

## PIG FACE IDENTIFICATION BASED ON IMPROVED ALEXNET MODEL

## / 基于改进 AlexNet 模型的生猪脸部识别

Hongwen Yan, Qingliang Cui\*, Zhenyu Liu<sup>1</sup>

College of Information Science and Engineering, Shanxi Agricultural University, Taigu / China

Tel: +86-0354-6289253; E-mail: qlcui@126.com

DOI: <https://doi.org/10.35633/inmateh-61-11>**Keywords:** AlexNet, Spatial attention, Pig face identification, Batch size**ABSTRACT**

Individual pig identification technology is the precondition of precise breeding. Taking pig face as the study point, this article puts forward a pig face identification method based on improved AlexNet model and explores the influence of training batch size on the performance of the model. Spatial attention module (SAM) is introduced in AlexNet model to compare the performance of the AlexNet model and the improved model on the training set and the validation set. The study shows that the improved AlexNet model can achieve higher precision rate under different training batch sizes and has higher convergence rate and robustness, with an identification precision rate reaching 98.11%, and a recall rate and f1 value reaching 98.03% and 98.05%. When the training batch sizes are 16, 32, and 64 respectively, the test time of the model, which represents its operating efficiency, improves by 1.99%, 2.36% and 10.31%, respectively, showing better performance in pig face identification. The test results show that different batch sizes have a certain influence on the prediction results of the model, while no fixed relationship.

**摘要**

生猪个体识别技术是实现精准养殖的前提，本文以猪脸为研究点提出基于改进 AlexNet 模型的猪脸识别方法并探究训练批大小对模型性能影响，在 AlexNet 模型中引入空间注意力模块（SAM），比较 AlexNet 与其改进模型在训练集及验证集上的性能，研究表明，改进的 AlexNet 模型在不同训练批大小情况下均可取得较高准确率，具有更快的收敛速度与鲁棒性，识别准确率达到 98.11%，召回率与 f1 值分别达到 98.03%、98.05%，其在训练批大小分别为 16、32、64 情况下，表征其运行效率的模型测试时间分别提高了 1.99%、2.36%、10.31%，可对生猪脸部进行更有效识别，试验结果表明，不同的批大小对模型预测结果有一定影响，不存在固定关系。

**INTRODUCTION**

Effective individual pig identification is the precondition of intelligent pig breeding. Pig face includes pig's eyes, pig's nose and other biological characteristics that are of natural identifiability. Based on these characteristics, pigs can be individually identified. The face identification technology has been widely used in the field of individual identification for its advantages such as non-invasiveness, low cost and operability.

Radio-frequency identification (RFID) is a widely-used technology in the field of individual pig identification (Maselyne et al., 2014; Hahnel et al., 2016). With its shortcomings becoming increasingly prominent, researchers turned to the machine learning method to conduct contour extraction and behavioural detection to study the pigs, laying emphasis on the aspects such as contour extraction (Ma et al., 2016; Guo et al., 2015, Li et al., 2017), climbing and attacking behaviours (Chen et al., 2017), standing and lying posture (Kim et al., 2017), and behaviour tracking (Peter et al., 2011).

Recently, some scholars have used deep learning technology to conduct image segmentation and individual target detection to pigs (Psota et al., 2019; Zhang et al., 2018, Ju et al., 2018).

However, as for individual pig identification based on convolution network, only Wang et al. (2018) effectively fused the feature integration method with features extracted from deep convolutional neural network such as DPN131 (Chen et al., 2017), InceptionV3 (Szegedy et al., 2016) and Xception (Chollet, 2017) based on transfer learning method, achieving an identification precision rate of 96.41%.

<sup>1</sup> Hongwen Yan, As. Ph.D. Lect. Eng.; Qingliang Cui\*, Prof. Ph.D. Eng.; Zhenyu Liu, Prof. Ph.D. Eng.

