CNN Performance on CIFAR-10

Introduction

This report evaluates the performance of convolutional neural networks (CNNs) on the CIFAR-10 dataset, comparing a baseline model against three improved architectures. The CIFAR-10 dataset consists of 60,000 32x32 RGB images across 10 classes, posing challenges in feature extraction and generalization due to its complexity and limited resolution.

The baseline model employed a shallow CNN with two convolutional layers and ReLU activations. Subsequent experiments introduced modifications such as deeper architectures, advanced activation functions, regularization techniques, and batch normalization to address overfitting and improve accuracy.

Methodology

Experiment 1 (baseline) utilized two convolutional layers (32 and 64 filters, 3x3 kernels), ReLU activations, and no dropout nor batch normalization. Training was performed over 10 epochs with the Adam optimizer. Three improved models were then tested:

Experiment 2 incorporated dropout layers (rates 0.1 and 0.2), batch normalization after convolutional layers, and data augmentation (rotations, shifts, and zooms) to mitigate overfitting.

Experiment 3 expanded network depth to three convolutional layers (64, 128, 256 filters) with larger kernels (up to 5x5), replaced ReLU with ELU activations for convolutional and GELU for hidden layers, applied batch normalization universally, and no dropout layers.

Experiment 4 retained Experiment 3's depth but adopted ELU activations for both convolutional and hidden layers, selective dropout (0.2 only in the final layer), and no data augmentation.

All models were evaluated using performance metrics (accuracy, loss, precision, recall). Training leveraged the categorical cross-entropy loss function, early stopping with a 5-epoch patience period and learning rate reducer.

Results

The baseline model achieved a test accuracy of **73%** with a loss of **0.80**, demonstrating moderate performance but significant overfitting. *Experiment 2*, despite introducing dropout and augmentation, underperformed with **71%** accuracy and higher loss **0.84**, suggesting overly aggressive regularization stifled feature learning. *Experiment 3* improved accuracy to **77%** and **0.68** loss through deeper layers and GELU activations, though precision plateaued at **83%**. *Experiment 4* delivered the best results: **81%** accuracy, **0.50** loss, **85%** precision, and **77%** recall, attributed to ELU's robust gradient propagation and strategic dropout placement.

Discussion

The superior performance of *Experiment 4* highlights the importance of balancing depth and regularization. ELU activations mitigated the "dying ReLU" problem in deeper layers, while selective dropout preserved critical features in early layers. Batch normalization stabilized training across all configurations, though its impact was most pronounced in *Experiment 4's* deeper architecture. *Experiment 2's* accuracy drop underscores the challenge of tuning dropout rates; excessive regularization eroded discriminative power. Computational costs increased with model depth, with *Experiment 4* requiring **30%** more training time than the baseline. A key limitation was the absence of data augmentation in *Experiment 4*, which may further enhance generalization.

Conclusion

The optimal model (*Experiment 4*) achieved **81%** test accuracy, a **10.9%** improvement over the baseline, by integrating three critical modifications: ELU activations for stable gradient flow, **selective dropout** to reduce overfitting without sacrificing feature extraction, and **progressive kernel sizing** (3x3 to 5x5) for hierarchical feature learning. Batch normalization ensured consistent convergence across layers.