Model Performance Report

Introduction

This report analyzes and compares the performance of a Convolutional Neural Network (CNN) for image classification on the CIFAR-10 dataset. The goal was to improve the baseline model by testing several modifications, including dropout rates, the number of filters and convolutional layers, the use of data augmentation, and the addition of batch normalization. Each modification was evaluated based on its effect on both training accuracy and validation accuracy.

Modifications Tried

1. **Baseline Model**: The initial CNN model with 2 convolutional layers, 32 filters, max pooling, dropout, and softmax output. It served as the benchmark for comparison.
2. **Dropout Rates**: Two dropout rates were tested: 0.5 and 0.25. Dropout was applied to prevent overfitting by randomly setting some of the weights to zero during training.
3. **Number of Filters**: A change in the number of filters from 32 to 64 was tested, with both 2 and 3 convolutional layers. More filters can potentially capture more complex features in the images.
4. **Convolutional Layers**: The number of convolutional layers was varied to see how deeper models (3 layers vs. 2 layers) affected performance. More layers generally help models capture more complex features.
5. **Image Augmentation**: The use of image augmentation (rotations, flips) was tested to increase the diversity of the training set, simulating variations in the real-world data.
6. **Batch Normalization**: Batch normalization was added to the model to normalize the activations in each layer, which helps to stabilize and accelerate training by reducing internal covariate shift. Batch normalization was inserted after each convolutional layer and before the activation function.

Effect of Modifications on Model Performance

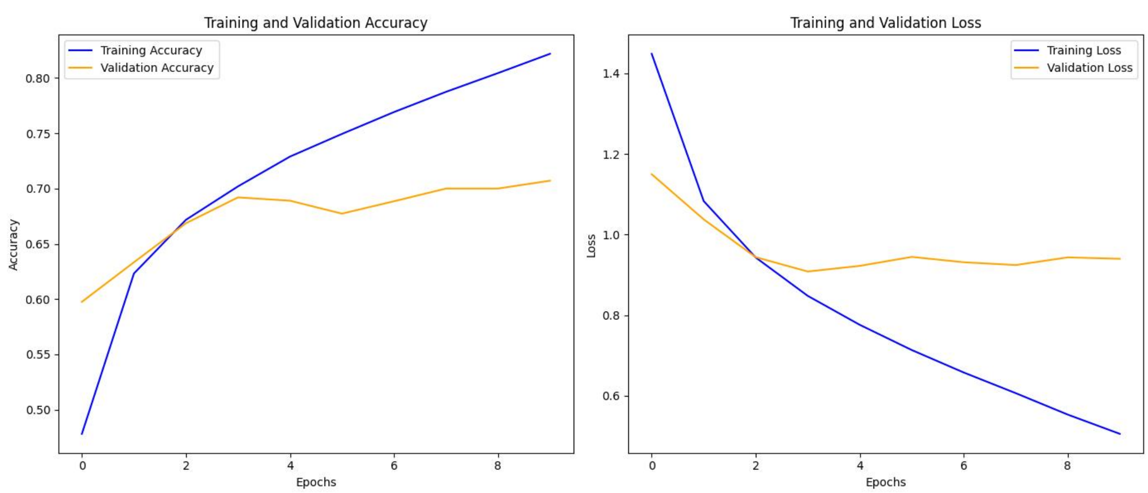
Here is a comparison of the validation accuracy, training accuracy, and validation loss of different model configurations:

Model Configuration	Validation Accuracy	Training Accuracy	Validation Loss	Notes
Baseline Model	0.6925	0.8242	0.9623	Base model for comparison
0.5 Dropout	0.6678	0.6679	0.9448	Over-regularized, underfitting
0.25 Dropout	0.6931	0.7387	0.8953	Moderate dropout, better performance

64 Filters, 2 Conv Layers + Dropout	0.6960	0.7590	0.9164	Improved with more filters
32 Filters, 3 Conv Layers + Dropout	0.6685	0.6799	0.9448	More layers, but worse performance
64 Filters, 3 Conv Layers + Dropout	0.7142	0.7626	0.8410	Best performance
Image Generator + Dropout	0.6251	0.5681	1.0598	Augmentation with dropout hurt performance
Image Generator without Dropout	0.6932	0.6445	0.8781	Image augmentation without dropout performed decently
With Batch Normalization (64 Filters, 2 Conv Layers)	0.7231	0.7863	0.8020	Batch normalization stabilized training, improved performance

