Enhanced Image Classification through Layer-Wise Feature Concatenation in Deep Neural Networks

Akshay Kumar Sharma, Kyung Ki Kim
Department of Electronics Engineering, Daegu University, South Korea
akshaysharma001@hotmail.com, kkkim@daegu.ac.kr

2024 21st International SoC Design Conference (ISOCC) | 979-8-3503-7708-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISOCC62682.2024.10762554

Abstract— **Image classification plays a pivotal role across various applications in sectors such as healthcare, agriculture, security, and surveillance. However, deploying large-scale models on resource-constrained edge devices remains a significant challenge. To address this issue, this paper introduces a novel, compact image classification framework that is specifically optimized for deployment on edge devices. The proposed model combines a simplified head block with four sequential stages, designed to balance accuracy with computational efficiency effectively. Evaluations on two datasets demonstrate the model's efficacy: it achieves an accuracy of 90.7% with only 0.11M parameters on the CIFAR-10 dataset, and 94.59% accuracy with 0.79M parameters on the Butterfly and Moth dataset. These results underscore the potential of the proposed framework as a practical solution for efficient, real-time image classification in environments where computational resources are limited.**

Keywords; Image classification, edge computing, deep learning, neural networks, computational efficiency, model optimization

I. INTRODUCTION

Image Classification (IC) is essential across sectors like security, agriculture, and healthcare, where it supports critical decision-making. Deploying large-scale IC models on resource-constrained edge devices poses significant challenges due to their limited resources. Lightweight image classifiers, which are computationally efficient and resource-conservative, have emerged as viable solutions for such environments. These classifiers are particularly suited for edge devices with restricted processing power and memory, and they facilitate effective resource management in systems with diverse devices through intelligent scheduling [1]. The importance of edge computing has grown with its applications in AI fields such as autonomous vehicles and augmented reality, which require minimal latency. By processing data closer to the source, edge computing enhances responsiveness and user experience, addressing latency issues associated with traditional cloud computing. This paper presents a novel Deep Neural Network (DNN) architecture tailored for edge devices. This compact yet powerful model is designed to provide accurate, real-time image classification without compromising on computational efficiency, bridging the gap between advanced IC capabilities and edge deployment constraints.

II. PROPOSED METHOD

The architecture of the proposed model is strategically designed to optimize computational efficiency while maintaining high classification accuracy. It consists of a head block followed by four distinct stages, each contributing uniquely to the model's overall performance.

A. Head Block

The head block serves as the initial processing layer, where key features are extracted from the input images. It is constructed using a 3x3 convolutional layer, which then splits the output into two separate pathways. The first pathway employs a Pointwise convolution followed by a Depthwise convolution, optimizing the layer's ability to filter and combine features with minimal parameter use. The second pathway incorporates a MaxPooling2D layer, which operates without downsampling the image, maintaining the original spatial dimensions. After processing, the outputs of both pathways are concatenated, followed by another Pointwise convolution. This design ensures a robust extraction of initial features while keeping the computational load manageable. The configuration of the head block is depicted in Fig. 1, which illustrates how the outputs from different layers converge and are integrated into subsequent stages.

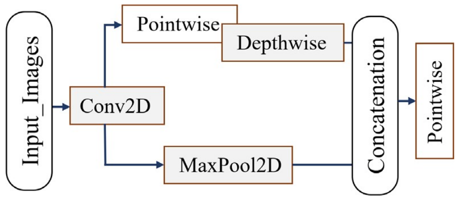

Figure 1. Head Block

B. Four Main Stages

Following the head block, the model includes four stages, each designed to further refine the feature maps extracted from the preceding layers. Unlike traditional approaches that heavily rely on computationally expensive 3x3 convolutions, our model utilizes a combination of Pointwise and Depthwise convolutions. This method significantly reduces the number of parameters involved while ensuring effective feature extraction and combination across the network.

Each of the four stages follows a consistent architectural theme but varies in the number of filters used. The stages start with fewer filters and gradually increase, allowing the network to process more complex features without a substantial increase in computational demand.

Figure 2 displays a typical stage within the model, highlighting the novel approach of concatenating feature maps from all previous layers. This technique facilitates the extraction of even the most subtle features from the input images, enhancing the model's ability to perform accurate classifications across varied datasets.

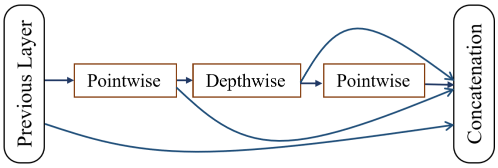

Figure 2. One Stage

The unique structure of the proposed model, combined with its innovative feature concatenation technique, provides a potent solution for image classification tasks, particularly when deployed on edge devices with limited computational resources. This methodological approach not only preserves the quality of image classification but also ensures that the model remains lightweight and efficient.

