Cow Face Detection and Recognition Based on Automatic Feature Extraction Algorithm

Liyao Yao
College of Mathematics and Informatics
South China Agricultural University, China
15555646429@163.com

Zexi Hu
School of Computer Science
University of Sydney, Sydney
huzexi@outlook.com

Caixing Liu
College of Mathematics and Informatics
South China Agricultural University, China
liu@scau.edu.cn

Hanxing Liu
College of Mathematics and Informatics
South China Agricultural University, China
lhx666@scau.edu.cn

Yingjie Kuang
College of Mathematics and Informatics
South China Agricultural University, China
kuangyj@scau.edu.cn

Yuefang Gao
College of Mathematics and Informatics
South China Agricultural University, China
gaoyuefang@scau.edu.cn

ABSTRACT

Automatic farm livestock detection and recognition have high importance in the management of livestock due to the increasing potentials in dairy cow welfare as well as production efficiency. In contrast to the general object (e.g., person, car and bird), the recognition of farm livestock still remains challenging due to the open complex scenarios, similar appearance, shape deformation, occlusion and insufficient annotated data and needs to be solved. In this paper, we discuss the problem of cow face detection and recognition by releasing a new large-scale cow dataset which containing about 50,000 annotated cow face detection data and probably 18,000 cow recognition data. Moreover, a cow face recognition framework is proposed which hybrids the detection and recognition model to improve the recognition performance. Experimental results show the superiority of the proposed method. The accuracy of the detection is 98.3%, and the accuracy of the cow face recognition is up to 94.1%.

CCS CONCEPTS

• Computing methodologies • Object detection; Object recognition; Deep neural networks.

KEYWORDS

Cow face detection, Cow face recognition, Deep neural network

ACM TURC 2019, May 17–19, 2019, Chengdu, China

ACM ISBN 978-1-4503-7158-2/19/05…$15.00
https://doi.org/10.1145/3321408.3322628
Yuefang Gao is the corresponding author

ACM Reference format:

Liyao Yao, Zexi Hu, Caixing Liu, Hanxing Liu, Yingjie Kuang and Yuefang Gao 2019.Cow Face Detection and Recongnition Based on Automatic Feature Extraction Algorithm. In ACM Turing Celebration Conference - China (ACM TURC 2019) (ACM TURC 2019), May 17–19, 2019, Chengdu, China.ACM,NewYork,NY,USA,5 pages. https://doi.org/10.1145/3321408.3322628

1 Introduction

Promotion of cow healthy breeding has been an important topic in recent year due to the increasing concern over animal welfare and production efficiency. It is suggested in [1, 2] that the distinguishing thousands of the cattle could be useful for improving livestock management. To this end, computer vision based methods have been proposed in the past decades [3,4,5]. This is because in contrary to direct visual observation and manual monitoring, it requires less labor cost and is feasible in large scale husbandry. In addition, compared to other measurement device based methods which make use of electronic sensors such as accelerometer, RFID and pedometer, computer vision based method does not have the disadvantage of measurement noise due to sensor malfunction and damage or loss of sensors due to contactless for cow. For example, Gaber et al [2] proposed a muzzle-based cattle recognition method which identified the head of cattle by using Weber local descriptor and the Adaboost classifier for each head image. Li et al.[6] introduced an identification approach of the individual dairy cow with the tail head image of cow. This method extracted shape features of the Region of Interest of the tail head image with Zernike moments, and classified with four classifiers alternatively. Guo et al. [7] used color edge detection and bilateral projection to detect the area of the cow's knees to determine whether the cow was lame. These methods are based on the color and contour features for cow detection and recognition. However, the heavy involvement of hand-crafted features prevents these approaches from application to complex scenarios.

