Detection of pig based on improved RESNET model in natural scene

Weixian Song¹, Junlong Fang¹, Runtao Wang¹, Kezhu Tan¹,†, Marwan Aouad²

¹ College of Electrical and Information, Northeast Agricultural University, Harbin, China
² Environmental Health Sciences, FCAS, CUD, Dubai, UAE

Abstract

The behaviours of the pig are often closely related to their health. Pig recognition is very important for pig behaviour analysis and digital breeding. Currently, the early signs and abnormal behaviours of sick pigs in breeding farms are mainly completed by human observation. However, visual inspection is labour intensive and time-consuming, and it suffers from the problems of individual experiences and varying environments. An improved ResNet model was proposed and applied to detect individual pigs in this study based on deep learning knowledge. The developed model captured the features of pigs applying across layer connections, and the ability of feature expression was improved by adding a new residual module. The number of layers was reduced to minimise the net complexity. Generally, the ResNet frame was developed by reducing the number of convolution layers, constructing different types of the residual module and adding the number of convolution kernels. The training accuracy and testing accuracy reached 98.2% and 96.4%, respectively, when using the improved model. The experiment results showed that the method proposed in this paper for checking living situations and disease prevention of commercial pigs in pig farms is potential.

Keywords: deep learning, object detection, ResNet model, feature classification

1 Introduction

Individual pig recognition is very important for pig behaviour analysis and digital breeding. It has been demonstrated that if a pig keeps constant postures such as standing with a raising head or laying for a longer time, it is likely tired or ill [1]. A further health check up and proper treatment are necessary in this case. If the pig keeps standing with its head raising, the pig could be stimulated and disturbed leading to panic or fear by the external environment. Therefore, it is necessary to examine the environment of the pig farm. To study the behaviour of pigs, the first step is to detect the pig target. However, the living situation and disease prevention of
commercial pigs in pig farms mainly depend on human eyes [2]. This visual inspection is labour intensive and time-consuming and usually suffers from the problems of individual experiences and varying environments [3].

In recent years, many researchers have focussed on developing computer vision systems for pig breeding. The application of digital breeding and video monitoring not only analysis the behaviour of pigs in time but also reduces the labour force. Based on 79 data sets of the back image area, actual area, body length, body width, body height, hip width and hip height of pigs, Li Zhuo et al. used linear regression, power regression, quadratic regression and RBF neural network to reconstruct 13 kinds of body mass estimation models, which can be used in the application of machine vision to estimate the body mass of pigs. Alicia Borowska et al. developed an automatic detection device to extract the pig image from the video frame sequence to realise the pig behaviour recognition. Bo Li proposed a novel pig detection method for video monitoring based on the top view in a complex social environment [4]. Liu Yanguo et al. used the inflexion point extraction based on the edge image to extract the key points of the contour map of the living pig obtained in advance [5]. The main problems mentioned by the above research works are high computation costs and difficult manual feature selection. Therefore, it is difficult to meet the real-time requirements.

To solve the problems of traditional recognition algorithms, many researchers have focussed on the method of target recognition using a neural network. CNN is the most representative model in deep learning, especially in the field of image recognition [6]. Since Alex Krizhevsky et al. proposed the AlexNet network, the idea of deep learning was applied to large-scale image classification [7], and the error rate was reduced to 16.4%. Zeiler MD et al. proposed a CNN visualisation method using a deconvolution neural network to visualise all layers of AlexNet and improve the network model of AlexNet and reduced the error rate to 11.7%. In 2015, the Microsoft Asia Research Institute team proposed the RESNET (deep residual network) and reduced the error rate to 3.57% in the image classification task. The depth of RESNET can be deepened easily even >1000 layers [8]. Gao Huang et al. proposed DenseNet based on RESNET structure, which is a kind of densely connected network. There are connections between any two layers, that is, the output of all layers in front of the network constitutes the input of each layer. The features learned in each layer will be directly transmitted to all layers in the back as input. Furthermore, a large amount of data are required for a traditional convolution neural network in the training step while data collection and label are difficult. Due to the longer training time, GPU is necessary for complex CNN, which makes the cost higher in practical applications. It is found that the accuracy will rise to saturation when adding the depth of convolution neural network, and then the accuracy will decrease.

