Image Classification Using Feature Fusion of ResNet18 and MobileNetV2

Kirk Patrick T. Vallecera

Graphics and Visual Computing

Abstract. This mini case study presents an image classification approach using feature fusion of two pretrained convolutional neural networks, ResNet18 and MobileNetV2. The goal is to improve image representation by combining complementary features extracted from both models.

1 Introduction

1.1 Problem Being Solved

The problem addressed in this study is image classification, where an image is assigned to a predefined class based on its visual content.

1.2 Why It Is Relevant

Image classification is a core task in computer vision and is widely used in realworld applications such as object recognition and automated visual analysis.

2 Dataset Description

2.1 Source

The CIFAR-10 dataset was used in this study and accessed through the PyTorch torchvision library. The dataset is publicly available and not sourced from Kaggle.

2.2 Size

The dataset contains 60,000 RGB images divided into 50,000 training images and 10,000 testing images across 10 classes.

2.3 Sample Images

Sample images were visualized to verify correct loading and preprocessing.

Fig.1. Sample images from the CIFAR-10 dataset.

3 Methodology

3.1 Architectures Used

ResNet18 and MobileNetV2 pretrained on ImageNet were used as feature extractors. ResNet18 provides deep hierarchical features, while MobileNetV2 offers efficient and lightweight feature extraction.

3.2 Fusion Strategy

The final classification layers of both networks were removed. Feature vectors from ResNet18 and MobileNetV2 were concatenated into a single fused feature vector. This fused representation was passed to a fully connected layer for final classification.

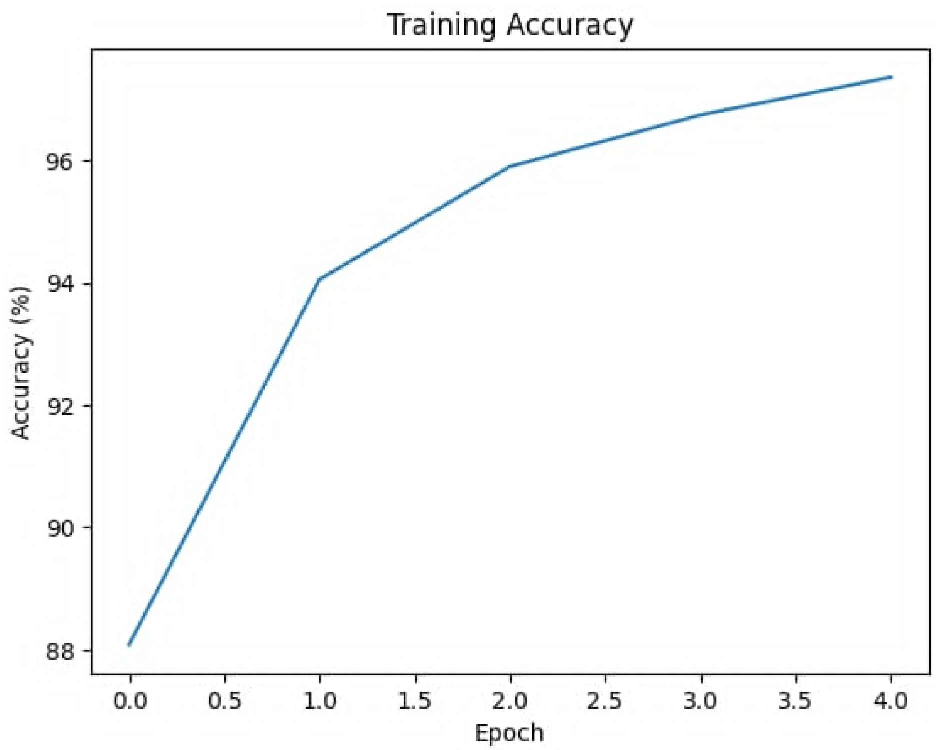
3.3 Preprocessing and Training Details

Images were resized to 224×224 pixels and normalized using ImageNet mean and standard deviation. The model was trained using cross-entropy loss and the Adam optimizer with a learning rate of 0.0003 for 5 epochs on Google Colab with GPU support.

4 Results and Visualizations

4.1 Accuracy and Loss Curves

Training accuracy increased while training loss decreased across epochs, indicating stable learning behavior.



