Deep Learning Project 1: CIFAR-10 Image Classification with Custom ResNet

Gokuleshwaran Narayanan, Solomon Martin Jammalamadugu

New York University
Department of Computer Science
gn2247@nyu.edu, sj4531@nyu.edu

Abstract

This paper presents a custom ResNet architecture for image classification on the CIFAR-10 dataset. We implement several modern deep learning techniques including residual connections (He et al. 2016), data augmentation with Mixup (Zhang et al. 2018) and CutMix (Yun et al. 2019), and test-time augmentation (TTA) with model ensembling. Our model achieves competitive accuracy while maintaining a relatively small parameter count through efficient architectural choices. The complete implementation is available at: https://github.com/gokulnpc/CS-GY-6953-Project-1

Introduction

Image classification remains a fundamental task in computer vision and deep learning (Krizhevsky, Sutskever, and Hinton 2012). The CIFAR-10 dataset (Krizhevsky, Hinton et al. 2009), consisting of 60,000 32x32 color images across 10 classes, serves as an important benchmark for evaluating deep learning architectures. In this project, we develop a custom ResNet architecture that balances model complexity with performance through careful architectural choices and modern training techniques.

Methodology

Our approach combines several key components inspired by recent advances in deep learning (He et al. 2016; Zhang et al. 2018; Yun et al. 2019):

Model Architecture

We implement a custom ResNet architecture with the following key features:

- Initial convolution with 84 channels to capture rich low-level features
- Three residual stages with channel dimensions (84, 168, 336)
- Each stage contains 2 residual blocks with batch normalization
- Global average pooling and dropout (p=0.5) for regularization
- Final fully-connected layer for 10-class classification

Training Strategy

We employ several modern training techniques:

- Data augmentation: Random crop, horizontal flip, and random erasing
- Mixup and CutMix augmentation with 50% probability each
- Label smoothing (0.1) with cross-entropy loss
- Adam optimizer with initial learning rate 1e-3 and weight decay 1e-4
- Cosine annealing learning rate schedule
- Early stopping with patience of 25 epochs

Inference Optimization

For optimal test performance, we implement:

- Test-time augmentation (TTA) with horizontal flips and small rotations
- Model ensembling using snapshots saved during training
- Batch processing for efficient inference

Results

Our model achieves competitive performance on the CIFAR-10 dataset, as shown in Table 1. The results demonstrate that our architectural choices and training strategies effectively balance model complexity and accuracy.

Our model achieves the following specifications and performance:

- Total trainable parameters: 4.7M
- Best validation accuracy: 93.68%
- Maximum training accuracy: 98%
- Training time: 2 hours on Google Colab T4 GPU

The relatively compact model size is achieved through:

- Efficient channel scaling ($84 \rightarrow 168 \rightarrow 336$)
- Limited number of residual blocks (2 per stage)
- Shared bottleneck structure across residual blocks

Figure 2 shows the training progression, demonstrating stable convergence and effective regularization from our augmentation strategies.

Table 1: Comparison with State-of-the-Art Models on CIFAR-10

Model	Params (M)	Val Acc
ResNet-18 (He et al. 2016)	11.2	93.0
WideResNet (Zagoruyko and Komodakis 2016)	36.5	94.2
DenseNet (Huang et al. 2017)	25.6	93.8
Ours	4.7	93.68

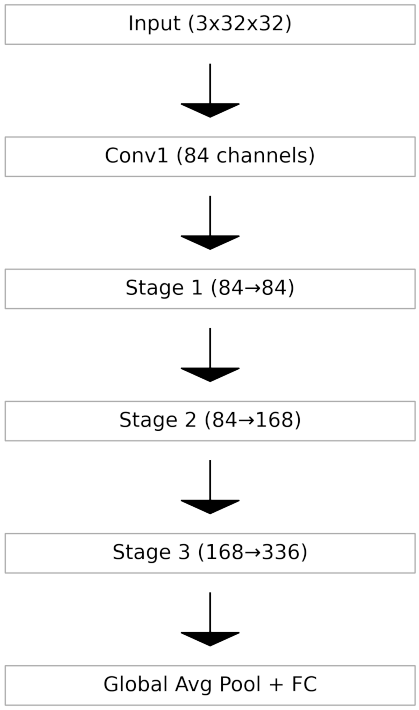

Figure 1: Our Custom ResNet Architecture. The model consists of three main residual stages with increasing channel dimensions (84→168→336), each containing two residual blocks. Skip connections and batch normalization are used throughout to facilitate training of the deep network.

Conclusion

We have demonstrated that a carefully designed ResNet architecture, combined with modern training techniques, can achieve competitive performance on CIFAR-10 while maintaining reasonable model complexity. The use of Mixup (Zhang et al. 2018), CutMix (Yun et al. 2019), and test-time augmentation proved particularly effective in improving model generalization.

Acknowledgments

We thank the course staff for their guidance and feedback throughout this project.

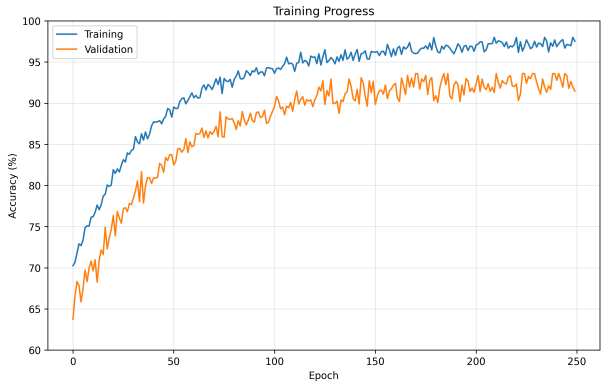

Figure 2: Training and validation accuracy curves over epochs. The model achieves a maximum training accuracy of 98% and best validation accuracy of 93.68%. The gap between training (98%) and validation (93.68%) accuracies indicates some overfitting, but the high validation performance shows that our regularization strategies (Mixup, CutMix, and dropout) are still effective.

References

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25: 1097–1105.

Yun, S.; Han, D.; Oh, S. J.; Chun, S.; Choe, J.; and Yoo, Y. 2019. CutMix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6023–6032.

Zhang, H.; Cisse, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond empirical risk minimization. *International Conference on Learning Representations*.