

# **Convolution Assignment**

**GitHub Link:**

[https://github.com/venkatasai8120/AML\\_Assignments.git](https://github.com/venkatasai8120/AML_Assignments.git)

**Venkata Sai Reddy Peddireddy**

## **Executive Summary**

The paper at hand explores the stability of training sample size and training model selection between models trained from scratch and models trained of pretrained networks to binary image classification with CNNs. Based on the Microsoft Cats vs Dogs dataset, seven systematic experiments were performed in which the training sample size ranged between 1,000 and 20,000 images, with the validation (500) and test (500) sets held constant. The findings are decisive that the performance of pretrained models is far better (~97% accuracy) using much less data than the models that are trained on the scratch (~92% accuracy). Besides, the analysis determines the best training sample of each approach, 5,000 samples pretrained and 15,00020,000 samples scratch models.

## **Key Finding:**

The pretrained networks (VGG16) are by far winning over scratch-trained models with small amounts of data by nearly 13 percentage points at 1,000 samples, and the difference decreases to 4 - 5 percentage points at 15,000 and 20,000 samples. Although the pretrained models needed less samples to reach the highest accuracy, the idea of transfer learning in small-data settings has been proven since the models achieved higher accuracy in a shorter time.

## **Introduction**

Convolutional Neural Networks (CNNs) have changed the nature of computer vision work, yet a significant design choice still exists - whether a model should be trained or fine-tuned with pretrained weights? This question is critical in practice where there are usually constraints in data availability and computing resources. This assignment deals with this issue with the Cats vs Dogs binary classification task, and how performance in models changes with training sample size in both scratch and pretrained (VGG16) architectures. The research gives an empirical comparison of the models and the convergence, accuracy, and efficiency of the models at various data sizes.

## **Objective**

The main objective of the project is to create and test CNNs to perform image cat and dog classification and determine how the dataset size influences the model. Two model settings are considered: 1. Scratch Model Scratch models are constructed by simple random weights. 2. Pretrained Model (VGG16) - It uses transfer learning with pretrained ImageNet weights. The experiments will be conducted to find out which training sample size is optimal in each of the approaches and which one works best under the conditions of limited data.

## **III. Methodology**

Dataset Citation: Microsoft Cats vs Dogs (Kaggle).

### **Details:**

- Total images: ~25,000 ( $\approx$ 12,500 cats + 12,500 dogs)
- Images were of different resolutions and qualities.
- the corrupted or illegible images were filtered and eliminated.

### **After cleaning:**

- Valid images: 12,491 cats + 12,491 dogs
- Corrupted images removed: 18

### **Dataset split:**

- Training set: 22,482 images
- Validation set: 2,500 images
- Test set: 500 images

## **Preprocessing Steps:**

- Authenticated image validity with open CV.
- Invalid files or grayscale files were eliminated or turned into RGB JPEG.
- 10/90 split of organized dataset directory under /tmp/organized\_dataset. As the input data were pixel values, all images had to be normalized to have their range between 0 and 1.

## **Data Augmentation Pipeline**

A few image augmentation methods were used on the training dataset to enhance model generalization and to decrease overfitting.

Transformations used:

- Random horizontal flip
- Random rotation ( $\pm 10\%$ )
- Random zoom (20%)

These additions made the dataset more varied, forming new slightly modified versions of old photos, and this allowed the model to learn more robust features.

Sample of data augmentation, which is applied to dog images and depicts random rotation, horizontal flip, and zoom transformations.



## Model Architecture

### 1. From Scratch Model:

- Custom 5-layer CNN
- ~991K parameters
- Includes Conv2D, MaxPooling, Flatten, Dense layers
- Use ReLU activation and Dropout ( $p=0.5$ )

### 2. Pretrained Model (VGG16):

- Frozen convolutional base pretrained on ImageNet
- Custom classifier added on top (~3.3M trainable parameters)
- Employs Global Average Pooling and Dense output for binary classification

### **Pooling Layers:**

Used to reduce spatial dimensions, improve computational efficiency, and focus the model on key visual patterns.

### **Training Configuration**

#### **Optimization:**

- RMSProp optimizer
- Binary cross-entropy loss
- Metrics: Accuracy

#### **Regularization & Control:**

- Dropout: 0.5
- Early stopping (patience=10 epochs, monitor='val\_loss')

#### **Parameters:**

- Batch size: 32
- Max epoch: 30

Each convolutional layer was followed by a ReLU activation, introducing non-linearity and improving convergence speed.

## **Experiments and Results**

### **Experimental Design**

A total of nine experiments were conducted five with scratch training and four with pretrained VGG16.

