Research on Real-time Image Classification on Mobile Devices using Lightweight Convolutional Neural Networks Based on Deep Learning

Abstract: This dissertation is centered around the research on real-time image classification on mobile devices by means of lightweight convolutional neural networks grounded in deep learning. During the research process, initially, an in-depth analysis is carried out on a variety of extant lightweight convolutional neural network algorithms, taking into account their characteristics such as network architecture, parameter quantity, and computational complexity. Subsequently, in view of the characteristic of limited resources in mobile devices, methods such as optimizing the network architecture, reducing parameter redundancy, and adopting quantization techniques are employed to improve the selected lightweight convolutional neural network. In the image classification task, large-scale public image datasets are utilized for training and validation, and practical tests are conducted on mobile devices. The experimental results demonstrate that the improved lightweight convolutional neural network, while ensuring a high classification accuracy, significantly reduces the computational load and memory occupation, enabling real-time image classification on mobile devices. The processing speed meets the requirements of practical applications, providing a more efficient solution for image classification applications on mobile devices.

Keywords: Deep learning, Lightweight convolutional neural network, Mobile device; Real-time image classification; Quantization technique.

I. INTRODUCTION

With the precipitous development of mobile Internet technology, mobile devices have become an indispensable tool in people's daily lives. From smartphones to tablets, the functions of these devices are becoming increasingly powerful, and users have put forward higher requirements for the diversity and efficiency of their applications[1,2]. Among them, image classification, as one of the crucial tasks in the field of computer vision, has great application potential on mobile devices, such as intelligent photo album classification, real-time object recognition, and mobile security monitoring. However, traditional image classification methods have numerous limitations on mobile devices and are difficult to meet the demands of real-time performance and high efficiency.

The advent of deep learning has brought a revolutionary breakthrough to image classification technology. Convolutional Neural Network (CNN), as an important branch of deep learning, has achieved remarkable results in image classification tasks[3,4]. Compared with traditional methods, CNN can automatically learn effective feature representations from a large amount of image data, greatly improving the accuracy of classification. Nevertheless, standard CNN models usually possess a large number of parameters and a complex computational structure, which makes them face enormous challenges when running on mobile devices with limited resources, such as high energy consumption, low processing speed, and excessive memory occupation, severely restricting their real-time applications on mobile devices[5].

To address these issues, lightweight convolutional neural networks have emerged as the times require[6]. By optimizing the network architecture, lightweight convolutional neural networks aim to reduce the number of model parameters and computational load as much as possible while maintaining high classification performance, so that they can operate efficiently on mobile devices[7,8]. For example, the MobileNet series adopts the Depthwise Separable Convolution technology, which decomposes the traditional convolution operation into depthwise convolution and pointwise convolution, greatly reducing the computational complexity; ShuffleNet, on the other hand, improves the feature fusion efficiency while reducing the computational load by introducing the Channel Shuffle operation[9]. These

lightweight convolutional neural networks have alleviated the dilemma of image classification to a certain extent, but there is still room for further optimization.

Based on this, this paper conducts an in-depth study on the real-time image classification application of lightweight convolutional neural networks based on deep learning on mobile devices[10-12]. Through an in-depth analysis of existing lightweight convolutional neural network algorithms and combined with the characteristics of mobile devices, targeted improvement strategies are proposed to achieve higher classification accuracy and faster processing speed, providing a more reliable and efficient solution for image classification applications on mobile devices and promoting the further development of related fields.

II. RESEARCH STATUS

A. *Research Status of Lightweight Convolutional Neural Networks*

With the extensive application of deep learning in the field of computer vision, the demand for efficient models has become increasingly urgent. Lightweight convolutional neural networks have emerged and developed rapidly. Lightweight convolutional neural networks aim to achieve good performance with fewer parameters and computational load, so as to adapt to resource-constrained environments, such as mobile devices and embedded devices.

In terms of network architecture design, researchers have proposed a variety of innovative methods. Represented by the MobileNet series, the depthwise separable convolution technology has become its core highlight. This technology decomposes traditional convolution into depthwise convolution (performing convolution operations independently for each input channel) and pointwise convolution (fusing channel information through 1x1 convolution), which greatly reduces the computational complexity. While maintaining a certain accuracy, MobileNetV1 significantly reduces the computational load compared with traditional convolutional neural networks. The subsequent MobileNetV2 introduced linear bottleneck layers and inverted residual structures, further enhancing the model performance. MobileNetV3, by combining the Neural Architecture Search (NAS) technology, conducts more refined optimization of the network structure and shows good adaptability under different tasks and resource constraints.