Hansen *et al.* (2018) combined the feature extraction results of VGG-Face model (Parkhi *et al.*, 2015) and used Grad-CAM (Ramprasaath *et al.*, 2016) activation-resembling mapping to distinguish the adhered pigs, achieving an identification precision rate of 96.7%. There is a lack of open data set in the studies of this direction, and no complete research direction has been formed yet.

Based on these, this paper used the principle of convolutional neural network to build and improve the AlexNet model, so as to explore the impact of batch size on performance of the model, and have individual identification of the pigs with pig face as the study point to provide reference for accurate breeding of pigs.

## MATERIALS AND METHODS

### Sample collection

As shown in Figure1, the experimental materials of this study were collected on a small farm in Wujiazhuang, Taigu County, Shanxi Province, China (112°53' E, 37°42' N), and the sampling date was in March 2018.

Pictures of a total of 10 pigs were collected, including 684 training samples, 77 validation samples and 254 test samples.



Fig. 1 - Pig Samples

To avoid memory overflow, batch training method was taken to compare the AlexNet model and its improved model on the training set and validation set. The training batch size was set to be 16, 32 and 64 respectively, and the batch size of validation set and that of training set were synchronous. The iteration round number was set to be 100. After each round of iteration, the model evaluation index values were calculated on the test set.

### Structure of base model

With an accuracy rate of 57.1% and a top-5 identification rate of 80.2%, AlexNet (Krizhevsky *et al.*, 2012) proposed by Alex Krizhevsky won the first prize in ImageNet competition in 2012. The structure of AlexNet model is as shown in Figure 2.

The model contained four layers, and each level was composed of different amounts of convolution layer and pooling layer. A total of 9 weight layers were there, among which 5 convolution layers and 4 fully connected layers were included.

AlexNet jointed the LRN layer after the 1st and 2nd convolution layer so as to do local normalization processing to the area near the activated neurons, but follow-up studies indicated that after joining LRN layer, the effect got reduced rather than increased.

Therefore, the LRN layer was removed in this experiment, the model's convolution and pooling layer feature map uses an effective filling mode, and the final convolution layer at each level was connected to the maximum pooling layer, and the each weight layer used ReLU activation function to nonlinearize linear operations after completion of the corresponding operation.

Since the fully connected layer contained many parameters, dropout mechanism was introduced in the fully connected layer to prevent model overfitting.

