Comparison Of Conventional and Lightweight Convolutional Neural Networks for Image Classification

Yulin Xue*

School of information Engineering, Chang'an University, Xi'an, Shannxi, 710000, China

* Corresponding Author Email: 2020904572@chd.edu.cn

Abstract. With the popularity of artificial intelligence, the use of deep learning is more and more extensive, but devices often need good performance to complete this task. However, for mobile devices and some devices with relatively poor performance, deep learning seems to have many obstacles. So, this article explores lightweight neural networks. I used two datasets to explore the problem of using lightweight neural networks to classify images. One is the issue of binary classification, which uses gender data sets. The other is the multi-classification problem, which uses bird data sets. I compared three models, one of which is a relatively complex neural network, and the other two are respectively MobileNet and ShuffleNet.In the first dataset, that is, the binary classification problem, the relatively complex neural network performs well, and the accuracy reaches 0.93. However, the accuracy of MobileNet with pre-training parameters and ShuffleNet without pre-training parameters also reached 0.86 and 0.84 respectively. Although the accuracy is reduced, it also shows that the lightweight neural network can well complete this problem. In the multi-classification problem, the relatively complex network performance has been greatly reduced. The accuracy is only 0.76, which may be due to the limited ability of standard convolution to process information. The accuracy of MobileNet with pre-training parameters is 0.87, and the accuracy of ShuffleNet without pre-training parameters is 0.62. So, the conclusion is that using lightweight neural networks should pay attention to the complexity of the problem, whether to use pre-training parameters and the selection of parameters.

Keywords: Deep learning; light weighted net; gender classification; bird classification.

1. Introduction

Nowadays deep learning is more and more popular [1, 2]. Neural network is a very important algorithm for deep learning, which can solve the problem of image classification well. Image classification is a very essential problem, which can be applied in many scenes, such as transportation and scientific research [3, 4]. Deep learning requires a lot of computation and high-performance graphics cards. In that case, Deep learning is hard to achieve. However, some lightweight neural networks can well solve the problems of large computation and high requirements for equipment [5, 6]. So, the lightweight neural networks could be explored. The lightweight neural networks are leveraged to compare relatively complex neural networks and explore the impact on the classification effect while reducing the amount of computation and parameters.

This research first compared three kinds of neural networks on the issue of gender classification. Two of them are lightweight models, and one is a relatively complex neural network. Gender classification is a classical binary classification. Gender classification is a good choice to compare neural networks. Because binary classification is not a complex classification problem, it can well show the characteristics of each model.

In order to make a more comprehensive comparison, in addition to gender classification, three models are compared on the data set of bird classification. Because this is a relatively complex classification problem, the performances of the models on complex problems could be compared. In this experiment, only ten kinds of birds were selected for classification.

In this experiment, the models with different complexity show great differences. The parameters of the three models are about 7 million, 2 million, and 1 million respectively. Because of the difference in parameter quantity, the training time of the three models is quite different. The relatively complex model needs a lot of time to train compared with the lightweight model. In simple classification problems, relatively complex models have higher accuracy. But on complex problems, there is little difference between the accuracy of light neural networks and that of relatively complex neural networks. The specific reasons will be discussed later.

2. Method

2.1. Datasets

Two datasets are leveraged to validate the effectiveness of the models. The former is a gender dataset and the latter is bird dataset.

The gender dataset is collected on Kaggle [7]. Gender datasets include 8000 male pictures and 8000 female pictures. The dataset is spited, where 80 percent for training data, and 20 percent for testing dataset. The training dataset consists 12800 pictures. The testing dataset consists 3200 pictures. Each picture has three channels, but the height and width of each picture are different.

The bird dataset is collected on Kaggle [8]. 10 kinds of birds in this dataset are selected. In the training set, there are about 150 pictures of birds of each category, and 30 pictures in the test set. The size of picture is 224*224*3.

2.2. Data pre-processing

In the image preprocessing stage, only image normalization and standardization are used. Normalization avoids some parameters being too large, otherwise it may lead to classification errors in the model and other problems, leading to experimental failure. In addition, normalization and standardization can reduce the time for convergence of the model without changing the data sequence. Because this experiment mainly focuses on the light neural network, no other preprocessing operations are carried out.

