Image Classification and Analysis on the CIFAR-10 Dataset

Dhruba Nandi
Statistics Department
University of Michigan
Ann Arbor, MI, USA
dnandi@umich.edu

***Abstract*—The CIFAR-10 dataset is a widely used benchmark in computer vision, comprising 60,000 32x32 RGB images categorized into ten classes. This study evaluates Microsoft's ResNet-152 and Google's EfficientNet-B2 models to compare their effectiveness in image classification tasks on CIFAR-10. Using fine-tuning, the project explores the trade-offs between deep residual learning and compound scaling strategies in terms of classification accuracy, computational efficiency, and generalization ability. Results demonstrate that EfficientNet-B2 outperforms ResNet-152 in both accuracy and computational efficiency, providing valuable insights for model selection in small-scale image classification tasks.**

I. Introduction

The CIFAR-10 dataset is a cornerstone in computer vision research, comprising 60,000 32x32 RGB images evenly distributed across 10 classes. It has been widely adopted as a benchmark for evaluating novel architectures and algorithms. This study evaluates two state-of-the-art models, ResNet-152 and EfficientNet-B2, for CIFAR-10 classification, focusing on their trade-offs in accuracy, computational efficiency, and robustness.

Over the years, numerous advancements have been made in image classification using CIFAR-10 as a benchmark:

- **Baseline CNNs:** Krizhevsky and Hinton introduced the CIFAR-10 dataset along with simple convolutional neural networks (CNNs), establishing the baseline for classification accuracy.
- **ResNet:** He et al. proposed residual networks (ResNet), which introduced residual connections to address degradation issues in very deep networks. ResNet variants like ResNet-20 and ResNet-110 showed substantial improvements on CIFAR-10.
- **DenseNet:** Huang et al. introduced DenseNet, which connects each layer to every subsequent layer to improve feature reuse and reduce the number of required parameters.
- **Capsule Networks:** Sabour et al. proposed Capsule Networks, which replaced pooling layers with dynamic routing mechanisms to preserve spatial hierarchies in feature maps.
- **Vision Transformers (ViT):** Dosovitskiy et al. demonstrated the potential of self-attention mechanisms for image classification, achieving state-of-the-art results on CIFAR-10 with Vision Transformers.

These advancements highlight the evolution of image classification techniques, transitioning from simple convolutional architectures to innovative designs such as residual connections, dense connectivity, and transformer-based models.

The motivation for this project lies in comparing the design philosophies of ResNet-152 and EfficientNet-B2:

- ResNet-152, proposed by He et al., focuses on deep residual learning, addressing vanishing gradients and facilitating the training of very deep networks.
- EfficientNet-B2, introduced by Tan and Le, employs compound scaling to balance depth, width, and resolution, optimizing accuracy and computational efficiency.

By fine-tuning these models on CIFAR-10, this study provides insights into their strengths, weaknesses, and applicability for small-scale image classification tasks.

II. Method

A. Problem Formulation

The task involves classifying CIFAR-10 images into ten categories. Each image is preprocessed to normalize pixel values and augmented with random cropping and flipping to enhance generalization.