Although a lot of work has been done for pig detection based on vision, there are still limitations in a natural complex environment. To solve these problems, this paper presents an improved RESNET model based on deep learning for individual pig detection. The specific goals are as follows:

1. Expanding the original dataset through the data enhancement method to extract more image features and reduce the overfitting phenomenon in the training stage.
2. Making a trade-off between detection accuracy and training time by improving ResNet structure.
3. Reducing the training time by removing the redundant connection and improving the learning ability through combining different residual blocks.

The rest of the paper is organised as follows: Section 2 introduces the experiment and proposes a method of improved RESNET model. Section 3 shows experimental results and analysis. Section 4 gives the conclusion and future work.

2 Materials and methods

2.1 Image acquisition

When using the deep convolution neural network for pig detection, whether in the training stage or the evaluating stage, a suitable data set is very important for extracting effective features. The image data used in this study were captured from a commercial pig breeding farm. While applying the model in a real pig breeding
farm, several factors were considered in acquiring image data. First, various postures of pigs were captured in the experiment. Because of certain illumination that severely affects the detection accuracy, different illumination conditions were also considered in the work. Second, the model would be difficult to recognise pig objects under occlusion circumstances. Therefore, occlusion pig images were included in the data set. To expand the data set, partial Internet public data sets were also considered. A total of 3000 original images under different scenes were collected to create the training set. All the collected image data were processed using Python language. Some pig images in the image data set are shown in Figure 1.

Fig. 1 Part dataset of this work.

2.2 Proposed method

To detect the pig targets more effectively, RESNET is considered to complete the recognising task. RESNET replaces the larger convolution kernel with several small convolution kernels, which reduces the parameters of the model, increases the number of nonlinear activation functions and the computation cost is effectively controlled. However, the small convolution kernel stack increases the depth of the network, which makes model training difficult and increases the training time.

Different deep RESNET models differ in convolution layers including ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. RESNET can be used to train the deep convolution neural network, and the network framework is relatively simple [9].

With the depth of the convolution neural network increases, the accuracy will rise to saturation, and then the accuracy will decrease. It is found that the error rate of the test set and training set is increasing, which demonstrates that this phenomenon is not due to overfitting. Because not only the error of the test set is increasing but also the error of the training set itself is increasing. To solve this problem, RESNET minimises the residual through fast connection to optimise the learning goal of the network [10]. RESNET’s main idea is to add a direct connection channel in the network, that is, the highway network. The former network structure is a nonlinear transformation of the input, while the highway network allows to retain a certain proportion of the previous network layer in the output. RESNET’s idea is very similar to that of the highway network, allowing the original
input information to be directly transmitted to the later layer, as shown in Figure 2.

![Fig. 2 Residual module.](image)

The output of the network is shown in Figure 3 and it can be expressed as:

$$H(x) = y = F(X, \{w_i\}) + X$$  \hspace{1cm} (1)

$H(x)$ is an elementary mapping that represents network output that needs to be matched by multiple superposition layers. $X$ indicates the input vector of the first layer and $y$ is output. $F(X,\{w_i\})$ represents the residual mapping that needs to be learned. $F=\omega_s s(\omega_1 X)$ indicates the activation function of ReLU [11]. In Eq. (1), the dimensions of the parameters ($X$ and $F$) are required to be the same. When this condition is not satisfied, dimension matching can be performed by linear projection parameters, as shown in Eq. (2)

$$H(x) = y = w_2 s(w_1 X) + \omega_3 X$$  \hspace{1cm} (2)

$\omega_3$ is the same square matrix as in Eq. (1), and $\omega_3$ is only used for matching dimensions.

