Assignment 2: Convolution

1. Executive Summary of Findings

The present experiment investigated the issue of sample size in terms of convolutional neural network (convnet) training, using a randomly chosen to set of images of cats and dogs (dataset of cats vs dogs) to train a convolutional neural network on the classification of images of all animals.

All models were trained and developed only with the given training dataset (x_train) and various sample sizes. Validation (500) as well as test (500) sets was the same across all experiments. The overall idea was to trace the effect of the number of training data available on the model accuracy, overfitting, and generalization in general.

To improve the performance of the models and minimize overfitting, several optimization and regularization techniques were introduced, such as data augmentation, dropout layers, batch normalization, and early stopping.

The scratch trained model used in the study to ensure model performance and reduce overfitting: several regularization mechanisms, such as data augmentation (Random Flip, Random Rotation, and Random Zoom) and Dropout layers to ensure network does not learn the data it was trained on, were implemented.

In case of the pretrained model, there was a two-phase training strategy. During the initial step, the training of a new classification head was done but the Xception base network was not trained but only the higher layers were updated to acquire task-specific features. The second step involved fine-tuning the whole network with very low learning rate to slowly fine-tune the already pretrained weights without overfitting and forgetting the previously trained representations.

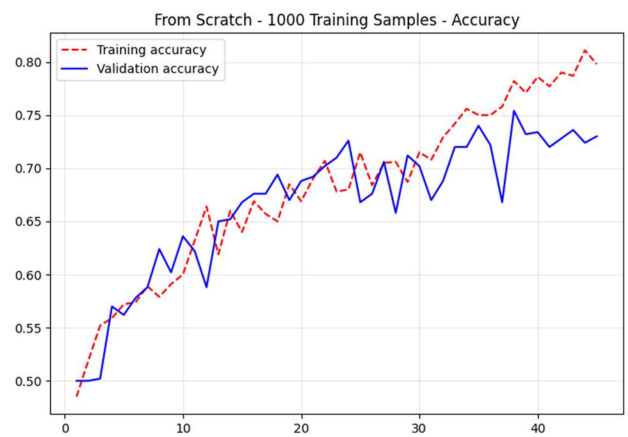
Results in Table

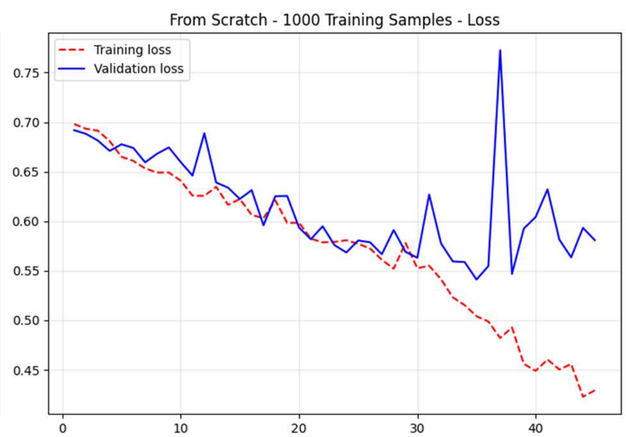
Train size	Scratch Test Accuracy	Pretrained Test Accuracy	Improvement
1000	0.7180	0.9500	+0.2320
1500	0.7660	0.9700	+0.2040
2000	0.8460	0.9660	+0.1200

