were used to extract the underlying pig facial features. The ResNAM module incorporated two 3×3 convolutional layers and an NAM to reduce the loss of pig facial features. When the stride was 2, the original feature maps in the jump–join branch were downsampled before feature fusion.

The ResNAM module incorporated the residual structure and the NAM attention mechanism. With the design idea of CBAM attention, NAM integrated channel attention and spatial attention submodules. Figure 4a shows the channel attention submodule of the NAM. First, a weight sparsity penalty was used on the input feature map. Second, the individual channel changes reacted according to the scaling factor in the BN. Third, the feature channels that the network was interested in were highlighted, and the background information was suppressed. The scaling factor in batch normalization was used to measure the channel variance and calibrate the importance of the channel features, as shown in Equation (1):

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta, \tag{1}$$

where μ_B and σ_B^2 are the mean and standard deviation of the smallest batch, respectively, and γ and β are trainable affine transformation parameters (scale and translocation). B_{in} is the feature map output from the previous layer.
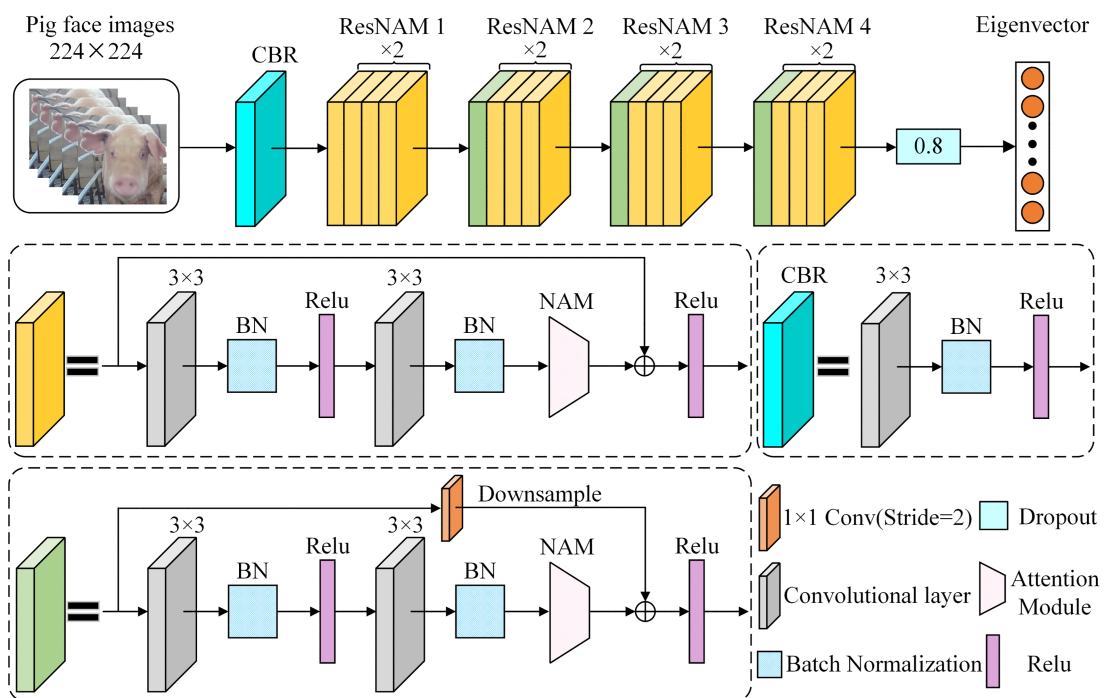

Figure 3. Structure of the backbone for feature extraction from pig face images.

The NAM channel attention submodule is shown in Equation (2):

$$\mathbf{M_c} = sigmoid(W_r(BN(\mathbf{F_1}))), \tag{2}$$

where

$$W_r = \frac{\gamma_i}{\sum_{j=0}^{\gamma_j}}, \tag{3}$$

where $\mathbf{F_1}$ is an input feature map, γ_i is a set of scale factors obtained from the input feature map through the BN layer, the weight W_r is calculated under the guidance of γ_i, and M_c is a weight factor after the processing of a *sigmoid* function.

Figure 4b shows the NAM spatial attention submodule, which is calculated as shown in Equation (5). The scaling factor of BN is applied to the spatial dimension to measure the importance of a pixel, which is called pixel normalization.

$$W_s = sigmoid(W_\lambda(BN_s(\mathbf{F_2}))), \tag{4}$$

where,

$$W_\lambda = \frac{\lambda_i}{\sum_{j=0}\lambda_j},$$ (5)

where the output is denoted as W_s, and λ is the scaling factor.

