Introduction

The CIFAR-10 dataset is a widely used benchmark for evaluating image classification models. It contains 60,000 32x32 colour images divided into 10 classes, making it ideal for testing the effectiveness of convolutional neural networks (CNNs). To do so, I conducted two experiments to assess the performance and impact of transfer learning compared to training a CNN from scratch for classifying CIFAR-10 images. The first involved training a CNN from scratch, while the second utilised a pre-trained ResNet-18 model in two different ways: (1) freezing all layers except the fully connected (fc) layer and (2) unfreezing the deeper layers (layer3, layer4, and fc) for finetuning. By comparing these approaches, this study examines how different levels of transfer learning influence classification performance and training efficiency.

Methodology

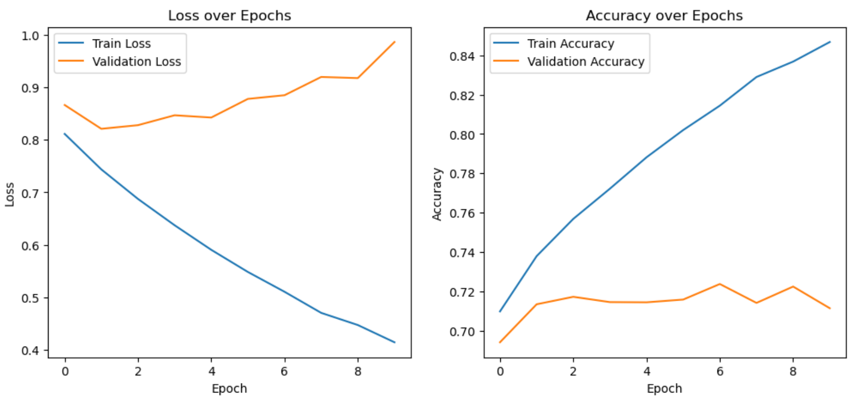
The dataset was split into 50000 training and 10000 test images. The CNN trained from scratched consisted of 2 convolutional layers to extract features, an activation function (ReLU), 2 pooling layers using max pooling and 2 fully connected layers for classification. ResNet-18 model on the other hand, kept all its layers intact except for the fc layer to output just 10 classes to align with the CIFAR-10 dataset. Images loaded into ResNet-18 were normalised with the mean and standard deviation values of the ImageNet dataset to ensure consistency between its original training and its application in this experiment. Meanwhile, the early layers were frozen so that only the later layers could be finetuned on CIFAR-10. In the first approach where only fc was trained, this was intended to assess how well the pre-trained feature extractor performed on CIFAR-10 with minimal adaptation. The second approach were layers 3, 4 and fc were unfrozen was done to allow the model to learn dataset-specific features while still leveraging the lower-level representations learned from ImageNet. This comparison was done to determine the extent of improvement in performance because of additional finetuning.

Results and Comparison

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

	CNN from scratch	**ResNet-18 (model 1)**	**ResNet-18 (model 2)**
Training time (mm:ss)	36:47	86:35	284:58
Test accuracy (%)	71.13	45.33	78.48

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. However, the results were contrary to expectations. Compared to the CNN model trained from scratch, both versions of ResNet-18 required significantly longer training times, with the model that had more layers unfrozen taking even longer. This can be attributed to ResNet-18's deeper architecture, which contains many layers and parameters, increasing both computational demands and memory usage. Even though transfer learning allows leveraging pre-trained weights, backpropagation through a more complex model still results in a slower training process, especially when more layers are unfrozen.

When only the fc layer was trainable, the majority of ResNet-18 acted as a fixed feature extractor, which significantly reduced the number of computations, making training relatively faster. However, because only the last layer was adapted to CIFAR-10, the model had limited ability to modify feature representations learned from ImageNet, which may not have aligned well with CIFAR-10's simpler objects. As a result, this approach yielded the lowest accuracy.

In contrast, when layers 3, 4, and fc were unfrozen, the model could update more complex feature representations, improving its ability to adapt to CIFAR-10, leading to better accuracy than the fully frozen ResNet-18 model. However, this came at the cost of much longer training times due to the need to update deeper feature maps. Despite this improvement, the performance was not significantly better than the CNN trained from scratch, which suggests that ResNet-18's pre-trained features were not an ideal match for CIFAR-10.

On the other hand, training a CNN from scratch meant that the model learned features specific to CIFAR-10 without being constrained by pre-trained representations. Since CIFAR-10 is relatively simple compared to ImageNet, learning from scratch allowed the model to focus on dataset-specific patterns, avoiding the risk of carrying over irrelevant features from ImageNet. This resulted in faster training and comparable accuracy than ResNet-18.

The findings highlight important considerations when applying transfer learning. The first approach, where only the fc layer was trained, was fast but underperformed due to limited feature adaptation. The second approach, which unfroze additional layers, improved accuracy but came at the cost of longer training times. However, neither approach led to a clear advantage over training a CNN from scratch, suggesting that transfer learning's benefits are not universal and depend on dataset complexity, similarity to the pre-trained dataset, and computational constraints.

Conclusion

These findings reveal that while transfer learning is a powerful tool, it is not always beneficial, particularly when working with relatively small and simple datasets like CIFAR-10. Training a model from scratch allows it to learn features specifically for the given dataset without inheriting potentially suboptimal pre-trained representations. Conversely, transfer learning is most effective when the target dataset is large and shares similarities with the pre-trained dataset. Thus, while transfer learning is often assumed to improve performance, it must be carefully evaluated based on dataset characteristics.