Assignment 3

Deep Learning for Computer Vision

1. Objective

The goal of this assignment was to apply **convolutional neural networks (CNNs)** to a real world image classification task and compare two approaches:

1. Training a CNN **from scratch**
2. Using a **pre trained network (transfer learning)** and optionally fine-tuning it

The dataset used is **Microsoft Cats vs Dogs**, obtained from Kaggle.
The assignment explores how model performance changes as the training-set size increases and how transfer learning can greatly improve accuracy and efficiency.

2. Methodology

Data Preparation

- Downloaded and extracted the dataset using Kaggle API.
- Removed corrupted and grayscale images using the Python PIL library.
- Converted all images to RGB and resized them to **224×224 pixels**.
- Created balanced splits for:
  - Training (varied by experiment)
  - Validation (250 per class)
  - Test (250 per class)
- Performed **data augmentation** using random flip, rotation, zoom, and contrast

Model 1 – CNN from Scratch

- 3 convolutional blocks with Batch Normalization and MaxPooling
- Global Average Pooling + Dropout (0.25)
- Sigmoid output for binary classification
- Optimizer: Adam (lr=1e-3)
- Loss: Binary Cross-Entropy with label smoothing (0.05)
- Regularization: L2 weight decay (1e-4) and Dropout
- Early Stopping and ReduceLROnPlateau callbacks

Model 2 – VGG16 Pre trained (Transfer Learning)

- Loaded **VGG16** without top layers, with ImageNet weights
- Added:
    1. Global Average Pooling
    2. Dropout (0.2)
    3. Dense(1, Sigmoid)
- Two phases:
    1. **Transfer only:** Base frozen, train only top layers
    2. **Fine-tune:** Unfrozen from block4_conv1 to adapt to cats/dogs
- Optimizer: Adam (1e-4 for transfer, 1e-5 for fine-tune)

3. Results

Model 1 – From Scratch

Training per Class	Total Train	Best Validation Accuracy	Test Accuracy
500	**1000**	**0.50**	**0.50**
2000	**4000**	**0.80**	**0.81**
4000	**8000**	**0.83**	**0.82**

Model 2 – VGG16 (Transfer Learning & Fine-tuning)

Run	Phase	Train per Class	Total Train	Best Val Acc	Test Acc
vgg_transfer_tpc500	Transfer Only	500	1,000	0.888	0.902
vgg_transfer_tpc2000	Transfer Only	2,000	4,000	0.954	0.966
vgg_transfer_tpc4000	Transfer Only	4,000	8,000	0.984	0.978
vgg_finetune_tpc500	Fine-tuned	500	1,000	0.980	0.972
vgg_finetune_tpc2000	Fine-tuned	2,000	4,000	0.974	0.988
vgg_finetune_tpc4000	Fine-tuned	4,000	8,000	0.986	0.992

Graphical Trend

Observation: Both models improve with larger datasets, but transfer learning outperforms the scratch CNN at every data size.

Training Images	Scratch CNN Test Acc	VGG16 Transfer Test Acc	VGG16 Fine-tune Test Acc
1000	0.50	0.90	0.97
4000	0.81	0.97	0.99
8000	0.82	0.98	0.99

Trend Summary:

- Accuracy increases with data size for both models.
- Transfer learning achieves higher accuracy even with limited data.
- Fine-tuning further improves performance.

---

4. Relationship Between Training Sample Size and Network Choice

1. **Training Sample Size:**
    - When the dataset is small (e.g., 1,000 total images), a CNN trained from scratch struggles to learn meaningful features.
    - The validation and test accuracy remain close to random (≈ 0.5).
2. **Choice of Network:**

- Pre-trained models like **VGG16** already know general visual features from large datasets such as ImageNet.
- Even with few training images, transfer learning can achieve high accuracy because the base layers already encode edges, shapes, and textures.
- Fine-tuning allows these features to adapt to the specific cat/dog domain, giving near-perfect accuracy.

3. **Conclusion:**
   - **From Scratch:** Needs thousands of samples to perform well.
   - **Transfer Learning:** Performs well even with small datasets.
   - **Fine-tuning:** Best overall performance and generalization.

---

5. Discussion and Insights

Aspect	From Scratch	Transfer Learning
Data Required	Large (8k+)	Small (~1k) works well
Training Time	Long	Short
Accuracy	~82 %	97–99 %
Generalization	Moderate	Excellent

Overfitting Risk	High on small data	Low, due to pretrained weights

Key Insight:
Transfer learning is both **data-efficient** and **performance-efficient**, especially when labeled data is limited.
As training sample size grows, both methods improve, but transfer learning consistently outperforms training from scratch.

---

6. Conclusion

This assignment demonstrated the effectiveness of transfer learning in image classification tasks.

- A CNN trained from scratch achieved **~82 %** accuracy with 8,000 images.
- A VGG16 transfer-learning model achieved **~99 %** accuracy even with smaller datasets.
- Increasing the training sample size consistently improved performance for both models.
- Fine-tuning gave the highest performance by adapting pretrained features to the new dataset.

Final takeaway:

> Transfer learning provides a significant advantage when dealing with limited training data and should be the preferred approach for most practical image classification problem.

7. Submission Items

1. cats_dogs_from_scratch.ipynb
2. cats_dogs_pretrained.ipynb

3. dogs_vs_cats_scratch_results_v2.csv
4. dogs_vs_cats_pretrained_results.csv
5. Assignment3_Report.pdf