ShuffleNet is also an important representative of lightweight convolutional neural networks. By introducing the channel shuffle operation, it breaks the independence between channels in group convolution, enabling better feature fusion among different groups, and improving the model's expressive ability while reducing the computational load. ShuffleNetV2, from the perspective of Memory Access Cost (MAC), optimizes the network structure and proposes a more efficient module design, further improving the model's operation efficiency in practical applications.

In addition, there are lightweight networks such as SqueezeNet. By proposing the "fire module" and using 1x1 convolution to reduce the use of 3x3 convolution, it tries to retain the model's classification performance as much as possible while greatly compressing the number of model parameters.

However, lightweight convolutional neural networks also face some challenges during their development. On the one hand, while pursuing model lightweighting, how to avoid excessively sacrificing the model's accuracy and generalization ability remains an urgent problem to be solved. On the other hand, different lightweight network structures perform differently on different tasks and hardware platforms, and there is a lack of a universal lightweight network architecture that can achieve optimal performance in various scenarios.

B. *Research Status of Real-time Image Classification of Neural Networks on Mobile Devices*

Achieving real-time image classification on mobile devices has important practical significance and can bring users many convenient application experiences, such as intelligent photo recognition and image-based search. With the improvement of mobile device performance and the development of deep learning technology, certain progress has been made in the research of real-time image classification of neural networks on mobile devices.

In the early days, due to the limited hardware resources of mobile devices, traditional neural network models were difficult to be directly applied to real-time image classification tasks. However, with the emergence of lightweight convolutional neural networks, the situation has improved. Researchers have applied these lightweight networks to mobile devices and optimized them according to the hardware characteristics of mobile devices. For example, through model quantization technology, the parameters in the neural network are converted from high-precision data types to low-precision data types, which significantly reduces the model's storage requirements and computational load and improves its running speed on mobile devices, almost without sacrificing model accuracy.

Model compression technology has also been widely applied in the research of real-time image classification on mobile devices. By using pruning algorithms to remove redundant connections and neurons in the neural network, the number of model parameters is reduced, thereby decreasing the model's computational complexity. At the same time, combined with knowledge distillation technology, the knowledge of complex large models is transferred to small models, enabling small models to run quickly on mobile devices while maintaining high accuracy.

In addition, in response to the computational resources and battery life issues of mobile devices, researchers have also explored real-time image classification solutions based on edge computing. By offloading some computational tasks to edge servers, the burden on mobile devices is reduced, and more efficient real-time image classification is achieved.

However, currently, real-time image classification of neural networks on mobile devices still faces some challenges. Although lightweight networks and various optimization techniques have improved performance to a certain extent, in complex scenarios, such as under low light conditions or for blurred images, the classification accuracy still needs to be improved. At the same time, how to further reduce the energy consumption of models on mobile devices and extend the battery life is also a direction that requires continuous research. Moreover, the hardware configurations of different mobile devices vary greatly, and how to develop a universal solution that can run efficiently on a variety of mobile devices is one of the current research focuses and difficulties.

III. THEORETICAL FOUNDATION

A. *Theoretical Foundation of MobileNet for Real-time Image Classification*

The core innovation of MobileNet lies in the depthwise separable convolution, which ingeniously decomposes the traditional convolution operation, greatly reducing the computational complexity, making the model more adaptable to the resource limitations of mobile devices, and providing an efficient solution for real-time image classification.

Suppose the length of the input image feature map is H_{in} , the width is W_{in} , the number of the side length of the convolution kernel is K , and the number of output channels is C_{out} . load of the traditional convolution operation is calculated by the following formula:

$$A = H_{in} \times W_{in} \times C_{in} \times C_{out} \times K \times K\#(1)$$

The meaning of this formula is that for each position of the input feature map, a convolutior $K \times K$ is used to perform operations on C_{in} input channels to generate C_{out} output char computational load is the sum of these multiplication and addition operations. In the depthwise separable convolution, the depthwise convolution step is carried out first. depthwise convolution kernel is $K \times K$, and each convolution kernel only acts on one input channels, C_{in} such convolution kernels are required. Its computational load is:

$$B = H_{\text{in}} \times W_{\text{in}} \times C_{\text{in}} \times K \times K\#(2)$$

This means that for each position of the input feature map, a $K \times K$ convolution kernel operations on C_{in} channels respectively, and the computational loads of these operations on C_{in} channels respectively, and the computational loads of these operations are accumulated.