Recently, the recent advancement in deep learning techniques and Convolutional Neural Network models (e.g., Fast R-CNN [8], Faster R-CNN [9], YOLO[10], SSD[11], Mask R-CNN[12], AlexNet[13], VGGNet[14] and ResNet[15]) have made a significant breakthrough on image detection and recognition through feature learning end to end. In [3], a deep convolutional neural network was introduced that learned the black and white pattern of the cow's body for recognition of individual cow. Zhang et al. [16] proposed a CNN-based detector to gain a trade-off between accuracy and speed with the features from prediction network.

In this paper, we propose a cow face recognition framework with coupling the Faster R-CNN detection and PANSNet-5 recognition model [17]. As shown in Figure 1, we combine the detection and recognition models, input a picture first through the cow face detection model to detect and crop out the cow face area in the picture, and then send the cropped image of the cow face area into the recognition model to confirm its specific number. In addition, a cow dataset is released and empirical study of the existing state-of-the-art image recognition methods is performed for performance analysis.

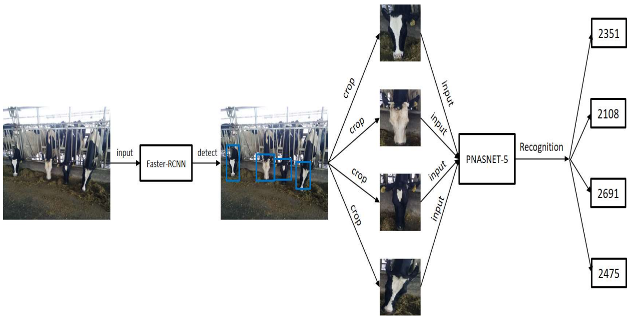

Figure 1: The architecture of cow recognition

2 The Cow Dataset

2.1 Data Collection and Processing

The original dataset is collected over a period of two months in a dairy cow farm in China. To ensure the quality of cow images, we manually use the mobile phone and digital camera to collect the original cow image data when they are fed.

For the cow face detection task, firstly, the dataset must be cleaned, and images that do not meet the requirement are removed. Secondly, the original image taken by the device such as a mobile phone is large in size which is difficult for the training, so the original image is compressed. Moreover, all images are converted to the jpg format, and then the LabelImg tool is used to label the location of the cow face in each image, which is saved with an XML file. Furthermore, the Python script is used to convert the tfrecord file into a network-compliant binary tfrecord file by using the image source file and the XML file.

The processing of the cow face recognition data is relatively simple. Firstly, according to the number of the ear tag in the cow face image, the same ear tag is divided into one class and placed in the same folder. The folder is renamed to the cow ear tag number. Secondly, the images are compressed and are converted into the JPG format. Moreover, the object detection CNN model is used to locate the region of cow face. Since the ear tag of the cow in the recognition dataset represents the final classification result that affects the recognition efficiency of the model, we additionally train the ear tag detection model to detect and code the ear tags in each image.

2.2 Data Description

The cow dataset is divided into the cow face detection dataset and the cow face recognition dataset. The cow face detection dataset contains 51,151 cow face images, which are divided according to the difficulty of the cow face object detection. As shown in Figure. 2, there are about 22,000 single cow face images at different angles, about 18,000 more than two cow faces in the images, about 4,000 occlusion images, and about 6,000 images affected by light changes.

The cow face recognition dataset is the cow face image of 200 cows cut in a cow bar, a total of 18,231 images, as shown in Figure 3, in which cows have about 9000 positive face images , 7,000 side face images and about 2000 occlusion images. These images are divided into 200 folders, and each folder contains about 90 samples of different angles.