2.3. Models

2.3.1 Proposed model

This neural network based on Keras. Keras is easy to operate and is a good choice. This model five convolution layers, four pooling layers, one flattens layers and two dense layers. The model has a total of 5 million parameters.

2.3.2 MobileNet

MobileNet [9] is a lightweight neural network. Different from the first model, MobileNet has only about 2 million parameters. In addition, the training time of MobileNet has also decreased significantly. MobileNet adopts the convolution method of depth separable convolution. It is efficient to reduce the parameters. This form of convolution splits the conventional convolution into a depth convolution. This design includes two technologies, the former is depthwise convolution and the latter is pointwise convolution. The difference between depthwise convolution and standard convolution is that the kernel of it is a single channel mode, and each input channel needs to be convolved, so that the output feature map has the same number of channels as the input. However, the number of output feature maps could be too small, which may affect the effectiveness of information. Pointwise convolution uses 1 * 1 convolution kernel to raise dimensions.

In order to verify the reduction of calculational consumption, the next step is to compare the parameter amount with the calculation amount. Assume that the kernel size of the depthwise separable convolution size is Dk * Dk, with M channels, so the parameter quantity is Dk * Dk * M. The point convolution is 1 * 1 * M, and the number is N, so there are M*N parameters. Dk * Dk * M+M*N is the total number of parameters.

Parameter quantity ratio: Dk * Dk * M+M*N/Dk*Dk*M*N=1/N+1/Dk*Dk.

MobileNet sacrifices very little accuracy but reduces a lot of computation.

2.3.2 ShuffleNet

ShuffleNet [10] is light weighted designed for devices with finite computer performance. The purpose of the designer to design ShuffleNet is to reduce the computational consumption while keeping the accuracy as much as possible. To achieve this goal, the designer mainly adopts two technologies, namely pointwise group evolution and channel shuffle. Point by point convolution refers to the use of 1 * 1 * 1 convolution kernel. However, this convolution method will lead to high computational complexity, because each pixel needs to be calculated. The convolution layer is replaced by group convolution to reduce computing resources. Group convolution also has its own disadvantages. Like inbreeding in genetics, there is no connection between each group of information that is detrimental to the results. In order to exchange the information of each group, the designer adopts the channel shuffle. It is equivalent that each group of information is composed of all groups. The results also demonstrate that channel shuffle can improve the classification performance.

2.4. Evaluation matrix

In this paper, the accuracy, precision, recall, f1-score, and confusion matrix are used to measure the effectiveness of the models.

Accuracy represents the percentage of correctly classified samples. Using examples of binary classifications to introduce. First introduce some important values, which will be used in the follow data is correctly classified in all samples, false negative (FN), false positive (FP), true negative (TN), and true positive (TN). However, such evaluation indicators are often incomplete. The analysis of the model cannot be comprehensive, so more indicators are needed to analyze the model. The calculation formula of accuracy is accuracy=(TP+TN)/(TP+TN+FP+FN).

Precision refers to how many data in this category really belong to this category. Precision is applicable to detecting all required categories correctly as far as possible, regardless of whether all data of the category has been detected. Taking gender classification for example, high precision means that most of the gender classification is correct, but it does not mean that most of the men and women are classified into the correct category. The calculation formula of precision is precision=TP/(TP+FP)

Recall is the proportion of correctly classified samples in the total number of samples. Also taking gender classification as an example. If there are 100 males and 70 of them are classified as males, then recall is 0.7. The scenarios that Recall applied are often different from those of precision. Recall and precision are contradictory. Usually, if the recall is high, the precision will decrease. If the precision is high, the recall will decrease. The recall is recall=TP/(TP+FN).

Since precision and recall are contradictory to some extent, they need to be considered comprehensively. One method is f1 score. It is the harmonic average of recall and precision. A high f1 score indicates the model works well. The calculation formula of precision is f1-score=2*precision*recall/(precision+recall).

Confusion matrix is a very intuitive and effective method to measure model classification. The confusion matrix is to put the values of correct classification and wrong classification in a table. This table is the confusion matrix.

