Optimizing Deep Learning Models for Image Classification on CIFAR-10

ECE-GY 7123 Deep Learning
Amarnadh Reddy Mettu , DevaKumar Katta , Jeel Patel

New York University Tandon School of Engineering
6 MetroTech Center
Brooklyn, New York 11201 USA

Abstract

This report focuses on optimizing a deep learning model for image classification on the CIFAR-10 dataset. We propose a modified ResNet architecture, leveraging techniques such as data augmentation and advanced optimization strategies to enhance model performance. The Adam optimizer with a learning rate scheduler is used to manage convergence efficiently. Our results show significant improvements in accuracy over the baseline ResNet model, demonstrating the effectiveness of our optimizations for image classification tasks. This work contributes to the development of efficient deep learning methodologies, offering a practical framework for similar applications.

1 Introduction

Image classification is a core task in computer vision, with the CIFAR-10 dataset serving as a widely used benchmark. Deep learning models, particularly ResNet architectures, have demonstrated strong performance on this dataset. However, increasing model complexity often leads to challenges such as overfitting, high computational costs, and inference latency.

This project aims to optimize a ResNet-based model for CIFAR-10, balancing accuracy and efficiency. We implement modifications to the standard ResNet architecture, leveraging PyTorch and GPU acceleration for faster training. To enhance generalization, we apply data augmentation techniques such as random cropping, horizontal flipping, color jittering, and random rotation. Additionally, we use the Adam optimizer with a learning rate scheduler to ensure stable convergence.

Our results demonstrate significant accuracy improvements over the baseline ResNet model while maintaining computational efficiency. This work provides insights into designing optimized deep learning models for image classification, contributing to more effective and practical implementations in real-world applications.

2 Methodology

This section explain the step by step process of the network, any techniques used and the hyperparameters chosen for the model.

2.1 Dataset

This project employs the CIFAR-10 dataset, a widely used benchmark for image classification. The dataset consists of 60,000 images of size pixels, categorized into 10 classes. The dataset is split as follows:

- Training Set: 50,000 images
- Test Set: 10,000 images (used for evaluation)
- Competition Test Set: `cifar_test_nolabel.pkl` containing unlabeled images

2.2 Data Preprocessing

To ensure consistency between training and inference, the following preprocessing transformations were applied using `torchvision.transforms`:

- Random Cropping (32, padding=4): Adds slight random cropping with padding to improve robustness.
- Random Horizontal Flip: Randomly flips images horizontally for data augmentation.
- Color Jitter: Randomly modifies brightness, contrast, saturation, and hue.
- Random Rotation (20 degrees): Adds random rotation up to 20 degrees to increase robustness.
- Random Affine Transformations: Randomly translates the image within a 10%range.
- Normalization: Standardizes pixel values using CIFAR-10 statistics:

$$\begin{aligned}\mu &= (0.4914, 0.4822, 0.4465)\\ \sigma &= (0.2023, 0.1994, 0.2010)\end{aligned} \tag{1}$$

3 Model Architecture

The ModifiedResNet model is based on the ResNet-18 architecture with key modifications aimed at optimizing performance on the CIFAR-10 dataset. Two variations of the architecture were developed to analyze the trade-off between model complexity and computational efficiency.

Model 1: Deeper Modified ResNet The first model is a deeper version of ResNet with an increased number of residual blocks, improving feature extraction and learning capacity. The architecture consists of:

- **Initial Convolution:** A 3x3 convolutional layer with 64 filters, stride=1, and padding=1.
- **Residual Blocks:**
  - **Layer 1:** Composed of 4 BasicBlocks, each operating on 64-channel feature maps, maintaining the spatial resolution of the input image while enriching feature representation.
  - **Layer 2:** Consists of 4 BasicBlocks, transitioning to 128-channel feature maps. The first convolutional layer in this block applies a stride of 2, effectively reducing the spatial dimensions by half.
  - **Layer 3:** Contains 3 BasicBlocks, further increasing the depth to 256 channels. This stage extracts high-level semantic information critical for classification, with additional downsampling to refine feature granularity.
- **Trainable Parameters:** 4,697,162
- **Memory Usage:** 36.56 MB

3.1 Model 2: Lighter Modified ResNet

This model follows the same general structure but with a reduced number of residual blocks, making it more lightweight and optimized for lower-resource environments.