Figure 2 Cow face detection data

Figure 3 Cow face recognition data

3 Experimental Method

3.1 Cow Face Detection Model

This paper proposes to use the Faster R-CNN detection algorithm to train the face detection model, and replaces the automatic feature extraction network with different depths in the face detection, such as VGGNet, InceptionV2, ResNet50 and ResNet101. As shown in Figure 4, we first input a cow image, and then scale the image according to the minimum side 600 and the maximum side 1024 to ensure that the image does not deform, and then obtain the feature map through the feature extraction network. The feature map is then fed into the RPN network. After 3*3 convolution and ReLU functions, one branch undergoes a 1*1 convolution and the linear activation function is used for frame regression. The other branch is a two-class network for distinguishing whether each anchor contains a target or no. If it is background, the input is 0, otherwise it is 1. Moreover, the regression-converted border and the real position information are subjected to the non-maximum suppression (NMS) using the proposal layer. Through the ROI(region of interest) pooling layer to extract the image features inside the bounding boxes, since the borders' sizes are different, and then the full connection layer requires the input size to be consistent. To unify the size of input data, the ROI pooling layer divides any feature map into 7 *7 small blocks which each small block utilizes the maximum pooling method. The last branch sends the feature map to the Softmax function to compute the class probability, and the other branch performs further position correction to output the position information of the final cow face.

3.2 Cow Face Recognition Model

In contrast to the cow face detection model, the cow face recognition model is much simpler. This paper exploits the progressive neural network search technology proposed in [17]. The PNASNet-5 image classification model is obtained by the AI automatic search and the method has achieved the positive precision in ImageNet. In this paper, the PNASNet-5 model is utilized as the pre-trained model on the ImageNet to initialize the parameters of the whole network, and then be fine-tuned further on the cow face recognition dataset to obtain the model of cow face recognition.

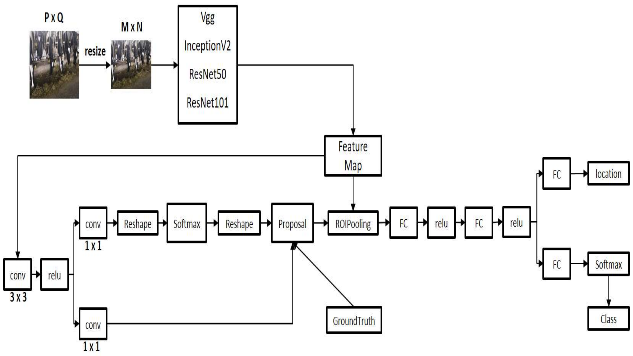

Figure 4 Cow face detection network structure

As shown in Figure.5 (a), the overall structure of the network models compose of several cell units of the same structure, which stride represents the step size. Figure. 5(b) shows the specific structure of the Cell, which is the best network structure searched by the progressive neural network structure search technology. The two sides of the "+" sign form a block, consisting of five different modules. These modules compose of separable convolutional layers (sep) of different sizes, a maximum pooling layer (max), and an identity transformation layer (identity). The outputs of these modules are ultimately connected together to get the input to the next layer of the network. The Cell structure also draws on the idea of a residual network, in which the input of the H^{C} layer is not only related to the output of its upper H^{C-1} layer, but also to the output of the H^{C-2} through the input of the block.

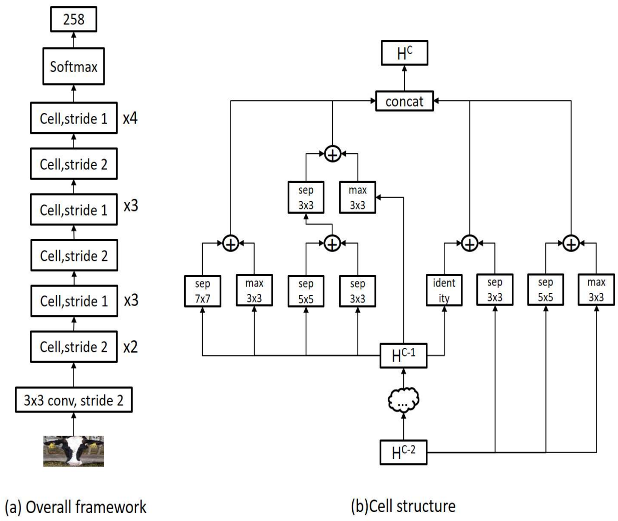

Figure 5 Cow face recognition network structure

4 Experiments Results

4.1 Experimental Setup and Criteria

The experimental environment is Ubuntu16.04 operating system and the deep learning framework is Tensorflow using NVIDIA TITAN X GPU and corresponding CUDA-9.0 and CUDNN-7.1 to accelerate the training and testing process. In the experiment, we divided the whole dataset by the ratio of the training set and the test set close to 8:2. In the cow face detection task, we use the Mean Average Precision (MAP) value to evaluate the detection performance of the model. As for the recognition task, accuracy is used to measure the recognition performance.

