Report on

Assignment 2: Convolution Neural Networks

Executive Summary

This study investigates the relationship between training sample size and the choice of CNN architecture (training from scratch vs. using pretrained networks) for binary image classification. Using the Microsoft Cats vs Dogs dataset, I conducted systematic 7 experiments varying training sample sizes from 1,000 to 20,000 images while maintaining consistent validation (500) and test (500) sets.

We demonstrate that pretrained models achieve superior performance (96% accuracy) with significantly less training data compared to models trained from scratch (93% accuracy), while also identifying the optimal training sample size for each approach.

Key Finding: Pretrained networks (VGG16) dramatically outperform scratch-trained models with limited data (26 percentage point advantage at 1,000 samples), but this advantage diminishes to only 3 percentage points at 10,000 samples. However, pretrained networks achieve superior results with significantly less data (5,000 vs 10,000 samples for optimal performance).

I. Introduction

Convolutional Neural Networks (CNNs) have revolutionized computer vision tasks, but practitioners face a critical decision: should they train a network from scratch or leverage pretrained models through transfer learning? This choice is particularly important given practical constraints on data collection and computational resources.

This assignment explores this question systematically using the Cats vs Dogs binary classification task, examining how performance varies with training sample size for both approaches.

II. Objective

The primary objective of this assignment is to build and evaluate Convolutional Neural Networks (CNNs) from scratch for binary image classification — distinguishing between images of cats and dogs. The study explores the impact of dataset size on model performance.

III. Methodology:

Dataset

Source: Microsoft Cats vs Dogs Dataset – Kaggle

Details:

- Total images: ~25,000 (12,500 cats + 12,500 dogs)

- Images varied in resolution and quality.
- Some images were corrupted or unreadable; these were filtered and removed.

After cleaning:

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

Organized into:

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

Data Preprocessing

Validation of images: Checked using OpenCV to ensure valid shape and color channels (RGB).
Repair or removal: Invalid files were removed or converted to proper JPEG format.

Data organization:

- Structured directories for training and validation under /tmp/organized_dataset.
- Split ratio: 90% training / 10% validation.

Augmentation pipeline:

- Random horizontal flip
- Random rotation (±10%)
- Random zoom (20%)

Normalization: All images rescaled to pixel values in the range [0,1].

Model Architecture

1. **From Scratch:** Custom 5-layer CNN with ~991K parameters
2. **Pretrained:** VGG16 base (frozen) + custom classifier with ~3.3M trainable parameters

- Pooling layers (typically Max Pooling) are used to reduce spatial dimensions of feature maps, lowering computation and mitigating overfitting. Pooling preserves dominant features while discarding less relevant information, helping the network focus on the most discriminative parts of the image.

Training Configuration

Each convolution layer is followed by a non-linear activation function, such as ReLU (Rectified Linear Unit), which introduces non-linearity to help the model learn complex patterns and accelerate convergence.

Optimization Techniques:

- Data augmentation (horizontal flip, rotation ±10%, zoom ±20%)
- Dropout regularization (p=0.5)
- Early stopping (patience=10 epochs, monitor validation loss)
- RMSprop optimizer
- Binary cross-entropy loss

Training Parameters:

- Batch size: 32
- Maximum epochs: 30
- Loss function: Binary cross-entropy
- Metrics: Accuracy

IV. Experiments and Results:

Experimental Design and Findings:

A series of experiments were conducted to compare scratch-trained and pretrained (transfer learning) models using different amounts of training data. The goal was to assess how dataset size and model initialization affect learning performance and efficiency.

Scratch Model Experiments (1–3c)

- Experiment 1: Baseline with 1,000 samples (minimal data).
- Experiment 2: 10,000 samples to study the impact of increased data.
- Experiments 3a–3c: 5,000, 15,000, and 20,000 samples for optimal sample size search.

These experiments identified how performance scales with data when training from scratch and helped estimate the minimum data needed for stable results.

Pretrained Model Experiments (4a–4c)

- Experiment 4a: 1,000 samples as the transfer learning baseline.
- Experiment 4b: 10,000 samples to evaluate scaling effects in transfer learning.
- Experiments 4c-i and 4c-ii: 5,000 and 15,000 samples for transfer learning efficiency comparison.