### 2.3 RESNET model improvement

The structure design is very important for convolution neural networks. Usually, the convolution neural network with a more complex structure would have a stronger ability of feature expression. However, the training time is longer when using the RESNET model with more layers. It is necessary to make a trade-off between detection accuracy and training time. To solve this problem, this paper presents an improved RESNET model, which not only improves the detection accuracy but also reduces the training time of the whole network in pig detection [12]. The idea of improving the model is to reduce the number of network’s layers, remove redundant connections and simplify the network’s structure. Therefore, the training time of the network is reduced through these means. The ability of learning and feature extraction is improved by adding a convolution layer in the fast connection to form a new type of residual block and introducing new convolution cores, respectively.

#### 2.3.1 Residual block design

There are two kinds of residual blocks in the improved RESNET structure in this paper. The first type keeps the original RESNET residual block unchanged. It is composed of two convolution layers. The output of the former layer serves as the convolution layer and the input of the fast connection. It is named class A residual block, and the structure is shown in Figure 3. The second type is composed of three convolution layers, which add the convolution layer at the fast connection. The output of the former layer serves as the input of the convolution layer at the fast connection. The output of the former layer is connected to the next convolution layer. The output of this convolution layer and the output of the convolution layer on the fast connection line are used as the input of the next stage [14]. Here we call it Class B residual block, and the structure is shown in Figure 4.

In Figures 3 and 4, Conv1 and Conv2 represent the first convolution layer and the second convolution layer, respectively. BN is a batch normalisation (BN) algorithm. ReLU1 and ReLU2 are action function. When $x>0$, ReLU(x)=$x$, and when $x<0$, ReLU(x)=0.
2.3.2 Improve the overall structure of the model

Due to single residual block cannot improve the classification accuracy, therefore, two types of residual blocks mentioned in Section 2.3.1 were combined to improve the CNN performance. Specifically, four A residual blocks and three B residual blocks were used in this work. To enhance the feature expression of the improved model, the number of convolution kernels in convolution layers is increased [15, 16]. The overall structure of the improved model is shown in Figure 5. In the improved model, the number of convolution cores in the convolution layer of conv1~conv9 remains unchanged. The number of convolution cores in conv10~conv17 is increased to 32. The size of convolution cores in the whole network structure is 3*3. ReLU was used as an activation function in the improved model.

3 Results and discussion

3.1 Detection result of the improved RESNET model

The improved RESNET model in this paper is compact with 32 convolution layers, and 3*3 convolution kernels are used in all convolution layers [17]. The original data set and the expanded data set are used for training the improved model, and the specific parameter settings in the training process are shown in Table 1.

| Parameter name | test_iter | test_interval | base_lr | weight_decay | gamma | max_iter | solver_mode |
|----------------|-----------|---------------|---------|--------------|-------|----------|-------------|
| Parameter setting | 120       | 500           | 0.1     | 0.0001       | 0.1   | 50,000   | GPU         |

In Table 1, test_iter indicates the times of forward pass in the test set, test_interval expresses that the test is executed every 500 times, base_lr shows the learning rate, weight_decay indicates the weight attenuation, gamma means the learning rate change index, max_iter is the maximum number of iterations [18], solver_mode indicates whether the training mode is GPU or CPU. GPU mode is used for training in this experiment. The improved RESNET model still applies the BN algorithm [19–21]. The idea of the BN algorithm is that the activation input value of a deep neural network gradually shifts or changes before making a nonlinear transformation. Therefore, the BN algorithm can transform the distribution of any neuron in each layer to the standard normal distribution with the mean value of 0 and the variance of 1. This means the activation input value would
Fig. 5 The overall structure of the improved model.

fall in the area where the nonlinear function is more sensitive to the input, therefore, the training speed can be greatly accelerated. The detection results applying the new improved model are shown in Figures 6 and 7.