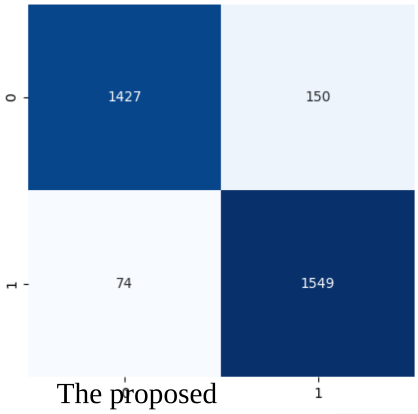
3. Result

3.1. Gender classification

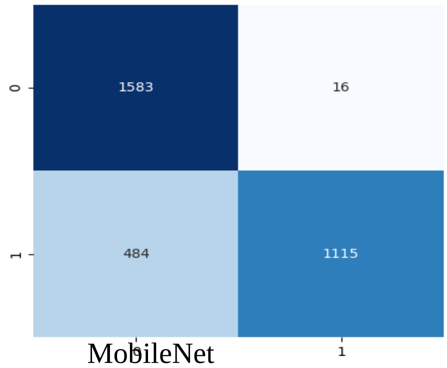
In the first experiment, to ensure the comparison of results is meaningful. I set the epochs and batches to the same number. MobileNet uses pre-training parameters, while the first model and ShufflNet do not use pre-training parameters. The first model takes the longest time. Each round takes eight minutes, MobileNet takes six and a half minutes, and ShuffleNet takes six minutes. The performance of each evaluation index of the first model is the best, while ShuffleNet is the lowest as shown in Table 1. Each evaluation index of MobileNet is between the proposed model and ShuffletNet.

Table 1. Classification results on gender dataset.

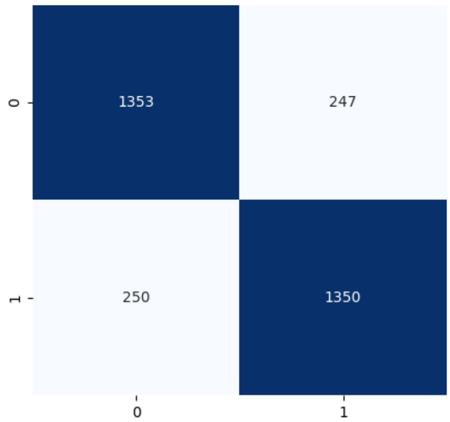
	Accuracy	Recall	Precision	F1-score
Proposed	**0.93**	**0.93**	**0.93**	**0.93**
MobileNet	0.86	0.86	0.88	0.86
ShuffleNet	0.84	0.84	0.84	0.84



The proposed



MobileNet



ShuffleNet

Fig. 1 Confusion matrix results of the three nets on gender dataset.

For the proposed network, the classification of the two categories in model one is relatively average. As for the MobileNet, through the confusion matrix, the MobileNet has a good classification effect for men, but the probability of wrong classification for women is very high. The recall of male reached 0.98, and the precision was only 0.79. The opposite is true for female. As for the ShuffleNet, although the overall evaluation index value of ShuffleNet is not high, the classification of each category is relatively average. Precision is basically equal to recall.

3.2. Bird classification

Table 2. Classification results of the proposed model and light weighted ones.

	Accuracy	Recall	Precision	F1-score
Proposed	0.76	0.78	0.78	0.77
MobileNet	**0.87**	**0.83**	**0.83**	**0.82**
ShuffleNet	0.62	0.62	0.63	0.62

For bird classification, first, the training time of each epoch of each model is introduced. The proposed model requires one and a half minutes, while MobileNet and ShuffleNet require one and a half minutes and fifty-five seconds, respectively. In the bird dataset, some changes have taken place in the performance of each model. In the first model, no matter what the indicators are, there has been a significant decline as shown in Table 2. For example, accuracy is only 0.76, and f1 score is only 0.77. MobileNet seems to adapt well to the changes in the data set, and each evaluation indicator has little change, which can play a good role in classification. Accuracy reached 0.87, and f1 score also reached 0.82. Each index of ShuffleNet is about 0.62. Next, the performance of each model is discussed.

