Classification and Recognition of Turtle Images Based on Convolutional Neural Network

Jiangchuan Liu*, Mantao Wang, Lie Bao, Xiaofan Li, Jun Sun, Yue Ming

Sichuan Agricultural University College of Information Engineering Ya’an China

*Corresponding author e-mail: liujiangchuan2018@163.com

Abstract. The identification of turtle species mainly depends on the recognition of turtle head and shell, but there is no relevant study on turtle image. In this paper, a turtle image recognition and classification system based on transfer learning is introduced. The system consists of four phases. First, turtle images need to be collected for data enhancement and dataset production. Secondly, the Inception-v3 network model is used to train and save parameters and model structure on the ImageNet dataset. In addition, the network model needs to be modified to change Softmax classifier into 5 categories. Finally, the tortoise dataset was used for training and saving the model, and the classification accuracy of five representative turtles was verified. The experiment proves that the network model of migration learning adopted in this paper is faster and more accurate than the one not adopted.

1. Introduction

Turtles are popular because of their small size and easy to raise. In the world, there are 235 species of turtles. Experts or experienced people can identify turtles by looking at the turtle shell, head and other parts [1]. However, the difference between many turtles is very little, which makes it easy to make mistakes in the identification of turtles. At this time, turtle enthusiasts can upload photos of turtles to the computer, through the identification system can quickly and accurately identify.

The goal of the turtle identification and classification system is to help non-expert users automatically identify turtles and reduce the time cost of identification. Turtle identification system can also be used to make intelligent instruments, and the proposed technology can be used in aquaculture and agricultural production automation to improve the economic benefits of agricultural products [2]. The application of computer vision and image processing technology in disease identification, species classification, product classification, is helpful for improving the yield and quality of aquatic products [3]. In recent years, machine learning algorithms have been widely applied in animal and plant identification. Blanc [4] et al. used binary linear support vector machine classifier to obtain considerable accuracy in fish recognition. MK Alsmadi [5] et al. also considered different image color and light environment, extracted features by measuring distance and geometric shape, and classified them by BP neural network. Shortis [6] et al. measured the length and width of fish, and used the nearest neighbor classifier to realize automatic identification of individual fish detection and identification. Because most of the turtle's shape and size are not different, it is of little significance to annotate the physical shape, and manually annotating the picture of the turtle is time-consuming, laborious and easy to make mistakes. The latest image contest shows that the method based on deep learning is a kind of neural network combined with...
automatic image feature classification [7], especially the classical neural network with deep convolutional neural network, which can achieve better performance in practical application. For example, Zeng [8] et al. added LRN (local response normalization) to lenet-5 to improve the accuracy. Zhang Jianhua [9] et al. improved VGG16 and applied it to cotton disease identification through transfer learning, and the identification effect reached 89.51%. Lin et al. [10] added a SVM classifier to the last layer of AlexNet network, and the classification accuracy reached 96.24%. Although deep learning has also been used to study the classification of aquatic animals [11], the processing is all species with morphological differences [12], while there is no relevant research on the application of turtle recognition. Therefore, this paper studies the application of image processing algorithm in turtle image automatic recognition, tests the performance of six network models, and achieves good results in recognition and classification. The research results in this paper can be used as a supplement to the existing research field and pave the way for further research on image classification of turtles.

2. Materials and methods

2.1. Experimental Materials

In the convolutional neural network, classification accuracy is highly dependent on the dataset used in the training stage, that is, the quality and quantity of the image set annotated by the expert. Therefore, a large enough data set should be prepared for the training and testing of the model before the experiment. This experimental dataset was created by the author himself. It was mainly obtained on the Internet by writing a crawler program, and the correct turtle image was left after screening. Of course, it is necessary to normalize the picture to make the turtle as far as possible in the center of the picture. The picture of the turtle is also selected from different environments. However, as there are few relevant pictures on the Internet, it is necessary to enhance the data of the picture, and the image is scaled, rotated and translated. The turtle dataset constructed according to the existing image data is named SciauTurtle5-2019, which contains a total of 6,212 images of five species of turtle, including 1536 images of Impressed Tortoise, 1,224 images of Cuora Trifasciata, 1,104 images of Star Tortoise, 1,300 images of Eyed Tortoise and 1,048 images of Olecranon turtle. Each image size is saved in JPG format, and each type of image contains various environmental factors, such as light, to enhance the robustness of the model. The dataset prepared in this experiment was named SicauTurtle5-2019, and all the images were judged and confirmed by relevant animal experts. The following Fig.1 is a partial presentation of the data, and the sequence from left to left is that of Impressed Tortoises, Cuora Trifasciata, Star Tortoises, Eyed Tortoises and Olecranon turtle.
2.2. Convolutional Neural Networks