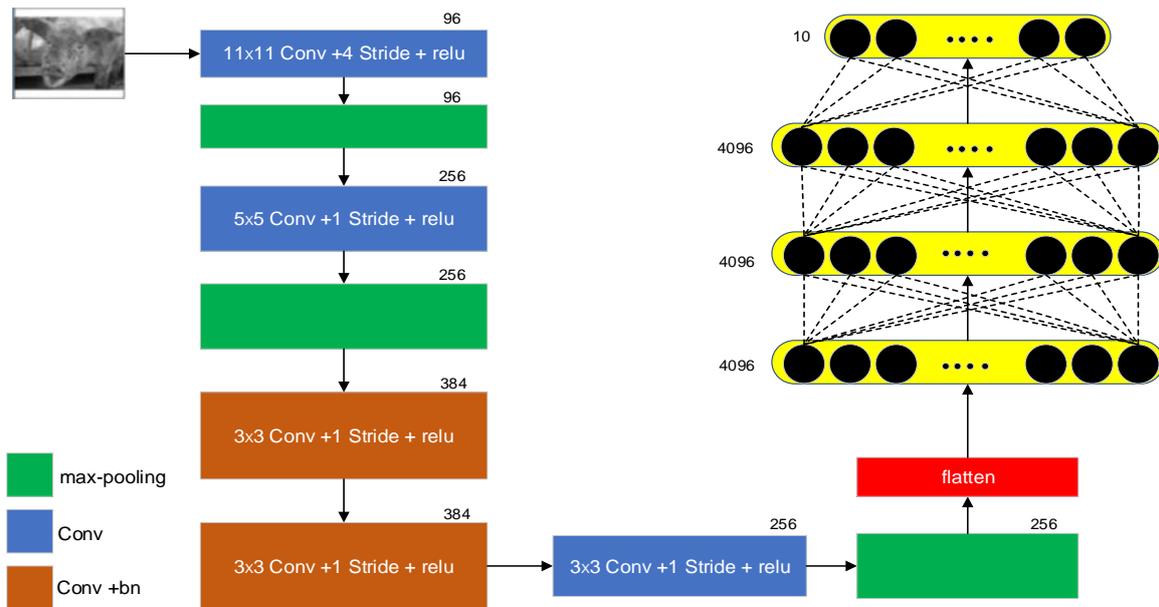


Fig. 2 - Structure of AlexNet Model

Note: max-pooling in the figure refers to the maximum pooling operation, conv refers to the conventional convolution operation, conv+bn refers to jointing normalized operation in batch behind at the end of the convolution operation is used. The Inside of each convolution module is represented in the form of 11x11 conv+4 stride+relu, in which 11x11 means both the width and the height of the convolution kernel are 11, while 4 means the step size of convolution operation, relu means that the linear result after convolution operation is nonlinearized by using relu activation function. The numbers 4,096 and 10 on the upper right indicate the number of nodes in the fully connected layer.

**SAM module, the improved part of the model**

Deep web contains rich semantic information and can guide surface web in choice of information, so it can get more accurate resolution ratio information. SAM module combined with soft attention mechanism can capture rich information associated to context, also it can give different weight values to different positions on a feature map to strengthen the effective features. As shown in Figure 3, its specific operation went through the following three stages.

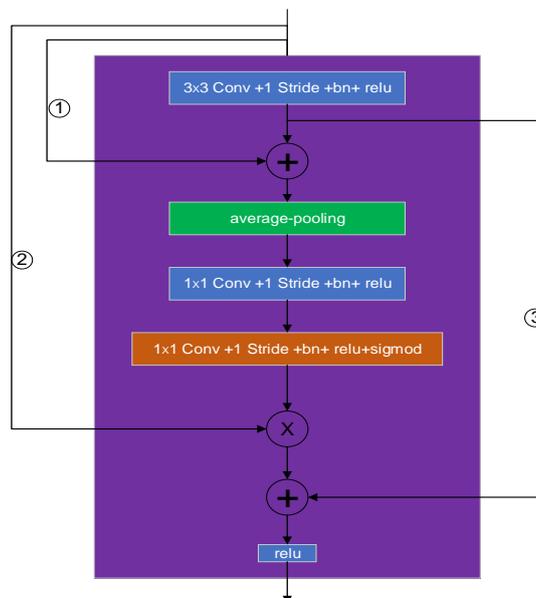


Fig. 3 - Spatial Attention Module (SAM)

Note: ① shows the residual of the original feature, ② shows the product between the original feature and the weighted feature map, which may recalibrate the original feature, ③ shows the original feature after one 3\*3 convolution, bn processing as well as relu nonlinear processing.

(1) The surface web input went through 1 operation of convolution with a size of  $3 \times 3$  and a step size of 1, batch normalization and activation function before the output of the middle layer was obtained. The surface layer input and the middle layer output were superimposed in operation 1 to be the input of step (2);

(2) After the global average pooling operation of output feature map in (1), the values at the same position of different channels were added and averaged to obtain the feature map with channel dimension of 1 and resolution size that was consistent with input to be the weight map of input feature map. In order to nonlinearize the global average pooling operation, the operations of two steps of  $1 \times 1$  convolution, batch normalization and activation function were introduced. In step 1, the number of convolution kernels took any value. In the test, the values taken were the same as the number of channels in the input feature map in step (2), and ReLU was used as the activation function. In step 2, for the  $1 \times 1$  convolution operation, Sigmoid activation function was selected to generate different regional characteristics. The number of convolution kernels must be the same as the number of channels of the input feature map in step (2) so as to facilitate the multiplication of subsequent feature maps. The final output result was taken as the output weight information of the middle layer, and the weighted surface layer input information was obtained by multiplying the weight information with the surface input via operation 2;

(3) Finally, via operation 3, the output of the middle layer and the result of step (2) were superimposed to get the output of SAM module, which was used as the input of subsequent operations.