- **Residual Blocks:**
  - **Layer 1:** 2 BasicBlocks operating on 64-channel feature maps, preserving spatial resolution while learning fundamental patterns.
  - **Layer 2:** 2 BasicBlocks processing 128-channel feature maps, introducing a stride of 2 in the first convolution to downsample the spatial dimensions.
  - **Layer 3:** 2 BasicBlocks increasing the depth to 256 channels, further refining high-level semantic representations required for classification.
- **Trainable Parameters:** 2,777,674
- **Memory Usage:** 21.11 MB

4 Training

With the model architecture and hyperparameters set, the next step is to train the Modified ResNet on the CIFAR-10 dataset. The data is loaded through PyTorch's DataLoader, incorporating the above mentioned augmentation techniques to enhance generalization and prevent overfitting.

The Cross-Entropy Loss function is chosen due to its effectiveness in classification tasks. Given the categorical nature of the CIFAR-10 dataset, Cross-Entropy Loss ensures a well-calibrated prediction distribution over the 10 classes.

For optimization, the Adam optimizer is used due to its adaptive learning rate properties. To further refine training stability, ReduceLROnPlateau is employed as a learning rate scheduler, dynamically adjusting the learning rate when validation loss stagnates.

The model is trained for 75 epochs, and L2 regularization (weight decay) is applied to penalize large weights, further enhancing the model's generalization performance. This technique ensures that the network does not overfit to specific training examples but instead learns a robust representation that performs well on unseen test data.

5 Results and Discussion

Our optimized ResNet models were trained and evaluated on the CIFAR-10 dataset. The deeper Modified ResNet ([4,4,3]) achieved a final test accuracy of 93.20%, while the lighter Modified ResNet ([2,2,2,2]) obtained an accuracy of 93.11%.

The performance of the Modified ResNet models was evaluated using the CIFAR-10 dataset. The evaluation was conducted on the test set, measuring accuracy and loss during training and testing. The primary metrics for assessment included training loss, test loss, and classification accuracy.

5.1 Training Performance

The models were trained for **75 epochs**, with loss and accuracy monitored on both the training and test datasets. The performance of both models was analyzed to understand the impact of depth on classification accuracy and computational efficiency.

Performance of Lighter Modified ResNet The first model, a shallower variation of ResNet, was trained and evaluated on CIFAR-10. The training and testing loss curves for this model are depicted in Figure 1, while the corresponding accuracy trends are shown in Figure 2.

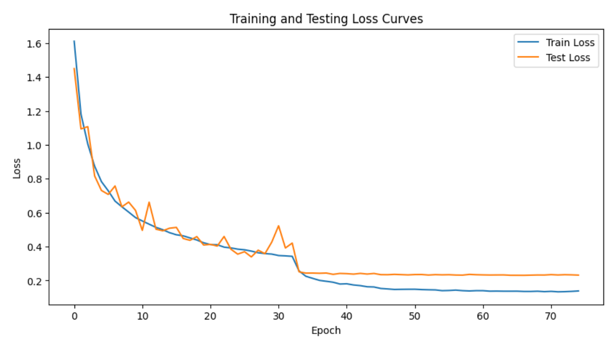

Figure 1: Train and test loss for the lighter ResNet model.