2.2.1. Basic Structure. Convolutional neural network is a multi-layer neural network inspired by human visual perception ability [13]. It integrates spatial context and weight sharing between pixels, and adopts effective representation of the original image [14]. Therefore, only a small amount of data preprocessing is required. Generally, convolutional neural network is composed of input layer, convolution layer, pooling layer, full connection layer and classifier.

2.2.2. Convolution and Pooling. In the field of deep learning, convolutional neural network can maintain the spatial relationship between pixels in the original image through the convolutional layer, thus replacing manual feature extraction [15]. Local perception and weight sharing can reduce the calculation of parameters, enhance the original signal features [16], and reduce the image noise. Through pooling, the adjacent pixels of feature graph can be merged and the feature value can be reduced. Convolution function calculation is shown in equation (1):

\[
X^l_j = f(\sum_{i \in M_j} X^{l-1}_{i} \ast K_{i,j} + b_j)
\]  

In the above equation, \(l\) represents the number of network layers, \(K\) represents the convolution kernel, \(M_j\) represents the combination of input feature graphs, and the bias of the output feature graph of each layer is \(b\).

2.2.3. Softmax classifier. At the last layer of the full connection layer is a classifier, whose function is to solve data classification problems. Commonly used classifiers include support vector machine (SVM) and Softmax classification [17]. SVM is generally used in linear classification, while Softmax classifier is more commonly used for dealing with multiple classification problems, which can limit the output to the range of \((0, 1)\) [18]. Softmax classifier can ensure the classification of image feature data processed by the full connection layer. The specific calculation formula is shown in formula (2) below:

\[
S_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}
\]  

Where, \(z_j\) represents the linear score function corresponding to classification \(j\), \(K\) represents the total number of categories of classification, and \(S_j\) represents the probability size of category \(j\) in all categories.

\[
Loss = -\sum_{k=1}^{n} \sum_{j=1}^{K} t_{k,j} \log(S_{k,j})
\]  

In the above equation, \(n\) is the total number of pixels of the single image, \(t_{k,j}\) is the probability that the pixel \(k\) belongs to category \(j\), and \(S_{k,j}\) is the probability that the model predicts the pixel \(k\) belongs to category \(j\).

2.3. Inception-v3 Based Migration Learning Model

2.3.1. Inception-v3 neural network. Inception network mainly dense matrix classification to approximate the optimal local sparse junction, and stitching of convolution kernel of different sizes can obtain different sizes of sensor field, so as to conduct feature fusion of different scales [19]. The purpose of Inception is not to manually select the filter size or determine whether the convolution layer and pooling layer need to be created, but to learn the parameters at the discretion of the network. Inception-v3 [20] is a new network structure proposed by Google after improvement on GoogleNet [21].
Fig. 2 Inception_v3 Network Structure Diagram

Fig. 2 is the network structure diagram of Inception-v3 used in this paper. The figure includes 6 convolution layers, 2 pooling layers, 1 full connection layer, 1 softmax classifier, and three different Inception modules used. The size of the convolution kernel used in the convolution layer is 3×3, where the step size of Conv_1 and Conv_5 is 2, and the step size of the remaining convolution kernel is 1. In the pooling layer, max_pooling is adopted, where Pool_1's pooling core size is 3×3 and step size is 2, Pool_2's pooling core size is 8×8 and step size is 1.

Inception model maintains a sparse network structure and improves computational performance through a dense matrix. The three Inception modules adopted are shown in Fig.3 below, each of which has four branching units containing a 1×1 convolution kernel, which can splicing the characteristic information of each channel to improve the robustness of the network. Each module also has a variety of convolution kernel and pooling layers of different sizes, which can extract different feature information and increase the adaptability of the network.