### Improved AlexNet module

Inspired by the application of attention mechanism in natural language processing and image segmentation field, SAM was introduced in the AlexNet model in the test to form the Attention-AlexNet model, so as to further improve the predictive performance of the model, whose structure is as shown in Figure 4.

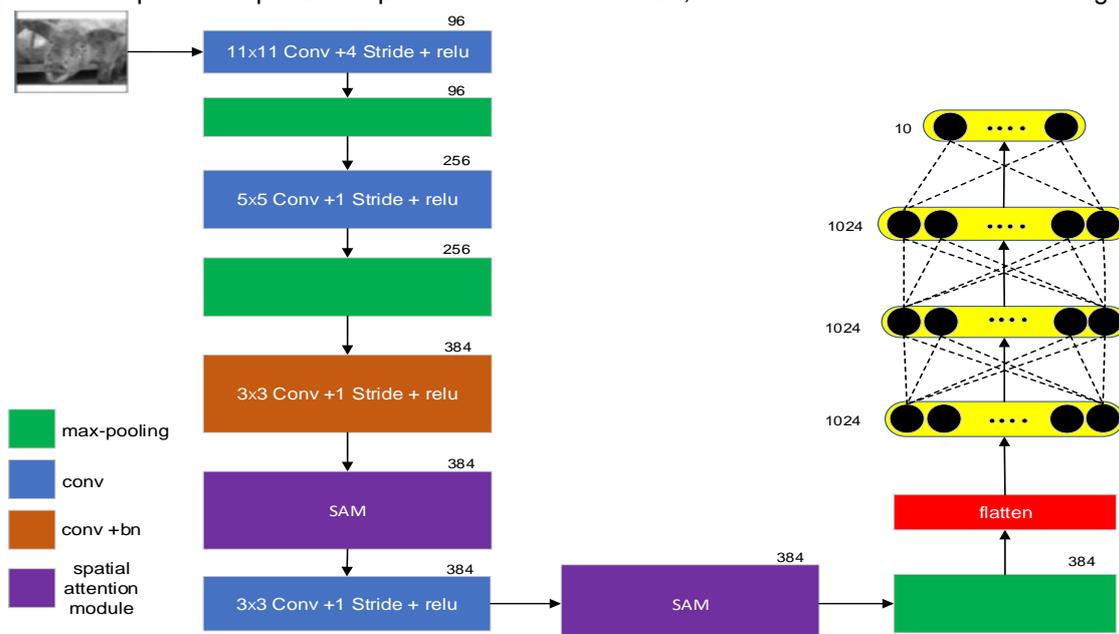


Fig. 4 - Structure of Attention-AlexNet Module

In Attention-AlexNet model, the structure of two levels in the front was the same as that of AlexNet model. After the front two-level operation, the input image was converted into 256 feature maps of size of  $26 \times 26$ . The Attention-AlexNet model contained 1 maximum pooling, 2 basic convolution operations and 2 SAM modules in the third level, where the second-level output was first made, a maximum pooling operation whose pooling kernel size was 3 and pooling step size was 2 before 256 feature maps of size of  $12 \times 12$  were obtained. Then, 384 conventional convolutions whose convolution kernel size was  $3 \times 3$  and step size was 1 were jointed, thus the feature map was converted to be 384 feature maps of size of  $10 \times 10$  to be the first SAM module input. After SAM processing, the feature map was consistent with the input in size and number. After the first SAM output, 384 conventional convolutions whose size was  $3 \times 3$  and step size was 1 were jointed to get 384 feature maps of size of  $8 \times 8$ .