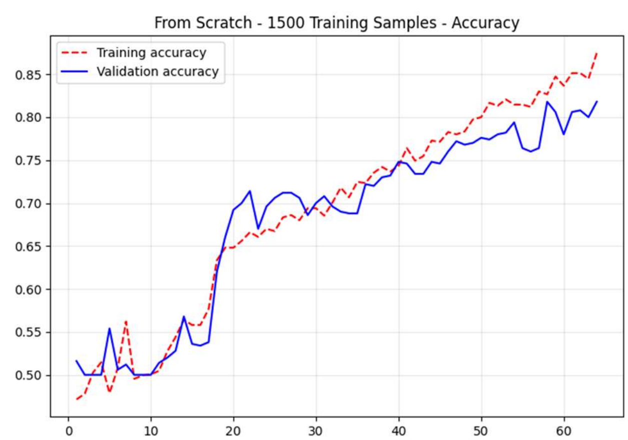
2. Performance Analysis

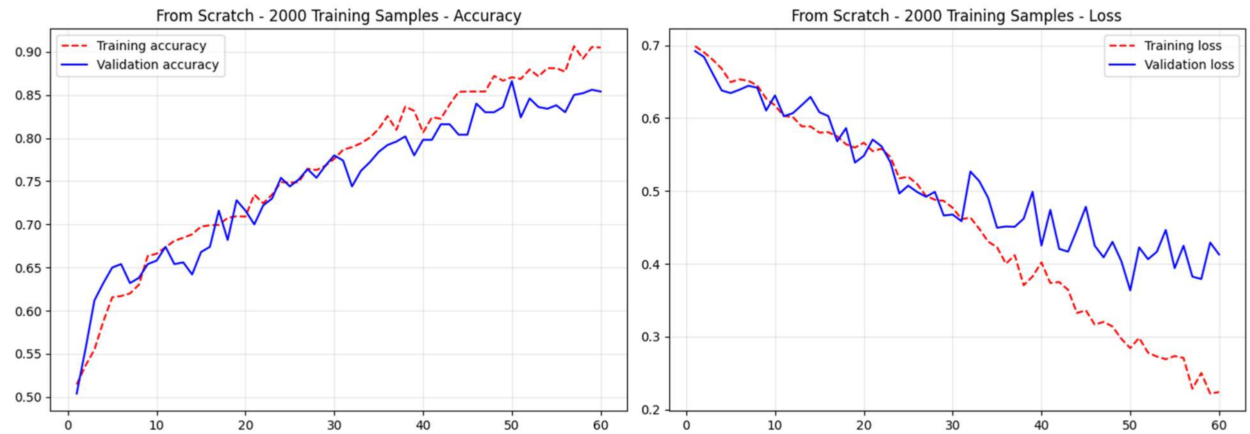
A. Training from scratch:

The purely trained model showed an evident and stable correlation between the model's performance and the size of the training sample. The level of validation as well as test accuracy improved significantly with more data implying that the model was becoming aware of more solid and generalizable visualizing patterns.









Performance summary:

There was maximum test accuracy of 84.60 by the model which was trained with 2,000 images. At smaller scales, the performance was much lower with 71.80% accuracy with 1,000 samples and 76.60% with 1,500 samples.

Interpretation:

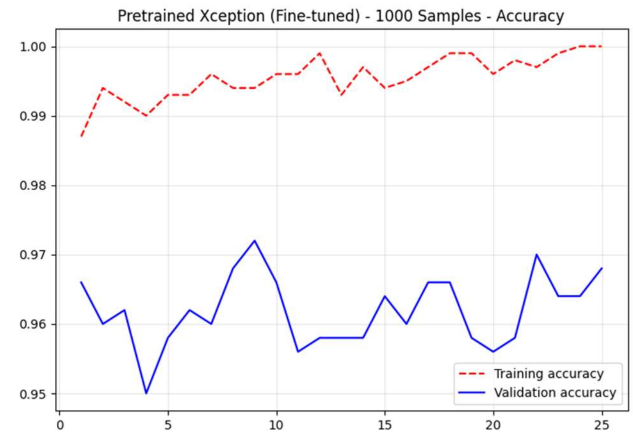
This is predictable as CNN trained from scratch uses no initial knowledge. It has to acquire all the visual features, such as simple shapes as edges and texture and more complicated ones like ears or snouts, only based on given data.

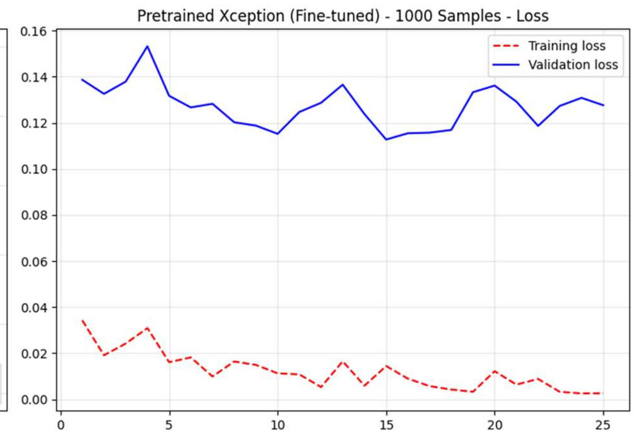
The network was under-trained on 1,000 training samples, and generalization was poor and overfitted as can be determined by training graphs (e.g., scratchtrain500). The more samples, 1,500 and 2,000, the model had more occurrences of different examples, which enabled it to build stronger internal representations and increase test accuracy.

