Introduction

The CIFAR-10 dataset is a widely used benchmark for evaluating image classification models. It contains 60,000 32x32 color images divided into 10 classes, making it ideal or testing the effectiveness of convolutional neural networks (CNNs). In this report, the performance of a CNN trained from scratch is compared against a pre-trained ResNet-18 model for classifying images from the CIFAR-10 dataset.

Methodology

The dataset was split into 50000 training images and 10000 test images. The CNN trained from scratched consists of 2 convolutional layers to extract features, an activation function (ReLU in this case), 2 pooling layers using max pooling and 2 fully connected layers for classification. ResNet-18, on the other hand, is a deep residual network trained on ImageNet dataset, which includes residual blocks that mitigate the vanishing gradient problem. To ensure compatibility with the CIFAR-10 dataset, the final fully connected layer was replaced with a new layer to output 10 classes. Images were also normalised with the mean and standard deviation values aligned with the ImageNet dataset that ResNet-18 was trained to ensure consistency between how its training and how it would be used. Meanwhile, the early layers were frozen so that only the last few layers are fine-tuned on CIFAR-10.

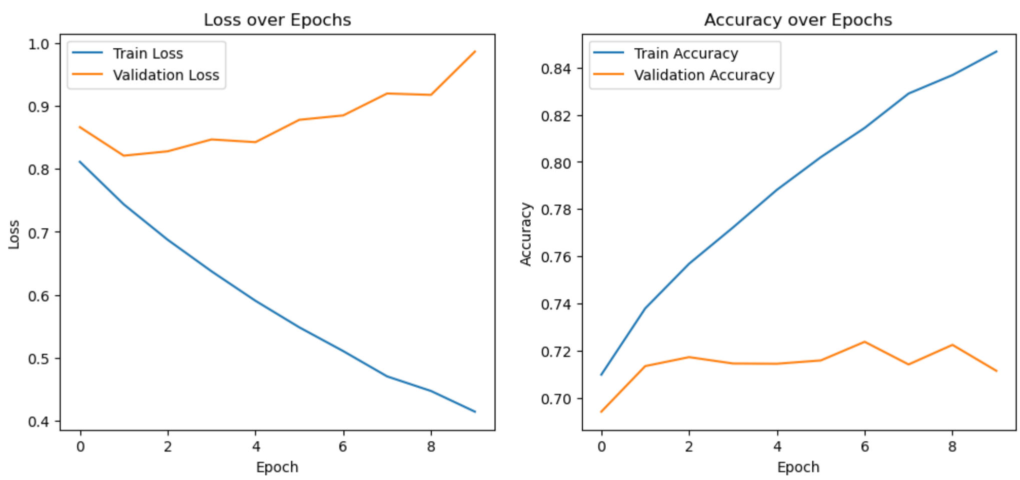
Both models were trained for 10 epochs with the following parameters: Adam optimiser, cross-entropy loss function, a batch size of 64, and a learning rate of 0.001, which was applied across all parameters for the CNN trained from scratch whereas the learning rate was applied only to the fully connected layer of ResNet-18 model.

Results and Comparison

The models are evaluated based on the following performance metrics: classification accuracy on the test set and training time. Metrics are as follows:

	CNN from scratch	**ResNet-18**
Training time (mm:ss)	36:47	86:35
Test accuracy (%)	71.13	45.33

Loss and validation accuracy of the CNN model trained from scratch are found as follows:



Based on the loss and accuracy plots, as the number of epochs increased, the gap between training and validation became wider, suggesting overfitting to the training data, which is a typical behaviour seen in CNN trained on image classification.

Analysis and Discussion

The ResNet-18 model, trained with transfer learning, was expected to require less time and yield better performance due to the pre-learned features from ImageNet. However, the results were contrary to expectations. Both training time and test accuracy of the ResNet-18 model were poorer compared to the CNN model trained from scratch. This discrepancy can be attributed to several factors.

First, ResNet-18 is a relatively deep model with large number of layers compared to the simple CNN built from scratch. Despite the advantage of transfer learning where earlier layers retain useful feature representations, the model's deep architecture requires substantial computational resources, leading to longer training time than expected. Additionally, the large number of parameters increases memory usage, slowing down training.

Second, ResNet-18's pre-trained weights are fine-tuned for complex datasets like ImageNet. However, CIFAR-10 is a relatively simpler dataset with fewer classes (10 classes of 32x32 images). The complexity of ResNet-18 might be too high for CIFAR-10, and though the pre-trained weights could be helpful, they may not transfer optimally to this new task. In this case, ResNet-18 may need more extensive fine-tuning to adapt to the simplicity of CIFAR-10, which can lead to longer training times and potential issues like overfitting, where the model becomes too specialised on the dataset. On the other hand, training a CNN from scratch means the model learns from the ground up, focusing on the specific features of CIFAR-10. Since the model is simpler, there is less chance of overfitting to complex features not relevant to the CIFAR-10 task. While this model might have lower capacity to capture nuanced features, the simpler architecture can often generalize better to simpler datasets.

Third, the learning rate might not have been the most suitable for the transfer learning process, leading to slower convergence. In contrast, the CNN from scratch may not face this issue because it starts with random weights and gradually learns.

The trade-off between transfer learning and training a CNN from scratch also revolves around adaptability. While transfer learning can accelerate training for complex datasets, it might not be as effective for simpler datasets. On the other hand, training a simpler CNN from scratch ensures the model is learning features tailored specifically for the CIFAR-10 dataset.

Conclusion

While transfer training is generally expected to reduce training time and improve performance, the specific characteristics of CIFAR-10, complexity of ResNet-18 and potential mismatches in fine-tuning approach likely contributed to the poorer performance of ResNet-18. To yield better results with transfer learning, more optimised training strategies might be necessary.