(a) Channel Attention Module
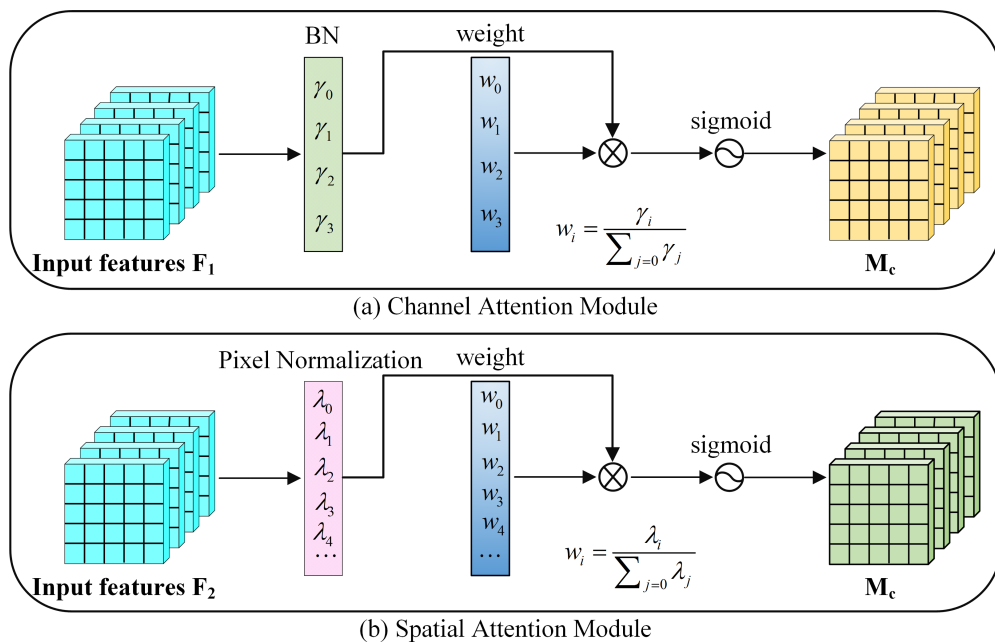

(b) Spatial Attention Module

Figure 4. Architecture of the NAM Attention Module.

4.3. An Open-Set Recognition Method for Pig Faces

Unlike other deep learning methods applied to pig faces for closed-set recognition, this study did not use the traditional softmax approach for classification learning. Instead, the last layer of the feature map was extracted and mapped into feature vectors, and metric methods were used to calculate the distance between the feature vectors to identify individual pigs. Therefore, an open-set pig face recognition method is proposed in this paper, and the specific process is shown in Figure 5.

Figure 5 shows the open-set pig face recognition method, which includes a feature extraction backbone network that incorporates a normalized attention mechanism called the ResNAM network, which was designed as described in Sections 3.1 and 3.2. In the open-set pig face recognition method, pig face images that were 224 pixels × 224 pixels in size were divided into a training set and a test set according to the process described in Section 3.2 and the test set consisted of 1800 pairs of images. The ResNAM model was trained by using the training set, the forward inference results were output after the random dropout and fully connected layers to calculate the loss values for pig face classification, and the parameters were adjusted by using stochastic gradient descent (SGD) to save the optimal model.

During model testing, the positive and negative pig face image pairs were fed into the trained ResNAM network to obtain paired feature vectors; the Euclidean distance or cosine distance was chosen as the measure for calculating the distance between the two vectors, and it was compared with the optimal threshold to determine whether the image pairs belonged to the same pig. In this case, the best threshold was calculated in the way described in the literature [14], and the test was chosen to obtain the average accuracy of the test set through ten-fold cross-validation. To further optimize the PigFaceNet model, the optimal loss function was obtained by selecting the most suitable loss function for pig face recognition from the existing loss functions—ArcFace, CosFace [25], and SphereFace [26]—and the optimal metric was found from the two measures of the Euclidean distance and cosine distance to preserve the optimal model.
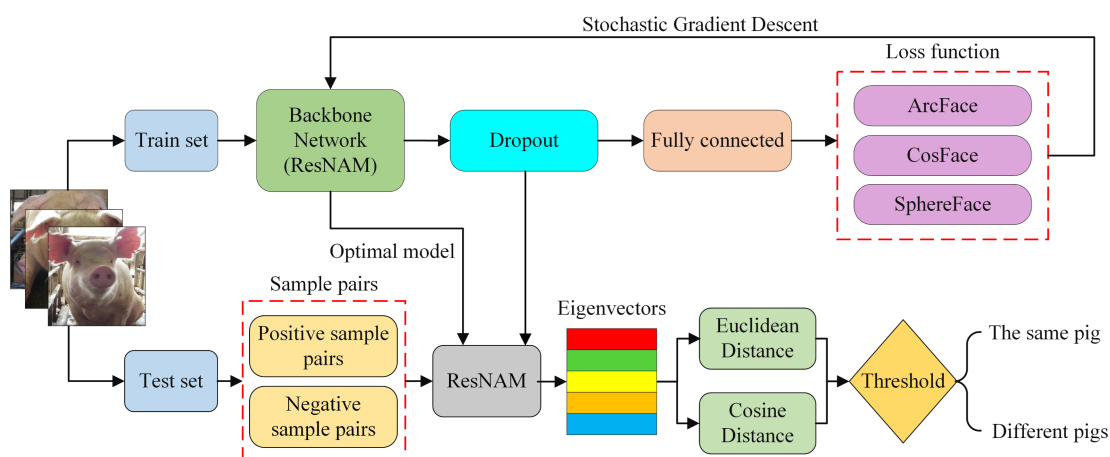
Figure 5. Open-set identification method for pig faces.