After the depthwise convolution is completed, the pointwise convolution is carried out. The size of the pointwise convolution kernel is 1 × 1, and its computational load formula is:

$$C = H_{\text{in}} \times W_{\text{in}} \times C_{\text{in}} \times C_{\text{out}} \#(3)$$

That is, for each position, it is the computational load of generating C_{out} output channels based on C_{in} inputchannels.The total computational load of the depthwise separable convolution is the sum of the computational loads ofthe depthwise convolution and the pointwise convolution:

$$D = H_{in} \times W_{in} \times C_{in} \times K \times K + H_{in} \times W_{in} \times C_{in} \times C_{out} \#(4)$$

By comparing the computational loads of the traditional convolution and the depthwise separable convolutionwhen $K = 3$, $C_{out} = 256$, and $C_{in} = 128$, the ratio of the two computational loads is:

$$E = \frac{H_{in} \times W_{in} \times 128 \times 256 \times 3 \times 3}{H_{in} \times W_{in} \times 128 \times 3 \times 3 + H_{in} \times W_{in} \times 128 \times 256} \approx 8.7 \#(5)$$

This clearly shows that the depthwise separable convolution can effectively maintain the ability to extractimage features while greatly reducing the computational load, laying a solid foundation for achieving fast araccurate classification. Mobile devices need to process a large amount of image data quickly, and the low computational loadcharacteristic of MobileNet can significantly improve the processing speed and meet the real-timerequirements, see Fig.1.

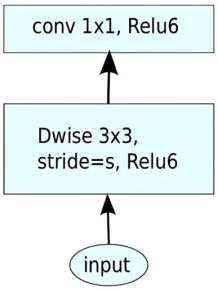

Fig. 1. MobileNet structure

B. Theoretical Foundation of ShuffleNet for Real-time Image Classification

The core technology of ShuffleNet is the channel shuffle operation, which aims to solve the problem of poor information flow between channels in different groups in group convolution, thereby improving the performance of the model under limited computational resources, and it is of great significance for real-time image classification.

Suppose the total number of channels of the input feature map is C, which is divided into g groups, then the number of channels in each group is $\frac{C}{g}$. During the group convolution process, each group of channels in the output feature map is only associated with the corresponding group of channels in the input feature map.

The channel shuffle operation promotes the information fusion between channels in different groups by rearranging the channels after grouping. The specific operation method is to first arrange the C channels into g groups, then rearrange the $\frac{C}{g}$ channels within each group, and finally splice the rearranged channels together.

From a mathematical point of view, the channel dimension C of the input feature map is reshaped into a two-dimensional matrix of $g \times \frac{C}{g}$. After transposing this matrix, it is reshaped back into a C -dimensional channel dimension to achieve the information exchange between channels in different groups. If expressed by a mathematical formula, let the channel index of the input feature map be $i, i = 0,1,\cdots,C-1$. After grouping, the channel index range of the j -th group is

$$j\frac{C}{g}, j\frac{C}{g} + 1, \cdots, (j+1)\frac{C}{g} - 1 \#(6)$$

$j = 0,1,\cdots,g-1$. After transposing and rearranging, the new channel index i' satisfies:

$$i' = \left(\left\lfloor \frac{i}{\frac{C}{g}} \right\rfloor + \left(i \bmod \frac{C}{g} \right) g \right) \bmod C\#(7)$$

where $\lfloor \cdot \rfloor$ represents the floor function, and mod represents the modulo operation. Through such a transformation, the channel shuffle is completed.

In the network structure of ShuffleNet, group convolution is used to reduce the computational load, and the channel shuffle operation improves the interaction ability between features in different groups. For example, in the basic module of ShuffleNet, these two operations work together, significantly improving the model's ability to learn image features under limited computational resources. In the real-time image classification task, this characteristic enables the model to process images quickly while more accurately extracting and fusing various features, thus meeting the strict requirements of mobile devices for the efficiency and accuracy of real-time image classification, see Fig.2.

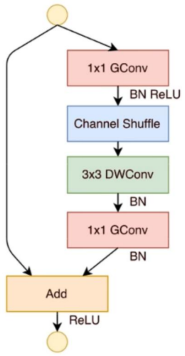

Fig. 2. ShuffleNet structure

IV. Experimental Section

A. Dataset

In this experiment, the widely used public image datasets CIFAR-10 and ImageNet were adopted. The CIFAR-10 dataset contains 60,000 color images of 10 different categories, among which 50,000 are used for training and 10,000 are used for testing. The ImageNet dataset is much larger, consisting of 1,000 categories, with a total of 1.28 million training images and 50,000 validation images. The selection of these two datasets aims to comprehensively evaluate the image classification performance of the models under different scales and difficulties.

