PAPER • OPEN ACCESS

Cattle Face Recognition Method Based on Parameter Transfer and Deep Learning

View the article online for updates and enhancements.

Cattle Face Recognition Method Based on Parameter Transfer and Deep Learning

Hongyu Wang[1], Junping Qin[1*], Qinqin Hou[1*] and Shaofei Gong[2]

[1]Collge of Data Science and Applications, Inner Mongolia University of Technology, Hohhot, Inner Mongolia, 010080, China

[2]Inner Mongolia Smart Animal Husbandry Technology Co., Ltd., Hohhot, Inner Mongolia, 010020, China

[*]Corresponding author's e-mail: W_huhongyu@163.com

Abstract. In order to accurately identify cattle, help insurance companies to deremine claim of Cattle Insurance, or realize inteligent breeding. We need a cattle face recognition technology. Due to the difficulty in data collection of cattle face and the small amount of data, it is difficult to apply deep learning method to cattle face recognition. So there is a method combined with tarnsfer learning effectively trains the model,or initializes the network weights by parametric transfer. This paper proposed using VGGFace dataset pre-training network to solve the problem of small sample face recognition. The VGG-16 deep convolutional neural network model was used to extract the cattle's face features. The model parameters obtained by training the VGGFace face dataset were used to initialize the network weights, and then the network was trained on the cattle face dataset. The experimental results show that the method can obtain 93% recognition accuracy under small sample data.

1. Introduction

Digital and refined farming based on the individual body condition of cattle has become the main development direction of modern cattle breeding. As the basis of intelligent agriculture management and cattle insurance industry, cattle individual identification currently uses radio frequency identification electronic ear tags to identify cattle individuals, but this method has the disadvantages of high cost, pain and easy replacement[1]. Biometric identifiers have invariant characteristics, and animal identification based on biometrics has become a development trend[2]. The cattle face identifier is unique and relatively easy to obtain, so the cattle face is mostly used for cattle identification. The existing cattle face recognition technology mostly uses traditional image feature extraction and classification methods. Cheng Cai, Jianqiao Li et al[3]proposed a face recognition method based on LBP feature. Traditional algorithms have poor robustness to illumination and occlusion, and low recognition in complex scenes. In recent years, deep learning algorithms have shown better performance. Deep learning is a multi-level learning that enhances the accuracy of classification or prediction by deepening the number of layers in the network, using high-level features to form high-level features, and using more abstract features to represent classifications[4]. Deep learning is based on massive data[5], but because of the incompatibility of cattle, the size of the face data is not large. Therefore, using the transfer learning technology, the effective parameters that have been learned are taken as the initial parameters , and learning the characteristics of the cattle face on this basis. The use of transfer learning requires that the source task be similar to the target task,

otherwise it is a negative transfer. The universal transfer method is based on the ImageNet pre-training network for fine-tuning, while the detection using the ImageNet dataset, the rough classification and the cattle face recognition are not similar tasks. In recent years, human face recognition technology has made breakthroughs, both human face and cattle face are facial recognition problems, and the two tasks are similar. Therefore, this paper proposes to use the trained network on the VGGFace face dataset as a pre-training network, and then train the network on the cattle face dataset. The obtained network model can be extracted to distinguish the cattle face features of different cattle individuals, and the feasibility and effectiveness of this method are verified by comparison with ImageNet pre-training network.

2. Dataset Preparation

Using the mobile phone camera to collect 36 cattle data by video, the experimental cattle are the breeding varieties of Simmental cattle introduced in China. The Simmental cattle are native to the Swiss Alps, and the coat color is yellow or white or reddish white. Since the cattle are not controlled and the environment is complicated, it is difficult to collect the image data of the cattle face, so the data is collected by video, and then the video data is framed into image data. When taking video, keep the distance within 1 meter from the cattle, adjust the camera position so that the field of view width is 3 to 4 cow face width, and the field of view height is 1.5 to 1.8 cow face height.

