PAPER

Facial Recognition of Dairy Cattle Based on Improved Convolutional Neural Network*

Zhi WENG[†,††], ***Member***, **Longzhen FAN**[††], **Yong ZHANG**[†a)], **Zhiqiang ZHENG**[††b)], **Caili GONG**[†,††], ***and*** **Zhongyue WEI**[†††], ***Nonmembers***

SUMMARY As the basis of fine breeding management and animal husbandry insurance, individual recognition of dairy cattle is an important issue in the animal husbandry management field. Due to the limitations of the traditional method of cow identification, such as being easy to drop and falsify, it can no longer meet the needs of modern intelligent pasture management. In recent years, with the rise of computer vision technology, deep learning has developed rapidly in the field of face recognition. The recognition accuracy has surpassed the level of human face recognition and has been widely used in the production environment. However, research on the facial recognition of large livestock, such as dairy cattle, needs to be developed and improved. According to the idea of a residual network, an improved convolutional neural network (Res_5_2Net) method for individual dairy cow recognition is proposed based on dairy cow facial images in this letter. The recognition accuracy on our self-built cow face database (3012 training sets, 1536 test sets) can reach 94.53%. The experimental results show that the efficiency of identification of dairy cows is effectively improved.

key words: *deep learning, convolutional neural network, facial recognition of dairy cattle, network architecture*

1. Introduction

Inner Mongolia has unique geographical conditions and policy support, which makes the development of the dairy industry prosperous. In Inner Mongolia, the number of dairy cattle is increasing, and the proportion of dairy products is ranked first in China. In 2018, there were 1,516,000 dairy cows in the Inner Mongolia Autonomous Region, with a milk output of 5.656 million tons. The proportion of large-scale breeding of more than 100 dairy cows was more than 80% [1].

Efficient and reliable individual identification technology of dairy cattle is important for the successful operation of large-scale ranches. Based on computer vision technology, dairy cow facial image recognition is an effective means to realize information management of dairy cows. It plays a vital role in the successful operation of large ranches, the control of infectious diseases and the operation of cow insurance [2], [3].

There are three types of traditional dairy cow identification methods: permanent identification methods (PIMs), semi-permanent identification methods (SIMs) and temporary identification methods (TIMs). Permanent identification methods include ear-tattooing, microchip and freeze branding. Semi-permanent identification usually uses ID-choker and ear tags to identify dairy cows. There are also temporary identification methods such as radio-frequency identification (RFID). The traditional methods of cow identification, such as ear lines, ear grooves, hot iron branding and frozen marking, are not reliable enough and not easily changed and copied. Moreover, they are only suitable for marking a small number of cattle, which is not suitable for medium- and large-scale pasture applications. After the cattle leave the pasture, they cannot be monitored and tracked.

Traditional dairy cattle identification approaches have major problems such as registration, traceability and breeding [4]. With the rapid development of computer vision technology, face recognition has increasingly become mature [5]. Similar to human faces, dairy cow faces can be used as the main biometric feature to recognize dairy cow identities because of their rich texture and facial features. The cow's face has short hair and rich facial patterns, which greatly reduces the difficulty of face recognition. In addition, cattle identification using face images, which is non-contact surveillance, is helpful from an animal welfare perspective. In this paper, we explore a network architecture suitable for cow recognition performance and compare the proposed architecture with the existing architecture through experiments to recognize cow individuals by extracting cow facial features by computer vision.

2. Construction of Dataset

Face photos of dairy cattle were collected from the Yibaikang dairy farm in Hohhot City from July to August 2019. A total of 4548 photos were taken of 50 Holstein cows. All photos were taken with a digital camera in a farmed environment. The collected dairy cow facial images were preprocessed by cutting, greying, denoising, and so on. Finally, each image was normalized to 224 × 224 facial images, and the dataset was established. Some images and

Manuscript received January 15, 2022.
Manuscript publicized March 2, 2022.

[†] The authors are with the College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China.

