Assignment 4

Name: Waleed Malik
Student Number: 400327771
Kaggle Username: **waleedmalik7**

COMPSCI 4ML3: Introduction to Machine Learning
December 2, 2024

Contents

Kaggle Report

Abstract

This report presents the implementation details and results of a deep learning model used for the CIFAR-100 image classification task on Kaggle.

After 100 epochs the final model achieved a train and test loss and accuracy of:

- Train loss: **1.154190**, Train accuracy: **66.81%**
- Test loss: **1.696981**, Test accuracy: **56.03%**

The model also achieved a Kaggle Score of 0.3184 when tested with the given test.csv

The model architecture consists of a simple convolutional neural network (CNN) with three convolutional layers, max-pooling, dropout for regularization, and two fully connected layers. The dataset was preprocessed using standard normalization techniques, with data augmentation methods such as random cropping and horizontal flipping applied to the training set to improve generalization.

The training process utilized the SGD optimizer, a learning rate of 0.05, and categorical cross-entropy as the loss function. Despite the simplicity of the model, it achieved competitive performance on CIFAR-100 given its complexity of 100 output classes.

Introduction

The CIFAR-100 dataset is a widely used benchmark for image classification tasks in computer vision. It consists of 60,000 color images of size 32x32 pixels, categorized into 100 distinct classes. Each image belongs to one of 20 superclasses, with each superclass containing 5 subcategories. The dataset is challenging due to its large number of classes and small image resolution, making it an excellent testbed for evaluating the effectiveness of deep learning models.

The task involves developing a machine learning model capable of classifying images into their respective categories with high accuracy. This requires designing an appropriate model architecture, preprocessing the data effectively, and fine-tuning hyperparameters to achieve the best possible performance.

In this competition, the primary objective is to maximize the classification accuracy on the test dataset, which consists of 10,000 unseen images. The model's performance is evaluated based on the percentage of correctly classified images, and the results are compared on the Kaggle leaderboard.

The Model

Inpiration for the Model

The model design for CIFAR-100 classification was inspired by previous experiments, particularly from Assignment 4, which involved classifying the SVHN dataset. The lessons learned during that task, such as the effectiveness of data augmentation, dropout for regularization, and leveraging convolutional neural networks, were applied to this more complex dataset. Given CIFAR-100's challenge of classifying images into 100 distinct categories, the architecture was carefully designed to balance simplicity, computational efficiency, and accuracy.

Preprocessing and Data Augmentation for the Model

Preprocessing and data augmentation played a critical role in improving model generalization. Data preprocessing involved normalizing the pixel values of the images using the CIFAR-100 dataset's channel-wise mean and standard deviation. This ensured that the input data was standardized, which is crucial for stabilizing and accelerating the convergence of the model during training.

In addition, data augmentation techniques such as **random horizontal flipping and random cropping** were added. The random horizontal flipping introduced symmetry to the training process, enabling the model to generalize better to variations in object orientation. Random cropping, combined with a padding of 4 pixels, allowed the model to learn robust features by simulating different object positions within the image.

Design of the Model

The model itself is a convolutional neural network (CNN) that processes the images through a sequence of convolutional layers, max-pooling, and fully connected layers. Three convolutional layers progressively increase the depth of feature maps, starting from 32 channels in the first layer, 64 channels in the second, and finally 128 channels in the third. These layers capture spatial and hierarchical features in the input images, enabling the model to learn both coarse and fine-grained patterns. After every two convolutional layers, a max-pooling operation is applied to reduce the spatial resolution, making the model more efficient by focusing on high-level features. Pooling also imparts translation invariance to the network, ensuring that the model remains robust to minor shifts in object position.

The fully connected layers at the end of the network further refine the features extracted by the convolutional layers. The flattened output from the final pooling layer, which represents a condensed version of the image's features, is passed to a fully connected layer with 256 neurons. This intermediate layer captures high-level representations of the data before being passed to the final output layer, which maps these features to the 100 output classes corresponding to CIFAR-100 categories.