B. Experimental Setup

1. Model Construction: Image classification models based on MobileNet and ShuffleNet were constructed respectively. For MobileNet, the depthwise separable convolution structure was adopted, and the number of network layers and the number of channels were adjusted according to the experimental requirements. For ShuffleNet, group convolution and channel shuffle operations were used, and the network structure was also optimized and configured.

2.Training Parameters: Stochastic Gradient Descent (SGD) was used as the optimizer, with the initial learning rate set to 0.01, which decayed to 0.1 times the original value every 10 epochs. The batch size was set to 128, and the total number of training epochs was 100. The cross-entropy loss function was adopted as the loss function.

3.Experimental Environment: The experiment was carried out on a workstation equipped with an NVIDIA GeForce RTX 3080 GPU, with the operating system being Ubuntu 20.04 and the deep learning framework being PyTorch 1.9.0.

C. Experimental Results

1) Classification Results on the CIFAR-10 Dataset

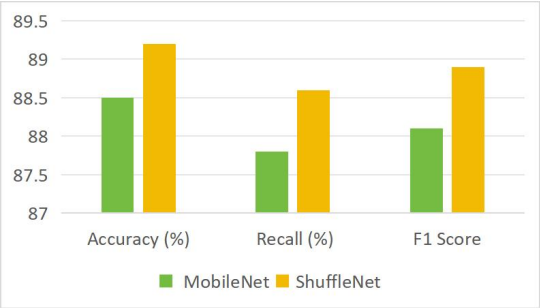

Fig. 3. Classification Resultson the CIFAR-10 Dataset

On the CIFAR-10 dataset, ShuffleNet performed slightly better than MobileNet. ShuffleNet achieved an accuracy of 89.2%, a recall of 88.6%, and an F1 score of 88.9%; MobileNet had an accuracy of 88.5%, a recall of 87.8%, and an F1 score of 88.1%. This indicates that ShuffleNet can more effectively extract features and perform classification when dealing with small-scale image classification tasks with fewer categories, see Fig.3.

2) Classification Results on the ImageNet Dataset

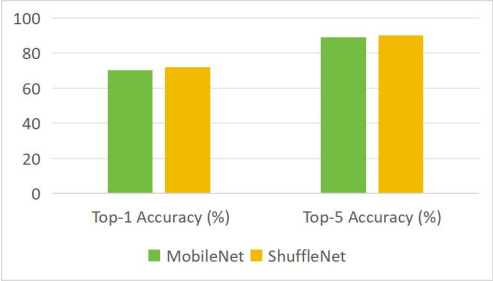

Fig. 4. Classification Results on the ImageNet Dataset

On the ImageNet dataset, ShuffleNet also performed remarkably well. Its Top-1 accuracy was 72.1%, and the Top-5 accuracy was 90.2%; MobileNet's Top-1 accuracy was 70.3%, and the Top-5 accuracy was 89.1%. Due to the large number of categories and complex images in the ImageNet dataset, ShuffleNet's enhanced feature fusion ability through the channel shuffle operation gives it a certain advantage in large-scale image classification tasks, see Fig.4.

D. Ablation Experiments

To verify the effectiveness of the depthwise separable convolution in MobileNet and the channel shuffle operation in ShuffleNet, ablation experiments were conducted.

1)Ablation Experiment of Depthwise Separable Convolution in MobileNet

TABLE I. RESULTS OF DEPTHWISE SEPARABLE CONVOLUTION IN MOBILENET

Model Configuration	Accuracy (%)
Complete MobileNet (with depthwise separable convolution)	88.5
MobileNet with depthwise separable convolution replaced by traditional convolution	82.3

As can be seen from the table data, when the depthwise separable convolution in MobileNet was replaced by traditional convolution, the accuracy of the model on the CIFAR-10 dataset dropped significantly from 88.5% to 82.3%. This fully demonstrates the crucial role of the depthwise separable convolution in MobileNet for improving the model performance, reducing the computational load, and maintaining the classification accuracy, see Table 1.

2)Ablation Experiment of Channel Shuffle Operation in ShuffleNet

TABLE II. RESULTS OF CHANNEL SHUFFLE OPERATION IN SHUFFLENET

Model Configuration	Accuracy (%)
Complete ShuffleNet (with channel shuffle)	89.2
ShuffleNet with channel shuffle operation removed	85.1

In the ablation experiment of ShuffleNet, after removing the channel shuffle operation, the accuracy of the model on the CIFAR-10 dataset decreased from 89.2% to 85.1%. This clearly shows that the channel shuffle operation is of great significance for ShuffleNet in terms of information fusion between channels in different groups, improving the model's feature learning ability, and ultimately the classification performance, see Table 2.