From Figures 6 and 7, it can be seen that the improved RESNET model in this paper recognises a single pig with higher accuracy. The overall outline of the pig can be detected clearly; however, when detecting a pig in the group, some pigs may be missed due to occlusion.

3.2 The influence of data enhancement on model performance

3.2.1 Expansion of data set

It is known that more features can be learned using more images when training the neural network. Because a large number of original pig images are difficult to collect, the data set needs to be expanded by the data enhancement method such as rotation, mirroring, stagger, shifting and size change [22]. Therefore, more image features can be extracted, and the overfitting phenomenon is also reduced in the training stage. In this paper, the rotation images were acquired using 45°, 90° and 180° transformation from existing images. Shifting images were created by horizontal and vertical transformations. Additional, cross-cutting, scale transformation and colour transformation were applied to expand the size of the data set as shown in Figure 8.

In this study, how the data set influences the detection accuracy, overfitting and generalisation ability was also explored when using the improved RESNET model. The original dataset was expanded by applying data enhancement technology [23]. To make a clear comparison, the original data set was expanded twice and eight times of the original training samples, and then individual pig detection was carried out. Tables 2 and 3 show the distribution of different data sets.
3.2.2 Impact of data set on detection

Figure 9 shows the performance of the improved RESNET model using different data sets. When the original data set was expanded twice, the performance curve is shown in Figure 10(a). The average test accuracy of the network reaches 91.2% after 8000 iterations. Figure 10(b) shows the performance curve when the original data set was expanded eight times. After 5000 iterations, the average test accuracy is 96.4%.

As can be seen from Figure 9, with the increase of the number of images, the accuracy of the model is improved, and the overfitting phenomenon disappears. These changes show that the expansion of the dataset can improve the performance of the model.
Table 2 Distribution of training and testing images for dataset 1.

| Pig posture                  | The number of original images | The number of images after expanding | The images in the training set | The images in the testing set |
|------------------------------|-------------------------------|-------------------------------------|-------------------------------|-----------------------------|
| Standing                     | 600                           | 1200                                | 983                           | 217                         |
| Standing with lowering the head | 200                           | 400                                 | 316                           | 84                          |
| Raising head                 | 240                           | 480                                 | 390                           | 90                          |
| Laying down                  | 55                            | 110                                 | 88                            | 22                          |

Table 3 Distribution of training and testing images for dataset 2.

| Pig posture                  | The number of original images | The number of images after expanding | The images in the training set | The images in the testing set |
|------------------------------|-------------------------------|-------------------------------------|-------------------------------|-----------------------------|
| Standing                     | 600                           | 4800                                | 3844                          | 956                         |
| Standing with lowering the head | 200                           | 1600                                | 1264                          | 336                         |
| Raising head                 | 240                           | 1920                                | 1560                          | 360                         |
| Laying down                  | 55                            | 440                                 | 352                           | 88                          |

Fig. 8 Rotation of driving wheels around the pipe axis.

3.2.3 The detection performance applying different network

To compare the detection performance of different models, AlexNet, GoogLeNet, RESNET and the new improved RESNET model were trained and tested. Figure 10 shows the experimental results of the original
AlexNet model. After about 13,000 iterations, the test accuracy and the loss of AlexNet converge. The average test accuracy and the average loss of AlexNet model are 90.4% and 22.1%, respectively.

![Training and testing results of the ResNet model](image1)

**Fig. 9** The training and testing results of the ResNet model. (a) Dataset 1 and (b) Dataset 2.

![Experimental results of the original AlexNet model](image2)

**Fig. 10** Experimental results of the original AlexNet model. (a) The relationship between test accuracy and the number of iterations and (b) The relationship between model loss and iterations.