5. Experiment

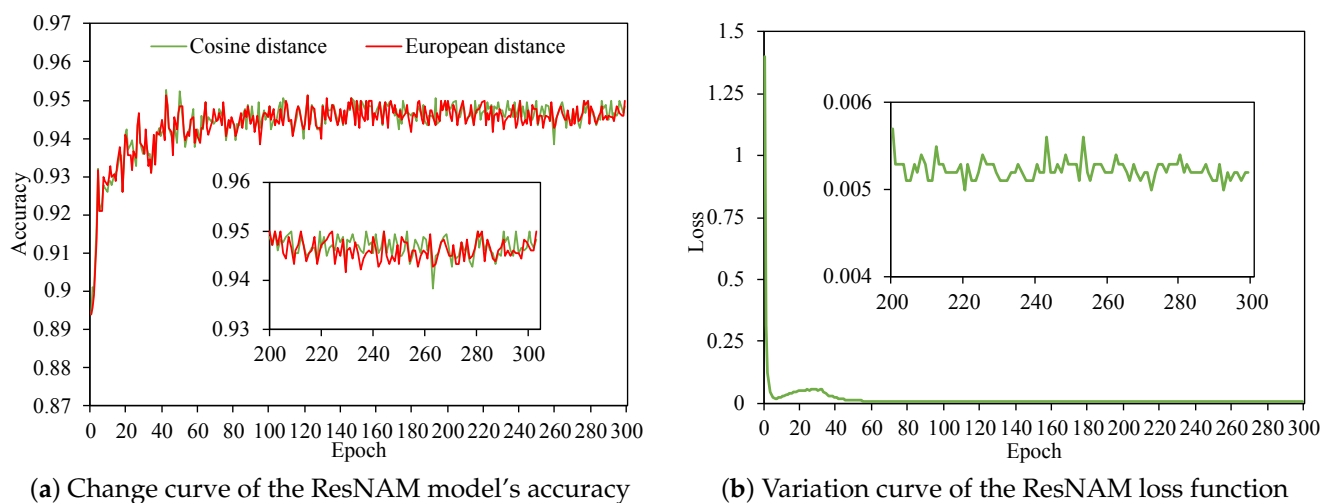### 5.1. Experimental Settings

For training, a 16 GB NVIDIA Tesla P100 graphics processor was used, and a deep learning algorithm training platform was built by using the Ubuntu 16.0 operating system, Python 3.8, and PyTorch 1.7.1. The CUDA API version was 10, and the CuDNN was 8.0.5. The training procedure had 300 training rounds, each training batch had 256 images, and the initial learning rate was set to 0.1. The learning rate was reduced to one-tenth of the original during the 5th, 60th, and 200th rounds.

5.2. Results of Training

This research used ResNet18 as the baseline model to construct a ResNAM network that incorporated the normalized attention mechanism. The ResNAM network described in this paper was trained by using the dataset shown in Table 1. The performance of our model with the cosine distance and Euclidean distance was evaluated in the test set after each epoch of training. Figure 6 depicts the changes in parameters throughout training.

Figure 6 depicts the accuracy and loss value curves when the Euclidean and cosine distances were employed during ResNAM training. Following the completion of each training epoch, the ResNAM network was used to infer test set image pairs to obtain feature vectors, and the open-set pig face recognition method from Section 4.3 was used to obtain the best threshold, test the accuracy with different measurement methods, and save the model weight with the highest accuracy. The recognition accuracy of the model displayed an overall rising trend as the number of training epochs increased, and this tended to be steady. The loss value initially dropped and then tended to stabilize. The accuracy and loss value of the model fluctuated substantially before 60 epochs; then, they progressively became stable after 60 epochs because the phased learning technique was used in this paper. The learning rate was dropped to one-tenth of the original between the fifth and 60th epochs, thus accelerating model convergence.

(**a**) Change curve of the ResNAM model's accuracy　　　(**b**) Variation curve of the ResNAM loss function

Figure 6. The change curve of each parameter during the training process.

The results of a comparison between the model in this paper and the baseline models of ResNet18 and ResNet50 are shown in Table 2. The ResNAM, ResNet18, and ResNet50 models were trained by using the same experimental environment and dataset. The test results indicated that when using the deep network as the backbone network to extract pig face features, the accuracy of the cosine-distance-based method was slightly higher than that of the Euclidean-distance-based method. The pig facial recognition architecture of this work incorporated three loss functions. Figure 7 depicts a comparison of the recognition results from various models with different loss functions. ResNet50's model size was double that of ResNAM when the loss function was ArcFace. ResNet50 had a measurement accuracy of 93.89% when the cosine distance was applied. The accuracy of ResNet50 was 93.67% when the Euclidean distance was utilized as the measurement technique. When the Euclidean distance or cosine distance was employed as the measurement method in the optimal model created by using CosFace and SphereFace as the loss functions, the accuracy of ResNAM reached 95.28% at best, exceeding those of ResNet18 and ResNet50.

When using the same feature extraction network, the model trained with SphereFace as the loss function provided the best accuracy and discrimination impact for pig facial features. To summarize, this paper combined the ResNAM network, the SphereFace loss function, and cosine distance measurement to produce the best open-set pig face recognition method, with an accuracy of 95.28%, which was significantly greater than that of the model before improvement. The experimental findings suggest that the ResNAM model developed in this paper substantially improved pig facial image feature extraction. They also demonstrate that the SphereFace loss function and cosine distance measurement could efficiently differentiate pig facial features by narrowing the intra-class gap and widening the interclass distance.