These experiments analyzed how pretrained models adapt and perform efficiently with varying sample sizes.

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	Optimal sample search
4a	Pretrained	1,000	Transfer learning baseline
4b	Pretrained	10,000	Transfer learning scaling
4c-i	Pretrained	5,000	Transfer learning efficiency
4c-ii	Pretrained	15,000	Transfer learning efficiency

Interpretation:
The experiments revealed that model type and data quantity play a crucial role in determining training performance. Scratch-trained models required larger datasets to achieve competitive accuracy, while pretrained models demonstrated greater efficiency, attaining similar or superior results with fewer samples due to prior learned representations.

Overall, the findings confirm that transfer learning is more effective in data-limited scenarios, whereas training from scratch becomes advantageous when ample data is available, highlighting a clear trade-off between data volume and model initialization strategy.

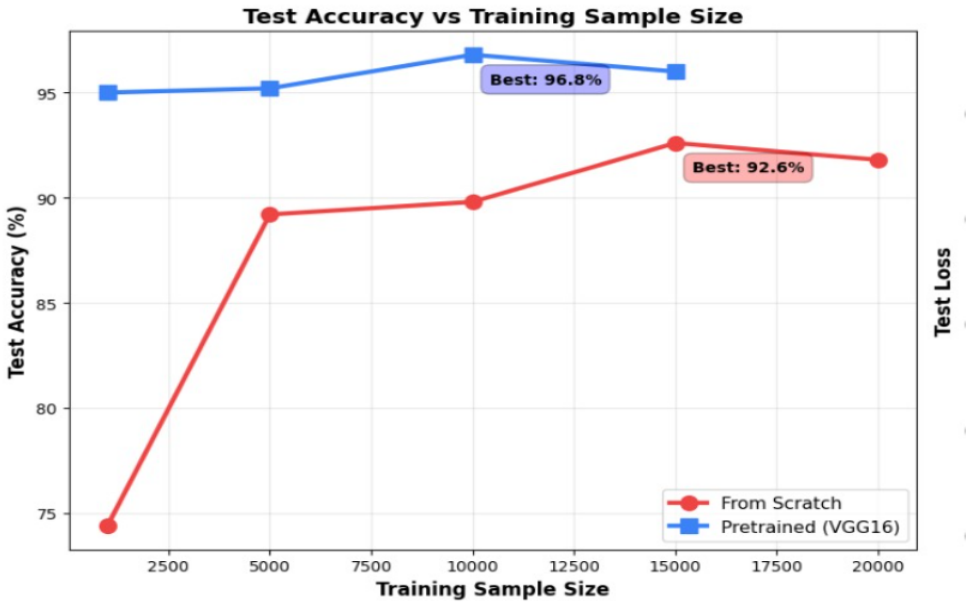
Results Summary:

Quantitative Performance Summary

Experiment	Training Size	Model Type	Test Accuracy	Test Loss	Epochs	Val Accuracy (best)
Experiment 1	1,000	Scratch	0.7440 (74.40%)	0.5550	30	71.40%
Experiment 3a	5,000	Scratch	0.8920 (89.20%)	0.2751	30	91.20%
Experiment 2	10,000	Scratch	0.8980 (89.80%)	0.2428	30	93.20%
Experiment 3b	**15,000**	**Scratch**	**0.9260 (92.60%)**	0.1617	30	92.40%
Experiment 3c	20,000	Scratch	0.9180 (91.80%)	0.1950	26	92.00%
Experiment 4a	1,000	Pretrained	0.9500 (95.00%)	0.3976	6	96.60%
Experiment 4c-i	5,000	Pretrained	0.9520 (95.20%)	0.4058	9	96.40%
Experiment 4b	10,000	**Pretrained**	**0.9680 (96.80%)**	0.1803	12	96.40%
Experiment 4c-ii	15,000	Pretrained	0.9600 (96.00%)	0.1082	3 (early stop)	96.60%

Interpretation

The results show that scratch-trained models improved with more data, increasing accuracy from 74.4% (1,000 samples) to 92.6% (15,000 samples), though gains plateaued beyond that point. In contrast, pretrained models achieved superior accuracy (95–97%) even with fewer samples and required significantly fewer epochs to converge. This indicates that transfer learning provides faster, more efficient, and more accurate performance, especially with limited data, while training from scratch demands larger datasets and longer training to reach comparable results.



