Bird image classification based on improved ResNet-152 image classification model

Huitong Song

School of Business, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, 999078, China

1609853pbb30001@must.edu.mo

Abstract. Bird image classification is an important research direction in the field of computer vision, and its main purpose is to automatically identify bird images through computers to achieve bird classification, recognition and monitoring. Birds are an important part of the planet's biodiversity and have an important impact on the balance of ecosystems and human survival. ResNet-152 is a deep convolutional neural network model, which is the deepest model in the ResNet family, with a depth of 152 layers. The model mainly solves the problems of gradient vanishing and gradient explosion in deep neural networks through residual learning, and improves the accuracy and generalization ability of the network. The ResNet-152 model has achieved good results in the fields of image classification, object detection and semantic segmentation. For the bird classification dataset BIRDS 525 SPECIES-IMAGE CLASSIFICATION, the ResNet-152 model was used for classification, and excellent results were obtained. The classification accuracy reached 96.5%, the accuracy reached 98.0%, the recall rate was 94.6%, the f1 score was 94.7%, and the AUC reached 95.6%. These indicators indicate that the ResNet-152 model has high classification accuracy and generalization ability for bird classification datasets. At the same time, it also shows that the ResNet-152 model has a good application prospect in processing large-scale image classification tasks.

Keywords: Image classification, ResNet-152, AUC.

1. Introduction

Bird image classification is an important research direction in the field of computer vision, and its main purpose is to automatically recognize bird images by computers to achieve bird classification, identification, and monitoring [1]. The background of this research direction is that birds are an important part of the planet's biodiversity and have an important impact on the balance of ecosystems and human survival [2]. However, the classification and monitoring of birds usually requires a lot of manpower and time, so the development of an automated bird image classification system can greatly improve the efficiency and accuracy of work [3].

Deep learning models are widely used in bird image classification, and the most commonly used models include convolutional neural networks (CNNs), residual neural networks (ResNets), and Inception networks [4]. The application of these models in bird image classification mainly includes two aspects: one is to use pre-trained models for transfer learning, and the other is to fine-tune the models [5].

Transfer learning using a pre-trained model refers to the application of a deep learning model that has been trained on a large-scale image dataset to a bird image classification task. This approach can greatly improve the classification accuracy of the model, while also reducing the training time and the need for data volume [6]. Commonly used pre-trained models include VGG, ResNet, and Inception. Among them, the ResNet model performed well in bird image classification tasks [7,8].

Fine-tuning the model refers to further training and tuning the model for bird image classification tasks on the basis of the pre-trained model [9]. This approach can further improve the classification accuracy of the model, but it requires more training time and data volume. During the fine-tuning process, the last layers of the model often need to be retrained to accommodate the new classification task [10].

The ResNet model is also widely used in bird image classification. The ResNet model has a very deep network structure, which can effectively solve the gradient vanishing problem in deep neural networks. In bird image classification tasks, the ResNet model can effectively improve the classification accuracy, and compared with other deep learning models, the training time and computational resource consumption of the ResNet model are relatively small. Therefore, the ResNet model is widely used in bird image classification tasks. In this paper, bird image classification is carried out based on the algorithm principle of the ResNet-152 model, which provides a theoretical basis for subsequent research.

2. Introduction to the dataset and the principle of the ResNet-152 model

Please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. This paragraph follows a section title so it should not be indented.

2.1. Introduction to datasets

The BIRDS 525 SPECIES- IMAGE CLASSIFICATION dataset is a dataset for bird image classification, containing 525 bird species with approximately 500 images per species. The dataset was created by the University of California, San Diego (UCSD) and Cornell University to provide a standard dataset for the study of bird image classification algorithms.

The images in this dataset are from bird photographers and bird watchers, and the quality and diversity of the images are high. Each image has a species tag associated with it, as well as other attribute information such as the source of the image, photographer, geographic location, etc. Applications of this dataset include bird species identification, bird species protection, ecological research and other fields.

The dataset has become a widely used standard for bird image classification, and many deep learning models have been trained and tested on it. For example, models such as ResNet, Inception, VGG, etc. have all been tested on this dataset with good results. Among them, ResNet's performance on this dataset is particularly outstanding, and its classification accuracy on this dataset can reach more than 90%. This is mainly due to the depth residual structure of ResNet, which enables the model to better solve the problems of gradient disappearance and gradient explosion during training, thus improving the accuracy of the model.