4.2 Performance Analysis with Different Detection Models

Because the number of cow face detection data set is large and the background environment is complex, in this experiment the original VGGNet feature extraction model in Faster R-CNN detection algorithm is replaced with a more complex feature extraction model(ResNet101) to improve the face detection MAP. In addition to VGGNet in the experiment, we compared Inceptionv2, Inceptionv3 with wider network width, and ResNet50 with deeper network depth, respectively. It can be seen from the Figure 6 that the final loss value of the ResNet101 feature extraction model is the smallest, the Inceptionv3 and ResNet50 loss values are equivalent, and the final convergence value of Inceptionv2 is slightly higher than the value of VGGNet. The MAP curves on the test set basically corresponds to the loss curves. The larger the loss, the smaller the MAP value. The Faster R-CNN detection algorithm with the original VGGNet feature extraction model has a poor detection effect with an MAP value up to 84.4%, while the Inceptionv2 feature extraction model has a MAP value of 90.7%. As for Inceptionv3 and ResNet50 feature extraction models, the cow detection model has the similar MAP value, with the high-

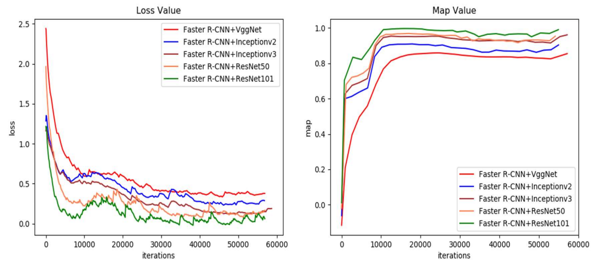

Figure 6 Detection Loss and MAP curves

est MAP 94.7% and 95.1% respectively. The detection algorithm with ResNet101 as the feature extraction model has the best performance, which the highest detection Map reaches 98.3%.

To better measure the performance of the cow face detection model, we separately calculated the detection MAP values of the five detection models under different conditions, as shown in Table 1. It can be seen from Table 1 that the five cow face detection models have higher MAP values for single cow face detection, but the cow face detection model is less effective under the conditions of partial occlusion and illumination changes.

Table 1 Map value of the cow face detection model under different conditions

Model	Single Face	Multiple Faces	Partial Occlusion	Light Change
Faster R-CNN+VGGNet	86.6	82.3	72.3	80.4
Faster R-CNN+ Inceptionv2	93.4	88.0	83.2	87.1
Faster R-CNN+ Inceptionv3	97.7	94.0	85.5	92.3
Faster R-CNN+ ResNet50	98.2	93.1	86.7	91.7
Faster R-CNN+ ResNet101	99.6	97.4	88.7	95.8

4.3 Performance Analysis with Different Recognition Models

Based on the cow face detection results using the above proposed detection model, we compared cow face recongnition approach(PNASNet-5) with the existing image recognition methods, such as LeNet-5[18], AlexNet, VGGNet, and ResNet50 by virtue of the recognition performance of cow face.

Figure 7 shows the loss and accuracy curves of the five recognition models in the cow face recognition dataset. An epoch represents training the entire dataset once. It can be seen from the smoothed loss curve that the loss values of the LeNet-5 and AlexNet network models are not only slower than the other models, but also converges to a relatively large value. The loss value of VGGNet, ResNet50 and PnasNet-5 dropped quickly, and the network models have basically converged after about 25 epochs. Among the final condensed loss values, PnasNet-5 is the lowest.

