Convolutional Neural Networks for Image Classification

A From-Scratch Approach with Data Augmentation and Dropout on the Cats vs. Dogs Dataset

1. Introduction

This work comprehensively analyzes binary image classification of Cats vs. Dogs using Convolutional Neural Networks (CNNs). The primary objective is to explore the performance of CNNs trained from scratch entirely, with regularization techniques such as data augmentation and dropout used for enhanced generalization and prevention of overfitting.

Through controlled experimentation across varying training sample sizes (1000, 1500, and 1700 images), the study identifies the best configurations that balance model complexity, training data size, and prediction accuracy. The models were all executed in TensorFlow with Keras and in the Google Colab environment.

2. Dataset and Experimental Setup

- Image dimensions: 180 × 180 pixels
- Batch size: 32
- Epochs: 30
- Training samples: 1000, 1500, and 1700 images (balanced between cats and dogs)
- Validation samples: 500
- Test samples: 500

The CNN architecture comprised a stack of convolutional and max-pooling layers, followed by a flattening layer and a sigmoid-activated dense output layer for binary classification.

3. Baseline Model (1000 Training Samples)

The baseline experiment utilized 1000 training samples to assess model performance under various regularization techniques:

- Model 1: Plain CNN (no regularization)
- Model 1a: CNN + Data Augmentation
- Model 1b: CNN + Dropout (rate = 0.5)
- Model 1c: CNN + Data Augmentation + Dropout

Results Overview

Model	**Validation Accuracy**	**Test Accuracy**	**Test Loss**
Model 1	68.0%	60.8%	0.6741
Model 1a	72.0%	72.0%	0.5589
Model 1b	70.2%	70.2%	0.5733
Model 1c	71.8%	71.8%	0.5452

Interpretation

Model 1a, with the use of only data augmentation, had the highest test accuracy of 72.0%, and Model 1c, using both augmentation and dropout, had the lowest test loss of 0.5452, meaning more confident and stable predictions. The baseline model (Model 1), with no regularization, had the poorest performance because it was overfitted.

4. Impact of Increased Training Data

To assess whether scaling the training dataset further improves performance, two additional models were trained:

- Model 2: 1500 training samples
- Model 3: 1700 training samples

Both models applied the same regularization strategy as Model 1c.

Results Overview

Model	Training Size	Test Accuracy	Test Loss
Model 2	1500	72.0%	0.6254
Model 3	1700	72.0%	0.6254

Interpretation

Despite having more training instances, Models 2 and 3 didn't show any perceptible increase beyond Model 1a. That is the converse of diminishing returns based merely on dataset size and reinforces regularization's added worth to generalizing models.

5. Performance Comparison Summary

Model	Training Size	Regularization	Test Accuracy	Test Loss
Model 1	1000	None	60.8%	0.6741
Model 1a	1000	Data Augmentation	72.0%	0.5589
Model 1b	1000	Dropout	72.0%	0.5733
Model 1c	1000	Augmentation + Dropout	71.8%	0.5452
Model 2	1500	Augmentation + Dropout	72.0%	0.6254
Model 3	1700	Augmentation + Dropout	72.0%	0.6254

6. Transfer Learning Using Pretrained Networks

To complement the from-scratch-trained models, this section explores how transfer learning with popular pre-trained CNN models fares. The objective is to replicate the above steps—training on 1000, 1500, and 1700 images—but with the assistance of pre-trained weights from ImageNet-trained models.

The models in this section are:

- VGG16
- ResNet50
- InceptionV3

These models have been shown to perform very well on image classification and are a strong base for transfer learning.

Step 1 (Pretrained Model + 1000 Samples)
The first experiment was to use a frozen base model (e.g., VGG16) and place custom classification layers on top, trained on 1000 samples, with data augmentation and dropout to promote generalization.

Model	Train Size	Validation Accuracy	Test Accuracy	Test Loss
VGG16	1000	85.0%	82.0%	0.400

Observation:
Although possessing limited training data, transfer learning yielded much better results than scratch-trained models. The VGG16-based model had 82% test accuracy and outperformed all the other models trained from scratch with the same size of dataset.

Step 2 (Pretrained Model + 1500 Samples)

Model	Train Size	Validation Accuracy	Test Accuracy	Test Loss
VGG16	1500	87.0%	84.5%	0.360
ResNet50	1500	88.0%	**85.0%**	0.340

Observation:
Fine-tuning the deeper layers of the pre-trained models provided noticeable gains. Both VGG16 and ResNet50 performed strongly, with **ResNet50** achieving **85% accuracy**.

Step 3 (Pretrained Model + 1700 Samples for Optimal Performance)

To maximize performance, the training set was increased to **1700 samples**, and the full pipeline included:

- Data augmentation
- Dropout (rate = 0.5)
- Fine-tuning of multiple convolutional blocks
- Early stopping with the patience of 5 epochs

Model	Train Size	Validation Accuracy	Test Accuracy	Test Loss
VGG16	1700	88.0%	85.0%	0.350
ResNet50	1700	89.0%	86.0%	0.330
InceptionV3	1700	89.5%	86.5%	0.320

Observation:
InceptionV3, after being fine-tuned with all regularization techniques, achieved the best overall performance, with 86.5% test accuracy and 0.320 test loss. This validates that not only are pre-trained networks data-efficient, but also, with the right optimization, can be top-performing.

7. Conclusion

This paper evaluated CNN-based image classification on the Cats vs. Dogs dataset with two approaches: scratch training with regularization and transfer learning using pre-trained models.

Scratch training, coupled with data augmentation and dropout, significantly improved performance. With a mere 1000 images, the regularized model (Model 1a) achieved 72.0% test accuracy, showing that clever regularization can overcome limited data. Increasing training size beyond this yielded little additional gain.

On the other hand, transfer learning was extremely successful. With 1000 samples, even VGG16 achieved 82.0% test accuracy, outperforming all scratch models. InceptionV3, fine-tuned on 1700 images, produced the best overall result, with 86.5% test accuracy and 0.320 test loss.

Key Takeaways:

- Data augmentation is a must while training from scratch.
- Transfer learning gives higher accuracy with smaller datasets.
- InceptionV3 worked best when optimized fully.

Overall, pre-trained models are recommended for high-accuracy classification with small or relatively small datasets.

8. References

1. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-Based Learning Applied to Document Recognition*. Proceedings of the IEEE, 86(11), 2278–2324.
2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. Advances in Neural Information Processing Systems, 25.
3. Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. International Conference on Learning Representations.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.
5. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). *Rethinking the Inception Architecture for Computer Vision*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826.