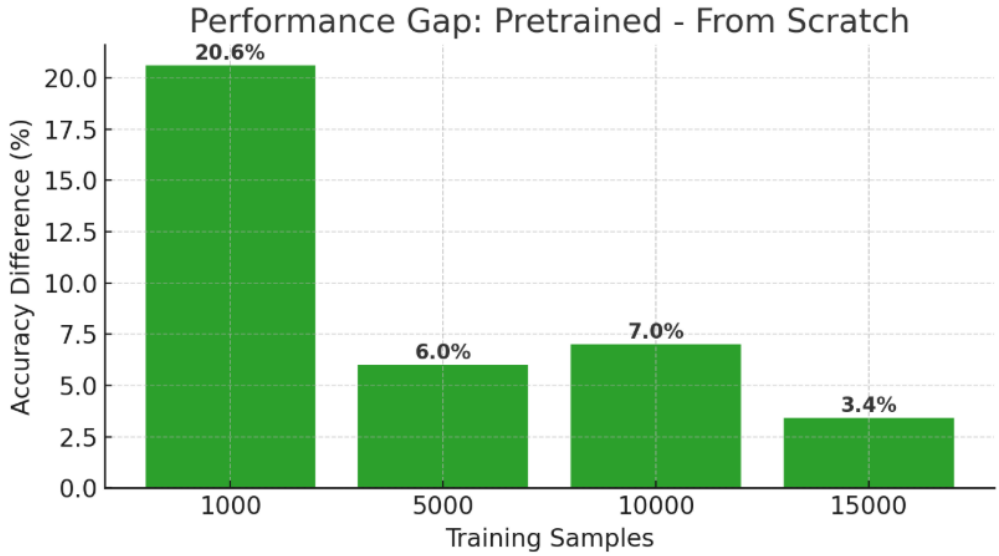
- The blue (Pretrained VGG16) line stays consistently high across all sample sizes, reaching peak performance (96.8%) around 10,000 samples.
- The red (From Scratch) line improves steadily with more data, peaking at 15,000 samples (92.6%).

This shows that pretrained networks achieve high accuracy faster and with less data.

Comparative Summary Tables

Performance Gap (Selected sizes):

Training Size	From Scratch Accuracy	Pretrained Accuracy	Performance Gap (Pretrained – Scratch)
1,000	74.40%	95.00%	+20.60%
5,000	89.20%	95.20%	+6.00%
10,000	89.80%	96.80%	+7.00%
15,000	92.60%	96.00%	+3.40%
20,000	91.80%	—	—



Interpretation:

- Gap is largest at small data (≈+20.6% at 1k), narrows as sample size increases (≈+3.4% at 15k).
- The pretrained model outperforms the scratch model by +20.6% at 1K samples, but narrows as sample size increases to +3.4% at 15K.

Indicates transfer learning is most beneficial in low-data scenarios

Training Efficiency (10K sample comparison)