Regularization was an essential component of this model, as it helped mitigate overfitting on the training data. A dropout layer with a rate of 50% was applied before the final output layer. By randomly disabling neurons during training, dropout ensures that the model does not overly depend on specific features, thereby improving its ability to generalize to unseen data.

In terms of design choices, this model was kept relatively simple to balance computational requirements and performance. A deeper architecture might have offered marginal improvements but at the cost of increased training time and the risk of overfitting. The inclusion of pooling layers and dropout was particularly significant, as these techniques ensured that the model remained efficient and robust despite the complexity of CIFAR-100. Overall, this architecture demonstrated the ability to effectively classify CIFAR-100 images, achieving competitive results while maintaining computational efficiency.

Results and Analysis

The results of the model training and testing are summarized in Figures 1 and 2, which show the loss and accuracy trends over 100 epochs. The training loss, as seen in Figure 1, consistently decreases over the epochs, indicating that the model effectively learned from the training data. By the 100th epoch, the training loss reaches a low value of approximately 1.154. The test loss, on the other hand, initially decreases alongside the training loss but begins to plateau after approximately 20 epochs, with fluctuations observed in later epochs. The test loss stabilizes around 1.697 by the end of training. These fluctuations suggest that while the model captures patterns in the data well, there are challenges in generalizing to unseen data, likely due to overfitting.

The accuracy trends, depicted in Figure 2, further support these findings. The training accuracy steadily improves throughout the 100 epochs, achieving a final value of 66.81%. The test accuracy follows a similar upward trend initially but plateaus around 56.03% with minor fluctuations after the 40th epoch. The 10% gap between training and test accuracy indicates overfitting, where the model performs better on the training set than on the test set. Despite this, the test accuracy demonstrates that the model has learned meaningful features from the CIFAR-100 dataset.

The fluctuations in the test loss and the gap between training and test accuracy indicate that further improvements could be made. Regularization techniques such as weight decay could help reduce overfitting, and additional data augmentation methods, such as random rotations or color jittering, might improve robustness. Furthermore, employing a learning rate scheduler or exploring more complex architectures, such as ResNet, could lead to better overall performance.