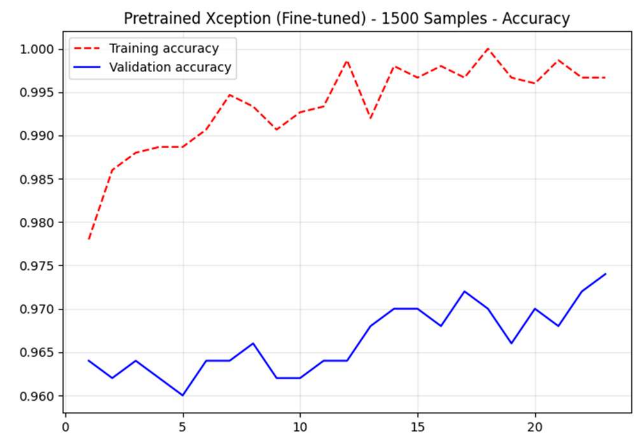
The uniform improvement of the performance with the growing sample size demonstrates that the scratch-trained model is data-hungry and would probably keep on improving even with the larger datasets. This approach simply meant that the amount of data had a direct effect on the level of learning and the ability to generalize.

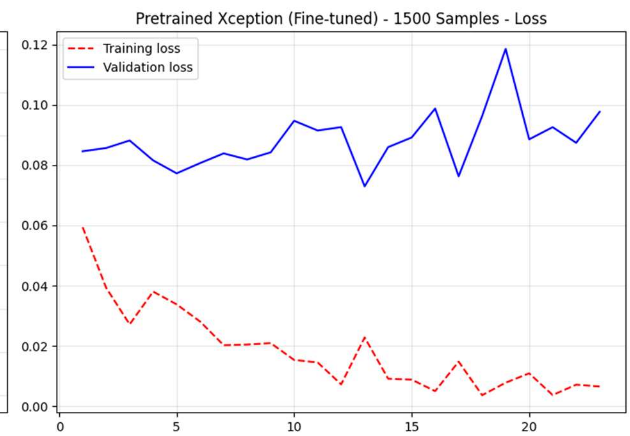
B. Pre-Trained Model (fine-tuned/ Xception):

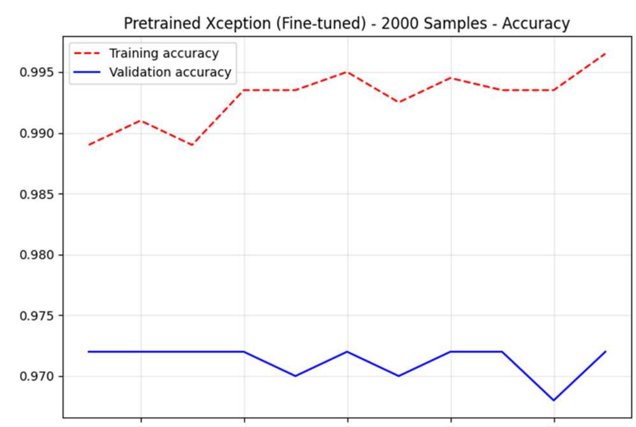
The transfer learning model, pretrained Xception, obtained outstanding results of all training sizes. In contrast to the scratch-trained model, it was highly performing even with a small amount of data and reached its highest test accuracy of 97.00% with only 1,500 training images.

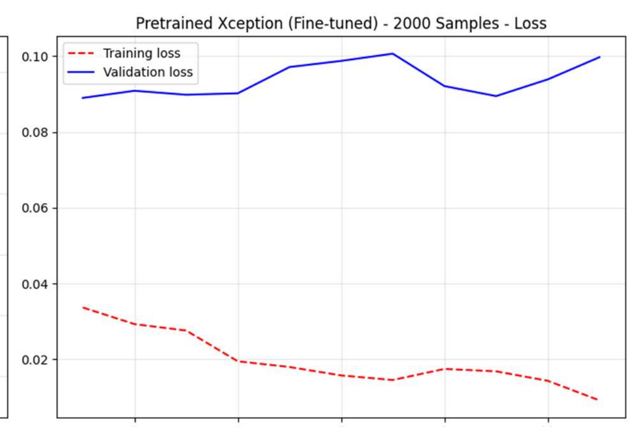












Performance summary:

The trained Xception model had a high performance in all dataset sizes. It had a test accuracy of 95.0 with 1,000 training samples following the fine-tuning. Raising the number of samples to 1,500 increased accuracies to 97.0 which is the optimal performance of the model. After 2,000 samples, the accuracy leveled at 96.6 which was the situation at which the model had effectively hit its near-optimal generalization.