[††] The authors are with the College of Electronic Information Engineering, Inner Mongolia University, Hohhot 010021, China.

[†††] The author is with the School of Mathematical Sciences, Inner Mongolia University, Hohhot 010021, China.

[*] This work was supported in part by the National Training Program of Innovation and Entrepreneurship for Undergraduates under Grant 201910126041, in part by the National Natural Science Foundation of China under Grant 31660678, and Grant 61966026.

a) E-mail: yongz@imau.edu.cn (Corresponding author)
b) E-mail: zqzheng@imu.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2022EDP7008

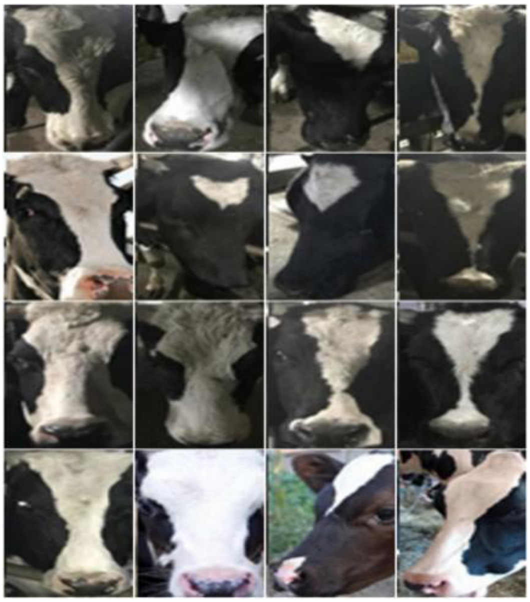


Fig. 1 Partial images of dairy cow facial database.

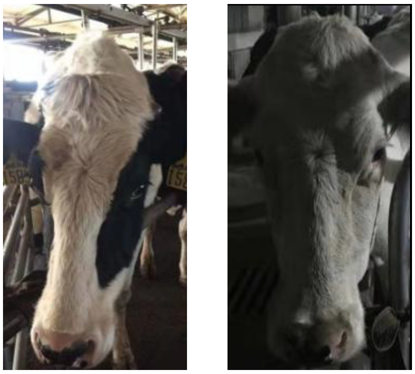


Fig. 2 Images of dairy cow faces with uneven light exposure.

processing results are shown in Fig. 1. In the 50 dairy cows we photographed, each dairy cow had different facial texture patterns.

In contrast to humans, the facial structure of dairy cows is scattered. Compared with human faces, the texture features of dairy cows are more complex, and the hair region is particularly obvious, which makes recognition more difficult. The problem of illumination is one of the most important factors that affect the accuracy of facial recognition in dairy cows. An uneven illumination of light will greatly improve the recognition error rate. The influence of the light distribution of facial images is shown in Fig. 2. In addition, the dairy cows constantly wagged their heads while the photos were being taken. This movement will cause the image to blur. Unsteadiness of the pose will also affect the accuracy of the recognition.

3. Proposed Methods

3.1 Construction of CNN

Deep learning has good anti-interference ability in computer vision and is robust for deformation, illumination and geometric transformation. CNN is a discriminant depth structure that successfully imitates the structure of biological neural networks. One of the most prominent features of CNNs is the local perceptive field. Some basic visual features in the input image are extracted by the local perceptive field, such as oriented edges and corners. Weight sharing is another feature of CNNs, which reduces the number of connection parameters to some extent. The combination of the two can effectively reduce the difficulty of training the network model [6], [7].

The convolution layer is the core part of the CNN. There are many convolution kernels in the convolution layer. The convolution kernel performs convolutions on the previous layer's feature map to extract convolution features. The weights of the same convolution kernel are shared. The weights of different convolution kernels are different, as are the extracted features [8]. The first layer of the convolutional layer can extract only some low-level image features, such as directional edges and corners, while more layers of the network can extract more complex features from low-level features [9].