The adjacent frame pictures obtained by video framing is very similar, so this paper uses SSIM to filter and select the picture data. SSIM (structural similarity index), structural similarity, is an indicator to measure the similarity of two images[6]. Given two images x and y, the structural similarity of the two images can be found by the following formula:

$$SSIM_{(x,y)} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ_x is the average of x, μ_y is the average of y, σ_x is the variance of x, σ_y is the variance of y, and σ_{xy} is the covariance of x and y. $c_1 c_2$ is a constant used to maintain stability. The structural similarity ranges from 0 to 1, and the SSIM has a value of 1 when the two images are identical. Each image is compared with all subsequent images, and the similarity exceeds the threshold to screen out one of them. After the frame is divided, 16967 original images are obtained, and 187 cattle face images are obtained after filtering and selection.

In order to solve the problem of insufficient data volume, this paper simulates natural conditions using four methods of image enhancement: rotation enhancement, blur enhancement, noise enhancement and flip enhancement. The image obtained by these enhancement methods has changed from the original image and belongs to two different images. After data enhancement, a total of 1087 images were used as training data for the neural network. Some of the original images and the enhanced image data are shown in Figure 1.
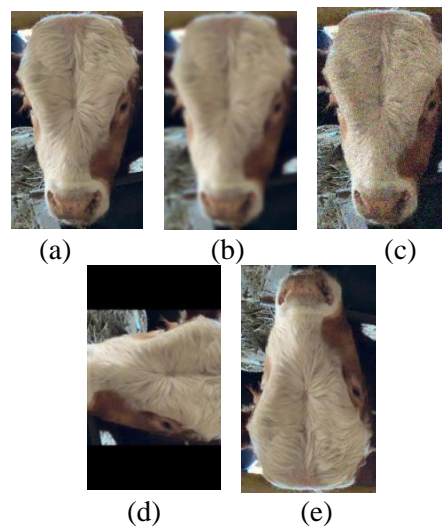
Fig 1. Partial original image and enhanced image.

3. VGGNet Deep Convolutional Neural Network Model

In the 2012 ILSVRC competition, the AlexNet model won the first place, and the final Top5 Error evaluation index was 10% lower than the second team. At the same time, the team using the shallow model uses a variety of methods to extract features to maximize the shallow model. The first model is only a newcomer to the deep model, it has gained great advantages, and the 2016 competition data is more indicative of the compaction of the shallow model by the deep model. The structure of the deep model is closer to the hierarchy of the human brain, so there is greater potential. Deep learning discards the steps of feature engineering, allowing the model to learn and grow according to the original state of the data, so it is easier to learn valuable information in the data.

VGGNet is a convolutional neural network proposed by the University of Oxford VGG visual geometry group Karen Simonyan and Andred Zisserman[7] in 2014. VGGNet has established a 19-layer deep network. The VGGNet network model has five Group convolutions, two layers of FC image features, and one layer of FC classification features. All convolutional layers of VGGNet are filters of the same size, the size is 3*3, the convolution interval is S=1, and the 3*3 convolution layer has one pixel padding. The characteristics of the VGGNet network model are as follows: 3*3 is the lowest size that captures the top, bottom, left, and center concepts; two 3*3 convolutional layers are 5*5, and three 3*3 are 7*7, which can substituted large filter size; multiple 3*3 convolution layers have more nonlinearity than a large size filter convolution layer, making the decision function more discriminative; multiple 3*3 convolution layers have fewer parameters than a large size filter. Although the classification accuracy of VGGNet model is slightly inferior to GoogLeNet, the task performance in many migration learning is better than GoogLeNet. Moreover, extracting CNN features from images, VGGNet is the preferred algorithm[8].