2.2. ResNet-152 model principle

Resnet-152 is a deep convolutional neural network model, which is the deepest of the ResNet family of networks, with 152 layers. The model principle of ResNet-152 is based on the idea of Residual Learning, and the problem of gradient disappearance and gradient explosion in deep convolutional neural networks is solved by introducing Residual Block.

In ResNet-152, each residual block contains two convolutional layers and a Shortcut Connection. Skip joins pass the input data directly to the output of the Residual block so that the model can learn about the residual. In this way, during the training process, the model can optimize the parameters of the network by learning the residual information, thus improving the accuracy of the model.

The model structure of ResNet-152 can be divided into several stages, each of which contains multiple residual blocks. Among them, the first four stages use different residual blocks, and the number of residual blocks in each stage is different. The fifth stage is the global average pooling layer and the fully connected layer, which are used to convert the convolution features into class probabilities.

ResNet-152 has achieved excellent performance in large-scale image classification tasks, and its accuracy exceeds the human level, making it one of the best image classification models available. The ResNet-152 is also widely used in other computer vision tasks, such as object detection and image segmentation.

Resnet-152 is a deep convolutional neural network model that is the deepest of the ResNet family of networks, with 152 layers. The main principle of ResNet-152 model is to solve the gradient disappearance problem in deep neural networks through the design of residual blocks, so that the model can better learn the feature information. In this paper, the model principle of ResNet-152 is introduced in detail from the aspects of network structure, residual block design, the function of skip connection and the function of global average pooling layer.

The network structure of the ResNet-152 model is very deep, with a total of 152 layers, including multiple convolutional layers, pooling layers, and fully connected layers. The ResNet-152 model uses a deep convolutional neural network, which contains multiple convolutional layers, each of which extracts different feature information.

Gradient disappearance is a very important problem in deep neural networks, which can cause the model to fail to learn the deep feature information, thus affecting the performance of the model. To solve this problem, the ResNet-152 uses a residual block design. The residual block is composed of two convolution layers and a jump link, which can directly add the input feature map to the output feature map, so as to solve the gradient disappearance problem and make the model learn the feature information better.

Skip connection is a very important part of ResNet-152 model, it can solve the gradient disappearance problem in deep neural networks, so that the model can learn the feature information better. The function of skip connection is to add the input feature map directly to the output feature map, so as to retain the information of the input feature map, so that the model can learn the feature information better.

Global average pooling layer is a very important component of ResNet-152 model, which can reduce the spatial information of feature graphs, reduce the number of parameters of the model, and improve the generalization ability of the model. The function of the global averaging pooling layer is to pass the output feature map of the last residual block into the global averaging pooling layer, average the values of each channel of the feature map, and get a feature vector. The global average pooling layer can reduce the dimension of the spatial information of the feature graph, reduce the number of parameters of the model, and improve the generalization ability of the model.

The size of the input image can be adjusted according to the specific situation, usually the size of the input image is 224x224 or 299x299, as shown below:
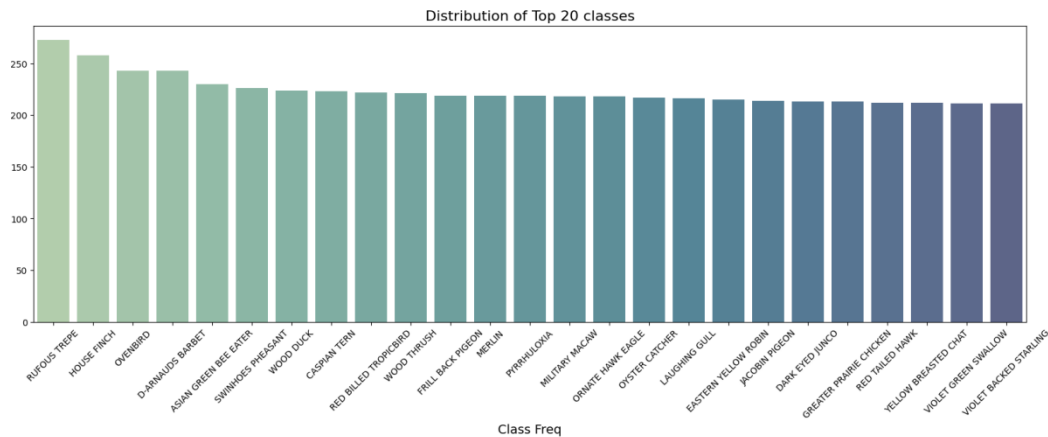

Figure 1. Data set introduction.
(Photo credit: Original)

Figure 2. Data set introduction.
(Photo credit: Original)

3. Deep learning classification

The input image is extracted through a series of convolutional layers to extract the feature information in the image. The ResNet-152 model uses a deep convolutional neural network, which contains multiple convolutional layers, each of which extracts different feature information.