4. Discussion

4.1. Gender classification

In gender classification, although Model 1 training takes the longest time, its performance is the best. In the training process, compared with the other two lightweight networks, the convergence speed of the first model is obviously faster. Compared with ShuffleNet, these two models do not use pre-training parameters, but model 1 will perform much better. Although the convergence speed of MobileNet is slow, it can also achieve good results due to the use of pre-training parameters. However, MobileNet has an obvious disadvantage in this dataset, that is, the accuracy and recall rate of female classification are relatively low. An important reason may be that the matching degree of pre-training parameters to gender dataset may be slightly poor. As the gender data set is a binary classification problem, it is not that complex. Although ShuffleNet does not use the pre-training parameters, it is only 0.01 lower than the indicators of MobileNet. It also shows that the improvement of shuffle algorithm is successful. With fewer calculations and fewer parameters, the same effect can be achieved or even better. Compared with MobileNet, ShuffleNet can converge faster in less time.

4.2. Bird classification

In the bird dataset, the performance of model one has dropped a lot. Because data sets become more complex, classification becomes more difficult. Once the model adopts the standard convolution method, it does not do any processing on the collected information, so the classification effect is poor. As reflected in the classification index, each value decreased by 0.15 on average. For mobile net, due to the use of pre-training parameters, the model classifiy well at the first epoch. Although the convergence speed is slow, it can still get good results. ShuffleNet does not use the pre training parameters to solve this problem, and can only improve the accuracy by adding more training rounds. Therefore, for lightweight networks, if the pre training parameters are not used, the classification effect is often not good in limited training rounds. Therefore, in order to obtain a better classification effect, lightweight networks should pay attention to the selection of training rounds and parameters. If the training round is too large, the training time will be too long and the advantage of less calculation will be lost. The training round is too small to obtain good classification effect. Or access the pre-training parameters in the device to achieve better classification results in less time.

5. Conclusion

This experiment mainly explores the advantages and disadvantages of lightweight neural networks in image classification. For some devices with limited computing power, the problem of lightweight network classification is very critical. This work explored this issue with two datasets. They are gender data set and bird data set. The reason for choosing these two data sets is that the author wants to compare lightweight networks from a binary problem and a relatively complex problem. This experiment realized the training process, advantages and disadvantages, and some use skills of lightweight networks. Lightweight networks can solve some simple problems well, such as gender classification in this experiment. No matter what kind of lightweight neural network it is, no matter whether the pre training parameters are used or not, it has achieved good classification results. For relatively complex classification problems, lightweight neural networks do not seem to have such good classification results. By observing the training process of each step, it is found that the lightweight network converges slowly. For a complex classification problem, if there is no good classification result in the first step, more time is needed to train the whole model. This is also the reason why lightweight networks have difficulty in dealing with relatively complex problems. Therefore, when using lightweight neural networks, researchers should first pay attention to the complexity of the problem. If it is a simple problem, they do not have to pay much attention to too many problems when using lightweight networks. If it is a complex classification problem, selecting pre training parameters is a good solution. It can solve the problem of slow convergence. The subsequent exploration of lightweight networks should not stop. In the future, the number of parameters and calculation could be reduced while considering the classification effect of the model to achieve a better balance.

References

[1] Abiodun O I, Jantan A, Omolara A E, et al. State-of-the-art in artificial neural network applications: A survey. Heliyon, 2018, 4(11): e00938.

[2] LeCun Y, Bengio Y, Hinton G. Deep learning. nature, 2015, 521(7553): 436-444.

[3] Lv Y, Duan Y, Kang W, et al. Traffic flow prediction with big data: a deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 2014, 16(2): 865-873.

[4] Reichstein M, Camps-Valls G, Stevens B, et al. Deep learning and process understanding for data-driven Earth system science. Nature, 2019, 566(7743): 195-204.

[5] Zhou Y, Chen S, Wang Y, et al. Review of research on lightweight convolutional neural networks. 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2020: 1713-1720.

[6] Kravchik M, Shabtai A. Efficient cyber-attack detection in industrial control systems using lightweight neural networks and pca. IEEE Transactions on Dependable and Secure Computing, 2021.

[7] Ashutosh C. Gender Classification Dataset. URL: https://www.kaggle.com/datasets/cashutosh/gender-classification-dataset. 2019.

[8] Gerry. Birds 450 Species Image Classification. URL: https://www.kaggle.com/datasets/gpiosenka/100-bird-species. 2022.

[9] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

[10] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.