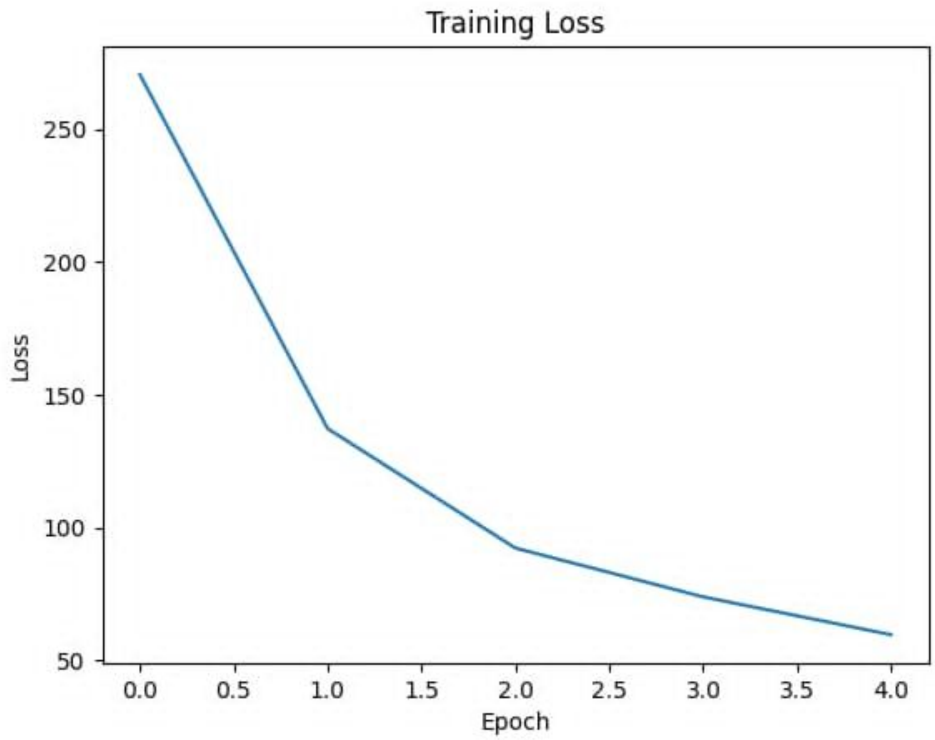


Fig.2. Training accuracy and loss curves of the fused model.

4.2 Sample Predictions

Sample predictions from the test dataset were visualized, showing both correct and incorrect classifications.

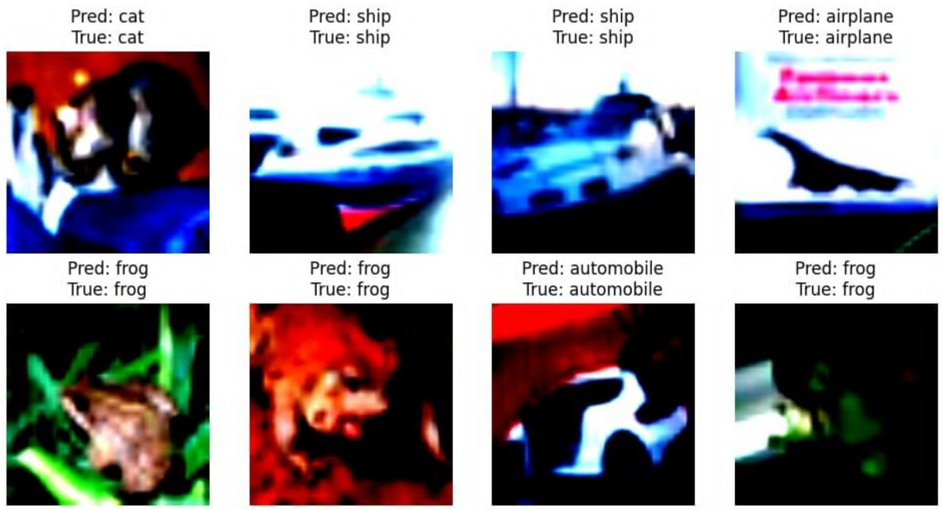
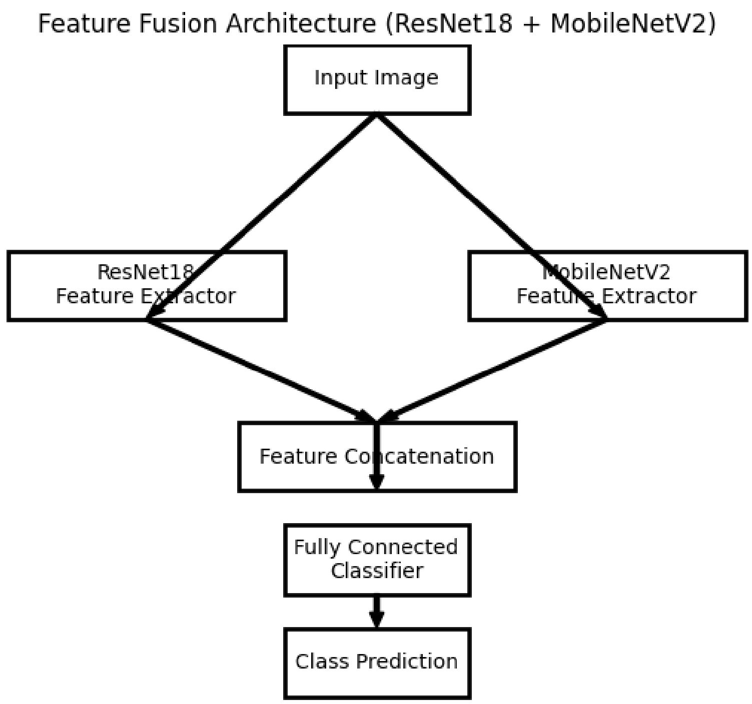


Fig.3. Sample test images with predicted and ground truth labels.

4.3 Architecture Visualization



A visual diagram was created to illustrate the feature fusion architecture.Image Classification Using Feature Fusion of ResNet18 and MobileNetV2

Fig.4. Feature fusion architecture combining ResNet18 and MobileNetV2.

5 Discussion

5.1 Fusion Contribution

Feature fusion combined complementary features from ResNet18 and MobileNetV2, resulting in richer image representations.

5.2 What Worked

Transfer learning enabled fast convergence and stable training. Feature concatenation was simple and effective.

5.3 What Did Not Work

Some misclassifications occurred between visually similar classes, and feature fusion increased computational cost.

6 Conclusion

This mini case study demonstrated that fusing features from multiple pretrained CNNs is an effective approach for image classification. The proposed model achieved reasonable performance while remaining simple and efficient.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition.CVPR (2016)
2. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.: MobileNetV2: InvertedResiduals and Linear Bottlenecks. CVPR (2018)
3. Krizhevsky, A.: Learning Multiple Layers of Features from Tiny Images (2009)