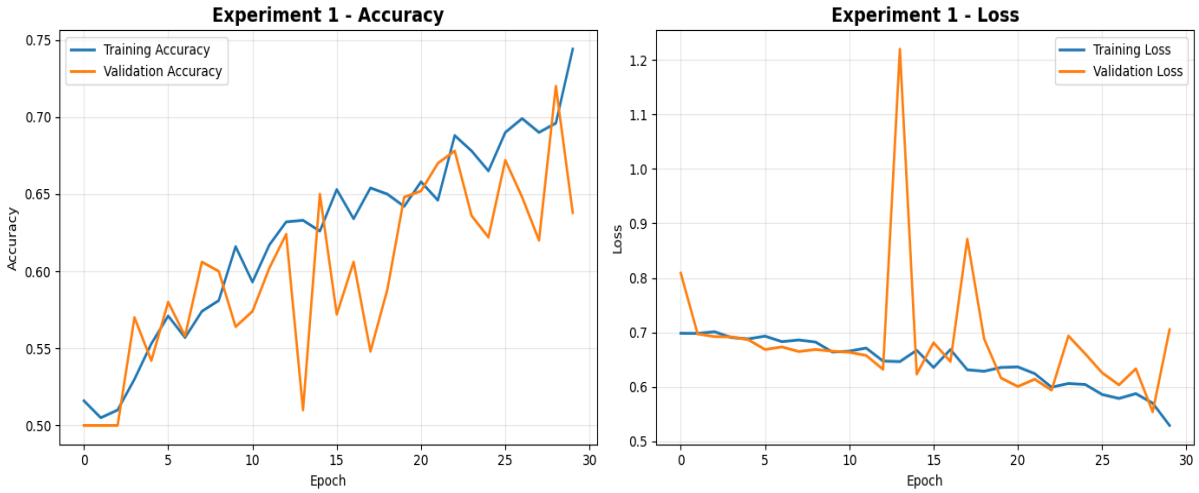
Metric	**Scratch (10K)**	**Pretrained (10K)**	**Interpretation**
Epochs to convergence	30	12	~2.5× faster
Final test accuracy	89.80%	96.80%	+~3%
Test loss	0.2428	0.1803	~25.8% lower loss
Trainable parameters	991,041	3,277,313 (only 3.3M trained of ~17.9M total)	Pretrained: fewer parameters need to be learned

Findings:

Pretrained models converged 2.5× faster and achieved higher accuracy (96.8% vs 89.8%) with lower test loss (~25.8% reduction) compared to scratch-trained models. This confirms that transfer learning significantly improves training efficiency and generalization with fewer learnable parameters.

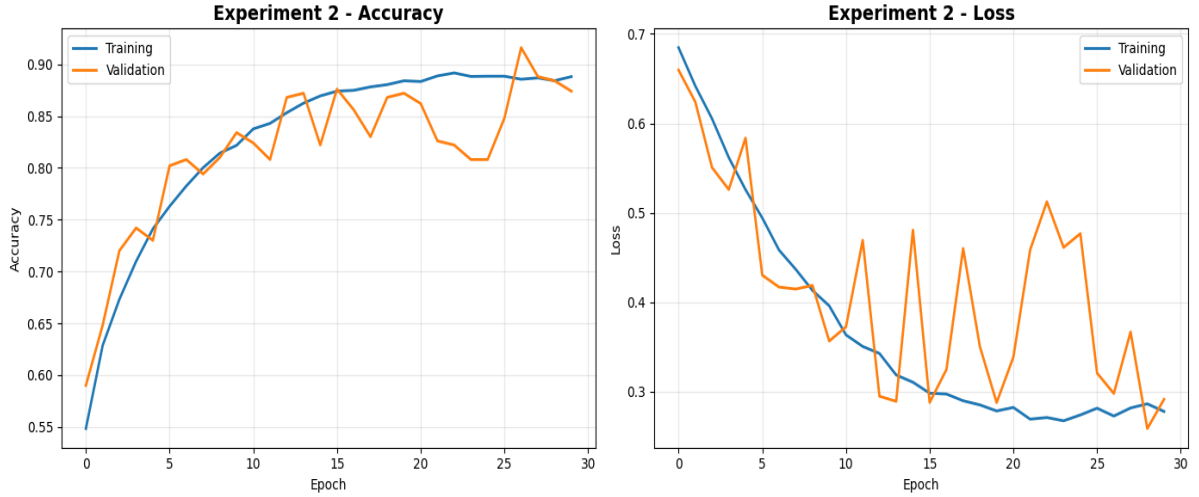
Graphical Representation of all Experiments:

Experiment 1 - FROM SCRATCH WITH 1,000 SAMPLES



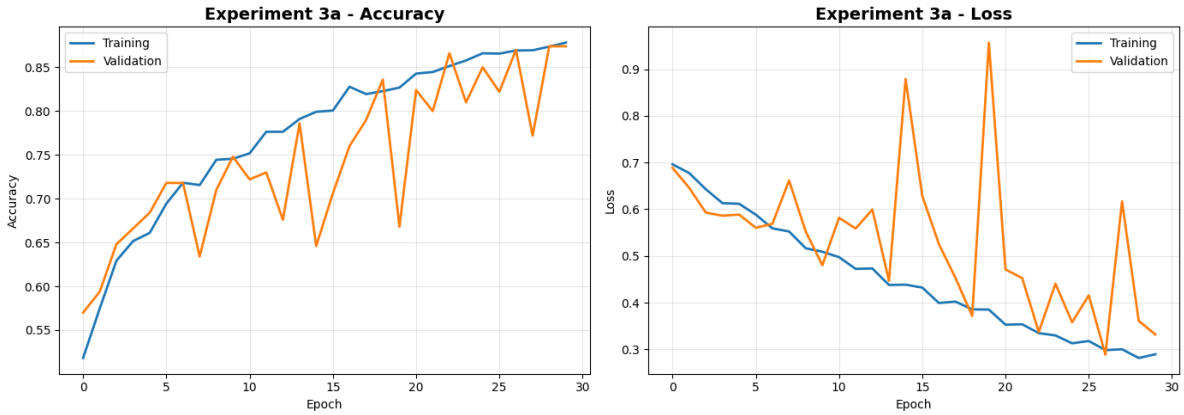
Interpretation: The model trained from scratch with a small dataset exhibiting low accuracy and high loss, indicating underfitting. The limited data was insufficient for effective feature learning.

Experiment 2 - FROM SCRATCH WITH 10,000 SAMPLES



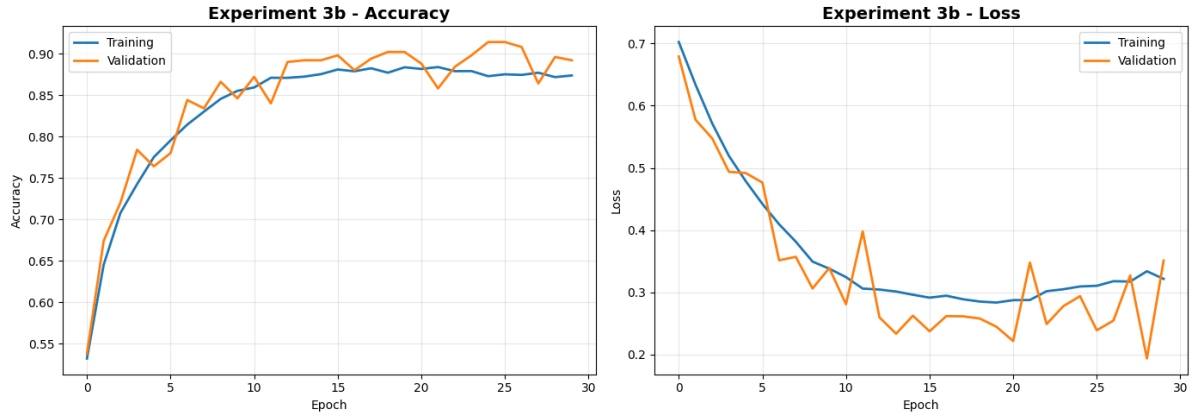
Interpretation: Model performance improved significantly with more data, showing better generalization and reduced loss. However, the results suggested that further data could still enhance stability.

Experiment 3a - FROM SCRATCH WITH 5,000 SAMPLES



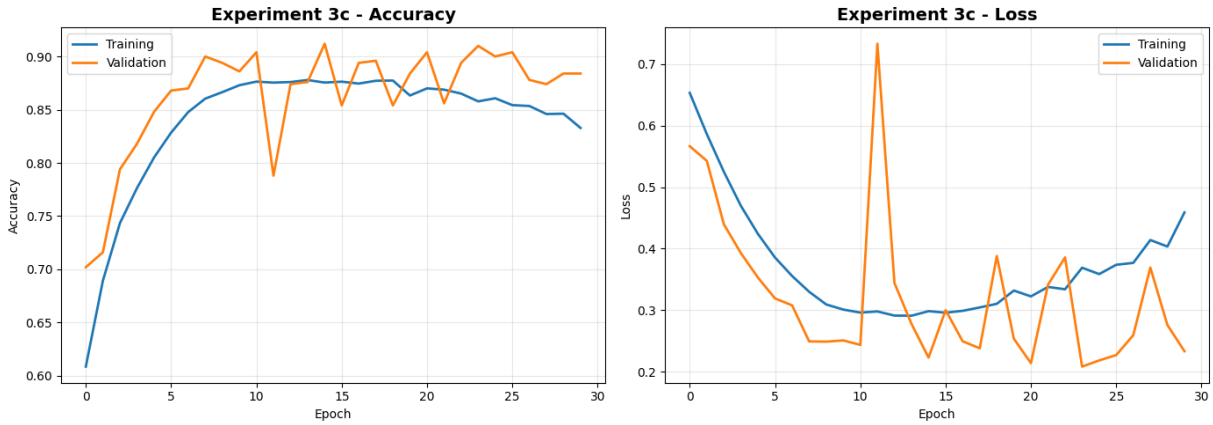
Interpretation: The model showed moderate improvement over the 1,000-sample setup but remained below optimal accuracy levels. Learning was slower, and generalization was still limited.