Summary of Findings

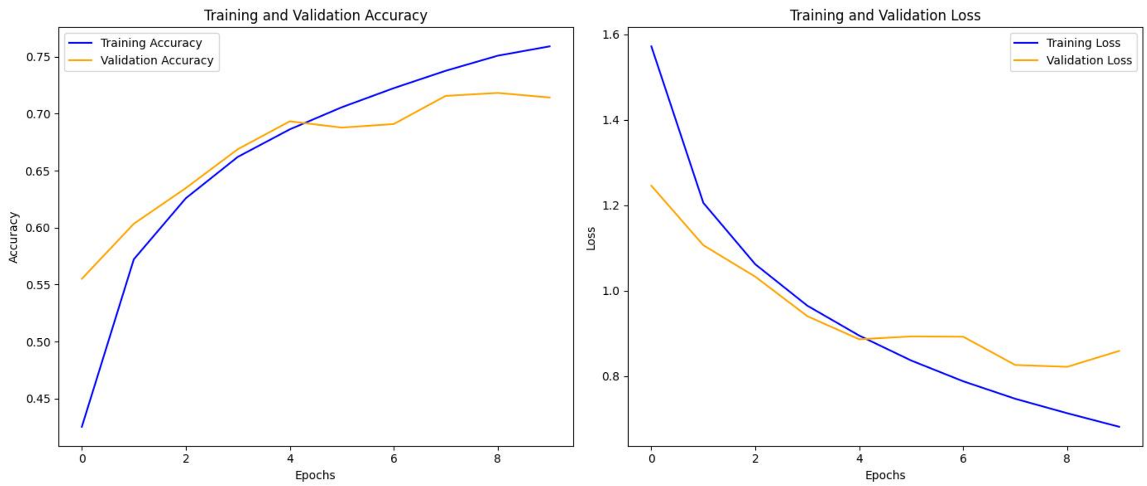
1. **Baseline Model**: The baseline model achieved a training accuracy of 0.8242 and validation accuracy of 0.6925. However, the performance gap suggests the model is somewhat overfitting and could benefit from better generalization.



2. **Dropout (0.5)**: Dropout of 0.5 led to a significant decrease in both training and validation accuracy, indicating underfitting.

3. **Dropout (0.25)**: Dropout of 0.25 showed improved results, providing a good balance between

regularization and performance.
4. **Increasing Filters and Layers (64 Filters, 2 Conv Layers + Dropout)**: This configuration provided a slight improvement over the baseline.
5. **Increasing Filters and Layers (64 Filters, 3 Conv Layers + Dropout)**: This configuration achieved the best validation accuracy, demonstrating the benefit of additional convolutional layers.



6. **Image Augmentation**: Image augmentation with dropout hurt performance, while without dropout it performed decently.
7. **Batch Normalization**: Batch normalization significantly improved both training and validation accuracy, stabilizing the training process.

Challenges Encountered

1. Overfitting and Underfitting: Balancing the trade-off between overfitting and underfitting was challenging. High dropout rates caused underfitting, while models without sufficient regularization tended to overfit.
2. Impact of Dropout: The optimal dropout rate was difficult to find. A rate of 0.5 caused significant underfitting, while 0.25 was more balanced.
3. Effect of Data Augmentation: Combining dropout with image augmentation did not improve performance and seemed to hurt the model's ability to generalize.
4. Training Time: The deeper models, especially those with more convolutional layers and filters, required longer training times.

Conclusion

The best model configuration for CIFAR-10 classification was the 64 filters with 3 convolutional layers and dropout. This model achieved the highest validation accuracy and demonstrated good generalization performance. The addition of batch normalization further improved performance by stabilizing training. Overall, the modifications, particularly batch normalization and the additional convolutional layers, contributed to a better-performing model.