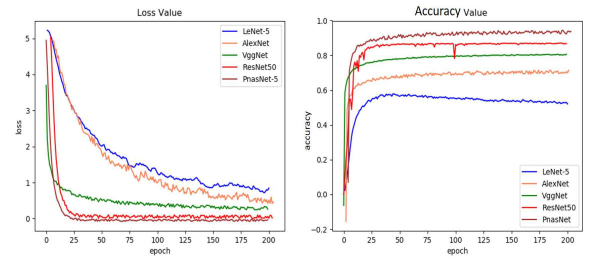

Figure 7 Recognition Loss and accuracy curves

As for the accuracy curve, the LeNet-5 model with the lowest number of network layers has the lowest precision on the test set, with a maximum accuracy of 58.5%. The recognition accuracy of the other models increases continuously before 25 epochs, and then it becomes flat, and the accuracy curve kept fluctuating up and down. Among these methods, AlexNet, VGGNet, and ResNet have the recognition accuracy with 71.4%, 81.1% and 87.7% respectively. While PnasNet-5 has the highest recognition accuracy rate of 94.1%. Table 2 shows the recognition accuracy of the five models under positive face, side face and occlusion conditions.

Table 2 The accuracy under different conditions

Model	Positive Face	Side Face	Occlusion
LeNet-5	58.3	57.6	49.1
AlexNet	71.1	71.7	62.2
VGGNet	82.6	81.2	71.9
ResNet50	88.4	86.4	70.4
PnasNet-5	94.7	93.1	75.3

5 Conclusion

Based on the deep neural network model of automatic learning features, this paper introduces a cow face recognition scheme that couples the CNN-based cow face detection model and the PnasNet-5 recognition model. Moreover, a large-scale cow dataset is released that contains more than 50,000 annotated cow face detection images and more than 18,000 cow recognition images. Extensive experiments and thorough analysis on cow face detection and recognition tasks demonstrate the superiority of the proposed method.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation of China (NSFC No. 61702196).

REFERENCES

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

[2] Gaber T, Tharwat A, Hassanien A E, et al. Biometric cattle identification approach based on weber's local descriptor and adaboost classifier[J]. Computers and Electronics in Agriculture, 2016, 122: 55-66.

[3] Zin T T, Phyo C N, Tin P. Image technology based cow identification system using deep learning [C]. Proceedings of the International Multi Conference of Engineers and Computer Scientists IMECS. 2018.

[4] Arcidiacono C, Porto S M C. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data [J]. Computers and electronics in agriculture, 2017, 134: 124-134

[5] 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

[6] Li W, Ji Z, Wang L, et al. Automatic individual identification of Holstein dairy cows using tailhead images [J]. Computers and Electronics in Agriculture, 2017, 142: 622-631.

[7] Guo Y, He D, Song H. Region detection of lesion area of knee based on colour edge detection and bilateral projection [J]. Biosystems Engineering, 2018, 173: 19-31.

[8] Girshick R. Fast r-cnn [C].Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[9] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [C]. Advances in neural information processing systems. 2015: 91-99

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788

[11] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]. European conference on computer vision,Springer, Cham, 2016: 21-37

[12] He K, Gkioxari G, Dollár P. Mask r-cnn [C]. Proceedings of the IEEE international conference on computer vision, 2017: 2961-2969

[13] Krizhevsky A, Sutskever I, Hinton G. E. Imagenet classification with deep convolutional neural networks [C]. Advances in neural information processing systems, 2012: 1097-1105

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014

[15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778

[16] Zhang L, Gray H. Automatic individual pig detection and tracking in surveillance videos [J]. arXiv preprint arXiv:1812.04901, 2018

[17] Liu C, Zoph B, Neumann M, et al. Progressive neural architecture search [C]. Proceedings of the European Conference on Computer Vision (ECCV). 2018: 19-34

[18] Le Cun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.