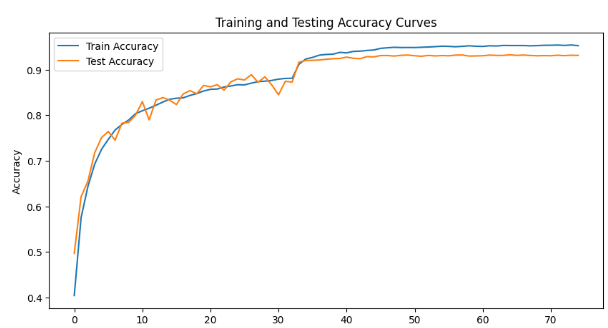

Figure 2: Train and test accuracy for the lighter ResNet model.

This model, with 2,777,674 parameters and a memory footprint of 21.11 MB, exhibited faster training times due to its reduced computational complexity. The lower number of parameters enabled efficient training while maintaining competitive accuracy. However, a trade-off was observed, as its performance was slightly lower compared to the deeper model.

Performance of Deeper Modified ResNet The second model, a deeper variation of ResNet, was trained and evaluated using the same methodology to compare its performance against the lighter model. The training and test loss curves for this model are shown in Figure 3, and the corresponding accuracy trends are illustrated in Figure 4.

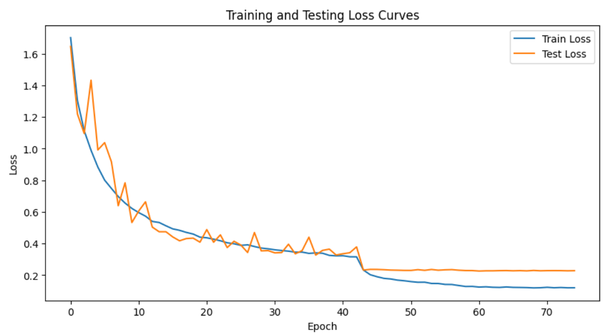

Figure 3: Train and test loss for the deeper ResNet model over 75 epochs.

Figure 4: Train and test accuracy for the deeper ResNet model over 75 epochs.

The deeper model, which contains 4,697,162 trainable parameters and requires an estimated 36.56 MB of memory, achieved improved accuracy compared to the shallower model. This model demonstrated superior feature extraction capabilities, making it better suited for applications where accuracy is prioritized over computational efficiency.

Overall, the study highlights the trade-off between model complexity and performance. The deeper model exhibits better classification accuracy, whereas the lighter variant offers improved efficiency, making it a suitable choice for resource-constrained environments.

6 Conclusion

This study demonstrates the effectiveness of the Modified ResNet models for image classification on the CIFAR-10 dataset. By incorporating data augmentation, learning rate scheduling, and L2 regularization, both models exhibited strong generalization capabilities. The deeper Modified ResNet ([4,4,3]) achieved a final test accuracy of 93.20%, while the lighter Modified ResNet ([2,2,2,2]) attained an accuracy of 93.11%.

The results highlight the trade-off between model depth and computational efficiency. The deeper model exhibited superior feature extraction, leading to slightly better accuracy, but required significantly more memory (36.56 MB vs. 21.11 MB) and had a higher parameter count (4.7M vs. 2.77M). In contrast, the lighter model offered a more efficient alternative for environments where computational resources are constrained while maintaining comparable accuracy.

7 Acknowledgments

We would like to thank Professor Chinmay Hegde for their guidance throughout this project. We also extend our gratitude to our peers at New York University for their valuable feedback and discussions. Lastly, we acknowledge the use of computational resources provided for training deep learning models efficiently.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, ”Deep Residual Learning for Image Recognition,” in extitProceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[2] D. Shuvam, ”Implementation of ResNet Architecture for CIFAR-10 and CIFAR-100 Datasets,” 2023. Available: https://medium.com/@shuvamd210/unveiling-themetrics-mystery-exploring-scg-and-ccg-for-image-modelevaluation-76301445f583.

[3] K. Liu, ”PyTorch CIFAR Models,” 2021. Available: https://github.com/kuangliu/pytorch-cifar.