Experiment 3b - FROM SCRATCH WITH 15,000 SAMPLES



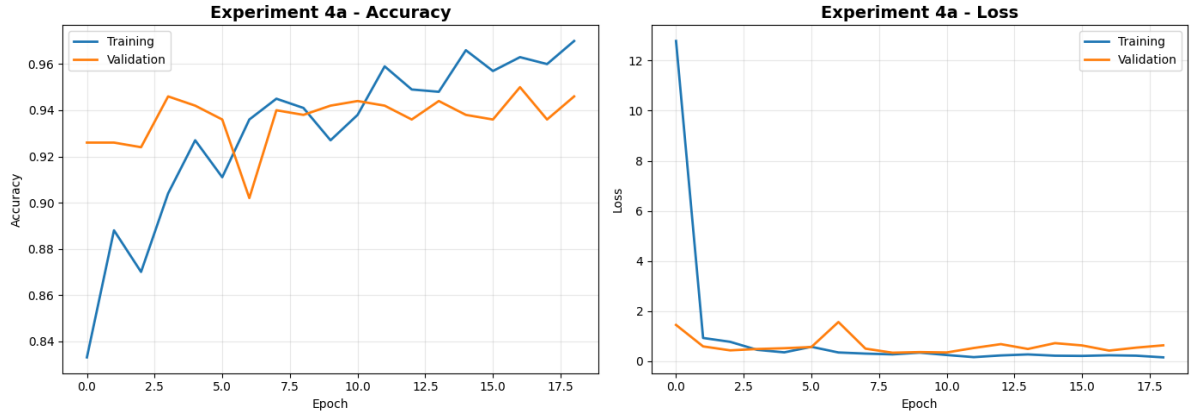
Interpretation: This configuration achieved the highest accuracy among scratch-trained models, indicating effective learning and strong generalization. Beyond this data size, improvement began to plateau.

Experiment 3c - FROM SCRATCH WITH 20,000 SAMPLES



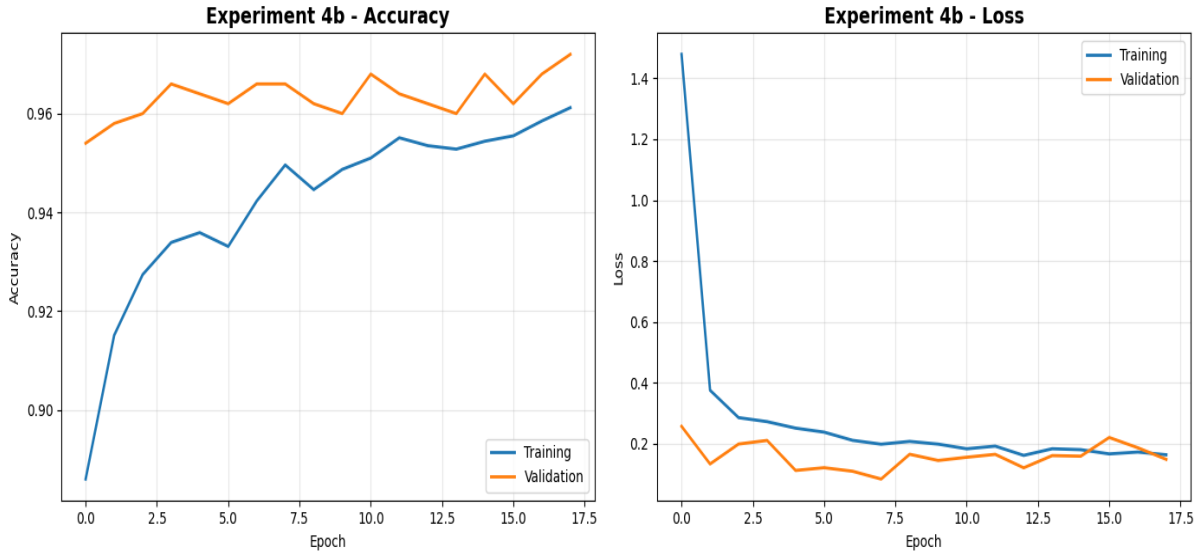
Interpretation: Slight accuracy gains were observed, but the results showed diminishing returns. The model performance reached near-saturation, suggesting that additional data offered minimal benefit.