Fig. 3 Three module diagrams for Inception-v3

2.3.2. Trans learning. In the application of deep learning, the training and optimization of network model need to rely on a large number of labels and data, and the production of data sets is time-consuming and laborious. VGG16 [22] and Inception-v3 [23] network have good classification ability in image classification, but it is time-consuming in the model training process, and the accuracy obtained in small datasets is not high. The application of transfer learning can effectively solve this problem, reduce the dependence on data, and improve the generalization ability of the model. Fig.4 shows the diagram of the migration learning model used in this paper. In the diagram, the whole structure is divided into two parts: the pre-training network nuclear migration network. The parameters of the pre-training model were obtained from the ImageNet data set with 1000 categories [24] and then migrated to the turtle classification task. The migrated data is only a small part of the migrated network and can be trained on a small dataset by resisting the turtle image. The network layer of the pre-training model is mainly used to extract features and obtain data to help the training convergence, which can greatly solve the training time.
3. Results and analysis

3.1. Data Distribution

The experimental platform of this identification system model is based on the Windows 10 operating system, which uses a GTX1050 Independent Graphics with 4G video memory, Intel i5 CPU and an HP Notebook computer with 12GB of running memory. The experimental program is based on python3.6.5 and runs on pyCharm2017 software with the Tensorflow1.0 framework. Before the experiment, SicauTurtle5-2019 dataset needs to be divided, 80% of which is divided into training set for model training and parameter learning, and 20% is used as validation dataset to optimize the model. During the training process, parameters are automatically fine-tuned according to the test results of the model. The specific dataset allocation data is shown in Table.1 below.

|                  | Total | Train | Validation |
|------------------|-------|-------|------------|
| Impressed Tortoise | 1536  | 1227  | 308        |
| Cuora Trifasciata | 1224  | 979   | 245        |
| Star Tortoise     | 1104  | 883   | 221        |
| Eyed Turtle       | 1300  | 1040  | 260        |
| Olecranon turtle  | 1048  | 838   | 210        |
| **Total**         | 6212  | 4967  | 1245       |

3.2. Experimental Network Parameter Setting

After the network model is selected, some parameters need to be set for the network structure in the training process, as well as the selection of the optimizer, and experiments need to be carried out for fine-tuning. In this paper, comparative experiments were conducted on Learning Rate, Dropout value and the optimizer, etc., to continuously optimize the network structure, and finally a better setting scheme was obtained: Learning Rate=0.01, weight decay Rate=0.8, iteration times 5000, Dropout=0.8, batch=128. In order to avoid over-fitting during network training, Dropout was adopted in the full connection layer to improve generalization ability, and only a part of neurons were trained in each iteration. If the network is divided into multiple child nodes, the single child node may be over-fitted to some extent, but the network will shield some nodes to improve the robustness of the network. The network model proposed in this paper adopts the mini-batch gradient descent method, which updates the gradient in batches to speed up the parameter update.
3.3. Model Evaluation

3.3.1. Parameters. The performance indicators to measure the image classification include f-measure, Accuracy, ROC (operating characteristic receiver) and AUC, etc. Accuracy is used in the experiment of this paper to evaluate the effect of the model. The accuracy and loss rate of the validation dataset are used in this experiment. The specific formula is shown in formula (4):

\[
\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{R_A}{R} \times 100\%
\]  

(4)

Where R represents the total number of verified images and \( R_A \) represents the number of properly classified images. TP represents the number of pictures accurately classified as non-a species, FP represents the number of pictures incorrectly classified as A species, TN represents the number of pictures accurately classified as A species, and FN represents the number of pictures of A species classified as non-a species.

3.3.2. Experimental visualization. The training data of the convolutional neural network (inception_v3-TL) based on transfer learning and Inception_v3 proposed in this paper were saved. The experiment experienced 5000 iterations, and the verification set accuracy and loss rate of each iteration were saved. With the number of iterations as the abscissa and the accuracy of the verification set or the loss rate as the ordinate, the curve images of the accuracy and loss rate of the turtle verification set were obtained, respectively, as shown in Fig.5 and Fig.6. It can be seen from the analysis of the curve image that the recognition accuracy increases significantly between 0 and 2800 steps in the number of iterations, while the curve gradually flattens after that. The accuracy reaches 90% at 1750 steps, and the best is 96.83% at 5000 steps. Similarly, the trend of the image of the loss rate curve is similar to that of the accuracy curve. The lowest loss rate is only 3.27%, which indicates that there is still room for reduction as long as appropriate improvement is made. According to the comprehensive analysis, there is no big fluctuation in the curve change of the accuracy and loss rate of the verification dataset, nor any fitting or underfitting phenomenon, which indicates that the turtle recognition and classification model proposed in this paper has good performance and meets the requirements.

