Convolutional Report

Alisha Ghatani

Introduction:

The purpose of this assignment is to build a convolutional neural network (CNN) to classify images of cats and dogs by applying convents to image data. The goal is to evaluate the relationship between training sample size and model performance, comparing training a network from scratch versus using a pretrained model.

Problem Definition:

The objectives binary classification: accurately identifying images as either a “cat” or “dog”. This task involves implementing and optimizing CNN architectures to enhance performance and reducing overfitting.

Methods:

Data Preparation:

The dataset used for this experiment is the Cats and Dogs dataset. The data was divided into three subsets:

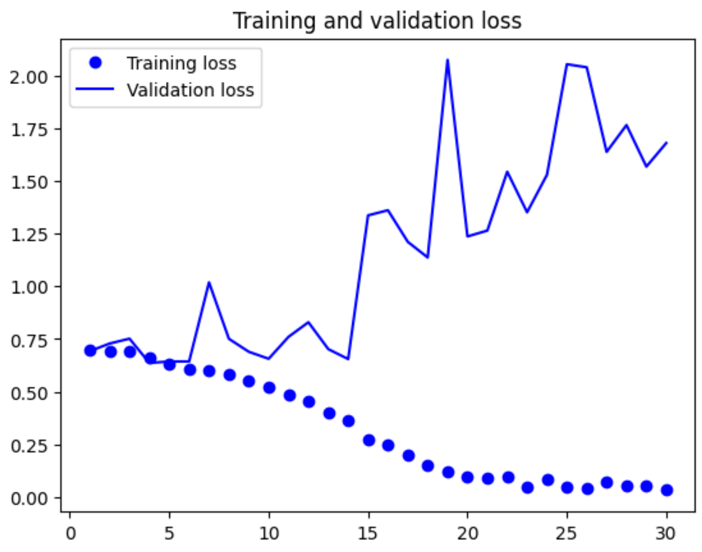
- Training set: 1000 images
- Validation set: 500 images
- Test set: 500 images

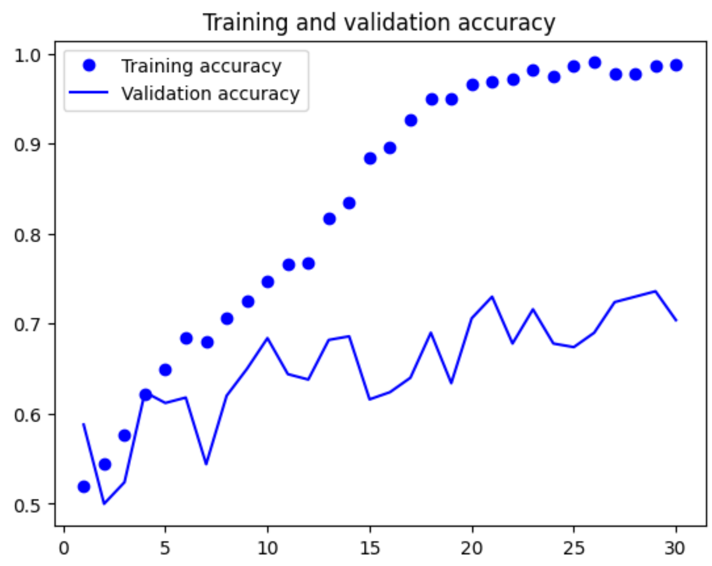
A CNN model was implemented from scratch using multiple Conv2D and Maxpooling2D layers for features extraction. To reduce overfitting , data augmentation techniques such as rotation, flipping, and zooming were applied. Images were resized to 150x150 pixels, converted to RGB pixel grids, and normalized by scaling pixel values to the [0,1] range to improve model performance.

Data Preprocessing:

Key steps include:

- Access the image files.
- Decode the JPEG content to obtain RGB pixel grids.
- Normalize the pixel values to fit within the [0, 1] range, as neural networks perform better with smaller input values. This is achieved by scaling the pixel values, which originally range from 0 to 255.



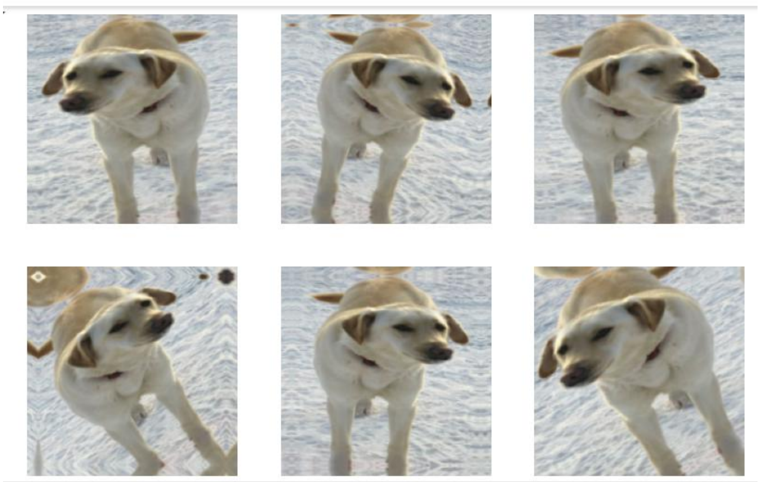


Increasing Training Sample Size

- Training sample size increased beyond 1000 images while keeping validation and test samples constant.
- Application of optimization techniques to fine-tune the model.

