Convolutional Neural Networks (ConvNets) for Image Classification

Cats vs. Dogs – From Scratch with Augmentation and Dropout

Introduction

This report presents a detailed evaluation of the application of Convolutional Neural Networks (ConvNets) for classifying binary images from the Cats vs. Dogs dataset. The goal is to examine the performance of several architectural strategies and training enhancements such as data augmentation, dropout regularization, and increased dataset size.

The main objective is to train and develop CNN models from the ground up and identify configurations that yield the best validation and test accuracy without overfitting. The task doesn't include using pre-trained models but how well the custom models can generalize with effective yet simple optimization techniques.

Methodology

Model Architecture and Configurations

Six CNN models were developed using TensorFlow and Keras:

- **Model 1**: Baseline architecture trained from scratch
- **Model 1a**: Model 1 + data augmentation
- **Model 1b**: Model 1 + dropout
- **Model 1c**: Model 1 + data augmentation + dropout
- **Model 2**: Same architecture with increased training data (1500 samples)
- **Model 3**: Same architecture with 1700 training samples

Each model consisted of stacked Conv2D and max-pooling layers, followed by flattening and a final dense layer with sigmoid activation for binary classification.

Dataset and Training Strategy

- **Image size**: 180 × 180 pixels
- **Batch size**: 32
- **Epochs**: 30
- **Training sizes**: 1000, 1500, and 1700 images

- **Validation set**: 500 images
- **Test set**: 500 images

All experiments were conducted in Google Colab.

Baseline Training from Scratch (Sample Size = 1000)

In the initial experiment, the model was trained from scratch with:

- Training set: 1000 images (500 cats, 500 dogs)
- Validation set: 500 images
- Test set: 500 images

Regularization Strategy

To mitigate overfitting, three models were explored:

- Model 1: Baseline CNN (no regularization)
- Model 1a: Data augmentation applied
- Model 1b: Dropout applied
- Model 1c: Combined augmentation and dropout

Results

Model	Val Accuracy	Test Accuracy	Test Loss
Model 1	71.8%	68.0%	1.150
Model 1a	73.6%	**70.7**	0.570
Model 1b	73.6%	66.4%	0.660
Model 1c	73.6%	69.6%	0.560

Training on 1000 samples from the start, Model 1a with data augmentation achieved the highest, at 70.7% test accuracy. Model 1c was also doing great with reduced test loss (0.560), indicating high confidence in predictions.

Increasing Training Sample Size (1500–1700 Images)

To evaluate the effect of larger training datasets, experiments were repeated with:

- Model 2: 1500 training images
- Model 3: 1700 training images
- Validation and test sets were unchanged (500 each)

Results

Model	Train Size	Val Accuracy	Test Accuracy	Test Loss
Model 2	1500	72.1%	70.5%	0.650
Model 3	1700	73.3%	70.5%	0.650

Increasing the training sample to 1700 provided some validation accuracy increase, but test accuracy was identical (70.5%) to the smaller-augmented models in Step 1. This suggests regularization contributed more to performance than the absolute amount of data at this stage.

Optimizing Sample Size for Best Performance

Based on the experiments, data augmentation + dropout combination (Model 1c) with 1000 samples trained yielded positive results with fewer test losses. Model 1a achieved the best test accuracy (70.7%) even though it did not have additional data.

The **ideal training sample size was 1000** when paired with effective regularization (augmentation). This smaller dataset with better training techniques outperformed larger sets without them.

Transfer Learning Using Pretrained Models

The same steps were repeated using **pre-trained ConvNet architectures** (VGG16, ResNet50, InceptionV3) with and without fine-tuning

Results (Pretrained Networks)

Model	Train Size	Val Accuracy	Test Accuracy	Test Loss
VGG16	1000	85.0%	82.0%	0.400
VGG16	1500	87.0%	84.5%	0.360
VGG16	1700	88.0%	85.0%	0.350
ResNet50	1700	89.0%	86.0%	0.330
InceptionV3	1700	89.5%	86.5%	0.320

Transfer learning greatly improved performance. VGG16 achieved 82% accuracy on the test set using only 1000 samples, beating all of the from-scratch models. Scaling up to 1700 samples increased accuracy even more to 85%–86.5%, and reduced test loss to 0.32.

Conclusion

This project showcases the trade-offs between training from scratch and transfer learning in image classification:

- From-scratch models can be ~70% accurate when well regularized, even with 1000 images.
- Bigger training sizes more than 1500 provide decreasing returns unless augmented with regularization.
- Transfer learning beats from-scratch models by a huge margin, with >85% accuracy on less data and lower test loss.
- Best overall performance was obtained using InceptionV3 trained on 1700 images, at 86.5% test accuracy and 0.32 loss.

References

1. Chollet, F. (2018). *Deep Learning with Python* (2nd ed.)
2. TensorFlow Developers. (2023). *TensorFlow API Documentation*.
3. Keras Documentation. (2023). *Image Classification*.
4. Brownlee, J. (2019). *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery.

5. Russakovsky, O., Deng, J., Su, H., et al. (2015). *ImageNet Large Scale Visual Recognition Challenge*. *International Journal of Computer Vision.*