A Transfer Learning Approach to Classifying the Intel Image Dataset

Stephen Kakuda

February 2025

1 Introduction

This assignment applies transfer learning techniques to perform image classification on the Intel Image Classification dataset. Transfer learning involves fine-tuning a pre-trained model on a new, specific dataset, enabling us to leverage the knowledge learned from a larger, related dataset. By applying this approach, we aim to achieve efficient and accurate classification results for the given task. The model of choice for this assignment was Resnet-50, a popular model for image recognition tasks. In this assignment, the accuracy of using the pre-trained model as a fixed feature extractor is compared to the accuracy of fine-tuning the entire CNN model.

2 The Data

The Intel Image Classification dataset was used for this assignment, which contains approximately 25,000 images with a total of 6 classes: building, forest, glacier, mountain, sea, and street. A custom Dataset class was implemented to apply the appropriate transforms and prepare the data for the pipeline. The provided prediction set was discarded due to the lack of labels, and only the train and test sets were used for this assignment. The train set containing approximately 14,000 images was shuffled and then divided into train and validation sets using a 85-15 split. In order to prepare the data, a series of transforms were applied to the images. The first transformation resized the image to be of shape (244, 244) which is one of the expected formats for the ResNet model. The last two layers converted the image to tensor format and normalized the values based on the Intel Image distribution.

3 Experimental Setup

In order to asses the effectiveness of transfer learning, two different models were trained. The first was a fixed feature extractor that only trained the final output layer of the ResNet-50 model. All the parameters but the final output layer were frozen before training. The second model used the same base ResNet-50 but all parameters were fine-tuned by the optimizer.

Given that the ResNet-50 model was pre-trained using ImageNet, a lower learning rate of 0.0001 and momentum of 0.9 were used to prevent gradient exploding. Another precaution to prevent exploding was to clip the gradients, which ended up having a significant effect. A standard Stochastic Gradient Descent optimizer was used with a Cross Entropy loss function to train the models. Batch sizes were set to 32 and 10 epochs were used per model. While larger batch sizes and more epochs could improve performance of the models, these settings were able to achieve respectable results while reducing total runtime.

4 Experimental Results

The following results were returned after 10 epochs of training and were tested using the validation set containing images that neither model had not seen yet.

Model Type	Fixed Feature Extractor	Finetuned CNN
Validation Accuracy	88.80%	90.96%
Final Loss Score	0.4717	0.3177

5 Discussion

Overall, both models achieved high validation accuracies after only 10 epochs of training. As expected, the fully fine-tuned model performed better than the fixed feature extractor. However, the models' validation accuracies were only separated by less than 3%. The loss function measures how different the model's prediction was from the expected values and here we can see a greater difference in performance. The fine-tuned CNN had a lower loss score by 0.154, indicating it was better able to learn the features of the dataset. Overall, both models demonstrated how transfer learning can have a significant impact leading to higher performance.

ResNet-50 was pre-trained on ImageNet, a dataset of more than 14 million images with over 20,000 classes. This gives ResNet-50 a strong advantage for image classification, as it has already learned many useful features, making a smaller datasets like the Intel Image Classification dataset easier to train for.

Future work with these models would include training with more epochs and comparing the effectiveness of using transfer learning with different models like AlexNet. Another interesting metric to measure would be how long it would take for a ResNet-50 model that is not pre-trained to reach the same level of accuracy as the pre-trained models.