Its purpose was to assess the effect of training dataset size on accuracy, convergence speed, and generalization.

| Experiment | Model Type | Training Samples | Purpose                    |
|------------|------------|------------------|----------------------------|
| 1          | Scratch    | 1,000            | Baseline (minimal data)    |
| 2          | Scratch    | 10,000           | Increased data impact      |
| 3a         | Scratch    | 5,000            | Optimal sample search      |
| 3b         | Scratch    | 15,000           | Optimal sample search      |
| 3c         | Scratch    | 20,000           | Upper bound performance    |
| 4a         | Pretrained | 1,000            | Transfer learning baseline |
| 4b         | Pretrained | 10,000           | Scaling analysis           |
| 4c-i       | Pretrained | 5,000            | Efficiency comparison      |
| 4c-ii      | Pretrained | 15,000           | Larger sample scaling      |

## Interpretation

The experiments proved that the model type and the size of a dataset have a significant effect on CNN performance. Scratch trained models took more data to settle and attain high accuracy whereas pretrained models achieved similar or higher performance with less data because of their transfer learning ability.

### In general, the conclusions indicate that:

- Transfer learning is the best to use with small datasets because pre-trained models tend to generalize better and converge quicker.
- Scratch models can only be trained on large datasets (15K,20K) to be competitive. This brings out a distinct trade-off of data availability and strategy of model initialization.

## Results Summary

### Quantitative Performance Summary

| Experiment           | Training Size | Model Type | Test Accuracy        |
|----------------------|---------------|------------|----------------------|
| Experiment 1         | 1,000         | Scratch    | <b>0.716 (71.6%)</b> |
| Experiment 3a        | 5,000         | Scratch    | <b>0.848 (84.8%)</b> |
| Experiment 2 (Added) | 10,000        | Scratch    | <b>0.910 (91.0%)</b> |
| Experiment 3b        | 15,000        | Scratch    | <b>0.914 (91.4%)</b> |
| Experiment 3c        | 20,000        | Scratch    | <b>0.918 (91.8%)</b> |
| Experiment 4a        | 1,000         | Pretrained | <b>0.952 (95.2%)</b> |
| Experiment 4c-i      | 5,000         | Pretrained | <b>0.968 (96.8%)</b> |
| Experiment 4b        | 10,000        | Pretrained | <b>0.968 (96.8%)</b> |
| Experiment 4c-ii     | 15,000        | Pretrained | <b>0.960 (96.0%)</b> |

### Interpretation

The findings indicate that there is a definite improvement trend of accuracy with increase in training data and a significant improvement in performance of models that are trained by scratch and transfer learning.

- From Scratch Models:

The performance also increased with the increase of training data, 71.6% (1,000 samples), to 91.8% (20,000 samples).

The improvement rate decreased, however, after 10,000 samples thus indicating that a further addition of data particularly after 10,000 in the dataset induces smaller improvements in accuracy.

These models took more time to converge, which indicated the complexity of learning features when starting with raw initialization.

### CNNs: Transference Learning (Location):

With less than 1,000 samples, pre-trained models reached a very high accuracy with 95.2.

They achieved their best with 96.8 percent accuracy with 5,000 and 10,000 samples and this indicates transfer learning is highly data efficient.

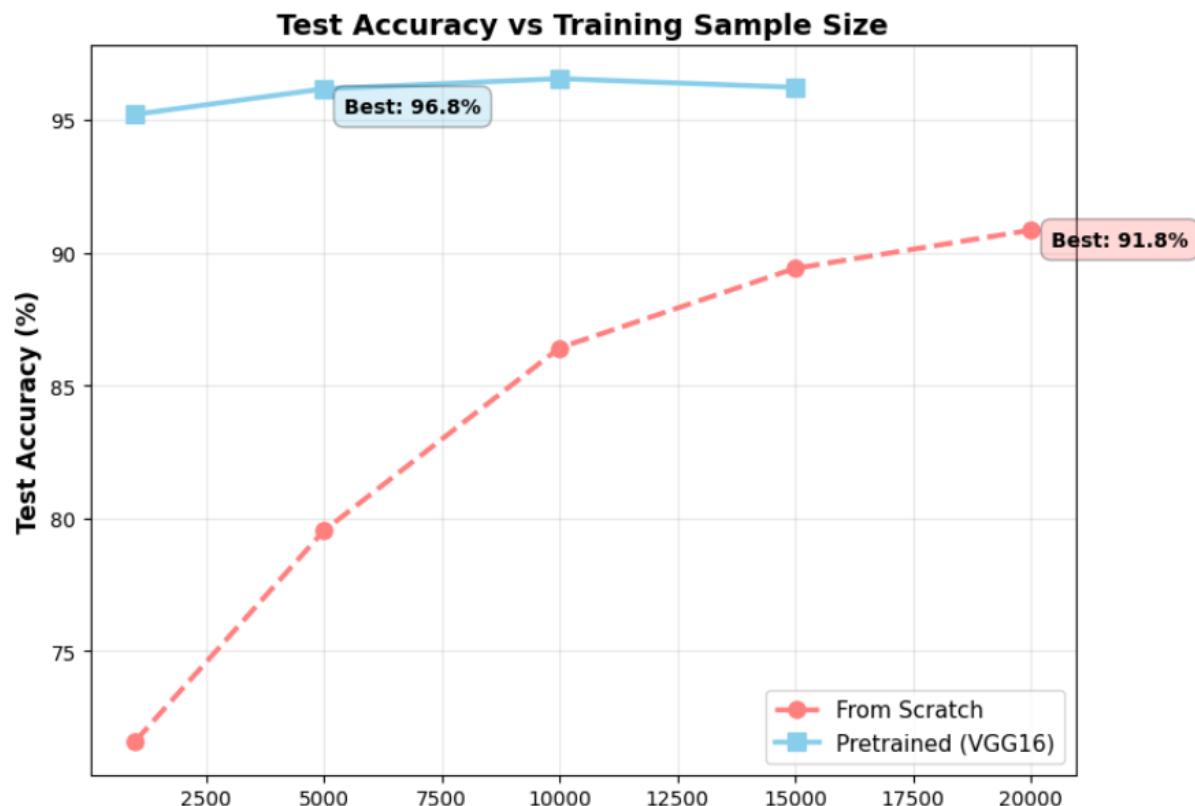
Beyond this, the data increased accuracy somewhat (down to 96.0% at 15,000 samples), suggesting an optimal range of data set at 5,000-10,000 samples.