III. IMPLEMENTATION DETAILS & RESULTS

A. Datasets Used

The efficacy of the proposed model was evaluated using two distinct datasets, chosen for their diversity in image size and complexity:

① **CIFAR-10 Dataset**: This dataset is composed of 60,000 images distributed across 10 classes, with a resolution of 32x32x3 pixels. It serves as a benchmark for evaluating the model's performance in recognizing small-scale images. The testing phase utilized 10,000 images, with the remaining 50,000 employed for training purposes.

② **Butterfly and Moth Dataset [7]**: Encompassing a broader and more complex range of images, this dataset includes 12,597 images across 100 classes, each measuring 224x224x3 pixels. This dataset was selected to test the model's capability with high-resolution images, using 1,000 images for testing and the rest for training.

B. Performance Analysis

The proposed model was rigorously tested with three different activation functions to determine the optimal configuration
for both accuracy and computational efficiency:

- **Rectified Linear Unit (ReLU)**
- **Leaky Rectified Linear Unit (LeakyReLU)**
- **Exponential Linear Unit (ELU)**

Among these, ELU demonstrated superior performance, likely due to its smooth and differentiable nature across all input values, which aids in gradient-based optimization. Performance comparisons were drawn between the proposed model and conventional models to highlight its efficiency and effectiveness.

The proposed model demonstrates substantial efficiency, using only 111,706 parameters and achieving an accuracy of 90.7% on the CIFAR-10 dataset, which underscores its potential for deployment in resource-constrained environments as shown in Table 1. On the Butterfly and Moth dataset, the proposed model achieves a remarkable accuracy of 94.59% with significantly fewer parameters (0.79M), highlighting its capability to handle complex image classification tasks effectively and efficiently as shown in Table 2. These results validate the proposed model's robustness and adaptability, confirming its suitability for real-time applications on edge devices with limited computational resources.

Table1.Comparsion between Conventional and Proposed Model

Model	Parameters(M)	Accuracy(%)
MobileNet[2]	4.2	83.9
Lightweight[2]	4.1	84.2
LMFRNet[3]	0.52	94.60
Proposed Model	0.11	90.7

Table2.Comparsion between Conventional and Proposed Model

Model	Parameters(M)	Accuracy(%)
ResNet50[4]	33.62	94.29
EfficientNetB0[5]	10.32	94.06
MobileNetV2[6]	8.53	95.64
Proposed Model	0.79M	94.59

IV. CONCLUSION

This paper substantiates the efficacy of a lightweight deep neural network for image classification, optimized for deployment on resource-constrained edge devices. The proposed model leverages an innovative feature concatenation approach, resulting in significant parameter reduction while preserving high accuracy levels. Tested against the CIFAR-10 and Butterfly and Moth datasets, the model achieved accuracies of 90.7% and 94.59% respectively, with minimal computational overhead. These results not only demonstrate the model's robust performance across varied image scales but also highlight its potential for real-time applications in edge computing. The findings contribute valuable insights to the field of efficient AI, advocating for models that balance performance with computational efficiency to facilitate broader and more sustainable AI deployment.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-0-02052) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation). This work was partly supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2022-0-01013, Development of DRAM PIM semiconductor technology for enhanced computing function for edge, 50%).The EDA tool was supported by the IC Design Education Center.

REFERENCES

[1] B. Weinger, J. O. Kim, A. Sim, et al., “Enhancing IoT anomaly detection performance for federated learning”, Digital communications and networks, 2022.

[2] W. Sun, X. Zhang, and X. He, “Lightweight image classifier using dilated and depthwise separable convolutions”, J. Cloud Comp., Vol. 9, No. 1, pp. 1-12, 2020.

[3] G. Wan and L. Yao, “LMFRNet: A Lightweight Convolutional Neural Network Model for Image Analysis”, Electronics, 13, 2024.

[4] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition”, Computer Vision and Pattern Recognition, 2015.

[5] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, 2020.

[6] [M. Sandler, A. Howard, et al., “MobileNetV2: Inverted Residuals and Linear Bottlenecks”, Computer Vision and Pattern Recognition, 2018.

[7] Butterfly & Moths Image Classification 100 species. (Last Accessed on 10th May 2024) (https://www.kaggle.com/datasets/gpiosenka/butterfly- images40-species)