In order to further strengthen the obviously characteristic area, SAM module was jointed again after the operation.

After the fourth layer, the flattening operation was jointed and three fully connected layers on which the number of neuron nodes was 1024, 1024 and 10 respectively, were jointed behind so as to greatly reduce the training parameters and improve the training speed.

## Tests and results analysis

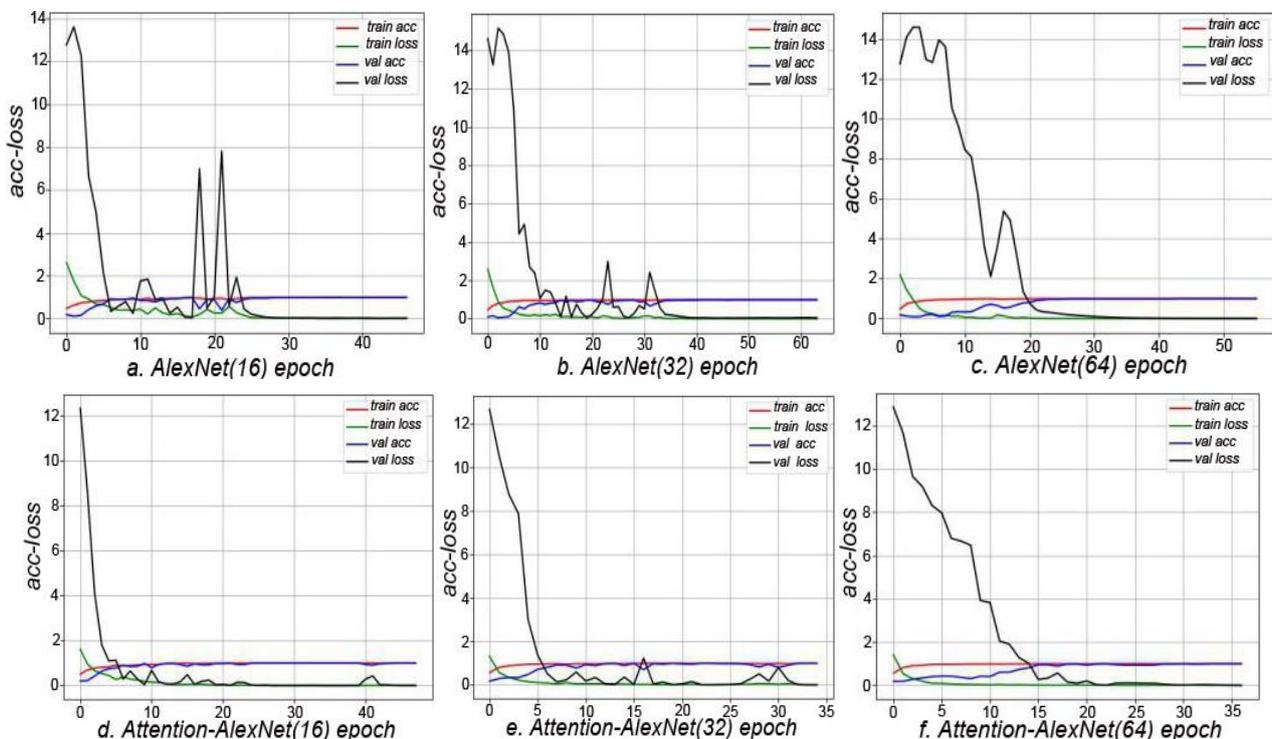
### Test parameter setting

In this study, python v3.5 language was used to construct the convolution network AlexNet and Attention-AlexNet models under the Keras framework with a video memory capacity of 6G. The construction was carried out in the GPU environment where the graphics card chipset was GeForce GTX TITAN, and the system adopted CentOS7.0.