![Validation dataset accuracy rate](image-url)  

Fig. 5 Validation dataset accuracy rate
3.4. Experimental Comparison

In order to verify the performance of inception_v3-TL model used in this paper in turtle classification, the comparative experiment adopted LeNet-5, AlexNet, VGG16, VGG16-TL and other four models of transfer learning. Similarly, the four models are trained in the same dataset, the training data are saved and drawn into a change curve as shown in Fig.7, where Fig.7 is the accuracy rate. As can be seen from the graph, the initial accuracy of the five models is relatively low, with the highest AlexNet model accuracy of 59.75% and the lowest inception_v3-TL accuracy of 28.11%. With the beginning of training, accuracy began to show an upward trend, Inception_v3-TL increased the fastest, lenet-5 did not change much, and the final accuracy was only 65.75%. As transfer learning models, Inception_v3-TL and VGG16-TL have little difference in curve change between 200-2200 steps, but the accuracy of Inception_v3-TL after 2200 steps is higher than that of VGG16-TL. The final accuracy of the five models was recorded in Table.2. The table shows that the accuracy of the models without transfer learning is lower than 90%, while the recognition accuracy of VGG16-TL and inception_v3-TL with transfer learning is improved, reaching 91.70% and 96.83% respectively.

![Fig. 6 Validation dataset loss rate](image)

**Fig. 6 Validation dataset loss rate**

![Fig. 7 Graph of classification accuracy of five network models](image)

**Table. 2 Experimental comparison results of verification set of the model**

| Model         | Loss (%) | Act (%) |
|---------------|----------|---------|
| LeNet-5       | 34.25    | 65.75   |
| AlexNet       | 18.06    | 81.94   |
| VGG16         | 13.44    | 86.56   |
| VGG16-TL      | 8.30     | 91.70   |
| Inception_v3-TL | 3.17    | 96.83   |
In order to further test the generalization ability of the five model, 500 pictures of turtles in different environments were taken to form the test dataset. 100 pictures of each kind of turtle were taken, and none of them participated in the training and validation optimization of the model. In order to maintain the final accuracy, the test dataset was input into five models for testing, and the test results were shown in Table.3. The table shows that the generalization ability of inception_v3-TL model is relatively strong, and the recognition and classification effect of five species of turtles is relatively good, with an average accuracy (ACC) of 96.4%.

Table. 3 Identification results of different Turtle species by different methods

|       | I   | II  | III | IV  | V   | ACC |
|-------|-----|-----|-----|-----|-----|-----|
| LeNet-5 | 67.0| 69.0| 62.0| 65.0| 63.0| 65.2|
| AlexNet | 83.0| 84.0| 82.0| 78.0| 76.0| 80.6|
| VGG16  | 86.0| 87.0| 84.0| 82.0| 83.0| 84.4|
| VGG16-TL | 91.0| 95.0| 90.0| 92.0| 89.0| 91.4|
| Inception-v3 | 88.0| 90.0| 85.0| 86.0| 87.0| 87.2|
| Inception_v3-TL | 98.0| 98.0| 93.0| 97.0| 96.0| 96.4|

Notes: I- Impressed Tortoise, II- Star Tortoise, III- Olecranon turtle, IV- Cuora Trifasciata, V- Eyed Turtle, ACC- Average Accuracy

4. Conclusion

Deep learning is widely used in animal and plant disease identification, image classification and other applications, which can effectively solve some practical research problems. In this paper, a turtle image recognition and classification system based on transfer learning was proposed, and a SicauTurtle5-2019 turtle dataset containing 6,212 images was created. In this system, 500 turtle images can be tested for effective classification and recognition, with an average accuracy of 96.83% and between 90% and 99% for a single category. Through comparative experiments with LeNet, AlexNet, VGG16, Inception_v3 and vgg16-TL, it is found that the performance of the currently used model is far better than that of the network model without using migration learning, and the accuracy rate is higher than that of vgg16-TL. The recognition effect of this system has met the application requirements. With the expansion of dataset, the improvement of image algorithm, and the identification of diseases, the recognition effect will become more superior. Later, the turtle classification model is transplanted to mobile devices to realize intelligent identification and classification of turtles in smart phones and enhance the application value of the model.

Acknowledgments

This work was supported by “Youth Fund Project of Sichuan Education Department” (18ZB0467) and “Agricultural Information Engineering” Key Laboratory of Sichuan Province University.