3.2 Improved Residual Network

3.2.1 Design of the Network Structure

The main idea of a residual network is to add shortcut connections in the network to form residual learning blocks, which is helpful in alleviating the problem of gradient disappearance and gradient explosion; these are caused by increasing the depth of the neural network [10].

Each residual block in the residual network can be described as

$$\begin{aligned} y_l &= h(x_l) + F(x_l, W_l) \\ x_{l+1} &= f(y_l) \end{aligned} . \tag{1}$$

Where x_l and x_{l+1} are the input and output of the lth residual learning block. F is a residual mapping function, $h(x_l)$ is the identity mapping function, and f is the activation function.

3.2.2 Res_5_2Net

Res_5_2Net is constructed using a two-level overlevel connection. We followed the basic architecture of ResNet and made a fair comparison with ResNet. ResNet's block is composed of 2-3 convolutional layers, while Res_5_2Net is composed of 5 convolutional layers and a large shortcut connection to form a large residual learning block. In the block of Res_5_2Net, the first layer is a convolution layer with a convolution kernel size of 3×3 and a stride of 2. The second

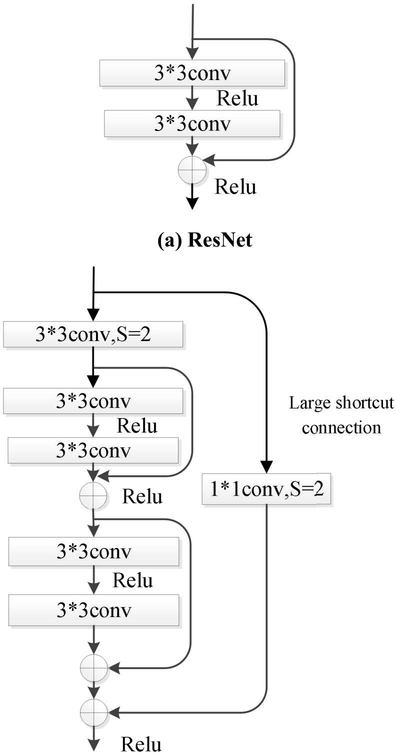


Fig. 3 The block structure of ResNet and Res_5_2Net

and third layers form a small residual learning block, and the fourth and fifth layers form another small residual learning block. To match the size, the large shortcut connection chooses to use a convolution layer with a convolution kernel size of 1 × 1 and a stride of 2. A large shortcut connection promotes information transfer from the high-level residual block to the low-level residual block and alleviates gradient disappearance. The difference between the block structure of ResNet and Res_5_2Net is shown in Fig. 3. Compared with the traditional ResNet, Res_5_2Net introduces more shortcut connections appropriately to enhance the accuracy of the network identification.

At the same time, we designed two improved networks from Res_5_2Net, namely, Improved Res_5_2Net (a) and Improved Res_5_2Net (b). In Improved Res_5_2Net (a), two blocks of Res_5_2Net are added to build a block. In Improved Res_5_2Net (b), we superimposed two blocks of Res_5_2Net in the channel dimension. After superimposition, the channel dimension will be increased, and thus, before concatenation, we first perform a 1 × 1 convolution to reduce the channel dimension to half of the original. The block structures of Improved Res_5_2Net (a) and Improved Res_5_2Net (b) are shown in Fig. 4.

The network structure is shown in Fig. 5. The improved CNN model (Res_5_2Net) uses three-stage cascaded convolution layers with a convolution kernel size of 3 × 3 to extract the underlying features rather than a convolution kernel size of 7 × 7. The receptive field of the three cascaded 3 × 3 convolution kernels is equivalent to that of the 7 × 7 convolution kernel, but the training weight parameter is significantly lower than that of the 7 × 7 convolution kernel. The shared module represents a block of Res_5_2Net or Improved Res_5_2Net (a) or Improved Res_5_2Net (b). The improved CNN model (Res_5_2Net) has good feature extraction ability, and it is easy to optimize the network and adjust its parameters. It can extract the facial feature information