Categorical\_crossentropy in Keras was adopted as cost function; dropout mechanism was introduced to restrain neuron nodes from participating in the process of back propagation at probability of 0.4; EarlyStopping mechanism was used to prevent overfitting; Adam optimization model was adopted and the initial value of learning rate parameter was set to be 0.0001; ModelCheckpoint mechanism was used, thus it was able to judge whether the precision rate got improved when a training was completed. If improved, the current parameters shall be saved, otherwise the next round of training shall be carried out. To automatically modify the learning rate, ReduceLROnPlateau was introduced, and it was set that the learning rate would be reduced to 0.9 times the current learning rate when the loss function value of the validation set did not reduce in the 5 rounds of iteration process. The optimal parameters and structure could be obtained at the end of the training, and the pig face identification test could be carried out by using the training results.

## RESULTS AND DISCUSSION

Figure 5 shows the relation curve between precision rate and loss functions of AlexNet model and Attention-AlexNet model under training set and validation set of different batch sizes, where Figure (5a)~Figure (5c) represents the performance of AlexNet model under batch sizes of 16, 32 and 64 respectively, and Figure (5d) ~ Figure (5f) represents the performance of Attention-AlexNet model under batch sizes of 16, 32 and 64 respectively.



**Fig. 5 - Iteration curve of AlexNet, Attention-AlexNet modules under different batch sizes**

Note: 'acc-loss' represents the accuracy rate and loss function curve, 'train acc' represents the accuracy rate of the model on the training set, 'val acc' represents the accuracy rate of the model on the validation set, 'train loss' represents the loss rate of the model on the training set, and 'val loss' represents the loss rate of the model on the validation set, 'val' represents validation set and 'train' represents the training set.

As can be seen in Figure (5a), Figure (5c), for AlexNet model, in early iterations, the precision rate of training sets was all relatively low at the beginning but well above the validation precision rate. With the iteration going on, the precision rate of training set quickly reached the convergence condition. In Figure (5a), the 5th round of iteration tended to be smooth and stable; and in Figure (5b), the 4th round tended to be stable; in Figure (5c), the convergence rate of AlexNet became faster, and in the third round it had tended to balance. For the precision rate of validation set, the iteration converged after a certain number of rounds. In the early stage of the convergence, fluctuations were there, but they tended to be stable on the whole, and the stationary value finally held the line with the precision rate of training set, which showed that the AlexNet model applied to the data set of pig face in this test. For loss function, the loss function values of the validation set were far higher than those of training set at the beginning of the training iteration, and in the iteration process, drastic fluctuation appeared. The greater the fluctuation was, the more obvious the fluctuation of precision rate on the corresponding validation set was. With the iteration going on, when the batch sizes were 16, 32 and 64 respectively, convergence took place on the 25th, 35th and 30th rounds. However, for loss function values in the training set, in the whole iteration process, under different batch sizes, excessive fluctuation did appear, and the trend had been in the direction of a decrease in its value. It can be seen from the relationship between precision rate and the value of the loss function that in the iteration rounds where the precision rate changed significantly, the value of loss function had significant change correspondingly and the two went in opposite directions.

As can be seen in Figure (5d)~Figure (5f), for Attention-AlexNet model, the precision rate of training set and the precision rate of validation set were increasing in the whole iteration process, and compared with AlexNet model, it converged more quickly. In terms of value change of loss function, in three kinds of cases, its value changed greatly in some parts in the iterative process, though compared with the AlexNet model, its fluctuation amplitude was relatively small, showing that the Attention-AlexNet model could achieve convergence condition much faster and it owned a small fluctuation amplitude.