5.3. Ablation Study

In this paper, ResNAM, a ResNet18-based pig facial image feature extraction model, was built, and an open-set pig face recognition method is proposed. To produce the best model for open-set pig face recognition, the framework incorporated ResNAM, the SphereFace loss function, and cosine distance measurement. A control variable method was utilized in Section 4.2 to create several model comparison experiments that validated the efficiency of the SphereFace loss function and cosine distance measurement methods. In this part, the creation of a backbone network by merging several attention modules is described, and the same training and testing procedures as those in Section 4.2 were employed to achieve the best model and pig face recognition results. Table 3 displays the experimental outcomes.
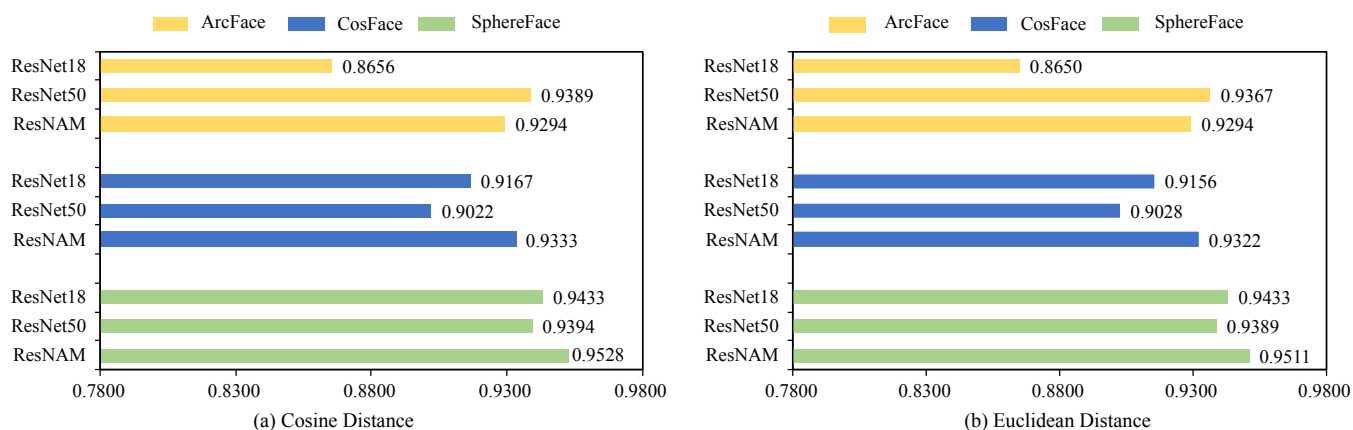
Figure 7. Structure of the backbone for pig facial image feature extraction.

Table 2. Experimental results of the base model.

| Model | Loss Function | Accuracy | | Model Weight/MB |
		Cosine Distance	Euclidean Distance	
ResNet18	ArcFace	86.56%	86.50%	42.78 MB
	CosFace	91.67%	91.56%	42.78 MB
	SphereFace	94.33%	94.33%	42.78 MB
ResNet50	ArcFace	93.89%	93.67%	90.27 MB
	CosFace	90.22%	90.28%	90.27 MB
	SphereFace	93.94%	93.89%	90.27 MB
ResNAM	ArcFace	92.94%	92.94%	42.83 MB
	CosFace	93.33%	93.22%	42.83 MB
	SphereFace	95.28%	95.11%	42.83 MB

Table 3 demonstrates how several attention processes, such as BAM [26] , CBAM [27], and NAM, were combined in the same backbone network to construct Nos. 1–3. The accuracy of the ResNAM model presented in this research was the greatest, achieving 95.28% with SphereFace as the loss function and cosine distance as the measurement technique. This model's accuracy was 0.17% and 1.34% greater than those of No. 2 and No. 3 with the identical loss function and measurement technique, respectively, and the ResNAM model was the smallest. ResNAM had the highest accuracy, achieving 93.22% when using CosFace as the loss function and the cosine distance as the measurement technique. This approach's accuracy was 1.72% and 3.33% greater than those of Nos. 2 and 3 with the identical loss function and measurement method, respectively. ResNAM had the highest accuracy, achieving 92.94% with ArcFace as the loss function and the cosine distance as the measurement technique. When compared to No. 2 and No. 3 with the identical loss function and measurement method, this strategy improved the accuracy by 2.5% and 1.05%, respectively.

To summarize, the accuracy of ResNAM was higher than that of the ResNet18 model that incorporated BAM and CBAM under the same loss function and measurement approach. When the same feature extraction network was used, the model trained using SphereFace as the loss function had the best pig face recognition performance. This demonstrated that, when compared to other loss functions, SphereFace produced features with high angular separability. It could cohere features from comparable pig face photos while also separating features from various pig face images, making it an enhancement suitable for constraining pig face features.

The ResNAM model extracted more pig face image characteristics. As a result, the best model generated by combining training with ResNAM and the SphereFace function was able to better discriminate pig facial features. The Euclidean and cosine distances had

the least impact on pig face recognition. Table 4 shows the results of the facial recognition model when it was used to identify pigs from the open-set recognition dataset for the pigs provided in this work. MobileFaceNet had a maximum accuracy of 92.67% when utilizing the Euclidean distance as the measurement technique. When the cosine distance was employed as the measurement technique, it had a maximum accuracy of 95.28%, which was 2.61% greater than that of MobileFaceNet. Due to the impact of inbreeding, there was minimal variation among pig individuals. However, there was a huge intra-class variability within the pig individuals due to the influence of light, angle, and posture, thus posing a significant obstacle to pig face recognition. The precision of applying a facial recognition model directly to a pig face was not optimal. As a result, the open-set pig face recognition technique suggested in this research successfully enhanced pig facial features, boosted pig face identification accuracy, and can be more effectively applied to pig farms.