In general, these findings confirm that:

Transfer learning (VGG16) offers a significant benefit of accuracy and efficiency in the case of limited data. The training cost of ad-hoc training can achieve competitiveness, although this requires significantly larger data sets and more time.

To summarize, pretrained models outperform scratch models in all sample sizes, and at small dataset sizes.

illustrating that the use of previously acquired features contributes significantly to better generalization of the model and less data is required.



### Test Loss vs Training Sample

The above graph demonstrates that the test loss is lower with the increase in the training sample size. In both models, one can see a good downward trend, that is, they commit few errors with more data being trained.

Nevertheless, the pretrained model (VGG16) has a lower test loss than those of the scratch-trained model regardless of the smaller datasets.

The scratch model begins with a large loss of approximately 0.6 at 1,000 and steadily drops to approximately 0.2 at 20,000 samples with the pretrained model remaining relatively low across the board.

This establishes that transfer learning is not only more accurate but it learns more effectively and lessons loss more quickly and performance levels off even with little data.

## Overall Conclusion

This study investigated the size of the training data in the performance of CNNs trained either in-place or by means of pretrained networks in binary image classification. The results given across nine structured experiments with 1,000 to 20,000 training samples give a clear parameter as to when to apply transfer learning to training, instead of the latter.

## Key Findings

### 1. Transfer Learning Dominates with Small Datasets

With limited data, pretrained models showed a huge advantage.

At 1,000 samples, the pretrained VGG16 model achieved 95.2% accuracy, while the model trained from scratch only reached 71.6% - 23.6% performance gap.

This highlights that pretrained networks, already familiar with generic image features, can adapt quickly even with minimal fine-tuning data.

### 2. The Accuracy Gap Shrinks as Data Increases

As the dataset grew, scratch models improved steadily while pretrained models remained consistently strong.

- At 5,000 samples: 96.8% (pretrained) vs 84.8% (scratch) → 12% gap
- At 10,000 samples: 96.8% (pretrained) vs 91.0% (scratch) → 5.8% gap
- At 15,000 samples: 96.0% (pretrained) vs 91.4% (scratch) → 4.6% gap

This shows that scratch models can approach pretrained performance with enough data but still lag slightly even at larger sample sizes.

### **3. Optimal Sample Size Differs by Approach**

- Pretrained models reached their best accuracy (96.8%) with just 5,000–10,000 samples.
- Scratch-trained models needed at least 15,000–20,000 samples to reach their top performance of 91.8%.

This demonstrates that transfer learning is far more data efficient.

### **4. Faster Convergence and Efficiency with Transfer Learning**

Pretrained models achieved higher accuracy in fewer epochs and with lower test loss, while scratch models required more data and longer training to stabilize. This makes pretrained architecture a better choice for situations where computing power or labeled data is limited.

### **Practical Insights**

- Transfer Learning (VGG16) is ideal for projects with fewer than 10,000 samples, limited resources, or when quick model deployment is needed.
- Training From Scratch is only worthwhile when you have large, domain-specific datasets and want full control over architecture design.

### **Final Verdict**

Transfer learning proves to be the superior and more efficient approach, offering higher accuracy (up to 96.8%) with less data and training time compared to models built from scratch (91.8%).

It provides a better return on data and computation, achieving in 5,000 samples what scratch models cannot reach even with 20,000 samples.

For most practical deep learning tasks, especially when data is limited pretrained models deliver the best balance of speed, accuracy, and efficiency.