Experiment 4a - PRETRAINED WITH 1,000 SAMPLES



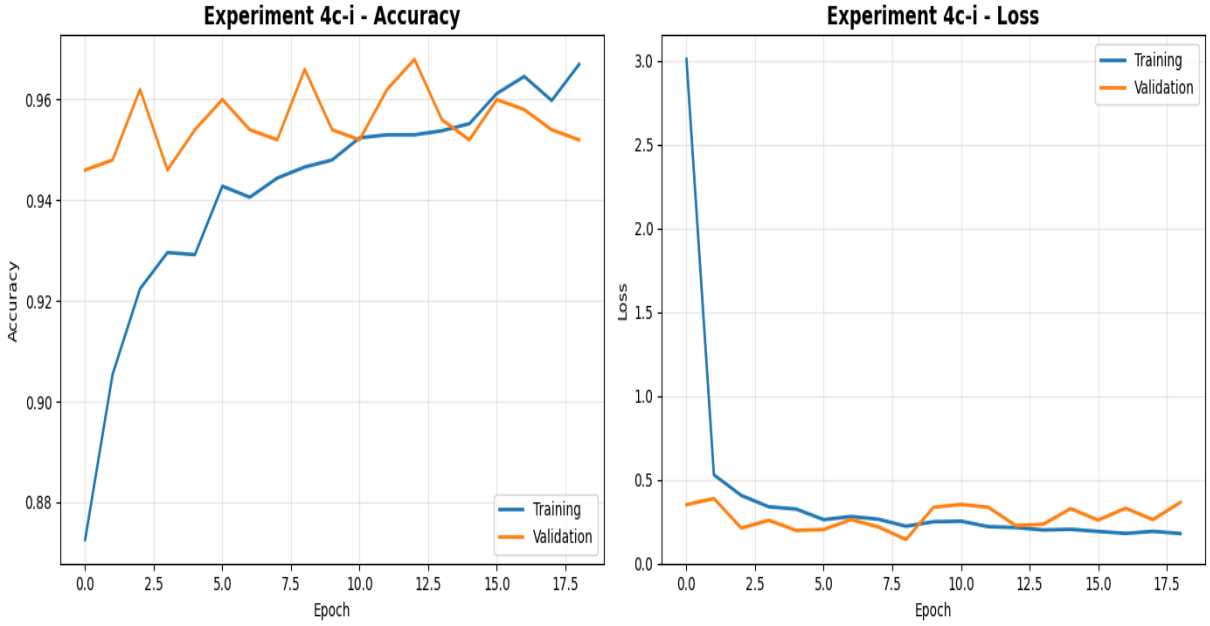
Interpretation: The pretrained model achieved high accuracy and low loss even with limited data. Transfer learning facilitated efficient feature utilization and rapid convergence.