Interpretation:

These better performance figures are attributed to the transfer learning because the Xception baseline network had been trained on the extensive ImageNet dataset. It is important to note that the model already knew rich and generalized feature representations for image recognition but then was modified in accordance with the Cats vs. Dogs task.

Phase 1 (Frozen Base):

The first experiment was done by initially freezing the Xception base and only training the new head of classifiers. At this point, the model was already quite accurate with 1,000 samples (92.0) and 1,500 samples (97.2). This shows that the already trained convolutional layers already learned highly transferable features that were pertinent to the separation between cats and dogs.

Phase 2 (Fine Tuning):

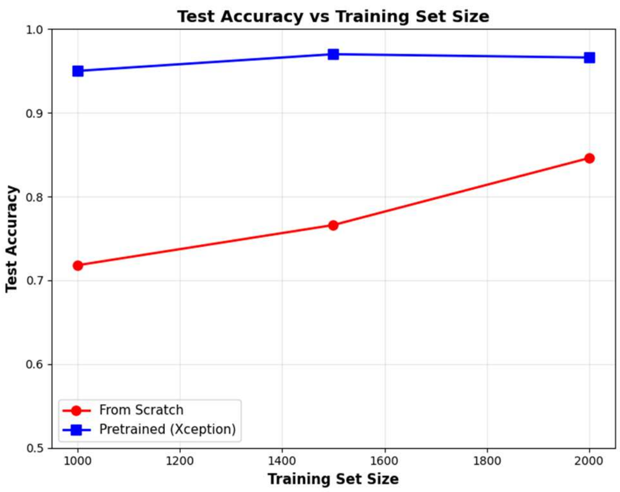
During the second phase, the whole model was unfrozen and recoined with extremely low learning rate (1e-5). This enabled making fine-tuning of the pretrained weights to suit the Cats vs. Dogs dataset. The fine-tuning step also made the model more accurate, e.g., the 1,000-sample model increased its accuracy by 2 to 92.0 percent, and the 2,000-sample model increased its accuracy by 2 to 96.0 percent.

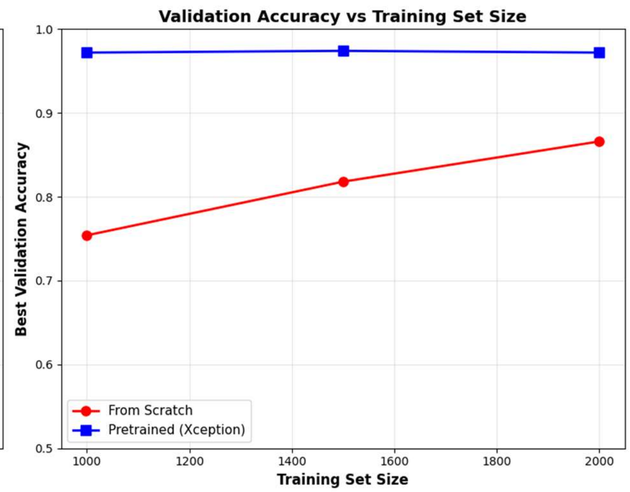
Performance format:

The pretrained model had a performance plateau during growth in the size of the datasets. It was 95 per cent accurate using only 1,000 samples, and further data did not show significant improvements. It implies that the pretrained Xception network already has good general visual knowledge, and only slightly new information is necessary to be reconfigured into a new classification task.

Accuracy of the Tests vs. The Size of the Training Set.

The Test Accuracy vs. Training Set Size graph is best suited to visualize the way the two models, trained starting with scratch as well as pretrained (Xception), react to higher training data volumes. The plot shows the two separate lines which are the red line of the model that is trained starting with a blank model and the second blue line of the model that is pretrained. Their combination creates a compelling story that incorporating the volume of data and transfer learning has on the performance of models.





3. Graph Analysis

The Performance Gap:

The gap between the blue and the red line is large and striking all through the graph. This gap is a visual depiction of the strength of transfer learning. Even based on the initial data point, the pretrained model was significantly more accurate than the scratch-trained model, and it obtained almost perfect accuracy with limited data.

