**Project Objective:**

To construct and evaluate convolutional neural network (CNN) models—both trained from scratch and utilize pretrained architectures—for the task of classifying images as either cats or dogs, using Kaggle's Dogs vs Cats dataset.

**Problem Statement**

To categorize 2000 images (with an equal number of cats and dogs) using CNN-based deep learning methods, tackling overfitting and optimizing accuracy with limited and augmented data.

**Key Objectives**

Main Aims Conceive CNNs that can efficiently differentiate between pictures of cats and dogs.  
  
Utilize data augmentation to enhance generalization.  
  
Compare models trained from the ground up versus those using transfer learning with pretrained networks (e.g., VGG16).

**Techniques and Implementation:  
i) Dataset Preparation**  
The Dogs vs Cats dataset was prepared into balanced training, validation, and test datasets of varying sample sizes, with equal representation from both classes.  
  
**ii) Model Architecture**  
Custom CNNs were created comprising multiple convolutional and pooling layers ending in fully connected layers for binary classification. Architectures were optimized for input size as well as performance.  
  
**iii) Data Augmentation**  
Methods like rotation, flipping, and zooming were applied to artificially augment the size of the training dataset to promote generalization and fight overfitting.

**iv) Pre-trained Models**  
Transfer learning with pre-trained models like VGG16 was utilized to leverage powerful feature extractors from ImageNet to gain improved accuracy even with small training sets.

**Results and Analysis:   
1**. **Model Performance:**

The evaluation of the models was conducted under various configurations, utilizing both CNNs trained from scratch and pretrained networks like VGG16. Each configuration was evaluated using validation accuracy and loss across different training sample sizes (ranging from 500 to 2000).

Models that were trained from the ground up exhibited moderate performance when using smaller datasets, according to initial results. When data augmentation and architectural improvements (such as padding and deeper layers) were implemented, there was considerable accuracy.  
Model-8, trained on 2000 images, achieved the best accuracy from scratch (82%).  
In small datasets, overfitting was common . However, it could be reduced by increasing the sample size and employing dropout/padding techniques.

Across all training sizes, the performance of pretrained models was significantly superior.  
Pretrained models achieved an accuracy of over 80% even with just 2,000 training samples.  
Model-2, which was trained on images and fine-tuned, attained the highest accuracy of 81%.  
Loss and accuracy tables clearly show that pretrained models, particularly when fine-tuned and augmented, are superior.

**2. Effect of Data Augmentation:**

Data augmentation played a critical role in enhancing the generalization and robustness of scratch-trained and pre-trained

models. Techniques such as random zooming, rotation, and flipping added controlled variations to the training images, reducing the likelihood of overfitting.

• In scratch models, the accuracy increased from 67% to 81% only by incorporating data augmentation and architecture tuning.

• Augmentation even improved generalization in pretrained models, especially in fine-tuned networks, with sub-1% loss.

Qualitative comparison of original images and augmented images showed more diversity without altering the semantics of label vital for model learning improvement.

**3.** **Comparison with Pre-trained Models:**

Comparison between scratch-trained CNNs and pretrained VGG16 models

Scratch-trained models:

Require greater data (2000 samples) and deeper networks to achieve reasonable accuracy (81%).

Tend to overfit when the sample size is small and the regularization is poor.

Pretrained models:

Achieved superior performance even with limited data.

Fine-tuning the pretrained layers with frozen base model produced excellent performance (up to 81 % accuracy). With the ability to save training time and speed up convergence due to feature learning transfer.

**Key Insights from Scratch Models:**

Overfitting was observed in early models that were trained with 1000 images, showing lower validation and test accuracy.

Data augmentation (zoom, rotation, random flip) greatly contributed to performance boost and reduction in overfitting.

Increase of training sample size from 1000 to 2000 enhanced accuracy from 70% to 77.6%, producing the benefit of greater data.

Padding in later models (Model-9) facilitated improvement to accuracy to 81%, confirming its importance in preservation of spatial data.

Subsidiary architecture of five convolutional layers and optimized hyperparameters (filter sizes and stride) provided the optimal accuracy of 83.4% (Model-7), which proved the capability of a well-structured CNN architecture.

A table with numbers and text

AI-generated content may be incorrect.

Conclusion

Overall, the Model-3 optimized using a pre-trained VGG16 network from 2000 training images is remarkable at achieving the highest accuracy of 81%, demonstrating the impact that increases in dataset size combined with transfer learning can have on performance. This model could be sustained by layer freezing and data augmentation, allowing it to prevent overfitting but at the same time promote effective fine-tuning for the classification task. Whereas previous models trained from scratch had decent results, the pretrained models, particularly with fine-tuning, significantly outperformed them. These results highlight the importance of taking advantage of pretrained networks and adapting training methods according to dataset size and task difficulty.

A graph with numbers and letters

AI-generated content may be incorrect.