VGGNet has five max-pooling layers, so there are five-stage convolution feature extraction. The number of convolutions per layer starts from 64 in the first phase, and each phase doubles until it reaches the highest 512, and then remains unchanged. VGGNet is famous for VGG-16 and VGG-19. The VGG-16 network structure used in this paper, because there are a total of 36 cattle training data, the number of FC nodes in the last layer is modified to 36, and the network structure is shown in Fig 2.
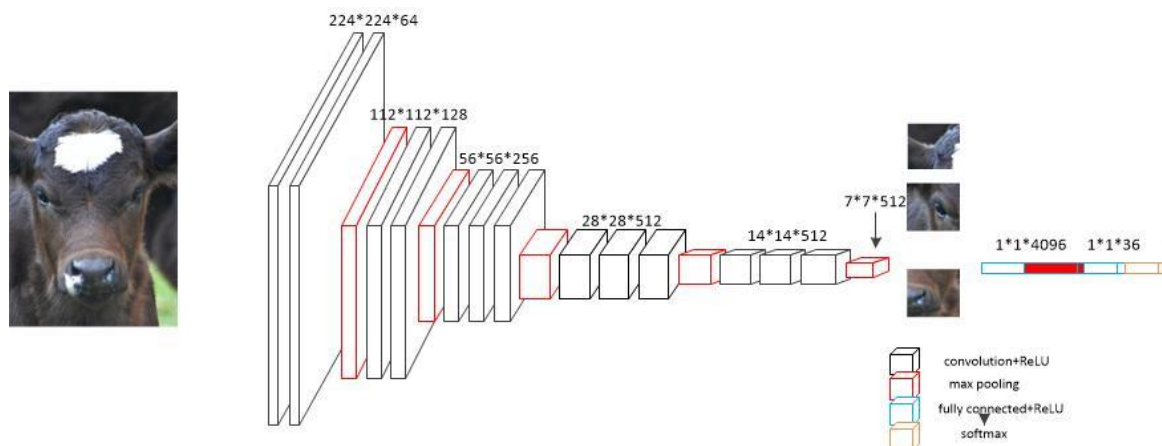
Fig 2. Neural network structure.

4. Parameter Transfer

Although the data has been expanded by the enhanced method, the size of the dataset is still far from the 100,000-level data required to train a neural network. Therefore, this paper uses the transfer learning method to make up for this deficiency. Transfer learning refers to the effect of one learning on another learning, or the impact of learned experience on completing other activities. Expressed in terms of neural networks, the parameters of each node in network transfer from a trained network to a completely new network, rather than training a neural network for each specific task from scratch. Transfer learning can solve the problem of small sample training and computing resources. However, the use of transfer learning requires that the two tasks of the target task and the source task have relevance. If the correlation is not large, it will cause negative transfer. If there is correlation, but it does not find a good transfer component, it is also a negative transfer. Many applications for transfer learning are based on the training of the ImageNet dataset (using the trained model parameters to initialize the network), and then fine-tuning the network parameters with the dataset of the target task. ImageNet is a large dataset containing 1.2 million annotation data and more than 20,000 categories[9]. Using ImageNet trained models and fine-tuning the network can solve the new image rough classification problem well, but this paper studies the cattle face recognition problem, let the neural network learning distinguish different cattles through the cattle face feature, the difference of the features of each cattle's face is subtle, and the effect of using the network model that can be roughly classified is not very good. In fact, the identification is accurate through experiments. The rate is extremely low. Therefore, this paper initializes the neural network weights using the model parameters trained on the face dataset.

The basic features of the bottom layer of the neural network are extracted. The high-level uses the basic features to form local features and then the global features. The cattle face image and the human face image have similar facial features. Therefore, the experiment takes the first two convolution groups of the frozen network, and only fine-tunes the neural network parameters of the next few layers.

5. Experiments and results analysis

The experiment completed the design and training of the neural network in the software environment of Ubantu16.04+Tensorflow combined with GPU acceleration calculation, and realized the visualization of the result through the matplotlib library.