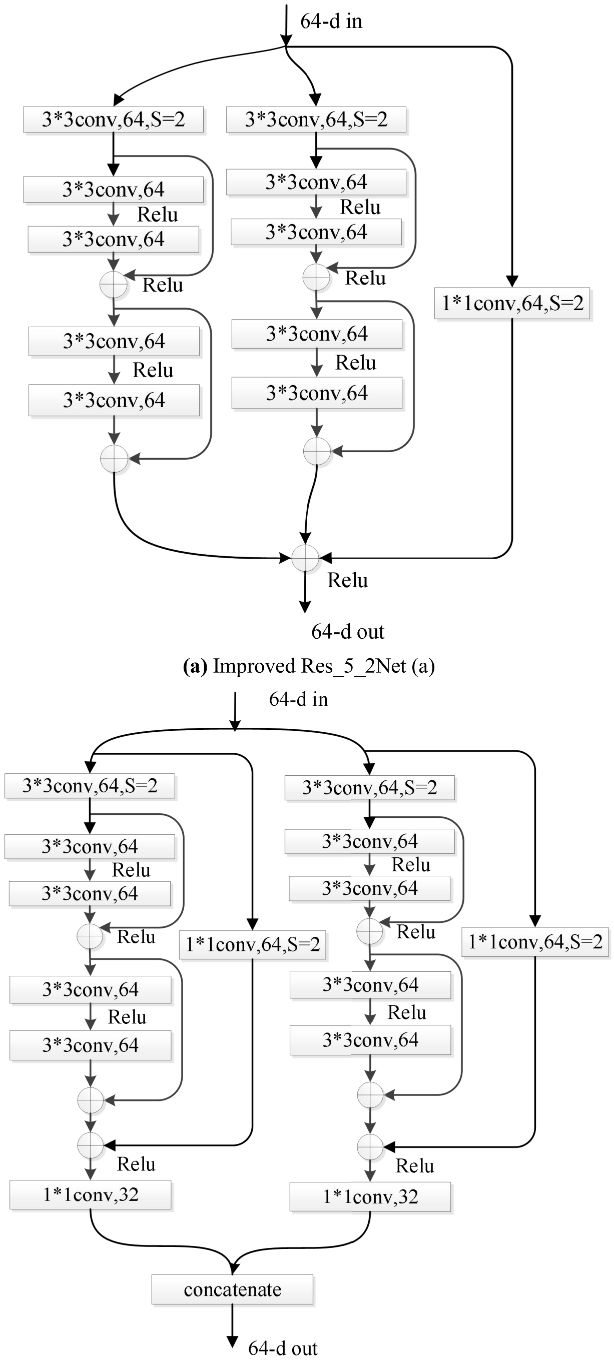


Fig. 4 The block structures of Improved Res_5_2Net (taking channel number C = 64 * 1 as an example)

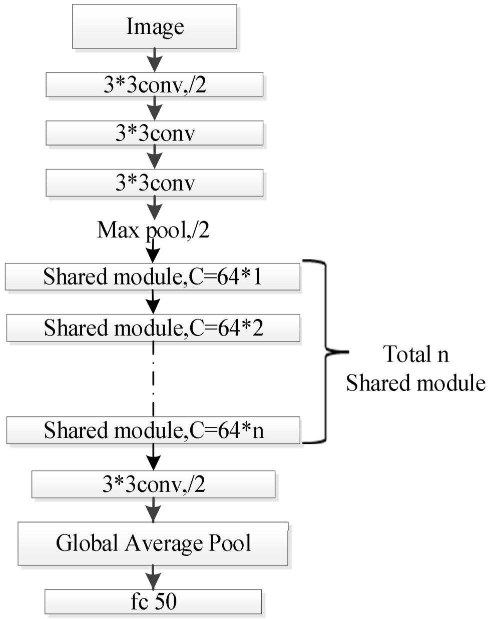


Fig. 5 Network structures of Res_5_2Net and Improved Res_5_2Net (a) and Improved Res_5_2Net (b). (C represents the number of channels)

Table 1 The number of training and testing samples for each cow (partial data)

Cow number	1	2	3	4	5	…	49	50
the number of training samples	53	62	65	59	63	…	54	59
the number of testing samples	25	26	23	28	34	…	31	32

of cows effectively, realize a variety of classifications, and finally identify different individual cows.