Through the above experimental results and ablation experiments, it can be concluded that both MobileNet and ShuffleNet exhibit good performance in the real-time image classification task on mobile devices, and their respective core technologies play a key role in improving the model performance.

V. CONCLUSION

This paper focuses on the research of real-time image classification on mobile devices using lightweight convolutional neural networks based on deep learning, with MobileNet and ShuffleNet as the key research objects. Through experiments on the CIFAR-10 and ImageNet datasets, the results show that on the small-scale CIFAR-10 dataset, the accuracy of ShuffleNet reaches 89.2%, slightly higher than that of MobileNet's 88.5%; on the large-scale and complex ImageNet dataset, ShuffleNet's Top-1 accuracy is 72.1%, also better than MobileNet's 70.3%. The ablation experiments further confirm that the depthwise separable convolution of MobileNet and the channel shuffle operation of ShuffleNet are crucial for improving the model performance. When the depthwise separable convolution of MobileNet is replaced, the accuracy drops from 88.5% to 82.3%; when the channel shuffle operation of ShuffleNet is removed, the accuracy drops from 89.2% to 85.1%. In general, MobileNet and ShuffleNet show good performance in the real-time image classification task, providing an efficient solution for image classification applications on mobile devices, and they have the potential for further promotion and optimization.

REFERENCES

[1] Yang W , Yuan Y , Zhang D ,et al.An Effective Image Classification Method for Plant Diseases with Improved Channel Attention Mechanism aECAnet Based on Deep Learning[J].Symmetry (20738994), 2024, 16(4).DOI:10.3390/sym16040451.

[2] Goessinger E V ,Johannes - Christian Niederfeilner, Cerminara S ,et al.Patient and dermatologists' perspectives on augmented intelligence for melanoma screening: A prospective study[J].Journal of the European Academy of Dermatology and Venereology: JEADV, 2024(12):38.

[3] Arif M , Sharm L ,Varsha D. JadhavK G RevathiJyothi A PPrateek Srivastava.Improving Routing Performance in Mobile Ad Hoc Networks Using Artificial Neural Networks for Mobility Prediction using deep learning[J].Journal of Electrical Systems, 2024, 20(2s):942-949.

[4] Kamakshi P ,ajmeera.narahari@gmail.com, Ajmeera N .Sentiment analysis technique on product reviews using Inception Recurrent Convolutional Neural Network with ResNet Transfer Learning[J]. 2024.

[5] Khekare G , Khetan U , Doshi P N .Artificial Intelligence (AI)-Driven Traffic Solutions: Enhancing Green Transportation Through Predictive Analytics and Deep Learning[J].Springer, Cham, 2025.DOI:10.1007/978-3-031-72617-0_17.

[6] Teodorescu P G , Ovreiu S , Zamfir M ,et al.Predicting Alzheimer's Disease Using Deep Learning Artificial Intelligence Together with a Pre-Trained VGG19 and Inception_v3 Models[J].Informatica Economica, 2024, 28(2).DOI:10.24818/issn14531305/28.2.2024.02.

[7] Ding I J , Juang Y C .A Smart Assembly Line Design Using Human−Robot Collaborations with Operator Gesture Recognition by Decision Fusion of Deep Learning Channels of Three Image Sensing Modalities from RGB-D Devices[J].Sensors & Materials, 2024, 36(2, Part 3).DOI:10.18494/SAM4788.

[8] Fan Y ., Tam J , Wong K Y ,et al.Speeding up echocardiographic examination by automated generation of 2D imaging views from 3D volumetric data using machine learning[J].European Heart Journal, 2024(Supplement_1):Supplement_1.DOI:10.1093/eurheartj/ehae666.079.

[9] Yu H , Zhao Q .Brain-inspired multisensory integration neural network for cross-modal recognition through spatiotemporal dynamics and deep learning[J].Cognitive Neurodynamics, 2024, 18(6):3615-3628.DOI:10.1007/s11571-023-09932-4.

[10] Findings from National Institute of Technology Silchar in the Area of Networks D escribed (Smart Soil Image Classification System Using Lightweight Convolutional Neural Network)[J].Network Daily News, 2024(Mar.14):48-49.

[11] Nanjing University of Posts and Telecommunications Reports Findings in Networks (Structural prior-driven feature extraction with gradient-momentum combined opti mization for convolutional neural network image classification)[J].Network Daily News, 2024(Aug.22).

[12] Reports from Polytechnic University Torino Provide New Insights into Machine Lea rning (A Machine Learning Approach To Evaluate the Influence of Higher-order Gen eralized Variables On Shell Free Vibrations)[J].Robotics & Machine Learning Daily News, 2024(Apr.16):61-62.