### **Model efficiency analysis**

In the performance evaluation of deep learning models, recall ratio, precision ratio and f1-score measurement indexes were adopted. In the classification task, the results were usually divided into true positive cases (TP), false positive cases (FP), true negative cases (TN) and false negative cases (FN). With the number of samples corresponding to TP, FP, TN and FN given, the precision ratio was defined as:

$$precision = \frac{TP}{TP + FP} \quad (1)$$

Recall ratio was defined as

$$recall = \frac{TP}{TP + FN} \quad (2)$$

Recall ratio was also called recall rate. Recall ratio and precision ratio changed in opposite trend. f1-score can measure the different preferences of these two indexes, and the formula was as follows:

$$f1-score = 2 \times \frac{precision \times recall}{precision + recall} \quad (3)$$

where: *TP* represents the number of positive samples that are actually positive samples, [a];  
*FP* represents the number of positive samples that are actually negative samples, [a];  
*FN* represents the number of negative samples that are actually positive samples, [a].

The model was tested on the test set. In order to effectively measure the operating efficiency of different models, the training and test time was counted, and the results were as shown in Table 1.

**Precision Rate, Recall Ratio, f1 Value, Training and Testing Time of Each Model** **Table 1**

Model	Precision [%]	Recall [%]	f1-score [%]	Test time[ms]	Training time[ms]
AlexNet(16)	97.19	98.03	98.05	5,123	211,951
AlexNet(32)	95.53	95.28	95.29	4,926	203,674
AlexNet(64)	97.48	97.24	97.28	4,518	198,631
Attention-AlexNet(16)	98.11	98.03	98.05	5,021	255,283
Attention-AlexNet(32)	97.48	97.24	97.29	4,810	194,913
Attention-AlexNet(64)	97.48	97.24	97.27	4,052	217,527

It can be seen in Table 1 that, in terms of precision rate, recall ratio and f1 value, under the same batch size, Attention-AlexNet model performed better than or not weaker than AlexNet model, and under the batch size of 16, the three evaluation indexes reached optimum, which benefited from the introduction of SAM module in Attention-AlexNet module. This module was able to give different weight values to element values at different locations in the feature map, so as to filter out the effective features for identification. In terms of operating efficiency, although the Attention-AlexNet model increased the training time of the model, after the Attention-AlexNet model is introduced into the SAM module, the model training time increases by 43,332ms, -8,761ms, and 18,896ms when the batch is 16, 32, and 64. However, the test time has been reduced by 102ms, 116ms, and 466ms under the three batch sizes. We can use the test time of the model to express its operating efficiency. The test efficiency of the model has been increased by 1.99%, 2.36%, and 10.31%, achieving good results.

## CONCLUSIONS

In this study, the AlexNet model was constructed and improved to form the Attention-AlexNet model. The pig face identification results were compared between the two models, also the impact of batch size on the model performance was discussed. The conclusions are as follows:

(1) With the improved Attention-AlexNet model, pig face identification test was carried out, and its precision rate, recall ratio and f1 value reached 98.11%, 98.03% and 98.05% at most, increasing by 0.92 percentage points, 0.0% and 0.0% respectively than that of the AlexNet model, also its operating efficiency got increased under different batch sizes, increasing by 10.31% at most, showing that the improved Attention-AlexNet model can be used to identify the pig face more effectively.

(2) The batch size had a certain influence on the prediction result of the model, but there was no fixed relationship between the batch size and the performance of the model.

## ACKNOWLEDGEMENTS

This research, titled 'Pig Face Identification Based on Improved AlexNet Model', was funded by the National Key Research and Development Plan of China (2016YFD0701801), the National Natural Science Foundation of China (31772651), the Doctor Scientific Research Foundation of Shanxi Agricultural University (2020BQ14). The authors are grateful and honoured to have obtained support from the Key Laboratory of Biomechanics.