3.2.3 Model Expanding the Hyperparameters

We choose the number n of shared modules as a hyperparameter, and we can change the depth of the model by adjusting the size of n. Therefore, we designed six models: we constructed Res_5_2Net-n3, Improved Res_5_2Net (a)-n3 and Improved Res_5_2Net (b)-n3, with the number n of shared modules being 3. Simultaneously, we constructed Res_5_2Net-n4, Improved Res_5_2Net (a)-n4 and Improved Res_5_2Net (b)-n4 with the number n of shared modules being 4.

3.2.4 Network Training

The collected data were transformed into 224 × 224 tagged dairy cow facial images by interpolation. The dataset was divided into the test set and training set. Seventy percent of each dairy cow was randomly selected as the training sample and thirty percent as the test sample. For the dataset, 3012 were obtained as the training set and 1536 as the test set. The number of training and testing samples for each cow is shown in Table 1.

Table 2 The experimental results

The network structure	Number of cattle	Total number of training samples	Total number of test samples	Accuracy of the training set	Accuracy of the test set
ResNet-50[10]	50	3012	1536	98.57%	**93.75%**
Res_5_2Net-n3	50	3012	1536	98.47%	92.71%
Res_5_2Net-n4	50	3012	1536	98.84%	93.56%
Improved Res_5_2Net (a)-n3	50	3012	1536	98.71%	93.29%
Improved Res_5_2Net (a)-n4	50	3012	1536	**98.80%**	**94.40%**
Improved Res_5_2Net (b)-n3	50	3012	1536	98.77%	93.36%
Improved Res_5_2Net (b)-n4	50	3012	1536	**98.97%**	**94.53%**

To optimize the dataset and improve the training quality of the samples, the dairy cow facial image was enhanced before training. It is mainly accomplished through image flipping, zooming, centre rotation and other data expansion. Network training adopts the method of batch training. The training batch size is 32. The initial value of the learning rate is 0.1, the decay rate is 1e-5, the traversal of all data in a training set is an iteration, and the training termination condition is 300 iterations.

4. Results and Analysis

The experimental software environment was based on 64-bit Windows 10, which has a hardware configuration of 16 GB random-access memory (RAM). The TensorFlow open source framework was adopted, and Python was used as the programming language. We used a graphical accelerated processing (GPU) of GeForce RTX 2080 with 8 GB RAM to accelerate the convolution operations used in the models. We trained Res_5_2Net and its improved model on the cow facial dataset and compared the results with other CNNs. After 300 training iterations, the test sample set (1536 pieces) was used for verification. The final experimental results are shown in Table 2.

In Table 2, we can note that the accuracy of Improved Res_5_2Net (a)-n4 and Improved Res_5_2Net (b)-n4 is 94.40% and 94.53%, respectively, which is approximately 0.65%-0.78% higher than that of ResNet-50 (93.75%).

5. Conclusions

In this paper, an improved convolutional neural network structure is designed, i.e., Res_5_2Net, which adopts the method of large shortcut connections to further alleviate the problem of gradient disappearance and to strengthen the feature propagation. In addition, we designed two improved networks from Res_5_2Net to further optimize the network structure. The recognition accuracy of the CNN network in the self-built cow facial image database can reach 94.53%. In the future, we will further optimize the network to reduce the number of network parameters and to improve the accuracy.

References

[1] Q. Shuang, Y.R. Wang, J.W. Zhong, and Z.M. Chen, "The development history and current situation of Inner Mongolia dairy industry," China Dairy Industry, vol.46, no.6, pp.32–35+60, 2018.

