Evaluating the Impact of Training Sample Size on Model Selection for Cats and Dogs Image Classification

1. Introduction

In image classification tasks, the choice of network architecture and training strategy significantly influences performance. This study focuses on classifying images of cats and dogs, using two main approaches:

1. Training a Convolutional Neural Network (CNN) from scratch.
2. Utilizing transfer learning with the VGG16 pretrained model, fine-tuning it for our specific classification task.

Our goal is to assess how training sample size impacts the effectiveness of each model, as well as to determine the optimal sample size and network choice for achieving high classification accuracy.

2. Methodology

2.1. Dataset Preparation

The dataset consists of labeled images of cats and dogs, divided into training, validation, and test sets. To evaluate the effect of training sample size, we used four different training sizes:

- 600 images
- 1200 images
- 3000 images
- 6000 images

Data augmentation techniques such as rotation, zoom, and horizontal flips were applied to the training images to increase variability, helping prevent overfitting and improve generalization.

2.2. Approach

The study was conducted in two stages:

- **Stage 1: Training a CNN from Scratch** We constructed a CNN with convolutional, max-pooling, and dropout layers, training it from scratch. This model's accuracy was tested across different training sizes to determine how sample size affects its learning capacity.
- **Stage 2: Transfer Learning with VGG16** VGG16, a pretrained CNN on the ImageNet dataset, was employed as a base model for transfer learning. Additional dense layers were added for classification, and fine-tuning was applied by unfreezing some VGG16 layers to adapt it for the specific task of cats and dogs classification.

To further optimize, we applied a learning rate scheduler to progressively reduce the learning rate, thereby stabilizing training and improving convergence.

3. Results

3.1. CNN Trained from Scratch

The CNN trained from scratch showed a clear correlation between training sample size and model accuracy:

- **600 images**: Test accuracy of 68.0%
- **1200 images**: Test accuracy of 69.8%
- **3000 images**: Test accuracy of 72.0%
- **6000 images**: Test accuracy of 80.2%

From these results, we observed that the CNN required a larger dataset to reach higher accuracy levels. The model showed incremental improvements with increased data, highlighting its dependency on sample size for effective learning.

3.2. Transfer Learning with VGG16

The VGG16 pretrained model demonstrated a significant performance boost, even with smaller sample sizes:

- **600 images**: Test accuracy of 88.4%
- **1200 images**: Test accuracy of 90.0%
- **3000 images**: Test accuracy of 94.2%
- **6000 images**: Test accuracy of 93.0%

The pretrained model achieved high accuracy even with fewer images (600 or 1200), suggesting that transfer learning with a pretrained network is highly effective for tasks with limited data. Interestingly, the highest test accuracy (94.2%) was achieved with 3000 images, with performance plateauing as the sample size increased further.

4. Analysis and Key Findings

4.1. Effect of Training Sample Size on Model Performance

- **CNN from Scratch**: The CNN required a substantial amount of training data to reach reasonable accuracy levels. As the training sample size increased, the model showed steady improvement, demonstrating that training from scratch benefits more from larger datasets. With smaller datasets, the model struggled to generalize effectively.
- **Transfer Learning with VGG16**: The pretrained model showed impressive performance across all sample sizes. VGG16's initial weights, trained on ImageNet, provided a strong foundation for feature extraction. Even with fewer training samples, the pretrained model captured essential features for the cats and dogs classification task, reaching high accuracy quickly.

4.2. Comparison of Model Choices Across Sample Sizes

- For smaller datasets (600 and 1200 images), the VGG16 pretrained model significantly outperformed the CNN trained from scratch. This indicates that transfer

learning is more suitable for tasks with limited data availability, as it leverages learned features from larger datasets.
- As the dataset size increased to 3000 images, the CNN's performance improved but remained lower than that of the pretrained model. The VGG16 model reached its peak performance (94.2% accuracy) with 3000 images, after which adding more data showed diminishing returns.
- For the largest sample size (6000 images), the CNN reached 80.2% accuracy, while the VGG16 model showed a slight drop in accuracy (93.0%) compared to the peak at 3000 images. This suggests that adding more data beyond a certain point may not necessarily enhance performance for the pretrained model, potentially due to overfitting or the saturation of learned features.

5. The Relationship between Sample Size and Model Choice

The results illustrate a clear relationship between training sample size and optimal model choice:

- **With Smaller Sample Sizes (<3000 images)**: The pretrained VGG16 model is a more effective choice. Its transfer learning capabilities allow it to perform well with limited data, leveraging pre-learned features to achieve high accuracy without extensive training. For datasets with fewer than 3000 images, transfer learning offers a substantial performance advantage over training from scratch.
- **With Larger Sample Sizes (>=3000 images)**: While the CNN model's accuracy improved with more data, it still lagged behind the pretrained VGG16. The VGG16 model achieved its highest accuracy at 3000 images, indicating that this sample size is likely optimal for fine-tuning pretrained networks on the cats-and-dogs task. Beyond this sample size, the pretrained model's performance gains were minimal, suggesting that, for larger datasets, it may be beneficial to use other architectures or regularization techniques to prevent overfitting.

6. Conclusion and Recommendations

The findings emphasize the importance of sample size in determining the appropriate model architecture and training strategy:

1. **Transfer Learning is Ideal for Smaller Datasets**: When dealing with smaller datasets, pretrained networks like VGG16 offer a significant advantage by achieving high accuracy with limited training samples. The pretrained model achieved optimal performance with only 3000 images, suggesting that transfer learning is highly suitable for tasks with constrained data resources.
2. **CNNs from Scratch Require Larger Datasets**: Training a CNN from scratch is feasible for larger datasets, but it demands substantial data (at least 6000 images in our case) to approach the performance of a pretrained model. For smaller datasets, training from scratch may not be the best choice, as it struggles to generalize effectively.
3. **Optimal Training Sample Size for Transfer Learning**: The results indicate that for the VGG16 model, a sample size of 3000 images was optimal, as it provided the highest accuracy. Beyond this point, performance plateaued, and even slightly

decreased, indicating that adding more data may not be beneficial for all pretrained models.

7. Future Directions

To further improve classification accuracy and generalization, future studies could explore:

- **Alternative Pretrained Models**: Testing architectures like ResNet or EfficientNet may offer further insights into transfer learning's effectiveness for this task.
- **Ensemble Models**: Combining predictions from multiple models (e.g., CNN trained from scratch and pretrained models) could enhance accuracy.
- **Hyperparameter Tuning**: More extensive hyperparameter tuning, including regularization techniques, could mitigate overfitting, especially for the larger dataset sizes.