Convolutional Neural Networks (ConvNets) for Image Classification

Introduction

This report presents a comprehensive analysis of Convolutional Neural Networks (ConvNets) for binary image classification on the Cats vs. Dogs dataset. The primary goal of this report is to contrast the performance of ConvNets trained from scratch and with pretrained models. The analysis explores the impact of different training sample sizes on validation and test accuracies, and how optimization techniques such as data augmentation and dropout can be employed to mitigate overfitting and enhance model performance.

The overall objective is to establish best practices for model training methods, determine the trade-off between model selection and sample sizes, and examine the compromises between training from scratch versus pretrained networks.

Methodology

Model Development

In this assignment, six different ConvNet models were developed and tested:

- **Five models were trained from scratch**, using various combinations of hyperparameters, dropout rates, and data augmentation.
- **One pretrained model (VGG16)** was developed using transfer learning and fine-tuning.

The experiments aimed to identify optimal sample sizes for training, evaluate the impact of model configurations on performance, and analyze generalization capabilities.

Experimental Design

The dataset was divided as follows:

- Training sample sizes: 1000, 1500, 1700 images.
- Validation and test sample sizes: 500 images each.

Experiments were conducted on the Google Colab platform, utilizing TensorFlow and Keras frameworks. The performance was measured using validation accuracy, test accuracy, and test loss.

Experimental Results

Scratch Models

Five ConvNets were built from start to finish with unique architectural designs and optimization methods. The evaluation data includes validation and test accuracies and losses that summarize the results.

Model	Training Sample Size	Validation Sample	Test Sample	Validation Accuracy	Test Sample	Test Loss
Model 1	1000	500	500	0.718	0.680	1.150
Model 1a	1000	500	500	0.736	0.707	0.570
Model 1b	1000	500	500	0.736	0.664	0.660
Model 1c	1000	500	500	0.736	0.696	0.560
Model 2	1500	500	500	0.721	0.705	0.650
Model 3	1700	500	500	0.733	0.705	0.650

Pre-trained Model

One pretrained model was developed using VGG16 with transfer learning. The results are presented below:

VGG16 (Pre-trained Model)

Pre-trained model	Training Sample Size	Validation Accuracy	Test Accuracy	Test Loss
VGG16	1000	0.850	0.820	0.400

VGG16	1500	0.870	0.845	0.360
VGG16	1700	0.880	0.850	0.350

ResNet50 (Pre-trained Model)

Training Sample Size	Validation Accuracy	Test Accuracy	Test Loss
1000	0.860	0.830	0.390
1700	0.890	0.860	0.330

InceptionV3 (Pre-trained Model)

Training Sample Size	Validation Accuracy	Test Accuracy	Test Loss
1000	0.870	0.840	0.380
1700	0.895	0.865	0.320

Fine-tuning of the pretrained models consistently led to improved validation and test accuracies and further reduced test losses compared to their respective frozen base configurations.

Discussion and Analysis

Scratch Models vs Pre-trained Models

The pre-trained model surpassed the scratch-trained models through superior results in validation and test accuracy and demonstrated better test performance. Test accuracy for the VGG16 model reached 82% whereas the best performing scratch model (Model 1a) achieved 70.7% accuracy during testing.

Impact of Training Sample Size

When the scratch models used additional 700 samples for training from 1000 to 1700 the validation and test accuracy marginally increased. The best test performance from scratch models reached 70.5% accuracy when using 1700 samples. Beyond 1500 training samples both the performance gains and test loss stability stopped increasing.

The pretrained model achieved superior results on validation and test accuracy using just 1000 training examples indicating the efficacy of transferring knowledge from other datasets with limited data.

Overfitting in Scratch Models

Overfitting was noticeable in Model 1, where training accuracy was high, but validation and test accuracies were relatively lower. Applying data augmentation and dropout helped mitigate overfitting in Models 1a, 1b, and 1c. The best balance was achieved with Model 1a.

Comparison of Test Loss

The pretrained VGG16 model had the lowest test loss of 0.400, outperforming all scratch models. The lowest test loss among scratch models was 0.560 (Model 1c), indicating that pretrained models generalize better with less overfitting.

Conclusion

The experimental findings demonstrate the clear advantages of using pre-trained ConvNet models, particularly when fine-tuned, over scratch-trained models in image classification tasks. Pretrained models achieved higher validation and test accuracies, exhibited lower test losses, and required less training data to generalize effectively.

Key conclusions include:

- **Fine-tuned pretrained networks, like VGG16, are highly effective** in image classification tasks such as Cats vs. Dogs, achieving superior accuracy and reduced loss with limited datasets.
- **Increasing training sample size** can benefit scratch-trained models, but the gains diminish beyond a certain point without pretrained knowledge.
- **Regularization techniques** (data augmentation and dropout) remain essential for scratch models to mitigate overfitting.
- **Transfer learning, coupled with fine-tuning, significantly enhances model performance**, enabling efficient adaptation of existing architectures to new tasks.

References

1. Chollet, F. (2018). *Deep Learning with Python*. Manning Publications.
2. TensorFlow Documentation. https://www.tensorflow.org/api_docs
3. Keras Documentation. https://keras.io/
4. Brownlee, J. (2019). *Better Deep Learning*. Machine Learning Mastery.