[2] S. Kumar, S.K. Singh, A.I. Abidi, D. Datta, and A.K. Sangaiah, "Group sparse representation approach for recognition of cattle on muzzle point images," International Journal of Parallel Programming, vol.46, no.5, pp.812–837, 2018.

[3] S. Kumar, S.K. Singh, R.S. Singh, A.K. Singh, and S. Tiwari, "Real-time recognition of cattle using animal biometrics," Journal of Real-Time Image Processing, vol.13, no.3, pp.505–526, 2017.

[4] S. Kumar, S. Tiwari, and S.K. Singh, "Face Recognition of Cattle: Can it be Done?" Proc. National Academy of Sciences, India Section A: Physical Sciences, vol.86, no.2, pp.137–148, 2016.

[5] R. Wu, and S. Kamata, "Sparse Graph Based Deep Learning Networks for Face Recognition," IEICE Trans. Information and Systems, vol.E101-D, no.9, pp.2209–2219, 2019.

[6] D. Chen, C.D. Yang, H. Ji, B.A. Jiang, and Z. Liu, "Application and Implementation of CNN in Artillery Countermeasure Training System," IOP Conference Series: Materials Science and Engineering, vol.612, no.3, p.032015 (6pp), 2019.

[7] Q. Weng, Z.Y. Mao, J.W. Lin and X.W. Liao, "Land-use scene classification based on a CNN using a constrained extreme learning machine," International Journal of Remote Sensing, vol.39, no.19, pp.6281–6299, 2018.

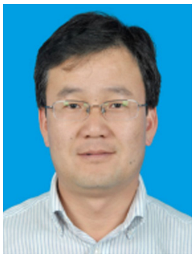
[8] S.H. Cheng, and B. Zhou, "Recognition of characters in aluminum wheel back cavity based on improved convolution neural network," Computer Engineering, vol.45, no.5, pp.182–186, 2019.

[9] M. Dyrmann, H. Karstoft, and H.S. Midtiby, "Plant species classification using deep convolutional neural network," Biosystems Engineering, vol.151, pp.72–80, 2016.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conferenceon Computer Visionand Pattern Recognition (CVPR), pp.770–778, June 2016.

Zhi Weng was born in Inner Mongolia, China in 1978. He received the B.E. degree from Inner Mongolia University, Hohhot, China, in 2000 and the M.S. degree from Inner Mongolia University of Technology, Hohhot, China, in 2003. He is a professor at College of Electronic Information Engineering, Inner Mongolia University. His research interests are related to Signal Processing and Intelligent System, Measurement Technology and Automatic Equipment.

Longzhen Fan was born in Jinan, Shandong Province, China in 1999. He is currently pursuing his B.S. degree in the College of Electronic Information Engineering, Inner Mongolia University. His research interests include neural networks, big data, etc.



Yong Zhang received his Ph.D. degree in Mechanical and Electronic Engineering from Xi'an Jiaotong University, China in 2008. He is a professor at College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University. His research interests are Agricultural Electrification and Automation.

Zhiqiang Zheng received his Ph.D. degree in Automation Science & Electrical Engineering from Beihang University, China in 2012. In 2012, he joined the College of Electronic Information Engineering, Inner Mongolia University, China. His research interests include image processing and pattern recognition, fault diagnosis and prognosis, system control and robot networks.

Caili Gong received the B.S. degree from Inner Mongolia University, Hohhot, China, in 2004, the M.S. degree from Inner Mongolia University, Hohhot, China, China, in 2007. She was a lecturer with the College of Electronic Information Engineering, Inner Mongolia University. Her current research interests include informatization of agriculture and animal husbandry.

Zhongyue Wei was born in Chifeng, Inner Mongolia Autonomous Region, China in 2000. She is currently studying at the School of Mathematical Sciences, Inner Mongolia university. She is now studying knowledge in statistics.