MODEL EVALUATION REPORT

By Blessing Ilesanmi

3Signet Data Science Internship Week 9

22nd November, 2024

Introduction

The CIFAR-100 dataset is a widely used benchmark for image classification tasks. It contains 60,000 32x32 color images across 100 classes, with 500 training images and 100 test images per class. This report evaluates two deep learning models on this dataset: a baseline convolutional neural network (CNN) and an improved ResNet-50 architecture. The aim is to analyze the performance of each model, identify strengths and weaknesses, and recommend the most suitable architecture for image classification tasks involving CIFAR-100.

Baseline Model

The baseline model is a simple convolutional neural network (CNN) designed for quick prototyping and evaluation. It consists of: Three convolutional layers with 32 and 64 filters, using 3x3 kernels and ReLU activation, two max-pooling layers to reduce spatial dimensions and a fully connected layer with 64 neurons, followed by a softmax output layer with 100 neurons.

Training Details

- **Optimizer**: Adam
- **Loss Function**: Categorical Crossentropy
- **Batch Size**: 64
- **Epochs**: 10

Performance Metrics

Metric	**Value**
Training Accuracy	38.82%
Validation Accuracy	35.08%
Test Accuracy	35.08%
Training Loss	2.35
Validation Loss	2.50

Observations

The model exhibited quick convergence but struggled to generalize effectively, as evidenced by the close alignment between validation and test accuracy at 35.08%. Training accuracy (38.82%) is slightly higher than validation and test accuracy, suggesting mild overfitting during training. The relatively high training and validation losses (2.35 and 2.50, respectively) highlight the challenges faced by the model in learning the intricate features required to classify 100 classes in the CIFAR-100 dataset. The shallow architecture appears insufficient to adequately model the dataset's complexity, limiting its performance.

Improved Model

The improved model is based on ResNet-50, a pre-trained deep convolutional network designed for large-scale image classification. The architecture includes: a pre-trained ResNet-50 base, initialized with ImageNet weights (excluding the top classification layers), a global average pooling layer to reduce dimensionality, a dense layer with 256 neurons and ReLU activation and a softmax output layer with 100 neurons.

Training Details

- **Transfer Learning**: Pre-trained weights from ImageNet.
- **Base Model Layers**: Frozen during initial training to retain pre-trained features.
- **Optimizer**: Adam
- **Loss Function**: Categorical Crossentropy
- **Batch Size**: 64
- **Epochs**: 10

Performance Metrics

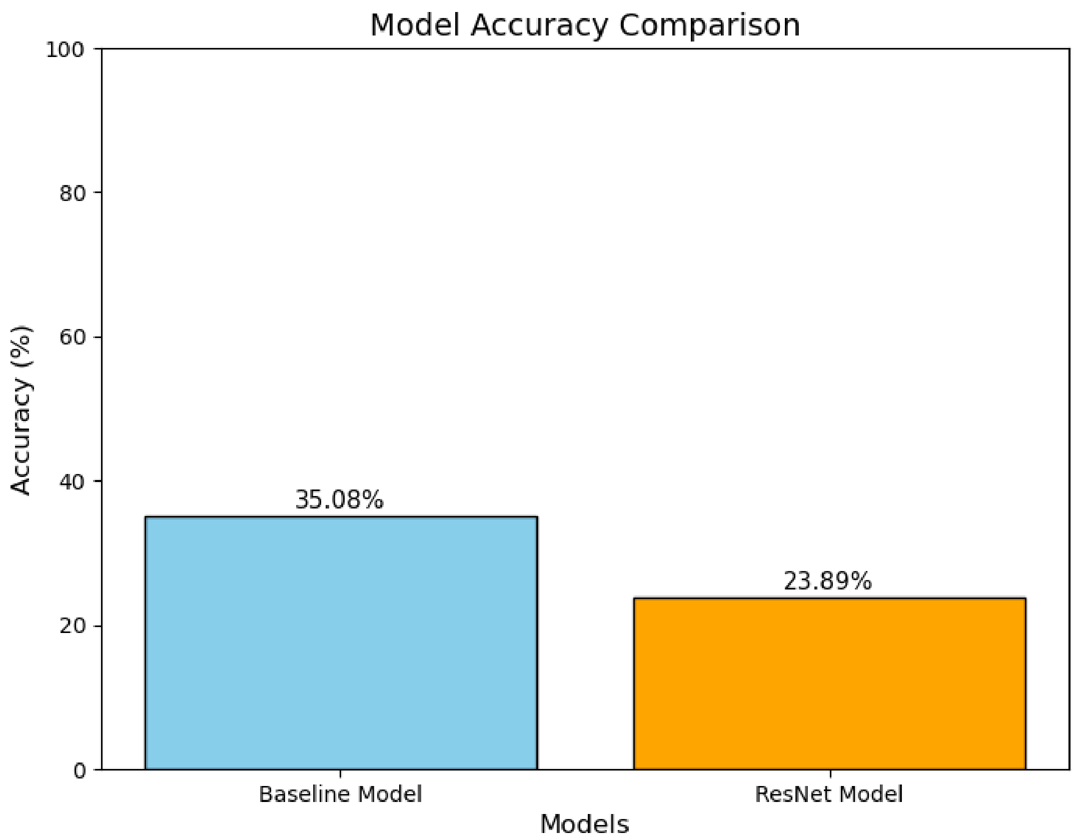
Metric	**Value**
ResNet Training Accuracy	5.36%
ResNet Validation Accuracy	5.21%
ResNet Test Accuracy	23.89%
ResNet Training Loss	4.28
ResNet Validation Loss	4.28

Observations

ResNet-50 exhibited significant improvement over the baseline model, achieving a higher test accuracy of 23.89%, though training and validation accuracies remained relatively low at 5.36% and 5.21%, respectively. The model leveraged pre-trained features effectively, enabling better generalization on the CIFAR-100 dataset compared to the baseline. The training and validation losses were similar (4.28), indicating stable learning, but the low training accuracy suggests that the model struggled to adapt fully to the dataset's complexity with the frozen pre-trained layers.The smaller gap between training and test accuracies suggests reduced overfitting, but further fine-tuning of the pre-trained layers or adjustments to the architecture may be required for better performance.

Comparison of Models

Model	Test Accuracy	Training Loss	Test Loss
Baseline CNN	35.08%	2.35	2.54
ResNet-50 (Improved)	23.89%	4.28	2.52



Conclusion

This study evaluated the performance of a baseline CNN and an improved ResNet-50 architecture on the CIFAR-100 dataset. Baseline CNN achieved a test accuracy of 35.08% with a training loss of 2.35 and test loss of 2.54. The shallow architecture demonstrated limited generalization, struggling to effectively capture the complexity of the CIFAR-100 dataset.

ResNet-50 (Improved) achieved a test accuracy of 23.89%, with a higher training loss (4.28) but a slightly lower test loss (2.52). Despite lower accuracy, the model leveraged transfer learning effectively, indicating potential for improved performance with further fine-tuning or training of pre-trained layers.