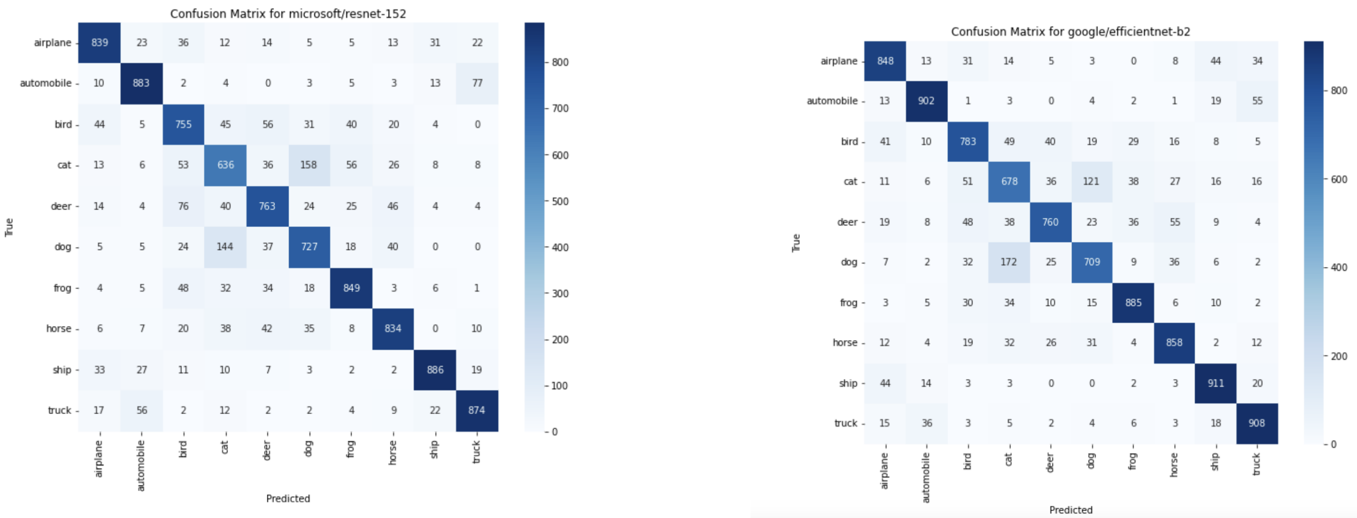
B. Dataset

The CIFAR-10 dataset contains 50,000 training images and 10,000 testing images. Validation subsets (20% of the training data) are created using stratified sampling to maintain class balance.

C. Model Architectures

ResNet-152: ResNet-152, introduced by He et al., features 152 layers and leverages residual blocks with skip connections to address the vanishing gradient problem, enabling the effective training of deep networks. Pretrained on ImageNet, it provides robust feature extraction but is computationally intensive, making it suitable for high-resource environments.

EfficientNet-B2: EfficientNet-B2, proposed by Tan and Le, uses compound scaling to balance depth, width, and resolution, optimizing performance and computational efficiency. It incorporates MBConv layers and squeeze-and-excitation blocks, offering state-of-the-art accuracy with reduced parameters and FLOPs. Its scalability makes it ideal for constrained hardware.



(a) ResNet-152 Confusion Matrix

(b) EfficientNet-B2 Confusion Matrix

Fig. 1: Confusion Matrices for ResNet-152 and EfficientNet-B2

Key Differences: While ResNet-152 focuses on depth and residual learning, EfficientNet-B2 emphasizes a balance between accuracy and efficiency, offering a lightweight and scalable alternative.

D. Training Pipeline

Both models were fine-tuned using PyTorch. Pretrained weights were loaded from Hugging Face's 'transformers' library. Training parameters included:

- **Optimizer:** AdamW with learning rates of 2×10^{-4} (EfficientNet-B2) and 1×10^{-4} (ResNet-152).
- **Batch Size:** 64 for all datasets.
- **Epochs:** 25 for EfficientNet-B2, 15 for ResNet-152.
- **Loss Function:** Cross-entropy.

The training loop included backpropagation and weight updates. Validation accuracy and loss were tracked after each epoch.

III. Results

A. Model Performance

Both models achieved competitive classification accuracy, with EfficientNet-B2 outperforming ResNet-152.

TABLE I: Performance Comparison

Metric	ResNet-152	EfficientNet-B2
Validation Accuracy	80.45%	83.00%
Test Accuracy	80.00%	82.00%
Precision	80%	82%
Recall	80%	82%

B. Confusion Matrix Analysis

Confusion matrices are shown in Fig. 1. Both models struggled with fine-grained distinctions, such as distinguishing cats from dogs. EfficientNet-B2 showed higher recall for minority classes, indicating robustness against class imbalances.

C. Computational Efficiency

EfficientNet-B2 required fewer computational resources, with faster training times and lower memory usage compared to ResNet-152, making it more suitable for resource-constrained applications.

IV. Conclusion

This study demonstrates that EfficientNet-B2 outperforms ResNet-152 in both accuracy and computational efficiency on CIFAR-10, showcasing the advantages of compound scaling for resource optimization. While ResNet-152 remains competitive, its deeper architecture incurs higher computational costs. Both models exhibit limitations in fine-grained classifications, suggesting the potential for improvement through additional data augmentation or hybrid architectures. EfficientNet-B2's scalability and robustness make it a strong candidate for small-scale image classification tasks, with future research opportunities including hybrid model designs and applications to larger datasets like ImageNet.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[2] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, PMLR, 2019, pp. 6105–6114.

[3] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Technical Report, 2009.

[4] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2016.

[5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[6] Hugging Face, "CIFAR-10 Dataset," 2024. [Online]. Available: https://huggingface.co/datasets/uoft-cs/cifar10

[7] Hugging Face, "ResNet-152 Pretrained Model," 2024. [Online]. Available: https://huggingface.co/microsoft/resnet-152

[8] Hugging Face, "EfficientNet-B2 Pretrained Model," 2024. [Online]. Available: https://huggingface.co/google/efficientnet-b2