The feature image is passed into multiple residual blocks, each containing two convolution layers and a skip connection. Jump connections can solve the problem of gradient disappearance in deep neural networks, so that the model can learn the feature information better.

The output feature map of the last residual block is passed into the global averaging pooling layer, and the values of each channel of the feature map are averaged to obtain a feature vector. The global average pooling layer can reduce the dimension of the spatial information of the feature graph, reduce the number of parameters of the model, and improve the generalization ability of the model.

Fully connected layer: The feature vector is passed into the fully connected layer to get the score value for each category. The fully connected layer usually contains multiple neurons, each corresponding to a class, and the score value for that class is calculated.

The score value of the full connection layer is passed through Softmax function to convert the score value into class probability. The Softmax function can convert the score value to the probability value, making the results of the model output more intuitive.

Output result: Output the predicted category and its probability value. The output result of the model usually includes the predicted category and the probability value of that category, which can be sorted according to the probability value to get the predicted result.

Figure 3. Forecast result.
(Photo credit: Original)

Table 1. Model evaluation parameter.

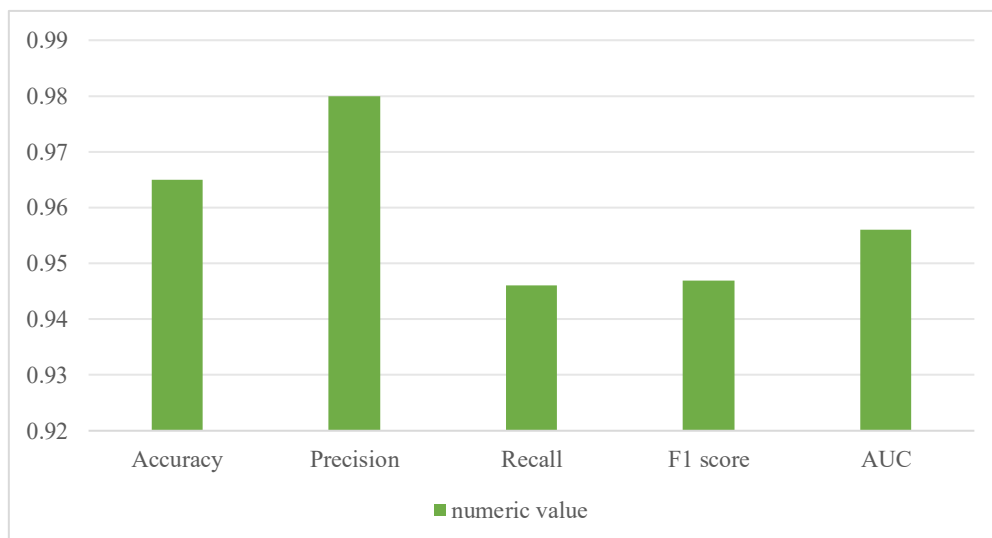| Accuracy	Precision	Recall	F1 score	AUC
0.965	0.980	0.946	0.947	0.956

Figure 4. Model evaluation parameter.
(Photo credit: Original)

According to the above results, the classification accuracy of the model reaches 96.5%, the accuracy reaches 98.0%, the recall rate is 94.6%, the f1 score is 94.7%, and the AUC reaches 95.6%.

4. Conclusion.

Resnet-152 is a deep convolutional neural network model, which is one of the ResNet series. It has a very deep network structure, which can solve the problem of gradient disappearance and gradient explosion in deep neural networks well, thus improving the performance of the model. On BIRDS 525 SPECIES-IMAGE CLASSIFICATION dataset, the classification based on ResNet-152 model has achieved excellent results, with classification accuracy up to 96.5%, accuracy up to 98.0% and recall rate up to 94.6%. f1 scored 94.7% and AUC achieved 95.6%.