The problem of the face recognition in the actual business is a comparison verification problem, which can be regarded as a two-category problem. The same cattle is 1, and the different cattle is 0. When there are new cattles, there is no need to retrain the neural network, input the picture into the trained neural network, output the 4096-dimensional vector as the feature of the image of the cattle face. Calculating the cosine similarity between the feature vector of the image to be identified and the registered image feature vector in the face database, the value is greater than 0.95 is considered to be

the same cattle. As shown in Fig 3, (a) (b) the similarity value is 0.98, which is the same cattle, and (a) (c) the similarity value is 0.83, which is not the same cattle. A total of three experiments were performed, finetuned network pre-trained on ImageNet dataset, trained network with VGGFace face dataset, finetuned network pre-trained on VGGFace face dataset. And used cattle face feature extracted by these network models respectively for comparison identification.

(a)   0.98

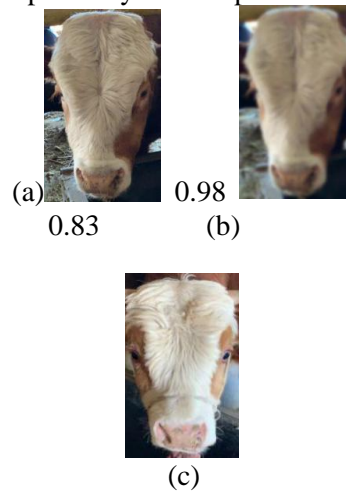0.83   (b)

(c)

Fig 3. Similarity comparison example.

The experiment used 122 pictures of 11 cattle to test, and all the test images were compared by the others, a total of 7381 test samples, including 649 positive samples and 6732 negative samples. Because the ratio of positive and negative samples in the test sample is not equal, three evaluation indicators are used: positive sample recall rate, negative sample recall rate, and accuracy rate. Positive sample recall rate refers to the correct proportion of all positive samples. The negative sample recall rate refers to the correct proportion of predictions in all negative samples. The accuracy rate refers to the ratio of the predicted correct number of samples to the total number of samples in all test samples.

Table 1 compares the results of the three sets of experiments. On the basis of the ImageNet dataset pre-trained network,then finetuned network. The cattle sample has a positive sample recall rate of 0 and a positive sample recall rate of 100%, indicating that the cattle face features extracted by the network are not discriminated. Sexuality, the cosine similarity values of all the test samples in the experiment are around 0.6 and 0.7, less than 0.95, so they are all predicted as negative samples, as shown in Figure 4 (original point indicates that some samples in the test sample, triangles indicate test samples), positive and negative samples can not be distinguished. Using the VGGFace dataset pre-trained network to extract features directly, the positive sample recall rate is 50%, which can distinguish different cattle faces, indicating that the network trained by the human face dataset has learned the common features of the face, but there are differences in cattle face and human face, so on the VGGFace pre-trained network, after using our cattle face dataset fine-tuned the network, the positive sample recall rate is 74%, the negative sample recovery rate is 95%, and the accuracy rate is 93%. The characteristics of the cattle face extracted by the network are discriminative for different cows, and the recognition accuracy is high. The experimental results also show that the cosine similarity value corresponding to the positive sample in the test sample is greater than 0.95, and the negative sample similarity value less than 0.95, as shown in Figure 5 (including partial positive and negative samples).

Table 1. Classification Result.

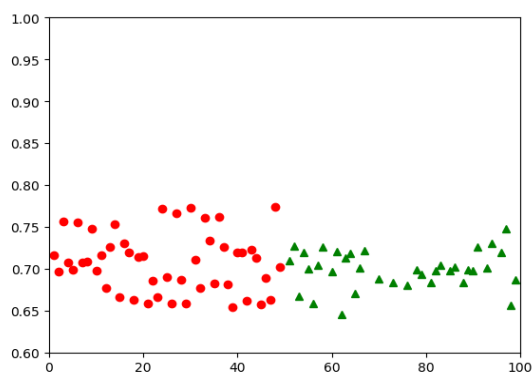|	positive recall	Negative recall	Accuracy
ImageNet pre-trained+fine-tuned	0	100%	91%
VGGFace pre-trained	50%	99%	95%
VGGFace pre-trained+fine-tuned	74%	95%	93%