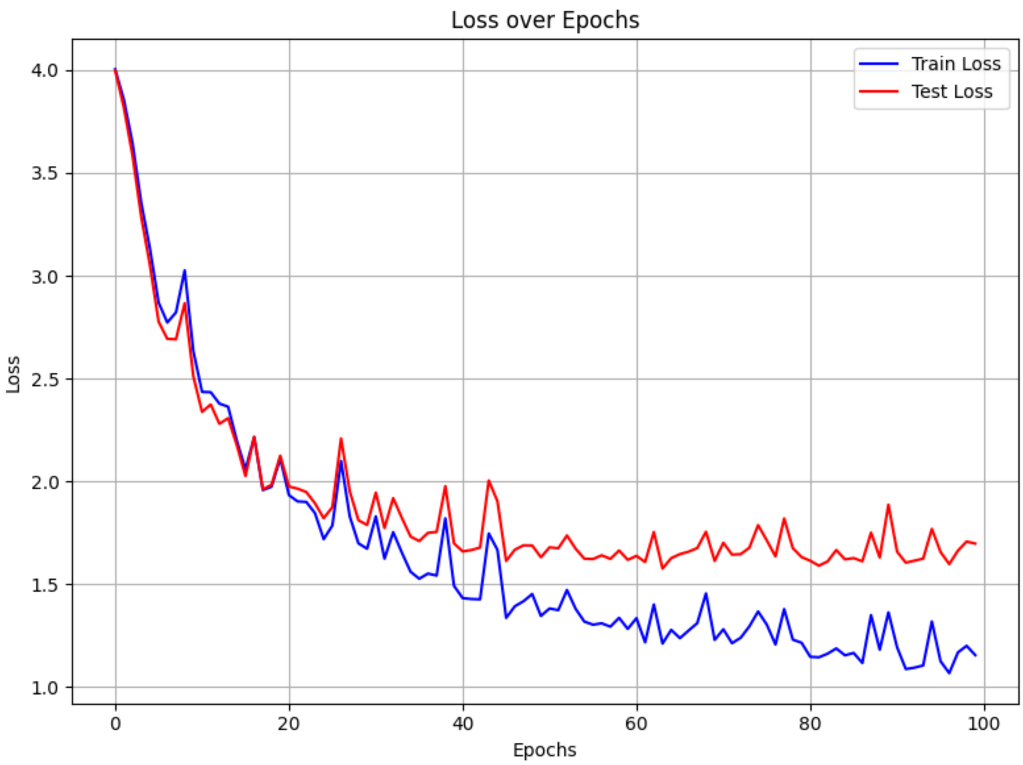

Figure 1: Training and Test Loss Trends Over 100 Epochs for CIFAR-100 Classification

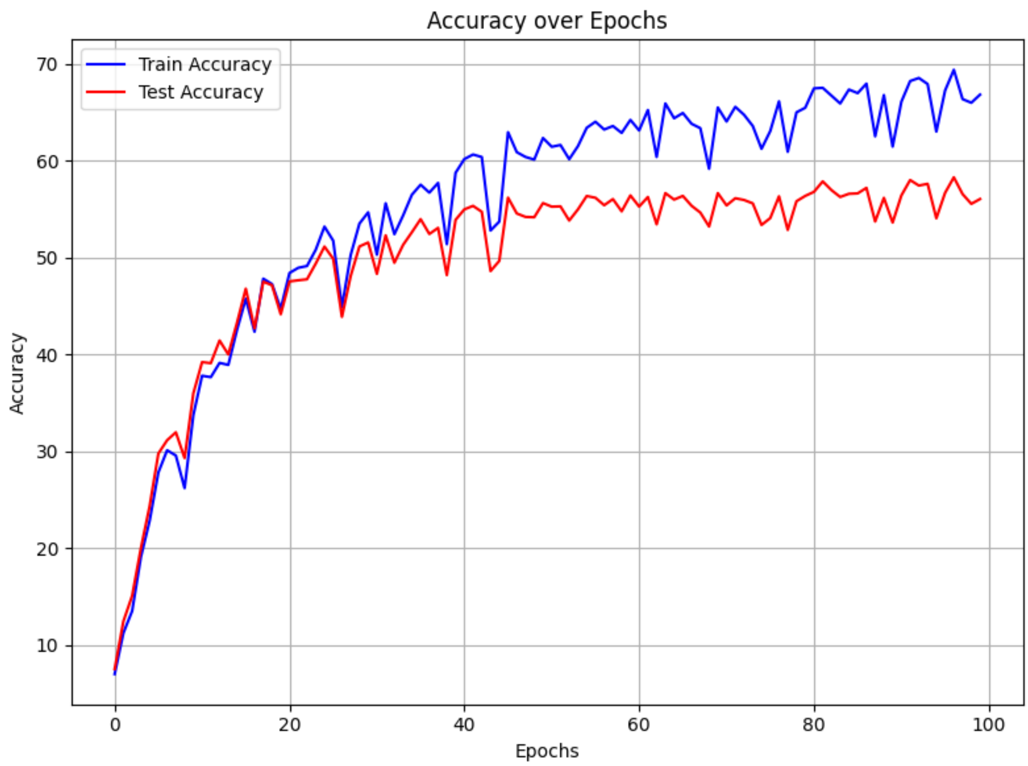

Figure 2: Training and Test Accuracy Trends Over 100 Epochs for CIFAR-100 Classification

Conclusion

This report presented the design and implementation of a convolutional neural network (CNN) for CIFAR-100 image classification, achieving a test accuracy of 56.03% and a Kaggle score of 0.3184. The model utilized data augmentation, dropout regularization, and a simple three-layer architecture to balance efficiency and performance. While effective in learning features, overfitting remains an area for improvement, which could be addressed through advanced regularization techniques or deeper architectures. Overall, this project provides a strong foundation for tackling the complexities of CIFAR-100 classification.