Data Augmentation

We aim to utilize data augmentation techniques to enhance the accuracy of our model. Data augmentation involves generating new data from existing training samples by introducing random variations, allowing for improved performance even with limited datasets. By applying this method, the model is exposed to different versions of images during training that it hasn't encountered before, facilitating better generalization. To achieve our objective, we intend to randomly apply various transformations such as flipping, rotating, and zooming to the images in the training set. This approach generates diverse versions of the original images, thereby enriching the dataset and strengthening the resilience of our model.



Optimal Training Sample Size

- Experimentation with different training sample sizes to identify the ideal size for optimal performance.

Training from Scratch Models Table

Training Sample	Models Type	Test accuracy	Test loss	Validation Accuracy	Validation loss
1000	Baseline Model 1	57.8%	0.644	67.0%	0.607
1000	Augmentation Model 1	75.6%	0.544	76.8%	0.498
1500	Baseline Model 2	66.6%	0.627	68.2%	0.585
1500	Augmentation Model 2	77.4%	0.508	80.2%	0.432
2000	Baseline Model 3	69.2%	0.637	73.4%	0.526
2000	Augmentation Model 3	82.2%	0.465	82.0%	0.418
1700	Baseline Model 4	61.0%	0.656	62.4%	0.637
1700	Augmentation Model 4	77.8%	0.50	78.8%	0.455
600	Baseline Model 5	62.0%	0.692	65.4%	0.617
600	Augmentation Model 5	71.0%	0.614	72.6%	0.538

Model 1 (1000 Training Samples)

It performs moderately, as the Baseline Model has a test accuracy of 57.8% and a validation accuracy of 67.0%, not very useful in distinguishing different testing data. The test loss is about 0.644 after the training, while validation loss is 0.607; this may be interpreted that the model is not generalizing well to new, unseen data. The Augmentation Model has performed exceptionally well by attaining test accuracy of 75.6% and validation accuracy of 76.8% with the help of data augmentation. The decrease in validation loss from 0.607 to 0.498 confirmed that augmentation has some impact on decreasing overfitting. This indicates that under a small dataset condition, augmentation is of utmost importance to enhance model performance.

Model 2 (1500 Training Samples)

Testing with a new training sample of 1500 images shows that Model 2 offers an advantage over Model 1. Consequently, base accuracy improved beyond expectation by reaching a test accuracy of 66.6% and validation accuracy of 68.2%. The test loss is 0.627 and the validation loss is 0.585, indicating that the model is slightly better at generalization. When augmentation is applied, the performance sees yet another boost now with the Augmentation Model reaching a test accuracy of 77.4% and a validation accuracy of 80.2%; the validation loss now drops from 0.585 to 0.432. This indicates that augmentation does indeed help the model learn robust features, hence preventing overfitting. This means that more training samples improve accuracy, but augmentation is still required for optimal performance.

Model 3 (2000 Training Samples)

At 2000 training samples, the Baseline Model realizes a test accuracy of 69.2% and a validation accuracy of 73.4%, with further improvement in the current round. The validation loss reduces to 0.526, which shows better generalization of this model compared to earlier ones. Applying augmentation allows the model to reach its highest performance so far, in which the Augmentation Model attains an impressive 82.2% test accuracy and 82.0% validation accuracy. The validation loss record is the lowest among all models (0.418), which makes it believable that this model generalizes well. The best balance primary accuracy and loss has been achieved by this model with 2000 training samples and augmentation, making it the most optimal choice above all others tested.

Model 4 (1700 Training Samples)

Interestingly enough, with 1,700 training samples, the Baseline Model performs worse than Model 3 by under-achieving at 61.0% test accuracy and 62.4% validation accuracy. Validation loss counts 0.637, which is higher than that of the model with 2000 samples, suggesting overfitting. Significant improvement of the Augmentation Model to 77.8% test accuracy and 78.8% validation accuracy. Validation loss reduces from 0.637 to 0.455, suggesting that augmentation is still an important factor that plays a role in reducing overfitting. Nevertheless, this model did not perform as well as the one from the augmentation with 2000 samples, downloading further augmenting with sample size establishes performance further.

Model 5 (600 Training Samples)

The Baseline Model flop with 600 training samples, resulting in 62.0% test accuracy and a validation accuracy of 65.4%. The validation loss counts to 0.617, which indicates a very pronounced overfitting tendency of the model. With augmentation, the Augmentation Model acquires 71.0% in test accuracy and 72.6% in validation accuracy, indicating significant improvement. This model still cannot hold up with large datasets, however. Results show 600 samples to be inadequate to train a CNN from scratch, even if augmented.