Table 3. Ablation experiment results for different models.

Number	Model	Loss Function	Accuracy		Model Weight/MB
			Cosine Distance	Euclidean Distance	
NO.1	ResNet18+BAM	ArcFace	90.44%	90.44%	42.91 MB
		CosFace	91.61%	91.50%	42.91 MB
		SphereFace	95.11%	95.22%	42.91 MB
NO.2	ResNet18+CBAM	ArcFace	91.89%	91.83%	43.16 MB
		CosFace	90.00%	90.00%	43.16 MB
		SphereFace	93.94%	93.83%	43.16 MB
NO.3	ResNet18+NAM (Ours: ResNAM)	ArcFace	92.94%	92.94%	42.83 MB
		CosFace	93.33%	93.22%	42.83 MB
		SphereFace	95.28%	95.11%	42.83 MB

Table 4. Results of a comparison with facial recognition models.

Model	Accuracy	
	Cosine Distance	Euclidean Distance
MobileFaceNet	92.44%	92.67%
Ours	95.28%	95.11%

5.3.1. Discussion and Analysis

To eliminate identification mistakes caused by unequal sample categorization, this research employed ten-fold cross-validation to compute the average accuracy and verify the model's resilience. We divided the test set into ten parts and took nine at a time to get the optimal threshold value. Then, by using the optimal threshold value, we tested the accuracy of the tenth piece ten times and determined the average accuracy of each test set as the test set's average accuracy. The cosine distance had a value range of [0,1], while the Euclidean distance had a value range of [0,4] [28]. Table 5 displays the test results of the ten-fold cross-validation of the best model in this article.

Table 5. Pig face recognition results with different attention mechanisms.

Number	Model	Loss Function	Accuracy		Model Weight/MB
			Cosine Distance	Euclidean Distance	
NO.1	ResNet18+BAM	ArcFace	90.44%	90.44%	42.91 MB
		CosFace	91.61%	91.50%	42.91 MB
		SphereFace	95.11%	95.22%	42.91 MB
NO.2	ResNet18+CBAM	ArcFace	91.89%	91.83%	43.16 MB
		CosFace	90.00%	90.00%	43.16 MB
		SphereFace	93.94%	93.83%	43.16 MB
NO.3	ResNet18+NAM (Ours: ResNAM)	ArcFace	92.94%	92.94%	42.83 MB
		CosFace	93.33%	93.22%	42.83 MB
		SphereFace	95.28%	95.11%	42.83 MB

It was discovered via ten-fold cross-validation that different test sets had different ideal thresholds for image capture, and the different optimal thresholds had varied test outcomes in each test set, with a maximum recognition rate of 97.778%. The best threshold value assessed with the cosine distance for pig face images was 0.745, while the best threshold value measured with the Euclidean distance was 0.510. The average accuracy of the test set was determined by taking the average accuracy of 10 test sets. The model's accuracy measured with the cosine distance was 95.278%, while the model's accuracy measured with the Euclidean distance was 95.111%. The testing findings revealed that the cosine distance was more suitable for determining the distances between pig face photos. Because the pig facial features recovered from the best ResNAM model with SphereFace as the loss function differentiated various pig individuals well, the accuracy of using the cosine distance as the measurement technique was only marginally greater than that of using the Euclidean distance. According to the results in Table 5, the ideal threshold value was 0.745, and the accuracy of the test set was the highest, reaching 95.278% when the model utilized the cosine distance as the measurement method. As a result, the cosine threshold was adjusted to 0.745, and the accuracy was computed for each pig in the test set, as shown in Figure 8.