In terms of model structure, ResNet-152 model is a very deep convolutional neural network model, which is composed of multiple convolutional layers and pooling layers, including residual blocks. The structure of ResNet-152 model is very complex, but it can handle the task of bird image classification well and has good generalization ability. Residuals in the ResNet-152 model can effectively improve the performance of the model, making it easier to train and optimize. ResNet-152 model has a very deep network structure, which can solve the problem of gradient disappearance and gradient explosion in deep neural networks, and improve the performance of the model.

In terms of the performance of the model, BIRDS 525 SPECIES-IMAGE CLASSIFICATION has achieved excellent results. The classification accuracy of the model reached 96.5%, the accuracy reached 98.0%, the recall rate was 94.6%, the f1 score was 94.7%, and the AUC reached 95.6%. These indicators show that the ResNet-152 model can identify bird images well, the prediction results are very accurate, can capture positive samples well, while striking a good balance between accuracy and recall rate, and can distinguish positive and negative samples well.

ResNet-152 model has a very deep network structure, which can solve the problem of gradient disappearance and gradient explosion in deep neural networks, and improve the performance of the model. Residuals in ResNet-152 model can effectively improve the performance of the model and make it easier to train and optimize. The ResNet-152 model performs very well on the dataset BIRDS 525 SPECIES-IMAGE CLASSIFICATION, with very high classification precision, accuracy, recall rate, f1 score and AUC. The ResNet-152 model has good generalization ability and can handle various bird image classification tasks well.

The ResNet-152 model also has some limitations. For example, the ResNet-152 model has a very deep network structure, which requires large computing resources and time to train and optimize the model. Secondly, ResNet-152 model is prone to overfitting in the training process, and some regularization methods are needed to alleviate this problem. Finally, the ResNet-152 model has a very complex network structure, and it is difficult to explain the decision-making process and feature extraction process of the model.

In conclusion, ResNet-152 model is a very good deep neural network model, which can handle the bird image classification task well. It has a very deep network structure and residuals, which can solve the problem of gradient disappearance and gradient explosion in deep neural networks, and improve the performance of the model. On the dataset BIRDS 525 SPECIES-IMAGE CLASSIFICATION, the ResNet-152 model is used to classify birds, and very good results are obtained, with high classification precision, accuracy, recall rate, f1 score and AUC. However, the ResNet-152 model also has some limitations, which need to be selected and optimized according to the specific situation in practical application.

References
[1] Thien-Nu H ,Daehee K . Supervised contrastive ResNet and transfer learning for the in-vehicle intrusion detection system[J]. Expert Systems With Applications,2024,238(PE).
[2] Serena S ,Ashish S ,Sreeram V P , et al. A refined ResNet18 architecture with Swish activation function for Diabetic Retinopathy classification[J]. Biomedical Signal Processing and Control,2024,88(PA).

[3]   Enrico C ,Alessio S ,Matteo L , et al. A Robust Initialization of Residual Blocks for Effective ResNet Training Without Batch Normalization.[J]. IEEE transactions on neural networks and learning systems,2023,PP.

[4]   Enrico C ,Alessio S ,Matteo L , et al. A Robust Initialization of Residual Blocks for Effective ResNet Training Without Batch Normalization.[J]. IEEE transactions on neural networks and learning systems,2023,PP.

[5]   A. V ,S. S . Resnet-Unet-FSOA based cranial nerve segmentation and medial axis extraction using MRI images[J]. The Imaging Science Journal,2023,71(8).

[6]   Esraa H ,Shamim M H ,Abeer S , et al. A quantum convolutional network and ResNet (50)-based classification architecture for the MNIST medical dataset[J]. Biomedical Signal Processing and Control,2024,87(PB).

[7]   Wiku L K ,Pingan W ,Hyun-Ho N , et al. Airborne hyperspectral imaging for early diagnosis of kimchi cabbage downy mildew using 3D-ResNet and leaf segmentation[J]. Computers and Electronics in Agriculture,2023,214.

[8]   Tiejun C ,SungJune B . Library-Based Raman Spectral Identification Using Multi-Input Hybrid ResNet.[J]. ACS omega,2023,8(40).

[9]   Li X ,Xu X ,He X , et al. Intelligent Crack Detection Method Based on GM-ResNet[J]. Sensors,2023,23(20).

[10]   Asel S ,Mo K ,Andrii K , et al. Hybrid quantum ResNet for car classification and its hyperparameter optimization[J]. Quantum Machine Intelligence,2023,5(2).