Overall, an increased training sample size improves model performance, but data augmentation is of constant benefit. The best-performing model among all the ones in the table is Augmentation Model 3 with 2000 training samples, attaining the maximum test accuracy (82.2%) and minimum validation loss (0.418). This suggests that 2000 training samples with augmentation have the best trade-off between accuracy and generalization. On the other hand, models with smaller datasets like Model 1 and Model 5 show how this practice limits overfitting and provides generalization when augmentation is not adopted.

Pretrained Network

Taking VGG16 with the implementation of data augmentations employs pretrained features from ImageNet and makes the model more loaded with increased variance in data due to transformations such as rotations and flips. This enhances generalization as well as reduces overfitting and results in increased accuracy, making the model more resilient in tasks like Cats vs Dogs classification.

Pretrained Models Table

Training Sample	Models	Optimization technique	Test Accuracy	Test Loss	Validation Accuracy	Validation Loss
1000	Model 6	Data Augmentation	96.8%	5.11	97.4%	1.987
1500	Model 7	Data Augmentation	97.0%	6.138	97.6%	2.840
2000	Model 8	Data Augmentation	97.1%	8.022	98.6%	1.540
1700	Model 9	Data Augmentation	97.2%	3.671	98.2%	1.802
600	Model 10	Data Augmentation	95.8%	5.620	98.0%	2.176

Starting with Model 6 trained with 1000 samples, pretraining with data augmentation gives a test accuracy of 96.8% and validation accuracy of 97.4%. The test loss was at 5.11 and the validation loss at 1.987. Performance on a small dataset is reasonable; augmentation must have helped in preventing overfitting in this case with such limited data.

Training with 1500 samples on Model 7 increased the test accuracy to 97.0% while validation accuracy was at 97.6%. However, the test loss became greater at 6.138 in conjunction with the validation loss being equal to 2.840. The accuracy therefore has increased with more samples, yet the loss started to climb, indicating possible instability or slight overfitting despite applied augmentation.

Model 8 had the highest test accuracy since 97.1% and the highest validation accuracy since 98.6% with 2000 training sample size and lowest validation loss since 1.540. Unfortunately, the test loss is at an extreme 8.022. This indicates that the model may be overfitting to the validation set or having difficulty generalizing to unseen test data despite augmentation measures.

Model 9 emerged with the best balance when trained with 1700 samples. It attained the greatest test accuracy while at the same time showing the least test loss, being 97.2% and 3.671 respectively. Validation accuracy at 98.2% with a loss of 1.802 clearly indicates benefits to this model by pretraining and data augmentation, in that the model generalizes best in that field.

Concluding, finally, Model 10 barely works at all. With 600 samples, tested with a 95.8% accuracy-lowest among them all-and a loss of 560.There seems to be frustration between its validation accuracy of 98% and high yet generalization inadequacy of 2.176. This is indeed proving the lack of sufficient data with at least augmentation and pretraining considerations.

Combining pretrained models and using data augmentation was found to improve performance considerably. However, there is a sweet spot in training sample size-Model 9 with 1700 samples that hits this balance perfectly. Model 9 generalizes well, avoids overfitting, and achieves the best overall performance. On the other hand, models with very few provided samples, especially like Model 10, become unreliable for actual deployment.

The comparison table shows the performance of a best model from scratch and pretrained model.

Training from scratch vs Pretrained Model

Training Sample	Models	Optimization technique	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
2000	Augmentation model 3	Data Augmentation	82.2%	0.465	82.0%	0.418
1700	Model 9	Data Augmentation	97.2%	3.671	98.2%	1.802

When compared, the Pretrained Model 9 completely outshines the Scratch Model 3 in the task of differentiating among Cats vs Dogs. It achieves a significantly higher validation accuracy (98.2%) rendering the scratch model's (82.0%)validation accuracy an insignificance when compared with it to show that it is way better generalization to identify critical patterns such as fur types and shapes. Much fewer samples can be processed by the pretrained model, which is 1700 against 2000 samples by the scratch model, and still signature higher performance because it learned from a larger dataset that it has to learn quickly and more efficiently from it.

However, the scratch one underfits; fails to accommodate complex patterns and requires a much larger amount of data to deliver lower performance, while the pretrained model is speedier in training, considering that it doesn't learn from scratch primitive features, making it resource-wise cost-effective.

Overall, Pretrained Model 9 is the better choice due to its higher accuracy , faster learning speed, and efficient use of fewer resources.

SUMMARY

In this study, we investigated the relationship between training sample size and the choice of network architecture in image classification tasks. Through experiments with the Cats & Dogs dataset, we demonstrated the impact of sample size on model performance and highlighted the advantages and limitations of training from scratch versus using pretrained networks.

CONCLUSION

The relationship between training sample size and choice of network architecture is crucial in determining model performance. While larger training sample sizes generally lead to better performance, the effectiveness of pretrained networks with smaller sample sizes highlights the importance of transfer learning. Experimentation and optimization are essential to identify the most effective approach for a specific task and dataset.