Figure 8 shows that the accuracy of the pig face recognition model that utilized the cosine distance as the measurement method to identify Pig4 was 100%, whereas Pig39 and Pig40 had poor recognition rates. Recognition mistakes occurred when Pig39 was combined with some images of Pig34, Pig46, Pig39, and Pig40 to produce an image pair, and recognition errors occurred when Pig40 was paired with some photographs of Pig34, Pig46, Pig39, and Pig40 to make an image pair. The number of negative samples with identification errors in the test sample pair of Pig39 was greater than the number of positive samples with identification errors, indicating that the distinction between classes for Pig39 was modest, and it was easily confused with other pigs.
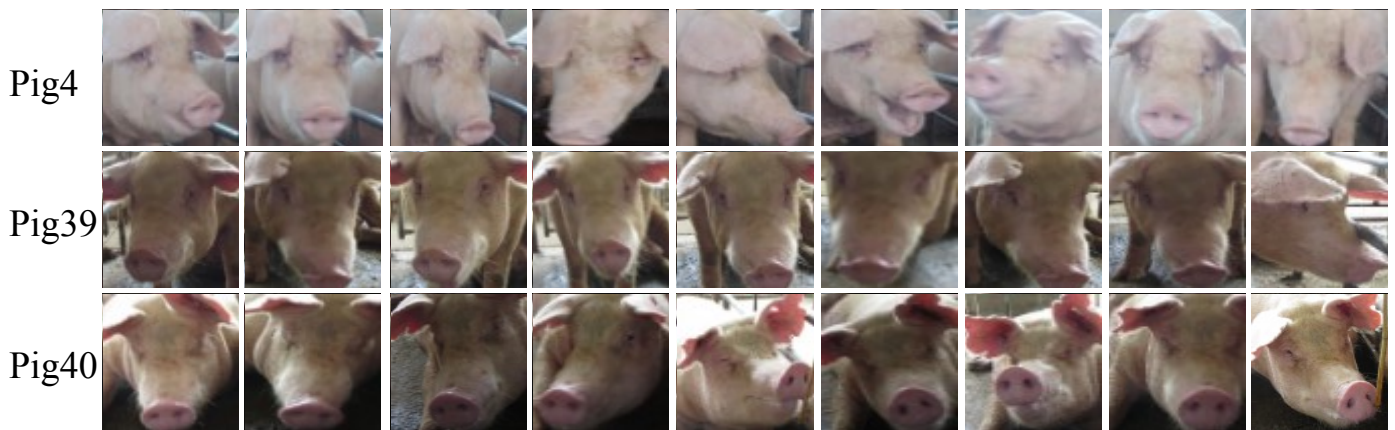
Negative sample pairs with identification errors accounted for 8.8% of the overall number of negative samples in the Pig40 test samples, while positive sample pairings with identification errors accounted for 11% of the total number of positive samples. Positive samples had a higher mistake rate than that of negative samples. As a result, big intraclass variances had a significant impact on the accuracy for Pig40. Figure 9 depicts various sample pairs with incorrect identification. Figure 9a shows that the test samples of Pig4, Pig39, and Pig40 had various angle and ear occlusion issues. Pig39 and Pig40 had much greater pig face angles and light effects than Pig4. However, the recognition rate for Pig4 was 100%, and those for Pig39 and Pig40 were 89.01% and 90.64%.
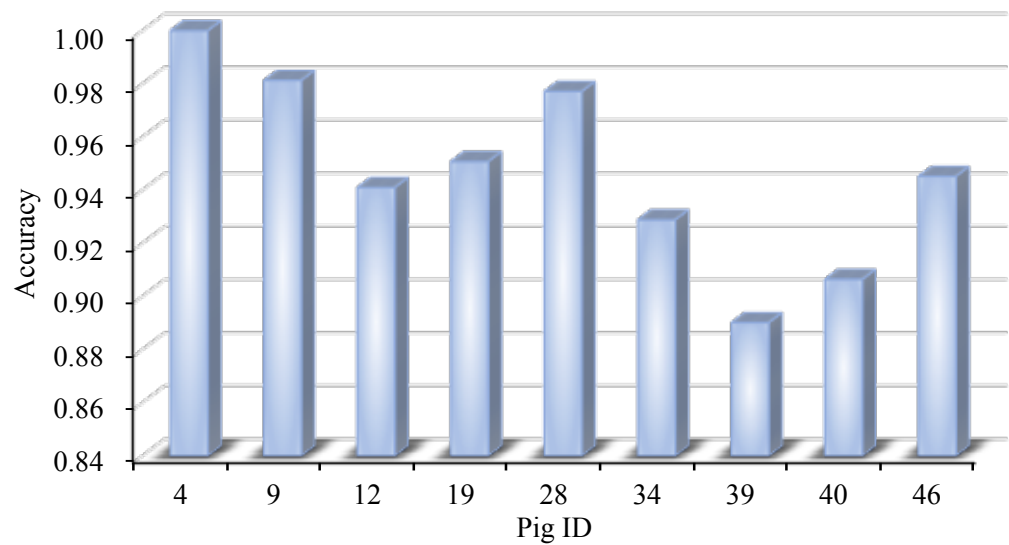
Figure 8. Accuracy at a fixed threshold.

On the one hand, this demonstrates that the pigs' ears, posture, angle, and illumination conditions, which are typical characteristics in pig face recognition, caused a high intraclass difference and a modest interclass difference. On the other hand, this demonstrated that the strategy in this research minimized the intraclass difference to some extent, which presents suggestions for boosting pig face recognition accuracy. The positive and negative examples of recognition mistakes, as shown in Figure 9b, had a strong similarity between the face images of Pig39 and those of Pig34 and Pig46, which were difficult for human eyes to identify, and the difference between classes was modest. However, there were many light and angle interferences in pig image 40, resulting in a large intra-class gap and low recognition rate for positive sample pairs. The method in this paper improved the accuracy of pig face recognition to a certain extent, but it has not been completely solved. In practical applications, open-set recognition can be used to compare new pig face images with the pig face images in a database one by one to determine whether the unknown pig has ever appeared in the database.

As a result, we can try to add pig face photographs with various perspectives, lighting, and attitudes to the database in order to increase the accuracy of pig face identification. Closed-set recognition, on the other hand, assumes that the pig to be identified is a pig in the database, and it cannot be used to identify pigs that have not been in the database. This problem cannot be solved by adding wealthy pig face photos. To summarize, the approach in this study increased the accuracy of pig face identification, addressed the problem of huge gaps within a class, overcame the interference of external factors to some extent, and serves as a reference for future pig face recognition research (Tables 6 and 7).

(a) Sample images of Pig4, 39, 40

(b) Examples of misidentified sample pairs

Figure 9. Examples of positive and negative sample pairs that identified errors.