Fig 4. Similarity value of partial test sample
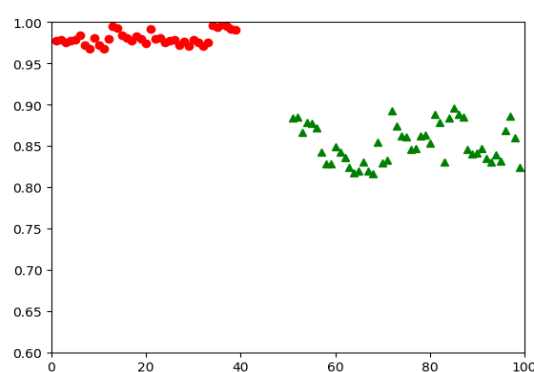(Pre-trained network on ImageNet).

Fig 5. Similarity value of partial test sample.
(Pre-trained network on VGGFace).

6. Conclusion

For small sample cattle face data, transfer learning method can be used, but the target task needs to have similarity with the source task. For example, using ImageNet pre-trained network parameters for poor transfer, this paper uses VGG face dataset pre-trained network. After the parameters are migrated and the self-made cattle face dataset is fine-tuned, the cattle face feature extracted by the network model has good discriminability. The network model is used to output the feature vector, calculate the cosine similarity value, and determine the same cattle according to the threshold value greater than 0.95, complete the alignment identification. However, there are still some shortcomings in this paper. The recognition accuracy needs to be further improved. The current recognition level can meet the application requirements of the pastoral insurance industry (70% accuracy rate). Other application scenarios have more positive sample rate and accuracy. High requirements. Therefore, we need to optimized the production of dataset, improve the performance of the model, and improve the recognition accuracy.

Acknowledgments

References

[1]   Lv C W, Zhang C M.Cattle face recognition method based on convolutional neural network and classifier model: China, CN201710675211. 9[P]. 2014-04-20.

[2]   Awad, Ismail A. From classical methods to animal biometrics: A review on cattle identification and tracking[J]. Computers and Electronics in Agriculture, 2016, 123: 423-435.

[3]   Cai C A , Li J B. Cattle face recognition using local binary pattern descriptor[C]// Signal & Information Processing Association Summit & Conference. IEEE, 2014.

[4]   (Zheng Yuanpan, Li Guangyang, Li Ye.Review of the Application of Deep Learning in ImageRecognition[J]. Computer Engineering and Applications, 2019: 1-1.

[5]   Jin Zhuxin, Qin Feiwei, Fang Meie. Early migration of Alzheimer's disease assisted by deep migration learning[J]. Computer Application and Software, 2019(36): 171-17.

[6]   Hansen M F, Smith M L, Smith L N, et al. Towards on-farm pig face recognition using convolutional neural networks[J]. Computers in Industry, 2018, 98: 145-152.

[7]   Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman. Deep Face Recognition[J].

[8]   Yue Y. Deep learning-Caffe's classic model detailed and practical[M]. Beijing, Electronic Industry Press, 2016, 146-147.

[9]   Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the Acm, 2012, 60(2): 2012.

[10]   Kumar S, Singh S K, Singh A K . Muzzle point pattern based techniques for individual cattle identification[J]. IET Image Processing, 2017, 11(10): 805-814.

[11]   Kumar, Santosh, Tiwari, et al. Face Recognition of Cattle: Can it be Done?[J]. Proceedings of the National Academy of Sciences India, 2016, 86(2): 137-148.

[12]   Kumar S, Tiwari S, Singh S K. Face recognition for cattle[C]// Third International Conference on Image Information Processing. IEEE, 2016:65-72.