Figure 11 shows the results of the GoogLeNet model when training and testing the image data set. After about 12,000 iterations, the test accuracy and the loss of the model converge. The average test accuracy of the GoogLeNet model is 93.1%, and the average loss is 16.7%.

Figure 12 shows the results of ResNet34 model when training and testing the image data set. After about 9000 iterations, the test accuracy and the loss of the model converge. The average test accuracy of the GoogLeNet model is 95.3%, and the average loss of the model is 10.7%.

Figure 13 shows the results of the improved RESNET model. After 10,000 iterations, the test accuracy and the loss of the model converge were detected. The average test accuracy of the network reaches 96.4%, and the average loss of the model is 7.68%.

The test accuracy and the loss of different CNN models are shown in Table 4.

It can be seen from Table 4 that when using the same training sets and test sets, the training accuracy of AlexNet structure is 94.2%, and the testing accuracy is 90.4%; the training accuracy of GoogLeNet model is...
96.7%, and the testing accuracy is 93.1%; the training accuracy of RESNET model is 97.7%, and the testing accuracy is 95.3%; the training accuracy of improved RESNET model is 98.2%, and the testing accuracy is 96.4%. In contrast, the performance of the improved RESNET model is better than other models, which shows that the improved model can effectively detect pigs. At the same time, it shows that the correlation between the depth and complexity of the model is not obvious.
Table 4 The training and testing results of dataset2 of different CNN models.

| CNN model    | Training accuracy (%) | Testing accuracy (%) | Loss (%) |
|--------------|-----------------------|----------------------|----------|
| AlexNet      | 94.2                  | 90.4                 | 22.1     |
| GoogLeNet    | 96.7                  | 93.1                 | 16.7     |
| ResNet       | 97.7                  | 95.3                 | 10.7     |
| Improved ResNet | 98.2              | 96.4                 | 7.7      |

4 Conclusion

A method for recognising individual pigs using an improved RESNET model was proposed in this work. In general, the ResNet frame was developed by reducing the number of convolution layers, constructing different types of residual modules and adding the number of convolution kernels. The major works in this study can be summarised as followings:

(1) A fundamental dataset was created using 3000 original images under different scenes. And then the original dataset was expanded twice and eight times of the original training samples applying data enhancement technology such as rotation, mirroring, staggering, shifting and size change. The method of data expansion proposed in this study improved the test accuracy, reaching 96.4%, and reduced the overfitting phenomenon in the training stage. Finally, the problem of a small dataset was resolved.

(2) The ResNet was improved by reducing the number of network’s layers, removing redundant connections and simplify the network’s structure. The ability of learning and feature extraction is improved through adding a convolution layer in the fast connection to form a new type of residual block and applying new convolution cores, respectively.

(3) Different models, AlexNet, GoogLeNet, RESNET, were compared with the new improved RESNET model. When the same number of training sets and testing sets were used to detect pigs, the training accuracy and testing accuracy of AlexNet structure were 94.2% and 90.4%, respectively. The training accuracy of GoogLeNet model was 96.7%, and the test accuracy was 93.1%. The training accuracy of the RESNET model was 97.7%, and the test accuracy was 95.3%. The training accuracy of the improved RESNET model was 98.2%, and the test accuracy was 96.4%.

Acknowledgments

The authors would like to thank Qingfeng pig breeding Farm in Hei Longjiang province. The authors also would like to thank the Natural Science Foundation of Heilongjiang Province (LH2020C003) and the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province under Grant (UNPYSCT-2018143) for financial support.

