

Assignment 3 Report – Convolutional Neural Networks and Transfer Learning

1. Objective

The key aim of this task is to familiarize ourselves with convolutional neural networks (CNNs), and transfer learning as they relate to image classification. In this assignment we constructed two separate models using the Cats vs Dogs dataset. One model was constructed from scratch while the other originated from a pretrained MobileNetV2 model. From there, we compared their results at various training data sizes to evaluate how much the sample size affects model accuracy, as well as data augmentation, dropout, regularization, and early stopping to lessen overfitting.

2. Dataset and Experimental Setup

The dataset utilized in this study was the Cats vs Dogs filtered dataset offered in TensorFlow. All images were resized to 160x160 pixels and normalized. The data was divided into 3 sections - the training data, validation, and test data. The validation and test data consisted of 500 images each while the training data varied across experiments with 1000, 1500, 1800, and 2000 images in order to experiment with the effect of dataset size on the model's accuracy.

For data augmentation, random flipping, rotation, and zooming were implemented to generate additional variety in the data and aid models' abilities to generalize. Adam optimizer and binary cross entropy were used as the loss function for both models. Early stopping, based on validation accuracy, was used to prevent further unnecessary training upon plateaued performance.

3. Model Architectures

We used two different models:

1. From-scratch CNN: This was a simple convolutional neural network built manually with several Conv2D and MaxPooling2D layers. It ended with fully connected dense layers and a dropout layer to reduce overfitting.
2. Pretrained MobileNetV2: This model was loaded with ImageNet weights and the top layers were removed. We added a global average pooling layer, a dropout layer, and a dense output layer. First, the base was kept frozen, and then it was unfrozen for fine-tuning using a smaller learning rate.

4. Results and Observations

The two models were trained with varying training sizes, and validation and test accuracy were recorded for each training scenario. The table below outlines the average performance obtained from running multiple training.

| Training Size | Model Type | Validation Accuracy | Test Accuracy | Overfitting | Remarks |
|---------------|-------------------------|---------------------|---------------|-------------|---|
| 1000 | From-scratch CNN | 70–75% | 68% | High | Small data, model memorized training data quickly |
| 1500 | From-scratch CNN | 78% | 76% | Medium | Better generalization with more data |
| 1800 | From-scratch CNN | 81% | 79% | Low | Performance improved with more samples |
| 2000 | From-scratch CNN | 82% | 80% | Low | Best result for scratch model |
| 1000 | Pretrained (Frozen) | 84% | 83% | Low | Good results even with small data |
| 1800 | Pretrained (Fine-tuned) | 88–89% | 87% | Low | Fine-tuning helped improve results |
| 2000 | Pretrained (Fine-tuned) | 90% | 89–90% | Very Low | Best overall performance |

Our results indicate that the validation and test accuracy increases as the training size increases for both models. The pretrained MobileNetV2 performed much better than the model created from training the MobileNetV2 architecture from scratch. This was particularly true when the dataset size was small. Because of the pretrained weights, fine-tuning improved average testing accuracy by a minor amount. This was over and above initial training using the pretrained weights, indicating some improvement resulted from fine-tuning the model to the new dataset.

5. Analysis and Interpretation

The CNN built from the ground up exhibited a great deal of reliance on the size of the training data. With just 1000 images, it quickly began to overfit to this data, and was unable to generalize to the more general data set it was tested on. With 2000 training samples, both validation and test accuracy improved, and potentially other performance metrics improved too.

The pretrained MobileNetV2 model was able to perform well even with a smaller dataset because it had already learned general visual patterns from the ImageNet dataset, which was a much larger combine of images than in the Cats vs Dogs setting. An appropriate selection in the fine-tuning phase with a lower learning rate helped the network adapt more accurately to the new data.

Overfitting was controlled for using data augmentation, dropout layers, and early stopping at some point too. The random dropout approach ensured the model did not simply memorize what it learned in the training, but made some sense of independent, authentic learning in the training data.

6. Conclusion

In this assignment, we compared two deep learning approaches for image classification: building a CNN from scratch, versus using a pretrained model. The results showed that transfer learning using MobileNetV2 performed better than the model trained from scratch in all training sizes. The pretrained model also achieved good accuracy with fewer training samples. Fine-tuning improved the results even further.

In summary, these results demonstrate that pretrained models are very beneficial when training with limited data. Because pretrained models use learned features from larger datasets, they save you time and offer increased accuracy. From-scratch models may have similar performance when enough data is available, but require careful regularization to avoid overfitting and will take much longer to train.

7. Summary of Learning

1. The size of training data is directly correlated with model performance.

Initially, the model struggled to generalize and exhibited some overfitting when there was a small training data set of approximately 1000 images. Continuing to increase the size of training samples at 1500, 1800, and 2000 images showed a steady increase in the validation and test accuracy.

It is evident here that increased diversity and representation of training samples enables the network to learn accurate and more stable patterns.

2. Transfer learning achieves more accurate results with fewer samples.

The pretrained MobileNetV2 model achieved greater accuracy when trained on 1000 images compared to the CNN model initially trained from scratch on 2000 images.

This shows the power of transfer learning in that the pretrained model already possessed rich visual features based on millions of images from the ImageNet dataset that allowed it to perform well on new and related tasks with little new data.

3. Fine-tuning can lead to performance improvements over a pretrained model.

Once we unfroze some of the deeper layers of the MobileNetV2 architecture and employed a smaller learning rate to fine-tune, the model began to specialize a little more to the Cats vs. Dogs dataset.

This fine-tuning was the final push the model needed to attain an accuracy nearly at 90%, which suggests that fine-tuning pretrained weights to the new domain has a better tendency of adapting and performing overall.

4. Overfitting can be overcome using many different methods, including data augmentation, dropout, and early stopping.

Although the model behaved as if it was memorizing the training data in the beginning, image augmentations were introduced (random flips, rotations and zooms), leading to a more diverse dataset.

Dropout regularization led to the model relying on general patterns in the data, instead of specific pixels.

Early stopping based on validation accuracy helped prevent extra epochs when the model had no potential of improvement, saving time and helping performance.

5. Models that have been pretrained (i.e., trained with supervised learning on large datasets for different modeling tasks) like MobileNetV2 are effective and efficient alternatives to using a pretrained generic model for small or medium dataset sizes.

Pretrained models are advantageous because they strike the right speed and accuracy balance, while also considering computing and hardware efficiency. Alternatively, it is possible to build a neural network from scratch, which would depend on a large dataset (or a composite of large datasets mirroring the problem) and an unlikely hyperparameter tuning with relatively good performance to achieve the performance/accuracy benchmark. Knowing some of the limitations associated with pretrained models, pretrained models will settle for an effective plug-and-play basis giving nearly state-of-the-art accuracy performance benchmarks with significantly less data and computational cost.