Scratch Model (Presented in Red line):

The steep uphill curve illustrates that the red line began with an accuracy of 71.8 at 1000 training samples and reached 84.6 at 2000 training samples. This sharp movement implies that the CNN that was trained on scratch is very data dependent. It acquires visual features in a ground up manner meaning that the better more examples are given to it, the more it will continue to improve its performance. The trend indicates that with more data, the accuracy would most probably keep increasing.

Pre-Trained Model (Presented in blue line):

Conversely the blue line begins very high - 95.0 percent accuracy at only 1,000 samples - and is relatively flat, reaching 97.0 percent accuracy at 1,500 samples and then slightly leveling off at 96.6 percent accuracy at 2,000 samples. This plateau shows that pretrained models are not restricted by the availability of data. Since the Xception base was trained on millions of various images, it will be able to extract intricate visual features even when using very small datasets.

Interpretation:

The graph brings out a basic understanding that, whereas the models trained fresh need a lot of data to generalize well, the rest of the models utilize prior knowledge to deliver outstanding performance with minimum training samples. The visual impact of the blue line as compared to red line is that transfer learning facilitates quick and efficient adaptation, and this is the reason why it is the best when it comes to data limited image classification.

4. Conclusion: The Connection Between Network Selection and Sample Size

The results of this experiment unequivocally show that the choice of network architecture and training sample size are directly and significantly correlated. In addition to fulfilling the assignment's goals, the outcomes provide useful information about how model selection and performance in deep learning-based image classification are influenced by data availability.

1. When data is scarce, trained models predominate.

The pretrained Xception model performed noticeably better when dealing with smaller datasets, which ranged from 1,000 to 2,000 images. In contrast to the model trained from scratch, which had test accuracies of only 71.8% to 84.6%, the pretrained approach had test accuracy ranging from 95.0% to 97.0%. The power of transfer learning, in which a pretrained network efficiently reuses rich, general-purpose visual representations acquired from a large dataset like ImageNet, is demonstrated by this striking contrast.

The pretrained model has a thorough understanding of basic image structures, including edges, shapes, textures, and color patterns, in contrast to a model trained from scratch, which begins with randomly initialized weights. It can therefore quickly adjust to new tasks with little further training. The pretrained Xception model demonstrated exceptional efficiency and robustness, achieving near-state-of-the-art accuracy even with sparse data. This makes it the best option in situations where gathering data is costly, time-consuming, or subject to other limitations.

The validation curves also showed that the pretrained model had low variance and little overfitting. This stability suggests that transfer learning fosters improved generalization to unknown data in addition to speeding up training.

2. Scratch Training Requires a Lot of Data and Resources

The scratch-trained model, on the other hand, demonstrated a significant reliance on the size of the training sample. With bigger datasets, its performance steadily improved, but it was always noticeably worse than the pretrained model. With 1,000 samples, the model's accuracy was only 71.8%; with 1,500 samples, it slightly improved to 76.6%; and with 2,000 samples, it peaked at 84.6%.

The data-hungry nature of deep learning models that are trained from scratch is reflected in this consistent but constrained improvement. Such models must extract all visual patterns, from simple edges to intricate object components, from the given data since they lack any pre-learned features. To get close to the accuracy attained by pretrained networks, they thus need tens or even hundreds of thousands of training examples.

Furthermore, training from scratch requires a lot of computation and time, and it frequently calls for more powerful hardware, longer convergence times, and extensive hyperparameter tuning. This approach is rarely cost-effective in real-world applications, such as early-stage research or industry projects with little data.

3. The Wider Consequence: Effectiveness via Transfer Learning

The findings of this study are consistent with a larger trend in contemporary computer vision, which is the use of pretrained networks as universal feature extractors rather than building large models from the ground up. Researchers and practitioners can achieve high accuracy with little data and computation thanks to transfer learning, which drastically reduces the need for large, labelled datasets.

This benefit is illustrated both visually and numerically in this experiment by the difference in performance between the pretrained and scratch-trained models. While the scratch-trained model continued to improve but fell behind, the pretrained Xception model quickly achieved optimal performance, with diminishing returns after 1,500 images. This relationship demonstrates that pretrained architectures are both feasible and scalable solutions for real-world image classification problems because they can effectively adapt to new tasks with little fine-tuning.