*Corresponding author: Wang Mantao, Associate Professor, Sichuan Agricultural University. E-mail: 545550965@qq.com.

References

[1] Zhi Jinsong. Turtle Artificial Breeding Technology. Chinese Aquatic Products, Beijing:2003, pp.39-40.
[2] Raghavan U N, Albert R, Kumara S. Near linear time algorithm to detect community structures in large scale networks. Physical Review, New York:2007, pp.96-106.
[3] Zhang Yanqi, Zhu Hong, Xu Chenghua. Research on the Development Status of Agricultural Product Classification at Home and Abroad [J]. China Standardization, Beijing:2014, pp.70-75.
[4] K. Blanc, D. Lingrand, F. Precioso. Fish species recognition from video using SVM classifier. In Proceedings of the 3rd ACM International Workshop on Multimedia Analysis for Ecological Data ACM (2014, November), pp.1-6.
[5] M.K. Alsmadi, K.B. Omar, S.A. Noah, I. Almarashdeh. Fish recognition based on robust features extraction from size and shape measurements using neural network. J. Comput. 2010, pp. 1088.

[6] M.R. Shortis, M. Ravanbakhsh, F. Shafait, A. Mian. Progress in the automated identification, measurement, and counting of fish in underwater image sequences. Mar. Technol. Soc. J., 2016, pp. 4-16.

[7] Mehmet Metin Ozguvena, Kemal Ademb. Automatic detection and classification of leaf spot disease in sugar beet using deep learning. Physica A: Statistical Mechanics and its Applications, 2019, pp. 535-537.

[8] Zeng Pingping, Li Linsheng. Research on fruit image classification and recognition based on convolution neural network. Mechanical design and research, 2019(35), pp. 23-26.

[9] Zhang Jianhua, Kong Fantao, Wu Jianzhai et al. Cotton disease identification model based on improved VGG convolution neural network. Journal of China Agricultural University, 2012, pp. 133-140.

[10] Lin Mingwang. Application of deep learning in fish image recognition and classification. Digital Technology and Application, 2017, pp. 96-97.

[11] Sébastien Villon, David Mouillot, Marc Chaumont, et al. A Deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecological Informatics, 2018, pp. 238-244.

[12] CC Yang, SO Prasher, JA Landry, H. Ramaswamy, A. Ditommaso. Application of artificial neural networks in image recognition and classification of crop and weeds. Canadian Agriculture Engineering, 2000, pp. 147-152.

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

[14] Ahmed El-Sawy, Hazem EL-Bakry. Mohamed Loey CNN for Handwritten Arabic Digits Recognition Based on LeNet-5. Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016. pp. 566-575.

[15] YIN S Y, OU Y P, DAI X, et al. An adaboost based face detection system using parallel configurable architecture with optimized computation. IEEE Systems Journal, 2017, pp. 260-271.

[16] Bengio Y. Learning deep architectures for AI. Foundations Trends Machine Learning, 2009, pp. 5-6.

[17] Liu W, Wen Y, Yu Z, et al. Large-Margin Softmax Loss for Convolutional Neural Networks. Proceedings of The 33rd International Conference on Machine Learning, 2016, pp. 507-516.

[18] Lian Xiaoin, Cheng Kiiyuan, An Sa, et al. Fruit Image Classification Based on Deep Learning and Trans Learning. Pattern Recognition and Artificial Intelligence, 2019, pp. 15-18.

[19] Alessandra Lumini, Loris Nanni. Deep learning and transfer learning features for plankton classification. Ecological Informatics, 2019, pp. 33-43.

[20] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818-2826.

[21] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [J]. 2014: 1-9.

[22] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks For Larger Scale Image Recognition. Published as a conference paper at ICLR, 2015, pp. 1067-1081.

[23] S. J. Pan, I. W. Tsang, J. T. Kwok and Q. Yang. Domain Adaptation via Transfer Component Analysis. IEEE Transactions on Neural Networks, 2011, pp. 199-210.

[24] Aydin Kaya, Ali Seydi Kececi, Cagatay Catal et al. Analysis of transfer learning for deep neural network based plant. Computers and Electronics in Agriculture, 2019, pp. 20-29.

[25] Yosinski, J, Clune, J, Bengio, Y, Lipson, H. How transferable are features in deep neural networks?. In Advances in neural information processing systems (NIPS 2014). Montreal, Canada: NIPS 2014, pp. 3320-3328.