Assignment 2

Convolution

Report

Introduction:

In this analysis, we evaluate the effectiveness of developing a convolutional neural network (CNN) using the Cats vs. Dogs dataset. The study focuses on determining the optimal sample sizes and approaches for model construction to enhance performance during the model-building phase.

Methodology:

We developed six models from scratch and three pre-trained models, experimenting with various configurations. These models differ in terms of the number of layers, neurons, optimizers, dropout rates, and other hyperparameters to assess the impact of each on performance.

Scratch Models

Validation Accuracy, Test Accuracy, and Test loss

Model no	Training sample size	Validation and Test sample size	Validation Accuracy	Test Accuracy	Test Loss
Model 1	1000	500,500	0.772	0.771	0.509
Model 1a	1000	500,500	0.833	0.811	0.394
Model 1b	1000	500,500	0.738	0.750	0.509
Model 1c	1000	500,500	0.823	0.823	0.407
Model 2	1000	500,500	0.883	0.879	0.654
Model 3	1000	500,500	0.894	0.889	0.475

Pre-Trained Models

Validation Accuracy, Test Accuracy, and Test Loss

Model	Validation Accuracy	Test Accuracy	Test Loss
Pre-trained Model - 1000 Training samples	0.985	0.981	0.11
Pre-trained Model - 5000 Training samples	0.97	0.977	0.13
Pre-trained Model - 10000 Training samples	0.99	0.995	0.02

Overall Observations and Conclusions

Scratch Models versus Pre-trained Models:

The pre-trained models perform considerably better, with very high margins from the scratch models, both at validation and test accuracy. This is evident by the highest test accuracy of 0.995 by the pre-trained model with 10,000 training samples against the best accuracy of the scratch model of 0.889 from Model 3.

Impact of Training Sample Size (Pre-trained Models):

In the case of pre-trained models, there is an increase in both the validation and test accuracy with increasing sample size of training. The validation accuracy was 0.99 and the test accuracy was 0.995 for 10,000 samples in a pre-trained model, again higher for all the models. Correspondingly, test loss also decreased significantly with increasing sample sizes from 0.11 to 0.02 with increased sample sizes from 1,000 to 10,000 samples, respectively.

Performance of Scratch Models:

For the scratch models, performance varied somewhat, with Model 3 having the best validation accuracy at 0.894, with a test accuracy of 0.889.
Model 2 achieved a comparable validation accuracy of 0.883 but resulted in a much higher test loss of 0.654, which could potentially indicate overfitting.
Meanwhile, Model 1a and Model 1c show a relatively balanced validation-test accuracy of about 0.81-0.83, while the respective losses for the test set are smaller and, hence more generalizable.

Overfitting in Scratch Models:

The accuracies for the scratch models are very similar in both the validations and testing. However, outstanding Model 2 has its test loss at 0.654, well higher from others, which directly indicates overfitting to the training data and not generalizing on unseen data.

Comparison of Test Loss:

Indeed, the pre-trained models perform with lower test losses across the board, for which the lowest is 0.02 in the case of 10,000 training samples. On the other side, scratch models have shown characteristics of having higher test losses; whereas Model 2 went out of the chart with a value of 0.654, the rest of them-1a and 1c-managed to stay above with values of 0.394 and 0.407, respectively.

Conclusion:

In general, even with small datasets, pre-trained models can outperform scratch models both for accuracy and test loss. If anything, pre-trained models are capable of higher generalization and a reduction in test loss with larger sample sizes. Scratch models showed relatively good results in the case of Model 1a and Model 1c, but still fall behind pre-trained nets, demonstrating how vital transfer learning is for image classification tasks such as cats-vs-dogs.