## REFERENCES

- [1] Chen C., Zhu W.X., Cao J., Ma C.H., Huang W.J., Ruan C.Z., (2017), Image motion feature extraction for recognition of aggressive behaviours among group-housed pigs. *Computers and Electronics in Agriculture*, Vol.142, pp.380-387, Oxford/England;
- [2] Chen Y.P., Li J.A., Xiao H.X., Jin X.J., Yan S.C., Feng J.S., (2017), Dual path networks. *Advances in Neural Information Processing Systems*, pp.4470-4478, New York/USA;
- [3] Chollet F., (2017), Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251-1258, Hawaii/USA;
- [4] Guo Y.Z., Zhu W.X., Jiao P.P., Ma C.H., Yang J.J., (2015), Multi-object extraction from topview group-housed pig images based on adaptive partitioning and multilevel thresholding segmentation. *Biosystems Engineering*, Vol.135, pp.54-60, San Diego/USA;
- [5] Hahnel D, Burgard W, Fox D, Fishkin K., Philipose M., (2004), Mapping and localization with RFID technology. *IEEE International Conference on Robotics and Automation*, pp. 1015-1020, New Orleans, LA/USA;
- [6] Hansen M F, Smith M L, Smith L N, (2018), Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry*, Vol.98, pp. 145–152, Elsevier B.V. /Netherlands;
- [7] Ju M, Choi Y, Seo J, (2018), A Kinect-Based Segmentation of Touching-Pigs for Real-Time Monitoring. *Sensors*, Vol.18, Issue 6, pp.1746, Basel/ Switzerland;
- [8] Kim J, Choi Y, Ju M, (2017), Lying-Pig Detection using Depth Information. *ICACS '17 Proceedings of the International Conference on Algorithms, Computing and Systems*, pp.40-43, New York/USA;
- [9] Krizhevsky A, Sutskever I, Hinton G E., (2012), ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pp.1097-1105, New York/USA;

- [10] Li Y, Sun L.Q., Zou Y.B., (2017), Detection method of moving object pig based on difference method and energy minimization. *International Agricultural Engineering Journal*, Vol.26, pp.428-438, Issue 3, Beijing/P.R.C.;
- [11] Ma C.H., Li M., Zhu W.X., Li H., Li X.C., (2016), Pig target extraction based on adaptive elliptic block and wavelet edge detection. *ICSPS 2016 Proceedings of the 8th International Conference on Signal Processing Systems*, pp.11-15, New York/USA;
- [12] Maselyne J, Saeys W, Ketelaere B D., (2014), Validation of a High Frequency Radio Frequency Identification (HF RFID) system for registering feeding patterns of growing-finishing pigs. *Computers and Electronics in Agriculture*, Vol.102, pp.10-18 , Oxford/England;
- [13] Parkhi O M, Vedaldi A, Zisserman A., (2015), Deep face recognition. *Proceedings of the British Machine Vision Conference (BMVC)*, Vol.1, Issue 3, pp.1-12, Swansea/UK;
- [14] Peter A, Gregersen T, Karstoft H., (2011), Development of a real-time computer vision system for tracking loose-housed pigs. *Computers and Electronics in Agriculture*, Vol.76, Issue 2, pp.169-174, Oxford/England;
- [15] Psota E T, Mittek M, Pérez L C., (2019), Multi-Pig Part Detection and Association with a Fully-Convolutional Network. *Sensors*, Vol.19, Issue 4, pp.852, Basel/ Switzerland;
- [16] Ramprasaath R S, Das A, Vedantam R., (2016), Grad-CAM: Why Did You Say That?. *arXiv preprint arXiv. 1611.07450*, New York/USA;
- [17] Szegedy C, Vanhoucke V, Ioffe S., (2016), Rethinking the inception architecture for computer vision. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2818-2826, Las Vegas/USA;
- [18] Wang J.Z., Liu A.Z., Xiao J., (2018), Video-Based Pig Recognition with Feature-Integrated Transfer Learning. *Biometric Recognition*, pp.620-631, Basel/Switzerland;
- [19] Zhang L., Gray H, Ye X.J., (2018), Automatic individual pig detection and tracking in surveillance videos. *arXiv preprint arXiv, 1812.04901*, New York/USA.