References

[1] Yanming Guo; Yu Liu; Erwin M. Bakker; Yuanhao Guo; Michael S. Lew. “CNN-RNN: a large-scale hierarchical image classification framework”, 2017/12/12, Volume 77, Issue 8
[2] C. S. Nandi, B. Tudu, C. Koley. “An automated machine vision based system for fruit sorting and grading”, 2012 Sixth International Conference on Sensing Technology (ICST), 2012
[3] Li X X, Zhang L X, et al. The corn disease remote diagnostic system in China [J]. Journal of Food Agriculture and Environment, 2012, 10(1).
[4] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors [J]. 1986, 323(6088)
[5] Kezhu Tan, Shuwen Wang, Yuzhu Song, Yao Liu, Zhenping Gong. “Estimating nitrogen status of rice canopy using hyperspectral reflectance combined with BPSO-SVR in cold region”, Chemometrics and Intelligent Laboratory Systems,
[6] Zhenyu Lu, Bin Xu, Le Sun, Tianming Zhan, Songze Tang. “3D Channel and Spatial Attention Based Multi-Scale Spatial Spectral Residual Network for Hyperspectral Image Classification”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020.

[7] Daral N. Histograms of Oriented Gradients for Human Detection [J]. Proc. of CVPR, 2005.

[8] Xu L, Xu X, Hu M, et al. Corn leaf disease identification based on multiple classifiers fusion [J]. Transactions of the Chinese Society of Agricultural Engineering, 2015.

[9] Yunhao Shi, Hua Xu, Lei Jiang, Yinghui Liu. “Few-Shot Modulation Classification Method Based on Feature Dimension Reduction and Pseudo-Label Training”, IEEE Access, 2020.

[10] Guo Y, Liu Y, Oerlemans A, et al. Deep learning for visual understanding review [J]. Neurocomputing, 2016, 187(C):27-48.

[11] Dechant C, Wiesner-Hanks T, et al. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning [J]. Phytopathology, 2017, 107(11):1426-1432.

[12] Qiang Guo, Xin Yu, Guoqing Ruan. “LPI Radar Waveform Recognition Based on Deep Convolutional Neural Network Transfer Learning”, Symmetry, 2019.

[13] Yang Yang, Jinhao Duan, Haitao Yu, Zhipeng Gao, Xuesong Qiu. “An Image Classification Method Based on Deep Neural Network with6 Energy Model”, Computer Modeling in Engineering and Sciences, 2018.

[14] Toto Haryanto, Ito Wasito, Heru Suhartanto. “Convolutional Neural Network (CNN) for gland images classification”, 2017 11th International Conference on Information and Communication Technology and System (ICTS), 2017.

[15] Kang, Chen. “Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards”, Sensors, 2019.

[16] Lixin Liao, Yao Zhao, Shikui Wei, Jingdong Wang, Ruoyu Liu. “Finding the Secret of CNN Parameter Layout under Strict Size Constraint”, Proceedings of the 2017 ACM on Multimedia Conference - MM ’17, 2017.

[17] Chunyao Fang, Kebin Jia, Pengyu Liu, Liang Zhang. “Research on Cloud Recognition Technology Based on Transfer Learning”, 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2019.

[18] Kai Yang, Zhiyi Sun, Anhong Wang, Ruizhen Liu, Qianlai Sun, Yin Wang. “Deep hashing network for material defect image classification”, IET Computer Vision, 2018.

[19] Yining Meng, Guoyin Wang, Qun Liu. “Multi-layer Convolutional Neural Network Model12 Based on Prior Knowledge of Knowledge Graph for Text Classification”, 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2019.

[20] Xu L, Xu X, Hu M, et al. Corn leaf disease identification based on multiple classifiers fusion [J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(14):194-201.

[21] Zhao QI, Zhaohui J, Chunhe Y, et al. Identification of maize leaf diseases based on image technology [J]. Journal of Anhui Agricultural University, 2016, 43(2):325-330.

[22] Kai S, Xiaoan S, Jianwei J. Corn leaf disease recognition based on support vector machine method [J]. Transactions of the Chinese Society of Agricultural Engineering, 2007, 23(1):155-157.

[23] Li C, Lanying W. Research on Application of Probability Neural Network in Maize Leaf Disease Identification [J]. Journal of Agricultural Mechanization Research, 2011, 33(6):145-148.