Table 6. Results of the ten-fold cross-validation test of the optimal model.

Test	Cosine Distance		Euclidean Distance	
	The Best Threshold	Accuracy	The Best Threshold	Accuracy
NO.1	0.745	96.667%	0.510	96.667%
NO.2	0.745	95.556%	0.510	95.556%
NO.3	0.745	94.444%	0.510	94.444%
NO.4	0.745	93.333%	0.500	91.667%
NO.5	0.745	95.556%	0.510	95.556%
NO.6	0.745	95.000%	0.510	95.000%
NO.7	0.745	94.444%	0.510	94.444%
NO.8	0.745	97.222%	0.510	97.222%
NO.9	0.745	92.778%	0.510	92.778%
NO.10	0.745	97.778%	0.510	97.778%
average accuracy	——	95.278%	——	95.111%

Table 7. Test results at different thresholds.

Data	Cosine Distance		Euclidean Distance	
	The Best Threshold	Accuracy	The Best Threshold	Accuracy
test set	0.745	95.278%	0.500 0.510	95.111% 95.278%

5.3.2. Comparison with Existing Studies

Table 8 shows the results of this study compared with those of other studies. The current research on individual recognition of livestock mainly contains both closed-set recognition and open-set recognition. The research on pig face recognition is focused on closed-set recognition. CNN backbone networks, such as LeNet-5, AlexNet, ResNet50, VGG, and DenseNet 121, have been used to extract image features and to achieve high accuracy. However, closed-set recognition can only recognize the identities of livestock that have appeared in the training set, so the recognition is less difficult and the recognition accuracy is higher. Research on open-set recognition has mainly focused on cow face recognition and cow pattern recognition, and less research has been done on open-set recognition of pig faces. In a test set containing seven cows, the accuracy of the model proposed by Andrew et al. reached 93.8%. Xu et al. proposed the CattleFaceNet model, which fused the Arcface function, metric learning, and a backbone network, and the accuracy of open-set recognition of cow faces reached 91.35%. Based on previous studies, a ResNAM backbone network combining the NAM attention mechanism with ResNet was proposed in this paper to extract image features. Then, an open-set pig face recognition framework integrating three loss functions and two metrics was proposed. The open-set pig face recognition model proposed in this paper finally achieved an accuracy of 95.28%, which was higher than those of the existing open-set recognition models and proved the effectiveness of the improved method proposed in this paper.

Table 8. Results of a comparison of individual recognition methods for livestock.

Method	Studies	Year	Species	Objects	Backbone	Accuracy
Closed-set recognition	Hansen et al. [15]	2018	pig	10 pigs	VGG	96.7%
	Marsot et al. [16]	2020	pig	10 pigs	Two Haar feature+ CNN	83.0%
	Salama et al. [8]	2019	sheep	52 sheeps	AlexNet	98.0%
	Wang et al. [17]	2020	pig	10 pigs	LeNet-5	97.6%
	Wang et al. [29]	2021	pig	46 pigs	ResNet50	97.66%
	Wang et al. [18]	2022	pig	28 pigs	DenseNet 121	94.04%
Open-set recognition	Andrew et al. [14]	2021	cow	46 cows	CNN	93.8%
	Xu et al. [20]	2022	cow	90 cows	CattleFaceNet	91.35%
	Ours (ResNAM)	2022	pig	46 pigs	ResNAM	95.28%

6. Conclusions

Firstly, to extract features from pig face images, the ResNAM backbone network was presented, which integrated a normalized attention mechanism and residual network and could fully extract key features from pig face images when they were disturbed by noise, occlusion, and other situations.

Secondly, when compared to the BAM and CBAM attention modules, NAM improved the performance of the pig facial recognition model more effectively and could extract richer high-level semantic data. The accuracy of pig face recognition when using NAM was higher than that when using BAM and CBAM with the identical loss function and measurement method.

Thirdly, an open-set pig face recognition framework was provided in this study, which integrated three loss functions and two measurement methods and accomplished open-set pig face recognition with non-overlapping individuals in the training and test

sets. ResNAM's accuracy was 95.28% with SphereFace as the loss function and the cosine distance as the measurement technique, which was 2.61% greater than that of the facial recognition model.

To summarize, deep learning may be used to perform open-set pig face recognition. The issue of only identifying pig individuals that have appeared in the training set was overcome through open-set pig face recognition. This paper proposed a corresponding open-set pig face recognition framework based on metric learning, and it made corresponding improvements in the backbone network used for feature extraction in the framework, which improved the accuracy of open-set pig face recognition and provided a new idea for future research on pig face recognition algorithms. In addition, in future work, we will try to combine open-set recognition with deep unsupervised active learning in order to improve the quality of learning and render it more semantic.

Author Contributions: Conceptualization, methodology, software, validation, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, all authors. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are presented in this article in the form of figures and tables.