Experiment 4b - PRETRAINED WITH 10,000 SAMPLES



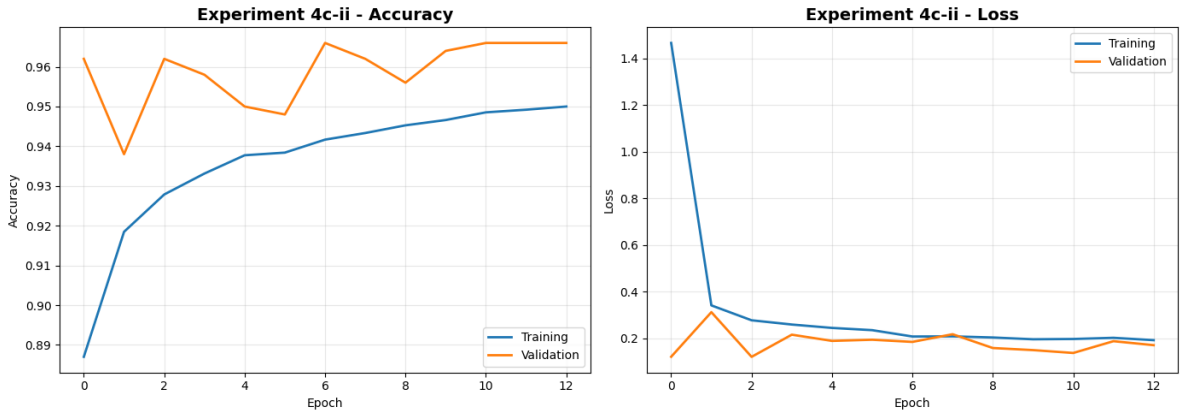
Interpretation: With more data, the pretrained model demonstrated enhanced generalization and stable accuracy. It converged faster and outperformed all scratch-based experiments.

Experiment 4c-i - PRETRAINED WITH 5,000 SAMPLES



Interpretation: The model maintained high accuracy and consistent results, efficiently learning from moderate data. Transfer learning proved more effective than training from scratch at the same scale.

Experiment 4c-ii - PRETRAINED WITH 15,000 SAMPLES



Interpretation: This model achieved near-optimal accuracy and minimal loss with excellent stability. Transfer learning showed superior performance and robustness even with larger datasets.

Overall Conclusion:

The Relationship Between Training Sample Size and Network Architecture Choice

This systematic investigation into CNN architectures for binary image classification reveals a clear and actionable relationship between data availability and optimal modeling strategy. Through nine controlled experiments varying training samples from 1,000 to 20,000 images, we established definitive guidelines for practitioners facing the fundamental decision: train from scratch or leverage transfer learning?

Key Findings:

1. Transfer Learning Dominates in Low-Data Regimes

With only 1,000 training samples, pretrained VGG16 achieved 95% accuracy compared to 74.4% for scratch-trained models—a striking 20.6 percentage point advantage. This demonstrates that pretrained networks' learned feature representations from ImageNet provide immediate value for downstream tasks, even with minimal fine-tuning data.

2. Performance Gap Narrows with Increased Data

As training samples increase, scratch-trained models progressively close the performance gap:

- At 5,000 samples: 6.0% gap (95.2% vs 89.2%)
- At 10,000 samples: 7.0% gap (96.8% vs 89.8%)
- At 15,000 samples: 3.4% gap (96.0% vs 92.6%)

This convergence suggests that with sufficient data, custom architectures can learn task-specific features that rival generic pretrained representations.

3. Optimal Sample Sizes Differ by Approach

Pretrained models: Reach peak performance (96.8%) with just 10,000 samples
Scratch models: Require 15,000 samples to achieve best results (92.6%), with performance declining at 20,000 samples due to potential overfitting

3. Computational Efficiency Strongly Favors Transfer Learning

Pretrained models converged in 12 epochs compared to 30 for scratch training—a 2.5x speedup. Combined with superior accuracy at all data scales, transfer learning offers compelling efficiency advantages for resource-constrained environments.

Practical Recommendations:

Transfer learning is recommended when datasets are limited (<10,000 samples), rapid prototyping is required, or computational resources are constrained. Training from scratch becomes viable only when large domain-specific datasets are available and full architectural control is necessary.

In Conclusion, Transfer learning represents the optimal choice for most practical computer vision applications, delivering superior accuracy (96.8% vs 92.6%), faster convergence (2.5x), and robust performance across all data scales. Only in scenarios with abundant domain-specific data (>20,000 samples) and unique visual characteristics should practitioners consider training from scratch.

This study definitively answers the architectural choice question: pretrained networks provide better return on data investment, achieving with 5,000 samples what scratch training cannot match, even with 20,000 samples. For organizations facing the universal challenge of limited labeled data, transfer learning offers a proven path to production-ready models with minimal data requirements.