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

References

1. Adrion, F.; Kapun, A.; Eckert, F.; Holland, E.M.; Staiger, M.; Götz, S.; Gallmann, E. Monitoring trough visits of growing-finishing pigs with UHF-RFID. *Comput. Electron. Agric.* **2018**, *144*, 144–153. [CrossRef]
2. Kashiha, M.; Bahr, C.; Ott, S.; Moons, C.P.H.; Niewold, T.A.; Ödberg, F.O.; Berckmans, D. Automatic identification of marked pigs in a pen using image pattern recognition. *Comput. Electron. Agric.* **2013**, *93*, 111–120. [CrossRef]
3. Zhao, K.; Jin, X.; Ji, J.; Wang, J.; Ma, H.; Zhu, X. Individual identification of Holstein dairy cows based on detecting and matching feature points in body images. *Biosyst. Eng.* **2019**, *181*, 128–139. [CrossRef]
4. Jiang, B.; Wu, Q.; Yin, X.; Wu, D.; Song, H.; He, D. FLYOLOv3 deep learning for key parts of dairy cow body detection. *Comput. Electron. Agric.* **2019**, *166*, 104982. [CrossRef]
5. Li, S.; Fu, L.; Sun, Y.; Mu, Y.; Chen, L.; Li, J.; Gong, H. Individual dairy cow identification based on lightweight convolutional neural network. *PLoS ONE* **2021**, *16*, e0260510. Publisher: Public Library of Science. [CrossRef] [PubMed]
6. Wu, D.; Wang, Y.; Han, M.; Song, L.; Shang, Y.; Zhang, X.; Song, H. Using a CNN-LSTM for basic behaviors detection of a single dairy cow in a complex environment. *Comput. Electron. Agric.* **2021**, *182*, 106016. [CrossRef]
7. Kumar, S.; Pandey, A.; Sai Ram Satwik, K.; 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]
8. Salama, A.; Hassanien, A.E.; Fahmy, A. Sheep Identification Using a Hybrid Deep Learning and Bayesian Optimization Approach. *IEEE Access* **2019**, *7*, 31681–31687. [CrossRef]
9. Hu, H.; Dai, B.; Shen, W.; Wei, X.; Sun, J.; Li, R.; Zhang, Y. Cow identification based on fusion of deep parts features. *Biosyst. Eng.* **2020**, *192*, 245–256. [CrossRef]
10. Shen, W.; Hu, H.; Dai, B.; Wei, X.; Sun, J.; Jiang, L.; Sun, Y. Individual identification of dairy cows based on convolutional neural networks. *Multimed. Tools Appl.* **2020**, *79*, 14711–14724. [CrossRef]
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the Advances in Neural Information Processing Systems*; Pereira, F., Burges, C.J., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Nice, France, 2012; Volume 25.
12. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
14. Andrew, W.; Gao, J.; Mullan, S.; Campbell, N.; Dowsey, A.W.; Burghardt, T. Visual identification of individual Holstein-Friesian cattle via deep metric learning. *Comput. Electron. Agric.* **2021**, *185*, 106133. [CrossRef]
15. Hansen, M.F.; Smith, M.L.; Smith, L.N.; Salter, M.G.; Baxter, E.M.; Farish, M.; Grieve, B. Towards on-farm pig face recognition using convolutional neural networks. *Comput. Ind.* **2018**, *98*, 145–152. [CrossRef]

16. Marsot, M.; Mei, J.; Shan, X.; Ye, L.; Feng, P.; Yan, X.; Li, C.; Zhao, Y. An adaptive pig face recognition approach using Convolutional Neural Networks. *Comput. Electron. Agric.* **2020**, *173*, 105386. [CrossRef]

17. Wang, K.; Chen, C.; He, Y. Research on pig face recognition model based on keras convolutional neural network. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2020; p. 032030.

18. Wang, Z.; Liu, T. Two-stage method based on triplet margin loss for pig face recognition. *Comput. Electron. Agric.* **2022**, *194*, 106737. [CrossRef]

19. Gao, J.; Burghardt, T.; Andrew, W.; Dowsey, A.W.; Campbell, N.W. Towards Self-Supervision for Video Identification of Individual Holstein-Friesian Cattle: The Cows2021 Dataset. *arXiv* **2021**, arXiv:2105.01938,

20. Xu, B.; Wang, W.; Guo, L.; Chen, G.; Li, Y.; Cao, Z.; Wu, S. CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss. *Comput. Electron. Agric.* **2022**, *193*, 106675. [CrossRef]

21. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. SphereFace: Deep Hypersphere Embedding for Face Recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6738–6746. [CrossRef]

22. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [CrossRef]

23. Deng, J.; Guo, J.; Yang, J.; Xue, N.; Kotsia, I.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 5962–5979. [CrossRef] [PubMed]

24. Liu, Y.; Shao, Z.; Teng, Y.; Hoffmann, N. NAM: Normalization-based Attention Module. *arXiv* **2021**, arXiv:2111.12419. https://doi.org/10.48550/arXiv.2111.12419.

25. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5265–5274. [CrossRef]

26. Park, J.; Woo, S.; Lee, J.Y.; Kweon, I.S. BAM: Bottleneck Attention Module. In Proceedings of the British Machine Vision Conference, Newcastle, UK, 3–6 September 2018.

27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Proceedings of the Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–19. _1. [CrossRef]

28. Chen, S.; Liu, Y.; Gao, X.; Han, Z. MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices. In *Proceedings of the Biometric Recognition*; Springer International Publishing: Cham, Switzerland, 2018; Lecture Notes in Computer Science, pp. 428–438. [CrossRef]

29. Wang, R.; Shi, Z.; Li, Q.; Gao, R.; Zhao, C.; Feng, L. Pig Face Recognition Model Based on a Cascaded Network. *Appl. Eng. Agric.* **2021**, *37*, 879–890. [CrossRef]